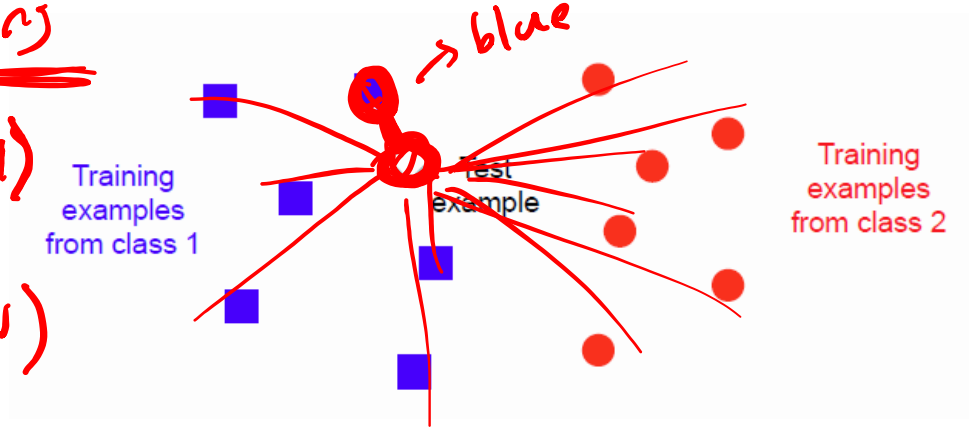
K-Nearest neighbors (KNNs)

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!



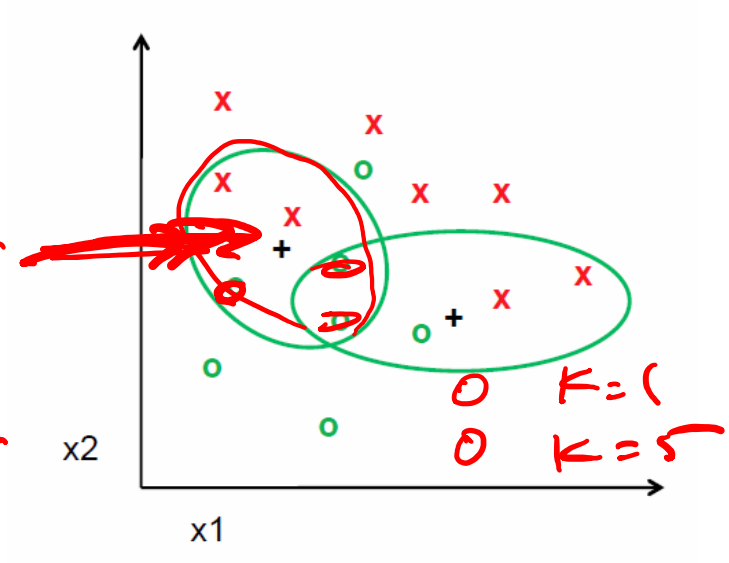
k-Nearest Neighbors

hyper parameter

- 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

5-KNN

1=k



3 O's
2 X's

K=5

K=1

K=5

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?
- How can we use a threshold on the number of votes to create a precision-recall graph?

Example

- Below is a PR-Graph made by varying $0 \leq t \leq 5$ where t is the number of votes need to be labeled as *"positive"*