# " Final Topics "

C211046

## Assignment - 01

| Serial | Age | Amount |
|---|---|---|
| A1 | 20 | 500 |
| A2 | 40 | 100 |
| A3 | 30 | 800 |
| A4 | 18 | 300 |
| A5 | 28 | 1200 |
| A6 | 35 | 1400 |
| A7 | 45 | 1800 |

Find out K-mean where $\boxed{K=2}$

Soln: Assume that

Centroid, $C_1 = A1 = (20, 500)$

Centroid, $C_2 = A2 = (40, 100)$

For, A3 $\Rightarrow$ $d_1(A_3, A_1) = \sqrt{(30-20)^2 + (800-500)^2} = 300$

$d_2(A_3, A_2) = \sqrt{(30-40)^2 + (800-100)^2} = 700$

As $\boxed{d_1 < d_2}$, it will go to Centroid $C_1$

Now, $C_1 = \dfrac{30+20}{2}, \dfrac{500+800}{2} = (25, 650)$

$\boxed{C211046}$

Centroid $C_1 = A1 = (25, 650)$

$C_2 = A2 = (40, 100)$

For A4 $\Rightarrow$ $d_1(A_4, C_1) = \sqrt{(18-25)^2 + (300-650)^2}$

$= 350$

$d_2(A_4, C_2) = \sqrt{(18-40)^2 + (300-100)^2}$

$= 201$

As, $[d_2 < d_1]$, it will go to Centroid C2

So, $C1 = A1 = (25, 650)$

$C2 = A2, A4 = \frac{40+18}{2}, \frac{100+300}{2} = (29, 200)$

For A5 $\Rightarrow$ $d_1(A5, C1) = \sqrt{(28-25)^2 + (1200-650)^2} = 550$

$d_2(A5, C2) = \sqrt{(28-29)^2 + (1200-200)^2} = 1000$

As $[d1 < d2]$, it will go to Centroid C1

Now, $C1 = A1, A5 = \frac{25+28}{2}, \frac{650+1200}{2} = (28, 925)$

$C2 = A_2 = (29, 200)$

C211046

For A6 $\Rightarrow$ $d_1(A6, C1) = \sqrt{(35-26)^2 + (1400-925)^2} = 475$

$d_2(A6, C2) = \sqrt{(35-29)^2 + (1400-200)^2} = 1200$

As $[d1 < d2]$, it will go to Centroid C1

So, $C1 = A1, A6 = \frac{26+35}{2}, \frac{925+1400}{2} = (31, 1163)$

$C2 = A2 = (29, 200)$

For A7 $\Rightarrow$ $d_1(A7, C1) = \sqrt{(45-31)^2 + (1800-1163)^2} = 637$

$d_2(A7, C2) = \sqrt{(45-29)^2 + (1800-200)^2} = 1600$

As $[d1 < d2]$, it will go to Centroid C1

Now, $C1 = A1, A7 = \frac{31+45}{2}, \frac{1800+1163}{2}$

$= (38, 1482)$

$C2 = A2 = (29, 200)$

9

C211046

# Assignment - 02

| ID | Att 1 | Att 2 | Class |
|---|---|---|---|
| 1 | 10 | 2 | Yes |
| 2 | 4 | 4 | No |
| 3 | 1 | 9 | Yes |
| 4 | 3 | 10 | Yes |
| 5 | 4 | 6 | No |
| 6 | 8 | 8 | No |
| 7 | 1 | 8 | Yes |

| ID | Att 1 | Att 2 | Class |
|---|---|---|---|
| 8 | 2 | 7 | ? |
| 9 | 7 | 7 | ? |
| 10 | 1 | 11 | ? |

$K = 3$

Find K-nn ?

Soln : Applying $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

| Euc. Dis. | 8 | 9 | 10 |
|---|---|---|---|
| 1 | 9.43 | 5.83 | 12.72 |
| 2 | 3.61 | 4.24 | 7.62 |
| 3 | 2.2 | 6.32 | 2 |
| 4 | 3.16 | 5 | 2.24 |
| 5 | 2.23 | 3.16 | 5.83 |
| 6 | 6.08 | 1.41 | 7.62 |
| 7 | 1.41 | 6.08 | 3 |

Now, 3NN(8) = (7, 3, 5) = (Yes, Yes, No) = Yes.

3NN(9) = (6, 5, 2) = (No, No, No) = No.

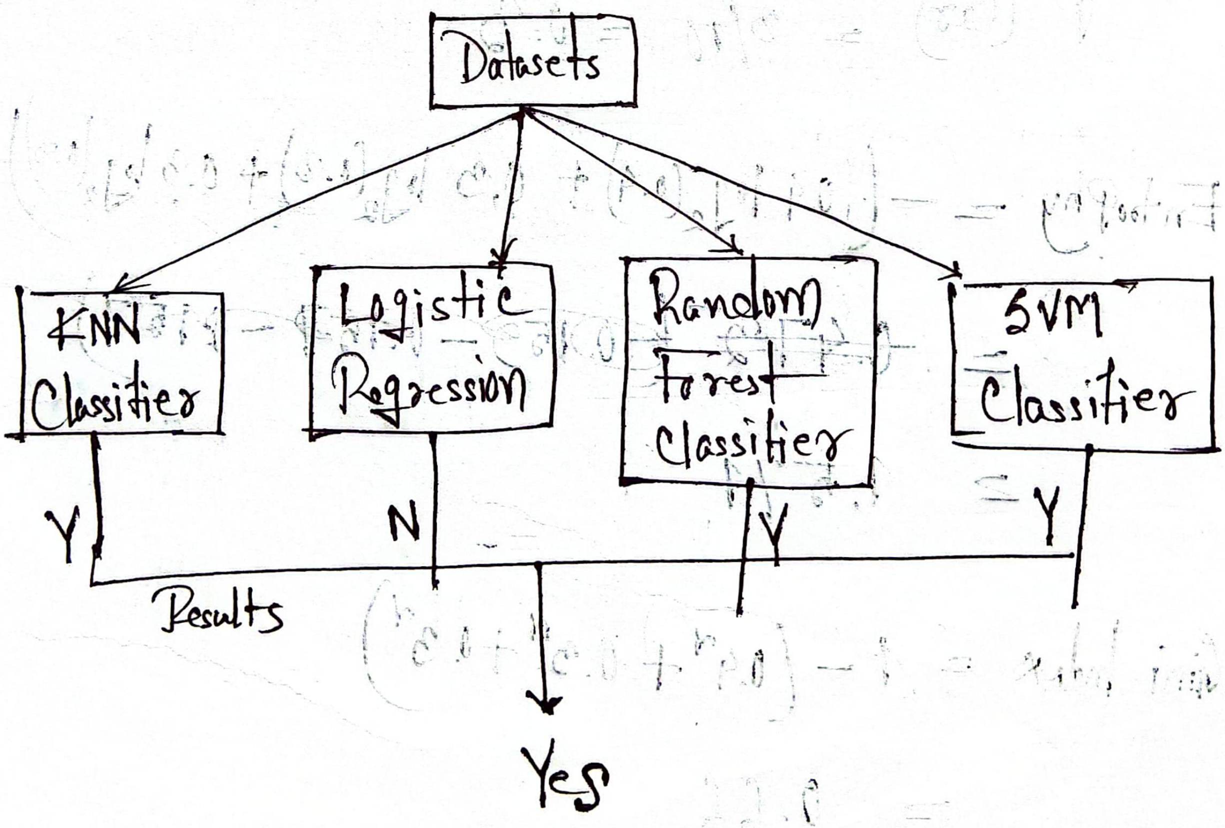
3-NN(10) = (3, 4, 7) = (Yes, Yes, Yes) = Yes.

# Naive Bayes Classification:

$$P(Y|X) = \frac{P(X|Y) * P(Y)}{P(X)}$$

$$P(Y|x_1, x_2, \ldots x_n) = \frac{P(x_1|Y) * P(x_2|Y) * \ldots P(x_n|Y) * P(Y)}{P(x_1) * P(x_2) * P(x_3) * \ldots (x_n)}$$

# Ensemble learning:



## Dicision Tree:

$$\text{Entropy} = \sum -P_j \log_2(P_j)$$

$$\text{Gini Index} = 1 - \sum P_j^2$$

**SP23(1b):**

Here, $P(\text{Bus}) = \frac{4}{10} = 0.4$

$P(\text{Train}) = \frac{3}{10} = 0.3$

$P(\text{Car}) = 3/10 = 0.3$

$$\text{Entrophy} = -\left(0.4 \log_2(0.4) + 0.3 \log_2(0.3) + 0.3 \log_2(0.3)\right)$$

$$= \cancel{0.472} \; \cancel{(-0.159 - 0.157 - 0.157)}$$

$$= 1.571$$

$$\text{Gini Index} = 1 - \left(0.4^2 + 0.3^2 + 0.3^2\right)$$

$$= 0.66$$

$$\text{Classification Error} = 1 - \max(P_j)$$

$$= 1 - \max(0.4, 0.3, 0.3) = 1 - 0.4$$

## Answer Part: Degree of impurity (Entropy)

**for Travel Cost:**

"Cheap" Travel Cost $\Rightarrow P(B) = \frac{4}{5} = 0.8$

$P(T) = 1/5 = 0.2$

$$\text{Entrophy} = -\left(0.8 \log_2(0.8) + 0.2 \log_2(0.2)\right)$$

$$= 0.722$$

"Expensive" Travel Cost $\Rightarrow P(C) = \frac{3}{3} = 1$

$$\text{Entropy} = -1 \log_2 1 = 0$$

"Standard" Travel Cost $\Rightarrow P(T) = 2/2 = 1$

$$\text{Entrophy} = -1 \log_2 1 = 0$$

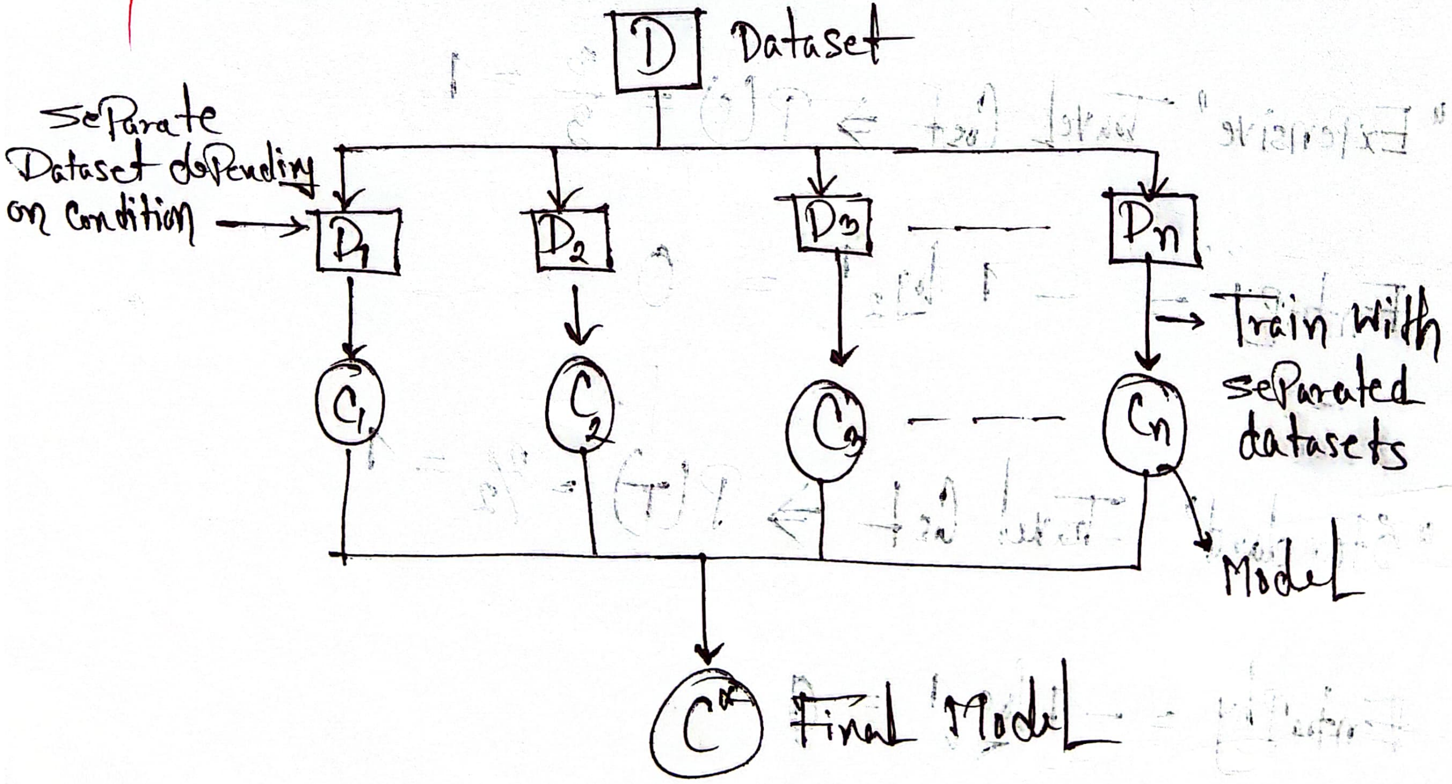
**Extra Part:**

$$\text{Information gain (Travel Cost)} = \text{Main Parent Entropy} - \text{Sum}\left(n_k/k * \text{Entropy of } k\right)$$

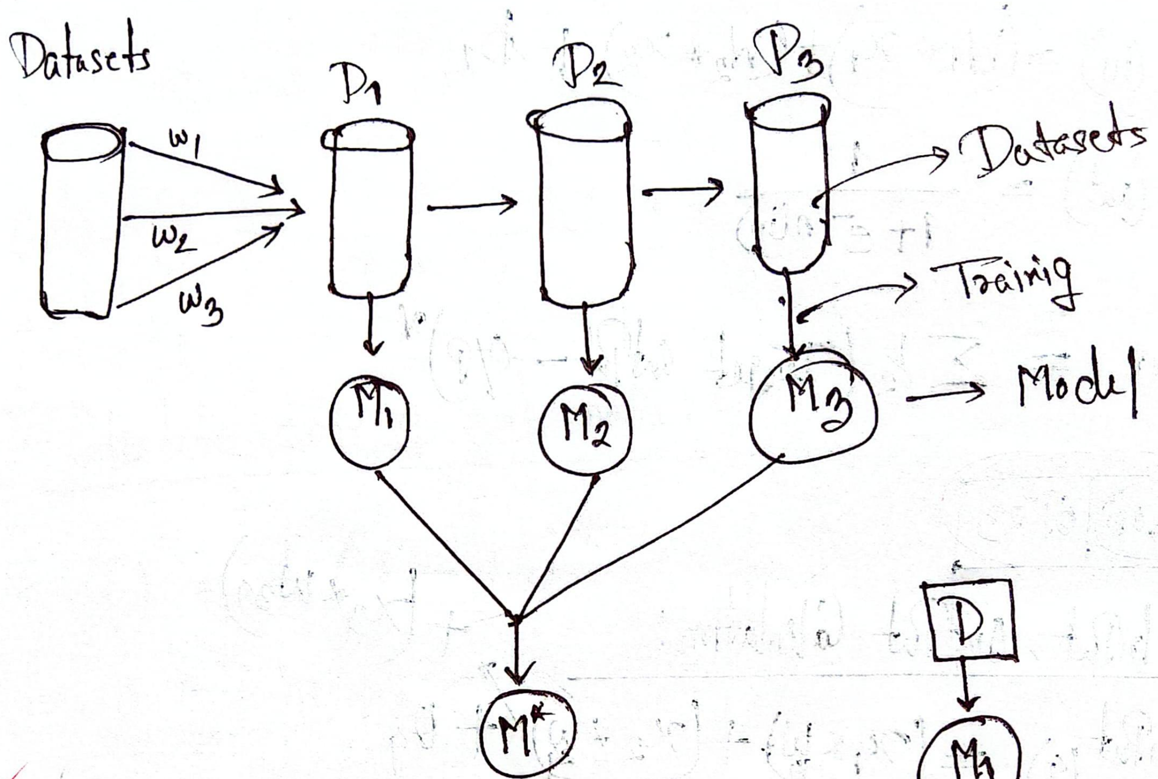
$$= 1.571 - \left(5/10 * 0.722 + 3/10 * 0 + 2/10 * 0\right)$$

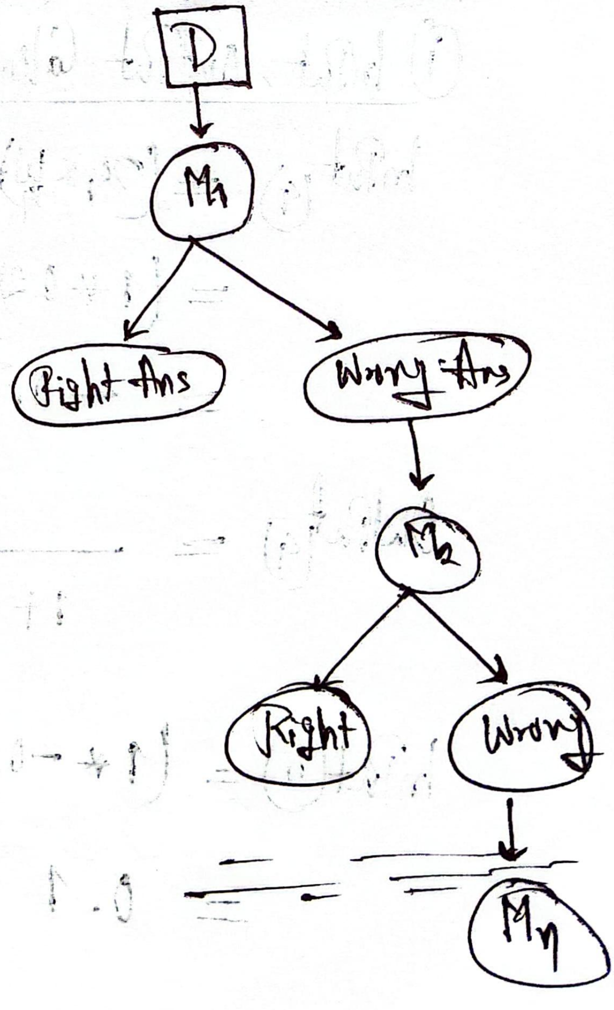
$$= 1.21$$

# Bagging/Bootstrap Aggregation of ensemble Learning:

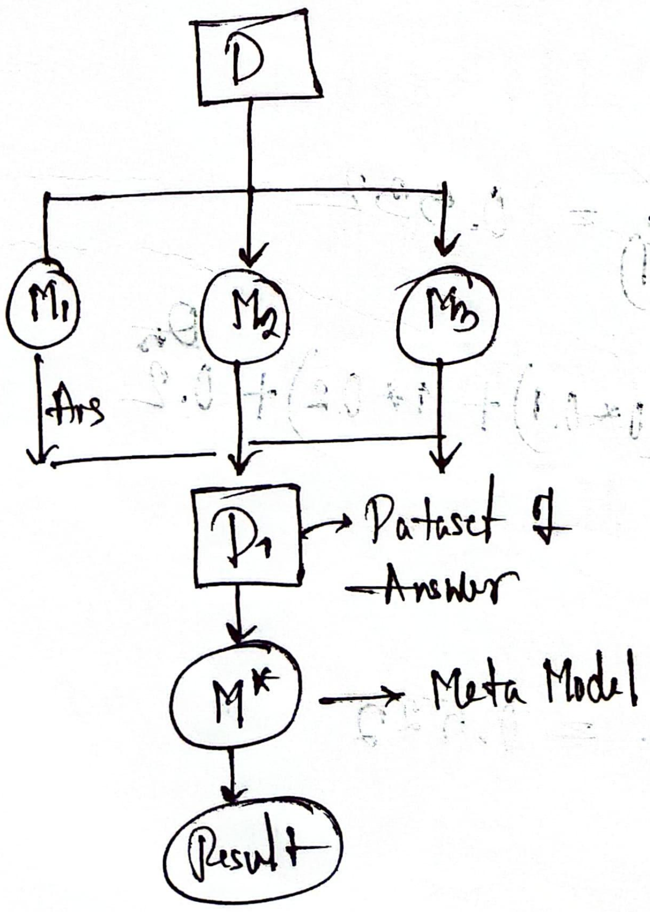


# Boosting of Ensemble Learning:




# Stacking:



# ~~#~~ Back ProPagation:

$$h_1(in) = (W_1 * x_1) + (W_2 * x_2) + b_1$$

$$h_1(out) = \frac{1}{1+e^{-h(in)}}$$

$$Error = \sum \tfrac{1}{2}(\text{Target output} - O/P)^2$$

$\boxed{\text{3b, 0n} \mid \text{SP'23}}$

## (i) InPut, OutPut Calculation:

$$InPut_{(4)} = (x_1 \times W_{14}) + (x_2 * W_{24}) + (x_3 * W_{34}) + \theta_4$$

$$= (1 * 0.2) + (0 * 0.4) + (-0.4) + (1 * -0.5)$$

$$= -0.7$$

$$OutPut_{(4)} = \frac{1}{1+e^{-(-0.7)}} = 0.332$$

$$InPut(5) = (1 * -0.3) + (0 * 0.1) + (1 * 0.2) + \underset{\theta_5}{0.2}$$

$$= 0.1$$

$$OutPut(5) = \frac{1}{1+e^{-0.1}} = 0.525$$

$$\text{Input}(6) = (0.332 * -0.3) + (0.525 * -0.2) + \overset{\theta_6}{0.1}$$

$$= -0.105$$

$$\text{Output}(6) = \frac{1}{1 + e^{.105}} = 0.479$$

## (ii) Calculate error of each node:

$$\text{Error}(6) = \frac{1}{2}(1 - 0.47)^2 \qquad \Big| \; y = 1 = \text{Target O/P}$$

$$= 0.13$$

$$\text{Error}(4) = O_4 * (1 - O_4) \cdot (E_6 * W_{46})$$

$$= 0.33 * 0.67 * 0.13 * -0.3$$

$$= -0.0086$$

$$\text{Error}(5) = O_5 * (1 - O_5) * (E_6 * W_{56})$$

$$= 0.525 * 0.975 * 0.13 * (-0.2)$$

$$= -0.0065$$

## (iii) UPdation of Weight & Bias :

For Node (6)

$$W_{46} = \eta * E_6 * O_4 + W_{46} \qquad \eta = 0.9$$

$$= 0.9 * 0.13 * 0.33 + (-0.3)$$

$$= -0.26$$

$$W_{56} = \eta * E_6 * O_5 + W_{56}$$

$$= 0.9 * 0.13 * 0.525 + (-0.2)$$

$$= -0.14$$

$$\theta_6 = \eta * E_6 + \theta_6 = .9 * .13 + 0.1$$

$$= 0.217$$

For Node (4) :

$$W_{14} = 0.9 * \overset{E_4}{(-0.0086)} * \overset{x}{1} + \overset{W_{14}}{0.2}$$

$$= 0.19$$

$$W_{24} = 0.9 * (-0.0086) * \overset{x_2}{0} + 0.24 = 0.24$$

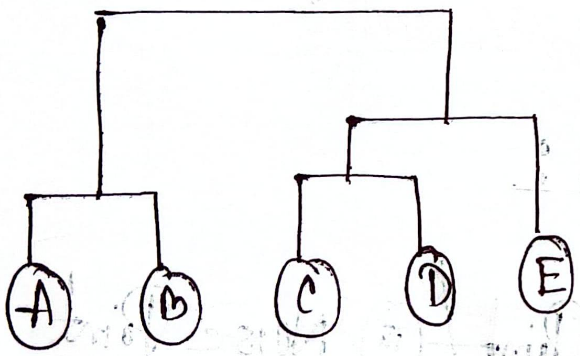
$$W_{34} = 0.9 * (-0.0086) * \overset{x_3}{1} + (-0.5) = -0.51$$

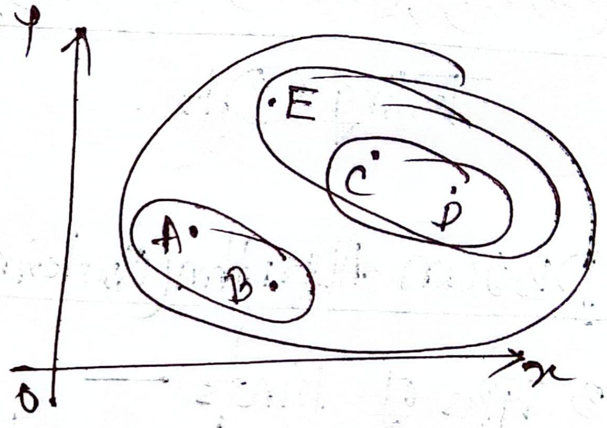
$\theta_4 = 0.9 * (-0.0086) + -0.4$

$= -0.41$

——— Same for Node 5

## # Hierarchical Clustering :




## # Cross-Validation :

| 2a | SP'23 |

Given,

(0,0,0,+), (0,0,1,−), (0,1,1,+), (0,1,1,−), (0,0,1,+), (1,0,1,−)

As K = 2

Data = 6

So, Partition = 6/2 = 3

Fold 1 :

$\underbrace{(0,0,0,+),(0,0,1,-),(0,1,1,+)}_{\text{Training Data}},\underbrace{(0,1,1,-),(0,0,1,+),(1,0,1,-)}_{\text{Testing Data.}}$

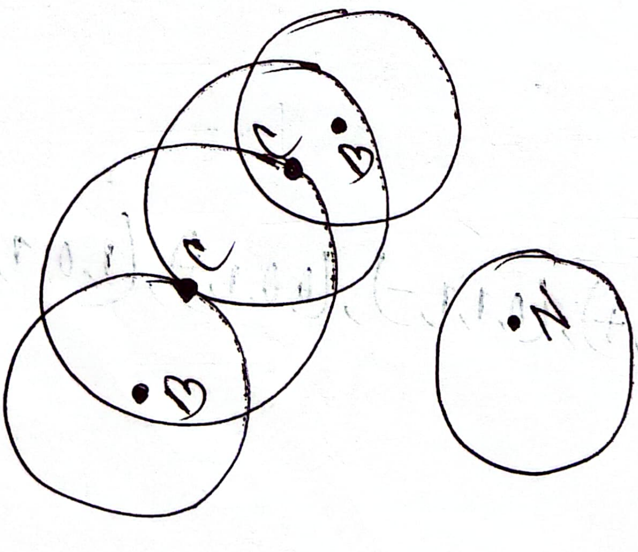
Fold 2 :

$\underbrace{(0,0,0,+),(0,0,1,-),(0,1,1,+)}_{\text{Testing Data.}},\underbrace{(0,1,1,-)(0,1,1,+),(1,0,1,-0)}_{\text{Training Data}}$

# DBSCAN Algorithm (Clustering) :

3 Types of Points —

(i) Core Point / (ii) Boundary Point / (iii) Noise Point



# Apriori Algorithm:

| Aut '22 | 5b (08) |
|---|---|

Given,

Treshold = 55%

Confidence = 75%

$$\text{Support} = \frac{55}{100} \times 5 = 2.75$$

| Items | Support |
|---|---|
| M | 3 |
| O | 4 |
| N | 2 |
| K | 5 |
| E | 4 |
| Y | 3 |
| D | 1 |
| U | 1 |
| C | 2 |
| I | 1 |

Greater than support:

| Items | Support |
|---|---|
| M | 3 |
| O | 4 |
| K | 5 |
| E | 4 |
| Y | 3 |

Generate Pair:

| Items | Support |
|---|---|
| M,O | 1 |
| M,K | 3 |
| ME | 2 |
| MY | 2 |
| OK | 3 |
| OE | 3 |
| OY | 2 |
| KE | 4 |
| KY | 3 |
| EY | 2 |

→

| ITems | Support |
|---|---|
| MK | 3 |
| OK | 3 |
| OE | 3 |
| KE | 4 |
| KY | 3 |

Tripple:

| | |
|---|---|
| MKO | 1 |
| MK~~Y~~ | 2 |
| MKE | 2 |
| OKE | 3 |
| OKY | 2 |
| OEM | 1 |
| OEY | 2 |
| KEY | 2 |

→

| Item | Support |
|---|---|
| OKE → | 3 |

Support

Appear

| Association Rule | Confidence |
|---|---|
| $O \wedge K \rightarrow E$ | $2.75/3 = .91$ — 91% |
| $O \wedge E \rightarrow K$ | $2.75/3 = .91$ — 91% |
| $K \wedge E \rightarrow O$ | $2.75/4 = .687$ — 69% |
| $E \rightarrow O \wedge K$ | $2.75/4 = .687$ — 69% |
| $K \rightarrow O \wedge E$ | $2.75/5 = .55$ — 55% |
| $O \rightarrow K \wedge E$ | $2.75/4 = .687$ — 69% |

Compare this with min-Confidence = 75%

| Rules | Confidence |
|---|---|
| $O \wedge K \rightarrow E$ | 91% |
| $O \wedge E \rightarrow K$ | 91% |

Ans //

# Non-linear SVM:

Given,

$B_1(1, 1)$ $R_1(2, 0)$

$B_2(-1, 1)$ $R_2(0, 2)$

$B_3(-1, -1)$ $R_3(-2, 0)$

$B_4(1, -1)$ $R_4(0, -2)$

$\boxed{4a \mid SP'23}$

$$\phi(x_1, x_2) = \begin{cases} \begin{pmatrix} 6 - x_1 + (x_1 - x_2)^2 \\ 6 - x_2 + (x_1 - x_2)^2 \end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} \geq 2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} & \text{otherwise} \end{cases}$$

B Points are no change since $\sqrt{x_1^2 + x_2^2} \leq 2$ for all the vectors.

For R Points,

$$\phi\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 6-2+4 \\ 6-0+4 \end{pmatrix} = \begin{pmatrix} 8 \\ 10 \end{pmatrix}$$

$$\phi\begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 6-0+4 \\ 6-2+4 \end{pmatrix} = \begin{pmatrix} 10 \\ 8 \end{pmatrix}$$

$$\phi\begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 6+2+4 \\ 6-0+4 \end{pmatrix} = \begin{pmatrix} 12 \\ 10 \end{pmatrix}$$

$$\phi\begin{pmatrix} 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 6-0+4 \\ 6+2+4 \end{pmatrix} = \begin{pmatrix} 10 \\ 12 \end{pmatrix}$$

Given, 3 support vector, $s_1 = \begin{pmatrix} 8 \\ 10 \end{pmatrix}$ $s_1' = \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix}$

$s_2 = \begin{pmatrix} 10 \\ 8 \end{pmatrix}$ $s_2' = \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix}$

$s_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $s_3' = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

Assume 3 Linear eqn,

$$\alpha_1 s_1' s_1' + \alpha_2 s_2' \cdot s_1' + \alpha_3 s_3' \cdot s_1' = 1$$

$$\alpha_1 s_1' s_2' + \alpha_2 s_2' \cdot s_2' + \alpha_3 s_3' \cdot s_2' = 1$$

$$\alpha_1 s_1' \cdot s_3' + \alpha_2 \cdot s_2' \cdot s_2' + \alpha_3 s_3' \cdot s_3' = -1$$

$$\Rightarrow 165\alpha_1 + 161\alpha_2 + 19\alpha_3 = 1$$

$$161\alpha_1 + 165\alpha_2 + 19\alpha_3 = 1$$

$$19\alpha_1 + 19\alpha_2 + 3\alpha_3 = -1$$

$$\therefore \alpha_1 = \alpha_2 = 0.859, \ \alpha_3 = -1.4219$$

Now,

$$\tilde{W} = \sum \alpha_i s_i'$$

$$= \alpha_1 \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$= (0.859) \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} + (0.859) \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} + (-1.4219) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0.1243 \\ 0.1243 \\ -1.2501 \end{pmatrix}$$

$y = Wx + b$ where $W = \begin{pmatrix} 0.1243 \\ 0.1243 \end{pmatrix} = \begin{pmatrix} 0.1243/0.1243 \\ 0.1243/0.1243 \end{pmatrix}$

$$= \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and, $b$ (offset) $= -1.2501/0.1243$

$$= -10.057$$

# Supervised vs unsupervised:

(i) Delta rule
(ii) BackPropagation
(iii) Linear Regression
(iv) SVR
(v) KNN
(vi) Decision Tree
(vii) ANN/CNN.

(i) K-mean
(ii) Hierarchical clustering
(iii) K-medoids

SP'23|2b

Here, Age : High, Low = 2 Values types

Income: Fair, Excelent = 2

Credit_Rating ; Low, High = 2

Since each features has 2 Possible value type, the value J, K = 2

Using k for Laplasian smoothing:

For Class "Yes":

Total 'Yes' = 2

$$P(Yes) = \frac{2}{5} = 0.4$$

Age - Low:

$$P(Low/Yes) = \frac{Count(Low, Yes) + 1}{Count(Yes) + K}$$

$$= \frac{0+1}{2+2} = 0.25$$

Income - Fair:

$$P(\cancel{High}\ Fair/Yes) = \frac{2+1}{2+2} = 0.75$$

Credit - Low:

$$P(Low/Yes) = \frac{1+1}{2+2} = 0.5$$

For Class No:

~~$P(Low/No) = 1 - P\left(\overset{Low}{\cancel{Yes}}/Yes\right) =$~~

$$P(No) = \frac{3}{5} = 0.6$$

Age → Low:

$$P(Low/No) = \frac{3+1}{3+2} = 0.8$$

Income → Fair:

$$P(Fair/No) = \frac{1+1}{3+2} = 0.4$$

Credit → Low:

$$P(Low/No) = \frac{0+1}{3+2} = 0.2$$

Now,

$$P(Yes/Low, Fair, Low) = \frac{P(Low/Yes) \cdot P(Fair/Yes) \cdot P(Low/Yes) \cdot P(Yes)}{\cancel{\ }}$$

$$= 0.25 * 0.75 * 0.5 * 0.4$$

$$= 0.0375$$

$P(NO \mid \cancel{Yes}\ Low, Fair, Low)$

$= P(Low|NO) \cdot P(Fair|NO) \cdot P(Low|NO) \cdot P(NO)$

$= 0.6 * 0.8 * 0.2 * 0.6$

$= 0.384$

since $(NO > Yes)$, so for (Low, fair, Low) sample the class label is "NO"

$\boxed{5a \mid SP'23}$

For Matrix 1,

$$Precision = \frac{TP}{TP+FP} = \frac{6959}{6959+96} = 0.99$$

$$Recall = \frac{TP}{TP+FN} = \frac{6959}{6945+912} = 0.94$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} = 0.96$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = 0.95$$

Do for Matrix (ii) also - ——

FAHIM REZA | C211046
fahim.bdctg51@gmail.com