UNIT-4

Linear Algebra, mean vector and correlation and precision matrices, Mahalanobis distance and multivariate normal distribution, Principal component analysis and finding the number of principal components, loadings and plots of principal components, Identification of key dimensions in multidimensional scaling.

Covariance Matrix

Covariance matrix is a type of matrix that is used to represent the covariance values between pairs of elements given in a random vector.

Variance covariance matrix is defined as a square matrix where the diagonal elements represent the variance and the off-diagonal elements represent the covariance. The covariance between two variables can be positive, negative, and zero. A positive covariance indicates that the two variables have a positive relationship whereas negative covariance shows that they have a negative relationship. If two elements do not vary together then they will display a zero covariance.

$$\begin{bmatrix} Var(x_1) & \cdots & Cov(x_n, x_1) \\ \vdots & \ddots & \vdots \\ Cov(x_n, x_1) & \cdots & Var(x_n) \end{bmatrix}$$

To determine the covariance matrix, the formulas for variance and covariance are required.

Population Variance: $var(x) = \frac{\sum_1^n (x_i - \mu)^2}{n}$

Population Covariance: $cov(x, y) = \frac{\sum_1^n (x_i - \mu_x)(y_i - \mu_y)}{n}$

Sample Variance: $var(x) = \frac{\sum_1^n (x_i - \bar{x})^2}{n-1}$

Sample Covariance: $cov(x, y) = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$

μ = mean of population data.

$\bar{x}$ = mean of sample data.

n = number of observations in the dataset.

x_i = observations in dataset x.

Covariance Matrix 2 × 2

A 2 × 2 matrix is one which has 2 rows and 2 columns. The formula for a 2 × 2 covariance matrix is given as follows:

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

Covariance Matrix 3 × 3

If there are 3 datasets, x, y, and z, then the formula to find the 3 × 3 covariance matrix is given below:

$$\begin{bmatrix} var(x) & cov(x,y) & cov(x,z) \\ cov(x,y) & var(y) & cov(y,z) \\ cov(x,z) & cov(y,z) & var(z) \end{bmatrix}$$

The number of variables determines the dimension of a variance-covariance matrix. For example, if there are two variables (or datasets) it indicates that the covariance matrix will be 2 dimensional.

Student	Math (X)	Science (Y)
1	92	80
2	60	30
3	100	70

The steps to calculate the covariance matrix for the sample are given below:

- **Step 1:** Find the mean of one variable (X). This can be done by dividing the sum of all observations by the number of observations. Thus, (92 + 60 + 100) / 3 = 84
- **Step 2:** Subtract the mean from all observations; (92 - 84), (60 - 84), (100 - 84)
- **Step 3:** Take the sum of the squares of the differences obtained in the previous step. $(92 - 84)^2 + (60 - 84)^2 + (100 - 84)^2$.
- **Step 4:** Divide this value by 1 less than the total to get the sample variance of the first variable (X). var(X) = $[(92 - 84)^2 + (60 - 84)^2 + (100 - 84)^2] / (3 - 1) = 448$
- **Step 5:** Repeat steps 1 to 4 to find the variances of all variables. Using these steps, var(Y) = 700.
- **Step 6:** Choose a pair of variables (X and Y).

- **Step 7:** Subtract the mean of the first variable (X) from all observations; (92 - 84), (60 - 84), (100 - 84).
- **Step 8:** Repeat step 7 for the second variable (Y); (80 - 60), (30 - 60), (70 - 60).
- **Step 9:** Multiply the corresponding observations. (92 - 84)(80 - 60), (60 - 84)(30 - 60), (100 - 84)(70 - 60).
- **Step 10:** Add these values and divide them by (n - 1) to get the covariance.
  cov(x, y) = cov(y, x) = [(92 - 84)(80 - 60) + (60 - 84)(30 - 60) + (100 - 84)(70 - 60)] / (3 - 1) = 520.
- **Step 11:** Repeat steps 6 to 10 for different pairs of variables.
- **Step 12:** Now using the general formula for covariance matrix arrange these values in matrix form. Thus, the variance covariance matrix for the example is given as $\begin{bmatrix} 448 & 520 \\ 520 & 700 \end{bmatrix}$.

The same steps can be followed while calculating the covariance matrix for a population. The only difference is that the population variance and covariance formulas will be applied.

NOTE:

The mean vector consists of the means of each variable and the variance-covariance matrix consists of the variances of the variables along the main diagonal and the covariances between each pair of variables in the other matrix positions.

The formula for computing the covariance of the variables X and Y is

$$\text{COV} = \frac{\sum_{i=1}^{n}(X_i - \bar{x})(Y_i - \bar{y})}{n-1},$$

with $\bar{x}$ and $\bar{y}$ denoting the means of X and Y, respectively.

variance-covariance matrix S is calculated by

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'$$

precision matrix:

Let X be multivariate normal with covariance matrix Σ.

The precision matrix, Ω, is simply defined to be the inverse of the covariance matrix:

$$\Omega := \Sigma^{-1}$$

NOTE:

Univariate data –

This type of data consists of **only one variable**.

Bivariate data –

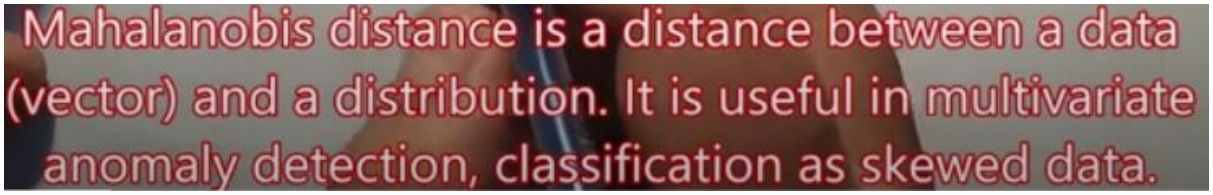
This type of data involves **two different variables**.

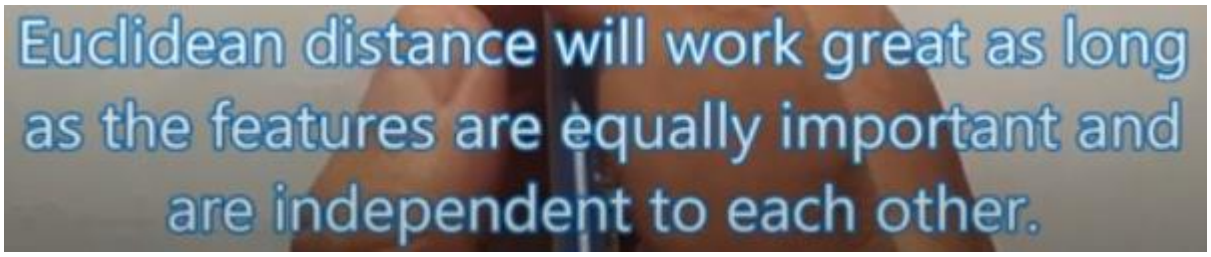
Multivariate data –

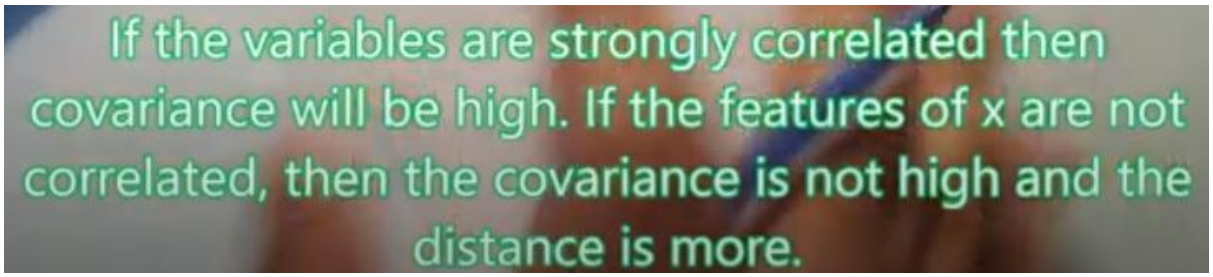
When the data involves **three or more variables**, it is categorized under multivariate.

Mahalanobis distance:

http://data-science.tokyo/ed-e/ede1-2-2-4-2.html







Euclidean distance is the commonly used straight line distance between two points. Euclidean distance will work fine as long as the dimensions are equally weighted and are independent of each other.

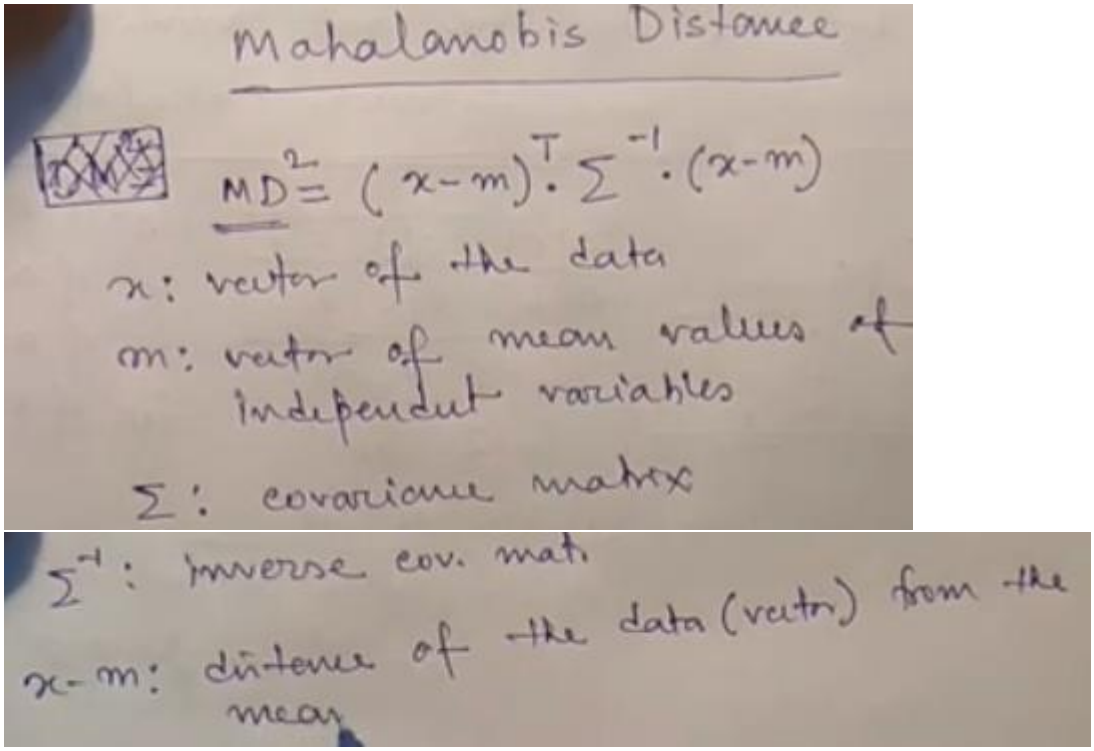
if the dimensions (columns in your dataset) are correlated to one another, which is typically the case in real-world datasets, the Euclidean distance between a point and the center of the points (distribution) can give little or misleading information about how close a point really is to the cluster.

Euclidean distance is a distance between two points only. It does not consider how the rest of the points in the dataset vary.

So, it cannot be used to really judge how close a point actually is to a distribution of points. What we need here is a more robust distance metric that is an accurate representation of how distant a point is from a *distribution*.

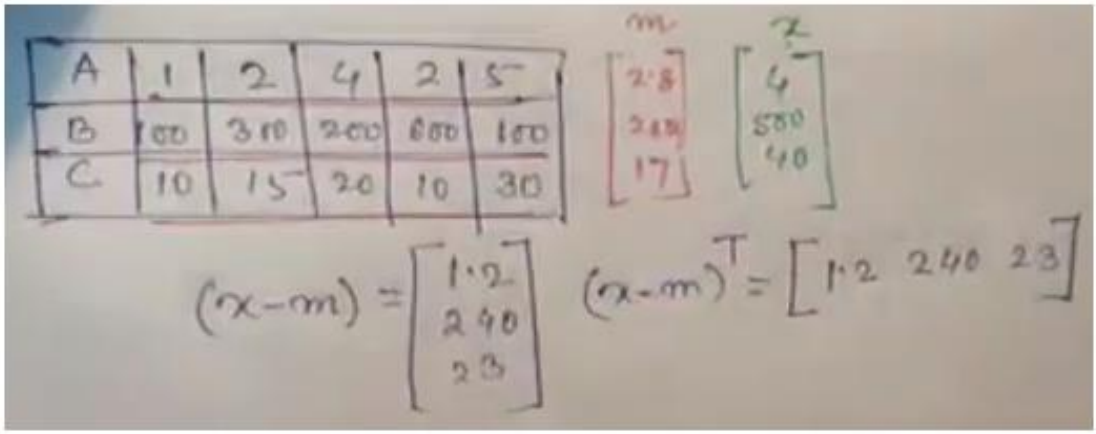
Mahalonobis distance is the distance between a point and a distribution. And not between two distinct points. It is effectively a multivariate equivalent of the Euclidean distance.

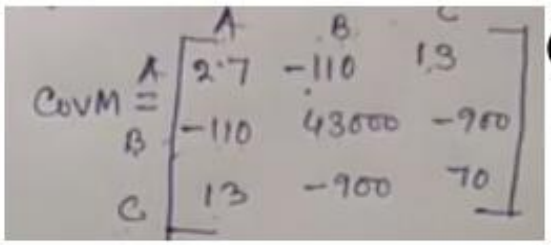
Mahalanobis distance is an effective multivariate distance metric that measures the distance between a point (vector) and a distribution.

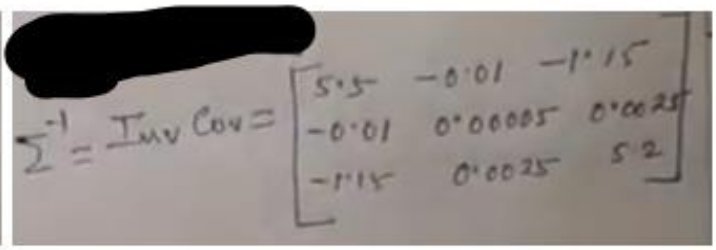


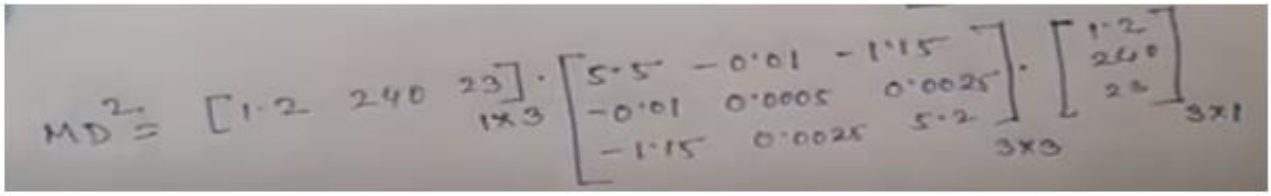
Example:

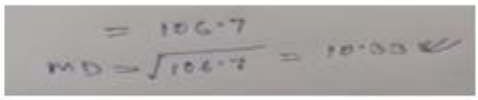
Consider the following table and data point x, find the distance between x and data.











Eigenvalues and Eigenvectors

Definition: Matrix

A system of *mn* numbers(elements) arranged in a rectangular arrangement along *m* rows and *n* columns and bounded by the brackets [ ] or ( ) is called an m by n matrix, which is written as $m \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & .. & .. & a_{1n} \\ a_{21} & a_{22} & .. & .. & a_{2n} \\ ... & .. & .. & .. & .. \\ .. & .. & .. & .. & .. \\ a_{m1} & a_{m2} & .. & .. & a_{mn} \end{bmatrix}$$

Characteristic polynomial

The determinant $|A-\lambda I|$ when expanded will give a polynomial, which we call as characteristic polynomial of matrix A.

Definition: Eigenvalues

$A = [a_{ij}]$ be a square matrix.

The characteristic equation of A is $|A-\lambda I| = 0$.

The roots of the characteristic equation are called Eigenvalues of A.

Definition: Eigenvectors

A = [a_{ij}] be a square matrix of order 'n'

If there exist a non zero vector X = $\begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{bmatrix}$

such that AX = λX , then the vector X is called an Eigenvector of A corresponding to the Eigenvalue λ .

Method of finding characteristic equation of a 3x3 matrix and 2x2 matrix

The characteristic equation of a 3x3 matrix is $\lambda^3 - S_1\lambda^2 + S_2\lambda - S_3 = 0$

Where, S1= sum of main diagonal elements.

S_2 = sum of minor of main diagonal elements.

S_3 = Det (A) =|A|

The characteristic equation of a 2x2 matrix is $\lambda^2 - S_1\lambda + S_2 = 0$

Where, S_1 = sum of main diagonal elements.

S_2 = Det (A) =|A|

Example 1: Find the eigenvalues and eigenvectors of the following matrix.

Solution:

$$A = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}, f_A(x) = |xI - A| = (x-2)(x-3) - 2 = x^2 - 5x + 4, \lambda_1, \lambda_2 = [5 \pm \sqrt{(25-16]}]/2 = 4,1.$$

$$[A - 4I \mid 0] = \begin{pmatrix} -2 & 2 & 0 \\ 1 & -1 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{is an eigen - vector for e.v. } \lambda_1 = 4.$$

$$[A - I \mid 0] = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 2 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow v_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix} \text{is an eigen - vector for e.v. } \lambda_2 = 1.$$

Example 2: Find all eigenvalues and corresponding eigenvectors for the matrix A if

$$\begin{pmatrix} 2 & -3 & 0 \\ 2 & -5 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Solution:

$$\det\left(\begin{pmatrix} 2 & -3 & 0 \\ 2 & -5 & 0 \\ 0 & 0 & 3 \end{pmatrix} - \lambda\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

$$\begin{pmatrix} 2 & -3 & 0 \\ 2 & -5 & 0 \\ 0 & 0 & 3 \end{pmatrix} - \lambda\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\lambda\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$$

$$= \begin{pmatrix} 2 & -3 & 0 \\ 2 & -5 & 0 \\ 0 & 0 & 3 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$$

$$= \begin{pmatrix} 2-\lambda & -3 & 0 \\ 2 & -5-\lambda & 0 \\ 0 & 0 & 3-\lambda \end{pmatrix}$$

$$= \det\begin{pmatrix} 2-\lambda & -3 & 0 \\ 2 & -5-\lambda & 0 \\ 0 & 0 & 3-\lambda \end{pmatrix}$$

$$= (2-\lambda)\det\begin{pmatrix} -5-\lambda & 0 \\ 0 & 3-\lambda \end{pmatrix} - (-3)\det\begin{pmatrix} 2 & 0 \\ 0 & 3-\lambda \end{pmatrix} + 0\cdot\det\begin{pmatrix} 2 & -5-\lambda \\ 0 & 0 \end{pmatrix}$$

$$= (2-\lambda)\left(\lambda^2+2\lambda-15\right) - (-3)\cdot 2(-\lambda+3) + 0\cdot 0$$

$$= -\lambda^3+13\lambda-12$$

$$-\lambda^3+13\lambda-12=0$$

$$-(\lambda-1)(\lambda-3)(\lambda+4)=0$$

The eigenvalues are :

$\lambda=1,\ \lambda=3,\ \lambda=-4$

Eigenvectors for $\lambda=1$

$$\begin{pmatrix} 2 & -3 & 0 \\ 2 & -5 & 0 \\ 0 & 0 & 3 \end{pmatrix} - 1\cdot\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -3 & 0 \\ 2 & -6 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

$$(A-1I)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & -3 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{Bmatrix} x-3y=0 \\ z=0 \end{Bmatrix}$$

Isolate

$$\begin{Bmatrix} z=0 \\ x=3y \end{Bmatrix}$$

Plug into $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$

$$\eta = \begin{pmatrix} 3y \\ y \\ 0 \end{pmatrix} \quad y \neq 0$$

Let $y=1$

$$\begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}$$

Similarly

Eigenvectors for $\lambda=3$: $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

Eigenvectors for $\lambda=-4$: $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$

The eigenvectors for $\begin{pmatrix} 2 & -3 & 0 \\ 2 & -5 & 0 \\ 0 & 0 & 3 \end{pmatrix}$

$$= \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

3. Find the Eigenvalues of $\begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$

characteristic equation is $\lambda^3-4\lambda^2+3\lambda-0=0$

$\lambda(\lambda^2-4\lambda+3)=0$

$\lambda=0,\ (\lambda^2-4\lambda+3)=0$

$(\lambda-1)(\lambda-3)=0$

$\lambda=1, 3$

The Eigen values are 1, 3, and 0.

1. Find all the Eigenvalues and Eigenvectors of the matrix $\begin{pmatrix}1 & -1 & 4\\ 3 & 2 & -1\\ 2 & 1 & -1\end{pmatrix}$

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

To solve the characteristic equation:

If $\lambda=1$ By synthetic division

1	1	-2	-5	6
	0	1	-1	-6
	1	-1	-6	0

Therefore the $\lambda=1$ and other roots are given by $\lambda^2-\lambda-6=0$

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

$\lambda=-2, 3$

Therefore Eigenvalues are 1,-2, 3

To find the Eigenvectors:

To get the Eigenvectors solve: $(A-\lambda I)X=0$

$$\left[\begin{pmatrix}1 & -1 & 4\\ 3 & 2 & -1\\ 2 & 1 & -1\end{pmatrix}-\lambda\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\right]\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$

$$\left[\begin{pmatrix}1-\lambda & -1 & 4\\ 3 & 2-\lambda & -1\\ 2 & 1 & -1-\lambda\end{pmatrix}\right]\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$

$$\left.\begin{aligned}(1-\lambda)x_1-x_2+4x_3&=0\\ 3x_1+(2-\lambda)x_2-x_3&=0\\ 2x_1+x_2+(-1-\lambda)x_3&=0\end{aligned}\right\}\quad\ldots.(1)$$

Case 1: Substitute $\lambda=1$ in, (1) we get

$0x_1-x_2+4x_3=0$ .... (2)

$3x_1+x_2-x_3=0$ .... (3)

$2x_1+x_2-2x_3=0$ .... (4)

Solving (2) and (3) by cross multiplication rule, we get

x_1	x_2	x_3
-1	4	-1
1	-1	1

$$\frac{x_1}{1-4}=\frac{x_2}{12-0}=\frac{x_3}{0+3}$$

$$\Rightarrow\frac{x_1}{-3}=\frac{x_2}{12}=\frac{x_3}{3}$$

$$\Rightarrow\frac{x_1}{-1}=\frac{x_2}{4}=\frac{x_3}{1}$$

Therefore $X_1=\begin{pmatrix}-1\\ 4\\ 1\end{pmatrix}$

Case 2: Substitute $\lambda = -2$ in (1), we get

$3x_1-x_2+4x_3=0$ ....(5)

$3x_1+4x_2-x_3=0$ ....(6)

$2x_1+x_2+x_3=0$ ....(7)

Solving (5) and (6) by cross multiplication rule we get

$$\begin{array}{ccccc} & x_1 & & x_2 & & x_3 \\ -1 & & 4 & & 3 & & -1 \\ 4 & & -1 & & 3 & & 4 \end{array}$$

$$\frac{x_1}{1-16} = \frac{x_2}{12+3} = \frac{x_3}{12+3}$$

$$\Rightarrow \frac{x_1}{-15} = \frac{x_2}{15} = \frac{x_3}{15}$$

$$\Rightarrow \frac{x_1}{-1} = \frac{x_2}{1} = \frac{x_3}{1}$$

Therefore $X_2 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$

Case 3: Substitute $\lambda = 3$ in (1) we get

$-2x_1-x_2+4x_3=0$ .... (8)

$3x_1-x_2-x_3=0$ .... (9)

$2x_1+x_2-4x_3=0$ .... (10)

Solving (8) and (9) by cross multiplication rule we get

$$\begin{array}{ccccc} & x_1 & & x_2 & & x_3 \\ -1 & & 4 & & -2 & & -1 \\ -1 & & -1 & & 3 & & -1 \end{array}$$

$$\frac{x_1}{1+4} = \frac{x_2}{12-2} = \frac{x_3}{2+3}$$

$$\Rightarrow \frac{x_1}{5} = \frac{x_2}{10} = \frac{x_3}{5}$$

$$\Rightarrow \frac{x_1}{1} = \frac{x_2}{2} = \frac{x_3}{1}$$

Therefore $X_3 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$

Result: The Eigen values of A are 1,-2, 3 and the Eigenvectors are $\begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$

univariate normal distribution:

A random variable X is normally distributed with mean μ and variance σ^2 if it has the probability density function of X as: $\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$

As shorthand notation we may use the expression below: $X \sim N(\mu, \sigma^2)$

Multivariate Normal Distribution:

If we have a p x 1 random vector $\mathbf{X}$ that is distributed according to a multivariate normal distribution with a population mean vector μ and population variance-covariance matrix Σ, then this random vector, $\mathbf{X}$, will have the joint density function as shown in the expression below:

$$\phi(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{p/2} |\Sigma|^{-1/2} \exp\{-\tfrac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\}$$

$|\Sigma|$ denotes the determinant of the variance-covariance matrix Σ and $\Sigma^{-1}$is just the inverse of the variance-covariance matrix Σ. Again, this distribution will take maximum values when the vector $\mathbf{X}$ is equal to the mean vector μ, and decrease around that maximum.

The shorthand notation, similar to the univariate version is, $\mathbf{X} \sim N(\mu, \Sigma)$

We use the expression that the vector $\mathbf{X}$ 'is distributed as' multivariate normal with mean vector μ and variance-covariance matrix Σ.

A random vector $\mathbf{X} = (X_1, ..., X_n)$ has a multivariate Normal distribution with mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and (symmetric and positive-definite) covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$, written $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, if it has the following joint PDF:

$$f_{\mathbf{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \qquad \boldsymbol{x} \in \mathbb{R}^n$$

Principal component analysis:

https://www.analyticsvidhya.com/blog/2022/07/principal-component-analysis-beginner-friendly/

https://www.turing.com/kb/guide-to-principal-component-analysis

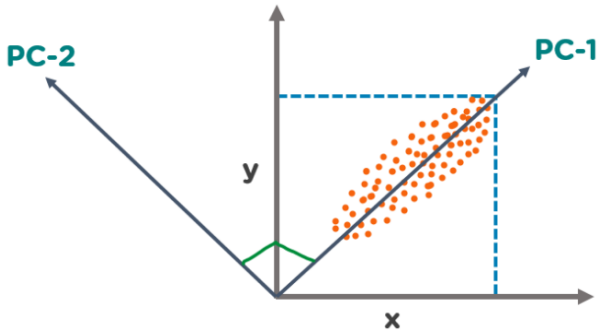
https://www.simplilearn.com/tutorials/machine-learning-tutorial/principal-component-analysis

The Principal Component Analysis is a popular unsupervised learning technique for reducing the dimensionality of large data sets. It increases interpretability yet, at the same time, it minimizes information loss. It helps to find the most significant features in a dataset and makes the data easy for plotting in 2D and 3D. PCA helps in finding a sequence of linear combinations of variables.

Dimensionality:

The term "dimensionality" describes the quantity of features or variables used in the analysis.

While reducing the number of variables in the dataset, dimensionality reduction methods like PCA are used to preserve the most crucial data. The original variables are converted into a new set of variables called principal components, which are linear combinations of the original variables, by PCA in order to accomplish this. The dataset's reduced dimensionality depends on how many principal components are used in the study. The objective of PCA is to select fewer principal components that account for the data's most important variation.



In the above figure, we have several points plotted on a 2-D plane. There are two principal components. PC1 is the primary principal component that explains the maximum variance in the data. PC2 is another principal component that is orthogonal to PC1.

Principal Component:

The Principal Components are a straight line that captures most of the variance of the data. They have a direction and magnitude. Principal components are orthogonal projections (perpendicular) of data onto lower-dimensional space.

Given a dataset with n observations and p variables represented by the n x p data matrix X, the goal of PCA is to transform the original variables into a new set of k variables called principal components that capture the most significant variation in the data. The principal components are defined as linear combinations of the original variables.

Steps for PCA Algorithm

1. Standardize the data: PCA requires standardized data, so the first step is to standardize the data to ensure that all variables have a mean of 0 and a standard deviation of 1.
2. Calculate the covariance matrix: The next step is to calculate the covariance matrix of the standardized data. This matrix shows how each variable is related to every other variable in the dataset.
3. Calculate the eigenvectors and eigenvalues: The eigenvectors and eigenvalues of the covariance matrix are then calculated. The eigenvectors represent the directions in which the data varies the most, while the eigenvalues represent the amount of variation along each eigenvector.
4. Choose the principal components: The principal components are the eigenvectors with the highest eigenvalues. These components represent the directions in which the data varies the most and are used to transform the original data into a lower-dimensional space.
5. Transform the data: The final step is to transform the original data into the lower-dimensional space defined by the principal components.

The steps involved for PCA are as follows-

1. Original Data
2. Normalize the original data (mean =0, variance =1)
3. Calculating covariance matrix
4. Calculating Eigen values, Eigen vectors, and normalized Eigenvectors
5. Calculating Principal Component (PC)
6. Plot the graph for orthogonality between PCs

Consider dataset,

Feature	sample1	sample2	sample3	sample4
a	4	8	13	7
b	11	4	5	14

Step1:

No. of features, n = 2 (a, b)

No. of samples, N = 4 (sample1, sample2, sample3, sample4)

Step2:

calculating mean,

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

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

Step3:

calculating covariance matrix, between features,

In the given dataset, ordered features are as,

(a,a), (a,b), (b,a), (b,b)

$$cov(a,a) = \frac{1}{N-1} \sum_{k=1}^{N} (a_i - \bar{a})(a_i - \bar{a})$$

$$= \frac{1}{N-1} \sum_{k=1}^{N} (a_i - \bar{a})^2 \rightarrow \text{for same feature}$$

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

$$= \frac{4^2 + 0 + 5^2 + 1^2}{3} = \frac{16+0+25+1}{3}$$

$$= \frac{42}{3} = 14.$$

$$cov(a,b) = \frac{1}{N-1} \sum_{k=1}^{N} (a_i - \bar{a})(b_i - \bar{b})$$

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

$$= \frac{1}{3} [(-4)(2.5) + (0) + (5)(-3.5) + (-1)(5.5)]$$

$$= \frac{1}{3} [-10 - 17.5 - 5.5] = \frac{-33}{3} = -11$$

$$cov(b,a) = \frac{1}{N-1}\sum_{K=1}^{N}(b_i - \bar{b})(a_i - \bar{a})$$

$$= cov(a,b)$$

$$= -11$$

$$cov(b,b) = \frac{1}{N-1}\sum_{K=1}^{N}(b_i - \bar{b})(b_i - \bar{b})$$

$$= \frac{1}{N-1}\sum_{K=1}^{N}(b_i - \bar{b})^2$$

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

$$= \frac{1}{3}\left[(2.5)^2 + (-4.5)^2 + (-3.5)^2 + (5.5)^2\right]$$

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

$$= \frac{69}{3} = \underline{\underline{23}}$$

Hence covariance matrix can be

$$S = \begin{bmatrix} cov(a,a) & cov(a,b) \\ cov(b,a) & cov(b,b) \end{bmatrix} = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$$

Step4:-

calculate Eigen value, Eigen vector, Normalized Eigen vector.

Inorder calculate Eigen value,

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

$$I(\text{Identity matrix}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\lambda I = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$\det\left(\begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = 0$$

$$\det\begin{pmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{pmatrix} = 0.$$

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

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

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

After rearranging,

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

'λ' can be calculated by quadratic equ,

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad \begin{matrix} a=1 \\ b=-37 \\ c=201 \end{matrix}$$

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

$$= \frac{37 \pm \sqrt{1369 - 804}}{2} = \frac{37 \pm \sqrt{565}}{2}$$

$$= \frac{37 \pm 23.76}{2} \Rightarrow \frac{37+23.76}{2}, \frac{37-23.76}{2}$$

$$= \frac{60.76}{2}, \frac{13.24}{2}$$

Eigen Values. $\boxed{\lambda_1 = 30.38, \lambda_2 = 6.62}$

So, while arranging in descending order,

$$\lambda_1 > \lambda_2 > \ldots\ldots$$

Hence, $\lambda_1 = 30.38$

$\lambda_2 = 6.62$

We are going to find out Eigenvectors for Eigen value, $\lambda = 30.38$.

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

S: covariance matrix; λ_1: 30.38; I: Identity matrix; U_1: Eigen vectors of λ_1

$$\left(\begin{pmatrix} 14 & -11 \\ -11 & 23 \end{pmatrix} - 30.38 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) U_1 = 0$$

Assume $U_1 = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$

Hence,

$$\begin{pmatrix} 14 & -11 \\ -11 & 23 \end{pmatrix} - \begin{pmatrix} 30.38 & 0 \\ 0 & 30.38 \end{pmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{pmatrix} 14-30.38 & -11 \\ -11 & 23-30.38 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{pmatrix} -16.38 & -11 \\ -11 & -7.38 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$-16.38\,v_1 - 11v_2 = 0 \rightarrow (1)$$
$$-11\,v_1 - 7.38v_2 = 0 \rightarrow (2)$$

So, from this if we need to calculate v_1 & v_2

$(1) \times 7.38 \Rightarrow 120.88v_1 - 81.18v_2 = 0$

$(2) \times -11. \Rightarrow +121\,v_1 + 81.18\,v_2 = 0$

$$0.12v_1 = 0$$

$$\boxed{v_1 = 0}$$

then, apply v_1 in (1), then

$$-16.38 \times 0 - 11v_2 = 0$$

$$\boxed{v_2 = 0}$$

this can't be possible, hence

$$\begin{bmatrix} (14-\lambda_1) & (-11) \\ (-11) & (28-\lambda_1) \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$(14-\lambda_1)v_1 - 11v_2 = 0 \rightarrow (a)$$
$$-11v_1 + (23-\lambda_1)v_2 = 0 \rightarrow (b)$$

from (a),

$$(14-\lambda_1)v_1 - 11v_2 = 0$$
$$(14-\lambda_1)v_1 = 11v_2.$$
$$\frac{v_1}{11} = \frac{v_2}{14-\lambda_1} = A \text{ (Assigning)}$$

Assume $A=1$,

$$\frac{v_1}{11} = \frac{v_2}{14-\lambda_1} = A = 1.$$

Hence, $\frac{v_1}{11} = 1 \Rightarrow v_1 = 11.$

$$\frac{v_2}{14-\lambda_1} = 1 \Rightarrow v_2 = 14-\lambda_1$$
$$= 14 - 30.38$$
$$= -16.38.$$

Hence Eigen vector for $\lambda_1 \Rightarrow \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.38 \end{bmatrix}$

Then if we want to normalize the eigen vector,

$$n_1 = \begin{bmatrix} 11/\sqrt{11^2+16.38^2} \\ -16.38/\sqrt{11^2+16.38^2} \end{bmatrix}$$ / dividing by its length.

$$= \begin{bmatrix} \frac{11}{19.73} \\ -16.38/19.73 \end{bmatrix} = \begin{bmatrix} 0.5575 \\ -0.8302 \end{bmatrix}$$

Now, calculate eigen vector for $\lambda_2 = 6.62$

$$(S - \lambda_2 I) v_2 = 0$$

$$\begin{bmatrix} (14-\lambda_2) & -11 \\ -11 & (23-\lambda_2) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$(14-\lambda_2)v_1 - 11v_2 = 0 \rightarrow (c)$$

$$-11v_1 - (23-\lambda_2)v_2 = 0 \rightarrow (d)$$

from (c),

$$(14-\lambda_2)v_1 - 11v_2 = 0$$

$$(14-\lambda_2)v_1 = 11v_2$$

$$\frac{v_1}{11} = \frac{v_2}{14-\lambda_2} = B \text{ (Assume)}$$

Assume $B = 1$,

$$\frac{v_1}{11} = \frac{v_2}{14-\lambda_2} = B = 1$$

Hence, $\frac{v_1}{11} = 1 \Rightarrow v_1 = 11$

$$\frac{v_2}{14-\lambda_2} = 1 \Rightarrow v_2 = 14-\lambda_2$$

$$= 14 - 6.62$$

$$= 7.38$$

Hence, Eigen Vector for $\lambda_2 = \begin{bmatrix} 11 \\ 7.38 \end{bmatrix}$

If we want to normalize eigen vector,

$$n_2 = \begin{bmatrix} 11/\sqrt{11^2+7.38^2} \\ 7.38/\sqrt{11^2+7.38^2} \end{bmatrix} = \begin{bmatrix} 11/13.24 \\ 7.38/13.24 \end{bmatrix}$$

$$= \begin{bmatrix} 0.8308 \\ 0.5574 \end{bmatrix}$$

Step 5: New dataset,

Feature	sample1	sample2	sample3	sample4
a	4	8	13	7
b	11	4	5	14

1st PC	P_{11}	P_{12}	P_{13}	P_{14}
	sample1	sample2	sample3	sample4

$$P_{11} = \eta_1^T \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5575 & -0.8302 \end{bmatrix} \begin{bmatrix} -4 \\ 2.5 \end{bmatrix}$$

$$= (-2.23 - 2.0755)$$

$$= -4.305 //.$$

$$P_{12} = \eta_1^T \begin{bmatrix} 8-8 \\ 4-8.5 \end{bmatrix} \Rightarrow (0.5575 - 0.8302) \begin{pmatrix} 0 \\ -4.5 \end{pmatrix}$$

$$= 0 + 3.7359 = 3.7359 //$$

$$P_{13} = \eta_1^T \begin{bmatrix} 13-8 \\ 5-8.5 \end{bmatrix} \Rightarrow (0.5575 - 0.8302) \begin{pmatrix} 5 \\ -3.5 \end{pmatrix}$$

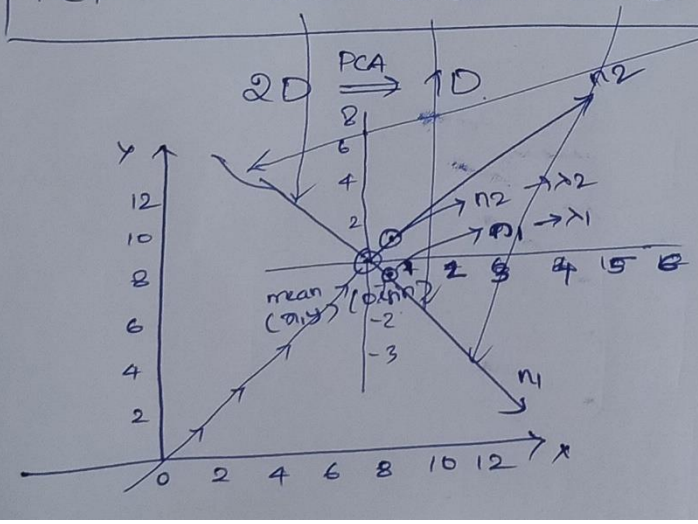
$$= 2.787 + 2.905 = \underline{5.692}.$$

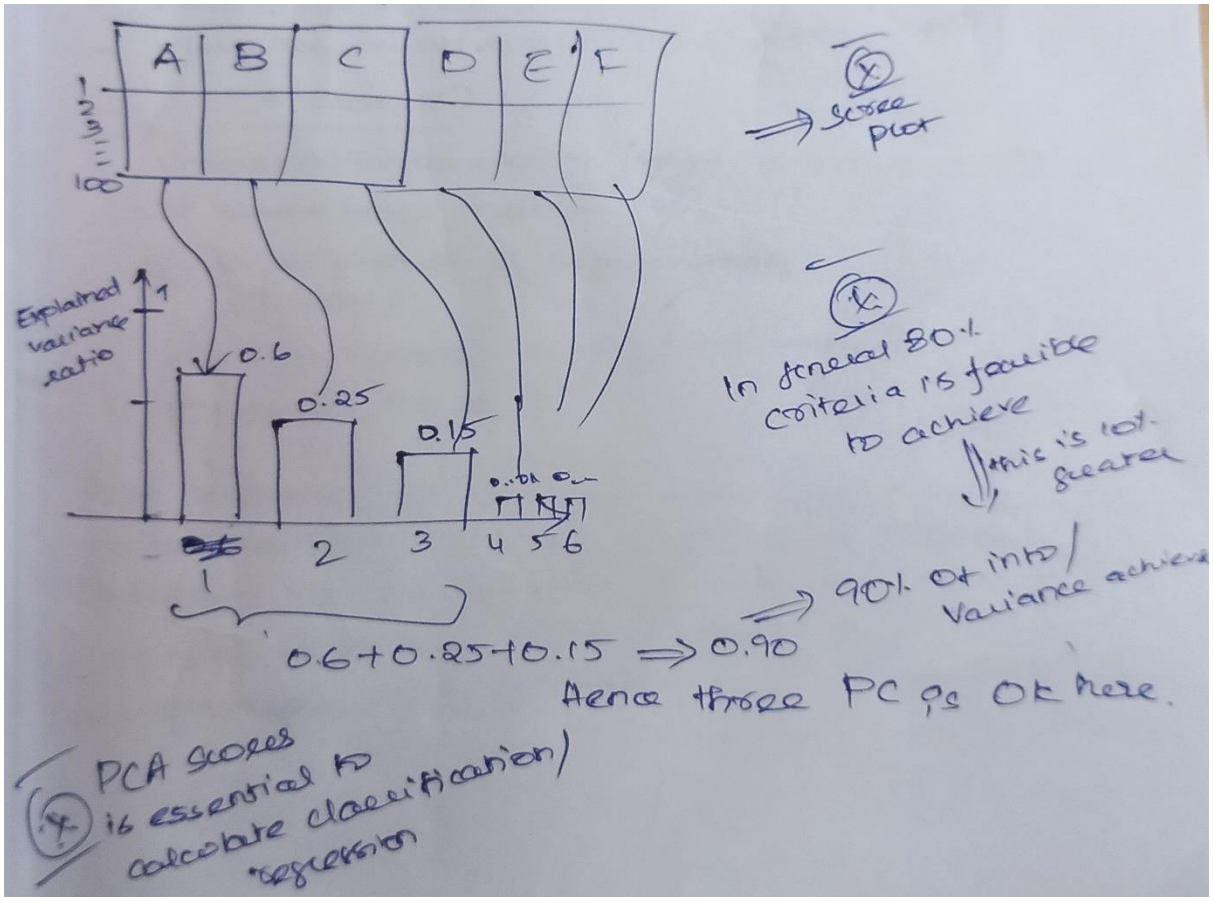
$$P_{14} = \eta_1^T \begin{bmatrix} 7-8 \\ 14-8.5 \end{bmatrix} = (0.5575 - 0.8302) \begin{pmatrix} -1 \\ 5.5 \end{pmatrix}$$

$$= -0.5575 - 4.5661$$

$$= -5.123 //.$$

PC1 =	-4.305	3.7359	5.692	-5.123





Identification of key dimensions in multidimensional scaling.

In statistics and data visualisation, Multidimensional Scaling (MDS) is a technique that is used to examine the structure of pairwise dissimilarities or distances between a group of items. Representing these items in a lower-dimensional space while retaining as much of the original pairwise distances as feasible is the aim of MDS. Axes in the reduced-dimensional space are referred to as dimensions in the context of MDS.

The emphasis is on interpreting and comprehending the meaning of these dimensions in relation to the data when selecting key dimensions in multidimensional scaling. The following actions and things to think about can help you find important MDS dimensions:

Analyse the Dimension Loads:

Examine each dimension's loadings after completing MDS. The contribution of each dimension to the total variance in the data is shown by loadings. Higher loading dimensions have a more significant role in illustrating the data's structure.

representing Dimensional Axes

Make an effort to understand what each dimension means. For instance, what does each axis represent in terms of the original variables or features if MDS reduces data to a 2D space? In order to comprehend the underlying structure, interpretability is essential.

Relationship between Dimensions and Original Variables

Examine the relationship that exists between each dimension and the dataset's original variables. Knowing which variables contribute more to each dimension and, thus, to the overall structure of the data, can be aided by this.

Perform an analysis of sensitivity:

Examine how adding or removing dimensions affects the MDS solution's configuration. Understanding the stability and robustness of the discovered characteristics can be gained from this sensitivity study.

Assessing Reproducibility

Verify whether the same dimensions consistently appear if you have numerous datasets or whether you can execute MDS with different subsets of your data. Confidence in the stability of the determined dimensions is increased when the dimensions are reproducible across many datasets.

Think about Outside Knowledge:

Use outside domain expertise to help with dimension interpretation. Verify whether expectations about the importance of particular variables, whether grounded in theory or practice, are consistent with the key dimensions that have been discovered.

Visualise the Outcomes:

Construct visual aids, like scatter plots or biplots, to investigate the connections between the original variables and the identified dimensions. Understanding the geometric arrangement of data points in the reduced-dimensional space can be aided by visualisation tools.

Assessing the Significance of Dimensions:

To determine the relevance of each dimension, apply resampling procedures or statistical tests. This can assist in differentiating between dimensions that may be the consequence of noise and those that are actually informative.

You can learn more about the important aspects of multidimensional scaling and enhance your comprehension of your data's structure in a reduced-dimensional space by implementing these procedures and taking these variables into account.

loadings and plots of principal components:

- Loadings:

• Loadings represent the coefficients of the original variables in the linear combinations that form the principal components. Each loading indicates the strength and direction of the relationship between a variable and a principal component.

- . Interpretation:

• A high positive or negative loading for a variable on a particular principal component suggests that the variable contributes significantly to the formation of that component. The sign of the loading indicates the direction of the relationship.

- Sum of Squares:

• The sum of the squared loadings for each variable across all principal components is equal to the variance of that variable in the original dataset.

- Variance Explained:

• The proportion of variance in the original data explained by a principal component is related to the sum of the squared loadings for that component. High loading values contribute more to the explanation of variance.

Principal Component Plots:

1. Scree Plot:

• The scree plot is a line plot that shows the eigenvalues of each principal component. The eigenvalues represent the amount of variance explained by each component. A sharp drop in eigenvalues helps in deciding how many principal components to retain.

2. Biplot:

• A biplot is a scatterplot that displays both the observations (data points) and the variables (loadings) simultaneously. Data points are represented in the reduced-dimensional space of the principal components, and vectors indicate the direction and magnitude of the loadings for each variable.

3. Interpretation of Biplot:

• In a biplot, the angle and length of a variable vector indicate its correlation with other variables. Variables pointing in the same direction are positively correlated, while those pointing in opposite directions are negatively correlated. Data points close to a variable vector have higher values for that variable.

4. Principal Component Scores:

• The scores of each observation on the principal components are obtained by projecting the data onto the principal component space. These scores can be used to understand the position of individual observations in the reduced-dimensional space.

5. Cumulative Variance Plot:

• A plot showing the cumulative proportion of variance explained by the first k principal components can help in deciding how many components to retain. Often, a threshold (e.g., 80% or 90% cumulative variance) is used as a criterion.

Understanding loadings and interpreting principal component plots is crucial for gaining insights into the structure and relationships within the data when applying PCA.