Machine learning@Data science, 2016–Exercises.

Set 2. Due February 12, 2016

Problem 7 Let the joint distribution of (X, Y) be such that X is uniform on the interval $[0, 1]$, and for all $x \in [0, 1]$, $\eta(x) = x$. Determine the prior probabilities $\mathbf{P}\{Y = 0\}, \mathbf{P}\{Y = 1\}$ and the class-conditional densities $f(x|Y = 0)$ and $f(x|Y = 1)$.

Calculate R^*, R_{1-NN}, and R_{3-NN} (i.e., the Bayes risk and the asymptotic risk of the 1-, and 3-nearest neighbor rules).

Problem 8 Let $X_1, \ldots, X_n$ be independent random variables taking values in $[0, 1]$. Denote $m = \mathbf{E}\sum_{i=1}^n X_i$. Prove that for any $t \geq m$,

$$\mathbf{P}\left\{\sum_{i=1}^n X_i \geq t\right\} \leq \left(\frac{m}{t}\right)^t e^{t-m} .$$

Hint: Use Chernoff's bounding technique. Use the fact that by convexity of $e^{\lambda x}$, $e^{\lambda x} \leq xe^{\lambda} + (1-x)$.

Problem 9 Let R_{k-NN} denote the asymptotic risk of the k-nearest neighbor classifier, where k is an odd positive integer. Use the expression of R_{k-NN} found in class to show that

$$R_{k-NN} - R^* \leq \sup_{p\in[0,1/2]} (1 - 2p)\mathbf{P}\{\text{Bin}(k, p)) > k/2\} .$$

Use Hoeffding's inequality to deduce from this that

$$R_{k-NN} - R^* \leq \frac{1}{\sqrt{ke}} .$$

Problem 10 (RADEMACHER AVERAGES.) Let A be a bounded subset of $\mathbb{R}^n$. Define the *Rademacher average*

$$R_n(A) = \mathbf{E} \sup_{a\in A} \frac{1}{n}\left|\sum_{i=1}^n \sigma_i a_i\right| ,$$

where $\sigma_1, \ldots, \sigma_n$ are independent random variables with $\mathbf{P}\{\sigma_i = 1\} = \mathbf{P}\{\sigma_i = -1\} = 1/2$ and $a_1, \ldots, a_n$ are the components of the vector a. Let $A, B \subset \mathbb{R}^n$ be bounded sets and let $c \in \mathbb{R}$ be a constant. Prove the following "structural" results:

$$R_n(A \cup B) \leq R_n(A) + R_n(B), \qquad R_n(c \cdot A) = |c|R_n(A) , \qquad R_n(A \oplus B) \leq R_n(A) + R_n(B)$$

where $c \cdot A = \{ca : a \in A\}$ and $A \oplus B = \{a + b : a \in A, b \in B\}$. Moreover, if $\text{absconv}(A) = \left\{\sum_{j=1}^N c_j a^{(j)} : N \in \mathbb{N},\ \sum_{j=1}^N |c_j| \leq 1, a^{(j)} \in A\right\}$ is the absolute convex hull of A, then

$$R_n(A) = R_n(\text{absconv}(A)) .$$

Problem 11 A half plane is a set of the form $H_{a,b,c} = \{(x, y) \in \mathbb{R}^2 : ax + by \geq c\}$ for some real numbers a, b, c. Detemine the n-the shatter coefficient of the classes

$$\mathcal{A}_0 = \{H_{a,b,0} : a, b \in \mathbb{R}\} \quad \text{and} \quad \mathcal{A} = \{H_{a,b,c} : a, b, c \in \mathbb{R}\} .$$