UNIVERSITY OF ASIA PACIFIC

Department of Computer Science & Engineering

Final Examination
Spring-2021

Student Name	:	Md. Sohanuzzaman Soad
Student ID	:	18101064
Section	:	B
Year	:	4th
Semester	:	1st
Course Code	:	CSE 427
Course Title	:	Topics of Current Interest
Date	:	27-November-2021

University of Asia Pacific

Admit Card

Final-Term Examination of Spring, 2021

Financial Clearance	PAID

Registration No : 18101064

Student Name : Md. Sohanuzzaman Soad

Program : Bachelor of Science in Computer Science and Engineering

SI.NO.	COURSE CODE	COURSE TITLE	CR.HR.	EXAM. SCHEDULE
1	CSE 400	Project / Thesis	3.00	
2	CSE 401	Mathematics for computer Science	3.00	
3	CSE 403	Artificial Intelligence and Expert Systems	3.00	
4	CSE 404	Artificial Intelligence and Expert Systems Lab	1.50	
5	CSE 405	Operating Systems	3.00	
6	CSE 406	Operating Systems Lab	1.50	
7	CSE 407	ICTLaw, Policy and Ethics	2.00	
8	CSE 410	Software Development	1.50	
9	CSE 427	Topics of Current Interest	3.00	

Total Credit: 21.50

1. Examinees are not allowed to enter the examination hall after 30 minutes of commencement of examination for mid semester examinations and 60 minutes for semester final examinations.

2. No examinees shall be allowed to submit their answer scripts before 50% of the allocated time of examination has elapsed.

3. No examinees would be allowed to go to washroom within the first 60 minutes of final examinations.

4. No student will be allowed to carry any books, bags, extra paper or cellular phone or objectionable items/incriminating paper in the examination hall.
Violators will be subjects to disciplinary action.

This is a system generated Admit Card. No signature is required.

ID: 48909064

Ans to the Ques No: 1(a)

Cross Validation: Cross validation is a statistical method used to estimate the skill of of machine learning models. It is commonly used in applied machine learning to compare and select a model for a given predictive modeling problem. because it is a easy to understand.

There are two types of cross validation:

1) K-Fold cross validation

2) 5 x 2 cross validation.

ID: 18909064

1) K-Fold cross Validation:

Dataset Divided into random size of part x_i, $i = 1 \ldots k$. To generate each pair.

Pair (v_i, T_i):

$v_1 = x_1 \quad T_1 = x_2 \cup x_3 \cup \ldots \cup x_K$

$V_2 = x_2 : T_2 = X_1 \cup X_3 \cup \ldots \cup X_K$

— — — — — — — — —

$V_K = X_K \quad T_K = X_1 \cup X_2 \cup \ldots \cup X_{K-1}$

2) 5X2 cross validation: Dataset Divided into 2 equal part $x_1^{(1)}$ and $x_1^{(2)}$. $x_1^{(1)}$ as training set and $x_1^{(2)}$ as validation set. Then swap these. This is repeat more than four times to get ten pair of training set and validation set.

ID: 18909064

$T_1 = x_1^{(1)}, \quad V_1 = x_1^{(2)}$

$T_2 = x_1^{(2)}, \quad V_2 = x_1^{(1)}$

$T_3 = x_2^{(1)}, \quad V_3 = x_2^{(2)}$

$T_4 = x_2^{(2)}, \quad V_4 = x_2^{(1)}$

$T_9 = x_5^{(1)}, \quad V_9 = x_5^{(2)}$

$T_{10} = x_5^{(2)}, \quad V_{10} = x_5^{(1)}$

ID:48109064

Ans to the Que. No; 1 (6)

Each record may be assigned a class ~~model~~ label 'relevent' or not 'relevant'. All the records a were tested for relevance. The test classified 55 records on 'relevant'. But only 45 of them were actually relevant. Hence we have following confusion matrix for search:

Fp = Number of False Positive.

	Actual Condition : True	Actual Condition False.
Predicted condition is true	TP	FP
Predicted condition is false	FN	TN

ID:48101064

	Actual: relevant	Actual Not relevant
Predicted 'relevant'	45 (TP)	10 (FP)
Predicted 'Not relevant'	20 (FN)	35 (TN)

Here,

$$TP = 45$$

$$FP = 55 - 45 = 10$$

$$FN = (65 - 45) = 20$$

The precision p is we know that,

$$P = \frac{TP}{TP + FP} = \frac{45}{45+10} = \frac{45}{55}$$

$$= 0.81$$

The Recall R is we know that

$$R = \frac{TP}{TP + FN} = \frac{45}{45+20} = \frac{45}{65}$$

$$= 0.69$$

ID:18101064

Ans to the Que. No:2(a)

Regression Trees:

A regression problem is a problem of determining of a relation one or more independent variable and output variable which is a real continious variable and then using the relation to predict the values of dependent variables. Trees can be also used to make such prediction. A tree used for making prediction of numerical variable is called prediction tree or a regression tree.

x_1	1	3	4	6	10	15	2	16	0
x_2	12	23	21	10	27	23	35	27	17
y	10.1	15.3	11.5	13.9	17.8	23.1	12.7	17.6	14.9

Table 1: Data for regression tree

ID: 18101064

Solution: We shall construct raw decision tree to predict the value of y corresponding to any untabulated values x_i and x_1

Step 1: Arbitrary split the values of x_1 into two set : one set is $x_1 < 6$ and other is $x_1 \geq 6$

x_1	1	2	4	2	0
x_2	12	23	21	35	17
y	10.1	15.3	11.5	12.7	14.9

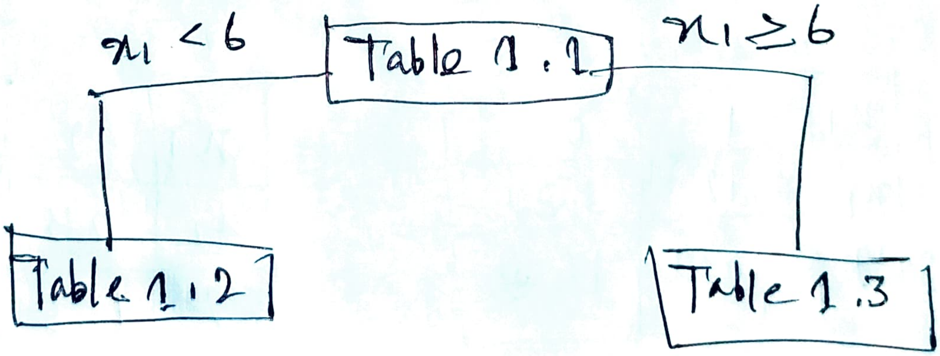
Table 7.9 : Data for Regression Tree

ID : 18104064

Step 2 : Now split the values of x_2 into two set : one defined by $x_2 < 21$ and other $x_2 \geq 21$.

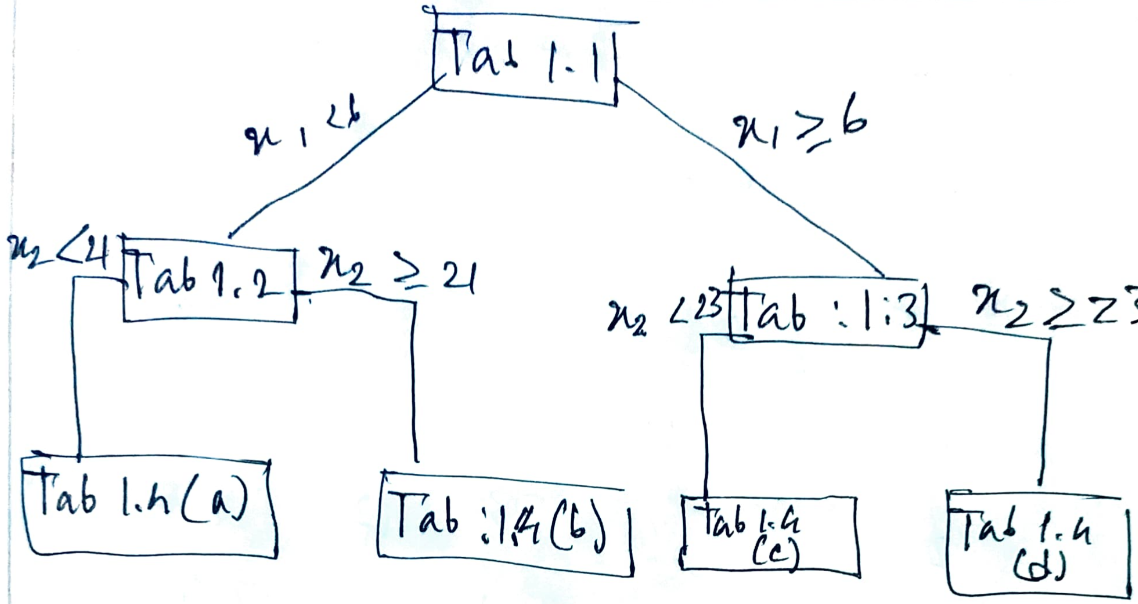
x_1	6	10	15	7	16
x_2	10	27	23	12	27
y	13.9	17.8	23.1	43.0	17.6

Table 1.2 : Data for Regression Tree



and one specified by $x_2 \geq 23$. the split data are given in Table 1.4(a) -(d)

ID: 18101064



Step 3: make the nodes spcified by table 1.4 (a) - — 1.4 (d) into leaf node

x_1	1	0
x_2	12	17
y	10.1	14.9

(a)

x_1	2	4	2
x_2	23	21	35
y	15.3	11.5	12.7

(b)

x_1	6	7
x_2	10	12
y	13.9	43.0

(c)

x_1	10	15	16
x_2	27	23	17
y	7.8	23.1	17.6

(d)

Table : 1.4

ID: 18101064

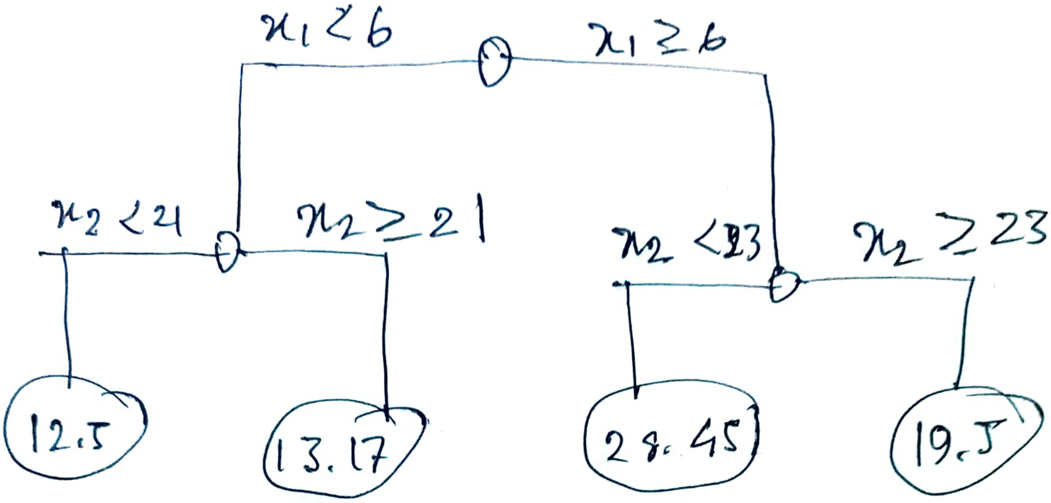


Fig: 1.5 : Regression Tree

Step 4: Fig 1.5 is the final Regression tree for predicting the values of y based on the data.

ID: 18101064

Ans to the Que No: 2(b)

Using ID3 Algorithm to construct a decision tree for the data in the following table:

Age	Competition	Type	Class
old	Yes	Software	Down
old	No	Software	Down
old	No	Hardware	Down
mid	Yes	Software	Down
mid	Yes	Hardware	Down
mid	No	Hardware	Up
mid	No	Software	Up
New	Yes	Software	Up
New	No	Hardware	Up
New	No	Software	

Solution: In the given data there first three attributes are predictor and the last attributes ~~is~~ is the target attribute.

ID: 18101064

Step 1: We first create a root node for the tree ~~root node~~

Root node

$\boxed{\text{Table}}$

Fig 1: Root node of the decision tree for Table-1

Step 2: All examples are positive (class label 'Down') and not all examples are negative (class lable 'up') also the number of features is not zero.

Step 3: We have to decide which feature is to be placed at the root node. For this we have to calculate the information gain corresponding to each of the three attributes.

ID: 18101064

i) Calculation of Entropy (S):

$$\text{Entropy}(S) = P_{down} \log_2 (P_{down}) - P_{up} \log_2 (P_{up})$$

$$= -\frac{5}{10} \log_2 \left(\frac{5}{10}\right) - \frac{5}{10} \log_2 \left(\frac{5}{10}\right)$$

$$= 0.5 + 0.5$$

$$= 1$$

ii) Calculational of Gain (S, Age): Here the values of the attributes 'Age' are 'old', 'mid' 'New' we have to calculate entropy of these too,

$$\text{Entropy}(S_{old}) = -\frac{3}{3} \log_2 \left(\frac{3}{3}\right) - \frac{0}{3} \log_2 \left(\frac{0}{3}\right)$$

$$= 0$$

$$\text{Entropy}(S_{mid}) = -\frac{2}{4} \log_2 \left(\frac{2}{4}\right) - \frac{2}{4} \log_2 \left(\frac{2}{4}\right)$$

$$= 0.5 + 0.5 = 1$$

$$\text{Entropy}(S_{new}) = -\frac{0}{3} \log_2 \left(\frac{0}{3}\right) - \frac{3}{3} \log_2 \left(\frac{3}{3}\right)$$

$$= 0$$

$\therefore$ Gain $(S, Age) = Entropy(S) - \frac{|Entropy(S_{old})|}{|Entropy(S)|} \times$

$Entropy(S_{old}) - \frac{|Entropy(S_{mid})|}{|Entropy(S)|} \times Entropy(S_{mid})$

$- \frac{|Entropy(S_{new})|}{|Entropy(S)|} \times Entropy(S_{new})$

$\Rightarrow 1 - \left(\frac{0}{1} \times 0\right) - \left(\frac{1}{1} \times 1\right) - \left(\frac{0}{1} \times 0\right)$

$\Rightarrow 1 - 0 - 1 - 0$

$\Rightarrow 0$

iii) Calculation of Gain (S, Competition):

$Entropy(S_{yes}) = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)$

$= 0.811$

$Entropy(S_{No}) = -\frac{2}{6}\log_2\left(\frac{2}{6}\right) - \frac{4}{6}\log_2\left(\frac{4}{6}\right)$

$= 0.906$

$Gain(S, competition) = 1 - \left(\frac{0.811}{1} \times 0.811\right) - \left(\frac{0.906}{1} \times 0.906\right)$

$\Rightarrow 1 - 0.657 - 0.820$

$\Rightarrow -0.477$

ID: 18101064

IV) Calculation of Gain (S, type):

Entropy ($S_{Software}$) $= -\frac{3}{6}\log_2(\frac{3}{6}) - \frac{3}{6}\log_2(\frac{3}{6})$

$= 0.5 + 0.5$

$= 1$

Entropy ($S_{hardware}$) $= -\frac{2}{4}\log(\frac{2}{4}) - \frac{2}{4}\log_2(\frac{2}{4})$

$= 0.5 + 0.5 = 1$

Gain (S, Type) $= 1 - (\frac{1}{1} \times 1) - (\frac{1}{1} \times 1)$

$= 1 - 1 - 1 = 1$

Step 4: Here we find the highest information gain which is the maximum among Gain (S, Age), Gain (S, competition), Gain (S, Type):

Highest Information gain $= \max\{0, -0.477, -1\}$

$= 0$

This corresponds to the feature 'Age' so we place Age at to root node and we now split the root node in the following

ID: 18101064

figure (1) into three branches.

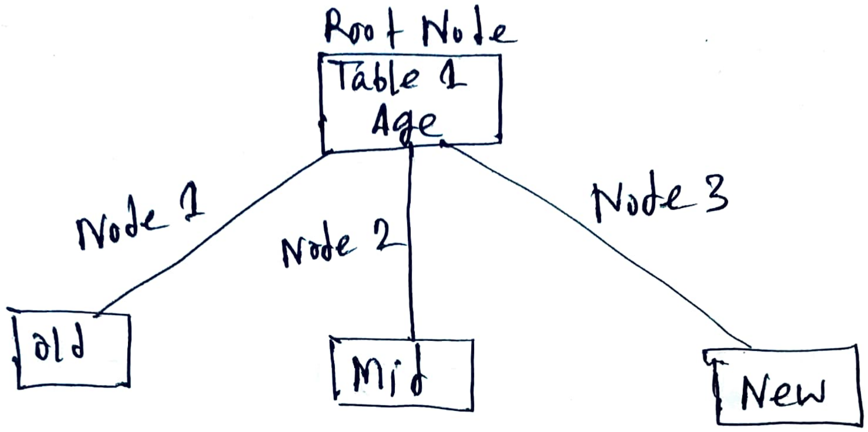


Figure 2: Decision Tree for data in Table-1 after selecting the branching feature at the root node.

Step 5: Now make a leaf node for old which is down on the basis of data information.

Age	Competition	Type	Class
old	Yes	Software	Down
old	No	Software	Down
old	No	Hardware	Down

So for 'old' age Target attributes all goes 'Down'. So we insert a leaf node for 'old' which is 'Down' in Figure-3

Now, for 'New' Age target attributes all go 'Up'

Age	Competition	Type	Class
New	Yes	Software	Up
New	No	Hardware	Up
New	NO	Software	Up

So, we now make a leaf node for 'New' Age which is upon the basis of our dataset information in the following figure 3:

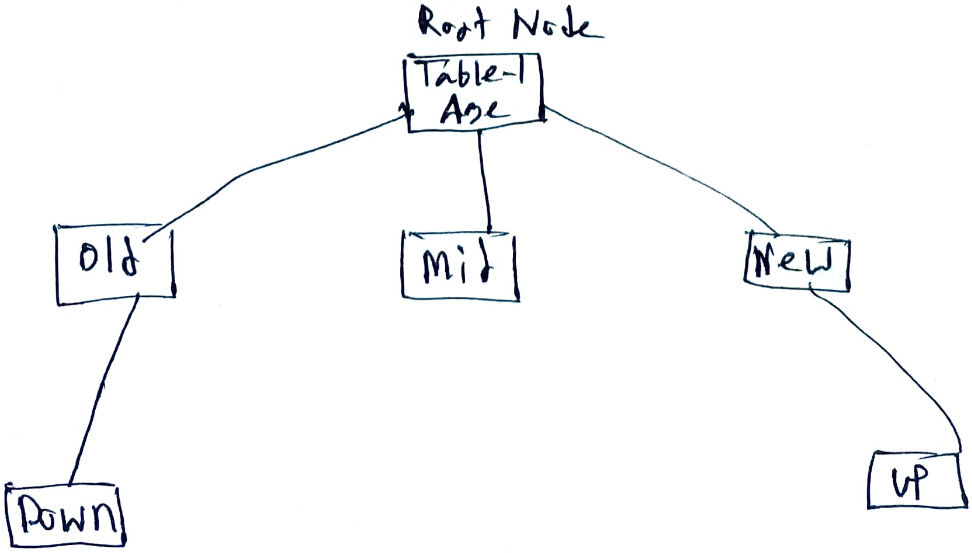


Figure3: Decision Tree

ID: 18101064

In the given dataset 'Mid' age there are two causes, so which will become the child node of 'Mid', we need to calculate the gain for 'competition' and 'Type' to find the child node 'Mid' Age.

Age	competition	Type	class
Mid	Yes	Software	Down
Mid	Yes	Hardware	Down
Mid	No	Hardware	up
Mid	No	Software	up

Let $S^{(1)} = S_{Age} = \text{Mid}$ we have $|S^{(1)}| = 4$ which is shown in the table

$\therefore \text{Gain}(S^{(1)}, \text{competition})$

$= \text{Entropy}(S^{(1)}) - \frac{|S^{(1)}_{\text{competition}=\text{Yes}}|}{|S^{(1)}|} \times \text{Entropy}(S^{(1)}_{\text{competition}=\text{yes}}) - \frac{|S^{(1)}_{\text{compitition}=\text{No}}|}{|S^{(1)}|} \times \text{Entropy}(S^{(1)}_{\text{compitition}=\text{No}})$

$= \{-\frac{2}{4}\log_2(\frac{2}{4}) - \log_2(\frac{2}{4})\} - \{\frac{2}{4} \times (-(\frac{2}{2}\log_2(\frac{2}{2})))\} - \{\frac{2}{4} \times (-(\frac{2}{2})\log_2(\frac{2}{2}))\}$

ID: 98909064

$= \{0.5 + 0.5\} - \{0\} - \{0\}$

$= 1$

Now, Gain $(S^{(1)}, Type)$

$= Entropy(S^{(1)}) - \frac{|S_{Type=Software}|}{|S^{(1)}|} \times Entropy(S^{(1)}_{Type=Software}) - \frac{|S^{(1)}_{Type=Hardware}|}{|S^{(1)}|} \times Entropy(S^{(1)}_{Type=Hardware})$

$= \{-\frac{2}{4}\log_2(\frac{2}{4}) - \frac{2}{4}\log_2(\frac{2}{4})\} - \{\frac{2}{4} \times (-(\frac{1}{2})\log_2(\frac{1}{2}) - (\frac{1}{2})\log(\frac{1}{2})\} - \{\frac{2}{4} \times (-(\frac{1}{2})\log_2(\frac{1}{2}) - (\frac{1}{2})\log_2(\frac{1}{2})\}$

$= \{0.5 + 0.5\} - \{\frac{2}{4} \times (\frac{1}{2} + \frac{1}{2})\} - \{\frac{2}{4}(\frac{1}{2} + \frac{1}{2}))\}$

$= 1 - \frac{2}{4} - \frac{2}{4}$

$= 0$

So, the maximum of Gain $(S^{(1)}$, competition) and Gain $(S^{(1)}, Type)$ is Gain $(S^{(1)}$, competition). As Gain of competition is greater

ID: 18404064

than gain of Type. So, 'competition' because the child node of 'Mid' which is show in Figure - 4

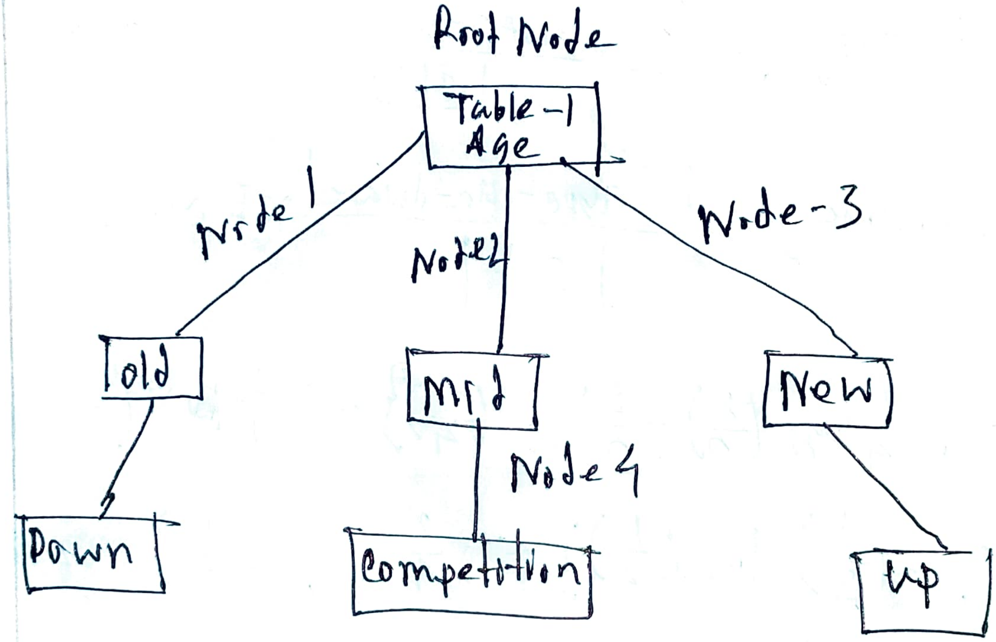


Figure: 4: Decision Tree

Step 6: Now the dataset as follows for competition:

Competition	Type	class
Yes	Software	Down
Yes	Hardware	Down
No	Hardware	Up
No	Software	Up

ID: 18109064

so, now we make a leaf note for competition if the 'competition' is 'Yes' then class label 'Down' else label is 'up'. So this is shown is Figure 5.

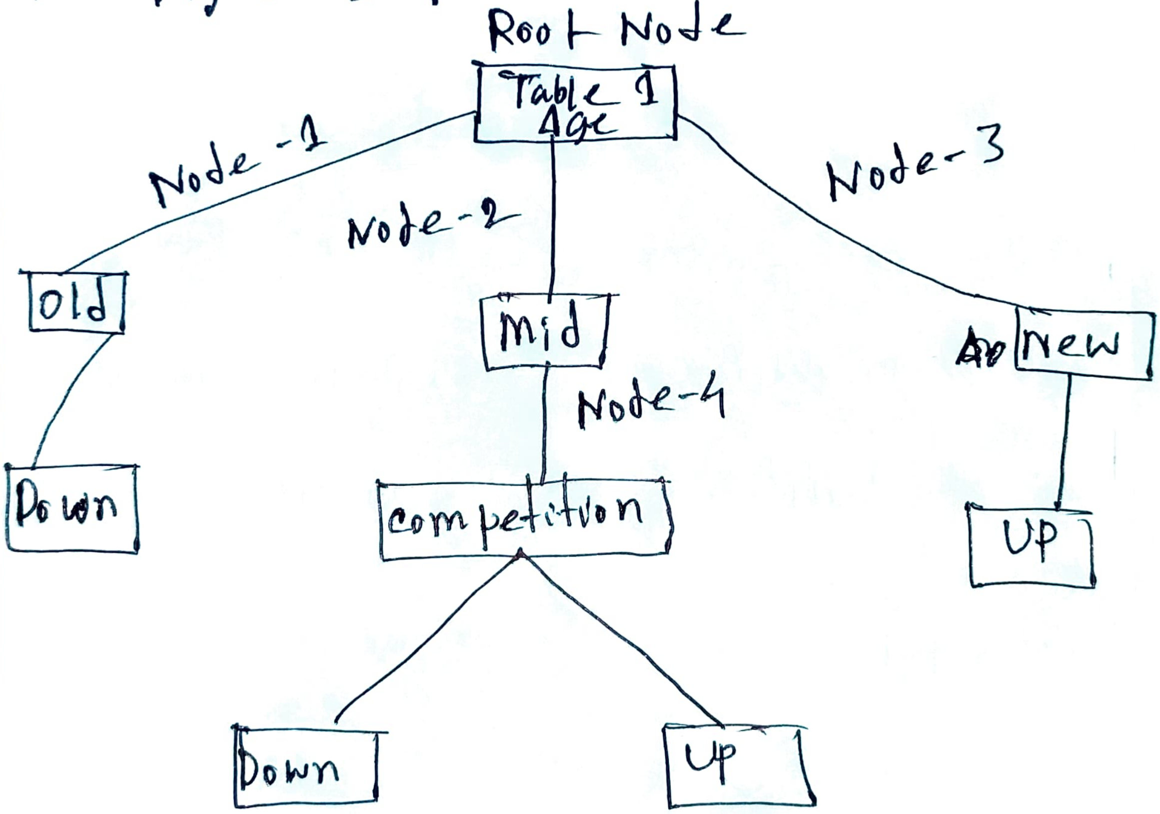


Figure 5: Final Decision Tree

So our decision Tree for given dataset is Constructed with ID3 Algorithm shown in figure: 5.

ID: 18101064

Ans to the Que-NO; 3(a)

Dimensionality reduction is useful because of several reasons:

1. In most learning Algorithm, the complexity depends on the number of input dimension d, as well as on size of sample, N. For reduce memory use this reduction method.

2. Avoid ~~unnesse~~ useless data from dataset.

3. Simpler models are more robust on small datasets.

4. When data can be explain dwith fewer feature.

ID: 18101064

In forward selection, we start with no variables and add them one by one, at each step adding the one that decrease the error the most untill any further addition does not decrease the error.

Procedure: Following Notations:

n: num of input variable

$x_1 \ldots x_n$: Input Variable

F_i: Subset of the set of input variable

$E(F_i)$: error incurred ~~one~~ on the validation sample.

ID: 18101064

1) Set ~~of~~ $F_0 = \emptyset$ and $E(F_0) = \infty$

2) For $i = 0, 1$ ___ repeat the following untill $E(F_{i+1}) \geq E(F_i)$.

(a) For all possible input variables x_j, train the model with the input ~~va~~ $F_i \cup \{x_j\}$ on calculate $E(F_i \cup \{x_j\})$ on validation set.

(b) choose that input x_m that has least error.

$$E(F_i \cup \{x_j\}) : m = \arg\min_j E(F_i \cup \{x_j\})$$

(c) Set $F_{i+1} = F_i \cup \{x_m\}$

(3) Then Set F_i is outputted as the best subset.

ID: 18101064

Ans to the Que. No: 3(b)

Principle Component Analysis:

principal component analysis (PCA) is a statistical procedure that we an algorithm for orthogonal transformation to convert a set of observation of possibly correlated variable into a set of values of linearly uncorrelated variable called principle components. Each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components.

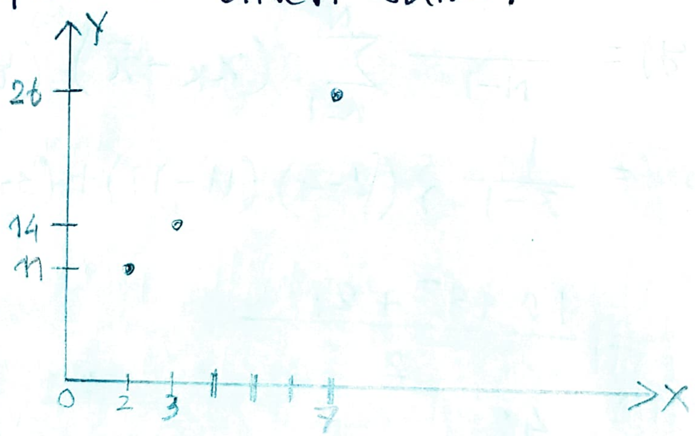
ID: 18101064

Given the following data:

$x = 2 \quad 3 \quad 7$

$y = 11 \quad 14 \quad 26$

Scatter plot of Given data:



computes the means of the variables:

$$\bar{x} = \frac{1}{3}(2+3+7) = \frac{12}{3} = 4$$

$$\bar{y} = \frac{1}{3}(11+14+26) = \frac{51}{3} = 17$$

Calculate the covariance matrix:

$$S = \begin{bmatrix} cov(x,x) & cov(x,y) \\ cov(y,x) & cov(y,y) \end{bmatrix}$$

ID: 18101064

$$\operatorname{cov}(x,x) = \frac{1}{N-1}\sum_{K=1}^{N}(x_K-\bar{x})^2$$
$$= \frac{1}{3-1}\left\{(2-4)^2+(3-4)^2+(7-4)^2\right\}$$
$$= \frac{4+1+9}{2}$$
$$= \frac{14}{2}$$
$$= 7$$

$$\operatorname{cov}(x,y) = \frac{1}{N-1}\sum_{K=1}^{N}(x_K-\bar{x})(y_K-\bar{y})$$
$$= \frac{1}{3-1}\left\{(2-4)(11-17)+(3-4)(14-17)+(7-4)(26-17)\right\}$$
$$= \frac{12+3+27}{2}$$
$$= \frac{42}{2}$$
$$= 21$$

$$\operatorname{cov}(y,x) = \frac{1}{N-1}\sum_{K=1}^{N}(y_K-\bar{y})(x_K-\bar{x})$$
$$= \frac{1}{3-1}\left\{(11-17)(2-4)+(14-17)(3-4)+(26-17)(7-4)\right\}$$
$$= \frac{12+3+27}{2}$$
$$= \frac{42}{2}$$
$$= 21$$

ID: 18101064

$$cov(y, y) = \frac{1}{N-1} \sum_{K=1}^{N} (y_K - \bar{y})^2$$

$$= \frac{1}{3-1} \{(11-17)^2 + (14-17)^2 + (26-17)^2\}$$

$$= \frac{36+9+81}{2}$$

$$= \frac{126}{2}$$

$$= 63$$

$$S = \begin{bmatrix} 7 & 21 \\ 21 & 63 \end{bmatrix}$$

Eigen values of the covariance Matrix:

The characteristic equation of the covariance matrix is :

$$\det(S - \lambda I) = 0$$

$$\Rightarrow \begin{vmatrix} 7-\lambda & 21 \\ 21 & 63-\lambda \end{vmatrix} = 0$$

$$\Rightarrow (7-\lambda)(63-\lambda) - (21)(21) = 0$$

$$\Rightarrow 441 - 7\lambda + \lambda^2 - 63\lambda - 441 = 0$$

$$\Rightarrow \lambda^2 - 70\lambda = 0 \quad \Rightarrow \lambda(\lambda - 70) = 0$$

$$\therefore \lambda_1 = 70 \qquad \therefore \lambda_2 = 0$$

Computation of the eigenvector:

The eigenvector corresponding to $\lambda = \lambda_1$ is a vector.

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

The following equation:

$$(S - \lambda_1 I)\, U = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} (7-\lambda_1) & 21 \\ 21 & (63-\lambda_1) \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} (7-\lambda_1)U_1 & 21U_2 \\ 21U_1 & (63-\lambda_1)U_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This I= equivalence to the following two equation

$$(7-\lambda_1)U_1 + 21U_2 = 0$$

$$21U_2 + (63-\lambda_1)U_2 = 0$$

using the theory of system of linear equations these equation are not independent and solution are given by:

$$(7-\lambda_1)U_1 + 21U_2 = 0$$

$$(7-\lambda_1) U_1 = -21 U_2$$

$$\frac{U_1}{U_2} = \frac{-21}{7-\lambda_1} = t$$

where t is real number, taking $t=1$ we get an eigenvector corresponding to λ_1

$$U = \begin{bmatrix} -21 \\ 7-\lambda_1 \end{bmatrix}$$

To find a unit eigenvector, we compute the length of U which is given by

$$\|U\| = \sqrt{(-21)^2 + (7-\lambda_1)^2}$$

$$= \sqrt{441 + (7-70)^2}$$

$$= \sqrt{441 + (-63)}$$

$$= 66.4078$$

Therefore a unit eigenvector corresponding to λ_1 is:

$$e_1 = \begin{bmatrix} -21/\|U\| \\ (7-\lambda_1)/\|U\| \end{bmatrix} = \begin{bmatrix} -21/66.41 \\ (7-70)/66.41 \end{bmatrix} = \begin{bmatrix} -0.3162 \\ -0.9487 \end{bmatrix}$$

ID: 18101064

Similar computation for the unit vector e_2

corresponding to the eigen ~~vector~~ value $\lambda = \lambda_2$

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

The following equation :

$$(S - \lambda_2 I) V = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} (7-\lambda_2) & 21 \\ 21 & (63-\lambda_2) \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} (7-\lambda_2)V_1 & 21V_2 \\ 21V_1 & (63-\lambda_2)V_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The following two equation :

$$(7-\lambda_2) V_2 + 21 V_2 = 0$$

$$\Rightarrow 21 V_1 + (63 - \lambda_2) \cancel{V_2} = 0$$

Now,

$$(7-\lambda_2) V_2 + 21 V_2 = 0$$

$$\Rightarrow (7-\lambda_2) V_2 = -21 V_2$$

$$\Rightarrow \frac{V_1}{V_2} = \frac{-21}{7-\lambda_2} = \ell$$

ID: 18101064

taking $t=1$ we get an eigenvector e_1 for λ_2

$$v = \begin{bmatrix} -21 \\ 7-\lambda_2 \end{bmatrix}$$

Now,

$$\|v\| = \sqrt{(-21)^2 + (7-\lambda_2)^2}$$

$$= \sqrt{(-21)^2 + (7-0)^2}$$

$$= \sqrt{441+49}$$

$$= 22.135$$

$$e_2 = \begin{bmatrix} -21/\|v\| \\ (7-\lambda_2)/\|v\| \end{bmatrix} = \begin{bmatrix} -21/22.135 \\ (7-0)/22.135 \end{bmatrix} = \begin{bmatrix} -0.9487 \\ -0.3162 \end{bmatrix}$$

Compution of first principle Components:

transpose of the matrix e_1^T

$$e_1^T = \begin{bmatrix} -0.3162 & -0.9487 \end{bmatrix}$$

calculate the first principle components.

$$e_1^T \cdot \begin{bmatrix} x_k - \bar{x} \\ y_k - \bar{y} \end{bmatrix}$$

ID: 18101064

example $\begin{bmatrix} 2 \\ 11 \end{bmatrix}$

$\therefore \begin{bmatrix} 0.3162 & -0.9487 \end{bmatrix} \begin{bmatrix} 2-4 \\ 11-17 \end{bmatrix}$

$\Rightarrow (-0.3162)(-2) + (-0.9487)(-6)$

$\Rightarrow 6.6324$

example $\begin{bmatrix} 3 \\ 14 \end{bmatrix}$

$\therefore \begin{bmatrix} -0.3162 & -0.9487 \end{bmatrix} \begin{bmatrix} 3-4 \\ 14-17 \end{bmatrix}$

$\Rightarrow (-0.3162)(-1) + (-0.9487)(-3) \Rightarrow 3.1623$

example $\begin{bmatrix} 7 \\ 26 \end{bmatrix}$

$\therefore \begin{bmatrix} -0.3162 & -0.9487 \end{bmatrix} \begin{bmatrix} 7-4 \\ 26-17 \end{bmatrix}$

$\Rightarrow (-0.3162)(3) + (-0.9487)(9) \Rightarrow -9.4864$

So,

x	2	3	7
y	11	14	26
First principle components	6.6324	3.1623	-9.4869

ID: 18101064

Ans to the Que-NO: 4 (OR) (a)

Hidden Markov Model : The General case

A hidden markov model (HMM) is characterized by the following :

1) The number of states in the model. Say N, Let the states S_1, S_2, S_N.

2) The number of distinct observation symbols. Say M. Let the observation symbols are $V_1, V_2, \ldots V_M$. (The observation symbol correspond to the physical outputs of the system)

3) The state transition probablities specified by an $N \times N$ matrix $A = [a_{ij}]$

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i)$$

for $i = 1, 2, \ldots N$

where q_t is the state at time t.

4) The observation symbol probablity distributions $b_j(k)$ for $j = 1, \_\_\_, N$ and $k = 1, \_\_\_, M$. $b_j(k)$ is the that at time t, the outcome is the symbol v_k given at the system is in state s_j:

$$b_j(k) = P(v_k \text{ at } t \mid q_t = s_j)$$

5) The initial probablities:

$$\pi = [\pi_i]$$

$$\pi = P(q_1 = s_i) \text{ for } i = 1, 2, \sim N$$

The values of N and M implictly defined by A, B and π. So, a HMM is completely defined by parameter set

ID: 18 10 1064

$\lambda = (A, B, \pi)$

There are three Basic Problems :

Problem 1: Evaluation Problem

Given the observation sequence

$O = O_1 O_2 \ldots O_T$

and a HMM model

$\lambda = (A, B, \pi)$

how do we efficiently compute

$P(O|\lambda)$

the probablity of the observation sequence

O given by the model λ?

Problem 2: Finding the state Sequence

Problem

given the observation sequence

$O = O_1 O_2 - O_T$

and HMM model

$\lambda = (A, B, \pi)$

ID: 18101064

how do we find the state sequence

$$Q = q_1 q_2 \text{———} , q_T$$

which has the highest probablity of generating O; that is how do we find Q^* that maximizes the probablity $P(Q \mid O, \lambda)$?

problem 3: Learning model parameter

Problem:

Given a training set X observation sequence how do we learn the model $\lambda = (A, B, \pi)$ that maximizes the probablity of generating X: that is how do we find λ^* that maximizes the probablity

$$P(X \mid \lambda)$$

ID: 18101064

Ans to the Que. No: 4(OR)(b)

A markov process is a random process indexed by time, and with the property that future is independent of the past given the present. The time space may be discrete taking the values 1, 2, _ or continious taking any non-negative number as a value.

Through this example we introduce the various elements that constitute discrete homogenous markov process:

1) System and States: Let us consider a highly simplified model of the different states a stock market is in a give a week. We assume that there are only three possible states:

S1 : Bull Market Trend

S2 : Bear Market Trend

S3 : Stagnant Market Trend

2) Transition Probablities: Week after week, the stock market moves from one state to other state

from previous data. It has been estimated that there are certain probablities associated with these movements. These probablities are called transition probablities.

3) Markov Assumption: We assume that the following statement regarding transition probablities is true:

Let the weeks be counted at 1, 2, — and let and arbitary week be the t-th week. Then, the state in week t+1 depends only on the state in week t, regardless of the states in the previous weeks. This corresponds to saying that given the presented state, the future is ~~inde~~ ~~tobe~~ Independent of the past.

4) Homogeneity Assumption: To simplify the computation we assume that following property called the homogeniety assumption is also true.

The probablity that the stock market is in a particular state in a particular week t+1 is given that it is a particular state in week t.

ID: 18101064

5) Representation of transition Probablites :

The transition probablities can be represented in two ways :

a) if the number of state is small, the state transition probablity can be depresent diagramm atically.

b) The state transition probablities can also be represented by a matrix called state transition matrix.

$$A = \begin{bmatrix} a_{11} & a_{12} & \text{---} & a_{1N} \\ a_{21} & a_{22} & \text{---} & a_{2N} \\ a_{N1} & a_{N2} & & a_{NN} \end{bmatrix}$$

6) Initia Probablities : We define the initial probablities π_i which is the probablity that the first state in the sequence is S_i;

$$\pi = P(q_1 = S_i)$$

We also write, $\pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_N \end{bmatrix}$

We must have

$$\sum_{i=1}^{N} \pi_i = 1$$

ID: 18101064

7) Discrete Markov Process : A system with the states $S_1, S_2, \ldots, S_N$ satifying the markov property is called a discreate Markov process. If is satisfies the homogeneity property, then it is called a homogenous ~~die~~ discrete Markov process.

consider a homogenous discrete Markov process with transition Matrix A and initial probablity vector π. A and π are the parameaters of the process. The following procedure may be applied to learn these paramters.

process :

step 1: obtain K observation sequences each of length T. Let q_{tk} be the observed state at a time t in the Kth observation sequence.

ID: 18104064

Step 2: Let $\hat{\pi}_i$ be the estimate of the initial probablity π_i. Then

$$\hat{\pi}_i = \frac{\text{Number of sequence starting with } S_i}{\text{total number of sequence}}$$

Step 3: Let $\hat{a}_{ij}$ be the estimate of a_{ij}. Then

$$\hat{a}_{ij} = \frac{\text{Number of transition from } S_i \text{ to } S_j}{\text{Number of transition from } S_i}$$

Example: Let there be a discrete Markov Process with three states S_1, S_2, S_3. Suppose we have the following 10 observation sequence each of length 5:

O_1 : $S_1 S_2 S_1 S_1 S_1$

O_2 : $S_2 S_1 S_1 S_3 S_1$

O_3 : $S_3 S_1 S_3 S_2 S_2$

O_4 : $S_1 S_3 S_3 S_1 S_1$

O_5 : $S_3 S_2 S_1 S_1 S_3$

O_6 : $S_3 S_1 S_1 S_2 S_1$

O_7 : $S_1 S_1 S_2 S_3 S_2$

O_8 : $S_2 S_3 S_1 S_2 S_2$

O_9 : $S_3 S_2 S_1 S_1 S_2$

O_{10} : $S_1 S_2 S_2 S_1 S_1$

ID: 18101064

We have:

$$\hat{\pi}_1 = \frac{\text{Number of Seq starting with } S_1}{\text{total number of Seq}}$$

$$= \frac{4}{10}$$

$$\hat{\pi}_2 = \frac{\text{Number of Seq starting with } S_2}{\text{total number of Seq}}$$

$$= \frac{2}{10}$$

$$\hat{\pi}_3 = \frac{\text{Number of Seq starting with } S_3}{\text{total number of Seq}}$$

$$= \frac{4}{10}$$

Therefore,

$$\pi = \begin{bmatrix} 4/10 \\ 2/10 \\ 4/10 \end{bmatrix}$$

We Illustrate the computation a_{ij}'s with an example

$$\hat{a}_{21} = \frac{\text{Num of transitions from } S_2 \text{ to } S_1}{\text{Num of transition from } S_2}$$

$$= \frac{6}{11}$$

ID: 18101064

$$\hat{a}_{22} = \frac{\text{Num of transitions from } S_2 \text{ to } S_2}{\text{Num of transition from } S_2}$$

$$= \frac{3}{11}$$

$$\hat{a}_{23} = \frac{\text{Num of transitions from } S_2 \text{ to } S_3}{\text{num. of transition from } S_2}$$

$$= \frac{2}{11}$$

The remaining transition probablities canbe estimated In a similar way.

Let there be a discrete markov process with two rather S_1 and S_2 given the following sequence of observation of these states. ~~are~~ We have 8 observation

states.

$O_1 : S_1 S_2$

$O_2 : S_2 S_2$

$O_3 : S_1 S_2$

$O_4 : S_2 S_1$

$O_5 : S_1 S_1$

$O_6 : S_2 S_1$

$O_7 : S_1 S_2$

$O_8 : S_1 S_1$

ID: 18101064

We have ,

$$\pi_1 = \frac{\text{Number of sequence with } S1}{\text{total number of seq}}$$

$$= \frac{5}{8}$$

$$\pi_2 = \frac{\text{Number of sequence with } S2}{\text{total number of seq}}$$

$$= \frac{3}{8}$$

Therefore,

$$\pi = \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} = \begin{bmatrix} 5/8 \\ 3/8 \end{bmatrix}$$

So, the transition matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

The computation of a_{ij}

$$a_{11} = \frac{\text{num of transition from } s_1 \text{ to } s_1}{\text{num of transition from } s_1}$$

$$= \frac{2}{5}$$

ID:18101064

$$a_{12} = \frac{\text{Num of transition from } S_1 \text{ to } S_2}{\text{Num of transition from } S_1}$$

$$= \frac{3}{5}$$

$$a_{21} = \frac{\text{Num of transition farm } S_2 \text{ to } S_1}{\text{Num of transition from } S_2}$$

$$= \frac{2}{3}$$

$$a_{22} = \frac{\text{Num of transition from } S_2 \text{ to } S_2}{\text{Num of transition from } S_2}$$

$$= \frac{1}{3}$$

Transition probablity:

$$A = \begin{bmatrix} 2/5 & 3/5 \\ 2/3 & 1/3 \end{bmatrix}$$

Initial Probablity:

$$\pi = \begin{bmatrix} \pi_1 \\ \hline \pi_2 \end{bmatrix} = \begin{bmatrix} 5/8 \\ 3/8 \end{bmatrix}$$