



Πολυτεχνείο Κρήτης

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Υπολογιστική πλατφόρμα βαθιάς μάθησης για
νευρωνικά δίκτυα καθοδηγούμενα από εξισώσεις
φυσικής

Διπλωματική Εργασία

του

Χρήστου Κυριάκου

Επιβλέπων: Γεώργιος Καρυστινός
Βασίλειος Σαμολαδάς
Γεώργιος Σταυρουλάκης

30 Αυγούστου 2023

Περίληψη

Η παρούσα διατριβή παρουσιάζει μια συγκριτική μελέτη των νευρωνικών δικτύων που βασίζονται σε φυσικούς νόμους (PINNs) και των απλών νευρωνικών δικτύων για την πρόβλεψη της θέσης ενός αρμονικού ταλαντωτή. Τα PINNs είναι ένα είδος νευρωνικού δικτύου που ενσωματώνει φυσικούς νόμους και περιορισμούς ως μέρος της διαδικασίας εκπαίδευσης, επιτρέποντας την ενσωμάτωση της φυσικής γνώσης που σχετίζεται με το πρόβλημα το οποίο επιλύεται (χαρακτηριστική διαφορική εξίσωση). Τα απλά νευρωνικά δίκτυα, από την άλλη πλευρά, δεν ενσωματώνουν ρητά τέτοια γνώση και βασίζονται αποκλειστικά στα δεδομένα για την εκπαίδευση.

Πραγματοποιήθηκαν δύο πειράματα για τη σύγκριση της απόδοσης αυτών των δικτύων. Το πρώτο πείραμα αφορούσε την πρόβλεψη της θέσης ενός μόνο αρμονικού ταλαντωτή δεδομένου ενός περιορισμένου σει παρατηρήσεων σε συγκεκριμένα χρονικά σημεία. Το δεύτερο πείραμα αφορούσε ένα πρόβλημα πολλαπλών διαστάσεων και είχε ως στόχο την πρόβλεψη της θέσης τεσσάρων διαφορετικών αρμονικών ταλαντωτών με διαφορετικά χαρακτηριστικά.

Τα αποτελέσματα έδειξαν ότι τα PINNs υπερτερούν των απλών νευρωνικών δικτύων σε απόδοση και ακρίβεια πρόβλεψης δεδομένων. Η χρήση των PINNs βελτιώνει την απόδοση του μοντέλου λόγω της ενσωμάτωσης περιορισμών και γνώσης σχετικά με το πρόβλημα της ταλάντωσης.

Συνολικά, η παρούσα διατριβή αποδεικνύει ότι η χρήση των PINNs είναι ένα αποδοτικό και ακριβές εργαλείο για την πρόβλεψη της θέσης του αρμονικού ταλαντωτή και άλλων σχετικών προβλημάτων.

Abstract

This thesis presents a comparative study of physics-informed neural networks (PINNs) and simple neural networks for the prediction of the position of a harmonic oscillator. PINNs are a type of neural network that incorporates physical laws and constraints as part of the training process, allowing for the incorporation of domain-specific knowledge into the model. Simple neural networks, on the other hand, do not explicitly incorporate such knowledge and rely on data alone for training.

Two experiments were conducted to compare the performance of these networks. The first experiment involved predicting the position of a single oscillator given a limited set of observed positions at specific timestamps. For this experiment, a simple feedforward neural network and a PINN were trained using data generated by solving the underlying differential equation governing the motion of the oscillator.

In the second experiment, a multidimensional problem was considered, involving the prediction of the positions of four oscillators. A PINN was again trained on data generated by solving the differential equations governing the oscillators' motion, while a feedforward neural network was trained on the raw input data.

The results of these experiments showed that PINNs outperformed simple neural networks in both cases, achieving higher accuracy and lower mean squared error. Furthermore, the PINN was able to learn the underlying physical laws governing the motion of the oscillators, while the simple neural network was not. This indicates that PINNs are more suitable for learning from limited data and can provide more accurate predictions in cases where physical constraints and domain-specific knowledge are important.

Overall, this study contributes to the understanding of the capabilities of PINNs and simple neural networks for the prediction of complex systems. The results suggest that PINNs are a promising approach for predicting the behavior of nonlinear dynamical systems and may have practical applications in fields such as physics and engineering.

Ευχαριστίες

Από αυτήν την θέση η οποία μου δίνετε, θα ήθελα να ευχαριστήσω θερμά μέσα από την καρδιά μου τους καθηγητές μου κ. Γεώργιο Καρυτινό, κ. Βασίλειο Σαμολαδά και κ. Γεώργιο Σταυρουλάκη για την εμπιστοσύνη που μου έδειξαν αναθέτοντάς μου το εν λόγω θέμα. Επίσης θα ήθελα να τους ευχαριστήσω για την συνεργασία μας καθ'όλη την διάρκεια της εργασίας μου καθώς και για τις γνώσεις που μου μετέδωσαν όλα αυτά τα χρόνια όντας φοιτητής υπό την επίβλεψή τους. Χάρη στην εργασία αυτή εισήλθα στον πολύ ενδιαφέρον χώρο των Νευρωνικών Δικτύων και της Μηχανικής Μάθησης. Εν συνεχεία θα ήθελα να ευχαριστήσω την οικογένειά μου, τους γονείς μου Κυριάκο και Ευαγγελία καθώς και τα αδέρφια μου Κωνσταντίνο και Έλενα για την στήριξη και την απέραντη αγάπη τους όλα αυτά τα χρόνια. Η περίοδος των σπουδών μου δεν θα ήταν το ίδιο ευχάριστη και δημιουργική αν δεν είχα δίπλα μου τους φίλους μου Νεκτάριο, Στέλιο, Ρολάντο, Αλέξανδρο και πολλούς άλλους τους οποίους αγαπώ και ευχαριστώ θερμά. Μια από τις πιο όμορφες εμπειρίες της ζωής μου έχει φτάσει αισίως στο τέλος της, γνωρίζοντας ότι μετά από εδώ θα ανοίξουν καινούριοι ορίζοντες στην ζωή μου. Η φοίτησή μου στο Πολυτεχνείο Κρήτης θα μείνει για πάντα χαραγμένη στην καρδιά και στην μνήμη μου. Ευχαριστώ από καρδιάς όλους τους συνοδιπόρους μου σ'αυτή την πανέμορφη διαδρομή.

Περιεχόμενα

1 Μηχανική Μάθηση	4
1.0.1 Εισαγωγή στη Μηχανική Μάθηση	4
2 Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ)	8
2.0.1 Μηχανική Μάθηση και Νευρωνικά Δίκτυα	8
2.0.2 Εισαγωγή στα Νευρωνικά Δίκτυα	9
2.0.3 Ιστορική Αναδρομή	10
2.0.4 Βιολογικά Νευρωνικά Δίκτυα	12
2.0.5 Το Μοντέλο του Τεχνητού Νευρώνα	14
2.0.6 Συνάρτηση Ενεργοποίησης	15
2.0.7 Αρχιτεκτονική Νευρωνικών Δικτύων	20
2.0.8 Διαδικασίες Μάθησης των ΤΝΔ	22
2.0.9 Εφαρμογές των Νευρωνικών Δικτύων	24
3 Physics Informed Neural Networks (PINNs)	27
3.0.1 Εισαγωγή στα PINN	27
3.0.2 Ερευνητική δουλειά στα PINN	29
3.0.3 Μοντελοποίηση και Υπολογισμός	30
4 Ορισμός προβλήματος του Αποσβενόμενου Αρμονικού Ταλαντωτή	33
4.0.1 Αποσβενόμενος Αρμονικός Ταλαντωτής	33
4.0.2 Η εξίσωση της κίνησης	34
4.0.3 Ο Εκθετικός Ansatz	35
4.0.4 Υπερκρίσιμη απόσβεση	36
4.0.5 Περίπτωση με κρίσιμη απόσβεση	38
4.0.6 Υποκρίσιμη απόσβεση	39
5 Επίλυση διαφορικής εξίσωσης Αποσβενόμενου Αρμονικού Ταλαντωτή με χρήση PINNs	43
5.0.1 Μονοβάθμιο πρόβλημα αρμονικού ταλαντωτή	46
5.0.2 Πολυβάθμιο πρόβλημα αρμονικού ταλαντωτή	53

6 Μελλοντική εργασία	61
Α΄ Ακρωνύμια και συντομογραφίες	62

Κεφάλαιο 1

Μηχανική Μάθηση

1.0.1 Εισαγωγή στη Μηχανική Μάθηση

Από τότε που εφευρέθηκαν οι ηλεκτρονικοί υπολογιστές οι άνθρωποι αναρωτιούνται, πότε θα είναι ικανοί να μαθαίνουν λειτουργίες και να αποκτήσουν δεξιότητες χρήσιμες για την καθημερινή ζωή των ανθρώπων. Ρομπότ που μιλάνε και κάνουν δουλειές του νοικοκυριού, αυτοκίνητα που οδηγούν μόνα τους κ.α. είναι μερικά από τα καινοτόμα μηχανήματα που πολλοί έχουν οραματιστεί. Με κίνητρο τέτοια επιστημονικά οράματα άρχισε σιγά σιγά να αναπτύσσεται ένας νέος επιστημονικός κλάδος: αυτός της Μηχανικής Μάθησης (Machine Learning). Οι επιστήμονες συνειδητοποίησαν ότι η εξέλιξη της τεχνολογίας και πολλές εκφάνσεις της καθημερινότητας των ανθρώπων θα έμπαιναν σε μία ραγδαία τροχιά, εάν μπορούσαμε να προγραμματίσουμε τα μηχανήματα να μαθαίνουν αυτοβούλως και μέσα από την εμπειρία. Έτσι, οι ερευνητές με γνώμονα την πρόοδο της ανθρωπότητας οραματιστήκαν την κατασκευή υπολογιστών, που θα μαθαίνουν από τα ιατρικά αρχεία ποιά θεραπεία είναι πιο αποτελεσματική για νέες ασθένειες ή ακόμα και σπιτιών που θα μπορούν εμπειρικά να βελτιστοποιούν το ενεργειακό τους κόστος, βασισμένα στις ανάγκες του κάθε ιδιοκτήτη. Μέσα λοιπόν από την επιτυχημένη εκπαίδευση των υπολογιστών θα γεννηθούν νέες χρήσεις τους, δίνοντας λύσεις σε διάφορα ζητήματα, που αντιμετωπίζει ο άνθρωπος στη σύγχρονη εποχή. Επιπρόσθετα, η λεπτομερής κατανόηση των πληροφοριών των αλγορίθμων επεξεργασίας για τη Μηχανική Μάθηση, ίσως οδηγήσουν στην εμβάθυνση και κατανόηση των ανθρώπινων ικανοτήτων μάθησης σε όλα τα ηλικιακά στάδια. Με την εξέλιξη της Μηχανικής Μάθησης ως επιστημονικού κλάδου της αποδόθηκαν ποικίλοι ορισμοί με βάση την οπτική που σημαντικοί επιστήμονες ανέπτυξαν με τη διαχρονική ενασχόληση πάνω σε αυτήν. Έτσι, γενικά αναφέρουμε ότι: Η Μηχανική Μάθηση είναι ένα πεδίο της Επιστήμης των Υπολογιστών που αναπτύχθηκε μέσα από την ενασχόληση και τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής μάθησης στην Τεχνητή Νοημοσύνη. Πραγματεύεται δηλαδή μεθόδους και τεχνικές που επιτρέπουν στους υ-

πολογιστές να 'μαθαίνουν'. Το 1959 η Μηχανική Μάθηση ορίζεται από τον Arthur Samuel ως το πεδίο μελέτης που δίνει στους υπολογιστές τη δυνατότητα να μαθαίνουν χωρίς να έχουν προγραμματιστεί ρητά. Η Μηχανική Μάθηση ορίζεται ουσιαστικά ως ένα σύνολο μεθόδων, οι οποίες μπορούν αυτόματα να ανιχνεύουν μοτίβα σε δεδομένα και στη συνέχεια να χρησιμοποιούν τα άγνωστα μοτίβα για να προβλέπουν μελλοντικά δεδομένα. Εξετάζει με πιο απλά λόγια τη μελέτη και κατασκευή αλγορίθμων, που μαθαίνουν μέσα από μια βάση δεδομένων και μπορούν να κάνουν προβλέψεις σχετικά με αυτά. Επίσης, η Μηχανική Μάθηση είναι στενά συνυφασμένη με την Υπολογιστική Στατιστική, έναν επιστημονικό κλάδο, που ομοίως χρησιμοποιεί τους υπολογιστές για να κάνει προβλέψεις. Ακόμα, συνεργάζεται στενά με τη Μαθηματική Βελτιστοποίηση, η οποία παρέχει τις μεθόδους, τη θεωρία και τους τομείς εφαρμογής. Ουσιαστικά, η χρήση της Μηχανικής Μάθησης αποκτά νόημα μέσα από την αδυναμία τόσο του ρητού προγραμματισμού των αλγορίθμων να υλοποιήσουν μια σειρά από υπολογιστικές εργασίες, όσο και της αντίστοιχης αδυναμίας που παρουσιάζει ο σχεδιασμός των εργασιών αυτών. Παραδείγματα τέτοιων εργασιών αποτελούν: οι μηχανές αναζήτησης, τα φίλτρα (spam filtering), η υπολογιστική όραση και η οπτική αναγνώριση χαρακτήρων (Pattern Recognition). Εργασίες όπου η Μηχανική Μάθηση αναλαμβάνει να επινοήσει πολύπλοκα μοντέλα και αλγορίθμους μέσα από τους οποίους, ερευνητές, αναλυτές δεδομένων καθώς και μηχανικοί θα οδηγούνται στην αξιόπιστη πρόβλεψη. Μέσα στην πληθώρα επιστημόνων, οι οποίοι ασχολήθηκαν με την Επιστήμη των Υπολογιστών και ιδιαίτερα της Μηχανικής Μάθησης ο Tom M. Mitchell πρότεινε έναν πιο επίσημο ορισμό για τη Μηχανική Μάθηση σύμφωνα με τον οποίο: "Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από την εμπειρία E ως προς μια κλάση εργασιών T και ένα μέτρο επίδοσης P , αν η επίδοση του σε εργασίες της κλάσης T , όπως αποτιμάται από το μέτρο P , βελτιώνεται με την εμπειρία E ". Με αυτόν τον ορισμό ο Mitchell δημιουργεί τις βάσεις για τον καθορισμό της Μηχανικής Μάθησης σε βασικό λειτουργικό πλαίσιο χωρίς τη χρήση γνωστικών όρων, βαδίζοντας στα χνάρια του Alan Turing στο επιστημονικό του έργο "Υπολογιστικές Μηχανές και Νοημοσύνη", όπου το ερώτημα αν μπορούν οι μηχανές να σκεφτούν, αντικαθίσταται με το ερώτημα αν μπορούν οι μηχανές να κάνουν αυτό, που οι άνθρωποι ως σκεπτόμενα πλάσματα κάνουμε. Σύμφωνα με όλα τα παραπάνω, λοιπόν, γίνεται εύκολα αντιληπτό ότι ένας υπολογιστής λέμε ότι καθίσταται πιο αποτελεσματικός ή ότι μαθαίνει με βάση την εμπειρία που αποκτάει μέσα από τους αλγόριθμους, τους οποίους τρέχει κατά τη διαδικασία της μάθησης. Έτσι, η Μηχανική Μάθηση κατηγοριοποιείται ευρέως σε Επιβλεπόμενη Μάθηση (Supervised Learning), σε Μη-Επιβλεπόμενη Μάθηση (Unsupervised Learning) και στην Ενισχυτική Μάθηση (Reinforcement Learning), ανάλογα με το είδος της εμπειρίας που προσφέρεται στον υπολογιστή, κατά τη διαδικασία της μάθησης. Η Επιβλεπόμενη Μάθηση (Supervised Learning) στοχεύει στην εκμάθηση μίας χαρτογράφησης από εισόδους (inputs) x και από εξόδους (outputs) y , δεδομένης μίας γνωστής σειράς ζευγών από εισόδους και ε-

ξόδους $D = (x_i, y_i)$, $i = 1, \dots, N$. Εδώ το D ονομάζεται εκπαιδευτικό σύνολο (training set) από εισόδους-εξόδους και N είναι ο αριθμός των εκπαιδευτικών παραδειγμάτων, που χρησιμοποιεί ο "δάσκαλος" κατά τη διαδικασία μάθησης του υπολογιστή. Έτσι, αναφέρουμε ότι, στην Επιβλεπόμενη Μάθηση ο υπολογιστής δέχεται τις παραδειγματικές εισόδους x και τα επιθυμητά αποτελέσματα y από έναν "δάσκαλο", με στόχο να μάθει έναν γενικό κανόνα (αλγόριθμο), προκειμένου να αντιστοιχίσει τις εισόδους με τα αντίστοιχα αποτελέσματα τους. Στην πιο απλή μορφή, κάθε εκπαιδευτική είσοδος (training input) x_i είναι ένα διάνυσμα αποτελούμενο από αριθμούς, όπως το ύψος και το βάρος ενός ανθρώπου, τα οποία ονομάζονται χαρακτηριστικά ή γνωρίσματα. Γενικά, όμως, τα γνωρίσματα αυτά μπορεί να αποτελούν ένα σύνθετα δομημένο αντικείμενο, όπως για παράδειγμα μία εικόνα, μία πρόταση, μία χρονοσειρά, ένα γράφημα κ.α. Ομοίως, η μορφή των εξόδων y_i , γνωστές και ως μεταβλητές απόκρισης, μπορεί να είναι οτιδήποτε. Παρόλα αυτά οι περισσότερες μέθοδοι υποθέτουν ότι οι έξοδοι y_i , είναι μία κατηγορηματική ή ονομαστική μεταβλητή από κάποιο πεπερασμένο σύνολο $y_i \in 1, \dots, C$ (άντρας ή γυναίκα) ή ότι τα y_i είναι μίας πραγματικής αξίας βαθμωτά μεγέθη, όπως το εισόδημα ενός νοικοκυριού. Όταν οι έξοδοι y_i είναι κατηγορηματικές μεταβλητές, το πρόβλημα είναι γνωστό ως Ταξινόμηση (Classification) ή ως αναγνώριση προτύπων (Pattern Recognition). Ενώ όταν οι έξοδοι y_i αποτελούν πραγματικά βαθμωτά μεγέθη το πρόβλημα ονομάζεται Οπισθοδρόμηση (Regression). Στη Μη-Επιβλεπόμενη Μάθηση (Unsupervised Learning), τώρα, παρέχονται στον εκπαιδευόμενο υπολογιστή μόνο είσοδοι $D = x_i$, $i = 1, \dots, N$ και ο στόχος είναι να βρεθούν "ενδιαφέροντα μοτίβα" στη βάση δεδομένων του διανύσματος D . Η διαδικασία αυτή ονομάζεται "Ανακάλυψη της γνώσης", καθώς σε αντίθεση με την Επιβλεπόμενη Μάθηση, όπου ο "δάσκαλος" προσφέρει στον εκπαιδευόμενο υπολογιστή ζεύγη εισόδων-εξόδων και αυτός ψάχνει να βρει τον αλγόριθμο, ο οποίος συνδέει τις γνωστές εισόδους με τις αντίστοιχες εξόδους, εδώ ο υπολογιστής γνωρίζει μόνο το διάνυσμα των εισόδων και προσπαθεί να βρει κρυμμένα μοτίβα στη βάση δεδομένων των εισόδων που γνωρίζει. Καταλήγουμε στο συμπέρασμα, λοιπόν, ότι η ΜηΕπιβλεπόμενη Μάθηση δεν είναι ένα καλά ορισμένο πρόβλημα, καθώς δε γνωρίζουμε τι είδους μοτίβων αναζητάμε και άρα δεν υπάρχει προφανές σφάλμα μέτρησης για να χρησιμοποιήσουμε στην αξιολόγηση των πιθανών λύσεων. Εκτός από τις δύο κατηγορίες Μηχανικής Μάθησης, οι οποίες παρουσιάστηκαν παραπάνω υπάρχει και η λεγόμενη Ενισχυτική Μάθηση (Reinforcement Learning), η οποία είναι λιγότερη διαδεδομένη. Η Ενισχυτική Μάθηση παρουσιάζεται ως μια ομάδα τεχνικών σύμφωνα με τις οποίες ένα σύστημα μάθησης προσπαθεί να εκπαιδευτεί μέσα από την αλληλεπίδραση με το περιβάλλον του. Η Ενισχυτική Μάθηση ως μέθοδος είναι εμπνευσμένη από τα μοντέλα της μάθησης με επιβράβευση και τιμωρία, που συναντώνται στην εκπαίδευση των έμβιων όντων π.χ. σκύλος, γάτα. Ο έλεγχος της κίνησης ρομπότ, η βελτιστοποίηση των εργασιών σε εργοστάσια, καθώς και η εκμάθηση επιτραπέζιων παιχνιδιών, όπως το σκάκι, από έναν υπολογιστή αποτελούν παραδείγματα εφαρμογών και

χρήσης της Ενισχυτικής Μάθησης. Το σύστημα δεν καθοδηγείται από κάποιον εξωτερικό επιβλέποντα για το ποια ενέργεια πρέπει να εκτελέσει, αλλά είναι ελεύθερο να ανακαλύψει μόνο του τις ενέργειες, που θα του αποφέρουν το μέγιστο κέρδος. Με τη χρήση των παραπάνω μεθόδων έχουμε φτάσει ένα βήμα πιο κοντά στο ερώτημα του Alan Turing, εάν οι μηχανές θα μπορούν να κάνουν αυτό που οι άνθρωποι ως νοήμον όντα μπορούν. Υπολογιστές νικάνε μερικούς από τους καλύτερους παίκτες σκάκι στον κόσμο, αυτοκίνητα μαθαίνουν να "οδηγούν" με χαρακτηριστικό παράδειγμα το σύστημα ALVINN του Pomerleau (1993), αλλά και άλλα τεχνολογικά επιτεύγματα, τα οποία παρότι δεν έχουν φτάσει όλα στο τελικό στάδιο εφαρμογής τους, δείχνουν την τεράστια εξέλιξη, καθώς και τις δυνατότητες που η Μηχανική Μάθηση μπορεί να προσφέρει στην ανθρωπότητα. Συνοψίζοντας, λοιπόν, καταλήγουμε στο συμπέρασμα ότι η Μηχανική Μάθηση έχει αποκτήσει περίοπτη θέση στην επιστημονική κοινότητα, γεγονός που διαφαίνεται και από την αύξηση του αριθμού των επιστημόνων που ασχολούνται με τον τομέα αυτόν.

Κεφάλαιο 2

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ)

2.0.1 Μηχανική Μάθηση και Νευρωνικά Δίκτυα

Η Μηχανική Μάθηση προσφέρει τεχνικές επίλυσης προβλημάτων προσανατολισμένες στα δεδομένα (data driven), οι οποίες βασίζονται κατά κύριο λόγο στην ικανότητα των συστημάτων που αναπτύσσονται, προσαρμόζονται με βάση τα εκάστοτε δεδομένα. Ο συνδυασμός του εύρους των ικανοτήτων του πεδίου της Μηχανικής Μάθησης και η αδυναμία των αυστηρών αλγοριθμικών μεθόδων να αντιμετωπίσουν προβλήματα υψηλότερου επιπέδου, όπως η ταξινόμηση οδήγησαν στις τεχνικές της Μηχανικής Μάθησης. Μία από της πιο σημαντικές μεθόδους της Μηχανικής Μάθησης, είναι τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ), τα οποία προσεγγίζουν τη λειτουργία του ανθρώπινου εγκεφάλου. Μέχρι σήμερα, χρησιμοποιούνται σε ένα ευρύ φάσμα επιστημονικών περιοχών για την επίλυση προβλημάτων ταξινόμησης ή πρόβλεψης, όπως η Ιατρική, η Βιολογία, η Γεωλογία και η Φυσική. Η αρχιτεκτονική τους στηρίζεται στην αρχιτεκτονική των βιολογικών νευρωνικών δικτύων και έτσι είναι ικανά να εκτελούν υπολογισμούς με μαζικό τρόπο. Σκοπός των ΤΝΔ είναι να μπορούν να εκτελέσουν τους υπολογισμούς, που εκτελεί ο ανθρώπινος εγκέφαλος, δηλαδή να μπορούν να μεταδίδουν πληροφορίες σχετικά με τα ερεθίσματα που λαμβάνουν. Τα ΤΝΔ εκπαιδεύονται είτε με επίβλεψη (Supervised Learning) είτε χωρίς επίβλεψη (Unsupervised Learning). Τα Νευρωνικά Δίκτυα δεν είναι αποτελεσματικά σε προβλήματα, όπου η αριθμητική ακρίβεια καθίσταται σημαντικός παράγοντας. Υπάρχουν πολλά μοντέλα νευρωνικών δικτύων ανάλογα με την αρχιτεκτονική τους, τον τρόπο λειτουργίας και εκπαίδευσης τους. Η ανάπτυξη του τομέα των ΤΝΔ έχει περάσει διαχρονικά από διάφορες φάσεις με αξιοσημείωτα και ενδιαφέροντα επιτεύγματα. Έτσι, τα τελευταία 30 χρόνια και μέσα από τη συνεχή εξέλιξή του, συναντάμε αυτόν τον επιστημονικό κλάδο, ολοένα και περισσότερο σε διάφορους τομείς της ζωής, όπως στην οικονομία, στην ασφάλεια και στην εκπαίδευση.

2.0.2 Εισαγωγή στα Νευρωνικά Δίκτυα

Τα Τεχνητά Νευρωνικά Δίκτυα, αποτελούν απλοποιημένα μοντέλα του κεντρικού νευρικού συστήματος των ζώντων οργανισμών και ειδικότερα του ανθρώπου. Οι λειτουργίες τους, μιμούνται αυτή των βιολογικών νευρώνων του εγκεφάλου και τη δομή των βιολογικών Νευρικών Δικτύων. Επιχειρούν δηλαδή, να συνδυάσουν τη λειτουργία του ανθρώπινου εγκεφάλου, με τον αυστηρά αφηρημένο μαθηματικό τρόπο σκέψης. Έτσι, για παράδειγμα, ένα Νευρωνικό Δίκτυο είναι ικανό να μαθαίνει, να εκπαιδεύεται, να θυμάται μία αριθμητική παράσταση κ.λ.π, πράξεις που μέχρι πρότινος αποδίδαμε μόνο στην ανθρώπινη σκέψη και ικανότητα. Επιπλέον, καθίστανται ικανά να μαθαίνουν πολύπλοκες συναρτήσεις και άλλα σύνθετα αντικείμενα της μαθηματικής Ανάλυσης. Ειδικότερα, αποτελούνται από διασυνδεδεμένα υπολογιστικά στοιχεία, τα οποία είναι ικανά να ανταποκρίνονται σε ερεθίσματα που δέχονται στην είσοδο τους και να μαθαίνουν να προσαρμόζονται στο περιβάλλον τους. Όπως γίνεται αντιληπτό, πηγή έμπνευσης στην έρευνα των ΤΝΔ, αποτελεί η Βιολογία και η Ιατρική. Οι ζωντανοί οργανισμοί, από τους πιο απλούς μέχρι τον πιο σύνθετο, εκτελούν διάφορες καθημερινές διεργασίες, όπως είναι η επαφή με τον εξωτερικό κόσμο, η μάθηση και η μνήμη. Υπεύθυνο για την εκτέλεση αυτών των εργασιών, είναι το Νευρικό Σύστημα, που όλοι οι έμβιοι οργανισμοί εκτός από τα φυτά διαθέτουν. Το νευρικό σύστημα των οργανισμών, αποτελείται από πολλά νευρωνικά δίκτυα, που είναι εξειδικευμένα σε διάφορες εργασίες. Τα νευρωνικά δίκτυα αυτά, τα ονομάζουμε βιολογικά νευρωνικά δίκτυα, καθώς είναι τα πρώτα δίκτυα που μελετήθηκαν και υπάρχουν στους ζώντες οργανισμούς. Πυρήνας του νευρικού συστήματος των οργανισμών αποτελεί ο εγκέφαλος, ο οποίος επίσης αποτελείται από νευρωνικά δίκτυα. Κάθε νευρωνικό δίκτυο αποτελείται από ένα μεγάλο αριθμό μονάδων, που ονομάζονται νευρώνες ή νευρώνια (Neurons). Οι νευρώνες αποτελούν το βασικό δομικό στοιχείο του εγκεφάλου. Είναι δηλαδή, η πιο μικρή ανεξάρτητη μονάδα του Δικτύου, όπως το άτομο είναι η πιο μικρή μονάδα της ύλης. Οι Νευρώνες επεξεργάζονται πληροφορίες συνεχώς και ασταμάτητα, παίρνοντας και στέλνοντας ηλεκτρικά σήματα σε άλλους νευρώνες. Κίνητρο για την έρευνα και τη μελέτη του νευρώνα και του νευρωνικού δικτύου, αποτέλεσε η ελπίδα ανακάλυψης ενός νέου πρότυπου υπολογιστικού μοντέλου με δικτυακή δομή, όμοια με αυτή του εγκεφάλου. Με αφετηρία τον στόχο της ανακάλυψης αυτής, δημιουργήθηκε μια καινοτόμα υπολογιστική πλατφόρμα γνωστή ως Connectionist Model, κατάλληλη για την ανάπτυξη ευφυών αλγόριθμων και διαδικασιών, όπως η μάθηση και η απομνημόνευση. Μέσα από αυτό το μοντέλο συνδέονται διάφοροι κλάδοι επιστημών όπως η Ψυχολογία, η Νευρολογία, η Ανθρωπολογία κ.α. πλάθοντας με αυτό τον τρόπο τη Γνωσιακή Επιστήμη (Cognitive Science) και κατέπεκταση το παραπάνω μοντέλο, το οποίο προσπαθεί να ενώσει τον ψυχικό κόσμο των έμβιων οργανισμών και τη νοημοσύνη, σε έναν υπολογιστή με τη χρήση των ΤΝΔ. Τα πιο διαδεδομένα Τεχνητά Νευρωνικά Δίκτυα, χρησιμοποιούν πολύ απλοποι-

ημένα μοντέλα νευρώνων, με συνέπεια να διατηρούν μόνο τα κύρια χαρακτηριστικά των μοντέλων που υπάρχουν στη Νευρολογία. Έτσι, καταλήγουμε στο συμπέρασμα, ότι τα Τεχνητά Νευρωνικά Μοντέλα έχουν ελάχιστη σχέση με τα αντίστοιχα Βιολογικά Νευρωνικά Δίκτυα. Ωστόσο, στην κατανόηση της ευφυούς συμπεριφοράς των Βιολογικών Νευρωνικών Δικτύων, οι διαφορές αυτές στα χαρακτηριστικά, δεν έχουν ιδιαίτερη αξία. Παρόλα αυτά, ακόμα, και τα απλά μοντέλα είναι ικανά να δημιουργήσουν ενδιαφέροντα δίκτυα, με την προϋπόθεση ότι πληρούν δύο βασικά χαρακτηριστικά : 1. Οι νευρώνες τους πρέπει να διαθέτουν κατάλληλες παραμέτρους, ώστε να γίνεται πιο εύκολη διαδικασία της μάθησης. 2. Το Δίκτυο είναι αναγκαίο να δομείται από πολλούς νευρώνες, έτσι ώστε να είναι εφικτός ο παραλληλισμός της επεξεργασίας και η κατανομή της πληροφορίας. Συνοψίζοντας όλα τα παραπάνω, μπορούμε να ορίσουμε τα Τεχνητά Νευρωνικά Δίκτυα ως ένα κύκλωμα διασυνδεδεμένων νευρώνων, το οποίο με την ανάπτυξη κατάλληλων αλγόριθμων εκπαίδευσης είναι ικανό να προσομοιώσει ευφυείς διαδικασίες. Όμως, για την υλοποίηση της διαδικασίας που ορίσαμε παραπάνω, είναι απαραίτητο να καθορίσουμε το περιβάλλον εκπαίδευσης.

2.0.3 Ιστορική Αναδρομή

Η πηγαία ανάγκη του ανθρώπου για συνεχή εξέλιξη, σε συνδυασμό με την εφευρετικότητα οδήγησαν στη σταδιακή δημιουργία μίας νέας επιστημονικής περιοχής, τα λεγόμενα Νευρωνικά Δίκτυα. Η ανάπτυξη αυτού του κλάδου ξεκίνησε σε παγκόσμιο επίπεδο, μόλις τις τελευταίες δεκαετίες του 20ού αιώνα, γνωρίζοντας μεγαλύτερη αναγνώριση στο κοινό μετά το 1980. Στηριζόμενοι στην ανάπτυξη του υλικού των Η/Υ, όσο και στην ανάπτυξη νέων αλγόριθμων εκπαίδευσης, οι επιστήμονες κατόρθωσαν να βάλουν τα θεμέλια στον τομέα των νευρωνικών δικτύων. Το 1943 ο νευροφυσιολόγος McCulloch και ο 18χρονος πρωτοετής φοιτητής μαθηματικών Pitts, παρουσιάζουν το πρώτο μοντέλο νευρωνικού δικτύου, στο οποίο προτείνουν τους νευρώνες, ως τη βασική μονάδα του δικτύου. Μάλιστα, οι επιστήμονες παρομοίασαν τον τρόπο λειτουργίας των νευρώνων και των συνδέσεων τους, με αυτή του ηλεκτρικού ρεύματος. Με βάση την εργασία των δύο ερευνητών, εισάγεται για πρώτη φορά η ιδέα ότι ένα νευρωνικό δίκτυο, αποτελείται από μια συλλογή μεγάλου αριθμού νευρώνων. Επιπρόσθετα, τεκμηριώνεται ο τρόπος με τον οποίο θα μπορούσαν να λειτουργούν οι νευρώνες, με τις διασυνδέσεις τους. Αυτή είναι η πρώτη ιστορικά καταγεγραμμένη εικόνα, ενός νευρωνικού δικτύου. Οι McCulloch και Pitts, δεν εφησυχάστηκαν στην πρώτη τους επιτυχία, συνέχισαν την ενασχόληση τους με τα νευρωνικά δίκτυα και κατόρθωσαν το 1947 να φτάσουν σε ένα νέο πιο εξελιγμένο πρότυπο, για την αναγνώριση σχημάτων. Στο καινούργιο αυτό μοντέλο, ο νευρώνας θεωρείται, ότι μπορεί να έχει δύο μόνο καταστάσεις. Ο κάθε νευρώνας επιτρέπεται να δέχεται πολλές εισόδους, αλλά μπορεί να δίνει αποκλειστικά μόνο μία έξοδο. Οι πληροφορίες της εξόδου ενός νευρώνα, δεν επιτρέπεται

να συνδεθούν με την έξοδο άλλου νευρώνα, αλλά κατευθύνονται ρητά, σε είσοδο άλλου νευρώνα. Οι απολήξεις των νευρώνων διαχωρίζονται σε διεγερτικές ή ανασταλτικές. Δηλαδή, ένας νευρώνας μπορεί είτε να πυροδοτεί, είτε να βρίσκεται σε ηρεμία. Όταν ένας νευρώνας διεγείρεται, στέλνει έναν παλμό με πληροφορίες, οι οποίες ελέγχονται μέσα στον νευρώνα, από τις πύλες. Οι πύλες είναι διεγερτικές ή ανασταλτικές, όπως και η έξοδος του εκάστοτε νευρώνα. Οι λειτουργίες που περιέγραψαν στη δεύτερη εργασία τους για τα νευρωνικά δίκτυα, οι δύο ερευνητές, αναφέρεται ότι γίνονται σε διακριτό χρόνο και έτσι θεωρητικά, όλοι οι νευρώνες αποκρίνονται ταυτόχρονα. Άρα, το σύστημα δρα συγχρονισμένα. Τα δίκτυα των McCulloch-Pitts, προσπαθούν να εισάγουν και να εξηγήσουν για πρώτη φορά, πως λειτουργεί η μνήμη. Η ιδέα για τη λειτουργία του μηχανισμού μνήμης, είναι η ύπαρξη κλειστών διαδρομών του σήματος μέσα στο δίκτυο. Αν δεν υπάρχει καμία τέτοια κλειστή διαδρομή, τότε το δίκτυο παραμένει σε κατάσταση ηρεμίας. Πιο συγκεκριμένα, δημιουργείται ένας μηχανισμός ανάδρασης (feedback), σύμφωνα με τον οποίο, μία ίνα, συνδέει την απόληξη ενός κυττάρου, με την είσοδο του ίδιου κυττάρου. Με αυτό τον τρόπο, αποστέλλονται οι πληροφορίες στις πύλες του κυττάρου και άρα επανατροφοδοτούνται στον νευρώνα. Αυτή η ροή δεδομένων στον νευρώνα, αποτελεί τον πρώτο μηχανισμό μνήμης. Λίγα χρόνια αργότερα, ο J. Von Neumann, χρησιμοποίησε τις εργασίες των McCulloch και Pitts, ως παράδειγμα για υπολογιστικές μηχανές, ωθώντας ουσιαστικά στην άντληση πληροφοριών από τα βιολογικά δίκτυα και τη δημιουργία των πρώτων νευρωνικών δικτύων. Σημαντικό στίγμα στην εξέλιξη των νευρωνικών δικτύων άφησε ο D. Hebb, ο οποίος το 1949 με το βιβλίο του "The Organization of Behavior" εισήγαγε τον κανόνα μάθησης του Hebb. Στο έργο του ο Hebb, αποτυπώνει σαν πυρήνα, τις συνδέσεις μεταξύ των μονάδων του συστήματος, δηλαδή των νευρώνων. Σύμφωνα με τον κανόνα μάθησης του Hebb, κάθε φορά που το δίκτυο χρησιμοποιεί τις νευρωνικές του συνδέσεις, αυτές ενισχύονται και το δίκτυο πλησιάζει όλο και πιο πολύ, στο να μάθει το πρότυπο, την επιθυμητή διεργασία. Το 1957, ο F. Rosenblatt, παρουσίασε για πρώτη φορά το μοντέλο του αισθητήρα (Perceptron). Το μοντέλο αυτό, είχε μόνο δύο επίπεδα, ένα εισόδου και ένα εξόδου, μέσα στα οποία το σήμα διέρχεται από την είσοδο στην έξοδο. Οι ερευνητές αυτοί, στο βιβλίο τους "Perceptrons", κατορθώνουν να εκφράσουν τη χρησιμότητα του προτύπου του αισθητήρα, καταγράφοντας μαζί και τους περιορισμούς που αυτό έχει. Την ίδια περίπου εποχή, με την ανάπτυξη του μοντέλου του Perceptron, δύο άλλοι επιστήμονες, οι Widrow και Hoff, ανέπτυξαν δύο νέα μοντέλα, το Adaline και το Madaline. Τα μοντέλα αυτά, ήταν τα πρώτα που χρησιμοποιήθηκαν σε πρακτικά προβλήματα της καθημερινότητας, όπως η εξάλειψη της ηχώ, στις τηλεφωνικές γραμμές. Για τα επόμενα 20 χρόνια, οι περιορισμοί και οι δυσκολίες που δημιουργούσαν αυτοί στην ανάπτυξη των νευρωνικών δικτύων, απέτρεψε πληθώρα ερευνητών να ασχοληθούν με το πεδίο αυτό. Το 1982, η ώθηση που θα έβαζε τέλος στο ερευνητικό τέλμα στο πεδίο των νευρωνικών δικτύων, δόθηκε από τον J. Hopfield. Ο Hopfield, έδειξε στο έργο του με αυστηρά μαθηματική απόδειξη, πως

ένα νευρωνικό δίκτυο, μπορεί να χρησιμοποιηθεί ως αποθηκευτικός χώρος (storage device), όπως επίσης πως μπορεί ένα δίκτυο να επανακτήσει όλη την πλατφόρμα ενός συστήματος, ακόμα και εάν δεν του δοθεί ολόκληρο το σύστημα, παραμόνο μερικά τμήματα. Ένα επόμενο σημαντικό βήμα, στην ανάπτυξη των νευρωνικών δικτύων, ήταν η πρόοδος που επιτελέσθηκε, στη διαδικασία εκπαίδευσης των δικτύων, όταν επινοήθηκε ο κανόνας διόρθωσης του σφάλματος (error correction learning). Σύμφωνα με τον κανόνα αυτόν λοιπόν, κατά την εκπαίδευση ενός δικτύου, σε όποια κατάσταση και αν βρίσκεται αυτό, μια δεδομένη στιγμή, σημασία έχει η απόκλιση που δίνει στην έξοδο του το δίκτυο από την αναμενόμενη τιμή ή τον στόχο που έχουμε θέσει. Η διαφορά που παράγει το δίκτυο τη δεδομένη στιγμή, ονομάζεται σφάλμα και ενεργοποιεί έναν μηχανισμό ελέγχου, με στόχο να επιφέρει μια σειρά από διορθωτικές αλλαγές στα βάρη των νευρώνων. Μέσα από αυτές τις αλλαγές αυτοβελτίωσης, το δίκτυο έρχεται ένα βήμα πιο κοντά στην εκπαίδευσή του. Το 1986, οι McClelland και Rumelhart, δημοσιεύουν το βιβλίο τους "Parallel Distributed Processing", στο οποίο εκφράζουν την καινοτόμα ιδέα τους, κατά την οποία ένα νευρωνικό δίκτυο, μπορεί να θεωρηθεί και να χρησιμοποιηθεί ως παράλληλος επεξεργαστής. Ένα βήμα πιο πέρα από το μοντέλο Perceptron, το έργο αυτό, επιτρέπει την ύπαρξη και άλλων επίπεδων νευρώνων, εκτός από την είσοδο και την έξοδο, που αποτελούν την εσωτερική δομή του δικτύου. Ως απόρροια των παραπάνω εξελίξεων στα δίκτυα, εισάγεται το 1986 μία νέα μέθοδος εκπαίδευσης, γνωστή και ως back-propagation. Ο αλγόριθμος μάθησης, Back Propagation, αποτελεί μέχρι και σήμερα την πιο χρήσιμη τεχνική εκπαίδευσης των δικτύων. Η ραγδαία εξέλιξη του πεδίου των νευρωνικών δικτύων, δε θα μπορούσε στην εποχή της πληροφορίας και της ανάπτυξης, να μην έχει τη θέση που της αρμόζει στον επιστημονικό κλάδο. Από το 1985 και μετά, καθιερώθηκαν συνέδρια, αφιερωμένα αποκλειστικά στα νευρωνικά δίκτυα, δημιουργήθηκαν περιοδικά με επίκεντρο τα νευρωνικά δίκτυα και εταιρείες κολοσσοί, οι οποίες προώθησαν τα προϊόντα τους στην αγορά.

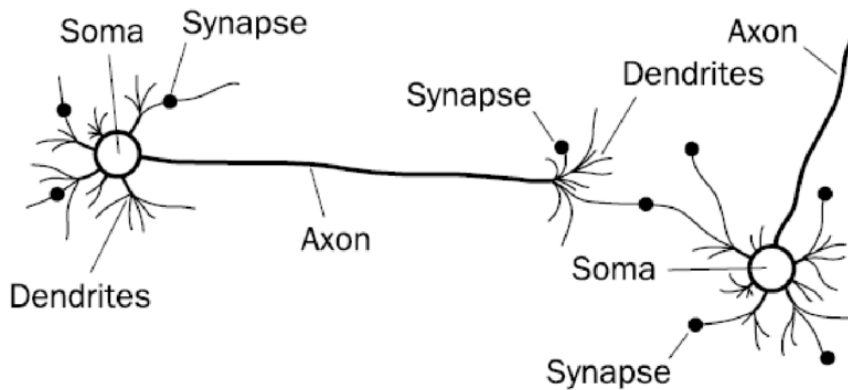
2.0.4 Βιολογικά Νευρωνικά Δίκτυα

Σημαντικό πυλώνα, στην ανάπτυξη των Τεχνητών Νευρωνικών Δικτύων, αποτέλεσε η κατανόηση της δομής και της λειτουργίας των Βιολογικών Νευρωνικών Δικτύων και ειδικότερα του ανθρώπινου εγκεφάλου. Ακόμα και μικρά παιδιά, είναι ικανά να εκτελούν απλές διεργασίες, τις οποίες ένας Η/Υ, παρά τις τεράστιες ταχύτητες που διαθέτει σε σχέση με τον άνθρωπο, δυσκολεύεται. Έτσι, η δομή και η λειτουργία των Βιολογικών Νευρωνικών Δικτύων, αποτέλεσε πηγή έμπνευσης και βασικό συστατικό στην εξέλιξη των ΤΝΔ. Κρίνεται, λοιπόν, απαραίτητο να κατανοήσουμε ορισμένα κύρια σημεία στη δομή και τη λειτουργία των Βιολογικών Νευρωνικών Δικτύων, πριν αναφερθούμε στην αντίστοιχη των Τεχνητών Νευρωνικών Δικτύων. Το ανθρώπινο νευρικό σύστημα, έχει ως δομική μονάδα του, τον νευρώνα. Ο νευρώνας, είναι ένα εξειδικευμένο είδος κυττάρου

που αποτελεί τον πυρήνα των συστημάτων επεξεργασίας πληροφοριών που απαρτίζουν το ανθρώπινο κεντρικό νευρικό σύστημα. Καθίσταται το βασικό δομικό στοιχείο του εγκεφάλου, τόσο του ανθρώπου, όσο και των ζώων. Υπολογίζεται ότι ο εγκέφαλος ενός ενήλικα, αποτελείται περίπου από 100 δισεκατομμύρια νευρώνες, κάθε ένας από τους οποίους συνδέεται, με περίπου 10.000 άλλους νευρώνες, μέσω των συνδέσεων. Οι συνδέσεις αυτές, ονομάζονται συνάψεις και ο αριθμός τους δεν είναι σταθερός, αλλά διαφέρει από κύτταρο σε κύτταρο. Ένα σύνολο νευρώνων μαζί με τις συνάψεις τους, αποτελούν ένα νευρωνικό δίκτυο (neural net) και όλο το πλήθος των νευρωνικών δικτύων που υπάρχουν σε έναν άνθρωπο, απαρτίζουν το λεγόμενο Κεντρικό Νευρικό Σύστημα (Central Nervous System). Τα κύτταρα των νευρώνων, είναι κοινά αποδεκτό, ότι δεν πολλαπλασιάζονται, ούτε αναπαράγονται. Έτσι, κάθε υγιής ενήλικας, χάνει καθημερινά έναν αριθμό νευρώνων, η ποσότητα των οποίων, επηρεάζεται κυρίως από την ηλικία και το αλκοόλ. Αντίθετα, με τους νευρώνες, οι συνάψεις βρίσκονται συνεχώς σε ισορροπία, καθώς δημιουργούνται νέες και καταστρέφονται παλιές. Με τους νευρώνες ως δομικά στοιχεία του νευρικού συστήματος, ο εγκέφαλος μεταφέρει πληροφορίες στα διάφορα άλλα συστήματα που ο ανθρώπινος οργανισμός διαθέτει, όπως για παράδειγμα το πεπτικό σύστημα. Η ανατομία του νευρώνα, δεν παρουσιάζει καμία διαφορά σε σχέση με τα υπόλοιπα κύτταρα του ανθρώπινου οργανισμού. Αποτελείται λοιπόν, από το σώμα, τον πυρήνα, τους δενδρίτες, τον άξονα και τις συνάψεις. Πιο συγκεκριμένα, το σώμα αποτελεί το κυρίως μέλος του νευρώνα, μέσα στο οποίο βρίσκεται ο πυρήνας του κυττάρου. Στον πυρήνα υπάρχουν όλες οι πληροφορίες, του γεννητικού υλικού του οργανισμού. Οι δενδρίτες, είναι το σημείο εισόδου των πληροφοριών, όπου ο κάθε νευρώνας δέχεται από τους γειτονικούς του νευρώνες. Η ροή δεδομένων, εισέρχεται στον κάθε νευρώνα και μέσω των δενδριτών, με τη μορφή ηλεκτρικών παλμών. Η εισερχόμενη στον εκάστοτε νευρώνα πληροφορία, εξέρχεται από αυτόν, μέσω ηλεκτρικών σημάτων, με τη λειτουργία του σημείου εξόδου του νευρώνα, τον άξονα. Κλείνοντας τη βασική δομή ενός νευρώνα, οι συνάψεις, είναι τα απειροελάχιστα κενά, στα οποία ενώνονται οι διακλαδώσεις του άξονα ενός νευρώνα, με τους δενδρίτες άλλων γειτονικών του. Η αμεσότητα, με την οποία μεταδίδεται η ηλεκτρική δραστηριότητα από έναν άξονα-αποστολέα στους δενδρίτες-παραλήπτες, εξαρτάται από το πλάτος της σύναψης και από την πυκνότητα του ηλεκτροχημικού υλικού. Η ποσότητα του ηλεκτρικού παλμού που μεταφέρεται τελικά στους δενδρίτες του νευρώνα - παραλήπτη, ονομάζεται συναπτικό βάρος.

Οι νευρώνες παρουσιάζουν, δύο πιθανές καταστάσεις. Κατά τη λειτουργία, του ένας νευρώνας μπορεί να είναι ενεργός ή μη-ενεργός. Όταν λέμε ότι ένας νευρώνας είναι ενεργός, εννοούμε ότι πυροδοτεί ένα ηλεκτρικό σήμα, μέσω του οποίου, μεταφέρονται τα δεδομένα που υπάρχουν στον νευρώνα, σε κάποιον γειτονικό του. Αντίστοιχα, όταν ένας νευρώνας παραμένει αδρανής, δηλαδή είναι μη-ενεργός, δεν παράγει ηλεκτρικό

Σχήμα 2.1: Βιολογικός νευρώνας

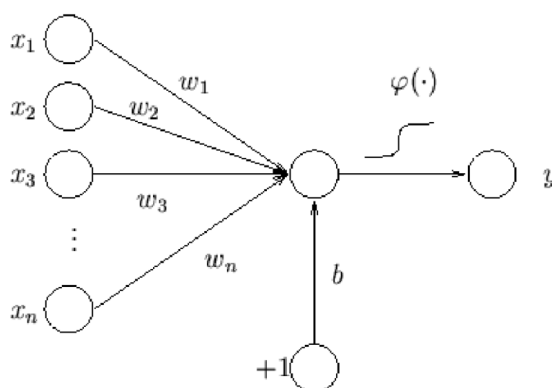


σήμα. Μέσω αυτής της ευαισθησίας του νευρώνα, να αντιδράει σε εξωτερικά ερεθίσματα (ηλεκτρομαγνητικά, χημικά, θερμικά κ.α.), παράγονται οι ηλεκτρικοί παλμοί που μεταφέρουν τις πληροφορίες. Ο εκάστοτε νευρώνας, συλλέγει όλο το ηλεκτρικό φορτίο που παραλαμβάνει, από όλες τις συνάψεις του, στους δενδρίτες του. Όσο πιο ισχυρός είναι ο δεσμός, σε μία σύναψη δύο νευρώνων, τόσο πιο έντονα συμμετέχει το συγκεκριμένο παραγόμενο φορτίο εισόδου, στο συνολικό άθροισμα φορτίου. Ένας νευρώνας είναι ενεργός, δηλαδή παράγει ηλεκτρικούς παλμούς, εάν το συνολικό άθροισμα του φορτίου του, είναι μεγαλύτερο από ένα κατώφλι. Σε αντίθετη περίπτωση, ο νευρώνας καθίσταται αδρανής. Έτσι, αναφέρουμε ότι ο νευρώνας, είναι ένα δυαδικό στοιχείο.

2.0.5 Το Μοντέλο του Τεχνητού Νευρώνα

Η πολυπλοκότητα που το ανθρώπινο νευρικό σύστημα παρουσιάζει, φανερώνει τις δυσκολίες που δημιουργούνται στην προσπάθεια απόδοσης του, στα ΤΝΔ. Παρόλα αυτά, οι επιστήμονες, με πηγή έμπνευσης τη δομή του βιολογικού νευρώνα, έχουν καταφέρει να δημιουργήσουν, ένα αντίστοιχο πρότυπο, το λεγόμενο μοντέλο του τεχνητού νευρώνα (artificial neuron). Ένας απλοποιημένος βιολογικός νευρώνας, λαμβάνει στους δενδρίτες του, σήματα εισόδου σε μορφή ηλεκτρικών παλμών, τα επεξεργάζεται και τέλος τα μεταφέρει μέσω του άξονα και των συνάψεων σε άλλους γειτονικούς νευρώνες, εάν βρίσκεται σε ενεργή κατάσταση. Στο αντίστοιχο μαθηματικό μοντέλο, ο νευρώνας λαμβάνει στους δενδρίτες του κάποια σήματα εισόδου x_1, x_2, \dots, x_n , τα οποία παίρνει από τον άξονα του προηγούμενου νευρώνα, πολλαπλασιαστικά και μέσω των συνάψεων (βάρη) w_i . Ο ρόλος των συναπτικών βαρών, είναι αντίστοιχος των συνάψεων του βιολογικού μοντέλου. Τα συναπτικά βάρη είναι εκπαιδευσιμα και ελέγχουν την επιρροή, που ασκεί ένας νευρώνας σε έναν άλλο. Ανάλογα με τις τιμές των βαρών, θετικές ή αρνητικές, η λειτουργία ενός νευρώνα είναι διεγερτική ή ανασταλτική αντίστοιχα, όσον αφορά τη μεταφορά των δεδομένων, ή όχι. Το παρακάτω σχήμα, παρουσιάζει το μοντέλο του τεχνητού νευρώνα.

Σχήμα 2.2: Το μοντέλο του τεχνητού νευρώνα



Η επεξεργασία των σημάτων, γίνεται στο σώμα του νευρώνα, όπου στα απλά τεχνητά νευρωνικά δίκτυα, χωρίζεται σε δύο μέρη: τον αθροιστή (sum) και τη συνάρτηση ενεργοποίησης (activation function). Ο αθροιστής, αθροίζει τα επηρεασμένα από τα βάρη σήματα εισόδου, παράγοντας ένα άθροισμα S . Η συνάρτηση ενεργοποίησης f , είναι ένα είδος φίλτρου που υπολογίζει την τελική τιμή του σήματος εξόδου y , συναρτήσει της ποσότητας S και της τιμής κατωφλίου θ . Ανάλογα με την τελική τιμή y του νευρώνα, ένας νευρώνας βρίσκεται σε ενεργή ή αδρανή κατάσταση, αντίστοιχα. Η τελική τιμή εξόδου y που μεταφέρει ένας ενεργός νευρώνας είναι μοναδική. Η μοναδικότητα των ηλεκτρικών σημάτων εξόδου, αφορά την τελική τιμή της και δεν έχει καμία σχέση με το πλήθος των γειτονικών νευρώνων, που λαμβάνουν το σήμα. Δηλαδή, ένας νευρώνας, είναι δυνατόν να έχει πολλές εξόδους μεταφοράς του σήματος σε γειτονικούς νευρώνες, αλλά όλες οι εξοδοι, θα έχουν την ίδια τελική τιμή.

2.0.6 Συνάρτηση Ενεργοποίησης

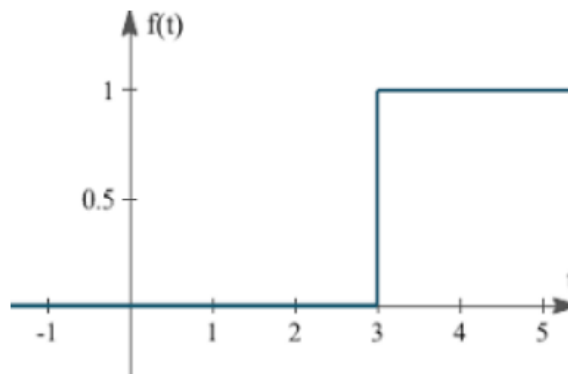
Η θεώρηση της μετάδοσης του ηλεκτρικού σήματος από έναν νευρώνα σε έναν άλλον, όπως περιγράφηκε παραπάνω, είναι γενική για όλα τα απλά τεχνητά νευρωνικά δίκτυα και δε μεταβάλλεται. Το μοναδικό κομμάτι που μπορεί να διαφέρει στη διαδικασία αυτή, είναι η συνάρτηση ενεργοποίησης του εκάστοτε νευρώνα, σύμφωνα με την οποία παράγει την τελική τιμή εξόδου του. Παρακάτω, παρατίθενται μερικές από τις πιο διαδεδομένες συναρτήσεις ενεργοποίησης και παρουσιάζονται θετικές και αρνητικές πτυχές, που αυτές έχουν στα Τεχνητά Νευρωνικά Δίκτυα:

1. Βηματική Συνάρτηση Ενεργοποίησης:

Η Βηματική, είναι μια συνάρτηση η οποία χρησιμοποιείται αρκετά, σε νευρωνικά δίκτυα με ένα μόνο νευρώνα. Η γενική της μορφή είναι η εξής: $f(net) = a$ αν $net < c$ ή b αν $net > c$ Η τιμή της ποσότητας net , πολλές φορές είναι ίση με το a , άλλες φορές ορίζεται ίση με το β και κάποιες άλλες με το $\frac{a+b}{2}$. Οι πιο συνη-

Θισμένες τιμές των a, β, ς είναι: $a = 0, b = 1, c = 0$ ή $a = -1, b = -1, c = 0$. Η δεύτερη περίπτωση Βηματικής συνάρτησης ονομάζεται και συνάρτηση Πρόσημου ((Signum Function). Η βηματική συνάρτηση, παρουσιάζει μία πολύ απλή λογική και γι'αυτό βρίσκει εφαρμογή σε απλά μοντέλα ενός νευρώνα. Ειδικότερα, αυτές οι συναρτήσεις, δομούνται στην ιδέα ύπαρξης ενός κατωφλιού c . Έτσι, όταν η τιμή της τελικής εξόδου του δικτύου είναι μικρότερη από την τιμή του κατωφλιού, τότε ο νευρώνας παραμένει αδρανής. Από την άλλη πλευρά, όταν η τιμή της τελικής απόκρισης του δικτύου, είναι μεγαλύτερη του κατωφλιού, τότε ο νευρώνας είναι ενεργός και διαδίδει το σήμα του, στους γειτονικούς νευρώνες. Παρόλη την απλή ιδέα και λειτουργία της, η Βηματική συνάρτηση δε θεωρείται χρήσιμη στα ΤΝΔ, καθώς παρουσιάζει σύμφωνα με τον Απειροστικό Λογισμό, ένα πολύ σημαντικό μειονέκτημα: Η παράγωγος της απειρίζεται. Για τον λόγο αυτόν οι ερευνητές στράφηκαν σε συναρτήσεις, των οποίων οι γραφικές παραστάσεις, προσεγγίζουν την αντίστοιχη της Βηματικής, αλλά είναι συνεχείς και παραγωγίσιμες στο πεδίο ορισμού τους. Χαρακτηριστικό παράδειγμα τέτοιας συνάρτησης, αποτελεί η σιγμοειδής συνάρτηση (Sigmoid function).

Σχήμα 2.3: Βηματική Συνάρτηση με $a=0, b=1$ και $c=3$

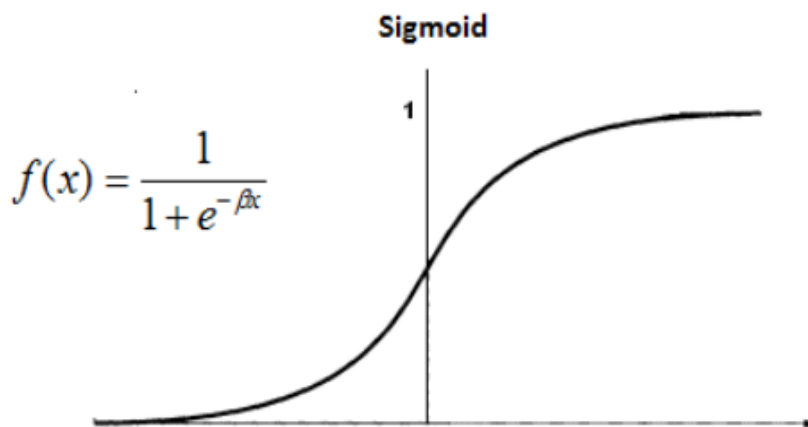


2. Σιγμοειδής Συνάρτηση Ενεργοποίησης:

Ο γενικός τύπος της σιγμοειδούς συνάρτησης είναι: $s(t) = L/(1 + e^{-k(t-t_0)})$, όπου L είναι η μέγιστη τιμή της συνάρτησης, t_0 το κεντρικό σημείο της καμπύλης και k μία παράμετρος, που καθορίζει την καμπυλότητα της γραφικής παράστασης της. Το σύνολο τιμών της, δηλαδή το σύνολο των σημάτων εξόδου που παράγει ένας νευρώνας με χρήση της σιγμοειδούς είναι το $[0, L]$. Είναι γνησίως αύξουσα και διαφορίσιμη συνάρτηση στο πεδίο ορισμού της, καλύπτοντας έτσι το κενό που αφήνει η Βηματική συνάρτηση. Στα νευρωνικά δίκτυα χρησιμοποιείται ευρέως μια ειδική περίπτωση της σιγμοειδούς, για $L = 1, t_0 = 0, k = 1$, σύμφωνα με την οποία η συνάρτηση παίρνει την τελική μορφή: $\sigma(t) = 1/(1 + e^t)$, με σύνολο τιμών το $[0, 1]$. Η σιγμοειδής συνάρτηση, αποτελεί την πιο ευρέως διαδεδομένη συνάρτηση ενεργοποίησης στα νευρωνικά δίκτυα, καθώς παραπέμπει μέσω του

συνόλου τιμών της, στις δύο καταστάσεις, όπου ένας νευρώνας μπορεί να επέλθει. Έτσι, έχουμε 0, όταν ο νευρώνας είναι πλήρως απενεργοποιημένος και 1, όταν ο νευρώνας είναι ενεργός στη μέγιστη συχνότητα. Στη σιγμοειδή συνάρτηση, η τιμή εξόδου, που παίρνει ο νευρώνας δεν είναι κεντραρισμένη γύρω από το 0, γεγονός το οποίο δημιουργεί ορισμένα προβλήματα. Δύο εκ των βασικών ζητημάτων που αντιμετωπίζονται στα νευρωνικά δίκτυα, με τη χρήση της σιγμοειδούς συνάρτησης, είναι τα εξής: α) Η συνάρτηση, παρουσιάζει όρια κορεσμού στο 0 και το 1, έτσι για πολύ μικρές ή πολύ μεγάλες τιμές εισόδου, η κλίση στις οποίες, θα είναι περίπου μηδέν. Έτσι, κατά τη χρήση της μεθόδου εκπαίδευσης της οπισθοδρόμησης, (της οποίας τα αποτελέσματα θα μελετήσουμε αργότερα), η τοπική κλίση του νευρώνα, θα έχει σχεδόν μηδενική τιμή, με αποτέλεσμα να μη διαδίδεται το σήμα από τον έναν νευρώνα στον άλλον προς τα πίσω, δηλαδή να σταματάει η μάθηση. β) Στη σιγμοειδή συνάρτηση, η αρχικοποίηση των βαρών, επιβάλλεται να είναι άκρως προσεκτική. Με βάση τον τύπο της, αν αρχικά οριστούν πολύ μεγάλες τιμές βαρών (αρνητικές ή θετικές), τότε ο νευρώνας μπορεί να οδηγηθεί σε κορεσμό, δηλαδή να παίρνει συνεχώς την τιμή 1 ως έξοδο.

Σχήμα 2.4: Η Σιγμοειδής Συνάρτηση

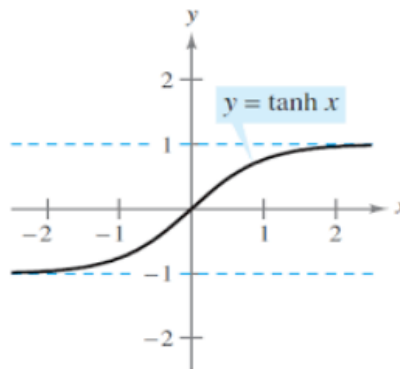


3. Υπερβολική Εφαπτομένη Συνάρτηση Ενεργοποίησης:

Η Υπερβολική Εφαπτομένη, αποτελεί μία συνεχής και διαφορίσιμη στο πεδίο ορισμού της, συνάρτηση ενεργοποίησης. Έτσι, ομοίως με τη σιγμοειδή, ξεπερνάει το εμπόδιο της απειρίζουσας παραγώγου, που παρουσιάζει η Βηματική συνάρτηση ενεργοποίησης. Ο τύπος της συνάρτησης είναι : $\tanh(t) = e^t - e^{-t} / (e^t + e^{-t})$. Σε αντίθεση με τη σιγμοειδή συνάρτηση, η τιμή εξόδου που παράγει ένας νευρώνας με χρήση της υπερβολικής συνάρτησης ενεργοποίησης, παραμένει κεντραρισμένη γύρω από το 0. Με βάση κυρίως το παραπάνω πλεονεκτήματα, η υπερβολική συνάρτηση προτιμάται, σε σχέση με τη σιγμοειδή. Παρόλα αυτά, το γεγονός ότι επέρχεται σε κορεσμό, οδήγησε με τον καιρό και με την ανάπτυξη νέων συναρ-

τήσεων, στη σταδιακή εγκατάλειψη της χρήσης της υπερβολικής εφαπτομένης, ως συνάρτηση ενεργοποίησης.

Σχήμα 2.5: Η Υπερβολική Εφαπτομένη



4. **Ανορθωμένη Γραμμική ή Συνάρτηση Ράμπας (ReLU):**

$$\alpha(t) = \max(0, t) \quad (2.1)$$

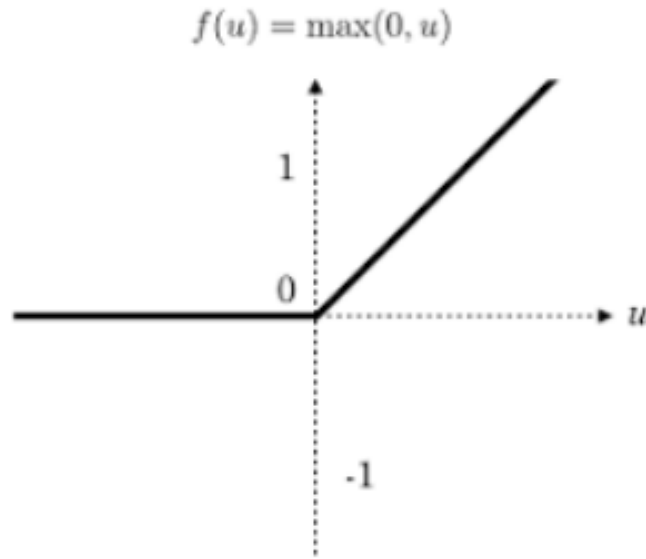
Ο γενικός τύπος της συνάρτησης ράμπας, όπως δομήθηκε, προσφέρει με γρήγορο και εύκολα υπολογίσιμο τρόπο την τελική τιμή εξόδου του νευρώνα. Η διαδικασία αυτή, επιτυγχάνεται ουσιαστικά με τη σύγκριση του αθροίσματος Σ των γινομένων των βαρών, επί τις τιμές εισόδου του νευρώνα με το 0. Έτσι, σε αντίθεση με τις προηγούμενες πολύπλοκες συναρτήσεις που αναφέραμε (\tanh , \exp) η συνάρτηση Ράμπας οδηγεί στην επιτυχή εκπαίδευση του νευρωνικού δικτύου. Ωστόσο, η συνάρτηση ράμπας, εκτός από την πληθώρα πλεονεκτημάτων που έχει, παρουσιάζει ένα βασικό μειονέκτημα. Εάν, κατά τη διαδικασία της μάθησης, εφαρμόζεται συνεχής ανανέωση των συνοπτικών βαρών, τότε υπάρχει ενδεχόμενο να παραχθεί αρνητικό άθροισμα εξόδου S , με αποτέλεσμα ο νευρώνας να παραμένει συνεχώς ανενεργός. Αυτό, γιατί στη σύγκριση μεταξύ του αρνητικού αθροίσματος S και του μηδενός, το μηδέν θα είναι πάντοτε μεγαλύτερο, από οποιαδήποτε αρνητική ποσότητα. Ως εκ τούτου, ο κάθε νευρώνας του νευρωνικού δικτύου, θα δημιουργεί μηδενική τιμή εξόδου, με αποτέλεσμα ο νευρώνας να παραμένει πρακτικά “νεκρός” και να μη μαθαίνει

5. **Διαρρέουσα Ανορθωμένη Γραμμική (Leaky ReLU) και Παραμετροποιήσιμη Συνάρτηση Ενεργοποίησης (pReLU):**

$$\hat{h}(t) = \max(0, t) + \min(0, t) = at, \text{ αν } t < 0 \text{ ή } t \geq 0 \quad (2.2)$$

Στην προσπάθεια τους οι ερευνητές, να διορθώσουν το πρόβλημα των ‘νεκρών’

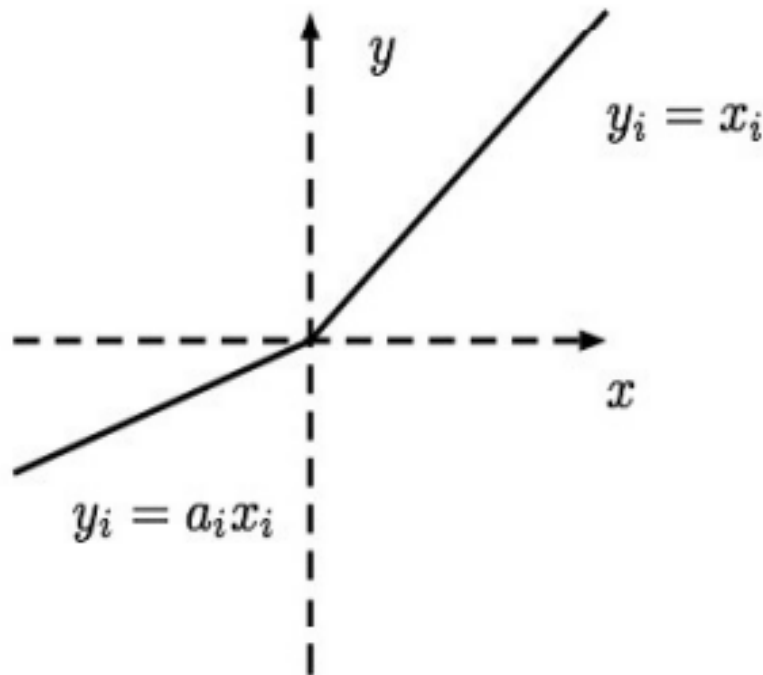
Σχήμα 2.6: Η συνάρτηση Ράμπας (ReLU)



νευρώνων, που παρουσιάστηκε στη συνάρτηση ενεργοποίησης της ράμπας, οδηγήθηκαν σε μία μικρή και συγχρόνως αποτελεσματική τροποποίηση του τύπου της συνάρτησης. Έτσι, καταφέρανε να αντισταθμίσουν την ύπαρξη του αρνητικού αθροίσματος S που παρουσιάστηκε στη συνάρτηση ράμπας, προσθέτοντας στο $\max(0, t)$ του ήδη υπάρχοντος τύπου το $\min(0, t)$ πολλαπλασιασμένο με μία επιλεγμένη σταθερά a . Η ποσότητα του a , προτάθηκε για πρώτη φορά ως μια επιλεγμένη σταθερά στο βιβλίο "Rectifier nonlinearities improve neural network acoustic models" των Maas, Hannun και Ng. Σε μία περαιτέρω προσπάθεια βελτίωσης του αλγόριθμου της διαδικασίας μάθησης, μία νέα ομάδα ερευνητών παρουσίασε το a ως παράμετρο μάθησης του νευρωνικού δικτύου. Σύμφωνα με τους συγγραφείς της διαφορετικής αυτής ιδέας, θεωρήθηκε ότι με τη χρήση της ποσότητας του a ως παράμετρο, βελτιώνεται κατά πολύ η διαδικασία μάθησης. Η χρήση της παραμέτρου a προσθέτει ένα μεγάλο σύνολο παραμέτρων προς μάθηση.

Γενικά, υπάρχουν δύο μεγάλες κατηγορίες συναρτήσεων ενεργοποίησης, οι οποίες είναι οι Γραμμικές (Linear) και οι Μη-Γραμμικές (Non-Linear). Γραμμικές, ονομάζονται οι συναρτήσεις, των οποίων οι γραφικές παραστάσεις, αποτυπώνονται με μια ευθεία γραμμή. Από την άλλη, Μη Γραμμικές, ονομάζονται οι συναρτήσεις των οποίων οι γραφικές παραστάσεις, είναι σύνθετα σχήματα. Έτσι, λοιπόν, οι συναρτήσεις των παραπάνω παραδειγμάτων, άλλες είναι γραμμικές (Βηματική, Ράμπας, PReLU) και άλλες είναι Μη-Γραμμικές, όπως η Σιγμοειδής και η Υπερβολική Εφαπτομένη.

Σχήμα 2.7: Η συνάρτηση Leaky ReLU/PReLU



2.0.7 Αρχιτεκτονική Νευρωνικών Δικτύων

Ένα σύνολο νευρώνων, μαζί με τις συνάψεις τους, δομούν ένα βιολογικό νευρωνικό δίκτυο. Μέσα από έναν τόσο απλό ορισμό και με μια γρήγορη ματιά στα βιολογικά νευρωνικά δίκτυα, μπορεί κανείς να λάβει μία πρώτη εικόνα, από τον κόσμο των τεχνητών νευρωνικών δικτύων. Τα νευρωνικά δίκτυα, διαχωρίζονται με βάση τη λειτουργία τους, τη μέθοδο εκπαίδευσης και την αρχιτεκτονική τους. Ο όρος αρχιτεκτονική ενός δικτύου, ορίζει με ακρίβεια τον τρόπο διάταξης των συνδέσεων των νευρώνων, όπως επίσης και τον αριθμό και τον τύπο τους. Σε ένα νευρωνικό δίκτυο, οι νευρώνες οργανώνονται σε μορφή επιπέδων και δομούνται με βάση τον αλγόριθμο μάθησης, που χρησιμοποιείται κατά την εκπαίδευση του εκάστοτε δικτύου. Τελικά, προκύπτουν οι παρακάτω κατηγορίες αρχιτεκτονικών δικτύων:

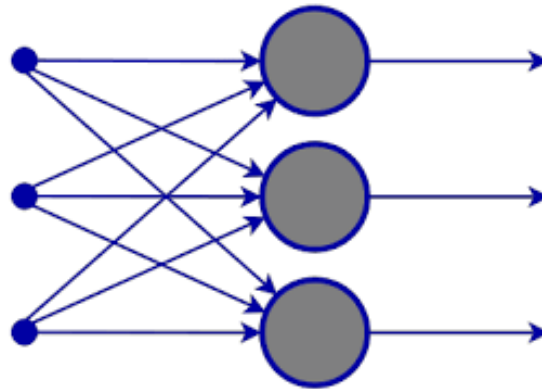
1. **Δίκτυα πρόσθιας τροφοδότησης ενός επιπέδου (feedforward):**

Τα δίκτυα αυτά, αποτελούν την πιο απλή μορφή που ένα δίκτυο μπορεί να πάρει. Η δομή τους, απαρτίζεται από ένα επίπεδο εισόδου, αποτελούμενο από πηγαίους κόμβους και από ένα επίπεδο νευρώνων εξόδου. Το ηλεκτρικό σήμα, που εισέρχεται στο επίπεδο εισόδου, μεταφέρεται μονόδρομα και προς το επίπεδο νευρώνων εξόδου. Ο ρόλος των νευρώνων εισόδου του δικτύου, περιορίζεται στη μεταφορά του σήματος στο επίπεδο εξόδου, χωρίς να πραγματοποιούν κάποια επεξεργασία στην πληροφορία που διαδίδεται. Ο χαρακτηρισμός 'ενός επιπέδου' προσανατολίζεται στο επίπεδο εξόδου, όπου και υλοποιείται ο υπολογισμός της τελικής τιμής, που το επίπεδο εξάγει. Η περιθωριοποίηση από τον χαρακτηρισμό

αυτόν, του επιπέδου εισόδου, είναι πλήρως δικαιολογημένη, καθώς ο μοναδικός του ρόλος είναι η απλή μεταφορά του σήματος.

Στο σχήμα 2.8 παρουσιάζεται ένα δίκτυο πρόσθιας τροφοδότησης ενός επιπέδου :

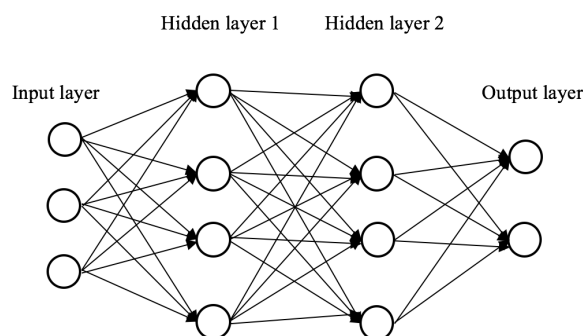
Σχήμα 2.8: Δίκτυο πρόσθιας τροφοδότησης ενός επιπέδου



2. Πολυεπίπεδα δίκτυα πρόσθιας τροφοδότησης :

Η ιδέα των πολυεπίπεδων δικτύων, δομήθηκε σε απόλυτη συνάφεια με την αντίστοιχη του δικτύου ενός επιπέδου, που μελετήσαμε παραπάνω, πραγματοποιώντας μια προσθήκη σε αυτό. Πιο συγκεκριμένα, ένα πολυεπίπεδο δίκτυο, αποτελείται από τα επίπεδα εισόδου και εξόδου, με την ενσωμάτωση ενός ή περισσότερων κρυφών επιπέδων. Οι νευρώνες των κρυφών επιπέδων, ονομάζονται κρυφοί. Σε κάθε επίπεδο, η είσοδος των νευρώνων, είναι τα σήματα εξόδου του προηγούμενου επιπέδου. Το σήμα, διαδίδεται μέσα στο δίκτυο, από το επίπεδο εισόδου και μέσω των κρυφών επιπέδων, προς το τελικό επίπεδο εξόδου. Όμοια με το δίκτυο πρόσθιας τροφοδότησης ενός επιπέδου, η λειτουργία των επιπέδων εισόδου, περιορίζεται στη μεταφορά του σήματος στα επόμενα επίπεδα, χωρίς να πραγματοποιούν καμία επεξεργασία. Αντίθετα, οι κρυφοί νευρώνες και οι νευρώνες εξόδου, είναι οι υπολογιστικοί νευρώνες, που το μοντέλο ακολουθεί.

Σχήμα 2.9: Δίκτυο πρόσθιας τροφοδότησης πολλών επιπέδων



2.0.8 Διαδικασίες Μάθησης των ΤΝΔ

Πρωταρχικός στόχος της λειτουργίας των τεχνητών νευρωνικών δικτύων, είναι να μπορούν να αντιμετωπίζουν συγκεκριμένα προβλήματα, που τους ανατίθενται ή να επιτελούν από μόνα τους ορισμένες διεργασίες, λόγω χάριν η αναγνώριση εικόνων. Ένα δίκτυο για να κατορθώσει να κατακτήσει την ικανότητα αυτή, επιβάλλεται να έχει εκπαιδευτεί κατάλληλα. Οι διαδικασίες της μάθησης ή της εκπαίδευσης, αποτελούν και το κύριο χαρακτηριστικό των νευρωνικών δικτύων. Τα ΤΝΔ, εκτελούν δύο βασικές λειτουργίες αυτές της μάθησης και της ανάκλασης. Με τον όρο ανάκλαση, ορίζεται η διαδικασία του υπολογισμού ενός διανύσματος εξόδου, για δεδομένο διάνυσμα εισόδου και τιμές συναπτικών βαρών. Από την άλλη, μάθηση ονομάζεται η διαδικασία, σύμφωνα με την οποία πραγματοποιείται η τροποποίηση της τιμής των βαρών, με σκοπό, δοθέντος συγκεκριμένης εισόδου, να παραχθεί συγκεκριμένη έξοδος. Μέσα από τη μάθηση, επιδιώκεται η συνεχής βελτίωση, της απόδοσης των νευρωνικών δικτύων. Έτσι, ένα δίκτυο φτάνει στην κατάκτηση της μάθησης, με τη χρήση παραδειγμάτων εκπαίδευσης και ενός αλγόριθμου εκπαίδευσης. Υπάρχει πληθώρα αλγόριθμων μέσω των οποίων, ένα δίκτυο επέρχεται σε σταδιακή μάθηση, όπως οι αλγόριθμοι οπισθοδρόμησης (Back Propagation), Ομαδοποίησης (Clustering) και Decision Tree. Οι αλγόριθμοι μάθησης, είναι επαναληπτικές διαδικασίες, με στόχο τη μείωση του σφάλματος, μεταξύ της επιθυμητής και της πραγματικής τιμής εξόδου, που το δίκτυο παράγει. Για την επίτευξη αυτού του στόχου, τα νευρωνικά δίκτυα εκπαιδεύονται με διάφορες μεθόδους, ώστε να αλλάζουν τις τιμές των ελεύθερων παραμέτρων του δικτύου (βάρη, κατώφλια).

Έτσι, τα ΤΝΔ ανάλογα με τους διαφορετικούς τρόπους αλλαγής των παραμέτρων που χρησιμοποιούνται, χωρίζονται σε 2 κατηγορίες μάθησης:

- (α') Επιβλεπόμενη Μάθηση (Supervised Learning)
- (β') Μη-Επιβλεπόμενη Μάθηση (Unsupervised Learning)

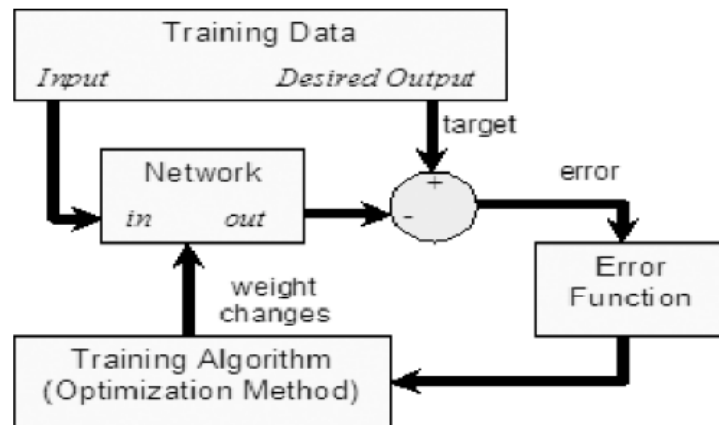
Επιγραμματικά, η διαδικασία της μάθησης που ακολουθεί ένα ΤΝΔ, αποτελείται από τα ακόλουθα βήματα:

- Αρχικά το ΤΝΔ, διεγείρεται από το περιβάλλον του.
- Στη συνέχεια, το ΤΝΔ ως αποτέλεσμα αυτής της διέγερσης, υλοποιεί μεταβολές στις παραμέτρους του.
- Τέλος, το ΤΝΔ, αλληλεπιδρά με το νέο περιβάλλον του με καινούριο τρόπο, λόγω των μεταβολών που επήλθαν, μέσω των αλλαγών των παραμέτρων του

Παρακάτω, περιγράφονται οι κατηγορίες μάθησης και ο τρόπος λειτουργίας τους:

(α') Η εποπτευόμενη μάθηση, αποτελεί την πιο διαδεδομένη μέθοδο, στην εκπαίδευση των νευρωνικών δικτύων. Στην επιβλεπόμενη μάθηση, ο 'εκπαιδευτής', παραθέτει στο δίκτυο τις τιμές των εισόδων και των στόχων, που επιθυμεί να μάθει το δίκτυο. Οι τιμές των στόχων επίτευξης και των εισόδων τους, αποτελούν τα λεγόμενα πρότυπα του εκπαιδευσιμου νευρωνικού δικτύου. Αρχικά, οι τιμές των συναπτικών βαρών w , είναι τυχαίες για τις τιμές των εισόδων. Κατά τη διαδικασία της εκπαίδευσης του δικτύου, οι τιμές των βαρών αλλάζουν, με σκοπό να περιοριστεί τελικά το σφάλμα (η διαφορά), που δημιουργείται μεταξύ της εξόδου στόχου και της τρέχουσας τιμής εξόδου. Στην προσπάθεια βελτιστοποίησης της αποδοτικότητας του δικτύου, επιβάλλεται μεγάλος αριθμός διαρθρώσεων στα βάρη και κατέπείκταση, απαιτείται μεγάλος υπολογιστικός χρόνος. Συνοψίζοντας, κατά τη μάθηση ενός δικτύου με επίβλεψη, ένας 'εκπαιδευτής' διεγείρει το περιβάλλον του δικτύου (βήμα 1), ωθώντας το σύστημα ουσιαστικά να τροποποιηθεί (βήμα 2), έτσι ώστε να φτάσει στη μέγιστη απόδοση όπου το δίκτυο είναι ικανό να φτάσει. (βήμα 3)

Σχήμα 2.10: Σχηματική πορεία της Επιβλεπομένης Μάθησης



(β') Στο μοντέλο της μη-επιβλεπόμενης μάθησης, καταργείται ο ρόλος του 'εκπαιδευτή', όπως καθιερώθηκε στο αντίστοιχο πρότυπο της επιβλεπόμενης μάθησης. Σε αντίθεση με τη διαδικασία της επιβλεπόμενης μάθησης, εδώ εισάγεται η πληροφορία στο δίκτυο, χωρίς όμως να προσφέρονται οι στόχοι εξόδων. Ως αποτέλεσμα αυτής της διαφοροποίησης, στη διαδικασία της εκπαίδευσης, δεν πραγματοποιείται κανένας έλεγχος ή σύγκριση για την πορεία του σφάλματος. Το δίκτυο δε χρησιμοποιεί καμία εξωτερική παράμετρο για την αλλαγή των βαρών, αλλά ακολουθεί μία συγκεκριμένη διαδικασία, η οποία οδηγεί στην εκπαίδευση του δικτύου. Μέσω ενός εσωτερικού ελέγχου, το δίκτυο ψάχνει να βρει κάποιες τάσεις/μοτίβα, στα σήματα εισόδου και στοχεύει οι έξοδοι να παρουσιάζουν τα ίδια χαρακτηριστικά, με αυτά των ει-

σόδων. Με χρήση αυτής της τεχνικής, το εκπαιδευόμενο δίκτυο, ελέγχει τον εαυτό του και διορθώνει στον πιο αποδοτικό βαθμό, τα σφάλματα στα δεδομένα, μέναν μηχανισμό ανάδρασης (feedback). Έτσι ουσιαστικά, το δίκτυο είναι αυτο-εκπαιδευόμενο, μεταβάλλοντας μόνο του τις τιμές των βαρών του.

2.0.9 Εφαρμογές των Νευρωνικών Δικτύων

Τα τελευταία χρόνια, τα τεχνητά νευρωνικά δίκτυα, βασισμένα στη δομή και τις λειτουργίες τους, κεντρίζουν ολοένα και περισσότερο το ενδιαφέρον των ερευνητών με τις χρησιμότητες τους. Σύμμαχοι στην ανάπτυξη των δικτύων, αποτελούν οι υπολογιστικές της ταχύτητες, η δυνατότητα αντιμετώπισης πολύπλοκων μη γραμμικών λειτουργιών, όπως επίσης και η ικανότητα αναγνώρισης σχέσεων, ανάμεσα σε ποσότητες, που είναι δύσκολο να μοντελοποιηθούν. Η χρησιμοποίηση των τεχνητών νευρωνικών δικτύων σε διάφορες εφαρμογές τους, προσδίδει ευκολία υλοποίησης, αρκετά μεγάλη αξιοπιστία στη λειτουργία τους, αλλά και άμεση απόκριση εάν έχουν υλοποιηθεί σε hardware. Αρκετές από αυτές τις εφαρμογές, υπάρχουν ήδη στην αγορά και είναι ιδιαίτερα διαδεδομένες στη χρήση τους. Μερικές από τις ικανότητες των νευρωνικών δικτύων, περιλαμβάνουν την αναγνώριση προτύπων, τον υπολογισμό συναρτήσεων, την πρόβλεψη, τη βελτιστοποίηση αλλά και τον αυτόματο έλεγχο. Παρακάτω, παρουσιάζονται μερικά πεδία εφαρμογής των ΤΝΔ, όπως η Οικονομία, η Επιστήμη των Υπολογιστών, Επεξεργασία Σημάτων και η Ιατρική. Οικονομία: Ένας τομέας που τα τεχνητά νευρωνικά δίκτυα έχουν μεγάλη χρησιμότητα, προσφέροντας λύσεις σε σοβαρά ζητήματα, είναι η Οικονομία. Συγκεκριμένα, βοηθούν στις τραπεζικές διεργασίες να υπολογισθεί ο παράγοντας επικινδυνότητας στην παροχή στεγαστικών δανείων. Μία τράπεζα προκειμένου να εμπιστευτεί σε έναν δανειολήπτη τα χρήματά της, πρέπει να είναι σίγουρη για την ικανότητα του να αποπληρώσει το δάνειο. Σε μια προσπάθεια να εκμηδενιστεί η πιθανότητα λάθους, η τράπεζα δημιουργεί διαγράμματα εισοδήματος-φερεγγυότητας προσαρμοσμένα στα συγκεκριμένα στοιχεία (οικονομική επιφάνεια), του υποψήφιου δανειζόμενου. Στα διαγράμματα αυτά, υπάρχουν δύο αντιδιαμετρικά αντίθετες περιοχές. Στη μία, ο δανειολήπτης έχει μεγάλο εισόδημα και είναι πολύ φερέγγυος, οπότε η απόφαση να του παραχωρηθεί το δάνειο, είναι προφανής, ενώ στην άλλη περιοχή, συμβαίνει ακριβώς το αντίθετο. Οι δύο αυτές περιοχές των διαγραμμάτων, χωρίζονται στη μέση, από μια γκρίζα περιοχή, όπου η λήψη απόφασης για την παροχή του δανείου ή όχι, είναι ιδιαίτερα πολύπλοκες. Στο σημείο αυτό, εισέρχονται τα νευρωνικά δίκτυα, για να δώσουν τη λύση. Ένα πρόγραμμα νευρωνικού δικτύου που λέγεται Νέστωρ (Nestor), εκπαιδεύεται σε χιλιάδες αιτήσεις δανείων, όπου οι μισές έχουν εγκριθεί από τα στελέχη της τράπεζας και οι άλλες μισές απορρίφθηκαν.

Το σύστημα, συγκρίνει τα δεδομένα μίας νέας αίτησης δανείου με τη βάση δεδομένων που έχει και αναζητεί να βρει στοιχεία για την επικινδυνότητα παροχής του δανείου ή όχι, λαμβάνοντας τελικά μια απόφαση. Φυσικά, το πρόγραμμα Νέστωρ, δεν είναι πάντα αποτελεσματικό, αλλά γενικά κρίνεται ιδιαίτερα επιτυχές για τις αποφάσεις που παίρνει. Επιστήμη των Υπολογιστών: Τα νευρωνικά δίκτυα παρουσιάζουν μία πολύ επιτυχημένη πορεία, στον τομέα της επεξεργασίας δεδομένων. Η λειτουργία που εμπίπτει σε αυτή την περιοχή, είναι η αναγνώριση προτύπων, εικόνων και κειμένων, γνωστή και ως αναγνώριση προτύπων (Pattern Recognition). Χαρακτηριστικό παράδειγμα αυτού του τρόπου λειτουργίας των νευρωνικών δικτύων, αποτελεί το πρόγραμμα Omnipage, που η εταιρεία Scansoft προώθησε το 1994. Το πρόγραμμα αυτό υλοποιείται σε έναν απλό ηλεκτρονικό υπολογιστή και προσανατολίζεται στην ανάγνωση κειμένων. Ένας σαρωτής, διαβάζει το δοθέν κείμενο και η εφαρμογή το μετατρέπει στον πρότυπο κώδικα, για ανταλλαγή πληροφοριών ASCII. Το Omnipage, παρουσιάζει ικανοποιητικά αποτελέσματα, ακόμα και αν τα γράμματα είναι αλλοιωμένα. Εκτός από την εφαρμογή της εταιρείας Scansoft, η εταιρεία Nestor, δημιούργησε ένα άλλο μοντέλο που αναγνωρίζει την ιαπωνική γραφή και τη μετατρέπει αυτόματα σε αγγλικά. Στην αρχική της εφαρμογή, το πρόγραμμα ήτανε ικανό να αναγνωρίζει 2500 λέξεις, με επιτυχία 92%. Τέλος, σταδιακά το πρόγραμμα αυτό, αναπτύχθηκε και για άλλες γραφές, όπως το κυριλλικό και το εβραϊκό αλφάβητο. Στην επιστήμη των υπολογιστών επίσης, παρουσιάζεται λύση, για το γνωστό πρόβλημα της μετατροπής κειμένου σε φωνή και το αντίστροφο. Με τη χρήση των νευρωνικών δικτύων, γεννήθηκε ένα καινοτόμο πρόγραμμα το NETtalk, που προσφέρει ακριβώς αυτές τις δυνατότητες. Το πρόγραμμα αυτό, αποτελείται από ένα δίκτυο με 309 νευρώνες, 18629 συνάψεις και 3 επίπεδα λειτουργίας. Στα πρώτα στάδια ανάπτυξης αυτής της εφαρμογής, η ομιλία ήτανε ακατανόητη, με την πάροδο των χρόνων όμως, έφτασε στο επίπεδο να είναι κατανοητή και αναγνωρίσιμη σε ποσοστό 95%. Μία πιο σύγχρονη εφαρμογή των νευρωνικών δικτύων στην επιστήμη των υπολογιστών αποτελεί η ανάπτυξη της υπολογιστικής πλατφόρμας της Google : Google Neural Machine Translation (GNMT). Η εφαρμογή αυτή παρουσιάστηκε τον Νοέμβριο του 2016 και χρησιμοποιεί νευρωνικά δίκτυα με σκοπό να αυξήσει την ακρίβεια και την ευχέρεια του Google Translate στη μετάφραση όχι μόνο λέξεων, αλλά και ολόκληρων προτάσεων ή και κειμένων. Επεξεργασία Σημάτων: Μία ιδιαίτερα χρήσιμη εφαρμογή των νευρωνικών δικτύων στην ανθρώπινη καθημερινότητα, καθίσταται η δυνατότητα που προσφέρουν, να καθαρίζει η τηλεφωνική γραμμή από παρεμβολές (ηχώ). Για τη λειτουργία αυτή, τοποθετήθηκε ένα φίλτρο, το οποίο συγχρόνως περιορίζει τα σφάλματα. Το πρώτο τέτοιο φίλτρο, εφευρέθηκε από τον ερευνητή B.Widrow και ονομάζεται Adaline. Η εφαρμογή αυτή, δομείται με βάση το μοντέλο του απλού αισθητήρα (Perceptron).

Ιατρική: Τα νευρωνικά δίκτυα ακόμα, παρουσιάζουν λύσεις σε μία πολύ σημαντική επι-
στήμη για την ανθρωπότητα, αυτή της Ιατρικής. Πιο συγκεκριμένα, χρησιμοποιούνται
με μεγάλη επιτυχία στην εκτέλεση δύσκολων εγχειρήσεων, αλλά και για την ανάγνωση
ακτινογραφιών. Επιπρόσθετα, αποτελούν εξαιρετικό εργαλείο στα χέρια των καρδιο-
λόγων, καθώς έχουν καταφέρει να αναλύουν τα επίπεδα των ενζύμων της καρδιάς, όπως
επίσης και να αναγνωρίζουν διάφορα καρδιακά συμπτώματα. Η έγκυρη διάγνωσή τους,
είναι σύμμαχος στην αντιμετώπισή τους, που πολλές φορές σώζει ζωές. Αξιοσημείωτη
λειτουργία των δικτύων αυτών, καθίσταται επίσης η χρήση τους στην ανάλυση ομιλίας
για τη δημιουργία ακουστικών βαρηκοΐας. Κλείνοντας, μία πολύ σημαντική εφαρμο-
γή των ΤΝΔ στην Ιατρική, είναι η ικανότητα να προβλέπει τις αντιδράσεις οργανισμών
σε διάφορες φαρμακευτικές αγωγές. Με οδηγό όλες τις παραπάνω ικανότητες και τα
πλεονεκτήματα, που προσφέρει η χρήση των τεχνητών νευρωνικών δικτύων, γίνεται κα-
τανοητό ότι η επιστημονική αυτή περιοχή, θα απασχολήσει σημαντικά τον άνθρωπο και
την καθημερινή του ζωή.

Κεφάλαιο 3

Physics Informed Neural Networks (PINNs)

3.0.1 Εισαγωγή στα PINN

Στο κεφάλαιο 2 αναφέρθηκαν λεπτομερώς τα πεδία εφαρμογής των τεχνητών νευρωνικών δικτύων καθώς και ο βασικός τρόπος λειτουργίας τους. Σε αυτό το κεφάλαιο θα αναλυθεί πώς τα νευρωνικά δίκτυα μπορούν να επιλύσουν ζητήματα φυσικής και μαθηματικών. Οι επιστήμονες διαρκώς αναζητούν νέους και ταυτόχρονα αποδοτικούς τρόπους να μοντελοποιήσουν το κόσμο μέσω μαθηματικών έτσι ώστε να μπορέσουν να τον μελετήσουν λεπτομερώς. Για παράδειγμα, η γεννήσεις ενός πληθυσμού καθώς και ο θάνατος τους επηρεάζεται από εξωτερικούς παράγοντες. Αν κάποιος καταφέρει να συσχετίσει με μαθηματικό τρόπο αυτά τα δύο τότε μπορεί να κάνει μακροχρόνιες προβλέψεις για το μέλλον συγκεκριμένων πληθυσμών στους διάφορους πλανήτες. Επίσης, ένας πυρηνικός φυσικός γνωρίζοντας πώς αλληλεπιδρά το καύσιμο ενός πυρηνικού αντιδραστήρα με τη θερμοκρασία που περιβάλλοντος, θα μπορεί να διατηρεί την ασφαλή λειτουργία του πυρηνικού εργοστασίου στα επιθυμητά επίπεδα. Αυτά και πολλά παραπάνω, μοντελοποιούνται με χρήση συνήθων διαφορικών εξισώσεων. Οι διαφορικές εξισώσεις προκύπτουν σε πολλά πλαίσια των μαθηματικών και της επιστήμης (κοινωνική καθώς και φυσική). Οι σύγχρονες μαθηματικές περιγραφές αλλάζουν τη χρήση των διαφορικών συμβολισμών. Πολλοί μαθηματικοί έχουν μελετήσει τις διαφορικές εξισώσεις και συνέβαλαν στον τομέα αυτό, μερικοί από αυτούς ήταν οι Isaac Newton, Gottfried Wilhelm Leibniz, οικογένεια Bernoulli, Riccati, Clairaut, Jean le Rond d'Alembert και Leonhard Euler. Ένα απλό παράδειγμα είναι ο δεύτερος νόμος του Νεύτωνα - η σχέση ανάμεσα στη μετατόπιση x και το χρόνο t ενός αντικειμένου με δύναμη F δίνεται από τη διαφορική εξίσωση

$$m \frac{\partial^2 x(t)}{\partial t^2} = F(x(t)) \quad (3.1)$$

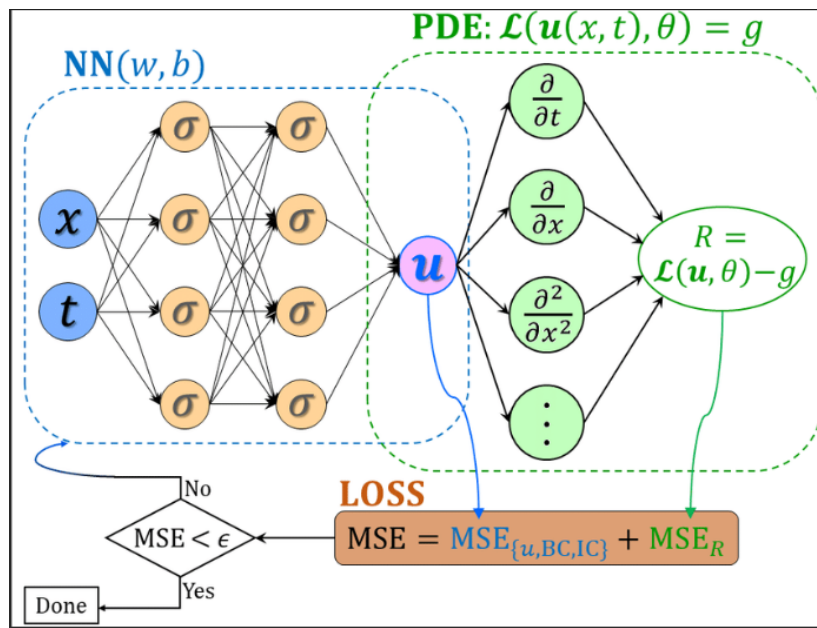
η οποία περιγράφει την κίνηση ενός σωματιδίου σταθερής μάζας m . Σε γενικές γραμμές, F είναι η συνάρτηση θέσης $x(t)$ του σωματιδίου στο χρόνο t . Η άγνωστη συνάρτηση $x(t)$ εμφανίζεται και στα δύο μέρη της διαφορικής εξίσωσης, και περιλαμβάνεται στην παράσταση $F(x(t))$. Οι περισσότεροι από τους φυσικούς νόμους που διέπουν τη δυναμική ενός συστήματος μπορούν να περιγραφούν με μερικές διαφορικές εξισώσεις. Για παράδειγμα, οι εξισώσεις Navier–Stokes [1] είναι ένα σύνολο μερικών διαφορικών εξισώσεων που προέρχονται από τους νόμους διατήρησης (δηλ. διατήρηση της μάζας, της ορμής και της ενέργειας) που διέπουν τη μηχανική των ρευστών. Η λύση των εξισώσεων Navier–Stokes με κατάλληλες αρχικές και οριακές συνθήκες επιτρέπει την ποσοτικοποίηση της δυναμικής ροής σε μια επακριβώς καθορισμένη γεωμετρία. Ωστόσο, αυτές οι εξισώσεις δεν μπορούν να λυθούν επακριβώς και επομένως πρέπει να χρησιμοποιηθούν αριθμητικές μέθοδοι (όπως πεπερασμένες διαφορές, πεπερασμένα στοιχεία και πεπερασμένοι όγκοι). Σε αυτή τη ρύθμιση, αυτές οι εξισώσεις που διέπουν πρέπει να επιλυθούν λαμβάνοντας υπόψη τις προηγούμενες παραδοχές, τη γραμμικοποίηση και την επαρκή διακριτοποίηση χρόνου και χώρου.

Πρόσφατα, η επίλυση των επιμέρους διαφορικών εξισώσεων των φυσικών φαινομένων με τη χρήση βαθιάς μάθησης έχει εμφανιστεί ως νέο πεδίο επιστημονικής μηχανικής μάθησης (SciML), αξιοποιώντας την καθολική προσέγγιση (universal approximation) [2] και την υψηλή απόδοση των νευρωνικών δικτύων. Γενικά, τα βαθιά νευρωνικά δίκτυα θα μπορούσαν να προσεγγίσουν οποιαδήποτε συνάρτηση υψηλών διαστάσεων δεδομένου ότι παρέχονται επαρκή δεδομένα εκπαίδευσης [3]. **Ωστόσο, τέτοια δίκτυα δε λαμβάνουν υπόψη τα φυσικά χαρακτηριστικά που κρύβουν το πρόβλημα και το επίπεδο ακρίβειας προσέγγισης που παρέχεται από αυτά εξακολουθεί να εξαρτάται σε μεγάλο βαθμό από τις προδιαγραφές της γεωμετρίας του προβλήματος καθώς και από τις αρχικές και οριακές συνθήκες. Χωρίς αυτές τις προκαταρκτικές πληροφορίες, η λύση δεν είναι μοναδική και μπορεί να χάσει τη φυσική ορθότητα.** Από την άλλη πλευρά, τα νευρωνικά δίκτυα με πληροφόρηση για τη φυσική (PINN) αξιοποιούν τις φυσικές εξισώσεις στην εκπαίδευση νευρωνικών δικτύων. Δηλαδή, τα PINN έχουν σχεδιαστεί για να εκπαιδεύονται ώστε να ικανοποιούν τα δεδομένα εκπαίδευσης καθώς και τις επιβαλλόμενες εξισώσεις που διέπουν. Με αυτόν τον τρόπο, ένα νευρωνικό δίκτυο μπορεί να καθοδηγηθεί με δεδομένα εκπαίδευσης που δε χρειάζεται απαραίτητα να είναι μεγάλα και πλήρη. Δυσνητικά, μια ακριβής λύση των μερικών διαφορικών εξισώσεων μπορεί να βρεθεί χωρίς να γνωρίζουμε τις οριακές συνθήκες [4]. Επομένως, με κάποια γνώση σχετικά με τα φυσικά χαρακτηριστικά του προβλήματος και κάποια μορφή δεδομένων εκπαίδευσης (ακόμη και αραιά και ελλιπή), το PINN μπορεί να χρησιμοποιηθεί για την εύρεση μιας βέλτιστης λύσης με υψηλή πιστότητα.

Τα PINN επιτρέπουν την αντιμετώπιση ενός ευρέος φάσματος προβλημάτων στην υπολο-

γιστική επιστήμη και αντιπροσωπεύουν μια πρωτοποριακή τεχνολογία που οδηγεί στην ανάπτυξη νέων κατηγοριών αριθμητικών επιλυτών για διαφορικές εξισώσεις (Partial Differential Equations (PDE)). Συγκεκριμένα, το εκπαιδευμένο δίκτυο PINN μπορεί να χρησιμοποιηθεί για την πρόβλεψη των τιμών σε πλέγματα προσομοίωσης διαφορετικών αναλύσεων χωρίς να χρειάζεται επανεκπαίδευση [5]. Επιπλέον, επιτρέπουν την εκμετάλλευση της αυτόματης διαφοροποίησης (AD) [6] για τον υπολογισμό των απαιτούμενων παραγώγων στις μερικές διαφορικές εξισώσεις, μια νέα κατηγορία τεχνικών διαφοροποίησης που χρησιμοποιούνται ευρέως για την εξαγωγή νευρωνικών δικτύων που αξιολογούνται ως ανώτερα από την αριθμητική ή συμβολική διαφοροποίηση.

Σχήμα 3.1: Σχηματική Αναπαράσταση ενός PINN



3.0.2 Ερευνητική δουλειά στα PINN

Στα σύγχρονα συγγράμματα ως η πρώτη ολοκληρωμένη πρόταση της χρήσης ΤΝΔ για την επίλυση διαφορικών εξισώσεων, συνοδευόμενη από υλοποίηση, αναγνωρίζεται η ως η εργασία με τίτλο "Artificial Neural Networks for Solving Ordinary and Partial Differential Equations" (Lagaris, Likas, & Fotiadis, 1997) [7], από το τμήμα Επιστήμης Ηλεκτρονικών Υπολογιστών του Πανεπιστημίου Ιωαννίνων. Βέβαια, είναι σημαντικό να αναφερθεί ότι η αρχική ιδέα της προσπάθειας να χρησιμοποιηθούν τα νευρωνικά δίκτυα για την επίλυση διαφορικών εξισώσεων είχε προταθεί από τις αρχές τις δεκαετίας του 1990 μέσα από θεωρητικές εργασίες όπου ορισμένες από αυτές αναφέρονται στη βιβλιογραφία. Ενδεικτικά ορισμένες από αυτές είναι η εργασία με τίτλο "Neural algorithms for solving differential equations" (Lee & Kang, 1990) [8] και "The numerical solution of Linear Ordinary Differential Equations by Feedforward Neural networks" (Meade Jr & Fernandez, 1994) [9]. Οι παραπάνω ερευνητές που αναφέρθηκαν προέρχονταν από

επιστημονικά πεδία κυρίως των εφαρμοσμένων μαθηματικών και της επιστήμης υπολογιστών. Η μέθοδος εισήχθη στο πεδίο της μηχανικής και συγκεκριμένα στο πεδίο της ανάλυσης κατασκευών μέσα από την εργασία με τίτλο "A neural network approach to the modelling, calculation and identification of semi-rigid connections in steel structures" (Stavrulakis, Avdelas, Abdalla, & Panagiotopoulos, 1997) [10]. Μια χαρακτηριστική εργασία η οποία παρουσιάζει τη μέθοδο PINNs με τη μορφή που έχει σήμερα, αποτελεί η εργασία με τίτλο "Physics Informed Deep Learning (Part I): Data - driven Solutions of Nonlinear Partial Differential Equations" (Raissi, Perdikaris, & Karniadakis, 2017) [11] στην οποία ορίζονται οι αρχές της μεθόδου, αναλύεται η προγραμματιστική υλοποίηση της και εφαρμόζεται στις μερικές διαφορικές εξισώσεις Burger's Equation και Schrödinger Equation. Η εργασία αυτή αποτελεί έναν πολύ καλό οδηγό για κάποιον που επιθυμεί να ξεκινήσει την ενασχόλησή του με το θέμα. Οι παραπάνω συγγραφείς έχουν συμμετάσχει στη δημοσίευση αρκετών εργασιών τα τελευταία χρόνια, όπως επίσης φαίνεται ότι ασχολούνται συστηματικά με τη μέθοδο και τις δυνατότητές της, ενώ δηλώνουν ιδιαίτερα αισιόδοξοι για την πορεία της. Πέρα όμως από τους ερευνητές που συναντάει κανείς πιο συχνά κατά τη μελέτη της μεθόδου, παράλληλα έχουν δημοσιευτεί και ορισμένες επιπλέον εργασίες. Ενδεικτικά μια ακόμα εργασία πάνω στο θέμα έχει τίτλο "IDRLnet: A Physics-Informed Neural Network Library" (Peng, και συν., 2021) [12] στην οποία μπορεί ένας αναγνώστης να βρει πιο λεπτομερείς πληροφορίες για την υλοποίηση της μεθόδου. - Physics Informed Neural Networks (PINNs)

Η μελέτη της μεθόδου φαίνεται να παρουσιάζει ιδιαίτερη ακμή από το 2017 και μετά, καθώς πέρα των χαρακτηριστικών εργασιών που αναφέρθηκαν παραπάνω, παράλληλα έχουν δημοσιευτεί αρκετές επιπλέον, από ερευνητές από όλο τον κόσμο. Ως πιο πρόσφατες εργασίες του τελευταίου χρόνου ενδεικτικά είναι η εργασία με τίτλο "Physics-informed machine learning" (Karniadakis, Kevrekidis, Lu, & Perdikaris, 2021) και "Physics-Informed Neural Networks for elastic plate problems" (Muradova & Stavrulakis, 2021) [13] όπου στην εργασία αυτή, εφαρμόζεται η μέθοδος για την επίλυση της διαφορικής εξίσωσης που αφορά το πρόβλημα της πλάκας υπό κάμψη (Kirchhoff Plate Bending). Όλες οι παραπάνω εργασίες και ορισμένες επιπλέον αποτέλεσαν πηγή πληροφόρησης για την κατανόηση της μεθόδου, την εφαρμογή της και για τη συγγραφή της παρούσας εργασίας.

3.0.3 Μοντελοποίηση και Υπολογισμός

Μια γενική μη γραμμική μερική διαφορική εξίσωση μπορεί να είναι:

$$u_t + N[u; \mathcal{I}] = 0, \quad x \in \Omega, \quad t \in [0, T],$$

που $u(t, x)$ υποδεικνύει τη λύση, $N[.; \mathcal{I}]$ είναι ένας μη γραμμικός τελεστής με παρα-

μέτρους \mathcal{I} , και Ω είναι ένα υποσύνολο του \mathbb{R}^D . Αυτή η γενική μορφή εξισώσεων συνοψίζει ένα ευρύ φάσμα προβλημάτων στη μαθηματική φυσική, όπως οι κινητικές εξισώσεις. Δεδομένου του θορύβου των μετρήσεων ενός γενικού δυναμικού συστήματος που περιγράφεται από την παραπάνω εξίσωση, τα PINN μπορούν να σχεδιαστούν για την επίλυση δύο κατηγοριών προβλημάτων:

- λύσεις οδηγούμενες από δεδομένα
- εντοπισμός οδηγούμενος από δεδομένα

των μερικών διαφορικών εξισώσεων.

Επίλυση μερικών διαφορικών εξισώσεων οδηγούμενες από δεδομένα

Η λύση PDE [11] που βασίζεται σε δεδομένα έχει ως αποτέλεσμα τον υπολογισμό της άγνωστης κατάστασης $u(t, x)$ του συστήματος, δεδομένου ότι υπάρχουν μετρήσεις z και συγκεκριμένοι παράμετροι \mathcal{I} .

$$u_t + N[u] = 0, \quad x \in \Omega, \quad t \in [0, T],$$

Ορίζουμε το $f(t, x)$ ως:

$$f := u_t + N[u] = 0,$$

και χρησιμοποιούμε νευρωνικό δίκτυο έτσι ώστε να ελαχιστοποιήσουμε τη συνάρτηση κόστους L_{tot} :

$$L_{tot} = L_u + L_f$$

Όπου, $L_u = u - z$, όπου u , η έξοδος του νευρωνικού δικτύου και όπου z , οι πραγματικές μετρήσεις. Αντίστοιχα όπου L_f ορίζεται το μέσο τετραγωνικό σφάλμα που προκύπτει από την αντικατάσταση των τιμών της εξόδου του νευρωνικού δικτύου στη διαφορική εξίσωση.

Αυτή η στρατηγική επιτρέπει τη συνένωση υπολογιστικά αποδοτικών μοντέλων βασισμένων στη φυσική, τα οποία μπορούν να βρουν εφαρμογή στην πρόβλεψη φυσικών διεργασιών βάσει δεδομένων [14].

Ανακάλυψη μερικών διαφορικών εξισώσεων βάσει δεδομένων

Δεδομένων των ελλιπών μετρήσεων z της κατάστασης του συστήματος, η βασισμένη σε δεδομένα ανακάλυψη της PDE [15] έχει ως αποτέλεσμα τον υπολογισμό της άγνωστης κατάστασης $u(t, x)$ και παραμέτρους μοντέλου μάθησης \mathcal{I} που περιγράφουν δεδομένα που έχουν παρατηρηθεί. Η διαφορική εξίσωση είναι η εξής:

$$u_t + N[u; \mathcal{I}] = 0, \quad x \in \Omega, \quad t \in [0, T],$$

Ορίζουμε ως $f(t, x)$ το,

$f := u_t + N[u; \mathcal{H}] = 0$, και χρησιμοποιούμε νευρωνικό δίκτυο έτσι ώστε να ελαχιστοποιήσουμε τη συνάρτηση κόστους L_{tot} :

$$L_{tot} = L_u + L_f$$

Όπου, $L_u = u - z$, όπου u , η έξοδος του νευρωνικού δικτύου και όπου z , οι πραγματικές μετρήσεις. Αντίστοιχα όπου L_f ορίζεται το μέσο τετραγωνικό σφάλμα που προκύπτει από την αντικατάσταση των τιμών της εξόδου του νευρωνικού δικτύου στη διαφορική εξίσωση.

Αυτή η στρατηγική επιτρέπει την ανακάλυψη δυναμικών μοντέλων που περιγράφονται από μη γραμμικά PDE που συναρμολογούν υπολογιστικά αποδοτικά και πλήρως διαφορίσιμα υποκατάστατα μοντέλα που μπορούν να βρουν εφαρμογή σε προβλήματα πρόβλεψης, τον έλεγχο και την κατανόηση δεδομένων. [16] [17] [18]

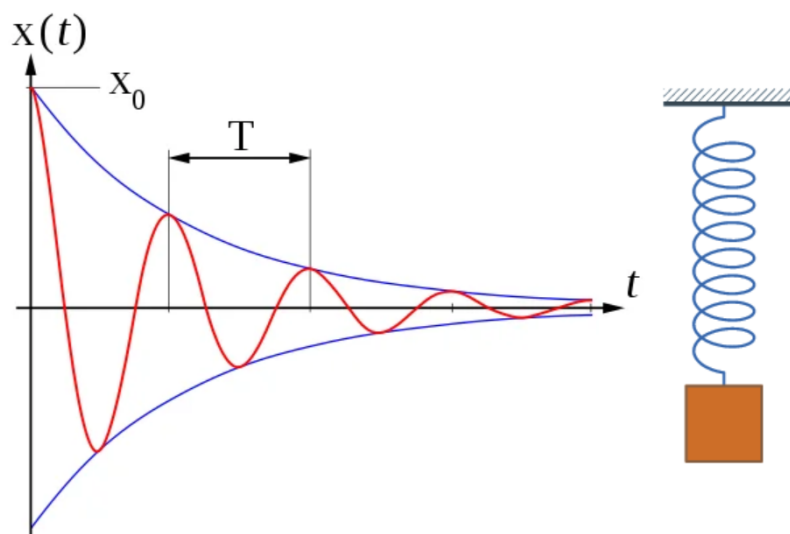
Κεφάλαιο 4

Ορισμός προβλήματος του Αποσβενόμενου Αρμονικού Ταλαντωτή

4.0.1 Αποσβενόμενος Αρμονικός Ταλαντωτής

Ο αποσβενόμενος αρμονικός ταλαντωτής είναι ένα κλασικό πρόβλημα στη μηχανική. Περιγράφει την κίνηση ενός μηχανικού ταλαντωτή (π.χ. εκκρεμές ελατηρίου) υπό την επίδραση μιας δύναμης επαναφοράς και της τριβής. Αυτό το υπό κεφάλαιο ασχολείται με την εξαγωγή της εξίσωσης ταλάντωσης για τον αποσβενόμενο ταλαντωτή.

Σχήμα 4.1: Σχηματική αναπαράσταση αποσβενόμενου αρμονικού ταλαντωτή. πηγή : <https://beltoforion.de/en/harmonicoscillator/>

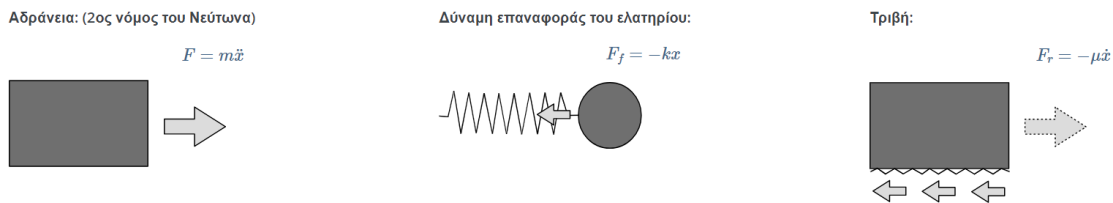


4.0.2 Η εξίσωση της κίνησης

Η εξίσωση κίνησης είναι μια μαθηματική εξίσωση που περιγράφει πλήρως τη χωρική και χρονική ανάπτυξη ενός δυναμικού συστήματος υπό την επίδραση εξωτερικών δυνάμεων. Κατά κανόνα, πρόκειται για διαφορικές εξισώσεις δεύτερης τάξης. Σε αυτή την ενότητα, εξετάζουμε πρώτα τις δυνάμεις που επηρεάζουν την κίνηση του εκκρεμο-ύς.

Η κίνηση του εκκρεμούς εξαρτάται από την ισορροπία τριών δυνάμεων: της αδράνειας, της δύναμης επαναφοράς του ελατηρίου και της τριβής.

Σχήμα 4.2: Σχηματική αναπαράσταση αποσβεσμένου αρμονικού ταλαντωτή. πηγή : <https://beltoforion.de/en/harmonicoscillator/>



Ο δεύτερος νόμος του Νεύτωνα δηλώνει ότι οποιαδήποτε αλλαγή στην κίνηση ενός σώματος με μάζα m είναι ανάλογη με τη δύναμη που ασκείται σε αυτό και ότι αυτή η αλλαγή στην κίνηση λαμβάνει χώρα πάντα στην ίδια κατεύθυνση στην οποία δρα η δύναμη.

Σε ένα εκκρεμές έχουμε μια δύναμη επαναφοράς που δρα στον ταλαντωτή. Τραβάει τη μάζα του εκκρεμούς προς τη θέση ηρεμίας. Αυτή η δύναμη είναι ευθέως ανάλογη με την απόκλιση του εκκρεμούς. Η διεύθυνση του είναι αντίθετη από την κατεύθυνση της απόκλισης του εκκρεμούς. Η δύναμη αυτή του υπολογίζεται πολλαπλασιάζοντας την απόσταση απόκλισης με μια σταθερά ελατηρίου k που εξαρτάται από το υλικό.

Τέλος, το εκκρεμές θα αντιμετωπίσει τριβή. Η δύναμη τριβής είναι ανάλογη με την ταχύτητα του εκκρεμούς. Η διεύθυνση του είναι αντίθετη από τη φορά κίνησης του εκκρεμούς. Η παράμετρος μ είναι ένας συντελεστής τριβής που εξαρτάται από το υλικό και το σχήμα. Η αδρανειακή δύναμη επομένως αντιτίθεται από δύο δυνάμεις:

$$m \frac{d^2x}{dt^2} = -\mu \frac{dx}{dt} - kx \quad (4.1)$$

που μπορεί να μετατραπεί σε:

$$\frac{d^2x}{dt^2} + \frac{\mu}{m} \frac{dx}{dt} + \frac{k}{m} x = 0 \quad (4.2)$$

Αυτή η εξίσωση είναι μια γραμμική, ομοιογενής, διαφορική εξίσωση δεύτερης τάξης με

σταθερούς συντελεστές. Ο προσδιορισμός του τύπου της διαφορικής εξίσωσης είναι σημαντικός επειδή η προσέγγιση που είναι απαραίτητη για την επίλυσή της εξαρτάται από αυτό. Το εκθετικό ansatz επιλέγεται συνήθως για την επίλυση διαφορικών εξισώσεων αυτού του τύπου.

4.0.3 Ο Εκθετικός Ansatz

Η εξίσωση είναι μια ομοιογενής γραμμική διαφορική εξίσωση δεύτερης τάξης με σταθερούς συντελεστές. Η υπόθεση είναι ότι η λύση σε μια τέτοια διαφορική εξίσωση είναι μια εκθετική συνάρτηση. Αυτή η μέθοδος ήταν η πρώτη η οποία περιγράφεται από τον Γερμανό μαθηματικό Leonard Euler και ονομάζεται 'εκθετικός Ansatz'. 4.2

Ο εκθετικός ansatz δηλώνει ότι μια ειδική λύση στη διαφορική εξίσωση παίρνει τη μορφή:

$$x(t) = Ce^{\hat{\lambda}t} \quad (4.3)$$

Σχηματίζοντας την πρώτη και τη δεύτερη παράγωγο του 4.3 παίρνουμε:

$$\frac{dx}{dt} = \hat{\lambda}Ce^{\hat{\lambda}t}, \quad \frac{dx^2}{dt} = \hat{\lambda}^2Ce^{\hat{\lambda}t} \quad (4.4)$$

Εισάγοντας τις εξισώσεις 4.3 και 4.4 στη διαφορική εξίσωση 4.2 παίρνουμε:

$$\hat{\lambda}^2Ce^{\hat{\lambda}t} + \frac{\mu}{m}\hat{\lambda}Ce^{\hat{\lambda}t} + \frac{k}{m}Ce^{\hat{\lambda}t} = 0$$

Διαιρώντας την πιο πάνω εξίσωση με $Ce^{\hat{\lambda}t}$ απλοποιείται σε:

$$\hat{\lambda}^2 + \frac{\mu}{m}\hat{\lambda} + \frac{k}{m} = 0 \quad (4.5)$$

Αυτή η εξίσωση ονομάζεται επίσης χαρακτηριστική εξίσωση. Είναι μια τετραγωνική εξίσωση σε κανονική μορφή. Ως εκ τούτου, υπάρχουν οι ακόλουθες δύο λύσεις για $\hat{\lambda}$:

$$\hat{\lambda}_{1,2} = -\frac{\mu}{2m} \pm \sqrt{\left(\frac{\mu}{2m}\right)^2 - \frac{k}{m}} \quad (4.6)$$

Για περαιτέρω απλοποίηση εισάγουμε τις νέες σταθερές δ , ω_0 όπου:

$$\delta = \frac{\mu}{2m}, \quad \omega_0 = \sqrt{\frac{k}{m}}$$

Ακολουθώντας εισάγοντας τις σταθερές δ και ω_0 στην εξίσωση 4.6 παίρνουμε την απλοποι-

ημένη :

$$\lambda_{1,2} = -\delta \pm \sqrt{\delta^2 - \omega_0^2} \quad (4.7)$$

Η πιο πάνω λύση εξαρτάται από τις τιμές του λ_1 και λ_2 . Ο όρος μέσα στην τετραγωνική ρίζα ονομάζεται διακρίνουσα. Ανάλογα με την επιλογή των σταθερών δ και ω_0 η διακρίνουσα μπορεί να είναι μεγαλύτερη από 0, μικρότερη από 0 ή ίση με 0. Ως εκ τούτου και μπορεί να είναι:

1. Δύο πραγματικά πολύτιμες λύσεις που διαφέρουν μεταξύ τους.
2. Δύο συζυγείς σύνθετες λύσεις.
3. Δύο πανομοιότυπες πραγματικά πολύτιμες λύσεις.

Κάθε μία από τις τρεις επιλογές απαιτεί διαφορετική προσέγγιση στη λύση. Η γενική λύση της ομοιογενούς διαφορικής εξίσωσης έχει τη μορφή:

$$x(t) = C_1 x_1(t) + C_2 x_2(t) \quad (4.8)$$

Οι συναρτήσεις $x_1(t)$ και $x_2(t)$ καθορίζονται από την τιμή της διακρίνουσας στην εξίσωση. Ας εξετάσουμε τις διάφορες περιπτώσεις όπου η διακρίνουσα μπορεί να είναι μεγαλύτερη από 0, μικρότερη από 0 ή ίση με 0.

4.0.4 Υπερκρίσιμη απόσβεση

Εάν $\delta > \omega_0$, έχουμε μια περίπτωση ισχυρής τριβής, τότε η διακρίνουσα στην εξίσωση είναι θετική και υπάρχουν δύο διαφορετικές πραγματικές λύσεις. Τότε η λύση της διαφορικής εξίσωσης μοιάζει με:

$$x_1(t) = C_1 e^{\lambda_1 t}, x_2(t) = C_2 e^{\lambda_2 t}$$

Εισάγοντας την πιο πάνω εξίσωση στην εξίσωση 4.8 λαμβάνουμε τη γενική λύση της διαφορικής εξίσωσης :

$$x(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}$$

Επίσης, η πιο πάνω διαφορική μπορεί να μετασχηματιστεί ως εξής εισάγοντας τα λ από την εξίσωση 4.7

$$x(t) = C_1 e^{(-\delta + \sqrt{\delta^2 - \omega_0^2}) t} + C_2 e^{(-\delta - \sqrt{\delta^2 - \omega_0^2}) t} \quad (4.9)$$

Επομένως, η εξίσωση 4.9 είναι μια προσθήκη δύο εκθετικά διασπώμενων συναρτήσεων, αυτά θα τα δούμε ξανά στο παρακάτω διάγραμμα. Για να απλοποιήσουμε την εξίσωση θα αντικαταστήσουμε την τετραγωνική ρίζα με μια νέα σταθερά:

$$a = \sqrt{\delta^2 - \omega_0^2} \quad (4.10)$$

Μετά την εισαγωγή της νέας σταθεράς στην εξίσωση 4.9 μπορούμε να ξαναγράψουμε την εξίσωση ως:

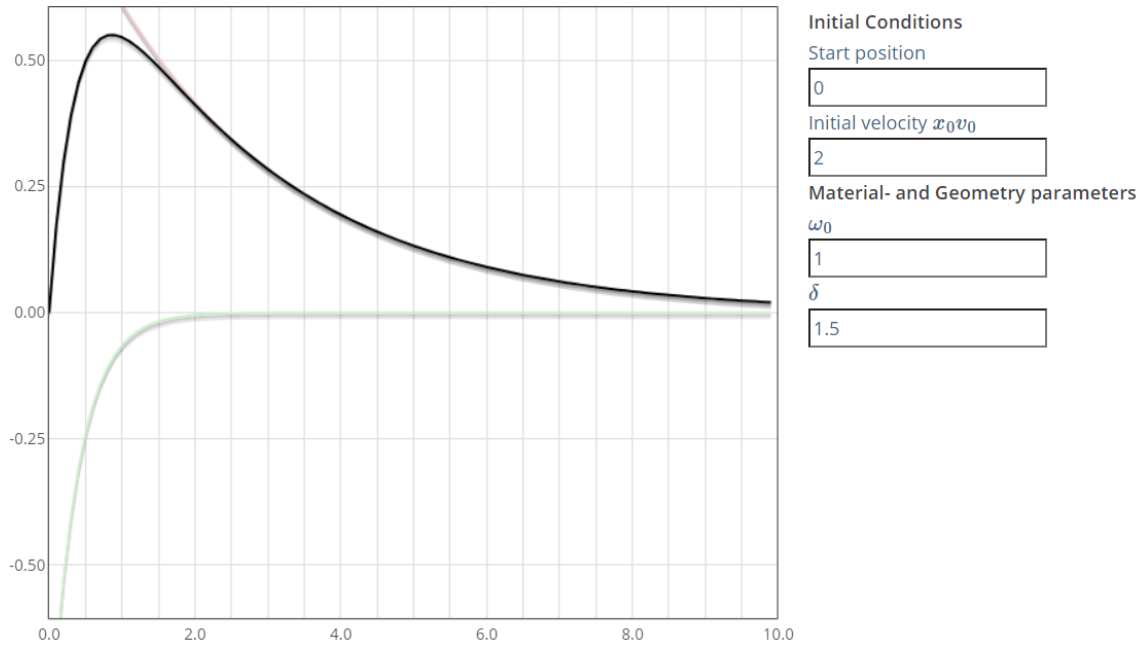
$$x(t) = e^{-\delta t} (C_1 e^{at} + C_2 e^{-at}) \quad (4.11)$$

Αυτή είναι η λύση για την υπερκρίσιμη απόσβεση. Οι σταθερές ολοκλήρωσης C_1 και C_2 για ένα συγκεκριμένο πρόβλημα μπορεί να προσδιοριστούν από δεδομένες αρχικές συνθήκες. Για παράδειγμα, θα μπορούσε να δοθεί η θέση εκκίνησης και η αρχική ταχύτητα του εκκρεμούς.

$$x(0) = x_0 \quad , \quad \frac{dx}{dt}(0) = v_0 \quad (4.12)$$

Η πιο κάτω γραφική, είναι η γραφική αναπαράσταση του γενικού αναλυτικού διαλύματος του υπερκρίσιμου αρμονικού ταλαντωτή. Το διάγραμμα δείχνει την εκτροπή του εκκρεμούς με την πάροδο του χρόνου. Η ανοιχτοπράσινη καμπύλη είναι η δεύτερη μερική λύση. Η μαύρη καμπύλη είναι το άθροισμα των δύο μερικών λύσεων και αντιπροσωπεύει τη λύση της διαφορικής εξίσωσης του υπερενισχυμένου αρμονικού ταλαντωτή για ένα δεδομένο σύνολο αρχικών συνθηκών. $C_1 e^{-\beta_1 t} + C_2 e^{-\beta_2 t}$

Σχήμα 4.3: Overdamped Harmonic Oscillator



4.0.5 Περίπτωση με κρίσιμη απόσβεση

Εάν $\delta = \omega_0$ τότε μιλάμε για κρίσιμη απόσβεση. Σε αυτή την περίπτωση η εξίσωση έχει μια ενιαία λύση για:

$$\hat{\lambda}_1 = \hat{\lambda}_2 = \hat{\lambda} = -\delta \quad (4.13)$$

Η χρήση του εκθετικού Ansatz σε αυτή την περίπτωση παραπέμπει στη χρήση της ακόλουθης προσέγγισης για την απόκτηση των δύο μερικών λύσεων στη διαφορική εξίσωση:

$$x_1(t) = C_1 e^{\hat{\lambda}t} \quad x_2(t) = t C_2 e^{\hat{\lambda}t} \quad (4.14)$$

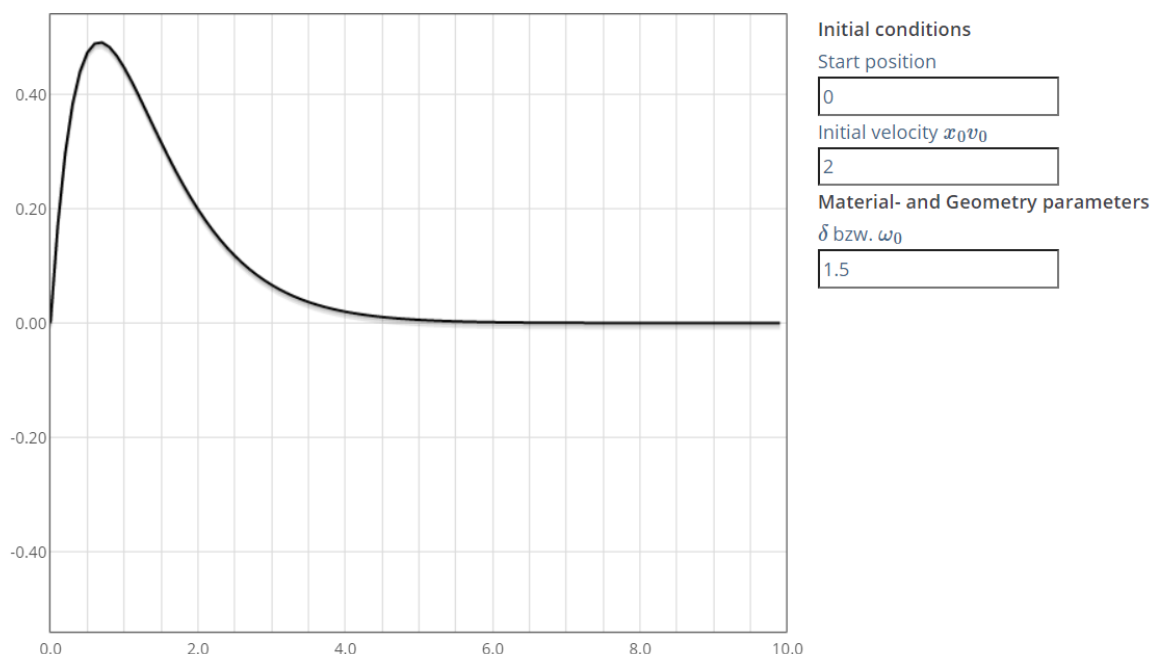
Η εισαγωγή της εξίσωσης 4.13 και 4.14 στη γενική λύση 4.8 δίνει τη λύση για την κρίσιμα αποσβεσμένη περίπτωση:

$$x(t) = e^{-\delta t} (C_1 + t C_2) \quad (4.15)$$

Όπως και στην περίπτωση της Υπερφορτωμένης περίπτωσης οι σταθερές ολοκλήρωσης C_1 και C_2 πρέπει να ληφθούν από δεδομένες αρχικές συνθήκες ενός συγκεκριμένου προβλήματος.

$$x(0) = x_0 \quad , \quad \frac{dx}{dt}(0) = v_0 \quad (4.16)$$

Σχήμα 4.4: Critically damped case



Γραφική αναπαράσταση της γενικής αναλυτικής λύσης της κρισίμως αποσβεσμένης περίπτωσης. Το διάγραμμα δείχνει την εκτροπή του εκκρεμούς με την πάροδο του χρόνου για τις αρχικές συνθήκες που δίνονται στο πλαίσιο. Οι λεπτομέρειες για τον υπολογισμό των σταθερών του πιο πάνω γραφήματος φαίνονται στον σύνδεσμο Wolfarm Alpha.

4.0.6 Υποκρίσιμη απόσβεση

Εάν $\delta < \omega_0$, τότε έχουμε αποσβεσμένη ταλάντωση. Στην περίπτωση αυτή, η διακρίνουσα στην εξίσωση είναι αρνητική, επομένως λ_1 και λ_2 είναι μιγαδικοί αριθμοί. Το εκθετικό ansatz χρησιμοποιείται και πάλι για την επίλυση της διαφορικής εξίσωσης $x(t) = Ce^{\lambda t}$.

$$x_1(t) = C_1 e^{\lambda_1 t} \quad x_2(t) = C_2 e^{\lambda_2 t}$$

Εισάγοντας το εκθετικό ansatz στην εξίσωση 4.8 και αντικαθιστώντας λ με την εξίσωση 4.7 παίρνουμε το εξής:

$$x(t) = e^{-\delta t} \left(C_1 e^{\sqrt{\delta^2 - \omega_0^2} t} + C_2 e^{-\sqrt{\delta^2 - \omega_0^2} t} \right) \quad (4.17)$$

Έχουμε ήδη δει αυτή την εξίσωση σε μια ελαφρώς τροποποιημένη μορφή ως εξίσωση 4.11. Είναι η γενική λύση στη διαφορική εξίσωση του αρμονικού ταλαντωτή. Τώρα

έχουμε να κάνουμε με μια σύνθετη λύση επειδή στην υποκρίσιμη απόσβεση ο όρος κάτω από την τετραγωνική ρίζα είναι αρνητικός. Επομένως, οι σταθερές C_1 και C_2 έχουν μιγαδικές τιμές. Για την αντιμετώπιση μιγαδικών αριθμών χρησιμοποιούμε τον τύπο του Euler. Αυτή η εξίσωση δημιουργεί μια σύνδεση μεταξύ μιγαδικών εκθετικών συναρτήσεων και τριγωνομετρικών συναρτήσεων:

$$e^{i\phi} = \cos\phi + i\sin\phi$$

Είναι χρήσιμο να αναδιατάξουμε την εξίσωση 4.17 με τέτοιο τρόπο ώστε το φανταστικό της μέρος να είναι απομονωμένο:

$$\sqrt{\delta^2 - \omega_0^2} = \sqrt{-1 * (\omega_0^2 - \delta^2)} = i \sqrt{\omega_0^2 - \delta^2} \quad (4.18)$$

Παίρνοντας έναν όρο $i\omega$ που αποτελείται από τη φανταστική μονάδα πολλαπλασιασμένη με μια πραγματική τετραγωνική ρίζα, για να απλοποιήσουμε περαιτέρω υπολογισμούς αντικαθιστούμε την τετραγωνική ρίζα με μια νέα σταθερά που ονομάζεται :

$$\omega = \sqrt{\omega_0^2 - \delta^2} \quad (4.19)$$

Όπως θα δούμε σύντομα, αυτή η σταθερά αντιπροσωπεύει τη φυσική συχνότητα του αρμονικού ταλαντωτή. Με τις σχέσεις από τις εξισώσεις 4.18, 4.19 και την εξίσωση 4.17 μπορεί τώρα να μετατραπούν σε :

$$x(t) = e^{-\delta t} \underbrace{\left(C_1 e^{i\omega t} + C_2 e^{-i\omega t} \right)}_{\hat{x}(t)} \quad (4.20)$$

Από φυσική άποψη, μόνο η καθαρά πραγματική πολύτιμη λύση παρουσιάζει ενδιαφέρον. Για να τα βρούμε αυτά πρέπει να διαχωρίσουμε τα πραγματικά και τα φανταστικά μέρη μεταξύ τους. Όπως ήδη αναφέρθηκε οι C_1 και C_2 είναι μιγαδικές σταθερές. Παρακάτω τα μετατρέπουμε στην πολική τους μορφή.

$$C_1 = \hat{C}_1 e^{i\phi_1}, C_2 = \hat{C}_2 e^{i\phi_2} \quad (4.21)$$

Το πρώτο μέρος της εξίσωσης 4.20 είναι μια εκθετική συνάρτηση. Επειδή η σταθερά δ δεν είναι μιγαδική τιμή, αυτό το μέρος της εξίσωσης δεν μπορεί να αποφέρει σύνθετα αποτελέσματα. Εστιάζουμε λοιπόν στο δεύτερο μέρος που ονομάσαμε $\hat{x}(t)$. Αντικαθιστούμε τις σταθερές C_1 και C_2 με αυτές που είναι γραμμένες σε πολική μορφή 4.21 και εφαρμόζουμε τον τύπο του Euler.

$$\begin{aligned}
\hat{x}(t) &= \hat{C}_1 e^{i\phi_1} e^{i\omega t} + \hat{C}_2 e^{i\phi_2} e^{-i\omega t} \\
&= \hat{C}_1 e^{i(\phi_1 + \omega t)} + \hat{C}_2 e^{i(\phi_2 - \omega t)} \\
&= \hat{C}_1 \cos(\phi_1 + \omega t) + i\hat{C}_1 \sin(\phi_1 + \omega t) + \hat{C}_2 \cos(\phi_2 - \omega t) + i\hat{C}_2 \sin(\phi_2 - \omega t) \\
&= \hat{C}_1 \cos(\phi_1 + \omega t) + \hat{C}_2 \cos(\phi_2 - \omega t) + i(\hat{C}_1 \sin(\phi_1 + \omega t) + \hat{C}_2 \sin(\phi_2 - \omega t)) \quad (4.22)
\end{aligned}$$

Στην εξίσωση 4.22 διαχωρίζονται πραγματικά και φανταστικά μέρη της εξίσωσης. Μας ενδιαφέρουν μόνο οι λύσεις για τις οποίες εξαφανίζεται το φανταστικό μέρος. Αυτό συμβαίνει όταν ισχύουν τα εξής:

$$\begin{aligned}
\hat{C}_1 \sin(\phi_1 + \omega t) + \hat{C}_2 \sin(\phi_2 - \omega t) &= 0 \\
\hat{C}_1 \sin(\phi_1 + \omega t) &= -\hat{C}_2 \sin(\phi_2 - \omega t) \\
\hat{C}_1 \sin(\phi_1 + \omega t) &= \hat{C}_2 \sin(-\phi_2 + \omega t) \quad (4.23)
\end{aligned}$$

Έτσι το φανταστικό μέρος της εξίσωσης 4.20 εξαφανίζεται αν:

$$\hat{C}_1 = \hat{C}_2, \quad \phi_1 = -\phi_2 \quad (4.24)$$

Ή με άλλα λόγια, οι σταθερές C_1 και C_2 πρέπει να είναι σύνθετες συζυγείς:

$$C_1 = C_2^*$$

Εισάγοντας την εξίσωση 4.24 στο υπόλοιπο πραγματικό μέρος της εξίσωσης 4.22 λαμβάνουμε για $\hat{x}(t)$:

$$\begin{aligned}
\hat{x}(t) &= \hat{C}_1 \cos(\phi_1 + \omega t) + \hat{C}_1 \cos(-\phi_1 - \omega t) \\
&= \hat{C}_1 \cos(\phi_1 + \omega t) + \hat{C}_1 \cos(\phi_1 + \omega t) \\
&= 2\hat{C}_1 \cos(\phi_1 + \omega t) \quad (4.25)
\end{aligned}$$

Χρησιμοποιώντας την εξίσωση 4.25 στην εξίσωση 4.20 λαμβάνουμε τη λύση της διαφορικής εξίσωσης για την υποκρίσιμη απόσβεση:

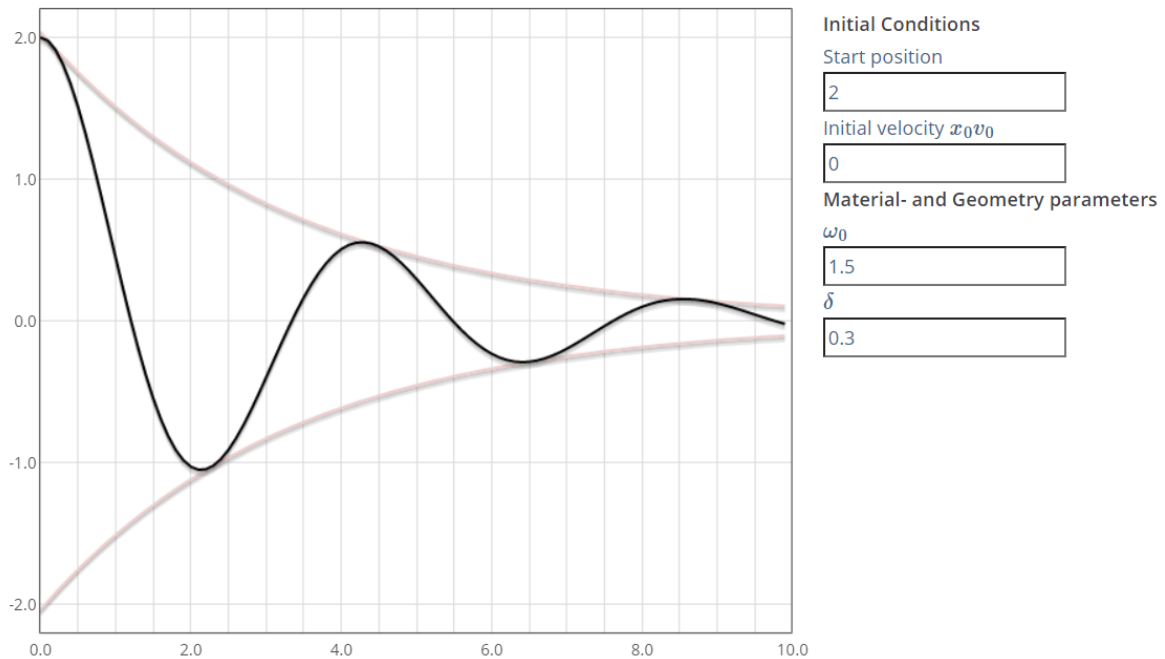
$$x(t) = e^{-\delta t} (2A \cos(\phi + \omega t)) \quad (4.26)$$

Οι σταθερές \hat{C}_1 και ϕ_1 έχουν μετονομαστεί σε A και ϕ αντίστοιχα, όπου A είναι ένα

πλάτος και φ η γωνία φάσης. Οι σταθερές μπορούν να υπολογιστούν από ένα δεδομένο σύνολο αρχικών συνθηκών, όπως:

$$x(0) = x_0 \quad , \quad \frac{dx}{dt}(0) = v_0 \quad (4.27)$$

Σχήμα 4.5: Underdamped case



Γραφική αναπαράσταση της γενικής αναλυτικής λύσης της υποκρίσιμης απόσβεσης. Το διάγραμμα δείχνει την παραμόρφωση του εκκρεμούς με την πάροδο του χρόνου για τις αρχικές συνθήκες που έχουν οριστεί στο πλαϊνό πλαίσιο. Οι λεπτομέρειες για τον υπολογισμό των σταθερών του πιο πάνω γραφήματος φαίνονται στον σύνδεσμο [Wolfarm Alpha](#).

Κεφάλαιο 5

Επίλυση διαφορικής εξίσωσης Αποσβενόμενου Αρμονικού Ταλαντωτή με χρήση PINNs

Σε αυτό το κεφάλαιο θα αναπτύξουμε τα κατάλληλα νευρωνικά δίκτυα και PINNs έτσι ώστε να επιλύσουμε το πρόβλημα του αποσβενόμενου αρμονικού ταλαντωτή για την υποκρίσιμη απόσβεση. Η λέξη 'επίλυση' ίσως δεν είναι η κατάλληλη για να περιγράψει τη διαδικασία που θα ακολουθήσει, διότι τα νευρωνικά δίκτυα δεν επιλύουν προβλήματα αλλά τείνουν να ελαχιστοποιήσουν το σφάλμα εκτίμησης. Όπως αναφέραμε και παραπάνω για να εκπαιδευτεί ένα οποιοδήποτε νευρωνικό δίκτυο με επίβλεψη είναι αναγκαίο να υπάρχουν δεδομένα εκπαίδευσης, με άλλα λόγια δεδομένα που οδηγούν το μοντέλο από μία συγκεκριμένη είσοδο σε μία καθορισμένη έξοδο. Στη δική μας περίπτωση τα δεδομένα εκπαίδευσης δημιουργήθηκαν από την παρακάτω διαδικασία:

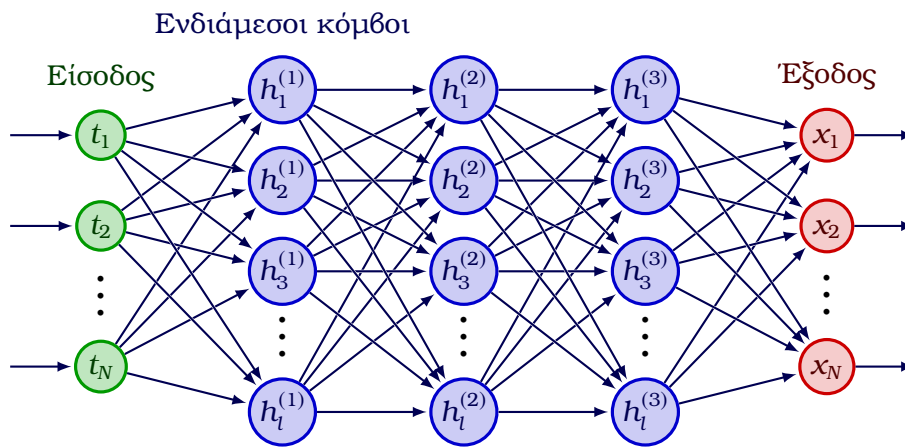
1. Κατασκευή προσομοιωτή ενός αρμονικού ταλαντωτή της υποκρίσιμης απόσβεσης με είσοδο τους παραμέτρους και σταθερές του ταλαντωτή (w_0, d) καθώς και τις χρονικές στιγμές t για τις οποίες θα παρατηρήσουμε την ταλάντωση. Για ευκολία οι χρονικές στιγμές αυτές θα κρατούνται σε ένα διάνυσμα \mathbf{t} , όπου $\mathbf{t} = [t_1, t_2, \dots, t_N]$
2. Καταγραφή της θέσης x του ταλαντωτή για όλες τις χρονικές στιγμές \mathbf{t} στο διάνυσμα $\mathbf{x} = [x(t_1), x(t_2), \dots, x(t_N)]$. Το διάνυσμα \mathbf{x} αποτελεί την επιθυμητή έξοδο του μοντέλου μηχανικής μάθησης.
3. Δειγματοληψία m χρονικών του διανύσματος \mathbf{t} , όπου $m < N$ και αποθήκευση αυτών στο διάνυσμα $\mathbf{t}' = [t_1, t_2, \dots, t_m]$. Το διάνυσμα \mathbf{t}' αποτελεί την είσοδο στο μοντέλο μηχανικής μάθησης.

Ουσιαστικά, τα δειγματοληπτημένα σημεία \mathbf{t}' θα πρέπει να μας οδηγούν στην εκτίμηση όλων των σημείων ταλάντωσης \mathbf{x} . Με άλλα λόγια το νευρωνικό δίκτυο ή το PINN θα

κατασκευάσει μέσω εκπαίδευσης μια συνάρτηση που θα προβάλει τα δεδομένα από τον διανυσματικό χώρο της εισόδου στο διανυσματικό χώρο της εξόδου, προσπαθώντας να ελαχιστοποιήσει το σφάλμα αυτής της προβολής. Τα πειράματα που θα ακολουθήσουν δείχνουν την απόδοση του απλού νευρωνικού δικτύου καθώς και του PINN, έχοντας ως μετρική το Root Mean Square Error (RMSE).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x'_i - x_i)^2} \quad (5.1)$$

όπου x'_i οι εκτιμώμενες θέσεις του ταλαντωτή, όπου x_i οι πραγματικές θέσεις του ταλαντωτή. Το απλό νευρωνικό δίκτυο που χρησιμοποιήθηκε έχει την παρακάτω μορφή :

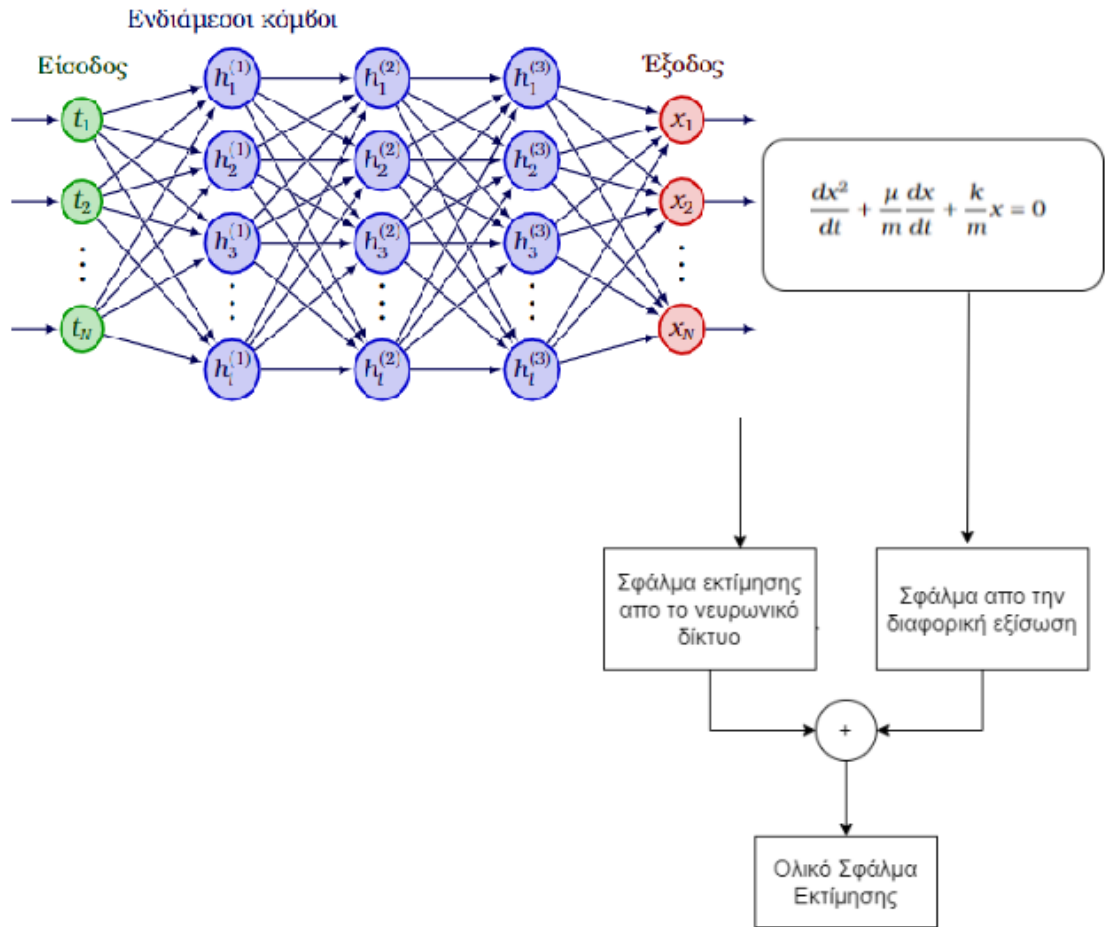


Σχήμα 5.1: Νευρωνικό δίκτυο πειραμάτων

Το l αντιπροσωπεύει τον αριθμό των κρυφών νευρώνων ανά στρώμα (hidden layers) που στα δικά μας πειράματα είναι ίσως με τριάντα δύο, ενώ ο αριθμός των κρυφών στρωμάτων ή ενδιάμεσων κόμβων είναι τρεις. Σε όλους του κόμβους χρησιμοποιήθηκε η υπερβολική εφαπτομένη ως συνάρτηση ενεργοποίησης εκτός του κόμβων εξόδου. Για την υλοποίηση του νευρωνικού δικτύου έγινε χρήση της βιβλιοθήκης pytorch.

Όπως αναφέρθηκε και παραπάνω, σε προβλήματα φυσικής τα PINN μπορούν να αποδειχθούν μια εξαιρετική προσέγγιση, διότι συνδυάζουν την πρότερη γνώση που έχει ο επιστήμονας για το πρόβλημα της αρμονικής ταλάντωσης, όπως για παράδειγμα τη διαφορική εξίσωση που περιγράφει το φαινόμενο, με την αποτελεσματικότητα των νευρωνικών δικτύων. Τα PINN δεν είναι τίποτα άλλο από μία επέκταση των απλών νευρωνικών δικτύων στον τρόπο με τον οποίο υπολογίζεται το σφάλμα. Στα απλά νευρωνικά δίκτυα το σφάλμα εκτίμησης υπολογίζεται από τη γεωμετρική απόσταση που έχει η εκτιμώμενη από την πραγματική τιμή. Τα PINN υπολογίζουν επιπλέον το σφάλμα που προκύπτει αντικαθιστώντας τις τυχαίες μεταβλητές της διαφορικής εξίσωσης που περιγράφει το πρόβλημα, με τις εξόδους του νευρωνικού δικτύου. Για παράδειγμα, στην περίπτωση

Σχήμα 5.2: PINN για την επίλυση του προβλήματος της αποσβεσμένης αρμονικής ταλάντωσης



της υποκρίσιμης αρμονικής ταλάντωσης η ομογενής διαφορική εξίσωση είναι ίση με το μηδέν. Αν οι εκτιμώμενη θέση του ταλαντωτή για μία δεδομένη χρονική στιγμή δεν οδηγεί το αποτέλεσμα της διαφορικής εξίσωσης σε μηδέν, τότε αυτό είναι μια χρήσιμη πληροφορία που το PINN την εκμεταλλεύεται για τον καλύτερο υπολογισμό των βαρών των νευρώνων. Τεχνικά μιλώντας, το ολικό σφάλμα του PINN είναι το άθροισμα του RMSE του νευρωνικού δικτύου (σφάλμα εκτίμησης) και το σφάλμα που προκύπτει από τη διαφορική εξίσωση όπως φαίνεται στο σχήμα 5.2. Το νευρωνικό δίκτυο που χρησιμοποιήθηκε για την κατασκευή του PINN έχει τριάντα δύο νευρώνες σε κάθε κρυφό στρώμα, ενώ ο αριθμός των κρυφών στρωμάτων είναι τρεις, όπως φαίνεται στην εικόνα 5.2. Σε όλους του κόμβους χρησιμοποιήθηκε η υπερβολική εφασπτομένη ως συνάρτηση ενεργοποίησης εκτός του κόμβων εξόδου. Για την υλοποίηση του PINN έγινε χρήση της βιβλιοθήκης pytorch.

5.0.1 Μονοβάθμιο πρόβλημα αρμονικού ταλαντωτή

Σε αυτό το κεφάλαιο θα μελετήσουμε την απόδοση του NN και του PINN στην πρόβλεψη της θέσης ενός ελατηρίου σε συγκεκριμένες χρονικές στιγμές.

ΔΕΙΓΜΑΤΑ	NN-RMSE (%)	PINN-RMSE (%)
Δέκα ισοκαταμεμημένα δείγματα στα πρώτα 0.3 sec / πρόβλεψη 1 sec	58.79	0.4552
Δέκα ισοκαταμεμημένα δείγματα στα πρώτα 0.75 sec / πρόβλεψη 2 sec	76.37	6.927
Δέκα ισοκαταμεμημένα δείγματα στα πρώτα 1.1 sec / πρόβλεψη 3 sec	30.12	17.13
Δέκα ισοκαταμεμημένα δείγματα στα πρώτα 1.5 sec / πρόβλεψη 4 sec	26.53	22.77
Δέκα ισοκαταμεμημένα δείγματα στα πρώτα 1.9 sec / πρόβλεψη 5 sec	21.24	12.62

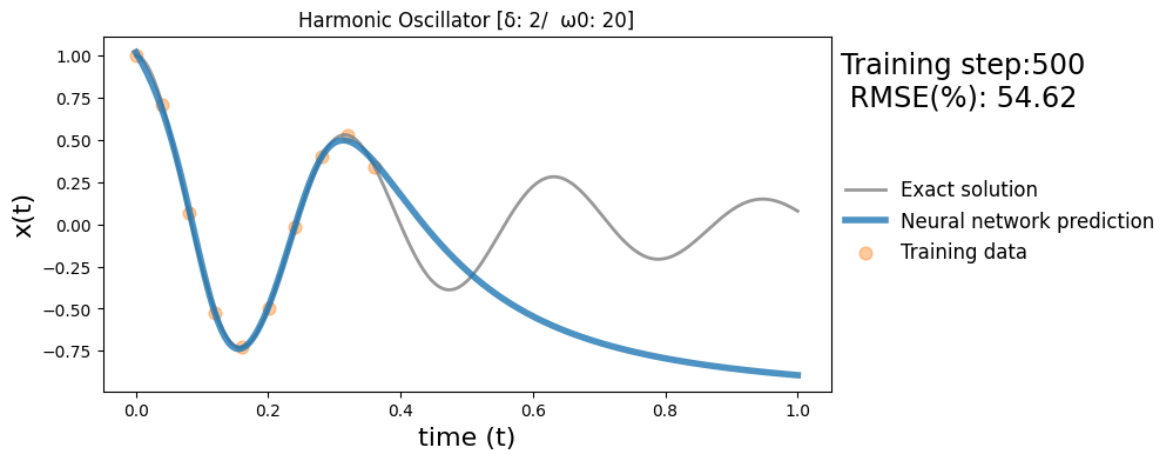
Ο πίνακας συγκρίνει την απόδοση δύο διαφορετικών μοντέλων, ενός απλού νευρωνικού δικτύου και ενός νευρωνικού δικτύου με πληροφορία φυσικής (PINN), με δέκα ομοιόμορφα καταμεμημένα δείγματα ως είσοδο. Η απόδοση του κάθε μοντέλου μετριέται χρησιμοποιώντας το ποσοστιαίο RMSE.

Ο πίνακας παρουσιάζει τα αποτελέσματα των μοντέλων σε πέντε διαφορετικά σαι δειγμάτων, το καθένα με διαφορετικό χρόνο πρόβλεψης. Για κάθε σύνολο δειγμάτων, το RMSE του απλού νευρωνικού δικτύου είναι σημαντικά υψηλότερο από αυτό του PINN. Για παράδειγμα, όταν τα δείγματα βρίσκονται στα πρώτα 0,3 δευτερόλεπτα και ο χρόνος πρόβλεψης είναι 1 δευτερόλεπτο, το απλό νευρωνικό δίκτυο έχει RMSE 58,79%, ενώ το PINN έχει RMSE μόνο 0,4552%. Ομοίως, όταν τα δείγματα είναι στα πρώτα 1,9 δευτερόλεπτα και ο χρόνος πρόβλεψης είναι 5 δευτερόλεπτα, το απλό νευρωνικό δίκτυο έχει RMSE 21,24%, ενώ το PINN έχει RMSE 12,62%. Γενικά, όσο μεγαλύτερο είναι το χρονικό διάστημα των παρατηρήσεων, τόσο χαμηλότερη είναι το RMSE των δύο μοντέλων. Αυτό οφείλεται στο γεγονός ότι οι εκτιμήσεις των νευρωνικών δεν είναι ακριβείς σε περιοχές που δεν υπάρχουν δείγματα.

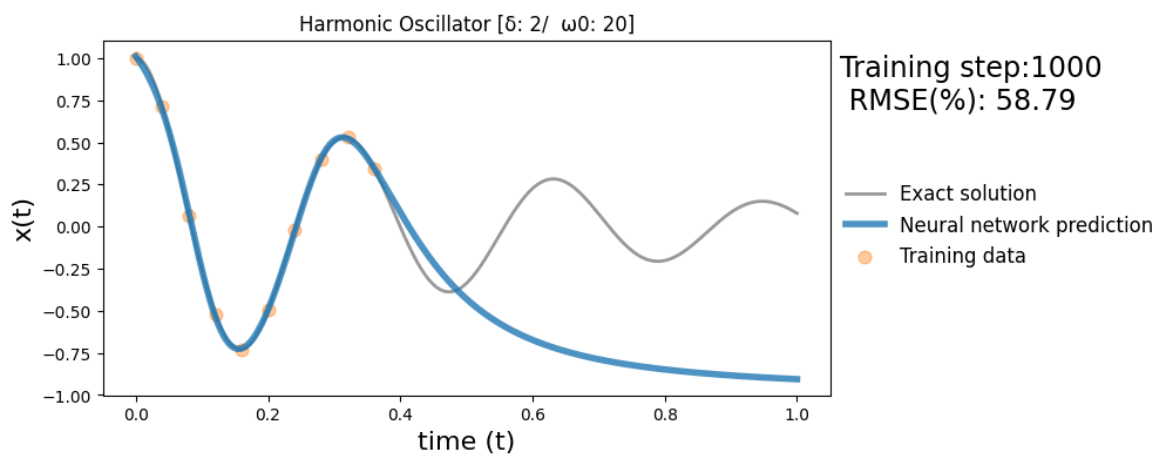
Συνολικά, αυτά τα αποτελέσματα δείχνουν ότι το μοντέλο PINN είναι σημαντικά πιο ακριβές από το απλό νευρωνικό δίκτυο στην πρόβλεψη των αποτελεσμάτων των δειγμάτων, ανεξάρτητα από το χρόνο πρόβλεψης. Αυτό υποδηλώνει ότι το μοντέλο PINN είναι σε θέση να ενσωματώσει καλύτερα τη γνώση των υποκείμενων φυσικών αρχών που διέπουν το σύστημα που μοντελοποιείται, οδηγώντας σε πιο ακριβείς προβλέψεις.

Τα παρακάτω γραφήματα δείχνουν την πορεία εκπαίδευσης του NN για πρόβλεψη 1 δευτερολέπτου

Σχήμα 5.3: Neural Network, Training step=500, Prediction period=1 sec

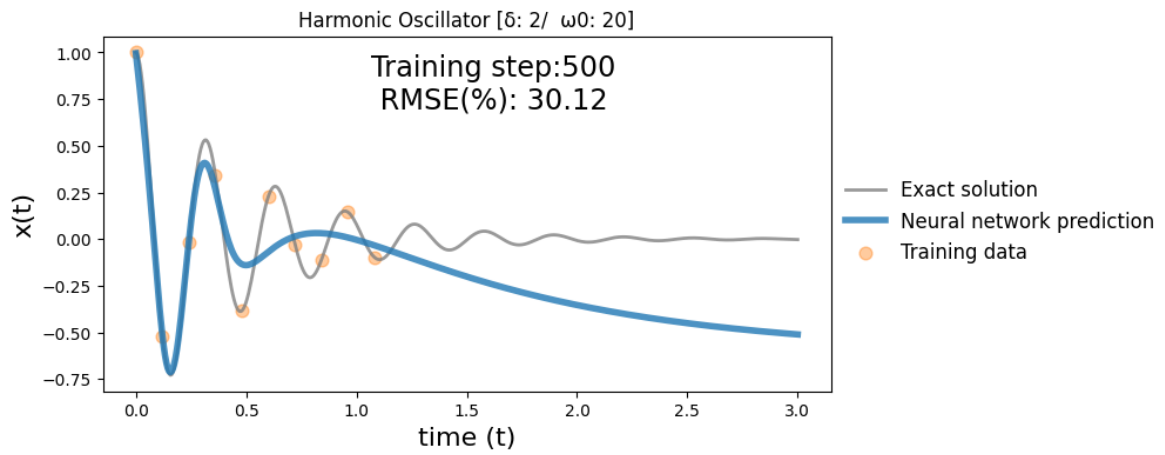


Σχήμα 5.4: Neural Network, Training step=1000, Prediction period=1 sec

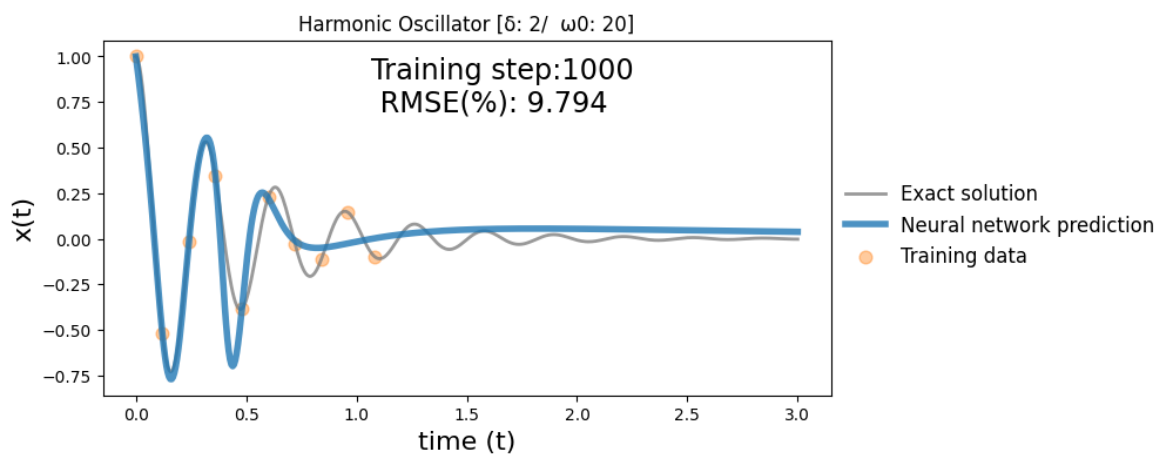


Τα παρακάτω γραφήματα δείχνουν την πορεία εκπαίδευσης του NN για πρόβλεψη 3 δευτερολέπτων

Σχήμα 5.5: Neural Network, Training step=500, Prediction period=3 sec

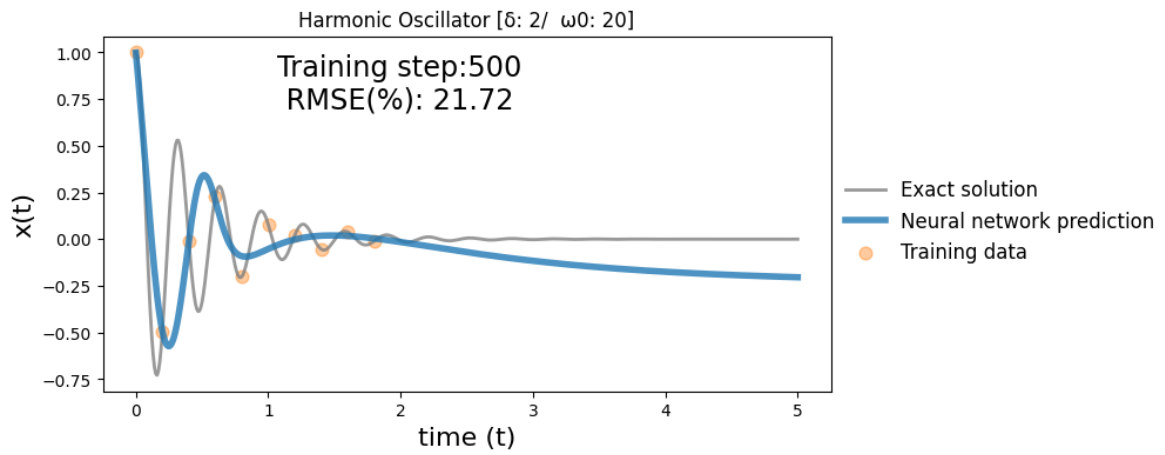


Σχήμα 5.6: Neural Network, Training step=1000, Prediction period=3 sec

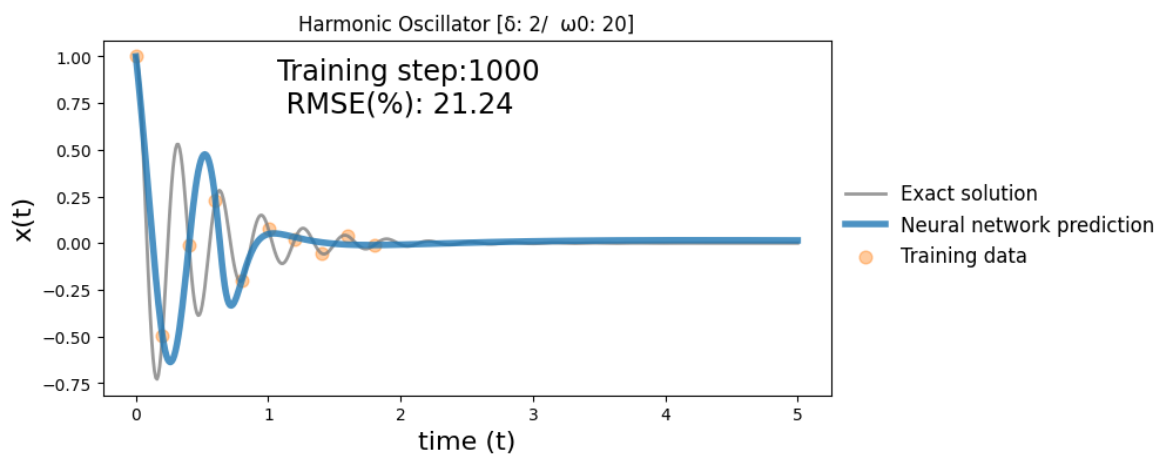


Τα παρακάτω γραφήματα δείχνουν την πορεία εκπαίδευσης του NN για πρόβλεψη 5 δευτερολέπτων

Σχήμα 5.7: Neural Network, Training step=500, Prediction period=5 sec

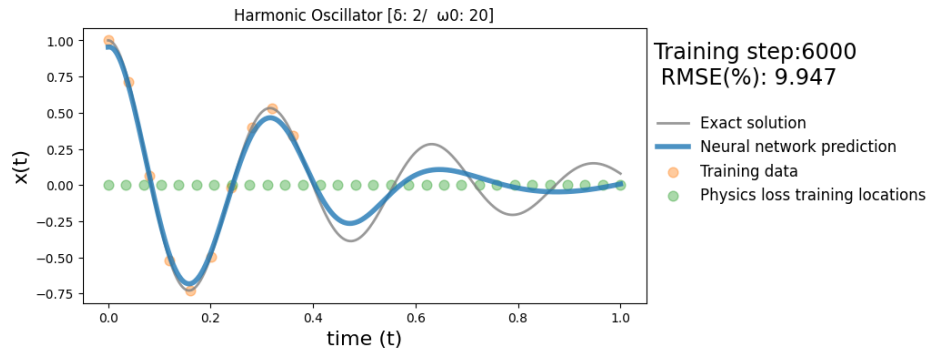


Σχήμα 5.8: Neural Network, Training step=1000, Prediction period=5 sec

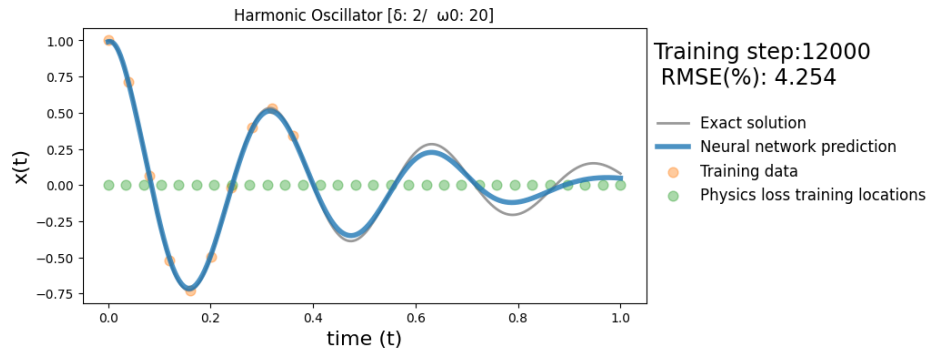


Τα παρακάτω γραφήματα δείχνουν την πορεία εκπαίδευσης του PINN για πρόβλεψη 1 δευτερολέπτου

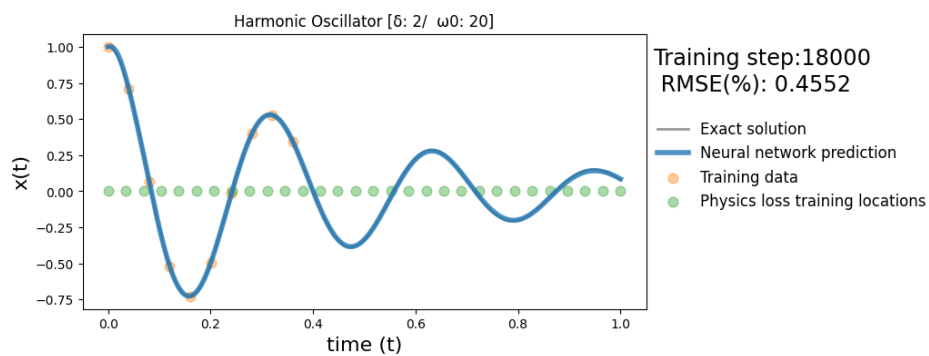
Σχήμα 5.9: PINN, Training step=6000, Prediction period=1 sec



Σχήμα 5.10: PINN, Training step=12000, Prediction period=1 sec

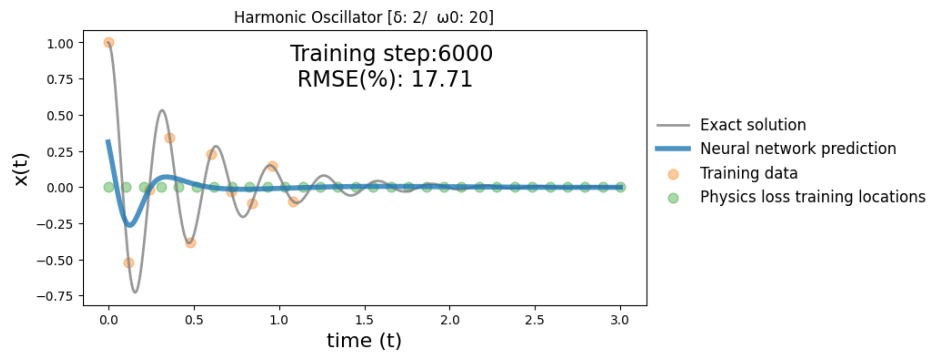


Σχήμα 5.11: PINN, Training step=18000, Prediction period=1 sec

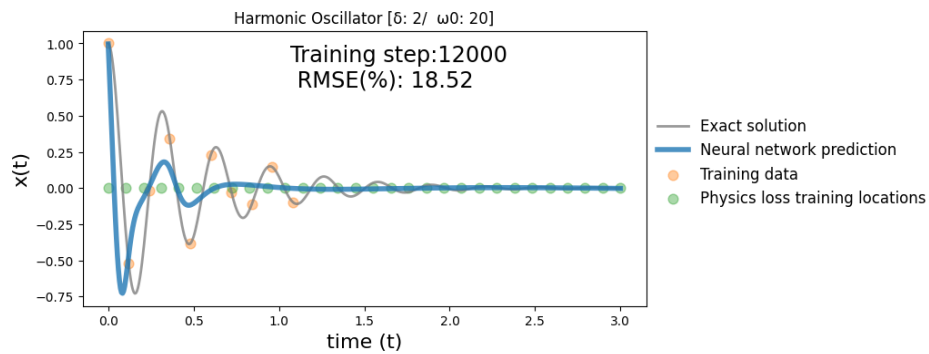


Τα παρακάτω γραφήματα δείχνουν την πορεία εκπαίδευσης του PINN για πρόβλεψη 3 δευτερολέπτων

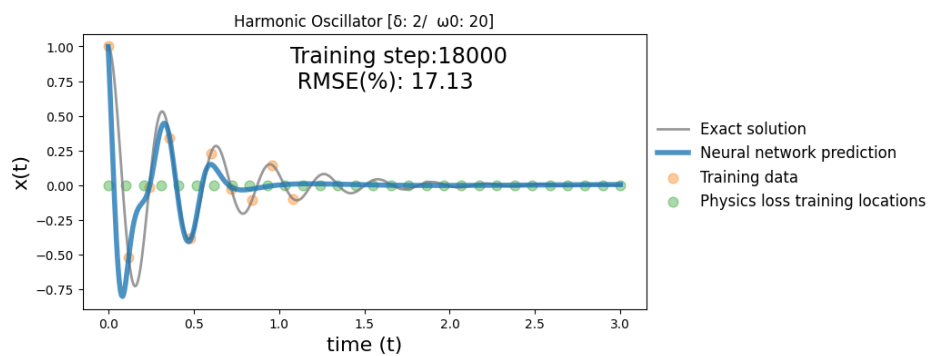
Σχήμα 5.12: PINN, Training step=6000, Prediction period=3 sec



Σχήμα 5.13: PINN, Training step=12000, Prediction period=3 sec

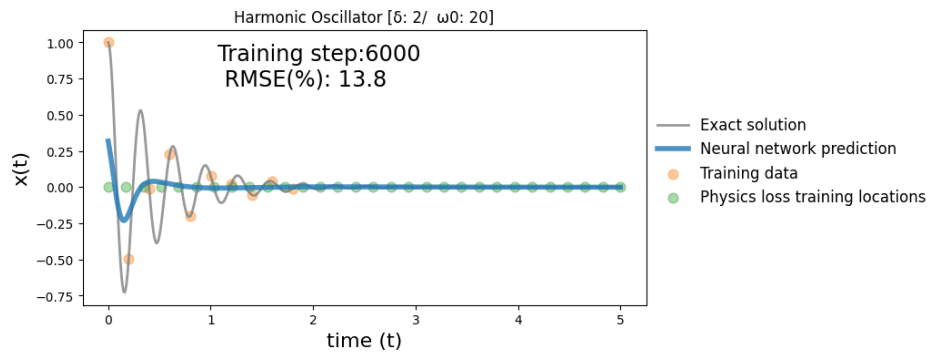


Σχήμα 5.14: PINN, Training step=18000, Prediction period=3 sec

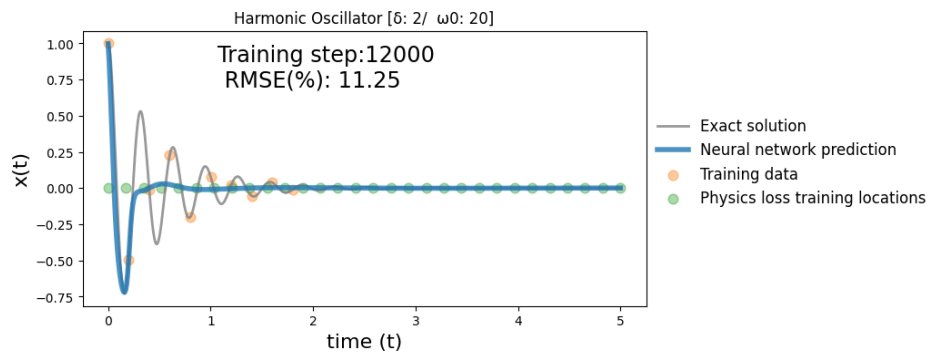


Τα παρακάτω γραφήματα δείχνουν την πορεία εκπαίδευσης του PINN για πρόβλεψη 5 δευτερολέπτων

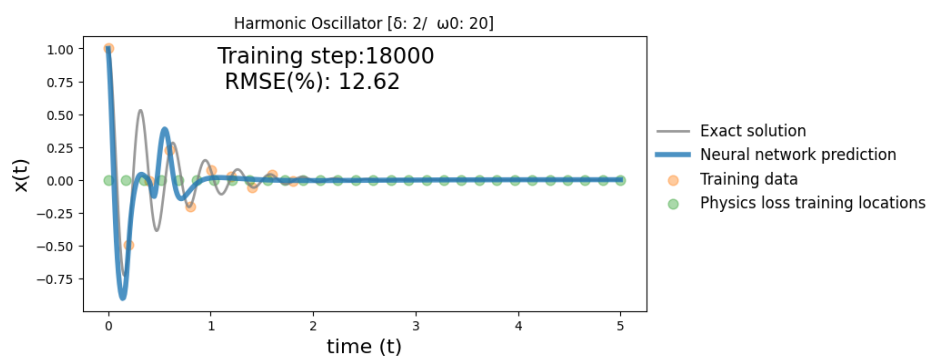
Σχήμα 5.15: PINN, Training step=6000, Prediction period=5 sec



Σχήμα 5.16: PINN, Training step=12000, Prediction period=5 sec



Σχήμα 5.17: PINN, Training step=18000, Prediction period=5 sec



5.0.2 Πολυβάθμιο πρόβλημα αρμονικού ταλαντωτή

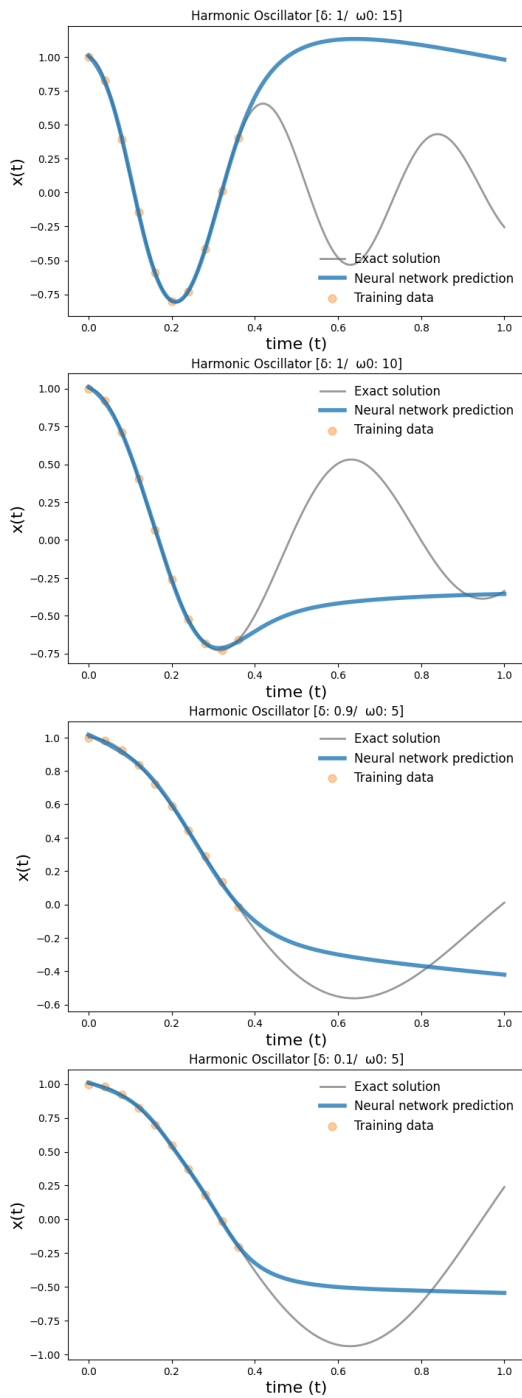
Σε αυτό το πρόβλημα, η έξοδος των μοντέλων είναι η θέση τεσσάρων ανεξάρτητων αρμονικών ταλαντωτών. Οι τέσσερις ταλαντωτές έχουν διαφορετικές παραμέτρους, με τον πρώτο ταλαντωτή να έχει λόγο απόσβεσης $\delta = 1$ και συχνότητα $\omega_0 = 15$, τον δεύτερο ταλαντωτή να έχει $\delta = 1$ και $\omega_0 = 10$, τον τρίτο ταλαντωτή να έχει $\delta = 0,9$ και $\omega_0 = 5$, και τέταρτος ταλαντωτής που έχει $\delta = 0,1$ και $\omega_0 = 5$. Αυτό σημαίνει ότι το νευρωνικό δίκτυο και το PINN πρέπει να μάθουν να προβλέπουν τη θέση κάθε ταλαντωτή ξεχωριστά, με βάση το χρόνο εισόδου και τις εσωτερικές παραμέτρους του μοντέλου. Αυτό θέτει ένα πιο δύσκολο πρόβλημα σε σύγκριση με την πρόβλεψη της θέσης ενός μεμονωμένου ταλαντωτή, καθώς το νευρωνικό δίκτυο πρέπει να είναι σε θέση να καταγράφει με ακρίβεια τη διακριτή δυναμική κάθε ταλαντωτή με τις μοναδικές του παραμέτρους. Επιπροσθέτως, Επειδή οι ταλαντωτές είναι ανεξάρτητοι και δεν αλληλεπιδρούν μεταξύ τους, το νευρωνικό δίκτυο πρέπει να μπορεί να γενικεύει τις προβλέψεις του σε νέους συνδυασμούς παραμέτρων ταλαντωτή. Συνολικά, αυτό το πολυδιάστατο πρόβλημα απαιτεί από το νευρωνικό δίκτυο να έχει την ικανότητα να συλλαμβάνει πολύπλοκες δυναμικές και να γενικεύει σε νέες καταστάσεις, οι οποίες μπορεί να είναι προκλητικές για τα παραδοσιακά NN.

Για να απεικονίσουμε την ακρίβεια πρόβλεψης των μοντέλων νευρωνικών δικτύων, παρουσιάζουμε 2 εικόνες που απεικονίζουν την εκτίμηση της θέσης για κάθε ταλαντωτή από το απλό νευρωνικό δίκτυο στα βήματα εκπαίδευσης 500 και 1000. Αυτές οι εικόνες μας επιτρέπουν να δούμε πώς το νευρωνικό δίκτυο είναι σε θέση να χρησιμοποιήσει τα δεδομένα εκπαίδευσης και προβλέπει τη θέση των ταλαντωτών σε διαφορετικές χρονικές κλίμακες. Επιπλέον, παρουσιάζουμε τέσσερις εικόνες που απεικονίζουν την εκτίμηση της θέσης για κάθε ταλαντωτή από το PINN στα βήματα εκπαίδευσης 10000, 40000 και 80000. Αυτές οι εικόνες παρέχουν πληροφορίες για τη μακροπρόθεσμη ακρίβεια πρόβλεψης του PINN, καθώς και την ικανότητά του να αποτυπώνει τη θέση των ταλαντωτών.

Τα παρακάτω γραφήματα δείχνουν την πορεία εκπαίδευσης του πολυβάθμιου NN για πρόβλεψη 1 δευτερολέπτου

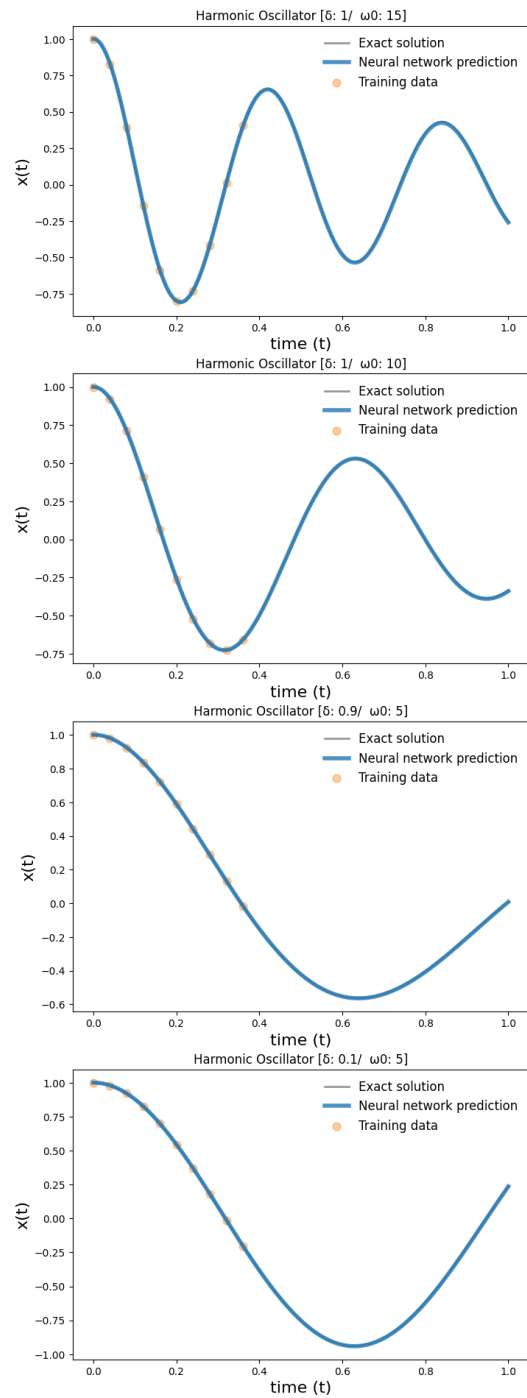
(α') Neural Network, Training step=80000, Prediction period=1 sec

Training step: 1000 RMSE (%) : 50.02



(β') PINN, Training step=80000, Prediction period=1 sec

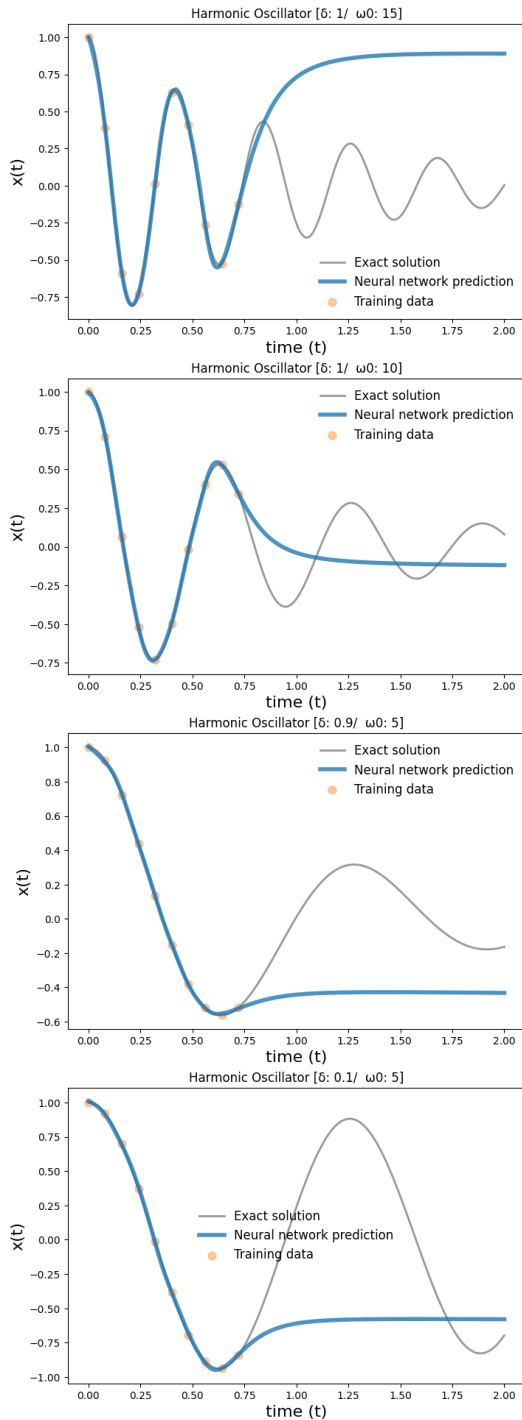
Training step: 80000 RMSE (%) : 0.2791



Τα παρακάτω γραφήματα δείχνουν την πορεία εκπαίδευσης του πολυβάθμιου NN για πρόβλεψη 2 δευτερολέπτων

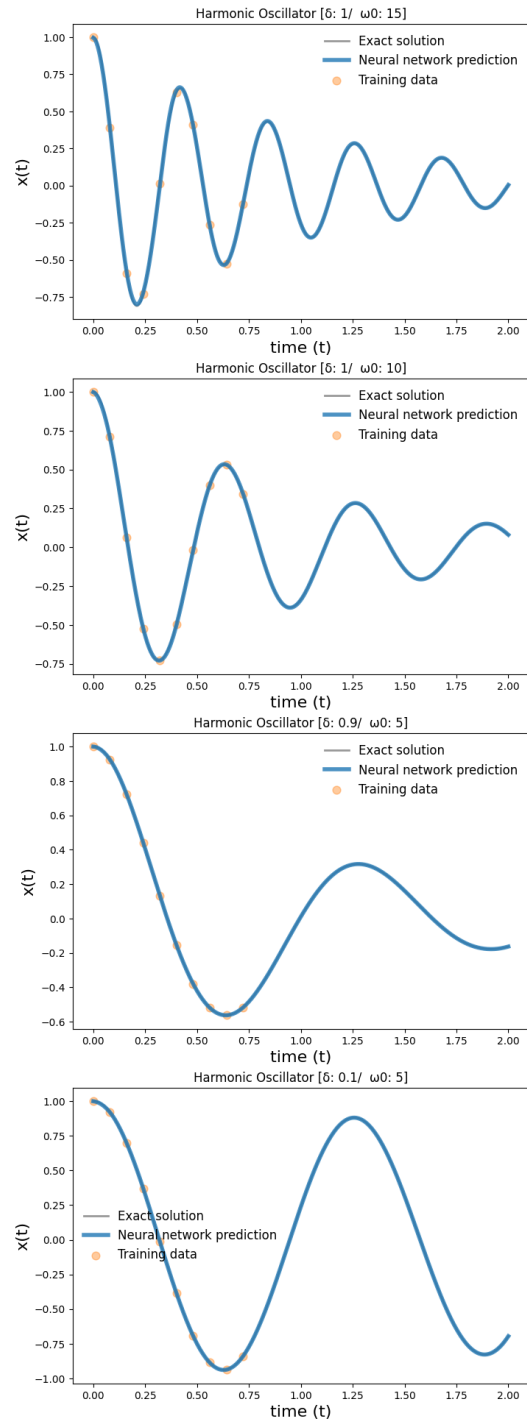
(α') Neural Network, Training step=80000, Prediction period=2 sec

Training step: 1000 RMSE (%) : 52.07



(β') PINN, Training step=80000, Prediction period=2 sec

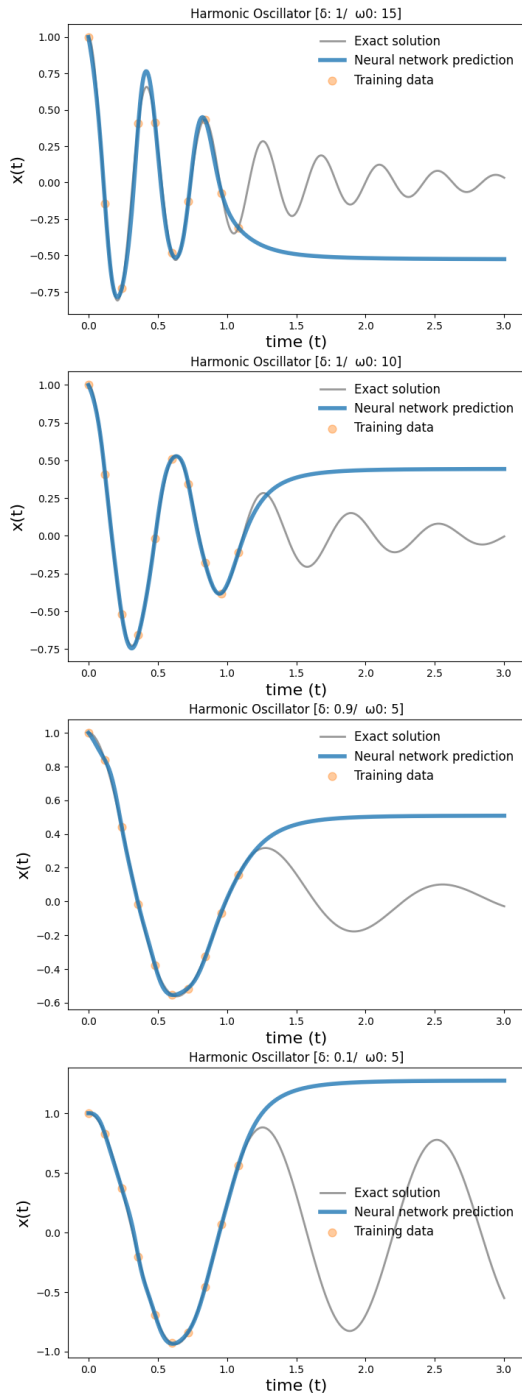
Training step: 80000 RMSE (%) : 0.3182



Τα παρακάτω γραφήματα δείχνουν την πορεία εκπαίδευσης του πολυβάθμιου NN για πρόβλεψη 3 δευτερολέπτων

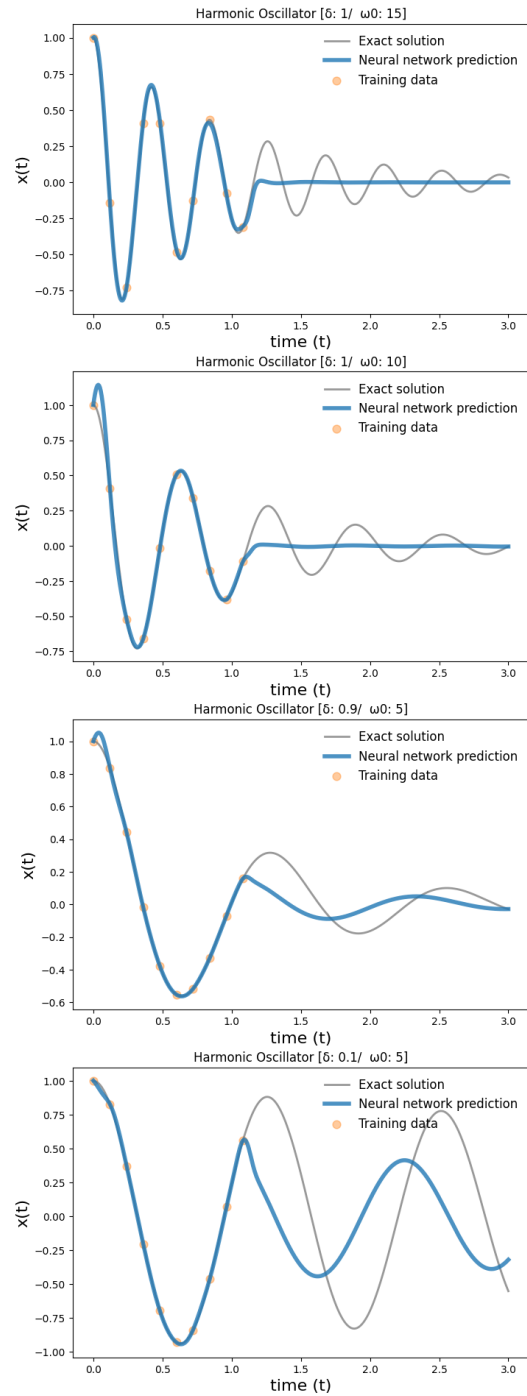
(α') Neural Network, Training step=80000, Prediction period=3 sec

Training step: 1000 RMSE (%) : 59.8



(β') PINN, Training step=80000, Prediction period=3 sec

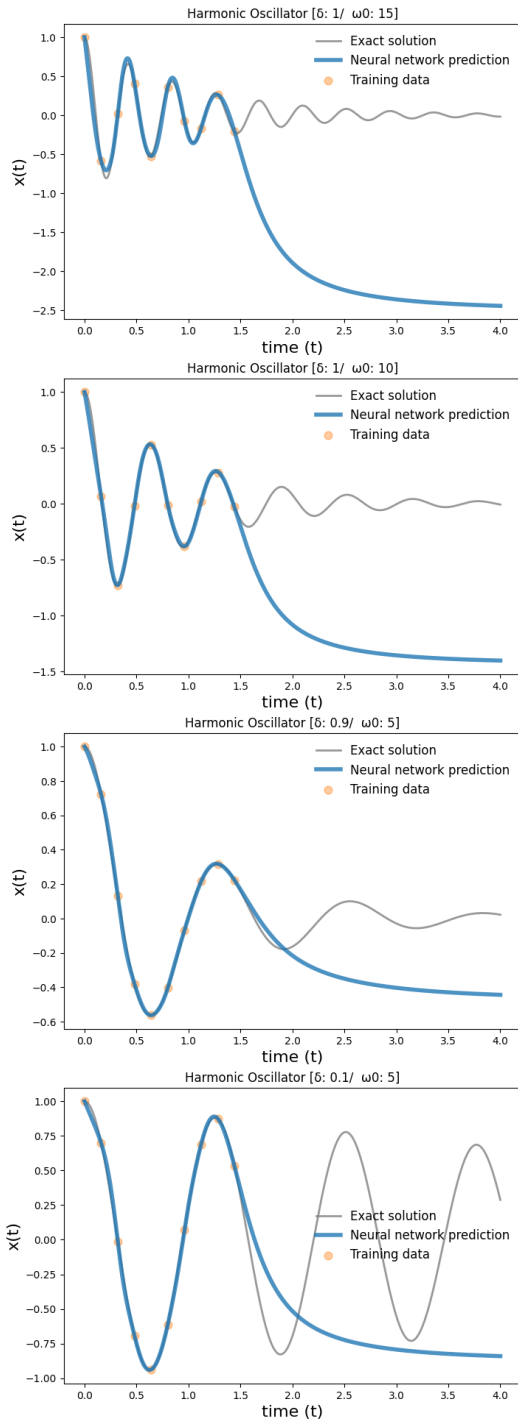
Training step: 80000 RMSE (%) : 24.54



Τα παρακάτω γραφήματα δείχνουν την πορεία εκπαίδευσης του πολυβάθμιου NN για πρόβλεψη 4 δευτερολέπτων

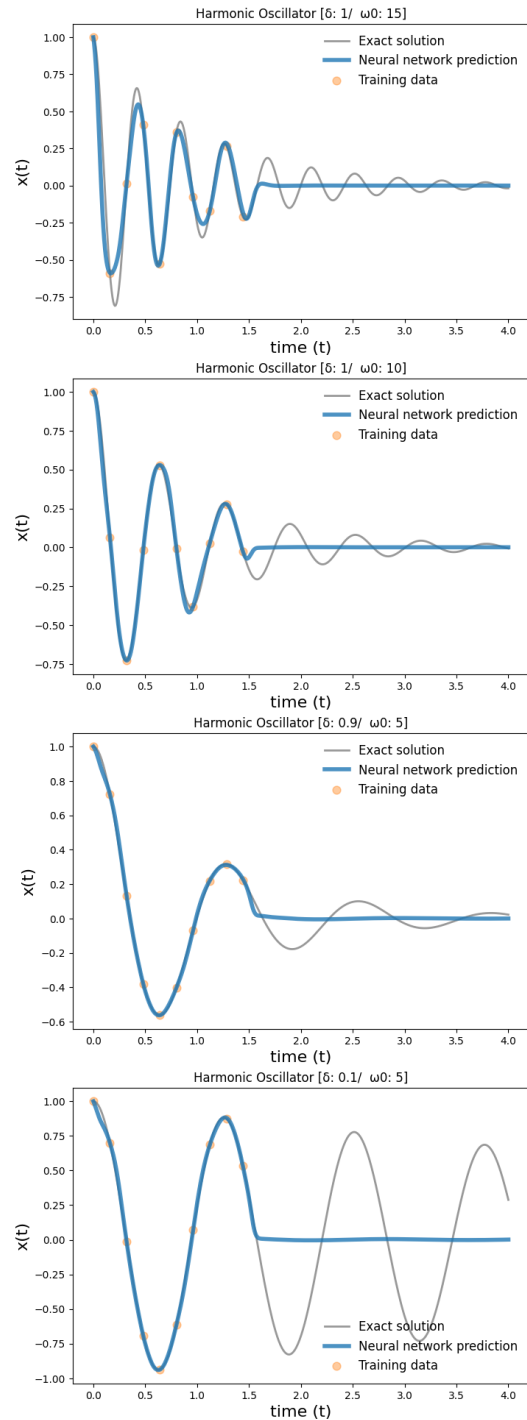
(α') Neural Network, Training step=80000, Prediction period=4 sec

Training step: 1000 RMSE (%) : 106.1



(β') PINN, Training step=80000, Prediction period=4 sec

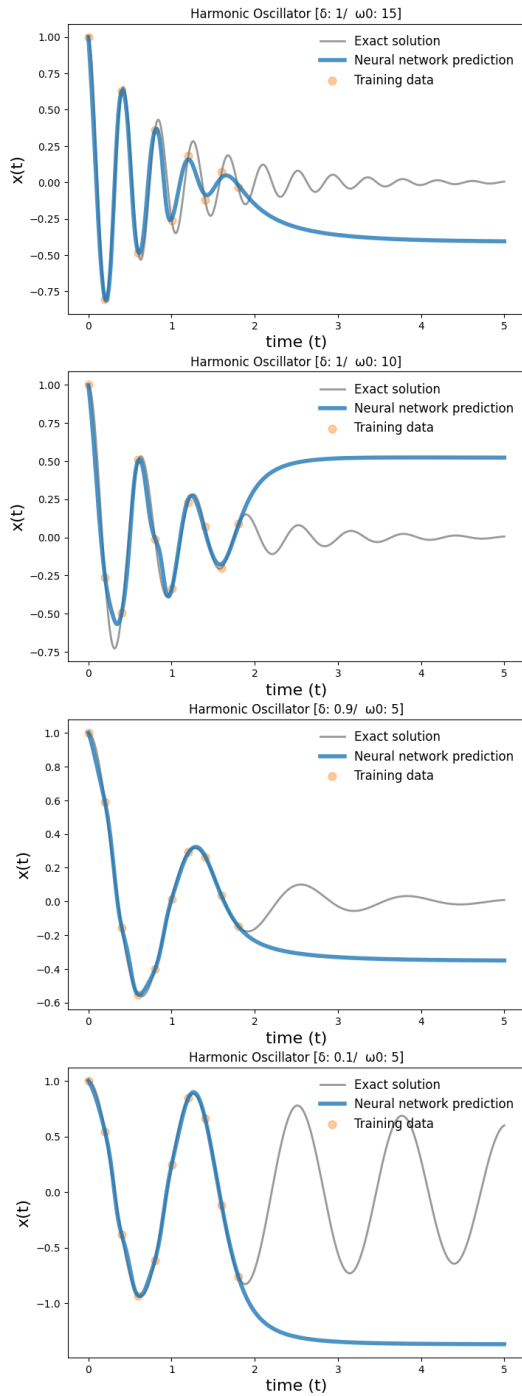
Training step: 80000 RMSE (%) : 22.15



Τα παρακάτω γραφήματα δείχνουν την πορεία εκπαίδευσης του πολυβάθμιου NN για πρόβλεψη 5 δευτερολέπτων

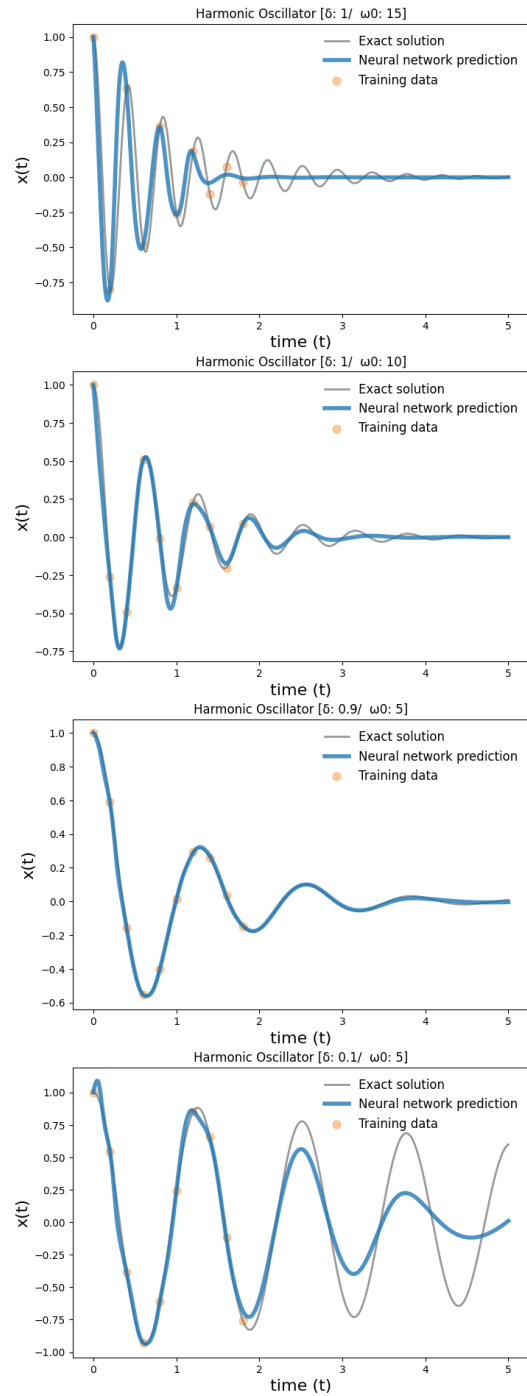
(α') Neural Network, Training step=80000, Prediction period=5 sec

Training step: 1000 RMSE (%) : 62.52



(β') PINN, Training step=80000, Prediction period=5 sec

Training step: 80000 RMSE (%) : 14.64



ΔΕΙΓΜΑΤΑ	NN-RMSE (%)	PINN-RMSE (%)
Δέκα ισοκαταμεμημένα δείγματα στα πρώτα 0.3 sec / πρόβλεψη 1 sec	50.02	0.2791
Δέκα ισοκαταμεμημένα δείγματα στα πρώτα 0.75 sec / πρόβλεψη 2 sec	52.07	0.3182
Δέκα ισοκαταμεμημένα δείγματα στα πρώτα 1.1 sec / πρόβλεψη 3 sec	59.8	24.54
Δέκα ισοκαταμεμημένα δείγματα στα πρώτα 1.5 sec / πρόβλεψη 4 sec	106.1	22.15
Δέκα ισοκαταμεμημένα δείγματα στα πρώτα 1.9 sec / πρόβλεψη 5 sec	62.52	14.64

Αυτός ο πίνακας δείχνει τα αποτελέσματα δύο διαφορετικών μοντέλων, ενός απλού νευρωνικού δικτύου και ενός νευρωνικού δικτύου με πληροφορίες φυσικής PINN για τον υπολογισμό της θέσης 4 διαφορετικών ελατηρίων. Τα δείγματα που αναλύθηκαν είναι δείγματα ομοιόμορφα καταμεμημένα μέσα στα πρώτα 0,3, 0,75, 1,1, 1,5, 1,9 δευτερόλεπτα παρατηρήσεων και η πρόβλεψη των μοντέλων είναι για μεγαλύτερο χρονικό διάστημα (1, 2, 3, 4, 5 δευτερόλεπτα αντίστοιχα). Το πεδίο "NN-RMSE" δείχνει την ακρίβεια του απλού νευρωνικού δικτύου και το πεδίο "PINN-RMSE" δείχνει την ακρίβεια του νευρωνικού δικτύου με πληροφορίες φυσικής. Η ακρίβεια μετράται χρησιμοποιώντας το RMSE (Root Mean Squared Error), μια μέτρηση που μετρά τον μέσο όρο των διαφορών στο τετράγωνο μεταξύ των προβλεπόμενων και των πραγματικών τιμών.

Μπορούμε να δούμε ότι, γενικά, το νευρωνικό δίκτυο με πληροφορία φυσικής έχει μεγαλύτερη ακρίβεια από το απλό νευρωνικό δίκτυο. Αυτό είναι ιδιαίτερα εμφανές στα δείγματα με παρατηρήσεις έως 0,75 δευτερόλεπτα και 1,1 δευτερόλεπτα, όπου το νευρωνικό δίκτυο με πληροφορία φυσικής έχει μηδενικό ή πολύ χαμηλό RMSE, ενώ το απλό νευρωνικό δίκτυο έχει υψηλό RMSE. Γενικά, όσο αυξάνεται η χρονική περίοδος των παρατηρήσεων, το RMSE και των δύο μοντέλων μειώνεται. Αυτό οφείλεται στο ότι καθώς έχουμε περισσότερες πληροφορίες για τη διαδικασία που προβλέπουμε, γίνεται ευκολότερο να προβλέψουμε το μέλλον.

Συνολικά, φαίνεται ότι το νευρωνικό δίκτυο με πληροφορίες φυσικής αποδίδει σημαντικά καλύτερα από το απλό νευρωνικό δίκτυο, ιδιαίτερα για μικρότερες χρονικές περιόδους παρατηρήσεων. Το νευρωνικό δίκτυο που ενημερώνεται για τη φυσική είναι σε θέση να κάνει καλύτερες προβλέψεις επειδή λαμβάνει υπόψη φυσικούς περιορισμούς και νόμους, οι οποίοι μπορεί να κάνουν τις προβλέψεις πιο ακριβείς. Από την άλλη πλευρά, το απλό νευρωνικό δίκτυο μπορεί να είναι περιορισμένο στην ικανότητά του να κάνει ακριβείς προβλέψεις επειδή δε λαμβάνει υπόψη φυσικούς περιορισμούς και νόμους. Αξίζει επίσης να σημειωθεί ότι υπάρχει σημαντική διαφορά στο RMSE των δύο μοντέλων για τα δείγματα με παρατηρήσεις έως 1,5 δευτερόλεπτο και 1,9 δευτερόλεπτα. Σε αυτές τις περιπτώσεις, το απλό νευρωνικό δίκτυο έχει πολύ χαμηλότερο RMSE

από το νευρωνικό δίκτυο με πληροφορίες φυσικής. Αυτό οφείλεται στην αύξηση της πολυπλοκότητας του προβλήματος που επιλύεται.

Συμπερασματικά, το νευρωνικό δίκτυο με πληροφορίες φυσικής αποδίδει σημαντικά καλύτερα από το απλό νευρωνικό δίκτυο για μικρότερες χρονικές περιόδους παρατηρήσεων, αλλά υπάρχει μια αξιοσημείωτη διαφορά στο RMSE για μεγαλύτερες χρονικές περιόδους. Συνολικά, αυτά τα αποτελέσματα υποδηλώνουν ότι το νευρωνικό δίκτυο που ενημερώνεται για τη φυσική μπορεί να είναι μια πολλά υποσχόμενη προσέγγιση για την πρόβλεψη ορισμένων τύπων διεργασιών, ιδιαίτερα εκείνων που υπόκεινται σε φυσικούς περιορισμούς και νόμους.

Κεφάλαιο 6

Μελλοντική εργασία

Στη διατριβή μας, έχουμε εξετάσει την απόδοση των PINNs και απλών νευρωνικών δικτύων για την πρόβλεψη της θέσης ενός αρμονικού ταλαντωτή. Ενώ τα αποτελέσματά μας έχουν δείξει ότι τα PINNs υπερτερούν των απλών νευρωνικών δικτύων, υπάρχει ακόμα πολλή δουλειά που πρέπει να γίνει σε αυτόν τον τομέα.

Μια περιοχή για μελλοντική έρευνα είναι η επέκταση της ανάλυσής μας για να περιλάβει πιο περίπλοκα και ρεαλιστικά σενάρια. Για παράδειγμα, θα ήταν ενδιαφέρον να διερευνήσουμε πώς λειτουργούν τα PINNs όταν τα δεδομένα εισόδου είναι θορυβώδη ή όταν το σύστημα που μοντελοποιείται είναι υπό την επίδραση εξωτερικών διαταραχών ή αβεβαιοτήτων.

Μια άλλη κατεύθυνση για μελλοντική έρευνα είναι η εξερεύνηση του δυναμικού των PINNs για άλλους τύπους φυσικών συστημάτων. Η χρήση των PINNs θα μπορούσε να είναι χρήσιμη σε πεδία όπως η μηχανική των ρευστών, η ηλεκτρομαγνητική και η κβαντοφυσική, όπου οι αναλυτικές λύσεις είναι δυσκολότερες ή αδύνατες να βρεθούν.

Επιπλέον, μια άλλη πιθανή κατεύθυνση είναι η εξέταση της απόδοσης των PINNs σε προβλήματα πολλαπλών φυσικών παραμέτρων. Αυτό μπορεί να επιτευχθεί με την ανάπτυξη μιας μεθόδου που επιτρέπει την αποτελεσματική παρακολούθηση πολλών μεταβλητών εξόδου ταυτόχρονα.

Συνολικά, οι προτεινόμενες κατευθύνσεις μελλοντικής έρευνας θα μπορούσαν να βοηθήσουν στη βελτίωση της κατανόησης των PINNs και στην εξέταση της απόδοσής τους σε διαφορετικά περιβάλλοντα και προβλήματα.

Παράρτημα Α΄

Ακρωνύμια και συντομογραφίες

NN Neural Network

PINN Physical Informed Neural Network

RMSE Root Mean Square Error

PDE Partial Differential Equations

ΤΝΔ Τεχνητά Νευρωνικά Δίκτυα

Bibliography

- [1] G. K. Batchelor, “An introduction to fluid dynamics (2nd pbk. ed.),” *Cambridge University Press*.
- [2] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0893608089900208>
- [3] A. Arzani and S. T. M. Dawson, “Data-driven cardiovascular flow modelling: examples and opportunities,” *Journal of The Royal Society Interface*, vol. 18, no. 175, feb 2021. [Online]. Available: <https://doi.org/10.1098%2Frsif.2020.0802>
- [4] A. Arzani, J.-X. Wang, and R. M. D'Souza, “Uncovering near-wall blood flow from sparse data with physics-informed neural networks,” *Physics of Fluids*, vol. 33, no. 7, p. 071905, jul 2021. [Online]. Available: <https://doi.org/10.1063%2F5.0055600>
- [5] S. Markidis, “The old and the new: Can physics-informed deep-learning replace traditional linear solvers?” 2021. [Online]. Available: <https://arxiv.org/abs/2103.09655>
- [6] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, “Automatic differentiation in machine learning: a survey,” 2015. [Online]. Available: <https://arxiv.org/abs/1502.05767>
- [7] I. Lagaris, A. Likas, and D. Fotiadis, “Artificial neural networks for solving ordinary and partial differential equations,” *IEEE Transactions on Neural Networks*, vol. 9, no. 5, pp. 987–1000, 1998. [Online]. Available: <https://doi.org/10.1109%2F72.712178>
- [8] H. Lee and I. S. Kang, “Neural algorithm for solving differential equations,” *Journal of Computational Physics*, vol. 91, no. 1, pp. 110–131, 1990. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/002199919090007N>

- [9] A. Meade and A. Fernandez, "The numerical solution of linear ordinary differential equations by feedforward neural networks," *Mathematical and Computer Modelling*, vol. 19, no. 12, pp. 1–25, 1994. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0895717794900957>
- [10] G. Stavroulakis, A. Avdelas, K. Abdalla, and P. Panagiotopoulos, "A neural network approach to the modelling, calculation and identification of semi-rigid connections in steel structures," *Journal of Constructional Steel Research*, vol. 44, no. 1, pp. 91–105, 1997, structural Steel Research in Greece. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0143974X97000394>
- [11] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations," 2017. [Online]. Available: <https://arxiv.org/abs/1711.10561>
- [12] W. Peng, J. Zhang, W. Zhou, X. Zhao, W. Yao, and X. Chen, "Idrlnet: A physics-informed neural network library," 2021. [Online]. Available: <https://arxiv.org/abs/2107.04320>
- [13] A. Muradova and G. Stavroulakis, "Physics-informed neural networks for elastic plate problems with bending and winkler-type contact effects," *Journal of the Serbian Society for Computational Mechanics*, vol. 15, pp. 45–54, 12 2021.
- [14] M. Raissi, A. Yazdani, and G. Karniadakis, "Hidden fluid mechanics: A navier-stokes informed deep learning framework for assimilating flow visualization data," 08 2018.
- [15] M. Raissi, P. Perdikaris, and G. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0021999118307125>
- [16] M. Raissi, A. Yazdani, and G. Karniadakis, "Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations," *Science*, vol. 367, p. eaaw4741, 01 2020.
- [17] S. Mishra and R. Molinaro, "Estimates on the generalization error of physics informed neural networks (pinns) for approximating a class of inverse problems for pdes," 2020. [Online]. Available: <https://arxiv.org/abs/2007.01138>
- [18] T. De Ryck, A. D. Jagtap, and S. Mishra, "Error estimates for physics informed neural networks approximating the navier-stokes equations," 2022. [Online].

Available: <https://arxiv.org/abs/2203.09346>