UNIT - I

Data Representation

Distance Measures play an Important role in machine Learning they provide the foundation for many popular and effective Machine of Learning Algorithms like k-nearest neighbors for Supervised Learning and k-means clustering for unsupervised Learning.

Different Distance Measures Must be choosen and used Depending on the types of the data. As such, it is important to know how to Implement and calculate a range of different popular distance measures and the Intutions for the resulting Scores.

In this tutorial, you will discover the distance measure in machine learning After Completing this tutorial you will know:

The role and Importance of distance Measures in Machine Learning.

After completing this tutorial you will know:-

The role and Importance of distance Measures in Machine learning algorithms

How to implement and calculate Hamming, Euclidean, and Manhattan distance measures.

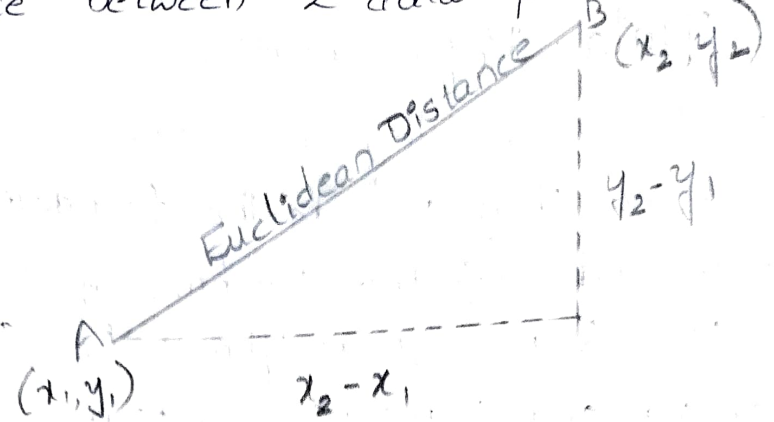
How to Implement and calculate the Minkowski distance that generalizes The Euclidean and Manhattan distance measures

Euclidean Distance is also known as the L_2 norm of a vector (length)

Euclidean Distance:-

Euclidean distance is the straight line

distance between 2 data points in a plane



It is calculated using the Minkowski Distance formula by setting 'p' value to 2, thus, also known as the l_2 norm distance Metric.

The formula is

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

This formula is similar to the pythagorean theorem formula, thus it is also known the pythagorean theorem.

MANHATTAN distance :-

Also called as the city block distance or L_1 norm of a vector; manhattan distance is calculated as the sum of absolute distances between two points

$$\text{Manhattan} = \sum_{i=1}^{n} |x_i - y_i|$$

CHEBYSHEV distance :-

It is calculated as the maximum of the absolute difference between the elements of the vectors

$$\text{chebysav} = \max(|x_i - y_i|)$$

It is also called the maximum value of distance

MINKOWSKI Distance :-

The Minkowski distance is just a generalized form of the above distances.

Minkowski distance is also called as p-norm of the vector.

$$\text{Minkowski} = \sqrt[p]{\sum_{i=1}^{n} |x_i - y_i|^p}$$

MINKOWSKI for Different values of p:-

for, p=1, the distance measure is the manhattan Measure

p=2, the distance measure is the Euclidean measure

p=∞, the distance measure is the chebyshev measure

HAMMING Distance:-

We use hamming distance if we need to deal with categorical attributes.

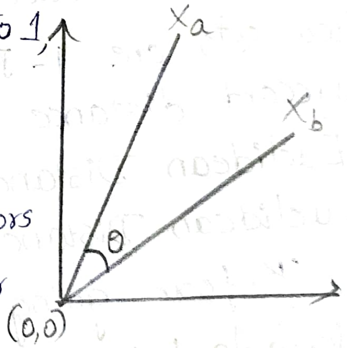
Hamming distance Measures whether the two attributes are different or not when they are equal, the distance is zero (0); otherwise it is 1.

We can use hamming distance only if the strings are of equal length.

for Example, let's take two strings "Hello world" and "Hallo warld"

The Hamming distance between these two strings is 2 as the string differ in the two places.

COSINE Similarity:-



Cosine Similarity ranges from 0 to 1, Where 1 means the two vectors are perfectly similar.

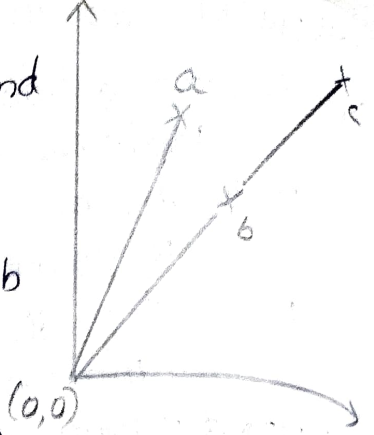
If the angle between the two vectors increases then they are less Similar

$$\text{Similarity}(a,b) = \frac{a.b}{||a|.|b||}$$

Cosine Similarity cares only about the Angle between the two vectors and not distance between them.

Assume there's another vector c in the direction of b.

What do you think the cosine Similarity would be between b and c ?



The cosine Similarity between b and c is 1. Since the angle b/w b and c is 0 and $\cos(0) = 1$.

Even though the distance between b&c is large comparing to a and b cosine Similarity cares about the direction of the vector and not the distance.

JACCARD Similarity and distance:-

In Jaccard Similarity instead of vectors, we will be using sets.

It is used to find the Similarity between two sets. Jaccard Similarity is defined as the ~~instruction~~ intersection of sets divided by their union.

Jaccard Similarity b/w A & B is, where A and B are two sets

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

JACCARD Distance :-

We use Jaccard distance to find how dissimilar two sets are. 1 - Jaccard - Similarity will give you the Jaccard distance.

Euclidean Distance :

Euclidean Distance formula:-

Before going to learn the Euclidean distance formula let us see what is Euclidean distance.

In coordinate geometry, Euclidean.

1) What are the applications of Euclidean distance formula ?

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

finding distance helps in providing the given vertices from a Square, rectangle, etc., (or) providing given vertices from an equilateral triangle, right-angled triangle etc

2) What is the difference between Euclidean distance formula and Manhattan Distance formula for any two points (x_1, y_1) and (x_2, y_2) on a plane

* The Euclidean distance formula says the distance between the above points is

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

* Manhatton destance formula says the distance between the above points is

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

Example - 1 :-

find the distance between points $P(3,2)$ and $Q(4,1)$.

Sol Given $P(3,2) = (x_1, y_1)$

$Q(4,1) = (x_2, y_2)$

Using Euclidean formula of distance

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

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

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

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

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

Example-2:- prove that points $A(0,4)$, $B(6,2)$ and $C(9,1)$ are collinear.

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

we will find the distance between Every pair of points using the Euclidean distance formula

$AB = \sqrt{(6-0)^2+(2-4)^2} = \sqrt{36+4} = \sqrt{40} = 2\sqrt{10}$

$BC = \sqrt{(9-6)^2+(1-2)^2} = \sqrt{9+1} = \sqrt{10}$

$CA = \sqrt{(0-9)^2+(4-1)^2} = \sqrt{81+9} = \sqrt{90} = 3\sqrt{10}$

Here we can see that $AB + BC = CA$

(This is because $2\sqrt{10} + \sqrt{10} = 3\sqrt{10}$

Example-3:-

Check that points $A(\sqrt{3},1)$ $B(0,0)$ and $C(2,0)$ are the vertices of an equilateral Triangle.

Let three vertices A, B and C are the vertices of an equilateral triangle if and only if $AB = BC = CA$.

Given, $A(\sqrt{3},1) = (x_1, y_1)$

$B(0,0) = (x_2, y_2)$

$C(2,0) = (x_3, y_3)$

Using Euclidean Distance formula

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

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

$= \sqrt{3+1}$

$= 2$

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

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

$= \sqrt{4}$

$= 2$

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

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

$= \sqrt{8-4\sqrt{3}}$

Here $AB = BC \neq CA$

A, B, C are not the vertices of Equilateral triangle.

* Euclidean-Distance
* Manhattan Distance
* Minkowski Distance

Distance between objects :-

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

a) compute the Euclidean distance between the two objects.

b) Compute the Manhattan distance between the two objects.

c) Compute the Minikowski distance between the two objects, using $p=3$

Euclidean Distance - Data pre processing

a) Compute the Euclidean distance between the two objects (22,1,42,10) and (20,0,36,8):

$$d(i,j) = \sqrt{(x_{i1}-y_{j1})^2+(x_{i2}-y_{j2})^2+\cdots+(x_{in}-y_{jn})^2}$$

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

$$= 6.71$$

Manhattan Distance - Data pre processing

b) Compute the Manhattan Distance between the two objects (22,1,42,10) and (20,0,36,8):

$$d(i,j) = |x_{i1}-x_{j1}|+|x_{i2}-x_{j2}|+\cdots+|x_{in}-x_{jn}|$$

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

$$= 11$$

Minkowski Distance - Data pre processing

c) Compute the Minkowski Distance between the two objects.

$(22,1,42,10)$ and $(20,0,36,8)$ using $p=3$

$$d(i,j) = \left(|x_{i1}-x_{j1}|^p + |x_{i2}-x_{j2}|^p + \dots + |x_{in}-x_{jn}|^p \right)^{1/p}$$

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

$$= 6.15$$

Hyperplane, Subspace and Half Space

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 dimension less.

So it's going to be 2 dimensions and a 2-dimensional entity in a 3D space would be a Single-dimensional geometric entity, Which would be a Line and so on.

- The hyperplane is usually Describe by an equation as follows

$$X^T n + b = 0 \qquad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad n = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_n \end{bmatrix} \quad [x_1, x_2, \dots x_n]$$

- If we Expand this out for n variables we will get Something like this

$$x_1 n_1 + x_2 n_2 + x_3 n_3 + \dots x_n n_n + b = 0$$

- In Just two dimensions we will get Something like this Which is nothing but an equation of a line

$$x_1 n_1 + x_2 n_2 + b = 0$$

Example :-

Let us consider a 2D Geometry with

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

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

Though it's 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

$y = mx + c$ equation

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

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 4 = 0$$

$$x_1 + 3x_2 + 4 = 0$$

So as you can see from the solution the hyperplane is the equation of a line.

2. Subspace :-

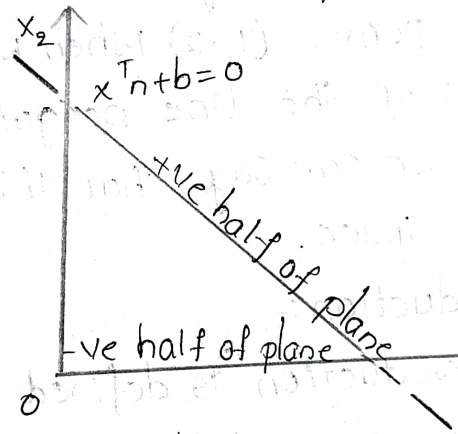
Hyperplanes, in general, are not sub-spaces. However, if we have hyper-plane of the form

$$x^T n = 0 \qquad x^T n = 0$$

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 and 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 $x^T n + b = 0$

So, for this two dimensions, we could write this line as we discussed previously.

let us consider the same Example that we have taken in hyperplane case. So by solving.

We got the equation as

$$x_1 + 3x_2 + 4 = 0$$

There may a rise 3 cases. let's discuss each case with an Example.

Case 1 :- $x_1 + 3x_2 + 4 = 0$; on the line.

Let us consider the two points (-1, -1) when we put this value on the equation of line we got the equation as 0.

So we can say that the point is on the hyper plane of the line.

Case 2 :- $x_1 + 3x_2 + 4 > 0$: positive half-space

Consider two points (1, -1) when we put this value on the equation of the line we got 2 which is greater than 0. So we can say that this point is on positive half space.

Case 3 :- $x_1 + 3x_2 + 4 < 0$: Negative half-space

Consider two points (1, -2) when we put this value on the equation of the line we got -1 which is less than 0. So we can say that this point is on negative half space.

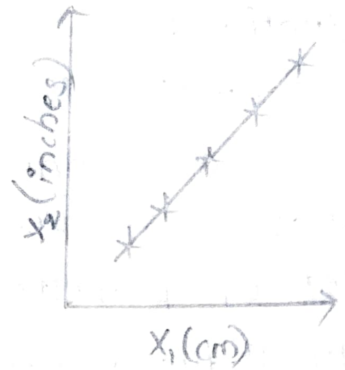
Dimension Reduction :-

Dimension reduction is defined as it is a process of converting a data set having vast dimensions into a data set with lesser Dimensions.

Example :-

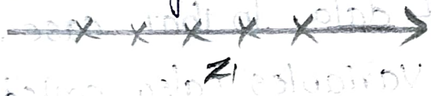
Consider the following example :-

- The following graph shows two dimensions x_1 and x_2
- x_1 Represents measure of several objects in cm.
- x_2 Represents measurement of several objects in inches.



In Machine Learning,

- Use both these dimensions convey similar Information.
- Also, they Introduce a lot of noise in the system.
- So, it is better to use just one dimension using dimension reduction techniques.
- We convert the dimensions of data from 2 dimensions (x_1 and x_2) to 1 dimension (z_1).
- It makes the data relatively easier to explain.



Benefits :-

Dimension Reduction offers several benefits Such as

* It compresses the data and thus reduces the storage space requirements.
* It reduces the time required for the computation Since less dimensions require less computation.
* It eliminates the reductant features
  no longer used.
* It Improves the model performance.

Dimension Reduction Technique :-

The two popular and well known dimension reduction techniques are

Dimension Reduction Techniques

- Principal component Analysis (PCA)
- fisher linear Discriminant Analysis (LDA)

Dimensionality Reduction :-

Dimensionality Reduction is a machine Learning (ML) or statistical technique of reducing the Amount of random variables in a problem by obtaining a set of principal variables

What is Dimensionality Reduction ? overview, objectives and popular techniques :-

What is Dimensionality Reduction?

Before we give a clear definition of dimensionality reduction we first need to understand dimensionality. If you have too many input variables, machine Learning Algorithm performance may degrade.

Suppose you use rows and columns, like those commonly found on a spread sheet, to represent your ML data. In that case, the columns become Input variables (also called features) find to a model predecting the target variable.

Additionally, we can treat the data column as dimensions on the n-dimensional features space, while the data rows are points located on the space. The process is known as interpreting a data set geometrically.

Unfortunately, if many dimensions reside in the feature space, that results in a large volume of data space. consequently only a tiny non representative sample.

This imbalance can negatively affect machine learning algorithm performance. This condition is known as the course of dimensionality". The bottom line, a data set with vast Input features.

Complicates the predicate machine task putting performance and Accuracy at risk.

Here's an Example to help visible visualize the problem. Assume you walked in a straight line for 50 yards, and some were a long that line, you dropped a quarter. you will area covers a square 50 yards by 50 yards. Now your search area covers will take days! But we're not done yet. Now make that search area a cube thats 50 by 50 by 50 yards you may want to say "good bye" to that quarter! the more dimensions Involved. the more complex and the longer it is to search.

How do we lift the curse of dimensionality By reduction reducing the number of dimensions in the feature space. Hence "dimensionality reduction."

To make a long story short, dimensionality reduction means reducing your feature set's dimension.

Why Dimensionality Reduction is Important :-

Dimensionality Reduction brings many advantages to your machine learning data including:

fewer features mean less complexity

you will need less storage space because you have fewer data.

fewer features require less computation time

Model Accuracy improves due to less misleading data. Algorithm train faster thanks to fewerdata.

Reducing The data sets ~~dat~~ features dimensions help visualize the data faster..

It removes noise and redundant features.

Approaches of Dimension Reduction :-

There are two ways to apply The dimensions reduction technique, which are given below.

feature selection:-

feature selection is the process of selecting the subset of the relevant features and leaving out the irrelevant features present in a data set to build a model of high Accuracy. In other words, it is a way of selecting the optimal features from the Input data type set.

Three Methods are used to for the feature selection.

1. filter Methods:-

In these Method, the data set is filtered and a Subset that contains only the relevant features is taken Some Common techniques of filter method are.

- correlation
- chi- square test
- ANova (Analysis of variance)
- Information Gain, etc.,

2. Wrappers Methods:-

The wrapper Method has the same goal as the filter method, but it takes a machine learning model for its evaluation. In this method, Some features are feed to The ML model and evaluate the performance.

The performance decides whether to add these features or remove to increase the accuracy of the model. This method is more Accurate than the filtering method but complex to work.

Some Common techniques to wrapper the methods are

* forward selection
* Backward selection
* Bi-directional Elimination

3. Embedded Methods:-

Embedded Methods check the different training Iterations of the machine learning model and Evaluate the importance of the each feature

Some Common techniques of Embedded Methods are :-

* LASSO
* Elastic Net
* Ridge Regression, etc.,

feature Extraction :-

feature Extraction is the process of transforming the space containing many dimensions into space with fewer dimensions. This approach is useful when we want to keep the whole Information but use fewer resources while processing the Information.

Some Common features Extraction techniques are :-

a. principal component Analysis
b. Linear Discriminant Analysis
c. kernel PCA (Kernel PCA)
d. Quadratic Discriminant Analysis.

4. Dimensionality Reduction techniques:-

Here are some techniques machine Learning professionals use.

1) Principal Component Analysis:-

PCA extends a new set of variables from an Existing more extensive set. The new set is called "principal components"

2) Backward feature Elimination:-

This five-step technique defines the optimal number of features required for a machine Learning Algorithm. By choosing the best model Performance and the maximum tolerable rate Error.

3) forward feature selection:-

This technique follows the Inverse of the backward feature Elimination process. Thus don't Eliminate the feature. Instead, We find the best features that produce The highest Increase in the model's of performance

4) Missing value Ratio:

This technique sets a threshold level of missing values if a variable Exceeds the three hold its dropped

5) Low Variance filter:-

Like The missing value Ratio technique. The low variance filter works with a treshold however in this case, its testing data columns. The method Calculate the variance of each variable. All data columns with variances falling below the treshold are dropped Since low variance features don't effect the target variable.

6) High correlation filter:-

This Method applies to two variables carrying the Same Information, thus of potentially degrading the model. In this method, we Identify the variables with the high correlation and use the Variance Inflation factor (VF) to choose one you can remove variable with a higher value (VIF > 5)

7) Decision trees:-

Decision trees are a popular Supervised learning algorithm that split data into a homo genous sets based on input variables.

This approach solves problems like data outliers, missing values, and Identifying the Significant variables.

8) Random forest:-

This method is like the decision tree strategy however. In this case, we generate a large set of trees (hence "forest") against the target variable. Then we find the feature subset with the help of each attribute usage statistics of each attributes.

9) factor Analysis:-

This method place highly correlated variable into their own group symbolizing a single factor or construct

Advantages of dimensionality Reduction:-

* It helps us data compression and hence reduced storage space
* It reduces computation time
* It also helps remove redundant features, if any

Disadvantages of Dimensionality Reduction:-

* It may lead to some Amount of data loss.
* PCA tends to find linear correlation b/w variables, which is sometimes undersirable.
* PCA fails in cases where mean and covariance are not enough to define the data sets.
* We may not know how many principle components to keep - in practice, some thumb rules are applied.

Principal Component Analysis:-

Principal component Analysis is an unsupervised learning algorithm that is used for the dimensionality reduction in machine learning

It is a stastical process that converts the observation is correlated features into a set of Linearly uncorre-lated features with the help of orthogonal transfor-mation. These new transformed features are called the principal components. It is one of the popular tools that is used for exploratory data Analysis and predictive modeling. It is a technique to draw strong patterns from the given dataset by reducing the variances

PCA generally tries to find the lower dimensional surface to project the high dimensional data

PCA works by considering the variance of each attribute because the high attribute shows the good split between the classes and hence it reduces the dimensionality. Some real-world Applications of PCA are Image processing, movie recommendation system. optimizing the power allocation in various communication channels. It is a feature Extraction technique, so it contains the Important variables and drops the least Important variable

The PCA Algorithm is based on some mathematical concepts such as

* Variance and covariance
* Eigen values and Eigen vectors.

Some Common term used in PCA Algorithm :-

Dimensionality:- It is the number of features or variables present in the given data set. More easily, it is the number of columns present in the dataset.

Correlation:-

It signifies that how strongly two variables are related to each other. Such as if one changed, the other variable also get changed. The correlation value ranges from -1 to +1 Here, -1 occurs if variable are Inversly proportional to each other and +1 indicates that variables are directly proportional to each other.

orthogonal :-

It defines that variables are correlated to each other and hence the correlation between the pair of variables is zero.

Eigen Vectors :-

If there is a square matrix M and a non zero vector v is given. Then v will be Eigen vector if Av is Scalar multiple of v

Covariance Matrix :-

A Matrix containing the covariance between the pair of variables is called the covariance matrix.

principal components of PCA :-

As desired described above, the transformed new features or the output of A "PCA" are the principal Components.

The number of the PCs are either equal or less than the Original features present in the dataset. Some properties of principal components are given below.

- The principal component must be the linear Combination of original features.
- These components are orthogonal i.e. The correlation between a pair of variables is zero.
- The importance of each other Component decreases when going to 1 to n. it means the 1st pc has the most Importance and npc will have the least Importance.

Remove less or unimportant of features from the new data set. The new features set has occured, so we will decode here what to keep and what to remove. It means we will only keep the relevant or Important features in the new data set and unimportance of features will be removed out.

Applications of principal component Analysis :-

- PCA is mainly used as the dimensionality reduction technique in various AI Applications Such as computer vision, Image compression etc,
- It can Also be used for finding hidden patterns if data has high dimensions. some fields where PCA is used are finance, data mining, psychology etc..

PCA Algorithm :-

The steps Involved in PCA Algorithm are as follows :

Step-01 : Get data set
Step-02 computation of mean of variables
Step-02 :. Compute the mean vector (M)
Step-03 computation of covariance matrix
Step-03 : Subtract Mean from the given data
Step-04 Eigen value, Eigen vector, Normalized eigen vector
Step-04 : Calculate the Covariance Matrix
Step-05 : Calculate the Eigen Vectors and Eigen values of the Covariance matrix
Step-06 : choosing components and forming a feature vector
Step-07 : Deriving the new data set (step-5).

Principal component Analysis :- (PCA)

Keypoints :-

* principal component Analysis is a well known dimension reduction technique
* It transforms the variables into a newset 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 of the data
* There can only two principal components for a two dimensional data set

problems :-

problems Based on Principal Component Analysis:

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

$x = 5, 6, 7$

$y = 6, 7, 8$

Solution :-

We use the above discussed PCA Algorithm

Step-01 :-

Given data

The given features vectors are -

- $x_1 = (2, 1)$
- $x_2 = (3, 5)$
- $x_3 = (4, 3)$

- $x_4 = (5,6)$
- $x_5 = (6,7)$
- $x_6 = (7,8)$

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

Step-02 :-

Calculate the value of mean vector (μ)

Mean Vector (μ)

$$= ((2+3+4+5+6+7)/6, (1+5+3+6+7+8)/6)$$

$$= (4.5, 5)$$

Thus,

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

Step-03 :-

Subtract Mean vector (μ) – the given feature vectors

- $x_1 - \mu = (2-4.5, 1-5) = (-2.5, -4)$
- $x_2 - \mu = (3-4.5, 5-5) = (-1.5, 0)$
- $x_3 - \mu = (4-4.5, 3-5) = (-0.5, -2)$
- $x_4 - \mu = (5-4.5, 6-5) = (0.5, 1)$
- $x_5 - \mu = (6-4.5, 7-5) = (1.5, 2)$
- $x_6 - \mu = (7-4.5, 8-5) = (2.5, 3)$

features vectors (x_i) After Subtracting Mean vector (μ) are

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

Step-04 :-

Calculate the covariance matrix

Covariance matrix is given by

$$\text{Covariance matrix} = \frac{\Sigma (x_i-\mu)(x_i-\mu)^t}{n}$$

Now,

$$m_1 = (x_1-\mu)(x_1-\mu)^t = \begin{bmatrix}-2.5\\-4\end{bmatrix}_{2\times1} \begin{bmatrix}-2.5 & -4\end{bmatrix}_{1\times2}$$

$$= \begin{bmatrix} 6.25 & 10 \\ 10 & 16 \end{bmatrix}$$

$$m_2 = (X_2 - \mu)(X_2 - \mu)^t = \begin{bmatrix} -1.5 \\ 0 \end{bmatrix} \begin{bmatrix} -1.5 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 2.25 & 0 \\ 0 & 0 \end{bmatrix}$$

$$m_3 = (X_3 - \mu)(X_3 - \mu)^t = \begin{bmatrix} -0.5 \\ -2 \end{bmatrix} \begin{bmatrix} -0.5 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} 0.25 & 1 \\ 1 & 4 \end{bmatrix}$$

$$m_4 = (X_4 - \mu)(X_4 - \mu)^t = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \begin{bmatrix} 0.5 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.25 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

$$m_5 = (X_5 - \mu)(X_5 - \mu)^t = \begin{bmatrix} 1.5 \\ 2 \end{bmatrix} \begin{bmatrix} 1.5 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 2.25 & 3 \\ 3 & 4 \end{bmatrix}$$

$$m_6 = (X_6 - \mu)(X_6 - \mu)^t = \begin{bmatrix} 2.5 \\ 3 \end{bmatrix}$$

Now,
Covariance Matrix $= (m_1 + m_2 + m_3 + m_4 + m_5 + m_6)/6$
on adding the above matrices and dividing by 6, we get

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

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

Step-05:-
Calculate the given Eigen values and Eigen vectors of the covariance matrix
λ is the Eigen value for a Matrix M if it is a solution of the characterstic 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

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

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

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

Solving the quadratic equation, we get

$$\lambda = 8.22,\ 0.38$$

Thus, two Eigen values are $\lambda_1 = 8.22$ and $\lambda_2 = 0.38$

Eigen values $\lambda_1 = 8.22$, $\lambda_2 = 0.38$

clearly, the second Eigen value is very small compared to the first Eigen value.

So, the second Eigen value 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.

$$MX = \lambda X$$

Where, M = Covariance Matrix

X = Eigen vector

λ = Eigen value

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

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

$3.67x_1 + 5.67x_2 = 8.22 \times 2$

on simplification, we get

$5.3x_1 = 3.67x_2$ — ①

$3.67x_1 = 2.55x_2$ — ②

from ① and ② $x_1 = 0.69x_2$

from ② the Eigen vector is

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

The principal component for the given data set is

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

$\therefore$ principal component is $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2.55 \\ 3.67 \end{bmatrix}$

Lastly, we project the data points on to the new subspace as

Covariance and correlation coefficient :-

In many fields of observational geoscience many variables are being monitored together as a function of space (or sample number) or time. The covariance is a measure of how variations in pairs of variables are linked to each other.

If we measure properties x_i and y_i for $i=1,2,\ldots n$, we can write the sample variances for x and y as

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \text{ and } S_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

We define the covariance between x and y

S_{xy} as

$$\therefore S_{xy}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

The covariance tells us how x and y values depend on each other The correlation coefficient r is a normalized version of the covariance and is given by

$$\therefore r = \frac{S^2xy}{S_x S_y}$$

The correlation coefficient is constrained for fall in the range ±1 A value of +1 tells us that the points (x_i, y_i) define a straight line with a positive slope

A value -1 tells us that the points (x_i, y_i) define a straight Line with a negative slope. A value of '0' shows that there is no dependence of y on x (or) vice versa (i.e., no correlation).

The quantity r^2 is the coefficient of determina-tion and it is a measure of the fraction of the variance of y that can be attributed to a relation-ship with x.

It is Important to note that 'r' Measures the strength of the Linear relationship of the variance between the two variables are Linearly related. It is easy to device non-linear relation ships that give a high correlation coefficient. It is Important to cook of the data and use common use.

Covariance formula:

Covariance formula is a Statistical formula which is used to Access the relationship between the two variables. In simple words, Covariance is one of the statistical Measurement to know the Relationship of the variance b/w the two variables.

The Covariance Indicates how two variables are related and also helps to know whether the data two variables vary together or change together. The Covariance is denoted as $cov(x,y)$ and the formulas for variance are given below.

Population Covariance formula $cov(x,y)$

$$= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N}$$

Sample Covariance formula $cov(x,y)$

$$= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Notations in covariance formulas

- x_i = data values of x
- y_i = data values of y
- $\bar{x}$ = means of x
- $\bar{y}$ = means of y
- N = number of data values

Example:-

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 and S&P returns have a positive or Inverse relationship. Before you compute the Covariance, calculate the mean of x and y.

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

$x = 2.1, 2.5, 4.0$ and 3.6 (Economic growth)

$y = 8, 12, 14$ and 10 (S&P 500 returns)

find $\bar{x}$ and $\bar{y}$

Solution:-

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2.1+2.5+4+3.6}{4}$$

$$\bar{x} = 3.1$$

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

$$\bar{y} = \frac{44}{4} = 11$$

Now, Substitute these values into the covariance formula to determine the relationship b/w economic growth and S&P 500 returns

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

$$\text{cov}(x,y) = \frac{(-1)(-3)+(-0.6)1+(0.9)3+(0.5)(-1)}{4-1}$$

$$= \frac{3-0.6+2.7-0.5}{3} = \frac{4.6}{3} = 1.53$$

Note:-

* positive covariance :- Indicates that Two Variables tend to move in Same direction.
* Negative covariance :- Reveals that Two variables tend To more in Inverse Directions.

Example:- for the following data find cov(x,y)

x	2, 4, 6, 8
y	1, 3, 5, 7

sol

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2+4+6+8}{4} = 5$$

Singular value Decomposition (SVD)

For $A \in R^{m \times n}$ with $m > n$ and rank$(A) < n$, the QR decomposition may be difficult to use, since numerically A may "Seem" to have rank n unlike earlier topics in this Semister we'll discuss what it is, and how to use it.

4.1 The SVD

We start with the main theoretical result to this Section (this should remaind you of your Introductory class, Where this was also presented)

Theorem 4.1.1 :- (Singular value Decomposition)

let $A \in C^{m \times n}$ be non-zero and Assume that rank $(A) = r > 0$ then

$$A = U \Sigma V^*$$

$A^T A = I$

Where $U \in C^{m \times m}$ and $V \in C^{n \times n}$ are unitary matrices $\Sigma \in R^{m \times n}$ is all zeros Except on The diagonal;

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}$$

and where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$

Singular value of Decomposition:-

Every m by n matrix can be factored in to $A = U \Sigma V^T$, Where U and V are orthogonal matrices and Σ is diagonal:

$$A = U\Sigma V^T = \underset{m \text{ by } m}{\begin{bmatrix} u_1 & u_2 & u_m \end{bmatrix}} \underset{m \text{ by } n}{\begin{bmatrix} \sigma_1 & \\ & \sigma_r \end{bmatrix}} \underset{n \text{ by } n}{\begin{bmatrix} v_1 & v_2 & v_n \end{bmatrix}^T}$$

The matrix Σ has the "Singular values" $\sigma_1 \ldots \sigma_r$ on its diagonal and is otherwise zero.

Compare with the Symmetric case $A = Q\Lambda Q^T$. The orthogonal matrices of U and V are no Longer the Same Q. The Input basis is not the Same as the output basis.

The Input basis starts with $v_1 \ldots v_r$ from the row space. It finishes with Any orthogonal basis $v_{r+1}, \ldots, v_n$ for the null space. Similarly the output basis

The Input basis starts with $v_1 \ldots v_r$ from the row space. It finishes with Any orthonormal basis $v_{r+1}, \ldots v_n$ for the null Space. Similarly the output basis starts with the good $u_1, \ldots u_r$ in the column Space and ends with Any orthonormal $u_{r+1} \ldots u_m$ in the left null space.

The Singular value Decomposition (SVD):- has $AV = U\Sigma$ with orthogonal matrices U and V then

$$A = U\Sigma V^{-1} = U\Sigma V^T$$

This is the new factorization:- Orthogonal times orthogonal. We have two matrices U and V instead of one matrix S. But there is a neat way to get U and out of the picture and see V by itself: Multiply A^T times A

$$A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T$$

$U^T U$ disappears because it equals. Then Σ^T is next to Σ multiplying those diagonal matrix gives σ_1^2 and σ_2^2. That leaves an ordinary factorization of the Symmetric Matrix $A^T A$

$$A^T A = V \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} V^T$$

In chapter 6 we would have called this $Q \Lambda Q^T$ The symmetric matrix was A itself Now symmetric matrix is $A^T A$! and the column of V are its Eigen Vectors.

This tells us how to find V. we Already to complete the Example.

Example 7.10 :-

Find the Singular value decomposition of $A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$

sol Compute $A^T A$ and its Eigen vectors

$$v_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \text{ and } v_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

The Eigen value of $A^T A$ are 2 and 8 The v's are perpendicular because Eigen vectors of every Symmetric matrix are perpendicular - and $A^T A$ is automatically symmetric

Example 7.12 :-

Find the vector SVD of the Singular matrix $A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$ the rank is $r = 1$. The row Space has only one basic vector v_1. The column Space has only one basic vector u_1. we can See those vector rs in A, and make them into unit vectors

$$v_1 = \text{multiple of row} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$u_1 = \text{multiple of column} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

that

Then AV_1 must be equal $\sigma_1 U_1$. It does with Singular value $\sigma_1 = \sqrt{10}$. The SVD could stop there (it usually doesn't):

$$\begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix} = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} [\sqrt{10}] [1/\sqrt{2} \ \ 1/\sqrt{2}]$$

It is customary for U and V to be square the matrices and need a Second column. The vector V_2 must be orthogonal to V_1, and U_2 must be orthogonal to U_1:

$$V_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ and } U_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

Notes:-

* This is called the Singular value Decomposition of A.
* The columns of U are orthonormal; these are called the left Singular vectors of A
* The columns of V are orthonormal; they are called the right Singular vector of A
* The values $\sigma_1, \dots \sigma_r$ are called the Singular values of A
* U and V are not unique; however Σ is unique

The vector V_2 is in the null space:-

It is perpendicular to V_1 in the row space. multiply by A to get $AV_2 = 0$. we would say that the Second singular value is $\sigma_2 = 0$ but this is against the rules, Singular values are like pivots - only the r non zeroes are counted.

If A is 2 by 2 then all three matrices U, Σ, V are 2 by 2 in the true SVD:

$$\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} = U \Sigma V^T = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix}\begin{pmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{pmatrix} 1/\sqrt{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

The matrices U and V contain orthonormal bases for all four fundamental Sub Spaces:

first r columns of V : row space of A
first $n-r$ columns of V : null space of A
Last r columns of U : column space of A
Last $m-r$ columns of U : null space of A^T

problem 1 :- find the singular values of the matrix

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

We compute AA^T (this is the smaller of two symmetric matrices associated with A) we get $AA^T = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 1 & 0 \\ 2 & 0 & 2 \end{pmatrix}$ we next find the Eigen values of this matrix. The characteristic polynomial is. $\lambda^3 - 6\lambda^2 + 6\lambda = \lambda|\lambda^2 - 6\lambda + 6|$ This gives three Eigen values: $\lambda = 3+\sqrt{3}$, $\lambda = 3-\sqrt{3}$ and $\lambda = 0$ Note that all are positive and that there are two non zero Eigen values, corresponding to the fact that A has rank 2

for the singular values of A, we now take the square roots of the Eigen values of AA^T, so

$\sigma_1 = \sqrt{3+\sqrt{3}}$ and $\sigma_2 = \sqrt{3-\sqrt{3}}$

(We don't have to mension the singular values which are zero).

Problem 2 :- Find the singular values of the matrix

$$B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

We use the same Approach : $AA^T = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$

This has characterstics polynomial : $\lambda^2 + 10\lambda + 9$, So $\lambda = 9$ and $\lambda = 1$ are the Eigen values. Hence the singular values are 3 and 1

Problem 3 :- find the singular values of $A = \begin{pmatrix} 0 & 1 & 1 \\ \sqrt{2} & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ and find the SDV decomposition of A.

We compute AA' and find $AA' = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 2 \end{bmatrix}$

The characteristic polynomial is

$$-\lambda^3 + 10\lambda^2 - 16\lambda = -\lambda(\lambda^2 - 10\lambda + 16)$$
$$= -\lambda(\lambda - 8)(\lambda - 2)$$

So the Eigen values of AA^T are $\lambda = 8, \lambda = 2, \lambda = 0$

Thus the Sigular values are

$\sigma_1 = 2\sqrt{2}$, $\sigma_2 = \sqrt{2}$ and $(\sigma_3 = 0)$

To The given the decomposition we consider the diagonal matrix of singular values

$$\Sigma_1' = \begin{bmatrix} 2\sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Next, we find an orthogonal set of Eigen vectors for AA^T for $\lambda = 8$, we find an Eigen vector $(1, 2, 1)$ normalizing gives $P_1 = \left(\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right)$

for $\lambda = 2$ we find $P_2 = \left(-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right)$ and finally for $\lambda = 0$ we get $P_3 = \left(\frac{1}{\sqrt{2}}, 0, -1/\sqrt{2}\right)$

This gives the matrix $P = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ 2/\sqrt{6} & 1/\sqrt{3} & 0 \\ 1/\sqrt{6} & -1/\sqrt{3} & -1/\sqrt{2} \end{bmatrix}$

finally we have to find an orthogonal set of Eigen vectors of $A^TA = \begin{bmatrix} 2 & 2\sqrt{2} & 0 \\ 2\sqrt{2} & 6 & 2 \\ 0 & 2 & 2 \end{bmatrix}$

This can be done in two ways we show both ways Starting with orthogonal diagonalization.

We already known that the Eigen values will be $\lambda = 8, \lambda = 2, \lambda = 0$. This gives Eigen vectors

$q_1 = \left(\frac{1}{\sqrt{6}}, \frac{3}{\sqrt{12}}, \frac{1}{\sqrt{12}}\right)$, $q_2 = \left(\frac{1}{\sqrt{3}}, 0, -\frac{2}{\sqrt{6}}\right)$ and

$q_3 = \left(\frac{1}{\sqrt{2}}, \frac{-1}{2}, \frac{1}{2}\right)$ put these together to get

$$Q = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{12}} & 0 & -\frac{1}{2} \\ \frac{1}{\sqrt{12}} & \frac{-2}{\sqrt{6}} & \frac{1}{2} \end{bmatrix}$$

for a quicker method, we calculate the columns of Q using those of P using the formula

$$P_i = \frac{1}{\sigma_i} A^T p_i$$

Thus we calculate

$$P_1 = \frac{1}{\sigma_1} A^T P_1 = \frac{1}{\sqrt{8}} \begin{bmatrix} 0 & \sqrt{2} & 0 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix} = q_1$$

And similarly for other two columns Either way we can now we verify that we have $A = P\Sigma Q^T$.

Gram-SCHMIDT Method :-

In Any linear products space, we can choose the basis in which to work

Innerproduct Space :-

It offers greatly simplifies calculations to work in an orthogonal basis for one thing.

if $S = \{V_1, V_2, --- V_n\}$ is an orthogonal basic for an Inner product space V, which means all pairs of distinct vector in S are orthogonal

$\langle V_i, V_j \rangle = 0$ for all $V_i, V_j \in S$

Orthogonal sets :-

let V be a vector space with an Inner product

Definition :-

Non zero vectors $V_1, V_2 --- V_k \in V$ from an orthogonal set it they are orthogonal to each other $\langle V_i, V_j \rangle = 0$ for $i \neq j$ if, in addition, all vectors are of unit norm

$\|V_i\| = 1$, then $V_1, V_2 --- V_c$ is called an orthonormal set

orthogonal projection:-

let v be an Inner product space

let $x, v \in V$, $v \neq 0$ then $p = \frac{\langle x, y \rangle}{\langle v, v \rangle}$ is the orthogonal projection of vector x on to the vector v. that is the remaining remainder $o = x - p$ is orthogonal to v.

If $v_1, v_2, \ldots v_n$ is an orthogonal set of vectors then $p = \frac{\langle x, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 + \frac{\langle x, v_2 \rangle}{\langle v_2, v_2 \rangle} + \cdots + \frac{\langle x, v_n \rangle}{\langle v_n, v_n \rangle} v_n$ is the orthogonal projection of vector 'x' on the subspace spanned by $v_1 \ldots v_n$. That is, the remainder $o = x - p$ is orthogonal to $v_1 \ldots v_n$

The Gram - Schmidt process :-

If $B = \{v_1, \ldots v_n\}$ is a basic for a subspace $H < R^m$ and $U_i = V_i = proj_{span}\{v_i \ldots v_{i-1}\} v_i$ for $i \leq i \leq n$, then $\{u_i\}_{i=1}^{m}$ is an orthogonal basis for H and $\{e_i = \hat{u}_i\}_{i=1}^{n}$ is an orthonormal basis for M

The Gram- Schmidt orthoganalization process

let v be a vector space with an Inner product Suppose $x_1, x_2 \ldots x_n$ is a basic for v. let $v_1 = x_1$,

$v_2 = x_2 - \frac{\langle x_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle x_3, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2$,

$v_n = x_n - \frac{\langle x_n, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \cdots - \frac{\langle x_n, v_{n-1} \rangle}{\langle v_{n-1}, v_{n-1} \rangle}$

Then $v_1, v_2, \ldots v_n$ is an orthogonal basis for v

Any basis $x_1, x_2, \ldots x_n$ — orthogonal basis $v_1, v_2, \ldots v_n$

key concepts:-

Given An Arbitrary basics $\{x_1, x_2, \ldots x_n\}$ for an n-dimensional Inner product space v, the

Gram-Schmidt Algorithm :-

Constructs An Algorithm basics $\{v_1, v_2 - - - v_n\}$ for v :

Step-1 :- let $v_1 = x_1$

Step-2 :- let $v_2 = x_2 - \frac{\langle x_2, v_1 \rangle}{\|v_1\|^2} v_1$

Step 3 :- let $v_3 = x_3 - \frac{\langle x_3, v_1 \rangle}{\|v_1\|^2} - \frac{\langle v_3 v_2 \rangle}{\|v_2\|^2} v_2$

Example :-

let $V = R^3$ with Euclidean Inner product we will Apply the Gram-Schmit Algorithm to orthogonalize the basis $\{(1,-1,1), (1,0,1), (1,1,2)\}$

Step 1 : $v_1 = (1,-1,1)$

Step 2 : $v_2 = (1,0,1) - \frac{(1,0,1).(1,-1,1)}{\|(1,-1,1)\|^2}(1,-1,1)$

$= (1,0,1) - \frac{2}{3}(1,-1,1)$

$\therefore v_2 = (1/3, 2/3, 1/3)$

Step 3 :- $v_3 = (1,1,2) - \frac{(1,1,2).(1,-1,1)}{\|(1,-1,1)\|^2}(1,-1,1) - \frac{(1,1,2).(1/3, 2/3, 1/3)}{\|1/3, 2/3, 1/3\|^2}$

$(1/2, 2/3, 1/3)$.

$= (1,1,2) - 2/3(1,-1,1) - 5/2(1/3, 2/3, 1/3)$

$v_3 = (-1/2, 0, 1/2)$

you can verify that $\{(1,-1,1), (1/3, 2/3, 1/3), (-1/2, 0, 1/2)\}$ forms an orthogonal basis for R^3. Normalizing the vectors in the orthogonal basis we obtain the orthonormal basis.

Properties of the Gram-Schmidt process :-

* $v_k = x_k - (\alpha_1 x_1 + - - - + \alpha_{k-1} x_{k-1})$, $1 \le k \le n$;
* The span of $v_1 - - - v_k$ is the same as the span of $x_1, - - - x_k$;
* v_k is orthogonal to $x_1, - - - x_{k-1}$;
* $v_k = x_k - p_k$, where p_k is the orthogonal

$||\alpha||$ is defined as the positive square root of $\langle \alpha, \alpha \rangle$.

Norm or length of $\alpha \in V = ||\alpha|| = \sqrt{\langle \alpha, \alpha \rangle}$

$\Rightarrow \boxed{||\alpha||^2 = \langle \alpha, \alpha \rangle}$

UNIT vector:-

let $V(F)$ be an inner product space. $\alpha \in V$ is called a unit vector if $||\alpha|| = 1$

if $\alpha \in V$ then $\frac{\alpha}{||\alpha||} \in V$ is unit vector

Orthogonal vector:-

let $V(F)$ be an inner product space and $\alpha, \beta \in V$. α is said to be orthogonal to β. If $\langle \alpha, \beta \rangle = 0$

α is orthogonal to $\beta \Leftrightarrow \langle \alpha, \beta \rangle = 0 \Leftrightarrow \langle \beta, \alpha \rangle = \overline{\langle \alpha, \beta \rangle} = 0$

(conjugate of 0 is 0)

$\Leftrightarrow \beta$ is orthogonal to α.

Ortho Normal Set:-

let 'S' be a non-empty subset of an inner product space $V(F)$. The set 'S' is said to be an orthonormal set if $||\alpha_i|| = 1$ for each $\alpha_i \in S$ and $\langle \alpha_i, \alpha_j \rangle = 0$ for $\alpha_i, \alpha_j \in S$, $i \neq j$

$S \subset V$ is an orthogonal set if $\langle \alpha_i, \alpha_j \rangle = 0$, for $i = j$ and $\langle \alpha_i, \alpha_j \rangle = 0$ for $i \neq j$ where $\alpha_i, \alpha_j \in S$

that is $\langle \alpha_i, \alpha_j \rangle = \delta_{ij}$

Gram-Schmidt orthogonalisation process:

A basis of an innerproduct space $V(F)$ which is also orthonormal is called "orthonormal basis of the inner product space.

Working method for finding orthogonal basis:-

let $\{\beta_1, \beta_2, \ldots \beta_n\}$ be the given linearly independent basis of $V(F)$. The orthogonal basis of $V(F)$ namely $\{\gamma_1, \gamma_2, \ldots \gamma_n\}$ is given by the following relations

$\gamma_1 = \beta_1$

$\gamma_2 = \beta_2 - \frac{\langle \beta_2, \gamma_1 \rangle}{\|\gamma_1\|^2} \gamma_1$

$\gamma_3 = \beta_3 - \frac{\langle \beta_3, \gamma_1 \rangle}{\|\gamma_1\|^2} \gamma_1 - \frac{\langle \beta_3, \gamma_2 \rangle}{\|\gamma_2\|^2} \gamma_2$

$\vdots$

$\gamma_n = \beta_n - \frac{\langle \beta_n, \gamma_1 \rangle}{\|\gamma_1\|^2} \gamma_1 - \cdots - \frac{\langle \beta_n, \gamma_{n-1} \rangle}{\|\gamma_{n-1}\|^2} \gamma_{n-1}$

Working method for finding orthonormal Basis:-

let $\{\beta_1, \beta_2, \cdots \beta_n\}$ be a given basis of a finite dimensional inner product space $V(F)$. The vectors $\alpha_1, \alpha_2, \cdots \alpha_n$ of the orthonormal basis of $V(F)$ are given by

$\alpha_1 = \frac{\gamma_1}{\|\gamma_1\|} = \frac{\beta_1}{\|\beta_1\|}$

$\alpha_2 = \frac{\gamma_2}{\|\gamma_2\|}$

$\alpha_3 = \frac{\gamma_3}{\|\gamma_3\|}$

$\vdots$

$\alpha_n = \frac{\gamma_n}{\|\gamma_n\|}$

problem 1 :-

1. Given $B = \{u_1, u_2, u_3\}$ where $u_1 = (1, 2, 1)$, $u_2 = (1, 1, 3)$ and $u_3 = (2, 1, 1)$, use the Gram-Schmidt procedure to obtain a corresponding orthonormal B.

Sol Given $B = \{u_1, u_2, u_3\}$

$u_1 = (1, 2, 1)$, $u_2 = (1, 1, 3)$, $u_3 = (2, 1, 1)$

By Gram - Schmidt procedure orthonormal basis be $\{\alpha_1, \alpha_2, \alpha_3\}$

$\alpha_1 = \frac{\gamma_1}{\|\gamma_1\|}$

$\gamma_1 = u_1 = (1, 2, 1)$

$\|\gamma_1\| = \sqrt{1^2 + 2^2 + 1^2} = \sqrt{6}$

$$\alpha_1 = \frac{v_1}{\|v_1\|} = \frac{(1,2,1)}{\sqrt{6}} = \left(\frac{1}{\sqrt6}, \frac{2}{\sqrt6}, \frac{1}{\sqrt6}\right)$$

$$\alpha_2 = \frac{v_2}{\|v_2\|}$$

$$v_2 = u_2 - \frac{\langle u_2, v_1\rangle}{\|v_1\|^2} v_1 \rightarrow (1)$$

$$\langle u_2, v_1\rangle = (1,1,3)(1,2,1) = 1+2+3 = 6$$

$$\|v_1\|^2 = (\sqrt6)^2 = 6$$

$$(1) \Rightarrow v_2 = (1,1,3) - \frac{6}{6}(1,2,1)$$

$$= (1,1,3) - (1,2,1) = (0,-1,2)$$

$$\|v_2\| = \sqrt{(-1)^2 + 2^2} = \sqrt5$$

$$\alpha_2 = \frac{v_2}{\|v_2\|} = \frac{(0,-1,2)}{\sqrt5}$$

$$\alpha_2 = \left(0, \frac{-1}{\sqrt5}, \frac{2}{\sqrt5}\right)$$

$$\alpha_3 = \frac{v_3}{\|v_3\|}$$

$$v_3 = u_3 - \frac{\langle u_3, v_1\rangle}{\|v_1\|^2} v_1 - \frac{\langle u_3, v_2\rangle}{\|v_2\|^2} v_2$$

$$\langle u_3, v_1\rangle = (2,1,1)(1,2,1)$$

$$= 2+2+1 = 5$$

$$\langle u_3, v_2\rangle = (2,1,1)\cancel{(\sqrt5)}(0,-1,2) = \cancel{2\sqrt5 + \sqrt5 + \sqrt5} = \cancel{4\sqrt5} = 0-1+2 = 1$$

$$\|v_2\|^2 = (\sqrt5)^2 = 5$$

$$v_3 = (2,1,1) - \frac{5}{6}(1,2,1) - \frac{\overset{1}{\cancel{4\sqrt5}}}{5}(0,-1,2)$$

$$= \left(2 - \frac{5}{6}, 1 - \frac{10}{6} + \frac{1}{5}, 1 - \frac{5}{6} - \frac{2}{5}\right)$$

$$\left(\frac{7}{6}, \frac{-7}{15}, \frac{-7}{30}\right)$$

$$\|\gamma_3\| = \sqrt{\left(\frac{7}{6}\right)^2+\left(\frac{-7}{15}\right)^2+\left(\frac{-7}{30}\right)^2} = \sqrt{\frac{49}{36}+\frac{49}{225}+\frac{49}{900}}$$

$$\frac{7\sqrt{30}}{30}$$

$$\alpha_3 = \frac{\gamma_3}{\|\gamma_3\|} = \frac{\left(\frac{7}{6}, \frac{-7}{15}, \frac{-7}{30}\right)}{\frac{7\sqrt{30}}{30}}$$

$$= \left(\frac{7}{6}\times\frac{30}{7\sqrt{30}}, \frac{-7}{15}\times\frac{30}{7\sqrt{30}}, \frac{-7}{30}\times\frac{30}{7\sqrt{30}}\right)$$

$$= \frac{5}{\sqrt{30}}, \frac{-2}{\sqrt{30}}, \frac{-1}{\sqrt{30}}$$

$\therefore$ The required orthogonal basis is

$$= \left\{\left(\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right), \left(0, \frac{-1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right), \left(\frac{5}{\sqrt{30}}, \frac{-2}{\sqrt{30}}, \frac{-1}{\sqrt{30}}\right)\right\}$$

2) construct the orthogonal basis for $\{(2,1,3)\ (1,2,3)\ (1,1,1)\}$

Given $B = \{B_1, B_2, B_3\}$

$B_1 = (2,1,3)$, $B_2 = (1,2,3)$, $B_3 = (1,1,1)$

By Gram-Schmidt procedure orthonormal basis be $\{\alpha_1, \alpha_2, \alpha_3\}$

$$\alpha_1 = \frac{\gamma_1}{\|\gamma_1\|}$$

$$\gamma_1 = B_1 = (2,1,3)$$

$$\|\gamma_1\| = \sqrt{2^2+1^2+3^2} = \sqrt{4+1+9} = \sqrt{14} =$$

$$\alpha_1 = \frac{\gamma_1}{\|\gamma_1\|} = \frac{(2,1,3)}{\sqrt{14}} = \left(\frac{2}{\sqrt{14}}, \frac{1}{\sqrt{14}}, \frac{3}{\sqrt{14}}\right)$$

$$\alpha_2 = \frac{\gamma_2}{\|\gamma_2\|}$$

$$\gamma_2 = B_2 - \frac{\langle B_2, \gamma_1\rangle}{\|\gamma_1\|^2}\gamma_1$$

$$\langle B_2, \gamma_1\rangle = (1,2,3)(2,1,3) = 2+2+9 = 13$$

$$\|\gamma_1\|^2 = (\sqrt{14})^2 = 14$$

$$q_2 = (1,2,3) - \frac{13}{14}(2,1,3)$$

$$= \frac{14(1,2,3) - 13(2,1,3)}{14}$$

$$= \left(1 - \frac{26}{14},\ 2 - \frac{13}{14},\ 3 - \frac{39}{14}\right)$$

28
13
15

42

$$= \left(-\frac{12}{14},\ \frac{15}{14},\ \frac{3}{14}\right)$$

UNIT - I

problems

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, $n = 2$

No. of samples, $N = 4$

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 covariance Matrix

ordered pairs are $(x,x),(x,y)(y,x)(y,y)$

$$cov(x,x) = \frac{1}{N-1}\sum_{N=1}^{N}(x_i-\bar{x})^2$$

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

$$= \frac{1}{3}\left[16+0+25+1\right] = \frac{42}{3} = 14$$

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

$$= \cancel{\tfrac{1}{3}}\left[(4-8)+(8-8)+(13-8)+(7-8)\right)$$

$$= \frac{1}{3}\Big[(4-8)(11-8.5)+(8-8)(4-8.5)+(13-8)(5-8.5)$$
$$+(7-8)(14-8.5)\Big]$$

$$= \frac{1}{3}\left[-10+0-17.5-5.5\right] = -11$$

$$\text{cov}(y,y) = \frac{1}{N-1} \sum (y_i - \bar{y})^2$$

$$= \frac{1}{3}\left[(11-8.5)^2 + (4-8.5)^2 + (5-8.5)^2 + (14-8.5)^2 \right]$$

$$= \frac{1}{3}\left[6.25 + 20.25 + 12.25 + 30.25 \right]$$

$$= \frac{1}{3}[69] = 23$$

$\therefore$ Covariance Matrix $S = \begin{array}{cc} \text{cov}(x,x) & \text{cov}(x,y) \\ \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix} & \\ \text{cov}(y,x) & \text{cov}(y,y) \end{array}$

Step 4:- calculate the Eigen vector, Eigen values and Normalize.

characteristic equation of S is $|S - \lambda I| = 0$

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

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

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

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

$$\lambda_1 = 30.38, \quad \lambda_2 = 6.62$$

$$\boxed{\lambda_1 > \lambda_2}$$

Eigen characteristic matrix of S is $(S - \lambda I)V = 0$

$$\begin{bmatrix} 14-\lambda & -11 \\ -11 & 23 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 0$$

Eigen vector corresponding to Eigen value $\lambda = 30.38$

put $\lambda = 30.38$

$$\begin{bmatrix} 14-30.38 & -11 \\ -11 & 23-30.38 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 0$$

$$\begin{bmatrix} -16.38 & -11 \\ -11 & -7.38 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 0$$

$-16.38V_1 - 11V_2 = 0 \times 11$ $\qquad$ $-180.18V_1 - 121V_2 = 0$

$-11V_1 - 7.38V_2 = 0 \times 16.38$ $\qquad$ $-180.18 - 120.8644V_2 = 0$

$0.1156 V_2 = 0$

$V_1 = 0$

$$\begin{pmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{pmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 0$$

$(14-\lambda)V_1 - 11V_2 = 0 \rightarrow (i)$

$-11V_1 + (23-\lambda)V_2 = 0 \rightarrow (ii)$

$(i) \Rightarrow (14-\lambda)V_1 = 11V_2$

$$\frac{V_1}{11} = \frac{V_2}{14-\lambda} = t \quad \text{(let)}$$

$V_1 = 11t \quad V_2 = (14-\lambda)t$

if $t = 1$

$V_1 = 11 \quad V_2 = 14-\lambda$

If $\lambda = 30.38$ then

$V_1 = 11 \quad V_2 = -16.38$

$$\therefore V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} 11 \\ -16.38 \end{pmatrix}$$

Norm (length) of V, $\|V\| = \sqrt{V_1^2 + V_2^2}$

$$\|V\| = \sqrt{11^2 + (-16.38)^2}$$

$$= 19.73$$

Normalization V, $e = \begin{pmatrix} \frac{V_1}{\|V\|} \\ V_2/\|V\| \end{pmatrix}$

$$= \begin{bmatrix} 11/19.73 \\ -16.38/19.73 \end{bmatrix}$$

$$e = \begin{bmatrix} 0.55 \\ -0.83 \end{bmatrix}$$

Step : 5 :- Derive new data set

		Ex_1	Ex_2	Ex_3	Ex_4
First PCA	(PC_1)	P_{11}	P_{12}	P_{13}	P_{14}

$P_{11} = e^T \begin{pmatrix} x_1 - \bar{x} \\ y_1 - \bar{y} \end{pmatrix}$

$= \begin{bmatrix} 0.55 & -0.83 \end{bmatrix} \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix}$

$= \begin{bmatrix} 0.55 & -0.83 \end{bmatrix} \begin{bmatrix} -4 \\ 2.5 \end{bmatrix}$

$= -2.2 - 2.07$

$P_{11} = -4.27$

$P_{12} = e^T \begin{pmatrix} x_2 - \bar{x} \\ y_2 - \bar{y} \end{pmatrix}$

$= \begin{bmatrix} 0.55 & -0.83 \end{bmatrix} \begin{bmatrix} 8-8 \\ 4-8.5 \end{bmatrix}$

$P_{12} = 3.735$

$P_{13} = e^T \begin{bmatrix} x_3 - \bar{x} \\ y_3 - \bar{y} \end{bmatrix}$

$= \begin{bmatrix} 0.55 & -0.83 \end{bmatrix} \begin{bmatrix} 13-8 \\ 5-8.5 \end{bmatrix}$

$= 5.655$

$P_{14} = e^T \begin{bmatrix} x_4 - \bar{x} \\ y_4 - \bar{y} \end{bmatrix}$

$= \begin{bmatrix} 0.55 & -0.83 \end{bmatrix} \begin{bmatrix} 7-8 \\ 14-8.5 \end{bmatrix}$

$= -5.11$

The new data set is

	ex_1	ex_2	ex_3	ex_4
PC_1	-4.27	3.735	5.65	-5.11

2)

Step-1 :- Given data set

x_i	2	3	4	5	6	7
y_i	1	5	3	6	7	8

No. of features, $n = 2$
no. of samples, $N = 6$

step-2 :- computing mean of variables

$$\bar{x} = \frac{2+3+4+5+6+7}{6} = 4.5$$

$$\bar{y} = \frac{1+5+3+6+7+8}{6} = 5$$

step-3 :- computation of covariance matrix

$$cov(x,x) = \frac{1}{N-1} \sum (x_i - \bar{x})^2$$

$$= \frac{1}{6-1}\left[(2-4.5)^2 + (3-4.5)^2 + (4-4.5)^2 + (5-4.5)^2 + (6-4.5)^2 + (7-4.5)^2\right]$$

$$cov(x,x) = 3.5$$

$$cov(x,y) = cov(y,x) = \frac{1}{N-1} \sum_{i=1}^{6} (x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{6-1}\left[(2-4.5)(1-5) + (3-4.5)(5-5) + (4-4.5)(3-5) + (5-4.5)(6-5) + (6-4.5)(7-5) + (7-4.5)(8-5)\right]$$

$$= \frac{1}{5}[24]$$

$$= 4.8$$

$$cov(y,y) = \frac{1}{N-1} \sum_{i=1}^{6} (y_i - \bar{y})^2$$

$$= \frac{1}{5}\left[(1-5)^2 + (5-5)^2 + (3-5)^2 + (6-5)^2 + (7-5)^2 + (8-5)^2\right]$$

$$= \frac{1}{5}[34]$$

$$= 6.5$$

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

$$S = \begin{bmatrix} 3.5 & 4.8 \\ 4.8 & 6.5 \end{bmatrix}$$

Step-04 :- eigen values, eigen vectors, normalized eigen vectors

characteristic equation of S is $|S - \lambda I| = 0$

$$\begin{vmatrix} 3.5-\lambda & 4.8 \\ 4.8 & 6.5-\lambda \end{vmatrix} = 0$$

$$(3.5-\lambda)(6.5-\lambda)-(4.8)(4.8)=0$$

$$22.75-3.5\lambda-6.5\lambda+\lambda^2-23.04=0$$

$$\lambda^2-10\lambda-0.29=0$$

$$\lambda_1=10.02 \quad \lambda_2=-0.028$$

$$\lambda_1>\lambda_2$$

$$\begin{pmatrix} 3.5-\lambda & 4.8 \\ 4.8 & 6.5-\lambda \end{pmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}=0$$

$$(3.5-\lambda)v_1+4.8v_2=0 \rightarrow (i)$$

$$4.8v_1+(6.5-\lambda)v_2=0 \rightarrow (ii)$$

$$(i) \Rightarrow (3.5-\lambda)v_1=-4.8v_2$$

$$(\lambda-3.5)v_1=4.8v_2$$

$$\frac{v_1}{4.8}=\frac{v_2}{\lambda-3.5}=t\,(\text{let})$$

let $t=1$ $\lambda=10.02$

$$\frac{v_1}{4.8}=\frac{v_2}{10.02-3.5}=1$$

$$v_1=4.8 \quad v_2=6.52$$

$$v=\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}=\begin{bmatrix} 4.8 \\ 6.52 \end{bmatrix}$$

Norm (length) of v $\|v\|=\sqrt{v_1^2+v_2^2}$

$$=\sqrt{(4.8)^2+(6.52)^2}$$

$$=8.096$$

Normalizing v, i.e $=\begin{bmatrix} v_1/\|v\| \\ v_2/\|v\| \end{bmatrix}$

$$=\begin{pmatrix} 4.8/8.09 \\ 6.52/8.09 \end{pmatrix}=\begin{bmatrix} 0.593 \\ 0.805 \end{bmatrix}$$

$$=\begin{bmatrix} 0.59 \\ 0.80 \end{bmatrix}$$

Step 5:- Derive new data set

first PCA P_{11} P_{12} P_{13} P_{14} P_{15} P_{16}

$$P_{11} = e^T \begin{pmatrix} x_1 - \bar{x} \\ y_1 - \bar{y} \end{pmatrix}$$

$$= \begin{bmatrix} 0.59 & 0.80 \end{bmatrix} \begin{bmatrix} 2-4.5 \\ 1-5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.59 & 0.80 \end{bmatrix} \begin{bmatrix} -2.5 \\ -4 \end{bmatrix}$$

$$= -1.475 - 3.2$$

$$= -4.675$$

$$P_{12} = e^T \begin{pmatrix} x_2 - \bar{x} \\ y_2 - \bar{y} \end{pmatrix}$$

$$= \begin{bmatrix} 0.59 & 0.80 \end{bmatrix} \begin{bmatrix} 3-4.5 \\ 5-5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.59 & 0.80 \end{bmatrix} \begin{bmatrix} -1.5 \\ 0 \end{bmatrix}$$

$$= -0.885$$

$$P_{13} = e^T \begin{bmatrix} x_3 - \bar{x} \\ y_3 - \bar{y} \end{bmatrix}$$

$$= \begin{bmatrix} 0.59 & 0.80 \end{bmatrix} \begin{bmatrix} 4-4.5 \\ 3-5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.59 & 0.80 \end{bmatrix} \begin{bmatrix} -0.5 \\ -2 \end{bmatrix}$$

$$= -0.295 - 1.6$$

$$= 1.895$$

$$P_{14} = e^T \begin{pmatrix} x_4 - \bar{x} \\ y_4 - \bar{y} \end{pmatrix}$$

$$= \begin{bmatrix} 0.59 & 0.80 \end{bmatrix} \begin{bmatrix} 5-4.5 \\ 6-5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.59 & 0.80 \end{bmatrix} \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$$

$$= 0.59 \times 0.5 + 0.80$$

$$= 1.095$$

$$P_{15} = e^T \begin{pmatrix} x_5 - \bar{x} \\ y_5 - \bar{y} \end{pmatrix}$$

$$= \begin{bmatrix} 0.59 & 0.80 \end{bmatrix} \begin{bmatrix} 6-4.5 \\ 7-5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.59 & 0.80 \end{bmatrix} \begin{bmatrix} 1.5 \\ 2 \end{bmatrix}$$

$$= 0.885 + 1.6$$

$$= 2.485$$

$$P_{16} = e^T \begin{pmatrix} x_6 - \bar{x} \\ y_6 - \bar{y} \end{pmatrix}$$

$$= \begin{bmatrix} 0.59 & 0.80 \end{bmatrix} \begin{bmatrix} 7-4.5 \\ 8-5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.59 & 0.80 \end{bmatrix} \begin{bmatrix} 2.5 \\ 3 \end{bmatrix}$$

$$= 1.475 + 2.4$$

$$= 3.875$$

The new data set is

PC_1	-4.675	-0.885	1.895	1.095	2.485	3.875

10M

1) obtain the singular value decomposition (SVD) of A, $U\Sigma V^T$, where $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}$

Sol Given matrix

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}_{3\times 2}$$

let $A = U\Sigma V^T$ be SVD of A

Step-1 :- compute v

consider the matrix $A^T.A = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}_{2\times 3} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}_{3\times 2}$

$$= \begin{bmatrix} 1+0+1 & 1+0-1 \\ 1+0-1 & 1+1+1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}_{2\times 2}$$

We have to find eigen values & eigen vectors of $A^T.A$

characteristic equation of A^TA is

$$|A^TA - \lambda I| = 0$$

$$\begin{vmatrix} 2-\lambda & 0 \\ 0 & 3-\lambda \end{vmatrix} = 0$$

$$(2-\lambda)(3-\lambda) = 0$$

$$\lambda = 2, 3$$

$\therefore \lambda = 3, 2$ are eigen values of $A^T.A$

characteristic matrix of $A^T.A$ is $(A^TA - \lambda I)v = 0$

$$\begin{bmatrix} 2-\lambda & 0 \\ 0 & 3-\lambda \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{—} \; ①$$

for $\lambda = 3$ eq① becomes

$$\begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$-v_1 + 0.v_2 = 0 \quad \text{— (i)}$$

$$0v_1 + 0.v_2 = 0 \quad \text{— (ii)}$$

for $v_1 = 0$ & $v_2 = 1$ eq (i) satisfied

$$V_1 = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

length of $V_1 = \sqrt{V_1^2+V_2^2} = \sqrt{0^2+1^2} = 1$

Normalized $V_1 = N(V_1) = \begin{bmatrix} 0/1 \\ 1/1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

for $\lambda = 2$ eq ① becomes

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$0.V_1 + 0.V_2 = 0 \longrightarrow$ (i)

$0.V_1 + V_2 = 0 \longrightarrow$ (ii)

for $V_2 = 0$ & $V_1 = 1$ eq ① satisfied

$$V_2 = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

length of $V_2 = \sqrt{1^2+0^2} = 1$

Normalized $V_2 = N(V_2) = \begin{bmatrix} 1/1 \\ 0/1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$$V = [N(V_1) \ N(V_2)] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$V^T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Step-2 :- Compute Σ

order of Σ = order of A

Now $\sigma_1 = \sqrt{\lambda_1} = \sqrt{3}$

$\sigma_2 = \sqrt{\lambda_2} = \sqrt{2}$

No. of non-zero Eigen values = Rank

Rank = 2

$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$ i.e Σ_1 is a diagonal matrix whose diagonal entries are σ_1 & σ_2

$\Sigma_1 = R \times R = 2 \times 2$

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix}_{3\times 2}$$

Step 3 :- Compute U

Consider the matrix $A.A^T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}$

$= \cancel{1+1} \quad \cancel{0+1} \quad -1+1$

$$= \begin{bmatrix} 1+1 & 0+1 & -1+1 \\ 0+1 & 0+1 & 0+1 \\ -1+1 & 0+1 & 1+1 \end{bmatrix}$$

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

We have to find Eigen values & Eigen vectors of AA^T

characteristic equation of AA^T is $|AA^T - \lambda I| = 0$

$$\begin{vmatrix} 2-\lambda & 1 & 0 \\ 1 & 1-\lambda & 1 \\ 0 & 1 & 2-\lambda \end{vmatrix} = 0$$

$$(2-\lambda)((1-\lambda)(2-\lambda)-1) - 1(2-\lambda) = 0$$

$$(2-\lambda)(2-\lambda-2\lambda+\lambda^2-1) - 2+\lambda = 0$$

$$(2-\lambda)(\lambda^2-3\lambda+1) - 2+\lambda = 0$$

$$2\lambda^2 - 6\lambda + \cancel{2} - \lambda^3 + 3\lambda^2 - \cancel{\lambda} - \cancel{2} + \cancel{\lambda} = 0$$

$$-\lambda^3 + 5\lambda^2 - 6\lambda = 0$$

$$\lambda(-\lambda^2+5\lambda-6) = 0$$

$$-\lambda(\lambda^2-5\lambda+6) = 0$$

$$\lambda = 0 \quad | \quad \lambda^2 - 5\lambda + 6 = 0$$

$$\lambda^2 - 2\lambda - 3\lambda + 6 = 0$$

$$\lambda(\lambda-2) - 3(\lambda-2) = 0$$

$$\lambda = 2, 3$$

$$\lambda = 0, 2, 3$$

$\therefore \lambda = 3, 2, 0$ are the Eigen values of AA^T

characteristic matrix of AA^T is $(AA^T - \lambda I)U = 0$

$$\begin{bmatrix} 2-\lambda & 1 & 0 \\ 1 & 1-\lambda & 1 \\ 0 & 1 & 2-\lambda \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$(2-\lambda)U_1 + U_2 = 0$$

$$U_1 + (1-\lambda)U_2 + U_3 = 0$$

$$U_2 + (2-\lambda)U_3 = 0$$

for $\lambda = 3$ eq ① becomes

$-U_1 + U_2 = 0$ — (i)

$U_1 - 2U_2 + U_3 = 0$ — (ii)

Solving (i) & (ii)

$$\begin{array}{cccc} U_1 & U_2 & U_3 & \\ 1 & 0 & -1 & 1 \\ -2 & 1 & 1 & -2 \end{array}$$

$$\frac{U_1}{1} = \frac{U_2}{1} = \frac{U_3}{1} = k$$

$U_1 = 1, \; U_2 = 1, \; U_3 = 1$

$$U_1 = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

length of $U_1 = \sqrt{U_1^2 + U_2^2 + U_3^2}$
$= \sqrt{1^2 + 1^2 + 1^2} = \sqrt{3}$

Normalized U_1, $N(U_1) = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$

for $\lambda = 0$ eq ① becomes

$2U_1 + U_2 + 0.U_3 = 0$ — (i)

$U_1 + U_2 + U_3 = 0$ — (ii)

$U_1 . 0 + U_2 + 2U_3 = 0$

Solving (i) & (ii)

$$\begin{array}{cccc} U_1 & U_2 & U_3 & \\ 1 & 0 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{array}$$

$$\frac{U_1}{1} = \frac{U_2}{-2} = \frac{U_3}{1}$$

$U_1 = 1, \; U_2 = -2, \; U_3 = 1$

$$U_3 = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

length of $U_3 = \sqrt{1^2 + (-2)^2 + 1^2}$
$= \sqrt{6}$

Normalize U_3, $N(U_3) = \begin{bmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$

for $\lambda = 2$ eq ① becomes

$0.U_1 + U_2 + 0.U_3 = 0$ — (i)

$U_1 - U_2 + U_3 = 0$ — (ii)

$0.U_1 + U_2 + 0.U_3 = 0$

Solving (i) & (ii)

$$\begin{array}{cccc} U_1 & U_2 & U_3 & \\ 1 & 0 & 0 & 1 \\ -1 & 1 & 1 & -1 \end{array}$$

$$\frac{U_1}{1} = \frac{U_2}{0} = \frac{U_3}{-1}$$

$U_1 = 1, \; U_2 = 0, \; U_3 = -1$

$$U_2 = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

length of $U_2 = \sqrt{1^2 + 0^2 + (-1)^2}$
$= \sqrt{2}$

Normalize U_2, $N(U_2) = \begin{bmatrix} 1/\sqrt{2} \\ 0/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

$\therefore U = [N(U_1) \; N(U_2) \; N(U_3)]$

$$= \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0/\sqrt{2} & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}$$

$\therefore A = U \Sigma V^T$

$$A = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

2) obtain the singular value decomposition (SVD) of A, $U \Sigma V^T$ where $A = \begin{bmatrix} 0 & -2 \\ 1 & 5 \end{bmatrix}$

3) obtain the SVD of A, $U \Sigma V^T$, where $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$

2) Given matrix $A = \begin{bmatrix} 0 & -2 \\ 1 & 5 \end{bmatrix}$

let $A = U \Sigma V^T$ be SVD of A

Step-1 :- compute V

consider the matrix $A^T A = \begin{bmatrix} 0 & 1 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} 0 & -2 \\ 1 & 5 \end{bmatrix}$

$$= \begin{bmatrix} 0+1 & 0+5 \\ 0+5 & 4+25 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 5 \\ 5 & 29 \end{bmatrix}$$

We have to find Eigen values & Eigen vectors A^TA

characteristic equation is $|A^TA - \lambda I| = 0$

$$\begin{vmatrix} 1-\lambda & 5 \\ 5 & 29-\lambda \end{vmatrix} = 0$$

$$(1-\lambda)(29-\lambda) - 25 = 0$$

$$29 - \lambda - 29\lambda + \lambda^2 - 25 = 0$$

$$\lambda^2 - 30\lambda + 4 = 0$$

$$\lambda_1 = 29.8 \qquad \lambda_2 = 0.133$$

$\lambda_1 = 29.8, 0.133$ are the Eigen values of A^TA

characteristic equation of A^TA is $|A^TA - \lambda I| V = 0$

$$\begin{pmatrix} 1-\lambda & 5 \\ 5 & 29-\lambda \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{—} ①$$

for $\lambda = 29.8$ Eq ① becomes

$$\begin{pmatrix} -28.8 & 5 \\ 5 & -0.8 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$-28.8V_1 + 5V_2 = 0 \quad \text{—} ②$$

$$5V_1 - 0.8V_2 = 0 \quad \text{—} ③$$

$$② \Rightarrow -28.8V_1 + 5V_2 = 0$$

$$-28.8V_1 = -5V_2 \Rightarrow \frac{V_1}{5} = \frac{V_2}{28.8} = k$$

$$V_1 = 5k, \quad V_2 = 28.8k$$

for $k=1$ then $V_1 = 5, \; V_2 = 28.8$

$$V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{bmatrix} 5 \\ 28.8 \end{bmatrix}$$

length of $V_1 = \|V_1\| = \sqrt{5^2 + (28.8)^2} = 29.23$

Normalised vector $N(V_1) = \begin{bmatrix} 5/29.23 \\ 28.8/29.23 \end{bmatrix} = \begin{bmatrix} 0.171 \\ 0.983 \end{bmatrix}$

for $\lambda = 0.133$ eq ① becomes

$$\begin{bmatrix} 0.867 & 5 \\ 5 & 28.86 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$0.867 V_1 + 5V_2 = 0$ —— ④

$5V_1 + 28.86 V_2 = 0$ —— ⑤

④ $\Rightarrow 0.867 V_1 + 5V_2 = 0$

$0.867 V_1 = -5V_2$

$\frac{V_1}{-5} = \frac{V_2}{0.867}$ if $V_1 = -5k$ and $V_2 = +0.867$

if $k=1$ then $V_1 = 5$ $V_2 = 0.867$.

$$V_2 = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} -5 \\ 0.867 \end{bmatrix}$$

length of $V_2 = \|V_2\| = \sqrt{(-5)^2 + (0.867)^2} = 5.074$

Normalised vector $N(V_2) = \begin{bmatrix} -5/5.074 \\ 0.867/5.074 \end{bmatrix} = \begin{bmatrix} -0.985 \\ 0.170 \end{bmatrix}$

$$V = \begin{bmatrix} N(V_1) & N(V_2) \end{bmatrix} = \begin{bmatrix} 0.171 & -0.985 \\ 0.985 & 0.170 \end{bmatrix}$$

$$V^T = \begin{bmatrix} 0.171 & 0.985 \\ -0.985 & 0.170 \end{bmatrix}$$

Step 2 :- compute Σ

order of Σ = order of A

Now $\sigma_1 = \sqrt{\lambda_1} = \sqrt{29.8}$, $\sigma_2 = \sqrt{\lambda_2} = \sqrt{0.133}$

No. of non-zero Eigen values = Rank

Rank = 2

$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$, Σ_1 is a diagonal matrix whose diagonal entries are σ_1 & σ_2

$\Sigma = R \times R = 2 \times 2$

$$\Sigma = \begin{bmatrix} \sqrt{29.8} & 0 \\ 0 & \sqrt{0.133} \end{bmatrix}_{2\times 2}$$

step 3 : compute U

consider the matrix $AA^T = \begin{bmatrix} 0 & -2 \\ 1 & 5 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -2 & 5 \end{bmatrix} = \begin{bmatrix} 4 & -10 \\ -10 & 26 \end{bmatrix}$

We have find the Eigen values and Eigen vectors of AA^T

characteristic eqn of AA^T is $|AA^T - \lambda I| = 0$

$$\begin{vmatrix} 4-\lambda & -10 \\ -10 & 26-\lambda \end{vmatrix} = 0$$

$$(4-\lambda)(26-\lambda) - 100 = 0$$

$$104 - 4\lambda - 26\lambda + \lambda^2 - 100 = 0$$

$$\lambda^2 - 30\lambda + 4 = 0$$

$$\lambda_1 = 29.8 \quad \lambda_2 = 0.133$$

$\lambda_1 = 29.8, 0.133$ are the Eigen values of A^TA

characteristic eq of A^TA is $(A^TA - \lambda I)v = 0$

$$\begin{pmatrix} 4-\lambda & -10 \\ -10 & 26-\lambda \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow (1)$$

for $\lambda = 29.8$ eq (1) becomes

$$\begin{pmatrix} -25.8 & -10 \\ -10 & -3.8 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$-25.8u_1 - 10u_2 = 0 \quad \text{—} (5)$$

$$-10u_1 - 3.8u_2 = 0 \quad \text{—} (6)$$

$(5) \Rightarrow \quad -25.8u_1 = 10u_2$

$$\frac{u_1}{10} = \frac{u_2}{-25.8} = k$$

$$u_1 = 10, \ u_2 = -25.8$$

$$u_1 = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 10 \\ 25.8 \end{pmatrix}$$

length of $u_1 = \|u_1\| = \sqrt{(10)^2 + (-25.8)^2} = 27.6$

Normalised vector $N(u_1) = \begin{bmatrix} 10/27.6 \\ -25.8/2 \end{bmatrix} = \begin{bmatrix} 0.362 \\ -0.93 \end{bmatrix}$

for $\lambda = 0.133$ eq (1) becomes

$$\begin{pmatrix} 3.867 & -10 \\ -10 & 25.86 \end{pmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$3.867u_1 - 10u_2 = 0 \rightarrow (7)$$

$$-10u_1 + 25.86u_2 = 0 \quad \text{—} (3)$$

$(7) \Rightarrow \quad 3.867u_1 - 10u_2 = 0$

$$3.867u_1 = 10u_2$$

$$\frac{u_1}{10} = \frac{u_2}{3.867} = k$$

if $k=1$ then $u_1 = 10 \quad u_2 = 3.867$

$$u_2 = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 10 \\ 3.867 \end{bmatrix}$$

length of $u_2 = \|u_2\| = \sqrt{(10)^2 + (3.867)^2} = 10.72$

Normalised vector $N(u_2) = \begin{bmatrix} 10/10.72 \\ 3.867/10.72 \end{bmatrix} = \begin{bmatrix} 0.932 \\ 0.360 \end{bmatrix}$

$$U = [N[u_1] \ N[u_2]] = \begin{bmatrix} 0.362 & 0.932 \\ -12.9 & 0.360 \end{bmatrix}$$

$$A = \begin{bmatrix} 0.362 & 0.932 \\ -0.93 & 0.360 \end{bmatrix} \begin{bmatrix} \sqrt{29.8} & 0 \\ 0 & \sqrt{0.133} \end{bmatrix} \begin{bmatrix} 0.171 & 0.985 \\ -0.985 & 0.170 \end{bmatrix}$$

3) Let $A = U \Sigma V^T$ be SVD of A. Given $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$

Step 1 : compute V

$$A^T = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix}_{3\times 2}$$

consider the matrix $A^TA = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} = \begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix}$

We have to find eigen values and eigen vectors of A^TA

characteristic eqn of $|A^TA - \lambda I|V = 0$

$$\begin{vmatrix} 13-\lambda & 12 & 2 \\ 12 & 13-\lambda & -2 \\ 2 & -2 & 8-\lambda \end{vmatrix} = 0 \longrightarrow (1)$$

$$(13-\lambda)\left[(13-\lambda)(8-\lambda)-4\right] - 12\left[12(8-\lambda)-(4)\right] + 2\left[12(-2)-2(13-\lambda)\right] = 0$$

$$(13-\lambda)\left[104-13\lambda-8\lambda+\lambda^2-4\right] - 12\left[96-12\lambda+4\right] + 2\left[-24-36+2\lambda\right] = 0$$

$$(13-\lambda)\left[\lambda^2-21\lambda+100\right] - 12\left[-12\lambda+100\right) + 2\left[2\lambda-50\right] = 0$$

$$13\lambda^2 - 273\lambda + 1300 - \lambda^3 + 21\lambda^2 - 100\lambda + 144\lambda - 1200 + 4\lambda - 100 = 0$$

$$-\lambda^3 + 34\lambda^2 - 225\lambda = 0$$

$$-\lambda(\lambda^2 - 34\lambda + 225) = 0$$

$$\lambda = 0 \quad \lambda^2 - 34\lambda + 225 = 0$$

$$\lambda = 0 \quad \lambda = 9, \ \lambda = 25$$

$\lambda = 25, 9, 0$ are the Eigen values of A^TA

characteristic eq of A^TA is $|A^TA - \lambda I| V = 0$

for $\lambda = 0$ Then eq (1) becomes

$$\begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$13V_1 + 12V_2 + 2V_3 = 0$$

$$12V_1 + 13V_2 - 2V_3 = 0$$

$$\begin{array}{cccc} V_1 & V_2 & V_3 & \\ 12 & 2 & 13 & 12 \\ 13 & -2 & 12 & 13 \end{array}$$

$$\frac{V_1}{-24-26} = \frac{V_2}{24+26} = \frac{V_3}{169-144} = K$$

$$\frac{V_1}{\cancel{-50}_{\,2}} = \frac{V_2}{\cancel{50}_{\,2}} = \frac{V_3}{\cancel{25}_{\,1}} = K$$

$V_1 = -2, \ V_2 = 2, \ V_3 = 1$

$$V_1 = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}$$

length of $V_1 = \sqrt{(-2)^2 + 2^2 + 1} = \sqrt{9} = 3$

Normalized vector $N(V_1) = \begin{bmatrix} -2/3 \\ 2/3 \\ 1/3 \end{bmatrix}$

for $\lambda = 25$ then eq (1) becomes

$$\begin{bmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$-12V_1 + 12V_2 + 2V_3 = 0 \longrightarrow (i)$$

$$12V_1 - 12V_2 - 2V_3 = 0$$

$$\begin{array}{cccc} V_1 & V_2 & V_3 & \\ 12 & 2 & -12 & 12 \\ -12 & -2 & 12 & -12 \end{array}$$

$$\frac{V_1}{-24+24} = \frac{V_2}{24-24} = \frac{V_3}{144-144}$$

$V_1 = 0, \ V_2 = 0, \ V_3 = 0$

length of $V_2 = ||V_2|| = \sqrt{0+0+0} = 0$

Normalized vector $N(V_2) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

for $\lambda = 9$ then Eq ① becomes

$$\begin{bmatrix} 4 & 12 & 2 \\ 12 & 4 & -2 \\ 2 & -2 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$4v_1 + 12v_2 + 2v_3 = 0$

$12v_1 + 4v_2 - 2v_3 = 0$

$$\begin{array}{cccc} v_1 & v_2 & v_3 & \\ 12 & 2 & 4 & 12 \\ 4 & -2 & 12 & 4 \end{array}$$

$$\frac{v_1}{-24-8} = \frac{v_2}{24+8} = \frac{v_3}{16-144}$$

$$\frac{v_1}{-3\cancel{2}} = \frac{v_2}{3\cancel{2}} = \frac{v_3}{-1\cancel{28}}$$

$v_1 = -1, \quad v_2 = 1, \quad v_3 = -4$

$$v_3 = \begin{bmatrix} -1 \\ 1 \\ -4 \end{bmatrix}$$

length of $v_3 = ||v_3|| = \sqrt{(-1)^2 + 1^2 + (-4)^2} = \sqrt{1+1+16} = \sqrt{18}$
$= 4.24$

Normalized vector $N(v_3) = \begin{bmatrix} -1/4.24 \\ 1/4.24 \\ -4/4.24 \end{bmatrix}$

$$N[v_1], N[v_2] \,\&\, N(v_3) = \begin{bmatrix} 2/3 & 0 & -1/4.24 \\ 2/3 & 0 & 1/4.24 \\ 1/3 & 0 & -4/4.24 \end{bmatrix}$$

Step 2 compute Σ

$\sigma_1 = \sqrt{\lambda_1} = \sqrt{25} = 5$

$\sigma_2 = \sqrt{\lambda_2} = \sqrt{9} = 3$

order of Σ = order of A

no get order 2×3 we add 0 column in Σ

$$\Sigma = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \end{bmatrix}_{2\times 3}$$

Step 3 compute U

compute the matrix $AA^T = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$

characteristic eq of AA^T is $|AA^T - \lambda I| = 0$

$$\begin{vmatrix} 17-\lambda & 8 \\ 8 & 17-\lambda \end{vmatrix} = 0 \Rightarrow (17-\lambda)(17-\lambda) - 64 = 0$$

$$\lambda_1 = 25, \lambda_2 = 9$$

$\lambda_1 = 9, 25$ are the Eigen values.

characteristic eq of A^TA is $(A^TA - \lambda I)v = 0$

$$\begin{pmatrix} 17-\lambda & 8 \\ 8 & 17-\lambda \end{pmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow (1)$$

for $\lambda = 9$ Eq (1) becomes

$$\begin{bmatrix} 8 & 8 \\ 8 & 8 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$8v_1 + 8v_2 = 0 \rightarrow (2)$

$8v_1 + 8v_2 = 0 \rightarrow (3)$

$8v_1 = -8v_2$

$\frac{v_1}{-8} = \frac{v_2}{8} = k$

$v_1 = -8, v_2 = 8$

length of $v_1 = \sqrt{(-8)^2 + 8^2}$

$= 11.3$

Normalized $N(v_1) = \begin{bmatrix} -8/11.3 \\ 8/11.3 \end{bmatrix}$

for $\lambda = 25$ then eq (1) becomes

$$\begin{bmatrix} -8 & 8 \\ 8 & -8 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$-8v_1 + 8v_2 = 0 \rightarrow (4)$

$8v_1 + -8v_2 = 0 \rightarrow (5)$

$-8v_1 = -8v_2$

$\frac{v_1}{-8} = \frac{v_2}{-8} = k$

$v_1 = -8, v_2 = -8 \quad v_2 = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} -8 \\ -8 \end{bmatrix}$

length of $v_2 = \sqrt{(-8)^2 + (-8)^2} = 11.3$

Normalized $N(v_2) = \begin{bmatrix} -8/11.3 \\ -8/11.3 \end{bmatrix}$

$A = U \Sigma V^T$

$$= \begin{bmatrix} -8/11.3 & -8/11.3 \\ 8/11.3 & -8/11.3 \end{bmatrix}\begin{bmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \end{bmatrix}\begin{bmatrix} -2/3 & 2/3 & 1/3 \\ 0 & 0 & 0 \\ -1/4.24 & 1/4.24 & -4/4.24 \end{bmatrix}$$

W_1, W_2, W_3. The desired distance will be $|w_3|$

$v_1 = (1,-1,1,-1)$, $v_2 = (0,2,2,0)$ as $z - x_0 = |-1,0,1,0|$

$W_1 = v_1 = (1,-1,1,-1)$

$W_2 = V_2 - \frac{\langle V_2, W_1 \rangle}{\langle W_1, W_1 \rangle} W_1 = V_2 = (0,2,2,0)$ as $V_2 \perp V_1$

$$W_3 = (z - x_0) - \frac{(z - x_0, w_1)}{\langle w_1, w_1 \rangle} w_1 - \frac{\langle z - x_0, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2$$

$= (-1,0,1,0) - \frac{0}{4}(1,-1,1,-1) - \frac{2}{8}(0,2,2,0)$

$= (-1, -1/2, 1/2, 0)$

$|w_3| = |(-1, -1/2, 1/2, 0)| = 1/2 \, |(-2,-1,1,0)|$

$= \frac{\sqrt{6}}{2} = \frac{\sqrt{3}}{2}$

Inner product Space:-

let $V(F)$ be a vector space where F is the field of real numbers or the field of complex number. The vector Space $V(F)$ is said to be an inner product Space if there is defined for any two vectors $\alpha, \beta \in V$, An element $\langle \alpha, \beta \rangle \in F$ Such that

1) $\langle \alpha, \beta \rangle = \overline{\langle \beta, \alpha \rangle}$ (conjugate Symmetry)

2) $\langle \alpha, \alpha \rangle > 0$ (zero element in F) for $\alpha \neq \bar{0}$

3) $\langle a\alpha + b\beta, \gamma \rangle = \langle a\alpha, \gamma \rangle + \langle b\beta, \gamma \rangle$
$= a\langle \alpha, \gamma \rangle + b\langle \beta, \gamma \rangle$ (Distributive law)

for any $\alpha, \beta, \gamma \in V$ & $a, b \in F$.

* If $F = R$, the field of real numbers then $V(F)$ is called Euclidean Space (Real Inner product Space)
* If $F = C$, the field of complex numbers then $V(F)$ is called unitary space (or) Complex inner product space.

Norm or length of a vector:-

let V be an inner product Space over the field F. The norm (length) where $\alpha \in V$ denoted by

$$V_3 = X_3 - \frac{\langle X_3, V_1 \rangle}{\langle V_1, V_1 \rangle} V_1 - \frac{\langle X_3, V_2 \rangle}{\langle V_2, V_2 \rangle} V_2$$

$$= (-3, 7, 1, 3) - \frac{-12}{4}(1, -1, 1, -1) - \frac{16}{8}(0, 2, 2, 0)$$

$$= (0, 0, 0, 0)$$

The Gram-Schmidt process can be used to check linear Independence of vectors!

The vectors x_3 is a linear combination of x_1 and x_2. π is a plane, not a 3-dimensional Subspace

We should orthogonalize vectors: x_1, x_2, y

$$V_4 = y - \frac{\langle y, V_1 \rangle}{\langle V_1, V_1 \rangle} V_1 - \frac{\langle y, V_2 \rangle}{\langle V_2, V_2 \rangle} V_2$$

$$= (0, 0, 0, 1) - -1/4 (1, -1, 1, -1) - 0/8 (0, 2, 2, 0)$$

$$= (1/4, -1/4, 1/4, 3/4)$$

$$|V_4| = |(1/4, -1/4, 1/4, 3/4)| = 1/4 |(1, -1, 1, 3)|$$

$$= \frac{\sqrt{12}}{4} = \sqrt{3}/2$$

Problem :- find the distance from the point $z = (0, 0, 1, 0)$ to the plane π that passes through the points to $x_0 = (1, 0, 0, 0)$ and i parallel to the vectors $v_1 = (1, -1, 1, -1)$ and $v_2 = (0, 2, 2, 0)$

The plane π is not a subspace of R^4 as it does not pass through the origin. Let $\pi_0 = \text{Span}(v_1, v_2)$

Then $\pi = \pi_0 + x_0$

Hence the distance from the point z to the plane π is the Same as the distance from the points $z - x_0$ to the plane π_0

We shall Apply the Gram-Schmidt process to vectors $v_1, v_2, z - x_0$ This will yield an orthogonal System.

Now $v_1 = (1,2,2)$, $v_2 = (-4/3, -2/3, 4/3)$

$v_3 = (2/9, -2/9, 1/9)$ is an orthogonal basis for R^3 while v_1, v_2 is an orthogonal basis for π it remains to normalize these vectors

$\langle v_1, v_1 \rangle = 9 \Rightarrow \|v_1\| = 3$

$\langle v_2, v_2 \rangle = 4 \Rightarrow \|v_2\| = 2$

$\langle v_3, v_3 \rangle = 1/9 \Rightarrow \|v_3\| = 1/3$

$w_1 = v_1 / \|v_1\| = (1/3, 2/3, 2/3) = 1/3(1,2,2)$

$w_2 = v_2 / \|v_2\| = (-2/3, -1/3, 2/3) = 1/3(-2,-1,2)$

$w_3 = v_3 / \|v_3\| = (2/3, -2/3, 1/3) = 1/3(2,-2,1)$

w_1, w_2 is an orthonormal basis for π w_1, w_2 is an orthonormal basis for R^3

problem:-

find the distance from the point $y = (0,0,0,1)$ to the subspace $\pi < R^4$ spanned by vectors $x_1 = (1,-1,1,-1)$, $x_2 = (1,1,3,-1)$ and $x_3 = (-3,7,1,3)$

Let us Apply The Gram-Schmidt process to vectors x_1, x_2, x_3, x_4. We should obtain An orthogonal system v_1, v_2, v_3, v_4.

The desired distance will be $|v_4|$

$x_1 = (1,-1,1,-1)$, $x_2 = (1,1,3,-1)$

$x_3 = (-3,7,1,3)$, $x_4 = (0,0,0,1)$

find the distance from the point $y = (0,0,0,7)$ to the subspace $\pi < R^4$ spanned by vectors

$x_1 = (1,-1,1,-1)$, $x_2 = (1,1,3,-1)$ and $x_3 = (-3,7,1,3)$

let us Apply The Gram-Schmidt process to vectors x_1, x_2, x_3, y we should obtain an orthogonal system v_1, v_2, v_3, v_4. The desired distance will be $|v_4|$

$x_1 = (1,-1,1,-1)$, $x_2 = (1,1,3,-1)$

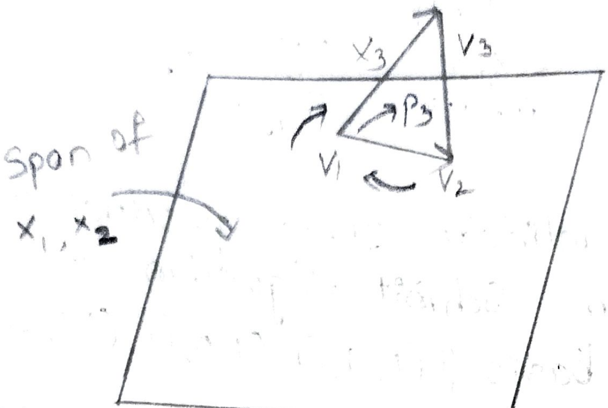
$x_3 = (-3,7,1,3)$, $y = (0,0,0,1)$

$v_1 = x_1 = (1,-1,1,-1)$

$v_2 = x_2 - \frac{\langle x_2, v_1 \rangle}{\langle v_1, v_1 \rangle} = (1,1,3,-1) - \frac{4}{4}(1,-1,1,-1) - (0,2,2,0)$

projection of the vector X_k on the Sub Space Spanned by $x_1, \dots X_{k-1}$;

* $\|V_k\|$ is the distance from X_k to the Sub Space Spanned by $x_1, \dots X_{k-1}$



Problem :- let π be the plane in R^3 Spanned by vectors $x_1 = (1, 2, 2)$ and $x_2 (-1, 0, 2)$

i) find an orthonormal basis for R^3

ii) Extend it to an orthonormal basis for R^3.

x_1, x_2 is a basis for the plane π. we can extend it to a basis for R^3 by adding one vector from the standard bias. for the instance, vectors x_1, x_2 and $x_3 = (0, 0, 1)$ form a basis for R^3 because

$$\begin{vmatrix} 1 & 2 & 2 \\ -1 & 0 & 2 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ -1 & 0 \end{vmatrix} = 2 \neq 0$$

Using The Gram - Schmidt process, we orthogonalize the basis $x_1 = (1, 2, 2)$, $x_2 = (-1, 0, 2)$ $x_3 = (0, 0, 1)$

$$V_1 = x_1 = (1, 2, 2)$$

$$V_2 = X_2 - \frac{\langle X_2, V_1 \rangle}{\langle V_1, V_1 \rangle} V_1 = (-1, 0, 2) - \frac{3}{9} (1, 2, 2)$$

$$= (-4/3, -2/3, 4/3)$$

$$V_3 = X_3 - \frac{\langle X_3, V_1 \rangle}{\langle V_1, V_1 \rangle} V_1 - \frac{\langle X_3, V_2 \rangle}{\langle V_2, V_2 \rangle} V_2$$

$$= (0, 0, 1) - \frac{2}{9} (1, 2, 2) - \frac{4/3}{4} (-4/3, -2/3, 4/3)$$

$$= (2/9, -2/9, 1/9)$$