Pattern Recognition
Non parametric approaches for real classifiers
K-NN

Francesco Tortorella

University of Cassino and Southern Latium

Cassino, Italy

Parametric or non parametric?

- In the parametric approach, we assume that the forms of the underlying density functions were known.
- Unfortunately, in most pattern recognition applications this assumption is weak: the common parametric forms rarely fit the densities actually encountered in practice.
- In the ***non parametric*** approaches, the classifier is built without the assumption that the forms of the underlying densities are known.
- In particular, we consider the ***Nearest Neighbor*** approach that allows us to directly estimating the a posteriori probabilities $P(\omega_j|x)$ and to go directly to decision functions.

Nearest Neighbor Approach

- Basic technique for estimating the probability functions of the problem.
- Simple to build.
- Suppose we have a set Ts of n *labeled* samples belonging to the different classes.
- Let n_i the number of samples of the class ω_i with

$$n = \sum_{i=1}^{n} n_i$$

- We want to classify a sample x $\notin$Ts.

K-Nearest Neighbor approach

- Choose in Ts the k samples nearest to x.
- Let $k_i \leq k$ the number of samples out of k belonging to class ω_i.
- From these values we can estimate
  - The local value of the likelihood

$$p(x|\omega_i) = \frac{k_i}{n_i}$$

  - The local value of the unconditional density

$$p(x) = \frac{k}{n}$$

K-NN classifier

- We can also estimate the priors as

$$P\left(\omega_i\right)=\frac{n_i}{n}$$

- Putting it all together, we can obtain an estimate of the post probabilities:

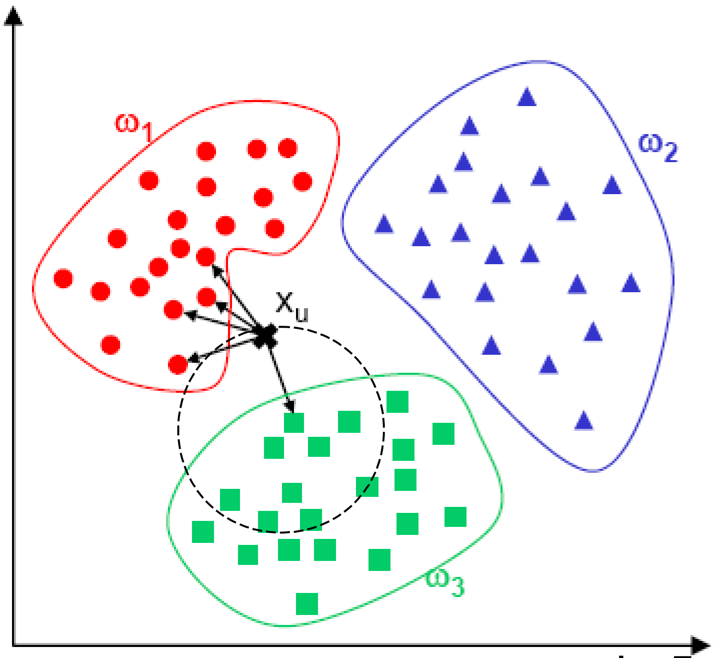
$$P\left(\omega_i | x\right)=\frac{p\left(x | \omega_i\right) P\left(\omega_i\right)}{p\left(x\right)} \simeq \frac{k_i}{n_i} \frac{n_i}{n} \frac{n}{k}=\frac{k_i}{k}$$

- k *Nearest Neighbor (k-NN) rule*

$$\alpha\left(x\right)=\underset{j=1,\ldots,C}{\arg\max} \frac{k_j}{k}$$

K-NN classifier

- In summary, the k-NN classifier assigns the sample x to the class most frequent near the sample x in the set Ts.
- Very simple to build.
- Ingredients:
  - k
  - Ts
  - A distance



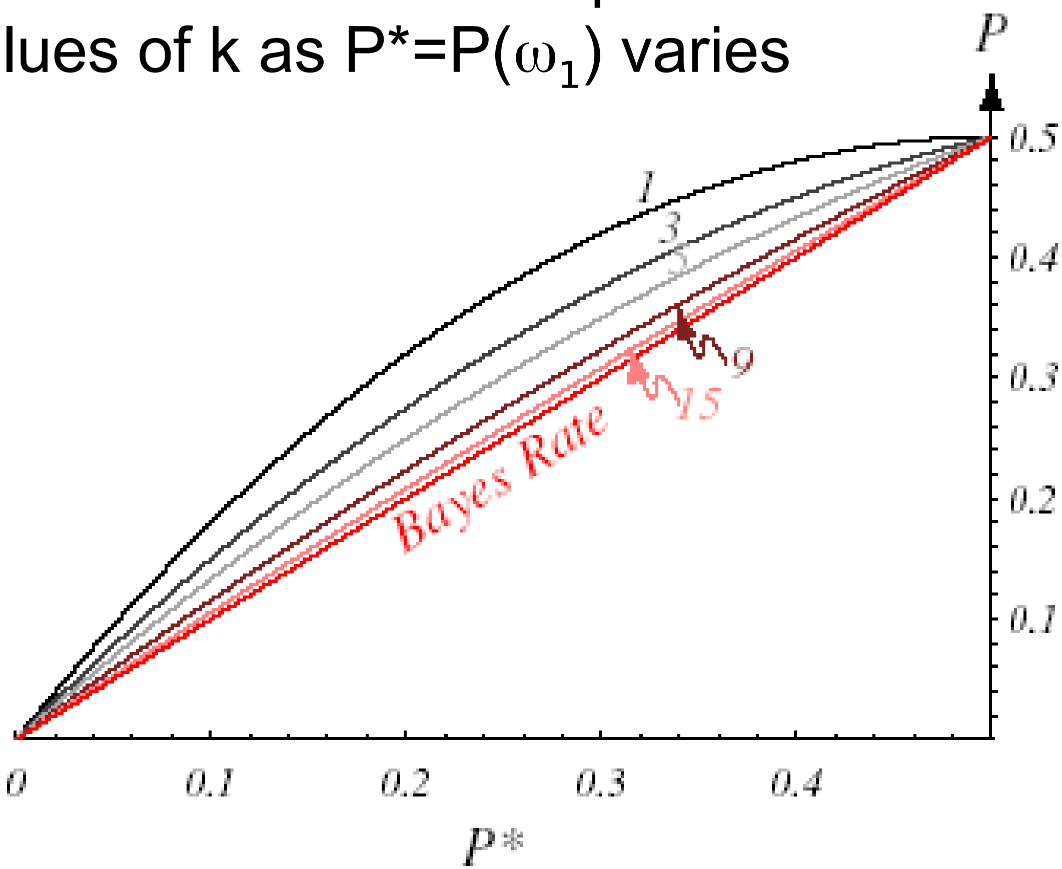
k=5

k-NN classifier performance

- K-NN is suboptimal: it does not guarantee the minimum probability of error provided by the bayesian classifier.
- It can be shown that, if $n \rightarrow \infty$, the probability of error of the k-NN approaches the optimal value for $k \rightarrow \infty$.

k-NN classifier performance

Probability of error for a 2 class problem for different values of k as $P^*=P(\omega_1)$ varies



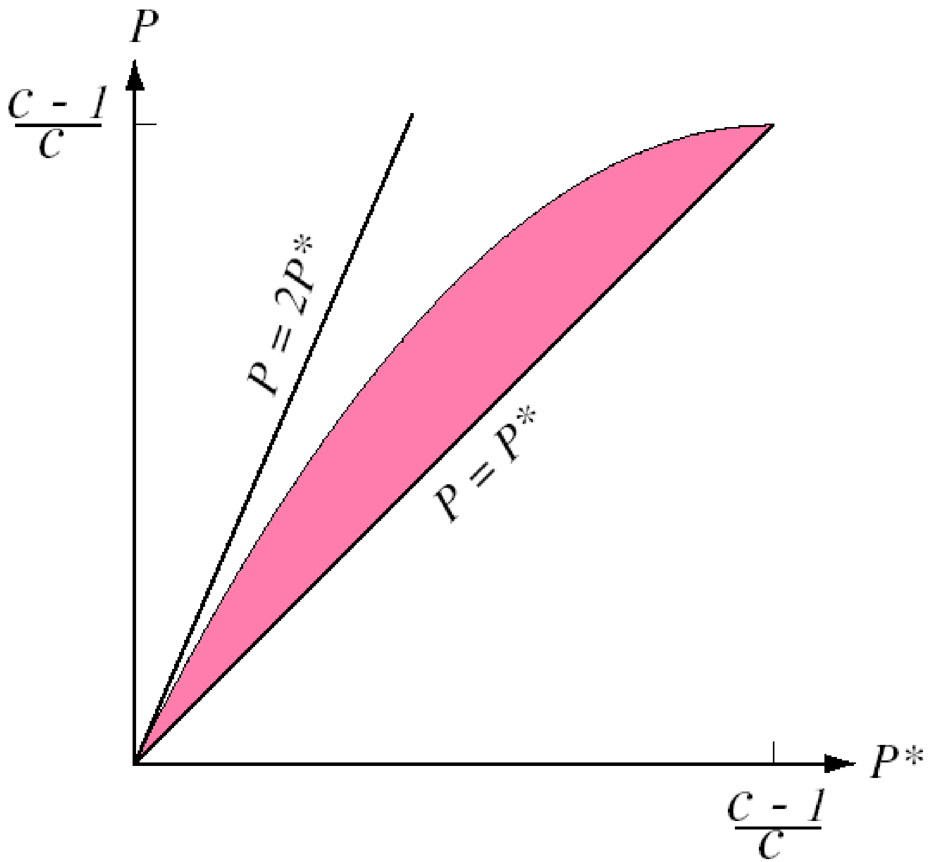
NN classifier

- A particular case is for k=1.
- *1-NN classifier* or *Nearest Neighbor classifier*
- The class chosen is the label of the sample in Ts nearest to x.
- Suboptimal classifier but:

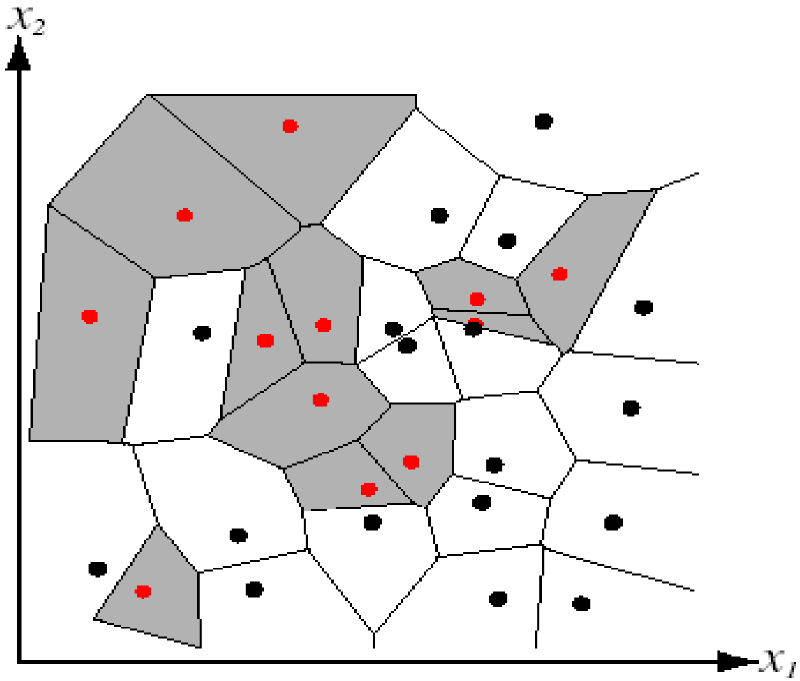
$$P_e^* \leq P_e \leq 2P_e$$

where P_e^* is the probability of error of the Bayesian classifier.

NN classifier performance



K-NN Pros and Cons



- Pro:
  - very simple to build
- Cons:
  - decision regions can be very complicated
  - affected by outliers
  - not efficient