



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Χημικών Μηχανικών
Τομέας II - Τομέας Ανάλυσης, Σχεδιασμού και Ανάπτυξης Διεργασιών και Συστημάτων
Μονάδα Αυτόματης Ρύθμισης και Πληροφορικής

Διπλωματική Εργασία

**Ανάπτυξη μαθηματικών μοντέλων πρόβλεψης ιδιοτήτων χημικών ουσιών με
τεχνικές βαθιάς μηχανικής μάθησης**

Χρήστος Ζαχαρόπουλος

Επιβλέπων Καθηγητής
Χαράλαμπος Σαρίμβεης

Αθήνα, Σεπτέμβριος 2021



National Technical University of Athens
School of Chemical Engineering
Department II: Department of Process Analysis and Plant Design
Unit of Process Control and Informatics

Diploma Thesis

Development of deep learning models for predicting properties of chemical compounds

Christos Zacharopoulos

Supervisor

Prof. Haralambos Sarimveis

Athens, September 2021

Περίληψη

Τα τελευταία χρόνια, η βαθιά μάθηση (deep learning) έχει βρει εφαρμογή σε ποικίλους τομείς της καθημερινότητας και της επιστήμης. Αρχιτεκτονικές βαθιών νευρωνικών δικτύων χρησιμοποιούνται για την εξαγωγή μοντέλων υπολογιστικής όρασης, αναγνώρισης ομιλίας, αλλά και ανακάλυψης φαρμάκων και ανάλυσης ιατρικών εικόνων. Στην παρούσα εργασία, αναπτύσσονται δύο μοντέλα βαθιών νευρωνικών δικτύων με σκοπό την ταξινόμηση χημικών ουσιών σε τρεις κλάσεις ανάλογα με το δυναμικό πρόσδεσής τους στους υποδοχείς οιστρογόνων του ανθρώπινου οργανισμού. Το πρώτο μοντέλο αφορά την επεξεργασία δεδομένων εισόδου με τη μορφή περιγραφικών μεταβλητών 2D και 3D μοριακής δομής μέσω ενός βαθιού νευρωνικού δικτύου εμπρόσθιας τροφοδότησης (μοντέλο Multilayer Perceptron - MLP). Η μεθοδολογία αυτή μπορεί να υπαχθεί στις κλασικές μεθοδολογίες Ποσοτικών Σχέσεων Δομής-Δράσης (Quantitative Structure-Activity Relationship - QSAR). Το δεύτερο μοντέλο νευρωνικών δικτύων δέχεται ως είσοδο χημικούς γράφους, δηλαδή μη-ευκλείδειες απεικονίσεις της δομής των χημικών μορίων, όπου οι κόμβοι αναπαριστούν τα άτομα και οι ακμές τους χημικούς δεσμούς μεταξύ των ατόμων (μοντέλο Graph Neural Networks - GNN). Με αυτόν τον τρόπο, επιτυγχάνεται η προτυποποίηση των σχέσεων μεταξύ δομικών χαρακτηριστικών και βιολογικής δράσης χωρίς να είναι αναγκαία η κατασκευή και επιλογή κατάλληλων περιγραφικών μεταβλητών σε μοριακό επίπεδο.

Τα παραπάνω μοντέλα βελτιστοποιούνται και εκπαιδεύονται με χρήση γλώσσας προγραμματισμού Python και υπολογιστικών πόρων από το Google Colaboratory, ενώ αξιολογούνται τόσο στα δεδομένα επαλήθευσης όσο και στα δεδομένα εκπαίδευσης. Ακόμη, επιχειρείται σύγκριση με μοντέλο προηγούμενης εργασίας, το οποίο επιλύει το ίδιο πρόβλημα ταξινόμησης τριών κλάσεων χρησιμοποιώντας βαθιά συνελκτικά νευρωνικά δίκτυα με είσοδο εικόνες χημικών μορίων (μοντέλο Convolutional Neural Networks - CNN). Συνοπτικά, τα δύο μοντέλα της παρούσας εργασίας προβλέπουν ικανοποιητικά την κλάση ενός ενδοκρινικού διαταράκτη (η ακρίβεια πρόβλεψης ξεπερνάει το 0.7), ενώ παρουσιάζουν καλύτερα αποτελέσματα από το μοντέλο Convolutional Neural Networks - CNN. Αναφορικά με την επιμέρους σύγκριση των μοντέλων, το μοντέλο Graph Neural Networks - GNN υπερτερεί, έστω και σε μικρό βαθμό, του μοντέλου Multilayer Perceptron - MLP σε όλες τις μετρικές τιμές αξιολόγησης. Τέλος, οι προοπτικές περαιτέρω βελτίωσης της προβλεπτικής ικανότητας του μοντέλου Graph Neural Networks - GNN μέσα από την ενσωμάτωση σύγχρονων και στιβαρότερων στρωμάτων νευρωνικής ανταλλαγής μηνυμάτων αναδεικνύουν το σημαντικό ερευνητικό ενδιαφέρον γύρω από την πρόβλεψη ιδιοτήτων χημικών ουσιών με αρχιτεκτονικές νευρωνικών δικτύων που επεξεργάζονται δομές γράφων.

Λέξεις-κλειδιά: βαθιά μάθηση, νευρωνικά δίκτυα εμπρόσθιας τροφοδότησης, νευρωνικά δίκτυα γράφων, συνελκτικά νευρωνικά δίκτυα, ποσοτικές σχέσεις δομής-δράσης, χημικός γράφος, ενδοκρινικοί διαταράκτες

Abstract

Over the last years, deep learning has found application in various fields of daily life and science. Deep neural networks architectures are being used for the extraction of models dealing with problems concerning a diverse range of challenging tasks, such as computer vision, speech recognition, medical images analysis and drug discovery. In this thesis, two models of deep neural networks that perform the task of 3-class classification of chemical compounds depending on their estrogen receptor binding strength are developed. The first model processes input data in the form of 2D and 3D molecular structure descriptors using a deep feedforward neural network (Multilayer Perceptron - MLP model). This approach can be seen as a regular Quantitative Structure-Activity Relationship (QSAR) approach. The second model's inputs are chemical graphs, i.e. non-euclidean illustrations of molecules, where nodes and edges represent atoms and bonds respectively (Graph Neural Networks - GNN model). In this way, modeling of the relationship between the structural features and the biological activity is achieved without the need for the generation and proper choice of molecular descriptors.

The above models have been optimized, trained and evaluated using Python and computational resources from Google Colaboratory. Furthermore, they are compared with a previous work's model, which is dealing with the same 3-class classification problem using molecular image-based deep convolutional neural networks (Convolutional Neural Networks - CNN model). Briefly, both of the models developed in the current thesis satisfactorily predict the class of an endocrine disrupting chemical (accuracy overcomes 0.7), while, in general, they are performing better than the Convolutional Neural Networks - CNN model. It is a fact that the Graph Neural Networks - GNN model outperforms Multilayer Perceptron - MLP model in all of the evaluation metrics, even in a small degree. Finally, the prospects for further improvement of Graph Neural Networks - GNN model's forecasting ability through integration with state-of-the-art and more powerful neural message passing layers highlight the importance of research surrounding chemical properties prediction using graph-processing neural networks structures.

Keywords: deep learning, feedforward neural networks, graph neural networks, convolutional neural networks, Quantitative Structure-Activity Relationship, chemical graph, endocrine disrupting chemicals

Ευχαριστίες

Η παρούσα διπλωματική εργασία με τίτλο "Ανάπτυξη μαθηματικών μοντέλων πρόβλεψης ιδιοτήτων χημικών ουσιών με τεχνικές βαθιάς μηχανικής μάθησης" ολοκληρώνει τον προπτυχιακό κύκλο σπουδών μου στη σχολή Χημικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου. Έναν κύκλο σπουδών, ο οποίος παρά τις πρωτόγνωρες και δύσκολες για όλους μας συνθήκες του τελευταίου ενάμιση χρόνου, μόνο θετικά στοιχεία μπορεί να μου προσκομίσει. Το πολύπλευρο γνωστικό αντικείμενο της σχολής, οι κοινωνικές και φιλικές σχέσεις που αναπτύχθηκαν κατά τη διάρκεια των σπουδών, το γενικότερο περιβάλλον της Πολυτεχνειούπολης έθεσαν τις βάσεις όχι μόνο για να γίνω ένας επιστήμονας μηχανικός, αλλά και συνολικά ένας καλύτερος άνθρωπος.

Αρχικά, οφείλω να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου, Χαράλαμπο Σαρίμβη, ο οποίος μου έδωσε την ευκαιρία να εισαχθώ στον κόσμο της έρευνας μέσα από τα πολύ ενδιαφέροντα θέματα ενασχόλησης του εργαστηρίου του. Επίσης, τον ευχαριστώ για την καθοδήγηση κατά την εκπόνηση της εργασίας και την άψογη συνεργασία μας από πλευράς του. Δεν θα μπορούσα να μην ευχαριστήσω, ακόμη, και το μέλος της ομάδας του εργαστηρίου Παντελή Καρατζά, ο οποίος με την εξειδίκευσή του πάνω στη βαθιά μηχανική μάθηση, καθώς και την ανιδιοτέλεια και τη συνεργατικότητα που τον διακρίνει, με βοήθησε να ενσωματώσω τις απαραίτητες δεξιότητες και γνώσεις για την εκπόνηση της εργασίας.

Επίσης, ευχαριστώ θερμά τους φίλους μου, και ιδιαίτερα τη φίλη μου Βέρα, για την πολύτιμη βοήθεια και υποστήριξη κατά τη διάρκεια των προπτυχιακών μου σπουδών.

Τέλος, το μεγαλύτερο ευχαριστώ οφείλω στους γονείς μου, στην αδερφή μου Νικολέττα, στον παππού και στην γιαγιά μου, δίχως την στήριξη των οποίων σε όλα τα επίπεδα δεν θα τα είχα καταφέρει.

Περιεχόμενα

Εισαγωγή (Introduction)	14
1 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks)	16
1.1 Τεχνητή Νοημοσύνη και Μηχανική Μάθηση (Artificial Intelligence and Machine Learning)	16
1.2 Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα (Introduction to Artificial Neural Networks)	17
1.2.1 Γραμμική Παλινδρόμηση (Linear Regression)	18
1.2.2 Αλγόριθμοι Βελτιστοποίησης (Optimization Algorithms)	18
1.2.3 Νευρωνικά Δίκτυα Εμπρόσθιας Τροφοδότησης (Feedforward Neural Networks)	23
1.2.4 Τεχνητοί Νευρώνες σε Σύγκριση με Βιολογικούς Νευρώνες (Artificial Neurons in Comparison to Biological Neurons)	25
1.3 Εκπαίδευση Τεχνητών Νευρωνικών Δικτύων (Artificial Neural Networks Training)	26
1.3.1 Αλγόριθμος Οπισθο-Διάδοσης (Back-Propagation Algorithm)	26
1.3.2 Συναρτήσεις Κόστους (Cost Functions)	28
1.3.3 Συναρτήσεις Ενεργοποίησης (Activation Functions)	29
1.3.4 Φθορά Βαρών και Dropout ως Μέθοδοι Αποφυγής Υπερπροσαρμογής (Weight Decay and Dropout as Methods to Prevent Overfitting)	33
1.4 Το Θεώρημα Καθολικής Προσέγγισης (Universal Approximation Theorem)	37
2 Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks)	39
2.1 Οι Εικόνες ως Δεδομένα Εισόδου σε Νευρωνικά Δίκτυα (Images as Neural Networks Input Data)	39
2.2 Η Διαδικασία της Συνέλιξης (Convolution Operation)	41
2.3 Στρώματα ενός Συνελκτικού Νευρωνικού Δικτύου (Layers of a Convolutional Neural Network)	45
2.3.1 Το Συνελκτικό Στρώμα (Convolutional Layer)	45
2.3.2 Το Στρώμα Υποδειγματοληψίας (Pooling Layer)	46
2.3.3 Το Πλήρως Συνδεδεμένο Στρώμα (Fully Connected Layer)	47
2.4 Μέθοδοι Βελτίωσης Εκπαίδευσης Βαθιών Νευρωνικών Δικτύων (Methods for Improving Training of Deep Neural Networks)	48
2.4.1 Κανονικοποίηση Παρτίδας (Batch Normalization)	48
2.4.2 Ενίσχυση Δεδομένων (Data Augmentation)	48
3 Νευρωνικά Δίκτυα Γράφων (Graph Neural Networks)	50
3.1 Θεωρία Γράφων (Graph Theory)	50

3.1.1	Γειτνίαση και Πρόσπτωση στους Γράφους (Graph Adjacency and Incidency)	51
3.1.2	Εκχώρηση Πληροφοριών στους Γράφους (Assigning Graph Attributes)	52
3.1.3	Στοιχεία Γραμμικής Άλγεβρας στους Γράφους (Graph Theory and Linear Algebra)	53
3.1.4	Άλλες Διακρίσεις των Γράφων (Various Graph Types)	54
3.2	Από τα Μοριακά Αποτυπώματα στα Συνελικτικά Δίκτυα σε Γράφους (From Molecular Fingerprints to Graph Convolutions)	55
3.2.1	Μοριακά Αποτυπώματα (Molecular Fingerprints)	55
3.2.2	Αποτυπώματα Duvenaud (Duvenaud Fingerprints)	58
3.3	Νευρωνική Ανταλλαγή Μηνυμάτων και Νευρωνικά Δίκτυα Γράφων (Neural Message Passing and Graph Neural Networks)	60
3.4	Μοντέλα Νευρωνικών Δικτύων Γράφων (Graph Neural Networks Models)	62
3.4.1	Συνελικτικό Δίκτυο Γράφων (Graph Convolutional Network)	62
3.4.2	Υποδειγματοληψία Γράφων (Graph Pooling)	67
3.4.3	Πλήρως Συνδεδεμένα Στρώματα (Fully-Connected layers)	67
4	Ποσοτικές Σχέσεις Δομής - Δράσης (Quantitative Structure Activity Relationships - QSAR)	68
4.1	Η Εξέλιξη των QSAR Μοντέλων (Evolution of QSAR Models)	68
4.2	QSAR Μοντέλα Βαθιάς Μηχανικής Μάθησης (Deep Learning QSAR Models)	70
5	Μελέτη Περίπτωσης - Ενδοκρινικοί Διαταράκτες (Case Study - Endocrine Disruptors)	72
5.1	Το Σύνολο των Δεδομένων (Dataset)	74
5.2	Το Σύστημα Απλοποιημένης Μοριακής Γραμμικής Γραφής (Simplified Molecular-Input Line-Entry System)	75
6	Ανάπτυξη Μοντέλων και Αποτελέσματα (Models Development and Results)	78
6.1	Μέτρα Αξιολόγησης (Evaluation Metrics)	78
6.1.1	Πίνακας Σύγχυσης (Confusion Matrix)	78
6.1.2	Ακρίβεια (Accuracy)	78
6.1.3	Συντελεστής Συσχέτισης Matthews (Matthews Correlation Coefficient)	79
6.2	Διασταυρούμενη Επικύρωση (Cross-Validation)	79
6.3	Προηγούμενη Εργασία (Previous Work)	80
6.4	Μοντέλο Νευρωνικών Δικτύων Εμπρόσθιας Τροφοδότησης (Feed-Forward Neural Networks Model)	82
6.4.1	Κατασκευή Δεδομένων Εισόδου (Input Data Construction)	82
6.4.2	Ανάπτυξη Μοντέλου (Model Development)	85
6.4.3	Αποτελέσματα (Results)	89
6.5	Μοντέλο Νευρωνικών Δικτύων Γράφων (Graph Neural Networks Model)	93
6.5.1	Κατασκευή Δεδομένων Εισόδου (Input Data Construction)	93
6.5.2	Ανάπτυξη Μοντέλου (Model Development)	94
6.5.3	Αποτελέσματα (Results)	103
7	Συμπεράσματα και Προτάσεις για Μελλοντική Έρευνα (Conclusions and Suggestions for Future Research)	107
7.1	Συμπεράσματα (Conclusions)	107
7.2	Προτάσεις για Μελλοντική Έρευνα (Suggestions for Future Research)	108
Α'	Ακρωνύμια και συντομογραφίες	110

Β' Μπεϋζιανή Βελτιστοποίηση (Bayes Optimization)	111
Βιβλιογραφία	114

Κατάλογος Σχημάτων

1.1	Κατηγορίες και εφαρμογές μηχανικής μάθησης	17
1.2	Ο αλγόριθμος Gradient Descent	19
1.3	Κρίσιμα σημεία συναρτήσεων 2 μεταβλητών	21
1.4	Η ορμή στους αλγόριθμος βελτιστοποίησης Gradient Descent	22
1.5	Το γραμμικό νευρωνικό δίκτυο	24
1.6	Το νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης	24
1.7	Βιολογικός vs τεχνητός νευρώνας	26
1.8	Υπολογιστικός γράφος MLP	27
1.9	Γραφική παράσταση της σιγμοειδούς συνάρτησης	30
1.10	Γραφική παράσταση της συνάρτησης υπερβολικής εφαπτομένης	31
1.11	Γραφική παράσταση της συνάρτησης ReLU	31
1.12	Γραφική παράσταση της συνάρτησης PReLU	32
1.13	Υποπροσαρμογή και υπερπροσαρμογή	33
1.14	Τιμή κόστους κατά την εκπαίδευση του νευρωνικού δικτύου	34
1.15	Εφαρμογή Dropout	37
1.16	Βαθύ νευρωνικό δίκτυο	38
2.1	X/Y πλέγμα εικονοστοιχείων	39
2.2	RGB κανάλια επιλεγμένων εικονοστοιχείων	40
2.3	Συνέλιξη χωρίς περιστροφή πυρήνα	43
2.4	Εφαρμογή πυρήνα(ή φίλτρου) σε φωτογραφία	43
2.5	Συνέλιξη με padding μιας σειράς	44
2.6	Συνέλιξη με επιλεγμένο stride	45
2.7	Συνελκτικό στρώμα	46
2.8	Max pooling	47
3.1	Παράδειγμα γράφου με 4 κόμβους και 6 ακμές	51
3.2	Παράδειγμα κατευθυνόμενου γράφου	51
3.3	Χημικός γράφος με χαρακτηριστικά	53
3.4	Παράδειγμα αποτυπώματος μονοπατιού	56
3.5	Παράδειγμα κυκλικού αποτυπώματος	57
3.6	Κυκλικές γειτονίες ενός μορίου	58
3.7	Δισδιάστατη συνέλιξη σε εικόνα vs μη-ευκλείδεια συνέλιξη σε γράφο	61
3.8	Στρώμα νευρωνικού δικτύου ανταλλαγής μηνυμάτων	62
3.9	Παράδειγμα απλού μη-κατευθυνόμενου γράφου	63
4.1	Βήματα ανάπτυξης κλασικού QSAR μοντέλου	70
5.1	Ενδοκρινική διαταραχή	73
6.1	Πίνακας σύγκρισης για πρόβλημα ταξινόμησης 3 κλάσεων.	79

6.2 Παραδείγματα εικόνων που χρησιμοποιούνται ως είσοδοι στο συνελκτικό νευρωνικό δίκτυο.	80
6.3 Μοντέλο CNN: αποτελέσματα ακρίβειας συναρτήσει εποχών στα δεδομένα εκπαίδευσης	81
6.4 Μοντέλο CNN: αποτελέσματα ακρίβειας συναρτήσει εποχών στα δεδομένα επαλήθευσης	81
6.5 Μοντέλο MLP: Ακρίβεια πρόβλεψης στα δεδομένα εκπαίδευσης για διάφορες επιλογές αλγορίθμων βελτιστοποίησης	86
6.6 Μοντέλο MLP: Ακρίβεια πρόβλεψης στα δεδομένα επαλήθευσης για διάφορες επιλογές αλγορίθμων βελτιστοποίησης	87
6.7 Μοντέλο MLP: Ακρίβεια πρόβλεψης στο σύνολο δεδομένων εκπαίδευσης για εκπαίδευση με μεγέθη παρτίδας 64, 128, 256	88
6.8 Μοντέλο MLP: Ακρίβεια πρόβλεψης στο σύνολο δεδομένων επαλήθευσης για εκπαίδευση με μεγέθη παρτίδας 64, 128, 256	88
6.9 Μοντέλο MLP: αποτελέσματα 5-fold cross validation ακρίβειας συναρτήσει εποχών στα δεδομένα εκπαίδευσης	90
6.10 Μοντέλο MLP: αποτελέσματα 5-fold cross validation ακρίβειας συναρτήσει εποχών στα δεδομένα επαλήθευσης	90
6.11 Διαδικασία κατασκευής δεδομένων εισόδου μοντέλου νευρωνικών δικτύων γράφων	94
6.12 Μοντέλο GNN: Ακρίβεια πρόβλεψης στα δεδομένα εκπαίδευσης για διάφορες επιλογές ρυθμών μάθησης	98
6.13 Μοντέλο GNN: Ακρίβεια πρόβλεψης στα δεδομένα επαλήθευσης για διάφορες επιλογές ρυθμών μάθησης	98
6.14 Μοντέλο GNN: Ακρίβεια πρόβλεψης στα δεδομένα εκπαίδευσης για επιλογές αλγορίθμων βελτιστοποίησης	99
6.15 Μοντέλο GNN: Ακρίβεια πρόβλεψης στα δεδομένα επαλήθευσης για επιλογές αλγορίθμων βελτιστοποίησης	100
6.16 Μοντέλο GNN: Ακρίβεια πρόβλεψης στα δεδομένα εκπαίδευσης για επιλογές πλήθους συνελκτικών στρωμάτων γράφων	101
6.17 Μοντέλο GNN: Ακρίβεια πρόβλεψης στα δεδομένα επαλήθευσης για επιλογές πλήθους συνελκτικών στρωμάτων γράφων	101
6.18 Αρχιτεκτονική μοντέλου GNN	103
6.19 Μοντέλο GNN: αποτελέσματα 5-fold cross validation ακρίβειας συναρτήσει εποχών στα δεδομένα εκπαίδευσης	104
6.20 Μοντέλο GNN: αποτελέσματα 5-fold cross validation ακρίβειας συναρτήσει εποχών στα δεδομένα επαλήθευσης	104

Κατάλογος Πινάκων

5.1	Πίνακας κλάσεων του προβλήματος ταξινόμησης ενδοκρινικών διαταρακτών	75
6.1	Πίνακας σύγχυσης του μοντέλου CNN	82
6.2	Μοριακές περιγραφικές μεταβλητές για τα δεδομένα εισόδου του μοντέλου MLP, όπως παράγονται μέσω του λογισμικού Mordred	83
6.3	Εύρος τιμών των υπερπαραμέτρων του μοντέλου MLP	85
6.4	Οι τελικές υπερπαραμέτροι του μοντέλου MLP	89
6.5	Αποτελέσματα ακρίβειας και MCC του μοντέλου MLP	91
6.6	Πίνακας σύγχυσης του μοντέλου MLP	91
6.7	Κανονικοποιημένος πίνακας σύγχυσης του μοντέλου MLP	92
6.8	Χαρακτηριστικά κόμβων/ατόμων του χημικού γράφου	93
6.9	Εύρος τιμών των υπερπαραμέτρων του μοντέλου GNN	97
6.10	Τελικές υπερπαραμέτροι του μοντέλου GNN	102
6.11	Αποτελέσματα ακρίβειας και MCC του μοντέλου GNN	104
6.12	Πίνακας σύγχυσης του μοντέλου GNN	105
6.13	Κανονικοποιημένος πίνακας σύγχυσης του μοντέλου GNN	105

Εισαγωγή (Introduction)

Με την τεράστια τεχνολογική πρόοδο των τελευταίων δεκαετιών, ο όγκος των πληροφοριών που είναι διαθέσιμες στο ευρύ κοινό όλο και αυξάνεται. Είναι αναπόφευκτη, λοιπόν, η ανάπτυξη υπολογιστικών εργαλείων, τα οποία θα επεξεργάζονται με κατάλληλο τρόπο τις πληροφορίες αυτές με σκοπό την εξαγωγή χρήσιμης γνώσης που θα βελτιώνει την ανθρώπινη ζωή. Φυσικά, τα εργαλεία αυτά της πληροφορικής δεν θα μπορούσαν να μην αφορούν και επιστημονικούς τομείς, όπως η χημεία, η βιολογία και η επιστήμη των υλικών. Έτσι, ο συνδυασμός των παραπάνω με αλγορίθμους τεχνητής νοημοσύνης (και συγκεκριμένα, του κλάδου της μηχανικής μάθησης) διαμόρφωσε υπολογιστικές ροές αναγνώρισης προτύπων μεταξύ ενός συνόλου δεδομένων χημικών μορίων, βιομορίων ή νανοϋλικών. Μάλιστα, πολύ συχνά οι ροές αυτές είναι ικανές να αντικαταστήσουν τις δαπανηρές, χρονοβόρες και συχνά αντιδεοντολογικές (λόγω των in-vivo δοκιμών σε ζώα) πειραματικές διαδικασίες. Ταυτόχρονα, τα παραγόμενα μοντέλα μηχανικής μάθησης μπορούν αποδεδειγμένα να παράξουν νέα γνώση, η οποία θα ήταν αδύνατο να παραχθεί εφαρμόζοντας κλασικές μεθοδολογίες.

Η πρόβλεψη ιδιοτήτων χημικών ουσιών με χρήση αλγορίθμων μηχανικής μάθησης αποτελεί ένα παράδειγμα σύμπραξης της πληροφορικής με τη χημεία και τη βιολογία. Μοντέλα, τα οποία συνδέουν διάφορα χαρακτηριστικά της 2D ή 3D δομής των μορίων με σκοπό την πρόβλεψη βιολογικής δραστηριότητας (QSAR μοντέλα) ξεκίνησαν ήδη να αναπτύσσονται στα μέσα του προηγούμενου αιώνα. Σήμερα, η ανάπτυξη τους ενθαρρύνεται από τον Ευρωπαϊκό Οργανισμό Χημικών (European Chemicals Agency - ECHA) υπό την αιγίδα της Ευρωπαϊκής Ένωσης σύμφωνα με τον κανονισμό REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals), αλλά και από άλλους οργανισμούς, προσφέροντας μάλιστα πληθώρα δεδομένων βιολογικής απόκρισης (π.χ. τοξικότητα) μέσω ηλεκτρονικών βάσεων δεδομένων.

Ταυτόχρονα, ένας κλάδος της μηχανικής μάθησης που καλείται βαθιά μάθηση (deep learning) προσφέρει τα τελευταία χρόνια όλο και περισσότερα υπολογιστικά εργαλεία που σχετίζονται με εφαρμογές σε πεδία, όπως η υπολογιστική όραση, η αναγνώριση ομιλίας, η επεξεργασία φυσικής γλώσσας αλλά και η βιοπληροφορική, η ανακάλυψη φαρμάκων και η ανάλυση ιατρικών εικόνων. Στην παρούσα εργασία επιχειρείται η χρήση της βαθιάς μάθησης μέσω διαφόρων τύπων νευρωνικών δικτύων για την δημιουργία μοντέλων που θα προβλέπουν τις ιδιότητες χημικών ουσιών λαμβάνοντας ως είσοδο δεδομένα που χαρακτηρίζουν δομικά τα εν λόγω μόρια. Στην περίπτωση, μάλιστα, του υπό μελέτη συνόλου δεδομένων, θα μπορούσαμε να κατατάξουμε τα μοντέλα που παράγονται στην κατηγορία των μοντέλων QSAR, αφού η ιδιότητα που προβλέπεται είναι το δυναμικό πρόσδεσης χημικών ουσιών στο ανθρώπινο ενδοκρινικό σύστημα, δηλαδή μια μορφή βιολογικής δραστηριότητας.

Πιο συγκεκριμένα, παράγονται μοντέλα βαθιών νευρωνικών δικτύων εμπρόσθιας τροφοδότησης και βαθιών νευρωνικών δικτύων γράφων, η απόδοση των οποίων αξιολογείται,

αλλά και συγκρίνεται με την απόδοση μοντέλου βαθιών συνελκτικών νευρωνικών δικτύων με είσοδο εικόνες χημικών μορίων. Τα δεδομένα εισόδου στο μοντέλο εμπρόσθιων νευρωνικών δικτύων είναι μοριακές περιγραφικές μεταβλητές, όπως ισχύει στο σύνολο των κλασικών QSAR μοντελοποιήσεων. Από την άλλη, στα ταχύτατα εξελισσόμενα νευρωνικά δίκτυα γράφων, είσοδο αποτελούν τα χημικά μόρια σε μορφή χημικών γράφων, όπου οι κόμβοι αποτελούν τα άτομα και οι ακμές τους δεσμούς, συγκροτώντας, έτσι, μια ρεαλιστική και στιβαρή μη-ευκλείδεια απεικόνιση της δομής χημικών ουσιών.

Η παρούσα εργασία ξεκινά με την αναλυτική παρουσίαση του θεωρητικού υπόβαθρου για τα είδη των νευρωνικών δικτύων, πάνω στα οποία βασίζονται τα μαθηματικά μοντέλα που αναπτύσσονται. Στη συνέχεια, δίνονται πληροφορίες σχετικά με τις μοντελοποιήσεις QSAR και τη συμβολή σε αυτές σύγχρονων υπολογιστικών ροών βαθιάς μηχανικής μάθησης. Επίσης, παρουσιάζεται το σύνολο δεδομένων ενδοκρινικών διαταρακτών και ο τρόπος κωδικοποίησής τους στην διαδικτυακή βάση δεδομένων. Τέλος, περιγράφεται η μεθοδολογία ανάπτυξης των μοντέλων, καθώς και η αξιολόγησή τους μέσα από διάφορες μετρικές τιμές. Η διεκπεραίωση του σκοπού ταξινόμησης των χημικών ουσιών έλαβε χώρα με χρήση του διαδικτυακού περιβάλλοντος Google Colaboratory, μέσα από το οποίο προσφέρονται δωρεάν (με ορισμένους περιορισμούς) προς χρήση ισχυρές μονάδες επεξεργαστών (CPU) και μονάδες καρτών γραφικών (GPU). Οι υπολογιστικοί πόροι αυτοί κρίθηκαν απαραίτητοι για την άρτια ανάπτυξη και επικύρωση των μοντέλων σε βιώσιμους υπολογιστικούς χρόνους.

Κεφάλαιο 1

Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks)

1.1 Τεχνητή Νοημοσύνη και Μηχανική Μάθηση (Artificial Intelligence and Machine Learning)

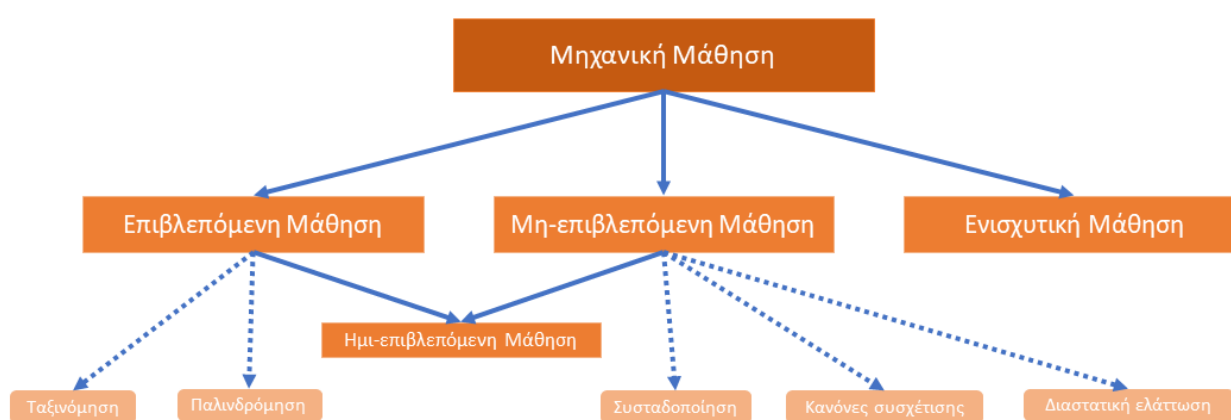
Για πολλές χιλιετίες οι άνθρωποι ονειρεύονταν την δημιουργία μηχανών που μπορούν να σκέφτονται. Χαρακτηριστικό παράδειγμα αποτελεί ο Τάλως, μια μυθική μεταλλική ανθρωπόμορφη μηχανή *προγραμματισμένη* από τον Ήφαιστο να φυλάει το νησί της Κρήτης. Σήμερα, η ικανότητα μίμησης από υπολογιστές πολλών στοιχείων της ανθρώπινης ευφυΐας, όπως επίτευξη συλλογιστικής πορείας, μάθησης, σχεδιασμού, αντίληψης και επεξεργασία φυσικής γλώσσας αποτελεί μια υπαρκτή και ταχύτατα εξελισσόμενη τεχνολογία. Η αναγνώριση ομιλίας και εικόνων, τα κοινωνικά δίκτυα, η ιατρική και η ρομποτική είναι ορισμένοι μόνο από τους τομείς της ανθρώπινης δραστηριότητας όπου η τεχνητή νοημοσύνη έχει βρει εφαρμογή τα τελευταία χρόνια.

Η μηχανική μάθηση είναι μια υποκατηγορία της τεχνητής νοημοσύνης, η οποία μελετά την ανάπτυξη αλγορίθμων που βελτιώνονται αυτόματα μέσω της εμπειρίας από τη χρήση δεδομένων. Οι αλγόριθμοι αυτοί κατασκευάζουν ένα μοντέλο, το οποίο είναι ικανό να "εκπαιδεύεται" από ένα σύνολο δεδομένων, γνωστό ως "δεδομένα εκπαίδευσης", προκειμένου να πραγματοποιεί συγκεκριμένες προβλέψεις ή να λαμβάνει αποφάσεις χωρίς να είναι ρητά προγραμματισμένο για αυτόν τον σκοπό. [1] Η μηχανική μάθηση χωρίζεται σε 3 βασικές κατηγορίες:

- **Επιβλεπόμενη Μάθηση (Supervised Learning)**, όπου το μοντέλο εκπαιδεύεται σε δεδομένα εκπαίδευσης, τα οποία περιλαμβάνουν τις εισόδους με τις αντίστοιχες επιθυμητές εξόδους. Στόχος του μοντέλου είναι να προβλέψει σωστά τις προκαθορισμένες εξόδους χρησιμοποιώντας διάφορες μεθόδους, όπως τα τεχνητά νευρωνικά δίκτυα.
- **Μη-επιβλεπόμενη Μάθηση (Unsupervised Learning)**, όπου παρέχονται μόνο δεδομένα εισόδου χωρίς τις αντίστοιχες εξόδους. Τα μοντέλα μη-επιβλεπόμενης μάθησης χρησιμοποιούν αλγορίθμους μηχανικής μάθησης με σκοπό την εύρεση κρυφών μοτίβων (ομοιοτήτων ή διαφορών) μεταξύ των δεδομένων εισόδου, καθώς και την ομαδοποίησή τους (προβλήματα συσταδοποίησης). [3]
 - **Ημί-επιβλεπόμενη Μάθηση (Semi-supervised Learning)**, η οποία είναι μια προσέγγιση της μηχανικής μάθησης που συνδυάζει στοιχεία από την επιβλεπόμενη και μη-επιβλεπόμενη μάθηση. Στην περίπτωση αυτή, το μοντέλο εκπαιδεύεται

με χρήση ενός μικρού συνόλου δεδομένων με ετικέτα (είσοδοι με τις αντίστοιχες επιθυμητές εξόδους) σε συνδυασμό με ένα μεγάλο σύνολο δεδομένων χωρίς ετικέτα έχοντας ως σκοπό την πρόβλεψη των άγνωστων ετικετών. [2]

- Ενισχυτική μάθηση (Reinforcement Learning), όπου το μοντέλο εκπαιδεύεται μέσα από άμεση αλληλεπίδραση με το περιβάλλον ανατροφοδοτούμενο από τα αποτελέσματα των δικών του ενεργειών, δηλαδή σήμα επιβράβευσης για "θετική συμπεριφορά" (σωστή πρόβλεψη) και σήμα τιμωρίας για "αρνητική συμπεριφορά" (λάθος πρόβλεψη). Η διαφορά με την επιβλεπόμενη μάθηση έγκειται στην πηγή της ανατροφοδότησης του μοντέλου (στην περίπτωση της επιβλεπόμενης μάθησης είναι οι προκαθορισμένες σωστές ετικέτες των δεδομένων εισόδου), ενώ η διαφορά με την μη-επιβλεπόμενη μάθηση έγκειται στους διαφορετικούς στόχους (η ενισχυτική μάθηση επιδιώκει το σχεδιασμό ενεργειών που θα μεγιστοποιήσουν την συνολική επιβράβευση) [2].



Σχήμα 1.1: Διαγραμματική απεικόνιση κατηγοριών και εφαρμογών της μηχανικής μάθησης

Η παρούσα εργασία στοχεύει στην ανάπτυξη μοντέλων επιβλεπόμενης μάθησης με χρήση Τεχνητών Νευρωνικών Δικτύων.

1.2 Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα (Introduction to Artificial Neural Networks)

Η ιδέα των τεχνητών νευρωνικών δικτύων προέρχεται, όπως φανερώνει η ονομασία τους, από το νευρικό σύστημα του ανθρώπου. Το νευρικό σύστημα είναι υπεύθυνο για τον έλεγχο και λειτουργία όλων των σωματικών και ψυχικών λειτουργιών, καθώς επίσης και για την αλληλεπίδραση του ανθρώπου με το περιβάλλον. Σε κυτταρικό επίπεδο, το νευρικό σύστημα αποτελείται από νευρώνες, οι οποίοι είναι δομές που επιτρέπουν την γρήγορη μεταφορά ηλεκτροχημικών σημάτων σε άλλα κύτταρα. Παίρνοντας ως παράδειγμα την αίσθηση της όρασης, το τμήμα του εγκεφάλου που λαμβάνει και επεξεργάζεται οπτικά ερεθίσματα ονομάζεται οπτικός φλοιός και αποτελείται από 5 περιοχές (πρωτοταγής οπτικός φλοιός ή V1, V2, V3, V4, V5). Η περιοχή V1 από μόνη της περιλαμβάνει περίπου 140 εκατομμύρια νευρώνες σε κάθε ημισφαίριο του εγκεφάλου με δεκάδες δισεκατομμύρια συνδέσεις μεταξύ τους. Και όμως, η επεξεργασία του οπτικού ερεθίσματος, δηλαδή της εικόνας που ένας άνθρωπος βλέπει χρησιμοποιώντας τα μάτια του, λαμβάνει μέρος και στις πέντε περιοχές με πολύπλοκες συνδέσεις προς τα εμπρός (feedforward connections) ή και συνδέσεις ανατροφοδότησης (feedback connections) μεταξύ των επιμέρους νευρώνων. Συμπεραίνει κανείς ότι το νευρικό σύστημα του ανθρώπου είναι ένας υπερυπολογιστής

που μπορεί να αντιλαμβάνεται ένα τεράστιο πλήθος οπτικών, και όχι μόνο, ερεθισμάτων, όπως επίσης, να μαθαίνει από αυτά και να προσαρμόζεται σε αυτά. Τα συστήματα νευρωνικών δικτύων, λοιπόν, προκύπτουν από την απλουστευμένη ψηφιακή μοντελοποίηση του νευρικού συστήματος και πιο συγκεκριμένα, των νευρώνων και σκοπεύουν στην επίτευξη προσαρμοστικής μάθησης αυτή τη φορά από τις μηχανές και όχι από τον ανθρώπινο εγκέφαλο. Δίνεται, λοιπόν, ένας πρωταρχικός ορισμός των νευρωνικών δικτύων:

Ένα (τεχνητό) νευρωνικό δίκτυο είναι ένα διασυνδεδεμένο συγκρότημα από απλά στοιχεία επεξεργασίας, τους κόμβους (ή νευρώνες), των οποίων η λειτουργία βασίζεται σε αυτή των ζωικών νευρώνων. Η ικανότητα επεξεργασίας του δικτύου είναι αποθηκευμένη στα βάρη των συνδέσεων μεταξύ των νευρώνων, τα οποία υπολογίζονται από μια διαδικασία προσαρμογής σε (ή μάθησης από) ένα σύνολο δεδομένων εκπαίδευσης [4].

1.2.1 Γραμμική Παλινδρόμηση (Linear Regression)

Ο όρος παλινδρόμηση αναφέρεται στις μεθόδους που προτυποποιούν τη σχέση μεταξύ μίας ή περισσότερων ανεξάρτητων μεταβλητών και μίας εξαρτημένης μεταβλητής. Στις φυσικές και κοινωνικές επιστήμες, σκοπός της παλινδρόμησης είναι, συνήθως, ο χαρακτηρισμός της σχέσης μεταξύ εισόδων (inputs) και εξόδων (outputs). Στην μηχανική μάθηση, από την άλλη πλευρά, η παλινδρόμηση χρησιμοποιείται κυρίως για πρόβλεψη (prediction). [5]

Για την περίπτωση της γραμμικής παλινδρόμησης γίνονται, αρχικά, οι παραδοχές ότι η σχέση που συνδέει τις ανεξάρτητες μεταβλητές \mathbf{x} και την εξαρτημένη μεταβλητή y είναι γραμμική, καθώς και ότι ο θόρυβος ακολουθεί κανονική (γκουαουσιανή) κατανομή. Ένα σύνολο δεδομένων εκπαίδευσης, τυπικά, περιλαμβάνει d ανεξάρτητες μεταβλητές που καλούνται χαρακτηριστικά (features), n σειρές από αριθμητικές τιμές που αντιστοιχούν στις ανεξάρτητες μεταβλητές και καλούνται παραδείγματα (examples), καθώς και n αριθμητικές τιμές που ανά μία αντιστοιχούν στην εξαρτημένη μεταβλητή και καλούνται ετικέτες (labels). Οι ετικέτες αποτελούν και τις πραγματικές τιμές εξόδου που το μοντέλο μέσω γραμμικής παλινδρόμησης προσπαθεί να προβλέψει. [5] Συγκεντρωτικά, τα παραδείγματα παρουσιάζονται μέσω του πίνακα σχεδιασμού (design matrix) $\mathbf{X} \in \mathbb{R}^{n \times d}$. Οι ετικέτες συμβολίζονται με το διάνυσμα $\mathbf{y} \in \mathbb{R}^n$, ενώ οι προβλέψεις με το διάνυσμα $\hat{\mathbf{y}} \in \mathbb{R}^n$. Σκοπός της παλινδρόμησης είναι η κατάλληλη επιλογή των βαρών (weights) $\mathbf{w} \in \mathbb{R}^d$ και του bias b έτσι ώστε οι προβλέψεις $\hat{\mathbf{y}}$ να είναι όσον το δυνατόν πιο κοντά στις ετικέτες \mathbf{y} . Με βάση τα παραπάνω οι προβλέψεις εκφράζονται σύμφωνα με την εξίσωση:

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{w} + b. \quad (1.1)$$

Η ποσοτικοποίηση της προσαρμογής του μοντέλου, δηλαδή το κατά πόσο οι προβλέψεις πλησιάζουν τις πραγματικές τιμές (ετικέτες) επιτυγχάνεται μέσω μιας συνάρτησης κόστους (loss ή cost function) L . Μια δημοφιλής επιλογή για προβλήματα παλινδρόμησης είναι το Μέσο Τετραγωνικό Σφάλμα (Mean-Squared Error ή MSE) [5]:

$$L = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} (\hat{\mathbf{y}}_i - \mathbf{y}_i)^2 = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} (\mathbf{w}^T \mathbf{x}_i + b - \mathbf{y}_i)^2. \quad (1.2)$$

1.2.2 Αλγόριθμοι Βελτιστοποίησης (Optimization Algorithms)

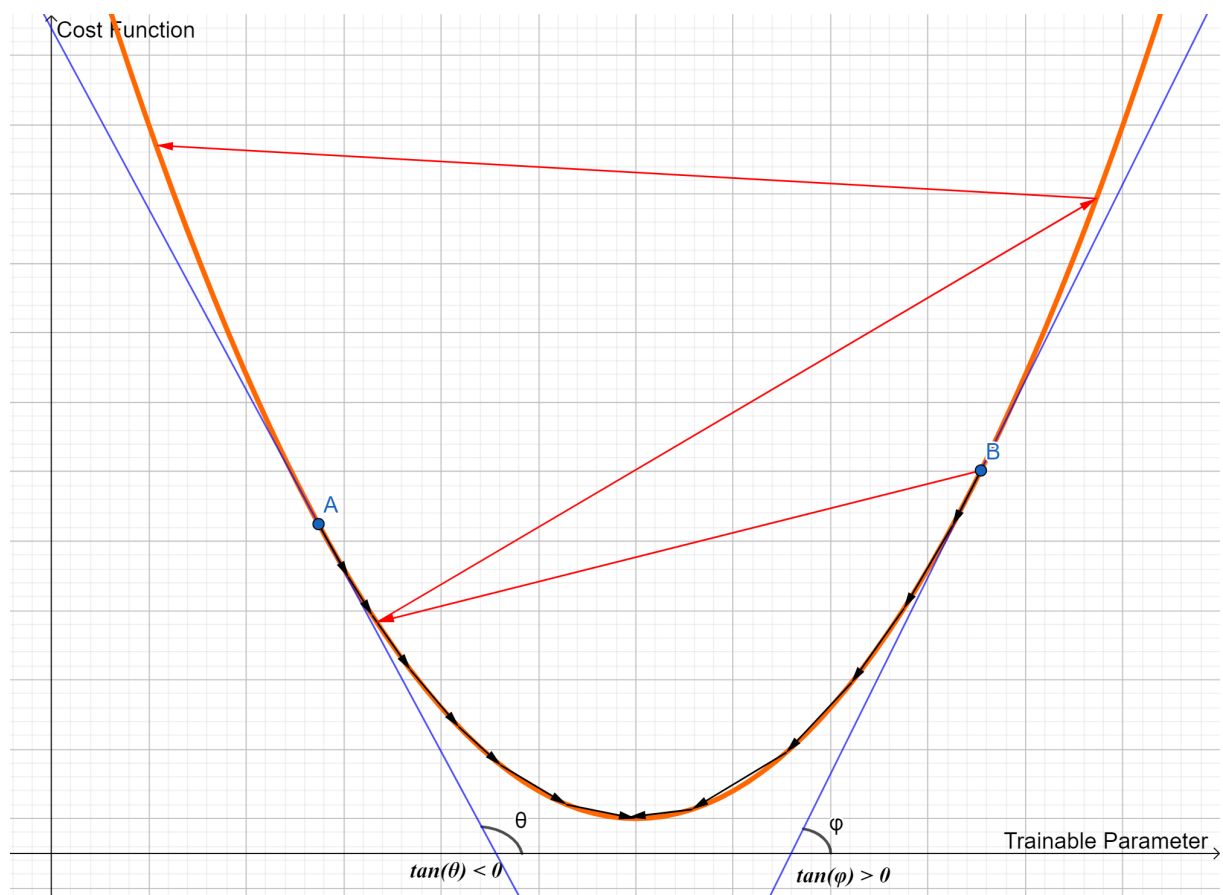
Στόχος είναι η εύρεση των παραμέτρων \mathbf{w} και b , για τις οποίες το μέσο τετραγωνικό σφάλμα ελαχιστοποιείται. Για το παράδειγμα της γραμμικής παλινδρόμησης εάν αμε-

ληθεί ο όρος b η λύση βρίσκεται εύκολα μέσω αναλυτικής παραγωγίσης και είναι ίση με $\mathbf{w}^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$. Παρ' όλα αυτά, στις περισσότερες περιπτώσεις μοντέλων μηχανικής μάθησης η αναλυτική εύρεση του ελάχιστου της συνάρτησης κόστους είναι αδύνατη. Έτσι, το επιθυμητό βάρος υπολογίζεται αριθμητικά με χρήση του αλγορίθμου βελτιστοποίησης κατάβασης βαθμίδας (Gradient Descent) ή παραλλαγών του. [6] Ο αλγόριθμος Gradient Descent, ο οποίος είναι ένας αλγόριθμος εύρεσης τοπικού ελάχιστου διαφορίσιμων συναρτήσεων, εκφράζεται μαθηματικά για το παράδειγμα της γραμμικής παλινδρόμησης ως εξής:

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \frac{\partial L(\mathbf{w}, b)}{\partial \mathbf{w}}, \quad (1.3)$$

$$b \leftarrow b - \eta \frac{\partial L(\mathbf{w}, b)}{\partial b}. \quad (1.4)$$

Σε κάθε επανάληψη οι εκπαιδευσιμες παράμετροι (trainable parameters) ανανεώνονται με αφαίρεση της μερικής παραγώγου της συνάρτησης κόστους ως προς αυτές πολλαπλασιασμένη με έναν θετικό αριθμό η που ονομάζεται ρυθμός μάθησης (learning rate). Ο αλγόριθμος τερματίζεται από τον χρήστη όταν η συνάρτηση κόστους λάβει μια μικρή θετική τιμή (συγκλίνει στο 0). Η ιδέα πίσω από τον αλγόριθμο βρίσκεται στην "κατηφορική" κίνηση κατά μήκος της επιφάνειας της συνάρτησης κόστους, δηλαδή με φορά που έχει αντίθετο πρόσημο από αυτό της κλίσης ως προς την εκπαιδευσιμη παράμετρο. [5] Διαγραμματικά, ο αλγόριθμος Gradient Descent για το απλοποιημένο παράδειγμα κυρτής συνάρτησης κόστους μιας ανεξάρτητης μεταβλητής (μονοδιάστατο πρόβλημα) αποτυπώνεται ως εξής:



Σχήμα 1.2: Απεικόνιση του αλγορίθμου βελτιστοποίησης Gradient Descent.

Όπως φαίνεται στο σχήμα 1.2, ξεκινώντας από 2 τυχαία σημεία A και B (με αρνητική και θετική κλίση ως προς την εκπαιδευσιμη παράμετρο αντίστοιχα) ο αλγόριθμος με κατάλληλη επιλογή της υπερπαραμέτρου (hyperparameter) η μπορεί να συγκλίνει στο ελάχιστο της συνάρτησης (μαύρα βέλη). Παρ' όλα αυτά, αν επιλεγεί μια αρκετά μεγάλη τιμή ρυθμού μάθησης ο αλγόριθμος αποκλίνει και με κάθε επανάληψη η λύση απομακρύνεται από την βέλτιστη (κόκκινα βέλη). Εξάγεται, λοιπόν, το συμπέρασμα ότι είναι πολύ σημαντική για την πορεία της εκπαίδευσης του μοντέλου η επιλογή του ρυθμού μάθησης, έτσι ώστε ούτε ο αλγόριθμος να συγκλίνει με πολύ αργό ρυθμό (με επιλογή πολύ μικρού η) ούτε να αποκλίνει (με επιλογή πολύ μεγάλου η). Ορισμένες φορές, μάλιστα, επιλέγεται ο ρυθμός μάθησης να μεταβάλλεται (συνήθως ελαττώνεται) κατά τη διάρκεια της εκπαίδευσης του μοντέλου. [6]

Η μαθηματική προέλευση του αλγορίθμου βρίσκεται στην προσέγγιση της συνάρτησης κόστους μέσω σειράς Taylor πρώτης τάξης. Για διαφορίσιμη συνάρτηση κόστους $f : \mathbb{R} \rightarrow \mathbb{R}$ ισχύει:

$$f(x + \epsilon) = f(x) + \epsilon f'(x) + O(\epsilon^2) . \quad (1.5)$$

Θεωρώντας $\epsilon = -\eta f'(x)$, όπου $\eta > 0$ προκύπτει ότι:

$$f(x - \eta f'(x)) = f(x) - \eta f'^2(x) + O(\eta^2 f'^2(x)) . \quad (1.6)$$

Εφόσον η παράγωγος είναι μη-μηδενική ($f'(x) \neq 0$), τότε με επιλογή μικρού η προκειμένου να μην ληφθούν υπόψη οι υπόλοιποι όροι της σειράς Taylor και με δεδομένο ότι $\eta f'^2(x) > 0$ προκύπτει ότι:

$$f(x - \eta f'(x)) \lesssim f(x) . \quad (1.7)$$

Επομένως, αποδεικνύεται ότι με ανανέωση σε κάθε επανάληψη, $x \leftarrow x - \eta f'(x)$, η τιμή της $f(x)$ πιθανότατα θα μειώνεται και ο στόχος της προσέγγισης του τοπικού ελαχίστου θα επιτυγχάνεται.

Ένα από τα κύρια μειονεκτήματα του αλγορίθμου Gradient Descent είναι το υπολογιστικό της κόστος. Πιο συγκεκριμένα, σε προβλήματα μηχανικής μάθησης η συνολική συνάρτηση κόστους ή αντικειμενική συνάρτηση είναι ο μέσος όρος των συναρτήσεων κόστους για κάθε παράδειγμα του συνόλου δεδομένων εκπαίδευσης. Έτσι, για σύνολο δεδομένων n παραδειγμάτων το υπολογιστικό κόστος σε κάθε επανάληψη για κάθε ανεξάρτητη μεταβλητή είναι $O(n)$, το οποίο αυξάνεται γραμμικά με αύξηση των παραδειγμάτων του συνόλου. Επομένως, για μεγάλα σύνολα δεδομένων που συνήθως χρησιμοποιούνται σε προβλήματα βαθιάς μάθησης θα απαιτείται πολύς χρόνος για την σύγκλιση του αλγορίθμου. [6] Λύση στο πρόβλημα φέρνει η παραλλαγή Stochastic Gradient Descent ή SGD που απεικονίζεται παρακάτω:

Αλγόριθμος SGD

Απαίτηση: η : Ρυθμός μάθησης

Απαίτηση: $f(\theta)$: Αντικειμενική συνάρτηση κόστους με παραμέτρους θ

Απαίτηση: θ_0 : Αρχικό τυχαίο διάνυσμα παραμέτρων

Απαίτηση: $t \leftarrow 0$: Αρχικό χρονικό βήμα

Όσο θ_t δεν συγκλίνει **κάνε:**

$$t \leftarrow t + 1$$

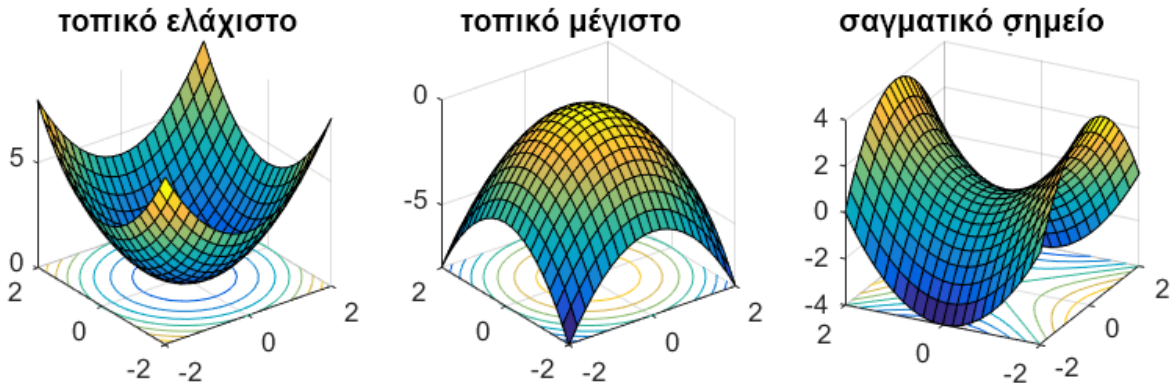
Για $i \in \{1 \dots N\}$ **διάλεξε** i σε τυχαία σειρά:

$$\theta_t \leftarrow \theta_{t-1} - \eta \frac{\partial f(\theta_{t-1}, X_i, y_i)}{\partial \theta_{t-1}}$$

**Τέλος επανάληψης για
Έλεγχε κριτήριο σύγκλισης
Τέλος επανάληψης όσο**

Σε κάθε επανάληψη του SGD, επιλέγεται ομοιόμορφα ένα και μόνο παράδειγμα του συνόλου δεδομένων εκπαίδευσης με βάση το οποίο ανανεώνονται οι εκπαιδευσιμοι παράμετροι. Έτσι, το υπολογιστικό κόστος σε κάθε επανάληψη είναι $O(1)$ αντί για $O(n)$. Για αύξηση της ακρίβειας υπολογισμού της παραγώγου και ταυτόχρονα όχι σημαντική αύξηση του υπολογιστικού κόστους μπορεί σε κάθε επανάληψη να υπολογίζεται ο μέσος όρος των παραγώγων μιας μικρής παρτίδας παραδειγμάτων (Mini-batch Gradient Descent). [6]

Ένα βασικό πρόβλημα που παρατηρείται στους αλγορίθμους Gradient Descent έχει να κάνει με τις συναρτήσεις, των οποίων η κυρτότητα μεταβάλλεται (non-convex functions), όπως είναι οι συναρτήσεις κόστους πολύπλοκων νευρωνικών δικτύων σαν και αυτών που θα μελετηθούν στην παρούσα εργασία. Οι συναρτήσεις αυτές παρουσιάζουν πολλά τοπικά ελάχιστα και επομένως, δυσκολεύει η εύρεση του ολικού ελάχιστου. Η πορεία του αλγορίθμου βελτιστοποίησης θα ήταν ελαττωματική εάν υπήρχαν πολλά τοπικά ελάχιστα με υψηλή τιμή κόστους συγκριτικά με το ολικό ελάχιστο. Παρ' όλα αυτά, παρατηρείται ότι σε προβλήματα πολλών διαστάσεων τα περισσότερα τοπικά ελάχιστα έχουν συνήθως μια χαμηλή τιμή κόστους και έτσι, δεν κρίνεται σημαντική η εύρεση του πραγματικού ολικού ελάχιστου [7]. Από την άλλη, η δυσκολία σύγκλισης του αλγορίθμου έγκειται, κυρίως, στην ύπαρξη σαγματικών σημείων, δηλαδή σημείων στο επίπεδο όπου η μια διεύθυνση έχει θετική κλίση και η άλλη αρνητική και συνήθως περιβάλλονται από περιοχές πλατώ με ίση και σταθερή τιμή συνάρτησης κόστους. Οι αλγόριθμοι Gradient Descent δυσκολεύονται αρκετά να "ξεφύγουν" από αυτές τις περιοχές αφού η παράγωγος είναι μηδενική προς όλες τις διευθύνσεις με αποτέλεσμα πολλές φορές να αποτυγχάνουν να συγκλίνουν σε μια ελάχιστη τιμή. [8]



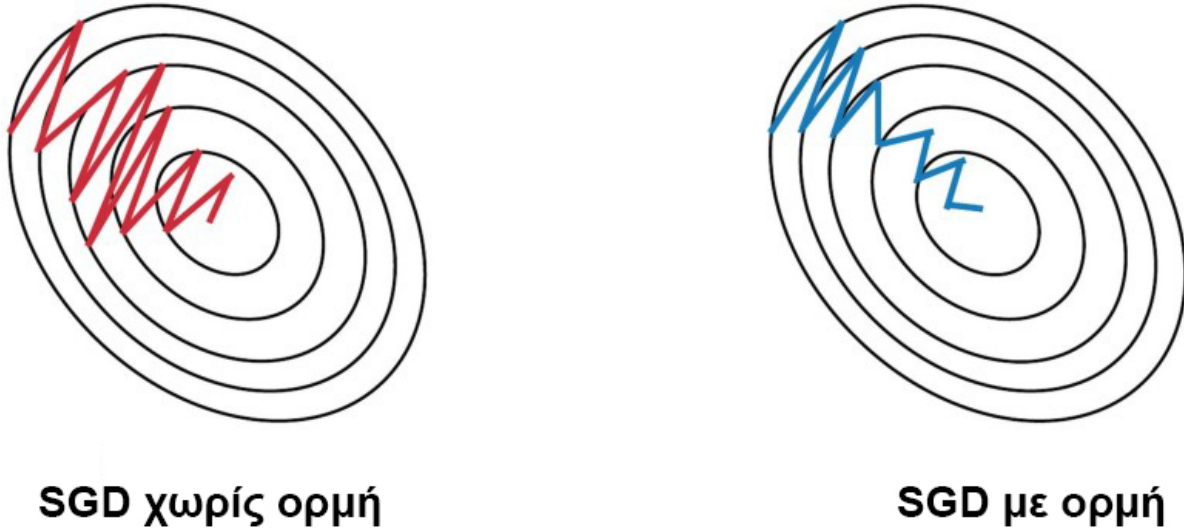
Σχήμα 1.3: Τα κρίσιμα σημεία συναρτήσεων 2 μεταβλητών, μεταξύ των οποίων διακρίνεται και το σαγματικό σημείο. [9]

Για την αντιμετώπιση των προβλημάτων των κλασικών αλγορίθμων Gradient Descent έχουν προταθεί διάφορες τροποποιήσεις. Μία από αυτές είναι η προσθήκη του όρου της Ορμής (Momentum) στο διάνυσμα ανανέωσης της παραμέτρου θ (στους Gradient Descent αλγορίθμους ίσο με $\eta \frac{\partial f(\theta_{t-1})}{\partial \theta_{t-1}}$) [10]:

$$\mathbf{u}_t \leftarrow \gamma \mathbf{u}_{t-1} + \eta \frac{\partial f(\theta_{t-1})}{\partial \theta_{t-1}}, \quad (1.8)$$

$$\theta_t \leftarrow \theta_{t-1} - \mathbf{u}_t . \quad (1.9)$$

Η προσθήκη της ορμής γίνεται μέσω ενός κλάσματος γ (συνήθως ίσο με 0.9) του διανύσματος ανανέωσης (\mathbf{u}) του προηγούμενου βήματος. Η ορμή προσφέρει στον αλγόριθμο υψηλότερες ταχύτητες σύγκλισης και μικρότερη ταλαντωτική συμπεριφορά, ιδιαίτερα σε περιοχές όπου η επιφάνεια καμπυλώνει πολύ πιο απότομα προς τη μία διάσταση απ' ότι στην άλλη. Οι περιοχές αυτές παρομοιάζονται με φαράγγια και βρίσκονται συνήθως γύρω από τοπικά ελάχιστα[6]. Η βελτίωση αυτή φαίνεται στο παρακάτω σχήμα:



Σχήμα 1.4: Βελτίωση στην ταχύτητα σύγκλισης και στην ταλαντωτική συμπεριφορά με χρήση του όρου ορμής [12]

Άλλη μια μέθοδος που βελτιώνει τους αλγορίθμους Gradient Descent είναι αυτή των προσαρμόσιμων ρυθμών μάθησης (adaptive learning rates). Χρησιμοποιήθηκε για πρώτη φορά στον αλγόριθμο Adagrad [13] και λειτουργεί ως εξής: Προσαρμόζει τον ρυθμό μάθησης στις εκπαιδευσιμες παραμέτρους, εκτελώντας μικρότερες ανανεώσεις (δηλαδή χρησιμοποιώντας χαμηλό ρυθμό μάθησης) σε παραμέτρους που συσχετίζονται με χαρακτηριστικά συχνά εμφανιζόμενα στο σύνολο δεδομένων, και μεγαλύτερες ανανεώσεις (δηλαδή με υψηλό ρυθμό μάθησης) σε παραμέτρους που συσχετίζονται με σπανιότερα χαρακτηριστικά. Ένας ακόμη αλγόριθμος βελτιστοποίησης που βασίζεται σε προσαρμόσιμους ρυθμούς μάθησης και εφαρμόζεται στην παρούσα εργασία είναι ο αλγόριθμος Adaptive Moment Estimation - Adam [14] και εκφράζεται με ψευδοκώδικα ως εξής:

Αλγόριθμος Adam

Απαίτηση: η : Ρυθμός μάθησης

Απαίτηση: b_1, b_2 : Συντελεστές εκθετικών κινητών μέσων όρων για εκτίμηση ροπών

Απαίτηση: $f(\theta)$: Αντικειμενική συνάρτηση κόστους με παραμέτρους θ

Απαίτηση: θ_0 : Αρχικό τυχαίο διάνυσμα παραμέτρων

Απαίτηση: $\mathbf{m}_0 \leftarrow \mathbf{0}$: Αρχικό διάνυσμα πρώτης ροπής

Απαίτηση: $\mathbf{u}_0 \leftarrow \mathbf{0}$: Αρχικό διάνυσμα δεύτερης ροπής

Απαίτηση: $t \leftarrow 0$: Αρχικό χρονικό βήμα

Όσο θ_t δεν συγκλίνει **κάνε:**

$$t \leftarrow t + 1$$

$\mathbf{g}_t \leftarrow \nabla_{\mathbf{a}} f(\boldsymbol{\theta}_{t-1})$: Υπολογισμός παραγώγου αντικειμενικής συνάρτησης σε χρόνο t

$\mathbf{m}_t \leftarrow \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \mathbf{g}_t$: Υπολογισμός πρώτης ροπής σε χρόνο t

$\mathbf{u}_t \leftarrow \beta_2 \mathbf{u}_{t-1} + (1 - \beta_2) \mathbf{g}_t^2$: Υπολογισμός δεύτερης ροπής σε χρόνο t

$\hat{\mathbf{m}}_t \leftarrow \mathbf{m}_t / (1 - \beta_1^t)$: Υπολογισμός διορθωμένης πρώτης ροπής σε χρόνο t

$\hat{\mathbf{u}}_t \leftarrow \mathbf{u}_t / (1 - \beta_2^t)$: Υπολογισμός διορθωμένης δεύτερης ροπής σε χρόνο t

$\boldsymbol{\theta}_t \leftarrow \boldsymbol{\theta}_{t-1} - \frac{\eta \hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{u}}_t + \epsilon}}$: Ανανέωση παραμέτρων

Έλεγε κριτήριο σύγκλισης

Τέλος επανάληψης όσο

Σε κάθε επανάληψη του Adam υπολογίζεται, αρχικά, η κλίση της αντικειμενικής συνάρτησης κόστους ως προς τις εκπαιδευσιμες παραμέτρους. Στη συνέχεια, υπολογίζονται εκτιμήσεις για την πρώτη και την δεύτερη ροπή της κλίσης (μέσος όρος και διακύμανση της συνάρτησης της κλίσης αντίστοιχα) μέσω εκθετικών κινητών μέσων όρων. Οι εκτιμήσεις για τις ροπές, έπειτα, διορθώνονται, καθώς η αρχικοποίηση τους ως μηδενικά διανύσματα οδηγεί σε κινητούς μέσους όρους κοντά στο 0, ιδιαίτερα σε αρχικά βήματα του αλγορίθμου. Τέλος, το νέο διάνυσμα εκπαιδευσιμων παραμέτρων υπολογίζεται μέσω προσαρμογής του ρυθμού μάθησης σε παρελθοντικές τιμές της κλίσης της αντικειμενικής συνάρτησης κόστους. Οι συγγραφείς του Adam προτείνουν για τις σταθερές του αλγορίθμου της εξής τιμές:

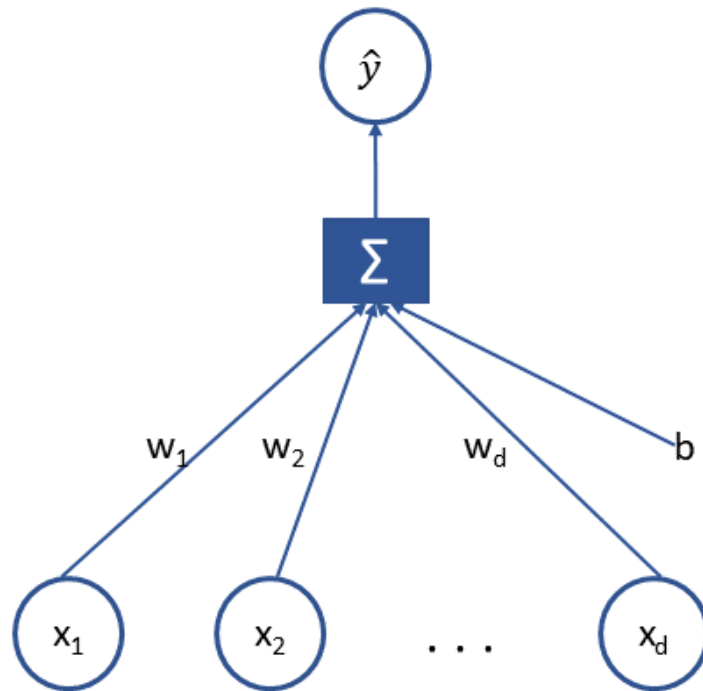
$$\eta = 0.001, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-9}$$

Σημειώνεται ότι η σταθερά ϵ είναι ένας όρος εξομάλυνσης (smoothing term) που χρησιμοποιείται για την αποφυγή διαίρεσης με το 0 στο τελικό βήμα του αλγορίθμου [11].

1.2.3 Νευρωνικά Δίκτυα Εμπρόσθιας Τροφοδότησης (Feedforward Neural Networks)

Ένα νευρωνικό δίκτυο μπορεί να μοντελοποιηθεί μαθηματικά ως ένας κατευθυνόμενος γράφος (περισσότερες πληροφορίες πάνω στην θεωρία γράφων παρατίθενται στο κεφάλαιο 3.1). Αποτελείται, δηλαδή, από κόμβους, οι οποίοι καλούνται νευρώνες και ακμές οι οποίες συνδέουν τους κόμβους μεταξύ τους. Οι νευρώνες κατανέμονται σε στρώματα (layers): το στρώμα εισόδου (input layer), το στρώμα εξόδου (output layer) και πιθανά κρυφά στρώματα (hidden layers). Οι νευρώνες ενός στρώματος δεν επικοινωνούν μεταξύ τους (δηλαδή δεν συνδέονται με ακμές), παρά μόνο με τους νευρώνες του προηγούμενου και του επόμενου στρώματος.

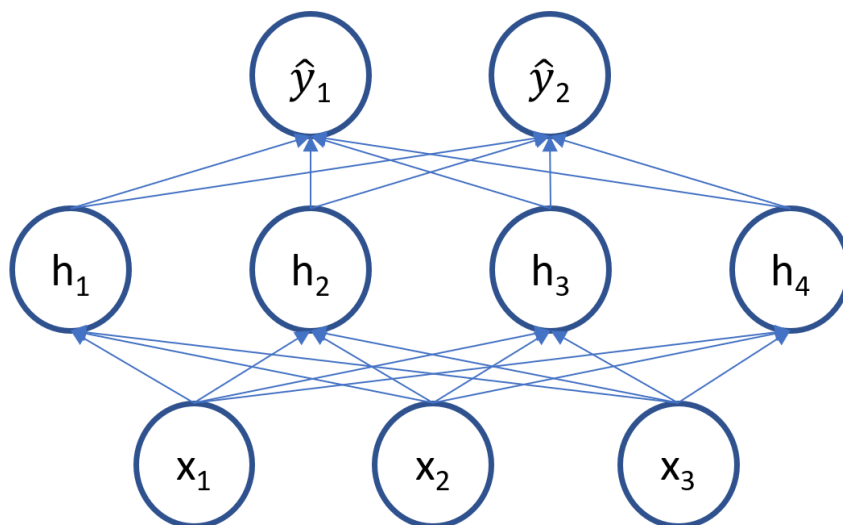
Το παράδειγμα της γραμμικής παλινδρόμησης μπορεί να αποτυπωθεί ως ένα νευρωνικό δίκτυο, με στρώμα εισόδου διάστασης d (αριθμός χαρακτηριστικών του συνόλου δεδομένων) και στρώμα εξόδου διάστασης 1. Πιο συγκεκριμένα, το στρώμα εισόδου αποτελείται από d ανεξάρτητες μεταβλητές x_1, x_2, \dots, x_d και το στρώμα εξόδου από μια εξαρτημένη μεταβλητή \hat{y} . Το στρώμα εισόδου που αποτελείται από d νευρώνες μπορεί να θεωρηθεί και ως ένα διάνυσμα \mathbf{x} του \mathbb{R}^d . Το γραμμικό νευρωνικό δίκτυο απεικονίζεται παρακάτω:



Σχήμα 1.5: Αρχιτεκτονική γραμμικού νευρωνικού δικτύου (με χρήση όρου bias)

Σημειώνεται ότι το κουτί που δηλώνει το άθροισμα συνήθως παραλείπεται από την απεικόνιση των αρχιτεκτονικών των διαφόρων νευρωνικών δικτύων.

Τα νευρωνικά δίκτυα εμπρόσθιας τροφοδότησης (feedforward neural networks ή Multi-Layer Perceptrons - MLPs) συνήθως αποτελούνται και από κρυφά στρώματα, εκτός από στρώματα εισόδου-εξόδου, ενώ όλα τα στρώματα είναι πλήρως συνδεδεμένα (fully-connected layers). Αυτό σημαίνει ότι κάθε κόμβος ενός στρώματος συνδέεται με όλους τους κόμβους του προηγούμενου στρώματος και όλους τους κόμβους του επόμενου στρώματος. Παρακάτω δίνεται η απεικόνιση ενός τυχαίου MLP με στρώμα εισόδου διάστασης 3, ένα κρυφό στρώμα διάστασης 4 και στρώμα εξόδου διάστασης 2:



Σχήμα 1.6: Αρχιτεκτονική δικτύου MLP (παράλειψη απεικόνισης βαρών και bias)

Για την μαθηματική έκφραση του MLP, τα παραδείγματα των δεδομένων εκπαίδευσης εκφράζονται μέσω πινάκων. Πιο συγκεκριμένα, ορίζεται ο πίνακας $\mathbf{X} \in \mathbb{R}^{n \times d}$, ο οποίος περιλαμβάνει τα n παραδείγματα d χαρακτηριστικών (ή εισόδων) (παραδείγματα του στρώματος εισόδου). Επίσης, ορίζεται ο πίνακας $\mathbf{H} \in \mathbb{R}^{n \times h}$, ο οποίος περιλαμβάνει τα παραδείγματα, όπως εξέρχονται από το κρυφό στρώμα αποτελούμενο από h κρυφά κανάλια. Τέλος, ορίζεται ο πίνακας $\mathbf{O} \in \mathbb{R}^{n \times q}$, ο οποίος περιλαμβάνει τις εξόδους από το στρώμα εξόδου, δηλαδή τις τελικές προβλέψεις του νευρωνικού δικτύου. Στο παραπάνω παράδειγμα MLP, ισχύει για τα επιμέρους κανάλια ότι $d = 3, h = 4, q = 2$. Με την προϋπόθεση γραμμικής σχέσης μεταξύ των στρωμάτων ισχύει:

$$\begin{aligned} \mathbf{H} &= \mathbf{X}\mathbf{W}^{(1)} + \mathbf{b}^{(1)}, \\ \mathbf{O} &= \mathbf{H}\mathbf{W}^{(2)} + \mathbf{b}^{(2)}, \end{aligned} \quad (1.10)$$

όπου $\mathbf{W}^{(1)} \in \mathbb{R}^{d \times h}$, $\mathbf{b}^{(1)} \in \mathbb{R}^{1 \times h}$ είναι τα βάρη και τα biases του (πλήρως συνδεδεμένου με το στρώμα εισόδου) κρυφό στρώμα και $\mathbf{W}^{(2)} \in \mathbb{R}^{h \times q}$, $\mathbf{b}^{(2)} \in \mathbb{R}^{1 \times q}$ είναι τα βάρη και τα biases του (πλήρως συνδεδεμένου με το κρυφό στρώμα) στρώμα εξόδου.

Αρκετά συχνά, η προσέγγιση συναρτήσεων μέσω γραμμικής παρεμβολής και άρα στην περίπτωση της επιβλεπόμενης μάθησης, η πρόβλεψη των ετικετών δεν είναι ικανοποιητική. Μπορεί να αποδειχθεί ότι η προσθήκη κρυφών στρωμάτων που ακολουθούν γραμμική σχέση δεν προσφέρει τίποτα περισσότερο από ένα απλό γραμμικό νευρωνικό δίκτυο. Παίρνοντας το παράδειγμα δικτύου του σχήματος 1.6 και ορίζοντας $\mathbf{W} = \mathbf{W}^{(1)}\mathbf{W}^{(2)}$ και $\mathbf{b} = \mathbf{b}^{(1)}\mathbf{W}^{(2)} + \mathbf{b}^{(2)}$ προκύπτει μέσω της εξίσωσης 1.10:

$$\mathbf{O} = (\mathbf{X}\mathbf{W}^{(1)} + \mathbf{b}^{(1)})\mathbf{W}^{(2)} + \mathbf{b}^{(2)} = \mathbf{X}\mathbf{W}^{(1)}\mathbf{W}^{(2)} + \mathbf{b}^{(1)}\mathbf{W}^{(2)} + \mathbf{b}^{(2)} = \mathbf{X}\mathbf{W} + \mathbf{b} \quad (1.11)$$

Για τον σκοπό της βελτίωσης της αποτελεσματικότητας ενός νευρωνικού δικτύου μέσω προσθήκης κρυφών στρωμάτων, απαιτείται η εφαρμογή μιας μη-γραμμικής συνάρτησης μετασχηματισμού σε κάθε νευρώνα του κρυφού στρώματος, η οποία καλείται συνάρτηση ενεργοποίησης (activation function) και συμβολίζεται με σ . Έτσι, η εξίσωση 1.10 παίρνει τη μορφή:

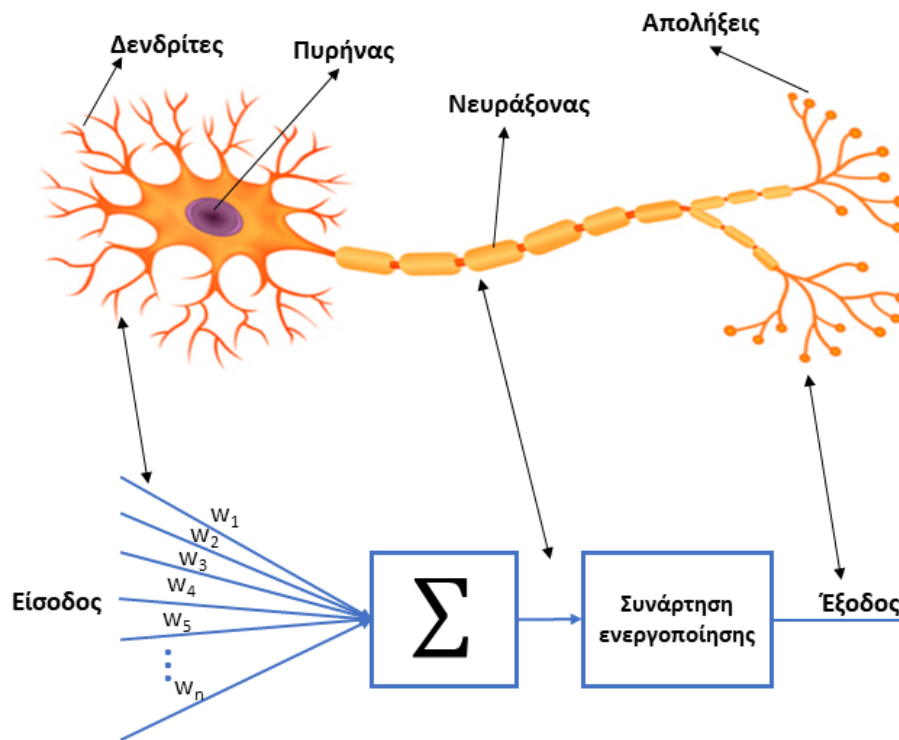
$$\begin{aligned} \mathbf{H} &= \sigma(\mathbf{X}\mathbf{W}^{(1)} + \mathbf{b}^{(1)}), \\ \mathbf{O} &= \mathbf{H}\mathbf{W}^{(2)} + \mathbf{b}^{(2)}. \end{aligned} \quad (1.12)$$

Παραδείγματα συναρτήσεων ενεργοποίησης που χρησιμοποιούνται κατά την εκπαίδευση νευρωνικών δικτύων θα αναφερθούν στην παράγραφο 1.3.3.

1.2.4 Τεχνητοί Νευρώνες σε Σύγκριση με Βιολογικούς Νευρώνες (Artificial Neurons in Comparison to Biological Neurons)

Όπως έχει ήδη αναφερθεί, τα τεχνητά νευρωνικά δίκτυα είναι εμπνευσμένα από τους νευρώνες του ανθρώπινου νευρικού συστήματος. Τυπικά, ένας (βιολογικός) νευρώνας αποτελείται από τους δενδρίτες και τον νευράξονα, ο οποίος καταλήγει στις νευραξονικές απολήξεις. Οι συνάψεις μεταξύ δενδριτών και νευραξονικών απολήξεων του προηγούμενου νευρικού κυττάρου τροποποιούν με διάφορους τρόπους το εισερχόμενο ηλεκτροχημικό σήμα. Στη συνέχεια, εάν το συνολικό τροποποιημένο, πλέον, σήμα ξεπεράσει ένα

προκαθορισμένο όριο τότε μεταφέρεται μέσω του νευράξονα προς το επόμενο συνδεδεμένο νευρικό κύτταρο. Κατ' αντίστοιχο τρόπο, ένας νευρώνας του MLP λαμβάνει ως είσοδο αριθμητικές τιμές, οι οποίες πολλαπλασιάζονται με τα βάρη. Έπειτα, πάνω στο άθροισμα των τροποποιημένων, πλέον, τιμών εφαρμόζεται μια συνάρτηση ενεργοποίησης και τέλος, η τιμή εξόδου τροφοδοτείται στον επόμενο νευρώνα [15]. Η σχηματική αναπαράσταση και σύγκριση βιολογικών και τεχνητών νευρώνων φαίνεται παρακάτω:



Σχήμα 1.7: Αναπαράσταση βιολογικού και τεχνητού νευρώνα. Γίνεται αντιστοίχιση των εισόδων του τεχνητού νευρώνα με τους δενδρίτες, της εφαρμογής αθροίσματος και συνάρτησης ενεργοποίησης με τον νευράξονα και της εξόδου με τις νευρικές απολήξεις.

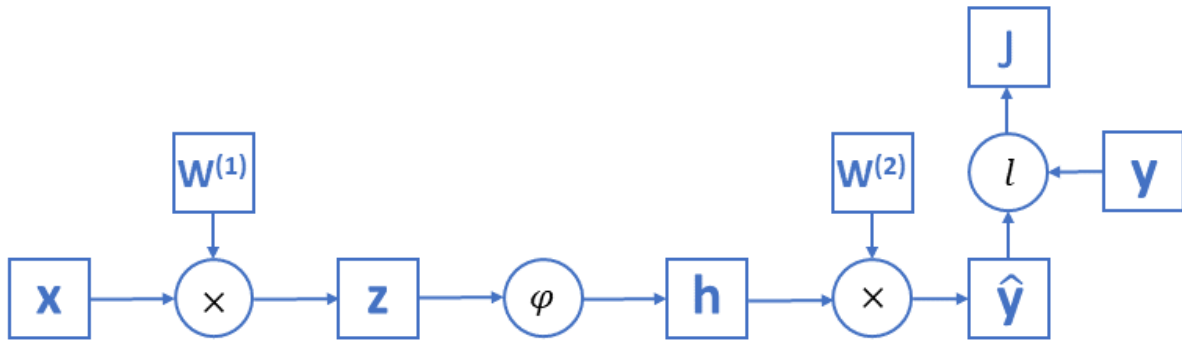
1.3 Εκπαίδευση Τεχνητών Νευρωνικών Δικτύων (Artificial Neural Networks Training)

1.3.1 Αλγόριθμος Οπισθο-Διάδοσης (Back-Propagation Algorithm)

Σκοπός των νευρωνικών δικτύων εμπρόσθιας τροφοδότησης είναι μέσω μιας εισόδου x να παράξουν στην έξοδο μια πρόβλεψη \hat{y} , με την πληροφορία να ρέει προς τα εμπρός κατά μήκος του δικτύου. Η είσοδος x λαμβάνει τις αρχικές πληροφορίες, τις οποίες διαδίδει, αρχικά, στα κρυφά στρώματα για επεξεργασία και τέλος, στην έξοδο όπου εξάγεται η τελική πρόβλεψη του μοντέλου. Η διαδικασία αυτή καλείται εμπρόσθια διάδοση (forward propagation). Τι συμβαίνει, όμως, όταν κατά τη διάρκεια της εκπαίδευσης είναι απαραίτητος ο υπολογισμός παραγώγων των εκπαιδευσιμων παραμέτρων (βλ. υποκεφάλαιο 1.2.2); Σε αυτό το κρίσιμο στάδιο της εκπαίδευσης επεμβαίνει ο αλγόριθμος παραγωγίσης Back-Propagation.

Όπως έχει ήδη αναφερθεί το νευρωνικό δίκτυο μπορεί να προτυποποιηθεί ως ένας κατευθυνόμενος γράφος. Ο γράφος αυτός καλείται υπολογιστικός γράφος, οι κόμβοι του οποίου

περιλαμβάνουν τετράγωνα για μεταβλητές και κύκλους για μαθηματικές πράξεις. [5] Για ένα απλό MLP με ένα κρυφό στρώμα ο υπολογιστικός γράφος είναι ο εξής (οι όροι bias αμελούνται για λόγους απλότητας):



Σχήμα 1.8: Υπολογιστικός γράφος MLP ενός κρυφού στρώματος, όπου ο κυκλικός κόμβος ϕ δηλώνει την εφαρμογή συνάρτησης ενεργοποίησης του κρυφού στρώματος και ο κυκλικός κόμβος l δηλώνει την εφαρμογή συνάρτησης κόστους. Η τελική αντικειμενική συνάρτηση προς ελαχιστοποίηση συμβολίζεται με J

Σκοπός του αλγορίθμου Back-Propagation στο παραπάνω παράδειγμα είναι ο υπολογισμός των παραγώγων της αντικειμενικής συνάρτησης ως προς τις εκπαιδευσιμες παραμέτρους $\nabla_{\mathbf{w}^{(1)}} J$ και $\nabla_{\mathbf{w}^{(2)}} J$. Για να επιτευχθεί αυτό χρησιμοποιείται ο κανόνας της αλυσίδας (chain rule), ο οποίος για βαθμωτά μεγέθη $x, y, z \in \mathbb{R}$ με $y = g(x)$, $z = f(g(x)) = f(y)$ όπου $f, g : \mathbb{R} \rightarrow \mathbb{R}$ ορίζεται ως εξής [5]:

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx}. \quad (1.13)$$

Ο κανόνας της αλυσίδας μπορεί να γενικευτεί και για μεγέθη μεγαλύτερων διαστάσεων. Έστω τα διανυσματικά μεγέθη $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$, το βαθμωτό μέγεθος $z \in \mathbb{R}$ και οι συναρτήσεις $g : \mathbb{R}^m \rightarrow \mathbb{R}^n$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ με $\mathbf{y} = g(\mathbf{x})$ και $z = f(\mathbf{y})$. Τότε ισχύει:

$$\frac{\partial z}{\partial x_i} = \sum_j \frac{\partial z}{\partial y_j} \frac{\partial y_j}{\partial x_i}, \quad (1.14)$$

ή ισοδύναμα χρησιμοποιώντας σημειογραφία διανυσμάτων:

$$\nabla_{\mathbf{x}} z = \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right)^T \nabla_{\mathbf{y}} z, \quad (1.15)$$

όπου $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ ο ιακωβιανός πίνακας της συνάρτησης g .

Κατ' αντιστοιχία, για τανυστές τρίτης τάξης \mathbf{X}, \mathbf{Y} και βαθμωτό μέγεθος z με $\mathbf{Y} = g(\mathbf{X})$ και $z = f(\mathbf{Y})$ ισχύει:

$$\nabla_{\mathbf{X}} z = \sum_j (\nabla_{\mathbf{X}} Y_j) \frac{\partial z}{\partial Y_j}. \quad (1.16)$$

Όπως γίνεται αντιληπτό, η παράγωγος ως προς μια μεταβλητή \mathbf{x} μπορεί να βρεθεί πολλαπλασιάζοντας έναν ιακωβιανό πίνακα $\frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{x}}$ με την παράγωγο ως προς την μεταβλητή \mathbf{y} (όμοια, για τανυστές τρίτης τάξης αθροίζοντας ως προς την μία διάσταση). Ο αλγόριθμος Back-Propagation, λοιπόν, πραγματοποιεί γινόμενα ιακωβιανού πίνακα-παραγώγου για κάθε μαθηματική πράξη (κυκλικό κόμβο) στον υπολογιστικό γράφο. Έτσι, για το παράδειγμα του σχήματος 1.8 θα ισχύει:

$$\nabla_{\mathbf{W}^{(2)}} J = \left(\frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{W}^{(2)}} \right)^T \nabla_{\hat{\mathbf{y}}} J = \mathbf{h} \nabla_{\hat{\mathbf{y}}} J, \quad (1.17)$$

$$\nabla_{\mathbf{z}} J = \left(\frac{\partial \mathbf{h}}{\partial \mathbf{z}} \right)^T \nabla_{\mathbf{h}} J = \nabla_{\mathbf{h}} J \odot \phi'(\mathbf{z}), \quad (1.18)$$

$$\nabla_{\mathbf{W}^{(1)}} J = \left(\frac{\partial \mathbf{z}}{\partial \mathbf{W}^{(1)}} \right)^T \nabla_{\mathbf{z}} J = \mathbf{x} \nabla_{\mathbf{z}} J = \mathbf{x} \nabla_{\mathbf{h}} J \odot \phi'(\mathbf{z}). \quad (1.19)$$

Παρατηρείται, λοιπόν, ότι η πορεία που χαράσσεται για τον υπολογισμό των απαιτούμενων παραγώγων ξεκινά από το τέλος του υπολογιστικού γράφου (αντικειμενική συνάρτηση προς βελτιστοποίηση) και καταλήγει στην αρχή (δεδομένα εισόδου).

Σημειώνεται ότι η συνάρτηση ενεργοποίησης ϕ εφαρμόζεται ανά στοιχείο του διανύσματος \mathbf{z} και έτσι, ο υπολογισμός της παραγώγου της εξίσωσης 1.18 γίνεται με πολλαπλασιασμό ανά στοιχείο (γινόμενο Hadamard).

Συμπερασματικά, κατά την εκπαίδευση νευρωνικών δικτύων η πληροφορία διαδίδεται αρχικά προς τα εμπρός με υπολογισμό όλων των μεταβλητών κατά μήκος του υπολογιστικού γράφου και στην συνέχεια, οι μεταβλητές αυτές χρησιμοποιούνται για τον υπολογισμό των απαιτούμενων παραγώγων ως προς τις εκπαιδευσιμες παραμέτρους, οι οποίες θα χρησιμοποιηθούν για την ελαχιστοποίηση της αντικειμενικής συνάρτησης. Κατόπιν, με εφαρμογή αλγορίθμων βελτιστοποίησης (π.χ. Gradient Descent) οι παράμετροι ανανεώνονται και ξεκινά εκ νέου η προς τα εμπρός διάδοση πληροφορίας. Επομένως, η εκπαίδευση νευρωνικών δικτύων στηρίζεται στην εναλασσόμενη και αλληλένδετη εφαρμογή προς τα εμπρός και προς τα πίσω διάδοσης (forward and backward propagation).

1.3.2 Συναρτήσεις Κόστους (Cost Functions)

Όπως αναφέρθηκε στο κεφάλαιο 1.2.1, οι συναρτήσεις κόστους αποτελούν ένα μέτρο ποσοτικοποίησης της επιτυχίας πρόβλεψης των ετικετών $\hat{\mathbf{y}}$ δεδομένων των εισόδων \mathbf{x} ενός συνόλου δεδομένων από το νευρωνικό δίκτυο. Οι συναρτήσεις κόστους δηλαδή βοηθούν τον υπολογιστή σε κάθε επανάληψη να καταλάβει το λάθος του και στη συνέχεια, με την διαδικασία της εκπαίδευσης (προς τα εμπρός και προς τα πίσω διάδοσης) να μάθει από αυτό. Προκειμένου η διαδικασία της προς τα πίσω διάδοσης να λειτουργεί, οι συναρτήσεις κόστους πρέπει να πληρούν τις εξής δύο προϋποθέσεις: [16]

- Η συνάρτηση κόστους L θα πρέπει να μπορεί να γραφεί ως ο μέσος όρος $L = \frac{1}{n} \sum_i L_{xi}$ συναρτήσεων κόστους για κάθε μεμονωμένο παράδειγμα του συνόλου δεδομένων εκπαίδευσης, \mathbf{x}_i .
- Η συνάρτηση κόστους θα πρέπει να εξαρτάται από τις εξόδους του νευρωνικού δικτύου, δηλαδή τις εξόδους του τελευταίου στρώματος του νευρωνικού δικτύου.

Η συνάρτηση του Μέσου Τετραγωνικού Σφάλματος (1.2) ικανοποιεί και τις δύο προϋποθέσεις και έτσι, μπορεί να αποτελέσει μια επιλογή για συνάρτηση κόστους του νευρωνικού δικτύου.

Για επίλυση προβλημάτων ταξινόμησης (classification problems), όπως το πρόβλημα της παρούσας εργασίας, χρησιμοποιείται ευρέως η συνάρτηση κόστους διασταυρούμενης εντροπίας (cross-entropy loss function). Έστω ένα πρόβλημα ταξινόμησης δύο κλάσεων βασισμένο σε n παραδείγματα $\{x_1, x_2, \dots, x_n\}$. Θεωρείται ότι οι δύο κλάσεις χωρίζονται σε αρνητική και θετική και κωδικοποιούνται με 0 και 1 αντίστοιχα και ότι το νευρωνικό δίκτυο εξαρτάται από την παράμετρο θ . Ορίζεται επίσης η πραγματική κατανομή P με συνάρτηση κατανομής πιθανότητας $p(x)$ και η προβλεπόμενη κατανομή Q με συνάρτηση κατανομής πιθανότητας $q(x)$. Εάν είναι επιθυμητό να βρεθεί η βέλτιστη τιμή θ ώστε $\hat{y}_i = p_\theta(y_i|x_i)$, εφαρμόζεται η προσέγγιση της μέγιστης πιθανοφάνειας (maximum likelihood approach) [5]. Πιο συγκεκριμένα, για πραγματικές ετικέτες y_i και προβλέψεις \hat{y}_i , η πιθανότητα για ταξινόμηση ως θετική κλάση είναι $\pi_i = p_\theta(y_i = 1|x_i)$. Για λόγους αριθμητικής βελτιστοποίησης χρησιμοποιείται ο λογάριθμος της πιθανοφάνειας:

$$\begin{aligned} l(\theta) &= -L(\theta) \\ &= -\ln(q(x)^{p(x)}) \\ &= -\ln \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \\ &= -\sum_{i=1}^n y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i), \end{aligned} \quad (1.20)$$

όπου $L(\theta)$ η συνάρτηση λογαρίθμου πιθανοφάνειας (log-likelihood function) όπως εφαρμόζεται σε ταξινόμηση δύο κλάσεων (δύο ανεξάρτητα ενδεχόμενα).

Παρατηρείται στην εξίσωση 1.20 ότι οι όροι $y_i \ln(\pi_i)$, $(1 - y_i) \ln(1 - \pi_i)$ είναι αρνητικοί αριθμοί. Η συνάρτηση κόστους, όμως, εκφράζει την απόκλιση της πρόβλεψης του μοντέλου από την πραγματική τιμή και συνηθίζεται να παίρνει θετική τιμή, η οποία για ένα καλά εκπαιδευμένο μοντέλο προσεγγίζει το 0. Για αυτόν τον λόγο την εξίσωση 1.20 χρησιμοποιείται το αρνητικό πρόσημο. Παράλληλα, η ελαχιστοποίηση της αρνητικής πιθανοφάνειας συνεπάγεται την μεγιστοποίηση της πιθανοφάνειας, η οποία είναι και η αρχή στην οποία στηρίζεται η συνάρτηση κόστους διασταυρούμενης εντροπίας.

Το κόστος διασταυρούμενης εντροπίας μπορεί να επεκταθεί και για περισσότερες των δύο (έστω k) κλάσεων:

$$l(\theta) = -\sum_{i=1}^n \sum_{j=1}^k y_{ij} \ln(\pi_j). \quad (1.21)$$

1.3.3 Συναρτήσεις Ενεργοποίησης (Activation Functions)

Οι συναρτήσεις ενεργοποίησης, όπως φανερώνει το όνομά τους, αποφασίζουν εάν ένας νευρώνας του νευρωνικού δικτύου "ενεργοποιείται" ή όχι μεταβιβάζοντας πληροφορίες στον επόμενο νευρώνα. Είναι διαφορεσιμες συναρτήσεις, όπως προϋποθέτει ο αλγόριθμος εκπαίδευσης Back-Propagation και εφαρμόζονται στα κρυφά στρώματα του νευρωνικού δικτύου. Η απλούστερη συνάρτηση ενεργοποίησης, όπως λογικά μπορεί να σκεφτεί κανείς, είναι η βηματική συνάρτηση (step function):

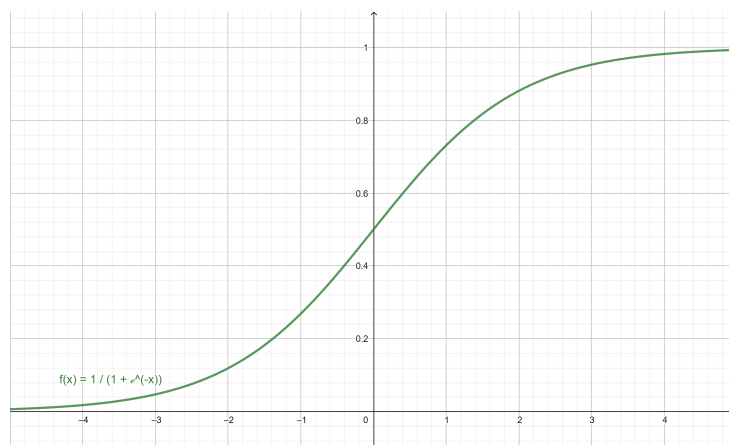
$$\sigma(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (1.22)$$

Η βηματική συνάρτηση επιτρέπει την μετάδοση της πληροφορίας αυτούσιας εάν αυτή δεν υπερβαίνει ένα συγκεκριμένο όριο (στην συγκεκριμένη περίπτωση το 0). Παρ' όλα αυτά, η παράγωγος της είναι παντού 0 και έτσι, ο αλγόριθμος βελτιστοποίησης Gradient Descent δεν θα μπορούσε να ανανεώσει τα βάρη και τα biases και η εκπαίδευση μέσω του Back-Propagation θα αποτύγχανε. Επιπλέον, όπως έχει ήδη αναλυθεί στο υποκεφάλαιο 1.2.3, ο μετασχηματισμός του σήματος μέσω μιας μη γραμμικής συνάρτησης βελτιώνει την αποτελεσματικότητα του νευρωνικού δικτύου. Άλλη μια ιδιότητα των συναρτήσεων ενεργοποίησης που είναι επιθυμητή για λόγους σταθερότητας των μεθόδων βελτιστοποίησης βασισμένες στον υπολογισμό παραγώγου (όπως ο Gradient Descent και οι παραλλαγές του) είναι το σύνολο τιμών τους να είναι εντός ενός πεπερασμένου συνόλου. Σε αντίθετη περίπτωση, επιλέγεται μικρός ρυθμός μάθησης για πιο αργή και σταθερή βελτιστοποίηση.

Με βάση τα παραπάνω δεδομένα, προτείνονται από τη βιβλιογραφία οι εξής συναρτήσεις ως επιλογές για συναρτήσεις ενεργοποίησης κρυφών στρωμάτων ενός τεχνητού νευρωνικού δικτύου [17]:

- Η σιγμοειδής ή λογιστική συνάρτηση (sigmoid function):

$$f(x) = \frac{1}{1 + e^{-x}} : \mathbb{R} \rightarrow (0, 1) , \quad (1.23)$$

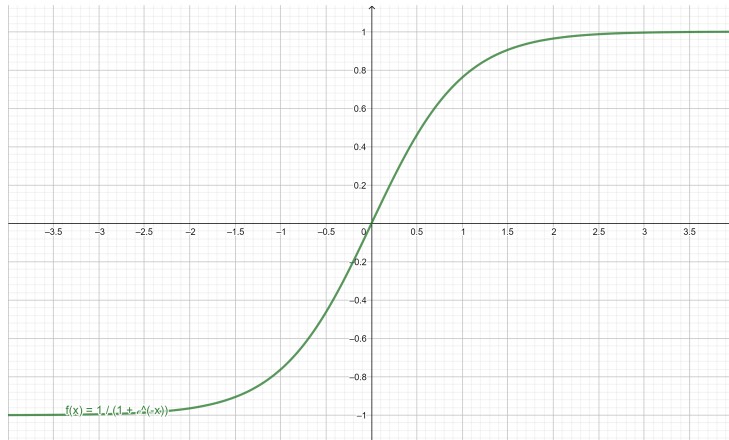


Σχήμα 1.9: Γραφική παράσταση της σιγμοειδούς συνάρτησης

η οποία είναι μια μη γραμμική, φραγμένη, διαφορίσιμη συνάρτηση με θετική πρώτη παράγωγο. Στο σχήμα 1.9 παρατηρείται ότι σε τιμές $|x| > 4$ η κλίση της συνάρτησης σχεδόν μηδενίζεται, γεγονός που οδηγεί στο πρόβλημα των βαθμίδων που εξαφανίζονται (vanishing gradients problem). [18] Αυτό σημαίνει ότι μόλις η εκπαίδευση μέσω του αλγόριθμου βελτιστοποίησης οδηγηθεί στα άκρα της συνάρτησης, αυτή πρακτικά σταματάει λόγω της εμφάνισης μηδενικών βαθμίδων. Το παραπάνω είναι και το μοναδικό μειονέκτημα της σιγμοειδούς συνάρτησης, η οποία παρ' όλα αυτά, χρησιμοποιείται ευρέως ως ενεργοποίηση του στρώματος εξόδου σε περιπτώσεις ταξινόμησης δύο κλάσεων. Λόγω του συνόλου τιμών της, οι πληροφορίες στην έξοδο μεταφράζονται σε μορφή πιθανοτήτων, το οποίο είναι και το ζητούμενο μιας εφαρμογής ταξινόμησης.

- Η συνάρτηση υπερβολικής εφαπτομένης (hyperbolic tangent function):

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} : \mathbb{R} \rightarrow (-1, 1) . \quad (1.24)$$

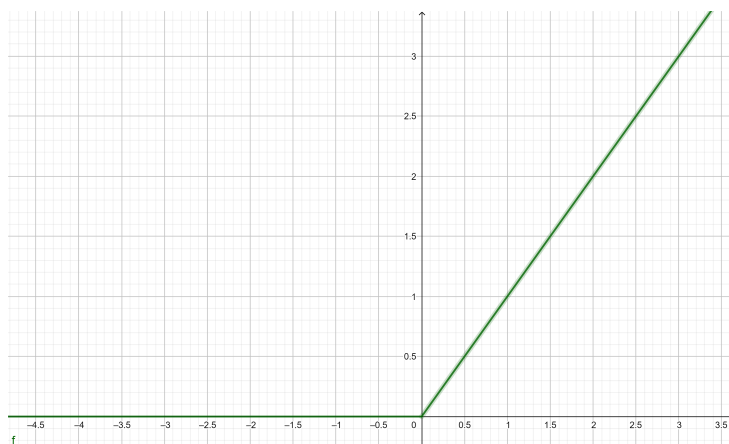


Σχήμα 1.10: Γραφική παράσταση της συνάρτησης υπερβολικής εφαπτομένης

Η διαφορά της υπερβολικής εφαπτομένης από την σιγμοειδή συνάρτηση έγκειται, εκτός από το σύνολο τιμών, στην κλίση τους στο διάστημα $[-2,2]$. Η υπερβολική εφαπτομένη έχει μεγαλύτερη κλίση σε αυτό το διάστημα. Επομένως, η επιλογή ανάμεσα στις δύο συναρτήσεις εξαρτάται από το πόσο μεγάλη κλίση χρειάζεται να επιβληθεί στο νευρωνικό δίκτυο στην εκάστοτε περίπτωση. Σε γενικές γραμμές, η υπερβολική εφαπτομένη προσφέρει καλύτερη εκπαίδευση, παρ' όλα αυτά, το πρόβλημα των βαθμίδων που εξαφανίζονται παραμένει.

- Η συνάρτηση διορθωμένης γραμμικής μονάδας (rectified linear unit - ReLU):

$$f(x) = \max(0, x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} : \mathbb{R} \rightarrow [0, +\infty) . \quad (1.25)$$



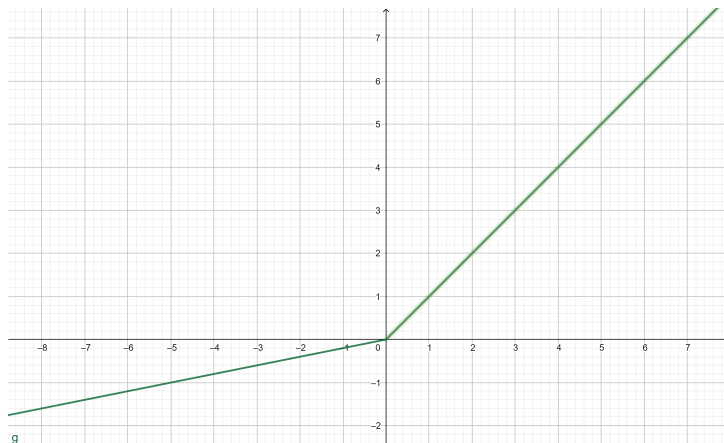
Σχήμα 1.11: Γραφική παράσταση της συνάρτησης ReLU

Η ReLU χρησιμοποιείται ευρύτατα στα νευρωνικά δίκτυα ως ενεργοποίηση κρυφών νευρώνων, καθώς η εφαρμογή της προσφέρει την απαραίτητη μη-γραμμικότητα στις πληροφορίες χωρίς την υπολογιστική επιβάρυνση των εκθετικών και των κλασμάτων των προηγούμενων συναρτήσεων. Ακόμη, δεν ενεργοποιεί όλους τους νευρώνες και έτσι, ένα νευρωνικό δίκτυο εκατοντάδων ή και χιλιάδων νευρώνων γίνεται πιο αραιό και ανάλαφρο και εξοικονομείται πολύς υπολογιστικός χρόνος. Παρ' όλα αυτά, ορισμένες φορές οι νεκροί νευρώνες (δηλαδή οι νευρώνες που δέχονται μηδενική ενεργοποίηση) μπορεί να επιφέρουν δυσκολίες στην ικανότητα επαλήθευσης του μοντέλου. Έτσι, προτείνονται στη βιβλιογραφία διάφορες παραλλαγές της ReLU που δεν μηδενίζουν αρνητικές τιμές εισόδου στην ενεργοποίηση του νευρώνα.

- Η συνάρτηση παραμετρικής διορθωμένης γραμμικής μονάδας (parametric rectified linear unit - PReLU):

$$f(x_i) = \begin{cases} x_i, & x_i \geq 0 \\ a_i x_i, & x_i < 0 \end{cases} : \mathbb{R} \rightarrow \mathbb{R}, \quad (1.26)$$

όπου i δηλώνει τον εκάστοτε νευρώνα του νευρωνικού δικτύου που δέχεται ενεργοποίηση μέσω PReLU και a_i εκπαιδευσιμη παράμετρος του νευρωνικού δικτύου.



Σχήμα 1.12: Γραφική παράσταση της συνάρτησης PReLU

Η PReLU λύνει το πρόβλημα των νεκρών νευρώνων που παρουσιάζεται στην ReLU, ενώ ταυτόχρονα προσθέτει μια παράμετρο σε κάθε νευρώνα με αρνητική είσοδο στην ενεργοποίηση που μπορεί να εκπαιδευτεί μέσω του back-propagation. Η χρήση της σε εφαρμογές ταξινόμησης εικόνων είχε εντυπωσιακά αποτελέσματα, αφού τα νευρωνικά δίκτυα που την χρησιμοποίησαν έγιναν τα πρώτα που ξεπέρασαν τον άνθρωπο σε ικανότητα οπτικής αναγνώρισης. [19]

- Η συνάρτηση Softmax

$$f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}, \quad (1.27)$$

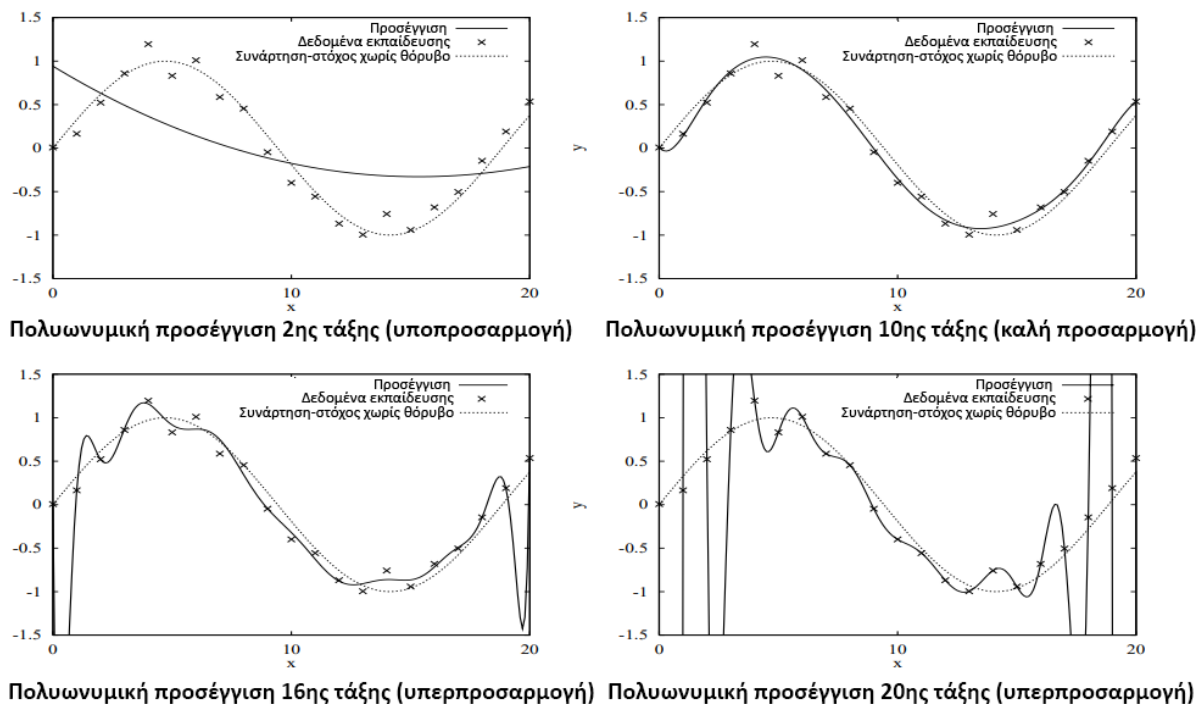
όπου i δηλώνει τον νευρώνα που εφαρμόζεται η ενεργοποίηση μέσω Softmax και j τους νευρώνες του στρώματος, στο οποίο βρίσκεται ο νευρώνας i .

Η Softmax χρησιμοποιείται μόνο σε νευρώνες του στρώματος εξόδου νευρωνικών δικτύων που επιδιώκουν ταξινόμηση περισσότερων των δύο κλάσεων. Αυτό που κατορθώνει είναι η μετατροπή αριθμητικών τιμών στην είσοδο της ενεργοποίησης σε πιθανότητες (αριθμούς που ανήκουν στο διάστημα $[0,1]$). Το άθροισμα των εξόδων των νευρώνων του στρώματος εξόδου έχει άθροισμα 1. Η κλάση με την μεγαλύτερη πιθανότητα αποτελεί και την τελική πρόβλεψη του νευρωνικού δικτύου.

1.3.4 Φθορά Βαρών και Dropout ως Μέθοδοι Αποφυγής Υπερπροσαρμογής (Weight Decay and Dropout as Methods to Prevent Overfitting)

Σκοπός των μοντέλων μηχανικής μάθησης είναι η ανακάλυψη μοτίβων που χαρακτηρίζουν έναν πληθυσμό, μέρος του οποίου είναι το σύνολο των δεδομένων εκπαίδευσης. Εάν τα μοτίβα αυτά είναι γενικευμένα, τότε το μοντέλο θα μπορέσει να αποδώσει καλά ακόμα και για άτομα του πληθυσμού που δεν έχει συναντήσει (δηλαδή πάνω στα οποία δεν έχει εκπαιδευτεί). Από την άλλη, εάν εκπαιδευτεί τέλεια για τα άτομα του συνόλου εκπαίδευσης, υπάρχει ο κίνδυνος το μοντέλο να αδυνατεί να παράξει προβλέψεις για διαφορετικά άτομα του πληθυσμού. Αυτό αποτελεί και το πρόβλημα της υπερπροσαρμογής και για την αποφυγή του στα νευρωνικά δίκτυα χρησιμοποιούνται τεχνικές κανονικοποίησης (regularization techniques), όπως η φθορά βαρών (weight decay) και το dropout.

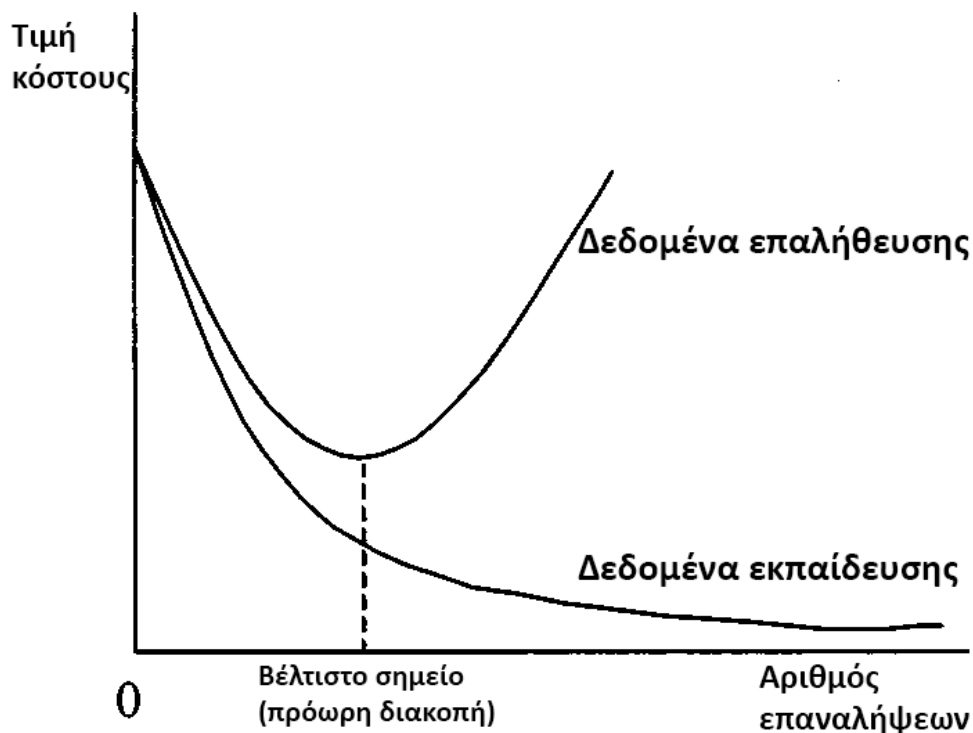
Το πρόβλημα απουσίας γενίκευσης του μοντέλου μέσω του overfitting μπορεί να γίνει αντιληπτό μέσω ενός παραδείγματος πολυωνυμικής προσέγγισης συνάρτησης [20]:



Σχήμα 1.13: Περιπτώσεις υποπροσαρμογής, καλής προσαρμογής και υπερπροσαρμογής μοντέλων πολυωνυμικής προσέγγισης σε δεδομένα εκπαίδευσης

Κατά τη διάρκεια εκπαίδευσης ενός μοντέλου νευρωνικών δικτύων είναι πολύ σημαντική η online παρακολούθηση της απόδοσης του μοντέλου σε άτομα του πληθυσμού που δεν ταυτίζονται με το σύνολο δεδομένων εκπαίδευσης. Έτσι, ορίζεται το σύνολο δεδομένων

επαλήθευσης, το οποίο συνήθως έχει μέγεθος 10 – 20% του μεγέθους του συνόλου δεδομένων εκπαίδευσης. Μάλιστα, τα δεδομένα εκπαίδευσης και επαλήθευσης μπορούν να εναλλάσσονται μέσω της διαδικασίας της διασταυρούμενης επαλήθευσης που θα αναλυθεί σε επόμενο κεφάλαιο της εργασίας (βλ. υποκεφάλαιο 6.2). Η προσαρμογή του μοντέλου στα δεδομένα εκπαίδευσης μετράται και πάλι μέσω της συνάρτησης κόστους. Η τιμή κόστους στα δεδομένα εκπαίδευσης και στα δεδομένα επαλήθευσης για ένα μοντέλο που παρουσιάζει υπερπροσαρμογή στα δεδομένα εκπαίδευσης συναρτήσει των κύκλων εκπαίδευσης θα έχει την παρακάτω μορφή [17]:



Σχήμα 1.14: Τιμή κόστους για δεδομένα εκπαίδευσης και επαλήθευσης συναρτήσει του αριθμού επαναλήψεων του αλγορίθμου εκπαίδευσης του νευρωνικού δικτύου

Ζητούμενο σε αυτές τις περιπτώσεις είναι είτε η εξάλειψη/περιορισμός του φαινομένου της υπερεκπαίδευσης είτε η πρόωρη διακοπή εκπαίδευσης του μοντέλου στο βέλτιστο σημείο, δηλαδή στο ελάχιστο της τιμής κόστους του συνόλου επαλήθευσης που ταυτίζεται με το σημείο μέγιστης γενίκευσης του μοντέλου.

Φθορά βαρών (weight decay)

Η φθορά βαρών ή κανονικοποίηση L2 (L2 regularization) είναι ίσως η πιο διαδεδομένη τεχνική για κανονικοποίηση παραμέτρων μοντέλων μηχανικής μάθησης. Κατά την εφαρμογή της, προστίθεται στην συνάρτηση κόστους η ευκλείδεια (ή L2) νόρμα του διανύσματος βαρών του εκάστοτε παραδείγματος i του συνόλου δεδομένων εκπαίδευσης. Οι νόρμες αποτελούν ένα μέτρο για το μέγεθος ενός διανύσματος ή πίνακα. Συγκεκριμένα, η ευκλείδεια νόρμα μετρά την ευκλείδεια απόσταση του αθροίσματος των στοιχείων ενός διανύσματος από το 0.

Στο σημείο αυτό ορίζεται η ευκλείδεια νόρμα ενός διανυσματικού μεγέθους $\mathbf{x} \in \mathbb{R}^n$ με στοιχεία x_1, x_2, \dots, x_n ως εξής:

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}, \quad (1.28)$$

ενώ ο γενικότερος ορισμός της L_p νόρμας για το ίδιο μέγεθος \mathbf{x} είναι ο εξής:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n x_i^p \right)^{1/p}. \quad (1.29)$$

Τέλος, για πίνακα $\mathbf{X} \in \mathbb{R}^{m \times n}$ ορίζεται κατ' αντιστοιχία της ευκλείδειας νόρμας για τα διανύσματα, η νόρμα Frobenius:

$$\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2}. \quad (1.30)$$

Επομένως, η νέα συνάρτηση κόστους με την προσθήκη του όρου κανονικοποίησης παίρνει την εξής μορφή:

$$L(\mathbf{w}, b)^{reg} = L(\mathbf{w}, b) + \lambda \|\mathbf{w}\|_2^2, \quad (1.31)$$

όπου λ ονομάζεται παράμετρος κανονικοποίησης και αποτελεί μια υπερπαράμετρο που καθορίζεται από τον χρήστη, μέσω της οποίας αποφασίζεται η βαρύτητα της κανονικοποίησης στην νέα αντικειμενική συνάρτηση προς ελαχιστοποίηση. Για τιμή $\lambda = 0$ η αντικειμενική συνάρτηση επιστρέφει στην πρότερη μορφή της (χωρίς τον όρο κανονικοποίησης), ενώ για μεγάλες τιμές του λ μπορεί να παρατηρηθεί υποπροσαρμογή του μοντέλου. Έτσι, το λ μπορεί να χαρακτηριστεί ως ένα απόβαρο μεταξύ καλής προσαρμογής του μοντέλου και παρουσίας μικρών τιμών παραμέτρων. [22] Επίσης, για λόγους εξοικονόμησης υπολογιστικού όγκου, οι ευκλείδειες νόρμες υψώνονται στο τετράγωνο, αφού χωρίς την ύπαρξη της ρίζας είναι πιο εύκολος ο υπολογισμός των παραγώγων τους.

Συνοπτικά, με την φθορά βαρών καθίσταται δυνατό κατά την διάρκεια της εκπαίδευσης ο αλγόριθμος βελτιστοποίησης να επικεντρωθεί στην ελαχιστοποίηση του αθροίσματος των ευκλείδειων νορμών των βαρών εφόσον τα βάρη πάρουν μεγάλες τιμές και όχι στην ελαχιστοποίηση της συνάρτησης κόστους αποτρέποντας με αυτόν τον τρόπο την υπερπροσαρμογή του μοντέλου. Σε περίπτωση που είναι επιθυμητή η επιλογή χαρακτηριστικών (feature selection), τότε επιλέγεται η L_1 κανονικοποίηση, η οποία χρησιμοποιεί την L_1 νόρμα των βαρών. Η νόρμα αυτή των βαρών ταυτίζεται με το άθροισμα των απόλυτων τιμών των στοιχείων που τα απαρτίζουν. Κατά την εκπαίδευση με L_1 κανονικοποίηση, ο αλγόριθμος βελτιστοποίησης συγκεντρώνει τα βάρη σε ένα μικρό σει χαρακτηριστικών θέτοντας τα υπόλοιπα ίσα με 0. Η L_1 κανονικοποίηση μπορεί να επιλεγεί σε περίπτωση που ένα σύνολο δεδομένων περιλαμβάνει πολλά χαρακτηριστικά, αρκετά εκ των οποίων δεν θεωρείται ότι βοηθούν στην εξαγωγή μοτίβων μέσω μοντέλων μηχανικής μάθησης.

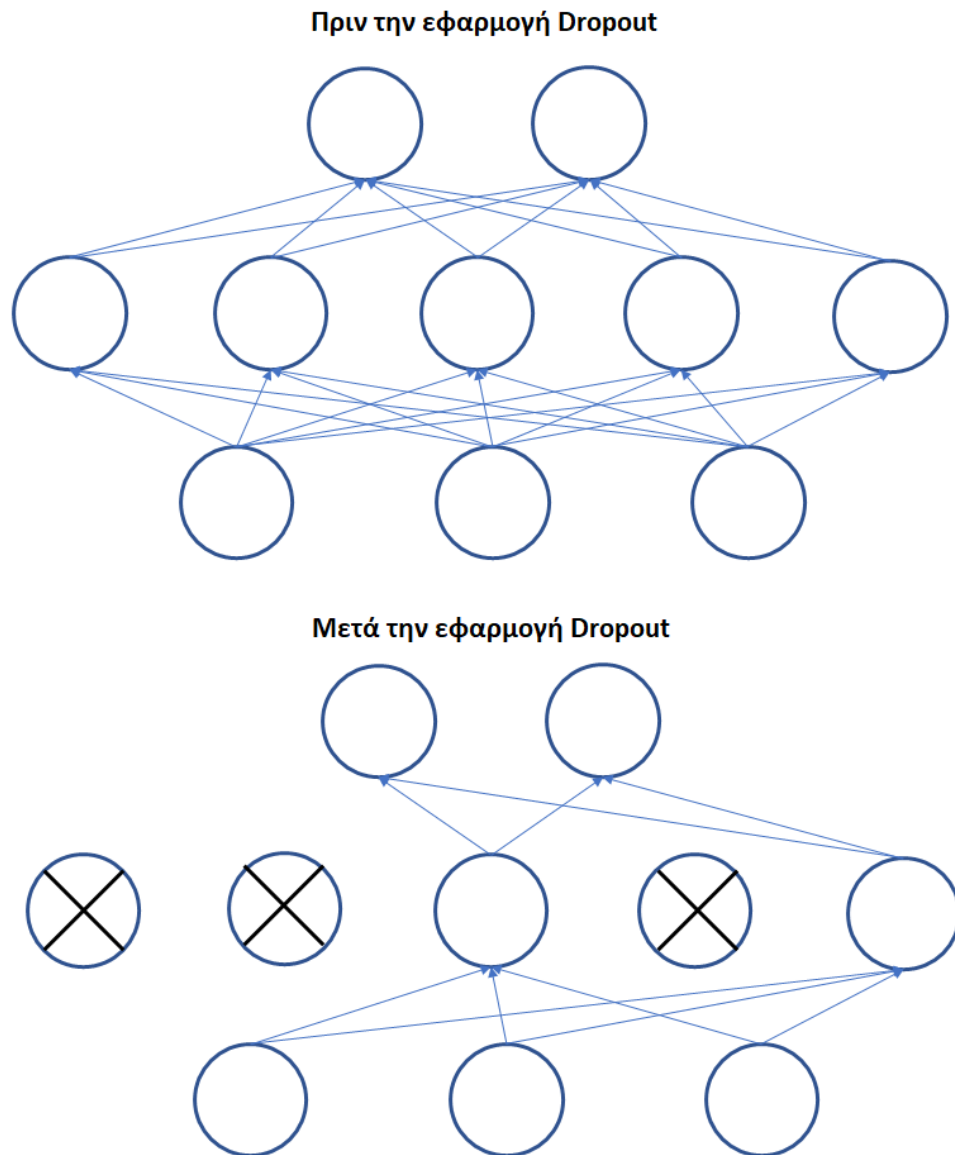
Dropout

Η εφαρμογή του Dropout σε ένα στρώμα του νευρωνικού δικτύου συνεπάγεται την απενεργοποίηση τυχαίου αριθμού νευρώνων του. Με άλλα λόγια, σήματα ενίσχυσης νευρώνων επιλέγονται κατά τυχαίο τρόπο ίσα με 0. Το Dropout στα νευρωνικά δίκτυα στηρίζεται στην

ιδέα του Bishop, ο οποίος το 1995 απέδειξε ότι η εκπαίδευση ενός νευρωνικού δικτύου με προσθήκη θορύβου (ο οποίος συλλέχθηκε από κανονική κατανομή $N(0, \sigma^2)$) στην είσοδο είναι ισοδύναμη της εκπαίδευσης με κανονικοποίηση Tikhonov, μια ειδική περίπτωση της οποίας είναι και η φθορά βαρών. [23] Η εισαγωγή του θορύβου στην περίπτωση του Dropout γίνεται στα στρώματα του νευρωνικού δικτύου θέτοντας με πιθανότητα p ενισχύσεις νευρώνων ίσες με 0 και διαιρώντας όλες τις υπόλοιπες ενισχύσεις με $1 - p$. Η διαίρεση των νευρώνων του στρώματος με το κλάσμα των νευρώνων που επιβίωσαν συνεπάγεται την κανονικοποίηση του στρώματος αφού το συνολικό άθροισμα των σημάτων ενίσχυσης του στρώματος θα είναι ίσο με αυτό που προέκυπτε πριν την εφαρμογή του Dropout. [24] Με βάση τα παραπάνω, οι ενισχύσεις των νευρώνων ενός κρυφού στρώματος με Dropout πιθανότητας p μεταβάλλονται ως εξής:

$$h' = \begin{cases} 0, & \text{με πιθανότητα } p, \\ \frac{h}{1-p}, & \text{αλλιώς.} \end{cases} \quad (1.32)$$

Σχηματικά, η εφαρμογή Dropout στο κρυφό στρώμα ενός νευρωνικού δικτύου με ένα κρυφό στρώμα φαίνεται παρακάτω:



Σχήμα 1.15: Απλό MLP πριν και μετά την εφαρμογή Dropout 60% στο κρυφό στρώμα

1.4 Το Θεώρημα Καθολικής Προσέγγισης (Universal Approximation Theorem)

Στην πορεία του κεφαλαίου αυτού, παρουσιάστηκαν οι βασικές μαθηματικές αρχές που διέπουν τη θεμελίωση και εκπαίδευση των τεχνητών νευρωνικών δικτύων εμπρόσθιας τροφοδότησης ή MLPs. Τα τεχνητά νευρωνικά δίκτυα χρησιμοποιούνται για την προσέγγιση οποιασδήποτε συναρτησιακής σχέσης και μάλιστα, αποδεικνύεται ότι η προσεγγιστική δύναμη τους είναι καθολική. Αυτό πρακτικά σημαίνει ότι οι δυνατότητες των νευρωνικών δικτύων ως προς την μοντελοποίηση προβλημάτων είναι απεριόριστες και για αυτό, μπορούν να αποτελέσουν ένα πολύ σημαντικό εργαλείο για τους κλάδους της χημικής μηχανικής, της χημείας, της βιολογίας και της ιατρικής. Στην συνέχεια της εργασίας θα γίνει παρουσίαση των συνελκτικών νευρωνικών δικτύων, καθώς και της επέκτασής τους στην μαθηματική αποτύπωση των γράφων που αυξάνουν ακόμα περισσότερο τις δυνατότητες αυτού του τομέα τεχνητής νοημοσύνης. Πιο συγκεκριμένα, όσον αφορά τα MLPs ισχύει

το θεώρημα καθολικής προσέγγισης, το οποίο παρουσιάζεται παρακάτω χωρίς απόδειξη λόγω της αυξημένης μαθηματικής περιπλοκότητάς του [25] [26]:

Θεώρημα

Κάθε συνεχής συνάρτηση ορισμένη στον n -διάστατο μοναδιαίο υπερκύβο (έστω f) μπορεί να προσεγγιστεί από ένα πεπερασμένο άθροισμα του τύπου :

$$\Psi(\mathbf{x}) = \sum_{i=1}^N v_i \phi(\mathbf{w}_i \mathbf{x} + b_i) , \quad (1.33)$$

όπου $v_i, b_i \in \mathbb{R}$, $\mathbf{w}_i \in \mathbb{R}^n$ και ϕ μια συνεχής μεροληπτική συνάρτηση ή ισοδύναμα :

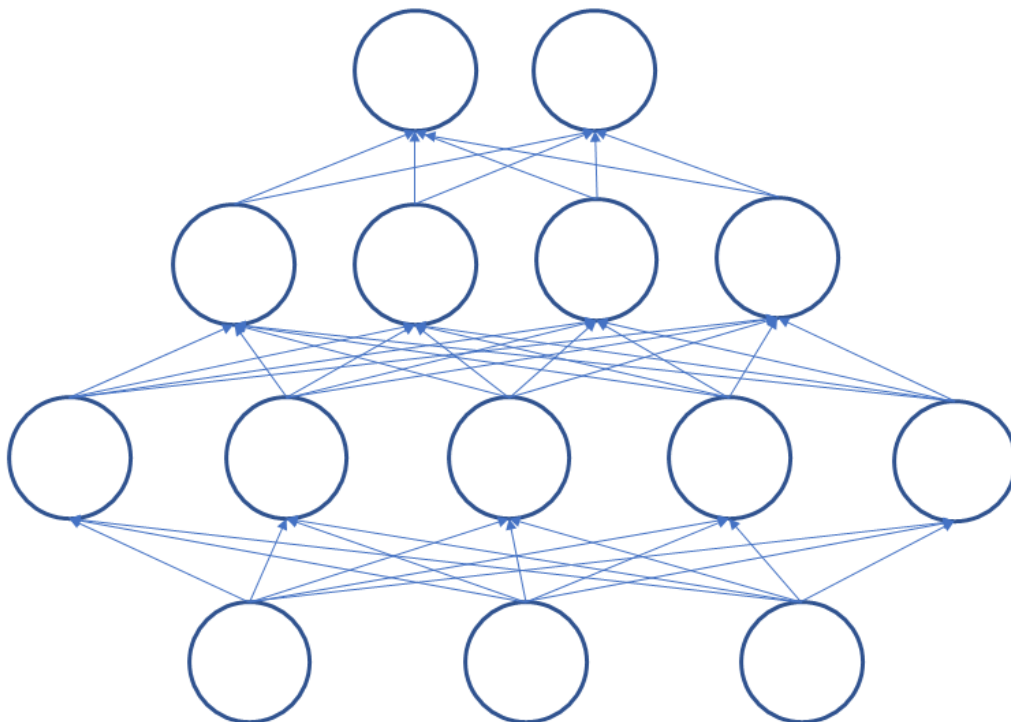
$$\Psi(\mathbf{x}) - f(\mathbf{x}) < \epsilon, \forall \epsilon \in \mathbb{R} . \quad (1.34)$$

Μια συνάρτηση λέγεται μεροληπτική (discriminatory) όταν ισχύει η συνεπαγωγή:

$$\int_{I_n} \phi(\mathbf{w}_i \mathbf{x} + b_i) d\mu(x) = 0 \Rightarrow \mu(x) = 0, \forall \mathbf{w}_i \in \mathbb{R}^n, b_i \in \mathbb{R} , \quad (1.35)$$

όπου $\mu(x)$ το μέτρο.

Σημειώνεται ότι το πεπερασμένο άθροισμα $\Psi(\mathbf{x})$ εκφράζει ένα νευρωνικό δίκτυο ενός κρυφού στρώματος N νευρώνων. Δηλαδή ένα απλό MLP με ένα κρυφό στρώμα είναι ικανό να προσεγγίσει οποιαδήποτε συνεχή συνάρτηση. Στην πραγματικότητα, χρησιμοποιούνται βαθύτερα νευρωνικά δίκτυα (περισσότερα του ενός κρυφά στρώματα), καθώς μπορούν να προσεγγίσουν συναρτήσεις με πιο σιβαρό τρόπο. Τα νευρωνικά δίκτυα με περισσότερα του ενός κρυφά στρώματα καλούνται βαθιά νευρωνικά δίκτυα και εκπροσωπούν την υποκατηγορία της μηχανικής μάθησης που καλείται βαθιά μάθηση (deep learning).



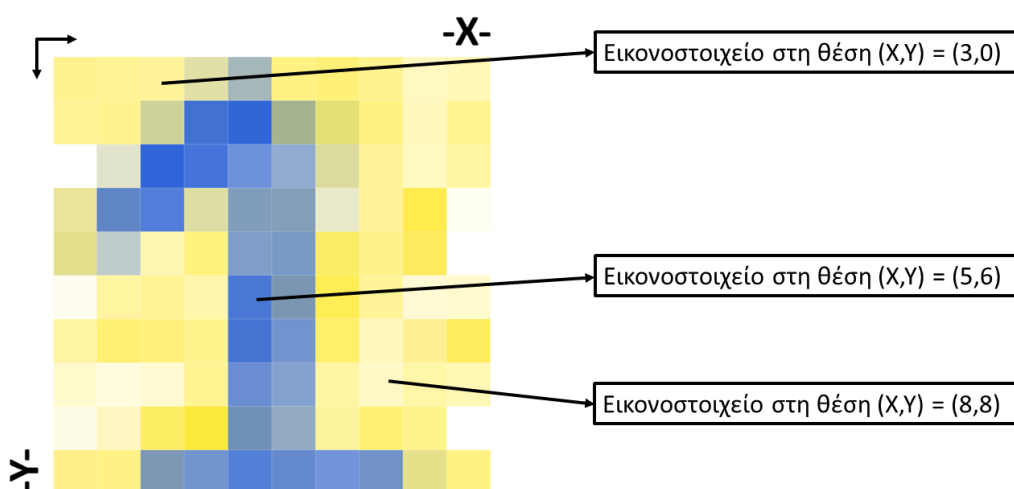
Σχήμα 1.16: Βαθύ νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης με 2 κρυφά στρώματα

Κεφάλαιο 2

Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks)

2.1 Οι Εικόνες ως Δεδομένα Εισόδου σε Νευρωνικά Δίκτυα (Images as Neural Networks Input Data)

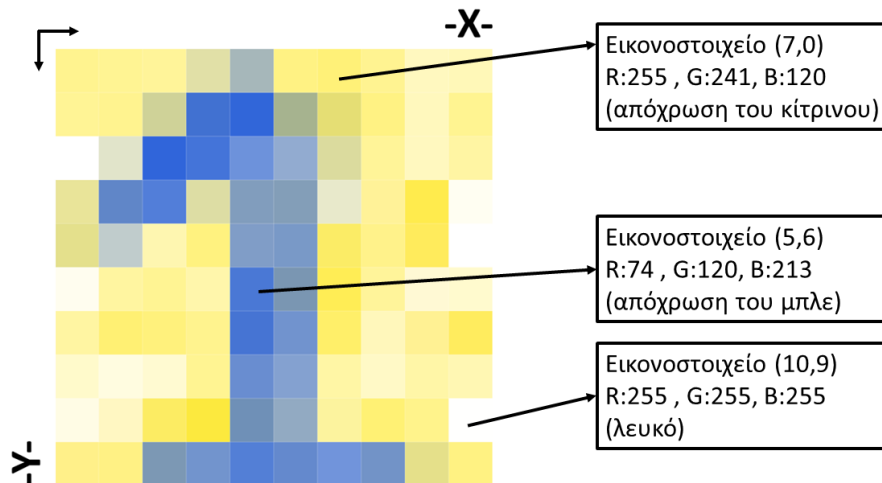
Τα συνελικτικά νευρωνικά δίκτυα (convolutional neural networks), τα οποία εισήχθησαν από τον LeCun το 1989 [27], αποτελούν ένα είδος νευρωνικών δικτύων που επεξεργάζονται δεδομένα με συγκεκριμένη τοπολογία υπό τη μορφή πλέγματος (data with known grid-like topology) [7]. Παραδείγματα τέτοιων δεδομένων είναι οι εικόνες, η επεξεργασία και εξαγωγή μοτίβων από τις οποίες αποτελεί, ίσως, το κυριότερο πεδίο εφαρμογής των συνελικτικών νευρωνικών δικτύων. Πράγματι, οι εικόνες μπορούν να θεωρηθούν ως ένα δισδιάστατο πλέγμα από εικονοστοιχεία (ή pixels), το οποίο έχει μια συγκεκριμένη διάταξη στον χώρο, όπως ορίζεται από την ίδια την φύση των εικόνων. Ως παράδειγμα δίδεται μια εικόνα 10×10 εικονοστοιχείων που παρουσιάζεται σε μεγέθυνση υπό την μορφή ενός X/Y πλέγματος:



Σχήμα 2.1: Απεικόνιση σε μεγεθυμένο X/Y πλέγμα μιας εικόνας 10×10 εικονοστοιχείων

Η πληροφορία που περιλαμβάνει ένα εικονοστοιχείο είναι, συνήθως, ένα διάνυσμα 3 στοι-

χείων που αντιστοιχούν στα 3 χρωματικά κανάλια: κόκκινο, πράσινο και μπλε (RGB). Τα κανάλια RGB λαμβάνουν τιμές από το 0 έως το 255, οι οποίες αντιστοιχούν στην φωτεινότητα του χρώματος που υποδηλώνουν. Με αυτόν τον τρόπο, μπορεί να καλυφθεί όλο το χρωματικό εύρος με σχετικά καλή ανάλυση. Ακολουθεί παράδειγμα απεικόνισης καναλιών με το σύστημα RGB επιλεγμένων εικονοστοιχείων του προηγούμενου παραδείγματος:



Σχήμα 2.2: RGB κανάλια και αντίστοιχη απόχρωση επιλεγμένων εικονοστοιχείων εικόνας 10 × 10 εικονοστοιχείων υπό μορφή X/Y πλέγματος

Σύμφωνα με τα παραπάνω, μια εικόνα θα μπορούσε μαθηματικά να αποτυπωθεί ως ένας τανυστής τρίτης τάξης, του οποίου οι 2 πρώτες διαστάσεις καθορίζονται από το μέγεθος της σε εικονοστοιχεία και η τρίτη διάστασή της καθορίζεται από τα χρωματικά κανάλια (στην περίπτωση του συστήματος RGB 3). Οι τανυστές τρίτης τάξης μπορούν να αποτελέσουν είσοδο ενός νευρωνικού δικτύου εμπρόσθιας τροφοδότησης σύμφωνα με τα όσα έχουν αναφερθεί στο κεφάλαιο 1. Τι συμβαίνει, όμως, με μια φωτογραφία που έχει μέγεθος 1 megapixel; Ο τανυστής τρίτης τάξης που την αντιπροσωπεύει θα έχει διάσταση της τάξης 10^6 . Ακόμα και μία δραστική μείωση στις 10^3 κρυφές διαστάσεις θα επέφερε τουλάχιστον $10^6 \times 10^3 = 10^9$ εκπαιδευσιμες παραμέτρους στο νευρωνικό δίκτυο. Η εκπαίδευση ενός τέτοιου νευρωνικού δικτύου προσδοκά τεράστια υπολογιστική ισχύ, ενώ μπορεί να καταστεί ακόμα και υπολογιστικά αδύνατη. Λύση σε αυτό το πρόβλημα επιχειρούν να δώσουν τα συνελκτικά νευρωνικά δίκτυα.

Συνήθως, σκοπός ενός νευρωνικού δικτύου, το οποίο λαμβάνει ως είσοδο μια εικόνα, είναι η ανίχνευση ενός αντικειμένου που περιλαμβάνεται σε αυτήν. Σε προβλήματα ταξινόμησης, το αντικείμενο αυτό αναπαριστά μια κλάση του προβλήματος που πρέπει να διακριθεί από τις υπόλοιπες. Παραδείγματα θα μπορούσαν να αποτελούν η αναγνώριση γάτας/σκύλου, όπου οι κλάσεις είναι 2 ή η αναγνώριση χειρόγραφων αριθμητικών ψηφίων, όπου οι κλάσεις είναι 10 (όσες και τα αριθμητικά ψηφία). Φαίνεται λογικό πως οποιαδήποτε μέθοδος ανίχνευσης αντικειμένων, όπως είναι και το ανθρώπινο μάτι, δεν θα νοιάζεται για την ακριβή θέση του αντικειμένου στην εικόνα. Για παράδειγμα, όταν τίθεται το πρόβλημα της αναγνώρισης της γάτας από τον σκύλο σε έναν άνθρωπο, αυτός δεν ενδιαφέρεται αν η γάτα είναι στην κορυφή ή σε μια γωνία της φωτογραφίας, αρκεί ότι υπάρχει σε αυτήν. Το συνελκτικό νευρωνικό δίκτυο προσπαθεί να συστηματοποιήσει την ιδέα αυτή

της χωρικής ανεξαρτησίας (spatial invariance) βασισμένο σε δύο άξονες [5]:

- Το νευρωνικό δίκτυο πρέπει να αντιμετωπίζει κάθε τμήμα της εικόνας με συγκεκριμένο τρόπο ανεξάρτητα από την θέση του σε αυτή.
- Το νευρωνικό δίκτυο πρέπει να επικεντρώνεται σε τοπικές περιοχές της εικόνας, χωρίς να ενδιαφέρεται για περιεχόμενα της εικόνας σε μακρινή απόσταση. Τελικά, αυτές οι τοπικές αναπαραστάσεις μπορούν να συσσωρευθούν και να παράξουν προβλέψεις σε επίπεδο ολόκληρης εικόνας.

Τα συνελκτικά νευρωνικά δίκτυα επιτυγχάνουν τα παραπάνω χρησιμοποιώντας την μαθηματική διαδικασία της συνέλιξης (convolution operation).

2.2 Η Διαδικασία της Συνέλιξης (Convolution Operation)

Η συνέλιξη μεταξύ δύο συναρτήσεων $f, g : \mathbb{R} \rightarrow \mathbb{R}$ συμβολίζεται με αστερίσκο (*) και ορίζεται ως εξής:

$$(f * g)(x) = \int f(z)g(x - z)dz . \quad (2.1)$$

Επομένως, η συνέλιξη υπολογίζει την επικάλυψη των συναρτήσεων f και g αν η g "αναστραφεί" ως προς τον άξονα y και μετατοπιστεί κατά x . Με βάση την ορολογία των συνελκτικών νευρωνικών δικτύων, η συνάρτηση f αναφέρεται ως είσοδος (input) και η συνάρτηση g ως πυρήνας (kernel). Για να γίνει αντιληπτή η φυσική σημασία της συνέλιξης δίνεται το εξής παράδειγμα:

Έστω ότι μετράται η θέση ενός διαστημόπλοιου με αισθητήρα λέιζερ. Ο αισθητήρας παράγει μια μοναδική έξοδο $h(t)$ που αντιστοιχεί στη θέση του διαστημόπλοιου σε χρόνο t . Έστω τώρα ότι το σήμα του αισθητήρα είναι θορυβώδες. Τότε για μια καλύτερη εκτίμηση της θέσης του διαστημόπλοιου είναι επιθυμητός ο υπολογισμός ενός μέσου όρου των μετρήσεων του αισθητήρα. Φυσικά, οι νεότερες μετρήσεις είναι πιο σχετικές και έτσι, επιλέγεται ο υπολογισμός ενός σταθμισμένου μέσου όρου με περισσότερο βάρος να δίδεται στις νεότερες μετρήσεις. Ορίζεται, λοιπόν, μια συνάρτηση βάρους $w(a)$ όπου a η ηλικία της μέτρησης. Ο επιθυμητός σταθμισμένος μέσος όρος μετρήσεων θα δίδεται από την συνέλιξη των συναρτήσεων h και w : $(h * w)(t) = \int h(a)w(t - a)da$.

Στο παραπάνω παράδειγμα, το σήμα που αντιστοιχεί στη μέτρηση του αισθητήρα είναι μια συνεχής συνάρτηση, κάτι που είναι μη-ρεαλιστικό. Στην πραγματικότητα, ένα όργανο παράγει μετρήσεις σε διακριτό χρόνο, δηλαδή ανά συγκεκριμένα χρονικά διαστήματα. Έτσι, το σήμα h έχει τη μορφή διανύσματος πεπερασμένου μήκους και η συνέλιξη δίνεται από την σχέση:

$$(h * w)(t) = \sum_a h(a)w(t - a) . \quad (2.2)$$

Όπως έχει ήδη αναφερθεί, οι είσοδοι στα νευρωνικά δίκτυα συνήθως είναι τανυστές τάξης μεγαλύτερης του 1, όπως τα διανύσματα του προηγούμενου παραδείγματος. Οι έγχρωμες εικόνες, λόγω της ύπαρξης των χρωματικών καναλιών RGB είναι τανυστές τρίτης τάξης. Από την άλλη, οι ασπρόμαυρες εικόνες μπορούν να αναπαρασταθούν ως τανυστές δεύτερης τάξης, όπου σε κάθε εικονοστοιχείο αντιστοιχεί μια (αντί για τρεις) τιμή φωτεινότητας από το 0 έως το 255. Για λόγους μαθηματικής απλοποίησης και ευκολίας απεικόνισης,

τα επόμενα παραδείγματα θα θεωρούν ασπρόμαυρες εικόνες, δηλαδή τανυστές δεύτερης τάξης ως δεδομένα εισόδου. Έστω, λοιπόν, η δισδιάστατη εικόνα \mathbf{I} και ο δισδιάστατος πυρήνας \mathbf{K} . Τότε για τη συνέλιξη τους ισχύει:

$$[\mathbf{I} * \mathbf{K}]_{ij} = \sum_m \sum_n [\mathbf{I}]_{m,n} [\mathbf{K}]_{i-m,j-n} , \quad (2.3)$$

όπου υπολογίζεται το στοιχείο του πίνακα που προκύπτει έπειτα από τη συνέλιξη στη θέση (i,j) .

Η αντιμεταθετική ιδιότητα ισχύει στην συνέλιξη και έτσι, ισχύει:

$$[\mathbf{K} * \mathbf{I}]_{ij} = \sum_m \sum_n [[\mathbf{I}]_{i-m,j-n} \mathbf{K}]_{m,n} . \quad (2.4)$$

Η αντιμεταθετική ιδιότητα προκύπτει από την περιστροφή του πυρήνα σε σχέση με την είσοδο, με την έννοια ότι καθώς το m αυξάνεται, τόσο ο δείκτης της εισόδου μειώνεται και ο δείκτης του πυρήνα αυξάνεται. Τα παραπάνω δεν έχουν κάποια πρακτική χρησιμότητα στην κατανόηση των συνελκτικών νευρωνικών δικτύων, αλλά είναι χρήσιμα σε μερικές μαθηματικές αποδείξεις. Παρ' όλα αυτά, πολλές βιβλιοθήκες νευρωνικών δικτύων αντί για τη συνέλιξη χρησιμοποιούν μια αντίστοιχη μαθηματική διαδικασία αλλά χωρίς την περιστροφή του πυρήνα, η οποία καλείται δια-συσχέτιση (cross-correlation) και συμβολίζεται με αστέρι (\star):

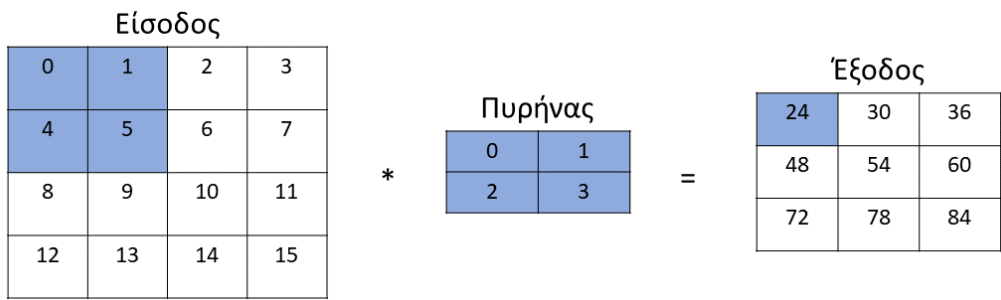
$$[\mathbf{K} \star \mathbf{I}]_{ij} = \sum_m \sum_n [[\mathbf{I}]_{i+m,j+n} \mathbf{K}]_{m,n} . \quad (2.5)$$

Εφεξής και όπως συνηθίζεται βιβλιογραφικά, η μαθηματική διαδικασία της δια-συσχέτισης θα καλείται συνέλιξη χωρίς περιστροφή πυρήνα.

Για τη συνέλιξη χωρίς περιστροφή πυρήνα θεωρώντας ως είσοδο έναν πίνακα $n_I \times m_I$ και ως πυρήνα έναν πίνακα $n_K \times m_K$, ορίζεται ο υποπίνακας του πίνακα εισόδου, παράθυρο συνέλιξης (convolution window), το μέγεθος του οποίου ισούται με το μέγεθος του πυρήνα. Η διαδικασία της συνέλιξης χωρίς περιστροφή πυρήνα είναι η εξής: ξεκινώντας το παράθυρο συνέλιξης από την πάνω αριστερή γωνία και κυλώντας το κατά μήκος του πίνακα εισόδου, εφαρμόζεται γινόμενο Hadamard μεταξύ του παραθύρου συνέλιξης και του πυρήνα. Το άθροισμα των στοιχείων κάθε νέου πίνακα που προκύπτει από το γινόμενο Hadamard αποτελεί και ένα στοιχείο του πίνακα εξόδου. Το μέγεθος του πίνακα εξόδου δίνεται από την εξής σχέση:

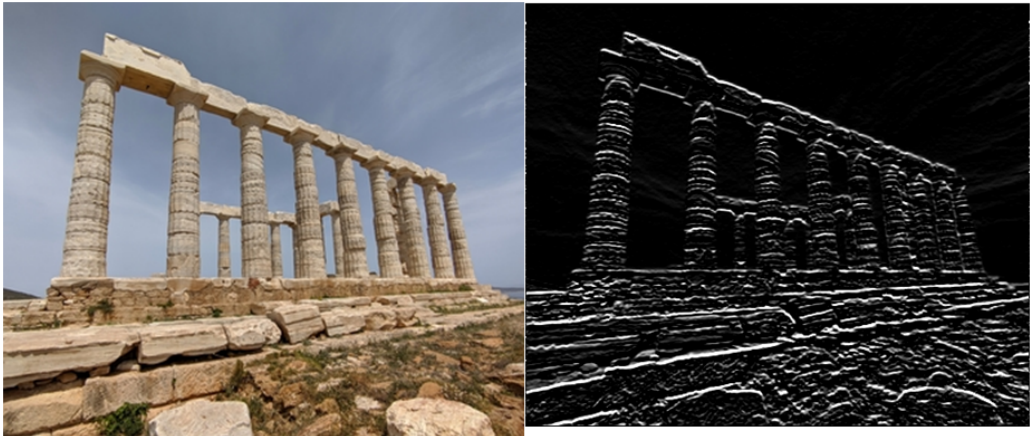
$$(n_I - n_K + 1) \times (m_I - m_K + 1) . \quad (2.6)$$

Ακολουθεί αριθμητικό παράδειγμα δισδιάστατης συνέλιξης χωρίς περιστροφή πυρήνα:



Σχήμα 2.3: Σχηματική απεικόνιση συνέλιξης χωρίς περιστροφή πυρήνα. Το παράθυρο συνέλιξης (γαλάζιο στον πίνακα εισόδου) πολλαπλασιάζεται ανά στοιχείο με τον πυρήνα και κατόπιν πρόσθεσης των στοιχείων του νέου πίνακα παράγει ένα στοιχείο του πίνακα εξόδου. Κάθε φορά μεταφέρεται κατά μία θέση δεξιά κατά μήκος του πίνακα εισόδου.

Μέσω της διαδικασίας της συνέλιξης επιτυγχάνεται ο εντοπισμός σημαντικών τοπικών χαρακτηριστικών των δεδομένων εισόδου που στην προκειμένη περίπτωση είναι εικόνες. Πιο συγκεκριμένα, μπορούν να επιτευχθούν διάφορες διαδικασίες επεξεργασίας της εικόνας ανάλογα με την επιλογή πυρήνα (ή φίλτρου όπως συνηθίζεται να αποκαλείται στην επεξεργασία εικόνας), όπως θάμπωμα (blurring), όξυνση (sharpening), εντοπισμός ακμών (edge detection) κλπ [28]. Ακολουθεί ένα παράδειγμα συνέλιξης για εντοπισμό ακμών:



Σχήμα 2.4: Αριστερά: φωτογραφία του Ναού του Ποσειδώνα στο Σούνιο μεγέθους 320 × 293 εικονοστοιχείων. Δεξιά: Η εικόνα μετά την συνέλιξη με επιλεγμένο πυρήνα για εντοπισμό ακμών. Η παραγωγή της επεξεργασμένης εικόνας έγινε μέσω του online λογισμικού <https://setosa.io/ev/image-kernels/>

Στο παράδειγμα του σχήματος 2.4 πραγματοποιείται συνέλιξη με τον εξής 3 × 3 πυρήνα:

-1	-2	-1
0	0	0
1	2	1

Ο πίνακας εξόδου, λοιπόν, θα έχει διάσταση $(320 - 3 + 1) \times (293 - 3 + 1) = 318 \times 291$. Επομένως, με τη διαδικασία της συνέλιξης γίνονται $318 \times 291 \times 10 = 925380$ πράξεις αριθμών κινητής υποδιαστολής (floating point numbers) (9 πολλαπλασιασμοί κατά το γινόμενο Hadamard και 1 πρόσθεση). Από την άλλη, εάν η είσοδος ήταν πλήρως συνδεδεμένη με την έξοδο, δηλαδή η εικόνα εισερχόταν σε ένα νευρωνικό δίκτυο εμπρόσθιας

τροφοδότησης, τότε σύμφωνα με όσα έχουν αναλυθεί στο κεφάλαιο 1, θα χρειάζονταν, λόγω του πολλαπλασιασμού πινάκων, $320 \times 293 \times 318 \times 291$ ή σχεδόν 9 δισεκατομμύρια πράξεις μεταξύ αριθμών κινητής υποδιαστολής. Επομένως, γίνεται κατανοητή η σπουδαία εξοικονόμηση υπολογιστικής ισχύος που επιφέρει η αραιή σύνδεση (sparse connectivity) των συνελκτικών νευρωνικών δικτύων, χωρίς ταυτόχρονα να χάνονται σημαντικά χαρακτηριστικά της αρχικής εικόνας.

Ένα πρόβλημα που μπορεί να προκύψει με την εφαρμογή πολλαπλών συνελκτικών στρώματων το ένα μετά το άλλο είναι η περικοπή της αρχικής εικόνας σε σημείο να χάνονται σημαντικές πληροφορίες της. Στο προηγούμενο παράδειγμα, με εφαρμογή πυρήνα 3×3 το μέγεθος της εικόνας ελαττώθηκε από 320×293 σε 318×291 . Όπως φαίνεται και στο σχήμα 2.4, στην συγκεκριμένη περίπτωση η εικόνα δεν έχασε σημαντική πληροφορία μετά την περικοπή που συνέβη λόγω της εφαρμογή συνέλιξης. Αυτό θα συνέβαινε, όμως, εάν η αρχιτεκτονική του μοντέλου περιλάμβανε, για παράδειγμα, 10 συνελκτικά στρώματα όπου εφαρμόζονταν ο ίδιος πυρήνας μεγέθους 3×3 . Μία άμεση λύση στο ζήτημα αυτό είναι η προσθήκη ενός επιπλέον στρώματος εικονοστοιχείων περιμετρικά της αρχικής εικόνας. Τυπικά, τα νέα εικονοστοιχεία έχουν την τιμή 0. Η τακτική αυτή συναντάται στη βιβλιογραφία με τον αγγλικό όρο padding [5]. Ακολουθεί ένα γραφικό παράδειγμα όπου εφαρμόζεται padding 2 σειρών και 2 στηλών νέων εικονοστοιχείων σε έναν πίνακα εισόδου:

Είσοδος με padding δύο γραμμών
και δύο στηλών

0	0	0	0	0	0
0	0	1	2	3	0
0	4	5	6	7	0
0	8	9	10	11	0
0	12	13	14	15	0
0	0	0	0	0	0

* Πυρήνας

0	1
2	3

= Έξοδος

0	3	8	13	6
12	24	30	36	14
28	48	54	60	22
44	72	78	84	30
12	13	14	15	0

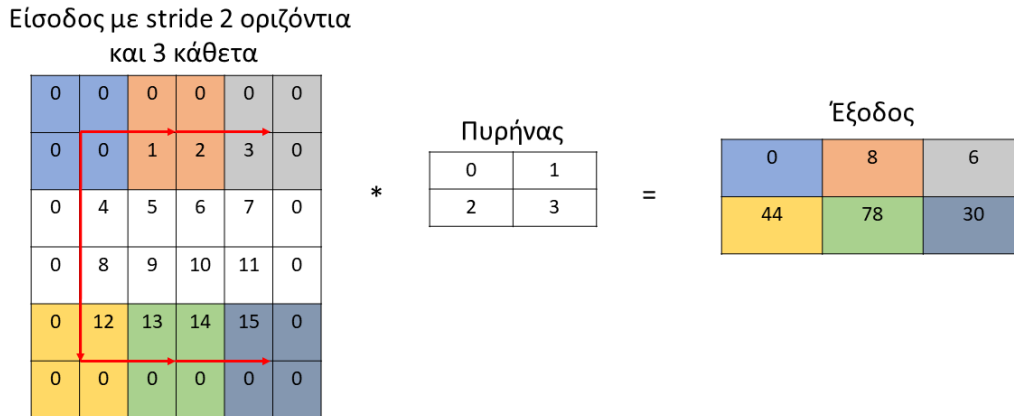
Σχήμα 2.5: Σχηματική απεικόνιση συνέλιξης χωρίς περιστροφή πυρήνα κατόπιν εφαρμογής padding μίας σειράς

Σημειώνεται ότι τυπικά κατά το padding επιλέγεται ζυγός αριθμός σειρών και στηλών καθώς οι μισές κατανέμονται από την μία πλευρά του πίνακα και οι άλλες μισές από την άλλη (δεξιά και αριστερά για στήλες, πάνω και κάτω για σειρές). Το μέγεθος του πίνακα εξόδου μετά την εφαρμογή padding n_p σειρών και m_p στηλών βρίσκεται από την τροποποιημένη εξίσωση 2.6:

$$(n_l - n_k + n_p + 1) \times (m_l - m_k + m_p + 1) . \quad (2.7)$$

Άλλη μια "τροποποίηση" της διαδικασίας της συνέλιξης που χρησιμοποιείται είτε για σκοπούς εξοικονόμησης υπολογιστικού κόστους είτε για σκοπούς αποκλιμάκωσης του αρχικού όγκου δεδομένων είναι το άλμα (εφεξής με τον αγγλικό όρο stride). Με την εφαρμογή stride, το παράθυρο συνέλιξης κινείται κατά μήκος του πίνακα εισόδου παραλείποντας συγκεκριμένο αριθμό στηλών ή σειρών που ορίζεται από τον χρήστη. Για παράδειγμα,

με stride 2 οριζόντια, παραλείπεται κάθε φορά 1 στήλη (ο πολλαπλασιασμός ανά στοιχείο και το άθροισμα γίνονται 2 στήλες μετά), ενώ με stride 3 κατακόρυφα, όταν το παράθυρο φτάσει στο τέλος μιας γραμμής παραλείπονται οι 2 επόμενες γραμμές(ο πολλαπλασιασμός ανά στοιχείο και το άθροισμα γίνονται 3 γραμμές μετά). Η εφαρμογή stride 2 οριζόντια και 4 κάθετα μπορεί να οπτικοποιηθεί με βάση το παράδειγμα πίνακα εισόδου του σχήματος 2.5 ως εξής:



Σχήμα 2.6: Σχηματική απεικόνιση συνέλιξης χωρίς περιστροφή πυρήνα κατόπιν εφαρμογής stride 2 οριζόντια και 4 κάθετα

Το μέγεθος του πίνακα εξόδου που προκύπτει μετά από εφαρμογή stride n_s οριζόντια και m_s κατακόρυφα βρίσκεται από τον εξής τύπο (ο οποίος περιλαμβάνει και την περίπτωση padding):

$$\frac{n_I - n_K + n_P + n_S}{n_S} \times \frac{m_I - m_K + m_P + m_S}{m_S} \quad (2.8)$$

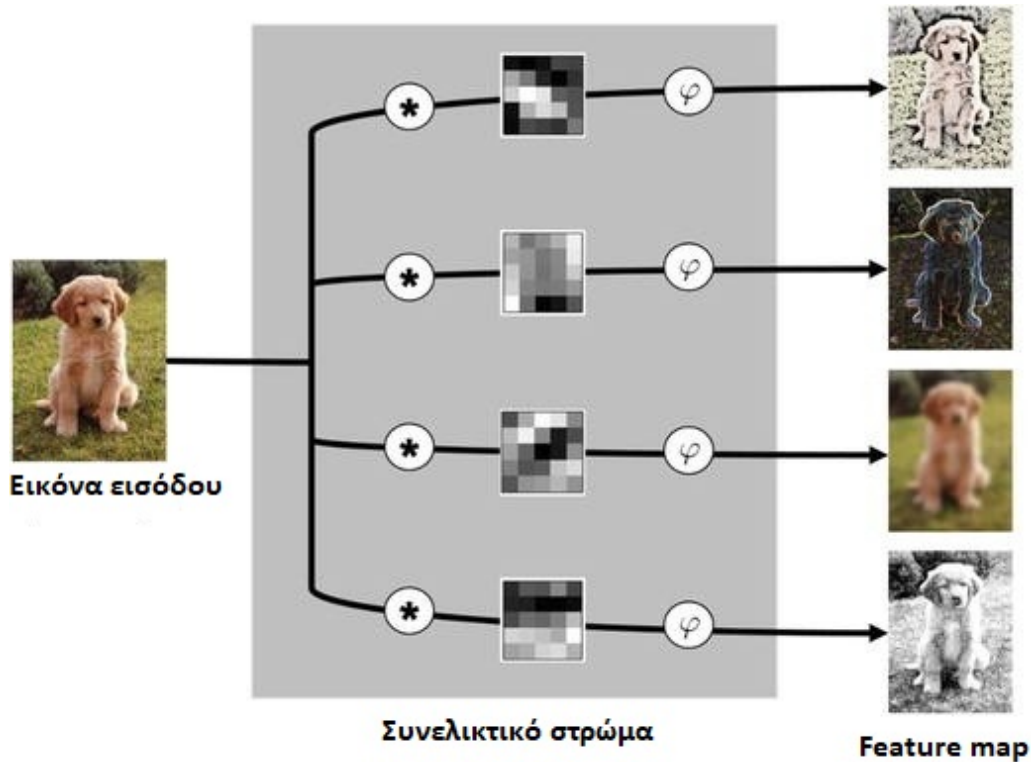
2.3 Στρώματα ενός Συνελικτικού Νευρωνικού Δικτύου (Layers of a Convolutional Neural Network)

Ένα τυπικό συνελικτικό νευρωνικό δίκτυο αποτελείται από ένα ή περισσότερα συνελικτικά στρώματα (convolutional layers), τα οποία συχνά ακολουθούνται από ένα ή περισσότερα στρώματα υποδειγματοληψίας (pooling layers) και πιθανόν, από ένα ή περισσότερα πλήρως συνδεδεμένα στρώματα (fully-connected layers). [29]

2.3.1 Το Συνελικτικό Στρώμα (Convolutional Layer)

Σε ένα συνελικτικό στρώμα εφαρμόζεται η διαδικασία της συνέλιξης μέσω ενός εκπαιδευμένου πυρήνα. Σε περίπτωση που η είσοδος είναι εικόνα τότε αυτή αποτελεί τανυστή δεύτερης ή τρίτης τάξης (ασπρόμαυρη ή έγχρωμη αντίστοιχα) και ο πυρήνας καλείται φίλτρο. Η επεξεργασμένη εικόνα που παράγεται καλείται χάρτης χαρακτηριστικών (feature map). Παράδειγμα feature map είναι η δεξιά εικόνα του σχήματος 2.4. Η ονομασία αυτή δεν είναι τυχαία, καθώς όπως έχει αναφερθεί στο υποκεφάλαιο 2.2, η συνέλιξη έχει το σκοπό της εξαγωγής σημαντικών τοπικών χαρακτηριστικών της εικόνας επιτυγχάνοντας ταυτόχρονα αραιές συνδέσεις μεταξύ εισόδου και εξόδου και άρα, εξοικονόμηση υπολογιστικής ισχύος. Ένα συνελικτικό στρώμα μπορεί να παράξει τόσα feature maps όσα και ο αριθμός των φίλτρων (ή πυρήνων) που περιλαμβάνει. Ακόμη, μετά την διαδικασία της

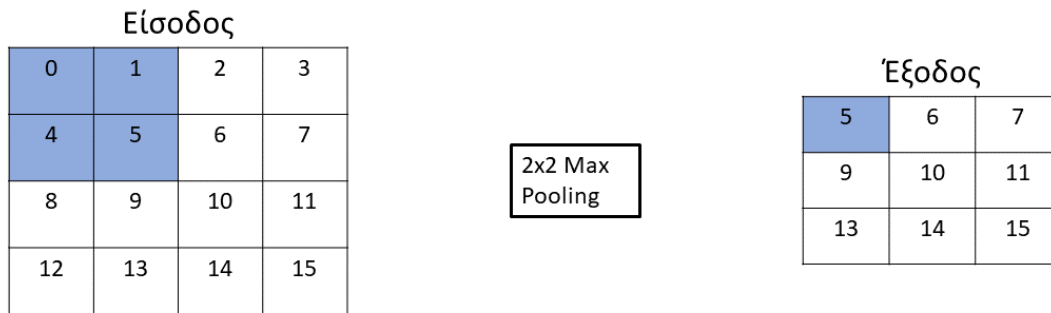
συνέλιξης και πριν την εξαγωγή του feature map εφαρμόζεται μια μη-γραμμική συνάρτηση ενεργοποίησης, ακριβώς όπως στην περίπτωση των πλήρως συνδεδεμένων στρωμάτων (στρώματα κλασσικών MLP - βλ. κεφάλαιο 1). Σε ένα συνεκτικό στρώμα, εκπαιδευσιμες παράμετροι είναι τα στοιχεία των πινάκων των φίλτρων, τα οποία σε κάθε επανάληψη ανανεώνονται με τους γνωστούς αλγορίθμους βελτιστοποίησης (βλ. υποκεφάλαιο 1.2.2). Συνοπτικά, η γραφική απεικόνιση ενός συνεκτικού στρώματος με 4 φίλτρα φαίνεται στο παρακάτω σχήμα :



Σχήμα 2.7: Συνεκτικό στρώμα 4 φίλτρων. Ο αστερίσκος δηλώνει την πράξη της συνέλιξης της αρχικής εικόνας με το εκάστοτε φίλτρο και το φ την μη-γραμμική συναρτήση ενεργοποίησης. [29]

2.3.2 Το Στρώμα Υποδειγματοληψίας (Pooling Layer)

Το στρώμα υποδειγματοληψίας ελαττώνει το μέγεθος μιας εικόνας, καθώς συνδυάζει γειτονικά εικονοστοιχεία σε μια περιοχή της εικόνας σε μία μοναδική αντιπροσωπευτική τιμή. Για παράδειγμα, η υποδειγματοληψία μεγίστου (max pooling), η οποία είναι η ευρύτερα χρησιμοποιούμενη συνάρτηση υποδειγματοληψίας, επιστρέφει την μέγιστη τιμή μιας ορθογώνιας γειτονιάς του πίνακα εισόδου του στρώματος. Η υποδειγματοληψία εφαρμόζεται στον πίνακα εισόδου με παρόμοιο τρόπο με την συνέλιξη, αφού και πάλι ένα παράθυρο κυλίνεται κατά μήκος του πίνακα εισόδου. Αυτή την φορά, όμως, η τιμή που εκχωρείται στην αντίστοιχη θέση του πίνακα εξόδου καθορίζεται από την συνάρτηση υποδειγματοληψίας (π.χ. μέγιστη για max pooling). Το μέγεθος του παραθύρου καθορίζεται μαζί με την συνάρτηση υποδειγματοληψίας. Εκτός από το max pooling, άλλες συναρτήσεις υποδειγματοληψίας είναι η υποδειγματοληψία αθροίσματος (sum pooling), η υποδειγματοληψία μέσου όρου (average pooling) και η υποδειγματοληψία L2 νόρμας (L2 norm pooling). Ακολουθεί ένα γραφικό παράδειγμα εφαρμογής 2×2 max pooling:



Σχήμα 2.8: Σχηματική απεικόνιση εφαρμογής 2×2 max pooling

Σκοπός του στρώματος υποδειγματοληψίας είναι η σταδιακή συσσώρευση πληροφοριών με ταυτόχρονη ελάττωση μεγέθους της εικόνας όσο το νευρωνικό δίκτυο βαθαίνει. Η εφαρμογή μιας στατιστικής συνάρτησης σε μια γειτονιά της εικόνας, καθιστά τις πληροφορίες που λαμβάνονται μη-εξαρτώμενες από μικρές αλλαγές των εισόδων που μπορεί να συμβαίνουν καθώς το νευρωνικό δίκτυο εκπαιδεύεται. Πράγματι, όπως αναλύθηκε στην αρχή του παρόντος κεφαλαίου (υποκεφάλαιο 2.1) κατά την επεξεργασία μιας εικόνας (αντίστοιχα, και οποιουδήποτε άλλου τύπου δεδομένο εισόδου σε συνελκτικό νευρωνικό δίκτυο), μας ενδιαφέρει περισσότερο η ίδια η παρουσία ενός σημαντικού χαρακτηριστικού ενός αντικειμένου, παρά η ακριβής θέση του. Έτσι, το δίκτυο πρέπει να είναι ικανό να εντοπίζει και να απομονώνει αυτό το χαρακτηριστικό ακόμα κι αν η θέση του δεν είναι σταθερή μετά από αλληπάλληλες εφαρμογές συνελκτικών φίλτρων.

2.3.3 Το Πλήρως Συνδεδεμένο Στρώμα (Fully Connected Layer)

Τα στρώματα των κλασικών νευρωνικών δικτύων εμπρόσθιας τροφοδότησης ή MLP χρησιμοποιούνται, συνήθως, στο τελικό στάδιο ενός συνελκτικού νευρωνικού δικτύου. Αποσκοπούν στην μετατροπή των feature maps, δηλαδή των εξόδων ενός συνελκτικού στρώματος ή ενός στρώματος υποδειγματοληψίας, σε κατανομή πιθανοτήτων κλάσεων. Για να το κατορθώσουν αυτό, αρχικά, το feature map μιας εικόνας, όπου για έγχρωμες εικόνες αποτελεί τανυστή τρίτης τάξης, μετασχηματίζεται σε τανυστή πρώτης τάξης, δηλαδή διάνυσμα. Το διάνυσμα αυτό έχει μήκος όσο και το πλήθος των στοιχείων του feature map και η διαδικασία με την οποία παράγεται συναντάται στη διεθνή βιβλιογραφία με τον όρο flattening. Κάθε στοιχείο του νέου διανύσματος αποτελεί και τιμή ενεργοποίησης ενός νευρώνα του στρώματος εισόδου του MLP. Η έξοδος του MLP, το οποίο πιθανότατα και σύμφωνα με τη θεωρία του κεφαλαίου 1 περιλαμβάνει κρυφά στρώματα, αποτελείται από τόσους νευρώνες όσες και οι κλάσεις του προβλήματος ταξινόμησης και η κάθε τιμή ενεργοποίησης των νευρώνων εξόδου έχει την μορφή πιθανότητας ($o_i \in [0, 1]$) που έχει προκύψει από εφαρμογή της συνάρτησης softmax (βλ. παράγραφο 1.3.3). Συμπερασματικά, ένα MLP, δηλαδή πλήρως συνδεδεμένα στρώματα σε σειρά, τοποθετημένο στο τέλος ενός συνελκτικού νευρωνικού δικτύου, μετατρέπουν δεδομένα εικόνας σε κατανομή πιθανοτήτων κλάσεων, κάτι που αποτελεί και το ζητούμενο ενός προβλήματος ταξινόμησης. [30]

2.4 Μέθοδοι Βελτίωσης Εκπαίδευσης Βαθιών Νευρωνικών Δικτύων (Methods for Improving Training of Deep Neural Networks)

2.4.1 Κανονικοποίηση Παρτίδας (Batch Normalization)

Η συνεχής ανανέωση των παραμέτρων στρώματων νευρωνικών δικτύων ως άμεσο επακόλουθο της εκπαίδευσης του νευρωνικού δικτύου, επιφέρει την συνεχή αλλαγή της κατανομής των εισόδων σε κάθε στρώμα. Έχει παρατηρηθεί ότι η παραπάνω συμπεριφορά επιβραδύνει την εκπαίδευση βαθιών νευρωνικών δικτύων, απαιτώντας μικρότερους ρυθμούς μάθησης και αρκετά προσεκτική επιλογή αρχικοποίησης των παραμέτρων. Λύση στο πρόβλημα μπορεί να δώσει η εφαρμογή κανονικοποίησης σε κάθε παρτίδα της εκπαιδευτικής διαδικασίας του μοντέλου. Όπως περιγράφηκε στην παράγραφο 1.2.2, επιλέγεται συχνά η εκπαίδευση ενός νευρωνικού δικτύου με χρήση σε κάθε επανάληψη μιας μικρής παρτίδας παραδειγμάτων προκαθορισμένου μεγέθους. Το βήμα κανονικοποίησης εφαρμόζεται στην είσοδο κάθε στρώματος ενός βαθιού νευρωνικού δικτύου, όπου εισάγονται οι κρυφές ενεργοποιήσεις της αρχικά επιλεγμένης παρτίδας παραδειγμάτων του συνόλου δεδομένων εκπαίδευσης. Ο μετασχηματισμός που διέπει την κανονικοποίηση παρτίδας στην είσοδο του στρώματος k ενός νευρωνικού δικτύου ορίζεται από τις ακόλουθες εξισώσεις [31]:

$$\hat{x}_i^{(k)} = \frac{x_i^{(k)} - \mu_B^{(k)}}{\sqrt{\sigma_B^{(k)2} + \epsilon}} \quad (2.9)$$

$$y_i^{(k)} = \gamma^{(k)} \hat{x}_i^{(k)} + \beta^{(k)} \quad (2.10)$$

όπου ο δείκτης $i \in [1, m]$ χρησιμοποιείται για να καθορίσει μια ενεργοποίηση της παρτίδας B μεγέθους m , μ_B ο μέσος όρος όλων των ενεργοποιήσεων της παρτίδας, σ_B^2 η διακύμανση των ενεργοποιήσεων της παρτίδας B , ϵ μικρή σταθερά που προτίθεται στον παρονομαστή για αριθμητική σταθερότητα, γ και β εκπαιδευσιμες παράμετροι του μοντέλου που ορίζουν την κανονικοποιημένη ενεργοποίηση y_i , οι οποίες ανανεώνονται μέσω του αλγορίθμου back-propagation.

2.4.2 Ενίσχυση Δεδομένων (Data Augmentation)

Η ενίσχυση δεδομένων είναι μια μέθοδος τεχνητής αύξησης του μεγέθους ενός συνόλου δεδομένων εκπαίδευσης συνελκτικών νευρωνικών δικτύων μέσα από την παραγωγή επεξεργασμένων εκδόσεων των εικόνων του συνόλου δεδομένων. Η μέθοδος αυτή βρίσκει εφαρμογή σε πολλούς τομείς, όπου δεν υπάρχει πρόσβαση σε μεγάλα δεδομένα (big data), όπως για παράδειγμα οι ιατρικές εικόνες. Μέσω της ενίσχυσης δεδομένων, εκτός από το μέγεθος αυξάνεται και η ποιότητα του συνόλου δεδομένων οδηγώντας με αυτόν τον τρόπο σε καλύτερη εκπαίδευση του μοντέλου. Σε αντίθεση με τεχνικές κανονικοποίησης της εκπαιδευτικής διαδικασίας για την αποφυγή υπερπροσαρμογής, όπως το Dropout ή το Batch Normalization, ενίσχυση δεδομένων αποτελεί μια μέθοδο που στοχεύει στην καρδιά του προβλήματος, την οποία αποτελεί το σύνολο δεδομένων εκπαίδευσης. Ορισμένες τεχνικές ενίσχυσης δεδομένων εικόνων είναι οι εξής [32]:

- Τυχαία οριζόντια ή κάθετη μετακίνηση των εικονοστοιχείων μιας εικόνας
- Περιστροφή μιας εικόνας

- Τυχαία επεξεργασία φωτεινότητας μιας εικόνας
- Τυχαία μεγέθυνση μιας εικόνας
- Προσθήκη θορύβου στην εικόνα
- Τυχαία διαγραφή εικονοστοιχείων της εικόνας
- Τυχαία αλλαγή χρωματικών καναλιών των εικονοστοιχείων της εικόνας
- Ανάμειξη εικόνων μέσα από εφαρμογή μέσου όρου στις τιμές των εικονοστοιχείων τους
- Προσθήκη feature maps από την διαδικασία εκπαίδευσης του συνελκτικού νευρωνικού δικτύου

Κεφάλαιο 3

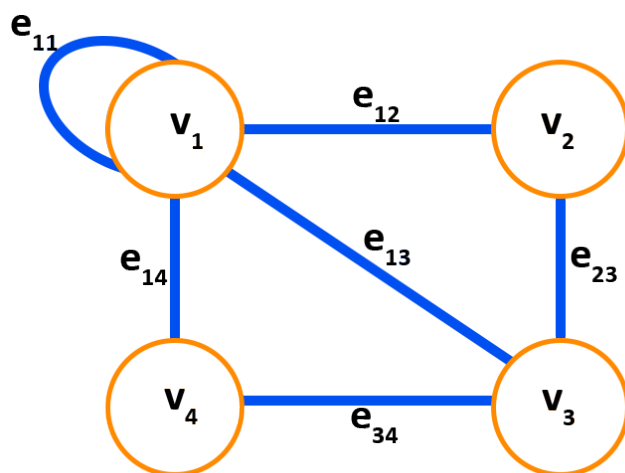
Νευρωνικά Δίκτυα Γράφων (Graph Neural Networks)

3.1 Θεωρία Γράφων (Graph Theory)

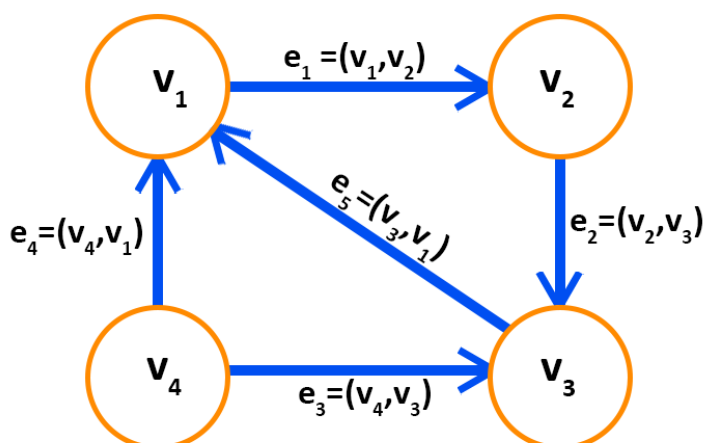
Στη σύγχρονη κοινωνία της πληροφορίας παρουσιάζεται σε όλο και περισσότερες πτυχές της ανθρώπινης ζωής η ανάγκη της μαθηματικής απεικόνισης ομάδων (δικτύων) αντικειμένων και των σχέσεων που αναπτύσσονται μεταξύ τους. Η αποτύπωση της πολυπλοκότητας τέτοιων δικτύων συσχετιζόμενων αντικειμένων μπορεί να γίνει με την απλή μαθηματική δομή του γράφου. Ένας γράφος αποτελείται από τους κόμβους (nodes/vertices), οι οποίοι αναπαρίστανται ως σημεία στον χώρο και εκφράζουν τα αντικείμενα ενός δικτύου και από τις ακμές (edges), οι οποίες αναπαρίστανται ως γραμμές που συνδέουν τους κόμβους και εκφράζουν τις σχέσεις μεταξύ των αντικειμένων. Παραδείγματα δικτύων αντικειμένων που μπορούν να κωδικοποιηθούν μέσω γράφων αντλούνται από αναρίθμητους τομείς της ανθρώπινης δραστηριότητας και μπορούν να είναι ένα κοινωνικό δίκτυο όπως το Facebook, ένας χάρτης πλοήγησης ή ένα χημικό μόριο. Στην περίπτωση του Facebook τους κόμβους μπορούν να αποτελούν οι χρήστες της πλατφόρμας και τις ακμές οι ενέργειες που επιτελούνται μεταξύ τους (π.χ. φιλία, like, ανταλλαγή μηνυμάτων, αποκλεισμός κ.λπ.), στον χάρτη πλοήγησης τους κόμβους αποτελούν τα διάφορα σημεία ενδιαφέροντος και τις ακμές οι δρόμοι, ενώ στο μόριο τους κόμβους αποτελούν τα άτομα και τις ακμές οι χημικοί δεσμοί. [33]

Ως γράφος ορίζεται το διατεταγμένο ζεύγος $G = (V, E)$, όπου V το σύνολο των κόμβων ή κορυφών και E το σύνολο των ακμών, δηλαδή διμερών σχέσεων μεταξύ των κόμβων του συνόλου V . Συνήθως, οι ακμές αναφέρονται σε δεσμούς μεταξύ δύο κόμβων του συνόλου V , οι οποίοι ονομάζονται παρακείμενοι ή γειτονικοί ή γειτνιάζοντες. [33] Παρ' όλα αυτά, σε έναν γράφο μπορεί να επιτραπεί ακμή μεταξύ ενός και μόνο κόμβου, η οποία ονομάζεται βρόχος.

Όταν όλα τα στοιχεία του συνόλου E ενός γράφου είναι διατεταγμένα ζεύγη (v_i, v_j) , όπου $v_i, v_j \in V$ ο γράφος ονομάζεται κατευθυνόμενος (Directed Graph). Τα διατεταγμένα ζεύγη (v_i, v_j) λέγονται τόξα ή διατεταγμένες ακμές. Ένα τόξο (v_i, v_j) κατευθύνεται από τον κόμβο v_i στον κόμβο v_j και απεικονίζεται με γραμμή που περιλαμβάνει το αντίστοιχο βέλος κατεύθυνσης. [33]



Σχήμα 3.1: Γράφος με 4 κόμβους και 6 ακμές όπου: $V = \{v_1, v_2, v_3, v_4\}$ και $E = \{e_{11}, e_{12}, e_{13}, e_{14}, e_{23}, e_{34}\}$. Σημειώνεται ότι στον γράφο αυτό υπάρχει πολλαπλή σύνδεση, δηλαδή τουλάχιστον δύο κόμβοι συνδέονται με περισσότερες από μία ακμές, καθώς επίσης η ακμή e_{11} είναι βρόχος.



Σχήμα 3.2: Κατευθυνόμενος γράφος με 4 κόμβους και 5 ακμές

3.1.1 Γειτνίαση και Πρόσπτωση στους Γράφους (Graph Adjacency and Incidency)

Δύο κόμβοι $v_i, v_j \in V$, οι οποίοι συνδέονται μέσω μίας ακμής $e_{ij} \in E$ λέγονται γειτνιάζοντες. Σε επίπεδο γράφου $G = (V, E)$, ορίζεται ο πίνακας γειτνίασης A (Adjacency matrix), ο οποίος είναι τετραγωνικός διαστάσεων $n \times n$, όπου n ο αριθμός των κόμβων ενός γράφου ή αλλιώς τάξη (order) του γράφου [33]:

$$A_{ij} = \begin{cases} 2, & \text{εάν } i = j \text{ και στον κόμβο } v_i \text{ υπάρχει βρόχος,} \\ 1, & \text{εάν οι κόμβοι } v_i, v_j, \text{ με } i \neq j, \text{ είναι γειτνιάζοντες,} \\ 0, & \text{αλλιώς.} \end{cases} \quad (3.1)$$

Για το παράδειγμα του σχήματος 3.1 ο πίνακας γειτνίασης είναι ο εξής:

$$A = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Μία ακμή $e_{ij} \in E$ λέγεται ότι προσπίπτει στους κόμβους $v_i, v_j \in V$. Αντίστοιχα με την σχέση γειτνίασης, ορίζεται ο πίνακας πρόσπτωσης διαστάσεων $m \times n$ (Incidence matrix), όπου m ο αριθμός των ακμών ενός γράφου ή αλλιώς μέγεθος (size) του γράφου:

$$B_{ij} = \begin{cases} 1, & \text{εάν ο κόμβος } v_i \text{ και η ακμή } e_j \text{ είναι προσπίπτοντες,} \\ 0, & \text{αλλιώς.} \end{cases} \quad (3.2)$$

Στην περίπτωση κατευθυνόμενου γράφου για τον πίνακα πρόσπτωσης ισχύει:

$$B_{ij} = \begin{cases} 1, & \text{εάν η ακμή } e_j \text{ έχει ως αρχή τον κόμβο } v_i, \\ -1, & \text{εάν η ακμή } e_j \text{ έχει ως τέλος τον κόμβο } v_i, \\ 0, & \text{αλλιώς.} \end{cases} \quad (3.3)$$

Για το παράδειγμα του σχήματος 3.2 ο πίνακας πρόσπτωσης είναι ο εξής:

$$B = \begin{bmatrix} 1 & 0 & 0 & -1 & -1 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

3.1.2 Εκχώρηση Πληροφοριών στους Γράφους (Assigning Graph Attributes)

Δύο βασικοί τύποι γράφων που περιλαμβάνουν εκχωρημένες πληροφορίες στους κόμβους ή/και στις ακμές τους είναι οι εξής [33]:

- Σταθμισμένοι Γράφοι (Weighted Graphs)
Σταθμισμένος ονομάζεται ο γράφος, του οποίου οι ακμές έχουν προσαρτημένα βάρη με τη μορφή ενός πραγματικού αριθμού. Τα βάρη ποσοτικοποιούν την σχέση μεταξύ των κόμβων που εκφράζουν οι εκάστοτε ακμές ενός γράφου.
- Γράφοι με Χαρακτηριστικά (Attributed Graphs)
Γράφος, του οποίου οι κόμβοι και οι ακμές περιλαμβάνουν προσαρτημένα χαρακτηριστικά (attributes), τα οποία μπορούν να είναι είτε ποιοτικά (π.χ. φύλο) είτε ποσοτικά (π.χ. ηλικία).

Το παράδειγμα του χημικού μορίου

Η αναπαράσταση του μορίου μιας χημικής ένωσης ως γράφος θα αποτελέσει ένα από τα κύρια αντικείμενα μελέτης της παρούσας εργασίας. Προς χάριν διευκόλυνσης της κατανόησης της έννοιας του γράφου με χαρακτηριστικά, παρατίθεται το παράδειγμα του μορίου του οξικού οξέος (CH_3COOH). Τα άτομα του μορίου εκτός από τα υδρογόνα (Hydrogen-suppressed graph) αποτελούν τους κόμβους του γράφου, ενώ οι δεσμοί μεταξύ των ατόμων αυτών αποτελούν τις ακμές του γράφου. Τα χαρακτηριστικά κόμβων και ακμών είναι ποιοτικά και ποσοτικά και για την περαιτέρω ανάλυση τους μέσω μηχανικής μάθησης ή και άλλων εφαρμογών προτιμάται η κωδικοποίησή τους σε μορφή διανυσμάτων. Σημειώνεται ότι η αναλυτική παρουσίαση της μεθοδολογίας χαρακτηρισμού των χημικών μορίων και

κόμβων, τότε για τα διαγώνια στοιχεία του πίνακα βαθμών $D \in R^{n \times n}$ ισχύει:

$$D_{ii} = \sum_j A_{ij} . \quad (3.4)$$

Ο Λαπλασιανός πίνακας (Laplacian matrix) γράφου $G = (V, E)$, ο οποίος συμβολίζεται ως $L(G)$, είναι ο πίνακας $n \times n$ που ορίζεται ως εξής:

$$L_{ij} = \begin{cases} -1, & \text{εάν ο κόμβος } v_i \text{ είναι γειτονικός του κόμβου } v_j, \\ 0, & \text{εάν ο κόμβος } v_i \text{ δεν είναι γειτονικός του κόμβου } v_j, \\ d_i, & \text{εάν } i = j. \end{cases} \quad (3.5)$$

Για τον υπολογισμό Λαπλασιανού πίνακα μη κατευθυνόμενου γράφου G ισχύει η παρακάτω εξίσωση:

$$L = D - A . \quad (3.6)$$

Επίσης, αν $B \in R^{m \times n}$ ο πίνακας πρόσπτωσης του γράφου G , τότε ισχύει:

$$L = BB^T . \quad (3.7)$$

Ο κανονικοποιημένος Λαπλασιανός πίνακας (Normalized Laplacian Matrix) \hat{L} ορίζεται ως εξής:

$$\hat{L}_{ij} = \begin{cases} 1, & \text{εάν } i = j \text{ και } d_i \neq 0, \\ -\frac{1}{\sqrt{d_i d_j}}, & \text{εάν } i \neq j \text{ και ο κόμβος } v_i \text{ είναι γειτονικός του κόμβου } v_j, \\ 0, & \text{αλλιώς.} \end{cases} \quad (3.8)$$

Για τον υπολογισμό του \hat{L} ισχύει:

$$\hat{L} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} . \quad (3.9)$$

3.1.4 Άλλες Διακρίσεις των Γράφων (Various Graph Types)

Μερικές κατηγορίες γράφων που δεν έχουν αποσαφηνιστεί μέχρι στιγμής είναι οι εξής [33]:

- **Τυπικός γράφος (Regular graph)**
Γράφος, του οποίου κάθε κόμβος έχει τον ίδιο αριθμό γειτονικών κόμβων, δηλαδή τον ίδιο βαθμό.
- **Πλήρης γράφος (Complete graph)**
Γράφος, του οποίου κάθε ζευγάρι κόμβων συνδέεται με μια ακμή.
- **Διμερής γράφος (Bipartite graph)**
Γράφος, του οποίου το σύνολο των κόμβων V μπορεί να διαιρεθεί σε δύο υποσύνολα W και X , για τα οποία ισχύει: $V = W \cup X$ και $W \cap X = \emptyset$ και κάθε ακμή συνδέει

έναν κόμβο του W με έναν κόμβο του X . Αν κάθε κόμβος του W γειτνιάζει με κάθε κόμβο του X τότε ο γράφος ονομάζεται Πλήρης Διμελής.

- Γραμμικός γράφος (Path graph)
Γράφος τάξης $n \geq 2$, του οποίου οι κόμβοι μπορούν να διαταχθούν σε μια σειρά v_1, v_2, \dots, v_n , σύμφωνα με την οποία οι ακμές είναι τα διατεταγμένα ζεύγη $\{v_i, v_{i+1}\}$, όπου $i = 1, 2, \dots, n - 1$.
- Κυκλικός γράφος (Cycle graph)
Γράφος τάξης $n \geq 3$, του οποίου οι κόμβοι μπορούν να διαταχθούν σε μια σειρά v_1, v_2, \dots, v_n , σύμφωνα με την οποία οι ακμές είναι τα διατεταγμένα ζεύγη $\{v_i, v_{i+1}\}$, όπου $i = 1, 2, \dots, n - 1$ μαζί με την ακμή $\{v_n, v_1\}$.

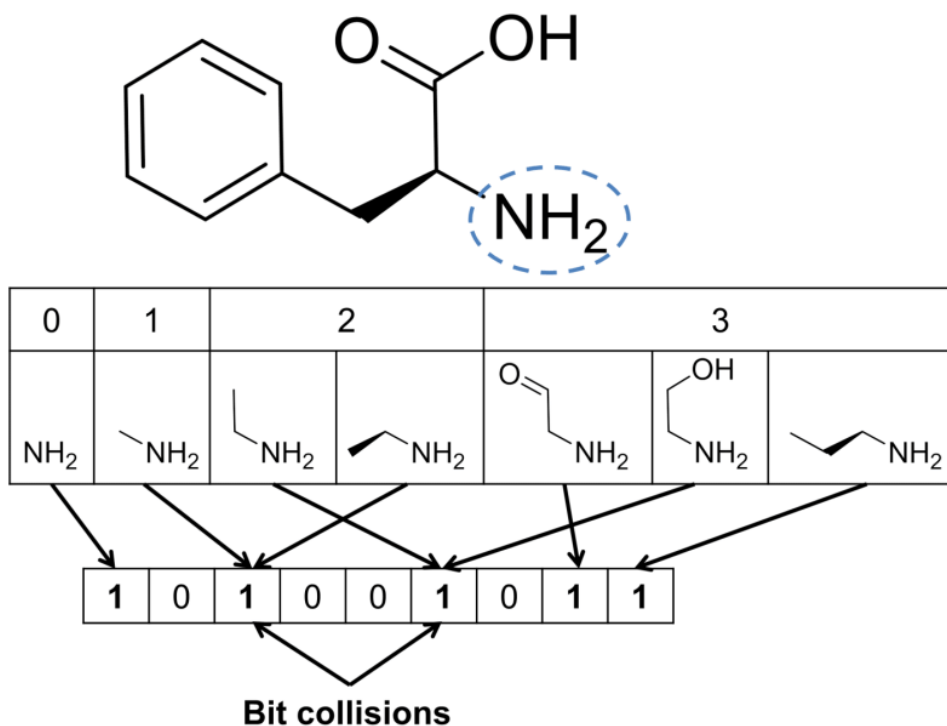
3.2 Από τα Μοριακά Αποτυπώματα στα Συνελικτικά Δίκτυα σε Γράφους (From Molecular Fingerprints to Graph Convolutions)

3.2.1 Μοριακά Αποτυπώματα (Molecular Fingerprints)

Τα μοριακά αποτυπώματα είναι μια μορφή κωδικοποίησης της δομής ενός χημικού μορίου σε διανύσματα δυαδικών ψηφίων (bits) με σκοπό την επίτευξη διαφόρων μορφών υπολογιστικής ανάλυσης. Αρχικά, βρήκαν ευρεία εφαρμογή στην αναζήτηση μέσω οργάνωσης σε βάση δεδομένων υπό-ομάδων των χημικών μορίων, όπως για παράδειγμα συγκεκριμένων αρωματικών δακτυλίων ή συνδυασμών αυτών. Πιο πρόσφατα, τα μοριακά αποτυπώματα χρησιμοποιήθηκαν σε εφαρμογές ταξινόμησης και συσταδοποίησης χημικών μορίων, καθώς και εύρεση ομοιοτήτων μεταξύ τους. Κατά κανόνα, προκύπτουν μέσα από την απαρίθμηση όλων των δυνατών τεμαχίων (fragments) του μορίου και κατόπιν, τη μετατροπή τους σε αριθμητικές τιμές μέσω μιας συνάρτησης κατατεμαχισμού (hash function). Οι αριθμητικές αυτές τιμές υποδεικνύουν τη θέση των bits στο κατατεμαχισμένο αποτύπωμα. Από την εφαρμογή της διαδικασίας κατατεμαχισμού παράγεται ένα καθορισμένου μήκους διάνυσμα bits από ένα χημικό μόριο αυθαίρετου μεγέθους και έτσι προκύπτουν "συγκρούσεις bits", όπου διαφορετικά τεμάχια του μορίου μετατρέπονται στην ίδια αριθμητική τιμή (και άρα τοποθετούνται στην ίδια θέση bit). Επομένως, δεν υπάρχει καμία απευθείας συσχέτιση μεταξύ των τεμαχίων και των bits του αποτυπώματος. [34]

Τα μοριακά αποτυπώματα κατηγοριοποιούνται σε αποτυπώματα μονοπατιού (path-based fingerprints) και σε κυκλικά αποτυπώματα (circular fingerprints).

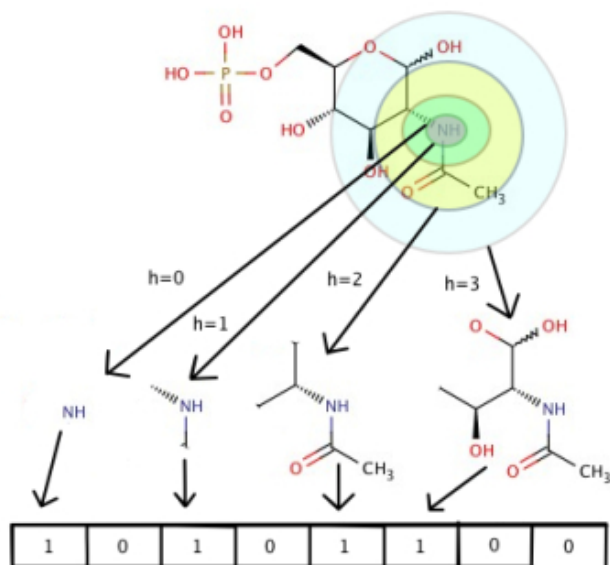
Τα αποτυπώματα μονοπατιού παράγονται ακολουθώντας μια, συνήθως, γραμμική διαδρομή κατά μήκος του μορίου έως έναν προκαθορισμένο αριθμό δεσμών. Τα ευρύτερα χρησιμοποιούμενα αποτυπώματα μονοπατιού είναι τα Daylight αποτυπώματα.



Σχήμα 3.4: Παράδειγμα αποτυπώματος μονοπατιού όπου παράγεται το διάνυσμα bits που αφορά το άτομο N του μορίου έως τον 3ο δεσμό μακριά από αυτό. Παρατηρούνται "συγκρούσεις" bits μεταξύ διαφορετικών τεμαχίων.

[35]

Τα κυκλικά αποτυπώματα παράγονται λαμβάνοντας υπόψιν τη "κυκλική" γειτονιά κάθε ατόμου έως μια συγκεκριμένη "διάμετρο" ή "ακτίνα" γειτνίασης. Σημαντικά κυκλικά αποτυπώματα, τα οποία χρησιμοποιούνται και για τους σκοπούς της παρούσας εργασίας, είναι τα αποτυπώματα ευρείας συνεκτικότητας (Extended-connectivity fingerprints - ECFPs) [36]. Ένα απλοποιημένο παράδειγμα κατασκευής ενός κυκλικού αποτυπώματος φαίνεται στην παρακάτω εικόνα:



Σχήμα 3.5: Στο παράδειγμα αυτό παράγεται το τμήμα του κυκλικού αποτυπώματος του μορίου της εικόνας που αφορά το άτομο N επιλέγοντας την "ακτίνα" γειτνίασης ίση με 3. [35]

Τα αποτυπώματα ευρείας συνεκτικότητας (εφεξής ECFPs) αποτελούν μια μεθοδολογία κυκλικών αποτυπωμάτων ειδικά ανεπτυγμένη για το σκοπό της αποτύπωσης χαρακτηριστικών των χημικών μορίων. Βρίσκουν εφαρμογή τόσο σε γενικές μεθόδους υπολογιστικής ανάλυσης, όπως εύρεση ομοιοτήτων, συσταδοποίηση και ταξινόμηση των μορίων, όσο και σε ειδικότερες εφαρμογές, όπως η πρόβλεψη της δράσης φαρμακευτικών ουσιών.

Στάδια παραγωγής ECFPs [36]

1. Αρχική εκχώρηση ταυτοποιητών ατόμων

Ένας ακέραιος αριθμός εκχωρείται σε κάθε άτομο του υπό διερεύνηση μορίου, ο οποίος έχει τον ρόλο του ταυτοποιητή (atom identifier). Ο ταυτοποιητής αποθηκεύει τοπικές πληροφορίες του προς ταυτοποίηση ατόμου, οι οποίες έχουν να κάνουν με ιδιότητες του (π.χ. ατομικός αριθμός, αριθμός άμεσων γειτόνων κ.λπ.) και μετατρέπονται σε έναν ακέραιο μέσω μιας συνάρτησης κατατεμαχισμού.

2. Επαναληπτική ενημέρωση των ταυτοποιητών

Κάθε άτομο αποθηκεύει σε μία λίστα τόσο τον τρέχοντα ταυτοποιητή του όσο και τους αντίστοιχους των γειτόνων του. Κατόπιν, η λίστα αυτή μετατρέπεται εκ νέου σε μοναδικό ακέραιο ταυτοποιητή μέσω μιας συνάρτησης κατατεμαχισμού. Σε κάθε επανάληψη αυξάνεται κατά μία μονάδα η ακτίνα της κυκλικής γειτονιάς από την οποία λαμβάνει πληροφορίες το κάθε άτομο του μορίου. Ο καθορισμός των κυκλικών γειτονιών γύρω από τυχαίο άτομο ενός παραδείγματος χημικού μορίου φαίνεται στην παρακάτω εικόνα:



Σχήμα 3.6: Κυκλικές γειτονίες που λαμβάνονται υπόψιν στην επαναληπτική διαδικασία του βήματος 2 για το μόριο της εικόνας. [35]

3. Αφαίρεση διπλότυπων

Το τελευταίο βήμα παραγωγής των ECFPs περιλαμβάνει την αφαίρεση πολλαπλών κοινών ακέραιων ταυτοποιητών (που έχουν προκύψει μετά την εφαρμογή της συνάρτησης κατατεμαχισμού).

Η αναπαράσταση των ECFPs μπορεί να γίνει είτε μέσω της λίστας ακέραιων ταυτοποιητών, όπως έχει ήδη περιγραφεί, είτε μέσω ενός προκαθορισμένου μήκους διανύσματος bits (κλασσικός τρόπος αναπαράστασης αποτυπωμάτων). Η "αναδίπλωση" ("folding") των ακέραιων ταυτοποιητών σε διανύσματα bits επιτυγχάνεται μέσω της εκ νέου εφαρμογής μιας συνάρτησης κατατεμαχισμού στους ακέραιους ταυτοποιητές. Η μετατροπή αυτή μειώνει αρκετά τον απαιτούμενο αποθηκευτικό χώρο, αυξάνοντας, όμως, την πιθανότητα "συγκρούσεων bits". Με αυτόν τον τρόπο, ένα μέρος της πληροφορίας χάνεται και η τελική αναπαράσταση του αποτυπώματος στερείται ποιότητας και ερμηνευσιμότητας σε σχέση με την αρχική αναπαράσταση των ακέραιων ταυτοποιητών.

3.2.2 Αποτυπώματα Duvenaud (Duvenaud Fingerprints)

Οι Duvenaud et al. το 2015 [37] εισήγαγαν ένα συνελικτικό νευρωνικό δίκτυο, το οποίο δέχεται ως είσοδο χημικούς γράφους αυθαίρετου μεγέθους και σχήματος. Η αρχιτεκτονική που χρησιμοποίησαν γενικεύει μεθόδους εξαγωγής χαρακτηριστικών μορίων που βασίζονται στα κυκλικά αποτυπώματα και πιο συγκεκριμένα, στα ECFPs. Η σημαντικότερη τροποποίηση της αρχικής διαδικασίας παραγωγής των ECFPs είναι η αντικατάσταση της συνάρτησης κατατεμαχισμού από ένα διαφορίσιμο νευρωνικό δίκτυο. Σκοπός του μοντέλου είναι η παραγωγή ενός νέου αποτυπώματος υπό τη μορφή διανύσματος πραγματικών αριθμών (και όχι bits όπως στα ECFPs), το οποίο στη συνέχεια θα εισαχθεί σε ένα MLP για εφαρμογές ταξινόμησης ή παλινδρόμησης.

Πιο συγκεκριμένα, το μοντέλο δέχεται ως είσοδο το μόριο υπό μορφή γράφου, του οποίου οι κόμβοι (δηλαδή τα άτομα του μορίου) χαρακτηρίζονται από ένα διάνυσμα χαρακτηριστικών \mathbf{r}_a (feature vector). Ακόμη, εισόδους στο μοντέλο αποτελούν η ακτίνα R , η οποία είναι ανάλογη της ακτίνας των κυκλικών αποτυπωμάτων, τα αρχικά τυχαία κρυφά βάρη \mathbf{H}_L^N και τα αρχικά τυχαία βάρη εξόδου \mathbf{W}_L , όπου L δηλώνει την απόσταση από τον κόμβο αναφοράς και N τον αριθμό των γειτόνων του κόμβου αναφοράς. Στην συνέχεια, κατ' αντιστοιχία

με τα ECFPs, για αυξανόμενη απόσταση L (με μέγιστη τιμή την ακτίνα R), εφαρμόζεται αντί για συνάρτηση κατατεμαχισμού νευρωνικό δίκτυο ενός στρώματος με είσοδο το άθροισμα των διανυσμάτων χαρακτηριστικών των κόμβων της γειτονιάς του κόμβου αναφοράς (η οποία ορίζεται από την απόσταση L). Η έξοδος του νευρωνικού δικτύου ανανεώνει το διάνυσμα χαρακτηριστικών του ίδιου του κόμβου αναφοράς. Επιτελείται, λοιπόν, μια διαδικασία "ανταλλαγής μηνυμάτων" μεταξύ του κόμβου αναφοράς και των γειτόνων του. Στη συνέχεια, εφαρμόζεται ένα ακόμη νευρωνικό δίκτυο με συνάρτηση ενεργοποίησης Softmax αυτή τη φορά στο ανανεωμένο διάνυσμα χαρακτηριστικών του κόμβου αναφοράς. Η έξοδος του νευρωνικού αυτού δικτύου προστίθεται στο διάνυσμα που εκφράζει το αποτύπωμα. Η παραπάνω διαδικασία επαναλαμβάνεται για κάθε δεδομένη απόσταση L και για όλα τα άτομα του υπό εξέταση μορίου, δηλαδή για όλους του κόμβους του υπό εξέταση γράφου. Αλγοριθμικά, η διαδικασία που ακολουθείται φαίνεται παρακάτω:

Αλγόριθμος κατασκευής αποτυπώματος Duvenaud

Είσοδος: $G = (A, B)$: Χημικός γράφος, όπου A το σύνολο κόμβων ή ατόμων και B το σύνολο των ακμών ή δεσμών

Είσοδος: R : Ακτίνα του αποτυπώματος

Είσοδος: H_L^N : Αρχικά τυχαία κρυφά βάρη

Είσοδος: W_L : Αρχικά τυχαία βάρη εξόδου

Απαίτηση: $f \leftarrow \mathbf{0}$: Αρχική τιμή διανύσματος αποτυπώματος

Για κάθε άτομο $a \in A$ στον χημικό γράφο **κάνε**:

$\mathbf{r}_a \leftarrow g(a)$, όπου g συνάρτηση χαρακτηρισμού του ατόμου a

Για $L \in \{1 \dots R\}$ **κάνε**:

Για κάθε άτομο $a \in A$ στον χημικό γράφο **κάνε**:

$\mathbf{v} \leftarrow \mathbf{r}_a + \sum_{i=1}^N \mathbf{r}_i$, όπου r_i τα γειτονικά άτομα του a

$\mathbf{r}_a \leftarrow \sigma(\mathbf{v}H_L^N)$, όπου σ συνάρτηση ενεργοποίησης

$\mathbf{i} \leftarrow \text{softmax}(\mathbf{r}_a W_L)$

$\mathbf{f} \leftarrow \mathbf{f} + \mathbf{i}$

Επίστρεψε: \mathbf{f}

Τέλος, τα υπολογισμένα μοριακά αποτυπώματα για τα διάφορα μόρια ενός συνόλου δεδομένων εισάγονται σε νευρωνικό δίκτυο ενός κρυφού στρώματος με σκοπό την εξαγωγή προβλέψεων για μια συγκεκριμένη χημική ιδιότητα (πρόβλημα παλινδρόμησης ή ταξινόμησης).

Τόσο τα αποτυπώματα Duvenaud, όσο και τα μοριακά αποτυπώματα αποτελούν συνελκτικές διαδικασίες, από την άποψη ότι εφαρμόζουν την ίδια διεργασία τοπικά παντού και στην συνέχεια, συσσωρεύουν πληροφορίες μέσω ενός βήματος υποδειγματοληψίας. Στην περίπτωση των αποτυπωμάτων Duvenaud η συνελκτική διεργασία αντιστοιχεί στην εφαρμογή ενός νευρωνικού δικτύου στα αθροισμένα χαρακτηριστικά μιας γειτονιάς μορίων. Οι έξοδοι του νευρωνικού δικτύου είναι παρόμοιες όταν η τοπική μοριακή δομή δεν διαφέρει σημαντικά, γεγονός που επιφέρει πλεονέκτημα έναντι των ECFPs στην ικανότητα εντοπισμού ενός σημαντικού τοπικού χαρακτηριστικού (υπενθυμίζεται ότι η αντιστοιχη συνάρτηση κατατεμαχισμού στα ECFPs παράγει διαφορετικό ταυτοποιητή για οποιαδήποτε διαφορά στην τοπική μοριακή δομή). Επιπλέον, η εφαρμογή του νευρωνικού δικτύου με την ενεργοποίηση Softmax στο ανανεωμένο, πλέον, διάνυσμα χαρακτηριστικών έχει τον ρόλο του βήματος υποδειγματοληψίας, αφού το διάνυσμα αποκτά καθορισμένο μήκος και μπορεί να προστεθεί στο διάνυσμα του αποτυπώματος. Η συνολική διαδικασία, λοιπόν, αποτελεί μια πρώιμη μορφή συνελκτικού νευρωνικού δικτύου σε γράφους, αφού λαμβάνει ως είσοδο γράφους, περιλαμβάνει παραμέτρους (H_L^N , W_L), οι οποίες εκπαιδεύονται μέσω

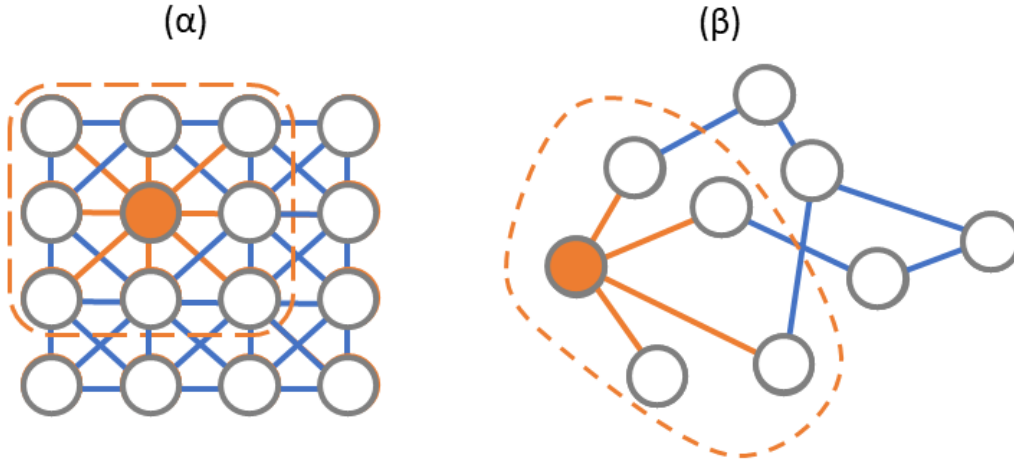
back-propagation και χρήση των γνωστών αλγορίθμων βελτιστοποίησης (π.χ. Adam) και οργανώνεται σε στρώματα, τα οποία καθορίζονται από το μήκος της ακτίνας R (δηλαδή 3 στρώματα για $R = 3$, 4 στρώματα για $R = 4$ κ.ο.κ.).

Συμπερασματικά, τα μοριακά αποτυπώματα του Duvenaud δείχνουν τον ρόλο που διαδραμάτισε η χημειοπληροφορική με τα κυκλικά μοριακά αποτυπώματα, καθώς και η "φυσική" αναπαράσταση του χημικού μορίου ως μαθηματικός γράφος στην εξέλιξη των νευρωνικών δικτύων γράφων. Οι βασικοί πυλώνες των νευρωνικών δικτύων γράφων, καθώς και η ιδέα πίσω από την διαδικασία "ανταλλαγής μηνυμάτων" μεταξύ των κόμβων θα παρουσιαστούν εκτενέστερα στο επόμενο υποκεφάλαιο.

3.3 Νευρωνική Ανταλλαγή Μηνυμάτων και Νευρωνικά Δίκτυα Γράφων (Neural Message Passing and Graph Neural Networks)

Όπως έχει περιγραφεί στο υποκεφάλαιο 3.1, ο γράφος αποτελεί έναν φυσικά πολυδύναμο τρόπο μοντελοποίησης δομών που έχουν τη μορφή δικτύου αντικειμένων. Η εξαγωγή μοτίβων από τέτοιες δομές με στόχο την πρόβλεψη ιδιοτήτων τους οδήγησε στην δημιουργία ενός νέου είδους νευρωνικών δικτύων, τα νευρωνικά δίκτυα γράφων (graph neural networks - GNNs). Οι προβλέψεις των GNNs μπορούν να γίνουν σε επίπεδο κόμβων (ταξινόμηση/παλινδρόμηση κόμβων - node classification/regression), σε επίπεδο γράφων (ταξινόμηση/παλινδρόμηση γράφων - graph classification/regression) ή και σε επίπεδο ακμών (πρόβλεψη συνδέσεων - link prediction) [38]. Στην συντριπτική πλειοψηφία τους, τα μοντέρνα GNNs ακολουθούν σε επίπεδο κόμβου μια στρατηγική συσσώρευσης πληροφοριών από τη γειτονιά. Πιο συγκεκριμένα, ο κάθε κόμβος του γράφου ανανεώνει την αναπαράστασή του συνδυάζοντας πληροφορίες από την αρχική του αναπαράσταση και από τις αναπαραστάσεις των γειτονικών του κόμβων, επιτελώντας, δηλαδή μια διαδικασία ανταλλαγής μηνυμάτων με τους γείτονές του. Με αυτόν τον τρόπο, ο κάθε κόμβος "μαθαίνει" από το περιβάλλον του συλλέγοντας δομικές πληροφορίες του γράφου.

Όπως φάνηκε στα αποτυπώματα Duvenaud, τα νευρωνικά δίκτυα γράφων στηρίζονται στην ιδέα της συνέλιξης, όπως εκφράστηκε στα συνελκτικά νευρωνικά δίκτυα. Τα συνελκτικά δίκτυα, όμως, βρίσκουν εφαρμογές σε δεδομένα που εκφράζονται στον ευκλείδειο χώρο, όπως εικόνες (δισδιάστατα πλέγματα) και κείμενο (μονοδιάστατη ακολουθία). Υπάρχει, όμως, μια πληθώρα δεδομένων που δεν μπορεί να αναπαρασταθεί σε ένα πλέγμα συγκεκριμένων διαστάσεων στον χώρο. Τα παραδείγματα είναι αναρίθμητα και μπορεί να αποτελούν ένα κοινωνικό δίκτυο, ένα δίκτυο τραπεζικών λογαριασμών, έναν χάρτη πλοήγησης ή μια λίστα φαρμακευτικών ουσιών μαζί με τις γνωστές θεραπευτικές λειτουργίες που επιτελούν. Όλες αυτές οι κυριολεκτικά άπειρες, λόγω της περίπλοκης ανθρώπινης φύσης, δομές δεδομένων μπορούν να προτυποποιηθούν ως γράφοι. Η γενίκευση της συνέλιξης, λοιπόν, σε μη-ευκλείδειες δομές και η εφαρμογή σε αυτούς τεχνικών βαθιάς μάθησης έχει συγκεντρώσει τεράστιο ερευνητικό ενδιαφέρον. [39] Μία οπτική σύγκριση της συνέλιξης στις εικόνες με την συνέλιξη στους γράφους μπορεί να επιτευχθεί εάν η εικόνα θεωρηθεί ως μια ειδική κατηγορία γράφου, όπου τα εικονοστοιχεία (κόμβοι) συνδέονται μέσω ακμών με γειτονικά εικονοστοιχεία:



Σχήμα 3.7: (α) Δισδιάστατη συνέλιξη σε εικόνα που έχει παρασταθεί σε μορφή γράφου με κάθε κόμβο να αντιπροσωπεύει ένα εικονοστοιχείο. Στην συγκεκριμένη περίπτωση η 3×3 συνέλιξη συμβαίνει με κεντρικό κόμβο αναφοράς τον χρωματισμένο με πορτοκαλί. Η γειτονιά ενός κόμβου είναι κάθε φορά προκαθορισμένου μεγέθους. (β) Μη-ευκλείδεια συνέλιξη σε γράφο. Η συνέλιξη αποσκοπεί στην ανανέωση πληροφοριών του χρωματισμένου πορτοκαλί κόμβου συναθροίζοντας πληροφορίες από την αρχική του αναπαράσταση και τις αναπαραστάσεις των γειτόνων του. Η γειτονιά ενός κόμβου σε αυτήν την περίπτωση δεν έχει προκαθορισμένο μέγεθος και σχήμα.

Οι Gilmer et al το 2017 κατόρθωσαν να αναδιατυπώσουν τα μέχρι τότε μοντέλα νευρωνικών δικτύων που δέχονταν ως είσοδο γράφους προτείνοντας το γενικό πλαίσιο των νευρωνικών δικτύων ανταλλαγής μηνυμάτων (message passing neural networks - MPNNs) [40]. Τα MPNNs, όπως και τα προγενέστερα αποτυπώματα Duvenaud, αναπτύχθηκαν για τους σκοπούς της πρόβλεψης χημικών μοριακών ιδιοτήτων, εκτελώντας με αυτόν τον τρόπο προβλέψεις σε επίπεδο γράφου. Το γενικό πλαίσιο των MPNNs βρίσκει εφαρμογή ακόμα και σε state-of-the-art μοντέλα που αφορούν ταξινόμηση ή παλινδρόμηση γράφων. Για την μαθηματική περιγραφή των MPNNs, αρχικά, ορίζονται ο γράφος $G(V, E)$ και τα διανύσματα \mathbf{x}_v και \mathbf{e}_{vw} , τα οποία αποτελούν τα διανύσματα χαρακτηριστικών (feature vectors) του κόμβου v και της ακμής που συνδέει τους κόμβους v και w αντίστοιχα. Η διαδικασία της εμπρόσθιας διάδοσης (forward propagation) περιλαμβάνει δύο στάδια, το στάδιο της ανταλλαγής μηνυμάτων (message passing phase) και το στάδιο της ανάγνωσης (readout phase). Το στάδιο ανταλλαγής μηνυμάτων τρέχει για T χρονικά βήματα και περιλαμβάνει την εφαρμογή μιας συνάρτησης μηνύματος (message function), M_t και μιας συνάρτησης ανανέωσης κόμβου, U_t . Οι κρυφές αναπαραστάσεις των κόμβων (hidden nodes representations), \mathbf{h}_v^t ανανεώνονται με βάση τα μηνύματα \mathbf{m}_v^{t+1} ως εξής:

$$\mathbf{m}_v^{t+1} = \sum_{w \in N(v)} M_t(\mathbf{h}_v^t, \mathbf{h}_w^t, \mathbf{e}_{vw}), \quad (3.10)$$

$$\mathbf{h}_v^{t+1} = U_t(\mathbf{h}_v^t, \mathbf{m}_v^{t+1}), \quad (3.11)$$

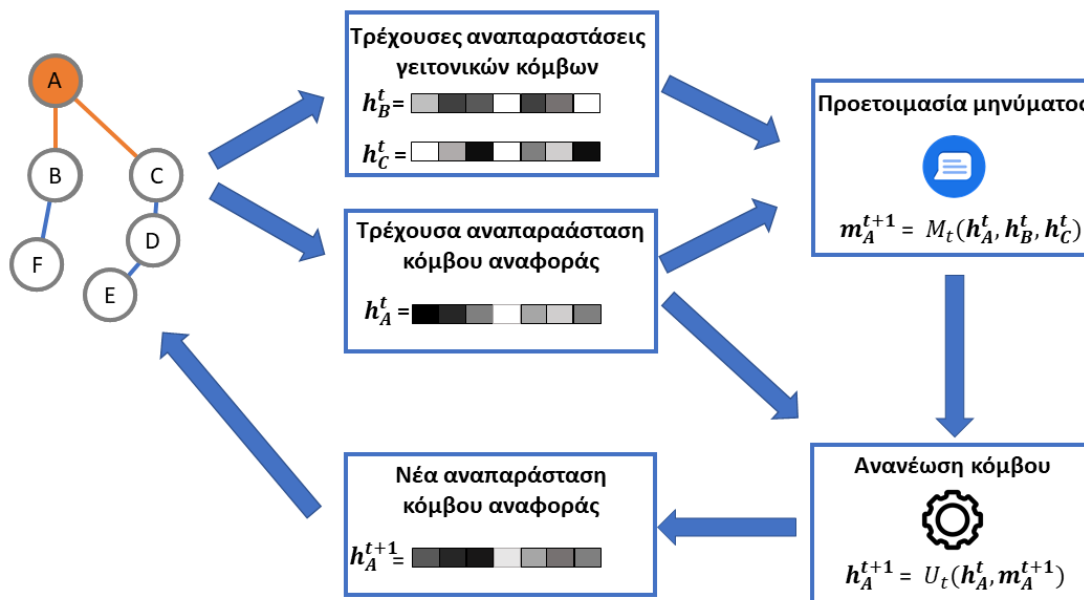
όπου $N(v)$ είναι το σύνολο των γειτονικών κόμβων του κόμβου v . Παραδείγματα συναρτήσεων μηνύματος μπορεί να είναι ο μέσος όρος (average) ή το μέγιστο (maximum), ενώ παράδειγμα συνάρτησης ανανέωσης κόμβου μπορεί να είναι ένα νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης ενός στρώματος.

Το στάδιο της ανάγνωσης περιλαμβάνει μια συνάρτηση που δεν εξαρτάται από μετάθεση του συνόλου των κόμβων, η οποία συνδυάζει τις τελικά υπολογισμένες κρυφές αναπαραστάσεις όλων των κόμβων του γράφου, \mathbf{h}_v^T , με σκοπό την παραγωγή ενός διανύσματος που χαρακτηρίζει ολόκληρο τον γράφο:

$$\hat{\mathbf{y}} = R(\{\mathbf{h}_v^T | v \in V\}) . \quad (3.12)$$

Σημειώνεται ότι εάν σκοπός του μοντέλου είναι η πρόβλεψη σε επίπεδο κόμβου, τότε το στάδιο της ανάγνωσης παραλείπεται.

Όλες οι παραπάνω συναρτήσεις (M_t, U_t, R) είναι διαφορίσιμες και άρα, μπορούν να χρησιμοποιηθούν για την εκπαίδευση ενός παραμετρικού νευρωνικού δικτύου μέσω back-propagation και αλγορίθμων βελτιστοποίησης (π.χ. Gradient Descent). Μια σχηματική αναπαράσταση της εφαρμογής MPNN που έχει ως είσοδο έναν απλό μη-κατευθυνόμενο γράφο είναι η ακόλουθη:



Σχήμα 3.8: Στρώμα νευρωνικού δικτύου ανταλλαγής μηνυμάτων με είσοδο απλό μη-κατευθυνόμενο γράφο. Ο κόμβος A ανανεώνεται με βάση τα μηνύματα των γειτόνων (B, C) και την προτερή του αναπαράσταση. Τα χαρακτηριστικά των ακμών έχουν αγνοηθεί για λόγους οικονομίας χώρου του σχήματος.

3.4 Μοντέλα Νευρωνικών Δικτύων Γράφων (Graph Neural Networks Models)

3.4.1 Συνελικτικό Δίκτυο Γράφων (Graph Convolutional Network)

Το συνελικτικό δίκτυο γράφων ή GCN είναι από τα ευρύτερα χρησιμοποιούμενα νευρωνικά δίκτυα γράφων και προτάθηκε από τους Kipf και Welling το 2017 [41]. Σύμφωνα με το γενικό πλαίσιο των MPNNs, εφαρμόζεται η συνάρτηση μηνύματος μέσω πολλαπλασιασμού πινάκων (η πλήρης διαδικασία θα αναλυθεί στη συνέχεια) και η συνάρτηση ανανέωσης κόμβου μέσω ενός MLP ενός στρώματος. Σημειώνεται ότι όσο περισσότερα στρώματα GCN χρησιμοποιούνται, τόσο πιο μακριά μπορεί να ταξιδέψει η πληροφορία ενός κόμβου. Για

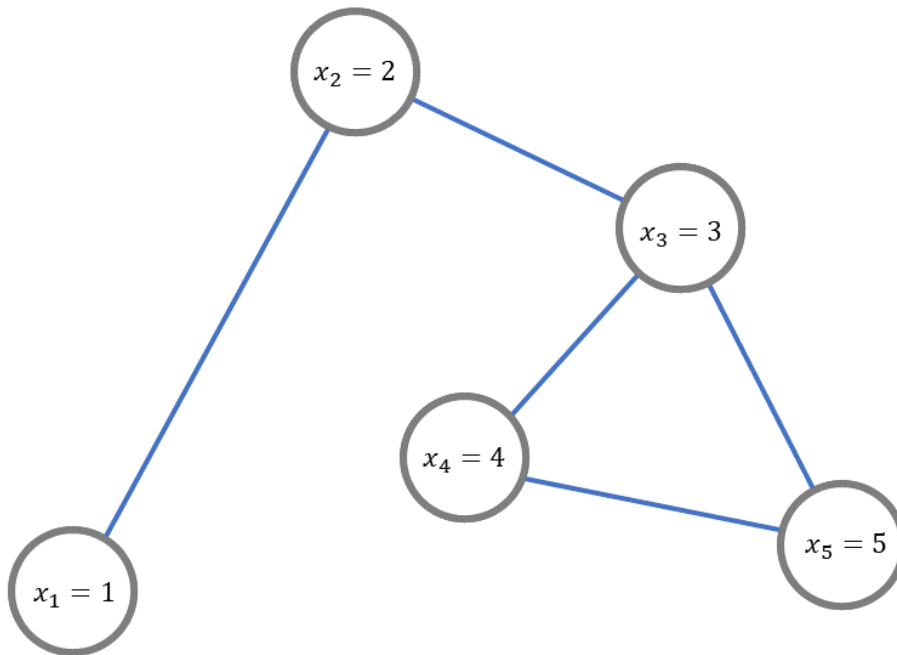
παράδειγμα, για ένα GCN ενός στρώματος ένας κόμβος ανανεώνεται με βάση την πληροφορία των άμεσα γειτονικών του κόμβων. Για δύο στρώματα GCN, όμως, θα ανανεωθεί μεν λαμβάνοντας πληροφορίες από τους άμεσα γειτονικούς του κόμβους, αλλά αυτοί στο προηγούμενο στρώμα θα έχουν, επίσης, λάβει πληροφορίες από τους γείτονές τους. Έτσι, για δύο στρώματα κάθε κόμβος ανταλλάσσει μηνύματα με τους γείτονες των γειτόνων του ή με την 2-hop γειτονιά του, για τρία στρώματα με την 3-hop γειτονιά του κ.ο.κ..

Για την μαθηματική προτυποποίηση του GCN, γίνεται χρήση του πίνακα γειτνίασης του γράφου ($\mathbf{A} \in \mathbb{R}^{N \times N}$) και του πίνακα βαθμών του γράφου ($\mathbf{D} \in \mathbb{R}^{N \times N}$) (όπου N ο αριθμός των κόμβων του γράφου) που ορίστηκαν στο υποκεφάλαιο 3.1. Ο κανόνας ανανέωσης του πίνακα κρυφών αναπαραστάσεων του γράφου, $\mathbf{H} \in \mathbb{R}^{N \times D}$, του οποίου οι γραμμές αντιστοιχούν στα διανύσματα κρυφών αναπαραστάσεων μήκους D των κόμβων είναι ο ακόλουθος:

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^l \mathbf{W}^l), \quad (3.13)$$

όπου l δηλώνει το αντίστοιχο στρώμα του GCN, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$, ο τροποποιημένος πίνακας γειτνίασης, ο οποίος προκύπτει εάν σε κάθε κόμβο θεωρηθεί ότι υπάρχει μια ακμή που τον συνδέει με τον εαυτό του (βρόχος - self-connection), $\tilde{\mathbf{D}}$, ο τροποποιημένος πίνακας βαθμών που προκύπτει κατ' αντιστοιχία με τον πίνακα βαθμών (\mathbf{D}) ως εξής: $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, σ μια συνάρτηση ενεργοποίησης, όπως π.χ. η relu .

Η εξίσωση 3.13 προκύπτει θεωρητικά μέσα από την εφαρμογή της διαδικασίας της συνέλιξης στην μη-ευκλείδεια δομή ενός γράφου. Η απόδειξη θα δειχθεί στη συνέχεια. Αρχικά, όμως, θα παρουσιαστεί το φυσικό νόημα της παραπάνω εξίσωσης μέσω αριθμητικών παραδειγμάτων που περιλαμβάνουν πολλαπλασιασμό πινάκων. Έστω ο γράφος του παρακάτω σχήματος:



Σχήμα 3.9: Απλός μη-κατευθυνόμενος γράφος. Τα χαρακτηριστικά των κόμβων είναι βαθμωτά μεγέθη.

Οι πίνακες γειτνίασης και βαθμών του παραπάνω γράφου θα ισούνται αντίστοιχα με :

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

Έστω ότι, προς χάριν ευκολίας, το χαρακτηριστικό κάθε κόμβου δεν δίνεται από διάνυσμα αλλά από βαθμωτό μέγεθος, δηλαδή $D = 1$. Επίσης, έστω ότι το διάνυσμα (πλέον) χαρακτηριστικών των κόμβων του γράφου παίρνει τις εξής τιμές :

$$\mathbf{X} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$$

Η πράξη πολλαπλασιασμού του πίνακα γειτνίασης \mathbf{A} με το διάνυσμα χαρακτηριστικών \mathbf{X} δίνει το εξής αποτέλεσμα :

$$\mathbf{AX} = \begin{bmatrix} 2 \\ 4 \\ 11 \\ 8 \\ 7 \end{bmatrix}$$

δηλαδή, ο παραπάνω πολλαπλασιασμός έχει ως αποτέλεσμα το άθροισμα χαρακτηριστικών των γειτονικών κόμβων του επιμέρους κόμβου αναφοράς. Επομένως, ο πολλαπλασιασμός πίνακα γειτνίασης με πίνακα χαρακτηριστικών επιφέρει τη συνάρτηση μηνύματος αθροίσματος με βάση όσα αναφέρθηκαν στο υποκεφάλαιο 3.3. Κατ' αντίστοιχο τρόπο, για συνάρτηση μηνύματος μέσου όρου θα πρέπει τα επιμέρους αθροίσματα να αναχθούν ως προς το μέγεθος της επιμέρους γειτονιάς, κάτι που μπορεί να επιτευχθεί με χρήση του πίνακα βαθμών D ως εξής :

$$\mathbf{D}^{-1}\mathbf{AX} = \begin{bmatrix} 2 \\ 2 \\ 3.63 \\ 4 \\ 3.5 \end{bmatrix}$$

Μέσω της τροποποίησης του πίνακα γειτνίασης και του πίνακα βαθμών με την προσθήκη self connections, η συνάρτηση ανανέωσης κόμβου σύμφωνα με τα περιεχόμενα του υποκεφαλαίου 3.3 ενσωματώνεται στην συνάρτηση μηνύματος, αφού πράγματι, η εξίσωση 3.13 υπολογίζει μια νέα κρυφή αναπαράσταση κόμβου. Η προσθήκη των self connections έχει και άλλη μια χρησιμότητα που αφορά την βελτιστοποίηση του νευρωνικού δικτύου και θα αναφερθεί παρακάτω στην απόδειξη της συνελκτικής διαδικασίας σε γράφους. Έστω ο πίνακας $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}$ (πρώτοι 3 όροι πολλαπλασιασμού στην εξίσωση 3.13). Μια άλλη μαθηματική απεικόνιση για τον υπολογισμό των στοιχείων του πίνακα $\hat{\mathbf{A}}$ είναι :

$$\hat{\mathbf{A}}_{ij} = \frac{1}{\sqrt{\tilde{a}_i \tilde{a}_j}} \tilde{\mathbf{A}}_{ij}, \quad (3.14)$$

όπου γίνεται αναγωγή του πίνακα γειτνίασης με βάση το μέγεθος της γειτονιάς τόσο του κόμβου-πηγής του μηνύματος όσο και του κόμβου-προορισμού του μηνύματος. Παρατηρείται, λοιπόν, ότι ο κανόνας ανανέωσης του πίνακα κρυφών αναπαραστάσεων των κόμβων με βάση το μοντέλο GCN έχει, διαισθητικά, φυσική σημασία και συμβαδίζει με το γενικό πλαίσιο των MPNNs. Το θεωρητικό υπόβαθρο του συνελκτικού στρώματος στους γράφους παρουσιάζεται συνοπτικά παρακάτω:

Θεωρητικό Υπόβαθρο GCN

Έστω ένας μη-κατευθυνόμενος γράφος $G(V, E)$ με πίνακα γειτνίασης $A \in \mathbb{R}^{N \times N}$. Έστω, επίσης, ένα σήμα γράφου $x : V \rightarrow \mathbb{R}$. Το σήμα γράφου αντιστοιχεί κάθε κόμβο του γράφου (σύνολο V) σε ένα πραγματικό αριθμό και μπορεί να εκφραστεί ως διάνυσμα του \mathbb{R}^N , όπου x_i η τιμή του \mathbf{x} στον i -οστό κόμβο. Ως παράδειγμα δίνεται το κοινωνικό δίκτυο Twitter, όπου οι χρήστες αποτελούν τους κόμβους, οι σχέσεις μεταξύ των χρηστών τις ακμές και οι δημοσιεύσεις που περιέχουν το hashtag #covid19 τις τελευταίες 6 εβδομάδες ένα τυχαίο σήμα. Ακόμη, ορίζεται ο κανονικοποιημένος λαπλασιανός πίνακας $\hat{\mathbf{L}}$ (βλ. υποκεφάλαιο 3.1), ο οποίος ως ένας πραγματικός συμμετρικός θετικά ημιορισμένος πίνακας, περιέχει ένα σύνολο ορθοκανονικών ιδιοδιανυσμάτων $\{\mathbf{u}_i\}_{i=0}^{N-1}$ και τις συσχετιζόμενες με αυτές πραγματικές μη-αρνητικές ιδιοτιμές $\{\lambda_i\}_{i=0}^{N-1}$. Πράγματι, ο κανονικοποιημένος λαπλασιανός πίνακας διαγωνιοποιείται ως εξής:

$$\hat{\mathbf{L}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T, \quad (3.15)$$

όπου $\mathbf{\Lambda}$ ο διαγώνιος πίνακας των ιδιοτιμών ($\mathbf{\Lambda} = \text{diag}(\lambda_0, \dots, \lambda_{n-1})$) και \mathbf{U} ο πίνακας των ιδιοδιανυσμάτων ($\mathbf{U} = [\mathbf{u}_0, \dots, \mathbf{u}_{n-1}]$).

Ο μετασχηματισμός Fourier γράφου (graph Fourier transform) του σήματος $\mathbf{x} \in \mathbb{R}^N$ ορίζεται ως εξής [42]:

$$\hat{\mathbf{x}} = \mathbf{U}^T \mathbf{x}. \quad (3.16)$$

Λόγω της φυσικής αδυναμίας έκφρασης της διαδικασίας της συνέλιξης στο πεδίο των κόμβων, ορίζεται η φασματική συνέλιξη σε γράφο (spectral graph convolution) στο πεδίο συχνοτήτων Fourier για δύο σήματα $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ ως εξής [42]:

$$\mathbf{x} *_G \mathbf{y} = \mathbf{U}(\hat{\mathbf{x}} \odot \hat{\mathbf{y}}) = \mathbf{U}((\mathbf{U}^T \mathbf{x}) \odot (\mathbf{U}^T \mathbf{y})), \quad (3.17)$$

όπου $*_G$ το σύμβολο της φασματικής συνέλιξης σε γράφο και \odot το γινόμενο Hadamard.

Η φασματική συνέλιξη γράφου ενός φίλτρου \mathbf{g}_θ που εξαρτάται από τις ιδιοτιμές του κανονικοποιημένου λαπλασιανού πίνακα ή αλλιώς, το φιλτράρισμα του σήματος \mathbf{x} από το \mathbf{g}_θ δίνεται από την ακόλουθη εξίσωση [41] [42]:

$$\mathbf{g}_\theta *_G \mathbf{x} = \mathbf{U} \mathbf{g}_\theta(\mathbf{\Lambda}) \mathbf{U}^T \mathbf{x}. \quad (3.18)$$

Ο υπολογισμός της εξίσωσης 3.18 είναι υπολογιστικά δαπανηρός, αφού ο πολλαπλασιασμός με τον πίνακα ιδιοδιανυσμάτων \mathbf{U} έχει πολυπλοκότητα $O(N^2)$. Για την λύση του προβλήματος αυτού, προτάθηκε από τους Hammond et al. [43] η προσέγγιση του φίλτρου $\mathbf{g}_\theta(\mathbf{\Lambda})$ με χρήση πολυωνύμων Chebysen $T_K(x)$ έως τάξης K :

$$\mathbf{g}_\theta(\mathbf{\Lambda}) \approx \sum_{k=0}^K \partial'_k T_k(\tilde{\mathbf{\Lambda}}), \quad (3.19)$$

όπου $\tilde{\mathbf{\Lambda}} = \frac{2}{\hat{\lambda}_{max}} \mathbf{\Lambda} - \mathbf{I}_N$, $\hat{\lambda}_{max}$ η μεγαλύτερη ιδιοτιμή του L .

Εφόσον ισχύει $(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T)^k = \mathbf{U}\mathbf{\Lambda}^k\mathbf{U}^T$, η εξίσωση 3.18 παίρνει την μορφή:

$$\mathbf{g}_\theta *_{\mathbf{G}} \mathbf{x} \approx \sum_{k=0}^K \partial'_k T_k(\tilde{\mathbf{L}}) \mathbf{x}. \quad (3.20)$$

Τα πολυώνυμα Chebysen ορίζονται ως $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$, $T_0(x) = 1$, $T_1(x) = x$. Για το μοντέλο GCN, επιλέχθηκε από τους συγγραφείς του τιμή του K ίση με 1, δηλαδή γραμμική συνάρτηση ως προς τον λαπλασιανό πίνακα. Με αυτόν τον τρόπο, γίνεται δυνατή η κατασκευή βαθύτερων δικτύων χωρίς τεράστια υπολογιστική επιβάρυνση. Ταυτόχρονα, στοιβάζοντας περισσότερα από ένα συνελκτικά στρώματα (και δημιουργώντας ένα βαθύτερο δίκτυο) παραμένει η δυνατότητα καλής προσέγγισης των φίλτρων χωρίς την ρητή παραμετροποίηση ενός στρώματος με προσέγγιση Chebysen υψηλής τάξης. Ακόμη, γίνεται προσέγγιση της τιμής $\hat{\lambda}_{max} \approx 2$. Με βάση τα παραπάνω η εξίσωση 3.20 απλοποιείται ως εξής:

$$\mathbf{g}_\theta *_{\mathbf{G}} \mathbf{x} \approx \partial'_0 \mathbf{x} + \partial'_1 (\mathbf{L} - \mathbf{I}_N) \mathbf{x} = \partial'_0 \mathbf{x} + \partial'_1 \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \mathbf{x}, \quad (3.21)$$

όπου ∂'_0, ∂'_1 ελεύθεροι παράμετροι. Για μείωση των παραμέτρων και κατά συνέπεια, πρακτική λύση στο πρόβλημα της υπερπροσαρμογής, όπως και ελαχιστοποίηση των πράξεων κατά την εφαρμογή ενός στρώματος, επιλέγεται $\partial = \partial'_0 = -\partial'_1$. Τέλος, για μεγαλύτερη αριθμητική σταθερότητα και αποφυγή του προβλήματος των βαθμίδων που εξαφανίζονται, προστίθενται self-connections. Με αυτόν τον τρόπο, προκύπτει, τελικά, η εξίσωση 3.13 του στρώματος GCN.

Συμπερασματικά, μέσω του μοντέλου GCN, επιτεύχθηκε η δημιουργία ενός στιβαρού μοντέλου νευρωνικών δικτύων με ικανοποιητικά αποτελέσματα σε προβλήματα πρόβλεψης τόσο σε επίπεδο κόμβων όσο και σε επίπεδο γράφων. Η απλότητα του συνελκτικού στρώματος προσδίδει αρκετά καλά ποιοτικά χαρακτηριστικά στο δίκτυο, όπως αποφυγή υπερπροσαρμογής, χαμηλό υπολογιστικό κόστος και δυνατότητα προσθήκης βάθους. Υστερεί, όμως, σε άλλους τομείς, όπως η αδυναμία πρόβλεψης σε επίπεδο ακμών αφού δεν δέχεται ως είσοδο διάνυσμα χαρακτηριστικών τους. Μάλιστα, το παραπάνω μπορεί σε αρκετές περιπτώσεις να αποτελεί και αδυναμία ως προς την μοντελοποίηση των δεδομένων εισόδου. Αυτό είναι και ένα από τα μειονεκτήματα των μοντέλων νευρωνικών δικτύων γράφων που βασίζονται στη φασματική συνέλιξη σε γράφους [39]. Είναι δεδομένο πως τα μοντέλα που επιχειρούν χωρική συνέλιξη σε γράφους, υπό την έννοια της ανταλλαγής μηνυμάτων στην γειτονιά ενός κόμβου, έχουν περισσότερες δυνατότητες. Το GCN, όμως, κατάφερε να γεφυρώσει το χάσμα μεταξύ φασματικής και χωρικής συνέλιξης, αφού όπως δείχθηκε στην αρχή του υποκεφαλαίου αυτού, πληρεί τις προϋποθέσεις των message passing neural networks. [38]

3.4.2 Υποδειγματοληψία Γράφων (Graph Pooling)

Η υποδειγματοληψία στα μοντέλα νευρωνικών δικτύων γράφων αποσκοπεί (α) στην μείωση των παραμέτρων του συστήματος μέσω της παραγωγής νέων μικρότερων αναπαραστάσεων κόμβων και άρα, αποφυγή προβλημάτων υπερπροσαρμογής και υπολογιστικής πολυπλοκότητας και (β) στην παραγωγή αναπαραστάσεων σε επίπεδο γράφου, οι οποίες βασίζονται στις αναπαραστάσεις των κόμβων (διαδικασία ανάγνωσης των κόμβων). Η πιο διαδεδομένη μέθοδος υποδειγματοληψίας είναι η υποδειγματοληψία αθροίσματος/μεγίστου/μέσου όρου, γνωστή και ως καθολική υποδειγματοληψία (global pooling), όπου μία από τις 3 παραπάνω συναρτήσεις εφαρμόζεται στις τελικές αναπαραστάσεις των κόμβων παράγοντας μια καθολική αναπαράσταση του γράφου ως εξής [38]:

$$\mathbf{h}_G = \text{sum/max/mean}(\mathbf{h}_1^K, \mathbf{h}_2^K, \dots, \mathbf{h}_N^K), \quad (3.22)$$

όπου K το τελευταίο στρώμα GNN του δικτύου, N ο αριθμός των κόμβων και \mathbf{h}_i^K τα τελικά διανύσματα χαρακτηριστικών των κόμβων του γράφου.

Πιο πρόσφατα, προτάθηκαν νέες μέθοδοι υποδειγματοληψίας που ελαττώνουν το μέγεθος του γράφου λαμβάνοντας υπόψιν τόσο χαρακτηριστικά των κόμβων όσο και δομικά χαρακτηριστικά του ίδιου του γράφου, οι οποίες συναντώνται ως ιεραρχικές μέθοδοι υποδειγματοληψίας (hierarchical pooling modules). [39] Μία από αυτές είναι η μέθοδος Diffpool, η οποία ομαδοποιεί τους κόμβους ενός γράφου όχι με καθολικό τρόπο αλλά μαθαίνοντας έναν πίνακα εκχώρησης ομάδων (cluster assignment matrix). [44] Ο υπολογισμός του πίνακα εκχώρησης ομάδων, S γίνεται μέσω ενός στρώματος GNN (π.χ. το στρώμα GCN) ως εξής:

$$\mathbf{S}^K = \text{softmax}(\text{GNN}(\mathbf{A}^K, \mathbf{T}^K))\mathbf{A}^{K+1} = (\mathbf{S}^K)^T \mathbf{A}^K \mathbf{S}^K, \quad (3.23)$$

όπου A και H οι πίνακες γειννίασης και χαρακτηριστικών του γράφου αντίστοιχα και K δηλώνει το στρώμα του νευρωνικού δικτύου.

Σημειώνεται ότι η παραπάνω διαδικασία είναι διαφορίσιμη και άρα, ο πίνακας S μπορεί να ανανεωθεί μέσω Back-propagation και ενός αλγορίθμου βελτιστοποίησης.

3.4.3 Πλήρως Συνδεδεμένα Στρώματα (Fully-Connected layers)

Πλήρως συνδεδεμένα στρώματα, δηλαδή στρώματα MLP χρησιμοποιούνται συχνά σε αρχιτεκτονικές νευρωνικών δικτύων γράφων τόσο για προεπεξεργασία των χαρακτηριστικών των κόμβων, δηλαδή πριν την εισαγωγή των γράφων σε στρώματα GNN, όσο για την τελική παραγωγή προβλέψεων του μοντέλου. Πιο συγκεκριμένα, όσον αφορά τις τελικές προβλέψεις, οι αναπαραστάσεις σε επίπεδο γράφου, οι οποίες παράγονται μετά την εφαρμογή μιας συνάρτησης ανάγνωσης (συνήθως συναρτήσεις global pooling), εισάγονται σε ένα MLP με σκοπό την τελική παραγωγή προβλέψεων (ταξινόμηση/παλινδρόμηση γράφων) [38] [45].

Κεφάλαιο 4

Ποσοτικές Σχέσεις Δομής - Δράσης (Quantitative Structure Activity Relationships - QSAR)

Οι Ποσοτικές Σχέσεις Δομής-Δράσης (QSARs) έχουν ως στόχο την παραγωγή μαθηματικών μοντέλων πρόβλεψης φυσικών, χημικών ή βιολογικών ιδιοτήτων χημικών ουσιών, βασισμένες στην υπόθεση ότι οι ιδιότητες αυτές εξαρτώνται από την στεreoχημική δομή του μορίου. Η προτυποποίηση της σχέσης δομής-δράσης επιτυγχάνεται, παραδοσιακά, με την χρήση αριθμητικών περιγραφικών μεταβλητών (numerical descriptors), οι οποίες μπορεί να είναι (α) εμπειρικές, όπως μια φυσικοχημική ιδιότητα του μορίου και (β) μη-εμπειρικές, οι οποίες τυπικά περιλαμβάνουν δομικές πληροφορίες του μορίου. Στο κεφάλαιο αυτό θα γίνει μια σύντομη ιστορική αναδρομή των QSAR μοντέλων και θα παρουσιαστεί η συμβολή των νευρωνικών δικτύων στην μοντελοποίηση ποσοτικών σχέσεων δομής-δράσης με ή χωρίς αριθμητικές περιγραφικές μεταβλητές.

4.1 Η Εξέλιξη των QSAR Μοντέλων (Evolution of QSAR Models)

Ο πρώτος ισχυρισμός περί σύνδεσης βιολογικής δράσης με χημική δομή μιας ουσίας προήλθε από τους Crum-Brown και Fraser το 1868 [46]. Μερικές δεκαετίες αργότερα, το 1898 και το 1901, οι Meyer και Overton ήταν οι πρώτοι που εξήγαγαν, ανεξάρτητα ο ένας από τον άλλο, μαθηματικές σχέσεις που συνέδεαν την αναισθητική δράση οργανικών ουσιών με την λιποφιλία τους. [47], [48] Στη συνέχεια, το 1937 ο Hammett εισήγαγε τις ηλεκτρονιακές σταθερές σ , μέσω των οποίων εξέφρασε την επίδραση που ασκούν οι υποκαταστάτες οργανικών μορίων στον ρυθμό μιας χημικής αντίδρασης.[4] Η προσέγγιση αυτή έδωσε το έναυσμα στους Hansch και Fujita να αναπτύξουν το 1964 το πρώτο QSAR μοντέλο, όπου η βιολογική δραστηριότητα μιας ομάδας χημικών ενώσεων παρόμοιας χημικής δομής μπορεί να περιγραφεί ως ακολούθως (γενική μορφή της εξίσωσης Hansch-Fujita):

$$\log\left(\frac{1}{C}\right) = a\sigma + b\log P + c, \quad (4.1)$$

όπου C η συγκέντρωση που έχει συγκεκριμένη βιολογική απόκριση σε συγκεκριμένο χρόνο, σ η ηλεκτρονιακή σταθερά Hammett και $\log P$ ο συντελεστής μερισμού της ουσίας που αποτελεί μέτρο της λιποφιλίας.

Οι συντελεστές a , b και η σταθερά c υπολογίζονται μέσω πολλαπλής γραμμικής ανάλυσης παλινδρόμησης. [50]

Την δεκαετία του 1980, με τα μοντέλα τύπου Hansch-Fujita πλέον εδραιωμένα, ο Klopman ανέπτυξε μια νέα προσέγγιση απεικόνισης των δομικών πληροφοριών ενός μορίου. Κατόρθωσε, λοιπόν, να αποσυνθέσει χημικά μόρια σε διδιάστατα τεμάχια και να συσχετίσει την συχνότητα εμφάνισης καθενός από αυτά με την βιολογική δραστηριότητα των μορίων. Με αυτόν τον τρόπο, έγινε ένα σημαντικό βήμα προς την κατεύθυνση ανάπτυξης καθολικών QSAR μοντέλων μέσω μιας υπολογιστικά αποδοτικής μεθόδου εξαγωγής και συσχέτισης εύκολα ερμηνεύσιμων δομικών χαρακτηριστικών για μεγάλο αριθμό χημικών μορίων. Ταυτόχρονα, καθώς τα σύνολα δεδομένων γίνονταν μεγαλύτερα και περισσότερο ποικιλόμορφα ως προς την μοριακή δομή, εισήχθησαν μέθοδοι χαρακτηρισμού που κατάφεραν, πατώντας στην φιλοσοφία των τεμαχίων Klopman να συλλαμβάνουν πιο γενικές σχέσεις δομής-δραστηριότητας βασισμένες πάντα στην αρχή ότι ολόκληρη η χημική δομή ενός μορίου καθορίζει τις ιδιότητές του. Ένα παράδειγμα τέτοιων μεθόδων είναι τα τεμάχια CASE (CASE fragments), ενώ σύγχρονες μέθοδοι που και αυτές βασίζονται στην αναζήτηση υπο-δομών του μορίου είναι τα μοριακά κυκλικά αποτυπώματα (βλ. υποκεφάλαιο 3.2.1). [51]

Σε γενικές γραμμές, για την ανάπτυξη ενός αξιόπιστου κλασικού QSAR μοντέλου τα βήματα που ακολουθούνται είναι τα εξής: [51]

1. Να επιλεγεί ένα σύνολο δεδομένων με χημικές ουσίες, οι οποίες παρουσιάζουν δομική συγγένεια και οι οποίες να έχουν καλά καταμερισμένες τιμές βιολογικής δράσης (ετικέτες του συνόλου δεδομένων).
2. Να προσδιοριστούν μέσω υπολογιστικών εργαλείων ή πειραμάτων οι κατάλληλες περιγραφικές μεταβλητές μοριακής δομής.
3. Να διαχωριστεί το σύνολο δεδομένων σε ομάδα εκπαίδευσης και ομάδα επαλήθευσης. Οι δύο ομάδες πρέπει να παρουσιάζουν ποικιλομορφία ετικετών και να είναι αντιπροσωπευτικές του αρχικού συνόλου, προκειμένου το μοντέλο να παρουσιάσει την καλύτερη δυνατή γενίκευση.
4. Να επιλεγεί η κατάλληλη μέθοδος μηχανικής μάθησης, η οποία θα παρουσιάζει την καλύτερη δυνατή επίδοση εκπαίδευσης και επαλήθευσης. Η ικανότητα πρόβλεψης να εκτιμάται με βάση στατιστικές μεθόδους αξιολόγησης (metrics).
5. Να διερευνηθεί και να καθοριστεί το πεδίο εφαρμογής του μοντέλου με χρήση δεδομένων εισόδου εκτός του αρχικού συνόλου δεδομένων.



Σχήμα 4.1: Βήματα ανάπτυξης κλασικού QSAR μοντέλου

4.2 QSAR Μοντέλα Βαθιάς Μηχανικής Μάθησης (Deep Learning QSAR Models)

Μια μοντελοποίηση QSAR με την κλασσική έννοια του όρου μπορεί να συνοψιστεί ως εξής:

$$\text{Βιολογική δραστηριότητα} = f\left(\sum (\text{Δομικές Ιδιότητες})\right) \quad (4.2)$$

Τα πρωταρχικά QSAR μοντέλα στηρίζονταν σε γραμμικές σχέσεις μεταξύ χημικής δομής και βιολογικής δραστηριότητας. Παρ' όλα αυτά, στη φύση οι σχέσεις αυτές είναι μη γραμμικές, αφού η παρατηρούμενη βιολογική απόκριση ενός χημικού μορίου είναι απόρροια τόσο πολύπλοκων σχέσεων μεταξύ δομικών χαρακτηριστικών και βιολογικής δραστηριότητας όσο και αλληλεπιδράσεων του μορίου μέσα σε ένα βιολογικό σύστημα. Επομένως, μη-γραμμικές προσεγγίσεις μοντελοποίησης μπορούν να παράξουν καλύτερα προβλεπτικά QSAR μοντέλα. Ως γνωστόν, τα τεχνητά νευρωνικά δίκτυα είναι πολύ αποτελεσματικά στην μοντελοποίηση μη-γραμμικών σχέσεων και έτσι, κατάφεραν και βρήκαν ευρεία εφαρμογή και στην περίπτωση των QSAR μοντελοποιήσεων. [52]

Τα νευρωνικά δίκτυα που χρησιμοποιήθηκαν, αρχικά, ήταν "αβαθή" (shallow neural networks), δηλαδή είχαν μόλις ένα κρυφό στρώμα. Ένα νευρωνικό δίκτυο ενός κρυφού στρώματος, όπως έχει αποδειχθεί, μπορεί να προσεγγίσει οποιαδήποτε συνεχή συνάρτηση (βλ. υποκεφάλαιο 1.4). Ταυτόχρονα, η έλλειψη υπολογιστικής ισχύος στην δεκαετία

του '90 και στις αρχές της δεκαετίας του '00 δεν επέτρεπε την βιώσιμη εκπαίδευση βαθιών νευρωνικών δικτύων. Επομένως, για την παραγωγή καλών προβλεπτικών μοντέλων, τα οποία δεν θα χαρακτηρίζονταν από υποεκπαίδευση ή υπερεκπαίδευση δινόταν μεγάλο βάρος στην επιλογή των κατάλληλων περιγραφικών μεταβλητών για κάθε σύνολο δεδομένων, όπως άλλωστε συνηθίζεται στην QSAR μοντελοποίηση. [53]

Τα τελευταία χρόνια, όμως, η ανάπτυξη της βαθιάς μάθησης, μέσα από την χρήση μοντέρνων CPU και GPU οδήγησε στην παραγωγή μοντέλων νευρωνικών δικτύων εμπρόσθιας τροφοδότησης με πολλαπλά κρυφά στρώματα και μεγάλο αριθμό νευρώνων (100) ανά κρυφό στρώμα. Μάλιστα, καινοτόμες ρυθμίσεις των νευρωνικών δικτύων, όπως η συνάρτηση ενεργοποίησης *relu*, η χρήση *dropout* για αποφυγή υπερπροσαρμογής, αλλά και μέθοδοι βέλτιστης επιλογής υπερπαραμέτρων, όπως η μπεϋζιανή βελτιστοποίηση, επέτρεψαν την εξαγωγή QSAR μοντέλων βαθιών νευρωνικών δικτύων που έφταναν ή ξεπερνούσαν σε δυνατότητες τα παραδοσιακά QSAR μοντέλα χρησιμοποιώντας ως είσοδο μεγάλο όγκο δομικών περιγραφικών μεταβλητών [54] [55] [56]. Πώς, όμως, σύγχρονες αρχιτεκτονικές βαθιών νευρωνικών δικτύων μπορούν να οδηγήσουν στην πλήρη ανεξαρτησία των QSAR από τις δομικές περιγραφικές μεταβλητές;

Οι descriptor-free προσεγγίσεις για QSAR μοντελοποίηση κινούνται σε 3 άξονες:

1. Μοντέλα long short-term memory (LSTM) νευρωνικών δικτύων, τα οποία δέχονται ως είσοδο μια ακολουθία διανυσμάτων προκαθορισμένου μεγέθους. Η ακολουθία αυτή παράγεται κατόπιν επεξεργασίας της συμβολοσειράς (*string*) που αντιπροσωπεύει ένα χημικό μόριο εάν αυτό γραφεί με ένα σύστημα γραφής για υπολογιστική χρήση, όπως το *smiles* (βλ. υποκεφάλαιο 5.2). [57]
2. Μοντέλα συνελκτικών νευρωνικών δικτύων, τα οποία δέχονται ως είσοδο εικόνες των μορίων. Το μοντέλο *DeepSnap* μοντελοποιεί σχέσεις δομής-βιολογικής δραστηριότητας μέσω φωτογραφικών στιγμιότυπων τρισδιάστατων μοριακών δομών. [58]
3. Μοντέλα νευρωνικών δικτύων γράφων, τα οποία δέχονται ως είσοδο χημικό γράφο. Ο χημικός γράφος αποτελεί μια φυσική αναπαράσταση του χημικού μορίου, αφού οι κόμβοι αποτελούν τα άτομα και οι ακμές τους δεσμούς. Ακόμη, χαρακτηριστικά μπορούν να εκχωρηθούν τόσο στους κόμβους όσο και στις ακμές καθιστώντας την μοντελοποίηση των μορίων ακόμα πιο ρεαλιστική. Τα νευρωνικά δίκτυα γράφων αντιμετωπίζουν αποτελεσματικά το ζήτημα της παραγωγής αρκετά διαφορετικών περιγραφικών μεταβλητών ή συμβολοσειρών *SMILES* ή εικόνων για μόρια παρόμοιας χημικής δομής [59] και σε αρκετές περιπτώσεις υπερνικούν σε απόδοση παραδοσιακά QSAR μοντέλα [59] [60].

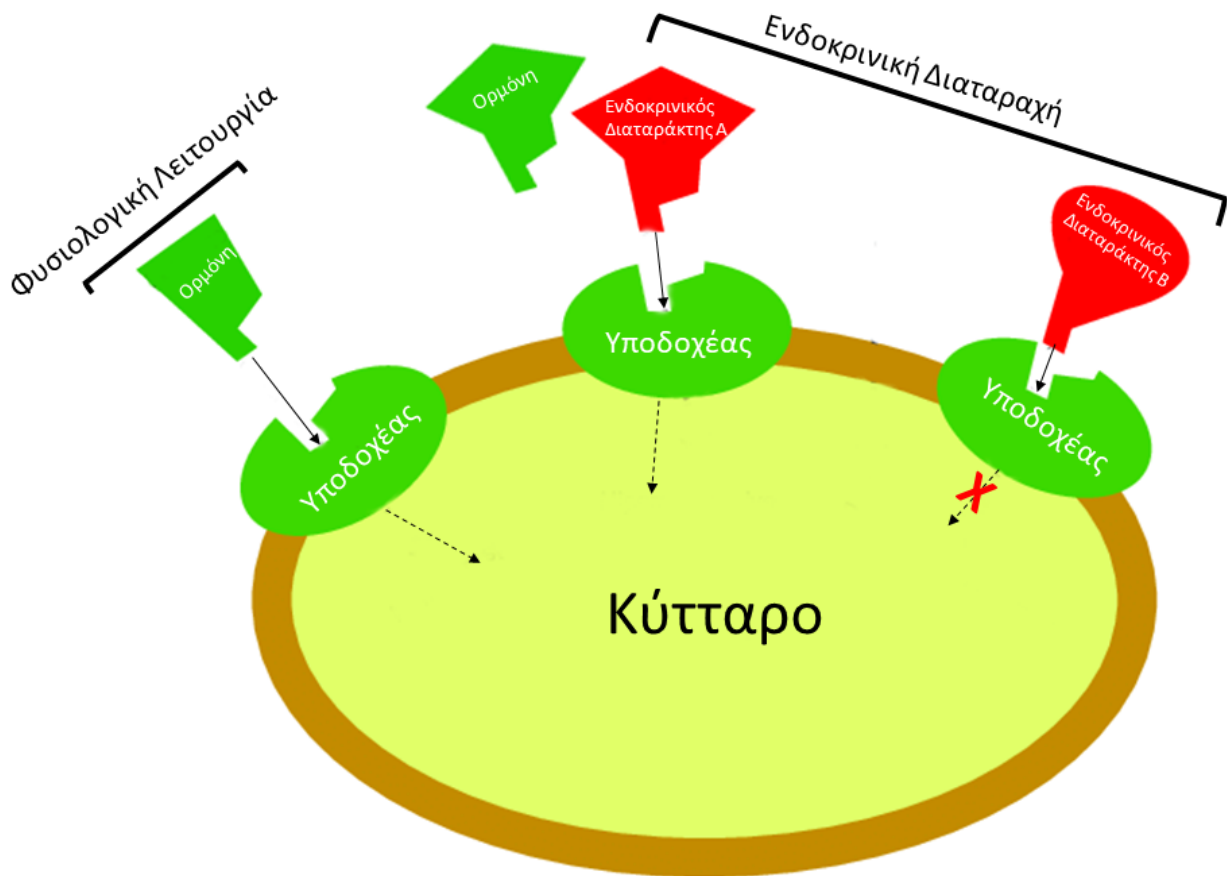
Κεφάλαιο 5

Μελέτη Περίπτωσης – Ενδοκρινικοί Διαταράκτες (Case Study – Endocrine Disruptors)

Πολυάριθμες χημικές ενώσεις που χρησιμοποιούνται στην καθημερινότητα ή και διοχετεύονται στο περιβάλλον λόγω της ανθρώπινης δραστηριότητας έχουν την δυνατότητα να διαταράσσουν το ενδοκρινικό σύστημα ανθρώπων και άγριας πανίδας. Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας, ως ενδοκρινικός διαταράκτης (endocrine disruptor - ED) χαρακτηρίζεται μια εξωγενής ουσία ή μείγμα, η οποία μεταβάλλει λειτουργίες του ενδοκρινικού συστήματος και ως αποτέλεσμα, προκαλεί δυσμενείς επιπτώσεις στην υγεία ενός ακέραιου οργανισμού, απογόνων του ή (υπο)πληθυσμούς [63]. Ειδικότερα, όσον αφορά την άγρια πανίδα, έκθεση σε EDs έχει συσχετιστεί με ανώμαλη λειτουργία του θυρεοειδούς σε πληθυσμούς πτηνών και ψαριών, με μειωμένη γονιμότητα σε πληθυσμούς πτηνών, ψαριών, οστρακοειδών και θηλαστικών, με προβληματική έκφραση χαρακτηριστικών του φύλου σε πληθυσμούς μαλάκιων, ψαριών, πτηνών και θηλαστικών, καθώς και άλλες διαταραχές [64]. Ορισμένες γνωστές επιδράσεις στην ανθρώπινη υγεία είναι η ελαττωματική λειτουργία αναπαραγωγικών οργάνων τόσο σε αρσενικά όσο και σε θηλυκά άτομα, ο καρκίνος του μαστού, του ενδομητρίου και των όρχεων, μεταβολές στην έναρξη της εφηβείας και μεταβολές στην αναλογία αρσενικών προς θηλυκών γεννήσεων [65]. Με ποιον τρόπο, όμως, οι EDs επιδρούν στο ενδοκρινικό σύστημα με αποτέλεσμα όλες τις ανεπιθύμητες συνέπειες που προαναφέρθηκαν;

Το ενδοκρινικό σύστημα αποτελείται από ειδικά όργανα, τους ενδοκρινείς αδένες, οι οποίοι παράγουν και εκκρίνουν τις ορμόνες, ειδικές ουσίες που έχουν τον ρόλο της μετάδοσης σημάτων σε ολόκληρο το σώμα του οργανισμού μέσω του αίματος στους ζωικούς οργανισμούς ή του οποίου στους φυτικούς οργανισμούς. Ακόμη, μέρος του ενδοκρινικού συστήματος αποτελούν και οι αντίστοιχοι υποδοχείς των σημάτων. Τα σημαντικότερα όργανα του ενδοκρινικού συστήματος του ανθρώπου είναι η υπόφυση, ο θυρεοειδής, ο παραθυρεοειδής, τα επινεφρίδια, ο θύμος, το πάγκρεας, καθώς επίσης, οι ωοθήκες στα θηλυκά άτομα και οι όρχεις στα αρσενικά. Το ενδοκρινικό σύστημα με την έκκριση ορμονών, μαζί με το νευρικό σύστημα, ουσιαστικά, ρυθμίζουν την λειτουργία του οργανισμού. Η παρουσία ή μη ορμονών επηρεάζει ποικίλες ενέργειες ή καταστάσεις του ανθρώπου, όπως τον μεταβολισμό, τα δευτερογενή χαρακτηριστικά του φύλου τόσο σε γυναίκες όσο και σε άνδρες, την συμπεριφορά, την ομοιόσταση, τα επίπεδα γλυκόζης στα κύτταρα κ.λπ..Οι EDs έχουν τη δυνατότητα είτε (α) να μιμούνται την επίδραση ενδογενών ορμονών και να προκαλούν επιβλαβείς μη-αναμενόμενες αντιδράσεις είτε (β) να εμποδίζουν την αλληλεπίδραση

των ορμονών με τους ενδογενείς υποδοχείς με ανεπιθύμητες συνέπειες στο αναπτυξιακό, αναπαραγωγικό, νευρολογικό και ανοσοποιητικό σύστημα. [66] [67]



Σχήμα 5.1: Σχηματική απεικόνιση ενδοκρινικής διαταραχής. Ο ενδοκρινικός διαταράκτης Α μιμείται την ορμόνη μεταδίδοντας μη φυσιολογικό σήμα στο κύτταρο. Ο ενδοκρινικός διαταράκτης Β μπλοκάρει τον υποδοχέα με αποτέλεσμα την αδυναμία πρόσδεσης με ορμόνη και μετάδοσης σήματος στο κύτταρο.

Μια σημαντική κατηγορία ορμονικών υποδοχέων, η δράση των οποίων αποτελεί αντικείμενο εκτενούς έρευνας είναι οι υποδοχείς οιστρογόνων (estrogen receptors - ERs). Ανήκουν στην ευρύτερη οικογένεια των πυρηνικών υποδοχέων (nuclear receptors), οι οποίοι βρίσκονται στον πυρήνα ή στο κυτταρόπλασμα και όταν ενεργοποιηθούν δρουν ως μεταγραφικοί ρυθμιστές στον πυρήνα. Υποδέχονται μικρά ή υδρόφοβα μόρια, τα οποία έχουν την ικανότητα να διαπερνούν την κυτταρική μεμβράνη, όπως οι στεροειδείς ορμόνες (ανάμεσά τους τα οιστρογόνα) και θυροειδείς ορμόνες (ανάμεσά τους η θυροξίνη). [66] Οι EDs μπορούν να προσδεθούν στους υποδοχείς οιστρογόνων και να παρέμβουν στην φυσιολογική μεταφορά σημάτων οιστρογόνων μέσα από γονιδιωματικές ή μη-γονιδιωματικές οδούς. Σε κάθε περίπτωση, η πρόσδεση μη φυσιολογικών οιστρογόνων (ή ξενοοιστρογόνων) μπορεί να οδηγήσει σε μεταβολές που δυνητικά μπορούν να διαταράξουν ολόκληρο το ενδοκρινικό σύστημα του ανθρώπου [67]. Σημαντικά ξενοοιστρογόνα είναι τα περιβαλλοντικά οιστρογόνα, τα οποία προέρχονται από τοξικά απορρίμματα και άλλες ρυπογόνες ουσίες (πλαστικά, βιομηχανικές επιφανειοδραστικές ουσίες, παρασιτοκτόνα) και καταλήγουν στο έδαφος, αλλά και σε υδάτινα οικοσυστήματα. Είναι χαρακτηριστικό ότι ένα πλαστικό μονομερές που δρα ως ξενοοιστρογόνο, η δισφαινόλη-Α (bisphenol A - BPA), βρίσκεται σε ανιχνεύσιμα επίπεδα σε πάνω από το 90% των Αμερικανών και άλλων πλη-

θυσμών. [68]

Η κύρια πηγή έκθεσης του ανθρώπινου οργανισμού σε EDs είναι η διατροφή του, καθώς, λόγω της λιποφιλικής τους φύσης, είναι πολύ πιθανή η συσσώρευσή τους στους λιπώδεις ιστούς ζώων που καταναλώνει (π.χ. ψάρια) [69]. Επίσης, το ξηνοοιστρογόνο BPA ανιχνεύεται σε όλα τα δείγματα φρούτων και λαχανικών, στα οποία εισρέει μέσω της άρδευσης. Εκτός από τα φαγητά, EDs μπορούν να προσληφθούν και από οικιακά προϊόντα, όπως καθαριστικά και προϊόντα περιποίησης. Συνοπτικά, οι EDs αποδεδειγμένα έχουν δυσμενείς επιπτώσεις στην ανθρώπινη υγεία και ιδιαίτερα, στην γονιμότητα, την εγκυμοσύνη και την ανάπτυξη του εμβρύου. Ιδιαίτερα, η έκθεση κατά την διάρκεια της εγκυμοσύνης και στα αρχικά στάδια της ζωής ενός ανθρώπου μπορεί να επηρεάσει την συμπεριφορά στην μετέπειτα ζωή τους οδηγώντας σε προβληματική κοινωνική συμπεριφορά [70]. Η συνεχής και αναπόφευκτη τη σύγχρονη εποχή έκθεση σε EDs τη στιγμή που πολλοί από αυτούς είναι ακόμη άγνωστοι λόγω ελλειπών πειραματικών δεδομένων δείχνει την σημαντικότητα ανάπτυξης υπολογιστικών εργαλείων για την πρόβλεψη του δυναμικού ενδοκρινικής διαταραχής χημικών ουσιών [71].

5.1 Το Σύνολο των Δεδομένων (Dataset)

Τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση των μαθηματικών μοντέλων βαθιάς μηχανικής μάθησης συλλέχθηκαν από την Βάση Δεδομένων Οιστρογονικής Δραστηριότητας (Estrogenic Activity Database - EADB) του Οργανισμού Τροφίμων και Φαρμάκων των ΗΠΑ (U.S. Food and Drug Administration - FDA). Η EADB περιλαμβάνει δεδομένα οιστρογονικής δραστηριότητας για 8212 χημικές ουσίες, τα οποία προήλθαν από *in vitro* και *in vivo* πειράματα σε 11 οργανισμούς. Όλες οι χημικές ενώσεις είναι διαθέσιμες σε κωδικοποίηση SMILES που επιτρέπει την ανάγνωση και κατανόηση των δομών από ηλεκτρονικούς υπολογιστές. Η κωδικοποίηση SMILES παρουσιάζεται λεπτομερώς στο επόμενο υποκεφάλαιο (υποκεφάλαιο 5.2). Στην παρούσα εργασία χρησιμοποιήθηκαν πειραματικά δεδομένα συγγένειας πρόσδεσης (binding affinity) στους υποδοχείς οιστρογόνων (ERs). Η συγγένεια πρόσδεσης αναφέρεται στην ισχύ αλληλεπίδρασης με τη μορφή της πρόσδεσης χημικών μορίων σε ορμονικούς υποδοχείς και συγκεκριμένα, στους ERs. Συνολικά, επιλέχθηκαν 1457 πειραματικές τιμές, οι οποίες προήλθαν από *in vitro* πειράματα σε ανθρώπινα κύτταρα.

Οι πειραματικές τιμές αναφέρονται στην σχετική συγγένεια πρόσδεσης (relative binding affinity), δηλαδή είναι κανονικοποιημένες ως προς την τιμή συγγένειας πρόσδεσης της 17-β οιστραδιόλης (estradiol - 2). Η οιστραδιόλη είναι ένα οιστρογόνο (στεροειδής ορμόνη) και αποτελεί την σημαντικότερη ορμόνη του γυναικείου φύλου. Έχει, δηλαδή, μια φυσιολογικά υψηλή τιμή συγγένειας πρόσδεσης στους ERs. Για την κανονικοποίηση, η RBA της E2 τέθηκε ίση με 100 ή $\log_{10}RBA = 2$. Σε γενικές γραμμές, όσο μεγαλύτερο το $\log RBA$ μιας ουσίας τόσο πιο ισχυρό είναι το δυναμικό πρόσδεσής της στους υποδοχείς οιστρογόνων. Το εύρος των πειραματικών αποκρίσεων $\log RBA$ κυμαίνεται στο διάστημα $[-3.328, 2.857]$. Οι τιμές αυτές αποτελούν και τις ετικέτες του συνόλου δεδομένων.

Τα μοντέλα βαθιάς επιβλεπόμενης μάθησης που αναπτύχθηκαν επιλύουν προβλήματα ταξινόμησης. Επομένως, τα πειραματικά δεδομένα χωρίζονται σε 3 κατηγορίες (κλάσεις):

Πίνακας 5.1: Πίνακας κλάσεων του προβλήματος ταξινόμησης ενδοκρινικών διαταρακτών

Κλάσεις	Εύρος πειραματικών αποκρίσεων	Αριθμός παραδειγμάτων	Ιδιότητα*
1	[-3.328, -0.26]	531	Ασθενής συγγένεια πρόσδεσης με ERs
2	[-0.259, 0.824]	445	Μέτρια συγγένεια πρόσδεσης με ERs
3	[0.826, 2.857]	481	Ισχυρή συγγένεια πρόσδεσης με ERs

*Σημειώνεται ότι ο χαρακτηρισμός της ισχύος πρόσδεσης σε ERs σε κάθε κλάση προκύπτει πάντα σε σύγκριση με την αντίστοιχη τιμή της οιστραδιόλης (τιμή αναφοράς), η οποία έχει μια σχετικά υψηλή συγγένεια πρόσδεσης. Παρ' όλα αυτά, σε κάθε περίπτωση όλες οι ουσίες του συνόλου δεδομένων είναι οιστρογονικά ενεργές και δυνητικά επικίνδυνες για οργανισμούς ανάλογα με την συγκέντρωσή τους σε αυτούς [67].

5.2 Το Σύστημα Απλοποιημένης Μοριακής Γραμμικής Γραφής (Simplified Molecular-Input Line-Entry System)

Το σύστημα απλοποιημένης μοριακής γραμμικής γραφής (simplified molecular-input line-entry system) ή SMILES είναι ένα σύστημα γραφής χημικών ενώσεων και αντιδράσεων ειδικά σχεδιασμένο για υπολογιστική χρήση από χημικούς. Εισήχθη από τον D. Weininger στα τέλη της δεκαετίας του '80 [61] και αναπτύχθηκε από την DAYLIGHT Chemical Information Systems. Το κυριότερο πλεονέκτημα της SMILES είναι ότι χρησιμοποιεί γλωσσικούς χαρακτήρες αντί για υπολογιστικές δομές δεδομένων, γεγονός που σε συνδυασμό με το απλό της "λεξιλόγιο" και τους λίγους κανόνες "γραμματικής" καθιστά εύκολη την ανάγνωση των χημικών ενώσεων και από τους ανθρώπους. Η επινόηση της SMILES στηρίχθηκε στην θεώρηση του χημικού μορίου ως γράφος, λαμβάνοντας υπόψιν τα άτομα ως κόμβους και τους δεσμούς ως ακμές. Η SMILES, λοιπόν, κατορθώνει να περιγράψει μοναδικά τον χημικό γράφο με τρόπο ευανάγνωστο από το ανθρώπινο μάτι προσεγγίζοντας ουσιαστικά την δισδιάστατη απεικόνιση των χημικών ενώσεων στο χαρτί. Κανένα δεδομένο σχετικά με την τρισδιάστατη διάταξη των ατόμων δεν λαμβάνεται υπόψιν.

Η γραφή SMILES αποτελείται από λατινικούς χαρακτήρες και διάφορα σύμβολα σε σειρά χωρίς κενά. Τα άτομα υδρογόνου μπορούν να παραλείπονται ή να περιλαμβάνονται ανάλογα με τη μορφή του χημικού γράφου, ο οποίος αποτυπώνεται σε γραφή SMILES (Hydrogen-suppressed ή Hydrogen-complete). Αρωματικές δομές μπορούν να γραφούν απευθείας ή σε μορφή Kekulé. Οι συνοπτικοί κανόνες της SMILES είναι οι εξής [62]:

- Άτομα
 - Κάθε άτομο, εκτός του υδρογόνου, αναπαρίσταται από το χημικό του σύμβολο κλεισμένο σε αγκύλες, [], με κεφαλαίο το πρώτο γράμμα και μικρό το δεύτερο (όπου συναντάται). Για παράδειγμα ο στοιχειακός χρυσός συμβολίζεται με [Au]
 - Τα εξής άτομα που συναντώνται σε οργανικές ενώσεις B, C, N, O, P, S, F, Cl,

Br, I μπορούν να γραφούν εκτός αγκύλων εάν ο αριθμός των προσαρτημένων σε αυτά υδρογόνων είναι ίσος με το χαμηλότερο τυπικό χημικό σθένος με εξαίρεση τα στοιχεία N, P και S (4 για τον C, 3 για το B, 3 ή 5 για το N, 2 για το O, 3 ή 5 για τον P, 2 ή 4 ή 6 για το S και 1 για τα αλογόνα). Παραδείγματα γραφής ενώσεων με τον παραπάνω κανόνα αποτελούν το C για το μεθάνιο (CH₄), το O για το νερό (H₂O) και το Cl για το υδροχλωρικό οξύ (HCl).

- Άτομα σε αρωματικούς δακτυλίους αντιπροσωπεύονται από μικρά γράμματα αντί για κεφαλαία.
- Τα προσαρτημένα υδρογόνα ή/και φορτία των χημικών ενώσεων πρέπει να καθορίζονται εντός αγκυλών. Ο αριθμός των υδρογόνων καθορίζεται από το σύμβολο H ακολουθούμενο από το αντίστοιχο ψηφίο, ενώ το φορτίο ανάλογα με τον τύπο του καθορίζεται από τα σύμβολα + ή - ακολουθούμενα από το αντίστοιχο ψηφίο. Παραδείγματος χάριν, τα κατιόντα αμμωνίου και δισθενούς σιδήρου συμβολίζονται με [NH₄⁺] και [Fe + 2] αντίστοιχα.

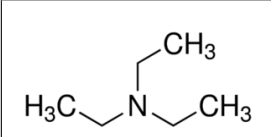
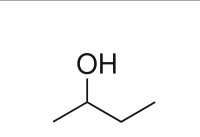
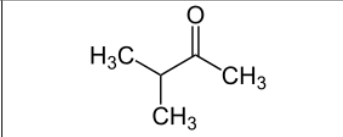
- Δεσμοί

Μονός, διπλός, τριπλός και αρωματικός δεσμός αντιπροσωπεύονται από τα σύμβολα -, =, και :, αντίστοιχα. Οι μονοί και οι αρωματικοί δεσμοί μπορούν να παραλείπονται από τον συμβολισμό.

π.χ. CC – Αιθάνιο, C = O – Φορμαλδεΰδη, O = C = O – Διοξείδιο του άνθρακα

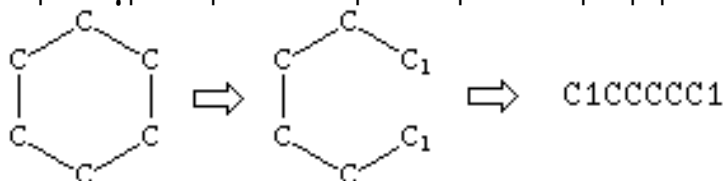
- Διακλαδώσεις

Οι διακλαδώσεις καθορίζονται μέσω της χρήσης παρενθέσεων. Η διακλαδική αλυσίδα συνδέεται με την κύρια αλυσίδα της ένωσης που περιγράφεται με το άτομο που βρίσκεται αριστερά της παρένθεσης. Παραδείγματα χημικών ενώσεων αποτελούν τα εξής:

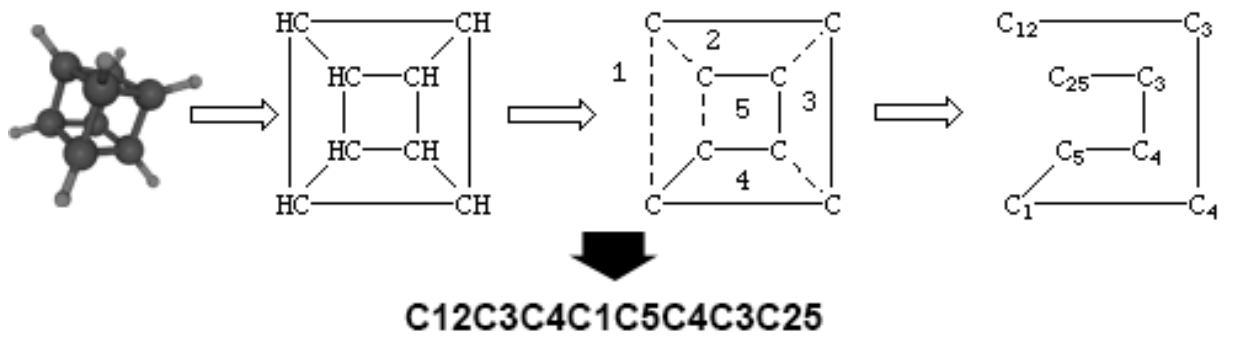
		
Τριαιθυλαμίνη	2-Βουτανόλη	3-Μέθυλο-2-βουτανόνη
CCN(CC)CC	CCC(C)O	CC(C)C(=O)C

- Κυκλικές δομές

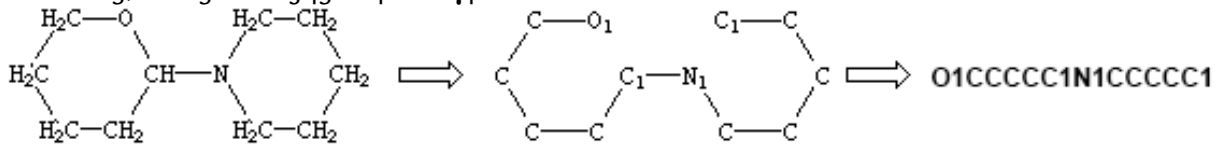
Οι κυκλικές δομές στη γραφή SMILES καθορίζονται με βάση τους κανόνες που περιγράφηκαν παραπάνω αφού προηγηθεί σπάσιμο του δακτυλίου σε ένα αυθαίρετο σημείο και δημιουργία με αυτόν τον τρόπο ενός γραμμικού (μη κυκλικού) χημικού γράφου. Οι αριθμοί 1 έως 9 χρησιμοποιούνται για να υποδείξουν το αρχικό και το τελικό άτομο ενός δακτυλίου και γράφονται αμέσως μετά τα αντίστοιχα άτομα. Ως παράδειγμα παρατίθεται η τυπική κυκλική δομή του κυκλοεξανίου:



Ένα άτομο που συσχετίζεται με περισσότερους από έναν δακτυλίους γράφεται με τη συνοδεία των αντίστοιχων αριθμών (παράδειγμα κυβανίου):



Τέλος, ο ίδιος αριθμός μπορεί να χρησιμοποιηθεί για περισσότερους από έναν δακτυλίους, όπως στο εξής παράδειγμα:



Κεφάλαιο 6

Ανάπτυξη Μοντέλων και Αποτελέσματα (Models Development and Results)

6.1 Μέτρα Αξιολόγησης (Evaluation Metrics)

Η αξιολόγηση της απόδοσης του μοντέλου είναι ένα αναπόσπαστο κομμάτι κάθε τεχνικής μηχανικής μάθησης. Οι δείκτες που ποσοτικοποιούν την αξιολόγηση αυτή συναντώνται με τον όρο *metrics* και ποικίλουν ανάλογα με το είδος του μοντέλου και τον σκοπό της επιδιωκόμενης πρόβλεψης. Τα *metrics* μπορούν να εκφράζουν την τελική απόδοση ενός μοντέλου, αλλά μπορούν, επίσης, να χρησιμοποιηθούν για την online παρακολούθηση της απόδοσης ενός μοντέλου κατά τη διάρκεια της εκπαίδευσής του. Τα *metrics* του προβλήματος ταξινόμησης 3 κλάσεων που χρησιμοποιούνται στην παρούσα εργασία είναι η ακρίβεια, ο πίνακας σύγχυσης και ο συντελεστής συσχέτισης Matthews [72].

6.1.1 Πίνακας Σύγχυσης (Confusion Matrix)

Ο πίνακας σύγχυσης (*confusion matrix*) ή πίνακας σφάλματος (*error matrix*) αποτελεί έναν τρόπο οπτικοποίησης της ικανότητας πρόβλεψης ενός μοντέλου ταξινόμησης σε κλάσεις. Οι σειρές του πίνακα αντιστοιχούν στις πραγματικές τιμές (ετικέτες του συνόλου δεδομένων), ενώ οι στήλες αντιστοιχούν στις προβλέψεις του μοντέλου. Ο πίνακας σύγχυσης για πρόβλημα ταξινόμησης 3 κλάσεων έχει την μορφή του σχήματος 6.1 [72].

Στη γενική του μορφή, για πρόβλημα ταξινόμησης n κλάσεων έχει μέγεθος $n \times n$. Οι σωστές προβλέψεις του μοντέλου αντιστοιχούν στα στοιχεία της κυρίας διαγωνίου του πίνακα, ενώ όλα τα υπόλοιπα στις λανθασμένες.

6.1.2 Ακρίβεια (Accuracy)

Η ακρίβεια (*accuracy*) μετράει τις σωστές προβλέψεις του μοντέλου και λαμβάνει τιμές από 0 έως 1. Ορίζεται ως ο λόγος των σωστών προβλέψεων του μοντέλου προς τον συνολικό αριθμό των παραδειγμάτων. Εάν C_{ij} τα στοιχεία της γενικής μορφής του πίνακα σύγχυσης, τότε η ακρίβεια ορίζεται ως εξής [72]:

$$Accuracy = \frac{\sum_{i=1}^n C_{ii}}{\sum_{i=1}^n \sum_{j=1}^n C_{ij}} \quad (6.1)$$

<div style="display: inline-block; transform: rotate(-45deg); transform-origin: center;"> Προβλέψεις μοντέλου Πραγματικές τιμές </div>	Κλάση 1	Κλάση 2	Κλάση 3
Κλάση 1	True "Class 1"	False "Class 1"	False "Class 1"
Κλάση 2	False "Class 2"	True "Class 2"	False "Class 2"
Κλάση 3	False "Class 3"	False "Class 3"	True "Class 3"

Σχήμα 6.1: Πίνακας σύγχυσης για πρόβλημα ταξινόμησης 3 κλάσεων.

6.1.3 Συντελεστής Συσχέτισης Matthews (Matthews Correlation Coefficient)

Ένα περισσότερο εκφραστικό μέτρο αξιολόγησης, συγκριτικά με την ακρίβεια, είναι ο συντελεστής συσχέτισης Matthews (Matthews Correlation Coefficient - MCC). Παίρνει τιμές από -1 έως 1 και λαμβάνει υπόψιν τις σωστές και λανθασμένες προβλέψεις σε επίπεδο κλάσης και όχι στο σύνολο των δεδομένων, όπως η ακρίβεια. Τιμές κοντά στο 1 αντιπροσωπεύουν μια τέλεια πρόβλεψη, τιμές κοντά στο 0 μια πρόβλεψη που δεν διαφέρει από την τυχαία κατανομή και τιμές κοντά στο -1 πλήρη αδυναμία συσχέτισης των προβλέψεων με τις αρχικές παρατηρήσεις. Με βάση τη γενική μορφή του πίνακα σύγχυσης ορίζεται ως εξής [72]:

$$MCC = \frac{c \cdot s - \sum_{k=1}^n p_k t_k}{\sqrt{\left(s^2 - \sum_{k=1}^n p_k^2\right) \left(s^2 - \sum_{k=1}^n t_k^2\right)}} \quad (6.2)$$

όπου k δηλώνει την κλάση,

$s = \sum_{i=1}^n \sum_{j=1}^n C_{ij}$, το συνολικό πλήθος των στοιχείων,

$c = \sum_{i=1}^n C_{ii}$, το συνολικό πλήθος των σωστών προβλέψεων,

$p_k = \sum_{i=1}^n C_{ki}$, το πλήθος των προβλέψεων κλάσης k ,

$t_k = \sum_{i=1}^n C_{ik}$, το πλήθος των πραγματικών ετικετών κλάσης k .

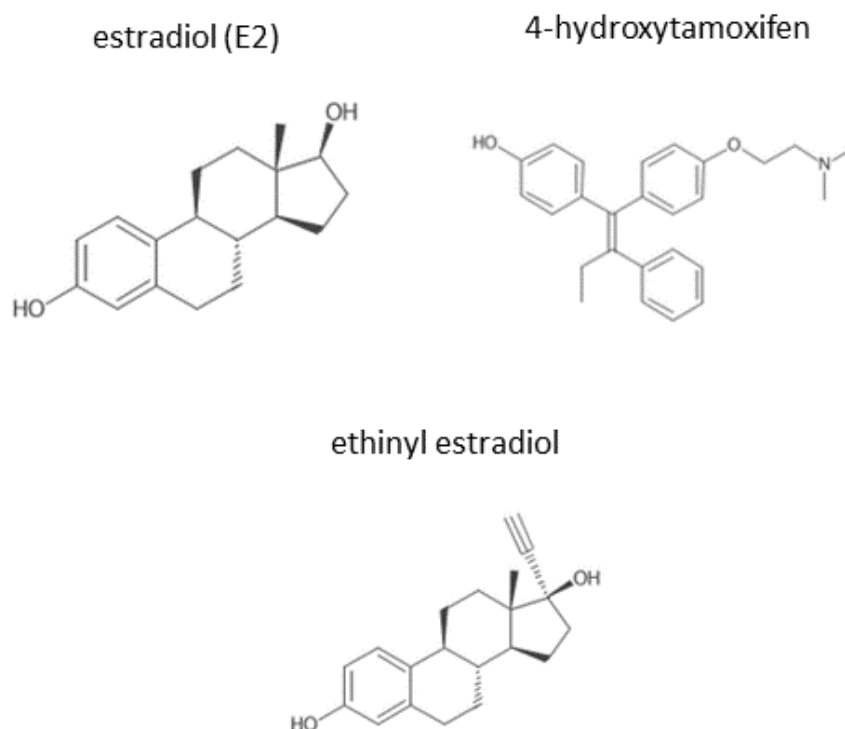
6.2 Διασταυρούμενη Επικύρωση (Cross-Validation)

Η διασταυρούμενη επικύρωση (cross validation) είναι μια στατιστική μέθοδος επικύρωσης της απόδοσης μοντέλων μηχανικής μάθησης. Κατά την εφαρμογή της, το σύνολο των δεδομένων χωρίζεται με τυχαίο τρόπο σε k ισομεγέθεις ομάδες (k -fold cross validation). Από τις k αυτές ομάδες, μια ομάδα επιλέγεται να αποτελεί το σύνολο δεδομένων επαλήθευσης και οι υπόλοιπες ($k - 1$) αποτελούν το σύνολο δεδομένων εκπαίδευσης. Το μοντέλο, κατόπιν, εκπαιδεύεται k φορές, όπου καθεμία από τις k ομάδες χρησιμοποιείται ακριβώς μια φορά σαν ομάδα επαλήθευσης. Με αυτόν τον τρόπο, το μοντέλο εκπαιδεύεται

και επαληθεύεται σε όλο το σύνολο δεδομένων, αυξάνοντας έτσι την γενίκευση των αποτελεσμάτων του. Τα συνολικά μέτρα αξιολόγησης του μοντέλου προκύπτουν ως ο μέσος όρος των k αποτελεσμάτων στα δεδομένα επαλήθευσης από κάθε ξεχωριστή εκπαιδευτική διαδικασία του μοντέλου [73].

6.3 Προηγούμενη Εργασία (Previous Work)

Μια προηγούμενη εφαρμογή βαθιάς μηχανικής μάθησης για την ταξινόμηση ενδοκρινικών διαταρακτών του υφιστάμενου συνόλου δεδομένων (βλ. κεφάλαιο 5) αναπτύχθηκε από τον υποψήφιο διδάκτορα της Μονάδας Αυτόματης Ρύθμισης και Πληροφορικής της Σχολής Χημικών Μηχανικών Παντελή Καρατζά στο πλαίσιο του ΔΠΜΣ "Εφαρμοσμένες Μαθηματικές Επιστήμες" [74]. Αφορά την επίλυση του προβλήματος ταξινόμησης μέσω της απεικόνισης των χημικών μορίων ως δισδιάστατες εικόνες και την επεξεργασία τους μέσω ενός μοντέλου συνελκτικών νευρωνικών δικτύων. Αρχικά, οι χημικές ενώσεις σε μορφή Smiles μετατρέπονται σε ασπρόμαυρες δισδιάστατες εικόνες μέσω της Java βιβλιοθήκης Chemistry Development Kit [75]. Το μέγεθος των εικόνων που παράγονται εξαρτάται από το μόριο που απεικονίζουν. Κατά κανόνα, οι εικόνες που αποτελούν είσοδο σε ένα συνελκτικό νευρωνικό δίκτυο έχουν ίδιο μέγεθος. Έτσι, κάθε εικόνα του συνόλου δεδομένων τοποθετείται στοιχισμένη στο κέντρο ενός λευκού φόντου μεγέθους 200×200 pixels.



Σχήμα 6.2: Παραδείγματα εικόνων που χρησιμοποιούνται ως είσοδοι στο συνελκτικό νευρωνικό δίκτυο.

Κατόπιν, το μέγεθος του συνόλου δεδομένων αυξάνεται με χρήση τεχνικής ενίσχυσης δεδομένων (data augmentation). Ειδικότερα, οι εικόνες του συνόλου δεδομένων περιστρέφονται διαδοχικά κατά 90, 180 και 270 μοίρες διατηρώντας την ετικέτα τους τετραπλασιάζοντας με αυτόν τον τρόπο τον όγκο δεδομένων, πάνω στον οποίο εκπαιδεύεται το συνελκτικό νευρωνικό δίκτυο. Η βέλτιστη αρχιτεκτονική περιλαμβάνει 3 συνελκτικά στρώματα. Στο τέλος κάθε συνελκτικού στρώματος παρεμβάλλεται ένα στρώμα 2×2

υποδειγματοληψίας μεγίστου. Στη συνέχεια, οι τανυστές τρίτης τάξης στην έξοδο του τελικού στρώματος υποδειγματοληψίας μετατρέπονται σε διανύσματα μέσω της διαδικασίας του flattening και εισάγονται σε 2 πλήρως συνδεδεμένα στρώματα 1024 κόμβων το καθένα. Τέλος, ένα στρώμα εξόδου 3 κόμβων με ενεργοποίηση softmax παράγει τις τελικές προβλέψεις κλάσεων. Σημειώνεται ότι τόσο τα συνελκτικά στρώματα όσο και τα κρυφά πλήρως συνδεδεμένα στρώματα του δικτύου ενεργοποιούνται με συνάρτηση relu, ενώ η βελτιστοποίηση γίνεται με τον αλγόριθμο Adam και ρυθμό μάθησης ίσο με 0.3 . Για την επικύρωση του μοντέλου το σύνολο δεδομένων χωρίστηκε με τυχαίο τρόπο σε σύνολο δεδομένων εκπαίδευσης (90% του αρχικού συνόλου) και σύνολο δεδομένων επαλήθευσης (10% του αρχικού συνόλου). Τα αποτελέσματα απόκρισης της ακρίβειας συναρτήσε των εποχών τόσο για τα δεδομένα επαλήθευσης όσο και για τα δεδομένα εκπαίδευσης φαίνεται στα παρακάτω σχήματα :



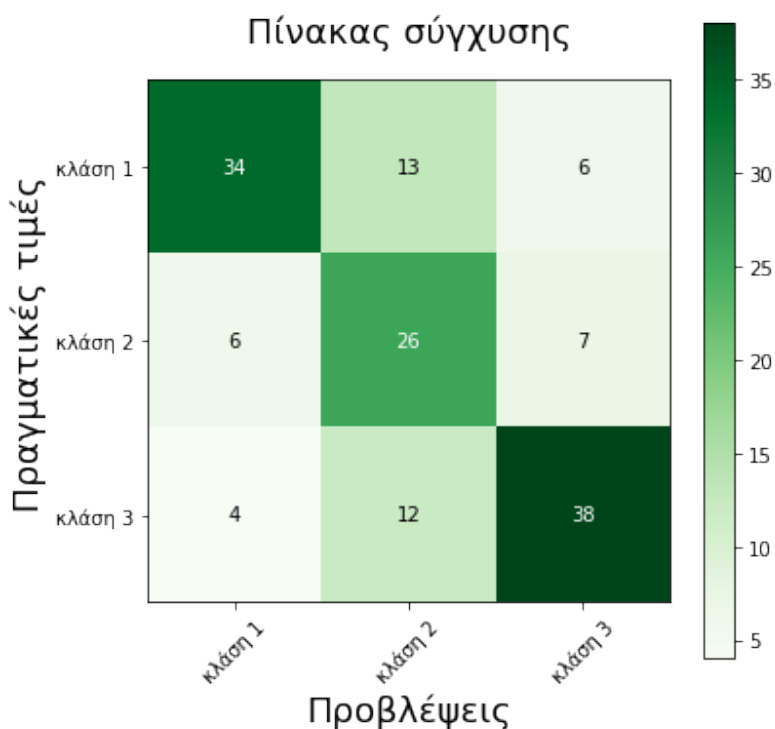
Σχήμα 6.3: Μοντέλο CNN: αποτελέσματα ακρίβειας συναρτήσε εποχών στα δεδομένα εκπαίδευσης



Σχήμα 6.4: Μοντέλο CNN: αποτελέσματα ακρίβειας συναρτήσε εποχών στα δεδομένα επαλήθευσης

Αναφορικά με την απόδοση του μοντέλου σε δεδομένα πάνω στα οποία δεν έχει εκπαιδευτεί (δεδομένα επαλήθευσης), η ακρίβειά του αγγίζει το 0.68, ενώ ο συντελεστής MCC το 0.51. Ακολουθεί ένας ενδεικτικός πίνακας σύγχυσης :

Πίνακας 6.1: Πίνακας σύγχυσης του μοντέλου CNN



6.4 Μοντέλο Νευρωνικών Δικτύων Εμπρόσθιας Τροφοδότησης (Feed-Forward Neural Networks Model)

6.4.1 Κατασκευή Δεδομένων Εισόδου (Input Data Construction)

Όπως αναφέρθηκε στο υποκεφάλαιο 1.2 της εργασίας, είσοδο στα νευρωνικά δίκτυα εμπρόσθιας τροφοδότησης ή MLP αποτελεί ένας πίνακας σχεδιασμού $\mathbf{X} \in \mathbb{R}^{n \times d}$ αποτελούμενος από n παραδείγματα και d ανεξάρτητες μεταβλητές. Οι ανεξάρτητες μεταβλητές θα έχουν την μορφή αριθμητικών περιγραφικών μεταβλητών (numerical descriptors), οι οποίοι παράγονται μέσω του λογισμικού Mordred [76]. Το Mordred είναι ένα ελεύθερα διαθέσιμο υπολογιστικό πακέτο υπολογισμού μοριακών περιγραφικών μεταβλητών (molecular descriptors). Συγκεκριμένα, υπολογίζει περισσότερες από 1600 ιδιότητες των μορίων βασισμένες σε διδιάστατες (2D) δομές και τρισδιάστατες (3D) δομές τους. Ως είσοδος του πακέτου, χρησιμοποιούνται χημικά μόρια σε μορφή SMILES. Σε αντίθεση με παλαιότερα λογισμικά παραγωγής molecular descriptors (π.χ. PaDEL, BlueDesc, PyDPI κ.λπ.), το Mordred προεπεξεργάζεται με αυτόματο τρόπο τα μόρια ανάλογα με τις απαιτήσεις του κάθε descriptor (προσθέτει ή αφαιρεί υδρογόνα, μετατρέπει τα SMILES σε μορφή Kekule, ανιχνεύει την αρωματικότητα των μορίων). Ταυτόχρονα, υπερτερεί στην αποδοτικότητα και ταχύτητα εκτέλεσης του αλγορίθμου, ακόμα και σε μόρια όσο μεγάλα όσο η τοξίνη MTX (το μεγαλύτερο μη-πολυμερές μόριο στη φύση με μοριακό βάρος 3422). Για παράδειγμα, στο λογισμικό PaDEL, αριθμητικές περιγραφικές μεταβλητές μεγάλων μορίων περιείχαν σε ορισμένες περιπτώσεις τιμές που έλειπαν εξαιτίας σφάλματος "λήξης χρόνου". Τέλος, το Mordred είναι το μοναδικό πακέτο που ελέγχει με αυτόματο τρόπο εάν οι τιμές των περιγραφικών μεταβλητών που παράγει βρίσκονται εντός μίας κατανομής από τιμές αναφοράς, οι οποίες προέρχονται από δημοσιευμένες μελέτες, αποτελέσματα άλλων λογισμικών και δημοσιευμένες τιμές από χειροκίνητους υπολογισμούς των αλγορίθμων υπολογισμού που εφαρμόζονται. Με αυτόν τον τρόπο, ελέγχεται η εγκυρότητα των

υπολογισμένων τιμών περιγραφικών μεταβλητών. Τα ονόματα και το πλήθος των μοριακών περιγραφικών μεταβλητών που παράγονται μέσω του Mordred λαμβάνοντας ως είσοδο χημικές ενώσεις σε μορφή SMILES είναι οι ακόλουθες:

Πίνακας 6.2: Μοριακές περιγραφικές μεταβλητές για τα δεδομένα εισόδου του μοντέλου MLP, όπως παράγονται μέσω του λογισμικού Mordred

Όνομα	Περιγραφή	Πλήθος Μεταβλητών
<i>2D</i> Δομικά Χαρακτηριστικά		
<i>ABCIndex</i>	Δείκτης συνδεσιμότητας ατόμου-δεσμού	2
<i>AcidBase</i>	Μετρητής όξινων ομάδων	2
<i>AjacencyMatrix</i>	Πίνακας γειτνίασης	13
<i>Aromatic</i>	Μετρητής αρωματικών ατόμων	2
<i>AtomCount</i>	Μετρητής ατόμων	16
<i>Autocorrelation</i>	Αυτοσυσχέτιση τοπολογικής δομής (ή αυτοσυσχέτιση <i>Moreau – Broto</i>)	606
<i>BCUT</i>	Επέκταση του πίνακα <i>Burden</i>	24
<i>BalabanJ</i>	Δείκτης <i>J Balaban</i> (προκύπτει από τη θεώρηση του μορίου ως μοριακός γράφος)	1
<i>BaryszMatrix</i>	Πίνακας <i>Barysz</i> (προκύπτει από τη θεώρηση του μορίου ως σταθμισμένος μοριακός γράφος)	104
<i>BertzCT</i>	Τοπολογικός δείκτης για την ποσοτικοποίηση της "περιπλοκότητας" των μορίων)	1
<i>Bondcount</i>	Μετρητής δεσμών	9
<i>CarbonTypes</i>	Μεταβλητές σχετικά με τον υβριδισμό των ανθράκων)	10
<i>Chi</i>	Μοριακή συνδεσιμότητα <i>X</i>	56
<i>Constitutional</i>	Καταστατικές περιγραφικές μεταβλητές <i>X</i> των ατόμων (π.χ. μοριακό βάρος)	16
<i>DetourMatrix</i>	Πίνακας <i>Detour</i> (προκύπτει από τη θεώρηση του μορίου ως μοριακός γράφος)	14
<i>DistanceMatrix</i>	Πίνακας απόστασης (προκύπτει από τη θεώρηση του μορίου ως μοριακός γράφος)	13
<i>EState</i>	Ηλεκτρο-τοπολογική κατάσταση ατόμων	316
<i>EccentricConnectivity Index</i>	Δείκτης που προκύπτει από την εκκεντρότητα των ατόμων	1
<i>ExtendedTopochemical Atom</i>	Ηλεκτρο-τοπολογική κατάσταση των ατόμων	45
<i>FragmentComplexity</i>	Δείκτης πολυπλοκότητας τεμαχίων	1
<i>Framework</i>	Λόγος μοριακού πλαισίου	1
<i>HydrogenBond</i>	Μετρητής ατόμων που σχηματίζουν δεσμούς υδρογόνου	2
<i>InformationContent</i>	Πληροφορίες γειτονικών ατόμων (προκύπτουν από τη θεώρηση του μορίου ως μοριακός γράφος)	42
<i>KappaShapeIndex</i>	Δείκτες μοριακού σχήματος <i>K</i> (προκύπτει από τη θεώρηση του μορίου ως μοριακός γράφος)	3
<i>Lipinski</i>	Κανόνας του 5 του <i>Lipinski</i>	2
<i>McGowanVolume</i>	Μοριακός όγκος <i>McGowan</i>	1
<i>MoeType</i>	Μεταβλητές τύπου <i>MolecularOperatingEnvironment</i>	53

	για υπολογισμούς εμβαδού επιφάνειας μορίου	
<i>MolecularDistance Edge</i>	Διάνυσμα μοριακής απόστασης-γωνίας	19
<i>PathCount</i>	Μετρητής μονοπατιών κατά μήκος του μοριακού γράφου	21
<i>Polarizability</i>	Ικανότητα πόλωσης του μορίου	2
<i>RingCount</i>	Μετρητής δακτυλίων του μορίου	138
<i>RotatableBond</i>	Μετρητής δεσμών που επιτρέπουν ελεύθερη περιστροφή	2
<i>SLogP</i>	Συντελεστής μερισμού <i>Wildman – Crippen</i>	2
<i>TopoPSA</i>	Εμβαδόν πολικής επιφάνειας μορίου	2
<i>TopologicalCharge</i>	Τοπολογικό φορτίο	21
<i>TopologicalCharge</i>	Τοπολογικοί δείκτες	4
<i>VdwVolumeABC</i>	Μοριακός όγκος <i>VanderWaals</i>	1
<i>VertexAdjacency Information</i>	Πληροφορίες γειτνίασης κόμβων (προκύπτουν από τη θεώρηση του μορίου ως μοριακός γράφος)	1
<i>WalkCount</i>	Μετρητής περιπάτων (προκύπτουν από τη θεώρηση του μορίου ως μοριακός γράφος)	21
<i>Weight</i>	Μοριακό βάρος	2
<i>WienerIndex</i>	Δείκτης <i>Wiener</i> (προκύπτει από τη θεώρηση του μορίου ως μοριακός γράφος)	2
<i>ZagrebIndex</i>	Δείκτης <i>Zagreb</i> (προκύπτει από τη θεώρηση του μορίου ως μοριακός γράφος)	2
3D Δομικά Χαρακτηριστικά		
<i>CPSA</i>	Μερικό εμβαδόν αρνητικά φορτισμένης επιφάνειας μορίου	43
<i>GeometricalIndex</i>	Γεωμετρική διάμετρος μορίου	4
<i>GravitationallIndex</i>	Χαρακτηρισμός κατανομής μάζας του μορίου	4
<i>MoRSE</i>	Μεταβλητές που σχετίζονται με την 3D αναπαράσταση μοριακών δομών βασισμένη σε περίθλαση ηλεκτρονίων	160
<i>MomentOfInertia</i>	Ροπή αδράνειας μορίου	3

Για τη χρήση του λογισμικού *mordred* χρησιμοποιήθηκε η αντίστοιχη βιβλιοθήκη της Python 3.7.11 [76], καθώς και η βιβλιοθήκη *rdkit* [77]. Συνολικά, παρήχθησαν 1613 μοριακές περιγραφικές μεταβλητές για καθεμία από τις 1457 χημικές ενώσεις του συνόλου δεδομένων. Επομένως, ο πίνακας σχεδιασμού **X** έχει μέγεθος 1457×1613 .

Σημειώνεται ότι πριν την εισαγωγή του στο μοντέλο νευρωνικών δικτύων MLP ο πίνακας σχεδιασμού υφίσταται κανονικοποίηση. Η κανονικοποίηση αποτελεί μια μέθοδο προεπεξεργασίας των δεδομένων εισόδου. Με την εφαρμογή της, αποτρέπονται προβλήματα που αφορούν την απόδοση των αλγορίθμων βελτιστοποίησης. Ειδικότερα, όταν τα δεδομένα εισόδου είναι σε διαφορετική κλίμακα, τότε στην ίδια κατάσταση θα βρίσκονται και οι εκπαιδευσιμες παράμετροι (βάρη και *biases*) του δικτύου που σχετίζονται με τα αντίστοιχα δεδομένα εισόδου. Κάτι τέτοιο μπορεί να οδηγήσει σε μια "αδέξια" τοπολογία της συνολικής συνάρτησης κόστους, με αποτέλεσμα ο αλγόριθμος βελτιστοποίησης να δίνει έμφαση

για τον υπολογισμό της παραγώγου σε συγκεκριμένες παραμέτρους. Κανονικοποιώντας τα δεδομένα εισόδου σε μια συγκεκριμένη κλίμακα, βελτιώνεται αρκετά η διαδικασία εκπαίδευσής του μέσω ενός αλγορίθμου βελτιστοποίησης βασισμένου στην μέθοδο gradient descent [83]. Στην συγκεκριμένη περίπτωση η κανονικοποίηση ενός στοιχείου x του πίνακα σχεδιασμού \mathbf{X} λαμβάνει χώρα ως εξής:

$$z = \frac{x - u}{s} \quad (6.3)$$

όπου z το κανονικοποιημένο στοιχείο, u και s ο μέσος όρος και η τυπική απόκλιση των στοιχείων του πίνακα \mathbf{X} αντίστοιχα.

Η κανονικοποίηση των δεδομένων εισόδου γίνεται με χρήση του πακέτου μηχανικής μάθησης scikit-learn [81].

6.4.2 Ανάπτυξη Μοντέλου (Model Development)

Η εκπαίδευση του μοντέλου εμπρόσθιων νευρωνικών δικτύων λαμβάνει χώρα με χρήση της βιβλιοθήκης της Python Keras [78] του πακέτου μηχανικής μάθησης Tensorflow [79]. Όλες οι προσομοιώσεις γίνονται στο εργαλείο εκτέλεσης προγραμμάτων Python Jupyter Notebook ή στο αντίστοιχο διαδικτυακό notebook της Google, Google Colaboratory. Τα διαγράμματα ακρίβειας πρόβλεψης συναρτήσει χρονικών βημάτων παράγονται μέσω του εργαλείου οπτικοποίησης του tensorflow, tensorboard.

Σε συμφωνία με την θεωρητική παρουσίαση του κεφαλαίου 1, το μοντέλο MLP εκπαιδεύεται με χρήση του αλγορίθμου οπισθοδιάδοσης (backpropagation) και κατάλληλη επιλογή αλγορίθμου βελτιστοποίησης (βλ. υποκεφάλαια 1.3.1, 1.2.2). Σε κάθε επανάληψη (ή εποχή), επιλέγεται η ανανέωση των παραμέτρων του συστήματος (βάρη \mathbf{w} και biases \mathbf{b}) να γίνεται με οπισθοδιάδοση του μέσου όρου μιας παρτίδας (batch) παραδειγμάτων. Με αυτόν τον τρόπο, επιτυγχάνεται καλύτερη ακρίβεια στην διαδικασία υπολογισμού των νέων παραμέτρων χωρίς ιδιαίτερη υπολογιστική επιβάρυνση. Το μέγεθος της παρτίδας (batch size) αποτελεί μια υπερπαραμέτρο του συστήματος, δηλαδή μια παράμετρο που ορίζεται από τον χρήστη πριν την έναρξη της εκπαίδευσης του μοντέλου. Τόσο αυτή, όσο και οι υπόλοιπες υπερπαραμέτροι πρέπει να καθοριστούν με τέτοιο τρόπο ώστε το μοντέλο να παράξει την καλύτερη δυνατή πρόβλεψη. Οι υπερπαραμέτροι του μοντέλου MLP μαζί με τις τιμές που μπορούν αυτές να λάβουν φαίνονται στον παρακάτω πίνακα:

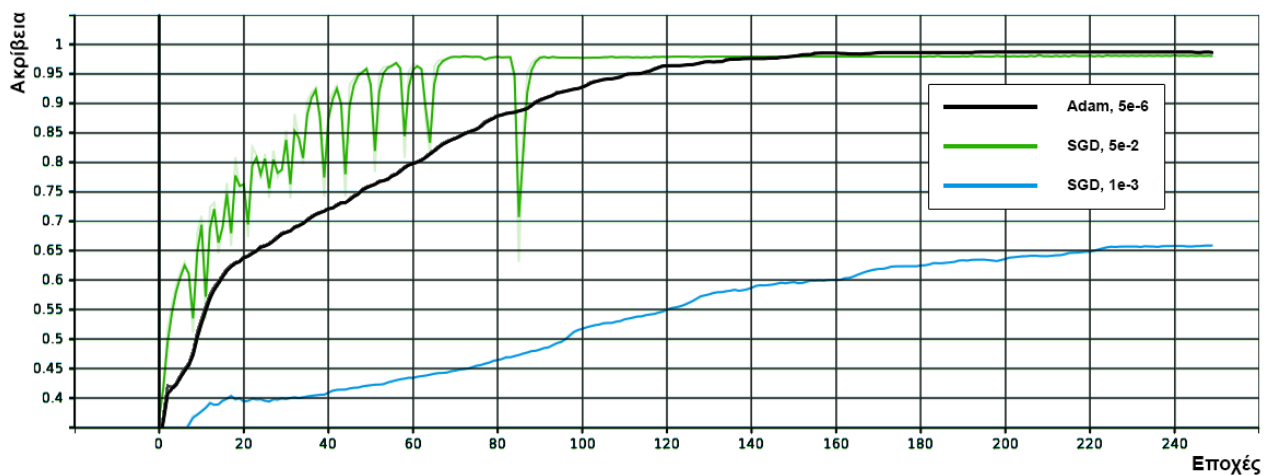
Πίνακας 6.3: Εύρος τιμών των υπερπαραμέτρων του μοντέλου MLP

Υπερπαραμέτρος	Εύρος τιμών
Αλγόριθμος βελτιστοποίησης	{Adam, SGD}
Μέγεθος παρτίδας	{64, 128, 256}
Συνάρτηση κόστους	Διασταυρούμενη εντροπία
Ρυθμός μάθησης	$[10^{-6}, 10^{-2}]$
Συνάρτηση ενεργοποίησης	{relu, σιγμοειδής, tanh}
Αριθμός κρυφών στρωμάτων	[1, 10]
Αριθμός κρυφών νευρώνων	[100, 1200]

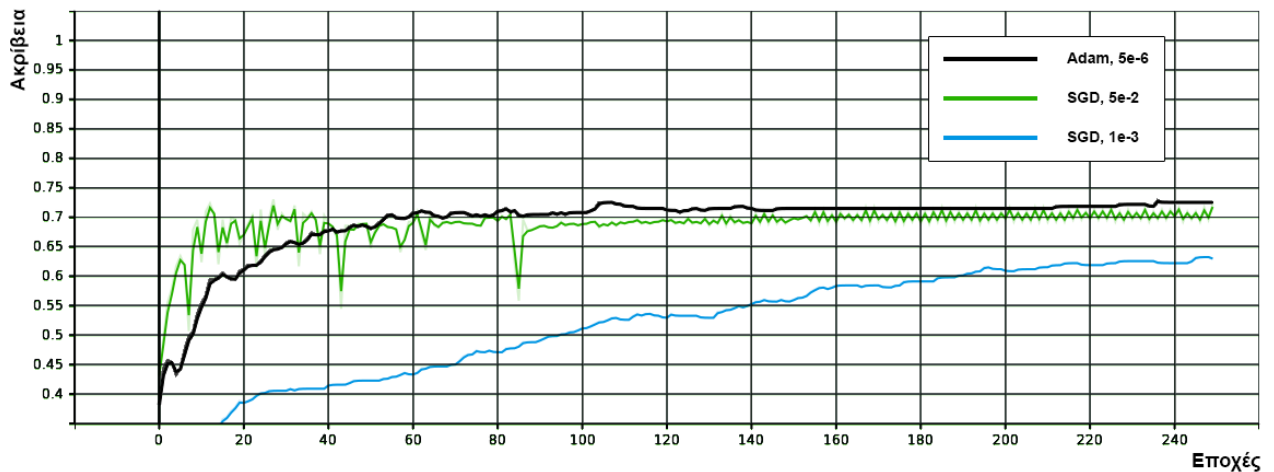
Όπως διακρίνεται στον πίνακα 6.10, ως συνάρτηση κόστους επιλέγεται το κόστος διασταυρούμενης εντροπίας (cross-entropy loss). Για την υπερπαραμέτρο αυτή δεν θεωρήθηκε αναγκαίο να γίνει δοκιμή κάποιας άλλης συνάρτησης κόστους, αφού η διασταυρούμενη

εντροπία χρησιμοποιείται ευρύτατα και με καλά αποτελέσματα σε προβλήματα ταξινόμησης μέσω νευρωνικών δικτύων. Μάλιστα, στο παρόν πρόβλημα εφαρμόζεται η επέκταση της διασταυρούμενης εντροπίας για 3 κλάσεις, η οποία στη γενική της μορφή συναντάται και ως κατηγορική διασταυρούμενη εντροπία (categorical cross-entropy) (για θεωρία βλ. παράγραφο 1.3.2).

Η επιλογή αλγορίθμου βελτιστοποίησης είναι ιδιαίτερα σημαντική για την σωστή εκπαίδευση του νευρωνικού δικτύου. Ειδικότερα, με μια λανθασμένη επιλογή μπορεί να προκύψουν σοβαρά προβλήματα, όπως αδυναμία σύγκλισης, εξαφανιζόμενες βαθμίδες ή και υπερεκπαίδευση του μοντέλου. Όπως περιγράφηκε στο υποκεφάλαιο 1.2.2 ο θεμελιώδης αλγόριθμος βελτιστοποίησης των νευρωνικών δικτύων είναι ο Gradient Descent. Μια παραλλαγή του, ο αλγόριθμος Adam αποτελεί το state-of-the-art στα μοντέλα βαθιάς μάθησης και έτσι, επιλέγεται ως υποψήφιος αλγόριθμος του παρόντος μοντέλου MLP. Για την σύγκριση των ανωτέρω αλγορίθμων, εκπαιδεύεται το ίδιο μοντέλο με χρήση των δύο αλγορίθμων για διάφορους ρυθμούς μάθησης. Συγκεκριμένα, παρακάτω παρατίθενται σε 2 διαγράμματα τα αποτελέσματα ακρίβειας συνόλου εκπαίδευσης και ακρίβειας συνόλου επαλήθευσης για μοντέλο 7 κρυφών στρωμάτων για εκπαίδευση με αλγόριθμο Adam και ρυθμό μάθησης $5 \cdot 10^{-6}$ και για εκπαίδευση με αλγόριθμο SGD και ρυθμούς μάθησης 10^{-3} , $5 \cdot 10^{-2}$:



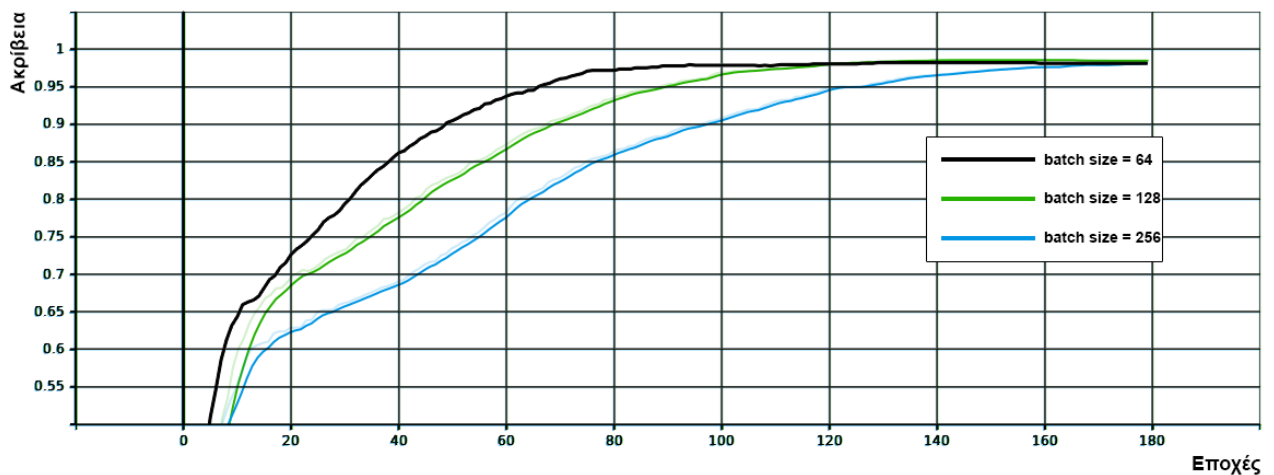
Σχήμα 6.5: Ακρίβεια πρόβλεψης στο σύνολο δεδομένων εκπαίδευσης για εκπαίδευση με αλγόριθμο Adam και ρυθμό μάθησης $5 \cdot 10^{-6}$ και για εκπαίδευση με αλγόριθμο SGD και ρυθμούς μάθησης 10^{-3} , $5 \cdot 10^{-2}$



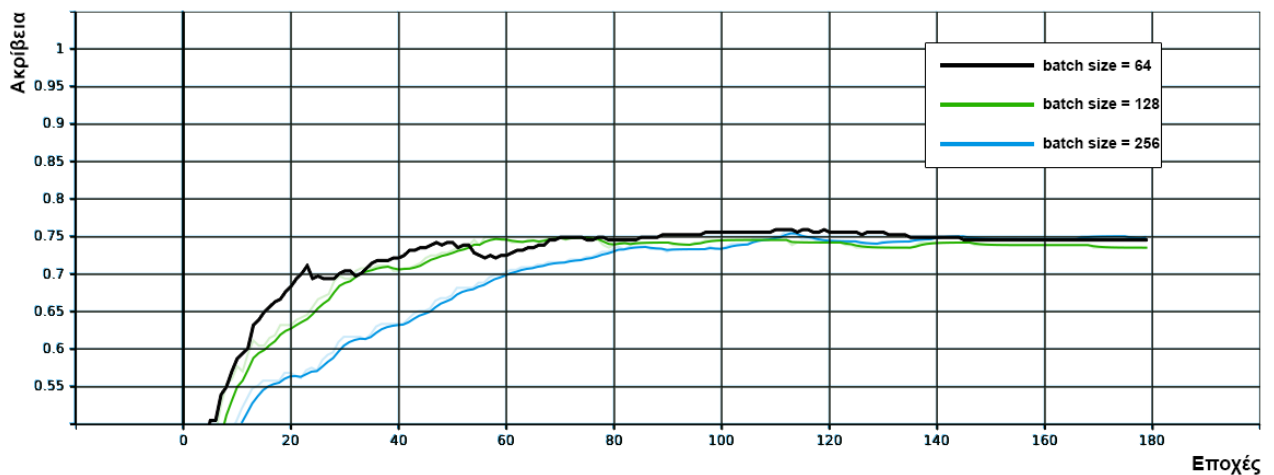
Σχήμα 6.6: Ακρίβεια πρόβλεψης στο σύνολο δεδομένων επαλήθευσης για εκπαίδευση με αλγόριθμο Adam και ρυθμό μάθησης $5 \cdot 10^{-6}$ και για εκπαίδευση με αλγόριθμο SGD και ρυθμούς μάθησης 10^{-3} , $5 \cdot 10^{-2}$

Παρατηρείται ότι το μοντέλο με χρήση του αλγορίθμου SGD για ρυθμό μάθησης ίσο με 10^{-3} εκπαιδεύεται πολύ αργά ενώ η καμπύλη ακρίβειας των δεδομένων εκπαίδευσης τείνει να παρουσιάσει μέγιστο, γεγονός που υποδηλώνει ότι ο αλγόριθμος θα συγκλίνει σε χαμηλές τιμές ακρίβειας. Από την άλλη, για ρυθμό μάθησης ίσο με $5 \cdot 10^{-2}$, το μοντέλο εκπαιδεύεται μεν σε ικανοποιητικές τιμές ακρίβειας στα δεδομένα εκπαίδευσης, παρουσιάζοντας, όμως, μεγάλη αστάθεια ειδικά στο ξεκίνημα της εκπαίδευσής του. Ακόμη, διαφαίνεται αστάθεια στην ικανότητα πρόβλεψης στα δεδομένα επαλήθευσης σε όλο το εύρος των εποχών. Με την χρήση του Adam, το μοντέλο εκπαιδεύεται σταθερότερα προσεγγίζοντας μεγάλες τιμές ακρίβειας στα δεδομένα εκπαίδευσης, ενώ υπερνικά σε ικανότητα πρόβλεψης στα δεδομένα επαλήθευσης και τις δύο περιπτώσεις χρήσης SGD. Η πολύ χαμηλή τιμή ρυθμού μάθησης στον Adam συγκριτικά με τον SGD έχει να κάνει με την φύση του αλγορίθμου και την μεγαλύτερη ευκολία σύγκλισης (περισσότερες θεωρητικές λεπτομέρειες στο υποκεφάλαιο 1.2.2). Με βάση τα παραπάνω, για την εκπαίδευση του μοντέλου MLP επιλέγεται ο αλγόριθμος βελτιστοποίησης Adam.

Με ανάλογο τρόπο αποφασίζεται και η υπερπαράμετρος του μεγέθους παρτίδας (batch size). Στα παρακάτω διαγράμματα παρουσιάζεται η ακρίβεια πρόβλεψης στα δεδομένα εκπαίδευσης και επαλήθευσης για εκπαίδευση μοντέλου με τα 3 διαφορετικά μεγέθη παρτίδας του πίνακα 6.10:



Σχήμα 6.7: Μοντέλο MLP: Ακρίβεια πρόβλεψης στο σύνολο δεδομένων εκπαίδευσης για εκπαίδευση με μεγέθη παρτίδας 64, 128, 256



Σχήμα 6.8: Μοντέλο MLP: Ακρίβεια πρόβλεψης στο σύνολο δεδομένων επαλήθευσης για εκπαίδευση με μεγέθη παρτίδας 64, 128, 256

Παρατηρείται ότι όσο μικρότερο το μέγεθος παρτίδας τόσο πιο γρήγορα συγκλίνει σε υψηλές τιμές ακρίβειας στα δεδομένα εκπαίδευσης το μοντέλο. Παρ' όλα αυτά, η ακρίβεια στα δεδομένα επαλήθευσης δεν αλλάζει όσον αφορά την μέγιστη τιμή της που είναι και το ζητούμενο του προβλήματος. Η επιλογή ενός υψηλού μεγέθους παρτίδας μπορεί να οδηγήσει σε μεγάλους υπολογιστικούς χρόνους. Στην περίπτωση του συγκεκριμένου προβλήματος MLP, όμως, ο όγκος των δεδομένων προς επεξεργασία με βάση τους διαθέσιμους πόρους (Google Colaboratory) δεν επιβαρύνει, ιδιαίτερα, τους χρόνους της προσομοίωσης. Συμπερασματικά, λοιπόν, το μέγεθος παρτίδας στο επιλεγμένο εύρος (το οποίο είναι το σύνηθες στη βιβλιογραφία) δεν κρίνεται ως σημαντική υπερπαραμέτρος για την απόδοση του μοντέλου. Για το παρόν μοντέλο, επιλέγεται μέγεθος παρτίδας 256, αφού η επεξεργασία των δεδομένων για περισσότερα χρονικά βήματα μπορεί τελικά να αποφέρει καλύτερα αποτελέσματα, τη στιγμή που οι άλλες μικρότερες τιμές πιθανόν να εμφάνιζαν υπερεκπαίδευση.

Ο καθορισμός των υπολειπόμενων υπερπαραμέτρων, δηλαδή του ρυθμού μάθησης, της συνάρτησης ενεργοποίησης, του αριθμού κρυφών στρωμάτων και του αριθμού νευρώνων σε κάθε κρυφό στρώμα, γίνεται με εφαρμογή μπεϋζιανής βελτιστοποίησης με γκαουσιανή

διαδικασία παλινδρόμησης (για θεωρητικές λεπτομέρειες βλέπε παράρτημα Β'). Το χαμηλό υπολογιστικό κόστος των συγκεκριμένων μοντέλων MLP επιτρέπουν εφαρμογές αλγορίθμων βελτιστοποίησης υπερπαραμέτρων, όπως η μπεϋζιανή βελτιστοποίηση, σε αντίθεση με βαθύτερα δίκτυα ή δίκτυα με μεγαλύτερο όγκο δεδομένων εισόδου, όπως συνελκτικά δίκτυα με εικόνες, τα οποία βελτιστοποιούνται κυρίως με εμπειρικό τρόπο ή βασίζονται σε αναγνωρισμένες για την απόδοσή τους αρχιτεκτονικές. Παρ' όλα αυτά, για να καθοριστούν τα σχετικά εύρη επιτρεπτών τιμών έγιναν δοκιμές με διάφορες αρχιτεκτονικές MLP. Συγκεκριμένα, για ρυθμό μάθησης μεγαλύτερο από 10^{-2} , ο αλγόριθμος βελτιστοποίησης εμφάνιζε απόκλιση με αδυναμία εκπαίδευσης του μοντέλου, ενώ για τιμές μικρότερες από 10^{-6} η εκπαίδευση του μοντέλου ήταν αρκετά αργή. Ως συναρτήσεις ενεργοποίησης επιλέχθηκαν οι ευρέως χρησιμοποιούμενες σιγμοειδής, υπερβολική εφαπτομένη και relu (βλ. υποκεφάλαιο 1.3.3). Η χρήση της prelu εμφάνιζε παρόμοια αποτελέσματα με την συνάρτηση relu και αποφασίστηκε να μην συμπεριληφθεί στο εύρος τιμών. Αναφορικά με τον αριθμό των κρυφών νευρώνων αποφασίστηκε να τεθεί το όριο, ώστε το δίκτυο να μην γίνει άσκοπα βαθύ για το συγκεκριμένο σύνολο δεδομένων, κάτι ενδεχομένως που θα προκαλούσε υπερεκπαίδευση και θα αύξανε την υπολογιστική επιβάρυνση. Τέλος, το εύρος [100,1200] για το αριθμό νευρώνων σε κάθε κρυφό στρώμα κρίθηκε λογικό με βάση τη διάσταση των στρώματος εισόδου που ισούται με 1613 νευρώνες.

Για την βέλτιστη επιλογή των παραπάνω υπερπαραμέτρων, λοιπόν, χρησιμοποιήθηκε η βιβλιοθήκη scikit-optimize [80] που βασίζεται στο πακέτο μηχανικής μάθησης sci-kit [81] σε συνδυασμό με τη βιβλιοθήκη keras. Ο αλγόριθμος της γκαουσιανής βελτιστοποίησης έτρεξε για 200 επαναλήψεις σε GPU του Google Colaboratory (NVIDIA Tesla K80 12 GB) αναζητώντας την καλύτερη ακρίβεια στα δεδομένα επαλήθευσης στο εύρος τιμών υπερπαραμέτρων που φαίνονται στον πίνακα 6.10. Ο υπολογιστικός χρόνος ανήλθε στις 2 ώρες και 13 λεπτά. Τα αποτελέσματα ήταν τα ακόλουθα:

- Ρυθμός μάθησης = $4.5 \cdot 10^{-6}$
- Αριθμός κρυφών στρωμάτων = 6
- Αριθμός νευρώνων σε κάθε κρυφό στρώμα = 650
- Συνάρτηση ενεργοποίησης = 'relu'

Συνοπτικά, οι υπερπαραμέτροι που καθορίζουν το μοντέλο ταξινόμησης MLP φαίνονται στο παρακάτω πίνακα:

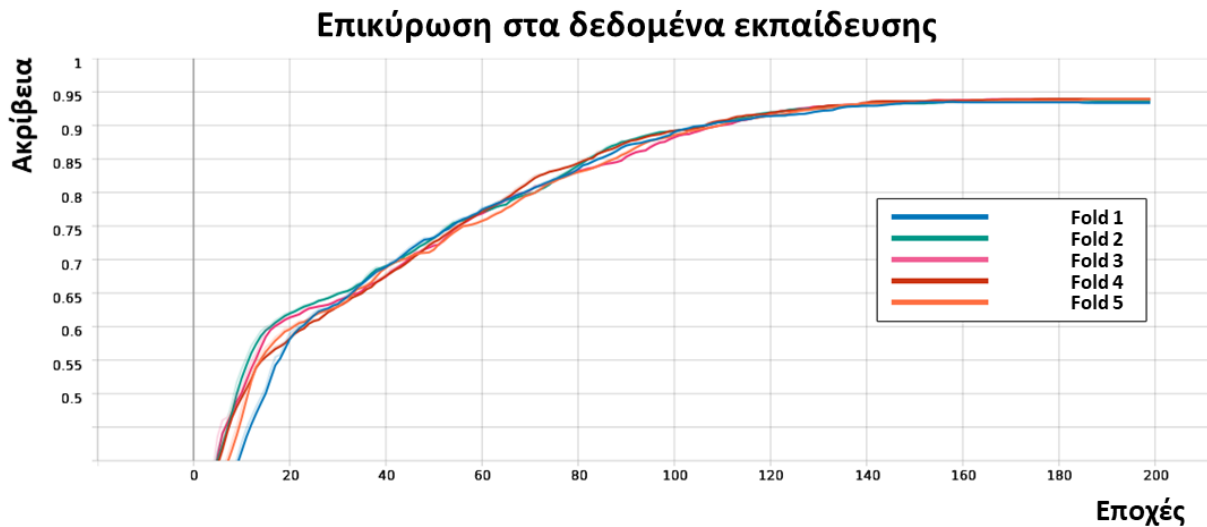
Πίνακας 6.4: Οι τελικές υπερπαραμέτροι του μοντέλου MLP

Υπερπαραμέτρος	Τιμή
Αλγόριθμος βελτιστοποίησης	<i>Adam</i>
Μέγεθος παρτίδας	256
Συνάρτηση κόστους	Διασταυρούμενη εντροπία
Ρυθμός μάθησης	$4.5 \cdot 10^{-6}$
Συνάρτηση ενεργοποίησης	<i>relu</i>
Αριθμός κρυφών στρωμάτων	6
Αριθμός κρυφών νευρώνων	650

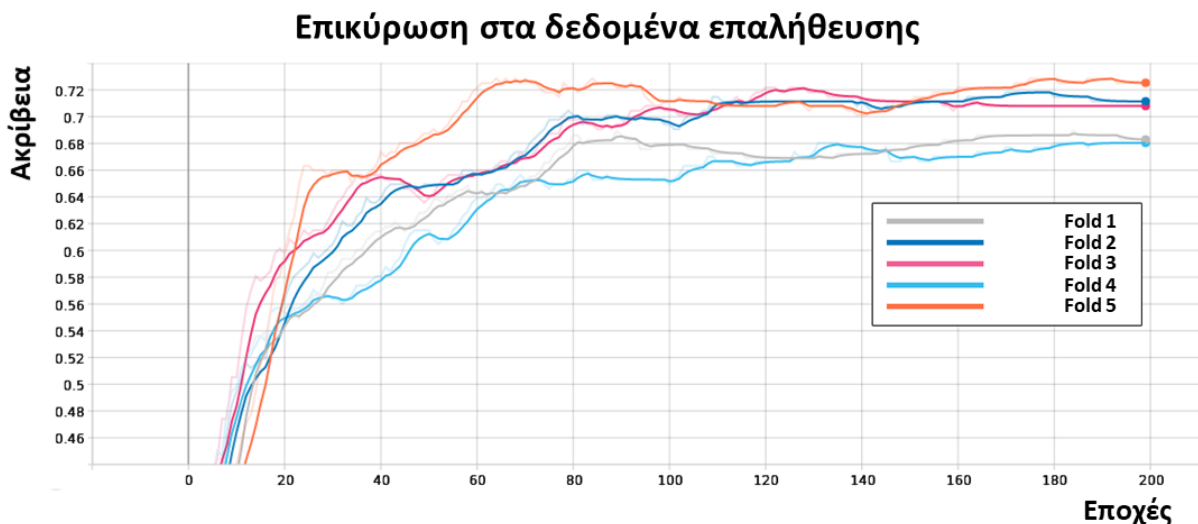
6.4.3 Αποτελέσματα (Results)

Το μοντέλο νευρωνικών δικτύων εμπρόσθιας τροφοδότησης, εκπαιδεύτηκε για 200 χρονικά βήματα (εποχές). Η αρχιτεκτονική και οι ρυθμίσεις του μοντέλου καθορίζεται από τις

τιμές υπερπαραμέτρων του πίνακα 6.8. Για την επικύρωση της απόδοσης του μοντέλου ακολουθείται η διαδικασία της διασταυρούμενης επικύρωσης, όπως περιγράφηκε στο υποκεφάλαιο 6.2. Ειδικότερα, επιλέχθηκε 5-fold cross validation. Επομένως, το σύνολο των δεδομένων, όπως ορίζεται από τον πίνακα σχεδιασμού $\mathbf{X} \in \mathbb{R}^{1457 \times 1613}$, χωρίζεται με τυχαίο τρόπο σε 5 ομάδες ίσου μεγέθους. Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε Intel Xeon CPU @ 2 GHz (Google Colaboratory) και οι υπολογιστικοί χρόνοι για εκπαίδευση 200 εποχών ανήλθαν σε περίπου 2,5 λεπτά. Ακολουθούν διαγράμματα ακρίβειας συναρτήσεϊ εποχών τόσο για επικύρωση στα δεδομένα εκπαίδευσης όσο και για επικύρωση στα δεδομένα επαλήθευσης με ταυτόχρονη απεικόνιση και των 5 folds που προέκυψαν από το cross-validation:



Σχήμα 6.9: Μοντέλο MLP: αποτελέσματα 5-fold cross validation ακρίβειας συναρτήσεϊ εποχών στα δεδομένα εκπαίδευσης



Σχήμα 6.10: Μοντέλο MLP: αποτελέσματα 5-fold cross validation ακρίβειας συναρτήσεϊ εποχών στα δεδομένα επαλήθευσης

Οι συγκεντρωτικές τιμές των μέτρων επικύρωσης στα δεδομένα επαλήθευσης, ακρίβειας

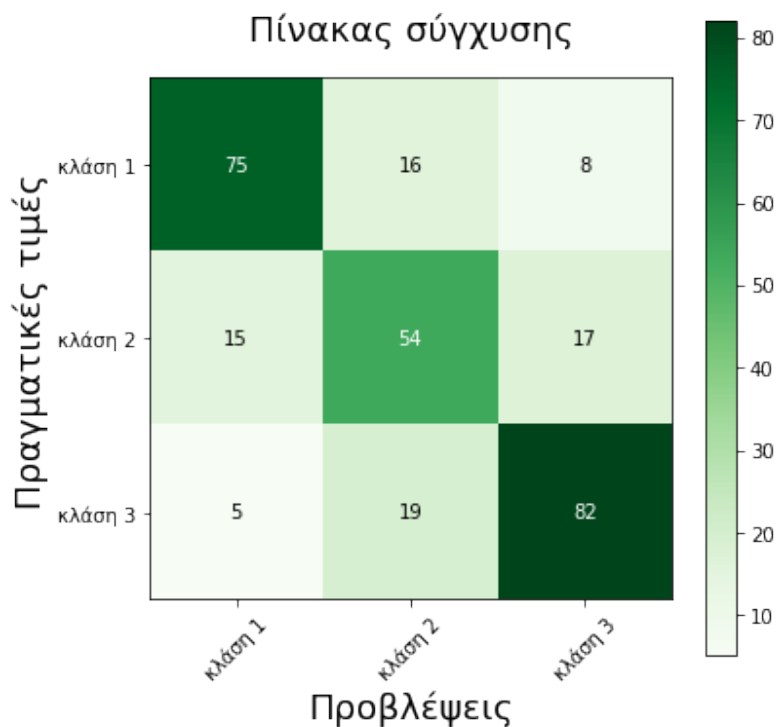
και MCC, μετά την εφαρμογή του 5-fold cross-validation είναι οι ακόλουθες:

Πίνακας 6.5: Αποτελέσματα ακρίβειας και MCC του μοντέλου MLP

	Ακρίβεια	MCC
<i>Fold 1</i>	0.69	0.53
<i>Fold 2</i>	0.72	0.58
<i>Fold 3</i>	0.71	0.57
<i>Fold 4</i>	0.68	0.52
<i>Fold 5</i>	0.73	0.59
Σύνολο	0.71 ± 0.02	0.56 ± 0.03

Ένας ενδεικτικός πίνακας σύγχυσης που προέκυψε από το 5ο fold του cross-validation είναι ο εξής:

Πίνακας 6.6: Πίνακας σύγχυσης του μοντέλου MLP



Όπως φαίνεται στο σχήμα 6.9, το μοντέλο MLP εκπαιδεύεται επιτυχώς στα δεδομένα εκπαίδευσης. Και τα 5 folds του cross validation παρουσιάζουν παρόμοια συμπεριφορά εκπαίδευσης και προσεγγίζουν ακρίβεια στα δεδομένα εκπαίδευσης που αγγίζει το 0.95. Η διαφορά μεταξύ των folds έγκειται στην επικύρωση του μοντέλου με βάση τα δεδομένα επαλήθευσης. Η ακρίβεια στα δεδομένα επαλήθευσης διαφέρει από fold σε fold, γεγονός που αναδεικνύει την σημασία εφαρμογής της διαδικασίας cross validation στην εξαγωγή των αποτελεσμάτων ενός μοντέλου μηχανικής μάθησης. Η τυχαιότητα, δηλαδή, μεταξύ του διαχωρισμού του συνόλου δεδομένων σε δεδομένα εκπαίδευσης και δεδομένα επαλήθευσης μπορεί να επιφέρει μεγάλες διαφορές στις μετρικές τιμές απόδοσης με κάθε τρέξιμο του μοντέλου. Στην συγκεκριμένη περίπτωση, ο μέσος όρος των τιμών ακρίβειας στα δεδομένα επαλήθευσης ανέρχεται σε 0.71 με τυπική απόκλιση ίση με 0.02, ενώ για τους συντελεστές MCC σε 0.56 με τυπική απόκλιση 0.03.

Τόσο ο συντελεστής MCC όσο και ο πίνακας σύγχυσης δίνουν σημαντικές πληροφορίες για την ικανότητα πρόβλεψης του μοντέλου σε επίπεδο κλάσης και μπορούν να οδηγήσουν σε ασφαλέστερα και σιβαρότερα συμπεράσματα σε σύγκριση με την "φτωχότερη" πληροφορία που προσφέρει η ακρίβεια. Όπως παρατηρείται στο σχήμα 6.6, το μοντέλο αποδίδει χειρότερα στην κλάση 2, δηλαδή σε ενδιάμεσες τιμές της συγγένειας πρόσδεσης ενδοκρινικών διαταρακτών (logRBA). Περισσότερο διαφωτιστικός είναι ο κανονικοποιημένος πίνακας σύγχυσης που προκύπτει με την κανονικοποίηση κάθε σειράς του πίνακα σύγχυσης ως προς το σύνολο των προβλέψεων ανά κλάση :

Πίνακας 6.7: Κανονικοποιημένος πίνακας σύγχυσης του μοντέλου MLP



Σύμφωνα με το σχήμα 6.7, λοιπόν, η ικανότητα πρόβλεψης στις κλάσεις 1 και 3 είναι, σχεδόν, όμοια με την ακρίβεια να αγγίζει το 76% και 77% αντίστοιχα. Η ακρίβεια, όμως, στην κλάση 2 είναι αρκετά μικρότερη και συγκεκριμένα, 63%. Τα παραπάνω δείχνουν ότι το μοντέλο δυσκολεύεται να ταξινομήσει ουσίες με μέτρια συγγένεια πρόσδεσης στους ERs, όπως αυτή ορίστηκε στον πίνακα 5.1. Η συμπεριφορά αυτή είναι αναμενόμενη αφού στα όρια των κλάσεων είναι δυνατό να χάνεται ο σαφής διαχωρισμός μεταξύ των κλάσεων ως προς το δυναμικό πρόσδεσης των ουσιών. Από την άλλη, όμως, η ικανότητα του μοντέλου στην διάκριση των ενδοκρινικών διαταρακτών σε ασθενείς και ισχυρούς (κλάσεις 1 και 3 αντίστοιχα) είναι σημαντική για πρακτικούς λόγους, αφού προβλέπεται με ικανοποιητική ακρίβεια πόσο επικίνδυνη δύναται να καταστεί μια ουσία για το ενδοκρινικό σύστημα και κατ' επέκταση την υγεία και φυσιολογική λειτουργία του ανθρώπινου οργανισμού.

6.5 Μοντέλο Νευρωνικών Δικτύων Γράφων (Graph Neural Networks Model)

6.5.1 Κατασκευή Δεδομένων Εισόδου (Input Data Construction)

Είσοδο στα νευρωνικά δίκτυα γράφων αποτελούν, συνήθως, γράφοι με χαρακτηριστικά (attributed graphs), οι πληροφορίες των οποίων εκφράζονται μέσω του πίνακα γειννίασης $A \in \mathbb{R}^{N \times N}$, του πίνακα χαρακτηριστικών κόμβων $X \in \mathbb{R}^{N \times D}$ και του πίνακα χαρακτηριστικών ακμών $E \in \mathbb{R}^{M \times F}$, όπου N ο αριθμός των κόμβων, D ο αριθμός των χαρακτηριστικών του κάθε κόμβου, M ο αριθμός των ακμών και F ο αριθμός των χαρακτηριστικών των ακμών. Στο παρόν πρόβλημα, ζητούμενο είναι η ταξινόμηση χημικών μορίων. Επομένως, είσοδο στο μοντέλο αποτελεί χημικός γράφος, του οποίου οι κόμβοι αντιστοιχούν στα άτομα της χημικής ένωσης (εκτός υδρογόνων) και οι ακμές στους δεσμούς της. Επειδή ο τύπος GNN που χρησιμοποιείται είναι το συνελκτικό δίκτυο γράφων (GCN), το οποίο δεν λαμβάνει υπόψιν τα χαρακτηριστικά των ακμών, ζητούμενο είναι η κατασκευή του χημικού γράφου με πίνακα γειννίασης και χαρακτηριστικά κόμβων.

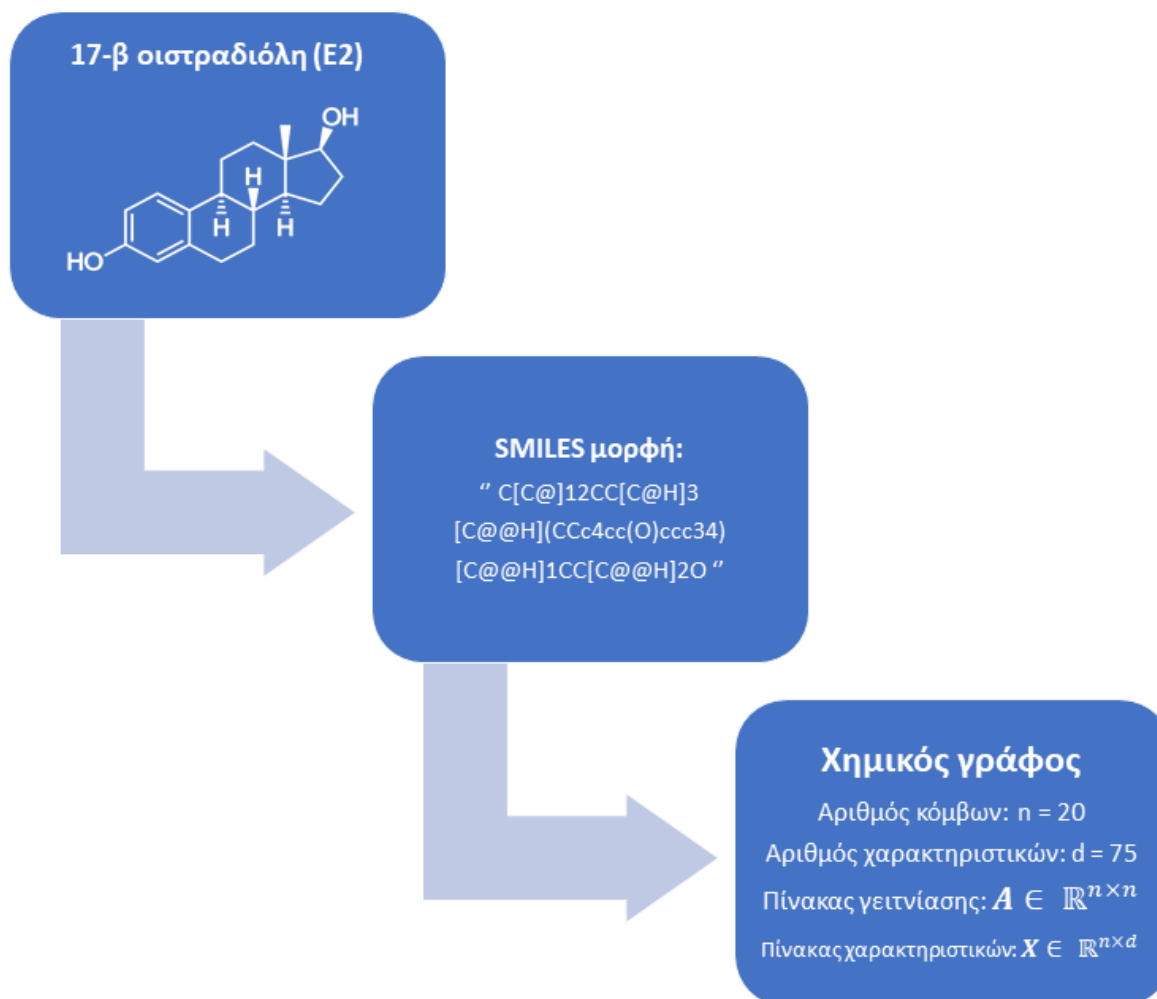
Και σε αυτήν την περίπτωση η είσοδος του μοντέλου προκύπτει από τις χημικές ενώσεις του συνόλου δεδομένων αποτυπωμένες στην γραφή SMILES. Η μετατροπή των SMILES σε χημικούς γράφους γίνεται μέσω της βιβλιοθήκης βαθιάς μάθησης για εφαρμογές χημειοπληροφορικής και βιοπληροφορικής `deerpchem` [82] και της βιβλιοθήκης `rdkit` [77]. Οι κόμβοι/άτομα χαρακτηρίζονται μέσω αριθμητικών, αλλά και ποιοτικών μεταβλητών, οι οποίες μετατρέπονται σε αριθμητικές μέσω `one-hot` κωδικοποίησης (`one-hot encoding`). Σύμφωνα με την μέθοδο της `one-hot` κωδικοποίησης, για κάθε ποιοτική μεταβλητή (categorical variable) δημιουργείται ένα διάνυσμα από 0 και 1 μήκους ίσου με το πλήθος των συμβολικών τιμών (string values) που μπορεί να λάβει η μεταβλητή αυτή. Η συμβολική τιμή που τελικά εκφράζει την μεταβλητή στον εκάστοτε κόμβο παίρνει την αριθμητική τιμή 1 και όλες οι υπόλοιπες την τιμή 0. Όλα τα διανύσματα από 0 και 1 που προκύπτουν συνδέονται σε σειρά συγκροτώντας το διάνυσμα χαρακτηριστικών του εκάστοτε κόμβου. Σημειώνεται ότι η `one-hot` κωδικοποίηση μπορεί να εφαρμοστεί και σε αριθμητικές μεταβλητές, οι οποίες λαμβάνουν καθορισμένες τιμές (συνήθως συγκεκριμένους ακέραιους αριθμούς). Οι αριθμητικές και οι ποιοτικές μεταβλητές, δηλαδή τα χαρακτηριστικά που εκχωρούνται στους κόμβους/άτομα μαζί με τις τιμές που μπορούν να πάρουν είναι οι ακόλουθες:

Πίνακας 6.8: Χαρακτηριστικά κόμβων/ατόμων του χημικού γράφου

Χαρακτηριστικά ατόμων	Πιθανές τιμές
Χημικό στοιχείο	[<i>C', N', O', S', F', Si', P', Cl', Br', Mg', Na', Ca', Fe', As', Al', I', B', V', K', Tl', Yb', Sb', Sn', Ag', Pd', Co', Se', Tl', Zn', Li', Ge', Cu', Au', Ni', Cd', In', Mn', Zr', Cr', Pt', Hg', Pb', Unknown'</i>]
Αριθμός γειτονικών ατόμων (βαθμός κόμβου χημικού γράφου)	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
Χημικό σθένος	[0, 1, 2, 3, 4, 5, 6]
Αριθμός μονήρων ηλεκτρονίων	Μη-αρνητικός ακέραιος αριθμός
Τυπικό φορτίο	Μη-αρνητικός ακέραιος αριθμός
Υβριδισμός	[<i>SP', SP²', SP³', SP³D, SP³D²</i>]
Αρωματικότητα	Αληθής ή Ψευδής (1 ή 0)

Οι τιμές της δεύτερης στήλης του πίνακα 6.8 που βρίσκονται σε αγκύλες εισέρχονται στο διάλυμα μέσω one-hot κωδικοποίησης, ενώ οι υπόλοιπες (αριθμός μονήρων ηλεκτρονίων, τυπικό φορτίο, αρωματικότητα) εισέρχονται αυτούσιες. Το τελικό διάλυμα χαρακτηριστικών ενός ατόμου έχει συνολικό μήκος 75 στοιχείων.

Ένα παράδειγμα χημικής ένωσης (17-β οιστραδιόλη) που υφίσταται την διαδικασία επεξεργασίας ώστε να αποτελέσει είσοδο στο μοντέλο νευρωνικών δικτύων γράφων βασισμένο στο GCN φαίνεται παρακάτω :



Σχήμα 6.11: Διαδικασία κατασκευής δεδομένων εισόδου μοντέλου νευρωνικών δικτύων γράφων

6.5.2 Ανάπτυξη Μοντέλου (Model Development)

Η εκπαίδευση του μοντέλου νευρωνικών δικτύων γράφων λαμβάνει χώρα με χρήση της βιβλιοθήκης Spektral, η οποία είναι μια εξειδικευμένη βιβλιοθήκη για ανάπτυξη νευρωνικών δικτύων γράφων βασισμένη στο Keras του tensorflow [84]. Όπως και στην περίπτωση του μοντέλου MLP, τα python scripts εκτελούνται στο περιβάλλον Google Colaboratory ή σε Jupyter notebook.

Το μοντέλο GNNs που αναπτύσσεται βασίζεται στο συνελκτικό στρώμα του GCN για την διαδικασία της νευρωνικής ανταλλαγής μηνυμάτων μεταξύ των κόμβων. Όπως περιγράφηκε στο υποκεφάλαιο 3.4, ένα μοντέλο νευρωνικών δικτύων γράφων, το οποίο παράγει προβλέψεις σε επίπεδο γράφων, περιλαμβάνει :

- στρώματα ανταλλαγής μηνυμάτων (ή συνελκτικά στρώματα), όπου οι αναπαραστάσεις των κόμβων ανανεώνονται χρησιμοποιώντας πληροφορίες της γειτονιάς τους με απώτερο σκοπό την συλλογή δομικών πληροφοριών του γράφου (στην συγκεκριμένη περίπτωση, χημικού μορίου).
- ένα στρώμα υποδειγματοληψίας, όπου οι αναπαραστάσεις των κόμβων αθροίζονται με χρήση μιας στατιστικής συνάρτησης για το σκοπό της παραγωγής μιας καθολικής αναπαράστασης του γράφου.
- πλήρως συνδεδεμένα στρώματα (στρώματα MLP), τα οποία είτε προεπεξεργάζονται τα χαρακτηριστικά των κόμβων εάν χρησιμοποιούνται πριν το πρώτο συνελκτικό στρώμα είτε επεξεργάζονται τις αναπαραστάσεις των γράφων με σκοπό την εξαγωγή της τελικής πρόβλεψης εάν χρησιμοποιούνται μετά το στρώμα υποδειγματοληψίας.

Όπως και στην περίπτωση του MLP, οι παράμετροι του μοντέλου ανανεώνονται με τη γνωστή διαδικασία της εκπαίδευσης μέσω οπισθο-διάδοσης (back-propagation) με χρήση ενός αλγορίθμου βελτιστοποίησης. Ο καθορισμός της αρχιτεκτονικής του μοντέλου γίνεται με την κατάλληλη επιλογή διαφόρων υπερπαραμέτρων, δηλαδή παραμέτρων του μοντέλου που ορίζονται από τον χρήστη πριν εκτελέσει την διαδικασία της εκπαίδευσης στα δεδομένα εισόδου. Ένα νευρωνικό δίκτυο γράφων με είσοδο τους 1457 χημικούς γράφους του συνόλου δεδομένων του παρόντος προβλήματος καταναλώνει σε σύγκριση με το αντίστοιχο μοντέλο MLP αρκετά μεγαλύτερη υπολογιστική ισχύ, με αποτέλεσμα να κρίνεται απαραίτητη η χρήση GPU. Οι περιορισμένοι υπολογιστικοί πόροι που προσφέρει το Google Colaboratory δεν επέτρεψαν την εφαρμογή ενός αλγορίθμου αυτόματης επιλογής υπερπαραμέτρων, αντίστοιχο με την μπεϋζιανή βελτιστοποίηση στο μοντέλο MLP. Έτσι, η επιλογή των υπερπαραμέτρων γίνεται μη αυτόματα και συγκεκριμένα, λαμβάνοντας υπόψιν αντίστοιχες τιμές υπερπαραμέτρων της βιβλιογραφίας και εκτελώντας πειράματα εύρεσης και επιλογής υπερπαραμέτρων μεταβάλλοντάς τες σε ένα καθορισμένο εύρος και διατηρώντας τις υπόλοιπες σταθερές. Οι υπερπαραμέτροι που καθορίζουν το παρόν μοντέλο GNNs είναι οι εξής:

- Αλγόριθμος βελτιστοποίησης: η συνηθέστερη επιλογή μεταξύ μοντέλων νευρωνικών δικτύων γράφων είναι ο αλγόριθμος Adam [41] [85] [86]. Παρ' όλα αυτά, δοκιμάζεται και ο θεμελιώδης αλγόριθμος βελτιστοποίησης Stochastic Gradient Descent.
- Μέγεθος παρτίδας: διάφορες τιμές συνηθίζονται στη βιβλιογραφία (64, 512 για το GraphSage [85], [32,128] για το GINConv [86]). Όπως και στην περίπτωση του μοντέλου MLP, το μέγεθος παρτίδας σχετίζεται κυρίως με την ταχύτητα εκπαίδευσης του μοντέλου χωρίς να επηρεάζει σημαντικά την απόδοσή του. Έτσι, αποφασίζεται το εύρος επιλογών να περιλαμβάνει τις ενδεικτικές τιμές {32, 64, 128, 256}.
- Ρυθμός μάθησης: εκτός από απλή επιλογή τιμής ρυθμού μάθησης, εφαρμόζεται και πρόγραμμα ρυθμού μάθησης (learning rate schedule), διαδικασία που θα αναλυθεί παρακάτω.
- Συνάρτηση κόστους: επιλέγεται η συνάρτηση κόστους διασταυρούμενης εντροπίας.
- Συνάρτηση ενεργοποίησης: συνηθέστερη επιλογή στην βιβλιογραφία είναι η relu [41] [85] [86]. Παρ' όλα αυτά, πολύ καλά αποτελέσματα έχει αποδειχθεί ότι παράγει και η παραλλαγή της relu, prelu [45]. Επομένως, το εύρος επιλογής περιλαμβάνει τις δύο αυτές συναρτήσεις.
- Αριθμός συνελκτικών στρωμάτων: η επιλογή ποικίλει ανάλογα με τον τύπο του στρώματος και το πρόβλημα πρόβλεψης. Όπως έχει περιγραφεί στην παράγραφο 3.4.1, όσο βαθύτερο το νευρωνικό δίκτυο γράφων, τόσο μεγαλύτερη είναι η ακτίνα

της γειτονιάς από την οποία οι κόμβοι λαμβάνουν πληροφορίες για την ανανέωσή τους. Για το συγκεκριμένο πρόβλημα επιλέγεται να δοκιμαστούν 2, 4, 6 και 8 στρώματα GCN.

- Τύπος στρώματος υποδειγματοληψίας: οι συνηθέστερες προτιμήσεις στατιστικών συναρτήσεων καθολικής υποδειγματοληψίας και αυτές που δοκιμάζονται στην παρούσα εργασία είναι το άθροισμα, ο μέσος όρος και το μέγιστο.
- Αριθμός πλήρως συνδεδεμένων στρωμάτων προεπεξεργασίας: αναφέρεται στον αριθμό στρωμάτων MLP που επεξεργάζονται τα χαρακτηριστικά των κόμβων πριν την εισαγωγή τους στα συνελκτικά στρώματα. Έχει αποδειχθεί ότι η δράση τους είναι ευεργετική για την εκπαίδευση και επικύρωση του μοντέλου [45]. Επιλέγεται να δοκιμαστούν 0, 1 και 2 πλήρως συνδεδεμένα στρώματα προεπεξεργασίας.
- Αριθμός πλήρως συνδεδεμένων στρωμάτων τελικής πρόβλεψης: αναφέρεται στον αριθμό στρωμάτων MLP που λαμβάνουν ως είσοδο την αναπαράσταση του γράφου και παράγουν την τελική πρόβλεψη (στην προκειμένη περίπτωση, ταξινόμηση του χημικού γράφου). Δοκιμάζονται 1, 2 και 3 πλήρως συνδεδεμένα στρώματα.
- Αριθμός κρυφών νευρώνων στα συνελκτικά στρώματα: η διάσταση που ορίζει το μέγεθος των κρυφών αναπαραστάσεων των κόμβων. Το αρχικό μοντέλο GCN χρησιμοποιεί 16 και 64 κρυφούς νευρώνες ανάλογα με το πρόβλημα. Το εύρος επιλογής στην συγκεκριμένη περίπτωση θα περιλαμβάνει 16, 32, 64 και 128 κρυφούς νευρώνες. Σημειώνεται ότι για χάρη απλότητας και οικονομίας υπολογιστικού χρόνου τυχόν παρεμβαλλόμενα πλήρως συνδεδεμένα στρώματα θα έχουν τον ίδιο αριθμό κρυφών νευρώνων (το τελικό στρώμα έχει πάντα αριθμό νευρώνων ίσο με τις κλάσεις, στην προκειμένη περίπτωση 3).
- Στρώμα κανονικοποίησης παρτίδας (batch normalization layer): τόσο στα κλασικά συνελκτικά νευρωνικά δίκτυα όσο και στα νευρωνικά δίκτυα γράφων συνηθίζεται η εφαρμογή ενός στρώματος κανονικοποίησης της εισόδου ενός συνελκτικού στρώματος (βλ. υποκεφάλαιο 2.4).
- Dropout: Εφαρμογή dropout για κανονικοποίηση κόστους (βλ. παράγραφο 1.3.4).

Οι υπερπαράμετροι και το εύρος επιτρεπτών τιμών τους συνοψίζονται στον ακόλουθο πίνακα:

Πίνακας 6.9: Εύρος τιμών των υπερπαραμέτρων του μοντέλου GNN

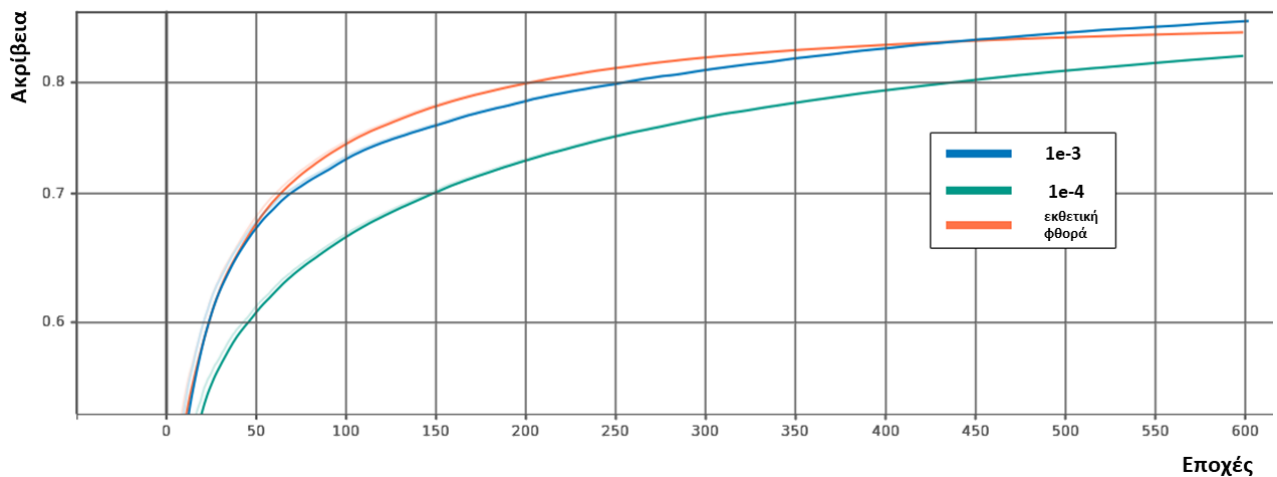
Υπερπαραμέτρος	Εύρος τιμών
Αλγόριθμος βελτιστοποίησης	{ <i>Adam, SGD</i> }
Μέγεθος παρτίδας	{32, 64, 128, 256}
Συνάρτηση κόστους	Διασταυρούμενη εντροπία
Ρυθμός μάθησης	$[10^{-5}, 10^{-2}]$ + Πρόγραμμα ρυθμού μάθησης
Συνάρτηση ενεργοποίησης	{ <i>relu, prelu</i> }
Αριθμός συνελκτικών στρωμάτων	{2, 4, 6, 8}
Στρώμα υποδειγματοληψίας	{ <i>sum, average, max</i> }
Αριθμός πλήρως συνδεδεμένων στρωμάτων προεπεξεργασίας	{0, 1, 2}
Αριθμός πλήρως συνδεδεμένων στρωμάτων τελικής πρόβλεψης	{1, 2, 3}
Αριθμός κρυφών νευρώνων	{16, 32, 64, 128}
Στρώματα κανονικοποίησης παρτίδας	{ <i>True, False</i> }
Στρώματα <i>Dropout</i>	{ <i>True, False</i> }

Όπως αναφέρθηκε στην παρουσίαση των υπερπαραμέτρων, εκτός από απλή επιλογή τιμών ρυθμού μάθησης, επιλέγεται και η εφαρμογή προγράμματος ρυθμού μάθησης (*learning rate schedule*), όρος που συναντάται στη βιβλιογραφία και ως φθορά ρυθμού μάθησης (*learning rate decay*). Ως γνωστόν, μια μικρή τιμή ρυθμού μάθησης μπορεί να αποφέρει καλά αποτελέσματα επικύρωσης, αλλά με πολύ αργή εκπαίδευση, τη στιγμή που μια μεγάλη τιμή μπορεί να επιταχύνει την σύγκλιση σε υψηλές τιμές ακρίβειας στα δεδομένα εκπαίδευσης με υπερπροσαρμογή και άσχημα αποτελέσματα στα δεδομένα επαλήθευσης. Πολλές φορές, προτιμάται η μεταβολή του ρυθμού μάθησης και συγκεκριμένα, η ελάττωση του με τα χρονικά βήματα της εκπαίδευσης. Κάτι τέτοιο επιτρέπει μεγάλες αλλαγές των εκπαιδευσιμων παραμέτρων στο ξεκίνημα της εκπαίδευσης και μικρότερες αλλαγές που έχουν τον χαρακτήρα της λεπτής ρύθμισης (*fine-tuning*) προς το τέλος της εκπαιδευτικής διαδικασίας [7]. Το πρόγραμμα ρυθμού μάθησης που εφαρμόζεται ακολουθεί την εκθετική φθορά και ορίζεται ως εξής:

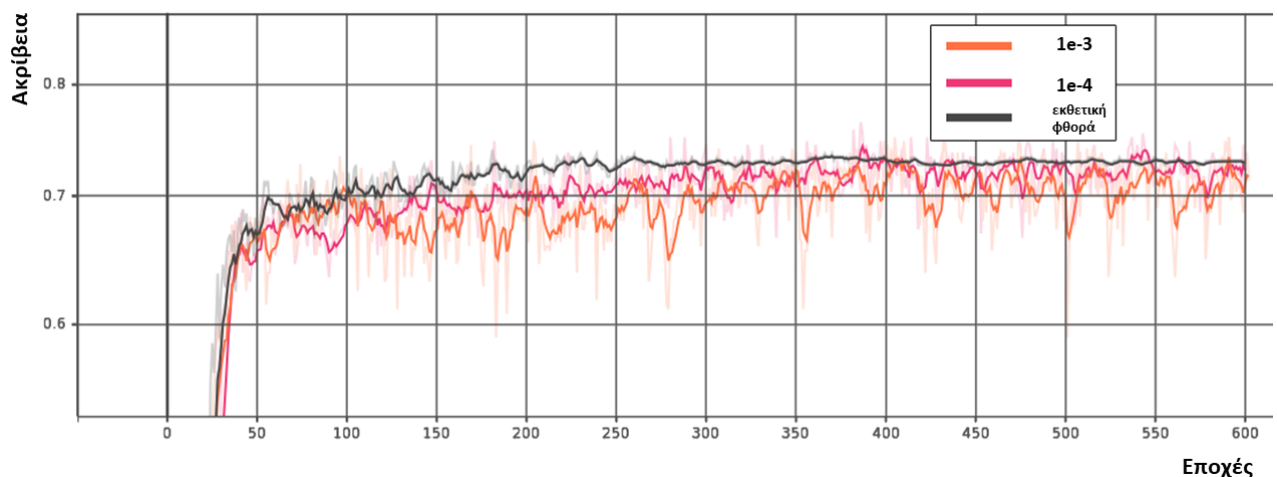
$$\eta = \eta_0 r_d \exp\left(-\frac{t}{t_d}\right) \quad (6.4)$$

όπου η_0 ο αρχικός ρυθμός μάθησης, r_d , t_d , ο ρυθμός φθοράς (*decay rate*) και οι εποχές φθοράς (*decay epochs*) αντίστοιχα, παράμετροι που καθορίζουν τον ρυθμό φθοράς και την τελική τιμή του ρυθμού μάθησης, και t το χρονικό βήμα (εποχή) της εκπαιδευτικής διαδικασίας.

Ακολουθεί πείραμα αναζήτησης κατάλληλου ρυθμού μάθησης με τιμές ρυθμού μάθησης 10^{-3} και 10^{-4} , καθώς και πρόγραμμα εκθετικής φθοράς με αρχική τιμή ρυθμού μάθησης το 10^{-3} (μια τιμή που επιτρέπει μεγάλες αλλαγές βαρών και εκπαιδεύει γρήγορα το δίκτυο), ρυθμό φθοράς 0.2 και εποχές φθοράς 1200. Σημειώνεται ότι το μοντέλο βελτιστοποιείται με *Adam*, περιέχει 4 συνελκτικά στρώματα GCN και εκπαιδεύεται για 600 εποχές.



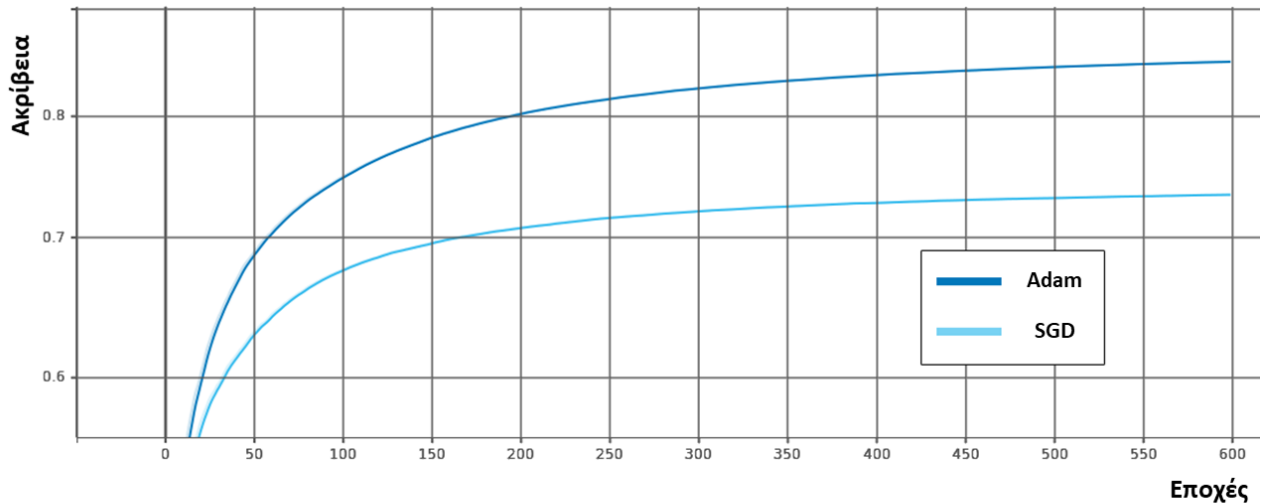
Σχήμα 6.12: Ακρίβεια πρόβλεψης στο σύνολο δεδομένων εκπαίδευσης για εκπαίδευση με αλγόριθμο Adam και ρυθμούς μάθησης 10^{-3} , 10^{-4} και εκθετική φθορά ρυθμού μάθησης με αρχικό ρυθμό μάθησης 10^{-3} , ρυθμό φθοράς 0, 2 και 1200 εποχές φθοράς.



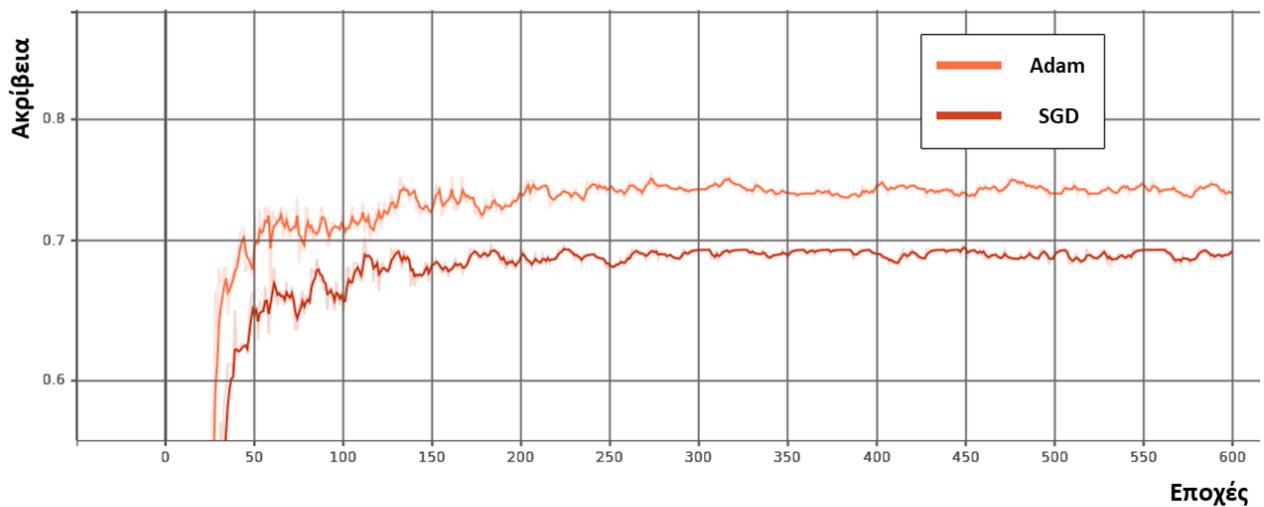
Σχήμα 6.13: Ακρίβεια πρόβλεψης στο σύνολο δεδομένων επαλήθευσης για εκπαίδευση με αλγόριθμο Adam και ρυθμούς μάθησης 10^{-3} , 10^{-4} και εκθετική φθορά ρυθμού μάθησης με αρχικό ρυθμό μάθησης 10^{-3} , ρυθμό φθοράς 0, 2 και 1200 εποχές φθοράς.

Όπως παρατηρείται στο σχήμα 6.12, η συμπεριφορά εκπαίδευσης στο σύνολο δεδομένων εκπαίδευσης είναι παρόμοια και για τις τρεις επιλογές ρυθμού μάθησης. Όπως αναμενόταν, για μικρότερο ρυθμό μάθησης (10^{-4}), η εκπαίδευση είναι πιο αργή και άρα, η ακρίβεια πρόβλεψης στις 600 εποχές μικρότερη. Όσον αφορά τα δεδομένα επαλήθευσης, σύμφωνα με το σχήμα 6.13, το μοντέλο με ρυθμό μάθησης 10^{-3} παρουσιάζει μεγάλη αστάθεια στην ακρίβεια πρόβλεψης σε όλο το εύρος των εποχών, γεγονός που υποδηλώνει εμφάνιση υπερκπαίδευσης. Η συμπεριφορά αυτή βελτιώνεται με ρυθμό μάθησης 10^{-4} , ενώ, σχεδόν, εξαλείφεται με την εκθετική φθορά του ρυθμού μάθησης. Μάλιστα, με την εκθετική φθορά το μοντέλο επιτυγχάνει τελικά την καλύτερη ακρίβεια πρόβλεψης στα δεδομένα επαλήθευσης. Επομένως, η επιλογή αυτή για πρόγραμμα ρυθμού μάθησης κρίνεται ικανοποιητική και εφαρμόζεται στο τελικό μοντέλο GNN. Σημειώνεται ότι δοκιμάστηκαν και άλλες δομές εκθετικής φθοράς ρυθμού μάθησης με αυτήν που παρουσιάζεται παραπάνω να έχει συγκριτικά τα καλύτερα αποτελέσματα στα δεδομένα επαλήθευσης.

Άλλη μια σημαντική υπερπαράμετρος του μοντέλου είναι ο αλγόριθμος βελτιστοποίησης. Οι επιλογές που ερευνώνται είναι ο αλγόριθμος SGD και ο αλγόριθμος Adam. Και στις δύο περιπτώσεις το μοντέλο 4 συνελκτικών στρωμάτων εκπαιδεύεται για 600 εποχές με ρυθμό μάθησης που προκύπτει από την εκθετική φθορά.



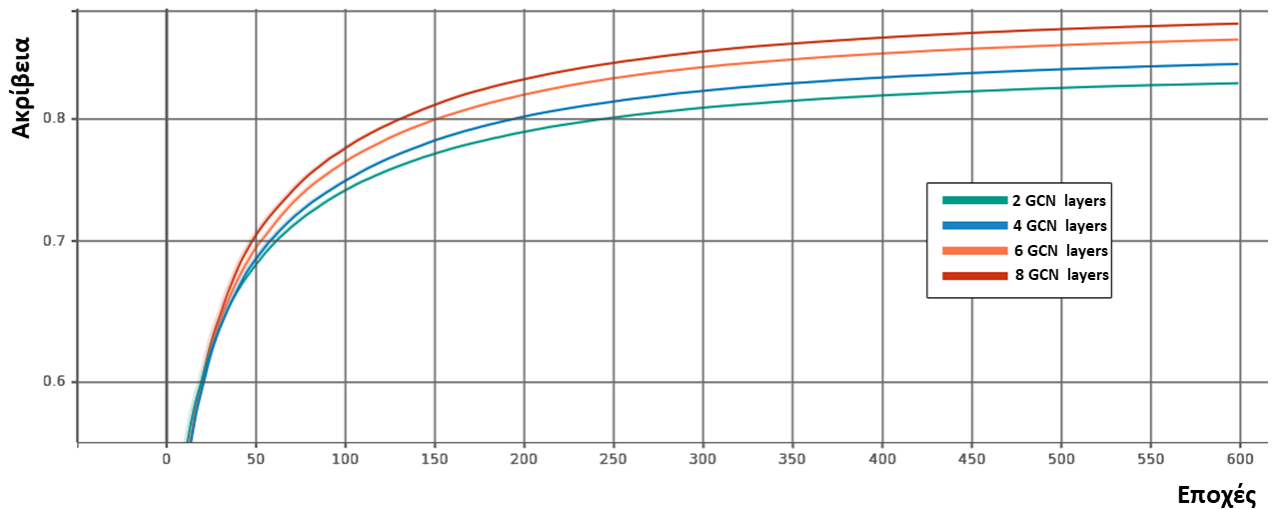
Σχήμα 6.14: Ακρίβεια πρόβλεψης στο σύνολο δεδομένων εκπαίδευσης για εκπαίδευση με αλγόριθμο Adam και αλγόριθμο SGD με ρυθμό μάθησης που προκύπτει από εκθετική φθορά με αρχικό ρυθμό μάθησης 10^{-3} , ρυθμό φθοράς 0,2 και 1200 εποχές φθοράς. Το μοντέλο περιέχει 4 συνελκτικά στρώματα και εκπαιδεύεται για 600 εποχές.



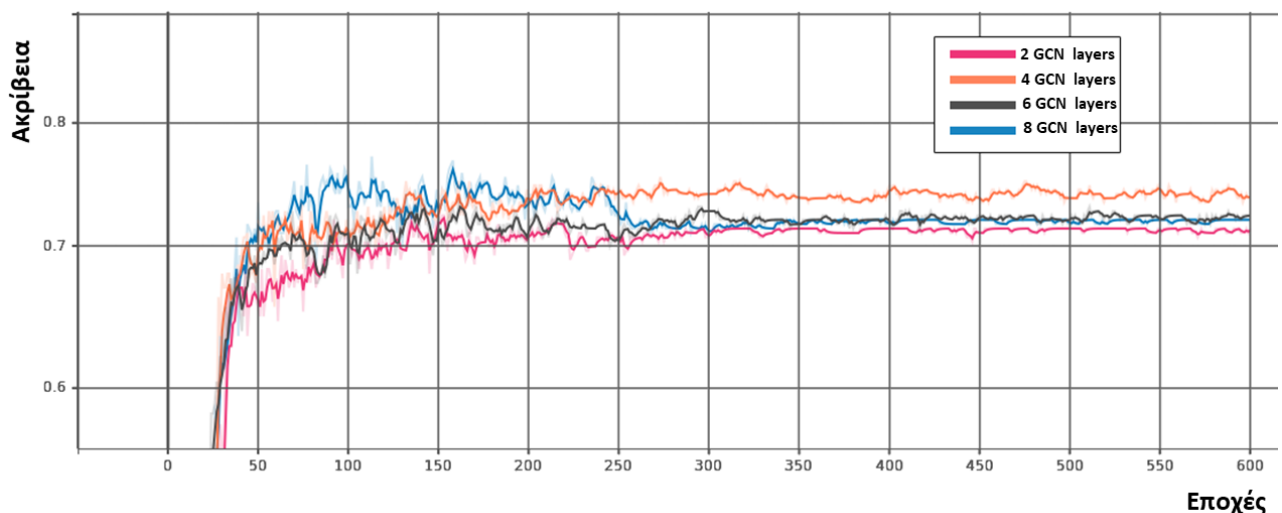
Σχήμα 6.15: Ακρίβεια πρόβλεψης στο σύνολο δεδομένων επαλήθευσης για εκπαίδευση με αλγόριθμο Adam και αλγόριθμο SGD με ρυθμό μάθησης που προκύπτει από εκθετική φθορά με αρχικό ρυθμό μάθησης 10^{-3} , ρυθμό φθοράς 0,2 και 1200 εποχές φθοράς. Το μοντέλο περιέχει 4 συνελκτικά στρώματα και εκπαιδεύεται για 600 εποχές.

Η υπεροχή του αλγορίθμου βελτιστοποίησης Adam είναι ξεκάθαρη, αφού όπως παρατηρείται στα σχήματα 6.14 και 6.15, το μοντέλο συγκλίνει σε υψηλότερες τιμές ακρίβειας τόσο στα δεδομένα εκπαίδευσης όσο και στα δεδομένα επαλήθευσης.

Τέλος, παρουσιάζεται σχηματικά και η σπουδαίας σημασίας για την δομή του μοντέλου GNN υπερπαραμέτρος του πλήθους των συνελκτικών στρωμάτων γράφων. Και εδώ, το μοντέλο εκπαιδεύεται για 600 εποχές με χρήση αλγορίθμου Adam και ρυθμό μάθησης που προκύπτει από την εκθετική φθορά. Παρουσιάζονται οι αποκρίσεις της ακρίβειας πρόβλεψης για 2, 4, 6 και 8 συνελκτικά στρώματα :



Σχήμα 6.16: Ακρίβεια πρόβλεψης στο σύνολο δεδομένων εκπαίδευσης για εκπαίδευση μοντέλου με 2, 4, 6 και 8 στρώματα GCN χρησιμοποιώντας αλγόριθμο Adam με ρυθμό μάθησης που προκύπτει από εκθετική φθορά με αρχικό ρυθμό μάθησης 10^{-3} , ρυθμό φθοράς 0, 2 και 1200 εποχές φθοράς. Το μοντέλο εκπαιδεύεται για 600 εποχές.



Σχήμα 6.17: Ακρίβεια πρόβλεψης στο σύνολο δεδομένων επαλήθευσης για εκπαίδευση μοντέλου με 2, 4, 6 και 8 στρώματα GCN χρησιμοποιώντας αλγόριθμο Adam με ρυθμό μάθησης που προκύπτει από εκθετική φθορά με αρχικό ρυθμό μάθησης 10^{-3} , ρυθμό φθοράς 0, 2 και 1200 εποχές φθοράς. Το μοντέλο εκπαιδεύεται για 600 εποχές.

Όσον αφορά τα δεδομένα εκπαίδευσης, από το σχήμα 6.16 συμπεραίνεται ότι όσο αυξάνονται τα συνελκτικά στρώματα γράφων τόσο αυξάνεται και η ακρίβεια πρόβλεψης στην οποία συγκλίνει το μοντέλο GNN στις 600 εποχές. Παρ' όλα αυτά, στα δεδομένα επαλήθευσης την καλύτερη απόκριση δίνει το μοντέλο με τα 4 στρώματα GCN. Ειδικότερα, συγκλίνει στην υψηλότερη τιμή ακρίβειας πρόβλεψης παρόλο που το μοντέλο με τα 8 συνελκτικά στρώματα επιτυγχάνει μεν σε απόλυτη τιμή μεγαλύτερη ακρίβεια πρόβλεψης, αλλά σε αρχικό χρονικό βήμα. Στην συνέχεια, το μοντέλο αυτό συγκλίνει σε αρκετά χαμηλότερη ακρίβεια, γεγονός που υποδηλώνει την εμφάνιση υπερεκπαίδευσης. Επομένως, ως βέλτιστος αριθμός συνελκτικών στρωμάτων γράφων επιλέγονται τα 4, δηλαδή οι κόμ-

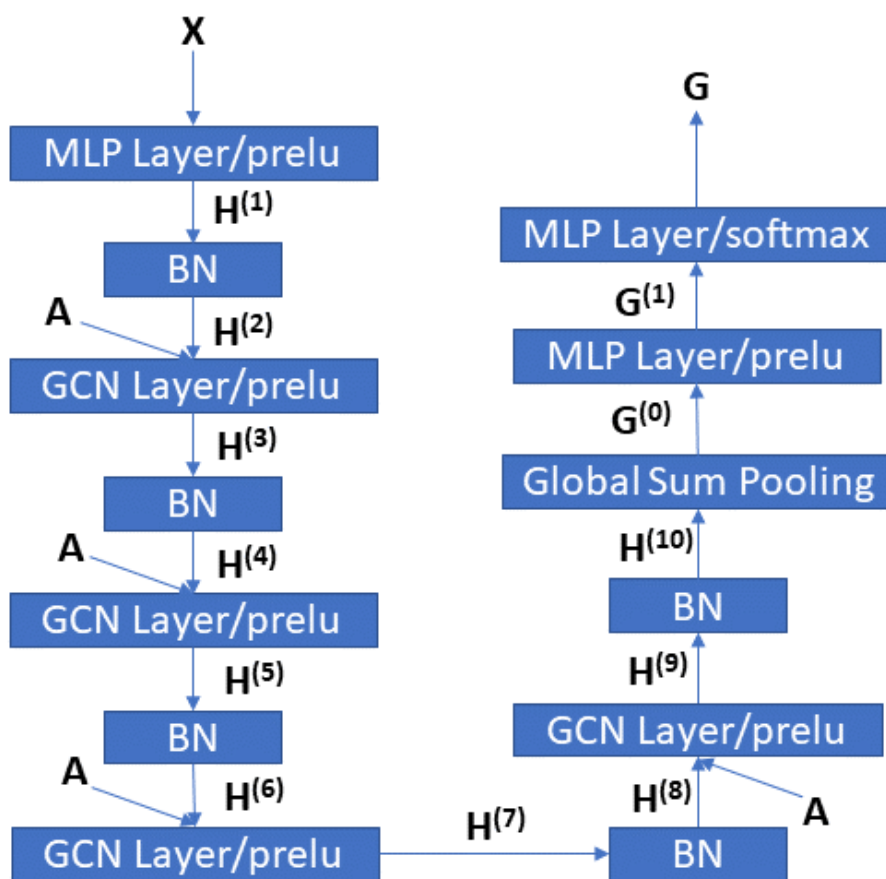
βοι των χημικών γράφων ανανεώνονται συλλέγοντας πληροφορίες από την 4-hop γειτονιά τους.

Με αντίστοιχο τρόπο, καθορίζονται και οι υπόλοιπες υπερπαράμετροι του μοντέλου GNN. Τα αποτελέσματα, τα οποία ορίζουν την αρχιτεκτονική του τελικού μοντέλου παρουσιάζονται στον ακόλουθο πίνακα:

Πίνακας 6.10: Τελικές υπερπαράμετροι του μοντέλου GNN

Υπερπαράμετρος	Τιμή
Αλγόριθμος βελτιστοποίησης	<i>Adam</i>
Μέγεθος παρτίδας	{128}
Συνάρτηση κόστους	Διασταυρούμενη εντροπία
Ρυθμός μάθησης	Πρόγραμμα ρυθμού μάθησης
Συνάρτηση ενεργοποίησης	<i>relu</i>
Αριθμός συνελικτικών στρωμάτων	4
Στρώμα υποδειγματοληψίας	<i>sum</i>
Αριθμός πλήρως συνδεδεμένων στρωμάτων προεπεξεργασίας	1
Αριθμός πλήρως συνδεδεμένων στρωμάτων τελικής πρόβλεψης	2
Αριθμός κρυφών νευρώνων	128
Στρώματα κανονικοποίησης παρτίδας	<i>True</i>
Στρώματα <i>Dropout</i>	<i>False</i>

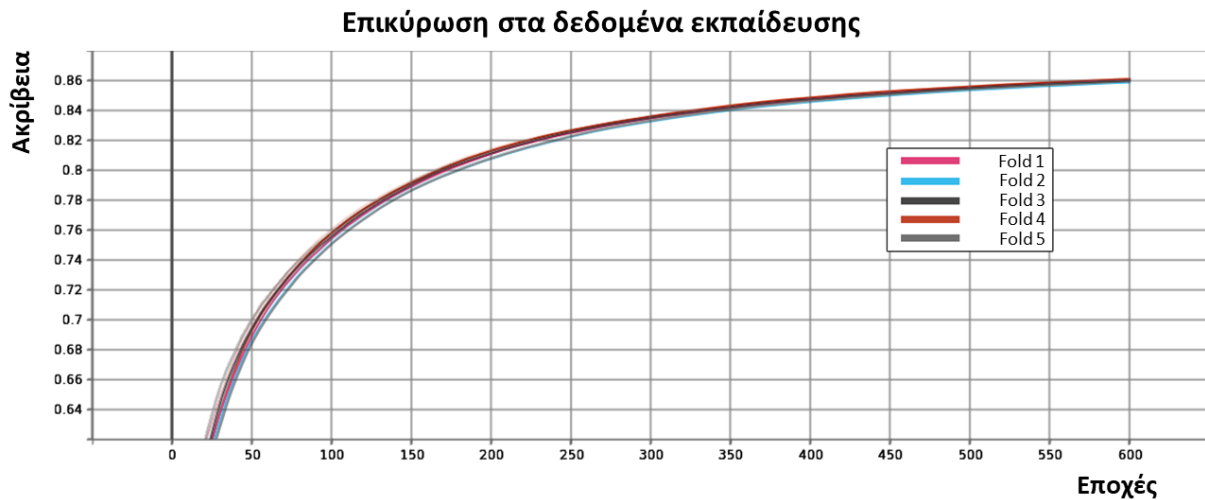
Η σχηματική αναπαράσταση της αρχιτεκτονικής του μοντέλου GNN φαίνεται παρακάτω:



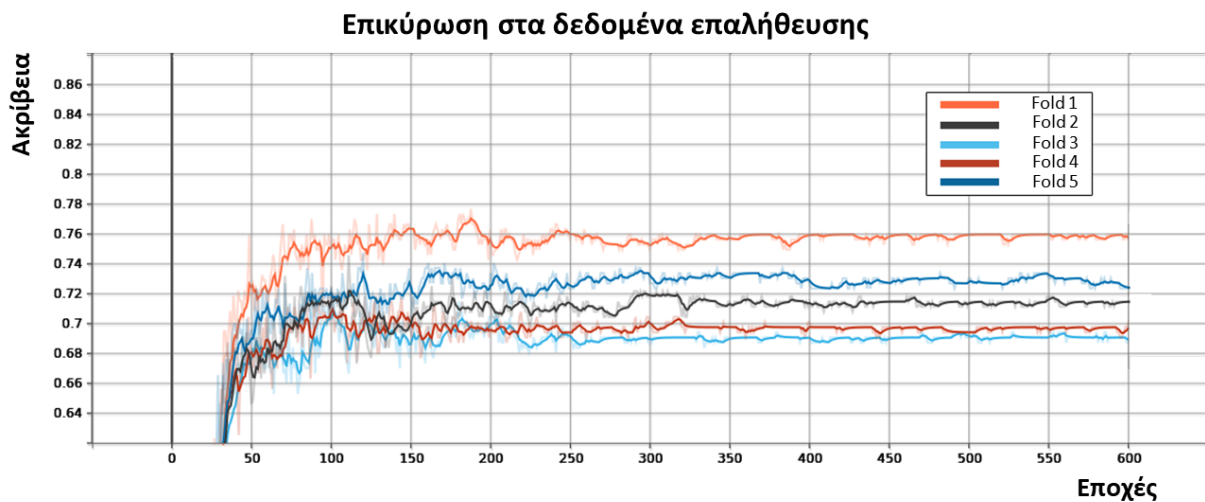
Σχήμα 6.18: Η αρχιτεκτονική των στρωμάτων του μοντέλου GNN. Με X συμβολίζονται τα αρχικά χαρακτηριστικά των κόμβων, με $U^{(i)}$ οι κρυφές αναπαραστάσεις των κόμβων που παράγονται από το στρώμα i , με A ο πίνακας γειτνίασης του γράφου, με $G^{(0)}$ και $G^{(1)}$ οι κρυφές αναπαραστάσεις του γράφου και με G η τελική πρόβλεψη σε επίπεδο γράφου.

6.5.3 Αποτελέσματα (Results)

Η αρχιτεκτονική και οι ρυθμίσεις του μοντέλου GNN προέκυψαν από τις τιμές των υπερπαραμέτρων του πίνακα 6.10. Η εκπαίδευση αποφασίστηκε να διαρκέσει 600 χρονικά βήματα, καθώς στο τέλος του χρονικού αυτού εύρους η απόδοση στα δεδομένα επαλήθευσης έχει σταθεροποιηθεί. Όπως και στην περίπτωση του μοντέλου MLP, η επικύρωση λαμβάνει χώρα με εφαρμογή 5-fold cross validation. Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε η GPU NVIDIA Tesla K80 12 GB μέσω Google Colaboratory. Οι υπολογιστικοί χρόνοι ανήλθαν σε περίπου 8 λεπτά για 600 εποχές εκπαίδευσης. Η απόκριση της ακρίβειας τόσο στα δεδομένα εκπαίδευσης όσο και στα δεδομένα επαλήθευσης με τα χρονικά βήματα φαίνεται στα ακόλουθα σχήματα:



Σχήμα 6.19: Μοντέλο GNN: αποτελέσματα 5-fold cross validation ακρίβειας συναρτήσεϊ εποχών στα δεδομένα εκπαίδευσης



Σχήμα 6.20: Μοντέλο GNN: αποτελέσματα 5-fold cross validation ακρίβειας συναρτήσεϊ εποχών στα δεδομένα επαλήθευσης

Οι συγκεντρωτικές τιμές των μέτρων επικύρωσης στα δεδομένα επαλήθευσης, ακρίβειας και MCC, μετά την εφαρμογή του 5-fold cross-validation είναι οι ακόλουθες:

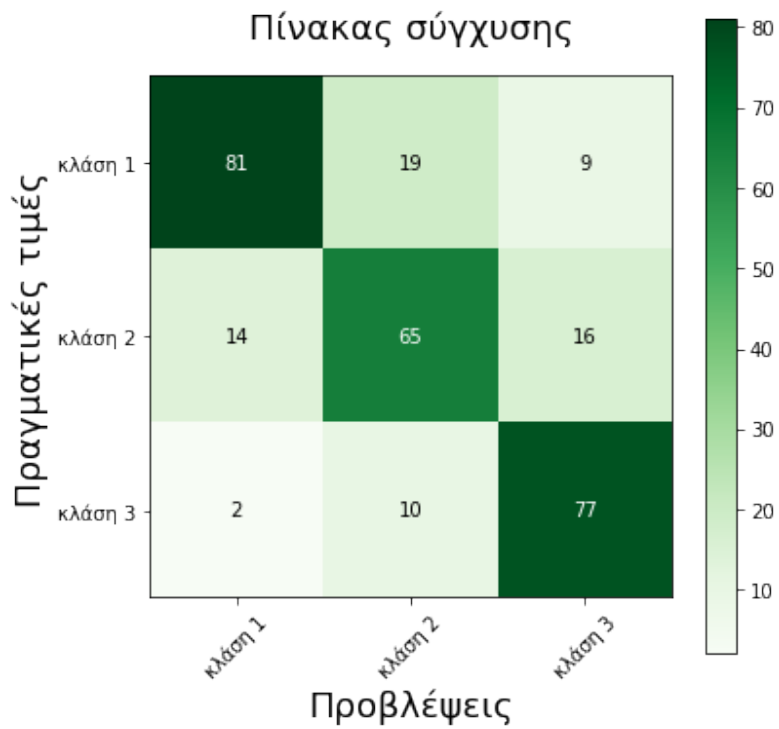
Πίνακας 6.11: Αποτελέσματα ακρίβειας και MCC του μοντέλου GNN

	Ακρίβεια	MCC
<i>Fold 1</i>	0.76	0.64
<i>Fold 2</i>	0.72	0.58
<i>Fold 3</i>	0.69	0.54
<i>Fold 4</i>	0.70	0.56
<i>Fold 5</i>	0.73	0.60
Σύνολο	0.72 ± 0.02	0.58 ± 0.03

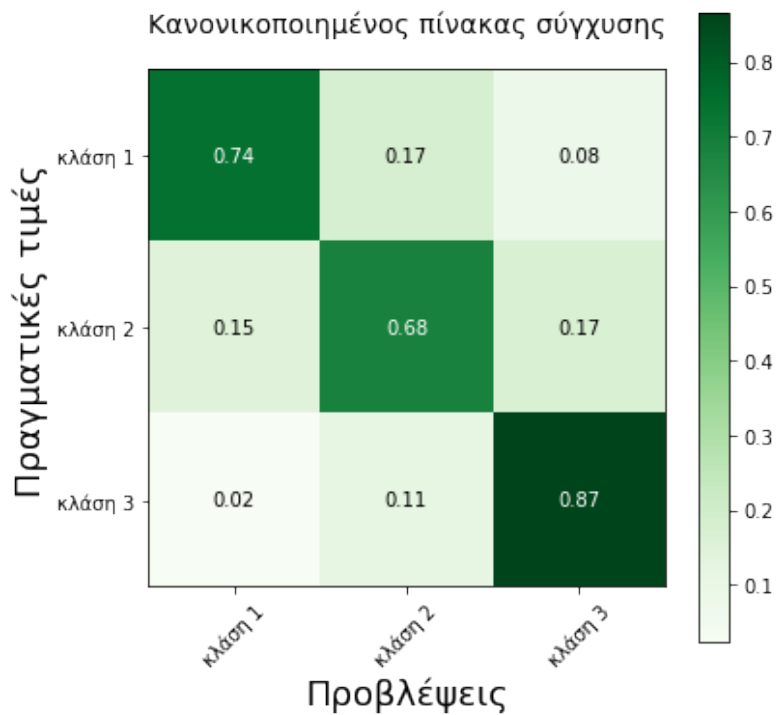
Επίσης, παρατίθενται ο πίνακας σύγχυσης, καθώς και η κανονικοποιημένη μορφή του που

αντιστοιχούν στην καλύτερη πρόβλεψη που προέκυψε από την διαδικασία επικύρωσης (1^ο fold του cross validation):

Πίνακας 6.12: Πίνακας σύγχυσης του μοντέλου GNN



Πίνακας 6.13: Κανονικοποιημένος πίνακας σύγχυσης του μοντέλου GNN



Από τα σχήμα 6.19 παρατηρείται ότι και τα 5 folds που προκύπτουν από τη διαδικασία του

cross validation έχουν όμοια συμπεριφορά στα δεδομένα εκπαίδευσης. Συγκεκριμένα, σε 600 εποχές επιτυγχάνουν ακρίβεια ίση 0.86. Περαιτέρω εκπαίδευση του μοντέλου για επίτευξη μεγαλύτερης ακρίβειας στα δεδομένα εκπαίδευσης δεν κρίνεται αναγκαία, καθώς σύμφωνα με το σχήμα 6.20 και στα 5 folds η απόδοση του μοντέλου στα δεδομένα επαλήθευσης έχει σταθεροποιηθεί και μεταβάλλεται ελάχιστα. Η μέση τιμή ακρίβειας στα δεδομένα επαλήθευσης του μοντέλου GNN αγγίζει το 0.72 με τυπική απόκλιση μεταξύ των 5 folds 0.02, ενώ η αντίστοιχη μέση τιμή συντελεστή συσχέτισης Matthews το 0.58 με τυπική απόκλιση 0.03. Παρατηρείται, λοιπόν, ότι η μέση απόδοση του μοντέλου GNN είναι ελάχιστα καλύτερη από την αντίστοιχη του μοντέλου MLP. Η επίδοση του πρώτου fold που προέκυψε από το cross-validation αποτελεί, μάλιστα, την καλύτερη που έχει παρατηρηθεί αναφορικά με το συγκεκριμένο σύνολο δεδομένων χημικών ουσιών. Σύμφωνα με το σχήμα 6.13, η ακρίβεια σε επίπεδο τρίτης κλάσης αγγίζει το 87%, σε επίπεδο πρώτης κλάσης το 74% και σε επίπεδο δεύτερης κλάσης το 68%. Η χαμηλότερη απόδοση στους ενδοκρινικούς διαταράκτες που ανήκουν στην δεύτερη κλάση (δηλαδή με μέτρια συγγένεια πρόσδεσης στους υποδοχείς οιστρογόνων) είναι αναμενόμενη για τους λόγους που εξηγήθηκαν και στα αποτελέσματα του μοντέλου MLP (παράγραφος 6.4.3). Παρ' όλα αυτά, σε σύγκριση με το μοντέλο MLP παρατηρείται πολύ μεγαλύτερη ικανότητα πρόβλεψης στην τρίτη κλάση (ισχυροί ενδοκρινικοί διαταράκτες), όπου βρίσκονται ουσίες που μπορούν δυνητικά να καταστούν οι πιο επικίνδυνες (μεταξύ του συνόλου δεδομένων) για την υγεία του ανθρώπινου οργανισμού.

Κεφάλαιο 7

Συμπεράσματα και Προτάσεις για Μελλοντική Έρευνα (Conclusions and Suggestions for Future Research)

Συνοψίζοντας, στην παρούσα διπλωματική εργασία αναπτύχθηκαν δύο προβλεπτικά μοντέλα βαθιάς μηχανικής μάθησης που ως σκοπό έχουν την ταξινόμηση χημικών ουσιών, οι οποίες διαταράσσουν το ενδοκρινικό σύστημα του ανθρώπου, ανάλογα με την πειραματική τιμή συγγένειας πρόσδεσης στους υποδοχείς οιστρογόνων. Το πρώτο μοντέλο δομήθηκε με χρήση βαθιάς αρχιτεκτονικής νευρωνικών δικτύων εμπρόσθιας τροφοδότησης (MLP), ενώ το δεύτερο με χρήση βαθιάς αρχιτεκτονικής νευρωνικών δικτύων γράφων (GNN) που έχουν ως βάση το συνελικτικό δίκτυο γράφων (GCN). Από τα παραπάνω μοντέλα, καθώς και από το μοντέλο συνελικτικών νευρωνικών δικτύων (CNN) που αναπτύχθηκε από τον υποψήφιο διδάκτορα Παντελή Καρατζά στο πλαίσιο της διπλωματικής του εργασίας για το ΔΠΜΣ "Εφαρμοσμένες Μαθηματικές Επιστήμες" εξάγονται χρήσιμα συμπεράσματα αναφορικά με τη χρήση της βαθιάς μηχανικής μάθησης σε εφαρμογές προτυποποίησης σχέσεων μεταξύ χημικής δομής και (βιολογικής) δράσης χημικών ουσιών.

7.1 Συμπεράσματα (Conclusions)

Οι δομικές ιδιότητες των ενδοκρινικών διαταρακτών του συνόλου δεδομένων, το οποίο ανλήθηκε από τον Οργανισμό Τροφίμων και Φαρμάκων των ΗΠΑ, προτυποποιήθηκαν (α) με τη χρήση περιγραφικών μεταβλητών που παρήγαγε το λογισμικό mordred και (β) με τη χρήση χημικών γράφων με χαρακτηριστικά στους κόμβους που παρήγαγαν τα λογισμικά deepchem και rdkit. Οι παραπάνω υπολογιστικές δομές αποτέλεσαν δεδομένα εισόδου σε βαθιά νευρωνικά δίκτυα εμπρόσθιας τροφοδότησης και γράφων αντίστοιχα. Το πρόβλημα που επιλύεται αποτελεί πρόβλημα ταξινόμησης σε 3 κλάσεις, οι οποίες ορίζονται ανάλογα με την σχετική συγγένεια πρόσδεσης των ουσιών (ενός μέτρου του δυναμικού πρόσδεσης) στους υποδοχείς οιστρογόνων με τιμή συγγένειας πρόσδεσης αναφοράς αυτήν της 17-β οιστραδιόλης ($\log RBA = 2$).

Η αρχιτεκτονική του μοντέλου νευρωνικών δικτύων εμπρόσθιας τροφοδότησης (μοντέλο MLP) καθορίζεται από εφαρμογή μπεϋζιανής βελτιστοποίησης σε σημαντικές υπερπαραμέτρους των νευρωνικών δικτύων. Το τελικό μοντέλο που παράχθηκε φαίνεται να εκπαιδεύεται ικανοποιητικά αφενός επειδή η εσωτερική ακρίβεια, δηλαδή η ακρίβεια στα δεδομένα εκπαίδευσης, προσεγγίζει το 1 και αφετέρου επειδή σε ένα εύρος 200 χρονικών βημάτων

εκπαίδευσης τα μέτρα αξιολόγησης του μοντέλου στα δεδομένα επαλήθευσης σταθεροποιούνται στην μέγιστή τους τιμή. Συγκεκριμένα, μέσω της διαδικασίας διασταυρούμενης επικύρωσης η μέση ακρίβεια υπολογίστηκε ίση με 0.71, ενώ ο μέσος συντελεστής συσχέτισης Matthews ίσος με 0.56. Ιδιαίτερα, ο συντελεστής MCC, όπως και ο πίνακας σύγχυσης, δίνουν πληροφορίες για την απόδοση του μοντέλου σε επίπεδο κλάσεων. Από αυτά τα μέτρα αξιολόγησης εξάγεται το συμπέρασμα ότι το μοντέλο επιτυγχάνει ικανοποιητική πρόβλεψη, ιδιαίτερα στις κλάσεις 1 και 3 που αφορούν ουσίες με ασθενές και ισχυρό δυναμικό πρόσδεσης αντίστοιχα.

Αναφορικά με το μοντέλο των ταχύτατα εξελισσόμενων και state of the art νευρωνικών δικτύων γράφων (μοντέλο GNN), η βέλτιστη αρχιτεκτονική του επιλέγεται χειροκίνητα μέσω αναλύσεων ευαισθησίας της απόκρισης εξωτερικής ακρίβειας ως προς τις διάφορες υπερπαραμέτρους. Το τελικό μοντέλο προσεγγίζει εσωτερική ακρίβεια ίση με 0.86 και εξωτερική μέση ακρίβεια ίση με 0.72. Τίθεται, λοιπόν, το ζήτημα της κατασκευής ενός μοντέλου που να εκπαιδεύεται με υπολογιστικά εφικτό τρόπο σε υψηλότερες τιμές εσωτερικής ακρίβειας με ταυτόχρονη βελτίωση των μέτρων αξιολόγησης στα δεδομένα επαλήθευσης. Παρ' όλα αυτά, το μοντέλο GNN παρουσιάζει συνολικά καλύτερη απόδοση από το μοντέλο MLP τόσο στην ακρίβεια όσο και στον συντελεστή MCC. Ταυτόχρονα, κατά την διασταυρούμενη επικύρωση παρατηρήθηκε και η καλύτερη μοντελοποίηση του συστήματος με την ακρίβεια σε επίπεδο τρίτης κλάσης (ισχυρών ενδοκρινικών διαταρακτών) να αγγίζει το 0.87.

Συμπερασματικά, τα μοντέλα που αναπτύχθηκαν αποδίδουν καλύτερα από το μοντέλο συνελκτικών νευρωνικών δικτύων με είσοδο εικόνες που αναπτύχθηκε σε προηγούμενη εργασία. Αυτό δείχνει ότι τα κλασικά νευρωνικά δίκτυα με είσοδο περιγραφικές μεταβλητές και τα νευρωνικά δίκτυα γράφων αποτελούν τις καλύτερες προσεγγίσεις για πρόβλεψη χημικών ιδιοτήτων με τεχνικές βαθιάς μηχανικής μάθησης. Επιπλέον, τα καλύτερα αποτελέσματα των δεικτών αξιολόγησης του μοντέλου GNN συγκριτικά με το μοντέλο MLP αποδεικνύουν την σιβαρότητα των νευρωνικών δικτύων γράφων στην πρόβλεψη ιδιοτήτων χημικών ουσιών. Η "φυσική" αναπαράσταση των χημικών μορίων ως γράφοι, η οποία συνοδεύεται από την απουσία ανάγκης κατασκευής και επιλογής μοριακών περιγραφικών μεταβλητών δημιουργούν ένα νέο πολλά υποσχόμενο περιβάλλον ανάπτυξης μαθηματικών μοντέλων πρόβλεψης χημικών ιδιοτήτων με δυνατότητα βαθιάς εκπαίδευσης σε μεγάλα σύνολα δεδομένων μέσω σύγχρονων τεχνολογικών μέσων, όπως οι GPU.

7.2 Προτάσεις για Μελλοντική Έρευνα (Suggestions for Future Research)

Το άμεσο πεδίο μελλοντικής έρευνας που σχετίζεται με την παρούσα εργασία αφορά την βελτίωση των προβλεπτικών μοντέλων που αναπτύχθηκαν. Τα περιθώρια βελτίωσης του μοντέλου νευρωνικών δικτύων εμπρόσθιας τροφοδότησης είναι μικρά, καθώς το μοντέλο καθορίστηκε με αυτόματη διαδικασία βελτιστοποίησης και περιλαμβάνει πολλές state of the art ρυθμίσεις νευρωνικών δικτύων, όπως ο βελτιστοποιητής Adam και η συνάρτηση ενεργοποίησης relu. Μια πιθανή προσέγγιση, παρ' όλα αυτά, θα ήταν η σύγκριση με ένα "αθαθές" νευρωνικό δίκτυο μικρού αριθμού νευρώνων, το οποίο θα προκύπτει από προσεκτική επιλογή των περιγραφικών μοριακών μεταβλητών. Αναφορικά με το μοντέλο νευρωνικών δικτύων γράφων η καλύτερη μοντελοποίηση του συστήματος μπορεί να επιτευχθεί με μοντέλα που στηρίζονται σε νεότερες και σιβαρότερες αρχιτεκτονικές νευρωνικής ανταλλαγής μηνυμάτων. Ιδιαίτερα, βελτίωση στην πρόβλεψη χημικών ιδιοτήτων μπορούν να φέρουν μοντέλα που λαμβάνουν υπόψιν τους κατά την ανανέωση των αναπαραστάσεων

των κόμβων/ατόμων τα χαρακτηριστικά των ακμών/δεσμών, κάτι που δεν επιτυγχάνεται μέσω του συνελκτικού δικτύου γράφων (GCN).

Επίσης, το πρόβλημα δύναται άμεσα να αποδεσμευτεί από τα όρια του προβλήματος ταξινόμησης, το οποίο προέκυψε ως συνέχιση προηγούμενης εργασίας ταξινόμησης εικόνων χημικών μορίων τριών κλάσεων, και να μετατραπεί σε πρόβλημα παλινδρόμησης. Με αυτόν τον τρόπο, θα είναι περισσότερο εφικτή η άμεση σύγκριση των μοντέλων βαθιάς μηχανικής μάθησης με παλαιότερα QSAR μοντέλα που προτυποποιούν το δυναμικό πρόσδεσης χημικών ουσιών στο ενδοκρινικό σύστημα.

Τέλος, ένα ευρύτερο πεδίο έρευνας αφορά την μοντελοποίηση μέσω βαθιών νευρωνικών δικτύων γράφων εκτός από των σχετικά μικρών διαστάσεων χημικών μορίων, και μεγαλύτερων βιομορίων, όπως πρωτεΐνες και δομές RNA και DNA, καθώς και άλλων δομών που συναντώνται στη βιολογία, στη χημεία και στην επιστήμη των υλικών (όπως νανοϋλικά). Ένας γράφος μπορεί να σχηματιστεί εάν τα άτομα θεωρηθούν ως κόμβοι, η σχετική θέση των οποίων και άρα, και η γειτονιά τους καθορίζεται από τις αντίστοιχες συντεταγμένες της 3D αναπαράστασης των μορίων. Επίσης, ενδιαφέρον παρουσιάζεται και στην προοπτική μοντελοποίησης της αλληλεπίδρασης μεταξύ διαφορετικών μορίων μέσω γράφων με εφαρμογές στην ανακάλυψη φαρμάκων (π.χ. αλληλεπίδραση πρωτεΐνης-πρωτεΐνης). Μάλιστα, και η περίπτωση του προβλήματος της παρούσας εργασίας θα μπορούσε να θεωρηθεί ως μια περίπτωση αλληλεπίδρασης μιας χημικής ουσίας με έναν πυρηνικό υποδοχέα (πρωτεΐνη), όπως ο υποδοχέας οιστρογόνων.

Παράρτημα Α΄

Ακρωνύμια και συντομογραφίες

Adam Adaptive Moment Estimation

ECHA European Chemicals Agency

EADB Estrogenic Activity Database

ED Endocrine Disruptor

ER Estrogen Receptor

GCN Graph Convolutional Network

GNN Graph Neural Network

MLP Multi-Layer Perceptrons

MPNN Message Passing Neural Networks

MSE Mean-Squared Error

PReLU Parametric Rectified Linear Unit

RBA Relative Binding Affinity

REACH Registration, Evaluation, Authorisation and Restriction of Chemicals

ReLU Rectified Linear Unit

SGD Stochastic Gradient Descent

SMILES Simplified Molecular-Input Line-Entry System

QSAR Quantitative Structure Activity Relationships

Παράρτημα Β΄

Μπεϋζιανή Βελτιστοποίηση (Bayes Optimization)

Κατά βάση, τα μοντέλα μηχανικής μάθησης περιέχουν παραμέτρους που ελέγχονται από τον χρήστη και σε μεγάλο βαθμό καθορίζουν την γρηγορότερη και ποιοτικότερη εκπαίδευσή τους. Οι παράμετροι αυτές καλούνται υπερπαραμέτροι και στα μοντέλα νευρωνικών δικτύων μπορεί να είναι ο ρυθμός μάθησης, η επιλογή αλγορίθμου βελτιστοποίησης, ο αριθμός κρυφών νευρώνων ή ο αριθμός κρυφών στρωμάτων. Προκειμένου να αποφευχθεί μια εμπειρική επιλογή των παραμέτρων αυτών, η οποία ενδεχομένως θα επιφέρει μια απόδοση του μοντέλου που θα αποκλίνει από την βέλτιστη, χρησιμοποιούνται μέθοδοι αυτόματης επιλογής υπερπαραμέτρων. Οι μέθοδοι αυτές αντιμετωπίζουν το παραπάνω πρόβλημα ως ένα πρόβλημα αριστοποίησης αντικειμενικής συνάρτησης, το οποίο προσπαθούν να λύσουν αλγοριθμικά. Στην παρούσα εργασία ο αλγόριθμος που χρησιμοποιείται για την βέλτιστη επιλογή υπερπαραμέτρων στο μοντέλο MLP είναι η Μπεϋζιανή Βελτιστοποίηση (Bayesian Optimization).

Το πρόβλημα βελτιστοποίησης, το οποίο λύνεται με μπεϋζιανή βελτιστοποίηση εκφράζεται ως εξής:

$$\mathbf{x}_{opt} = \underset{\mathbf{x} \in \mathbb{X}}{\operatorname{argmax}} f(\mathbf{x}), \quad (\text{B'.1})$$

όπου $\mathbb{X} \subset \mathbb{R}^d$, $f : \mathbb{R}^d \rightarrow \mathbb{R}$ η αντικειμενική συνάρτηση προς μεγιστοποίηση, \mathbf{x} το διάνυσμα των παραμέτρων που μεταβάλλονται μέσω επαναληπτικής διαδικασίας και \mathbf{x}_{opt} το δείγμα που μεγιστοποιεί την αντικειμενική συνάρτηση f .

Η θεμελιώδης φιλοσοφία για τον υπολογισμό του ολικού μεγίστου, η οποία πηγάζει από το θεώρημα του Bayes, συνοψίζεται πίσω από την χρήση όλων των διαθέσιμων πληροφοριών από πρότερους υπολογισμούς του $f(\mathbf{x})$, δηλαδή την αξιοποίηση της πρότερης γνώσης (prior belief). Αυτό έχει ως αποτέλεσμα την μεγιστοποίηση (ή ελαχιστοποίηση ανάλογα με τον στόχο) "δύσκολων" συναρτήσεων, των οποίων η κυρτότητα μεταβάλλεται (non-convex functions), με σχετικά λίγες επαναλήψεις υπολογισμών, αλλά με κατανάλωση μεγάλης υπολογιστικής ισχύος στον καθορισμό του επόμενου σημείου δοκιμής [87]. Πράγματι, σύμφωνα με το θεώρημα του Bayes (εξίσωση B'.2), η ύστερη πιθανότητα ενός ενδεχομένου A, δεδομένης εμπειρίας B είναι ανάλογη του γινομένου της εμπειρίας B δεδομένου του ενδεχομένου A επί την πρότερη πιθανότητα του ενδεχομένου A.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (\text{B'.2})$$

Η μπεϋζιανή βελτιστοποίηση αποτελείται από δύο κύρια συστατικά: ένα μπεϋζιανό στατιστικό μοντέλο που προτυποποιεί την αντικειμενική συνάρτηση και μια συνάρτηση εκμάθησης (acquisition function). Το μπεϋζιανό στατιστικό μοντέλο είναι, ουσιαστικά, μια Γκαουσιανή Διαδικασία Παλινδρόμησης (Gaussian Process Regression). Για την περιγραφή της γκαουσιανής διαδικασίας παλινδρόμησης έστω ένα πεπερασμένο σύνολο δειγμάτων $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^d$ και το διάνυσμα τιμών της αντικειμενικής συνάρτησης που προκύπτει: $[f(\mathbf{x}_1), \dots, f(\mathbf{x}_k)]$. Κάθε φορά που μια ποσότητα είναι άγνωστη στην μπεϋζιανή στατιστική, όπως το παραπάνω διάνυσμα, θεωρείται ότι αντλήθηκε τυχαία από μια πρότερη κατανομή πιθανοτήτων. Η γκαουσιανή διαδικασία παλινδρόμησης θεωρεί ότι η πρότερη αυτή κατανομή είναι πολυμεταβλητή κανονική, με ένα συγκεκριμένο διάνυσμα μέσης τιμής και ένα συγκεκριμένο πίνακα συνδιακύμανσης. Το διάνυσμα μέσης τιμής κατασκευάζεται υπολογίζοντας τη μέση τιμή κάθε δείγματος $\mathbf{x}_i, i = 1, \dots, k$, ενώ ο πίνακας συνδιακύμανσης εφαρμόζοντας συνάρτηση συνδιακύμανσης ή πυρήνα Σ_0 σε κάθε ζευγάρι δειγμάτων $(\mathbf{x}_i, \mathbf{x}_j), i, j = 1, \dots, k$. Ο πυρήνας επιλέγεται με τέτοιο τρόπο ώστε τα δείγματα $\mathbf{x}_i, \mathbf{x}_j$ που βρίσκονται πιο κοντά στον χώρο των δειγμάτων εισόδου να έχουν μια μεγάλη θετική συσχέτιση, ενσωματώνοντας την ιδέα ότι οι αντίστοιχες τιμές της αντικειμενικής συνάρτησης πρέπει να είναι πιο παρεμφερείς σε σχέση με αντίστοιχες, των οποίων τα δείγματα βρίσκονται μακριά. Ο πυρήνας πρέπει, επίσης, να οριστεί με τέτοιο τρόπο ώστε ο πίνακας συνδιακύμανσης που παράγεται να είναι θετικά ημιορισμένος, ανεξάρτητα από το σύνολο δειγμάτων που επιλέχθηκε. Με βάση τα παραπάνω, η πρότερη κατανομή πιθανοτήτων του διανύσματος τιμών της αντικειμενικής συνάρτησης προσεγγίζεται ως εξής [88]:

$$f(\mathbf{x}_{1:k}) \sim \mathcal{N}(\boldsymbol{\mu}_0(\mathbf{x}_{1:k}), \boldsymbol{\Sigma}_0(\mathbf{x}_{1:k}, \mathbf{x}_{1:k})) \quad (\text{B'.3})$$

Έστω ότι παρατηρείται μια κατανομή $f(\mathbf{x}_{1:n})$ χωρίς θόρυβο και είναι επιθυμητός ο υπολογισμός της αντικειμενικής συνάρτησης f σε ένα νέο σημείο \mathbf{x} . Για τον σκοπό αυτό, τίθεται $k = n + 1$ και $\mathbf{x}_k = \mathbf{x}$, έτσι ώστε η πρότερη κατανομή πιθανοτήτων του διανύσματος $[f(\mathbf{x}_{1:n}), f(\mathbf{x})]$ να υπολογίζεται μέσω της εξίσωσης B'.3. Η δεσμευμένη κατανομή πιθανοτήτων $f(\mathbf{x})|f(\mathbf{x}_{1:n})$ καλείται ύστερη κατανομή πιθανοτήτων και υπολογίζεται είτε μέσω απευθείας εφαρμογής του θεωρήματος Bayes είτε γρηγορότερα και αριθμητικά σταθερότερα, μέσω αποσύνθεσης Cholesky. Η ανάλυση των παραπάνω μαθηματικών διαδικασιών είναι εκτός του πεδίου της παρούσας εργασίας.

Η χρήση της συνάρτησης εκμάθησης αποσκοπεί στην εύρεση ενός σημείου \mathbf{x} το οποίο θα στοχεύει στην βελτιστοποίηση της αντικειμενικής συνάρτησης με βάση την συνολική πρότερη γνώση. Η μπεϋζιανή βελτιστοποίηση, αρχικά, υπολογίζει την ύστερη κατανομή λαμβάνοντας υπόψιν όλα τα διαθέσιμα δείγματα και κατόπιν, ένα δείγμα \mathbf{x}_n επιλέγεται από την συνάρτηση εκμάθησης, η οποία δρα στην τρέχουσα ύστερη κατανομή. Η τιμή της συνάρτησης $f\mathbf{x}_n$ εκχωρείται στην μνήμη και το n αυξάνεται προκειμένου να ξεκινήσει μια νέα επαναληπτική διαδικασία. Τελικά, η μπεϋζιανή βελτιστοποίηση τερματίζει όταν το n γίνει ίσο με ένα οριακό N και επιστρέφει στον χρήστη το δείγμα \mathbf{x}_{opt} , δηλαδή το δείγμα για το οποίο παρατηρήθηκε η βέλτιστη τιμή της αντικειμενικής συνάρτησης [88].

Ο ψευδοκώδικας της μπεϋζιανής βελτιστοποίησης φαίνεται παρακάτω:

Μπεϋζιανή Βελτιστοποίηση με Γκαουσιανή Διαδικασία Παλινδρόμησης

Εκτέλεσε πρότερη Γκαουσιανή διαδικασία στην f χρησιμοποιώντας τα σημεία εξερεύνησης

Παρατήρησε f σε n_0 σημεία ανάλογα με τις αρχικές τιμές των παραμέτρων του πειράματος

Θέσε $n = n_0$

Απαίτησε N μέγιστο πλήθος δειγμάτων

Όσο $n \leq N$ **κάνε**:

Ενημέρωσε ύστερη κατανομή της f χρησιμοποιώντας όλα τα διαθέσιμα δεδομένα

Επίλεξε x_n από την συνάρτηση εκμάθησης, η οποία υπολογίζεται πάνω στην τρέχουσα ύστερη κατανομή

Παρατήρησε $y_n = f(x_n)$

Αύξησε n

Τέλος επανάληψης όσο

Επίστρεψε $x_{opt} = \operatorname{argmax}_{x \in \mathbb{X}} f(x)$

Βιβλιογραφία

- [1] M. I. Jordan and T. M. Mitchell (2015). *Machine learning: Trends, perspectives, and prospects*, Science, vol. 349, issue 6245, pp. 255-260, available at: <https://doi.org/10.1126/science.aaa8415>
- [2] Christopher M. Bishop (2006). *Pattern Recognition and Machine Learning*, Springer-Verlag, Berlin, Heidelberg
- [3] Liu B, Hsu W, Ma Y, Ma B (1998). *Integrating Classification and Association Rule Mining*, KDD'98: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, pp. 80-86, available at: <https://dl.acm.org/doi/10.5555/3000292.3000305>
- [4] Kevin Gurney. (1997). *An introduction to neural networks*, UCL Press
- [5] Zhang, Aston and Lipton, Zachary C. and Li, Mu and Smola, Alexander J. (2021). *Dive into Deep Learning*, arXiv preprint arXiv:2106.11342, Published online at: <https://d2l.ai/>
- [6] Sebastian Ruder (2016). *An overview of gradient descent optimization algorithms*. CoRR, abs/1609.04747, Available at: <http://arxiv.org/abs/1609.04747>
- [7] Ian Goodfellow et al. (2016). *Deep Learning*, MIT Press, Published online at: <https://www.deeplearningbook.org/>
- [8] Yann N. Dauphin and Razvan Pascanu and Caglar Gulcehre and Kyunghyun Cho and Surya Ganguli and Yoshua Bengio (2016). *Identifying and attacking the saddle point problem in high-dimensional non-convex optimization*. CoRR, abs/1406.2572, Available at: <http://arxiv.org/abs/1406.2572>
- [9] Rong Ge (2016). *Escaping from Saddle Points*. Off the convex path, Available online at: <https://www.offconvex.org/2016/03/22/saddlepoints/>
- [10] NingQian (1999). *On the momentum term in gradient descent learning algorithms*. Neural Networks, Volume 12, Issue 1, Pages 145-151, Available at: [https://doi.org/10.1016/S0893-6080\(98\)00116-6](https://doi.org/10.1016/S0893-6080(98)00116-6)
- [11] Sebastian Ruder (2016). *An overview of gradient descent optimization algorithms*. Sebastian Ruder, Available online at: <https://ruder.io/optimizing-gradient-descent/>
- [12] Bisong E. (2019). *More on Optimization Techniques*. Building Machine Learning and Deep Learning Models on Google Cloud Platform, Apress, Berkeley, CA. Available at: <https://doi.org/10.1007/978-1-4842-4470-833>
- [13] John Duchi and Elad Hazan and Yoram Singer (2011). *Adaptive Sub-gradient Methods for Online Learning and Stochastic Optimization*. Jour-

- nal of Machine Learning Research, vol.12, pg. 2121-2159. Available at: <https://dl.acm.org/doi/10.5555/1953048.2021068>
- [14] Diederik P. Kingma, Jimmy Ba (2017). *Adam: A Method for Stochastic Optimization*. CoRR, abs/1412.6980, Available at: <https://arxiv.org/abs/1412.6980>
- [15] Eric Roberts (2000). *Neural Networks: The perceptron*. Stanford University, Available at: <https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/Neuron/index.html>
- [16] Michael A. Nielsen (2015). *Neural Networks and Deep Learning*. Determination Press, Published online at: <http://neuralnetworksanddeeplearning.com/>
- [17] Chigozie Nwankpa and Winifred Ijomah and Anthony Gachagan and Stephen Marshall (2018). *Activation Functions: Comparison of trends in Practice and Research for Deep Learning*. CoRR, abs/1811.03378, Available at: <http://arxiv.org/abs/1811.03378>
- [18] John F. Kolen; Stefan C. Kremer (2001). *Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies*. in A Field Guide to Dynamical Recurrent Networks, IEEE, 2001, pp.237-243. Available at: <https://doi.org/10.1109/9780470544037.ch14>
- [19] Kaiming He and Xiangyu Zhang and Shaoqing Ren and Jian Sun (2015). *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*. CoRR, abs/1502.01852, Available at: <http://arxiv.org/abs/1502.01852>
- [20] Steve Lawrence and C. Lee Giles and Ah Chung Tsoi(1997). *Lessons in Neural Network Training: Overfitting May be Harder than Expected*. Proceedings of the Fourteenth National Conference on Artificial Intelligence, AAAI-97, AAAI Press, Menlo Park, California, pp. 540-545, 1997, Available at: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.48.8002rep=rep1type=pdf>
- [21] Ramazan Gençay and Min Qi (2001). *Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging*. IEEE Transactions on Neural Networks, vol.12, issue 4, pp. 726 - 734. Available at: <http://dx.doi.org/10.1109/72.935086>
- [22] Andrew Y. Ng (2004). *Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance*. Association for Computing Machinery. Available at: <https://doi.org/10.1145/1015330.1015435>
- [23] Christopher M. Bishop (1995). *Training with Noise is Equivalent to Tikhonov Regularization*. Neural Computation, vol.7, issue 1, pp. 108-116. Available at: <https://doi.org/10.1162/neco.1995.7.1.108>
- [24] Nitish Srivastava and Geoffrey Hinton and Alex Krizhevsky and Ilya Sutskever and Ruslan Salakhutdinov (2014). *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*. Journal of Machine Learning Research, vol.15, issue 56, pp. 1929-1958. Available at: <http://jmlr.org/papers/v15/srivastava14a.html>
- [25] G. Cybenko (1989). *Approximation by superpositions of a sigmoidal function*. Mathematics of Control, Signals, and Systems, vol.2, issue 4, pp. 303-314. Available at: <https://doi.org/10.1007/BF02551274>

- [26] Kurt Hornik (1991). *Approximation capabilities of multilayer feedforward networks*. Neural Networks, vol.4, issue 2, pp. 251-257. Available at: [https://doi.org/10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T)
- [27] Y. LeCun et al. (1989). *Backpropagation Applied to Handwritten Zip Code Recognition*. Neural Computation, vol.1, issue 4, pp. 541-551. Available at: <https://doi.org/10.1162/neco.1989.1.4.541>
- [28] Jamie Ludwig. *Image Convolution*. Satellite Digital Image Analysis, Portland State University, Available at: <http://web.pdx.edu/~jduh/courses/Archive/geog481w07/Students/LudwigImageConvolution.pdf>
- [29] Kim P. (2017). *Convolutional Neural Network*. In: MATLAB Deep Learning. Apress, Berkeley, CA. Available at: <https://doi.org/10.1007/978-1-4842-2845-66>
- [30] Anon. *Layers of a Convolutional Neural Network*. Convolutional Neural Networks for Image and Video Processing, Technische Universität München, Available at: <https://wiki.tum.de/display/lfdv/Layers+of+a+Convolutional+Neural+NetworkLayersofaConvolutionalNeuralNetwork-FullyConnectedLayer>
- [31] Sergey Ioffe and Christian Szegedy (2015). *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. CoRR, abs/1502.03167, Available at: <http://arxiv.org/abs/1502.03167>
- [32] Connor Shorten and Taghi M. Khoshgoftaar (2019). *A survey on Image Data Augmentation for Deep Learning*. Journal of Big Data, Vol. 6, Article number: 60. Available at: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0>
- [33] Jonathan L. Gross, Jay Yellen, Mark Anderson (2018). *Graph Theory and Its Applications*. Chapman and Hall, Boca Raton. Available at: <https://doi.org/10.1201/9780429425134>
- [34] Anon (2019). *Molecular Descriptors*. Available at: <https://chem.libretexts.org/@go/page/192626>
- [35] Anon. *Extended Connectivity Fingerprint ECFP*. Available at: <https://docs.chemaxon.com/display/docs/extended-connectivity-fingerprint-ecfp.mdsrc-1806333-extendedconnectivityfingerprintecfp-propertiesofecfps>
- [36] Rogers D and Hahn M (2010). *Extended-connectivity fingerprints*. Journal of chemical information and modeling, Vol. 50, Issue: 5, pp. 742-754. Available at: <https://doi.org/10.1021/ci100050t>
- [37] David Duvenaud and Dougal Maclaurin and Jorge Aguilera (2019). *Convolutional Networks on Graphs for Learning Molecular Fingerprints*. CoRR, abs/1509.09292. Available at: <http://arxiv.org/abs/1509.09292>
- [38] Zonghan Wu and Shirui Pan and Fengwen Chen and Guodong Long and Chengqi Zhang and Philip S. Yu (2019). *A Comprehensive Survey on Graph Neural Networks*. CoRR, abs/1901.00596. Available at: <http://arxiv.org/abs/1901.00596>
- [39] Jie Zhou and Ganqu Cui and Zhengyan Zhang and Cheng Yang and Zhiyuan Liu and Maosong Sun (2018). *Graph Neural Networks: A Review of Methods and Applications*. CoRR, abs/1812.08434. Available at: <http://arxiv.org/abs/1812.08434>

- [40] Justin Gilmer and Samuel S. Schoenholz and Patrick F. Riley and Oriol Vinyals and George E. Dahl (2017). *Neural Message Passing for Quantum Chemistry*. CoRR, abs/1704.01212. Available at: <http://arxiv.org/abs/1704.01212>
- [41] Thomas N. Kipf and Max Welling (2016). *Semi-Supervised Classification with Graph Convolutional Networks*. CoRR, abs/1609.02907. Available at: <http://arxiv.org/abs/1609.02907>
- [42] Michael Defferrard and Xavier Bresson and Pierre Vandergheynst (2016). *Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering*. CoRR, abs/1606.09375. Available at: <http://arxiv.org/abs/1606.09375>
- [43] David K Hammond, Pierre Vandergheynst, Rémi Gribonval (2009). *Wavelets on Graphs via Spectral Graph Theory*. CoRR, abs/0912.3848. Available at: <https://arxiv.org/abs/0912.3848>
- [44] Rex Ying and Jiaxuan You and Christopher Morris and Xiang Ren and William L. Hamilton and Jure Leskovec (2018). *Hierarchical Graph Representation Learning with Differentiable Pooling*. CoRR, abs/1806.08804. Available at: <http://arxiv.org/abs/1806.08804>
- [45] Jiaxuan You and Rex Ying and Jure Leskovec (2020). *Design Space for Graph Neural Networks*. CoRR, abs/2011.08843. Available at: <https://arxiv.org/abs/2011.08843>
- [46] JA. Crum-Brown, T.R. Fraser (1868-1869). *On the connection between chemical constitution and physiological action. Part 1. On the physiological action of the ammonium bases*. Trans. R. Soc. Edinburgh, 25, 151
- [47] Meyer, H (1899). Naunyn-Schmiedeberg's Arch. Exp. Patkol. Pharmakof. 42,109-118
- [48] C.E. Overton (1901). Studien Uber die Narkose Fischer, Jena, Germany.
- [49] Hammett LP. (1937). *The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives..* J Am Chem Soc. Available at: <https://doi.org/10.1021/ja01280a022>
- [50] Corwin. Hansch and Toshio. Fujita (1964). *$\rho - \sigma - \pi$ Analysis. A Method for the Correlation of Biological Activity and Chemical Structure*. J. Am. Chem. Soc., vol. 86, issue 8, pp. 1616-1626. Available at: <https://doi.org/10.1021/ja01062a035>
- [51] Cherkasov, Artem Muratov, Eugene Fourches, Denis Varnek, Alexandre Baskin, Igor Cronin, Mark Dearden, John Gramatica, Paola Martin, Yvonne Todeschini, Roberto Consonni, Viviana Kuz'min, Victor Cramer, Richard Benigni, Romualdo Yang, Chihae Rathman, James Terfloth, Lothar Gasteiger, Johann Richard, Ann Tropsha, Alexander (2013). *QSAR Modeling: Where have you been? Where are you going to?*. Journal of Medicinal Chemistry, vol. 57, issue 12. Available at: <http://dx.doi.org/10.1021/jm4004285>
- [52] Gupta, Manish K Gupta, Swati Rawal, Ravindra (2016). *Impact of Artificial Neural Networks in QSAR and Computational Modeling*. In book: Artificial Neural Network for Drug Design, Delivery and Disposition (pp.153-179) Available at: <http://dx.doi.org/10.1016/B978-0-12-801559-9.00008-9>

- [53] askin, Igor Palyulin, Vladimir Zefirov, Nikolai (2008). *Neural Networks in Building QSAR Models*. In book: *Artificial Neural Networks Methods and Applications* (pp.133-154) Available at: <http://dx.doi.org/10.1007/978-1-60327-101-18>
- [54] Koutsoukas, A., Monaghan, K.J., Li, X. et al. (2017). *Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data*. *Journal of Cheminformatics*, vol. 9. issue 42. Available at: <https://doi.org/10.1186/s13321-017-0226-y>
- [55] George E. Dahl and Navdeep Jaitly and Ruslan Salakhutdinov (2014). *Multi-task Neural Networks for QSAR Predictions*. CoRR, abs/1406.1231. Available at: <https://arxiv.org/abs/1406.1231>
- [56] Tsou, L.K., Yeh, SH., Ueng, SH. et al. (2020). *Comparative study between deep learning and QSAR classifications for TNBC inhibitors and novel GPCR agonist discovery*. *Scientific Reports*, vol. 10. issue: 1. Available at: <https://doi.org/10.1038/s41598-020-73681-1>
- [57] Chakravarti, S. K., Alla, S. (2019). *Descriptor Free QSAR Modeling Using Deep Learning With Long Short-Term Memory Neural Networks*. *Frontiers in artificial intelligence*, vol. 2. issue: 17. Available at: <https://doi.org/10.3389/frai.2019.00017>
- [58] Matsuzaka, Y., Uesawa, Y. (2021). *Molecular Image-Based Novel Quantitative Structure-Activity Relationship Approach, Deepsnap-Deep Learning and Machine Learning*. *Current issues in molecular biology*, vol. 42. pp. 455–472. Available at: <https://doi.org/10.21775/cimb.042.455>
- [59] Hung, C., Gini, G. (2021). *QSAR modeling without descriptors using graph convolutional neural networks: the case of mutagenicity prediction*. *Molecular diversity*, vol. 25, issue: 3 pp. 1283–1299. Available at: <https://doi.org/10.1007/s11030-021-10250-2>
- [60] Joseph D. Romano, Yun Hao, Jason H. Moore (2021). *Improving QSAR Modeling for Predictive Toxicology using Publicly Aggregated Semantic Graph Data and Graph Neural Networks*. bioRxiv 2021.08.08.455550. Available at: <https://doi.org/10.1101/2021.08.08.455550>
- [61] David Weininger (1988). *SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules*. *Journal of Chemical Information and Computer Sciences*, vol. 28, issue: 1, pp. 31-36. Available at: <https://doi.org/10.1021/ci00057a005>
- [62] Anon. *SMILES - A Simplified Chemical Language*. Available at: <https://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>
- [63] R. Thomas Zoeller, T. R. Brown, L. L. Doan, A. C. Gore, N. E. Skakkebaek, A. M. Soto, T. J. Woodruff, F. S. Vom Saal (2012). *Endocrine-Disrupting Chemicals and Public Health Protection: A Statement of Principles from The Endocrine Society*. *Endocrinology*, Volume 153, Issue 9, pp. 4097-4110. Available at: <https://doi.org/10.1210/en.2012-1422>
- [64] T Colborn, F S vom Saal, and A M Soto (1993). *Developmental effects of endocrine-disrupting chemicals in wildlife and humans*. *Environmental Health Perspectives*, Volume 101, Issue 5, pp. 378–384. Available at: <https://dx.doi.org/10.1289%2Fehp.93101378>

- [65] Yilmaz, B., Terekeci, H., Sandal, S., Kelestimur, F. (2020). *Endocrine disrupting chemicals: exposure, effects on human health, mechanism of action, models for testing and strategies for prevention*. Reviews in endocrine metabolic disorders, Volume 21, Issue 1, pp. 127–147. Available at: <https://doi.org/10.1007/s11154-019-09521-z>
- [66] Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter (2014). *Essential Cell Biology 4th Edition*. Garland Publishing
- [67] Shen, J., Xu, L., Fang, H., Richard, A. M., Bray, J. D., Judson, R. S., Zhou, G., Colatsky, T. J., Aungst, J. L., Teng, C., Harris, S. C., Ge, W., Dai, S. Y., Su, Z., Jacobs, A. C., Harrouk, W., Perkins, R., Tong, W., Hong, H. (2013). *EADB: an estrogenic activity database for assessing potential endocrine activity*. Toxicological sciences : an official journal of the Society of Toxicology, vol. 135, issue 2, pp. 277–291. Available at: <https://doi.org/10.1093/toxsci/kft164>
- [68] Watson, C. S., Jeng, Y. J., Guptarak, J. (2011). *Endocrine disruption via estrogen receptors that participate in nongenomic signaling pathways*. The Journal of steroid biochemistry and molecular biology, vol. 127, issues 1-2, pp. 44–50. Available at: <https://doi.org/10.1016/j.jsbmb.2011.01.015>
- [69] Schecter, A., Haffner, D., Colacino, J., Patel, K., Pöpke, O., Opel, M., Birnbaum, L. (2010). *Polybrominated diphenyl ethers (PBDEs) and hexabromocyclodecane (HBCD) in composite U.S. food samples*. Environmental health perspectives, vol. 118, issue 3, pp. 357–362. Available at: <https://doi.org/10.1289/ehp.0901345>
- [70] Paterni, I., Granchi, C., Minutolo, F. (2017). *Risks and benefits related to alimentary exposure to xenoestrogens*. Critical reviews in food science and nutrition, vol. 57, issue 16, pp. 3384–3404. Available at: <https://doi.org/10.1080/10408398.2015.1126547>
- [71] Wagner M. (2017). *Know thy unknowns: why we need to widen our view on endocrine disruptors*. Journal of Epidemiology and Community Health, vol. 71, pp. 209–212. Available at: <http://dx.doi.org/10.1136/jech-2016-207259>
- [72] Margherita Grandini and Enrico Bagli and Giorgio Visani (2020). *Metrics for Multi-Class Classification: an Overview*. CoRR, abs/2008.05756. Available at: <https://arxiv.org/abs/2008.05756>
- [73] Douglas M. Hawkins, Subhash C. Basak, and Denise Mills (2003). *Assessing Model Fit by Cross-Validation*. Journal of Chemical Information and Computer Sciences, vol. 43, issue 2, pp. 579–586. Available at: <https://doi.org/10.1021/ci025626i>
- [74] Καρατζάς Παντελής (2020). *Ταξινόμηση χημικών δομών με χρήση βαθιάς μηχανικής μάθησης*. Εθνικό Μετσόβιο Πολυτεχνείο, Σχολή Εφαρμοσμένων Μαθηματικών και Επιστημών, Αθήνα
- [75] Willighagen, E.L., Mayfield, J.W., Alvarsson, J. et al. (2017). *The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching*. Journal of Cheminformatics, vol. 9, issue 33. Available at: <https://doi.org/10.1186/s13321-017-0220-4>
- [76] Moriwaki, H., Tian, YS., Kawashita, N. et al. (2018). *Mordred: a molecular descriptor calculator*. Journal of Cheminformatics, vol. 10, issue 4. Available at: <https://doi.org/10.1186/s13321-018-0258-y>

- [77] Greg Landrum (2019). *RDKit Documentation*. Available at: <https://buildmedia.readthedocs.org/media/pdf/rdkit/latest/rdkit.pdf>
- [78] Keras. Available at: <https://keras.io/about/>
- [79] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng (2016). *TensorFlow: A System for Large-Scale Machine Learning*. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) Available at: <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>
- [80] *scikit-optimize: Sequential model-based optimization in Python*. Available at: <https://scikit-optimize.github.io/stable/>
- [81] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, vol. 12, issue 85, pp. 28252830 Available at: <https://www.jmlr.org/papers/v12/pedregosa11a.html>
- [82] Ramsundar, Bharath (2018). *Molecular machine learning with DeepChem*. Stanford University, Computer Science Department Available at: <https://purl.stanford.edu/js264hd4826>
- [83] Christopher M. Bishop (1995). *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford
- [84] Daniele Grattarola and Cesare Alippi (2020). *Graph Neural Networks in TensorFlow and Keras with Spektral*, CoRR, abs/2006.12138. Available at: <https://arxiv.org/abs/2006.12138>
- [85] William L. Hamilton and Rex Ying and Jure Leskovec (2017). *Inductive Representation Learning on Large Graphs*, CoRR, abs/1706.02216. Available at: <https://arxiv.org/abs/1706.02216>
- [86] Keyulu Xu and Weihua Hu and Jure Leskovec and Stefanie Jegelka (2018). *How Powerful are Graph Neural Networks?*, CoRR, abs/1810.00826. Available at: <https://arxiv.org/abs/1810.00826>
- [87] Jasper Snoek and Hugo Larochelle and Ryan P. Adams (2012). *Practical Bayesian Optimization of Machine Learning Algorithms*, CoRR, abs/1206.2944. Available at: <https://arxiv.org/abs/1206.2944>
- [88] Peter I. Frazier (2018). *A Tutorial on Bayesian Optimization*, CoRR, abs/1807.02811. Available at: <https://arxiv.org/abs/1807.02811>