HW3

↳ Hardik Prajapati (2678294168)

1. AML Pb. 1.7

Given:-

$P[H] = \mu$ (Probability of error)

$P[k \mid N, \mu] = \binom{N}{k} \mu^k (1-\mu)^{N-k}$

Training error $\nu = \frac{k}{N}$

a) Sample Size $(N) = 10$

$\mu = 0.05$

$P[\text{At least one coin will have } \nu = 0] = ?$

Case:- I) one coin.

$P[\text{At least one coin will have } \nu = 0] = P[0 \mid 10, 0.05]$

$= \binom{10}{0} (0.05)^0 (0.95)^{10-0}$

$= (1)(1)(0.95)^{10}$

$= 0.5987$

Case:- II) 1000 coins.

→ As coins generate samples independently,

$P[\text{At least one coin will have } \nu = 0] = 1 - P[\text{No coin will have } \nu = 0]$

$$\therefore P[\text{At least one coin will have } \nu=0] = 1 - \{P[\text{a coin have } \nu>0]\}^{1000}$$

$$= 1 - \{1 - P[\text{a coin will have } \nu=0]\}^{1000}$$

$$= 1 - \{1 - P[0 \mid 10, 0.05]\}^{1000}$$

$$= 1 - \left\{1 - \binom{10}{0}(0.05)^0(0.95)^{10-0}\right\}^{1000}$$

$$= 1 - \{1 - (0.95)^{10}\}^{1000}$$

$$= 1$$

Case :- III) 10 00 000 Coins.

Similarly,

$$\therefore P[\text{At least one coin will have } \nu=0] = 1 - \{1 - (0.95)^{10}\}^{1000000}$$

$$= 1$$

Now, $\mu = 0.8$.

Case :- 1 coin,

$$P[\text{At least one coin will have } \nu=0] = 1 - \{1 - (0.2)^{10}\}^{1}$$

$$= 1.024 \times 10^{-7}$$

Case :- 1000 coins

$$P[\text{Atleast one coin will have } \nu = 0] = 1 - \{1 - (0.2)^{10}\}^{1000}$$

$$= 0.00010239$$

Case :- 1000000 coins

$$P[\text{Atleast one coin will have } \nu = 0] = 1 - \{1 - (0.2)^{10}\}^{1000000}$$

$$= 0.09733$$

Q1 b) 1000 coins, $\mu = 0.05$

$E_{out}(h) = 0.05$

i) 1000 coins represent $\rightarrow$ 1000 datasets (independent) of size 10 drawn from input space with replacement.

ii) $P[\text{At least one coin will have } \nu = 0] \rightarrow P[\text{At least one dataset will have Zero training Error}]$

Q2. a) i) $E_{D(10)}(h) = 0.3$ (Simulation)

↳ No, $E_{D(10)}(h) = 0.2$

↳ This is because, samples in datset are drawn at random with replacement.

↳ Also, Dataset size is too small, for Error rate to be more accurate to actual Error value

ii) $\mu = 0.2$, $\nu = \frac{k}{N}$

$$P[E_{D(10)}(h) = 0.2] = P[k=2 \mid 10, 0.2]$$

$$= \binom{10}{2}(0.2)^2(0.8)^8$$

$$= 0.3019$$

$$= 0.30$$

Q2. b) i) Simulation

ii) Fraction of runs with $E_D(h) \neq 0.2 = \frac{71}{100} = 0.71$

$$P[E_{D(10)}(h) = 0.2] = 0.3$$

$$\Rightarrow P[E_{D(10)}(h) \neq 0.2] = 0.7$$

Values are closeby.

iii) Simulation

Q2. c) i) $\mu = 0.2$, $N = 100$

$$P[E_{D(100)}(h) = 0.2] = P[k=20 \mid 100, 0.2]$$
$$= \binom{100}{20}(0.2)^{20}(0.8)^{80}$$
$$= 0.0993$$
$$= 0.10$$

$$\Rightarrow P[E_{D(100)}(h) \neq 0.2] = 0.90$$

Almost equal

$\rightarrow$ Fraction of runs with $E_D(h) \neq 0.2 = \frac{92}{100} = 0.92$

ii) $\mu = 0.5$, $N = 10$

$$P[E_{D(10)}(h) = 0.5] = P[k = 5 \mid 10, 0.5]$$
$$= \binom{10}{5}(0.5)^{5}(0.5)^{5}$$
$$= 0.2460$$
$$= 0.25$$

$$\Rightarrow P[E_{D(10)}(h) \neq 0.5] = 0.75$$

Small Difference

$\rightarrow$ Fraction of runs with $E_D(h) \neq 0.5 = \frac{70}{100} = 0.70$

iii) $\mu = 0.5$, $N = 100$

$$P[E_{D(100)}(h) = 0.5] = P[k = 50 \mid 100, 0.5]$$
$$= \binom{100}{50} (0.5)^{50} (0.5)^{50}$$

$$= 0.0795$$
$$= \underline{\underline{0.08}}$$

$$\Rightarrow P[E_{D(100)}(h) \neq 0.5] = \underline{\underline{0.92}}$$ Closeby.

↳ Fraction of runs with $E_D(h) \neq 0.5 = \frac{89}{100} = \underline{\underline{0.89}}$

Q2. d. i) With increase in N, error-estimate from a test dataset improves & become more accurate
↳ from Simulation results, it's evident that when N changes from $10 \to 100$ for a particular μ, Sample Mean Error rate improves with Sample Standard Deviation taking Low value.

ii) For classifier c) ii) $N = 10$, $\mu = 0.5$:-
↳ # of test datasets with $E_{d(10)}(h) \leq 0.45 = \underline{\underline{40}}$

For classifier c) iii) $N = 100$, $\mu = 0.5$:-
↳ # of test datasets with $E_{d(100)}(h) \leq 0.45 = \underline{\underline{22}}$

↳ As $E_d(h) \leq 0.45$ {lesser than true value = 0.5},

⇒ Datasets has less no. of wrong classified samples

⇒ Classifier is moving towards more better classification

⇒ Classifier is learning.

3.

AML Pb. 2.1

$\delta = 0.03$

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

Eq. 2.1 :- $E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$

a) $M = 1$, $\epsilon \leq 0.05$

$\Rightarrow \sqrt{\frac{1}{2N} \ln \frac{2(1)}{0.03}} \leq 0.05$

$\Rightarrow N \geq 839.941$

$\Rightarrow N = 840$ or greater

b) $M = 100$, $\epsilon \leq 0.05$

$\Rightarrow \sqrt{\frac{1}{2N} \ln \frac{2(100)}{0.03}} \leq 0.05$

$\Rightarrow N \geq 1760.98$

$\Rightarrow N = 1761$ or greater

c) $M = 10{,}000$, $\varepsilon \leq 0.05$

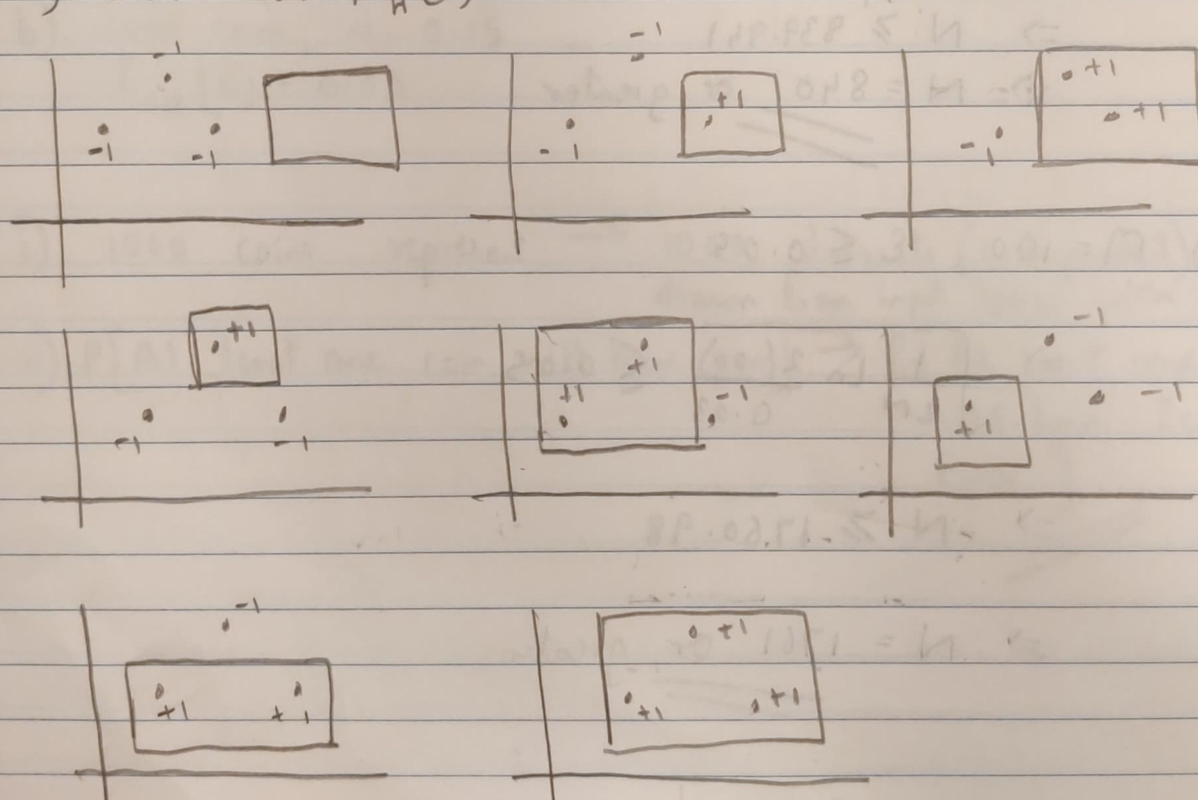
$$\Rightarrow \sqrt{\frac{1}{2N} \ln \frac{2(10000)}{0.03}} \leq 0.05$$

$$\Rightarrow N \geq 2682.01$$

$$\Rightarrow N = 2683 \text{ or greater}$$

4. [AML Pb 2.2.]

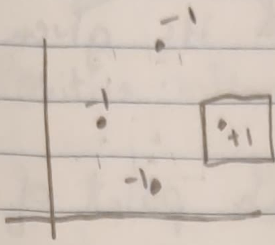
a) Show that $m_H(3) = 2^3$



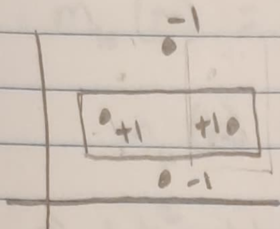
$$\Rightarrow m_H(3) = 8 = 2^3$$

b

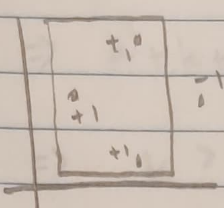
$m_H(4) = 2^4$



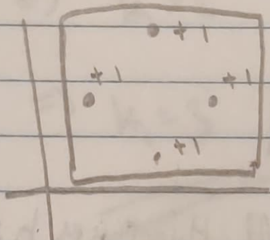
$\rightarrow$ Similar 3 others, So total $= 4$
(taking single point inside/on Rectangle boundary)



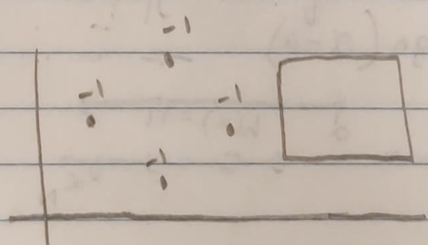
$\rightarrow$ Similar 5 others So total $= 6$
(taking 2 points' inside/on the Rectangle boundary)



$\rightarrow$ Similar 3 others, So total $= 4$



$\rightarrow$ One



$\rightarrow$ One

$\Rightarrow$ Total Dichotomies $= 16$

$\Rightarrow m_H(4) = 2^4 = 16.$

c. Given :- $m_H(5) < 2^5$

Since, H can shatter 4 points & its given that $m_H(5) < 2^5$,

$\Rightarrow k_0 = 5$ is the smallest break point of H.

$\Rightarrow d_{vc}(H) = k_0 - 1 = 4.$

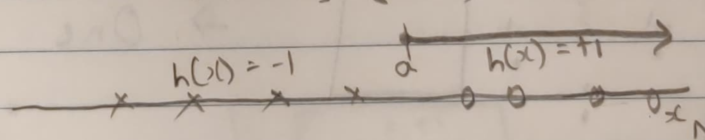
$$\Rightarrow m_H(N) \leq N^{d_{vc}} + 1$$

$$\Rightarrow m_H(N) \leq N^4 + 1$$

Ans:- Upper Bound $= N^4 + 1$

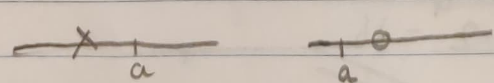
5. [Based on AML Exercise 2.1]

a) Positive Rays: H consists of all Hypotheses $h: \mathbb{R} \rightarrow \{-1, +1\}$
$h(x) = \text{sign}(x - a)$

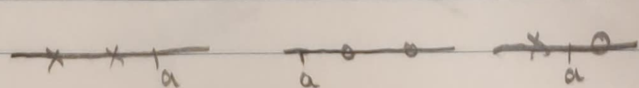


$$m_H(N) = N + 1$$

$N = 1$,



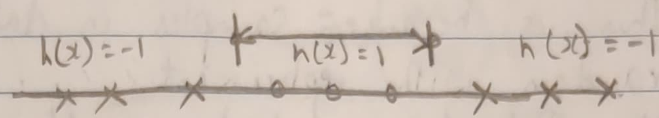
$N = 2$,



$\Rightarrow$ Only 3 dichotomies

$\Rightarrow k_0 = 2$ is the smallest break point.

5. b Positive intervals :-



$$m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$$

$$m_H(k) < 2^k \quad , \text{ for breakpoint}$$

$$\Rightarrow \frac{1}{2}k^2 + \frac{1}{2}k + 1 < 2^k$$

$$\Rightarrow k^2 + k + 2 < 2^{k+1}$$

$$\Rightarrow k > 2 \quad \text{or} \quad 0 < k < 1$$

$$\Rightarrow k > 2$$

$\Rightarrow k_0 = 3$ is the smallest break point