Machine Learning Foundations HW2

B06502152, 許書銓

December 14, 2020

1

The answer is (c). The problem ask to find which set could be shattered by the hypothesis. From the lecture slide, we have the conclusion that the VC dimension for N-D perceptron is $N + 1$. Hence, the VC dimension for 3-D is 4. There is no possiblility for (3) to be shatted, since there are 5 points in that set. For other choice, we are looking for whose rank is exactly 4. (a) rank is 3. (b) rank is 3. (c) rank is 4. (d) rank is 3.

2

The answer should be (d). Since in the scenrio of $N = 4$, there are 14 dichotomies, fulfiliing the condition when $4N - 2$.

3

The answer should be (b). For this problem, we have to divide the problem into 2 parts. First, we have to prove that H can't generate 2^N dichotomies, and second, we have to find that there exist some set of N inputs, which can be shatted by the H. To begin with, we assume that the VC dimension is 2.

From the definition of 2D perceptron, we realive that

$$h_w(x) = sign(w_0 \mathbf{x_0} + w_1 \mathbf{x_1} + w_2 \mathbf{x_2})$$

$$= sign(\begin{bmatrix} \mathbf{x_0} \\ \mathbf{x_1} \\ \mathbf{x_2} \end{bmatrix} * \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix})$$

any set of 3 distinct inputs is not shattered by H

First, we examine that the inputs are linear independent, and we choose that $x_0 = 1$

$$h_w(x) = sign(\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} * \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}) \quad = sign(\begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix})$$

From the upper scenerio, it is obvious that the $sign(y_0)$ is simply fixed since the positive 2D perceptron require that $w_0 > 0$. Hence, we can not find that the output condition of $\begin{bmatrix} -sign(w_0) \\ sign(y_1) \\ sign(y_2) \end{bmatrix}$

some set of 2 distinct inputs is shattered by H

4

The answer should be (b). Since the meaning of $x_1^2 + x_2^2 + x_3^2$ is the $distance^2$ between the origin and the point. Hence, the idea of the hypothesis is like a 3D version of the hypothesis of positive interval in 2D., which means we chould choose 2 points from N+1 interval and plus the condition that choose 2 points from the same interval.

5

The answer should be (b). Since the growth function of the hypothesis is $\frac{1}{2}N^2 + \frac{1}{2}N + 1$ and the break point is N = 3. Hence, the VC dimension is 2.

6

The answr should be (c). From the lecture note, we have that

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N}ln(\frac{4m_H(2N)}{\delta})}$$

And, since the cheating hypothesis's

$$E_{out}(g_*) = E_{in}(g_*) = E_{in}(g)$$

From the upper equations, we can conclude that the upper bond is

$$E_{out}(g) - E_{out}(g_*) \leq \sqrt{\frac{8}{N}ln(\frac{4m_H(2N)}{\delta})}$$

7

The answer should be (d). Since if we have M hypothesis, we chould have M dichotomies. From the defnition of the VC dimension, we realize that $M \leq 2^N$. Hence, VC dimension should be $\lfloor lg(M) \rfloor$.

8

The answer should be (c). We use the hypothesis that count the number of 1's. There are k number in a set; hence, we can shatter all dichotomies, which are 2^k dichotomies. Thus, the VC dimension is k.

9

The answer should be (b). The correct condition is that some set of d distinct inputs is shattered by H and that any set of $d+1$ distinct inputs is not shattered by H.

10

The answer should be (a). Since if we draw a circle, we can determine infinte recantangles, whose end points are on the circle.

11

The answer should be (d). We would like to know the edited out sample error. The possibility of the edited out sample error is that the original out sample error times the possibility not to flip and the possibility of the original correct one times the possibility to flip. Mathematically,

$$E_{out}(h, \tau) = E_{out}(h, 0) * (1 - \tau) + (1 - E_{out}(h, 0)) * \tau$$

$$E_{out}(h, \tau) = (1 - 2\tau) * E_{out}(h, 0) + \tau$$

$$E_{out}(h, 0) = \frac{E_{out}(h, \tau) - \tau}{1 - 2\tau}$$

12

The answer should be (b).

	f(x) = 1	f(x) = 2	f(x) = 3
y = 1	0.7	0.2	0.1
y = 2	0.1	0.7	0.2
y = 3	0.2	0.1	0.7

$$E_{out}(f) = \frac{1}{3}\Sigma_{k=1}^{3} E_{out}(k)$$

$$E_{out}(f) = \frac{1}{3}[(1 * 0.1 + 4 * 0.2) + (1 * 0.2 + 1 * 0.1) + (4 * 0.1 + 1 * 0.2)]$$

$$E_{out}(f) = 0.6$$

13

The answer should be (b).
For f(x) = 1, the $f_*(x) = 1 * 0.7 + 2 * 0.1 + 3 * 0.2 = 1.5$
For f(x) = 2, the $f_*(x) = 1 * 0.2 + 2 * 0.7 + 3 * 0.1 = 1.9$
For f(x) = 3, the $f_*(x) = 1 * 0.1 + 2 * 0.2 + 3 * 0.7 = 2.6$

$$\Delta(f, f_*) = \frac{1}{3}[(1 - 1.5)^2 + (2 - 1.9)^2 + (3 - 2.6)^2] = 0.14$$

14

The answer should be (d). Since $\delta = 4(4N) * e^{\frac{-1}{8}\epsilon^2 N}$

For $N = 6000$, $\delta = 53.096$

For $N = 8000$, $\delta = 5.811$

For $N = 10000$, $\delta = 0.59$

For $N = 12000$, $\delta = 0.058$

For $N = 14000$, $\delta = 0.005624$
Hence, the smallest N for δ to be smaller than 0.1 is 12000.

15

The answer should be (b).

When s = 1,

$$E_{out}(h, \tau) = P[(x > 0) \cap (x - \theta < 0)] + P[(x < 0) \cap (x - \theta > 0)]$$

$$E_{out}(h, \tau) = \frac{1}{2}\frac{\theta}{2} + \frac{1}{2}\frac{|\theta|}{2}$$

$$E_{out}(h, \tau) = \frac{|\theta|}{2}$$

16-20

(d) -> (b) -> (d) -> (b) -> (a)