Set 3. Due February 26, 2021

Problem 9 A circle in the plane is a set of the form $C_{\boldsymbol{c},r} = \{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x} - \boldsymbol{c}\| \leq r\}$ for some $\boldsymbol{c} \in \mathbb{R}^2$ and $r \geq 0$.

Determine the VC dimension of the class $\mathcal{A} = \{C_{\boldsymbol{c},r} : \boldsymbol{c} \in \mathbb{R}^2, r \geq 0\}$ 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 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 $\mathrm{absconv}(A) = \left\{ \sum_{j=1}^{N} c_j a^{(j)} : N \in \mathbb{N},\ \sum_{j=1}^{N} |c_j| \leq 1, a^{(j)} \in A \right\}$ is the absolute convex hull of A, then

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

Problem 11 (SPARSE LINEAR CLASSIFIERS.) Let $k < d$ and consider the class $\mathcal{A}_k$ of sets of the form

$$\{x \in \mathbb{R}^d : x^T w \geq 0\}$$

where $w \in \mathbb{R}^d$ is such that it has at most k nonzero components. Give an upper bound of the VC dimension of $\mathcal{A}_k$. Compare this to the VC dimension of the class of all linear half-spaces (without any restriction on the components of w).

Problem 12 Write a program that generates n uniformly distributed points $X_1, \ldots, X_n$ in the d-dimensional cube $[-2^{1/d}, 2^{1/d}]^d$. Assign labels such that $Y_i = 1$ if $X_i \in [-1, 1]^d$ and $Y_i = 0$ otherwise. (Thus, about half of the points have label 1.)

Train two different classifiers, both performing empirical risk minimization as follows.

The first classifier selects the smallest cube of form $[-a, a]^d$ (for some $a \geq 0$) that contains all points with label 1 and classifies with 1 inside the cube and with 0 outside.

The second classifier selects the smallest rectangle of form $[a_1, b_1] \times \cdots \times [a_d, b_d]$ (for arbitrary real numbers $a_i \leq b_i$, $i = 1, \ldots, d$) that contains all points with label 1 and classifies with 1 inside the rectangle and with 0 outside.

Try a wide range of values of d and n and plot the test error (measured on a large independent test set) for both classifiers. Explain what you see.