CS 383/613 – Machine Learning

K-Nearest Neighbors

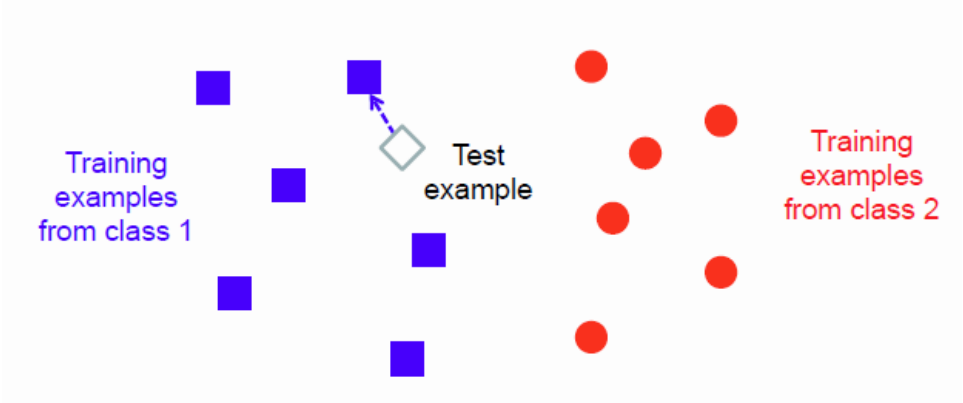
Slides adapted from material created by E. Alpaydin
Prof. Mordohai, Prof. Greenstadt, Pattern Classification (2nd Ed.),
Pattern Recognition and Machine Learning

Objectives

- K-Nearest Neighbors

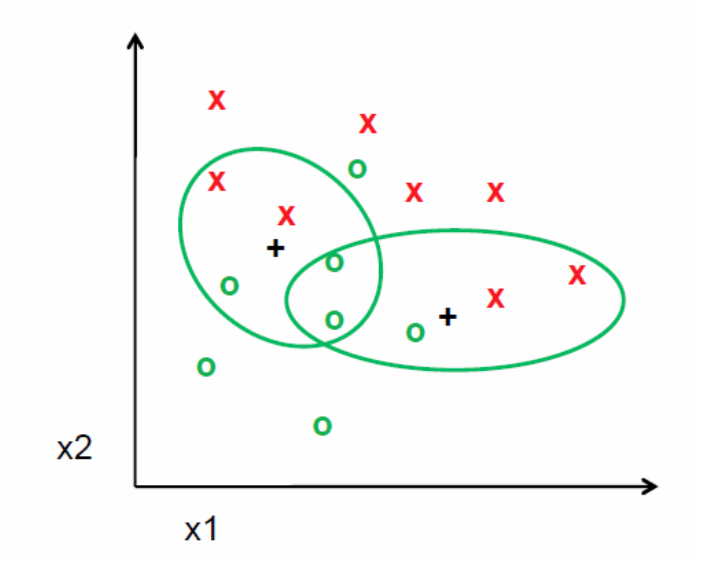
Nearest Neighbor Classifier

- Idea: Assign label to x according to the label of the training example nearest $x' \in trainingset$
- Simple Algorithm
  - All we need is distance/similarity function
  - No training required!



k-Nearest Neighbors

- Just using a single nearest neighbor is susceptible to noise.
- So maybe use k-nearest neighbors
  - And choose class that gets the most votes
- Example: 5-nearest neighbors



Other thoughts

- Intuitively we'd assign the label of the mode of the neighbors' labels.
  - What should we do if k is even?
- How can we decide on k?
  - How can we use a validation set to do this?
- Should we standardize our features?