

---

# MNIST DATASET

## OF HANDWRITTEN DIGITS

---

Χριστίνα ΙΣΑΚΟΓΛΟΥ - 2056  
christci@csd.auth.gr

### Εφαρμογή των αλγορίθμων **k-nearest** και **nearest centroid**

Σε πρώτη φάση υλοποιήθηκαν σε MATLAB οι αλγόριθμοι **k-nearest** και **nearest centroid** και εφαρμόστηκαν σε ολόκληρο το σετ δεδομένων(60.000 training samples και 10.000 test samples).

Η κάθε εικόνα-ψηφίο αναπαρίσταται, μετά από την αρχική ανάγνωση από τα αρχεία που περιέχουν τα δεδομένα, από έναν πίνακα  $28 \times 28$ , όπου υπάρχει και η αντιστοιχία κάθε pixel με στοιχείο πίνακα. Στη συνέχεια ο πίνακας αυτός μετασχηματίζεται σε ένα διάνυσμα γραμμή  $1 \times 784$ , όπου πλέον κάθε pixel αντιστοιχίζεται σε στήλη, για τη διευκόλυνση της μετέπειτα αναγκαίας αναζήτησης πλησιέστερων διανυσμάτων χαρακτηριστικών.

- **k-nearest** : Ως μετρική απόστασης γειτόνων χρησιμοποιείται η ευκλείδεια απόσταση μεταξύ των διανυσμάτων χαρακτηριστικών που αντιπροσωπεύουν κάθε εικόνα(γραμμή του μετασχηματισμένου πίνακα). Η εύρεση του κοντινότερου γείτονα γίνεται μέσω της ενσωματωμένης συνάρτησης του MATLAB knnsearch. Στην περίπτωση του 1-nearest η κλάση του κοντινότερου train sample του εκάστοτε test sample γίνεται αυτομάτως και κλάση του ίδιου, ενώ στην περίπτωση του 3-nearest η συχνότερα εμφανιζόμενη κλάση ανάμεσα στους τρεις κοντινότερους γείτονες είναι και αυτή στην οποία κατατάσσεται το test sample . Αντίθετα, σε περίπτωση ισοψηφίας(αναφορικά με την εμφάνιση των κλάσεων) ο 3-nearest ανάγεται στον 1-nearest. Τα αποτελέσματα της εφαρμογής των αλγορίθμων για  $k = 1, 3$  είναι τα παρακάτω:

```
>> k_nearest_neighbor (60000,10000,1);  
Time needed 973.224655 seconds  
9691 test samples out of 10000 were classified correctly  
Test error rate : 3.09 %
```

```
>> k_nearest_neighbor (60000,10000,3);  
Time needed 923.831615 seconds  
9717 test samples out of 10000 were classified correctly  
Test error rate : 2.83 %
```

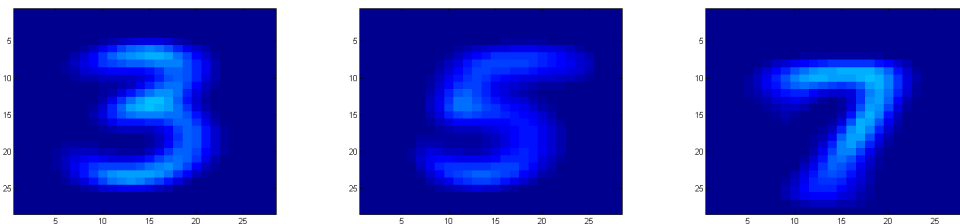
- **nearest centroid** : Η προεπεξεργασία δεδομένων που είναι αναγκαία για αυτό τον αλγόριθμο είναι η εξαγωγή της 'μέσης εικόνας-ψηφίου'. Έτσι, δημιουργούνται συνολικά δέκα κλάσεις και από αυτές η κοντινότερη σε κάθε test sample θα είναι και αυτή που θα της αποδώσει το label της. Τα αποτελέσματα του αλγορίθμου είναι τα παρακάτω:

```
>> [ means,labels_nearest ] = nearest_centroid(60000,10000);
Time needed 5.099676
6968 test samples out of 10000 were classified correctly
Test error rate : 30.32 %
```

Πίνακας 1: Συνοπτική περιγραφή των επιδόσεων των αλγορίθμων

| Αλγόριθμος       | Χρόνος(σε λεπτά) | Ποσοστό λάθους | Δείγματα ταξινομημένα σωστά |
|------------------|------------------|----------------|-----------------------------|
| 1-nearest        | 16               | 3.09%          | 9691                        |
| 3-nearest        | 16               | 2.83%          | 9717                        |
| nearest centroid | 0.08             | 30.32%         | 6968                        |

Σχήμα 1: Παράδειγμα μέσω εικόνων για τα ψηφία 3,5,7.



## Εφαρμογή πολυστρωματικού Perceptron

Σε δεύτερη φάση υλοποιήθηκε ο Multilayer Perceptron[2], με την ύπαρξη ενός κρυφού στρώματος και εφαρμόστηκε με διαφορετικές παραμέτρους σε πειράματα που θα αναλυθούν, για την εξαγωγή συμπερασμάτων. Ο αλγόριθμος στον οποίο βασίζεται η εκπαίδευση είναι ο Back-propagation, κάνοντας προώθηση των σφαλμάτων προς τα πίσω, στους νευρώνες, και έχοντας στη διάθεση του διανύσματα στόχων (μάθηση με επίβλεψη) για τα δείγματα εκπαίδευσης. Το πρόγραμμα που περιλαμβάνει για λόγους συνοχής τις συναρτήσεις που διαβάζουν τα δεδομένα, υλοποιούν την εκπαίδευση και ελέγχουν την εγκυρότητα αυτής τόσο στα δείγματα εκπαίδευσης, όσο και στα δείγματα τεστ, είναι αυτό με όνομα **mlp\_run.m**.

Ιδιαίτερη σημασία για τον αλγόριθμο εκπαίδευσης έχει ο πίνακας στον οποίο είναι αποθηκευμένα τα διανύσματα χαρακτηριστικών των δειγμάτων εκπαίδευσης, τα οποία εισέρχονται ως είσοδοι στους νευρώνες μηδενικού στρώματος, και τα διανύσματα στόχων, που χρησιμεύουν στην αναπροσαρμογή των βαρών του δικτύου. Ο πίνακας αυτός έχει τη μορφή

$$\mathbf{P} = [input_1 \ input_2 \ \dots \ target_1 \ target_2],$$

όπου καθένα από τα στοιχεία αποτελεί ένα διάνυσμα γραμμής. Στην περίπτωση της εισόδου αποθηκεύονται και τα 784 pixels της κάθε εικόνας, κανονικοποιημένα σε 0-1, χωρίς την εξαγωγή χαρακτηριστικών. Αντίστοιχα, στην περίπτωση των στόχων στις 10 πρώτες θέσεις αποθηκεύονται οι τιμές που είναι επιθυμητό να εξάγει το δίκτυο σε καθένα από τους 10 νευρώνες εξόδου του, με την παραδοχή ότι κάθε νευρώνας παραπέμπει σε μία από τις 10 κλάσεις ψηφίων που ζητείται να αναγνωριστούν. Στην θέση 11 του διανύσματος στόχου αποθηκεύεται η κλάση στην οποία ανήκει το δείγμα για λόγους ευκολίας κατά τον έλεγχο των αποτελεσμάτων. Έτσι για παράδειγμα έχουμε ένα δείγμα εκπαίδευσης να ανήκει στην κλάση 1. Το διάνυσμα στόχου του θα είναι το [0.05 0.95 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 1.0]. Οι αριθμοί που χρησιμοποιούνται για την επιθυμητή έξοδο των νευρώνων βλέπουμε ότι αποφεύγουν να είναι οι μέγιστες τιμές που μπορεί να πιάσει η συνάρτηση ενεργοποίησης που χρησιμοποιείται (σιγμοειδής συνάρτηση) ούτως ώστε να μην επέλθει κορεσμός.

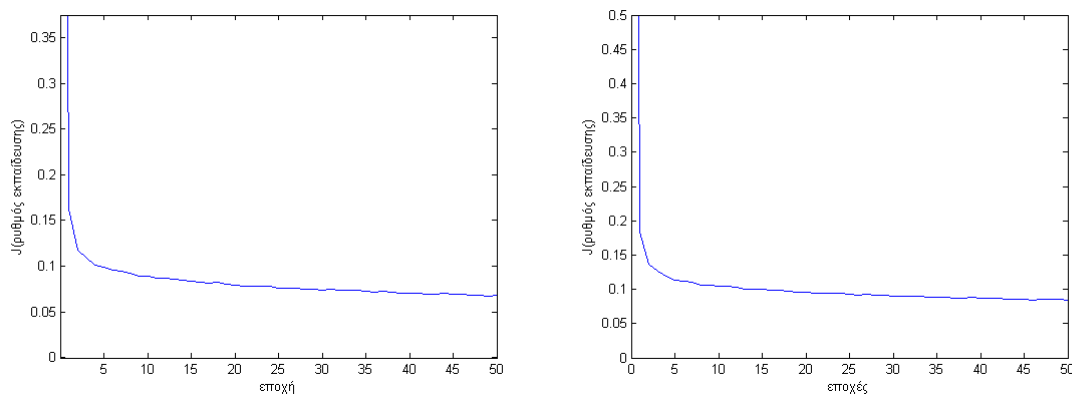
Στον παρακάτω πίνακα αναγράφονται οι επιδόσεις δύο διαφορετικών πειραμάτων του αλγορίθμου με παραμέτρους  $n = 20$  (για τον αριθμό των νευρώνων),  $\alpha = 0.9$  (ορμή) και  $\eta = 0.4$  (ρυθμός μάθησης) στην πρώτη

περίπτωση και  $n=15$ ,  $\alpha=0.7$  και  $\eta=0.45$  στη δεύτερη. Τα εκπαιδευμένα αυτό δίκτυα(επαλήθευση των αποτελεσμάτων μπορεί να γίνει με την ανάκληση του matlab workspace *training\_1.mat* για το πρώτο - *training\_2.mat* για το δεύτερο και την κλήση της συνάρτησης *test\_mlp* τόσο για τα δεδομένα εκπαίδευσης, όσο και τεστ) σημείωσαν τις υψηλότερες επιδόσεις ανάμεσα στα άλλα που δοκιμάστηκαν με διαφορετικές παραμέτρους. Ο ρυθμός εκπαίδευσης τους εμφανίζεται στην Σχήμα 2, όπου και κάποιος μπορεί να παρατηρήσει τη γρηγορότερη σύγκλιση στην πρώτη περίπτωση λόγω της αυξημένης ορμής(αναλύεται παρακάτω).

Πίνακας 2: Επιδόσεις εκπαίδευσης μετά από 200 εποχές για  $n=20, \alpha=.9$  και  $\eta=.4$ ,  $n=15, \alpha=.7$ ,  $\eta=.45$

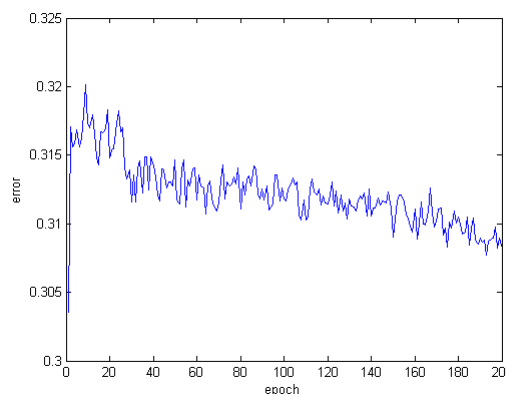
| Ποσοστό επιτυχίας εκπαίδευσης | Ποσοστό επιτυχίας τεστ | Χρόνος(σε λεπτά) |
|-------------------------------|------------------------|------------------|
| 97.65%                        | 94.92%                 | 111              |
| 97.02%                        | 94.19%                 | 74               |

Σχήμα 2: Ρυθμοί εκπαίδευσης των παραπάνω εκπαιδευμένων δικτύων στο διάστημα των 50 πρώτων εποχών.



Ενδεικτικά παρατίθεται, στο Σχήμα 3, η εξέλιξη του συνολικού λάθους των εξόδων ανά εποχή στην εκπαίδευση του πρώτου δικτύου, προκειμένου να φανεί η μείωση αυτού.

Σχήμα 3: Ρυθμοί εκπαίδευσης των παραπάνω εκπαιδευμένων δικτύων.



Αναφορικά με τις επιδόσεις του παρόντος αλγορίθμου σε σύγκριση με αυτές της προηγούμενης ενότητας, αξίζει να σημειωθεί ότι ενώ μετά το πέρας 200 εποχών ο perceptron φαίνεται να υστερεί, δεδομένου ότι στην εκπαίδευση τα πηγαίνει καλύτερα, αλλά στα δειγμάτα των τεστ χειρότερα, έχει αποδειχτεί πως με μεγαλύτερη εκπαίδευση, και συγκεκριμένα 1000 εποχών εφαρμόζοντας τις παραμέτρους της δεύτερης περίπτωσης, ο αλγόριθμος υπερτερεί έναντι του 3-NN[1].

Στη συνέχεια θα γίνουν κάποιες παρατηρήσεις και συμπεράσματα που εξάχθηκαν με βάση διαφορετικές τροποποιήσεις στις διάφορες παραμέτρους του δικτύου και την επιρροή που αυτές είχαν τόσο κατα τη δια-

δικασία της εκπαίδευσης όσο και στη διαδικασία ελέγχου της σωστής κατηγοριοποίησης των δειγμάτων τεστ, έχοντας ως κριτήριο αναφοράς αυτό της καλύτερης εκπαίδευσης που επιτεύχθηκε.

Τέλος, θα αναλυθούν κάποιες περιπτώσεις παραδείγματα που το δίκτυο κατατάσσει εσφαλμένα τα δείγματα προκειμένου να αναδειχτεί η δυσκολία που αντιμετωπίζει και το οδηγεί σε λάθος.

### Αριθμός νευρώνων

Στο συγκεκριμένο πρόβλημα όπως ήδη διατυπώθηκε ένας αριθμός 20 νευρώνων μαζί με τις παραμέτρους που αναφέρθηκαν οδηγεί σε μια καλή εκπαίδευση και ταυτόχρονα σε μια καλή κατάταξη αγνωστων δειγμάτων (η κλάση των δειγμάτων τεστ είναι άγνωστη σε αντίθεση με τα δείγματα που χρησιμοποιούνται στην εκπαίδευση. Αντίθετα, η σημαντική μείωση των αριθμών νευρώνων οδηγεί σε ελαττωματική εκπαίδευση και έχει και ως αποτέλεσμα την κατά μεγαλύτερο ποσοστό εσφαλμένη κατάταξη δειγμάτων τεστ. Χαρακτηριστικά αναφέρονται οι επιδόσεις του αλγορίθμου για αριθμό νευρώνων 5 και 1 όπου παρατηρούμε πως το ποσοστό σωστής κατηγοριοποίησης έχει μειωθεί αισθητά, όπως όμως και ο χρόνος εκπαίδευσης, όπου για παράδειγμα στους 10 νευρώνες έχει μειωθεί στα 28 λεπτά.

Πίνακας 3: Επιδόσεις εκπαίδευσης μετά από 200 εποχές για  $n=10, \alpha=.9$  και  $\eta=.4$ ,  $n=5, \alpha=.9$ ,  $\eta=.4$

| Ποσοστό επιτυχίας εκπαίδευσης | Ποσοστό επιτυχίας τεστ |
|-------------------------------|------------------------|
| 94.09%                        | 92.69%                 |
| 85.53%                        | 83.58%                 |

Αντίστοιχα, η εκπαίδευση με περισσότερους νευρώνες μπορεί να οδηγήσει σε υπερκπαίδευση και το δίκτυο θα καταλήξει να ταξινομεί πολύ καλά τα δείγματα εκπαίδευσης, αλλά θα αποτυγχάνει σημαντικά στο να ταξινομεί δείγματα τεστ, που δεν γνωρίζει.

### Βήμα εκπαίδευσης και Ορμή

Το βήμα εκπαίδευσης αποτελεί ένα μέτρο του κατά πόσο γρήγορα θα συγκλίνει ο αλγόριθμος και πόσο γρήγορος θα είναι ο ρυθμός εκπαίδευσης. Οι τιμές που δίνονται στο βήμα κινούνται στην κλιμακα 0-1, με τις τιμές κοντά στο 0 να οδηγούν σε αργή σύγκλιση και χαμηλό ρυθμό εκπαίδευσης, ενώ με μεγάλες τιμές σε κακή σύγκλιση, καθώς ο αλγόριθμος οδηγείται σε ταλαντώσεις στην καλύτερη περίπτωση, ενώ στη χειρότερη οι τιμές των βαρών οδηγούνται στο άπειρο, όπως στην περίπτωση μας αν το δίκτυο εκπαιδευτεί με βήμα ίσο με 0.9.[3]

Στο δίλημμα που αντιμετωπίστηκε πιο πάνω για την κατάλληλη επιλογή του βήματος εκπαίδευσης, τη λύση δίνει μια επιπλέον παράμετρος, αυτή της ορμής. Με βασικό σκοπό να αποφευχθούν οι ταλαντώσεις, ενώ ταυτόχρονα επιταχύνεται η σύγκλιση, η ορμή αυτό που καταφέρνει να κάνει είναι να συνδέσει τη διόρθωση ενός βάρους σε μια επανάληψη με τη διόρθωση που εφαρμόστηκε πάνω του στην προηγούμενη επανάληψη. Με αυτόν τον τρόπο διατηρείται η πορεία της κατάβασης.[3]

## Αναφορές

- [1] K. Roy, C. Chaudhuri, M. Kundu, M. Nasirupi, D. K. Basu, "Comparison of the Multi Layer Perceptron and the Nearest Neighbor Classifier for Handwritten Numeral Recognition", Journal of information science and engineering 21, 1247-1259, 2005.
- [2] S. Haykin, "Neural Networks: A comprehensive foundation", Pearson Education, 2nd edition, 1999.
- [3] Κ. Διαμαντάρας, "Τεχνητά Νευρωνικά Δίκτυα", Κλειδάριθμος, Αθήνα, 2007.