ANLY-601

Advanced Pattern Recognition

Spring 2018

L17 --- Nonparametric Classifiers

GEORGETOWN UNIVERSITY

Non-Parametric Classifiers

Kernel or Parzen Classifiers

The obvious approach is to use a kernel density estimate for each of the class-conditional densities and proceed with a likelihood ratio test

$$\frac{\hat{p}(x\,|\,\omega_1)}{\hat{p}(x\,|\,\omega_2)} = \frac{\frac{1}{N_1}\sum_{i=1}^{N_1} \kappa_1(x - x_j^{(1)})}{\frac{1}{N_2}\sum_{i=1}^{N_2} \kappa_2(x - x_j^{(2)})} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} t$$

Error estimation – use leave-one-out (see x-validation) to set kernel width. Leave-one-out is **cheap** for kernel estimators:

$$\hat{p}_L(x_k) = \frac{1}{N-1}\left[\left(\sum_{j=1}^{N} \kappa(x_k - x_i)\right) - \kappa(0)\right]$$

KNN Classifiers

Two paradigms:

- KNN density estimates followed by likelihood ratio test

- Voting KNN. This is what people usually mean when they refer to a KNN classifier.

For each test sample, select K nearest neighbors from the training data (all classes). The sample is classified according to the majority class of the k neighbors.

Asymptotics for Large N
Error Rates for Voting KNN

K=1 Classify the point *x* as the class of its nearest neighbor

x •

• $x_{NN} \in \omega_i \;\Rightarrow\; x \in \omega_i$

The conditional risk is (classification error rate given x and x_{NN})

$$r_1 = \text{Prob}\left[x \in \omega_1 \text{ and } x_{NN} \in \omega_2 \mid x,\ x_{NN}\right] + \text{Prob}\left[x \in \omega_2 \text{ and } x_{NN} \in \omega_1 \mid x,\ x_{NN}\right]$$
$$= P(\omega_1 \mid x)\, P(\omega_2 \mid x_{NN}) + P(\omega_2 \mid x)\, P(\omega_1 \mid x_{NN})$$

In the limit $N \to \infty$ the data points become closer and closer together, and we can replace $P(\omega_i \mid x_{NN})$ with $P(\omega_i \mid x)$ leaving

$$r_1^* = \lim_{N\to\infty} r_1 = 2\, p(\omega_1 | x)\ p(\omega_2 | x)$$

GEORGETOWN UNIVERSITY

For Odd Number of Neighbors in General

Asymptotically, and for odd number of neighbors, the error rate at x becomes

$$r_{2k-1}^{*} = \lim_{N\to\infty} r_{2k-1} = \sum_{i=1}^{k} \tfrac{1}{i} \binom{2i-2}{i-1} \left[p(\omega_1 | x)\ p(\omega_2 | x)\right]^{i} + \tfrac{1}{2} \binom{2k}{k} \left[p(\omega_1 | x)\ p(\omega_2 | x)\right]^{k}$$

$$r_1^{*} = 2\ p(\omega_1 \mid x)\ p(\omega_2 \mid x)$$

$$r_3^{*} = p(\omega_1 \mid x)\ p(\omega_2 \mid x) + 4\left(p(\omega_1 \mid x)\ p(\omega_2 \mid x)\right)^2$$

$$r_5^{*} = p(\omega_1 \mid x)\ p(\omega_2 \mid x) + \left(p(\omega_1 \mid x)\ p(\omega_2 \mid x)\right)^2 + 12\left(p(\omega_1 \mid x)\ p(\omega_2 \mid x)\right)^3$$

$$r_7^{*} = p(\omega_1 \mid x)\ p(\omega_2 \mid x) + \left(p(\omega_1 \mid x)\ p(\omega_2 \mid x)\right)^2 + 2\left(p(\omega_1 \mid x)\ p(\omega_2 \mid x)\right)^3 + 40\left(p(\omega_1 \mid x)\ p(\omega_2 \mid x)\right)^4$$

Bayes vs KNN Error

Recall that the **Bayes risk** at *x* is

$$r_B = \min\left(p(\omega_1|x),\ p(\omega_2|x)\right) = \tfrac{1}{2}\left(1 - \sqrt{1 - 4\,p(\omega_1|x)\,p(\omega_2|x)}\right)$$

or by Taylor series expansion

$$r_B = \sum_{i=1}^{\infty} \tfrac{1}{i}\binom{2i-2}{i-1}\left(p(\omega_1|x)\,p(\omega_2|x)\right)^i$$

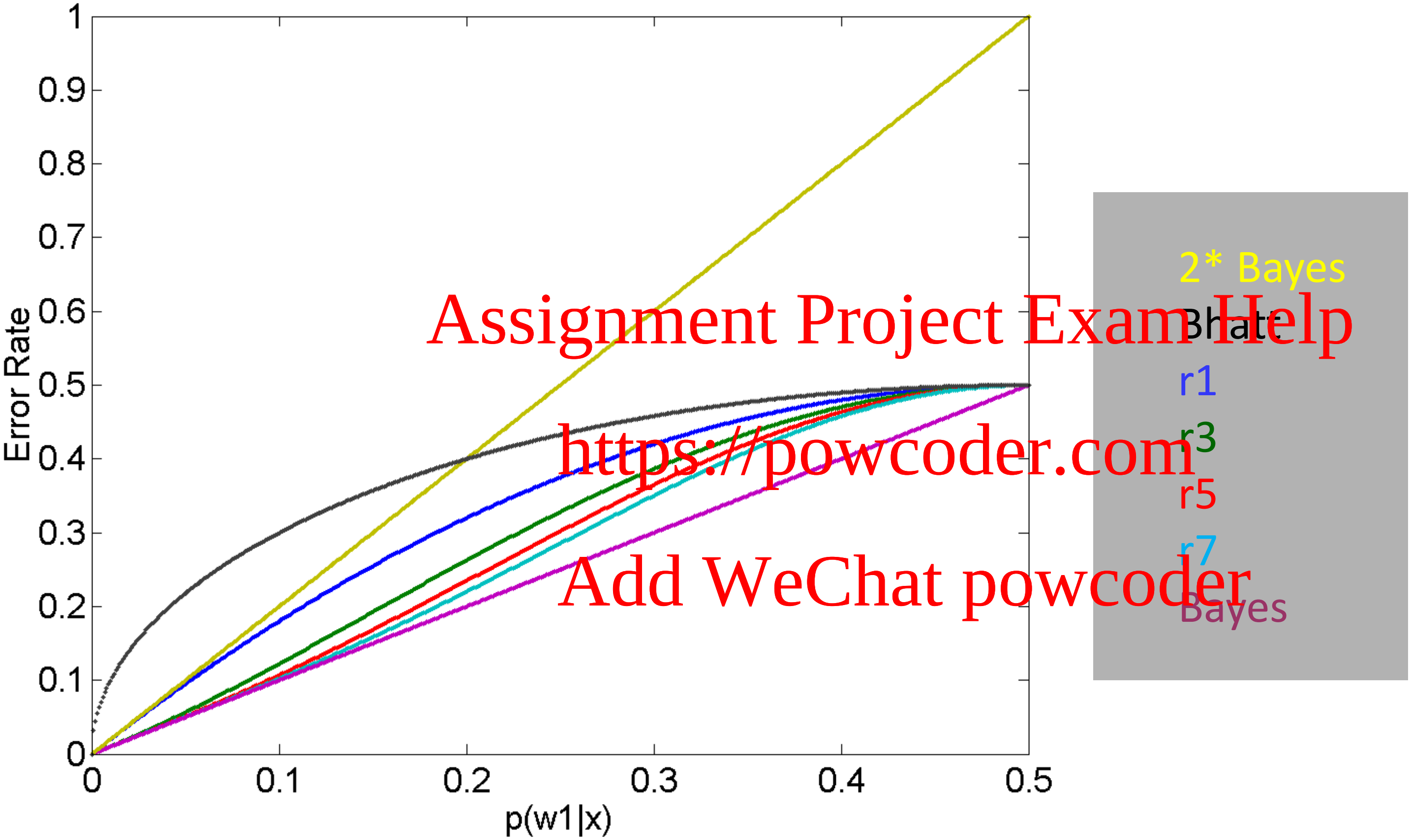
and the error rate for (odd) KNN for infinitely many training points

$$r^*_{2k-1} = \sum_{i=1}^{k} \tfrac{1}{i}\binom{2i-2}{i-1}\left[p(\omega_1|x)\,p(\omega_2|x)\right]^i + \tfrac{1}{2}\binom{2k}{k}\left[p(\omega_1|x)\ p(\omega_2|x)\right]^k$$

One can use these expressions to find the famous bound

$$\boxed{r_B \le \ldots \le r_5^* \le r_3^* \le r_1^* \le 2\,r_B}$$

Bayes vs KNN Error



The KNN bound is tighter than the Bhattacharya bound $\sqrt{p(\omega_1 \mid x)\ p(\omega_2 \mid x)}$

Error Rate for Even Number of Neighbors

With <u>even</u> number of neighbors, we can have a tie (for two classes). When there's a tie, we don't make a choice. This leads to a slightly lower error rate

$$r_{2k}^{*} = \sum_{i=1}^{k} \frac{1}{i} \binom{2i-2}{i-1} \left[p(\omega_1 | x)\, p(\omega_2 | x) \right]^i$$

and a sequence of bounds

$$\tfrac{1}{2} r_B \le r_2^{*} \le r_4^{*} \le r_6^{*} \le \ldots \le \quad r_B \quad \le \ldots \le r_5^{*} \le r_3^{*} \le r_1^{*} \le 2\, r_B$$

GEORGETOWN UNIVERSITY