ECE 2372 - Homework 4

<< — VC BOUND — >>

Please upload your solutions to Canvas by February 25, 2024

Jiyang Liu 4731134

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 $d_{VC}(\mathcal{X}_1) \leq d_{VC}(\mathcal{X}_2)$.

Sol.

For x_1, the VC dimension is $d_{VC}(x_1)$, which means in x_1 there are $d_{VC}(x_1)$ points can be shattered.
$x_1 \subseteq x_2$, x_2 has all the points of x_1. Therefore, x_2 has at least $d_{VC}(x_1)$ points can be shattered. Which means

$$d_{VC}(x_1) \leq d_{VC}(x_2)$$

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 $c_0, \ldots c_d \in \mathbb{R}$, 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]

a)

Assume that

$$c = (c_0, c_1, \ldots, c_d)^T$$
$$x = (x_0, x_1, \ldots, x_d)^T$$

And we rewrite H:

$$H = \{h: h(x) = sign(c^T x)\}$$

Suppose there are $d+1$ points in x,

$$x = \begin{pmatrix} - & \left(x^{(1)}\right)^T & - \\ - & \left(x^{(2)}\right)^T & - \\ \ldots & \ldots & \ldots \\ - & \left(x^{(d+1)}\right)^T & - \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_{d+1} \end{pmatrix}$$

$Let\ x\ be:$

$$x = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 1 \end{pmatrix}$$

$x\ is\ non-sigular. Therefore\ for\ \forall y, we\ can\ always\ find$

$$c = x^{-1}y$$

$that\ saitisfied:$

$$sign(xc) = y$$

$And\ we\ proved\ that\ there\ are\ d+1\ points\ shattered\ by\ H.$

b)

$$x_2 = x_1 \cup \{new\ points\}$$

$From\ problem\ 1,$

$$x_1 \subseteq x_2$$
$$d_{VC}(x_1) \le d_{VC}(x_2)$$

$Suppose$

$$x = \begin{pmatrix} - & \left(x^{(1)}\right)^T & - \\ - & \left(x^{(2)}\right)^T & - \\ \dots & \dots & \dots \\ - & \left(x^{(d+1)}\right)^T & - \\ - & \left(x^{(d+2)}\right)^T & - \end{pmatrix}$$

$Since\ x^{(j)} \in R^{d+1}, for\ d+2\ points, there\ are\ at\ least\ one\ point\ x^{(j)}\ can\ be\ represented\ as$

$$x^{(j)} = \sum_{i \ne j} \alpha_i x^i$$

And

$$y^{(j)} = sign\left(c^T x^{(j)}\right) = \sum_{i \ne j} \alpha_i sign\left(c^T x^{(i)}\right)$$

$For\ a\ given\ x^{(i)}, x^{(j)}\ \left(or\ y^{(j)}\right)\ is\ determined\ by\ \ x^{(i)}\ and\ not\ free.$

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:

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

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

a)

$$d_{VC} = 3, \delta = 1 - 0.95 = 0.05$$

$$R(h^*) \leq \hat{R}(h^*) + \sqrt{\frac{8d_{VC}}{n} log \frac{8n}{\delta}}$$

$n = 100$:

$$Upper\ Bound = \sqrt{\frac{8*3}{100} log \frac{8*100}{0.05}} = 1.524$$

$n = 1000$:

$$Upper\ Bound = \sqrt{\frac{8*3}{1000} log \frac{8*1000}{0.05}} = 0.536$$

$n = 10000$:

$$Upper\ Bound = \sqrt{\frac{8*3}{10000} log \frac{8*10000}{0.05}} = 0.185$$

b)

$$\sqrt{\frac{8*3}{n} log \frac{8*n}{0.05}} = 0.01$$

$If\ n = 10^6, Bound \approx 0.021\ , if\ n = 10^7, Bound \approx 0.007\ , therefore,$

$$n \in (10^6, 10^7)$$

$Compare\ to\ the\ 'rule\ of\ thumb', we\ see\ that\ for\ every\ n,$

$$n \geq 10d_{vc} = 30$$