Machine Learning Theorems

Aimee Barciauskas

29 March 2016

Admissibility of the Nearest Neighbor Rule

There exist distributions for which for all n, the 1-nn rule is better than the k-nn rule for any $k \geq 3$

Proof:

Let S_0 and S_1 be two spheres of radius 1 centered at a and b, where $\|a - b\| > 1$. Given $Y = 1$, X is uniform on S_1, while given $Y = 0$, X is uniform on S_0, whereas $P\{Y = 1\} = P\{Y = 0\} = 1/2$. We note that given n observations, with the 1-nn rule:

$\mathbb{E}\{L_n\} = P\{Y = 0, Y_1 = ... = Y_n = 1\} + P\{Y = 1, Y_1 = ... = Y_n = 0\} = \frac{1}{2^n}$

For the $k - nn$ rule, k being odd, we have:

$\mathbb{E}\{L_n\} = P\left\{Y = 0, \sum_{i=1}^{n} \mathbb{I}_{Y_i=0} \leq \frac{k}{2}\right\} P\left\{Y = 1, \sum_{i=1}^{n} \mathbb{I}_{Y_i=1} \leq \frac{k}{2}\right\}$

$= \mathbb{P}\{Bin(n, 1/2) \leq \frac{k}{2}\}$

$= \frac{1}{2^n} \sum_{j=0}^{k/2} \binom{n}{j} > \frac{1}{2^n}$ when $k \geq 3$

Hence, the $k - nn$ rule is worse than the $1 - nn$ rule for every n when the distribution is given above. We refer to the exercises regarding some interesting admissibility questions for k-nn rules.

Theorem 32.4

Sometimes the cost of guessing zero while the true value of Y is one is different from the cost of guessing one, while Y = 0. These situations may be handled as follows. Define the costs:

$C(m, l), m, l = 0, 1$

Here $C(Y, g(X))$ is the cost of deciding on $g(X)$ when the true label is Y. The risk of a decision function g is defined as the expected value of the cost:

$R_g = \mathbb{E}\{C(Y, g(X))\}$

Note that if:

$$C(m, l) = \begin{cases} 1 & \text{if } m \neq l \\ 0 & \text{otherwise} \end{cases}$$

then the risk is just the probability of error. Introduce the notation:

$Q_m(x) = \eta(x)C(1, m) + (1 - \eta(x))C(0, m), m = 0, 1$

Then we have the following extension of Theorem 2.1:

Define:

$$\hat{g}(x) = \begin{cases} 1 & \text{if } Q_1(x) \geq Q_0(x) \\ 0 & \text{otherwise} \end{cases}$$

Then for all decision functions g we have $R_{\hat{g}} \leq R_g$