Name: Norhan Reda Abdelwahed Ahmed

Sec: 2

BN: 31

Code: 9203639

Subject: Date / /

Problem 2.16

ⓐ Consider $D+1$ distinct points $x_0, x_1, \dots, x_D$
$X$ will be

$$X = \begin{pmatrix} 1 & x_0^1 & \cdots & x_0^D \\ 1 & x_1^1 & \cdots & x_1^D \\ \vdots & & \cdots & \vdots \\ 1 & x_D^1 & & x_D^D \end{pmatrix}$$

We need to show that for any binary labels $(y_0, y_1, \dots, y_D)^T \in \{-1, 1\}^{D+1}$ there exists a set of coefficients $c_0, c_1, \dots, c_D$ such that $h_c(x_i) = y_i$ for each point $x_i$

The function $h_c(x) = \text{Sign}\left(\sum_{i=0}^{D} c_i x^i\right)$ becomes a linear system
So let $c = (c_0, \dots, c_D)^T = X^{-1} y$
and $X$ is a Vandermonde matrix because it has distinct points so the determinant of $X \neq 0$
$\therefore X$ is invertable and $X^{-1}$ exists
so there will be a solution for the linear equation

Subject: Date / /

$\therefore Xc = y$

$\therefore h_c(x_n) = \text{sign}\left(\sum_{i=0}^{D} c_i x_n^i\right) = y_n$

for all $n = 0, \ldots, D$

there exists a set of coefficients $c_0, c_1, \ldots, c_D$ to satisfy any given labeling $y_0, y_1, \ldots, y_D$ implying that $H$ can shatter $D+1$ points

So $m_H(D+1) = 2^{D+1}$

and $d_{vc}(H) \geq D+1$

CamScanner الممسوحة ضوئيا بـ

Subject: ______ Date / /

(b)

Now consider $D+2$ distinct Points
$x_0, x_1, \ldots, x_{D+1}$ vectors
So we will have $D+2$

$x_n^0, x_n^1, \ldots, x_n^D$
$n = 0, \ldots, D+1$
We have $D+2$ vectors in $D+1$ dimensions

To shatter $D+2$ distinct Points
the function would need to alternate
between labels $\{-1, 1\}$ for each Point
However a Polynomial of degree $D$ can have
at most $D$ sign changes

Given $D+2$ district points the function
would need to change its sign at
least $D+1$ times to correctly classify
all Possible labelings since a polynomial
of degree $D$ can only alternate between
$-1$ and $1$ up to $D$ times it will be
impossible for the function to correctly
classify $D+2$ distinct Points with any
binary labeling
$\therefore$ H cannot shatter $D+2$ Points

Problem 2.24

ⓐ $\bar{g}(x) = E_D\left(g^{(D)}(x)\right)$

$g^D(x)$ is the hypothesis $g$ that makes $E_{in}$ to be minimum applied to dataset $D$

$$E_{in}(g) = \sum_{i=1}^{2} \left(f(x_i) - h(x_i)\right)^2$$

$$E_{in}(g) = \sum_{i=1}^{2} \left(x_i^2 - (ax_i + b)\right)^2$$

~~$\partial E_{in}(g)$~~

get the min $E_{in}$ to get $a, b$ values

$$\frac{\partial E_{in}(g)}{\partial a} = 0 \quad , \quad \frac{\partial E_{in}(g)}{\partial b} = 0$$

$$\frac{\partial E_{in}(g)}{\partial a} = -2 \sum_{i=1}^{2} x_i\left(x_i^2 - ax_i - b\right) = 0 \rightarrow ①$$

$$\frac{\partial E_{in}(g)}{b} = -2 \sum_{i=1}^{2} \left(x_i^2 - ax_i - b\right) = 0 \rightarrow ②$$

By expanding the summation of each equation

$$\frac{\partial E_{in}(g)}{\partial a} = x_1(x_1^2 - a x_1 - b) + x_2(x_2^2 - a x_2 - b) = 0 \rightarrow (3)$$

$$\frac{\partial E_{in}(g)}{\partial b} = (x_1^2 - a x_1 - b) + (x_2^2 - a x_2 - b) = 0 \rightarrow (4)$$

~~by~~

$x_1 * (4) - (3)$

$$(x_1 - x_2)(x_2^2 - a x_2 - b) = 0$$

$$x_2^2 - a x_2 - b = 0$$

$$b = x_2^2 - a x_2 \rightarrow (5)$$

$x_2 * (4) - (3)$

$$(x_2 - x_1)(x_1^2 - a x_1 - b) = 0$$

$$x_1^2 - a x_1 - b = 0 \rightarrow (6)$$

by substituting by (5) in (6)

$$x_1^2 - a x_1 - x_2^2 + a x_2 = 0$$

$$a(x_2 - x_1) = x_2^2 - x_1^2$$

$$a\cancel{(x_2 - x_1)} = \cancel{(x_2 - x_1)}(x_2 + x_1)$$

$$a = x_2 + x_1$$

by substituting in (5)

$$b = x_2^2 - (x_2 + x_1)x_2 = \cancel{x_2^2} - \cancel{x_2^2} - x_1 x_2 = -x_1 x_2$$

$$\therefore a = x_2 + x_1 \quad , \quad b = -x_1 x_2$$

$\therefore g^{D}(x) = ax + b = (x_2 + x_1)\, x - x_1 x_2 = x_2 x + x_1 x - x_1 x_2$

$\bar{g}(x) = E_D(g^{D}(x)) = E_D(x_2 x + x_1 x - x_1 x_2)$

$= E_D(x_2 x) + E_D(x_1 x) - E_D(x_1 x_2)$

due to the independence of points $x_1$, $x_2$

$\bar{g}(x) = E_D(x_2)\, x + E_D(x_1)\, x - E_D(x_1)\, E_D(x_2)$

Since the data follows a uniform distribution in the interval $[-1, 1]$

$E_D(x) = \frac{1}{2}\int_{-1}^{1} x \, dx = \frac{1}{2}\frac{1}{2}\left[x^2\right] = 0$

$E_D(x_1) = 0 \quad , \quad E_D(x_2) = 0$

$\therefore \bar{g}(x) = 0 * x + 0 * x - 0 * 0 = 0$

(d)

$$\text{Variance} = E_x\left[E_D\left[\left(g^D(x) - \bar{g}(x)\right)^2\right]\right]$$

$$= E_x\left[E_D\left[\left((x_2+x_1)x - x_1x_2 - 0\right]^2\right]\right]$$

$$= E_x\left[E_D\left[(x_1+x_2)^2x^2 + x_1^2x_2^2 - 2x_1x_2(x_1+x_2)x\right]\right]$$

$$= E_x\left[x^2 E_D(x_1^2+x_2^2+2x_1x_2) + E_D(x_1^2x_2^2) - 2x\,E_D(x_1^2x_2 + x_1x_2^2)\right]$$

since the data follows a random distribution in $[-1,1]$

$$E_D(x^2) = \frac{1}{2}\int_{-1}^{1} x^2\,dx = \frac{1}{2}\times\frac{1}{3}\left[x^3\right]_{-1}^{1} = \frac{1}{3}$$

$$\text{Variance} = E_x\left[x^2\left(\frac{1}{3}+\frac{1}{3}\right) + \left(\frac{1}{3}\times\frac{1}{3}\right)\right]$$

$$= E_x\left[\frac{2}{3}x^2 + \frac{1}{9}\right] = \frac{2}{3}\times\frac{1}{3} + \frac{1}{9} = \frac{1}{3}$$

$$\therefore \boxed{\text{Variance} = \frac{1}{3}}$$

CamScanner الممسوحة ضوئيا بـ

$$\text{bias} = E_x\left[\left(\bar{g}(x) - f(x)\right)^2\right] = E_x\left[\left(0 - x^2\right)^2\right]$$

$$= E_x\left[x^4\right]$$

Since the data follows a uniform distribution in the interval $[-1, 1]$

$$E_x(x^4) = \frac{1}{2}\int_{-1}^{1} x^4\,dx = \frac{1}{2} * \frac{1}{5}\left[x^5\right]_{-1}^{1} = \frac{1}{5}$$

$$\therefore \boxed{\text{bias} = \frac{1}{5}}$$

$$E_{out}(E_D g^{(D)}) = E_x\left[E_D\left(g^{(D)}(x) - f(x)\right)^2\right]$$

from slides

$$= E_x(\text{var}(x) + \text{bias}(x))$$

$$E_{out} = \frac{1}{3} + \frac{1}{5} = \frac{8}{15}$$

$$\therefore \boxed{E_{out} = \frac{8}{15}}$$