CS-534 Machine Learning

Written Assignment - 2

Submitted By - Akash Agarwal
OSU ID No :- 933-471-097

Ques1. Let p_a, p_c, p_g, p_t be those unknown frequencies. Assume that we have obtained a strand of DNS sequences & we want to estimate the unknown frequencies. Let n_a, n_c, n_g, n_t be the corresponding no. of bases that you observe for A, C, T & G respectively. Derive the maximum likelihood estimates for the unknown parameters p_a, p_c, p_g & p_t.

Ans1. A.T.Q. for the DNA Bases A, C, T & G, the corresponding no. of bases observed are: n_a, n_c, n_g & n_t respectively.

The likelihood $L(p_a, p_c, p_g \& p_t) = p_a^{n_a} \times p_c^{n_c} \times p_g^{n_g} \times p_t^{n_t}$

$$\Rightarrow L(p_a, p_c, p_g, p_t) = \prod_{i=1}^{4} p_i^{n_i}$$

Let log likelihood be $l(p_a, p_c, p_g, p_t)$

$$\Rightarrow l(p_a, p_c, p_g, p_t) = \log(L(p_a, p_c, p_g, p_t))$$

$$= \sum_{i=1}^{4} n_i \log p_i$$

Also, we have a constraint that $-$ $p_a + p_c + p_g + p_t = 1$

In cases where we need to deal with constraints, we choose Lagrange Multiplier for constraint optimization.

Now, constraint : $p_a + p_c + p_g + p_t = 1$

$$\Rightarrow 1 - (p_a + p_c + p_g + p_t) = 0$$

Applying Lagrangian

$$L(x, \alpha) = f(x) + \alpha g(x) \quad \forall \alpha \geq 0$$

$$\Rightarrow L(x, \alpha) = (n_a \log p_a + n_b \log p_b + n_g \log p_g + n_t \log p_t) + \alpha(1 - p_a - p_c - p_g - p_t)$$

differentiating $L(x, \alpha)$ w.r.t p_a, p_b, p_g & p_t respectively

$$\frac{\partial L}{\partial p_a} = \frac{n_a}{p_a} - \alpha \quad \text{—(I)}$$

$$\frac{\partial L}{\partial p_c} = \frac{n_c}{p_c} - \alpha \quad \text{—(II)}$$

$$\frac{\partial L}{\partial p_g} = \frac{n_g}{p_g} - \alpha \quad \text{—(III)}$$

$$\frac{\partial L}{\partial p_t} = \frac{n_t}{p_t} - \alpha \quad \text{—(IV)}$$

To maximize, i.e. ~~to~~ find MLE

$$\frac{\partial L}{\partial p_i} = 0 \quad \forall \quad i = \cancel{a, b}, a, c, g \,\&\, t$$

$\therefore$ Equating eq^n (I), (II), (III) & (IV) to 0.

$$\frac{n_a}{p_a} = \frac{n_c}{p_c} = \frac{n_g}{p_g} = \frac{n_t}{p_t} = \alpha$$

$$\Rightarrow \alpha = \frac{n_a + n_c + n_g + n_t}{p_a + p_c + p_g + p_t}$$

A.T.Q. $\quad p_a + p_c + p_g + p_t = 1$

$$\Rightarrow \alpha = n_a + n_c + n_g + n_t$$

$$\therefore \; p_a = \frac{n_a}{n_a + n_c + n_g + n_t}$$

$$p_c = \frac{n_c}{n_a + n_c + n_g + n_t}$$

$$p_g = \frac{n_g}{n_a + n_c + n_g + n_t}$$

$$\& \; p_t = \frac{n_t}{n_a + n_c + n_g + n_t}$$

Ques2. (Naive Bayes Classifier) Consider the following set:

A	B	C	Y
0	1	1	0
1	1	1	0
0	0	0	0
1	1	0	1
0	1	0	1
1	0	1	1

(A) Learn a Naive Bayes classifier by estimating all necessary probabilities.

(B) Compute the probability $P(y=1 \mid A=1, B=0, C=0)$

(C) Suppose we know that A, B, C are independent random variables, can we say that the Naive Bayes assumption is valid?

Ans2. (A) The estimation of all necessary probabilities is as follows :—

$P(y=1) = \frac{3}{6} = \frac{1}{2}$

$P(A=0|y=1) = \frac{1}{3}$

$P(B=0|y=1) = \frac{1}{3}$

$P(C=0|y=1) = \frac{2}{3}$

$P(A=0|y=0) = \frac{2}{3}$

$P(B=0|y=0) = \frac{1}{3}$

$P(C=0|y=0) = \frac{1}{3}$

$P(y=0) = 1 - P(y=1) = \frac{1}{2}$

$P(A=1|y=1) = 1 - P(A=0|y=1) = \frac{2}{3}$

$P(B=1|y=1) = 1 - P(B=0|y=1) = \frac{2}{3}$

$P(C=1|y=1) = 1 - P(C=0|y=1) = \frac{1}{3}$

$P(A=1|y=0) = 1 - P(A=0|y=0) = \frac{1}{3}$

$P(B=1|y=0) = 1 - P(B=0|y=0) = \frac{2}{3}$

$P(C=1|y=D) = 1 - P(C=0|y=0) = \frac{2}{3}$

2(B). To solve: $P(y=1 \mid A=1, B=0, C=0)$

Sol^n: Using Naive Bayes Theorem:

$$P(y=1 \mid A=1, B=0, C=0) = \frac{P(A=1, B=0, C=0 \mid y=1)\, P(y=1)}{P(A=1, B=0, C=0)}$$

Using Naive Bayes conditional independence assumption:-

$$P(y=1 \mid A=1, B=0, C=0) = \frac{P(A=1 \mid y=1)\, P(B=0 \mid y=1)\, P(C=0 \mid y=1)\, P(y=1)}{P(A=1 \mid y=1)\, P(B=0 \mid y=1)\, P(C=0 \mid y=1)\, P(y=1) + P(y=0)\, P(A=1 \mid y=0)\, P(B=0 \mid y=0)\, P(C=0 \mid y=0)}$$

$$= \frac{\frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{1}{2}}{\frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3}} = \frac{\frac{4}{54}}{\frac{5}{54}} = \frac{4}{5}$$

$$\Rightarrow \boxed{P(y=1 \mid A=1, B=0, C=0) = 0.8} \quad \text{Ans.}$$

2(C). No, if A, B and C are independent random variables then the Naive Bayes assumption is not valid.
This is because, Conditional Independence $\neq$ Independence
i.e. $p(A,B,C) = p(A)\,p(B)\,p(C)$ doesnot imply that $p(A,B,C \mid y) = p(A \mid y)\,p(B \mid y)\,p(C \mid y)$

eg. If we have 2 variables A & B representing Rock, paper, scissor for 2 opponents & C is the outcome of the game then $P(A \& B \mid C) \neq P(A \& B)$

Ques3. (Maximum A Posteriori) As discussed in class, consider using a beta prior Beta(2,2) for estimating p, the probability of head for a weighted coin. What is the posterior distribution of p after we observe 5 coin tosses & 2 of them are head? What is the posterior distribution of p after we observe 50 coin tosses & 20 of them are head? Plot the pdf function of these two posterior distributions. Assume that $p = 0.4$ is the true probability, as we observe more & more coin tosses from this coin, what do you expect to happen to posterior?

Ans3. Let us consider the prior probability of getting a head be $P(\theta)$.

A.T.Q. prior probability follows a beta distribution

$$\therefore \quad P(\theta) \sim Beta(B_H, B_T)$$

where, B_H = no. of hallucinated heads

B_T = no. of hallucinated tails

$$\therefore P(\theta) = \frac{\theta^{B_H - 1}(1-\theta)^{B_T - 1}}{B(B_H, B_T)}$$

A.T.Q. $P(\theta) \sim Beta(2,2)$

$\therefore$ $B_H = 2$ & $B_T = 2$

$$\therefore P(\theta) = \frac{\theta(1-\theta)}{B(2,2)}$$

Now, $P(\theta|D) = \dfrac{P(D|\theta)\,P(\theta)}{P(D)}$

D ↳ Data

$$\cancel{\propto \frac{P(D|\theta)\,P(\theta)}{1}}$$

$$\Rightarrow \quad P(\theta|D) \propto P(D|\theta)\,P(\theta)$$

Now, $P(D|\theta) = \theta^{\alpha_H}(1-\theta)^{\alpha_T}$

$\forall$ α_H = No. of observed heads

α_T = No. of observed tails.

The Posterior Distribution of θ also follows Beta distribut

$\therefore$ $P(\theta|D) \sim \text{Beta}(\alpha_H + \beta_H, \ \alpha_T + \beta_T)$

$\therefore$ In case 1, where we have 5 coin tosses with 2 being heads,

$\alpha_H = 2 \quad \alpha_T = 3$

$\beta_H = 2 \quad \beta_T = 2$

$\therefore$ $P(\theta|D) \sim \text{Beta}(2+2, 3+2)$

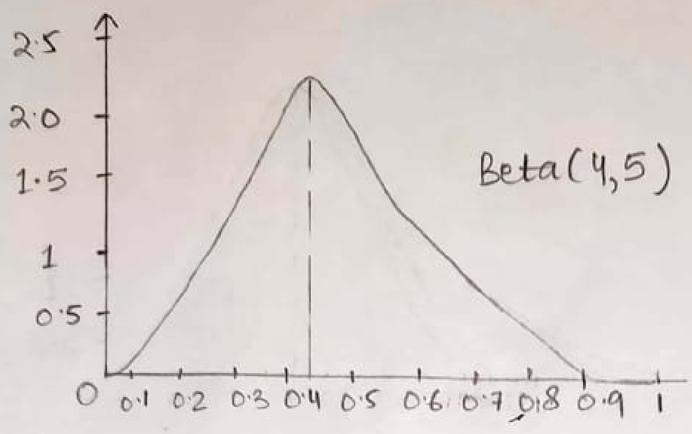
$\sim \text{Beta}(4, 5)$

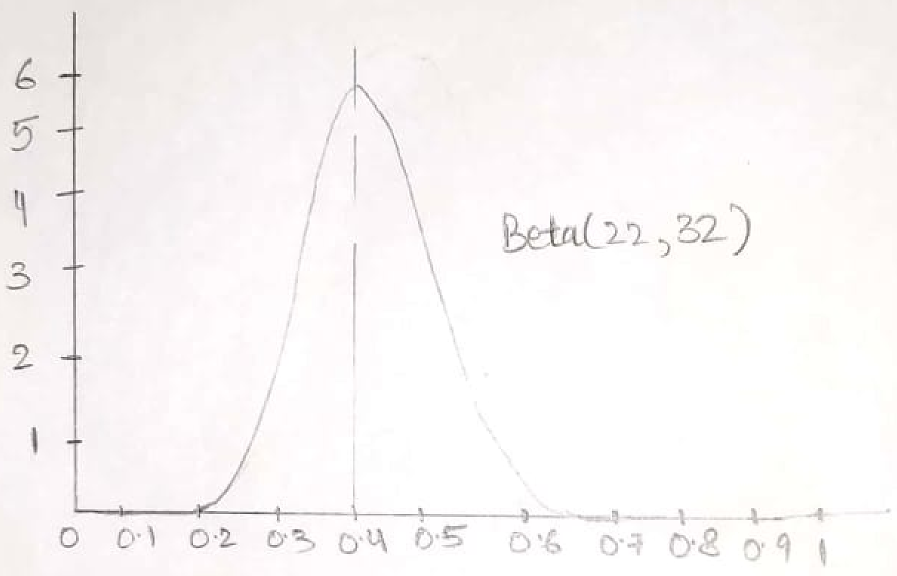
In case 2, where we have 50 coin tosses with 20 being heads.

$\alpha_H = 20 \quad \alpha_T = 30$

$\beta_H = 2 \quad \beta_T = 2$

$\therefore$ $P(\theta|D) \sim \text{Beta}(20+2, 30+2) \sim \text{Beta}(22, 32)$





$\Rightarrow$ We can observe that, if we increase the no. of coin tosses, then the probability of getting a head becomes more & more nearer to the true probability i.e. 0.4.

Thus, as the no. of coin tosses become a large no. such that $\alpha_H \gg \beta_H$ & $\alpha_T \gg \beta_T$, then $P(\theta|D) \approx 0.4$.

This can also be observed by the plots of Beta(4,5) & Beta(22,32).

Ques 4. (Perceptron) The perceptron algorithm will only converge if the data is linearly separable. It is possible to force your data to be linearly separable as follows: If you have N data points in D dimensions, map data point $\vec{x}_n$ to the (D+N) dimensional point $\langle \vec{x}_n, e_n \rangle$ where e_n is a N-dimensional vector of all zeros but one 1 at the n^{th} position.

(A) Show that if you apply this mapping the data becomes linearly separable.

(B) How does this mapping affect generalisation?

Ans 4. Given: N data points, each having D dimensions.

$\Rightarrow$ We have a vector X, representing all the data points of the form N×D.

$$X = \underbrace{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}_{D} \Big\} \text{N - Data points}$$

each point having d dimensions.

It is given that X is linearly inseparable.

Now, we map each point in $\vec{X}$ such that each point $\vec{x}_n$ gets mapped into point $\langle \vec{x}_n, e_n \rangle$ where $\langle \vec{x}_n, e_n \rangle$ is a point in (D+N) dimensions.

Thus, after adding N dimensions to the data,

$$\vec{X} = \begin{bmatrix} 1 & x_1 & x_2 & x_3 & \cdots & x_d & 1 & 0 & 0 & \cdots & 0 \\ 1 & x_1 & x_2 & x_3 & \cdots & x_d & 0 & 1 & 0 & \cdots & 0 \\ 1 & & & & & & & & & & \\ \vdots & & & & & & & & & & \\ 1 & x_1 & x_2 & x_3 & \cdots & x_d & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \updownarrow N \text{ point}$$

$$\longleftarrow d+n+1 \text{ dimensions} \longrightarrow$$

let L_p be the perceptron loss

$$\therefore \quad L_p(g(\omega,x),y) = \max(0, -y\omega^T x)$$

$\therefore$ As shown in the matrix x above, the $\vec{X}$ is now of dimension $N \times (D+N+1)$ where each data point is of $D+N+1$ dimensions.

If we break the $\vec{X}$ into two classes of positive & negative examples A & B respectively then for each point $x_i \in A$ there must exist a weight vector w_i such that $\sum_{i=0}^{n} w_i x_i > k$ & for each point $x_j \in B$, there must exist some weight vector w_j such that $\sum_{j=0}^{m} w_j x_j < -k$ where k is some constant.

Ans 4 (B) This mapping will affect generalisation such that it will not allow the method to be generalized. for each data point, we need to uniquely address the weight to make it linearly separable. And this mapping also depends on the size of e_n.

Ques 5. (Kernels) In class, we showed that the quadratic Kernel $K(x_i, x_j) = (x_i \cdot x_j + 1)^2$ was equivalent to mapping each $x = (x_1, x_2) \in R^2$ into a higher dimensional space, where

$$\phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1)$$

Now consider the cubic Kernel $K(x_i, x_j) = (x_i \cdot x_j + 1)^3$. What is corresponding ϕ function?

Ans 5. Given $K(x_i, x_j) = (x_i \cdot x_j + 1)^3$

Using $(a+b)^3 = a^3 + b^3 + 3a^2b + 3ab^2$

$$(x_i \cdot x_j + 1)^3 = (x_i \cdot x_j)^3 + 1 + 3(x_i \cdot x_j)^2 + 3(x_i \cdot x_j)$$

Now, each vector x consists of x_1 & x_2

$$\cancel{\Rightarrow (x_i \cdot x_j + 1)^3 = (x_{i1} \cdot x_{j1})^3 + (x_{i2}\, x_{j2})^3 + 1 +}$$

$$\Rightarrow (x_i \cdot x_j + 1)^3 = (x_{i1}x_{j1} + x_{i2}x_{j2})^3 + 1 + 3(x_{i1}x_{j1} + x_{i2}x_{j2})^2 + 3(x_{i1}x_{j1} + x_{i2}x_{j2}) + 1$$

$$= (x_{i1}x_{j1})^3 + (x_{i2}x_{j2})^3 + 3(x_{i1}x_{j1})^2(x_{i2}x_{j2}) + 3(x_{i1}x_{j1})(x_{i2}x_{j2})^2 + 3(x_{i1}x_{j1})^2 + 3(x_{i2}x_{j2})^2 + 6x_{i1}x_{j1}x_{i2}x_{j2} + 3(x_{i1}x_{j1}) + 3(x_{i2}x_{j2}) + 1$$

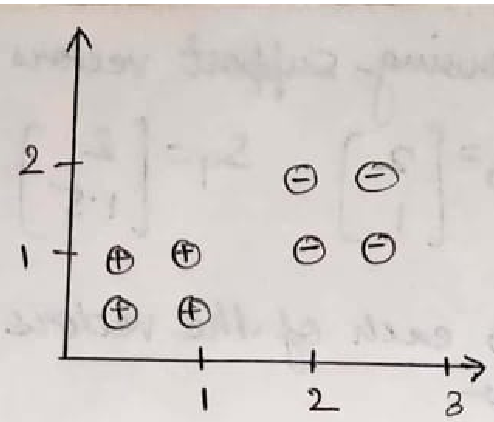
From the above eq^n, we can observe that the cubic kernel of $K(x_i, x_j) = (x_i \cdot x_j + 1)^3$ can be represented as a dot product of vectors $\phi(x_i)$ & $\phi(x_j)$ such that —

$$\underbrace{\begin{bmatrix} x_{i1}^3 \\ x_{i2}^3 \\ \sqrt{3}\, x_{i1}^2 x_{i2} \\ \sqrt{3}\, x_{i1} x_{i2}^2 \\ \sqrt{3}\, x_{i1}^2 \\ \sqrt{3}\, x_{i2}^2 \\ \sqrt{6}\, x_{i1} x_{i2} \\ \sqrt{3}\, x_{i1} \\ \sqrt{3}\, x_{i2} \\ 1 \end{bmatrix}}_{\phi(x_i)} \cdot \underbrace{\begin{bmatrix} x_{j1}^3 \\ x_{j2}^3 \\ \sqrt{3}\, x_{j1}^2 x_{j2} \\ \sqrt{3}\, x_{j1} \cdot x_{j2}^2 \\ \sqrt{3}\, x_{j1}^2 \\ \sqrt{3}\, x_{j2}^2 \\ \sqrt{6}\, x_{j1} x_{j2} \\ \sqrt{3}\, x_{j1} \\ \sqrt{3}\, x_{j2} \\ 1 \end{bmatrix}}_{\phi(x_j)}$$

Thus, $\boxed{\phi(x) = [x_1^3, x_2^3, \sqrt{3}x_1^2x_2, \sqrt{3}x_1x_2^2, \sqrt{3}x_1^2, \sqrt{3}x_2^2, \sqrt{6}x_1x_2, \sqrt{3}x_1, \sqrt{3}x_2, 1]}$

Ans

Ques6.

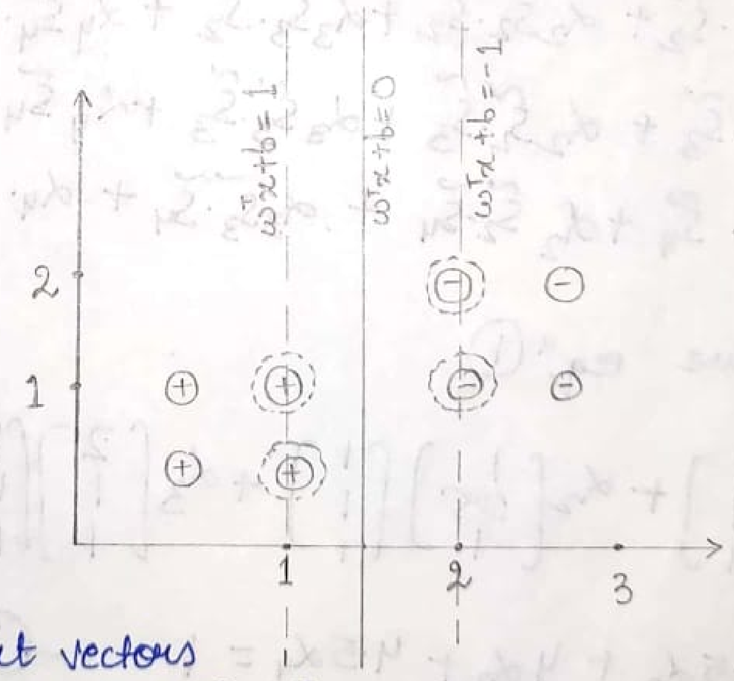


(Linear SVM) Apply linear SVM without soft margin to the following problem. Note that the two rightmost positive points are $(1, 0.5)$ & $(1, 1)$. The two leftmost negative points are $(2, 1)$ & $(2, 1.5)$.

(A). Please markout the support vectors, the decision boundary ($w^T x + b = 0$) & ($w^T x + b = 1$) & ($w^T x + b = -1$).

(B) Please solve for w & b based on the support vectors you identified in (a).

Ans6. (A)



We have 4 support vectors

$S_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $S_2 = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$ $S_3 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ $S_4 = \begin{bmatrix} 2 \\ 1.5 \end{bmatrix}$

Ans 6. (B) We have the following support vectors :-

$$S_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad S_2 = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} \quad S_3 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad S_4 = \begin{bmatrix} 2 \\ 1.5 \end{bmatrix}$$

We'll add a bias term in each of the vectors and use these augmented vectors-

$$\therefore \quad \tilde{S}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \tilde{S}_2 = \begin{bmatrix} 1 \\ 0.5 \\ 1 \end{bmatrix} \quad \tilde{S}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \quad \tilde{S}_4 = \begin{bmatrix} 2 \\ 1.5 \\ 1 \end{bmatrix}$$

let $\alpha_1, \alpha_2, \alpha_3$ & α_4 be the parameters for the system of linear equations formed by these augmented support vectors.

$$\therefore \quad \alpha_1 \tilde{S}_1 \cdot \tilde{S}_1 + \alpha_2 \tilde{S}_2 \cdot \tilde{S}_1 + \alpha_3 \tilde{S}_3 \cdot \tilde{S}_1 + \alpha_4 \tilde{S}_4 \cdot \tilde{S}_1 = 1 \quad - (1)$$

$$\alpha_1 \tilde{S}_1 \cdot \tilde{S}_2 + \alpha_2 \tilde{S}_2 \cdot \tilde{S}_2 + \alpha_3 \tilde{S}_3 \cdot \tilde{S}_2 + \alpha_4 \tilde{S}_4 \cdot \tilde{S}_2 = 1 \quad - (2)$$

$$\alpha_1 \tilde{S}_1 \cdot \tilde{S}_3 + \alpha_2 \tilde{S}_2 \cdot \tilde{S}_3 + \alpha_3 \tilde{S}_3 \cdot \tilde{S}_3 + \alpha_4 \tilde{S}_4 \cdot \tilde{S}_3 = -1 \quad - (3)$$

$$\alpha_1 \tilde{S}_1 \cdot \tilde{S}_4 + \alpha_2 \tilde{S}_2 \cdot \tilde{S}_4 + \alpha_3 \tilde{S}_3 \cdot \tilde{S}_4 + \alpha_4 \cdot \tilde{S}_4 \cdot \tilde{S}_4 = -1 \quad - (4)$$

Lets solve eq^n (1)

$$\alpha_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 0.5 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \alpha_3 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \alpha_4 \begin{bmatrix} 2 \\ 1.5 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 1$$

$$\Rightarrow 3\alpha_1 + 2.5\alpha_2 + 4\alpha_3 + 4.5\alpha_4 = 1 \quad - (i)$$

Lets solve eq^n (2)

for eq^n ②

$$\alpha_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0.5 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 0.5 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0.5 \\ 1 \end{bmatrix} + \alpha_3 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0.5 \\ 1 \end{bmatrix} + \alpha_4 \begin{bmatrix} 2 \\ 1.5 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0.5 \\ 1 \end{bmatrix} = 1$$

$$2.5\alpha_1 + 2.25\alpha_2 + 3.5\alpha_3 + 3.75\alpha_4 = 1 \quad \text{—(ii)}$$

for eq^n ③

$$\alpha_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 0.5 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \alpha_3 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \alpha_4 \begin{bmatrix} 2 \\ 1.5 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = -1$$

$$4\alpha_1 + 3.5\alpha_2 + 6\alpha_3 + 6.5\alpha_4 = -1 \quad \text{—(iii)}$$

for eq^n ④

$$\alpha_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1.5 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 0.5 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1.5 \\ 1 \end{bmatrix} + \alpha_3 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1.5 \\ 1 \end{bmatrix} + \alpha_4 \begin{bmatrix} 2 \\ 1.5 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1.5 \\ 1 \end{bmatrix} = -1$$

$$4.5\alpha_1 + 3.75\alpha_2 + 6.5\alpha_3 + 7.25\alpha_4 = -1 \quad \text{—(iv)}$$

Therefore, we have 4 eq^{n}s of parameters —

$$3\alpha_1 + 2.5\alpha_2 + 4\alpha_3 + 4.5\alpha_4 = 1$$
$$2.5\alpha_1 + 2.25\alpha_2 + 3.5\alpha_3 + 3.75\alpha_4 = 1$$
$$4\alpha_1 + 3.5\alpha_2 + 6\alpha_3 + 6.5\alpha_4 = -1$$
$$4.5\alpha_1 + 3.75\alpha_2 + 6.5\alpha_3 + 7.25\alpha_4 = -1$$

$\Rightarrow$ In Matrix form, the system of linear eq^n's can be represented as :-

$$\begin{bmatrix} 3 & 2.5 & 4 & 4.5 \\ 2.5 & 2.25 & 3.5 & 3.75 \\ 4 & 3.5 & 6 & 6.5 \\ 4.5 & 3.75 & 6.5 & 7.25 \end{bmatrix}_{4\times4} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}_{4\times1} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}_{4\times1}$$

This is of the form $AX = B$

$$\Rightarrow X = A^{-1} B$$

On solving –

$$\therefore \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = \begin{bmatrix} 2.125 \\ 6.25 \\ -5.25 \\ 0.25 \end{bmatrix}$$

$\therefore$ $\alpha_1 = 2.125$; $\alpha_2 = 6.25$; $\alpha_3 = -5.25$; $\alpha_4 = 0.25$

Now, lets calculate the weight matrix $\tilde{\omega}$

$$\tilde{\omega} = \sum_{i=1}^{4} \alpha_i S_i = 2.125 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 6.25 \begin{bmatrix} 1 \\ 0.5 \\ 1 \end{bmatrix} - 5.25 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + 0.25 \begin{bmatrix} 2 \\ 1.5 \\ 1 \end{bmatrix}$$

$$\Rightarrow \tilde{\omega} = \begin{bmatrix} 2.125 \\ 2.125 \\ 2.125 \end{bmatrix} + \begin{bmatrix} 6.25 \\ 3.125 \\ 6.25 \end{bmatrix} + \begin{bmatrix} -10.5 \\ -5.25 \\ -5.25 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0.375 \\ 0.25 \end{bmatrix}$$

$$\Rightarrow \tilde{\omega} = \begin{bmatrix} -1.625 \\ 0.375 \\ 3.375 \end{bmatrix}$$

; This $\tilde{\omega}$ is the augmented weight matrix that contains the bias b.

$\therefore$ $\boxed{\omega = \begin{bmatrix} -1.625 \\ 0.375 \end{bmatrix}}$ $\boxed{b = 3.375}$ Ans.