



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ ΚΑΙ ΑΝΑΛΥΤΙΚΗ**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

*Βαθιά Νευρωνικά Δίκτυα με Μηχανές Διανυσμάτων
Υποστήριξης στο Επίπεδο Εξόδου*

Όνομ/μο: Παναγιώτης Τζιμής

A.M.: ME 1616

Υπεύθυνος Καθηγητής: Επίκ. Καθ. Τελέλης Ορέστης

Περίληψη

Η παρούσα εργασία ασχολείται με την μελέτη της επίδοσης των νευρωνικών δικτύων με τη χρήση μηχανών διανυσμάτων υποστήριξης στο επίπεδο εξόδου. Η χρήση των μηχανών διανυσμάτων υποστήριξης στο επίπεδο εξόδου, επιτυγχάνονται καλύτερα ποσοστά ακρίβειας αφού γίνεται προσπάθεια εύρεσης ενός βέλτιστου ορίου απόφασης μεταξύ των δεδομένων. Η υλοποίηση των πειραμάτων πραγματοποιήθηκε σε νευρωνικά δίκτυα Multilayer Perceptron και σε συνελκτικά νευρωνικά δίκτυα. Χρησιμοποιήθηκαν τρία σύνολα δεδομένων όπως: MNIST, CIFAR-10 και CIFAR-100 και αποτελούνταν από πέντε ενδιάμεσα επίπεδα με διαφορετικό αριθμό κόμβων. Τα καλύτερα ποσοστά ακρίβειας που επιτυγχάνονται για το σύνολο δεδομένων MNIST με τη χρήση δύο επιπέδων, είναι για το MLP 0.9838 και με την χρήση του SVM ταξινομητή 0.9848. Για το σύνολο δεδομένων CIFAR-10 το καλύτερο ποσοστό ακρίβειας με τη χρήση τριών επιπέδων για το νευρωνικό δίκτυο MLP 0.454 και με την χρήση του SVM ταξινομητή 0.49.

Ακόμα, για το σύνολο δεδομένων CIFAR-100 το καλύτερο ποσοστό ακρίβειας που επιτυγχάνεται είναι με τη χρήση δύο ενδιάμεσων επιπέδων για το MLP 0.1857 και με την χρήση του SVM ταξινομητή 0.2201. Τέλος, εκτελέστηκαν πειράματα και σε αρχιτεκτονικές συνελκτικών δικτύων επιτυγχάνοντας ποσοστά σφάλματος για τον SVM ταξινομητή 2.39% και για το συνελκτικό δίκτυο 2.46%.

Abstract

In the present study, we examine the progress of the neural networks using the Support Vector Machines (SVM) in the output layer. SVM can provide accurate rates by determining the optimum decision boundary for data classification in two classes. The experiments were processed in two types of neural networks like Multilayer Perceptron (MLP) and convolutional neural network, using datasets MNIST, CIFAR-10 and CIFAR-100, within five hidden layers of numerous nodes. For the MNIST dataset, the accuracy that achieved based on two hidden layers were for the Multilayer Perceptron 0.9838 and for SVM 0.9848. For the CIFAR-10 dataset, the accuracy that achieved based on three hidden layers were for the MLP 0.454 and for SVM 0.49. Furthermore, for the CIFAR-100 dataset the accuracy that achieved based on two hidden layers were for the MLP 0.1857 and for SVM 0.2201. Finally, experiments were also conducted in the architecture of convolutional neural networks obtaining test errors (%) of 2.39% in the SVM and 2.46% in the convolutional network.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον κύριο Τελέλη Ορέστη, Επίκουρο Καθηγητή του τμήματος Ψηφιακών Συστημάτων του Πανεπιστημίου Πειραιώς για τη συνεχή καθοδήγηση και βοήθειά του, στη περάτωση της διπλωματικής αυτής εργασίας.

Πίνακας περιεχομένων

Κεφάλαιο 1 Εισαγωγή	1
1.1 Δομή εργασίας	1
1.2 Μηχανική Μάθηση: Σύνοψη τεχνικών	2
1.2.1 Μέθοδοι εκπαίδευσης	2
1.2.2 Γραμμική Παλινδρόμηση (Linear Regression).....	2
1.2.3 Λογιστική παλινδρόμηση (Logistic Regression)	3
1.2.4 Αλγόριθμος Βαθμωτής Καθόδου (Gradient Descent).....	4
1.3 Μηχανές Διανυσμάτων Υποστήριξης (SVM).....	6
1.3.1 Soft Margin	7
1.3.2 Διανύσματα Υποστήριξης και Συναρτήσεις Πυρήνα.....	8
1.4 Βασικές Αρχές Νευρωνικών Δικτύων	10
1.4.1 Συναρτήσεις Ενεργοποίησης.....	10
1.4.2 Αρχιτεκτονική νευρωνικών δικτύων	12
1.4.3 Μοντέλο Perceptron	13
1.4.4 Αλγόριθμος Οπίσθιας διάδοσης σφάλματος (Backpropagation)	15
1.4.5 Κριτήρια τερματισμού	16
Κεφάλαιο 2 Συνελκτικά Νευρωνικά Δίκτυα.....	18
2.1 Χαρακτηριστικά Συνελκτικών Δικτύων	19
2.1.1 Επίπεδο Συνέλιξης.....	19
2.1.2 Επίπεδο Υποδειγματοληψίας.....	22
2.1.3 Πλήρως συνδεδεμένα επίπεδα.....	24
2.1.4 Τοπικά συνδεδεμένα δίκτυα	24
2.2 Αλγόριθμος Backpropagation στα συνελκτικά δίκτυα.....	24
2.3 Γενικοποίηση Δικτύου	25
2.4 Τεχνικές επεξεργασίας δεδομένων εικόνας	25
Κεφάλαιο 3 Βιβλιογραφική Επισκόπηση	27
3.1 Βασικές εργασίες.....	27
3.2 Επιπλέον εργασίες.....	29
3.3 Συγκρίσεις συνόλων δεδομένων.....	30
3.2.1 MNIST	30
3.2.2 CIFAR-10	31
3.2.3 CIFAR-100	31
Κεφάλαιο 4 Πειραματική Αποτίμηση.....	32

4.1 Βιβλιοθήκη Keras.....	32
4.2 Πειράματα	32
4.2.1 Multilayer Perceptron (MLP).....	33
4.2.2 Συνελκτικά Δίκτυα με Μηχανές Διανυσμάτων Υποστήριξης.....	47
Κεφάλαιο 5 Συμπεράσματα	51
5.1 Μετέπειτα πειράματα και έρευνα	51
Βιβλιογραφία	1

Εικόνα 1 Αναπαράσταση Γραμμικής Παλινδρόμησης με εύρεση ευθείας γραμμής (κόκκινη) για τον καλύτερο κατανομημένο διαχωρισμό των δεδομένων.....	3
Εικόνα 2 Αναπαράσταση σιγμοειδής συνάρτησης	3
Εικόνα 3 Αναπαράσταση Gradient Descent	5
Εικόνα 4 Αναπαράσταση του ρυθμός μάθησης σε δύο περιπτώσεις.....	5
Εικόνα 5 Αναπαράσταση του SVM ταξινομητή, με το όριο απόφασης ως την μέγιστη απόσταση από τα δεδομένα των 2 κλάσεων	6
Εικόνα 6 Αναπαράσταση μη γραμμικώς διαχωρίσιμα δεδομένα.....	7
Εικόνα 7 Αναπαράσταση SoftMargin	7
Εικόνα 8 Δυσκό πρόβλημα.....	8
Εικόνα 9 Απεικόνιση συνάρτησης Φ σε χώρο μεγαλύτερων διαστάσεων	10
Εικόνα 10 Αναπαράσταση ταυτοτική συνάρτησης	11
Εικόνα 11 Αναπαράσταση συνάρτησης RELU	11
Εικόνα 12 Αναπαράσταση συνάρτησης tanh	12
Εικόνα 13 Αναπαράσταση εμπρόσθιας τροφοδότησης νευρωνικού δικτύου.....	13
Εικόνα 14 Αναπαράσταση ανατροφοδότησης νευρωνικού δικτύου	13
Εικόνα 15 Αναπαράσταση κόμβου με 7 κόμβους εισόδου και 3 κόμβους εξόδου	14
Εικόνα 16 Αναπαράσταση κόμβου (MLP) με 2 ενδιάμεσα επίπεδα	15
Εικόνα 17 Αναπαράσταση αλγορίθμου Backpropagation.....	16
Εικόνα 18 Αρχιτεκτονική συνελκτικού δικτύου.....	18
Εικόνα 19 Αναπαράσταση συνέλιξης	19
Εικόνα 20 Αναπαράσταση συνέλιξης σε εικόνα 5x5 για την θέση $\Gamma(0,0)$	20
Εικόνα 21 Αναπαράσταση συνέλιξης σε εικόνα 5x5 για την θέση $\Gamma(0,1)$	21
Εικόνα 22 Αναπαράσταση συνέλιξης για την τελική εικόνα.	22
Εικόνα 23 Διαδικασία εύρεσης τοπικού μεγίστου.	23
Εικόνα 24 Αναπαράσταση στοχαστικής υποδειγματοληψίας για χαρακτηριστικά που βρίσκονται σε διαφορετικές περιοχές με εμφάνιση διαφορετικών πιθανοτήτων ως επιλογή για το επόμενο επίπεδο.	23

Πίνακας 1 Ποσοστά ακρίβειας.....	28
Πίνακας 2 Αναπαράσταση χαμηλότερων ποσοστών σφάλματος με την χρήση διαφορετικών συναρτήσεων ενεργοποίησης.....	30
Πίνακας 3 Αναπαράσταση ποσοστών σφάλματος σε αρχιτεκτονικές που διαφέρουν στον αριθμό των ενδιάμεσων επιπέδων.....	31
Πίνακας 4 Ποσοστά ακρίβειας (MLP και ταξινομητή SVM) σε διαφορετικές αρχιτεκτονικές για το σύνολο δεδομένων MNIST	35
Πίνακας 5 Ποσοστά ακρίβειας (MLP και ταξινομητή SVM) σε διαφορετικό αριθμό εποχών και μέτρηση του χρόνου υλοποίησης(min/h) για το σύνολο δεδομένων MNIST	37
Πίνακας 6 Ποσοστά ακρίβειας (MLP και SVM ταξινομητή) σ για το σύνολο δεδομένων CIFAR-10	40
Πίνακας 7 Ποσοστά ακρίβειας (MLP και ταξινομητή SVM) σε διαφορετικό αριθμό εποχών και μέτρηση του χρόνου υλοποίησης(min/h) για το σύνολο δεδομένων CIFAR-10	41
Πίνακας 8 Ποσοστά ακρίβειας (MLP και SVM ταξινομητή) για το σύνολο δεδομένων CIFAR-100.....	44
Πίνακας 9 Ποσοστά ακρίβειας (MLP και ταξινομητή SVM) σε διαφορετικό αριθμό εποχών και μέτρηση του χρόνου υλοποίησης(min/h) για το σύνολο δεδομένων CIFAR-100	45
Πίνακας 10 Ποσοστά σφάλματος συνελκτικού δικτύου με ένα φίλτρο (πυρήνα) 5x5 στο επίπεδο συνέλιξης, ένα ενδιάμεσο επίπεδο και διαφορετικό αριθμό κόμβων για το σύνολο δεδομένων MNIST.....	48
Πίνακας 11 Αναπαράσταση ποσοστών σφάλματος (συνελκτικού δικτύου) με ένα φίλτρο (πυρήνα) 5x5 στο επίπεδο συνέλιξης, δύο ενδιάμεσα επίπεδα και διαφορετικό αριθμό κόμβων για το σύνολο δεδομένων MNIST.....	49
Πίνακας 12 Ποσοστά σφάλματος (συνελκτικού δικτύου) με ένα φίλτρο (πυρήνα) 5x5 στο επίπεδο συνέλιξης, πέντε ενδιάμεσα επίπεδα και διαφορετικό αριθμό κόμβων για το σύνολο δεδομένων MNIST.....	49
Πίνακας 13 Αναπαράσταση ποσοστών σφάλματος (συνελκτικού δικτύου και SVM ταξινομητή) με ένα φίλτρο (πυρήνα) 5x5 στο επίπεδο συνέλιξης, πέντε ενδιάμεσα επίπεδα και διαφορετικό αριθμό κόμβων για το σύνολο δεδομένων MNIST	50

Κεφάλαιο 1 Εισαγωγή

Η παρούσα διπλωματική εργασία ασχολείται με την μελέτη και την υλοποίηση νευρωνικών δικτύων για την επεξεργασία εικόνας τόσο σε θεωρητικό όσο και σε πρακτικό επίπεδο. Η αρχιτεκτονική που αναπτύσσεται είναι με τη χρήση της επιβλεπόμενης μάθησης (Supervised Learning) τόσο σε νευρωνικά δίκτυα Multilayer Perceptron όσο και σε συνελκτικά δίκτυα (Convolutional Networks). Στις αρχιτεκτονικές των μοντέλων που υλοποιήθηκαν γίνεται η χρήση ενός SVM ταξινομητή στο τελευταίο επίπεδο του δικτύου για τον καλύτερο διαχωρισμό των δεδομένων σε υπερεπίπεδα με αποτέλεσμα την καλύτερη ταξινόμηση των δεδομένων. Στο παρόν κεφάλαιο γίνεται αναφορά σε κάποιες βασικές αρχές της Μηχανικής μάθησης, στις μηχανές διανυσμάτων υποστήριξης (SVM) καθώς και κάποιες βασικές αρχές των νευρωνικών δικτύων.

1.1 Δομή εργασίας

Η δομή της εργασίας παρουσιάζεται στα εξής κεφάλαια:

Στο Κεφάλαιο 1 αναφέρονται κάποιες βασικές έννοιες Μηχανικής μάθησης όπως οι μέθοδοι εκπαίδευσης, η γραμμική παλινδρόμηση (Linear Regression), η λογιστική παλινδρόμηση (Logistic Regression) και ο αλγόριθμος βαθμωτής καθόδου (Gradient Descent). Στη συνέχεια γίνεται αναφορά στις μηχανές διανυσμάτων υποστήριξης (SVM) ως ένα από τα κύρια χαρακτηριστικά της αρχιτεκτονικής του νευρωνικού δικτύου καθώς και μια σύντομη περιγραφή στις συναρτήσεις πυρήνα (Kernel Methods). Τέλος, αναφέρονται κάποιες βασικές αρχές νευρωνικών δικτύων όπως οι συναρτήσεις ενεργοποίησης, η αρχιτεκτονική των νευρωνικών δικτύων, το μοντέλο Perceptron καθώς και ο αλγόριθμος οπίσθιας διάδοσης (Backpropagation).

Στο κεφάλαιο 2 περιγράφονται βασικές αρχές συνελκτικών δικτύων (Convolutional Networks) όπως το επίπεδο Συνέλιξης (Convolution layer), το επίπεδο υποδειγματοληψίας (Pooling layer) και τα πλήρως συνδεδεμένα επίπεδα (fully connected layers). Στη συνέχεια, γίνεται αναφορά στον αλγόριθμο Back-propagation καθώς και σε κάποιες τεχνικές γενίκευσης (generalization) και επεξεργασίας δεδομένων εικόνας.

Στο κεφάλαιο 3 περιγράφονται κάποιες από τις πιο σύγχρονες εργασίες που έχουν υλοποιηθεί με αντικείμενο την απόδοση είτε των νευρωνικών δικτύων (MLP) είτε των συνελκτικών δικτύων. Γίνεται χρήση τεχνικών που αφορούν στην βελτιστοποίηση της απόδοσης των νευρωνικών δικτύων. Ακόμα, γίνεται μία σύντομη αναφορά σε αποτελέσματα που έχουν προκύψει από εργασίες στα σύνολα δεδομένων MNIST, CIFAR-10 και CIFAR-100 τα οποία χρησιμοποιούμε και στην παρούσα εργασία.

Στο κεφάλαιο 4 παρουσιάζονται αναλυτικά όλα τα αποτελέσματα που προκύπτουν από την παρούσα εργασία. Τα αποτελέσματα αυτά, περιλαμβάνουν αναλυτικούς πίνακες απόδοσης του Multilayer Perceptron και του συνελκτικού δικτύου σε σύγκριση με τον SVM ταξινομητή στο τελευταίο επίπεδο. Ακόμα, υπάρχουν γραφικές παραστάσεις για την απόδοση των νευρωνικών δικτύων και του ταξινομητή SVM.

Στο κεφάλαιο 5 παρουσιάζονται συγκεντρωτικά τα συμπεράσματα που προκύπτουν από την μελέτη και την υλοποίηση αρχιτεκτονικών νευρωνικών δικτύων. Παράλληλα προκύπτουν και προτάσεις για μελλοντικές εργασίες.

1.2 Μηχανική Μάθηση: Σύνοψη τεχνικών

Στη συγκεκριμένη υποενότητα θα αναλυθούν κάποιες βασικές έννοιες της μηχανικής μάθησης όπως οι μέθοδοι εκπαίδευσης: μάθηση με επίβλεψη (Supervised Learning), μάθηση χωρίς επίβλεψη (Unsupervised Learning), η Γραμμική Παλινδρόμηση (Linear Regression) που χρησιμοποιείται για βασική πρόβλεψη τιμών, η Λογιστική Παλινδρόμηση (Logistic Regression) που χρησιμοποιείται και αυτή στην πρόβλεψη τιμών από ένα πλήθος ανεξάρτητων μεταβλητών προσεγγίζοντας διακριτή συνάρτηση. Ακόμα γίνεται αναφορά στον αλγόριθμο βαθμωτής καθόδου (Gradient Descent) όπου είναι μέρος του αλγορίθμου Backpropagation που χρησιμοποιείται στα νευρωνικά δίκτυα.

1.2.1 Μέθοδοι εκπαίδευσης

Οι μέθοδοι μηχανικής μάθησης που έχουν αναπτυχθεί, χρησιμοποιούνται ανάλογα με την φύση του προβλήματος και αναφέρονται ως εξής:

Μάθηση με επίβλεψη

Κατά την επιβλεπόμενη μάθηση το σύστημα προσπαθεί να υλοποιήσει μια συνάρτηση που ονομάζεται συνάρτηση στόχος (target function). Η συνάρτηση στόχος χρησιμοποιείται για την πρόβλεψη μιας μεταβλητής, που ονομάζεται εξαρτημένη μεταβλητή ή μεταβλητή εξόδου, μέσα από ένα σύνολο μεταβλητών, που ονομάζονται ανεξάρτητες μεταβλητές ή μεταβλητές εισόδου. Στην μάθηση με επίβλεψη διακρίνονται 2 κατηγορίες προβλημάτων: τα προβλήματα ταξινόμησης (Classification) που αφορούν τη δημιουργία μοντέλων πρόβλεψης διακριτών τιμών (κλάσεις/κατηγορίες) και τα προβλήματα παρεμβολής (Regression) που αφορούν τη δημιουργία μοντέλων πρόβλεψης αριθμητικών τιμών.

Μάθηση χωρίς επίβλεψη

Στη μάθηση χωρίς επίβλεψη το σύστημα έχει στόχο να ανακαλύψει συσχετίσεις και ομάδες από τα δεδομένα χωρίς να γνωρίζει κάποιες από τις τιμές εισόδου. Στη συγκεκριμένη μάθηση υπάρχουν προβλήματα κατηγοριοποίησης (Clustering) όπου το αποτέλεσμα είναι η δημιουργία ομάδων (Clusters) από τα δεδομένα.

1.2.2 Γραμμική Παλινδρόμηση (Linear Regression)

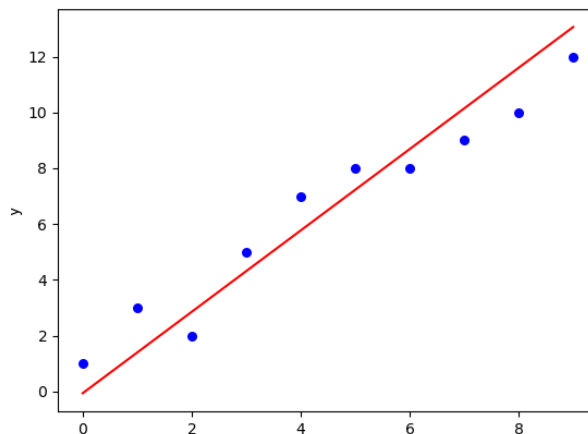
Η Γραμμική Παλινδρόμηση είναι η συσχέτιση της διακριτής εξαρτώμενης μεταβλητής (Y) από μια ή περισσότερες ανεξάρτητες μεταβλητές (X). Ο στόχος μας είναι να βρεθεί συνάρτηση $y=h(x)$ τέτοια ώστε $y^{(i)} = h(x^{(i)})$ για κάθε τιμή εκπαίδευσης. Για την εύρεση της συνάρτησης γίνεται υπόθεση ότι $h_w(x) = \sum_j w_j x_j = w^T x$ και αποτελεί μια πρόβλεψη των τιμών του y. Στόχος είναι να βρεθεί κάποιο θ τέτοιο ώστε η $h_w(x^{(i)})$ να προσεγγίζει καλύτερα την $y^{(i)}$.

Προκειμένου να ελαχιστοποιήσουμε το σφάλμα των προβλέψεων της $h_w(x)$ ως προς τις πραγματικές τιμές της y στο σύνολο δεδομένων μας ελαχιστοποιούμε το άθροισμα των τετραγώνων των διαφορών, που ορίζεται ακολούθως:

$$J(w) = \frac{1}{2} \sum_i (h_w(x^{(i)}) - y^{(i)})^2 = \frac{1}{2} \sum_i (w^T(x^{(i)}) - y^{(i)})^2$$

Αυτή η συνάρτηση ονομάζεται συνάρτηση κόστους (cost function) και υπολογίζει το σφάλμα στην πρόβλεψη του $y^{(i)}$ για συγκεκριμένο w .

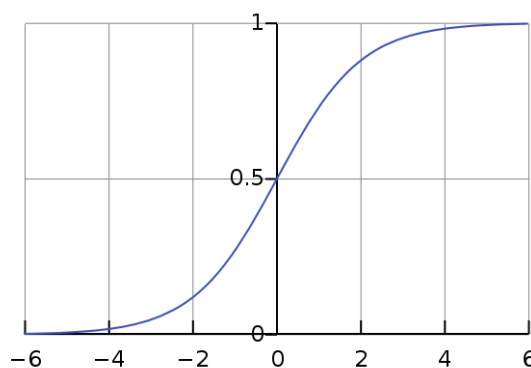
Στην παρακάτω εικόνα 1 γίνεται αναπαράσταση της ευθείας Γραμμικής Παλινδρόμησης (κόκκινο) για ένα απεικονιζόμενο σύνολο δεδομένων σημείων (μπλέ). Σκοπός είναι να βρεθεί ευθεία γραμμή $f(x) = y = ax + \beta$ για την καλύτερη προσέγγιση των σημείων.



Εικόνα 1 Αναπαράσταση Γραμμικής Παλινδρόμησης με εύρεση ευθείας γραμμής (κόκκινη) για τον καλύτερο κατανομημένο διαχωρισμό των δεδομένων

1.2.3 Λογιστική παλινδρόμηση (Logistic Regression)

Η Λογιστική Παλινδρόμηση (Logistic Regression) αποτελεί ένα μοντέλο ταξινόμησης των τιμών μιας δυαδικής μεταβλητής απόκρισης. Στο μοντέλο αυτό η μεταβλητή Y λαμβάνει συνήθως τιμές 0 ή 1 και ο στόχος μας είναι η πρόβλεψη της τιμής της από ένα πλήθος ανεξάρτητων μεταβλητών. Σε σχέση με την Γραμμική Παλινδρόμηση η διαφορά είναι ότι προσεγγίζει διακριτή συνάρτηση.



Εικόνα 2 Αναπαράσταση σιγμοειδής συνάρτηση

Το μοντέλο πρόβλεψης της Λογιστικής Παλινδρόμησης είναι το εξής:

$$P(y = 1|x) = h_w(x) = \frac{1}{1 + e^{(-w^T x)}} = \sigma(w^T x)$$

$$P(y = 0|x) = 1 - P(y = 1|x) = 1 - h_w(x)$$

Η συνάρτηση $\sigma(z) = \frac{1}{1+e^{(-z)}}$ ονομάζεται σιγμοειδής (sigmoid) ή λογιστική (logistic) συνάρτηση και απεικονίζεται στην εικόνα 2. Σκοπός είναι η εύρεση των παραμέτρων w έτσι ώστε η πιθανότητα $P(y = 1|x) = h_w(x)$ να είναι μεγάλη όταν το x ανήκει στην κλάση 1 και μικρή όταν ανήκει στην κλάση 0.

Οι παράμετροι w προκύπτουν από την ελαχιστοποίηση της ακόλουθης συνάρτησης σφάλματος:

$$J(w) = - \sum_i [y^{(i)} \log(h_w(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_w(x^{(i)}))]$$

Όταν $y^{(i)}$ είναι 0 ή 1 τότε μόνο ένας από τους δύο όρους στο άθροισμα είναι μη μηδενικός για κάθε παράδειγμα εκπαίδευσης. Όταν $y^{(i)} = 1$ τότε προκύπτει η ελαχιστοποίηση της συνάρτησης κόστους με αύξηση του $h_w(x^{(i)})$ και στην περίπτωση που $y^{(i)} = 0$ τότε υπάρχει αύξηση $1 - h_w$.

Αν $P(y = 1|x) > P(y = 0|x)$ τότε ανήκει στην κλάση 1 ενώ αν $P(y = 1|x) \leq P(y = 0|x)$ ανήκει στην κλάση 0.

1.2.4 Αλγόριθμος Βαθμωτής Καθόδου (Gradient Descent)

Ο αλγόριθμος βαθμωτής καθόδου είναι ένας αλγόριθμος βελτιστοποίησης και χρησιμοποιείται στον υπολογισμό του τοπικού ελάχιστου μιας συνάρτησης. Αυτή η βελτιστοποίηση έχει οδηγήσει σε έναν από τους πιο γνωστούς αλγόριθμους μάθησης τον Backpropagation.

Χρήση του αλγόριθμου βαθμωτής καθόδου για την επιλογή των βαρών w και την ελαχιστοποίηση του σφάλματος που παράγεται από την συνάρτηση κόστους $J(w)$. Η συνάρτηση κόστους $J(w)$ θα ελαχιστοποιηθεί ώστε να συγκλίνει στις κατάλληλες τιμές. Η επιλογή των βαρών επιλέγεται τυχαία αρχικά και στη συνέχεια μεταβάλλονται.

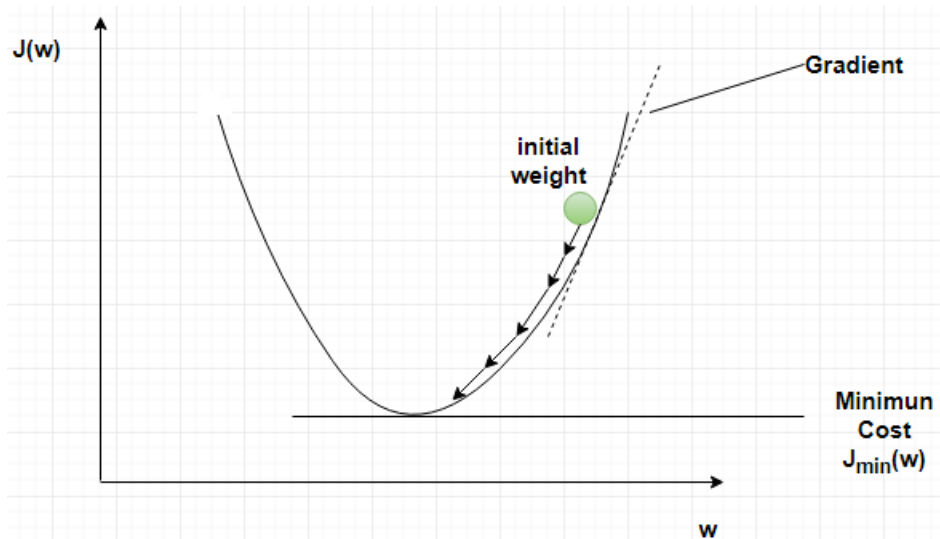
Η μεταβολή αυτή δίνεται παρακάτω:

$$w_j := w_j - \alpha \frac{\partial J(w)}{\partial w_j}$$

όπου α είναι ο βαθμός μάθησης (learning rate).

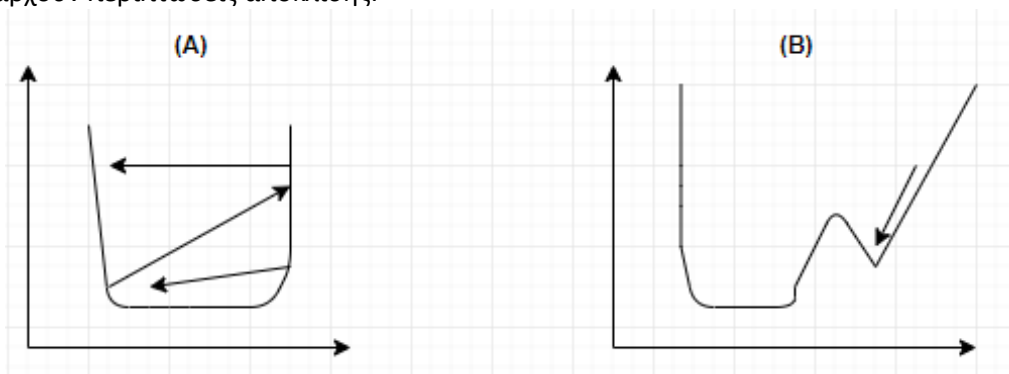
Ο αλγόριθμος σαρώνει το dataset των δεδομένων σε κάθε βήμα του, και αργότερα μεταβάλλει τα βάρη μέσω της συνάρτησης κόστους. Μια παράμετρος που επηρεάζει την ταχύτητα εκμάθησης και το τελικό ποσοστό σφάλματος είναι ο ρυθμός μάθησης α (learning rate).

Στην εικόνα 3 γίνεται η αναπαράσταση του αλγορίθμου Gradient Descent με την αρχικοποίηση των βαρών και την μεταβολή τους με συνεχείς επαναλήψεις για την ελαχιστοποίηση του σφάλματος.



Εικόνα 3 Αναπαράσταση Gradient Descent

Όταν η τιμή του ρυθμού μάθησης είναι μικρή, τότε η μεταβολή των βαρών και το σφάλμα που προκύπτει είναι μικρό με αποτέλεσμα η διαδικασία εκμάθησης να αργεί ενώ στην περίπτωση που ο ρυθμός μάθησης έχει μεγάλη τιμή η μεταβολή των βαρών είναι μεγάλη και υπάρχουν περιπτώσεις απόκλισης.



Εικόνα 4 Αναπαράσταση του ρυθμός μάθησης σε δύο περιπτώσεις

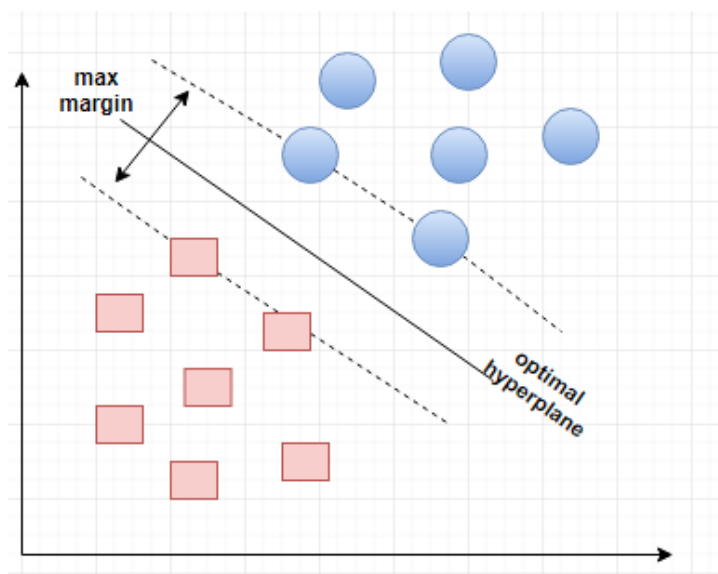
(A): Μεγάλος ρυθμός εκπαίδευσης (overshooting)

(B): Μικρός ρυθμός εκπαίδευσης: επαναλήψεις και παγίδευση σε τοπικό ελάχιστο

1.3 Μηχανές Διανυσμάτων Υποστήριξης (SVM)

Παρακάτω ακολουθεί μια σύντομη περιγραφή των βασικών χαρακτηριστικών των Μηχανών Διανυσμάτων Υποστήριξης. Κρίνεται απαραίτητο να αναφερθούν, καθώς οι ταξινομήσεις των πειραμάτων που υλοποιούνται στα επόμενα κεφάλαια, υλοποιούνται με τη χρήση SVM για την καλύτερη απόδοση των νευρωνικών δικτύων.

Οι Μηχανές Διανυσμάτων Υποστήριξης είναι μοντέλα μάθησης με επίβλεψη τα οποία χρησιμοποιούνται σε εφαρμογές κατηγοριοποίησης και αναγνώρισης δεδομένων. Το κύριο χαρακτηριστικό τους είναι ότι έχουν γεωμετρικό υπόβαθρο σε σχέση με τη Γραμμική και Λογιστική Παλινδρόμηση. Πιο συγκεκριμένα, σε ένα σύνολο δεδομένων εκπαίδευσης με ένα ταξινομητή SVM γίνεται προσπάθεια εύρεσης ενός βέλτιστου ορίου απόφασης (decision boundary) μεταξύ των δεδομένων, καθορίζοντας σε ποια κλάση θα ανήκει η κάθε τιμή. Το όριο αυτό είναι η γραμμή που διχοτομεί την μεγαλύτερη απόσταση μεταξύ των δεδομένων των 2 κλάσεων. Η απόσταση μεταξύ του ορίου απόφασης και του εγγύτερου παραδείγματος εκπαίδευσης από κάθε κλάση ονομάζεται περιθώριο (margin) και οι τιμές της κάθε κλάσης που βρίσκονται πλησιέστερα σε αυτό το όριο ονομάζονται διανύσματα υποστήριξης (support vectors). Για περισσότερες λεπτομέρειες σε θεωρία και αποδείξεις για τους ταξινομητές SVM παραπέμπουμε τον αναγνώστη στο βιβλίο του Flach [2] (Κεφάλαιο 7 ενότητα 3).



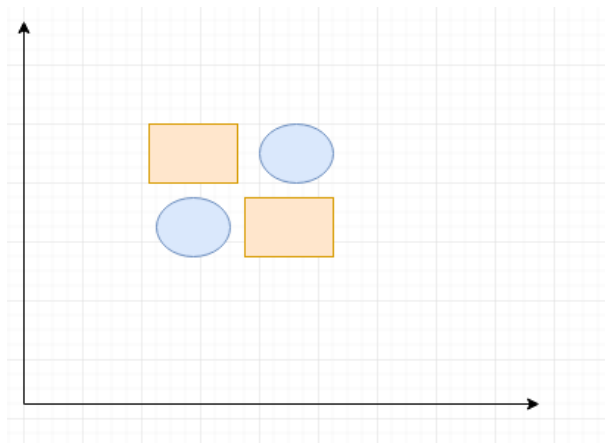
Εικόνα 5 Αναπαράσταση του SVM ταξινομητή, με το όριο απόφασης ως την μέγιστη απόσταση από τα δεδομένα των 2 κλάσεων.

Ας θεωρήσουμε αρχικά την περίπτωση τα παραδείγματα εκπαίδευσης (2 κλάσεων, 1 και -1) είναι γραμμικά διαχωρίσιμα από ευθεία $w \cdot x + b$. Αυτό σημαίνει ότι υπάρχουν παράμετροι w και b , έτσι ώστε για κάθε παράδειγμα εκπαίδευσης $i=1 \dots m$ το πρόσημο του $w \cdot x^{(i)} + b$ ταυτίζεται με το πρόσημο του $y^{(i)}$. Τότε οι τιμές των παραμέτρων w και b δίνονται από την λύση του ακόλουθου μαθηματικού προγράμματος:

$$w^*, b^* = \underset{w, b}{\operatorname{argmin}} \frac{1}{2} \|w\|^2 \quad \text{έτσι ώστε } y_i(w \cdot x_i - b) \geq 1, 1 \leq i \leq m$$

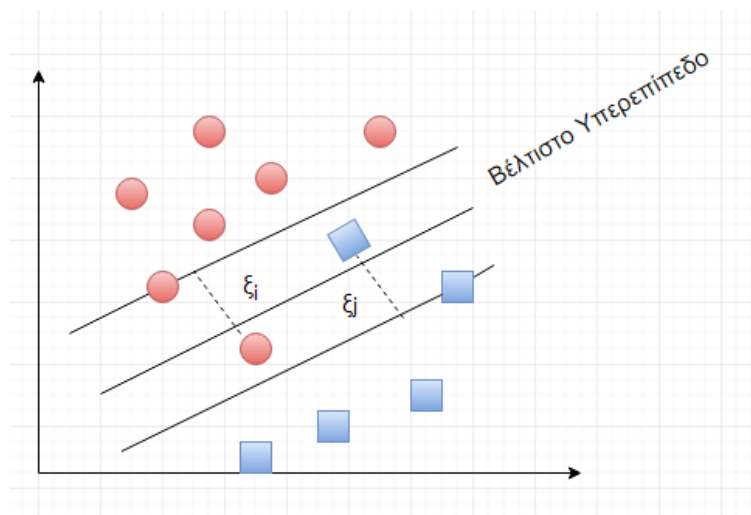
1.3.1 Soft Margin

Όταν το σύνολο εκπαίδευσης είναι μη γραμμικώς διαχωρίσιμο και οι κλάσεις δεν μπορούν να διαχωριστούν η λύση στο πρόβλημα είναι η εισαγωγή μεταβλητών απόκλισης ξ_i για την κάθε τιμή. Παρακάτω αναπαρίσταται η εικόνα 6 που δείχνει τη δυσκολία των μη γραμμικά διαχωρίσιμων δεδομένων.



Εικόνα 6 Αναπαράσταση μη γραμμικώς διαχωρίσιμα δεδομένα

Στην εισαγωγή μεταβλητών απόκλισης ξ_i για την κάθε τιμή, κάποιες είναι μέσα στο περιθώριο (margin) και κάποιες άλλες στη λάθος πλευρά του ορίου απόφασης, οι οποίες ονομάζονται σφάλματα περιθωρίου (margin errors).



Εικόνα 7 Αναπαράσταση SoftMargin

Με την προσθήκη των μεταβλητών απόκλισης ξ_i στη συνάρτηση $w \cdot x_i - t \geq 1 - \xi_i$ το αποτέλεσμα είναι η βελτιστοποίηση του προβλήματος περιθωρίου.

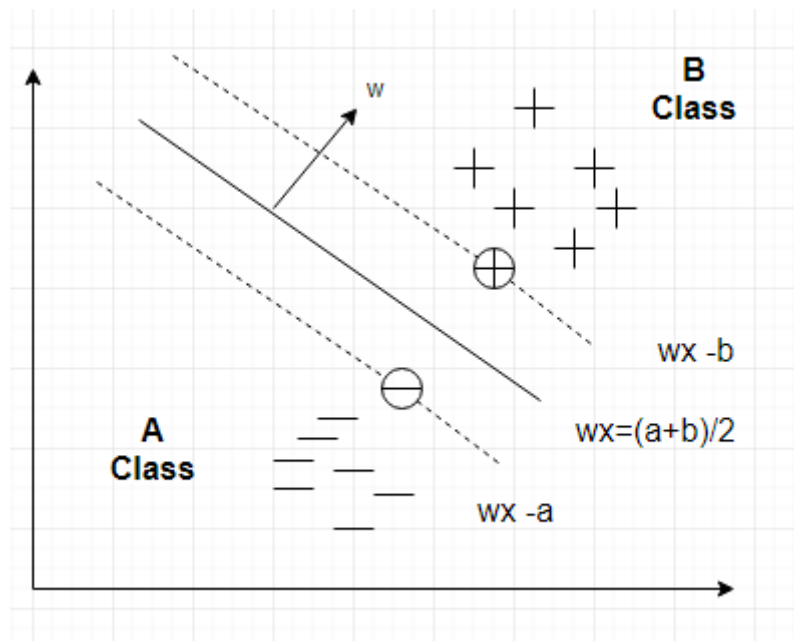
$$w^*, t^*, \xi_i^* = \operatorname{argmin}_{w, t, \xi_i} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i$$

με $y_i (w \cdot x_i - t) \geq 1 - \xi_i$ και $\xi_i \geq 0, 1 \leq i \leq n$

Το C είναι μια παράμετρος που καθορίζεται από τον χρήστη για την μεγιστοποίηση του περιθωρίου έναντι της ελαχιστοποίησης της μεταβλητής απόκλισης. Όταν το C έχει μεγάλη τιμή τότε τα σφάλματα περιθωρίου επιφέρουν μεγάλη ποινή ενώ όταν η τιμή C είναι μικρή τότε δημιουργείται ένα μεγάλο περιθώριο.

1.3.2 Διανύσματα Υποστήριξης και Συναρτήσεις Πυρήνα

Στην αναπαράσταση παράλληλων επιπέδων όπως φαίνεται και στο παρακάτω σχήμα τα δεδομένα της κλάσης A είναι τοποθετημένα ώστε να ικανοποιούν την σχέση $w \cdot x + b \geq a$ και τουλάχιστον ένα δείγμα να ικανοποιεί την σχέση $w \cdot x + b = a$. Όμοια με τα δεδομένα της κλάσης B να ικανοποιούν την σχέση $w \cdot x + b \leq b$ και τουλάχιστον ένα δείγμα να ικανοποιεί την ισότητα $w \cdot x + b = b$.



Εικόνα 8 Διυικό πρόβλημα

Το πρόβλημα μεγιστοποίησης των υπερεπιπέδων ορίζεται ως:

$$\min_{w,a,b} \frac{1}{2} \|w\|^2 - (a - b)$$

Η απόσταση των δύο επιπέδων είναι η απόσταση μεταξύ των δύο κλάσεων. Τα κοντινότερα σημεία μεταξύ των κλάσεων θα πρέπει να είναι και σημεία των δύο παράλληλων επιπέδων και η ευθεία που ενώνει τα δύο σημεία θα πρέπει να είναι ορθογώνια ως προς τα δύο επίπεδα. Στην περίπτωση που δεν είναι, σημαίνει πως τα επίπεδα δεν έχουν επιλεγεί στην μέγιστη απόσταση τους ή τα σημεία δεν είναι τα κοντινότερα. Αυτό είναι ένα χαρακτηριστικό παράδειγμα του διυικού προβλήματος (dual problem).

Μία βέλτιστη λύση στο πρόβλημα είναι η μεταβλητή απόκλισης ξ_i να είναι μηδέν που προκύπτει από $C - \alpha_i - \beta_i = 0$ για όλα τα i και ο προστιθέμενος όρος να εξαφανίζεται από το

δυσκό πρόβλημα. Όταν τα a_i και β_i είναι θετικά, το a_i δεν μπορεί να είναι μεγαλύτερο από το C , το οποίο είναι το ανώτερο όριο στο a_i για το δυσκό πρόβλημα.

$$a_1^*, \dots, a_n^* = \operatorname{argmax}_{a_1, \dots, a_n} - \frac{1}{2} \left(\sum_{i=1}^n \sum_{j=1}^n a_i a_j x_i x_j y_i y_j \right) + \sum_{i=1}^n a_i$$

$$\mu \varepsilon 0 \leq a_i \leq C \text{ και } \sum_{i=1}^n a_i y_i = 0$$

Περιορίζοντας τις μεταβλητές απόκλισης να είναι θετικές και προσθέτοντας τις στη συνάρτηση $w x_i - t \geq 1 - \xi_i$ γίνεται ελαχιστοποίηση του περιθωρίου. Οι μεταβλητές απόκλισης εφαρμόζονται στην έννοια της απώλειας άρθρωσης (hinge loss) όπου όταν ένα περιθώριο $z > 1$ δεν συνεπάγεται με κάποιο σφάλμα και όταν ένα περιθώριο $z = 1 - \xi \leq 1$ συνεπάγεται με κάποιο σφάλμα $\xi = 1 - z$ [2].

Μια λύση στο πρόβλημα βελτιστοποίησης περιθωρίου είναι ο διαχωρισμός των παραδειγμάτων εκπαίδευσης σε τρεις περιπτώσεις:

- $a_i = 0$ αφορούν τα έξω και πάνω στο περιθώριο
- $0 < a_i < C$ αφορούν τις μηχανές διανυσμάτων υποστήριξης πάνω στο περιθώριο
- $a_i = C$ αφορούν αυτά που είναι πάνω ή μέσα στο περιθώριο

Συναρτήσεις πυρήνα (Kernel Methods). Οι συναρτήσεις πυρήνα (Kernel Functions) είναι απεικονίσεις των διανυσμάτων εισόδου x και ανήκουν στο \mathbb{R} . Μετασχηματίζουν κατάλληλα τα διανύσματα εισόδου ώστε να επιτύχουν μια γενικεύσιμη λύση στο πρόβλημα ταξινόμησης. Αυτό επιτυγχάνεται με τον μετασχηματισμό από τον χώρο σε έναν χώρο χαρακτηριστικών μεγαλύτερης διάστασης με μια μη γραμμική απεικόνιση $\Phi(x)$.

Η συνάρτηση πυρήνα ορίζεται ως εξής:

$$k(x, x') = \Phi^T(x) \Phi(x') = \langle \Phi(x), \Phi(x') \rangle$$

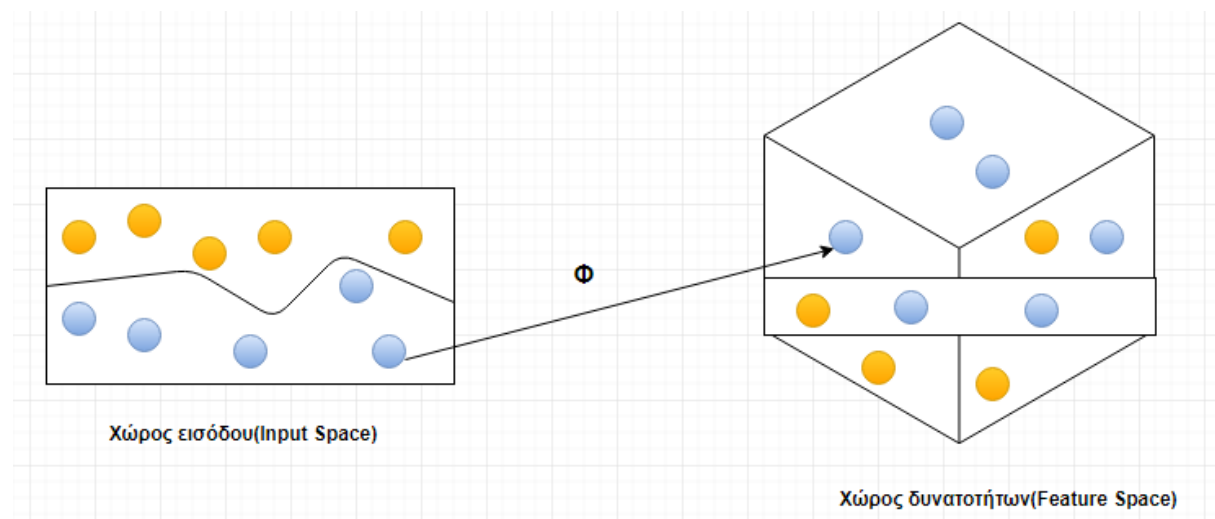
η οποία είναι ένα εσωτερικό γινόμενο μεταξύ των διανυσμάτων $\Phi(x)$ και $\Phi(x')$ στον καινούριο χώρο χαρακτηριστικών.

Η πιο απλή συνάρτηση πυρήνα που προκύπτει για την μοναδιαία απεικόνιση $\Phi(x) = x$ είναι:

$$k(x, x') = x^T x$$

η οποία είναι το εσωτερικό γινόμενο μεταξύ των δύο διανυσμάτων.

Η κεντρική ιδέα είναι η εύρεση απλούστερων συνόρων απόφασης (υπερεπίπεδων) από τον χώρο εισόδου στο χώρο χαρακτηριστικών υψηλότερης διάστασης που κάνει το πρόβλημα γραμμικά διαχωρίσιμο.



Εικόνα 9 Απεικόνιση συνάρτησης Φ σε χώρο μεγαλύτερων διαστάσεων

Παρακάτω υπάρχουν κάποιες συναρτήσεις πυρήνα:[26]

$$k(x, y) = e^{-|x-y|^2/(2\sigma^2)} \text{ Γκαουσιανή RBF}$$

$$k(x, y) = [X^T y + \theta] \text{ Πολυωνυμική}$$

$$k(x, y) = \tanh(ax^T y + \theta) \text{ Σιγμοειδής}$$

1.4 Βασικές Αρχές Νευρωνικών Δικτύων

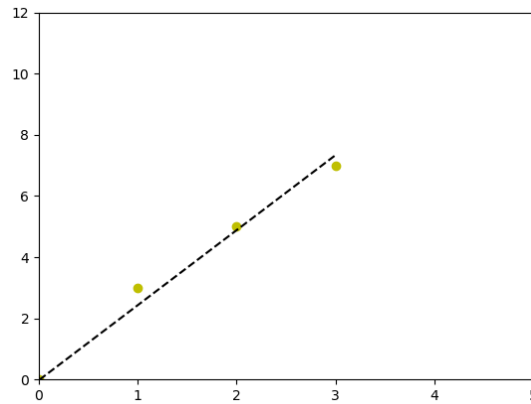
1.4.1 Συναρτήσεις Ενεργοποίησης

Υπάρχουν αρκετές συναρτήσεις ενεργοποίησης (activation functions) ή συναρτήσεις μεταφοράς (transfer functions) που χρησιμοποιούνται στα νευρωνικά δίκτυα για τον μετασχηματισμό των βαρών και η επιλογή τους εξαρτάται από την αρχιτεκτονική του νευρωνικού δικτύου. Παρακάτω παρουσιάζονται μερικές από τις συναρτήσεις ενεργοποίησης.

- Ταυτοτική Συνάρτηση

Είναι η πιο απλή περίπτωση εξόδου του νευρωνικού δικτύου διότι αποτελείται από τον πολλαπλασιασμό των βαρών με το άθροισμα των biases. Αυτού του είδους οι γραμμικές συναρτήσεις εξάγουν γραμμικό μετασχηματισμό του διανύσματος εισόδου και χρησιμοποιείται κυρίως σε προβλήματα Γραμμικής Παλινδρόμησης.

$$f(x) = y_i = h(x, w) = w^T x = \sum_{i=0}^m w_i x_i + b$$



Εικόνα 10 Αναπαράσταση ταυτοτική συνάρτησης

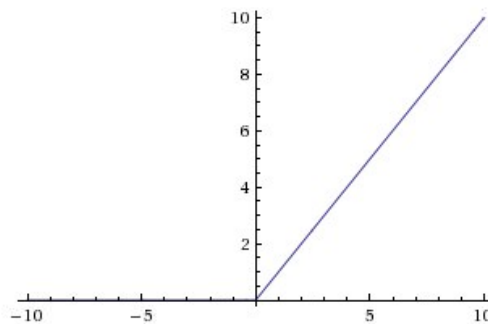
Η γραμμική συνάρτηση μεταφοράς είναι η πιο απλή συνάρτηση μεταφοράς όπου m είναι το σύνολο των τιμών εκπαίδευσης, w είναι τα βάρη που αντιπροσωπεύουν την κλίση της ευθείας και b η μετατόπιση της ευθείας από την αρχή των αξόνων.

- Rectified Linear Unit (RELU)

Όταν στην είσοδο του δικτύου υπάρχουν αρνητικές τιμές τότε η συνάρτηση επιστρέφει μηδέν ενώ όταν οι τιμές εισόδου είναι θετικές τότε η συνάρτηση επιστρέφει την τιμή πίσω. Αυτό σημαίνει πως όταν $x < 0$ τότε $f(x) = 0$ και όταν $x \geq 0$ τότε $f(x) = x$.

Η συνάρτηση είναι:

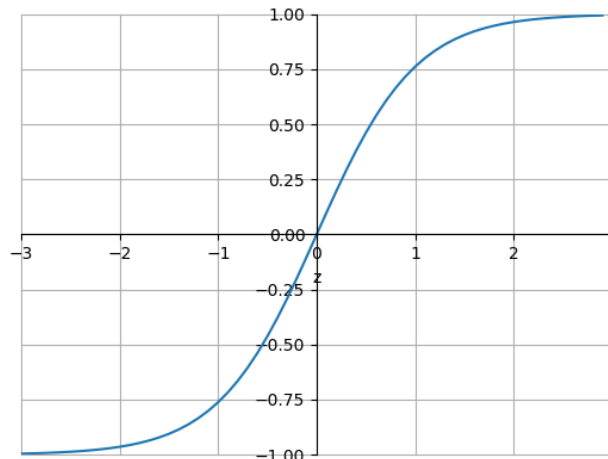
$$f(x) = \max(0, x)$$



Εικόνα 11 Αναπαράσταση συνάρτησης RELU

- Tanh (Hyperbolic tangent)

Όταν η σιγμοειδής συνάρτηση χρησιμοποιείται, η υπερβολική tanh έχει καλύτερη απόδοση στη κατηγοριοποίηση 2 κλάσεων. Το εύρος τιμών της συνάρτησης είναι από -1 έως 1. Η ταυτοτική συνάρτηση μοιάζει με την tanh κοντά στο 0 με την έννοια ότι $\tanh(0) = 0$ ενώ $\sigma(0) = \frac{1}{2}$, έτσι ώστε όταν εκπαιδεύεται ένα νευρωνικό δίκτυο να μοιάζει με το γραμμικό μοντέλο ενεργοποίησης εφόσον οι ενεργοποιήσεις είναι μικρές. Αυτό έχει σαν αποτέλεσμα την ευκολότερη εκπαίδευση του tanh δικτύου.



Εικόνα 12 Αναπαράσταση συνάρτησης tanh

1.4.2 Αρχιτεκτονική νευρωνικών δικτύων

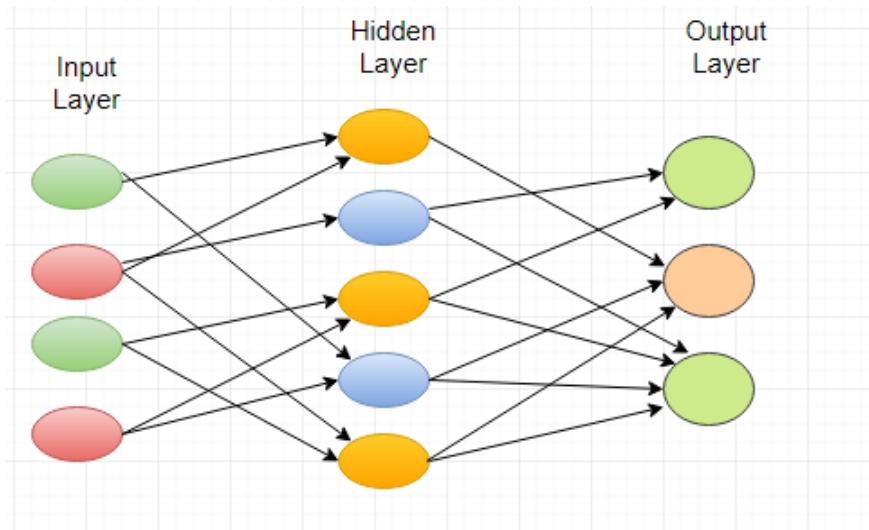
Τα επίπεδα ενός νευρωνικού δικτύου αναφέρονται ως εξής:

- **Επίπεδο εισόδου (input layer):** είναι το πρώτο επίπεδο ενός δικτύου και σε αυτό εισάγονται τα δεδομένα εισόδου όπου το πλήθος των κόμβων του είναι ίσος με τις τιμές εισόδου (Features).
- **Ενδιάμεσο επίπεδο (hidden layer):** είναι το ενδιάμεσο επίπεδο που συνδέεται με το προηγούμενο και το επόμενο από αυτό επίπεδο (μπορεί να υπάρχουν πολλά ενδιάμεσα επίπεδα). Ο αριθμός των ενδιάμεσων επιπέδων καθορίζεται από τον χρήστη ανάλογα με το πρόβλημα ταξινόμησης.
- **Επίπεδο εξόδου (output layer):** είναι το τελευταίο επίπεδο του δικτύου που από αυτό προκύπτουν τα αποτελέσματα μετά την εκπαίδευση του.

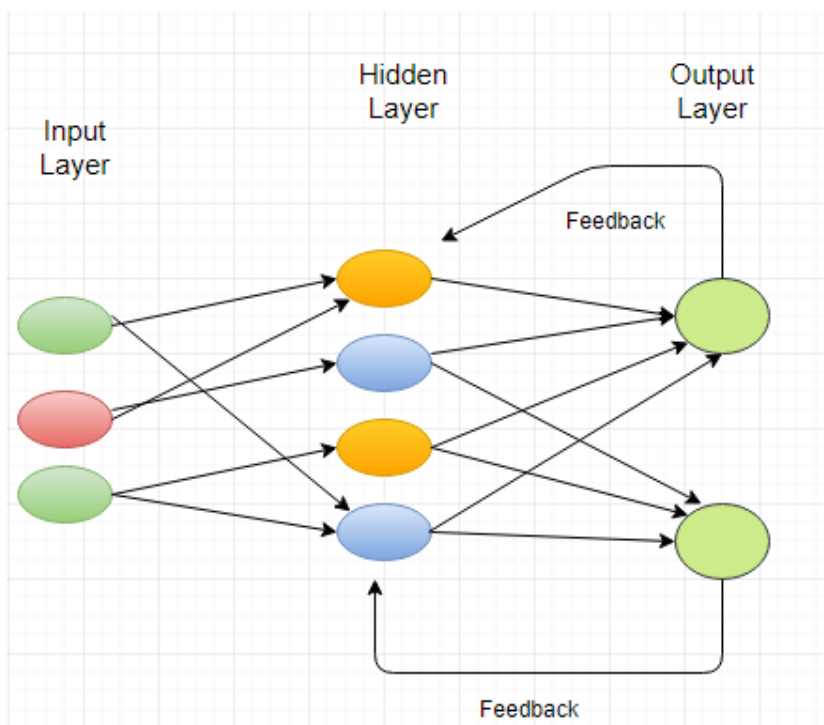
Τα νευρωνικά δίκτυα κατηγοριοποιούνται ανάλογα με την αρχιτεκτονική τους και τον τρόπο που συνδέονται οι κόμβοι μεταξύ τους.

- **Εμπρόσθια τροφοδότηση (feedforward):** η πληροφορία μεταφέρεται από την είσοδο έως στην έξοδο του δικτύου χωρίς να μπορεί να μεταφερθεί αντίστροφα. Αυτά τα δίκτυα διακρίνονται σε μονοεπίπεδα (Perceptron) ή πολυεπίπεδα (Multilayer Perceptron - MLP) και περιλαμβάνουν ένα ή περισσότερα ενδιάμεσα επίπεδα.

- **Ανατροφοδότηση (feedback):** είναι τα δίκτυα που περιέχουν τουλάχιστον έναν κόμβο ανατροφοδότησης έτσι ώστε η πληροφορία στην έξοδο να τροφοδοτείται στις εισόδους των κόμβων των προηγούμενων επιπέδων.



Εικόνα 13 Αναπαράσταση εμπρόσθιας τροφοδότησης νευρωνικού δικτύου

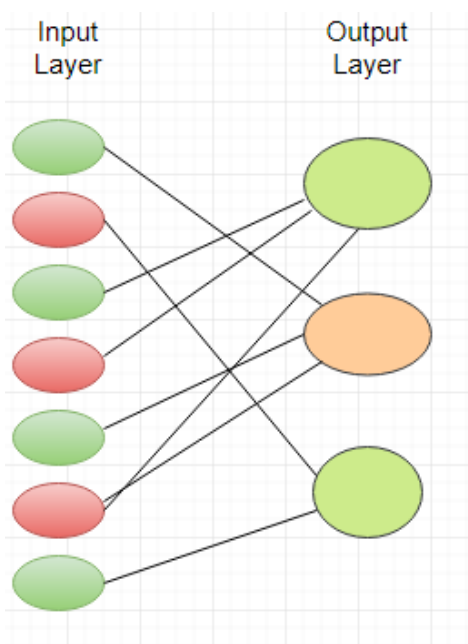


Εικόνα 14 Αναπαράσταση ανατροφοδότησης νευρωνικού δικτύου

1.4.3 Μοντέλο Perceptron

Το συγκεκριμένο μοντέλο ήταν από τα πρώτα μοντέλα που αναπτύχθηκαν στην δεκαετία του 1950. Αποτελείται από ένα επίπεδο εισόδου και ένα επίπεδο εξόδου. Το απλό αυτό μοντέλο χρησιμοποιεί ως συνάρτηση ενεργοποίησης τη ταυτοτική συνάρτηση και χρησιμοποιείται για το διαχωρισμό ή την ταξινόμηση των δεδομένων σε 2 κλάσεις. Στην παρακάτω εικόνα 15

γίνεται αναπαράσταση κόμβου με 7 κόμβους εισόδου και 3 κόμβους εξόδου.



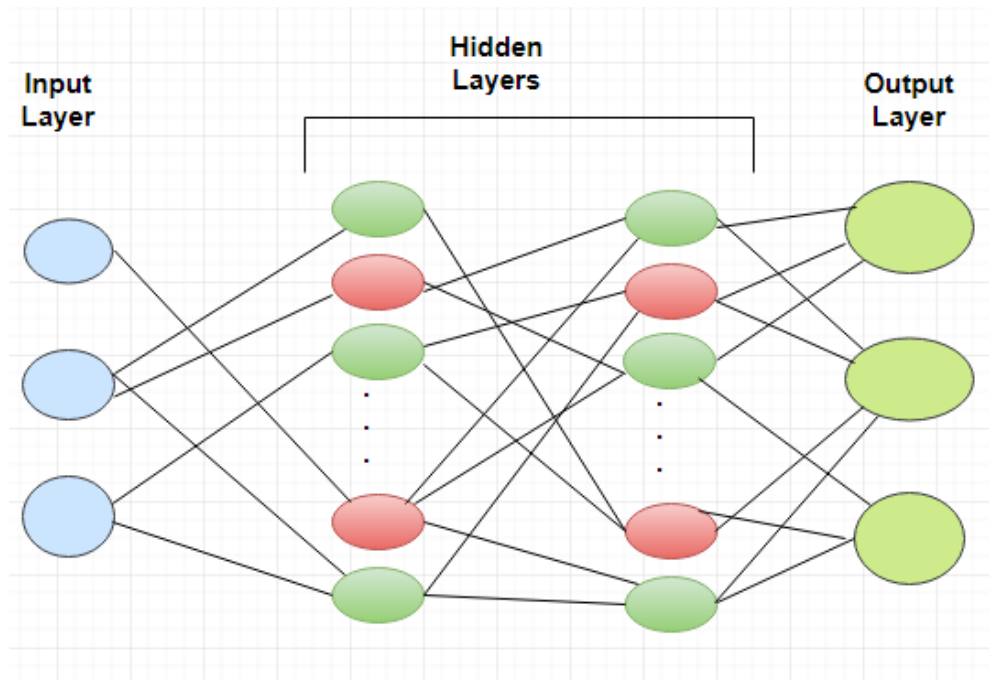
Εικόνα 15 Αναπαράσταση κόμβου με 7 κόμβους εισόδου και 3 κόμβους εξόδου

Η διαδικασία εκπαίδευσης ενός κόμβου είναι η μάθηση με επίβλεψη και υλοποιείται με το σύνολο τιμών εκπαίδευσης. Σκοπός της εκπαίδευσης είναι ο υπολογισμός των βαρών κατάλληλα, έτσι ώστε το δίκτυο να έχει το ελάχιστο ποσοστό σφάλματος και το καλύτερο ποσοστό ακρίβειας στα αποτελέσματα εξόδου.

Στην περίπτωση του κόμβου και της ύπαρξης 2 πιθανών κλάσεων, τα βάρη τροποποιούνται συνεχώς μέχρι η κάθε επιμέρους διαφορά να γίνει 0. Αυτή η διαδικασία ονομάζεται Κανόνας Δέλτα (Delta Rule), με δ να ορίζεται ως η διαφορά μεταξύ του επιθυμητού αποτελέσματος εξόδου και του πραγματικού αποτελέσματος.

Το πιο σημαντικό πρόβλημα που μπορεί να αντιμετωπίσει το Perceptron είναι ότι μπορεί να επιλύσει μόνο γραμμικώς διαχωρίσιμα προβλήματα. Στην αντιμετώπιση αυτού του προβλήματος έχουν δημιουργηθεί πιο πολύπλοκες αρχιτεκτονικές οι οποίες ανταποκρίνονται σε πιο σύνθετα προβλήματα όπως το MLP .

Multilayer Perceptron. Οι αρχιτεκτονικές MLP σε σχέση με αυτές του Perceptron είναι η ύπαρξη περισσότερων ενδιάμεσων επιπέδων μεταξύ των επιπέδων εισόδου και εξόδου. Πιο απαιτητικά προβλήματα αντιμετωπίζονται με την ύπαρξη περισσότερων ενδιάμεσων επιπέδων διότι υπάρχει αύξηση του αριθμού των κόμβων στο νευρωνικό δίκτυο άρα και της πληροφορίας των βαρών. Η ροή των δεδομένων πραγματοποιείται από το επίπεδο εισόδου προς το επίπεδο εξόδου. Αυτό έχει σαν αποτέλεσμα, η μετάβαση στο επίπεδο εξόδου να μπορεί να πραγματοποιηθεί προς μια μόνο κατεύθυνση και όχι προς την αντίθετη φορά. Στην εικόνα 16 γίνεται αναπαράσταση του κόμβου με 2 ενδιάμεσα επίπεδα.



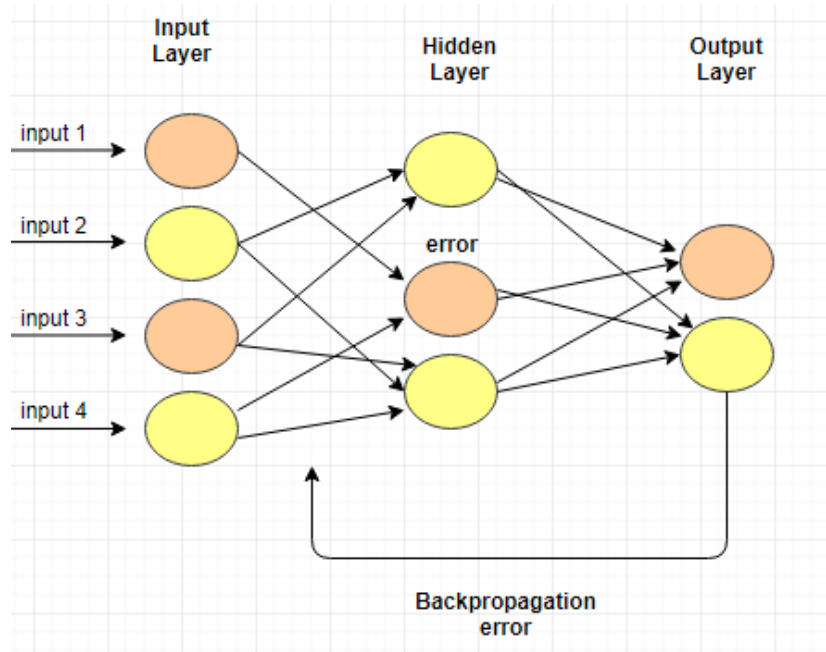
Εικόνα 16 Αναπαράσταση κόμβου (MLP) με 2 ενδιάμεσα επίπεδα

υ

1.4.4 Αλγόριθμος Οπίσθιας διάδοσης σφάλματος (Backpropagation)

Για να χρησιμοποιηθεί ο συγκεκριμένος αλγόριθμος θα πρέπει το νευρωνικό δίκτυο να περιέχει ένα ή περισσότερα ενδιάμεσα επίπεδα και η συνάρτηση ενεργοποίησης να είναι η σιγμοειδής. Η διαδικασία εκπαίδευσης ενός δικτύου με αυτή την μέθοδο είναι η παρακάτω:

- 1) Οι τιμές των βαρών θέτονται τυχαία κατά την έναρξη της διαδικασίας της εκπαίδευσης στο νευρωνικό δίκτυο.
- 2) Κατά την εξαγωγή σφάλματος στο αποτέλεσμα κατά την επανάληψη, τα βάρη μετασχηματίζονται ώστε να γίνει ελαχιστοποίηση του σφάλματος που είχε προκύψει.
- 3) Η διαδικασία επαναλαμβάνεται μέχρι το σφάλμα να μηδενιστεί ή να θεωρηθεί αρκετά μικρό.



Εικόνα 17 Αναπαράσταση αλγορίθμου Backpropagation

Ένα πολύ σημαντικό χαρακτηριστικό του συγκεκριμένου αλγορίθμου είναι ότι αποτελείται από δύο βασικές λειτουργίες. Η μια είναι ότι η πληροφορία μεταφέρεται προς τα εμπρός από το επίπεδο εισόδου προς το επίπεδο εξόδου μέσω των ενδιάμεσων επιπέδων. Από αυτή την μεταφορά προκύπτει κάποια τιμή εξόδου, όπου λόγω κάποιας διαφοράς που μπορεί να υπάρχει μεταξύ της επιθυμητής και της πραγματικής εξόδου, η πληροφορία μεταδίδεται και προς την αντίθετη κατεύθυνση. Αυτή είναι η δεύτερη λειτουργία, η οποία επιδιώκει την τροποποίηση των βαρών με κύριο στόχο την ελαχιστοποίηση του σφάλματος. Η διαδικασία της ελαχιστοποίησης του σφάλματος γίνεται με την χρήση του αλγορίθμου της βαθμωτής καθόδου.

1.4.5 Κριτήρια τερματισμού

Παραπάνω έγινε περιγραφή του τρόπου υλοποίησης της διαδικασίας εκπαίδευσης του νευρωνικού δικτύου με τον αλγόριθμο οπίσθιας διάδοσης σφάλματος. Στο συγκεκριμένο σημείο θα προσδιοριστούν συγκεκριμένα κριτήρια για τον τερματισμό της διαδικασίας της εκπαίδευσης.

1) **Πρόωρη διακοπή (early stopping):** στην πρόωρη διακοπή θα πρέπει να γίνει αναφορά στο σύνολο δεδομένων για επικύρωση (validation set). Με τα συγκεκριμένα δεδομένα υπολογίζεται ένα σφάλμα επικύρωσης (validation error) σε κάθε εποχή (epoch) το οποίο είναι διαφορετικό από το σφάλμα εκπαίδευσης. Υπάρχουν περιπτώσεις που το σφάλμα εκπαίδευσης μειώνεται σε κάθε εποχή ενώ το σφάλμα επικύρωσης αυξάνεται. Το αποτέλεσμα είναι ο αλγόριθμος να αντιμετωπίζει πρόβλημα υπερεκπαίδευσης (overfitting) και να τερματίζεται πρόωρα με διακοπή της διαδικασίας εκπαίδευσης.

2) **Σύγκλιση αλγορίθμου:** η σύγκλιση αλγορίθμου συμβαίνει όταν η μεταβολή του

σφάλματος είναι μικρή.

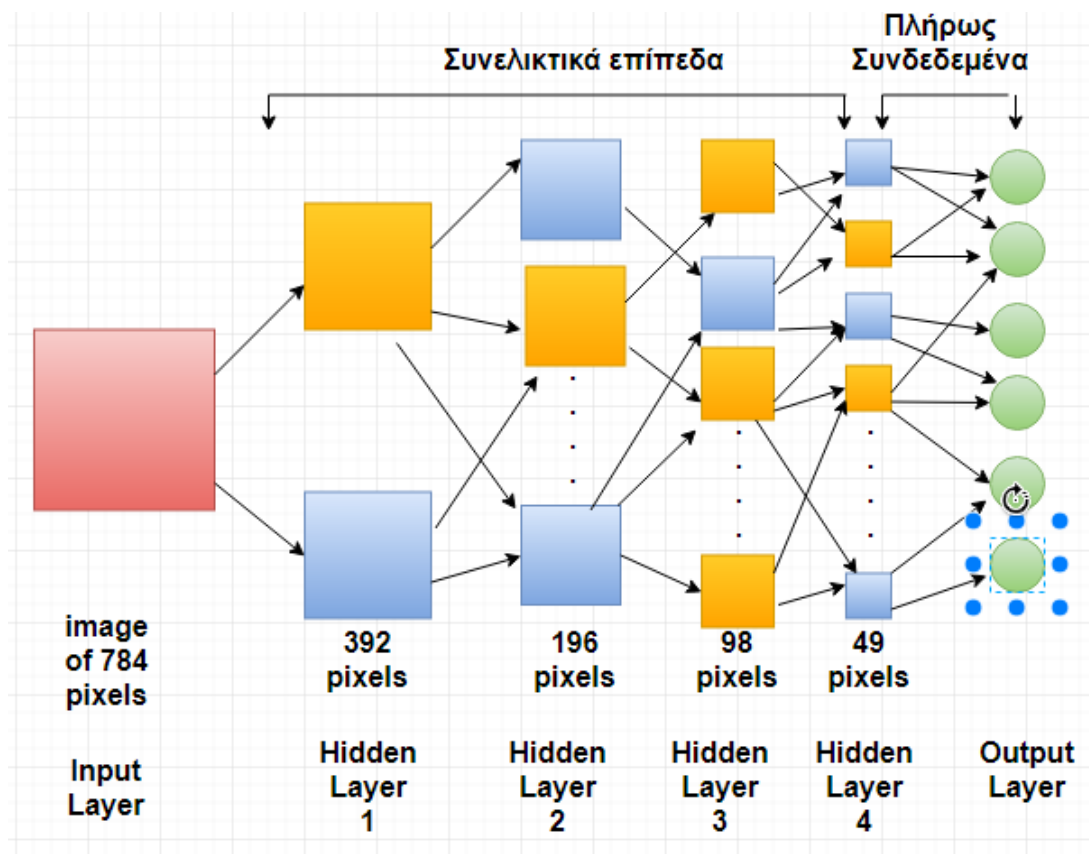
- 3) **Τερματισμός με προκαθορισμένο όριο επαναλήψεων:** γίνεται τερματισμός της εκπαίδευσης του νευρωνικού δικτύου με συγκεκριμένο αριθμό επαναλήψεων στην περίπτωση που το σφάλμα μεταβάλλεται ελάχιστα. Οι επαναλήψεις στο νευρωνικό δίκτυο αφορούν την έννοια των εποχών.

Κεφάλαιο 2 Συνελικτικά Νευρωνικά Δίκτυα

Τα συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks, ΣΝΔ) είναι αρχιτεκτονικές νευρωνικών δικτύων εμπρόσθιας τροφοδότησης (feedforward) που εφαρμόζονται κυρίως σε εφαρμογές αναγνώρισης εικόνας και ήχου. Ο όρος «Συνελικτικό» υποδηλώνει ότι το δίκτυο χρησιμοποιεί την πράξη της συνέλιξης σε τουλάχιστον ένα επίπεδο (επίπεδο συνέλιξης) του δικτύου, για τη μείωση των διαστάσεων των χαρακτηριστικών (features) της εικόνας που δέχεται στο επίπεδο εισόδου [1].

Σε αυτό το κεφάλαιο θα αναλυθούν τα βασικά χαρακτηριστικά των ΣΝΔ όπως το συνελικτικό επίπεδο (convolutional layer), το επίπεδο χωρικής υποδειγματοληψίας (pooling layer), τα πλήρως συνδεδεμένα επίπεδα (fully connected layers) και οι συναρτήσεις σύνδεσης (link functions).

Στην εικόνα 18 γίνεται αναπαράσταση μια γενικής μορφής ενός συνελικτικού δικτύου του επιπέδου εισόδου του που είναι μια εικόνα με 784 pixels, τα επόμενα επίπεδα είναι 4 συνελικτικά επίπεδα, που μειώνουν τις διαστάσεις της εικόνας διαδοχικά σε 392 pixels, 196 pixels, 98 pixels και 49 pixels. Το τελευταίο επίπεδο αποτελείται από κόμβους που είναι πλήρως συνδεδεμένοι με το προηγούμενο συνελικτικό επίπεδο.



Εικόνα 18 Αρχιτεκτονική συνελικτικού δικτύου

2.1 Χαρακτηριστικά Συνελικτικών Δικτύων

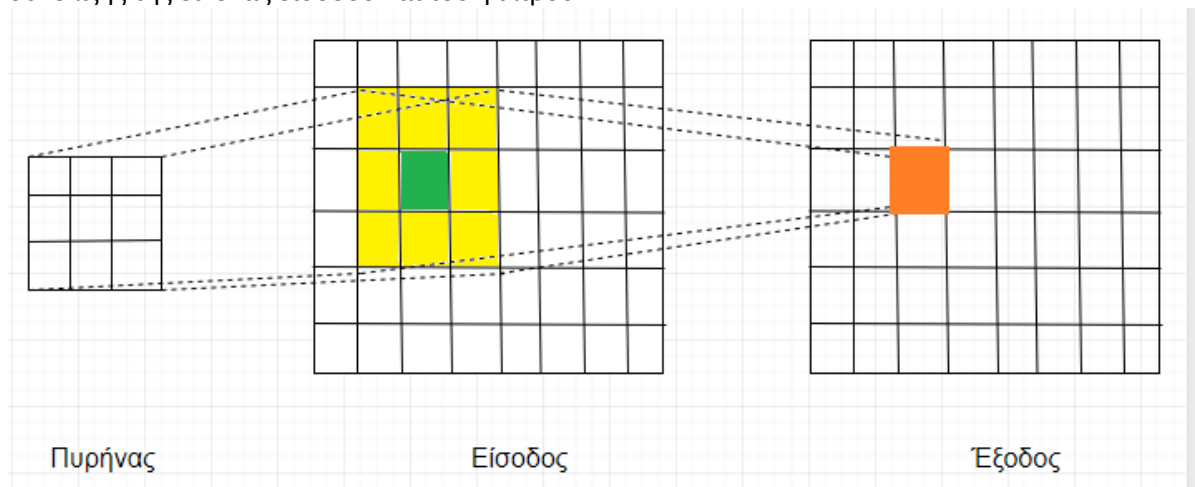
Η διαφορά των ΣΝΔ σε σχέση με τα MLP είναι ότι μειώνουν την διάσταση των δεδομένων στο επίπεδο εισόδου με διαδοχικά επίπεδα συνέλιξης. Όταν τα δεδομένα εισόδου στο συνελικτικό δίκτυο είναι έγχρωμες εικόνες, οι διαστάσεις τους είναι τρεις: ύψος, πλάτος και βάθος με τρία επίπεδα χρωμάτων (κόκκινο, πράσινο, μπλέ) τα οποία είναι τα χαρακτηριστικά που τροφοδοτούνται στο επίπεδο εισόδου. [16].

2.1.1 Επίπεδο Συνέλιξης

Κάθε συνελικτικό δίκτυο αποτελείται από ένα τουλάχιστον επίπεδο συνέλιξης στα δεδομένα εισόδου του. Αυτό το επίπεδο αποτελείται από εκπαιδευόμενα φίλτρα (συνελικτικός πυρήνας) και μέσω των βαρών του κάθε κόμβου, είναι συνδεδεμένο με την διάσταση βάθους της εικόνας εισόδου. Το κάθε φίλτρο μετακινείται κατά μήκος των διαστάσεων πλάτους και ύψους της εικόνας με αποτέλεσμα την απεικόνιση ενός δισδιάστου «χάρτη χαρακτηριστικών» (feature map). Η λειτουργία αυτή υλοποιεί την πράξη της συνέλιξης του φίλτρου με την εικόνα εισόδου. Το αποτέλεσμα της συνέλιξης προκύπτει από το εσωτερικό γινόμενο των τιμών της εικόνας εισόδου με τις τιμές του φίλτρου.

Για τον υπολογισμό των διαστάσεων του πλάτους και του ύψους της εικόνας εξόδου πρέπει να οριστούν δύο παράμετροι στο συνελικτικό επίπεδο: το βήμα (stride) και το γέμισμα (padding). Το βήμα καθορίζει το πόσο πυκνή θα είναι η δειγματοληψία στην είσοδο, δηλαδή ο αριθμός των εικονοστοιχείων που μετακινούνται οριζόντια και κάθετα. Όταν το βήμα είναι ίσο με 1 τότε η σάρωση γίνεται πυκνή ενώ για μεγαλύτερα βήματα η σάρωση στην είσοδο θα είναι αραιή. Το γέμισμα χρησιμοποιείται για την προσθήκη μηδενικών στην είσοδο, συμβάλλοντας στην διατήρηση των χωρικών διαστάσεων (ύψος και πλάτος) εισόδου ώστε να εφαρμοστεί συγκεκριμένο μέγεθος φίλτρου για τη σάρωση των χαρακτηριστικών [37].

Για παράδειγμα όταν η συνέλιξη εφαρμοστεί σε μια έγχρωμη εικόνα εισόδου $32 \times 32 \times 3$ (RGB) με φίλτρο $5 \times 5 \times 3$, το αποτέλεσμα της εξόδου θα είναι $28 \times 28 \times 3$ με τις χωρικές διαστάσεις (ύψος και πλάτος) να μειώνονται. Στην παρακάτω εικόνα 19 γίνεται αναπαράσταση της συνέλιξης της εικόνας εισόδου και του φίλτρου.



Εικόνα 19 Αναπαράσταση συνέλιξης

Στις εικόνες 20, 21, 22 απεικονίζεται η διαδικασία της συνέλιξης σε μια εικόνα διαστάσεων 5x5x3 με τη χρήση ενός φίλτρου 3x3x3 και σαν αποτέλεσμα εξόδου, μια εικόνα 3x3x3. Η διαδικασία της συνέλιξης πραγματοποιείται μεταξύ των τιμών της εικόνας και του φίλτρου σε κάθε περιοχή. Αφού γίνει ο υπολογισμός της συνέλιξης για την περιοχή με το φίλτρο, το φίλτρο μετατοπίζεται κατά βήμα 1 στην επόμενη περιοχή.

Στο συγκεκριμένο παράδειγμα ο υπολογισμός της συνέλιξης για την κίτρινη περιοχή γίνεται με τον πολλαπλασιασμό των τιμών της εικόνας (A πίνακας) με το φίλτρο (B πίνακας) και με αποτέλεσμα την εικόνα μετά την συνέλιξη (Γ πίνακας).

Ο υπολογισμός της συνέλιξης με το φίλτρο για την συγκεκριμένη περιοχή και η τιμή της εικόνας στην θέση Γ(0,0) μετά την συνέλιξη θα είναι:

$$\begin{aligned}
 A(0,0) \times B(0,0) &= 1 \times 1 = 1 \\
 A(0,1) \times B(0,1) &= 1 \times 0 = 0 \\
 A(0,2) \times B(0,2) &= 1 \times 1 = 1 \\
 A(1,0) \times B(1,0) &= 0 \times 0 = 0 \\
 A(1,1) \times B(1,1) &= 1 \times 1 = 1 \\
 A(1,2) \times B(1,2) &= 1 \times 0 = 0 \\
 A(2,0) \times B(2,0) &= 0 \times 1 = 0 \\
 A(2,1) \times B(2,1) &= 0 \times 0 = 0 \\
 A(2,2) \times B(2,2) &= 1 \times 1 = 1
 \end{aligned}$$

$$\begin{aligned}
 \Gamma(0,0) &= A(0,0) \times B(0,0) \\
 &+ A(0,1) \times B(0,1) \\
 &+ A(0,2) \times B(0,2) \\
 &+ A(1,0) \times B(1,0) \\
 &+ A(1,1) \times B(1,1) \\
 &+ A(1,2) \times B(1,2) \\
 &+ A(2,0) \times B(2,0) \\
 &+ A(2,1) \times B(2,1) \\
 &+ A(2,2) \times B(2,2) \\
 &= 1 + 0 + 1 + 0 + 1 + 0 + 0 + 0 + 0 + 1 = 4
 \end{aligned}$$

A

	0	1	2	3	4
0	1 <small>x1</small>	1 <small>x0</small>	1 <small>x1</small>	0	0
1	0 <small>x0</small>	1 <small>x1</small>	1 <small>x0</small>	1	0
2	0 <small>x1</small>	0 <small>x0</small>	1 <small>x1</small>	1	1
3	0	0	1	1	0
4	0	1	1	0	0

Image 5x5x3

Γ

	0	1	2
0	4		
1			
2			

Convolved feature

B

	0	1	2
0	1	0	1
1	0	1	0
2	1	0	1

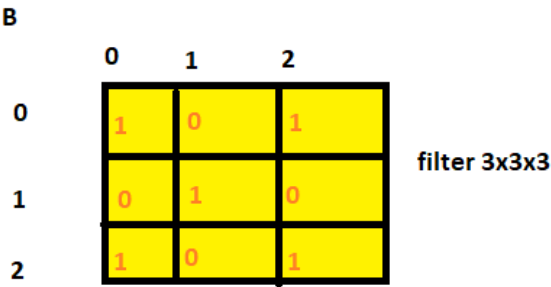
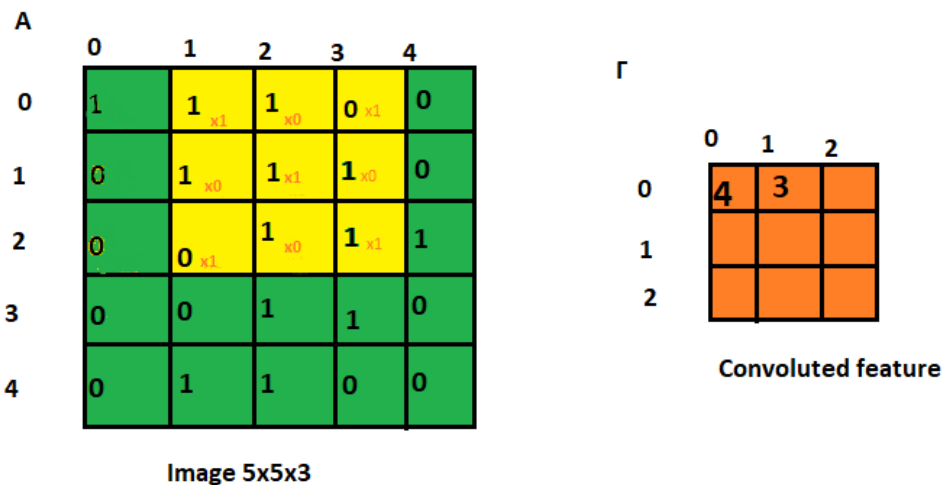
filter 3x3x3

Εικόνα 20 Αναπαράσταση συνέλιξης σε εικόνα 5x5 για την θέση Γ(0,0).

Στη συνέχεια υλοποιείται η συνέλιξη στην επόμενη περιοχή της εικόνας (μετατόπιση βήματος κατά 1) όπως φαίνεται και από το παρακάτω σχήμα με την τιμή της τελικής εικόνας στην θέση $\Gamma(0,1)$ να υπολογίζεται:

$$\begin{aligned}
 A(0,1) \times B(0,0) &= 1 \times 1 = 1 \\
 A(0,2) \times B(0,1) &= 1 \times 0 = 0 \\
 A(0,3) \times B(0,2) &= 0 \times 1 = 0 \\
 A(1,1) \times B(1,0) &= 1 \times 0 = 0 \\
 A(1,2) \times B(1,1) &= 1 \times 1 = 1 \\
 A(1,3) \times B(1,2) &= 1 \times 0 = 0 \\
 A(2,1) \times B(2,0) &= 0 \times 1 = 0 \\
 A(2,2) \times B(2,1) &= 1 \times 0 = 0 \\
 A(2,3) \times B(2,2) &= 1 \times 1 = 1
 \end{aligned}$$

$$\begin{aligned}
 \Gamma(0,1) &= A(0,1) \times B(0,0) \\
 &+ A(0,2) \times B(0,1) \\
 &+ A(0,3) \times B(0,2) \\
 &+ A(1,1) \times B(1,0) \\
 &+ A(1,2) \times B(1,1) \\
 &+ A(1,3) \times B(1,2) \\
 &+ A(2,1) \times B(2,0) \\
 &+ A(2,2) \times B(2,1) \\
 &+ A(2,3) \times B(2,2) \\
 &= 1 + 0 + 0 + 0 + 1 + 0 + 0 + 0 + 0 + 1 = 3
 \end{aligned}$$

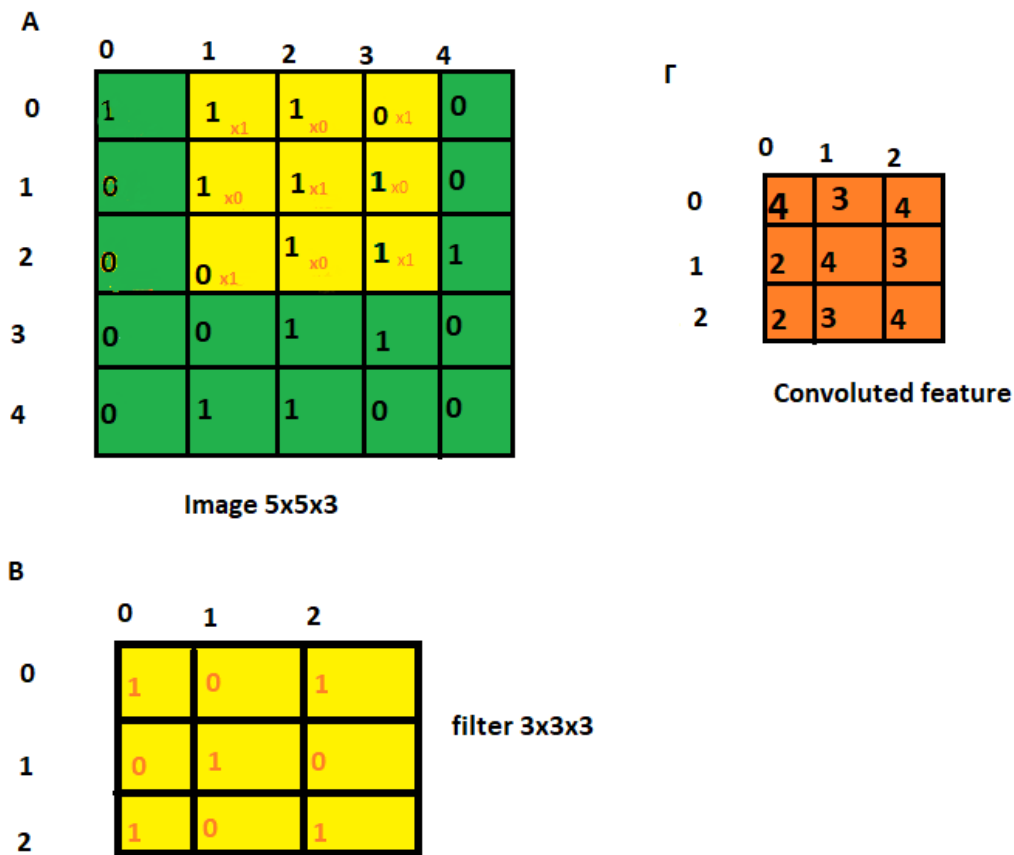


Εικόνα 21 Αναπαράσταση συνέλιξης σε εικόνα 5x5 για την θέση $\Gamma(0,1)$.

Με τον ίδιο ακριβώς τρόπο θα γίνει ο υπολογισμός της συνέλιξης στην τελική εικόνα και για τις υπόλοιπες θέσεις:

$$\begin{aligned}
 \Gamma(0,2) &= 4 & \Gamma(1,0) &= 2 & \Gamma(1,1) &= 4 & \Gamma(1,2) &= 3 \\
 \Gamma(2,0) &= 2 & \Gamma(2,1) &= 3 & \Gamma(2,2) &= 4 & &
 \end{aligned}$$

Στην εικόνα 22 γίνεται αναπαράσταση της τελικής εικόνας μετά τον υπολογισμό της συνέλιξης:



Εικόνα 22 Αναπαράσταση συνέλιξης για την τελική εικόνα.

2.1.2 Επίπεδο Υποδειγματοληψίας

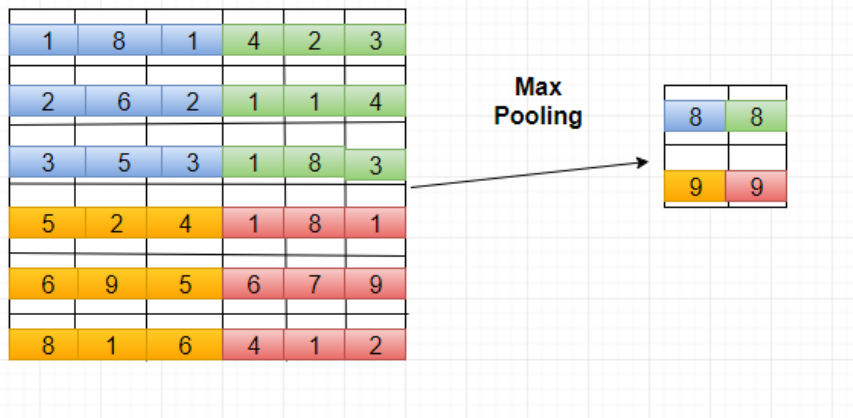
Η διαφορά του συνελκτικού επιπέδου με το επίπεδο υποδειγματοληψίας είναι ότι και τα δύο επίπεδα μειώνουν τις διαστάσεις τις εικόνας αλλά το συνελκτικό επίπεδο ανιχνεύει μοτίβα σε υποπεριοχές της εικόνας εισόδου ενώ το επίπεδο υποδειγματοληψίας χρησιμεύει στη σταδιακή μείωση της εικόνας εισόδου, στη μείωση του χρόνου εκπαίδευσης του δικτύου και στην αποφυγή της υπερεκπαίδευσης.

Το επίπεδο υποδειγματοληψίας συνήθως βρίσκεται ενδιάμεσα σε δύο συνελκτικά επίπεδα. Χρησιμοποιείται για τη μείωση των χωρικών διαστάσεων (ύψος και πλάτος) της εικόνας και δέχεται παραμέτρους για τη δήλωση του μεγέθους του φίλτρου και του βήματος σάρωσης.

Κάποιες συναρτήσεις για την επίτευξη της χωρικής υποδειγματοληψίας που υπάρχουν είναι: η εξαγωγή του τοπικού μεγίστου από τις περιοχές (Max Pooling), η εξαγωγή του μέσου όρου των περιοχών (Average Pooling), καθώς και η στοχαστική χωρική υποδειγματοληψία (Stochastic Pooling).

- i. **Max Pooling:** η λειτουργία των επιπέδων Max Pooling βασίζεται στην εξαγωγή της μέγιστης τιμής που εντοπίζεται σε μια περιοχή. Η εξαγωγή του τοπικού μεγίστου βρέθηκε πως υπερισχύει έναντι της εξαγωγής του μέσου όρου που συζητιέται παρακάτω, διότι επιτρέπει στα πιο δυνατά χαρακτηριστικά να διαρρέουν εντός του δικτύου [38].

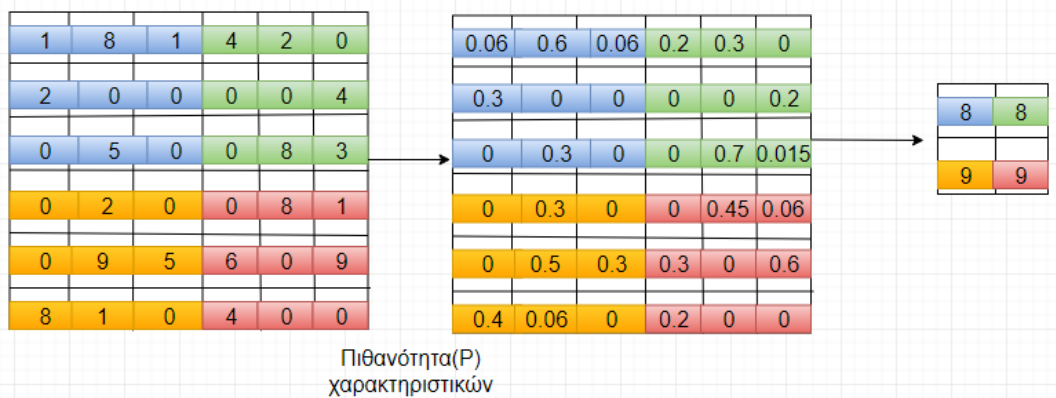
Στην εικόνα 24 απεικονίζεται η διαδικασία εύρεσης του τοπικού μεγίστου όπου από κάθε περιοχή επιλέγεται το τοπικό μέγιστο. Στο συγκεκριμένο παράδειγμα, η εικόνα εισόδου είναι μια εικόνα 6x6 και χωρίζεται σε 4 υπο-περιοχές για να προκύψει η τελική εικόνα 2x2 με τις μέγιστες τιμές από την κάθε υπο-περιοχή της εικόνας εισόδου.



Εικόνα 23 Διαδικασία εύρεσης τοπικού μεγίστου.

- ii. **Average Pooling:** η συγκεκριμένη συνάρτηση χρησιμοποιεί τον μέσο όρο της περιοχής όπου θα γίνει η υποδειγματοληψία.
- iii. **Stochastic Pooling:** η στοχαστική υποδειγματοληψία κατά την διάρκεια της εκπαίδευσης των ΣΝΔ δειγματοληπτεί πιθανοτικά την κάθε υπο-περιοχή εισόδου και πραγματοποιεί κανονικοποίηση του μοντέλου (model averaging) [39].

Στην εικόνα 24 απεικονίζεται η διαδικασία της στοχαστικής υποδειγματοληψίας σε μια εικόνα 6x6 για τα χαρακτηριστικά περιοχών με διαφορετικές πιθανότητες εμφάνισης. Η τελική εικόνα 2x2 που προκύπτει είναι οι τιμές που είχαν την μεγαλύτερη πιθανότητα στην κάθε περιοχή.



Εικόνα 24 Αναπαράσταση στοχαστικής υποδειγματοληψίας για χαρακτηριστικά που βρίσκονται σε διαφορετικές περιοχές με εμφάνιση διαφορετικών πιθανοτήτων ως επιλογή για το επόμενο επίπεδο.

2.1.3 Πλήρως συνδεδεμένα επίπεδα

Τα πλήρως συνδεδεμένα επίπεδα τοποθετούνται στα τελικά επίπεδα ενός συνελκτικού δικτύου. Κάθε κόμβος των πλήρως συνδεδεμένων επιπέδων συνδέεται με όλους τους κόμβους του προηγούμενου επιπέδου όπως και στα MLP. Η σύνδεσης των κόμβων με τους κόμβους του προηγούμενου επιπέδου συνδράμει στη μάθησης των pixels όλης της εικόνας [40].

2.1.4 Τοπικά συνδεδεμένα δίκτυα

Μια απλή λύση στο πρόβλημα είναι τα τοπικά συνδεδεμένα δίκτυα (locally connected networks) τα οποία περιορίζουν τις συνδέσεις μεταξύ των κόμβων εισόδου και των κόμβων στα ενδιάμεσα επίπεδα [41].

2.2 Αλγόριθμος Backpropagation στα συνελκτικά δίκτυα

Όπως στα MLP η εκπαίδευση γίνεται με την χρήση του αλγορίθμου Backpropagation για την ενημέρωση των βαρών από το επίπεδο εξόδου προς το επίπεδο εισόδου, έτσι και στα ΣΝΔ, το σφάλμα κατηγοριοποίησης διαδίδεται από τα πλήρως συνδεδεμένα επίπεδα, στα αραιότερα συνδεδεμένα επίπεδα φτάνοντας στο επίπεδο εισόδου. Ιδιαίτερα σημαντικό είναι ότι ο κάθε κόμβος λαμβάνει το σφάλμα μόνο από τους κόμβους που συμμετέχουν στην διαδικασία της συνέλιξης και η προώθηση γίνεται μέσω των βαρών. Στην συγκεκριμένη περίπτωση το σφάλμα $\delta^{m,\lambda-1}_{x,y}$, είναι το σφάλμα που αντιστοιχεί στο pixel m στη θέση x,y για το επίπεδο λ-1. Το $k^{nm}_{i,j}$ είναι το pixel που βρίσκεται στην θέση i,j και ενώνει τα pixel m,n, f η συνάρτηση ενεργοποίησης, k_o το πλήθος των pixels, M_o είναι το μέγεθος των εικόνων και N είναι το πλήθος των εικόνων του επιπέδου λ.

$$\delta^{m,\lambda-1}_{x,y} = \sum_{n=1}^N \sum_{i=d}^x \sum_{j=g}^y \delta^{n,\lambda}_{i,j} k^{nm}_{x-i,y-j} f'(u^{n,\lambda}_{i,j})$$

$$d = \max(0, x - k_o + 1), g = \max(0, y - k_o + 1)$$

Στον παρακάτω τύπο ορίζεται η συνάρτηση ενημέρωσης των βαρών:

$$k^n_{x,y} = k^0_{i,j} - e \left(\sum_{i=0}^{M_o} \sum_{j=0}^{M_o} \delta^\lambda_{i,j} a^{\lambda-1}_{i+x,j+y} \right)$$

Πολύ σημαντικό στοιχείο του Backpropagation στα ΣΝΔ είναι το upscaling. Αν έχει εκτελεστεί υποδειγματοληψία για την καλύτερη αντιστοίχιση των σφαλμάτων για το Backpropagation, θα πρέπει οι διαστάσεις της εικόνας να επανέλθουν στο αρχικό τους μέγεθος πριν την υποδειγματοληψία. Αυτό πραγματοποιείται με τους εξής τρόπους:

- Στην υποδειγματοληψία με Average Pooling, το σφάλμα του κάθε pixel μοιράζεται στα γειτονικά του.

- Στην υποδειγματοληψία με Max Pooling, το σφάλμα του κάθε pixel , δίνεται στο pixel που είχε το μεγαλύτερο σφάλμα κατά την διαδικασία της προώθησης με τα υπόλοιπα pixel. Αυτή η διαδικασία είναι πιο πολύπλοκη διότι θα πρέπει να είναι γνωστή η θέση του pixel με το μεγαλύτερο σφάλμα.

Παρατηρείται ότι το σφάλμα μειώνεται όσο γίνεται η προώθηση του στα αρχικά επίπεδα του δικτύου και το upscaling δεν μπορεί να δώσει σωστές τιμές σφάλματος στα pixel. Αυτό έχει σαν αποτέλεσμα η εκπαίδευση του δικτύου να συγκλίνει αργά και να μειώνονται φαινόμενα overfitting.

2.3 Γενικοποίηση Δικτύου

Ο πιο σημαντικός τρόπος επίτευξης ενός δικτύου γενικοποίησης (generalized) είναι η αποφυγή της υπερεκπαίδευσης για ένα συγκεκριμένο σύνολο δεδομένων. Η λύση στο συγκεκριμένο πρόβλημα είναι η στοχαστικότητα. Η τεχνική Dropout από τους Hinton, et al.[17] είχε σκοπό την αποφυγή σύνθετων αρχιτεκτονικών στα δεδομένα εκπαίδευσης με την απενεργοποίηση ενός συγκεκριμένου ποσοστού που ορίζεται στους κόμβους κάθε ενδιάμεσου επιπέδου. Στη συνέχεια ένας ακόμη τρόπος μείωσης του σφάλματος στα δεδομένα ελέγχου, είναι η τυχαία απόρριψη κάποιων κόμβων στα επίπεδα του δικτύου με την χρήση της διαδικασίας Dropout. Ο τύπος της διαδικασίας Dropout είναι ο παρακάτω, όπου r είναι η έξοδος του αντίστοιχου επιπέδου, m είναι ο αριθμός των κόμβων και f η συνάρτηση ενεργοποίησης με τα βάρη των κόμβων W_u :

$$r = m f(W_u)$$

Μια άλλη τεχνική είναι η DropConnect που χρησιμοποιήθηκε από τους Wan, et al. [18]. Είναι η εξέλιξη της τεχνικής Dropout, καθώς αντί να επιδιώκει την απενεργοποίηση κάθε κόμβου στα επίπεδα με ένα ποσοστό, απενεργοποιεί τις συνδέσεις μεταξύ των κόμβων. Και οι δύο τεχνικές έχουν πολλά κοινά στοιχεία μεταξύ τους καθώς εισάγουν μια αραιώση (sparsity) μέσα στο μοντέλο. Η τεχνική DropConnect εισάγει αυτή την αραιώση στα βάρη και όχι στις τιμές εξόδου. Έτσι το δίκτυο στην αρχή της εκπαίδευσης του έχει πλήρως συνδεδεμένα επίπεδα και στη συνέχεια έχει αραιότερα συνδεδεμένα επίπεδα.

2.4 Τεχνικές επεξεργασίας δεδομένων εικόνας

Οι συγκεκριμένες τεχνικές αφορούν τα δεδομένα εισόδου που εφαρμόζονται είτε σε εικόνες κλίμακας γκρι (grayscale) με αναπαράσταση γραμμάτων είτε σε έγχρωμες εικόνες τριών διαστάσεων (RGB) .

Επεξεργασία εικόνας κλίμακας γκρι. Η συγκεκριμένη επεξεργασία αφορά εικόνες με αναπαράσταση συμβόλων που κύριο στόχο έχει τα σύμβολα να τοποθετηθούν στο κέντρο της εικόνας πλαισίου και, όπου δεν υπάρχουν σύμβολα στην υπόλοιπη εικόνα, να υπάρχουν μηδενικές τιμές. Στην αρχική εικόνα μπορεί να γίνει αύξηση των διαστάσεων της με την προσθήκη θορύβου ή την μεταφορά εικονοστοιχείων σε τυχαίες θέσεις [42].

Επεξεργασία εικόνας με αντικείμενα. Βασίζεται στην εξαγωγή των τιμών από τα εικονοστοιχεία, μειώνοντας το πλήθος των καναλιών των έγχρωμων εικόνων (RGB) από 3 κανάλια σε 1 κανάλι κλίμακας γκρι. Μια άλλη διαδικασία για την καλύτερη εκπαίδευση του δικτύου είναι ο υπολογισμός της μέσης τιμής και η αφαίρεση της από την κάθε εικόνα έτσι ώστε να έχει μηδενική μέση τιμή [19].

Κεφάλαιο 3 Βιβλιογραφική Επισκόπηση

Στο κεφάλαιο αυτό περιγράφονται κάποιες σύγχρονες εργασίες που έχουν υλοποιηθεί στα MLP και στα συνελκτικά δίκτυα. Αναφέρονται τεχνικές που βελτιστοποιούν την απόδοση των δικτύων στα σύνολα δεδομένων MNIST, CIFAR-10 και CIFAR-100 [3],[4]. Η ταξινόμηση εικόνας με την χρήση νευρωνικών δικτύων αποτελεί μια από τις σημαντικότερες και πιο συνήθεις εφαρμογές τα τελευταία χρόνια. Έχει δοκιμαστεί πληθώρα αλγορίθμων και προσεγγίσεων με σκοπό την επίτευξη καλύτερων αποτελεσμάτων ακρίβειας με τη χρήση νευρωνικών δικτύων και μηχανών διανυσμάτων υποστήριξης.

3.1 Βασικές εργασίες

Η εργασία των Rumelhart et al. [23] αναφέρεται στην επιτάχυνση της εκπαίδευσης του δικτύου με χρήση του μετασχηματισμού Fourier. Ο μετασχηματισμός Fourier μπορεί να χρησιμοποιηθεί στα επίπεδα όπου υπάρχει συνέλιξη έτσι ώστε να μετατραπεί σε πολλαπλασιασμό και πρόσθεση αλλά δεν εξαρτάται από την συνέλιξη και πιο συγκεκριμένα από τον πυρήνα της.

Το 2012 οι Krizhevsky, Sutskever και Hinton στην εργασία τους [9] είχαν σκοπό την αναγνώριση και ταξινόμηση αντικειμένων και προτύπων σε έγχρωμες εικόνες. Το μοντέλο ήταν ένα συνελκτικό δίκτυο με τη χρήση επιπέδων μέγιστης υποδειγματοληψίας και την σύνδεση τους με 3 πλήρως συνδεδεμένα επίπεδα. Τα σύνολα δεδομένων που χρησιμοποιήθηκαν ήταν τα ILSVRC-2010 και ILSVRC-2012 και με την εφαρμογή του παραπάνω μοντέλου επιτεύχθηκαν τα καλύτερα ποσοστά ακρίβειας και σφάλματος 65 %.

Το 2013 οι Zeiler, Matthew, και Fergus στην εργασία τους [20] πέτυχαν καλύτερο ποσοστό ακρίβειας κατά 1,7% και 1,6% λιγότερο σε ποσοστό σφάλματος για την κατηγοριοποίηση 1000 εικόνων. Τον προηγούμενο χρόνο (2012) οι Krizhevsky, Sutskever και Hinton στον διαγωνισμό ImageNet πέτυχαν ακρίβεια 65%. Η τεχνική που χρησιμοποιήθηκε ήταν η αποσυνέλιξη που υλοποιείται από ένα συνελκτικό δίκτυο και χρησιμοποιεί το επίπεδο συνέλιξης και το επίπεδο υποδειγματοληψίας αλλά από την αντίθετη φορά, έτσι ώστε αντί να κάνει απεικόνιση των pixel στα χαρακτηριστικά, κάνει το αντίθετο. Με την αναγνώριση προτύπων και αντικειμένων ασχολήθηκαν στην εργασία τους οι Girshick et al. [10]. Καταγράφηκαν αποτελέσματα κατά 30% πιο ακριβή σε σχέση με τις ταξινομήσεις που είχαν γίνει στο σύνολο δεδομένων VOC 2012. Το μοντέλο ήταν ένα συνελκτικό δίκτυο με τη χρήση ταξινομητή SVM πετυχαίνοντας μεγάλα ποσοστά ακρίβειας έναντι των παλαιότερων μεθόδων εκμάθησης στα σύνολα δεδομένων όπως το PASCAL VOC.

Μια διαφορετική προσέγγιση στην εργασία τους είχαν οι Gong, et al. [21] με τη χρήση συνελκτικών δικτύων για την κατηγοριοποίηση εικόνας. Χρησιμοποίησαν τη συνάρτηση κόστους WARP στο σύνολο δεδομένων NUS-WIDE επιτυγχάνοντας καλύτερα ποσοστά ακρίβειας κατά 10%.

Ακόμη μια εργασία των Huang, et al. [14] πραγματοποιήθηκε την ανίχνευση και την αναγνώριση χαρακτηριστικών εικόνας με νευρωνικά δίκτυα σε συνδυασμό ενός ταξινομητή SVM. Το συνελκτικό δίκτυο εκπαιδεύεται για την ανίχνευση απλών αντικειμένων εικόνας και ο πυρήνας Gaussian-Kernel SVM εκπαιδεύεται από την έξοδο του προτελευταίου επιπέδου του δικτύου. Τα αποτελέσματα που προκύπτουν αφορούν την αναγνώριση αντικειμένων 6

κατηγοριών (άνθρωποι, αεροπλάνα, φορτηγά, κ.α. Στο δοκιμαστικό σύνολο δεδομένων το οποίο περιέχει διαφορετικές περιπτώσεις αντικειμένων σε σχέση με το σύνολο εκπαίδευσης, ο SVM ταξινομητής αποδίδει ποσοστό σφάλματος 43,3% , το συνελκτικό δίκτυο πετυχαίνει ποσοστό 7,2% και ο SVM ταξινομητής στο τελευταίο επίπεδο επιτυγχάνει 5,9%. Με το ίδιο σύνολο δεδομένων ασχολήθηκαν οι Chen et al. στην εργασία τους [12] εφαρμόζοντας την δική τους μέθοδο σε ένα νευρωνικό δίκτυο ενός επιπέδου και μία μη γραμμική συνάρτηση ενεργοποίησης με την ονομασία «extreme learning machine». Πριν την εκπαίδευση του δικτύου προηγήθηκε μια επεξεργασία των δεδομένων για την εξαγωγή φασματικών και χωρικών χαρακτηριστικών με τα φίλτρα Gabor.

Στη συνέχεια, ο Tang [25] χρησιμοποιεί στο ανώτερο επίπεδο τον ταξινομητή του τετραγωνικού αθροίσματος των αραιών μεταβλητών (L2-SVM) [27] για την εκπαίδευση των νευρωνικών δικτύων με κύριο στόχο το επίπεδο εξόδου να συμβάλει στην μάθηση των παραμέτρων των κατώτερων επιπέδων. Τα βάρη των κατώτερων επιπέδων εκπαιδεύονται με την τεχνική Backpropagation εξαρτώμενα από το υψηλότερο επίπεδο. Το προτελευταίο επίπεδο πρέπει να συμβάλει όπου η είσοδος X αντικαθίσταται με την ενεργοποίηση h του προτελευταίου επιπέδου. Σε ένα διαγωνισμό ICML το 2013, το σύνολο δεδομένων που ασχολήθηκε ο Tang αποτελούνταν από 28.709 εικόνες 48x48 από εκφράσεις προσώπων. Η καλύτερη λύση στο πρόβλημα επιτεύχθηκε με ποσοστά στη μέση ακρίβεια εγκυρότητας (cross validation accuracy) με πάνω από 8 χωρίσματα (splits) με ποσοστό 69.4% και στο test score με ποσοστό 71.2% μεγαλύτερο κατά 2% από την δεύτερη ομάδα σε κατάταξη στον διαγωνισμό. Το μοντέλο του περιλαμβάνει τη χρήση ενός απλού συνελκτικού δικτύου με την χρήση γραμμικών διανυσμάτων υποστήριξης one vs all στην κορυφή του. Τα ποσοστά ακρίβειας που προέκυψαν ήταν με τη χρήση μηχανών διανυσμάτων υποστήριξης L2-SVM και softmax .

Στον πίνακα 1 αναφέρονται τα ποσοστά ακρίβειας που πέτυχε ο Tang σε διαγωνισμούς Kaggle, ICML.

Πίνακας 1 Ποσοστά ακρίβειας

	Softmax	L2-SVM
Ποσοστό ακρίβειας	67.6%	68.9%
Ποσοστό νικητή Kaggle	69.3%	69.4%
Ποσοστό νικητή διαγωνισμού ICML	70.1%	71.2%

Ο Tang ακόμα υλοποίησε πειράματα σε σύνολα δεδομένων MNIST και CIFAR-10 με την χρήση ενός νευρωνικού μοντέλου για το MNIST και ενός συνελκτικού δικτύου για το CIFAR-10 με την χρήση μηχανών διανυσμάτων υποστήριξης L2-SVM στο επίπεδο εξόδου του δικτύου.

MNIST. Ο σχεδιασμός που ακολουθήθηκε ήταν ένα απλό νευρωνικό μοντέλο όπου σαν είσοδο δέχεται 784 κόμβους και ακολουθούν 2 ενδιάμεσα επίπεδα όπου το καθένα έχει 512 κόμβους και στη συνέχεια μια συνάρτηση softmax ή ένας μηχανισμός διανυσμάτων υποστήριξης L2-SVM.. Η εκπαίδευση έγινε με το αλγόριθμο βαθμωτής καθόδου στα 300 minibatches για 400 εποχές με συνολικά 120.000 ενημερώσεις των βαρών. Για την επίτευξη καλύτερων αποτελεσμάτων, ο Gaussian θόρυβος τοποθετείται στη αρχή του δικτύου. Η τοποθέτηση του θορύβου στην αρχή για καλύτερα αποτελέσματα έχει παρθεί από τις εργασίες [4], [5]. Τα αποτελέσματα που επιτυγχάνονται με την χρήση softmax 0.99 και με την χρήση μηχανών διανυσμάτων υποστήριξης 0.87 .

CIFAR-10. Ο σχεδιασμός του μοντέλου που ακολουθήθηκε ήταν ένα απλό συνελικτικό δίκτυο με δύο επίπεδα υποδειγματοληψίας και ένα επίπεδο συνέλιξης. Το πρώτο επίπεδο περιέχει 325×5 φίλτρα με την εφαρμογή της συνάρτησης ενεργοποίησης ReLU στο ενδιάμεσο επίπεδο και το δεύτερο επίπεδο περιέχει $64 \times 5 \times 5$ φίλτρα. Και τα δύο στρώματα υποδειγματοληψίας χρησιμοποιούν την μέγιστη υποδειγματοληψία. Το προτελευταίο επίπεδο περιέχει 3072 κόμβους και χρησιμοποιεί την συνάρτηση ενεργοποίησης RELU με ρυθμό απόρριψης (Dropout) ποσοστό 0.2. Τα ποσοστά σφάλματος με την χρήση softmax είναι 14.0% ενώ με την χρήση μηχανών διανυσμάτων υποστήριξης το ποσοστό ελαχιστοποιείται στο 11.9% . Από τις παρατηρήσεις του, προέκυψε ότι όσο μειώνεται η εντροπία τόσο αυξάνεται το σφάλμα στο μεσαίο επίπεδο.

3.2 Επιπλέον εργασίες

Με την εργασία των Hinton, Osindero και Teh [8] ξεκίνησε μια πολύ σημαντική ανάπτυξη στον σχεδιασμό των νευρωνικών δικτύων με αρχιτεκτονικές βαθιάς μάθησης (Deep learning). Με τη χρήση ενός άπληστου αλγορίθμου (greedy) η ενημέρωση των βαρών πραγματοποιήθηκε γρήγορα. Το μοντέλο του δικτύου και η ύπαρξη τριών ενδιάμεσων επιπέδων αποτέλεσε ένα πολύ καλό γενικό μοντέλο για την ταξινόμηση συνόλων δεδομένων που αποτελούνταν από ψηφία και σύμβολα. Αυτές οι αρχιτεκτονικές επιτέλεσαν σπουδαίο ρόλο στο χώρο των νευρωνικών δικτύων με δυνατότητες εξαγωγής πιο σύνθετων χαρακτηριστικών και είχε σαν αποτέλεσμα τον επαναπροσδιορισμό των μεθόδων ταξινόμησης στα υπερφασματικά (hyper-spectral) δεδομένα εικόνων.

Σε μια άλλη εργασία των Sigtia και Dixon [13] που ασχολήθηκαν με την μουσική, ένα σημαντικό πρόβλημα των νευρωνικών δικτύων ήταν ότι μεγάλα σύνολα δεδομένων ο χρόνος εκπαίδευσης τους ήταν απαγορευτικός. Εξετάστηκαν τρόποι για βελτίωση της μάθησης σε ένα σύνολο ηχητικών δεδομένων και μια από αυτές τις βελτιώσεις, ήταν η χρήση της ReLU αντί της σιγμοειδούς συνάρτησης. Στην συνέχεια, για την κανονικοποίηση των δεδομένων χρησιμοποιήθηκε η τεχνική της απόρριψης (Dropout) για την βελτιστοποίηση της εκπαίδευσης των σιγμοειδών κόμβων στο δίκτυο. Η μέθοδος απόρριψης παρέχει σημαντικές βελτιώσεις στο χρόνο εκπαίδευσης του δικτύου πετυχαίνοντας ποσοστό ακρίβειας 83% στο σύνολο δεδομένων GTZAN.

Στη συνέχεια, με την εργασία των Yadan et al. [22] γίνεται αναφορά στην επιτάχυνση της εκπαίδευσης του δικτύου από τη χρήση μονάδων επεξεργασίας εικόνων (GPU). Η συγκεκριμένη μονάδα επεξεργασίας περιλαμβάνει έναν επεξεργαστή, κυκλώματα μνήμης τυχαίας προσπέλασης (RAM) και έναν συνδυασμό πολλών πυρήνων κάνοντας παραλληλοποίηση της διαδικασίας της εκπαίδευσης. Το πιο σημαντικό επίτευγμα είναι η χρήση μονάδων επεξεργασίας εικόνων (GPU) που επιτρέπει περίπλοκες τεχνικές παραλληλοποίησης είτε σε επίπεδο του υπάρχοντος μοντέλου είτε σε επίπεδο της πληροφορίας. Στο επίπεδο μοντέλου υπήρχε διαχωρισμός των επιπέδων του νευρωνικού δικτύου σε 2 μονάδες επεξεργασίας εικόνων (GPU) ενώ στο επίπεδο πληροφορίας, οι εικόνες χωρίστηκαν και αυτές σε 2 GPU. Το αποτέλεσμα ήταν να υπάρχει παραλληλοποίηση σε 4 GPU με απόδοση μεγαλύτερη κατά 2,2 φορές από την ύπαρξη μιας GPU.

3.3 Συγκρίσεις συνόλων δεδομένων

Παρακάτω αναφέρονται αποτελέσματα εργασιών σε σύνολα δεδομένων που έχουν χρησιμοποιηθεί στην παρούσα εργασία με τη χρήση διάφορων τεχνικών και μεθόδων.

3.2.1 MNIST

Οι Wan et al. στην εργασία τους [18] με την χρήση της τεχνικής DropConnect απενεργοποιούν τις συνδέσεις μεταξύ των κόμβων με αποτέλεσμα να υπάρχει μια αραιώση (sparsity) στο μοντέλο. Η τεχνική αυτή χρησιμοποιείται στις συνδέσεις των βαρών και όχι στις τιμές εξόδου του μοντέλου με αποτέλεσμα το νευρωνικό δίκτυο κατά την εκπαίδευση του να έχει αραιά συνδεδεμένα επίπεδα. Η αρχιτεκτονική περιλαμβάνει δύο πλήρως συνδεδεμένα επίπεδα που το καθένα είχε 800 κόμβους. Η είσοδος είχε τα εικονοστοιχεία της εικόνας και η έξοδος τροφοδοτείται σε ταξινομητή softmax 10 κλάσεων. Στα πειράματα που έγιναν, χρησιμοποιήθηκαν συναρτήσεις ενεργοποίησης όπως ReLu, σιγμοειδής και tanh. Η τεχνική DropConnect μπορεί να είναι σχετικά αργή αλλά συμβάλει σε χαμηλότερο ποσοστό σφάλματος.

Μερικά από τα καλύτερα ποσοστά χαμηλότερους σφάλματος που προέκυψαν με την χρήση των συναρτήσεων ενεργοποίησης ReLu, σιγμοειδούς και tanh στα πειράματα που υλοποιήθηκαν, προκύπτουν στον πίνακα 2.

Πίνακας 2 Χαμηλότερα ποσοστά σφάλματος με την χρήση διαφορετικών συναρτήσεων ενεργοποίησης

Activation Function	Model	% Test Error
ReLu	DropConnect	1.12
Sigmoid	DropConnect	1.36
Tanh	DropConnect	1.35

Για την βελτίωση του αποτελέσματος της ταξινόμησης χρησιμοποιείται η συνάρτηση ενεργοποίησης ReLu σε ένα συνελικτικό δίκτυο δύο επιπέδων με το πρώτο επίπεδο να αποτελείται με 32 χαρακτηριστικά χάρτη και το δεύτερο επίπεδο με 64 χαρακτηριστικά. Το επίπεδο εισόδου αποτελείται από 784 κόμβους και το επίπεδο εξόδου αποτελείται από 150 κόμβους και σε όλο το δίκτυο εφαρμόζεται η τεχνική DropConnect. Το ποσοστό σφάλματος που επιτεύχθηκε ήταν 0.21% που αποδείχθηκε καλύτερο από το ποσοστό σφάλματος 0.23% που είχαν επιτύχει οι Ciresan et al. [32] μέχρι τότε. Οι Ciresan D.C et al. [33] στην εργασία τους υλοποίησαν πειράματα σε 5 αρχιτεκτονικές νευρωνικών δικτύων για να καταλήξουν στο καλύτερο ποσοστό σφάλματος 0.35% όταν οι Ranzato et al. [34] είχαν πετύχει ποσοστό σφάλματος 0.39% και οι Simard et al. [35] ποσοστό σφάλματος 0.40%. Στον παρακάτω πίνακα 3 αναφέρεται η αρχιτεκτονική και το ποσοστό σφάλματος που προκύπτει με διαφορετικό αριθμό ενδιάμεσων επιπέδων και κόμβων.

Πίνακας 3 Ποσοστά σφάλματος σε αρχιτεκτονικές που διαφέρουν στον αριθμό των ενδιάμεσων επιπέδων.

Hidden layer1	Hidden layer2	Hidden layer3	Hidden layer4	Hidden layer5	Hidden layer6	Hidden layer7	Hidden layer8	Hidden layer9	Output layer	% Test error
1000	500								10	0.49
1500	1000	500							10	0.46
2000	1500	1000	500						10	0.41
2500	2000	1500	1000	500					10	0.35
1000	1000	1000	1000	1000	1000	1000	1000	1000	10	0.44

3.2.2 CIFAR-10

Οι Wan et al. στην ίδια εργασία [18] με τη χρήση της τεχνικής DropConnect και με αρχιτεκτονική συνελκτικού δικτύου που περιλαμβάνει δύο επίπεδα συνέλιξης και δύο πλήρως συνδεδεμένα επίπεδα πέτυχαν ποσοστό σφάλματος 9.41%. Στην εργασία του ο Tang [25] χρησιμοποιεί στο ανώτερο επίπεδο τον ταξινομητή του τετραγωνικού αθροίσματος των αραιών μεταβλητών (L2-SVM) [27] για την εκπαίδευση με κύριο στόχο το τελευταίο επίπεδο να συμμετέχει στην μάθηση των κατώτερων επιπέδων. Τα βάρη που βρίσκονται στο τελευταίο επίπεδο εκπαιδεύονται με την τεχνική Backpropagation. Το μοντέλο ήταν ένα συνελκτικό δίκτυο που περιλάμβανε δύο επίπεδα υποδειγματοληψίας και ένα επίπεδο συνέλιξης. Το πρώτο επίπεδο αποτελούνταν από 325 x 5 φίλτρα και το δεύτερο επίπεδο 64 x 5 φίλτρα. Το προτελευταίο επίπεδο αποτελούνταν από 3072 κόμβους και η τεχνική απόρριψης DropOut με ποσοστό 20%. Με την χρήση της Softmax στο τελευταίο επίπεδο επιτυγχάνονται ποσοστά σφάλματος 14% και με την χρήση μηχανών διανυσμάτων υποστήριξης υπάρχει ελαχιστοποίηση του ποσοστού σφάλματος 11%.

3.2.3 CIFAR-100

Στην εργασία τους [36] οι Clevert D. et al. χρησιμοποιούν την εκθετική συνάρτηση (Exponential Linear Units-ELU) η οποία επιταχύνει την εκπαίδευση ενός νευρωνικού δικτύου και πετυχαίνει μεγάλα ποσοστά ταξινόμησης. Η αρχιτεκτονική τους ήταν ένα συνελκτικό δίκτυο που αποτελούνταν από 11 συνελκτικά επίπεδα, 1 επίπεδο υποδειγματοληψίας με φίλτρο 2 x 2 και χρήση της τεχνικής DropOut για την κανονικοποίηση του μοντέλου. Στο συγκεκριμένο συνελκτικό δίκτυο δοκιμάστηκε και η ReLU εκτός από την ELU, με καλύτερο ποσοστό σφάλματος να είναι της ELU με 28.75% έναντι της ReLU που πέτυχε ποσοστό 29%.

Κεφάλαιο 4 Πειραματική Αποτίμηση

Στο παρόν κεφάλαιο παρουσιάζονται αποτελέσματα πειραμάτων σε αναλυτικούς πίνακες του Multilayer Perceptron και του συνελικτικού δικτύου σε σύγκριση με τον SVM ταξινομητή στο επίπεδο εξόδου. Επίσης, παρουσιάζονται γραφικές παραστάσεις που απεικονίζουν την απόδοση των νευρωνικών δικτύων και του ταξινομητή SVM.

4.1 Βιβλιοθήκη Keras

Χρησιμοποιήθηκε η βιβλιοθήκη Keras [28] για την υλοποίηση μοντέλων βαθιών αρχιτεκτονικών με πολλαπλά ενδιάμεσα επίπεδα και μεγάλο αριθμό κόμβων. Είναι μια υψηλού επιπέδου διεπαφή προγραμματισμού (API), υλοποιημένη σε γλώσσα προγραμματισμού Python και εκτελείται πάνω από άλλες γνωστές βιβλιοθήκες όπως το Tensorflow, Theano και το Cntk.

Μερικά από τα χαρακτηριστικά της βιβλιοθήκης keras είναι τα εξής:

- Υποστηρίζει τόσο τα συνελικτικά δίκτυα όσο και τα αναδρομικά δίκτυα.
- Το μοντέλο του νευρωνικού δικτύου που σχεδιάζεται είναι μια ακολουθία αυτόνομων ενοτήτων (modules). Συγκεκριμένα, τα επίπεδα του δικτύου, οι συναρτήσεις κόστους (cost functions), οι βελτιστοποιητές (optimizers) και οι συναρτήσεις ενεργοποίησης (activation functions) είναι ανεξάρτητα στοιχεία μεταξύ τους και μπορούν να συνδυαστούν ώστε να δημιουργήσουν καινούρια μοντέλα.
- Οι καινούριες ενότητες είναι εύκολο να προστεθούν στην λειτουργικότητα των υπάρχουσών ενοτήτων που υπάρχουν στο μοντέλο [28].

4.2 Πειράματα

Στην υλοποίηση των πειραμάτων χρησιμοποιήθηκε η βιβλιοθήκη Keras και σύνολα δεδομένων όπως το MNIST, CIFAR-10, CIFAR-100 σε μοντέλα βαθιάς μηχανικής μάθησης όπως το Multilayer Perceptron και τα συνελικτικά δίκτυα, με τη χρήση ταξινομητή SVM στο επίπεδο εξόδου. Η αρχιτεκτονική του δικτύου μεταβάλλεται είτε προσθέτοντας επιπλέον ενδιάμεσα επίπεδα είτε τροποποιώντας τον αριθμό των κόμβων στα επίπεδα για την εύρεση του καλύτερου ποσοστού ακρίβειας των δεδομένων ελέγχου, του ταξινομητή SVM αλλά και της σύγκρισης των δύο ποσοστών.

4.2.1 Multilayer Perceptron (MLP)

Σύνολο Δεδομένων MNIST [3]. Το συγκεκριμένο σύνολο δεδομένων αποτελείται από 10 κλάσεις χειρόγραφων συμβόλων, 60.000 παραδείγματα συνόλου εκπαίδευσης και 10.000 παραδείγματα συνόλου δοκιμών. Τα ψηφία έχουν κανονικοποιηθεί και τοποθετηθεί στο κέντρο σε μια εικόνα σταθερού μεγέθους 28x28.



Εικόνα 25 Σύνολο δεδομένων MNIST Πηγή: Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition" Proceedings of the IEEE, 86(11):2278-2324, 1998.

Οι αρχικές ασπρόμαυρες εικόνες που προκύπτουν από το σύνολο δεδομένων του NIST¹ είναι κανονικοποιημένες ως προς το μέγεθος τους σε κλίμακα από 0 έως 1 και περιέχουν επίπεδα γκρι λόγω της τεχνικής που χρησιμοποιεί ο αλγόριθμος κατά της αλλοίωσης του χρώματος. Οι εικόνες τοποθετήθηκαν στο κέντρο του πλαισίου 28x28 βάση του υπολογισμού του συνόλου των εικονοστοιχείων .

Για το συγκεκριμένο σύνολο δεδομένων έχουμε στο επίπεδο εισόδου 28x28 =784 κόμβους, η έξοδος αποτελείται από 10 κλάσεις πρόβλημα ταξινόμησης πολλών κλάσεων. Η softmax συνάρτηση ενεργοποίησης χρησιμοποιείται στο επίπεδο εξόδου έτσι ώστε να μετατρέψει τις εξόδους σε τιμές πιθανοτήτων, και να επιτρέψει την επιλογή μιας από τις 10 κλάσεις ως πρόβλεψη εξόδου στο μοντέλο. Ακόμα, χρησιμοποιείται ο αλγόριθμος βελτιστοποίησης ADAM για την μάθηση των βαρών [30]. Μετά την εκπαίδευση του δικτύου, το τελευταίο επίπεδο με την softmax συνάρτηση αντικαθίσταται από τις μηχανές διανυσμάτων υποστήριξης όπου τροφοδοτούνται από του κόμβους του προτελευταίου επιπέδου. Με την συγκεκριμένη αρχιτεκτονική, το νευρωνικό δίκτυο ασχολείται με τις διαστάσεις που υπάρχουν στα επίπεδα μέχρι να γίνει η τροφοδότηση στα SVM σε λιγότερα επίπεδα και τα SVM να κάνουν τον διαχωρισμό σε υπερεπίπεδα.

Το καλύτερο ποσοστό ακρίβειας που επιτυγχάνεται με την χρήση ενός ενδιάμεσου επιπέδου και αριθμό κόμβων 500 για 10 επαναλήψεις, είναι για τα δεδομένα ελέγχου 0,9419 και του ταξινομητή svm 0,9849. Παρατηρήθηκε πως αυξάνοντας ή μειώνοντας τον αριθμό των κόμβων στο συγκεκριμένο επίπεδο, υπάρχει μείωση του ποσοστού ακρίβειας είτε στο MLP

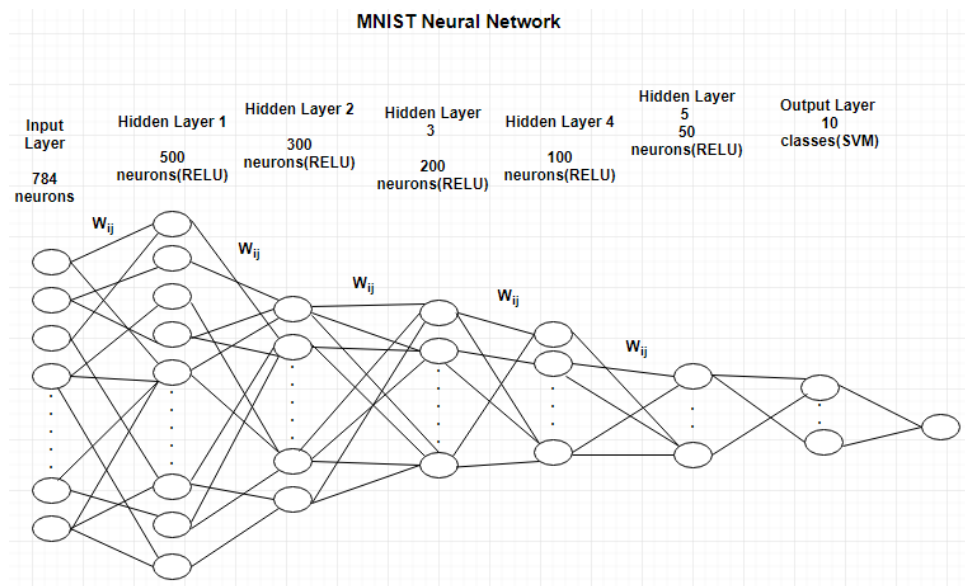
¹ NIST: National Institute of Standards and Technology

είτε με την χρήση του ταξινομητή SVM στο τελευταίο επίπεδο.

Με την προσθήκη ενός δεύτερου ενδιάμεσου επιπέδου και με συνολικό αριθμό κόμβων 300, τα ποσοστά ακρίβειας και στις δύο περιπτώσεις έχουν μειωθεί με ακρίβεια δεδομένων ελέγχου: 0,9380 και ακρίβεια ταξινομητή svm 0,98 . Όπως προτείνει ο Stathakis στο [31], ένα βαθύ νευρωνικό δίκτυο έχει την τάση ο αριθμός των κόμβων να είναι ίσος ή μικρότερος από το προηγούμενο επίπεδο και τα ποσοστά ακρίβειας να μειώνονται.

Στη συνέχεια προστέθηκε ένα τρίτο επίπεδο με αριθμό κόμβων 100 ως τον ιδανικότερο αριθμό κόμβων για αυτό το επίπεδο. Τα ποσοστά ακρίβειας που προκύπτουν για τα δεδομένα ελέγχου 0,9215 και για τον ταξινομητή svm 0,9750. Στην προσθήκη ενός τέταρτου επιπέδου με 25 κόμβους τα καλύτερα ποσοστά ακρίβειας για τα δεδομένα ελέγχου είναι 0,926 και για τον ταξινομητή svm είναι 0,94, με τα ποσοστά να έχουν μειωθεί ελάχιστα σε σχέση με το προηγούμενο βήμα. Τέλος, το συγκεκριμένο σύνολο δεδομένων δοκιμάστηκε στην προσθήκη και πέμπτου επιπέδου με 50 κόμβους όπου τα ποσοστά ακρίβειας για τα δεδομένα ελέγχου είναι 0,856 και για τον ταξινομητή svm 0,869. Γενικότερα, παρατηρήθηκε πως όσο αυξάνονται τα ενδιάμεσα επίπεδα σε ένα δίκτυο τόσο τα ποσοστά ακρίβειας μειώνονται.

Στην εικόνα 26 αναπαρίστανται η αρχιτεκτονική του νευρωνικού δικτύου για το σύνολο δεδομένων MNIST.



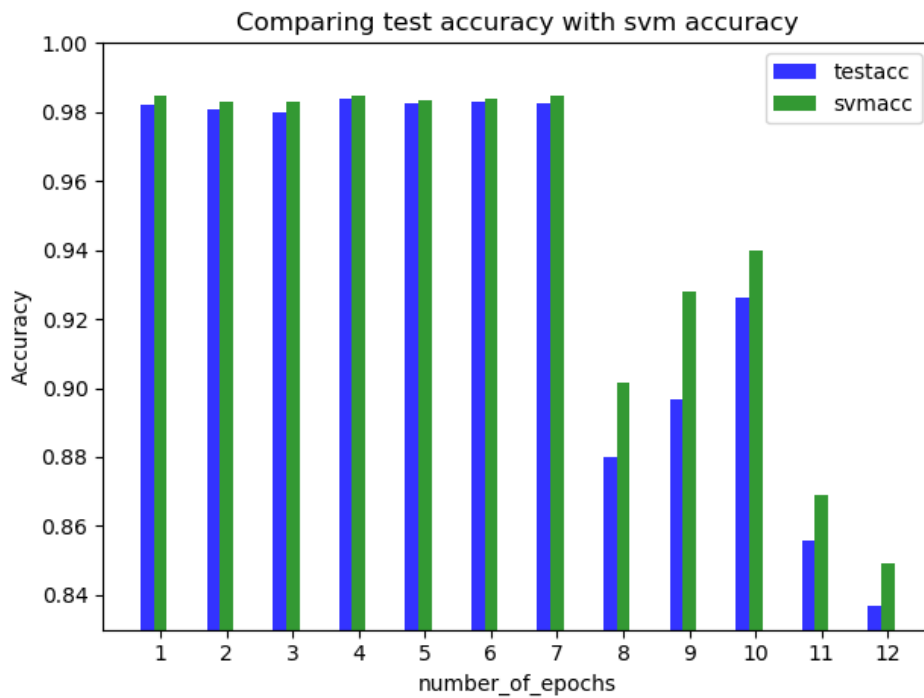
Εικόνα 26 Αρχιτεκτονική νευρωνικού δικτύου για το σύνολο δεδομένων MNIST

Στον πίνακα 4 αναφέρονται τα ποσοστά ακρίβειας του νευρωνικού δικτύου MLP αλλά και του ταξινομητή SVM στο επίπεδο εξόδου για το MNIST.

Πίνακας 4 Ποσοστά ακρίβειας (MLP και ταξινομητή SVM) σε διαφορετικές αρχιτεκτονικές για το σύνολο δεδομένων MNIST

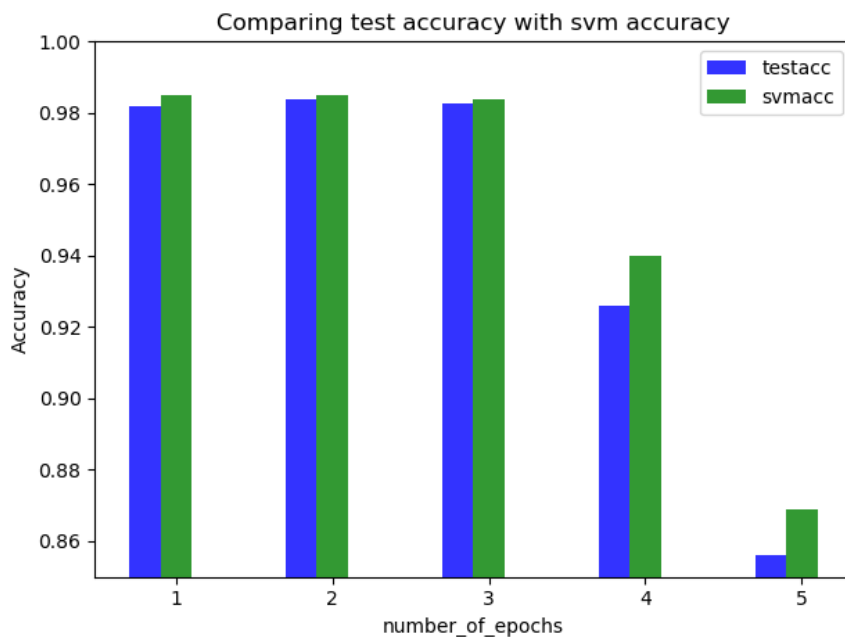
Number of epochs	Hidden layer1	Hidden layer2	Hidden layer3	Hidden layer4	Hidden layer5	Ακρίβεια δεδομένων ελέγχου	Ακρίβεια svm
10	500					0.9819	0.9849
10	250					0.9307	
10	200					0.9387	
10	300					0.9384	
10	400					0.9427	
10	450					0.9458	
10	500					0.9434	
10	550					0.9461	
10	550					0.9432	
10	600					0.9532	
10	500	550				0.9806	0.9829
10	500	500				0.9801	0.9832
10	500	300				0.9838	0.9848
10	500	300	150			0.9825	0.9835
10	500	300	100			0.9828	0.9839
10	500	300	50			0.9825	0.9846
10	500	300	200	100		0.88	0.9015
10	500	300	200	50		0.8965	0.928
10	500	300	200	25		0.926	0.94
10	500	300	200	100	50	0.856	0.869
10	500	300	200	100	25	0.837	0.849

Στην εικόνα 27 γίνεται αναπαράσταση των 12 καλύτερων ποσοστών ακρίβειας για το σύνολο δεδομένων MNIST με τον άξονα yγ' να έχει εύρος 0.84-1 για την καλύτερη απεικόνιση των αποτελεσμάτων διότι οι τιμές της ακρίβειας του svm ταξινομητή και της ακρίβειας των δεδομένων ελέγχου είναι πολύ κοντινές.



Εικόνα 27 Σύγκριση 12 πειραμάτων για το MNIST ως προς τα ποσοστά ακρίβειας και ακρίβειας svm

Στην εικόνα 28 γίνεται αναπαράσταση των 5 καλύτερων ποσοστών ακρίβειας για το σύνολο δεδομένων MNIST με τον άξονα yγ' να έχει εύρος 0.98-1 για την καλύτερη απεικόνιση των αποτελεσμάτων διότι οι τιμές της ακρίβειας του svm ταξινομητή και της ακρίβειας των δεδομένων ελέγχου είναι πολύ κοντινές.



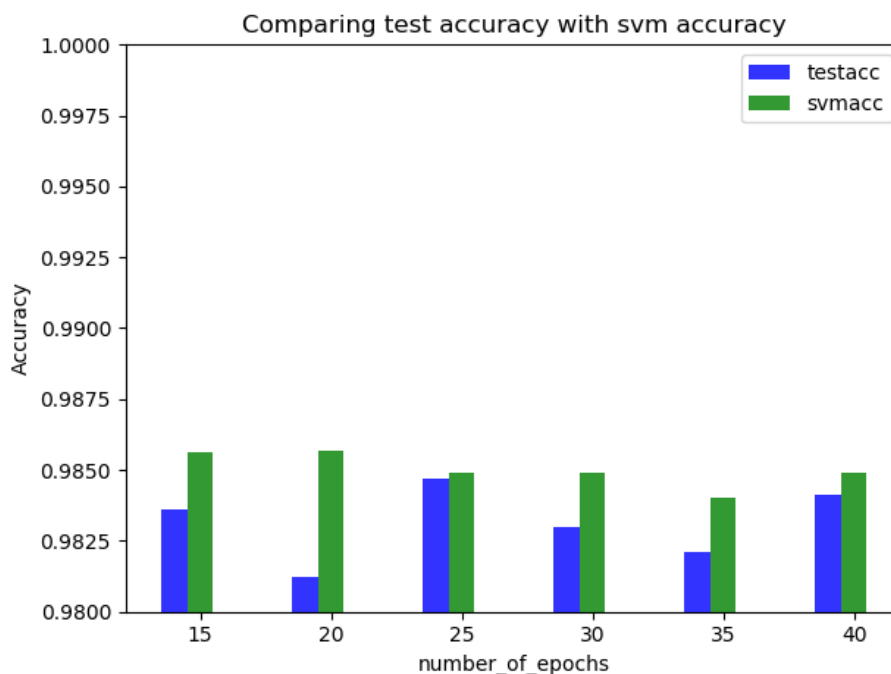
Εικόνα 28 Σύγκριση των 5 καλύτερων αποτελεσμάτων για το MNIST ως προς τα ποσοστά ακρίβειας και ακρίβειας svm

Στον πίνακα 5 αναφέρονται τα ποσοστά ακρίβειας του νευρωνικού δικτύου MLP και του ταξινομητή SVM στο επίπεδο εξόδου αλλά και ο χρόνος εκπαίδευσης του δικτύου για το MNIST.

Πίνακας 5 Ποσοστά ακρίβειας (MLP και ταξινομητή SVM) σε διαφορετικό αριθμό εποχών και μέτρηση του χρόνου υλοποίησης(min/h) για το σύνολο δεδομένων MNIST

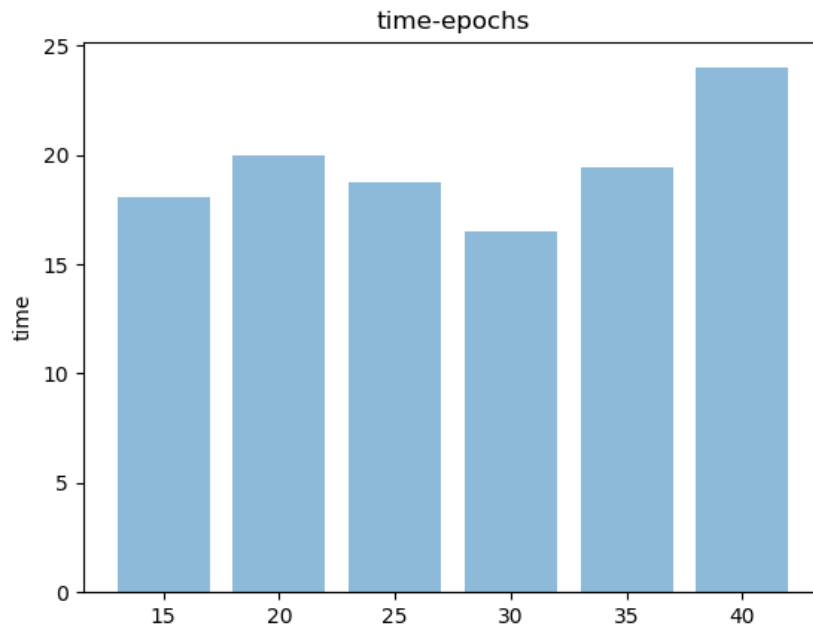
Number of epochs	Hidden layer1	Hidden layer2	Hidden layer3	Hidden layer4	Ακρίβεια δεδομένων ελέγχου	Ακρίβεια svm	TIME(Min/h)
15	500				0.9836	0.9856	18.0522
20	500				0.9812	0.9857	19.9693
25	500	300			0.9847	0.9849	18.7494
30	500	300	150		0.983	0.9849	16.5319
35	500	300	100		0.9821	0.984	19.4142
40	500	300	100	50	0.9841	0.9849	23.9800

Στην παρακάτω εικόνα 29 η αναπαράσταση του άξονα γγ' έχει εύρος 0.98-1 για την καλύτερη απεικόνιση των αποτελεσμάτων διότι οι τιμές της ακρίβειας των δεδομένων ελέγχου και του svm ταξινομητή είναι πολύ κοντινές.



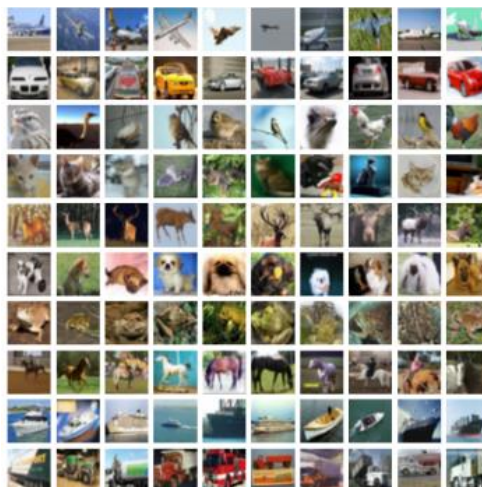
Εικόνα 29 Σύγκριση αποτελεσμάτων ως προς την ακρίβεια δεδομένων ελέγχου και την ακρίβεια svm σε εποχές για το MNIST

Στην εικόνα 30 γίνεται αναπαράσταση του χρόνου υλοποίησης σε σχέση με τον αριθμό των επαναλήψεων για το σύνολο δεδομένων MNIST.



Εικόνα 30 Χρόνος υλοποίησης σε σχέση με τον αριθμό των εποχών για το σύνολο δεδομένων MNIST

Σύνολο Δεδομένων CIFAR-10 [4]. Το συγκεκριμένο σύνολο δεδομένων έχει συλλεχθεί από ένα σύνολο 80.000.0000 μικρών εικόνων από τους Krizhevsky, Nair και Hinton [29]. Αποτελείται από 60.000 έγχρωμες εικόνες 32x32 και 10 κλάσεις που η κάθε κλάση έχει 6.000 εικόνες. Οι κλάσεις αποτελούνται από κατηγορίες όπως αεροπλάνο, ελικόπτερο, πουλί, γάτα, ελάφι, σκύλος, κ.α. , με τις εικόνες εκπαίδευσης να είναι συνολικά 50.000 και οι εικόνες των δοκιμών να είναι 10.000. Διαχωρίζεται από πέντε ομάδες(batch) εκπαίδευσης και μια ομάδα δοκιμών. Η ομάδα δοκιμών περιέχει 1000 τυχαίες επιλεγμένες εικόνες από την κάθε κατηγορία. Οι ομάδες εκπαίδευσης περιέχουν τις υπόλοιπες εικόνες με τυχαία σειρά αλλά μερικές ομάδες εκπαίδευσης περιέχουν περισσότερες εικόνες από μια κατηγορία σε σχέση με μια άλλη. Ακόμα οι ομάδες εκπαίδευσης περιέχουν 5.000 εικόνες από την κάθε κατηγορία.



Εικόνα 31 Σύνολο δεδομένων CIFAR-10 Πηγή: T. Raiko, H. Valpola, and Y. LeCun, " Deep learning made easier by linear transformations in perceptrons". Journal of Machine Learning Research - Proceedings Track, (22):924-932, (2012).

Ο κώδικας που χρησιμοποιήθηκε για την εκπαίδευση του μοντέλου είναι περίπου ίδιος με αυτόν που χρησιμοποιήθηκε στο σύνολο δεδομένων MNIST με κάποιες μικρές παραλλαγές. Μια πολύ σημαντική διαφορά είναι ότι το σύνολο δεδομένων CIFAR-10 έχει τριπλάσια πληροφορία στην είσοδο του, λόγω των έγχρωμων εικόνων οι οποίες αποτελούνται από τρεις διαστάσεις: ύψος, πλάτος, βάθος. Ακόμα, τα δεδομένα αυτού του CIFAR-10 είναι πιο ποικιλόμορφα από τα δεδομένα MNIST, καθώς περιέχουν αντικείμενα διαφόρων μεγεθών και χρωμάτων. Αυτό έχει σαν αποτέλεσμα το ποσοστό ακρίβειας να είναι μικρότερο για τα πειράματα που θα διεξαχθούν σε σχέση με τα πειράματα του MNIST.

Στα πειράματα που πραγματοποιήθηκαν στο νευρωνικό δίκτυο, το επίπεδο εισόδου δέχεται 3072bytes από την εικόνα, εκ των οποίων τα πρώτα 1024 αφορούν το κόκκινο φάσμα της εικόνας, τα επόμενα 1024 αφορούν το πράσινο και τα τελευταία 1024 αφορούν για το μπλε φάσμα.

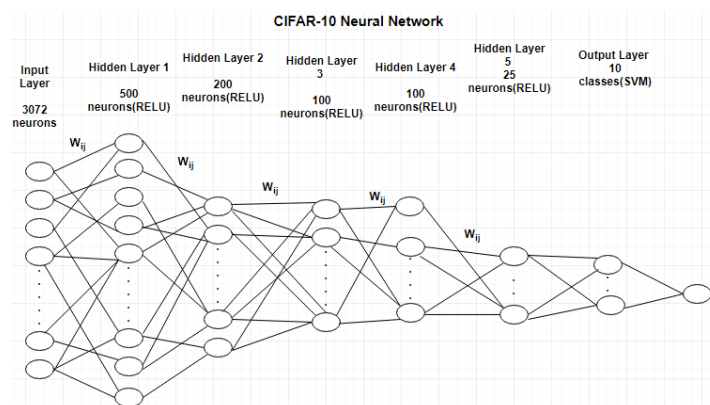
Το καλύτερο ποσοστό με την χρήση ενός ενδιάμεσου επιπέδου με αριθμό κόμβων 500 και με σύνολο εποχών 10 επιτυγχάνονται με ποσοστά ακρίβειας για τα δεδομένα ελέγχου 0,45 και με την χρήση ταξινομητή svm 0,488 .

Το δεύτερο καλύτερο ποσοστό με την προσθήκη δεύτερου ενδιάμεσου επιπέδου και με συνολικό αριθμό κόμβων 300 , επιτυγχάνονται ποσοστά ακρίβειας για τα δεδομένα ελέγχου 0,4501 και για τον ταξινομητή svm 0,467. Παρατηρείται πως με την προσθήκη του συγκεκριμένου επιπέδου τα ποσοστά ακρίβειας μειώνονται σε σχέση με τα ποσοστά του προηγούμενου επιπέδου.

Στη συνέχεια, προστίθεται τρίτο επίπεδο και με αριθμό κόμβων 150 και τα ποσοστά ακρίβειας για τα δεδομένα ελέγχου είναι 0,454 και για τον ταξινομητή svm είναι 0,49 . Με την προσθήκη αυτού του επιπέδου παρατηρείται μια μικρή αύξηση των ποσοστών ακρίβειας σε σχέση με τα δύο προηγούμενα πειράματα.

Προστίθεται ακόμα ένα τέταρτο επίπεδο με αριθμό κόμβων 50 που τα ποσοστά ακρίβειας μειώνονται σε σχέση με τον συνδυασμό των προηγούμενων 3 επιπέδων. Έτσι το ποσοστό ακρίβειας για το νευρωνικό δίκτυο είναι 0,4328 και με την χρήση ταξινομητή svm 0,4629.

Τέλος, με την προσθήκη πέμπτου επιπέδου με αριθμό κόμβων 25, τα ποσοστά ακρίβειας που επιτυγχάνονται είναι ακρίβεια δεδομένων ελέγχου 0,4165 και με την χρήση ταξινομητή svm 0,4553. Στην εικόνα 32 αναπαρίσταται η αρχιτεκτονική του νευρωνικού δικτύου για το σύνολο δεδομένων CIFAR-10.



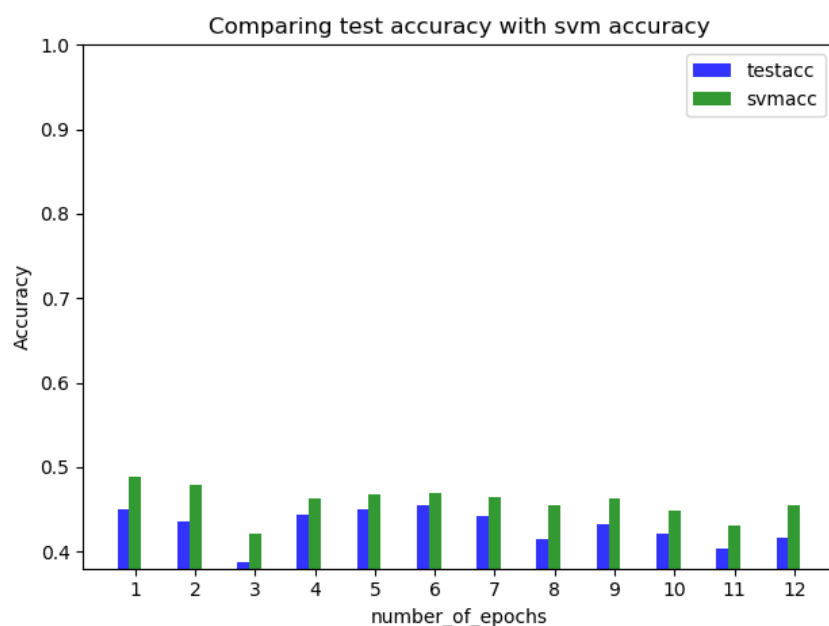
Εικόνα 32 Αρχιτεκτονική νευρωνικού δικτύου για το σύνολο δεδομένων CIFAR-10

Στον πίνακα 6 αναφέρονται τα ποσοστά ακρίβειας του νευρωνικού δικτύου MLP αλλά και του ταξινομητή SVM στο επίπεδο εξόδου για το CIFAR-10.

Πίνακας 6 Ποσοστά ακρίβειας (MLP και SVM ταξινομητή) σ για το σύνολο δεδομένων CIFAR-10

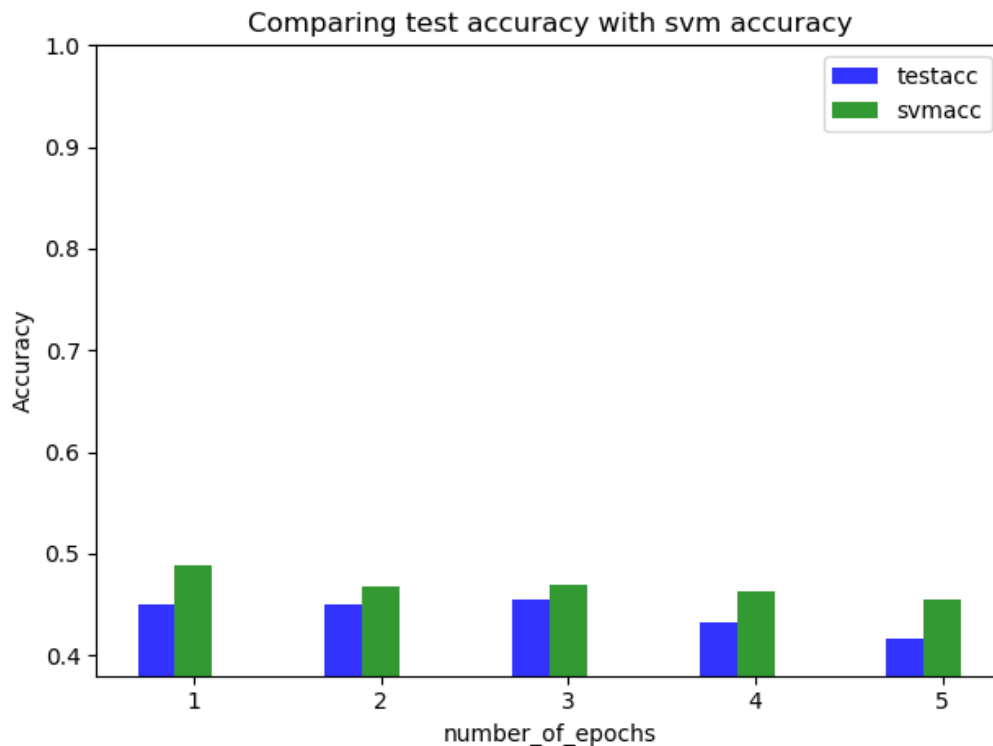
Number of epochs	Hidden layer1	Hidden layer2	Hidden layer3	Hidden layer4	Hidden layer5	Ακρίβεια δεδομένων ελέγχου	Ακρίβεια svm
2	500					0.381	0.4461
10	500					0.45	0.4888
10	300					0.4362	0.4784
2	500	550				0.3881	0.4204
10	500	550				0.4432	0.4634
10	500	300				0.4501	0.467
10	500	200	150			0.454	0.49
10	500	200	100			0.4418	0.4652
10	500	200	50			0.4154	0.4541
10	500	200	100	50		0.4328	0.4629
10	500	200	100	25		0.4216	0.4485
10	500	200	100	100	50	0.4030	0.430
10	500	200	100	100	25	0.4165	0.4553

Στην εικόνα 33 γίνεται αναπαράσταση των 12 καλύτερων ποσοστών ακρίβειας για το σύνολο δεδομένων CIFAR-10 με τον άξονα γγ' να έχει εύρος 0.34-1 για την καλύτερη απεικόνιση των αποτελεσμάτων διότι οι τιμές της ακρίβειας του svm ταξινομητή και της ακρίβειας των δεδομένων ελέγχου είναι πολύ κοντινές.



Εικόνα 33 Σύγκριση 12 αποτελεσμάτων για το CIFAR-10 ως προς τα ποσοστά ακρίβειας και ακρίβειας svm

Στην εικόνα 34 γίνεται αναπαράσταση των 5 καλύτερων ποσοστών ακρίβειας για το σύνολο δεδομένων CIFAR-10 με τον άξονα yγ' να έχει εύρος 0.98-1 για την καλύτερη απεικόνιση των αποτελεσμάτων διότι οι τιμές της ακρίβειας του svm ταξινομητή και της ακρίβειας δεδομένων ελέγχου είναι πολύ κοντινές.



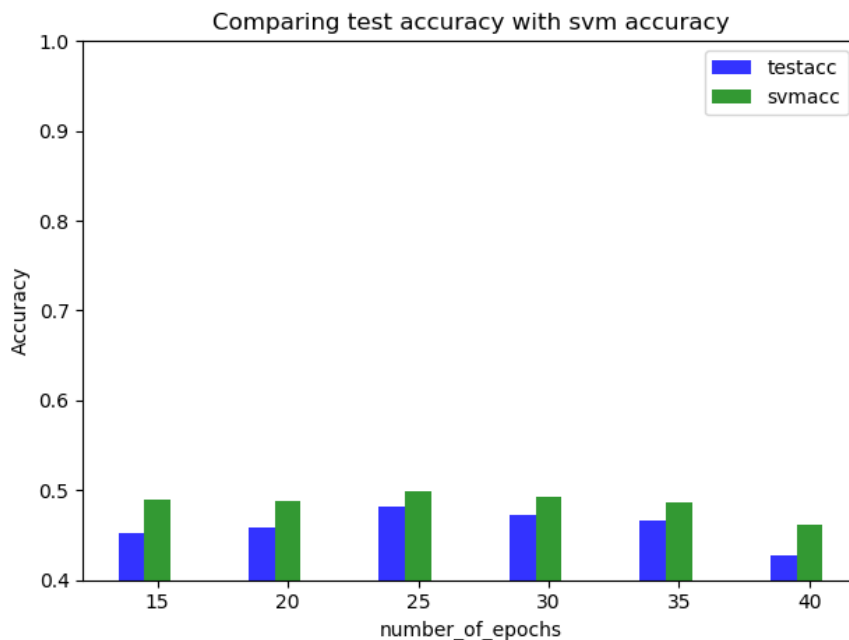
Εικόνα 34 Σύγκριση των 5 καλύτερων αποτελεσμάτων για το CIFAR-10 ως προς τα ποσοστά ακρίβειας και ακρίβειας svm

Στον πίνακα 7 αναφέρονται τα ποσοστά ακρίβειας του νευρωνικού δικτύου MLP και του ταξινομητή SVM στο επίπεδο εξόδου αλλά και ο χρόνος εκπαίδευσης του δικτύου για το CIFAR-10.

Πίνακας 7 Ποσοστά ακρίβειας (MLP και ταξινομητή SVM) σε διαφορετικό αριθμό εποχών και μέτρηση του χρόνου υλοποίησης(min/h) για το σύνολο δεδομένων CIFAR-10

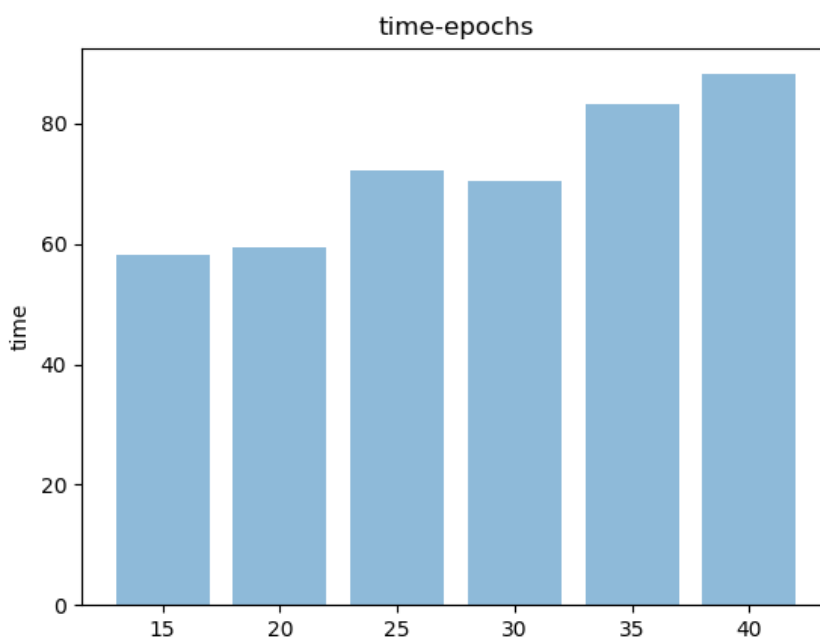
Number of epochs	Hidden layer1	Hidden layer2	Hidden layer3	Hidden layer4	Hidden layer5	Ακρίβεια δεδομένων ελέγχου	Ακρίβεια svm	TIME(Min/h)
15	500					0.4518	0.4898	58.0612
25	500	300				0.4583	0.4873	59.5073
30	500	300	150			0.4818	0.499	72.2942
35	500	300	100			0.4725	0.4923	70.5253
40	500	300	75	50		0.4657	0.4863	83.1649
45	500	250	125	75	25	0.4279	0.4617	88.1708

Στην παρακάτω εικόνα 35 η αναπαράσταση του άξονα γγ' έχει εύρος 0.4-1 για την καλύτερη απεικόνιση των αποτελεσμάτων λόγω ότι οι τιμές της ακρίβειας των δεδομένων ελέγχου και του του svm ταξινομητή είναι πολύ κοντινές.



Εικόνα 35 Σύγκριση αποτελεσμάτων ως προς την ακρίβεια δεδομένων ελέγχου και την ακρίβεια svm σε εποχές για το CIFAR-10

Στην εικόνα 36 γίνεται αναπαράσταση του χρόνου υλοποίησης σε σχέση με τον αριθμό των επαναλήψεων για το σύνολο δεδομένων CIFAR-10.



Εικόνα 36 Χρόνος υλοποίησης σε σχέση με τον αριθμό των εποχών για το σύνολο δεδομένων CIFAR-10

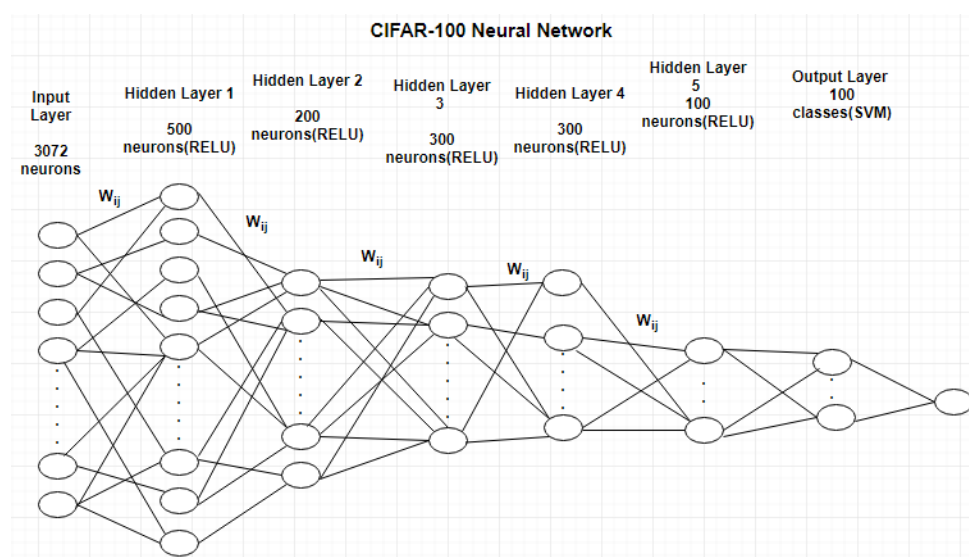
Σύνολο Δεδομένων CIFAR-100 [3]. Το συγκεκριμένο σύνολο δεδομένων είναι παρόμοιο με το προηγούμενο CIFAR-10 αλλά διαφέρει στο ότι έχει 100 κλάσεις και η κάθε κλάση περιέχει 600 εικόνες. Υπάρχουν 500 εικόνες εκπαίδευσης και 100 εικόνες δοκιμών και οι 100 κλάσεις ομαδοποιούνται σε 20 υπερ-κλάσεις. Κάθε εικόνα αποτελείται από ένα σημάδι που δείχνει την κλάση από την οποία προέρχεται και ένα ακόμη σημάδι της υπερ-κλάσης στην οποία ανήκει. Οι υπερ-κλάσεις είναι: ψάρια, λουλούδια, φρούτα και λαχανικά, ερπετά, μικρά θηλαστικά, άνθρωποι, δέντρα, κ.α.

Τα πειράματα πραγματοποιήθηκαν με την ίδια λογική όπως στο προηγούμενο σύνολο δεδομένων CIFAR-10, με την μόνη διαφορά ότι στην έξοδο του νευρωνικού δικτύου υπάρχουν 100 κλάσεις και όχι 10. Αυτή η αύξηση του αριθμού των κλάσεων έχει σαν αποτέλεσμα να υπάρχουν μικρότερα ποσοστά ακρίβειας και το δίκτυο να χρειάζεται περισσότερο χρόνο να εκπαιδευτεί.

Το νευρωνικό δίκτυο στο επίπεδο εισόδου δέχεται 3072 pixels και με την χρήση ενός ενδιάμεσου επιπέδου 500 κόμβων τα ποσοστά ακρίβειας που επιτυγχάνονται για τα δεδομένα ελέγχου είναι 0,1642 και με την χρήση svm ταξινομητή είναι 0,2011 . Παρατηρείται πως τα ποσοστά ακρίβειας έχουν μειωθεί σε σχέση με το σύνολο δεδομένων CIFAR-10.

Το καλύτερο ποσοστό ακρίβειας στην προσθήκη δεύτερου ενδιάμεσου επιπέδου 300 κόμβων για τα δεδομένα ελέγχου είναι 0,1861 και με την χρήση svm ταξινομητή είναι 0,2182. Στη συνέχεια, με την προσθήκη τρίτου ενδιάμεσου επιπέδου 300 κόμβων, τα ποσοστά ακρίβειας για τα δεδομένα ελέγχου είναι 0,1857 και με την χρήση svm ταξινομητή 0,2201. Ακόμα, στην προσθήκη τέταρτου επιπέδου τα καλύτερα ποσοστά ακρίβειας που επιτυγχάνονται για τα δεδομένα ελέγχου είναι 0,1704 και για τον svm ταξινομητή είναι 0,2133. Τέλος, στην προσθήκη πέμπτου επιπέδου 100 κόμβων τα ποσοστά ακρίβειας για τα δεδομένα ελέγχου είναι 0,1569 και με την χρήση svm ταξινομητή είναι 0,184 .

Στην εικόνα 37 αναπαρίσταται η αρχιτεκτονική του νευρωνικού δικτύου για το σύνολο δεδομένων CIFAR-100.



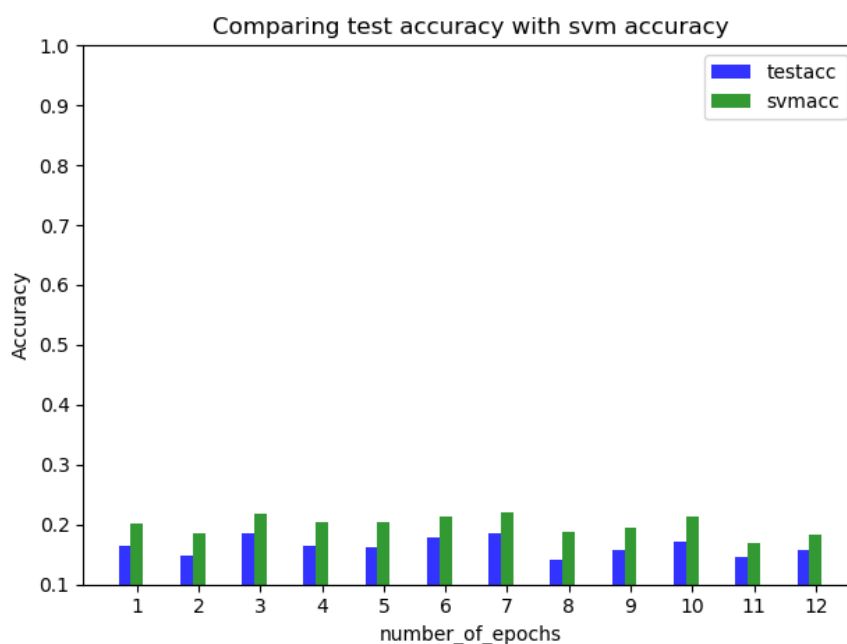
Εικόνα 37 Αρχιτεκτονική MLP για το σύνολο δεδομένων CIFAR-100

Στον πίνακα 8 αναφέρονται τα ποσοστά ακρίβειας του νευρωνικού δικτύου MLP αλλά και του ταξινομητή SVM στο επίπεδο εξόδου για το CIFAR-100.

Πίνακας 8 Ποσοστά ακρίβειας (MLP και SVM ταξινομητή) για το σύνολο δεδομένων CIFAR-100

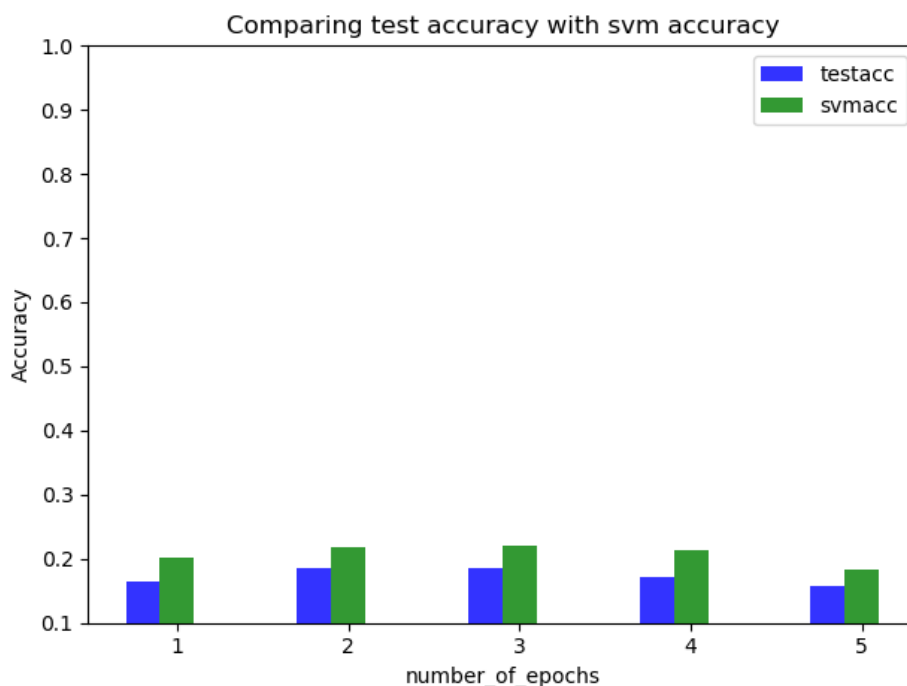
Number of epochs	Hidden layer1	Hidden layer2	Hidden layer3	Hidden layer4	Hidden layer5	Ακρίβεια δεδομένων ελέγχου	Ακρίβεια svm
2	500					0.1204	0.1736
10	500					0.1642	0.2011
10	300					0.1487	0.1858
10	500	300				0.1861	0.2182
10	500	200	150			0.165	0.2047
10	500	200	100			0.1624	0.2048
10	500	200	200			0.1774	0.2126
10	500	200	300			0.1857	0.2201
10	500	200	300	100		0.1417	0.1879
10	500	200	300	200		0.1568	0.1955
10	500	200	300	300		0.1704	0.2133
10	500	200	300	300	200	0.145	0.168
10	500	200	300	300	100	0.1569	0.184

Στην εικόνα 38 γίνεται αναπαράσταση των 12 καλύτερων ποσοστών ακρίβειας για το σύνολο δεδομένων CIFAR-100 με τον άξονα yγ' να έχει εύρος 0.1-1 για την καλύτερη απεικόνιση των αποτελεσμάτων διότι οι τιμές της ακρίβειας του svm ταξινομητή και της ακρίβειας των δεδομένων ελέγχου είναι πολύ κοντινές.



Εικόνα 38 Σύγκριση 12 αποτελεσμάτων για το CIFAR-100 ως προς τα ποσοστά ακρίβειας και ακρίβειας svm

Στην εικόνα 39 γίνεται αναπαράσταση των 5 καλύτερων ποσοστών ακρίβειας για το σύνολο δεδομένων CIFAR-100 με τον άξονα yγ' να έχει εύρος 0.1-1 για την καλύτερη απεικόνιση των αποτελεσμάτων διότι οι τιμές της ακρίβειας του svm ταξινομητή και της ακρίβειας των δεδομένων ελέγχου είναι πολύ κοντινές.



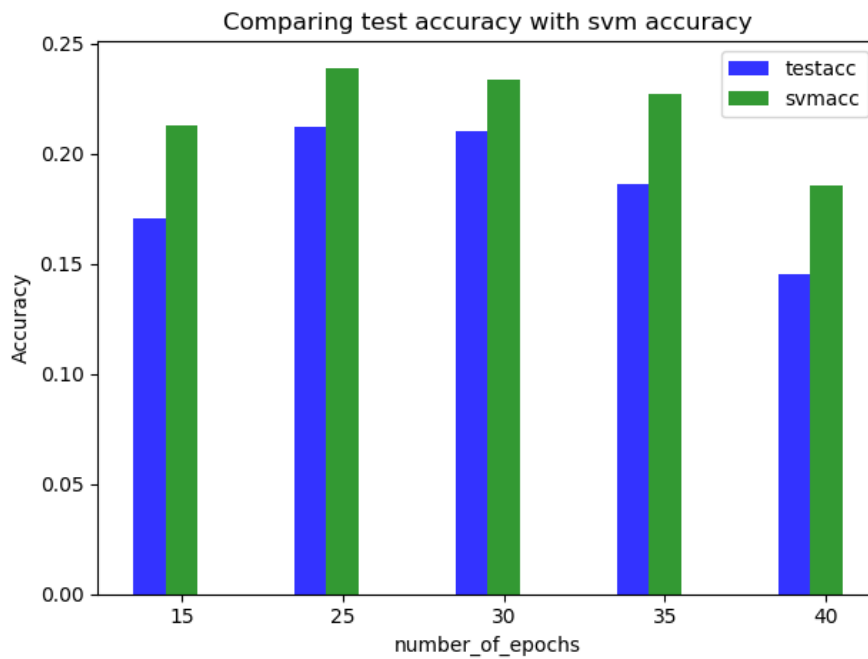
Εικόνα 39 Σύγκριση των 5 καλύτερων αποτελεσμάτων για το CIFAR-100 ως προς τα ποσοστά ακρίβειας και ακρίβειας svm

Στον πίνακα 9 αναφέρονται τα ποσοστά ακρίβειας του νευρωνικού δικτύου MLP και του ταξινομητή SVM στο επίπεδο εξόδου αλλά και ο χρόνος εκπαίδευσης του δικτύου για το CIFAR-100.

Πίνακας 9 Ποσοστά ακρίβειας (MLP και ταξινομητή SVM) σε διαφορετικό αριθμό εποχών και μέτρηση του χρόνου υλοποίησης(min/h) για το σύνολο δεδομένων CIFAR-100

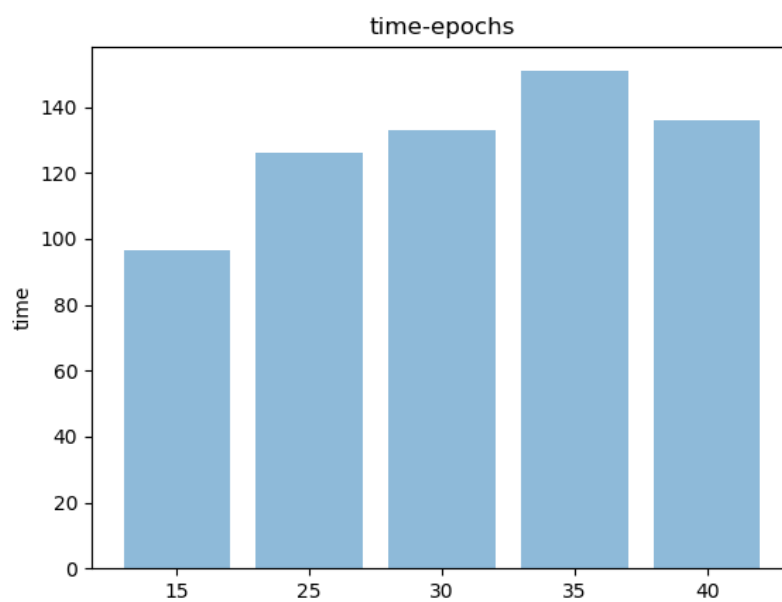
Number of epochs	Hidden layer1	Hidden layer2	Hidden layer3	Hidden layer4	Hidden layer5	Ακρίβεια δεδομένων ελέγχου	Ακρίβεια svm	TIME(Min/h)
15	500					0.1708	0.2131	96.5115
25	500	300				0.2123	0.239	126.2417
30	500	300	150			0.2104	0.2334	132.7578
35	500	300	250	200		0.1862	0.2271	150.8455
40	500	300	250	200	150	0.1451	0.1858	135.7774

Στην παρακάτω εικόνα 40 η αναπαράσταση της ακρίβειας για τα δεδομένα ελέγχου του svm ταξινομητή.



Εικόνα 40 Σύγκριση αποτελεσμάτων ως προς την ακρίβεια δεδομένων ελέγχου και την ακρίβεια svm σε εποχές για το CIFAR-100

Στην εικόνα 41 γίνεται αναπαράσταση του χρόνου υλοποίησης σε σχέση με τον αριθμό των επαναλήψεων για το σύνολο δεδομένων CIFAR-10.



Εικόνα 41 Χρόνος υλοποίησης σε σχέση με τον αριθμό των εποχών για το σύνολο δεδομένων CIFAR-100

4.2.2 Συνελκτικικά Δίκτυα με Μηχανές Διανυσμάτων Υποστήριξης

Ο συνδυασμός των συνελκτικικών δικτύων και των SVM επιλύουν προβλήματα κατηγοριοποίησης μεγάλων συνόλων δεδομένων. Παρακάτω περιγράφεται αναλυτικά η εκπαίδευση του δικτύου με συνδυασμό των δύο αρχιτεκτονικών:

- I. Γίνεται εισαγωγή των εικόνων στο επίπεδο εισόδου του νευρωνικού δικτύου.
- II. Εκπαιδεύονται οι δυαδικές κλάσεις των SVM με τις εικόνες που έχουν εισαχθεί αλλά και με τις τιμές του προτελευταίου επιπέδου του δικτύου.
- III. Μέσω του δικτύου υπολογίζονται οι τιμές για το προτελευταίο επίπεδο από τα εικονοστοιχεία των εικόνων που έχουν εισαχθεί.
- IV. Στη συνέχεια οι υπολογισμένες τιμές εισάγονται στους ταξινομητές SVM για να προκύψουν οι έξοδοι τους.
- V. Οι τιμές εξόδων των SVM ορίζονται σαν τιμές ενεργοποίησης του τελευταίου επιπέδου.
- VI. Εκτελείται ο αλγόριθμος Backpropagation για να προωθηθούν τα σφάλματα κατηγοριοποίησης στα προηγούμενα επίπεδα και να γίνουν μεταβολές στα βάρη ελαχιστοποιώντας το σφάλμα.

Σύνολο Δεδομένων MNIST. Γίνεται υλοποίηση ενός συνελκτικού δικτύου για την επίτευξη καλύτερου ποσοστού ακρίβειας σε σχέση με το MLP. Τα συνελκτικά δίκτυα πετυχαίνουν καλύτερα ποσοστά ακρίβειας επειδή εντοπίζουν τα μοτίβα σε εικόνες και εξάγουν τα χαρακτηριστικά της εικόνας.

Το ΣΝΔ που υλοποιήθηκε αποτελείται από το επίπεδο εισόδου που είναι το επίπεδο συνέλιξης, με συνολικό αριθμό 32 χαρακτηριστικών χάρτη μεγέθους 5 x 5 και το επίπεδο εισόδου δέχεται 784 κόμβους. Το επόμενο επίπεδο είναι το επίπεδο υποδειγματοληψίας, το οποίο μειώνει την εικόνα που έχει προκύψει από την συνέλιξη του προηγούμενου επιπέδου με τη χρήση της συνάρτησης υποδειγματοληψίας μεγέθους 2 x 2. Το επόμενο επίπεδο είναι το επίπεδο της κανονικοποίησης (regularization layer) που χρησιμοποιείται η τεχνική DropOut με ποσοστό 20% για την αποφυγή υπερεκπαίδευσης. Στη συνέχεια, το επόμενο επίπεδο μετατρέπει τον δυοδιάστατο πίνακα σε ένα διάνυσμα έτσι ώστε να επιτρέπει την επεξεργασία της εξόδου από πλήρως συνδεδεμένα επίπεδα. Το επόμενο επίπεδο είναι ένα πλήρως συνδεδεμένο επίπεδο που αποτελείται από συγκεκριμένο αριθμό κόμβων. Τέλος, το τελευταίο επίπεδο αποτελείται από 10 κλάσεις και ένα ταξινομητή SVM για τον διαχωρισμό της κάθε κλάσης. Ο αλγόριθμος βελτιστοποίησης που χρησιμοποιείται είναι ο Adam για τον καλύτερο διαχωρισμό των βαρών. Το συνελκτικό δίκτυο εκπαιδεύεται για 10 εποχές όπως και το MLP.

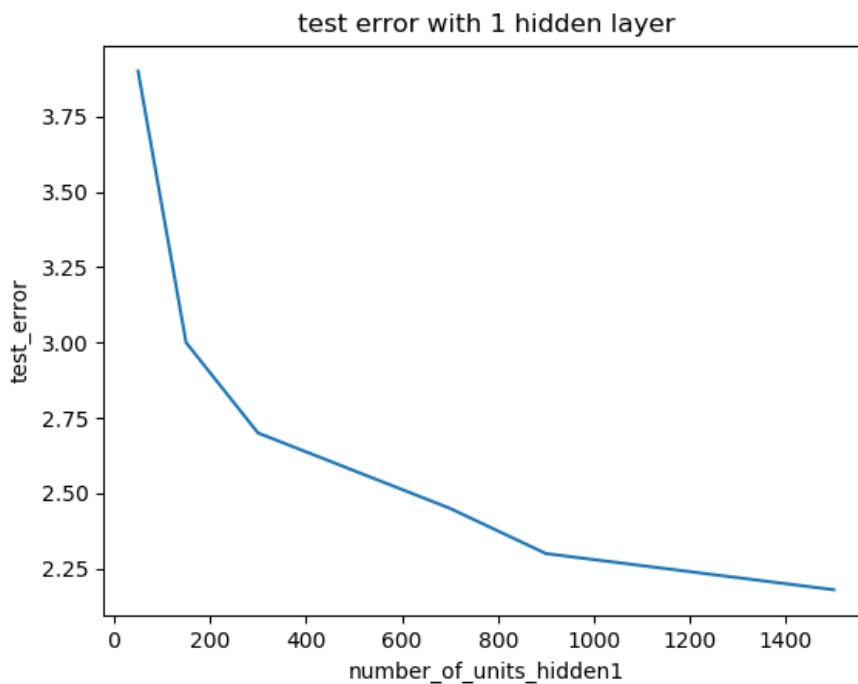
Παρακάτω προκύπτουν τα αποτελέσματα ποσοστών σφαλμάτων (% test error) με την ύπαρξη ενός ενδιάμεσου επιπέδου με το καλύτερο ποσοστό σφάλματος 2.18% να προκύπτει από το συνολικό αριθμό κόμβων στο ενδιάμεσο επίπεδο 2000.

Στον πίνακα 10 αναφέρονται τα ποσοτά σφάλματος που προκύπτουν με ένα φίλτρο 5x5 στο επίπεδο συνέλιξης και ένα ενδιάμεσο επίπεδο.

Πίνακας 10 Ποσοστά σφάλματος συνελκτικού δικτύου με ένα φίλτρο (πυρήνα) 5x5 στο επίπεδο συνέλιξης, ένα ενδιάμεσο επίπεδο και διαφορετικό αριθμό κόμβων για το σύνολο δεδομένων MNIST

Input layer	Pooling layer	Hidden layer1	Output layer	%test error
32	5 x 5	100	10	3.9
32	5 x 5	200	10	3
32	5 x 5	500	10	2.7
32	5 x 5	1000	10	2.45
32	5 x 5	1500	10	2.3
32	5 x 5	2000	10	2.18

Στην εικόνα 42 γίνεται η αναπαράσταση του ποσοστού σφάλματος με την χρήση ενός ενδιάμεσου επιπέδου.



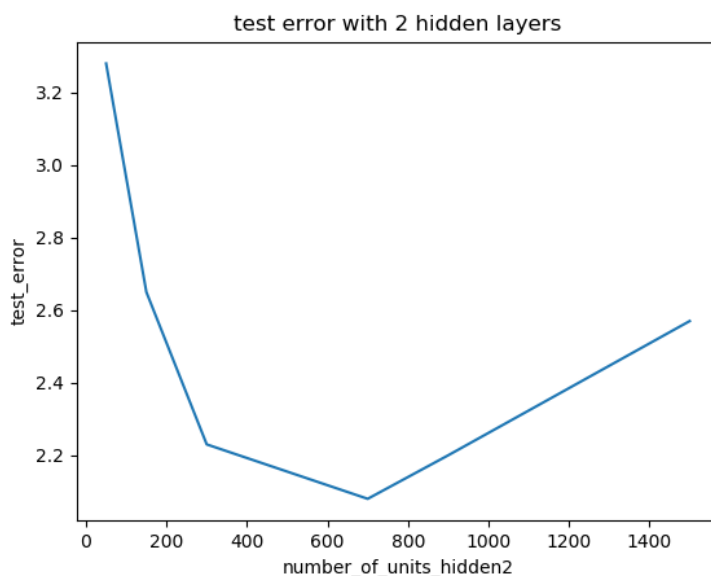
Εικόνα 42 Αναπαράσταση ποσοστού σφάλματος με ένα ενδιάμεσο επίπεδο

Στον παρακάτω πίνακα 11 προκύπτουν τα αποτελέσματα των ποσοστών σφάλματος με τη χρήση δύο ενδιάμεσων επιπέδων και με καλύτερο ποσοστό σφάλματος αν είναι 2.08% .

Πίνακας 11 Αναπαράσταση ποσοστών σφάλματος (συνελικτικού δικτύου) με ένα φίλτρο (πυρήνα) 5x5 στο επίπεδο συνέλιξης, δύο ενδιάμεσα επίπεδα και διαφορετικό αριθμό κόμβων για το σύνολο δεδομένων MNIST

Input layer	Pooling layer	Hidden layer1	Hidden layer2	Output layer	%test error
32	5 x 5	100	50	10	3.28
32	5 x 5	200	150	10	2.65
32	5 x 5	500	300	10	2.23
32	5 x 5	1000	700	10	2.08
32	5 x 5	1500	900	10	2.20
32	5 x 5	2000	1500	10	2.57

Στην εικόνα 43 γίνεται η αναπαράσταση του ποσοστού σφάλματος με την χρήση δύο ενδιάμεσων επιπέδων.



Εικόνα 43 Αναπαράσταση ποσοστού σφάλματος με δύο ενδιάμεσα επίπεδα

Στον παρακάτω πίνακα 12 προκύπτουν τα αποτελέσματα των ποσοστών σφάλματος με τη χρήση πέντε ενδιάμεσων επιπέδων και το καλύτερο ποσοστό σφάλματος να είναι 2.46% .

Πίνακας 12 Ποσοστά σφάλματος (συνελικτικού δικτύου) με ένα φίλτρο (πυρήνα) 5x5 στο επίπεδο συνέλιξης, πέντε ενδιάμεσα επίπεδα και διαφορετικό αριθμό κόμβων για το σύνολο δεδομένων MNIST

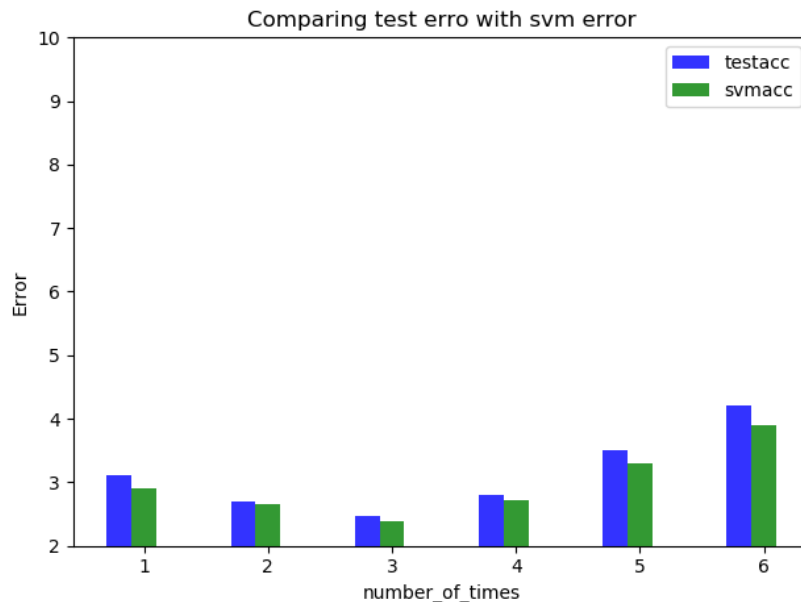
Input layer	Pooling layer	Hidden layer1	Hidden layer2	Hidden layer3	Hidden layer4	Hidden layer5	Output layer	%test error
32	5 x 5	100	80	65	50	30	10	3.1
32	5 x 5	200	190	150	130	80	10	2.7
32	5 x 5	500	420	350	260	130	10	2.46
32	5 x 5	1000	730	520	410	250	10	2.8
32	5 x 5	1500	1100	840	520	150	10	3.5
32	5 x 5	2000	1500	1000	650	360	10	4.2

Στον παρακάτω πίνακα 13 προκύπτουν τα ποσοστά σφάλματος με τη χρήση ταξινομητή SVM και το καλύτερο ποσοστό σφάλματος να είναι 2.39%.

Πίνακας 13 Αναπαράσταση ποσοστών σφάλματος (συνελικτικού δικτύου και SVM ταξινομητή) με ένα φίλτρο (πιρήνα) 5x5 στο επίπεδο συνέλιξης, πέντε ενδιάμεσα επίπεδα και διαφορετικό αριθμό κόμβων για το σύνολο δεδομένων MNIST

Input layer	Pooling layer	Hidden layer1	Hidden layer2	Hidden layer3	Hidden layer4	Hidden layer5	Output layer	SVM	%test error
32	5 x 5	100	80	65	50	30	10	2.9	3.1
32	5 x 5	200	190	150	130	80	10	2.65	2.7
32	5 x 5	500	420	350	260	130	10	2.39	2.46
32	5 x 5	1000	730	520	410	250	10	2.72	2.8
32	5 x 5	1500	1100	840	520	150	10	3.3	3.5
32	5 x 5	2000	1500	1000	650	360	10	3.9	4.2

Στην εικόνα 44 αναπαρίστανται το ποσοστό σφάλματος του νευρωνικού δικτύου και του SVM ταξινομητή μετά την υλοποίηση 6 πειραμάτων.



Εικόνα 44 Σύγκριση 6 αποτελεσμάτων για το MNIST ως προς το ποσοστό σφάλματος του νευρωνικού δικτύου – ποσοστό σφάλματος svm

Κεφάλαιο 5 Συμπεράσματα

Η μελέτη της βιβλιογραφίας αλλά και τα πειράματα που υλοποιήθηκαν, οδηγήθηκαν στο συμπέρασμα ότι τα νευρωνικά δίκτυα, είτε τα MLP, είτε τα συνελκτικά είναι μια πολύ αποδοτική αρχιτεκτονική ειδικά για προβλήματα κατηγοριοποίησης εικόνων σε πολλές κλάσεις. Αυτές οι αρχιτεκτονικές είναι πολλά υποσχόμενες καθώς αφήνουν πολλά περιθώρια βελτίωσης ειδικά στην λύση πιο σύνθετων προβλημάτων.

Βελτίωση στην απόδοση των αποτελεσμάτων εξόδου ήταν ο συνδυασμός του νευρωνικού δικτύου με τις μηχανές διανυσμάτων υποστήριξης. Ο συνδυασμός αυτών των αρχιτεκτονικών υλοποιήθηκε σε τρία σετ δεδομένων με τα αποτελέσματα εξόδου να δείχνουν βελτίωση ως προς την ακρίβεια ειδικά με την προσθήκη του SVM ταξινομητή στο τελευταίο επίπεδο του νευρωνικού δικτύου. Μια τέτοια αρχιτεκτονική θα μπορούσε να χρησιμοποιηθεί για την αύξηση της απόδοσης σε ένα ήδη υπάρχον σύστημα ή ακόμα και σε ένα νέο. Ακόμα, αξίζει να παρατηρηθεί πως αυτός ο συνδυασμός είναι πολύ αποδοτικός ως προς την εκπαίδευση ενός δικτύου, διότι, όπως έχει αναφερθεί και πιο πάνω, το δίκτυο θα κάνει την εξαγωγή των χαρακτηριστικών από τα δεδομένα εισόδου και θα τα τροφοδοτήσει στα SVM με λιγότερες διαστάσεις για τον διαχωρισμό.

5.1 Μετέπειτα πειράματα και έρευνα

Σύμφωνα με τα αποτελέσματα που προέκυψαν από τα πειράματα, η συγκεκριμένη αρχιτεκτονική αποδείχθηκε πως είναι ιδιαίτερα αποδοτική με περιθώρια βελτίωσης. Το μοντέλο που υλοποιήθηκε θα μπορούσε να χρησιμοποιηθεί σε πολύ περισσότερα σύνολα δεδομένων πιο μεγάλα και πιο περίπλοκα. Ένα ακόμη στοιχείο που θα μπορούσε να δοκιμαστεί είναι η ύπαρξη περισσότερων ενδιάμεσων επιπέδων στο νευρωνικό δίκτυο, τα οποία θα απαιτούσαν μεγαλύτερη υπολογιστική ισχύ αλλά και χρόνο για την εκπαίδευση των δεδομένων. Έτσι, θα μπορούσε να δοκιμαστεί η σωστή λειτουργία του μοντέλου σε διάφορες συνθήκες.

Τέλος, θα μπορούσαν να δοκιμαστούν πολλές αρχιτεκτονικές νευρωνικών δικτύων με κύριο σκοπό να βρεθούν αυτές που μειώνουν τις διαστάσεις για τη τροφοδότηση των SVM. Ακόμα θα μπορούσαν να δοκιμαστούν σύνολα δεδομένων διαφορετικής μορφής από εικόνες όπως ήχος ή βίντεο ώστε να προκύψουν αποτελέσματα απόδοσης και να κριθεί αν αυτή η αρχιτεκτονική θα μπορούσε να φανεί ιδιαίτερα χρήσιμη και σε άλλα προβλήματα πιο γενικά εκτός από την κατηγοριοποίηση εικόνας.

Βιβλιογραφία

- 1) Goodfellow, Y. Bengio, A. Courville, "Deep Learning. MIT Press, Adaptive Computation and Machine Learning Series ", 2017.
- 2) P. Flach, "Machine Learning: The Art and Science of Algorithms that Make Sense of Data", Cambridge University Press,2012.
- 3) Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition." *Proceedings of the IEEE*, 86(11):2278-2324, 1998.
- 4) T. Raiko, H. Valpola, and Y. LeCun, " Deep learning made easier by linear transformations in perceptrons". *Journal of Machine Learning Research - Proceedings Track*, (22):924-932, (2012).
- 5) S. Rifai, Dauphin, Yann, P.Vincent, Y. Bengio, and X.Muller, "The manifold tangent classifier. " In NIPS, 2294-2302, (2011a).
- 6) D. E. Rumelhart, G. E. Hinton and R. J.Williams, "Learning representations by back-propagating errors".*Nature* (323): (1986).
- 7) M. Collins, R. E. Schapire, and Y. Singer, "Logistic regression, Adaboost and Bregman distances. " *Machine Learning*, (48):253-285(2002).
- 8) G. E. Hinton , S. Osindero , Y.W. Teh, "A fast learning algorithm for deep belief nets ", *Journal in Neural Computation*,: 1527-1554(2006).
- 9) A. Krizhevsky , I. Sutskever , G.E Hinton, " ImageNet Classification with Deep Convolutional Neural Networks ", *Advances in Neural Information Processing Systems Conference*, 2012.
- 10) R. Girshick , J. Donahue , T. Darrell , J. Malik , "Rich feature hierarchies for accurate object detection and semantic segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015
- 11) J. Li, J.M. Bioucas-Dias, A. Plaza , "Spectral–Spatial Classification of Hyperspectral Data Using Loopy Belief Propagation and Active Learning", *IEEE transactions on geoscience and remote sensing*, vol. 51, no. 2, 2013.
- 12) C. Chen, W. Li, H. Su, K. Liu, "Spectral-Spatial Classification of Hyperspectral Image Based on Kernel Extreme Learning Machine", *Remote Sensing (MDPI)*, vol. 6(6):5795-5814(2014).
- 13) S. Sigtia, S. Dixon, "IMPROVED MUSIC FEATURE LEARNING WITH DEEP NEURAL NETWORKS", *International Conference on Acoustics, Speech and Signal Processing(IEEE)*, 6959 – 6963(2014).

- 14) F. J. Huang and Y. LeCun, "Large-scale learning with SVM and Convolutional for generic object categorization. " In CVPR, 284-291(2006).
- 15) Y. Lecun and Y. Bengio, "Convolutional Networks for Images, Speech and Time Series", The MIT Press, 255-258(1995).
- 16) M. Browne and S. S. Ghidary, "Convolutional neural networks for image processing: An application in robot vision. ", Australian Conference on Artificial Intelligence, Springer, 641-652(2003).
- 17) G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors. " CoRR, 2012.
- 18) L. Wan, M. D. Zeiler, S. Zhang, Y. LeCun, and R. Fergus, "Regularization of neural networks using dropconnect. " JMLR Proceedings(ICML), 1058-1066(2013).
- 19) A. G. Howard, "Some improvements on deep Convolutional neural network based image classification" CoRR, 2013.
- 20) M. D. Zeiler, and R. Fergus, "Visualizing and understanding Convolutional networks. " CoRR, 2013.
- 21) Y. Gong, Y. Jia, T. Leung, A. Toshev, and S. Ioffe, " Deep Convolutional ranking for multilabel image annotation. " CoRR, 2013.
- 22) O. Yadan, K. Adams, Y. Taigman, and M. Ranzato, "Multi-gpu training of convnets. " CoRR, 2013.
- 23) D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Parallel distributed processing: " Explorations in the microstructure of cognition", Chapter Learning Internal Representations by Error Propagation(MIT Press), 318-362(1986).
- 24) F. Huang, Y. LeCun, " Large-scale learning with svm and Convolutional nets for generic object categorization. " 2006.
- 25) Y. Tang, " Deep learning using linear support vector machines. " Workshop on Challenges in Representation Learning, ICML, 2013.
- 26) A. Shigeo, " Support Vector Machines for Pattern Classification ", Springer, Advances in Pattern Recognitions, 2005.
- 27) Y. Gong, Y. Jia, T. Leung, A. Toshev, S. Ioffe, "Comparison of L1 and L2 Support Vector Machines ", Proceedings of the International joint Conference, 2054-2059(2003)
- 28) <https://keras.io/>
- 29) A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images", 2009.

- 30)** P. K. Diederik and L. Ba Jimmy, "ADAM: A Method For Stochastic Optimization", International Conference on Learning Representations(ICLR),2015
- 31)** D. Stathakis, "How many hidden layers and nodes? ", International Journal of Remote Sensing, (30):2133-2147(2009).
- 32)** D. Cirean, U. Meier, and J. Schmidhuber, " Multi-column deep neural networks for image classification" , Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3642-3649:2012.
- 33)** D.C. Cirean, U. Meier , L.C. Gambardella, J. Schmidhuber, " Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition", Neural Competition,22(12):2010
- 34)** M. Ranzato, C. Poultney, S. Chopra and Y. LeCun, " Efficient Learning of Sparse Representations with an Energy-Based Model " Advances in Neural Information Processing Systems (NIPS):2006.
- 35)** P.Y. Simard, D. Steinkraus and J.C. Platt, "Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis", Intl. Conf. Document Analysis and Recognition, 958 – 962:200
- 36)** D. Clevert , T. Unterthiner , S. Hochreiter, "Fast And Accurate Deep Network Learning By Exponential Linera Units (ELUS) ",Internation Conference on Learning Representations (ICLR):2016
- 37)** M. Lin, Q. Chen, and S. Yan, " Network in network. " ICLR : 2014
- 38)** J. Nagi, F. Ducatelle, G. A. D. Caro, D. Cires, U. Meier, A. Giusti, and L. M. Gambardella, " Max-Pooling Convolutional Neural Networks for Vision-based Hand Gesture Recognition, " in IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 342–347:2011
- 39)** M. D. Zeiler, R. Fergus, " Stochastic pooling for regularization of deep convolutional neural networks. " CoRR, 2013.
- 40)** J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation. " CoRR, 2014.
- 41)** J. Bruna , W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs. " ,ICLR:2014.
- 42)** M. Couprie, F. Bezerra, and G. Bertrand, "Topological operators for grayscale image processing, " J. Electron. Imag., (4):1003–1015:2001.