Machine learning@Data science, 2019–Exercises.

Set 3. Due February 28, 2019

Problem 8 A circle in the plane is a set of the form $C_{\boldsymbol{c},r} = \{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x} - \boldsymbol{c}\| \le r\}$ for some $\boldsymbol{c} \in \mathbb{R}^2$ and $r \ge 0$.

Determine the VC dimension of the class $\mathcal{A} = \{C_{\boldsymbol{c},r} : \boldsymbol{c} \in \mathbb{R}^2, r \ge 0\}$ of all circles.

What is the VC dimension of the class $\mathcal{A}_1 = \{C_{\boldsymbol{c},1} : \boldsymbol{c} \in \mathbb{R}^2\}$ of all circles of radius one?

Problem 9 A half plane is a set of the form $H_{a,b,c} = \{(x,y) \in \mathbb{R}^2 : ax + by \ge c\}$ for some real numbers a, b, c. Detemine the n-the shatter coefficient of the classes

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

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:

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

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

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

Problem 11 Write a program that implements the perceptron algorithm and test the running time and the probability of error of the resulting linear classifier. Consider both linearly separable and non-separable data.

For the linearly separable case, you may generate (X, Y) pairs (with $X \in \mathbb{R}^d$ and $Y \in \{-1, 1\}$) such that $\mathbf{P}\{Y = 1\} = 1/2$ and, conditionally on $Y = 1$, X is uniformly distributed in $[-1, -a] \times [-1, 1] \times \cdots \times [-1, 1]$ (i.e., the first component of X is uniform between -1 and a and the rest are uniform in $[-1, 1]$) while, conditionally on $Y = -1$, X is uniformly distributed in $[a, 1] \times [-1, 1] \times \cdots \times [-1, 1]$. Here $a \in [0, 1]$ is a parameter, playing the role of the margin. Try different values of a, d and different sample sizes. Run the algorithm many times for each data set to estimate the number of updates until the algorithm stops and generate independent data to estimate the probability of error.

For the non-separable case, generate the two class-conditional distributions as multivariate normals with means $(0, \ldots, 0)$ and $(m, 0, 0, \ldots, 0)$ and identity covariance matrix. Here the algorithm never stops making updates so run it by passing through the data set several times and define a stopping time. Again, test the performance of the classifier for several sample sizes, dimensions, and choices of m. Compare the obtained probability of error to the Bayes risk.