Advanced Machine Learning

Types of ML :-
* Supervised learning
* Unsupervised learning
* Semisupervised learing
* Reinforcemend learing

1) Supervised :-
* It is based on supervision
* We can train the machine using the labelled datasets.
* Types
  → Classification = True / false, Hot / Cold
    → Binary Classification
    → Multiclass classification
  → Regression

2) Unsupervised learing :-
* The labels are unknown
* No need of supervision.
* It includes algorithms like K-Mean and Principle Component Analysis (PCA)
* Learning types
  → Clustering = we wand to discover the pattern and grouping of data
  → Association = we wand to discover rule that describes large portions of your data.

3) Semisupervised learning:-

* It is the combination of supervised and unsupervised learning
* Large amount of unlabelled dataset is given for the model training.

4) Reinforcement learning:-

* Learning by interacting with the environment.
* Ex: Chess game.
* In this the reward is included for better result.
* Components ⟶ Agent, Action, Environment, State, Reward.

Issues in ML:-

* Lack of data
* Noisy or poor quality data
* Non-representative training data
* Overfitting and underfitting
* Which algorithm performs best for which types of problems and representation
* Lack of skilled resourses.

Feature Selection:-

Selecting the most usefull features to train one among existing feature.

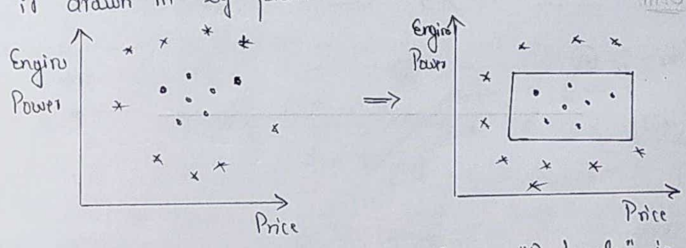
Feature Extraction:-

Combining existing features to produce a most useful features.

Hypothesis:-

1) The Classification Problem:-

* Task ⟶ Learn Class C 'family Car'
* Car has the features like color, price, ~~hors~~ Engine power, model, etc.
* Feature selection: Price and Engine Power.
* Label: 1 ⟶ Family Car
  0 ⟶ Not a family car } = r
* X1 = Price , X2 = Engine Power
* Any thing that seperate positive and negative example / label or class is called hypothesis.
* If suppose for the above example the dataset given and it drawn in xy plane.



∴ Choosen "Rectangle" is a hypothesis.

* The most specific hypothesis is one of the tightest.
* The largest is called most general hypothesis.
* The middle part b/n most specific & general hypothesis called Version Space.

If h_m possible hypothesis then

h_m : IF $x \geq m$ THEN "1" ELSE "0"

<u>Example:</u>- Consider the training data

Feature	Classes
27	1
15	0
25	1
30	1
35	1
17	0
12	0
31	1
C	0

<u>Soln:</u>-

<u>Occam Razor Principle</u> :-

VC Dimension $\rightarrow$ Maximum no. of data points which can be shattered by hypothesis space H.

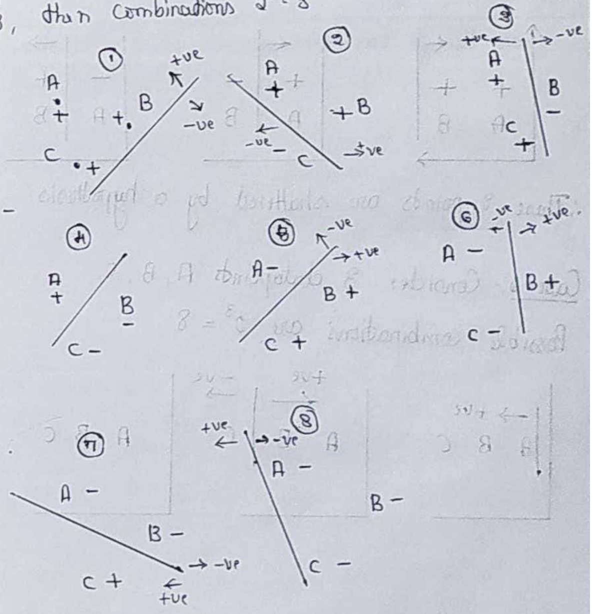
Shattering $\rightarrow$ Correctly classifying all labelling combination of the dataset.

* If we have 'n' points then maximum no. of combination is 2^n

Ex. Dichotomy $\rightarrow$ It is the partition of +ve & −ve classes.

<u>Ex:</u>- If $n=3$, then combinations $2^3 = 8$

+ + +
+ + −
+ − +
~~+ + +~~ + − −
− + +
− + −
− − +
− − −



1) Let 'X' set of all points (x, y) in a plane.

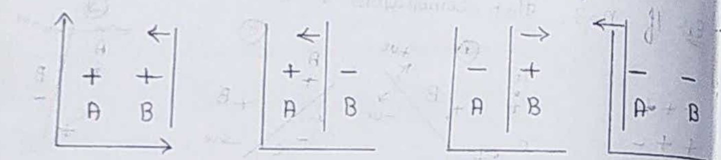
$$H = \{ h_{a,b,c}(x,y) \mid a, b, c \text{ are realnumbers} \}$$

then $h_{a,b,c}(x,y) = \begin{cases} 1, & \text{if } ax+by+c > 0 \\ 0, & \text{otherwise} \end{cases}$

Soln:- Case-1:-

Consider 'D' = {A, B} so two points.

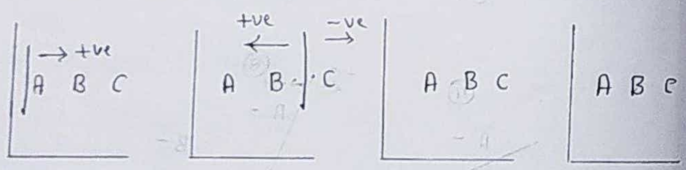
Possible combinations are $2^2 = 4$ (Dichotomy)



∴ These 2 points are shattered by a hypothesis.

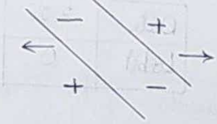
Case-2:- Consider 3 datapoints A, B, C.

Possible combinations are $2^3 = 8$



Case-3:- Consider 4 points A, B, C, D.

We can see that there is no hypothesis which is consistent with the following dichotomy



So the 4 element cannot be shattered by H.

∴ So VC(H) = 3

2) Let the instance space X be the set of all real number.

$H = \{ h_m : m \text{ is a real number} \}$ and h_m is

h_m : If $x \geq m$ then 1 else 0

Find all the consistent hypothesis and VC dimension of H.

Soln:- Case-1:-

Let D = {3.25} => $2^1 = 2$

(a)

Data	3.25
Label	0

(b)

Data	3.25
Label	1

Let m = 4

3.25 ≥ 4 => False = 0

Let m = 3

3.25 ≥ 3 => True = 1

~~3.25~~

∴ h_4 is consistent hypothesis for 'a' and h_3 is consistent hypothesis for 'b'

So this dataset shattered by H.

Case-2:- Let $D = \{3.25, 4.75\} \Rightarrow 2^2 = 4$.

(a)

Data	3.25	4.75
Label	0	0

(b)

Data	3.25	4.75
Label	0	1

(c)

Data	3.25	4.75
Label	1	0

(d)

Data	3.25	4.75
Label	1	1

For (a) Let $m = 5$

$3.25 \geq 5 \Rightarrow$ False $= 0$

$4.75 \geq 5 \Rightarrow$ False $= 0$

(b) Let $m = 4$

$3.25 \geq 4 \Rightarrow$ False $= 0$

$4.75 \geq 4 \Rightarrow$ True $= 1$

(c) Not consistant hypothesis.

(d) Let $m = 2$.

$3.25 \geq 2 \Rightarrow$ True $= 1$

$4.75 \geq 2 \Rightarrow$ True $= 1$

The hypothesis h_5 is consistant with dichotomy (a), h_4 is consistant with (b), h

but no consistant hypothesis with dichotomy (c)

Thus the data point $\{3.25, 4.75\}$ is not shattered with the hypothesis.

$\therefore$ VC(H) = 1

3) Find VC dimension of H,

$$h_{(a,b)}(x) = \begin{cases} 1, & \text{if } a < x < b \\ 0, & \text{otherwise.} \end{cases}$$

Soln:- Here 2 elements are there, so $2^2 = 4$.

Let $D = \{3.25, 4.75\}$.

(a)

Data	3.25	4.75
Label	0	0

Let $(a, b) = (5, 6)$ then

$5 < 3.25 < 6 \Rightarrow F \Rightarrow 0$

$5 < 4.75 < 6 \Rightarrow F \Rightarrow 0$

(b)

Data	3.25	4.75
Label	0	1

Let $(a, b) = (4, 5)$

$4 < 3.25 < 5 \rightarrow F \Rightarrow 0$

$4 < 4.75 < 5 \rightarrow T \Rightarrow 1$

Data	3.25	4.75
Label	1	0

Let $(a, b) = (3, 4)$

$3 < 3.25 < 4 \rightarrow T \Rightarrow 1$

$3 < 4.75 < 4 \rightarrow F \Rightarrow 0$

Data	3.25	4.75
Label	1	1

Let $(a, b) = (3, 5)$

$3 < 3.25 < 5 \Rightarrow T \Rightarrow 1$

$5 < 4.75 < 5 \Rightarrow T \Rightarrow 1$

The hypothesis $h_{5,6}$ is consistand, $h_{4,5}$ is consistand, $h_{3,4}$ is consistand and $h_{3,5}$ is consistand with ⓐ ⓑ, ⓒ & ⓓ respectively.

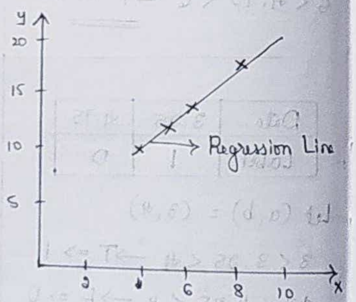
$\therefore VC(H) = 2$

Classification and Regression:-

* Classification is supervised machine learning method where model is tries to predict correct label.
* Characteristics → Discrete Output, Supervised Learning, Common Algorithm, Evaluation Metrics.

* Regression aims to predict continuous numerical values.
* In regression, the output is a real number that can fall within a range.
* It is used to find the relationship b/n input features & the target variable.

Exper(x)	Salary (y)
4	10
5	12
6	14
8	18



∴ This line eqn → $\overset{2x+2}{mx+b} = y$.

* If we want to find value for 10, then, $y = 2\times 10 + 2 = \underline{22}$

1> Simple Linear Regression:

$$Y = ax + b$$

$$a = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$b = \frac{\sum y - a\sum x}{n}$$

$x \to$ Independent variable

$y \to$ Dependent variable

$n \to$ no. of data instands.

Model Evaluation:-

* Best fit line is selected based on the least "residual error"
* Sum of Squared Error (SSE),

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Ex:-

Temp (°c)	Ice Cream Sales (lt)
20	13
25	21
30	25
35	35
40	38

Find a regression line that predict the ice cream sales at 28°C.

Soln:-

Step-1:- For each data instance or point (x,y) calculate x^2 & xy

Step-2:- Find the sum & $n = 5$

x	y	x^2	xy
20	13	400	260
25	21	625	525
30	25	900	750
35	35	1225	1225
40	38	1600	1520
$\Sigma =$ 150	132	4750	4280

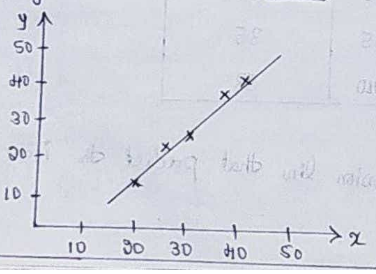
$$a = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{5 \times 4280 - (150 \times 132)}{5 \times 4750 - (150)^2}$$

$$= \cancel{12.412}\ 1.28$$

$$b = \frac{\Sigma y - a\Sigma x}{n} = \frac{132 - 1.28 \times 150}{5} = -12$$

$\therefore\ y = ax + b = 1.28x - 12$

$\therefore\ x = 28^\circ C$, $y = 1.28 \times 28 - 12 = 23.84$ Litres.



Find the equation on regression lines for the following data.

Regration Equation of X on Y:

$$X - \bar{X} = b_{xy}(Y - \bar{Y})$$

$$b_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma Y^2 - (\Sigma Y)^2}$$

Regression Equation of Y on X:-

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$b_{yx} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma X^2 - (\Sigma X)^2}$$

i) Calculate regression coefficient & obtain the lines of regression for the following data

X	1	2	3	4	5	6	7
Y	9	8	10	12	11	13	14

Soln:-

X	Y	X^2	Y^2	XY
1	9	1	81	9
2	8	4	64	16
3	10	9	100	30
4	12	16	144	48
5	11	25	121	55
6	13	36	169	78
7	14	49	196	98
$\Sigma =$ 28	77	140	875	334

Regression equation of X on Y

$$b_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma Y^2 - (\Sigma Y)^2} = \frac{7\times 334 - 28\times 77}{7\times 875 - (77)^2}$$

$$= \frac{13}{14}$$

$$\therefore X - \bar{X} = b_{xy}(Y - \bar{Y}) = \frac{13}{14}(Y - \bar{Y}) \quad \text{—— (1)}$$

Regression equation of Y on X

$$b_{yx} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma X^2 - (\Sigma X)^2} = \frac{7\times 334 - 28\times 77}{7\times 140 - (28)^2}$$

$$= \frac{13}{14} = 0.929$$

$$\therefore Y - \bar{Y} = b_{yx}(X - \bar{X}) = \frac{13}{14}(X - \bar{X}) \quad \text{—— (2)}$$

$$= \cancel{\frac{13}{14}\left(\frac{13}{14}(Y-\bar{Y})\right)}$$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{28}{7} = 4, \quad \bar{Y} = \frac{\Sigma Y}{N} = \frac{77}{7} = 11$$

$$\therefore (1) \Rightarrow X - 4 = 0.929(Y - 11) = 0.929Y - 10.219$$

$$\therefore X = 0.929Y - 6.219$$

∴ This is the equation of X on Y.

$$\therefore (2) \quad Y - \bar{Y} = 0.929(X - \bar{X})$$

$$Y - 11 = 0.929(X - 4) = 0.929X - 3.716$$

$$Y = 0.929X + 7.284$$

∴ This is the regression equation of Y on X

2) For 5 pairs of observation the following results are obtained.

$\Sigma X = 15$, $\Sigma Y = 25$, $\Sigma X^2 = 55$, $\Sigma Y^2 = 135$, $\Sigma XY = 83$.

Find the equation of lines of regression and estimate the value of X on the first line when $Y = 12$ and value of Y on the 2nd line if $X = 8$.

Soln:-

$$\bar{X} = \frac{\Sigma X}{N} = \frac{15}{5} = 3, \quad \bar{Y} = \frac{\Sigma Y}{N} = \frac{25}{5} = 5$$

∴ Regression equation of X on Y:-

$$b_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma Y^2 - (\Sigma Y)^2} = \frac{5\times 83 - 15\times 25}{5\times 135 - 25^2} = 0.8$$

$$\therefore X - \bar{X} = b_{xy}(Y - \bar{Y})$$

$$X - 3 = 0.8(Y - 5) \Rightarrow X - 3 = 0.8Y - 4$$

$$\therefore X = 0.8Y - 1$$

when $Y = 12$,

$$X = 0.8\times 12 - 1 = 8.6$$

Regression equation of Y on X:-

$$b_{yx} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma X^2 - (\Sigma X)^2} = \frac{5\times 83 - 15\times 25}{5\times 55 - (15)^2} = 0.8$$

$\therefore Y - \overline{Y} = b_{yx}(X - \overline{X})$

$Y - 5 = 0.8(X - 3) \Rightarrow Y - 5 = 0.8X - 2.4$

$\therefore \underline{Y = 0.8X + 2.6}$

when $X = 8$,

$Y = 0.8\times 8 + 2.6 = \underline{9}$

2) Multiple Linear Regression:-

- In this one dependent variable based on two or more independent variable

$$\underbrace{Y}_{\text{Dependent variable}} = \underbrace{\beta_0}_{\text{Y-intercept}} + \underbrace{\beta_1 X_1 + \beta_2 X_2}_{\text{Slope coefficient}} + \dots + \underbrace{\epsilon}_{\text{Error rate}}$$

Independent variable

- Multicollinearity is the strong relationship between the variables

Ex:- Age - DOB.

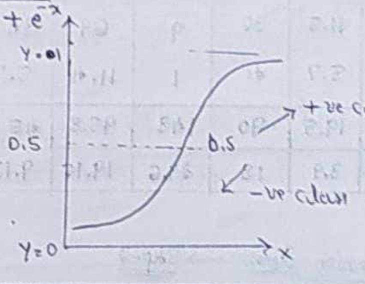
3) Polynomial Regression:-

$$Y = b_0 + b_1 x_1 + b_2 x^2 + \dots + b_n x^n$$

4) Logistic regression:-

- Supervised learning algorithm that accomplishes binary classification tasks by predicting the probability of an outcome, event or observation
- Outcomes are $\Rightarrow$ 1/0, True/False, Yes/No

$$g(x) = \frac{1}{1+e^{-x}}$$



Predict the value of Y for subject 6 from the given dataset that contains values for X_1, X_2 and Y by using Multiple Regression Model.

Subj	X_1	X_2	Y
1	3	8	-3.7
2	4	5	3.5
3	5	7	2.5
4	6	3	11.5
5	2	1	5.7
6	3	2	?

Soln:

Step-1:

	X_1	X_2	Y	X_1^2	X_2^2	X_1Y	X_2Y	X_1X_2
	3	8	-3.7	9	64	-11.1	-29.6	24
	4	5	3.5	16	25	14	17.5	20
	5	7	2.5	25	49	12.5	17.5	35
	6	3	11.5	36	9	69	34.5	18
	2	1	5.7	4	1	11.4	5.7	2
Σ:	20	24	19.5	90	148	95.8	45.6	99
Mean	4	4.8	3.9	18	29.6	19.16	9.12	19.8

Calculate Regression Sums: $\longrightarrow$ Step-2

$$\Sigma x_1^2 = \Sigma X_1^2 - \frac{(\Sigma X_1)^2}{N} = 90 - \frac{(20)^2}{5} = \underline{10}$$

$$\Sigma x_2^2 = \Sigma X_2^2 - \frac{(\Sigma X_2)^2}{N} = 148 - \frac{(24)^2}{5} = \underline{32.8}$$

$$\Sigma x_1 y = \Sigma X_1 Y - \frac{\Sigma X_1 \Sigma Y}{N} = 95.8 - \frac{20 \times \overset{19.5}{\cancel{24}}}{5} = \underline{17.8}$$

$$\Sigma x_2 y = \Sigma X_2 Y - \frac{\Sigma X_2 \Sigma Y}{N} = 45.6 - \frac{24 \times 19.5}{5} = \underline{-48}$$

$$\Sigma x_1 x_2 = \Sigma X_1 X_2 - \frac{\Sigma X_1 \Sigma X_2}{N} = 99 - \frac{20 \times 24}{5} = \underline{3}$$

Step-3: Calculate β_0, β_1 and β_2

$$\therefore Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad \text{—— (1)}$$

$$\therefore \beta_1 = \frac{(\Sigma x_2^2)(\Sigma x_1 y) - (\Sigma x_1 x_2)(\Sigma x_2 y)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

$$= \frac{32.8 \times 17.8 - 3 \times (-48)}{10 \times 32.8 - 3^2} = \underline{2.2816}$$

$$\therefore \beta_2 = \frac{(\Sigma x_1^2)(\Sigma x_2 y) - (\Sigma x_1 x_2)(\Sigma x_1 y)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

$$= \frac{10 \times (-48) - 3 \times 17.8}{10 \times 32.8 - 3^2} = \underline{-1.6721}$$

$$\therefore \beta_0 = \overline{Y} - \beta_1 \overline{X}_1 - \beta_2 \overline{X}_2$$

$$= 3.9 - 2.2816 \times 4 - (-1.6721) \times 4.8$$

$$= \underline{2.7997}$$

$\therefore$ (1) $\Rightarrow$

$$Y = \underline{2.7997 + 2.2816\, X_1 - 1.6721\, X_2}$$

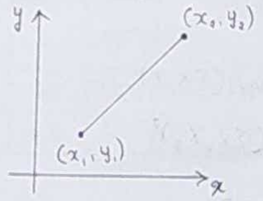
For $X_1 = 3$ and $X_2 = 2$

$$\therefore Y = 2.7997 + 2.2816 \times 3 - 1.6721 \times 2$$

$$= \underline{6.3003}$$

K-Nearest Neighbor (KNN) Algorithm:-

* It is a non parametric supervised learning method.
* The K values are the odd numbers, $K = 3, 5, 7, \ldots$
* The KNN works by finding the K-nearest neighbors to a given data point based on a distance metric, such as Euclidean Distance.



Manhattan Distance,

$$L_1 = |x_2 - x_1| + |y_2 - y_1|$$

Euclidean Distance

$$L_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Solve using KNN algorithm using Euclidean distance and $K=3$

Name	Age	Gender	Sport
A	32	M	Football
B	40	M	Neither
C	16	F	Cricket
D	34	F	Cricket
E	55	M	Neither
F	40	M	Cricket
G	20	F	Neither
H	15	M	Cricket
I	55	F	Football
J	15	M	Football
Anjali	15	F	?

Soln:- Given, $X_1 = 5$, $Y_1 = 1$ & $M = 0$, $F = 1$

Name	Age	Gender	Distance	Sport	Min values of distance
A	32	0	27.018	Football	
B	40	0	35.014	Neither	
C	16	1	11.~~045~~	Cricket	3
D	34	1	29	Cricket	
E	55	0	50.009	Neither	
F	40	0	35.014	Cricket	
G	20	1	15	Neither	
H	15	0	10.049	Cricket	2
I	55	1	50	Football	
J	15	0	10.049	Football	1

∴ Select the small 3 distance because $k = 3$.

∴ Majority of those 3 is Cricket.

∴ Sport is calculated for Anjali, 15, F is "Cricket"

Perform KNN-Classification Algorithm on following dataset and predict the class for $P_1 = 3$ and $P_2 = 7$ and $k = 3$.

P_1	P_2	Class	Distance	
7	7	False	4	3
7	4	False	5	
3	4	True	3	2
1	4	True	3.605	1

∴ Majority is True ∴ Predicted Class is "True"

Feature Selection:-

* Selecting the best features helps the model to perform well.
* Used to
  → Reduce the dimensionality of feature space
  → To speedup the learning algorithm
  → Improve the accuracy.
  → Avoid overfitting

Methods:-

1) Filter Method:-

* Filter out the irrelevant features and redundant columns from the dataset

* Techniques:-
  * Missing values → Some feature selection methods can handle missing data by ignoring it, impiting it or using it as a feature.
  * Information Gain → The amount of information provided by the feature for identifying the target values.
  * Chi-Square Test → Is generally used to test the relationship b/n categorical variables.

$$\chi^2 = \sum \frac{(\text{Observed value} - \text{Expected value})^2}{\text{Expected value}}$$

2) Wrapper Method:-

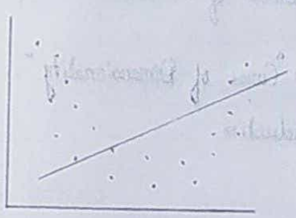
* Also refered as greedy algorithm train the algorithm by using a subset of features in an iterative manner.

* Techniques:-
  * Forward Selection → Initially start with an null set of features & keep adding a feature which best improve the model.
  * Backward selection
  * Bidirectional Elimination

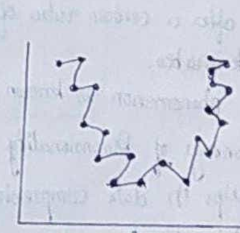
Overfitting, Underfitting:-

Underfitted → It cannot identified the dominating trend in the data resulting in training mistakes and poor model performance.

Overfitting → In training the model shows good accuracy and in testing the model shows poor accuracy



Underfitting



Overfitting

* The bias in known as the difference between the predicted value and the actual value in ML model.
* High bias (error) → Underfitting
* Variance is the spread between the numbers in a data set.

How to prevent overfitting?

* Hold-out
* Cross-validation
* Data augmentation
* Feature Selection
* L1/L2 regularization
* Remove layers
* Drop out

Hyper Parameter → The feature which is fixed by a person not the ML model.

Dimensionality Reduction :-

* The number of attributes, features or input variable of a data set is known as dimensionality
* The predictive power of any classifier increases as the no. of dimensions increases.
* But after a certain value of number of dimensions the performance deteriorates.
* This phenomenon is known as "Curse of Dimensionality"
* Advantages of Dimensionality Reduction
  → Helps in data compression
  → Reduced storage space
* Methods are,
  → Principle Component Analysis (PCA)
  → Linear Discriminant Analysis (LDA)
  → Generalized Discriminant Analysis (GDA)

1

PCA:-

* It is a unsupervised method.
* Used to reduce the dimensionality of large data set.
* Steps:
  → Standardize the data
  → Calculate the covariance matrix
  → Find the eigenvalues and eigenvectors of the covariance matrix
  → Plot the eigenvectors / principal components over the scaled data

Given the following data, use PCA to reduce the dimension from 2 to 1.

Feature	Ex_1	Ex_2	Ex_3	Ex_4
x	4	8	13	7
y	11	4	5	14

Soln:-

Step-1:- Dataset

Feature	Ex_1	E_{x_2}	E_{x_3}	E_{x_4}
x	4	8	13	7
y	11	4	5	14

No. of features = 2 = n

No. of samples, N = 4

Step-2: Computation of mean of variables.

$$\bar{x} = \frac{4+8+13+7}{4} = \underline{8}$$

$$\bar{y} = \frac{11+4+5+14}{4} = \underline{8.5}$$

Step-3: Computation of covariance matrix

Ordered Pairs = $(x,x), (x,y), (y,x), (y,y)$

i) $$\boxed{\therefore Cov(x_i, y_j) = \frac{1}{N-1}\sum_{K=1}^{N}(x_{iK}-\bar{x}_i)(x_{jK}-\bar{x}_j)}$$

$$Cov(x,x) = \frac{1}{4-1}\sum_{K=}\left[(4-8)^2+(8-8)^2+(13-8)^2+(7-8)^2\right]$$

$$= \underline{14}$$

$$Cov(y,y) = \frac{1}{4-1}\left[(11-8.5)^2+(4-8.5)^2+(5-8.5)^2+(14-8.5)^2\right]$$

$$= \underline{23}$$

$$Cov(x,y) = \frac{1}{4-1}\left[(4-8)(11-8.5)+(8-8)(4-8.5)+(13-8)(5-8.5)+(7-8)(14-8.5)\right]$$

$$= \underline{-11}$$

$\therefore Cov(y,x) = Cov(x,y)$

ii) Covariance Matrix, $S = \begin{bmatrix} Cov(x,x) & Cov(x,y) \\ Cov(y,x) & Cov(y,y) \end{bmatrix}$

$$\therefore S = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$$

Step-4: Find Eigen value, Eigen vector and Normalized Eigen vectors.

i) Eigen Value,

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

$$S-\lambda I = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix}$$

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

$$\begin{vmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{vmatrix} = 0$$

$$(14-\lambda)(23-\lambda) - 121 = 0$$

$$322 - 14\lambda - 23\lambda + \lambda^2 - 121 = 0$$

$$\therefore \lambda^2 - 37\lambda + 201 = 0$$

$$\therefore \lambda_1 = \underline{30.3849} \quad \& \quad \lambda_2 = \underline{6.6151}$$

$\therefore$ Arrange value in descending order so that λ_1 get larger value.

ii) Eigen Vector,

$\underline{\lambda_1 = 30.3849} \Rightarrow (S-\lambda_1 I)U_1 = 0$

$$\begin{bmatrix} 14-30.3849 & -11 \\ -11 & 23-30.3849 \end{bmatrix}\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = 0$$

$$\begin{bmatrix} -16.3849 & -11 \\ -11 & -7.3849 \end{bmatrix}\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = 0$$

$-16.3849 U_1 - 11 U_2 = 0 \Rightarrow -16.3849 U_1 = 11 U_2$ — ①

$-11 U_1 - 7.3849 U_2 = 0 \Rightarrow -11 U_1 = 7.3849 U_2$ — ②

① $\Rightarrow$ $\frac{-16.3849}{U_2} = \frac{11}{U_1} = t$ (assume)

if $t = 1$

$\frac{-16.3849}{U_2} = \frac{11}{U_1} = 1$

$\therefore \frac{11}{U_1} = 1 \Rightarrow U_1 = 11$, $U_2 = -16.3849$

$\therefore \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.3849 \end{bmatrix}$

iii) Normalize the eigen vector U_1

$$e_1 = \begin{bmatrix} \frac{11}{\sqrt{11^2 + (-16.3849)^2}} \\ \frac{-16.3849}{\sqrt{11^2 + (-16.3849)^2}} \end{bmatrix} = \begin{bmatrix} 0.5574 \\ 0.8303 \end{bmatrix}$$

Step-5:- Derive the new dataset

	E_{x_1}	E_{x_2}	E_{x_3}	E_{x_4}
First Principal Component	P_{11}	P_{12}	P_{13}	P_{13}

$\therefore P_{11} = e_1^T \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} -4 \\ 2.5 \end{bmatrix}$

$= -4.3054$

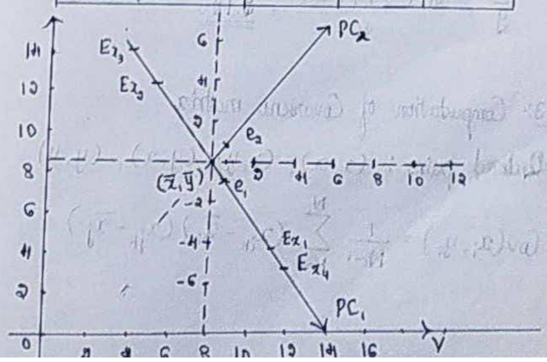
$\therefore P_{12} = e_1^T \begin{bmatrix} 8-8 \\ 4-8.5 \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 0 \\ -4.5 \end{bmatrix} = 3.7363$

$\therefore P_{13} = e_1^T \begin{bmatrix} 13-8 \\ 5-8.5 \end{bmatrix} =$

$\therefore P_{14} =$

$\therefore$ Final Dataset

PC_1	E_{x_1}	E_{x_2}	E_{x_3}	E_{x_4}
	-4.3052	3.7363	5.6928	-5.1238



Solve the given dataset using PCA method to reduce dimension 2 to 1.

Feature	E_{x_1}	E_{x_2}	E_{x_3}	E_{x_4}
x	2	1	0	-1
y	4	3	1	0.5

Soln:-

Step-1:- Dataset

Features	E_{x_1}	E_{x_2}	E_{x_3}	E_{x_4}
x	2	1	0	-1
y	4	3	1	0.5

∴ No of features, $n = 2$

∴ No. of samples, $N = 4$

Step-2:- Computation of Means of variables.

$$\bar{x} = \frac{2+1+0+(-1)}{4} = \underline{0.5}$$

$$\bar{y} = \frac{4+3+1+0.5}{4} = \underline{2.125}$$

Step-3:- Computation of Covariance matrix

Ordered pairs : (x,x), (x,y), (y,x), (y,y)

i) $$Cov(x_i, y_i) = \frac{1}{N-1}\sum_{k=1}^{N}(x_{ik}-\bar{x}_i)(x_{jk}-\bar{x}_j)$$

$$\therefore Cov(x,x) = \frac{1}{4-1}\left[(2-0.5)^2+(1-0.5)^2+(0-0.5)^2+(-1-0.5)^2\right]$$

$$= \underline{\underline{1.6667}}$$

$$\therefore Cov(y,y) = \frac{1}{4-1}\left[(4-2.125)^2+(3-2.125)^2+(1-2.125)^2+(0.5-2.125)^2\right]$$

$$= \underline{\underline{2.7292}}$$

$$\therefore Cov(x,y) = Cov(y,x) = \frac{1}{4-1}\left[(2-0.5)(4-2.125)+(1-0.5)(3-2.125)+(0-0.5)(1-2.125)+(-1-0.5)(0.5-2.125)\right]$$

$$= \underline{\underline{2.0833}}$$

ii) Covariance Matrix,

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

$$= \begin{bmatrix} 1.6667 & 2.0833 \\ 2.0833 & 2.7292 \end{bmatrix}$$

Step-3:- Find the Eigen value and Eigen vectors and Normalize the Eigen vectors.

i) Eigen Values, $\det(S - \lambda I) = 0$

$$S - \lambda I = \begin{bmatrix} 1.6667 & 2.0833 \\ 2.0833 & 2.7292 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 1.6667-\lambda & 2.0833 \\ 2.0833 & 2.7292-\lambda \end{bmatrix}$$

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

$$\begin{vmatrix} 1.6667-\lambda & 2.0833 \\ 2.0833 & 2.7292-\lambda \end{vmatrix} = 0$$

$$(1.6667-\lambda)(2.7292-\lambda) - (2.0833)^2 = 0$$

$$4.5488 - 1.6667\lambda - 2.7292\lambda + \lambda^2 - 4.3401 = 0$$

$$\lambda^2 - 4.3959\lambda + 0.2087 = 0$$

$$\therefore \underline{\lambda_1 = 4.3479} \quad \& \quad \underline{\lambda_2 = 0.048}$$

ii) Eigen Vector, $\lambda_1 = 4.3479$

$$(S - \lambda I)U_1 = 0$$

$$\begin{bmatrix} 1.6667 - 4.3479 & 2.0833 \\ 2.0833 & 2.7292 - 4.3479 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = 0$$

$$\begin{bmatrix} -2.6812 & 2.0833 \\ 2.0833 & -1.6187 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = 0$$

$$-2.6812U_1 + 2.0833U_2 = 0 \quad \text{— (1)}$$

$$2.0833U_1 - 1.6187U_2 = 0 \quad \text{— (2)}$$

(1) $\Rightarrow$ $2.6812\,U_1 = 2.0833\,U_2$

$$2.6812/U_2 = 2.0833/U_1 = t \quad (\text{assume } t)$$

if $t = 1$, $\dfrac{2.6812}{U_2} = \dfrac{2.0833}{U_1} = 1$

$\therefore \underline{U_1 = 2.0833}$ & $\underline{U_2 = 2.6812}$

$$\therefore \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 2.0833 \\ 2.6812 \end{bmatrix}$$

iii) Normalize the Eigen vector,

$$\therefore e_1 = \begin{bmatrix} \dfrac{2.0833}{\sqrt{2.0833^2 + 2.6812^2}} \\ \dfrac{2.6812}{\sqrt{2.0833^2 + 2.6812^2}} \end{bmatrix} = \underline{\begin{bmatrix} 0.6136 \\ 0.7896 \end{bmatrix}}$$

Step-5:- Derive the new dataset

	Ea_1	Ea_2	Ea_3	Ea_4
First Principle Component	P_{11}	P_{12}	P_{13}	P_{14}

$$\therefore P_{11} = e_1^T \begin{bmatrix} 2-0.5 \\ 4-2.125 \end{bmatrix} = \begin{bmatrix} 0.6136 & 0.7896 \end{bmatrix} \begin{bmatrix} 1.5 \\ 1.875 \end{bmatrix} = \underline{\underline{2.4009}}$$

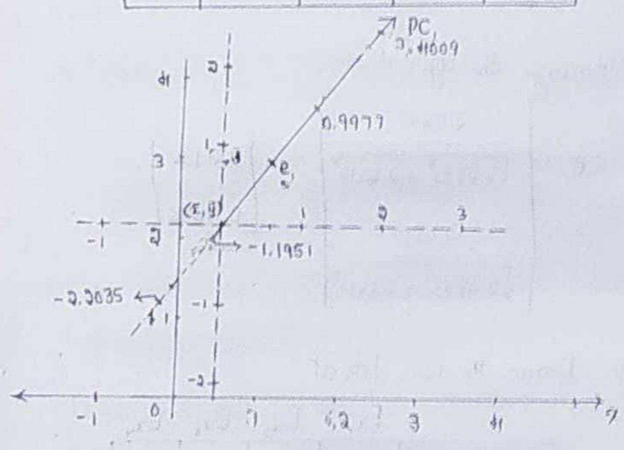
$$\therefore P_{12} = e_1^T \begin{bmatrix} 1-0.5 \\ 3-2.125 \end{bmatrix} = \begin{bmatrix} 0.6136 & 0.7896 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.875 \end{bmatrix} = \underline{\underline{0.9977}}$$

$$\therefore P_{12} = e_1^T \begin{bmatrix} 0-0.5 \\ 1-2.125 \end{bmatrix} = \begin{bmatrix} 0.6136 & 0.7896 \end{bmatrix} \begin{bmatrix} -0.5 \\ -1.125 \end{bmatrix} = \underline{\underline{-1.1951}}$$

$$\therefore P_{14} = e_1^T \begin{bmatrix} -1-0.5 \\ 0.5-2.125 \end{bmatrix} = \begin{bmatrix} 0.6136 & 0.7896 \end{bmatrix} \begin{bmatrix} -1.5 \\ -1.625 \end{bmatrix} = \underline{-2.2035}$$

$\therefore$ Final Dataset

PC_1	Ex_1	Ex_2	Ex_3	Ex_4
	2.41009	0.9977	-1.1951	-2.2035



Linear Discriminant Analysis (LDA):

* LDA is a technique that transforms a set of features or variable into smaller set of new feature
* Supervised learning is holds
* It seperate multiple classes with multiple labels
* Use → Dimensionality reduction
  ↳ Feature extraction
  ↳ Improving classification accuracy
  ↳ Handle multiple class problem
  ↳ Reducing overfitting

* Two criteria
  ↳ Maximize the distance b/n mean of 2 variable
  ↳ Minimize the variation within each classes

* Between-class scatter matrix

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

* Within-class scatter matrix

$$S_1 = \sum_{x \in C} (x-\mu_1)(x-\mu_1)^T$$

$$J(w) = \frac{|\overline{\mu_1} - \overline{\mu_2}|^2}{\tilde{S}_1^2 + \tilde{S}_2^2}$$

$$\therefore S_W = S_1 + S_2$$

* Steps
  ↳ Compute the mean vectors of different class
  ↳ Compute the scatter matrix (between & within class scatter matrix)
  ↳ Compute eigen vectors & eigen values for scatter matrix
  ↳ Sort eigen vectors by decreasing eigenvalues & choose k eigenvectors with largest eigenvalues
  ↳ Use this eigenvector matrix to transform the samples onto the new subspace

Problem - LDA:

$C_1 \rightarrow X_1 = \{(4,1)(2,4),(2,3),(3,6),(4,4)\} = (x_1, x_2)$

$C_2 \rightarrow X_2 = \{(9,10),(6,8),(9,5),(8,7),(10,8)\} = (x_1, x_2)$

Soln:

Step-1: Compute the mean μ_1 & μ_2

$$\therefore \mu_1 = \left\{\frac{4+2+2+3+4}{5}, \frac{1+4+3+6+4}{5}\right\} = \{3, 3.6\}$$

$$\therefore \mu_2 = \left\{\frac{9+6+9+8+10}{5}, \frac{10+8+5+7+8}{5}\right\} = \{8.4, 7.6\}$$

Step-2: Compute within-class scatter matrix. S_w

$$\therefore S_w = S_1 + S_2$$

So lets now find the covariance matrices for each class

$$S_1 = \sum_{x \in C_1} (x-\mu_1)(x-\mu_1)^T$$, where μ_1 is the mean of class

$$X_1 - \mu_1 = \begin{bmatrix} 1 & -1 & -1 & 0 & 1 \\ -2.6 & 0.4 & -0.6 & 2.4 & 0.4 \end{bmatrix}$$

Now for each x, we are going to calculate $(x-\mu_1)(x-\mu_1)^T$.
So we will have 5 such matrices

$$\begin{bmatrix} 1 \\ -2.6 \end{bmatrix} \begin{bmatrix} 1 & -2.6 \end{bmatrix} = \begin{bmatrix} 1 & -2.6 \\ -2.6 & 6.76 \end{bmatrix} \quad \text{--- (1)}$$

$$\begin{bmatrix} -1 \\ 0.4 \end{bmatrix} \begin{bmatrix} -1 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & -0.4 \\ -0.4 & 0.16 \end{bmatrix} \quad \text{--- (2)}$$

$$\begin{bmatrix} -1 \\ -0.6 \end{bmatrix} \begin{bmatrix} -1 & -0.6 \end{bmatrix} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 0.36 \end{bmatrix} \quad \text{--- (3)}$$

$$\begin{bmatrix} 0 \\ 2.4 \end{bmatrix} \begin{bmatrix} 0 & 2.4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 5.76 \end{bmatrix} \quad \text{--- (4)}$$

$$\begin{bmatrix} 1 \\ 0.4 \end{bmatrix} \begin{bmatrix} 1 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 0.16 \end{bmatrix} \quad \text{--- (5)}$$

Adding (1) + (2) + (3) + (4) + (5) and take the average and we will get the covariance matrix

$$\therefore S_1 = \frac{1}{5}\begin{bmatrix} 4 & -2 \\ -2 & 13.2 \end{bmatrix} = \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 2.64 \end{bmatrix}$$

$$\therefore S_2 = \sum_{x \in C_2} (x-\mu_2)(x-\mu_2)^T$$

$$\therefore X_2 - \mu_2 = \begin{bmatrix} 0.6 & -2.4 & 0.6 & -0.4 & 1.6 \\ 2.4 & 0.4 & -2.6 & -0.6 & 0.4 \end{bmatrix}$$

$$\therefore \begin{bmatrix} 0.6 \\ 2.4 \end{bmatrix} \begin{bmatrix} 0.6 & 2.4 \end{bmatrix} = \begin{bmatrix} 0.36 & 1.44 \\ 1.44 & 5.76 \end{bmatrix} \quad \text{--- (6)}$$

$$\begin{bmatrix} -2.4 \\ 0.4 \end{bmatrix} \begin{bmatrix} -2.4 & 0.4 \end{bmatrix} = \begin{bmatrix} 5.76 & -0.96 \\ -0.96 & 0.16 \end{bmatrix} \quad \text{--- (7)}$$

$$\begin{bmatrix} 0.6 \\ -2.6 \end{bmatrix} \begin{bmatrix} 0.6 & -2.6 \end{bmatrix} = \begin{bmatrix} 0.36 & -1.56 \\ -1.56 & 6.76 \end{bmatrix} \quad \text{--- (8)}$$

$$\begin{bmatrix} -0.4 \\ -0.6 \end{bmatrix} \begin{bmatrix} -0.4 & -0.6 \end{bmatrix} = \begin{bmatrix} 0.16 & 0.24 \\ 0.24 & 0.36 \end{bmatrix} \quad \text{--- (9)}$$

$$\begin{bmatrix} 1.6 \\ 0.4 \end{bmatrix} \begin{bmatrix} 1.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 2.56 & 0.64 \\ 0.64 & 0.16 \end{bmatrix} \quad \text{--- (10)}$$

Add (6) + (7) + (8) + (9) + (10) and take the average

$$\therefore S_2 = \frac{1}{5} \begin{bmatrix} 9.2 & -0.2 \\ -0.18 & 13.2 \end{bmatrix} = \begin{bmatrix} 1.84 & -0.04 \\ -0.036 & 2.64 \end{bmatrix}$$

$$\therefore S_W = S_1 + S_2 = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$$

Step-3:- Compute between-class scatter matrics.

$$\therefore S_B = [\mu_1 - \mu_2](\mu_1 - \mu_2)^T$$

$$= \begin{pmatrix} -5.4 \\ -4 \end{pmatrix} (-5.4 \quad -4)$$

$$= \begin{bmatrix} 29.16 & 21.6 \\ 21.6 & 16 \end{bmatrix}$$

Step-4:- Find the best LDA projection vector. Similar to PCA we have to find Eigen values & vectors & largest value to be considered.

$$\therefore |S_W^{-1} S_B - \lambda I| = 0$$

$$S_W^{-1} = \begin{bmatrix} 0.384 & 0.03 \\ 0.03 & 0.19 \end{bmatrix}$$

$$\therefore S_W^{-1} S_B - \lambda I = \begin{bmatrix} 0.38 & 0.03 \\ 0.03 & 0.19 \end{bmatrix} \begin{bmatrix} 29.16 & 21.6 \\ 21.6 & 16 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$= \begin{bmatrix} 11.89 & 8.81 \\ 5.08 & 3.76 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$= \begin{bmatrix} 11.89 - \lambda & 8.81 \\ 5.08 & 3.76 - \lambda \end{bmatrix}$$

$$\therefore S_W^{-1} S_B - \lambda I = 0$$

$$(11.89 - \lambda)(3.76 - \lambda) - 44.75 = 0$$

$$44.71 - 11.89\lambda - 3.76\lambda + \lambda^2 - 44.75 = 0$$

$$\lambda^2 - 15.65 - 0.04 = 0$$

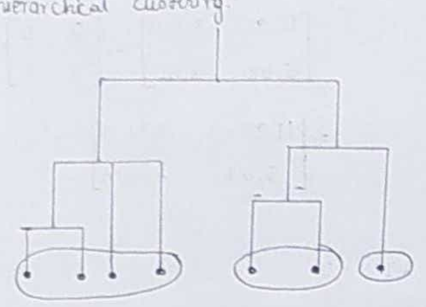
$$\therefore \lambda_1 = 15.65, \quad \lambda_2 = -0.002$$

Clustering:-

- The task of grouping a set of objects in such a way that objects in the same group.
- It is a unsupervised learning.
- Similarity is high in intra-class and low in inter-class
- Major Clustering approach → Partitioned based, hierarchical based, density-based, Grid based and model based clustering.

Dendrogram:-

• It is a representation through which we define the clusters in hierarchical clustering.



K-Means Clustering:-

• It is a partitional clustering algorithm.

• Algorithm
- ↳ Given k
- ↳ Randomly pick k instance as the initial ~~cluster~~ centers
- ↳ Assign the rest of instances to closest one of k clusters
- ↳ Recalculate the mean of each cluster
- ↳ Repeat 3 & 4 until means don't change.

• Advantages → Ease of implementation & high-speed performance
↳ Measurable and efficient in large data.

• Disadvantages → Selection of optimal number of clusters (k) is difficult
↳ Selection of the initial ~~condition~~ centroids is random.

Use K-Mean clustering algorithm to divide the following data into clusters. Use Euclidean ~~cluster~~ distance measure. Use (2,1) & (2,3) as initial centroids.

x_1	1	2	2	3	4	5
x_2	1	1	3	2	3	5

Soln:-

∴ Given k = 2, (centroids)

Iteration-1:-

minimum among C_1 & C_2

Data Points	Dist. from (2,1) (C_1)	Dist. from (2,3) (C_2)	Assigned clusters
(1,1)	1	2.24	C_1
(2,1)	0	2	C_1
(2,3)	2	0	C_2
(3,2)	1.41	1.41	C_1
(4,3)	2.83	2	C_2
(5,5)	5	3.61	C_2

∴ Cluster-1 elements => (1,1), (2,1), (3,2)

∴ Cluster-2 elements => (2,3), (4,3), (5,5)

∴ Mean of Cluster-1 => (2, 1.33)

∴ Mean of Cluster-2 => (3.67, 3.67)

Iteration - 2:-

Data Points	Dist. from (2, 1.33)	Dist. from (3.67, 3.67)	Assigned Cluster
(1, 1)	1.05	3.78	C_1
(2, 1)	0.33	3.15	C_1
(2, 3)	1.67	1.79	C_1
(3, 2)	1.2	1.79	C_1
(4, 3)	2.61	0.75	C_2
(5, 5)	4.74	1.88	C_2

∴ Cluster-1 elements ⇒ (1,1) (2,1) (2,3) (3,2) ⇒ (2, 1.75)

∴ Cluster-2 elements ⇒ (4,3), (5,5) ⇒ (4.5, 4)

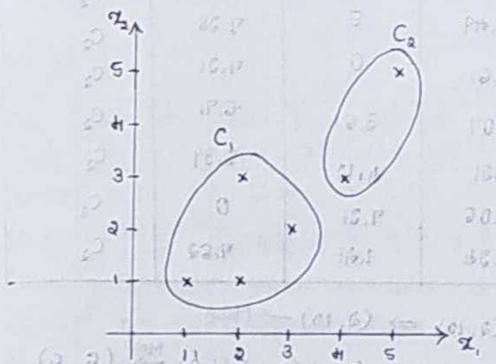
Iteration - 3:-

Data Points	Dist. from (2, 1.75)	Dist. from (4.5, 4)	Assigned Clusters
(1, 1)	1.25	4.61	C_1
(2, 1)	0.75	3.91	C_1
(2, 3)	1.25	2.69	C_1
(3, 2)	1.03	2.5	C_1
(4, 3)	2.36	1.12	C_2
(5, 5)	4.42	1.12	C_2

∴ Cluster-1 ⇒ (1,1), (2,1), (2,3), (3,2) ⇒ (2, 1.75)

∴ Cluster-2 ⇒ (4,3), (5,5) ⇒ (4.5, 4)

∴ We got same clusters from iteration 2 & 3. So we need to stop the iteration.



Use K-Mean clustering algorithm to divide the data points into 3-clusters and the points are (2,10), (2,5), (8,4), (5,8), (7,5), (6,4), (1,2), (4,9). Use the Euclidean distance function with the initial centroids as (2,10), (5,8) and (1,2).

Soln:-

Given, k = 3.

Data Points	C_1 Dist. from (2, 10)	C_2 Dist. from (5, 8)	C_3 Dist. from (1, 2)	Assigned Clusters
(2,10)	0	3.61	8.06	C_1
(2,5)	5	4.24	3.16	C_3
(8,4)	8.49	5	7.28	C_2
(5,8)	3.61	0	7.21	C_2
(7,5)	7.07	3.61	6.71	C_2
(6,4)	7.21	4.12	5.39	C_2
(1,2)	8.06	7.21	0	C_3
(4,9)	2.24	1.41	7.62	C_2

∴ Cluster-1 ⟹ (2, 10) ⟹ (2, 10) ⟶ Mean

∴ Cluster-2 ⟹ (8, 4) (5,8) (7,5) (6, 4) (4,9) $\xRightarrow{\text{Mean}}$ (6, 6)

∴ Cluster-3 ⟹ (2,5) (1,2) $\xRightarrow{\text{Mean}}$ (1.5, 3.5)

Iteration-2:-

Data Points	Dist. from (2, 10)	Dist. from (6, 6)	Dist. from (1.5, 3.5)	Assigned Cluster
(2,10)	0	5.66	6.52	C_1
(2,5)	5	4.12	1.58	C_3
(8,4)	8.4	2.83	6.52	C_2
(5,8)	3.61	2.24	5.7	C_2
(7,5)	7.07	1.41	5.7	C_2
(6,4)	7.21	2	4.53	C_2
(1,2)	8.06	6.4	1.58	C_3
(4,9)	2.24	3.61	6.04	C_1

∴ Cluster-1 ⟶ (2,10), (4,9) $\xrightarrow{\text{Mean}}$ (3, 9.5)

∴ Cluster-2 ⟶ (8, 4) (5, 8) (7,5) (6, 4) $\xrightarrow{\text{Mean}}$ (6.5, 5.25)

∴ Cluster-3 ⟶ (2,5) (1,2) $\xrightarrow{\text{Mean}}$ (1.5, 3.5)

Iteration-3:-

Data Points	Dist. from (3, 9.5)	Dist. from (6.5, 5.25)	Dist. from (1.5, 3.5)	Assigned Cluster
(2,10)	1.12	6.54	6.52	C_1
(2,5)	4.61	4.51	1.58	C_3
(8,4)	7.43	1.95	6.52	C_2
(5,8)	2.5	3.13	5.7	C_1
(7,5)	6.02	0.56	5.7	C_2
(6,4)	6.26	1.35	4.53	C_2
(1,2)	7.76	6.39	1.58	C_3
(4,9)	1.12	4.51	6.04	C_1

∴ Cluster-1 ⟶ (2, 10) (5, 8) (4, 9)

∴ Mean = (<u>3.67, 9</u>)

∴ Cluster-2 ⟶ (8, 4) (7, 5) (6, 4)

∴ Mean = (<u>7, 4.33</u>)

∴ Cluster-3 ⟹ (2,5) (1,2)

∴ Mean ⟹ (<u>1.5, 3.5</u>)

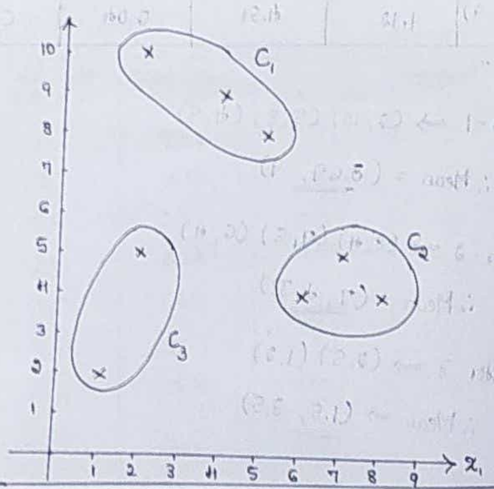
Iteration - 4 :-

Data Points	Dist. from (3.67, 9)	Dist. from (7, 4.33)	Dist. from (1.5, 3.5)	Assigned Clusters
(2, 10)	1.95	7.56	6.52	C_1
(2, 5)	4.33	5.04	1.58	C_3
(8, 4)	6.61	1.05	6.52	C_2
(5, 8)	1.66	4.18	5.7	C_1
(7, 5)	5.2	0.67	5.7	C_2
(6, 4)	5.52	1.05	4.53	C_2
(1, 2)	7.49	6.44	1.58	C_3
(4, 9)	0.33	5.55	6.04	C_1

∴ Cluster-1 ⟹ (2,10) (5,8) (4,9) $\xrightarrow{\text{Mean}}$ (3.67, 9)

∴ Cluster-2 ⟹ (8,4) (7,5) (6,4) $\xrightarrow{\text{Mean}}$ (7, 4.33)

∴ Cluster-3 ⟶ (2,5) (1,2) $\xrightarrow{\text{Mean}}$ (1.5, 3.5)



Elbow Method:-

- It is used to calculate the k clusters in dataset.
- Within - Cluster - sum - of - square (WCSS) on X-axis
- Number of clusters (k) on Y-axis.
- The optimal k value is the point at which the graph forms an elbow.
- WCSS → Sum of squared distance b/n centroid and each points.

Hierarchical Clustering:-

- The tree structure represents the relationship between the objects and shows how objects are clustered at different levels.
- This diagram is also known as Dendrogram.
- Two methods
  - ↳ Agglomerative ⟶ AGNES ⟶ bottom-up
  - ↳ Divisive ⟶ DIANA ⟶ top-down

Given the dataset {a, b, c, d, e} and the following distance matrix, construct a dendrogram by complete linkage hierarchical clustering, using aglomerative method.

	a	b	c	d	e
a	0	9	3	6	11
b	9	0	7	5	10
c	3	7	0	9	2
d	6	5	9	0	8
e	11	10	2	8	0

Soln:-

Step-1:- Assigning each data items to its own cluster so that we have N = 5, clusters, each containing just one item.

Dataset = { a, b, c, d, e }

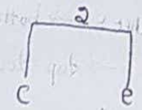
Initial Clusters are, C_1 : {a}, {b}, {c}, {d}, {e}

Step-2:- Find the closest pair of clusters and merge them into a single cluster so that now we have one less cluster.

∴ Minimum distance between {c} and {e}

$\phi(\{c\}, \{e\}) = 2.$

∴ New set of clusters C_2 = {a}, {b}, {d}, {c, e}



Step-3:- Compute the distance between new cluster & each of the old cluster.

	{a}	{b}	{d}	{c, e}
{a}	0	9	6	11
{b}	9	0	5	10
{d}	6	5✓	0	9
{c, e}	11	10	9	0

Here complete linkage is used method. So we maximum distance between the data points.

∴ $d(\{a\}, \{c, e\}) = \max(d\{a, c\}, d\{a, e\})$

$= \max(3, 11) = \underline{11}$

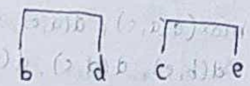
$d(\{b\}, \{c, e\}) = \max(d\{b, c\}, d\{b, e\}) = \underline{10}$

$d(\{d\}, \{c, e\}) = \max(d\{d, c\}, d\{d, e\}) = \underline{9}$

Step-4:- Repeat step-2 & 3 until all items are clustered into a single cluster.

∴ Minimum distance, $\phi(\{b\}, \{d\}) = 5$

∴ New set of cluster C_3 : {a}, {b, d}, {c, e}

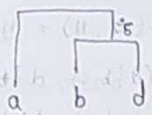


	{a}	{b, d}	{c, e}
{a}	0	9	11
{b, d}	9✓	0	10
{c, e}	11	10	0

∴ $d(\{a\}, \{b, d\}) =$

∴ Minimum distance, $\phi(\{a\},\{b,d\}) = 9$

∴ New cluster $C_4 = \{a,b,d\}, \{c,e\}$

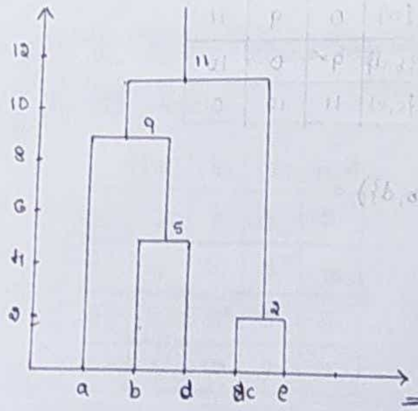


	{a,b,d}	{c,e}
{a,b,d}	0	11
{c,e}	11	0

$d(\{a,b,d\},\{c,e\}) = max(d(a,c), d(a,e), d(b,c), d(b,e), d(d,c), d(d,e))$

$= (3,$

$= 11$



Solve using complete linkage aglomerative clustering and draw the dendrogram.

	a	b	c	d	e
a	0	4	7	9	1
b	4	0	3	5	3
c	7	3	0	2	6
d	9	5	2	0	8
e	1	3	6	8	0

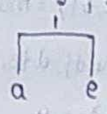
Soln:-

Dataset = {a, b, c, d, e}

Initial Clusters C_1: {a} {b} {c}; {d}; {e}

∴ Minimum distance is : 1 & cluster is : {a, e}

∴ New clusters are, C_2: {a, e}, {b}, {c}, {d}



	{a,e}	{b}	{c}	{d}
{a,e}	0	4	6	9
{b}	4	0	3	5
{c}	6	3	0	2
{d}	9	5	2	0

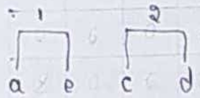
$d(\{a,e\},\{b\}) = max(d\{a,b\}, d\{e,b\}) = max(4, \cancel{8}\ 3) = \cancel{5}\ 4$

$d(\{a,e\},\{c\}) = max(d\{a,c\}, d\{e,c\}) = max(1, 6) = 6$

$d(\{a,e\},\{d\}) = \max(d\{a,d\}, d\{d,e\}) = \max(9,8) = 9$

Now minimum distance : 2 & clusters : {d}, {c}

∴ New clusters, C_3 : {a, e} {b} {c, d}



	{a,e}	{b}	{c,d}
{a,e}	0	4	9
{b}	4 ✓	0	5
{c,d}	9	5	0

$d(\{a,e\}, b) = \max(d\{a,b\}, d\{e,b\}) = \max(4,3) = 4$

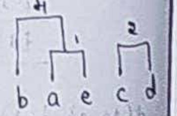
$d(\{a,e\},\{c,d\}) = \max(d\{a,c\}, d\{a,d\}, d\{e,c\}, d\{e,d\})$

$= \max(7, 9, 6, 8) = 9$

$d(\{b\},\{c,d\}) = \max(d\{b,c\}, d\{b,d\}) = \max(3,5) = 5$

∴ Now minimum distance is : 4

Clusters considered : {a, e} & {b}.



∴ New Clusters C_4 : {a, b, e}, {c, d}

	{a,b,e}	{c,d}
{a,b,e}	0	9
{c,d}	9 ✓	0

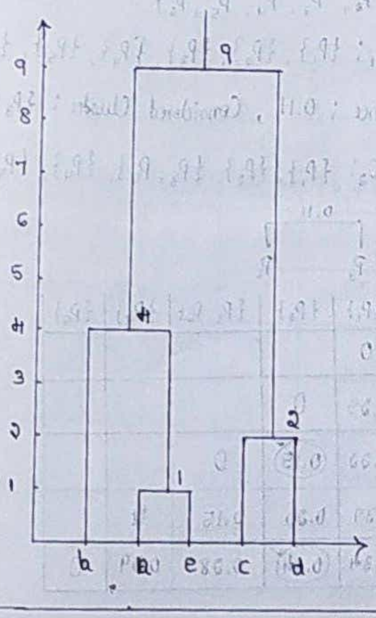
$d(\{a,b,e\}, \{c,d\}) = \max(d\{a,c\}, d\{a,d\}, d\{b,c\}, d\{b,d\}, d\{e,c\}, d\{e,d\})$

$= \max(7, 9, 3, 5, 6, 8) = 9$

∴ Now minimum distance = 9

Clusters considered, C_5 : {a, b, e} & {c, d}

∴ New Cluster, C_5 : {a, b, c, d, e}



Solve using Single linkage aglomerative (min method) and draw the dendrogram.

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0					
P_2	0.25	0				
P_3	0.22	0.15	0			
P_4	0.37	0.20	0.15	0		
P_5	0.34	0.14	0.28	0.29	0	
P_6	0.23	0.25	0.11	0.22	0.39	0

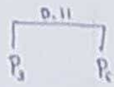
Soln:-

Datasets : $\{P_1, P_2, P_3, P_4, P_5, P_6\}$

Initial Clusters C_1 : $\{P_1\}, \{P_2\}, \{P_3\}, \{P_4\}, \{P_5\}, \{P_6\}$

$\therefore$ Minimum distance : 0.11 , Considered Clusters : $\{P_3\}$ & $\{P_6\}$

$\therefore$ New Clusters, C_2 : $\{P_1\}, \{P_2\}, \{P_3, P_6\}, \{P_4\}, \{P_5\}$



	$\{P_1\}$	$\{P_2\}$	$\{P_3, P_6\}$	$\{P_4\}$	$\{P_5\}$
$\{P_1\}$	0				
$\{P_2\}$	0.25	0			
$\{P_3, P_6\}$	0.22	0.15	0		
$\{P_4\}$	0.37	0.20	0.15	0	
$\{P_5\}$	0.34	0.14	0.28	0.29	0

$\therefore d(\{P_3, P_6\}, \{P_1\}) = \min(d\{P_3, P_1\}, d\{P_6, P_1\}) = \min(0.22, 0.23)$
$= \underline{0.22}$

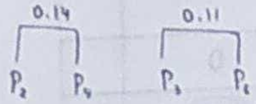
$\therefore d(\{P_3, P_6\}, \{P_2\}) = \min(d\{P_3, P_2\}, d\{P_6, P_2\}) = \min(0.15, 0.25)$
$= \underline{0.15}$

$\therefore d(\{P_4\}, \{P_3, P_6\}) = \min(d\{P_4, P_3\}, d\{P_4, P_6\}) = \min(0.15, 0.22)$
$= \underline{0.15}$

$\therefore d(\{P_5\}, \{P_3, P_6\}) = \min(d\{P_5, P_3\}, d\{P_5, P_6\}) = \min(0.28, 0.39)$
$= \underline{0.28}$

$\therefore$ Minimum distance : 0.14 & Considered Cluster : $\{P_2\}$ & $\{P_5\}$

$\therefore$ New Clusters, C_3 : $\{P_1\}, \{P_2, P_5\}, \{P_3, P_6\}, \{P_4\}$



	$\{P_1\}$	$\{P_2, P_5\}$	$\{P_3, P_6\}$	$\{P_4\}$
$\{P_1\}$	0			
$\{P_2, P_5\}$	0.25	0		
$\{P_3, P_6\}$	0.22	0.15	0	
$\{P_4\}$	0.37	0.20	0.15	0

$d(\{P_2, P_5\}, \{P_1\}) = \min(d\{P_2, P_1\}, d\{P_5, P_1\}) = \min(0.25, 0.34) = \underline{0.25}$

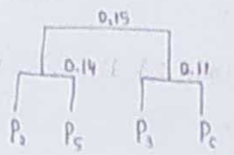
$d(\{P_3, P_6\}, \{P_2, P_5\}) = \min(d\{P_3, P_2\}, d\{P_3, P_5\}, d\{P_6, P_2\}, d\{P_6, P_5\})$
$= \min(0.15, 0.28, 0.25, 0.39) = \underline{0.15}$

$d(\{P_4\}, \{P_3, P_6\}) = \min(d\{P_4, P_3\}, d\{P_4, P_6\}) = \min(0.20, 0.24)$

$= \underline{0.20}$

$\therefore$ Minimum distance : 0.15 & Considered Clusters : $\{P_3, P_6\}$ & $\{P_2, P_5\}$

$\therefore$ New Clusters, C_4 : $\{P_1\}, \{P_2, P_3, P_5, P_6\}, \{P_4\}$



	$\{P_1\}$	$\{P_2, P_3, P_5, P_6\}$	$\{P_4\}$
$\{P_1\}$	0		
$\{P_2, P_3, P_5, P_6\}$	0.22	0	
$\{P_4\}$	0.37	0.15 ✓	0

$d(\{P_2, P_3, P_5, P_6\}, \{P_1\}) = \min(d\{P_2, P_1\}, d\{P_3, P_1\}, d\{P_5, P_1\}, d\{P_6, P_1\})$

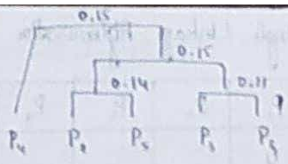
$= \min(0.25, 0.22, 0.37, 0.23) = \underline{0.22}$

$d(\{P_4\}, \{P_2, P_3, P_5, P_6\}) = \min(d\{P_4, P_2\}, d\{P_4, P_3\}, d\{P_4, P_5\}, d\{P_4, P_6\})$

$= \min(0.20, 0.15, 0.29, 0.22) = \underline{0.15}$

$\therefore$ Minimum distance : 0.15 & Considered Clusters : $\{P_4\}$ & $\{P_2, P_3, P_5, P_6\}$

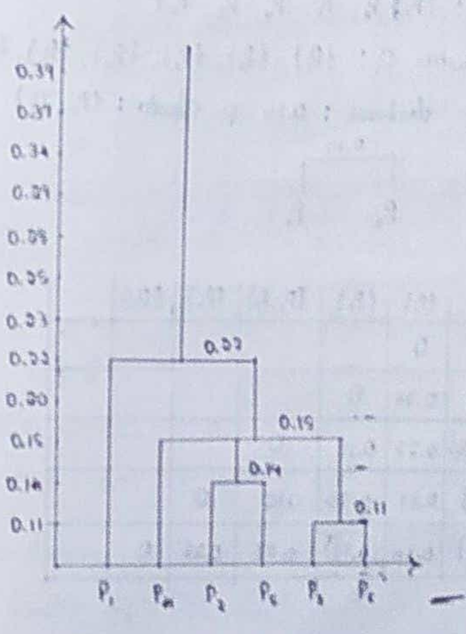
$\therefore$ New Clusters, C_5 : $\{P_1\}, \{P_2, P_3, P_4, P_5, P_6\}$



	$\{P_1\}$	$\{P_2, P_3, P_4, P_5, P_6\}$
$\{P_1\}$	0	
$\{P_2, P_3, P_4, P_5, P_6\}$	0.22	0

$\therefore$ Minimum distance : 0.22

$\therefore$ New Cluster C_6 : $\{P_1, P_2, P_3, P_4, P_5, P_6\}$



Solve using Single Linkage Agglomerative.

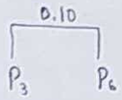
	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0					
P_2	0.24	0				
P_3	0.22	0.15	0			
P_4	0.37	0.20	0.16	0		
P_5	0.34	0.13	0.28	0.28	0	
P_6	0.24	0.25	0.10	0.22	0.39	0

Soln:-

Datasets : $\{P_1, P_2, P_3, P_4, P_5, P_6\}$

Initial Clusters C_1 : $\{P_1\}, \{P_2\}, \{P_3\}, \{P_4\}, \{P_5\}, \{P_6\}$

∴ Minimum distance : 0.10 & cluster : $\{P_3, P_6\}$



	$\{P_1\}$	$\{P_2\}$	$\{P_3, P_6\}$	$\{P_4\}$	$\{P_5\}$
$\{P_1\}$	0				
$\{P_2\}$	0.24	0			
$\{P_3, P_6\}$	0.22	0.15	0		
$\{P_4\}$	0.37	0.20	0.16	0	
$\{P_5\}$	0.34	0.13	0.28	0.28	0

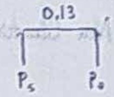
distor

$d(\{P_3, P_6\}, \{P_1\}) = \min(d\{P_3, P_1\}, d\{P_6, P_1\}) = \min(0.22, 0.24)$
$= 0.22$

∴ Minimum distance : 0.13 & clusters are $\{P_5\}$ & $\{P_2\}$

∴ New Clusters, C_2 : $\{P_1\}, \{P_2, P_5\}, \{P_3, P_6\}, \{P_4\}$

	$\{P_1\}$	$\{P_2, P_5\}$	$\{P_3, P_6\}$	$\{P_4\}$
$\{P_1\}$	0			
$\{P_2, P_5\}$	0.24	0		
$\{P_3, P_6\}$	0.22	0.15	0	
$\{P_4\}$	0.37	0.20	0.16	0



$d(\{P_2, P_5\}, \{P_3, P_6\}) = \min(d\{P_2, P_3\}, d\{P_2, P_6\}, d\{P_5, P_3\}, d\{P_5, P_6\})$
$= \min(0.15, 0.25, 0.28, 0.39) = 0.15$

∴ Minimum distance : 0.15 & clusters are : $\{P_3, P_6\}$ & $\{P_2, P_5\}$

∴ New clusters ; $C_3 = \{P_1\}, \{P_2, P_3, P_5, P_6\}, \{P_4\}$

	$\{P_1\}$	$\{P_2, P_3, P_5, P_6\}$	$\{P_4\}$
$\{P_1\}$	0		
$\{P_2, P_3, P_5, P_6\}$	0.22	0	
$\{P_4\}$	0.37	0.16	0

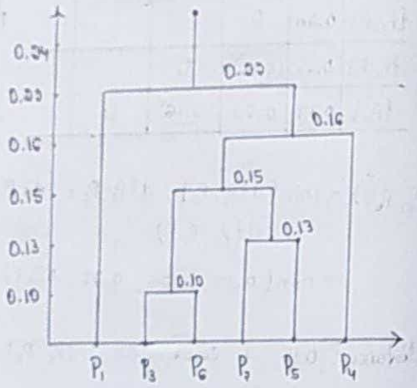
∴ Minimum distance : 0.16 & clusters are : $\{P_4\}$ & $\{P_2, P_3, P_5, P_6\}$

∴ New Clusters : C_4 : $\{P_1\}, \{P_2, P_3, P_4, P_5, P_6\}$

	$\{P_1\}$	$\{P_2, P_3, P_4, P_5, P_6\}$
$\{P_1\}$	0	
$\{P_2, P_3, P_4, P_5, P_6\}$	0.22	0

∴ Minimum distance : 0.22 & clusters are : $\{P_1\}$ & $\{P_2, P_3, P_4, P_5, P_6\}$

∴ New Cluster is : $C_6 = \{P_1, P_2, P_3, P_4, P_5, P_6\}$



Naïve Bayes Classifier :-

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$P(B|A) \longrightarrow$ Likely hood

$P(A) \longrightarrow$ Prior

Problem : 1

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cold	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Soln :-

$N = 14$

$Yes = 9$

$No = 5$

$\therefore P(Yes) = 9/14$

$\therefore P(No) = 5/14$

Frequency Table:-

		Play Tenis Yes	Play Tenis No
Outlook	Sunny	2	3
	Overcast	4	0
	Rainy	3	2

		Play Tenis Yes	Play Tenis No
Temprature	Hot	2	2
	Cold	3	1
	Mild	4	2

		Play Tenis Yes	Play Tenis No
Humidity	High	3	4
	Normal	6	1

		Play Tenis Yes	Play Tenis No
Wind	Strong	3	3
	Weak	6	2

∴ Find

Outlook	Temp	Humidity	Wind	PlayTenis
Rain	Cool	High	Strong	?

$\Rightarrow P(Yes/X) = P(Rain/Yes) \times P(Cool/Yes) \times P(High/Yes) \times P(Strong/Yes) \times P(Yes)$

$= 3/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14$

$= \underline{0.0079}$

$P(No/X) = P(Rain/No) \times P(Cool/No) \times P(High/No) \times P(Strong/No) \times P(No)$

$= 2/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14$

$= \underline{0.0137}$ ✓

∴ For the given instance, We cannot play tenis.
∴ The result is "No".

Problem - 2:- We have a "Red, Domestic, SUV".

No.	Color	Type	Origin	Stolen
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Soln:- $N = 10$, Yes $= 5$, No $= 5$, $P(Yes) = 5/10$, $P(No) = 5/10$

		Stolen	
		Yes	No
Color	Red	3	2
	Yellow	2	3

$\Rightarrow$ Likelyhood

Stolen	
Yes	No
3/5	2/5
2/5	3/5

		Stolen	
		Yes	No
Type	Sports	4	2
	SUV	1	3

$\Rightarrow$ Likelyhood

Stolen	
Yes	No
4/5	2/5
1/5	3/5

		Stolen	
		Yes	No
Origin	Domestic	2	3
	Imported	3	2

$\Rightarrow$ Likelyhood

Stolen	
Yes	No
2/5	3/5
3/5	2/5

$\therefore P(Yes/X) = P(Red/Yes) \times P(Domestic/Yes) \times P(SUV/Yes) \times P(Yes)$

$= 3/5 \times 2/5 \times 1/5 \times 5/10$

$= \underline{0.024}$

$P(No/X) = P(Red/No) \times P(Domestic/No) \times P(SUV/No) \times P(No)$

$= 2/5 \times 3/5 \times 3/5 \times 5/10$

$= \underline{0.072}$

$\therefore$ For the given data the stolen is "No"

Support Vector Machine (SVM):-

- Supervised machine learning
- Used for both classification & regression problem.
- Primarily it is mostly used for classification problem.
- The core of SVM consist of maximum marginal hyperplane.

$w^T x + b = 0 \rightarrow$ optimal hyperplane

$w^T x + b = 1 \rightarrow$ positive class

$w^T x + b = -1 \rightarrow$ negative class

where, w = weight vector

x = input feature

b = bias

- Types $\rightarrow$ Linear & Non-Linear (Kernel - trick)
- Kernel Trick is a kernel function.
- Types Kernel $\rightarrow$ Linear $\Rightarrow K(x_1, x_2) = (x_1 \cdot x_2)$

$\rightarrow$ Polynomial $\Rightarrow K(x_1, x_2) = (x_1 \cdot x_2 + 1)^d$

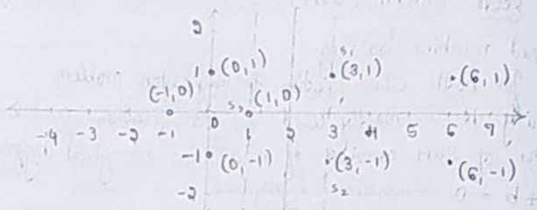
$\Rightarrow$ RBSF , $\Rightarrow$ Sigmoid

Problem:-

Positively labelled data points $(3,1), (3,-1), (6,1), (6,-1)$ and negatively labelled data points $(1,0), (0,1), (0,-1), (-1,0)$. Use SVM to find weight w, and bias b.

Soln:-

Step-1:- Identify the data points closest to both sides of the classes SV are $(1,0), (3,1), (3,-1)$.



<u>Step-2</u>:- Augment '1' as bias input to each vector.

$$S_1 = \begin{pmatrix}1\\0\end{pmatrix}, \quad S_2 = \begin{pmatrix}3\\1\end{pmatrix} \quad S_3 = \begin{pmatrix}3\\-1\end{pmatrix}$$

$$\Downarrow$$

$$\overline{S}_1 = \begin{pmatrix}1\\0\\1\end{pmatrix} \quad \overline{S}_2 = \begin{pmatrix}3\\1\\1\end{pmatrix}, \quad \overline{S}_3 = \begin{pmatrix}3\\-1\\1\end{pmatrix}$$

<u>Step-3</u>:- Write the equations needed to calculate the weighted vector.

$$\alpha_1 \overline{S}_1 \overline{S}_1 + \alpha_2 \overline{S}_2 \overline{S}_1 + \alpha_3 \overline{S}_3 \overline{S}_1 = -1 \quad \text{— ①}$$

$$\alpha_1 \overline{S}_1 \overline{S}_2 + \alpha_2 \overline{S}_2 \overline{S}_2 + \alpha_3 \overline{S}_3 \overline{S}_2 = 1 \quad \text{— ②}$$

$$\alpha_1 \overline{S}_1 \overline{S}_3 + \alpha_2 \overline{S}_2 \overline{S}_3 + \alpha_3 \overline{S}_3 \overline{S}_3 = 1 \quad \text{— ③}$$

<u>Step-4</u>:- Substitute the values into the equation.

$$① \Rightarrow \alpha_1 \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix} + \alpha_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix} = -1$$

$$② \Rightarrow \alpha_1 \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix} + \alpha_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix} = 1$$

$$③ \Rightarrow \alpha_1 \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix} + \alpha_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix} = 1$$

$$① \Rightarrow \alpha_1 (1+0+1) + \alpha_2 (3+0+1) + \alpha_3 (3+0+1) = -1$$

$$2\alpha_1 + 4\alpha_2 + 4\alpha_3 = -1$$

$$② \Rightarrow \alpha_1 (3+0+1) + \alpha_2 (9+1+1) + \alpha_3 (9-1+1) = 1$$

$$4\alpha_1 + 11\alpha_2 + 9\alpha_3 = 1$$

$$③ \Rightarrow \alpha_1 (3+0+1) + \alpha_2 (9-1+1) + \alpha_3 (9+1+1) = 1$$

$$4\alpha_1 + 9\alpha_2 + 11\alpha_3 = 1$$

Using ①, ② & ③

$$\therefore \alpha_1 = \underline{-3.5}, \quad \alpha_2 = \underline{0.75}, \quad \alpha_3 = \underline{0.75}$$

<u>Step-5</u>:- Calculate the weight vector.

$$\overline{w} = \sum_i \alpha_i \overline{S}_i$$

$$= \alpha_1 \overline{S}_1 + \alpha_2 \overline{S}_2 + \alpha_3 \overline{S}_3$$

$$= -3.5 \begin{pmatrix}1\\0\\1\end{pmatrix} + 0.75 \begin{pmatrix}3\\1\\1\end{pmatrix} + 0.75 \begin{pmatrix}3\\-1\\1\end{pmatrix}$$

$$= \begin{pmatrix}-3.5\\0\\-3.5\end{pmatrix} + \begin{pmatrix}2.25\\0.75\\0.75\end{pmatrix} + \begin{pmatrix}2.25\\-0.75\\0.75\end{pmatrix}$$

$$= \begin{bmatrix}1\\0\\-2\end{bmatrix}$$

$$\therefore w = \begin{pmatrix}1\\0\end{pmatrix} \text{ and } b = -2$$

if $w = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ the line is parallel to x_2 axis |

$= \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ the line is parallel to x_1 axis —

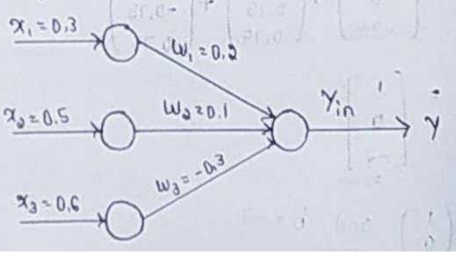
$= \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ the line make 45° w.r.t. to x_1 & x_2 axis \

Soft margin SVM:-

- If the data is not linearly separable, the earlier method (hard margin) will not work.
- We define a "Slack Variable" ξ_i which stores the deviation from the origin.
- ξ_i are just "slack variable" optimization.
- $0 < \xi_i < 1 \rightarrow X_i$ lies on the correct side of hyperplane but inside the margin.
- $\xi_i > 0 \longrightarrow X_i$ is misclassified.
- Soft margin SVM also known as C-SVM (C- for regularization parameter)

Artificial Neural Networks (ANN):-

For the following networks calculate the net input to the output neuron.



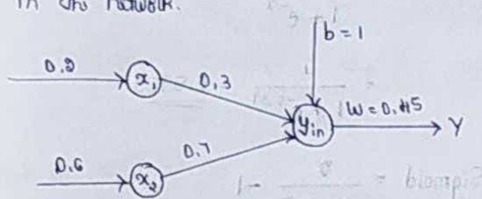
Soln:-

$Y_{in} = \sum w_i x_i$

$= (0.3 \times 0.2) + (0.5 \times 0.1) + (0.6 \times -0.3)$

$= \underline{\underline{-0.07}}$

Calculate the net input for the below network with the bias induced in the network.



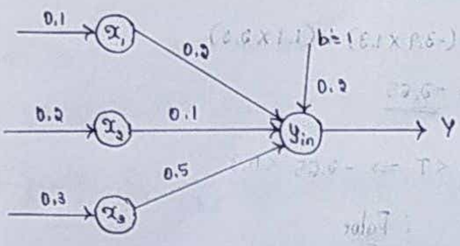
Soln:-

$Y_{in} = \sum w_i x_i + b$

$= [(0.2 \times 0.3) + (0.6 \times 0.7)] + (1 \times 0.45)$

$= \boxed{0.93}$

Calculate the net input for the following network with the activation function sigmoid and bipolar sigmoid.



Soln:-

$$y_{in} = b + \sum w_i x_i$$

$$= (1\times 0.2) + [(0.1\times 0.3) + (0.2\times 0.1) + (0.3\times 0.5)]$$

$$= \underline{0.39}$$

$$\therefore \text{Sigmoid function} = \frac{1}{1+e^{-x}}$$

$$= \frac{1}{1+e^{-0.39}} = \underline{0.5963}$$

$$\therefore \text{Bipolar Sigmoid} = \frac{2}{1+e^{-x}} - 1$$

$$= \frac{2}{1+e^{-0.39}} - 1 = \underline{0.1926}$$

A perceptron has input weight $w_1 = -3.9$ and $w_2 = 1.1$ with a threshold value, $T = 0.3$ what output does it give from input $x_1 = 1.3$ & $x_2 = 2.2$.

Soln:-

$$y_{in} = \sum w_i x_i$$

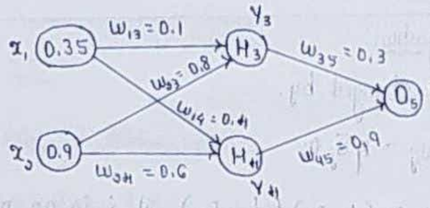
$$= (-3.9 \times 1.3) + (1.1 \times 2.2)$$

$$= \underline{-2.65}$$

$$-2.65 < T \Rightarrow -2.65 < 0.3$$

$\therefore$ False.

Assume that the neuron has a sigmoid activation function, perform a forward pass & backward pass on the network. Assume that the actual output of y is 0.5 and $lr = 1$. Find out the error & weight updation after 1st iteration.



Soln:-

First find net input at Y_3

$$Y_3 = (0.35 \times 0.1) + (0.9 \times 0.8) = \underline{0.755}$$

$$\text{Sigmoid function, } H_3 = \frac{1}{1+e^{-x}} = \frac{1}{1+e^{-0.755}} = \underline{0.6803}$$

Net input at Y_4

$$Y_4 = (0.35 \times 0.4) + (0.9 \times 0.6) = \underline{0.68}$$

$$\text{Sigmoid function, } H_4 = \frac{1}{1+e^{-x}} = \frac{1}{1+e^{-0.68}} = \underline{0.6637}$$

Net input at Y_5,

$$Y_5 = (0.6803 \times 0.3) + (0.6637 \times 0.9) = \underline{0.8014}$$

$$\text{Sigmoid function, } O_5 = \frac{1}{1+e^{0.8014}} = \underline{0.6903}$$

Actual output = 0.5

Error = O_5 - Actual o/p

$= 0.6903 - 0.5$

$= \underline{0.1903}$

Back Propagation:-

Each weight changed by,

$$\Delta w_{ij} = \eta \delta_j O_i$$

$\delta_j = O_j(1-O_j)(d_j - O_j)$ if j is an o/p unit

$\delta_j = O_j(1-O_j)\sum \delta_k w_{kj}$ if j is an hidden unit

where d_j = correct o/p

δ_j = Error measure.

Backward Pass:-

Compute $\delta_3, \delta_4, \delta_5$

$\delta_5 = y_{actual}(1-y_{act})(y_{target} - y_{act})$

$= 0.6903(1-0.6903)(0.5-0.6903)$

$= \underline{-0.0407}$

For hidden units

$\delta_3 = y_3(1-y_3)w_{35}\delta_5$

$= 0.6903(1-0.6903) \times 0.3 \times (-0.0407)$

$= \underline{-0.00265}$

$\delta_4 = y_4(1-y_4)w_{45}\delta_5$

$= (0.6637)(1-0.6637)(0.9 \times -0.0407)$

$= \underline{0.00818}$

$\Delta w_{45\,new} = \eta \delta_5 y_4 = 1 \times -0.0406 \times 0.6637$

$= \underline{-0.0269}$

$w_{45\,new} = \Delta w_{45} + w_{45\,old}$

$= -0.0269 + 0.9$

$= \underline{0.0731}$

$\Delta w_{35} = \eta \delta_5 y_3 = 1 \times -0.0406 \times 0.6803$

$= \underline{-0.0276}$

$w_{35\,new} = \Delta w_{35} + w_{35\,old}$

$= -0.0276 + 0.3$

$= \underline{0.2724}$

Recurrent Neural Network (RNN):-

* It is a type of artificial neural network which uses sequential data or time series data.

Recurrent (Feedback link)

Input → $\boxed{NN}$ → Output

$\therefore h_t = f_w(h_{t-1}, x_t)$

x_t → New input

h_{t-1} → Old state

f_w → Activation function

h_t → New state

Decision Tree :-

Algorithm:

↳ Choose the best feature as root node (based on Entropy & information gain)

↳ Split the dataset

↳ Repeat

↳ Stop when the leaf node are pure.