Name : Binod Kumar
Class : M-Tech CSE
Roll no. : 23203006
Subject : Machine Learning
Assignment: 02

Q.1 Consider an artificial example of building a decision tree classification model to classify bank loan application by assigning applications to one of thee risk classes.

Owns home	Married	Gender	Employed	Credit-rating	Risk class
Yes	Yes	Male	yes	A	B
No	No	Female	Yes	A	A
Yes	yes	Female	yes	B	C
Yes	No	Male	No	B	B
No	yes	Female	yes	B	C
No	NO	Female	yes	B	A
No	NO	Male	No	B	B
Yes	NO	Female	Yes	A	A
No	Yes	Female	Yes	A	C
Yes	Yes	Female	Yes	A	C

1.a. Compute the entropy of the training data with respect to the Risk class

Sol^n

Total sample =10

Poss values of Target = A, B, C

where, $n(A) = 3$

$n(B) = 3$

$n(C) = 4$

$$Entropy(S) = -\sum_{j=1}^{v} P_j \log_v P_j$$

$$= -\frac{3}{10}\log_3(3/10) - \frac{3}{10}\log_3(3/10) - \frac{4}{10}\log_3[4/10]$$

$$= 0.328 + 0.328 + 0.333$$

$$= 0.989$$

1.b) Compute the information gain of all attributes. Write all attributes and necessary expressions used in the computations and show all the steps neatly.

Soln

~~Info~~ $Gain(S, Attribute) = Entropy(S) - \sum_{v \in V} \frac{|S_v|}{|S|} Entropy(S_v)$

where. V : attribute
v : values of attribute

Step-1: Calculate Entropy

$Entropy(S) = 0.989$

Step-2 For attribute "own home":

There are two possible value Yes or NO

own home	Risk class		
	A	B	C
Yes	1	2	2
NO	2	1	2

$$\text{Info-Gain(ownhome)} = \underbrace{\frac{5}{10} I(1,2,2)}_{\text{For 'Yes'}} + \underbrace{\frac{5}{10} I(2,1,2)}_{\text{For NO}}$$

$$\text{Gain(Own-home)} = \frac{5}{10}\, I(1,2,2) + \frac{5}{10}\, I(2,1,2)$$

$$= \frac{5}{10}\left[-\frac{1}{5}\log_3 1/5 - \frac{2}{5}\log_3 2/5 - \frac{2}{5}\log_3 2/5\right]$$

$$+ \frac{5}{10}\left[-\frac{2}{5}\log_3 2/5 - \frac{1}{5}\log_3 1/5 - \frac{2}{5}\log_3 2/5\right]$$

Info.Gain = 0.958

Gain(Own home) = Entropy(S) − Info.Gain = 0.989 − 0.958

So $\boxed{\text{Gain(own.home)} = 0.031}$

Step-3 for attribute Married

Gain(Married)

	Risk class		
Married	A	B	C
Yes	0	1	4
NO	3	2	0

$$\text{Gain(Married)} = \frac{5}{10}\, I(0,1,4) + \frac{5}{10}\, I(3,2,0)$$

$$= \left(\frac{5}{10}\times 0\right) + \left(\frac{5}{10}\times 0\right) = 0$$

$\boxed{\text{Gain(Married} = 0.0}$, ~~In~~ = 0.989 − 0

$\boxed{\text{Gain(married)} = 0.989}$

Step-3: For Attribute 'Gender'

	Risk class		
Gender	A	B	C
Male	0	3	0
Female	3	0	4

$$\text{InfoGain(Married)} = \frac{3}{10} I(0,3,0) + \frac{7}{10} I(3,0,4)$$
$$= 0 + 0$$
$$= 0$$

$$\therefore \text{Gain(Married)} = \text{Entropy}(S) - 0$$
$$= 0.989 - 0$$
$$= 0.989$$

Step-4: for Employed

Employed	A	B	C
yes	3	1	4
NO	0	2	0

$$\text{Info\_Gain(Employed)} = \frac{8}{10} I(3,1,4) + \frac{2}{10} I(0,2,0)$$
$$= \frac{8}{10} I(3,1,4) + 0$$
$$= \frac{8}{10}\left[-\frac{3}{8}\log_3 3/8 - \frac{1}{8}\log_3 1/8 - \frac{4}{8}\log_3 4/8\right]$$
$$= 0.708$$

$$\therefore \text{Gain(Employed)} = \text{Entropy}(S) - 0.708$$
$$= 0.989 - 0.708$$
$$= 0.281$$

For Q. credit Rating

credit Rate	A	B	C
A	2	1	2
B	1	2	2

$$\text{Info}(\text{credit rate}) = \frac{5}{10} I(2,1,2) + \frac{5}{10} I(1,2,2)$$

$$= \frac{5}{10}\left[-\frac{2}{5}\log 2/5 - \frac{1}{5}\log 1/5 - \frac{2}{5}\log 2/5\right]$$

$$+ \frac{5}{10}\left[-\frac{1}{5}\log 1/5 - \frac{2}{5}\log 2/5 - \frac{2}{5}\log 2/5\right)$$

$$= 0.958$$

$$\therefore \text{Gain}(\text{credit rate}) = \text{Entropy}(s) - 0.958$$
$$= 0.031$$

Now, we have

Gain (own home) = 0.031
Gain (Married) = 0.989
Gain(Gender) = 0.989
Gain(credit rating) = 0.031
Gain (Employed) = 0.281

Q.1.C Draw the complete decision tree. Justify your answer.

Soln We have maximum Gain for attribute 'Gender' as well as Married. So you can consider one of them as root.

ID3 Algo

Step.1 Create a root node for the tree

Step.2: If all examples are posi'A', return the single node tree root, with level risk 'A'.

Step.3: If all examples are 'B', Return the single node tree root, with lebel B.

Step.4 If all attributes is empty, return the single-node tree root, with label = most common value of target-attribute in example.

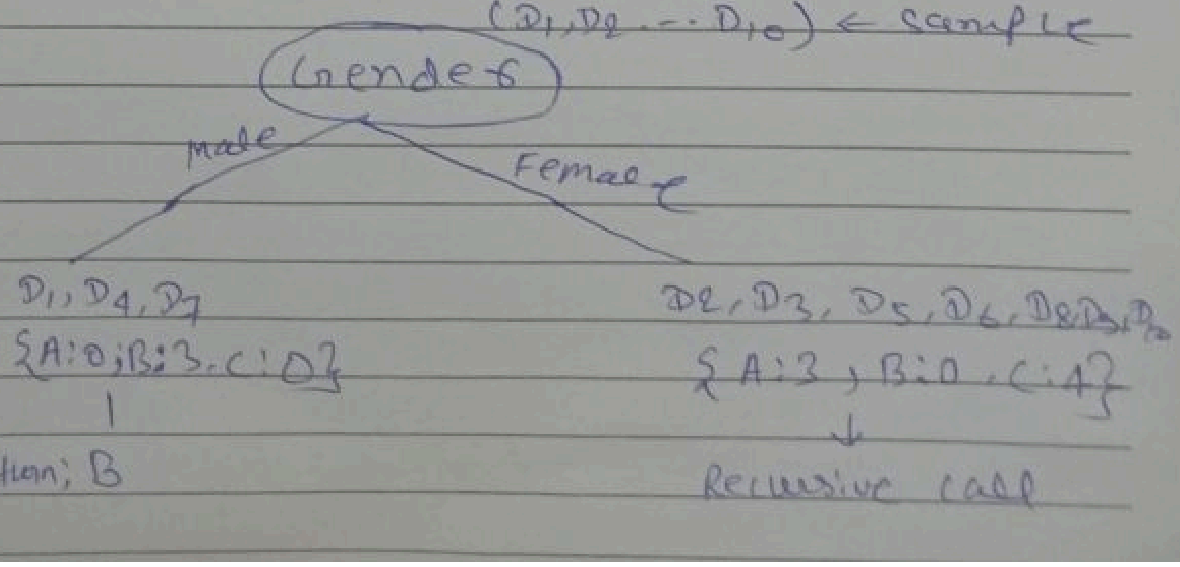
Otherwise Begin

- A ← the attribute from attributes that best * classifies examples
- The decision attribute for root ← A
- ~~for~~

```
For each possible value of A.
  a> Add a new branch below
      root, corresponding to the
      test A = Vi
  b> let Examples vi be the subset
      of examples that have value
      of vi for A.
  c> if Examples_vi is empty
      - Then below this new branch
        add a leaf node with label
        which are most common
        value of Target-~~att~~ribute in
        example.
Else
   a> below this new branch add the
       subtree
   b> ID3 (Examples_vi, Target_attributes,
                Attributes - {A}))
```

This is a recursive call to generate or build Decision tree.

Since, "Gender" has Max gain. So, make it as root.



Now. remaining attribute
{own home, Married, Gender, Employed, Credit rate}
— {Gender}

	Own home	married	Employed ~~Gender~~	Credit rate	Risk class
D2	NO	NO	yes	A	A
D3	yes	yes	yes	B	C
D5	NO	yes	yes	B	C
D6	NO	NO	yes	B	C
D8	yes	NO	yes	A	A
D9	NO	yes	yes	A	C
D10	yes	yes	yes	A	C

Again, we will calculate the gain each attribute & consider maximum gain as next branch. $n(A) = 2$, $n(C) = 5$

Entropy (Risk class) $= I(2,5)$

$$= \left(-\frac{2}{7}\log_2 2/7\right) - \frac{5}{7}\log_2 5/7$$

$$= 0.516 + 0.346$$

$$= 0.862$$

Gain (own-home) $= 0.862 - \left[\frac{2}{7} I(1,2) + \frac{5}{7} I(1,2)\right]$

home	A	C
Yes	1	2
No	1	2

$$= 0.862 - \left[\frac{2}{7} \times 1 + \frac{5}{7}\left(-\frac{1}{5}\log 1/5 - \frac{4}{5}\log\frac{4}{5}\right)\right)$$

$$= 0.862 - \left(\frac{2}{7} + \frac{5}{7} \times 0.721\right)$$

$$= 0.862 - 0.299 = 0.608$$

~~= 0.062~~

$$\text{Gain (Married)} = 0.862 - \left[\frac{2}{7} I(0,2) + \frac{5}{7} I(2,1)\right]$$

Married	A	C
Yes	0	4
No	2	1

$$= 0.862 - \left[0 + \frac{5}{7}\left(-\frac{2}{3}\log \frac{2}{3} - \frac{1}{3}\log \frac{1}{3}\right)\right]$$

$$= 0.862 - 0.655$$

$$= 0.207$$

Gain(Emploeed)

Emplo	A	C
Yes	2	5
No	0	0

$$= 0.862 - \left[\frac{2}{7} I(2,5) + \frac{5}{7}(0,0)\right]$$

$$= 0.862 - \frac{2}{7}\left[-\frac{2}{7}\log 2/7 - \frac{5}{7}\log 5/7\right]$$

$$= 0.862 - 0.246$$

$$\text{Gain(Emplye)} = 0.616$$

Gain (credit rate)

Credit	A	C
A	2	2
B	0	3

$$\text{Gain (Credit)} = 0.862 - \left[\frac{2}{7} I(2,2) + \frac{5}{7} I(0,3)\right]$$

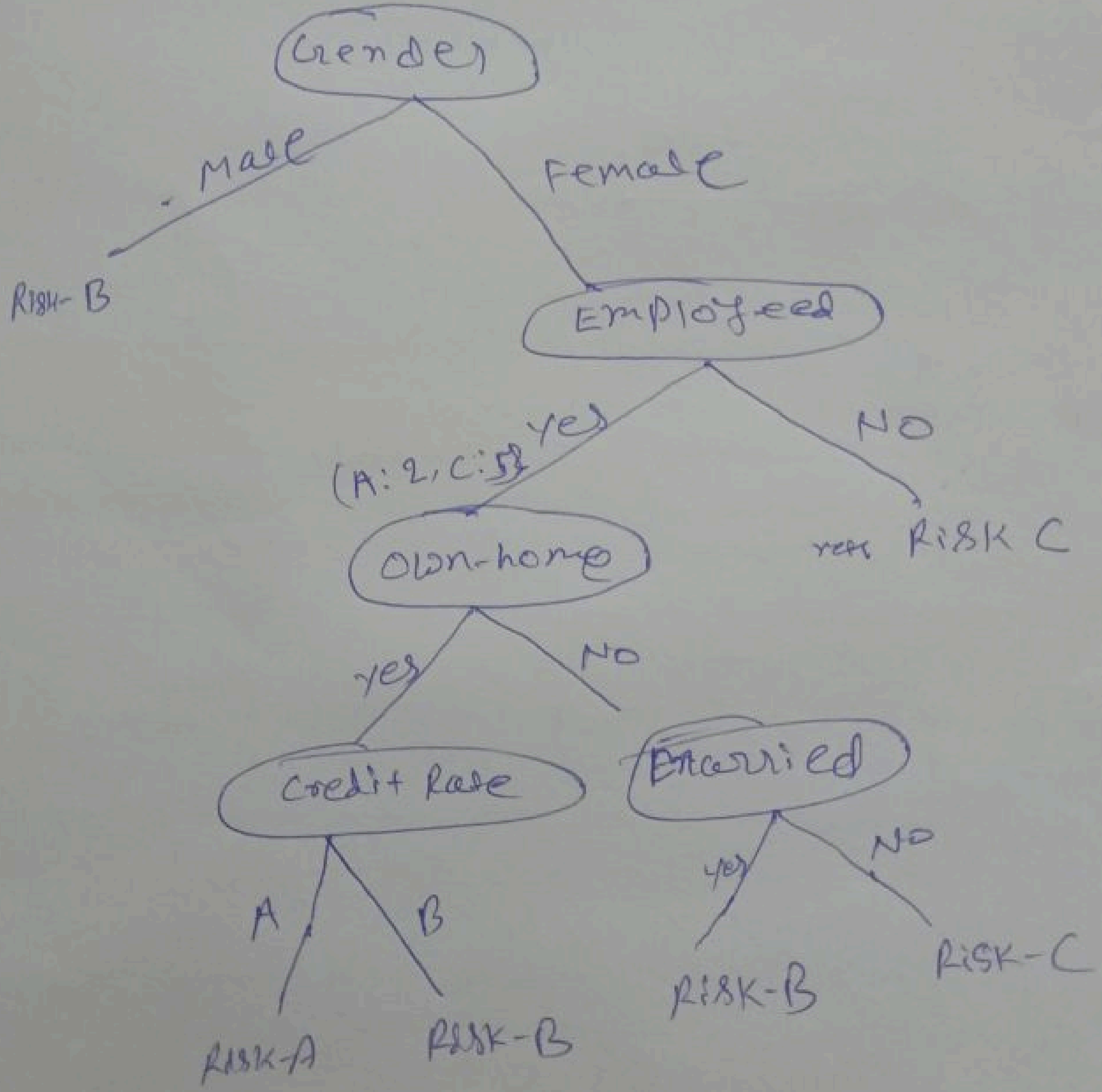
$$= 0.862 - \left[\frac{2}{7} \times 1 + 0\right]$$

$$\text{Gain(credit)} = 0.576$$

So, we have, tree

we have, Gain(credit) = 0.576
Gain(own home) = 0.608
Gain(married) = 0.207
Gain(Employed) = 0.616

Here "Employed" has maximum gain. so, next branch will be Employed.



Q.2. A sample of 6 person was selected the value of their age (x variable) and their weight is demonstrated. Find the regression equation and what the predicted weight is when age is 8.5 years.

serial No.	Age(x)	weight(y)
1	7	12
2	6	8
3	8	12
4	5	10
5	6	11
6	9	13

Solⁿ

	Age(x)	weight(y)	xy	x^2	y^2
	7	12	84	49	144
	6	8	48	36	64
	8	12	96	64	144
	5	10	50	25	100
	6	11	66	36	121
	9	13	117	81	169
Total	41	66	461	291	742

$$\bar{x} = \frac{41}{6} = 6.83 \quad , \quad \bar{y} = \frac{66}{6} = 11$$

$$b = \frac{461 - \frac{41 \times 66}{6}}{291 - \frac{(41)^2}{6}} = 0.92$$

Regression eqn

$$y = 11 + 0.9(x - 6.83)$$

$$y(x) = 4.675 + 0.92x$$

$$\therefore y(8.5) = 4.675 + 0.92 * 8.5$$

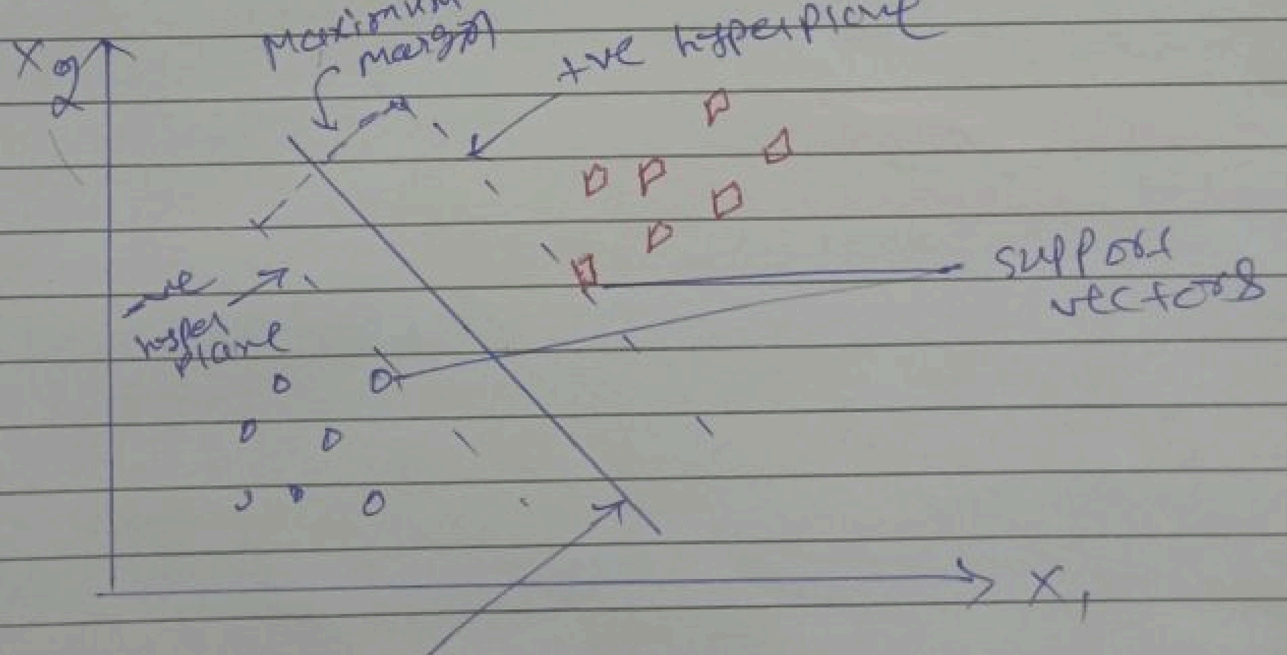
$$= 12.50 \text{ kg} \ \underline{\text{Ans}}$$

Q3. Explain support vector machine. mention at least two advantages and dis-advantages of support vector machines.

Ans.

support vector machine is a supervised machine learning algorithm used for classification and regression tasks. It works by finding the optimal hyperplane that separates data points belonging to different classes in a high-dimensional space.

The key idea is to maximize the margin between classes, which is the difference between the hyperplane and the nearest data points.



Maximum
Margin
hyperplane.

Advantages

i) Effective in high dimensional space

ii) Robousst overfitting.

svm are less prone to overfitting.

Dis-advantages

i) computational intensity

ii) sensitivity to noise.

svm are sensitive to noise in the training data, which can affect the placement of the hyperplane.

Signature