



## ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ

Διδάσκων: Δ. Κουτσομητρόπουλος

Ακαδημαϊκό Έτος 2020-2021

### Εργαστηριακή Άσκηση Μέρος Α'

#### Α. Αναγνώριση χειρόγραφων χαρακτήρων (αριθμητικών ψηφίων) με Χρήση Νευρωνικών Δικτύων

Συστήματα επεξεργασίας εικόνας (image processing systems) εφαρμόζουν μεθόδους πάνω σε μια ψηφιακή εικόνα με σκοπό την εξαγωγή χαρακτηριστικών. Η χρήσιμη αυτή πληροφορία που θα προκύψει στην έξοδο του συστήματος, θα περιέχει στοιχεία άμεσα σχετιζόμενα με την αντίστοιχη εικόνα που θα δοθεί ως είσοδος. Μια τέτοια χαρακτηριστική περίπτωση με ιδιαίτερο επιστημονικό ενδιαφέρον αποτελεί και η προσπάθεια αναγνώρισης χειρόγραφων χαρακτήρων, όπως είναι τα αριθμητικά ψηφία. Συγκεκριμένα, δοθέντων εικόνων που αναπαριστούν ακριβώς ένα ακέραιο ψηφίο που ανήκει στο διάστημα  $[0,9]$ , θα πρέπει να υπάρξει υλοποίηση που να απαντά το ερώτημα του ποιος τελικά είναι αυτός ο αριθμός. Για παράδειγμα, στην ακόλουθη εικόνα αποτυπώνονται 100 χειρόγραφα ψηφία, με το κάθε ένα να έχει ανάλυση  $28 \times 28 = 784$  pixels.

```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
```

Διάφορες τεχνικές έχουν προταθεί κατά καιρούς στη βιβλιογραφία, μεταξύ άλλων και η χρήση μοντέλων μηχανικής μάθησης για την εύρεση αυτών των προβλέψεων. Επομένως, στην παρούσα εργασία θα εξετάσετε τη χρήση ενός πολυεπίπεδου ΤΝΔ για την πρόβλεψη της ακέραιας τιμής που αναπαριστά η κάθε εικόνα. Για τον σκοπό αυτό θα αξιοποιηθεί το σύνολο δεδομένων *MNIST*<sup>1</sup> που συγκεντρώνει 60K δείγματα για εκπαίδευση και 10K δείγματα για έλεγχο (<https://www.kaggle.com/oddrational/mnist-in-csv>). Τα δεδομένα αυτά περιέχονται στο αρχείο *mnist\_train.csv* και *mnist\_test.csv*. Η κάθε εγγραφή - γραμμή εκτός από πραγματική ετικέτα περιλαμβάνει και το σύνολο των 784 pixels που σχηματίζουν την ίδια την εικόνα (το κάθε pixel λαβαίνει τιμές από 0 έως 255). Οπότε σχηματικά έχουμε τις παρακάτω στήλες, χωρισμένες με κόμμα.

label , pixel\_1 , pixel\_2 , ... , pixel\_784

Για την υλοποίηση των αλγορίθμων μπορείτε να χρησιμοποιήσετε οποιοδήποτε περιβάλλον, βιβλιοθήκη ή γλώσσα προγραμματισμού κρίνετε σκόπιμο. Ενδεικτικά αναφέρονται *MatLab*, *WEKA*, *Azure ML Studio*, *Google Colaboratory*, *TensorFlow*, *Keras*, *SciKit-Learn*.

Το ζητούμενο στην εργασία αυτή είναι να κατασκευαστεί και να εκπαιδευτεί ένα ΤΝΔ που να ταξινομεί εισόδους (εικόνες με χειρόγραφα ψηφία) σε δέκα διαφορετικές κλάσεις (ακέραιες τιμές).

<sup>1</sup> [https://en.wikipedia.org/wiki/MNIST\\_database](https://en.wikipedia.org/wiki/MNIST_database)

## A1. Προεπεξεργασία και Προετοιμασία δεδομένων [15 μονάδες]

Προσοχή: Ό,τι μετασχηματισμοί εφαρμοστούν στα δεδομένα του συνόλου εκπαίδευσης, οι ίδιοι θα πρέπει να εφαρμοστούν και στα δεδομένα του συνόλου ελέγχου ή εναλλακτικά να αντιστραφούν πρώτου μετρηθούν οι μετρικές αξιολόγησης παρακάτω.

α) Οι τιμές των pixels της εικόνας, όπως αναφέρθηκε, είναι ακέραιες και κινούνται στο διάστημα  $[0,255]$ . Με δεδομένη την συγκεκριμένη αποτύπωση και την πιθανή ανάγκη προσαρμογής των τιμών αυτών σε διαφορετική κλίμακα (εξάλειψη ενδεχόμενης πόλωσης), παρουσιάζονται οι εξής τρεις μέθοδοι:

- *Κεντράρισμα (Centering)*: Με την μέθοδο αυτή αφαιρούμε τον μέσο όρο των τιμών των pixel μιας εικόνας από όλες τις τιμές που έχουν ήδη αποδοθεί.
- *Κανονικοποίηση (Normalization)*: Με την μέθοδο αυτή μεταφέρουμε το εύρος τιμών των pixel των δειγμάτων σε νέα κλίμακα πχ  $[0,1]$ .
- *Τυποποίηση (Standardization)*: Με την μέθοδο αυτή παρέχουμε στο δείγμα ιδιότητες όπως μηδενική μέση τιμή και μοναδιαία διακύμανση (Gaussian).

Εξετάστε τη χρησιμότητα των ανωτέρω μεθόδων για το συγκεκριμένο πρόβλημα και εφαρμόστε τη/τις στα δεδομένα εκπαίδευσης, αν κρίνετε σκόπιμο. [10]

β) Διασταυρούμενη Επικύρωση (*cross-validation*): Βεβαιωθείτε ότι έχετε διαχωρίσει τα δεδομένα σας σε σύνολα εκπαίδευσης και ελέγχου, ώστε να χρησιμοποιήσετε 5-fold CV για όλα τα πειράματα. [5]

## A2. Επιλογή αρχιτεκτονικής [55 μονάδες]

Όσον αφορά την τοπολογία των ΤΝΔ για την εκπαίδευση τους με τον Αλγόριθμο Οπισθοδιάδοσης του Σφάλματος (*back-propagation*), θα χρησιμοποιήσετε ΤΝΔ με ένα κρυφό επίπεδο και θα πειραματιστείτε με τον αριθμό των κρυφών κόμβων. Για την εκπαίδευση του δικτύου χρησιμοποιήστε αρχικά ρυθμό μάθησης  $\eta = 0.001$ .

α) Η εκπαίδευση και αξιολόγηση των μοντέλων σας μπορεί να γίνει με χρήση *Cross-Entropy* (CE), καθώς και *Μέσου Τετραγωνικού Σφάλματος* (MSE) για τις τιμές που περιέχονται στα σύνολα εκπαίδευσης και ελέγχου<sup>2</sup>. Να εξηγήσετε με απλά λόγια ποια είναι η σημασία των παραπάνω μετρικών για το συγκεκριμένο πρόβλημα. [5]

β) Πόσες εισόδους θα χρειαστείτε στο ΤΝΔ, δεδομένου ότι μια είσοδος πρέπει να αναπαριστά μια εικόνα 28X28; [5]

γ) Πόσους νευρώνες θα χρειαστείτε στο επίπεδο εξόδου, δεδομένου του ζητούμενου της ταξινόμησης σε διαφορετικές κλάσεις; [5]

δ) Να επιλέξετε κατάλληλη συνάρτηση ενεργοποίησης για τους κρυφούς κόμβους και να τεκμηριώσετε την επιλογή σας. [5]

ε) Ποια συνάρτηση ενεργοποίησης θα χρησιμοποιήσετε για το επίπεδο εξόδου; Σιγμοειδή, Softmax ή κάποια άλλη; [5]

---

<sup>2</sup> Έστω δύο διανύσματα  $\mathbf{p}$  και  $\mathbf{q}$  διάστασης  $K$ , όπου  $K$  ο αριθμός των κλάσεων,  $\mathbf{p}$  επιθυμητό και  $\mathbf{q}$  πραγματικό διάνυσμα εξόδου. Έστω  $\mathbf{x}$  το διάνυσμα εισόδου και  $n$  το πλήθος των δειγμάτων.  $CE = \frac{1}{n} \sum_x \sum_{i=1}^K p_i(\mathbf{x}) \log(q_i(\mathbf{x}))$  και επιστρέφει μόνο θετικές τιμές. Όσες από αυτές είναι μικρές, θα αναφέρονται σε μεγάλη ομοιότητα. Επίσης,  $MSE = \frac{1}{2n} \sum_x \sum_{i=1}^K (p_i(\mathbf{x}) - q_i(\mathbf{x}))^2$ . Το MSE ισοδυναμεί με το τετράγωνο της ευκλείδειας απόστασης των δύο διανυσμάτων, με μικρότερες τιμές να αναφέρονται σε μικρότερο σφάλμα.

στ) Πειραματιστείτε με 3 διαφορετικές τιμές για τον αριθμό των νευρώνων του κρυφού επιπέδου και συμπληρώστε τον παρακάτω πίνακα. Εμπειρικά ενδεδειγμένες τιμές για τον αριθμό των κρυφών κόμβων βρίσκονται στο διάστημα  $[O, I+O]$  ( $I$  αριθμός εισόδων,  $O$  αριθμός εξόδων,  $H$  αριθμός κόμβων στο κρυφό επίπεδο). Να συμπεριλάβετε και τις γραφικές παραστάσεις σύγκλισης (M.O.) ανά κύκλο εκπαίδευσης. Διατυπώστε τα συμπεράσματά σας σχετικά με (i) τον αριθμό των κρυφών κόμβων, (ii) την επιλογή της συνάρτησης κόστους και (iii) την ταχύτητα σύγκλισης ως προς τις εποχές εκπαίδευσης. [15]

Αριθμός νευρώνων στο κρυφό επίπεδο	CE loss	MSE
$H_1 = O$		
$H_1 = (I+O)/2$		
$H_1 = I+O$		

ζ) Με τον βέλτιστο αριθμό κρυφών κόμβων που βρήκατε στο στ) δοκιμάστε να προσθέσετε ένα ακόμα κρυφό επίπεδο  $H_2$  στο δίκτυο. Πειραματιστείτε με τον αριθμό των κόμβων του  $H_2$ . Περιγράψτε μια λογική για τη στοίχιση των κρυφών επιπέδων (είναι καλό να έχουν τον ίδιο αριθμό κόμβων; Μειούμενο; Αυξανόμενο;). Να αναφέρετε CE και MSE και να διατυπώσετε τα συμπεράσματα σχετικά με την προσθήκη κρυφών επιπέδων. [10]

Αριθμός νευρώνων στο κρυφό επίπεδο	CE loss	MSE
$H_2 =$		
$H_2 =$		
$H_2 =$		

η) Κριτήριο τερματισμού. Επιλέξτε και τεκμηριώστε κατάλληλο κριτήριο τερματισμού της εκπαίδευσης κάθε φορά (για κάθε fold). Μπορεί να χρησιμοποιηθεί η τεχνική του πρόωρου σταματήματος (early stopping); [5]

Προσοχή: σε όλα τα πειράματα θα χρησιμοποιήσετε 5-fold cross validation (5-fold CV).

### A3. Μεταβολές στον ρυθμό εκπαίδευσης και σταθεράς ορμής [15 μονάδες]

Επιλέγοντας την τοπολογία που δίνει το καλύτερο αποτέλεσμα βάσει του προηγούμενου ερωτήματος, να πραγματοποιήσετε βελτιστοποίηση των υπερπαραμέτρων ρυθμού εκπαίδευσης  $\eta$  και σταθεράς ορμής  $m$  με χρήση CV και να συμπληρώσετε τον παρακάτω πίνακα. Να συμπεριλάβετε και τις γραφικές παραστάσεις σύγκλισης (M.O.) ως προς τους κύκλους εκπαίδευσης που θα χρειαστούν. Να τεκμηριώσετε θεωρητικά γιατί  $m < 1$ .

$\eta$	$m$	CE loss	MSE
0.001	0.2		
0.001	0.6		
0.05	0.6		
0.1	0.6		

Να διατυπώσετε σύντομα τα συμπεράσματα που προκύπτουν από τα 4 πειράματα.

### A4. Ομαλοποίηση [15 μονάδες]

Μια μέθοδος για την αποφυγή υπερπροσαρμογής του δικτύου και βελτίωση της γενικευτικής του ικανότητας είναι η ομαλοποίηση του διανύσματος των βαρών (regularization). Να εφαρμόσετε  $L2$  ομαλοποίηση (φθορά βαρών) και να επανεκπαιδεύσετε το δίκτυό σας, όπως προέκυψε από το A3, αξιολογώντας διάφορες τιμές για τον συντελεστή φθοράς  $r$ .

i)  $r = 0.1$  ii)  $r = 0.5$  iii)  $r = 0.9$

Συμπληρώστε τον παρακάτω πίνακα για κάθε μία από τις παραπάνω περιπτώσεις με χρήση 5-fold CV. Να συμπεριλάβετε και τις γραφικές παραστάσεις σύγκλισης (M.O.) ανά κύκλο εκπαίδευσης.

Συντελεστής φθοράς	CE loss	MSE
0.1		
0.5		
0.9		

Διατυπώστε τα συμπεράσματά σας σχετικά με την επίδραση της μεθόδου στη γενικευτική ικανότητα του δικτύου.

#### **A5. Convolutional Neural Network.** [προαιρετικό ερώτημα - 10 μονάδες bonus]

Τα Convolutional Neural Networks (CNN) αποτελούν μοντέλα βαθιάς μάθησης που βρίσκουν εφαρμογή σε διάφορους τομείς, ενώ είναι ιδιαίτερα αποδοτικά σε εφαρμογές επεξεργασίας εικόνας. Σας ζητείται να πειραματιστείτε με διαφορετικές αρχιτεκτονικές CNN δικτύων (τουλάχιστον δύο), υλοποιώντας CNN μοντέλα που μετά από εκπαίδευση, να είναι ικανά να αναγνωρίζουν χειρόγραφα ψηφία. Για την υλοποίηση των επιπέδων του CNN μπορείτε να χρησιμοποιήσετε οποιαδήποτε από τις προτεινόμενες βιβλιοθήκες. Περιγράψτε αναλυτικά την αρχιτεκτονική δομή των μοντέλων και τις παραμέτρους εκπαίδευσης. Να αναφέρετε τιμές για CE και MSE, σχολιάστε τα αποτελέσματα και συγκρίνετέ τις υλοποιήσεις μεταξύ τους, αλλά και με την υλοποίηση των προηγούμενων ερωτημάτων.

#### **Παραδοτέα**

Η αναφορά που θα παραδώσετε θα πρέπει να περιέχει εκτενή σχολιασμό των πειραμάτων σας, καθώς και πλήρη καταγραφή των αποτελεσμάτων και των συμπερασμάτων σας, ανά υπό-ερώτημα. Επίσης, πρέπει να συμπεριλάβετε στην αρχή της αναφοράς σας ένα link προς τον κώδικα που έχετε χρησιμοποιήσει (σε κάποια file sharing υπηρεσία ή code repo).

Μην ξεχάσετε να συμπληρώσετε τα στοιχεία σας στην αρχή της 1<sup>ης</sup> σελίδας.

#### **Αξιολόγηση**

Η απάντηση των ερωτημάτων Α και Β έχει βαρύτητα 20% στον τελικό βαθμό του μαθήματος (το σύνολο και των δύο μερών της εργασίας έχει βαρύτητα 40%). Ο βαθμός του Bonus (10%) προστίθεται στο παραπάνω ποσοστό 40%.

#### **Παρατηρήσεις**

1. Η αναφορά, σε ηλεκτρονική μορφή, πρέπει να αναρτηθεί στο e-class μέχρι τη Μ. Δευτέρα, 26/4/2021, στις 23:59.
2. Για οποιαδήποτε διευκρίνηση / ερώτηση μπορείτε να χρησιμοποιείτε το σχετικό forum στο eclass του μαθήματος.