2.1

$$\epsilon(M, N, \delta) = \sqrt{\frac{1}{2N} \ln\left(\frac{2M}{\delta}\right)}$$

A.1) a) $M = 1$, $\epsilon \leq 0.05$, $\delta = 0.03$

using the given equation we have

$N = 839.941$

b) $M = 100$
$N = 1760.9750$

c) $M = 1000$
$N = 2682.00908$

2.2

For N=4, we can pick points: (1,3),(2,4),(3,1),(4,2). It's easy to see that these points are shattered by positive rectangles. So mH(4)=24.

The idea is that for any two points, if we draw a rectangle using them as diagnoal points, the rectangle should NOT contain any other point. Otherwise, whenever the two diagnoal points have values 1, the middle point will have value 1 as well, which excludes the possibility of having -1.

For N=5, if we draw horizontal and vertical lines through each of the four points above, the plane is divided into grids. The four points enclusing a 9-grid area. It's clear that the fifth point can't lie within the 9-grid area. Otherwise, there'll always a rectangle (constructed by two points) contains the fifth point.

In the same way, if we place the fifth point outside the 9-grid area, it's easy to see that the point will always lie below or above at least two points (in either x or y direction). These three points construct a rectangle which contains a point in it. This shows that mH(5)<25.

We have the VC dimension dVC(H)=4, and mH(N)≤∑i=04(Ni).

2.3

$$vc\ bound = \sqrt{\frac{8 \times \ln\left(\frac{4 \times growth(2N)}{\delta}\right)}{N}}$$

A3 $N = 100$

~~vc bound = $\sqrt{8 \times \ln}$~~

$vc\ bound = 0.848159$

$N = 10000$

$vc\ bound = 0.104278$

2.4

```python
#!/usr/bin/env python3

import numpy as np


def get_sample_size(dvc, deviation, epsilon, n):
    max_iterations = 1000
    tolerance = 10

    for _ in range(max_iterations):
        rhs = 8 * np.log(4 * (pow(2 * n, dvc) + 1) / deviation)
        rhs = rhs / epsilon / epsilon

        result = n - rhs
        if np.abs(result) <= tolerance:
            break

        n = int(rhs)
    return n


if __name__ == "__main__":
    dvc = 10
    deviation = 0.05
    epsilon = 0.05

    # Starting value for N
    n = 1000

    sample_data = get_sample_size(dvc, deviation, epsilon, n)
    print(f"Sample Size: {sample_data}")

```

Sample Size: 45290

2.5

A-5) i) Assume we have $h(x) = ax + b$, then

$$E_x[(f(x) - h(x))^2] = E_x[(\sin(\pi x) - (ax+b))^2]$$

$$= \int_{-1}^{1} [\sin(\pi x) - (ax+b)]^2 \, p(x) dx$$

$$= \int_{-1}^{1} \frac{1}{2} [\sin(\pi x) - (ax+b)]^2 dx$$

$$\frac{\delta E}{\delta a} = -\int_{-1}^{1} x [\sin(\pi x) - (ax+b)] dx = 0$$

$$\frac{\delta E}{\delta b} = -\int_{-1}^{1} [\sin(\pi x) - (ax+b)] dx = 0$$

Solving, $\frac{2}{\pi} - \frac{2}{3} a = 0$

$b = 0$

$a = \frac{3}{\pi}$

Best hypothesis that approx. f in the mean squared error :

a) $h(x) = \frac{3}{\pi} x$; ($h(x) = ax + b$)

b) $h(x) = \frac{3}{\pi} x$; $h(x) = ax$

c) $h(x) = 0$; $h(x) = b$

ii) Exp. value 'D' of hypothesis produces two points x_1 & x_2

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

$$= \sum_{i=1}^{N} (\sin(\pi x_i) - (a x_i + b))^2$$

wrt a & b (Taking derivative)

$$\frac{\delta E_{in}(g)}{\delta a} = -2 \sum_{i=1}^{N} x_i (\sin(\pi x_i) - (a x_i + b)) = 0$$

$$\frac{\delta E_{in}(g)}{\delta b} = -2 \sum_{i=1}^{N} (\sin(\pi x_i) - (a x_i + b))$$

Solving a & b

$$(x_2 - x_1)(\sin \pi x_2 - a x_2 - b) = 0$$

$$(x_1 - x_2)(\sin \pi x_1 - a x_1 - b) = 0$$

We assume $x_1 \neq x_2$

$$a = \frac{\sin \pi x_2 - \sin \pi x_1}{x_2 - x_1}$$

$$b = \frac{x_2 \sin \pi x_1 - x_1 \sin \pi x_2}{x_2 - x_1}$$

if $x_1 = x_2$, let $x_2 = x_1 + k$

substituting in the above equation with $k \to 0$

we can obtain hypothesis tangential to f

$$g^{D}(x) = \frac{\sin \pi x_2 - \sin \pi x_1}{x_2 - x_1} \cdot x + \frac{x_2 \sin \pi x_1 - x_1 \sin \pi x_2}{x_2 - x_1}$$

for hypothesis set $h(x) = ax$

$$g^{0}(x) = \frac{x_1 \sin \pi x_1 + x_2 \sin \pi x_2}{x_1^2 + x_2^2} \cdot x$$

for $h(x) = b$

$$g^{0}(x) = \frac{1}{2}(\sin \pi x_1 + \sin \pi x_2)$$

x is uniformly distributed between $[-1, 1]$
bias, variance can be calculated by
$h(x) = b$

$$E_D[g^{D}(x)] = E_D\left[\frac{1}{2}(\sin \pi x_1 + \sin \pi x_2)\right]$$

$$= 0$$

bias component

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

$$= E_x[(0 - \sin(\pi x))^2]$$

$$= E_x[(\sin(\pi x))^2] = \frac{1}{2}$$

Variance Component

$$\text{variance} = E_x[E_D[(g^D(x) - \bar{g}(x))^2$$

$$= E_x[E_D(\tfrac{1}{2}(\sin \pi x_1 + \sin \pi x_2) -$$

$$= E_x[\tfrac{1}{4} E_D (\sin \pi x_1 + \sin \pi x$$

$$= E_x[E_D(\sin^2 \pi x_1 + \sin^2 \pi x_2 + 2 \sin \pi x_1 \sin \pi x_2)]$$

$$= E_x\left[\tfrac{1}{4}\left(\tfrac{1}{2} + \tfrac{1}{2} + 0\right)\right]$$

$$= 0.25$$

Out of sample error = Variance + bias

$$= \frac{1}{2} + \frac{1}{4} = 0.75$$

2.6

To show that k is a break point for H, we need to show H cannot shatter any set of k points x1, xk.

• If k is a break point, then mH (k) < 2k,

• In general, it is easier to find a break point for H than to compute the full growth function for that H.

• So the correct option is (d)

2.7

M T W T F S S

Page No :

Date

YOUVA

A.7.) Let $f(x) = \begin{cases} 1 & , \text{if } \omega^T x + b > 0 \\ -1 & , \text{otherwise} \end{cases}$

Consider $d+1$ points $x^{(0)} = (0, \ldots, 0)^T$, $x^{(1)} = (1, 0, \ldots 0)$
$x^{(2)} = (0, 1, \ldots, 0)^T, \ldots, x^{(d)} = (0, 0, \ldots, 1)^T$

After these $d+1$ points being arbitrarily labelled: $y = (y_0, y_1, \ldots, y_d)^T \in \{-1, 1\}^{d+1}$

Let $b = 0.5\, y_0$ and $w = (w_1, w_2, \ldots, w_d)$ where $w_i = y_i$, $i \in \{1, 2, \ldots, d\}$. Thus $f(x)$ can label all these $d+1$ points correctly. So, the VC dimensions of perceptron is at least $d+1$. —— (1)

Expand $x \in R^d$ to $X \in R^{d+1}$ by letting $(X)^T = (x^T, 1)$, and let $W^T = (w^T, b)$

Thus, $f(X) = \begin{cases} 1 & \text{if } W^T X > 0 \\ -1 & \text{otherwise} \end{cases}$

Assume, there exists $d+2$ points the perceptron in R^d can shatter, namely $x^{(1)}, x^{(2)}, \ldots, x^{(d+2)} \in R^d$ corressponding to $X^{(1)}, X^{(2)}, \ldots, (X)^{(d+2)} \in R^d$

Since $d+2$ points in R^{d+1}, there exist certain i such that
$$X^{(i)} = \sum_{j \neq i} a_j \cdot X^{(j)}$$

where at least one $a_j \neq 0$.

Page No.:
Date:

Let $S = \{ j \mid j \neq i,\ a_j \neq 0 \}$.

$\forall j \in S$, we give $x^{(j)}$ the label $\text{Sign}(a_j)$, and give $x^{(i)}$ a label -1

By our assumption, there exists W that makes $f(x)$ label those $d+2$ points correctly.

So, $\forall j \in S$, we have

$$a_j \cdot W^T x^{(j)} > 0$$

and

$$W^T x^{(i)} \leq 0$$

Also,

$$W^T x^{(i)} = W^T \left(\sum_{j \neq i} a_j \cdot x^{(j)} \right)$$

$$= W^T \left(\sum_{j \in S} a_j \cdot x^{(j)} \right)$$

$$= \sum_{j \in S} a_j \cdot W^T x^{(j)} > 0$$

So our assumption is false. The VC dimension of perceptron in $\mathbb{R}^d$ is at most $d+1$. ——— (2)

~~Thus, from the above, we can conclude~~

From (1) & (2) we can conclude that VC dimension of perceptron in $\mathbb{R}^d$ is $d+1$

2.8

M T W T F S S
Page No.:
Date:
YOUVA

)a) $E_{out}(g) \le E_{in}(g) + \sqrt{\frac{1}{2N} \ln\left(\frac{2M}{\delta}\right)}$

$\epsilon = 0.05$
$N = 200$

Test bounds

$\text{test bound} = \sqrt{\frac{1}{2N} \ln\left(\frac{2}{\epsilon}\right)}$

$= \sqrt{\frac{1}{2 \times 200} \times \ln\left(\frac{2}{0.05}\right)} = 0.096$

Train bound

$M = 1000$
$N = 400$

$\text{train bound} = \sqrt{\frac{1}{2N} \ln\left(\frac{2M}{\epsilon}\right)}$

$= \sqrt{\frac{1}{2 \times 400} \times \ln\left(\frac{2 \times 1000}{0.05}\right)} = 0.115$

$E_{in}(g) = 0.115 \quad E_{test}(g) = 0.096$

Thus, error bar on in simple error is higher than the error from test error

b) If more samples for testing is used then the samples for training are less

We may end up with a hypothesis that is not as good as we could have arrived if using more training samples.

So $E_{test}(g)$ might be ~~[illegible]~~ way too off even the error bar on it is small.