Proofs and correction exercises on *The Elements of Statistical Learning*

11 mai 2017

Chapter 2 : Overview of Supervised Learning

Exercise 1

In our context, each of our output variables $(Y_1, ..., Y_n)$ is categorical and belongs to some set of cardinal K. Let denote them as follows :

For $i = 1, ..., n$, $Y_i \in \{0,1\}^K$ and $\sum_j Y_i(j) = 1$. Given some new input $x \in \mathbb{R}^p$, we want to predict the group it belongs to. Therefore, we estimate the vector of probabilities $\widehat{p} := (\widehat{p}_1, ..., \widehat{p}_K)$, where $\widehat{p}_k$ denotes the (estimated) probability that x belongs to group k, with $k = 1, ..., K$. The predicted class is the integer $\bar{k}$ that maximizes the $\widehat{p}_k$-s. Furthermore, for all integer $k = 1, ..., K$, we denote by t_k a vector of all zeros except a 1 at k-th position.

Show that $\bar{k} = \arg\max_k \widehat{p}_k = \arg\min_k \|t_k - \widehat{p}\|$.

Exercise 1 - Solution

$$\|t_k - \widehat{p}\|^2 = \sum_{j=1}^{K} (t_k(j) - \widehat{p}_j)^2 = (t_k(k) - \widehat{p}_k)^2 = (1 - \widehat{p}_k)^2$$

is a 2-d degree polynomial defined on $[0,1]$ and decreasing on that set. Therefore, $\|t_k - \widehat{p}\|$ is minimal when $\widehat{p}_k$ is maximal.

Exercise 3 - Solution

For all generated points $x_1, ..., x_N$, define $r_i := \|x_i\|$ their distance to the origin. By definition, the median distance m from the closest point to the origin must satisfy $\mathbb{P}(\min_i r_i \geq m) = \frac{1}{2}$. By independence of the generation of points,

$$\begin{aligned}\mathbb{P}(\min_i r_i \geq m) = \prod_{i=1}^{N} \mathbb{P}(r_i \geq m) = \prod_{i=1}^{N} (1 - \mathbb{P}(r_i \leq m)) = \prod_{i=1}^{N} (1 - \frac{\lambda_p(\mathbf{B}(0,m))}{\lambda_p(\mathbf{B}(0,1))}) \\ = (1 - \frac{\pi^{p/2} m^p}{\Gamma(p/2+1)} \frac{\Gamma(p/2+1)}{\pi^{p/2} 1^p})^N = (1 - m^p)^N\end{aligned}$$

By solving $(1 - m^p)^N = 1/2$ we find the solution $m = (1 - 2^{-1/N})^{1/p}$.