Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρικών Βιομηχανικών Διαταξεων &
Συστηματων Αποφασεων

Καθορισμός επενδύσεων με χρήση αλγορίθμων μηχανικής και ενισχυτικής μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΑΥΡΟΓΙΑΝΝΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ-ΓΕΩΡΓΙΟΣ

Επιβλέπων : Ασημακόπουλος Βασίλειος
Καθηγητής Ε.Μ.Π.

Υπεύθυνος : Καλτσούνης Αναστάσιος
Υποψήφιος Διδάκτωρ Ε.Μ.Π
Σπηλιώτης Ευάγγελος
Διδάκτωρ Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2024

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρικών Βιομηχανικών Διαταξεων &
Συστηματων Αποφασεων

Καθορισμός επενδύσεων με χρήση αλγορίθμων μηχανικής και ενισχυτικής μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΑΥΡΟΓΙΑΝΝΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ-ΓΕΩΡΓΙΟΣ

Επιβλέπων : Ασημακόπουλος Βασίλειος
Καθηγητής Ε.Μ.Π.

Υπεύθυνος : Καλτσούνης Αναστάσιος
Υποψήφιος Διδάκτωρ Ε.Μ.Π
Σπηλιώτης Ευάγγελος
Διδάκτωρ Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την ΗΜΕΡΟΜΗΝΙΑ

(Υπογραφή)	(Υπογραφή)	(Υπογραφή)
……	……	……
Βασίλειος Ασημακόπουλος	Ιωάννης Ψαρράς	Δημήτριος Ασκούνης

Αθήνα, Φεβρουάριος 2024

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων & Συστημάτων Αποφάσεων

(Υπογραφή)

.....

Μαυρογιάννης Κωνσταντίνος - Γεώργιος, Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Ηλεκτρονικών Υπολογιστών

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια των ερευνητικών δραστηριοτήτων της Μονάδας Προβλέψεων και Στρατηγικής του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου Αθηνών κατά το ακαδημαϊκό έτος 2023-2024. Η μονάδα υπάγεται στον Τομέα Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων της σχολής.

Θα ήθελα να ευχαριστήσω θερμα:

Τον επιβλέποντα καθηγητή κ. Βασίλειο Ασημακόπουλο για την ανάθεση και επίβλεψη της διπλωματικής αυτής.

Τον Καθηγητή κ. Ι. Ψαρρά και τον Καθηγητή κ. Δημήτρη Ασκούνη για τη συμμετοχή τους στην εξεταστική επιτροπή της εργασίας.

Τους υπευθύνους της διπλωματικής εργασίας, τον κ. Ε. Σπηλιώτη, Διδάκτωρ Ε.Μ.Π, και ιδιαίτερα τον κ. Α. Καλτσούνη, Υποψήφιο Διδάκτορα Ε.Μ.Π, για την εύρυθμη συνεργασία και την άμεση διαθεσιμότητα τους καθ'ολη την διάρκεια της διαδικασίας.

Τέλος, τους δικούς μου ανθρώπους και την οικογένειά μου που ήταν μαζί μου σε όλη μου την ακαδημαϊκή διαδρομή.

Περίληψη

Η παρούσα διπλωματική εργασία έχει σαν στόχο την ανάπτυξη μεθοδολογίας για τον σχεδιασμό και την υλοποίηση ενός πράκτορα Ενισχυτικής Μάθησης ο οποίος είναι ικανός να διαχειρίζεται και να βελτιστοποιεί χαρτοφυλάκια μετοχών αυτοματοποιημένα.

Η βελτιστοποίηση χαρτοφυλακίου είναι ένας κλάδος που απασχολεί ερευνητές εδώ και δεκαετίες με τη δημιουργία ενός μοντέλου ή αλγορίθμου που μεγιστοποιεί το κέρδος και ελαχιστοποιεί το ρίσκο να είναι στο κέντρο του ενδιαφέροντος. Με την άνθηση της επιστήμης των υπολογιστών και συγκεκριμένα της Μηχανικής Μάθησης, δημιουργήθηκαν νέοι μέθοδοι ανάλυσης δεδομένων ικανοί να επεξεργαστούν και να αξιοποιήσουν τεράστιους όγκους δεδομενων για πρώτη φορά αποδοτικά. Η επανάσταση αυτή οδήγησε στην μελέτη τρόπων που θα μπορούσαν αυτοί οι μέθοδοι να ενταχθούν στον τομέας της βελτιστοποίησης χαρτοφυλακίου. Η Ενισχυτική Μάθηση, δηλαδή η εκπαίδευση ενός πράκτορα με σκοπό τη λήψη βέλτιστων αποφάσεων, αποτελεί μία από τις πιο σημαντικές εξελίξεις στον χώρο της Μηχανικής Μάθησης και είναι ένα από τα πιο υποσχόμενα εργαλεία στο πεδίο αυτό.

Μέσα στην εργασία γίνεται μια ανασκόπηση των παραδοσιακών μεθόδων διαχείρισης χαρτοφυλακίου και των μετρικών αξιολόγησης αυτών. Κατόπιν, η εργασία εστιάζει στις βασικές τεχνικές Μηχανικής Μάθησης και ειδικότερα στην Ενισχυτική Μάθηση. Μέσα από μια σειρά πειραμάτων, αναπτύσσεται και δοκιμάζεται η μεθοδολογία που επιτρέπει την ανάπτυξη ενός πράκτορα Ενισχυτικής Μάθησης, ικανό να παίρνει αποφάσεις με σκοπό τη μεγιστοποίηση του κέρδους.

Αρχικά, παρουσιάζονται αναλυτικά οι σχεδιαστικές επιλογές όπως ο αλγόριθμος Διπλών Βαθιών Q-Δικτύων (DDQN) και η επιλογή του χαρτοφυλακίου τεχνολογικών μετοχών που θα μελετηθεί. Στη συνέχεια, εκτελείται πληθώρα πειραμάτων τα οποία έχουν ως σκοπό την μελέτη των διαφορετικών παραμέτρων του πράκτορα και την βελτιστοποίηση τους. Συγκεκριμένα, στα πρώτα πειράματα γίνεται μελέτη της επίδρασης των υπερ-παραμέτρων εκπαίδευσης και επιλέγονται οι βέλτιστες, αξιολογώντας διαφορετικές μετρικές. Τα επόμενα πειράματα έχουν ως σκοπό την εύρεση της βέλτιστης κατάστασης-εισόδου του συστήματος χρησιμοποιώντας τις βέλτιστες υπερ-παραμέτρους εκπαίδευσης που βρέθηκαν προηγουμένως. Στη συνέχεια, μελετάται η επίδραση της αρχιτεκτονικής του δικτύου και συγκεκριμένα, γίνεται σύγκριση της προβλεπτικής ικανότητας των Συνελικτικών Νευρωνικών Δικτύων (CNN) και των Δικτύων Προς τα Εμπρός (FFN). Τέλος, για την επαλήθευση των αποτελεσμάτων δοκιμάζεται ο βελτιστοποιημένος πράκτορας έναντι νέων χαρτοφυλακίων με διαφορετικά χαρακτηριστικά και τα αποτελέσματα συγκρίνονται με τα αρχικά.

Λέξεις κλειδιά: Ενισχυτική Μάθηση, Βελτιστοποίηση Χαρτοφυλακίου, Εμπόριο Μετοχών

Abstract

This thesis aims to develop a general methodology for the design and implementation of a Reinforcement Learning agent capable of managing and optimizing stock portfolios automatically.

Portfolio optimization is a topic that has occupied researchers for decades, having as a main goal the creation of a model or algorithm that maximizes profit and minimizes the risk. With the boom in computer science and specifically Machine Learning, new data analysis methods were created capable of processing and exploiting massive amounts of data efficiently for the first time. This revolution led to the study of ways in which these methods could be incorporated into the field of portfolio optimization. Reinforcement Learning, i.e. training an agent to make optimal decisions, is one of the most important developments in the field of Machine Learning and is one of the most promising tools in this field.

The paper reviews traditional portfolio management methods and their evaluation metrics. Then, the paper focuses on the basic Machine Learning techniques and in particular on Reinforcement Learning. Through a series of experiments, the methodology is developed and tested that allows the development of a Reinforcement Learning agent, capable of making decisions in order to maximize profit.

First, design options such as the Double Deep Q-Networks (DDQN) algorithm and the selection of the technology stock portfolio to be studied are presented in detail. Afterwards, a multitude of experiments are performed which aim to study the different parameters of the agent and optimize them. Specifically, in the first experiments, the effect of training hyper-parameters is studied and the optimal ones are selected, evaluating different metrics and graphical representations. The following experiments are aimed at finding the optimal input-state of the system using the previously found optimal training hyper-parameters. Then, the effect of the network architecture is studied and specifically, the predictive ability of Convolutional Neural Networks (CNN) and Feedforward Networks (FFN) is compared. Finally, to verify the results, the optimized agent is tested against new portfolios with different characteristics and the results are compared with the original ones.

Keywords: Reinforcement Learning, Portfolio Optimization, Stock Trading

Πίνακας Περιεχομένων

Κεφάλαιο 1. Εισαγωγή

1.1 Αντικείμενο εργασίας

Οι χρηματιστηριακές συναλλαγές αποτελούν τον ακρογωνιαίο λίθο των παγκόσμιων χρηματοπιστωτικών αγορών για αιώνες, προσφέροντας σε άτομα και ιδρύματα την ευκαιρία να επενδύσουν σε εταιρείες και να δημιουργήσουν πλούτο. Ωστόσο, η επιστήμη της αγοραπωλησίας μετοχών και της διαχείρισης χαρτοφυλακίου έχουν εξελιχθεί σημαντικά με τα χρόνια δημιουργώντας ένα όλο και πιο ανταγωνιστικό περιβάλλον. Οι επενδυτές αντιμετωπίζουν τώρα ένα ιδιαίτερα περίπλοκο τοπίο χρηματοπιστωτικών μέσων, δυναμικής της αγοράς και οικονομικών παραγόντων που μπορούν να επηρεάσουν την απόδοση των επενδύσεών τους.

Η βελτιστοποίηση χαρτοφυλακίου, μια θεμελιώδης πτυχή των σύγχρονων χρηματιστηριακών επενδύσεων, στοχεύει στην επίτευξη ισορροπίας μεταξύ κινδύνου και απόδοσης, δημιουργώντας ένα βέλτιστο μείγμα περιουσιακών στοιχείων σε ένα επενδυτικό χαρτοφυλάκιο. Ο στόχος είναι να μεγιστοποιηθούν οι αποδόσεις με παράλληλη άμβλυνση των κινδύνων που σχετίζονται με τις διακυμάνσεις της αγοράς. Η επίτευξη αυτής της ισορροπίας δεν είναι εύκολο κατόρθωμα και παραδοσιακά βασίζεται στην ανθρώπινη τεχνογνωσία και στα οικονομικά μοντέλα.

Τα τελευταία χρόνια, το τοπίο της διαπραγμάτευσης μετοχών και της βελτιστοποίησης χαρτοφυλακίου έχει μεταμορφωθεί με την ενσωμάτωση τεχνικών αυτοματισμού, Τεχνητής Νοημοσύνης (Artificial Intelligence - AI), Μηχανικής Μάθησης (Machine Learning - ML) και Ενισχυτικής Μάθησης (Reinforcement Learning - RL). Αυτές οι τεχνολογίες επέτρεψαν την ανάπτυξη ευφυών πρακτόρων, ικανών να λαμβάνουν αποφάσεις βάσει δεδομένων σε πραγματικό χρόνο, εγκαινιάζοντας μια νέα εποχή αυτοματοποιημένης διαπραγμάτευσης μετοχών και διαχείρισης χαρτοφυλακίου. Οι αυτοματοποιημένες συναλλαγές, που συχνά αναφέρονται ως αλγοριθμικές συναλλαγές (algorithmic trading), αξιοποιούν προγράμματα υπολογιστών για την εκτέλεση συναλλαγών με ταχύτητα και ακρίβεια που ξεπερνούν τις ανθρώπινες δυνατότητες. Αυτοί οι αλγόριθμοι μπορούν να αναλύουν τεράστια σύνολα δεδομένων, να εντοπίζουν μοτίβα και να εκτελούν συναλλαγές σε κλάσματα δευτερολέπτου, καθιστώντας τους ιδιαίτερα κατάλληλους τόσο σε περιβάλλοντα όπου η ταχύτητα είναι απαραίτητη για την επίτευξη κέρδους, όπως το εμπόριο υψηλής συχνότητας (high frequency trading) αλλά και σε περιβάλλοντα όπου η λήψη απόφασης είναι δύσκολη λόγω του μεγάλου όγκου δεδομένων.

Η βελτιστοποίηση χαρτοφυλακίου, ως αναπόσπαστο μέρος αυτής της επανάστασης αυτοματισμού, επιδιώκει να αξιοποιήσει τη δύναμη των αλγορίθμων για τη δημιουργία χαρτοφυλακίων που ευθυγραμμίζονται με συγκεκριμένους επενδυτικούς στόχους. Είτε ο στόχος είναι η διατήρηση του κεφαλαίου είτε η δημιουργία εισοδήματος, οι αυτοματοποιημένοι αλγόριθμοι βελτιστοποίησης χαρτοφυλακίου στοχεύουν στη δημιουργία ενός διαφοροποιημένου

μείγματος περιουσιακών στοιχείων που μεγιστοποιούν τις αποδόσεις ελαχιστοποιώντας παράλληλα την έκθεση σε αδικαιολόγητους κινδύνους.

Για να γίνουν καλύτερα κατανοητές οι επαναστατικές δυνατότητες της Τεχνητής Νοημοσύνης , της Μηχανικής Μάθησης και της Ενισχυτικής Μάθησης στον τομέα της αυτοματοποιημένης διαπραγμάτευσης μετοχών και της βελτιστοποίησης χαρτοφυλακίου, είναι απαραίτητο να κατανοήσουμε τις θεμελιώδεις αρχές αυτών των τεχνολογιών.

Η Τεχνητή Νοημοσύνη, ή AI, αναφέρεται στη δημιουργία ευφυών πρακτόρων ή συστημάτων που μπορούν να εκτελέσουν εργασίες που απαιτούν συνήθως ανθρώπινη νοημοσύνη. Αυτό περιλαμβάνει ένα ευρύ φάσμα τεχνικών, από πιο απλά συστήματα βασισμένα σε κανόνες έως τη Μηχανική Μάθηση. Στον πυρήνα της, η Τεχνητή Νοημοσύνη επιδιώκει να μιμηθεί την ανθρώπινη διαδικασία μάθησης για να λαμβάνει τεκμηριωμένες αποφάσεις σε διάφορους τομείς.

Η Μηχανική Μάθηση είναι ένα υποσύνολο της Τεχνητής Νοημοσύνης που εστιάζει στην ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν στα συστήματα να μαθαίνουν και να βελτιώνονται από δεδομένα. Σε αντίθεση με τον παραδοσιακό προγραμματισμό, όπου οι κανόνες προβλέπουν ξεκάθαρα τον τρόπο εκτέλεσης του προγράμματος, τα μοντέλα Μηχανικής Μάθησης μπορούν να αναγνωρίσουν μοτίβα και σχέσεις στα δεδομένα, επιτρέποντάς τους να κάνουν προβλέψεις ή αποφάσεις χωρίς ρητό προγραμματισμό. Η Επιβλεπόμενη (Supervised), η μη επιβλεπόμενη (Unsupervised) και η Ενισχυτική Μάθηση είναι όλες υποκατηγορίες της Μηχανικής Μάθησης.

Η Ενισχυτική Μάθηση, όπου είναι και το αντικείμενο μελέτης αυτής της εργασίας, είναι η διαδικασία μάθησης όπου ένας πράκτορας αλληλεπιδρά με ένα περιβάλλον για να μεγιστοποιήσει μια σωρευτική ανταμοιβή. Μέσω δοκιμής και λάθους, ο πράκτορας μαθαίνει τις βέλτιστες στρατηγικές και πολιτικές για την επίτευξη των στόχων του. Στο πλαίσιο της διαπραγμάτευσης μετοχών, η Ενισχυτική Μάθηση προσφέρει μεγάλη προσαρμοστικότητα και τη δυνατότητα βελτιστοποίησης χαρτοφυλακίων σε δυναμικές, αβέβαιες αγορές. Οι παραδοσιακές μέθοδοι βελτιστοποίησης χαρτοφυλακίου βασίζονται συχνά σε στατικά μοντέλα που μπορεί να μην προσαρμόζονται καλά στη συνεχώς μεταβαλλόμενη δυναμική των χρηματοπιστωτικών αγορών. Αντίθετα, η Ενισχυτική Μάθηση προσφέρει ένα δυναμικό πλαίσιο που επιτρέπει στους πράκτορες να μαθαίνουν και να προσαρμόζουν τις στρατηγικές τους σε πραγματικό χρόνο, καθιστώντας το ένα πολλά υποσχόμενο εργαλείο για τη βελτιστοποίηση των χαρτοφυλακίων μετοχών.

Η εργασία αυτή θα εμβαθύνει στην πρακτική εφαρμογή αλγορίθμων Ενισχυτικής Μάθησης για βελτιστοποίηση χαρτοφυλακίου, διερευνώντας θέματα όπως αναπαραστάσεις κατάστασης και ενεργειών, συναρτήσεις ανταμοιβής και τις προκλήσεις που σχετίζονται με την ανάπτυξη στον πραγματικό κόσμο. Στόχος είναι, να αναπτυχθεί μια γενικότερη μεθοδολογία κατασκευής ενός πράκτορα Ενισχυτικής Μάθησης για βελτιστοποίηση χαρτοφυλακίου καθώς και αξιολόγησης

του. Τέλος, εφαρμόζοντας την μεθοδολογία που αναπτύχθηκε σε ένα σύνολο δεδομένων, θα εξετάσουμε εμπειρικά αποτελέσματα μέσω πινάκων και γραφικών παραστάσεων για να ρίξουμε φως στην αποτελεσματικότητα και τους πιθανούς περιορισμούς της Ενισχυτικής Μάθησης στη βελτιστοποίηση των χαρτοφυλακίων μετοχών.

1.2 Δομή εργασίας

Στο δεύτερο Κεφάλαιο, εξηγούνται τα βασικά για το εμπόριο μετοχών. Επιπλέον, γίνεται μια ανασκόπηση στα βασικές μεθόδους εμπορίας μετοχών και τα κύρια χαρακτηριστικά της κάθε μεθόδου ενώ παρουσιάζονται και οι πιο συνήθης μετρικές αξιολόγησης επενδυτικών κινήσεων. Στη συνέχεια, γίνεται εισαγωγή στην έννοια της διαχείρισης χαρτοφυλακίου μετοχών και μελετώνται οι παραδοσιακο τρόποι βελτιστοποίησης ενός χαρτοφυλακίου.

Το τρίτο Κεφάλαιο αποτελεί εισαγωγή στην Μηχανική Μάθηση. Αρχικά, παρουσιάζονται βασικές έννοιες της Μηχανικής Μάθησης και κάποιοι βασικοί αλγόριθμοί της. Στη συνέχεια, ακολουθεί μια περιγραφή των Νευρωνικών Δικτύων και του τρόπου λειτουργία τους. Έπειτα, γίνεται εκτενής ανάλυση γύρω από την Ενισχυτική Μάθηση, τα βασικά χαρακτηριστικά της και τους πιο σημαντικούς αλγορίθμους της. Τέλος, παρουσιάζονται και αναλύονται σύγχρονοι μέθοδοι διαχείρισης χαρτοφυλακίου βασισμένοι στην ανάλυση δεδομένων και την Μηχανική Μάθηση.

Στο τέταρτο Κεφάλαιο παρουσιάζεται αναλυτικά η προτεινόμενη μεθοδολογία δημιουργίας ενός πράκτορα Ενισχυτικής Μάθησης που διαχειρίζεται και βελτιστοποιεί ένα χαρτοφυλάκιο μετοχών. Η μεθοδολογία περιλαμβάνει τον ορισμό του περιβάλλοντος του πράκτορα, την επεξεργασία των δεδομένων εισόδου, την επιλογή του αλγορίθμου και τέλος των μετρικών αξιολόγησης των πειραμάτων.

Στο πέμπτο Κεφάλαιο γίνεται ανάλυση και παρουσίαση των σχεδιαστικών επιλογών και των πειραματικών αποτελεσμάτων. Γίνεται διερεύνηση της επίδρασης των βέλτιστων υπερ-παραμέτρων Ενισχυτικής Μάθησης και προσδιορίζονται οι βέλτιστες τιμές τους. Επιπλέον, μελετάται η επίδραση της κατάστασης-εισόδου στην αποτελεσματικότητα του πράκτορα κατά τη διαδικασία εκπαίδευσης αλλά και στην ικανότητά του να γενικεύει σε άγνωστα δεδομένα με σκοπό την επιλογή της βέλτιστης εισόδου. Διερευνάται επίσης η επίδραση της αρχιτεκτονικής στην αποτελεσματικότητα του πράκτορα δοκιμάζοντας διαφορετικές αρχιτεκτονικές και επιλέγοντας τη βέλτιστη. Τέλος, επαληθεύεται η βελτιστοποιημένη διαδικασία μάθησης έναντι διαφορετικών χαρτοφυλακίων ώστε να αξιολογηθεί η ικανότητα του πράκτορα να εκπαιδεύεται σε δεδομένα με διαφορετικά χαρακτηριστικά.

Τέλος, στο έκτο Κεφάλαιο, παρουσιάζονται τα συμπεράσματα της παραπάνω μελέτης, και επιπλέον γίνονται προτάσεις για περαιτέρω βελτίωσης των επιδόσεων του πράκτορα.

Κεφάλαιο 2. Εμπόριο μετοχών και διαχείριση χαρτοφυλακίου

2.1 Εισαγωγή

Το εμπόριο μετοχών (stock trading) αναφέρεται στην αγοραπωλησία μεριδίων δημόσιων εταιρειών. Κάθε μερίδιο, που αντιπροσωπεύει ένα μικρό μέρος της ιδιοκτησίας, εμπορεύεται στη χρηματιστηριακή αγορά, με την τιμή να καθορίζεται από την προσφορά και τη ζήτηση. Η χρηματιστηριακή αγορά αποτελεί ουσιαστικό μέρος του χρηματοπιστωτικού συστήματος, διαδραματίζοντας καθοριστικό ρόλο στην κατανομή πόρων, την παροχή ρευστότητας και τη διαχείριση κινδύνου.

Η ιστορία του εμπορίου μετοχών μπορεί να αναδρομολογηθεί στον 17ο αιώνα, με την ίδρυση του Χρηματιστηρίου του Άμστερνταμ, που είναι γνωστό σήμερα ως Euronext Amsterdam. Ήταν σε αυτό το χρηματιστήριο που τα μερίδια της Ολλανδικής Ανατολικής Ινδικής Εταιρείας (VOC) διαπραγματεύονταν ενεργά, σηματοδοτώντας τη γέννηση της πρώτης επίσημης αγοράς μετοχών.

Στις Ηνωμένες Πολιτείες, το New York Stock Exchange (NYSE) ιδρύθηκε το 1792, καθιστώντας το ένα από τα παλαιότερα και μεγαλύτερα χρηματιστήρια σε παγκόσμιο επίπεδο. Το Χρηματιστήριο της Νέας Υόρκης δημιουργήθηκε για να προσφέρει μια κεντρική τοποθεσία για το εμπόριο μετοχών και ομολόγων εταιρειών. Ξεκίνησε με πέντε αξιόγραφα, συμπεριλαμβανομένων τριών κρατικών ομολόγων και δύο μετοχών τραπεζών.

Στη σύγχρονη εποχή, το εμπόριο μετοχών έχει εξελιχθεί τεράστια με την πρόοδο της τεχνολογίας. Οι τεχνολογικές εξελίξεις έχουν καταστήσει το εμπόριο πιο προσιτό, γρηγορότερο και οικονομικότερο. Για παράδειγμα, οι ηλεκτρονικές πλατφόρμες εμπορίου έχουν αντικαταστήσει το παραδοσιακό εμπόριο, επιτρέποντας γρήγορη εκτέλεση και ανάλυση συναλλαγών.

Υπάρχουν διάφοροι τύποι στρατηγικών εμπορίου μετοχών που εφαρμόζουν οι επενδυτές, όπως το ημερήσιο εμπόριο, το εμπόριο ταλάντωσης (swing) και η στρατηγική αγοράς και κράτησης. Οι ημερήσιοι εμπόροι στοχεύουν στο κέρδος από τις βραχυπρόθεσμες κινήσεις των τιμών εντός μιας εμπορικής ημέρας. Αντίθετα, οι έμποροι ταλάντωσης προσπαθούν να αποκομίσουν κέρδη κρατώντας μετοχές για λίγες ημέρες ή εβδομάδες. Αντίθετα, οι επενδυτές αγοράς και κράτησης επιδιώκουν το κέρδος από τη μακροπρόθεσμη αύξηση των τιμών και τα μερίσματα.

Το εμπόριο μετοχών είναι ένα σύνθετο πεδίο που απαιτεί βαθιά κατανόηση των χρηματοπιστωτικών αγορών. Η επιτυχημένη εμπορία μετοχών περιλαμβάνει προσεκτική ανάλυση, στρατηγικό σχεδιασμό και διαχείριση κινδύνου.

2.2 Μέθοδοι εμπορίας μετοχών

Οι μέθοδοι εμπορίας μετοχών μπορούν γενικά να κατηγοριοποιηθούν σε τρεις ευρείς τύπους: την θεμελιώδη ανάλυση (fundamental analysis), την ποσοτική ανάλυση (quantitative analysis) και την τεχνική ανάλυση (technical analysis). Κάθε μία από αυτές τις μεθόδους παρέχει μια διαφορετική προσέγγιση για την κατανόηση και την πρόβλεψη της συμπεριφοράς της χρηματιστηριακής αγοράς.

2.2.1 Θεμελιώδης ανάλυση

Η θεμελιώδης ανάλυση περιλαμβάνει την εκτίμηση της εγγενούς αξίας μιας εταιρείας αναλύοντας διάφορους μακροοικονομικούς και εταιρικούς παράγοντες. Αυτή η μέθοδος υποθέτει ότι η χρηματιστηριακή αγορά μπορεί να τιμολογήσει λανθασμένα μια μετοχή βραχυπρόθεσμα, αλλά η σωστή τιμή θα αντικατοπτριστεί τελικά μακροπρόθεσμα.

Οι αναλυτές χρησιμοποιούν διάφορες χρηματοοικονομικές μετρήσεις και αναλογίες για να αξιολογήσουν την οικονομική υγεία μιας εταιρείας, όπως το πολλαπλασιαστικό Τιμής προς Τζίρος (Price/Earnings), η αναλογία Χρέους προς Ίδια Κεφάλαια (Debt/Equity) και η Απόδοση Ιδίων Κεφαλαίων (Return on Equity). Εκτός από την ανάλυση των χρηματοοικονομικών καταστάσεων, οι θεμελιώδεις αναλυτές λαμβάνουν υπόψη ευρύτερους μακροοικονομικούς παράγοντες, όπως οικονομικούς δείκτες, τάσεις της βιομηχανίας και γεωπολιτικά γεγονότα.

Ο στόχος της θεμελιώδους ανάλυσης είναι να εντοπίσει υποτιμημένες ή υπερτιμημένες μετοχές και να λάβει επενδυτικές αποφάσεις με βάση τις «ανεπάρκειες» της αγοράς. Για παράδειγμα, αν η εγγενής αξία μιας εταιρείας εκτιμάται ότι είναι υψηλότερη από την τρέχουσα αγοραία της τιμή, ένας θεμελιώδης αναλυτής θα το δει ως ευκαιρία αγοράς.

2.2.2 Τεχνική ανάλυση

Αντίθετα με τη θεμελιώδη ανάλυση, η τεχνική ανάλυση επικεντρώνεται στις στατιστικές τάσεις βασισμένες στην αγοραία δραστηριότητα, όπως οι κινήσεις των τιμών και ο όγκος. Οι τεχνικοί αναλυτές, πιστεύουν ότι η ιστορική εμπορική δραστηριότητα και οι αλλαγές των τιμών μπορούν να είναι σημαντικοί δείκτες των μελλοντικών τάσεων των τιμών.

Η τεχνική ανάλυση χρησιμοποιεί διάφορα σχήματα διαγραμμάτων, δείκτες και στατιστικά εργαλεία για να προβλέψει τις κινήσεις των τιμών. Κοινές τεχνικές περιλαμβάνουν κινητές μέσες τιμές, γραμμές τάσης, επίπεδα υποστήριξης και αντίστασης και διάφορους δείκτες ορμής. Μερικές από τις πιο συνήθεις τεχνικές είναι οι ακόλουθες:

Κινητός Μέσος Όρος - Moving Average (MA)

Οι Κινητοί Μέσοι Όροι είναι στατιστικοί υπολογισμοί που χρησιμοποιούνται για την ανάλυση δεδομένων σε ένα συγκεκριμένο χρονικό διάστημα. Χρησιμοποιούνται κυρίως για την εντοπισμό των τάσεων, την ανωμαλία των δεδομένων τιμών και την υποστήριξη των επενδυτών στη λήψη αποφάσεων σχετικά με την αγορά και την πώληση περιουσιακών στοιχείων. Υπάρχουν δύο κοινοί τύποι κινητών μέσων:

Απλός Κινητός Μέσος Όρος (ΑΚΜΟ) - Simple Moving Average (SMA)

Ο Απλός Κινητός Μέσος Όρος υπολογίζεται προσθέτοντας ένα σύνολο τιμών κλεισίματος για ένα συγκεκριμένο χρονικό διάστημα και στη συνέχεια διαιρώντας αυτό το άθροισμα με τον αριθμό των χρονικών διαστημάτων:

$$SMA = \frac{x_1+x_2+x_3+\ldots+x_n}{n},$$

όπου x_i η τιμή κλεισίματος τη χρονική στιγμή i και n το πλήθος των παρατηρήσεων.

Για παράδειγμα, ο ημερήσιος ΑΚΜΟ-50 θα προσθέτει τις τιμές κλεισίματος των τελευταίων 50 ημερών και στη συνέχεια θα διαιρεί αυτό το άθροισμα με 50 για να πάρει την τιμή ΑΚΜΟ.

Οι ΑΚΜΟ χρησιμοποιούνται για την εντοπισμό της γενικής κατεύθυνσης μιας τάσης και την ανωμαλία των τιμών σε μικρή χρονική κλίμακα. Οι επενδυτές συχνά αναζητούν τη διασταύρωση μεταξύ των κινητών μέσων σε διαφορετικά χρονικά διαστήματα ως σημάδια για αλλαγές στην τάση.

Εκθετικός Κινητός Μέσος Όρος (ΕΚΜΟ) - Exponential Moving Average (EMA)

Ο ΕΚΜΟ είναι ένας δημοφιλής τεχνικός δείκτης που χρησιμοποιείται στην τεχνική ανάλυση για να αναλύσει και να προβλέψει την κατεύθυνση και τη δυναμική μιας μετοχής. Είναι ένας τύπος κινούμενου μέσου όρου που δίνει μεγαλύτερο βάρος στα πρόσφατα δεδομένα τιμών, καθιστώντας πολύ ανταποκριτικό στις πρόσφατες κινήσεις τιμών σε σύγκριση με τον Απλό Κινούμενο Μέσο (SMA). Ο τύπος υπολογισμού είναι ο εξής:

$$EKMO_t = \frac{P_t + \sum_{i=1}^{n} (1-a)^i \cdot x_{t-i}}{1 + \sum_{i=1}^{n} (1-a)^i},$$

όπου n το μέγεθος του παραθύρου, α ο παράγοντας εξομάλυνσης και
x_i η τιμή της i παρατήρησης

Ο ΕΚΜΟ χρησιμοποιείται κυρίως για να αναγνωρίσει και να επιβεβαιώσει τις τάσεις στις κινήσεις των τιμών. Όταν ο ΕΚΜΟ ανεβαίνει, υποδηλώνει ανοδική τάση, και όταν πέφτει, υποδηλώνει καθοδική τάση. Οι έμποροι συνήθως χρησιμοποιούν μικρότερες περιόδους ΕΚΜΟ (π.χ. 10, 20 ημερών) για σύντομες τάσεις και μεγαλύτερες περιόδους ΕΚΜΟ (π.χ. 50, 200 ημερών) για μακροπρόθεσμες τάσεις. Οι διασταυρώσεις των παραπάνω ΕΚΜΟ είναι από τα πιο δημοφιλή σήματα εμπορίου. Όταν ένας ΕΚΜΟ με μικρότερη περίοδο (π.χ. 10 περίοδοι) διασταυρώνεται πάνω από ένας ΕΚΜΟ με μεγαλύτερη περίοδο (π.χ. 50 περίοδοι), σηματοδοτεί μια πιθανή ανοδική τάση. Αντίστοιχα, όταν το μικρότερο ΕΚΜΟ διασταυρώνεται κάτω από το μεγαλύτερο ΕΚΜΟ, σηματοδοτεί μια πιθανή καθοδική τάση. Οι έμποροι συνήθως χρησιμοποιούν αυτές τις διασταυρώσεις για να εισέλθουν ή να βγουν από θέσεις.

Δείκτης Σχετικής Δύναμης (ΔΣΔ) - Relative Strength Index (RSI)

Το RSI αντιπροσωπεύει τον Δείκτη Σχετικής Δύναμης και είναι ένα δημοφιλές τεχνικό εργαλείο που χρησιμοποιείται στις χρηματοοικονομικές αγορές, ιδιαίτερα στον τομέα της τεχνικής ανάλυσης. Ο ΔΣΔ χρησιμοποιείται για την αξιολόγηση της ισχύος και της δυναμικής των κινήσεων των τιμών μιας μετοχής. Δημιουργήθηκε από τον J. Welles Wilder το 1978 και έχει από τότε γίνει ένα διαδεδομένο εργαλείο μεταξύ των εμπόρων και των αναλυτών. Ο ΔΣΔ υπολογίζεται ως εξής:

$$\text{ΔΣΔ} = 100 - \frac{100}{1 + RS}$$

$$\text{όπου } RS = \frac{\text{Μέσος Όρος Θετικών Κλεισιμάτων}}{\text{Μέσος Όρος Αρνητικών Κλεισιμάτων}}$$

Η πιο συνηθισμένη περίοδος που χρησιμοποιείται για τον υπολογισμό του ΔΣΔ είναι 14 ημέρες. Ο ΔΣΔ χρησιμοποιείται συνήθως για την ανίχνευση συνθηκών υπεραγοράς και υπερπώλησης στην τιμή μιας μετοχής. Όταν ο ΔΣΔ ανεβαίνει πάνω από το όριο υπεραγοράς (π.χ. 70), υποδηλώνει ότι η μετοχή μπορεί να είναι υπερτιμημένη και ενδέχεται να αναμένεται διόρθωση τιμής. Αντίστροφα, όταν ο ΔΣΔ πέφτει κάτω από το όριο υπερπώλησης (π.χ. 30), υποδηλώνει ότι η μετοχή μπορεί να είναι υποτιμημένη και ενδέχεται να αναμένεται αντίστροφη κίνηση.

Κινητός Μέσος Όρος Σύγκλισης/Απόκλισης (ΚΜΟΣΑ) - Moving Average Convergence/Divergence (MACD)

Ο Κινητός Μέσος Όρος Σύγκλισης - Απόκλισης, είναι ένας άλλος δημοφιλής τεχνικός δείκτης που χρησιμοποιείται στις χρηματοοικονομικές αγορές, ιδιαίτερα στην τεχνική ανάλυση. Δημιουργήθηκε από τον Gerald Appel στα τέλη της δεκαετίας του 1970 και χρησιμοποιείται για την ανάλυση των τάσεων, τον εντοπισμό δυνητικών σημάτων αγοράς και πώλησης, καθώς και

την εκτίμηση της ισχύος και της δυναμικής των κινήσεων των τιμών σε ένα χρηματοοικονομικό ενεργητικό.

Ο ΚΜΟΣΑ υπολογίζεται χρησιμοποιώντας τα ακόλουθα στοιχεία:

1. Γραμμή ΚΜΟΣΑ (Γρήγορη Γραμμή): Η γραμμή ΚΜΟΣΑ είναι η διαφορά μεταξύ της Εκθετικής Κινητής Μέσης Τιμής 12 περιόδων (ΕΜΑ) και της ΕΜΑ 26 περιόδων. Η γραμμή ΚΜΟΣΑ θεωρείται η "γρήγορη γραμμή" επειδή αντιδρά πιο γρήγορα στις αλλαγές των τιμών.
2. Γραμμή Σήματος (Αργή Γραμμή): Η γραμμή σήματος, γνωστή και ως "αργή γραμμή," είναι η ΕΜΑ 9 περιόδων της γραμμής ΚΜΟΣΑ. Καθιστά τη γραμμή ΚΜΟΣΑ πιο ομαλή και χρησιμοποιείται για τη δημιουργία σημάτων αγοράς ή πώλησης.

Η πιο συνηθισμένη χρήση του ΚΜΟΣΑ είναι η ανίχνευση των διασταυρώσεων μεταξύ της γραμμής ΚΜΟΣΑ και της γραμμής σήματος. Όταν η γραμμή ΚΜΟΣΑ «περνάει» πάνω από τη γραμμή σήματος, δημιουργεί ένα ανοδικό σήμα, υποδεικνύοντας ότι ενδέχεται να είναι καλή στιγμή για αγορά. Αντίστροφα, όταν η γραμμή ΚΜΟΣΑ «περνάει» κάτω από τη γραμμή σήματος, δημιουργεί έναν καθοδικό σήμα, υποδηλώνοντας ότι ενδέχεται να είναι καλή στιγμή για πώληση.

Όγκος Ισορροπίας (OI) - On-Balance Volume (OBV)

Ο Όγκος Ισορροπίας (OBV) είναι ένας δείκτης που χρησιμοποιείται στις χρηματοοικονομικές αγορές, ιδιαίτερα στην τεχνική ανάλυση, για να μετρήσει τη θετική και αρνητική τάση μια μετοχής. Δημιουργήθηκε από τον Joseph Granville στη δεκαετία του 1960 και βασίζεται στην αρχή ότι ο όγκος μπορεί να παρέχει σημαντικές ενδείξεις για τις κινήσεις των τιμών. Ο OI προσπαθεί να αποτυπώσει τη σχέση μεταξύ του όγκου και των αλλαγών των τιμών.

Ο υπολογισμός του OI είναι σχετικά απλός. Αναθέτει μια θετική ή αρνητική τιμή σε κάθε ημέρα συναλλαγών, ανάλογα με το αν η κλείσιμο της τιμής είναι υψηλότερο ή χαμηλότερο από το κλείσιμο της προηγούμενης ημέρας. Εάν το κλείσιμο της τιμής είναι υψηλότερο από το κλείσιμο της προηγούμενης ημέρας, ο ημερήσιος όγκος θεωρείται θετικός και προστίθεται στον OI. Εάν το κλείσιμο της τιμής είναι χαμηλότερο από το κλείσιμο της προηγούμενης ημέρας, ο ημερήσιος όγκος θεωρείται αρνητικός και αφαιρείται από τον OI. Εάν το κλείσιμο της τιμής είναι αναλλοίωτο από την προηγούμενη ημέρα, τότε συνήθως αγνοείται ο όγκος.

Ο OI ξεκινά με μια αρχική τιμή μηδέν και συγκεντρώνει αυτές τις θετικές ή αρνητικές τιμές του όγκου με την πάροδο του χρόνου για να δημιουργήσει μια συσσωρευτική γραμμή OI.

Όταν η γραμμή OI αυξάνεται, υποδεικνύει ότι ο όγκος αγοράς αυξάνεται, πράγμα που μπορεί να

υποδηλώνει ανοδική δυναμική της τιμής. Οι εμπόροι συνήθως ερμηνεύουν την αύξηση του OI ως ανοδικό σήμα, υποδεικνύοντας μια πιθανή ανοδική τάση της τιμής. Αντιστρόφως, όταν η γραμμή OI φθίνει, υποδεικνύει ότι ο όγκος πώλησης αυξάνεται, πράγμα που μπορεί να υποδηλώνει καθοδική τάση της τιμής.

Ζώνες Bollinger - Bollinger Bands

Οι Ζώνες Bollinger είναι ένα δημοφιλές τεχνικό εργαλείο ανάλυσης που βοηθά τους επενδυτές να αξιολογήσουν την μεταβλητότητα (volatility), να εντοπίσουν πιθανά σημεία αναστροφής και να κατανοήσουν τις κινήσεις των τιμών στις χρηματοοικονομικές αγορές. Δημιουργήθηκαν από τον John Bollinger στις αρχές της δεκαετίας του 1980 και αποτελούνται από τρεις γραμμές:

1. **Κεντρική Ζώνη (SMA)**: Η κεντρική ζώνη είναι συνήθως ένας Απλός Κινητός Μέσος Όρος (SMA) της τιμής του περιουσιακού στοιχείου για ένα συγκεκριμένο χρονικό διάστημα. Η πιο συνηθισμένη περίοδος που χρησιμοποιείται είναι 20 ημέρες, αλλά οι επενδυτές μπορούν να την προσαρμόσουν ανάλογα με τις προτιμήσεις και την ανάλυσή τους.
2. **Άνω Ζώνη**: Η άνω ζώνη υπολογίζεται προσθέτοντας ένα συγκεκριμένο αριθμό τυπικών αποκλίσεων στην κεντρική ζώνη. Οι τυπικές αποκλίσεις μετρούν την μεταβλητότητα της τιμής του περιουσιακού στοιχείου. Η συνήθης πρακτική είναι να χρησιμοποιούνται δύο τυπικές αποκλίσεις, οι οποίες περιλαμβάνουν περίπου το 95% των δεδομένων τιμών μέσα στις λωρίδες.
3. **Κάτω Λωρίδα**: Η κάτω λωρίδα υπολογίζεται αφαιρώντας τον ίδιο αριθμό τυπικών αποκλίσεων από την κεντρική λωρίδα.

Εδώ είναι τα βασικά χαρακτηριστικά και οι χρήσεις των Ζωνών Bollinger:

1. Αξιολόγηση της Μεταβλητότητας:

- Όταν οι λωρίδες συστέλλονται, υποδεικνύει χαμηλή μεταβλητότητα.
- Όταν οι λωρίδες διαστέλλονται, σημαίνει υψηλή μεταβλητότητα.

2. Αναγνώριση Τάσης:

- Όταν η τιμή ενός περιουσιακού στοιχείου κινείται πάνω από την άνω λωρίδα, μπορεί να υποδεικνύει ότι το περιουσιακό στοιχείο είναι υπεραγορασμένο και μια πτώση της τιμής είναι πιθανή
- Όταν η τιμή κινείται κάτω από την κάτω λωρίδα, μπορεί να υποδεικνύει ότι το περιουσιακό στοιχείο είναι υπερπουλημένο και μια άνοδος της τιμής είναι πιθανή.

Όλες οι παραπάνω μέθοδοι βασίζονται στην θεμελιώδη παραδοχή της τεχνικής ανάλυσης, ότι όλες οι πληροφορίες της αγοράς αντικατοπτρίζονται ήδη στην τιμή, και έτσι, η μελέτη των μοτίβων και των τάσεων των τιμών είναι ο πιο αποτελεσματικός τρόπος για την πρόβλεψη των μελλοντικών κινήσεων των τιμών.

2.2.3 Ποσοτική Ανάλυση

Η Ποσοτική Ανάλυση (Quantitative Analysis) είναι μια πιο σύγχρονη μέθοδος που περιλαμβάνει τη χρήση μαθηματικού και στατιστικού μοντελοποίησης για την κατανόηση και την πρόβλεψη της συμπεριφοράς στις χρηματοοικονομικές αγορές. Αυτή η μέθοδος περιλαμβάνει πολύπλοκους αλγόριθμους και υπολογιστές υψηλής ταχύτητας για τη δημιουργία και την εκτέλεση αυτοματοποιημένων στρατηγικών εμπορίας (algorithmic trading).

Οι ποσοτικοί αναλυτές, χρησιμοποιούν αυτή τη μέθοδο για την εντοπίσει ευκαιρίες εμπορίας μέσω της ανάλυσης μεγάλων ποσοτήτων δεδομένων. Αυτοί οι αλγόριθμοι συνήθως βασίζονται σε ιστορικά δεδομένα και προσπαθούν να προβλέψουν τις μελλοντικές τάσεις των τιμών με βάση τα στατιστικά μοτίβα του παρελθόντος.

2.3 Μέθοδοι Αξιολόγησης Συναλλαγών Μετοχών

Η αξιολόγηση της απόδοσης των στρατηγικών συναλλαγών μετοχών είναι ένα κρίσιμο βήμα για τη βελτίωση και την τελειοποίηση των αποτελεσμάτων των συναλλαγών. Αυτή η διαδικασία συνήθως βασίζεται σε έναν συνδυασμό διαφορετικών μετρήσεων και στατιστικών δεικτών, οι οποίοι εξυπηρετούν διάφορους σκοπούς και φωτίζουν διάφορες πτυχές της απόδοσης των συναλλαγών. Παρακάτω παρουσιάζονται κάποιες κοινά χρησιμοποιούμενες μέθοδοι αξιολόγησης:

2.3.1 Απόδοση Επένδυσης (ΑΕ) - Return On Investment (ROI)

Η ΑΕ είναι μια μέτρηση της κέρδους ή απώλειας που έγινε σε μια επένδυση σε σχέση με το ποσό των χρημάτων που επενδύθηκαν. Είναι συνήθως εκφρασμένο ως ποσοστό και υπολογίζεται ως

$$\text{ΑΕ} = \frac{\text{Καθαρό Κέρδος}}{\text{Κόστος Επένδυσης}} \cdot 100\%$$

Η ΑΕ είναι ένας ευρέως χρησιμοποιούμενος δείκτης κερδοφορίας, επειδή είναι απλός και ευέλικτος. Ωστόσο, δεν λαμβάνει υπόψη τον κίνδυνο που σχετίζεται με μια επένδυση, ούτε λαμβάνει υπόψη τη διάρκεια της επένδυσης.

2.3.2 Ετήσιες Αποδόσεις

Οι ετήσιες αποδόσεις αντιπροσωπεύουν τον γεωμετρικό μέσο όρο των χρημάτων που κερδίζεται από μια επένδυση κάθε χρόνο κατά τη διάρκεια ενός δεδομένου χρονικού διαστήματος. Υπολογίζεται ως η γεωμετρική μέση των αποδόσεων, κλιμακούμενη σε ετήσια βάση, που επιτρέπει τη σύγκριση των αποδόσεων από διάφορες επενδύσεις σε διάφορα χρονικά διαστήματα:

$$\text{Ετήσια Απόδοση} = \left(\frac{\text{Τελική Αξία}}{\text{Αρχική Αξία}}\right)^{1/\text{Αριθμός Χρόνων}} - 1$$

2.3.3 Δείκτης Απώλειας - Drawdown

Ο Δείκτης Απώλειας μετρά την πτώση από ένα ιστορικό υψηλό σε κάποια μεταβλητή (συνήθως το σωρευτικό κέρδος ή η συνολική ανοιχτή κεφαλαιακή επένδυση μιας χρηματοοικονομικής στρατηγικής). Είναι μια σημαντική μέτρηση κινδύνου, ιδιαίτερα στο εμπόριο παραγώγων και για τις στρατηγικές που στοχεύουν στην ελαχιστοποίηση της πιθανότητας και της μείωσης των σοβαρών απωλειών.

2.3.4 Μέγιστη Απώλεια (MDD)

Η Μέγιστη Απώλεια (MDD) είναι η μέγιστη ζημία από το υψηλότερο σημείο στο χαμηλότερο σημείο πριν νέα ρεκόρ υψηλά. Αυτή η μετρική βοηθά στην αναγνώριση του ρίσκου και της απόδοσης της παρούσας επένδυσης.

$$\Delta\text{Α} = \frac{\text{Ελάχιστο} - \text{Μέγιστο}}{\text{Μέγιστο}}$$

2.3.5 Συντελεστής Sharpe - Sharpe Ratio

Ο Συντελεστής Sharpe, που αναπτύχθηκε από τον κάτοχο βραβείου Νόμπελ William F. Sharpe, μετρά την απόδοση που προσαρμόζεται στον κίνδυνο. Υπολογίζεται αφαιρώντας το άνευ κινδύνου επιτόκιο (risk free rate) - συνήθως εκείνο ενός ομολόγου των ΗΠΑ 10 ετών - από την ποσοστιαία επιστροφή για ένα χαρτοφυλάκιο και διαιρώντας το αποτέλεσμα με την τυπική απόκλιση των επιστροφών (returns) του χαρτοφυλακίου.

$$\text{Συντελεστής } Sharpe = \frac{R_\chi - R_f}{\sigma_\chi}$$

όπου R_χ η ποσοστιαία επιστροφή χαρτοφυλακίου , R_f η άνευ κινδύνου επιτόκιο και

σ_{χ} η τυπική απόκλιση των επιστροφών τον χαρτοφυλακίου

Ο συντελεστής Sharpe παρέχει μια εικόνα της απόδοσης μιας επένδυσης σε σχέση με τον κίνδυνό της.

2.3.6 Συντελεστής Sortino - Sortino Ratio

Ο συντελεστής Sortino, μια παραλλαγή του συντελεστή Sharpe, διακρίνει την επιβλαβή ταλάντωση από την συνολική συνολική ταλάντωση, χρησιμοποιώντας την τυπική απόκλιση των αρνητικών επιστροφών του χαρτοφυλακίου - την αρνητική απόκλιση - αντί για τη συνολική τυπική απόκλιση των επιστροφών του χαρτοφυλακίου.

$$Sortino\ Ratio\ = \frac{R_{\chi} - R_{f}}{\sigma_{d\chi}},$$

όπου R_{χ} η ποσοστιαία επιστροφή χαρτοφυλακίου , R_{f} η άνευ κινδύνου επιτόκιο και

σ_{χ} η τυπική απόκλιση των αρνητικών επιστροφών τον χαρτοφυλακίου

Αυτός ο λόγος είναι μια καλύτερη μέτρηση όταν αναλύουμε χαρτοφυλάκια που έχουν μη κανονικές κατανομές επιστροφών ή ασύμμετρα προφίλ κινδύνου.

2.4 Διαχείριση χαρτοφυλακίου

Η διαχείριση χαρτοφυλακίου, στο πλαίσιο των χρηματοοικονομικών επενδύσεων, αναφέρεται στη διαδικασία διαχείρισης ενός χαρτοφυλακίου επενδύσεων με τρόπο που αποσκοπεί στη μεγιστοποίηση των αποδόσεων και την ελαχιστοποίηση του κινδύνου, δεδομένων των συγκεκριμένων στόχων και περιορισμών του επενδυτή. Η διαχείριση χαρτοφυλακίου περιλαμβάνει τη λήψη αποφάσεων σχετικά με τον συνδυασμό και την πολιτική επενδύσεων, την κατανομή του διαθέσιμου κεφαλαίου για άτομα και θεσμικούς επενδυτές, την ταύτιση των επενδύσεων με τους στόχους και την ισορροπία του κινδύνου έναντι της απόδοσης (Markowitz, 1952).

Το θεωρητικό πλαίσιο της διαχείρισης χαρτοφυλακίου επαναπροσδιορίστηκε από τη Σύγχρονη Θεωρία Χαρτοφυλακίου (MPT), που παρουσίασε ο Harry Markowitz στο άρθρο του το 1952, "Portfolio Selection". Η θεωρία αυτή υποστήριζε τα οφέλη της διαφοροποίησης (diversification) των επενδύσεων σε διάφορα αγαθά για τη μείωση του κινδύνου. Η θεωρία προτείνει ότι τα χαρακτηριστικά κινδύνου και απόδοσης μιας επένδυσης δεν πρέπει να εξετάζονται μόνα τους,

αλλά πρέπει να αξιολογούνται ανάλογα με το πώς η επένδυση επηρεάζει τον κίνδυνο και την απόδοση του συνολικού χαρτοφυλακίου (Markowitz, 1952).

Γενικά υπάρχουν δύο τύποι στρατηγικών διαχείρισης χαρτοφυλακίου: ενεργητικές και παθητικές. Η ενεργητική διαχείριση χαρτοφυλακίου περιλαμβάνει την αγοραπωλησία τίτλων με σκοπό την καλύτερη απόδοση από έναν επενδυτικό δείκτη αναφοράς. Οι ενεργητικοί διαχειριστές βασίζονται σε αναλυτική έρευνα, προβλέψεις, και τη δική τους κρίση και εμπειρία στη λήψη αποφάσεων επένδυσης για το ποιοι τίτλοι θα αγοραστούν, θα κρατηθούν και θα πωληθούν.

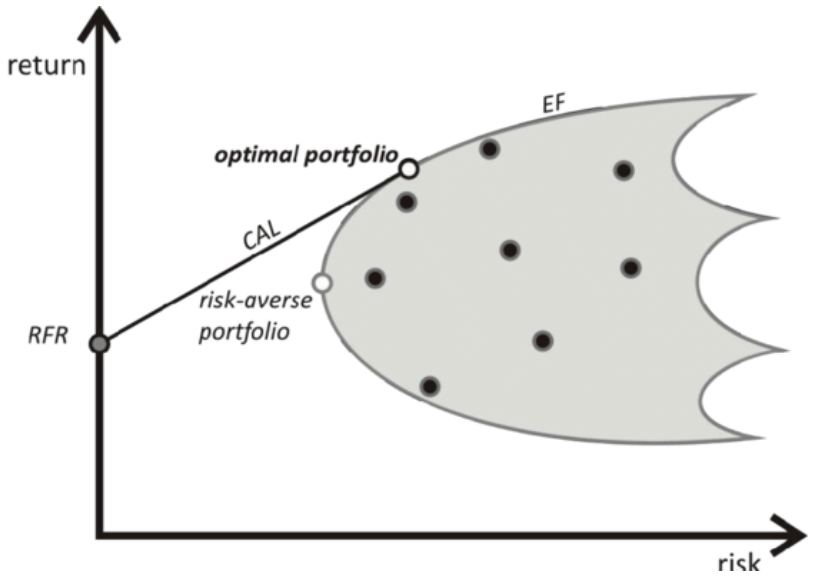
Από την άλλη πλευρά, η παθητική διαχείριση χαρτοφυλακίου στοχεύει στην ταύτιση των αποδόσεων της αγοράς ή ενός τομέα αυτής, συνήθως ακολουθόντας έναν δείκτη αναφοράς. Οι παθητικοί διαχειριστές δεν επιδιώκουν να υπερβούν την αγορά, αντ' αυτού, επιδιώκουν να αναπαράγουν την απόδοση ενός συγκεκριμένου δείκτη ή μέτρου όσο το δυνατόν πιο κοντά, κρατώντας όλα (ή ένα αντιπροσωπευτικό δείγμα) των τίτλων που συνθέτουν τον δείκτη. Η πιο κοινή μέθοδος παθητικής διαχείρισης είναι μέσω δεικτοποιημένων κεφαλαίων ή exchange-traded funds (ETFs).

Στα τελευταία χρόνια, οι τεχνολογικές εξελίξεις έχουν οδηγήσει στην εμφάνιση των αυτοματοποιημένων «συμβούλων επενδύσεων», οι οποίοι χρησιμοποιούν αλγόριθμους για τη διαχείριση των χαρτοφυλακίων επενδύσεων. Αυτές οι πλατφόρμες συχνά χρησιμοποιούν παθητικές επενδυτικές στρατηγικές, προσφέροντας στους επενδυτές μια λύση χαμηλού κόστους για τη διαχείριση χαρτοφυλακίου.

2.5 Βελτιστοποίηση Χαρτοφυλακίου

Οι πρώτες μέθοδοι βελτιστοποίησης του χαρτοφυλακίου βρίσκουν τις ρίζες τους στη Σύγχρονη Θεωρία Χαρτοφυλακίου (MPT) που πρότεινε ο Harry Markowitz το 1952, η οποία παρέχει μια ποσοτική διαδικασία για την σύνθεση ενός χαρτοφυλακίου, με στόχο είτε τη μεγιστοποίηση της επιστροφής για ένα δεδομένο επίπεδο κινδύνου ή εναλλακτικά την ελαχιστοποίηση του κινδύνου για ένα δεδομένο επίπεδο επιστροφής (Markowitz, 1952).

Η προσέγγιση του Markowitz, γνωστή ως βελτιστοποίηση μέσου-διασποράς, διατυπώνει το πρόβλημα βελτιστοποίησης του χαρτοφυλακίου ως πρόβλημα τετραγωνικής βελτιστοποίησης, όπου επιδιώκεται να μεγιστοποιηθεί η αναμενόμενη απόδοση του χαρτοφυλακίου για ένα δεδομένο επίπεδο διασποράς, ή αντιστρόφως, η διασπορά (που αντιπροσωπεύει τον κίνδυνο) ελαχιστοποιείται για ένα δεδομένο επίπεδο αναμενόμενης απόδοσης. Αυτή η μέθοδος χρησιμοποιεί τον πίνακα συνδιασποράς των αποδόσεων των περιουσιακών στοιχείων, ο οποίος κωδικοποιεί τόσο τις ατομικές διασπορές των αποδόσεων των περιουσιακών στοιχείων όσο και τις ζευγαρικές συσχετίσεις τους.



Σχήμα 2.1 Βέλτιστο χαρτοφυλάκιο - Όριο αποδοτικότητας

Μια άλλη παραδοσιακή προσέγγιση στη βελτιστοποίηση του χαρτοφυλακίου είναι το Μοντέλο Αποτίμησης Περιουσιακών Στοιχείων (CAPM), που αναπτύχθηκε από τους William Sharpe (1964) και John Lintner (1965). Το ΜΑΠΣ επεκτείνει το μοντέλο του Markowitz εξετάζοντας τη σχέση μεταξύ του κινδύνου και της απόδοσης ενός περιουσιακού στοιχείου και εκείνης ενός χαρτοφυλακίου αγοράς. Σύμφωνα με το ΜΑΠΣ, η αναμενόμενη απόδοση ενός περιουσιακού στοιχείου είναι ανάλογη του βήτα, μιας μέτρησης του συστηματικού κινδύνου του, ο οποίος είναι ο κίνδυνος που δεν μπορεί να μειωθεί με μεθόδους διαφοροποίησης.

Μια ακόμη προσέγγιση είναι το μοντέλο Black-Litterman (Black και Litterman, 1992), το οποίο υπερβαίνει κάποιους περιορισμούς του μοντέλου μέσου-διασποράς ενσωματώνοντας τις απόψεις των επενδυτών σχετικά με τις μελλοντικές αποδόσεις των περιουσιακών στοιχείων και την εμπιστοσύνη σε αυτές τις απόψεις. Αυτό το μοντέλο παράγει μια πιο διαισθητική, ανθεκτική κατανομή χαρτοφυλακίου σε σύγκριση με τη βελτιστοποίηση της μέσης διασποράς.

Η στρατηγική "Αγορά και Κράτηση" (Buy and Hold) αποτελεί την πιο διαδεδομένη τεχνική διαχείρισης χαρτοφυλακίου. Αντιπροσωπεύει μια διαχειριστική προσέγγιση στον χρηματοοικονομικό τομέα, όπου ο επενδυτής αγοράζει χρηματοοικονομικά εργαλεία, όπως μετοχές ή ομόλογα, και τα κρατά για μακριά χρονικά διαστήματα, ανεξάρτητα από τις καθημερινές αγοραπωλησίες ή τις συναλλαγές. Η βασική ιδέα είναι να επωφεληθεί ο επενδυτής από τη μακροπρόθεσμη αύξηση της αξίας των επενδύσεών του. Αυτή η στρατηγική είναι γνωστή για την απλότητά της, καθώς δεν απαιτεί συχνό και ενεργό εμπλοκή του επενδυτή στις αγορές. Ο κεντρικός στόχος είναι η αποφυγή υπερβολικών συναλλαγών και η αξιοποίηση της μακροπρόθεσμης απόδοσης των επενδύσεων. Έτσι, ο επενδυτής διατηρεί την ιδιοκτησία των χρηματοοικονομικών του εργαλείων, ακόμα και σε περιόδους αγοραίας αναστάτωσης, με στόχο τη μακροπρόθεσμη απόδοση και τη μείωση του χρονικού ορίζοντα απόφασης. Αυτή η

αποτελεσματική προσέγγιση μειώνει την επίδραση της αγοραπωλησίας και των συναλλαγών, μειώνοντας έτσι το κόστος συναλλαγών. Η ανοχή σε μακροπρόθεσμες ανοδικές τάσεις συμβάλλει στη μείωση του κινδύνου της αγοράς και τη διατήρηση της αξίας του χαρτοφυλακίου σε διάφορες καταστάσεις αγοράς.

Η μεθοδολογία του Ακολούθησε τον Ηγέτη (Follow the Leader - FtL) στη βελτιστοποίηση χαρτοφυλακίου αντιπροσωπεύει μια πρακτική προσέγγιση, κατά την οποία οι επενδυτές μιμούνται τις επενδυτικές αποφάσεις ενός καθορισμένου ηγέτη. Συγκεκριμένα, μια απλοποιημένη προσαρμογή αυτής της μεθόδου περιλαμβάνει έναν επενδυτή που επιλέγει συνεχώς την καλύτερη μετοχή μέσα στο χαρτοφυλάκιο. Σε αυτό το πλαίσιο, οι ακόλουθοι αντιγράφουν την απόφαση του ηγέτη να επενδύσει μόνο στη μετοχή που εκδηλώνει τις υψηλότερες μετρήσεις απόδοσης, επιδιώκοντας να επωφεληθούν από την ιστορική επιτυχία του ηγέτη. Αυτή η προσέγγιση βασίζεται στην υπόθεση ότι ο ηγέτης διαθέτει γνώση της αγοράς το οποίο επιτρέπει μια καλή απόδοση (Devenow & Welch, 1996).

Η στρατηγική Σταθερής Επαναζύγισης Χαρτοφυλακίου (Constant Rebalance Portfolio - CRP) αντιπροσωπεύει μια προσέγγιση στη διαχείριση χαρτοφυλακίου, όπου ο επενδυτής τακτικά προβαίνει σε επανισορροπήσεις των κεφαλαίων προκειμένου να διατηρήσει σταθερή την αρχική κατανομή των επενδύσεων. Κατά τη διάρκεια αυτών των επανισορροπήσεων, πωλούνται ή αγοράζονται χρηματοοικονομικά εργαλεία ώστε να επιστρατευτεί η αρχική αναλογία μεταξύ διαφορετικών κατηγοριών επενδύσεων. Η τεχνική της Σταθερής Επαναζύγισης Χαρτοφυλακίου (Constant Rebalance Portfolio - CRP) παρουσιάζει πλεονεκτήματα στον τομέα της διαχείρισης χαρτοφυλακίου. Ένα από τα κύρια οφέλη είναι η διαχείριση του κινδύνου, καθώς η στρατηγική διατηρεί μια διαφοροποιημένη σύνθεση περιουσιακών στοιχείων. Αυτή η διαφοροποίηση συνεισφέρει στη μείωση των επιπτώσεων των χαμηλών αποδόσεων κάποιων περιουσιακών στοιχείων στο συνολικό χαρτοφυλάκιο. Επιπλέον, η συνεχής επανισορρόπηση του χαρτοφυλακίου διατηρεί σταθερό το προφίλ κινδύνου, αποτρέποντας το χαρτοφυλάκιο από το να εκτίθεται υπερβολικά σε συγκεκριμένες κατηγορίες περιουσιακών στοιχείων. Η στρατηγική επιτυγχάνει τη διαχείριση του ρίσκου με μια πειθαρχημένη προσέγγιση, ενισχύοντας την ικανότητα του επενδυτή να διατηρεί τους μακροπρόθεσμους του στόχους επενδυτικής απόδοσης.

Παρά το γεγονός ότι αυτές οι παραδοσιακές μέθοδοι θεωρούνται αξιόπιστες και έχουν υπηρετήσει καλά τους επενδυτές στην πράξη, έχουν ορισμένους περιορισμούς. Βασίζονται σε μεγάλο βαθμό στις ακριβείς εκτιμήσεις των μέσων, των διασπορών και των συσχετίσεων των αποδόσεων των περιουσιακών στοιχείων, τα οποία είναι δύσκολο να εκτιμηθούν με ακρίβεια στην πράξη. Επιπλέον, αυτές οι μέθοδοι υποθέτουν μια κανονική κατανομή των επιστροφών και αγνοούν τις ακραίες τιμές της σκευότητας και της κυρτότητας, οι οποίες μπορούν να είναι κρίσιμες στις αποδόσεις των περιουσιακών στοιχείων που εκδηλώνουν ασυμμετρία και βαριές ουρές. Αυτό οδήγησε στην ανάπτυξη εναλλακτικών μεθόδων βελτιστοποίησης του χαρτοφυλακίου, οι οποίες θα συζητηθούν στη συνέχεια.

Κεφάλαιο 3. Μηχανική Μάθηση

3.1 Εισαγωγή

Η μηχανική μάθηση αποτελεί έναν σημαντικό κλάδο της τεχνητής νοημοσύνης, όπου ερευνώνται και αναπτύσσονται συστήματα και μέθοδοι που μπορούν να "μάθουν". Αυτό σημαίνει ότι τα συστήματα μηχανικής μάθησης χρησιμοποιούν νέα δεδομένα για να βελτιώσουν την απόδοσή τους σε συγκεκριμένες εργασίες.

Η αρχή της μηχανικής μάθησης μπορεί να εντοπιστεί στις πρωτοπόρες ιδέες του Alan Turing, ο οποίος με το καθοριστικό άρθρο του το 1950, "Υπολογιστικά Μηχανήματα και Νοημοσύνη", έθεσε τις θεμελιώδεις βάσεις για την τεχνητή νοημοσύνη και την προοπτική των μηχανών που μπορούν να μαθαίνουν και να προσομοιάζουν την ανθρώπινη σκέψη. Πάνω σε αυτές τις ιδέες, το άρθρο του Arthur Samuel το 1959, "Μερικές Μελέτες στην Μηχανική Μάθηση χρησιμοποιώντας το Παιχνίδι των Ντάμας", παρουσίασε στον κόσμο τον όρο της αυτο-βελτίωσης στους υπολογιστές μέσω της επαναληπτικής μάθησης, αφήνοντας ένα σημαντικό σημείο αναφοράς που προέτρεψε την ανάπτυξη σύγχρονων αλγορίθμων μηχανικής μάθησης.

Οι αλγόριθμοι που χρησιμοποιούνται σε αυτήν την περιοχή δημιουργούν μοντέλα που εκπαιδεύονται με δεδομένα από παρατηρήσεις και καταγραφές. Τα μοντέλα αυτά χρησιμοποιούνται για να παράγουν προβλέψεις και αποφάσεις, χωρίς να χρειάζεται να προγραμματιστούν εξαρχής για κάθε εργασία. Αυτή η δυνατότητα επίλυσης προβλημάτων με δεδομένα είναι ιδιαίτερα χρήσιμη σε πολύπλοκες εργασίες όπου ο συμβατικός προγραμματισμός είναι δύσκολος ή αδύνατος.

Ένα ειδικό πεδίο εντός της μηχανικής μάθησης αναπτύσσεται γύρω από τα νευρωνικά δίκτυα, τα οποία είναι μοντέλα που προσομοιώνουν τη λειτουργία του ανθρώπινου εγκεφάλου. Τα πρώτα βήματα στην ανάπτυξη των τεχνητών νευρωνικών δικτύων ξεκίνησαν στη δεκαετία του 1940 από τους Warren S. McCulloch και Walter Pitts, οι οποίοι δημιούργησαν ένα υπολογιστικό μοντέλο για νευρωνικά δίκτυα. Από τότε, η έρευνα σε αυτό τον τομέα έχει σημειώσει σημαντική πρόοδο, με την εφαρμογή νευρωνικών δικτύων σε πολλούς τομείς.

Σήμερα, η μηχανική μάθηση και τα νευρωνικά δίκτυα βρίσκονται σε πρωτοφανή ανάπτυξη χρησιμοποιούνται ευρέως σε διάφορες εφαρμογές. Αυτές οι εφαρμογές περιλαμβάνουν αναγνώριση προτύπων, υπολογισμό συναρτήσεων, βελτιστοποίηση, πρόβλεψη, αυτόματο έλεγχο, και πολλές άλλες. Οι εφαρμογές αυτές εκτείνονται σε πολλούς τομείς της καθημερινής ζωής, όπως αυτόνομα οχήματα και αναγνώριση φωνής, και αναμένεται ότι θα επεκταθούν ακόμα περισσότερο στο μέλλον.

3.2 Επιβλεπόμενη - Μη Επιβλεπόμενη Μηχανική Μάθηση

Η επιβλεπόμενη (supervised) και η μη επιβλεπόμενη (unsupervised) μηχανική μάθηση αποτελούν δύο θεμελιώδεις πυλώνες εντός του ευρύτερου πεδίου της τεχνητής νοημοσύνης και της επιστήμης των δεδομένων, καθένα από τα οποία διαδραματίζει έναν κεντρικό ρόλο στην επίλυση ποικίλων και πολύπλοκων πραγματικών προβλημάτων.

Η επιβλεπόμενη μηχανική μάθηση περιλαμβάνει τη χρήση ετικεταρισμένων δεδομένων (labeled data), όπου ο αλγόριθμος εκπαιδεύεται σε ένα σύνολο δεδομένων που περιλαμβάνει χαρακτηριστικά εισόδου και τις αντίστοιχες επιθυμητές ετικέτες. Ο κύριος στόχος είναι να μάθει μια συνάρτηση αντιστοίχισης από τις εισόδους στις επιθυμητές εξόδους, επιτρέποντας στο μοντέλο να κάνει προβλέψεις ή ταξινομήσεις σε δεδομένα με ακρίβεια. Η διαδικασία αυτή όπως γίνεται αντιληπτό απαντάται σε πληθώρα προβλημάτων όπως την αναγνώριση εικόνων, την επεξεργασία φυσικής γλώσσας και την προβλεπτική ανάλυση. Μερικοί από τους πιο σημαντικούς αλγορίθμους στην επιβλεπόμενη μάθηση περιλαμβάνουν τη γραμμική παλινδρόμηση, τα δέντρα απόφασης, τις μηχανές διανυσμάτων υποστήριξης και τα βαθιά νευρωνικά δίκτυα. Η επιβλεπόμενη μάθηση χρησιμοποιείται συνήθως σε προβλήματα όπου διατίθενται καλά καθορισμένες ετικέτες αναφοράς, επιτρέποντας στον αλγόριθμο να μάθει και να γενικεύσει αποτελεσματικά σε άγνωστα δεδομένα.

Από την άλλη πλευρά, η μη επιβλεπόμενη μηχανική μάθηση λειτουργεί χωρίς ετικεταρισμένα δεδομένα, βασιζόμενη σε ενδογενείς μοτίβα και δομές εντός των δεδομένων εισόδου. Στόχος της είναι να ανακαλύψει κρυμμένες συσχετίσεις, να ομαδοποιήσει παρόμοια δεδομένα ή να μειώσει τη διαστατικότητα πολύπλοκων συνόλων δεδομένων. Τεχνικές ομαδοποίησης και μείωσης διαστατικότητας αποτελούν παραδείγματα μη επιβλεπόμενης μάθησης. Ο αλγόριθμος K-means, η ιεραρχική ομαδοποίηση και η ανάλυση των κύριων συνιστωσών (PCA) είναι συνήθεις εργαλεία σε αυτόν τον τομέα. Η μη επιβλεπόμενη μάθηση είναι ιδιαίτερα χρήσιμη σε σενάρια όπου τα δεδομένα δεν διαθέτουν σαφείς ετικέτες ή όπου ο στόχος είναι η ανακάλυψη κρυμμένων δομών, όπως η τμηματοποίηση πελατών, η ανίχνευση ανωμαλιών και η δημιουργία συστημάτων προτάσεων.

Και η επιβλεπόμενη και η μη επιβλεπόμενη μηχανική μάθηση έχουν τα μοναδικά τους πλεονεκτήματα και εφαρμογές. Η επιλογή μεταξύ αυτών των παραδειγμάτων εξαρτάται από τη φύση του προβλήματος, τη διαθεσιμότητα των ετικετών και τους συγκεκριμένους στόχους της ανάλυσης δεδομένων ή του προβλήματος μοντελοποίησης που εκτελείται. Καθώς το πεδίο συνεχίζει να εξελίσσεται, οι αλληλεπιδράσεις μεταξύ αυτών των παραδειγμάτων, μαζί με την ενσωμάτωση της ενισχυτικής μάθησης, μια νεότερη μορφή της τεχνητής νοημοσύνης, υπόσχεται να ξεκλειδώσει ακόμη πιο εξειδικευμένες και ευέλικτες λύσεις/

3.3 Επιβλεπόμενη Μηχανική Μάθηση

Η επιβλεπόμενη μηχανική μάθηση είναι ένας τύπος μεθόδου μηχανικής μάθησης όπου το μοντέλο μαθαίνει να κάνει προβλέψεις από ετικετοποιημένα δεδομένα εκπαίδευσης. Ο όρος "εποπτευόμενος" αναφέρεται στην παρουσία ενός "επόπτη" ή ενός "δασκάλου" που παρέχει στον αλγόριθμο τις σωστές απαντήσεις κατά τη διάρκεια της εκπαίδευσης. Ο αλγόριθμος επαναληπτικά κάνει προβλέψεις στα δεδομένα εκπαίδευσης και διορθώνεται από τον δάσκαλο. Η διαδικασία μάθησης συνεχίζεται μέχρι το μοντέλο να πετύχει ένα αποδεκτό επίπεδο επίδοσης.

3.3.1 Βασικά επιβλεπόμενης Μηχανικής Μάθησης

Στην επιβλεπόμενη μάθηση, ο στόχος είναι να μάθουμε μια συνάρτηση αντιστοίχισης από τις εισόδους x στις εξόδους y, με βάση τα ζεύγη είσοδος-έξοδος στα δεδομένα εκπαίδευσης. Μπορούμε να συμβολίσουμε αυτήν τη συνάρτηση αντιστοίχισης ως $f(x) = y$. Τα προβλήματα εποπτευόμενης μάθησης κατηγοριοποιούνται σε προβλήματα παλινδρόμησης και ταξινόμησης.

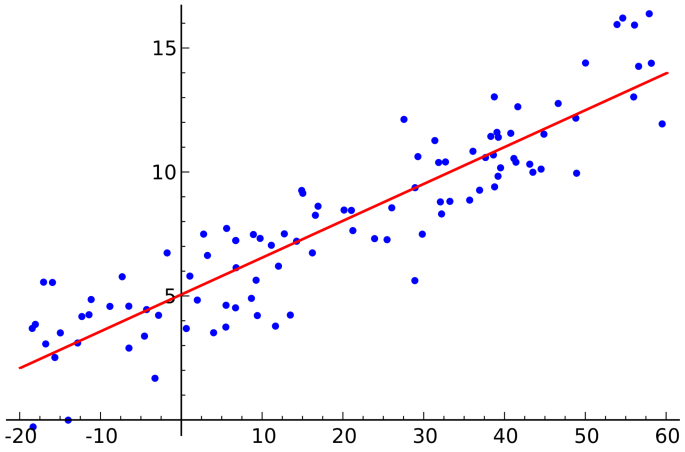
- Σε ένα πρόβλημα παλινδρόμησης, οι έξοδοι y είναι πραγματικές τιμές. Παραδείγματα περιλαμβάνουν την πρόβλεψη της τιμής ενός σπιτιού με βάση διάφορους παράγοντες όπως η έκτασή του, τον αριθμό των δωματίων, κ.λπ.
- Σε ένα πρόβλημα ταξινόμησης, οι έξοδοι y είναι διακριτές τιμές ή ετικέτες. Παραδείγματα περιλαμβάνουν την πρόβλεψη εάν ένα email είναι επικίνδυνο ή όχι, ή τη διάγνωση μιας ασθένειας με βάση τα συμπτώματα ενός ασθενή.

Η επιβλεπόμενη μάθηση χρησιμοποιείται ευρέως σε διάφορες εφαρμογές όπως η αναγνώριση ομιλίας, η ταξινόμηση εικόνων και η επεξεργασία φυσικής γλώσσας.

3.3.2 Αλγόριθμοι εποπτευόμενης Μηχανικής Μάθησης

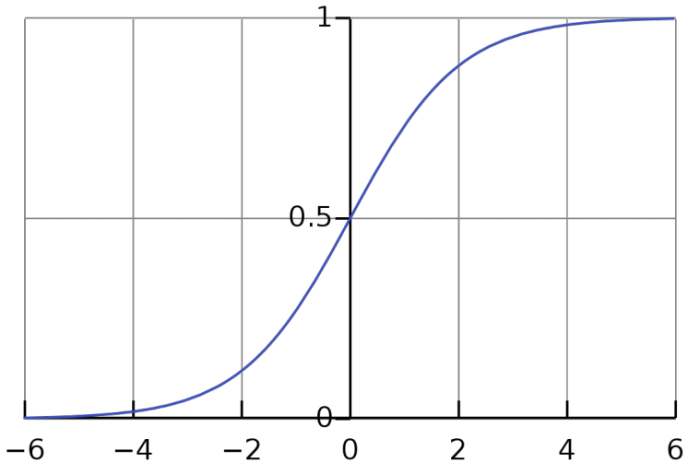
Υπάρχει ένας τεράστιος αριθμός αλγορίθμων εποπτευόμενης μάθησης διαθέσιμοι, καθένας με τα δυνατά και αδύνατα σημεία του. Η επιλογή του αλγορίθμου εξαρτάται από τον τύπο και την ποσότητα των διαθέσιμων δεδομένων, την πολυπλοκότητα του προβλήματος και τους υπολογιστικούς πόρους που διαθέτουμε. Εδώ είναι μερικοί συχνά χρησιμοποιούμενοι αλγόριθμοι:

- Γραμμική Παλινδρόμηση: Αυτή είναι μια στατιστική μέθοδος που χρησιμοποιείται για προγνωστική ανάλυση. Είναι ένας εποπτευόμενος αλγόριθμος μηχανικής μάθησης που προβλέπει συνεχείς τιμές.



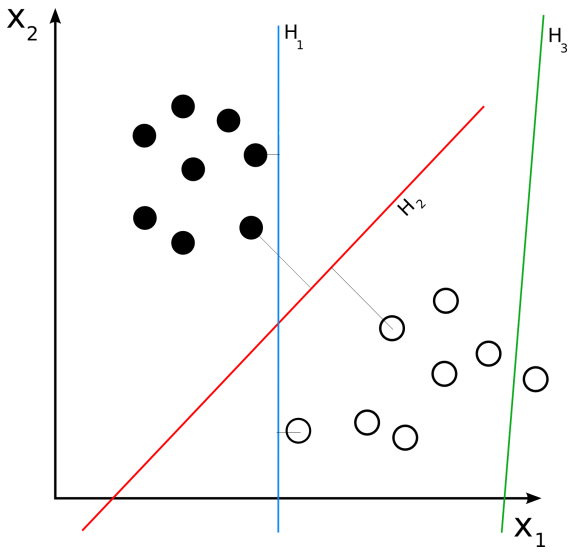
Σχήμα 3.1 Γραμμική Παλινδρόμηση

- Λογιστική Παλινδρόμηση: Παρά το όνομά της, η λογιστική παλινδρόμηση χρησιμοποιείται για δυαδικά προβλήματα ταξινόμησης. Προβλέπει την πιθανότητα εμφάνισης ενός γεγονότος προσαρμόζοντας τα δεδομένα σε μια λογιστική συνάρτηση.



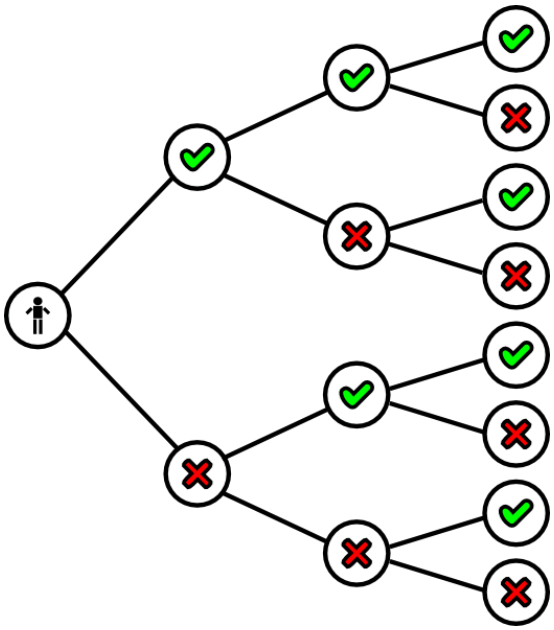
Σχήμα 3.2 Λογιστική Παλινδρόμηση

- Μηχανές Διανυσματικής Υποστήριξης (ΜΔΥ - Support Vector Machines SVMs): Οι ΜΔΥ χρησιμοποιούνται για τόσο προβλήματα παλινδρόμησης όσο και για προβλήματα ταξινόμησης. Στοχεύουν στον εντοπισμό ενός υπερεπιπέδου σε έναν N-διάστατο χώρο που ταξινομεί με σαφήνεια τα δεδομένα.



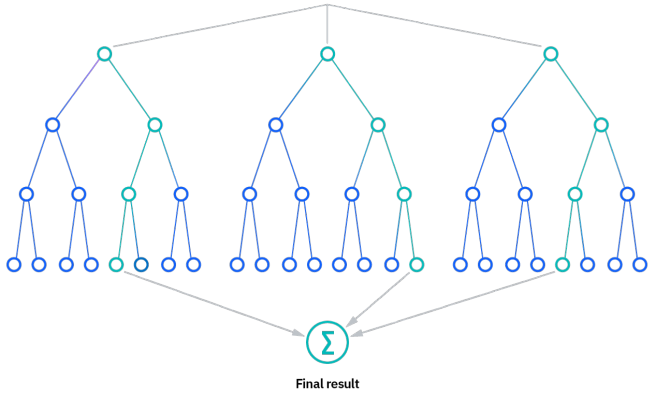
Σχήμα 3.3 Μηχανές Διανυσματικής Υποστήριξης

- Δέντρα Αποφάσεων - Decision Trees: Αυτός ο αλγόριθμος παίρνει αποφάσεις με βάση τις συνθήκες. Χρησιμοποιεί ένα δενδροειδές μοντέλο αποφάσεων για την ταξινόμηση και πρόβλεψη δεδομένων.



Σχήμα 3.4 Δέντρα Αποφάσεων

- Τυχαίο Δάσος - Random Forest: Το Τυχαίο Δάσος είναι ένας τύπος μεθόδου μάθησης συνόλου (ensemble) που λειτουργεί δημιουργώντας μια πληθώρα δέντρων αποφάσεων κατά το χρόνο εκπαίδευσης και εξάγοντας την κατηγορία που είναι η μέθοδος των κατηγοριών ή η μέση πρόβλεψη των ατομικών δένδρων.



Σχήμα 3.5 Τυχαίο Δάσος

- Gradient Boosting: Αυτή είναι μια μέθοδος ενίσχυσης που λειτουργεί δημιουργώντας σειριακά μοντέλα, όπου το κάθε επόμενο μοντέλο αποκαθιστά τα λάθη του προηγούμενου.

3.4 Νευρωνικά Δίκτυα και Βαθιά Μάθηση

3.4.1 Εισαγωγή

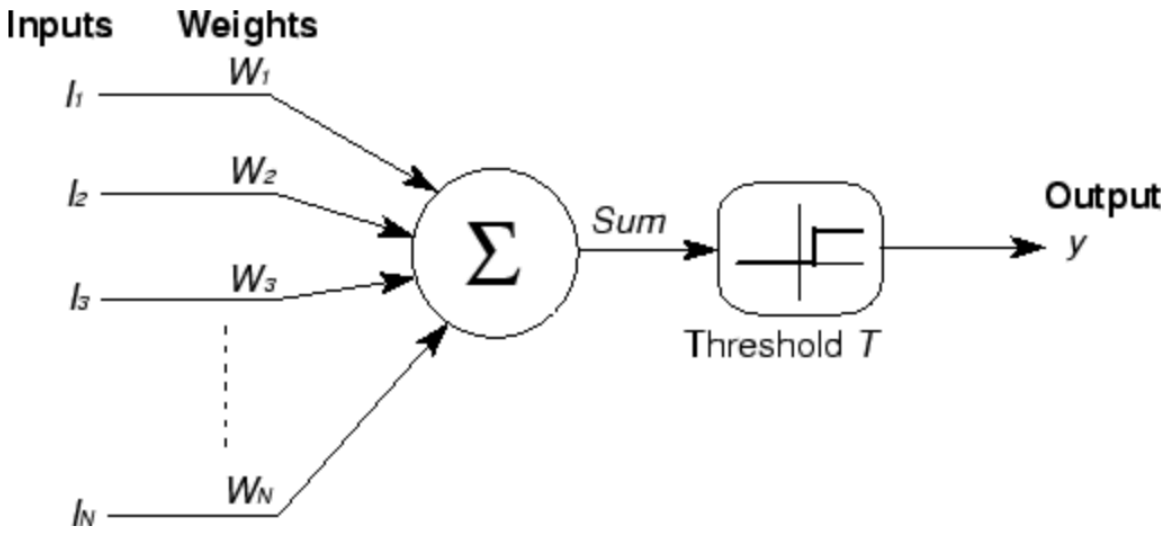
Η Βαθιά Μάθηση, ένα υποπεδίο της Μηχανικής Μάθησης, χρησιμοποιεί αλγόριθμους που προσπαθούν να μοντελοποιήσουν υψηλού επιπέδου αφαιρέσεις στα δεδομένα μέσω της χρήσης αρχιτεκτονικών που αποτελούνται από πολλαπλές μη γραμμικές μετατροπές. Στην καρδιά της Βαθιάς Μάθησης βρίσκεται η αξιοποίηση της δύναμης των νευρωνικών δικτύων, με το "βάθος" στη Βαθιά Μάθηση να υποδηλώνει την πολυπλοκότητα του δικτύου, η οποία καθορίζεται από τον αριθμό των στρωμάτων που περιέχει.

Οι αλγόριθμοι της Βαθιάς Μάθησης είναι σχεδιασμένοι για να μαθαίνουν από τεράστιες ποσότητες δεδομένων, όπου οι παραδοσιακοί αλγόριθμοι Μηχανικής Μάθησης συχνά αποτυγχάνουν. Η εμφάνιση των Μεγάλων Δεδομένων (Big Data) έχει προσφέρει στη Βαθιά Μάθηση τις τεράστιες ποσότητες πληροφοριών που χρειάζεται για να παράγει σημαντικά αποτελέσματα. Ως εκ τούτου, έχει βρει εφαρμογή σε μια ποικιλία τομέων, συμπεριλαμβανομένης της αναγνώρισης εικόνων και ομιλίας, της επεξεργασίας φυσικής γλώσσας, της ιατρικής διάγνωσης και τις επενδύσεις.

3.4.2 Μοντέλο τεχνητού νευρώνα

Ο τεχνητός νευρώνας είναι το δομικό υλικό των νευρωνικών δικτύων και διαδραματίζει κρίσιμο ρόλο στη λειτουργία τους. Σχεδιασμένος με έμπνευση από το βιολογικό νευρώνα που βρίσκεται

στον ανθρώπινο εγκέφαλο, ο τεχνητός νευρώνας είναι μια υπολογιστική μονάδα που μιμείται τις θεμελιώδεις ικανότητες επεξεργασίας πληροφοριών του βιολογικού αντίστοιχου.



Σχήμα 3.6 Ένας τεχνητός νευρώνας

Στην ουσία του, ένας τεχνητός νευρώνας λειτουργεί δεχόμενος πολλά σήματα εισόδου, καθένα από τα οποία συσχετίζεται με ένα συγκεκριμένο βάρος που αντικατοπτρίζει τη σημασία ή την επίδρασή του στον υπολογισμό. Τα σήματα εισόδου αυτά στη συνέχεια συνδυάζονται γραμμικά, λαμβάνοντας υπόψη τα αντίστοιχα βάρη τους. Το συνολικό άθροισμα των συνδυασμένων εισόδων περνά από μια συνάρτηση ενεργοποίησης, η οποία εισάγει μη γραμμικότητα στη διαδικασία. Αυτή η συνάρτηση ενεργοποίησης βοηθά το νευρώνα να ανιχνεύει πολύπλοκα μοτίβα και σχέσεις μέσα στα δεδομένα που επεξεργάζεται.

Η συνάρτηση ενεργοποίησης, ένα σημαντικό στοιχείο του τεχνητού νευρώνα, καθορίζει εάν ο νευρώνας θα "εκπέμψει" ή θα παράγει ένα σήμα εξόδου βάσει του συνολικού αθροίσματος των εισόδων του με βάση τα βάρη τους. Συνήθεις συναρτήσεις ενεργοποίησης περιλαμβάνουν τη συνάρτηση βήματος, τη συνάρτηση σιγμοειδούς καμπύλης και τη συνάρτηση γραμμικής ρύθμισης (Rectified Linear Unit - ReLU). Κάθε συνάρτηση ενεργοποίησης έχει τα δικά της χαρακτηριστικά και είναι κατάλληλη για διαφορετικές κατηγορίες εργασιών. Για παράδειγμα, η συνάρτηση βήματος είναι δυαδική και παράγει είτε 0 είτε 1 ως έξοδο, ενώ η συνάρτηση σιγμοειδούς καμπύλης παράγει τιμές ανάμεσα στο 0 και το 1, κάνοντάς την κατάλληλη για εργασίες που σχετίζονται με πιθανότητες.

Ένα από τα κύρια χαρακτηριστικά των τεχνητών νευρώνων και των νευρωνικών δικτύων είναι η δυνατότητά τους να μαθαίνουν από τα δεδομένα. Κατά τη διάρκεια της διαδικασίας εκπαίδευσης, οι παράμετροι του τεχνητού νευρώνα, όπως τα βάρη και οι πολώσεις (bias) του, προσαρμόζονται επανειλημμένα για να ελαχιστοποιηθεί το σφάλμα μεταξύ της προβλεπόμενης

εξόδου και της πραγματικής επιθυμητής εξόδου. Αυτή η προσαρμογή γίνεται συνήθως χρησιμοποιώντας αλγόριθμους βελτιστοποίησης όπως η κλίση καθόδου (gradient descent) κατά την κατεύθυνση της κλίσης της ευθείας της παραγώγου. Μέσω αυτής της διαδικασίας μάθησης, ο τεχνητός νευρώνας προσαρμόζεται στα δεδομένα που συναντά, επιτρέποντάς του να ανιχνεύει περίπλοκα μοτίβα και να κάνει όλο και πιο ακριβείς προβλέψεις.

Οι τεχνητοί νευρώνες δεν είναι απομονωμένες μονάδες αλλά συνδέονται για να σχηματίσουν νευρωνικά δίκτυα. Αυτά τα δίκτυα μπορεί να κυμαίνονται από απλά μονό-επίπεδα μέχρι βαθιά νευρωνικά δίκτυα με πολλά κρυμμένα επίπεδα. Η διάταξη και η συνδεσιμότητα των νευρώνων σε ένα δίκτυο καθορίζουν την αρχιτεκτονική του και, επομένως, τη δυνατότητά του να αντιμετωπίσει συγκεκριμένες εργασίες. Τα βαθιά νευρωνικά δίκτυα, με τα πολλά επίπεδα τεχνητών νευρώνων, έχουν επαναπροσδιορίσει τα πεδία όπως η όραση υπολογιστών, η επεξεργασία φυσικής γλώσσας και η αναγνώριση φωνής, επιτυγχάνοντας εντυπωσιακά αποτελέσματα σε εργασίες που κάποτε θεωρούνταν εξαιρετικά δύσκολες για τους υπολογιστές.

3.4.3 Αρχιτεκτονικές Νευρωνικών δικτύων

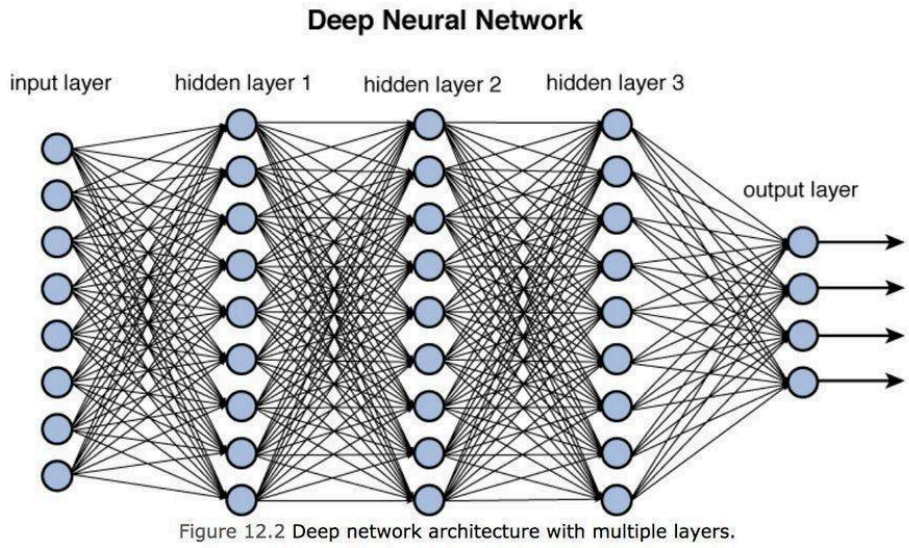
Η Βαθιά Μάθηση περιλαμβάνει διάφορους τύπους νευρωνικών δικτύων, με κάθε ένα να ειδικεύεται σε συγκεκριμένα προβλήματα.

Δίκτυα Προς τα Εμπρός (ΔΠΕ - FNNs)

Τα Δίκτυα Προς τα Εμπρός (FNNs), που επίσης αναφέρονται ως Πολυστρωματικά Perceptron (Multi-layer Perceptrons - MLPs), είναι ο πιο απλός τύπος τεχνητού νευρωνικού δικτύου. Σε αυτά τα δίκτυα, η πληροφορία κινείται σε μία κατεύθυνση - από την είσοδο προς την έξοδο - χωρίς να γίνεται επαναφορά. Αναφέρονται ως δίκτυα προς τα εμπρός επειδή η πληροφορία μεταφέρεται μόνο προς τα εμπρός στο δίκτυο, χωρίς βρόγχους ή κύκλους.

Τα ΔΠΕ αποτελούνται από τρία είδη στρωμάτων:

- **Είσοδος:** Το πρώτο στρώμα (layer) του δικτύου, όπου τα ακατέργαστα δεδομένα εισάγονται για περαιτέρω επεξεργασία από τους τεχνητούς νευρώνες του δικτύου.
- **Κρυφά Στρώματα:** Τα στρώματα μετά το στρώμα εισόδου είναι γνωστά ως κρυφά στρώματα, όπου γίνεται η πραγματική επεξεργασία μέσω ενός συστήματος συνδέσεων με βάρη. Τα κρυφά στρώματα εφαρμόζουν μετασχηματισμούς στα δεδομένα εισόδου για τον υπολογισμό της εξόδου. Αυτά τα στρώματα χρησιμοποιούν τα βάρη και τις προκαθορισμένες τιμές (biases), που ενημερώνονται κατά τη διάρκεια της εκπαίδευσης, για να βοηθήσουν το μοντέλο να μάθει πολύπλοκα μοτίβα.
- **Έξοδος:** Το τελικό στρώμα, όπου γίνεται η τελική επεξεργασία και παρέχεται η έξοδος.



Σχήμα 3.7 Fully-Connected FFN

Κάθε νευρώνας σε ένα στρώμα συνδέεται με κάθε άλλο νευρώνα στο επόμενο στρώμα, καθιστώντας τα ΔΠΕ πλήρως συνδεδεμένα δίκτυα. Αυτό είναι γνωστό ως τοπολογία πλήρους σύνδεσης (Σχήμα 3.7).

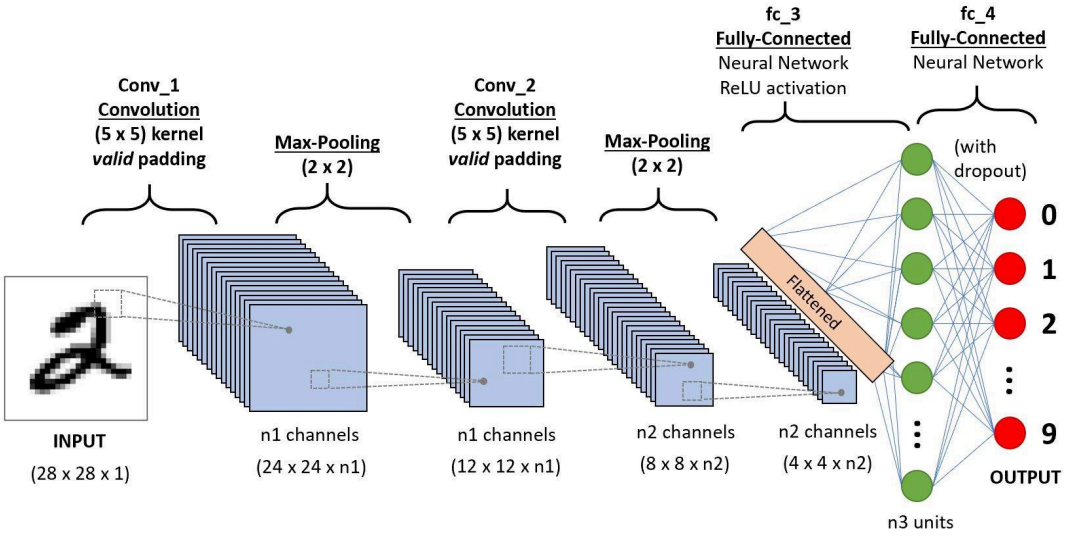
Το κυρίως μειονέκτημα των ΔΠΕ είναι ότι μπορεί να γίνουν πολύπλοκα και δύσκολα στην εκπαίδευση όταν υπάρχουν πολλά κρυφά στρώματα. Αυτό είναι γνωστό ως το πρόβλημα εξαφανιζόμενων κλίσεων (vanishing gradient problem), το οποίο μπορεί να κάνει δύσκολη την εκπαίδευση του δικτύου, επειδή οι παράγοντες βαθμού γίνονται πολύ μικροί καθώς περνάμε προς τα εμπρός μέσα στο δίκτυο.

Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ - CNNs)

Τα Συνελικτικά Νευρωνικά Δίκτυα (CNNs), που παρουσιάστηκαν για πρώτη φορά από τους LeCun κ.ά., το 1989, είναι ένα εξειδικευμένο είδος νευρωνικών δικτύων για την επεξεργασία δεδομένων που έχουν τοπολογία πίνακα. Το όνομα «συνελικτικό» υπονοεί ότι το δίκτυο χρησιμοποιεί μια μαθηματική λειτουργία που ονομάζεται συνέλιξη, ένα εξειδικευμένο είδος γραμμικής πράξης. Τα συνελικτικά δίκτυα είναι απλά νευρωνικά δίκτυα που χρησιμοποιούν συνέλιξη αντί για γενική πολλαπλασιαστική πράξη μήτρας τουλάχιστον σε ένα από τα επίπεδά τους.

Ένα τυπικό ΣΝΔ αποτελείται από τρία είδη επιπέδων: συνελικτικά, συγκεντρωτικά και πλήρως συνδεδεμένα. Συνήθως διοργανώνονται σε αυτήν την ακολουθία:

Convolutional -> Pool -> Convolutional -> Pool -> Fully Connected -> Fully Connected.



Σχήμα 3.8 Τυπικό CNN

1. Συνελικτικό Επίπεδο**:** Τα συνελικτικά επίπεδα εφαρμόζουν μια συνεκτική πράξη στην είσοδο, περνώντας το αποτέλεσμα στο επόμενο επίπεδο. Η συνέλιξη προσομοιώνει την απάντηση ενός μεμονωμένου νευρώνα στα οπτικά ερεθίσματα. Αυτά τα συνελικτικά επίπεδα έχουν παραμέτρους (βάρη και προκαθορισμένες τιμές-bias) που εκπαιδεύονται κατά τη διαδικασία της εκπαίδευσης. Κάθε νευρώνας στο συνελικτικό επίπεδο είναι συνδεδεμένος με μια μικρή περιοχή στον εισερχόμενο όγκο των δεδομένων, μέσω ενός συνόλου εκπαιδεύσιμων βαρών που ονομάζεται φίλτρο ή πυρήνας. Με το να διασχίζουμε (συνελικτικά) αυτά τα φίλτρα στο πλάτος και το ύψος του εισερχόμενου όγκου, παράγεται ένας δισδιάστατος χάρτης ενεργοποίησης. Ο χάρτης ενεργοποίησης δείχνει τις τοποθεσίες όπου ενεργοποιήθηκε το φίλτρο (δηλαδή βρέθηκε ένα μοτίβο που έψαχνε).

2. Συγκεντρωτικό Επίπεδο (Pooling Layers)**:** Τα συγκεντρωτικά επίπεδα μειώνουν τις διαστάσεις των δεδομένων συνδυάζοντας τις εξόδους των προηγούμενων επιπέδων νευρώνων σε ένα επίπεδο σε έναν μεμονωμένο νευρώνα στο επόμενο επίπεδο. Υπάρχουν διάφορες τύποι συγκεντρωτικών λειτουργιών, με το max και average pooling να είναι ένα τα πιο συνηθισμένα. Το max pooling διαιρεί την εισερχόμενη εικόνα σε ένα σύνολο μη επικαλυπτόμενων ορθογώνιων και, για κάθε τέτοια υποπεριοχή, εξάγει το μέγιστο.

3. Πλήρως Συνδεδεμένο Επίπεδο**:** Τα πλήρως συνδεδεμένα επίπεδα συνδέουν κάθε νευρώνα σε ένα επίπεδο με κάθε νευρώνα σε άλλο επίπεδο. Αυτό είναι το ίδιο με την τακτική λειτουργία του κανονικού νευρωνικού δικτύου. Το πλήρως συνδεδεμένο επίπεδο χρησιμοποιείται συνήθως για το τελικό μέρος ταξινόμησης του ΣΝΔ, όπου κάθε νευρώνας αντιπροσωπεύει μια συγκεκριμένη κατηγορία.

Τα CNNs χρησιμοποιούνται πιο συνηθισμένα για την ανάλυση οπτικής εικονογραφίας και είναι επίσης ευρέως χρησιμοποιημένα σε άλλες εφαρμογές, όπως η επεξεργασία φυσικής γλώσσας, οι χρηματοοικονομικές χρονοσειρές (Liang, 2018) και η ιατρική διάγνωση.

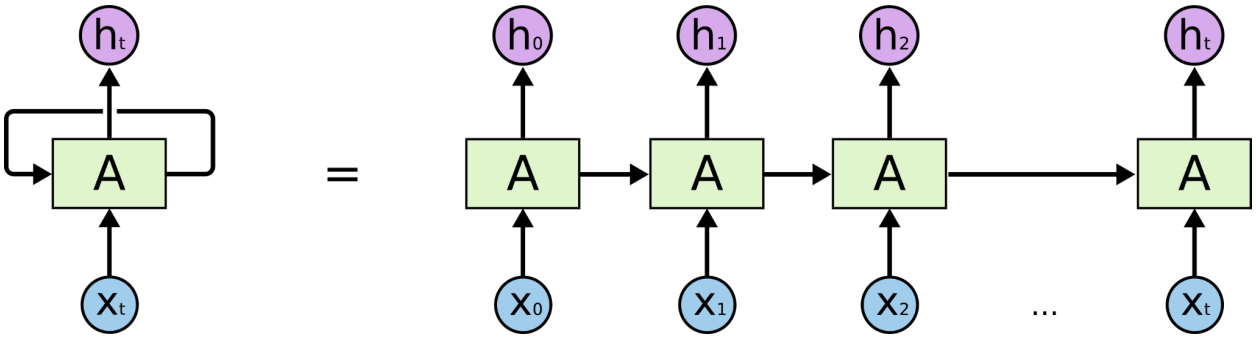
Αυτά τα δίκτυα είναι ιδιαίτερα καλά στην αναγνώριση μοτίβων που είναι χωρικά και ιεραρχικά κατανεμημένα στον δισδιάστατο χώρο. Αυτή η ικανότητα επιτυγχάνεται μέσω της συνέλιξης και του συγκεντρωτικού επιπέδου, που μειώνουν την πολυπλοκότητα των δεδομένων και εξάγουν τα πιο σημαντικά χαρακτηριστικά.

Αναδρομικά Νευρωνικά Δίκτυα (ΑΝΔ - RNNs)

Τα Αναδρομικά Νευρωνικά Δίκτυα (RNNs) αποτελούν μια κατηγορία τεχνητών νευρωνικών δικτύων όπου οι συνδέσεις μεταξύ των κόμβων δημιουργούν ένα κατευθυνόμενο γράφημα κατά την χρονική ακολουθία. Σε αντίθεση με τα προς τα εμπρός νευρωνικά δίκτυα, τα ΑΝΔ έχουν κυκλικές συνδέσεις που τα καθιστούν ισχυρά για την μοντελοποίηση ακολουθιών. Ονομάζονται "αναδρομικά" γιατί εκτελούν το ίδιο καθήκον για κάθε στοιχείο μιας ακολουθίας, με την έξοδο να εξαρτάται από τους προηγούμενους υπολογισμούς (Elman, 1990).

Τα ΑΝΔ έχουν μια εσωτερική κατάσταση που μπορεί να αναπαριστά πληροφορίες σχετικά με την παρελθούσα ακολουθία των εισόδων. Αυτή η ιδιότητα καθιστά τα ΑΝΔ πολύ κατάλληλα για εργασίες που σχετίζονται με σειριακά δεδομένα, όπως πρόβλεψη σειρών χρόνου, επεξεργασία φυσικής γλώσσας και αναγνώριση ομιλίας.

Η πιο απλή μορφή ενός RNN μπορεί να οπτικοποιηθεί ως εξής:



Σχήμα 3.9 Αναδρομικό Νευρωνικό Δίκτυο

Από το Σχήμα 3.9, βλέπουμε ότι ένα ΑΝΔ είναι σαν πολλαπλά αντίγραφα του ίδιου δικτύου, όπου κάθε ένα περνάει ένα μήνυμα σε έναν διάδοχο.

Αυτο-ενσωματωτές (AEs)

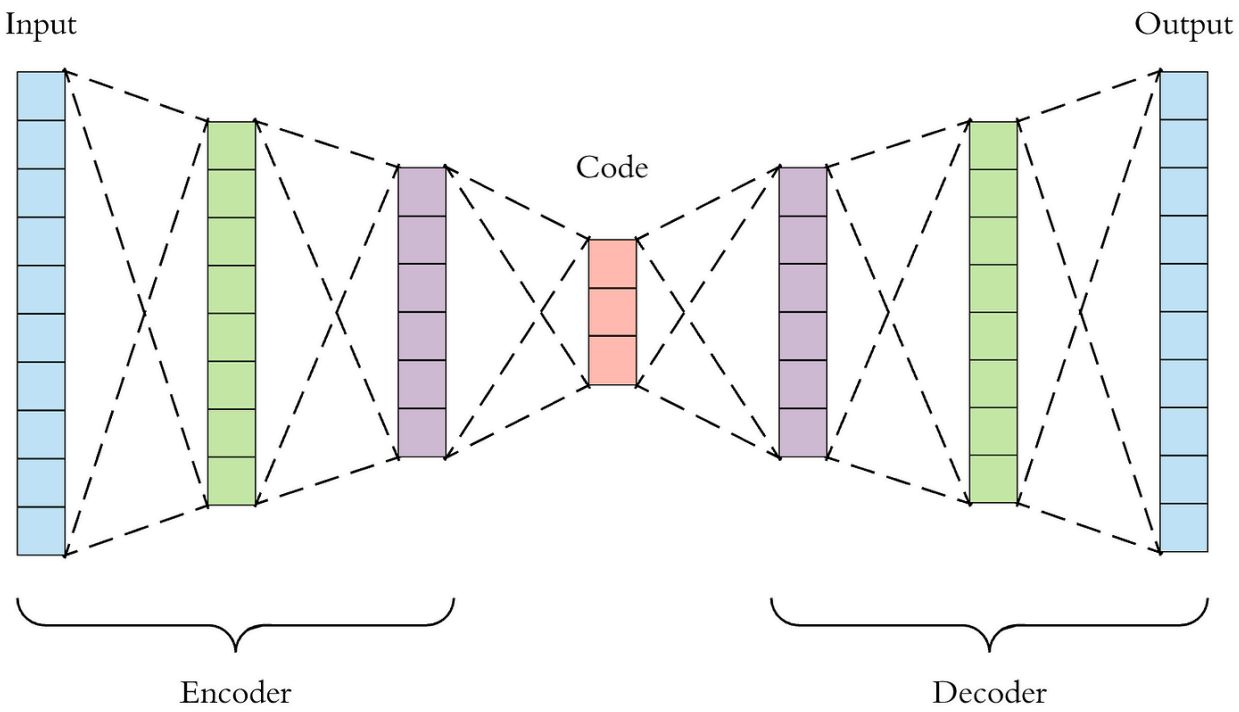
Οι αυτοενσωματωτές είναι ένας συγκεκριμένος τύπος των feedforward νευρωνικών δικτύων που χρησιμοποιούνται γενικά για την εκμάθηση αποτελεσματικών κωδικοποιήσεων δεδομένων με μη εποπτευόμενο τρόπο. Ο στόχος ενός αυτοενσωματωτή είναι να μάθει μια αναπαράσταση (κωδικοποίηση) για ένα σύνολο δεδομένων, συνήθως για μείωση της διαστατικότητας.

Η αρχιτεκτονική ενός αυτοενσωματωτή είναι συμμετρική με μια κεντρική κρυφή στρώση (το στρώμα κωδικοποίησης) που περιγράφει την είσοδο, περιβαλλόμενη από στρώσεις στις δύο πλευρές (ο κωδικοποιητής και ο αποκωδικοποιητής). Τα δεδομένα περνούν μέσα από τις στρώσεις του κωδικοποιητή μέχρι να φτάσουν στο στρώμα κωδικοποίησης και στη συνέχεια περνούν μέσα από τις στρώσεις του αποκωδικοποιητή για να αναδημιουργήσουν την αρχική είσοδο.

Ένας αυτοενσωματωτής αποτελείται από δύο κύρια μέρη:

1. **Κωδικοποιητής:** Αυτό το μέρος του δικτύου συμπιέζει την είσοδο σε μια αναπαράσταση μικρότερης διάστασης. Μπορεί να αναπαρασταθεί από μια κωδικοποιητική συνάρτηση h=f(x).
2. **Αποκωδικοποιητής:** Αυτό το μέρος στοχεύει στην ανακατασκευή της εισόδου από την επεξεργασμένη αναπαράσταση. Μπορεί να αναπαρασταθεί από μια αποκωδικοποιητική συνάρτηση r=g(h).

Η απλούστερη μορφή ενός αυτοενσωματωτή είναι ένα feedforward, μη αναδρομικό νευρωνικό δίκτυο παρόμοιο με το πολύπλευρο περσεπτρόν (MLP) - αλλά με το στρώμα εξόδου να έχει τον ίδιο αριθμό κόμβων με το στρώμα εισόδου, και με τον σκοπό της ανακατασκευής των δικών της εισόδων. Το δίκτυο μπορεί να οπτικοποιηθεί ως εξής:



Σχήμα 3.10 Αυτοενσωματωτής

Ο στόχος ενός αυτοενσωματωτή είναι να ελαχιστοποιήσει το σφάλμα ανακατασκευής το οποίο μπορεί να οριστεί ως η διαφορά μεταξύ της εισόδου και της εξόδου. Η διαφορά υπολογίζεται χρησιμοποιώντας μια συνάρτηση απώλειας, η οποία συχνά είναι το μέσο τετραγωνικό σφάλμα (MSE).

Οι αυτοενσωματωτές μπορούν να χρησιμοποιηθούν αποτελεσματικά για την επίλυση προβλημάτων όπως η ανίχνευση ασυνήθιστων σημείων ή η αποθορυβοποίηση, όπου οι μετασχηματισμοί δεδομένων που εκτελούνται από τον αυτοενσωματωτή μπορούν να χρησιμοποιηθούν για την ανίχνευση ασυνήθιστων σημείων δεδομένων ή την αφαίρεση θορύβου από τα δεδομένα.

Γενετικά Ανταγωνιστικά Δίκτυα (GANs)

Τα Γενετικά Ανταγωνιστικά Δίκτυα (GANs) παρουσιάστηκαν από τον Ian Goodfellow και άλλους το 2014. Πρόκειται για μια τεχνική μηχανικής μάθησης όπου δύο νευρωνικά δίκτυα αντιπαρατίθενται μεταξύ τους, πράγμα που οδηγεί στη δημιουργία νέων, τεχνιτών δειγμάτων που μοιάζουν με τα δεδομένα εκπαίδευσής.

Τα δύο νευρωνικά δίκτυα είναι:

1. **Δημιουργός (Generator):** Αυτό το δίκτυο λαμβάνει τυχαίο θόρυβο ως είσοδο και παράγει δεδομένα (όπως μια εικόνα).

2. **Διακριτής (Discriminator):** Αυτό το δίκτυο λαμβάνει πραγματικά δείγματα δεδομένων και την έξοδο του δημιουργού ως είσοδο, και έχει το έργο να κατατάξει αν κάθε περίπτωση είναι πραγματική (προέρχεται από το πραγματικό σύνολο δεδομένων) ή ψεύτικη (δημιουργήθηκε από τον δημιουργό).

Ο δημιουργός προσπαθεί να ξεγελάσει τον διακριτή και ο διακριτής προσπαθεί να αποφύγει την απάτη βελτιώνοντας την ικανότητά του να διακρίνει πραγματικά δεδομένα από τα δεδομένα που παράγει ο δημιουργός. Ο δημιουργός εκπαιδεύεται για να κάνει τον διακριτή να ταξινομεί την έξοδό του ως πραγματική.

Αυτά τα αντιπαλεύοντα δίκτυα συνεχίζουν να μαθαίνουν το ένα από το άλλο, βελτιώνοντας την ποιότητα των παραγόμενων περιπτώσεων. Ο δημιουργός μαθαίνει να παράγει όλο και πιο ρεαλιστική έξοδο, προσπαθώντας να ξεγελάσει τον διακριτή, και ο διακριτής γίνεται όλο και καλύτερος στη διάκριση μεταξύ πραγματικών και ψεύτικων.

Αυτός ο είδος αρχιτεκτονικής έχει δει επιτυχημένες εφαρμογές στη δημιουργία ρεαλιστικών εικόνων, βίντεο, φωνής, και ακόμα και στην προσομοίωση φυσικών συστημάτων.

3.4.4 Βελτιστοποιητές - Optimizers

Οι βελτιστοποιητές νευρωνικών δικτύων παίζουν καίριο ρόλο στην εκπαίδευση μοντέλων βαθιάς μάθησης, επιτρέποντάς τους να μάθουν πολύπλοκα μοτίβα και να κάνουν ακριβείς προβλέψεις σε διάφορους τομείς, από την αναγνώριση εικόνας έως την επεξεργασία φυσικής γλώσσας. Βρίσκονται στην καρδιά της διαδικασίας εύρεσης της βέλτιστης λύσης και προσαρμόζουν επαναληπτικά τις παραμέτρους του μοντέλου για να ελαχιστοποιήσουν τη συνάρτηση απώλειας και να βελτιώσουν την απόδοση.

Η βελτιστοποίηση βασισμένη στην κατάβαση κλίσης (gradient descent) αποτελεί τον πυρήνα της εκπαίδευσης των νευρωνικών δικτύων. Χρησιμοποιεί την παράγωγο της συνάρτησης απώλειας ως προς τις παραμέτρους του μοντέλου για να προσαρμόσει αυτές τις παραμέτρους επαναλαμβανόμενα. Ο πιο θεμελιώδης βελτιστοποιητής είναι η κατάβαση κλίσης, που ενημερώνει τις παραμέτρους του μοντέλου προς την κατεύθυνση που οδηγούν στην πιο απότομη πτώσης της συνάρτησης απώλειας. Ενώ η τεχνική αυτή αποτελεί τον θεμέλιο λίθο της βελτιστοποίησης στα νευρωνικά δίκτυα, έχει αρκετά προβλήματα όπως την ευαισθησία στην επιλογή του ρυθμού μάθησης (learning rate) και τα προβλήματα σύγκλισης σε βαθιά δίκτυα.

Για να αντιμετωπίσει αυτές τις προκλήσεις, έχουν αναπτυχθεί διάφοροι προηγμένοι αλγόριθμοι βελτιστοποίησης. Ένας τέτοιος βελτιστοποιητής είναι η Στοχαστική Κατάβαση Κλίσης (Stochastic Gradient Descent - SGD), που εισάγει τυχαιότητα επιλέγοντας ένα τυχαίο μικρό πακέτο δεδομένων (mini-batch) σε κάθε επανάληψη. Αυτή η τυχαιότητα βοηθά να αποφεύγονται τοπικά ελάχιστα και να επιταχύνεται η σύγκλιση. Ένας άλλος δημοφιλής βελτιστοποιητής, το

Momentum, ενσωματώνει έναν κινούμενο μέσο όρο των προηγούμενων παραγώγων για να επιταχύνει τη σύγκλιση στη σχετική κατεύθυνση, με αποτέλεσμα να μειώνονται οι ταλαντώσεις.

Οι βελτιστοποιητές που προσαρμόζουν τον ρυθμο μάθησης, όπως ο Adagrad, ο RMSprop και ο Adam, έχουν κερδίσει ευρεία δημοτικότητα λόγω της ικανότητάς τους να προσαρμόζουν το ρυθμό μάθησης για κάθε παράμετρο. Ο Adagrad διατηρεί έναν διαφορετικό ρυθμό μάθησης για κάθε παράμετρο, επιτρέποντας την ταχεία μάθηση για τις παραμέτρους που ενημερώνονται σπάνια και την αργή μάθηση για τις παραμέτρους που ενημερώνονται συχνά. Ο RMSprop, από την άλλη, χρησιμοποιεί έναν κινούμενο μέσο όρο των τετραγώνων των παραγώγων για να προσαρμόζει το ρυθμό μάθησης ανεξάρτητα για κάθε παράμετρο. Ο Adam συνδυάζει τα οφέλη και του Momentum και του Adagrad, καθιστώντας τον έναν από τους πιο διαδεδομένους βελτιστοποιητές στην πράξη.

Ένας άλλος βελτιστοποιητής που αξίζει να αναφερθεί είναι ο Επιταχυνόμενης Κλίσης Nesterov (Nesterov Accelerated Gradient - NAG), που αποτελεί βελτίωση έναντι του κανονικού Momentum. Ο NAG υπολογίζει την παράγωγο της γραμμής απώλειας ως προς την προβλεπόμενη μελλοντική θέση των παραμέτρων αντί της τρέχουσας θέσης, με αποτέλεσμα να γίνονται πιο ακριβείς ενημερώσεις και να επιταχύνεται η σύγκλιση.

Μια πρόσφατη προσθήκη στον τομέα των βελτιστοποιητών είναι ο βελτιστοποιητής L-BFGS (Limited-memory Broyden-Fletcher-Goldfarb-Shanno). Ο L-BFGS είναι ένας προσεγγιστικός αλγόριθμος του Μπρόιντεν-Φλέτσερ-Γκόλντφαρμπ-Σάννο (Broyden-Fletcher-Goldfarb-Shanno), που προσεγγίζει τον πίνακα Hessian της συνάρτησης απώλειας, επιτρέποντάς του να κάνει πιο ακριβείς ενημερώσεις. Παρόλο που είναι υπολογιστικά ακριβός και απαιτητικός σε μνήμη, είναι ιδιαίτερα χρήσιμος για μικρότερα δίκτυα με περιορισμένα δεδομένα εκπαίδευσης.

Για τους βελτιστοποιητές νευρωνικών δικτύων δεν υπάρχει μία προσέγγιση που να ταιριάζει σε όλες τις περιπτώσεις, και η επιλογή του βελτιστοποιητή εξαρτάται συχνά από το συγκεκριμένο πρόβλημα και την αρχιτεκτονική του νευρωνικού δικτύου.

3.5 Ενισχυτική Μάθηση

3.5.1 Διαδικασίες Μάρκοφ

Στο πλαίσιο της Ενισχυτικής Μάθησης, το θέμα των διαδικασιών Μάρκοφ είναι θεμελιώδες. Μια διαδικασία Μάρκοφ, ή αλυσίδα Μάρκοφ, που πήρε το όνομά της από τον ρώσο μαθηματικό Αντρέι Μάρκοφ, είναι ένα στοχαστικό μοντέλο που περιγράφει μια ακολουθία δυνατών γεγονότων στα οποία η πιθανότητα κάθε γεγονότος εξαρτάται μόνο από την κατάσταση που βρισκόταν το σύστημα στο προηγούμενο γεγονός (Sutton, 2018).

Τα δύο κύρια συστατικά μιας διαδικασίας Μάρκοφ είναι:

1. **Κατάσταση (S)**: Μια κατάσταση αντιπροσωπεύει την τρέχουσα κατάσταση του συστήματος. Για παράδειγμα, η τρέχουσα θέση ενός πιονιού σε μια σκακιέρα, ή οι τρέχουσες τιμές μιας σειράς μετοχών.
2. **Πιθανότητα Μετάβασης (P)**: Η πιθανότητα μετάβασης περιγράφει τους κανόνες που διέπουν τη μετάβαση από τη μία κατάσταση στην επόμενη. Συγκεκριμένα, περιγράφει την πιθανότητα η επόμενη κατάσταση του συστήματος να είναι η s' δεδομένου ότι γίνεται η ενέργειας a στην τρέχουσα κατάσταση s.

Οι διαδικασίες Μάρκοφ είναι "χωρίς μνήμη" (memoryless), που σημαίνει ότι οι μελλοντικές καταστάσεις εξαρτώνται μόνο από την παρούσα κατάσταση και όχι από την ακολουθία των γεγονότων που προηγήθηκαν. Αυτή η ιδιότητα ονομάζεται ιδιότητα Μάρκοφ.

Στην ενισχυτική μάθηση, οι αλυσίδες Μάρκοφ συχνά επεκτείνονται σε ένα πιο γενικό πρότυπο γνωστό ως Markov Decision Process (MDP), όπου η πιθανότητα μετάβασης εξαρτάται τόσο από την τρέχουσα κατάσταση όσο και από τις ενέργειες που λαμβάνει ο πράκτορας.

Επίσης, μια κατάσταση στη διαδικασία Markov είναι μια εννοιολογική αναπαράσταση της κατάστασης ενός συστήματος, περιβάλλοντος ή πράκτορα, που αναμένεται να αλλάζει με την πάροδο του χρόνου. Η κατάσταση περιλαμβάνει όλες τις σχετικές πληροφορίες που χαρακτηρίζουν την τρέχουσα κατάσταση ή πλαίσιο του συστήματος ή του πράκτορα. Στην ενισχυτική μάθηση, η κατάσταση χρησιμοποιείται από τον πράκτορα για να λάβει αποφάσεις σχετικά με τις ενέργειες που πρέπει να πάρει. Το σύνολο όλων των πιθανών καταστάσεων στο περιβάλλον αναπαρίσταται από τον χώρο της κατάστασης, που συμβολίζεται ως S.

Σε μια διαδικασία Μάρκοφ, υποτίθεται ότι η τρέχουσα κατάσταση έχει την ιδιότητα Μάρκοφ. Αυτό σημαίνει ότι περιέχει όλες τις απαραίτητες πληροφορίες που χρειάζονται για να προβλεφθεί το μέλλον, δεδομένης της τρέχουσας κατάστασης και αγνοώντας την ιστορία. Αυτή είναι μια κύρια υπόθεση στις Διαδικασίες Απόφασης Μάρκοφ (MDPs), όπου η πιθανότητα μετάβασης σε οποιαδήποτε συγκεκριμένη κατάσταση καθορίζεται μόνο από την τρέχουσα

κατάσταση και τις ενέργειες του λήπτη των αποφάσεων. Η υπόθεση αυτή αποτελεί και τον θεμέλιο λίθο της Ενισχυτικής Μάθησης.

Οι Διαδικασίες Απόφασης Μάρκοφ (MDP) αποτελούν τον πυρήνα πολλών αλγορίθμων ενισχυτικής μάθησης. Το πλαίσιο MDP παρέχει μια μαθηματική προσέγγιση στα διαδοχικά προβλήματα λήψης αποφάσεων, όπου τα αποτελέσματα είναι εν μέρει τυχαία και εν μέρει υπό τον έλεγχο ενός λήπτη αποφάσεων, πράκτορα στην Ενισχυτική Μάθηση.

Οι MDP χαρακτηρίζονται από μια 5-άδα (S, A, P, R, γ) όπου:

- S είναι ένα πεπερασμένο σύνολο καταστάσεων.
- A είναι ένα πεπερασμένο σύνολο ενεργειών.
- P είναι ένας πίνακας πιθανοτήτων μετάβασης καταστάσεων, P(s'|s, a), ο οποίος καθορίζει την πιθανότητα μετάβασης στην κατάσταση s' δεδομένης της τρέχουσας κατάστασης s και της ενέργειας a.
- R είναι μια συνάρτηση ανταμοιβής, R(s, a, s'), η οποία δίνει την αναμενόμενη άμεση ανταμοιβή που λαμβάνεται μετά τη μετάβαση από την κατάσταση s στην κατάσταση s', λόγω της ενέργειας a.
- γ είναι ο παράγοντας έκπτωσης ο οποίος καθορίζει την τρέχουσα αξία των μελλοντικών ανταμοιβών. Κυμαίνεται μεταξύ 0 και 1, όπου γ κοντά στο 0 κάνει τον πράκτορα "βραχυπρόθεσμο" (επικεντρωμένο στις άμεσες αμοιβές), ενώ γ κοντά στο 1 κάνει τον πράκτορα "μακροπρόθεσμο" (επικεντρωμένο στις μακροπρόθεσμες αμοιβές).

Σε μια δεδομένη κατάσταση s, ο πράκτορας επιλέγει μια ενέργεια a βάσει μιας πολιτικής π(a|s), η οποία είναι μια πιθανοτική κατανομή πάνω στο σύνολο ενεργειών. Ο στόχος της Ενισχυτικής Μάθησης είναι να υπολογίσει την βέλτιστη πολιτική π* που μεγιστοποιεί την αναμενόμενη συσσωρευτική ανταμοιβή δεδομένου της συνάρτησης R.

Κάποιος μπορεί να λύσει τις Διαδικασίες Απόφασης Μάρκοφ χρησιμοποιώντας τεχνικές δυναμικού προγραμματισμού όταν το μοντέλο είναι πλήρως γνωστό, ή να χρησιμοποιήσει μεθόδους ενισχυτικής μάθησης όπως η Q-learning όταν το μοντέλο δεν είναι γνωστό.

3.5.2 Βασικά χαρακτηριστικά της Ενισχυτικής Μάθησης

Η Ενισχυτική Μάθηση (RL) είναι ένας τύπος μηχανικής μάθησης, όπου ένας πράκτορας (agent) μαθαίνει πώς να λαμβάνει σωστές αποφάσεις μέσω αλληλεπιδράσεων με το περιβάλλον του. Αυτό το πλαίσιο μάθησης διαφέρει από την επιβλεπόμενη και την μη επιβλεπόμενη μάθηση, καθώς ο πράκτορας δεν διαθέτει ετικέτες ή υποδείξεις για το πώς πρέπει να ενεργήσει. Αντίθετα, ο πράκτορας πρέπει να διερευνήσει το περιβάλλον και να συμπεράνει από τις ανταμοιβές που λαμβάνει (Sutton, 1998).

Οι βασικοί συντελεστές μιας διαδικασίας ενισχυτικής μάθησης περιλαμβάνουν:

1. **Πράκτορας (Agent)**: Αυτή είναι η οντότητα (π.χ., ρομπότ, αλγόριθμος) που προσπαθεί να μάθει από την αλληλεπίδρασή του με το περιβάλλον.
2. **Περιβάλλον (Environment)**: Αναφέρεται σε όλα τα άλλα που δεν ανήκουν στον πράκτορα. Αυτό μπορεί να περιλαμβάνει τόσο φυσικά στοιχεία (π.χ., τοιχώματα, αντικείμενα) όσο και αφηρημένα στοιχεία (π.χ., τιμές των μετοχών, ειδήσεις).
3. **Χώρος Ενεργειών (Action Space)**: Ο χώρος ενεργειών αντιπροσωπεύει το σύνολο όλων των δυνατών ενεργειών που μπορεί να πραγματοποιήσει ο πράκτορας ανάλογα με τις παρατηρήσεις του. Είναι σημαντικό να καθοριστεί η διακριτή ή συνεχής φύση αυτού του χώρου και το εύρος ή το σύνολο των έγκυρων ενεργειών. Στο πρόβλημα της Διαχείρισης Χαρτοφυλακίου στην ουσία ο πράκτορας πρέπει να αποφασίζει τον τρόπο που θα κατανείμει το διαθέσιμο κεφάλαιο στις μετοχές του χαρτοφυλακίου. Οι διαθέσιμες ενέργειες στο γενικό πρόβλημα επενδυτικής απόφασης είναι η αγορά, η πώληση καθως τίποτα από τα δύο. Έτσι, σε κάθε στιγμή απόφασης ο πράκτορας θα πρέπει να αποφασίσει τα βάρη w_i για τις n μετοχές του χαρτοφυλακίου, όπου το βάρος $|w_i|$ καθορίζει το ποσοστό του ενεργητικού που επενδύεται στην i μετοχή:

$$-1 \leq w_i \leq 1$$

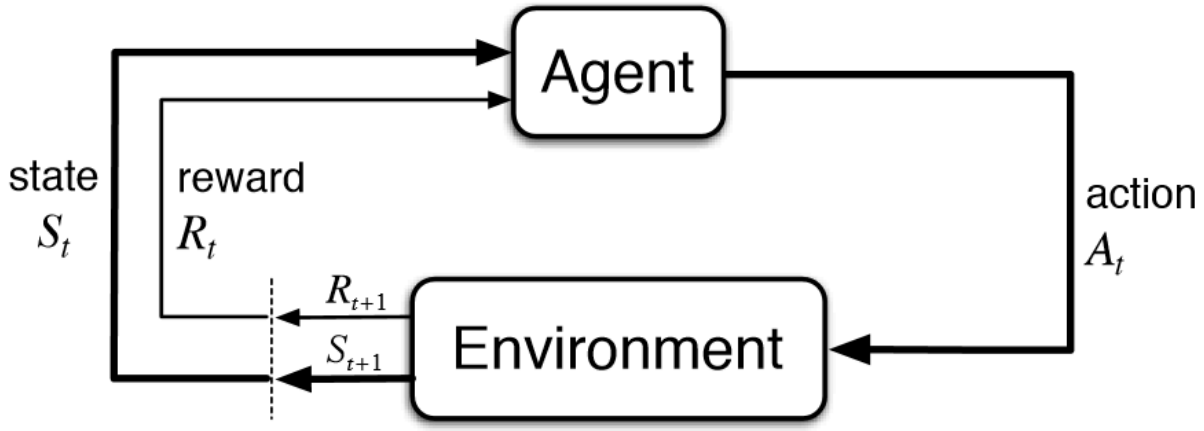
και

$$\sum_{i=1}^{n} |w_i| = 1$$

Αν $w_i = -1$ τότε αυτό σημαίνει ότι ο πράκτορας αποφάσισε να χρησιμοποιήσει όλο το διαθέσιμο ενεργητικό για να πουλήσει τη μετοχή i. Αντίστοιχα, αν $w_i = 1$ τότε αυτό σημαίνει ότι ο πράκτορας αποφάσισε να χρησιμοποιήσει όλο το διαθέσιμο ενεργητικό για να αγοράσει τη μετοχή i. Η γενική αυτή περιγραφή του χώρου ενεργειών του προβλήματος Διαχείρισης Χαρτοφυλακίου, όπως εύκολα γίνεται αντιληπτό δημιουργεί έναν συνεχή Χώρο Ενεργειών ο οποίος καθιστά το πρόβλημα αρκετά πολύπλοκο. Υπάρχει η δυνατότητα να απλοποιηθεί μειώνοντας τις διαθέσιμες ενέργειες, συγκεκριμένα για την υπολογιστική απλοποίηση του προβλήματος, στα παρακάτω πειράματα γίνεται η παραδοχή πως οι μετοχές θα έχουν πάντα το ίδιο «βάρος» στο χαρτοφυλάκιο. Αυτό σημαίνει πως σε όσες μετοχές αποφασιστεί η αγορά ή η πώληση τους (BUY, SELL) ο πράκτορας θα ισομοιράσει σε αυτές το διαθέσιμο κεφάλαιο. Δηλαδή εκτός από τους παραπάνω περιορισμούς για τα βάρη w_i της κάθε μετοχής θα ισχύει επιπλέον ότι:

$$|w_i| = |w_j| \text{ για κάθε } i, j \text{ αν } |w_i| \neq 0 \text{ και } |w_j| \neq 0$$

4. **Ενέργεια (Action)**: Είναι η επιλογή που κάνει ο πράκτορας μεταξύ των των διαθέσιμων ενεργειών από τον αντίστοιχο χώρο για να αλληλεπιδράσει με το περιβάλλον του. Ο πράκτορας διαλέγει τις ενέργειες με βάση την πολιτική του.
5. **Χώρος παρατηρήσεων (Observation Space):** Καθορίζει το τι μπορεί να αντιληφθεί ο πράκτορας από το περιβάλλον. Περιλαμβάνει όλες τις σχετικές πληροφορίες που μπορεί ο πράκτορας να χρησιμοποιήσει για να λαμβάνει αποφάσεις. Στο πρόβλημα της Διαχείρισης Χαρτοφυλακίου ο χώρος παρατηρήσεων μπορεί να περιλαμβάνει τις τιμές των μετοχών, τις τιμές άλλων δεικτών της αγοράς όπως τιμές ομολόγων, νέα της αγοράς (τα οποία μπορούν να αξιοποιηθούν χρησιμοποιώντας τεχνικές Επεξεργασίας Φυσικής Γλώσσας) κ.α. Επιπλέον στον χώρο παρατηρήσεων του συγκεκριμένου προβλήματος περιλαμβάνονται και χαρακτηριστικά του ίδιου του χαρτοφυλακίου όπως η σύνθεσή του, το διαθέσιμο ενεργητικό και άλλοι τυχόν περιορισμοί.
6. **Κατάσταση (State)**: Αυτή είναι η πληροφορία που ο πράκτορας διαθέτει ο πράκτορας για το περιβάλλον κατά τη στιγμή απόφασης της ενέργειας. Η κατάσταση μπορεί να ταυτίζεται με τον χώρο παρατηρήσεων ή μπορεί να είναι ένα υποσύνολο αυτού. Η κατάσταση μπορεί να είναι πλήρης (π.χ., ο πράκτορας ξέρει ακριβώς που βρίσκεται) ή μερική (π.χ., ο πράκτορας ξέρει μόνο μέρος του περιβάλλοντος).
7. **Ανταμοιβή (Reward)**: Αυτή είναι μια συμβολική ποσότητα που το περιβάλλον παρέχει στον πράκτορα για κάθε ενέργεια που κάνει. Ο στόχος του πράκτορα είναι να μεγιστοποιήσει την συνολική ανταμοιβή που λαμβάνει.



Σχήμα 3.11 Διαδικασία Ενισχυτικής Μάθησης

Ο πράκτορας, μέσω της αλληλεπίδρασης με το περιβάλλον, επιδιώκει να μάθει μια πολιτική που θα του επιτρέψει να επιλέξει τις καλύτερες ενέργειες σε κάθε κατάσταση με στόχο, την μεγιστοποίηση της αναμενόμενης ανταμοιβής.

3.5.3 Εξερεύνηση έναντι Εκμετάλλευσης

Το πρόβλημα της εξερεύνησης έναντι της εκμετάλλευσης αποτελεί ένα κεντρικό πρόβλημα στον τομέα της Ενισχυτικής Μάθησης, ενσωματώνοντας την ουσία της λήψης αποφάσεων υπό αβεβαιότητα. Στην ουσία, αυτή η διλημματική κατάσταση συνοψίζει την αντίθεση μεταξύ δύο θεμελιωδών στόχων που ένα ευφυές σύστημα πρέπει να διαχειριστεί κατά την αλληλεπίδρασή του με το περιβάλλον: η εξερεύνηση, που συνεπάγεται την αναζήτηση νέων και, πιθανόν, πολύτιμων πληροφοριών, και η εκμετάλλευση, που περιλαμβάνει την αξιοποίηση της υπάρχουσας γνώσης για τη μεγιστοποίηση των άμεσων ανταμοιβών. Η επίτευξη της ιδανικής ισορροπίας μεταξύ αυτών των αντικρουόμενων στόχων είναι κρίσιμη για την επιτυχία των αλγορίθμων Ενισχυτικής Μάθησης, καθώς η λήψη υποβέλτιστων αποφάσεων κατά τη διάρκεια της διαδικασίας μάθησης μπορεί να οδηγήσει σε αργή σύγκλιση ή ακόμη και στην αποτυχία να ανακαλύψει βέλτιστες πολιτικές.

Στα πρώτα στάδια της μάθησης, ένας πράκτορας συνήθως διαθέτει περιορισμένη ή καθόλου γνώση για το περιβάλλον και τη δυναμική του. Για να αντιμετωπίσει αυτό το έλλειμμα γνώσης, η εξερεύνηση γίνεται ιδιαίτερα σημαντική. Ο πράκτορας πρέπει να αναζητήσει ενεργά ανεξερεύνητες καταστάσεις και ενέργειες για να συλλέξει δεδομένα που θα βελτιώσουν την κατανόησή του για το περιβάλλον και θα του επιτρέψουν να λαμβάνει καλύτερες αποφάσεις στο μέλλον. Ωστόσο, η εξερεύνηση ενέχει αναγκαστικά κινδύνους. Η λήψη μη δοκιμασμένων ενεργειών μπορεί να οδηγήσει σε ανεπιθύμητα αποτελέσματα, προκαλώντας πιθανώς τον πράκτορα να υποστεί κυρώσεις που πιθανόν να δυσκολεύουν τη διαδικασία μάθησής του.

Αντίστροφα, η εκμετάλλευση αποσκοπεί στο να αξιοποιήσει ο πράκτορας την υπάρχουσα γνώση του και να εκμεταλλευτεί ενέργειες που έχουν δείξει ότι παρέχουν υψηλές ανταμοιβές. Η υπερβολική εκμετάλλευση όμως, μπορεί να οδηγήσει σε πρόωρη δέσμευση σε μη βέλτιστες ενέργειες ή καταστάσεις, αποτρέποντας τον πράκτορα από το να ανακαλύψει πιο ανταμοιβοφόρες επιλογές που ίσως υπάρχουν πέρα από την τρέχουσα γνώση του.

Υπάρχουν διάφορες στρατηγικές και αλγόριθμοι που αντιμετωπίζουν το πρόβλημα της εξερεύνησης έναντι της εκμετάλλευσης. Μια κοινή προσέγγιση είναι η πολιτική Epsilon-Greedy, που συνδυάζει εξερεύνηση και εκμετάλλευση επιλέγοντας την ενέργεια με την υψηλότερη εκτιμώμενη αξία με πιθανότητα (1 - ε) και επιλέγοντας μια τυχαία ενέργεια με πιθανότητα εψιλον (ε). Δηλαδή, όταν το ε είναι 1 η ενέργεια που επιλέγεται είναι πλήρως τυχαία ενώ όταν το ε είναι 0 ο πράκτορας επιλέγει την ενέργεια που εκτιμά ότι θα έχει την μεγαλύτερη ανταμοιβή. Οι ενδιάμεσες τιμές εξασφαλίζουν ότι ο πράκτορας θα επιλέγει κάποιες ενέργειες τυχαία, και κάποιες ενέργειες σύμφωνα με την πρότερη γνώση του. Η παράμετρος εψιλον ελέγχει τον συμβιβασμό μεταξύ εξερεύνησης και εκμετάλλευσης, και η τιμή της μπορεί να προσαρμοστεί για να προσαρμοστεί η συμπεριφορά του πράκτορα με τον χρόνο. Μια συνήθης πρακτική, είναι η η παράμετρος εψιλον να αρχικοποιείται με 1 και με τη πάροδο του χρόνου να μειώνεται είτε γραμμικά είτε εκθετικά μέχρι να μηδενιστεί. Ο πράκτορας τότε, θα περιοριστεί αποκλειστικά

στην εκμετάλλευση υποθέτοντας ότι έχει μάθει όλες τις απαραίτητες πληροφορίες που του χρειάζονται για να μεγιστοποιήσει την αναμενόμενη ανταμοιβή.

Μια άλλη σημαντική προσέγγιση είναι η μέθοδος Upper Confidence Bound (UCB), που επιδιώκει να ισορροπήσει την εξερεύνηση διαθέτοντας μια εκτίμηση της αβεβαιότητας για την αξία κάθε ενέργειας. Οι ενέργειες με μεγαλύτερη αβεβαιότητα προτιμώνται για να προωθήσουν την εξερεύνηση, ενώ οι ενέργειες με μικρότερη αβεβαιότητα επιλέγονται για εκμετάλλευση. Οι αλγόριθμοι UCB έχουν ευρύτατη εφαρμογή σε προβλήματα τύπου bandit και σενάρια Ενισχυτικής Μάθησης, προσφέροντας ένα αποτελεσματικό μέσο για την αντιμετώπιση του προβλήματος εξερεύνησης-εκμετάλλευσης.

Το Thompson sampling, βασισμένο στην πιθανότητα Bayes, είναι μια άλλη ισχυρή τεχνική που έχει αποκτήσει ευρεία εφαρμογή στην αντιμετώπιση του προβλήματος αυτού. Σε αυτήν την προσέγγιση, ο πράκτορας διατηρεί μια πιθανοτική κατανομή στις πραγματικές αξίες των ενεργειών και επιλέγει δείγματα από αυτήν την κατανομή για να λαμβάνει αποφάσεις. Το Thompson sampling αξιοποιεί την πιθανοτική αβεβαιότητα για να ισορροπήσει φυσικά την εξερεύνηση και την εκμετάλλευση.

3.5.4 Q-Learning

Η Q-Learning είναι ένας αλγόριθμος ενισχυτικής μάθησης που δεν βασίζεται σε μοντέλο και εισήχθη από τον Christopher Watkins το 1989. Ο στόχος της Q-Learning είναι να βρει μια συνάρτηση τιμής δράσης, γνωστή επίσης ως Q-function, η οποία μπορεί να καθοδηγήσει έναν πράκτορα για να επιλέξει την καλύτερη δράση σε κάθε κατάσταση προκειμένου να μεγιστοποιήσει την αναμενόμενη συσσωρευτική αμοιβή.

Η Q-function αντιπροσωπεύει την αναμενόμενη απόδοση (άθροισμα αμοιβών) μιας δράσης δεδομένης μιας κατάστασης, εκφραζόμενη ως Q(s, a), όπου 's' είναι η κατάσταση, και 'a' είναι η δράση.

Ο αλγόριθμος Q-learning λειτουργεί εξερευνώντας το περιβάλλον μέσω δοκιμής και σφάλματος και ενημερώνει τις Q-τιμές με βάση τις αμοιβές που λαμβάνει για τις δράσεις που λαμβάνει. Ο κανόνας ενημέρωσης, γνωστός επίσης ως εξίσωση Bellman, είναι ο εξής:

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$$

Σε αυτήν την εξίσωση:

- 's' είναι η τρέχουσα κατάσταση
- 'a' είναι η δράση που λαμβάνεται
- 'r' είναι η άμεση αμοιβή μετά τη λήψη της δράσης 'a' στην κατάσταση 's'

- 's'' είναι η νέα κατάσταση μετά τη λήψη της δράσης 'a' στην κατάσταση 's'
- 'a'' είναι η δράση που έχει την υψηλότερη Q-τιμή στη νέα κατάσταση 's''
- 'α' είναι το ποσοστό μάθησης ($0 \leq \alpha \leq 1$), που καθορίζει πόση σημασία δίνεται στις νέες πληροφορίες σε σχέση με τις παλιές πληροφορίες
- 'γ' είναι ο παράγοντας έκπτωσης ($0 \leq \gamma \leq 1$), που καθορίζει πόση σημασία δίνεται στις μελλοντικές αμοιβές σε σχέση με τις άμεσες αμοιβές

Αρχικά, το Q-table (ένας πίνακας που αποθηκεύει τις Q-τιμές για κάθε ζεύγος κατάστασης-δράσης) μπορεί να αρχικοποιηθεί αυθαίρετα, και στη συνέχεια ενημερώνεται επαναληπτικά χρησιμοποιώντας την παραπάνω εξίσωση μέχρι να συγκλίνει στην βέλτιστη Q-function, Q*.

Η βέλτιστη Q-function, Q*, υπακούει στην Εξίσωση του Bellman:

$$Q^*(s, a) = r + \gamma \max_{a'} Q^*(s', a')$$

Μόλις έχει μάθει η βέλτιστη Q-function Q*, ο πράκτορας μπορεί να ακολουθήσει την βέλτιστη πολιτική επιλέγοντας την δράση που έχει την υψηλότερη Q-τιμή σε κάθε κατάσταση.

3.5.5 Βαθύ Q-Δίκτυο - Deep Q-Network

Ο αλγόριθμος Deep Q-Network (DQN) αποτελεί μια σημαντική πρόοδο στον τομέα της Ενισχυτικής Μάθησης, ιδιαίτερα όσον αφορά τα προβλήματα που σχετίζονται με υψηλής διάστασης χώρους καταστάσεων και πολύπλοκους χώρους ενεργειών. Αναπτύχθηκε ως μια επέκταση βαθιάς μάθησης του κλασικού αλγορίθμου Q-learning, χρησιμοποιώντας τη δύναμη των βαθιών νευρικών δικτύων για να προσεγγίσει τις τιμές Q, που ποσοτικοποιούν τις αναμενόμενες συσσωρευτικές ανταμοιβές που συνδέονται με τη λήψη συγκεκριμένων ενεργειών σε διάφορες καταστάσεις.

Στον πυρήνα του αλγορίθμου DQN βρίσκεται μια αρχιτεκτονική νευρικού δικτύου που λαμβάνει τις καταστάσεις ως εισόδους και παράγει τις τιμές Q για κάθε πιθανή ενέργεια ως εξόδους. Αυτό το δίκτυο, συχνά αναφερόμενο ως το Q-δίκτυο, εκπαιδεύεται επαναληπτικά για να ελαχιστοποιήσει το σφάλμα χρονικής διαφοράς μεταξύ των προβλεπόμενων τιμών Q.

Για να σταθεροποιήσει περαιτέρω την εκπαίδευση και να βελτιώσει την αποτελεσματικότητα των δεδομένων, το DQN ενσωματώνει την συσσώρευση εμπειριών. Αυτή η τεχνική περιλαμβάνει την αποθήκευση προηγούμενων εμπειριών, που αποτελούνται από τα ζεύγη κατάστασης-ενέργειας-ανταμοιβής-επόμενης κατάστασης, σε έναν αποθηκευτικό χώρο. Κατά τη διάρκεια της εκπαίδευσης, επιλέγονται τυχαία μικρά πακέτα εμπειριών από τον αποθηκευτικό χώρο, διακόπτοντας τις χρονικές συσχετίσεις στα δεδομένα και μειώνοντας τον αντίκτυπο των

ακραίων τιμών. Η συσσώρευση εμπειριών βελτιώνει σημαντικά την ικανότητα του αλγορίθμου να μαθαίνει από προηγούμενες εμπειρίες, καθιστώντας τον πιο αποδοτικό και ανθεκτικό.

Ένα από τα κυριότερα επιτεύγματα που εισήγαγε το DQN είναι η δυνατότητά του να χειρίζεται χώρους καταστάσεων υψηλών διαστάσεων, όπως εικόνες από οθόνες βιντεοπαιχνιδιών. Χρησιμοποιώντας συνελικτικά νευρωνικά δίκτυα (CNNs) ως αρχιτεκτονική για το Q-δίκτυο, το DQN μπορεί να εξάγει αυτόματα χαρακτηριστικά από τα δεδομένα εισόδου, επιτρέποντάς του να μαθαίνει απευθείας από τις αισθητήριες εισόδους. Αυτή η ικανότητα έχει καταστήσει το DQN ιδιαίτερα επιτυχημένο σε εργασίες όπως η παιχνιδοκονσόλα Atari 2600, όπου έχει επιτύχει απόδοση σε ανθρώπινο και υπεράνθρωπο επίπεδο.

Το DQN ενσωματώνει επίσης μια στρατηγική εξερεύνησης-εκμετάλλευσης χρησιμοποιώντας την πολιτική επιλογής epsilon-greedy. Κατά τη διάρκεια της εκπαίδευσης, ο αλγόριθμος επιλέγει την ενέργεια με τη υψηλότερη εκτιμηθείσα τιμή Q με πιθανότητα (1 - ε) και εξερευνά μια τυχαία ενέργεια με πιθανότητα ε. Αυτή η ισορροπία μεταξύ εξερεύνησης και εκμετάλλευσης είναι κρίσιμη για την μάθηση μιας βέλτιστης πολιτικής, καθώς επιτρέπει στον αλγόριθμο να εξερευνά νέες ενέργειες ενώ παράλληλα επικεντρώνεται στις ενέργειες που φαίνεται ότι είναι οι καλύτερες σύμφωνα με την τρέχουσα γνώση του.

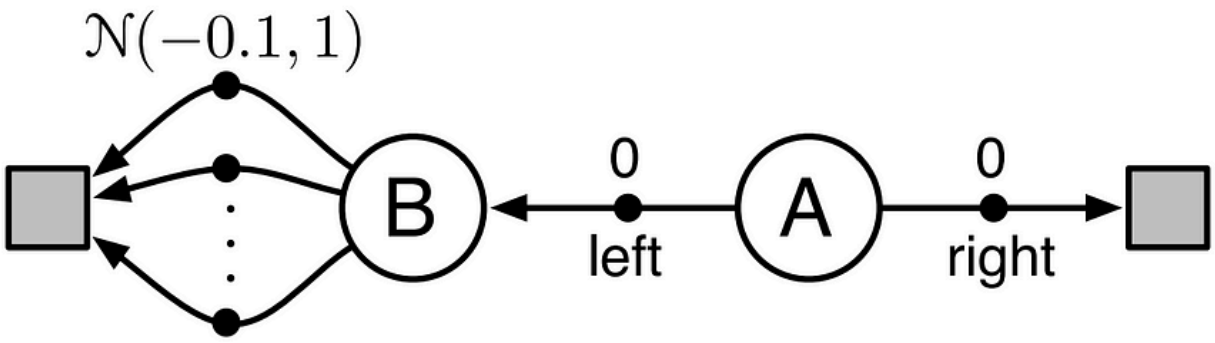
Το DQN έχει ένα βαθύ αντίκτυπο στον τομέα της Ενισχυτικής Μάθησης και έχει εμπνεύσει πολλές παραλλαγές και επεκτάσεις. Η επιτυχία του έχει επεκταθεί πέρα από τα περιβάλλοντα παιχνιδιών σε πραγματικές εφαρμογές, συμπεριλαμβανομένης της ρομποτικής, των αυτόνομων οχημάτων και των οικονομικών προβλέψεων. Παρά τις επιτυχίες του, το DQN αντιμετωπίζει και και προκλήσεις, καθώς μπορεί να είναι ευαίσθητο στις υπερ-παραμέτρους. Παρόλα αυτά, οι θεμελιώδεις συνεισφορές του στη Βαθιά Ενισχυτική Μάθηση έχουν ανοίξει το δρόμο για περαιτέρω προόδους στον τομέα.

3.5.6 Διπλά Βαθιά Q-Δίκτυα - Double Deep Q-Networks (DDQN)

Τα Διπλά Βαθιά Q-Δίκτυα (Double Deep Q-Networks ή DDQN) αποτελούν μια σημαντική πρόοδο στον τομέα της βαθιάς Ενισχυτικής Μάθησης, προσφέροντας μια ισχυρή λύση για να αντιμετωπιστούν ορισμένοι περιορισμοί των παραδοσιακών Βαθιών Q-Δικτύων (Deep Q-Networks ή DQN). Στον τομέα της τεχνητής νοημοσύνης και της Ενισχυτικής Μάθησης, το DDQN έχει αναδειχθεί ως ένας αναγκαίος αλγόριθμος για την εκπαίδευση πρακτόρων στο να λαμβάνουν βέλτιστες αποφάσεις σε πολύπλοκα περιβάλλοντα.

Στην ουσία, το DDQN είναι μια επέκταση του αρχικού αλγορίθμου DQN, που ήταν ιδιαίτερα αποτελεσματικός στο να προσεγγίζει τις τιμές Q των ζευγαριών κατάστασης-ενέργειας χρησιμοποιώντας βαθιά νευρωνικά δίκτυα. Ωστόσο, το DQN έχει ένα γνωστό πρόβλημα: τείνει να υπερεκτιμά τις τιμές των ενεργειών, πράγμα που μπορεί να οδηγήσει σε μη βέλτιστες

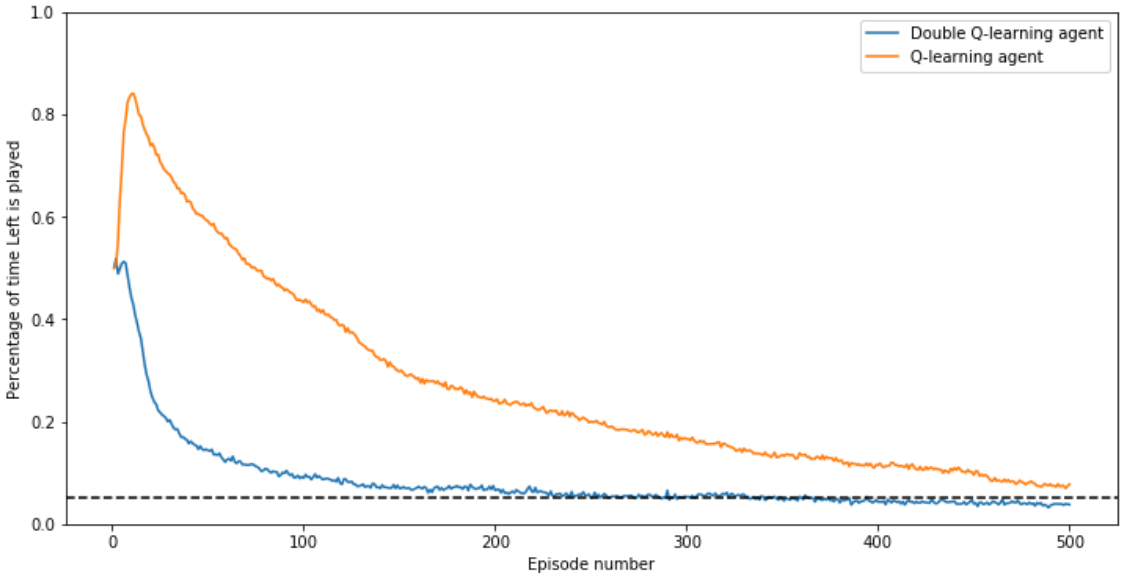
πολιτικές αποφάσεων κατά τη διάρκεια της εκπαίδευσης. Αυτή η υπερεκτίμηση μπορεί να είναι ιδιαίτερα προβληματική κατά την εκπαίδευση σε περιβάλλοντα με μεγάλους χώρους ενεργειών ή υψηλών διαστάσεων χώρους κατάστασης.



Σχήμα 3.12 Παράδειγμα Τάση Μεγιστοποίησης - Διάταξη

Η τάση μεγιστοποίησης (maximization bias) που από κατασκευής έχει το απλό DQN είναι ένα πρόβλημα το οποίο το κάνει να υπερεκτιμά το Q. Για να γίνει πιο κατανοητό το πρόβλημα παρατηρώντας το σχήμα 3.11, ένας πράκτορας εκκινώντας από την Κατάσταση A μπορεί να πάει δεξιά στην τελική κατάσταση και να έχει ανταμοιβή 0 και αριστερά στην Κατάσταση B και από εκεί μέσω έναν μεγάλο αριθμό ενεργειών με ανταμοιβή τυχαία από την N(-0.1, 1), δηλαδή κανονική κατανομή με μέση τιμή -0.1 και διασπορά 1, στη τελική κατάσταση. Η αναμενόμενη Ανταμοιβή θα έπρεπε να καθοδηγεί τον πράκτορα να επιλέγει μόνο το δεξί μονοπάτι αφού η συνολική Ανταμοιβή είναι 0 ενώ του αριστερού -0,1, η διασπορά όμως των ανταμοιβών δημιουργεί αυτό που ονομάζεται maximization bias. Όπως εξηγήθηκε παραπάνω, ο κανόνας υπολογισμού των τιμών Q είναι ο εξής:

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$$



Σχήμα 3.13 Παράδειγμα Τάση Μεγιστοποίησης - Αποτελέσματα

Η επιλογή της ενέργειας με το μεγαλύτερο μέσο reward στο παρελθόν σε συνδυασμό με τη διασπορά, δημιουργεί συχνά μια αισιόδοξη πρόβλεψη για την πιθανή ανταμοιβή που κάνει τον πράκτορα να μαθαίνει πολύ πιο αργά. Στο σχήμα 3.12 φαίνεται η διαφορά μεταξύ του DQN και του DDQN στο παραπάνω απλό πρόβλημα.

Το DDQN αντιμετωπίζει το πρόβλημα της υπερεκτίμησης με μια εξυπνη τροποποίηση του κανόνα ενημέρωσης Q-learning. Αντί να εξαρτώνται αποκλειστικά από ένα μόνο νευρωνικό δίκτυο για την εκτίμηση των τιμών στόχων Q και της τρέχουσας τιμής Q, το DDQN χρησιμοποιεί δύο ξεχωριστά δίκτυα: το δίκτυο στόχου (target network) και το τρέχον δίκτυο (policy network). Αυτή η αποσύνδεση των δικτύων συμβάλλει στην αντιμετώπιση της υπερεκτίμησης που υπάρχει στο DQN.

Το δίκτυο στόχου χρησιμοποιείται για τον υπολογισμό των τιμών στόχου Q, οι οποίες χρησιμοποιούνται στην ενημέρωση Q-learning. Ωστόσο, οι παράμετροι του δικτύου στόχου δεν ενημερώνονται τόσο συχνά όσο αυτές του τρέχοντος δικτύου. Αντί αυτού, οι παράμετροι του δικτύου στόχου ενημερώνονται περιοδικά με μια "ήπια" ενημέρωση (soft update). Αυτή η ήπιαενημέρωση εξασφαλίζει ότι το δίκτυο στόχου θα έχει μια πιο σταθερή και με λιγότερες διακυμάνσεις εκτίμηση των βέλτιστων τιμών Q.

Η κύρια ιδέα πίσω από το DDQN είναι ότι με την αποσύνδεση των δικτύων στόχου και τρέχοντος και τη χρήση του δικτύου στόχου με καθυστέρηση στις ενημερώσεις, γίνεται λιγότερο πιθανό να προκύπτει υπερεκτίμηση των τιμών Q. Αυτό οδηγεί σε πιο σταθερή και αξιόπιστη εκπαίδευση, με αποτέλεσμα πράκτορες που μαθαίνουν καλύτερες πολιτικές, λαμβάνουν λιγότερες υπο-βέλτιστες αποφάσεις και συγκλίνουν πιο αποτελεσματικά.

3.5.7 Deep Deterministic Policy Gradient (DDPG)

Ο αλγόριθμος Deep Deterministic Policy Gradient (DDPG) αποτελεί ένα σημαντικό ορόσημο στον τομέα της Ενισχυτικής Μάθησης, ιδιαίτερα όσον αφορά την αντιμετώπιση ενός από τα πιο δύσκολα κομμάτια αυτού του πεδίου - την αντιμετώπιση καταστάσεων υψηλών διαστάσεων σε συνεχείς χώρους ενεργειών. Ο DDPG αντιπροσωπεύει μια ισχυρή και αποτελεσματική προσέγγιση που συνδυάζει τα πλεονεκτήματα των βαθιών νευρικών δικτύων και των πολιτικών γενετικών κανόνων, καθιστώντας τον μια ανθεκτική επιλογή για εργασίες όπου οι παραδοσιακοί αλγόριθμοι Ενισχυτικής Μάθησης αποτυγχάνουν λόγω των μεγάλων διαστάσεων (Curse of dimensionality).

Στον πυρήνα της, η αρχιτεκτονική του DDPG περιλαμβάνει δύο βασικά νευρωνικά δίκτυα: τον "Δράστη" (Actor) και τον Κριτή (Critic). Ο Παράγοντας διαδραματίζει καθοριστικό ρόλο στην προσέγγιση της βέλτιστης πολιτικής. Απεικονίζει απευθείας τις καταστάσεις σε συνεχείς ενέργειες, επιτρέποντας στον πράκτορα να δρα με ακρίβεια στο χώρο των συνεχών ενεργειών. Αυτή η ικανότητα είναι ιδιαίτερα χρήσιμη σε πραγματικά περιβάλλοντα, όπως η ρομποτική και ο αυτόνομος έλεγχος, όπου οι ενέργειες υπάρχουν σε συνεχείς χώρους, επιτρέποντας μεγαλύτερο έλεγχο και πιο ομαλές αλληλεπιδράσεις με το περιβάλλον.

Αντίθετα, το δίκτυο Επικριτή αξιολογεί την ποιότητα των ενεργειών που επιλέγονται από τον Actor, εκτιμώντας την αναμενόμενη συσσωρευτική ανταμοιβή. Μέσω της εκτίμησης των Q-τιμών για ζεύγη κατάστασης-ενέργειας, ο Επικριτής παρέχει ένα κρίσιμο σήμα ανατροφοδότησης στον Παράγοντα, βοηθώντας στην βελτίωση της πολιτικής. Αυτή η διάκριση μεταξύ των πολιτικών και των κριτικών δικτύων βελτιώνει τόσο την απόδοση όσο και την ευστάθεια της διαδικασίας μάθησης, καθώς τα δίκτυα μπορούν να εκπαιδευτούν ανεξάρτητα και παράλληλα.

Ο DDPG χρησιμοποιεί δύο κύριες τεχνικές για τη σταθεροποίηση και την επιτάχυνση της διαδικασίας μάθησης - τη συσσώρευση εμπειριών και τα δίκτυα στόχων. Η συσσώρευση εμπειριών περιλαμβάνει την αποθήκευση παλαιότερων εμπειριών σε έναν αποθηκευτικό χώρο και τη δειγματοληψία μικρών πακέτων κατά την εκπαίδευση. Αυτή η προσέγγιση περιορίζει τις χρονικές συσχετίσεις στα δεδομένα, μειώνει την επίδραση των ακραίων τιμών και βελτιώνει σημαντικά την αποτελεσματικότητα των δεδομένων, με αποτέλεσμα την πιο σταθερή μάθηση. Τα δίκτυα στόχων, τόσο για τον Παράγοντα όσο και για τον Επικριτή, μειώνουν μέσω του περιοδικού κανόνα ενημέρωσης τους τον κίνδυνο απόκλισης κατά τη διάρκεια της εκπαίδευσης, καθώς και τη διακύμανση στις εκτιμήσεις των Q-τιμών εξασφαλίζοντας έτσι ομαλότερη σύγκλιση.

Ένα σημαντικό σημείο του DDPG είναι η εξάρτησή του από μια πολιτική κανόνων. Αντίθετα με πολλούς άλλους αλγόριθμους Ενισχυτικής Μάθησης που χρησιμοποιούν στοχαστικές πολιτικές, ο DDPG χρησιμοποιεί μια πολιτική κανόνων, απλοποιώντας έτσι τη διαδικασία της

εξερεύνησης. Η εξερεύνηση στον DDPG επιτυγχάνεται προσθέτοντας θόρυβο στην έξοδο του Δράστη κατά τη διάρκεια της εκπαίδευσης. Αυτός ο θόρυβος επιτρέπει στον πράκτορα να εξερευνήσει διάφορες ενέργειες, ενώ ταυτόχρονα διατηρεί τα οφέλη μιας πολιτικής κανόνων κατά την εκμετάλλευση. Αυτή η προσέγγιση προσφέρει έναν καλύτερο συμβιβασμό μεταξύ εξερεύνησης και εκμετάλλευσης, καθώς η πολιτική κανόνων μπορεί να παρέχει σταθερές και αξιόπιστες ενέργειες, ενώ ο θόρυβος εξερεύνησης προσθέτει την απαραίτητη τυχαιότητα.

Ο DDPG έχει επιδείξει σημαντική επιτυχία στην επίλυση ευρείας γκάμας προκλητικών εργασιών. Οι εφαρμογές του εκτείνονται από τον ρομποτικό έλεγχο και την αυτόνομη πλοήγηση έως τα πολύπλοκα παιχνίδια.

3.5.8 Trust Region Policy Optimization (TRPO)

Ο TRPO είναι ένας κορυφαίος και πολύπλοκος αλγόριθμος Ενισχυτικής Μάθησης που έχει αποκτήσει μεγάλη δημοφιλία στον τομέα της τεχνητής νοημοσύνης και της ρομποτικής. Αναπτύχθηκε από τον John Schulman το 2015 και δίνει λύση σε αρκετές προκλήσεις που αντιμετωπίζουν οι παραδοσιακοί μέθοδοι βελτίωσης πολιτικής, όπως η αστάθεια και η αργή σύγκλιση. Στην ουσία, ο TRPO επικεντρώνεται στη βελτιστοποίηση των συναρτήσεων πολιτικής για πολύπλοκες εργασίες, καθιστώντας τον ένα απαραίτητο εργαλείο για την εκπαίδευση πρακτόρων σε διάφορους τομείς, από την αυτόνομη ρομποτική έως τα βιντεοπαιχνίδια.

Μία από τις θεμελιώδεις καινοτομίες του TRPO είναι η χρήση περιοχών εμπιστοσύνης, οι οποίες παρέχουν κάποιους περιορισμούς στο μέγεθος των ενημερώσεων της πολιτικής. Αυτός ο περιορισμός βοηθά να εξασφαλιστεί ότι η νέα βελτιωμένη πολιτική παραμένει κοντά στην αρχική πολιτική, αποτρέποντας καταστροφικές αλλαγές που θα μπορούσαν να διαταράξουν τη μάθηση. Με τον περιορισμό της αλλαγής της πολιτικής εντός μιας περιοχής εμπιστοσύνης, ο TRPO αποσκοπεί στο να βρει μια ισορροπία μεταξύ εξερεύνησης και εκμετάλλευσης, καταλήγοντας τελικά σε πιο σταθερή και αξιόπιστη εκπαίδευση.

Η έννοια της περιοχής εμπιστοσύνης στον TRPO υλοποιείται μέσω μαθηματικών περιορισμών στην ενημέρωση της πολιτικής. Συγκεκριμένα, ο TRPO προσπαθεί να μεγιστοποιήσει την προσδοκώμενη συσσωρευτική ανταμοιβή, υπό τον περιορισμό όμως, ότι η νέα πολιτική πρέπει να παραμένει εντός μια συγκεκριμένης απόστασης, η οποία μετριέται χρησιμοποιώντας μετρικές όπως η απόκλιση Kullback-Leibler (KL), από την αρχική πολιτική. Αυτός ο περιορισμός εξασφαλίζει ότι οι αλλαγές στην πολιτική είναι αρκετά μικρές ώστε ο αλγόριθμος να συγκλίνει και παράλληλα να αποτραπεί ο πράκτορας από τη λήψη υπερβολικά ριψοκίνδυνες ενεργειες.

Ένα από τα βασικά πλεονεκτήματα του TRPO είναι ότι παρέχει θεωρητικές εγγυήσεις για τη βελτίωση της πολιτικής. Αυτό σημαίνει ότι, αντίθετα με ορισμένες άλλες μεθόδους Ενισχυτικής

Μάθησης, ο TRPO εξασφαλίζει ότι η νέα πολιτική είναι τουλάχιστον τόσο καλή όσο και η παλιά πολιτική, ακόμα και όταν οι ενημερώσεις πραγματοποιούνται με περιορισμένο τρόπο εντός της περιοχής εμπιστοσύνης. Αυτή η θεωρητική εγγύηση καθιστά τον TRPO μια καλή επιλογή σε περιπτώσεις όπου η ασφάλεια και η αξιοπιστία είναι προτεραιότητες, όπως στην αυτόνομη οδήγηση ή τον ρομποτικό έλεγχο.

Παρά τη θεωρητική του κομψότητα, ο TRPO συνοδεύεται από ορισμένες υπολογιστικές προκλήσεις. Ο αλγόριθμος περιλαμβάνει την επίλυση ενός προβλήματος βελτιστοποίησης υπό περιορισμούς, που μπορεί να είναι υπολογιστικά δαπανηρό, ιδίως σε υψηλές διαστάσεις του χώρου των ενεργειών. Για να το αντιμετωπίσει αυτό, ο TRPO αξιοποιεί προηγμένες τεχνικές βελτιστοποίησης, όπως η Μέθοδος Συζυγών Κλίσεων, για να βρει αποδοτικά την ενημέρωση της πολιτικής που μεγιστοποιεί τον στόχο διατηρώντας τον περιορισμό της περιοχής εμπιστοσύνης.

Ένα άλλο σημαντικό στοιχείο του TRPO είναι η δυνατότητά του να λειτουργεί τόσο με διακριτούς όσο και με συνεχείς χώρους ενεργειών, καθιστώντας τον ευέλικτο για μια ευρεία γκάμα εφαρμογών. Το χαρακτηριστικό του αυτό τον καθιστά ιδιαίτερα χρήσιμο σε περιπτώσεις όπου οι ενέργειες δεν είναι διακριτές, αλλά αντιστοιχούν σε συνεχείς παραμέτρους, όπως ο έλεγχος ταχύτητας και κατεύθυνσης ενός ρομποτικού βραχίονα.

3.5.9 Proximal Policy Optimization (PPO)

Ο Proximal Policy Optimization (PPO), είναι ένα αλγόριθμος Ενισχυτικής Μάθησης που έχει κερδίσει σημαντική προσοχή και δημοτικότητα στην κοινότητα της βαθιάς Ενισχυτικής Μάθησης. Σχεδιάστηκε για να αντιμετωπίσει ορισμένες από τις προκλήσεις και τους περιορισμούς των προηγούμενων μεθόδων βελτιστοποίησης πολιτικής, όπως ο Trust Region Policy Optimization (TRPO) και οι αρχικές μέθοδοι βελτίωσης πολιτικής.

Ο PPO κατατάσσεται ως αλγόριθμος βελτιστοποίησης πολιτικής, πράγμα που σημαίνει ότι επικεντρώνεται στη βελτίωση της πολιτικής ενός πράκτορα για να μεγιστοποιήσει την αναμενόμενη συσσωρευτική ανταμοιβή σε ένα συγκεκριμένο περιβάλλον. Αντίθετα με τις μεθόδους που βασίζονται στην αξιολόγηση της αξίας διάφορων ενεργειών ή καταστάσεων, οι αλγόριθμοι βελτιστοποίησης πολιτικής, όπως το PPO, βελτιστοποιούν απευθείας την πολιτική του πράκτορα.

Μία από τις κεντρικές ιδέες πίσω από τον PPO είναι η έννοια της "κοντινής" βελτιστοποίησης. Αυτό σημαίνει ότι το PPO περιορίζει το μέγεθος των ενημερώσεων της πολιτικής σε κάθε επανάληψη για να εξασφαλίσει τη σταθερότητα κατά τη διάρκεια της εκπαίδευσης. Αυτό το πετυχαίνει εισάγοντας ένα περιορισμό στην ενημέρωση της πολιτικής, προστατεύοντας την από το να απομακρύνεται υπερβολικά από την προηγούμενη πολιτική όπως ακριβώς και ο TRPO. Αυτός ο περιορισμός βοηθά να αποφεύγονται ενημερώσεις της πολιτικής που μπορεί να

οδηγήσουν σε καταστροφική απόδοση, ένα πρόβλημα που αντιμετωπίζουν ορισμένοι προηγούμενοι αλγόριθμοι. Η καινοτομία που εισάγει ο PPO είναι ότι χρησιμοποιεί μια εξυπνη αντικειμενική συνάρτηση που συνδυάζει δύο βασικά στοιχεία: μια περικομμένη αντικειμενική συνάρτηση αντικατάστασης (clipped surrogate objective) και έναν όρο ρύθμισης της εντροπίας.

Η περικομμένη αντικειμενική συνάρτηση αντικατάστασης προτρέπει την πολιτική να κινείται προς την κατεύθυνση που αυξάνει την πιθανότητα οι ενέργειες να έχουν υψηλότερες αποδόσεις, μειώνοντας ταυτόχρονα την πιθανότητα επιλογής ενεργειών με χαμηλότερες αποδόσεις. Αυτό το στοιχείο εξασφαλίζει ότι η ενημέρωση της πολιτικής δεν αποκλίνει υπερβολικά από την τρέχουσα πολιτική.

Ο όρος ρύθμισης της εντροπίας, από την άλλη πλευρά, προτρέπει την εξερεύνηση, αυξάνοντας την εντροπία της πολιτικής. Αυτό εμποδίζει την πολιτική από το να γίνει υπερβολικά προκαθορισμένη, προωθώντας την εξερεύνηση διαφόρων ενεργειών και βελτιώνοντας τη δυνατότητα του πράκτορα να μαθαίνει σε περίπλοκα περιβάλλοντα. Η ισορροπία μεταξύ εξερεύνησης και εκμετάλλευσης είναι κρίσιμο στοιχείο της Ενισχυτικής Μάθησης, και ο PPO το επιτυγχάνει αποτελεσματικά.

Η συμβατότητα του PPO με διακριτούς και συνεχείς χώρους ενεργειών ενισχύει την ευελιξία του. Μπορεί να εφαρμοστεί σε μια ευρεία γκάμα προβλημάτων Ενισχυτικής Μάθησης, από απλά παιχνίδια έως πολύπλοκα προβλήματα ελέγχου ρομπότ. Αυτή η προσαρμοστικότητα έχει καταστήσει τον PPO μια πολυ δημοφιλή επιλογή.

3.6 Μηχανική Μάθηση στην βελτιστοποίηση χαρτοφυλακίου

3.6.1 Επιβλεπόμενη Μηχανική Μάθηση

Η Μηχανική Μάθηση παρέχει ισχυρά εργαλεία για την αντιμετώπιση της πολυδιάστατης φύσης της βελτιστοποίησης χαρτοφυλακίου. Το μεγάλο πλεονέκτημα της Μηχανικής Μάθησης βρίσκεται στην ικανότητά της να αξιοποιεί τεράστιες ποσότητες δεδομένων και να αναγνωρίζει πρότυπα που ενδέχεται να παραβλέπονται από τις παραδοσιακές μεθόδους, ή ακόμα και από πολύπειρους αναλυτές.

Μία από τις κοινές εφαρμογές της μηχανικής μάθησης στη βελτιστοποίηση χαρτοφυλακίου είναι η πρόβλεψη των τιμών των περιουσιακών στοιχείων (όπως μετοχές) ή των αποδόσεών τους. Αλγόριθμοι εποπτευόμενης μάθησης όπως η γραμμική παλινδρόμηση, οι μηχανές διανυσμάτων υποστήριξης (SVM), και τα μοντέλα δέντρων αποφάσεων μπορούν να εκπαιδευτούν για την πρόβλεψη μελλοντικών αποδόσεων με βάση τα παρελθοντικά και τωρινά δεδομένα της αγοράς (Huang et al., 2005).

Η γραμμική παλινδρόμηση, για παράδειγμα, βρίσκει μια γραμμική σχέση μεταξύ των χαρακτηριστικών εισόδου (όπως παρελθόντες τιμές, όγκος συναλλαγών, οικονομικοί δείκτες κ.λπ.) και της εξόδου (όπως η μελλοντική τιμή). Αυτή η μέθοδος υποθέτει ότι υπάρχει μια γραμμική σχέση μεταξύ της εισόδου και της εξόδου, και προσπαθεί να βρει την βέλτιστες παραμέτρους της γραμμικής σχέσης αυτής που ελαχιστοποιούν την απόσταση μεταξύ των προβλεπόμενων και των πραγματικών τιμών.

Οι Μηχανές Διανυσμάτων Υποστήριξης (SVM), από την άλλη πλευρά, είναι ικανές να χειριστούν μη γραμμικές σχέσεις μεταξύ μεταβλητών. Λειτουργούν μεταφέροντας τα δεδομένα εισόδου (input) σε έναν υψηλότερο διάστασης χώρο όπου επιχειρούν να βρουν ένα υπερεπίπεδο που διαχωρίζει καλύτερα τις διάφορες κατηγορίες δεδομένων (όπως επικερδείς εναντίον μη επικερδών συναλλαγών). Αυτή η τεχνική χρησιμοποιείται συχνά σε σενάρια όπου η σχέση μεταξύ των μεταβλητών εισόδου και εξόδου είναι μη γραμμική, όπως τα δεδομένα της αγοράς και η μελλοντική τιμή του περιουσιακού στοιχείου (Cao et al., 2003).

Τα δέντρα αποφάσεων είναι ένα άλλο ισχυρό εργαλείο μηχανικής μάθησης. Κατασκευάζουν ένα μοντέλο αποφάσεων βάσει πραγματικών παρελθοντικών τιμών των δεδομένων. Οι αποφάσεις διαχωρίζονται σε δομή δέντρου μέχρι να επιτευχθεί μια πρόβλεψη αποτελέσματος. Στο πλαίσιο της βελτιστοποίησης χαρτοφυλακίου, τα δέντρα αποφάσεων θα μπορούσαν να χρησιμοποιηθούν για την λήψη αποφάσεων σχετικά με την αγορά, την πώληση ή την κράτηση συγκεκριμένων αποθεμάτων (Kumar and Thenmozhi, 2016).

Εντούτοις, όπως ισχύει για κάθε τεχνική, οι μέθοδοι μηχανικής μάθησης δεν είναι απαλλαγμένες από αδυναμίες. Αρχικά, υπάρχει κίνδυνος υπερεκπαίδευσης, όπου το μοντέλο προσαρμόζεται υπερβολικά καλά στα δεδομένα εκπαίδευσης και επομένως επιδεικνύει χαμηλή απόδοση σε νέα, δεδομένα.

Επιπλέον, τα μοντέλα μηχανικής μάθησης συχνά απαιτούν εκτενή προεπεξεργασία των δεδομένων για την εκπαίδευση και την επικύρωση των μοντέλων. Αυτό μπορεί να συμπεριλαμβάνει την εξαγωγή χαρακτηριστικών, την αντιμετώπιση των τιμών που απουσιάζουν από τα δεδομένα και την κανονικοποίηση των δεδομένων. Αυτές οι διαδικασίες μπορεί να είναι χρονοβόρες, και ενδέχεται να μην είναι πρακτικές σε ένα περιβάλλον υψηλής συχνότητας όπου οι αποφάσεις πρέπει να ληφθούν σε κλάσματα δευτερολέπτου.

Τέλος, τα μοντέλα μηχανικής μάθησης ενδέχεται να είναι «μαύρα κουτιά», με λίγη ή καμία διαφάνεια στο πώς λαμβάνουν τις προβλέψεις τους. Αυτό μπορεί να είναι πρόβλημα σε περιπτώσεις όπου η ερμηνεία του μοντέλου είναι σημαντική όπως σε αποφάσεις επενδυτικού ενδιαφέροντος.

3.6.2 Βαθιά Μάθηση στην βελτιστοποίηση χαρτοφυλακίου

Η Βαθιά Μάθηση (Deep Learning) είναι υποσύνολο της μηχανικής μάθησης που μιμείται τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου στην επεξεργασία δεδομένων και τη δημιουργία μοτίβων για λήψη αποφάσεων. Έχει σχεδιαστεί για να μαθαίνει αυτόματα και προσαρμοστικά από τις αναπαραστάσεις των δεδομένων, και έχει χρησιμοποιηθεί με επιτυχία σε τομείς όπου η ανακάλυψη των καλύτερων χαρακτηριστικών είναι δύσκολη, όπως η αναγνώριση εικόνων, η αναγνώριση ομιλίας, η επεξεργασία φυσικής γλώσσας και η βιοπληροφορική (LeCun, Bengio, & Hinton, 2015).

Η βαθιά μάθημα έχει τη δυνατότητα να ξεπεράσει τις παραδοσιακές μεθόδους Μηχανικής Μάθησης (Machine Learning) στην βελτιστοποίηση χαρτοφυλακίου για αρκετούς λόγους:

1. Μη γραμμικότητα και πολυπλοκότητα: Τα μοντέλα Βαθιάς Μάθησης, ειδικά τα νευρωνικά δίκτυα με πολλά κρυφά επίπεδα, μπορούν να συλλάβουν πολύπλοκες και μη γραμμικές σχέσεις στα χρηματοοικονομικά δεδομένα. Η βελτιστοποίηση χαρτοφυλακίου συχνά εμπλέκει περίπλοκες αλληλεπιδράσεις μεταξύ των αποδόσεων των περιουσιακών στοιχείων, που ενδεχομένως να μην κατανοούνται επαρκώς από μοντέλα που χρησιμοποιούνται συνήθως στην παραδοσιακή μηχανική μάθηση.

2. Εξαγωγή χαρακτηριστικών: Τα μοντέλα Βαθιάς Μάθησης έχουν τη δυνατότητα να εξάγουν αυτόματα χαρακτηριστικά από τα ακατέργαστα χρηματοοικονομικά δεδομένα, όπως οι τιμές των μετοχών, οι όγκοι συναλλαγών, οι οικονομικοί δείκτες και πολλά άλλα. Αυτή η ικανότητα εξαγωγής χαρακτηριστικών μπορεί να οδηγήσει σε βελτιωμένα αποτελέσματα βελτιστοποίησης χαρτοφυλακίου, καθώς μειώνει την ανάγκη για χειροκίνητη μηχανική εξαγωγή χαρακτηριστικών.

3. Αναπαράσταση δεδομένων: Τα μοντέλα Βαθιάς Μάθησης μπορούν να χειριστούν διάφορους τύπους δεδομένων, συμπεριλαμβανομένων των χρονοσειρών και των δεδομένων χωρίς δομή όπως η φυσική γλώσσα, που είναι ουσιώδη για τη λήψη ενημερωμένων αποφάσεων επενδυτικής στρατηγικής. Τα παραδοσιακά μοντέλα μηχανικής μάθησης ενδέχεται να αντιμετωπίζουν δυσκολίες στον αποτελεσματικό συνδυασμό τέτοιων ποικίλων πηγών δεδομένων.

4. Ευελιξία: Τα μοντέλα Βαθιάς Μάθησης είναι αρκετά ευέλικτα και μπορούν να προσαρμοστούν σε διάφορα χαρακτηριστικά και περιορισμούς του εκάστοτε χαρτοφυλακίου. Η δυνατότητα σχεδιασμού διαφορετικών αρχιτεκτονικών τα καθιστά πιο προσαρμόσιμα από τα παραδοσιακά μοντέλα μηχανικής μάθησης, καθώς γίνεται πλέον δυνατός ο σχεδιασμών μοντέλων τα οποία υπηρετούν συγκεκριμένες επενδυτικές στρατηγικές.

Στο πλαίσιο της βελτιστοποίησης χαρτοφυλακίου, οι τεχνικές βαθιάς μάθησης μπορούν να παράσχουν καινοτόμες λύσεις, κυρίως μέσω της χρήσης διαφόρων αρχιτεκτονικών νευρωνικών δικτύων. Τα νευρωνικά δίκτυα αξιοποιούν μια σειρά αλγορίθμων που αναζητούν να αναγνωρίσουν τις σχέσεις σε ένα σύνολο δεδομένων μέσα από μια διαδικασία που μιμείται τη λειτουργία του ανθρώπινου εγκεφάλου. Έτσι, ένα νευρωνικό δίκτυο μπορεί να προσαρμόζει τις εσωτερικές του παραμέτρους βάσει των πληροφοριών που επεξεργάζεται, μαθαίνοντας από το περιβάλλον και βελτιώνοντας την προβλεπτική του ακρίβεια με την πάροδο του χρόνου.

Οι αυτοενσωματωτές, μία τεχνική βαθιάς μάθησης, έχουν χρησιμοποιηθεί για τη μείωση της διάστασης των δεδομένων εισόδων στο πρόβλημα της διαχείρισης χαρτοφυλακίου, καθιστώντας το πιο εύκολα υπολογίσιμο (Dixon, 2016). Με τη μείωση του αριθμού των μεταβλητών που λαμβάνονται υπόψη στη βελτιστοποίηση χαρτοφυλακίου μειώνεται η υπερεκπαίδευση καθώς και, αυξάνεται η δυνατότητα γενίκευσης του μοντέλου το οποίο το καθιστά καλύτερο σε νέα δεδομένα.

Τα Αναδρομικά Νευρωνικά Δίκτυα (ΑΝΔ - RNNs), ιδιαίτερα τα Δίκτυα Μακροχρόνιας Μνήμης (ΔΜΜ - LSTM), είναι άλλη μια μέθοδος βαθιάς μάθησης που έχει εφαρμοστεί στη βελτιστοποίηση χαρτοφυλακίου. Τα ΔΜΜ είναι ιδιαίτερα ισχυρά σε προβλήματα πρόβλεψης χρονοσειράς γιατί μπορούν να αποθηκεύουν παλαιότερες σειριακές πληροφορίες. Αυτό είναι σημαντικό στη βελτιστοποίηση χαρτοφυλακίου, όπου οι χρονικές ακολουθίες και η σειρά των παλαιότερων τιμών της αγοράς και άλλων οικονομικών δεικτών μπορεί να περιέχουν χρήσιμες προβλεπτικές πληροφορίες. Ο Heaton (2017) εφάρμοσε ΔΜΜ για να προβλέψει μελλοντικές επιστροφές χαρτοφυλακίου, προσφέροντας μια νέα προσέγγιση στο πρόβλημα της βελτιστοποίησης χαρτοφυλακίου.

Τα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ - CNNs), γνωστά για την εξαιρετική τους απόδοση σε προβλήματα ταξινόμησης εικόνων, μπορούν επίσης να εφαρμοστούν σε προβλήματα βελτιστοποίησης χαρτοφυλακίων. Μετατρέποντας τις οικονομικές χρονοσειρές σε δομές πίνακα παρόμοιες με εικόνες, μια μελέτη από τους Sezer και Ozbayoglu (2018) παρουσίασε πώς τα ΣΝΔ μπορούν να χρησιμοποιηθούν για την πρόβλεψη της μελλοντικής κίνησης της αγοράς μετοχών, ανοίγοντας τον δρόμο σε νέες προσεγγίσεις στη βελτιστοποίηση χαρτοφυλακίου με βάση τα ΣΝΔ.

Η ανώτερη προβλεπτική απόδοση των τεχνικών Βαθιάς Μάθησης στις χρηματοοικονομικές αγορές τις καθιστά ένα ισχυρό εργαλείο για τη βελτιστοποίηση χαρτοφυλακίου. Ωστόσο, πρέπει να τονιστεί ότι οι σύνθετες αρχιτεκτονικές τους, η πληθώρα παραμέτρων που πρέπει να εκπαιδευτούν και η ανάγκη για μεγάλη ποσότητα δεδομένων, καθιστούν τη βαθιά μάθηση μια τεχνολογία που απαιτεί σημαντική υπολογιστική ισχύ και εξειδικευμένες γνώσεις για την εφαρμογή της.

3.6.3 Ενισχυτική Μάθηση στη βελτιστοποίηση χαρτοφυλακίου

Η Ενισχυτική Μάθηση (Reinforcement Learning - RL), μια ακόμα υποκατηγορία της μηχανικής μάθησης, εισάγει μια διαδικασία μάθησης όπου ένας πράκτορας (agent) αποκτά γνώση πραγματοποιώντας ενέργειες (actions) στο περιβάλλον του (environment) και μαθαίνοντας από τα λάθη και τις επιτυχίες του μέσω μιας συνάρτησης ανταμοιβής (reward function). Πρόκειται για μια προσέγγιση που είναι βασισμένη στα πολύ βασικά χαρακτηριστικά του πώς μαθαίνουν οι άνθρωποι και τα ζώα: μέσω της προσπάθειας μεγιστοποίησης των ανταμοιβών και ελαχιστοποίησης των ποινών.

Σε ένα μοντέλο Ενισχυτικής Μάθησης, η διαδικασία μάθησης είναι επαναληπτική. Ο πράκτορας θα διενεργήσει μια ενέργεια εντός του δεδομένου περιβάλλοντος, και η ποιότητα αυτής της ενέργειας θα μεταδοθεί πίσω στον πράκτορα με τη μορφή μιας ανταμοιβής. Ο κύριος στόχος του πράκτορα σε ένα περιβάλλον Ενισχυτικής Μάθησης είναι να μάθει μια βέλτιστη πολιτική, δηλαδή μια στρατηγική, που καθοδηγεί την επιλογή της ενέργειας σε κάθε κατάσταση (state) για να μεγιστοποιηθεί η συνολική ανταμοιβή κατά τη διάρκεια ενός καθορισμένου χρονικού διαστήματος (episode). Η υλοποίηση τέτοιων συστημάτων μάθησης έχει δει μια πληθώρα επιτυχημένων εφαρμογών σε περιοχές όπως η ρομποτική, τα συστήματα ελέγχου, τα παιχνίδια και πιο πρόσφατα, οι χρηματοοικονομικές επενδύσεις και η διαχείριση χαρτοφυλακίων (Moody and Saffell, 2001). Αναλυτική παρουσίαση του τρόπου λειτουργίας των μοντέλων Ενισχυτικής Μάθησης αλλά και των διαφόρων τεχνικών της θα γίνει στο επόμενο κεφάλαιο.

Η Διαχείριση Χαρτοφυλακίου είναι μια διαδικασία λήψης αποφάσεων σε χρονική σειρά και επιδιώκει να διανείμει τα διαχειριζόμενο κεφάλαιο με τρόπο που μεγιστοποιεί το κέρδος περιορίζοντας όμως το ρίσκο, το οποίο ανάλογα με το επενδυτικό προφιλ μπορεί να είναι περισσότερο ή λιγότερο ανεκτό. Αυτή η διαδικασία απόφασης σε σειρά συμβαδίζει ιδιαίτερα καλά με το πλαίσιο της Ενισχυτικής Μάθησης, καθιστώντας το ένα υποσχόμενο εργαλείο για τη βελτιστοποίηση των χρηματοοικονομικών χαρτοφυλακίων.

Οι πρώιμες εφαρμογές της Ενισχυτικής Μάθησης στη διαχείριση χαρτοφυλακίου χρησιμοποιούσαν σχετικά απλούς αλγόριθμους και επικεντρώνονταν κυρίως στην αγορά μετοχών. Για παράδειγμα, οι Moody και Saffell (2001) διερεύνησαν την άμεση ενίσχυση (direct reinforcement) για να παρακάμψουν την ανάγκη πρόβλεψης των επιστροφών και των τιμών, δείχνοντας σημαντικές βελτιώσεις σε σχέση με τη στρατηγική αγοράς και κρατήματος στο δείκτη S&P 500. Ο παράγοντάς τους έμαθε να κάνει εμπόριο με την άμεση μεγιστοποίηση μιας ανταμοιβής που προσαρμόζεται στον κίνδυνο. Σχεδιάστηκε για να λαμβάνει διακριτές αποφάσεις αγοράς ή πώλησης σε κάθε χρονικό βήμα ενώ η ανταμοιβή ήταν η επιστροφή του χαρτοφυλακίου.

Η ραγδαία πρόοδος στις τεχνικές μηχανικής μάθησης, ιδιαίτερα η εισαγωγή των μεθόδων βαθιάς μάθησης, έχει επηρεάσει σημαντικά τις εφαρμογές της ενισχυτικής μάθησης στη διαχείριση

χρηματοοικονομικών χαρτοφυλακίων. Η Βαθιά Ενισχυτική Μάθηση (Deep Reinforcement Learning - DRL) συνδυάζει την ισχύ της βαθιάς μάθησης για εξαγωγή χαρακτηριστικών και την ικανότητα της ενισχυτικής μάθησης να μαθαίνει από την εμπειρία, προσφέροντας τη δυνατότητα δημιουργίας συστημάτων που μπορούν να εκπαιδευτούν να εκτελούν σύνθετες εργασίες επίσης στον χώρο των επενδύσεων και της διαχείρισης χαρτοφυλακίου.

Μια σημαντική πρόοδος έγινε από την ομάδα του Deng και συνεργατών του (2016), οι οποίοι προσάρμοσαν τον αλγόριθμο της Q-Learning (ειδική μέθοδος ενισχυτικής μάθησης) στη βελτιστοποίηση χαρτοφυλακίου. Το μοντέλο τους, γνωστό ως Financial Market Q-Learning (FMQL), εκμεταλλευόταν τις αλλαγές της τάσης των χρηματοοικονομικών αγορών για την απόφαση συναλλαγής. Η εφαρμογή της Βαθιάς Ενισχυτικής Μάθησης σε προβλήματα όπως αυτά οδήγησε στην εκμάθηση πολιτικών που υπερβαίνουν τις προσεγγίσεις που βασίζονται σε παραδοσιακές μεθόδους.

Οι πρόσφατες προσεγγίσεις έχουν χρησιμοποιήσει πιο προηγμένες τεχνικές, όπως τα Διπλά Βαθιά Q-Δίκτυα (Double Deep Q-Networks - DDQN), για την εξερεύνηση των χρηματοοικονομικών αγορών και τη βελτιστοποίηση των χαρτοφυλακίων. Σε συνδυασμό με τεχνικές όπως το βαθύ νευρωνικό δίκτυο, το συνεχές σχέδιο ελέγχου και την αντιμετώπιση της αστάθειας των χρηματοοικονομικών αγορών, οι εφαρμογές της Βαθιάς Ενισχυτικής Μάθησης στη βελτιστοποίηση των χρηματοοικονομικών χαρτοφυλακίων αναμένεται να συνεχίσουν να επεκτείνονται και να βελτιώνονται στο μέλλον.

Κεφάλαιο 4. Προτεινόμενη Μεθοδολογία

4.1 Εισαγωγή

Σκοπός της παρούσας διπλωματικής εργασίας, είναι η υλοποίηση και αξιολόγηση διαφορετικών Πρακτόρων Βαθιάς Ενισχυτικής Μάθησης για το πρόβλημα της διαχείρισης χαρτοφυλακίου. Στο πλαίσιο αυτό, αναπτύσσονται πράκτορες με διαφορετικές παραμέτρους, αρχιτεκτονικές και εισόδους ώστε να γίνει η επιλογή αυτού που επιτυχάνει την καλύτερη απόδοση. Η παρακάτω μεθοδολογική προσέγγιση ακολουθήθηκε ώστε να δημιουργηθεί ένα σύστημα όπου θα επιτρέπει στον πράκτορα, δυναμικά, για οποιοδήποτε χαρτοφυλάκιο, να συλλέγει τα δεδομένα, να τα επεξεργάζεται και στη συνέχεια να «εκπαιδεύεται» πάνω σε αυτά.

4.2 Ορισμος Προβλήματος

Πρώτο βήμα, όπως είναι φυσικό, αποτελεί ο σαφής ορισμός του προβλήματος το οποίο θα καθορίσει και το σχεδιασμό των υπολοίπων βημάτων της μεθοδολογίας. Το συγκεκριμένο βήμα είναι απαραίτητο για τους παρακάτω λόγους. Πρώτον, η επιλογή των δεδομένων. τα οποία θα χρησιμοποιηθούν στα πειράματα, καθορίζεται σημαντικά από τον ορισμό του προβλήματος καθώς και από τα συμπεράσματα τα οποία επιθυμούμε να εξάγουμε. Δεύτερον, η προγραμματιστική υλοποίηση του συστήματος εκτέλεσης των πειραμάτων θα πρέπει να είναι σε θέση να υποστηρίξει όλες τις λειτουργίες οι οποίες θα χρειαστούν κατά την πραγματοποίηση των πειραμάτων. Τέλος, ο πλήρης και σαφής ορισμός του προβλήματος αποτελεί κομβικό βήμα για την τελική διαδικασία αξιολόγησης τον πειραμάτων και βελτίωσης αυτών μέχρι να επιτευχθεί το επιθυμητό αποτέλεσμα.

Η παρούσα εργασία, επικεντρώνεται στη μελέτη και εξαγωγή συμπερασμάτων των παρακάτω προβλημάτων. Αρχικά, επιλέγοντας ένα συγκεκριμένο χαρτοφυλάκιο μετοχών μελετώνται μέσω διαδοχικών πειραμάτων με διαφορετικές υπερ-παραμέτρους, το πως επηρεάζει η κάθε υπερ-παράμετρος την εκπαίδευση του μοντέλου και την αποτελεσματικότητά του τόσο στα δεδομένα εκπαίδευσης όσο και σε άγνωστα δεδομένα. Από τα πειράματα αυτά σκοπός είναι η επιλογή των βέλτιστων υπερ-παραμέτρων του πράκτορα. Στη συνέχεια, χρησιμοποιώντας τις πληροφορίες για τις υπερπαραμέτρους που συλλέχθηκαν προηγουμένως, υλοποιούνται 4 διαφορετικοί πράκτορες και διεξάγονται πειράματα ώστε να βρεθεί η απόδοσή τους τόσο στα δεδομένα εκπαίδευσης όσο και σε άγνωστα δεδομένα (test set).

Συγκεκριμένα για τους πρώτους τρεις πράκτορες επιλέγονται διαφορετικά δεδομένα εισόδου και μελετάται το πως αυτά, και οι διάφοροι συνδυασμοί τους, επηρεάζουν την αποτελεσματικότητα του πράκτορα (agent). Ο πρώτος πράκτορας χρησιμοποιεί ως δεδομένα εισόδου τις τιμές Ανοίγματος, Κλεισίματος, Χαμηλή και Υψηλή. Ο δεύτερος πράκτορας χρησιμοποιεί

αποκλειστικά τις ποσοστιαίες μεταβολές των τιμών κλεισίματος ενώ ο τρίτος πράκτορας χρησιμοποεί πέρα από τις ποσοστιαίες μεταβολές των τιμών και διάφορους στατιστικούς δείκτες (statistical indicators). Ο τελευταίος πράκτορας αξιοποιεί συνελικτικά δίκτυα και συγκεκριμένα την αρχιτεκτονική που προτείνει ο Liang και οι συνεργάτες του (Liang, 2018) με σκοπό να ελεγχθεί, αν χρησιμοποιώντας συνελικτικές τεχνικές αυξάνεται η προβλεπτική ισχύ. Όλοι οι πράκτορες πέρα από παρελθοντικές τιμές των μετοχών λαμβάνουν και σαν είσοδο την τωρινή κατάσταση του χαρτοφυλακίου. Λεπτομέρειες θα αναλυθούν στο επόμενο κεφάλαιο.

Μετά το πέρας όλων των παραπάνω πειραμάτων, συγκρίνονται τα αποτελέσματα των διαφορετικών μοντέλων ώστε να εκτιμηθεί η αποδοτικότητα των μεθόδων που χρησιμοποιήθηκαν.

4.3 Παρουσίαση Μεθοδολογίας

Αρχικά, αφού πρόκειται για πρόβλημα Ενισχυτικής Μάθησης, κομβικής σημασίας για την δημιουργία του συστήματος είναι πρώτα ο καθορισμός του περιβάλλοντος (environment) του πράκτορα, το οποίο εμπεριέχει μεταξύ άλλων που θα αναλυθούν παρακάτω τον ορισμό της κατάστασης (state), όπου δρα ως είσοδος στο σύστημα απόφασης, τον ορισμό της συνάρτησης μετάβασης (transition function) καθώς και η επιλογή των διαθέσιμων ενεργειών (actions).

Στη συνέχεια έχοντας καθορίσει σαφώς το περιβάλλον σειρά έχει η επιλογή των μετοχών οι οποίες θα καταρτίσουν το χαρτοφυλάκιο των πειραμάτων. Αφού γίνει η επιλογη, ακολουθεί η συλλογή των δεδομένων (data collection), ο «καθαρισμός» των δεδομένων (data cleansing), η μελέτη και κατανόηση των δεδομένων (data exploration) και η επεξεργασία τους (data processing).

Το επόμενο βήμα, στον σχεδιασμό ενός συστήματος ενισχυτικής μάθησης, είναι η επιλογή του αλγορίθμου (policy) και η εύρεση μέσω πειραμάτων των διαφόρων παραμέτρων αυτού.

Τέλος, απαραίτητο για τον σχεδιασμό ενός συστήματος διαχείρισης χαρτοφυλακίου είναι ο καθορισμός των μετρικών αξιολόγησης των αποτελεσμάτων των πειραμάτων, η χρήση τους για την εξαγωγή συμπερασμάτων και η οπτικοποίηση τους.

4.4 Καθορισμός περιβάλλοντος

Η καθορισμός του περιβάλλοντος ενός συστήματος Ενισχυτικής Μάθησης (reinforcement learning - RL) αποτελεί έναν κρίσιμο βήμα στον σχεδιασμό και την υλοποίηση του. Το περιβάλλον λειτουργεί ως το πλαίσιο στο οποίο λειτουργεί ο πράκτορας, παρέχοντας στον πράκτορα τις απαραίτητες πληροφορίες και προκλήσεις για την εκμάθηση και τη λήψη

αποφάσεων. Παρακάτω παρουσιάζονται οι παράμετροι ενισχυτικής μάθησης που επιλέχθηκαν στα πλαίσια αυτής της μεθοδολογίας:

1. **Χώρος Παρατηρήσεων (Observation Space)**: Στα πειράματα που θα ακολουθήσουν, ο χώρος παρατηρήσεων αποτελείται από τα δεδομένα που παρουσιάστηκαν παραπάνω. Συγκεκριμένα, είναι οι ημερήσιες τιμές Ανοίγματος, Κλεισίματος, Χαμηλή και Υψηλή τιμή των μετοχών του χαρτοφυλακίου, ο ημερήσιος όγκος συναλλαγών καθώς και επεξεργασμένες μορφές των παραπάνω όπως χρηματιστηριακοί δείκτες π.χ. RSI, MACD κ.α. Επιπλέον, επειδή ο πράκτορας επιχειρεί με διαδοχικές ενέργειες να βελτιστοποιήσει ένα χαρτοφυλάκιο, στο χώρο παρατηρήσεων ανήκει η παρούσα κατάσταση του χαρτοφυλακίου δηλαδή το πλήθος των μετοχών που έχουμε αγοράσει ή πουλήσει για κάθε μία από τις μετοχές.

2. **Κατάσταση (State)**: Η κατάσταση του συστήματος καθορίζεται σημαντικά από τον χώρο παρατηρήσεων. Ο ρόλος της Κατάστασης είναι να επιλεχθούν μεταξύ των διαθέσιμων πληροφοριών (Χώρος Παρατηρήσεων) αυτές οι οποίες είναι χρήσιμες και μπορούν να αξιοποιηθούν για την επίλυσή του. Στα πειράματα παρακάτω γίνεται μελέτη της βέλτιστης κατάστασης εισόδου του συστήματος ώστε η πολυπλοκότητα να μην είναι ούτε υπερβολικά μεγάλη αλλά ούτε να υπάρχει έλλειψη πληροφορίας που θα οδηγήσει σε υποβέλτιστη εκπαίδευση. Δοκιμάζονται λοιπόν διαφορετικά υποσύνολα του χώρου παρατηρήσεων ως είσοδος-κατάσταση του συστήματος.

3. **Χώρος Ενεργειών (Action Space)**: Όπως εξηγήθηκε και στη θεωρία στο πρόβλημα της διαχείρισης χαρτοφυλακίου ο χώρος ενεργειών στο πρόβλημα της διαχείρισης χαρτοφυλακίου μπορεί να είναι είτε συνεχής είτε διακριτός. Στα πειράματα που θα ακολουθήσουν αποφασίστηκε πως ο χώρος ενεργειών θα είναι διακριτός και ο πράκτορας (agent) θα επιλέγει για τη κάθε μετοχή, μία από τις 3 διαθέσιμες ενέργειες (BUY, HOLD και SELL). Έτσι τελικά, ο χώρος ενεργειών που δημιουργει αυτός ο σχεδιασμός για n μετοχές είναι 3^n.

4. **Δομή του Επεισοδίου (Episode Structure)**: Καθορίζει πώς δομούνται τα επεισόδια μέσα στο περιβάλλον. Η Ενισχυτική Μάθηση συνήθως προϋποθέτει την εκμάθηση μέσα από αλληλεπιδράσεις πάνω σε πολλά επεισόδια, και η κατανόηση της δομής του επεισοδίου μπορεί να επηρεάσει τον σχεδιασμό του αλγορίθμου μάθησης. Στα πλαίσια της διαχείρισης χαρτοφυλακίου είναι σημαντικό να αποφασιστεί η επιθυμητή συχνότητα λήψης της επενδυτικής απόφασης για να καθοριστεί η δομή του επεισοδίου.

   Στα πειράματα παρακάτω αποφασίστηκε πως οι πράκτορες θα αποφασίζουν μία φορά την ημέρα στο τέλος της ημέρας. Επιπλέον, γίνεται η παραδοχή πως όλες οι ενέργειες

αγοράς και πώλησης γίνονται με την Προσαρμοσμένη Τιμή Κλεισίματος της μετοχής.

5. **Δυναμική Μετάβασης Κατάστασης (State Transition Dynamics)**: Η κατανόηση του πώς το περιβάλλον ανταποκρίνεται στις ενέργειες του πράκτορα είναι κομβική. Γι' αυτό καθορίζονται σαφώς οι κανόνες που διέπουν τον τρόπο με τον οποίο η κατάσταση του περιβάλλοντος αλλάζει όταν ο πράκτορας πραγματοποιεί συγκεκριμένες ενέργειες.

   Συγκεκριμένα, στα πειράματα που θα ακολουθήσουν ο πράκτορας επιλέγει κάθε μέρα πως θα κατανέμει το διαθέσιμο κεφάλαιο στις μετοχές του χαρτοφυλακίου. Οι ενέργειες επηρεάζουν την κατάσταση του συστήματος μεταβάλλοντας τη σύνθεση του χαρτοφυλακίου, παρ' όλα αυτά το εξωτερικό περιβάλλον, δηλαδή οι μελλοντικές τιμές των μετοχών, δεν επηρεάζεται (υποθέτουμε ότι ο όγκος συναλλαγών είναι μικρός, και συνεπώς δεν επηρεάζουν την αγορά).

6. **Συνάρτηση Ανταμοιβής (Reward Function)**: Η συνάρτηση ανταμοιβής είναι ένα κρίσιμο στοιχείο του ορισμού του περιβάλλοντος. Αξιολογεί την άμεση ανατροφοδότηση που λαμβάνει ο πράκτορας μετά από κάθε ενέργεια. Στο πρόβλημα που εξετάζεται εδώ ο πράκτορας, αποσκοπεί μέσω των επιλογών του να μεγιστοποιήσει το κέρδος, συνεπώς η συνάρτηση ανταμοιβής θα πρέπει να περιγράφει τον τρόπο που κάθε επιλογή του πράκτορα επηρεάζει την κεφαλαιοποίηση του χαρτοφυλακίου. Όπως είναι λογικό λοιπόν, αφού ο πράκτορας αποφασίζει καθημερινά τη σύσταση του χαρτοφυλακίου, η ανταμοιβή του θα είναι η ημερήσια απόδοση της ενέργειας που έκανε. Για να γίνει η προσομοίωση όσο το δυνατόν πιο ρεαλιστική, σημαντικό είναι να συμπεριληφθεί στη συνάρτηση ανταμοιβής και τα κόστη συναλλαγών. Έτσι η συνάρτηση ανταμοιβής παίρνει την εξής μορφή:

$$\text{Ανταμοιβή} = \frac{\text{Κεφαλαιοποίηση Χαρτοφυλακίου} + \text{Ημερήσιο Κέρδος}}{\text{Κεφαλαιοποίηση Χαρτοφυλακίου}}$$

$$\text{Ημερήσιο Κέρδος} = \sum_{i=1}^{n} (\text{Ημερήσιο κέρδος μετοχής}_i - \text{Κόστη συναλλαγής μετοχής}_i)$$

$$\text{Ημερήσιο κέρδος μετοχής} = (\text{Τιμή Επόμενης Ημέρας} - \text{Τιμή Ημέρας}) * \text{Νέα Ποσότητα}$$

$$\text{ΚΣΜ} = \text{Τιμή Ημέρας} * \text{Παράγοντας Κόστους} * |\text{Νέα Ποσότητα} - \text{Τωρινή Ποσότητα}|$$

οπου ο παράγοντας κόστους καθορίζεται παρακάτω και n ο αριθμός των μετοχών

7. **Συνθήκες Τερματισμού (Termination Conditions)**: Καθορίζουν τις συνθήκες υπό τις οποίες ένα επεισόδιο μέσα στο περιβάλλον τερματίζεται. Αυτό μπορεί να περιλαμβάνει τον επίτευξη ενός στόχου, την υπέρβαση ενός χρονικού ορίου ή την αντιμετώπιση συγκεκριμένων γεγονότων. Στα πειράματα που θα ακολουθήσουν σαν επεισόδιο ορίζουμε την προσπέλαση όλων των δεδομένων εκπαίδευσης. Επιπλέον, μια ακόμα συνθήκη τερματισμού που προστίθεται στα πειράματα είναι η απώλεια όλου του διαθέσιμου ενεργητικού.

8. **Στρατηγική Εξερεύνησης (Exploration Strategy)**: Στα πειράματα που θα αναλυθούν παρακάτω, εφαρμόζεται η στρατηγική εξερεύνησης epsilon-greedy οπου το epsilon αρχικοποιείται σε 1, δηλαδή ο πράκτορας στην αρχή κάνει μόνο τυχαίες ενέργειες, και σε κάθε επεισόδιο μειώνεται εκθετικά μέχρι να εξαφανιστεί σχεδόν πλήρως η τυχαιότητα. Η ταχύτητα μείωσης του epsilon είναι κάτι το οποίο θα μελετηθεί στη συνέχεια.

4.5 Επιλογή μετοχών και διαστήματος πρόβλεψης

Μετά τον ορισμό του προβλήματος, σειρά έχει η επιλογή των μετοχών στις οποίες θα γίνουν τα πειράματα. Η επιλογή των μετοχών, όπως είναι λογικό, προηγείται της συλλογής των δεδομένων ώστε να γίνει διαλογή μεταξύ των διαφόρων πηγών δεδομένων ανάλογα με το αν διαθέτουν όλα όσα είναι απαραίτητα για τα πειράματα.

Επιπλέον, πριν τη συλλογή των δεδομένων είναι σημαντικό να αποφασιστεί το διάστημα εκπαίδευσης και πρόβλεψης ώστε να αξιολογηθεί και αυτό στην επιλογή της πηγής των δεδομένων. Το χρονικό διάστημα εκπαίδευσης διαφέρει ανάλογα με τη πολυπλοκότητα του προβλήματος και την επιλογή του μοντέλου πρόβλεψης. Πολύπλοκα μοντέλα όπως νευρωνικά δίκτυα απαιτούν μεγάλο αριθμό δεδομένων ώστε να μπορέσουν να βρουν συσχετίσεις και να γενικεύσουν, ενώ απλότερα μοντέλα απαιτούν πολύ λιγότερα δεδομένα.

Στα πειράματα που θα ακολουθήσουν παρακάτω, χρησιμοποιήθηκαν μετοχές από τον ίδιο τομέα της οικονομίας και συγκεκριμένα αυτόν της τεχνολογίας. Η πρακτική αυτή είναι ιδιαίτερα σύνηθης στον χώρο των επενδύσεων αφού συχνά επενδυτές επιλέγουν μετοχές ή δείκτες ενός τομέα της οικονομίας και τοποθετούν το κεφάλαιό τους σε αυτούς.

4.6 Συλλογή Δεδομένων

Έχοντας, πλέον, επιλέξει τις μετοχές που θα χρησιμοποιηθούν στα πειράματα, σειρά έχει η αναζήτηση της πηγής των δεδομένων αλλά και η δημιουργία ενός εύκολου και συστηματικού τρόπου συλλογής τους.

Στα πλαίσια της διαδικασίας αυτής αξιολογούνται διαφορετικές πηγές όπως Yahoo Finance, MarketWatch, Investing.com κ.α. Κριτήρια κατά την επιλογή της πηγής των δεδομένων είναι:

1. Πληρότητα δεδομένων: Όπως είναι λογικό, το πιο σημαντικό κριτήριο για την επιλογή της πηγής των δεδομένων είναι να περιέχει όσο το δυνατόν περισσότερα από τα επιθυμητά δεδομένα. Η χρήση των λιγότερων δυνατών διαφορετικών πηγών είναι πολύ σημαντική καθώς διαφορετικές πηγές μπορεί να χρησιμοποιούν διαφορετική μορφοποίηση, διαφορετικό τρόπο εξαγωγής δεδομένων ακόμη και μικρές διαφορές στα δεδομένα λόγω διαφορετικών παραδοχών. Όλα αυτά επιβραδύνουν και προσθέτουν πολυπλοκότητα στην πειραματική διαδικασία και για αυτό γίνεται προσπάθεια να αποφευχθούν.
2. Τρόπος εξαγωγής των δεδομένων: Είναι επίσης σημαντικό οι πηγές που τελικα θα επιλεχθούν να προσφέρουν στον χρήστη εύκολους τρόπους εξαγωγής των δεδομένων όπως API, τα οποία μπορούν να χρησιμοποιηθούν από προγραμματιστικά εργαλεία. Στη παρούσα εργασία δίνεται αυξημένη βαρύτητα στον παράγοντα αυτό, καθώς ένας από τους στόχους της εργασίας είναι η δημιουργία ενός δυναμικού τρόπου επιλογής χαρτοφυλακίου.

4.7 Καθαρισμός και επεξεργασία δεδομένων

Ο καθαρισμός και η επεξεργασίας των δεδομένων, είναι η διαδικασία αναγνώρισης και διόρθωσης σφαλμάτων, αντιφάσεων και ανακριβειών σε σύνολα δεδομένων ή βάσεις δεδομένων. Ο στόχος του καθαρισμού δεδομένων είναι η βελτίωση της ποιότητας, της ακρίβειας και της αξιοπιστίας των δεδομένων, καθιστώντας τα πιο κατάλληλα για ανάλυση, αναφορά και άλλες δραστηριότητες που βασίζονται σε δεδομένα. Εδώ είναι μερικές κύριες πτυχές του καθαρισμού δεδομένων:

1. **Ανίχνευση Σφαλμάτων:** Ο καθαρισμός δεδομένων ξεκινά ανιχνεύοντας διάφορα είδη σφαλμάτων και ανωμαλιών στα δεδομένα. Αυτά τα σφάλματα μπορεί να περιλαμβάνουν τυπογραφικά σφάλματα, διπλές εγγραφές, λείπουσες τιμές, προβλήματα στη μορφοποίηση και μη αναμενόμενες τιμές. Στην προεπεξεργασία δεδομένων που γίνεται μετά τη συλλογή των δεδομένων για τα πειράματα έγινε έλεγχος όλων των παραπάνω χρησιμοποιώντας προγραμματιστικά εργαλεία.

2. **Κατάργηση Διπλοεγγραφών:** Αφαιρούνται διπλές εγγραφές για να αποφευχθεί η περιττή αναπαραγωγή και να διασφαλιστεί ότι κάθε δεδομένο σημείο είναι μοναδικό.

3. **Διαχείριση Λειψόντων Δεδομένων:** Τα ελλείποντα δεδομένα στα πειράματα που θα ακολουθήσουν θα αντιμετωπιστούν εξαιρώντας τα από τη μελέτη.

4. **Εντοπισμός και Διαχείριση Μη Αναμενόμενων Τιμών (Outliers):** Οι «ακραίες» τιμές είναι δεδομένα τα οποία είναι σημαντικά διαφορετικά από την πλειονότητα από τη πλειοψηφία. Στα πειράματα που θα ακολουθήσουν η μεθοδολογία που ακολουθείται για τις ακραίες τιμές είναι η διασταύρωση της εγκυρότητας αυτών με διαφορετικές πηγές δεδομένων.

4.8 Επιλογή Αλγορίθμου

Με βάση τα χαρακτηριστικά που καθορίστηκαν παραπάνω επιλέχθηκε το Διπλό Βαθύ Q-Δίκτου (Double Deep Q-Network - DDQN) ως η βέλτιστη λύση για το πρόβλημα βελτιστοποίησης χαρτοφυλακίου με διακριτό χώρο ενεργειών. Η επιλογή αυτή έγινε λόγω της αποτελεσματικότητάς του στην αντιμετώπιση αρκετών σημαντικών προκλήσεων που προκύπτουν σε περιβάλλοντα Ενισχυτικής Μάθησης στον χρηματοοικονομικό τομέα.

Καταρχάς, το DDQN, όπως εξηγήθηκε παραπάνω, βασίζεται στην παραδοσιακή αρχιτεκτονική του Βαθούς Q-Δικτύου (Deep Q-Network - DQN), αλλά με την βελτίωση ότι αντιμετωπίζει το πρόβλημα υπερεκτίμησης των τιμών Q. Σε εφαρμογές όπως η βελτιστοποίηση ενός χαρτοφυλακίου με μετοχές, όπου η ακρίβεια είναι κρίσιμη, η υπερεκτίμηση της αξίας των ενεργειών μπορεί να οδηγήσει σε μη βέλτιστες αποφάσεις. Το DDQN αντιμετωπίζει αυτό το πρόβλημα χρησιμοποιώντας δύο ξεχωριστά δίκτυα: ένα για την επιλογή των ενεργειών και ένα άλλο για την αξιολόγηση αυτών των ενεργειών.

Επιπλέον, οι χρηματοοικονομικές αγορές συχνά εμφανίζουν μη σταθερή συμπεριφορά, όπου οι υποκείμενες δυναμικές μπορούν να αλλάξουν ραγδαία. Το DDQN είναι κατάλληλο για τη διαχείριση αυτής της μη σταθερότητας, καθώς χρησιμοποιεί την ανασκόπηση εμπειριών (replay memory) και το Q-Δίκτυο στόχου. Η ανασκόπηση εμπειριών επιτρέπει στον πράκτορα να μαθαίνει από ένα σύνολο προηγούμενων εμπειριών, βοηθώντας τον να προσαρμοστεί στις μεταβαλλόμενες συνθήκες της αγοράς με την πάροδο του χρόνου. Το δίκτυο στόχου, με τις καθυστερημένες ενημερώσεις του, παρέχει έναν πιο σταθερό στόχο για την εκτίμηση της αξίας Q, που είναι ζωτικής σημασίας σε ασταθείς χρηματοοικονομικές συνθήκες.

Ένα ακόμα σημαντικό πλεονέκτημα του DDQN είναι ο αποτελεσματικός του χειρισμός μεγάλων χώρων ενεργειών. Στον παρών σχεδιασμό, όπου οι διαθέσιμες ενέργειες σε κάθε στιγμή απόφασης είναι 27, η διαδικασία της μάθησης και της εξερεύνησης είναι δύσκολη και απαιτεί προσεκτικό σχεδιασμό. Η συμβατότητα του DDQN με διακριτούς χώρους ενεργειών, σε συνδυασμό με τεχνικές όπως η εξερεύνηση epsilon-greedy, επιτρέπει τη συστηματική εξερεύνηση του χώρου ενεργειών χωρίς να γίνεται υπολογιστικά απαγορευτικό.

Συνοψίζοντας, ο αλγόριθμος DDQN πληροί όλα τα χαρακτηριστικά που είναι απαραίτητα για την επίλυση του προβλήματος όπως έχει οριστεί, συνεπώς στα παρακάτω πειράματα θα χρησιμοποιηθεί αυτός για την εξαγωγή συμπερασμάτων.

4.9 Σχεδιασμός και Αξιολόγηση πειραμάτων

Έχοντας ολοκληρώσει όλα τα παραπάνω βήματα, σειρά έχει η διεξαγωγή των πειραμάτων και η αξιολόγησή τους.

Αρχικά, όπως ορίστηκε στο πρόβλημα παραπάνω, θα εκτελεστούν πειράματα τα οποία έχουν σκοπό την εύρεση των βέλτιστων υπερπαραμέτρων Ενισχυτικής Μάθησης. Για το σκοπό αυτό, θα διεξαχθούν πειράματα με διαφορετικές τιμές για τη κάθε παράμετρο ώστε να βρεθούν οι τιμές που μεγιστοποιούν την απόδοση του πράκτορα. Κατά τη διεξαγωγή των πειραμάτων η αρχιτεκτονική του Q-δικτύου παραμένει σταθερή και το μόνο που μεταβάλλεται είναι οι τιμές των υπερπαραμέτρων ώστε τα αποτελέσματα να επηρεάζονται αποκλειστικά από τις αλλαγές των τιμών αυτών.

Οι υπερ-παράμετροι που επιλέχθηκαν ως οι κρισιμότεροι για την εκπαίδευση του πράκτορα και συνεπώς θα βελτιστοποιηθούν μέσα από μια σειρά πειραμάτων είναι:

- Συχνότητα εκπαίδευσης: Έχει εξηγηθεί παραπάνω πως ο πράκτορας, λαμβάνει την απόφαση για την κατανομή του διαθέσιμου κεφαλαίου μία φορά την ημέρα. Αυτό όμως, δεν σημαίνει πως ο πράκτορας θα πρέπει να εκπαιδεύεται κάθε μέρα καθώς οι διαθέσιμες πληροφορίες που έχει, αλλάζουν ελάχιστα. Εισάγουμε στα πειράματά μας τη παράμετρο που καθορίζει τη συχνότητα της εκπαίδευσης ώστε να βρεθεί η τιμή που επιτυγχάνει τη βέλτιστη απόδοση.

- Συχνότητα ενημέρωσης Δικτύου Στόχου (Target Network): Το δίκτυο στόχου όπως έχει αναλυθεί εκτενώς παραπάνω είναι αυτό που υπολογίζει τις τιμές Q και ενημερώνεται περιοδικά ώστε να αποφευχθεί πρόβλημα μεγιστοποίησης (maximization bias) και να γίνει η μάθηση πιο σταθερή. Για το λόγο αυτό, εισάγεται στα πειράματα η παράμετρος που καθορίζει τη συχνότητα ενημέρωσης του Δικτύου Στόχου ώστε να μελετηθεί η επίδραση που έχει στην εκπαίδευση και να βρεθεί η βέλτιστη τιμή του.

- Μείωση παράγοντα epsilon (epsilon greedy): Για τα πειράματα χρησιμοποιείται ο αλγόριθμος DDQN ο οποίος χρησιμοποιεί την τεχνική epsilon greedy για να βρεθεί η ισορροπία μεταξύ εξερεύνησης και εκμετάλλευσης. Ο ρυθμός μείωσης του epsilon επηρεάζει σημαντικά το πόσο ο παράγοντας συνεχίζει να εξερευνεί, επιλέγοντας τυχαίες ενέργειες, και πόσο εκμεταλλεύεται την γνώση που έχει αποκτήσει ήδη. Γρήγορη μείωση

του epsilon μπορεί να οδηγήσει σε γρήγορη σύγκλιση αλλά υπό βέλτιστα αποτελέσματα ενώ πολύ αργή μείωση του μπορεί να οδηγήσει σε πολύ αργή ή καθόλου σύγκλιση.

- Παράγοντας Έκπτωσης (γ): Ο παράγοντας έκπτωσης γ είναι θεμελιώδης στην Ενισχυτική Μάθηση αφού καθορίζει την επίδραση των πιθανών μελλοντικών καταστάσεων στην λήψη της απόφασης. Υψηλότερος παράγοντας γ σημαίνει πως οι μελλοντικές καταστάσεις έχουν πολύ μεγάλη επίδραση στην λήψη της απόφασης ενώ χαμηλότερος παράγοντας γ μικρότερη επίδραση.

Για την αξιολόγηση των παραπάνω πειραμάτων θα χρησιμοποιηθούν διάφορες μετρικές και γραφικές παραστάσεις όπου θα συγκριθούν μεταξύ τους για να αναδειχθεί η βέλτιστη παραμετροποίηση.

4.9.1 Εύρεση βέλτιστης κατάστασης εισόδου

Έχοντας βρει τις βέλτιστες υπερ-παραμέτρους του πράκτορα, σειρά έχει η εύρεση της βέλτιστης κατάστασης εισόδου. Θα εκτελεστούν πειράματα για 3 πράκτορες όπου σαν είσοδος θα χρησιμοποιηθούν διαφορετικά χαρακτηριστικά των μετοχών, σε διαφορετικά χρονικά παράθυρα όπως:

- Τιμή Ανοίγματος
- Χαμηλή τιμή
- Υψηλή τιμή
- Τιμή Κλεισίματος
- Ποσοστιαία απόδοση
- Διάφοροι τεχνικοί δείκτες όπως RSI, MACD κ.α.

Συγκεκριμένα, επιλέγοντας διαφορετικούς συνδυασμούς των παραπάνω χαρακτηριστικών υλοποιούνται οι παρακάτω 3 πράκτορες:

Μοντέλο 1

Ο πρώτος πράκτορας χρησιμοποιεί αποκλειστικά τις τιμές των μετοχών των W προηγούμενων ημερών καθώς και την τρέχουσα σύνθεση του χαρτοφυλακίου. Συγκεκριμένα:

1. Τιμή Ανοίγματος W ημερών
2. Τιμή Κλεισίματος W ημερών
3. Υψηλή Τιμή W ημερών
4. Χαμηλή Τιμή W ημερών
5. Αριθμός μετοχών στη παρούσα σύνθεση του χαρτοφυλακίου, π.χ 1 αν ο πράκτορας έχει αγοράσει 1 μετοχή ή -1 αν έχει πουλήσει μια μετοχή

Μοντέλο 2

Ο δεύτερος πράκτορας, αντικαθιστά στην είσοδο-κατάσταση του τις τιμές των μετοχών με τις ποσοστιαίες μεταβολές τους. Η αλλαγή αυτή γίνεται με σκοπό την διευκόλυνση του πράκτορα να εντοπίσει μοτίβα σε όλο το διάστημα εκπαίδευσης τα οποία μπορεί να ταυτοποιήσει σε διαφορετικά χρονικά παράθυρα ανεξαρτήτως των τιμών των μετοχών. Συγκεκριμένα, η είσοδος του Μοντέλου 2 είναι η εξής:

1. Ημερήσιες ποσοστιαίες μεταβολές τιμής W ημερών
2. Αριθμός μετοχών στη παρούσα σύνθεση του χαρτοφυλακίου, π.χ 1 αν ο πράκτορας έχει αγοράσει 1 μετοχή ή -1 αν έχει πουλήσει μια μετοχή

Μοντέλο 3

Για το μοντέλο 3 θα ακολουθηθεί παρόμοια τακτική με το μοντέλο 2 αλλά θα προστεθούν στην είσοδο-κατάσταση του συστήματος και κάποιοι στατιστικοί δείκτες ώστε να διαπιστωθεί αν αυτοί μπορούν να βελτιώσουν την ικανότητα του μοντέλου να γενικεύει σε άγνωστα δεδομένα. Έτσι, η είσοδος είναι η εξής:

1. Ημερήσιες ποσοστιαίες μεταβολές τιμής W ημερών
2. Δείκτης Σχετικής Δύναμης (RSI)
3. Όγκος Ισορροπίας (OBV)
4. Κινητός Μέσος Όρος Σύγκλισης/Απόκλισης (MACD)
5. Εκθετικός Μέσος Όρος W Ημερών
6. Εκθετικός Μέσος Όρος W Ημερών
7. Αριθμός μετοχών στη παρούσα σύνθεση του χαρτοφυλακίου, π.χ 1 αν ο πράκτορας έχει αγοράσει 1 μετοχή ή -1 αν έχει πουλήσει μια μετοχή

Για τους πράκτορες αυτούς, εκτελείται μια σειρά πειραμάτων χρησιμοποιώντας τις βέλτιστες υπερ παραμέτρους ώστε να διαπιστωθεί η αποδοτικότητά τους.

Τα αποτελέσματα των πειραμάτων αυτών θα αξιολογηθούν έναντι άλλων παραδοσιακών τεχνικών διαχείρισης χαρτοφυλακίου χρησιμοποιώντας κοινές μετρικές για την εξαγωγή συμπερασμάτων.

4.9.2 Σύγκριση αρχιτεκτονικής δικτύου

Μοντέλο 4

Μετά το πέρας των πειραμάτων για τη βελτιστοποίηση της κατάσταση είσοδο σειρά έχει η διεξαγωγή πειραμάτων με σκοπό την αξιολόγηση της αρχιτεκτονικής του δικτύου.

Η σύγκριση γίνεται, διεξάγοντας πειράματα με νέα αρχιτεκτονική και συγκρίνοντας τα αποτελέσματα με τα προηγούμενα μοντέλα. Σε αυτά πειράματα αυτά μας ενδιαφέρει, η απόδοση των μοντέλων τόσο στα δεδομένα εκπαίδευσης όσο και στα άγνωστα δεδομένα test set.

Οι προηγούμενοι πράκτορες χρησιμοποιούσαν απλά δίκτυα perceptron για τον υπολογισμό των Q-τιμών. Στην νέα προσέγγιση, θα διερευνηθεί η αξιοποίηση συνελικτικών δικτύων για τον υπολογισμό των Q-τιμών ακολουθώντας την αρχιτεκτονική που έχει προταθεί από τον Liang και τους συνεργάτες του (Liang, 2018). Συγκεκριμένα, η προτεινόμενη αρχιτεκτονική χρησιμοποιεί ένα συνελικτικό στρώμα εισόδου μεγέθους ίσου με την είσοδο-κατάσταση, ένα κρυφό συνελικτικό στρώμα και ένα στρώμα εξόδου μεγέθους ίσο με τις διαθέσιμες ενέργειες για την επιλογή της βέλτιστης.

Τα συνελικτικά δίκτυα συνήθως χρησιμοποιούνται σε προβλήματα όπου η είσοδος είναι εικόνα καθώς είναι ιδιαίτερα αποδοτικά στο να αναλύουν τέτοιου είδους δεδομένα και να εξάγουν τα χαρακτηριστικά της εικόνας που έχουν το μεγαλύτερο προβλεπτικό ενδιαφέρον. Γνωρίζοντας αυτό το πλεονέκτημα των αρχιτεκτονικών αυτών μετασχηματίζουμε τα δεδομένα μας με έναν τρόπο ώστε η είσοδος να είναι παρόμοια με αυτή μια εικόνας.

Έτσι, η είσοδος-κατάσταση στο σύστημα, ακολουθώντας και πάλι την μεθοδολογία του παραπάνω άρθρου περιέχει την τιμή κλεισίματος, υψηλή και χαμηλή. Συγκεκριμένα:

1. Τιμή Κλεισίματος W ημερών
2. Τιμή Υψηλή W ημερών
3. Τιμή Χαμηλή W ημερών

Από τα παραπάνω γίνεται φανερό πως το μέγεθος της εισόδου είναι:

$$\text{Μέγεθος εισόδου} = n \times W \times f,$$

όπου n ο αριθμός των μετοχών στο χαρτοφυλάκιο, W το μέγεθος του παραθύρου των δεδομένων και f ο αριθμός των χαρακτηριστικών (features)

Άρα η είσοδος που περιγράφεται παραπάνω έχει μέγεθος 3 x 50 x 3, το οποίο θυμίζει μια RGB

Αξιοποιώντας τη γνώση που έχουμε αποκτήσει από τα προηγούμενα πειράματα, χρησιμοποιούμε για την εκπαίδευση τις βέλτιστες υπερ-παραμέτρους και κατασκευάζουμε το Μοντέλο 4 (Συνελικτικά Νευρωνικά Δίκτυα).

Κεφάλαιο 5. Πειραματική διαδικασία

5.1 Εισαγωγή

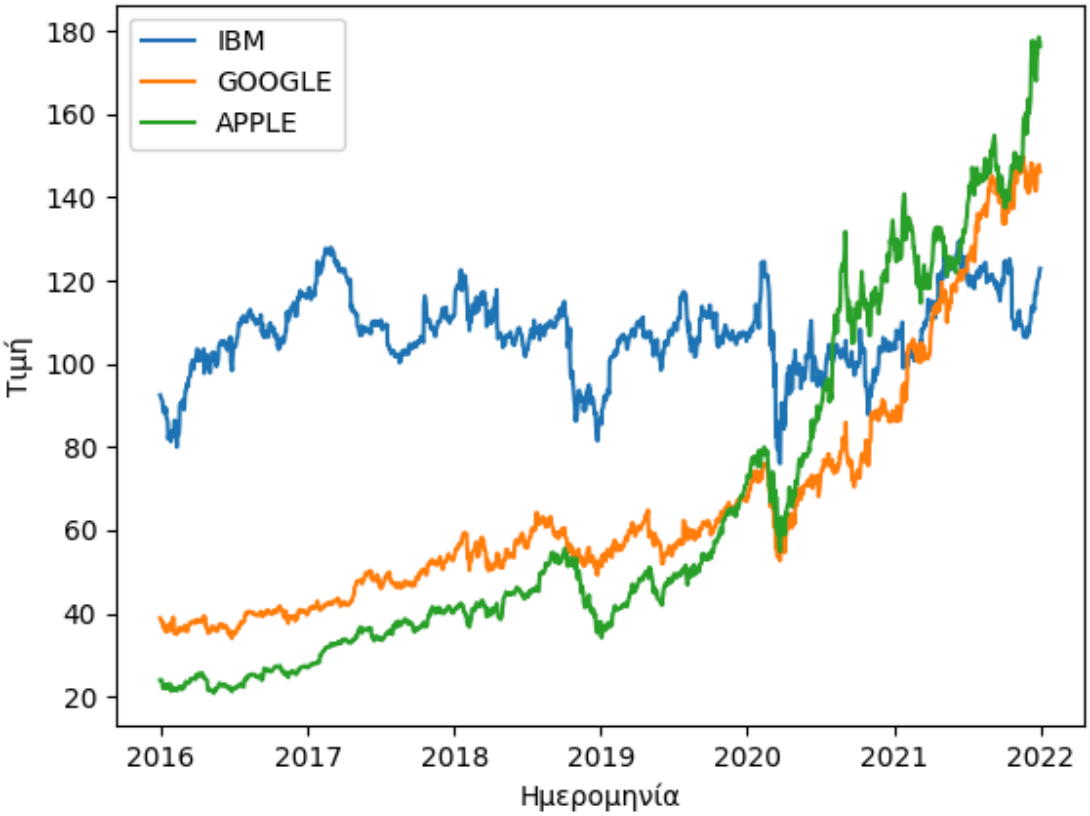
Για την μελέτη του προβλήματος που ορίστηκε παραπάνω, αναπτύχθηκε κώδικας προγραμματισμού ο οποίος επιτελεί όλες τις λειτουργίες που είναι απαραίτητες για την εκτέλεση των πειραμάτων και την εξαγωγή συμπερασμάτων. Η ανάπτυξη ενός προγραμματιστικού πλασίου, το οποίο μπορεί με ταχύτητα και ευκολία να εκτελεί τα πειράματα, επιταχύνει την διαδικασία και προσφέρει μεγάλη ευελιξία στη συλλογή αποτελεσμάτων.

Παρακάτω θα παρουσιαστούν αναλυτικά τα βήματα της εργασίας και θα παρουσιαστούν τα αποτελέσματα των πειραμάτων.

5.2 Επιλογή μετοχών

Αρχικά, αποφασίστηκε πως το χαρτοφυλάκιο που θα μελετηθεί θα αποτελείται από 3 μετοχές λαμβάνοντας υπόψη τόσο τη περιγραφή του προβλήματος, δηλαδή ότι πρόκειται για μια μελέτη πάνω στη Διαχείριση Χαρτοφυλακίου, όσο και την πολυπλοκότητα του. Αποφασίστηκε έτσι, πως για ένα πρόβλημα διαχείρισης χαρτοφυλακίου 2 μετοχές είναι πολύ λίγες καθώς αυτές οι περιπτώσεις έχουν μελετηθεί εκτενώς στη βιβλιογραφία και χρησιμοποιούνται τεχνικές όπως Εμπορίας Ζευγών (Pair Trading) οι οποίες επικεντρώνονται στην μεταξύ αλληλεπίδραση των μετοχών αυτών. Για παραπάνω μετοχές, η πολυπλοκότητα του προβλήματος αυξάνεται εκθετικά με αποτέλεσμα οι απαιτούμενοι πόροι να είναι αδύνατο να διατεθούν. Έτσι, αποφασίστηκε πως ο ιδανικός αριθμός των μετοχών για το χαρτοφυλάκιο που θα μελετηθεί είναι 3 ώστε να επιτυγχάνεται ο σκοπός μελέτης ενός προβλήματος διαχείρισης χαρτοφυλακίου χωρίς όμως η πολυπλοκότητα να αποτελεί εμπόδιο στην διεξαγωγή των πειραμάτων.

Για το βασικό χαρτοφυλάκιο, όπου θα εκτελεστούν τα πρώτα πειράματα βελτιστοποίησης του πράκτορα, επιλέχθηκαν οι μετοχές των Apple (APPL), Google (GOOGL) και IBM (IBM). Οι συγκεκριμένες μετοχές επιλέχθηκαν καθώς είναι μετοχές που έχουν μελετηθεί εκτενώς στη βιβλιογραφια, συνεπώς είναι εύκολο, μελλοντικά, να συγκριθούν τα αποτελέσματα στην παρούσα έρευνα με άλλες παρόμοιες, και επιπλέον εμφανίζουν ενδιαφέροντα χαρακτηριστικά και αντιθέσεις στη διάρκεια των χρόνων, όπως μεγάλη μεταβλητότητα, γεγονός που θα διευκολύνει την εξαγωγή συμπερασμάτων.



Σχήμα 5.1 Τιμές μετοχών τομέα τεχνολογίας

Από το σχήμα 4.1 μπορούμε να δούμε πως τόσο η μετοχή της Google όσο και της Apple εμφανίζουν σημαντική ανοδική τάση σχεδόν στο σύνολο της περιόδου ενώ η μετοχή της IBM δεν φαίνεται να παρουσιάζει κάποια τάση αλλά μεγάλη μεταβλητότητα.

5.3 Επιλογή διαστήματος πρόβλεψης

Όπως αναφέρθηκε παραπάνω, το διάστημα που επιλέχθηκε σαν δεδομένα εκπαίδευσης είναι το διάστημα 01/01/2016 - 31/12/2020 δηλαδή συνολικά 5 χρόνια. Σαν τεστ δεδομένα επιλέχθηκε η περίοδος 01/01/2021 - 31/12/2021, δηλαδή ένας χρόνος.

Σχήμα 5.2 Δείκτης S&P 500, 2016-2021

Το διάστημα που μελετάται, έχει την ιδιομορφία ότι ενώ πρόκειται για μία γενικά καλή περίοδος για την οικονομία, όπως φαίνεται στο Σχήμα 5.1, στις αρχές του 2020 λόγω της πανδημίας του κορονοϊού υπήρξε μεγάλη πτώση. Ενδιαφέρον έχει, λοιπόν, να παρατηρηθεί κατά πόσο το σύστημα διαχείρισης χαρτοφυλακίου που αναπτύσσουμε μπορεί να χειριστεί τέτοιες καταστάσεις.

5.4 Συλλογή δεδομένων

Λαμβάνοντας υπόψη τα κριτήρια που αναλύθηκαν παραπάνω, σαν βέλτιστη επιλογή, μεταξύ των πηγών που αξιολογήθηκαν, κρίθηκε το Yahoo Finance για τους παρακάτω λόγους. Πρωτίστως, το Yahoo Finance περιέχει δεδομένα για όλες τις μετοχές που αναφέρθηκαν παραπάνω. Επιπλέον, η συλλογή των δεδομένων για όλες τις μετοχές γίνεται πολύ γρήγορα μέσω Διεπαφής Προγραμματισμού Εφαρμογών (API) το οποίο είναι εύκολα προσβάσιμο. Για την προγραμματιστική συλλογή των δεδομένων χρησιμοποιήθηκε το πακέτο yfinance της Python και στην συνέχεια αποθηκεύτηκαν σε μορφή csv.

Ενδεικτικά, παρακάτω φαίνεται η περιγραφή των δεδομένων για τη μετοχή της Apple μετά την εξαγωγή τους. Για τη παρουσίαση και επεξεργασία των δεδομένων χρησιμοποιείται η δομή dataframe του πακέτου pandas της python.

	Open	**High**	**Low**	**Close**	**Adj Close**	**Volume**
count	1511	1511	1511	1511	1511	1511
mean	66.46	67.18	65.79	66.52	64.59	127000000

std	41.19	41.69	40.69	41.22	41.30	59592658
min	22.5	22.91	22.36	22.58	20.82	41000000
25%	37.59	38.01	37.29	37.63	35.58	87053200
50%	47.97	48.55	47.76	48.14	46.15	111000000
75%	91.26	92.79	90.73	91.42	89.63	147000000
max	181.12	182.13	178.53	180.33	178.52	533000000

Πίνακας 5.1 Περιγραφή δεδομένων Apple

5.5 Ανάλυση και επεξεργασία δεδομένων

5.5.1 Κατανόηση δεδομένων

Για να ξεκινήσει η διαδικασία των πειραμάτων αρχικά εκτυπώνονται τα δεδομένα ώστε μελετηθούν.

Date	Open	High	Low	Close	Adj Close	Volume
1/4/2016	25.65	26.34	25.5	26.33	24.04	270597600
1/5/2016	26.43	26.46	25.6	25.67	23.43	223164000
1/6/2016	25.14	25.59	24.96	25.17	22.98	273829600
1/7/2016	24.67	25.03	24.1	24.11	22.01	324377600
1/8/2016	24.63	24.77	24.19	24.24	22.12	283192000

Σχήμα 5.2 Δείγμα Δεδομένων για την μετοχή Apple

Όπως φαίνεται στο σχήμα 5.2 για κάθε μέρα του διαστήματος μελέτης έχουμε διαθέσιμη την τιμή ανοίγματος, την υψηλότερη τιμή της ημέρας, τη χαμηλότερη τιμή της ημέρας, τη τιμή κλεισίματος, τη προσαρμοσμένη τιμή κλεισίματος και τον όγκο συναλλαγών. Η προσαρμοσμένη τιμή κλεισίματος είναι αυτή που θα χρησιμοποιείται στα πειράματα ως τιμή της μετοχής καθώς έχει γίνει προσαρμογή σε γεγονότα όπως διάσπαση μετοχών (stock split) και διαμοιρασμό μερισμάτων με σκοπό η τιμή να εμφανίζει συνέχεια στη πάροδο του χρόνου.

5.5.2 Επεξεργασία δεδομένων

Χρησιμοποιώντας τα παραπάνω δεδομένα θα κατασκευάσουμε νέα χαρακτηριστικά (features), όπως δείκτες τάσης και όγκου οι οποίοι έχουν χρησιμοποιηθεί εκτενώς στη βιβλιογραφία. Από αυτά που αναφέρθηκαν παραπάνω χρησιμοποιούμε τα πιο σημαντικά. Τα features αυτά θα χρησιμοποιηθούν σε διάφορους συνδυασμούς στα πειράματα παρακάτω:

Δείκτης Σχετικής Δύναμης (ΔΣΔ - RSI): Χρησιμοποιώντας τον τύπο του ΔΣΔ παράγουμε τις τιμές του για τις παραπάνω μετοχες για το διάστημα μελέτης. Παρακάτω φαίνονται οι τιμές του ΔΣΔ για την Apple τις πρώτες μέρες του διαστήματος μελέτης:

Date	RSI
1/4/2016	24.04148
1/5/2016	23.43903
1/6/2016	22.98033
1/7/2016	22.01045

Κινητός Μέσος Όρος Σύγκλισης - Απόκλισης (ΚΜΟΣΑ - MACD): Παρομοίως για τον ΚΜΟΣΑ μπορούμε να να υπολογίσουμε τη διαφορά μεταξύ της γραμμής ΚΜΟΣΑ και της γραμμής Σήματος. Παρακάτω φαίνονται οι πρώτες μέρες της διαφοράς για τη μετοχή της Apple:

Date	MACD
1/4/2016	-0.08143
1/5/2016	-0.10796
1/6/2016	-0.14319
1/7/2016	-0.21433

Όγκος Ισορροπίας - On-Balance Volume (OI - OBV): Χρησιμοποιώντας την στήλη του όγκου των συναλλαγών υπολογίζουμε τον OI. Παρακάτω φαίνονται οι τιμές του OI για την Apple τις πρώτες μέρες του διαστήματος μελέτης:

Date	Obv
1/4/2016	-1674713600
1/5/2016	-1897877600
1/6/2016	-2171707200
1/7/2016	-2496084800

5.6 Καθορισμός περιβάλλοντος

Ο χώρος ενεργειών όπως εξηγήθηκε παραπάνω για n μετοχές είναι 3^n έτσι για το χαρτοφυλάκιο που θα εξεταστεί, δηλαδή για 3 μετοχές, οι δυνατές ενέργειες θα είναι $3^3 = 27$.

5.7 Σχεδιαστικές επιλογές πειραμάτων

5.7.1 Βελτιστοποίηση υπερ-παραμέτρων

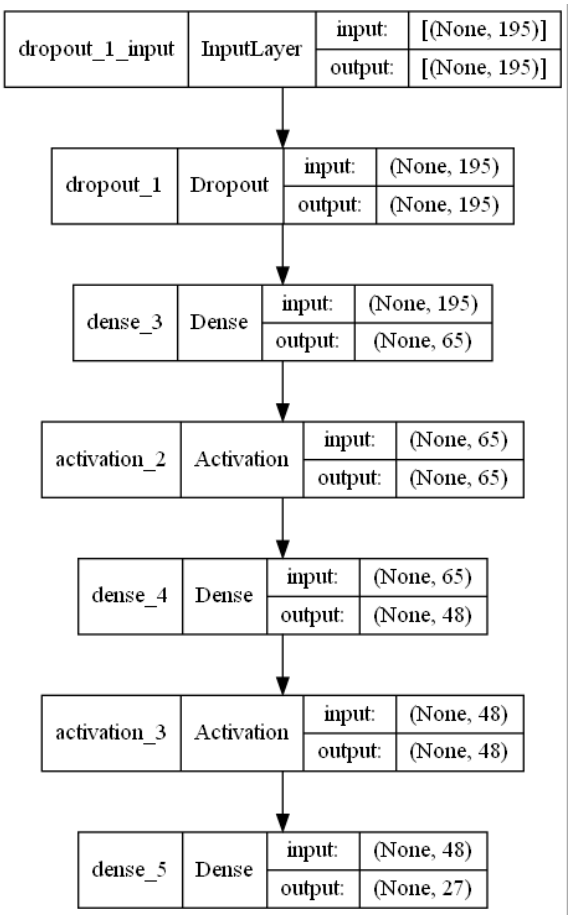
Οι υπερ-παράμετροι που επιλέχθηκαν στα παρακάτω πειράματα βελτιστοποίησης είναι:

- Η συχνότητα εκπαίδευσης
- Η συχνότητα ενημέρωσης του Δικτύου Στόχου (Target Network)
- Η ταχύτητα μείωσης του παράγοντα epsilon (epsilon greedy)
- Ο παράγοντας Έκπτωσης (γ)

5.7.2 Καθορισμός αρχιτεκτονικής και παραμέτρων

Σκοπός της παρούσας εργασίας είναι η μελέτη του προβλήματος Ενισχυτικής Μάθησης συνεπώς κρίθηκε σκόπιμο όλα τα πειράματα που θα ακολουθήσουν να χρησιμοποιούν την ίδια αρχιτεκτονική Νευρωνικού Δικτύου για τον υπολογισμό των Q τιμών, η οποία δεν

μεταβάλλεται. Αποφασίστηκε αυτό έτσι ώστε, μεταβάλλοντας αποκλειστικά τις παραμέτρους Ενισχυτικής Μάθησης, να μελετηθεί το πως επηρεάζει η κάθε παράμετρος την ικανότητα του πράκτορα, πρώτον να μαθαίνει στα δεδομένα εκπαίδευσης, και στη συνέχεια να γενικεύει σε άγνωστα δεδομένα. Παρακάτω λοιπόν παρουσιάζονται οι επιλογές που έγιναν αναφορικά με την αρχιτεκτονική του δικτύου και τις παραμέτρους μάθησης του:



5.3 Αρχιτεκτονική νευρωνικού δικτύου πειραμάτων

Συγκεκριμένα, η αρχιτεκτονική του Νευρωνικού Δικτύου που χρησιμοποιήθηκε για τα Μοντέλα 1, 2 και 3 φαίνεται στο σχήμα 5.4 και είναι η ακόλουθη:

1. Dropout στρώμα εισόδου με παράμετρο 0.25. Η είσοδος του νευρωνικού είναι ίση με το μέγεθος της κατάστασης του συστήματος. Παρακάτω θα αναλυθούν οι διαφορετικές καταστάσεις που χρησιμοποιήθηκαν στα πειράματα.

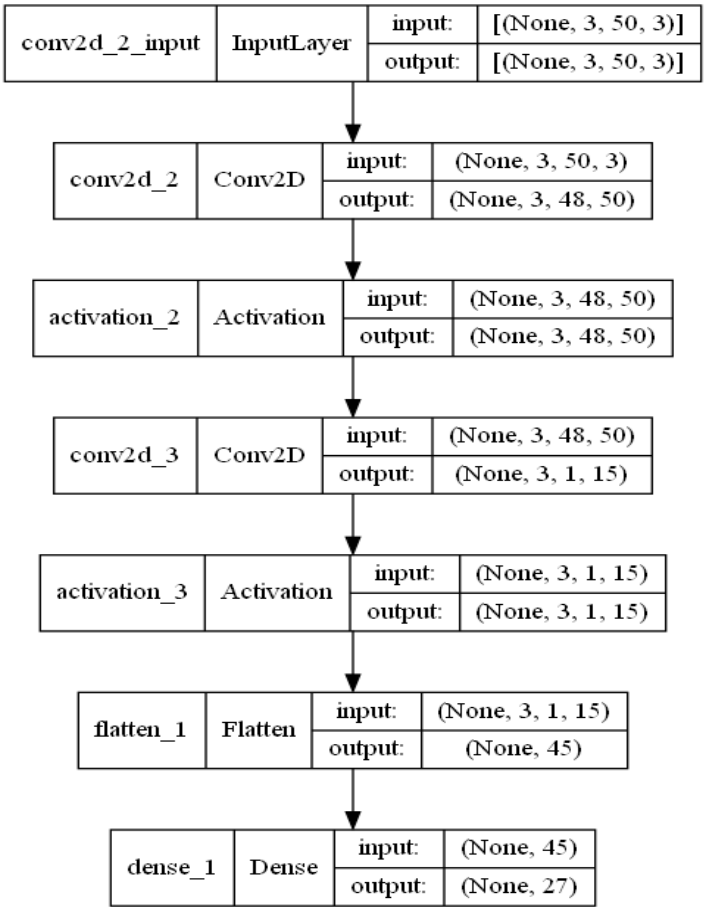
2. Πυκνό κρυφό στρώμα με μέγεθος το 1/3 του μεγέθους της εισόδου και συνάρτηση ενεργοποίησης τη συνάρτηση γραμμικής ρύθμισης (Rectified Linear Unit - ReLU).

3. Πυκνό κρυφό στρώμα με μέγεθος το 1/4 του μεγέθους της εισόδου και συνάρτηση ενεργοποίησης τη συνάρτηση γραμμικής ρύθμισης (Rectified Linear Unit - ReLU).

4. Πυκνό στρώμα εξόδου με μέγεθος ίσο με τις διαθέσιμες ενέργειες, δηλαδή 27 και γραμμική (linear) συνάρτηση ενεργοποίησης.

Επιλέχθηκε μια αρχιτεκτονική η οποία προσαρμόζεται στο μέγεθος της εισόδου ώστε μεταβάλλοντας την κατάσταση εισόδου του συστήματος να προσαρμόζεται και το δίκτυο σε αυτή.

Ο βελτιστοποιητής που επιλέχθηκε για το σύστημα είναι ο Adam με ρυθμό μάθησης 0.001 ο οποίος είναι από τους πιο αποδοτικούς βελτιστοποιητές καθώς συνδυάζει τα πλεονεκτήματα του Adagrad, δηλαδή τη προσαρμογή του ρυθμού μάθησης για κάθε παράμετρο, και του Momentum που χρησιμοποιεί μέσους όρους των προηγούμενων παραγώγων για να επιταχύνει τη σύγκλιση.



Σχήμα 5.4 Αρχιτεκτονική δικτύου Μοντέλου ΣΝΔ

Το Μοντέλο 4 όπως εξηγήθηκε και παραπάνω χρησιμοποιεί Νευρωνικό Δίκτυο που αποτελείται από συνελικτικά στρώματα για την εκτίμηση των Q τιμών καθώς στόχος είναι να αξιολογηθεί η αποδοτικότητα της χρήσης της αρχιτεκτονικής που φαίνεται στο Σχήμα 5.4 έναντι της προηγούμενης η οποία αποτελείται από απλά feed-forward στρώματα.

5.7.3 Μετρικές αξιολόγησης αποτελεσμάτων

Οι τεχνικές που αποφασίστηκε να χρησιμοποιηθούν ως αναφορά (benchmark) για την αξιολόγηση είναι:

1. Universal Buy and Hold - UBH: Το διαθέσιμο κεφάλαιο μοιράζεται ισόποσα αγοράζοντας όλες τις μετοχές του χαρτοφυλακίου και διατηρείται έτσι μέχρι το τέλος του διαστήματος μελέτης.

2. Efficient Frontier - EF: Η σύνθεση του χαρτοφυλακίου ισορροπείται ανά τακτά διαστήματα με σκοπό να μεγιστοποιηθεί η απόδοση προσαρμοσμένη με την έκθεση σε ρίσκο.

3. Follow the Leader - FtL: Σε κάθε βήμα απόφασης, επιλέγεται η μετοχή του χαρτοφυλακίου με τη μεγαλύτερη άνοδο και όλο το κεφάλαιο διατίθεται για την αγορά αυτής.

4. Follow the Leader - FtL (Buy - Sell): Ίδια με τη FtL αλλά αυτή τη φορά επιλέγεται η μετοχή με τη μεγαλύτερη πτώση ή άνοδο και ακολουθείται αυτή.

5. Universal Constant Rebalance - UCR: Σε κάθε βήμα απόφασης το διαθέσιμο κεφάλαιο ανακατανέμεται ισόποσα σε όλες τις μετοχές του χαρτοφυλακίου με σκοπό ανά πάσα στιγμή όλες οι μετοχές του χαρτοφυλακίου να έχουν την ίδια κεφαλαιοποίηση.

Επιπλέον, μετρικές που κρίθηκαν ως πιο κατάλληλες για την αξιολόγηση των αποτελεσμάτων είναι είναι:

- Απόδοση επένδυσης (ROI) μετά από τη διαδικασία εκπαίδευσης
- Συντελεστής Sharpe (Sharpe Ratio) μετά από τη διαδικασία εκπαίδευσης

Ενώ οι ακόλουθες γραφικές παραστάσεις θα χρησιμοποιηθούν για να διαπιστωθεί η σταθερότητα της εκπαίδευσης, η ικανότητα γενίκευσης και η ταχύτητα σύγκλισης:

- Γραφική παράσταση μέσης ανταμοιβής ενέργειας ανά επεισόδιο
- Ιστόγραμμα των αποδόσεων επένδυσης ανά επεισόδιο

5.8 Αποτελέσματα πειραμάτων

Τα πρώτα πειράματα που εκπονήθηκαν είχαν ως σκοπό την εύρεση των βέλτιστων υπερ-παραμέτρων εκπαίδευσης. Για το σκοπό αυτό αποφασίστηκε η είσοδος-κατάσταση, του συστήματος να είναι η ίδια σε όλα τα πειράματα και οι μόνες αλλαγές να αφορούν τις υπερ-παραμέτρους εκπαίδευσης. Έτσι στα πειραματα χρησιμοποιήθηκαν τα παρακάτω δεδομένα εισόδου για κάθε μετοχή του χαρτοφυλακίου:

1. Τιμή Ανοίγματος 16 ημερών
2. Τιμή Κλεισίματος 16 ημερών
3. Υψηλή Τιμή 16 ημερών
4. Χαμηλή Τιμή 16 ημερών
5. Αριθμός μετοχών στη παρούσα σύνθεση του χαρτοφυλακίου, π.χ 1 αν ο πράκτορας έχει αγοράσει 1 μετοχή ή -1 αν έχει πουλήσει μια μετοχή

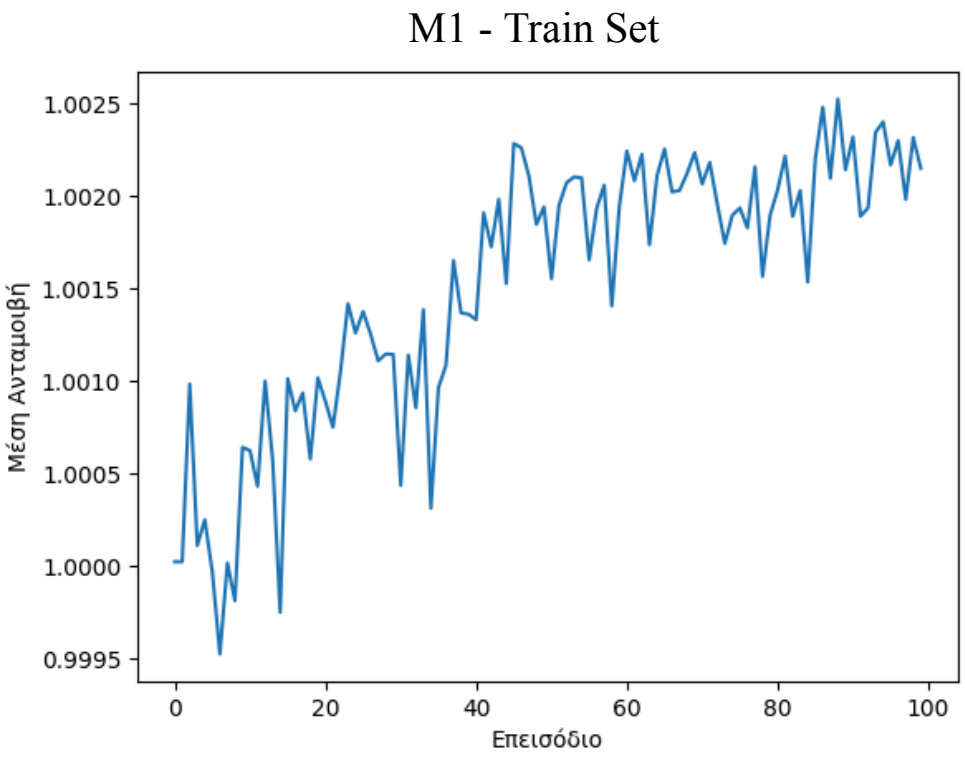
Χρησιμοποιώντας τις παραπάνω εισόδους εκτελέστηκε μία σειρά πειραμάτων. Οι μετρικές που χρησιμοποιήθηκαν για την αξιολόγηση των βέλτιστων τιμών των υπερπαραμέτρων είναι η Απόδοση Επένδυσης (ROI) του εκπαίδευμένου πράκτορα στα δεδομένα εκπαίδευσης και το Sharpe Ratio του εκπαιδευμένου πράκτορα στα δεδομένα εκπαίδευσης.

Από τα αποτελέσματα του πίνακα Π.1 (Παράρτημα), είναι φανερό πως ο συνδυασμός παραμέτρων 9 επιφέρει τα καλύτερα αποτελέσματα και είναι:

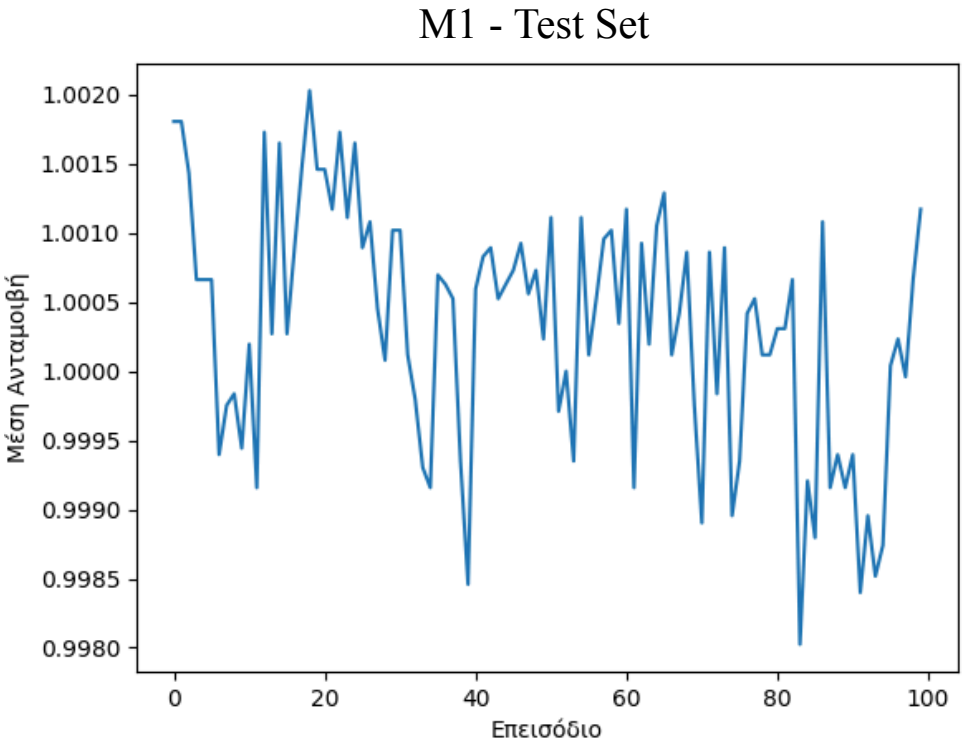
- Συχνότητα εκπαίδευσης πράκτορα (μέρες) - **20**
- Συχνότητα ενημέρωσης δικτύου στόχου (αριθμός εκπαιδεύσεων) - **1200**
- Μέγεθος μνήμης εμπειριών (παρατηρήσεις) - **4000**
- Παράγοντας μείωσης epsilon - **0.99**
- Παράγοντας έκπτωση - **0.99**

5.8.1 Μοντέλο 1 - Open, High, Low, Close

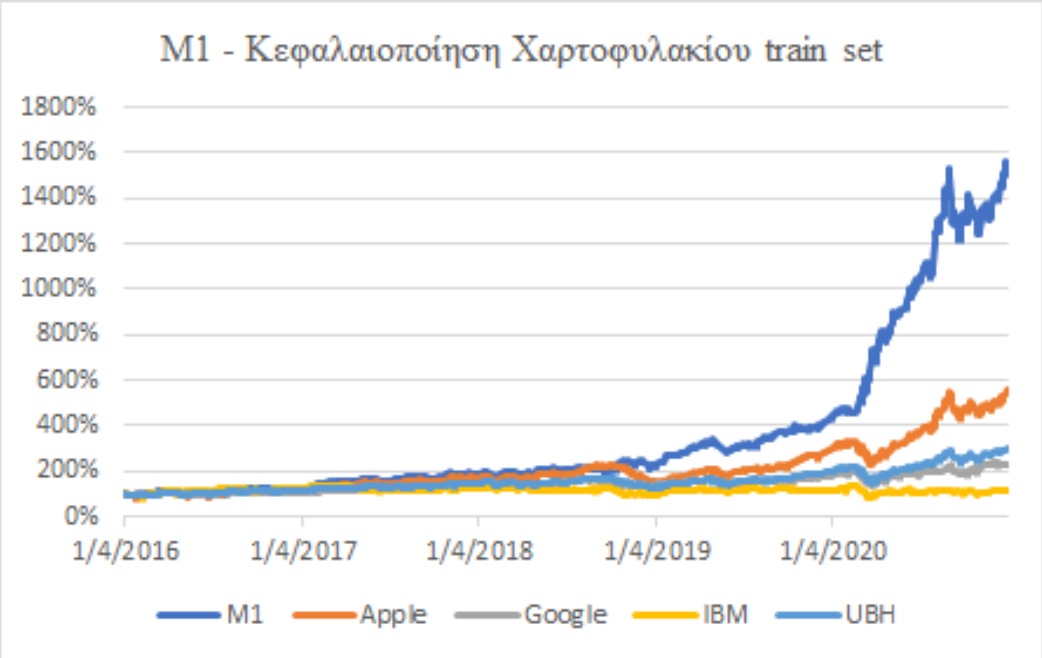
Χρησιμοποιώντας τις παραπάνω παραμέτρους προχωράμε στην υλοποίηση των πρακτόρων που θα μελετηθούν. Ο πρώτος πράκτορας χρησιμοποιεί σαν είσοδο αυτή που περιγράφηκε παραπάνω. Για κάθε μετοχή τα δεδομένα εισόδου θα αποτελούνται από 65 τιμές, άρα για τις τρεις μετοχές του χαρτοφυλακίου θα έχουμε 195 δεδομένα εισόδου το οποίο αποτελεί και το μέγεθος του χώρου κατάστασης.



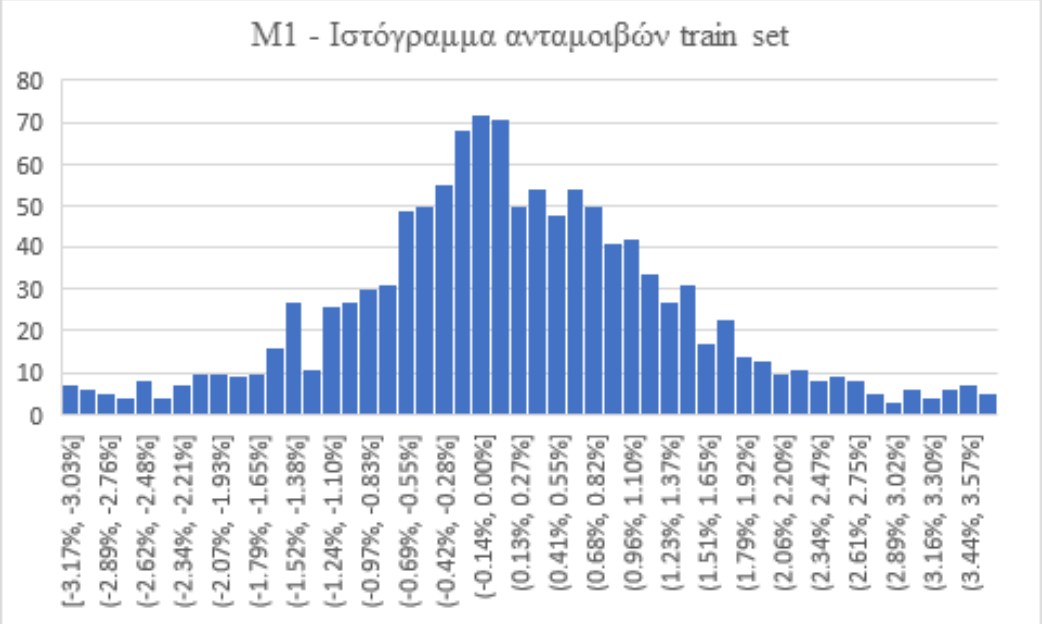
Σχήμα 5.4 M1 - Μέση ανταμοιβή ανά επεισόδιο στα δεδομένα εκπαίδευσης



Σχήμα 5.5 Μοντέλο M1 - Μέσης ανταμοιβή ανά επεισόδιο στα δεδομένα τεστ



Σχήμα 5.6 M1 - Κεφαλαιοποίηση Χαρτοφυλακίου Train set



Σχήμα 5.7 M1 - Ιστόγραμμα ανταμοιβών Train set

Train set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
APPLE	448%	-	1.44	-
GOOGLE	129%	-	0.73	-
IBM	14%	-	0.11	-
Χαρτοφυλάκιο				
UBH	197%	5	1.08	3
EF	324%	3	0.83	5
FtL	426%	2	1.16	2
FtL (BUY-SELL)	-71%	6	-0.75	6
UCR	259%	4	0.98	4
M1 (O-H-L-C)	**1432%**	**1**	**2.8**	**1**

Πίνακας 5.2 M1 - Σύγκριση με παραδοσιακές μεθόδους διαχείρισης χαρτοφυλακίου - Train set

Test set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
APPLE	39%	-	1.34	-
GOOGLE	69%	-	2.15	-
IBM	19%	-	0.81	-
Χαρτοφυλάκιο				
UBH	42%	2	2.03	2
EF	39%	4	1.34	4
FtL	40%	3	1.52	3
FtL (BUY-SELL)	-37%	6	-2.13	6
UCR	**42%**	**1**	**2.04**	**1**
M1 (O-H-L-C)	34%	5	1.22	5

Πίνακας 5.3 M1 - Σύγκριση με παραδοσιακές μεθόδους διαχείρισης χαρτοφυλακίου - Test set

Είναι φανερό πως το M1 εκπαιδεύεται πολύ καλα στο train set και επιτυγχάνει πολύ καλή απόδοση παρόλα αυτά στο test set το μοντέλο δεν είναι σε θέση να γενικεύσει όσα έμαθε μέσα από τη διαδικασία εκπαίδευσης στο test set.

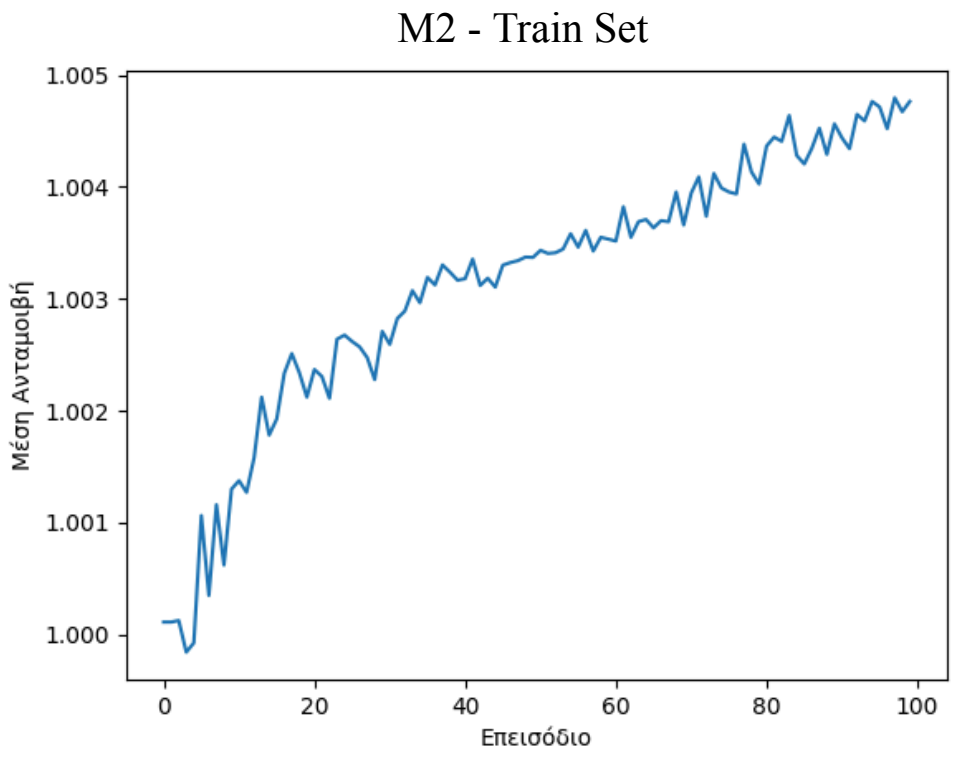
Συγκεκριμένα, η διαδικασία εκπαίδευσης, όπως φαίνεται και στο Σχήμα 5.5, γίνεται επιτυχημένα και το μοντέλο συνεχώς βελτιώνει την απόδοση του μέχρι που συγκλίνει. Στον πίνακα 5.1 είναι φανερό πως μετά το πέρας της εκπαίδευσης τόσο η επιστροφή όσο και το Sharpe Ratio είναι πολύ καλύτερο από όλες τις υπόλοιπες παραδοσιακές τεχνικές διαχείρισης χαρτοφυλακίου, κάτι που όμως δεν φαίνεται να ισχύει στο test set καθώς είναι φανερό πως στο test set το M1 είναι χειρότερο από τις υπόλοιπες τεχνικές και όπως φαίνεται και στο Σχήμα 5.6 η εκπαίδευση δεν φαίνεται να προσφέρει στο μοντέλο γνώσεις που του επιτρέπουν να γενικεύσει σε άγνωστα δεδομένα.

5.8.2 Μοντέλο 2 - Daily Percentage Change

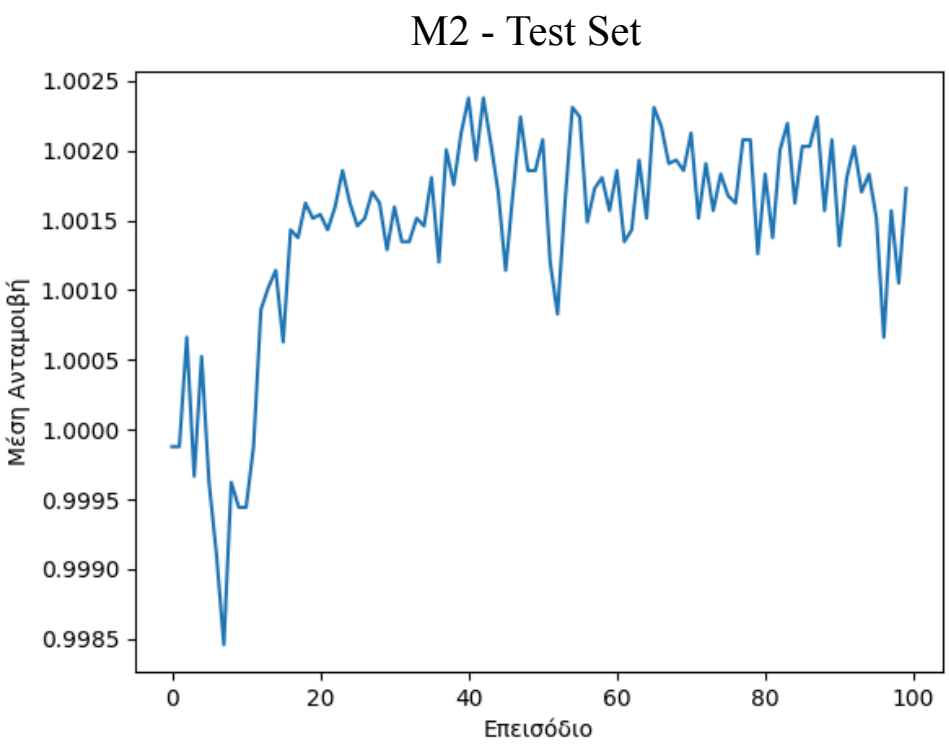
Η είσοδος-κατάσταση του Μοντέλου 2 είναι:

1. Ημερήσιες ποσοστιαίες μεταβολές τιμής 60 ημερών
2. Αριθμός μετοχών στη παρούσα σύνθεση του χαρτοφυλακίου, π.χ 1 αν ο πράκτορας έχει αγοράσει 1 μετοχή ή -1 αν έχει πουλήσει μια μετοχή

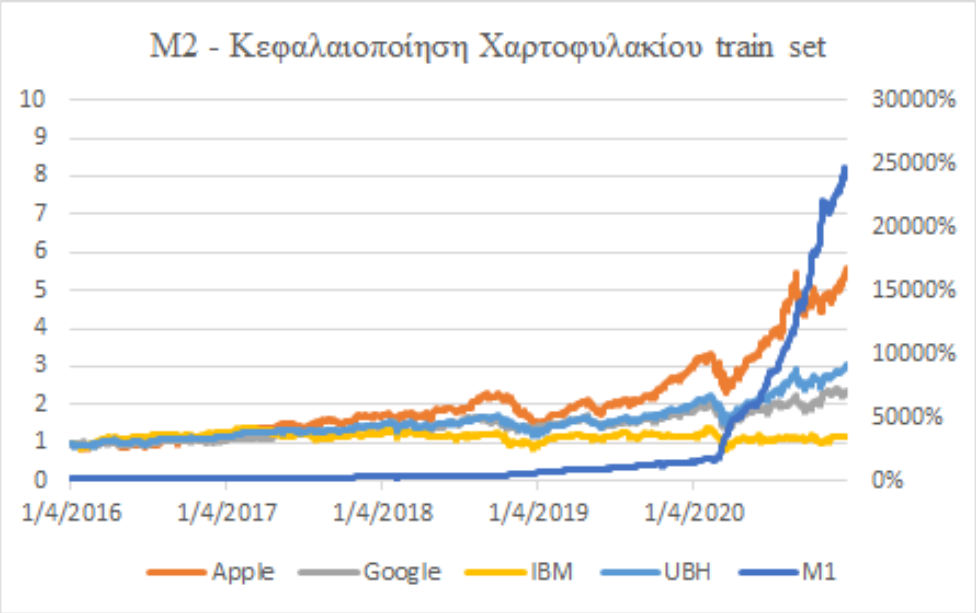
Για κάθε μετοχή τα δεδομένα εισόδου θα αποτελούνται από 61 τιμές, άρα για τις τρεις μετοχές του χαρτοφυλακίου θα έχουμε 183 δεδομένα εισόδου το οποίο αποτελεί και το μέγεθος του χώρου κατάστασης.



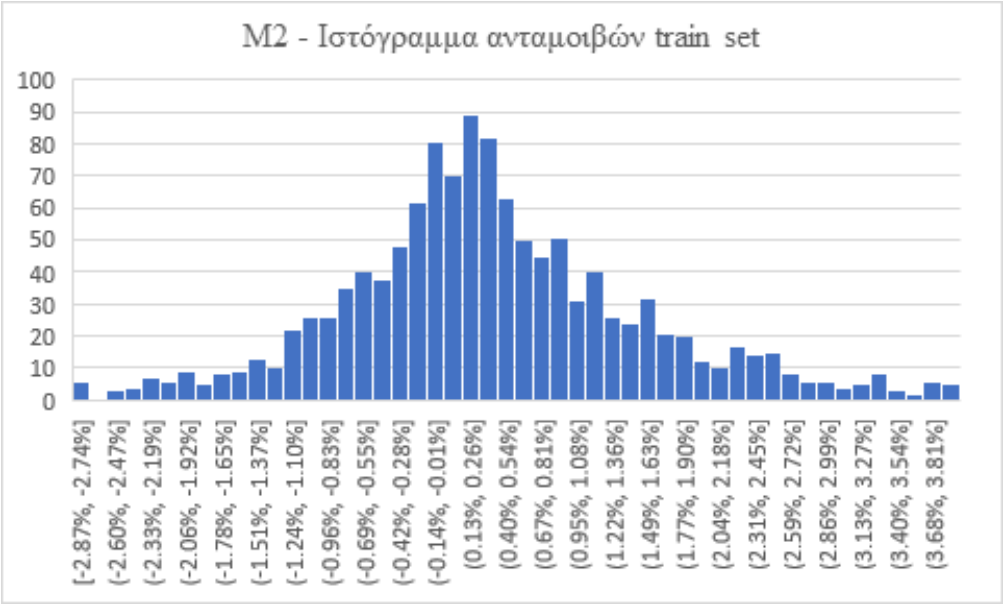
Σχήμα 5.8 M2 - Μέση ανταμοιβή ανά επεισόδιο στο train set



Σχήμα 5.9 M2 - Μέση ανταμοιβή ανά επεισόδιο στο test set



Σχήμα 5.10 M2 - Κεφαλαιοποίηση Χαρτοφυλακίου Train Set



Σχήμα 5.11 M2 - Ιστόγραμμα ανταμοιβών train set

Train set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
APPLE	448%	-	1.44	-
GOOGLE	129%	-	0.73	-
IBM	14%	-	0.11	-
Χαρτοφυλάκιο				
UBH	197%	5	1.08	3
EF	324%	3	0.83	5
FtL	426%	2	1.16	2
FtL (BUY-SELL)	-71%	6	-0.75	6
UCR	259%	4	0.98	4
M2 (DPC)	**24590%**	**1**	**10.57**	**1**

Πίνακας 5.5 M2 - Σύγκριση με παραδοσιακές μεθόδους διαχείρισης χαρτοφυλακίου - Train set

Test set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
APPLE	39%	-	1.34	-
GOOGLE	69%	-	2.15	-
IBM	19%	-	0.81	-
Χαρτοφυλάκιο				
UBH	42%	3	2.03	3
EF	39%	5	1.34	5
FtL	40%	4	1.52	4
FtL (BUY-SELL)	-37%	6	-2.13	6
UCR	42%	2	2.04	2
M2 (DPC)	**61%**	**1**	**2.2**	**1**

Πίνακας 5.6 M2 - Σύγκριση με παραδοσιακές μεθόδους διαχείρισης χαρτοφυλακίου - Test set

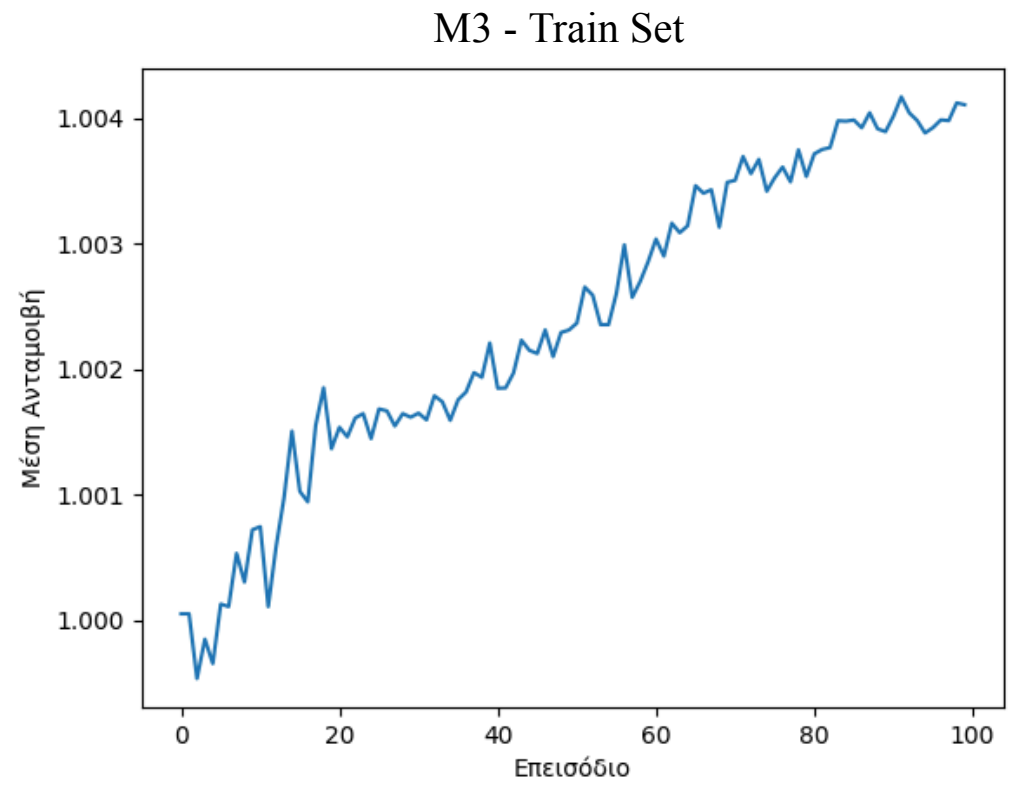
Όπως και το προηγούμενο μοντέλο έτσι και το M2 εκπαιδεύεται πολύ καλά στο train set και επιτυγχάνει πολύ καλή απόδοση. Φαίνεται επιπλέον ότι και στο test set το μοντέλο είναι πλέον σε θέση να γενικεύσει όσα έμαθε μέσα από τη διαδικασία εκπαίδευσης επιτυγχάνοντας πολύ καλα αποτελέσματα.

Συγκεκριμένα, η διαδικασία εκπαίδευσης, όπως φαίνεται και στο Σχήμα 5.10, βοηθάει το μοντέλο να μαθαίνει όλο και καλύτερα επιτυγχάνοντας όπως φαίνεται και στον πίνακα 5.3 πολύ καλή επιστροφή και υψηλό Sharpe Ratio όπως συνέβαινε και στο προηγούμενο μοντέλο. Το πλεονέκτημα του M2 φαίνεται στο test set όπου είναι φανερό πως αποδίδει πολύ καλύτερα από τις υπόλοιπες τεχνικές διαχείρισης χαρτοφυλακίου κάτι που σημαίνει ότι M2 είναι σε θέση να γενικεύσει αυτά που έμαθε κατά τη διάρκεια της εκπαίδευσης σε άγνωστα δεδομένα. Αυτό επιβεβαιώνεται και στο σχήμα 5.10 όπου φαίνεται πως το μοντέλο γρήγορα βελτιώνει την απόδοση του σε άγνωστα δεδομένα και συγκλίνει.

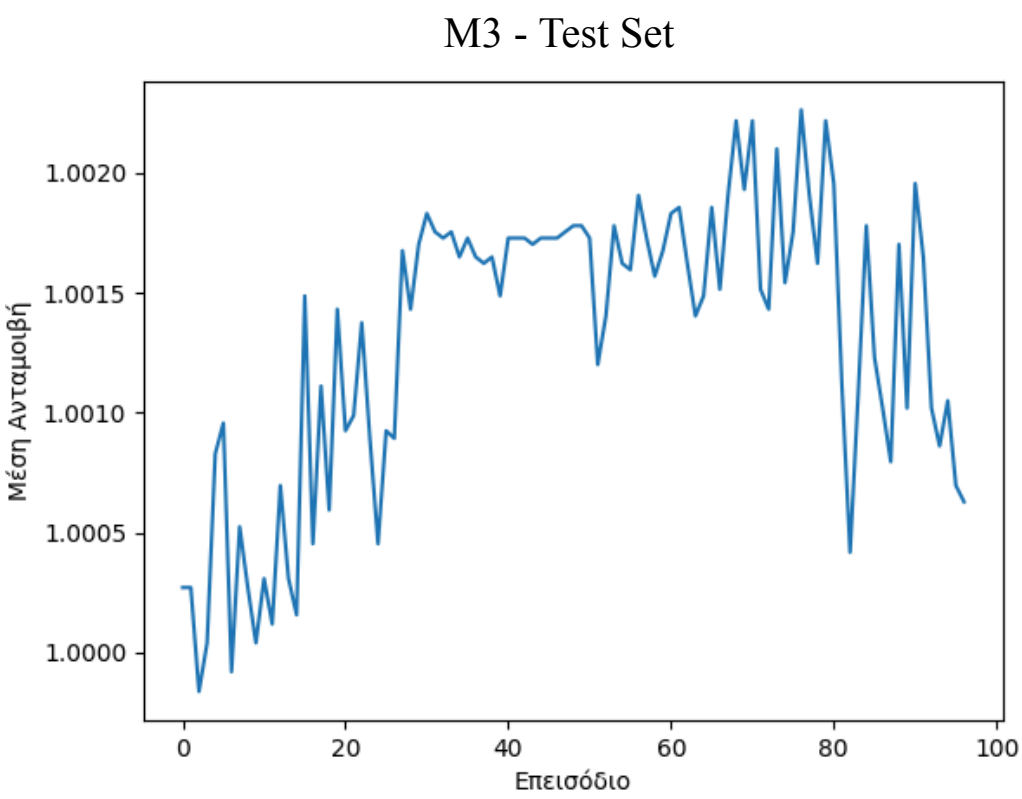
5.8.2 Μοντέλο 3 - Daily Percentage Changes and Statistical Indicators

Ο πράκτορας M3 (Ημερήσιες Ποσοστιαίες Μεταβολές - Στατιστικοί Δείκτες), χρησιμοποιεί τα παρακάτω δεδομένα εισόδου:

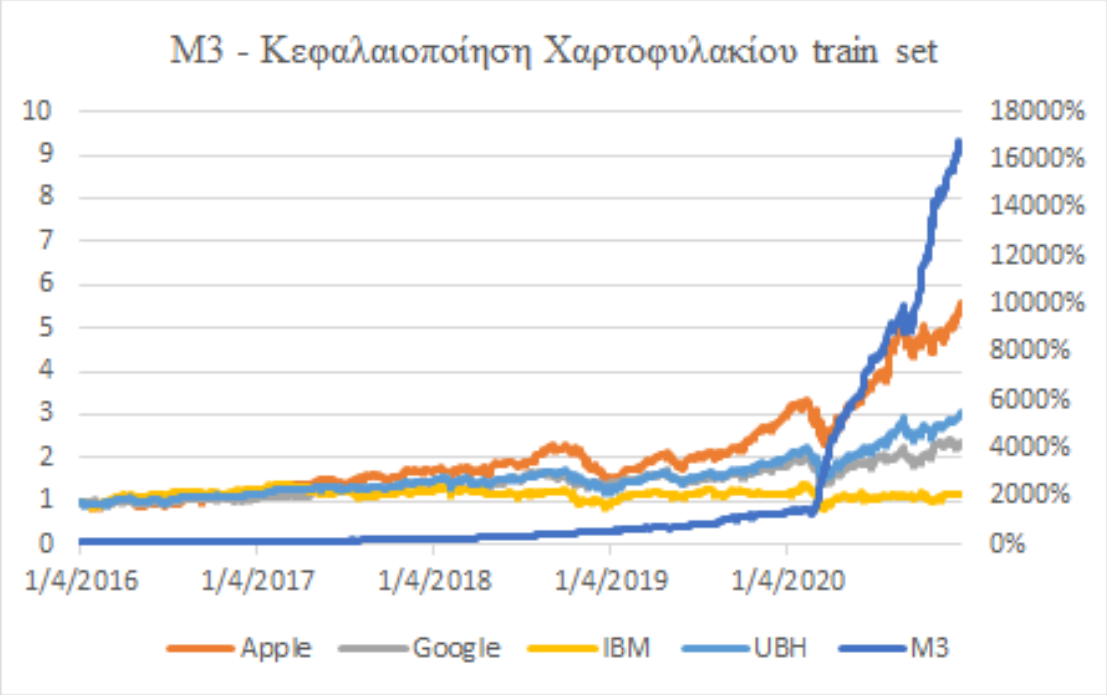
1. Ημερήσιες ποσοστιαίες μεταβολές τιμής 60 ημερών
2. Δείκτης Σχετικής Δύναμης (RSI)
3. Όγκος Ισορροπίας (OBV)
4. Κινητός Μέσος Όρος Σύγκλισης/Απόκλισης (MACD)
5. Εκθετικός Μέσος Όρος 30 Ημερών
6. Εκθετικός Μέσος Όρος 150 Ημερών
7. Αριθμός μετοχών στη παρούσα σύνθεση του χαρτοφυλακίου, π.χ 1 αν ο πράκτορας έχει αγοράσει 1 μετοχή ή -1 αν έχει πουλήσει μια μετοχή



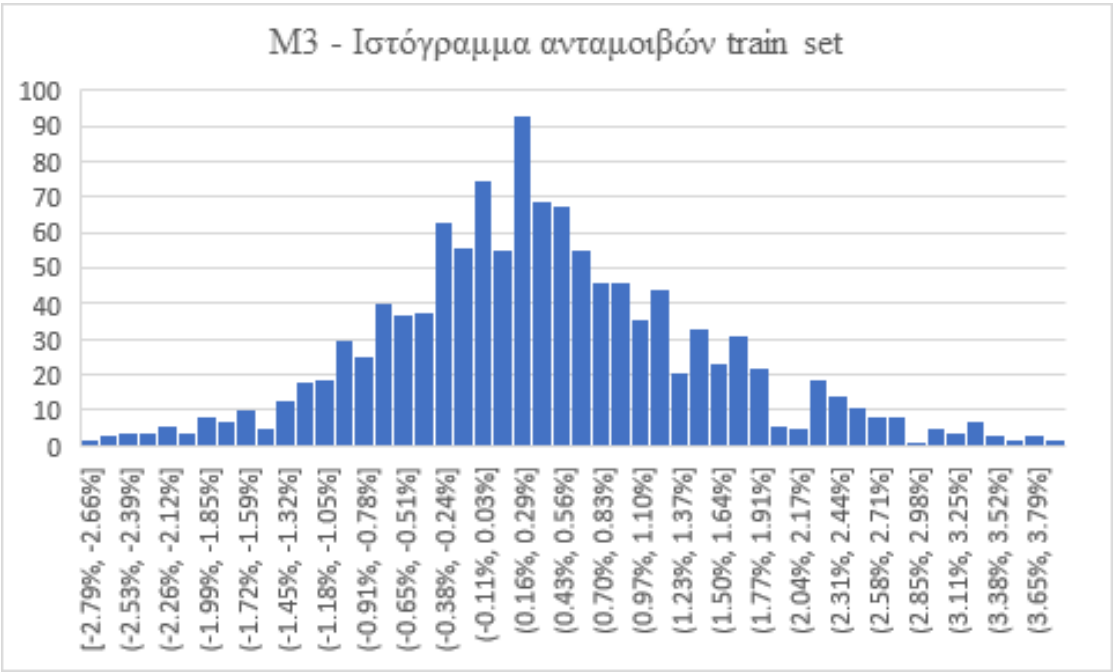
Σχήμα 5.12 M3 - Μέση ανταμοιβή ανά επεισόδιο στο train set



Σχήμα 5.13 M3 - Μέση ανταμοιβή ανά επεισόδιο στο test set



Σχήμα 5.14 M3 - Κεφαλαιοποίηση χαρτοφυλακίου train set



Σχήμα 5.15 M3 - Ιστόγραμμα ανταμοιβών train set

Train set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
APPLE	448%	-	1.44	-
GOOGLE	129%	-	0.73	-
IBM	14%	-	0.11	-
Χαρτοφυλάκιο				
UBH	197%	5	1.08	3
EF	324%	3	0.83	5
FtL	426%	2	1.16	2
FtL (BUY-SELL)	-71%	6	-0.75	6
UCR	259%	4	0.98	4
M3 (DPC&SI)	**16593%**	**1**	**8.09**	**1**

Πίνακας 5.7 M3 - Σύγκριση με παραδοσιακές μεθόδους διαχείρισης χαρτοφυλακίου - Train set

Test set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
APPLE	39%	-	1.34	-
GOOGLE	69%	-	2.15	-
IBM	19%	-	0.81	-
Χαρτοφυλάκιο				
UBH	42%	2	2.03	2
EF	39%	4	1.34	5
FtL	40%	3	1.52	3
FtL (BUY-SELL)	-37%	6	-2.13	6
UCR	**42%**	**1**	**2.04**	**1**
M3 (DPC&SI)	34%	5	1.43	4

Πίνακας 5.8 M3 - Σύγκριση με παραδοσιακές μεθόδους διαχείρισης χαρτοφυλακίου - Test set

Όπως και στα προηγούμενα μοντέλα παρατηρούμε πως και στο Μ3, το μοντέλο εκπαιδεύεται επιτυχώς στα δεδομένα εκπαίδευσης. Παρόλο όμως που η απόδοση του στο test set είναι αρκετά καλή, όπως φαίνεται και στον πίνακα 5.6 δεν απέχει πολύ από τις καλύτερες τεχνικές, είναι φανερό από το Σχήμα 5.14 πως η απόδοση του στο test set έχει μεγάλη μεταβλητότητα στο πέρας της εκπαίδευσης.

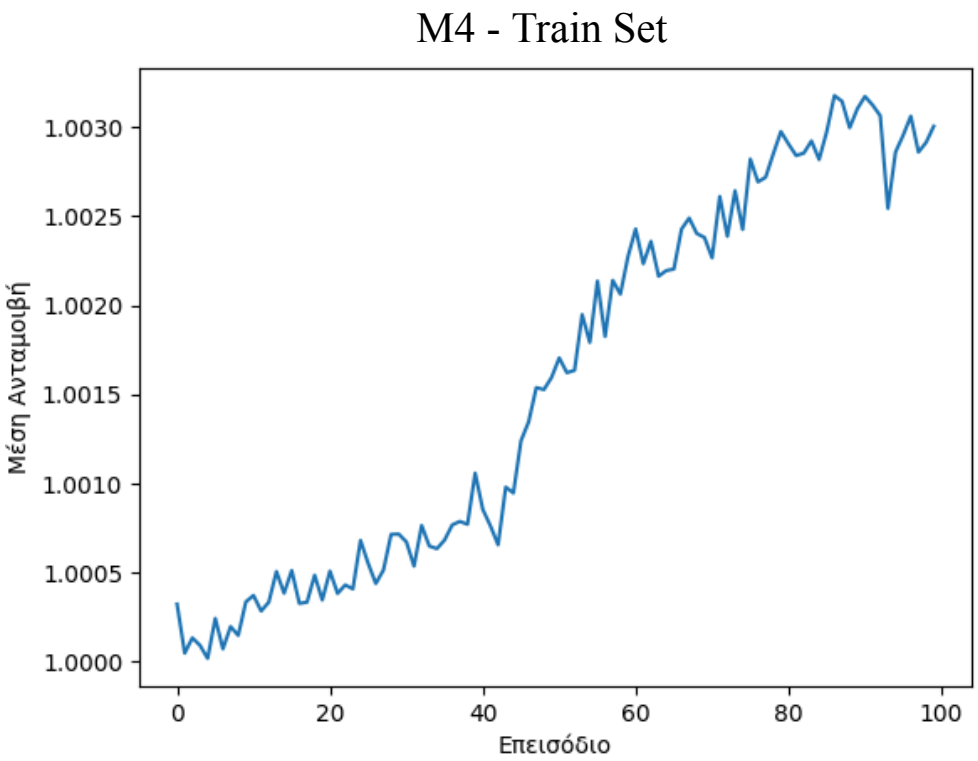
Παρατηρούμε λοιπόν, πως προσθέτοντας στατιστικούς δείκτες το Μ3 δεν είναι σε θέση να αποδόσει όσο καλά όσο το Μ2 χωρίς αυτούς, υποδεικνύοντας πως οι δείκτες αυτοί φαίνεται πως είτε είναι δύσκολο να αξιοποιηθούν από ένα απλό μοντέλο είτε δεν εμπεριέχουν κάποια προβλεπτική πληροφορία.

5.8.3 Μοντέλο 4 - Convolutional Neural Networks

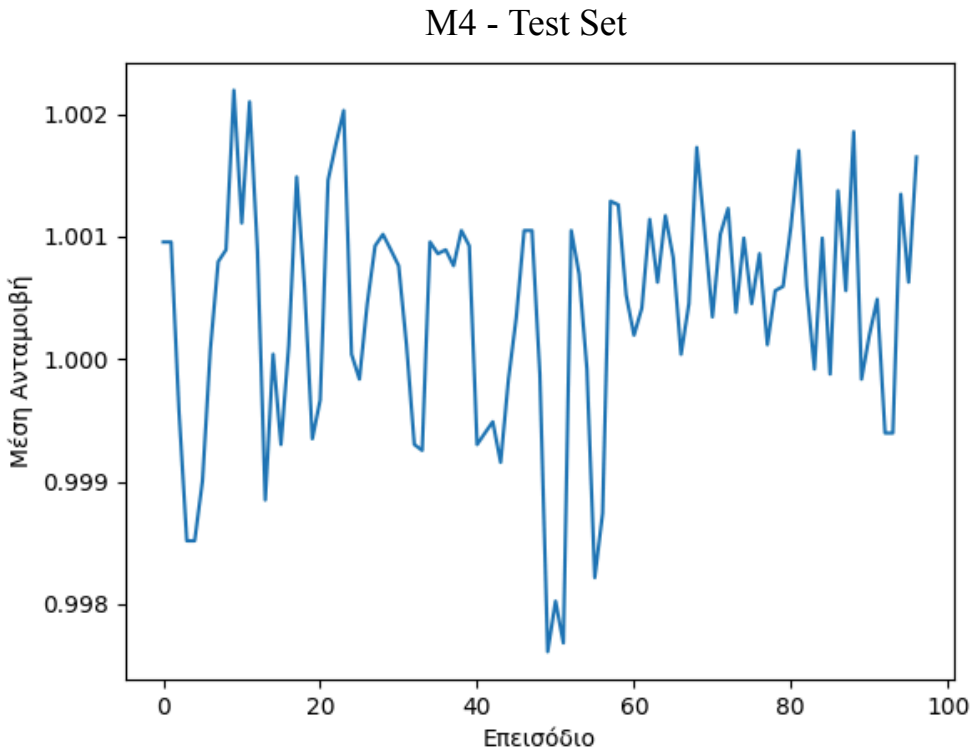
Η είσοδος-κατάσταση αποτελείται από την τιμή κλεισίματος, υψηλή και χαμηλή. Συγκεκριμένα:

1. Τιμή Κλεισίματος 50 ημερών
2. Τιμή Υψηλή 50 ημερών
3. Τιμή Χαμηλή 50 ημερών

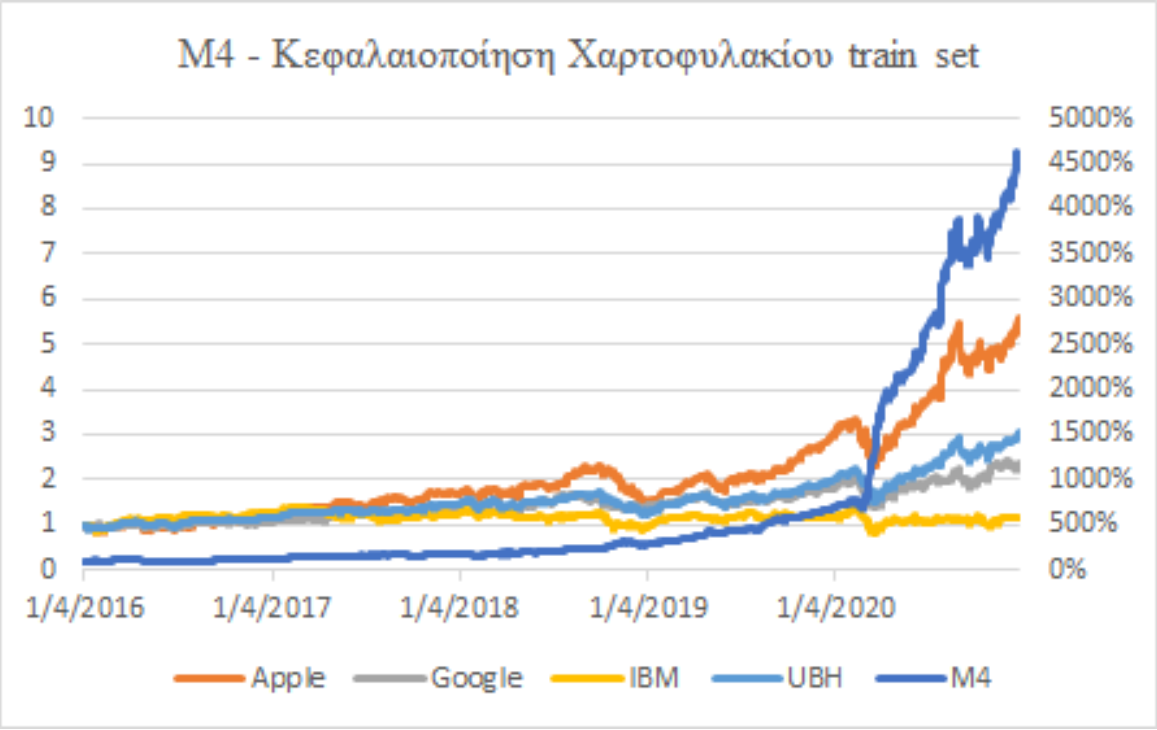
Άρα η είσοδος που περιγράφεται παραπάνω έχει μέγεθος 3 x 50 x 3, το οποίο θυμίζει μια RGB εικόνα μεγέθους 3Χ50. Έχοντας μετασχηματίσει την είσοδο σε κάτι το οποίο γνωρίζουμε πως η αρχιτεκτονική αυτή μπορεί να αντιμετωπίσει αποδοτικά προχωράμε στην εκτέλεση των πειραμάτων. Στο Σχήμα 5.11 φαίνεται η αρχιτεκτονική του δικτύου του Μοντέλου 4.



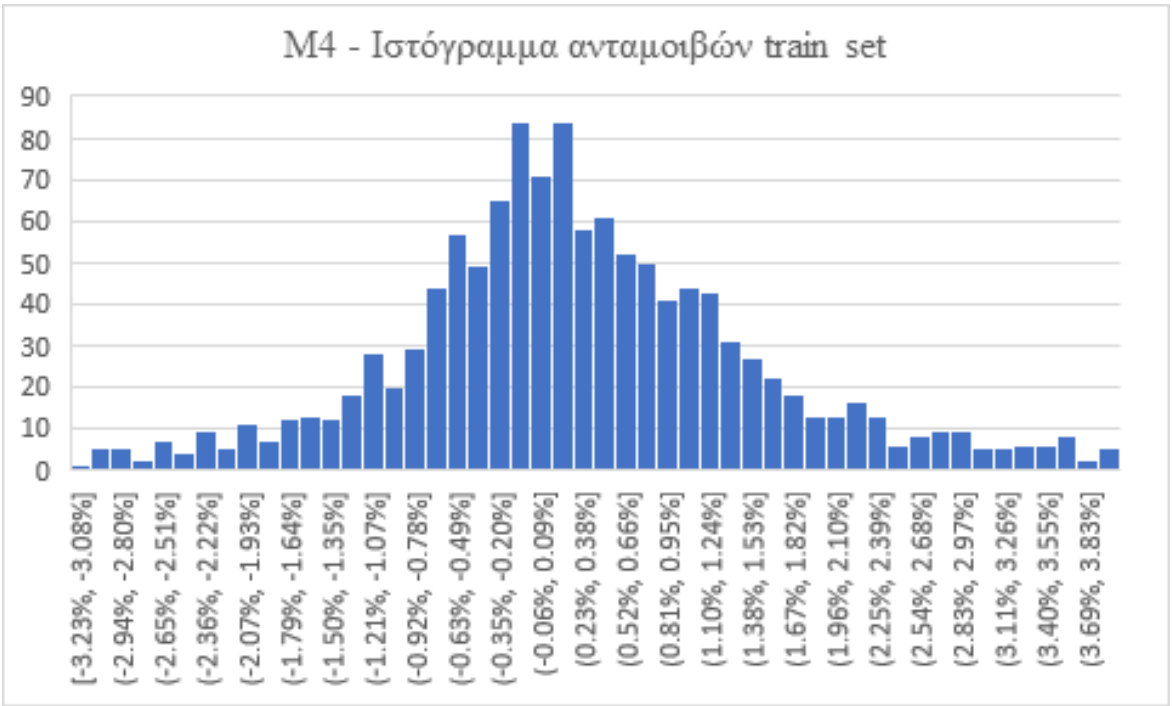
Σχήμα 5.17 M4 - Μέση ανταμοιβή ανά επεισόδιο στο train set



Σχήμα 5.18 M4 - Μέση ανταμοιβή ανά επεισόδιο στο test set



Σχήμα 5.19 M4 - Κεφαλαιοποίηση χαρτοφυλακίου train set



Σχήμα 5.20 M4 - Ιστόγραμμα ανταμοιβών train set

Train set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
APPLE	448%	-	1.44	-
GOOGLE	129%	-	0.73	-
IBM	14%	-	0.11	-
Χαρτοφυλάκιο				
UBH	197%	5	1.08	3
EF	324%	3	0.83	5
FtL	426%	2	1.16	2
FtL (BUY-SELL)	-71%	6	-0.75	6
UCR	259%	4	0.98	4
M4 (CNN)	**7072%**	**1**	**4.74**	**1**

Πίνακας 5.9 M4 - Σύγκριση με παραδοσιακές μεθόδους διαχείρισης χαρτοφυλακίου - Train set

Test set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
APPLE	39%	-	1.34	-
GOOGLE	69%	-	2.15	-
IBM	19%	-	0.81	-
Χαρτοφυλάκιο				
UBH	42%	3	2.03	3
EF	39%	5	1.34	5
FtL	40%	4	1.52	4
FtL (BUY-SELL)	-37%	6	-2.13	6
UCR	42%	2	2.04	2
M4 (CNN)	**49%**	**1**	**2.12**	**1**

Πίνακας 5.10 M4 - Σύγκριση με παραδοσιακές μεθόδους διαχείρισης χαρτοφυλακίου - Test set

Παρατηρώντας τον Πίνακα 5.7 γίνεται φανερό πως και σε αυτή τη περίπτωση το μοντέλο εκπαιδεύεται πολύ καλά (Σχήμα 5.17). Επιπλέον, φαίνεται πως η αλλαγή στην αρχιτεκτονική οδηγεί το M4 στο να έχει αρκετά καλή ικανότητα γενίκευσης κάτι το οποίο φαίνεται και από τον Πίνακα 5.8 όπου φαίνεται πως το μοντέλο υπερτερεί έναντι άλλον παραδοσιακών τεχνικών διαχείρισης χαρτοφυλακίου.

5.9 Αξιολόγηση σε διαφορετικά χαρτοφυλάκια

Σκοπός του κεφαλαίου αυτού είναι να εξεταστεί αν οι γνώσεις που αποκτήθηκαν στα προηγούμενα κεφάλαια μπορούν να μεταφερθούν και σε άλλα χαρτοφυλάκια με διαφορετικές ιδιότητες. Το κομμάτι αυτό είναι ιδιαίτερα σημαντικό καθώς σε αυτό το σημείο εξετάζεται αν ο πράκτορας που βελτιστοποιήθηκε στο συγκεκριμένο χαρτοφυλάκιο με τις μετοχές της Google, Apple και IBM είναι πράγματι ικανός να εντοπίσει και να αξιοποιήσει μοτίβα σε οποιοδήποτε χαρτοφυλάκιο ή λειτουργεί αποκλειστικά σε αυτό που βελτιστοποιήθηκε.

Train					**Test**				
Μετοχές	**ROI**		**Sharpe Ratio**		**Μετοχές**	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**		**Τιμή**	**Κατάταξη**	**Τιμή**	**Κατάταξη**
M1	1432%	4	2.8	4	M1	34%	8	1.22	8
M2	**24590%**	**1**	**10.57**	**1**	**M2**	**61%**	**1**	**2.2**	**1**
M3	16593%	2	8.09	2	M3	34%	7	1.43	6
M4	7072%	3	4.74	3	M4	49%	2	2.12	2
UBH	197%	8	1.08	6	UBH	42%	4	2.03	4
EF	324%	6	0.83	8	EF	39%	6	1.34	7
FtL	426%	5	1.16	5	FtL	40%	5	1.52	5
FtL (BUY-SELL)	-71%	9	-0.75	9	FtL (BUY-SELL)	-37%	9	-2.13	9
UCR	259%	7	0.98	7	UCR	42%	3	2.04	3

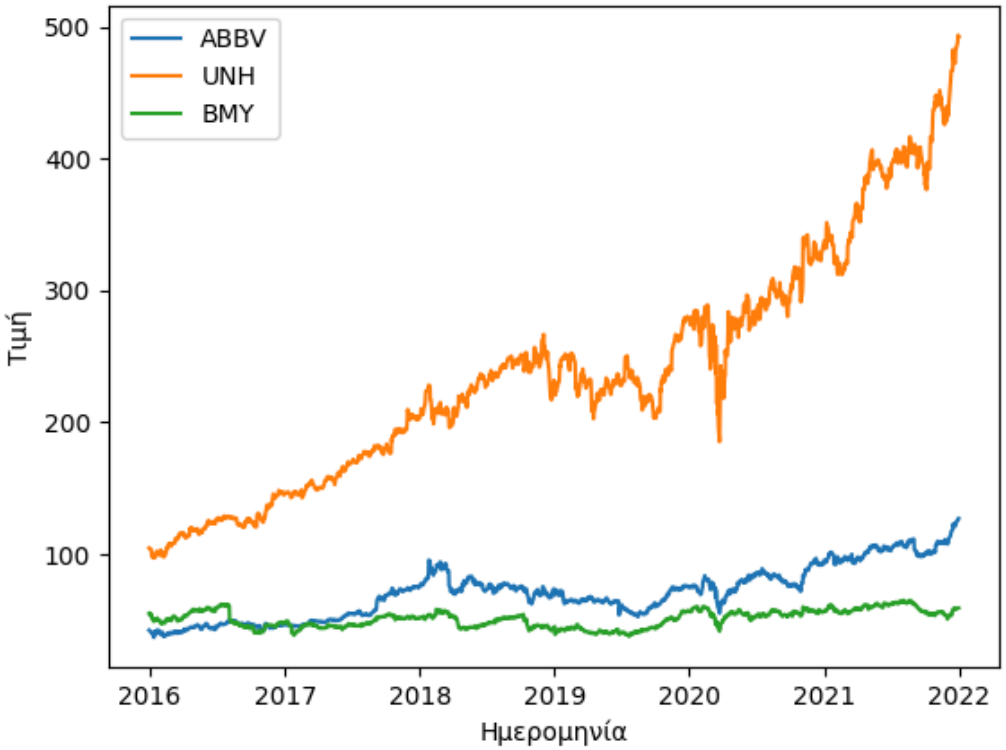
Πίνακας 5.11 Συγκριτικά αποτελέσματα μοντέλων

Χρησιμοποιώντας τα αποτελέσματα από τον Πίνακα 5.9 παρατηρούμε πως τόσο το M2 όσο και το M4 υπερτερούν έναντι των υπολοίπων παραδοσιακών μεθόδων Διαχείρισης Χαρτοφυλακίου ενώ το M2 φαίνεται πως υπερτερεί σημαντικά όλων. Συγκεκριμένα, παρατηρούμε ότι πετυχαίνει

τόσο την καλύτερη απόδοση όσο και το υψηλότερο Sharpe Ratio γεγονός που δείχνει ότι πετυχαίνει καλή απόδοση παίρνοντας χαμηλό ρίσκο.

Αξιοποιώντας λοιπόν τα παραπάνω συμπεράσματα επιλέγεται ο παράγοντας M2 ως ο βέλτιστος που θα χρησιμοποιηθεί ώστε να διαπιστωθεί αν η προτεινόμενη αρχιτεκτονική είναι ικανή να χρησιμοποιηθεί και σε πολλαπλά χαρτοφυλάκια. Όπως εξηγήθηκε παραπάνω, ο βελτιστοποιημένος πράκτορας εκπαιδεύτηκε σε μετοχές Τεχνολογίας, ενώ για τα παρακάτω πειράματα θα χρησιμοποιηθούν μετοχές από τους κλάδους της Υγείας, Ενέργειας και Οικονομικών. Παρακάτω παρουσιάζονται οι μετοχές και οι τιμές τους στην πάροδο του χρόνου:

1. **Τομέας Υγείας - ABBV, UNH, BMY**



Σχήμα 5.21 Τιμές μετοχών τομέα υγείας

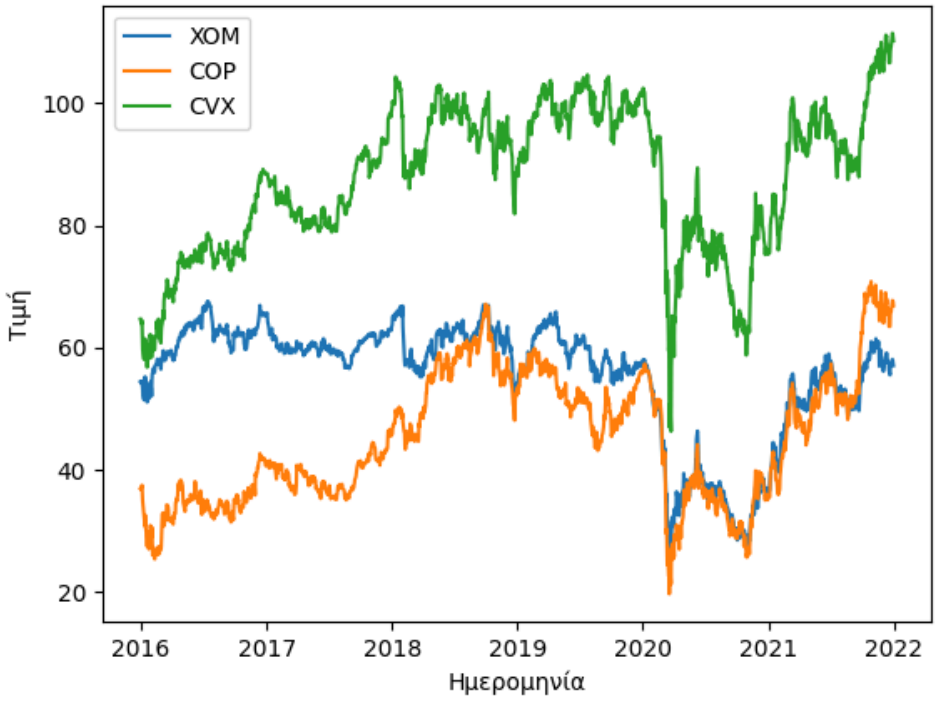
Train set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
ABBV	128%	-	0.66	-
BMY	6%	-	0.05	-
UNH	220%	-	0.99	-
Χαρτοφυλάκιο				
UBH	118%	2	0.78	3
EF	118%	3	0.75	4
FtL	74%	5	0.51	5
FtL (BUY-SELL)	-74%	6	-0.84	6
UCR	114%	4	0.8	2
M2 (DPC)	**4253%**	**1**	**5.34**	**1**

Πίνακας 5.12 Αποτελέσματα τομέα υγείας - Train set

Test set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
ABBV	35%	-	1.48	-
BMY	4%	-	0.28	-
UNH	46%	-	1.93	-
Χαρτοφυλάκιο				
UBH	28%	4	1.64	4
EF	34%	1	1.67	3
FtL	24%	5	1.25	5
FtL (BUY-SELL)	-22%	6	-1.17	6
UCR	28%	3	**1.75**	**1**
M2 (DPC)	**32%**	**2**	1.71	2

Πίνακας 5.13 Αποτελέσματα τομέα υγείας - Test set

2. Τομέας Ενέργειας - XOM, COP, CVX:



Σχήμα 5.22 Τιμές μετοχών τομέα ενέργειας

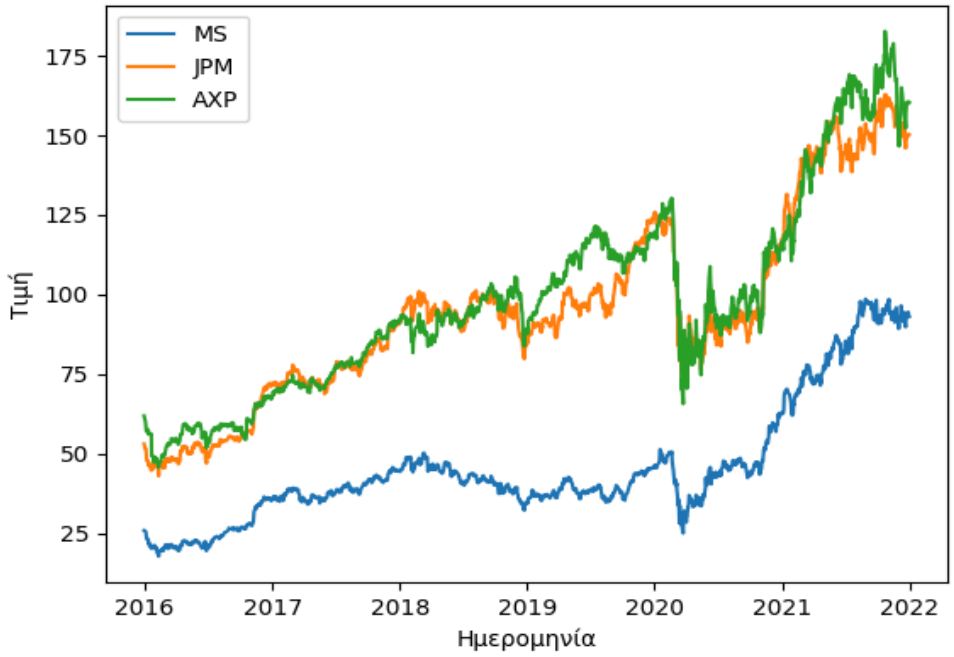
Train set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
COP	-3%	-	-0.02	-
CVX	19%	-	0.12	-
XOM	-32%	-	-0.28	-
Χαρτοφυλάκιο				
UBH	-5%	3	-0.04	4
EF	-18%	4	-0.05	5
FtL	-52%	5	-0.03	3
FtL (BUY-SELL)	-91%	6	-1.13	6
UCR	-4%	2	-0.01	2
M2 (DPC)	**38656%**	**1**	**7.9**	**1**

Πίνακας 5.14 Αποτελέσματα τομέα ενέργειας - Train set

Test set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
COP	89%	-	1.83	-
CVX	46%	-	1.56	-
XOM	57%	-	1.56	-
Χαρτοφυλάκιο				
UBH	64%	2	1.73	3
EF	46%	5	1.67	5
FtL	63%	4	1.72	4
FtL (BUY-SELL)	-20%	6	-0.51	6
UCR	63%	3	**1.85**	**1**
M2 (DPC)	**73%**	**1**	1.78	2

Πίνακας 5.15 Αποτελέσματα τομέα ενέργειας - Test set

3. Τομέας Οικονομικών - MS, JPM, AXP



Σχήμα 5.23 Τιμές μετοχών τομέα οικονομικών

Train set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
AXP	92%	-	0.44	-
JPM	125%	-	0.62	-
MS	148%	-	0.6	-
Χαρτοφυλάκιο				
UBH	122%	4	0.6	5
EF	111%	5	0.63	4
FtL	162%	2	0.77	2
FtL (BUY-SELL)	-75%	6	-0.6	6
UCR	131%	3	0.7	3
M2 (DPC)	**42383%**	**1**	**8.11**	**1**

Πίνακας 5.16 Αποτελέσματα τομέα οικονομικών - Train set

Test set				
Μετοχές	**ROI**		**Sharpe Ratio**	
	Τιμή	**Κατάταξη**	**Τιμή**	**Κατάταξη**
AXP	40%	-	1.27	-
JPM	29%	-	1.22	-
MS	47%	-	1.51	-
Χαρτοφυλάκιο				
UBH	39%	3	1.48	4
EF	46%	2	1.6	2
FtL	**80%**	**1**	**2.36**	**1**
FtL (BUY-SELL)	-7%	6	-0.16	6
UCR	38%	4	1.57	3
M2 (DPC)	37%	5	1.4	5

Πίνακας 5.17 Αποτελέσματα τομέα οικονομικών - Test set

Από τα παραπάνω είναι εμφανές πως ο πράκτορας δεν αποδίδει το ίδιο καλά σε όλα τα δεδομένα. Η αποδοτικότητα του πράκτορα σε μετοχές τεχνολογίας δεν μεταφράζεται στις μετοχές οικονομικών όπως γίνεται εμφανές από τον Πίνακα 5.15. Αυτό σημαίνει πως η αρχιτεκτονική που έχει επιλεχθεί καθώς και η βελτιστοποίηση υπερ-παραμέτρων που έγινε είναι ευαίσθητη σε αλλαγές στη φύση του χαρτοφυλακίου.

Παρ'όλα αυτά, όπως φαίνεται από τους Πίνακες 5.11 και 5.13, ο πράκτορας αποδίδει καλά σε δύο από τα τρία νέα χαρτοφυλάκια που δοκιμάστηκε, καταφέρνοντας να είναι μεταξύ των κορυφαίων μεθόδων σε αυτά. Το γεγονός αυτό αποδεικνύει πως ο βελτιστοποιημένος πράκτορας καταφέρνει και εντοπίζει μοτίβα στην αγορά τα οποία του επιτρέπουν σε πολλές περιπτώσεις να ξεπερνάει τις παραδοσιακές τεχνικές διαχείρισης χαρτοφυλακίου.

Κεφάλαιο 6. Συμπεράσματα και Προεκτάσεις

Στη παρούσα διπλωματική εργασία, μελετήθηκαν διάφοροι παράγοντες που επηρεάζουν την αποδοτικότητα ενός πράκτορα Ενισχυτικής Μάθησης στο πρόβλημα της βελτιστοποίησης χαρτοφυλακίου. Για το σκοπό αυτό, χρησιμοποιήθηκε ένα χαρτοφυλάκιο μετοχών τεχνολογίας (APPLE, GOOGLE, IBM) και επιλέχθηκε ο αλγόριθμος DDQN καθώς και διακριτός χώρος ενεργειών (BUY, SELL, HOLD). Αρχικά, έγινε μελέτη σχετικά με τις βέλτιστες τιμές των υπερ-παραμέτρων ενισχυτικής μάθησης, στη συνέχεια χρησιμοποιώντας τις τιμές αυτές μελετήθηκε η κατάσταση - είσοδος του συστήματος και πως επηρεάζει αυτή την απόδοση του. Έπειτα, μελετήθηκαν διαφορετικές αρχιτεκτονικές για το δίκτυο πρόβλεψης των Q τιμών χρησιμοποιώντας feed-forward δίκτυα αλλά και πιο σύνθετα που αξιοποιούν συνελικτικά στρώματα. Τέλος, επιλέχθηκαν νέα χαρτοφυλάκια, από διαφορετικούς τομείς της οικονομίας, ώστε να επαληθευθούν τα αρχικά αποτελέσματα για τον βελτιστοποιημένο μοντέλο και να ελεγχθεί η ικανότητα του στο να γενικεύει σε νέα δεδομένα με διαφορετικά χαρακτηριστικά.

Καθόλη τη διάρκεια της πειραματικής διαδικασίας, έγινε προφανές πως ο πράκτορας είναι ιδιαίτερα ικανός στην εκμάθηση των δεδομένων εκπαίδευσης πετυχαίνοντας πολύ υψηλά ποσοστά επιστροφών σε αυτά. Το φαινόμενο αυτό συνεχίστηκε παρόλες τις προσπάθειες προσθήκης πολυπλοκότητας ή μείωσης τόσο της συχνότητας εκπαίδευσης όσο και το πλήθος των δειγμάτων και ο πράκτορας, από τα πρώτα κιολας επεισόδια ήταν σε θέση να πετυχαίνει ιδιαίτερα υψηλές αποδόσεις. Σε πολλές περιπτώσεις όμως, όπως στα μοντέλα M1 και M3, η μεγάλη επιτυχία στα δεδομένα εκπαίδευσης δεν μεταφράστηκε σε παρόμοια απόδοση στα άγνωστα δεδομένα υποδεικνύοντας πως τα μοντέλα αυτά στερούνταν την ικανότητα γενίκευσης και τα μοτίβα που είχαν αναγνωρίσει κατά τη διαδικασία εκπαίδευσης δεν απέδιδαν στα νέα δεδομένα. Στα μοντέλα M2 και M4, από την άλλη πλευρά, ο πράκτορας στάθηκε ικανός να μεταφέρει την εμπειρία του από την εκπαίδευση και στα άγνωστα δεδομένα πετυχαίνοντας επιδόσεις καλύτερες από όλες τις παραδοσιακές τεχνικές διαχείρισης χαρτοφυλακίου που χρησιμοποιήθηκαν ως αναφορά καταδεικνύοντας πως υπό συνθήκες, και με προσεκτική βελτιστοποίηση, η Ενισχυτική Μάθηση μπορεί να αποτελέσει αξιόπιστο εργαλείο διαχείρισης χαρτοφυλακίων μετοχών.

Τα προβλήματα που εντοπίστηκαν κατά την διαδικασία εκπαίδευσης αφορούν κυρίως την ευαισθησία του πράκτορα στις μεγάλες διακυμάνσεις της αγοράς καθώς και στην τυχαιότητα που πολλές φορές υπάρχει, οθώντας τα μοντέλα να εντοπίζουν, κατά την εκπαίδευση, μοτίβα τα οποία δεν μπορούν να μεταφερθούν σε άγνωστα δεδομένα. Επιπλέον, από τα πειράματα στα νέα χαρτοφυλάκια έγινε φανερό πως η βελτιστοποίηση σε ένα συγκεκριμένο χαρτοφυλάκιο δεν αποτελεί εγγύηση για καλή απόδοση σε οποιοδήποτε άλλο χαρτοφυλάκιο γεγονός που δείχνει πως ο πράκτορας θα πρέπει να βελτιστοποιείται ανάλογα με τα χαρακτηριστικά του χαρτοφυλακίου που καλείται να διαχειριστεί. Παρόλα αυτά, ο πράκτορας καταφέρνει να επιτύχει καλή επίδοση σε δύο από τα τρία νέα χαρτοφυλάκια το οποίο σημαίνει ότι η είσοδος-κατάσταση

και η αρχιτεκτονική που επιλέχθηκε στα παραπάνω πειράματα αποτελούν καλή λύση για το πρόβλημα της διαχείρισης χαρτοφυλακίου.

Η ελλιπής δυνατότητα γενίκευσης που παρατηρήθηκε σε κάποια από τα μοντέλα καταδεικνύει πως ο πράκτορας θα επωφελούνταν από προσθήκη πολυπλοκότητας. Ένας τρόπος για να επιτευχθεί αυτό, είναι η μεταφορά του προβλήματος σε συνεχή χώρο ενεργειών όπου ο πράκτορας πλέον δεν μοιράζει το διαθέσιμο κεφάλαιο ισόποσα σε όλες τις μετοχές του χαρτοφυλακίου αλλα επιλέγει την ποσοστιαία κατανομή του σε κάθε μία. Η προσέγγιση αυτή αυξάνει σημαντικά την πολυπλοκότητα του προβλήματος, ωθώντας όμως τον πράκτορα να αναζητά ενέργειες οι οποίες μειώνουν το ρίσκο και αυξάνουν την απόδοση. Η μεταφορά του προβλήματος σε συνεχή χώρο ενεργειών θα σήμανε πλέον, την ανάγκη αξιοποίησης άλλων αλγορίθμων Ενισχυτικής Μάθησης καθώς ο αλγόριθμος DDQN που χρησιμοποιήθηκε στην παρούσα διπλωματική έχει τον περιορισμό ότι απευθύνεται αποκλειστικά σε προβλήματα με διακριτό χώρο ενεργειών. Οι αλγόριθμοι DDPG και PPO αποτελούν ιδανικές λύσεις για το πρόβλημα συνεχούς χώρου και ειδικά ο αλγόριθμος PPO όπου εμπεριέχει μέσα την έννοια της κοντινής βελτιστοποίησης μπορεί να αποτελέσει καλή λύση στο πρόβλημα αυτό αφού συχνά οι μεγάλες μεταβολές σε μια επενδυτική στρατηγική δημιουργούν και μεγάλα κόστη συναλλαγών.

Μια ακόμα προσέγγιση στο πρόβλημα της μειωμένης ικανότητας γενίκευσης είναι η προσθήκη μεγαλύτερου όγκου πληροφορίας ως είσοδο στο μοντέλο. Από τα πειράματα (μοντέλο M4) έγινε φανερό πως, η χρήση συνελικτικών νευρωνικών δικτύων στο συγκεκριμένο πρόβλημα έχει ελπιδοφόρα αποτελέσματα αφού από τις μετρικές αξιολόγησης φάνηκε πως το M4 υπερέχει έναντι των παραδοσιακών τεχνικών διαχείρισης χαρτοφυλακίου. Τα συνελικτικά δίκτυα όπως εξηγήθηκε και παραπάνω είναι ιδιαίτερα ικανά στην εξαγωγή χαρακτηριστικών από μεγάλους όγκους δεδομένων συνεπώς μια προσέγγιση που μπορεί να εξεταστεί στο συγκεκριμένο πρόβλημα θα ήταν η χρήση γραφημάτων με την απόδοση της μετοχής σε ένα δεδομένο χρονικό παράθυρο ως είσοδο στο μοντέλο. Εναλλακτικά, καθώς τα συνελικτικά νευρωνικά δίκτυα είναι ιδιαίτερα ικανά στο να διαχειρίζονται δεδομένα σε ακολουθία όπως κείμενο, η χρήση έγκυρων δημοσιογραφικών πηγών ή δημοσιεύσεις από κοινωνικά δίκτυα ως είσοδο στα μοντέλο μπορεί να προσφέρει μεγαλύτερη κατανόηση των συνθηκών της αγοράς και καλύτερες αποφάσεις. Ιδιαίτερα ικανές, επίσης, στην ανάλυση δεδομένων σε ακολουθία είναι και άλλες αρχιτεκτονικές όπως για παράδειγμα τα Δίκτυα Μακροχρόνιας Μνήμης (LSTM) οι οποίες μπορούν μέσω της πολυπλοκότητας που προσφέρουν να φέρουν σημαντικά αποτελέσματα.

Άρα συνοψίζοντας::

- Η Ενισχυτική Μάθηση μπορεί να αποτελέσει, υπό προϋποθέσεις, χρήσιμο εργαλείο στον τομέα της διαχείρισης χαρτοφυλακίου. Στα παραπάνω πειράματα διαπιστώθηκε πως με σωστή βελτιστοποίηση των υπερπαραμέτρων εκπαίδευσης και της κατάστασης-εισόδου είναι εφικτό ένα μοντέλο να πετύχει επιδόσεις καλύτερες από πολλές παραδοσιακές τεχνικές (M2 και M4).

- Η μεγάλη μεταβλητότητα και τυχαιότητα στα δεδομένα εκπαίδευσης, κάνει τη διαδικασία εκπαίδευσης ιδιαίτερα απαιτητική κάτι το οποίο οδηγεί σε υποβέλτιστα μοντέλα τα οποία αποδίδουν χειρότερα από παραδοσιακές τεχνικές διαχείρισης χαρτοφυλακίου (M1 και M3).
- Η χρήση ενός πράκτορα Ενισχυτικής Μάθησης που έχει βελτιστοποιηθεί σε ένα συγκεκριμένο χαρτοφυλάκιο για την διαχείριση ενός νέου με διαφορετικά χαρακτηριστικά, δεν εγγυάται καλά αποτελέσματα σε όλες τις περιπτώσεις.
- Όπως και στα προβλήματα Μηχανικής Μάθησης, η προσεκτική επεξεργασία των δεδομένων και επιλογή αρχιτεκτονικών στα μοντέλα έχει σημαντική επίδραση στις επιδόσεις τους.
- Τα συνελικτικά νευρωνικά δίκτυα που χρησιμοποιήθηκαν στο Μοντέλο 4 δείχνουν ελπιδοφόρα αποτελέσματα στον χώρο της διαχείρισης χαρτοφυλακίου και η χρήση αυτών σε συνδυασμό με δεδομένα εικόνας ή κειμένου μπορούν να προσφέρουν στον πράκτορα πολύτιμη πληροφορία για τη λήψη βέλτιστων αποφάσεων.

Λαμβάνοντας υπόψη όλα τα παραπάνω κάποια επόμενα βήματα στη μελέτη θα μπορούσαν να είναι τα εξής:

- Μεταφορά του προβλήματος σε συνεχή χώρο ενεργειών και χρήση κατάλληλων αλγορίθμων Ενισχυτικής Μάθησης όπως DDPG και PPO.
- Διερεύνηση και βελτιστοποίηση της συνάρτησης ανταμοιβής, όπως για παράδειγμα ένταξη του ρίσκου σε αυτή.
- Ενσωμάτωση μεγαλύτερου όγκου πληροφορίας όπως εικόνες ή κείμενο τα οποία περιέχουν πληροφορίες για την αγορά.
- Μελέτη διαφορετικών αρχιτεκτονικών όπως LSTM οι οποίες είναι ιδανικές σε δεδομένα χρονοσειράς.

Βιβλιογραφία

1. O'Hara, M. (2003). Presidential Address: Liquidity and Price Discovery. The Journal of Finance, 58(4), 1335-1354
2. Neal, L. (1990). The Rise of Financial Capitalism: International Capital Markets in the Age of Reason. Cambridge University Press
3. NYSE, (n.d.). History of The New York Stock Exchange
4. Madhavan, A. (2002). Market Microstructure: A Survey. Journal of Financial Markets, 3(3), 205-258
5. Barber, B. M., & Odean, T. (2013). The Behavior of Individual Investors. Handbook of the Economics of Finance, 2, 1533-1570
6. Graham, B., & Dodd, D. (2009). Security Analysis: The Classic 1940 Edition. McGraw-Hill Education
7. Damodaran, A. (2012). Investment Valuation: Tools and Techniques for Determining the Value of any Asset, Wiley
8. Murphy, J.J. (1999). Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance
9. Chan, E.P. (2013). Quantitative Trading: How to Build Your Own Algorithmic Trading Business. Wiley
10. Sharpe, W.F. (1994). "The Sharpe Ratio." Journal of Portfolio Management, 21(1), 49–58
11. Sortino, F.A., & Price, L.N. (1994). "Performance measurement in a downside risk framework." The Journal of Investing, 3(3), 59-64
12. Investopedia: "Maximum Drawdown (MDD)."
13. Markowitz, H. (1952). Portfolio Selection. The Journal of Finance, 7(1), 77-91.
14. Black, F., & Litterman, R. (1992). Global Portfolio Optimization. Financial Analysts Journal, 48(5), 28-43.
15. Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. The Review of Economics and Statistics, 47(1), 13-37.
16. Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. The Journal of Finance, 19(3), 425-442.
17. Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers & Operations Research, 32(10), 2513-2522.
18. Cao, L., Tay, F. E. (2001). Financial forecasting using support vector machines. Neural Computing & Applications, 10, 184-192.
19. Kumar, D., & Thenmozhi, M. (2016). Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets. Pacific-Basin Finance Journal, 40, 122-139.
20. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

21. Dixon, M., Klabjan, D., & Bang, J. H. (2016). Classification-based financial markets prediction using deep neural networks. Algorithms, 9(3), 41.
22. Heaton, J., Polson, N., & Witte, J. (2017). Deep learning in finance. arXiv preprint arXiv:1602.06561.
23. Sezer, O. B., & Ozbayoglu, A. M. (2018). Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. Applied Soft Computing, 70, 525-538.
24. Moody, J., and Saffell, M. (2001). Learning to trade via direct reinforcement. IEEE Transactions on Neural Networks, 12(4), 875-889.
25. Jiang, Z., Liang, J., Li, D., and Wu, Z. (2017). Deep reinforcement learning with modulated policy for multi-objective portfolio optimization. Proceedings of the Genetic and Evolutionary Computation Conference Companion - GECCO ’17, 1133–1140.
26. Deng, Y., Bao, F., Kong, Y., Ren, Z., and Dai, Q. (2019). Deep Direct Reinforcement Learning for Financial Signal Representation and Trading. IEEE transactions on neural networks and learning systems, 30(3), 653-668.
27. Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 2nd edition. O’Reilly Media, 2019
28. Bishop, Christopher M. Pattern Recognition and Machine Learning. Springer, 2006
29. James, Gareth, et al. An Introduction to Statistical Learning: with Applications in R. Springer, 2013
30. P. Bruce and A. Bruce, Practical Statistics for Data Scientists. O'Reilly, 2017
31. Hosmer Jr, David W., Stanley Lemeshow, and Rodney X. Sturdivant. Applied logistic regression. Vol. 398. John Wiley & Sons, 2013
32. Cortes, Corinna, and Vladimir Vapnik. Support-vector networks. Machine learning 20.3 (1995): 273-297
33. Quinlan, J. R. C4. 5: programs for machine learning. Elsevier, 2014
34. Breiman, Leo. Random forests. Machine learning 45.1 (2001): 5-32
35. Friedman, Jerome H. Greedy function approximation: a gradient boosting machine. Annals of statistics (2001): 1189-1232
36. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press
37. Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (pp. 249-256)
38. Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford university press.
39. Liang, Z. et al. (2018) Adversarial Deep Reinforcement Learning in Portfolio Management, arXiv.org
40. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural Computation, vol. 1, no. 4, pp. 541–551, 1989

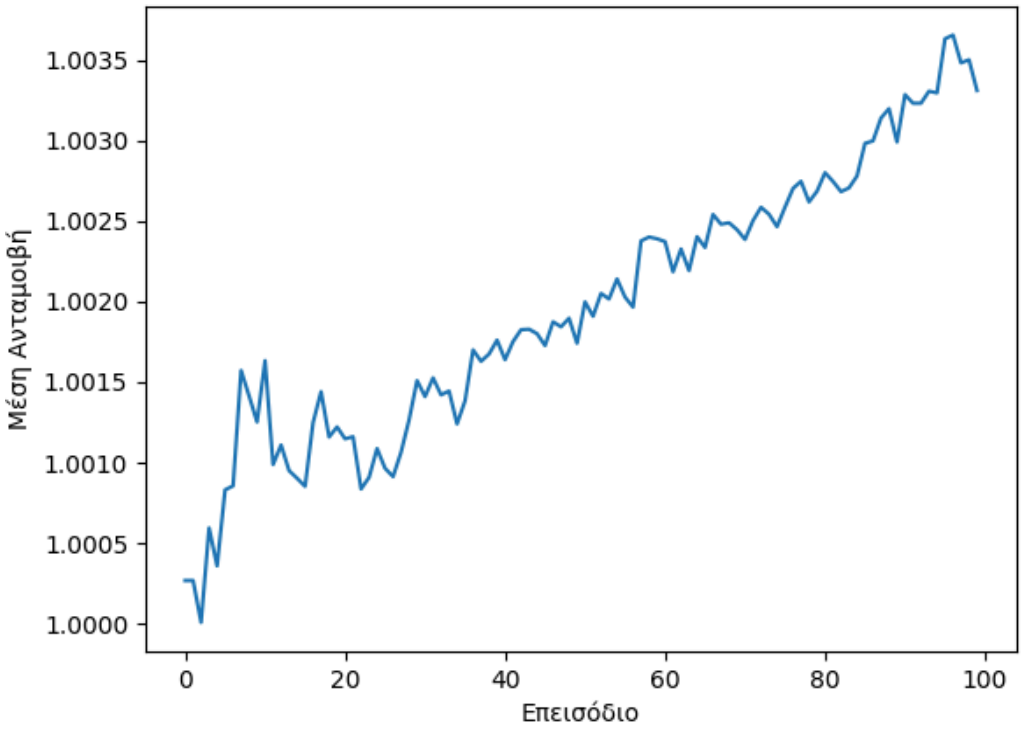
41. O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458
42. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324
43. Elman, J. L. (1990). Finding structure in time. Cognitive Science, 14(2), 179–211
44. Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. science, 313(5786), 504-507
45. Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. arXiv preprint arXiv:1312.6114
46. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680)
47. Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA, USA
48. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., 2015. Human-level control through deep reinforcement learning. Nature 518, 529–533
49. Abbeel, P., Ng, A.Y., 2004. Apprenticeship learning via inverse reinforcement learning. In: Proceedings of the twenty-first international conference on Machine learning. ACM, New York, NY, USA, p. 1
50. Lattimore, T., & Szepesvári, C. (2020). Bandit Algorithms. Cambridge University Press.
51. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., ... & Wierstra, D. (2016). Continuous control with deep reinforcement learning
52. Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529-533.

Παράρτημα

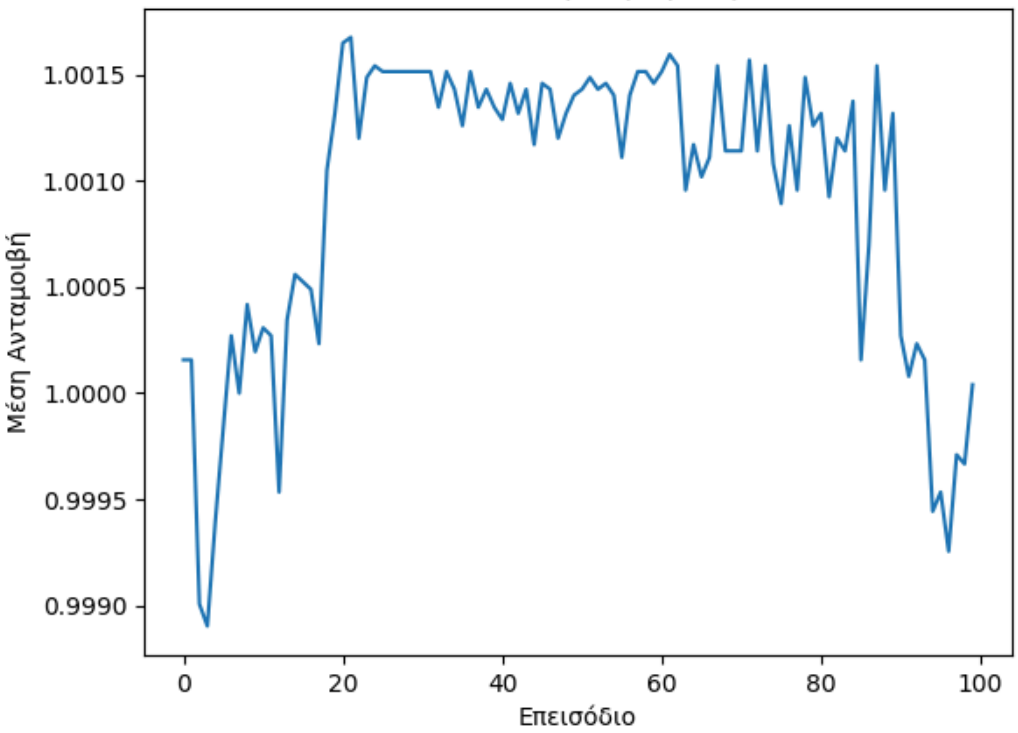
Πίνακες

No	Συχνότητα εκπαίδευσης (Μέρες)	Συχνότητα ενημέρωσης Δικτύου Στόχου (Εκπαιδεύσεις)	Μέγεθος Μνήμης Εμπειριών	Παράγοντας μείωσης epsilon (e-greedy)	Παράγοντας Έκπτωσης (γ)	ROI (%)	Sharpe
1	90	1900	4000	0.99	0.99	502	0.08
2	50	1900	4000	0.99	0.99	900	0.13
3	20	1900	4000	0.99	0.99	854	0.12
4	20	500	4000	0.99	0.99	1238	0.12
5	50	500	4000	0.99	0.99	629	0.11
6	90	500	4000	0.99	0.99	872	0.11
7	90	1200	4000	0.99	0.99	424	0.09
8	50	1200	4000	0.99	0.99	671	0.12
9	20	1200	4000	0.99	0.99	**2478**	**0.16**
10	20	1200	4000	0.98	0.99	1618	**0.16**
11	50	1200	4000	0.98	0.99	787	0.11
12	90	1200	4000	0.98	0.99	306	0.09
13	20	500	4000	0.98	0.99	613	0.1
14	50	500	4000	0.98	0.99	1217	0.13
15	90	500	4000	0.98	0.99	287	0.08
16	20	1200	8000	0.99	0.99	2145	0.15
17	20	1200	12000	0.99	0.99	1213	0.13

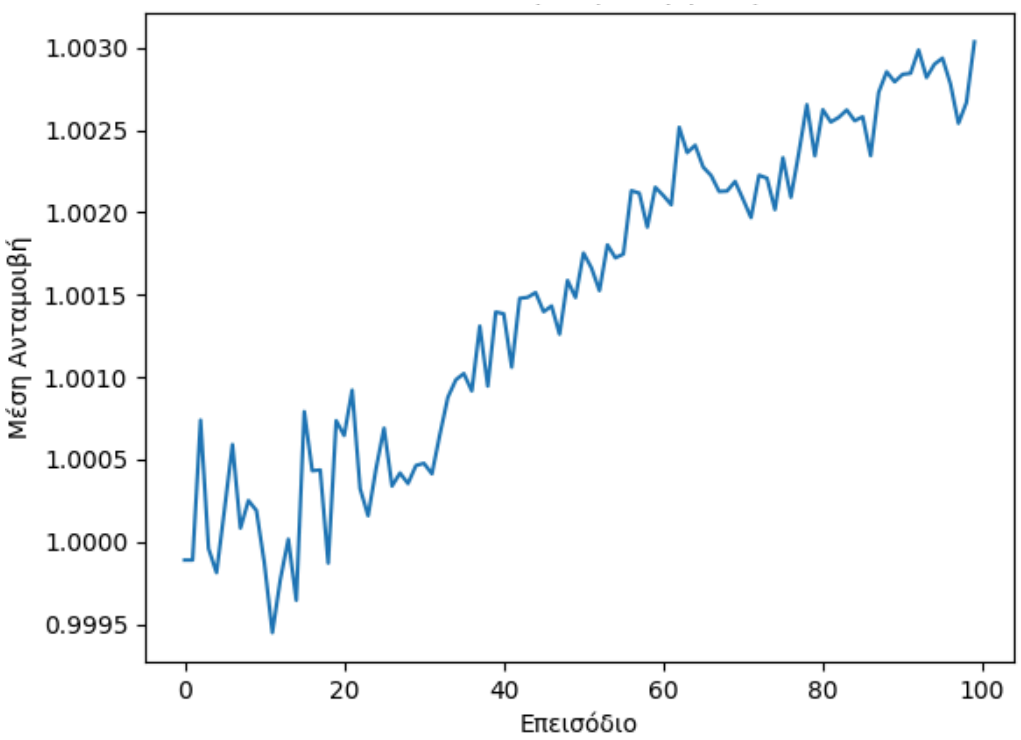
Πίνακας 5.1 Αποτελέσματα πειραμάτων βελτιστοποίησης υπερ-παραμέτρων



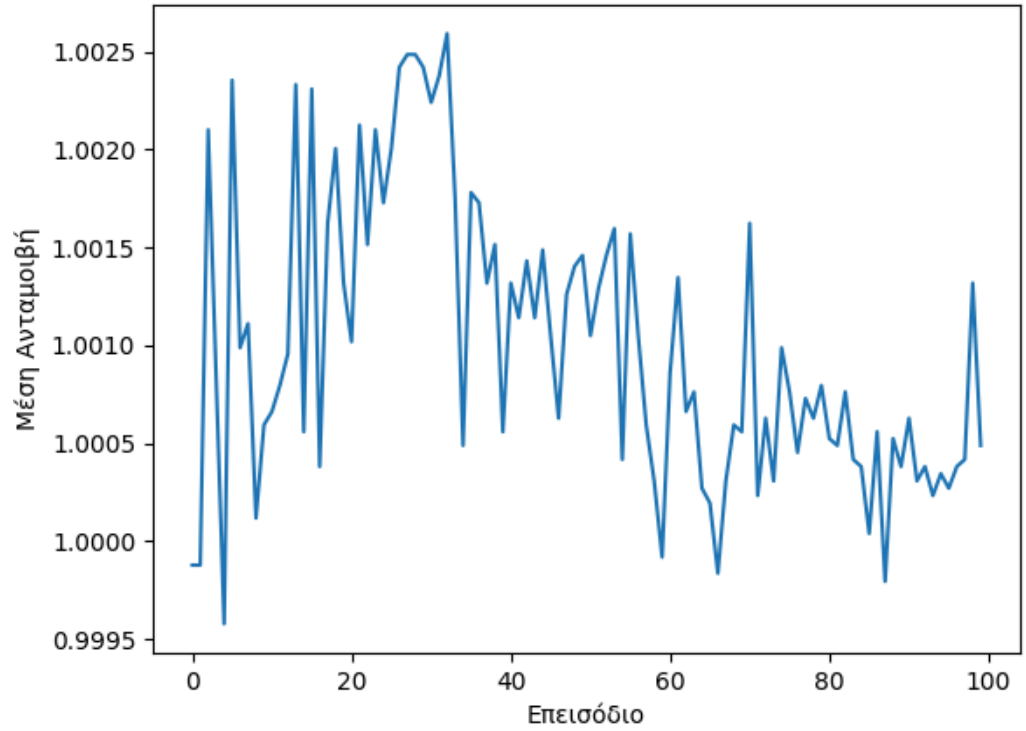
Σχήμα 5.10 M2 - Τομέας Υγείας Train set



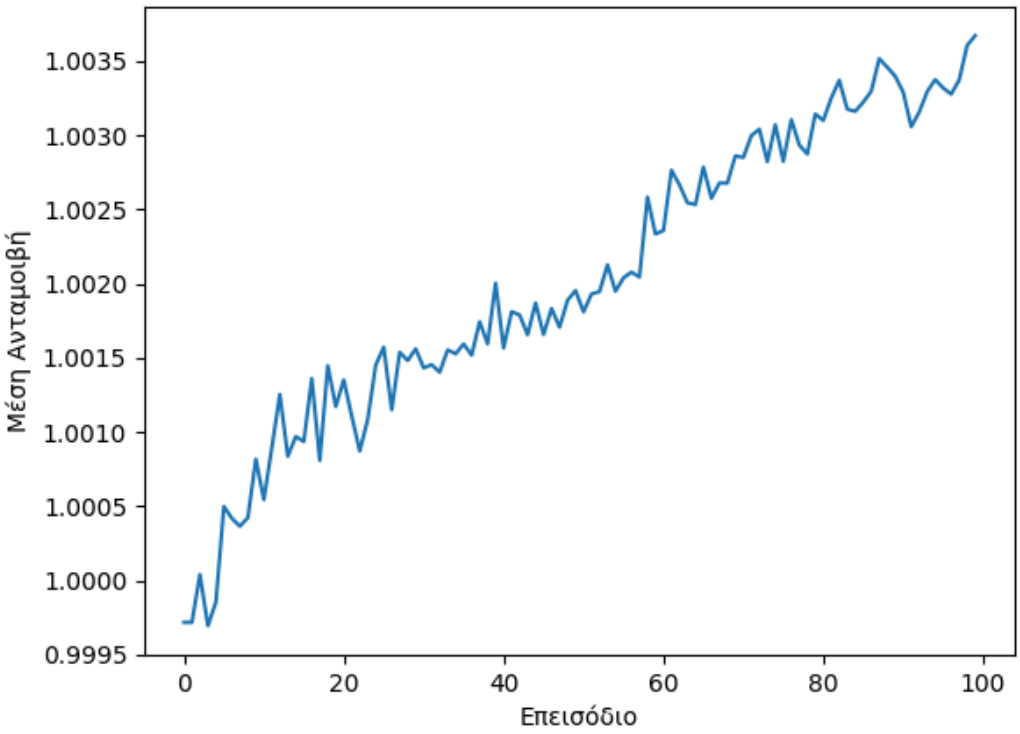
Σχήμα 5.10 M2 - Τομέας Υγείας Test set



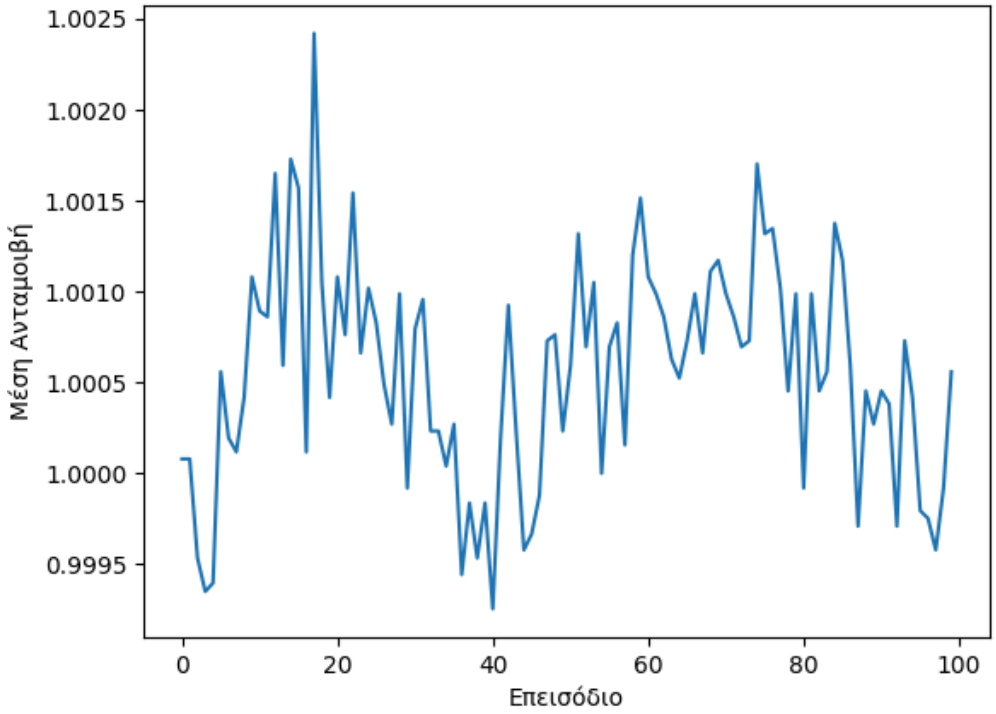
Σχήμα 5.10 M2 - Τομέας Ενέργειας Train set



Σχήμα 5.10 M2 - Τομέας Ενέργειας Test set



Σχήμα 5.10 M2 - Τομέας Οικονομικών Train set



Σχήμα 5.10 M2 - Τομέας Οικονομικών Train set