2η Εργασία Τεχνητής Νοημοσύνης

## Περιγραφή

Στην εργασία αυτή υλοποιούνται οι παραλλαγές του αλγορίθμου επιβλεπόμενης διανυσματικής κβάντισης LVQ2 και LFM στο MATLAB. Στην συνέχεια δοκιμάζονται αυτοί οι αλγόριθμοι στα δεδομένα από το Iris και το Wine για την ταξινόμησή τους. Κατά την διάρκεια των δοκιμών, τα δεδομένα χωρίζονται σε 10 ίσα μέρη και δημιουργούνται τρία σύνολα: δεδομένα συνόλου εκπαίδευσης (70%), δεδομένα συνόλου τεκμηρίωσης (20%) και δεδομένα συνόλου ελέγχου (10%). Με αυτόν τον τρόπο και με κυκλική ολίσθηση των τριών αυτών μερών, οι αλγόριθμοι θα εκπαιδεύονται και θα δοκιμάζονται σε διαφορετικά σύνολα δεδομένων ώστε να επιτευχθεί όσο το δυνατόν περισσότερη η γενίκευσή τους και όχι η υπερπροσαρμογή τους.

Οι δοκιμές για τους δύο αλγορίθμους έγιναν παρομοίως με εμφωλευμένες λούπες που εξέταζαν κάθε συνδυασμό από τις τιμές των παραμέτρων που ζητούνται (Ppc=[1, 2, 5], Epoch=[5, 10, 15], Ka=[0.01, 0.1]), όπως επίσης και με κέρδος προσαρμογής που ελαττώνεται με βάση την συνάρτηση α(t) = α0/(1+Καt) για α0 = 0.5. Τα βάρη αρχικοποιούνται με τυχαία επιλογή δεδομένων από το εκάστοτε αντιπροσωπευτικό διάνυσμα που αρχικοποιείται από το σύνολο εκπαίδευσης.

## Υλοποίηση LVQ2

Στον αλγόριθμο LVQ2 για κάθε εποχή εξετάζονται όλα τα διανύσματα από το σύνολο εκπαίδευσης ξεχωριστά. Αρχικά υπολογίζεται η απόσταση του διανύσματος από τα βάρη, μετά εντοπίζονται τα δύο πιο κοντινά βάρη και εξετάζεται εάν το κοντινότερο ανήκει σε άλλη κλάση και το επόμενο κοντινότερο ανήκει στην ίδια κλάση. Τέλος αν και τα δύο βάρη βρίσκονται μέσα στα όρια που ορίζονται από τις παρακάτω σχέσεις (LVQ2.1):

,

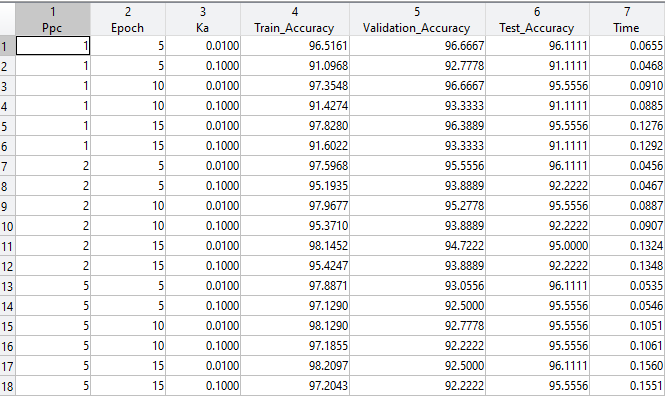
Tο κοντινότερο βάρος απομακρύνεται και το επόμενο κοντινότερο πλησιάζει με βάση του εξής τύπους:

W(:,id1)=W(:,id1) - at \* ( X(:,j) - W(:,id1))

W(:,id2)=W(:,id2) + at \* ( X(:,j) - W(:,id2))

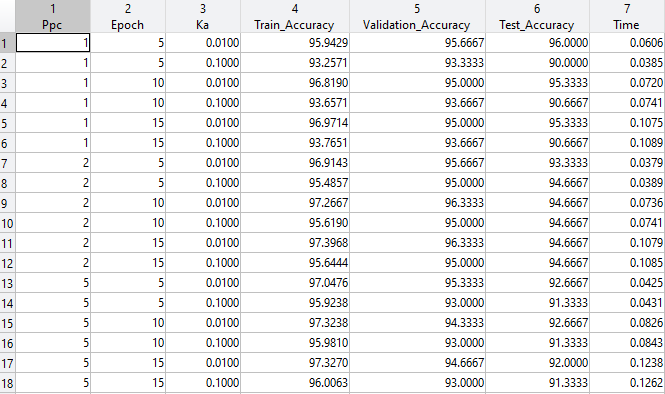
Μόλις τελειώσει η εκπαίδευση για όλο το σύνολο δεδομένων εκπαίδευσης υπολογίζεται το ποσοστό ακρίβειας του αλγορίθμου για εκείνη την εποχή. Μετά την ολοκλήρωση των εποχών που ζητήθηκαν υπολογίζεται το ποσοστό ακρίβειας για το σύνολο τεκμηρίωσης και το σύνολο ελέγχου.

Για το wine dataset τα αποτελέσματα για κάθε δοκιμή παρουσιάζονται στην παρακάτω εικόνα μαζί με τον χρόνο εκτέλεσης, με ποσοστά από 91-98% για την ακρίβεια εκπαίδευσης και 91-96% για το σύνολο ελέγχου.



Εικόνα 1 Αποτελέσματα LVQ2 για το Wine Dataset

Στα δεδομένα του Iris Dataset τα αποτελέσματα για κάθε δοκιμή βρίσκονται στο 93-97% για το σύνολο εκπαίδευσης και 90-96% για το σύνολο ελέγχου με τον αλγόριθμο του LVQ2:

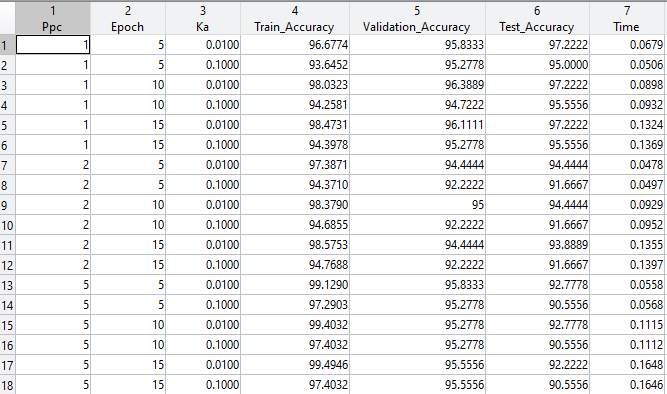


Εικόνα 2 Αποτελέσματα LVQ2 για το Iris Dataset

## Υλοποίηση LFM

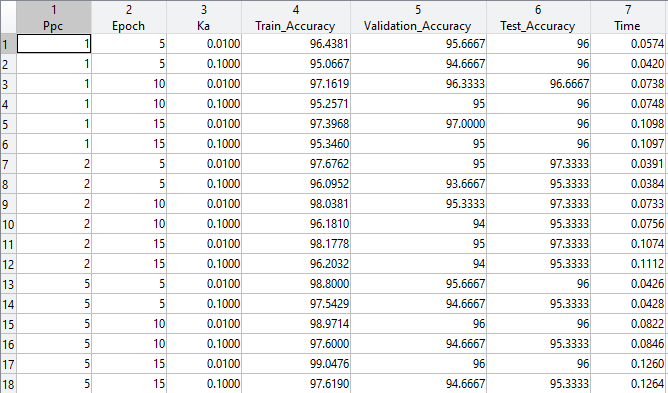
Ο αλγόριθμος LFM ακολούθησε την ίδια δομή με τον LVQ2 προηγουμένως, με την διαφορά ότι στις συνθήκες για την προσαρμογή των βαρών πλέον δεν εξεταζόταν το όριο το οποίο θα έπρεπε τα βάρη να βρίσκονται μέσα, εξεταζόταν μόνο αν το κοντινότερο βάρος είναι άλλης κλάσης και απομακρυνόταν και έψαχνε το βάρος που είναι πιο κοντά και ανήκε στην ίδια κλάση ώστε να το φέρει πιο κοντά με τους ίδιους υπολογισμούς.

Για το Wine Dataset τα αποτελέσματα για κάθε δοκιμή παρουσιάζονται στην παρακάτω εικόνα μαζί με τον χρόνο εκτέλεσής τους με ποσοστά από 93-99% για το σύνολο εκπαίδευσης και 90-97% για το σύνολο ελέγχου.



Εικόνα 3 Αποτελέσματα LFM για το Wine Dataset

Και για το Iris Dataset τα αποτελέσματα των πειραμάτων με τον LFM αλγόριθμο εμφανίζουν ποσοστά ακρίβειας εκπαίδευσης 95-99% και ελέγχου 95-97%.



Εικόνα 4 Αποτελέσματα LFM για το Iris Dataset

## Σύγκριση αποτελεσμάτων

Όπως φαίνεται και στις παραπάνω εικόνες από τα αποτελέσματα των δύο αλγορίθμων LVQ2 και LFM, οι διαφορές στα ποσοστά που πετυχαίνουν δεν είναι μεγάλες. Στα δεδομένα του wine οι δύο αλγόριθμοι πετυχαίνουν ποσοστό εκπαίδευσης 91-99% και 93-99% αντίστοιχα και ποσοστά ελέγχου 91-96% και 90-97%. Για τα δεδομένα του Iris τα ποσοστά των δύο αλγορίθμων είναι 93-97% για το σύνολο εκπαίδευσης και 90-96% για το σύνολο ελέγχου για τον LVQ2, και 95-99% εκπαίδευσης και 95-97% ελέγχου για τον LFM.

Στις επιμέρους δοκιμές με αλλαγές στις παραμέτρους Ka, του αριθμού των νευρωνίων (ppc) και του αριθμού των εποχών της εκπαίδευσης, παρατηρείται ότι από πολύ νωρίς και με λίγα βάρη και εποχές τα ποσοστά είναι αρκετά ικανοποιητικά και μάλιστα πολύ κοντά στον μέσο όρο που επιτυγχάνεται στις δοκιμές αυτές.