

Νευρωνικά Δίκτυα – Βαθιά Μάθηση

1^η Ενδιάμεση Εργασία

Ονοματεπώνυμο: Κωνσταντίνος Πέρρος

AEM: 3713

Εντός του αρχείου κώδικα βρίσκονται και αντίστοιχα σχόλια.

Αρχικά παίρνουμε δεδομένα από την βάση δεδομένων MNIST από το keras (βιβλιοθήκη ανοιχτού κώδικα με πολλά datasets) και τα αποθηκεύουμε σε αντίστοιχα train και test data. Τα train δεδομένα είναι 60000 28*28 εικόνες από αριθμούς 0 έως 9, ενώ τα test δεδομένα είναι 10000 αντίστοιχες εικόνες.

Χρησιμοποιώντας την scikit-learn βιβλιοθήκη εφαρμόζουμε κατηγοριοποίηση πλησιέστερου γείτονα (KNN) με 1 γείτονα αρχικά και με 3 αργότερα. Έπειτα εφαρμόζουμε κατηγοριοποίηση πλησιέστερου κέντρου (Nearest Centroid) και συγκρίνουμε τα αποτελέσματα. Για να γίνουν οι κατηγοριοποιήσεις αυτές κάνουμε αρχικά reshape τα δεδομένα από τρισδιάστατους πίνακες σε δισδιάστατους, μετατρέποντας τα δισδιάστατα δεδομένα 28*28 κελιών σε μονοδιάστατους πίνακες 784 κελιών.

Για περαιτέρω σύγκριση κάνουμε και KNN κατηγοριοποίηση σε 5 και 7 γείτονες αντίστοιχα.

KNN (n=1)	KNN (n=3)	Nearest Centroid	KNN (n=5)	KNN (n=7)
0.9691	0.9705	0.8203	0.9688	0.9694

Όπως παρατηρούμε στα αποτελέσματα, η κατηγοριοποίηση με 3 γείτονες είναι καλύτερη της κατηγοριοποίησης με 1 γείτονα, που είναι γενικώς πολύ πιο ακριβής της κατηγοριοποίησης πλησιέστερου κέντρου. Επίσης, οι πειραματικοί κατηγοριοποιητές με 5 και 7 γείτονες αντίστοιχα έχουν επίσης πολύ καλά αποτελέσματα αλλά χειρότερα του KNN με n=3.

Μπορούμε να αυξήσουμε τις διαστάσεις των πινάκων για να βελτιώσουμε το accuracy τους καθώς και να συμπίεσουμε τα δεδομένα χρησιμοποιώντας τον αλγόριθμο PCA (Principal Component Analysis) από την ίδια βιβλιοθήκη, για μια καλύτερη σύγκριση των δεδομένων. Προφανώς με χαμηλές διαστάσεις το accuracy παραμένει ίδιο ή χειροτερεύει, ενώ ανεβάζοντας τις διαστάσεις φτάνουμε σε σημείο όπου δεν μπορεί να υπάρξει περαιτέρω βελτίωση στο accuracy. Για καλύτερα αποτελέσματα χωρίς όμως περαιτέρω πλεονασμό θα αυξήσουμε τις διαστάσεις σε 50.

KNN (n=1)	KNN (n=3)	Nearest Centroid	KNN (n=5)	KNN (n=7)
1.0	0.9807	0.8213	0.9751	0.9701

Παρατηρούμε σημαντική βελτίωση σε κάθε κατηγοριοποιητή, με μεγάλες αλλαγές στην προηγούμενη κατάταξη, καθώς και τέλεια αποτελέσματα από τον KNN με n=1, καθιστώντας την εφαρμογή του PCA σημαντική για καλύτερη σύγκριση αποτελεσμάτων.