K-Nearest Neighbor Classifier

A non-parametric density estimation based classifier

Introduction

- K-Nearest Neighbor Classifier (k-NNC) is a simple classifier.
- It does not have a design phase, except for the value of k and the distance measure.
- Many people think that this is a weak classifier
  - That is, not a good one.
- In contrary, this is a strong classifier with well established asymptotic bounds.

- Given x to be classified –
- Let the set of classes be $\Omega = \{\omega_1, \dots, \omega_c\}$.
- We find $P(\omega_i|x)$ for all classes, i.e., $i = 1,2, \dots, c$.
- Note that $P(\omega_i|x)$ is the Posterior Probability that x belong to the class ω_i
- The classifier's decision is the class-label ω_{max} where $P(\omega_{max}|x) \geq P(\omega_i|x)$ for all i.

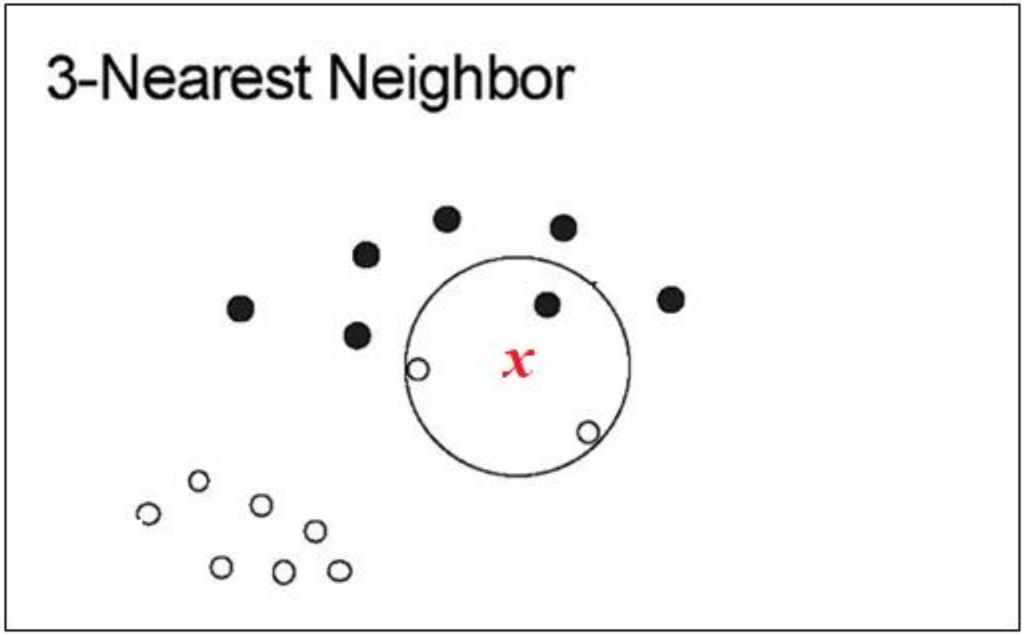
Posterior Probability

- We first find class-conditional densities, in order to find the Posterior Probability.
- To estimate the class-conditional $p(x|\omega_i)$ we do:

Class-conditional

- To estimate the class-conditional $p(x|\omega_i)$ we do:
  - For the given positive integer k , draw a hyper-sphere at x such that exactly k points are within the sphere.
  - count the number of points from class ω_i falling in the sphere. Let this be k_i
  - Let the number of training examples in class ω_i be n_i
  - Let the total number of training examples be n

- Then $p(x|\omega_i) = \frac{k_i}{n_i \cdot V}$
- Here, V is the volume of the hyper-sphere.



In the example,

$$p(x|Black) = \frac{1}{7 \cdot V}$$

$$p(x|White) = \frac{2}{9 \cdot V}$$

Apriori Probabilities

- We do: $P(\omega_i) = \frac{n_i}{n}$
- So in the example $P(Black) = \frac{7}{16}$, and $P(White) = \frac{9}{16}$.

Posterior Probability

- $P(\omega_i|x) = \dfrac{p(x|\omega_i)P(\omega_i)}{p(x)} = \dfrac{\frac{k_i}{n_i V} \cdot \frac{n_i}{n}}{p(x)}$

  $= \dfrac{k_i}{Vnp(x)}$

- Decision : $argmax_{\omega}\{P(\omega_1|x), \ldots, P(\omega_c|x)\}$
- Decision : $argmax_{\omega}\{\dfrac{k_1}{Vnp(x)}, \ldots, \dfrac{k_c}{Vnp(x)}\}$

- Decision : $argmax_{\omega}\{\frac{k_1}{Vnp(x)}, \dots, \frac{k_c}{Vnp(x)}\}$
- Decision : $argmax_{\omega}\{k_1, \dots, k_c\}$

k-NNC is an approximation of Bayes classifier!

Mathematically it can be shown that when $n \to \infty$, $k \to \infty$ and $k/n \to 0$, k-NNC is exactly the Bayes classifier.

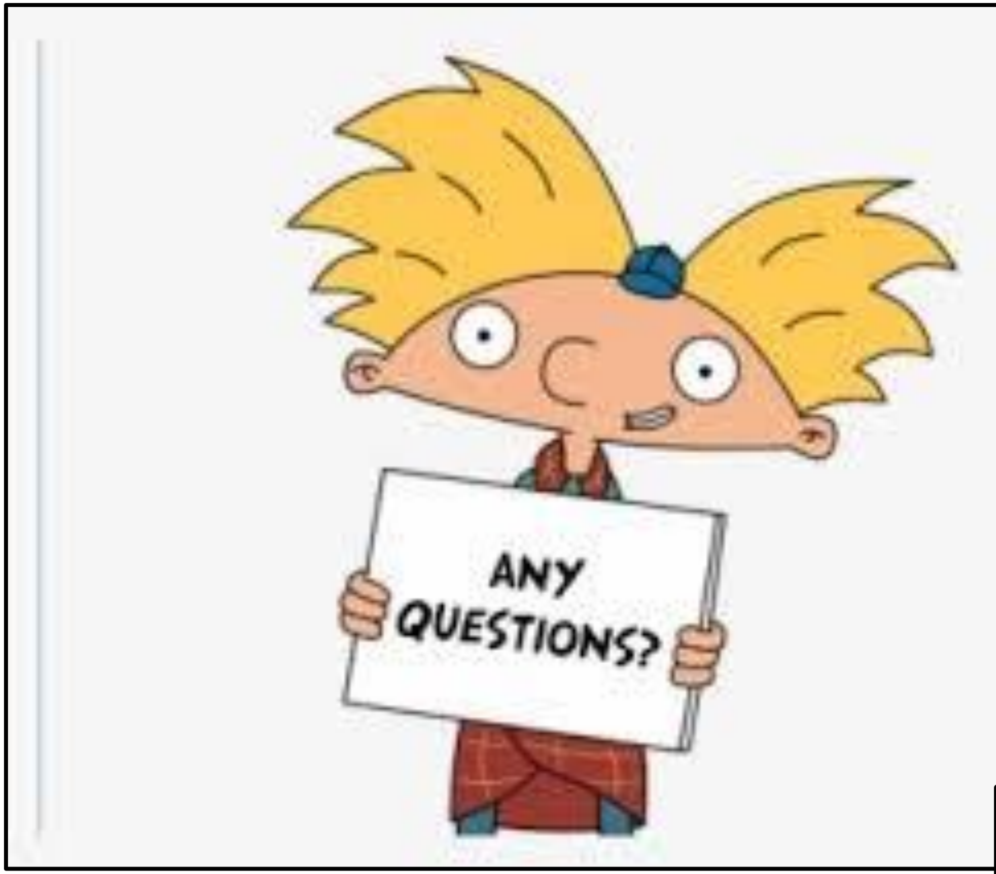
When $k = 1$, k-NNC is simply called the *nearest neighbor classifier (NNC)*.

- Asymptotically, it can be shown that the error of the NNC is less than twice the error of the Bayes Classifier.

- Cross-validation can be used to find the value of k.

Problems with k-NNC

- Test time.
  - Scanning of the entire training set is required.
- Test space
  - Entire training set has to be stored.
- Both time and space complexities of k-NNC (assuming that the k value is a small constant): $O(n)$.
- This is quiet heavy. For Neural networks this is $O(1)$.


ANY
QUESTIONS?