

# Αλγόριθμος

## k-Πλησιέστεροι Γείτονες

Λίγα λόγια σχετικά με τον αλγόριθμο KNN .

Ο αλγόριθμος βασίζεται στη χρήση μέτρων βασισμένων στην απόσταση . Η τεχνική του KNN προ υποθέτει ότι το σύνολο εκπαίδευσης δεν περιλαμβάνει μόνο τα δεδομένα αλλά επίσης και την επιθυμητή κατηγοριοποίηση για κάθε στοιχείο. Αυτό έχει σαν αποτέλεσμα τα δεδομένα εκπαίδευσης να αποτελούν το μοντέλο. Όταν πρόκειται να γίνει μια κατηγοριοποίηση για ένα καινούργιο στοιχείο , πρέπει να καθοριστεί η απόσταση του από κάθε στοιχείο του συνόλου εκπαίδευσης. Μόνο οι  $K$  πλησιέστερες εκχωρήσεις στο σύνολο εκπαίδευσης λαμβάνονται υπόψη στη συνέχεια . Το νέο στοιχείο τοποθετείται μετά στη κατηγορία που περιέχει τα περισσότερα στοιχεία από το σύνολο των  $K$  κοντινότερων στοιχείων .

### Η προσέγγιση KNN

- Το σύνολο δεδομένων εκπαίδευσης περιλαμβάνει τις κλάσεις.
- Για να αναθέσουμε ένα νέο στοιχείο σε μια κλάση εξετάζουμε τα  $K$  πλησιέστερα σ' αυτό σημεία.
- Τοποθετούμε το νέο στοιχείο στην κλάση που έχει την πλειοψηφία μέσα στα κοντινά στοιχεία.
- Πολυπλοκότητα  $O(q)$  για κάθε νέο στοιχείο ( $q$  είναι το μέγεθος του συνόλου δεδομένων εκπαίδευσης).

### Πλεονεκτήματα και μειονεκτήματα αλγορίθμου .

Οι κατηγοριοποιητές k-NN διαθέτουν αξιόλογα πλεονεκτήματα:

- Είναι αποτελεσματικοί όταν υπάρχουν σύνθετες εξαρτήσεις μεταξύ των μεταβλητών.
- Διαθέτουν απλό αλγόριθμο.
- Σε πολλές περιπτώσεις επέτυχαν υψηλές επιδόσεις κατηγοριοποίησης.

Ορισμένα από τα βασικά μειονεκτήματα τους είναι τα ακόλουθα:

- Το γεγονός ότι γίνονται πολλές συγκρίσεις μεταξύ παρατηρήσεων απαιτεί πολύ αποτελεσματικές τεχνικές καταλογοποίησης (indexing).
- Η κατηγοριοποίηση νέων παρατηρήσεων διαρκεί πολύ περισσότερο χρόνο, ειδικά στις περιπτώσεις όπου ο αριθμός των εν δυνάμει «γειτόνων» είναι μεγάλος.
- Τα αποτελέσματα τους μπορούν να επηρεαστούν σε σημαντικό βαθμό από το πλήθος των γειτόνων  $k$ .
- Είναι ευαίσθητοι σε τοπικά χαρακτηριστικά των δεδομένων.
- Είναι ευαίσθητοι στην ύπαρξη μη σημαντικών μεταβλητών εισόδου.