	R20 Regulations **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPUR** **(Established by Govt. of A.P., ACT No.30 of 2008)** **ANANTAPUR – 515 002 (A.P) INDIA**

Course Code	**Deterministic & Stochastic Statistical Methods**	**L**	**T**	**P**	**C**
20A54404		**3**	**0**	**0**	**3**

Course Objectives

- Study of various Mathematical Methods and Statistical Methods which is needed for Artificial Intelligence, Machine Learning, and Data Science and also for Computer Science and engineering problems.

Course outcomes (CO) : After completion of the course, the student can able to

CO-1: Apply logical thinking to problem-solving in context.

CO-2: Employ methods related to these concepts in a variety of data science applications.

CO-3: Use appropriate technology to aid problem-solving and data analysis.

CO-4: The Bayesian process of inference in probabilistic reasoning system.

CO-5: Demonstrate skills in unconstrained optimization.

Syllabus

UNIT - I- Data Representation

Distance measures, Projections, Notion of hyper planes, half-planes. Principal Component Analysis-Population Principal Components, sample principal coefficients, covariance, matrix of data set, Dimensionality reduction, Singular value decomposition, Gram Schmidt process.

UNIT - II - Single Variable Distribution

Random variables (discrete and continuous), probability density functions, properties, mathematical expectation Probability distribution - Binomial, Poisson approximation to the binomial distribution and normal distribution their properties-Uniform distribution-exponential distribution.

UNIT – III- Stochastic Processes And Markov Chains:

Introduction to Stochastic processes- Markov process. Transition Probability, Transition Probability Matrix, First order and Higher order Markov process, step transition probabilities, Markov chain, Steady state condition, Markov analysis.

UNIT – IV- Multivariate Distribution Theory

Multivariate Normal distribution – Properties, Distributions of linear combinations, independence, marginal distributions, conditional distributions, Partial and Multiple correlation coefficient. Moment generating function. BAYESIAN INFERENCE AND ITS APPLICATIONS: Statistical tests and Bayesian model comparison, Bit, Surprisal, Entropy, Source coding theorem, Joint entropy, Conditional entropy, Kullback- Leibler divergence.

UNIT – V- Optimization

Unconstrained optimization, Necessary and sufficiency conditions for optima, Gradient descent methods, Constrained optimization, KKT conditions, Introduction to non-gradient techniques, Introduction to least squares optimization, Optimization view of machine learning. Data Science Methods: Linear regression as an exemplar function approximation problem, linear classification problems.

Textbooks:

1. Mathematics for Machine Learning by A. Aldo Faisal, Cheng Soon Ong, and Marc Peter Deisenroth
2. Dr.B.S Grewal, Higher Engineering Mathematics, 45th Edition, Khanna Publishers.
3. Operations Research, S.D. Sharma

Reference Books:

1. Operations Research, An Introduction, Hamdy A. Taha, Pearson publishers.
2. A Probabilistic Theory of Pattern Recognition by Luc Devroye,. Laszlo Gyorfi, Gabor Lugosi.

①

Distance Measures

Many algorithms whether supervised (or) unsupervised make use of distance Measures

These measures such as Euclidean distance (or) Cosine similarity Can often be found in algorithms such as K-NN, UMAP, HDBSCAN etc.

Understanding the field of distance measure is more importance than you might realize.

Distance measures plays an important role in machine learning

They Provide the foundation for many Popular and effective machine learning algorithms like K- nearest neighbours for supervised learning and K- means clustering for unsupervised learning.

" Knowing when to use which distance Can help you go from a Poor classifier to an accurate model "

There are many distance measures which ~~are~~ explore how and when they best Can be used.

Some of the main distance measures are follows below.

(1) Euclidean distance

(2) Manhattan distance

(3) Minkowski distance

(4) Cosine Index (or)
Cosine Similarity

(5) Hamming distance.

(6) Chebyshev distance.

(7) Jaccard Index

(2)

- (1) <u>Euclidean distance</u> :-

Euclidean distance is the distance between two points (or) the straight line distance.

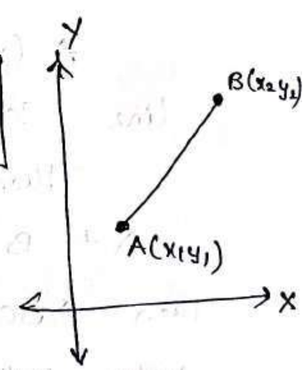
To find the two points on a plane, the length of a segment connecting the two points is measured.

we derive the Euclidean distance formula by using the pythagoras theorem.

<u>Euclidean distance formula</u> :-

Let us assume that (x_1, y_1) & (x_2, y_2) are the two points in a two-dimensional plane. Then the Euclidean distance formula is

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



where

(x_1, y_1) are Co-ordinates of one point

(x_2, y_2) are Co-ordinates of other point

d is the distance between $(x_1 y_1)$ & (x_2, y_2)

(1) What is Euclidean distance formula ?

(A) The Euclidean distance formula is used to find the distance between two points on a plane.

This formula says the distance between the two points (x_1, y_1) & (x_2, y_2) is

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

(2) How to derive Euclidean distance formula ?

(A) To derive the Euclidean distance formula Consider the two Points

$A(x_1, y_1)$ & $B(x_2, y_2)$ and Join them by a line segment.

Then draw horizontal & vertical lines from A to B to meet at C.

Then ABC is a Right angled Δle and hence we can apply pythagorous theorem to it

Then we get $AB^2 = AC^2 + BC^2$

$$d^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$$

Taking square root on both sides

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

3

(3) What are the applications of Euclidean distance formula?

A) The Euclidean distance formula is used to find the length of a line segment given two points on a plane.

Finding distance helps in proving the given vertices form a Square, Rectangle, etc (or) Proving given vertices form an equilateral Δle Right angled Δle etc.

(4) What is the difference between Euclidean distance formula & Manhattan distance formula.

Sol:- For any two points $(x_1 y_1)$ & $(x_2 y_2)$ on a plane

→ The Euclidean distance formula says, the distance between the above points is

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

→ The Manhattan distance formula says, the distance between the above points is

$$\boxed{d = |x_2 - x_1| + |y_2 - y_1|}$$

4

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

Hence the Euclidean distance formula is derived.

Problems.

(1) Find the distance between points $P(3,2)$ & $Q(4,1)$

Sol: given $P(3,2)$ $Q(4,1)$
$x_1\ y_1$ $x_2\ y_2$

Using Euclidean distance formula we have

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

$$PQ = \sqrt{(4-3)^2 + (1-2)^2}$$

$$= \sqrt{(1)^2 + (-1)^2}$$

$$PQ = \sqrt{2} \text{ units}$$

∴ The Euclidean distance between points $A(3,2)$ $B(4,1)$ is $\sqrt{2}$ units.

(2) Prove that points $A(0,4)$ $B(6,2)$ & $C(9,1)$ are Collinear

Sol: To Prove the given three points to be Collinear it is sufficient to prove that the sum of the distances between two pairs of points is equal to the distance between the third pair.

now we will find the distance between every pair of points using the Euclidean distance formula.

$$AB = \sqrt{(6-0)^2 + (2-4)^2}$$

$$= \sqrt{(6)^2 + (-2)^2}$$

$$= \sqrt{36+4} = \sqrt{40} = \underline{\underline{2\sqrt{10}}}.$$

$$BC = \sqrt{(9-6)^2 + (1-2)^2}$$

$$= \sqrt{(3)^2 + (-1)^2} = \sqrt{9+1} = \underline{\underline{\sqrt{10}}}.$$

$$CA = \sqrt{(0-9)^2 + (4-1)^2}$$

$$= \sqrt{(9)^2 + (3)^2}$$

$$= \sqrt{81+9} = \sqrt{90} = \underline{\underline{3\sqrt{10}}}.$$

Here we can see that

$$AB + BC = CA.$$

$$\underline{\underline{2\sqrt{10} + \sqrt{10} = 3\sqrt{10}}}.$$

$\therefore$ we proved that $\underline{\underline{A, B, C}}$ are Collinear.

(3) Check that Points $A(\sqrt{3}, 1)$, $B(0,0)$ $C(2,0)$ are the vertices of an Equilateral Δle.

Sol :- Three vertices A, B & C are vertices of an equilateral Δle $\Longleftrightarrow$ If $AB = BC = CA$.

given $A(\sqrt{3}, 1)$ $B(0,0)$ $C(2,0)$

$\quad\quad x_1 \; y_1 \quad\quad x_2 \; y_2 \quad\quad x_3 \; y_3$

Using Euclidean distance formula.

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

$$= \sqrt{(0-\sqrt{3})^2+(0-1)^2}$$

$$= \sqrt{(3+1} = \sqrt{4} = \underline{\underline{2}}.$$

$$BC = \sqrt{(x_3-x_2)^2+(y_3-y_2)^2}$$

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

$$= \sqrt{4+0} = \sqrt{4} = \underline{\underline{2}}.$$

$$CA = \sqrt{(x_3-x_1)^2+(y_3-y_1)^2}$$

$$= \sqrt{(2-\sqrt{3})^2+(0-1)^2}$$

$$= \sqrt{(9+25)}$$

$$= \sqrt{34}$$

Here $AB = BC \neq CA$.

$\therefore$ A, B & C are not the vertices of an equilateral Δle.

***

4) Difference between Euclidean Distance formula and Manhattan Distance formula?

For any two points $(x_1 y_1)$ & $(x_2 y_2)$ on a plane

(1) The Euclidean distance formula says, the distance between the above points is $d = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$

(2) The Manhattan distance formula says, the distance between the above points is $d = |x_2-x_1| + |y_2-y_1|$

5) Calculate the Euclidean distance between ~~Skoopsu~~ ~~ABBAQ~~ ~~where~~ the points (1,1,0) & (4,5,0) A in XY plane.

Sol: distance between points

$(1,1,0)$ $(4,5,0)$
$x_1 y_1 z_1$ $x_2 y_2 z_2$

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

$= \sqrt{(4-1)^2 + (5-1)^2}$

$= \sqrt{(3)^2 + (4)^2} = \sqrt{9+16} = \sqrt{25} = \underline{5 \text{ Units}}$

6) Calculate the distance between the two points

A $(-5,2,4)$ & B $(-2,2,0)$
$x_1 y_1 z_1$ $x_2 y_2 z_2$.

$d = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2}$

$= \sqrt{(-2+5)^2 + (2-2)^2 + (0-4)^2}$

$= \sqrt{(3)^2 + (0)^2 + (-4)^2}$

$= \sqrt{9+16} = \sqrt{25} = \underline{5 \text{ Units}}$.

6

7) The distance between $(1,2,3)$ & $(4,5,6)$ will be

Sol :- $(1,2,3)$ $(4,5,6)$
$x_1 y_1 z_1$ $x_2 y_2 z_2$

$$D = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}$$

$$= \sqrt{(4-1)^2+(5-2)^2+(6-3)^2}$$

$$= \sqrt{(3)^2+(3)^2+(3)^2}$$

$$= \sqrt{9+9+9} = \sqrt{27} \text{ Units}$$

(8) The distance between $P(-1,2,3)$ & $(4,0,-3)$ is

Sol : $(-1,2,3)$ $(4,0,-3)$
$x_1 y_1 z_1$ $x_2 y_2 z_2$

$$D = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}$$

$$= \sqrt{(4+1)^2+(0-2)^2+(-3-3)^2}$$

$$= \sqrt{(5)^2+(-2)^2+(-6)^2}$$

$$= \sqrt{25+4+36} = \sqrt{65} \text{ units}$$

(9) The distance of the point $(5,0,12)$ from the Origin $(0,0,0)$ is

Sol :- $(5,0,12)$ $(0,0,0)$
$x_1 y_1 z_1$ $x_2 y_2 z_2$

$$D = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2} = \sqrt{\cancel{[illegible]}}$$

$(5, 0, 12)$. $(0\ 0\ 0)$
$x_1\ y_1\ z_1 \quad x_2\ y_2\ z_2$

$$D = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2}$$

$$= \sqrt{(0-5)^2 + (0-0)^2 + (0-12)^2}$$

$$= \sqrt{(-5)^2 + (0)^2 + (-12)^2}$$

$$= \sqrt{25+144} = \sqrt{169} = 13 \text{ units}$$

$$\begin{array}{r} 13 \\ 13 \\ \hline 39 \\ 13\times \\ \hline 169 \end{array}$$

10) The distance between $A(2,-1)$ & $B(2,3)$ is

Sol :- $(2,-1)\ (2,3)$
$x_1\ y_1 \quad x_2\ y_2$

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

$$= \sqrt{(2-2)^2 + (3+1)^2}$$

$$= \sqrt{(0)^2 + (4)^2} = \sqrt{16} = 4 \text{ units}$$

Manhattan distance :- Formula is.

For 2 vectors $\boxed{d = |x_1 - x_2| + |y_1 - y_2|}$

For 3 vectors. $\boxed{d = |x_1 - x_2| + |y_1 - y_2| + |z_1 - z_2|}$

7

(d) Cosine - Correlation distance :-

$$\boxed{\cos(A,B) = \frac{A \cdot B}{\|A\| \cdot \|B\|}}$$

$A \cdot B = (1)(2) + (0)(1) + (2)(0) + (5)(3) + (3)(-1)$

$A \cdot B = \underline{14}$.

$\|A\| = \sqrt{1^2 + 0^2 + 2^2 + 5^2 + 3^2} = \underline{6.24}$

$\|B\| = \sqrt{2^2 + 1^2 + 0^2 + 3^2 + (-1)^2} = \underline{3.87}$..

$\therefore \cos(A,B) = \dfrac{A \cdot B}{\|A\| \cdot \|B\|} = \dfrac{14}{(6.24)(3.87)} = \underline{0.57}$.

~~Distance Measure :-~~

~~Distance measures play an important role in machine learning.~~

~~They provide the foundation for many popular and effective machine learning algorithms like K-nearest neighbours for supervised learning and K-Means clustering for unsupervised learning.~~

---

(4) Cosine - Correlation distance :-

$$\boxed{Cos(A,B) = \frac{A \cdot B}{||A|| \cdot ||B||}}$$

$A = (1, 0, 2, 5, 3)$
$B = (2, 1, 0, 3, -1)$

$$A \cdot B = (1)(2) + (0)(1) + (2)(0) + (5)(3) + (3)(-1) = \underline{14}.$$

$$||A|| = \sqrt{1^2 + 0^2 + 2^2 + 5^2 + 3^2} = \underline{6.24}$$

$$||B|| = \sqrt{2^2 + 1^2 + 0^2 + 3^2 + (-1)^2} = \underline{3.87}.$$

$$\therefore Cos(A,B) = \frac{14}{6.24 * 3.87} = \underline{0.57}.$$

→ Cosine distance measure for clustering determines the cosine of the angle between two vectors given by the formula $Cos(A,B) = \frac{A \cdot B}{||A|| \cdot ||B||}$

8

1) $X_1 = (1, 2, 2)$ $x_1\ y_1\ z_1$ $X_2 = (2, 5, 3)$ $x_2\ y_2\ z_2$

Manhattan (L_1) :-

$$L_1 = |1-2| + |2-5| + |2-3|$$
$$= |-1| + |-3| + |-1|$$
$$= 1 + 3 + 1 = \underline{5}.$$

Euclidean :- (L_2) :-

$$L_2 = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2}$$
$$= \sqrt{(2-1)^2 + (5-2)^2 + (3-2)^2}$$
$$= \sqrt{(1)^2 + (3)^2 + (1)^2}$$
$$= \sqrt{1+9+1}$$
$$= \underline{\sqrt{11}}\ \text{Units}.$$

(1) Given 5 Dimensional Samples

$A = (1, 0, 2, 5, 3)$

$B = (2, 1, 0, 3, -1)$ then Find.

(1) Euclidean distance between Points :- $\boxed{\sqrt{(A_K - B_K)^2}}$

$= \sqrt{(1-2)^2 + (0-1)^2 + (2-0)^2 + (5-3)^2 + (3+1)^2}$

$= \underline{\underline{5.09}}$.

(2) City block / Manhattan distance :-

$$\boxed{\sum_{K=1}^{n} |x_{iK} - x_{jK}|}$$

$d_{AB} = |1-2| + |0-1| + |2-0| + |5-3| + |3+1|$

$= |-1| + |-1| + |2| + |2| + |4|$

$= 1 + 1 + 2 + 2 + 4$

$= \underline{\underline{10}}$

(3) Minkowski distance :-

given external Variable $\boxed{P = 3}$

$= \left[\sum |A_K - B_K|^3 \right]^{1/3}$

$= \left[|1-2|^3 + |0-1|^3 + |2-0|^3 + |5-3|^3 + (3+1)^3 \right]^{1/3}$

$= \underline{\underline{4.34}}$.

$$\boxed{\left[\sum |A_K - B_K|^P \right]^{1/P}}$$

(9)

(2) Manhattan distance :-

This determines the absolute difference among the Pair of the Coordinates.

Suppose we have two points P and Q to determine the distance between these Points we simply have to Calculate the Perpendicular distance of the points from X-axis & Y- axis In a plane with P at Coordinate $(x_1\ y_1)$ and Q at $(x_2\ y_2)$

Manhattan distance between P & Q is.

$$\boxed{d = |x_2 - x_1| + |y_2 - y_1|}$$

The Manhattan distance, often Called as "Taxicab distance" (or) City Block distance. Calculates the distance between real - Valued vectors Imagine. vectors that describe objects on a Uniform grid such as a Chessboard.

Manhattan distance then refers to the distance between two vectors If they Could only move right angles. There is no diagonal movement involved. in Calculating the distance.

The Manhatten distance between two points $(x_1\ y_1)$ & $(x_2\ y_2)$ is given as

$$|x_1 - x_2| + |y_1 - y_2| \quad \text{(or)} \quad |x_2 - x_1| + |y_2 - y_1|$$

(1) Find the Manhatten distance between the points given below

(1) $(1, 2)\ (3, 4)$
$\quad x_1\ y_1 \quad x_2\ y_2$

$\Longrightarrow |3-1| + |4-2|$

$\Longrightarrow 2+2 = \underline{\underline{4}}$

(2) $(-4, 6)\ (3, -4)$
$\quad x_1\ y_1 \quad x_2\ y_2$

Manhatten.

$\Longrightarrow |3+4| + |-4-6|$

$\Longrightarrow |7| + |-10|$

$\Longrightarrow 7+10 = \underline{\underline{17}}$.

$\Rightarrow$ Manhatten distance is the most Preferable for high dimensional applications.

Thus Manhatten distance is Preferred over the Euclidean distance metric as the dimension of the data increases.

$\Rightarrow$ If we need to calculate the distance between two data points in a grid-like path we use Manhattan.

(10)

(3) Calculate the Manhattan distance from the points given below

$X_1 = (1, 2, 3, 4, 5, 6)$

$X_2 = (10, 20, 30, 1, 2, 3)$

$\Rightarrow |10-1| + |20-2| + |30-3| + |1-4| + |2-5| + |3-6|$

$\Rightarrow 9 + 18 + 27 + 3 + 3 + 3$

$\Rightarrow \underline{\underline{63}}$.

(3) Minkowski distance :-

Minkowski distance is a distance measured between two points in N- dimensional space.

It is basically a generalization of the Euclidean distance and Manhattan distance :

It is widely used in field of machine Learning especially in the Concept to find the optimal Correlation or Classification of data

Minkowski distance is used in Certain algorithms like K- Nearest Neighbors, LVQ. (Learning Vector Quantization), SOM (Self Organizing Map) and K- Means clustering

→ Let us Consider 2-dimensional space having three Points

$P_1(x_1 y_1)$, $P_2(x_2 y_2)$ $P_3(x_3 y_3)$

The Minkowski distance is given by

$$\left(|x_1 - y_1|^p + |x_2 - y_2|^p + |x_3 - y_3|^p \right)^{1/p}$$

(or)

The formula for Minkowski distance is given as

$$D(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}}.$$

Most interestingly about this distance measure is use of Parameter P.

we can use this Parameter to manipulate the distance metrics to closely resemble Others.

Common values of p are :-

(1) $P = 1 \implies$ Manhattan distance.

(2) $P = 2 \implies$ Euclidean distance

(3) $P = \infty \implies$ Chebyshev distance.

11

Problem

(1) Given 5 dimensional samples

$A = (1, 0, 2, 5, 3)$

$B = (2, 1, 0, 3, -1)$ & external variable $P = 3$.

(or) parameter $P = 3$

Sol :- $D(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^P \right)^{\frac{1}{P}}$

$= \left[|1-2|^3 + |0-1|^3 + |2-0|^3 + |5-3|^3 + |3+1|^3 \right]^{1/3}$

Here $P = 3$

$= \underline{4.34}$

(2) we are given two vectors vector A & vector B as

$A = (4, 2, 6, 8)$

$B = (5, 1, 7, 9)$ Find Minkowski distance for $\boxed{P = 2}$

Sol :- $\left[|4-5|^2 + |2-1|^2 + |6-7|^2 + |8-9|^2 \right]^{1/2}$

$= \underline{2}$.

(3) Calculate the Minkowski distance between two vectors using a power of $\boxed{P=3}$

$A = (2, 4, 4, 6)$

$B = (5, 5, 7, 8)$

Ans :- 3.979057

(4) $A = (2, 4, 4, 6)$

$B = (5, 5, 7, 8)$

$C = (9, 9, 9, 8)$

$D = (4, 2, 3, 3)$

Note :- Each vector in the matrix should be the same length.

The Minkowski distance between.

A & B is 3.98.

The Minkowski distance between

A & C is 8.43

The Minkowski distance between

A & D is 3.33

The Minkowski distance between

B & C is 5.14

B & D is 6.54

C & D is 10.61.

12

- Problem :-

1) Given two Objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8)

a) Compute the Euclidean distance between two Objects

b) Compute the Manhattan distance between two Objects

c) Compute the Minkowski distance between the two Objects using p=3

Sol :- a) Euclidean distance :-

(22, 1, 42, 10) (20, 0, 36, 8)

$$= \sqrt{|22-20|^2 + |1-0|^2 + |42-36|^2 + |10-8|^2}$$

$$= \underline{\underline{6.71}}.$$

b) Manhattan distance :-

(22, 1, 42, 10) (20, 0, 36, 8)

$$= |22-20| + |1-0| + |42-36| + |10-8|$$

$$= \underline{\underline{11}}.$$

(c) Minkowski distance :-

$$\left(|22-20|^3 + |1-0|^3 + |42-36|^3 + |10-8|^3\right)^{1/3}$$

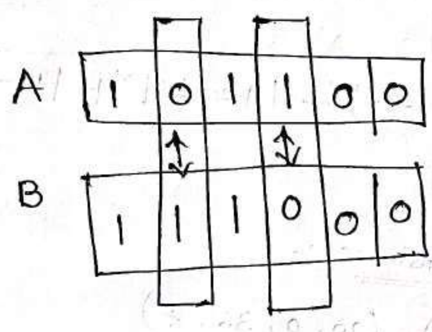
$$= \underline{\underline{6.15}}$$

(4) Hamming distance :-

Hamming distance is the number of values that are different between two vectors.

It is typically used to Compare two binary strings of equal length

It can also be used for strings to Compare how Similar they are to each other by Calculating the number of Characters that are different from each other.



13

(1) Find the Hamming distance between the Code words of

$C = \{(0000), (0101)\ (1011)\ (0111)\}$

Sol :- Let

$x = 0000$

$y = 0101$

$z = 1011$

$w = 0111$

$d(x, y) = \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{matrix} = \underline{\underline{2}}$

$d(x, z) = \begin{matrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{matrix} = \underline{\underline{3}}$

$d(x, w) = \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{matrix} = \underline{\underline{3}}$

$d(y, z) = \begin{matrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{matrix} = \underline{\underline{3}}$

$d(y, w) = \begin{matrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{matrix} = \underline{\underline{1}}$

$d(z, w) = \begin{matrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{matrix} = \underline{\underline{2}}$

(15) Chebyshev distance :-

Chebyshev distance is defined as the greatest of difference between two vectors along any Coordinate dimension.

In Otherwords, it is Simply the maximum distance along One axis

Due to its nature, it is often referred as Chessboard distance since the minimum number of moves needed by a King to go from one square to another is equal to Chebyshev distance.

$$\boxed{D(x,y) = \max_i \left(|x_i - y_i| \right)}$$

→ Consider two Points P_1 & P_2 with Coordinates as follows

$$P_1 = (P_1, P_2, P_3 \ldots P_N)$$
$$P_2 = (q_1, q_2, q_3 \ldots q_N)$$

Then the Chebyshev distance between the two Points P_1 & P_2 is

$$\text{Chebyshev distance} = \max \left(|P_i - q_i| \right)$$

14

- 1) The Point A has Coordinate (0, 3, 4, 5) and Point B has Coordinate (7, 6, 3, −1)

The Chebyshev distance between Point A & B is

$$d_{AB} = \text{Max}\{|0-7|, |3-6|, |4-3|, |5+1|\}$$

$$= \text{Max}\{7, 3, 1, 6\} = \underline{\underline{7}}.$$

2) distance $(A,B) = \text{Max}(|x_A - x_B|, |y_A - y_B|)$

distance $(A,B) = \text{Max}(|70-330|), |40-220|)$

distance $(A,B) = \text{max}(|-260|, |-188|)$

distance $(A,B) = \text{max}(260, 188)$

$\boxed{\text{distance } (A,B) = 260.}$

(6) Jaccard Index :-

The Jaccard distance measures the Similarity of the two data set items as the Intersection of those items divided by the Union of the data items

$$\boxed{J(A, B) = \frac{|A \cap B|}{|A \cup B|}}$$

where, J = Jaccard distance
A = Set-1
B = Set-2

→ To Calculate the Jaccard distance we Simply subtract the Jaccard index from '1'

$$\boxed{D(x, y) = 1 - \frac{|A \cap B|}{|A \cup B|}}$$

15

Hyperplane, Subspace & Halfspace

(1) Hyperplane :-

Geometrically, a hyperplane is a geometric entity whose dimension is one less than that of its ambient space.

what does it mean ?

It means the following

For example ;

If you take the 3D space then hyperplane is a geometric entity that is '1' dimensionless so its going to be 2 dimensions and a 2 dimensional entity in a 3D space would be a plane.

Now If you take 2 dimensions, then '1' dimensionless would be a single-dimensional geometric entity, which would be a line and so on.

(1) The hyperplane is usually described by an equation as follows

$$\boxed{X^T n + b = 0}$$

(2) If we expand this out for 'n' Variables we will get Something like this

$$X_1 n_1 + X_2 n_2 + X_3 n_3 + \cdots + X_n n_n + b = 0$$

(3) In just two dimensions we will get Something like this which is nothing but an equation of a line

$$\boxed{X_1 n_1 + X_2 n_2 + b = 0.}$$

Ex:- Let us Consider a 2D geometry with

$$n = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \& \quad b = 4$$

Though its a 2D geometry the Value of X will be $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

So according to the equation of hyperplane it Can be Solved as

$$X^T n + b = 0$$

$$[x_1 \; x_2] \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 4 = 0$$

$$\boxed{x_1 + 3x_2 + 4 = 0}$$

So as you Can see from the Solution the hyperplane is the equation of a line.

(16)

(2) Subspace :-

Hyper-planes, in general, are not subspaces.

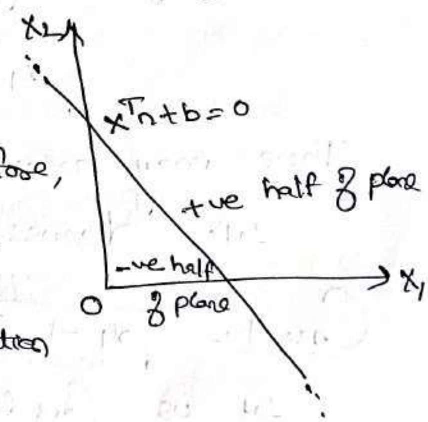
However, If we have hyper-planes of the form

$$\boxed{x^T n = 0}$$

That is if the plane goes through the Origin then a hyperplane also becomes a Subspace.

(3) Half-space :-

Consider this 2-dimensional Picture given below



so here we have a 2-dimensional space in X_1 & X_2 and as we have discussed before, an equation in two dimensions would be a line which would be a hyperplane. so the equation to the line is written as

$$\boxed{X^T n + b = 0.}$$

so, for this two dimensions, we could write this line as we discussed Previously

$$\boxed{X_1 n_1 + X_2 n_2 + b = 0.}$$

You Can notice from the above graph that this whole two-dimensional space is broken into two spaces.

one on this side (+ve half of plane) of a line and the other one on this side (-ve half of the plane) of a line. Now these two spaces are called as Half-spaces.

Example :- Let us consider the same example that we have taken in hyperplane case.

So by solving, we got the equation as

$$\boxed{x_1 + 3x_2 + 4 = 0}$$

There may arise 3 Cases

Let's discuss each case with an example.

Case-1 :- $\boxed{x_1 + 3x_2 + 4 = 0} \rightarrow$ On the line

Let us consider two points $(-1, -1)$, when we put this value on the equation of line we got '0' so we can say that this point is on the hyperplane of the line.

17

- Case-2 :- Hey $\boxed{x_1 + 3x_2 + 4 > 0} \rightarrow$ Positive Half space.

Consider two points $(1, -1)$, when we put this value on the equation of line we got '2' which is greater than '0' so we can say that this point is on the Positive Half space.

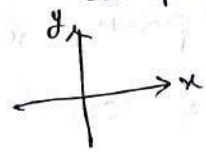
Case-3 :- Hey $\boxed{x_1 + 3x_2 + 4 < 0} \rightarrow$ Negative Half-space

Consider two points $(1, -2)$, when we put this value on the equation of line we got -1 which is less than '0' so we can say that this point is on the Negative Half space.

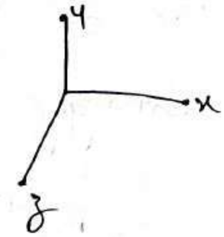
(18)

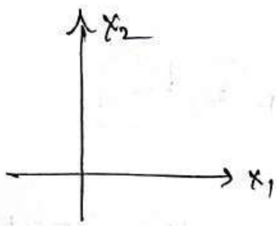
Hyper plane :-

2D space



3D space





2D

x_1 & x_2 hyperplane is $\boxed{c_1x_1 + c_2x_2 = K}$ x_1 x_2 value set or

x_1 & x_2, x_3 hyper plane is $\boxed{c_1x_1 + c_2x_2 + c_3x_3 = K}$ x_1 x_2 x_3 Collection is hyperplane

satisfying $x_1 x_2 x_3$ value set or Collection is called as hyper plane.

Hyperplane :-

In R^n (ie:- n-dimensional space) the set of Points $x = (x_1 x_2 \dots x_n)$ satisfying the equation

$$c_1x_1 + c_2x_2 + \dots + c_nx_n = K \longrightarrow (1)$$

(not all $c_i = 0$)

is Called a hyper plane for given value of c^is

⟶ As Particular case in R^3 (ie - 3-dimensional space) the set of Points $x = (x_1 x_2 x_3)$ satisfying the equation $\boxed{c_1x_1 + c_2x_2 + c_3x_3 = K}$.

→ As Particular Case in $\mathbb{R}^4$ (ie:- 4 dimensional space)

the set of points $x = (x_1, x_2, x_3, x_4)$

Satisfying the equation

$$\boxed{c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 = K}$$

Notes:- In a Linear Programming Problem the Objective function and the Constraints equations represents the hyperplanes

Notes:- If $\boxed{K=0}$ then the hyperplane is said to passes through the Origin, and then its equation can be written in the form

$$\boxed{c_1x_1 + c_2x_2 + c_3x_3 + \cdots + c_nx_n = 0}$$

Notes:- In matrix notation the equation of the hyperplane ① can be written as $\boxed{CX = K}$

where C is row vector

$$C = [c_1 \, c_2 \, \cdots \, c_n]$$

and X is Column vector

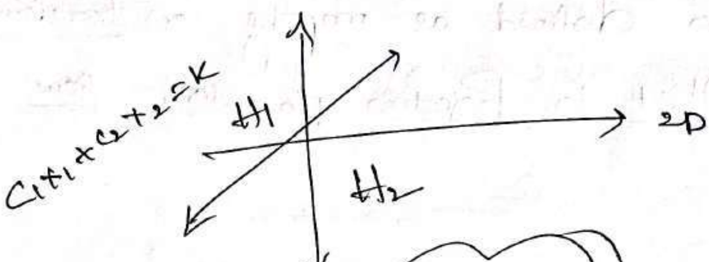
$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [x_1 \, x_2 \, \cdots \, x_n]$$

and K is a Constant matrix

→ If the hyperplane passes through the origin then its equation is $\boxed{Cx = 0}$

→ If a hyperplane divides R^n into two half space which can be denoted by

$$H_1 = \{ x \mid Cx \geq K \}$$
$$H_2 = \{ x \mid Cx \leq K \}$$



$Cx = K$

$Cx \geq K$

$Cx \leq K$

H_1 is the Half space

ie:- that portion of R^n that contain the vectors x for which $\boxed{Cx \geq K}$ and

H_2 is the Halfspace

ie:- that contain the vector x for which $\boxed{Cx \leq K}$.

Projection :-

Representing n- dimensional Object into (n-1) dimension is Known as Projection.

→ It is the Process of Converting a 3D Object into a 2D Object

→ It is also defined as mapping (or) transformation of the Object in Projection plane (or) View plane.

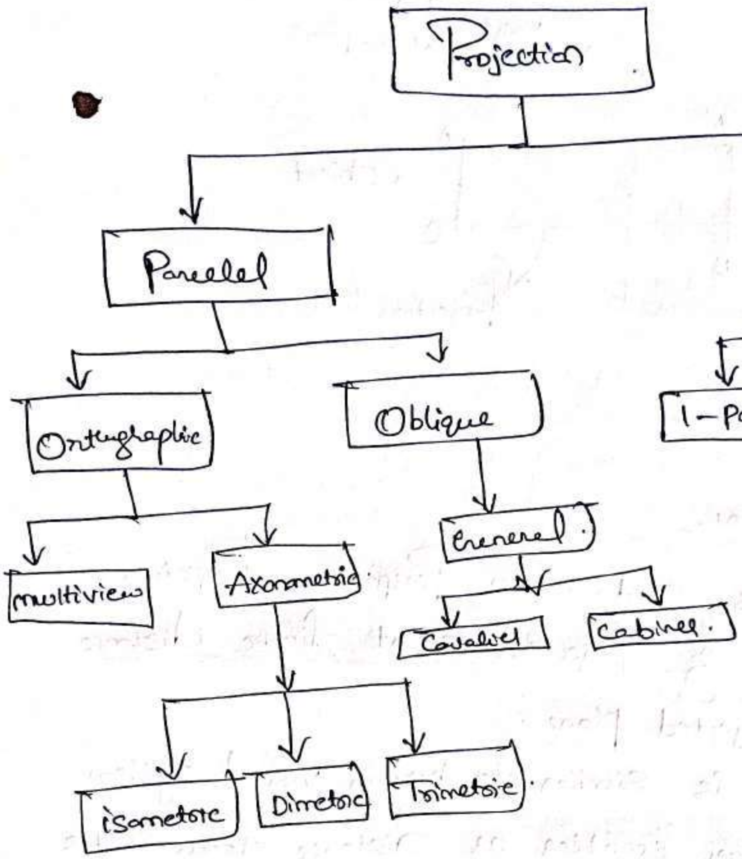
(3D becomes 2D
3D − ~~1~~ = 2D
(n − 1))

Projection are of two types

(1) Parallel Projection

(2) Perspective Projection

20



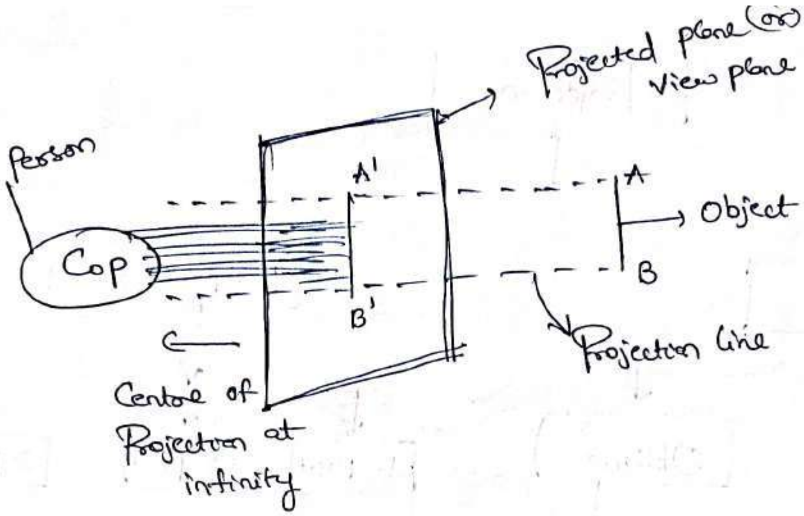
Overview of a Projection.

(1) Parallel Projection:

In this, Coordinate Positions are transformed to the View plane along Parallel lines.

→ A Projection is said to be Parallel If Centre of Projection is at infinite distance from the Projected plane.

→ The Projection lines are Parallel to each other and extended from the Object and intersect the view plane.



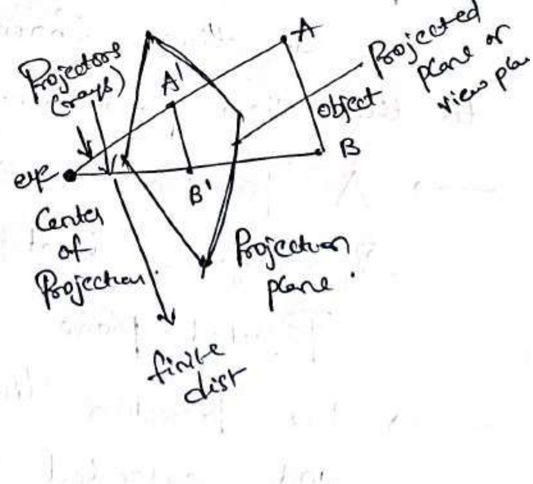
(2) Perspective Projection :-

→ The Projection is said to be Perspective Projection, if the Center of Projection is at finite distance from the Projected plane.

→ Visual effect is similar to human visual system

→ Objects appear smaller as distance from Center of Projection (Cop) (eye of Observer) increases.

→ Difficult to determine exact size and shape of the Object.



Dimension Reduction Techniques :-

The two popular and well-known dimension reduction techniques are

Dimension Reduction Techniques

- Principal Component Analysis (PCA)
- Fisher Linear Discriminant Analysis (LDA)

(1) Principal Component Analysis :- (PCA)

→ Principal Component Analysis is a Well-Known dimension reduction technique.

→ It transforms the Variables into a new set of Variables Called as Principal Components

→ These Principal Components are Linear Combination of Original variables and are Orthogonal

→ The First Principal Component accounts for most of the possible Variation of Original data.

→ The second Principal Component does its best to Capture the Variance in the data.

→ There Can be only two Principal Components for a two-dimensional data set.

PCA Algorithm :-

The steps involved in PCA Algorithm are as follows

Step-1 :- Get data

Step-2 :- Compute the mean vector (μ)

Step-3 :- Subtract mean from the given data

Step-4 :- Calculate the Co-variance matrix.

Step-5 :- Calculate the eigen vectors & eigen values of the Co-variance matrix

Step-6 :- Choosing Components and forming a feature vector

Step-7 :- Deriving the new data set.

(22)

Problems

(1) Given data = $\{2, 3, 4, 5, 6, 7 ; 1, 5, 3, 6, 7, 8\}$
Compute the Principal Component using PCA Algorithm.

(or)

Consider the two dimensional patterns
(2,1) (3,5) (4,3) (5,6) (6,7) (7,8)
Compute the Principal Component using PCA Algorithm

(or)

Compute the Principal Component of following data

Class-1 :- $X = 2, 3, 4$
$Y = 1, 5, 3$

Class-2 :- $X = 5, 6, 7$
$Y = 6, 7, 8$

Sol:- we use the above discussed PCA Algorithm

Step-1 :- Get data

The given feature vectors are

$X_1 = (2,1)$ $\quad X_4 = (5,6)$
$X_2 = (3,5)$ $\quad X_5 = (6,7)$
$X_3 = (4,3)$ $\quad X_6 = (7,8)$

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} \begin{bmatrix} 6 \\ 7 \end{bmatrix} \begin{bmatrix} 7 \\ 8 \end{bmatrix}$$

Step-2 :- Compute the Mean vector (μ)

Calculate the Mean vector (μ)

Mean vector (μ) =

$$= \frac{(2+3+4+5+6+7)}{6}, \frac{(1+5+3+6+7+8)}{6}$$

$$= (4.5, 5)$$

Thus Mean vector (μ) = $\begin{bmatrix} 4.5 \\ 5 \end{bmatrix}$

Step-3 :- Subtract mean vector (μ) from the given feature vectors

$$X_1 - \mu = (2-4.5,\ 1-5) = (-2.5, -4)$$
$$X_2 - \mu = (3-4.5,\ 5-5) = (-1.5, 0)$$
$$X_3 - \mu = (4-4.5,\ 3-5) = (-0.5, -2)$$
$$X_4 - \mu = (5-4.5,\ 6-5) = (0.5, 1)$$
$$X_5 - \mu = (6-4.5,\ 7-5) = (1.5, 2)$$
$$X_6 - \mu = (7-4.5,\ 8-5) = (2.5, 3)$$

Feature vectors (X_i) after subtracting mean vector (μ) are

$$\begin{bmatrix} -2.5 \\ -4 \end{bmatrix} \begin{bmatrix} -1.5 \\ 0 \end{bmatrix} \begin{bmatrix} -0.5 \\ -2 \end{bmatrix} \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \begin{bmatrix} 1.5 \\ 2 \end{bmatrix} \begin{bmatrix} 2.5 \\ 3 \end{bmatrix}$$

23

Step-4 :-

Calculate the Covariance matrix

Covariance matrix is given by

$$\boxed{\text{Covariance Matrix} = \frac{\sum (x_i - \mu)(x_i - \mu)^t}{n}}$$

Now

$$m_1 = (x_1 - \mu)(x_1 - \mu)^t = \begin{bmatrix} -2.5 \\ -4 \end{bmatrix} \begin{bmatrix} -2.5 & -4 \end{bmatrix} = \begin{bmatrix} 6.25 & 10 \\ 10 & 16 \end{bmatrix}$$

$$m_2 = (x_2 - \mu)(x_2 - \mu)^t = \begin{bmatrix} -1.5 \\ 0 \end{bmatrix} \begin{bmatrix} -1.5, & 0 \end{bmatrix} = \begin{bmatrix} 2.25 & 0 \\ 0 & 0 \end{bmatrix}$$

$$m_3 = (x_3 - \mu)(x_3 - \mu)^t = \begin{bmatrix} -0.5 \\ -2 \end{bmatrix} \begin{bmatrix} -0.5 & -2 \end{bmatrix} = \begin{bmatrix} 0.25 & 1 \\ 1 & 4 \end{bmatrix}$$

$$m_4 = (x_4 - \mu)(x_4 - \mu)^t = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \begin{bmatrix} 0.5 & 1 \end{bmatrix} = \begin{bmatrix} 0.25 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

$$m_5 = (x_5 - \mu)(x_5 - \mu)^t = \begin{bmatrix} 1.5 \\ 2 \end{bmatrix} \begin{bmatrix} 1.5 & 2 \end{bmatrix} = \begin{bmatrix} 2.25 & 3 \\ 3 & 4 \end{bmatrix}$$

$$m_6 = (x_6 - \mu)(x_6 - \mu)^t = \begin{bmatrix} 2.5 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 2.5 & 3 \end{bmatrix} = \begin{bmatrix} 6.25 & 7.5 \\ 7.5 & 9 \end{bmatrix}$$

Now Covariance Matrix

$$= \frac{(m_1 + m_2 + m_3 + m_4 + m_5 + m_6)}{6}$$

On adding the above matrices and dividing by 6, we get

Covariance Matrix $= \frac{1}{6}\begin{bmatrix} 17.5 & 22 \\ 22 & 34 \end{bmatrix}$

Covariance Matrix $= \begin{bmatrix} 2.92 & 3.67 \\ 3.67 & 5.67 \end{bmatrix}$

Step-5 :-

Calculate the eigen values and eigen vectors of the Covariance matrix.

λ is an eigen value for a matrix M if it is a solution of Characteristic equation

$|M - \lambda I| = 0.$

So, we have

$$\begin{vmatrix} 2.92 & 3.67 \\ 3.67 & 5.67 \end{vmatrix} - \begin{vmatrix} \lambda & 0 \\ 0 & \lambda \end{vmatrix} = 0$$

$$\begin{vmatrix} 2.92-\lambda & 3.67 \\ 3.67 & 5.67-\lambda \end{vmatrix} = 0$$

From here

$(2.92-\lambda)(5.67-\lambda) - (3.67 \times 3.67) = 0$

$16.56 - 2.92\lambda - 5.67\lambda + \lambda^2 - 13.47 = 0$

$\lambda^2 - 8.59\lambda + 3.09 = 0$

On solving this quadratic equation we get

$\lambda = 8.22, 0.38$

24

Thus two eigen values are

$\lambda_1 = 8.22$

$\lambda_2 = 0.38$

Clearly the second eigen value is very small compared to the first eigen value

So, the second eigen vector can be left out

Eigen vector corresponding to the greatest eigen value is the Principal Component for the given data set

So we find the eigen vector corresponding to eigen value λ_1

we use the following equation to find the eigen vector $\boxed{MX = \lambda X}$

where M = Covariance Matrix

X = Eigen vector

λ = Eigen value.

On Substituting the values in the above equation we get

$$\begin{bmatrix} 2.92 & 3.67 \\ 3.67 & 5.67 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = 8.22 \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

Solving these we get

$$2.92\,X_1 + 3.67\,X_2 = 8.22\,X_1$$

$$3.67\,X_1 + 5.67\,X_2 = 8.22\,X_2$$

on Simplification we get

$$5.3\,X_1 = 3.67\,X_2 \longrightarrow (1)$$

$$3.67\,X_1 = 2.55\,X_2 \longrightarrow (2)$$

From (1) & (2)

$x_1 = 0.69 x_2$

From (2) the eigen vector is

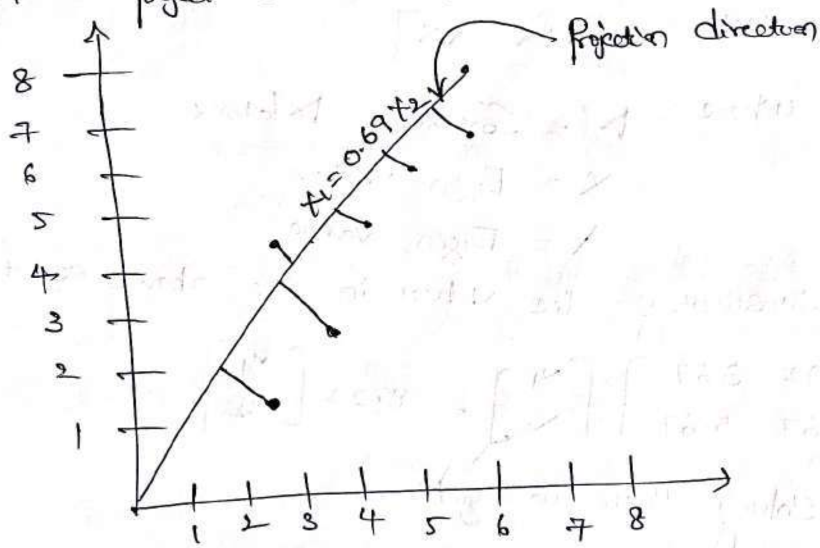
Eigen vector $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2.55 \\ 3.67 \end{bmatrix}$

Thus, Principal Component for the given data set is

Principal Component

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2.55 \\ 3.67 \end{bmatrix}$$

Lastly, we project the data points onto the new subspace as



(2) Use PCA Algorithm to transform the pattern (2,1) onto the eigen vector in the Previous question (25)

Sol: The given feature vector is (2,1)

Given Feature vector :- $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$

The Feature vector gets transformed to =

Transpose of eigen vector X (Feature vector − Mean vector)

$$= \begin{bmatrix} 2.55 \\ 3.67 \end{bmatrix}^T \times \left(\begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 4.5 \\ 5 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 2.55 & 3.67 \end{bmatrix} \times \begin{bmatrix} -2.5 \\ -4 \end{bmatrix}$$

$$= \underline{\underline{-21.055}}$$

Problem on PCA

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

Feature	Example 1	Example 2	Example 3	Example 4
x	4	8	13	7
y	11	4	5	14

Sol :- Step-1 :- Data set :-

Feature	Example 1	Example 2	Example 3	Example 4
x	4	8	13	7
y	11	4	5	14

No of features, $n = 2$ (x, y)

No of Sample $N = 4$

Step-2 :- Computation of Mean of Variables

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

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

Step-3 :- Computation of Co-variance Matrix

Ordered Pairs are (x, y) $n^2 = 2^2 = 4$

n variables

(x, x) (x, y) (y, x) (y, y)

(26)

1) Co-variance of all Ordered Pairs

$$Cov(x_i, x) = \frac{1}{N-1} \sum_{K=1}^{N} (x_{iK} - \overline{x}_i)(x_{jK} - \overline{x}_j)$$

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

$$= \underline{\underline{14}}.$$

$Cov(x, x) \longrightarrow$ If two variables are same then

$$\boxed{Cov(x, x) = \frac{1}{N-1} \sum_{K=1}^{N} (x_i - \overline{x})^2}$$

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

$$= \underline{\underline{-11}}.$$

$$Cov(y, x) = Cov(x, y) = -11$$

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

$$= \underline{\underline{23}}$$

$\therefore$ we have

$$\boxed{\begin{aligned} Cov(x, x) &= 14 \\ Cov(x, y) &= -11 \\ Cov(y, x) &= -11 \\ Cov(y, y) &= 23. \end{aligned}}$$

Using all these 4 Co-varience values we are going to Construct Co-varience Method of size $n \times n$, 2×2.

$$S = \begin{bmatrix} Cov(x,x) & Cov(x,y) \\ Cov(y,x) & Cov(y,y) \end{bmatrix}$$ Co-verience matrix.

Covarience matrix values

Covarience Matrix $S = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$ This is Co-varience Matrix S.

~~Fisch~~

Step 4 :- Eigen value, Eigen vector, Normalized Eigen vector.

1) Eigen value.

$$|S - \lambda I| = 0 \qquad I = 2\times 2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0 \qquad \lambda I = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

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

$$\Rightarrow \left| \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} \right| = 0 \,;$$

27

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

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

$\Rightarrow (14-\lambda)(23-\lambda) - (-11 \times -11) = 0$

$\Rightarrow \lambda^2 - 37\lambda + 201 = 0.$

Now we find roots ∴ roots are.

$\lambda = \underline{30.3849,}\ 6.6151$

$\boxed{\lambda_1 > \lambda_2}$

$\lambda_1 = 30.3849$

$\lambda_2 = 6.6151$

Roots of Quadratic eqn

$\boxed{\frac{1}{2a} \sqrt{b^2 - 4ac}}$

$\frac{1}{2(1)} \sqrt{(-37)^2 - 4(1)(201)}$

$\Rightarrow \lambda.$

Here

$a = 1$

$b = -37$

$c = 201$

The Largest eigen value by Principle.

∴ $\boxed{\lambda_1 = 30.3849}$ is Largest eigen value.

(ii) Eigen vector of λ_1

$\boxed{(S - \lambda_1 I)\, U_1 = 0}$

~~Hence~~ Where.

S = Co-Var matrix

I = Identity matrix

λ_1 = eigen value Largest

U_1 = Eigen vector of λ_1

$\boxed{U_1 = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}_{2\times1} \text{ matrix.}}$

$$\begin{bmatrix} 14-\lambda_1 & -11 \\ -11 & 23-\lambda_1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} (14-\lambda_1)u_1 - 11u_2 \\ -11u_1 + (23-\lambda_1)u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\Rightarrow$

$$\begin{bmatrix} (14-\lambda_1)u_1 - 11u_2 \\ -11u_1 + (23-\lambda_1)u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\Rightarrow (14-\lambda_1)u_1 - 11u_2 = 0$

$\cancel{-11u_1 + (23-\lambda_1)}$

$-11u_1 + (23-\lambda_1)u_2 = 0.$

now we have to find value of u_1, u_2 from these Linear eqn.

$$\frac{u_1}{11} = \frac{u_2}{14-\lambda_1} = t \text{ (say)}$$

when $\boxed{t=1} \Rightarrow u_1 = 11$
$u_2 = 14-\lambda_1$

$\therefore$ Eigen vector U_1 of $\lambda_1 = \begin{bmatrix} 11 \\ 14-\lambda_1 \end{bmatrix}$

$$= \begin{bmatrix} 11 \\ 14-30.3849 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.3849 \end{bmatrix}$$ eigen vector

(iii) Normalize the eigen vector U_1

$$e_1 = \begin{bmatrix} 11 \\ -16.3849 \end{bmatrix} = \begin{bmatrix} 11 / \sqrt{11^2 + (-16.3849)^2} \\ \dfrac{-16.3849}{\sqrt{11^2 + (-16.3849)^2}} \end{bmatrix}$$ (Length)

$$e_1 = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$ Normalized eigen vector

28

For λ_2

$$e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$$

Step-5 :- Derive New data set.

1st Principal Component,

	EX-1	EX-2	EX-3	EX-4
First Principal Component PC1	P_{11} = ?	P_{12} = ?	P_{13} = ?	P_{14} = ?

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

$$= -4.3052 \quad 1\times1 \text{ matrix}$$

$$P_{12} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 8-8 \\ 4-8.5 \end{bmatrix} = 3.7361$$

Ily.

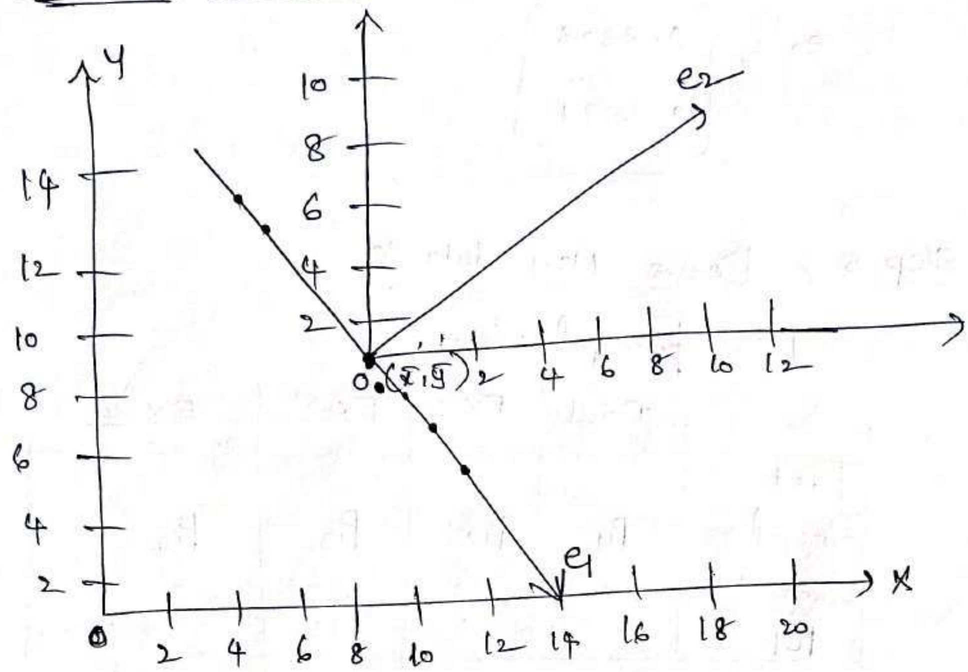
$P_{13} = 5.6928$

$P_{14} = -5.1238$

	EX-1	EX-2	EX-3	EX-4
PC1	P_{11} -4.3052	P_{12} 3.7361	P_{13} 5.6928	P_{14} -5.1238

New data set with reduced dimension !

Co-ordinate System for Principal Components.



Mean Value of $x = 8$

" " " $y = 8.5$

$$e_1 = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$

$$e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$$

Dimensionality Reduction

- Dimension of an Instance ? (or) Length of Instance
  → Number of variables of instance.

Dimensionality reduction :-

Dimensionality reduction is the Process of reducing the number of Variables under Consideration by Obtaining a Smaller set of Principal Variables.

Advantage of reducing dimension :-

→ Decreases the Complexity of the algorithm
→ Saves the Cost of extracting an unnecessary input
→ Simple models Can be choosed
→ Simplifying the Knowledge extraction
→ Easy to plot and analyse

Dimensionality Reduction

Feature Selection

Find 'K' of d dimensions and discard (d−k) dimensions

→ Subset Selection.

Feature extraction

1) Find a new K dimensions that are Combinations of 'd' dimensions
2) PCA (Principal Component Analysis
3) LDA (Linear Discriminant Analysis)

Subset Selection :-

→ Also known as Variable Selection, attribute selection, feature selection

→ {A, B, C}

{A} {B} {C} {AB} {BC} {AC} {ABC} {∅}

→ Advantages :-

→ Simplification of Models
→ Shorter training times
→ Enhanced generalization
→ To avoid the Curse of dimensionality.

Subset Selection

Forward Selection

→ Start with no Variables and add them one by one. at each step adding the One that decrease the error the most until any further addition does not decrease the error

Backward Selection.

Start with all Variables and remove them one by one at each step till the error become minimum

PCA (Principal Component Analysis) (30)

(used in Machine Learning)

1) Find the PCA

x	2.5	0.5	2.2	1.9	3.1	2.3	2.0	1.0	1.5	1.1
y	2.4	0.7	2.9	2.2	3.0	2.7	1.6	1.1	1.6	0.9

$\boxed{n=10}$ x & y are 2 Variables

$$\bar{x} = \frac{\Sigma x}{n_1} = \underline{\underline{1.81}} \qquad \bar{y} = \frac{\Sigma y}{n_2} = \underline{\underline{1.91}}$$

$$\therefore \text{Mean} \Rightarrow \boxed{\begin{matrix}\bar{x} = 1.81 \\ \bar{y} = 1.91\end{matrix}}$$

$$\text{Co-variance Matrix} = \begin{bmatrix} Cov(x,x) & Cov(x,y) \\ Cov(x,y) & Cov(y,y) \end{bmatrix}$$

$$Cov(x,x) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

(A) $x_i - \bar{x}$	(B) $y_i - \bar{y}$	AB	A^2	B^2
0.69	0.49	0.3381		
-1.31	-1.21	1.5851		
0.39	0.99	0.3861		
0.09	0.29	0.0261		
1.29	1.09	1.4061		
0.49	0.79	0.3871		
0.19	-0.31	-0.0589		
-0.81	-0.81	0.6561		
-0.31	-0.31	0.0961		
-0.71	-1.01	0.7171		
		$\boxed{5.539}$		

(A) $x_i - \bar{x}$	(B) $y_i - \bar{y}$	AB	A^2	B^2
0.69	0.49	0.3381	0.4761	0.2401
−1.31	−1.21	1.5851	1.7161	1.4641
0.39	0.99	0.3861	0.1521	0.9801
0.09	0.29	0.0261	0.0081	0.0841
1.29	1.09	1.4061	1.6641	1.1881
0.49	0.79	0.3871	0.2401	0.6241
0.19	−0.31	−0.0589	0.0361	0.0961
−0.81	−0.81	0.6561	0.6561	0.6561
−0.31	−0.31	0.0961	0.0961	0.0961
−0.71	−1.01	0.7171	0.5041	1.0201
		5.539	5.549	6.449

$$\text{Co-variance Matrix (A)} = \frac{\begin{bmatrix} 5.549 & 5.539 \\ 5.539 & 6.449 \end{bmatrix}}{n-1}$$

$10-1=9$

$$= \begin{bmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166 \end{bmatrix}$$

31

Eigen values $|A - \lambda I| = 0$.

$$\Rightarrow (0.6166 - \lambda)(0.7166 - \lambda) - (0.6154)^2 = 0$$

$$\Rightarrow 0.4418 - 0.6166\lambda - 0.7166\lambda + \lambda^2 - 0.3787 = 0$$

$$\lambda^2 - 1.3332\lambda + 0.0631 = 0$$

$ax^2 + bx + c$.

$$\boxed{\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}}$$

$\boxed{\begin{array}{l}\lambda_1 = 1.2841 \\ \lambda_2 = 0.0491\end{array}}$ are eigen values

$$\begin{bmatrix} 0.6166-\lambda & 0.6154 \\ 0.6154 & 0.7166-\lambda \end{bmatrix}$$

Now we substitute $\lambda_1\ \lambda_2$ to find eigen vectors.

(In place of $\lambda \rightarrow \lambda_1$ value
$\lambda \rightarrow \lambda_2$ value.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -0.6634 & 0.6154 \\ 0.6154 & -0.5631 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

(λ_1 substituted)

$$\left.\begin{array}{l} -0.6634\, a_{11} + 0.6154\, a_{12} = 0 \\ 0.6154\, a_{11} - 0.5631\, a_{12} = 0. \end{array}\right\} \text{Linear eqns}$$

$Z_1 = a_{11}x_1 + a_{12}x_2$
$Z_2 = a_{21}x_1 + a_{22}x_2$

$$\left.\begin{array}{l} 0.6634 a_{11} = 0.6154 a_{12} \\ 0.6154 a_{11} = 0.5631 a_{12} \end{array}\right\} \text{add eqns both}$$

$$a_{11} = 1.1785\, a_{12}$$

$$a_{11} = \frac{1.5785}{1.2788} a_{12}$$

$$a_{11} = 0.9215 a_{12}$$

$$\begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} \Rightarrow \begin{bmatrix} 0.9215 \\ 1 \end{bmatrix} \begin{pmatrix} \text{eigen} \\ \text{vector} \end{pmatrix} \Rightarrow \sqrt{(0.9215)^2 + 1^2}$$

$$= \sqrt{0.8491 + 1}$$

$$\Rightarrow \begin{bmatrix} \frac{0.9215}{1.3598} \\ \frac{1}{1.3598} \end{bmatrix} \qquad = \sqrt{1.8491}$$

$$= \underline{\underline{1.3598}}$$

$$\text{Eigen vectors} = \begin{bmatrix} 0.6677 \\ 0.7354 \end{bmatrix} \qquad \sqrt{(a_{11})^2 + (a_{12})^2} = \underline{\underline{1.3598}}$$

$$\begin{bmatrix} 0.5675 & 0.6154 \\ 0.6154 & 0.6675 \end{bmatrix} \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$(\lambda_2 \text{ substitute})$

$$0.5675\, a_{21} = -0.6154\, a_{22}$$

$$0.6154\, a_{21} = -0.6675\, a_{22} \quad (\text{add})$$

$$1.1829\, a_{21} = -1.2829\, a_{22}$$

$$a_{21} = \frac{-1.2829}{1.1829} a_{22} = \underline{\underline{-1.0845\, a_{22}}}$$

$$\begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} = \begin{bmatrix} -1.0845 \\ 1 \end{bmatrix} \Rightarrow \sqrt{(1.0845)^2 + 1^2}$$

$$= \sqrt{1.1761 + 1}$$

$$= \sqrt{2.1761} = \underline{\underline{1.4751}}$$

32

- $\begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{-1.0845}{1.4751} \\ \frac{1}{1.4751} \end{bmatrix}$

Eigen vectors = $\begin{bmatrix} -0.7352 \\ 0.6779 \end{bmatrix}$

Total variance % = Eigen vector 1   Eigen vector 2

$\left(\frac{\lambda_1}{\lambda_1+\lambda_2} \quad \frac{\lambda_2}{\lambda_1+\lambda_2} \right) \Rightarrow \frac{1.2841}{1.2841+0.0491} \qquad \frac{0.0491}{1.2841+0.0491}$

$= \frac{1.2841}{1.3332} \qquad = \frac{0.0491}{1.3332}$

$= 0.96 \qquad = 0.036$

(96%) ✓   (= 3.6%) ✗

High value   Low value.

of   of

Eigen vector 1   Eigen vect 2

(reduce the data value available)

(33)

Principal Component Analysis

(PCA)

1) Given two attributes X & Y with values given in the table below

X	2.5	0.5	2.2	1.9	3.1	2.3	2	1	1.5	1.1
Y	2.4	0.7	2.9	2.2	3.0	2.7	1.6	1.1	1.6	0.9

Find the eigen vector & Principal Component from the given data.

Sol:- · PCA Algorithm

(1) Get data
(2) Subtract the mean (Subtract mean from data)
(3) Calculate the Covariance matrix
(4) Calculate the eigen vector & eigen values of the Covariance matrix
(5) Choosing Components & forming a feature vector
(6) Deriving the new data set, this is final step in PCA.

Sol :- (1) Get data (given data)

(2) Mean :-

$$\bar{X} = \frac{2.5 + 0.5 + 2.2 + 1.9 + 3.1 + 2.3 + 2 + 1 + 1.5 + 1.1}{10}$$

$$\boxed{\bar{X} = 1.81}$$

$$\bar{Y} = \frac{2.4 + 0.7 + 2.9 + 2.2 + 3.0 + 2.7 + 1.6 + 1.1 + 1.6 + 0.9}{10}$$

$$\boxed{\bar{Y} = 1.91}$$

(3) Covariance Matrix : $\boxed{n = 10}$

no of terms .

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

$$\boxed{Cov(x,y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}}$$

$$\boxed{Cov(x,x) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(x_i - \bar{x})}{n-1}}$$

34

$$\boxed{Cov(x,x) = \sum_{i=1}^{n} \frac{(x_i-\bar{x})(x_i-\bar{x})}{n-1}}$$

$\boxed{\bar{x} = 1.81}$

x	$(x_i-\bar{x})$	$(x_i-\bar{x})(x_i-\bar{x})$
2.5	0.69	0.4761
0.5	−1.31	1.7161
2.2	0.39	0.1521
1.9	0.09	0.0081
3.1	1.29	1.6641
2.3	0.49	0.2401
2	0.19	0.0361
1	−0.81	0.6561
1.5	−0.31	0.0961
1.1	−0.71	0.5041
		$\sum(x_i-\bar{x})(x_i-\bar{x}) = 5.549 = \text{Sum}$

$$Cov(x\,x) = \frac{5.549}{n-1} \qquad \boxed{n=10}$$

$$= \frac{5.549}{9}$$

$$= 0.6165$$

$$\boxed{Cov(x,x) = 0.6165}$$

||ly. $Cov(x,y)$

$Cov(y,x)$

$Cov(y,y)$

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

$$C = \begin{bmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{bmatrix}$$

(4) Eigen value & Eigen vector :- (Co-variance matrix)

$$\boxed{C - \lambda I = 0}$$

$$\begin{bmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$$

$$\begin{bmatrix} 0.6165 - \lambda & 0.6154 \\ 0.6154 & 0.7165 - \lambda \end{bmatrix} = 0$$

$|C - \lambda I| = 0$

$$\Rightarrow (0.6165)[0.7165 - \lambda] - \lambda[0.7165 - \lambda] - 0.6154 \times 0.6154 = 0$$

$$\Rightarrow 0.6165 \times 0.7165 - 0.6165\lambda - 0.7165\lambda + \lambda^2 - (0.6154)^2 = 0$$

$$\Rightarrow \lambda^2 - 1.333\lambda + 0.4417 - 0.3737 = 0$$

$$\Rightarrow \lambda^2 - 1.333\lambda + 0.1063 = 0.$$

$\therefore$ $\boxed{\lambda_1 = 0.0490}$, $\boxed{\lambda_2 = 1.2840}$ } are Eigen values

Eigen Vector :- For $\boxed{\lambda_1 = 0.0490}$

$$\begin{bmatrix} 0.6165 - 0.0490 & 0.6154 \\ 0.6154 & 0.7165 - 0.0490 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} 0.5674 & 0.6154 \\ 0.6154 & 0.6674 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

35

$$\begin{bmatrix} 0.5674 & 0.6154 \\ 0.6154 & 0.6674 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\implies 0.5674x_1 + 0.6154y_1 = 0 \longrightarrow (1)$$

$$0.6154x_1 + 0.6674y_1 = 0 \longrightarrow (2)$$

$$\implies \boxed{x^2 + y^2 = 1}$$ on solving (1) & (2)

$$y_1 = -\frac{0.5674}{0.6154} x_1$$

$$x_1^2 + \left(-\frac{0.5674}{0.6154} x_1\right)^2 = 1$$

$$\boxed{x_1 = 0.7351}$$

$$\boxed{y_1 = -0.6778}$$

For $\boxed{\lambda_2 = 1.2840}$

$$\begin{bmatrix} -0.6675 & 0.6154 \\ 0.6154 & -0.5675 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$-0.6675x_2 + 0.6154y_2 = 0$$

$$0.6154x_2 - 0.5675y_2 = 0.$$

$$y_2 = \frac{0.6675}{0.6154} x_2$$

now using $\boxed{x^2 + y^2 = 1}$ $\boxed{x_2 = 0.6773}$

$\cancel{x_2 = 0.66}$ $\boxed{y_2 = 0.7351}$

$$\begin{array}{cc} x_1 & y_1 \\ \left[\begin{array}{cc} 0.7351 & -0.6778 \\ 0.6778 & 0.7351 \end{array}\right] \\ x_2 & y_2 \end{array}$$

(5)

$\lambda_1 = 0.0490$

$\lambda_2 = 1.2840$ $\quad \lambda_2 > \lambda_1$ Clearly.

$$\begin{array}{c} \lambda_1 \\ \lambda_2 \end{array} \begin{bmatrix} 0.7351 & -0.6778 \\ \boxed{0.6778 \quad 0.7351} \end{bmatrix}$$

eigen vector

PCA

PCA :-

⟶ It is a way of identifying patterns in data and expressing the data in such a way to highlight their similarities & differences

⟹ Dimensionality Reduction.

36

- **Principal Component Analysis (PCA)**

is a statistical Procedure that is used to reduce the dimensionality.

It uses an Orthogonal transformation to Convert a Set of Observations of Possibly Correlated Variables into a set of Values of Linearly UnCorrelated Variables Called Principal Components. It is often used as a dimensionality reduction technique.

Steps involved in the PCA :-

Step-1 :- Standardize the data set.

Step-2 :- Calculate the Covariance matrix for the features in the dataset.

Step-3 :- Calculate the eigen Values and eigen vectors for the Covariance matrix.

Step-4 :- Sort eigen values and their Corresponding eigen vectors.

Step-5 :- pick K eigen values and form a matrix of eigen vectors

Step-6 :- Transform the Original matrix.

(37)

Advantages of Dimensionality Reduction :-

(1) Dimensionality reduction helps in data compression, and hence reduced the storage space.

(2). It reduces Computation time.

(3) It also helps remove redundant features If any.

(4) Dimensionality Reduction helps in data Compressing and reducing the Storage space required.

(5) It fastens the time required for Performing same Computations.

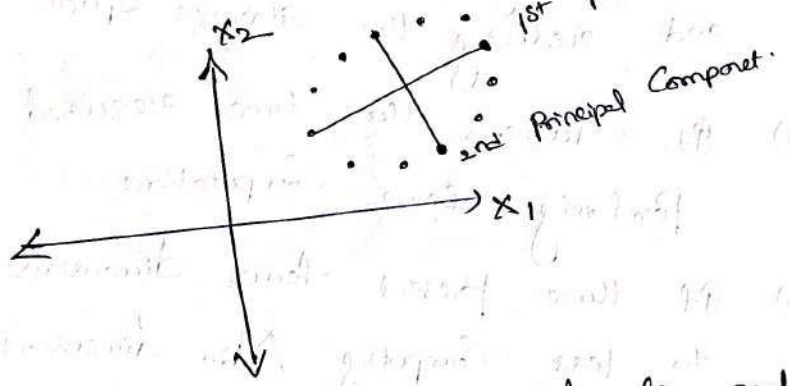
(6) If there present fewer dimensions then it leads to less Computing. Also dimensions Can allow usage of algorithms unfit for a large number of dimensions.

(7) It takes Care of multiCollinearity that improves the model Performance. It removes redundant features.

for example, there is no point in storing a value in two different units (meters & inches)

(8) Reducing the dimensions of data to 2D or 3D may allow us to plot and visualize it precisely.
You can then observe patterns more clearly.

Below you can see that, how a 3D data is converted into 2D. First it has identified the 2D plane then represented the points on these two new axes Z_1 and Z_2.



It is helpful in noise removal also and as a result of that, we can improve the performance of models.

38

Dis-advantages of Dimensionality Reduction :-

(1) Basically, it may lead to some amount of data loss.

(2) Although, PCA tends to find linear correlations between variables, which is sometimes undesirable

(3) Also, PCA fails in cases where mean and covariance are not enough to define datasets.

(4) Further, we may not know how many principal components to keep - in practice some thumb rules are applied.

Importance of Dimensionality Reduction :-

1) why is Dimension Reduction is important in machine learning Predictive modeling ?

A) The Problem of unwanted increase in dimension is closely related to other.
That was to fixation of measuring / recording data at a far granular level then it was done in past.
This is no way suggesting that this is a ~~~~ recent Problem..
It has started gaining more importance lately due to a surge in data.

- **Dis-advantages of Dimensionality Reduction :-**

1) It may lead to some amount of data loss
2) PCA tends to find linear Correlations between Variables, which is sometimes undesirable.
3) PCA fails in Cases where mean and Covariance are not enough to define datasets.

Advantages of Dimensionality Reduction :-

1) It helps in data Compression and hence reduced storage space.
2) It reduces Computation time.
3) It also helps remove redundant features, If any

Machine Learning :- Machine Learning is nothing but a field of study which allows Computers to "Learn" like humans without any need of explicit Programming.

What is Predictive Modeling :-

Predictive modeling is a Probabilistic Process that allows us to forecast outcomes, on the basis of some Predictors.

These Predictors are basically features that come into play when deciding the final result of the Model.

ie :- the outcome

What is Dimensionality Reduction ?

In machine learning classification problems, there are often too many factors on the basis of which the final classification is done.

These factors are basically variables called features.

The higher the number of features, the harder it gets to visualize the training set and then work on it.

Sometimes most of these features are correlated and hence redundant. That is where dimensionality reduction algorithms come into play.

(40)

Dimensionality reduction is the Process of reducing the number of random variables under Consideration, by Obtaining a set of Principal Variables.

It Can be divided into feature selection and feature extraction.

Components of Dimensionality Reduction :-

There are two Components of dimensionality reduction

1) **Feature Selection** :- In this, we try to find a subset of the Original set of Variables, or features, to get a smaller subset which can be used to model the Problem

It usually involves Three ways.

1) Filter
2) wrapper
3) Embedded.

2) **Feature Extraction** :- This reduces the data in a high dimensional space to a lower dimension space

ie:- a Space with lesser no of dimensions.

Methods of Dimensionality Reduction :

The various methods used for dimensionality reduction include

(1) Principal Component Analysis (PCA)

2) Linear Discriminant Analysis (LDA)

3) Generalized Discriminant Analysis (GDA)

Dimensionality reduction may be both linear or non-Linear, depending upon the method used.

41

- → **Principal Component Analysis (or) PCA**

  is a dimensionality reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

→ Based on the dataset find a new set of orthogonal feature vectors in such a way that the data spread is maximum in the direction of the feature vector (or) dimension.

(42)

Covariance Formula :-

Covariance formula is a statistical formula which is used to assess the relationship between two variables.

In simple words, Covariance is one of the statistical measurement to know the relationship of the variance between the two variables.

The Covariance indicates how two variables are related and also helps to know whether the two variables vary together or change together.

The Covariance is denoted by Cov(X,Y) and the formula of Covariance are given below

Population Covariance formula:

$$\boxed{Cov(X,Y) = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{N}}$$

Sample Covariance formula :-

$$\boxed{Cov(x,y) = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{N-1}}$$

These are the formulas to find sample and population covariance.

Notations in Covariance formulas :

X_i = data value of X

Y_i = data value of Y

$\bar{X}$ = mean of X

$\bar{Y}$ = mean of Y

N = Number of data values.

43

Co-variance

$$\boxed{Cov(x,y) = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{N}}$$

1)

x_i	10	12	14	8
y_i	40	48	56	32

$\bar{x} = \frac{10+12+14+8}{4} = \frac{44}{4} = 11$ Mean of $x_i = 11$

$\therefore \bar{x} = 11$.

$\bar{y} = \frac{40+48+56+32}{4} = \frac{176}{4} = 44$ Mean of $y_i = 44$

x_i	$x_i - \bar{x}$ ($\bar{x}=11$)	y_i	$y_i - \bar{y}$ ($\bar{y}=44$)
10	-1	40	-4
12	1	48	4
14	3	56	12
8	-3	32	-12

$$Cov(xy) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{N}$$

$$= \frac{(-1)(-4) + (1)(4) + (3)(12) + (-3)(-12)}{4}$$

$$= \frac{(4) + (4) + (36) + (36)}{4}$$

$$= \frac{80}{4} = 20$$

$\therefore Cov(xy) = 20$

Co-Variance :-

If X & Y are two random variables then Covariance between them is defined as

$$Cov(XY) = \sigma_{XY} = E(XY) - E(X)E(Y)$$

Proof :- $Cov(XY) = E[(X - E(X))(Y - E(Y))]$

$$= E[XY - XE(Y) - YE(X) + E(X)E(Y)]$$

$$= E(XY) - E(X)E(Y) - E(Y)\cancel{E(X)} + E(X)\cancel{E(Y)}$$

$$= \underline{E(XY) - E(X)E(Y)}$$

$$\left(\begin{array}{l} E(K) = K \\ E[E(X)] = E(X) \end{array} \right)$$

Properties of Covariance :-

(1) If X & Y are Independent, then

$E(XY) = E(X)E(Y)$ and hence

$$Cov(XY) = E(X)E(Y) - E(X)E(Y) = 0$$

2) $Cov(X,X) = Var(X)$

3) $Cov(aX, bY) = ab\, Cov(XY)$

4) $Cov(X+a, Y+b) = Cov(X, Y)$

(44)

- (5) $\mathrm{Cov}\left(\frac{x-\bar{x}}{\sigma_x}, \frac{y-\bar{y}}{\sigma_y}\right)$

$$= \frac{1}{\sigma_x \sigma_y} \mathrm{Cov}(x, y)$$

6) $\mathrm{Cov}(x+y, z) = \mathrm{Cov}(x, z) + \mathrm{Cov}(y, z)$

Covariance of (x, y)

In Bi-variate distribution of (x_i, y_i) take values

$(x_1, y_1), (x_2, y_2) \cdots (x_n, y_n)$

$$Cov(x, y) = \frac{1}{n} \sum (x - \bar{x})(y - \bar{y})$$

$$= \frac{1}{n} \sum [xy - x\bar{y} - \bar{x}y + \bar{x}\bar{y}]$$

$$= \frac{1}{n} \sum xy - \bar{y}\frac{\sum x}{n} - \bar{x}\frac{\sum y}{n} + \bar{x}\bar{y}$$

$$= \frac{1}{n} \sum xy - 2\bar{x}\bar{y} + \bar{x}\bar{y}$$

$$= \frac{1}{n} \sum xy - \bar{x}\bar{y}$$

$$\therefore \boxed{Cov(x, y) = \frac{1}{n} \sum xy - \bar{x}\bar{y}}$$

45

(2)

x	1	2	3	4	5
y	2	3	4	6	10

$$\boxed{Cov(xy) = \frac{1}{n}\Sigma xy - \bar{x}\bar{y}}$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = \underline{\underline{3}}$$

$$\therefore \boxed{\bar{x} = 3}$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{2+3+4+6+10}{5} = \frac{25}{5} = \underline{\underline{5}}$$

$$\therefore \boxed{\bar{y} = 5}$$

x	y	xy
1	2	2
2	3	6
3	4	12
4	6	24
5	10	50
		$\Sigma xy = 94$

$$Cov(xy) = \frac{1}{n}\Sigma xy - \bar{x}\bar{y}$$

$$= \frac{1}{5} \times (94) - (3)(5)$$

$$= 18.8 - 15$$

$$= \underline{\underline{3.8}}$$

Problem of Covariance formula :-

1) The table below describes the rate of economic growth (x_i) and the rate of return on the S&P 500 (y_i)

Using the Covariance formula, determine whether economic growth & S&P 500 returns have a positive or inverse relationship

Before you compute the Covariance, Calculate the Mean of X & Y.

Economic growth % (x_i)	2.1	2.5	4.0	3.6
S&P 500 Returns % (y_i)	8	12	14	10

Sol :- given x_i = 2.1, 2.5, 4.0 & 3.6 (economic growth)

y_i = 8, 12, 14, 10 (S&P 500 returns)

Find $\bar{x}$ & $\bar{y}$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{2.1 + 2.5 + 4.0 + 3.6}{4} = \frac{12.2}{4} = 3.1$$

$$\therefore \boxed{\bar{x} = 3.1}$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{8 + 12 + 14 + 10}{4} = \frac{44}{4} = 11$$

$$\therefore \boxed{\bar{y} = 11}$$

(46)

Now substitute these values into the Covariance formula to determine the relationship between economic growth & S&P 500 returns.

$\overline{X} = 3.1$ $\overline{Y} = 11$

x_i	y_i	$x_i - \overline{x}$	$y_i - \overline{y}$
2.1	8	−1	−3
2.5	12	−0.6	1
4.0	14	0.9	3
3.6	10	0.5	−1

$$\therefore \; Cov(x,y) = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{N}$$

$$= \frac{(-1)(-3) + (-0.6)(1) + (0.9)(3) + (0.5)(-1)}{4}$$

$$= \frac{4.6}{4} = \underline{1.15}.$$

(47)

1) Co-Variance

What is Covariance in relation to Variance & Correlation

Two Data sets   5 elements data set

$X = (2, 4, 6, 8, 10)$

$Y = (1, 3, 8, 11, 12)$

Variance = S^2 = A measure of how spread out the numbers of a data set are

X Average $(\bar{X}) = \frac{\sum x_i}{n} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6$

Y Average $(\bar{Y}) = \frac{\sum y_i}{n} = \frac{1+3+8+11+12}{5} = \frac{35}{5} = 7$

(X) Variance $(S_x^2) = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{(2-6)^2 + (4-6)^2 + \cdots + (10-6)^2}{5}$

$= \frac{16+4+0+4+16}{5} = \frac{40}{5} = 8.$

(Y) Variance $(S_y^2) = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{(1-7)^2 + (3-7)^2 + \cdots + (12-7)^2}{5}$

$= \frac{36+16+1+16+25}{5} = \frac{94}{5} = 18.8$

Co-Variance :- Cov(X Y) = A measure of how the trends of 2 data sets are related

$Cov(X Y) = \frac{\sum (x-\bar{x})(y-\bar{y})}{n} = \frac{(-4)(-6) + (-2)(-4) + (0)(0) + (2)(4) + (4)(5)}{5}$

$= \frac{24+8+0+8+20}{5} = \frac{60}{5} = 12.$

Correlation :- (r) = A measure of how the trends of 2 data sets are related $-1 \leq r \leq 1$

$$r = \frac{Cov(X,Y)}{S_X S_Y} = \frac{12}{\sqrt{8} \cdot \sqrt{18.8}} = \underline{\underline{0.98}}$$

$$r = \frac{Cov(X,Y)}{\sqrt{S_X^2} \cdot \sqrt{S_Y^2}} = \underline{\underline{0.98}}$$ (Strong relationship between 2 sets)

2) How to Calculate the Variance :-

Data set $Z = (-4, -1, 5, 12, 18)$ (The whole Population)

Step-1 : Find the Mean (Average)

$$\bar{Z} = \frac{\sum_i Z_i}{n} = \frac{(-4)+(-1)+(5)+(12)+(18)}{5}$$

$$\bar{Z} = \frac{30}{5} = \underline{\underline{6}}$$

Step-2 : Find the Variance :

$$S_Z^2 = \frac{\sum_i (Z_i - \bar{Z})^2}{n} = \frac{(-4-6)^2 + (-1-6)^2 + (5-6)^2 + (12-6)^2 + (18-6)^2}{5}$$

$\therefore S_Z^2 = 66$

$$= \frac{(-10)^2 + (-7)^2 + (-1)^2 + (6)^2 + (12)^2}{5}$$

$$= \frac{100 + 49 + 1 + 36 + 144}{5} = \frac{330}{5} = \underline{\underline{66}}$$

(48)

3) <u>Population</u> <u>vs</u> <u>Sample</u> <u>Variance</u> :-

$X = (2, 4, 6, 8, 10)$

$\bar{X} = \frac{\sum_i x_i}{n} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = \underline{6}$.

$\therefore \underline{\bar{X} = 6}$.

<u>Sample</u> <u>Variance</u> :-

$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1} = \frac{(2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2}{5-1}$

$= \frac{16+4+0+4+16}{4} = \frac{40}{4} = \underline{10}$.

$\therefore \underline{s^2 = 10}$.

$\sigma^2 = \frac{\sum_i (x_i - \bar{x})^2}{n} = \frac{(2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2}{5}$

$= \frac{16+4+0+4+16}{5} = \frac{40}{5} = \underline{8}$.

4) How to Calculate the Co-Variance :-

we have 2 data sets.

$X = (2, 4, 6, 8, 10)$
$Y = (12, 11, 8, 3, 1)$

Step-1 :- Find the Mean (average) of both sets.

$$\bar{X} = \frac{\sum_i X_i}{n} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{12+11+8+3+1}{5} = \frac{35}{5} = 7$$

Step-2 :- Find the Variance of both sets.

$$\sigma_X^2 = \frac{\sum_i (X_i - \bar{X})^2}{n} = \frac{(2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2}{5}$$

$$= \frac{4^2 + 2^2 + 0^2 + 2^2 + 4^2}{5}$$

$$= \frac{16+4+0+4+16}{5} = \frac{40}{5} = 8.$$

$$\sigma_Y^2 = \frac{\sum_i (Y_i - \bar{Y})^2}{n} = \frac{(12-7)^2 + (11-7)^2 + (8-7)^2 + (3-7)^2 + (1-7)^2}{5}$$

$$= \frac{5^2 + 4^2 + 1^2 + 4^2 + 6^2}{5}$$

$$= \frac{25+16+1+16+36}{5} = \frac{94}{5} = 18.8.$$

49

Step-3 :- Find the Covariance :-

$$Cov(x\,y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$= \frac{(-4)(5) + (-2)(4) + (0)(1) + (2)(-4) + (4)(-6)}{5}$$

$$= \frac{-20 - 8 + 0 - 8 - 24}{5} = \frac{-60}{5} = \underline{\underline{-12}}.$$

(5) Covariance :- What is the Covariance Matrix

The Covariance Matrix is an $n \times n$ matrix (where n = no of data sets) such that the diagonal elements represent the variances of each data set and the off-diagonal elements represent the covariance between the data sets.

1) $X = 2, 4, 6, 8, 10$ $\quad \bar{X} = 6 \quad Var(X) = \sigma_X^2 = 8$

$Y = 1, 3, 8, 11, 12$ $\quad \bar{Y} = 7 \quad Var(Y) = \sigma_Y^2 = 18.8$

$Cov(X\,Y) = 12 = Cov(Y, X)$

$$Var(X) = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}$$

$$Cov(X\,Y) = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{N}$$

Covariance Matrix =

$$\begin{array}{c c} & \begin{array}{cc} X & Y \end{array} \\ \begin{array}{c} X \\ Y \end{array} & \begin{bmatrix} Var(X) & Cov(X\,Y) \\ Cov(Y\,X) & Var(Y) \end{bmatrix} \end{array}$$

$$= \begin{bmatrix} 8 & 12 \\ 12 & 18.8 \end{bmatrix} //.$$

2) Example of Co-Variance Matrix :

$X = 2, 4, 6, 8, 10$.

$Y = 7, 3, 5, 1, 9$.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = \underline{\underline{6}}. \quad \therefore \boxed{\bar{x} = 6}$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{7+3+5+1+9}{5} = \frac{25}{5} = \underline{\underline{5}}. \quad \therefore \boxed{\bar{y} = 5}$$

$$\sigma_x^2 = Var(x) = \frac{\sum_i (x_i - \bar{x})^2}{n} = \frac{(2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2}{5}$$

$$\sigma_x^2 = \underline{\underline{8}}.$$

$$\sigma_y^2 = Var(y) = \frac{\sum_i (y_i - \bar{y})^2}{n} = \frac{(7-5)^2 + (3-5)^2 + (5-5)^2 + (1-5)^2 + (9-5)^2}{5}$$

$$\sigma_y^2 = \underline{\underline{8}}.$$

$$Cov(x\,y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$= \frac{(2-6)(7-5) + (4-6)(3-5) + (6-6)(5-5) + (8-6)(1-5) + (10-6)(9-5)}{5}$$

$$= \frac{(-4)(2) + (-2)(-2) + (0)(0) + (2)(-4) + (4)(4)}{5}$$

$$= \frac{4}{5} = \underline{\underline{0.8}}.$$

$\therefore Cov(x\,y) = \underline{\underline{0.8}}$.

50

Co-variance Matrix = $\begin{bmatrix} Var(x) & Cov(x\,y) \\ Cov(y\,x) & Var(y) \end{bmatrix}$

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

$= \begin{bmatrix} 8 & 0.8 \\ 0.8 & 8 \end{bmatrix}$

3) Example of Covariance Matrix :-

$x = 2, 4, 6, 8, 10$

$y = 10, 8, 6, 4, 2$

$\bar{x} = \frac{\sum_i x_i}{n} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6 \Rightarrow \boxed{\bar{x} = 6}$

$\bar{y} = \frac{\sum_i y_i}{n} = \frac{10+8+6+4+2}{5} = \frac{30}{5} = 6 \Rightarrow \boxed{\bar{y} = 6}$

$\sigma_x^2 \, (Var(x)) = \frac{\sum_i (x_i - \bar{x})^2}{n} = \frac{(2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2}{5}$

$= 8$.

$\sigma_y^2 \; Var(y) = \frac{\sum_i (y_i - \bar{y})^2}{n} = \frac{(10-6)^2 + (8-6)^2 + (6-6)^2 + (4-6)^2 + (2-6)^2}{5}$

$= 8$.

$$Cov(xy) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$= \frac{(2-6)(10-6) + (4-6)(8-6) + (6-6)(6-6) + (8-6)(4-6) + (10-6)(2-6)}{5}$$

$$= \frac{(-4)(4) + (-2)(2) + (0)(0) + (2)(-2) + (4)(-4)}{5}$$

$$= \frac{-40}{5} = \underline{\underline{-8}}.$$

$$\text{Covariance Matrix} = \begin{bmatrix} Var(x) & Cov(xy) \\ Cov(yx) & Var(y) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_x & Cov(xy) \\ Cov(yx) & \sigma_y \end{bmatrix}$$

$$= \begin{bmatrix} 8 & -8 \\ -8 & 8 \end{bmatrix}$$

51

8) Covariance :-

What is Correlation Coefficient :-

$x = 2, 4, 6, 8, 10 \qquad \bar{x} = 6$

$y = 1, 3, 8, 11, 12 \qquad \bar{y} = 7.$

$$\text{Var}(x) = \sigma_x^2 = \frac{\sum_i (x_i - \bar{x})^2}{n} = \underline{\underline{8}}.$$

$$\text{Var}(y) = \sigma_y^2 = \frac{\sum_i (y_i - \bar{y})^2}{n} = \frac{94}{5} = \underline{\underline{18.8}}$$

$$\text{Cov}(x\,y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$= \frac{(2-6)(1-7) + (4-6)(3-7) + (6-6)(8-7) + (8-6)(11-7) + (10-6)(12-7)}{5}$$

$$= \frac{60}{5} = \underline{\underline{12}}.$$

$$\text{Cov Matrix} = \begin{bmatrix} 8 & 12 \\ 12 & 18.8 \end{bmatrix}$$

$$\text{Correlation Co-eff } r = \frac{\text{Cov}(x\,y)}{\sqrt{\text{Var}(x)}\,\sqrt{\text{Var}(y)}} = \frac{\text{Cov}(x\,y)}{\sqrt{\sigma_x^2}\cdot\sqrt{\sigma_y^2}}$$

$$= \frac{12}{\sqrt{8}\cdot\sqrt{18.8}} = \underline{\underline{0.98}} \qquad \underline{\underline{(-1 \le r \le 1)}}$$

9) Calculate the Correlation Co-efficient.

$x = 2, 4, 6, 8, 10$

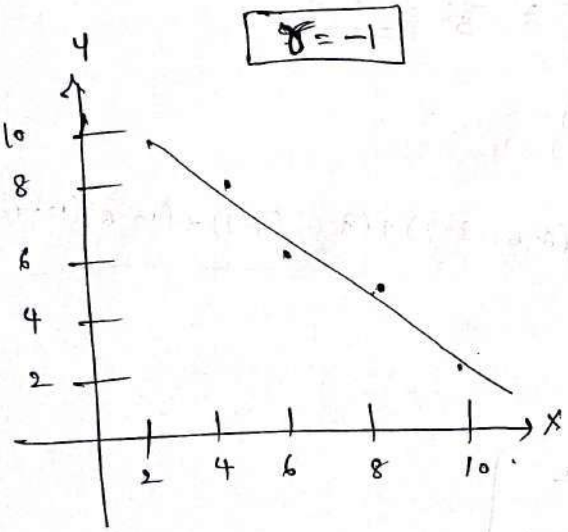
$y = 10, 8, 6, 4, 2$

$\bar{x} = 6 \quad \bar{y} = 6 \quad \sigma_x = 8 \quad \sigma_y = 8$

$Cov(xy) = -8$

$$r = \frac{Cov(x, y)}{\sqrt{\sigma_x^2} \cdot \sqrt{\sigma_y^2}} = \frac{-8}{\sqrt{8} \cdot \sqrt{8}}$$

$$\boxed{r = -1}$$



There is strong Correlation (line)

$x = 2, 4, 6, 8, 10$

$y = 7, 3, 5, 1, 9$

$\bar{x} = 6, \bar{y} = 5, \sigma_x = 8 \quad \sigma_y = 8$

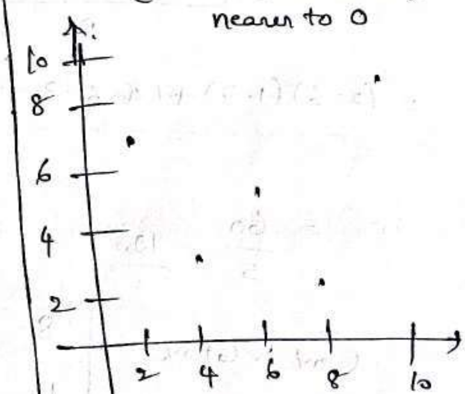
$Cov(xy) = 0.8$

$$r = \frac{Cov(xy)}{\sqrt{\sigma_x^2} \cdot \sqrt{\sigma_y^2}} = \frac{0.8}{\sqrt{8} \cdot \sqrt{8}}$$

$$= \frac{0.8}{8}$$

$$\boxed{r = 0.1}$$

(little Correlation)
nearer to 0



There is no strong Correlation Co-eff
(no line)

(52)

(10) Covariance :- Covariance Matrix with 3 data sets.

$X = 2, 4, 6, 8, 10 \quad \Sigma X = 30 \quad \bar{X} = 6$

$Y = 3, 6, 9, 12, 15 \quad \Sigma Y = 45 \quad \bar{Y} = 9$

$Z = 9, 7, 5, 3, 1 \quad \Sigma Y = 25 \quad \bar{Z} = 5$

$$Var(X) = \frac{\sum_{1} (x_i - \bar{x})^2}{n} = \frac{(2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2}{5}$$

$$= \frac{40}{5} = \underline{\underline{8}}$$

$$Var(Y) = \frac{\sum_{1} (y_i - \bar{y})^2}{n} = \frac{(3-9)^2 + (6-9)^2 + (9-9)^2 + (12-9)^2 + (15-9)^2}{5}$$

$$= \frac{90}{5} = \underline{\underline{18}}$$

$$Var(Z) = \frac{\sum_{1} (z_i - \bar{z})^2}{n} = \frac{(9-5)^2 + (7-5)^2 + (5-5)^2 + (3-5)^2 + (1-5)^2}{5}$$

$$= \frac{40}{5} = \underline{\underline{8}}.$$

$$\text{Cov Matrix} = \begin{array}{c|ccc} & X & Y & Z \\ \hline X & Var(X) & Cov(XY) & Cov(XZ) \\ Y & Cov(YX) & Var(Y) & Cov(Y,Z) \\ Z & Cov(ZX) & Cov(ZY) & Var(Z) \end{array}$$

$$= \begin{bmatrix} 8 & & \\ & 18 & \\ & & 8 \end{bmatrix}$$