STATISTICAL MACHINE LEARNING

SIDDHARTH VARSANI

M. Tech(Bennett University)

What is machine Learning

- **Machine learning** (**ML**) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalise to unseen data, and thus perform tasks without explicit instructions

Type of ML

- Supervised Machine Learning
  - It is a type of machine learning where a model is trained on a labeled dataset. In this approach, the dataset includes input-output pairs, where the inputs are features (or independent variables) and the outputs are the corresponding target values (or labels).
- Unsupervised Machine Learning
  - It is a type of machine learning where the model is trained on a dataset that does not have labeled outputs. In this approach, the algorithm learns to identify patterns, relationships, or structures within the input data without explicit guidance.

Supervised Machine Learning

- Linear Regression
- Logistic Regression
- Support Vector Machine(SVM)
- Decision Tree
- Random Forest
- K-Nearest Neighbour(k-NN)
- Naive Bayes

Unsupervised Machine Learning

- K-Means
- Hierarchical clustering
- DBSCAN
- Gaussian Mixture Model
- Principal Component Analysis(PCA)
- T- Distributed stochastic Neighbour Embedding

Todays topic

- Supervised Machine Learning
  - NAÏVE BAYES
  - K NEAREST NEIGHBOUR
- Unsupervised machine Learning
  - PCA

Basic terms in ML

- **Nominal Data**:- Nominal data is a type of categorical data where the categories are **labels or names** that do not have any inherent order or ranking. It is used to classify data into distinct groups, but the groups do not have a logical sequence.

- **Ordinal Data**:- Ordinal data is also a type of categorical data, but the categories have a **meaningful order or ranking**. However, the intervals between the categories are not necessarily equal or measurable.

- **Bias** :- Bias refers to the error introduced by approximating a real-world problem, which may be complex, with a simplified model. It reflects the model's **assumptions** and how well it captures the underlying patterns.
  - **High Bias**: The model oversimplifies the problem, leading to under fitting. It fails to capture the data's complexity.
  - **Low Bias**: The model makes fewer assumptions and is more flexible in capturing the underlying patterns.

- **Variance**:- Variance refers to the model's **sensitivity** to small changes in the training data. High variance models are highly flexible and can fit training data well but often fail to generalise to unseen data.
  - **High Variance**: The model overfits the training data, capturing noise along with the actual patterns.
  - **Low Variance**: The model is stable and less sensitive to changes in the training data.

- **Underfitting :-** Under fitting occurs when the model is too simple to capture the underlying structure of the data. It results in poor performance on both the training data and unseen data.
  - **Symptoms**:- High training error and High validation/testing error
- Overfitting :- Overfitting occurs when the model is too complex and captures noise in the training data, along with the actual patterns. It performs very well on the training data but poorly on unseen data.
  - **Symptoms**:- Low training error and high Validation/test error

NAÏVE BAYES (CLASSIFIER)

- WHAT IS NAÏVE BYASE
  - Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.
- **Assumption of Naive Bayes**
  - **Feature independence:** The features of the data are conditionally independent of each other, given the class label.
  - **Continuous features are normally distributed:** If a feature is continuous, then it is assumed to be normally distributed within each class.
  - **Discrete features have multinomial distributions:** If a feature is discrete, then it is assumed to have a multinomial distribution within each class.
  - **Features are equally important:** All features are assumed to contribute equally to the prediction of the class label.
  - **No missing data:** The data should not contain any missing values.

QUESTION :

Suppose we want our model to predict whether a "childless" family having age group of "young " and "medium" income will buy a car or not ?

WHY WE NEED

Bag 1 contains 2 Red and 3 Green balls, Bag 2 Contains 3 Red and 4 Green balls,One ball is drawn at random from one of the bags and its Red. Find the probability that it is drawn from Bag 1.

Bag 1

- **2 Red**
- **3 Green**

Bag 2

- **3 Red**
- **4 Green**

New Bag

Red

Bayes' Theorem

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Data

FAMILY	AGE GROUP	IMCOME	OWN A PET	BUY A CAR
Nuclear	Young	Low	Yes	No
Extended	Old	Low	Yes	No
Childless	Middle	Low	No	No
Childless	Young	High	No	Yes
Extended	Middle	Medium	Yes	Yes
Nuclear	Old	High	No	Yes
Nuclear	Middle	Medium	Yes	Yes
Extended	Middle	High	No	Yes
Single	Old	Low	No	No
Childless	Young	Medium	Yes	Yes

QUESTION :

Suppose we want our model to predict whether a “childless” family having age group of “young ” and “medium” income will buy a car or not ?

Question :

P(Buy a car | childless, young, medium) = ?

P(Buy a car | childless, young, medium) =

$$\left[\frac{P\left(\frac{\text{childless, young, medium}}{\text{Buy a car}}\right) \times P(\text{Buy a car})}{P(\text{childless, young, medium})}\right]$$

$$P\left(\frac{\text{childless, young, medium}}{\text{Buy a car}}\right) = P\left(\frac{\text{childless}}{\text{Buy a car}}\right) \times P\left(\frac{\text{young}}{\text{Buy a car}}\right) \times P\left(\frac{\text{Medium}}{\text{Buy a car}}\right)$$

From given data we can say

There is 6 “yes” and 4 “No”

Lets calculate probability for YES $P\left(\frac{\text{Yes}}{\text{Total data}}\right) = \frac{6}{10}$

Lets calculate probability for NO $P\left(\frac{\text{NO}}{\text{Total data}}\right) = \frac{4}{10}$

From data

Likelihood YES = $\frac{\text{Count}}{\text{Total YES}}$ Likelihood NO = $\frac{\text{Count}}{\text{Total NO}}$

		Frequency			Likelihood	
		YES	NO	TOTAL	YES	NO
FAMILY	Nuclear	2	1	3	2/6	1/4
	Extended	2	1	3	2/6	1/4
	Childless	2	1	3	2/6	1/4
	Single	1	0	1	1/6	0/4
AGE	Young	2	1	3	2/6	1/4
	Old	1	2	3	1/6	2/4
	Middle	1	3	4	1/6	3/4
SALARY	Low	0	4	4	0/6	4/4
	Medium	3	0	3	3/6	0/4
	High	3	0	3	3/6	0/4

$$P\left(\frac{\text{childless, young, medium}}{\text{Buy a car}}\right) = P\left(\frac{\text{childless}}{\text{Buy a car}}\right) \times P\left(\frac{\text{young}}{\text{Buy a car}}\right) \times P\left(\frac{\text{Medium}}{\text{Buy a car}}\right)$$

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

$$= \frac{1}{18}$$

Lets use this value for our main equation

$$\left[\frac{P\left(\frac{\text{childless, young, medium}}{\text{Buy a car}}\right) \times P(\text{Buy a car})}{P(\text{childless, young, medium})}\right]$$

$$= \frac{1}{18} \times \frac{6}{10} = \frac{6}{180}$$

Now calculate for not buying a car

$$\left[\frac{P\left(\frac{\text{childless, young, medium}}{\text{Not Buy a car}}\right) \times P(\text{not buy a car})}{P(\text{childless, young, medium})} \right]$$

where

$$P\left(\frac{\text{childless, young, medium}}{\text{Not Buy a car}}\right) = P\left(\frac{\text{childless}}{\text{Not Buy a car}}\right) \times P\left(\frac{\text{young}}{\text{Not Buy a car}}\right) \times P\left(\frac{\text{Medium}}{\text{Not Buy a car}}\right)$$

$$P\left(\frac{\text{childless, young, medium}}{\text{Not Buy a car}}\right) = P\left(\frac{\text{childless}}{\text{Not Buy a car}}\right) \times P\left(\frac{\text{young}}{\text{Not Buy a car}}\right) \times P\left(\frac{\text{Medium}}{\text{Not Buy a car}}\right)$$

$$= \frac{1}{4} \times \frac{2}{4} \times \frac{0}{4}$$

$$= 0$$

Lets use this value for our main equation

$$\left[\frac{P\left(\frac{\text{childless, young, medium}}{\text{Not Buy a car}}\right) \times P(\text{not buy a car})}{P(\text{childless, young, medium})} \right]$$

where

$$= \quad 0 \quad \times \quad \frac{4}{10} \quad = \quad 0$$

Here we are comparing for YES and NO so we

can Ignore P (childless, young, medium)

ANS :- our model will predict "childless" family having age group of "young " and "medium" income YES they will buy a car

TRY TO SOLVE BY YOUR SELF

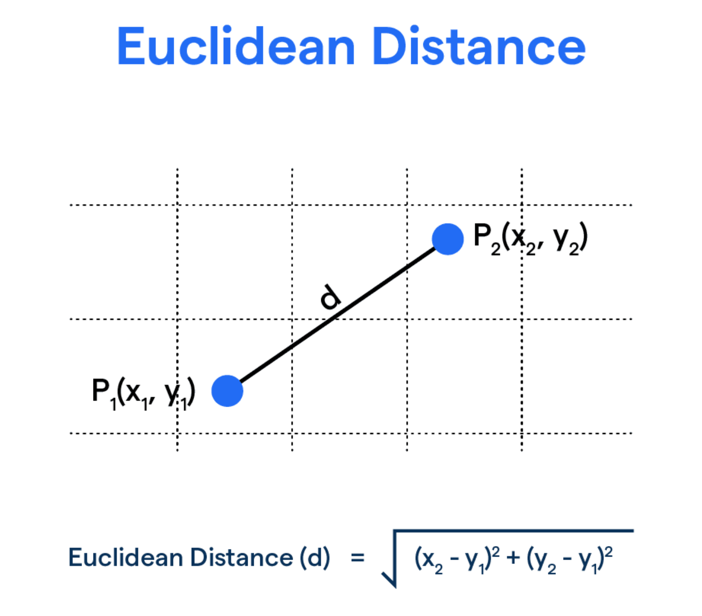
QUESTION :

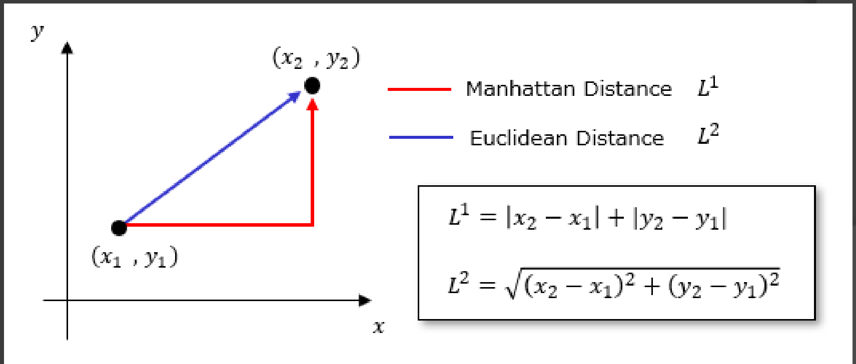
Suppose we want our model to predict whether a “extended” family having age group of “not have pat ” and “High” income and will buy a car or not ?

ANS :- our model will predict “Extended” family having age group of “not have pat” and “High” income YES they will buy a car

K-Nearest Neighbour

- Find Distance
  - Euclidean distance
  - Manhattan distance





Data

BMI	AGE	SUGAR
33.6	50	1
26.6	30	0
23.4	40	0
43.1	67	0
35.3	23	1
35.9	67	1
36.7	45	1

Apply K-NN to predict the diabetic patient with given features BMI, age of training examples assume K=3.

Test Example = BMI = 43.6 , AGE = 40 SUGAR = ?

$$\sqrt{(X1-X2)^2 + (Y1-Y2)^2}$$

$D1 = \sqrt{(43.6 - 33.6)^2 + (40 - 50)^2} = 14.14$

$D2 = \sqrt{(43.6 - 26.6)^2 + (40 - 30)^2} = 19.72$

$D3 = \sqrt{(43.6 - 23.4)^2 + (40-40)^2} = 20.2$

$D4 = \sqrt{(43.6 - 43.1)^2 + (40-67)^2} = 27.0$

$D5 = \sqrt{(43.6 - 35.3)^2 + (40-23)^2} = 18.92$

$D6 = \sqrt{(43.6 - 35.9)^2 + (40-67)^2} = 28.07$

$D7 = \sqrt{}(43.6 - 36.7)^2 + (40-45)^2 = 8.52$

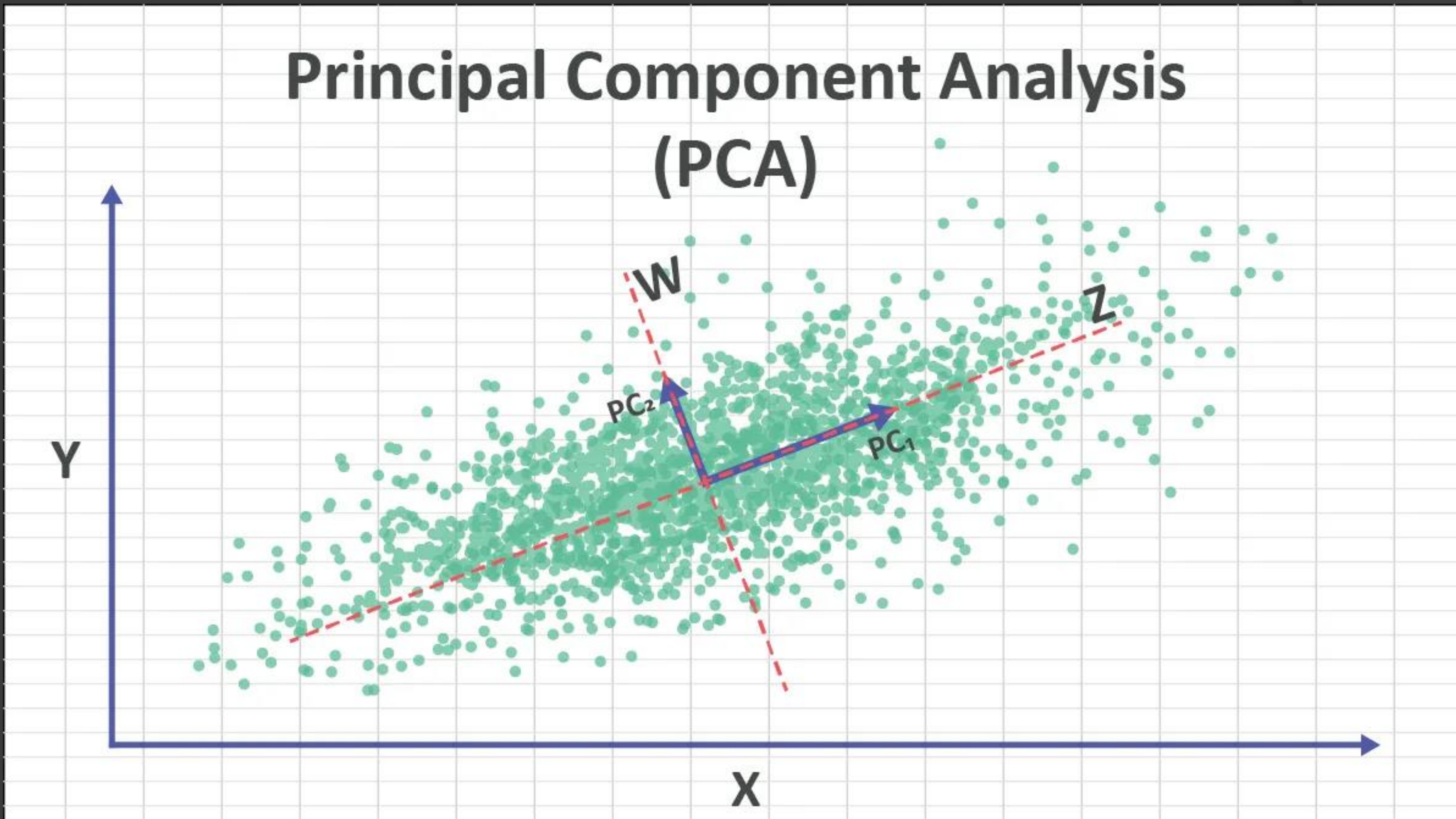
D1 = 14.4	Rank - 2	1
D2= 19.72		0
D3 = 20.2		0
D4 = 27.0		0
D5 = 18.92	Rank - 3	1
D6 = 27.07		1
D7 = 8.52	Rank - 1	1

The minimum distance between data points d1,d5,d7 is minimum and all the points in the data targets towards the sugar = 1 ,

Principle component Analysis

Principal Component Analysis (PCA) is a **dimensionality reduction technique** used in machine learning and statistics. It transforms a dataset with many features into a smaller set of features (called principal components) while retaining as much of the original data's variation as possible.

town	area	bathroom	plot	trees nearby	price
monroe	2600	2	8500	2	550000
monroe	3000	3	9200	2	565000
monroe	3200	3	8750	2	610000
monroe	3600	4	10200	2	680000
monroe	4000	4	15000	2	725000
west windsor	2600	2	7000	2	585000
west windsor	2800	3	9000	2	615000
west windsor	3300	4	10000	1	650000
west windsor	3600	4	10500	1	710000


Principal Component Analysis
(PCA)
W
Z
PC_2
PC_1
Y
X

Given the data in table, reduce the dimension from 2 to 1 using PCA

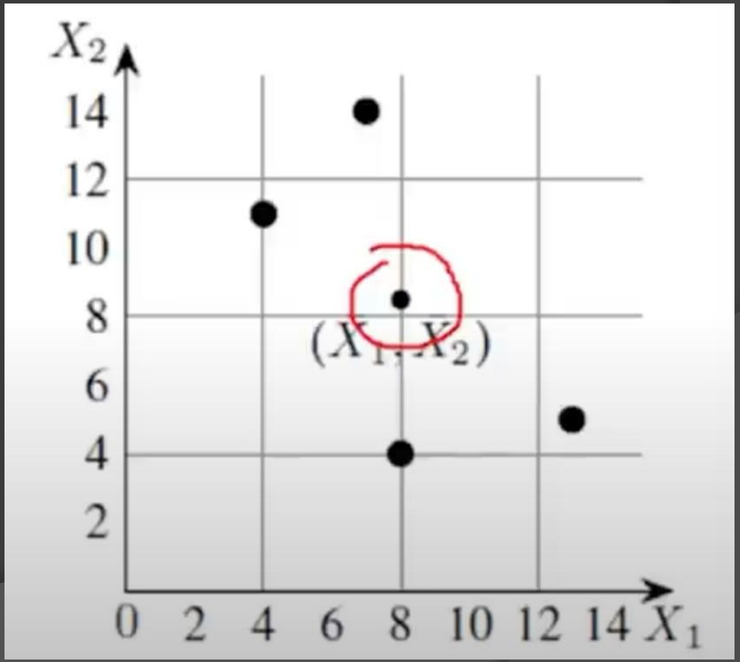
Features	Example 1	Example 2	Example 3	Example 4
X1	4	8	13	7
X2	11	4	5	14

- Step -1 calculate the mean.

X1 = (4+8+13+7)/4 = 8

X1 = (11+4+5+14)/4 = 8.5

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14



Step -2 calculation of the covariance matrix.

$$\begin{bmatrix} COV(X1, X1) & COV(X1, X2) \\ COV(X2, X1) & COV(X2, X2) \end{bmatrix}$$

$$COV(X1, X1) = \frac{1}{N-1} \sum_{K=1}^{N} (X1_K - \overline{X1})(X1_K - \overline{X1})$$

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$\overline{X1} = 8$

$\overline{X2} = 8.5$

- Step -2 calculation of the covariance matrix.

$$\begin{bmatrix} \text{COV }(X1, X1) & \text{COV }(X1, X2) \\ \text{COV }(X2, X1) & \text{COV }(X2, X2) \end{bmatrix}$$

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$$\text{COV }(X1, X1) = \frac{1}{N-1} \sum_{K=1}^{N} (X1_K - \overline{X1})(X1_K - \overline{X1})$$

$$= \frac{1}{3} \left((4-8)^2 + (4-8)^2 + (13-8)^2 + (7-8)^2\right)$$

$\overline{X1} = 8$

$\overline{X2} = 8.5$

$= 14$

Step -2 calculation of the covariance matrix.

$$\begin{bmatrix} COV(X1, X1) & COV(X1, X2) \\ COV(X2, X1) & COV(X2, X2) \end{bmatrix}$$

$$COV(X1, X2) = \frac{1}{N-1} \sum_{K=1}^{N} (X1_K - \overline{X1})(X2_K - \overline{X2})$$

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$\overline{X1} = 8$

$\overline{X2} = 8.5$

Step -2 calculation of the covariance matrix.

$$\begin{bmatrix} COV(X1, X1) & COV(X1, X2) \\ COV(X2, X1) & COV(X2, X2) \end{bmatrix}$$

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$$COV(X1, X2) = \frac{1}{N-1} \sum_{K=1}^{N} (X1_K - \overline{X1})(X2_K - \overline{X2})$$

$$= \frac{3}{3} ((4 - 8)+(11 – 8.5) + (8 - 8)^2 + (4 – 8.5) + (13 - 8)(5 – 8.5) + (7 – 8)(14 – 8.5))$$

$\overline{X1} = 8$

$\overline{X2} = 8.5$

$= -11$

COV (X1, X2) = COV (X1, X2)

= -11

- Step -2 calculation of the covariance matrix.

$$\begin{bmatrix} COV(X1, X1) & COV(X1, X2) \\ COV(X2, X1) & COV(X2, X2) \end{bmatrix}$$

$$COV(X2, X2) = \frac{1}{N-1} \sum_{K=1}^{N} (X2_K - \overline{X2})(X2_K - \overline{X2})$$

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$\overline{X1} = 8$

$\overline{X2} = 8.5$

- Step -2 calculation of the covariance matrix.

$$\begin{bmatrix} COV(X1, X1) & COV(X1, X2) \\ COV(X2, X1) & COV(X2, X2) \end{bmatrix}$$

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$$COV(X2, X2) = \frac{1}{N-1} \sum_{K=1}^{N} (X2_K - \overline{X2})(X2_K - \overline{X2})$$

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

$$= 23$$

$\overline{X1} = 8$

$\overline{X2} = 8.5$

Step -2 calculation of the covariance matrix.

$$\begin{bmatrix} COV(X1, X1) & COV(X1, X2) \\ COV(X2, X1) & COV(X2, X2) \end{bmatrix}$$

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

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$\overline{X1} = 8$

$\overline{X2} = 8.5$

Step - 3 Eigenvalues of the covariance matrix.

The characteristic equation of the covariance matrix is

$$0 = \text{determine}(s - \lambda I)$$

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

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$\overline{X1} = 8$

$\overline{X2} = 8.5$

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

Step - 3 Eigenvalues of the covariance matrix.

The characteristic equation of the covariance matrix is

$$0 = \text{determine}(s - \lambda I)$$

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

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

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

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$$\overline{X1} = 8$$

$$\overline{X2} = 8.5$$

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

Step - 3 Eigenvalues of the covariance matrix.

The characteristic equation of the covariance matrix is

$$0 = \text{determine}(s - \lambda I)$$

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

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

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

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

$\lambda = 1/2\,(37 \pm \sqrt{565})$

$\lambda = 30.3849\ ,\ 6.6151$

Quadratic formula

$$ax^2 + bx + c = 0$$

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

$\overline{X1} = 8$

$\overline{X2} = 8.5$

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

⊙ Step - 4 Compute Eigenvectors

$$U = \begin{bmatrix} U1 \\ U2 \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 0 \end{bmatrix} = (S - \lambda I)\, U$$

$$\begin{bmatrix} 14 - \lambda & -11 \\ -11 & 23 - \lambda \end{bmatrix} \begin{bmatrix} U1 \\ U2 \end{bmatrix}$$

$$\begin{bmatrix} (14 - \lambda)u1 - 11u2 \\ -11u1 + (23 - \lambda)u2 \end{bmatrix}$$

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$$\overline{X1} = 8$$
$$\overline{X2} = 8.5$$

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

$\lambda 1 = 30.3849$,

$\lambda 2 = 6.6151$

⦿ Step - 4 Compute Eigenvectors

$$\begin{bmatrix} (14 - \lambda)u1 - 11u2 \\ -11u1 + (23 - \lambda)u2 \end{bmatrix}$$

$(14 - \lambda)u1 - 11u2 = 0$

$-11u1 + (23 - \lambda)u2 = 0$

$(14 - \lambda)u1 = 11u2$

$$\frac{U1}{11} = \frac{U2}{14 - \lambda} = t$$

$U1 = 11t,$

$U2 - (14 - \lambda)t$

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$\overline{X1} = 8$

$\overline{X2} = 8.5$

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

$\lambda1 = 30.3849$,

$\lambda2 = 6.6151$

⊙ Step - 4 Compute Eigenvectors

$$\begin{bmatrix} 14 - \lambda & -11 \\ -11 & 23 - \lambda \end{bmatrix} \begin{bmatrix} 11 \\ 14 - \lambda 1 \end{bmatrix}$$

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

To find unit eigenvector, we compute the length of u1 which is

$$|| U1 || = \sqrt{(11^2 (14 - \lambda)}\ ^2$$

$$|| U1 || = \sqrt{(11^2 (14 - 30}.3849)\ ^2$$

$$|| U1 || = 19.7348$$

$\overline{X1} = 8$

$\overline{X2} = 8.5$

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

λ1 = λ 30.3849 ,

λ2 = 6.6151

Step - 5 finally Compute Eigenvectors

$$e1 = \begin{bmatrix} 11 / \|u1\| \\ (14 - \lambda)/\|u1\| \end{bmatrix}$$

$$\begin{bmatrix} 11 / 19.7348 \\ (14 - 30.3849)/ 19.7348 \end{bmatrix}$$

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

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$$e1 = \begin{bmatrix} 0.5574 \\ - 0.8303 \end{bmatrix}$$

$$\overline{X1} = 8$$

$$\overline{X2} = 8.5$$

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

$$e2 = \begin{bmatrix} 0.8308 \\ 0.5574 \end{bmatrix}$$

$\lambda1 = \lambda\ 30.3849$,

$\lambda2 = 6.6151$

Step - 6 computation of first principal

$$e1^T \begin{bmatrix} X1k - \overline{X1} \\ X2k - \overline{X2} \end{bmatrix}$$

$$= [\,0.5574 - \ 0.8303\,] \begin{bmatrix} X11 - \overline{X1} \\ X21 - \overline{X2} \end{bmatrix}$$

$$= [\,0.5574\ (X11 - \overline{X1}) - \ 0.8303\ (X21 - \overline{X2}\,]$$

$$= [\,0.5574\ (4 - 8\,) - \ 0.8303\ (11 - 8.5\,]$$

$$= - 4.30535$$

Do the dame for all the other examples

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14

$$e1 = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$

$$\overline{X1} = 8$$

$$\overline{X2} = 8.5$$

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

$$e2 = \begin{bmatrix} 0.8308 \\ 0.5574 \end{bmatrix}$$

$\lambda 1 = \lambda\ 30.3849$,

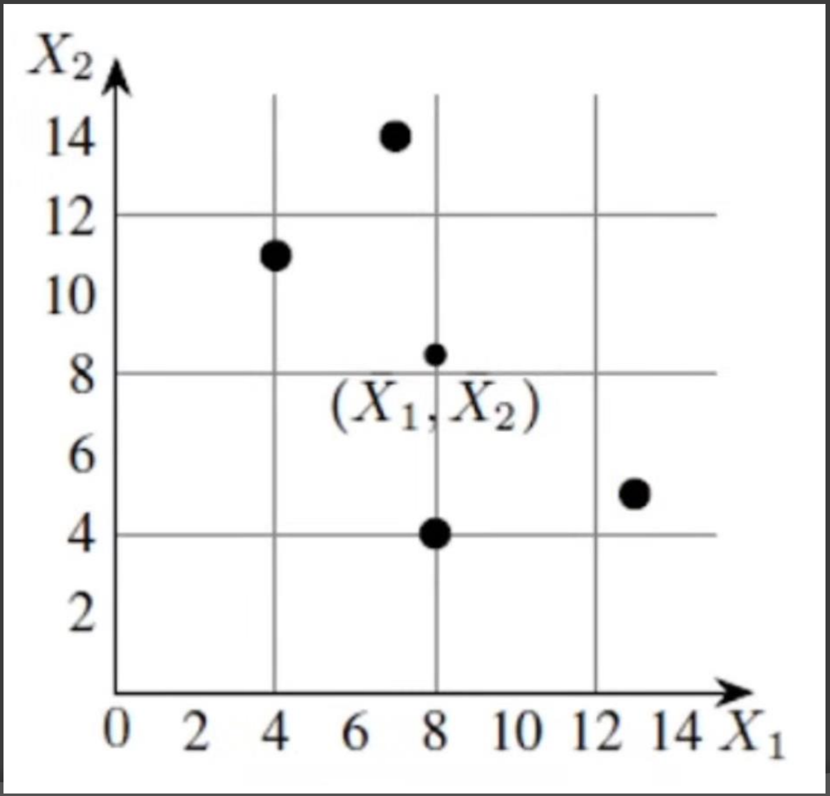
$\lambda 2 = 6.6151$

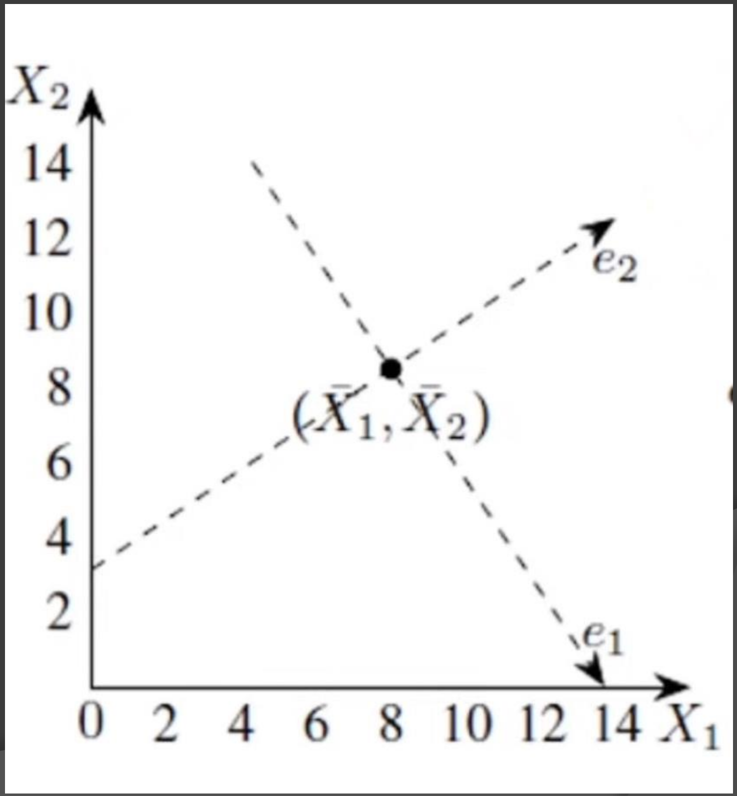
- Step - 6 computation of first principal

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14
PCA	- 4.3052	3.7361	5.6928	- 5.1238

⊙ Step - 6 computation of first principal

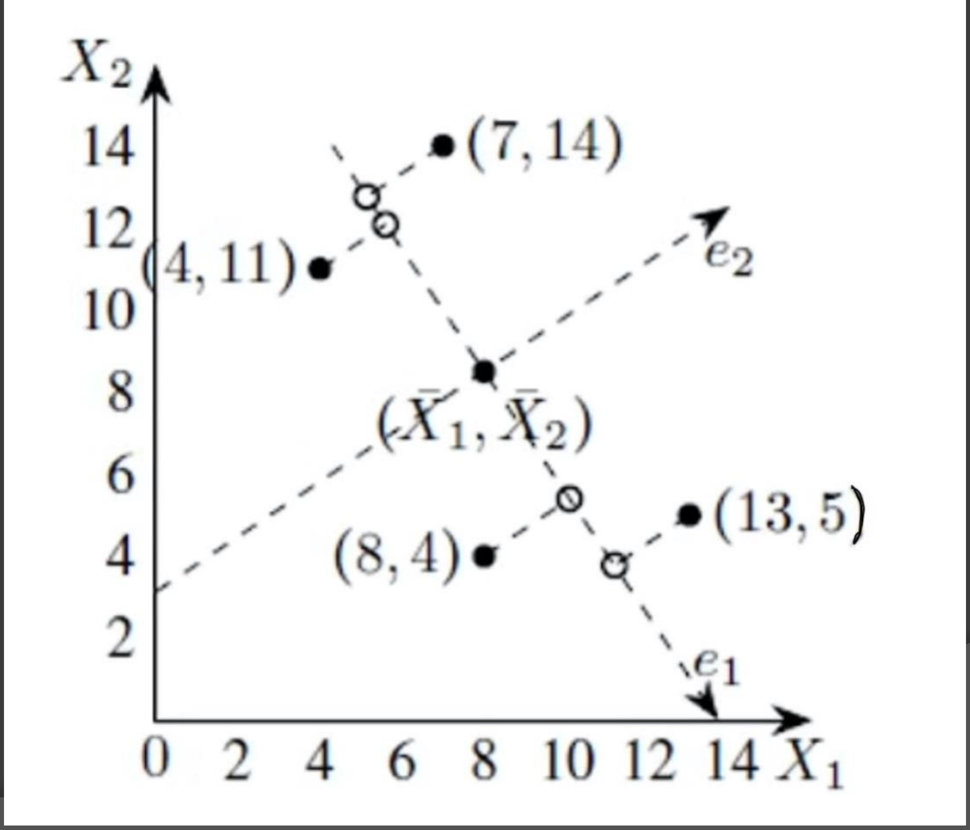
F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14
PCA	- 4.3052	3.7361	5.6928	- 5.1238





◉ Step - 6 computation of first principal

F	Ex. 1	Ex. 2	Ex. 3	Ex. 4
X1	4	8	13	7
X2	11	4	5	14
PCA	- 4.3052	3.7361	5.6928	- 5.1238



Thank you