

Πανεπιστήμιο Δυτικής Μακεδονίας
Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών
Υπολογιστών

Σύγκριση συναρτήσεων ενεργοποίησης
για την εκπαίδευση νευρωνικών
δικτύων

Δημήτριος Μπάτσιος (ΑΜ: 1102)
Επιβλέπων Καθηγητής: Θωμάς Κυριακίδης

26 Φεβρουαρίου 2024

Περίληψη

Στον τομέα της Τεχνητής Νοημοσύνης τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks) αποτελούν μια προσέγγιση που έχει συγκεντρώσει τεράστιο επιστημονικό ενδιαφέρον. Ιδιαίτερα τα τελευταία χρόνια, όπου το πλήθος των διαθέσιμων δεδομένων έχει εκτοξευθεί λόγω του διαδικτύου, οι δυνατότητες και η αποτελεσματικότητα των ΤΝΔ έχουν αναδειχθεί σημαντικά. Οι συναρτήσεις ενεργοποίησης (Activation Functions), ως ο κύριος ρυθμιστής της βασικής μονάδας επεξεργασίας δεδομένων - του νευρώνα - διαδραματίζουν κομβικό ρόλο στην απόδοση του δικτύου. Στόχος της παρούσας διπλωματικής είναι να μελετήσει και να συγκρίνει τις γνωστές συναρτήσεις ενεργοποίησης αλλά και να εξαχθούν χρήσιμα συμπεράσματα σχετικά με την καταλληλότητα της κάθε συνάρτησης ανάλογα με το είδος του προβλήματος που καλείται να λυθεί.

Λέξεις κλειδιά: Τεχνητή Νοημοσύνη, Τεχνητά Νευρωνικά Δίκτυα, Βελτιστοποίηση, Συναρτήσεις Ενεργοποίησης

Abstract

In the field of Artificial Intelligence, Artificial Neural Networks have garnered significant scientific interest, particularly in recent years. The availability of vast amounts of data, thanks to the internet, has enhanced the potential of Artificial Neural Networks to excel. Activation functions acting as neurons main regulator affect greatly networks performance. The objective of this thesis is to systematically study and compare well-established activation functions in order to extract useful knowledge about the applicability of each function in different case scenarios.

Keywords: Artificial Intelligence, Artificial Neural Networks, Optimization, Activation Functions

Δήλωση Πνευματικών Δικαιωμάτων

Δήλωση Πνευματικών Δικαιωμάτων Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν. 1599/1986 και τα άρθρα 2,4,6 παρ. 3 του Ν. 1256/1982, η παρούσα Διπλωματική Εργασία με τίτλο "Σύγκριση συναρτήσεων ενεργοποίησης για την εκπαίδευση νευρωνικών δικτύων" καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας και αναφέρονται ρητώς μέσα στο κείμενο που συνοδεύουν, και η οποία έχει εκπονηθεί στο Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών του Πανεπιστημίου Δυτικής Μακεδονίας, υπό την επίβλεψη του μέλους του Τμήματος κ. Θωμά Κυριακίδη αποτελεί αποκλειστικά προϊόν προσωπικής εργασίας και δεν προσβάλλει κάθε μορφής πνευματικά δικαιώματα τρίτων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή / και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και μόνο.

Copyright (C) Δημήτριος Μπάτσιος & Θωμάς Κυριακίδης, 2024, Κοζάνη

Υπογραφή Φοιτητή

Περιεχόμενα

1	Εισαγωγή	10
1.1	Ορισμός του Προβλήματος	10
1.2	Κίνητρα και Στόχοι Υλοποίησης	11
1.3	Διάρθρωση Κειμένου	11
2	Τεχνητή Νοημοσύνη και Νευρωνικά Δίκτυα	13
2.1	Σύντομη Ιστορική Αναδρομή	13
2.2	Βιολογικός Νευρώνας και Νευρωνικά Δίκτυα	14
2.3	Τεχνητά Νευρωνικά Δίκτυα	15
2.3.1	Perceptron	15
2.3.2	Πολυεπίπεδο Perceptron - MLP	17
2.3.3	Συνελικτικά και Αναδρομικά Νευρωνικά Δίκτυα	19
3	Συναρτήσεις Ενεργοποίησης	22
3.1	Γενικά για τις Συναρτήσεις	22
3.2	Πρόβλημα Εξαφανιζόμενων Κλίσεων	23
3.3	Παρουσίαση Γνωστών Συναρτήσεων Ενεργοποίησης	24
3.3.1	Γραμμική - Linear	24
3.3.2	Βηματική - Step	25
3.3.3	Ανορθωμένη Γραμμική Μονάδα - ReLU	25
3.3.4	Παραλλαγές ReLU	27
3.3.5	Sigmoid	29
3.3.6	Tanh και Penalized Tanh	30
3.3.7	Softmax	32
3.3.8	Swish	33
3.3.9	Softplus	34

3.4	Βιβλιογραφική Ανασκόπηση	35
3.4.1	Συγκριτική Ανάλυση των Συναρτήσεων Ενεργοποίησης στα Τεχνητά Νευρωνικά Δίκτυα [1]	35
3.4.2	Ανασκόπηση και Σύγκριση Γνωστών Συναρτήσεων Ενεργοποίησης για Βαθιά Νευρωνικά Δίκτυα [2]	39
3.4.3	Συναρτήσεις Ενεργοποίησης στη Βαθιά Μάθηση: Περιεκτική Επισκόπηση και Συγκριτική Αξιολόγηση [3]	42
3.4.4	Συμπεράσματα Ευρύτερης Μελέτης	45
4	Υλοποίηση Πειραμάτων και Αποτελέσματα	46
4.1	Αναγνώριση Εικόνας	46
4.1.1	Αποτελέσματα Cifar-10	47
4.1.2	Αποτελέσματα Cifar-100	50
4.1.3	Αποτελέσματα Vegetable-Image	52
4.2	Εξόρυξη Γνώμης	57
4.2.1	Αποτελέσματα Movie-Reviews	59
4.2.2	Αποτελέσματα 20-Newsgroup	60
4.3	Δεδομένα σε Μορφή Πίνακα	63
4.3.1	Churn Modeling	65
4.3.2	House Price	65
4.3.3	Hotel Reviews	68
4.4	Δεδομένα Χρονοσειρών	71
4.4.1	Google Stock	73
4.4.2	Amazon Stock	74
4.5	Συμπεράσματα Πειραμάτων	78
4.5.1	Αναγνώριση Εικόνας	78
4.5.2	Εξόρυξη Γνώμης	81
4.5.3	Δεδομένα σε Μορφή Πίνακα	82
4.5.4	Δεδομένα Χρονοσειρών	85
5	Συμπεράσματα	88

Κατάλογος σχημάτων

2.1	Αναπαράσταση Βιολογικού Νευρώνα	14
2.2	Perceptron	16
2.3	Πολυεπίπεδος Perceptron	18
2.4	Παράδειγμα Συνελικτικού Νευρωνικού Δικτύου	19
2.5	Παράδειγμα Αναδρομικού Νευρωνικού Δικτύου	20
3.1	Παράδειγμα Αλυσιδωτής Παραγωγίσιμης για το w_5	23
3.2	Η Σιγμοειδής Συνάρτηση και η Παράγωγος της	24
3.3	Η Γραφική Παράσταση της Γραμμικής	25
3.4	Η Γραφική Παράσταση Βηματικής	26
3.5	Η Γραφική Παράσταση της ReLU	26
3.6	Η Γραφική Παράσταση της Leaky ReLU	28
3.7	Η Γραφική Παράσταση της Parametric ReLU	29
3.8	Η Γραφική Παράσταση Sigmoid	30
3.9	Η Γραφική Παράσταση της Tanh	31
3.10	Η Γραφική Παράσταση της Penalized Tanh	32
3.11	Η Γραφική Παράσταση της Softmax	33
3.12	Η Γραφική Παράσταση της Swish	34
3.13	Η Γραφική Παράσταση της Softplus	35
3.14	Το ΤΝΔ της [1]	36
3.15	Linear-Linear [1]	37
3.16	Linear-ReLU [1]	37
3.17	ReLU-Linear [1]	37
3.18	ReLU-ReLU [1]	38
3.19	Sigmoid-Linear [1]	38
3.20	Sigmoid-ReLU [1]	39

3.21 Το ΣΝΔ της [2]	40
3.22 Γραφική Παράσταση Ακρίβειας [2]	41
3.23 Χρόνος Εκπαίδευσης [2]	42
3.24 Χρόνος Κατάταξης στο Σύνολο Ελέγχου [2]	42
3.25 ΑΝΔ της [3]	43
4.1 Alexnet-Cifar10 Ακρίβεια	48
4.2 Alexnet-Cifar10-Χρονομέτρηση	48
4.3 VGG16-Cifar10 Ακρίβεια	49
4.4 VGG16-Cifar10 Χρονομέτρηση	49
4.5 CNN3-Cifar10 Ακρίβεια	50
4.6 CNN3-Cifar10 Χρονομέτρηση	50
4.7 Alexnet-Cifar100 Ακρίβεια	51
4.8 Alexnet-Cifar100 Χρονομέτρηση	52
4.9 VGG16-Cifar100 Ακρίβεια	52
4.10 VGG16-Cifar100 Χρονομέτρηση	53
4.11 CNN3-Cifar100 Ακρίβεια	53
4.12 CNN3-Cifar100 Χρονομέτρηση	54
4.13 Alexnet-Vegetables Ακρίβεια	54
4.14 Alexnet-Vegetables Χρονομέτρηση	55
4.15 VGG16-Vegetables Ακρίβεια	55
4.16 VGG16-Vegetables Χρονομέτρηση	56
4.17 CNN3-Vegetables Ακρίβεια	56
4.18 CNN3-Vegetables Χρονομέτρηση	57
4.19 Η Αρχιτεκτονική του TextCNN 4.19	58
4.20 Η Αρχιτεκτονική του LSTM1 4.20	58
4.21 TextCNN-Movie Reviews Ακρίβεια	59
4.22 TextCNN-Movie Reviews Χρονομέτρηση	60
4.23 LSTM1-Movie Reviews Ακρίβεια	60
4.24 LSTM1-Movie Reviews Χρονομέτρηση	61
4.25 TextCNN 20-Newsgroup Ακρίβεια	61
4.26 TextCNN 20-Newsgroup Χρονομέτρηση	62
4.27 LSTM1 20-Newsgroup Ακρίβεια	62

4.28 LSTM1 20-Newsgroup Χρονομέτρηση	63
4.29 Η Αρχιτεκτονική του DNN1 4.29	64
4.30 Η αρχιτεκτονική του DNN2 4.30	64
4.31 DNN1 Churn Modeling Ακρίβεια	66
4.32 DNN1 Churn Modeling Χρονομέτρηση	66
4.33 DNN2 Churn Modeling Ακρίβεια	67
4.34 DNN2 Churn Modeling Χρονομέτρηση	67
4.35 DNN1 House Price R2-Score	68
4.36 DNN1 House Price Χρονομέτρηση	68
4.37 DNN2 House Price R2-Score	69
4.38 DNN2 House Price Χρονομέτρηση	69
4.39 DNN1 Hotel Reviews R2-Score	70
4.40 DNN1 Hotel Reviews Χρονομέτρηση	70
4.41 DNN2 Hotel R2-Score	71
4.42 DNN2 Hotel Reviews Χρονομέτρηση	71
4.43 Η Αρχιτεκτονική του RNN1 4.43	72
4.44 Η Αρχιτεκτονική του RNN2 4.44	73
4.45 RNN1 Google Stock ΜΑΣ	74
4.46 RNN1 Google Stock Χρονομέτρηση	74
4.47 RNN2 Google Stock ΜΑΣ	75
4.48 RNN2 Google Stock Χρονομέτρηση	75
4.49 RNN1 Amazon Stock ΜΑΣ	76
4.50 RNN1 Amazon Stock Χρονομέτρηση	76
4.51 RNN2 Amazon Stock ΜΑΣ	77
4.52 RNN2 Amazon Stock Χρονομέτρηση	77

Κατάλογος πινάκων

3.1	Αποτελέσματα Ακρίβειας Κατάταξης Δεδομένων	41
3.2	Μετάφραση Κειμένου	44
3.3	Αναγνώριση Ομιλίας	44
4.1	Αναγνώριση Εικόνας - Τορ6	78
4.2	Αναγνώριση Εικόνας: Τορ6 Κατάταξη Ακρίβειας	79
4.3	Αναγνώριση Εικόνας: Χρονικές Επιδόσεις Τορ6	79
4.4	Αναγνώριση Εικόνας: Τορ6 Χρονική Κατάταξη	79
4.5	Εξόρυξη Γνώμης - Τορ6	81
4.6	Εξόρυξη Γνώμης: Κατάταξη Ακρίβειας	81
4.7	Εξόρυξη Γνώμης - Χρονικές Επιδόσεις	81
4.8	Εξόρυξη Γνώμης: Χρονική Κατάταξη	82
4.9	Δεδομένα σε Μορφή Πίνακα - Τορ6	83
4.10	Δεδομένα σε Μορφή Πίνακα - Κατάταξη Ακρίβειας	83
4.11	Δεδομένα σε Μορφή Πίνακα - Χρονικές Επιδόσεις	83
4.12	Δεδομένα σε Μορφή Πίνακα - Χρονικές Επιδόσεις	83
4.13	Δεδομένα Χρονοσειρών - Τορ6	85
4.14	Δεδομένα Χρονοσειρών - Κατάταξη Ακρίβειας	86
4.15	Δεδομένα σε Χρονοσειρών - Χρονικές Επιδόσεις	86
4.16	Δεδομένα Χρονοσειρών - Χρονική Κατάταξη	86

Κεφάλαιο 1

Εισαγωγή

1.1 Ορισμός του Προβλήματος

Η τεχνητή νοημοσύνη γνωρίζει ραγδαία ανάπτυξη την τελευταία δεκαετία. Το εύρος των τομέων στους οποίους εφαρμόζεται αλλά και τα θεαματικά αποτελέσματα που παράγει σε επιστήμες όπως η ιατρική, γεννούν αισιοδοξία σχετικά με την αντιμετώπιση προβλημάτων που απασχολούν την ανθρωπότητα. Ακόμα ένας τομέας όπου η τεχνητή νοημοσύνη βρίσκει εφαρμογή είναι οι χρηματιστηριακές προβλέψεις. Από αυτό γίνεται σαφές πως αναφερόμαστε σε ένα ισχυρό εργαλείο το οποίο αλλάζει τα δεδομένα και δεν επιτρέπει σε όσους συμμετέχουν σε τέτοια ανταγωνιστικά πεδία - όπως το χρηματιστήριο - να μείνουν ουραγοί των εξελίξεων. Επίσης, τα τελευταία χρόνια βλέπουμε πως η τεχνητή νοημοσύνη δεν σταματά στην εφαρμογή της σε εξειδικευμένους τομείς - όπως αυτούς που αναφέραμε παραπάνω - αλλά ενσωματώνεται στην καθημερινότητα όλο και περισσότερων ανθρώπων, οι οποίοι τη χρησιμοποιούν υποστηρικτικά ακόμα και στα επαγγελματικά τους καθήκοντα. Χαρακτηριστικό παράδειγμα είναι το ChatGPT, μια εφαρμογή τεχνητής νοημοσύνης η οποία κυκλοφόρησε τον Νοέμβριο του 2022 από την OpenAI.

Καταλαβαίνουμε πως είναι αναμενόμενο το ενδιαφέρον της επιστημονικής κοινότητας για τη μελέτη του τομέα της τεχνητής νοημοσύνης και των Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ), με σκοπό την ανάπτυξη και την αντιμετώπιση των προκλήσεων που παρουσιάζονται σε αυτά. Εκ των βασικών παραμέτρων των ΤΝΔ και αντικείμενο μελέτης της παρούσας διπλωματικής είναι οι συναρτήσεις ενεργοποίησης, οι οποίες μεταβάλλουν την είσοδο που εισέρχεται εντός κάθε νευρώνα, ο οποίος πυροδοτεί το αποτέλεσμα της συνάρτησης προς τους επόμενους νευρώνες. Έτσι

οι συναρτήσεις ενεργοποίησης διαδραματίζουν πρωταγωνιστικό ρόλο στην προβλεπτική ικανότητα του δικτύου αλλά και στη διαδικασία της εκπαίδευσής του, καθώς με γνώμονα την παράγωγο της συνάρτησης καθορίζονται οι αλλαγές που απαιτούνται στις παραμέτρους του δικτύου με σκοπό την καλύτερη προσαρμογή του στο πρόβλημα που καλείται να αντιμετωπίσει.

1.2 Κίνητρα και Στόχοι Υλοποίησης

Στόχος αυτής της διπλωματικής είναι να παρουσιάσει τις γνωστότερες συναρτήσεις ενεργοποίησης και να μελετήσει - μέσω της διαθέσιμης βιβλιογραφία ς- τις περιπτώσεις χρήσης τους στις διαφορετικές αρχιτεκτονικές ΤΝΔ που υπάρχουν και σε διαφορετικά σύνολα δεδομένων (datasets). Επίσης, θα πραγματοποιηθούν πειράματα για τη σύγκριση μεταξύ των συναρτήσεων ενεργοποίησης, με σκοπό την εξαγωγή χρήσιμων συμπερασμάτων σχετικά με την καταλληλότητα της κάθε συνάρτησης ανά περίπτωση και τη σχέση της με τις άλλες παραμέτρους του δικτύου. Ευελπιστούμε μέχρι το τέλος να έχουν δοθεί αρκετά και τεκμηριωμένα δεδομένα ώστε η παρούσα διπλωματική να συμβάλει στην κατανόηση του αναγνώστη πάνω στο θέμα, την παροχή γνώσης για την κατάλληλη επιλογή συνάρτησης ενεργοποίησης, και την ανάδειξη της σημαντικότητας του θέματος με αποτέλεσμα την παρακίνηση για περαιτέρω έρευνα.

1.3 Διάρθρωση Κειμένου

Στα επόμενα κεφάλαια ακολουθούν τα εξής: Στο δεύτερο κεφάλαιο γίνεται μια σύντομη ιστορική αναδρομή στον τομέα της τεχνητής νοημοσύνης και των νευρωνικών δικτύων. Στη συνέχεια παρουσιάζεται ο βιολογικός νευρώνας ο οποίος ήταν η έμπνευση για την δημιουργία των ΤΝΔ. έπειτα το Perceptron. Το πρώτο ΤΝΔ το οποίο αποτέλεσε τη βάση για τις επόμενες αρχιτεκτονικές που αναφέρουμε.

Στο τρίτο κεφάλαιο αναλύονται οι παρακάτω συναρτήσεις ενεργοποίησης:

- Step
- Linear
- ReLU

-
- Leaky ReLU
 - Parametric ReLU
 - Sigmoid
 - Tanh
 - Softmax
 - Swish
 - Softplus

Στη συνέχεια γίνεται μια ανασκόπηση της διαθέσιμης βιβλιογραφίας, σχολιάζονται σχετικές δημοσιεύσεις και εξάγονται συμπεράσματα από τα ευρήματα αυτών.

Το τέταρτο κεφάλαιο αφορά τα πειράματα που πραγματοποιήθηκαν. Παρουσιάζεται η διαδικασία των πειραμάτων, παρατίθενται τα αποτελέσματα και γίνεται ανάλυση τους. Στο πέμπτο κεφάλαιο συνοψίζονται τα όσα παρουσιάστηκαν, εξάγονται γενικά συμπεράσματα και αναφέρονται πιθανές προεκτάσεις του θεμάτων της διπλωματικής εργασίας.

Κεφάλαιο 2

Τεχνητή Νοημοσύνη και Νευρωνικά Δίκτυα

2.1 Σύντομη Ιστορική Αναδρομή

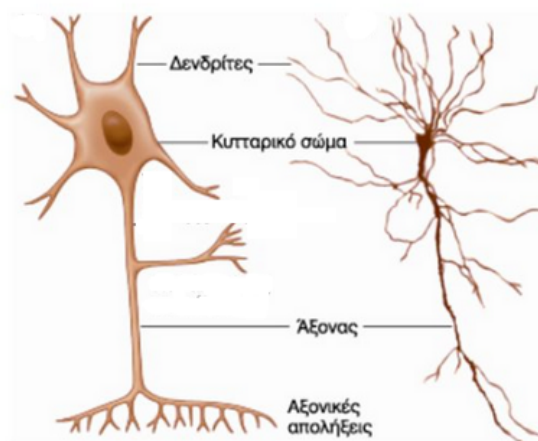
Από την αρχαιότητα υπήρχαν μύθοι που αφορούσαν νοήμονες μηχανές. Ωστόσο χρειάστηκε να φτάσουμε στη δεκαετία του 1940-50 και την έλευση της πρώτης γενιάς υπολογιστών ώστε να περάσουμε από τον μύθο στο όραμα. Συγκεκριμένα ήταν το 1943 όπου ο νευροφυσιολόγος Warren McCulloch και ο μαθηματικός Walter Pitts έγραψαν ένα άρθρο [4] περιγράφοντας τη λειτουργία των νευρώνων με τη βοήθεια της Άλγεβρας Boole. Επηρεασμένος από αυτές τις ιδέες ο Frank Rosenblatt δημιούργησε το Perceptron [5], το πρώτο Τεχνητό Νευρωνικό Δίκτυο (Artificial Neural Network - ANN). Παρά την αισιοδοξία που γέννησε το έργο του Rosenblatt, περιορισμοί όπως οι χαμηλές επιδόσεις των επεξεργαστών της εποχής καθώς και η αδυναμία του Perceptron να διαχειριστεί μη γραμμικά διαχωρίσιμα προβλήματα, οδήγησαν στην υποχρηματοδότηση και την έλλειψη ενδιαφέροντος για τον τομέα. Η περίοδος που ακολούθησε έμεινε γνωστή ως “Ο Χειμώνας της Τεχνητής Νοημοσύνης”. Η αναθέρμανση του ενδιαφέροντος ήρθε στα μέσα της δεκαετίας του 1970-80 λόγω των εμπορικών εφαρμογών που απέκτησαν τα έμπειρα συστήματα (Expert Systems). Έπειτα, με τη συνεχή βελτίωση του αρχικού μοντέλου Perceptron και την εισαγωγή του αλγόριθμου οπισθοδρόμησης (Backpropagation) -ο οποίος αντιμετώπισε με επιτυχία τα προβλήματα που προέκυπταν με στο μονοεπίπεδο Perceptron- οι ερευνητές έστρεψαν και πάλι το ενδιαφέρον τους στο πεδίο. Με το πέρασμα στον 21ο αιώνα, η ραγδαία αύξηση της υπολογιστικής ισχύς και η διάθεση μεγάλων συνό-

λων δεδομένων (datasets) ανοίγει τον δρόμο για τα Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks) και βλέπουμε για πρώτη φορά αρχιτεκτονικές όπως τα Συνελικτικά Νευρωνικά Δίκτυα (Convolution Neural Networks - CNN) και τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks - RNN). Τα πεδία όπου η τεχνητή νοημοσύνη βρίσκει εφαρμογή είναι αμέτρητα. Πλέον οι υπολογιστές αναγνωρίζουν και ταξινομούν εικόνες, αναλύουν τον προφορικό λόγο, δημιουργούν τέχνη και νικούν τον άνθρωπο σε παίγνια αφηρημένης στρατηγικής (AlphaGo).

2.2 Βιολογικός Νευρώνας και Νευρωνικά Δίκτυα

Όπως είναι αναμενόμενο, έμπνευση για τη δημιουργία των ΤΝΔ υπήρξαν τα βιολογικά νευρωνικά δίκτυα, τα οποία ανακαλύφθηκαν από την επιστήμη της νευροφυσιολογίας στη προσπάθεια μελέτης και κατανόησης του εγκεφάλου. Βασικό δομικό στοιχείο του οποίου είναι ο νευρώνας (Σχήμα 2.1). Οι Νευρώνες είναι κύτταρα τα οποία βρίσκονται στον εγκέφαλο και αποτελούνται από τα εξής τμήματα: τους δενδρίτες, τον πυρήνα, τον άξονα και τις συνάψεις.

Σχήμα 2.1: Αναπαράσταση Βιολογικού Νευρώνα



<https://player.slideplayer.gr/41/11206117/data/images/img1.jpg>

Ένας νευρώνας δέχεται ηλεκτρικά σήματα και οι δενδρίτες είναι οι υποδοχείς των σημάτων αυτών. Στη συνέχεια το σήμα περνάει στο κεντρικό τμήμα του νευρώνα - το σώμα - και από εκεί μέσω του άξονα φτάνει στις συνάψεις όπου και πυροδοτείται σε άλλους νευρώνες. Το "ταξίδι" αυτό των ηλεκτρικών σημάτων εντός των νευρώνων και μεταξύ αυτών δεν είναι τόσο απλό. Εάν η διαδικασία που πε-

ριγράψαμε παραπάνω δεν αποτελούσε μια απλοποίηση και απέδιδε πλήρως την λειτουργία των νευρώνων τότε θα αντιλαμβανόμασταν τους νευρώνες ως απλούς αγωγούς ηλεκτρικών σημάτων. Κάτι τέτοιο ωστόσο δεν ισχύει. Πιο συγκεκριμένα κάθε νευρώνας συλλέγει όλο το ηλεκτρικό φορτίο που δέχεται στους δενδρίτες του. Έπειτα μόνο αν το άθροισμα του φορτίου υπερβαίνει κάποιο κατώφλι (threshold) τότε ο άξονας του πυροδοτεί ηλεκτρικούς παλμούς με μεγάλη συχνότητα, διαφορετικά παράγει παλμούς πολύ αραιά και σε τυχαίες χρονικές στιγμές. Στη 2η περίπτωση λέμε πως ο νευρώνας είναι αδρανής. Ενώ στη περίπτωση που το φορτίο υπερβαίνει το κατώφλι τότε ο ηλεκτρικός παλμός ταξιδεύει μέσω του άξονα όπου φτάνει στις συνάψεις οι οποίες αποτελούν τα σημεία ένωσης μεταξύ του άξονα του νευρώνα και των δενδριτών των άλλων νευρώνων. Σε αυτό το σημείο όπου το σήμα περνάει από έναν νευρώνα στον επόμενο υπάρχουν παράγοντες που επηρεάζουν τη μετάδοση αυτή. Το πλάτος της σύναψης, η απόσταση της από τον δενδρίτη του άλλου νευρώνα και η πυκνότητα του ηλεκτροχημικού υλικού είναι οι παράγοντες αυτοί. Το ποσοστό της ηλεκτρικής δραστηριότητας που τελικά μεταδίδεται ονομάζεται συναπτικό βάρος. Πρέπει να σημειωθεί πως οι συνάψεις χωρίζονται σε ενισχυτικές (excitatory) και ανασταλτικές (inhibitory) ανάλογα με το εάν το φορτίο που μεταδίδουν διεγείρει τον νευρώνα-υποδοχέα για να παράγει παλμούς με μεγαλύτερη συχνότητα ή τον καταστέλλει εμποδίζοντας τον να παράγει παλμούς.

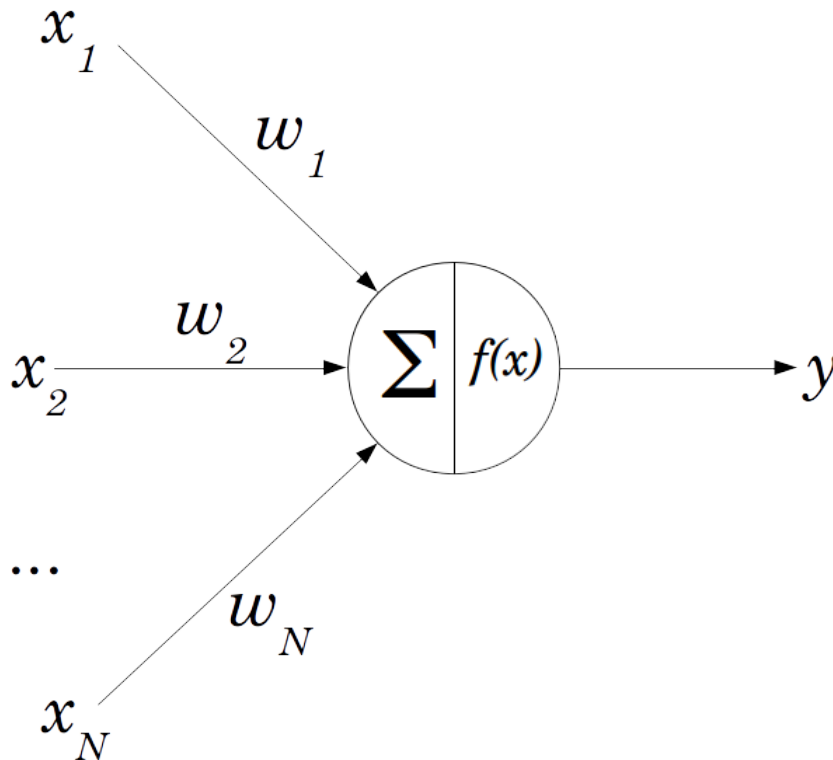
2.3 Τεχνητά Νευρωνικά Δίκτυα

2.3.1 Perceptron

Το Perceptron το οποίο αναφέραμε παραπάνω έχει απλή δομή η οποία φαίνεται καθαρά στο Σχήμα 2.2.

Σε ευθεία αναλογία με ένα βιολογικό δίκτυο νευρώνων τα ηλεκτρικά σήματα που εισέρχονται σε έναν νευρώνα μέσω των δενδριτών, μπορούν μαθηματικά να αναπαρασταθούν ως ανεξάρτητες μεταβλητές σε μια συνάρτηση, με τιμή της κάθε ανεξάρτητης μεταβλητής ανάλογη της έντασης του σήματος. Άρα αν N ο αριθμός των σημάτων που εισέρχονται στον νευρώνα μπορούμε να αναπαραστήσουμε το καθένα ως x_1, x_2, \dots, x_n . Δεν πρέπει να ξεχνάμε πως υπάρχουν παράγοντες που επηρεάζουν τη μετάδοση ενός σήματος κάτι το οποίο μπορεί να αναπαρασταθεί με

Σχήμα 2.2: Perceptron



<https://lucidar.me/en/neural-networks/files/perceptron.png>

τον πολλαπλασιασμό της κάθε μεταβλητής με έναν παράγοντα W που ονομάζεται βάρος. Φτάνοντας τα σήματα στο σώμα του νευρώνα αθροίζονται. Καθώς το κάθε σήμα έχει διαφορετικό βάρος το άθροισμα αυτών αναπαριστάται ως: $\sum_{i=1}^n x_i \times w_i + b$ όπου b είναι η πόλωση (bias), ένας σταθερός παράγοντας ανεξάρτητος από τις εισόδους. Τέλος, το αποτέλεσμα της άθροισης περνάει από μια συνάρτηση (γνωστή ως συνάρτηση ενεργοποίησης) για να δώσει την τελική έξοδο. Στον Perceptron η συνάρτηση αυτή είναι η βηματική (Step Function), η οποία περιγράφεται από τον τύπο :

$$\text{binary}(x, y) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Όπως καταλαβαίνουμε από τον τύπο της βηματικής συνάρτησης η τελική έξοδος του Perceptron είναι 0 ή 1.

Μέχρι εδώ παρουσιάστηκε το πως λαμβάνονται τα δεδομένα, επεξεργάζονται και δίνουν ένα αποτέλεσμα. Αλλά πως επιτυγχάνεται η διαδικασία της μάθησης;

Κάθε πρότυπο (datapoint) εισέρχεται στο δίκτυο με τη μορφή $x_1, x_2 \dots x_n$ και το δίκτυο - αφού επεξεργαστεί τα δεδομένα όπως αναφέρεται παραπάνω - παράγει μια έξοδο, στην περίπτωση του Perceptron 0 ή 1. Εάν η έξοδος είναι σωστή δεν εφαρμόζονται αλλαγές στο δίκτυο. Εάν η έξοδος είναι λάθος τότε πραγματοποιείται αλλαγή των βαρών $w_1, w_2 \dots w_n$. Ο κανόνας της αλλαγής αυτής περιγράφεται από τη σχέση :

$$w_i \leftarrow w_i + \eta \cdot (y - \hat{y}) \cdot x_i$$

w_i : Η τιμή του βάρους

η : Η παράμετρος η είναι ο ρυθμός μάθησης. Συνήθως είναι ένας μικρός ακέραιος και χρησιμοποιείται για να ρυθμίσει το μέγεθος της διόρθωσης που θα εφαρμοστεί στα βάρη.

y : Η σωστή έξοδος που θα θέλαμε το δίκτυο να δώσει

\hat{y} : Η λανθασμένη έξοδος που έδωσε το δίκτυο.

Επίσης, ανανεώνεται και η παράμετρος της πόλωσης σύμφωνα με τη σχέση:

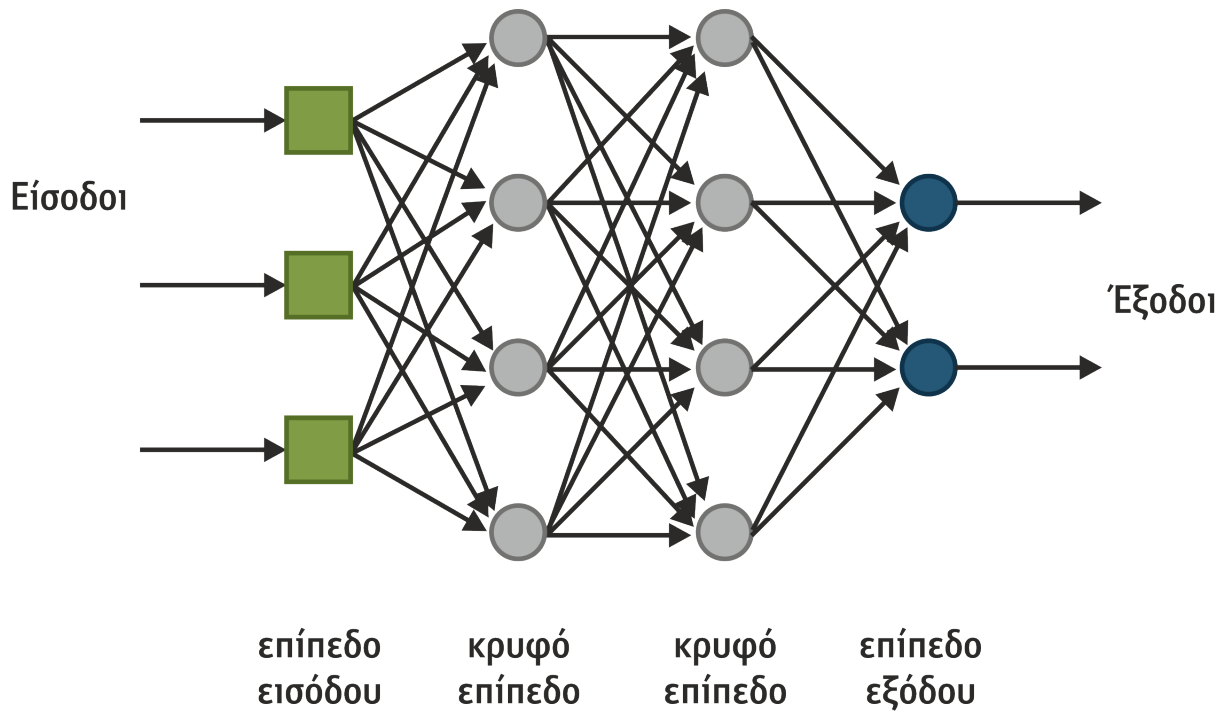
$$b \leftarrow b + \eta \cdot (y - \hat{y})$$

Σε κάθε πρότυπο εφαρμόζεται η παραπάνω διαδικασία. Όταν όλα τα πρότυπα περάσουν μία φορά από το δίκτυο, λέμε πως πέρασε μια εποχή (Epoch). Έτσι, ένα απλό δίκτυο όπως το Perceptron μαθαίνει από την εμπειρία του με σκοπό να επιτύχει καλύτερα ποσοστά πρόβλεψης. Ωστόσο όπως έχει αναφερθεί, το Perceptron δεν καταφέρνει να λύσει όλα τα προβλήματα, κάτι που απέδειξαν ο Marvin Minsky και ο Seymour Papert. Η δημοσίευση τους [6], η οποία αναδεικνύει την αδυναμία του Perceptron να διαχειριστεί μη γραμμικά διαχωρίσιμα προβλήματα, άνοιξε τον δρόμο για τη δημιουργία πιο εκλεπτυσμένων αρχιτεκτονικών όπως το πολυεπίπεδο Perceptron (Multi-Layer Perceptron MLP), ικανών να ξεπεράσουν τις δυσκολίες που αντιμετώπισε το Perceptron.

2.3.2 Πολυεπίπεδο Perceptron - MLP

Η αρχιτεκτονική ενός πολυεπίπεδου Perceptron με δύο κρυφά επίπεδα, φαίνεται στο Σχήμα 2.3.

Σχήμα 2.3: Πολυεπίπεδος Perceptron



http://repfiles.kallipos.gr/html_books/93/img_book/sxima_4,21.png

Αποτελεί αναβάθμιση του απλού Perceptron ενώ διατηρεί τις βασικές αρχές του προκατόχου του. Μια εκ των βασικών διαφορών είναι πως έχει κρυφά επίπεδα, κάθε ένα εκ των οποίων έχει νευρώνες-κόμβους. Έτσι η αρχική είσοδος περνάει από το πρώτο επίπεδο νευρώνων, οι οποίοι πυροδοτούν προς το δεύτερο επίπεδο νευρώνων οι οποίοι με τη σειρά πυροδοτούν προς το επόμενο κ.ο.κ. μέχρι το επίπεδο εξόδου. Συνεπώς η λέξη δίκτυο αποκτά μεγαλύτερο νόημα στον MLP καθώς μιλάμε για συνδεδεμένους νευρώνες που συνεργατικά παράγουν το αποτέλεσμα και όχι για έναν μοναδικό νευρώνα που επεξεργάζεται την είσοδο και παράγει την έξοδο όπως στη περίπτωση του μονοεπίπεδου Perceptron.

Κατά την εκπαίδευσή του, το δίκτυο MLP χρησιμοποιεί τον αλγόριθμο οπισθοδρόμησης (Backpropagation) προκειμένου να ανανεώσει τα βάρη με σκοπό να μειώσει το σφάλμα εξόδου. Αυτό το κάνει εξετάζοντας τα βάρη ξεκινώντας από το τελευταίο K επίπεδο, με το οποίο μόλις ολοκληρώσει τις αλλαγές κινείται ανάποδα προς το προ-τελευταίο $K - 1$ επίπεδο κ.ο.κ μέχρι να φτάσει και να ανανεώσει τα βάρη στο 1ο επίπεδο, τότε ο αλγόριθμος ολοκληρώνεται. Η ανανέωση των βαρών πραγματοποιείται υπολογίζοντας την κλίση της συνάρτησης απωλειών (loss

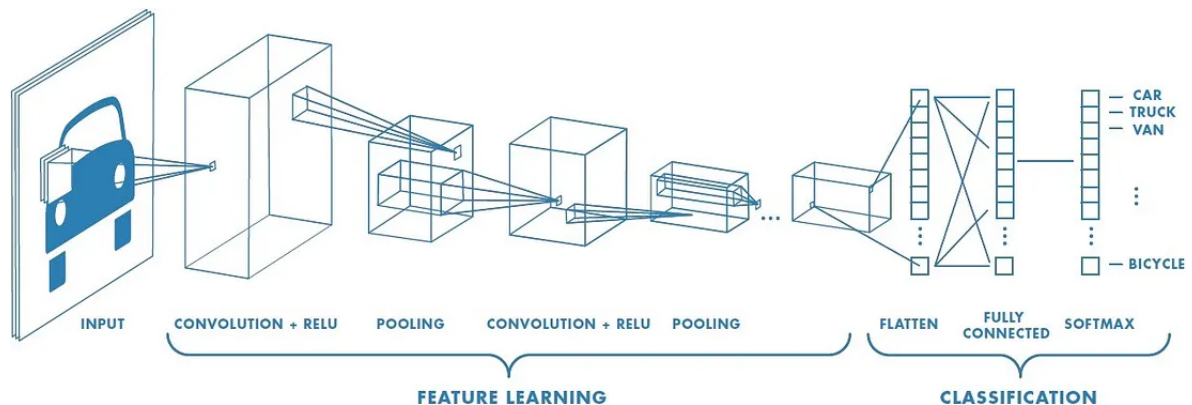
function) με τη μέθοδο της βαθμωτής κατάβασης (Gradient Descent). Θεωρώντας τα βάρη του επιπέδου που εξετάζεται ως τις ανεξάρτητες μεταβλητές της συνάρτησης απωλειών, η αρνητική κλίση μας δείχνει τις μεταβολές στα βάρη οι οποίες ελαχιστοποιούν τις απώλειες.

Από τα παραπάνω γίνεται σαφές πως πρέπει να υπολογιστεί η παράγωγος προκειμένου να επιτευχθεί σωστά η διαδικασία της μάθησης στον MLP. Έτσι είναι απαραίτητη προϋπόθεση η συνάρτηση ενεργοποίησης να είναι παραγωγίσιμη, κάτι το οποίο δεν ισχύει στην περίπτωση της βηματικής συνάρτησης. Παρακάτω θα παρουσιάζονται αναλυτικά οι γνωστότερες συναρτήσεις ενεργοποίησης που χρησιμοποιούνται σήμερα. Πριν από αυτό αξίζει να γίνει αναφορά σε δύο αρχιτεκτονικές ΤΝΔ που χρησιμοποιούνται τα τελευταία χρόνια να με σκοπό να αντιμετωπίσουν αποτελεσματικά συγκεκριμένα προβλήματα.

2.3.3 Συνελικτικά και Αναδρομικά Νευρωνικά Δίκτυα

Τα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ) είναι κατάλληλα για την αναγνώριση εικόνας (image recognition) και έχουν σημειώσει θεαματικά αποτελέσματα [7]. Η αρχιτεκτονική τους φαίνεται στο Σχήμα 2.4.

Σχήμα 2.4: Παράδειγμα Συνελικτικού Νευρωνικού Δικτύου



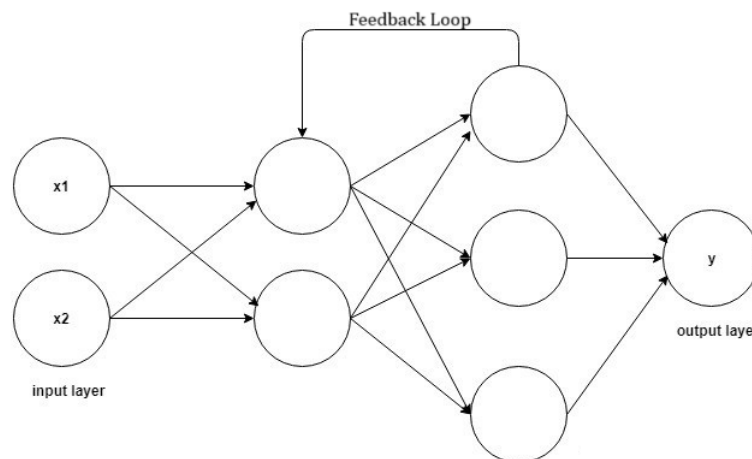
https://miro.medium.com/v2/resize:fit:1100/format:webp/1*vkQ0hXDaQv57sALXAJquxA.jpeg

Στα ΣΝΔ παίρνουμε συνήθως ως είσοδο μια εικόνα $M \times N \times C$ διαστάσεων, όπου C είναι ο αριθμός των καναλιών. Η διαφορά με ένα δίκτυο MLP, είναι η παρουσία συνελικτικών επιπέδων. Τα συνελικτικά επίπεδα εφαρμόζουν φίλτρα στην εικόνα. Η διαδικασία εφαρμογής του φίλτρου στα μαθηματικά είναι η πράξη της συνέλιξης. Ένα φίλτρο μπορεί παραδείγματος χάρη να μετατρέπει την εικόνα από πολύχρωμη σε ασπρόμαυρη. Αυτό εξυπηρετεί στη μείωση του όγκου δεδομένων της εικόνας

από τρία κανάλια (RGB) σε ένα. Σε κάθε επίπεδο μπορούν να εφαρμοστούν και πάνω από ένα φίλτρο. Το αποτέλεσμα της εφαρμογής κάθε φίλτρου ονομάζεται χάρτης χαρακτηριστικών. Κάθε χάρτης περνάει στο επόμενο συνελικτικό επίπεδο όπου εφαρμόζεται η ίδια διαδικασία. Η εφαρμογή των φίλτρων επιτρέπει στο ΣΝΔ να μειώνει τον όγκο των δεδομένων και να εντοπίσει χαρακτηριστικά στην εικόνα όπως φωτεινά σημεία, γραμμές ή καμπύλες κτλ. Αυτό βοηθά στον εντοπισμό πιο αφηρημένων χαρακτηριστικών όπως μάτια, αυτιά κλπ τα οποία με τη σειρά τους βοηθούν στον εντοπισμό και αναγνώριση πιο αφηρημένων προτύπων όπως άνθρωποι, δέντρα κ.ο.κ.

Τέλος, παρουσιάζονται τα Αναδρομικά Νευρωνικά Δίκτυα (ΑΝΔ) η αρχιτεκτονική των οποίων φαίνεται στο Σχήμα 2.5.

Σχήμα 2.5: Παράδειγμα Αναδρομικού Νευρωνικού Δικτύου



https://ambrapaliaidata.blob.core.windows.net/ai-storage/articles/1_UGxpbXV.jpg

Τα ΑΝΔ είναι κατάλληλα για προβλήματα χρονοσειρών. Ένα παράδειγμα τέτοιου προβλήματος είναι η μετάφραση κειμένου. Στη μετάφραση κειμένου το αποτέλεσμα-έξοδος που θα μας δώσει το δίκτυο δεν εξαρτάται μόνο από την είσοδο-λέξη αλλά και από προηγούμενα δεδομένα. Για παράδειγμα στην πρόταση “Is Maria brave?” η λέξη “brave” θα πρέπει να μεταφραστεί ως “γενναία” και όχι ως “γενναίος” καθώς αναφέρεται στη Maria. Άρα η λέξη “brave” ως είσοδος δεν είναι από μόνη της αρκετή ώστε να παράγει το σωστό αποτέλεσμα. Ένα ΑΝΔ αντιμετωπίζει τέτοια προβλήματα με το βρόγχο ανατροφοδότησης (feedback loop). Σε ένα ΑΝΔ, το αποτέλεσμα που παράγεται από ένα κόμβο-νευρώνα δεν ακολουθεί μόνο την

έξοδο αλλά και ανατροφοδοτείται στο δίκτυο καθώς αποτελεί χρήσιμη είσοδο για τις επόμενες προβλέψεις. Έτσι, στο παράδειγμα της παραπάνω πρότασης η λέξη “Maria” μεταφράστηκε ως “Μαρία” και ανατροφοδοτήθηκε στο δίκτυο καθώς είναι απαραίτητη για τη μετάφραση της επόμενης λέξης “brave”.

Κλείνοντας αυτό το κεφάλαιο πρέπει να αναφερθεί πως σε κάθε αρχιτεκτονική που περιγράφηκε, πυρήνας της επιτυχίας των ΤΝΔ είναι η διαδικασία της μάθησης. Οι αλγόριθμοι των ΤΝΔ έχουν την ικανότητα να αναπροσαρμόζουν τις παραμέτρους του δικτύου προκειμένου να βελτιώσουν τις προβλέψεις τους. Το βασικό χαρακτηριστικό από το οποίο εξαρτάται η αποτελεσματικότητα αλλά και ο χρόνος αναπροσαρμογής και μάθησης του δικτύου είναι οι συναρτήσεις ενεργοποίησης. Στο επόμενο κεφάλαιο επισημαίνεται ο σημαντικός ρόλος που διαδραματίζουν. Παρουσιάζονται και αναλύονται οι γνωστότερες συναρτήσεις και ακολουθεί εκτενής σχολιασμός της σχετικής διαθέσιμης βιβλιογραφίας.

Κεφάλαιο 3

Συναρτήσεις Ενεργοποίησης

3.1 Γενικά για τις Συναρτήσεις

Τα σήματα τα οποία έχουν εισέλθει εντός του νευρώνα, αρχικά αθροίζονται. Στη συνέχεια το άθροισμα αυτό περνά από μια συνάρτηση ενεργοποίησης προτού ο νευρώνας πυροδοτήσει το αποτέλεσμα της συνάρτησης στους νευρώνες του επόμενου επιπέδου. Η συνάρτηση ενεργοποίησης αυτή μεταβάλλει σε μεγάλο βαθμό το άθροισμα των σημάτων που εισήλθαν εντός του νευρώνα. Παραδείγματος χάρη ένα άθροισμα $a = 42$ σε μια κλασική βηματική συνάρτηση, μας δίνει $F(a) = 1$ ενώ για $a = -64$ παίρνουμε $F(a) = 0$. Εύκολα γίνεται κατανοητή η κρισιμότητα επιλογής της σωστής συνάρτησης ενεργοποίησης. Αλλά το ζήτημα δεν τελειώνει εδώ. Αν αναλογιστούμε πως ο υπολογισμός της παραγώγου -ως προς τα βάρη- της συνάρτησης απωλειών μας δείχνει ποιες αλλαγές βαρών πρέπει να συντελεστούν στο δίκτυο, τότε γίνεται σαφές ότι είναι απαραίτητο η συνάρτηση ενεργοποίησης να είναι παραγωγίσιμη. Επίσης, τεράστια σημασία στον χρόνο εκπαίδευσης έχει η δυσκολία υπολογισμού της παραγώγου της συνάρτησης. Το πλήθος των βαρών αυξάνεται εκθετικά όσο προσθέτουμε επίπεδα στο δίκτυο και συνεπώς ο απαιτούμενος χρόνος για τον αλγόριθμο οπισθοδρόμησης εκτοξεύεται. Μια παράγωγος που υπολογίζεται γρήγορα συντελεί σε μικρότερο χρόνο εκπαίδευσης σε αντίθεση με μια παράγωγο που είναι χρονικά πιο απαιτητική. Παρακάτω παρουσιάζεται το βασικό πρόβλημα που προκύπτει κατά την εκπαίδευση ενός ΤΝΔ. Το πρόβλημα των εξαφανιζόμενων κλίσεων (Vanishing Gradient Problem) που εμφανίζεται εντονότερα στα δίκτυα με βαθιά δομή. Στη συνέχεια εξετάζονται οι γνωστότερες συναρτήσεις ενεργοποίησης. Αναλύονται τα χαρακτηριστικά τους, οι συνήθεις περιπτώσεις χρήσης τους και τυχόν

αδυναμίες σε κάθε μια.

3.2 Πρόβλημα Εξαφανιζόμενων Κλίσεων

Το πρόβλημα των εξαφανιζόμενων κλίσεων αποτελεί το βασικό πρόβλημα πολλών συναρτήσεων ενεργοποίησης, έχοντας εντονότερη επίδραση στην εκπαίδευση δικτύων με βαθιά αρχιτεκτονική. Αναδείχθηκε από τον Sepp Hochreiter [8].

Κατά τη διαδικασία της εκπαίδευσης, ένα ΤΝΔ αναπροσαρμόζει τα βάρη του σύμφωνα με τη μέθοδο της βαθμωτής κατάβασης. Κατά την οπισθοδρόμηση η διόρθωση των βαρών ξεκινάει από το τελευταίο επίπεδο κινούμενη προς τα πίσω μέχρι το πρώτο. Στο πρόβλημα των εξαφανιζόμενων κλίσεων η μεταβολή των βαρών ολοένα και μειώνεται καθώς οδεύουμε προς τα πρώτα επίπεδα του δικτύου. Αυτό συμβαίνει καθώς ο υπολογισμός της παραγώγου της συνάρτησης απωλειών ως προς κάθε βάρος πραγματοποιείται σύμφωνα με τον κανόνα αλυσιδωτής παραγωγής. Έτσι αν η παράγωγος μιας συνάρτησης ενεργοποίησης είναι μεταξύ $(0, 1)$ τότε το μέγεθος της μεταβολής του βάρους μειώνεται. Δίνοντας ένα παράδειγμα, στο Σχήμα 3.1 βλέπουμε την παράγωγο της συνάρτησης απωλειών ως προς ένα τυχαίο βάρος w_5 , το οποίο βρίσκεται στο τρίτο επίπεδο από το τέλος.

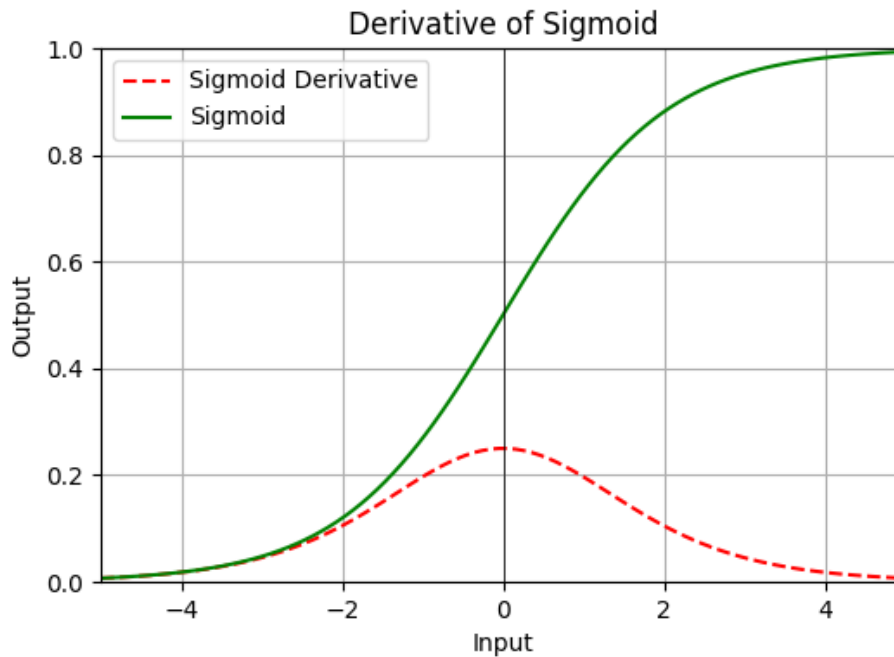
Σχήμα 3.1: Παράδειγμα Αλυσιδωτής Παραγωγής για το w_5

$$\frac{\partial L}{\partial W_5} = \frac{\partial L}{\partial y_4} \times \frac{\partial y_4}{\partial o_4} \times \boxed{\frac{\partial o_4}{\partial y_3}} \times \frac{\partial y_3}{\partial o_3} \times \boxed{\frac{\partial o_3}{\partial W_5}}$$

Κυκλωμένες είναι οι παράγωγοι των συναρτήσεων ενεργοποίησης που χρησιμοποιήθηκαν μεταξύ των δύο τελευταίων επιπέδων. Το πλήθος των παραγώγων που επηρεάζουν το αποτέλεσμα είναι αντιστρόφως ανάλογο του βάθους στο οποίο βρίσκεται το βάρος που εξετάζουμε. Έτσι ένα βάρος που βρίσκεται ένα επίπεδο νωρίτερα από το w_5 θα έχει ακόμα μια παράγωγο συνάρτησης ενεργοποίησης που επηρεάζει τη μεταβολή του. Στο Σχήμα 3.2 μπορούμε να δούμε τη γραφική παράσταση της παραγώγου της σιγμοειδούς συνάρτησης, μιας συνάρτησης ενεργοποίησης που επηρεάζεται σε μεγάλο βαθμό από το πρόβλημα των εξαφανιζόμενων κλίσεων.

Βλέπουμε πως η παράγωγος της σιγμοειδούς συνάρτησης έχει μέγιστη τιμή 0.25. Έτσι, ακόμα και αν οι παράγωγοι που κυκλώσαμε είχαν και στις δύο περιπτώσεις

Σχήμα 3.2: Η Σιγμοειδής Συνάρτηση και η Παράγωγος της



αυτή τη μέγιστη τιμή, η μεταβολή του w_5 θα ήταν εξαιρετικά μικρή. Κατά συνέπεια τα βάρη στα πρώτα επίπεδα καταλήγουν να μην ανταποκρίνονται στην εκπαίδευση.

3.3 Παρουσίαση Γνωστών Συναρτήσεων Ενεργοποίησης

3.3.1 Γραμμική - Linear

Η γραμμική συνάρτηση ενεργοποίησης -ή ταυτότητα- είναι από τις απλούστερες συναρτήσεις. Η χρήση της είναι σπάνια και συναντάται κυρίως στο επίπεδο εξόδου σε προβλήματα παλινδρόμησης. Περιγράφεται από τον εξής τύπο

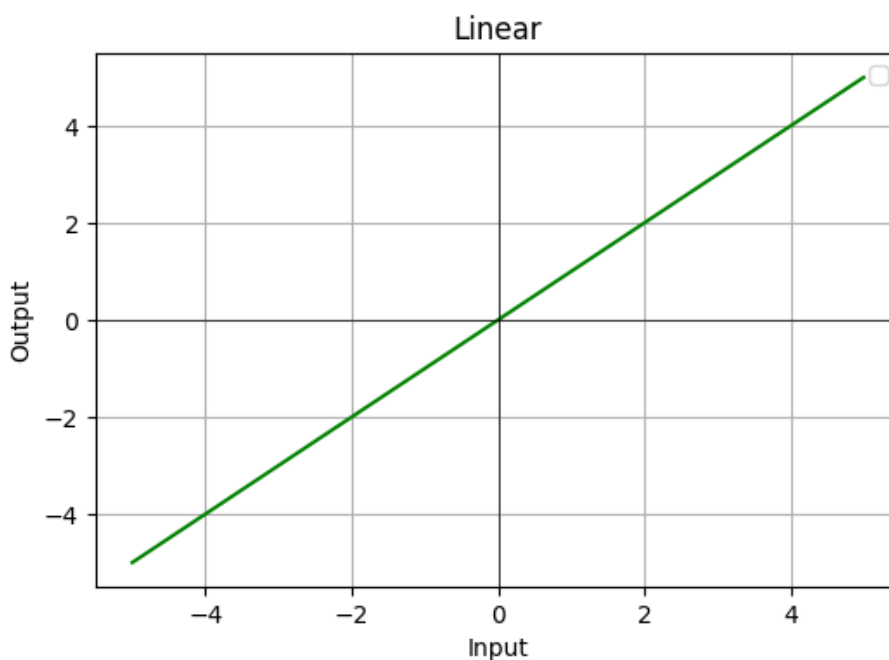
$$f(x) = ax + b$$

$$f'(x) = a$$

και η γραφική της παράσταση φαίνεται στο Σχήμα 3.3.

Όπως είναι εμφανές δεν επιφέρει αλλαγές στην είσοδο. Στα θετικά της καταλογίζεται ότι έχει απλή παράγωγο, ενώ το βασικό αρνητικό της γνώρισμα είναι θα δυσκολέψει το δίκτυο στο να αναπαραστήσει μη γραμμικές σχέσεις στα δεδομένα.

Σχήμα 3.3: Η Γραφική Παράσταση της Γραμμικής



3.3.2 Βηματική - Step

Η βηματική συνάρτηση - η οποία έχει ήδη αναφερθεί - δίνει έξοδο τιμές 0 ή 1 και χρησιμοποιείται πολύ σπάνια στα νευρωνικά δίκτυα πολλαπλών επιπέδων. Η γραφική της παράσταση φαίνεται στο Σχήμα 3.4. Διαχρονικά χρησιμοποιούταν για δυαδική ταξινόμηση αλλά λόγω του ότι δεν είναι παραγωγίσιμη δεν χρησιμοποιείται πλέον.

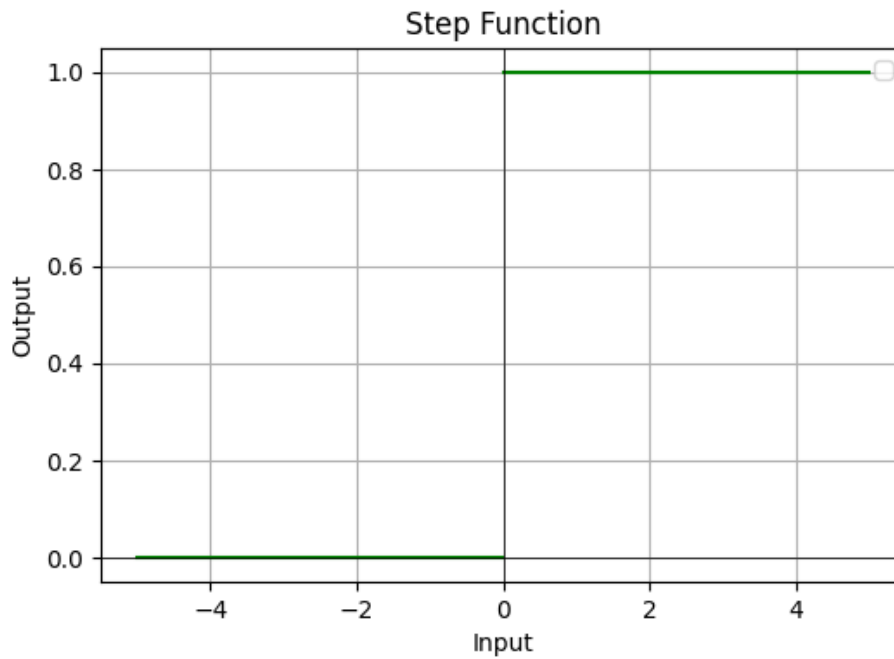
3.3.3 Ανορθωμένη Γραμμική Μονάδα - ReLU

Η ReLU είναι μια τμηματικά γραμμική συνάρτηση ενεργοποίησης. Θεωρείται η πιο διαδεδομένη και χρησιμοποιείται σε όλες τις αρχιτεκτονικές ενώ είναι ιδιαίτερος αποτελεσματική και στα βαθιά νευρωνικά δίκτυα. Παρακάτω βλέπουμε τον τύπο και τη παράγωγο της ενώ στο Σχήμα 3.5 φαίνεται η γραφική της παράσταση.

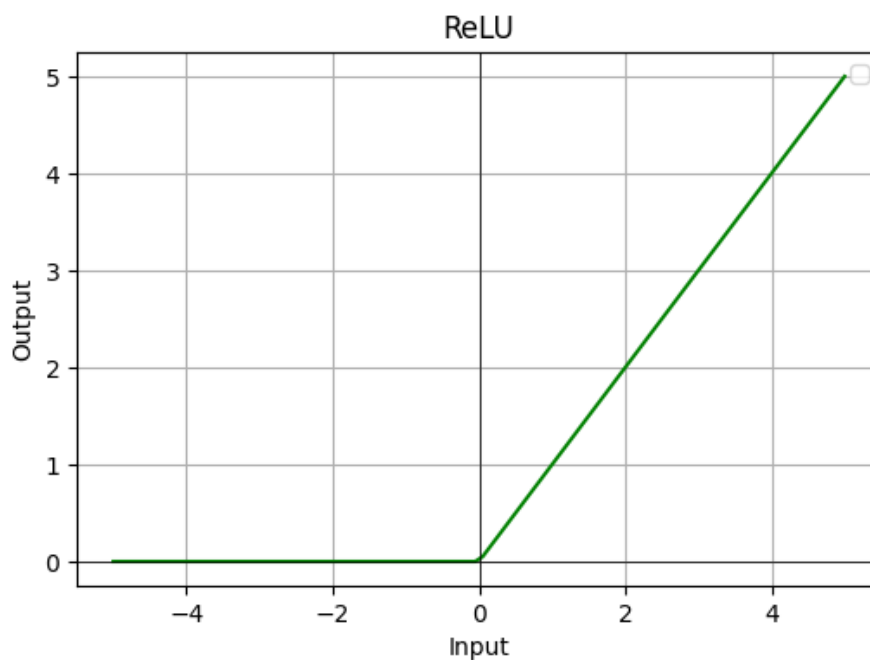
$$ReLU(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$ReLU'(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Σχήμα 3.4: Η Γραφική Παράσταση Βηματικής



Σχήμα 3.5: Η Γραφική Παράσταση της ReLU



Η απλή μορφή της παραγώγου της, επιτρέπει τη γρήγορη εκπαίδευση του δικτύου. Επίσης φαίνεται πως αντιμετωπίζει καλύτερα το πρόβλημα των εξαφανιζόμενων κλίσεων σε σχέση με άλλες συναρτήσεις, με αποτέλεσμα να επιταχύνεται ταχύτερα η σύγκλιση (convergence). Το βασικό της πρόβλημα είναι γνωστό ως dying ReLU. Στο φαινόμενο αυτό κάποιοι νευρώνες νεκρώνουν με αποτέλεσμα να μη συμ-

μετέχουν πλέον στο δίκτυο. Αξίζει να σημειωθεί πως κάτι παρόμοιο συμβαίνει και στα βιολογικά νευρωνικά δίκτυα. Αυτό δεν έχει απαραίτητα μόνο αρνητική χροιά καθώς νεκρώνουν οι νευρώνες που είναι ανενεργοί με αποτέλεσμα να μην σπαταλώνται πόροι και χρόνος. Έτσι θα λέγαμε πως ο εγκέφαλος χάνει σε εύρος και κερδίζει σε αποδοτικότητα. Κλείνοντας σχετικά με τη ReLU να αναφέρουμε πως ένα ακόμα πρόβλημα που παρουσιάζει είναι ότι στις αρνητικές εισόδους δεν είναι παραγωγίσιμη και δίνει σταθερά 0 ως έξοδο. Κάτι τέτοιο μπορεί να δημιουργήσει και άλλα προβλήματα -εκτός του dying ReLU- καθώς έτσι δεν αποτυπώνονται οι αρνητικές συσχετίσεις που υπάρχουν σε κάποια σετ δεδομένων.

3.3.4 Παραλλαγές ReLU

Έχουν γίνει αρκετές προσπάθειες υλοποίησης παραλλαγών της ReLU προκειμένου να αντιμετωπιστούν τα προβλήματα που συναντά. Εδώ θα παρουσιάσουμε δύο εκ των διασημότερων παραλλαγών, την Διαρρέουσα ReLU (Leaky ReLU - LReLU) και την Παραμετρική ReLU (Parametric ReLU - PReLU).

Διαρρέουσα ReLU: Στη διαρρέουσα ReLU, της οποίας τον τύπο και τη παράγωγο βλέπουμε

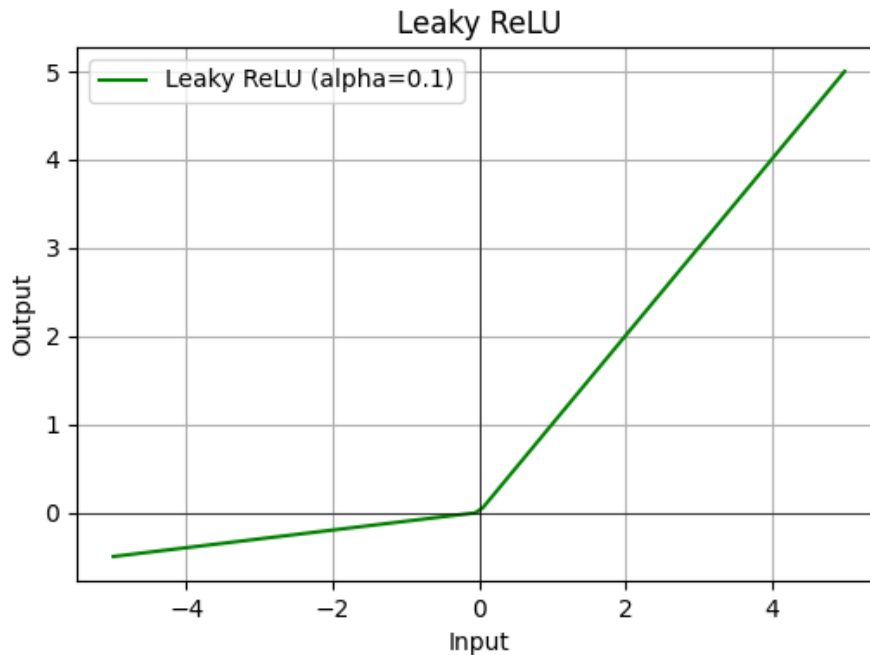
$$LReLU(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha \cdot x, & \text{otherwise} \end{cases}$$

$$LReLU'(x) = \begin{cases} 1, & \text{if } x > 0 \\ \alpha, & \text{otherwise} \end{cases}$$

ενώ η γραφική της παράσταση φαίνεται στο 3.6, αντί να παίρνουμε μηδέν για εισόδους μικρότερες του μηδενός, πολλαπλασιάζουμε τις αρνητικές εισόδους με ένα παράγοντα α , τον οποίο έχουμε ορίσει εκ των προτέρων και ο οποίος συνήθως είναι μικρότερος του ένα. Αυτή η πρόταση έγινε με σκοπό να αντιμετωπιστεί το φαινόμενο Dying ReLU. Η παράγωγος της είναι απλή και εύκολα υπολογίσιμη, κάτι το οποίο δεν αποτρέπει τη χρήση της. Χρησιμοποιείται κυρίως στα βαθιά νευρωνικά δίκτυα, σε περιπτώσεις όπου το φαινόμενο Dying ReLU εμφανίζεται και είναι σημαντικό να αποφευχθεί, όπως την περίπτωση που υπάρχουν αρνητικές τιμές σε ένα σύνολο δεδομένων.

Ωστόσο πρέπει να αναφέρουμε πως η ρύθμιση της παραμέτρου a πρέπει να γίνει με προσοχή καθώς η λανθασμένη επιλογή της μπορεί να φέρει μη επιθυμητά αποτελέσματα, π.χ. για $a = 1$ η LReLU μετατρέπεται σε ταυτότητα.

Σχήμα 3.6: Η Γραφική Παράσταση της Leaky ReLU



Παραμετρική ReLU: Η παραμετρική ReLU μοιάζει με τη διαρρέουσα ReLU όπως βλέπουμε από τον τύπο και τη παράγωγο της.

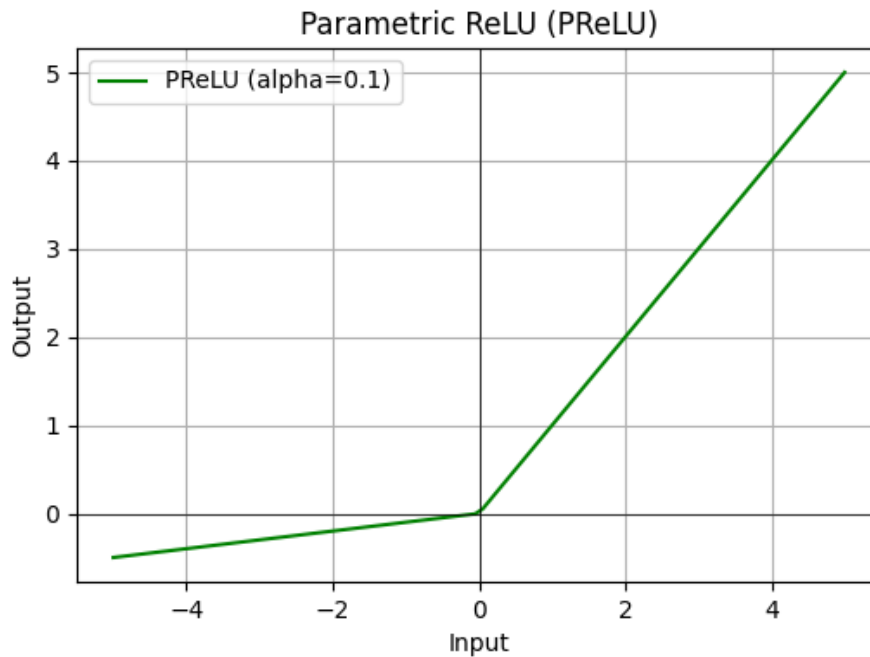
$$PReLU(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha \cdot x, & \text{otherwise} \end{cases}$$

$$PReLU'(x) = \begin{cases} 1, & \text{if } x > 0 \\ \alpha, & \text{otherwise} \end{cases}$$

Η γραφική της παράσταση φαίνεται στο Σχήμα 3.7. Η διαφορά μεταξύ τους εντοπίζεται στην παράμετρο a . Ενώ στη διαρρέουσα ReLU η τιμή της a αρχικοποιείται και μένει σταθερή καθ' όλη τη διάρκεια της εκπαίδευσης, στην παραμετρική ReLU η a είναι δυναμική καθώς προσαρμόζεται στα δεδομένα κατά τη διάρκεια της εκπαίδευσης.

Είναι λιγότερο δημοφιλής από τις δύο άλλες περιπτώσεις ReLU που παρουσιάσαμε. Χρησιμοποιείται εκεί όπου η διαρρέουσα ReLU δεν ανταπεξέρχεται στο πρό-

Σχήμα 3.7: Η Γραφική Παράσταση της Parametric ReLU



βλημα καθώς απαιτείται προσαρμοστικότητα της a , όπως την περίπτωση που διαφορετικά επίπεδα νευρώνων χρειάζονται διαφορετική τιμή της παραμέτρου. Ωστόσο η προσαρμοστικότητα της a δεν έρχεται χωρίς κόστος καθώς προσθέτει πολυπλοκότητα κατά την εκπαίδευση του δικτύου κάτι που επιβραδύνει τη διαδικασία.

3.3.5 Sigmoid

Η σιγμοειδής είναι μια μη γραμμική συνάρτηση ενεργοποίησης που χρησιμοποιούνται ευρέως κατά το παρελθόν. Ακολουθεί ο τύπος και η παράγωγο της

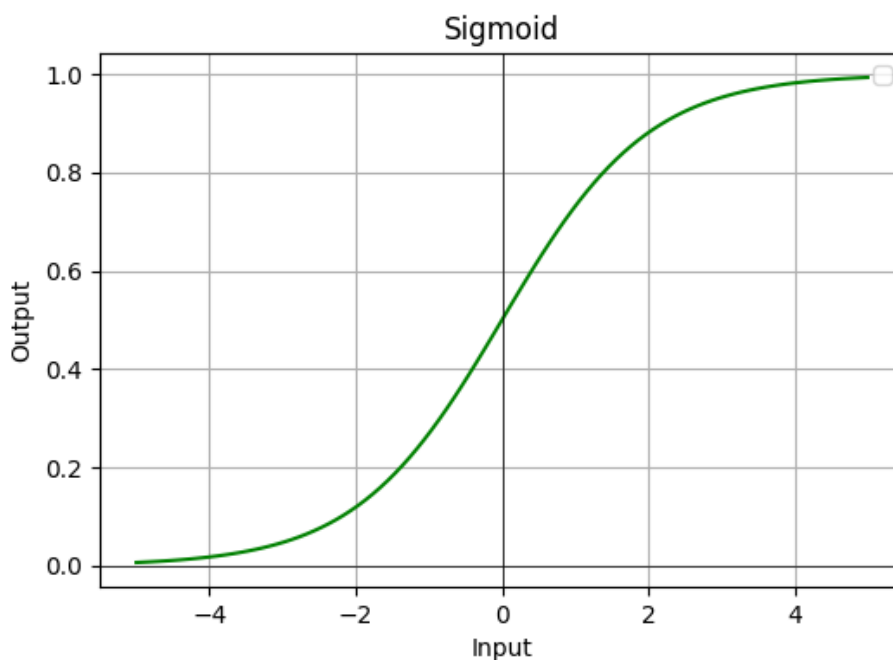
$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\sigma'(x) = \sigma(x) \cdot (1 - \sigma(x))$$

ενώ η γραφική της παράσταση φαίνεται στο Σχήμα 3.8.

Πλέον τη συναντάμε κυρίως στο επίπεδο εξόδου προβλημάτων δυαδικής ταξινόμησης με την έξοδο της να ερμηνεύεται ως πιθανότητα στο διάστημα $(0,1)$. Λόγω της εκθετικής της μορφής ο υπολογισμός της παραγώγου της είναι χρονικά πιο απαιτητικός από γραμμικές και τμηματικά γραμμικές συναρτήσεις, ωστόσο παραμένει χαμηλός. Λόγω του εύρους των εξόδων που δίνει, βρίσκεται αντιμέτωπη με

Σχήμα 3.8: Η Γραφική Παράσταση Sigmoid



το βασικό της μειονέκτημα, το πρόβλημα των εξαφανιζόμενων κλίσεων. Κάτι που την κάνει να μην προσαρμόζεται επαρκώς στα δεδομένα της εκπαίδευσης, με αποτέλεσμα η χρήση της τελευταία να έχει περιοριστεί σημαντικά καθώς υπάρχουν καλύτερες εναλλακτικές.

3.3.6 Tanh και Penalized Tanh

Η υπερβολική εφαρμοπτομένη είναι μια μη γραμμική συνάρτηση ενεργοποίησης η οποία δίνει τιμές στο διάστημα $(-1,1)$. Ακολουθεί ο τύπος και η παράγωγος της

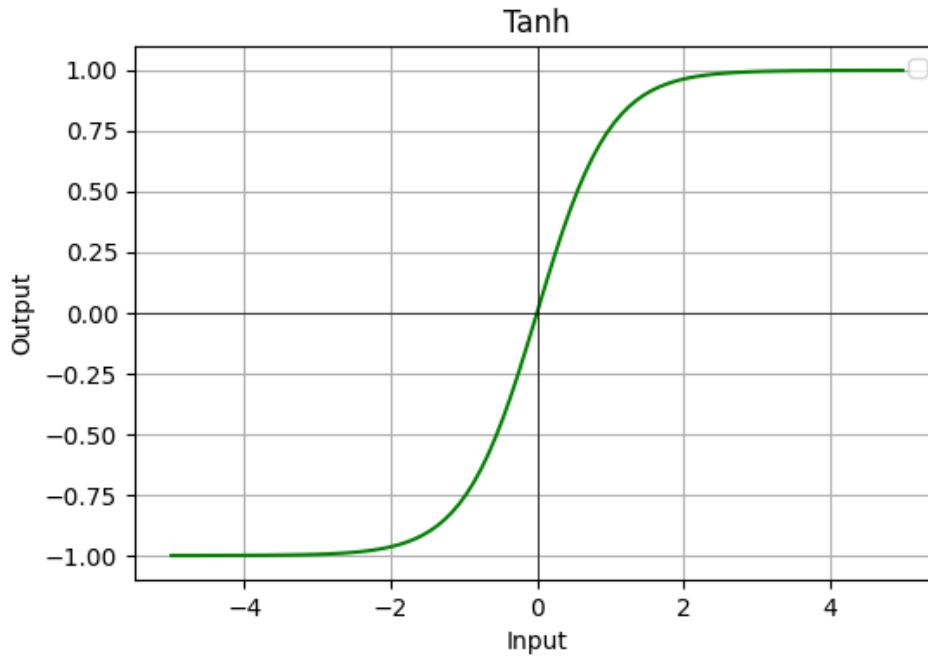
$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\tanh'(x) = 1 - \tanh^2(x)$$

ενω η γραφική της παράσταση αποτυπώνεται στο Σχήμα 3.9.

Παρά τη φθίνουσα πορεία της δημοφιλίας της, παραμένει η επικρατούσα επιλογή σε μερικές περιπτώσεις όπως για παράδειγμα στην αρχιτεκτονική των αναδρομικών νευρωνικών δικτύων, αλλά και περιπτώσεις όπου προτιμάται η έξοδος να έχει κεντρική τιμή το μηδέν. Βέβαια -όπως και η Sigmoid- παρουσιάζει το πρόβλημα των εξαφανιζόμενων κλίσεων σε εισόδους με πολύ μεγάλη ή πολύ μικρή τιμή. Τέλος,

Σχήμα 3.9: Η Γραφική Παράσταση της Tanh



καθώς περιέχει εκθετικά, η παράγωγος της είναι δυσκολότερα υπολογίσιμη από άλλες συναρτήσεις όπως η γραμμική.

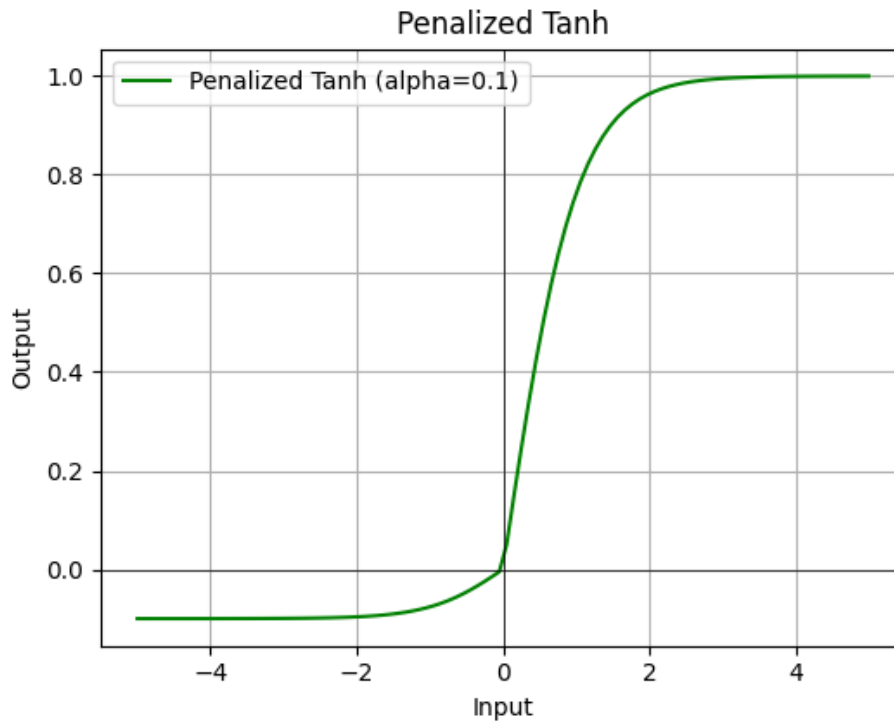
Penalized Tanh: Εδώ αξίζει να αναφερθεί η Penalized Tanh, μια παραλλαγή της Tanh που πρότειναν οι Xu et al. [9]. Στην παραλλαγή αυτή, για τις αρνητικές τιμές εισόδου πολλαπλασιάζουμε την Tanh με έναν σταθερό παράγοντα, όπως συμβαίνει μεταξύ των περιπτώσεων ReLU και Leaky ReLU. Ακολουθεί ο τύπος και η παράγωγος της:

$$PTanh(x) = \begin{cases} \tanh(x), & \text{if } x > 0 \\ a \cdot \tanh(x), & \text{if } x \leq 0 \end{cases}$$
$$PTanh'(x) = \begin{cases} (1 - \tanh^2(x)), & \text{if } x > 0 \\ a \cdot (1 - \tanh^2(x)), & \text{if } x \leq 0 \end{cases}$$

Στο Σχήμα 3.10 φαίνεται η γραφική της παράσταση.

Η Penalized Tanh, παρά τα θετικά και πολύ υποσχόμενα αποτελέσματα που έχει δείξει [10], η αναγνωρισιμότητά της παραμένει σε χαμηλά επίπεδα. Θα αναλυθεί εκτενέστερα στον σχολιασμό της βιβλιογραφίας αλλά και στη πειραματική διαδικασία.

Σχήμα 3.10: Η Γραφική Παράσταση της Penalized Tanh

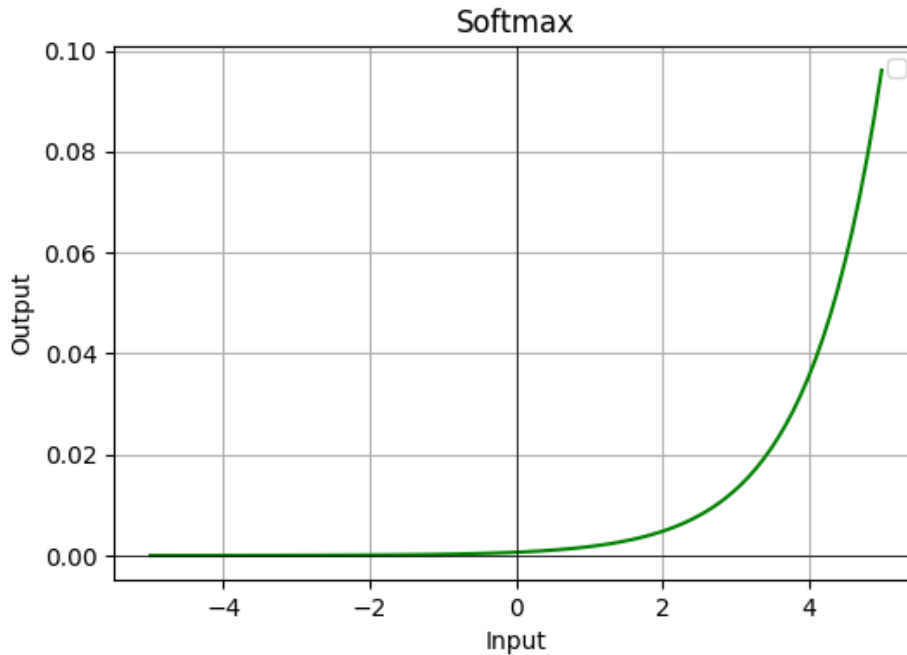


3.3.7 Softmax

Η Softmax είναι μια συνάρτηση ενεργοποίησης που χρησιμοποιείται στο τελευταίο επίπεδο ενός ΤΝΔ, για προβλήματα ταξινόμησης πολλών κλάσεων. Η γραφική της παράσταση φαίνεται στο Σχήμα 3.11.

Ένα πρόβλημα στο οποίο είναι ιδανική η χρήση της Softmax είναι η ταξινόμηση εικόνας. Σε ένα παράδειγμα ενός τέτοιου προβλήματος η είσοδος στο δίκτυο μπορεί να είναι μια εικόνα στην οποία αποτυπώνεται χειρόγραφα ένα ψηφίο, ενώ ως έξοδο παίρνουμε την πιθανότητα η εικόνα να ανήκει σε κάθε μια εκ των 10 πιθανών κλάσεων (10 ψηφία 0-9). Σε αυτό το παράδειγμα η Softmax θα μετατρέψει τις εισόδους στους 10 νευρώνες του τελευταίου επιπέδου σε πιθανότητες οι οποίες θα αθροίζονται στο 1. Άλλα προβλήματα στα οποία η Softmax είναι αποτελεσματική είναι η εξόρυξη γνώμης (Sentiment Analysis) και η γλωσσική μοντελοποίηση (Language Modeling). Ο υπολογισμός της παραγώγου της είναι υπολογιστικά απαιτητικός, ωστόσο το βασικό πρόβλημα που αντιμετωπίζει είναι η αριθμητική αστάθεια που προκαλείται λόγω ακραίων τιμών. Οι τιμές αυτές οδηγούν σε υπερχείλιση (Overflow) καθώς η ακρίβεια αναπαράστασης αριθμών στους Η/Υ είναι περιορισμένη. Τεχνι-

Σχήμα 3.11: Η Γραφική Παράσταση της Softmax



κές όπως η κανονικοποίηση(normalization) χρησιμοποιούνται για την αντιμετώπιση αυτού του προβλήματος.

3.3.8 Swish

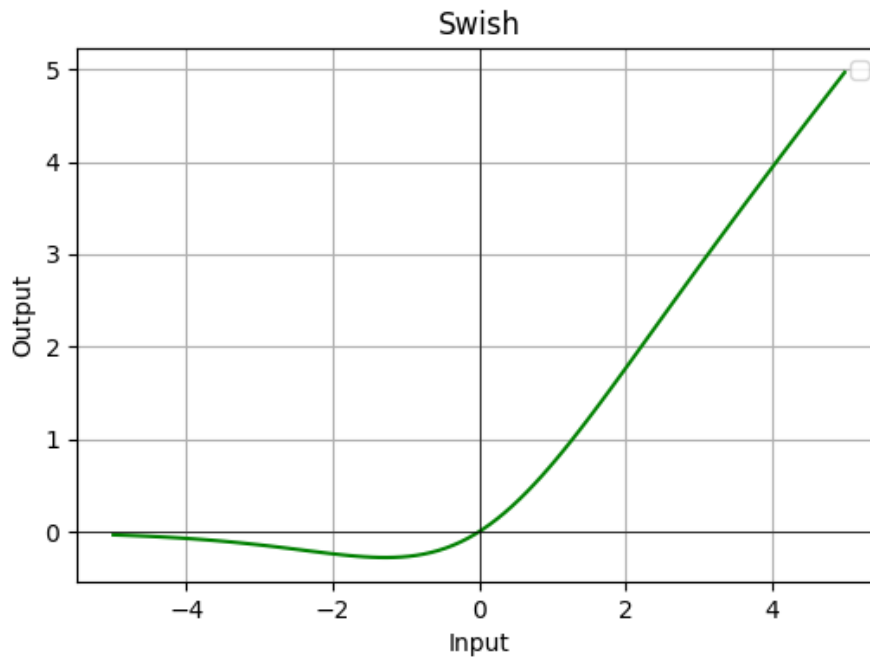
Η Swish είναι μια αρκετά νέα συνάρτηση ενεργοποίησης καθώς προτάθηκε μόλις το 2017 από ερευνητές της Google. Παρακάτω φαίνεται ο τύπος της και η γραφική της παράσταση στο Σχήμα 3.12.

$$Swish(x) = x \cdot \sigma(x)$$

$$Swish'(x) = \sigma(x) + x \cdot \sigma'(x)$$

Είναι μη γραμμική και δεν είναι μονότονη. Τα ιδιαίτερα χαρακτηριστικά της προσέλκυσαν την προσοχή των ερευνητών ωστόσο προς το παρόν δεν χαίρει μεγάλης δημοφιλίας. Το βασικό αρνητικό γνώρισμα της, είναι ότι η παράγωγος της είναι πολύπλοκη κάτι που την καθιστά υπολογιστικά απαιτητική, με αποτέλεσμα να επιβραδύνει την εκπαίδευση του δικτύου. Ωστόσο γίνονται πειραματισμοί με τη Swish σε δίκτυα με βαθιά δομή όπου απαιτούνται μη γραμμικές συναρτήσεις για να εντοπιστούν πολύπλοκα μοτίβα στα δεδομένα. Παρακάτω παρουσιάζονται

Σχήμα 3.12: Η Γραφική Παράσταση της Swish



δημοσιεύσεις στις οποίες έδειξε εξαιρετικά αποτελέσματα.

3.3.9 Softplus

Η Softplus είναι μια μη γραμμική συνάρτηση ενεργοποίησης η οποία παράγει μόνο θετικές τιμές.

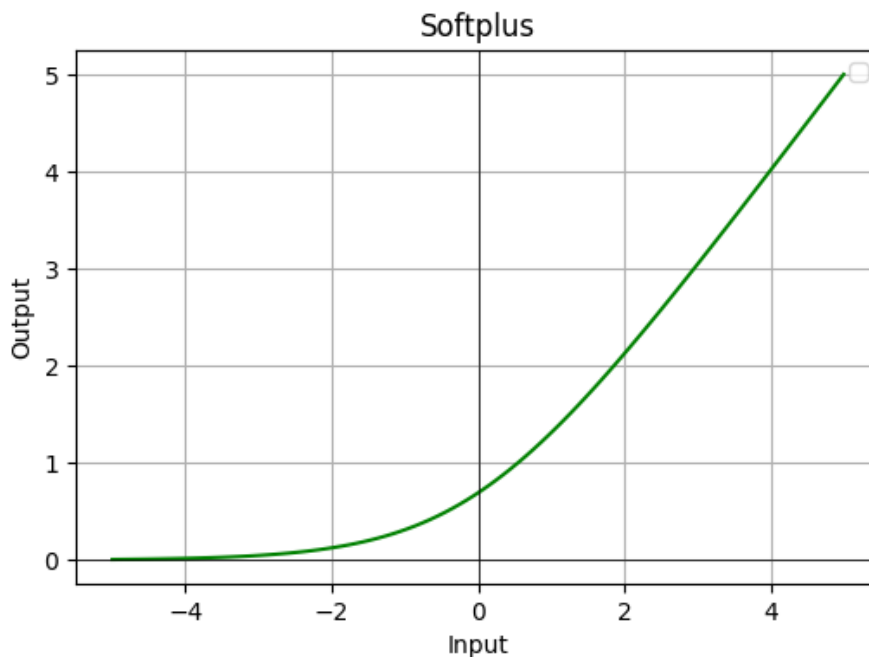
$$SPlus(x) = \ln(1 + e^x)$$

$$SPlus'(x) = \sigma(x)$$

Μοιάζει με την ReLU όπως μπορούμε να διακρίνουμε από τη γραφική της παράσταση (Σχήμα 3.13) ωστόσο στις εισόδους κοντά στο μηδέν προσφέρει ομαλότερη μετάβαση.

Υπάρχουν πειράματα όπου δείχνει πολύ θετικές επιδόσεις, στα οποία μας επιτρέπεται να τη λογίζουμε ως μια αξιόπιστη επιλογή. Από τη άλλη μεριά υπάρχουν περιπτώσεις που τα αποτελέσματα της δεν είναι τόσο θετικά. Οι Dubey et al. [3] έδειξαν και τις δύο κατηγορίες προβλημάτων. Στη αναγνώριση εικόνας η Softplus φάνηκε να καταφέρνει παρόμοιες και σε πολλές περιπτώσεις καλύτερες επιδόσεις από τη ReLU - τόσο σε ακρίβεια όσο και σε χρόνο - ενώ στη μετάφραση κειμένου δε τα πήγε εξίσου καλά.

Σχήμα 3.13: Η Γραφική Παράσταση της Softplus



3.4 Βιβλιογραφική Ανασκόπηση

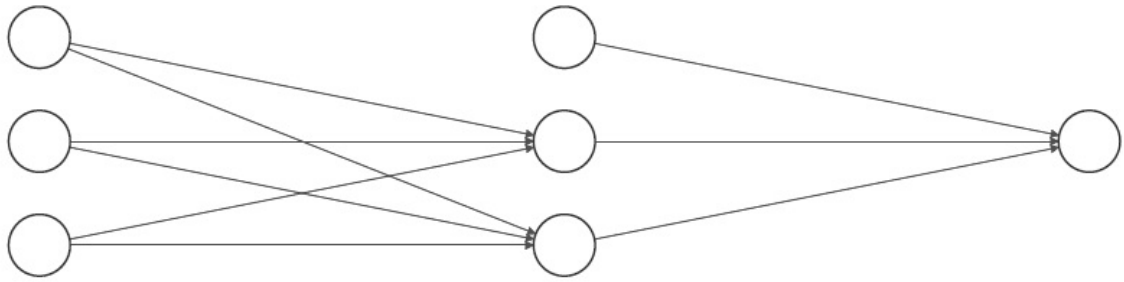
Αρκετές έρευνες εστιάζουν στις συναρτήσεις ενεργοποίησης και στην αποτελεσματικότητά τους κατά την εκπαίδευση ενός ΤΝΔ. Σε αυτήν την ενότητα γίνεται ανασκόπηση και σχολιασμός σχετικών δημοσιεύσεων.

3.4.1 Συγκριτική Ανάλυση των Συναρτήσεων Ενεργοποίησης στα Τεχνητά Νευρωνικά Δίκτυα [1]

Οι Kamalov et al. [1] μελετούν τρεις συναρτήσεις (Linear, ReLU, Sigmoid) πάνω σε ένα ΤΝΔ τριών επιπέδων (Layers), εισόδου, κρυφό, εξόδου. Στα δύο πρώτα επίπεδα υπάρχουν δύο κόμβοι - και η πόλωση (bias) - ενώ στην έξοδο υπάρχει ένας κόμβος. Το σύνολο δεδομένων που χρησιμοποιήθηκε είναι το Housing Data και είναι διαθέσιμο εδώ: <https://www.kaggle.com/chenzhiliang/housing-data>. Ως είσοδο έχουμε την Έκταση (FloorArea) και τον Αριθμό Δωματίων (Bedrooms) και ως έξοδο παίρνουμε την Αξία του Σπιτιού (Value of the house). Στο Σχήμα 3.14 φαίνεται το δίκτυο που περιγράψαμε παραπάνω.

Η διάρκεια εκπαίδευσης ήταν 30 εποχές και η αξιολόγηση της απόδοσης έγινε με γνώμονα το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error). Οι περιπτώσεις που

Σχήμα 3.14: Το TNA της [1]



μελετήθηκαν και συγκρίθηκαν ήταν:

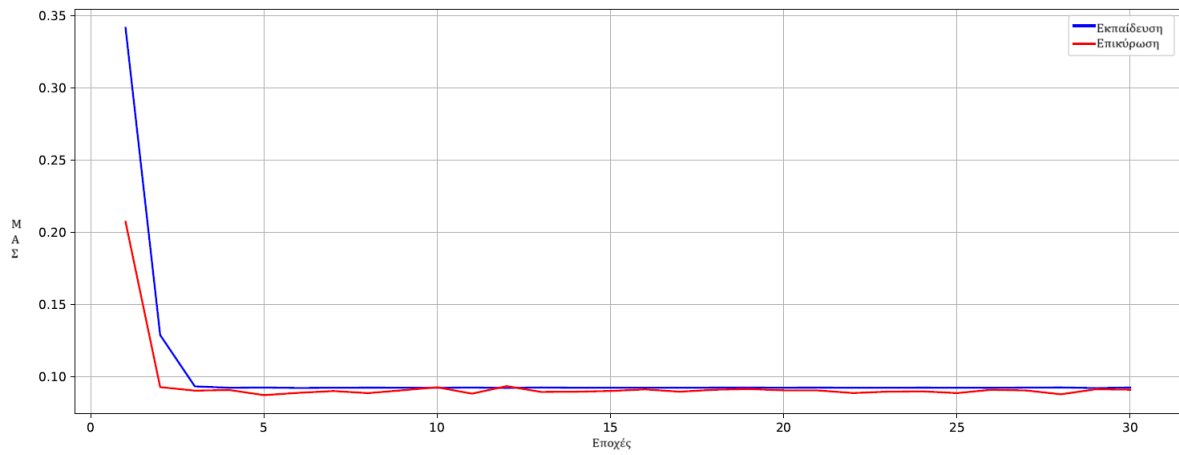
- Linear-Linear (σε κρυφό επίπεδο και επίπεδο εξόδου αντίστοιχα)
- Linear-ReLU
- ReLU-Linear
- ReLU-ReLU
- Sigmoid-Linear
- Sigmoid-ReLU

Στην περίπτωση Linear-Linear (Σχήμα 3.15), στα σετ Δεδομένων Εκπαίδευσης (Training Set) και Επικύρωσης (Validation set), το Μέσο Απόλυτο Σφάλμα (ΜΑΣ) μειώνεται απότομα στις τρεις πρώτες εποχές, απ' όπου και έπειτα σταθεροποιείται μέχρι το τέλος της εκπαίδευσης με ΜΑΣ 0.09137. Όμοια αποτελέσματα φαίνονται και στην περίπτωση Linear-ReLU (Σχήμα 3.16).

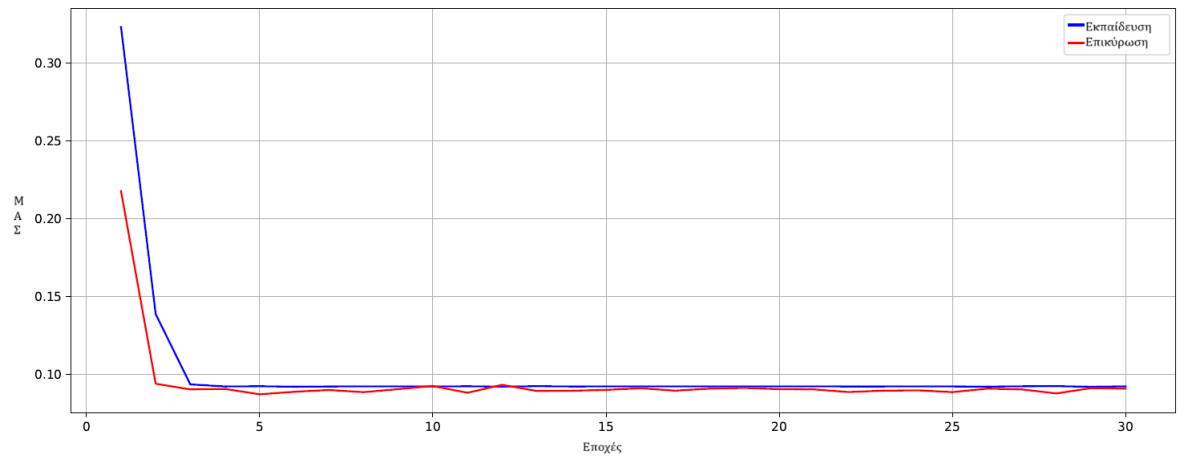
Στην περίπτωση ReLU-Linear (Σχήμα 3.17) φαίνεται και πάλι πως οι δύο καμπύλες - Εκπαίδευσης και Επικύρωσης - ταυτίζονται μεταξύ τους. Όπως και στις δύο πρώτες περιπτώσεις παρατηρείται απότομη μείωση του ΜΑΣ στις τρεις πρώτες εποχές και σχετική σταθεροποίηση από εκεί και έπειτα με τελική τιμή 0.08767. Από τα Σχήματα 3.17 και 3.18 θα φανεί πως οι περιπτώσεις ReLU-Linear και ReLU-ReLU είναι πανομοιότυπες.

Τελευταίες μελετήθηκαν οι περιπτώσεις Sigmoid-Linear και Sigmoid-ReLU. Και εδώ, όπως φαίνεται από το Σχήμα 3.19 και το Σχήμα 3.20, οι δύο αυτές περιπτώσεις είναι μεταξύ τους πανομοιότυπες. Ωστόσο υπάρχουν διαφορές στις γραφικές

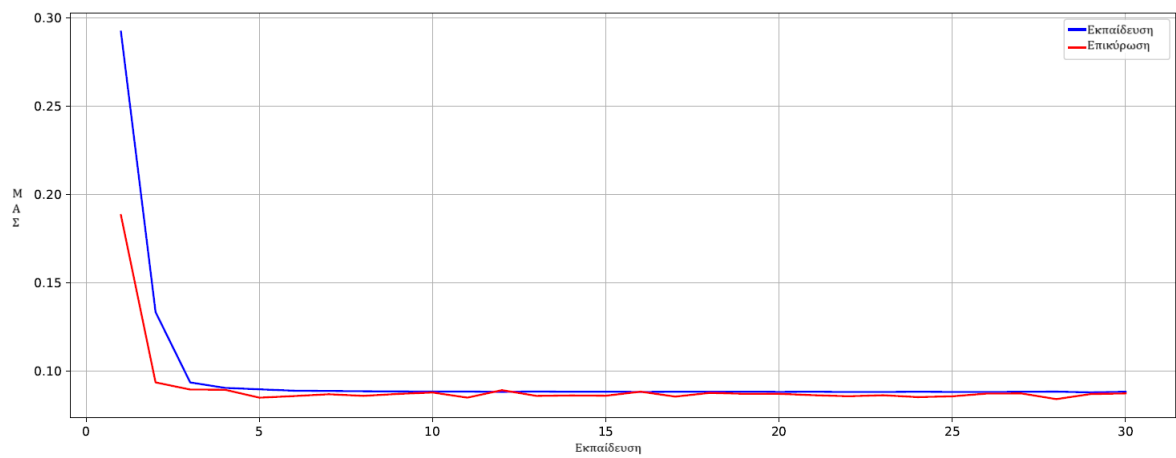
Σχήμα 3.15: Linear-Linear [1]



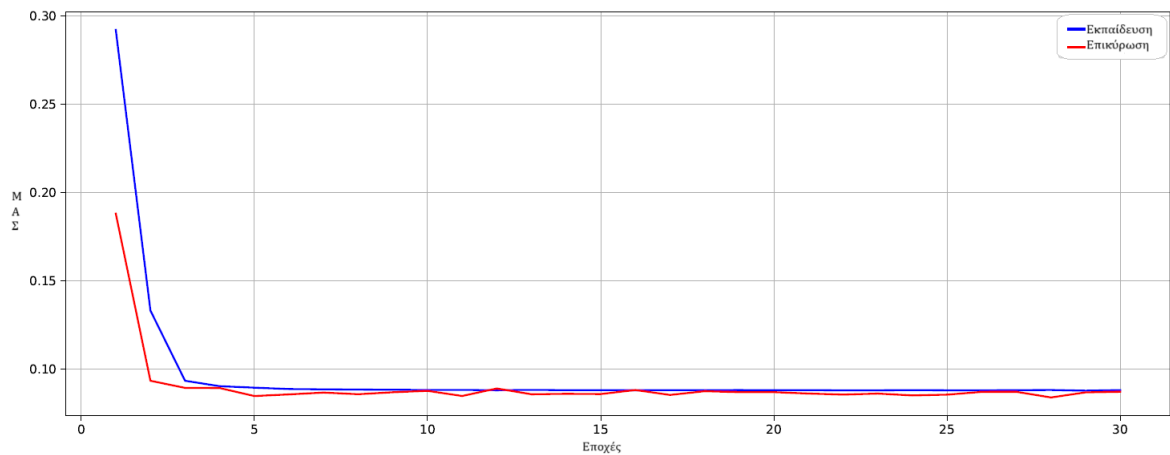
Σχήμα 3.16: Linear-ReLU [1]



Σχήμα 3.17: ReLU-Linear [1]

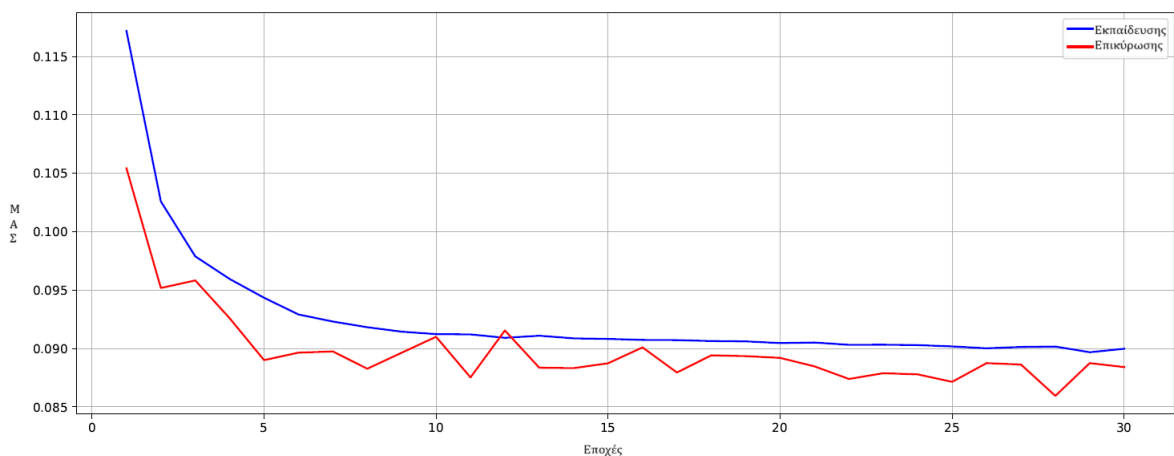


Σχήμα 3.18: ReLU-ReLU [1]



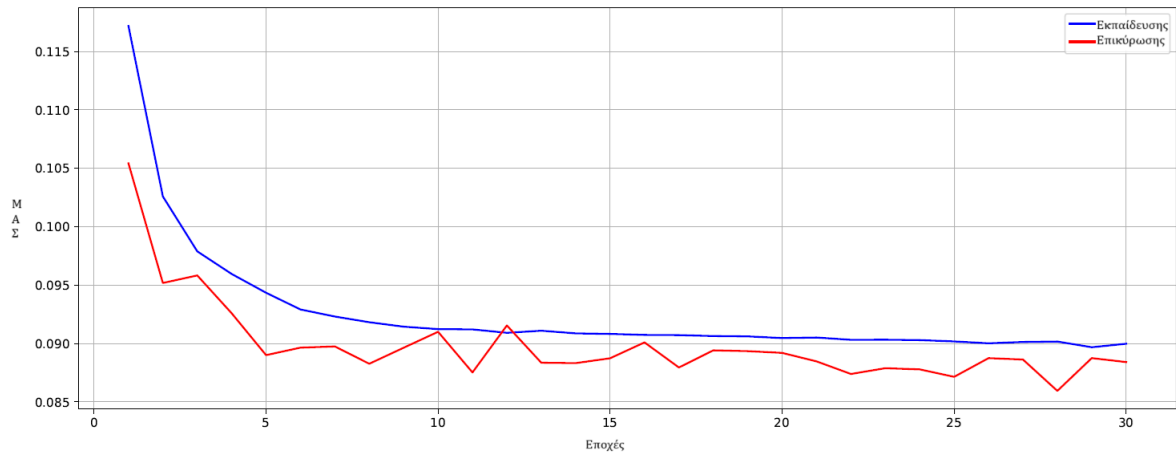
παραστάσεις σε σχέση με όλες τις παραπάνω μετρήσεις. Η Καμπύλη Εκπαίδευσης φαίνεται να φθίνει με μειούμενο ρυθμό καθ' όλη τη διάρκεια εκπαίδευσης με τελικό ΜΑΣ 0.08885 ενώ η καμπύλη επικύρωσης ενώ φθίνει, έχει συνεχείς διακυμάνσεις.

Σχήμα 3.19: Sigmoid-Linear [1]



Σε όλες τις περιπτώσεις, η εναλλαγή μεταξύ ReLU και Linear στο επίπεδο εξόδου δεν έδειξε καμία διαφορά. Αξίζει να σημειωθεί ότι το σύνολο δεδομένων έχει μόνο θετικές τιμές, ως εκ τούτου η παραπάνω παρατήρηση ήταν αναμενόμενη. Οι διαφορές ΜΑΣ σε όλες τις μετρήσεις ήταν μικρές μεταξύ τους, κάτι το οποίο επίσης δεν προκαλεί έκπληξη δεδομένου του μεγέθους του ΤΝΔ. Παρόλα αυτά φάνηκε πως η ReLU δίνει τη χαμηλότερη τιμή ΜΑΣ. Η διαφορά αναμένεται μεγαλύτερη σε ένα μεγαλύτερο ΤΝΔ. Τέλος, αξίζουν προσοχή τα ευρήματα σχετικά με την Sigmoid.

Σχήμα 3.20: Sigmoid-ReLU [1]



Όταν χρησιμοποιήθηκε στο κρυφό επίπεδο - παρά την τελικώς καλή απόδοση μετά από 30 εποχές - φάνηκε πως η Καμπύλη Εκπαίδευσης φθίνει καθ' όλη τη διάρκεια της, σε αντίθεση με τις άλλες περιπτώσεις που πέτυχαν την καλύτερη τους απόδοση από τις τρεις πρώτες εποχές. Αυτό δείχνει πως η Sigmoid χρειάζεται περισσότερη εκπαίδευση για να επιτύχει τη σύγκλιση. Κλείνοντας, πρέπει να γίνει αναφορά στην Καμπύλη Επικύρωσης. Βλέποντας τις διακυμάνσεις στην τιμή του ΜΑΣ προκαλείται προβληματισμός σχετικά με την ικανότητα του ΤΝΔ να γενικεύσει τη συμπεριφορά του και να προβλέψει άγνωστες τιμές.

3.4.2 Ανασκόπηση και Σύγκριση Γνωστών Συναρτήσεων Ενεργοποίησης για Βαθιά Νευρωνικά Δίκτυα [2]

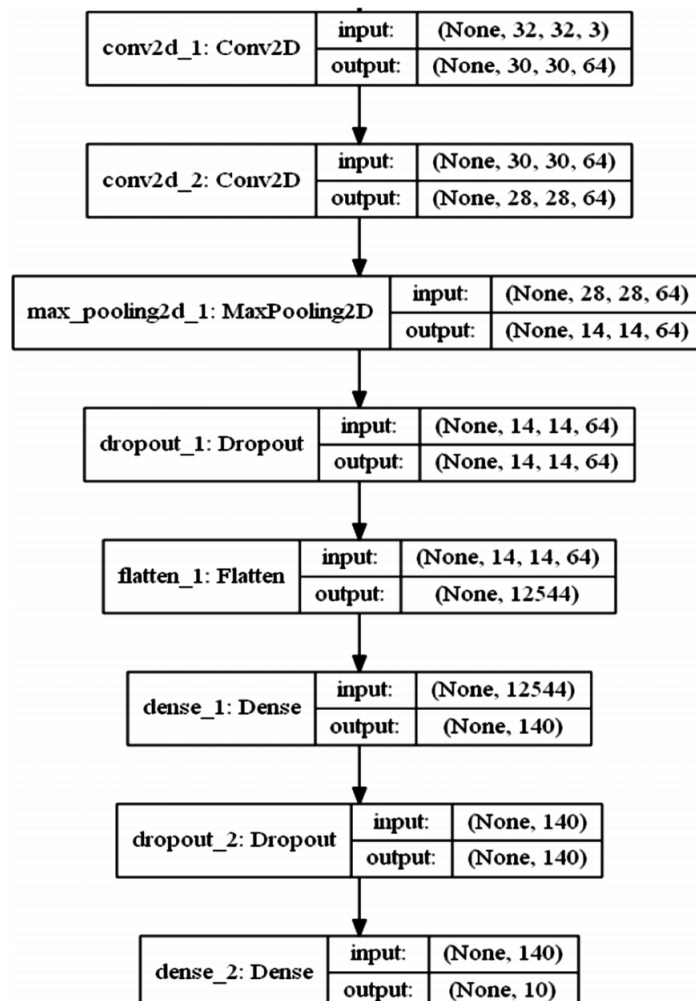
Ο Szandafła [2] συνέκρινε τις παρακάτω συναρτήσεις ως προς την ακρίβεια και αλλά τους χρόνους εκπαίδευσης και ελέγχου (test):

- Sigmoid
- Tanh
- ReLU
- Leaky-ReLU
- Swish

- Softsign
- Softplus

Το σύνολο δεδομένων που χρησιμοποιήθηκε είναι το CIFAR-10. Στόχος του ΣΝΔ της έρευνας είναι να κατατάξει σωστά τα δεδομένα, τα οποία χωρίστηκαν στο σύνολο εκπαίδευσης (50,000 εικόνες) και σύνολο ελέγχου (10,000 εικόνες). Η διάρκεια εκπαίδευσης για κάθε πείραμα ήταν 25 εποχές. Το δίκτυο έχει δύο συνελικτικά επίπεδα και η αρχιτεκτονική του παρουσιάζεται με λεπτομέρεια στο Σχήμα 3.21 Σε κάθε επίπεδο χρησιμοποιήθηκε μόνο η προς μελέτη συνάρτηση ανά πείραμα, εκτός από το επίπεδο εξόδου, όπου σε κάθε περίπτωση χρησιμοποιούνταν η Softmax. Για να ελαττωθεί η επίδραση του παράγοντα της τύχης, έγιναν τρεις μετρήσεις για κάθε μια από τις συναρτήσεις που αναφέρθηκαν. Τα αποτελέσματα που φαίνονται παρακάτω είναι ο μέσος όρος των τριών αυτών μετρήσεων.

Σχήμα 3.21: Το ΣΝΔ της [2]



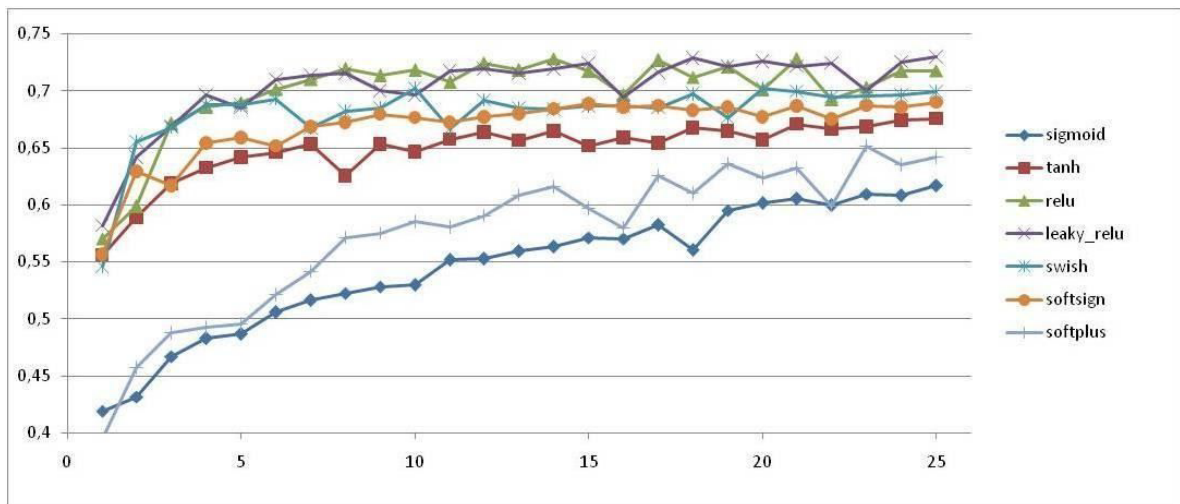
Τα αποτελέσματα ως προς την ακρίβεια κατάταξης:

Πίνακας 3.1: Αποτελέσματα Ακρίβειας Κατάταξης Δεδομένων

Συνάρτηση	Sigmoid	Tanh	ReLU	Leaky-ReLU	Swish	Softsing	Softplus
Ακρίβεια	61.66%	67.57%	71.79%	72.95%	69.89%	69.01%	65.98%

Από τον Πίνακα 3.1 φαίνεται πως η Leaky-ReLU έχει τη μεγαλύτερη ακρίβεια με την ReLU να ακολουθεί με μικρή διαφορά, ενώ τελευταία έρχεται η Sigmoid. Στο Σχήμα 3.22 αποτυπώνονται οι γραφικές παραστάσεις της εκπαίδευσης σε κάθε περίπτωση.

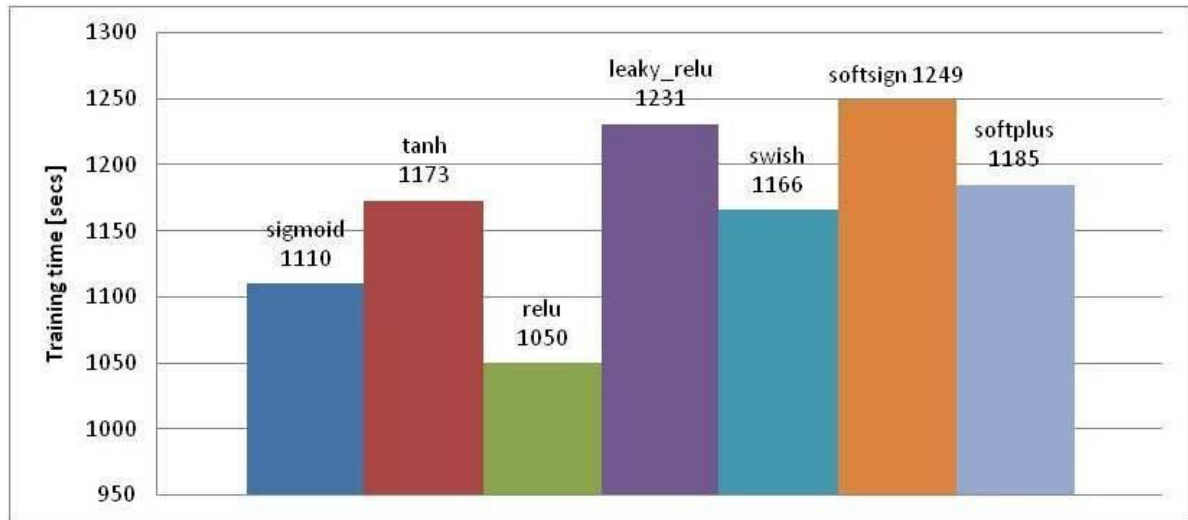
Σχήμα 3.22: Γραφική Παράσταση Ακρίβειας [2]



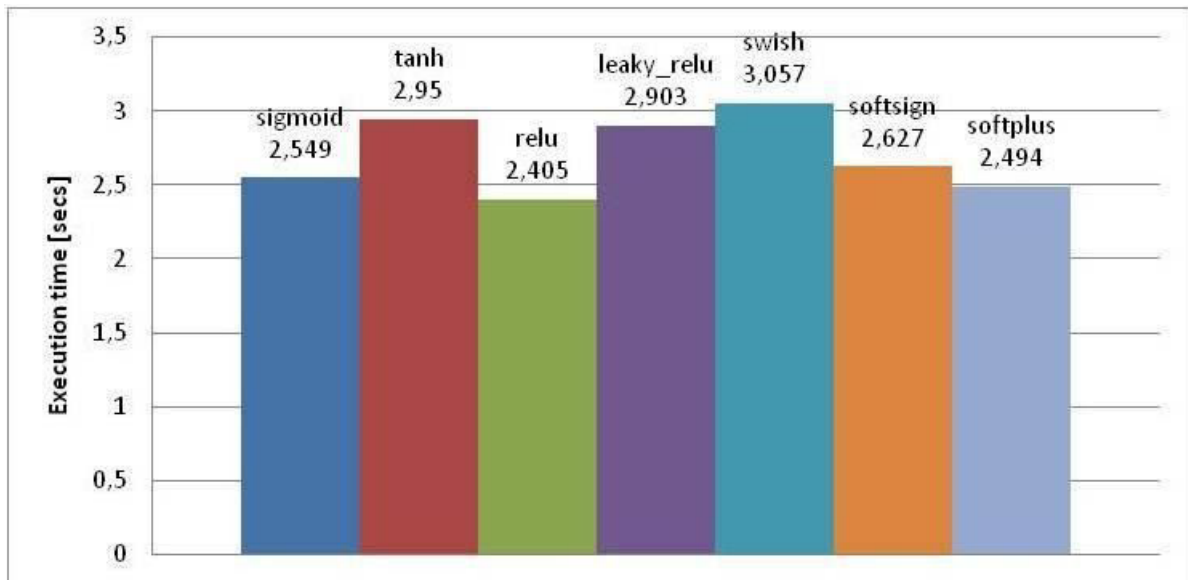
Όσον αφορά τους χρόνους εκπαίδευσης (Σχήμα 3.23) και κατάταξης (Σχήμα 3.24), παρατηρείται πως η ReLU είναι η γρηγορότερη. Ιδιαίτερα σε σχέση με τη Leaky-ReLU - η οποία πέτυχε ελαφρώς καλύτερα αποτελέσματα στην ακρίβεια - υπάρχει διαφορά περίπου τριών λεπτών (+22%) στον χρόνο εκπαίδευσης.

Με γνώμονα το ποσοστό ακρίβειας και την ταχύτητα εκπαίδευσης, τα αποτελέσματα της συγκεκριμένης έρευνας προκρίνουν τη ReLU ως την καλύτερη επιλογή. Παρόλο που η παραλλαγή της, η Leaky-ReLU πέτυχε ελαφρώς καλύτερα αποτελέσματα όσο αφορά την ακρίβεια (+1.16%), η διαφορά τους στις χρονικές απαιτήσεις δε δίνει κίνητρο να ώστε να προτιμηθεί, εκτός των περιπτώσεων όπου ο παράγοντας της ταχύτητας έχει ελάχιστη σημασία. Τέλος, όσον αφορά τις Sigmoid, Tanh, Swish, Softsing και Softplus, οι αποδόσεις τους στην έρευνα δεν προτρέπουν στη χρήση τους καθώς όλες είχαν χαμηλότερο ποσοστό ακρίβειας και χρειάστηκαν περισσότερο χρόνο εκπαίδευσης από τη ReLU.

Σχήμα 3.23: Χρόνος Εκπαίδευσης [2]



Σχήμα 3.24: Χρόνος Κατάταξης στο Σύνολο Ελέγχου [2]



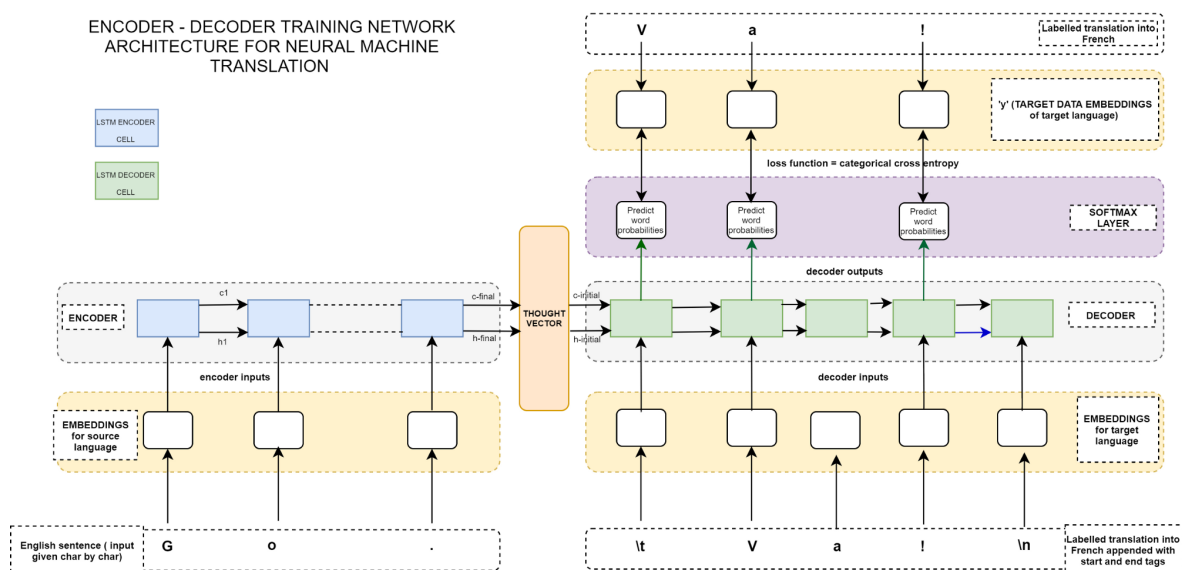
3.4.3 Συναρτήσεις Ενεργοποίησης στη Βαθιά Μάθηση: Περιεκτική Επισκόπηση και Συγκριτική Αξιολόγηση [3]

Οι Dubey et al. [3] συνέκριναν 18 συναρτήσεις ενεργοποίησης σε μια σειρά προβλημάτων. Παρακάτω παρουσιάζονται τα αποτελέσματα εστιάζοντας στις γνωστότερες συναρτήσεις. Το μοντέλο του αναδρομικού νευρωνικού δικτύου (Σχήμα 3.25) και το σύνολο δεδομένων (Multi30k) που χρησιμοποιήθηκαν για τη μετάφραση κειμένου (από Γερμανικά σε Αγγλικά) είναι διαθέσιμα εδώ:

<https://www.kaggle.com/code/parthplc/pytorch-seq2seq-machine-translation>

Οι παράμετροι της εκπαίδευσης:

Σχήμα 3.25: ANΔ της [3]



- Διάρκεια: 50 εποχές
- Ρυθμός μάθησης (Learning Rate): 0.001
- Μέγεθος δεσμίδας (Batch Size): 256
- Εγκατάλειψη (Drouout): 0.5
- Συνάρτηση απωλειών (Loss Function): Διασταυρούμενη Εντροπία (Cross Entropy)
- Βελιστοποιητής (Optimizer): Adam

Στον Πίνακα 3.2 φαίνεται ο μέσος όρος της Αξιολόγησης Δίγλωσσης Αντικατάστασης τεσσάρων λέξεων (Bleu Score with 4-gram) και η τυπική απόκλιση των πέντε μετρήσεων που έγιναν για κάθε συνάρτηση. Σχετικά με την αναγνώριση ομιλίας, το μοντέλο του δικτύου που χρησιμοποιήθηκε περιέχει δύο υπολειμματικά συνελκτικά επίπεδα (Residual Convolution Layers) και δύο επίπεδα αμφίδρομης Ανατροφοδοτούμενης Μονάδας με πύλη (GRU) με μέγεθος διάστασης 512 για το καθένα εκ των δύο επιπέδων GRU. Ως έξοδο δίνει 29 κλάσεις. Η εκπαίδευση έγινε με 100 ώρες ηχογραφημένης ομιλίας στα Αγγλικά, από το σύνολο δεδομένων LibriSpeech. Οι παράμετροι εκπαίδευσης είναι οι εξής:

- Διάρκεια: 10 εποχές
- Ρυθμός μάθησης: 0.0005

- Μέγεθος δεσμίδας: 10
- Εγκατάλειψη: 0.5

Στη μετάφραση κειμένου τα καλύτερα αποτελέσματα τα δίνει η Tanh με Bleu Score 20.93 με την PReLU να ακολουθεί με 20.04, ενώ η Elliot φαίνεται η λιγότερο κατάλληλη επιλογή με score 14.49. Όσον αφορά την αναγνώριση ομιλίας, στο σφάλμα χαρακτήρα αξιολογήθηκαν ως καλύτερες η ReLU με τις δύο παραλλαγές της - LReLU και PReLU - μαζί με τη Swish. Στο σφάλμα λέξης, η Swish μαζί με τη PReLU έδειξαν τα καλύτερα αποτελέσματα, με τις ReLU και LReLU να ακολουθούν με μικρή διαφορά. Στις δύο μετρικές σχετικές με την αναγνώριση ομιλίας, η Sigmoid ήρθε τελευταία.

Πίνακας 3.2: Μετάφραση Κειμένου

Συνάρτηση	Bleu Score
Sigmoid	14.59 ± 0.47
Tanh	20.93 ± 0.91
Elliot	14.49 ± 0.96
ReLU	1.88 ± 0.86
LReLU	18.89 ± 0.82
PReLU	20.04 ± 0.98
Softplus	16.78 ± 0.84
Swish	19.51 ± 0.97

Πίνακας 3.3: Αναγνώριση Ομιλίας

Συνάρτηση	CER	WER
Sigmoid	0.53 ± 0.18	1.19 ± 0.39
Tanh	0.26 ± 0	0.68 ± 0
Elliot	0.40 ± 0.01	0.93 ± 0.01
ReLU	0.24 ± 0.01	0.66 ± 0.01
LReLU	0.24 ± 0	0.66 ± 0.01
PReLU	0.24 ± 0	0.65 ± 0
Softplus	0.30 ± 0.01	0.76 ± 0.02
Swish	0.24 ± 0.01	0.65 ± 0.01

Τα αποτελέσματα αυτής της μελέτης δείχνουν για ακόμα μια φορά πως η ReLU αλλά και οι παραλλαγές της έχουν κορυφαίες επιδόσεις. Παρόλα αυτά φάνηκε πως σε ορισμένα προβλήματα υπάρχουν και επιλογές που τις ξεπερνούν. Στη μετάφραση κειμένου είδαμε πως η Tanh αναδείχθηκε κορυφαία επιλογή ενώ στην αναγνώριση

ομιλίας η Swish έδειξε ίδια και σε μερικές περιπτώσεις καλύτερα αποτελέσματα από την “οικογένεια” των ReLU.

3.4.4 Συμπεράσματα Ευρύτερης Μελέτης

Όπως φάνηκε στις τρεις παραπάνω δημοσιεύσεις των οποίων τα πειράματα αναλύθηκαν, αλλά και από περαιτέρω έρευνα της διαθέσιμης σχετικής βιβλιογραφίας, η ReLU και οι παραλλαγές της κάθε περίπτωση είναι από τις πιο αξιόπιστες επιλογές, αν όχι η πιο αξιόπιστη (όπως δείχνουν οι Abdulwahed Salam et al. [11]). Στη σύγκριση μεταξύ τους φαίνεται πως υπάρχουν μικρές διαφορές στα ποσοστά ακρίβειας, κάτι το οποίο αξίζει περαιτέρω έρευνας όπως συνιστούν οι Xu et al. [12]. Ένας επίσης σημαντικός παράγοντας αξιολόγησης είναι η ταχύτητα, και όπως μας δείχνουν και οι Dasgupta et al. [13] η ReLU φάνηκε ο αδιαμφισβήτητος πρωταθλητής. Τέλος, επισημάνεται πως τα ΤΝΔ καλούνται να λύσουν μεγάλο εύρος προβλημάτων. Έτσι, υπάρχουν διαφορετικές προσεγγίσεις οι οποίες προσαρμόζουν την αρχιτεκτονική των δικτύων για την αντιμετώπιση διαφορετικού τύπου προβλημάτων. Η επιλογή συνάρτησης ενεργοποίησης, ως βασική παράμετρος ενός ΤΝΔ παράγει διαφορετικά αποτελέσματα ανάλογα με το είδος του προβλήματος και την αρχιτεκτονική του δικτύου. Υπάρχουν δημοσιεύσεις όπως των Eger et al. [10] στις οποίες, συναρτήσεις όπως η Tanh ή η Penalized Tanh έδειξαν καλύτερα ή σταθερότερα αποτελέσματα από τη ReLU και τις παραλλαγές της. Συμπεραίνεται λοιπόν πως ενώ η βιβλιογραφία που αναλύθηκε, παρέχει πολύ σημαντικά στοιχεία για τις συναρτήσεις ενεργοποίησης και αποτελεί για τον μελετητή αδιαμφισβήτητη πηγή γνώσης, κάτι τέτοιο δεν έρχεται σε αντίθεση με την ανάγκη για περαιτέρω έρευνα πάνω στο πεδίο.

Κεφάλαιο 4

Υλοποίηση Πειραμάτων και Αποτελέσματα

Στο κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα που προέκυψαν από μια σειρά πειραμάτων. Όλες οι μετρήσεις πραγματοποιήθηκαν στο Kaggle.com με χρήση του GPU P100. Δοκιμάστηκαν οι προς μελέτη συναρτήσεις ενεργοποίησης σε ένα πλήθος συνόλων δεδομένων που αφορούν μερικά απ τα γνωστότερα είδη προβλημάτων που αντιμετωπίζουν τα ΤΝΔ. Τέτοια προβλήματα είναι η αναγνώριση εικόνας, η εξόρυξη γνώμης, προβλήματα παλινδρόμησης, καθώς και χρονοσειρών. Για κάθε πρόβλημα χρησιμοποιήθηκαν πάνω από ένα μοντέλο ΤΝΔ ώστε να εξαχθούν ασφαλέστερα συμπεράσματα.

4.1 Αναγνώριση Εικόνας

Για την αναγνώριση εικόνας χρησιμοποιήθηκαν τα εξής τρία μοντέλα συνελκτικών νευρωνικών δικτύων:

- Το Alexnet [7], ένα ΣΝΔ το οποίο περιέχει πέντε συνελκτικά επίπεδα, δύο κρυφά επίπεδα και ένα επίπεδο εξόδου.
- Το VGG16 [14], ένα ΣΝΔ με 13 συνελκτικά επίπεδα, δύο κρυφά επίπεδα και ένα επίπεδο εξόδου.
- Το CNN3, ένα δίκτυο που δημιουργήθηκε για τις ανάγκες των πειραμάτων. Περιέχει πέντε συνελκτικά επίπεδα, ένα κρυφό και ένα επίπεδο εξόδου.

Σε κάθε πείραμα χρησιμοποιήθηκε σε όλα τα επίπεδα του δικτύου η προς μελέτη συνάρτηση, εκτός από το τελευταίο επίπεδο όπου σε κάθε περίπτωση εφαρμόστηκε η Softmax. Κάθε μέτρηση επαναλήφθηκε τρεις φορές και τα αποτελέσματα που παρουσιάζονται αποτελούν τον μέσο όρο των αρχικών παρατηρήσεων. Μετά τη συλλογή των αποτελεσμάτων πραγματοποιήθηκε μια ακόμα μέτρηση με υβριδικό χαρακτήρα στην οποία εφαρμόστηκαν δύο συναρτήσεις, η μια στα συνελικτικά επίπεδα και η άλλη στα κρυφά. Τα σύνολα δεδομένων που χρησιμοποιήθηκαν είναι το Cifar-10 το οποίο περιέχει 10 κλάσεις, το Cifar-100 με 100 κλάσεις και το vegetable-image-dataset με 15 κλάσεις.

Οι παράμετροι της εκπαίδευσης είναι:

- Διάρκεια: 25 εποχές
- Ρυθμός μάθησης: 0.001
- Μέγεθος δεσμίδας: 32
- Συνάρτηση απωλειών: Κατηγοριακή Διασταυρούμενη Εντροπία (Categorical Cross Entropy)
- Μετρική: Ακρίβεια

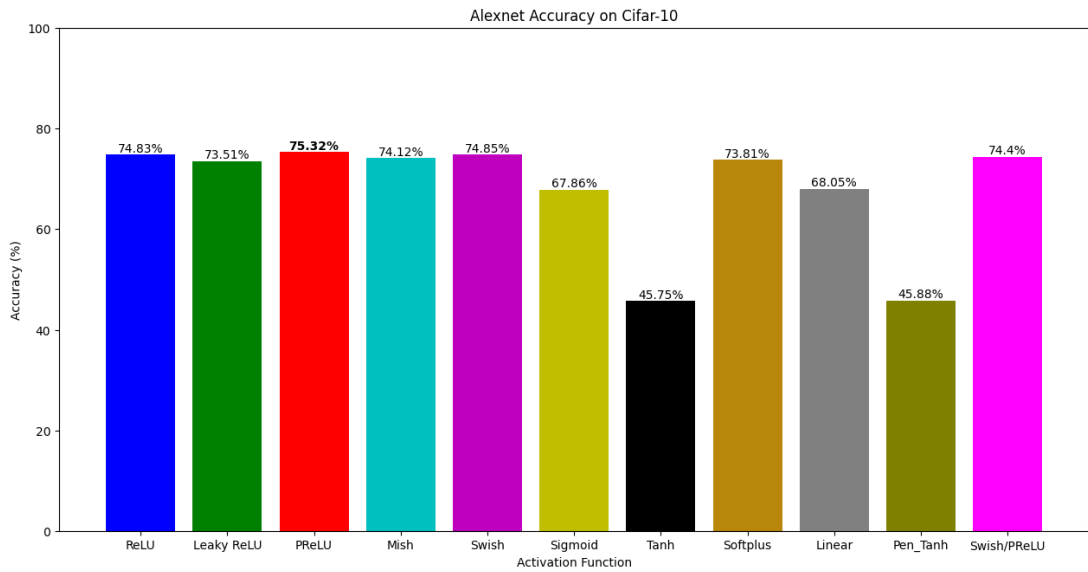
4.1.1 Αποτελέσματα Cifar-10

Στο Σχήμα 4.1 παρουσιάζονται τα ποσοστά σωστής ταξινόμησης που πέτυχε το Alexnet στο Cifar10. Φαίνεται πως η PReLU έρχεται πρώτη με 75.32% με τις Swish και ReLU να ακολουθούν με διαφορά μικρότερη του 1%. Οι Tanh, Penalized Tanh, Sigmoid και Linear έχουν χαμηλές επιδόσεις και η χρήση τους δε συνιστάται.

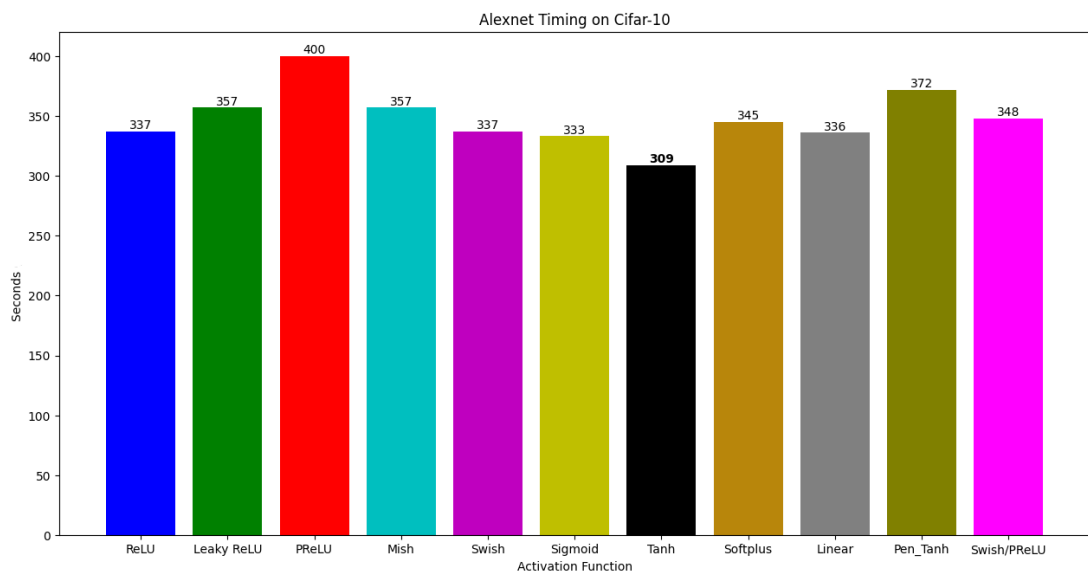
Οι χρόνοι εκπαίδευσης φαίνονται στο Σχήμα 4.2. Εδώ η Tanh έχει την καλύτερη χρονική απόδοση με 309 δευτερόλεπτα. Από τις συναρτήσεις που πέτυχαν υψηλά ποσοστά ακρίβειας, η ReLU και η Swish πέτυχαν γρηγορότερη εκπαίδευση με 337 δευτερόλεπτα.

Οι επιδόσεις ακρίβειας του VGG16 στο Cifar10 φαίνονται στο Σχήμα 4.3, ενώ οι χρόνοι εκπαίδευσης στο Σχήμα 4.4. Η Leaky Relu έρχεται πρώτη σε ακρίβεια με 85.75%, ενώ οι Penalized Tanh, Tanh, Sigmoid και Linear βρίσκονται στο τέλος της κατάταξης με μεγάλες διαφορές.

Σχήμα 4.1: Alexnet-Cifar10 Ακρίβεια

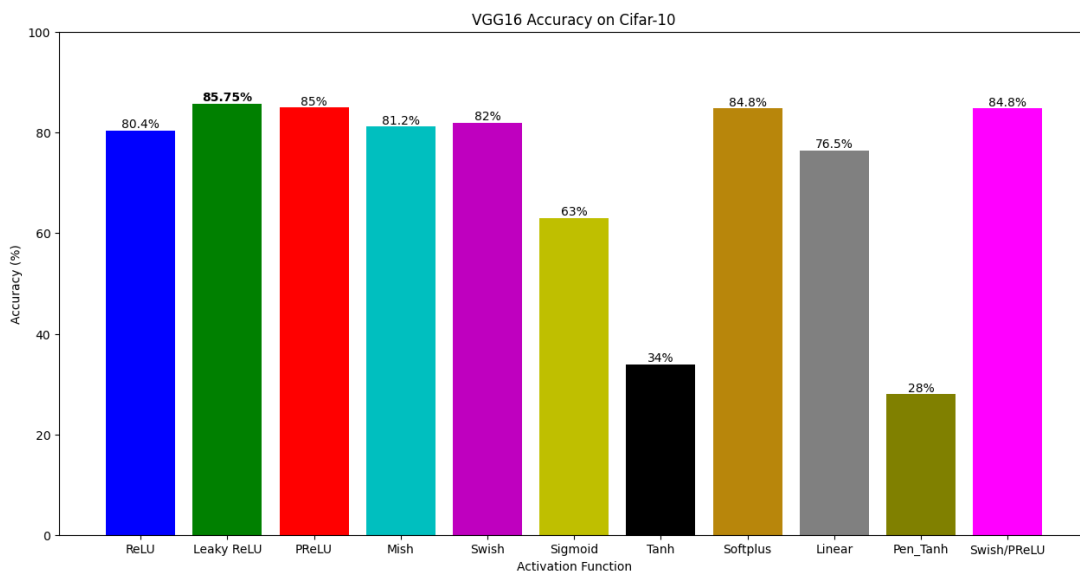


Σχήμα 4.2: Alexnet-Cifar10-Χρονομέτρηση

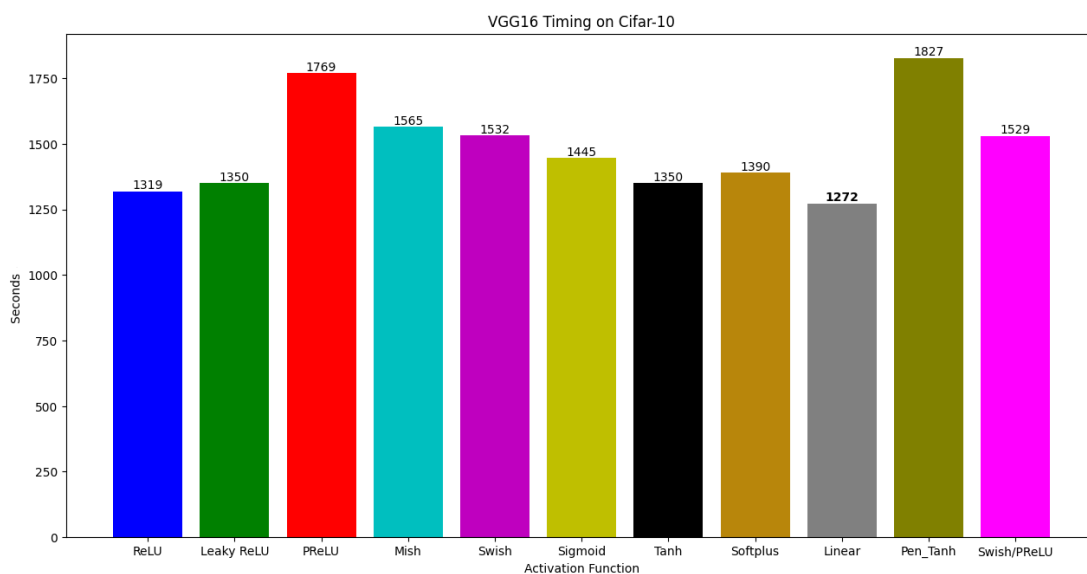


Σχετικά με τους χρόνους εκπαίδευσης, η Linear έρχεται πρώτη με 1,272 δευτερόλεπτα με τη ReLU να ακολουθεί με 1,319, ενώ η Leaky ReLU χρειάστηκε 1,350 δευτερόλεπτα. Τέλος, για CNN3 στο Σχήμα 4.5 φαίνεται πως η Leaky ReLU έρχεται πρώτη σε ποσοστά ακρίβειας με 82%, ενώ η Sigmoid με 15% έδειξε πως η εκπαίδευση απέτυχε να αποδώσει. Σχετικά με τους χρόνους εκπαίδευσης, στο Σχήμα 4.6

Σχήμα 4.3: VGG16-Cifar10 Ακρίβεια

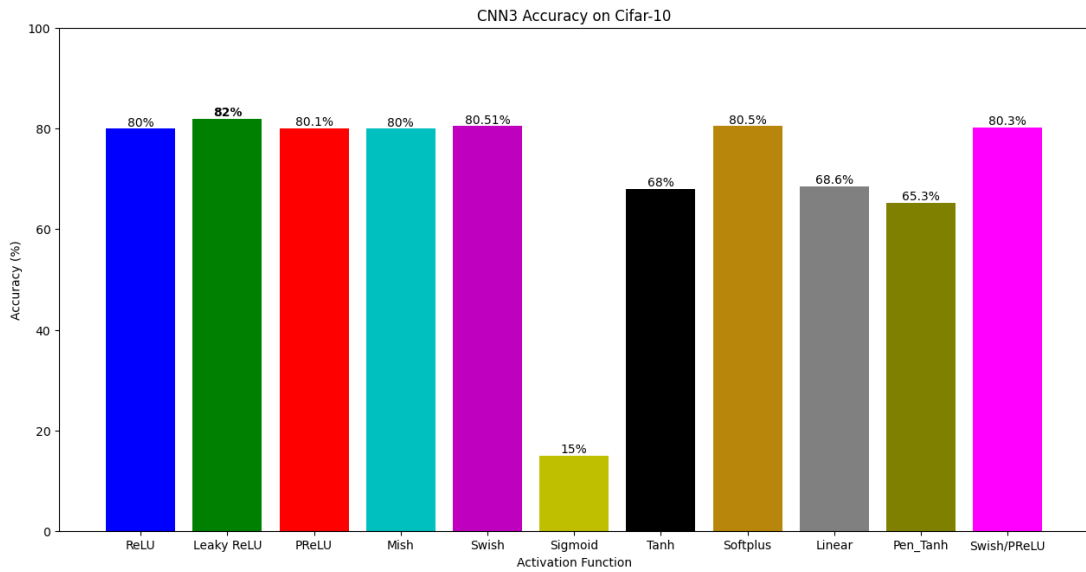


Σχήμα 4.4: VGG16-Cifar10 Χρονομέτρηση

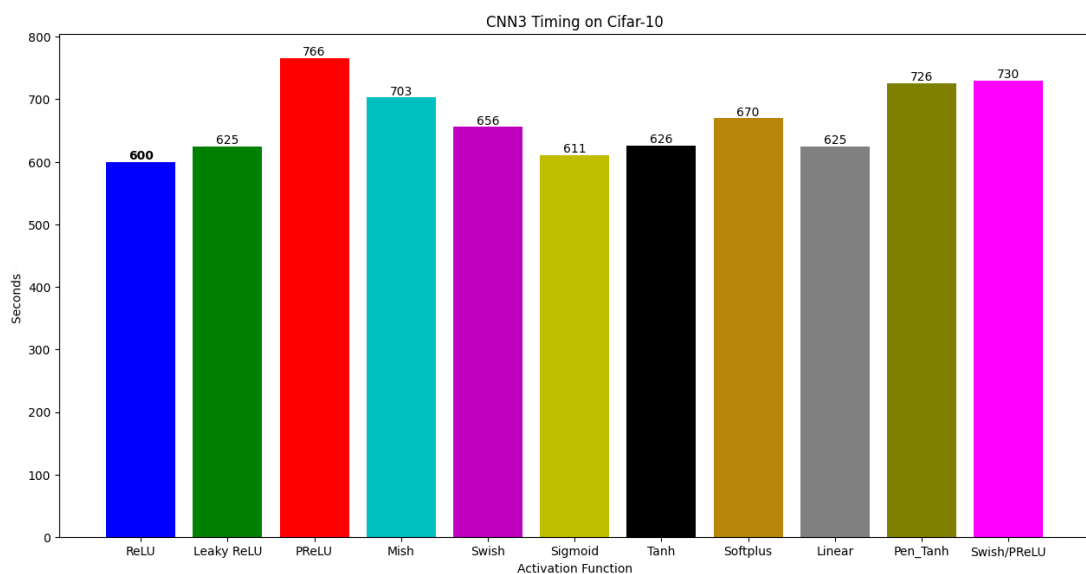


η ReLU ήταν η ταχύτερη με χρόνο εκπαίδευσης 600 δευτερόλεπτα, ενώ η Leaky ReLU χρειάστηκε 625.

Σχήμα 4.5: CNN3-Cifar10 Ακρίβεια



Σχήμα 4.6: CNN3-Cifar10 Χρονομέτρηση

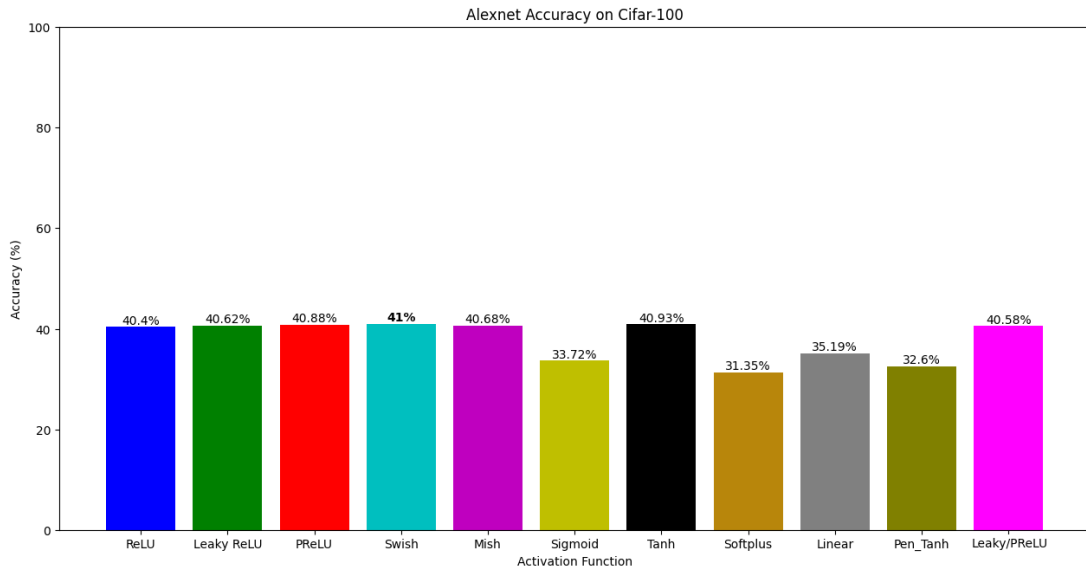


4.1.2 Αποτελέσματα Cifar-100

Οι επιδόσεις του Alexnet στο Cifar-100 αποτυπώνονται στο Σχήμα 4.7 και οι χρόνοι εκπαίδευσης στο Σχήμα 4.8. Η Swish με 41% πετυχαίνει το υψηλότερο ποσοστό ακρίβειας, έχοντας ωστόσο διαφορά μικρότερη του 1% με τις περισσότερες

συναρτήσεις.

Σχήμα 4.7: Alexnet-Cifar100 Ακρίβεια

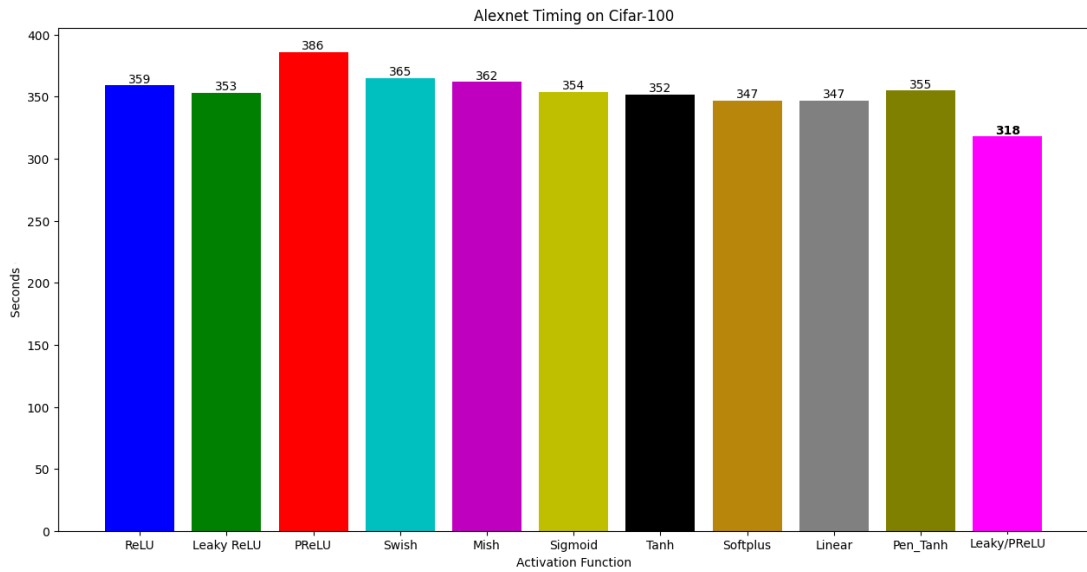


Σχετικά με τους χρόνους εκπαίδευσης ο συνδυασμός Leaky ReLU/PReLU απέδωσε καλύτερα καθώς η εκπαίδευση διήρκεσε 19 δευτερόλεπτα λιγότερο από τις επόμενες σε ταχύτητα Linear και Softplus, καθώς και 37 δευτερόλεπτα λιγότερα από τη Swish.

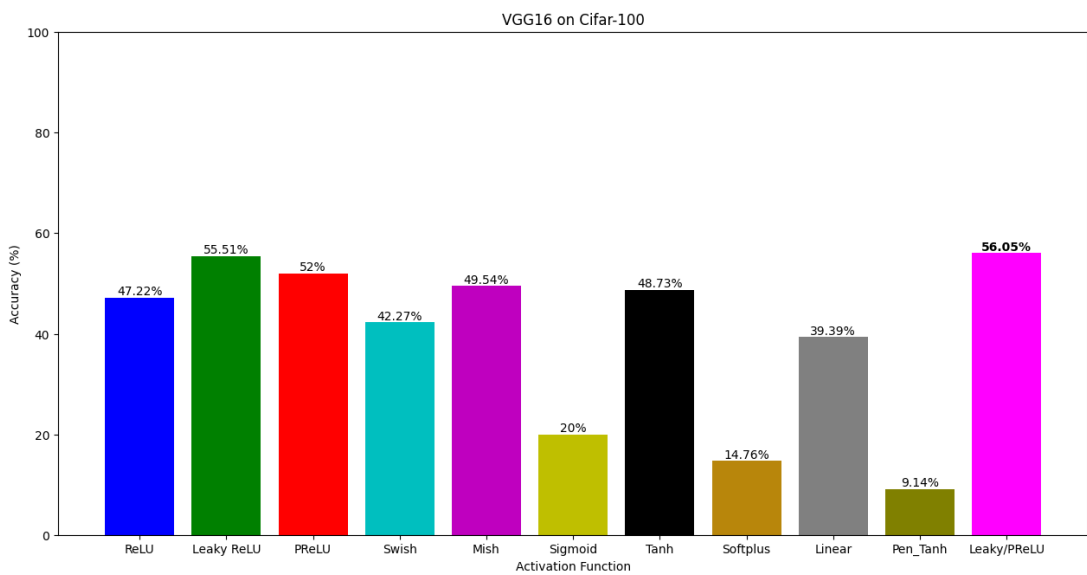
Σχετικά με το VGG16 φαίνονται στο Σχήμα 4.9 τα ποσοστά ακρίβειας των συναρτήσεων. Εδώ οι διακυμάνσεις είναι μεγαλύτερες σε σχέση με το Alexnet. Ο συνδυασμός Leaky/PReLU φαίνεται να κατακτά την πρωτιά με 56.05% με την LeakyReLU να ακολουθεί με διαφορά περίπου 0.5%. Σχετικά με τις χρονικές απαιτήσεις στο Σχήμα 4.10, η ReLU έρχεται εύκολα πρώτη με χρόνο 962 δευτερόλεπτα ενώ η Leaky Relu χρειάστηκε 1,012 και ο συνδιασμός Leaky/PReLU 1,023.

Στο Σχήμα 4.11 αποτυπώνονται οι επιδόσεις του CNN3 στο Cifar-100. Ξανά πρώτος σε ακρίβεια διακρίνεται ο συνδιασμός Leaky/PReLU με 49.61% και διαφορά μικρότερη από 0.1% από τη Leaky ReLU. Εντύπωση προκαλεί και το γεγονός πως το δίκτυο με τη ReLU δεν κατάφερε να αποδώσει στην εκπαίδευση. Σχετικά με τη ταχύτητα εκπαίδευσης, φαίνεται στο Σχήμα 4.12 πως πρώτη έρχεται η Linear με 388 δευτερόλεπτα, ενώ η Leaky ReLU με 425 είναι ταχύτερη κατά 12 δευτερόλεπτα από τον συνδυασμό Leaky/PReLU.

Σχήμα 4.8: Alexnet-Cifar100 Χρονομέτρηση



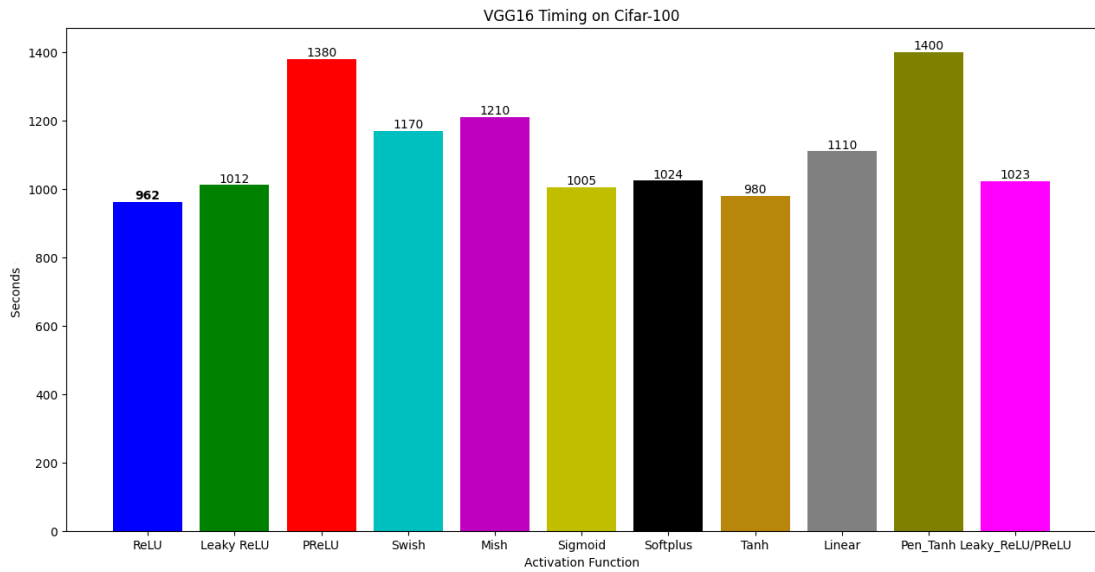
Σχήμα 4.9: VGG16-Cifar100 Ακρίβεια



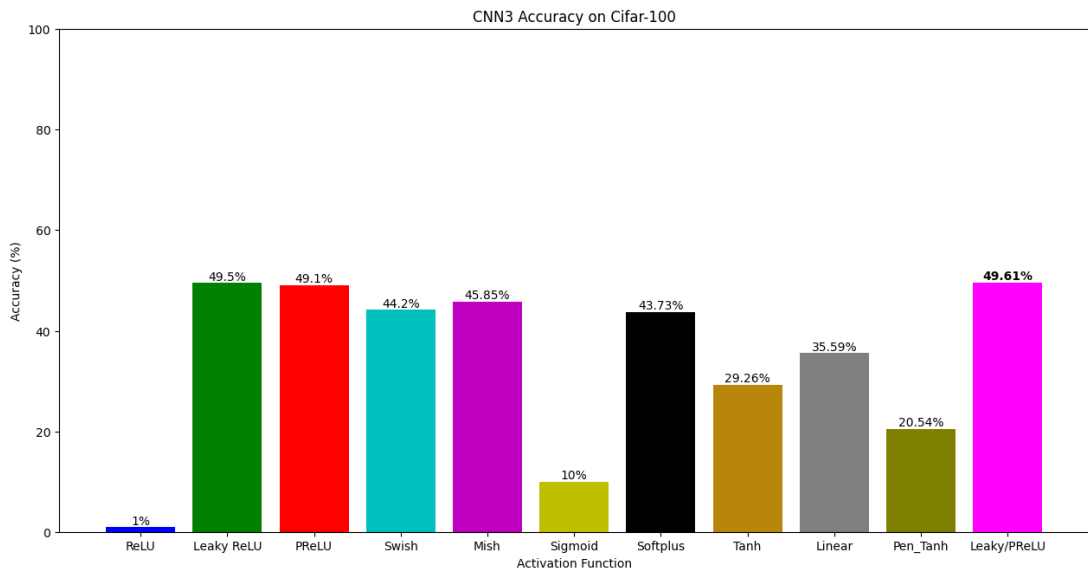
4.1.3 Αποτελέσματα Vegetable-Image

Στο Σχήμα 4.13 αποτυπώνονται τα ποσοστά ακρίβειας των συναρτήσεων στο Alexnet για το Vegetables-Image-Dataset. Η Softplus έρχεται πρώτη με ποσοστό 95.6% με τη ReLU να ακολουθεί με διαφορά μόλις 0.1%. Όσον αφορά τους χρόνους

Σχήμα 4.10: VGG16-Cifar100 Χρονομέτρηση



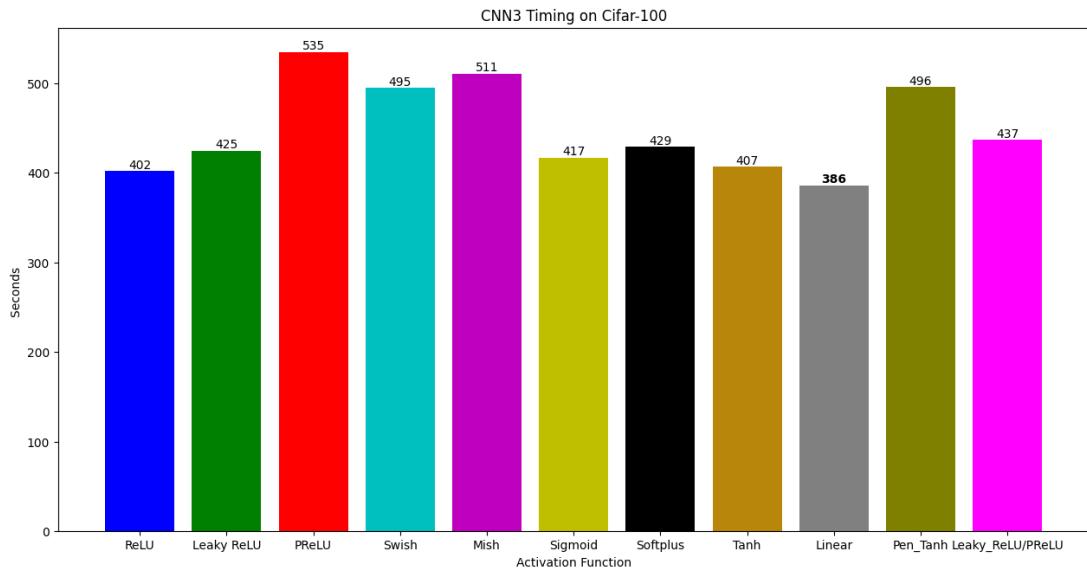
Σχήμα 4.11: CNN3-Cifar100 Ακρίβεια



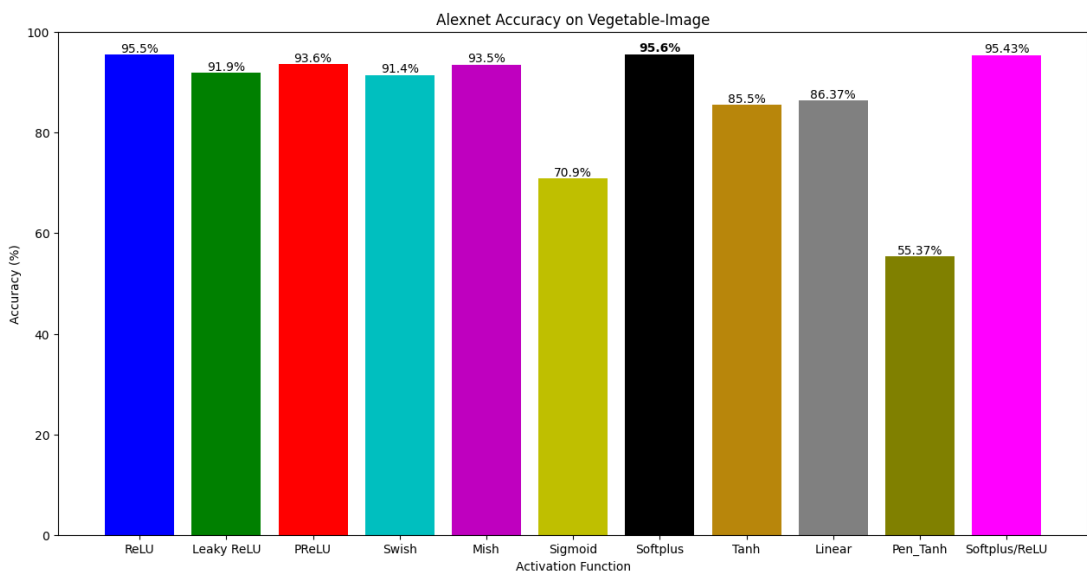
εκπαίδευσης στο Σχήμα 4.14 η Tanh έρχεται πρώτη με 422 δευτερόλεπτα και η Softplus με 428 αφήνει πολύ πίσω την ReLU η οποία χρειάστηκε περίπου 50 δευτερόλεπτα παραπάνω.

Για την ακρίβεια του VGG16, στο Σχήμα 4.15 η Swish με ποσοστό ακρίβειας 96.3% έρχεται πρώτη με σημαντική διαφορά -περίπου 2.4%- από τη δεύτερη Linear.

Σχήμα 4.12: CNN3-Cifar100 Χρονομέτρηση



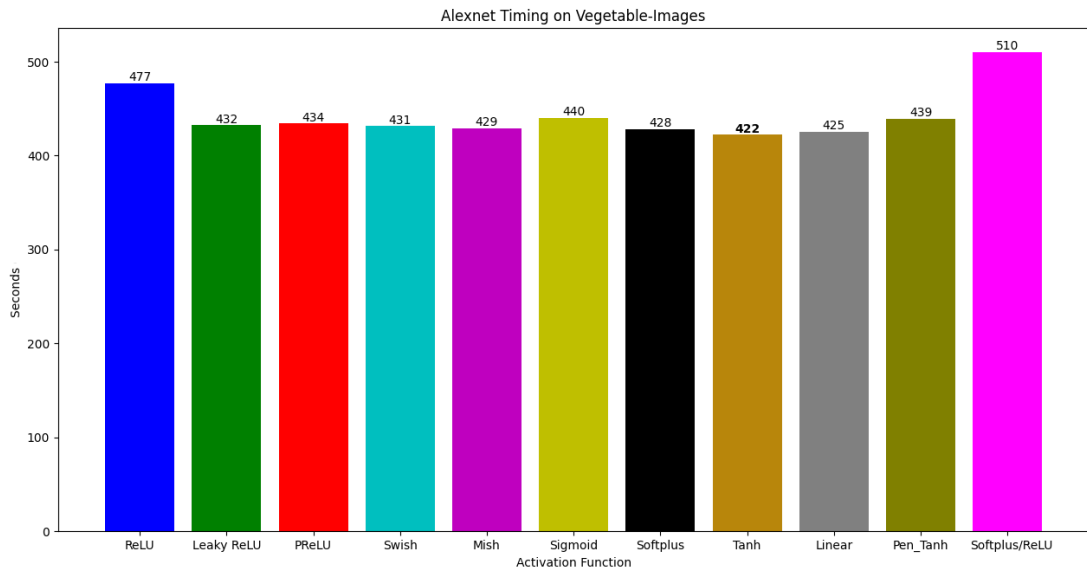
Σχήμα 4.13: Alexnet-Vegetables Ακρίβεια



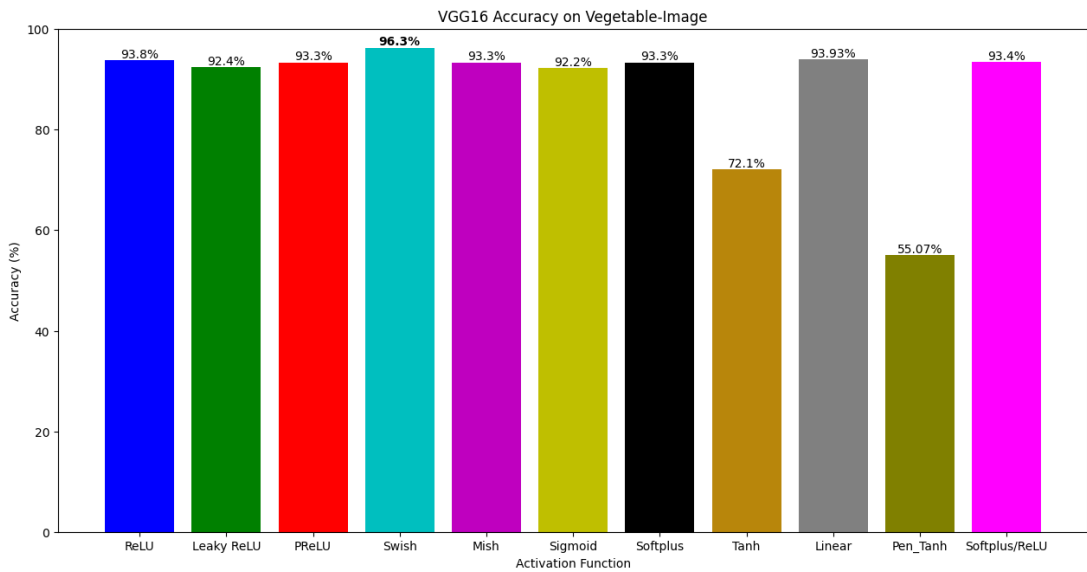
Στο Σχήμα 4.16 που αφορά τους χρόνους εκπαίδευσης οι δύο αυτές συναρτήσεις είναι και πάλι πρωταγωνίστριες με αντεστραμμένους ρόλους αυτή τη φορά. Πιο συγκεκριμένα η Linear είναι η ταχύτερη με 531 δευτερόλεπτα και η Swish ακολουθεί με 539.

Περνώντας στο CNN3, στο Σχήμα 4.17, μεγαλύτερη ακρίβεια πετυχαίνει η Softplus

Σχήμα 4.14: Alexnet-Vegetables Χρονομέτρηση



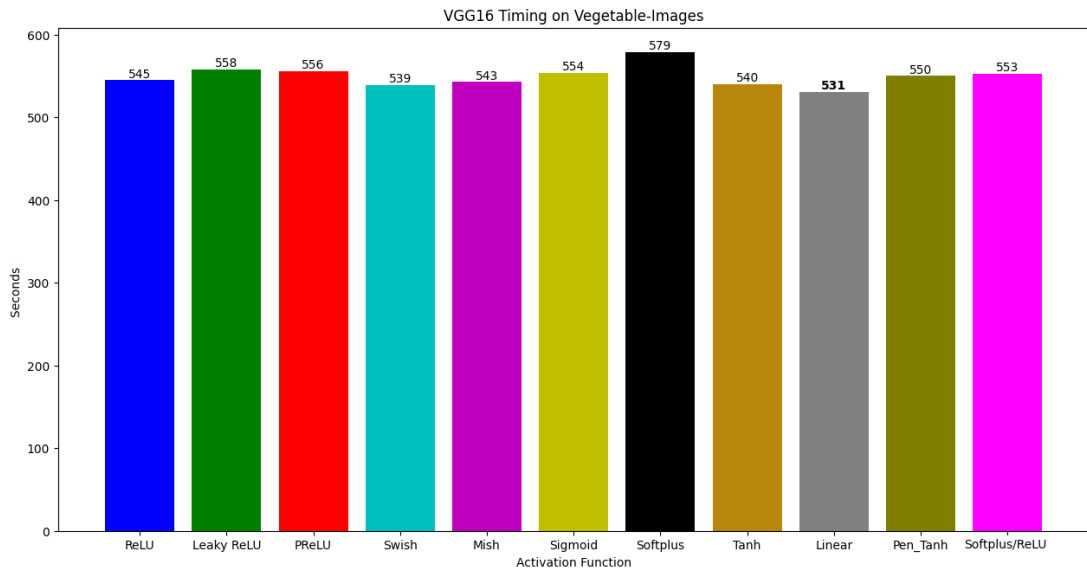
Σχήμα 4.15: VGG16-Vegetables Ακρίβεια



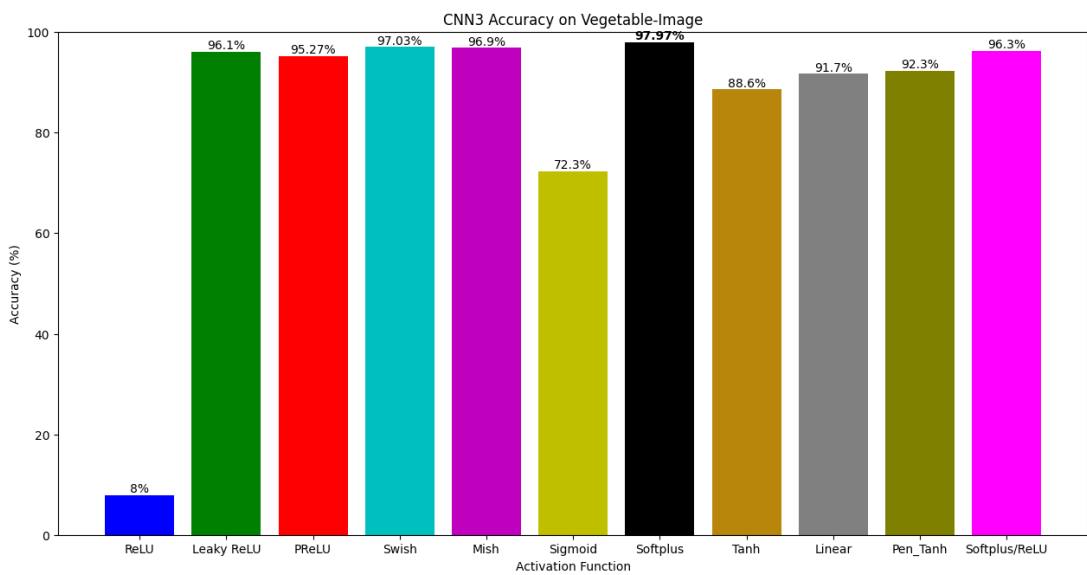
με ποσοστό 97.97% με τις Swish και Mish να ακολουθούν με διαφορά κοντά στο 1%. Ξαναβλέπουμε πως η ReLU δεν ανταποκρίθηκε καθόλου στα δεδομένα εκπαίδευσης δίνοντας ποσοστό 8% το οποίο ανταποκρίνεται σε τυχαίες προβλέψεις.

Από το Σχήμα 4.18 φαίνεται πως η Softplus αναδεικνύεται ξεκάθαρος νικητής στο συγκεκριμένο πείραμα καθώς πέρα απ το καλύτερο ποσοστό, χρειάστηκε και

Σχήμα 4.16: VGG16-Vegetables Χρονομέτρηση

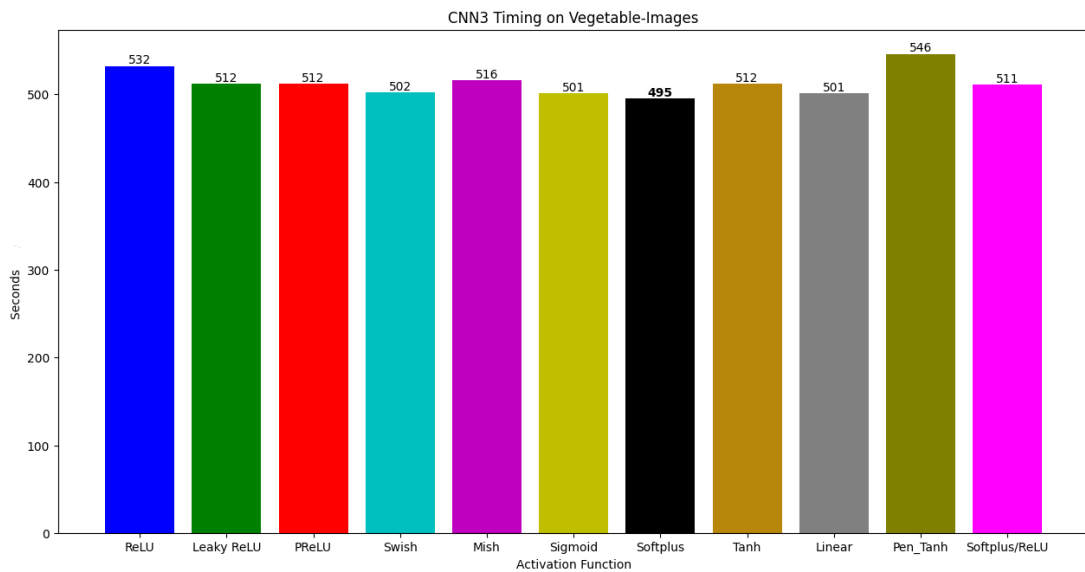


Σχήμα 4.17: CNN3-Vegetables Ακρίβεια



τον μικρότερο χρόνο εκπαίδευσης με 495 δευτερόλεπτα. Η Swish βρίσκεται σε απόσταση 7 δευτερολέπτων.

Σχήμα 4.18: CNN3-Vegetables Χρονομέτρηση



4.2 Εξόρυξη Γνώμης

Για την εξόρυξη γνώμης χρησιμοποιήθηκαν δύο δίκτυα με διαφορετική αρχιτεκτονική. Το TextCNN [15], ένα συνελκτικό νευρωνικό δίκτυο η αρχιτεκτονική του οποίου φαίνεται με λεπτομέρεια στο Σχήμα 4.19.

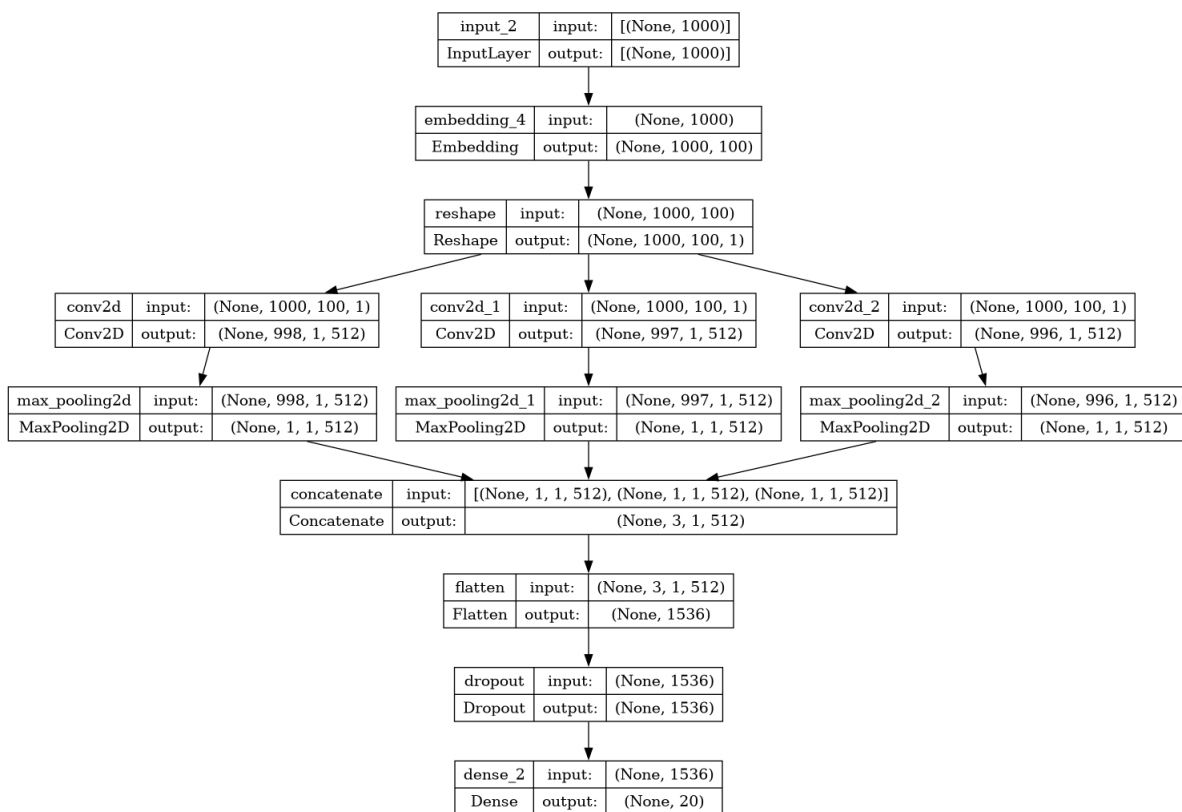
Ένα μικρό επαναληπτικό δίκτυο, το LSTM1, που δημιουργήθηκε για τις ανάγκες των πειραμάτων και αποτελείται από μόλις ένα επίπεδο LSTM και ένα επίπεδο εξόδου. Ακολουθεί η αρχιτεκτονική του στο Σχήμα 4.20.

Για τις μετρήσεις χρησιμοποιήθηκαν δύο σύνολα δεδομένων:

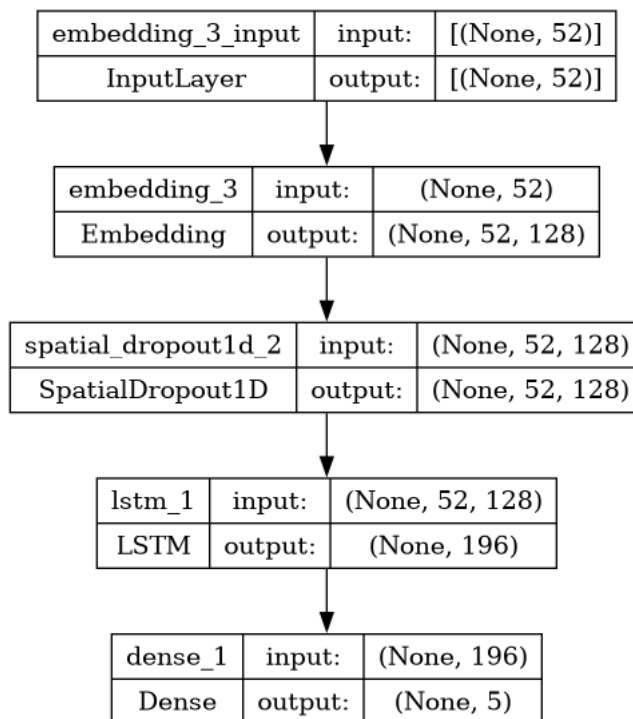
- Το Movie Review, το οποίο περιέχει κριτικές ταινιών. Το δίκτυο καλείται να εφαρμόσει γλωσσική ανάλυση με σκοπό να κατατάξει την κάθε κριτική σε μία εκ των πέντε κατηγοριών: Αρνητική, Λίγο Αρνητική, Ουδέτερη, Θετική, Πολύ Θετική.
- Το Newsgroup-20, που περιέχει μηνύματα τα οποία προέρχονται από 20 διαφορετικές ομάδες ειδήσεων (π.χ. Πολιτική, Αθλητισμός κλπ). Το δίκτυο καλείται να κατατάξει το κάθε μήνυμα στη σωστή κατηγορία απ την οποία προέρχεται.

Οι παράμετροι εκπαίδευσης είναι οι εξής:

Σχήμα 4.19: Η Αρχιτεκτονική του TextCNN 4.19



Σχήμα 4.20: Η Αρχιτεκτονική του LSTM1 4.20



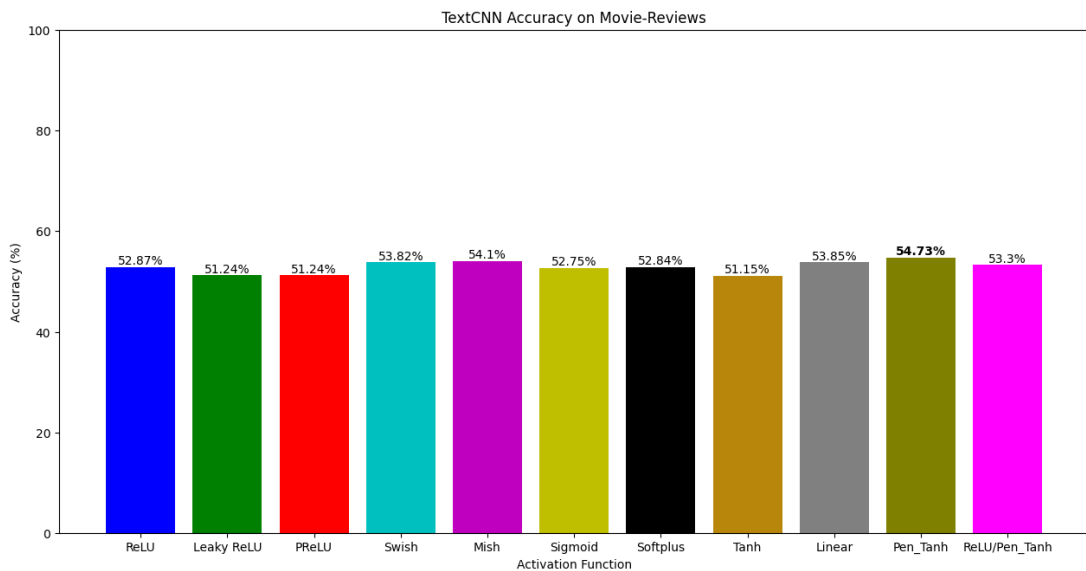
- Διάρκεια Εκπαίδευσης: 25 εποχές
- Ρυθμός Μάθησης: 0.001

- Μέγεθος δεσμίδας: Στο Movie-Review το μέγεθος δεσμίδας ήταν 32 ενώ στο 20-Newsgroup ήταν 256.
- Συνάρτηση Απωλειών: Κατηγοριακή Διασταυρούμενη Εντροπία
- Μετρική: Ακρίβεια

4.2.1 Αποτελέσματα Movie-Reviews

Στο Σχήμα 4.21 παρουσιάζονται οι επιδόσεις της κάθε συνάρτησης στο TextCNN.

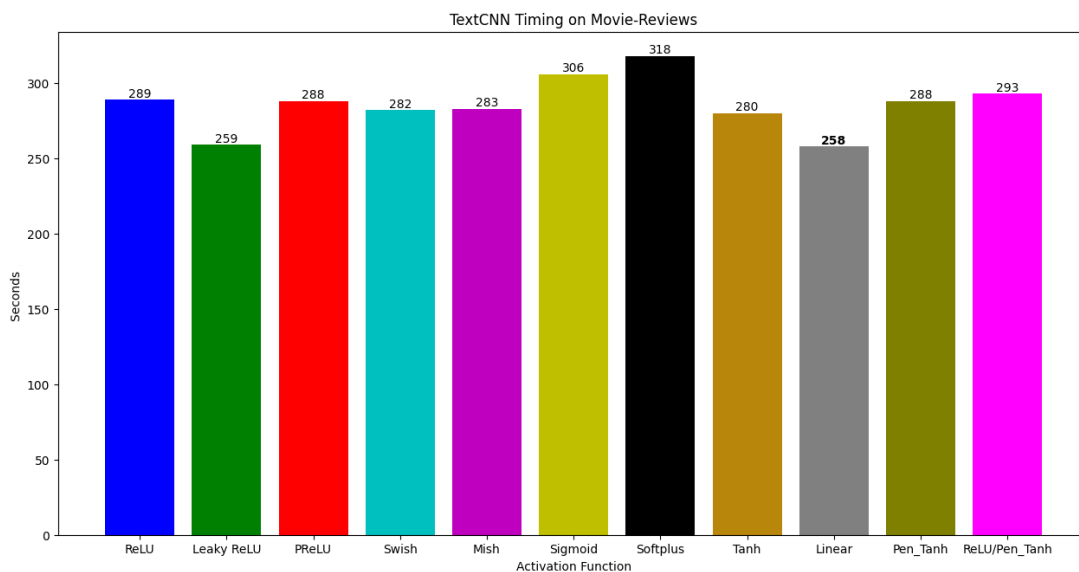
Σχήμα 4.21: TextCNN-Movie Reviews Ακρίβεια



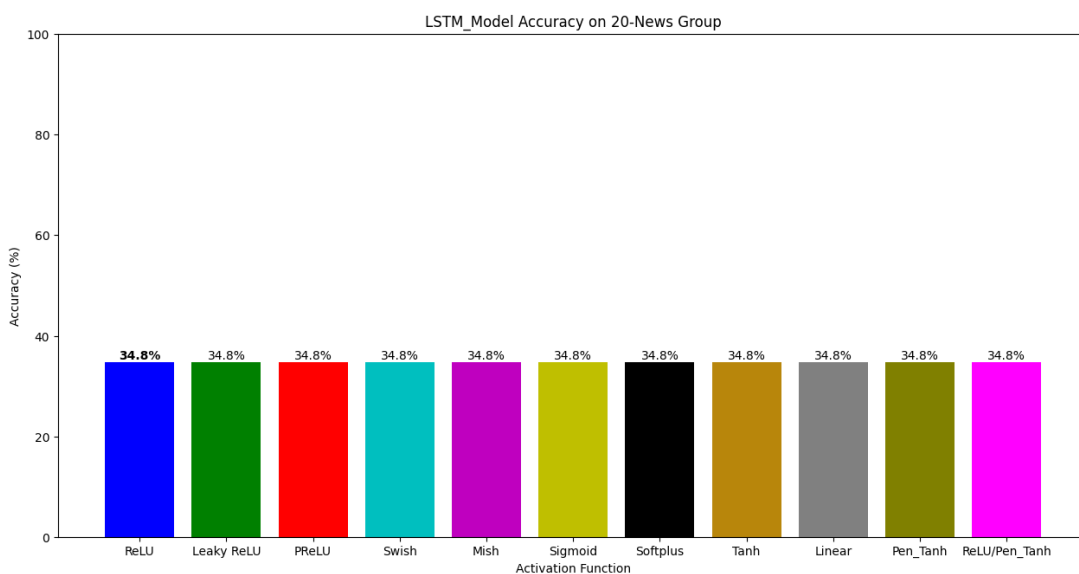
Παρατηρείται πως οι διαφορές από τη πρώτη έως την τελευταία συνάρτηση είναι μικρές. Πρώτη έρχεται η Penalized Tanh με 54.73%, ενώ τελευταία η Tanh με 51.15%. Στους χρόνους εκπαίδευσης (Σχήμα 4.22) η Linear είναι η ταχύτερη με 258 δευτερόλεπτα με τη Leaky ReLU να ακολουθεί ένα δευτερόλεπτο πίσω, ενώ η Penalized Tanh χρειάστηκε μισό λεπτό παραπάνω από την Linear.

Σχετικά με το LSTM1 στο Σχήμα 4.23, για πρώτη φορά το δίκτυο παρήγαγε το ίδιο αποτέλεσμα ανεξαρτήτως συνάρτησης. Στις χρονικές επιδόσεις τα πράγματα αλλάζουν καθώς στο Σχήμα 4.24 φαίνεται πως η Tanh είναι η ταχύτερη με 509 δευτερόλεπτα με την Penalized Tanh να ακολουθεί με διαφορά ενός δευτερολέπτου. Τελευταία έρχεται η PReLU με 595 δευτερόλεπτα.

Σχήμα 4.22: TextCNN-Movie Reviews Χρονομέτρηση



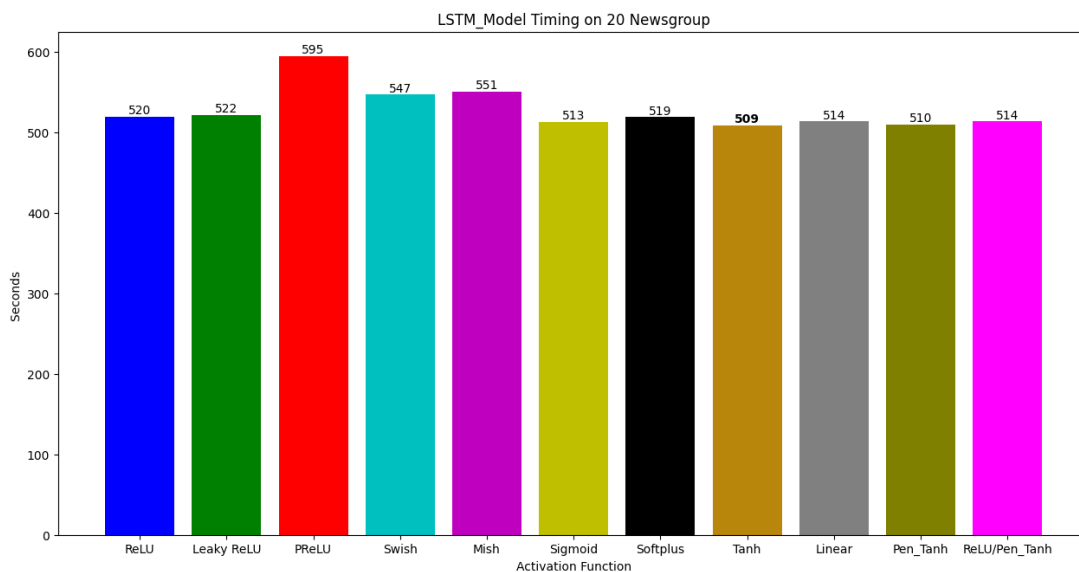
Σχήμα 4.23: LSTM1-Movie Reviews Ακρίβεια



4.2.2 Αποτελέσματα 20-Newsgroup

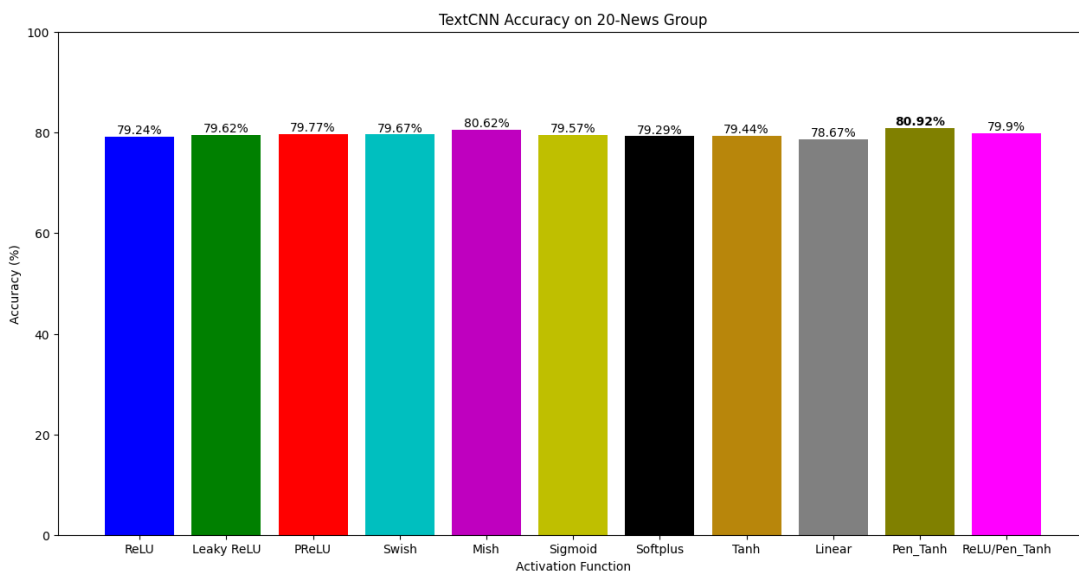
Τα αποτελέσματα που μας έδωσε το TextCNN στο σύνολο δεδομένων 20-NewsGroup μοιάζουν με τα αποτελέσματα στο Movie-Reviews. Από το Σχήμα 4.25 είναι εμφανές πως οι διαφορές μεταξύ των συναρτήσεων είναι μικρές. Καλύτερη απόδοση πέτυχε

Σχήμα 4.24: LSTM1-Movie Reviews Χρονομέτρηση



και πάλι η Penalized Tanh με 80.92%, ενώ τελευταία ήρθε η Linear με 78.67%.

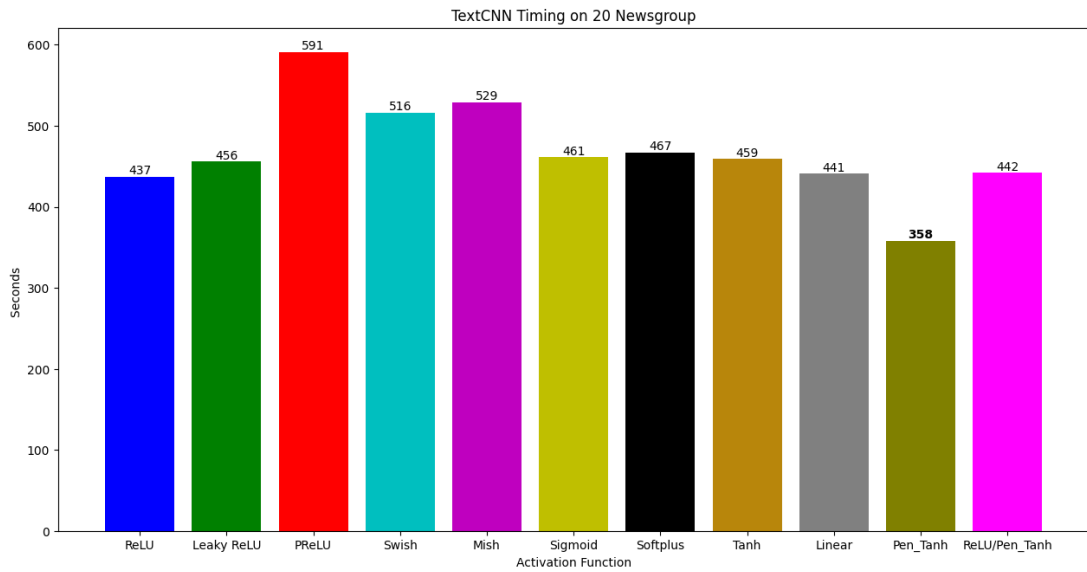
Σχήμα 4.25: TextCNN 20-Newsgroup Ακρίβεια



Χρονικά, στο Σχήμα 4.26 φαίνεται πως η Penalized Tanh έρχεται πρώτη με 358 δευτερόλεπτα και με μεγάλη απόσταση -79 δευτερολέπτων- από τη δεύτερη ReLU. Τελευταία έρχεται η PReLU με 591 δευτερόλεπτα.

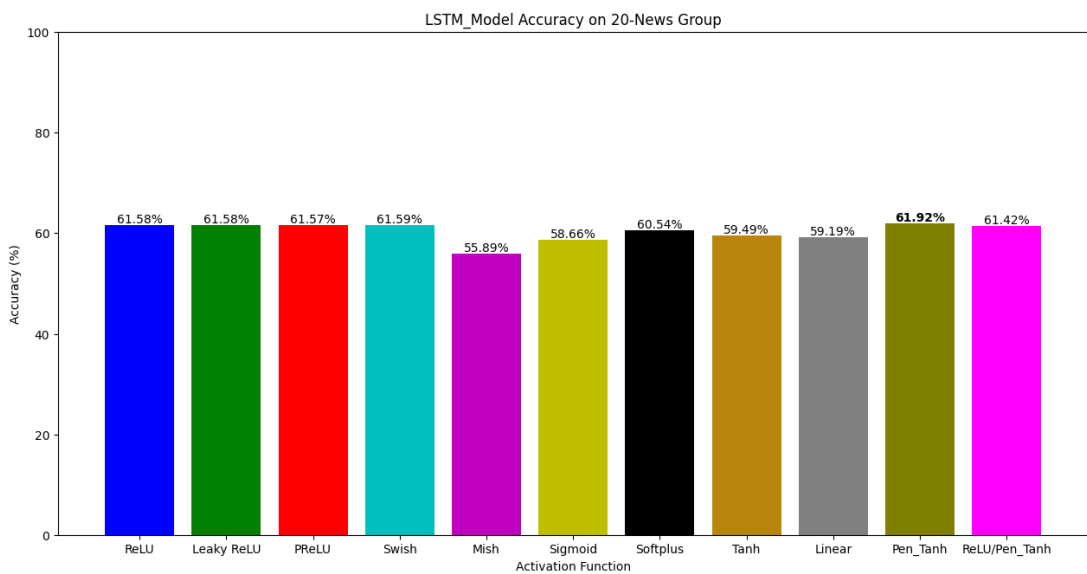
Σχετικά με το LSTM1, βλέποντας το Σχήμα 4.27 διακρίνεται πως για μια ακόμη

Σχήμα 4.26: TextCNN 20-Newsgroup Χρονομέτρηση

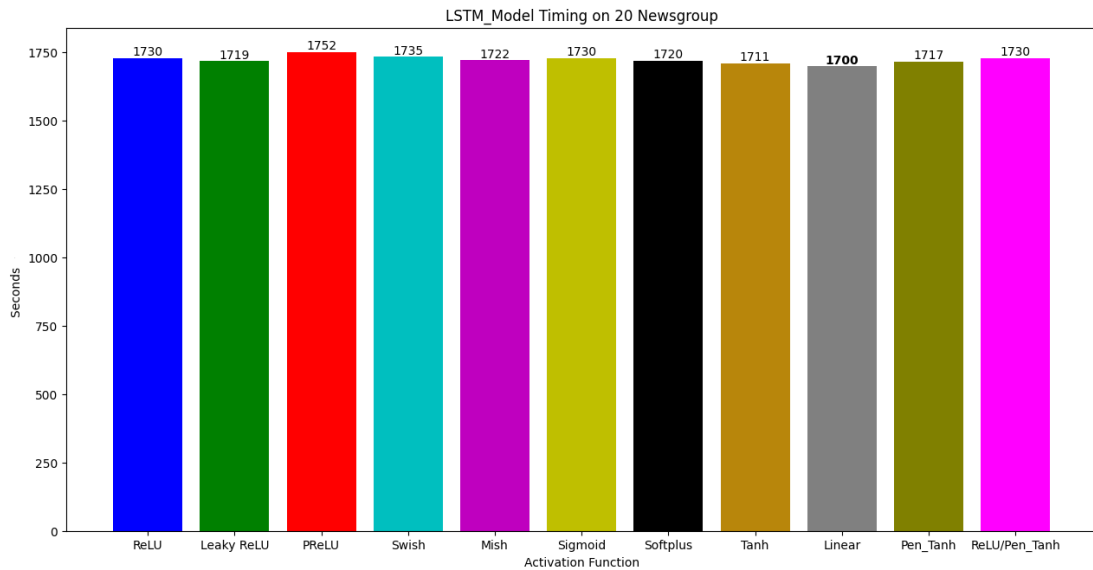


φορά την πρωτιά καταλαμβάνει η Penalized Tanh με ποσοστό 61.92%. Για τους χρόνους στο Σχήμα 4.28 οι διαφορές είναι μικρές. Πρώτη έρχεται η Linear με 1,700 δευτερόλεπτα, με Tanh και Penalized Tanh να ακολουθούν με 11 και 17 δευτερόλεπτα, αντίστοιχα. Θα δούμε και πάλι πως η PReLU έρχεται τελευταία με 1,752 δευτερόλεπτα.

Σχήμα 4.27: LSTM1 20-Newsgroup Ακρίβεια



Σχήμα 4.28: LSTM1 20-Newsgroup Χρονομέτρηση



4.3 Δεδομένα σε Μορφή Πίνακα

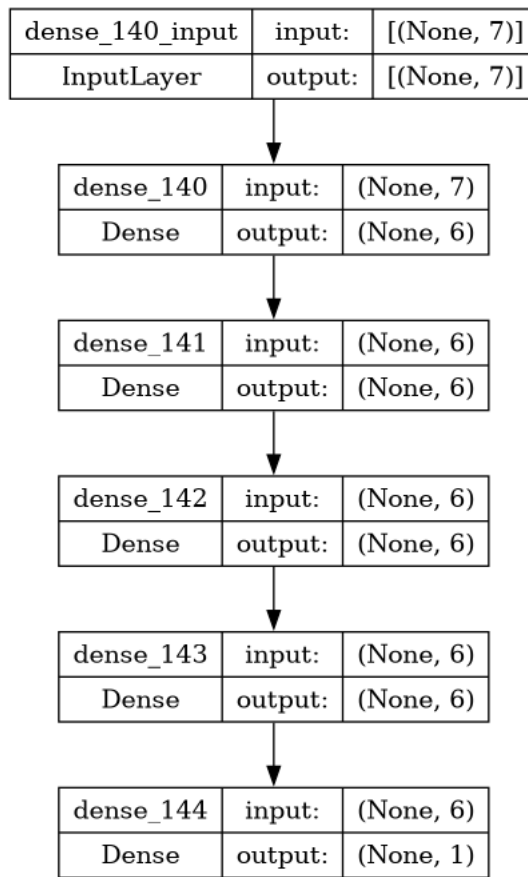
Για τα δεδομένα σε πίνακες χρησιμοποιήθηκαν δυο μοντέλα τεχνητών νευρωνικών δικτύων, τα DNN1 και DNN2. Το DNN1 έχει τέσσερα κρυφά επίπεδα και η αρχιτεκτονική του φαίνεται με λεπτομέρεια στο Σχήμα 4.29.

Το DNN2 έχει δύο κρυφά επίπεδα, και η αρχιτεκτονική του παρουσιάζεται στο Σχήμα 4.30.

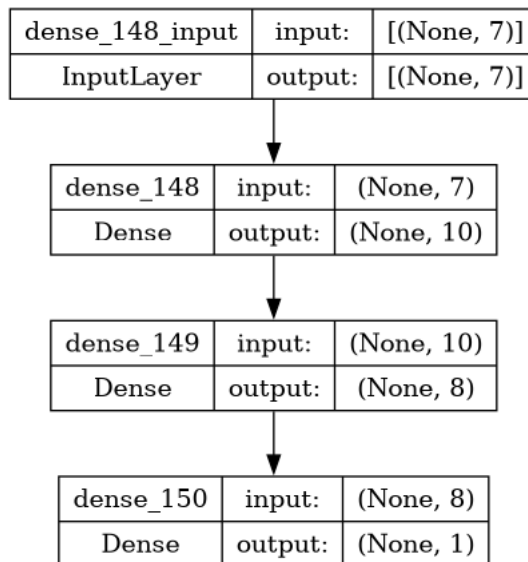
Για τα πειράματα χρησιμοποιήθηκαν τα εξής 3 σύνολα δεδομένων:

- Churn Modeling: Αυτό το σύνολο δεδομένων περιέχει στοιχεία πελατών μιας τράπεζας όπως η ηλικία, το ύψος καταθέσεων, το ύψος μισθού κλπ. Με βάση 10 σχετικά στοιχεία γίνεται πρόβλεψη εάν ένας πελάτης θα εγκαταλείψει τη τράπεζα.
- House Price: Το συγκεκριμένο σύνολο δεδομένων περιέχει μερικά χαρακτηριστικά ενός σπιτιού όπως έτος δόμησης, έκταση κλπ. Με γνώμονα πέντε σχετικά χαρακτηριστικά γίνεται εκτίμηση της τιμής του.
- Hotel-Reviews: Αυτό το σύνολο δεδομένων περιέχει κριτικές πελατών ξενοδοχείων καθώς και στοιχεία που σχετίζονται με τους πελάτες και με την κράτηση

Σχήμα 4.29: Η Αρχιτεκτονική του DNN1 4.29



Σχήμα 4.30: Η αρχιτεκτονική του DNN2 4.30



τους. Από 10 σχετικές πληροφορίες για κάθε κριτική γίνεται προσπάθεια πρόβλεψης της κριτικής που θα δώσει ο πελάτης (στο εύρος 1-10).

Οι παράμετροι εκπαίδευσης είναι οι εξής:

-
- Διάρκεια Εκπαίδευσης: 25 εποχές
 - Ρυθμός Μάθησης : 0.001
 - Μέγεθος δεσμίδας: 32
 - Συνάρτηση Απωλειών: Στο Churn Modeling χρησιμοποιήθηκε η δυαδική διασταυρούμενη εντροπία (Binary Crossentropy) ενώ στα άλλα δύο σύνολα δεδομένων χρησιμοποιήθηκε το τετράγωνο του μέσου απόλυτου σφάλματος (Mean Squared Error)
 - Μετρική: Στο Churn Modeling Χρησιμοποιήθηκε η Ακρίβεια ενώ στα άλλα δύο Dataset το R2-Score.

4.3.1 Churn Modeling

Όπως φαίνεται στο Σχήμα 4.31 που σχετίζεται με τις επιδόσεις του DNN1 στο Churn Modeling, η Leaky ReLU έρχεται πρώτη με 86.55% με πολύ μικρή διαφορά (0.01) από τη δεύτερη Swish. Είναι εμφανές επίσης πως όλες οι συναρτήσεις βρίσκονται σε εύρος διαφοράς μικρότερο του 1.4% εξαιρουμένων των Tanh και Linear με 80.55% και 80.15% αντίστοιχα.

Στους χρόνους στο Σχήμα 4.32, η Sigmoid να πρώτη με 38 δευτερόλεπτα, ενώ η Leaky ReLU δε βρίσκεται πολύ πίσω με μόλις ένα δευτερόλεπτο διαφορά.

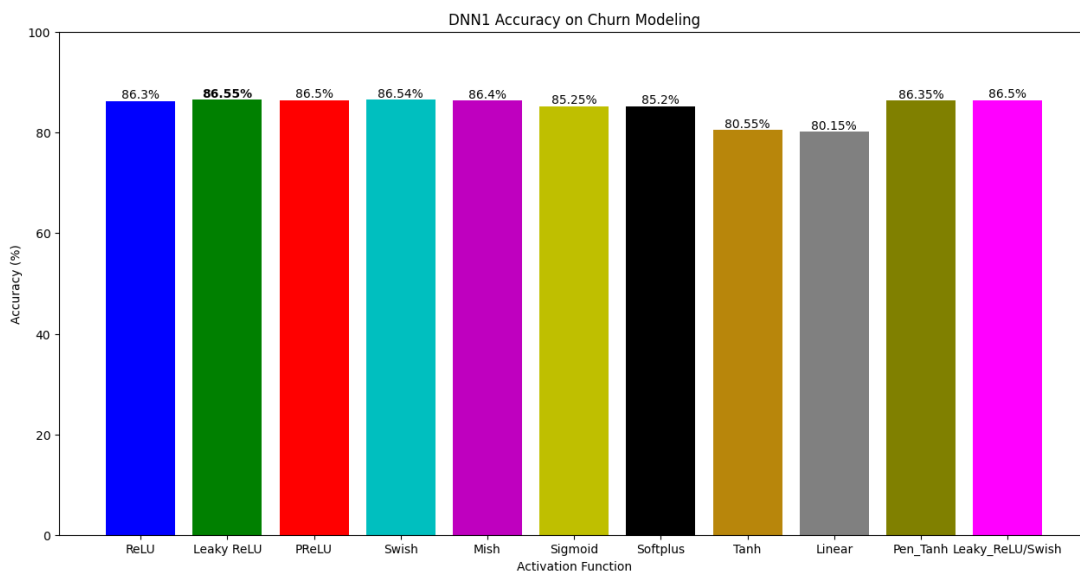
Σχετικά με το DNN2 στο Σχήμα 4.33, η εικόνα δε διαφέρει σημαντικά. Η Swish αυτή τη φορά αναδεικνύεται πρώτη με 86.85% με όλες τις υπόλοιπες συναρτήσεις να ακολουθούν με μικρές διαφορές εκτός της Linear που έδωσε 80.8%.

Χρονικά, στο Σχήμα 4.34 η Linear έρχεται πρώτη με 32 δευτερόλεπτα. Η Swish φαίνεται πως χρειάζεται περισσότερο χρόνο από τις περισσότερες συναρτήσεις καθώς ολοκλήρωσε την εκπαίδευση στα 34 δευτερόλεπτα, ενώ για μια ακόμη φορά πως η PReLU ήρθε τελευταία με 39 δευτερόλεπτα.

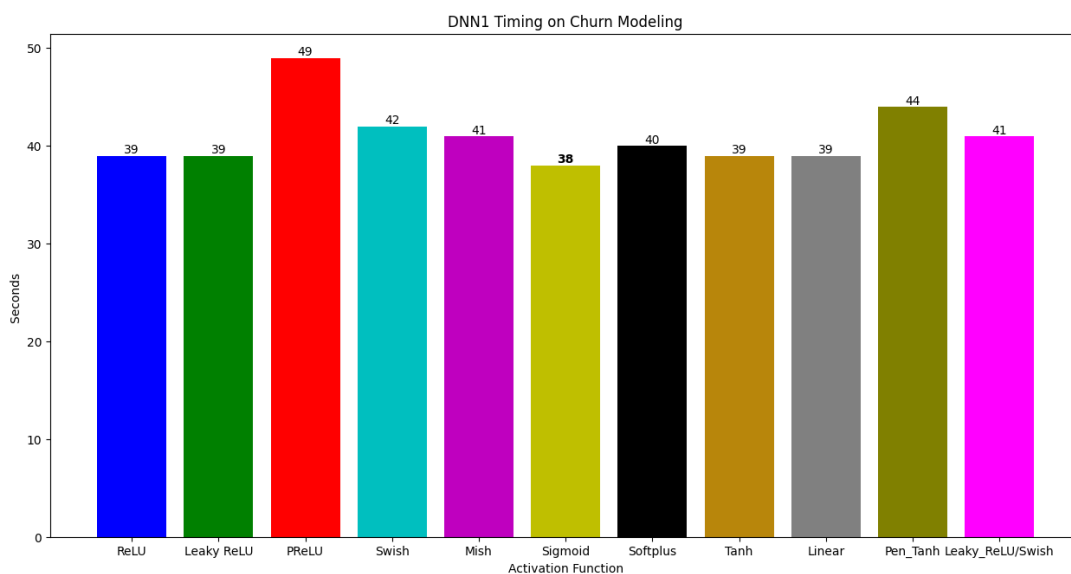
4.3.2 House Price

Στο Σχήμα 4.35 φαίνονται οι επιδόσεις του DNN1 στο σύνολο δεδομένων House Price. Πρώτη έρχεται η Linear με 57.04%, ενώ σχεδόν όλες οι υπόλοιπες συναρτήσεις ακολουθούν με διαφορά έως 0.51%. Ενδιαφέρον παρουσιάζει πως οι 3 συναρτήσεις

Σχήμα 4.31: DNN1 Churn Modeling Ακρίβεια



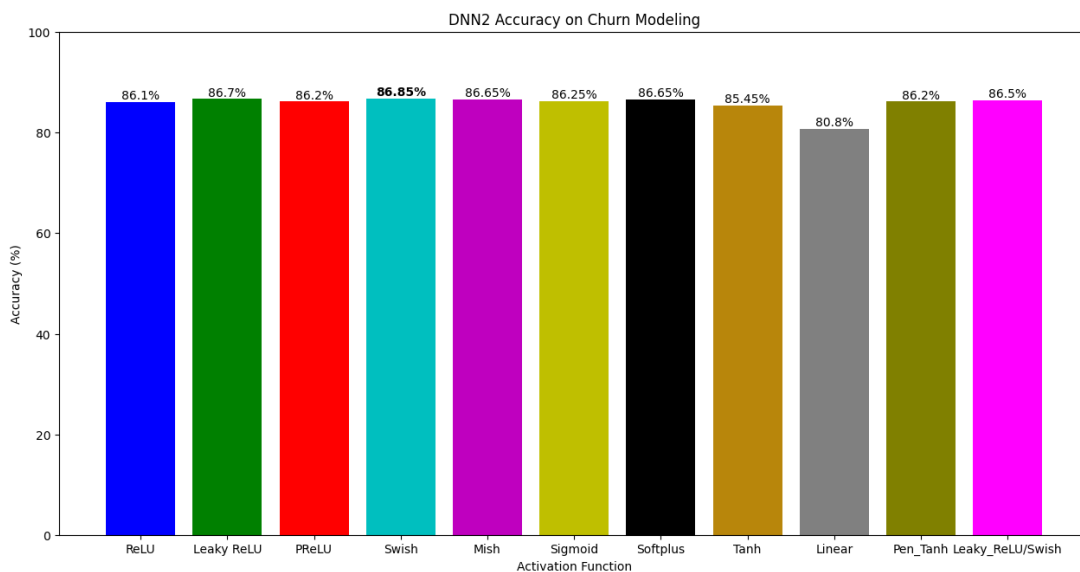
Σχήμα 4.32: DNN1 Churn Modeling Χρονομέτρηση



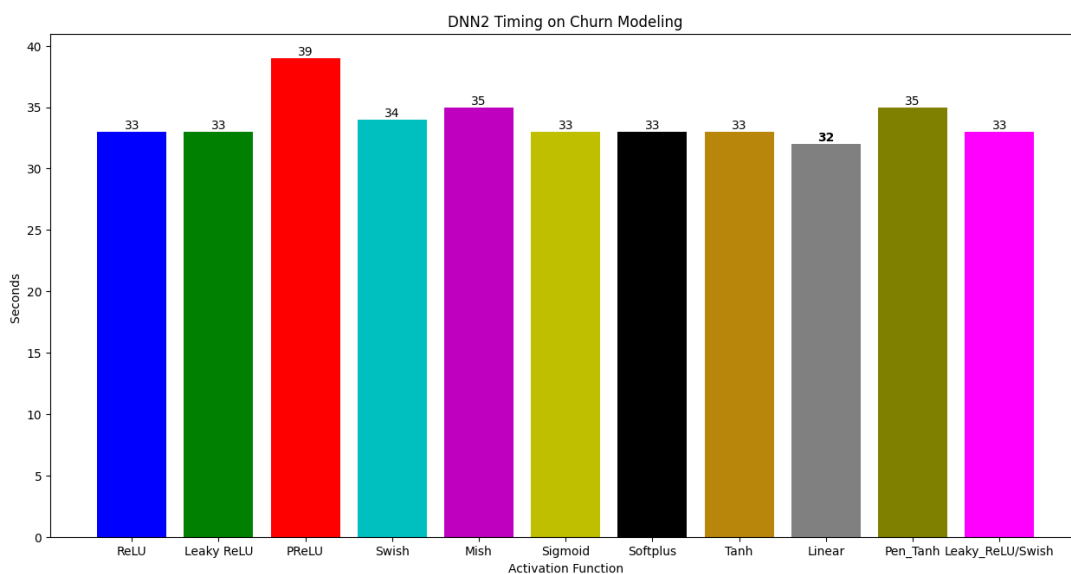
Sigmoid, Tanh, Penalized Tanh δεν ανταποκρίθηκαν καθόλου στην εκπαίδευση. Στις χρονικές απαιτήσεις στο Σχήμα 4.36 η Linear και πάλι αναδεικνύεται η καλύτερη επιλογή καθώς χρειάστηκε 29 δευτερόλεπτα, ένα λιγότερο από τις συναρτήσεις που την ακολουθούν.

Η ίδια ιστορία παρουσιάζεται και στο Σχήμα 4.37 που αφορά τις επιδόσεις του

Σχήμα 4.33: DNN2 Churn Modeling Ακρίβεια

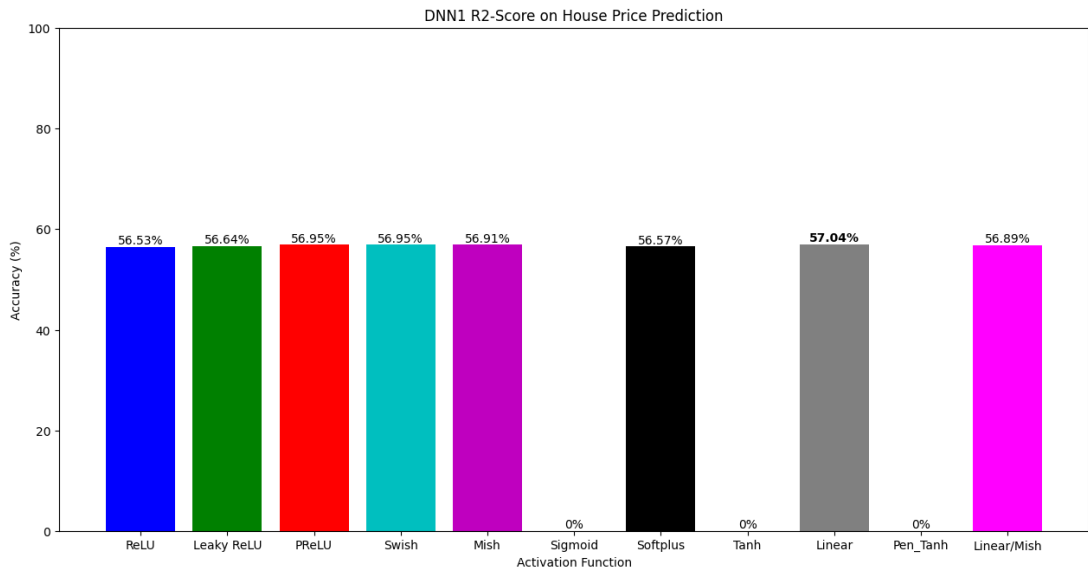


Σχήμα 4.34: DNN2 Churn Modeling Χρονομέτρηση

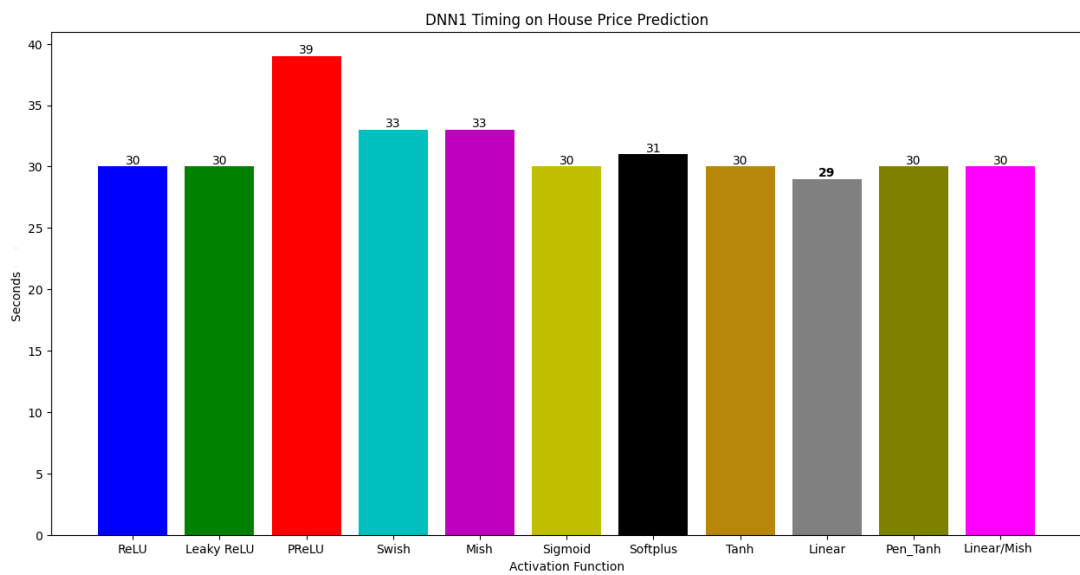


μοντέλου DNN2. Η Linear είναι πρώτη με ποσοστό 57.06% και μικρή διαφορά από τις υπόλοιπες εκτός των τριών Sigmoid, Tanh, Penalized Tanh. Χρονικά ωστόσο στο Σχήμα 4.38 οι ReLU και Leaky ReLU με 48 δευτερόλεπτα είναι ταχύτερες κατά 1 δευτερόλεπτο από τη Linear.

Σχήμα 4.35: DNN1 House Price R2-Score



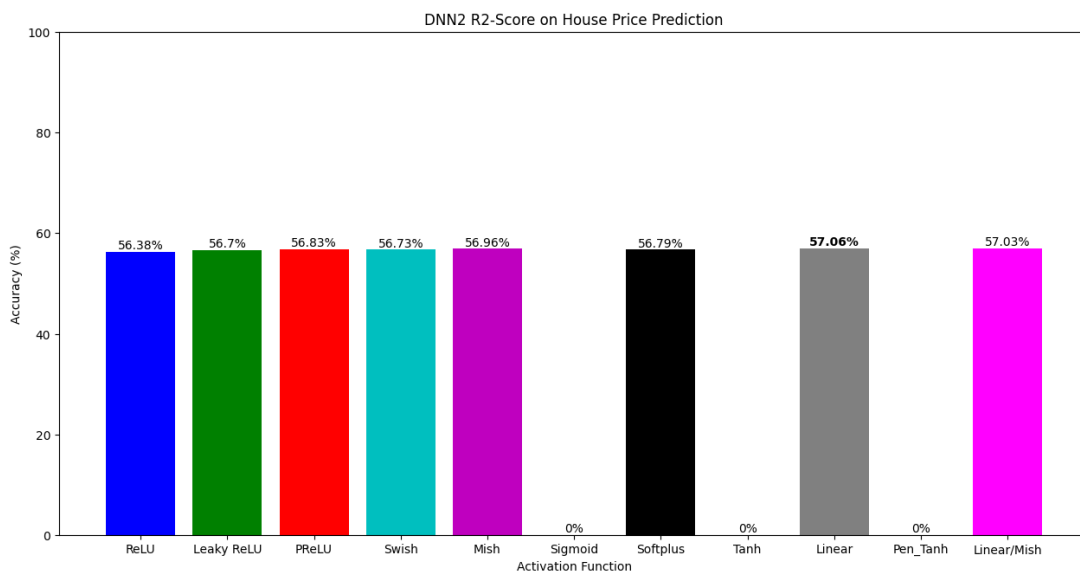
Σχήμα 4.36: DNN1 House Price Χρονομέτρηση



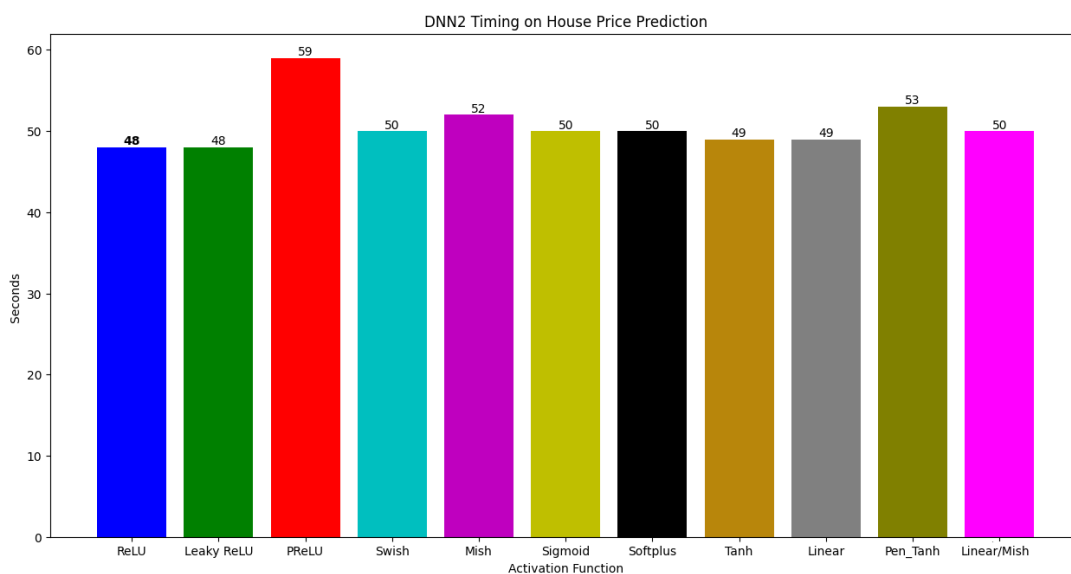
4.3.3 Hotel Reviews

Από το Σχήμα 4.39 που αφορά τις επιδόσεις του DNN1 στο σύνολο δεδομένων Hotel Reviews παρατηρείται πως το μοτίβο των αποτελεσμάτων του House Price επαναλαμβάνεται. Πρώτη είναι η Swish χωρίς μεγάλες διαφορές από τις υπόλοιπες

Σχήμα 4.37: DNN2 House Price R2-Score



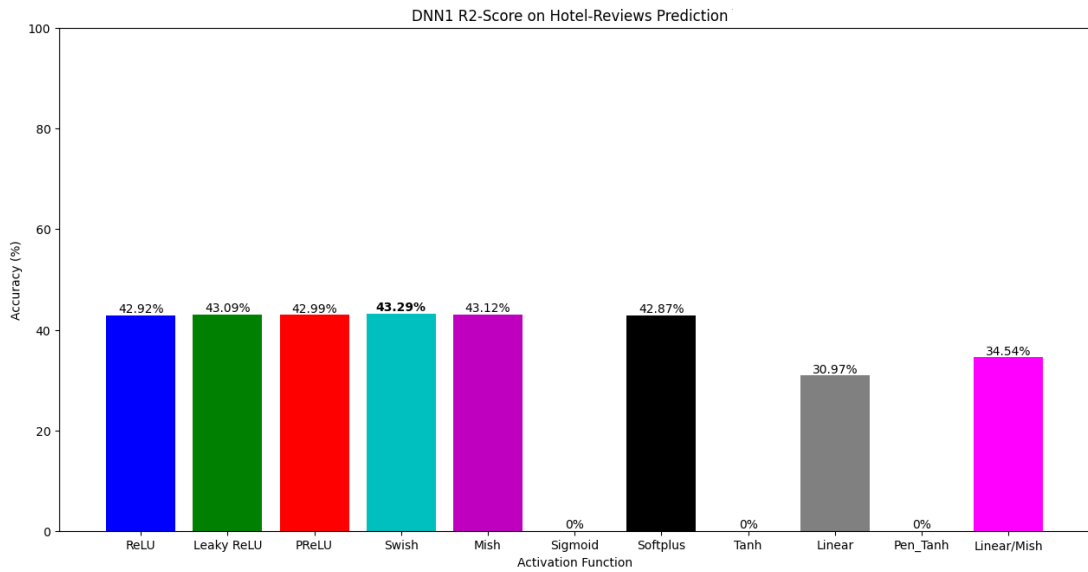
Σχήμα 4.38: DNN2 House Price Χρονομέτρηση



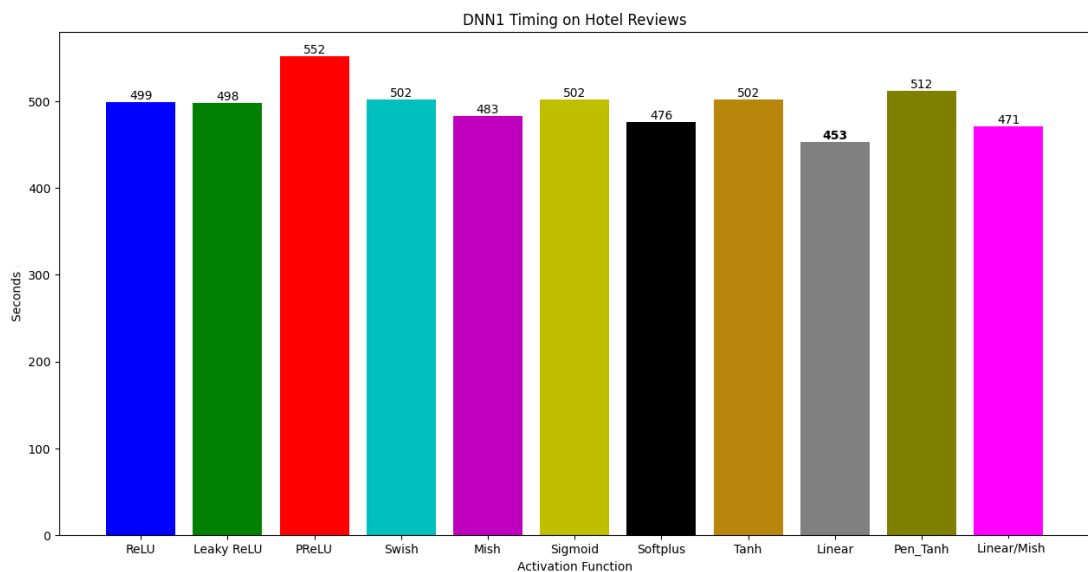
εκτός τη Linear, ενώ για μια ακόμη φορά η εκπαίδευση δεν απέδωσε στις Sigmoid, Tanh και Penalized Tanh. Χρονικά στο Σχήμα 4.40, η Linear με 453 δευτερόλεπτα είναι η ταχύτερη, ενώ από τις συναρτήσεις που πέτυχαν το καλύτερο MAΣ, η Softplus προηγείται με 476 δευτερόλεπτα.

Στο Σχήμα 4.41 αποτυπώνονται οι επιδόσεις του DNN2. Πρώτη με 43.19% έρ-

Σχήμα 4.39: DNN1 Hotel Reviews R2-Score

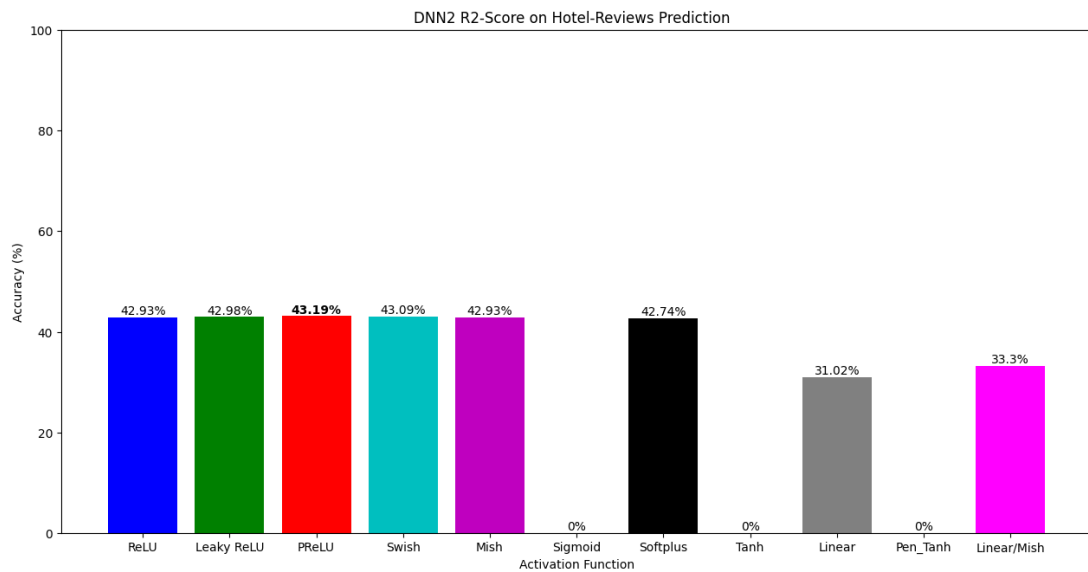


Σχήμα 4.40: DNN1 Hotel Reviews Χρονομέτρηση

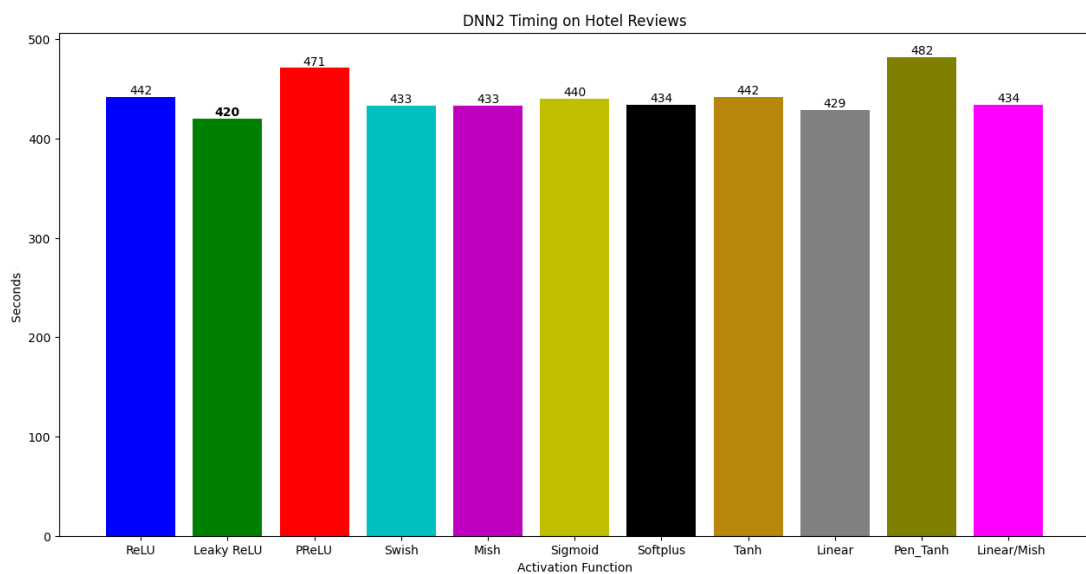


χεται η PReLU, ενώ παρατηρείται η ίδια συμπεριφορά για τις τρεις συναρτήσεις που δεν ανταποκρίνονται στην εκπαίδευση. Χρονικά, φαίνεται στο Σχήμα 4.42 πως πρώτη έρχεται η Leaky ReLU με 420 δευτερόλεπτα και διαφορά 51 δευτερολέπτων από την PReLU.

Σχήμα 4.41: DNN2 Hotel R2-Score



Σχήμα 4.42: DNN2 Hotel Reviews Χρονομέτρηση



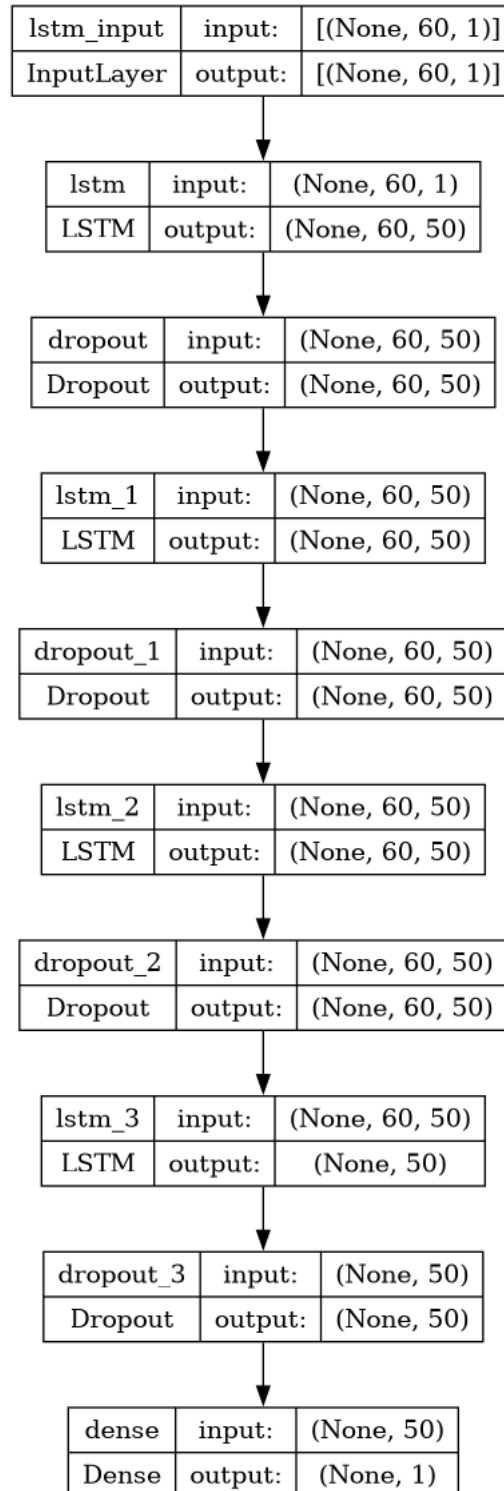
4.4 Δεδομένα Χρονοσειρών

Για τα δεδομένα χρονοσειρών χρησιμοποιήθηκαν δύο μοντέλα

- Το RNN1 με τέσσερα LSTM επίπεδα και ένα επίπεδο εξόδου. Η αρχιτεκτονική παρουσιάζεται με λεπτομέρεια στο Σχήμα 4.43.

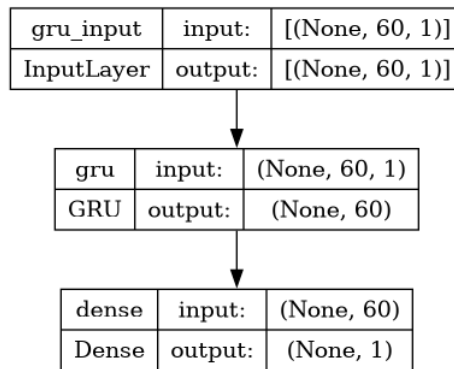
- Το RNN2 με ένα επίπεδο GRU και ένα επίπεδο εξόδου. Στο Σχήμα 4.44 φαίνεται η αρχιτεκτονική του.

Σχήμα 4.43: Η Αρχιτεκτονική του RNN1 4.43



Τα σύνολα δεδομένων που χρησιμοποιήθηκαν είναι το Google Stock Price και το Amazon Stock Price. Το δίκτυο καλείται να ανακαλύψει τα μοτίβα αυξομειώ-

Σχήμα 4.44: Η Αρχιτεκτονική του RNN2 4.44



σης της αξίας μια μετοχής ώστε να προβλέψει τη μελλοντική τιμή της με βάση τις προηγούμενες τιμές.

Οι παράμετροι της εκπαίδευσης είναι οι εξής:

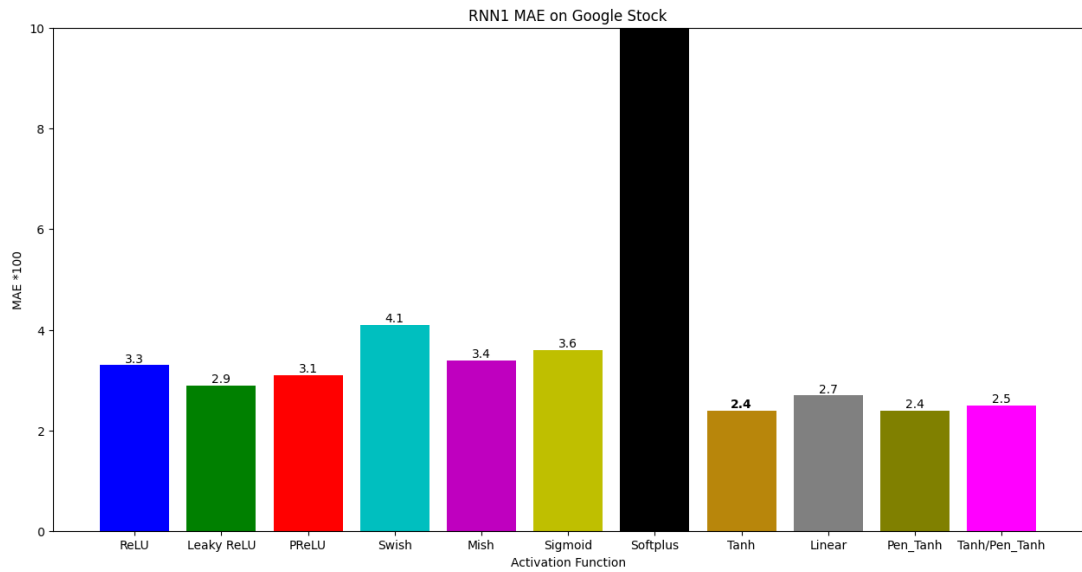
- Διάρκεια Εκπαίδευσης: 100 εποχές
- Ρυθμός Μάθησης: 0.001
- Μέγεθος δεσμίδας: 32
- Συνάρτηση Απωλειών: Το τετράγωνο του μέσου απόλυτου σφάλματος
- Μετρική: Μέσο απόλυτο σφάλμα *Σημειώνεται πως οι τιμές του ΜΑΣ που φαίνονται στα διαγράμματα πολλαπλασιασμένες επι 100.

4.4.1 Google Stock

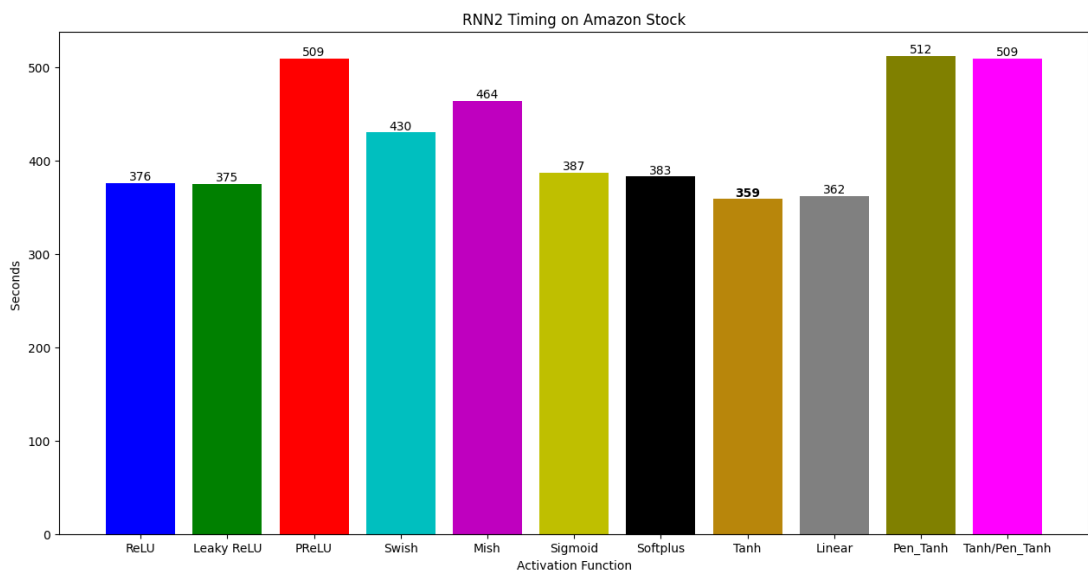
Στο Σχήμα 4.45 αποτυπώνεται η απόδοση του RNN1 στο σύνολο δεδομένων Google Stock Price. Γρήγορα γίνεται εμφανές πως η Softplus δεν κατάφερε να προσαρμοστεί στα δεδομένα εκπαίδευσης. Την καλύτερη απόδοση πετυχαίνει η Tanh και η Penalized Tanh με ΜΑΣ 2.4. Όπως φαίνεται στο Σχήμα 4.46, η Tanh είναι η ταχύτερη με 359 δευτερόλεπτα αφήνοντας πολύ πίσω την Penalized Tanh η οποία χρειάστηκε 512.

Για το RNN2, στο Σχήμα 4.47 οι διακυμάνσεις είναι μικρότερες. Πρώτες έρχονται οι Linear με τη PReLU με ΜΑΣ 1. Χρονικά στο Σχήμα 4.48, επικρατεί με ευκολία η Linear με 95 δευτερόλεπτα, ενώ η PReLU με 127 μένει μακριά από τον ανταγωνισμό.

Σχήμα 4.45: RNN1 Google Stock ΜΑΣ



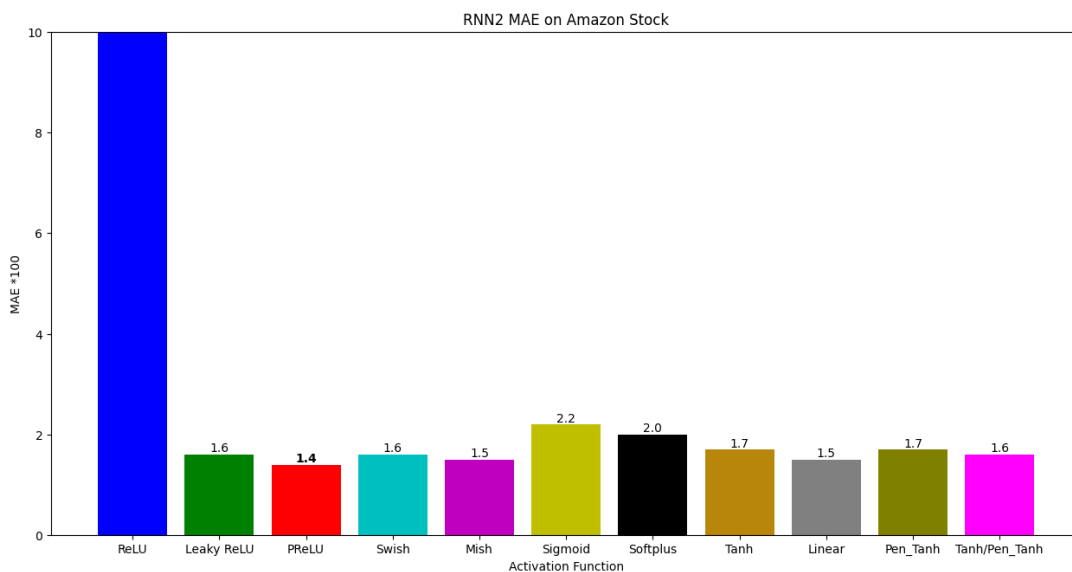
Σχήμα 4.46: RNN1 Google Stock Χρονομέτρηση



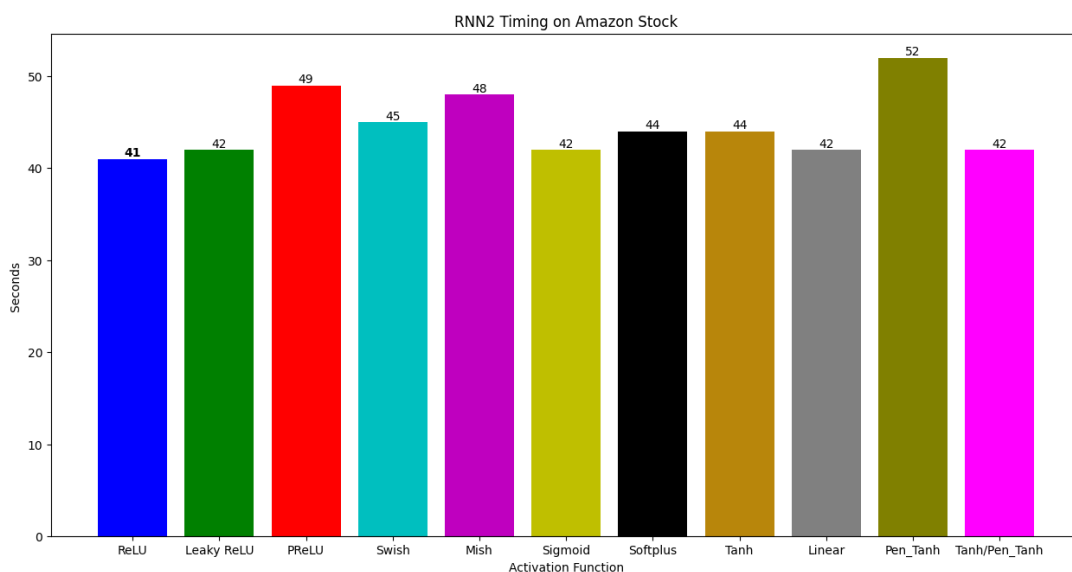
4.4.2 Amazon Stock

Για το RNN1, στο σύνολο δεδομένων Amazon Stock Price παρατηρείται παρόμοιο μοτίβο με το σύνολο δεδομένων που αφορούσε μετοχές της Google. Από το Σχήμα 4.49 φαίνεται πως η Softplus απέτυχε να προσαρμοστεί στα δεδομένα, ενώ

Σχήμα 4.47: RNN2 Google Stock ΜΑΣ



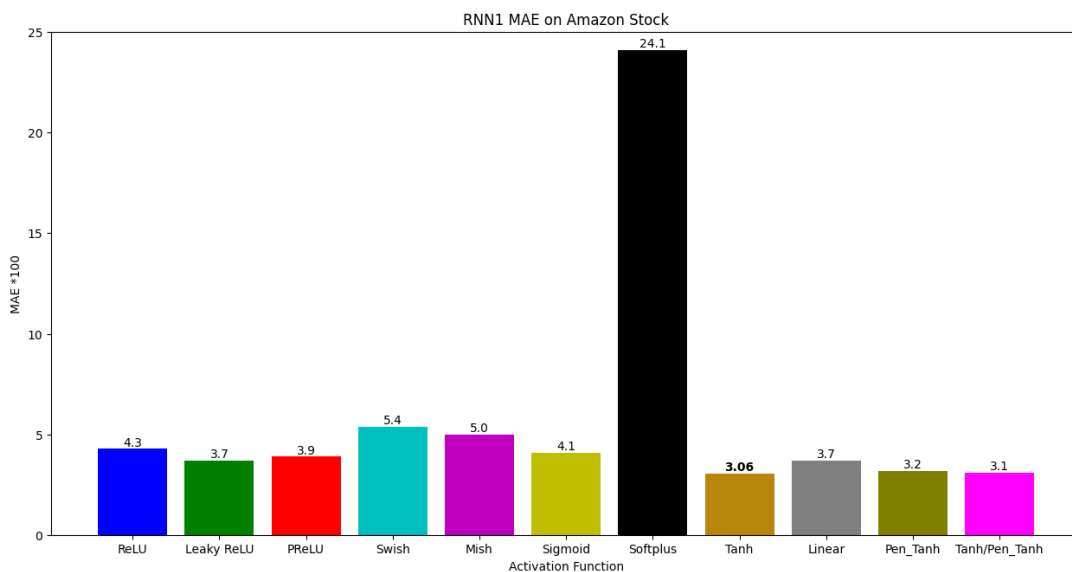
Σχήμα 4.48: RNN2 Google Stock Χρονομέτρηση



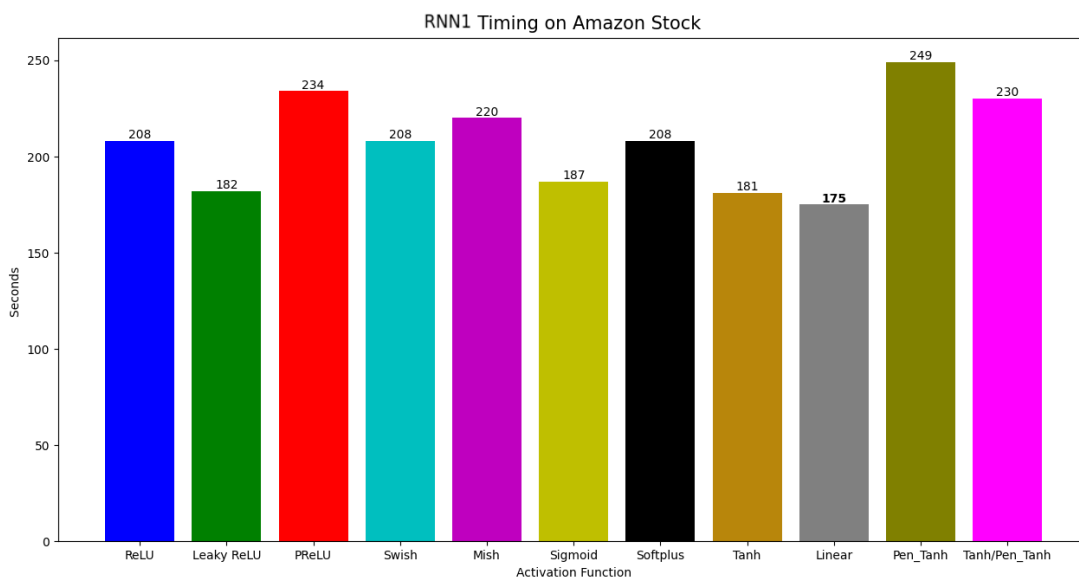
η Tanh με ΜΑΣ 3.06 έρχεται πρώτη. Στο Σχήμα 4.50, χρονικά η Linear είναι η ταχύτερη με 175 δευτερόλεπτα ενώ η Tanh ακολουθεί με 181.

Στο RNN2, αυτή τη φορά είναι η ReLU που απέτυχε να προσαρμοστεί στην εκπαίδευση, ενώ από το Σχήμα 4.51 φαίνεται πως το χαμηλότερο ΜΑΣ πετυχαίνει η PReLU με 1.4. Οι Linear και Mish ακολουθούν με ΜΑΣ 1.5.

Σχήμα 4.49: RNN1 Amazon Stock ΜΑΣ

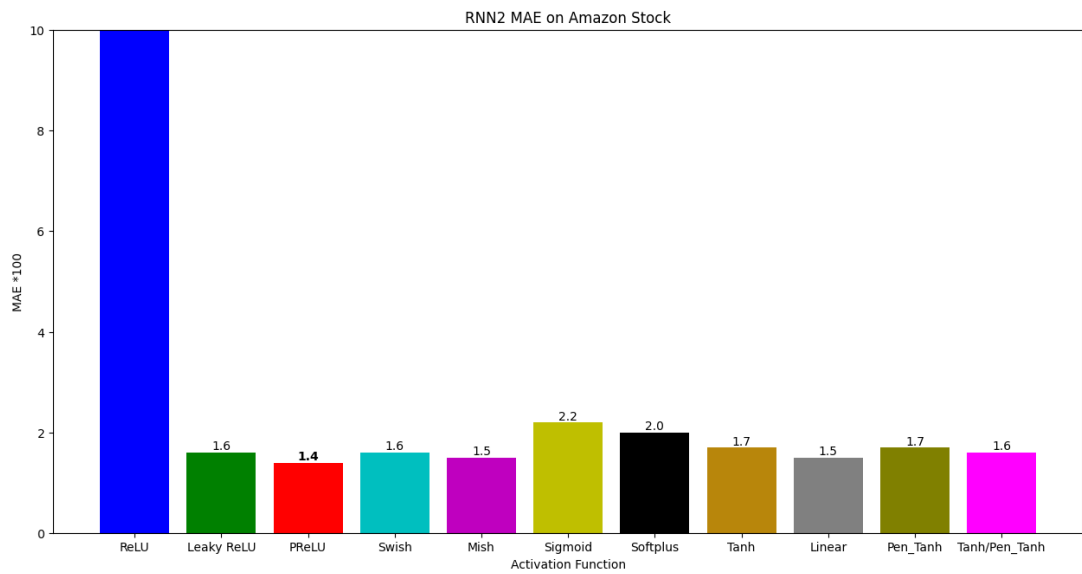


Σχήμα 4.50: RNN1 Amazon Stock Χρονομέτρηση



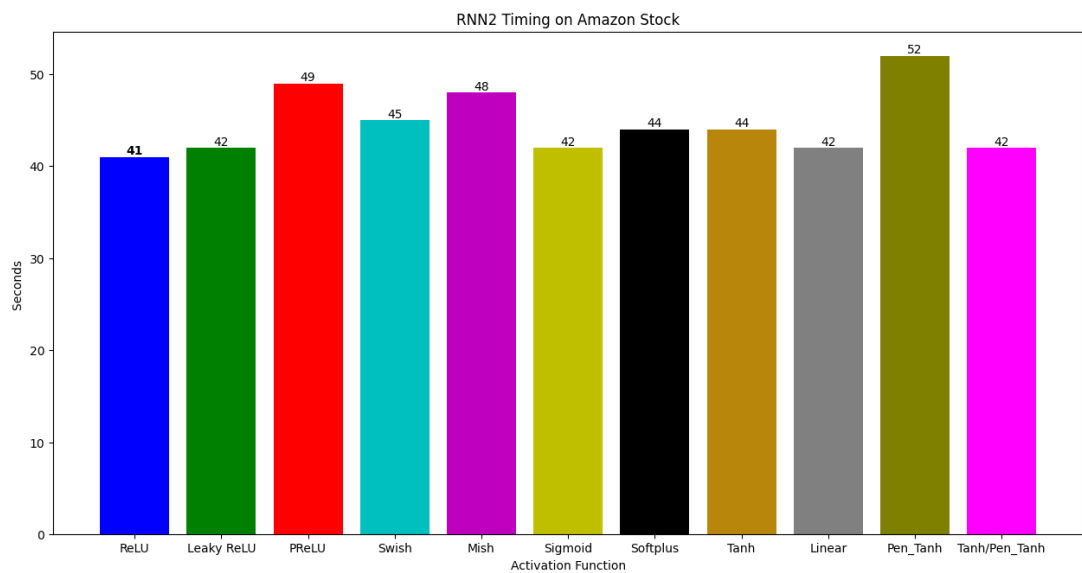
Βλέποντας τους χρόνους εκπαίδευσης στο Σχήμα 4.52 παρατηρείται πως η ReLU έρχεται πρώτη με 41 δευτερόλεπτα. Καθώς δεν κατάφερε να προσαρμοστεί στην εκπαίδευση θα δώσουμε προσοχή στις αμέσως καλύτερες οι οποίες με 1 δευτερόλεπτο διαφορά είναι οι Linear, Sigmoid και ο συνδιασμός Tanh/Penalized Tanh. Η καλύτερη σε ΜΑΣ PReLU φαίνεται για ακόμη μια φορά πως χρονικά βρίσκεται

Σχήμα 4.51: RNN2 Amazon Stock ΜΑΣ



αρκετά πίσω καθώς χρειάστηκε 49 δευτερόλεπτα.

Σχήμα 4.52: RNN2 Amazon Stock Χρονομέτρηση



4.5 Συμπεράσματα Πειραμάτων

Σε αυτή την ενότητα, αναλύονται τα αποτελέσματα που συλλέχθηκαν από τα πειράματα και εξάγονται συμπεράσματα με βάση αυτά. Καθώς είναι σαφές βλέποντας και μόνο τα αποτελέσματα πως δεν υπάρχει μια ιδανική λύση σε κάθε περίπτωση, επιλέχθηκε διαφορετική προσέγγιση. Πρωταρχικός στόχος είναι να αναδειχθούν συναρτήσεις που παράγουν σταθερά κορυφαία αποτελέσματα χωρίς να είναι ευαίσθητες στις μεταβολές των μοντέλων ή των δεδομένων. Έτσι έγινε η εξής ανάλυση: Για κάθε ένα εκ των τεσσάρων προβλημάτων (Αναγνώριση εικόνας, Εξόρυξη γνώμης, Πρόβλεψη σε δεδομένα πινάκων, Πρόβλεψη σε δεδομένα χρονοσειρών) θα υπολογίστηκε ο μέσος όρος ακρίβειας που πέτυχε η κάθε συνάρτηση σε κάθε σύνολο δεδομένων ξεχωριστά. Έτσι αν πχ η ReLU για το Cifar10 πέτυχε ποσοστά ακρίβειας σε Alexnet, VGG16 και CNN3 80%, 89% και 86% αντίστοιχα, ο μέσος όρος της για το Cifar-10 είναι 85%. Με αυτό τον τρόπο αναδεικνύονται οι έξι καλύτερες συναρτήσεις σε κάθε σύνολο δεδομένων. Από αυτές, μόνο οι συναρτήσεις που βρίσκονται στο top έξι κάθε συνόλου δεδομένων που αφορά το συγκεκριμένο πρόβλημα θεωρούνται ικανοποιητικά σταθερές. Έτσι αν πχ η ReLU στην Αναγνώριση εικόνας έρθει πρώτη κατά μέσο όρο στο Cifar-10 αλλά στο Cifar-100 έρθει έβδομη δε θα την αναλυθεί περαιτέρω. Με αυτή τη μέθοδο φιλτράρονται οι σταθερές συναρτήσεις. Στη συνέχεια, για τις συναρτήσεις που θα μείνουν θα αναλυθούν οι χρονικές τους επιδόσεις. Θα καταταχθούν σε κάθε σύνολο δεδομένων ξεχωριστά οι επιδόσεις τους και στο τέλος υπολογίζεται η μέση χρονική κατάταξη για κάθε συνάρτηση.

4.5.1 Αναγνώριση Εικόνας

Παρακάτω ακολουθεί ο πίνακας 4.1 με τις έξι συναρτήσεις που πέτυχαν τους κορυφαίους μέσους όρους ανά σύνολο δεδομένων.

Πίνακας 4.1: Αναγνώριση Εικόνας - Top6

Συνάρτηση	Cifar10 %	Cifar100 %	Vegetables %
Leaky ReLU	80.42	48.54	93.46
PReLU	80.14	47.3	94
Softplus	79.7	44.46	95.62
Swish	79	44.1	94.56
Mish	78.43	45.35	94.9
Hybrid	79.7	48.74	95

Παρατηρείται πως οι έξι συναρτήσεις είναι κοινές σε όλα τα σύνολα δεδομένων ωστόσο με διαφορετική κατάταξη. Στον πίνακα 4.2 φαίνεται η μέση κατάταξη της κάθε συνάρτησης -με βάση τα ποσοστά ακρίβειας- σε όλα τα σύνολα δεδομένων.

Πίνακας 4.2: Αναγνώριση Εικόνας: Top6 Κατάταξη Ακρίβειας

Συνάρτηση	Hybrid	Leaky ReLU	Softplus	PReLU	Swish	Mish
Μέση κατάταξη	2.3	3	3	3.3	4.7	4.7

Στη συνέχεια στον πίνακα 4.3 παρουσιάζονται οι αντίστοιχοι πίνακες που αφορούν τις χρονικές επιδόσεις των έξι αυτών συναρτήσεων. Με αυτά τα δεδομένα στον πίνακα 4.4 υπολογίζεται η μέση χρονική κατάταξη της κάθε συνάρτησης.

Πίνακας 4.3: Αναγνώριση Εικόνας: Χρονικές Επιδόσεις Top6

Συνάρτηση	Cifar10 %	Cifar100 %	Vegetables %
Leaky ReLU	777	596	494
PReLU	978	767	500
Softplus	801	600	501
Swish	841	678	490
Mish	875	694	496
Hybrid	869	592	525

Πίνακας 4.4: Αναγνώριση Εικόνας: Top6 Χρονική Κατάταξη

Συνάρτηση	Leaky ReLU	Swish	Softplus	Hybrid	Mish	PReLU
Μέση κατάταξη	1.7	2.6	3	3.3	4.3	5.3

Από την παραπάνω ανάλυση των αποτελεσμάτων βγαίνουν τα εξής συμπεράσματα σχετικά με τις επιδόσεις των συναρτήσεων στην αναγνώριση εικόνας:

- Η ReLU έδειξε να μην εκπαιδεύεται στα δύο από τα τρία σύνολα δεδομένων του CNN3. Το πρόβλημα δε φάνηκε να αφορά τα συγκεκριμένα σύνολα δεδομένων καθώς σε Alexnet και VGG16 έδειξε ικανοποιητικά αποτελέσματα. Παρόλο που η ReLU δε πέρασε το test σταθερότητας, δε πρέπει να πάψει να συμπεριλαμβάνεται στην γκάμα των επιλογών αλλά να χρησιμοποιείται με προσοχή, καθώς η μη ύπαρξη παραγώγου σε αρνητικές τιμές εισόδου μπορεί να δημιουργήσει προβλήματα σε δίκτυα με βαθιά δομή.
- Από την άλλη η PReLU ενώ πετυχαίνει κορυφαία ποσοστά ακρίβειας, δείχνει πως της στοιχίζει χρονικά η εκπαιδευσιμη παράμετρος a καθώς σχεδόν σε

κάθε μέτρηση βρέθηκε τελευταία σε ταχύτητα. Εάν ζητούμενο είναι μόνο η ακρίβεια τότε η PReLU αποτελεί μια καλή λύση.

- Οι Swish και Mish δείχνουν σταθερά σε κάθε μέτρηση δύο αξιόπιστες επιλογές. Μπορεί να μην πετυχαίνουν την καλύτερη επίδοση -αν και συμβαίνει σε μερικές περιπτώσεις (VGG16/Vegetable)- ωστόσο συνήθως δεν απέχουν πολύ από την κορυφή. Ενώ όμως παρέχουν σταθερότητα στην ακρίβεια, οι χρόνοι τους φαίνονται αρκετά ασταθείς με τις περισσότερες φορές να κλίνουν προς τις τελευταίες θέσεις. Έτσι αν η ταχύτητα είναι κύριο ζητούμενο συνιστάται να αποφεύγονται.
- Η Softplus έδειξε σταθερά κορυφαία αποτελέσματα με μέση κατάταξη χρόνου και ακρίβειας 3. Σίγουρα αποτελεί μια επιλογή που αξίζει να δοκιμαστεί καθώς ορισμένες φορές πέτυχε την καλύτερη επίδοση συνολικά. Ωστόσο πρέπει να ληφθεί υπ όψη το γεγονός ότι είναι λιγότερο σταθερή χρονικά από άλλες συναρτήσεις όπως η Leaky ReLU.
- Είναι σαφές πως η Leaky ReLU με την ακρίβεια που πετυχαίνει αλλά και την ταχύτητα της στην εκπαίδευση είναι η πρώτη επιλογή που θα πρέπει να δοκιμαστεί σε προβλήματα αναγνώρισης εικόνας.
- Κλείνοντας πρέπει να γίνει αναφορά στις υβριδικές λύσεις. Φάνηκε πως ο πειραματισμός με δύο συναρτήσεις ενεργοποίησης παράγει θετικά αποτελέσματα. Μπορεί η βελτίωση να μην είναι θεαματική, ωστόσο δε παύει να αποτελεί σε πολλές περιπτώσεις βελτίωση της προσέγγισης όπου μια μοναδική συνάρτηση εφαρμόζεται σε όλα τα επίπεδα.
- Στο πρόβλημα της αναγνώρισης εικόνας φάνηκε πως οι υβριδικές λύσεις δίνουν κατά μέσο όρο τα καλύτερα ποσοστά ακρίβειας χωρίς να μένουν πίσω χρονικά.
- Εάν υπάρχει χρόνος για δοκιμές, συστήνεται ο πειραματισμός έως ότου βρεθούν οι καλύτερες επιλογές για ένα υβριδικό μοντέλο. Ωστόσο η επιλογή της Leaky ReLU φαίνεται να είναι το ιδανικό σημείο εκκίνησης και ιδιαίτερα εάν η ταχύτητα του μοντέλου έχει πρωταρχική σημασία είναι πιθανό να μη βρεθεί κάποιος ταχύτερος συνδυασμός.

4.5.2 Εξόρυξη Γνώμης

Για την εξόρυξη γνώμης, σχετικά με το σύνολο δεδομένων Movie-Reviews, θα ληφθούν υπ όψη οι επιδόσεις ακρίβειας των συναρτήσεων μόνο στο μοντέλο TextCNN καθώς όπως φαίνεται στο Σχήμα4.23, στο LSTM1 όλες οι συναρτήσεις συνέκλιναν στο ίδιο ποσοστό.

Στον Πίνακα4.5 ακολουθούν οι έξι συναρτήσεις με τα υψηλότερα ποσοστά ακρίβειας στα δύο σύνολα δεδομένων.

Πίνακας 4.5: Εξόρυξη Γνώμης - Top6

Συνάρτηση	Movie-Review %	20-Netgroup %
Penalized Tanh	54.73	71.42
Linear	53.85	–
Mish	54.1	–
Swish	53.8	70.63
Hybrid	53.3	70.66
ReLU	52.87	70.41
PReLU	–	48.74
Leaky ReLU	–	70.6

Φαίνεται αυτή τη φορά πως η τομή των δύο συνόλων δεδομένων είναι οι Penalized Tanh, Swish, Hybrid, ReLU. Ο Πίνακας4.6 αφορά τη μέση κατάταξη των συναρτήσεων αυτών με βάση την ακρίβεια.

Πίνακας 4.6: Εξόρυξη Γνώμης: Κατάταξη Ακρίβειας

Συνάρτηση	Penalized Tanh	Hybrid	Swish	ReLU
Μέση κατάταξη	1	2.5	2.5	4

Στον Πίνακα4.7 φαίνεται η ταχύτητα εκπαίδευσης των τεσσάρων αυτών συναρτήσεων, και με βάση αυτά τα στοιχεία στον Πίνακα4.8 βλέπουμε τη μέση κατάταξη με βάση τους χρόνους.

Πίνακας 4.7: Εξόρυξη Γνώμης - Χρονικές Επιδόσεις

Συνάρτηση	Movie-Review	20-Netgroup
Penalized Tanh	398	1,038
Hybrid	404	1,086
Swish	415	1,126
ReLU	405	1,084

Από την ανάλυση των αποτελεσμάτων διακρίνονται τα εξής:

Πίνακας 4.8: Εξόρυξη Γνώμης: Χρονική Κατάταξη

Συνάρτηση	Penalized Tanh	ReLU	Hybrid	Swish
Μέση κατάταξη	1	2.5	2.5	4

- Στο συγκεκριμένο πρόβλημα τα πράγματα είναι πιο ξεκάθαρα ως προς την καλύτερη επιλογή καθώς η Penalized Tanh έδειξε πως έρχεται πρώτη σε όλες τις κατηγορίες.
- Δεν φάνηκαν -όπως στο πρόβλημα της αναγνώρισης εικόνας- ενδείξεις για το ποιες συναρτήσεις συνιστάται να αποφεύγονται. Η διακύμανση των αποτελεσμάτων είναι μικρή ως εκ τούτου δε δικαιολογείται να αποτρέψουμε τη δοκιμή καμίας συνάρτησης ενεργοποίησης.
- Η Swish, και σε αυτό το πρόβλημα είναι μια εκ των κορυφαίων λύσεων όσο αφορά την ακρίβεια. Αυτό την καθιστά τη μοναδική συνάρτηση που βρίσκεται στην κορυφή για τα δυο προβλήματα που αναλύθηκαν μέχρι στιγμής. Ωστόσο δεν πρέπει να ξεχάμε πως χρονικά δεν τα πηγαίνει εξίσου καλά.
- Το Hybrid μοντέλο δείχνει να παράγει καλά αποτελέσματα ωστόσο έδειξε υποδεέστερο από την Penalized Tanh. Έτσι, φαίνεται πως ο πειραματισμός που απαιτείται για να βρεθεί ο βέλτιστος συνδυασμός δεν έχει το ανάλογο αντίκρισμα.

4.5.3 Δεδομένα σε Μορφή Πίνακα

Στον Πίνακα 4.9 ακολουθούν τα αποτελέσματα των συναρτήσεων στα τρία σύνολα δεδομένων που χρησιμοποιήθηκαν. Παρατηρείται πως η τομή των τριών συνόλων δεδομένων είναι οι συναρτήσεις Swish, Mish και PReLU.

Στον Πίνακα 4.10 φαίνεται η μέση κατάταξη ακρίβειας των συναρτήσεων αυτών. Για μια ακόμη φορά η Swish βρίσκεται στις σταθερότερες με μέση κατάταξη θέσης στα τρία σύνολα δεδομένων 2.3 .

Στον Πίνακα 4.11 με λεπτομέρεια παρουσιάζεται η χρονική απόδοση των τριών συναρτήσεων και στον Πίνακα 4.12 η μέση χρονική τους κατάταξη. Η Swish είναι η ταχύτερη, με τη Mish να ακολουθεί ενώ η PReLU και πάλι αναδεικνύεται ουραγός σε αυτή τη κατηγορία.

Πίνακας 4.9: Δεδομένα σε Μορφή Πίνακα - Top6

Συνάρτηση	Churn Modeling (R2)	House Price (R2)	Hotel Reviews (R2)
Swish	86.69	56.84	43.19
Leaky ReLU	86.62	-	43.04
Mish	86.52	56.94	43.03
Hybrid	86.5	56.96	-
PReLU	86.35	56.9	43.09
Penalized Tanh	86.27	-	-
Linear	-	57.05	-
Softplus	-	56.68	42.81
ReLU	-	-	42.93

Πίνακας 4.10: Δεδομένα σε Μορφή Πίνακα - Κατάταξη Ακρίβειας

Συνάρτηση	Swish	Mish	PReLU
Μέση κατάταξη	2.3	3.3	3.6

Πίνακας 4.11: Δεδομένα σε Μορφή Πίνακα - Χρονικές Επιδόσεις

Συνάρτηση	Churn Modeling (sec)	House Price (sec)	Hotel Reviews (sec)
Swish	38	41	467
Mish	38	42	464
PReLU	44	49	511

Πίνακας 4.12: Δεδομένα σε Μορφή Πίνακα - Χρονικές Επιδόσεις

Συνάρτηση	Swish	Mish	PReLU
Μέση κατάταξη	1	2	3

Με βάση τη παραπάνω ανάλυση βγαίνουν τα παρακάτω συμπεράσματα:

- Από το σύνολο των αποτελεσμάτων στο συγκεκριμένο πρόβλημα, διακρίνονται δύο κατηγορίες συναρτήσεων με βάση την απόδοσή τους. Αυτές που κατάφεραν να αντεπεξεχθούν στα προβλήματα παλινδρόμησης (πρόβλεψη τιμής σπιτιού και πρόβλεψη κριτικής σε ξενοδοχείο) και αυτές που δεν κατάφεραν.
- Στην δεύτερη κατηγορία ανήκουν οι συναρτήσεις Sigmoid, Tanh και Penalized Tanh. Κοινό τους χαρακτηριστικό είναι η χαμηλή έξοδος, με τιμές κοντά στο μηδέν και ανώτατο όριο το ένα. Προτείνεται να μη χρησιμοποιούνται σε προβλήματα παλινδρόμησης συναρτήσεις με παρόμοια χαρακτηριστικά. Ωστόσο στο στη δυαδική ταξινόμηση δε φάνηκε να υστερούν σημαντικά, συγκριτικά με άλλες συναρτήσεις.
- Σχετικά με τις συναρτήσεις που απέδωσαν στην εκπαίδευση, παρατηρείται πως οι διαφορές μεταξύ τους είναι αρκετά μικρές.
- Φαίνεται πως η Leaky ReLU ενώ δεν πέρασε τον έλεγχο σταθερότητας αφού βρέθηκε έβδομη στο House Prediction σύνολο δεδομένων (μόλις 0.01 πίσω απ τη Softplus) αποτελεί μια αξιόπιστη λύση που δεν πρέπει να παραλείπεται απ την γκάμα των επιλογών καθώς παραμένει πολύ κοντά στα κορυφαία ποσοστά ενώ παράλληλα πετυχαίνει ικανοποιητικές χρονικές επιδόσεις.
- Το ίδιο ισχύει και για τη ReLU. Ωστόσο σε κάθε μέτρηση πετυχαίνει παρόμοια αποτελέσματα με τη Leaky ReLU αλλά ελαφρώς κατώτερα. Έτσι, ενώ δεν αποτελεί μια κακή επιλογή, επισκιάζεται από τη Leaky ReLU χωρίς να φανεί κάποια περίπτωση που δικαιολογεί τη προτίμηση της.
- Η PReLU βλέπουμε για μια ακόμη φορά πως αποδεικνύεται σταθερή και ακριβής αλλά αρκετά αργή. Δε προτείνεται αν η ταχύτητα είναι βασικό ζητούμενο καθώς υπάρχουν λύσεις εξίσου ακριβείς και σταθερές αλλά κατά πολύ ταχύτερες.
- Οι Swish και Mish και πάλι αποτελούν εξαιρετικές επιλογές. Συγκεκριμένα για στη Swish -η οποία αναδείχθηκε πρώτη στο τεστ σταθερότητας- αναγνωρίζεται πως είναι η μόνη συνάρτηση με σταθερή παρουσία στη κορυφή

και στα τρία προβλήματα που εξετάστηκαν μέχρι στιγμής (αναγνώριση εικόνας, εξόρυξη γνώμης, δεδομένα σε πίνακες). Προτείνεται πάντα για δοκιμή αλλά πρέπει να λαμβάνεται υπόψη πως υπάρχουν ταχύτερες επιλογές που δίνουν εξίσου ικανοποιητικά αποτελέσματα, όπως η Leaky ReLU και η Softplus.

- Η Linear έδειξε σε ένα εκ των τριών συνόλων δεδομένων που χρησιμοποιήθηκε (House Price) πως είναι η καλύτερη επιλογή τόσο σε ακρίβεια όσο και σε ταχύτητα. Ωστόσο δεν φάνηκε σταθερή. Συστήνεται ο πειραματισμός κυρίως σε ένα υβριδικό μοντέλο με χρήση της στο επίπεδο εξόδου.
- Το υβριδικό μοντέλο δεν πέρασε το τεστ σταθερότητας ενώ δεν φάνηκε σε καμία περίπτωση να προσθέτει κάτι παραπάνω απ' ό,τι οι υπόλοιπες επιλογές. Ωστόσο πρέπει να αναφερθεί πως τα σύνολα δεδομένων σε πίνακα αφορούν προβλήματα με μεγάλη γκάμα χαρακτηριστικών, και περιέχουν τιμές που ποικίλουν σημαντικά. Έτσι, ενώ συστήνονται επιλογές όπως είτε η Swish και η Mish -αν κύριος στόχος είναι η ακρίβεια- είτε η Leaky ReLU και η Softplus -αν απαιτείται και ταχύτητα- θα πρέπει να έχουμε κατά νου πως απαιτείται πειραματισμός χωρίς να αποκλείεται μια υβριδική λύση να αποτελεί την καλύτερη επιλογή.

4.5.4 Δεδομένα Χρονοσειρών

Σχετικά με τα δεδομένα χρονοσειρών φαίνεται πως δεν υπάρχει διαφορά στα δύο σύνολα δεδομένων όσο αφορά την τη καλύτερη εξάδα 4.13, ενώ από τη κατάταξη 4.14 είναι ξεκάθαρο πως διαμορφώνεται ένα αποτέλεσμα αρκετά διαφορετικό απ' ό,τι στα τρία παραπάνω προβλήματα. Εδώ Tanh και Penalized Tanh πρωταγωνιστούν ενώ η Linear δείχνει να τα καταφέρνει καλύτερα στο Google Stock απ' ό,τι στο Amazon Stock.

Πίνακας 4.13: Δεδομένα Χρονοσειρών - Top6

Συνάρτηση	Google Stock	Amazon Stock
Hybrid	3.6	4.7
Penalized Tanh	3.6	4.9
Tanh	3.7	4.76
Linear	3.7	5.2
Leaky ReLU	4.0	5.3
PReLU	4.1	5.3

Πίνακας 4.14: Δεδομένα Χρονοσειρών - Κατάταξη Ακρίβειας

Συνάρτηση	Linear	Tanh	Leaky ReLU	Hybrid	Penalized Tanh	PReLU
Μέση κατάταξη	3	2	4	1	2	4.5

Βλέποντας τους χρόνους 4.15 αλλά και την κατάταξη με βάση αυτούς 4.16, η Linear αναδεικνύεται εύκολα η ταχύτερη και στα δύο σύνολα δεδομένων με τις Leaky ReLU και Tanh να ακολουθούν.

Πίνακας 4.15: Δεδομένα σε Χρονοσειρών - Χρονικές Επιδόσεις

Συνάρτηση	Google Stock (sec)	Amazon Stock (sec)
Linear	228	108
Leaky ReLU	242	112
Tanh	256	112.5
Penalized Tanh	317	150
PReLU	318	141
Hybrid	323	136

Πίνακας 4.16: Δεδομένα Χρονοσειρών - Χρονική Κατάταξη

Συνάρτηση	Linear	Tanh	Leaky ReLU	Hybrid	Penalized Tanh	PReLU
Μέση κατάταξη	1	3	2	5	5	5

Με βάση τη παραπάνω ανάλυση των αποτελεσμάτων για τα προβλήματα χρονοσειρών παρατηρούνται τα εξής:

- Η ReLU έδειξε και πάλι σε μια περίπτωση (RNN2 - Amazon Stock) να μην ανταπεξέρχεται στην εκπαίδευση. Ομοίως η Softplus παρουσίασε το ίδιο πρόβλημα στο (RNN1 - Amazon Stock). Αμφότερες κρίνονται ως ασταθείς και προτείνεται να αποφεύγονται για δεδομένα χρονοσειρών.
- Η Leaky ReLU βρίσκεται στη καλύτερη εξάδα ωστόσο απέχει αρκετά απ την κορυφή χωρίς να δείχνει τέτοια υπεροχή στην ταχύτητα που θα δικαιολογούσε τη χρήση της. Κάτι παρόμοιο ισχύει και για τη PReLU η οποία ως συνήθως έδειξε πολύ αργή στην εκπαίδευση.
- Η Penalized Tanh είναι μια κορυφαία επιλογή όντας πρώτη σε ακρίβεια στο Google Stock και τρίτη στο Amazon Stock. Ωστόσο μεγάλο της μειονέκτημα φάνηκε ο απαιτούμενος χρόνος εκπαίδευσης φέροντας την στην τελευταία θέση στη μέση κατάταξη.

-
- Από τη συνολική εικόνα των αποτελεσμάτων φαίνεται πως η Tanh είναι η πρώτη επιλογή που πρέπει να δοκιμάζεται. Όσο αφορά το RNN1 που περιλαμβάνει κόμβους LSTM είναι η καλύτερη επιλογή τόσο σε ακρίβεια όσο σε ταχύτητα. Στο RNN2 με κόμβους GRU -χωρίς να έχει δείξει απογοητευτικά αποτελέσματα- υπάρχουν και καλύτερες συναρτήσεις.
 - Η Linear στο RNN2 αποτελεί την καλύτερη επιλογή καθώς μαζί με την PReLU πέτυχαν τις καλύτερες επιδόσεις ακρίβειας. Ωστόσο η εξαιρετική της ταχύτητα -και στα δύο μοντέλα- την κάνει πολύ πιο ελκυστική.
 - Το υβριδικό μοντέλο φάνηκε να κατατάσσεται συνολικά πρώτο όσο αφορά την ακρίβεια. Ωστόσο η χαμηλή ταχύτητα εκπαίδευσης δε του επιτρέπει να προκριθεί ως βέλτιστη επιλογή. Κλείνοντας επισημαίνεται το εξής. Στα προβλήματα χρονοσειρών συναντώνται η ιδιαιτερότητα ότι υπάρχουν πολλές αρχιτεκτονικές κόμβων που τα αντιμετωπίζουν, ενώ εδώ εξετάστηκαν μόνο δύο (LSTM, GRU). Καθώς παρατηρήθηκαν σημαντικές διαφορές μεταξύ τους, πρέπει να σημειωθεί πως τα αποτελέσματα δεν μπορούν εύκολα να γενικευτούν για όλες τις αρχιτεκτονικές και ενδεχομένως να διαφέρουν σε αρχιτεκτονικές που δεν εξετάστηκαν.

Κεφάλαιο 5

Συμπεράσματα

Στη παρούσα διπλωματική αναλύθηκαν οι συναρτήσεις ενεργοποίησης και ο ρόλος τους στα τεχνητά νευρωνικά δίκτυα. Μελετήθηκε η διαθέσιμη βιβλιογραφία και πραγματοποιήθηκαν πειράματα στα οποία συγκρίθηκαν μερικές από τις γνωστότερες συναρτήσεις σε διαφορετικά προβλήματα και σύνολα δεδομένων και εξετάστηκε η απόδοσή τους πάνω σε μια πληθώρα μοντέλων. Όπως φάνηκε, τα αποτελέσματα ποικίλουν ανά περίπτωση. Έτσι, δε μπορούμε να περάσουμε σε γενικές προτάσεις που ισχύουν καθολικά. Από τη μελέτη της βιβλιογραφίας, ως κοινή τομή προκρίθηκε η ReLU ως πρώτη και πιθανώς καλύτερη πρόταση. Ωστόσο από τα αποτελέσματα των πειραμάτων του τέταρτου κεφαλαίου είναι εμφανές πως τα πράγματα δεν είναι τόσο απλά. Στην παρακάτω λίστα γίνεται μια απόπειρα να ανακεφαλαίωση των όσων παρουσιάστηκαν για κάθε συνάρτηση ξεχωριστά.

- **ReLU:** Η ReLU φάνηκε στις περισσότερες περιπτώσεις ότι αποδίδει αρκετά ικανοποιητικά χωρίς να απέχει σημαντικά από την κορυφή. Συνδιάζοντας αυτό με τους χαμηλούς χρόνους εκπαίδευσης, φαίνεται μια συνάρτηση που θα αποτελούσε εξαιρετική επιλογή. Το μεγάλο της πρόβλημα έγκειται στην αστάθεια που έδειξε σε μερικά μοντέλα, μη καταφέροντας καθόλου να αντεπεξέλθει στα δεδομένα εκπαίδευσης. Συγκεκριμένα στο CNN3, σε δύο dataset η ακρίβεια της αναλογούσε σε τυχαίες προβλέψεις, ενώ στο RNN2 με τους κόμβους GRU δε κατάφερε καθόλου να προσεγγίσει της τιμές μέσου απόλυτου σφάλματος που έδωσαν οι άλλες συναρτήσεις. Το μη παραγωγίσιμο τμήμα της στις αρνητικές τιμές εισόδου φαίνεται πως μπορεί να προκαλέσει σημαντικά προβλήματα στην εκπαίδευση. Συνεπώς θα συνιστάται να γίνει πειραματισμός του μοντέλου πάνω στο οποίο θα χρησιμοποιηθεί και αν αποφευχθεί η χρήση

της σε σύνολα δεδομένων με αρνητικές συσχετίσεις.

- **Leaky ReLU:** Η Leaky ReLU από την άλλη φάνηκε εξαιρετική επιλογή στην πλειονότητα των προβλημάτων. Παρά την φαινομενικά μικρή διαφορά της από τη ReLU, είναι ξεκάθαρο πως η διαφορά κλίσης που έχουνε μεταξύ τους στις αρνητικές τιμές βοηθά τη Leaky ReLU να ξεπεράσει το πρόβλημα της προκατόχου της. Έδειξε εξαιρετικά ποσοστά ακρίβειας στις περισσότερες περιπτώσεις, ταχύτατους χρόνους εκπαίδευσης και από τις ικανοποιητικότερες σταθερότητες. Προτίνεται να δοκιμαστεί σε κάθε πρόβλημα, ακόμα και αν δε κατέχει πάντα τη πρωτιά.
- **PReLU:** Ιδιαίτερη περίπτωση, με τους δείκτες ακρίβειας και ταχύτητας να δίνουν αντικρουόμενα μηνύματα ως προς την καταλληλότητα χρήσης της. Έδειξε εξαιρετική ακρίβεια, με περιπτώσεις που αναδείχθηκε πρώτη ακόμα και σε προβλήματα χρονοσειρών, με αρχιτεκτονικές (RNN2/GRU - Amazon Stock) στις οποίες συνηθίζεται η χρήση συναρτήσεων με διαφορετικό προφίλ όπως η Tanh. Ωστόσο εύκολα διακρίνει κανείς πως είναι σταθερά ουραγός στη κατηγορία της ταχύτητας με σημαντικές αποστάσεις από την κορυφή. Συμπερασματικά λοιπόν, σημειώνεται πως η εκπαιδύσιμη παράμετρος για τη κλίση της συνάρτησης στις αρνητικές τιμές δεν αποδίδει ικανοποιητικά σε σχέση με τη χρονική επιβάρυνση που προσθέτει στο σύστημα. Ωστόσο προσδίδει σταθερότητα και προσαρμοστικότητα. Προτείνεται η χρήσης της μόνο στις περιπτώσεις όπου ο παράγοντας της ταχύτητας έχει χαμηλή προτεραιότητα ενώ διαφορετικά συστήνεται να αποφεύγεται.
- **Swish:** Η Swish κερδίζει τον τίτλο της σταθερότερης συνάρτησης καθώς σε τρία από τα τέσσερα προβλήματα που εξετάστηκαν βρισκόταν μέσα στις κορυφαίες συναρτήσεις. Τα ποσοστά ακρίβειας της είναι εξαιρετικά ενώ οι ταχύτητες της πως είναι ικανοποιητικές, χωρίς να είναι εξαιρετικές αλλά ούτε απαγορευτικές. Εάν ο χαμηλός χρόνος εκπαίδευσης δεν είναι το κύριο ζητούμενο τότε η χρήση της προτείνεται ανεπιφύλακτα εκτός από προβλήματα χρονοσειρών όπου οι επιδόσεις της έδειξαν αποτρεπτικές.
- **Mish:** Η Mish έδειξε αποτελέσματα πολύ όμοια με αυτά της Swish, όπως άλλωστε είναι αναμενόμενο αν παρατηρήσει κανείς τις γραφικές παραστάσεις

αυτών των δύο συναρτήσεων. Σε σύγκριση μεταξύ τους, γενικά η Swish δείχνει ελαφρώς ταχύτερη και ακριβέστερη χωρίς ωστόσο να ξεχνάμε το πρόβλημα της εξόρυξης γνώμης καθώς και αρκετές περιπτώσεις αναγνώρισης εικόνας όπου η Mish έδειξε μεγαλύτερη ακρίβεια.

- **Sigmoid:** Η Sigmoid φάνηκε αρκετά ξεκάθαρα πως είναι μια συνάρτηση της οποίας η χρήση -στη πλειονότητα των περιπτώσεων- πρέπει να αποφεύγεται να εφαρμόζεται σε όλα επίπεδο ενός δικτύου. Σε αρκετές περιπτώσεις δεν κατάφερε να προσαρμοστεί καθόλου στα δεδομένα εκπαίδευσης ενώ σε αυτές που έδειξε κάποια προσαρμογή, παρέμεινε μια κλίμακα κάτω από τις περισσότερες συναρτήσεις. Το μόνο σενάριο που δικαιολογείται η χρήση της, είναι σε προβλήματα δυαδικής ταξινόμησης (Churn Modeling) όπου δε φάνηκε να υστερεί σημαντικά σε ακρίβεια ενώ έδειξε εξαιρετικές ταχύτητες εκπαίδευσης (όπως και σε όλα σχεδόν τα προβλήματα). Έτσι αν η ταχύτητα είναι βασικό ζητούμενο μπορεί να δοκιμαστεί, αλλά με πιο βάσιμες προσδοκίες στα προβλήματα δυαδικής ταξινόμησης. Τέλος δε πρέπει να ξεχνάμε την περίπτωση των υβριδικών λύσεων όπου θα μπορούσε να αποτελεί μια εκ των επιλογών, ιδιαιτέρως στα επίπεδα που δεν απέχουν πολύ απ το επίπεδο εξόδου ώστε να αποφευχθεί το πρόβλημα των εξαφανιζόμενων κλίσεων.
- **Tanh:** Τα δεδομένα της Tanh έχουν μεγάλη ομοιότητα με αυτά της Sigmoid. Στις περισσότερες περιπτώσεις δε συστήνεται η χρήσης της, ωστόσο η βασική της διαφορά εντοπίζεται σε προβλήματα χρονοσειρών. Εκεί φάνηκε η καλύτερη λύση στο δίκτυο με LSTM κόμβους, τόσο σε ταχύτητα όσο και σε ακρίβεια. Έτσι είναι μια συνάρτηση η οποία δεν χρησιμοποιείται καθολικά αλλά έχει εξαιρετικές επιδόσεις σε συγκεκριμένο είδος προβλημάτων, όπου και προτείνεται ανεπιφύλακτα.
- **Penalized Tanh:** Η Penalized Tanh είναι η λιγότερο γνωστή συνάρτηση από αυτές που μελετήθηκαν. συμπεριλήφθηκε στη λίστα των δοκιμών καθώς έδειξε θετικά αποτελέσματα σε άλλες δημοσιεύσεις [10] που παρουσιάστηκε. Από τα πειράματα του Κεφαλαίου 4 φαίνεται πως πρόκειται για μια συνάρτηση που εφαρμόζεται ιδανικά σε συγκεκριμένα προβλήματα ενώ σε άλλα δε συνιστάται η χρήση της. Δε φαίνεται για παράδειγμα να δικαιολογείται η χρήση της σε

προβλήματα αναγνώρισης εικόνας ή δεδομένων σε πίνακα. Μπορεί στα προβλήματα αυτά να έδειξε πως ξεπερνά αρκετά την Tanh -της οποίας αποτελεί παραλλαγή- ωστόσο όχι επαρκώς ώστε να θεωρείται ικανοποιητική λύση. Ένα επίσης δεδομένο που παρατηρήθηκε είναι ότι γενικώς πρόκειται για αρκετά αργή συνάρτηση. Έτσι ενώ έδειξε αρκετά ικανοποιητικά αποτελέσματα ακρίβειας σε δεδομένα χρονοσειρών, οι χρονικές απαιτήσεις της δε τη καθιστούν ανταγωνιστική σε σχέση με συναρτήσεις όπως η Tanh και η Linear σε σχέση με τις οποίες χρειάζεται ως και +60% σε χρόνο εκπαίδευσης. Εκεί που τα δεδομένα ανατρέπονται είναι τα προβλήματα εξόρυξης γνώμης. Αναδείχθηκε πρώτη συγκεντρωτικά αλλά και σε κάθε μέτρηση ξεχωριστά ενώ οι χρόνοι της ήταν εξαιρετικοί με περιπτώσεις όπου ήταν και η ταχύτερη με σημαντική διαφορά. Συνεπώς, δε μπορούμε παρά να προτείνουμε της χρήση για προβλήματα εξόρυξης γνώμης.

- **Linear:** Η Linear είναι μια πολύ γνωστή συνάρτηση της οποίας ωστόσο η χρήση σπανίζει. Το κύριο χαρακτηριστικό που παρατηρήθηκε είναι η μεγάλη ταχύτητα της η οποία τη συνοδεύει σε όλες τις μετρήσεις. Σε προβλήματα αναγνώρισης εικόνας φάνηκε να στέκεται ένα επίπεδο πίσω από τις επιδόσεις των κορυφαίων συναρτήσεων. Αντίθετα στο σύνολο δεδομένων House Price ήρθε πρώτη σε R2 σκορ ενώ στα δεδομένα χρονοσειρών οι πολύ ικανοποιητικές της επιδόσεις σε συνδιασμό με τη ταχύτητα της την καθιστούν κορυφαία λύση. Εάν ο χαμηλός χρόνος εκπαίδευσης είναι βασικό ζητούμενο προτείνεται να δοκιμαστεί, μπορεί να αποτελέσει και επιλογή σε υβριδικές λύσεις ιδιαίτερα για προβλήματα παλινδρόμησης.

Κύριος στόχος των πειραμάτων ήταν να δοκιμαστούν οι συναρτήσεις ενεργοποίησης σε μεγάλο εύρος προβλημάτων ώστε να αναδειχθεί η σταθερότητα της κάθε μίας πέρα από την ακρίβεια και τη ταχύτητα. Καθώς οι αρχιτεκτονικές των μοντέλων είναι τόσες σε πλήθος που δε επιτρέπεται στα πλαίσια της διπλωματικής εργασίας να καλυφθούν όλες αλλά ούτε και να γίνει επαρκής εμβάθυνση για κάθε περίπτωση, κλείνοντας και ευχαριστώντας τους αναγνώστες, γίνεται η σχετική παράτρυνση σε όσους που βρήκαν ενδιαφέροντα και διαφωτιστικά τα ευρήματα που παρουσιάστηκαν να πραγματοποιήσουν τη δική τους έρευνα. Θα ήταν ιδιαίτερα χρήσιμο να διερευνηθούν σε μεγαλύτερο βάθος υβριδικές περιπτώσεις μοντέλων

που χρησιμοποιούν πάνω από μια συναρτήσεις ενεργοποίησης καθώς και να γίνουν προσπάθειες για προτάσεις νέων συναρτήσεων. Σε κάθε περίπτωση αποδίδονται ευχαριστίες σε όσους βοήθησαν στην ολοκλήρωση της παρούσας Διπλωματικής Εργασίας με ευχές και ελπίδες να παρέχει στους αναγνώστες γνώση χρήσιμη για τη συνέχεια του έργου τους.

Βιβλιογραφία

- [1] F. Kamalov, A. Nazir, M. Safaraliev, A. K. Cherukuri, and R. Zgheib, “Comparative analysis of activation functions in neural networks,” in *2021 28th IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*, pp. 1–6, 2021.
- [2] T. Szandała, “Review and comparison of commonly used activation functions for deep neural networks,” in *Bio-inspired Neurocomputing*, vol. 903, 2021.
- [3] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, “A comprehensive survey and performance analysis of activation functions in deep learning,” *CoRR*, vol. abs/2109.14545, 2021.
- [4] W. S. McCulloch and W. Pitts, “A logical calculus of ideas immanent in nervous activity,” *Bulletin of Mathematical Biophysics*, Vol. 5, 1943.
- [5] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain.,” *Cornell Aeronautical Laboratory, Psychological Review*, Vol 65, 1958.
- [6] M. Minsky and S. Papert, *Perceptrons*. Cambridge, MA: MIT Press, 1969.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, 2012.
- [8] S. Hochreiter, “The vanishing gradient problem during learning recurrent neural nets and problem solutions,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, pp. 107–116, 04 1998.
- [9] B. Xu, R. Huang, and M. Li, “Revise saturated activation functions,” 2016.
- [10] S. Eger, P. Youssef, and I. Gurevych, “Is it time to swish? comparing deep learning activation functions across NLP tasks,” vol. abs/1901.02671, 2019.
- [11] A. Salam, A. El Hibaoui, and A. Saif, “A comparison of activation functions in multilayer neural network for predicting the production and consumption of electricity power,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 1, p. 163, 2021.
- [12] B. Xu, N. Wang, T. Chen, and M. Li, “Empirical evaluation of rectified activations in convolutional network,” *arXiv preprint arXiv:1505.00853*, 2015.
- [13] R. Dasgupta, Y. S. Chowdhury, and S. Nanda, “Performance comparison of benchmark activation function relu, swish and mish for facial mask detection using convolutional neural network,” in *Intelligent Systems*, (Singapore), pp. 355–367, Springer Singapore, 2021.

-
- [14] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.
- [15] Y. Kim, “Convolutional neural networks for sentence classification,” *arXiv preprint arXiv:1408.5882*, 2014.