

---

# MNIST Dataset of Handwritten Digits

---

Εργασία στα  
Νευρωνικά Δίκτυα

---

Χριστίνα Ισάκογλου – 2056  
christci@csd.auth.gr

---

## Εισαγωγή

Το πρόβλημα που επιλύθηκε είναι αυτό της αναγνώρισης χειρόγραφων ψηφίων στη βάση MNIST. Η υλοποίηση του RBF δικτύου έγινε σε MATLAB και τα αποτελέσματα διαφόρων πειραμάτων και παρατηρήσεις παρατίθενται παρακάτω.

## Δομή προγράμματος

Η ροή των εντολών προκειμένου να εκπαιδευτεί και να αξιολογηθεί το νευρωνικό δίκτυο βρίσκεται στο MATLAB file `rbf_run.m`. Αυτό κάνει χρήση τεσσάρων συναρτήσεων (`create_data_matrix`, `fdim`, `getCenters`, `rbf_correct`) και τα βήματα που ακολουθούν τα παραπάνω είναι τα εξής:

- Αρχικό διάβασμα των δεδομένων και αποθήκευση τους σε πίνακες εικόνων εκπαίδευσης και αξιολόγησης (οι ετικέτες και οι στόχοι περιέχονται στους προαναφερόμενους πίνακες) μέσω της `create_data_matrix`. Σε περίπτωση που η αξιολόγηση του νευρωνικού γίνεται σε υποσύνολο της MNIST βάσης, η συνάρτηση φροντίζει να εξάγει ισάριθμο πλήθος εικόνων από κάθε ψηφίο για να υπάρχουν επαρκείς αντιπρόσωποι από κάθε κλάση.
- Μείωση διάστασης με την `fdim` (οι default διαστάσεις στις οποίες μειώνονται τα δεδομένα είναι οι 160).
- Εκπαίδευση του κρυφού επιπέδου χωρίς επίβλεψη εντός της `rbf_run.m`: υπολογισμός της παραμέτρου  $\sigma$ , υπολογισμός κέντρων με επιλογή της κατάλληλης μεθόδου και κλήση της `getCenters` και τέλος υπολογισμός αποστάσεων του κάθε δείγματος από το κάθε κέντρο.
- Εκπαίδευση με επίβλεψη του στρώματος εξόδου: υπολογισμός της παραμέτρου  $\lambda$  και εκπαίδευση στην ουσία ενός γραμμικού `perceptron` με ένα στρώμα καθώς οι έξοδοι του προηγούμενου επιπέδου αποτελούν τις τιμές ενεργοποίησης του δικτύου.
- Αξιολόγηση του εκπαιδευμένου πια δικτύου στα 10.000 άγνωστα δεδομένα.
- Καταγραφή της απόδοσης του δικτύου τόσο κατά τη διαδικασία εκπαίδευσης όσο και κατά τη διαδικασία αξιολόγησης.

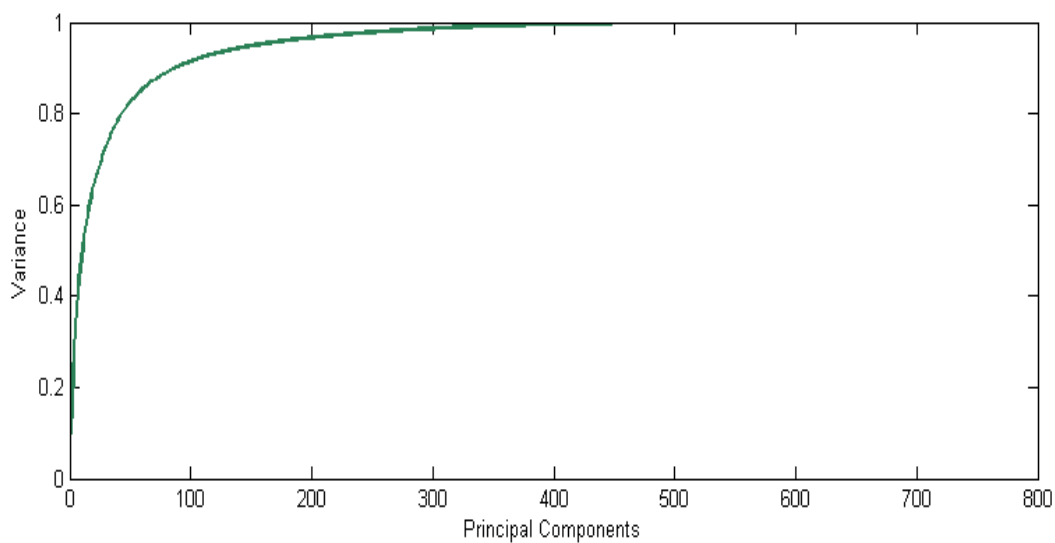
## Μείωση διάστασης

Οι 784 διαστάσεις του προβλήματος οδηγούν αργότερα σε μεγάλους πίνακες των οποίων ο υπολογισμός του επιβραδύνει την εκπαίδευση αρκετά. Για το λόγο αυτό προτιμήθηκε να γίνει μείωση των διαστάσεων του προβλήματος στις 160 με τη μέθοδο `pca` και διατηρώντας το 95% της αρχικής πληροφορίας.

Η μέθοδος `Principal Components Analysis` κάνει χρήση του πίνακα αυτοσυσχέτισης των διανυσμάτων παρατήρησης προκειμένου να βρεθούν πόσο συσχετισμένα είναι

μεταξύ τους τα χαρακτηριστικά των διανυσμάτων. Οι κύριες συνιστώσες που εξάγονται και αποτελούν το χώρο στον οποίο προβάλλουμε τα δεδομένα μας, αποτελούν γραμμικούς συνδυασμούς αρχικών χαρακτηριστικών και μέτρο του κατά πόσο διατηρείται η αρχική πληροφορία στις κύριες συνιστώσες που εξάγονται ενυπάρχει στις ιδιοτιμές του πίνακα αυτοσυσχέτισης.

Στο παρακάτω διάγραμμα αποτυπώνεται στην κλίμακα [0,1] το μέγεθος της πληροφορίας που αναλογεί σε διαφορετικό αριθμό κύριων συνιστωσών και έτσι δικαιολογείται η επιλογή των 160 πρώτων καθώς περιέχουν το 95.28%(0.9528). Βλέπουμε επίσης ότι και 784 κύριες συνιστώσες διατηρούν το σύνολο της πληροφορίας, όπως είναι αναμενόμενο.



Ο χώρος στον οποίο προβάλλονται τα δείγματα τεστ προβάλλονται στον ίδιο χώρο που προβλήθηκαν αυτά της εκπαίδευσης, κάνοντας χρήση του ίδιου γραμμικού συνδυασμού των αντίστοιχων χαρακτηριστικών.

Στη συνέχεια δίνεται ένα παράδειγμα εκπαίδευσης δικτύου την πρώτη φορά κάνοντας χρήση όλων των διαστάσεων του, ενώ τη δεύτερη κάνοντας χρήση του μειωμένου χώρου διαστάσεων που προκύπτει από τη PCA.

| Πίνακας (1)          |                      |            |                                    |  |                      |
|----------------------|----------------------|------------|------------------------------------|--|----------------------|
| Ακρίβεια εκπαίδευσης | Ακρίβεια αξιολόγησης | Διαστάσεις | Χρόνος εκπαίδευσης κρυφού επιπέδου | Χρόνος εκπαίδευσης εξωτερικού επιπέδου | Επιλογή κέντρων      |
| 95.58%               | 95.72%               | 784        | 176 sec                            | 239 sec                                | 1000 με τυχαίο τρόπο |
| 95.72%               | 95.74%               | 160        | 54 sec                             | 92 sec                                 | 1000 με τυχαίο τρόπο |

Παρατηρούμε δηλαδή ότι η απώλεια πληροφορίας επιφέρει από τη μία σημαντική επιτάχυνση της διαδικασίας και από την άλλη παρόμοια αποτελέσματα στην απόδοση. Η μείωση των

## Ανομοιογένεια στην εκπαίδευση

### Εκπαίδευση κρυφού στρώματος

Η εκπαίδευση του κρυφού στρώματος αποτελεί το τμήμα της εκπαίδευσης χωρίς επίβλεψη στο νευρωνικό. Οι παράμετροι που πρέπει να εκπαιδευτούν είναι το κέντρο κάθε συνάρτησης και το εύρος αυτής.

Ο κάθε νευρώνας του κρυφού επιπέδου αντιστοιχεί σε ένα από τα κέντρα που επιλέγονται. Συνεπώς όσα τα κέντρα που επιλέγονται τόσοι και οι νευρώνες του κρυφού επιπέδου. Για να πετύχει ένα RBF δίκτυο επίδοση παρόμοια με εκείνη ενός perceptron πολλών επιπέδων θα πρέπει να είναι υψηλής τάξης[1], το κρυφό του επίπεδο δηλαδή να εκπροσωπείται από πολλά κέντρα. Αυτό συμβαίνει λόγω της τοπικότητας των συγκεκριμένων δικτύων που αποτελεί εμπόδιο στην προσπάθεια μας οι συναρτήσεις ενεργοποίησης να καλύπτουν όλο τον χώρο στον οποίο ορίζονται τα δεδομένα μας, καθώς ο αριθμός των κέντρων εμφανίζει μια εκθετική εξάρτηση από τη διάσταση του χώρου εισόδου. (curse of dimensionality – κατάρα της διαστατικότητας).

Από το παραπάνω χαρακτηριστικό προκύπτουν τα παρακάτω ζητήματα τα οποία και αναλύονται περισσότερο κατά την έκθεση αποτελεσμάτων από πειράματα με διαφορετικές μεθόδους εξαγωγής κέντρων: αφενός ο αριθμός των κέντρων θα πρέπει να είναι μεγάλος προκειμένου το δίκτυο να πετυχαίνει υψηλή ακρίβεια, αφετέρου όμως όχι τόσο μεγάλος ώστε η ταχύτητα του να αντισταθμίζει την απόδοση του. Ακόμη, καθώς γίνεται φανερό ότι σε ένα πρόβλημα 60.000 δειγμάτων εκπαίδευσης η επιλογή και των 60.000 ως κέντρα δε θα χαρακτηριζόταν ως αποδοτική προκύπτει η σημασία της επιλογής των κέντρων με σκοπό κάθε κλάση να καλύπτεται επαρκώς.

Παρακάτω αναλύονται τρεις διαφορετικοί τρόποι επιλογής κέντρων:

..... Επιλογή όλων των δειγμάτων .....

Πειράματα με βάση αυτόν τον τρόπο επιλογής κέντρων έγιναν σε υποσύνολα του αρχικού σετ με σκοπό να δοκιμαστεί η επάρκεια της μεθόδου.

*Η βασικά παρατήρηση στη συγκεκριμένη μέθοδο είναι ότι εφόσον αντιστραφεί ο τρόπος απόδοσης κλάσης με βάση την έξοδο στο νευρωνικό(rbf\_correct) προκειμένου αντί για τη μέγιστη τιμή να ψάχνει την ελάχιστη τιμή της εξόδου και αυτή να θεωρεί ως label τα αποτελέσματα που προκύπτουν είναι: σε σετ δεδομένων 1000 δειγμάτων με 1000 κέντρα η εκπαίδευση είναι 97.8% σωστή και η αξιολόγηση 85.6, ενώ στο ίδιο υποσύνολο με τα διπλάσια κέντρα η εκπαίδευση είναι άριστη και η αξιολόγηση 86.1%.*

..... Επιλογή τυχαίων δειγμάτων .....

Η επιλογή τυχαίων δειγμάτων οδηγεί στο ακόλουθο πρόβλημα : σε περίπτωση που δεν εξάγουμε αρκετά κέντρα από τα δείγματα μας υπάρχει η περίπτωση να υπάρχουν δείγματα που δεν έχουν στα κέντρα κάποιον εκπρόσωπο της κλάσης τους με αποτέλεσμα οι αποστάσεις που θα παρθούν από όλα τα κέντρα να είναι μειωμένες και να μη μπορούν να βγουν συμπεράσματα για το συγκεκριμένο δείγμα.

Όταν όμως τα δείγματα είναι αρκετά βλέπουμε στον Πίνακα (1) ότι η απόδοση προσεγγίζει όντως αυτή ενός *percepton* πολλών στρωμάτων. Σε διαφορετική περίπτωση, για παράδειγμα με την επιλογή μόνο 100 τυχαίων κέντρων η εκπαίδευση είναι σωστή κατά 87.75, ενώ η αξιολόγηση κατά 88.45%.

..... Αλγόριθμος ομαδοποίησης.....

Το πρόβλημα της σωστής αντιπροσώπευσης της κάθε κλάσης από το κέντρο της προσεγγίζει η λογική της εφαρμογής κάποιου αλγορίθμου ομαδοποίησης. Ο *k-means* όντας ένας τέτοιος αλγόριθμος αναδεικνύει κέντρα τα οποία σε αντίθεση με τις προηγούμενες περιπτώσεις δεν ανήκουν στα δείγματα εκπαίδευσης, καθώς υπολογίζονται από επαναληπτικούς μέσους όρους των δειγμάτων που εντάσσονται σε κάθε επανάληψη σε κάθε κλάση, ως κοντινότερα στο κέντρο αυτής.

Μια πρώτη σκέψη είναι από τη στιγμή που ο αλγόριθμος μας εξασφαλίζει ότι θα πάρει χαρακτηριστικά δείγματα για κάθε κέντρο μπορούμε να θεωρήσουμε ότι τα κέντρα μας μπορούν να είναι 10.

Προκειμένου να αποφύγουμε την αναζήτηση 10 μέσων σε ένα σετ δεδομένων 60.000 που θα οδηγούσε σε μεγάλο αριθμό επαναλήψεων, παίρνουμε τυχαία ένα υποσύνολο 1000 δειγμάτων και υπολογίζουμε εκεί τους μέσους. Το αποτέλεσμα ήταν το δίκτυο να εκπαιδευτεί σε συνολικό χρόνο 10 δευτερολέπτων ενώ η ακρίβεια του να περιορίζεται σε 63.56% στην εκπαίδευση και σε 64.38% στην αξιολόγηση.

Αυξάνοντας το μέγεθος του τυχαίου υποσυνόλου σε 10.000, αλλά και τον αριθμό επαναλήψεων (από 100 σε 2000) δεν οδηγούμαστε σε αισθητά καλύτερα αποτελέσματα (66.30% στην εκπαίδευση, 67.58% στην αξιολόγηση και σε χρόνο 34 δευτερολέπτων. Εφαρμόζοντας *k-means* σε όλο το σετ δεδομένων εκπαίδευσης η απόδοση δε βελτιώνεται (64.07% στην εκπαίδευση, 64.79 στην αξιολόγηση) ενώ ο χρόνος εκπαίδευσης επιβραδύνεται κατά 2 λεπτά.

Οι παρατηρήσεις αυτές μας οδηγούν στο συμπέρασμα ότι 10 κέντρα δεν επαρκούν στη σωστή αντιπροσώπευση των δεδομένων, καθώς θα υπάρχουν πάντα ψηφία διαφορετικά από αυτό που θεωρείται πρότυπο για το εκάστοτε ψηφίο και τα συγκεκριμένα ψηφία με τις ιδιαιτερότητες τους όντας *outliers* του υπόλοιπου σετ θα αποτυγχάνουν τόσο στη φάση της εκπαίδευσης όσο και στη φάση της αξιολόγησης. Έτσι, θα μπορούσε κάποιος να ισχυριστεί ότι το ποσοστό των ψηφίων που ακολουθούν τους “κανόνες” και συμφωνούν με τη νόρμα γραφής κάθε ψηφίου είναι περίπου το 65% ενώ το υπόλοιπο αποτελείται από ψηφία με χαρακτηριστικά που τα διαφοροποιούν από τα υπόλοιπα της κλάσης τους.

Έτσι, συμπεραίνουμε ότι αν αυξήσουμε τον αριθμό των μέσων που υπολογίζουμε τα ποσοστά επιτυχίας μας θα μεγαλώσουν καθώς θα λαμβάνουμε υπόψη μας περισσότερες παραλλαγές ενός ψηφίου. Πράγματι εφαρμόζοντας τον k-means για 50 κέντρα το ποσοστό σωστής εκπαίδευσης ανέρχεται στο 88.22%, ενώ αυτό της αξιολόγησης σε 89.08%. Η εκπαίδευση ολοκληρώνεται σε 101 δευτερόλεπτα και βλέπουμε ότι η γενίκευση του δικτύου είναι αρκετά καλή.

| Πίνακας (2)          |                      |            |                                    |  |                 |
|----------------------|----------------------|------------|------------------------------------|--|-----------------|
| Ακρίβεια εκπαίδευσης | Ακρίβεια αξιολόγησης | Διαστάσεις | Χρόνος εκπαίδευσης κρυφού επιπέδου | Χρόνος εκπαίδευσης εξωτερικού επιπέδου | Επιλογή κέντρων |
| 91.56%               | 92.01%               | 160        | 155.49 sec                         | 3.67                                   | 100 με kmeans   |
| 94.52%               | 94.64%               | 160        | 382.62 sec                         | 10.61 sec                              | 300 με kmeans   |

.....

Ζητούμενο των παραπάνω είναι, όπως αναφέρθηκε, ο υπολογισμός των αποστάσεων του κάθε δείγματος από το κέντρο που αντιπροσωπεύει κάθε νευρώνας. Οι αποστάσεις αυτές υπολογίζονται με βάση τον τύπο της ευκλείδειας απόστασης και αποθηκεύονται σε έναν πίνακα  $G$ , ο οποίος είναι διαστάσεων  $n \times m$ , όπου  $n$  τα δείγματα που εισέρχονται στο δίκτυο και  $m$  ο αριθμός των κέντρων. Ο χαρακτηρισμός των δικτύων ως δίκτυα συναρτήσεων βάσης ακτινικού τύπου γίνεται στο βήμα ακολούθως του υπολογισμού του  $G$ . Οι αποστάσεις αυτές μετασχηματίζονται με χρήση της συνάρτησης Gauss  $f(x) = \exp(-||x-c||^2/2\sigma^2)$ , με αποτέλεσμα ο πίνακας πλέον να περιέχει τιμές, όπου η κάθε μία υποδηλώνει πως όσο μεγαλύτερη είναι τόσο πιο μικρή είναι η απόσταση μεταξύ του δείγματος (που αντιπροσωπεύεται από τη γραμμή του κελιού) και του κέντρου (που αντιπροσωπεύεται από τη στήλη του κελιού). Οι συναρτήσεις RBF, δηλαδή που αποτελούν τις συναρτήσεις ενεργοποίησης που εξάγονται από αυτό το βήμα της εκπαίδευσης, έχουν την ενδογενή δυνατότητα να αναγνωρίζουν εκείνα τα ορίσματα που είναι κοντά στο κέντρο της.

## Εκπαίδευση εξωτερικού στρώματος

Η εκπαίδευση του εξωτερικού στρώματος αποτελείται ουσιαστικά από την εκμάθηση των βαρών των συναρτήσεων ενεργοποίησης (συναρτήσεων ακτινωτής βάσης) και είναι αυτό που επιβραδύνει περισσότερο την εκπαίδευση (εκτός της περίπτωσης που στην εκπαίδευση του κρυφού στρώματος εφαρμόζεται αλγόριθμος ομαδοποίησης).

Ο συντελεστής που έχει ιδιαίτερα σημασία σε αυτό το στάδιο της εκπαίδευσης είναι ο η παράμετρος κανονικοποίησης  $\lambda$ . Ένα πρόβλημα είναι δυνατόν να είναι ένα κακώς-ορισμένο πρόβλημα (ill-posed problem), γεγονός που σημαίνει πως ένα μεγάλο σετ δεδομένων περιέχει μικρό μέγεθος πληροφορίας για το πρόβλημα που επιθυμούμε να

λύσουμε[4]. Έτσι, το δίκτυο καταλήγει να υπερεκπαιδευτεί στα δεδομένα που του δίνονται και να αποτυγχάνει σε όσα καινούρια του δοθούν.

Μια μη ικανοποιητική λύση του παραπάνω προβλήματος θα ήταν ο περιορισμός των νευρώνων του κρυφού επιπέδου. Η παράμετρος  $\lambda$  όμως χρησιμεύει ως ένα μέτρο ένδειξης για εμάς της σπουδαιότητας της αυστηρής κατασκευής του υπερεπιπέδου χωρίς λάθη έναντι της αποδοχής της μεγαλύτερης καμπυλότητας αυτού προκειμένου να περικλείει και δείγματα που δεν εμφανίζονται στην εκπαίδευση.

## Σύγκριση αποτελεσμάτων

Στη συνέχεια παρατίθενται συνοπτικά οι συγκρίσεις μεταξύ 3 ταξινομητών και το καλύτερο αποτέλεσμα του δικτύου που υλοποιήθηκε.

Βλέπουμε ότι η απόσταση αποτέλεσε ένα καλό κριτήριο ταξινόμησης με τον 3-κοντινότερο γείτονα, που είχε όμως ως μειονέκτημα τον μεγάλο χρόνο εκπαίδευσης. Στο RBF νευρωνικό καταφέραμε να οδηγηθούμε σε παρόμοια υψηλό αποτέλεσμα με καλή γενίκευση, κάνοντας χρήση των αποστάσεων και πετυχαίνοντας πολύ καλύτερους χρόνους.

| Πίνακας(3)           |            |                    |                    |
|----------------------|------------|--------------------|--------------------|
| Ακρίβεια αξιολόγησης | Διαστάσεις | Χρόνος εκπαίδευσης | Ταξινομητής        |
| 97.17%               | 784        | 16min              | 3-nearest neighbor |
| 69.68%               | 784        | 0.08min            | Nearest centroid   |
| 94.92%               | 784        | 111 min            | Mlp perceptron     |
| 95.74%               | 160        | 146 sec            | RBF neural network |

## Αναφορές

- [1] S. Theodoridis, K. Koutroumpas, “Αναγνώριση Προτύπων”
- [2] K. Γουλιανός, “Εργαστηριακές Ασκήσεις Νευρωνικών Δικτύων”
- [3] K. Διαμαντάρας, “Τεχνητά Νευρωνικά Δίκτυα”
- [4] S. Haykin, “Neural Networks: A Comprehensive Foundation”