

ΕΠΛ 442: Μηχανική Μάθηση

ΕΡΓΑΣΙΑ 1: Back Propagation (Μέρος 2)

Χαρακτήρες (παράδοση μέχρι 29/10/2020)

Χρησιμοποιήστε το πρόγραμμα που κατασκευάσατε στο Μέρος 1 για να υλοποιήσετε ένα νέο νευρωνικό δίκτυο τύπου Back Propagation το οποίο θα εκπαιδευθεί στην αναγνώριση χειρόγραφων χαρακτήρων. Τα δεδομένα για την εκπαίδευση και έλεγχο του δικτύου περιγράφονται στη παράγραφο «Δεδομένα». Ακολουθήστε τις οδηγίες που δόθηκαν στο Μέρος 1 όσο αφορά τα παραδοτέα.

Δεδομένα (Data)

Το παρεχόμενο αρχείο περιέχει 20,000 χειρόγραφους χαρακτήρες. Η πρώτη στήλη περιέχει το χαρακτήρα και οι υπόλοιπες 16 στήλες περιέχουν το διάνυσμα που αντιστοιχεί στο συγκεκριμένο χαρακτήρα και που θα χρησιμοποιήσετε για είσοδο στο νευρωνικό σας δίκτυο. Οι στήλες διαχωρίζονται με κόμμα. Οι τιμές του διανύσματος κυμαίνονται στο εύρος 0-15.

Παρακαλώ όπως διαβάσετε (τουλάχιστο) το πρώτο μέρος της πιο κάτω δημοσίευσης (σελ. 161-165) PIPIN την υλοποίηση του δικτύου. Κρίνεται ιδιαίτερα σημαντικό για να καταλάβετε τις διάφορες μεθόδους που χρησιμοποιήθηκαν για την προετοιμασία των δεδομένων.

P.W. Frey and D.J. Slate, Letter Recognition Using Holland style Adaptive Classifiers, Machine Learning Vol 6, 161-182, (1991). <http://springerlink.metapress.com/content/j53651412w267015/fulltext.pdf>

Ενδεικτικά δεδομένα: T,2,8,3,5,1,8,13,0,6,6,10,8,0,8,0,8 I,5,12,3,7,2,10,5,5,4,13,3,9,2,8,4,10

Το αρχείο βρίσκεται στο: <http://www.cs.ucy.ac.cy/~cchrist/epl442/letter-recognition.txt>

Έξοδος Προγράμματος

1. Αρχείο (errors.txt) με τρεις στήλες η πρώτη θα είναι ο μετρητής επαναλήψεων, η δεύτερη το λάθος εκμάθησης (training error) στο τέλος κάθε επανάληψης και η τρίτη το λάθος ελέγχου (testing error) στο τέλος κάθε επανάληψης.
2. Αρχείο (successrate.txt) με τρεις στήλες: η πρώτη θα είναι ο μετρητής επαναλήψεων, η δεύτερη το ποσοστό ορθών αποτελεσμάτων κατά την εκμάθηση (training phase) και η τρίτη το ποσοστό ορθών αποτελεσμάτων κατά τον έλεγχο (testing phase).

Παραδοτέα

- Γραπτό κείμενο μέχρι 3 σελίδες (σε PDF). Να περιλαμβάνει τις μεταβλητές και τα αποτελέσματα συγκεκριμένης εκπαίδευσης όπως αυτά παράχθηκαν από το πρόγραμμα σας. Ανάλυση επιλογών σας όσο αφορά το Νευρωνικό Δίκτυο. Ανάλυση αποτελεσμάτων. **Γραφική παράσταση του αρχείου errors.txt και successrate.txt της συγκεκριμένης εκπαίδευσης.**
- Κώδικας με σχόλια, training.txt, test.txt, parameters.txt Αποστολή στη διεύθυνση magath06@cs.ucy.ac.cy μέχρι τις **29/10/2020**. Όλα τα ζητούμενα αρχεία να σταλούν σε μορφή .zip και το όνομα του αρχείου να είναι: **EPL442_Assign2_<Student ID>.zip**
- Εκτέλεση προγράμματος: Το πρόγραμμα σας πρέπει να ψάχνει για το parameters.txt από όπου θα παίρνει όλες τις πληροφορίες για περαιτέρω εκτέλεση.

Σημειώσεις:

Η ακριβής μορφή των δικτύων, οι παραμέτροι εισόδου όπως και διάφορες λεπτομέρειες της υλοποίησης του κώδικα είναι στη δική σας δικαιοδοσία. Η τελική σας βαθμολογία θα εξαρτηθεί τόσο από την ικανότητα του δικτύου σας να ομαδοποιήσει τους χαρακτήρες όσο και από την πληρότητα της αναφοράς σας (όπου θα δικαιολογήσετε τις επιλογές σας) και την ποιότητα του κώδικα σας.