Machine Learning

HOMEWORK-V

1.) a.) AdaBoost chooses 'hi' in the first iteration.

b) $\alpha_1 \rightarrow$

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

As given, h_1 mistakes 1 out of a set of 17 points here.

$\therefore \epsilon_m = 1/17$.

$$\alpha_m = \ln\left\{\frac{1-\epsilon_m}{\epsilon_m}\right\}$$

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

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

c) Case 1: The learner made an error while choosing the points.

$$w' = 1/17 \; (1/N)$$

$$w_n^{(m+1)} = w_n^{(m)} \exp\left\{\alpha_m I(y_m(x_n) \neq t_n)\right\}$$

$$w_n^2 = w_n^1 e\left\{\ln(16) I(y_1(x_n) \neq t_n\right\}$$

$$= \frac{1}{17} e^{\ln(16)} = \frac{16}{17}$$

Case 2: No error made

$$w_n^{(m+1)} = w_n^{(m)} = \frac{1}{17}$$

2.) The samples here are : Heart.
The decision stumps are A, B, C

AdaBoost chooses 'A' as the first.

It will now choose B because the one example misclassified by A is correctly classified by B and as we know the misclassified examples are assigned higher weight in the next iteration.

B does make one another error, but C makes 2 errors.

In the first iteration,

$\epsilon_1 = \frac{1}{7}$ (mistakes made by A = 1, Total points = 7)

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

3.) 1D data with a mixture of 2 Guassians
1D data points $x = [1, 10, 20]$
Output of the E step

$$R = \begin{bmatrix} 1 & 0 \\ 0.4 & 0.6 \\ 0 & 1 \end{bmatrix}$$

a) Likelihood function:

$$P(D|\theta) = \prod_{i=1}^{3} \sum_{K=1}^{2} \pi_K \mathcal{N}(x_i | \mu_K, \sigma_K)$$

b.) Missing weights

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

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

c) Means

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

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

4) K-means with $K=2$ applied to the given data

After first iteration two classes / clusters are formed with initially defined centroids as (5,3) and (6,3). Based on the euclidean distance of each point from those we find new centroids for the two clusters.

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

New centroid for cluster 1

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

$$X = 25/5 = \underline{5} \qquad \underline{\underline{Y = 1.5}}$$

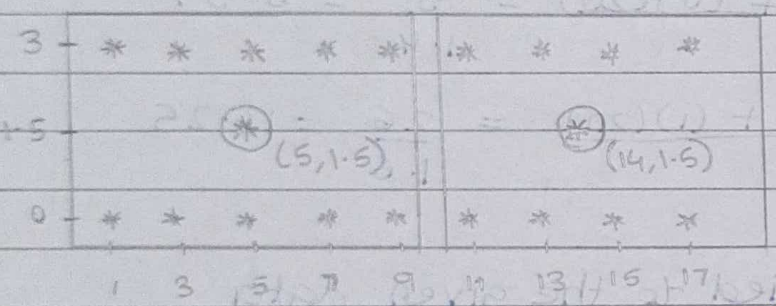
New centroid for cluster 2

$$X = \frac{2(11+13+15+17)}{8} = \frac{56}{4} \qquad Y = \frac{4(0+3)}{8} = \frac{3}{2}$$

$$\underline{\underline{X = 14}} \qquad \underline{\underline{Y = 1.5}}$$

After the second iteration, all the points get classified into the same cluster as before as the euclidean distance of each point comes to be the minimum for the new centroid.

This means there is no change in the two clusters and this makes the end of iterations on the points & the final clusters obtained are as follows



5) 2 input perceptron with $\epsilon_{max} = 5\%$ & confidence $= 90\%$

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

VC dimension is 3 (2+1) for a 2 input perceptron.

ϵ - error $= 0.05$

1-confidence $\Rightarrow \delta = 1 - 0.9 = 0.1$

$\therefore m \geq \frac{1}{0.05}\left(4\log_2\left(\frac{2}{0.1}\right) + 8 \times 3 \times \log_2\left(\frac{13}{0.05}\right)\right)$

$m \geq \frac{1}{0.05}\left(4(4.32) + 24(8.022)\right)$

$m \geq \frac{209.808}{0.05} \Rightarrow m \geq \lceil 4196.16 \rceil$

$\therefore \boxed{m \geq 4197}$

The upper bound on the number of training examples is 4197.

6.) VC dimension is always less than size of the hypothesis space. - True

The maximum value that can be computed for the VC dimension of a hypothesis is the log of the hypothesis space.

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

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

$d = 1 - 0.99 = 0.01$ & $\epsilon = 0.06$

$|H| \rightarrow$ 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 points along x_1 and x_2.

$x_1 \rightarrow [0, 199] \rightarrow$ total 200 points

$x_2 \rightarrow [0, 99] \rightarrow$ total 100 points

$x_1 \rightarrow {}^{200}C_2 = \frac{200!}{198!\,2!} = \frac{200 \times 199}{2} = 19900$

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

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

$\therefore m \geq \frac{1}{0.05}\left(\ln\left(\frac{1}{0.01}\right) + \ln(9850500)\right)$

$m \geq 20\,(4.605 + 18.4056)$

$m \geq 20\,(23.0106)$

$m \geq \lceil 460.21 \rceil$

$m \geq 461$.

The number of training examples sufficient is 461.

b.) We use the formula:

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

where, $d = 0.05$, $\epsilon = 0.01$

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

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

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

$$\geq 100\left[4 \times (5.3219) + 48(10.3443)\right]$$

$$\geq 100(517.814)$$

$\therefore m \geq \lceil 51781.4 \rceil$

$\therefore$ $\underline{\underline{m \geq 51782}}$

The number of training examples sufficient to satisfy the required conditions is 51782.