ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

«ΣΥΣΤΗΜΑΤΑ ΑΥΤΟΜΑΤΙΣΜΟΥ»

Μεταπτυχιακή Εργασία

Σύγκριση μοντέλων μηχανικής μάθησης για την παρακολούθηση διαγραμμάτων ελέγχου ποιότητας

Ζιάκας Παναγιώτης

Επιβλέπων Καθηγητής: Πανώριος Μπενάρδος

ΑΘΗΝΑ 2021

Περίληψη

Ο στατιστικός έλεγχος ποιότητας είναι μία μέθοδος που αποσκοπεί στην βελτίωση της ποιότητας των παραγόμενων προϊόντων. Τα διαγράμματα ελέγχου ποιότητας, χρησιμοποιούνται για τον στατιστικό έλεγχο της παραγωγικής διαδικασίας, υποδεικνύοντας αν αυτή εξελίσσεται ομαλά. Η εμφάνιση τάσεων, δηλαδή μη τυχαίων χρονοσειρών στα διαγράμματα ελέγχου, παραπέμπει στην επίδραση συστηματικών αιτιών που προκαλούν μεταβλητότητα, επιφέροντας μείωση της ποιότητας και καθιστώντας απαραίτητη την πραγματοποίηση διορθωτικών κινήσεων. Συνεπώς η αναγνώριση τάσεων σε διαγράμματα ελέγχου ποιότητας αποτελεί σημαντικό κομμάτι του στατιστικού ελέγχου ποιότητας. Τα τελευταία χρόνια ποικίλες μέθοδοι έχουν εφαρμοστεί για την αυτοματοποίηση αυτής της διαδικασίας. Μία από αυτές, με όλο και μεγαλύτερη απήχηση είναι τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ). Η παρούσα εργασία αφορά την αναγνώριση τάσεων σε διαγράμματα ελέγχου ποιότητας με την χρήση ΤΝΔ. Επιπλέον, προχωρά στην εφαρμογή μίας δεύτερης μεθόδου, αυτήν της κατηγοριοποίησης των k-Κοντινότερων Γειτόνων, με σκοπό την σύγκριση της απόδοσής τους. Αξιοποιεί δεδομένα τα οποία κατασκευάζονται και αναφέρονται σε 8 κατηγορίες τάσεων. Μέσω της βιβλιογραφίας εξετάζονται οι εκδοχές των παραμέτρων των ΤΝΔ σε ανάλογες εφαρμογές και επιλέγεται η διερεύνηση της επίδρασης ορισμένων εξ αυτών, που σχετίζονται με την αρχιτεκτονική, τον αλγόριθμο εκπαίδευσης, την συνάρτηση κόστους και την συνάρτηση ενεργοποίησης των νευρώνων, στο αποτέλεσμα. Συγκεκριμένα διαλέγονται, για την αρχιτεκτονική ο αριθμός των δύο και τριών κρυμμένων επιπέδων με 8,16,20,25 ή 30 κόμβους το κάθε ένα, για τον αλγόριθμος εκπαίδευσης, ο Levenberg-Marquardt και ο Scaled Conjugated Gradient Descent, για την συνάρτηση κόστους, το μέσο τετραγωνικό σφάλμα και η Cross Entropy και για την συνάρτηση ενεργοποίησης, η logsig και η Softmax. Το κάθε μοντέλο που δημιουργείται εκπαιδεύεται 7 φορές. Ομοίως για την μέθοδο κατηγοριοποίησης των k-Κοντινότερων Γειτόνων διαλέγονται οι τιμές 1, 2, 3 και 10 για την παράμετρο k. Τα αποτελέσματα αξιολογούνται ως προς την ολική ακρίβεια τους, την ανάκληση κάθε κλάσης και την ακρίβεια κάθε κλάσης. Τα ακριβέστερα μοντέλα κάθε μεθόδου συγκρίνονται μεταξύ τους. Επιλογικά, κατασκευάζονται δεδομένα μεικτών τάσεων και μελετάται η απόδοση των βέλτιστων μοντέλων ΤΝΔ με αυτά. Τα συμπεράσματα που προκύπτουν είναι τα ΤΝΔ υπερέχουν σε ακρίβεια και ικανότητα γενίκευσης, αλλά η απόδοσή τους επηρεάζεται σημαντικά από περισσότερες παραμέτρους, σε σχέση με τα μοντέλα knn.

Abstract

Statistical Process Control (SPC) is a method intending to improve product quality. Control charts play an important role in SPC as they indicate whether a manufacturing process is progressing normally. Trends, which are systematic patterns, in control charts imply the existence of systematic causes of variations, that reduce quality and make corrective actions necessary. Consequently, control chart pattern recognition is crucial in SPC. In recent years many methods have been applied to automate this process. One of them, with growing popularity is Artificial Neural Networks (ANNs). This paper deals with pattern recognition in control charts with the use of ANNs. Additionally, it applies a second method, this of the classification of k-Nearest Neighbors (kNN) with the purpose of comparing their performance. It utilizes simulation data referring to 8 classes of trends. Though the examination of various researches the effect, in this process, of the parameters of ANNs as, architecture, training algorithm, loss function and neuron's activation function, is considered. As a result, the ANNs formed are composed of two and three hidden layers, with either the Levenberg-Marquardt or Scaled Conjugated Gradient Descent training algorithm, either mean squared error or Cross Entropy loss function and either logsig or Softmax activation function. Each model is trained 7 times. Similarly, kNN models are formed with the parameter of k being either 1, 2, 3 or 10. The results are evaluated for their overall accuracy and each class's accuracy and precision. Additionally, each method's best models are compared. Finally, the best ANN models are tested with simulated mixed trends data. The conclusions made are that ANNs are superior in accuracy combined with generalization ability, but their performance is significantly depended on more parameters, than knn models.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον κ.Πανώριο Μπενάρδο για την καθοδήγησή του κατά την διάρκεια της εργασίας. Επίσης, ευχαριστώ εγκάρδια την οικογένειά μου για την υποστήριξή της.

Περιεχόμενα

Περίληψη 2
Abstract 3
Κατάλογος σχημάτων 7
Κατάλογος Πινάκων 11
1 Εισαγωγή 13
2 Θεωρητικό Μέρος 15
2.1 Η σημασία του ποιοτικού ελέγχου 15
2.1.1 Ιστορική αναδρομή 15
2.1.2 Διαγράμματα ελέγχου 15
2.1.3 Μορφές χρονοσειρών και τάσεις 17
2.2 Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα 22
2.2.1 Γενικά 22
2.2.2 Ιστορική αναδρομή 23
2.2.3 Βιολογικά Νευρωνικά Δίκτυα 24
2.2.4 Τεχνητά Νευρωνικά Δίκτυα 26
2.2.4.1 Αρχιτεκτονική 28
2.2.4.2 Εκπαίδευση 31
2.2.4.3 Οπισθόδρομη διάδοση σφάλματος 38
2.2.4.4 Συναρτήσεις ενεργοποίησης 43
2.2.4.5 Ειδικά 46
2.2.4.5.1 Συναρτήσεις ενεργοποίησης 46
2.2.4.5.2 Συνάρτηση Κόστους 48
2.2.4.5.3 Αλγόριθμος Εκπαίδευσης 49
2.2.4.5.4 Πίνακας Σύγχυσης 57
2.2.4.6 Εφαρμογές Τεχνητών Νευρωνικών Δικτύων 58
2.3 Αλγόριθμος k-Κοντινότερων Γειτόνων (k-nearest neighbors) 59
3 Βιβλιογραφική Ανασκόπηση 65
3.1 Αλγόριθμοι κατηγοριοποίησης σε διαγράμματα ποιότητας 65
3.2 Αναγνώριση τάσεων σε διαγράμματα ποιότητας με ΤΝΔ στην βιβλιογραφία 67
4 Ανάπτυξη Δεδομένων 72
4.1 Διανύσματα εκπαίδευσης και ελέγχου 72
4.1.1 Μητρικά διανύσματα 72
4.1.2 Παραγωγή διανυσμάτων. 75
4.2 Διανύσματα στόχοι 78
4.3 Διαχωρισμός των διανυσμάτων εκπαίδευσης και ελέγχου 78
4.4 Συμπληρωματικά διανύσματα ελέγχου 79
5 Ανάπτυξη μοντέλων αναγνώρισης τάσεων - Αποτελέσματα 81
5.1 Ανάπτυξη μοντέλων ΤΝΔ 81
5.1.1 Αντιστοίχιση εξόδου-τάσης 82
5.2 Αποτελέσματα ΤΝΔ 83
5.2.1 Αποτελέσματα κύριας βάσης δεδομένων 83
5.2.1.1 Κριτήριο μέσης ακρίβειας αρχιτεκτονικής 83
5.2.1.2 Κριτήριο βέλτιστης ακρίβειας αρχιτεκτονικής 86

5.2.2 Αποτελέσματα με την δευτερεύουσα βάση δεδομένων 89
5.2.3 Αποτελέσματα με τα διανύσματα μεικτών τάσεων 90
5.3 Ανάπτυξη μοντέλων knn 91
5.3.1 Αποτελέσματα knn 91
5.3.1.1 Αποτελέσματα κύριας βάσης δεδομένων 91
5.3.1.2 Αποτελέσματα δευτερεύουσας βάσης δεδομένων 92
6 Σχολιασμός αποτελεσμάτων και συμπεράσματα 94
6.1 ΤΝΔ 94
6.1.1 Σχολιασμός αποτελεσμάτων της κύριας βάσης δεδομένων 94
6.1.1.1 Αλγόριθμος εκπαίδευσης 94
6.1.1.2 Συνάρτηση Κόστους 97
6.1.1.3 Συνάρτηση ενεργοποίησης του επιπέδου εξόδου 99
6.1.1.4 Αριθμός κρυμμένων επιπέδων και αριθμός κόμβων 103
6.1.2 Σχολιασμός αποτελεσμάτων της δευτερεύουσας βάσης δεδομένων 105
6.2 KNN 107
6.2.1 Σχολιασμός αποτελεσμάτων της κύριας βάσης δεδομένων 107
6.2.2 Σχολιασμός αποτελεσμάτων της δευτερεύουσας βάσης δεδομένων 108
6.3 Σύγκριση ΤΝΔ knn και βιβλιογραφίας 110
6.3.1 ΤΝΔ και knn 110
6.3.2 ΤΝΔ knn και βιβλιογραφία 110
6.4 Προτάσεις για βελτίωση 111
Βιβλιογραφία 112
Παράρτημα Α 115
Παράρτημα Β 142
Παράρτημα Γ 143
Παράρτημα Δ 148

Κατάλογος σχημάτων

Εικόνα 1: Τυπικό διάγραμμα ελέγχου [3]. 17
Εικόνα 2: Σημείο εκτός των ορίων ελέγχου. 18
Εικόνα 3: Σημείο πολύ κοντά στα όρια ελέγχου. 18
Εικόνα 4: Διαδοχικά σημεία από την ίδια πλευρά της κεντρικής γραμμής. 19
Εικόνα 5: Μονότονη αύξουσα τάση. 20
Εικόνα 6: Κυκλική τάση. 20
Εικόνα 7: Άλμα τιμών. 21
Εικόνα 8: Μεγάλο ποσοστό σημείων κοντά στα όρια ελέγχου. 21
Εικόνα 9: Διαδοχικά σημεία σε μικρή απόσταση από την κεντρική γραμμή. 22
Εικόνα 10: Βιολογικοί νευρώνες και τα μέρη από τα οποία αποτελούνται [7]. 25
Εικόνα 11: Η συναπτική σύνδεση μεταξύ δύο νευρώνων [8]. 25
Εικόνα 12: Αντιστοιχία βιολογικού με τεχνητού νευρώνα [8]. 26
Εικόνα 13: Μοντέλο τεχνητού νευρώνα [9]. 27
Εικόνα 14: Νευρωνικό δίκτυο με τρία επίπεδα [10]. 28
Εικόνα 15: ΤΝΔ απλού επιπέδου [11]. 29
Εικόνα 16: ΤΝΔ πολλαπλών επιπέδων [11]. 29
Εικόνα 17: ΤΝΔ με παρακάμψεις [11]. 30
Εικόνα 18: ΤΝΔ τύπου Hopfield με τρείς νευρώνες [11]. 31
Εικόνα 19: Τρόποι εκπαίδευσης και κατηγορίες προβλημάτων στις οποίες εφαρμόζονται. 32
Εικόνα 20: Χαρακτηριστικές περιπτώσεις εκπαιδευμένων ΤΝΔ [11]. 33
Εικόνα 21: Πολλαπλές κλάσεις [12]. 35
Εικόνα 22: Η θεωρία μάθησης του Thorndike, που εξαρτά την μάθηση από το ερέθισμα που ακολουθεί, ως συνέπεια της συμπεριφοράς που εκδηλώνεται [13]. 36
Εικόνα 23: Σχηματική απεικόνιση της συνάρτησης σφάλματος σε σχέση με τα συναπτικά βάρη στον αλγόριθμο gradient descent [14]. 38
Εικόνα 24: Σχηματική αναπαράσταση του τρόπου με τον οποίο μειώνεται το σφάλμα μέσω της ανανέωσης των συντελεστών βαρύτητας στην οπισθόδρομη διάδοση σφάλματος [15]. 39
Εικόνα 25: Η φορά της οπισθόδρομης διάδοσης σφάλματος [15]. 39
Εικόνα 26: Εύρεση της βέλτιστης λύσης με την αρχή της κατεύθυνση μείωσης του σφάλματος (gradient descent) και εφαρμογή του κανόνα του Δέλτα [16]. 43
Εικόνα 27: Ταυτοτική συνάρτηση [30]. 44
Εικόνα 28: Βηματική συνάρτηση. 44
Εικόνα 29: Δυαδική σιγμοειδής συνάρτηση [30]. 45
Εικόνα 30: Διπολική σιγμοειδής συνάρτηση [30]. 45
Εικόνα 31: Συνάρτηση ReLU [31]. 46
Εικόνα 32: Παράδειγμα εφαρμογής της Softmax [17]. 47
Εικόνα 33: Παράδειγμα εφαρμογής της Cross Entropy. Αριστερά φαίνεται η πρόβλεψη του ΤΝΔ για την πιθανότητα κάθε κλάσης ενώ δεξιά η πραγματική πιθανότητα κάθε κλάσης. Η είσοδος εμφανώς ανήκει στην πρώτη κλάση [17]. 49

Εικόνα 34: Ο αλγόριθμος Levenberg-Marquardt (LM) ξεπερνάει σε επίδοση αλλά και σε απαιτήσεις μνήμης τους υπολοίπους [16]. 51
Εικόνα 35: Αριστερά, η τρισδιάστατη απεικόνιση μίας συνάρτησης δύο μεταβλητών. Με βέλη φαίνονται τα διανύσματα μέγιστης κλίσης σε διάφορα σημεία. Δεξιά, διάφορες ισοσταθμικές καμπύλες πάνω στις οποίες η τιμή της συνάρτησης διατηρείται σταθερή. [18] 52
Εικόνα 36: Σύγκριση του τρόπου μετακίνησης των δύο αλγορίθμων, πάνω στις ισοσταθμικές καμπύλες προς την κατεύθυνση μείωσής τους [19]. 53
Εικόνα 37: Η τετραγωνική μορφή της συνάρτησης σφάλματος που είναι παραβολοειδής, όταν ο πίνακας A είναι θετικά ορισμένος. Η ελάχιστη τιμή βρίσκεται στο κατώτατο σημείο [20]. 53
Εικόνα 38: Αριστερά, η τομή του κάθετου επιπέδου, που περιέχει το r_i με την συνάρτηση σφάλματος [20]. Δεξιά, η καμπύλη που προκύπτει. Στόχος αποτελεί η εύρεση του κατώτατου σημείου της [20]. 54
Εικόνα 39: Το σφάλμα e_1 είναι ορθογώνια συζευγμένο με το διάνυσμα d_0 που σημαίνει ότι στην κατεύθυνση του d_0 δεν χρειάζεται να πραγματοποιηθεί άλλο βήμα [20]. 56
Εικόνα 40: Κατηγοριοποίηση με βάση την πλησιέστερη κλάση [21]. 60
Εικόνα 41: (α) Η πράσινη κουκίδα αν k<=3 εκχωρείται στις Μπλε ενώ αν k>3 στις Πορτοκαλί. (β) Για μικρές τιμές του k οι περιοχές εμφανίζουν απότομες γωνίες ενώ (γ) για μεγαλύτερες τιμές είναι πιο ομαλές. 62
Εικόνα 42: Δέντρο απόφασης (Decision Tree) 66
Εικόνα 43: Η μεθοδολογία που ακολουθήθηκε στην εργασία των Tunchanit Bunjongjit και Wimalin Laosiritaworn [22]. 67
Εικόνα 44: Δομή του δικτύου των Guh και Hsieh [24]. Στο επίπεδο N1 ελέγχεται με το ΤΝΔ A η είσοδος και αν αναγνωριστεί ως μία από τις εξεταζόμενες τάσεις, περνά στο επίπεδο B εκχωρούμενη στο ανάλογο ΤΝΔ που δίνει την εκτίμηση των παραμέτρων της. 69
Εικόνα 45: Μεταβολή της ακρίβειας με την αύξηση του αριθμού των κόμβων του κρυμμένου επιπέδου. 70
Εικόνα 46: Το δίκτυο που μορφοποιούν οι [29] με επιμέρους ΤΝΔ για την αναγνώριση μεικτών τάσεων. 71
Εικόνα 47: Οι τρείς μήτρες της κάθε τάσης, αποτελούμενες από 16 στοιχεία. 74
Εικόνα 48: Παράγωγα διανύσματα των μητρικών διανυσμάτων Cycle pattern 2, Instability 3 και Monotonous 1. 76
Εικόνα 49: Υπολογίζονται 2 τυπικές αποκλίσεις για κάθε διάνυσμα άλματος και κρατείται η μικρότερη. 77
Εικόνα 50: Μικρή μετακίνηση των σημείων κατά τυχαίο τρόπο. 77
Εικόνα 51: Δημιουργία χρονοσειράς από την μήτρα Cycle pattern 3, μέσω μεταβολής της που εμπεριέχει τυχαιότητα. 78
Εικόνα 52: Οι αρχιτεκτονικές ΤΝΔ που αναπτύχθηκαν. 82
Εικόνα 53: Πίνακες σύγχυσης των βέλτιστων μοντέλων από τα 7 που απέδωσαν την μέγιστη μέση ακρίβεια, για τους αλγορίθμους εκπαίδευσης LM και SCG. Στο δεξί τμήμα φαίνεται η ανάκληση και στο κάτω η ακρίβεια κάθε κλάσης. (α) Πίνακας σύγχυσης του

μοντέλου LM mse logsig 16-30-16 με ακρίβεια 97.22% (β) Πίνακας σύγχυσης του μοντέλου SCG mse logsig 8-25-16 με ακρίβεια 52.78%. 85
Εικόνα 54: Πίνακες σύγχυσης των ακριβέστερων μοντέλων των αλγορίθμων εκπαίδευσης LM και SCG. (α) Πίνακας σύγχυσης του μοντέλου LM mse logsig 8-30-16 με ακρίβεια 100%. (β) Πίνακας σύγχυσης του μοντέλου SCG crossentropy softmax 8-30 με ακρίβεια 73.61%..... 88
Εικόνα 55: Πίνακας σύγχυσης των ακριβέστερων μοντέλων κάθε αλγορίθμου εκπάιδευσης, από τις δοκιμές με την δευτερεύουσα βάση δεδομένων. (α) Πίνακας σύγχυσης του μοντέλου LM mse logsig 8-25-25 με ακρίβεια 100%. (β) Πίνακας σύγχυσης του μοντέλου SCG crossentropy softmax 8-30 με ακρίβεια 66.67%. 90
Εικόνα 56: Μοντέλα knn που αναπτύχθηκαν. 91
Εικόνα 57: Πίνακας σύγχυσης της ακριβέστερης περίπτωσης μοντέλου knn, με k=1 για την κύρια βάση δεδομένων. 92
Εικόνα 58: Πίνακας σύγχυσης της ακριβέστερης περίπτωσης μοντέλου knn, με k=1 για την δευτερεύουσα βάση δεδομένων. 93
Εικόνα 59: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών κάθε αλγορίθμου εκπαίδευσης, με την μέση ακρίβεια των αντίστοιχων αρχιτεκτονικών με τον έτερο αλγόριθμο εκπαίδευσης. Αριστερά: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών LM με τις αντίστοιχες SCG. Δεξιά: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών SCG με τις αντίστοιχες LM. 95
Εικόνα 60: Σύγκριση ως προς την μέση ακρίβεια του συνόλου των αρχιτεκτονικών (α) LM και SCG mse logsig (β) LM και SCG mse softmax. 97
Εικόνα 61: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών κάθε συνάρτησης κόστους, με την μέση ακρίβεια των αντίστοιχων αρχιτεκτονικών με την ετέρα συνάρτηση κόστους. Αριστερά: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών SCG mse softmax με τις αντίστοιχες SCG crossentropy softmax. Δεξιά: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών SCG crossentropy softmax με τις αντίστοιχες SCG mse softmax. 98
Εικόνα 62: Σύγκριση ως προς την μέση ακρίβεια του συνόλου των αρχιτεκτονικών SCG mse και crossentropy softmax..... 99
Εικόνα 63: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών κάθε συνάρτησης ενεργοποίησης του επιπέδου εξόδου, με την μέση ακρίβεια των αντίστοιχων αρχιτεκτονικών με την ετέρα συνάρτηση ενεργοποίησης. (α) Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών LM mse logsig με τις αντίστοιχες LM mse softmax. (β) Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών SCG mse logsig με τις αντίστοιχες SCG mse softmax. (γ) Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών LM mse softmax με τις αντίστοιχες LM mse logsig. (δ) Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών SCG mse softmax με τις αντίστοιχες SCG mse logsig. 102
Εικόνα 64: Σύγκριση ως προς την μέση ακρίβεια του συνόλου των αρχιτεκτονικών (α) LM mse logsig και softmax (β) SCG mse logsig και softmax..... 103

Εικόνα 65: Επίδραση του αριθμού κρυμμένων επιπέδων και κόμβων στις τιμές της μέσης ακρίβειας κάθε αρχιτεκτονικής. 104
Εικόνα 66: Σύγκριση βέλτιστων μοντέλων για την κύρια βάση δεδομένων με τα αποτελέσματά τους για την δευτερεύουσα. 107
Εικόνα 67: Σύγκριση μοντέλων knn για την κύρια και την δευτερεύουσα βάση δεδομένων. 109
Εικόνα 68: Τα 19 διανύσματα μεικτών τάσεων που δημιουργήθηκαν. Από αριστερά προς τα δεξιά οι τάσεις που συνδυάζονται είναι: Περιοδικότητα & Σημείο πολύ κοντά στα όρια ελέγχου, Περιοδικότητα & Διακύμανση, Περιοδικότητα & Μονότονη τάση, Περιοδικότητα & Σημείο εκτός ορίων ελέγχου, Περιοδικότητα & Έλλειψη μεταβλητότητας, Σημείο πολύ κοντά στα όρια ελέγχου & Ομαδοποίηση, Σημείο πολύ κοντά στα όρια ελέγχου & Άλμα αλλαγής επιπέδου, Σημείο πολύ κοντά στα όρια ελέγχου & Μονότονη τάση, Σημείο πολύ κοντά στα όρια ελέγχου & Έλλειψη μεταβλητότητας, Ομαδοποίηση & Διακύμανση, Ομαδοποίηση & Άλμα αλλαγής επιπέδου, Ομαδοποίηση & Μονότονη τάση, Ομαδοποίηση & Σημείο εκτός ορίων ελέγχου, Ομαδοποίηση & Έλλειψη μεταβλητότητας, Άλμα αλλαγής επιπέδου & Μονότονη τάση, Άλμα αλλαγής επιπέδου & Σημείο εκτός ορίων ελέγχου, Άλμα αλλαγής επιπέδου & Έλλειψη μεταβλητότητας, Μονότονη τάση & Σημείο εκτός ορίων ελέγχου, Σημείο εκτός ορίων ελέγχου & Έλλειψη μεταβλητότητας 144
Εικόνα 69: Πίνακες σύγχυσης για τα διανύσματα των μεικτών τάσεων των βέλτιστων μοντέλων του Πίνακα 6. 147

Κατάλογος Πινάκων

Πίνακας 1: Γενική μορφή πίνακα σύγχυσης. 57
Πίνακας 2: Στοιχεία δανειοληπτών. 61
Πίνακας 3: Στοιχεία δανειοληπτών με κανονικοποίηση. 63
Πίνακας 4: Αποτελέσματα με βάση την μέση ακρίβεια, των τριών πιο επιτυχημένων συνδυασμών κρυμμένων επιπέδων και κόμβων, κάθε αρχιτεκτονικής. 84
Πίνακας 5: Μέσες ακρίβειες και ανακλήσεις των κλάσεων για την αρχιτεκτονική LM mse logsig 16-30-16, που έχει την μεγαλύτερη μέση ακρίβεια. 86
Πίνακας 6: Τα ακριβέστερα μοντέλα, των τριών πιο επιτυχημένων συνδυασμών κρυμμένων επιπέδων και κόμβων, κάθε αρχιτεκτονικής. 87
Πίνακας 7: Εφαρμογή των διανυσμάτων ελέγχου της δευτερεύουσας βάσης στα ακριβέστερα μοντέλα, των τριών πιο επιτυχημένων συνδυασμών κρυμμένων επιπέδων και κόμβων, κάθε αρχιτεκτονικής. 89
Πίνακας 8: Αποτελέσματα μοντέλων knn για τα 72 διανύσματα ελέγχου της κύριας βάσης δεδομένων. 91
Πίνακας 9: Αποτελέσματα μοντέλων knn για τα 24 διανύσματα ελέγχου της δευτερεύουσας βάσης δεδομένων. 92
Πίνακας 10: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών LM, με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές SCG. 94
Πίνακας 11: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών SCG, με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές LM. 95
Πίνακας 12: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών SCG mse softmax με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές SCG crossentropy softmax. 97
Πίνακας 13: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών SCG crossentropy softmax με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές SCG mse softmax. 98
Πίνακας 14: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών LM mse logsig, με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές LM mse softmax. 100
Πίνακας 15: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών SCG mse logsig, με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές SCG mse softmax. 100
Πίνακας 16: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών LM mse softmax, με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές LM mse logsig. 100
Πίνακας 17: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών SCG mse softmax, με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές LM mse logsig. 101
Πίνακας 18: Τα βέλτιστα μοντέλα της κύριας βάσης δεδομένων και τα αποτελέσματά τους με την δευτερεύουσα. 106
Πίνακας 19: Ακρίβεια μοντέλων knn με την κύρια βάση δεδομένων. 108
Πίνακας 20: Σύγκριση ακρίβειας μοντέλων knn με την κύρια και την δευτερεύουσα βάση δεδομένων. 109
Πίνακας 21: Μέση και βέλτιστη ακρίβεια των 750 αρχιτεκτονικών. 115
Πίνακας 22: Τιμές της μέσης ακρίβειας και ανάκλησης της κάθε κλάσης, για όλες τις αρχιτεκτονικές του Πίνακα 4. 142
Πίνακας 23: Αποτελέσματα των ακριβέστερων μοντέλων, για τα διανύσματα ελέγχου των μεικτών τάσεων. 145
Πίνακας 24: Οι ακριβείς τιμές των τριών μητρικών διανυσμάτων της κάθε τάσης. Φαίνονται επίσης, η τυπική απόκλιση (τ.α.) και η παράμετρος θέσης των ορίων ελέγχου (k). 148
Πίνακας 25: Διάνυσμα στόχος κάθε τάσης. 150

1 Εισαγωγή

Ο στατιστικός έλεγχος ποιότητας είναι ένα κύριο τμήμα της παραγωγικής διαδικασίας που αποβλέπει στην διατήρηση της επιθυμητής ποιότητας. Τα διαγράμματα ελέγχου ποιότητας αποτελούν αναπόσπαστο κομμάτι του, αφού η μελέτη τους οδηγεί σε συμπεράσματα που αφορούν την ομαλότητα της παραγωγικής διαδικασίας. Η τελευταία κρίνεται άρτια όταν οι χρονοσειρές των διαγραμμάτων ελέγχου παρουσιάζουν τυχαιότητα. Αντιθέτως, η εμφάνιση σε αυτά τάσεων, δηλαδή μη τυχαίων χρονοσειρών αποτελεί ισχυρή ένδειξη ύπαρξης συστηματικής αιτίας, που παρεμβαίνει στην παραγωγική διαδικασία, και μεταβάλλοντας τα μετρούμενα χαρακτηριστικά υποβαθμίζει την ποιότητα. Κατά συνέπεια είναι απαραίτητη η πραγματοποίηση διορθωτικών ενεργειών. Η εξέταση των διαγραμμάτων ποιότητας πραγματοποιείται πλέον αυτοματοποιημένα, με μεθόδους που αποφαίνονται για την ύπαρξη συστηματικών αιτιών μέσω του εντοπισμού των πιθανών τάσεων. Μία από αυτές, με όλο και μεγαλύτερη απήχηση είναι τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ).

Το αντικείμενο της παρούσας εργασίας είναι η αναγνώριση τάσεων σε διαγράμματα ελέγχου ποιότητας μέσω ΤΝΔ αλλά και η σύγκρισή τους μία δεύτερη μέθοδο που εφαρμόζεται για τον συγκεκριμένο σκοπό, αυτήν των k-Κοντινότερων Γειτόνων (knn). Απώτερος στόχος αποτελεί, η σύγκριση της αποτελεσματικότητας των δύο μεθόδων, έχοντας όμως προηγηθεί, η μεγιστοποίηση κατά το δυνατόν της απόδοσης τους, μέσω της διερεύνησης των παραγόντων που επηρεάζουν την κάθε μία. Οι παράγοντες αυτοί για τα ΤΝΔ είναι η αρχιτεκτονική, η εκπαίδευση, η συνάρτηση κόστους και η συνάρτηση ενεργοποίησης των νευρώνων, ενώ για την μέθοδο knn είναι η τιμή της παραμέτρου k.

Μέσω της βιβλιογραφίας διενεργείται μία επισκόπηση των πιο διαδεδομένων εκδοχών αυτών των παραγόντων και επιλέγονται συγκεκριμένες από αυτές ώστε να εφαρμοστούν. Συγκεκριμένα διαλέγονται, για την αρχιτεκτονική ο αριθμός των δύο και τριών κρυμμένων επιπέδων με 8,16,20,25 ή 30 κόμβους το κάθε ένα, για τον αλγόριθμος εκπαίδευσης, ο Levenberg-Marquardt και ο Scaled Conjugated Gradient Descent, για την συνάρτηση κόστους, το μέσο τετραγωνικό σφάλμα και η Cross Entropy και για την συνάρτηση ενεργοποίησης, η logsig και η Softmax. Για την παράμετρο k διαλέγονται οι τιμές 1,2,3 και 10. Η κάθε μία από αυτές τις εκδοχές του κάθε παράγοντα συνδυάζεται με όλες τις εκδοχές των υπολοίπων οδηγώντας στην δημιουργία 750 διαφορετικών περιπτώσεων ΤΝΔ και 4 knn. Ταυτόχρονα όμως, κρίνεται σκόπιμη η διαφοροποίηση της εργασίας από τις κύριες κατευθυντήριες γραμμές της βιβλιογραφίας. Έτσι, η κάθε περίπτωση ΤΝΔ εκπαιδεύεται όχι μία αλλά 7 φορές, οδηγώντας στην δημιουργία 5250 εκπαιδευμένων μοντέλων, με στόχο την εκλογή του ακριβέστερου και την σύγκρισή του με το ακριβέστερο μοντέλο knn.

Τα δεδομένα εκπαίδευσης και ελέγχου που χρησιμοποιούνται, είναι κατασκευασμένα και όχι πειραματικά, όπως και στην πλειοψηφία των σχετικών ερευνών, ωστόσο ο τρόπος κατασκευής και ανάπτυξής τους διαφέρει. Επιλέγεται μία διαφορετική προσέγγιση από την αυστηρή μαθηματική δομή των τάσεων της βιβλιογραφίας, εστιάζοντας περισσότερο στην μορφή τους και στην διεύρυνση της ικανότητας αναγνώρισή τους από τα μοντέλα ΤΝΔ και knn. Με αυτήν την λογική, ο αριθμός των

τάσεων διαφοροποιείται από τις 3 επικρατούσες τάσεις της βιβλιογραφίας, που είναι η κυκλική, η μονότονη και το άλμα αλλαγής επιπέδου, σε 8, δηλαδή στις αναφερθείσες και σε 5 επιπρόσθετες. Αυτός είναι ένας τρόπος αύξησης της δυσκολίας, καθώς αυξάνεται η πολυπλοκότητα. Στην ίδια λογική, επενεργείται διπλός έλεγχος της απόδοσής, τόσο με μία κύρια βάση δεδομένων όσο και με μία δευτερεύουσα, η οποία δομείται με τέτοιο τρόπο ώστε να αποτελεί πρόκληση για τα μοντέλα της κάθε μεθόδου, που αποδεικνύονται πιο αποδοτικά με την πρώτη.

Όλα τα προαναφερθέντα, μαζί με την τυχαιότητα που προστίθεται στην κατασκευή των δεδομένων, δημιουργούν ένα σύνολο συνισταμένων, τέτοιων ώστε το μοντέλο που θα καταφέρει να ανταποκριθεί σε αυτές να είναι αξιόπιστο. Η αξιολόγηση των αποτελεσμάτων πραγματοποιείται με βάση την συνολική ακρίβεια, την ανάκληση και την ακρίβεια κάθε κλάσης. Αυτά είναι τα κριτήρια με τα οποία εξετάζεται η συνεισφορά του κάθε παράγοντα, ξεχωριστά. Επιπλέον, πέραν του βέλτιστου μοντέλου, αξιολογούνται τα 7 μοντέλα της κάθε περίπτωσης ομαδικά, με βάση την μέση απόδοσή τους έτσι ώστε να εκτιμηθεί αν οι υψηλές αποδόσεις έχουν ευρύτερη ισχύ, ή οφείλονται σε μεμονωμένες περιπτώσεις. Τέλος, τα βέλτιστα μοντέλα ΤΝΔ δοκιμάζονται με δεδομένα μεικτών τάσεων που κατασκευάζονται για τον σκοπό αυτό.

Τα συμπεράσματα που αποκομίζονται είναι ότι η ακρίβεια των ΤΝΔ στην συγκεκριμένη εφαρμογή παρουσιάζει μεγάλες μεταβολές ανάλογα με τις παραμέτρους που χρησιμοποιούνται, κάτι που ισχύει και για τα μοντέλα knn. Η σύγκριση των βέλτιστων μοντέλων που δημιουργούνται από κάθε μέθοδο, βρίσκει τα ΤΝΔ να επικρατούν τόσο εξαιτίας της υπερέχουσας ακρίβειάς τους όσο και λόγω της ικανότητας γενίκευσής τους, στην οποία υστερούν τα μοντέλα knn και ελέγχεται με την δευτερεύουσα βάση δεδομένων. Ωστόσο οι δύο αυτές συνιστώσες, των ΤΝΔ που αναπτύχθηκαν, δεν αντικατοπτρίζονται άμεσα στον έλεγχο με τα δεδομένα μεικτών τάσεων.

2 Θεωρητικό Μέρος

2.1 Η σημασία του ποιοτικού ελέγχου

2.1.1 Ιστορική αναδρομή

Η ποιότητα σαν χαρακτηριστικό οπουδήποτε προϊόντος συνοψίζεται στο κατά πόσο αυτό είναι απαλλαγμένο από ελαττώματα και ικανοποιεί τις προδιαγραφές. Ο έλεγχος ποιότητας είναι το σύνολο των τεχνικών που εφαρμόζονται ώστε να ελεγχθεί αν πληρούνται οι προϋποθέσεις της ποιότητας [1].

Από τις αρχές του 19^{ου} αιώνα μέχρι περίπου το 1920 η διασφάλιση ποιότητας βασίζεται αποκλειστικά στην επιθεώρηση και μέτρηση των παραγόμενων προϊόντων. Χωρίς να επεκτείνεται σε προσπάθειες αναγνώρισης και διόρθωσης των αιτιών κακής ποιότητας, περιορίζεται στον εντοπισμό και επισκευή των ελαττωματικών.

Ο στατιστικός έλεγχος ποιότητας (Statistical Quality Control) έχει τις αρχές του στην δεκαετία του 1920 στα Bell Telephone Laboratories των Η.Π.Α. όπου αναπτύχθηκαν ταυτόχρονα δύο μεγάλες περιοχές: η δειγματοληψία αποδοχής και ο έλεγχος παραγωγικής διαδικασίας.

Ο έλεγχος ποιότητας αποδοχής (acceptance sampling) ξεκίνησε από την απλή διαπίστωση ότι ο 100% έλεγχος δεν είναι ο περισσότερο αποτελεσματικός τρόπος, από οικονομική άποψη, διαχωρισμού μεταξύ καλών και ελαττωματικών προϊόντων.

Ο έλεγχος παραγωγικής διαδικασίας (process control) εισάγει την έννοια της πρόληψης στον έλεγχο ποιότητας. Η αρχική του ανάπτυξη οφείλεται στον Walter Shewhart, ο οποίος συνειδητοποίησε ότι η διασπορά τιμών ενός χαρακτηριστικού ποιότητας (π.χ. μιας διάστασης) είναι αναπόφευκτη κατά την παραγωγή, αλλά ένα μέρος της οφείλεται σε τυχαίες, μη ελεγχόμενες αιτίες και ένα άλλο μέρος οφείλεται σε συγκεκριμένα αίτια (συστηματικές μεταβολές), τα οποία μπορούν να εντοπιστούν και να διορθωθούν. Για να διευκολύνει τη διάκριση ανάμεσα σε τυχαίες και συστηματικές μεταβολές, ο Shewhart σχεδίασε απλές στατιστικές τεχνικές και αντίστοιχα διαγράμματα ελέγχου και πρότεινε τρόπους βελτίωσης της ποιότητας με εξάλειψη των αιτιών των συστηματικών μεταβολών. Τα διαγράμματα ελέγχου που πρότεινε ο Shewhart εξακολουθούν να είναι ευρύτατα χρησιμοποιούμενα εργαλεία για τον έλεγχο ομαλής λειτουργίας των παραγωγικών διαδικασιών [2].

2.1.2 Διαγράμματα ελέγχου

Ο ρόλος των γραφικών μεθόδων είναι να μετασχηματίζουν μεγάλο όγκο αριθμητικών δεδομένων σε απλές γραφικές παραστάσεις. Οι κυριότερες από αυτές είναι τα διαγράμματα διαδικασίας, αιτίας-αποτελέσματος, διασποράς, διαγράμματα Pareto, φύλλα καταχώρησης, ιστογράμματα και διαγράμματα ελέγχου.

Το διάγραμμα ελέγχου (control chart) είναι ένα ειδικό διάγραμμα που απεικονίζει την πορεία των χαρακτηριστικών ποιότητας σε συνάρτηση με τον χρόνο παραγωγής. Συγκεκριμένα, δείγματα λαμβάνονται από την παραγωγική διαδικασία σε τακτά χρονικά διαστήματα και οι τιμές ορισμένης εκτιμήτριας (π.χ. του ποσοστού ελαττωματικών ή της

μέσης τιμής ενός μεγέθους) σημειώνονται διαδοχικά στο διάγραμμα. Στην ορολογία του στατιστικού ελέγχου παραγωγικής διαδικασίας η εκτιμήτρια ονομάζεται στατιστική δείγματος.

Η χρήση των διαγραμμάτων ελέγχου αποσκοπεί στον εντοπισμό μεταβολών στην παραγωγική διαδικασία και στην απάλειψη των αιτιών που τις προκαλούν. Οι αποκλίσεις από τις επιθυμητές τιμές έστω και σε ένα μικρό ποσοστό της συνολικής παραγωγής είναι αναπόφευκτες. Η προέλευσή τους ενδέχεται να έχει τις ρίζες της είτε σε τυχαίες είτε σε συστηματικές αιτίες. Τυχαίες αιτίες είναι εκείνες οι οποίες είναι πρακτικά αδύνατο να ελεγχθούν όπως π.χ. ξαφνικές αλλαγές του εξωτερικού περιβάλλοντος. Συστηματικές αιτίες είναι εκείνες οι οποίες είναι εφικτό να ελέγχονται, εφόσον εντοπιστούν από κάποια μέθοδο ελέγχου. Παράδειγμα συστηματικής αιτίας είναι η φθορά του κοπτικού εργαλείου σε μια κατεργασία κοπής μετάλλων.

Εκείνο που διαφοροποιεί τις διαδικασίες ως προς την ποιότητά τους δεν είναι η παρουσία ή απουσία μεταβλητότητας αλλά το μέγεθός της. Αυτό είναι που κάνει τον εντοπισμό των συστηματικών αιτιών καίριας σημασίας, καθώς η επίπτωσή τους στη διασπορά των τιμών είναι γενικά μεγαλύτερη από την επίπτωση των τυχαίων αιτιών και κατά κανόνα οδηγεί σε μη αποδεκτή στάθμη ποιότητας της διαδικασίας. Όταν μια παραγωγική διαδικασία είναι απαλλαγμένη από την παρουσία και επίδραση συστηματικών αιτιών, λέγεται ότι βρίσκεται σε κατάσταση στατιστικού ελέγχου.

Στη συνηθέστερη μορφή του διαγράμματος ελέγχου, η οποία φαίνεται παρακάτω, απεικονίζονται η κεντρική γραμμή (ΚΓ), το άνω όριο ελέγχου (ΑΟΕ) και το κάτω όριο ελέγχου (ΚΟΕ).

Οι γενικές σχέσεις που προσδιορίζουν την κεντρική γραμμή και τα όρια ελέγχου του απλού διαγράμματος Shewhart για οποιαδήποτε στατιστική δείγματος w είναι:

$$AOE = \mu_w + k\sigma_w \tag{1.1}$$

$$K\Gamma = \mu_w \tag{1.2}$$

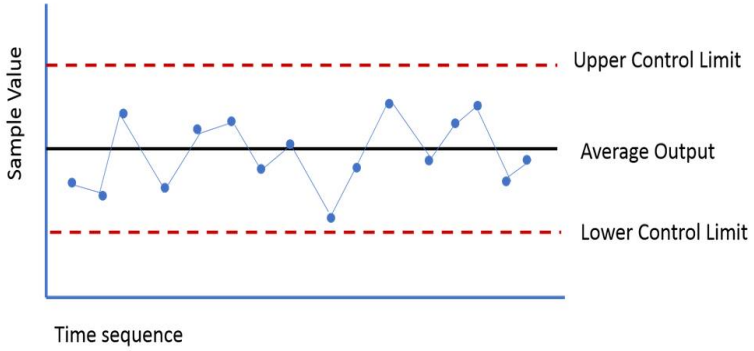
$$KOE = \mu_w - k\sigma_w \tag{1.3}$$

όπου μ_w και σ_w είναι αντίστοιχα η μέση τιμή και η τυπική απόκλιση της w σε κατάσταση στατιστικού ελέγχου και k είναι η παράμετρος θέσης των ορίων ελέγχου που καθορίζει την απόσταση των ορίων από την κεντρική γραμμή σε αριθμό τυπικών αποκλίσεων σ_w.

Κατά πλειοψηφία τα διαγράμματα ελέγχου έχουν όρια 3 τυπικών αποκλίσεων (k=3) και σε αυτήν την περίπτωση το άνω και κάτω όριο ονομάζονται φυσικά όρια ανοχών και εμπεριέχουν, για ένα χαρακτηριστικό ποιότητας που ακολουθεί κανονική κατανομή, το 99,73% των τιμών της τυχαίας μεταβλητής.

Η κεντρική γραμμή αντιστοιχεί στην τιμή της παραμέτρου όταν η διαδικασία βρίσκεται σε στατιστικό έλεγχο. Εκφράζει δηλαδή την κατάσταση ομαλής λειτουργίας, όπου η μεταβλητότητα οφείλεται μόνο σε τυχαίες αιτίες. Όταν τα σημεία του διαγράμματος βρίσκονται εντός των ορίων ελέγχου, η διαδικασία θεωρείται ότι βρίσκεται σε στατιστικό έλεγχο. Σημείο εκτός των ορίων ελέγχου αποτελεί ένδειξη ότι έχει επέλθει μεταβολή

στην παραγωγική διαδικασία που οφείλεται σε επίδραση συστηματικής αιτίας, όποτε είναι σκόπιμη η διερεύνηση της αιτίας αυτής της μεταβολής.



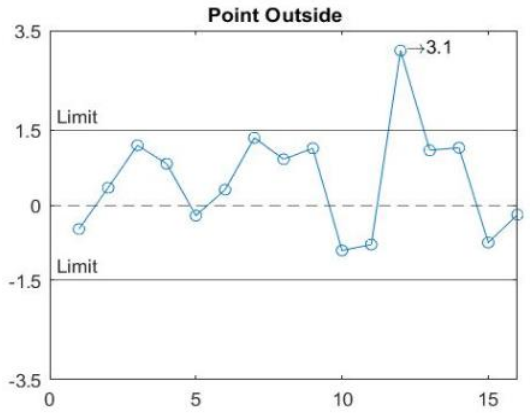
Εικόνα 1: Τυπικό διάγραμμα ελέγχου [3].

Η λειτουργία του διαγράμματος ελέγχου δεν εξαντλείται όμως σε αυτόν τον απλό κανόνα. Είναι σύνηθες όλα τα σημεία του διαγράμματος να βρίσκονται εντός των ορίων ελέγχου, αλλά η διάταξή τους να αποκαλύπτει την παρουσία μιας συστηματικής αιτίας. Αν π.χ. 15 διαδοχικά σημεία είναι πάνω από την κεντρική γραμμή αλλά κάτω από το άνω όριο ελέγχου, είναι προφανές ότι η κατανομή του χαρακτηριστικού ποιότητας έχει μετατοπιστεί και επομένως η διαδικασία βρίσκεται εκτός στατιστικού ελέγχου. Όταν η διαδικασία είναι πραγματικά σε στατιστικό έλεγχο, τα σημεία του διαγράμματος εμφανίζονται κατά απολύτως τυχαίο τρόπο, με το μεγαλύτερο ποσοστό τους κοντά στην κεντρική γραμμή αλλά και από τις δύο πλευρές της.

Παρότι το διάγραμμα ελέγχου είναι φαινομενικά μια απλή γραφική τεχνική, υποστηρίζεται από συγκεκριμένη στατιστική θεωρία, αφού αποτελεί ουσιαστικά εφαρμογή και γραφική απεικόνιση του στατιστικού ελέγχου υποθέσεων. Η σύγκριση κάθε σημείου με τα όρια ελέγχου του διαγράμματος είναι στατιστικός έλεγχος της υπόθεσης ότι η κατανομή του χαρακτηριστικού ποιότητας παραμένει εκείνη ακριβώς που αναμένεται όταν η μεταβλητότητα οφείλεται αποκλειστικά σε τυχαίες αιτίες [2]. Τα διαγράμματα ελέγχου της παρούσας εργασίας στηρίζονται στα αντίστοιχα του *SPC pattern recognition using LVQ neural nets* [4] τα οποία χρησιμοποιούνται ως πηγές αριθμητικών δεδομένων.

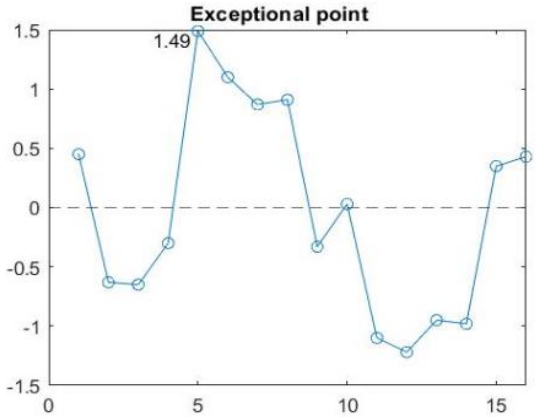
2.1.3 Μορφές χρονοσειρών και τάσεις

Ο απλούστερος κανόνας ένδειξης διαδικασίας εκτός ελέγχου αφορά την περίπτωση που η στατιστική δείγματος βρίσκεται εκτός των ορίων ελέγχου (**point outside control limits**), όπως φαίνεται στην Εικόνα 2. Τα όρια ελέγχου επισημαίνονται με τις δύο γραμμές «Limit».



Εικόνα 2: Σημείο εκτός των ορίων ελέγχου.

Ο κανόνας κατά επέκταση συμπεριλαμβάνει και την περίπτωση όπου κάποιο σημείο βρίσκεται οριακά εντός των ορίων και σχεδόν πάνω σε κάποιο από τα όρια ελέγχου (**exceptional point**), όπως παρουσιάζεται στην Εικόνα 3. Τα όρια ελέγχου, σε αυτήν την περίπτωση όπως και σε όσες ακολουθούν, ταυτίζονται με τα όρια του διαγράμματος.



Εικόνα 3: Σημείο πολύ κοντά στα όρια ελέγχου.

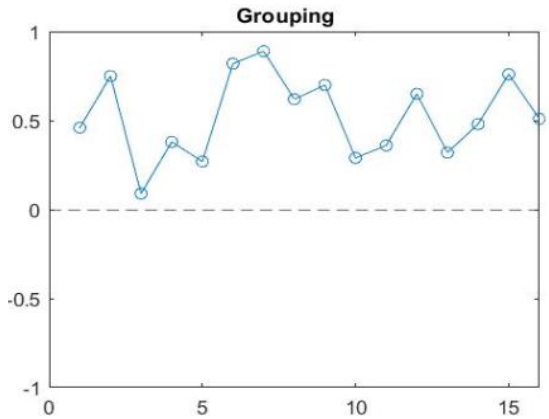
Ωστόσο αυτός ο κανόνας είναι επαρκής μόνο όταν οι συστηματικές αιτίες προκαλούν σχετικά μεγάλες μεταβολές στην παραγωγική διαδικασία. Ορισμένες συστηματικές αιτίες αντιθέτως, δεν προκαλούν τόσο ευκρινείς και μόνιμες μεταβολές όσο μεταβαλλόμενες και σταδιακές, που χαρακτηρίζονται από κάποιο μοτίβο όπως περιοδικότητα ή τάση. Κατά συνέπεια για να εντοπιστούν απαιτείται η εξέταση της συνολικής εικόνας μίας χρονοσειράς η οποία παρέχει περισσότερες πληροφορίες από όσες παρέχει κάθε μεμονωμένο σημείο του διαγράμματος.

Οι κυριότερες μορφές μη τυχαίων χρονοσειρών που παραπέμπουν σε επίδραση συστηματικής αιτίας κατατάσσονται σε δύο κύριες κατηγορίες ανάλογα με τον τύπο της μεταβολής που τις δημιουργεί:

α) χρονοσειρές που οφείλονται σε στιγμιαία μεταβολή της τιμής μιας παραμέτρου, η οποία στην συνέχεια διατηρείται σταθερή στην νέα της τιμή (π.χ. αύξηση του ποσοστού ελαττωματικών εξαιτίας παρτίδας ελαττωματικής πρώτης ύλης)

β) χρονοσειρές που οφείλονται σε συνεχή μεταβολή μιας παραμέτρου (π.χ. συνεχή αύξηση της μέσης ταχύτητας σε κατεργασία κοπής μετάλλου εξαιτίας της σταδιακής φθοράς του κοπτικού εργαλείου).

Οι χρονοσειρές που οφείλονται σε στιγμιαία αλλά παραμένουσα μεταβολή μιας παραμέτρου χαρακτηρίζονται από σχετικά μεγάλο αριθμό διαδοχικών σημείων από τη μία μόνο πλευρά της κεντρικής γραμμής (**grouping**), όπως φαίνεται στην Εικόνα 4



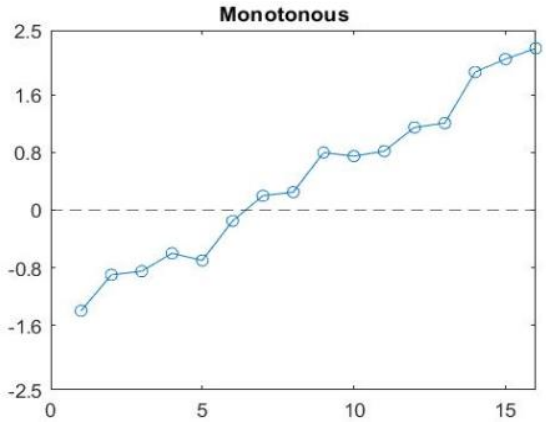
Εικόνα 4: Διαδοχικά σημεία από την ίδια πλευρά της κεντρικής γραμμής.

Όταν η διαδικασία είναι σε στατιστικό έλεγχο και η κατανομή της στατιστικής δείγματος είναι συμμετρική ως προς την κεντρική γραμμή, το κάθε σημείο του διαγράμματος είναι πάνω από την κεντρική γραμμή με πιθανότητα 0.5. Συνεπώς η πιθανότητα να δημιουργηθεί σειρά από 7 ή περισσότερα συνεχόμενα σημεία από την μία πλευρά της κεντρικής γραμμής είναι μικρότερη από 1% (0.5^7=0.0078) όταν η διαδικασία είναι πράγματι σε στατιστικό έλεγχο. Παραγωγικά εξάγεται το συμπέρασμα ότι κάποια συστηματική αιτία οδήγησε την διαδικασία εκτός στατιστικού ελέγχου [2].

Οι χρονοσειρές που οφείλονται σε συνεχή μεταβολή μιας παραμέτρου χαρακτηρίζονται γενικά από μη τυχαίες διατάξεις των σημείων στο διάγραμμα ελέγχου. Οι κυριότερες μορφές χρονοσειρών που εμφανίζονται σε διαγράμματα ελέγχου είναι οι εξής:

<u>Τάση αύξουσα ή φθίνουσα (**monotonous trend**)</u>

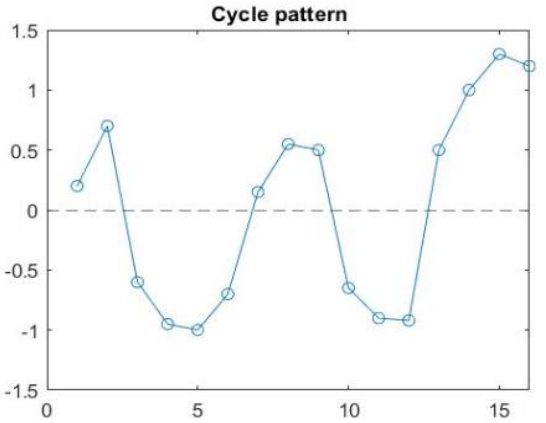
Διαδοχικά σημεία αποκλίνουν όλο και περισσότερο προς την ίδια πλευρά της κεντρικής γραμμής.



Εικόνα 5: Μονότονη αύξουσα τάση.

<u>Περιοδικότητα – επαναλαμβανόμενοι κύκλοι (**cycle pattern**)</u>

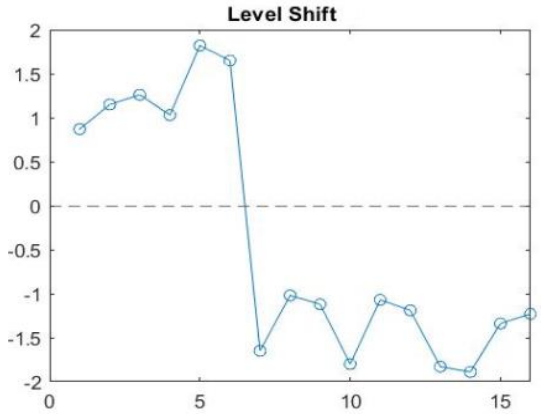
Εναλλασσόμενα σημεία με περιοδικότητα γύρω από την κεντρική γραμμή.



Εικόνα 6: Κυκλική τάση.

<u>Άλματα αλλαγής επιπέδου (**level shift**)</u>

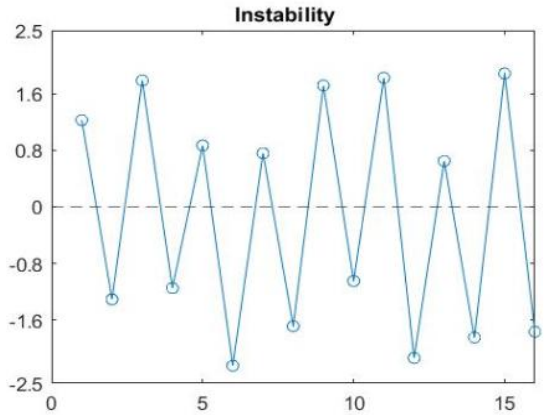
Ομάδες σημείων εκατέρωθεν της κεντρικής γραμμής στων οποίων την ένωση παρουσιάζεται άλμα στο μέσο επίπεδο τιμών.



Εικόνα 7: Άλμα τιμών.

<u>Μεγάλο ποσοστό σημείων κοντά στα όρια ελέγχου – μεγάλη διακύμανση (**instability**)</u>

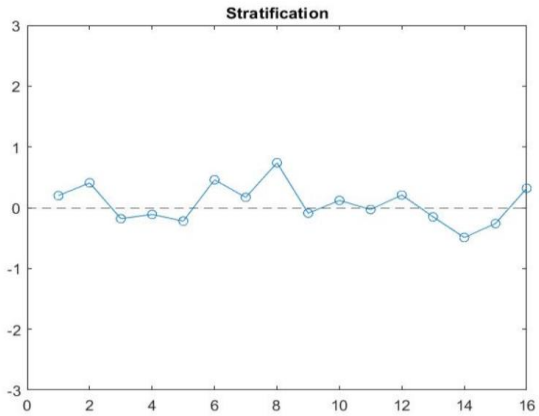
Διαδοχικά σημεία που εναλλάσσονται εκατέρωθεν της κεντρικής γραμμής κοντά στα όρια ελέγχου.



Εικόνα 8: Μεγάλο ποσοστό σημείων κοντά στα όρια ελέγχου.

<u>Έλλειψη μεταβλητότητας (**stratification**)</u>

Απουσία της αναμενόμενης διασποράς σύμφωνα με την φυσική μεταβλητότητα της διαδικασίας.



Εικόνα 9: Διαδοχικά σημεία σε μικρή απόσταση από την κεντρική γραμμή.

Ο εντοπισμός χρονοσειρών, σε διαγράμματα ποιότητας, που παραπέμπουν σε επίδραση συστηματικής αιτίας, αποτελεί πρόσφορο έδαφος ανάπτυξης μεθόδων, που στοχεύουν στην αυτοματοποίηση της διαδικασίας. Τα ΤΝΔ και η μέθοδος knn που εφαρμόζονται στην παρούσα εργασία ανήκουν σε αυτές.

2.2 Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα

2.2.1 Γενικά

Η σύγκριση των ικανοτήτων του ανθρώπινου εγκεφάλου με αυτές ενός υπολογιστή είναι συνήθης και διαχρονική. Οι βιολογικοί εγκέφαλοι είναι πολύ ευπροσάρμοστοι. Μπορούν να μαθαίνουν μέσω της εμπειρίας, να αξιολογούν προκειμένου να αποφασίσουν όταν αντιμετωπίζουν πρωτόγνωρες συνθήκες και να γενικεύουν. Σε αντίθεση οι υπολογιστές είναι ικανότατοι στο να εκτελούν ταχύτατα υπολογισμούς και εργασίες για τις οποίες έχουν προγραμματιστεί.

Ο ανθρώπινος εγκέφαλος περιέχει κατά προσέγγιση 100 δισεκατομμύρια νευρικά κύτταρα, 3.2 εκατομμύρια χιλιόμετρα «καλωδίων», ένα 10^{15} συνδέσεις, σε χώρο 1.5 λίτρου με βάρος 1.5 kg και κατανάλωση μόλις 10 Watt. Η οποιαδήποτε σύγκριση με ανθρώπινο κατασκεύασμα φαντάζει άτοπη, ωστόσο μία απόπειρα προσέγγισης πραγματοποιείται μέσω των Τεχνητών Νευρωνικών Δικτύων.

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) είναι απλοποιημένα υπολογιστικά μοντέλα που αναπτύχθηκαν βασισμένα στην λειτουργία του κεντρικού νευρικού συστήματος του ανθρώπου. Συνιστούν ουσιαστικά συστήματα επεξεργασίας της πληροφορίας που μιμούνται τις αρχές και την δομή των βιολογικών νευρώνων του εγκεφάλου. Κατά επέκταση ένα ΤΝΔ αποτελεί ένα σύνολο από υπολογιστικούς κόμβους (νευρώνες), δηλαδή δομικά στοιχεία επεξεργασίας πληροφοριών, με κύριο χαρακτηριστικό του την ικανότητα να ανταποκρίνεται σε ερεθίσματα που δέχεται στην είσοδό του αλλά και να προσαρμόζεται σε αυτά ώστε να ανταποκρίνεται μεταγενέστερα με μεγαλύτερη ακρίβεια και ταχύτητα. Αντικατοπτρίζοντας έτσι εκ θεμελίων την διαδικασία της μάθησης στον εγκέφαλο ενός βιολογικού όντος, επεκτείνεται και στις παράγωγες δυνατότητές του όπως η γενίκευση, η μνήμη και η ομαδοποίηση προτύπων, η οποίες

αποτελούν και τα κύρια πλεονεκτήματά του. Ταυτόχρονα όμως ακολουθείται και από την πολυπλοκότητα της λειτουργίας του ανθρώπινου εγκεφάλου, καθιστώντας σχεδόν αδύνατη την παρακολούθηση της ακριβούς διαδικασίας λήψης της κάθε απόφασης.

Παραλληλίζοντας τις διαδικασίες, η μάθηση σε ένα ΤΝΔ παίρνει πολύ χρόνο σε σχέση με τον βιολογικό εγκέφαλο, καθώς ενώ υπερέχει σε ταχύτητα διάδοσης του σήματος υστερεί καθοριστικά σε πολυπλοκότητα.

2.2.2 Ιστορική αναδρομή

Το πρώτο μοντέλο ΤΝΔ, το οποίο εξ ορισμού βασίζει την λειτουργία του σε θεμελιακές υπολογιστικές μονάδες προτείνεται από τους McCulloch και Pitts το 1943. Στην εργασία τους παρουσιάζεται για πρώτη φορά η ιδέα ενός δικτύου που απαρτίζεται από πλήθος αλληλοσυνδεόμενων κόμβων αλλά και ο τρόπος λειτουργίας του.

Λίγα χρόνια αργότερα πάνω στην εργασία τους βασίζονται οι Rudolph Ortvay και J. Von Neumann. Ο πρώτος πρότεινε ότι «υπάρχει μία σύνδεση μεταξύ του εγκεφάλου και των ηλεκτρονικών υπολογιστικών μηχανών» ενώ ο δεύτερος προχώρησε στην μελέτη αυτής της σύνδεσης για να συσχετίσει υπολογιστικά συστήματα και ζωντανούς οργανισμούς.

Το 1949 ο D. O. Hebb στο βιβλίο του «The Organization of Behavior» δημιουργεί μια υπόθεση μάθησης βασισμένη στον μηχανισμό της νευροπλαστικότητας, δηλαδή της ικανότητας του κεντρικού νευρικού συστήματος να αλλάζει τη δομή και τη λειτουργία του με σκοπό την αφομοίωση νέων πληροφοριών, που έγινε γνωστή ως κανόνας μάθησης του Hebb. Το μοντέλο του Hebb έχει ως βάση τις διασυνδέσεις των μονάδων του συστήματος, τους νευρώνες. Ο κανόνας ορίζει ότι κάθε φορά που το δίκτυο χρησιμοποιεί τις νευρωνικές του συνδέσεις, οι συνδέσεις αυτές ενισχύονται και το δίκτυο γίνεται όλο και αποτελεσματικότερο στην διεκπεραίωση της εκάστοτε εργασίας. Οι Farley and Wesley A. Clark ήταν οι πρώτοι που χρησιμοποίησαν υπολογιστικές μηχανές προσομοιώνοντας το δίκτυο του Hebb.

Μετά το πέρας σχεδόν μίας δεκαετίας ο Frank Rosenblatt δημιούργησε τον αλγόριθμο perceptron, ο οποίος αφορά δυαδική κατηγοριοποίηση μέσω επιτηρούμενης μάθησης. Η δυαδική κατηγοριοποίηση πραγματοποιείται όταν η είσοδος, αποτελούμενη από ένα διάνυσμα αριθμών, στο σύστημα αξιολογείται μέσω κάποιας συνάρτησης ως προς το αν ανήκει σε κάποια συγκεκριμένη κλάση ή όχι. Η εργασία του αυτή ευοδώθηκε με την κατασκευή του Mark I Perceptron το 1960, ουσιαστικά τον πρώτο υπολογιστή που μπορούσε να μαθαίνει νέες δεξιότητες μέσω δοκιμής και λάθους, χρησιμοποιώντας έναν τύπο νευρωνικού δικτύου που προσομοιώνει την ανθρώπινη διαδικασία σκέψης. [5]

Την ίδια περίοδο οι Henry J. Kelley (1960) και Arthur E. Bryson (1961) ασχολήθηκαν με την οπισθόδρομη διάδοση σφάλματος χρησιμοποιώντας αρχές δυναμικού προγραμματισμού. Το σφάλμα είναι η απόκλιση μεταξύ της εξόδου του νευρωνικού δικτύου κάθε στιγμή και της επιθυμητής εξόδου και στόχος αποτελεί η μείωση του μέσα από μία σειρά διορθωτικών αλλαγών που διατρέχουν οπισθόδρομα το δίκτυο.

Τα πρώτα λειτουργικά δίκτυα με πολλαπλά επίπεδα παρουσιάστηκαν από τους Ivakhnenko και Lapa το 1965. Μία δεκαετία σχεδόν αργότερα οι Dreyfus και Werbos βασίστηκαν πάνω στην οπισθόδρομη διάδοση σφάλματος (backpropagation), ο πρώτος

για να προσαρμόσει τις παραμέτρους ελεγκτών αναλογικά με την κλίση του σφάλματος και ο δεύτερος για να μορφοποιήσει έναν αλγόριθμο που επιτρέπει την εκπαίδευση τέτοιων δικτύων.

Το 1986 παρουσιάστηκε ολοκληρωμένη η μέθοδος της οπισθοδιάδοσης σφάλματος από τους McClelland και Rumelhart, η οποία είναι πλέον η πιο διαδεδομένη μέθοδος εκπαίδευσης δικτύων. Στην εργασία τους «Parallel Distributed Processing» προτείνεται η ιδέα ότι ένα νευρωνικό δίκτυο μπορεί να χρησιμοποιηθεί ως παράλληλος επεξεργαστής. Το έργο αυτό κάνει ένα σημαντικό βήμα πέρα από το Perceptron, με το να επιτρέπει την ύπαρξη και άλλων επιπέδων νευρώνων, εκτός της εισόδου και της εξόδου, που διαρθρώνουν την εσωτερική δομή του δικτύου.

Το επόμενο σημαντικό βήμα πραγματοποιήθηκε το 1992 και αφού η υπολογιστική ισχύς είχε αυξηθεί αρκετά, με την εισαγωγή της τεχνικής μέγιστης ομαδοποίησης (max-pooling). Αυτή αφορά την επεξεργασία σήματος και κυρίως την αναγνώριση εικόνας μέσω της κατάτμησής της σε μικρότερες εικόνες. Ο Schmidhuber την ίδια εποχή χρησιμοποίησε δίκτυα πολλών επιπέδων εκπαιδεύοντας ένα επίπεδο κάθε φορά μέσω οπισθόδρομης διάδοσης σφάλματος.

Από το 1985 αρχίζουν και τα πρώτα συνέδρια που είναι αφιερωμένα αποκλειστικά σε νευρωνικά δίκτυα από την American Physical Society και από την IEEE. Ταυτόχρονα δημιουργούνται και διεθνείς οργανισμοί όπως ο INNS (International Neural Network Society).

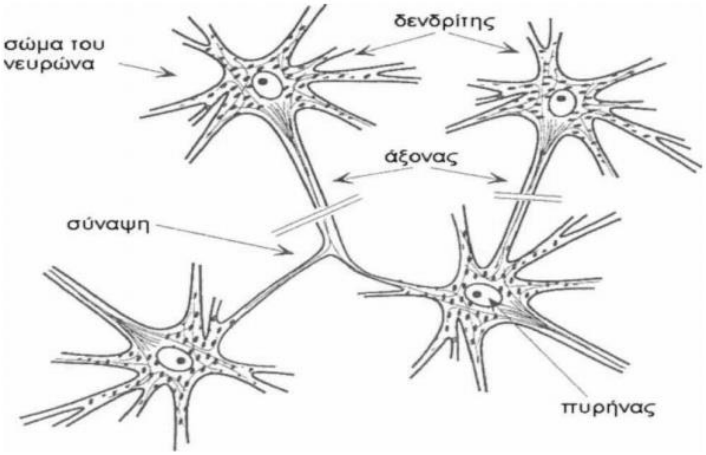
Το 2012 οι Andrew Yan-Tak Ng και Jeffrey Adgate Dean δημιούργησαν ένα δίκτυο ικανό να αναγνωρίζει υψηλών απαιτήσεων μοτίβα όπως εικόνες από γάτες. Η συνεχώς αυξανόμενη υπολογιστική ισχύς επέτρεψε την χρήση όλο και μεγαλύτερων δικτύων κυρίως στους τομείς αναγνώρισης εικόνας και ήχου δημιουργώντας τον κλάδο της βαθιάς μάθησης (deep learning).

Πλέον τα τεχνητά νευρωνικά δίκτυα αποτελούν αναπόσπαστο κομμάτι της τεχνητής νοημοσύνης που έχει ξεπεράσει τόσο ανθρώπινες επιδόσεις όσο και τις επιδόσεις προγραμμάτων που επωφελούνται μόνο από την υπολογιστική ισχύ. Χαρακτηριστικό παράδειγμα το πρόγραμμα AlphaZero της εταιρίας Deepmind που τον Δεκέμβριο του 2017 κατάφερε μέσα σε 24 ώρες να εκπαιδευτεί και να κυριαρχήσει στα παιχνίδια σκάκι, shogi και go νικώντας τα προγράμματα που ήταν πρωταθλητές μέχρι εκείνη την στιγμή. [6]

2.2.3 Βιολογικά Νευρωνικά Δίκτυα

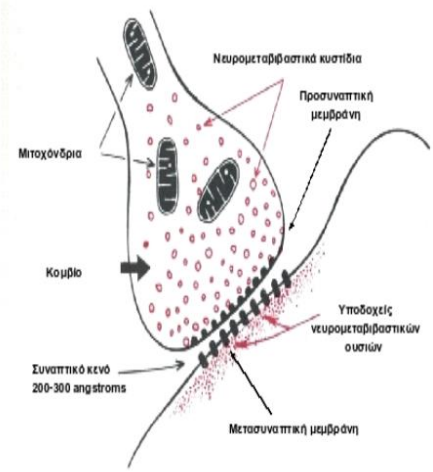
Ο ανθρώπινος εγκέφαλος αποτελείται από περίπου 100 δισεκατομμύρια νευρικά κύτταρα που ονομάζονται νευρώνες. Ο νευρώνας αποτελεί τη βασική λειτουργική μονάδα του νευρικού συστήματος. Όλοι οι νευρώνες είναι διαφορετικοί μεταξύ τους και δεν υπάρχουν μέσα σε αυτό το πλήθος ούτε δύο ολόιδιοι. Το κεντρικό νευρικό σύστημα είναι ικανό να αλλάζει τη δομή και τη λειτουργία του ως απάντηση στη μάθηση νέων πληροφοριών, ικανότητα που ονομάζεται νευροπλαστικότητα. Η νευροπλαστικότητα θεωρείται ότι αποτελεί το νευρολογικό υπόβαθρο της μάθησης.

Ο κάθε νευρώνας αποτελείται από τρία τμήματα, τις διακλαδισμένες αποφύσεις ή δενδρίτες (dendrites), το σώμα (soma) και την αξονική απόφυση (axon).



Εικόνα 10: Βιολογικοί νευρώνες και τα μέρη από τα οποία αποτελούνται [7].

Οι νευρώνες επικοινωνούν μεταξύ τους μέσω ηλεκτρικών σημάτων που εμφανίζονται ως παλμοί και διαβιβάζονται με την διαμεσολάβηση ηλεκτροχημικών συνδέσεων που ονομάζονται συνάψεις (synapses) και παρουσιάζονται στις διακλαδισμένες αποφύσεις του κυττάρου, τους δενδρίτες (dendrites). Ο κάθε νευρώνας έχει περί τις 10 χιλιάδες συνάψεις. Κατά τα 2 πρώτα χρόνια ζωής του ανθρώπου δημιουργούνται περίπου 1 εκατομμύριο συνάψεις το δευτερόλεπτο. Κάθε φορά που αποκτάται νέα γνώση, η συναπτική διαβίβαση μεταξύ των νευρώνων που συμμετέχουν είτε ενισχύεται είτε καταστρέφεται για να ξαναδημιουργηθεί βελτιωμένη. Καλύτερη επικοινωνία μεταξύ των νευρώνων σημαίνει ότι τα ηλεκτρικά σήματα ταξιδεύουν πιο αποτελεσματικά κατά μήκος του νέου δρόμου.



Εικόνα 11: Η συναπτική σύνδεση μεταξύ δύο νευρώνων [8].

Υπεύθυνη για τη διαβίβαση είναι μία χημική ουσία, η οποία ονομάζεται νευροδιαβιβαστής (neurotransmitter) και που μπορεί να λειτουργεί είτε ανασταλτικά

είτε διεγερτικά για κάθε σήμα. Το κάθε νευρικό κύτταρο πραγματοποιεί χιλιάδες συνδέσεις με τα άλλα κύτταρα οπότε και δέχεται συνεχώς ένα μεγάλο πλήθος από εισερχόμενα ηλεκτρικά σήματα τα οποία τελικά καταφθάνουν στο κυρίως σώμα του κυττάρου (cell body).

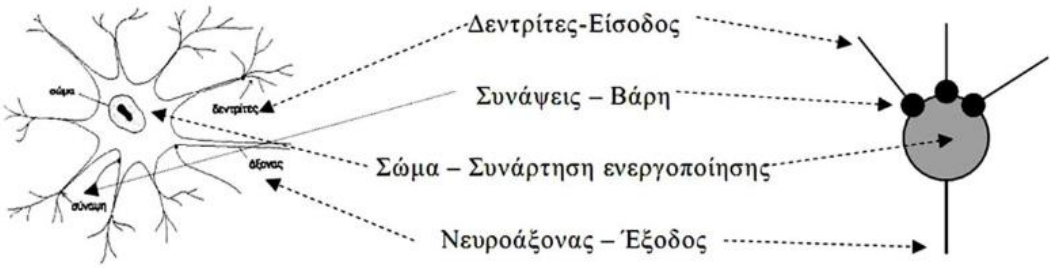
Στο σώμα αθροίζονται τα εισερχόμενα σήματα και αν το αποτέλεσμα ξεπερνά μια οριακή τιμή, διαφορετική για κάθε κύτταρο τότε ο νευρώνας ενεργοποιείται δημιουργώντας έναν ηλεκτρικό παλμό που διαδίδεται διαμέσου της αξονικής απόφυσης προς τις συνάψεις με τους επόμενους νευρώνες που είναι συνδεδεμένοι με αυτόν. Ορισμένα από τα σήματα εισόδου συνεισφέρουν αρνητικά στην ενεργοποίηση του κυττάρου ενώ άλλα θετικά. Αν εξεταστεί κάθε χρονική στιγμή ένας νευρώνας μπορεί να είναι είτε ενεργοποιημένος είτε όχι χωρίς αυτό να σημαίνει όμως ότι η συχνότητα ενεργοποίησής του είναι σταθερή.

Ο κάθε νευρώνας σαν μονάδα είναι ικανός να λαμβάνει, να τροποποιεί και να στέλνει σήματα. Όταν όμως πολλοί νευρώνες συνενωθούν τα αποτελέσματα είναι εντυπωσιακά. Εξάλλου ο ανθρώπινος εγκέφαλος αποτελεί ένα δίκτυο από δισεκατομμύρια τέτοιων νευρώνων ενωμένων μεταξύ τους.

2.2.4 Τεχνητά Νευρωνικά Δίκτυα

Η αρχιτεκτονική και ο τρόπος λειτουργίας των βιολογικών νευρώνων αποτελεί το πρότυπο για την κατασκευή των τεχνητών τους ισοδύναμων.

Σε ένα ΤΝΔ η επεξεργασία της πληροφορίας γίνεται σε πολλές απλές μονάδες, οι οποίες ονομάζονται νευρώνες. Ένας τεχνητός νευρώνας δέχεται στην είσοδό του πολλά διαφορετικά σήματα που διαβιβάζονται μεταξύ των νευρώνων μέσω συνδέσεων (connection links). Κάθε σύνδεση έχει έναν αντιστοιχισμένο συντελεστή βαρύτητας (weight), ο οποίος πολλαπλασιάζει το σήμα που διέρχεται από αυτήν και κατά αντιστοιχία με τον ρόλο της σύναψης στον βιολογικό νευρώνα, η τιμή του βάρους μπορεί να είναι είτε θετική ενισχύοντας την είσοδο είτε αρνητική εξασθενίζοντάς την.

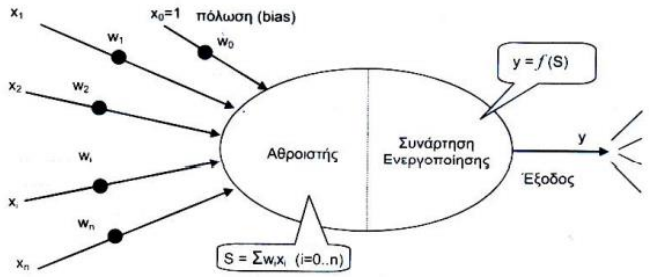


Εικόνα 12: Αντιστοιχία βιολογικού με τεχνητού νευρώνα [8].

Το σώμα του τεχνητού νευρώνα χωρίζεται σε δύο μέρη, τον αθροιστή, όπου αθροίζονται τα σταθμισμένα σήματα και τη συνάρτηση ενεργοποίησης, η οποία δέχεται ως όρισμα το αποτέλεσμα του αθροιστή και καθορίζει αν θα ενεργοποιηθεί ή όχι ο νευρώνας ανάλογα με το κατώφλι της και στην περίπτωση που ενεργοποιηθεί, ποια θα είναι η έξοδός του. Ένας ενεργοποιημένος νευρώνας εκπέμπει ένα σήμα, το οποίο είναι η τιμή

της συνάρτησης ενεργοποίησης του και ονομάζεται δραστηριότητα (activation) του νευρώνα. Διευκρινίζεται ότι ένας νευρώνας μπορεί να εκπέμψει ένα μόνο σήμα σε κάθε χρονική στιγμή και ότι το σήμα αυτό εκπέμπεται προς πολλούς άλλους νευρώνες ταυτόχρονα αλλά η επίδρασή του σε κάθε έναν από αυτούς είναι διαφορετική καθώς σταθμίζεται με διαφορετικούς συντελεστές βαρύτητας.

Παρακάτω παρουσιάζεται η δομή ενός νευρώνα με n συνδέσεις στην είσοδό του.



Εικόνα 13: Μοντέλο τεχνητού νευρώνα [9].

Όπως φαίνεται και στην Εικόνα 13 ένας νευρώνας δέχεται τα σήματα x_i στις συνδέσεις της εισόδου του (i=1,2,…,n) κάθε ένα από τα οποία πολλαπλασιάζεται με τον αντίστοιχο συντελεστή βαρύτητας w_i . Επίσης δέχεται μια εξωτερικά εφαρμοζόμενη πόλωση (bias) ίση με το συναπτικό βάρος w_0 της σταθερής εισόδου x_0=1. Η πόλωση είναι εξωτερική παράμετρος του νευρώνα και έχει ως αποτέλεσμα την αύξηση ή την μείωση της διέγερσης της συνάρτησης ενεργοποίησης, ανάλογα με τον αν είναι θετική ή αρνητική αντίστοιχα. Συνεπώς η συνολική είσοδος είναι:

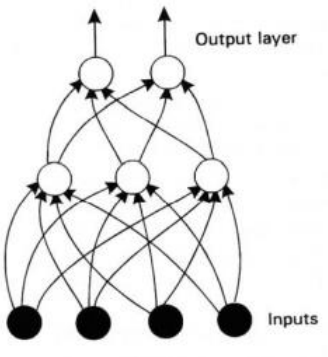
$$x_{in} = w_0 + \sum_{i=1}^{n} w_i x_i$$

Το άθροισμα αυτό εισάγεται στην συνάρτηση ενεργοποίησης του νευρώνα, η οποία έστω ότι είναι η $y = f(x)$ και το αποτέλεσμα στην έξοδο του είναι:

$$y = f(x_{in})$$

Παρά το γεγονός ότι το εκπεμπόμενο σήμα είναι το ίδιο προς όλους τους νευρώνες που συνδέονται με αυτόν, πολλαπλασιάζεται με διαφορετικούς συντελεστές βαρύτητας σε κάθε έναν από αυτούς και έχει κατά συνέπεια διαφορετική επίδραση.

Ένα ΤΝΔ αποτελεί ουσιαστικά ένα σύμπλεγμα από τέτοιους νευρώνες που επικοινωνούν μεταξύ τους. Μία απλή περίπτωση τέτοιας δομής παρουσιάζεται στην Εικόνα 14:



Εικόνα 14: Νευρωνικό δίκτυο με τρία επίπεδα [10].

Ο τρόπος με τον οποίο οργανώνονται οι νευρώνες σε επίπεδα (τουλάχιστον δύο, ένα εισόδου και ένα εξόδου) ονομάζεται αρχιτεκτονική του δικτύου (network architecture) και αποτελεί ίσως το πιο σημαντικό γνώρισμα ενός νευρωνικού δικτύου γιατί ανάλογα με αυτόν καθορίζονται και οι δυνατότητες που έχει το νευρωνικό δίκτυο.

Μία εικόνα από αυτές τις υπολογιστικές δυνατότητες δίνεται από την μελέτη των Hornik et al. που απέδειξαν ότι, με κατάλληλη προσαρμογή των βαρών και επαρκή αριθμό κόμβων, ένα νευρωνικό δίκτυο τριών επιπέδων μπορεί να προσεγγίσει οποιαδήποτε μη γραμμική συνάρτηση [10]. Αναλογιζόμενοι ότι τα προβλήματα κατηγοριοποίησης και πρόβλεψης μπορούν να ιδωθούν και ως προβλήματα προσέγγισης συναρτήσεων (function approximation) η διαπίστωση αυτή είναι αποκαλυπτική ως προς την δυναμική των ΤΝΔ.

Επαγωγικά καθοριστικό στοιχείο, που επηρεάζει την απόδοση του δικτύου είναι και οι τιμές των συντελεστών βαρύτητας στις συνδέσεις του δικτύου. Οι τιμές αυτές διαμορφώνονται κατά την εκμάθηση πρότυπων παραδειγμάτων, μία εγγενή ικανότητα του όπως προαναφέρθηκε. Κατά συνέπεια ένας δεύτερος κρίσιμος παράγοντας είναι η μέθοδος που χρησιμοποιείται για να καθορίζονται οι τιμές αυτών των συντελεστών, η οποία ονομάζεται αλγόριθμος εκπαίδευσης.

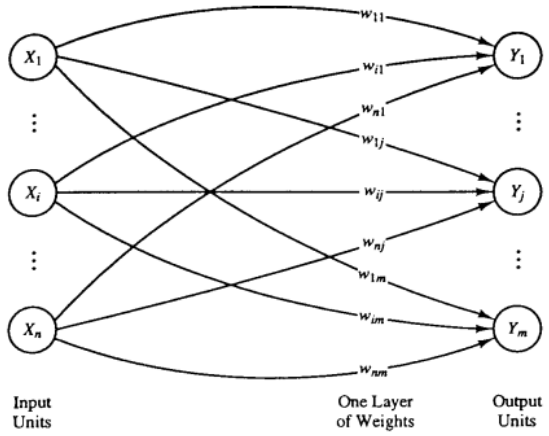
Τρίτον σημαντική επιρροή στο αποτέλεσμα έχει και ο τύπος της χρησιμοποιούμενης συνάρτησης ενεργοποίησης των νευρώνων.

Λόγω της σπουδαιότητας των στοιχείων αυτών, αναλύονται λεπτομερέστερα σε ξεχωριστή ενότητα το κάθε ένα.

2.2.4.1 Αρχιτεκτονική

Στην πλειοψηφία των ΤΝΔ, οι νευρώνες οργανώνονται σε επίπεδα όπως στην περίπτωση της Εικόνας 14. Οι νευρώνες που βρίσκονται στο ίδιο επίπεδο συμπεριφέρονται όμοια τις περισσότερες περιπτώσεις.

Η απλούστερη περίπτωση είναι το ΤΝΔ να αποτελείται από ένα επίπεδο εισόδου και ένα επίπεδο εξόδου όπως φαίνεται στην Εικόνα 15.



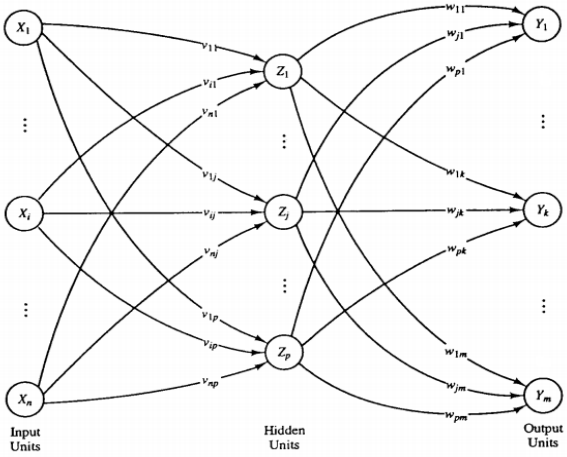
Εικόνα 15: ΤΝΔ απλού επιπέδου [11].

Το επίπεδο εισόδου χρησιμοποιείται για να εισαχθεί το σήμα στο δίκτυο, δηλαδή δεν υπάρχει πολλαπλασιασμός με συντελεστές βαρύτητας. Για τον λόγο αυτό στον προσδιορισμό του αριθμού των επιπέδων, το επίπεδο εισόδου εξαιρείται. Στην Εικόνα 15 ο αριθμός των νευρώνων εισόδου και εξόδου είναι n και m αντίστοιχα. Ο πίνακας W των συντελεστών βαρύτητας έχει συνεπώς διαστάσεις n×m και το διάνυσμα b των σταθερών πόλωσης 1×m. Το διάνυσμα των αποτελεσμάτων Y έχει m στοιχεία, οι τιμές των οποίων προκύπτουν ως εξής:

$$Y_{1\times m} = f(X_{1\times n} * W_{n\times m} + b_{1\times m})$$

όπου f η συνάρτηση ενεργοποίησης των νευρώνων του επιπέδου εξόδου.

Εκτός από τα νευρωνικά δίκτυα απλού επιπέδου (single layer) υπάρχουν και τα πολλαπλών επιπέδων (multilayer). Στην πλειοψηφία των τελευταίων ο κάθε νευρώνας ενός επιπέδου είναι συνδεδεμένος με τους νευρώνες του προηγούμενου και του επόμενου επιπέδου αλλά όχι με τους νευρώνες που βρίσκονται στο ίδιο επίπεδο με αυτόν. Ένα δίκτυο πολλαπλών επιπέδων απεικονίζεται στην Εικόνα 16.



Εικόνα 16: ΤΝΔ πολλαπλών επιπέδων [11].

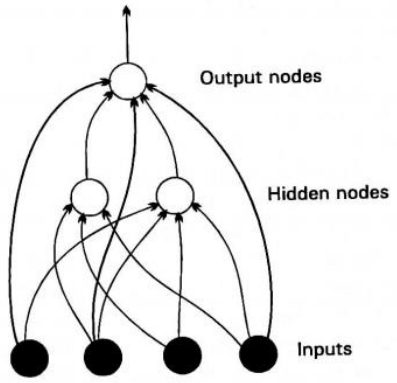
Τα ΤΝΔ πολλαπλών επιπέδων εμπεριέχουν τουλάχιστον ένα κρυμμένο επίπεδο. Στην περίπτωση που το κρυμμένο επίπεδο είναι ένα, όπως στην Εικόνα 16, με αριθμό νευρώνων p τότε ακολουθώντας την ίδια λογική με προηγουμένως το αποτέλεσμα εξόδου είναι:

$$Y_{1\times m} = f_2\left(f_1(X_{1\times n} * W1_{n\times p} + b1_{1\times p}) * W2_{p\times m} + b2_{1\times m}\right)$$

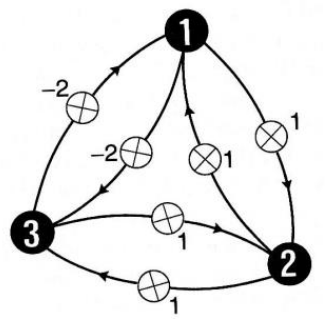
όπου f_1 και f_2 η συνάρτηση ενεργοποίησης, b1 και b2 το διάνυσμα των σταθερών πόλωσης και W1 και W2 οι πίνακες των συντελεστών βαρύτητας, του κρυμμένου επιπέδου και του επιπέδου εξόδου αντίστοιχα.

Οι παραπάνω περιπτώσεις ανήκουν στην κατηγορία των δικτύων πρόσθιας τροφοδότησης (feedforward), τα οποία χαρακτηρίζονται από έλλειψη ανάδρασης στις εισόδους τους δηλαδή το σήμα ρέει από το επίπεδο εισόδου προς το επίπεδο εξόδου και οι νευρώνες σε κάθε επίπεδο έχουν ως εισόδους μόνο τα σήματα εξόδου του προηγουμένου επιπέδου.

Υπάρχουν όμως και πιο πολύπλοκες αρχιτεκτονικές, όπως στην περίπτωση των δικτύων που παρακάμπτουν ένα ή περισσότερα κρυμμένα επίπεδα (by passing) όπως φαίνεται στην Εικόνα 17: ΤΝΔ με παρακάμψεις ., ή των αναδρομικών δικτύων (recurrent neural network) όπως η περίπτωση του δικτύου Hopfield της Εικόνας 18, τα οποία δέχονται ανάδραση από τα αποτελέσματά τους. Αυτό σημαίνει ότι τουλάχιστον σε έναν νευρώνα, το σήμα εξόδου επηρεάζει το σήμα που εισέρχεται στην είσοδό του. Συγκεκριμένα τα δίκτυα τύπου Hopfield αποτελούνται από μόνο ένα επίπεδο νευρώνων, το οποίο λειτουργεί και ως επίπεδο εισόδου και ως επίπεδο εξόδου, όπου κάθε νευρώνας συνδέεται με όλους τους υπολοίπους αλλά όχι με τον εαυτό του. Επιπλέον, οι συντελεστές βαρύτητας είναι συμμετρικοί μεταξύ τους.



Εικόνα 17: ΤΝΔ με παρακάμψεις [11].



Εικόνα 18: ΤΝΔ τύπου Hopfield με τρείς νευρώνες [11].

2.2.4.2 Εκπαίδευση

Ένα νευρωνικό δίκτυο μαθαίνει για το περιβάλλον του μέσω μίας επαναληπτικής διαδικασίας μεταβολής των συντελεστών βαρύτητας και των σταθερών πόλωσης στις συνδέσεις του. Οι μεταβλητές αυτές συνιστούν τα σημεία στα οποία αποθηκεύονται οι πληροφορίες που χρησιμοποιεί το νευρωνικό δίκτυο, για να λύσει το πρόβλημα που του τίθεται και συμπυκνώνουν όλη την γνώση που αυτό διαθέτει για το περιβάλλον του. Οι βέλτιστες τιμές τους διαφέρουν αναλόγως με το πρόβλημα και δεν μπορούν να υπολογιστούν με ακρίβεια λόγω του πλήθους των αγνώστων. Μπορούν όμως να βελτιώνονται δίνοντας ακριβέστερα αποτελέσματα. Συνεπώς εκπαίδευση ονομάζεται η διαδικασία μεταβολής αυτών των τιμών, προς την κατεύθυνση της βελτιστοποίησης της απόδοσης. Μετά την ολοκλήρωσή της δεν υφίσταται καμία μεταβολή τους.

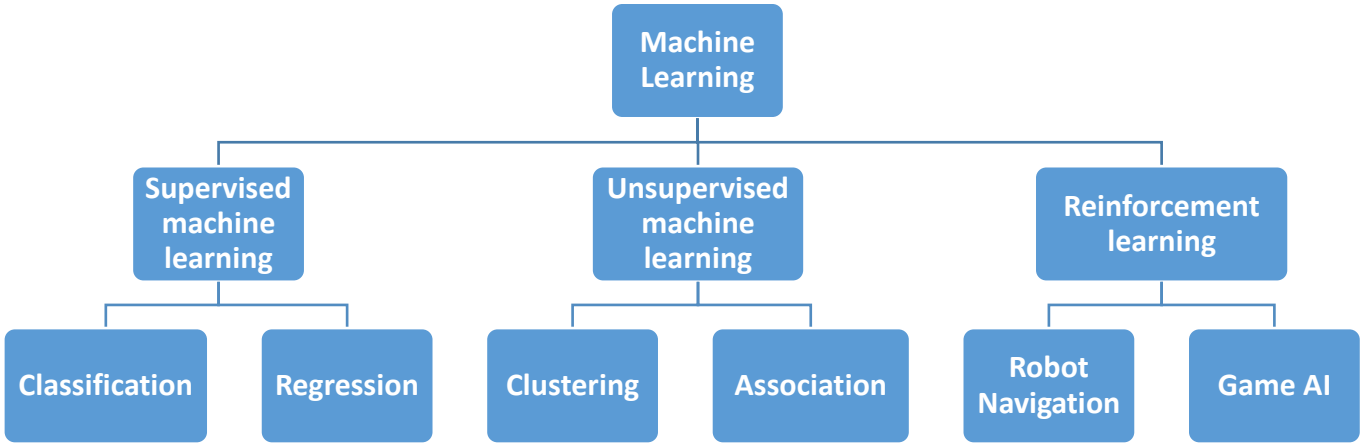
Αναλυτικότερα, ο απώτερος σκοπός της εκπαίδευσης, είναι να επιτευχθεί μία ισορροπία ανάμεσα στην ικανότητα του ΤΝΔ να αποκρίνεται σωστά στα διανύσματα εισόδου (απομνημόνευση) και στην ικανότητα να δίνει ικανοποιητικές αποκρίσεις σε διανύσματα που είναι παρόμοια, αλλά όχι ολόιδια, με αυτά που χρησιμοποιήθηκαν κατά την εκπαίδευση (γενίκευση). Η ισορροπία αυτή εξαρτάται από την πολυπλοκότητα τόσο της αρχιτεκτονικής, όσο και της βάσης δεδομένων εκπαίδευσης και αν δεν επιτευχθεί τότε υπάρχουν τα εξής ενδεχόμενα:

1. Το δίκτυο διατηρεί την ικανότητα της απομνημόνευσης δηλαδή ανταποκρίνεται σωστά όταν του παρουσιάζονται δεδομένα με τα οποία έχει εκπαιδευτεί αλλά έχει απολέσει την ικανότητα γενίκευσης δηλαδή αποκρίνεται άστοχα σε δεδομένα τα οποία διαφέρουν έστω μερικώς από αυτά με τα οποία εκπαιδεύτηκε. Το φαινόμενο αυτό ονομάζεται «overfitting», που σημαίνει ότι το δίκτυο είναι υπερβολικά προσαρμοσμένο στα δεδομένα εκπαίδευσης, τόσο που δεν είναι σε θέση να γενικεύσει. Οι συνηθέστεροι τρόποι επίλυσης του είναι η μείωση της πολυπλοκότητας της αρχιτεκτονικής (π.χ. με την αφαίρεση ενός επιπέδου), και η αύξηση του όγκου των δεδομένων εκπαίδευσης (π.χ. αύξηση του αριθμού των διανυσμάτων).
2. Το δίκτυο αδυνατεί τόσο να γενικεύσει όσο και να απομνημονεύσει επαρκώς, δηλαδή δεν αποκρίνεται με επιτυχία στα δεδομένα με τα οποία έχει εκπαιδευτεί. Το φαινόμενο αυτό ονομάζεται «underfitting», που σημαίνει ότι δεν υπάρχει επαρκής προσαρμογή στα δεδομένα εκπαίδευσης και επιλύεται συνήθως με την

αύξηση της πολυπλοκότητας της αρχιτεκτονικής (π.χ. με την προσθήκη νευρώνων σε ένα επίπεδο) και την αύξηση της διάκρισης μεταξύ των δεδομένων εκπαίδευσης (π.χ. περισσότερες κατηγορίες ταξινόμησης).

Αν και η εκπαίδευση απαιτεί κάποιο μεγάλο χρονικό διάστημα, ένα εκπαιδευμένο νευρωνικό δίκτυο μπορεί να παράγει το διάνυσμα εξόδου του, ταχύτατα.

Διακρίνονται τρείς διαφορετικοί τρόποι εκπαίδευσης, η επιτηρούμενη (supervised), η μη επιτηρούμενη (non supervised) και η ενισχυτική (reinforcement). Στην πραγματικότητα η ενισχυτική μάθηση αποτελεί υποκατηγορία της μη επιτηρούμενης μάθησης αλλά παρουσιάζεται ξεχωριστά λόγω των ιδιαιτεροτήτων της. Τόσο η επιτηρούμενη όσο και η μη επιτηρούμενη μάθηση λόγω των δομικών διαφορών τους, είναι αποτελεσματικές σε συγκεκριμένες κατηγορίες προβλημάτων. Η επιτηρούμενη εκπαίδευση εφαρμόζεται σε προβλήματα παλινδρόμησης (regression) και κατηγοριοποίησης (classification) ενώ η μη επιτηρούμενη εκπαίδευση είναι κατάλληλη για την επίλυση προβλημάτων ομαδοποίησης (clustering) και συσχέτισης (association).



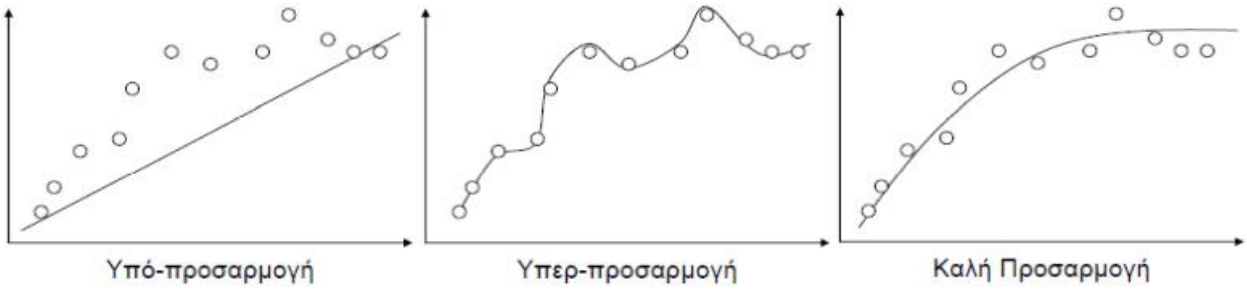
Εικόνα 19: Τρόποι εκπαίδευσης και κατηγορίες προβλημάτων στις οποίες εφαρμόζονται.

Επιτηρούμενη μάθηση

Σε αυτήν την περίπτωση τα διανύσματα εμφανίζονται σε ζευγάρια, εισόδου και επιθυμητής εξόδου. Το δίκτυο τροφοδοτείται με το κάθε διάνυσμα εισόδου και το αποτέλεσμα που δίνει συγκρίνεται με το εκάστοτε επιθυμητό. Η εκπαίδευση επικεντρώνεται κυρίως στην εύρεση με συστηματικό τρόπο, των συντελεστών βαρύτητας και σταθερών πόλωσης που ελαχιστοποιούν κάποιο μέτρο του σφάλματος μεταξύ της επιθυμητής και της πραγματικής εξόδου του δικτύου. Κατά την έναρξη της διαδικασίας, το αποτέλεσμα στην έξοδο διαφέρει αισθητά από το επιθυμητό, δηλαδή το σφάλμα (error) είναι μεγάλο. Ως μέτρο για την αυτήν την αξιολόγηση που αφορά την απόδοση χρησιμοποιείται μία συνάρτηση κόστους. Με βάση αυτό το σφάλμα αναπροσαρμόζονται τα βάρη και οι σταθερές πόλωσης σύμφωνα με μία μέθοδο η οποία ονομάζεται αλγόριθμος εκπαίδευσης και εξασφαλίζει ότι σε βάθος ορισμένων επαναλήψεων το σφάλμα μειώνεται. Ένα πέρασμα από όλα τα ζευγάρια διανυσμάτων

εισόδου και επιθυμητής εξόδου ονομάζεται εποχή. Η αναπροσαρμογή των βαρών και των σταθερών πόλωσης συμβαίνει είτε μετά την χρήση καθενός από τα διανύσματα εκπαίδευσης, όπου πρόκειται για εκπαίδευση ανά πρότυπο εισόδου (online training) είτε στο τέλος κάθε εποχής αφού αθροιστεί η μεταβολή που προκύπτει από κάθε διάνυσμα, όπου πρόκειται για εκπαίδευση ανά ομάδα προτύπων εισόδου (batch training).

Συνεπώς καθώς προχωράει η εκπαίδευση, η σύγκλιση μεταξύ της εκάστοτε εισόδου και της αντίστοιχης επιθυμητής εξόδου αυξάνεται και ο αλγόριθμος τερματίζει είτε όταν αυτή φτάσει σε ικανοποιητικά επίπεδα, είτε μετά από προκαθορισμένο αριθμό εποχών. Ο απαραίτητος αριθμός εποχών εξαρτάται από το εκάστοτε πρόβλημα.



Εικόνα 20: Χαρακτηριστικές περιπτώσεις εκπαιδευμένων ΤΝΔ [11].

Όπως αναφέρθηκε τα προβλήματα στα οποία εστιάζει η επιτηρούμενη εκπαίδευση είναι αυτά της παλινδρόμησης (regression) και κατηγοριοποίησης (classification).

Αμφότερα ανήκουν στην κατηγορία προσέγγισης συνάρτησης (function approximation) όπου κύριος στόχος είναι η απόκριση με ακρίβεια σε μία δεδομένη είσοδο ή διαφορετικά, η χαρτογράφηση της άγνωστης συνάρτησης f(x) που δίνει την επιθυμητή απάντηση y όταν τροφοδοτείται με μία δεδομένη μεταβλητή x, όπως στην εξίσωση $y = f(x)$. Η κύρια διαφορά τους είναι ότι στα προβλήματα παλινδρόμηση η έξοδος y είναι μία συνεχής αριθμητική τιμή. Αντιθέτως στα προβλήματα κατηγοριοποίησης η έξοδος είναι μία διακριτή μεταβλητή κατηγοριοποίησης. Στα τελευταία, η έξοδος συχνά είναι έχει την μορφή πιθανότητας, η οποία μετατρέπεται στην συνέχεια σε διακριτή μεταβλητή, προκειμένου να καταστεί δυνατή η κατηγοριοποίηση.

Εξαιτίας της χρησιμοποίησης της επιτηρούμενης εκπαίδευσης στην παρούσα εργασία οι δύο αυτές κατηγορίες προβλημάτων αξίζει να εξεταστούν αναλυτικότερα:

<u>Παλινδρόμηση (Regression)</u>

Η απόκριση σε αυτήν την κατηγορία είναι ένας πραγματικός αριθμός, που αντιπροσωπεύει συνήθως ποσότητα ή μέγεθος. Για παράδειγμα από μία βάση δεδομένων που συσχετίζει τιμές με τετραγωνικά σπιτιών, ζητείται η πρόβλεψη της τιμής κάποιων άλλων σπιτιών. Η απόκριση είναι μία συνεχής μεταβλητή. Το σφάλμα της αφορά την απόστασή της από την πραγματική τιμή.

Κατηγοριοποίηση (Classification)

Σε αυτήν την περίπτωση η απόκριση είναι μία διακριτή, μη συνεχής τιμή που απαντάει στο αν η κάθε κατηγορία ή αλλιώς κλάση του προβλήματος, αντιστοιχεί στην είσοδο ή όχι. Ο στόχος είναι η επιλογή της σωστής κατηγορίας ταξινόμησης της εισόδου μέσα από ένα πλήθος κατηγοριών. Για παράδειγμα από την ίδια βάση δεδομένων που συσχετίζει τιμές με τετραγωνικά σπιτιών, ζητείται η πρόβλεψη του αν κάποιο σπίτι κοστολογείται ως «Ακριβό» ή «Φτηνό». Η έξοδος είναι μία διακριτή τιμή, που ορίζει αν η κάθε κλάση απορρίπτεται ή γίνεται δεκτή, και ενδέχεται να παίρνει την μορφή 0 ή 1, με τις τιμές αυτές όμως, να είναι αντιπροσωπευτικές και όχι ποσοτικές. Συχνά στα προβλήματα κατηγοριοποίησης η έξοδος είναι η πιθανότητα να ανήκει η είσοδος στην κάθε κατηγορία. Σε αυτές τι περιπτώσεις είναι συνεχής μεταβλητή, που μπορεί να μετατραπεί σε διακριτή με την επιλογή της υψηλότερης πιθανότητας και της αντίστοιχης κατηγορίας. Στο παράδειγμα των σπιτιών, αν η έξοδος είναι 0.7 και 0.3 σημαίνει ότι το σπίτι έχει 70% πιθανότητες να είναι «Ακριβό» και 30% πιθανότητες να είναι «Φτηνό». Το σφάλμα της είναι το αν η επιλεγόμενη κλάση είναι η σωστή ή όχι. Η παρούσα εργασία ασχολείται με την αναγνώριση και κατηγοριοποίηση τάσεων και για αυτό κρίνεται σκόπιμη η διεξοδικότερη ανάλυση της τελευταίας.

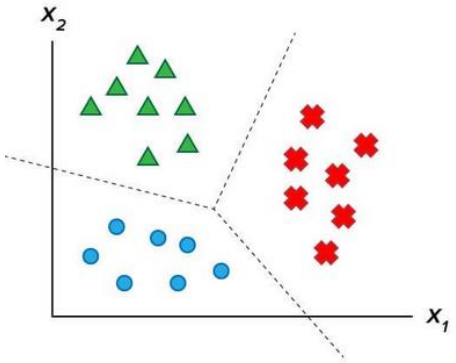
Τα προβλήματα κατηγοριοποίησης ταξινομούνται σε 4 κατηγορίες:

Δύο κλάσεων

Στα προβλήματα δύο κλάσεων υπάρχουν δύο κατηγορίες, στις οποίες τα δεδομένα ταξινομούνται. Κάθε αντικείμενο ανατίθεται σε μία από τις δύο κλάσεις, χωρίς να μπορεί να υπαχθεί ταυτόχρονα και στις δύο. Αυτό πραγματοποιείται συνήθως μέσω ενός δυαδικού δείκτη, που παίρνει τιμές 0 και 1, σε κάθε μία από τις δύο υπάρχουσες θέσεις που αντιστοιχίζονται με τις δύο κλάσεις, υποδεικνύοντας σε ποια από τις δύο ανήκει το αντικείμενο. Για παράδειγμα αν το τελικό αποτέλεσμα είναι το 01, συνεπάγεται ότι η πρόβλεψη του αλγορίθμου είναι ότι το αντικείμενο ανήκει στην δεύτερη κλάση. Αντίστοιχα το 10, επισημαίνει την πρώτη κλάση.

Πολλαπλών κλάσεων

Όπως φανερώνει η ονομασία τους τα προβλήματα πολλαπλών κλάσεων διαφέρουν από τα αντίστοιχα της δυαδικής κατηγοριοποίησης ως προς τον αριθμό των υπαρχόντων κλάσεων. Επεξηγηματικά, υπάρχουν περισσότερες των δύο κλάσεων και κάθε αντικείμενο ταξινομείται σε μία από όλες. Αυτό συνεπάγεται ότι τα ψηφία του δυαδικού δείκτη στην τελική πρόβλεψη για κάθε αντικείμενο, είναι ισάριθμα με τις κλάσεις και η θέση της μονάδες καθορίζει την κλάση η οποία επιλέγεται. Για παράδειγμα αν το τελικό αποτέλεσμα είναι το 0010, από τις συνολικά τέσσερις κλάσεις, η τρίτη κλάση είναι αυτή που έχει επιλεχθεί ως αντιπροσωπευτική του αντικειμένου.



Εικόνα 21: Πολλαπλές κλάσεις [12].

Πολλαπλής ανάθεσης

Στα προβλήματα πολλαπλής ανάθεσης υπάρχει η δυνατότητα το κάθε αντικείμενο να ανήκει σε περισσότερες από μία κλάσεις. Κατά συνέπεια η μονάδα ενδέχεται να εμφανίζεται περισσότερες από μία φορές στον δυαδικό δείκτη, πράγμα που δεν συνέβαινε έως τώρα. Για παράδειγμα αν το τελικό αποτέλεσμα είναι το 1010, το αντικείμενο έχει κατηγοριοποιηθεί στην πρώτη και στην τρίτη κλάση.

Ανισορροπίας

Τα προβλήματα ανισορροπίας είναι αυτά στα οποία εμφανίζεται απόκλιση στον αριθμό των δεδομένων, δηλαδή των γνωστών παραδειγμάτων, που υπάρχουν για κάθε κλάση και στα οποία βασίζεται το μοντέλο προκειμένου να κάνει προβλέψεις. Η απόκλιση υπάρχει περίπτωση να αφορά από μονοψήφιο αριθμό μέχρι πολλές τάξει μεγέθους. Για παράδειγμα μπορεί να υπάρχει ένα δεδομένο για την κλάση που μειοψηφεί, ενώ οι υπόλοιπες να διαθέτουν εκατοντάδες ή χιλιάδες δεδομένα. Αυτό, αν δεν προσεγγιστεί αναλόγως, οδηγεί σε αυξημένη πιθανότητα σφάλματος στις προβλέψεις που αφορούν την μειοψηφική κατηγορία.

Επιπλέον, αναφορικά με την επιτηρούμενη εκπαίδευση, σημειώνεται ότι αποτελεί πάγια τακτική η βάση δεδομένων να διαχωρίζεται σε δύο ή τρία μέρη. Το πρώτο και κυριότερο αποτελεί το 70%-80% των δεδομένων και χρησιμοποιείται για την εκπαίδευση του δικτύου. Το δεύτερο, αν υπάρχει, χρησιμοποιείται για την επικύρωση του δικτύου, δηλαδή τον έλεγχο της απόδοσής του και την πραγματοποίηση διορθώσεων για την βελτίωσή της. Αποτελεί το 10%-15% των δεδομένων. Το τρίτο χρησιμοποιείται για τον τελικό έλεγχο της απόδοσης του δικτύου και αποτελεί το υπολειπόμενο 10%-15%. Ο έλεγχος της απόδοσης συνοψίζεται στο ότι το ΤΝΔ τροφοδοτείται με δεδομένα που δεν έχει συναντήσει κατά την εκπαίδευση αλλά των οποίων τα ορθά αποτελέσματα είναι γνωστά. Έτσι, συγκρίνοντας τα αποτελέσματα που δίνει το ΤΝΔ σε αυτά τα δεδομένα, με τα ορθά, εξάγονται συμπεράσματα για την απόδοσή του.

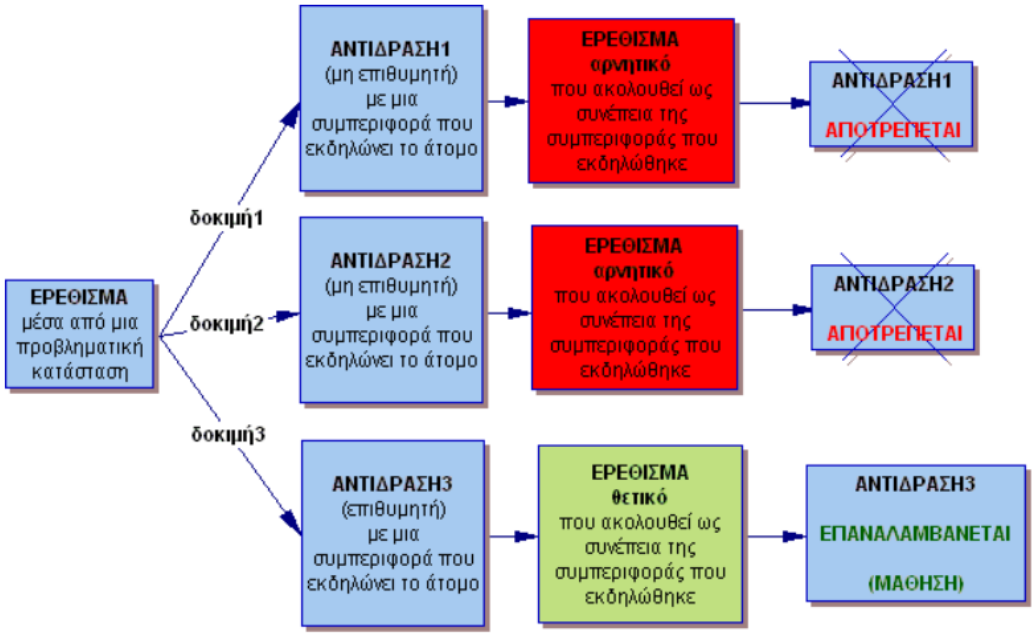
Μη επιτηρούμενη μάθηση

Η μη επιτηρούμενη μάθηση εφαρμόζεται στις πιο εξειδικευμένες μορφές δικτύων όπως τα αυτοδιοργανούμενα (self-organizing) ΤΝΔ. Τα διανύσματα σε αυτήν την περίπτωση δεν εμφανίζονται ανά ζεύγη καθώς εκλείπουν τα επιθυμητά διανύσματα εξόδου. Το

δίκτυο έχει την ικανότητα να κατηγοριοποιεί τα δεδομένα σε ομάδες, χωρίς την ύπαρξη προτύπων που να υποδεικνύουν το πώς είναι το αντιπροσωπευτικό μέλος κάθε ομάδας. Αυτό το επιτυγχάνει μεταβάλλοντας τους συντελεστές βαρύτητας έτσι ώστε τα πιο όμοια διανύσματα εισόδου να αναθέτονται στον ίδιο νευρώνα εξόδου. Παράλληλα, το ίδιο το νευρωνικό δίκτυο είναι που δημιουργεί ένα αντιπροσωπευτικό διάνυσμα για κάθε ομάδα.

Ενισχυτική μάθηση

Στην ενισχυτική, όπως και στην μη επιτηρούμενη μάθηση, τα διανύσματα εισόδου δεν αντιστοιχίζονται με επιθυμητές αποκρίσεις. Εδώ η απόκριση αξιολογείται με έναν ορισμένο δείκτης συμπεριφοράς, που ονομάζεται ενισχυτικό σήμα (reinforcement signal) και με βάση αυτόν αναπροσαρμόζονται οι συντελεστές βαρύτητας. Συγκεκριμένα, ανάλογα με τον δείκτη συμπεριφοράς εκπέμπεται και ανατροφοδοτείται στο δίκτυο ένα σήμα το οποίο είναι ενισχυτικό στην περίπτωση που η απόκριση αξιολογείται ως εύστοχη, ενώ είναι επιτιμητικό στην περίπτωση που αξιολογείται ως άστοχη. Πρόκειται για έναν μηχανισμό εξερεύνησης έτσι ώστε να προκύπτουν όλο και πιο εύστοχες αντιδράσεις. Έτσι το δίκτυο οδηγείται σταδιακά στην ορθή συμπεριφορά βασιζόμενο στον νόμο του αποτελέσματος του Edward Thorndike, που υποδεικνύει ότι μία αντίδραση που εμφανίζεται λίγο πριν από ένα θετικό ερέθισμα (αποτέλεσμα) έχει την τάση να εμφανίζεται και πάλι σε μελλοντικές παρόμοιες καταστάσεις.



Εικόνα 22: Η θεωρία μάθησης του Thorndike, που εξαρτά την μάθηση από το ερέθισμα που ακολουθεί, ως συνέπεια της συμπεριφοράς που εκδηλώνεται [13].

Η ενισχυτική μάθηση διακρίνεται σε συσχετιστική και μη συσχετιστική. Στην πρώτη η απόδοση αξιολογείται ταυτόχρονα ως εύστοχη σε σχέση με κάποιες αιτίες και ως άστοχη σε σχέση με κάποιες άλλες αντί να χαρακτηρίζεται μονοσήμαντα από ένα και μόνο μέτρο, τον δείκτη συμπεριφοράς, πράγμα που συμβαίνει στην δεύτερη περίπτωση, η

οποία αναλύθηκε παραπάνω. Σε κάθε περίπτωση όταν το δίκτυο σταματάει να αλλάζει τις τιμές των βαρών, η εκπαίδευση θεωρείται ολοκληρωμένη.

Ο τρόπος με τον οποίο υπολογίζεται και βελτιώνεται η απόδοση, μέσω της αναπροσαρμογής των βαρών και των σταθερών πόλωσης, ονομάζεται κανόνας μάθησης. Η εφαρμογή του εκάστοτε κανόνα μάθησης πραγματοποιείται επαναληπτικά και ξεκινώντας από τις υπάρχουσες συνθήκες της αρχιτεκτονικής και των αρχικών τιμών των βαρών και των σταθερών πόλωσης καθοδηγεί το ΤΝΔ ώστε να βελτιώσει την απόδοσή του. Παρακάτω εξετάζονται οι βασικότεροι κανόνες μάθησης των ΤΝΔ:

Κανόνας του Hebb (Hebbian Learning Rule)

Είναι ο πρώτος κανόνας μάθησης που εμφανίστηκε και αναπτύχθηκε το 1949 από τον Donald Hebb. Υποδεικνύει ότι αν δύο γειτονικοί νευρώνες, ενεργοποιούνται και απενεργοποιούνται συγχρόνως και επανειλημμένα, τότε η ισχύς της συνάψεως που τους συνδέει, δηλαδή η τιμή του συντελεστή βαρύτητάς της πρέπει να αυξηθεί. Σε νευρώνες που λειτουργούν αντίστροφα, η τιμή του συντελεστή βαρύτητάς πρέπει να μειωθεί, ενώ αν δεν υπάρχει συσχέτιση μεταξύ τους πρέπει να παραμείνει η ίδια.

Αυτό έχει ως αποτέλεσμα ισχυρά θετικούς συντελεστές στους κόμβους που λειτουργούν συγχρονισμένα και ισχυρά αρνητικούς στην αντίθετη περίπτωση. Κατά την εκκίνηση όλα τα βάρη λαμβάνουν μηδενική τιμή. Οι απόλυτες τιμές στις οποίες θα καταλήξουν οι τιμές των συντελεστών εξαρτώνται από το χρόνο εκπαίδευσης, πράγμα που συνιστά μείζον μειονέκτημα.

Εφόσον οι επιθυμητές αποκρίσεις των νευρώνων δεν εμπεριέχονται στην διαδικασία, πρόκειται για κανόνα μάθησης που εφαρμόζεται σε μη επιτηρούμενη εκπαίδευση.

Κανόνας του Perceptron (Perceptron Learning Rule)

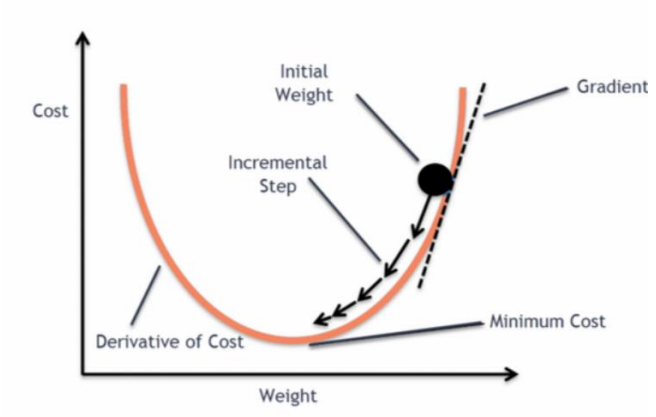
Πρόκειται για κανόνα διόρθωσης σφαλμάτων, που χρησιμοποιείται στην επιτηρούμενη εκπαίδευση. Σε αυτόν τον κανόνα η ανανέωση των συναπτικών βαρών και των σταθερών πόλωσης πραγματοποιείται, συγκρίνοντας τις αποκρίσεις του δικτύου με τις επιθυμητές. Το συνολικό σφάλμα υπολογίζεται σαν άθροισμα των τετραγώνων των σφαλμάτων που προκύπτουν από τους νευρώνες εξόδου. Συνεπώς πρόκειται για κανόνα επίλυσης ενός προβλήματος ελαχίστων τετραγώνων, που όμως χρησιμοποιεί βηματικές συναρτήσεις ενεργοποίησης που όπως εξηγείται παρακάτω δεν υποστηρίζουν την οπισθόδρομη διάδοση λάθους.

Κανόνας Συσχέτισης (Correlation Learning Rule)

Ο κανόνας συσχέτισης βασίζεται σε μία παρόμοια αρχή με τον κανόνα του Hebb, με την διαφορά ότι υλοποιείται σε επιτηρούμενη εκπαίδευση και συνεπώς χρησιμοποιεί την επιθυμητή έξοδο σαν οδηγό. Επισημαίνει ότι οι συντελεστές βαρύτητας μεταξύ νευρώνων που συγχρονίζονται, ανταποκρίνομενοι ο ένας στο άλλον πρέπει να γίνουν ισχυρά θετικοί ενώ αντίθετα, οι συντελεστές μεταξύ νευρώνων με αντίθετες αντιδράσεις πρέπει να γίνουν ισχυρά αρνητικοί. Όπως και στον κανόνα του Hebb κατά την εκκίνηση όλα τα βάρη λαμβάνουν μηδενική τιμή.

Κανόνας του Δέλτα

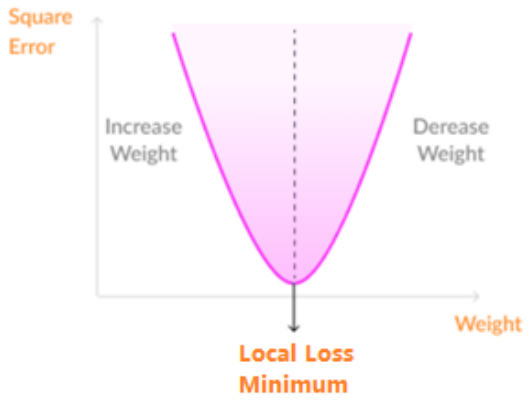
Ο κανόνας του Δέλτα είναι ένας ευρύτατα χρησιμοποιούμενος κανόνας ανανέωσης βαρών και σταθερών πόλωσης στην επιτηρούμενη εκπαίδευση, που στηρίζεται στην αρχή της κατεύθυνση μείωσης του σφάλματος (gradient descent). Αυτή αποτελεί έναν αλγόριθμο βελτιστοποίησης που προσεγγίζει ένα τοπικό ελάχιστο της συνάρτησης κόστους ή συνάρτησης σφάλματος, κάνοντας μικρά βήματα αντίθετα με την κλίση της συνάρτησης στο εκάστοτε σημείο. Στο κάθε βήμα ανανεώνονται οι συντελεστές βαρύτητας και οι σταθερές πόλωσης μέσω της οπισθόδρομης διάδοσης σφάλματος μειώνοντας την συνάρτηση σφάλματος. Η επαναληπτική διαδικασία ολοκληρώνεται όταν η κλίση της συνάρτησης μηδενιστεί.



Εικόνα 23: Σχηματική απεικόνιση της συνάρτησης σφάλματος σε σχέση με τα συναπτικά βάρη στον αλγόριθμο gradient descent [14].

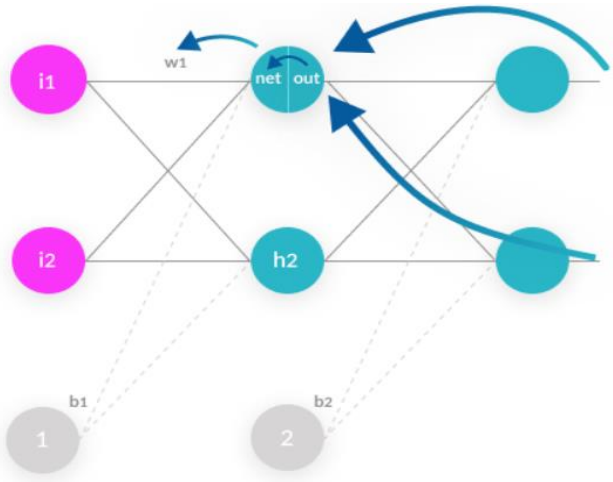
2.2.4.3 Οπισθόδρομη διάδοση σφάλματος

Ο αλγόριθμος οπισθόδρομης διάδοσης σφάλματος αποτελεί τον πλέον διαδεδομένο τρόπο εκπαίδευσης ΤΝΔ πρόσθιας τροφοδότησης με επιτηρούμενη μάθηση. Βασίζεται στην αρχή της κατεύθυνσης μείωσης του σφάλματος (gradient descent), που σημαίνει ότι ανανεώνει τα βάρη και τις σταθερές πόλωσης σε κάθε επανάληψη, προς την κατεύθυνση μείωσης του σφάλματος. Αυτό πραγματοποιείται ανεξαρτήτως, τόσο της συνάρτησης κόστους που επιλέγεται, η οποία είναι υπεύθυνη για τον τρόπο υπολογισμού του σφάλματος όσο του κανόνα ανανέωσης των βαρών και των σταθερών πόλωσης.



Εικόνα 24: Σχηματική αναπαράσταση του τρόπου με τον οποίο μειώνεται το σφάλμα μέσω της ανανέωσης των συντελεστών βαρύτητας στην οπισθόδρομη διάδοση σφάλματος [15].

Πιο συγκεκριμένα, υπολογίζεται το σφάλμα και η συνεισφορά του κάθε συντελεστή βαρύτητας σε αυτό διατρέχοντας τα επίπεδα νευρώνων αντίστροφα, ξεκινώντας από το επίπεδο εξόδου και καταλήγοντας στο επίπεδο εισόδου.



Εικόνα 25: Η φορά της οπισθόδρομης διάδοσης σφάλματος [15].

Στην συνέχεια αναπροσαρμόζονται οι συντελεστές βαρύτητας και σταθερές πόλωσης έτσι ώστε να συνεισφέρουν λιγότερο στο συνολικό σφάλμα και εισάγεται το επόμενο διάνυσμα εισόδου για να υπολογιστεί εκ νέου το σφάλμα και να επαναληφθεί η διαδικασία.

Η οπισθόδρομη διάδοση σφάλματος με εφαρμογή του κανόνα του Δέλτα στη περίπτωση ΤΝΔ απλού επιπέδου (βλ. Εικόνα 15) παρουσιάζεται αναλυτικά παρακάτω.

Ορίζονται ως:

x_i , το σήμα που εκπέμπει ο νευρώνας i του επιπέδου εισόδου

w_{ij} , ο συντελεστής βαρύτητας που συνδέει τον νευρώνα i του επιπέδου εισόδου με τον νευρώνα j του επιπέδου εξόδου

w_{0j} ,η σταθερά πόλωσης του νευρώνα j του επιπέδου εξόδου

y_j , το σήμα που εκπέμπει ο νευρώνας j του επιπέδου εξόδου

$g(x)$, η συνάρτηση ενεργοποίησης του νευρώνα j του επιπέδου εξόδου

t_j , η επιθυμητή έξοδος του νευρώνα j του επιπέδου εξόδου

n/m , ο αριθμός νευρώνων του επιπέδου εισόδου/εξόδου

z_j , το σήμα που εκπέμπει ο νευρώνας j του κρυμμένου επιπέδου

v_{ij} , ο συντελεστής βαρύτητας που συνδέει τον νευρώνα i του επιπέδου εισόδου με τον νευρώνα j του κρυμμένου επιπέδου

v_{0j} , η σταθερά πόλωσης του νευρώνα j του κρυμμένου επιπέδου

$f(x)$, η συνάρτηση ενεργοποίησης του νευρώνα j του κρυμμένου επιπέδου

Έστω y_{in_j} , το σήμα που λαμβάνει ο j νευρώνας του επιπέδου εξόδου. Ισχύει:

$$y_{in_j} = w_{0j} + \sum_{i=1}^{n} x_i w_{ij}$$

Η έξοδός του είναι:

$$y_j = g(y_{in_j})$$

Το σφάλμα του j νευρώνα του επιπέδου εξόδου μπορεί να οριστεί ως:

$$E_j = \frac{1}{2}(t_j - y_j)^2$$

Επιλέγεται αυτός ο ορισμός έτσι ώστε το σφάλμα να είναι πάντα θετικό ή μηδέν και κατά αντιστοιχία με την Εικόνα 23 η μείωση της απόλυτης τιμής της κλίσης του σφάλματος να οδηγεί σε τοπικό ελάχιστο.

Τότε το σφάλμα Ε των νευρώνων του επιπέδου εξόδου αποτελεί άθροισμα τετραγώνων:

$$E = \sum_{j=1}^{m} E_j = \sum_{j=1}^{m} \frac{1}{2}(t_j - y_j)^2$$

Η συνεισφορά των συντελεστών βαρύτητας στην κλίση του σφάλματος είναι:

$$\frac{\partial E}{\partial w_{ij}}, \qquad i = 1{:}n, \;\; j = 1{:}m$$

Εστιάζοντας σε έναν συγκεκριμένο συντελεστή w_{ij} η συνεισφορά του είναι:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial\left(\frac{1}{2}(t_j - y_j)^2\right)}{\partial w_{ij}} = \frac{\partial\left(\frac{1}{2}(t_j - y_j)^2\right)}{\partial y_j} * \frac{\partial y_j}{\partial w_{ij}} = -\left(t_j - y_j\right) * \frac{\partial y_j}{\partial w_{ij}}$$
$$= -\left(t_j - y_j\right) * \frac{\partial y_j}{\partial y_{in_j}} * \frac{\partial y_{in_j}}{\partial w_{ij}} = -\left(t_j - y_j\right) * g'\left(y_{in_j}\right) * \frac{\partial \sum_{i=1}^{n} x_i w_{ij}}{\partial w_{ij}}$$

Επειδή αναφερόμαστε στον συγκεκριμένο συντελεστή:

$$\frac{\partial \sum_{i=1}^{n} x_i w_{ij}}{\partial w_{ij}} = x_i$$

Άρα η εξίσωση της κλίσης του σφάλματος είναι:

$$\frac{\partial E}{\partial w_{ij}} = -\left(t_j - y_j\right) * g'\left(y_{in_j}\right) * x_i$$

Ο αλγόριθμος gradient descent υποδεικνύει η μεταβολή κάθε συντελεστή βαρύτητας να είναι ανάλογη της κλίσης, αλλά προς την αντίθετη κατεύθυνση. Συνεπώς εξαλείφοντας το μείον από την εξίσωση, ούτως ώστε ο συντελεστής w_{ij} να κινηθεί προς την κατεύθυνση μείωσης του σφάλματος και διαλέγοντας μία αναλογική θετική τιμή α που ονομάζεται ρυθμός μάθησης, προκύπτει η εξίσωση μεταβολής του συντελεστή:

$$\Delta w_{ij} = a * \left(t_j - y_j\right) * g'\left(y_{in_j}\right) * x_i$$

Ομοίως προκύπτει και η εξίσωση μεταβολής των σταθερών πόλωσης:

$$\Delta w_{0j} = a * \left(t_j - y_j\right) * g'\left(y_{in_j}\right)$$

Η τιμή του ρυθμού μάθησης α μπορεί να είναι είτε σταθερή είτε να αλλάζει σε κάθε επανάληψη. Γενικά προτιμάται η χρησιμοποίηση της βέλτιστης τιμής της που λαμβάνεται από την ελαχιστοποίηση της γραμμής (line minimization) σε κάθε βήμα.

Από τις παραπάνω σχέσεις φαίνεται η εξάρτηση από το σφάλμα, από την είσοδο και από την παράγωγο της συνάρτησης ενεργοποίησης του νευρώνα. Για αυτόν τον λόγο συναρτήσεις ενεργοποίησης με σταθερή ή μηδενική παράγωγο δεν χρησιμοποιούνται στην οπισθόδρομη διάδοση λάθους.

Γενίκευση της ανωτέρω περίπτωσης αποτελεί, η οπισθόδρομη διάδοση σφάλματος σε ένα ΤΝΔ με ένα κρυμμένο επίπεδο (βλ. Εικόνα 16), με εφαρμογή του κανόνα του Δέλτα.

Κάθε νευρώνας εισόδου λαμβάνει μία συνιστώσα του διανύσματος εκπαίδευσης εισόδου x_i και εκπέμπει την συνιστώσα αυτή, προς όλους τους νευρώνες του επόμενου επιπέδου καθώς συνάρτηση ενεργοποίησής των νευρώνων του επιπέδου είναι η ταυτοτική συνάρτηση και οι συντελεστές πόλωσης τους είναι μηδενικοί.

Έστω z_{in_j} το σήμα που λαμβάνει ο j νευρώνας του κρυμμένου επιπέδου. Ισχύει:

$$z_{in_j} = v_{0j} + \sum_{i=1}^{n} x_i v_{ij}$$

Η δραστηριότητά του είναι:

$$z_j = f\left(z_{in_j}\right)$$

Αυτή είναι η τιμή που εκπέμπει προς όλους τους νευρώνες του επόμενου επιπέδου, δηλαδή του επιπέδου εξόδου. Κάθε νευρώνας του επιπέδου εξόδου αθροίζει στην είσοδό του τα σταθμισμένα σήματα που λαμβάνει από όλους τους νευρώνες του κρυμμένου επιπέδου:

$$y_{in_k} = w_{0k} + \sum_{j=1}^{p} z_j w_{jk}$$

και εφαρμόζει την συνάρτηση δραστηριοποίησής του για να υπολογίσει την δραστηριότητά του:

$$y_k = g(y_{in_k})$$

Κάθε νευρώνας εξόδου λαμβάνει μία συνιστώσα του διανύσματος στόχου εξόδου, το οποίο αντιστοιχεί στο διάνυσμα εκπαίδευσης εισόδου και με βάση τον κανόνα του Δέλτα που αναλύθηκε παραπάνω υπολογίζεται ο παράγοντας σφάλματός του:

$$\delta_k = (t_k - y_k) * g'(y_{in_k})$$

υπολογίζεται ο παράγοντας διόρθωσης του συντελεστή βαρύτητας w_{jk}:

$$\Delta w_{jk} = a * \delta_k * z_j$$

υπολογίζεται ο παράγοντας διόρθωσης της σταθεράς πόλωσης w_{0k}:

$$\Delta w_{0k} = a * \delta_k$$

και εκπέμπεται αυτός ο παράγοντας σφάλματος στους νευρώνες του προηγούμενου επιπέδου δηλαδή του κρυμμένου επιπέδου. Εκεί σε κάθε νευρώνα αθροίζονται οι σταθμισμένοι παράγοντες σφάλματος:

$$\delta_{in_j} = \sum_{k=1}^{m} \delta_k * w_{jk}$$

υπολογίζεται ο παράγοντας διόρθωσης του σφάλματός του:

$$\delta_j = \delta_{in_j} * f'(z_{in_j})$$

υπολογίζεται ο παράγοντας διόρθωσης του συντελεστή βαρύτητας v_{ij}:

$$\Delta v_{ij} = a * \delta_j * x_i$$

και υπολογίζεται ο παράγοντας διόρθωσης της σταθεράς πόλωσης v_{0j}:

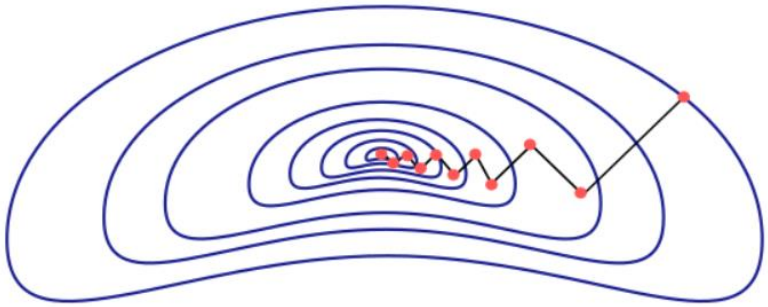
$$\Delta v_{0j} = a * \delta_j$$

Είναι προφανές ότι αν υπάρχουν περισσότερα κρυμμένα επίπεδα οι παράγοντες σφάλματος συνεχίζουν την οπισθόδρομη διάδοσή τους με τον ίδιο ακριβώς τρόπο μέχρι και το πρώτο κρυμμένο επίπεδο.

Όταν αυτή η διαδικασία ολοκληρωθεί ακολουθεί η διόρθωση των συντελεστών βαρύτητας και των σταθερών πόλωσης:

$$w_{jk}(new) = w_{jk}(old) + \Delta w_{jk} \quad , \quad w_{0k}(new) = w_{0k}(old) + \Delta w_{0k}$$
$$v_{ij}(new) = v_{ij}(old) + \Delta v_{ij} \quad , \quad v_{0j}(new) = v_{0j}(old) + \Delta v_{0j}$$

Στην συνέχεια εισάγεται το επόμενο διάνυσμα εισόδου. Η διαδικασία επαναλαμβάνεται για κάθε διάνυσμα εισόδου και επιθυμητής εξόδου. Σε κάθε επανάληψη ο κανόνας του Δέλτα, που στηρίζεται αρχή της κατεύθυνση μείωσης του σφάλματος (gradient descent), ωθεί την αναπροσαρμογή των βαρών και των συντελεστών πόλωσης στην κατεύθυνση όλο και μικρότερου σφάλματος. Όπως έχει αναφερθεί ένα πέρασμα από όλα τα ζευγάρια διανυσμάτων εισόδου και επιθυμητής εξόδου ονομάζεται εποχή και συνήθως χρειάζεται ένας μεγάλος αριθμός εποχών για την εκπαίδευση ενός ΤΝΔ οπισθόδρομης διάδοσης.



Εικόνα 26: Εύρεση της βέλτιστης λύσης με την αρχή της κατεύθυνση μείωσης του σφάλματος (gradient descent) και εφαρμογή του κανόνα του Δέλτα [16].

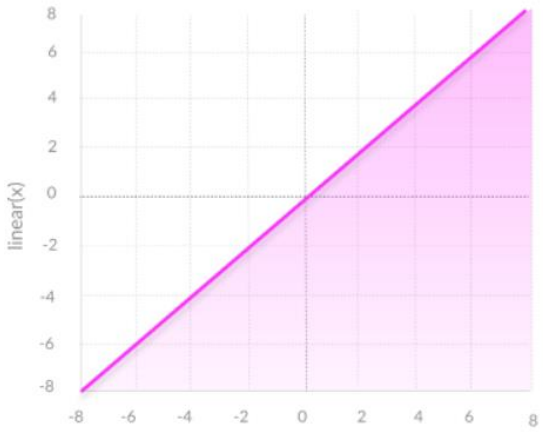
Στον αλγόριθμο που περιγράφηκε, απαιτούνται πολλές επαναλήψεις και συνεπώς αργή σύγκλιση στην επιθυμητή λύση. Επίσης η διόρθωση των συντελεστών βαρύτητας γίνεται μετά από κάθε παρουσίαση ενός μόνο ζευγαριού διανύσματος εκπαίδευσης εισόδου και διανύσματος στόχου εξόδου (online training). Μία τυπική παραλλαγή αυτού, είναι η μαζική διόρθωση, στην οποία οι παράγοντες διόρθωσης συγκεντρώνονται για μία ολόκληρη εποχή, πριν χρησιμοποιηθούν (batch training).

2.2.4.4 Συναρτήσεις ενεργοποίησης

Όπως έχει αναφερθεί, η συνάρτηση ενεργοποίησης κάθε νευρώνα δέχεται ως όρισμα το αποτέλεσμα του αθροιστή στην είσοδο του νευρώνα και καθορίζει την δραστηριότητά του. Αυτό συμπεριλαμβάνει τόσο το αν πρόκειται να υπάρξει εκπεμπόμενο σήμα από τον νευρώνα, όσο και το ποιο θα είναι το μέγεθός του.

Οι συναρτήσεις ενεργοποίησης που χρησιμοποιεί ένα τεχνητό νευρωνικό δίκτυο επηρεάζουν τόσο το αποτέλεσμα και την ακρίβεια όσο και την ικανότητά μάθησής του. Επιπλέον έχουν καταλυτική συνεισφορά στην δυνατότητα και στην ταχύτητα σύγκλισης. Σημειώνεται ότι συνήθως όλοι οι νευρώνες του ίδιου επιπέδου σε ένα ΤΝΔ πολλαπλών επιπέδων χρησιμοποιούν κοινή συνάρτηση ενεργοποίησης. Οι ευρύτερα χρησιμοποιούμενες χωρίζονται σε τρείς κατηγορίες, την ταυτοτική, την βηματική και τις μη γραμμικές συναρτήσεις και παρουσιάζονται παρακάτω.

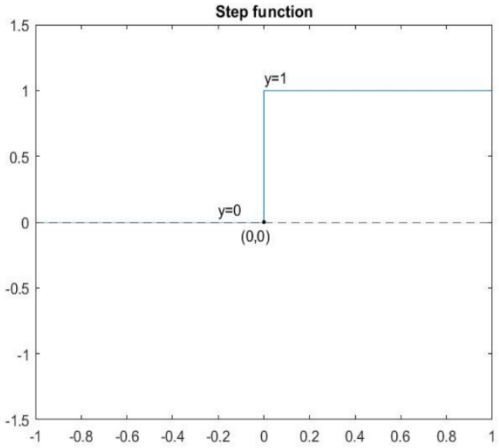
Ταυτοτική συνάρτηση (linear transfer function)



Εικόνα 27: Ταυτοτική συνάρτηση [30].

- Χρησιμοποιείται για τους νευρώνες του επιπέδου εισόδου.
- Δεν χρησιμοποιείται σε περισσότερα επίπεδα καθώς έχει την ιδιότητα να τα «συμπυκνώνει» σε ένα. Αυτό σημαίνει ότι ανεξαρτήτως αριθμού επιπέδων, το αποτέλεσμα του τελευταίου είναι μία γραμμική συνάρτηση του πρώτου (γραμμικός συνδυασμός γραμμικών συναρτήσεων είναι γραμμική συνάρτηση).
- Δεν χρησιμοποιείται στην οπισθόδρομη διάδοση λάθους καθώς έχει σταθερή παράγωγο.

Βηματική συνάρτηση (step function)



Εικόνα 28: Βηματική συνάρτηση.

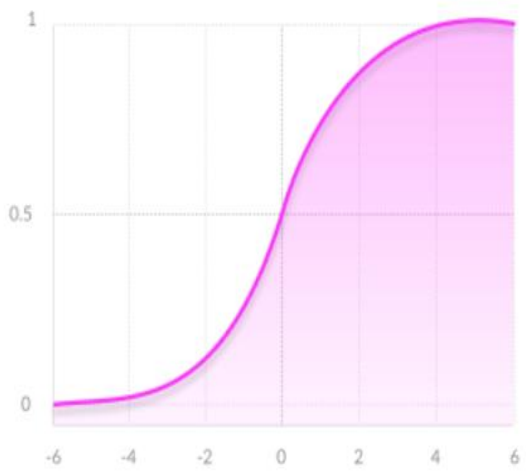
- $f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$
- Μετατρέπει το συνεχές σήμα εισόδου σε δυαδικό (0 ή 1) ή σε διπολικό (-1 ή 1) σήμα εξόδου.
- Διαθέτει το κατώφλι θ, που καθορίζει αν ο νευρώνας πρόκειται να ενεργοποιηθεί, στην περίπτωση που η είσοδος το ξεπερνάει ή αν πρόκειται να παραμείνει αδρανής, στην περίπτωση που η είσοδος δεν το ξεπερνάει.
- Δεν επιτρέπει εξόδους διαφορετικών τιμών.
- Δεν χρησιμοποιείται στην οπισθόδρομη διάδοση λάθους καθώς έχει μηδενική μη συνεχή παράγωγο.

Οι μη γραμμικές συναρτήσεις ενεργοποίησης επιτρέπουν πολύπλοκες διασυνδέσεις μεταξύ εισόδων και εξόδων που είναι ουσιαστικές για την εκμάθηση σύνθετων δεδομένων. Η απλή σχέση ανάμεσα στην τιμή των συναρτήσεων ενεργοποίησης και των παραγώγων τους σε ένα σημείο, απλοποιεί κατά πολύ τη διαδικασία εκπαίδευσης και μικραίνει τον υπολογιστικό φόρτο. Ένα ΤΝΔ μπορεί να προσεγγίσει σχεδόν οποιαδήποτε διαδικασία, με την χρήση μη γραμμικών συναρτήσεων ενεργοποίησης και αυτό γιατί επιτρέπουν:

1. Την οπισθόδρομη διάδοση λάθους, αφού η παράγωγός τους είναι συνάρτηση.
2. Την χρήση πολλαπλών επιπέδων νευρώνων που είναι απαραίτητα για την εκμάθηση πολύπλοκων δεδομένων, με υψηλή ακρίβεια.

Οι συχνότερα εμφανιζόμενες μη γραμμικές συναρτήσεις ενεργοποίησης είναι οι σιγμοειδείς καμπύλες:

Δυαδική σιγμοειδής συνάρτηση (sigmoid function)



Εικόνα 29: Δυαδική σιγμοειδής συνάρτηση [30].

- $f(x) = \frac{1}{1+e^{-\sigma * x}}$, σ=συντελεστής κλίσης.

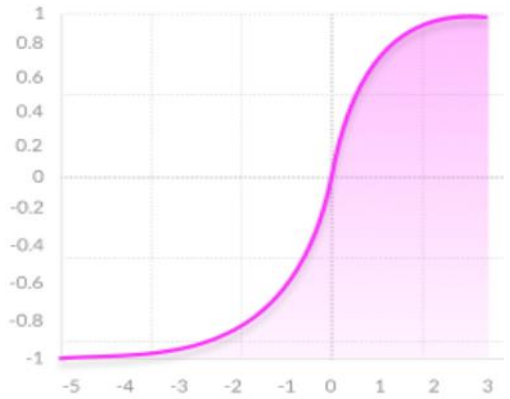
Πλεονεκτήματα:

- Ομαλή παράγωγος, που αποτρέπει άλματα στις τιμές εξόδου.
- Οι τιμές κυμαίνονται μεταξύ 0 και 1, κανονικοποιώντας την έξοδο κάθε νευρώνα.

Μειονεκτήματα:

- Εξαλειπτόμενη παράγωγος, όπου για πολύ μεγάλες και μικρές τιμές του X, δεν υπάρχει μεταβολή στην πρόβλεψη. Αυτό συμβάλει στην αργή σύγκλιση.
- Υψηλό υπολογιστικό φορτίο.

Διπολική σιγμοειδής-Συνάρτηση υπερβολικής εφαπτομένης (hyperbolic tangent)



Εικόνα 30: Διπολική σιγμοειδής συνάρτηση [30].

- $f(x) = \frac{1-e^{-\sigma * x}}{1+e^{-\sigma * x}}$, σ=συντελεστής κλίσης.

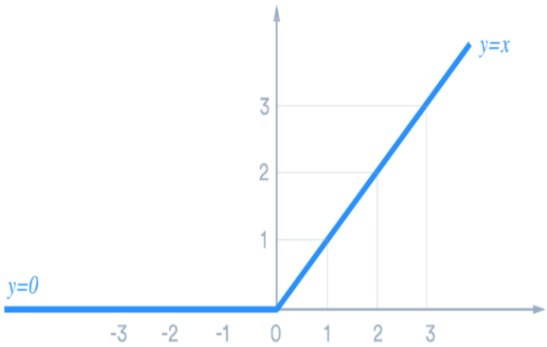
Πλεονεκτήματα:

- Συμμετρική ως προς το 0 και με πεδίο τιμών το $[-1,1]$ επιτρέπει την μοντελοποίηση αυστηρά αρνητικών ή θετικών εισόδων.
- Όμοια με την δυαδική σιγμοειδή συνάρτηση.

Μειονεκτήματα:

- Όμοια με την δυαδική σιγμοειδή συνάρτηση.

Συνάρτηση ReLU (Rectified Linear Unit)



Εικόνα 31: Συνάρτηση ReLU [31].

- $f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$

Πλεονεκτήματα:

- Επιτρέπει την ταχεία σύγκλιση.
- Η παράγωγός της δεν είναι σταθερή συνάρτηση, επιτρέποντας την οπισθόδρομη διάδοση λάθους.

Μειονεκτήματα:

- Όταν οι είσοδοι παίρνουν αρνητικές τιμές ή 0, η παράγωγός της μηδενίζεται μη επιτρέποντας την οπισθόδρομη διάδοση λάθους.

2.2.4.5 Ειδικά

Σε αυτήν την ενότητα εξετάζονται κάποια ειδικά θέματα, που αποτελώντας προεκτάσεις των όσων έχουν αναφερθεί, βρίσκουν εφαρμογή στην παρούσα εργασία. Πρόκειται για πιο εξειδικευμένες περιπτώσεις των όσων θεμελιακών πτυχών έχουν παρουσιαστεί και χρησιμοποιούνται για να προσδώσουν ταχύτητα και ακρίβεια στις αναλύσεις, αλλά και να τις προσαρμόσουν στην φύση της εργασίας. Αφορούν κύρια δομικά στοιχεία ενός ΤΝΔ δηλαδή τις συναρτήσεις ενεργοποίησης, τις συναρτήσεις κόστους και τον αλγόριθμο εκπαίδευσης.

2.2.4.5.1 Συναρτήσεις ενεργοποίησης

Αναφορικά με τις συναρτήσεις ενεργοποίησης στην εργασία χρησιμοποιούνται κατά κύριο λόγο οι παρακάτω:

Συνάρτηση Softmax

$$\boldsymbol{Softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{k} e^{x_j}} \quad , j = 1, \dots, k}$$

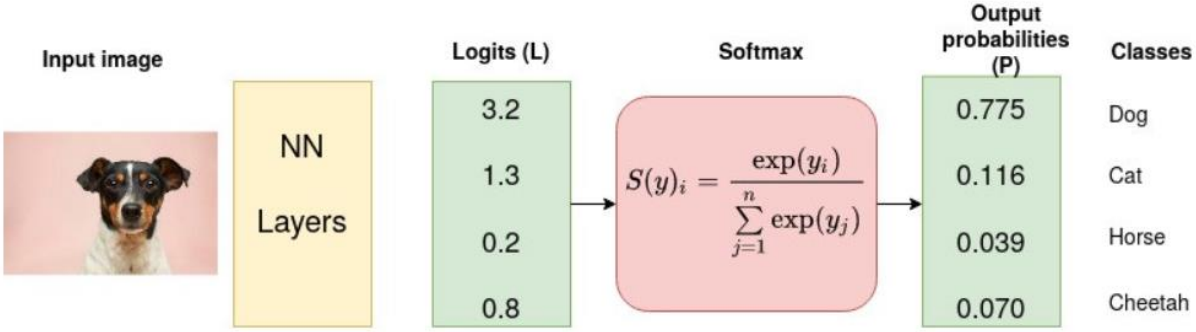
Η συνάρτηση Softmax ή κανονικοποιημένη εκθετική συνάρτηση χρησιμοποιείται συχνά ως συνάρτηση ενεργοποίησης κυρίως του επιπέδου εξόδου στα προβλήματα κατηγοριοποίησης (classification) με περισσότερες από δύο κατηγορίες. Ο λόγος είναι ότι μετατρέπει την έξοδο του ΤΝΔ σε πιθανότητα να ανήκει η εκάστοτε είσοδος στην κάθε κατηγορία. Αυτό σημαίνει ότι η έξοδος είναι ένα διάνυσμα με τιμές που ανήκουν στο [0,1] αλλά και που έχουν άθροισμα 1 καθώς αντιπροσωπεύουν πιθανότητες. Σημειώνεται ότι για να έχει νόημα μια τέτοια έξοδος οι κατηγορίες πρέπει να είναι αμοιβαία αποκλειόμενες.

Τα ορίσματα της συνάρτησης και κατά επέκταση του δικτύου μπορεί να είναι οποιοιδήποτε πραγματικοί αριθμοί. Η συνάρτηση είναι αύξουσα άρα για μεγαλύτερες

τιμές δίνει μεγαλύτερες πιθανότητες. Μία απλή εφαρμογή της φαίνεται στο ακόλουθο παράδειγμα:

$$\begin{array}{ccccc} \text{Είσοδος} & & \text{Softmax} & & \text{Έξοδος} \\ \begin{bmatrix} 8 \\ 5 \\ 0 \end{bmatrix} & \rightarrow & \frac{e^{x_i}}{\sum_{j=1}^{k} e^{x_j}} & \rightarrow & \begin{bmatrix} 0.952 \\ 0.047 \\ 0.001 \end{bmatrix} \end{array}$$

Όλες οι τιμές εξόδου βρίσκονται μεταξύ 0 και 1 και έχουν άθροισμα ίσο με 1. Η τιμή 8 έχει επικρατήσει στην συνάρτηση και έχει συμπιέσει τις τιμές 5 και 0 σε πολύ χαμηλές πιθανότητες. Από αυτό καταδεικνύεται ότι χρειάζεται προσοχή κατά την εφαρμογή της καθώς η συνάρτηση παράγει τιμές κοντά στο 0 και στο 1 για μεγάλο εύρος του πεδίου ορισμού της και ενδέχεται τα αποτελέσματα να υποδεικνύουν μία κατηγορία ως κυρίαρχη ενώ στην πραγματικότητα να μην επικρατεί με τόσο μεγάλη διαφορά.



Εικόνα 32: Παράδειγμα εφαρμογής της Softmax [17]

Συνάρτηση logsig

Πρόκειται για δυαδική σιγμοειδή συνάρτηση με συντελεστή κλίσης ίσο με 1. Έχει τα χαρακτηριστικά που περιγράφηκαν παραπάνω. Παρότι δίνει τιμές μεταξύ 0 και 1 αυτές δεν έχουν άθροισμα την μονάδα και συνεπώς δεν αποτελούν πιθανότητες. Σε αυτό το σημείο είναι και η κυριότερη διαφορά της logsig από την Softmax καθώς η πρώτη δέχεται σαν όρισμα βαθμωτό μέγεθος ενώ η δεύτερη διάνυσμα. Μπορεί να λεχθεί ότι αποτελεί υποπερίπτωσή της καθώς από την Softmax προκύπτει η logsig για το διάνυσμα [x,0] καθώς η πιθανότητα του x είναι:

$$Softmax([x, 0]) = \frac{e^x}{e^x + 1} = \frac{1}{1 + e^{-x}} = logsig(x)$$

Το επόμενο καίριο σημείο είναι οι συναρτήσεις κόστους που χρησιμοποιούνται στην παρούσα εργασία.

Η συνάρτηση κόστους αφορά το τρόπο υπολογισμού του σφάλματος, του οποίου η ελαχιστοποίηση αποτελεί κατά κανόνα, τον απώτερο σκοπό της εκπαίδευσης. Ο ρόλος της είναι σημαίνων αφού διυλίζει τις διάφορες πτυχές του μοντέλου καταλήγοντας σε έναν αριθμό τέτοιον ώστε, οποιαδήποτε μείωση, σε αυτόν τον αριθμό να αποτελεί

αναμφισβήτητη βελτίωση του μοντέλου. Συνεπώς είναι αυτή που κρίνει σε ποιο βαθμό το μοντέλο έχει προσαρμοστεί στην βάση δεδομένων.

Η καταλληλότητα της συνάρτησης κόστους εξαρτάται από το πρόβλημα. Σε γενικό επίπεδο η επιλογή της εξαρτάται από το αν διεκπεραιώνει παλινδρόμηση (regression) ή κατηγοριοποίηση (classification).

2.2.4.5.2 Συνάρτηση Κόστους

Στα προβλήματα παλινδρόμησης η ευρύτερα χρησιμοποιούμενη συνάρτηση κόστους είναι το μέσο τετραγωνικό σφάλμα.

Μέσο τετραγωνικό σφάλμα (Mean Squared Error)

$$MSE = \frac{\sum_{k=1}^{m}(y_k - t_k)^2}{m}$$

Το μέσο τετραγωνικό σφάλμα (mse) είναι η μέση τιμή του αθροίσματος των τετραγώνων των διαφορών μεταξύ των προβλέψεων και των πραγματικών τιμών. Κατά την εκπαίδευση των ΤΝΔ οι προβλέψεις αναφέρονται στις εξόδους του και οι πραγματικές τιμές στα διανύσματα στόχους. Όπως φαίνεται και από τον τύπο, λαμβάνονται υπ'όψην μόνο οι αποστάσεις των σφαλμάτων και όχι η διεύθυνσή τους, δηλαδή το αν υπερβαίνουν ή υπολείπονται των πραγματικών τιμών, αγνοώντας μία σημαντική πληροφορία. Επίσης λόγω της τετραγωνικής δύναμης τα μεγαλύτερα σφάλματα έχουν πολύ μεγαλύτερη συνεισφορά από τα μικρότερα. Το μέσο τετραγωνικό σφάλμα χρησιμοποιείται και στην γραμμική παλινδρόμηση για την εύρεση της βέλτιστης ευθείας που περιγράφει μία σειρά σημείων.

Στα προβλήματα κατηγοριοποίησης η συνάρτηση κόστους που χρησιμοποιείται στην πλειοψηφία των περιπτώσεων είναι η Cross Entropy.

Cross Entropy Loss

$$L_{CE} = -\sum_{k=1}^{m} t_k * log(p_k)$$

όπου,

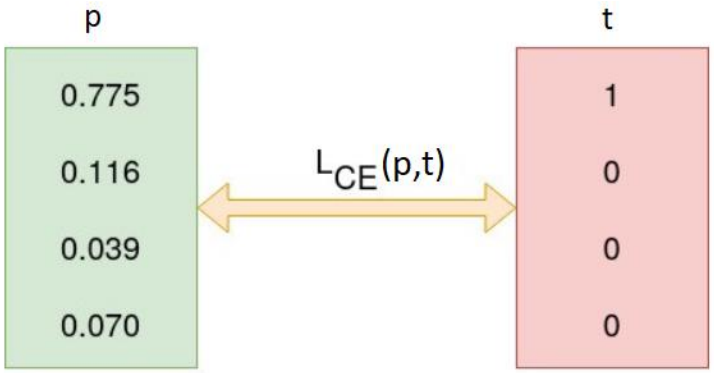
m είναι ο συνολικός αριθμός κλάσεων

p_k είναι η πιθανότητα που δίνει το ΤΝΔ στο να ανήκει η είσοδος στην κλάση k

t_k είναι 1 όταν η είσοδος αντιστοιχεί στην κλάση k και 0 σε όλες τις άλλες περιπτώσεις

Ονομάζεται και συνάρτηση λογαριθμικού κόστους και συνδυάζεται με την συνάρτηση Softmax, ως συνάρτηση ενεργοποίησης του επιπέδου εξόδου. Η πιθανότητα που προβλέπεται για κάθε κλάση πολλαπλασιάζεται με την πραγματική πιθανότητα της κλάσης , που είναι 0 ή 1. Συνεπώς τελικά απομένει μόνο ο λογάριθμος της πιθανότητας που έχει αποδώσει το ΤΝΔ στην σωστή κλάση. Όσο περισσότερο η πρόβλεψη πλησιάζει

την μονάδα τόσο η τιμή της συνάρτησης μειώνεται. Ένα μοντέλο με απόλυτη επιτυχία στις προβλέψεις του μηδενίζει αυτήν την τιμή. Ένα παράδειγμα φαίνεται στην Εικόνα 33 που αποτελεί συνέχεια της προηγούμενης.



Εικόνα 33: Παράδειγμα εφαρμογής της Cross Entropy. Αριστερά φαίνεται η πρόβλεψη του ΤΝΔ για την πιθανότητα κάθε κλάσης ενώ δεξιά η πραγματική πιθανότητα κάθε κλάσης. Η είσοδος εμφανώς ανήκει στην πρώτη κλάση [17].

Η τιμής της συνάρτησης σε αυτό το παράδειγμα είναι:

$$L_{CE} = -[1 * \log(0.775) + 0 * \log(0.116) + 0 * \log(0.039) + 0 * \log(0.07) = 0.2549$$

Όσο προχωράει η εκπαίδευση με τον αλγόριθμο gradient descent αυξάνεται η σύγκλιση μεταξύ προβλέψεων και πραγματικών τιμών αυξάνεται και η τιμή της συνάρτησης μειώνεται.

Στα προβλήματα κατηγοριοποίησης με πολλαπλές κλάσεις μπορεί να χρησιμοποιηθεί και το μέσο τετραγωνικό σφάλμα ως συνάρτηση κόστους αλλά δεν προτιμάται γιατί δίνει περισσότερη έμφαση στις λανθασμένες προβλέψεις και συγκλίνει στην επιθυμητή λύση πιο αργά από την Cross Entropy, λόγω μικρότερης κλίσης στο [0,1] και υψηλότερου υπολογιστικού φορτίου.

2.2.4.5.3 Αλγόριθμος Εκπαίδευσης

Τρίτο αλλά εξίσου σημαντικό στοιχείο είναι ο αλγόριθμος εκπαίδευσης που εφαρμόζεται. Η οπισθόδρομη διάδοση σφάλματος με κάποιον κανόνα ανανέωσης των συντελεστών βαρύτητας που βασίζεται στην αρχή της κατεύθυνσης μείωσης του σφάλματος (gradient descent), αποτελεί τον πλέον διαδεδομένο τρόπο εκπαίδευσης ΤΝΔ πρόσθιας τροφοδότησης με επιτηρούμενη μάθηση. Παρά ταύτα, ο κανόνας του Δέλτα έχει εξελιχθεί σε πιο εξειδικευμένους αλγορίθμους ανανέωσης των συντελεστών βαρύτητας, που συνεισφέρουν τόσο στην ταχύτητα όσο και στην ακρίβεια των αποτελεσμάτων. Παρακάτω παρουσιάζονται αναλυτικά όσοι από αυτούς υλοποιούνται στην παρούσα εργασία.

Levenberg-Marquardt algorithm

Ο αλγόριθμος Levenberg-Marquardt έχει σχεδιαστεί για να χρησιμοποιείται με συναρτήσεις κόστους που έχουν την μορφή αθροίσματος τετραγώνων, όπως το μέσο τετραγωνικό σφάλμα (mse). Χρησιμοποιείται στην επίλυση προβλημάτων μη γραμμικής

παλινδρόμησης. Είναι πιο εύρωστος από άλλους αλγορίθμους που σημαίνει ότι βρίσκει ένα τοπικό ελάχιστο ακόμα και αν ξεκινώντας, απέχει πολύ από αυτό. Η εύρεση το ολικού ελαχίστου εξαρτάται από το πόσο απέχει από αυτό κατά την εκκίνηση.

Παρακάτω παρουσιάζεται αναλυτικά για την περίπτωση ενός επιπέδου.

Το συνολικό σφάλμα σε κάθε επανάληψη ως άθροισμα τετραγώνων, όπως ορίστηκε παραπάνω είναι:

$$E = \sum_{j=1}^{m} \frac{1}{2}(t_j - y_j)^2$$

Ορίζοντας το σφάλμα του κάθε ενός από τους m νευρώνες εξόδου, $e_i = t_i - y_i$ προκύπτει

$$E = \frac{1}{2}\sum_{i=1}^{m} e_i^2 \quad i = 1, \dots, m \quad (*)$$

Εάν το σφάλμα του ίδιου νευρώνα στο επόμενο διάνυσμα εισόδου έχει μικρή διαφορά με το τωρινό τότε μπορεί να εκφραστεί με την μορφή σειράς Taylor:

$$e_i(new) = e_i(old) + (\nabla e_i)^T\big(w(new) - w(old)\big)$$

όπου,

$$\nabla e_i = \begin{bmatrix} \frac{\partial e_i}{\partial w_1} \\ \vdots \\ \frac{\partial e_i}{\partial w_n} \end{bmatrix}_{n\times 1} \quad \text{και} \quad w = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}_{n\times 1}$$

ο πίνακας w περιέχει όλους τους συντελεστές βαρύτητας και τις σταθερές πόλωσης, έστω n σε αριθμό και ο πίνακας της κλίσης του σφάλματος όλες τις μερικές παραγώγους, ως προς τον κάθε συντελεστή.

Σε μητρωική μορφή η παραπάνω σχέση γράφεται:

$$e(new) = e(old) + J_e * \Delta w$$

όπου,

$$e = \begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix}_{m\times 1}, \qquad J_e = \begin{bmatrix} \frac{\partial e_i}{\partial w_1} & \dots & \frac{\partial e_i}{\partial w_n} \\ \vdots & \dots & \vdots \\ \frac{\partial e_m}{\partial w_1} & \dots & \frac{\partial e_m}{\partial w_n} \end{bmatrix}_{m\times n}, \Delta w = \begin{bmatrix} \Delta w_1 \\ \vdots \\ \Delta w_n \end{bmatrix}_{n\times 1}$$

ο πίνακας J_e είναι η Ιακωβιανή μήτρα του διανύσματος των σφαλμάτων e.

Το συνολικό σφάλμα μπορεί πλέον να γραφεί ως:

$$E = \frac{1}{2}\sum_{i=1}^{m}[e_i + (\nabla e_i)^T \Delta w]^2 = \frac{1}{2}||e + J_e * \Delta w||^2$$

Η συνάρτηση αυτή ελαχιστοποιείται στο σημείο μηδενισμού της παραγώγου της που είναι:

$$e + J_e * \Delta w = 0 \leftrightarrow J_e^T * e + J_e^T * J_e * \Delta w = 0 \leftrightarrow \Delta w = -(J_e^T * J_e)^{-1} J_e^T * e$$

με

$$J_{e\,n\times m}^T * J_{e\,m\times n} = H_{E\,n\times n}$$

H_E είναι η Εσσιανή μήτρα της συνάρτησης Ε αγνοώντας τις παραγώγους τρίτης τάξης.

Έτσι η τελική μορφή της λύσης είναι:

$$\Delta w = -H_E{}^{-1} J_e^T * e$$

Συνεπώς η ανανέωση των συντελεστών βαρύτητας και των σταθερών πόλωσης περιλαμβάνει την αντιστροφή της Εσσιανής μήτρας σε κάθε επανάληψη. Για να μπορεί να ελεγχθεί το βήμα μετατόπισης των τιμών, αφού όσο μικρότερο είναι τόσο καλύτερη η προσέγγιση, η συνάρτηση κόστους Ε τροποποιείται ελαφρώς:

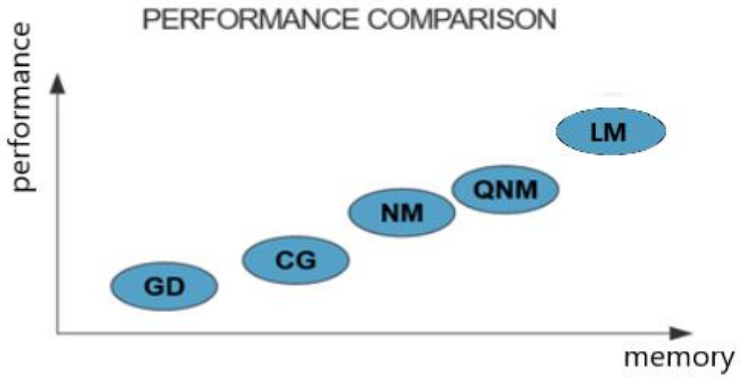
$$E = \frac{1}{2}(||e + J_e * \Delta w||^2 + \lambda * ||\Delta w||^2)$$

και η τελική λύση γίνεται:

$$\Delta w = -(H_E + \lambda I)^{-1} J_e^T * e$$

Με την παράμετρο λ να καθορίζει το βήμα μετατόπισης. Μεγάλες τιμές του λ οδηγούν στον αλγόριθμο Steepest Gradient Descent (GD) με μικρό ρυθμό εκπαίδευσης, ενώ πολύ μικρές στον αλγόριθμο Gauss-Newton, που χρησιμοποιεί την προσέγγιση της Εσσιανής μήτρας. Η μέθοδος Gauss-Newton είναι γρηγορότερη και πιο ακριβής κοντά σε ένα ελάχιστο της συνάρτησης συνεπώς ο στόχος είναι η ταχύτερη κατά το δυνατόν εναλλαγή από την πρώτη μέθοδο στην δεύτερη. Αυτό επιτυγχάνεται με την μείωση της παραμέτρου λ σε κάθε επανάληψη και την αύξησή της μόνο εάν σε κάποια από αυτές η συνάρτηση κόστους αυξηθεί. Έτσι επιταχύνεται η σύγκλιση στο τοπική ελάχιστη τιμή.

Ο αλγόριθμος Levenberg-Marquardt είναι ταχύτερος σε σχέση με την πλειοψηφία των υπολοίπων όμως δεν εφαρμόζεται, σε προβλήματα με συναρτήσεις κόστους που δεν έχουν την μορφή αθροίσματος τετραγώνων, όπως η Cross Entropy, καθώς επίσης και σε προβλήματα με μεγάλες βάσεις δεδομένων όπου ο υπολογισμός της Ιακωβιανής μήτρας απαιτεί πολύ υπολογιστική μνήμη.



Εικόνα 34: Ο αλγόριθμος Levenberg-Marquardt (LM) ξεπερνάει σε επίδοση αλλά και σε απαιτήσεις μνήμης τους υπολοίπους [16].

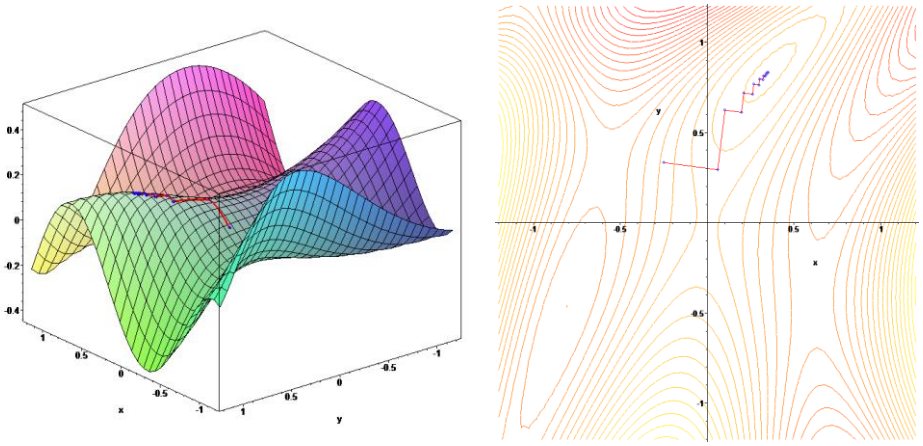
Steepest Gradient Descent (GD) & Scaled Conjugated Gradient Descent (SCG) algorithms

Ο αλγόριθμος Steepest Gradient Descent (GD) όπως περιγράφηκε με τον κανόνα του Δέλτα, μπορεί να εκφραστεί σε μητρωική μορφή ως:

$$\Delta w = -a\nabla E$$

όπου $\nabla E = \begin{bmatrix} \frac{\partial E}{\partial w_1} \\ \vdots \\ \frac{\partial E}{\partial w_n} \end{bmatrix}_{n\times 1}$, $w_1, \dots, w_n$ όλοι οι συντελεστές βαρύτητας και οι σταθερές πόλωσης, έστω n σε αριθμό και a ο ρυθμός εκπαίδευσης.

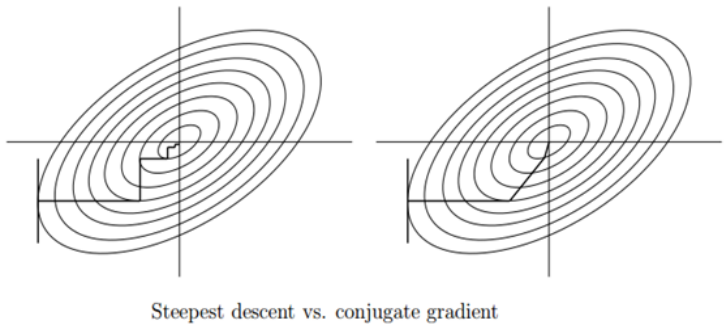
Σε κάθε επανάληψη το διάνυσμα $-\nabla E$ ορίζει την κατεύθυνση μέγιστης μείωσης του σφάλματος, η οποία στον αλγόριθμο είναι πάντα κάθετη στην προηγούμενη. Ο ρυθμός εκπαίδευσης ορίζει το μέγεθος βήματος σε αυτήν την κατεύθυνση. Έτσι σε βάθος ορισμένων επαναλήψεων ο αλγόριθμος συγκλίνει σε κάποιο τοπικό ελάχιστο.



Εικόνα 35: Αριστερά, η τρισδιάστατη απεικόνιση μίας συνάρτησης δύο μεταβλητών. Με βέλη φαίνονται τα διανύσματα μέγιστης κλίσης σε διάφορα σημεία. Δεξιά, διάφορες ισοσταθμικές καμπύλες πάνω στις οποίες η τιμή της συνάρτησης διατηρείται σταθερή. [18]

Όπως φαίνεται στο δεξιό τμήμα της Εικόνας 35 που αφορά μία συνάρτηση δύο μεταβλητών για απλούστευση, ο αλγόριθμος Steepest Gradient Descent (GD) κινείται προς κάποιο τοπικό ελάχιστο βηματικά, με την κατεύθυνση του κάθε βήματος να είναι κάθετη στην προηγούμενη, δημιουργώντας σκαλοπάτια.

Είναι προφανές ότι υπάρχει γρηγορότερος τρόπος μετακίνησης προς το τοπικό ελάχιστο. Μία πολύ γρηγορότερη εναλλακτική είναι το πρώτο βήμα να έχει τέτοιο μέγεθος ώστε το δεύτερο βήμα όντας κάθετο στο πρώτο να οδηγήσει στο τοπικό ελάχιστο. Αυτό εφαρμόζει ο αλγόριθμος Conjugated Gradient Descent, που συγκλίνει στο ολικό ελάχιστο σε βάθος το πολύ n βημάτων.



Εικόνα 36: Σύγκριση του τρόπου μετακίνησης των δύο αλγορίθμων, πάνω στις ισοσταθμικές καμπύλες προς την κατεύθυνση μείωσής τους [19].

Παρακάτω επεξηγείται ο μαθηματικός φορμαλισμός των δύο μεθόδων [20].

Αρχικά γίνονται κάποιες παραδοχές για την απλούστευση της παρουσίασης της μεθόδου, οι οποίες όμως δεν είναι δεσμευτικές καθώς υπάρχει η δυνατότητα γενίκευσης.

Έστω ότι η τιμή του σφάλματος μπορεί να γραφεί στην τετραγωνική μορφή:

$$E = \frac{1}{2} w^T A w - b^T w + c$$

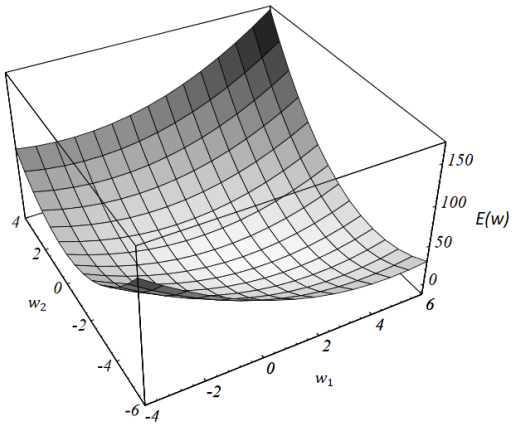
όπου, A είναι πίνακας διαστάσεων n×n, b διάνυσμα διαστάσεων n×1 και c σταθερά.

Αν ο πίνακας A είναι συμμετρικός και θετικά ορισμένος τότε η κλίση του σφάλματος είναι:

$$E' = \nabla E = \frac{1}{2} A^T w + \frac{1}{2} A w - b = A w - b$$

και από το σημείο μηδενισμού της προκύπτει το ολικό, σε αυτήν την ειδική περίπτωση του A, ελάχιστο:

$$A w - b = 0 \leftrightarrow A w = b$$



Εικόνα 37: Η τετραγωνική μορφή της συνάρτησης σφάλματος που είναι παραβολοειδής, όταν ο πίνακας A είναι θετικά ορισμένος. Η ελάχιστη τιμή βρίσκεται στο κατώτατο σημείο [20].

Η διαδικασία εύρεσής του ελαχίστου ξεκινά από ένα τυχαίο σημείο w_0 και προχωρά με διαδοχικά βήματα $w_1, w_2, \ldots$ μέχρι κάποιο σημείο που βρίσκεται ικανοποιητικά κοντά στο βέλτιστο *w.*

Από κάθε σημείο το επόμενο βήμα γίνεται στην κατεύθυνση μέγιστης μείωσης του σφάλματος που είναι η $r_i = -\nabla E_i = b - Aw_i$. Θέτοντας ως w^* το τελικό *w* που είναι ο στόχος για να ελαχιστοποιηθεί το σφάλμα Ε, προκύπτει το σφάλμα του w_i σε κάθε επανάληψη ως:

$$e_i = w_i - w^*$$

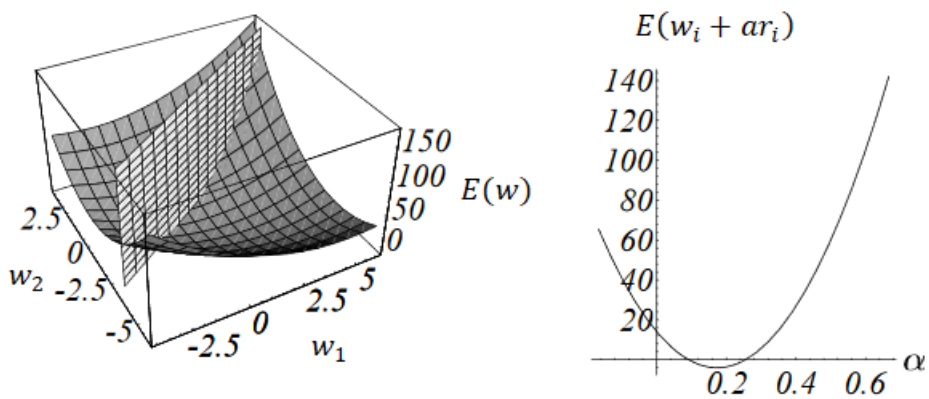
τότε

$$r_i = b - A(e_i + w^*) = b - Aw^* - Ae_i = -Ae_i$$

Συνεπώς για το πρώτο βήμα:

$$w_1 = w_0 + ar_0$$

Η τιμή της μεταβλητής *α*, που καθορίζει το μέγεθος του βήματος στην κατεύθυνση του διανύσματος r_i υπολογίζεται ελαχιστοποιώντας το σφάλμα *E* κατά μήκος μίας καμπύλης. Πρόκειται για την καμπύλη που προκύπτει από την τομή του κάθετου επιπέδου που περιέχει το διάνυσμα r_i με την συνάρτηση του σφάλματος.



Εικόνα 38: Αριστερά, η τομή του κάθετου επιπέδου, που περιέχει το r_i με την συνάρτηση σφάλματος [20]. Δεξιά, η καμπύλη που προκύπτει. Στόχος αποτελεί η εύρεση του κατώτατου σημείου της [20].

Η τιμή του *α* είναι τέτοια ώστε να οδηγεί στο κατώτατο σημείο της καμπύλης, στο οποίο η κατευθυνόμενη παράγωγος $\frac{dE(w_{i+1})}{da}$ μηδενίζεται.

$$\frac{dE(w_{i+1})}{da} = 0 \leftrightarrow (\nabla E_{i+1})^T * \frac{dw_{i+1}}{da} = 0 \leftrightarrow (\nabla E_{i+1})^T * r_i = 0 \leftrightarrow -r_{i+1}^T * r_i = 0$$

Όπως φαίνεται από την παραπάνω εξίσωση η τιμή του α είναι τέτοια ώστε το διάνυσμα κατεύθυνσης κάθε βήματος να είναι ορθογώνιο σε σχέση με το προηγούμενο. Αυτό σημαίνει ότι στο κατώτατο σημείο της κάθε καμπύλης το διάνυσμα μέγιστης μείωσης του σφάλματος Ε είναι κάθετο στο επίπεδο της καμπύλης. Έτσι δικαιολογείται η εμφάνιση σκαλοπατιών στον αλγόριθμο Steepest Gradient Descent.

Άρα για τον υπολογισμό του *α* στην πρώτη επανάληψη:

$$r_1^T r_0 = 0 \leftrightarrow (b - Aw_1)^T r_0 = 0 \leftrightarrow \left(b - A(w_0 + ar_0)\right)^T r_0 = 0 \leftrightarrow (b - Aw_0)^T r_0$$
$$= a(Ar_0)^T r_0 \leftrightarrow r_0^T r_0 = ar_0^T A r_0 \leftrightarrow a = \frac{r_0^T r_0}{r_0^T A r_0}$$

Έτσι προσδιορίζεται το κατώτατο σημείο και γίνεται το βήμα προς εκείνο. Στη συνέχεια η διαδικασία επαναλαμβάνεται. Συνεπώς συνολικά ο αλγόριθμος Steepest Gradient Descent είναι:

$$r_i = b - Aw_i$$
$$a_i = \frac{r_i^T r_i}{r_i^T A r_i}$$
$$w_{i+1} = w_i + a_i r_i$$

Σημειώνεται ότι ο αλγόριθμος λειτουργεί με μικρές παραλλαγές και στην περίπτωση που ο πίνακας A δεν είναι συμμετρικός. Ακόμα λειτουργεί και σε αρκετές περιπτώσεις όπου ο A δεν είναι θετικά ορισμένος.

Όπως αναφέρθηκε παραπάνω μία εξέλιξη του αλγορίθμου Steepest Gradient Descent αποτελεί ο αλγόριθμος Scaled Conjugated Gradient Descent, ο οποίος επιχειρεί να εξαλείψει τα πολλαπλά βήματα που γίνονται προς την ίδια κατεύθυνση κάνοντας σε κάθε κατεύθυνση ένα βήμα με το σωστό συνολικό μήκος [20].

Έτσι επιλέγεται ένα σύνολο από n ορθογώνια συζευγμένα διανύσματα: $d_0, d_1, \dots, d_{n-1}$ τέτοια ώστε να ισχύει:

$$d_i^T A d_j = 0 \quad \mu\varepsilon \quad i, j = 1, \dots, n-1 \ \& \ i \neq j$$

Στην κατεύθυνση καθενός από αυτά πραγματοποιείται ακριβώς ένα βήμα με το κατάλληλο μήκος. Μετά το πέρας n βημάτων, ο αλγόριθμος καταλήγει στην επιθυμητή λύση.

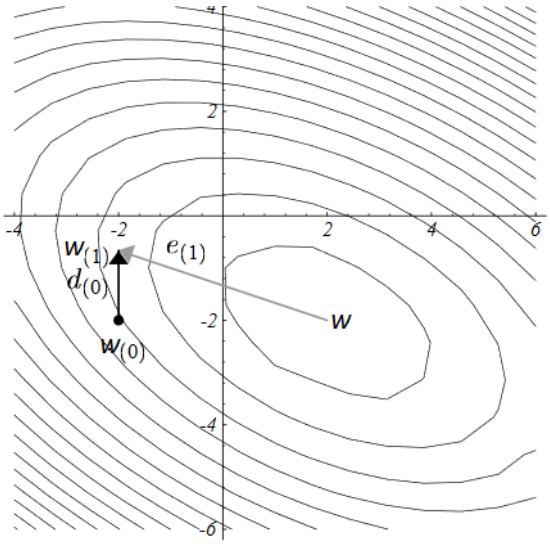
Το κάθε επόμενο σημείο επιλέγεται ως εξής:

$$w_{i+1} = w_i + a_i d_i$$

Η τιμή της μεταβλητής α, που καθορίζει το μέγεθος του βήματος στην κατεύθυνση του διανύσματος d_i, υπολογίζεται όπως και πριν, με την εύρεση του τοπικού ελαχίστου της καμπύλης που προκύπτει από την τομή του κάθετου επιπέδου που εμπεριέχει το διάνυσμα με την συνάρτηση του σφάλματος. Στο σημείο εκείνο η κατευθυνόμενη παράγωγος $\frac{dE(w_{i+1})}{da}$ μηδενίζεται.

$$\frac{dE(w_{i+1})}{da} = 0 \leftrightarrow (\nabla E_{i+1})^T * \frac{dw_{i+1}}{da} = 0 \leftrightarrow -r_{i+1}^T d_i = 0 \leftrightarrow (-Ae_{i+1})^T d_i = 0$$
$$\leftrightarrow e_{i+1}^T A d_i = 0 \leftrightarrow d_i^T A e_{i+1} = 0$$

Συνεπώς αποδεικνύεται ότι σφάλμα e_{i+1} είναι ορθογώνια συζευγμένο με το d_i στο τοπικό ελάχιστο της καμπύλης, που σημαίνει ότι από εκείνο το σημείο στην κατεύθυνση του d_i δεν χρειάζεται άλλο βήμα.



Εικόνα 39: Το σφάλμα e_1 είναι ορθογώνια συζευγμένο με το διάνυσμα d_0 που σημαίνει ότι στην κατεύθυνση του d_0 δεν χρειάζεται να πραγματοποιηθεί άλλο βήμα [20].

Συνεχίζοντας,

$$d_i^T A e_{i+1} = 0 \leftrightarrow d_i^T r_{i+1} = 0 \leftrightarrow d_i^T (b - A w_{i+1}) = 0 \leftrightarrow d_i^T (b - A(w_i + a_i d_i)) = 0$$
$$\leftrightarrow d_i^T (b - A w_i - A a_i d_i) = 0 \leftrightarrow d_i^T r_i - a_i d_i^T A d_i = 0 \leftrightarrow a_i = \frac{d_i^T r_i}{d_i^T A d_i}$$

Προφανώς αν τα ορθογώνια συζευγμένα διανύσματα d_i ταυτιστούν με τα διανύσματα της κλίσης, r_i τότε ο αλγόριθμος ταυτίζεται με τον προηγούμενο. Για να βρεθούν τα d_i ακολουθείται η μέθοδος Gram-Schmidt κατά την οποία:

‘Εστω n γραμμικώς ανεξάρτητα διανύσματα $u_0, u_1, \dots, u_{n-1}$. Ορίζεται $d_0 = u_0$. Για να κατασκευαστεί το κάθε διάνυσμα d_i αφαιρούνται από το διάνυσμα u_i όλες οι συνιστώσες που δεν είναι ορθογώνια συζευγμένες με τα προηγούμενα d διανύσματα. Συνεπώς το κάθε διάνυσμα d_i που δημιουργείται είναι ορθογώνια συζευγμένο με τα προηγούμενα:

$$d_i = u_i + \sum_{j=1}^{i-1} \beta_{ij} d_j$$

όπου το β_{ij} με $i > j$,προκύπτει ως εξής:

$$d_i^T A d_j = u_i^T A d_j + \underbrace{\sum_{k=1}^{i-1} \beta_{ik} d_k^T A d_j}_{\gamma\iota\alpha\ k \neq j\ \iota\sigma\chi\acute{\upsilon}\varepsilon\iota\ d_k^T A d_j = 0} \leftrightarrow 0 = u_i^T A d_j + \beta_{ij} d_j^T A d_j \leftrightarrow \beta_{ij} = -\frac{u_i^T A d_j}{d_j^T A d_j}$$

Συνεπώς συνολικά ο αλγόριθμος Scaled Conjugated Gradient Descent είναι:

$$w_{i+1} = w_i + a_i d_i$$
$$a_i = \frac{d_i^T r_i}{d_i^T A d_i}$$

$$r_i = b - Aw_i$$

$$d_i = u_i + \sum_{j=1}^{i-1} \beta_{ij} d_j$$

$$\beta_{ij} = -\frac{u_i^T A d_j}{d_j^T A d_j}$$

με $u_0, u_1, \dots, u_{n-1}$ γραμμικώς ανεξάρτητα.

Ο αλγόριθμος Scaled Conjugated Gradient Descent είναι γρηγορότερος από τον Steepest Gradient Descent και δεν απαιτεί τον υπολογισμό της Εσσιανής μήτρας σε κάθε επανάληψη όπως ο αλγόριθμος Levenberg-Marquardt.

2.2.4.5.4 Πίνακας Σύγχυσης

Ο πίνακας σύγχυσης είναι ένας τετραγωνικός πίνακας $\mathrm{N} \times \mathrm{N}$, όπου N ο αριθμός των κλάσεων, που χρησιμοποιείται στα προβλήματα κατηγοριοποίησης και επιτρέπει την οπτικοποίηση της απόδοσης του αλγορίθμου. Κάθε γραμμή του αναπαριστά την πραγματική κατηγορία στην οποία ανήκουν τα δεδομένα, ενώ κάθε στήλη του την κατηγορία στην οποία έχει εκτιμηθεί από τον αλγόριθμο ότι αυτά ανήκουν. Κατά συνέπεια εάν η πρόβλεψη συμπίπτει με την πραγματικότητα τότε αθροίζεται στο αντίστοιχο κελί της διαγωνίου του πίνακα. Αυτό σημαίνει ότι για την κατηγορία i (i=1,2,…,N) το σύνολο των σωστών προβλέψεων αναγράφεται στο κελί (i,j), με j=i. Σε κάθε ένα από τα υπόλοιπα κελιά της γραμμής i αναγράφονται οι λανθασμένες προβλέψεις των στοιχείων της κλάσης i που καταχωρήθηκαν στην κλάση της στήλης j με $\mathrm{j} \neq \mathrm{i}$. Αντίστοιχα σε κάθε ένα από τα υπόλοιπα κελιά της στήλης i αναγράφονται το πλήθος των στοιχείων της κλάσης j, με $\mathrm{j} \neq \mathrm{i}$, τα οποία καταχωρήθηκαν λανθασμένα στην κλάση i. Μία γενική μορφή πίνακα σύγχυσης φαίνεται παρακάτω:

Πίνακας 1: Γενική μορφή πίνακα σύγχυσης.

		Κατηγορία Πρόβλεψης					
		1	**2**	**…**	**j**	**…**	**N**
Πραγματική Κατηγορία	**1**	a_{11}	a_{12}	…	a_{1j}	…	a_{1N}
	2	a_{21}	a_{22}	…	a_{2j}	…	a_{2N}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	i	a_{i1}	a_{i2}	…	a_{ij}	…	a_{iN}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	N	a_{N1}	a_{N2}	…	a_{Nj}	…	a_{NN}

Με κίτρινο χρώμα, στην διαγώνιο του Πίνακα 1, φαίνονται οι σωστές προβλέψεις της κάθε κλάσης. Για παράδειγμα το στοιχείο α_{22} είναι το σύνολο των φορών που οι καταχωρήσεις δεδομένων στην κατηγορία 2 ήταν ορθές. Αντίθετα, με γρι χρώμα απεικονίζονται οι λανθασμένες προβλέψεις. Για παράδειγμα το στοιχείο α_{21} είναι το σύνολο των φορών που καταχωρήθηκαν δεδομένα στην κλάση 1 ενώ ανήκουν στην κλάση 2. Συνεπώς επαγωγικά προκύπτει ότι:

- **Σύνολο προβλέψεων** = $\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}$
- **Σύνολο ορθών προβλέψεων** = $\sum_{i=1}^{N} a_{ii}$

Με τα παραπάνω, ορίζονται οι ακόλουθοι δείκτες:

- **Ακρίβεια Πίνακα (Precision):** Συχνότητα ορθών προβλέψεων

$$\boldsymbol{P = \frac{O.\Pi.}{\Sigma.\Pi.}}$$

- **Ακρίβεια Κλάσης i:** Συχνότητα με την οποία προβλέπεται σωστά ότι ένα στοιχείο ανήκει στην κλάση i

$$\boldsymbol{P_i = \frac{a_{ii}}{\sum_{j=1}^{N} a_{ji}}}$$

- **Ανάκληση Κλάσης i (Recall):** Συχνότητα με την οποία ένα στοιχείο της κλάσης i καταχωρείται σωστά

$$\boldsymbol{R_i = \frac{a_{ii}}{\sum_{j=1}^{N} a_{ij}}}$$

Γενικότερα όταν σε μία γραμμή υπάρχουν πολλά κελιά με μη μηδενικές τιμές, εκτός του διαγώνιου, αυτό σημαίνει ότι ο αλγόριθμος κατηγοριοποιεί λανθασμένα τα δεδομένα της κατηγορίας και συνεπώς χρειάζεται γενίκευση γιατί δεν έχει εκπαιδευτεί επαρκώς σε αυτήν (underfitting). Αντιθέτως όταν σε μία στήλη υπάρχουν πολλά κελιά με μη μηδενικές τιμές, εκτός του διαγώνιου, αυτό σημαίνει ότι ο αλγόριθμος κατηγοριοποιεί περισσότερα δεδομένα στην κατηγορία αυτή από όσα πρέπει (overfitting) και συνεπώς χρειάζεται εξειδίκευση.

2.2.4.6 Εφαρμογές Τεχνητών Νευρωνικών Δικτύων

Τα τεχνητά νευρωνικά δίκτυα χρησιμοποιούνται πλέον σε πολλούς τομείς λόγω της ευελιξίας, της ικανότητας αντιμετώπισης πολύπλοκων μη γραμμικών φαινομένων για τα οποία δεν υπάρχει ακριβής περιγραφή και λύση και της ευκολίας υλοποίησής τους. Οι αλγόριθμοι εκπαίδευσης έχουν εξελιχθεί ώστε να ανταποκρίνονται στις ιδιαιτερότητες της αρχιτεκτονικής και των επιμέρους χαρακτηριστικών καθιστώντας έτσι δυνατή την εφαρμογή των ΤΝΔ σε όλο και πιο σύνθετα προβλήματα. Μερικοί από τους τομείς εφαρμογής τους είναι οι εξής:

- Αναγνώριση προτύπων (π.χ. εξαγωγή χαρακτηριστικών από κείμενα, αναγνώριση φωνής και προσώπου)
- Επεξεργασία σήματος (π.χ. αφαίρεση θορύβου τηλεφωνικού σήματος, ανάλυση σεισμικού σήματος, επεξεργασία εικόνας και μηχανική όραση)

- Ιατρική (π.χ. ανάλυση ηλεκτροκαρδιογραφήματος, ιατρική διάγνωση και θεραπεία με βάση τα συμπτώματα, πρόταση κατάλληλης φαρμακευτικής αγωγής)
- Άμυνα (π.χ. κωδικοποίηση σημάτων ραντάρ, αναγνώριση στόχων, υποβρύχια ανίχνευση ναρκών)
- Οικονομία (π.χ. πρόβλεψη τιμής μετοχών, εκτίμηση φερεγγυότητας δανειζόμενου πελάτη, εκτίμηση αξίας ακινήτων)
- Περιβάλλον (π.χ. πρόγνωση καιρού, ανάλυση γεωλογικών δεδομένων για εξεύρεση πετρελαίου)
- Μεταφορές (π.χ. αυτόματοι πιλότοι, ανίχνευση ελαττωμάτων σε τμήματα της μηχανής και της ατράκτου)
- Βιομηχανία (π.χ. ρομπότ, έλεγχος ποιότητας, αυτοματισμοί)
- Δυναμικά εξελισσόμενα συστήματα (π.χ. πρόβλεψη χρονοσειρών, εκτίμηση κατάστασης σήματος)
- Τεχνητή νοημοσύνη (π.χ. επικοινωνία ανθρώπου-υπολογιστή)

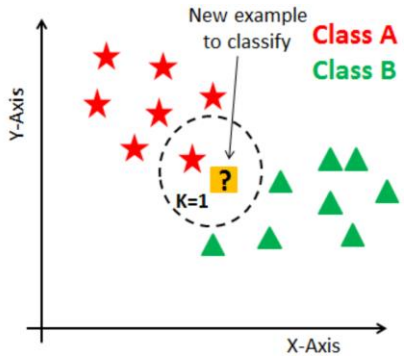
Ως μειονέκτημα των ΤΝΔ μπορεί αν θεωρηθεί το γεγονός ότι δεν υπάρχει γενικός τρόπος ερμηνείας της εσωτερικής λειτουργία τους, γεγονός που κάνει την εκ των προτέρων γνώση για την επιτυχία τους σε κάθε συγκεκριμένη εφαρμογή, ελλιπή. Αντιθέτως η επιτυχία τους έγκειται στην προσαρμογή τους στην κάθε περίπτωση.

2.3 Αλγόριθμος k-Κοντινότερων Γειτόνων (k-nearest neighbors)

Ο αλγόριθμος k-κοντινότερων γειτόνων (knn) αποτελεί μία ξεχωριστή μέθοδο μηχανικής μάθησης, διαφορετική από τα ΤΝΔ, για την επίλυση προβλημάτων επιτηρούμενης εκπαίδευσης, που όπως έχει αναφερθεί διακρίνονται σε προβλήματα κατηγοριοποίησης (classification) και παλινδρόμησης (regression). Παρουσιάστηκε το 1951 από τον Joseph Hodges και την Evelyn Fix. Είναι μη παραμετρικός αλγόριθμος, με την έννοια ότι δεν απαιτεί την προσαρμογή παραμέτρων στα δεδομένα, για την επίλυση του προβλήματος, όπως απαιτούν τα ΤΝΔ. Συνεπώς ο αλγόριθμος δεν μαθαίνει από τα δεδομένα και το μόνο που χρειάζεται κατά την εκπαίδευση είναι η αποθήκευση τους. Παράλληλα δεν απαιτεί τον σχεδιασμό ενός μοντέλου. Κατά κύριο λόγο εφαρμόζεται σε προβλήματα κατηγοριοποίησης, που αποτελούν αντικείμενο της παρούσας εργασίας και αναλύονται εκτενέστερα.

Βασίζεται στην λογική ότι παρόμοια δεδομένα έχουν μεγάλη εγγύτητα, δηλαδή βρίσκονται κοντά το ένα στο άλλο. Η κάθε νέα είσοδος κατηγοριοποιείται με βάση τα ήδη υπάρχοντα δεδομένα που βρίσκονται σε κοντινή απόσταση. Συμπερασματικά, η ομοιότητα που μεταφράζεται σε μικρότερη απόσταση της εισόδου με κάποια από τα δεδομένα είναι ο καθοριστικός παράγοντας που κρίνει το αποτέλεσμα. Ο αριθμός των χαρακτηριστικών που έχουν οι κλάσεις καθορίζει την διάσταση του χώρου απεικόνισης των δεδομένων. Στο επίπεδο αποτυπώνονται δεδομένα προβλημάτων, με δύο χαρακτηριστικά. Για παράδειγμα στην Εικόνα 40, υπάρχουν οι κλάσεις Α και Β (με δύο

χαρακτηριστικά x και y), που καταλαμβάνουν διαφορετικές περιοχές και η κάθε νέα είσοδος ταξινομείται σε μία από τις δύο, αναλόγως με το ποια βρίσκεται πλησιέστερα.



Εικόνα 40: Κατηγοριοποίηση με βάση την πλησιέστερη κλάση [21].

Στον αλγόριθμο knn κύριο ρόλο παίζει η παράμετρος k, η οποία καθορίζει το πλήθος των κοντινότερων δεδομένων των οποίων οι κλάσεις λαμβάνονται υπόψιν προκειμένου να κατηγοριοποιηθεί η εκάστοτε είσοδος. Εν τέλη η είσοδος εκχωρείται στην κλάση που πλειοψηφεί ανάμεσα στις κλάσεις των k πλησιέστερων δεδομένων. Προφανώς όταν k=1 η είσοδος εκχωρείται στην κλάση του κοντινότερου γείτονα. Επίσης, όταν k είναι άρτιος αριθμός τότε υπάρχει η πιθανότητα δύο κλάσεις να ισοψηφήσουν, για αυτό και επιλέγονται περιττοί αριθμοί.

Στα προβλήματα παλινδρόμησης η τιμή του αποτελέσματος είναι η μέση τιμή των k κοντινότερων γειτόνων.

Το μέτρο της απόστασης που καθορίζει την απόσταση μεταξύ των εισόδων και των κλάσεων είναι συνήθως η Ευκλείδεια απόσταση. Εναλλακτικές επιλογές, που χρησιμοποιούνται σπανιότερα, αποτελούν οι αποστάσεις Manhattan και Minkowski:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad Ευκλείδια\ απόσταση$$

$$\sum_{i=1}^{n} |\, x_i - y_i| \qquad Απόσταση\ Manhattan$$

$$\left(\sum_{i=1}^{n}(|x_i - y_i|)^q\right)^{\frac{1}{q}} \qquad Απόσταση\ Minkowski$$

όπου

n: ο αριθμός των χαρακτηριστικών των κλάσεων

$x_{1\times n}$: το διάνυσμα προς κατηγοριοποίηση

$y_{1\times n}$: κάποιο προϋπάρχον διάνυσμα

Παρακάτω παρατίθεται ένα απλουστευμένο πρόβλημα κατηγοριοποίησης, που αφορά την φερεγγυότητα ενός δανειολήπτη. Τα δεδομένα έχουν την μορφή πίνακα, πράγμα που αποτελεί απαίτηση του αλγορίθμου.

Με κριτήριο την ηλικία και το ποσό του δανείου επιχειρείται η αξιολόγηση του κατά πόσο ο δανειολήπτης πρόκειται να αποπληρώσει το δάνειο. Η εκτίμηση βασίζεται σε υπάρχοντα δεδομένα που περιλαμβάνουν τις ηλικίες και τα ποσά άλλων δανειοληπτών όπως και το αν τελικά αποπλήρωσαν το δάνειο ή όχι. Σε ένα πραγματικό πρόβλημα ενδεχομένως να περιλαμβάνονται περισσότερα κριτήρια, όπως το ποσό του μηνιαίου εισοδήματος του, ίσως και περισσότερες κατηγορίες, όπως «Φερέγγυος», «Πολύ φερέγγυος», «Καθόλου φερέγγυος», με στόχο την μεγαλύτερη ακρίβεια, όμως ο τρόπος υπολογισμού των αποστάσεων παραμένει ο ίδιος.

Πίνακας 2: Στοιχεία δανειοληπτών.

Αρίθμηση	Ηλικία	Ποσό (€)	Αποπληρωμή	Απόσταση
1	25	40000	Όχι	102000
2	35	60000	Όχι	82000
3	45	80000	Όχι	62000
4	20	20000	Όχι	122000
5	35	120000	Όχι	22000
6	52	18000	Όχι	124000
7	23	95000	Ναι	47000
8	40	62000	Ναι	80000
9	60	100000	Όχι	42000
10	48	220000	Ναι	78000
11	33	150000	Ναι	8000
12	48	142000	?	

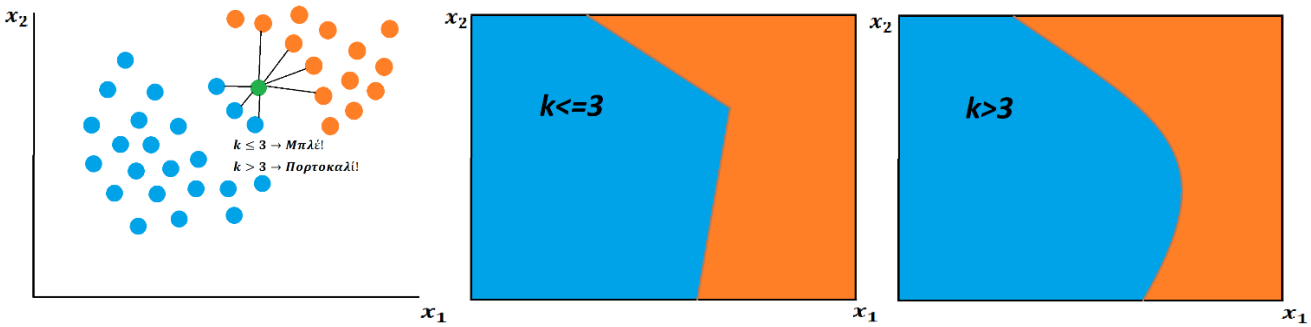
Στην τελευταία γραμμή του Πίνακα 2 φαίνονται η ηλικία και το δάνειο του τελευταίου δανειολήπτη. Στην δεξιότερη στήλη υπολογίζεται η ευκλείδεια απόστασή του από τους προηγούμενους. Για παράδειγμα:

$$d(x_2, y_2) = \sqrt{(48-35)^2 + (142000-60000)^2} \approx 82000$$

Στην περίπτωση που k=1 η κοντινότερη απόσταση είναι αυτή του δανειολήπτη 11 ο οποίος αποπλήρωσε το δάνειο. Συνεπώς προβλέπεται ότι και ο τελευταίος δανειολήπτης πρόκειται να πληρώσει το δάνειο.

Αντιθέτως αν k=3 τότε οι 3 κοντινότεροι γείτονες είναι οι δανειολήπτες 11, 5 και 9. Οι δύο τελευταίοι δεν έχουν αποπληρώσει τα δάνειά τους, με αποτέλεσμα η εκτίμηση τώρα να είναι ότι και ο τελευταίος δεν θα αποπληρώσει το δικό του.

Συνεπώς σε αυτήν την απλή περίπτωση φαίνεται πόσο μεγάλη επίδραση έχει η επιλογή της παραμέτρου k. Μικρές τιμές του k αυξάνουν την επίδραση του θορύβου στην λύση, που συνεπάγεται αστάθεια. Αυτό σημαίνει ότι τα όρια των περιοχών είναι δαντελωτά, και δημιουργούνται τοπικά μικρές περιοχές άλλων κατηγοριών μέσα στην κύρια περιοχή μιας κατηγορίας. Επίσης υπάρχει η πιθανότητα για «overfitting» (βλ. § 2.2.4.2 Εκπαίδευση). Όσο αυξάνεται η τιμή του k αυξάνεται η ικανότητα γενίκευσης και μειώνεται ο θόρυβος, που συνεπάγεται ευστάθεια στις λύσεις με αποτέλεσμα αμιγείς περιοχές, με ομαλά όρια χωρίς αιχμές. Όμως ταυτόχρονα αυξάνεται το υπολογιστικό φορτίο όπως και η πιθανότητα σφάλματος από κάποιες τιμές του k και έπειτα.



(α) (β) (γ)

Εικόνα 41: (α) Η πράσινη κουκίδα αν k<=3 εκχωρείται στις Μπλε ενώ αν k>3 στις Πορτοκαλί. (β) Για μικρές τιμές του k οι περιοχές εμφανίζουν απότομες γωνίες ενώ (γ) για μεγαλύτερες τιμές είναι πιο ομαλές.

Η εύστοχη επιλογή της τιμής του k εξαρτάται από το πρόβλημα. Μία συχνή επιλογή είναι η τετραγωνική ρίζα του αριθμού των χαρακτηριστικών των κλάσεων:

$$k = \sqrt{n}$$

Μία εναλλακτική τακτική είναι να δοκιμάζονται τιμές του k μεταξύ του 1% και 10% του αριθμού των δεδομένων. Ο έλεγχος της απόδοσης του αλγορίθμου για διαφορετικές τιμές του k ονομάζεται Διασταυρούμενη Επικύρωση (Cross Validation).

Στην περίπτωση που οι τιμές που λαμβάνουν τα χαρακτηριστικά διαφέρουν σε τάξη μεγέθους μεταξύ τους, όπως στο παραπάνω παράδειγμα, κρίνεται σκόπιμη η κανονικοποίηση τους, προκειμένου όλα τα δεδομένα να συνεισφέρουν επαρκώς στο αποτέλεσμα. Ένας απλός τρόπος κανονικοποίησης είναι από κάθε χαρακτηριστικό, δηλαδή από κάθε στήλη να εξαχθεί η μέγιστη και η ελάχιστη τιμή και όλες οι τιμές να αναχθούν στο διάστημα [0,1] ως εξής:

$$x_{ij} \rightarrow \frac{x_{ij} - \min_j (x_{ij})}{\max_j(x_i) - \min_j (x_i)} \quad , \quad i = 1{:}\,n \quad , \quad j = 1{:}\,\textit{αριθμός δεδομένων} \qquad (1.4)$$

Με αντίστοιχο τρόπο είναι δυνατή η αναγωγή στο διάστημα [-1,1] ως εξής:

$$x_{ij} \rightarrow \frac{2x_{ij} - \left(\max_j(x_i) + \min_j (x_{ij})\right)}{\max_j(x_i) - \min_j (x_i)} \quad , \qquad i = 1{:}\,n, j = 1{:}\,\textit{αριθμός δεδομένων} \; (1.5)$$

Εάν εφαρμοστεί αυτή η πρώτη μέθοδος στο παραπάνω παράδειγμα τότε ο Πίνακας 2 παίρνει την μορφή:

Πίνακας 3: Στοιχεία δανειοληπτών με κανονικοποίηση.

Αρίθμηση	Ηλικία	Ποσό (€)	Αποπληρωμή	Απόσταση	Αναγωγή Ηλικίας	Αναγωγή Ποσού	Απόσταση
1	**25**	40000	Όχι	102000	0,125	0,11	0,77
2	**35**	60000	Όχι	82000	0,375	0,21	0,52
3	**45**	80000	Όχι	62000	0,625	0,31	0,32
4	**20**	20000	Όχι	122000	0	0,01	0,92
5	**35**	120000	Όχι	22000	0,375	0,50	0,34
6	**52**	18000	Όχι	124000	0,8	0,00	0,62
7	**23**	95000	Ναι	47000	0,075	0,38	0,67
8	**40**	62000	Ναι	80000	0,5	0,22	0,44
9	**60**	100000	Όχι	42000	1	0,41	0,37
10	**48**	220000	Ναι	78000	0,7	1,00	0,39
11	**33**	150000	Ναι	8000	0,325	0,65	0,38
12	**48**	142000	?	?	0,7	0,61	?
Max	**60**	**220000**					
Min	**20**	**18000**					

Η Ευκλείδεια απόσταση της τελευταίας στήλης υπολογίστηκε με τον ίδιο τρόπο.

Πλέον αν k=1 η κοντινότερη απόσταση είναι αυτή του δανειολήπτη 3 ο οποίος δεν αποπλήρωσε το δάνειο. Συνεπώς προβλέπεται ότι και ο τελευταίος δανειολήπτης δεν πρόκειται να πληρώσει το δάνειο, που αποτελεί διαφορετική πρόβλεψη σε σχέση με αντίστοιχη περίπτωση χωρίς κανονικοποίηση. Η παρούσα είναι ασφαλέστερη, γιατί με την κανονικοποίηση η εγγύτητα της ηλικίας παίζει καθοριστικότερο ρόλο σε σχέση με πριν.

Στην περίπτωση που k=3 τότε οι 3 κοντινότεροι γείτονες είναι οι δανειολήπτες 3, 5 και 9. Και οι τρεις δεν έχουν αποπληρώσει το δάνειο, πράγμα που κάνει την πρόβλεψη για αποπλήρωση του δανείου από τον τελευταίο δανειολήπτη αρνητική. Η πρόβλεψη αυτή ταυτίζεται με την αντίστοιχη της περίπτωσης χωρίς κανονικοποίηση.

Επιπλέον, σε ορισμένες εφαρμογές κάποια κριτήρια κρίνονται σπουδαιότερης σημασίας από κάποια άλλα. Σε αυτές τις περιπτώσεις είναι δυνατή η εισαγωγή συντελεστών βαρύτητας, που πολλαπλασιάζουν το κάθε κριτήριο. Επιπλέον, οι συντελεστές βαρύτητας μπορούν να χρησιμοποιηθούν για να κάνουν τις κοντινές αποστάσεις να συνεισφέρουν περισσότερο στο τελικό αποτέλεσμα. Για παράδειγμα στην παραπάνω εφαρμογή, εάν k=5 και 3 από τους κοντινότερους γείτονες έχουν αποπληρώσει το δάνειο ενώ 2 όχι, αλλά οι 2 αυτοί έχουν τις μικρότερες αποστάσεις τότε η βαρύτητά τους μπορεί

να ενισχυθεί με ένα συντελεστή π.χ. $2 * 0.7 = 1.4$ και $3 * 0.3 = 0.9$ αλλάζοντας την τελική πρόβλεψη.

Συγκεντρωτικά, στα πλεονεκτήματα του αλγορίθμου περιλαμβάνονται η ευκολία στην χρήση, η ευρωστία στον θόρυβο των δεδομένων, η ευελιξία στην στάθμισή τους και η ικανότητα της διαχείρισης μεγάλων βάσεων δεδομένων. Στα μειονεκτήματά του συγκαταλέγεται η δυσκολία της εύρεσης της κατάλληλης τιμής του k που είναι μείζονος σημασίας και το αυξημένο υπολογιστικό φορτίο, που είναι ανάλογο της βάσης δεδομένων αφού για κάθε νέα είσοδο υπολογίζονται οι αποστάσεις με όλα τα δεδομένα, μειώνοντας καθοριστικά την ταχύτητα.

Τέλος ο αλγόριθμος λειτουργεί και με μη αριθμητικά δεδομένα, περίπτωση που δεν εξετάζεται εκτενέστερα καθώς δεν αφορά την παρούσα εργασία.

3 Βιβλιογραφική Ανασκόπηση

3.1 Αλγόριθμοι κατηγοριοποίησης σε διαγράμματα ποιότητας

Με σκοπό την αναγνώριση τάσεων σε διαγράμματα ποιότητας, πέραν των ΤΝΔ και της μεθόδου knn, έχουν αναπτυχθεί ποικίλοι αλγόριθμοι κατηγοριοποίησης. Σε αρκετές έρευνες, συγκρίνεται η ακρίβεια των ΤΝΔ με την ακρίβεια ορισμένων εξ αυτών και για τον λόγο αυτό παρουσιάζονται επιγραμματικά οι κυριότεροι.

Λογιστική παλινδρόμηση (Logistic Regression)

Ο αλγόριθμος χρησιμοποιεί μία λογιστική συνάρτηση όπως η Δυαδική σιγμοειδής συνάρτηση (βλ. § 2.2.4.4 Συναρτήσεις ενεργοποίησης) και η συνάρτηση Softmax (βλ. § 2.2.4.5.1 Συναρτήσεις ενεργοποίησης) για να υπολογίσει την πιθανότητα ενός αποτελέσματος. Βρίσκει εφαρμογή τόσο σε προβλήματα δύο κλάσεων όσο και σε προβλήματα πολλαπλών κλάσεων. Σε αμφότερα υπολογίζει την πιθανότητα κάθε κλάσης και ταξινομεί το αντικείμενο σε αυτήν με την μεγαλύτερη πιθανότητα. Το άθροισμα των πιθανοτήτων όλων των κλάσεων ισούται με 1. Σημειώνεται ότι οι κατηγορίες πρέπει να είναι αμοιβαία αποκλειόμενες.

Naive Bayes

Ο αλγόριθμος Naive Bayes στηρίζεται στο θεώρημα Naïve Bayes και χρησιμοποιείται σε περιπτώσεις όπου ο αριθμός των δεδομένων υπερβαίνει κατά πολύ τον αριθμό των κλάσεων. Το θεώρημα του Bayes στηρίζεται στις πιθανότητες υπό συνθήκη και υποθέτει ότι τα ενδεχόμενα, δηλαδή οι κατηγορίες, είναι αμοιβαία αποκλειόμενα. Η πιθανότητα υπό συνθήκη είναι η πιθανότητα να συμβεί ένα ενδεχόμενο με δεδομένο ότι έχει συμβεί προηγουμένως ένα άλλο. Για παράδειγμα αν H_1 , H_2 και H_3 είναι αμοιβαία αποκλειόμενες κατηγορίες, όπως τρείς ξεχωριστές μηχανές παραγωγής, η πιθανότητα ένα ελαττωματικό προϊόν να έχει παραχθεί από την πρώτη μηχανή είναι πιθανότητα υπό συνθήκη και ορίζεται ως $P(H_1/E)$. Η πιθανότητα αυτή μπορεί να υπολογιστεί αν είναι γνωστές οι πιθανότητες της κάθε κατηγορίας $P(H_i)$ όπως και οι υπό συνθήκη πιθανότητες $P(E/H_i)$, οι οποίες στο παράδειγμα που αναφέρθηκε είναι οι πιθανότητες ένα προϊόν να προέρχεται από την κάθε μηχανή και οι πιθανότητες η κάθε μηχανή να παράξει ελαττωματικό προϊόν, αντίστοιχα. Τότε ισχύει:

$$P(H_i/E) = \frac{P(H_i)P(E/H_i)}{\sum_{i=1}^{n} P(H_i)P(E/H_i)}$$

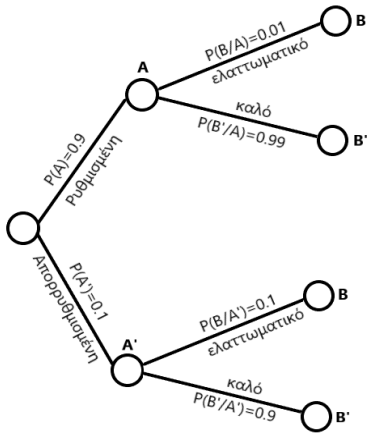
όπου n είναι ο αριθμός των κλάσεων.

Αν τα γεγονότα H_i χαρακτηριστούν ως αιτίες, το θεώρημα του Bayes μπορεί να θεωρηθεί ως η σχέση που δίνει την πιθανότητα το γεγονός E, που έχει ήδη συμβεί, να συνέβη εξαιτίας της H_i. Ο αλγόριθμος Naive Bayes χρησιμοποιεί αυτόν τον τρόπο για να υπολογίσει τις πιθανότητες να ανήκει κάποιο αντικείμενο σε κάθε κατηγορία. Έχει το πλεονέκτημα της ταχύτητας και της απλότητας αλλά το μειονέκτημα ότι υποθέτει την ανεξαρτησία των κλάσεων και δεν μαθαίνει τις σχέσεις μεταξύ τους.

Δέντρο απόφασης

Ένα δέντρο απόφασης είναι ένα διάγραμμα στο οποίο παριστάνονται τα δυνατά αποτελέσματα μίας διαδικασίας. Αποτελείται από κλάδους και κάθε κλάδος βρίσκεται ανάμεσα σε δύο κόμβους, εκείνου της αρχής και του πέρατός του. Οποιαδήποτε διαδρομή στο δέντρο, που αρχίζει από τον κόμβο της αρχής και καταλήγει σε ένα κόμβο πέρατος ονομάζεται απλή διαδρομή. Ακολουθώντας μία τέτοια διαδρομή και παίρνοντας σε κάθε κόμβο μία απόφαση μέχρι τον τελευταίο, πραγματοποιείται η τελική ανάθεση στην αντίστοιχη κατηγορία.

Τα δέντρα απόφασης εφαρμόζονται στις περιπτώσεις δύο κλάσεων, πολλαπλών κλάσεων και ανισορροπίας. Έχουν το πλεονέκτημα της εύκολης προετοιμασίας των δεδομένων αλλά και το μειονέκτημα της πολυπλοκότητας αλλά και της αστάθειας αφού μία μικρή αλλαγή σε έναν ενδιάμεσο κόμβο μπορεί να οδηγήσει σε ένα εντελώς διαφορετικό διάγραμμα.



Εικόνα 42: Δέντρο απόφασης (Decision Tree)

Μηχανή Διανύσματος Υποστήριξης (Support Vector Machine)

Μία μηχανή διανύσματος υποστήριξης χρησιμοποιεί ένα υπερεπίπεδο για να διαχωρίσει τα δεδομένα και να τα κατηγοριοποιήσει. Στην περίπτωση που τα δεδομένα έχουν ένα μόνο χαρακτηριστικό, δηλαδή μία διάσταση το υπερεπίπεδο εκφυλίζεται σε σημείο. Ομοίως, όταν έχουν δύο ή τρία χαρακτηριστικά, μετατρέπεται σε γραμμή και επίπεδο αντίστοιχα. Η πλευρά του υπερεπιπέδου στην οποία εμφανίζεται κάθε νέο σημείο καθορίζει και την κατηγορία στην οποία ταξινομείται. Δεδομένα που βρίσκονται σε διαφορετικές πλευρές του υπερεπιπέδου κατηγοριοποιούνται σε διαφορετικές κλάσεις. Η επιλογή του υπερεπιπέδου γίνεται με τέτοιο τρόπο ώστε να μεγιστοποιείται η απόσταση του από όλες τις κλάσεις. Στα πλεονεκτήματα της μεθόδου συγκαταλέγονται η ευστάθεια και η ακρίβεια όταν υπάρχουν μεγάλες βάσεις δεδομένων.

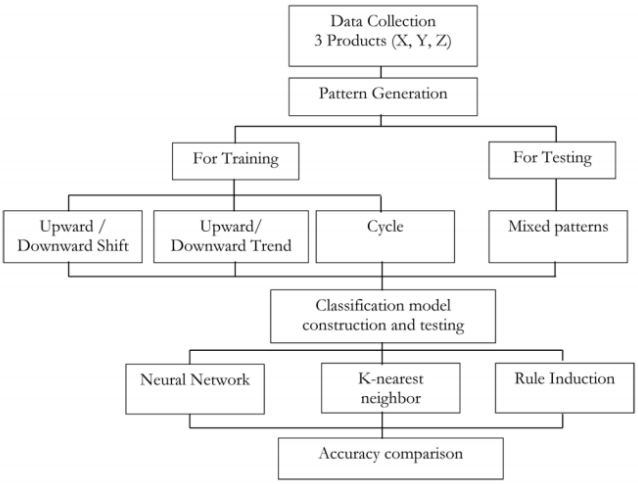
3.2 Αναγνώριση τάσεων σε διαγράμματα ποιότητας με ΤΝΔ στην βιβλιογραφία

Οι έρευνες που αφορούν την αναγνώριση τάσεων σε διαγράμματα ποιότητας με ΤΝΔ στην βιβλιογραφία είναι πολυάριθμες. Παρακάτω παρουσιάζονται παρακάτω ορισμένες ενδεικτικές περιπτώσεις εξ αυτών, που έχουν κοινά χαρακτηριστικά με την παρούσα εργασία.

Μία χαρακτηριστική περίπτωση είναι η εργασία των Tunchanit Bunjongjit και Wimalin Laosiritaworn με θέμα την κατηγοριοποίηση τάσεων που εμφανίζονται σε διαγράμματα ελέγχου μεταλλικών περιβλημάτων επενεργητών [22]. Σε αυτήν εφαρμόζονται τρείς μέθοδοι κατηγοριοποίησης: ΤΝΔ, knn και δέντρα απόφασης. Τα δεδομένα στα οποία στηρίζονται είναι πραγματικά από την γραμμή παραγωγής αλλά και κατασκευασμένα και αφορούν τρείς τάσεις: την Μονότονη αύξουσα και φθίνουσα τάση, την Περιοδική και τα Άλματα αλλαγής επιπέδου.

Για την κατασκευή των δεδομένων εκπαίδευσης, ακολουθείται η μέθοδος της μεταβολής ενός διανύσματος, που έχει προκύψει από δείγμα της παραγωγικής διαδικασίας σε κατάσταση στατιστικού ελέγχου, κατά μία ποσότητα που σχετίζεται με την τυπική της απόκλιση (S). Συγκεκριμένα προστίθεται, για το άλμα αλλαγής επιπέδου, σε ένα τμήμα του διανύσματος, μία ποσότητα που ανήκει στο [-0.2*S, 2.5*S], για την μονότονη τάση, σε ολόκληρο το διάνυσμα, μία ποσότητα που ανήκει στο [-0.22*S, 0.22*S] πολλαπλασιαζόμενη με τον χρόνο και για την κυκλική τάση, η ποσότητα $asin\left(\frac{2\pi t}{\Omega}\right)$, με το α να ανήκει στο [S, 2.5*S], όπου Ω η περίοδος του κύκλου. Το κάθε διάνυσμα έχει 36 στοιχεία.

Ο έλεγχος απόδοσης πραγματοποιείται με δεδομένα μεικτών τάσεων, στα οποία συνδυάζονται δύο ή περισσότερες τάσεις και διαθέτουν αριθμό στοιχείων, που λαμβάνει τις τιμές 6, 9 και 12.



Εικόνα 43: Η μεθοδολογία που ακολουθήθηκε στην εργασία των Tunchanit Bunjongjit και Wimalin Laosiritaworn [22].

Αναφορικά με το ΤΝΔ, χρησιμοποιείται η δυαδική σιγμοειδής συνάρτηση ενεργοποίησης με συντελεστή κλίσης ίσο με 1, για τα κρυμμένα επίπεδα και το επίπεδο εξόδου. Δοκιμάζονται 21 αρχιτεκτονικές με 3, 4 και 5 κρυμμένα επίπεδα των οποίων ο αριθμός κόμβων κυμαίνεται μεταξύ 6 και 20.

Στο μοντέλο knn επιλέγεται η τιμή του k να είναι ίση με την μονάδα.

Η ακρίβεια την οποία επιτυγχάνουν, με τα ΤΝΔ αγγίζει το 91.73%, με το μοντέλο knn το 97.44% και με τα δέντρα απόφασης, το 85.59%. Επίσης, καταλήγουν στο συμπέρασμα ότι η ακρίβεια επηρεάζεται σημαντικά από τον αριθμό των στοιχείων των διανυσμάτων ελέγχου.

Με την αναγνώριση των τάσεων κυκλική, μονότονη (αύξουσα και φθίνουσα) και άλμα αλλαγής επιπέδου (ανοδικό και καθοδικό) σε διαγράμματα ελέγχου παραγωγικής διαδικασίας ασχολούνται και οι Pham και Wani [23]. Για τον σκοπό αυτό, υλοποιούν ΤΝΔ οπισθόδρομης διάδοσης σφάλματος και δέντρα απόφασης. Η αρχιτεκτονική των ΤΝΔ αποτελείται από ένα κρυμμένο επίπεδο με 13 νευρώνες. Τα επίπεδα εισόδου και εξόδου απαρτίζονται από 9 και 6 νευρώνες, που καθορίζονται από το μέγεθος των διανυσμάτων εισόδου και τον αριθμό των κλάσεων της εφαρμογής, προσθέτοντας σε αυτές την απουσία τάσης, αντίστοιχα. Ο αριθμός των νευρώνων του κρυμμένου επιπέδου επιλέγεται εμπειρικά. Ως συνάρτηση ενεργοποίησης χρησιμοποιείται η δυαδική σιγμοειδής.

Η εκπαίδευση πραγματοποιείται με 83 διανύσματα για κάθε κλάση και ο έλεγχος αποτελεσμάτων με 167. Όλα τα διανύσματα εκπαίδευσης και ελέγχου είναι κατασκευασμένα και δημιουργούνται με βάση την μέση τιμή και την τυπική απόκλιση της παραγωγικής διαδικασίας, σε κατάσταση στατιστικού ελέγχου. Το κάθε στοιχείο των διανυσμάτων εξόδου λαμβάνει τιμές από 0 έως 1 και το στοιχείο με την μεγαλύτερη τιμή καθορίζει την κλάση στην οποία καταχωρείται το διάνυσμα.

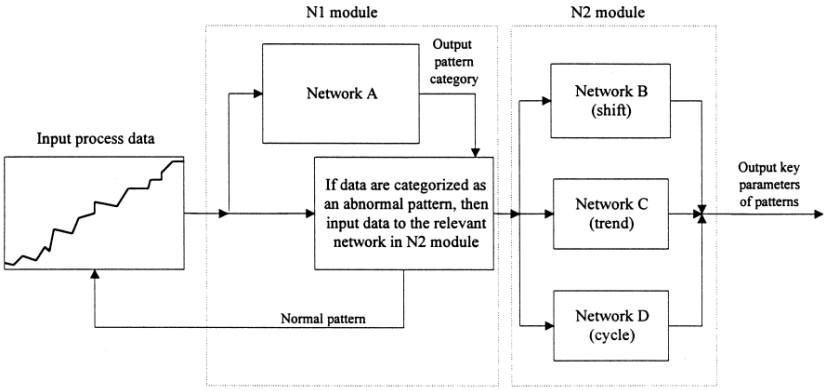
Η μέγιστη ακρίβεια των αποτελεσμάτων τους είναι 99% και επιτυγχάνεται με τα ΤΝΔ. Καταλήγουν στο συμπέρασμα ότι τα ΤΝΔ είναι λειτουργικότερα από τα δέντρα απόφασης, καθώς έχουν μεγαλύτερη απόδοση και δεν απαιτούν τον καθορισμό κανόνων απόφασης, παρά το γεγονός ότι καθιστούν την διαδικασία κατηγοριοποίησης μη ορατή στον χρήστη.

Ομοίως, οι Guh και Hsieh [24] πραγματεύονται την αναγνώριση των τάσεων, κυκλική, μονότονη (αύξουσα και φθίνουσα) και άλμα αλλαγής επιπέδου με την χρήση ΤΝΔ αλλά μέσω αυτών προχωρούν και στην εκτίμηση ορισμένων παραμέτρων της κάθε μίας όπως, πλάτος και μήκος κύκλου, κλίση και μέγεθος άλματος, αντίστοιχα. Όπως αναφέρουν, η εκπαίδευση ξεχωριστών ΤΝΔ για κάθε λειτουργία υποβοηθά την εκπαίδευση. Όλα τα διανύσματα εκπαίδευσης και ελέγχου είναι κατασκευασμένα και δομούνται με βάση την μέση τιμή και την τυπική απόκλιση της παραγωγικής διαδικασίας, σε κατάσταση στατιστικού ελέγχου. Επίσης, επιχειρείται την αναγνώριση μεικτών τάσεων.

Για τον σκοπό αυτό, υλοποιούν ΤΝΔ οπισθόδρομης διάδοσης σφάλματος με ένα κρυμμένο επίπεδο με 35 νευρώνες, ενώ τα επίπεδα εισόδου και εξόδου απαρτίζονται από 56 και 4 νευρώνες, όσο το μέγεθος των διανυσμάτων εισόδου και ο αριθμός των κλάσεων. Αναφέρουν ότι, σύμφωνα με τις παρατηρήσεις τους, όταν οι νευρώνες του

κρυμμένου επιπέδου ξεπερνούν τους 35 δεν βελτιώνεται η εκπαίδευση, αλλά γίνεται πιο χρονοβόρα. Χρησιμοποιούν ως συνάρτηση ενεργοποίησης, του κρυμμένου επιπέδου την δυαδική σιγμοειδή και του επιπέδου εξόδου την συνάρτηση υπερβολικής εφαπτομένης. Για την εκπαίδευση του ΤΝΔ Α (βλ. Εικόνα 44) χρησιμοποιούνται 3360 διανύσματα συνολικά, ενώ για την εκπαίδευση των B,C και D χρησιμοποιούνται 880, 800 και 2400 διανύσματα, αντίστοιχα.

Τα διανύσματα ελέγχου είναι 8000 συνολικά και από αυτά το 94% αναγνωρίζονται σωστά από το ΤΝΔ Α. Επίσης, οι παράμετροι των τάσεων εκτιμώνται με ικανοποιητική ακρίβεια, όπως σχολιάζεται. Τέλος, πραγματεύονται μόνο μία κατηγορία μεικτής τάσης, τον συνδυασμό μονότονης και κυκλικής. Τα διανύσματα που της αντιστοιχούν είναι 5000 και χρησιμοποιούνται μόνο για τον έλεγχο και όχι για την εκπαίδευση, με ακριβή αποτελέσματα όσον αφορά την αναγνώρισή τους.

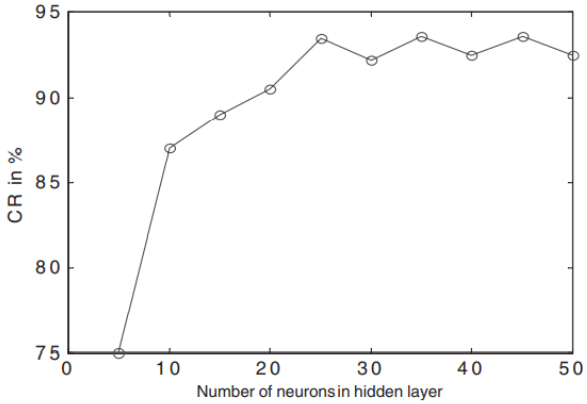


Εικόνα 44: Δομή του δικτύου των Guh και Hsieh [24]. Στο επίπεδο N1 ελέγχεται με το ΤΝΔ Α η είσοδος και αν αναγνωριστεί ως μία από τις εξεταζόμενες τάσεις, περνά στο επίπεδο Β εκχωρούμενη στο ανάλογο ΤΝΔ που δίνει την εκτίμηση των παραμέτρων της.

Ακολούθως, οι Pham και Sahran [25] υλοποιούν ακιδωτά ΤΝΔ (Spiking Neural Networks) για την αναγνώριση των ιδίων τάσεων (κυκλική, μονότονη αύξουσα και φθίνουσα και άλμα αλλαγής επιπέδου ανοδικό και καθοδικό) σε διαγράμματα ελέγχου παραγωγικής διαδικασίας. Το είδος αυτό των ΤΝΔ αποτελείται από νευρώνες, με βηματική συνάρτηση ενεργοποίησης, μιμούμενο τα βιολογικά νευρωνικά δίκτυα. Η αρχιτεκτονική τους αποτελείται από ένα κρυμμένο επίπεδο με 6 νευρώνες, ενώ τα επίπεδα εισόδου και εξόδου απαρτίζονται από 60 και 6 νευρώνες, που καθορίζονται από το μέγεθος των διανυσμάτων εισόδου και τον αριθμό των κλάσεων της εφαρμογής, προσθέτοντας σε αυτές την απουσία τάσης, αντίστοιχα. Η εκπαίδευση πραγματοποιείται με κατασκευασμένα 498 διανύσματα, δηλαδή με 83 για κάθε κατηγορία ενώ ο έλεγχος απόδοσης με 540, δηλαδή με 90 κατασκευασμένα διανύσματα για κάθε κατηγορία. Η ακρίβεια που επιτυγχάνεται είναι 97.85%. Επιπλέον, με σκοπό την σύγκριση, αναπτύσσουν ένα ΤΝΔ οπισθόδρομης διάδοσης σφάλματος καθώς και ένα δίκτυο LVQ (Learning Vector Quantization), το οποίο αποτελεί μία ειδική κατηγορία ΤΝΔ που

εφαρμόζει τον κανόνα μάθησης του Hebb (βλ. § 2.2.4.2 Εκπαίδευση). Η ακρίβεια που επιτυγχάνεται με τα τελευταία είναι 95% και 97.7% αντίστοιχα.

Ο Yousef Al-Assaf [26] κατασκευάζει δεδομένα με την εισαγωγή τυχαιότητας. Στοχεύει στην αναγνώριση των ιδίων τάσεων αλλά και 7 μεικτών. Πραγματοποιεί κυματική ανάλυση η οποία εξάγει τα χαρακτηριστικά των τάσεων, τροφοδοτώντας με αυτά το ΤΝΔ κατά την εκπαίδευσή του, προκειμένου να την διευκολύνει. Χρησιμοποιεί 5250 διανύσματα εκπαίδευσης και ισάριθμα διανύσματα ελέγχου, αποτελούμενα από 5 κόμβους, σε ένα νευρωνικό δίκτυο με ένα κρυμμένο επίπεδο. Όπως αναφέρει, ένα κρυμμένο επίπεδο είναι αρκετό για να επιτευχθεί ικανοποιητική ακρίβεια, με την προϋπόθεση, ο αριθμός κόμβων του να είναι ανάλογος του μεγέθους του διανύσματος εισόδου. Για τα διανύσματα ελέγχου απλών τάσεων, δοκιμάζοντας διαφορετικούς αριθμούς κόμβων συμπεραίνει ότι, αρχικά η ακρίβεια αυξάνεται με την αύξησή τους, όμως όταν ξεπεράσουν τους 25 η βελτίωση της ακρίβειας σταματάει καθώς αυξάνεται η πολυπλοκότητα (βλ. Εικόνα 45). Σχολιάζει ότι, η αυτή συμπεριφορά συμφωνεί με την βιβλιογραφία, στην οποία προτείνεται οι νευρώνες του κρυμμένου επιπέδου να είναι 3 έως 4 φορές περισσότεροι από αυτούς του διανύσματος εισόδου.



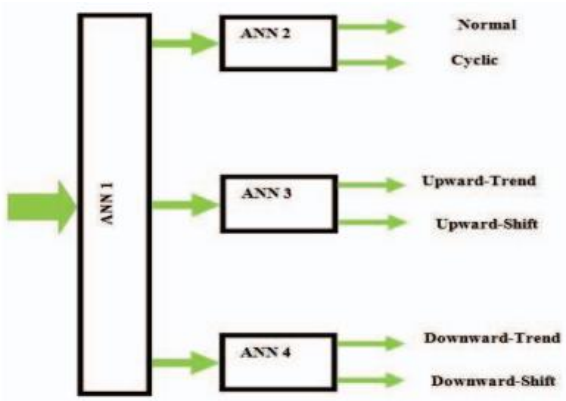
Εικόνα 45: Μεταβολή της ακρίβειας με την αύξηση του αριθμού των κόμβων του κρυμμένου επιπέδου.

Η ακρίβεια που επιτυγχάνει με τα διανύσματα ελέγχου μεικτών τάσεων δεν ξεπερνά το 15% και καταλήγει στο συμπέρασμα ότι η εκπαίδευση με διανύσματα απλών τάσεων, που εφαρμόζει, δεν επαρκεί για την αναγνώριση των μεικτών. Αντιθέτως είναι απαραίτητη η εκπαίδευση με διανύσματα μεικτών τάσεων για αυτόν τον σκοπό.

Οι Vahid Ranaee και Ata Ebrahimzadeh [28] πειραματίστηκαν με διαφορετικούς αλγορίθμους εκπαίδευσης ΤΝΔ για την αναγνώριση 6 τάσεων σε διαγράμματα ελέγχου ποιότητας. Ανάμεσά τους οι LM και SCG. Εκπαίδευσαν ΤΝΔ με δύο βάσεις δεδομένων. Αμφότερες περιλαμβάνουν τα διανύσματα εκπαίδευσης όμως η δεύτερη περιλαμβάνει επιπλέον τα χαρακτηριστικά των τάσεων, αφού χρησιμοποιήθηκε ένα σύστημα εξαγωγής τους. Τα διανύσματα εκπαίδευσης και ελέγχου τους είναι κατασκευασμένα και βασίζονται στην μέση τιμή και την τυπική απόκλιση της διαδικασίας σε κατάσταση στατιστικού ελέγχου. Επίσης εισάγουν τυχαιότητα σε αυτά. Οι αρχιτεκτονικές που δοκιμάζουν ποικίλουν. Καταλήγουν σε τέσσερις, οι δύο με ένα κρυμμένο επίπεδο με 15

και 25 νευρώνες αντίστοιχα και οι άλλες δύο με δύο κρυμμένα επίπεδα, η μία με 15 και 20 και η άλλη με 20 και 25 νευρώνες. Σχολιάζουν ότι όσον αφορά τις τελευταίες δύο αρχιτεκτονικές, είναι αυτές που επέφεραν τα καλύτερα αποτελέσματα από όσες χρησιμοποίησαν, με δύο κρυμμένα επίπεδα. Τα διανύσματα εισόδου έχουν 12 κόμβους. Ως συνάρτηση ενεργοποίησης του κρυμμένου επιπέδου όρισαν την συνάρτηση υπερβολικής εφαπτομένης, ενώ του επιπέδου εξόδου την ταυτοτική. Η εκπαίδευση πραγματοποιείται με 4500 διανύσματα ενώ ο έλεγχος απόδοσης με 1500. Τα αποτελέσματα που επιτυγχάνουν έχουν μέση ακρίβεια 92.14% για τις εκπαιδεύσεις χωρίς τα χαρακτηριστικά των τάσεων και 99.15% με αυτά για τον αλγόριθμο LM. Τόσο για τον τελευταίο, όσο και για τον αλγόριθμο SCG η ακρίβεια είναι μεγαλύτερη του 99% για όλες τις αρχιτεκτονικές όταν συμπεριλαμβάνονται τα χαρακτηριστικά των τάσεων στην εκπαίδευση.

Τέλος, μία διαφορετική προσέγγιση για την αναγνώριση μεικτών τάσεων πραγματοποιείται από τους Jalil Addeh, Ata Ebrahimzadeh και Vahid Ranaee [29]. Για την αναγνώριση 6 τάσεων, απλών και μεικτών, χρησιμοποιούν 4 ΤΝΔ τα οποία χωρίζονται σε 2 επίπεδα. Στο πρώτο επίπεδο το ΤΝΔ 1 διαχωρίζει τις τάσεις σε ζευγάρια πιθανών τάσεων. Στο δεύτερο επίπεδο όποιο από τα ΤΝΔ 2, 3 ή 4 τροφοδοτηθεί, αναγνωρίζει την κατηγορία της τάσης. Όλα τα ΤΝΔ αποτελούνται από ένα κρυμμένο επίπεδο με 23 νευρώνες, αριθμός στον οποίο κατέληξαν μετά από δοκιμές. Η βάση δεδομένων τους αποτελείται από 600 διανύσματα με το 60% να χρησιμοποιείται για την εκπαίδευση και το 40% για τον έλεγχο. Η συνάρτηση ενεργοποίησης του κρυμμένου επιπέδου είναι η συνάρτηση υπερβολικής εφαπτομένης, ενώ του επιπέδου εξόδου η ταυτοτική. Η μέση ακρίβεια των 50 δοκιμών που διεξάγει είναι 99.21%.



Εικόνα 46: Το δίκτυο που μορφοποιούν οι [29] με επιμέρους ΤΝΔ για την αναγνώριση μεικτών τάσεων.

4 Ανάπτυξη Δεδομένων

4.1 Διανύσματα εκπαίδευσης και ελέγχου

Τα δεδομένα της παρούσας εργασίας δεν κατασκευάζονται με αυστηρά φορμαλιστικό τρόπο όπως στη βιβλιογραφία ([22], [23], [24], [28]), στην οποία η κατασκευή τους προκύπτει από σχέσεις που έχουν ως βάση την μέση τιμή και την τυπική απόκλιση της διαδικασίας σε κατάσταση στατιστικού ελέγχου. Αυτό συγκεκριμενοποιεί πολύ το εύρος τιμών τους. Αντιθέτως στην εργασία αυτή, τα διανύσματα αναπτύσσονται με πιο ελεύθερο, ως προς το εύρος τιμών τους, τρόπο, εστιάζοντας κυρίως στην μορφή τους. Επίσης προστίθεται σε αυτά τυχαιότητα ([26], [28]) που σκοπό έχει τόσο να τα πολλαπλασιάσει όσο και να τα διαφοροποιήσει.

Αρχικά, οι τιμές τους αντλούνται από τα διαγράμματα ελέγχου του [4] που χρησιμοποιείται ως πηγή. Μέσω αυτού, μορφοποιούνται οι συνηθέστερες μορφές μη τυχαίων χρονοσειρών, που παραπέμπουν σε επίδραση συστηματικής αιτίας, όπως παρουσιάζονται στην παράγραφο 2.1.3.

Συνοπτικά περιλαμβάνονται οι χρονοσειρές με:

1. Περιοδικότητα (cycle pattern)
2. Σημείο πολύ κοντά στα όρια ελέγχου (exceptional point)
3. Διαδοχικά σημεία από τη μία μόνο πλευρά της κεντρικής γραμμής/Ομαδοποίηση (grouping)
4. Μεγάλη διακύμανση (instability)
5. Άλματα αλλαγής επιπέδου (level shift)
6. Μονότονη αύξουσα ή φθίνουσα τάση (monotonous trend)
7. Σημείο εκτός των ορίων ελέγχου (point outside of control limits)
8. Έλλειψη μεταβλητότητας (stratification)

Οι παραπάνω 8 κατηγορίες αναφέρονται και ως τάσεις και η σειρά με την οποία παρουσιάζονται είναι αλφαβητική ως προς την αγγλική τους ονομασία.

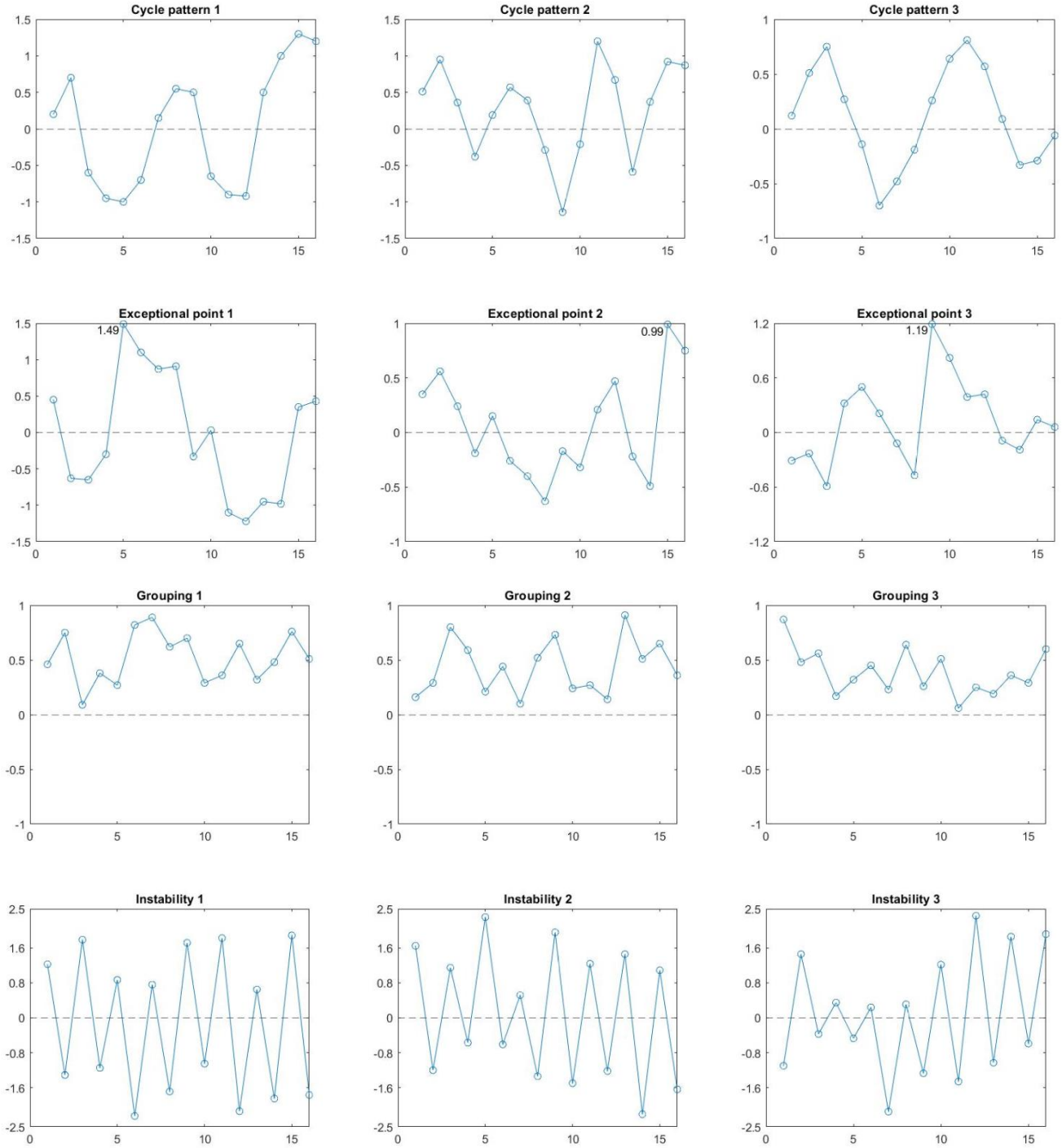
4.1.1 Μητρικά διανύσματα

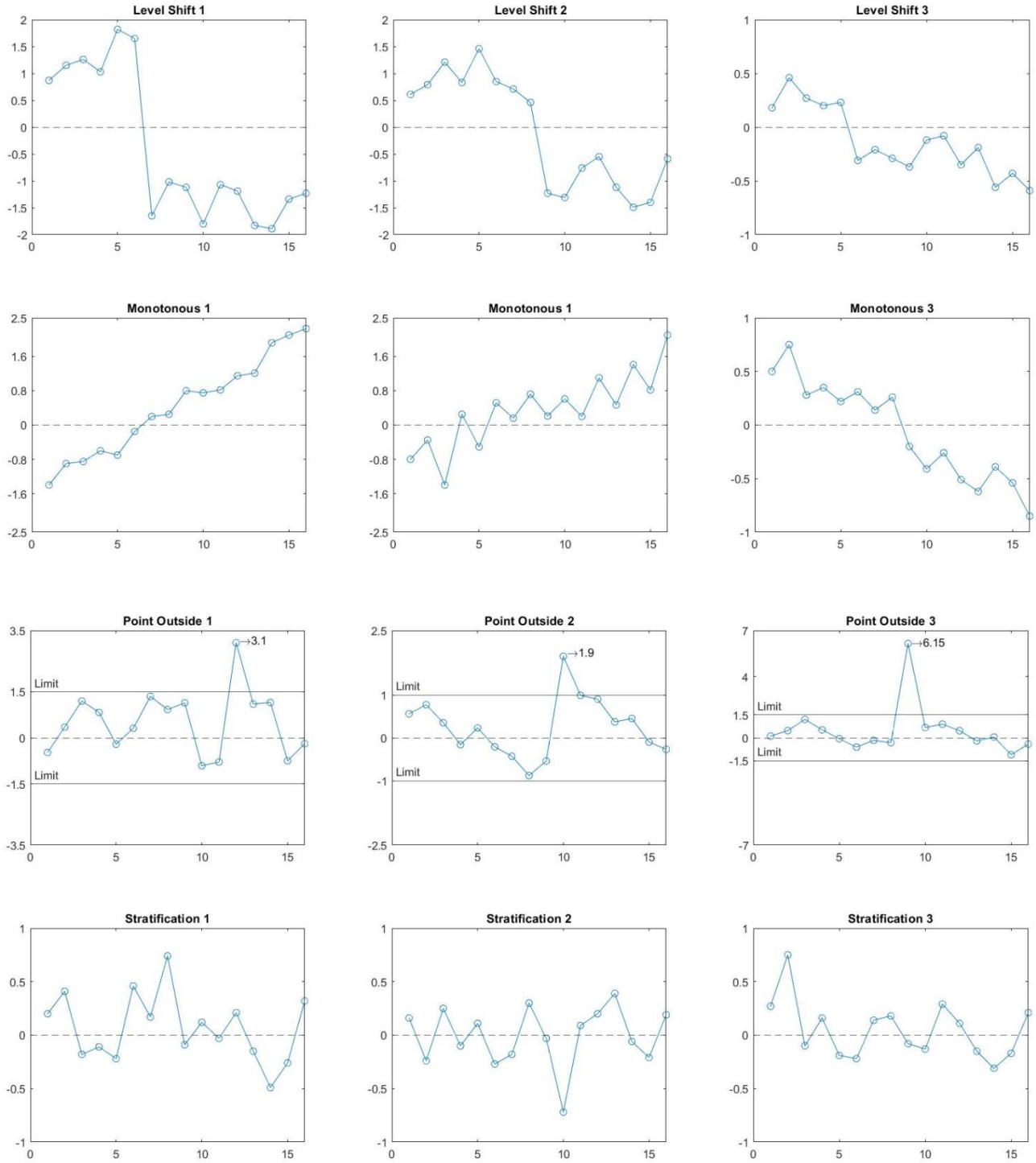
Τα διαγράμματα ελέγχου του [4] από όπου προέρχονται οι χρονοσειρές, απαρτίζονται από 16 σημεία το κάθε ένα. Οι τιμές αυτών των σημείων εξάγονται από κάθε διάγραμμα, διαμορφώνοντας ένα αρχικό διάνυσμα με 16 στοιχεία για κάθε κατηγορία. Για παράδειγμα το αρχικό διάνυσμα της κυκλικής τάσης είναι:

[0.2 0.7 -0.6 -0.95 -1 -0.7 0.15 0.55 0.5 -0.65 -0.9 -0.92 0.5 1 1.3 1.2]

Το διάνυσμα αυτό, με βάση τους κανόνες ορισμού κάθε τάσης και διατηρώντας την τάξη μεγέθους του παραλλάσσεται δύο φορές. Έτσι δημιουργούνται δύο επιπλέον διανύσματα 16 στοιχείων για κάθε τάση, που μαζί με το αρχικό ονομάζονται στο εξής μήτρες της τάσης. Κατά συνέπεια σε πρώτο επίπεδο οι 8 κατηγορίες μη τυχαίων χρονοσειρών, διαθέτουν 3 μητρικά διανύσματα 16 στοιχείων το κάθε ένα, που

παρουσιάζονται παρακάτω. Αναλυτικά οι τιμές τους, όπως και η τυπική απόκλιση τους φαίνονται στον Πίνακα 24 του Παραρτήματος Δ.





Εικόνα 47: Οι τρείς μήτρες της κάθε τάσης, αποτελούμενες από 16 στοιχεία.

Από την εξέταση των παραπάνω διαγραμμάτων εξάγεται μία σημαντική διαπίστωση. Ορισμένα διανύσματα διαφορετικών κατηγοριών, έχουν δομηθεί με τέτοιον τρόπο, ώστε η μορφή τους ενδέχεται να είναι παρεμφερής [27]. Για παράδειγμα:

- Οι μήτρες των σημείων πολύ κοντά στα όρια ελέγχου έχουν παρόμοια μορφή με τις μήτρες των σημείων εκτός ορίων ελέγχου.

- Οι μήτρες της έλλειψη μεταβλητότητας έχουν ομοιότητες με τις μήτρες της κυκλικής τάσης.
- Η μονότονες τάσεις και τα άλματα έχουν αμφότερες δύο ομάδες σημείων, μία στην κάθε πλευρά της κεντρικής γραμμής.

Όπως είναι λογικό είναι απαραίτητη η προσθήκη κάποιου παράγοντα στα διανύσματα, που συγκεκριμενοποιεί την κατηγορία στην οποία ανήκουν, διαχωρίζοντάς την με σαφήνεια από τις υπόλοιπες. Αυτός ο παράγοντας είναι τα όρια ελέγχου, τα οποία προστίθενται ως δύο επιπλέον στοιχεία στην αρχή κάθε διανύσματος με σειρά min max και μετατρέπουν τον συνολικό αριθμό στοιχείων του από 16 σε 18. Τα όρια ελέγχου έχουν επιλεγεί με βάση τις εξισώσεις (1.1) έως (1.3) και φαίνονται στα παραπάνω διαγράμματα ως οι δύο ακραίες τιμές του κατακόρυφου άξονα, με εξαίρεση τις περιπτώσεις των σημείων εκτός ορίων ελέγχου στις οποίες ορίζονται από τα σημεία τομής του κατακόρυφου άξονα με τις οριζόντιες γραμμές «Limit». Για παράδειγμα στο διάγραμμα Instability 1 τα όρια ελέγχου είναι -2.5, 2.5 ενώ στο διάγραμμα Point Outside 1 είναι -1.5, 1.5.

Υπογραμμίζεται ότι για όλα τα μητρικά διανύσματα ισχύει: $K\Gamma = 0$.

4.1.2 Παραγωγή διανυσμάτων.

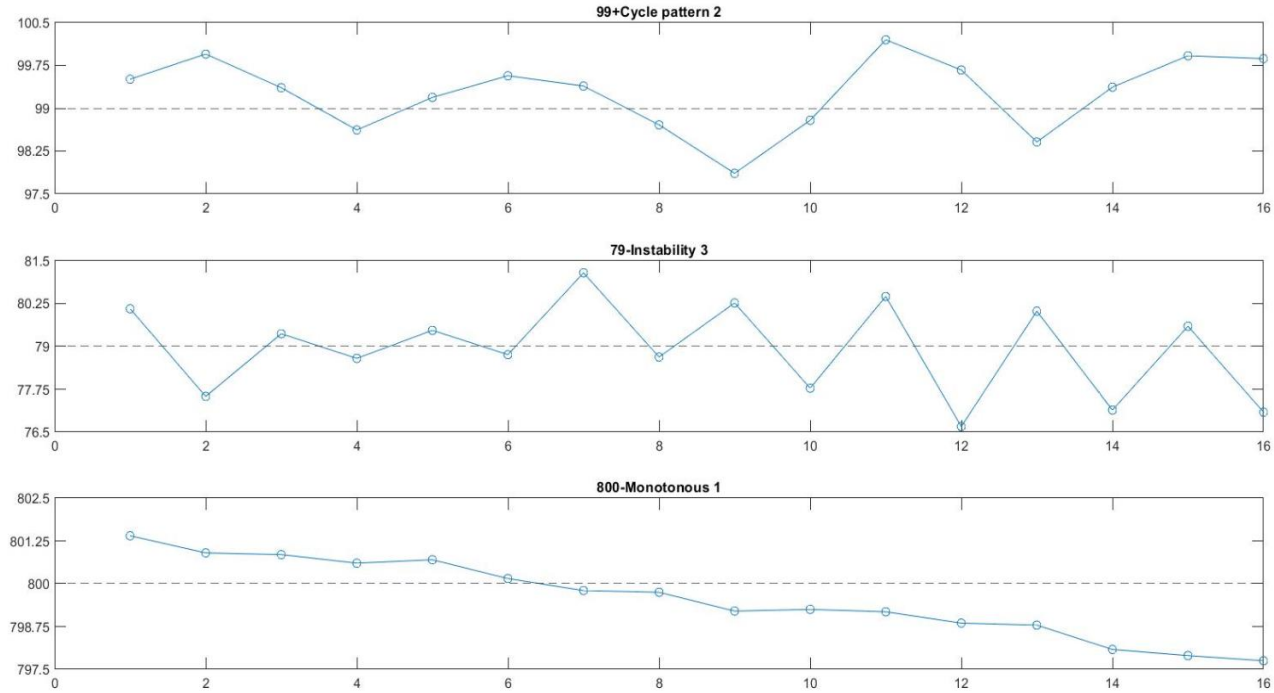
Σε δεύτερο επίπεδο τα διανύσματα κάθε τάσης αυξάνονται στα 18, που αντιστοιχούν στα 3 μητρικά και σε 15 παράγωγά τους. Τα 15 επιπλέον διανύσματα, δημιουργούνται από τα 3 μητρικά με την εφαρμογή της ακόλουθης μεθοδολογίας:

1. Οι μήτρες πολλαπλασιάζονται επί -1 (+3 διανύσματα)
2. Οι μήτρες πολλαπλασιάζονται επί 2,3 και 5 (+3 διανύσματα)
3. Προστίθεται ένας αριθμός της τάξης μεγέθους του 1 στην κάθε μήτρα. Για παράδειγμα 5+Cycle pattern 1. Έτσι μεταφέρεται η μέση τιμή της (+3 διανύσματα)
4. Προστίθεται ένας αριθμός της τάξης μεγέθους του 10 στην κάθε μήτρα. Για παράδειγμα 83+Stratification 2 (+3 διανύσματα)
5. Προστίθεται ένας αριθμός της τάξης μεγέθους του 100 στην κάθε μήτρα. Για παράδειγμα 235+Instability 3 (+3 διανύσματα)

Τόσο οι παράγοντες των γινομένων, όσο και οι σταθερές των αθροισμάτων επιλέγονται με τέτοιο τρόπο ώστε να διευρύνουν και να μετακινήσουν τις τάσεις μεγεθύνοντας το πεδίο τιμών τους, πέρα από το [-1,1].

Οι τελευταίοι τρεις κανόνες ενδέχεται να συνδυάζονται με τους δύο πρώτους, δημιουργώντας έτσι παράγωγα διανύσματα ποικίλων γραμμικών συνδυασμών της κάθε μήτρας όπως για παράδειγμα το 63-Monotonous 3.

Με αυτόν τον τρόπο, εκχωρώντας 18 διανύσματα σε κάθε τάση, προκύπτουν συνολικά 144 διανύσματα με 18 στοιχεία το κάθε ένα.



Εικόνα 48: Παράγωγα διανύσματα των μητρικών διανυσμάτων Cycle pattern 2, Instability 3 και Monotonous 1.

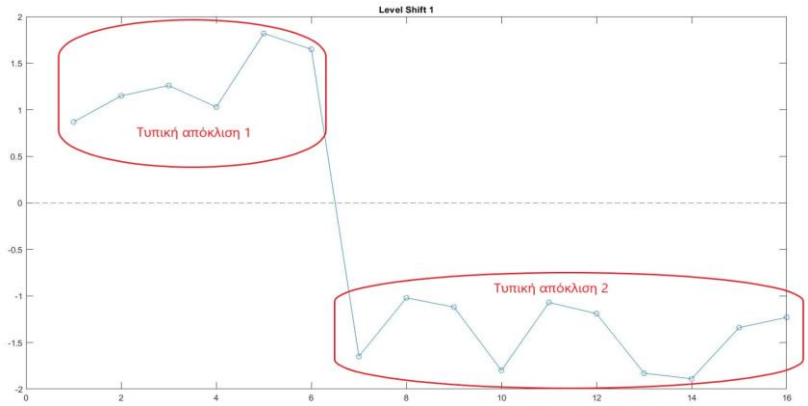
Προχωρώντας στο τρίτο στάδιο της ανάπτυξης των δεδομένων, υπολογίζεται η τυπική απόκλιση κάθε διανύσματος. Συνεπώς αντιστοιχίζονται συνολικά 144 τυπικές αποκλίσεις. Η τυπική απόκλιση για ένα δείγμα n στοιχείων δίνεται από τον τύπο:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

όπου $\bar{x}$ η μέση τιμή των στοιχείων.

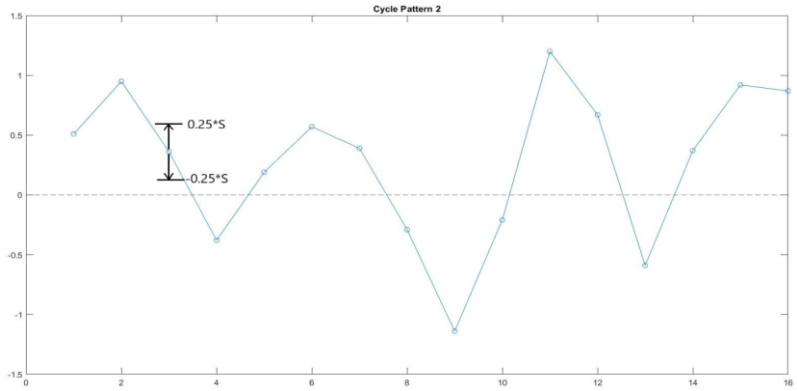
Σημειώνεται ότι κατά τον υπολογισμό της τυπικής απόκλισης κάθε διανύσματος, δεν συμμετέχουν τα όρια ελέγχου, που καταλαμβάνουν τις δύο πρώτες θέσεις.

Για τα διανύσματα της κατηγορίας level shift υπολογίζονται δύο τυπικές αποκλίσεις για κάθε διάνυσμα, μία για κάθε ομάδα σημείων εκατέρωθεν της ΚΓ, που διαχωρίζονται από το άλμα της τάσης και από αυτές κρατείται η μικρότερη.



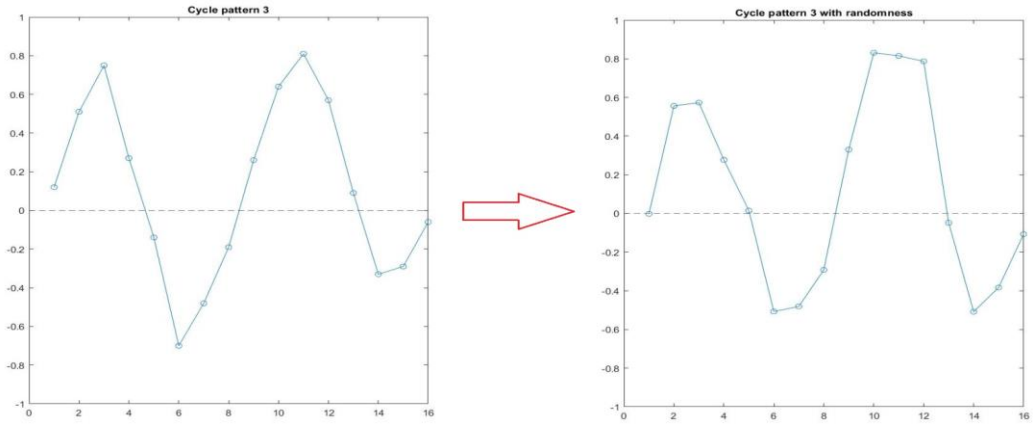
Εικόνα 49: Υπολογίζονται 2 τυπικές αποκλίσεις για κάθε διάνυσμα άλματος και κρατείται η μικρότερη.

Στην συνέχεια για κάθε ένα από τα 18 διανύσματα δημιουργείται ένα δεύτερο, θυγατρικό, προσθέτοντας σε κάθε στοιχείο του, με εξαίρεση τα όρια ελέγχου, έναν τυχαίο αριθμό που ανήκει στο $[-0.25 * \text{τυπική απόκλιση}\ , 0.25 * \text{τυπική απόκλιση}\ ]$ με την τυπική απόκλιση να είναι αυτή που υπολογίστηκε για το συγκεκριμένο διάνυσμα.



Εικόνα 50: Μικρή μετακίνηση των σημείων κατά τυχαίο τρόπο.

Με αυτόν τον τρόπο το κάθε σημείο μετακινείται με τυχαίο τρόπο σε μικρή απόσταση, ικανή να διαφοροποιήσει την χρονοσειρά, χωρίς όμως να αλλάξει την τάση στην οποία ανήκει.



Εικόνα 51: Δημιουργία χρονοσειράς από την μήτρα Cycle pattern 3, μέσω μεταβολής της που εμπεριέχει τυχαιότητα.

Έτσι, έχοντας προσθέσει τυχαιότητα στην μορφή των διανυσμάτων, αυτά διπλασιάζονται σε αριθμό και πλέον έχει δημιουργηθεί μια βάση δεδομένων που αποτελείται από 288 διανύσματα, δηλαδή 8 τάσεις, που ταξινομούνται με αλφαβητική σειρά ως προς την αγγλική τους ονομασία και εμπεριέχει 36 διανύσματα η κάθε μία.

4.2 Διανύσματα στόχοι

Η επιτηρούμενη εκπαίδευση, που χρησιμοποιείται στην παρούσα εργασία προϋποθέτει την ύπαρξη διανυσμάτων επιθυμητής εξόδου, που λειτουργούν ως στόχοι. Ο ρόλος τους είναι να υποδεικνύουν την τάση στην οποία ανήκει το κάθε διάνυσμα εισόδου. Ως εκ τούτου δημιουργούνται 288 διανύσματα στόχοι, ένα για κάθε διάνυσμα των δεδομένων.

Το κάθε διάνυσμα στόχος έχει την μορφή $[1\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$. Άρα διαθέτει 8 στοιχεία, δηλαδή 8 θέσεις, που αντιστοιχούν στις 8 τάσεις. Οι τάσεις έχουν ταξινομηθεί με αλφαβητική σειρά ως προς την αγγλική τους ονομασία, όπως παρουσιάζεται παραπάνω. Το κάθε στοιχείο παίρνει την τιμή 0 ή 1, αναλόγως με το αν η τάση στην οποία αντιστοιχεί είναι η σωστή ή όχι. Η θέση της μονάδας υποδεικνύει την εκάστοτε ορθή τάση. Για παράδειγμα το διάνυσμα στόχος για ένα διάνυσμα της μονότονης τάσης, που είναι 6η αλφαβητικά, είναι το $[0\ 0\ 0\ 0\ 0\ 1\ 0\ 0]$. Αναλυτικότερα τα διανύσματα στόχοι και οι τάσεις στις οποίες αντιστοιχούν παρουσιάζονται στον Πίνακα 25 του Παραρτήματος Δ.

Συμπερασματικά η βάση η βάση δεδομένων συμπληρώνεται με τα διανύσματα στόχους, επαναλαμβανόμενα 36 φορές για κάθε τάση. Δηλαδή αρχικά επαναλαμβάνεται 36 φορές το διάνυσμα $[1\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$ της κυκλικής τάσης, στην συνέχεια ισάριθμες φορές το διάνυσμα $[0\ 1\ 0\ 0\ 0\ 0\ 0\ 0]$ του σημείου πολύ κοντά στα όρια ελέγχου κ.ο.κ.

4.3 Διαχωρισμός των διανυσμάτων εκπαίδευσης και ελέγχου

Από τα 288 διανύσματα δεδομένων επιλέγεται το 75%, δηλαδή 216, να χρησιμοποιηθεί για την εκπαίδευση και το υπόλοιπο 25%, δηλαδή 72 για τον έλεγχο απόδοσης. Αυτό αναλύεται σε 27 διανύσματα για την εκπαίδευση και 9 για τον έλεγχο απόδοσης, από κάθε τάση. Γίνεται αντιληπτό ότι δεν υφίσταται τμήμα επικύρωσης των αποτελεσμάτων,

όπως αναφέρεται στην παράγραφο της επιτηρούμενης μάθησης και ο λόγος είναι η αποφυγή της αυτόματης πραγματοποίησης αλλαγών στο ΤΝΔ. Αντιθέτως προτιμάται η στενότερη παρακολούθηση και παρέμβαση στην εκάστοτε δομή του δικτύου και στα αποτελέσματα που επιτυγχάνει.

Αν ο ποσοστιαίος διαχωρισμός 75%-25% εφαρμοστεί συνολικά στο δεδομένα τότε, οι δύο τελευταίες τάσεις (Point outside of control limits, Stratification) θα απουσιάζουν από την εκπαίδευση και θα κυριαρχούν κατά αποκλειστικότητα στον έλεγχο, πράγμα ανεπιθύμητο. Αντιθέτως είναι απαραίτητο όλες οι τάσεις να συμμετέχουν, κατά αναλογία, ισόποσα τόσο στην εκπαίδευση όσο και στον έλεγχο. Αυτό πραγματοποιείται με τον εξής τρόπο:

Τα 36 διανύσματα κάθε τάσης αναδιατάσσονται εσωτερικά με τυχαίο τρόπο. Οι τάσεις όμως, ως σύνολα, παραμένουν στις ίδιες θέσεις, δηλαδή δεν αλλάζει η σειρά εμφάνισης τους. Στην συνέχεια από κάθε τάση επιλέγονται τα πρώτα 27 διανύσματα, για να συγκροτήσουν το συνολικό διάνυσμα εκπαίδευσης και τα υπόλοιπα 9, για να συγκροτήσουν το συνολικό διάνυσμα ελέγχου. Με αυτόν τον τρόπο διασφαλίζεται η ισομερής συνεισφορά μεταξύ των τάσεων, στην εκπαίδευση και στον έλεγχο αποτελεσμάτων και επιπλέον εισάγεται τυχαιότητα στην επιλογή των διανυσμάτων κάθε μίας.

Αντίστοιχα, με την επανάληψη 27 φορών του διανύσματος στόχου κάθε τάσης, διαμορφώνεται το συνολικό διάνυσμα στόχος της εκπαίδευσης ενώ με την επανάληψη 9 φορών, το συνολικό διάνυσμα στόχος του ελέγχου απόδοσης.

4.4 Συμπληρωματικά διανύσματα ελέγχου

Με σκοπό να ελεγχθεί περεταίρω η εγκυρότητα των αποτελεσμάτων, που προκύπτουν από την πρώτη, κύρια βάση δεδομένων, αναπτύσσεται και μία δευτερεύουσα που παίζει συμπληρωματικό ρόλο. Δημιουργείται προσθέτοντας στα τρία μητρικά διανύσματα κάθε τάσης έναν τυχαίο αριθμό που ανήκει στο [0,1000]. Δηλαδή, αποτελείται από 27 διανύσματα, που είναι τα μητρικά, μετατοπισμένα. Η λογική πίσω από αυτήν την κατασκευή, βασίζεται στον τρόπο λειτουργείας του αλγορίθμου knn, που είναι ο υπολογισμός των αποστάσεων (βλ 2.3 Αλγόριθμος k-Κοντινότερων Γειτόνων (k-nearest neighbors). Η μορφοποίηση έχει στόχο να διασταυρώσει κατά πόσο το κάθε μοντέλο είναι σε θέση να αναγνωρίσει μία τάση όταν η τάξη μεγέθους της δεν συμπίπτει με την τάξη μεγέθους κάποιου από τα διανύσματα με τα οποία εκπαιδεύτηκε, αλλά μάλλον συμπίπτει με αυτήν κάποιας άλλης τάσης κατά την εκπαίδευση. Για παράδειγμα:

Έστω ότι κάποιο διάνυσμα της κυκλικής τάσης έχει δημιουργηθεί από την μήτρα Cycle pattern 1 με την πρόσθεση της σταθεράς 920. Αυτό σημαίνει ότι η κυκλική τάση διαθέτει ένα διάνυσμα κατά την εκπαίδευση γύρω από την περιοχή του 920, που πιθανώς κάποια άλλη τάση, έστω η μονότονη να μην διαθέτει. Είναι ικανός ο εκάστοτε αλγόριθμος να αναγνωρίσει μία μονότονη τάση που προκύπτει στην περιοχή του 920 ή λόγω εγγύτητας την κατηγοριοποιεί ως κυκλική; Η απάντηση σε αυτό το ερώτημα αναζητείται μέσω της δευτερεύουσας βάσης δεδομένων.

Σημειώνεται ότι τα δεδομένα αυτά χρησιμοποιούνται μόνο για τον έλεγχο αποτελεσμάτων και όχι για την εκπαίδευση, η οποία πραγματοποιείται με τα διανύσματα εκπαίδευσης της κύριας βάσης δεδομένων. Παράλληλα με τον τρόπο που περιγράφηκε παραπάνω δημιουργούνται και τα κατάλληλα συνολικά διανύσματα στόχοι της εκπαίδευσης, με 27 διανύσματα.

Τέλος, μορφοποιούνται επιπλέον 19 διανύσματα μεικτών τάσεων τα οποία παρουσιάζονται μαζί με τα αντίστοιχα αποτελέσματα στο Παράρτημα Γ.

5 Ανάπτυξη μοντέλων αναγνώρισης τάσεων - Αποτελέσματα

5.1 Ανάπτυξη μοντέλων ΤΝΔ

Μέχρι σήμερα δεν υπάρχει θεωρητική μέθοδος καθορισμού της ιδανικής διαμόρφωσης ΤΝΔ για την αναγνώριση τάσεων σε διαγράμματα ελέγχου. Τα μοντέλα των ΤΝΔ που αναπτύσσονται έχουν οριοθετηθεί από τους εξής παράγοντες:

1. Τα ιδιαίτερα χαρακτηριστικά του προβλήματος, δηλαδή, το πλήθος και το μέγεθος των δεδομένων, τον αριθμό των κλάσεων κ.α.
2. Τις υπάρχουσες υπολογιστικές δυνατότητες
3. Την βιβλιογραφία

Με σκοπό να διερευνηθεί η επίδραση ποικίλων παραμέτρων στην απόδοση, δοκιμάζονται μοντέλα τα οποία διαφέρουν μεταξύ τους σε τουλάχιστον ένα από τα εξής:

1. Αλγόριθμος εκπαίδευσης

   Οι αλγόριθμοι εκπαίδευσης που υλοποιούνται είναι οι Levenberg-Marquardt (LM) και Scaled Conjugated Gradient Descent (SCG). Ο πρώτος συνδυάζεται με το μέσο τετραγωνικό σφάλμα ως συνάρτηση κόστους ενώ ο δεύτερος τόσο με το μέσο τετραγωνικό σφάλμα όσο με την Cross Entropy.

2. Αρχιτεκτονική, δηλαδή αριθμός επιπέδων και πλήθος κόμβων.

   Συγκεκριμένα δοκιμάζονται αρχιτεκτονικές με 2 και 3 κρυμμένα επίπεδα. Όπως αναφέρθηκε παραπάνω, ένα νευρωνικό δίκτυο τριών επιπέδων με επαρκή αριθμό κόμβων, μπορεί να προσεγγίσει οποιαδήποτε μη γραμμική συνάρτηση [10]. Ο αριθμός των κόμβων των κρυμμένων επιπέδων παίρνει μία από τις τιμές 8, 16, 20, 25 και 30. Αυτό είναι το εύρος που χρησιμοποιείται κατά κύριο λόγο στην βιβλιογραφία. Το σύνολο των πιθανών συνδυασμών των 5 αυτών τιμών σε δυάδες και τριάδες είναι 150. Συνεπώς, αναφορικά με τον αριθμό των κρυμμένων επιπέδων και τους κόμβους τους, εξετάζονται 150 συνδυασμοί.

   Ο αριθμός κόμβων του επιπέδου εισόδου είναι 16, αφού τόσα είναι τα στοιχεία των διανυσμάτων εισόδου ενώ ο αριθμός κόμβων του επιπέδου εξόδου είναι 8, εφόσον τόσες είναι οι κλάσεις.

3. Συνάρτηση ενεργοποίησης, με έμφαση σε αυτήν του επιπέδου εξόδου.

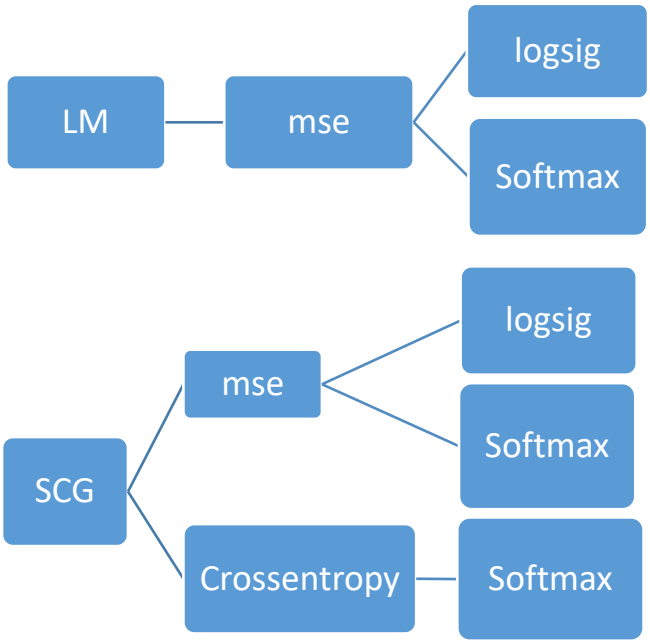
   Ως συνάρτηση ενεργοποίησης του επιπέδου εξόδου επιλέγεται είτε η συνάρτηση Softmax είτε η logsig. Υπενθυμίζεται ότι η πρώτη είναι απαραίτητη όταν η συνάρτηση κόστους είναι η Cross Entropy ενώ, η δεύτερη συνδυάζεται τόσο με την Cross Entropy όσο και με το μέσο τετραγωνικό σφάλμα. Σε όλα τα μοντέλα η συνάρτηση ενεργοποίησης των κρυμμένων επιπέδων είναι η logsig και του επιπέδου εισόδου η ταυτοτική συνάρτηση.

Η αρχιτεκτονικές που μελετώνται, αναφορικά με τον αλγόριθμο εκπαίδευσης, την συνάρτηση κόστους και την συνάρτηση ενεργοποίησης του επιπέδου εξόδου, παρουσιάζονται στην Εικόνα 52. Στο σύνολό τους είναι 5. Η κάθε μία από αυτές συνδυάζεται με τους 150 συνδυασμούς κρυμμένων επιπέδων και κόμβων και συνεπώς προκύπτουν συνολικά 750 αρχιτεκτονικές.

Η κάθε αρχιτεκτονική εκπαιδεύεται 7 φορές και από αυτές, υπολογίζεται η μέση και η βέλτιστη τιμή της ακρίβειας. Ως «μοντέλο» αναφέρεται το κάθε εκπαιδευμένο ΤΝΔ μεμονωμένα. Συνεπώς από κάθε αρχιτεκτονική προκύπτουν 7 μοντέλα. Εφόσον οι αρχιτεκτονικές είναι 750 προκύπτουν συνολικά 5250 μοντέλα ΤΝΔ.

Η δημιουργία και η διαλογή των διανυσμάτων εκπαίδευσης και ελέγχου, πραγματοποιούνται πριν την έναρξη των δοκιμών και δεν μεταβάλλονται, καθ' όλη την διάρκειά τους. Συνεπώς όλα τα μοντέλα ΤΝΔ εκπαιδεύονται με τα ίδια διανύσματα εκπαίδευσης. Ως μέγιστος αριθμός εποχών ορίζονται οι 1500 καθώς αυτό καθιστά το υπολογιστικό φορτίο ανεκτό και την ακρίβεια επαρκή, αφού τα περισσότερα μοντέλα δεν απαιτούν περισσότερες από 1000.

Σε όλες τις περιπτώσεις, η εκπαίδευση πραγματοποιείται με οπισθόδρομη διάδοση σφάλματος και η επισκόπηση των αποτελεσμάτων γίνεται μέσω του πίνακα σύγχυσης. Όλες οι αναλύσεις πραγματοποιήθηκαν με το λογισμικό Matlab 2020a.



Εικόνα 52: Οι αρχιτεκτονικές ΤΝΔ που αναπτύχθηκαν.

5.1.1 Αντιστοίχιση εξόδου-τάσης

Στην παράγραφο 4.2 επεξηγείται ο τρόπος με τον οποίο αντιστοιχίζονται τα διανύσματα στόχοι, που κάθε στοιχείο τους έχει την τιμή 0 ή 1, με τις τάσεις. Στην παρούσα εργασία, οι συναρτήσεις εξόδου που χρησιμοποιούνται διαμορφώνουν διανύσματα εξόδου, με τιμές των στοιχείων τους, που ανήκουν στο [0,1]. Προκειμένου να καταστεί δυνατή η

κατηγοριοποίηση, επιλέγεται το μεγαλύτερο στοιχείο του διανύσματος εξόδου και στρογγυλοποιείται στο 1, ενώ τα υπόλοιπα στρογγυλοποιούνται στον κοντινότερο ακέραιο. Με αυτόν τον τρόπο αποφεύγεται ο μηδενισμός όλων των στοιχείων ταυτόχρονα και τουλάχιστον ένα από αυτά, το μεγαλύτερο, λαμβάνει την τιμή 1. Επομένως, υπάρχουν δύο πιθανές εκδοχές:

- Η τιμή 1 να εμφανίζεται μόνο μία φορά, στην θέση του μεγαλύτερου στοιχείου. Αυτό συμβαίνει αν όλα τα υπόλοιπα στοιχεία, έχουν πριν την στρογγυλοποίηση τιμή ίση ή κάτω του 0.5. Σε αυτήν την περίπτωση το διάνυσμα εξόδου, έχοντας την μορφή κάποιου διανύσματος στόχου, κατηγοριοποιείται στην αντίστοιχη τάση.
- Η τιμή 1 να εμφανίζεται περισσότερες από μία φορές. Αυτό συμβαίνει αν κάποιο από τα υπόλοιπα στοιχεία, πέραν του μεγαλύτερου, λαμβάνει πριν την στρογγυλοποίηση τιμή μεγαλύτερη του 0.5. Σε αυτήν την περίπτωση το διάνυσμα εξόδου ταξινομείται στην ειδική κατηγορία που κατασκευάζεται για αυτόν τον σκοπό, με το όνομα «Mixed Trends».

  Προφανώς για να συμβεί κάτι τέτοιο, ως συνάρτηση εξόδου δεν χρησιμοποιείται η Softmax που αντιπροσωπεύει πιθανότητα, καθώς θα ήταν αδύνατο το δεύτερο μεγαλύτερο στοιχείο να ξεπερνά το 0.5.

Κατά συνέπεια στο διάγραμμα σύγχυσης εμφανίζονται 9 κατηγορίες, οι 8 τάσεις και η κατηγορία «Mixes Trends».

5.2 Αποτελέσματα ΤΝΔ

5.2.1 Αποτελέσματα κύριας βάσης δεδομένων

5.2.1.1 Κριτήριο μέσης ακρίβειας αρχιτεκτονικής

Η κύρια βάση δεδομένων αποτελείται από 288 διανύσματα εκπαίδευσης, δηλαδή 8 τάσεις με 36 διανύσματα η κάθε μία, ισάριθμα διανύσματα στόχους και 72 διανύσματα ελέγχου, δηλαδή 9 από κάθε τάση.

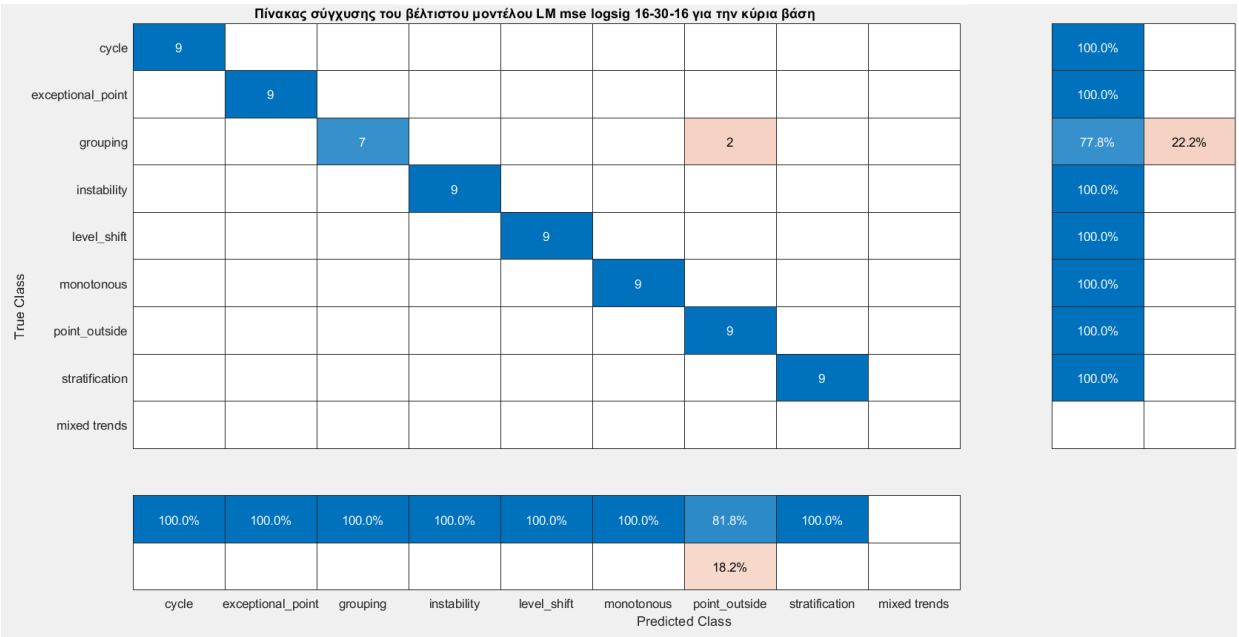
Στον Πίνακα 21 του παραρτήματος Α παρουσιάζονται η μέση και η βέλτιστη ακρίβεια που προκύπτουν από τις 7 εκπαιδεύσεις της κάθε αρχιτεκτονικής, για τα 72 διανύσματα ελέγχου της κύριας βάσης.

Για κάθε μία από τις 5 αρχιτεκτονικές στην Εικόνα 52 επιλέγονται οι τρείς συνδυασμοί αριθμού κρυμμένων επιπέδων και κόμβων, που αποδίδουν την μεγαλύτερη μέση ακρίβεια και παρουσιάζονται παρακάτω. Εάν σε κάποια θέση ισοβαθμούν περισσότερα μοντέλα τότε χρησιμοποιείται σαν δεύτερο κριτήριο η βέλτιστη ακρίβεια της κάθε αρχιτεκτονικής.

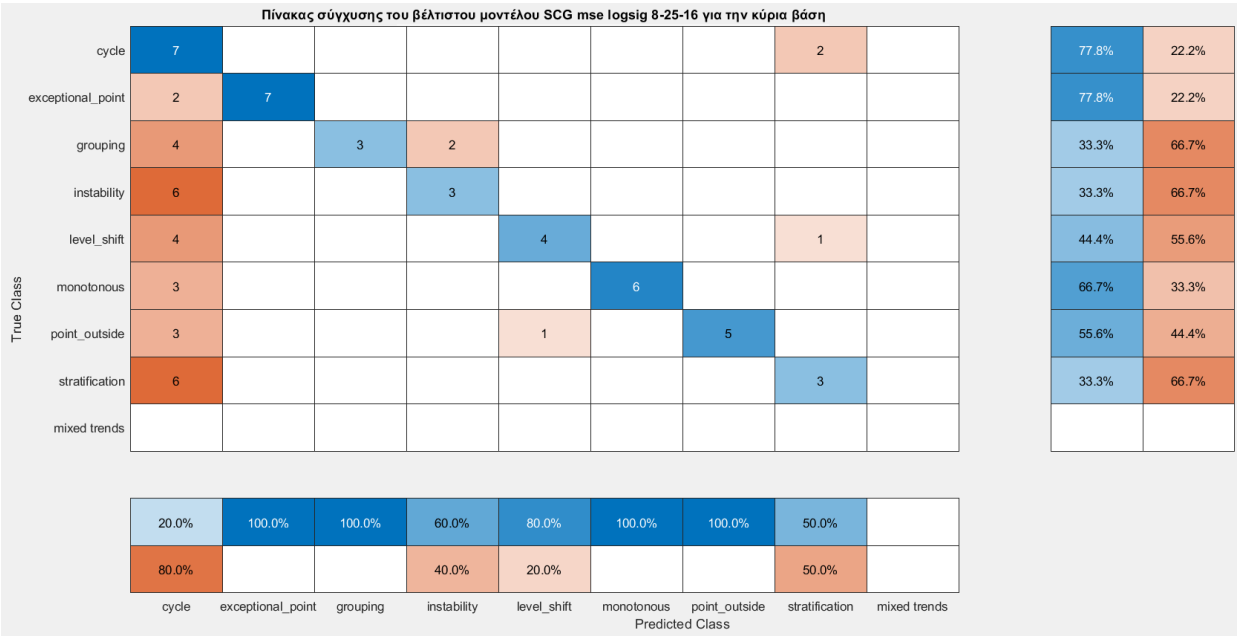
Πίνακας 4: Αποτελέσματα με βάση την μέση ακρίβεια, των τριών πιο επιτυχημένων συνδυασμών κρυμμένων επιπέδων και κόμβων, κάθε αρχιτεκτονικής.

Αρίθμηση	Αλγόριθμος Εκπαίδευσης	Συν.Κόστους	Συν. Ενεργ. Επ. Εξόδου	Κόμβοι Κρυμμένων Επιπέδων	Μέση Ακρίβεια (%)	Βέλτιστη Ακρίβεια (%)
1	LM	mse	logsig	16-30-16	94.64	97.22
2	LM	mse	logsig	8-30-20	94.44	98.61
3	LM	mse	logsig	8-30-16	93.05	100
4	LM	mse	softmax	8-20-30	93.25	97.22
5	LM	mse	softmax	8-25-20	93.25	97.22
6	LM	mse	softmax	16-30-30	92.86	100
7	SCG	mse	softmax	8-25-30	36.51	43.06
8	SCG	mse	softmax	8-30	34.13	50
9	SCG	mse	softmax	16-16	33.73	52.78
10	SCG	mse	logsig	8-25-16	42.46	52.78
11	SCG	mse	logsig	8-30-25	41.27	52.78
12	SCG	mse	logsig	8-20-30	39.48	59.72
13	SCG	crossentropy	softmax	8-30	41.67	73.61
14	SCG	crossentropy	softmax	8-30-30	38.89	52.78
15	SCG	crossentropy	softmax	8-25-25	38.49	54.17

Για κάθε έναν από τους 2 αλγορίθμους εκπαίδευσης, παρουσιάζεται ο πίνακας σύγχυσης, του μοντέλου που απέδωσε την βέλτιστη ακρίβεια, από το σύνολο των 7 μοντέλων που απέδωσαν την μέγιστη μέση ακρίβεια.



(α)



(β)

Εικόνα 53: Πίνακες σύγχυσης των βέλτιστων μοντέλων από τα 7 που απέδωσαν την μέγιστη μέση ακρίβεια, για τους αλγορίθμους εκπαίδευσης LM και SCG. Στο δεξί τμήμα φαίνεται η ανάκληση και στο κάτω η ακρίβεια κάθε κλάσης. (α) Πίνακας σύγχυσης του μοντέλου LM mse logsig 16-30-16 με ακρίβεια 97.22% (β) Πίνακας σύγχυσης του μοντέλου SCG mse logsig 8-25-16 με ακρίβεια 52.78%.

Στο Παράρτημα Β φαίνεται ο Πίνακας 22 με τις τιμές της μέσης ακρίβειας και ανάκλησης της κάθε κλάσης, για όλες τις αρχιτεκτονικές του Πίνακα 4.

Για την πρώτη αρχιτεκτονική του Πίνακα 4 (LM mse logsig 16-30-16) που είναι η ακριβέστερη με βάση την μέση ακρίβεια, παρουσιάζονται στον Πίνακα 5 η μέση ακρίβεια και η μέση ανάκληση της κάθε κλάσης.

Πίνακας 5: Μέσες ακρίβειες και ανακλήσεις των κλάσεων για την αρχιτεκτονική LM mse logsig 16-30-16, που έχει την μεγαλύτερη μέση ακρίβεια.

	Μέση Ακρίβεια Κλάσης (%)	**Μέση Ανάκληση Κλάσης(%)**
Cycle pattern	94.23	93.65
Exceptional point	95.6	98.41
Grouping	98.21	77.78
Instability	94.81	96.83
Level shift	97.4	100
Monotonous trend	97.4	100
Point outside control limits	94.23	96.83
Stratification	95.4	93.65

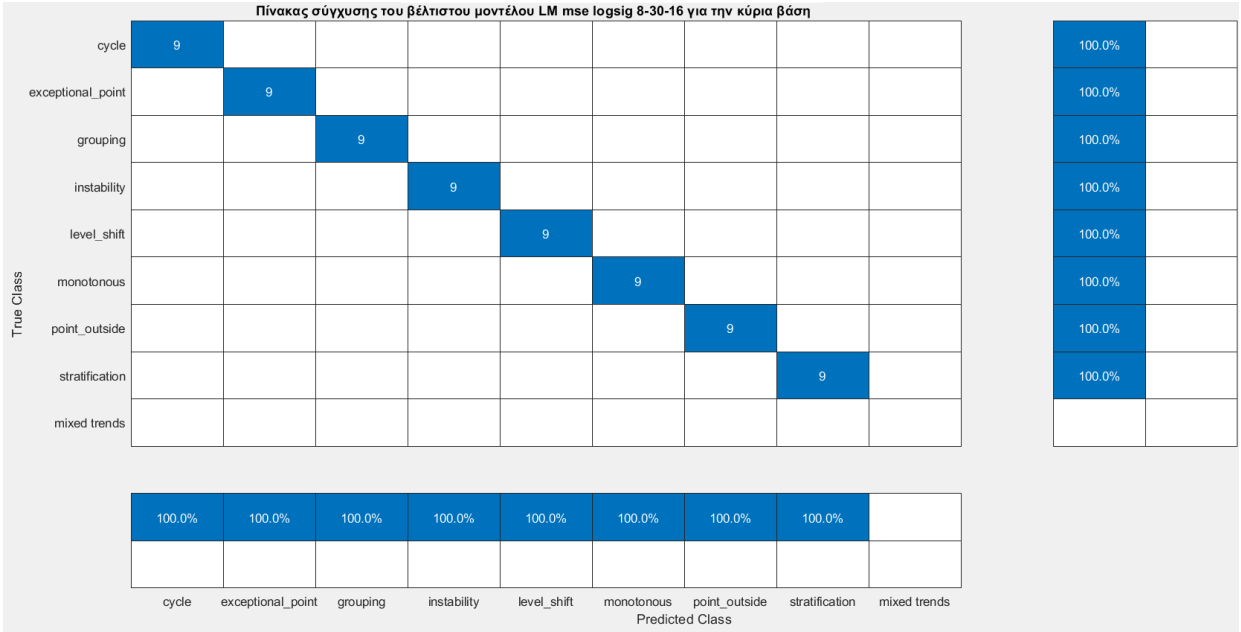
5.2.1.2 Κριτήριο βέλτιστης ακρίβειας αρχιτεκτονικής

Για κάθε μία από τις 5 αρχιτεκτονικές στην Εικόνα 52 επιλέγονται οι τρείς συνδυασμοί αριθμού κρυμμένων επιπέδων και κόμβων, από τις οποίες προκύπτουν τα μοντέλα που αποδίδουν τις τρείς μεγαλύτερες τιμές της βέλτιστης ακρίβειας. Το κριτήριο της βέλτιστης ακρίβειας είναι σημαντικό, καθώς τα μοντέλα που την επιτυγχάνουν, από κάθε αρχιτεκτονική, είναι αυτά που επιλέγονται στην συνέχεια για να δοκιμαστούν με την δευτερεύουσα βάση δεδομένων αλλά και με τις μεικτές τάσεις. Εάν σε κάποια θέση ισοβαθμούν περισσότερα μοντέλα τότε χρησιμοποιείται σαν δεύτερο κριτήριο η μέση ακρίβεια της κάθε αρχιτεκτονικής. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 6.

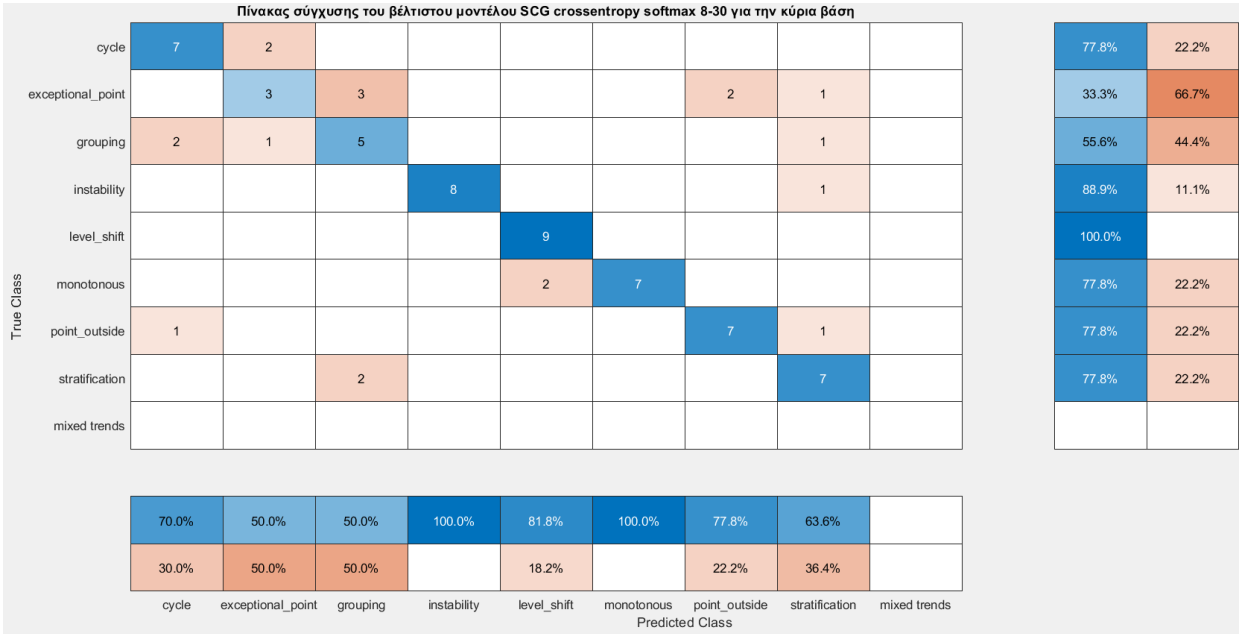
Πίνακας 6: Τα ακριβέστερα μοντέλα, των τριών πιο επιτυχημένων συνδυασμών κρυμμένων επιπέδων και κόμβων, κάθε αρχιτεκτονικής.

Αλγόριθμος Εκπαίδευσης	Συν.Κόστους	Συν. Ενεργ. Επ. Εξόδου	Κόμβοι Κρυμμένων Επιπέδων	Βέλτιστη Ακρίβεια (%)	Μέση Ακρίβεια Αρχιτεκτονικής (%)
LM	mse	logsig	8-30-16	100	93.05
LM	mse	logsig	8-25-25	100	91.86
LM	mse	logsig	8-30-25	100	91.07
LM	mse	softmax	16-30-30	100	92.86
LM	mse	softmax	20-25-8	100	67.86
LM	mse	softmax	8-8-30	98.61	71.23
SCG	mse	softmax	8-16	62.5	31.75
SCG	mse	softmax	8-30-8	59.72	33.14
SCG	mse	softmax	8-8-8	58.33	31.35
SCG	mse	logsig	8-16-16	61.11	31.15
SCG	mse	logsig	8-20-30	59.72	39.48
SCG	mse	logsig	16-20-20	59.72	39.28
SCG	crossentropy	softmax	8-30	73.61	41.67
SCG	crossentropy	softmax	8-30-16	62.5	38.29
SCG	crossentropy	softmax	8-30-25	59.72	34.92

Για κάθε έναν από τους 2 αλγορίθμους εκπαίδευσης, παρουσιάζεται ο πίνακας σύγχυσης, του ακριβέστερου μοντέλου.



(α)



(β)

Εικόνα 54: Πίνακες σύγχυσης των ακριβέστερων μοντέλων των αλγορίθμων εκπαίδευσης LM και SCG. (α) Πίνακας σύγχυσης του μοντέλου LM mse logsig 8-30-16 με ακρίβεια 100%. (β) Πίνακας σύγχυσης του μοντέλου SCG crossentropy softmax 8-30 με ακρίβεια 73.61%.

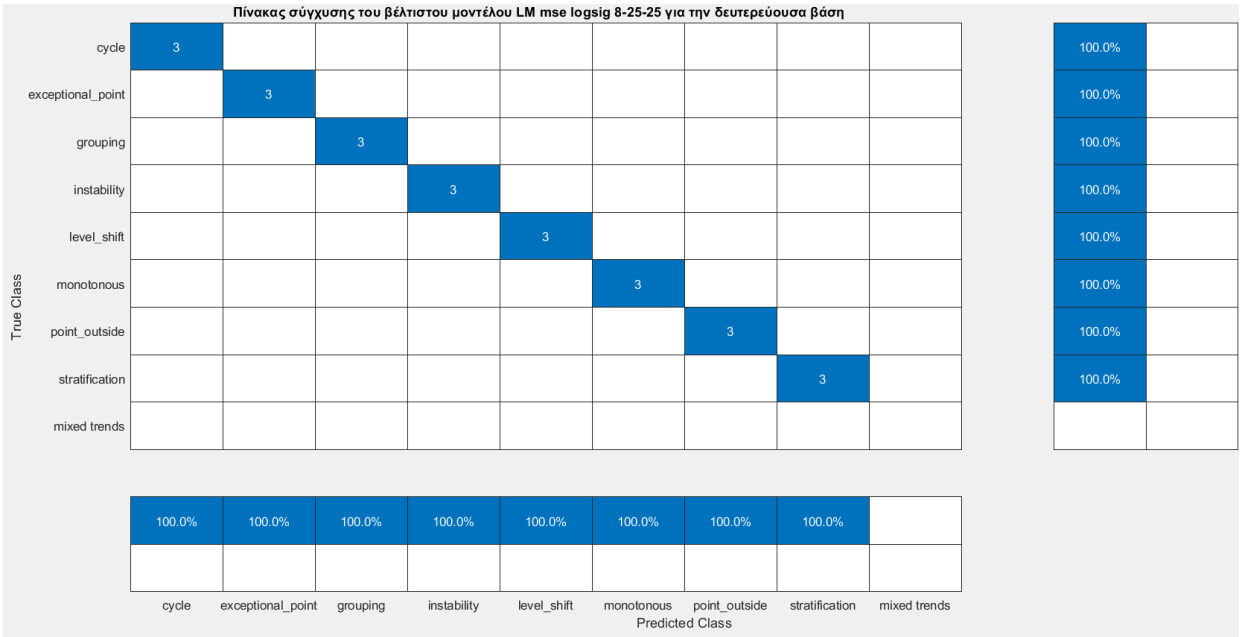
5.2.2 Αποτελέσματα με την δευτερεύουσα βάση δεδομένων

Προκειμένου να ελεγχθεί η ικανότητα γενίκευσής τους, τα μοντέλα του Πίνακα 6 ελέγχονται με τα 27 διανύσματα ελέγχου της δευτερεύουσας βάσης δεδομένων και στον Πίνακα 7 παρουσιάζονται τα αποτελέσματα.

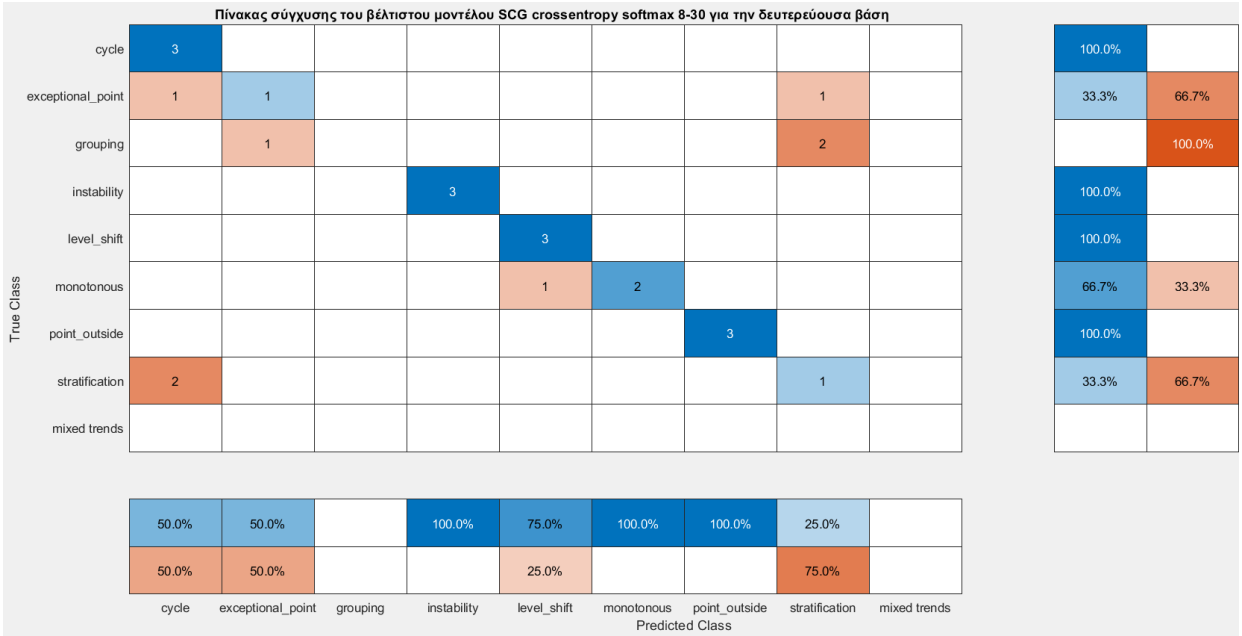
Πίνακας 7: Εφαρμογή των διανυσμάτων ελέγχου της δευτερεύουσας βάσης στα ακριβέστερα μοντέλα, των τριών πιο επιτυχημένων συνδυασμών κρυμμένων επιπέδων και κόμβων, κάθε αρχιτεκτονικής.

Αλγόριθμος Εκπαίδευσης	Συν.Κόστους	Συν. Ενεργ. Επ. Εξόδου	Κόμβοι Κρυμμένων Επιπέδων	Ακρίβεια (%)
LM	mse	logsig	8-30-16	95.83
LM	mse	logsig	8-25-25	100
LM	mse	logsig	8-30-25	70.83
LM	mse	softmax	16-30-30	83.33
LM	mse	softmax	20-25-8	50
LM	mse	softmax	8-8-30	70.83
SCG	mse	softmax	8-16	41.67
SCG	mse	softmax	8-30-8	12.5
SCG	mse	softmax	8-8-8	12.5
SCG	mse	logsig	8-16-16	12.5
SCG	mse	logsig	8-20-30	12.5
SCG	mse	logsig	16-20-20	12.5
SCG	crossentropy	softmax	8-30	66.67
SCG	crossentropy	softmax	8-30-16	12.5
SCG	crossentropy	softmax	8-30-25	12.5

Παρακάτω παρουσιάζεται ο πίνακας σύγχυσης του ακριβέστερου μοντέλου του Πίνακα 7 για κάθε αλγόριθμο εκπαίδευσης.



(α)



(β)

Εικόνα 55: Πίνακας σύγχυσης των ακριβέστερων μοντέλων κάθε αλγορίθμου εκπάιδευσης, από τις δοκιμές με την δευτερεύουσα βάση δεδομένων. (α) Πίνακας σύγχυσης του μοντέλου LM mse logsig 8-25-25 με ακρίβεια 100%. (β) Πίνακας σύγχυσης του μοντέλου SCG crossentropy softmax 8-30 με ακρίβεια 66.67%.

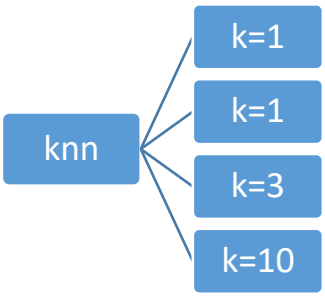
5.2.3 Αποτελέσματα με τα διανύσματα μεικτών τάσεων

Η δημιουργία των διανυσμάτων μεικτών τάσεων και τα αποτελέσματα των δοκιμών που πραγματοποιήθηκαν με αυτά, παρουσιάζονται στο Παράρτημα Δ.

5.3 Ανάπτυξη μοντέλων knn

Τα μοντέλα knn που αναπτύσσονται αφορούν τιμές του k 1, 2, 3 και 10, που επιλέχθηκαν βάση των όσων αναφέρονται στην παράγραφο 2.3. Σε όλα χρησιμοποιείται η Ευκλείδεια απόσταση.

Όπως παρουσιάστηκε στην παράγραφο 2.3 μία συνήθης τακτική για την βελτίωση της εγκυρότητας της μεθόδου είναι η κανονικοποίηση των δεδομένων. Η μέθοδος αυτή δεν είναι αποτελεσματική στα δεδομένα της παρούσας εργασίας και αυτό οφείλεται στον τρόπο με τον οποίο έχουν δημιουργηθεί τα 18 παράγωγα διανύσματα από τα μητρικά. Εξ' αιτίας του ότι έχουν προκύψει ως γραμμικοί συνδυασμοί της κάθε μήτρας εάν αναχθούν στο διάστημα [-1,1] μέσω της εξίσωσης (1.5) γίνονται ταυτόσημα με τα μητρικά τους.



Εικόνα 56: Μοντέλα knn που αναπτύχθηκαν.

5.3.1 Αποτελέσματα knn

5.3.1.1 Αποτελέσματα κύριας βάσης δεδομένων

Στον Πίνακα 8 παρουσιάζονται τα αποτελέσματα των δοκιμών για τα μοντέλα που αναπτύχθηκαν και εφαρμόστηκαν στην κύρια βάση δεδομένων. Τα διανύσματα εκπαίδευσης και ελέγχου είναι τα ίδια με αυτά των δοκιμών των ΤΝΔ. Το κάθε μοντέλο knn εκπαιδεύεται μία φορά, εφόσον η απόδοσή του παραμένει σταθερή για σταθερά δεδομένα εκπαίδευσης και ελέγχου.

Πίνακας 8: Αποτελέσματα μοντέλων knn για τα 72 διανύσματα ελέγχου της κύριας βάσης δεδομένων.

Μοντέλο	Τιμή k	Ακρίβεια (%)
KNN	1	89
KNN	2	58
KNN	3	43
KNN	10	24

Παρακάτω παρουσιάζεται ο πίνακας σύγχυσης της ακριβέστερης δοκιμής μοντέλου knn.

Εικόνα 57: Πίνακας σύγχυσης της ακριβέστερης περίπτωσης μοντέλου knn, με k=1 για την κύρια βάση δεδομένων.

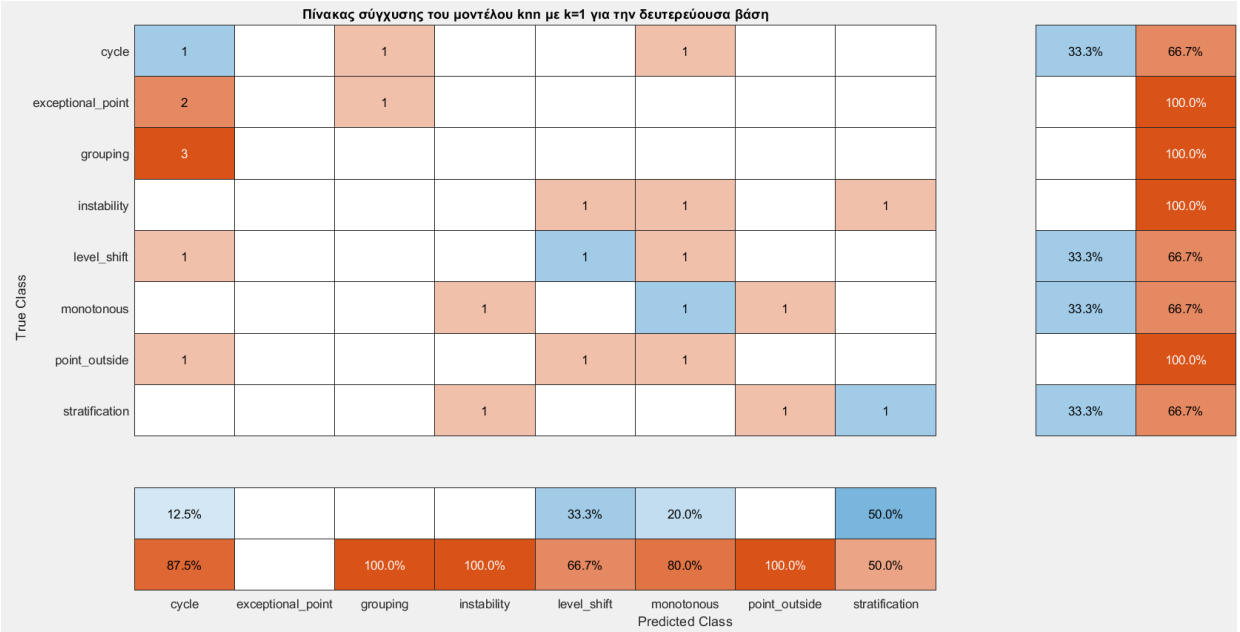
5.3.1.2 Αποτελέσματα δευτερεύουσας βάσης δεδομένων

Τα μοντέλα knn που δοκιμάστηκαν με την κύρια βάση δεδομένων, ελέγχονται και με τα 24 διανύσματα ελέγχου της δευτερεύουσας βάσης δεδομένων. Στον Πίνακα 9 παρουσιάζονται τα αποτελέσματα.

Πίνακας 9: Αποτελέσματα μοντέλων knn για τα 24 διανύσματα ελέγχου της δευτερεύουσας βάσης δεδομένων.

Μοντέλο	Τιμή k	Ακρίβεια (%)
KNN	1	16.67
KNN	2	12.5
KNN	3	12.5
KNN	10	8.33

Παρακάτω παρουσιάζεται ο πίνακας σύγχυσης του ακριβέστερου μοντέλου με k=1.



Εικόνα 58: Πίνακας σύγχυσης της ακριβέστερης περίπτωσης μοντέλου knn, με k=1 για την δευτερεύουσα βάση δεδομένων.

6 Σχολιασμός αποτελεσμάτων και συμπεράσματα

Από τα αποτελέσματα των δοκιμών, προκύπτουν ποικίλα συμπεράσματα, που αφορούν την συνεισφορά των διαφόρων παραμέτρων. Με σκοπό την εξαγωγή συμπερασμάτων, εξετάζεται κάθε παράμετρος ξεχωριστά, για τις διάφορες εκδοχές της, καθώς οι υπόλοιπες παράμετροι παραμένουν σταθερές. Με αυτόν τον τρόπο είναι δυνατή η εξέταση της επίδρασης της κάθε μίας στο αποτέλεσμα.

Αρχικά εξετάζονται οι παράμετροι των ΤΝΔ και στη συνέχεια του αλγορίθμου knn σε σχέση με τα αποτελέσματα που προκύπτουν από την κύρια βάση δεδομένων.

6.1 ΤΝΔ

6.1.1 Σχολιασμός αποτελεσμάτων της κύριας βάσης δεδομένων

6.1.1.1 Αλγόριθμος εκπαίδευσης

Όπως προκύπτει από τον Πίνακα 4 η επίδραση του αλγορίθμου εκπαίδευσης είναι καθοριστική. Για να εξεταστεί αυτή η επίδραση, απαιτείται η σύγκριση μεταξύ αρχιτεκτονικών που διαφέρουν μόνο στον αλγόριθμο εκπαίδευσης. Κατά συνέπεια, οι αρχιτεκτονικές με συνάρτηση κόστους την crossentropy παραβλέπονται, καθώς αυτή χρησιμοποιείται μόνο με τον αλγόριθμο εκπαίδευσης SCG.

Αρχικά, αντιπαραβάλλονται, οι τρεις ακριβέστερες με βάση την μέση ακρίβεια, αρχιτεκτονικές του κάθε αλγορίθμου εκπαίδευσης με τις αντίστοιχες αρχιτεκτονικές του εταίρου αλγορίθμου εκπαίδευσης μέσω του Πίνακα 4.

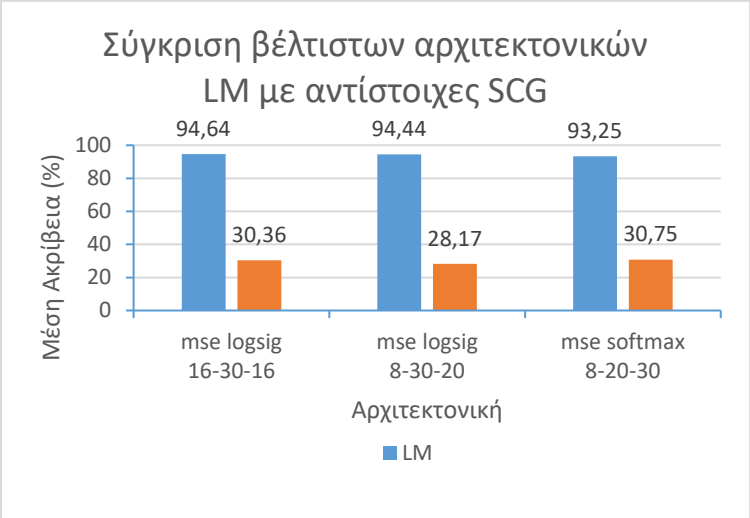
Πίνακας 10: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών LM, με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές SCG.

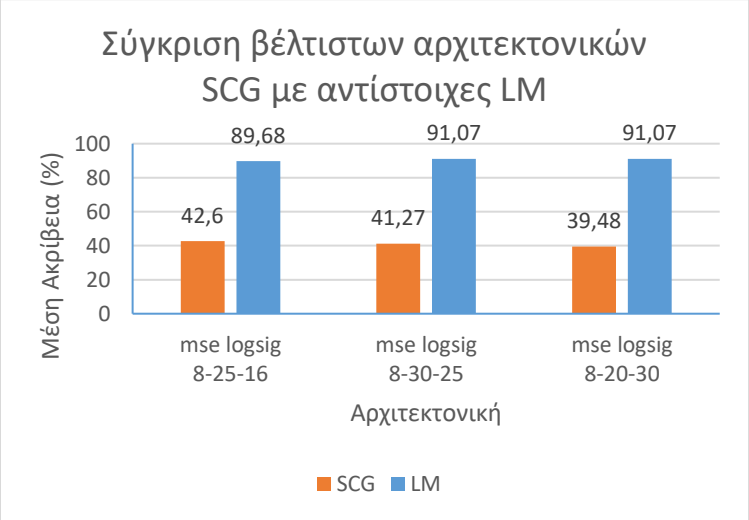
Συν.Κόστους	Συν. Ενεργ. Επ. Εξόδου	Κόμβοι Κρυμμένων Επιπέδων	Μέση Ακρίβεια (%)		Διαφορά (%)
			Ακριβέστερες LM	SCG	
mse	logsig	16-30-16	94.64	30.36	64.28
mse	logsig	8-30-20	94.44	28.17	66.27
mse	softmax	8-20-30	93.25	30.75	62.5
				Μέση διαφορά (%):	**64.35**

Πίνακας 11: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών SCG, με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές LM.

Συν.Κόστους	Συν. Ενεργ. Επ. Εξόδου	Κόμβοι Κρυμμένων Επιπέδων	Μέση Ακρίβεια (%)		Διαφορά (%)
			Ακριβέστερες SCG	LM	
mse	logsig	8-25-16	42.60	89.68	-47.08
mse	logsig	8-30-25	41.27	91.07	-49.8
mse	logsig	8-20-30	39.48	91.07	-51.59
				Μέση διαφορά (%):	**-49.49**

Παρακάτω φαίνονται τα αντίστοιχα γραφήματα.



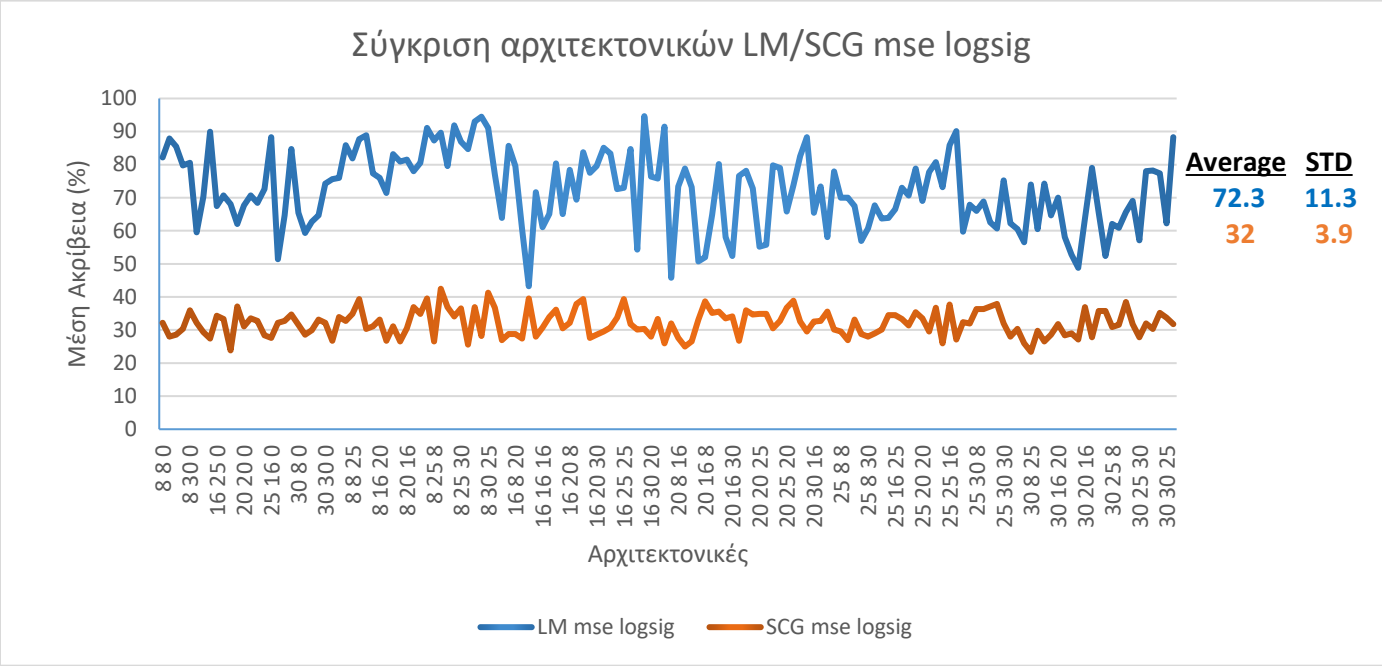


Εικόνα 59: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών κάθε αλγορίθμου εκπαίδευσης, με την μέση ακρίβεια των αντίστοιχων αρχιτεκτονικών με τον έτερο αλγόριθμο εκπαίδευσης. Αριστερά: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών LM με τις αντίστοιχες SCG. Δεξιά: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών SCG με τις αντίστοιχες LM.

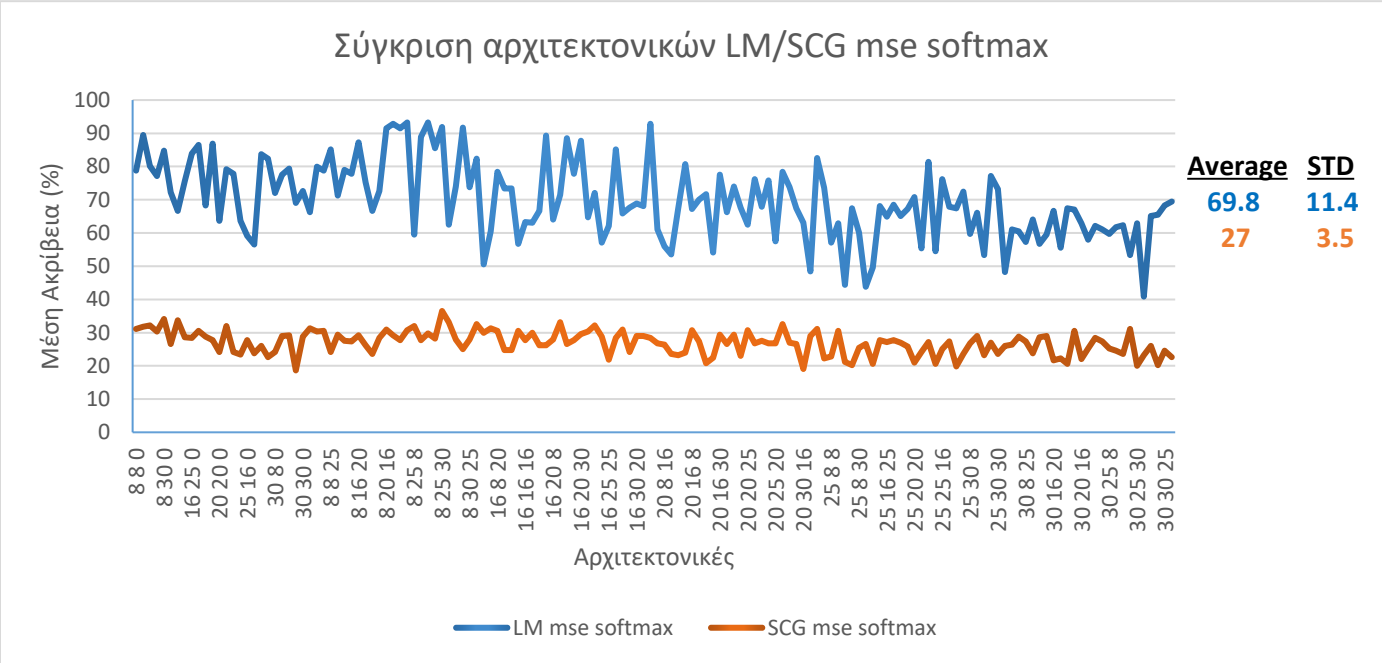
Από τα παραπάνω υποδηλώνεται ότι, οι βέλτιστες αρχιτεκτονικές με LM υπερέχουν των ιδίων αρχιτεκτονικών με SCG κατά 64.35% κατά μέσο όρο, αναφορικά με την μέση ακρίβεια. Αντιθέτως, οι βέλτιστες αρχιτεκτονικές με SCG υπολείπονται των ιδίων με LM κατά 49.49% κατά μέσο όρο.

Συμπληρωματικά παρατίθενται τα γραφήματα σύγκρισης της μέσης ακρίβειας, των 150 αρχιτεκτονικών LM mse logsig με τις αντίστοιχες SCG mse logsig καθώς και των

ισάριθμων αρχιτεκτονικών LM mse softmax με τις αντίστοιχες SCG mse softmax. Στον οριζόντιο άξονα οι αρχιτεκτονικές παρατίθενται με αύξουσα σειρά κρυμμένων επιπέδων και κόμβων κάθε κρυμμένου επιπέδου ξεκινώντας από το τελευταίο, προς το πρώτο. Επίσης σημειώνονται η μέση τιμή και η τυπική απόκλιση κάθε περίπτωσης.



(α)



(β)

Εικόνα 60: Σύγκριση ως προς την μέση ακρίβεια του συνόλου των αρχιτεκτονικών (α) LM και SCG mse logsig (β) LM και SCG mse softmax.

Συνεπώς από τα παραπάνω, εξάγεται το συμπέρασμα ότι, ο αλγόριθμος LM έχει σε γενικό επίπεδο υπερδιπλάσια ακρίβεια, από τον αλγόριθμο SCG στην συγκεκριμένη εφαρμογή.

Ωστόσο, ο μέσος χρόνος υπολογισμού, που απαιτεί στις δοκιμές, είναι τουλάχιστον δύο τάξεις μεγέθους μεγαλύτερος από τον αντίστοιχο των δοκιμών με τον αλγόριθμο SCG.

Τόσο η επίδοση όσο και η ταχύτητα του αλγορίθμου, που είναι αντιστρόφως ανάλογη με την απαίτηση σε μνήμη, είναι σύμφωνες με τα όσα αναφέρθηκαν στην θεωρία (βλ. Εικόνα 34).

6.1.1.2 Συνάρτηση Κόστους

Στην περίπτωση του αλγορίθμου LM η συνάρτηση κόστους δεν μεταβάλλεται και συνεπώς δεν μπορούν να εξεταστούν οι εκδοχές της. Στον αλγόριθμο SCG όπου μεταβάλλεται, συγκρίνονται οι δύο εκδοχές της, mse και crossentropy, πάντα με συνάρτηση ενεργοποίησης του επιπέδου εξόδου την softmax.

Αρχικά μέσω του Πίνακα 4 αντιπαρατίθενται ως προς την μέση ακρίβεια, οι ακριβέστερες αρχιτεκτονικές της κάθε περίπτωσης, με τις αντίστοιχες αρχιτεκτονικές της ετέρας συνάρτησης κόστους.

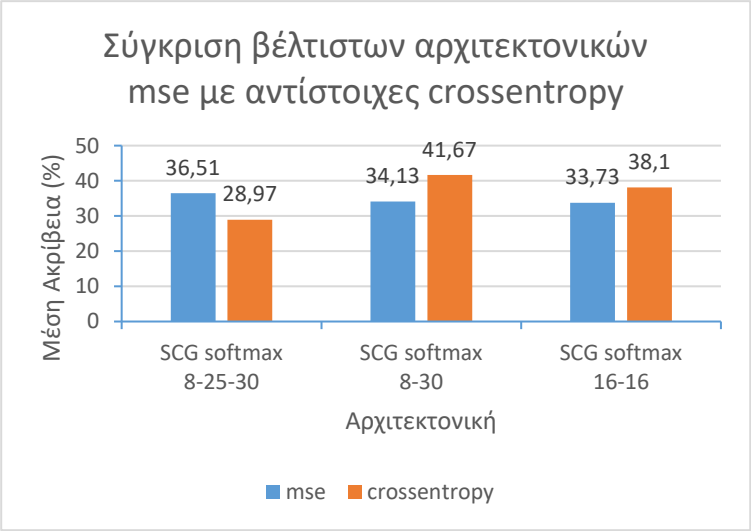
Πίνακας 12: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών SCG mse softmax με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές SCG crossentropy softmax.

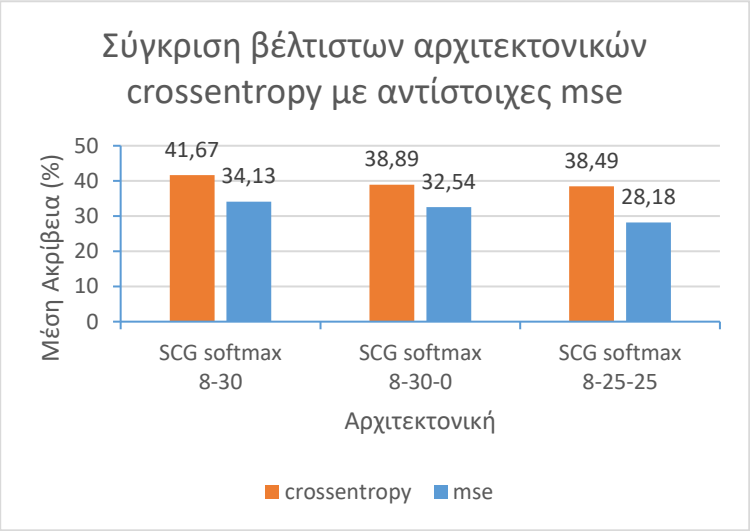
Αλγόριθμος Εκπαίδευσης	**Συν. Ενεργ. Επ. Εξόδου**	**Κόμβοι Κρυμμένων Επιπέδων**	**Μέση Ακρίβεια (%)**		**Διαφορά (%)**
			Ακριβέστερες mse	**crossentropy**	
SCG	softmax	8-25-30	36.51	28.97	7.54
SCG	softmax	8-30	34.13	41.67	-7.54
SCG	softmax	16-16	33.73	38.10	-4.37
				Μέση διαφορά (%):	**-1.46**

Πίνακας 13: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών SCG crossentropy softmax με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές SCG mse softmax.

Αλγόριθμος Εκπαίδευσης	Συν. Ενεργ. Επ. Εξόδου	Κόμβοι Κρυμμένων Επιπέδων	Μέση Ακρίβεια (%)		Διαφορά (%)
			Ακριβέστερες crossentropy	mse	
SCG	softmax	8-30	41.67	34.13	7.54
SCG	softmax	8-30-30	38.89	32.54	6.35
SCG	softmax	8-25-25	38.49	28.18	10.31
				Μέση διαφορά (%):	8.06

Παρακάτω φαίνονται τα αντίστοιχα γραφήματα.





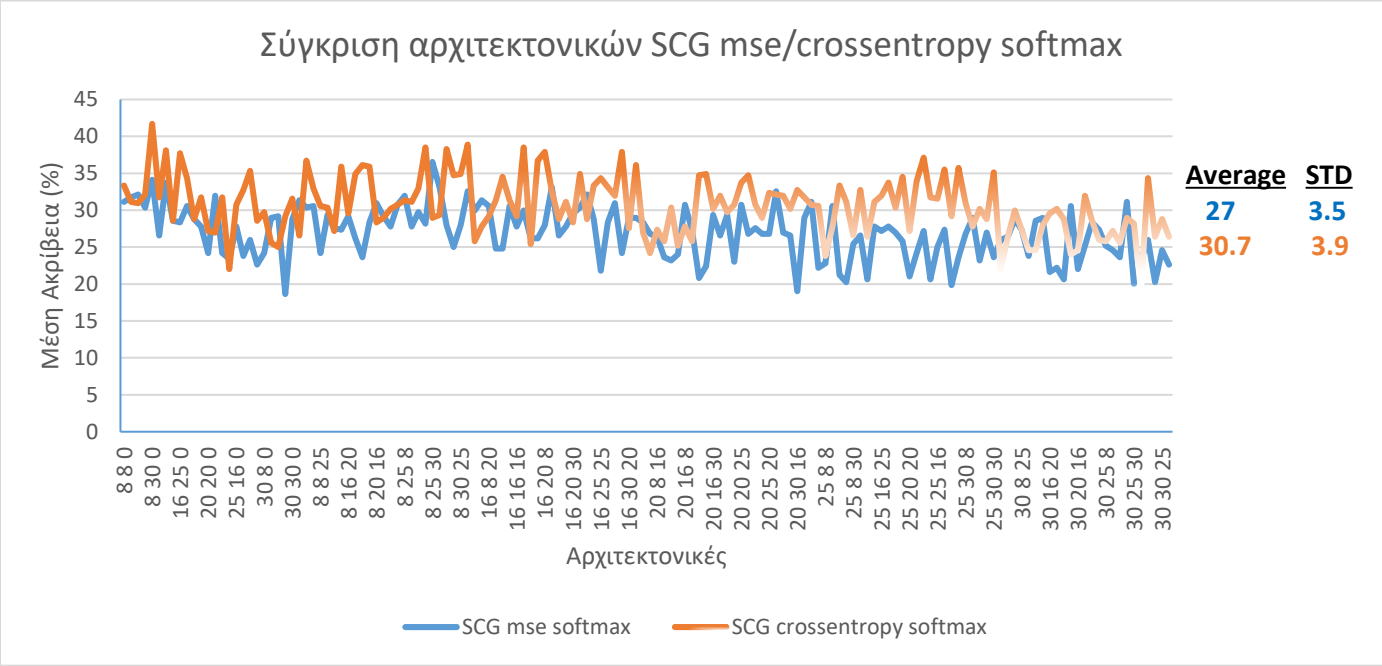
Εικόνα 61: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών κάθε συνάρτησης κόστους, με την μέση ακρίβεια των αντίστοιχων αρχιτεκτονικών με την ετέρα συνάρτηση κόστους. Αριστερά: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών SCG mse softmax με τις αντίστοιχες SCG crossentropy softmax. Δεξιά: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών SCG crossentropy softmax με τις αντίστοιχες SCG mse softmax.

Από τα παραπάνω δηλώνεται ότι οι βέλτιστες αρχιτεκτονικές με mse υπολείπονται, κατά μέσο όρο κατά 1.46%, αναφορικά με τη μέση ακρίβεια, σε σχέση με τις ίδιες αρχιτεκτονικές με crossentropy. Αντίστοιχα οι βέλτιστες αρχιτεκτονικές με crossentropy υπερέχουν, κατά μέσο όρο κατά 8.06%, αναφορικά με τη μέση ακρίβεια, σε σχέση με τις αντίστοιχες με mse.

Συμπληρωματικά, στην Εικόνα 62 παρατίθεται το γράφημα σύγκρισης της μέσης ακρίβειας, των 150 αρχιτεκτονικών SCG mse softmax με τις αντίστοιχες SCG crossentropy softmax. Αναγράφονται η μέση τιμή και η τυπική απόκλιση κάθε περίπτωσης.

Συνολικά, εξάγεται το συμπέρασμα ότι, η συνάρτηση κόστους crossentropy επιφέρει αποτελέσματα με μεγαλύτερη κατά 5%-10% σε γενικό επίπεδο ακρίβεια, από την συνάρτηση mse στην συγκεκριμένη εφαρμογή.

Επίσης ο χρόνος υπολογισμού παρατηρείται να είναι ο ίδιος και για τις δύο περιπτώσεις.



Εικόνα 62: Σύγκριση ως προς την μέση ακρίβεια του συνόλου των αρχιτεκτονικών SCG mse και crossentropy softmax.

6.1.1.3 Συνάρτηση ενεργοποίησης του επιπέδου εξόδου

Η εξέταση της επίδρασης των δύο διαφορετικών συναρτήσεων του επιπέδου εξόδου, της softmax και της logsig, είναι δυνατή μόνο στην περίπτωση που συνάρτηση κόστους είναι η mse, η οποία συνδυάζεται και με τις δύο. Για αυτήν την συνάρτηση κόστους, εξετάζονται οι δύο εκδοχές της συνάρτησης του επιπέδου εξόδου, για κάθε αλγόριθμο εκπαίδευσης.

Μέσω του Πίνακα 4 αντιπαρατίθενται ως προς την μέση ακρίβεια, οι ακριβέστερες αρχιτεκτονικές της κάθε συνάρτησης ενεργοποίησης του επιπέδου εξόδου, με τις αντίστοιχες αρχιτεκτονικές της ετέρας συνάρτησης ενεργοποίησης.

Πίνακας 14: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών LM mse logsig, με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές LM mse softmax.

Αλγόριθμος Εκπαίδευσης	Συν.Κόστους	Κόμβοι Κρυμμένων Επιπέδων	Μέση Ακρίβεια (%)		Διαφορά (%)
			Ακριβέστερες logsig	softmax	
LM	mse	16-30-16	94.64	67.66	26.98
LM	mse	8-30-20	94.44	91.67	2.77
LM	mse	8-30-16	93.05	74.01	19.04
				Μέση διαφορά (%):	**16.26**

Πίνακας 15: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών SCG mse logsig, με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές SCG mse softmax.

Αλγόριθμος Εκπαίδευσης	Συν.Κόστους	Κόμβοι Κρυμμένων Επιπέδων	Μέση Ακρίβεια (%)		Διαφορά (%)
			Ακριβέστερες logsig	softmax	
SCG	mse	8-25-16	42.46	27.78	14.68
SCG	mse	8-30-25	41.27	27.98	13.29
SCG	mse	8-20-30	39.48	30.75	8.73
				Μέση διαφορά (%):	**12.23**

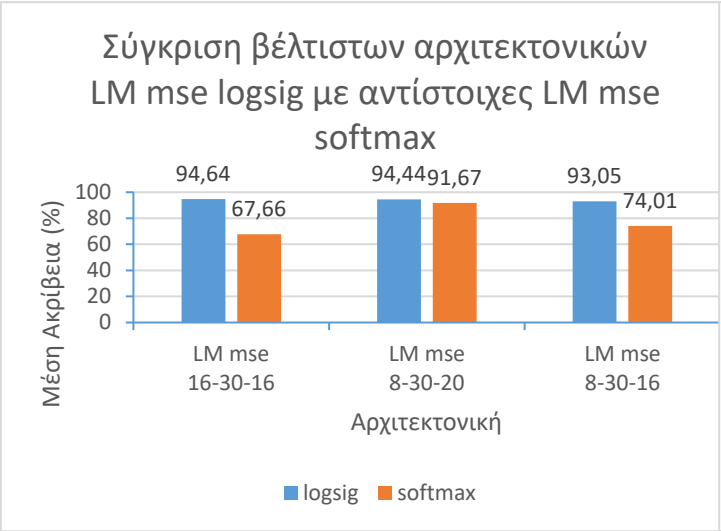
Πίνακας 16: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών LM mse softmax, με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές LM mse logsig.

Αλγόριθμος Εκπαίδευσης	Συν.Κόστους	Κόμβοι Κρυμμένων Επιπέδων	Μέση Ακρίβεια (%)		Διαφορά (%)
			Ακριβέστερες softmax	logsig	
LM	mse	8-20-30	93.25	91.07	2.18
LM	mse	8-25-20	93.25	79.56	13.69
LM	mse	16-30-30	92.86	91.47	1.39
				Μέση διαφορά (%):	**5.75**

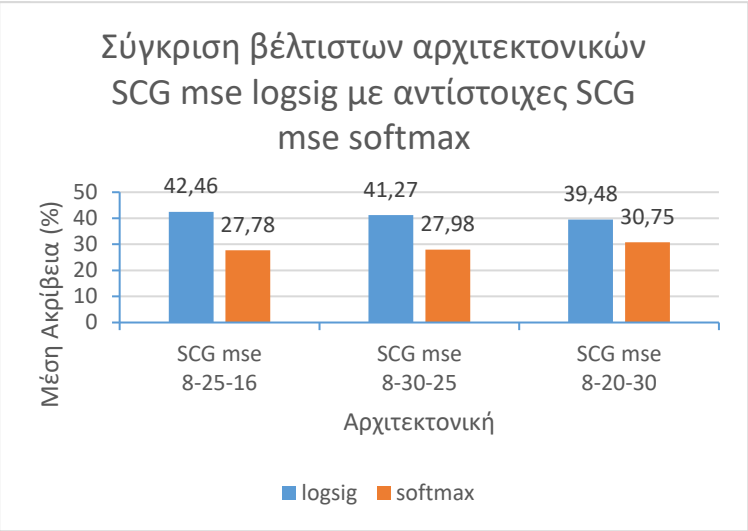
Πίνακας 17: Πίνακας σύγκρισης των ακριβέστερων αρχιτεκτονικών SCG mse softmax, με βάση την μέση ακρίβεια, με τις αντίστοιχες αρχιτεκτονικές LM mse logsig.

Αλγόριθμος Εκπαίδευσης	Συν.Κόστους	Κόμβοι Κρυμμένων Επιπέδων	Μέση Ακρίβεια (%)		Διαφορά (%)
			Ακριβέστερες softmax	logsig	
SCG	mse	8-25-30	36.51	36.51	0
SCG	mse	8-30	34.13	35.91	-1.78
SCG	mse	16-16	33.73	29.37	4.36
				Μέση διαφορά (%):	**0.86**

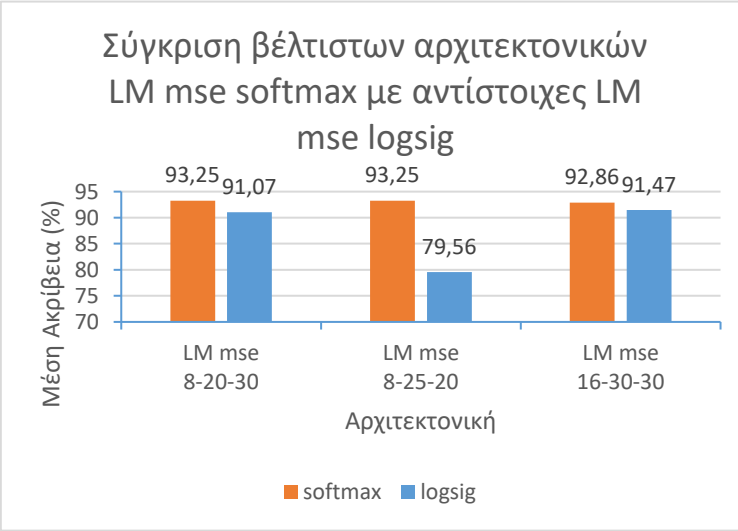
Παρακάτω φαίνονται τα αντίστοιχα γραφήματα.



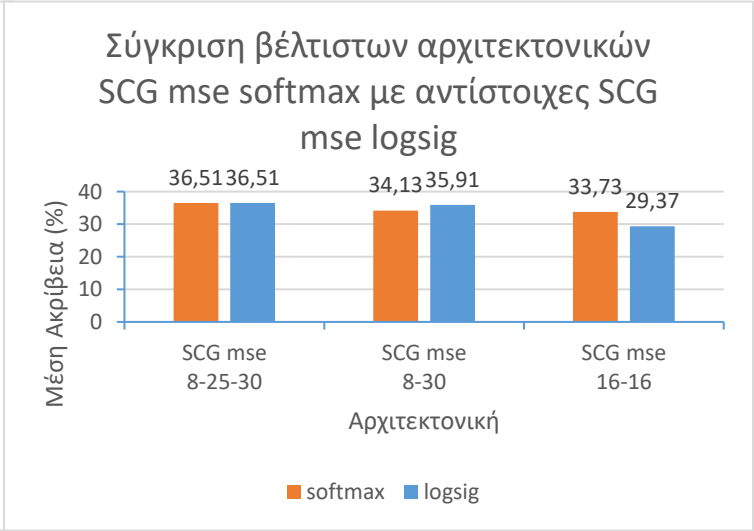
(α)



(β)



(γ)

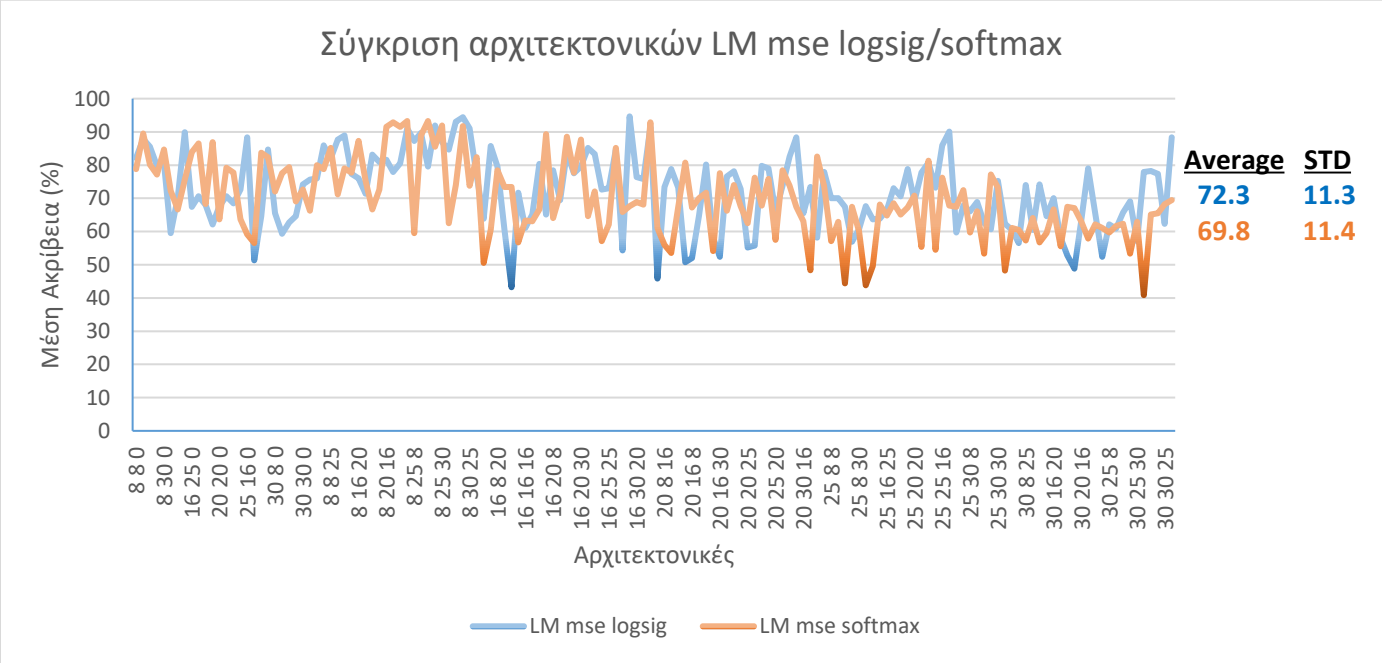


(δ)

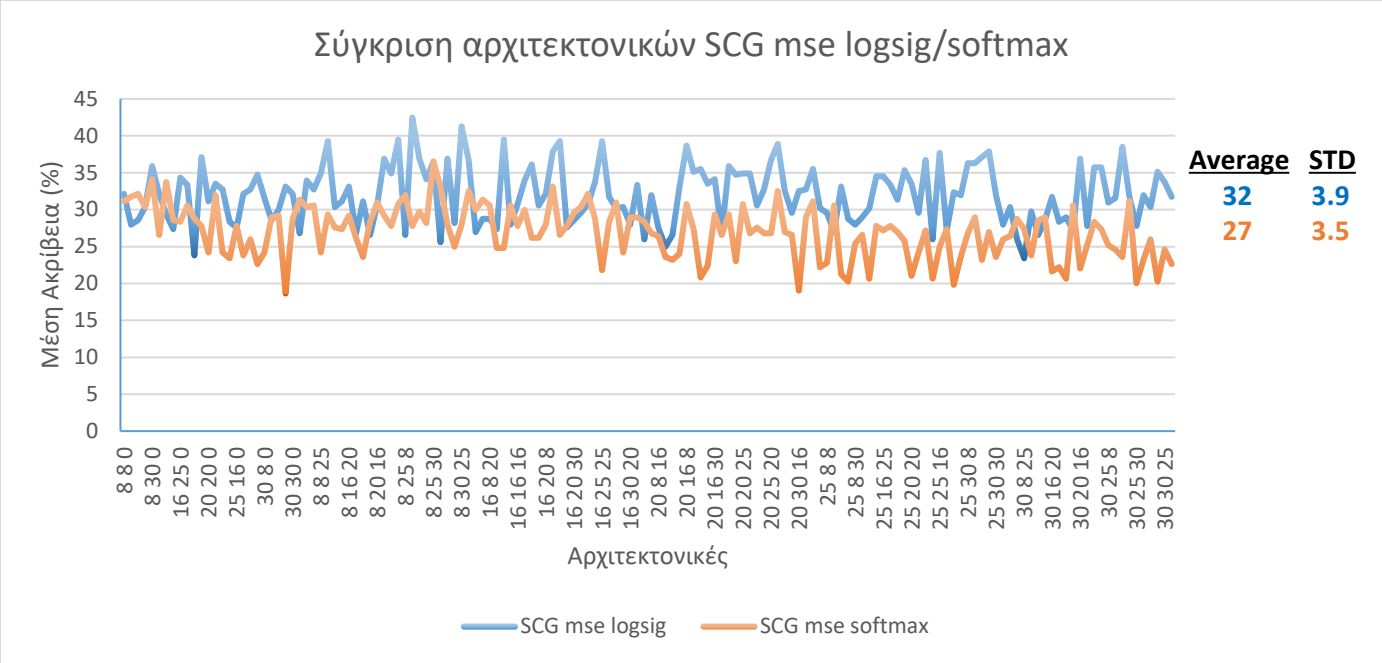
Εικόνα 63: Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών κάθε συνάρτησης ενεργοποίησης του επιπέδου εξόδου, με την μέση ακρίβεια των αντίστοιχων αρχιτεκτονικών με την ετέρα συνάρτηση ενεργοποίησης. (α) Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών LM mse logsig με τις αντίστοιχες LM mse softmax. (β) Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών SCG mse logsig με τις αντίστοιχες SCG mse softmax. (γ) Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών LM mse softmax με τις αντίστοιχες LM mse logsig. (δ) Σύγκριση μέσης ακρίβειας των βέλτιστων αρχιτεκτονικών SCG mse softmax με τις αντίστοιχες SCG mse logsig.

Από τα παραπάνω γίνεται εμφανές ότι στην σύγκριση των βέλτιστων με βάση την μέση ακρίβεια αρχιτεκτονικών, LM mse logsig και SCG mse logsig με τις ίδιες αρχιτεκτονικές, LM mse softmax και SCG mse softmax, οι πρώτες υπερέχουν κατά μέσο όρο, κατά 16% και 12% αντίστοιχα. Αντιθέτως, οι βέλτιστες με βάση την μέση ακρίβεια αρχιτεκτονικές, LM mse softmax και SCG mse softmax σε σύγκριση με τις ίδιες αρχιτεκτονικές, LM mse logsig και SCG mse logsig, υπερέχουν κατά μέσο όρο, κατά 6% και 1% αντίστοιχα.

Συμπληρωματικά παρατίθενται τα γραφήματα σύγκρισης της μέσης ακρίβειας, των 150 αρχιτεκτονικών LM mse logsig με τις αντίστοιχες LM mse softmax καθώς και των ισάριθμων αρχιτεκτονικών SCG mse logsig με τις αντίστοιχες SCG mse softmax. Αναγράφονται, η μέση τιμή και η τυπική απόκλιση κάθε περίπτωσης.



(α)



(β)

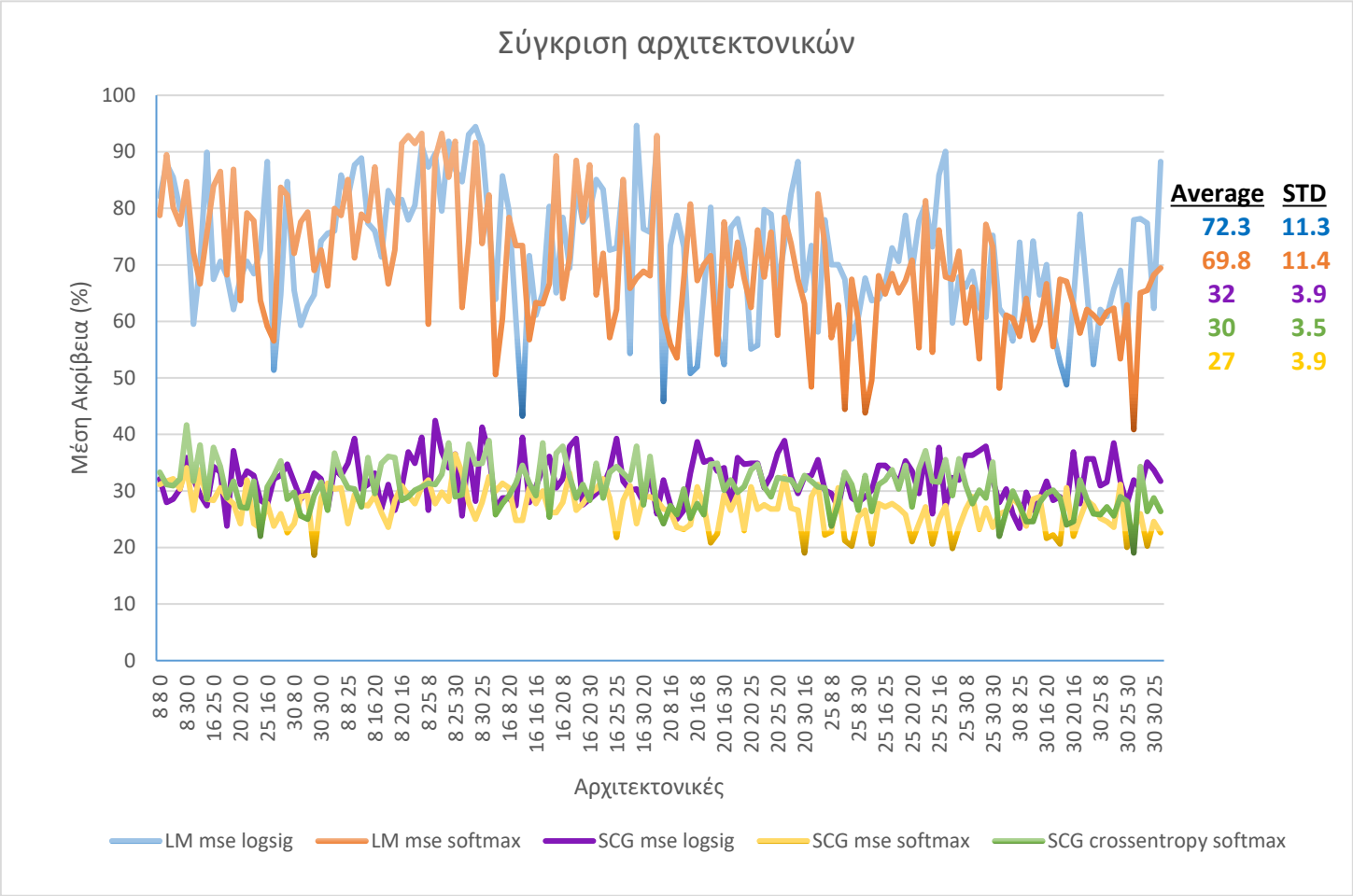
Εικόνα 64: Σύγκριση ως προς την μέση ακρίβεια του συνόλου των αρχιτεκτονικών (α) LM mse logsig και softmax (β) SCG mse logsig και softmax.

Συνεπώς από τα παραπάνω, εξάγεται το συμπέρασμα ότι, η logsig ως συνάρτηση ενεργοποίησης του επιπέδου εξόδου επιφέρει αποτελέσματα με μεγαλύτερη κατά 1%-10% σε γενικό επίπεδο ακρίβεια, από την συνάρτηση softmax στην συγκεκριμένη εφαρμογή. Μάλιστα η διαφορά τους φαίνεται να διευρύνεται υπερ της logsig όσο αυξάνεται ο αριθμός επιπέδων και κόμβων.

Επίσης ο χρόνος υπολογισμού παρατηρείται ο ίδιος και για τις δύο περιπτώσεις.

6.1.1.4 Αριθμός κρυμμένων επιπέδων και αριθμός κόμβων

Προκειμένου να εξεταστεί η συνεισφορά του κάθε συνδυασμού κρυμμένων επιπέδων και κόμβων παρουσιάζονται τα αποτελέσματα της μέσης ακρίβειας για το σύνολο των 150 συνδυασμών, για κάθε αρχιτεκτονική που αναπτύχθηκε (βλ. Εικόνα 52).



Εικόνα 65: Επίδραση του αριθμού κρυμμένων επιπέδων και κόμβων στις τιμές της μέσης ακρίβειας κάθε αρχιτεκτονικής.

Από τα παραπάνω διαγράμματα προκύπτει ότι οι ακραίοι συνδυασμοί αριθμού κρυμμένων επιπέδων και κόμβων δεν επιφέρουν τα βέλτιστα αποτελέσματα αναφορικά με την μέση ακρίβεια. Μάλιστα οι συνδυασμοί από το μέσο του γραφήματος (16-30-30) και έπειτα δείχνουν να επιφέρουν σε κάθε περίπτωση, μείωση της μέσης ακρίβειας σε σχέση με το πρώτο μισό. Πιο συγκεκριμένα, οι βέλτιστες αποδόσεις της μέσης ακρίβειας προέρχονται κατά πλειοψηφία από την περιοχή που περιλαμβάνει τους συνδυασμούς 20-20 έως 16-30-30. Αυτό το συμπέρασμα συμφωνεί με την βιβλιογραφία (Guh και Hsieh [24]) και αιτιολογείται από το γεγονός ότι, η υπερβολική αύξηση του αριθμού των κρυμμένων επιπέδων και των κόμβων τους, παύει να ενισχύει την ικανότητα μάθησης και αυξάνει την πολυπλοκότητα, οδηγώντας σε μείωση της απόδοσης.

Επίσης από τον Πίνακα 6 προκύπτει ότι τα περισσότερα μοντέλα με απόλυτη ακρίβεια προέρχονται από την αρχιτεκτονική LM mse logsig με τους 150 συνδυασμούς κρυμμένων επιπέδων και κόμβων της. Μάλιστα τα μοντέλα με ακρίβεια 100% για την συγκεκριμένη

αρχιτεκτονική είναι συνολικά 7, με την αρχιτεκτονική LM mse logsig 20-30-30 να διαθέτει δύο μοντέλα με απόλυτη ακρίβεια.

Ακόμα από τον Πίνακα 5 προκύπτει ότι για την ακριβέστερη, από την άποψη της μέσης ακρίβειας, αρχιτεκτονική LM mse logsig 16-30-16, η κλάση Grouping με την μεγαλύτερη μέση ακρίβεια, έχει την χαμηλότερη μέση ανάκληση. Αυτό σημαίνει ότι για την κλάση αυτή, σε σχέση με τις υπόλοιπες, το ΤΝΔ καταχωρεί με μεγαλύτερη συχνότητα διανύσματα της σε άλλες κλάσεις, αλλά όταν προβλέπει ότι ένα διάνυσμα ανήκει σε αυτήν είναι σε μεγάλο βαθμό ακριβές. Η κλάση με την χαμηλότερη μέση ακρίβεια είναι η Point outside control limits και αυτές με την υψηλότερη μέση ανάκληση είναι οι Level shift και Monotonous trend.

Είναι σημαντικό να τονιστεί ότι δεν προκύπτει ούτε μία περίπτωση, που να αποδίδεται κάποιο διάνυσμα στην κλάση mixed trends. Αυτό σημαίνει ότι στο διάνυσμα εξόδου των ΤΝΔ πέραν του μέγιστου στοιχείου, το οποίο στρογγυλοποιείται στο 1, κανένα άλλο δεν υπερβαίνει το 0.5.

Από την σκοπιά του χρόνου υπολογισμού και με βάση τις παρατηρήσεις που έγιναν, μπορεί να ειπωθεί ότι αυτό που αυξάνει τις χρονικές απαιτήσεις είναι η αύξηση του συνολικού αριθμού των συντελεστών βαρύτητας, που επηρεάζεται από τα κρυμμένα επίπεδα και τους κόμβους τους. Για παράδειγμα στις αρχιτεκτονικές 30-30 και 20-20-20, όπου ο συνολικός αριθμός κόμβων είναι ο ίδιος, ταχύτερη είναι η δεύτερη. Αυτό είναι λογικό αφού ο αριθμός των συντελεστών βαρύτητας και κατά επέκταση του υπολογιστικού φορτίου της πρώτης αρχιτεκτονικής είναι 1620 $(16 * 30 + 30 * 30 + 30 * 8)$ ενώ της δεύτερης είναι 1280 $(16 * 20 + 20 * 20 + 20 * 20 + 20 * 8)$.

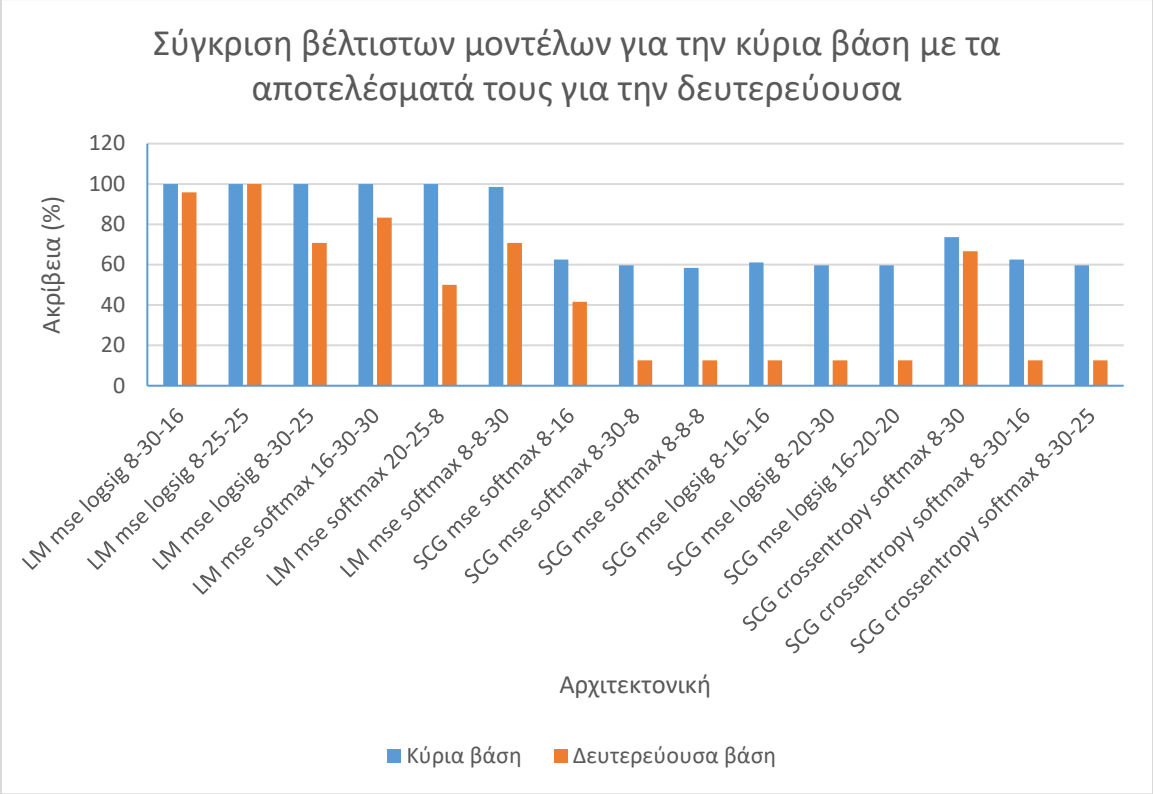
6.1.2 Σχολιασμός αποτελεσμάτων της δευτερεύουσας βάσης δεδομένων

Τα βέλτιστα μοντέλα, από την άποψη της μέσης ακρίβειας, που δοκιμάστηκαν με την δευτερεύουσα βάση δεδομένων, διατήρησαν σε γενικό επίπεδο, την ιεραρχία τους πράγμα που γίνεται αντιληπτό συνδυάζοντας τους Πίνακες 6 και 7. Αυτό σημαίνει ότι το μοντέλο ΤΝΔ με την μεγαλύτερη ακρίβεια, στην εφαρμογή με την κύρια βάση δεδομένων, διαθέτει την μεγαλύτερη ακρίβεια και με την δευτερεύουσα βάση. Το ίδιο συμβαίνει και με την σειρά της πλειοψηφίας των υπολοίπων.

Πίνακας 18: Τα βέλτιστα μοντέλα της κύριας βάσης δεδομένων και τα αποτελέσματά τους με την δευτερεύουσα.

Αρχιτεκτονική	Ακρίβεια με την κύρια βάση (%)	Ακρίβεια με την δευτερεύουσα βάση (%)
LM mse logsig 8-30-16	100	95.83
LM mse logsig 8-25-25	100	100
LM mse logsig 8-30-25	100	70.83
LM mse softmax 16-30-30	100	83.33
LM mse softmax 20-25-8	100	50
LM mse softmax 8-8-30	98.61	70.83
SCG mse softmax 8-16	62.5	41.67
SCG mse softmax 8-30-8	59.72	12.5
SCG mse softmax 8-8-8	58.33	12.5
SCG mse logsig 8-16-16	61.11	12.5
SCG mse logsig 8-20-30	59.72	12.5
SCG mse logsig 16-20-20	59.72	12.5
SCG crossentropy softmax 8-30	73.61	66.67
SCG crossentropy softmax 8-30-16	62.5	12.5
SCG crossentropy softmax 8-30-25	59.72	12.5

Παρακάτω παρατίθεται το συγκριτικό διάγραμμα.



Εικόνα 66: Σύγκριση βέλτιστων μοντέλων για την κύρια βάση δεδομένων με τα αποτελέσματά τους για την δευτερεύουσα.

Το γεγονός αυτό είναι μία θετική ένδειξη αξιοπιστίας των αποτελεσμάτων, με την κύρια βάση δεδομένων. Ωστόσο παρότι η ιεραρχία διατηρείται, η ακρίβεια μειώνεται στο σύνολο σχεδόν των περιπτώσεων. Συνεπώς η ικανότητα γενίκευσης φαίνεται να είναι επαρκής μόνο στο ένα τρίτο των περιπτώσεων.

Εξαίρεση αποτελεί, το βέλτιστο μοντέλο της αρχιτεκτονικής LM mse logsig 8-25-25, το οποίο κατάφερε να επιτύχει απόλυτη ακρίβεια τόσο με την κύρια όσο και με την δευτερεύουσα βάση δεδομένων.

6.2 KNN

6.2.1 Σχολιασμός αποτελεσμάτων της κύριας βάσης δεδομένων

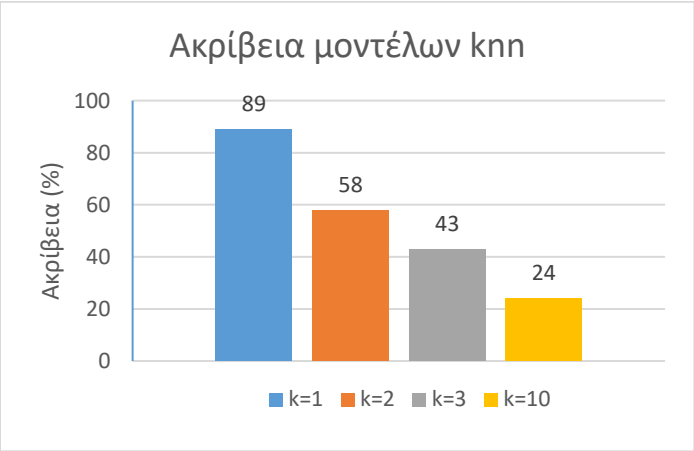
Η αύξηση της τιμής του k στον αλγόριθμο οδηγεί σε μείωση της απόδοσης, στην συγκεκριμένη εφαρμογή. Αξίζει να σημειωθεί ότι για κάθε τιμή του k η μέση ακρίβεια των μοντέλων knn, ταυτίζεται με την βέλτιστη. Αυτό συμβαίνει διότι για σταθερή τιμή του k και σταθερά διανύσματα εκπαίδευσης και ελέγχου, οι αποστάσεις που υπολογίζει ο αλγόριθμος (βλ. § 2.3 Αλγόριθμος k-Κοντινότερων Γειτόνων (k-nearest neighbors)) δεν

μεταβάλλονται και η κατηγοριοποίηση πραγματοποιείται με το ίδιο τρόπο, επιφέροντας το ίδιο αποτέλεσμα.

Για την τιμή αυτή της παραμέτρου k, η ακρίβεια και η ανάκληση κάθε κλάσης φαίνονται στην Εικόνα 57. Η κλάση με την μικρότερη ακρίβεια είναι η Monotonous trend και αυτές με την μικρότερη ανάκληση είναι οι Cycle pattern, Instability, Level Shift και Stratification. Η ανάκληση των υπολοίπων κλάσεων είναι μέγιστη (100%) ενώ μέγιστη ακρίβεια επιτυγχάνεται στις κλάσεις Exceptional point, Grouping, Instability, Level Shift και Point outside control limits.

Συμπερασματικά ο αλγόριθμος knn με k=1 αποδίδει, κάθε φορά πανομοιότυπο, υψηλής ακρίβειας αποτέλεσμα και μάλιστα ταχύτατα.

Πίνακας 19: Ακρίβεια μοντέλων knn με την κύρια βάση δεδομένων.



Το γεγονός αυτό καθιστά το μοντέλο του αλγορίθμου knn με k=1, το πιο αποτελεσματικό από οποιοδήποτε άλλο που εξετάστηκε. Ταυτόχρονα όμως, διεγείρει και το ερώτημα αν η αποτελεσματικότητά του αυτή συνδέεται με την συγκεκριμένη βάση δεομένων ή μπορεί να γενικευτεί. Γι' αυτό και δοκιμάζεται μαζί με τα τρία πιο ακριβή μοντέλα ΤΝΔ, με την δευτερεύουσα βάση δεδομένων και προκύπτουν τα παρακάτω συμπεράσματα.

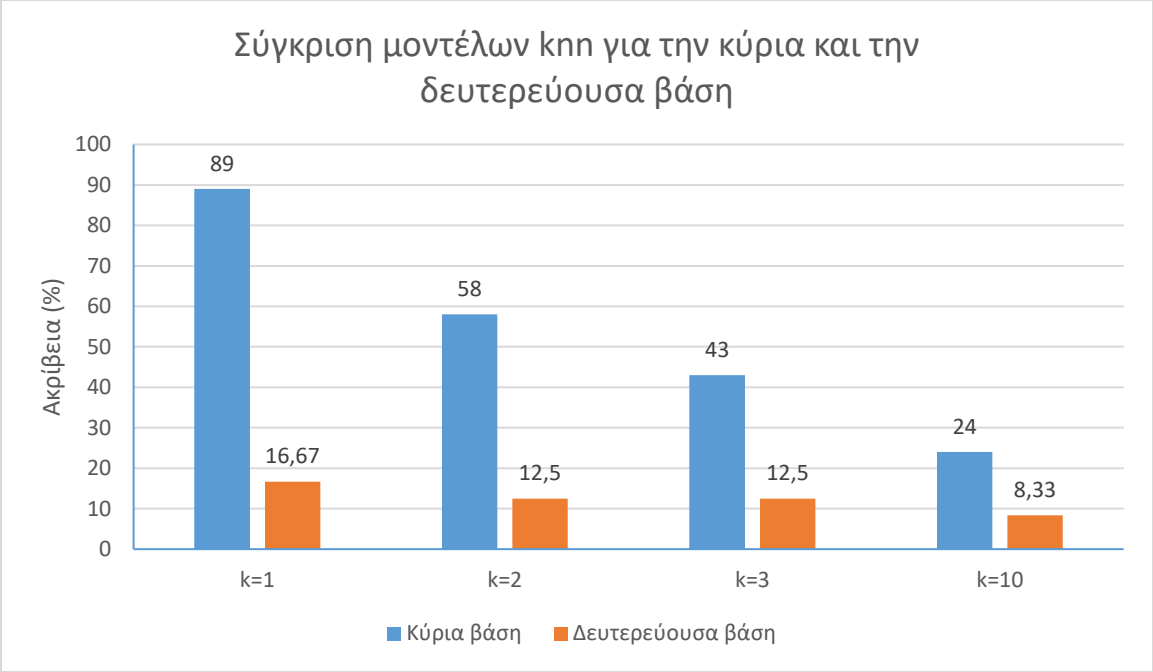
6.2.2 Σχολιασμός αποτελεσμάτων της δευτερεύουσας βάσης δεδομένων

Όπως φαίνεται από τον Πίνακα 20 η απόδοση του αλγορίθμου knn με αυτήν την βάση δεδομένων είναι διαμετρικά χαμηλότερη, σε σχέση με την προηγούμενη. Αυτό σημαίνει ότι η επιτυχία της πρώτης εφαρμογής οφείλεται στον τρόπο δημιουργίας και διαχωρισμού των διανυσμάτων εκπαίδευσης και ελέγχου και ότι δεν επεκτείνεται στην δεύτερη εφαρμογή, όπου οι αποστάσεις των τάσεων είναι διαφορετικές από αυτές με τις οποίες έχει εκπαιδευτεί ο αλγόριθμος να αναγνωρίζει την κάθε τάση. Άρα συμπεραίνεται ότι ο αλγόριθμος knn δεν διαθέτει δυνατότητα γενίκευσης, μεταξύ των συγκεκριμένων εφαρμογών.

Πίνακας 20: Σύγκριση ακρίβειας μοντέλων knn με την κύρια και την δευτερεύουσα βάση δεδομένων.

Μοντέλο	Τιμή k	Ακρίβεια με την κύρια βάση (%)	Ακρίβεια με την δευτερεύουσα βάση (%)	Διαφορά (%)
KNN	1	89	16.67	72.33
KNN	2	58	12.5	45.5
KNN	3	43	12.5	30.5
KNN	10	24	8.33	15.67
			Μέση διαφορά (%):	**41**

Παρακάτω παρατίθεται το αντίστοιχο διάγραμμα.



Εικόνα 67: Σύγκριση μοντέλων knn για την κύρια και την δευτερεύουσα βάση δεδομένων.

6.3 Σύγκριση ΤΝΔ knn και βιβλιογραφίας

6.3.1 ΤΝΔ και knn

Στον Πίνακα 18 φαίνεται ότι το ακριβέστερο από όλα τα μοντέλα ΤΝΔ ανήκει στην αρχιτεκτονική LM mse logsig 8-25-25 και επιτυγχάνει ακρίβεια 100% τόσο με την κύρια όσο και με την δευτερεύουσα βάση δεδομένων. Αντιστοίχως στον Πίνακα 20 αναγράφεται ότι το ακριβέστερο μοντέλο knn είναι αυτό με την τιμή της παραμέτρου k ίση με 1 και επιτυγχάνει ακρίβεια 89% με την κύρια και 16.67% με την δευτερεύουσα βάση δεδομένων. Συνεπώς το πιο επιτυχημένο μοντέλο ΤΝΔ επικρατεί του πιο επιτυχημένου μοντέλου knn τόσο στην συνολική ακρίβεια όσο και στην ακρίβεια ανά κλάση, με αμφότερες τις βάσεις δεδομένων. Αξίζει να σημειωθεί ότι, εξετάζοντας τον Πίνακα 18 προκύπτουν επιπλέον 5 μοντέλα ΤΝΔ, πέραν του ακριβέστερου, με ακρίβεια μεγαλύτερη από το συγκεκριμένο μοντέλο knn, με αμφότερες τις βάσεις δεδομένων. Επιπροσθέτως, προκύπτουν συνολικά 544 μοντέλα ΤΝΔ με ακρίβεια μεγαλύτερη από αυτήν του μοντέλου knn με k=1, με την κύρια βάση δεδομένων, εκ των οποίων τα 340 ανήκουν στην αρχιτεκτονική LM mse logsig και τα 204 στην αρχιτεκτονική LM mse softmax.

6.3.2 ΤΝΔ knn και βιβλιογραφία

Σε σύγκριση με την βιβλιογραφία, στην παρούσα εργασία, παρατηρούνται ομοιότητες αλλά και διαφοροποιήσεις.

Στην βιβλιογραφία υλοποιούνται κυρίως αρχιτεκτονικές με ένα κρυμμένο επίπεδο ([23], [24], [25], [26]) χωρίς να αποκλείονται όμως και εφαρμογές με περισσότερα από τρία [22]. Το πλήθος των δύο και τριών κρυμμένων επιπέδων που χρησιμοποιούνται στην παρούσα εργασία αποτελούν μία ενδιάμεση επιλογή. Το ίδιο ισχύει και για τον αριθμό των κόμβων των κρυμμένων επιπέδων, που στις αναφορές ποικίλει από 6 ([22], [25]) έως 35 [24] ενώ στην εργασία λαμβάνει τιμές από 8 έως 30.

Κύρια διαφοροποίηση αποτελεί το γεγονός ότι στην παρούσα εργασία ο αριθμός των κλάσεων είναι 8 ενώ στην πλειοψηφία των βιβλιογραφικών αναφορών είναι 4 έως 6 ([22], [23], [24], [25]). Όσο αυξάνεται ο αριθμός των κλάσεων αυξάνεται και η πολυπλοκότητα του προβλήματος, δυσχεραίνοντας την εκπαίδευση. Μία ακόμα σημαντική διαφορά είναι ότι, στην βιβλιογραφία πραγματοποιείται κατά κόρων μία εκπαίδευση για κάθε μοντέλο, ενώ στην παρούσα εργασία πραγματοποιούνται 7, επιφέροντας μικρά αλλά υπολογίσιμα αποτελέσματα στην αύξηση της απόδοσης. Ακόμα στις περισσότερες των περιπτώσεων χρησιμοποιείται ο αλγόριθμος οπισθόδρομης διάδοσης σφάλματος ([23], [24], [26]) ενώ σε λιγότερες επιχειρείται σύγκριση δύο η περισσοτέρων διαφορετικών αλγορίθμων [28].

Από την άλλη πλευρά κύρια ομοιότητα συνιστά το ότι, η πλειοψηφία των βιβλιογραφικών αναφορών πραγματεύεται με κατασκευασμένα διανύσματα εκπαίδευσης και ελέγχου, παρότι ο τρόπος κατασκευής τους διαφέρει, αφού στηρίζεται στην μέση τιμή και στην τυπική απόκλιση της διαδικασίας παραγωγής σε κατάσταση στατιστικού ελέγχου ([22], [23], [24], [25]). Τα διανύσματα εκπαίδευσης που χρησιμοποιούνται στην παρούσα εργασία είναι 216, αριθμός μικρότερος σε σχέση με

την βιβλιογραφία, όπου ποικίλει από 83 ([23], [25]) έως 5250 [26]. Τα διανύσματα εκπαίδευσης και ελέγχου αποτελούνται από 16 κόμβους, που είναι κατά κύριο λόγο λιγότεροι από αυτούς της βιβλιογραφίας ([24], [25]) χωρίς όμως αυτό να αποκλείει και την περίπτωση να είναι περισσότεροι ([23], [26]).

Επιπλέον, ταύτιση της εργασίας με τις αναφορές, διαπιστώνεται ως προς τις συναρτήσεις ενεργοποίησης του κρυμμένου επιπέδου και του επιπέδου εξόδου ([22] (, [23], [24]) με τις εξαιρέσεις να χρησιμοποιούν περισσότερο εξειδικευμένες συναρτήσεις ([25]).

Η ακρίβεια που επιτυγχάνεται στην βιβλιογραφία είναι το ίδιο υψηλή με την εργασία αυτήν. Μία χαρακτηριστική περίπτωση ομοιότητας είναι το 99% της ακρίβειας των Pham και Wani [23] αναγνωρίζοντας 4 κλάσεις με 83 διανύσματα εκπαίδευσης για κάθε μία. Ελαφρώς χαμηλότερο ποσοστό επιτυχίας, 94%, πετυχαίνουν οι Guh και Hsieh [24] με ισάριθμες κλάσεις και συνολικά 8000 διανύσματα ελέγχου. Οι δοκιμές των οι Pham και Sahran [25] τελεσφορούν με ακρίβεια 97.85% με 90 διανύσματα ελέγχου, για κάθε κατηγορία, ενώ οι Tunchanit Bunjongjit και Wimalin Laosiritaworn [22] επιτυγχάνουν το ποσοστό 91.73% με ΤΝΔ και 97.44% με μοντέλο knn με την παράμετρο k ίση με την μονάδα. Εδώ υπάρχει διαφοροποίηση με την παρούσα εργασία καθώς στην τελευταία, το ακριβέστερο μοντέλο ΤΝΔ υπερέχει του ακριβέστερου knn.

6.4 Προτάσεις για βελτίωση

Σε πρώτο επίπεδο αξιόλογη κρίνεται η προσπάθεια σύγκρισης των αποτελεσμάτων των αρχιτεκτονικών που δοκιμάστηκαν, με κάποιον διαφορετικό αλγόριθμο εκπαίδευσης, όπως για παράδειγμα ο stochastic gradient descent. Επίσης, ενδιαφέρον παρουσιάζει η εξέταση διαφορετικού εύρους αριθμού κρυμμένων επιπέδων και κόμβων.

Επιπλέον σημαντική επίδραση πιθανόν να έχει η εκπαίδευση με μεγαλύτερο αριθμό διανυσμάτων εκπαίδευσης, δηλαδή η διεύρυνση της βάσης δεδομένων. Ουσιαστική κρίνεται η εξαγωγή ορισμένων χαρακτηριστικών των τάσεων και η συμμετοχή τους στην εκπαίδευση.

Τέλος, εξαιτίας της μικρής αποτελεσματικότητας στην αναγνώριση μεικτών τάσεων (βλ. Παράρτημα Γ) ουσιαστική είναι η επενέργεια στην κατεύθυνση αύξησής της, που μπορεί να πραγματοποιηθεί είτε με την εκπαίδευση με διανύσματα μεικτών τάσεων είτε με την εξειδικευμένη εκπαίδευση περισσοτέρων ΤΝΔ και την σύνδεσή τους σε ένα μεγαλύτερο δίκτυο.

Βιβλιογραφία

[1] «Επιχειρησιακό σχέδιο παραγωγής φαρμακευτικών σκευασμάτων στο νοσοκομειακό φαρμακείο,» ΠΕΦΝΙ, 2012.

[2] Γ. Ν. Ταγαράς, Στατιστικός Έλεγχος Ποιότητας, Θεσσαλονίκη: Εκδόσεις ΖΗΤΗ, 2001.

[3] G. Jacobson, «KaiNexus,» 4 Feb 2021. [Ηλεκτρονικό]. Available: https://blog.kainexus.com/improvement-disciplines/lean/control-charts/an-introduction-to-process-control-charts.

[4] G. V. a. P. B. A. Avgikou, «SPC pattern recognition using LVQ neural nets,» National Technical University of Athens, School of Mechanical Engineering, Manufacturing Technology Division, GR-15780, Athens, Greece.

[5] «Wikipedia,» [Ηλεκτρονικό]. Available: https://en.wikipedia.org/wiki/Frank_Rosenblatt.

[6] «Wikipedia,» [Ηλεκτρονικό]. Available: https://en.wikipedia.org/wiki/Artificial_neural_network.

[7] Π. Ιωάννης, «Μελέτη γήρανσης μονώσεων με χρήση τεχνολογίας ασαφούς λογικής και νευρωνικών δικτύων,» Πάτρα.

[8] Σ. Δήμητρα, «Σύγκριση Μεθόδων Εκπαίδευσης Τεχνητών Νευρωνικών Δικτύων,» Πάτρα, 2010.

[9] Θ.-Κ. Λάουρα, «Τεχνητά Νευρωνικά Δίκτυα και εφαρμογές στα Συστήματα Αυτόματου Ελέγχου,» Πάτρα, 2013.

[10] L. Fausett, «Fundamentals of neural networks: architectures, algorithms, and applications.,» Prentice-Hall, Inc., USA, 1994.

[11] M. S. H. W. Kurt Hornik, «Multilayer feedforward networks are universal approximators.,» *Neural Networks,* pp. 359-366, 1989.

[12] Π. Στυλιανός, «Το νεύρο-ασαφές μοντέλο ANFIS, Πρόβλεψη τιμών συναλλάγματος με νευρο-ασαφή συστήματα,» Χανιά, 2014.

[13] Y. Gavrilova, «serokell,» [Ηλεκτρονικό]. Available: https://serokell.io/blog/classification-algorithms.

[14] «theoriesmathisi.weebly,» [Ηλεκτρονικό]. Available: https://theoriesmathisi.weebly.com/thorndike.html.

[15] D. Kapil, «medium» Available: https://medium.com/@divakar_239/stochastic-vs-batch-gradient-descent-8820568eada1.

[16] «missinglink.ai» Available: https://missinglink.ai/guides/neural-network-concepts/backpropagation-neural-networks-process-examples-code-minus-math/.

[17] A. Quesada, «neuraldesigner» Available: https://www.neuraldesigner.com/blog/5_algorithms_to_train_a_neural_network.

[18] K. E. Koech, «towardsdatascience» Available: https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e.

[19] «Wikipedia» Available: https://en.wikipedia.org/wiki/Gradient_descent.

[20] «stackexchange» Available: https://math.stackexchange.com/questions/1020008/why-is-the-conjugate-direction-better-than-the-negative-of-gradient-when-minimi.

[21] J. R. Shewchuk, «An Introduction to the Conjugate Gradient Method Without the Agonizing Pain» Pittsburgh, PA 15213, 1994.

[22] «morioh» Available: https://morioh.com/p/620a3a2faf6c.

[23] W. L. Tunchanit Bunjongjit, «Classification Techniques for Control Chart Pattern Recognition: A Case of Metal Frame for Actuator Production» *Journal of Science,* October 2013.

[24] M. A. Wani D.T. Pham, «Feature-based control chart pattern recognition» *International Journal of Production Research,* 1997.

[25] Y.-C. H. Ruey-Shy Guh, «A neural network based model for abnormal pattern recognition of control charts» Computers & Industrial Engineering, 1999.

[26] S. S. D.T. Pham, «Control Chart Pattern Recognition Using Spiking Neural Networks» Elsevier Ltd, Cardiff, 2006.

[27] Y. Al-Assaf, «Multi-Resolution Wavelets Analysis for the Recognition of Concurrent Control Chart Patterns» Taylor and Francis, Sharja, 2007.

[28] A. E. Vahid Ranaee, «Control chart patterns recognition using neural networks and efficient features: a comparative study» Iran, 2010.

[29] A. E. V. R. Jalil Addeh, «Control Chart Recognition Using Adaptive Back-Propagation Artificial Neural Networks and Efficient Features» Babol, 2011.

[30] C.-S. Cheng, «A neural network approach for the analysis of control chart patterns» Taylor & Francis, 1997.

[31] «wikipedia» Available: https://el.wikipedia.org/wiki/%CE%9C%CE%B7%CF%87%CE%B1%CE%BD%CE%B9%CE%BA%CE%AE_%CE%BC%CE%AC%CE%B8%CE%B7%CF%83%CE%B7.

[32] G. V. P. B. A. Avgikou, «SPC pattern recognition using LQV neural nets» GR-15780, Athens, Greece.

[33] «wikipedia» Available: https://en.wikipedia.org/wiki/Confusion_matrix.

[34] «missinglink.ai» Available: https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/.

[35] S. Sivarajkumar, «medium» Available: https://medium.com/@sonish.sivarajkumar/relu-most-popular-activation-function-for-deep-neural-networks-10160af37dda.

[36] «educationalresearchtechniques» Available: https://educationalresearchtechniques.com/2015/11/.

Παράρτημα Α

Πίνακας 21: Μέση και βέλτιστη ακρίβεια των 750 αρχιτεκτονικών.

Αρίθμηση	Αλγ. Εκπ.	Συν.Κόστους	Συν. Ενεργ. Επ. Εξόδου	Κόμβοι κρυμμένων επιπέδων	Μέση Ακρίβεια (%)	Βέλτιστη Ακρίβεια (%)
1	LM	mse	logsig	8-8-0	82.14	90.28
2	LM	mse	logsig	8-16-0	87.89	97.22
3	LM	mse	logsig	8-20-0	85.52	97.22
4	LM	mse	logsig	8-25-0	79.76	97.22
5	LM	mse	logsig	8-30-0	80.56	97.22
6	LM	mse	logsig	16-8-0	59.52	80.56
7	LM	mse	logsig	16-16-0	70.24	93.06
8	LM	mse	logsig	16-20-0	89.88	97.22
9	LM	mse	logsig	16-25-0	67.46	95.83
10	LM	mse	logsig	16-30-0	70.64	95.83
11	LM	mse	logsig	20-8-0	68.05	95.83
12	LM	mse	logsig	20-16-0	62.10	94.44
13	LM	mse	logsig	20-20-0	67.66	95.83
14	LM	mse	logsig	20-25-0	70.63	94.44
15	LM	mse	logsig	20-30-0	68.45	100.00
16	LM	mse	logsig	25-8-0	72.62	88.89
17	LM	mse	logsig	25-16-0	88.29	94.44
18	LM	mse	logsig	25-20-0	51.39	97.22
19	LM	mse	logsig	25-25-0	64.68	91.67
20	LM	mse	logsig	25-30-0	84.72	93.06
21	LM	mse	logsig	30-8-0	65.48	91.67
22	LM	mse	logsig	30-16-0	59.33	87.50

23	LM	mse	logsig	30-20-0	62.70	95.83
24	LM	mse	logsig	30-25-0	64.68	97.22
25	LM	mse	logsig	30-30-0	74.21	94.44
26	LM	mse	logsig	8-8-8	75.60	90.28
27	LM	mse	logsig	8-8-16	75.99	93.06
28	LM	mse	logsig	8-8-20	85.91	93.06
29	LM	mse	logsig	8-8-25	81.94	97.22
30	LM	mse	logsig	8-8-30	87.70	93.06
31	LM	mse	logsig	8-16-8	88.89	97.22
32	LM	mse	logsig	8-16-16	77.38	95.83
33	LM	mse	logsig	8-16-20	75.99	97.22
34	LM	mse	logsig	8-16-25	71.43	97.22
35	LM	mse	logsig	8-16-30	83.13	97.22
36	LM	mse	logsig	8-20-8	80.95	98.61
37	LM	mse	logsig	8-20-16	81.55	98.61
38	LM	mse	logsig	8-20-20	77.98	97.22
39	LM	mse	logsig	8-20-25	80.56	93.06
40	LM	mse	logsig	8-20-30	91.07	97.22
41	LM	mse	logsig	8-25-8	87.30	94.44
42	LM	mse	logsig	8-25-16	89.68	95.83
43	LM	mse	logsig	8-25-20	79.56	97.22
44	LM	mse	logsig	8-25-25	91.86	100.00
45	LM	mse	logsig	8-25-30	86.91	100.00
46	LM	mse	logsig	8-30-8	84.72	98.61
47	LM	mse	logsig	8-30-16	93.05	100.00
48	LM	mse	logsig	8-30-20	94.44	98.61

49	LM	mse	logsig	8-30-25	91.07	100.00
50	LM	mse	logsig	8-30-30	76.98	98.61
51	LM	mse	logsig	16-8-8	63.89	87.50
52	LM	mse	logsig	16-8-16	85.71	90.28
53	LM	mse	logsig	16-8-20	79.56	94.44
54	LM	mse	logsig	16-8-25	60.51	86.11
55	LM	mse	logsig	16-8-30	43.25	87.50
56	LM	mse	logsig	16-16-8	71.63	94.44
57	LM	mse	logsig	16-16-16	61.11	91.67
58	LM	mse	logsig	16-16-20	65.08	90.28
59	LM	mse	logsig	16-16-25	80.36	95.83
60	LM	mse	logsig	16-16-30	65.08	97.22
61	LM	mse	logsig	16-20-8	78.37	93.06
62	LM	mse	logsig	16-20-16	69.44	95.83
63	LM	mse	logsig	16-20-20	83.73	97.22
64	LM	mse	logsig	16-20-25	77.58	95.83
65	LM	mse	logsig	16-20-30	79.56	95.83
66	LM	mse	logsig	16-25-8	85.12	97.22
67	LM	mse	logsig	16-25-16	83.33	95.83
68	LM	mse	logsig	16-25-20	72.62	93.06
69	LM	mse	logsig	16-25-25	73.02	97.22
70	LM	mse	logsig	16-25-30	84.72	97.22
71	LM	mse	logsig	16-30-8	54.36	94.44
72	LM	mse	logsig	16-30-16	94.64	97.22
73	LM	mse	logsig	16-30-20	76.39	95.83
74	LM	mse	logsig	16-30-25	75.79	94.44

75	LM	mse	logsig	16-30-30	91.47	97.22
76	LM	mse	logsig	20-8-8	45.83	77.78
77	LM	mse	logsig	20-8-16	73.41	87.50
78	LM	mse	logsig	20-8-20	78.77	98.61
79	LM	mse	logsig	20-8-25	73.21	87.50
80	LM	mse	logsig	20-8-30	50.79	97.22
81	LM	mse	logsig	20-16-8	51.98	93.06
82	LM	mse	logsig	20-16-16	64.88	93.06
83	LM	mse	logsig	20-16-20	80.16	98.61
84	LM	mse	logsig	20-16-25	58.14	91.67
85	LM	mse	logsig	20-16-30	52.38	97.22
86	LM	mse	logsig	20-20-8	76.59	93.06
87	LM	mse	logsig	20-20-16	78.17	90.28
88	LM	mse	logsig	20-20-20	72.82	94.44
89	LM	mse	logsig	20-20-25	55.16	90.28
90	LM	mse	logsig	20-20-30	55.75	93.06
91	LM	mse	logsig	20-25-8	79.76	94.44
92	LM	mse	logsig	20-25-16	78.97	95.83
93	LM	mse	logsig	20-25-20	65.87	94.44
94	LM	mse	logsig	20-25-25	73.81	87.50
95	LM	mse	logsig	20-25-30	82.54	95.83
96	LM	mse	logsig	20-30-8	88.29	97.22
97	LM	mse	logsig	20-30-16	65.48	94.44
98	LM	mse	logsig	20-30-20	73.41	95.83
99	LM	mse	logsig	20-30-25	58.14	93.06
100	LM	mse	logsig	20-30-30	77.97	100.00

101	LM	mse	logsig	25-8-8	70.04	84.72
102	LM	mse	logsig	25-8-16	70.04	88.89
103	LM	mse	logsig	25-8-20	67.46	90.28
104	LM	mse	logsig	25-8-25	56.95	91.67
105	LM	mse	logsig	25-8-30	60.71	87.50
106	LM	mse	logsig	25-16-8	67.66	94.44
107	LM	mse	logsig	25-16-16	63.69	94.44
108	LM	mse	logsig	25-16-20	63.89	88.89
109	LM	mse	logsig	25-16-25	66.47	90.28
110	LM	mse	logsig	25-16-30	73.02	97.22
111	LM	mse	logsig	25-20-8	70.64	93.06
112	LM	mse	logsig	25-20-16	78.77	95.83
113	LM	mse	logsig	25-20-20	69.05	93.06
114	LM	mse	logsig	25-20-25	77.78	94.44
115	LM	mse	logsig	25-20-30	80.75	91.67
116	LM	mse	logsig	25-25-8	73.21	94.44
117	LM	mse	logsig	25-25-16	85.91	94.44
118	LM	mse	logsig	25-25-20	90.08	98.61
119	LM	mse	logsig	25-25-25	59.72	97.22
120	LM	mse	logsig	25-25-30	67.86	90.28
121	LM	mse	logsig	25-30-8	66.07	90.28
122	LM	mse	logsig	25-30-16	68.85	91.67
123	LM	mse	logsig	25-30-20	62.50	91.67
124	LM	mse	logsig	25-30-25	60.71	95.83
125	LM	mse	logsig	25-30-30	75.20	91.67
126	LM	mse	logsig	30-8-8	62.30	84.72

127	LM	mse	logsig	30-8-16	60.52	94.44
128	LM	mse	logsig	30-8-20	56.55	81.94
129	LM	mse	logsig	30-8-25	74.01	90.28
130	LM	mse	logsig	30-8-30	60.51	84.72
131	LM	mse	logsig	30-16-8	74.21	95.83
132	LM	mse	logsig	30-16-16	64.68	93.06
133	LM	mse	logsig	30-16-20	70.04	90.28
134	LM	mse	logsig	30-16-25	58.13	90.28
135	LM	mse	logsig	30-16-30	52.78	93.06
136	LM	mse	logsig	30-20-8	48.81	87.50
137	LM	mse	logsig	30-20-16	63.69	91.67
138	LM	mse	logsig	30-20-20	78.97	94.44
139	LM	mse	logsig	30-20-25	65.48	94.44
140	LM	mse	logsig	30-20-30	52.38	83.33
141	LM	mse	logsig	30-25-8	62.10	91.67
142	LM	mse	logsig	30-25-16	60.91	94.44
143	LM	mse	logsig	30-25-20	65.68	94.44
144	LM	mse	logsig	30-25-25	69.05	94.44
145	LM	mse	logsig	30-25-30	57.14	86.11
146	LM	mse	logsig	30-30-8	77.98	95.83
147	LM	mse	logsig	30-30-16	78.18	91.67
148	LM	mse	logsig	30-30-20	77.38	95.83
149	LM	mse	logsig	30-30-25	62.30	97.22
150	LM	mse	logsig	30-30-30	88.29	95.83
151	LM	mse	softmax	8-8-0	78.77	94.44
152	LM	mse	softmax	8-16-0	89.48	97.22

153	LM	mse	softmax	8-20-0	80.16	97.22
154	LM	mse	softmax	8-25-0	77.18	94.44
155	LM	mse	softmax	8-30-0	84.72	97.22
156	LM	mse	softmax	16-8-0	72.22	94.44
157	LM	mse	softmax	16-16-0	66.67	91.67
158	LM	mse	softmax	16-20-0	75.59	95.83
159	LM	mse	softmax	16-25-0	83.93	95.83
160	LM	mse	softmax	16-30-0	86.51	91.67
161	LM	mse	softmax	20-8-0	68.26	91.67
162	LM	mse	softmax	20-16-0	86.90	95.83
163	LM	mse	softmax	20-20-0	63.69	90.28
164	LM	mse	softmax	20-25-0	79.17	91.67
165	LM	mse	softmax	20-30-0	77.78	91.67
166	LM	mse	softmax	25-8-0	63.69	83.33
167	LM	mse	softmax	25-16-0	59.13	93.06
168	LM	mse	softmax	25-20-0	56.55	79.17
169	LM	mse	softmax	25-25-0	83.73	93.06
170	LM	mse	softmax	25-30-0	82.34	97.22
171	LM	mse	softmax	30-8-0	72.02	88.89
172	LM	mse	softmax	30-16-0	77.58	93.06
173	LM	mse	softmax	30-20-0	79.36	90.28
174	LM	mse	softmax	30-25-0	69.05	84.72
175	LM	mse	softmax	30-30-0	72.62	91.67
176	LM	mse	softmax	8-8-8	66.27	88.89
177	LM	mse	softmax	8-8-16	79.96	95.83
178	LM	mse	softmax	8-8-20	78.77	93.06

179	LM	mse	softmax	8-8-25	85.12	93.06
180	LM	mse	softmax	8-8-30	71.23	98.61
181	LM	mse	softmax	8-16-8	78.97	94.44
182	LM	mse	softmax	8-16-16	77.78	97.22
183	LM	mse	softmax	8-16-20	87.30	94.44
184	LM	mse	softmax	8-16-25	75.20	93.06
185	LM	mse	softmax	8-16-30	66.67	94.44
186	LM	mse	softmax	8-20-8	72.62	98.61
187	LM	mse	softmax	8-20-16	91.47	97.22
188	LM	mse	softmax	8-20-20	92.86	95.83
189	LM	mse	softmax	8-20-25	91.47	97.22
190	LM	mse	softmax	8-20-30	93.25	97.22
191	LM	mse	softmax	8-25-8	59.52	95.83
192	LM	mse	softmax	8-25-16	88.89	97.22
193	LM	mse	softmax	8-25-20	93.25	97.22
194	LM	mse	softmax	8-25-25	85.52	97.22
195	LM	mse	softmax	8-25-30	91.87	98.61
196	LM	mse	softmax	8-30-8	62.50	95.83
197	LM	mse	softmax	8-30-16	74.01	97.22
198	LM	mse	softmax	8-30-20	91.67	95.83
199	LM	mse	softmax	8-30-25	73.81	93.06
200	LM	mse	softmax	8-30-30	82.34	93.06
201	LM	mse	softmax	16-8-8	50.59	81.94
202	LM	mse	softmax	16-8-16	60.52	93.06
203	LM	mse	softmax	16-8-20	78.37	88.89
204	LM	mse	softmax	16-8-25	73.41	97.22

205	LM	mse	softmax	16-8-30	73.41	98.61
206	LM	mse	softmax	16-16-8	56.75	90.28
207	LM	mse	softmax	16-16-16	63.30	91.67
208	LM	mse	softmax	16-16-20	63.10	93.06
209	LM	mse	softmax	16-16-25	66.67	91.67
210	LM	mse	softmax	16-16-30	89.29	97.22
211	LM	mse	softmax	16-20-8	64.09	86.11
212	LM	mse	softmax	16-20-16	71.43	93.06
213	LM	mse	softmax	16-20-20	88.49	95.83
214	LM	mse	softmax	16-20-25	77.78	97.22
215	LM	mse	softmax	16-20-30	87.70	93.06
216	LM	mse	softmax	16-25-8	64.68	80.56
217	LM	mse	softmax	16-25-16	72.02	97.22
218	LM	mse	softmax	16-25-20	57.14	95.83
219	LM	mse	softmax	16-25-25	62.10	93.06
220	LM	mse	softmax	16-25-30	85.12	97.22
221	LM	mse	softmax	16-30-8	65.87	91.67
222	LM	mse	softmax	16-30-16	67.66	88.89
223	LM	mse	softmax	16-30-20	68.85	94.44
224	LM	mse	softmax	16-30-25	68.06	90.28
225	LM	mse	softmax	16-30-30	92.86	100.00
226	LM	mse	softmax	20-8-8	61.11	90.28
227	LM	mse	softmax	20-8-16	55.95	83.33
228	LM	mse	softmax	20-8-20	53.57	80.56
229	LM	mse	softmax	20-8-25	67.46	83.33
230	LM	mse	softmax	20-8-30	80.75	91.67

231	LM	mse	softmax	20-16-8	67.26	93.06
232	LM	mse	softmax	20-16-16	70.04	91.67
233	LM	mse	softmax	20-16-20	71.63	93.06
234	LM	mse	softmax	20-16-25	54.17	80.56
235	LM	mse	softmax	20-16-30	77.58	91.67
236	LM	mse	softmax	20-20-8	66.27	88.89
237	LM	mse	softmax	20-20-16	74.01	94.44
238	LM	mse	softmax	20-20-20	67.66	91.67
239	LM	mse	softmax	20-20-25	62.50	84.72
240	LM	mse	softmax	20-20-30	76.19	88.89
241	LM	mse	softmax	20-25-8	67.86	100.00
242	LM	mse	softmax	20-25-16	75.79	93.06
243	LM	mse	softmax	20-25-20	57.54	80.56
244	LM	mse	softmax	20-25-25	78.37	94.44
245	LM	mse	softmax	20-25-30	73.81	91.67
246	LM	mse	softmax	20-30-8	67.46	97.22
247	LM	mse	softmax	20-30-16	63.09	90.28
248	LM	mse	softmax	20-30-20	48.41	86.11
249	LM	mse	softmax	20-30-25	82.54	90.28
250	LM	mse	softmax	20-30-30	73.61	95.83
251	LM	mse	softmax	25-8-8	57.14	94.44
252	LM	mse	softmax	25-8-16	62.90	95.83
253	LM	mse	softmax	25-8-20	44.44	52.78
254	LM	mse	softmax	25-8-25	67.46	84.72
255	LM	mse	softmax	25-8-30	60.12	88.89
256	LM	mse	softmax	25-16-8	43.85	66.67

257	LM	mse	softmax	25-16-16	49.60	75.00
258	LM	mse	softmax	25-16-20	68.06	81.94
259	LM	mse	softmax	25-16-25	64.88	88.89
260	LM	mse	softmax	25-16-30	68.45	88.89
261	LM	mse	softmax	25-20-8	65.08	90.28
262	LM	mse	softmax	25-20-16	67.26	88.89
263	LM	mse	softmax	25-20-20	70.83	90.28
264	LM	mse	softmax	25-20-25	55.36	93.06
265	LM	mse	softmax	25-20-30	81.35	93.06
266	LM	mse	softmax	25-25-8	54.56	88.89
267	LM	mse	softmax	25-25-16	76.19	93.06
268	LM	mse	softmax	25-25-20	67.86	80.56
269	LM	mse	softmax	25-25-25	67.46	91.67
270	LM	mse	softmax	25-25-30	72.42	91.67
271	LM	mse	softmax	25-30-8	59.72	86.11
272	LM	mse	softmax	25-30-16	66.07	80.56
273	LM	mse	softmax	25-30-20	53.37	76.39
274	LM	mse	softmax	25-30-25	77.18	87.50
275	LM	mse	softmax	25-30-30	73.22	88.89
276	LM	mse	softmax	30-8-8	48.22	80.56
277	LM	mse	softmax	30-8-16	61.11	91.67
278	LM	mse	softmax	30-8-20	60.52	76.39
279	LM	mse	softmax	30-8-25	57.34	97.22
280	LM	mse	softmax	30-8-30	64.09	88.89
281	LM	mse	softmax	30-16-8	56.74	83.33
282	LM	mse	softmax	30-16-16	59.52	83.33

283	LM	mse	softmax	30-16-20	66.67	88.89
284	LM	mse	softmax	30-16-25	55.56	84.72
285	LM	mse	softmax	30-16-30	67.46	91.67
286	LM	mse	softmax	30-20-8	67.06	88.89
287	LM	mse	softmax	30-20-16	62.90	81.94
288	LM	mse	softmax	30-20-20	57.94	84.72
289	LM	mse	softmax	30-20-25	62.10	75.00
290	LM	mse	softmax	30-20-30	61.11	76.39
291	LM	mse	softmax	30-25-8	59.72	75.00
292	LM	mse	softmax	30-25-16	61.70	88.89
293	LM	mse	softmax	30-25-20	62.30	90.28
294	LM	mse	softmax	30-25-25	53.37	77.78
295	LM	mse	softmax	30-25-30	62.90	83.33
296	LM	mse	softmax	30-30-8	40.87	68.06
297	LM	mse	softmax	30-30-16	65.08	91.67
298	LM	mse	softmax	30-30-20	65.48	87.50
299	LM	mse	softmax	30-30-25	68.25	90.28
300	LM	mse	softmax	30-30-30	69.44	83.33
301	SCG	mse	softmax	8-8-0	31.15	37.50
302	SCG	mse	softmax	8-16-0	31.75	62.50
303	SCG	mse	softmax	8-20-0	32.14	44.44
304	SCG	mse	softmax	8-25-0	30.36	43.06
305	SCG	mse	softmax	8-30-0	34.13	50.00
306	SCG	mse	softmax	16-8-0	26.59	31.94
307	SCG	mse	softmax	16-16-0	33.73	52.78
308	SCG	mse	softmax	16-20-0	28.57	43.06
309	SCG	mse	softmax	16-25-0	28.37	36.11
310	SCG	mse	softmax	16-30-0	30.55	50.00

311	SCG	mse	softmax	20-8-0	28.77	36.11
312	SCG	mse	softmax	20-16-0	27.78	36.11
313	SCG	mse	softmax	20-20-0	24.21	34.72
314	SCG	mse	softmax	20-25-0	31.94	52.78
315	SCG	mse	softmax	20-30-0	24.21	34.72
316	SCG	mse	softmax	25-8-0	23.41	38.89
317	SCG	mse	softmax	25-16-0	27.78	48.61
318	SCG	mse	softmax	25-20-0	23.81	31.94
319	SCG	mse	softmax	25-25-0	25.99	36.11
320	SCG	mse	softmax	25-30-0	22.62	40.28
321	SCG	mse	softmax	30-8-0	24.21	41.67
322	SCG	mse	softmax	30-16-0	28.97	37.50
323	SCG	mse	softmax	30-20-0	29.17	40.28
324	SCG	mse	softmax	30-25-0	18.65	23.61
325	SCG	mse	softmax	30-30-0	28.77	37.50
326	SCG	mse	softmax	8-8-8	31.35	58.33
327	SCG	mse	softmax	8-8-16	30.36	50.00
328	SCG	mse	softmax	8-8-20	30.56	41.67
329	SCG	mse	softmax	8-8-25	24.21	37.50
330	SCG	mse	softmax	8-8-30	29.36	38.89
331	SCG	mse	softmax	8-16-8	27.58	36.11
332	SCG	mse	softmax	8-16-16	27.38	34.72
333	SCG	mse	softmax	8-16-20	29.17	43.06
334	SCG	mse	softmax	8-16-25	26.19	36.11
335	SCG	mse	softmax	8-16-30	23.61	44.44
336	SCG	mse	softmax	8-20-8	28.37	41.67
337	SCG	mse	softmax	8-20-16	30.95	52.78
338	SCG	mse	softmax	8-20-20	29.17	50.00
339	SCG	mse	softmax	8-20-25	27.78	40.28
340	SCG	mse	softmax	8-20-30	30.75	45.83

341	SCG	mse	softmax	8-25-8	31.94	48.61
342	SCG	mse	softmax	8-25-16	27.78	45.83
343	SCG	mse	softmax	8-25-20	29.76	45.83
344	SCG	mse	softmax	8-25-25	28.18	40.28
345	SCG	mse	softmax	8-25-30	36.51	43.06
346	SCG	mse	softmax	8-30-8	33.14	59.72
347	SCG	mse	softmax	8-30-16	27.97	36.11
348	SCG	mse	softmax	8-30-20	25.00	36.11
349	SCG	mse	softmax	8-30-25	27.98	40.28
350	SCG	mse	softmax	8-30-30	32.54	54.17
351	SCG	mse	softmax	16-8-8	29.96	38.89
352	SCG	mse	softmax	16-8-16	31.35	48.61
353	SCG	mse	softmax	16-8-20	30.55	47.22
354	SCG	mse	softmax	16-8-25	24.80	36.11
355	SCG	mse	softmax	16-8-30	24.80	36.11
356	SCG	mse	softmax	16-16-8	30.56	37.50
357	SCG	mse	softmax	16-16-16	27.78	54.17
358	SCG	mse	softmax	16-16-20	29.96	40.28
359	SCG	mse	softmax	16-16-25	26.19	47.22
360	SCG	mse	softmax	16-16-30	26.19	40.28
361	SCG	mse	softmax	16-20-8	27.97	34.72
362	SCG	mse	softmax	16-20-16	33.13	50.00
363	SCG	mse	softmax	16-20-20	26.59	47.22
364	SCG	mse	softmax	16-20-25	27.78	43.06
365	SCG	mse	softmax	16-20-30	29.56	44.44
366	SCG	mse	softmax	16-25-8	30.36	43.06
367	SCG	mse	softmax	16-25-16	32.14	45.83
368	SCG	mse	softmax	16-25-20	28.77	38.89
369	SCG	mse	softmax	16-25-25	21.82	36.11
370	SCG	mse	softmax	16-25-30	28.37	38.89

371	SCG	mse	softmax	16-30-8	30.95	43.06
372	SCG	mse	softmax	16-30-16	24.21	27.78
373	SCG	mse	softmax	16-30-20	28.97	41.67
374	SCG	mse	softmax	16-30-25	28.97	40.28
375	SCG	mse	softmax	16-30-30	28.37	34.72
376	SCG	mse	softmax	20-8-8	26.79	40.28
377	SCG	mse	softmax	20-8-16	26.39	37.50
378	SCG	mse	softmax	20-8-20	23.61	34.72
379	SCG	mse	softmax	20-8-25	23.21	31.94
380	SCG	mse	softmax	20-8-30	24.01	33.33
381	SCG	mse	softmax	20-16-8	30.75	47.22
382	SCG	mse	softmax	20-16-16	27.38	33.33
383	SCG	mse	softmax	20-16-20	20.83	30.56
384	SCG	mse	softmax	20-16-25	22.42	33.33
385	SCG	mse	softmax	20-16-30	29.36	43.06
386	SCG	mse	softmax	20-20-8	26.59	36.11
387	SCG	mse	softmax	20-20-16	29.37	43.06
388	SCG	mse	softmax	20-20-20	23.02	40.28
389	SCG	mse	softmax	20-20-25	30.75	41.67
390	SCG	mse	softmax	20-20-30	26.78	34.72
391	SCG	mse	softmax	20-25-8	27.58	43.06
392	SCG	mse	softmax	20-25-16	26.79	37.50
393	SCG	mse	softmax	20-25-20	26.79	34.72
394	SCG	mse	softmax	20-25-25	32.54	48.61
395	SCG	mse	softmax	20-25-30	26.99	31.94
396	SCG	mse	softmax	20-30-8	26.59	34.72
397	SCG	mse	softmax	20-30-16	19.05	34.72
398	SCG	mse	softmax	20-30-20	28.97	37.50
399	SCG	mse	softmax	20-30-25	31.15	41.67
400	SCG	mse	softmax	20-30-30	22.22	31.94

401	SCG	mse	softmax	25-8-8	22.82	31.94
402	SCG	mse	softmax	25-8-16	30.56	52.78
403	SCG	mse	softmax	25-8-20	21.23	29.17
404	SCG	mse	softmax	25-8-25	20.24	26.39
405	SCG	mse	softmax	25-8-30	25.40	36.11
406	SCG	mse	softmax	25-16-8	26.59	37.50
407	SCG	mse	softmax	25-16-16	20.64	27.78
408	SCG	mse	softmax	25-16-20	27.78	36.11
409	SCG	mse	softmax	25-16-25	27.18	37.50
410	SCG	mse	softmax	25-16-30	27.78	52.78
411	SCG	mse	softmax	25-20-8	26.98	45.83
412	SCG	mse	softmax	25-20-16	25.79	37.50
413	SCG	mse	softmax	25-20-20	21.03	25.00
414	SCG	mse	softmax	25-20-25	24.21	36.11
415	SCG	mse	softmax	25-20-30	27.18	37.50
416	SCG	mse	softmax	25-25-8	20.64	30.56
417	SCG	mse	softmax	25-25-16	25.00	37.50
418	SCG	mse	softmax	25-25-20	27.38	44.44
419	SCG	mse	softmax	25-25-25	19.84	25.00
420	SCG	mse	softmax	25-25-30	23.61	31.94
421	SCG	mse	softmax	25-30-8	26.79	31.94
422	SCG	mse	softmax	25-30-16	28.97	37.50
423	SCG	mse	softmax	25-30-20	23.21	33.33
424	SCG	mse	softmax	25-30-25	26.98	44.44
425	SCG	mse	softmax	25-30-30	23.61	31.94
426	SCG	mse	softmax	30-8-8	25.99	36.11
427	SCG	mse	softmax	30-8-16	26.39	44.44
428	SCG	mse	softmax	30-8-20	28.77	43.06
429	SCG	mse	softmax	30-8-25	27.38	38.89
430	SCG	mse	softmax	30-8-30	23.81	34.72

431	SCG	mse	softmax	30-16-8	28.57	43.06
432	SCG	mse	softmax	30-16-16	28.97	37.50
433	SCG	mse	softmax	30-16-20	21.63	27.78
434	SCG	mse	softmax	30-16-25	22.22	27.78
435	SCG	mse	softmax	30-16-30	20.63	33.33
436	SCG	mse	softmax	30-20-8	30.55	38.89
437	SCG	mse	softmax	30-20-16	22.02	29.17
438	SCG	mse	softmax	30-20-20	25.20	33.33
439	SCG	mse	softmax	30-20-25	28.37	37.50
440	SCG	mse	softmax	30-20-30	27.38	48.61
441	SCG	mse	softmax	30-25-8	25.20	36.11
442	SCG	mse	softmax	30-25-16	24.60	37.50
443	SCG	mse	softmax	30-25-20	23.61	41.67
444	SCG	mse	softmax	30-25-25	31.15	50.00
445	SCG	mse	softmax	30-25-30	20.04	29.17
446	SCG	mse	softmax	30-30-8	23.21	34.72
447	SCG	mse	softmax	30-30-16	25.99	40.28
448	SCG	mse	softmax	30-30-20	20.24	31.94
449	SCG	mse	softmax	30-30-25	24.60	36.11
450	SCG	mse	softmax	30-30-30	22.62	38.89
451	SCG	mse	logsig	8-8-0	32.14	51.39
452	SCG	mse	logsig	8-16-0	27.98	36.11
453	SCG	mse	logsig	8-20-0	28.57	44.44
454	SCG	mse	logsig	8-25-0	30.35	44.44
455	SCG	mse	logsig	8-30-0	35.91	52.78
456	SCG	mse	logsig	16-8-0	32.14	41.67
457	SCG	mse	logsig	16-16-0	29.37	38.89
458	SCG	mse	logsig	16-20-0	27.38	37.50
459	SCG	mse	logsig	16-25-0	34.32	56.94
460	SCG	mse	logsig	16-30-0	33.33	40.28

461	SCG	mse	logsig	20-8-0	23.81	47.22
462	SCG	mse	logsig	20-16-0	37.10	51.39
463	SCG	mse	logsig	20-20-0	31.15	41.67
464	SCG	mse	logsig	20-25-0	33.53	51.39
465	SCG	mse	logsig	20-30-0	32.74	47.22
466	SCG	mse	logsig	25-8-0	28.37	41.67
467	SCG	mse	logsig	25-16-0	27.58	38.89
468	SCG	mse	logsig	25-20-0	32.14	52.78
469	SCG	mse	logsig	25-25-0	32.74	43.06
470	SCG	mse	logsig	25-30-0	34.72	47.22
471	SCG	mse	logsig	30-8-0	31.55	40.28
472	SCG	mse	logsig	30-16-0	28.57	41.67
473	SCG	mse	logsig	30-20-0	29.96	43.06
474	SCG	mse	logsig	30-25-0	33.13	41.67
475	SCG	mse	logsig	30-30-0	32.14	41.67
476	SCG	mse	logsig	8-8-8	26.79	52.78
477	SCG	mse	logsig	8-8-16	33.93	47.22
478	SCG	mse	logsig	8-8-20	32.74	54.17
479	SCG	mse	logsig	8-8-25	34.92	54.17
480	SCG	mse	logsig	8-8-30	39.29	56.94
481	SCG	mse	logsig	8-16-8	30.36	51.39
482	SCG	mse	logsig	8-16-16	31.15	61.11
483	SCG	mse	logsig	8-16-20	33.14	52.78
484	SCG	mse	logsig	8-16-25	26.79	43.06
485	SCG	mse	logsig	8-16-30	31.15	44.44
486	SCG	mse	logsig	8-20-8	26.59	31.94
487	SCG	mse	logsig	8-20-16	30.56	51.39
488	SCG	mse	logsig	8-20-20	36.90	51.39
489	SCG	mse	logsig	8-20-25	34.92	48.61
490	SCG	mse	logsig	8-20-30	39.48	59.72

491	SCG	mse	logsig	8-25-8	26.59	34.72
492	SCG	mse	logsig	8-25-16	42.46	52.78
493	SCG	mse	logsig	8-25-20	36.91	48.61
494	SCG	mse	logsig	8-25-25	34.13	54.17
495	SCG	mse	logsig	8-25-30	36.51	51.39
496	SCG	mse	logsig	8-30-8	25.60	37.50
497	SCG	mse	logsig	8-30-16	36.91	52.78
498	SCG	mse	logsig	8-30-20	28.17	48.61
499	SCG	mse	logsig	8-30-25	41.27	52.78
500	SCG	mse	logsig	8-30-30	36.71	51.39
501	SCG	mse	logsig	16-8-8	26.99	40.28
502	SCG	mse	logsig	16-8-16	28.77	43.06
503	SCG	mse	logsig	16-8-20	28.77	38.89
504	SCG	mse	logsig	16-8-25	27.38	38.89
505	SCG	mse	logsig	16-8-30	39.48	47.22
506	SCG	mse	logsig	16-16-8	27.98	41.67
507	SCG	mse	logsig	16-16-16	30.75	37.50
508	SCG	mse	logsig	16-16-20	33.93	45.83
509	SCG	mse	logsig	16-16-25	36.11	55.56
510	SCG	mse	logsig	16-16-30	30.55	44.44
511	SCG	mse	logsig	16-20-8	32.14	47.22
512	SCG	mse	logsig	16-20-16	37.90	50.00
513	SCG	mse	logsig	16-20-20	39.28	59.72
514	SCG	mse	logsig	16-20-25	27.58	33.33
515	SCG	mse	logsig	16-20-30	28.57	40.28
516	SCG	mse	logsig	16-25-8	29.56	41.67
517	SCG	mse	logsig	16-25-16	30.76	43.06
518	SCG	mse	logsig	16-25-20	33.73	47.22
519	SCG	mse	logsig	16-25-25	39.29	50.00
520	SCG	mse	logsig	16-25-30	31.75	43.06

521	SCG	mse	logsig	16-30-8	30.16	40.28
522	SCG	mse	logsig	16-30-16	30.36	45.83
523	SCG	mse	logsig	16-30-20	27.98	37.50
524	SCG	mse	logsig	16-30-25	33.33	51.39
525	SCG	mse	logsig	16-30-30	25.99	33.33
526	SCG	mse	logsig	20-8-8	31.94	44.44
527	SCG	mse	logsig	20-8-16	27.58	45.83
528	SCG	mse	logsig	20-8-20	25.00	38.89
529	SCG	mse	logsig	20-8-25	26.59	36.11
530	SCG	mse	logsig	20-8-30	33.14	41.67
531	SCG	mse	logsig	20-16-8	38.69	48.61
532	SCG	mse	logsig	20-16-16	35.12	59.72
533	SCG	mse	logsig	20-16-20	35.52	58.33
534	SCG	mse	logsig	20-16-25	33.53	52.78
535	SCG	mse	logsig	20-16-30	34.13	44.44
536	SCG	mse	logsig	20-20-8	26.79	36.11
537	SCG	mse	logsig	20-20-16	35.91	45.83
538	SCG	mse	logsig	20-20-20	34.72	51.39
539	SCG	mse	logsig	20-20-25	34.92	44.44
540	SCG	mse	logsig	20-20-30	34.92	45.83
541	SCG	mse	logsig	20-25-8	30.56	40.28
542	SCG	mse	logsig	20-25-16	32.74	48.61
543	SCG	mse	logsig	20-25-20	36.71	43.06
544	SCG	mse	logsig	20-25-25	38.89	51.39
545	SCG	mse	logsig	20-25-30	32.54	48.61
546	SCG	mse	logsig	20-30-8	29.56	40.28
547	SCG	mse	logsig	20-30-16	32.54	44.44
548	SCG	mse	logsig	20-30-20	32.74	45.83
549	SCG	mse	logsig	20-30-25	35.52	45.83
550	SCG	mse	logsig	20-30-30	30.16	40.28

551	SCG	mse	logsig	25-8-8	29.56	44.44
552	SCG	mse	logsig	25-8-16	26.99	40.28
553	SCG	mse	logsig	25-8-20	33.14	50.00
554	SCG	mse	logsig	25-8-25	28.77	40.28
555	SCG	mse	logsig	25-8-30	27.98	41.67
556	SCG	mse	logsig	25-16-8	28.97	47.22
557	SCG	mse	logsig	25-16-16	30.16	41.67
558	SCG	mse	logsig	25-16-20	34.52	41.67
559	SCG	mse	logsig	25-16-25	34.52	45.83
560	SCG	mse	logsig	25-16-30	33.33	40.28
561	SCG	mse	logsig	25-20-8	31.35	43.06
562	SCG	mse	logsig	25-20-16	35.32	55.56
563	SCG	mse	logsig	25-20-20	33.53	44.44
564	SCG	mse	logsig	25-20-25	29.56	44.44
565	SCG	mse	logsig	25-20-30	36.71	54.17
566	SCG	mse	logsig	25-25-8	25.99	34.72
567	SCG	mse	logsig	25-25-16	37.70	55.56
568	SCG	mse	logsig	25-25-20	27.18	34.72
569	SCG	mse	logsig	25-25-25	32.34	45.83
570	SCG	mse	logsig	25-25-30	31.95	47.22
571	SCG	mse	logsig	25-30-8	36.31	40.28
572	SCG	mse	logsig	25-30-16	36.31	55.56
573	SCG	mse	logsig	25-30-20	37.10	48.61
574	SCG	mse	logsig	25-30-25	37.90	48.61
575	SCG	mse	logsig	25-30-30	31.94	55.56
576	SCG	mse	logsig	30-8-8	27.98	34.72
577	SCG	mse	logsig	30-8-16	30.36	40.28
578	SCG	mse	logsig	30-8-20	25.99	33.33
579	SCG	mse	logsig	30-8-25	23.41	27.78
580	SCG	mse	logsig	30-8-30	29.76	43.06

581	SCG	mse	logsig	30-16-8	26.59	34.72
582	SCG	mse	logsig	30-16-16	28.57	41.67
583	SCG	mse	logsig	30-16-20	31.75	37.50
584	SCG	mse	logsig	30-16-25	28.37	34.72
585	SCG	mse	logsig	30-16-30	28.97	44.44
586	SCG	mse	logsig	30-20-8	27.18	40.28
587	SCG	mse	logsig	30-20-16	36.91	47.22
588	SCG	mse	logsig	30-20-20	27.78	33.33
589	SCG	mse	logsig	30-20-25	35.71	44.44
590	SCG	mse	logsig	30-20-30	35.71	45.83
591	SCG	mse	logsig	30-25-8	30.95	48.61
592	SCG	mse	logsig	30-25-16	31.55	43.06
593	SCG	mse	logsig	30-25-20	38.49	55.56
594	SCG	mse	logsig	30-25-25	31.75	43.06
595	SCG	mse	logsig	30-25-30	27.78	51.39
596	SCG	mse	logsig	30-30-8	31.95	40.28
597	SCG	mse	logsig	30-30-16	30.36	47.22
598	SCG	mse	logsig	30-30-20	35.12	44.44
599	SCG	mse	logsig	30-30-25	33.73	54.17
600	SCG	mse	logsig	30-30-30	31.75	43.06
601	SCG	crossentropy	softmax	8-8-0	33.33	51.39
602	SCG	crossentropy	softmax	8-16-0	31.15	40.28
603	SCG	crossentropy	softmax	8-20-0	30.95	36.11
604	SCG	crossentropy	softmax	8-25-0	32.14	56.94
605	SCG	crossentropy	softmax	8-30-0	41.67	73.61
606	SCG	crossentropy	softmax	16-8-0	31.74	37.50
607	SCG	crossentropy	softmax	16-16-0	38.10	50.00
608	SCG	crossentropy	softmax	16-20-0	28.57	38.89
609	SCG	crossentropy	softmax	16-25-0	37.70	55.56
610	SCG	crossentropy	softmax	16-30-0	34.33	43.06

611	SCG	crossentropy	softmax	20-8-0	28.77	40.28
612	SCG	crossentropy	softmax	20-16-0	31.74	44.44
613	SCG	crossentropy	softmax	20-20-0	27.18	36.11
614	SCG	crossentropy	softmax	20-25-0	26.98	36.11
615	SCG	crossentropy	softmax	20-30-0	31.75	41.67
616	SCG	crossentropy	softmax	25-8-0	22.03	30.56
617	SCG	crossentropy	softmax	25-16-0	30.75	41.67
618	SCG	crossentropy	softmax	25-20-0	32.74	40.28
619	SCG	crossentropy	softmax	25-25-0	35.32	45.83
620	SCG	crossentropy	softmax	25-30-0	28.57	41.67
621	SCG	crossentropy	softmax	30-8-0	29.76	38.89
622	SCG	crossentropy	softmax	30-16-0	25.60	36.11
623	SCG	crossentropy	softmax	30-20-0	25.00	30.56
624	SCG	crossentropy	softmax	30-25-0	29.17	38.89
625	SCG	crossentropy	softmax	30-30-0	31.55	38.89
626	SCG	crossentropy	softmax	8-8-8	26.59	45.83
627	SCG	crossentropy	softmax	8-8-16	36.71	52.78
628	SCG	crossentropy	softmax	8-8-20	32.94	48.61
629	SCG	crossentropy	softmax	8-8-25	30.55	37.50
630	SCG	crossentropy	softmax	8-8-30	30.36	38.89
631	SCG	crossentropy	softmax	8-16-8	27.18	43.06
632	SCG	crossentropy	softmax	8-16-16	35.91	43.06
633	SCG	crossentropy	softmax	8-16-20	29.56	44.44
634	SCG	crossentropy	softmax	8-16-25	34.92	55.56
635	SCG	crossentropy	softmax	8-16-30	36.11	48.61
636	SCG	crossentropy	softmax	8-20-8	35.91	43.06
637	SCG	crossentropy	softmax	8-20-16	28.37	37.50
638	SCG	crossentropy	softmax	8-20-20	28.97	47.22
639	SCG	crossentropy	softmax	8-20-25	30.16	37.50
640	SCG	crossentropy	softmax	8-20-30	30.75	45.83

641	SCG	crossentropy	softmax	8-25-8	31.35	51.39
642	SCG	crossentropy	softmax	8-25-16	31.15	47.22
643	SCG	crossentropy	softmax	8-25-20	32.94	41.67
644	SCG	crossentropy	softmax	8-25-25	38.49	54.17
645	SCG	crossentropy	softmax	8-25-30	28.97	40.28
646	SCG	crossentropy	softmax	8-30-8	29.36	45.83
647	SCG	crossentropy	softmax	8-30-16	38.29	62.50
648	SCG	crossentropy	softmax	8-30-20	34.72	50.00
649	SCG	crossentropy	softmax	8-30-25	34.92	59.72
650	SCG	crossentropy	softmax	8-30-30	38.89	52.78
651	SCG	crossentropy	softmax	16-8-8	25.79	38.89
652	SCG	crossentropy	softmax	16-8-16	27.78	44.44
653	SCG	crossentropy	softmax	16-8-20	29.17	45.83
654	SCG	crossentropy	softmax	16-8-25	31.35	45.83
655	SCG	crossentropy	softmax	16-8-30	34.52	52.78
656	SCG	crossentropy	softmax	16-16-8	31.35	44.44
657	SCG	crossentropy	softmax	16-16-16	29.17	36.11
658	SCG	crossentropy	softmax	16-16-20	38.49	48.61
659	SCG	crossentropy	softmax	16-16-25	25.40	34.72
660	SCG	crossentropy	softmax	16-16-30	36.71	45.83
661	SCG	crossentropy	softmax	16-20-8	37.90	48.61
662	SCG	crossentropy	softmax	16-20-16	32.54	38.89
663	SCG	crossentropy	softmax	16-20-20	28.77	38.89
664	SCG	crossentropy	softmax	16-20-25	31.15	55.56
665	SCG	crossentropy	softmax	16-20-30	28.37	40.28
666	SCG	crossentropy	softmax	16-25-8	34.92	52.78
667	SCG	crossentropy	softmax	16-25-16	28.77	38.89
668	SCG	crossentropy	softmax	16-25-20	33.33	45.83
669	SCG	crossentropy	softmax	16-25-25	34.33	52.78
670	SCG	crossentropy	softmax	16-25-30	33.14	41.67

671	SCG	crossentropy	softmax	16-30-8	31.94	38.89
672	SCG	crossentropy	softmax	16-30-16	37.90	50.00
673	SCG	crossentropy	softmax	16-30-20	27.58	43.06
674	SCG	crossentropy	softmax	16-30-25	36.11	47.22
675	SCG	crossentropy	softmax	16-30-30	26.98	43.06
676	SCG	crossentropy	softmax	20-8-8	24.21	33.33
677	SCG	crossentropy	softmax	20-8-16	27.38	37.50
678	SCG	crossentropy	softmax	20-8-20	25.79	34.72
679	SCG	crossentropy	softmax	20-8-25	30.36	43.06
680	SCG	crossentropy	softmax	20-8-30	25.20	34.72
681	SCG	crossentropy	softmax	20-16-8	27.78	40.28
682	SCG	crossentropy	softmax	20-16-16	25.79	33.33
683	SCG	crossentropy	softmax	20-16-20	34.72	45.83
684	SCG	crossentropy	softmax	20-16-25	34.92	44.44
685	SCG	crossentropy	softmax	20-16-30	30.16	37.50
686	SCG	crossentropy	softmax	20-20-8	31.94	45.83
687	SCG	crossentropy	softmax	20-20-16	29.76	45.83
688	SCG	crossentropy	softmax	20-20-20	30.76	51.39
689	SCG	crossentropy	softmax	20-20-25	33.73	45.83
690	SCG	crossentropy	softmax	20-20-30	34.72	56.94
691	SCG	crossentropy	softmax	20-25-8	30.75	36.11
692	SCG	crossentropy	softmax	20-25-16	28.97	47.22
693	SCG	crossentropy	softmax	20-25-20	32.34	43.06
694	SCG	crossentropy	softmax	20-25-25	32.14	37.50
695	SCG	crossentropy	softmax	20-25-30	31.94	38.89
696	SCG	crossentropy	softmax	20-30-8	30.16	41.67
697	SCG	crossentropy	softmax	20-30-16	32.74	44.44
698	SCG	crossentropy	softmax	20-30-20	31.75	37.50
699	SCG	crossentropy	softmax	20-30-25	30.75	40.28
700	SCG	crossentropy	softmax	20-30-30	30.56	43.06

701	SCG	crossentropy	softmax	25-8-8	23.81	37.50
702	SCG	crossentropy	softmax	25-8-16	27.78	40.28
703	SCG	crossentropy	softmax	25-8-20	33.33	44.44
704	SCG	crossentropy	softmax	25-8-25	31.15	43.06
705	SCG	crossentropy	softmax	25-8-30	26.59	37.50
706	SCG	crossentropy	softmax	25-16-8	32.74	43.06
707	SCG	crossentropy	softmax	25-16-16	26.39	34.72
708	SCG	crossentropy	softmax	25-16-20	31.15	41.67
709	SCG	crossentropy	softmax	25-16-25	31.95	44.44
710	SCG	crossentropy	softmax	25-16-30	33.73	52.78
711	SCG	crossentropy	softmax	25-20-8	30.36	40.28
712	SCG	crossentropy	softmax	25-20-16	34.52	54.17
713	SCG	crossentropy	softmax	25-20-20	27.18	31.94
714	SCG	crossentropy	softmax	25-20-25	33.93	37.50
715	SCG	crossentropy	softmax	25-20-30	37.10	45.83
716	SCG	crossentropy	softmax	25-25-8	31.75	40.28
717	SCG	crossentropy	softmax	25-25-16	31.55	38.89
718	SCG	crossentropy	softmax	25-25-20	35.51	50.00
719	SCG	crossentropy	softmax	25-25-25	29.17	44.44
720	SCG	crossentropy	softmax	25-25-30	35.71	41.67
721	SCG	crossentropy	softmax	25-30-8	30.95	47.22
722	SCG	crossentropy	softmax	25-30-16	27.78	34.72
723	SCG	crossentropy	softmax	25-30-20	30.16	40.28
724	SCG	crossentropy	softmax	25-30-25	28.77	40.28
725	SCG	crossentropy	softmax	25-30-30	35.12	44.44
726	SCG	crossentropy	softmax	30-8-8	22.02	30.56
727	SCG	crossentropy	softmax	30-8-16	26.39	37.50
728	SCG	crossentropy	softmax	30-8-20	29.96	38.89
729	SCG	crossentropy	softmax	30-8-25	27.38	43.06
730	SCG	crossentropy	softmax	30-8-30	24.60	31.94

731	SCG	crossentropy	softmax	30-16-8	24.60	38.89
732	SCG	crossentropy	softmax	30-16-16	27.98	34.72
733	SCG	crossentropy	softmax	30-16-20	29.56	44.44
734	SCG	crossentropy	softmax	30-16-25	30.16	41.67
735	SCG	crossentropy	softmax	30-16-30	28.77	45.83
736	SCG	crossentropy	softmax	30-20-8	24.01	29.17
737	SCG	crossentropy	softmax	30-20-16	24.60	30.56
738	SCG	crossentropy	softmax	30-20-20	31.94	48.61
739	SCG	crossentropy	softmax	30-20-25	28.57	38.89
740	SCG	crossentropy	softmax	30-20-30	25.99	36.11
741	SCG	crossentropy	softmax	30-25-8	25.79	33.33
742	SCG	crossentropy	softmax	30-25-16	27.18	40.28
743	SCG	crossentropy	softmax	30-25-20	25.59	31.94
744	SCG	crossentropy	softmax	30-25-25	28.97	41.67
745	SCG	crossentropy	softmax	30-25-30	28.17	48.61
746	SCG	crossentropy	softmax	30-30-8	19.05	26.39
747	SCG	crossentropy	softmax	30-30-16	34.33	41.67
748	SCG	crossentropy	softmax	30-30-20	26.39	45.83
749	SCG	crossentropy	softmax	30-30-25	28.77	44.44
750	SCG	crossentropy	softmax	30-30-30	26.39	41.67

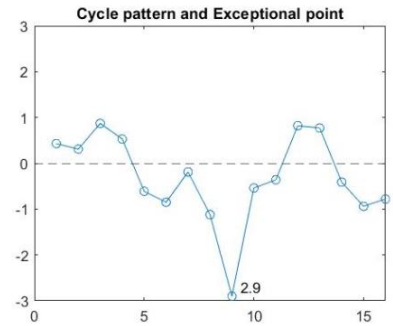
Παράρτημα Β

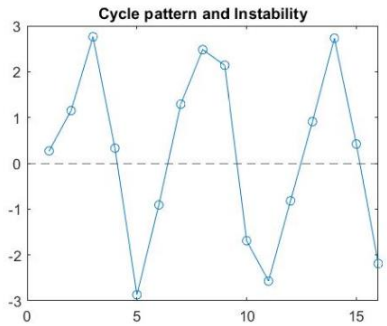
Πίνακας 22: Τιμές της μέσης ακρίβειας και ανάκλησης της κάθε κλάσης, για όλες τις αρχιτεκτονικές του Πίνακα 4.

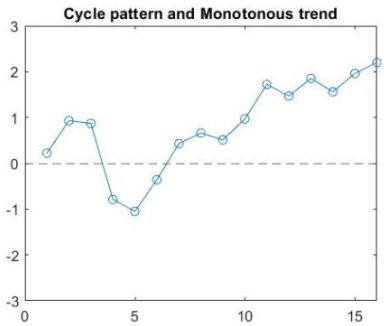
	cycle pattern		exceptional point		grouping		instability		level shift		monotonous trend		point outside of control limits		stratification	
	Ακρίβεια (%)	Ανάκληση (%)	Ακρίβεια (%)	Ανάκληση (%)	Ακρίβεια (%)	Ανάκληση (%)	Ακρίβεια (%)	Ανάκληση (%)	Ακρίβεια (%)	Ανάκληση (%)	Ακρίβεια (%)	Ανάκληση (%)	Ακρίβεια (%)	Ανάκληση (%)	Ακρίβεια (%)	Ανάκληση (%)
1	94.2	93.7	95.6	98.4	98.2	77.8	94.8	96.8	97.4	100.0	97.4	100.0	94.2	96.8	95.4	93.7
2	96.0	96.8	90.7	98.4	96.8	84.1	96.4	93.7	96.8	96.8	98.6	96.8	96.8	92.1	96.8	96.8
3	93.2	88.9	91.8	96.8	98.6	87.3	97.1	82.5	96.8	96.8	91.6	95.2	94.6	100.0	89.3	96.8
4	91.5	88.9	98.6	98.4	95.7	82.5	89.1	98.4	93.1	96.8	92.4	92.1	97.4	93.7	94.0	95.2
5	91.4	93.7	91.2	98.4	100.0	81.0	96.0	95.2	100.0	96.8	92.6	95.2	90.9	92.1	92.8	93.7
6	91.5	85.7	97.4	95.2	94.6	82.5	97.0	95.2	95.7	96.8	88.3	93.7	90.1	96.8	95.4	96.8
7	32.9	60.3	22.7	30.2	27.0	22.2	58.8	36.5	33.4	41.3	54.0	36.5	59.0	34.9	30.6	30.2
8	32.5	58.7	30.7	17.5	24.6	20.6	40.5	22.2	50.3	42.9	41.7	42.9	44.4	36.5	24.3	31.7
9	31.7	61.9	31.8	19.1	17.3	20.6	70.2	33.3	46.8	31.8	40.7	50.8	39.9	23.8	20.7	28.6
10	32.9	63.5	68.4	38.1	27.6	25.4	58.4	42.9	59.4	38.1	70.3	42.9	79.4	52.4	31.8	36.5
11	27.2	55.6	39.7	41.3	56.4	33.3	71.3	42.9	46.7	25.4	56.8	46.0	52.9	46.0	22.2	39.7
12	28.6	55.6	37.8	30.2	36.1	28.6	65.0	38.1	70.2	41.3	63.7	41.3	54.0	44.5	23.8	36.5
13	34.3	68.3	42.3	25.4	37.3	27.0	70.2	39.7	52.9	41.3	67.7	42.9	66.2	52.4	28.2	36.5
14	30.9	63.5	24.9	23.8	28.5	28.6	70.6	44.4	66.4	34.9	54.9	36.5	61.5	33.3	33.9	46.0
15	40.6	61.9	45.8	23.8	27.3	22.2	43.6	34.9	55.3	41.3	59.3	39.7	47.9	41.3	22.9	42.9

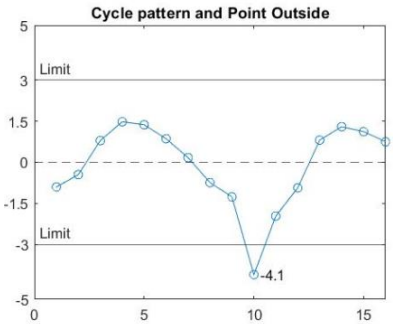
Παράρτημα Γ

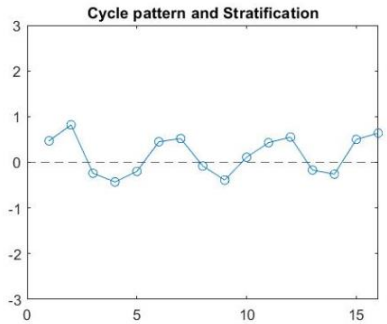
Με σκοπό να εξεταστεί η δυνατότητα ανίχνευσης μεικτών τάσεων, αντλούνται από το [4] τα αντίστοιχα διαγράμματα ελέγχου, που μετατρέπονται σε 19 διανύσματα. Αυτά αποτελούνται από 16 κόμβους και σε αυτούς προστίθενται στην αρχή οι 2 κόμβοι των ορίων. Τα όρια σε όλες τις περιπτώσεις είναι -3 και 3. Παρακάτω παρουσιάζονται τα διανύσματα και οι τάσεις οι οποίες συνδυάζονται σε κάθε ένα.

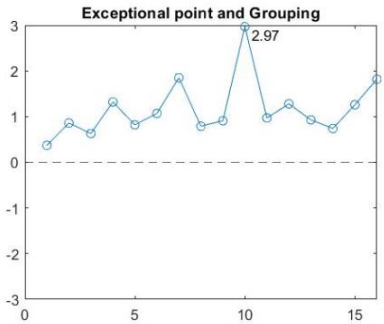


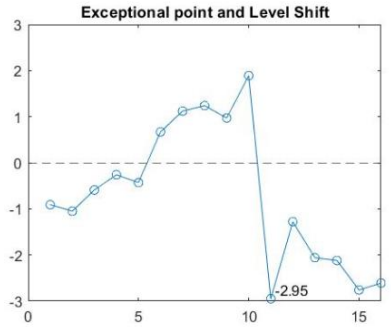


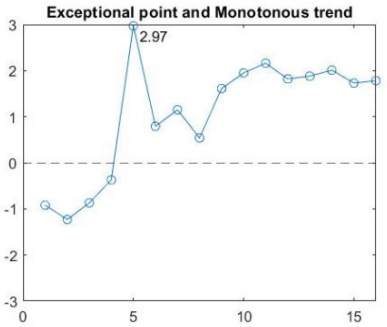


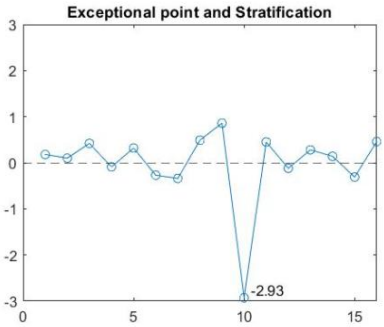


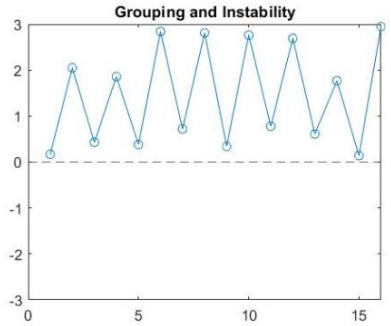


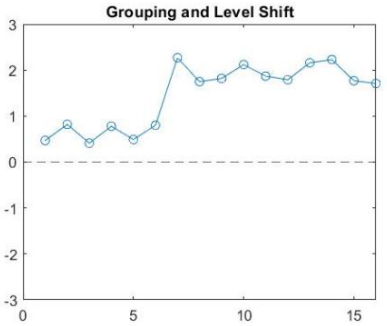


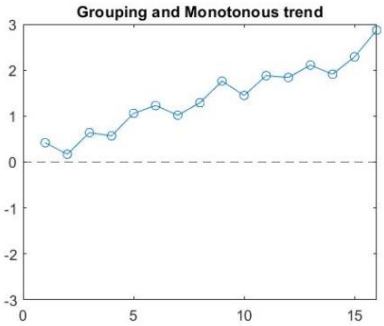


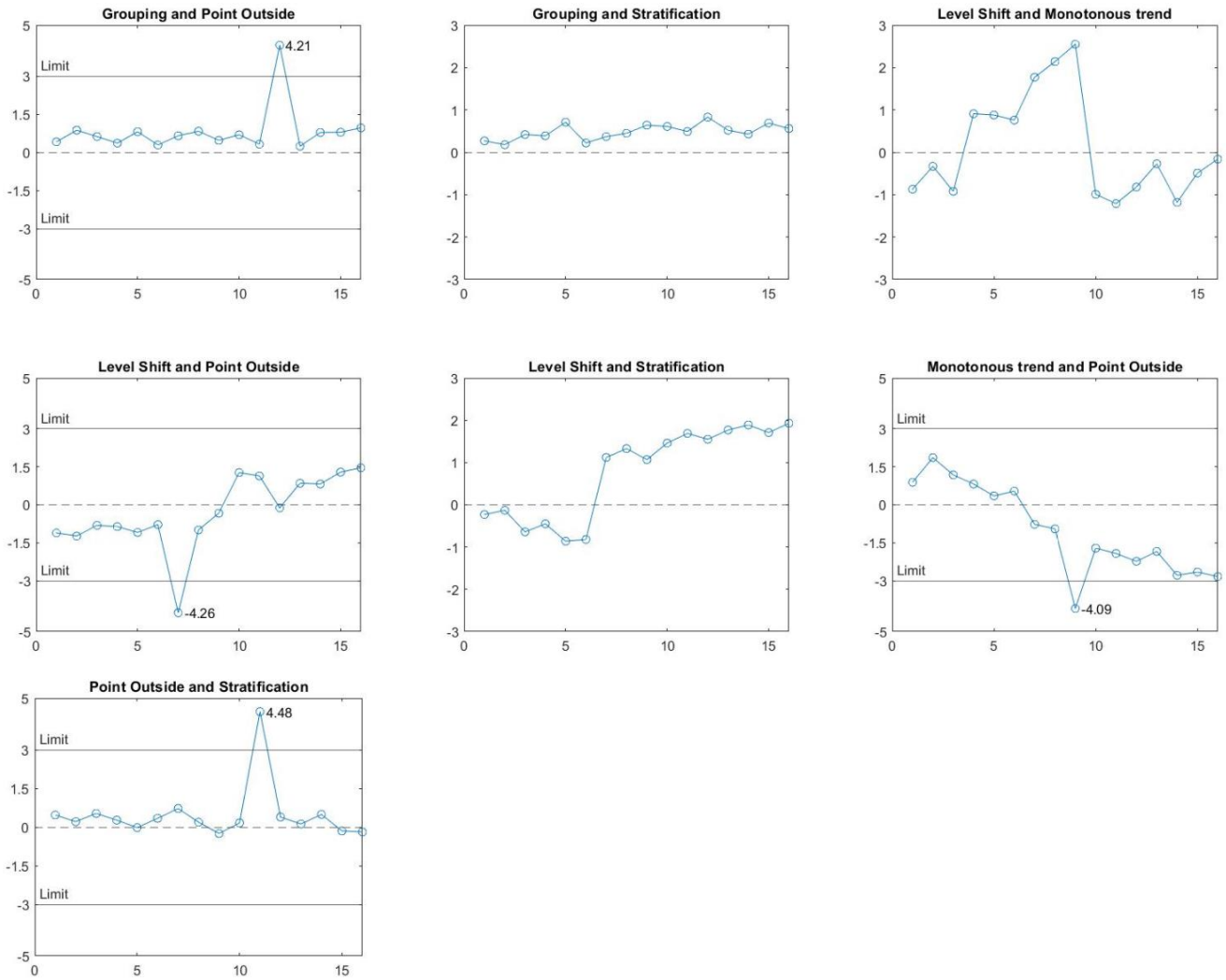












Εικόνα 68: Τα 19 διανύσματα μεικτών τάσεων που δημιουργήθηκαν. Από αριστερά προς τα δεξιά οι τάσεις που συνδυάζονται είναι: Περιοδικότητα & Σημείο πολύ κοντά στα όρια ελέγχου, Περιοδικότητα & Διακύμανση, Περιοδικότητα & Μονότονη τάση, Περιοδικότητα & Σημείο εκτός ορίων ελέγχου, Περιοδικότητα & Έλλειψη μεταβλητότητας, Σημείο πολύ κοντά στα όρια ελέγχου & Ομαδοποίηση, Σημείο πολύ κοντά στα όρια ελέγχου & Άλμα αλλαγής επιπέδου, Σημείο πολύ κοντά στα όρια ελέγχου & Μονότονη τάση, Σημείο πολύ κοντά στα όρια ελέγχου & Έλλειψη μεταβλητότητας, Ομαδοποίηση & Διακύμανση, Ομαδοποίηση & Άλμα αλλαγής επιπέδου, Ομαδοποίηση & Μονότονη τάση, Ομαδοποίηση & Σημείο εκτός ορίων ελέγχου, Ομαδοποίηση & Έλλειψη μεταβλητότητας, Άλμα αλλαγής επιπέδου & Μονότονη τάση, Άλμα αλλαγής επιπέδου & Σημείο εκτός ορίων ελέγχου, Άλμα αλλαγής επιπέδου & Έλλειψη μεταβλητότητας, Μονότονη τάση & Σημείο εκτός ορίων ελέγχου, Σημείο εκτός ορίων ελέγχου & Έλλειψη μεταβλητότητας

Τα 19 διανύσματα που κατασκευάζονται, χρησιμοποιούνται ως διανύσματα ελέγχου στα πιο ακριβή, με βάση την βέλτιστη ακρίβεια, μοντέλα ΤΝΔ. Αυτά, ανήκουν στις αρχιτεκτονικές που φαίνονται στις 5 πρώτες θέσεις του Πίνακα 6.

Κατά αναλογία με όσα αναφέρονται στην παράγραφο 5.1.1 για να πραγματοποιηθεί η κατηγοριοποίηση, επιλέγονται τα δύο μεγαλύτερα στοιχεία του διανύσματος εξόδου στρογγυλοποιούνται στο 1, ενώ τα υπόλοιπα στρογγυλοποιούνται στο 0.

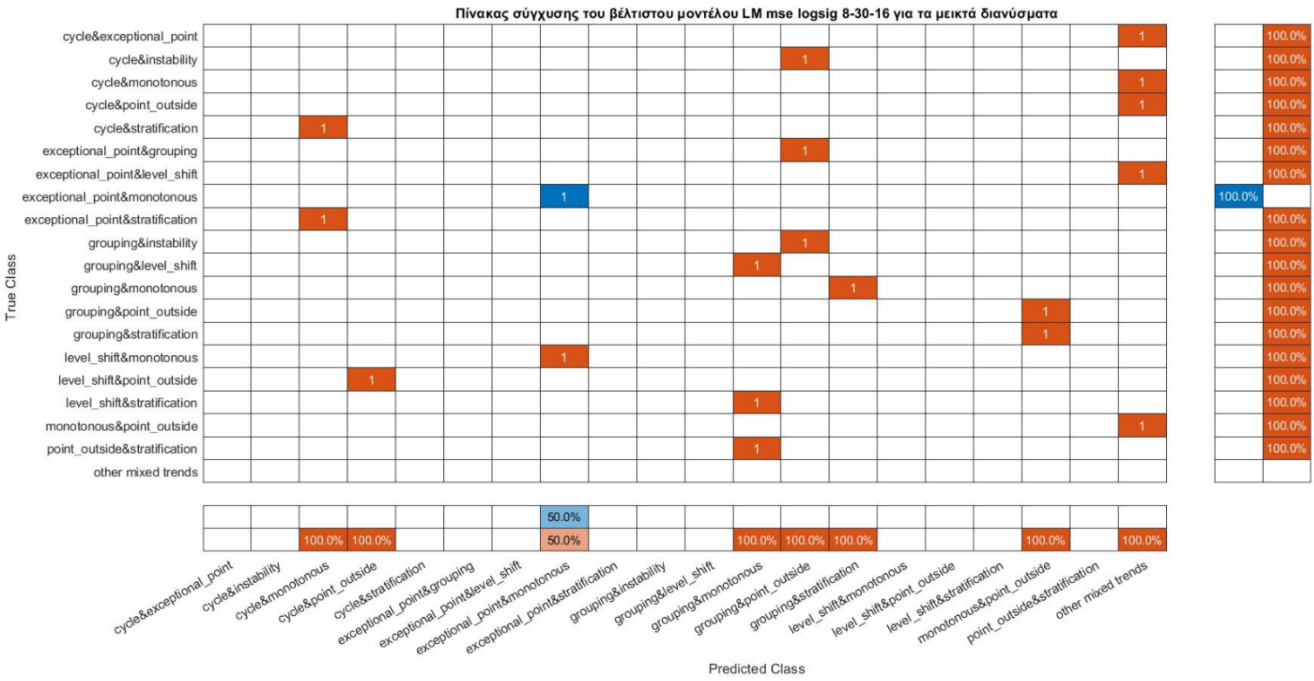
Εφόσον οι αρχικές, μονές, τάσεις είναι 8 το σύνολο των δυνατών συνδυασμών είναι 28. Συνεπώς επειδή μορφοποιούνται 19 μεικτές τάσεις, υπολείπονται 9, η οποίες συμπεριλαμβάνονται σε μία κατηγορία με το όνομα other mixed trends.

Στον Πίνακα 23 παρουσιάζονται τα αποτελέσματα των δοκιμών.

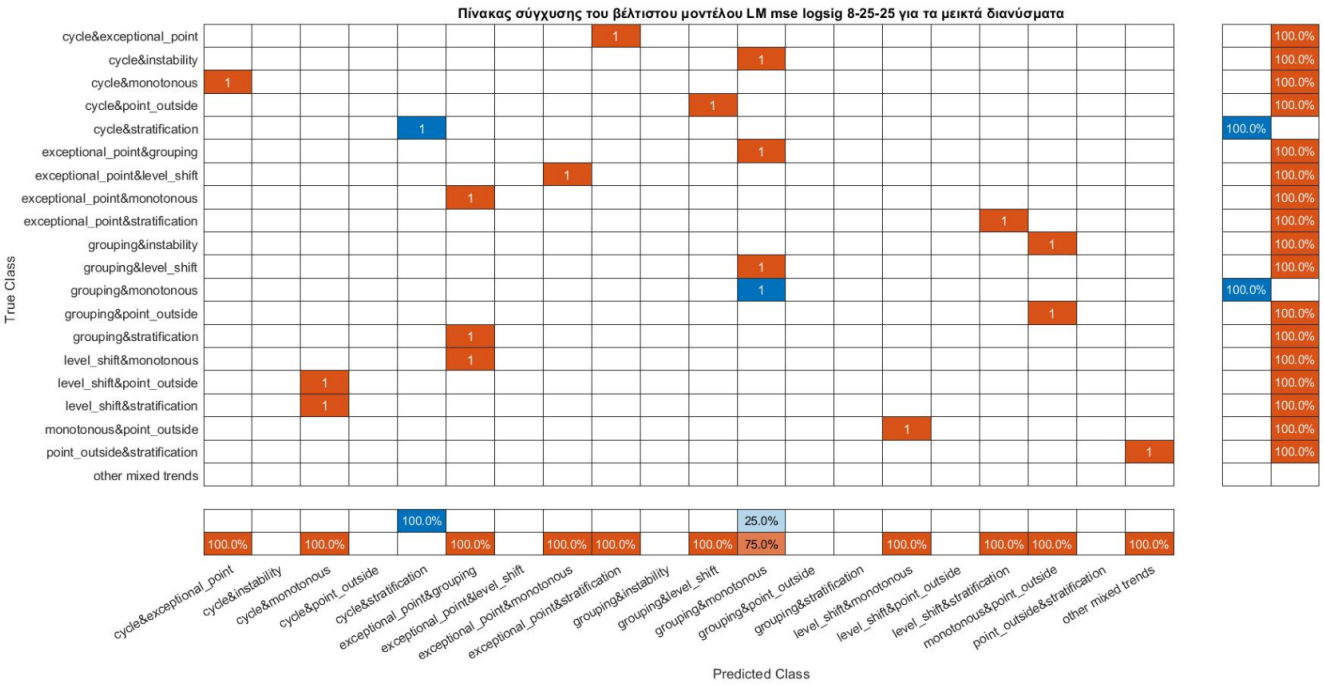
Πίνακας 23: Αποτελέσματα των ακριβέστερων μοντέλων, για τα διανύσματα ελέγχου των μεικτών τάσεων.

Αλγόριθμος Εκπαίδευσης	Συν.Κόστους	Συν. Ενεργ. Επ. Εξόδου	Κόμβοι Κρυμμένων Επιπέδων	Ακρίβεια (%)
LM	mse	logsig	8-30-16	5.26
LM	mse	logsig	8-25-25	10.53
LM	mse	logsig	8-30-25	5.26
LM	mse	softmax	16-30-30	21.05
LM	mse	softmax	20-25-8	15.79

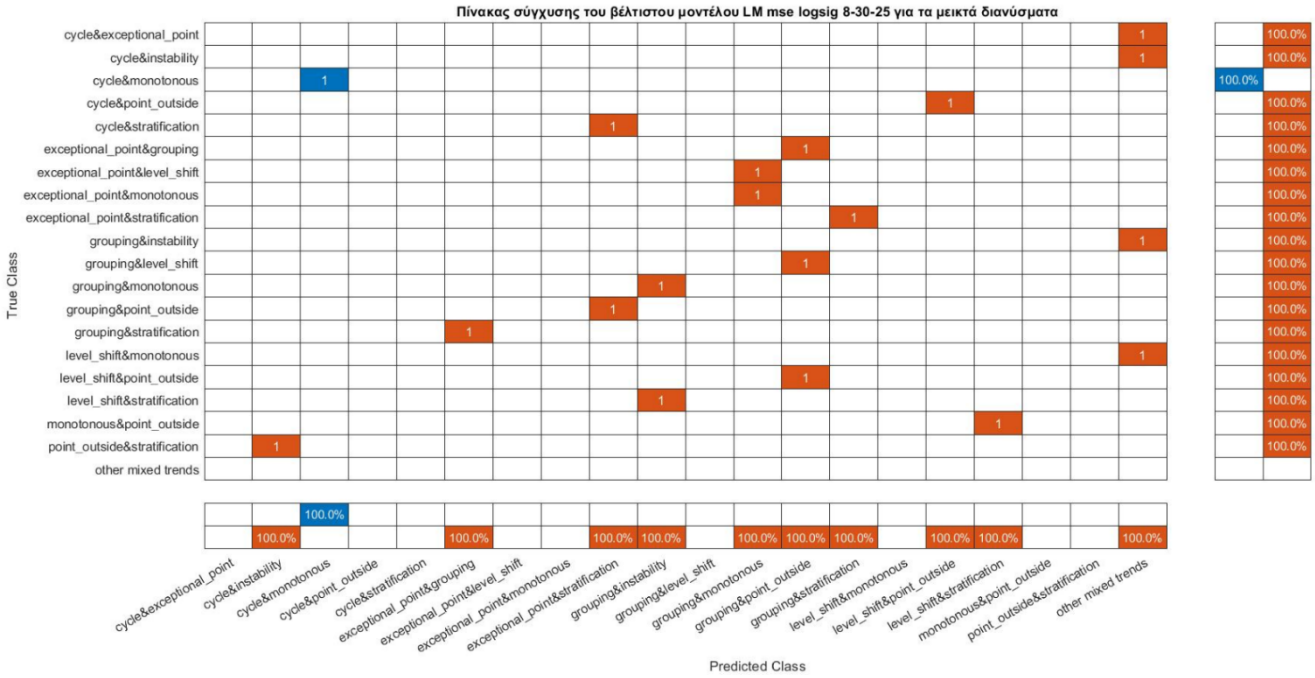
Παρακάτω παρουσιάζονται οι αντίστοιχοι πίνακες σύγχυσης:



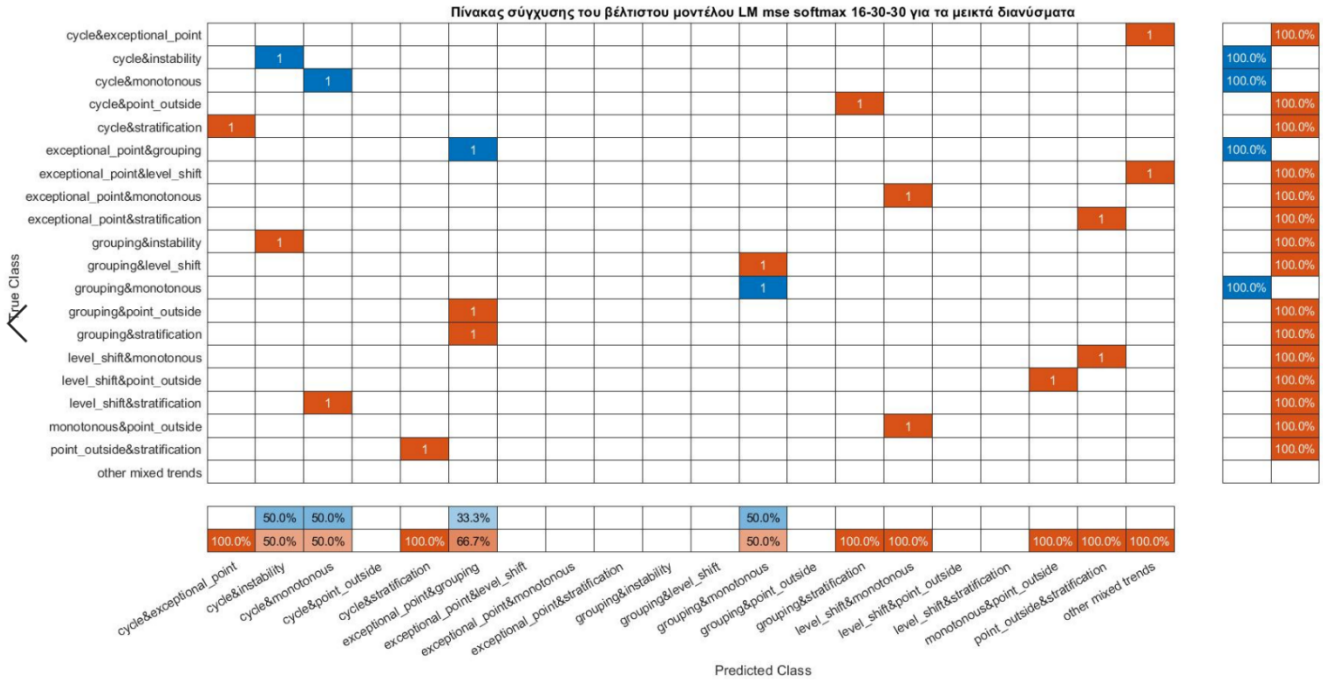
(α)


Πίνακας σύγχυσης του βέλτιστου μοντέλου LM mse logsig 8-25-25 για τα μεικτά διανύσματα
True Class
Predicted Class
cycle&exceptional_point
cycle&instability
cycle&monotonous
cycle&point_outside
cycle&stratification
exceptional_point&grouping
exceptional_point&level_shift
exceptional_point&monotonous
exceptional_point&stratification
grouping&instability
grouping&level_shift
grouping&monotonous
grouping&point_outside
grouping&stratification
level_shift&monotonous
level_shift&point_outside
level_shift&stratification
monotonous&point_outside
point_outside&stratification
other mixed trends

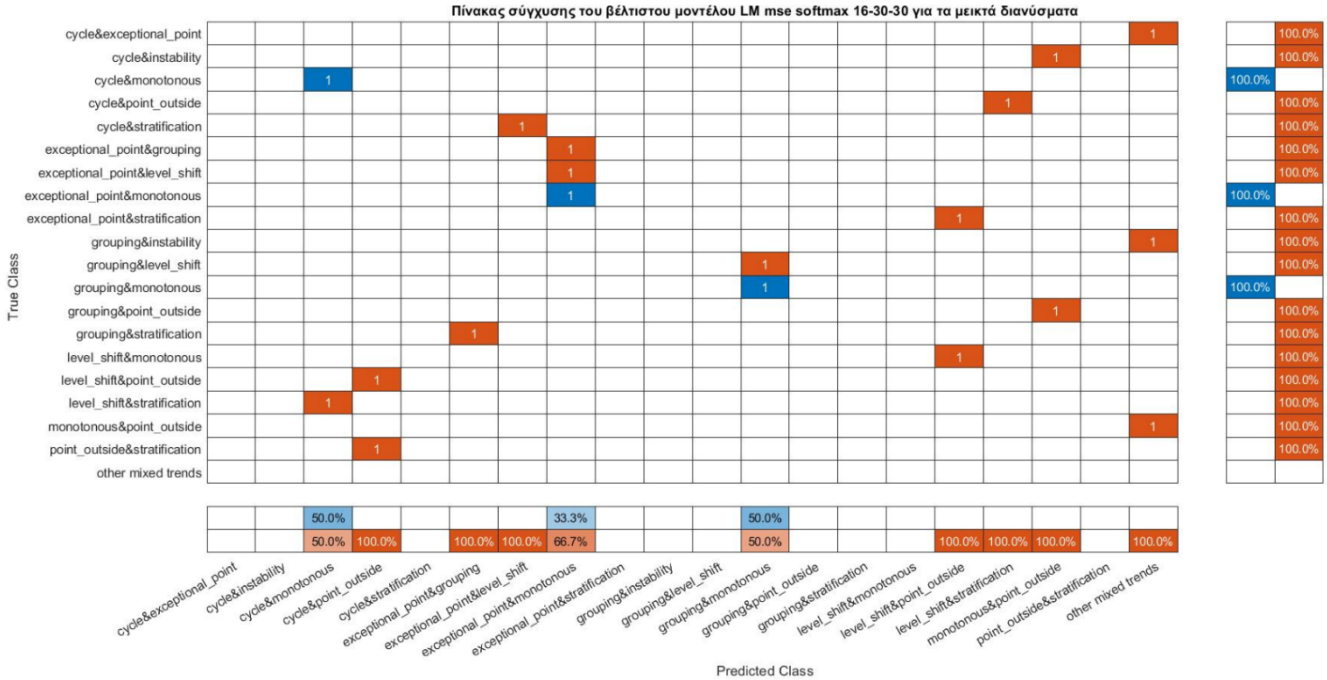
(β)


Πίνακας σύγχυσης του βέλτιστου μοντέλου LM mse logsig 8-30-25 για τα μεικτά διανύσματα
True Class
Predicted Class
cycle&exceptional_point
cycle&instability
cycle&monotonous
cycle&point_outside
cycle&stratification
exceptional_point&grouping
exceptional_point&level_shift
exceptional_point&monotonous
exceptional_point&stratification
grouping&instability
grouping&level_shift
grouping&monotonous
grouping&point_outside
grouping&stratification
level_shift&monotonous
level_shift&point_outside
level_shift&stratification
monotonous&point_outside
point_outside&stratification
other mixed trends

(γ)



(δ)



(ε)

Εικόνα 69: Πίνακες σύγχυσης για τα διανύσματα των μεικτών τάσεων των βέλτιστων μοντέλων του Πίνακα 6.

Από τα αποτελέσματα προκύπτει ότι η ικανότητα των μοντέλων στην αναγνώριση μεικτών τάσεων είναι περιορισμένη [26].

Παράρτημα Δ

Πίνακας 24: Οι ακριβείς τιμές των τριών μητρικών διανυσμάτων της κάθε τάσης. Φαίνονται επίσης, η τυπική απόκλιση (τ.α.) και η παράμετρος θέσης των ορίων ελέγχου (k).

Cycle pattern																τ.α.
0.20	0.70	-0.60	-0.95	-1.00	-0.70	0.15	0.55	0.50	-0.65	-0.90	-0.92	0.50	1.00	1.30	1.20	0.93
0.51	0.95	0.36	-0.38	0.19	0.57	0.39	-0.29	-1.40	-0.21	1.20	0.67	-0.59	0.37	0.92	0.87	0.80
0.12	0.51	0.75	0.27	-0.14	-0.70	-0.48	-0.19	0.26	0.64	0.81	0.57	0.09	-0.33	-0.29	-0.06	0.55

Exceptional point																τ.α.
0.45	-0.63	-0.65	-0.30	1.49	1.10	0.87	0.91	-0.33	0.03	-1.10	-1.22	-0.95	-0.98	0.35	0.43	0.96
0.35	0.56	0.24	-0.19	0.15	-0.26	-0.40	-0.63	-0.17	-0.32	0.21	0.47	-0.22	-0.49	0.99	0.75	0.56
-0.31	-0.23	-0.59	0.32	0.50	0.21	-0.12	-0.47	1.19	0.82	0.39	0.42	-0.09	-0.19	0.14	0.06	0.61

Grouping																τ.α.
0.46	0.75	0.09	0.38	0.27	0.82	0.89	0.62	0.70	0.29	0.36	0.65	0.32	0.48	0.76	0.51	0.44
0.16	0.29	0.80	0.59	0.21	0.44	0.10	0.52	0.73	0.24	0.27	0.14	0.91	0.51	0.65	0.36	0.44
0.87	0.48	0.56	0.17	0.32	0.45	0.23	0.64	0.26	0.51	0.06	0.25	0.19	0.36	0.29	0.60	0.42

Instability																τ.α.
1.23	-1.31	1.79	-1.15	0.87	-2.25	0.76	-1.69	1.72	-1.05	1.83	-2.14	0.65	-1.85	1.89	-1.77	1.74
1.65	-1.20	1.15	-0.57	2.31	-0.61	0.52	-1.34	1.96	-1.50	1.24	-1.22	1.46	-2.21	1.09	-1.64	1.65
-1.10	1.46	-0.37	0.35	-0.47	0.24	-2.15	0.31	-1.27	1.22	-1.46	2.34	-1.03	1.86	-0.59	1.92	1.54

Level shift																τ.α.
0.87	1.15	1.26	1.03	1.82	1.65	-1.65	-1.02	-1.12	-1.80	-1.07	-1.19	-1.83	-1.89	-1.34	-1.23	1.49
0.61	0.79	1.21	0.83	1.46	0.85	0.71	0.46	-1.23	-1.31	-0.76	-0.55	-1.12	-1.49	-1.40	-0.59	1.20
0.18	0.46	0.27	0.20	0.23	-0.31	-0.21	-0.29	-0.37	-0.12	-0.08	-0.35	-0.19	-0.56	-0.43	-0.59	0.46

Monotonous trend																τ.α.
-1.40	-0.90	-0.85	-0.60	-0.70	-0.15	0.20	0.25	0.80	0.75	0.82	1.15	1.21	1.92	2.10	2.25	1.38
-0.80	-0.35	-1.40	0.25	-0.51	0.52	0.16	0.72	0.21	0.61	0.20	1.10	0.47	1.41	0.82	2.10	1.18
0.50	0.75	0.28	0.35	0.22	0.31	0.14	0.26	-0.20	-0.41	-0.26	-0.51	-0.62	-0.39	-0.54	-0.85	0.56

Point outside of control limits																τ.α.
-0.48	0.35	1.20	0.83	-0.21	0.31	1.35	0.92	1.14	-0.91	-0.79	3.10	1.10	1.15	-0.75	-0.19	1.13
0.56	0.77	0.35	-0.16	0.23	-0.21	-0.43	-0.88	-0.54	1.90	0.99	0.90	0.37	0.45	-0.10	-0.27	0.74
0.11	0.47	1.20	0.52	-0.06	-0.61	-0.17	-0.32	6.15	0.67	0.89	0.46	-0.20	0.05	-1.10	-0.41	1.62

Stratification																τ.α.
0.20	0.41	-0.18	-0.11	-0.22	0.46	0.17	0.74	-0.09	0.12	-0.03	0.21	-0.15	-0.49	-0.26	0.32	0.46
0.16	-0.24	0.25	-0.10	0.11	-0.27	-0.18	0.30	-0.03	-0.72	0.09	0.20	0.39	-0.06	-0.21	0.19	0.43
0.27	0.75	-0.10	0.16	-0.19	-0.22	0.14	0.18	-0.08	-0.13	0.29	0.11	-0.15	-0.31	-0.17	0.21	0.43

Πίνακας 25: Διάνυσμα στόχος κάθε τάσης.

Τάση	Διάνυσμα στόχος							
Cycle pattern	1	0	0	0	0	0	0	0
Exceptional point	0	1	0	0	0	0	0	0
Grouping	0	0	1	0	0	0	0	0
Instability	0	0	0	1	0	0	0	0
Level shift	0	0	0	0	1	0	0	0
Monotonous trend	0	0	0	0	0	1	0	0
Point outside of control limits	0	0	0	0	0	0	1	0
Stratification	0	0	0	0	0	0	0	1