Machine Learning, Spring 2023
Homework 4

WeiLiang Sun 2020533010

Due on 23:59 April 27, 2023

1 Understanding VC dimension (50 points)

In this part, you need to complete some mathematical proofs about VC dimension. Suppose the hypothesis set

$$\mathcal{H} = \{f(x, \alpha) = \text{sign } (\sin(\alpha x))|, \alpha \in \Re\}$$

where x and f are feature and label, respectively.

- Show that $\mathcal{H}$ cannot shatter the points $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4$. (20 points)

  (Key: Mathematically, you need to show that there exists y_1, y_2, y_3, y_4, for any $\alpha \in \Re$, $f(x_i) \neq y_i, i = 1, 2, 3, 4$, for example, $+1, +1, -1, +1$)

- Show that the VC dimension of $\mathcal{H}$ is ∞. (Note the difference between it and the first question) (30 points)

  (Key: Mathematically, you have to prove that for any label sets $y_1, \cdots, y_m, m \in \mathbb{N}$, there exists $\alpha \in \Re$ and $x_i, i = 1, 2, \cdots, m$ such that $f(x; \alpha)$ can generate this set of labels. Consider the points $x_i = 10^{-i}$...)

Solution:

1.

We prove it by contradiction: We suppose there exists an α making that the four points labeling $--+-$, meaning that $sin(\alpha) < 0, sin(2\alpha) < 0, sin(3\alpha) \geq 0, sin(4\alpha) < 0$

$$sin(4\alpha) = 2sin(2\alpha)cos(2\alpha) < 0 \ and \ sin(2\alpha) < 0$$
$$\Rightarrow cos(2\alpha) = 1 - 2sin^2(\alpha) > 0$$
$$\Rightarrow sin^2(\alpha) < \frac{1}{2}$$

Meanwhile, we also have:

$$sin(3\alpha) = 3sin(\alpha) - 4sin^3(\alpha)\ and\ sin(3\alpha) \geq 0, sin(\alpha) < 0$$

$$\Rightarrow 3 - 4sin^2(\alpha) \leq 0 \Rightarrow sin^2\alpha \geq \frac{3}{4}$$

We both have that $sin^2(\alpha) < \frac{1}{2}$ and $sin^2\alpha > \frac{3}{4}$ which makes a contradiction. Thus we show that $\mathcal{H}$ cannot shatter the points $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4$

2.

For any $m > 0$ we can take the set of points $(x_1, x_2, ..., x_m)$ into consideration, whose labels are $y_1, y_2, ..., y_m$ respectively. Also we set the label $\in \{-1, +1\}^m$. What we need to do in the next step is to find a parameter which can make the hypothesis set $\mathcal{H}$ shatter the points.

We choose the parameter α as the following:

$$\alpha = \pi(1 + \sum_{i=1}^{m} 2^i y_i^{'})$$

In this condition, we can get the result as following:

$$\alpha x_j = \alpha 2^{-j} = \pi(2^{-j} + \sum_{i=1}^{m} 2^{i-j} y_i^{'})$$

$$= \pi(2^{-j} + (\sum_{i=1}^{j-1} 2^{i-j} y_i^{'} +) + y_j^{'} + (\sum_{i=1}^{m-j} 2^i y_i^{'}))$$

so according to the formula we can have:

$$\pi y_j^{'} < \pi(2^{-j} + (\sum_{i=1}^{j-1} 2^{i-j} y_i^{'} +) + y_j^{'}) \leq \pi(\sum_{i=1}^{j} 2^{-i} + y_j^{'}) < \pi(1 + y_j^{'})$$

which means that if $y_j = 1$ then $y_j^{'} = 0$ leading to $0 < \alpha x_j < \pi$, so $sign(\alpha x_j) = 1$. We can also have that when $y_j = -1$, $sign(\alpha x_j) = -1$

2 Bias-variance decomposition (50 points)

When there is noise in the data, $E_{out}(g^{(\mathcal{D})}) = \mathbb{E}_{\mathbf{x},y}\left[(g^{(\mathcal{D})}(\mathbf{x}) - y(\mathbf{x}))^2\right]$, where $y(\mathbf{x}) = f(\mathbf{x}) + \epsilon$. If ϵ is a zero-mean noise random variable with variance σ^2, show that the bias-variance decomposition becomes

$$\mathbb{E}_{\mathcal{D}}\left[E_{out}(g^{(\mathcal{D})})\right] = \sigma^2 + \text{bias} + \text{var}$$

Solution:

$$\mathbb{E}_{\mathcal{D}}\left[E_{out}(g^{(\mathcal{D})})\right] = \mathbb{E}_{\mathcal{D}}\left[\mathbb{E}_{\mathbf{x},y}\left[(g^{(\mathcal{D})}(\mathbf{x}) - y(\mathbf{x}))^2\right]\right]$$

$$= \mathbb{E}_{\mathbf{x},y}\left[\mathbb{E}_{\mathcal{D}}\left[(g^{(\mathcal{D})}(\mathbf{x}) - y(\mathbf{x}))^2\right]\right]$$

$$= \mathbb{E}_{\mathbf{x},y}\left[\mathbb{E}_{\mathcal{D}}\left[(g^{(\mathcal{D})}(\mathbf{x})^2 - 2\overline{g}(\mathbf{x})y(\mathbf{x}) + y(\mathbf{x})^2\right]\right]$$

We use Fubini's theorem to get:

$$\mathbb{E}_{\mathcal{D}}\left[(g^{(\mathcal{D})}(\mathbf{x})^2 - 2\overline{g}(\mathbf{x})y(\mathbf{x}) + y(\mathbf{x})^2\right]$$

$$= \mathbb{E}_{\mathcal{D}}\left[(g^{(\mathcal{D})}(\mathbf{x})^2\right] - 2\overline{g}(\mathbf{x})(f(\mathbf{x}) + \epsilon) + (f(\mathbf{x}) + \epsilon)^2$$

$$= \mathbb{E}_{\mathcal{D}}\left[(g^{(\mathcal{D})}(\mathbf{x})\right] - \overline{g}(\mathbf{x})^2) + (\overline{g}(\mathbf{x})^2 - 2\overline{g}(\mathbf{x})f(\mathbf{x}) + f(\mathbf{x})^2) + \epsilon^2 - 2\epsilon(\overline{g}(\mathbf{x}) - f(\mathbf{x})$$

$$= \mathbb{E}_{\mathcal{D}}\left[(g^{(\mathcal{D})}(\mathbf{x}) - \overline{g}(\mathbf{x}))^2\right] + (\overline{g}(\mathbf{x}) - f(\mathbf{x}))^2 + \epsilon^2 - 2\epsilon(\overline{g}(\mathbf{x}) - f(\mathbf{x})$$

$$= var(\mathbf{x}) + bias(\mathbf{x}) + \epsilon^2 - 2\epsilon(\overline{g}(\mathbf{x}) - f(\mathbf{x})$$

Now put this formula back to the original formula:

$$\mathbb{E}_{\mathcal{D}}\left[E_{out}(g^{(\mathcal{D})})\right] = \mathbb{E}_{\mathbf{x},y}[var(\mathbf{x})] + \mathbb{E}_{\mathbf{x},y}[bias(\mathbf{x})] + \mathbb{E}_{\mathbf{x},y}[\epsilon^2] - 2\mathbb{E}_{\mathbf{x},y}\epsilon[\overline{g}(\mathbf{x}) - f(\mathbf{x})]$$

$$= var + bias + \mathbb{E}_{\curvearrowleft}[\mathbb{E}_{\epsilon}[\epsilon^2]] - 2\mathbb{E}_{\mathbf{x}}[\mathbb{E}_{y|x}[(\overline{g}(\mathbf{x}) - f(\mathbf{x}))\epsilon|\mathbf{x}]]$$

$$= var + bias + Var_{\epsilon}[\epsilon] - 2\mathbb{E}_{\mathbf{x}}[(\overline{g}(\mathbf{x}) - f(\mathbf{x}))\mathbb{E}_{\epsilon}[\epsilon]]$$

$$= var + bias + \sigma^2$$