Machine Learning

Name: V Deekshitha
Rollno: 208W1A1259

Home Assignment -2

2.

S.NO	y	x
1	14	24
2	16	23
3	18	21
4	12	21
5	13	16
6	8	18
7	6	22
8	2	17

Find univariate regression equation in the format $y = w * x + b$ based on the above data

x	y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
24	14	3.75	2.87	10.76	14.06
23	16	2.75	4.88	13.42	7.56
21	18	0.75	6.88	5.16	0.56
21	12	0.75	0.88	0.66	0.56
16	13	-4.25	1.88	-7.99	18.06
18	8	-2.25	-3.12	7.02	5.06
22	6	1.75	-5.12	-8.96	3.06
17	2	-3.25	-9.12	29.64	10.56
20.25	11.12			Σ 49.71	Σ 59.48

$$w = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{49.71}{59.48}$$

$$= 0.835$$

$$b = y - w * x$$

$$= 11.12 - (0.835)(20.25)$$

$$b = -5.788$$

$$y = 0.835 * x - 5.788$$

5.

S.NO	y	x_1	x_2
1	14	24	12
2	16	23	8
3	18	21	14
4	12	21	8
5	13	16	3
6	8	18	5
7	6	22	7
8	2	17	9

Find multivariate regression equation in the format $y = w_0 + w_1 * x_1 + w_2 * x_2$ based on the above data.

$$b_1 = \frac{[(\Sigma x_2^2)(\Sigma x_1 y) - (\Sigma x_1 x_2)(\Sigma x_2 y)]}{[(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2]}$$

$$b_2 = \frac{[(\Sigma x_1^2)(\Sigma x_2 y) - (\Sigma x_1 x_2)(\Sigma x_1 y)]}{[(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2]}$$

$$b_0 = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2$$

	y	x_1	x_2	x_1^2	x_2^2	$x_1 y$	$x_2 y$	$x_1 x_2$
	14	24	12	576	144	336	168	288
	16	23	8	529	64	368	128	184
	18	21	14	441	196	378	252	294
	12	21	8	441	64	252	96	168
	13	16	3	256	9	208	39	48
	8	18	5	324	25	144	40	90
	6	22	7	484	49	132	42	154
	2	17	9	289	81	34	18	153
Sum	89	162	66	3340	632	1852	783	1379
Mean	11.125	20.25	8.25					

$$\Sigma x_1^2 = \Sigma x_1^2 - (\Sigma x_1)^2/n$$

$$= 3340 - \frac{(162)^2}{8}$$

$$= 59.5$$

$$\Sigma x_2^2 = \Sigma x_2^2 - (\Sigma x_2)^2/n$$

$$= 632 - \frac{(66)^2}{8}$$

$$= 87.5$$

$$\Sigma x_1 y = \Sigma x_1 y - (\Sigma x_1 \Sigma y)/n$$

$$= 1852 - \frac{(162)(89)}{8}$$

$$= 49.75$$

$$\Sigma x_2 y = \Sigma x_2 y - (\Sigma x_2 \Sigma y)/n$$

$$= 783 - \frac{(66)(89)}{8}$$

$$= 48.75$$

$$\Sigma x_1 x_2 = \Sigma x_1 x_2 - \frac{(\Sigma x_1 \Sigma x_2)}{n}$$

$$= 1379 - \frac{(162)(66)}{8}$$

$$= 42.5$$

$$b_1 = \frac{(87.5)(49.75) - (42.5)(48.75)}{(59.5)(87.5) - (42.5)^2}$$

$$= \frac{2281.25}{3400}$$

$$= 0.6709$$

$$b_2 = \frac{(59.5)(48.75) - (42.5)(49.75)}{(59.5)(87.5) - (42.5)^2}$$

$$= \frac{786.25}{3400}$$

$$= 0.2312$$

$$b_0 = 11.125 - (0.6709)(20.25) - (0.2312)(8.25)$$

$$= -4.3681$$

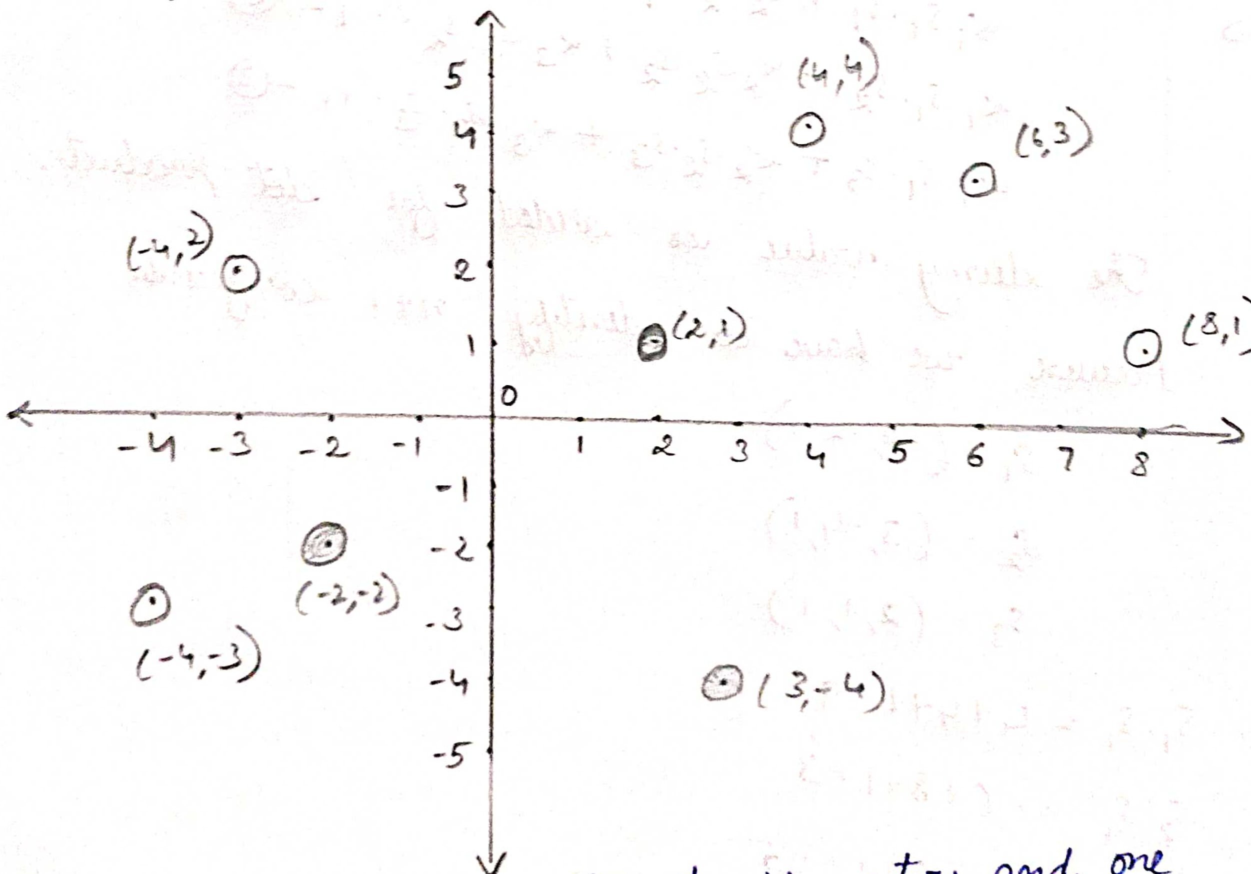
$$y = b_0 + b_1 x_1 + b_2 x_2$$

$$y = -4.3681 + (0.6709) x_1 + (0.2312) x_2$$

7.

S.No	Attribute 1	Attribute 2	Class
1	4	4	+
2	6	3	+
3	8	1	+
4	2	1	+
5	3	-4	−
6	-2	-2	−
7	-4	2	−
8	-4	-3	−

Identify the separating boundary of positive and negative samples using SVM algorithm.



3 support vectors - a Pair of -Ve vectors and one +Ve instance which are near ~~ch~~ combine comparid to others

2 → -Ve support vector

1 → +Ve support vector

The line equation is derived only by considering only. Support vectors remaining are negleted.

$S_1 = (-2, -2)$

$S_2 = (3, -4)$

$S_3 = (2, 1)$

Standard format of line equation $y = mx + c$

$$y = w_1 x_1 + w_2 x_2 + w_0$$

→ 3 co-efficients can be derived using 3 support vectors

→ $\alpha_1 S_1 \cdot S_1 + \alpha_2 S_2 \cdot S_1 + \alpha_3 S_3 \cdot S_1 = -1$ — ①

$\alpha_1 S_1 \cdot S_2 + \alpha_2 S_2 \cdot S_2 + \alpha_3 S_3 \cdot S_2 = -1$ — ②

$\alpha_1 S_1 \cdot S_3 + \alpha_2 S_2 \cdot S_3 + \alpha_3 S_3 \cdot S_3 = +1$ — ③

The dumy value is added for dot product, because we have to identify $n+1$ coefficients

$S_1 = (-2, -2, 1)$

$S_2 = (3, -4, 1)$

$S_3 = (2, 1, 1)$

$S_1 \cdot S_1 = 4 + 4 + 1 = 9$

$S_2 \cdot S_1 = -6 + 8 + 1 = 3$

$S_3 \cdot S_1 = -4 - 2 + 1 = -5$

$S_2 \cdot S_2 = 9 + 16 + 1 = 26$

$S_3 \cdot S_2 = 6-4+1 = 3$

$S_3 \cdot S_3 = 4+1+1 = 6$

$$9\alpha_1 + 3\alpha_2 - 5\alpha_3 = -1 \quad \text{--- (1)}$$

$$3\alpha_1 + 26\alpha_2 + 3\alpha_3 = -1 \quad \text{--- (2)}$$

$$-5\alpha_1 + 3\alpha_2 + 6\alpha_3 = +1 \quad \text{--- (3)}$$

By solving the eqn's we get

$$\alpha_1 = -0.0472$$

$$\alpha_2 = 0.0718$$

$$\alpha_3 = -0.2419$$

$$\bar{w} = \sum_i \alpha_i S_i$$

$$= -0.0472\begin{pmatrix} -2 \\ -2 \\ 1 \end{pmatrix} + 0.0718\begin{pmatrix} 3 \\ -4 \\ 1 \end{pmatrix} + (-0.2419)\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0.094 + 0.2154 - 0.4838 \\ 0.094 - 0.2872 - 0.2419 \\ -0.0472 + 0.0718 - 0.2419 \end{pmatrix}$$

$$\bar{w} = \begin{pmatrix} 0.48 \\ -0.43 \\ -0.21 \end{pmatrix}$$

8. Consider the training examples shown in the following table for a binary classification. The table shows a training set for a problem of predicting whether a loan applicant will repay his / her loan obligation or defaulting on his / her loan.

Tid	Home Owner	Marital Status	Annual Income	Default Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Using the tree approach that we discussed in the class, predict the class label for this test example, X = (Home Owner = No, Marital Status = Married, Income = $120K).

$$Entropy(S) = \sum_{i=1}^{c} -P_i \log_2 P_i$$

$$= \frac{-7}{10} \log_2\left(\frac{7}{10}\right) - \frac{3}{10} \log_2\left(\frac{3}{10}\right)$$

$$= 0.3602 + 0.5211$$

$$= 0.8813$$

$$Entropy_{HO}(S) = \frac{3}{10}[I(0,3)] + \frac{7}{10}[I(3,4)]$$

$$= \frac{7}{10}\left[\frac{-3}{7} \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \log_2\left(\frac{4}{7}\right)\right]$$

$$= \frac{7}{10}[+0.5239 + 0.4613]$$

$$= 0.6896$$

$$Gain(S, A) = 0.1917$$

$$Entropy_{MS}(S) = \frac{4}{10}[I(2,2)] + \frac{4}{10}[I(0,4)] + \frac{2}{10}[I(1,1)]$$

$$= \frac{4}{10}\left[\log_2\left(\frac{2}{4}\right)\right] + \frac{2}{10}\left[\log_2\left(\frac{1}{2}\right)\right]$$

$$= 0.4 + 0.2$$

$$= 0.6$$

$$Gain(S, A) = 0.2813$$

$$Entropy_{AI}(S) = \frac{7}{10}[I(1,6)] + \frac{3}{10}[I(0,3)]$$

$$= \frac{7}{10}\left[\frac{-1}{7} \log_2\left(\frac{1}{7}\right) - \frac{6}{7} \log_2\left(\frac{6}{7}\right)\right]$$

$$= \frac{7}{10}[0.4010 + 0.1906]$$

$$= 0.414$$

$$Gain(S, A) = 0.4673$$

Annual Income

≤ 100

HO	MS	DB
N	M	N
N	S	N
N	D	y
N	M	N
N	S	y
N	M	N
N	S	y

> 100

HO	MS	DB
y	S	N
y	M	N
y	D	N

Default Borrower No.

$$\text{Entropy}(S) = \frac{-4}{7}\log_2\left(\frac{4}{7}\right) - \frac{3}{7}\log_2\left(\frac{3}{7}\right)$$

$$= +0.461 + 0.5239$$

$$= 0.9849$$

$$\text{Entropy}_{HO}(S) = \frac{7}{7}\left[I(4,3)\right]$$

$$= -\frac{4}{7}\log_2\left(\frac{4}{7}\right) - \frac{3}{7}\log_2\left(\frac{3}{7}\right)$$

$$= 0.4613 + 0.5239$$

$$= 0.9852$$

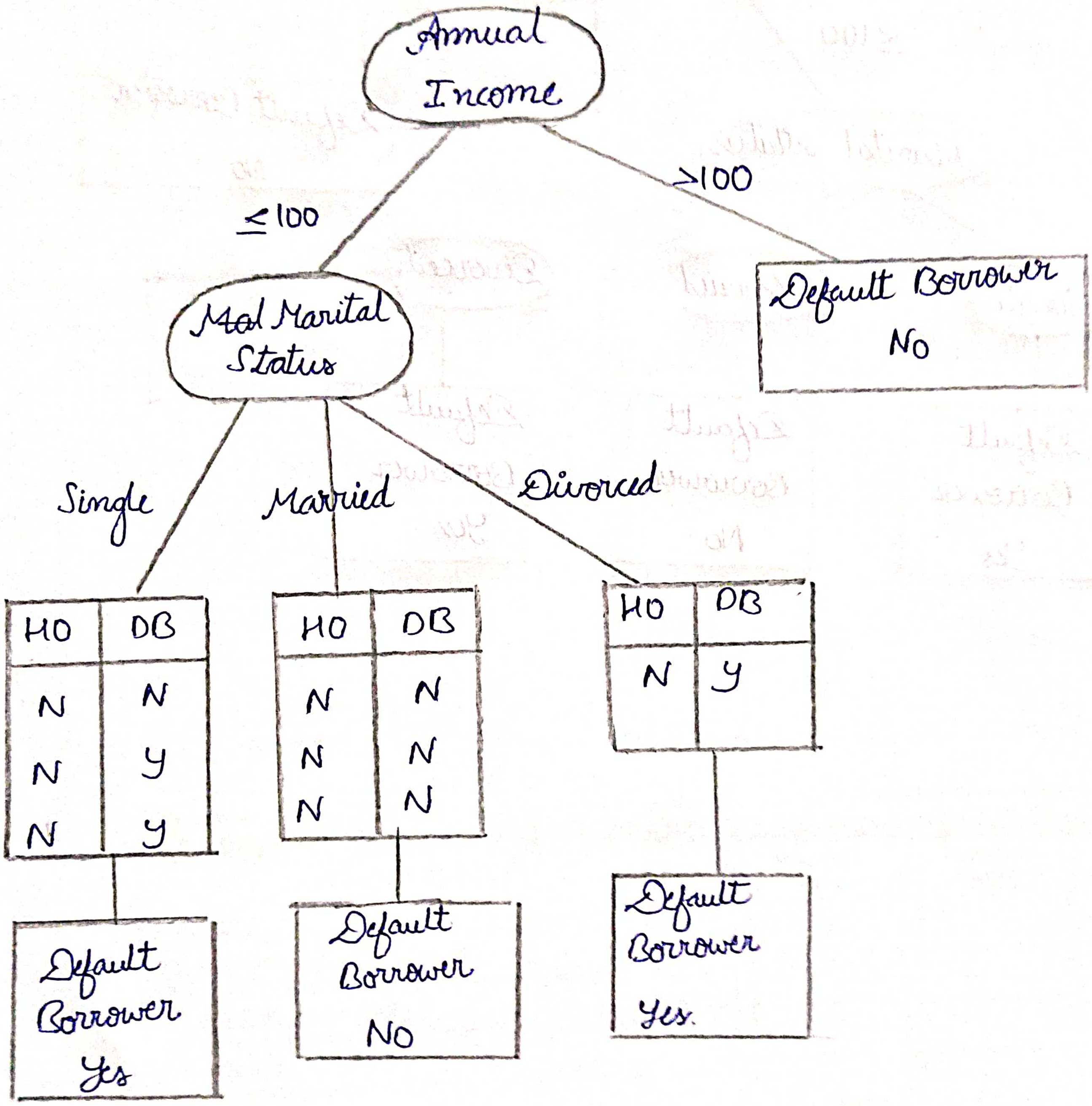
$$\text{Entropy}_{MS}(S) = \frac{3}{7}\left[I(0,3)\right] + \frac{1}{7}\left[I(1,0)\right] + \frac{3}{7}\left[I(2,1)\right]$$

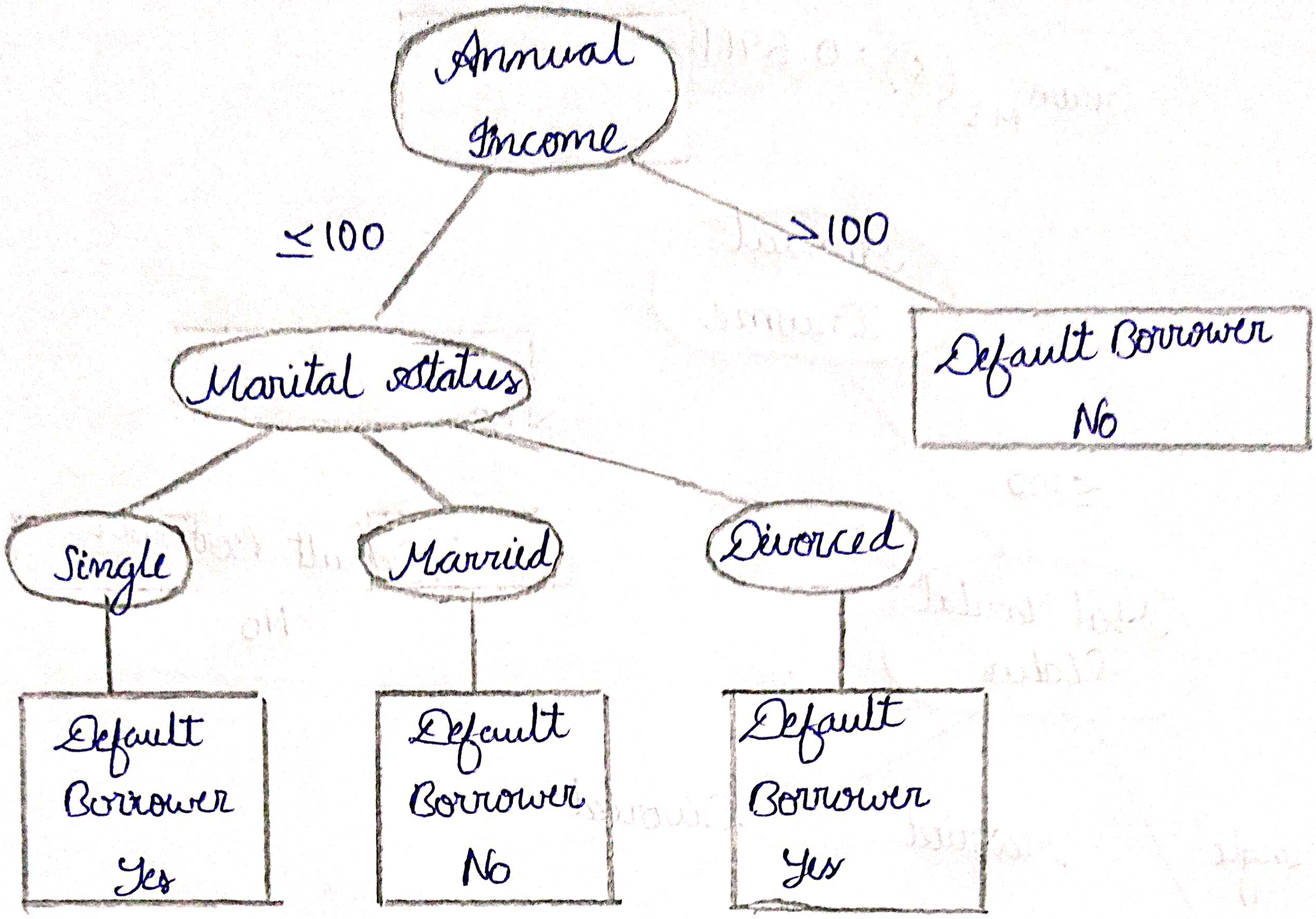
$$= \frac{3}{7}\left[\frac{-2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right]$$

$$= \frac{3}{7}\left[0.39 + 0.5288\right]$$

$= 0.39377$

$Gain_{MS}(S) = 0.5911$




Annual Income
≤100
>100
Default Borrower No
Marital Status
Single
Married
Divorced
Default Borrower Yes
Default Borrower No
Default Borrower Yes