STAT760_Homework1

Natalie Bladis

2023-01-25

Exercise 1. (ESL Ex. 2.1) – 5 pts

Suppose each of K classes has an associated target tk, which is a vector of all zeros, except a one in the kth position. Show that classifying to the largest element of ˆy amounts to choosing the closest target, mink ‖tk −ˆy‖, if the elements of ˆy sum to one.

Solution:

Need to show $\underset{k}{argmin}||t_k - \hat{y}|| = \underset{k}{argmax}(y_k)$:

$$\begin{aligned}
\underset{k}{argmin}||t_k - \hat{y}|| &= \underset{k}{argmin}||t_k - \hat{y}||^2 && \text{a norm is always postitive and } x \to x^2 \ is\ monotonic, \\
&= \underset{k}{argmin}\sqrt{\sum_{i=1}^{k}(y_i - (t_k)_i)^2} && \text{by the definition of norm,} \\
&= \underset{k}{argmin}\sum_{i=1}^{k}(y_i - (t_k)_i)^2 && \text{argmin will be equivalent, } \sqrt{}\text{can be ignored,} \\
&= \underset{k}{argmin}\sum_{i=1}^{k}(y_i^2 - 2y_i(t_k)_i) + (t_k)_i^2) && \text{distibution} \\
&= \underset{k}{argmin}[\sum_{i=1}^{k}y_i^2 + \sum_{i=1}^{k}(-2y_i(t_k)_i) + (t_k)_i^2)] && \text{properties of summation} \\
&= \underset{k}{argmin}\sum_{i=1}^{k}(-2y_i(t_k)_i) + (t_k)_i^2) && \text{since the sum } \sum_{i=1}^{k}y_i^2 \text{is the same for all classes k,} \\
&= \underset{k}{argmin}(-2y_k + 1) && \text{since for each k, } \sum_{i=1}^{k}(t_k)_i^2 = 1, \text{and } \sum_{i=1}^{k}(-2y_i(t_k)_i) = y_k \\
&= \underset{k}{argmin}(-2y_k) \\
&= \underset{k}{argmax}(y_k).
\end{aligned}$$

Exercise 2. (ESL Ex. 2.3) – 5 pts

Consider N data points uniformly distributed in a p-dimensional unit ball centered at the origin. Suppose we consider a nearest-neighbor estimate at the origin. The median distance from the origin to the closest data point is given by the expression

$$d(p, N) = (1 - \frac{1}{2}^{\frac{1}{N}})^{\frac{1}{p}}$$

. Derive equation (1).,

#Solution:

Let m be the median distance from the origin to the closest data point. Therefore,

$$P(\text{All N points are further than m from the origin}) = \frac{1}{2}$$

by the definition of the median. The points X_i are indepenednelt distibuted, so

$$\prod_{i=1}^{N} P(||x_i|| > m)$$

Because the points x_i are uniformly distributed in the unit ball,

$$\begin{aligned} \prod_{i=1}^{N} P(||x_i|| > m) &= 1 - P(||x_i|| \leq m) \\ &= 1 - \frac{Km^p}{K} \\ &= 1 - m^p \end{aligned}$$

And therefore,

$$\frac{1}{2} = (1 - m^p)^N$$

Solving for m yields,

$$m = (1 - \frac{1}{2}^{\frac{1}{N}})^{\frac{1}{p}}$$

#Exercise 3. (ESL Ex. 2.8) – 20 pts Note: Please write your own code. Don’t use libraries or packages. Compare the classification performance of linear regression and k-nearest neighbor clas- sification on the zipcode data. In particular, consider only the 2’s and 3’s, and k = 1, 3, 5, 7 and 15. Show both the training and test error for each choice. The zipcode data are available from the book website https://www.hastie.su.domains/ElemStatLearn/ (https://www.hastie.su.domains/ElemStatLearn/).