ECE 2372 - Homework 4

<< — VC BOUND — >>

Please upload your solutions to Canvas by February 25, 2024

Problem 1:

The VC dimension, rather implicitly, depends both on the input space $\mathcal{X}$ and $\mathcal{H}$– since $\mathcal{H}$ is defined as the mapping of $\mathcal{X}$ to a binary set of $\{-1, +1\}$.

For a fixed $\mathcal{H}$, consider two input spaces $\mathcal{X}_1 \subseteq \mathcal{X}_2$. Let $d_{VC}(\mathcal{X}_1)$ and $d_{VC}(\mathcal{X}_2)$ denote the VC dimension of $\mathcal{H}$ with respect to these input spaces. Show that $d_{VC}(\mathcal{X}_1) \leq d_{VC}(\mathcal{X}_2)$.

Problem 2:

Suppose $\mathcal{H}$ is obtained by evaluating some polynomial of degree d and comparing the result to a threshold, i.e.,

$$\mathcal{H} = \left\{ h : h(x) = sign\left(\sum_{i=0}^{d} c_i x_i \right) \right\}$$

for some $c_0, ...c_d \in \mathbb{R}$, and $\mathcal{X} \in \mathbb{R}$ as well. Prove that d_{VC} of this $\mathcal{H}$ is exactly $d+1$ by showing that

(a) There are $d+1$ points which are shattered by $\mathcal{H}$.

(b) There are no $d+2$ points which are shattered by $\mathcal{H}$. [Hint: Try relating this to a linear classifier in d dimensions and use the result from first problem]

Problem 3: Suppose the VC dimension of our $\mathcal{H}$ is $d_{VC} = 3$ (e.g., linear classifiers in $\mathbb{R}^2$). And suppose that we have an algorithm for selecting some $h^* \in \mathcal{H}$ based on a training sample size n.

(a) Using the VC generalization bound, give an upper bound on $R(h^*)$ that holds with probability at least 0.95 when $n = 100$. Repeat this for $n = 1,000$ and $10,000$.

(b) How large does n need to be to obtain a bound of the form:

$$R(h^*) \leq \hat{R}(h^*) + 0.01$$

that holds with probability at least 0.95? How does this compare to the "rule of thumb" that we stated in the class?

[Bonus] Problem 4: Usually, the VC dimension corresponds to the number of parameters or degrees of freedom in the hypothesis set. However this is not *always* true. This problem is an example of the not-always case. Consider the hypothesis set fr $x \in \mathbb{R}$ with

$$\mathcal{H} = \left\{ h : h(x) = (-1)^{\lfloor \alpha x \rfloor} \right\} \tag{1}$$

for some $\alpha \in \mathbb{R}$, and $\lfloor \cdot \rfloor$ is the floor function (i.e., $\lfloor \alpha x \rfloor$ is the largest integer less than α. Show that even though this hypothesis set has a single parameter, it has an infinite VC dimension! [Hint: Consider the inputs of $x_i = 10^i$ for $i = 1, .., n$ and show how to choose α to realize an arbitrary dichotomy.]