Name - Anushree Jumade - (axj174430).
Machine Learning (CS6375).
Homework 6.

Q.1. AdaBoost Algorithm — Chapter 14

\# a) AdaBoost choose 'h_1' in the first iteration.

b) $\alpha_1 \rightarrow$

$$\epsilon_m = \frac{\sum_{n=1}^{N} w_n^{(m)} I(y_m(x_n) \neq t_n)}{\sum_{n=1}^{N} w_n^{(m)}} \rightarrow \text{Eq. (14.16) from Bishop}$$

Here as given h_1 mistakes one out of a set of 17 pts.

$$\therefore \epsilon_m = \frac{1}{17}$$

$$\alpha_m = \ln\left\{\frac{1-\epsilon_m}{\epsilon_m}\right\} \quad \text{— Eq}^n \text{ 14.17 from Bishop.}$$

$$\therefore \alpha_m = \ln\left(\frac{1-1/17}{1/17}\right) = \ln\left(\frac{16}{1}\right)$$

$$\boxed{\alpha_m = \ln(16)}$$

c) W_2

① Case I :- points on which chosen learner made a mistake. $\boxed{w_n^1 = 1/17 \ \left(\frac{1}{N}\right)}$

$$w_n^{(m+1)} = w_n^{(m)} \exp\left\{\alpha_m I(y_m(x_n) \neq t_n)\right\} \quad \text{Eq. 14.18 Bishop}$$

↑ error made.

$$w_n^2 = w_n^1 e^{\{\ln(16) I(y_m(x_n) \neq t_n)\}}$$
$$= \frac{1}{17} e^{\ln(16)} = \frac{16}{17}.$$

② Case II :-

$$w_n^{(m+1)} = w_n^{(m)} \rightarrow \text{No error made.}$$
$$= 1/17$$

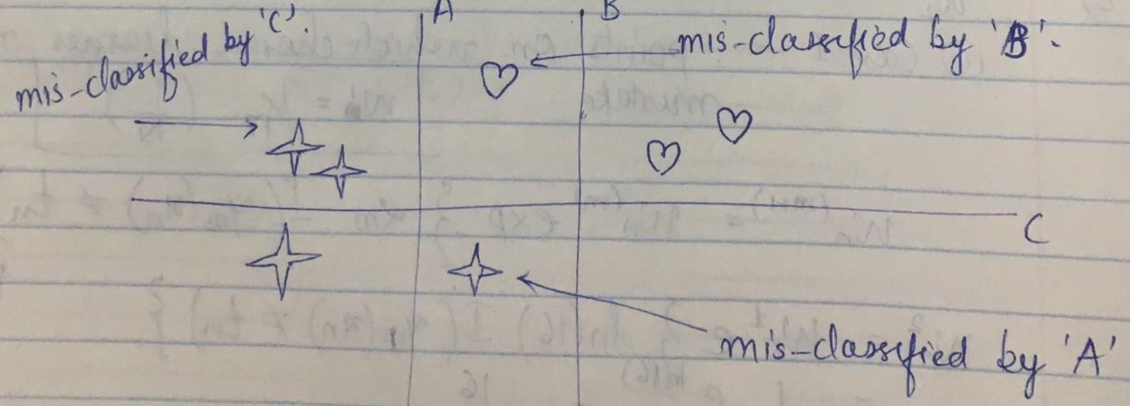
Ans - case I - 16/17
case II - 1/17.

Q.2. Heart → +ve samples.
A, B, C → Three decision stumps.
AdaBoost chooses 'A' as the first.

It will choose B because the one example mis-classified by A is correctly classified by B.
AdaBoost algorithm assigns higher weights to the mis-classified examples in the next iteration.
Also
mistakes made by B → 1.
mistakes made by C → 2

In the first iteration,
$$\epsilon_1 = \frac{1}{7}$$ (mistakes made by A = 1 & total data points = 7)

$$\alpha = \ln\left(\frac{1-\epsilon_m}{\epsilon_m}\right) = \ln\left(\frac{1-1/7}{1/7}\right) = \ln(6).$$



Q.3. 1D data with a mixture of 2 gaussians

1D data pts $x = [1 \quad 10 \quad 20]$.

O/P of E step

$$\underset{(r_{i,c})}{R} = \begin{bmatrix} 1 & 0 \\ 0.4 & 0.6 \\ 0 & 1 \end{bmatrix} \rightarrow \text{probability of } x_i \text{ belonging to cluster 'C'.}$$

a). Likelihood function :-

$$P(D|\theta) = \prod_{i=1}^{3} \sum_{k=1}^{2} \pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k).$$

b) Missing weights

$$\pi_1 = \frac{1 + 0.4 + 0}{3} = \frac{1.4}{3} = 0.466.$$

$$\pi_2 = \frac{0 + 0.6 + 1}{3} = \frac{1.6}{3} = 0.53$$

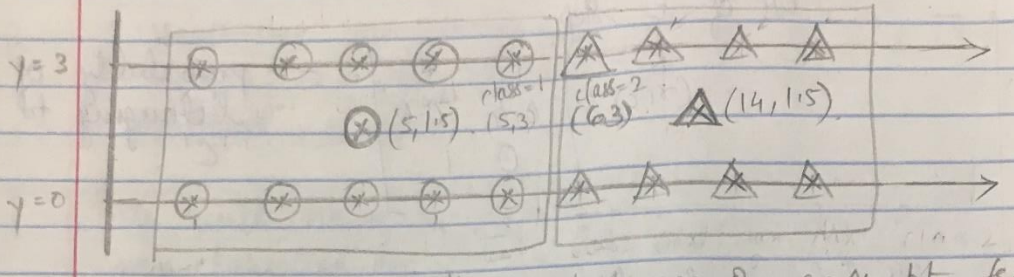
c). Means :-

$$\mu_1 = \frac{(1)1 + 0.4(10) + 0(20)}{1.4} = \frac{5}{1.4} = 3.57.$$

$$\mu_2 = \frac{0(1) + 0.6(10) + 1(20)}{1.6} = \frac{26}{1.6} = 16.25.$$

Q.4. K-means with K=2 applied to the given data.

After 1st iteration



Based on euclidean distance of each pt from the initial defined centroids, the given data pts form the above 2 clusters.

New centroid for cluster 1 :-

$$X = \frac{\sum_{i=1}^{N} x_i}{N}, \quad Y = \frac{\sum_{i=1}^{N} y_i}{N}$$

$\rightarrow X = \frac{2(1+3+5+7+9)}{10}$ $\quad Y = \frac{5(0+3)}{10} = \frac{3}{2} = 1.5$

$X = \frac{25}{5} = 5.$

New centroid for cluster 2 :

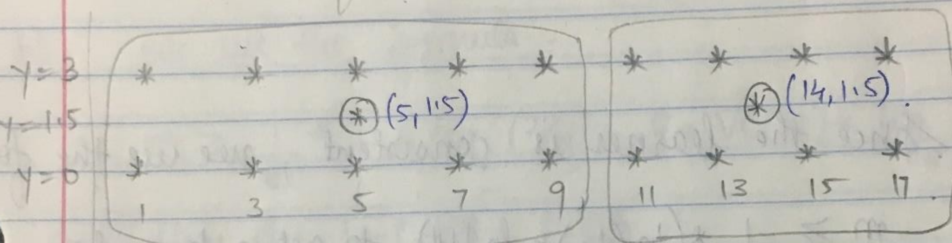
$X = \frac{2(11+13+15+17)}{8}$ or $\frac{56}{4} = 14$

$Y = \frac{4(0+3)}{8} = \frac{3}{2} = 1.5$

After 2nd iteration :-
All the pts get classified into the same cluster as before as the euclidean distance of each pt comes to be the minimum for the new centroid of its original assigned cluster.

This means there is no change in the two clusters & this marks the end of iterations on the pts.

& the final cluster centroid is as



Q.5. 2 I/P perception . $\epsilon_{max} = 5\%$ confidence – 90%.

We have .

$$m \geq \frac{1}{\epsilon}\left(4\log_2\left(2/\delta\right) + 8\,VC(H)\log_2\left(13/\epsilon\right)\right)$$

VC dimension is 3 (2+1) for a 2 I/P perceptron.

$\epsilon \rightarrow$ error $= 0.05$

1 - confidence $\rightarrow \delta = 1 - 0.9 = 0.1$.

$$\therefore m \geq \frac{1}{0.05}\left(4\log_2\left(\underbrace{2/0.1}_{=20}\right) + 8\times3\times\log_2\left(\underbrace{13/0.05}_{260}\right)\right)$$

$$m \geq \frac{1}{0.05}\left(4\times4.32 + 24\times8.022\right)$$

$$m \geq \frac{1}{0.05}(209.808) \rightarrow m \geq \lceil 4196.16 \rceil$$

$$m \geq 4197$$

Ans. Upper bound on the number of training examples is $\underline{\underline{4197}}$.

Q.6. VC dimension is always less than size of the hypothesis space — $\underline{\underline{\text{True}}}$.

The VC dimension of a hypothesis is at the most the log of the hypothesis space

Q.7.
a) Since the learner is consistent, we use the formula

$m \geq \frac{1}{\varepsilon} * \left(\ln\left(\frac{1}{d}\right) + \ln|H|\right)$ to get a bound on 'm'

where $d = 1 - 0.99 = 0.01$
& $\varepsilon = 0.05$.

$|H| \rightarrow$ no of rectangles :-
$x_1 \in [0, 199]$
$x_2 \in [0, 99]$.
possible rectangles in the space covered by x_1 & x_2 - choosing 2 distinct end-pts along x_1 & x_2.
$x_1 \rightarrow [0, 199] \rightarrow$ total 200 pts.
$x_2 \rightarrow [0, 99] \rightarrow$ total 100 pts.

$x_1 \rightarrow$ $^{200}C_2 = \frac{200!}{(200-2)!\,2!} = \frac{200 \times 199}{2} = 19900$

$x_2 \rightarrow$ $^{100}C_2 = \frac{100!}{98!\,2!} = \frac{100 \times 99}{2} = 4950$

Total possible rectangles $= 19900 \times 4950 = 98505000$

$\therefore\ m \geq \frac{1}{0.05} * \left(\ln \frac{1}{0.01} + \ln(98505000) \right).$

$\frac{1}{0.01} = 100$

$m \geq 20 * (4.605 + 18.4056).$

$m \geq 20 * (23.0106).$

$m \geq \lceil 460.217 \rceil.$

$m \geq 461$

Ans. No of training examples sufficient is 461.

b) We use the formula :-

$$m \geq \frac{1}{\epsilon} * \left[4 \log_2 (2/d) + 8 * VC(H) * \log_2 \left(\frac{13}{\epsilon} \right) \right].$$

Here $d = 1 - 0.95 = 0.05.$

$\epsilon = 0.01$

$VC(H) = 2 \times$ no of dimensions $= 2 \times 3 = 6$

$$m \geq \frac{1}{0.01} \left[4 \times \log_2 \left(\frac{2}{0.05} \right) + 8 \times 6 \times \log_2 \left(\frac{13}{0.01} \right) \right].$$

$$m \geq \frac{1}{0.01} \left[4 \times \log_2 40 + 48 \times \log_2 1300 \right]$$

$$m \geq 100 \times \left[4 \times 5.3219 + 48 \times 10.3443 \right].$$

$m \geq 100 \times 517.814.$

$m \geq \lceil 51781.4 \rceil$

$m \geq$ ~~51780~~ 51782.

Ans - No of training examples sufficient to satisfy the required condition is 51782.