MACHINE - LEARNING

HOME - ASSIGNMENT - II

CODE : 20IT6302

Submitted By :

2081A1220

IT - B

Submitted To :

Dr. T. Anuradha

Department of IT

MACHINE - LEARNING

HOME - ASSIGNMENT - 2

M. J. N. V. Sai
2O8WIAI2AO
IT - B

(3) Find Out the Univariate Regression in the format of $Y = w * x + b$ Based on the Given Data.

	Hours	Scores
0	2.5	21
1	5.1	47
2	3.2	27
3	8.5	75
4	3.5	30

General Formula is

$$Y = w * x + b$$

where x : hours
Y = Scores
w : Slope of Line
b : Y- intercept

$$w = \frac{(N \cdot \Sigma xy - \Sigma x \cdot \Sigma y)}{N \cdot \Sigma x^2 - (\Sigma x)^2}$$

$$b = \frac{\Sigma Y - w \cdot \Sigma x}{N}$$

$N = 5$ $\Sigma y = 200$

$\Sigma x = 23.8$ $\Sigma xy = 843.2$

$\Sigma x^2 = 105.67$

$$w = \frac{(5843.2 - 23.8200)}{5 * 105.67 - (23.8)^2}$$

$w = 9.7758$

$$b = \frac{(200 - 9.7758 * 23.8)}{5}$$

$b = -0.3581$

∴ final Univariate Regression is

$$\boxed{y = 9.7758 * x - 0.3581}$$

→ The above equation can be used to predict the scores based on the hours studied.

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 the multivariate Regression Equation in the format of

$Y = w0 + w1 * x1 + w2 * x2$ based on Above Data.

i) Calculate the Mean for Y, x_1, x_2

$\sum Y = (14 + 16 + 12 + 12 + 13 + 8 + 6 + 2)/8 = 11.125$

$\sum x_1 = (24 + 23 + 21 + 21 + 16 + 18 + 22 + 17)/8 = 19.5$

$\sum x_2 = (12 + 8 * 2 + 14 + 3 * 5 + 7 + 9)/8 = 8.125$

2) Calculate the Sum of Squares of Deviations for Y, x_1, x_2

$$SS(\text{Variable}) = \sum (x - \text{Mean}(\text{variable}))^2$$

$$SS(Y) = (14 - 11.125)^2 + (16 - 11.125)^2 + (18 - 11.125)^2 + (12 - 11.125)^2 + (13 - 11.125)^2 + (8 - 11.125)^2 + (6 - 11.125)^2 + (2 - 11.125)^2$$

$$= 137.375$$

$$SS(x_1) = (24 - 19.5)^2 + (23 - 19.5)^2 + (21 - 19.5)^2 + (21 - 19.5)^2 + (16 - 19.5)^2 + (18 - 19.5)^2 + (26 - 19.5)^2 + (17 - 19.5)^2$$

$$= 157.5$$

$$SS(x_2) = (12 - 8.125)^2 + (8 - 8.125)^2 + (14 - 8.125)^2 + (8 - 8.125)^2 + (3 - 8.125)^2 + (5 - 8.125)^2 + (7 - 8.125)^2 + (9 - 8.125)^2$$

$$= 50.875$$

3) Calculate the Sum of products of Deviations for x_1 & x_2.

$$SP(x_1, x_2) = (24 - 19.5) \cdot (16 - 8.125) + (23 - 19.5) \cdot (8 - 8.125)$$

$$+ (21 - 19.5) \cdot (14 - 8.125) + (21 - 19.5) \cdot (8 - 8.125) +$$

$$(16 - 19.5) \cdot (3 - 8.125) + (13 - 19.5) \cdot (5 - 8.125) +$$

$$(22 - 19.5) \cdot (7 - 8.125) + (17 - 19.5) \cdot (9 - 8.25)$$

$$= -50.25$$

4) Calculate the Regression coefficients w_1, w_2, w_0

$$w_1 = \frac{[SP(x_1, y) \cdot SS(x_2) - SP(x_2, y) \cdot SP(x_1, x_2)]}{SS(x_1) \cdot SS(x_2) - SP(x_1, x_2)^2}$$

$$= \frac{(304.5 \times 50.875) - (104.5 \times (-50.25))}{(157.5 * 50.875) - (-50.25)^2}$$

$$= 0.501$$

$$w_2 = \frac{[SP(x_2, y) \cdot SS(x_1) - SP(x_1, y) \cdot SP(x_1, x_2)]}{SS(x_1) \cdot SS(x_2) - SP(x_1, x_2)^2}$$

$$w_2 = \frac{(-104.5 * 157.5) - (304.5 * (-50.25))}{157.5 * 50.875 - (-50.25)^2}$$

$\therefore W_2 = -0$

$W_0 = \text{Mean}(Y) - W_1 * \text{Mean}(x_1) - W_2 * \text{Mean}(x_2)$

$= 11.125 - 0.501 * 19.5 - 0 * 8.125$

$\therefore W_0 = 1.123$

$\therefore$ The multivariate regression equation in format of

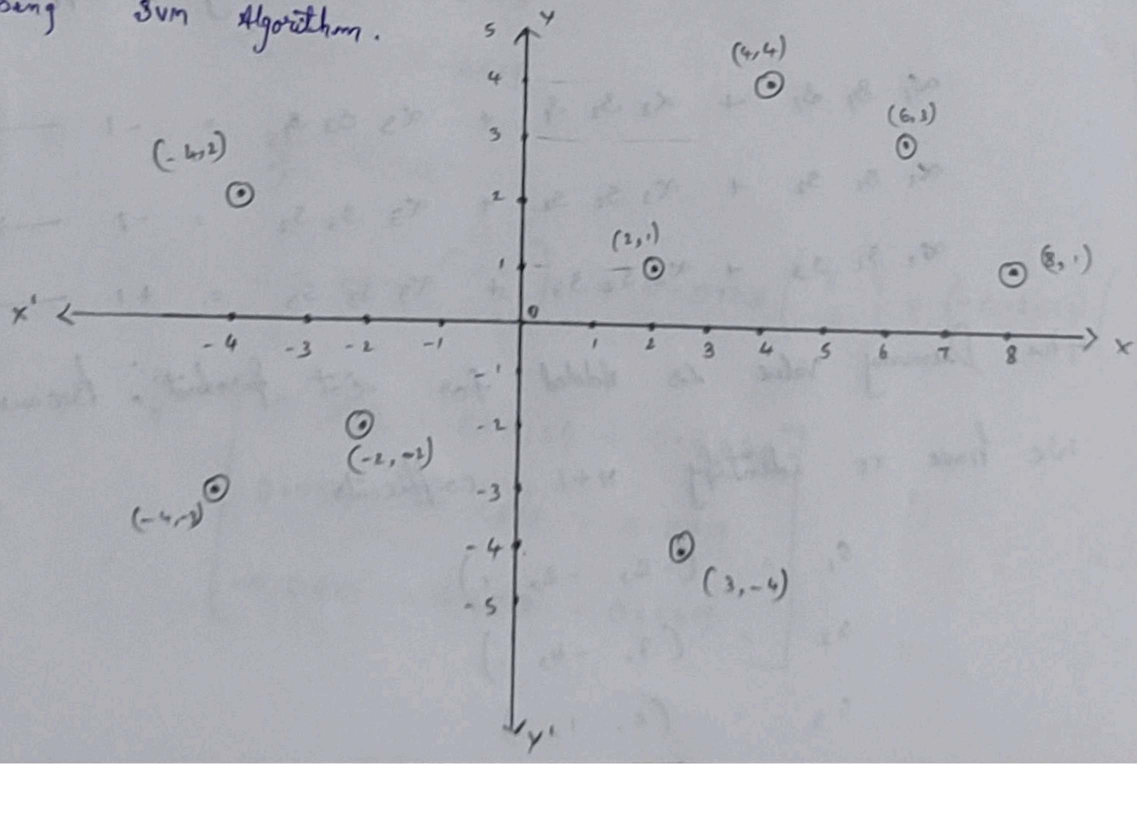
$Y = W_0 + W_1 \cdot x_1 + W_2 \cdot x_2$ is

$$\boxed{Y = 1.123 + 0.501 * x_1 + 0 * x_2}$$

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 seperating boundary of positive & negative samples using SVM Algorithm.



3 Support Vectors - A pair -ve vectors And One +ve instance which are near Comparid to other.

2 → -ve Support vector

1 → +ve Support vector

The Line Equation is Derived only by considering only. Support vector Remaining Are neglected.

$s_1 = (-2, -2)$

$s_2 = (3, -4)$

$s_3 = (2, 1)$

* Standard format of Line Equation $y = mx + c$

$$y = W_1 x_1 + W_2 x_2 + W_0$$

→ A 3 co-efficients can be Derived Using 3 Support Vectors.

$$\alpha_1 s_1 s_1 + \alpha_2 s_2 s_1 + \alpha_3 s_3 s_1 = -1 \longrightarrow (1)$$

$$\alpha_1 s_1 s_2 + \alpha_2 s_2 s_2 + \alpha_3 s_3 s_2 = -1 \longrightarrow (2)$$

$$\alpha_1 s_1 s_3 + \alpha_2 s_2 s_3 + \alpha_3 s_3 s_3 = +1 \longrightarrow (3)$$

The Dummy Value is Added for Dot product; Because We have to identify n+1 coefficents

$s_1 = (-2, -2, 1)$

$s_2 = (3, -4, 1)$

$s_3 = (2, 1, 1)$

$s_1 \cdot s_1 = 4+4+1 = 9$

$s_2 \cdot s_1 = -6+8+1 = 3$

$s_3 \cdot s_1 = -4-2+1 = -5$

$s_2 \cdot s_2 = 9+16+1 = 26$

$s_3 \cdot s_2 = 6-4+1 = 3$

$s_3 \cdot s_3 = 4+1+1 = 6$

$$9\alpha_1 + 3\alpha_2 - 5\alpha_3 = -1 \longrightarrow (1)$$

$$3\alpha_1 + 26\alpha_2 + 3\alpha_3 = -1 \longrightarrow (2)$$

$$-5\alpha_1 + 3\alpha_2 + 6\alpha_3 = +1 \longrightarrow (3)$$

By solving the equations we get

$\alpha_1 = -0.0472$

$\alpha_2 = 0.0718$

$\alpha_3 = -0.2419$

$$\overline{w} = \sum_i \alpha_i s_i$$

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

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

$$\therefore \quad \overline{W} = \begin{bmatrix} 0.48 \\ -0.43 \\ -0.21 \end{bmatrix}$$

(8) Consider the Training Samples 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 on Defaulting on his/her Loan.

Tid	Home Owner	Martial	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,

$X = [\text{Home Owner} = \text{NO}, \ \text{Martial Status} = \text{Married}, \ \text{Income} = 120K]$

$$\text{Entropy}(S) = \sum -P_i \log_2 \cdot P_i$$

$$= -\frac{7}{10} \cdot \log_2 \cdot \frac{7}{10} - \frac{3}{10} \log_2 \frac{3}{10}$$

$$= 0.3602 + 0.5211$$

$$= 0.8813$$

$$\text{Entropy}_{HO}(S) = \frac{3}{10}\left[I(0,3)\right] + \frac{7}{10}\left[I(3,4)\right]$$

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

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

$$= 0.6896$$

$$\text{Gain}(S, A) = 0.1917$$

$$\text{Entropy}_{MS}(S) = \frac{4}{10}\left[I(2,2)\right] + \frac{4}{10}\left[I(0,4)\right] + \frac{2}{10}\left[I(1,1)\right]$$

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

$$= 0.4 + 0.2$$

$$= 0.6$$

$$\text{Gain}(S, A) = 0.8813$$

$$\text{Entropy}_{A1}(S) = \frac{7}{10}\left[I(1,6)\right] + \frac{3}{10}\left[I(0,3)\right]$$

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

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

$$= 0.414$$

$$\text{Gain}(S, A) = 0.4673$$

Annual Income

≤ 100

HO	MS	OB
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

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

$$= 0.461 + 0.5239$$

$$= 0.9849$$

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

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

$$= 0.4613 + 0.5239$$

$$= 0.9852$$

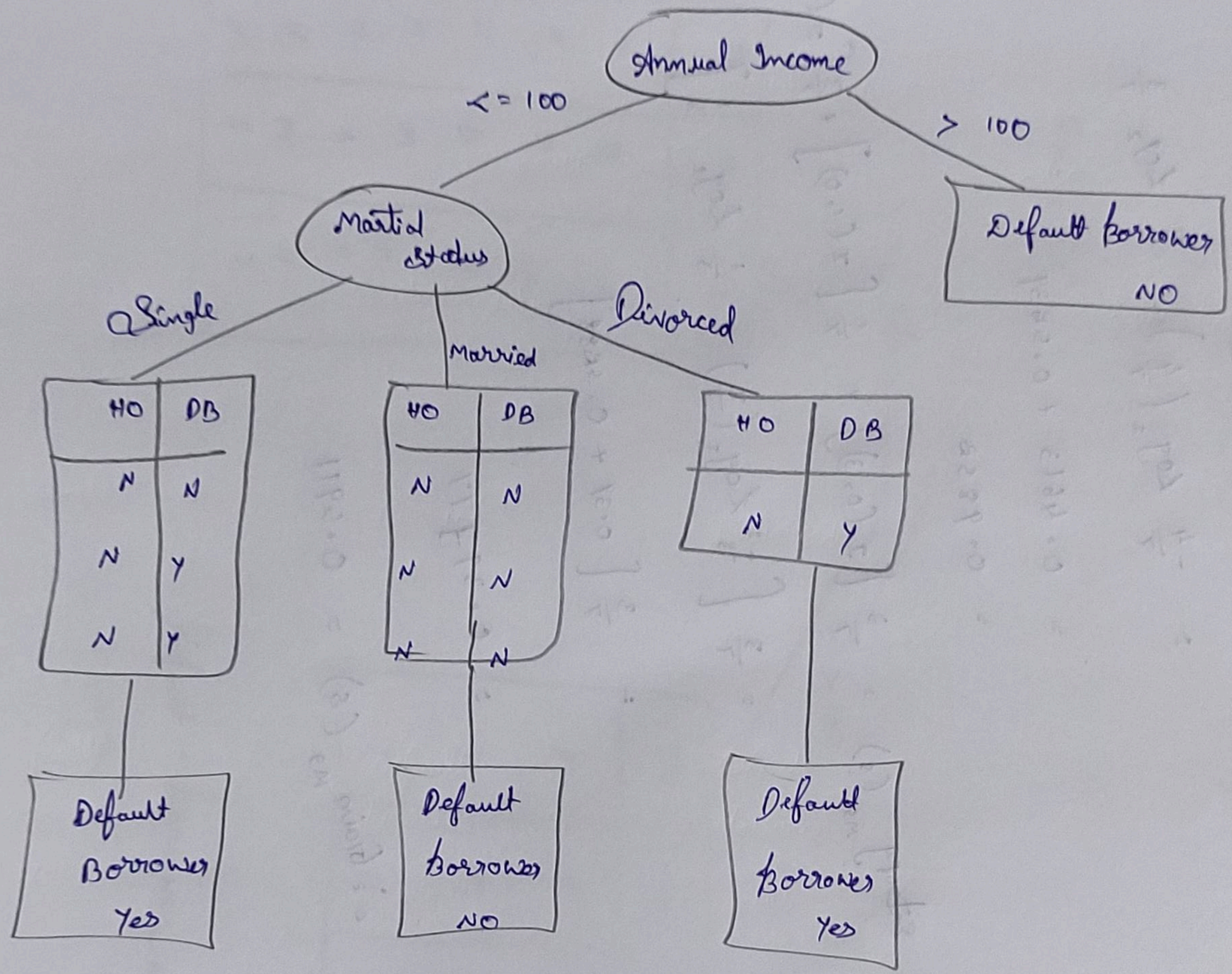
$$\text{Entropy}_{MS}(S) = \frac{3}{7}\left[I(0,3)\right] + \frac{1}{7}\left[I(1,0)\right] + \frac{3}{7}\left[I(2,1)\right]$$

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

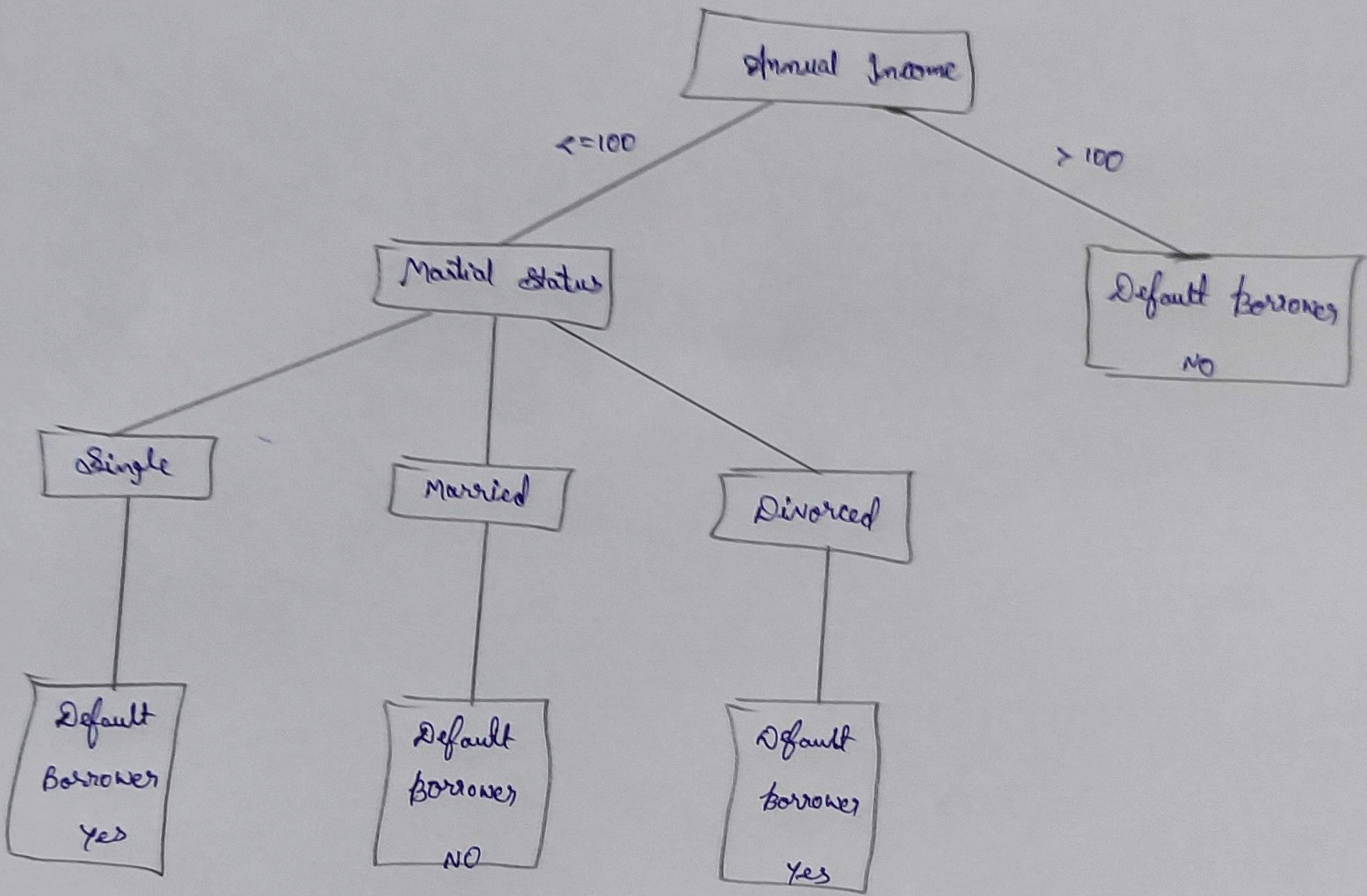
$$= \frac{3}{7}\left[0.39 + 0.5288\right]$$

$$= 0.39377$$

$$\therefore \text{Gain}_{MS}(S) = 0.5911$$



Annual Income
<= 100
> 100
Default Borrower NO
Martial Status
Single
Married
Divorced
HO DB
N N
N Y
N Y
HO DB
N N
N N
N N
HO DB
N Y
Default Borrower Yes
Default Borrower NO
Default Borrower Yes


Annual Income
<=100
> 100
Marital Status
Default Borrower
NO
Single
Married
Divorced
Default
Borrower
Yes
Default
Borrower
NO
Default
Borrower
Yes