

ΑΛΓΟΡΙΘΜΟΙ ΕΚΜΑΘΗΣΗΣ: Bayes, Κανόνες Συσχέτισης, Αδρανής Εκμάθηση & Ομαδοποίηση

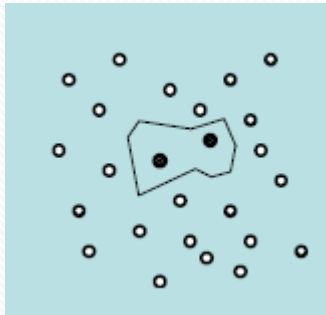
- **Απλοϊκός Bayes για ταξινόμηση:** Εύρεση πιθανότητας της κλάσης δεδομένου του υποδείγματος
 - Αριθμητικά χαρακτηριστικά $\sim N(\mu, \sigma^2)$
 - **Πρόβλημα:** Προσθήκη πολλών περιττών χαρακτηριστικών
 - @ weka: classify->classifier->bayes
 - Αλγόριθμοι: NaiveBayes, HNB, BayesNet κ.α
- **Κανόνες Συσχέτισης (association rules)**
 - Προβλέπουν οποιοδήποτε χαρακτηριστικό και όχι μόνο την τάξη
 - Περιορισμός στον αριθμό συσχετίσεων: **support & confidence**
 - Support: Αριθμός υποδειγμάτων που υποστηρίζει ο κανόνας
 - Confidence: Πόσες από τις προβλέψεις ήταν επιτυχείς
 - @weka->associate->assosiator
 - Αλγόριθμοι: Apriori, PredictiveApriori

- **Απεικόνιση με βάση υποδείγματα**


- **Βασική ιδέα:** Τα υποδείγματα εκπαίδευσης ερευνώνται για την εύρεση του περισσότερο όμοιου προς το νέο παράδειγμα υποδείματος
- **Μέθοδοι:** Πλησιέστερος γείτονας (NN), κ-πλησιέστεροι γείτονες (KNN)
- Για να ορίσουμε τον γείτονα, ορίζουμε την απόσταση μεταξύ των υποδειγμάτων.
 - Αριθμητικά Χαρακτηριστικά: Ευκλείδεια Νόρμα
 - Ονομαστικά Χαρακτηριστικά: 0 για ίδιες τιμές, 1 για διαφορετικές

ΕΚΜΑΘΗΣΗ ΠΡΟΤΟΤΥΠΩΝ

Εντοπίζω τα υποδείγματα που βρίσκονται κοντά στο όριο που αλλάζει το label και ελέγχω μόνο αυτά



Αλγόριθμοι: Classify-> Classifier-> Lazy-> IB1, Ibk, Kstar

- 
- **Ομαδοποίηση (Clustering)**
 - Έχουμε μάθηση χωρίς επίβλεψη (unsupervised learning), δηλαδή δεν γνωρίζουμε την τιμή στόχο
 - Προσπαθούμε μόνοι μας να κάνουμε την ομαδοποίηση των δεδομένων
 - **Ωφέλεια τάξης(category utility, CU):** εκφράζει τη συνολική ποιότητα του διαχωρισμού των υποδειγμάτων σε ομάδες
 - Αλγόριθμοι ομαδοποίησης: EM Cluster, Coweb/Classit
 - @ weka-> cluster->clusterer