MU

Sem 6

AIDS, CSE (DS), CSE (AIML), AIML, DE

DLOC II (ELECTIVE)

Distributed Computing

Dr. Nilesh Madhukar Patil
Dr. Pratik Kanani
Mr. Aniket
(SVKM's D. J. Sanghvi COE, Mumbai)

TECH-NEO PUBLICATIONS
Where Authors Inspire Innovation
A Sachin Shah Venture

With Solved Latest
UNIVERSITY QUESTION PAPERS.

www.techneobooks.in
info@techneobooks.in

DLOC II (ELECTIVES)

High Performance Computing CSDL(O6011)

Distributed Computing

Syllabus...

University of Mumbai

Machine Learning

Prerequisite : Data Structures, Basic Probability and Statistics, Algorithms	
Course Objectives	
1.	To introduce Machine learning concepts.
2.	To develop mathematical concepts required for Machine learning algorithms.
3.	To understand various Regression techniques.
4.	To understand Clustering techniques.
5.	To develop Neural Network based learning models.

Course Outcomes

After successful completion of the course students will be able to :

1. Comprehend basics of Machine Learning.
2. Build Mathematical foundation for machine learning.
3. Understand various Machine learning models.
4. Select suitable Machine learning models for a given problem.
5. Build Neural Network based models.
6. Apply Dimensionality Reduction techniques.

Module No.	Unit No.	Contents	Hrs.
1		**Introduction to Machine Learning**	6
	1.1	Introduction to Machine Learning, Issues in Machine Learning, Application of Machine Learning, Steps of developing a Machine Learning Application.	
	1.2	Supervised and Unsupervised Learning : Concepts of Classification, Clustering and prediction, Training, Testing and validation dataset, cross validation, overfitting and under fitting of model. **(Refer Chapter 1)**	
	1.3	Performance Measures : Measuring Quality of model - Confusion Matrix, Accuracy, Recall, Precision, Specificity, F1 Score, RMSE.	
2		**Mathematical Foundation for ML**	5
	2.1	System of Linear equations, Norms, Inner products, Length of Vector, Distance between vectors, Orthogonal vectors.	
	2.2	Symmetric Positive Definite Matrices, Determinant, Trace, Eigenvalues and vectors, Orthogonal Projections, Diagonalization, SVD and its applications. **(Refer Chapter 2)**	

MODULE 1

CHAPTER 1

Introduction to Machine Learning

Syllabus

Introduction to Machine Learning, Issues in Machine Learning, Application of Machine Learning, Steps of developing a Machine Learning Application.

Supervised and Unsupervised Learning : Concepts of Classification, Clustering and prediction, Training, Testing and validation dataset, cross validation, overfitting and under fitting of model.

Performance Measures : Measuring Quality of model - Confusion Matrix, Accuracy, Recall, Precision, Specificity, F1 Score, RMSE.

1.1 MACHINE LEARNING

> **UQ.** What is Machine learning? How it is different from data mining ?
>
> **(Ref. MU (Comp.) - May 17, 5 Marks, May 19, 5 Marks)**
>
> **UQ.** Define Machine learning and explain with example importance of Machine Learning.
>
> **(Ref. MU (Comp.) - Dec. 19, 5 Marks)**

- A machine that is intellectually capable as much as humans, have always attracted writers and early computer scientist who were excited about artificial intelligence and machine learning.
- The first machine learning system was developed in the 1950s. In 1952, Samuel has developed a program to play checkers. The program was able to observe positions at game and learn the model that gives better moves for machine player.
- In 1957, Frank Rosenblatt designed the Perceptron, which is a simple classifier but when it is combined in large numbers, in a network, it became a powerful tool.
- Minsky in 1960, came up with limitation of perceptron. He showed that the X-OR problem could not be represented by perceptron and such inseparable data distribution cannot be handled and following this Minsky's work neural network research went to dormant until 1980s.
- Machine learning became very famous in 1990s,due to the introduction of statistics. Computer science and statistics combination lead to probabilistic approaches in Artificial intelligence. This area is further shifted to data driven techniques. As Huge amount of data is available, scientists started to design intelligent systems that are able to analyze and learn from data.
- Machine learning is a category of Artificial Intelligence. In machine learning computers has the ability to learn themselves, explicit programming is not required.
- Machine focuses on the study and development of algorithms that can learn from data and also make predictions on data.
- Machine learning is defined by Tom Mitchell as "A program learns from experience 'E' with respect to some class of tasks 'T' and performance measure 'P', if its performance on tasks in 'T' as measured by 'P' improves with 'E'." Here 'E' represents the past experienced data and 'T' represents the tasks such as prediction, classification, etc. Example of 'P', we might want to increase accuracy in prediction.
- Machine learning mainly focuses on the design and development of computer programs that can teach themselves to grow and change when exposed to new data.
- Using machine learning we can collect information from a dataset by asking the computer to make some sense from data. Machine learning is turning data into information.

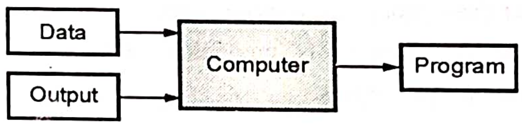


Fig. 1.1.1 : Machine Learning

- The Fig. 1.1.1 is the schematic representation of the ML system. ML system takes the training data and background knowledge as the input. Background knowledge and data helps the Learner program to provide a solution for a particular task or problem. Performance corresponding to the solution can be also measured. ML system comprises of mainly two components, Learner and a Reasoner. Learner use the training data and background knowledge to build the model and this can be used by reasoner to provide the solution for a task.
- Machine learning can be applied to many applications such as politics to geosciences. It is a tool that can be applied to many problems. Any application which needs to extract some information from data and also takes some action on data, can benefit from machine learning methods.

- Some of the applications are spam filtering in email, face recognition, product recommendations from Amazon.com and handwriting digit recognition.
- In detecting spam email, if you check for the occurrence of single word it will not be very helpful. But checking the occurrences of certain words used together and combined this with the length of the email and other parameters, you could get a much clearer idea of whether the email is spam or not.
- Machine learning is used by most of the companies to increase productivity, forecast weather, to improve business decisions, detect disease and do many more things.
- Machine learning uses statistics. There are many problems where the solution is not deterministic.
- There are certain problems for which we don't have that much information and also don't have that much computing power to properly model the problem.
- For these problems we need statistics, example of such type of problem is prediction of motivation and behavior of humans.
- The behavior and motivation of humans is a problem that is currently very difficult to model.

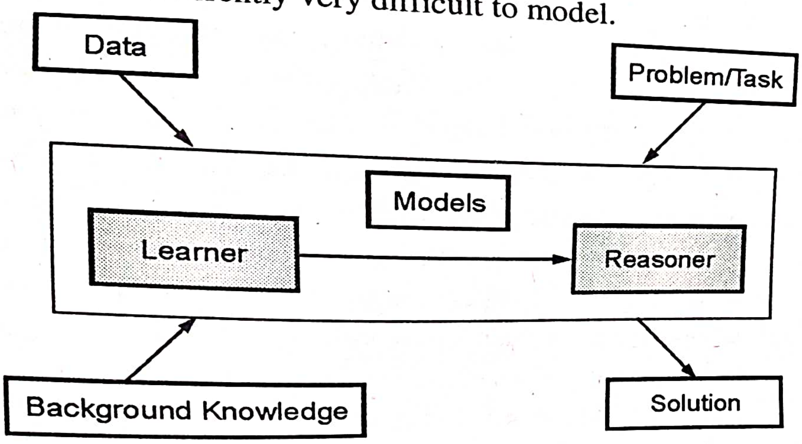


Fig. 1.1.2 : Schematic diagram of Machine Learning

Machine learning = Take data + understand it + process it + extract value from it + visualize it + communicate it

1.2 KEY TERMINOLOGY

UQ. What are the key tasks of Machine Learning?
(Ref. MU (Comp.) - May 16, 5 M

Expert System

- Expert system is a system which is developed some training set, testing set, and knowl representation, features, algorithm and classific terminology.
  - (i) ***Training Set :*** A training set comprises of trai examples which will be used to train mach learning algorithms.
  - (ii) ***Testing Set :*** To test machine learning algorith what's usually done is to have a training set data and a separate dataset, called a test set.
  - (iii) ***Knowledge Representation :*** Knowled representation may be stored in the form of a s of rules. It may be an example from the trainin set or a probability distribution.
  - (iv) ***Features :*** Important properties or attributes.
  - (v) ***Classification :*** We classify the data based on features.
- **Process:** Suppose we want to use a machine learning algorithm for classification. The next step is to train the algorithm, or allows it to learn. To train the algorithm we give as a input a quality data called as training set.
- Each training example has some features and one targe variable. The target variable is what we will be tryin to predict with our machine learning algorithms. In training dataset the target variable is known. Th machine learns by finding some relationship betwee the target variable and the features. In the classificatio tasks the target variables are known as classes. It i assumed that there will be a limited number of classes.
- The class or target variable that the training exampl belongs to is then compared to the predicted value, a we can get a idea about the accuracy of the algorithm.

- **Example :** First we will see some terminologies that are frequently used in machine learning methods. Let's take an example that we want to design a classification system that will classify the instances in to either Acceptable or Unacceptable. This kind of system is a fascinating topic often related with machine learning called *expert systems*.
- Four features of the various cars are stored in Table 1.2.1. The features or the attributes selected are Buying_Price, Maintenance_Price, Lug_Boot and Safety. Examples belong to Table 1.2.1 represents a record comprises of features.
- In Table 1.2.1 all the features are categorical in nature and takes limited disjoint values. The first two features represent the buying price and maintenance price of a car such as high, medium and low. Third feature shows the luggage capacity of a car as small, medium or big. Fourth feature represents whether the car has safety measures or not, which takes the value as low, medium or high.
- Classification is one of the important task in machine learning. In this application we want to evaluate the car out of a group of other cars. Suppose we have all information about car's Buying_Price, Maintenance_Price, Lug_Boot and Safety. Classification method is used to evaluate a given car as Acceptable or Unacceptable. Many machine learning algorithms are there that can be used for classification. The target or the response variable in this example is the evaluation of a car.
- Suppose we have selected a machine learning algorithm to use for classification. The main task in the classification is to train the algorithm, or allow it to learn. We give the experienced data as the input to train the algorithm which is called as training data.
- Let's assume training dataset contains 14 training records in Table 1.2.1. Suppose each training record has four features and one target or the response variable as shown in Fig. 1.2.1.The machine learning algorithm is used to predict the target variable.
- In classification task the target variable takes a discrete value, and in the task of regression its value could be continuous.
- In a training dataset we have the value of target variable. The relationship that exists between the features and the target variable is used by machine for learning. The target variable is the evaluation of the car.
- Classes are the target variables in the classification task. In classification systems it is assumed that classes are to be of limited number.
- Attributes or features are the individual values that, when combined with other features, make up a training example. This is usually columns in a training or test set.
- A training dataset and a testing dataset, is used to test machine learning algorithms. First the training dataset is given as input to the program. Program uses this data to learn. Next, the test set is given to the program.
- The program decides which instance of test data belongs to which class.
- The predicted output is compared with the actual output of the program, and we can get a idea about the accuracy of the algorithm.
- There are best ways to use all the information in the training dataset and test dataset.
- Assume in car evaluation classification system, we have tested the program and it meets the desired level of accuracy.
- Knowledge representation is used to check what the machine has learned. There are many ways in which knowledge can be represented.
- We can use set of rules or a probability distribution to represent the knowledge.
- Many algorithms represent the knowledge which is more interpretable to humans than others.

- In some situations we may not want to build an expert system but we are interested only in the knowledge representation that's acquired from training a machine learning algorithm.

Table 1.2.1 : Car evaluation classification based on four features

Buying_Price	Maintenance_ Price	Lug_Boot	Safety	Evaluation ?
High	High	Small	High	Unacceptable
High	High	Small	Low	Unacceptable
Medium	High	Small	High	Acceptable
Low	Medium	Small	High	Acceptable
Low	Low	Big	High	Acceptable
Low	Low	Big	Low	Unacceptable
Medium	Low	Big	Low	Acceptable
High	Medium	Small	High	Unacceptable
High	Low	Big	High	Acceptable
Low	Medium	Big	High	Acceptable
High	Medium	Big	Low	Acceptable
Medium	Medium	Small	Low	Acceptable
Medium	High	Big	High	Acceptable
Low	Medium	Small	Low	Unacceptable

Buying_Price	Maintenance_Price	Lug_Boot	Safety	Evaluation
Low	Low	Big	High	Acceptable
Features				Target Variables

Fig. 1.2.1 : Features and target variable identified

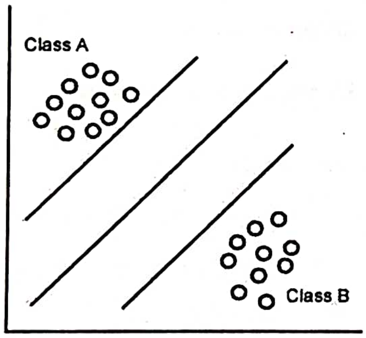
1.3 TYPES OF MACHINE LEARNING

1.3.1 Supervised Learning

> **GQ.** What is supervised learning?
>
> **GQ.** Explain supervised learning with the help of an example.
>
> **GQ.** How supervised learning works?

- Learning that takes place based on a class of examples is referred to as supervised learning. It is learning based on labelled data. In short, while learning, the system has knowledge of a set of labelled data. This is one of the most common and frequently used learning methods
- The supervised learning method is comprised of a series of algorithms that build mathematical models of certain data sets that are capable of containing both inputs and the desired outputs for that particular machine.
- The data being inputted into the supervised learning method is known as training data, and essentially consists of training examples which contain one or more inputs and typically only one desired output. This output is known as a "supervisory signal."
- In the training examples for the supervised learning method, the training example is represented by an array, also known as a vector or a feature vector, and the training data is represented by a matrix.
- The algorithm uses the iterative optimization of an objective function to predict the output that will be associated with new inputs.
- Ideally, if the supervised learning algorithm is working properly, the machine will be able to correctly determine the output for the inputs that were not a part of the training data.
- Supervised learning uses classification and regression techniques to develop predictive models. Classification techniques predict categorical responses,
- Regression techniques predict continuous responses, for example, changes in temperature or fluctuations in power demand. Typical applications include electricity load forecasting and algorithmic trading.
- Let us begin by considering the simplest machine-learning task : supervised learning for classification. Let us take an example of classification of documents. In this particular case a learner learns based on the available documents and their classes. This is also referred to as labelled data.

- The program that can map the input documents to appropriate classes is called a classifier, because it assigns a class (i.e., document type) to an object (i.e., a document). The task of supervised learning is to construct a classifier given a set of classified training examples. A typical classification is depicted in Fig. 1.3.4.
- Fig. 1.3.4 represents a hyperplane that has been generated after learning, separating two classes - class A and class B in different parts. Each input point presents input–output instance from sample space. In case of document classification, these points are documents.

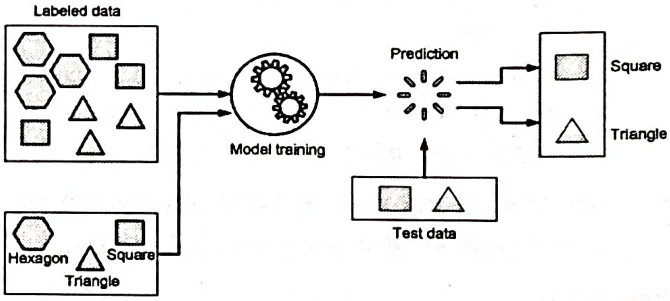


(1D1)**Fig. 1.3.4 : Supervised learning**

- Learning computes a separating line or hyperplane among documents. An unknown document type will be decided by its position with respect to a separator.
- There are a number of challenges in supervised classification such as generalization, selection of right data for learning, and dealing with variations. Labelled examples are used for training in case of supervised learning. The set of labelled examples provided to the learning algorithm is called the *training set.*
- Supervised learning is not just about classification, but it is the overall process that with guidelines maps to the most appropriate decision.

1.3.1(A) How Supervised Learning Works?

- In supervised learning, models are trained using labelled dataset, where the model learns about each type of data. Once the training process is completed, the model is tested on the basis of test data (a subset of the training set), and then it predicts the output.
- The working of Supervised learning can be easily understood by the below example and diagram (Fig. 1.3.5).



(1D2)**Fig. 1.3.5 How Supervised learning works?**

- Suppose we have a dataset of different types of shapes which includes square, rectangle, triangle, and Polygon. Now the first step is that we need to train the model for each shape.
  - o If the given shape has four sides, and all the sides are equal, then it will be labelled as a **Square**.
  - o If the given shape has three sides, then it will be labelled as a **triangle**.
  - o If the given shape has six equal sides then it will be labelled as **hexagon**.
- Now, after training, we test our model using the test set, and the task of the model is to identify the shape.
- The machine is already trained on all types of shapes, and when it finds a new shape, it classifies the shape on the bases of a number of sides, and predicts the output.
- Following are the steps involved in Supervised Learning :
  - o First Determine the type of training dataset
  - o Collect/Gather the labelled training data.

- Split the training dataset into training **dataset, test dataset, and validation dataset**.
- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
- Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
- Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

- Supervised learning can be further divided into two types of problems: Regression and Classification.

Regression

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc. Below are some popular Regression algorithms which come under supervised learning :

Linear Regression	Regression Trees	Non-Linear Regression
Bayesian Linear Regression	Polynomial Regression	

Classification

Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

Random Forest	Logistic Regression
Decision Trees	Support vector Machines

1.3.1(B) Advantages of Supervised Learning

(1) With the help of supervised learning, the model can predict the output on the basis of prior experiences.

(2) In supervised learning, we can have an exact idea about the classes of objects.

(3) Supervised learning model helps us to solve various real-world problems such as **fraud detection, spam filtering**, etc.

1.3.1(C) Disadvantages of Supervised Learning

(1) Supervised learning models are not suitable for handling the complex tasks.

(2) Supervised learning cannot predict the correct output if the test data is different from the training dataset.

(3) Training required lots of computation times.

(4) In supervised learning, we need enough knowledge about the classes of object.

1.3.2 Unsupervised learning

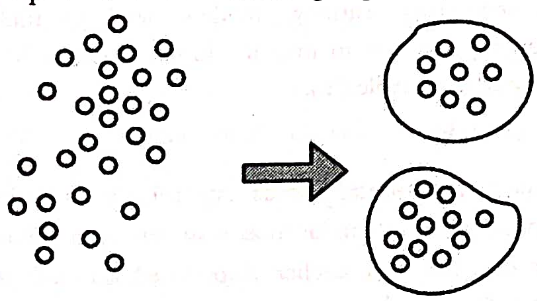
GQ. What is Unsupervised Learning?

GQ. What are the types of unsupervised learning?

GQ. What are the advantages and disadvantages of unsupervised learning?

- Unsupervised learning refers to learning from unlabeled data. It is based more on similarity and differences than on anything else. In this type of learning, all similar items are clustered together in a particular class where the label of a class is not known.
- It is not possible to learn in a supervised way in the absence of properly labeled data. In these scenarios there is need to learn in an unsupervised way.
- Here the learning is based more on similarities and differences that are visible. These differences and similarities are mathematically represented in unsupervised learning.
- Given a large collection of objects, we often want to be able to understand these objects and visualize their relationships.

- For an example based on similarities, a kid can separate birds from other animals. It may use some property or similarity while separating, such as the birds have wings.
- The criterion in initial stages is the most visible aspects of those objects. Linnaeus devoted much of his life to arranging living organisms into a hierarchy of classes, with the goal of arranging similar organisms together at all levels of the hierarchy.
- Many unsupervised learning algorithms create similar hierarchical arrangements based on similarity-based mappings.
- The task of hierarchical clustering is to arrange a set of objects into a hierarchy such that similar objects are grouped together.
- Non-hierarchical clustering seeks to partition the data into some number of disjoint clusters. The process of clustering is depicted in Fig. 1.3.6.
- A learner is fed with a set of scattered points, and it generates two clusters with representative centroids after learning. Clusters show that points with similar properties and closeness are grouped together.



(1D3)**Fig. 1.3.6 : Unsupervised learning**

- Unsupervised learning is a set of algorithms where the only information being uploaded is inputs.
- The device itself, then, is responsible for grouping together and creating ideal outputs based on the data it discovers. Often, unsupervised learning algorithms have certain goals, but they are not controlled in any manner.
- Instead, the developers believe that they have created strong enough inputs to ultimately program the machine to create stronger results than they themselves possibly could.
- The idea here is that the machine is programmed to run flawlessly to the point where it can be intuitive and inventive in the most effective manner possible.
- The information in the algorithms being run by unsupervised learning methods is not labelled, classified, or categorized by humans. Instead, the unsupervised algorithm rejects responding to feedback in favour of identifying commonalities in the data. It then reacts based on the presence, or absence, of such commonalities in each new piece of data that is being inputted into the machine itself.
- It is used to draw inferences from datasets consisting of input data without labelled responses. Clustering is the most common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns or groupings in data.
- Applications for clustering include gene sequence analysis, market research, and object recognition.

1.3.2(A) Types of Unsupervised Learning Algorithm

The unsupervised learning algorithm can be further categorized into two types of problems:

1. Clustering	2. Association

(1) Clustering

- Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group.
- Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those commonalities.

(2) Association

- An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database.
- It determines the set of items that occurs together in the dataset. Association rule makes marketing strategy more effective.
- Such as people who buy X item (suppose a bread) are also tend to purchase Y (Butter/Jam) item. A typical example of Association rule is Market Basket Analysis.

Below is the list of some popular unsupervised learning algorithms :

K-means clustering	Neural Networks
KNN (k-nearest neighbors)	Principle Component Analysis
Hierarchal clustering	Independent Component Analysis
Anomaly detection	Apriori algorithm
	Singular value decomposition

1.3.2(B) Advantages of Unsupervised Learning

(1) Unsupervised learning is used for more complex tasks as compared to supervised learning because, in unsupervised learning, we don't have labeled input data.

(2) Unsupervised learning is preferable as it is easy to get unlabeled data in comparison to labeled data.

1.3.2(C) Disadvantages of Unsupervised Learning

(1) Unsupervised learning is intrinsically more difficult than supervised learning as it does not have corresponding output.

(2) The result of the unsupervised learning algorithm might be less accurate as input data is not labeled, and algorithms do not know the exact output in advance.

In practical scenarios there is always need to learn from both labeled and unlabeled data. Even while learning in an unsupervised way, there is the need to make the best use of labeled data available. This is referred to as semi supervised learning. Semi supervised learning is making the best use of two paradigms of learning - that is, learning based on similarity and learning based on inputs from a teacher. Semi supervised learning tries to get the best of both the worlds.

1.3.2(D) Difference between Supervised and Unsupervised Learning

- Supervised and Unsupervised learning are the two techniques of machine learning. But both the techniques are used in different scenarios and with different datasets. Below the explanation of both learning methods along with their difference table is given.
- Supervised learning is a machine learning method in which models are trained using labeled data.
- In supervised learning, models need to find the mapping function to map the input variable (X) with the output variable (Y).

  $Y = f(X)$
- Supervised learning needs supervision to train the model, which is similar to as a student learns things in the presence of a teacher. Supervised learning can be used for two types of problems: Classification and Regression.
- **Example :** Suppose we have an image of different types of fruits. The task of our supervised learning model is to identify the fruits and classify them accordingly. So to identify the image in supervised learning, we will give the input data as well as output for that, which means we will train the model by the shape, size, color, and taste of each fruit. Once the training is completed, we will test the model by giving

the new set of fruit. The model will identify the fruit and predict the output using a suitable algorithm.

- Unsupervised learning is another machine learning method in which patterns inferred from the unlabeled input data. The goal of unsupervised learning is to find the structure and patterns from the input data. Unsupervised learning does not need any supervision. Instead, it finds patterns from the data by its own.
- Unsupervised learning can be used for two types of problems: Clustering and Association.
- **Example :** To understand the unsupervised learning, we will use the example given above. So unlike supervised learning, here we will not provide any supervision to the model. We will just provide the input dataset to the model and allow the model to find the patterns from the data. With the help of a suitable algorithm, the model will train itself and divide the fruits into different groups according to the most similar features between them.

UQ. Explain how supervised learning is different from unsupervised learning.

(Ref. MU (Comp.) - May 17, 5 Marks)

- The main differences between Supervised and Unsupervised learning are given below :

Table : 1.3.1

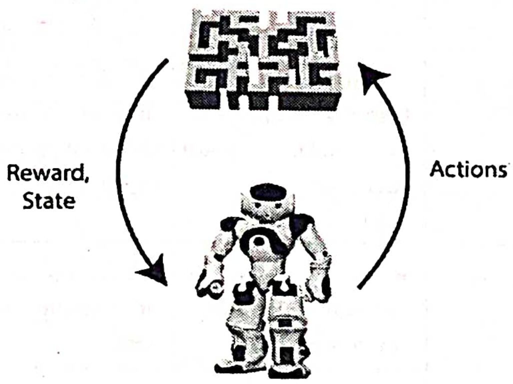
Supervised Learning	Unsupervised Learning
Supervised learning algorithms are trained using labeled data.	Unsupervised learning algorithms are trained using unlabeled data.
Supervised learning model takes direct feedback to check if it is predicting correct output or not.	Unsupervised learning model does not take any feedback.
Supervised learning model predicts the output.	Unsupervised learning model finds the hidden patterns in data.
In supervised learning, input data is provided to the model along with the output.	In unsupervised learning, only input data is provided to the model.
The goal of supervised learning is to train the model so that it can predict the output when it is given new data.	The goal of unsupervised learning is to find the hidden patterns and useful insights from the unknown dataset.
Supervised learning needs supervision to train the model.	Unsupervised learning does not need any supervision to train the model.
Supervised learning can be categorized in Classification and Regression problems.	Unsupervised Learning can be classified in Clustering and Associations problems.
Supervised learning can be used for those cases where we know the input as well as corresponding outputs.	Unsupervised learning can be used for those cases where we have only input data and no corresponding output data.
Supervised learning model produces an accurate result.	Unsupervised learning model may give less accurate result as compared to supervised learning.

Supervised Learning	Unsupervised Learning
Supervised learning is not close to true Artificial intelligence as in this, we first train the model for each data, and then only it can predict the correct output.	Unsupervised learning is more close to the true Artificial Intelligence as it learns similarly as a child learns daily routine things by his experiences.
It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, etc.	It includes various algorithms such as Clustering, KNN, and Apriori algorithm.

1.3.3 Reinforcement Learning

GQ. What is Reinforcement Learning? Explain with an example.

- Reinforcement Learning is a feedback-based Machine learning technique in which an agent learns to behave in an environment by performing the actions and seeing the results of actions. For each good action, the agent gets positive feedback, and for each bad action, the agent gets negative feedback or penalty.
- In Reinforcement Learning, the agent learns automatically using feedbacks without any labeled data, unlike supervised learning.
- Since there is no labelled data, so the agent is bound to learn by its experience only.
- RL solves a specific type of problem where decision making is sequential, and the goal is long-term, such as game-playing, robotics, etc.
- The agent interacts with the environment and explores it by itself. The primary goal of an agent in reinforcement learning is to improve the performance by getting the maximum positive rewards.
- The agent learns with the process of hit and trial, and based on the experience, it learns to perform the task in a better way. Hence, we can say that "Reinforcement learning is a type of machine learning method where an intelligent agent (computer program) interacts with the environment and learns to act within that." How a Robotic dog learns the movement of his arms is an example of Reinforcement learning.
- It is a core part of Artificial intelligence, and all AI agent works on the concept of reinforcement learning. Here we do not need to pre-program the agent, as it learns from its own experience without any human intervention.
- **Example :** Suppose there is an AI agent present within a maze environment, and his goal is to find the diamond. The agent interacts with the environment by performing some actions, and based on those actions, the state of the agent gets changed, and it also receives a reward or penalty as feedback.
- The agent continues doing these three things (take action, change state/remain in the same state, and get feedback), and by doing these actions, he learns and explores the environment.
- The agent learns that what actions lead to positive feedback or rewards and what actions lead to negative feedback penalty. As a positive reward, the agent gets a positive point, and as a penalty, it gets a negative point.



(1D4) **Fig. 1.3.7**

- For machine learning, the environment is typically represented by an "MDP" or Markov Decision Process.
- These algorithms do not necessarily assume knowledge, but instead are used when exact models are infeasible. In other words, they are not quite as precise or exact, but they will still serve a strong method in various applications throughout different technology systems.
- The key features of Reinforcement Learning are mentioned below.
  - o In RL, the agent is not instructed about the environment and what actions need to be taken.
  - o It is based on the hit and trial process.
  - o The agent takes the next action and changes states according to the feedback of the previous action.
  - o The agent may get a delayed reward.
  - o The environment is stochastic, and the agent needs to explore it to reach to get the maximum positive rewards.

1.3.3(A) Approaches to Implement Reinforcement Learning

GQ. What are the approaches for Reinforcement Learning?

There are mainly three ways to implement reinforcement-learning in ML, which are :

1. **Value-based :** The value-based approach is about to find the optimal value function, which is the maximum value at a state under any policy. Therefore, the agent expects the long-term return at any state(s) under policy π.
2. **Policy-based :** Policy-based approach is to find the optimal policy for the maximum future rewards without using the value function. In this approach, the agent tries to apply such a policy that the action performed in each step helps to maximize the future reward. The policy-based approach has mainly two types of policy :

   Deterministic : The same action is produced by the policy (π) at any state.

   Stochastic : In this policy, probability determines the produced action.
3. **Model-based :** In the model-based approach, a virtual model is created for the environment, and the agent explores that environment to learn it. There is no particular solution or algorithm for this approach because the model representation is different for each environment.

Here are important characteristics of reinforcement learning

- There is no supervisor, only a real number or reward signal
- Sequential decision making
- Time plays a crucial role in Reinforcement problems
- Feedback is always delayed, not instantaneous
- Agent's actions determine the subsequent data it receives

RL can be used in almost any application. It is a learning based on experience algorithm, a decision maker algorithm, an algorithm that learns autonomously, an optimization algorithm that over time learns to maximize its reward, the reward can be defined by the engineer to reach the objective of the problem.

1.3.3(B) Challenges of Reinforcement Learning

Here are the major challenges you will face while doing Reinforcement earning :

(1) Feature/reward design which should be very involved

(2) Parameters may affect the speed of learning.

(3) Realistic environments can have partial observability.

(4) Too much Reinforcement may lead to an overload of states which can diminish the results.

(5) Realistic environments can be non-stationary.

1.3.3(C) Applications of Reinforcement Learning

Here are applications of Reinforcement Learning :

(1) Robotics for industrial automation.

(2) Business strategy planning

(3) Machine learning and data processing

(4) Aircraft control and robot motion control

(5) It helps you to create training systems that provide custom instruction and materials according to the requirement of students.

1.3.3(D) Reinforcement Learning vs. Supervised Learning

GQ. What is the difference between Reinforcement Learning and Supervised Learning ?

Parameters	Reinforcement Learning	Supervised Learning
Decision style	Reinforcement learning helps you to take your decisions sequentially.	In this method, a decision is made on the input given at the beginning.
Works on	Works on interacting with the environment.	Works on examples or given sample data.
Dependency on decision	In RL method learning decision is dependent. Therefore, you should give labels to all the dependent decisions.	Supervised learning the decisions which are independent of each other, so labels are given for every decision.
Best suited	Supports and work better in AI, where human interaction is prevalent.	It is mostly operated with an interactive software system or applications.
Example	Chess game	Object recognition

1.4 ISSUES IN MACHINE LEARNING

UQ. What are the issues in Machine learning?

(Ref. MU (Comp.) - May 15, 5 Marks)

1. Which algorithm we have to select to learn general target functions from specific training dataset? What should be the settings for particular algorithms, so as to converge to the desired function, given sufficient training data? Which algorithms perform best for which type of problems and representations?

2. How much training data is sufficient? What should be the general amount of data that can be found to relate the confidence in learned hypotheses to the amount training experience and the character of the learner's hypothesis space?

3. Prior knowledge held by the learner is used at which time and manner to guide the process of generalizing from examples? If we have approximately correct knowledge, will it helpful even when it is only approximately correct?
4. What is the best strategy for choosing a useful next training experience, and how does the choice of this strategy after the complexity of the learning problem?
5. To reduce the task of learning to one or more function approximation problems, what will be the best approach? What specific functions should the system attempt to learn? Can this process itself be automated?
6. To improve the knowledge representation and to learn the target function, how can the learner automatically alter its representation?

1.5 HOW TO CHOOSE THE RIGHT ALGORITHM ?

UQ. Explain the steps required for selecting the right machine learning algorithm.

(Ref. MU (Comp.) - May 16, 8 Marks)

With all the different algorithms available in machine learning, how can you select which one to use ? First you need to focus on your goal. What are you trying to get out of this? What data do you have or can you collect ? Secondly you have to consider the data.

1. **Goal :** If you are trying to predict or forecast a target value, then you need to look into supervised learning. Otherwise, you have to use unsupervised learning.

   (a) If you have chosen **supervised** learning, then next you need to focus on what's your **target** value?

   If **target value** is **discrete** (e.g. Yes/ No, 1 /2/3, A/B/C), then use **Classification.**

   If **target value** is continuous i.e. Number of values (e.g. 0 – 100, – 99 to 99), then use **Regression.**

   (b) If you have chosen **unsupervised** learning, then next you need to focus on what is your **aim?**

   If you want to **fit your data** into some **discrete groups,** then use **Clustering**

   If you want to **find numerical estimate** of how strong the fit into each group, then use **density estimation algorithm**

2. **Data :** Are the features continuous or nominal ? Are there missing values in features? If yes, what is a reason for missing values? Are there outliers in the data? To narrow the algorithm selection process, all of these features of your data can help you.

Table 1.5.1 : Selection of Algorithm

	Supervised Learning	Unsupervised Learning
Discrete	Classification	Clustering
Continuous	Regression	Density Estimation

1.6 STEPS IN DEVELOPING A MACHINE LEARNING APPLICATION

UQ. Explain the steps of developing Machine Learning applications. **(Ref. MU (Comp.) - May 19, 10 Marks)**

1. Collection of Data

You could collect the samples from a website and extracting data.

- From RSS feed or an API
- From device to collect wind speed measurement
- Publicly available data.

2. Preparation of the input data

- Once you have the input data, you need to check whether it's in a useable format or not.
- Some algorithm can accept target variables and features as string; some need them to be integers.
- Some algorithm accepts features in a special format.

3. **Analyse the input data**

- Looking at the data you have passed in a text editor to check collection and preparation of input data steps are properly working and you don't have a bunch of empty values.
- You can also check at the data to find out if you can see any patterns or if there is anything obvious, such as a few data points greatly differ from remaining set of the data.
- Plotting data in 1, 2 or 3 dimensions can also help.
- Distil multiple dimensions down to 2/3 so that you can visualize the data.

4. The importance of this step is that it makes you understand that you don't have any garbage value coming in.

5. **Train the algorithm**

- Good clean data from the first two steps is given as input to the algorithm. The algorithm extracts information or knowledge. This knowledge is mostly stored in a format that is readily useable by machine for next 2 steps.
- In case of unsupervised learning, training step is not there because target value is not present. Complete data is used in the next step.

6. **Test the algorithm**

- In this step the information learned in the previous step is used. When you are checking an algorithm, you will test it to find out whether it works properly or not. In supervised case, you have some known values that can be used to evaluate the algorithm.
- In case of unsupervised, you may have to use some other matrices to evaluate the success. In either case, if you are not satisfied, you can again go back to step 4, change some things and test again.
- Mostly problem occurs in collection or preparation of data and you will have to go back to step 1.

7. **Use it**

In this step a real program is developed to do some task, and once again it is checked if all the previous steps worked as you expected. You might encounter some new data and have to revisit step 1-5.

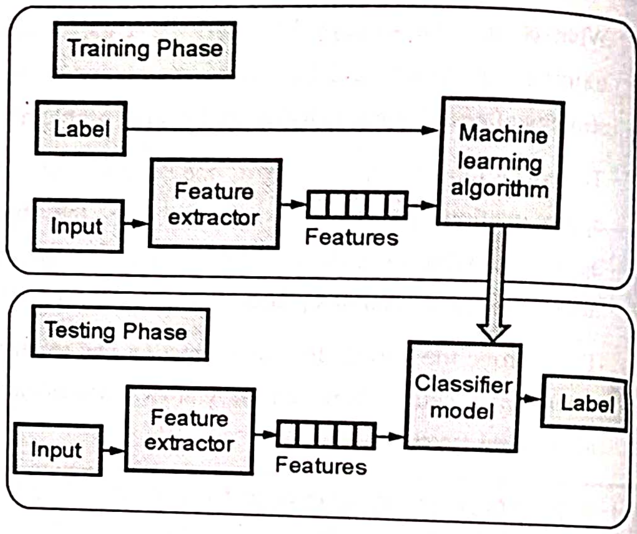


Fig. 1.6.1 : Typical example of Machine Learning Application

1.7 APPLICATIONS OF MACHINE LEARNING

UQ. Write short note on: Machine learning applications.
(Ref. MU (Comp.) - May 16, May 17, 10 Marks)

1. **Learning Associations**

- A supermarket chain-one an example of retail application of machine learning is basket analysis, which is finding associations between products bought by customers:
- If people who buy P typically also buy Q and if there is a customer who buys Q and does not buy P, he or she is a potential P customer. Once we identify such customers, we can target them for cross-selling.
- In finding an association rule, we are interested in learning a conditional probability of the form P (Q/P) where Q is the product we would like to condition on P, which are the product / products which we know that customer has already purchased.

P (Milk / Bread) = 0.7

- It implies that 70% of customers who buy bread also buy milk

2. Classification

- A credit is an amount of money loaned by a financial institution. It is important for the bank to be able to predict in advance the risk associated with a loan. Which is the probability that the customer will default and not pay the whole amount back?
- In credit scoring, the bank calculates the risk given the amount of credit and the information about the customer. (Income, savings, collaterals, profession, age, past financial history). The aim is to infer a general rule from this data, coding the association between a customer's attributes and his risk.
- Machine Learning system fits a model to the past data to be able to calculate the risk for a new application and then decides to accept or refuse it accordingly.

If income > Q_1 and savings > Q_2

Then low – risk ELSE high – risk

- Other classification examples are Optical character recognition, face recognition, medical diagnosis, speech recognition and biometric.

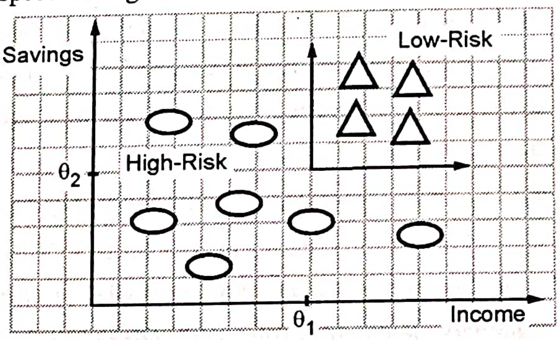


Fig. 1.7.1 : Classification for credit scoring

3. Regression

- Suppose we want to design a system that can predict the price of a flat. Let's take the inputs as the area of the flat, location and purchase year and other information that affects the rate of flat. The output is the price of the flat. The applications where output is numeric are regression problems.
- Let X represents flat features and Y is the price of flat. We can collect training data by surveying past purchased transactions and the Machine Learning algorithm fits a function to this data to learn Y as a function of X for the suitable values of W and W_0.

$$Y = w^*x + w_0$$

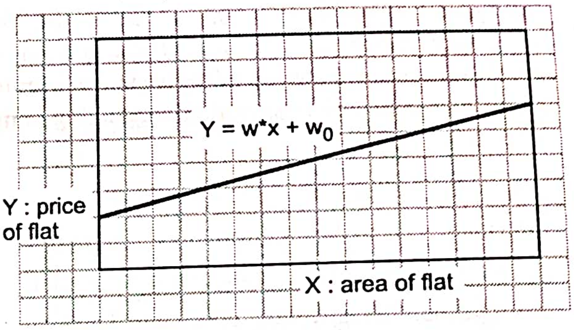


Fig. 1.7.2 : Regression for prediction of price of flat

4. Unsupervised Learning

- One of the important unsupervised learning problem is clustering. In clustering dataset is partitioned in to meaningful sub classes known as clusters. For example, suppose you want to decorate your home using given items.
- Now you will classify them using unsupervised learning (no prior knowledge) and this classification can be on the basis of color of items, shape of items, material used for items, type of items or whatever way you would like.

5. Reinforcement Learning

- There are some of the applications where output of system is a sequence of actions. In such applications the sequence of correct actions instead of single action is important in order to reach goal. An action is said to be good if it is part of good policy. Machine learning program generates a policy by learning previous good action sequences. Such methods are called reinforcement methods
- A good example of reinforcement learning is chess playing. In artificial intelligence and machine learning, one of the most important research area is game

playing. Games can be are easily described but at the same time, they are quite difficult to play well.

- Let's take a example of chess that has limited number of rules, but the game is very difficult because for each state there can be large number of possible moves.
- Another application of reinforcement learning is robot navigation. The robot can move in all possible directions at any point of time. The algorithm should reach goal state from an initial state by learning the correct sequence of actions after conducting number of trial runs.
- When the system has unreliable and partial sensory information, it makes reinforcement learning complex. Let's take an example of robot with incomplete camera information. Here robot does not know its exact location.

1.8 TRAINING ERROR AND GENERALIZATION ERROR

1.8.1 Training, Testing and Validation Dataset

When a labeled dataset is used to train machine learning models, it is common to break up the dataset into three parts :

- **Training** : used to directly improve the model's parameters.
- **Validation** : used to evaluate a model's performance while optimizing the model's hyperparameters.
- **Test**: used to evaluate a model after hyperparameter optimization is complete.

1.8.2 Why the Validation and Test Datasets are Necessary

- During the process of training a machine learning model, it is common for a model's parameters to over fit to the training dataset. These models report artificially high accuracy against the training dataset, but they perform poorly against data not in the training dataset.
- To realistically measure a model's performance, it is better to evaluate it against a validation dataset that was not used when training the model.
- You can train multiple models with different hyperparameters and compare them with the same validation dataset. But this creates a new problem: just like a model's parameters can overfit to the training dataset, a model's parameters and hyperparameters can overfit to the validation dataset.
- To realistically measure a set of models, it is better to evaluate them against a test dataset not used for training or validation. The test dataset is used to measure the performance of your various models at the end of the training process. Be careful not to repeatedly use the test dataset to re-train models or choose models, otherwise you risk creating models that have overfit to the test dataset.

1.8.3 Picking the Size of the Validation and Test Datasets

- The validation and test datasets need to be larger than a certain minimum size. Otherwise, the model's validation and test accuracy will not be representative of the "real-world" acccuracy. If your input dataset is very small, you can use cross-validation to train and evaluate a model against many different training/validation/test splits.
- For medium-sized datasets, it is typical for the validation and test datasets to *each* be 10%-30% of the total amount of data. For example, a common training/validation/test split is 60%/20%/20%.
- However, the validation and test datasets do not need to be larger than a certain absolute size. Above that size, adding more validation or test data does not make the model's performance metrics more realistic. If your input dataset contains millions of data samples, then you may only need about 1% each for the validation and test datasets.

1.8.4 How to Balance the Validation and Test Datasets

Preserve imbalanced classes

- If you are working on a classification problem with imbalanced classes–such as a dataset where one class is 99% of the dataset and the other class is 1% of the dataset–then you might consider improving the training process by oversampling the smaller class. But for your validation and test datasets, you want to measure your model's performance against the same class balance that your model would encounter in the real world.
- Validation and test datasets should have "newer" samples
- If you are training a model on time series data, typically your goal is to predict something about the future using data from the past or present.
- In order to properly evaluate a time series model, your training/validation/test split must obey the "arrow of time":
- All of the data samples in your validation dataset should be newer than your training dataset.
- All of the data samples in your test dataset should be newer than your validation dataset.
- If your training dataset contains data samples that are newer than your validation dataset, then your model's validation accuracy will be misleadingly high. Your model is effectively *traveling backward in time* if it trains on new data and evaluates on old data.
- Many people make the mistake of randomly shuffling the input dataset *before* splitting it into a training, validation, and test datasets–effectively violating the arrow of time.

Do not apply data augmentation

- Data augmentation is the use of computer algorithms to create or modify training data samples. The goal of data augmentation is to increase the size of the training dataset and to act as a regularizer–something that reduces a model's ability to overfit.
- However, when a machine learning model is deployed to the "real world" and is making predictions, typically the model will not perform any augmentation or regularization on its input. To mirror the real world, a model should not perform augmentation or regularization on the validation or test dataset.
- There are a few exceptions to the rule:
  - If the validation and/or test datasets are too small for a model to reliably evaluate, then it might make sense to use data augmentation to add data samples.
  - If the entire training dataset is computer-generated–like a dataset of images generated from a video game–then it may be reasonable for the validation and test datasets to also be entirely computer-generated.

Cross Validation

- In machine learning, we couldn't fit the model on the training data and can't say that the model will work accurately for the real data. For this, we must assure that our model got the correct patterns from the data, and it is not getting up too much noise. For this purpose, we use the cross-validation technique.
- Cross-validation is a technique in which we train our model using the subset of the data-set and then evaluate using the complementary subset of the data-set.
- The three steps involved in cross-validation are as follows:
  1. Reserve some portion of sample data-set.
  2. Using the rest data-set train the model.
  3. Test the model using the reserve portion of the data-set.

Methods of Cross Validation

Validation

In this method, we perform training on the 50% of the given data-set and rest 50% is used for the testing purpose.

The major drawback of this method is that we perform training on the 50% of the dataset, it may possible that the remaining 50% of the data contains some important information which we are leaving while training our model i.e higher bias.

LOOCV (Leave One Out Cross Validation)

- In this method, we perform training on the whole data-set but leaves only one data-point of the available data-set and then iterates for each data-point. It has some advantages as well as disadvantages also.
- An advantage of using this method is that we make use of all data points and hence it is low bias.
- The major drawback of this method is that it leads to higher variation in the testing model as we are testing against one data point.
- If the data point is an outlier it can lead to higher variation. Another drawback is it takes a lot of execution time as it iterates over 'the number of data points' times.

K-Fold Cross Validation

- In this method, we split the data-set into k number of subsets(known as folds) then we perform training on the all the subsets but leave one(k-1) subset for the evaluation of the trained model. In this method, we iterate k times with a different subset reserved for testing purpose each time.
- It is always suggested that the value of k should be 10 as the lower value of k is takes towards validation and higher value of k leads to LOOCV method.

Example

- The Fig 1.8.1 shows an example of the training subsets and evaluation subsets generated in k-fold cross-validation. Here, we have total 25 instances.
- In first iteration we use the first 20 percent of data for evaluation, and the remaining 80 percent for training ([1-5] testing and [5-25] training) while in the second iteration we use the second subset of 20 percent for evaluation, and the remaining three subsets of the data for training([5-10] testing and [1-5 and 10-25] training), and so on.

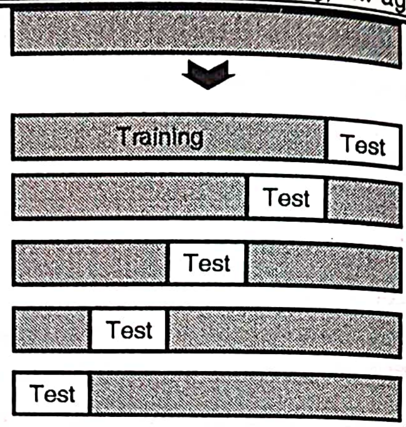


Fig:1.8.1: Cross Validation

1.8.5 Advantages of train/test split

(1) This runs K times faster than Leave One Out cross-validation because K-fold cross-validation repeats the train/test split K-times.

(2) Simpler to examine the detailed results of the testing process.

1.8.6 Advantages of cross-validation

(1) More accurate estimate of out-of-sample accuracy.

(2) More "efficient" use of data as every observation is used for both training and testing.

1.8.7 Training Error

In **machine learning, training** a predictive model means finding a function which maps a set of values x to a value y. If we apply the model to the data it was trained on we are calculating the **training error**.

If we calculate the **error** on data which was unknown in the **training** phase, we are calculating the test error. Training error is calculated as follows:

$$E_{train} = \frac{1}{n}\sum_{i=1}^{n} \text{error}\,(f_D(X_i), Y_i)$$

In the above equation n represents the number of training examples. $f_D(X_i)$ represents the predicted value and Y_i represents the true or actual values, error $(f_D(X_i), Y_i)$ is used to represent that these two values are same or not and if not then these values differs by how much.

1.8.8 Generalization Error

For supervised learning applications in machine learning and statistical learning theory, **generalization error** is a measure of how accurately an algorithm is able to predict outcome values for previously unseen data. Generalization error is calculated as follows:

- $E_{gen} = \int error\ (f_D\ (X_i),\ Y_i)\ P\ (Y, X)\ dX$

In the above equation error is calculated over all possible values of X and Y. $error\ (f_D(X_i), Y_i)$ is used to represent that these two values are same or not and if not then these values differes by how much. P(X, Y) represents how often we expect to see such X and Y.

1.8.9 Training Error versus Generalization Error

- The training error is the error of our model as calculated on the training dataset, while generalization error is the expectation of our model's error were we need to apply it to an infinite stream of additional data examples drawn from the same underlying data distribution as our original sample.
- Problematically, we can never calculate the generalization error exactly. That is because the stream of infinite data is an imaginary object. In practice, we must estimate the generalization error by applying our model to an independent test set constituted of a random selection of data examples that were withheld from our training set.
- Let's see an example. Consider a college student trying to prepare for his final exam.
- A diligent student will strive to practice well and test his abilities using exams from previous years. Nonetheless, doing well on past exams is no guarantee that he will excel when it matters.
- For instance, the student might try to prepare by rote learning the answers to the exam questions. This requires the student to memorize many things. She might even remember the answers for past exams perfectly.
- Another student might prepare by trying to understand the reasons for giving certain answers. In most cases, the latter student will do much better.
- Let's see one more example, consider the problem of trying to classify the outcomes of coin tosses (class 0: heads, class 1: tails) based on some contextual features that might be available.
- Suppose that the coin is fair. No matter what algorithm we come up with, the generalization error will always be 1/2. However, for most algorithms, we should expect our training error to be considerably lower, depending on the luck of the draw, even if we did not have any features! Consider the dataset {0, 1, 1, 1, 0, 1}.
- Our feature-less algorithm would have to fall back on always predicting the majority class, which appears from our limited sample to be 1. In this case, the model that always predicts class 1 will incur an error of 1/3, considerably better than our generalization error.
- As we increase the amount of data, the probability that the fraction of heads will deviate significantly from 1/2 diminishes, and our training error would come to match the generalization error.
- When we train our models, we attempt to search for a function that fits the training data as well as possible. If the function is so flexible that it can catch on to spurious patterns just as easily as to true associations, then it might perform *too well* without producing a model that generalizes well to unseen data. This is precisely what we want to avoid or at least control.
- Many of the techniques in deep learning are heuristics and tricks aimed at guarding against over fitting.
- When we have simple models and abundant data, we expect the generalization error to resemble the training error. When we work with more complex models and fewer examples, we expect the training error to go down but the generalization gap to grow.

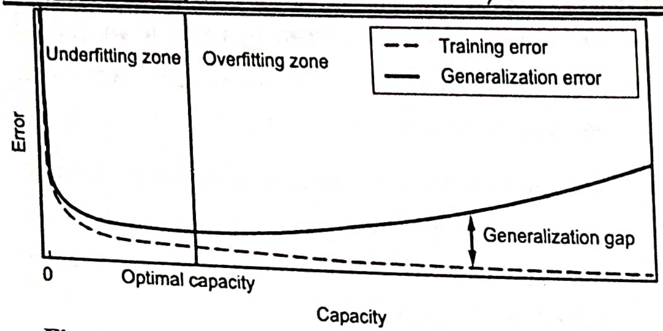


Fig. 1.8.2 : Training Error and Generalization Error

1.9 UNDERFITTING, OVERFITTING, BIAS AND VARIANCE TRADE OFF

- Let us consider that we are designing a machine learning model. A model is said to be a good machine learning model if it generalizes any new input data from the problem domain in a proper way. This helps us to make predictions in the future data, that data model has never seen.
- Now, suppose we want to check how well our machine learning model learns and generalizes to the new data. For that we have over fitting and under fitting, which are majorly responsible for the poor performances of the machine learning algorithms.
- Before diving further let's understands two important terms :
- Bias - Assumptions made by a model to make a function easier to learn. (The algorithms error rate on the training set is algorithms bias.)Variance - If you train your data on training data and obtain a very low error, upon changing the data and then training the same previous model you experience high error, this is variance. (How much worse the algorithm does on the test set than the training set is known as the algorithms variance.)

Under fitting

- A statistical model or a machine learning algorithm is said to have under fitting when it cannot capture the underlying trend of the data.
- Under fitting destroys the accuracy of our machine learning model. Its occurrence simply means that our model or the algorithm does not fit the data well enough.
- It usually happens when we have less data to build an accurate model and also when we try to build a linear model with a non-linear data. In such cases the rules of the machine learning model are too easy and flexible to be applied on such minimal data and therefore the model will probably make a lot of wrong predictions.
- Under fitting can be avoided by using more data and also reducing the features by feature selection.

In a nutshell, Under fitting - High bias and low variance

Techniques to reduce under fitting :

1. Increase model complexity
2. Increase number of features, performing feature engineering
3. Remove noise from the data.
4. Increase the number of epochs or increase the duration of training to get better results.

Over fitting

- A statistical model is said to be over fitted, when we train it with a lot of data . When a model gets trained with so much of data, it starts learning from the noise and inaccurate data entries in our data set. Then the model does not categorize the data correctly, because of too many details and noise.
- The causes of over fitting are the non-parametric and non-linear methods because these types of machine learning algorithms have more freedom in building the model based on the dataset and therefore they can really build unrealistic models.
- A solution to avoid over fitting is using a linear algorithm if we have linear data or using the parameters like the maximal depth if we are using decision trees.

In a nutshell, **Overfitting - High variance and low bias**

Techniques to reduce overfitting:

1. Increase training data.
2. Reduce model complexity.

3. Early stopping during the training phase (have an eye over the loss over the training period as soon as loss begins to increase stop training).
4. Ridge Regularization and Lasso Regularization
5. Use dropout for neural networks to tackle over fitting.

- Ideally, the case when the model makes the predictions with 0 error, is said to have a *good fit* on the data. This situation is achievable at a spot between over fitting and under fitting.
- In order to understand it we will have to look at the performance of our model with the passage of time, while it is learning from training dataset.
- With the passage of time, our model will keep on learning and thus the error for the model on the training and testing data will keep on decreasing.
- If it will learn for too long, the model will become more prone to overfitting due to the presence of noise and less useful details. Hence the performance of our model will decrease.
- In order to get a good fit, we will stop at a point just before where the error starts increasing. At this point the model is said to have good skills on training datasets as well as our unseen testing dataset.

☞ Bias-variance trade-off

So what is the right measure? Depending on the model at hand, a performance that lies between over fitting and under fitting is more desirable. This trade-off is the most integral aspect of Machine Learning model training. As we discussed, Machine Learning models fulfil their purpose when they generalize well. Generalization is bound by the two undesirable outcomes - high bias and high variance. Detecting whether the model suffers from either one is the sole responsibility of the model developer.

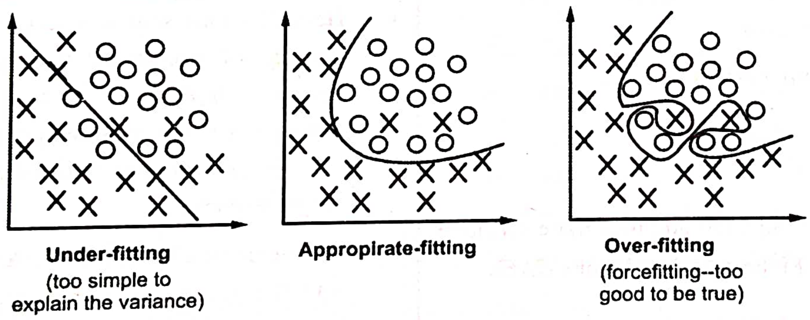


Fig. 1.9.1 : Underfitting and Overfitting

So what is the right measure? Depending on the model at hand, a performance that lies between overfitting and underfitting is more desirable. This trade-off is the most integral aspect of Machine Learning model training. As we discussed, Machine Learning models fulfil their purpose when they generalize well. Generalization is bound by the two undesirable outcomes — high bias and high variance. Detecting whether the model suffers from either one is the sole responsibility of the model developer.

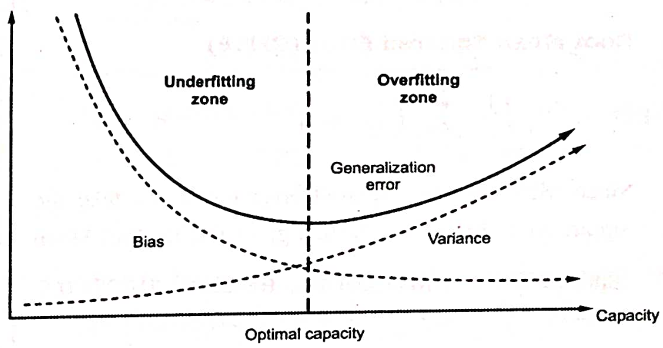


Fig. 1.9.2 : Bias variance Tradeoff as a function of model capacity

1.10 PERFORMANCE METRICS

1.10.1 Performance Metrics for Regression

Regression analysis is a subfield of supervised machine learning. It aims to model the relationship between a certain number of features and a continuous target variable. Following are the performance metrics used for evaluating a regression model :

1. Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} | y_i - \hat{y}_i |$$

Where y_i is the actual expected output and $\hat{y}_i$ is the model's prediction. It is the simplest evaluation metric for a regression scenario and is not much popular compared to the following metrics.

Say, y_i = [5,10,15,20] and $\hat{y}_i$ = [4.8,10.6,14.3,20.1]

Thus, $\text{MAE} = \frac{1}{4} * (|5 - 4.8| + |10 - 10.6| + |15 - 14.3| + |20 - 20.1|) = 0.4$

2. Mean Squared Error (MSE)

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Here, the error term is squared and thus more sensitive to outliers as compared to Mean Absolute Error (MAE).

Thus, $\text{MSE} = \frac{1}{4} * (|5 - 4.8| \hat{2} + |10 - 10.6| \hat{2} + |15 - 14.3| \hat{2} + |20 - 20.1| \hat{2}) = 0.225$

3. Root Mean Squared Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n} (y_j - \hat{y}_j)^2}$$

Since MSE includes squared error terms, we take the square root of the MSE, which gives rise to Root Mean Squared Error (RMSE). Thus, RMSE = (0.225) $\hat{0.5}$ = 0.474

4. R-Squared

$$R_2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

- R-squared is calculated by dividing the sum of squares of residuals (SS_{res}) from the regression model by the total sum of squares (SS_{tot}) of errors from the average model and then subtract it from 1. R-squared is also known as the Coefficient of Determination. It *explains the degree to which the input variables explain the variation of the output / predicted variable.*
- A R-squared value of 0.81, tells that the input variables explains 81 % of the variation in the output variable. The higher the R squared, the more variation is explained by the input variables and better is the model. Although, there exists a limitation in this metric, which is solved by the Adjusted R-squared.

5. Adjusted R-squared

$$\text{Adjusted } R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

- Here, N - total sample size (number of rows) and p - number of predictors (number of columns). The limitation of R-squared is that it will *either stay the same or increases with the addition of more variables, even if they do not have any relationship with the output variables.
- To overcome this limitation, Adjusted R-square comes into the picture as it penalizes you for adding the variables which do not improve your existing model. Hence, if you are building Linear regression on multiple variables, it is always suggested that you use Adjusted R-squared to judge the goodness of the model. If there exists only one input variable, R-square and Adjusted R squared are same.

1.10.2 Performance Metrics for Classification

Classification is the problem of identifying to which of a set of categories/classes a new observation belongs, based on the training set of data containing records whose class label is known. Following are the performance metrics used for evaluating a classification model :

- To understand different metrics, we must understand the Confusion matrix. A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

TN- True negatives (actual 0 predicted 0) & TP- True positives (actual 1 predicted 1)

FP- False positives (actual 0 predicted 1) & FN- False Negatives (actual 1 predicted 0)

- Consider the following values for the confusion matrix-

  True negatives (TN) = 300

  True positives (TP) = 500

  False negatives (FN) = 150

  False positives (FP) = 50

1. Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

- Accuracy is defined as the ratio of the number of correct predictions and the total number of predictions. It lies between [0,1]. In general, higher accuracy means a better model (TP and TN must be high).
- However, accuracy is not a useful metric in case of an imbalanced dataset (datasets with uneven distribution of classes). Say we have a data of 1000 patients out of which 50 are having cancer and 950 not, a dumb model which always predicts as no cancer will have the accuracy of 95%, but it is of no practical use since in this case, we want the number of False Negatives as a minimum. Thus, we have different metrics like recall, precision, F1-score etc.
- Thus, Accuracy using above values will be (500+300)/(500+50+150+300) = 800/1000 = 80%

2. Precision and Recall

$$\text{Precision} = \frac{TP}{TP + FP} \qquad \text{Recall} = \frac{TP}{TP + FN}$$

- Recall is a useful metric in case of cancer detection, where we want to minimize the number of False negatives for any practical use since we don't want our model to mark a patient suffering from cancer as safe.
- On the other hand, predicting a healthy patient as cancerous is not a big issue since, in further diagnosis, it will be cleared that he does not have cancer. Recall is also known as Sensitivity.
- Thus, Recall using above values will be 500/(500+150) = 500/650 = 76.92%
- Precision is useful when we want to reduce the number of False Positives. Consider a system that predicts whether the e-mail received is spam or not. Taking spam as a positive class, we do not want our system to predict non-spam e-mails (important e-mails) as spam, i.e., the aim is to reduce the number of False Positives.
- Thus, Precision using above values will be 500/(500+50) = 500/550 = 90.90%

3. Specificity

- Specificity is defined as the ratio of True negatives and True negatives + False positives. We want the value of specificity to be high. Its value lies between [0,1].

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

- Thus, Specificity using above values will be 300/(300+50) = 300/350 = 85.71%
- 4. F1-score

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

- F_1-score is a metric that combines both Precision and Recall and equals to the harmonic mean of precision and recall. Its value lies between [0,1] (more the value better the F1-score).
- Using values of precision=0.9090 and recall=0.7692, F1-score = 0.8333 = 83.33%

5. AUC-ROC

- AUC (Area Under The Curve)- ROC (Receiver Operating Characteristics) curve is one of the most important evaluation metrics for checking any classification model's performance.
- It is plotted between FPR (X-axis) and TPR (Y-axis). If the value is less than 0.5 than the model is even worse than a random guessing model.

$$\text{True Positive Rate (TPR)} = \frac{TP}{FP + FN}$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN}$$

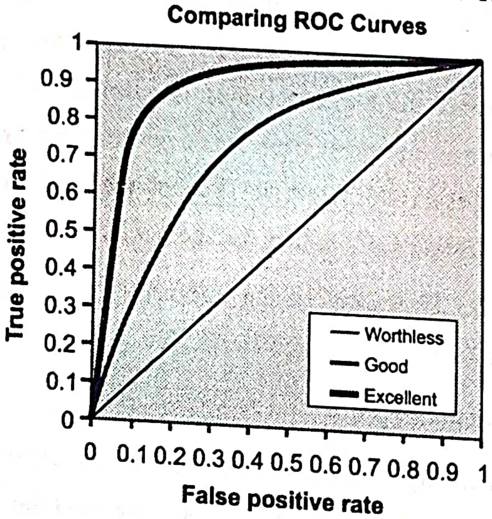


Fig. 1.10.1

1.10.3 Kappa Statistics

- It is a statistic which measures inter-rater agreement for qualitative (categorical) items. It is generally thought to be a more robust measure than simple percent agreement calculation, since k takes into account the agreement occurring by chance. Cohen's kappa measures the agreement between two raters who each classify N items into C mutually exclusive categories.
- Cohen's kappa coefficient is defined and given by the following function :

$$K = \frac{p_0 - p_e}{1 - p_e}$$

Where :

- o p_0 = relative observed agreement among raters.
- o p_e = the hypothetical probability of chance agreement.

p_0 and p_e are computed using the observed data to calculate the probabilities of each observer randomly saying each category. If the raters are in complete agreement then k = 1. If there is no agreement among the raters other than what would be expected by chance (as given by p_e), $k \leq 0$.

Example

Ex. 1.10.1 : Suppose that you were analyzing data related to a group of 50 people applying for a grant. Each grant proposal was read by two readers and each reader either said "Yes" or "No" to the proposal. Suppose the disagreement count data were as follows, where A and B are readers, data on the diagonal slanting left shows the count of agreements and the data on the diagonal slanting right, disagreements :

		B	
		Yes	No
A	Yes	20	5
	No	10	15

Calculate Cohen's kappa coefficient.

Soln. :

Note that there were 20 proposals that were granted by both reader A and reader B and 15 proposals that were rejected by both readers. Thus, the observed proportionate agreement is

$$p_o = \frac{20 + 15}{50} = 0.70$$

To calculate p_e (the probability of random agreement) we note that :

- Reader A said "Yes" to 25 applicants and "No" to 25 applicants. Thus reader A said "Yes" 50% of the time.
- Reader B said "Yes" to 30 applicants and "No" to 20 applicants. Thus reader B said "Yes" 60% of the time.

Using formula P (A and B) = P (A) × P (B) where P is probability of event occuring.

The probability that both of them would say "Yes" randomly is 0.50 × 0.60 = 0.30 and the probability that both of them would say "No" is 0.50 × 0.40 = 0.20. Thus the overall probability of random agreement is

$$p_e \; p_e = 0.3 + 0.2 = 0.5.$$

So now applying our formula for Cohen's Kappa we get:

$$k = 0.70 - \frac{0.50}{1 - 0.50} = 0.40$$

Chapter Ends...

□□□

MODULE 2

CHAPTER 2

Mathematical Foundation for ML

Syllabus

System of Linear equations, Norms, Inner products, Length of Vector, Distance between vectors, Orthogonal vectors, Symmetric Positive Definite Matrices, Determinant, Trace, Eigen values and vectors, Orthogonal Projections, Diagonalization, SVD and its applications.

2.1 SYSTEM OF LINEAR EQUATIONS

- A system of linear equations (or linear system) is a collection of one more linear equations involving the same **variables.**
- A linear system in three variables determines a collection of planes. The intersection point is the solution :

  For example,

$$\begin{aligned} x - y + 2z &= -1 \\ 3x + 2y - z &= 1 \\ -2x + y - 2z &= 1 \end{aligned}$$

  is a system of three equations in three variables x, y and z.

- A solution to a linear system is a assignment of values to the variables such that all the equations are simultaneously satisfied.

  A solution of the above **system** is :

$$(x, y, z) = (1, -2, -2)$$

- The word **system** impels that the equations are to be considered collectively, and not individually.

2.1.1 Linear Systems is a Part of Linear Algebra

- The theory of linear systems is the basis and a fundamental part of **linear algebra.**
- It is used in most part of modern mathematics.
- Computational algorithms for finding the solutions are an important part of **numerical linear algebra,** and play a prominent role in engineering, computer science, economics etc.
- A system of non-linear equations can be **approximated** by a linear system. This technique helps while making a mathematical model or computer simulation of a relatively **complex system.**

2.1.2 Examples of Linear Equations

(A) General form

A general system of m linear equations with n **unknowns** and **coefficients** can be written as :

$$a_{11} x_1 + a_{12} x_2 + \dots \dots + a_{1n} x_n + b_1 = 0$$

$$a_{21} x_1 + a_{22} x_2 + \dots \dots + a_{2n} x_n + b_2 = 0$$

$$\vdots$$

$$\vdots$$

$$a_{m1} x_1 + a_{m2} x_2 + \dots \dots + a_{mn} x_n + b_m = 0,$$

where $x_1, x_2, \dots \dots, x_n$ are unknowns, $a_{11}, a_{12}, \dots \dots, a_{mn}$ are coefficients of the systems such that

$a_{11} + a_{12} + \dots \dots + a_{1n} \neq 0$ and $b_1, b_2, \dots, b_m$ are constant terms.

- Generally the coefficients and unknowns are real or complex numbers.

(B) Vector Equation

- We can regard each unknown as a weight for a column vector in a linear combination :

$$x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \dots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = 0 \qquad \dots(1)$$

- The collection of all possible linear combinations of the vectors on the left - hand side is called their **span,** and the equations have a solution just when the right – hand side is within that span.

(C) Matrix Equation

- The vector equation is equivalent to a matrix equation of the form :
- AX = b, where A is an m × n matrix, X is a column vector with n entries and b is a column vector with m entries.

$$\text{i.e., } A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix},$$

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

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

2.1.3 Solution Set

- A solution of a linear systems is an assignment of values to the variables $x_1, x_2, \ldots\ldots, x_n$ such that each of the equations is satisfied.
- The set of all solutions is called the **solution set.**
- A linear system may possess any one of the three types of solutions :
  (1) The system has **infinitely** many solutions.
  (2) The system has a **unique** solution.
  (3) The system has **no solution.**

2.1.4 Geometric Interpretation

(1) A system involving two variables, say, x and y, each linear equation determines a line an XY – Plane. And the solution set is the intersection of these linear.

It is either a line, a single point or the empty set.

(2) For three variables, each linear equations determines a **plane in three - dimensional** space.

The solution set is the intersection of these planes. Thus the solution set may be a plane, a line a single point, or the empty set.

If there are three parallel planes, then they do not have a common point, and the solution set is empty.

If the planes pass through a point i.e., it has unique solution.

If the planes pass through two point then the planes pass through a line and in that case, it has infinitely many solutions.

(3) For **n-variables,** each linear equation determines a **hyperplane in n - dimensional space.**

The solution set is the intersection of the hyperplanes.

2.2 NORMS

- A norm is a way to measure the size of a vector, matrix or a tensor.
- In other words, norms are a class of functions that enable to quantify the magnitude of a vector.
- If norm of x is greater than 0, then x is not equal to zero vector and if norm is equal to 0, then x is a zero vector.
- Thus the norm of a vector is simply the square - root of the sum of each component squared.

2.2.1 The Norm and Value

- The difference between a value and a norm is that a value is general, where as norm specifies certain things that have to be done.
- Values can be operationalised in specifying norms norms are justified by underlying value.

2.2.2 L1 and L2 Norm

- The L1 norm is calculated as the sum of the absolute values of the vectors.
- The L2 norm is calculated on the square – root of the sum of the squared vector values.

2.2.3 (Mathematical) Definition of Norm

Given a vector space X over a subfield F of real number R (or complex number C), a **norm** on X is a real valued function $P : X \rightarrow R$ with the following properties :

- **(1) Subadditivity / Triangle inequality**

  $p(x + y) \leq p(x) + p(y)$ for all $x, y \in X$

- **(2) Absolute Homogeneity**

  $p(sx) \leq |s| p(x)$ for all $x \in X$ and all scalars s

  (where $|s|$ denotes the absolute value of a scalar s)

- **(3) Positive definiteness**

For all $x \in X$, if $p(x) = 0$, then $x = 0$

2.2.4 Semi Norm

- A seminorm on X is a function $p : X \rightarrow R$ that possesses properties (1) and (2).
- Thus every norm is a subnorm.

2.2.5 Equivalent Norms

- Let p and q be two norms (or seminorms) on a vector space X.
- Then p and q are called **equivalent,** if there exist two real constants a and b with $a > 0$ such that for every vector $x \in X$,

$$aq(x) \leq p(x) \leq bq(x)$$

Remark

Norm of a vector $x \in X$ is denoted by $\| x \|$.

2.2.6 Euclidean Norm

On n - dimensional Euclidean space R^n, the length of the vector

$$x = (x_1, x_2, \dots, x_n) \text{ is given as}$$

$$\| x \| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

- This is the **Euclidean norm** and it gives the distance from the origin.
- This operation is also referred to as "SRSS", which is "the square root of the sum of squares".

2.3 INNER PRODUCT

- An inner product is a generalisation of the **dot product.**
- In a vector space, it is a way to multiply vectors together, and the result of this multiplication is a scalar.
- For a real vector space, an inner product $<\cdot, \cdot>$ satisfies the following four properties.

Let u, v and w be vectors and α be a scalar, then

1. $< u + v, w > = < u, w > + < v, w >$
2. $< \alpha v, w > = \alpha < v, w >$
3. $< v, w > = < w, v >$
4. $< v, v > \geq 0$ and $< v, v > = 0$ if $v = 0$

Remark

- The forth condition is know as **positive definite** condition.

2.3.1 Examples of Inner Product

(1) The real numbers R, where the inner product is given by

$$< x, y > = xy$$

(2) The Euclidean space R^n , where the inner product is given by the dot product

$$< (x_1, x_2, \dots x_n), (y_1, y_2, \dots y_n) >$$

$$= x_1 y_1 + x_2 y_2 + \dots x_n y_n$$

(3) The vector space of real – valued functions whose domain is closed interval [a, b] with inner product.

$$< f, g > = \int_a^b f \cdot g \cdot dx = \int_a^b f(x) \cdot g(x) \cdot dx$$

2.3.2 Norm Properties

- Every inner product space induces a norm, called its canonical form, and is denoted by

$$\| x \| = \sqrt{< x, x >}$$

- With this norm, every inner product space becomes a **normed vector space.**
- With this norm, every inner product space becomes a **normed vector space.**
- Hence, every general property of normed vector space applies to inner product spaces.

In particular, the following properties are valid :

- **(i) Absolute homogeneity**

$$\| ax \| = | a | \| x \|$$

For every $x \in V$ and $a \in F$.

- **(ii) Triangle inequality**

$$\| x + y \| \leq \| x \| + \| y \|$$

For $x, y, \in V$.

- **(iii) Cauchy – Schwarz Inequality**

$$|< x, y >| \leq \| x \| \| y \|$$

For every x, y ∈ V ; and equality holds if and only if x and y are linearly dependent.

- **(iv) Parallelogram law**

  $\| x+y \|^2 + \| x-y \|^2 = 2[\| x \|^2 + \| y \|^2]$,

  For every x, y ∈ V ,

- **(v) Polarization Identity**

  $\| x+y \|^2 = \| x \|^2 + \| y \|^2 + 2\text{Re} < x, y >$,

  For every x, y ∈ V.

- **(vi) Ptolemy's Inequality**

  $\| x-y \| \| z \| + \| y-z \| \| x \| \geq \| x-z \| \| y \|$,

  For every x, y, z ∈ V.

- **(vii) Orthogonality**

  Two vectors x and y are said to be orthogonal, written as $x \perp y$, if their inner product is zero, that is $< x, y > = 0$.

2.4 LENGTH OF VECTOR

- The length of the vector is the distance of the vector in the standard position from the origin. The length of a vector is also known as the magnitude of the vector.
- Suppose $\bar{u} = xi + yj$, then be can have the length of $\bar{u}$, using the formula

$$|\bar{u}| = \sqrt{x^2 + y^2}$$

- If a vector has three components,

$$\bar{u} = xi + yj + zk,$$

Then $$|\bar{u}| = \sqrt{x^2 + y^2 + z^2}$$

2.4.1 Arc length of a Vector Function

- We can extend the concept of length to the length of the vector function $\bar{r}(t)$, within the interval [a, b], can be calculated as follows

(i) $\bar{r}(t) = < x(t), y(t) >$

$$\text{arc length} = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} \cdot dt$$

(ii) if $\bar{r}(t) = < x(t), y(t), z(t) >$, then

$$\text{arc length} = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} \cdot dt$$

2.4.2 Solved Examples on Length of Vector

Ex. 2.4.1 : The vector $\bar{u}$ has an initial point at P (– 2, 1, 0) and endpoint at Q (4, – 2, 3), what is the vector length.

☑ Soln. :

We have $\bar{u} = \overline{PQ}$

$$= < (4-(-2)), (-2-1), (3-0) > = < 6, -3, 3 >$$

$$\therefore |\bar{u}| = \sqrt{(6)^2 + (-3)^2 + (3)^2} = \sqrt{36+9+9}$$

$$= \sqrt{54} = 3\sqrt{6}$$

Ex. 2.4.2 : Calculate the arc length of the vector valued function $\bar{r}(t) = < 2\cos t, 2\sin t, 4t >$, if t is within the interval [0, 2π]

☑ Soln. :

Here we have to use the formula for the arc-length of the vector function,

$$\text{Arc length} = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} \cdot dt \quad ...(i)$$

Now, $x(t) = 2\cos t$,

$$\therefore \frac{dx}{dt} = x'(t) = -2\sin t$$

$$y(t) = 2\sin t$$

$$\frac{dy}{dt} = y'(t) = 2\cos t, = z(t) = 4t$$

$$\therefore \frac{dz}{dt} = z'(t) = 4$$

$$\therefore \text{Arc length} = \left|\frac{d\bar{r}}{dt}\right|$$

$$= \int_0^{2\pi} \sqrt{(-2\sin t)^2 + (2\cos t)^2 + (4)^2} \cdot dt$$

$$= \int_0^{2\pi} \sqrt{4(\sin^2 t + \cos^2 t) + 16} \cdot dt$$

$$= \int_0^{2\pi} \sqrt{20}\, dt = \sqrt{20}\,(2\pi) = (4\sqrt{5}\,\pi)$$

2.5 DISTANCE BETWEEN VECTORS

(i) In two dimension

Let $\bar{u} = x_1 i + y_1 j$ and

$\bar{v} = x_2 i + y_2 j$ then

Euclidean distance between two vectors $\bar{u}$ and $\bar{v}$ is given by

$$|\bar{u}-\bar{v}| = \sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$$

(ii) In three dimension

Let $\bar{u} = x_1 i + y_1 j + z_1 k$

and $\bar{v} = x_2 i + y_2 j + z_2 k$,

then $|\bar{u}-\bar{v}| = \sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2}$

2.6 ORTHOGONAL VECTORS

- **Orthogonality** is the generalisation of the notion of **perpendicularity** to the linear algebra of bilinear forms.
- Two elements u and v of a vector space with bilinear form B are **orthogonal** if B(u, v) = 0

2.6.1 Definitions of Orthogonal Vectors

(1) In geometry, two vectors are **orthogonal** it they are perpendicular, i.e., they form a right angle.

(2) Two vectors $\bar{u}$ and $\bar{v}$ in an inner product space V are orthogonal if their inner product is zero. The relationship is denoted by $\bar{u} \perp \bar{v}$.

(3) An orthogonal matrix is a matrix whose column vectors are orthonormal to each other.

(4) Two vector subspaces, A and B, of an inner product space V, are called orthogonal subspaces if each vector in A is orthogonal to each vector in B.

The largest subspace of V that is orthogonal to a given subspace is its orthogonal complements.

(5) In Euclidean space, two vectors are orthogonal if and only if their dot product is zero, i.e., they make an angle of (90°) or one of the vectors is zero.

Thus orthogonality of vectors is an extension of the concept of **perpendicular** vectors to any dimension

(6) The orthogonal complement of a subspace is the space of all vectors that are orthogonal to every vector in the subspace.

Remark

We note that the geometric concept of two planes being perpendicular does not correspond to the orthogonal complement, since in three dimensions a pair of vectors, one from each of a pair of perpendicular planes, might meet at any angle.

(7) In four - dimensional Euclidean space, the orthogonal complement of a line is the **hyper plane** and vice – versa and that of a plane is a plane.

2.7 SYMMETRIC POSITIVE DEFINITE MATRICES

A given matrix A is symmetric positive definite if the following conditions are satisfied.

(i) Every leading principal minor det (AK), is positive.

(ii) The eigen values of A are all positive.

Remark

(1) The first condition implies in particular that det (A) > 0, which also follows from the second condition since the determinant is the product of eigenvalues.

(2) We note that, just a determinant is positive does not imply that A is symmetric positive, e.g.

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}; \text{ here } |A| = 1$$

But A is not positive definite.

2.7.1 Properties of Symmetric Positive Definite

(i) A^{-1} is positive definite.

(ii) A has a unique symmetric positive definite square root X, where a square root is a matrix X such that $X^2 = A$.

(iii) A has a unique factorisation ,

(called as cholesky factorisation)

i.e. $A = R^T R$, where R is upper triangular with positive diagonal elements.

2.7.2 Examples of Symmetric Positive Definite Matrices

(1) Hilbert matrix of order 4×4

$$H_4 = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{bmatrix}$$

(2) Pascal Matrix

$$P_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix}$$

(3) A 4×4 symmetric positive definite matrix that was used as a **test – matrix** of digital computing is the Wilson – matrix

$$W = \begin{bmatrix} 5 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix}$$

Remark

Computing the eigen values and checking their positivity is reliable method to check whether a matrix is symmetric positive definite.

2.7.3 Another Definition of Positive Definite Symmetric Matrix

Let A be $n \times n$ matrix. A is symmetric positive definite if :

(i) It is symmetric (A is equal to its transpose A^T)

And

(ii) $X^T A X > 0$ for all non – zero rector X (X is non – zero column matrix of order $n \times 1$)

Remark

By making particular choices of X. We can derive the inequalities :

$$a_{ii} > 0, \text{ for all } i \text{ ; and}$$

$$a_{ij} < \sqrt{a_{ii} \cdot a_{jj}} \text{ for all } i \neq j$$

- But these inequalities are not sufficient for positive definiteness.
- For example, the matrix

$$A = \begin{bmatrix} 1 & \frac{3}{4} & 0 \\ \frac{3}{4} & 1 & \frac{3}{4} \\ 0 & \frac{3}{4} & 1 \end{bmatrix}$$

- Satisfies all the above inequalities but

$$X^T AX < 0 \text{ for } X = \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix}$$

- Thus a sufficient condition for a symmetric matrix to be positive definite is that it has,

(i) Positive diagonal elements ; and

(ii) It is diagonally dominant,

that is $a_{ii} > \sum_{j \neq i} |a_{ij}|$ for all i.

2.8 DETERMINANT

- Determinant is a scalar value. It is a function of the entries of a **square** matrix. It allows characterising some properties of the matrix and the **linear map** represented by the matrix.
- In particular, the determinant is non - zero if and only if the matrix is invertible. The linear map represented by the matrix is an **isomorphism.**
- The determinant of a product of matrices is the product of their determinants. The determinant of a matrix A is denoted by det (A), det A or | A |.
- In the case of 2×2 matrix, the determinant can be defined as

$$|A| = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

- For a 3×3 matrix A, the determinant is

$$|A| = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

$$= a\begin{bmatrix} e & f \\ h & i \end{bmatrix} - b\begin{bmatrix} d & f \\ g & i \end{bmatrix} + c\begin{bmatrix} d & e \\ g & h \end{bmatrix}$$

$= aei + bfg + cdh - ceg - bdi - afh$

- Each determinant of a 2×2 matrix in this equation is called a **minor** of the matrix A.
- This method can be extended for the determinant of an $n \times n$ matrix, and it is known as Laplace expansion.

2.8.1 Uses of Determinants

(i) Determinants occur throughout mathematics. For example, a matrix is often used to represent the **coefficients** in a system of linear equations, and determinants can be used to solve these equations. (cramer's rule)

(ii) Determinants are used for defining the characteristic polynomial of a matrix, whose roots are the eigenvalues.

(ii) Jacobian determinant is used for change of variables in multiple integrals.

(iv) The absolute value of (ad – bc) is the area of the parallelogram and thus represents the scale factor by which areas are transformed by A.

2.8.2 Properties of the Determinant

1. det (I) = 1, where I is an identity matrix.
2. The determinant is alternating.
   Whenever two columns of a matrix are identical, its determinant is 0.
3. The determinant is a **homogeneous function,**
   i.e. det. $(CA) = C^n \det(A)$ (for $n \times n$ matrix A)
4. Interchanging any pair of columns of a matrix multiplies its determinant by (– 1).
5. If some column can be expressed as a linear combination of the other columns (i.e. the columns of the matrix form a linearly dependent set), the determinant is 0.
6. Adding a scalar multiple of one column to another column does not change the value of the determinant.
7. If A is a triangular matrix, i.e. $a_{ij} = 0$, whenever $i > j$ or if $i < j$, then its determinant equals the product of the diagonal entries :

$$\det(A) = a_{11}\, a_{22} \ldots a_{nn} = \prod_{i=1}^{n} a_{ii}$$

8. The determinant of the transpose of A equals the determinant of A,
   i.e. $\det(A^T) = \det(A)$

2.9 TRACE

- The trace of a square matrix A, denoted tr (A) is defined to be the sum of elements on the main diagonal, of A.
- The trace is only defined for a square matrix $(n \times n)$.

Thus,

$$\text{tr}(A) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \ldots + a_{nn}$$

- where a_{ii} denotes the entry on i^{th} row and i^{th} column of A.

2.9.1 Properties

(1) The trace is a linear mapping :

i.e. $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$

$\text{tr}(CA) = C\,\text{tr}(A)$

for all square matrices A and B and all scalars C.

(2) A matrix and its transpose have the same trace :

$\text{tr}(A) = \text{tr}(A^T)$

(3) Let A be any $(n \times n)$ real or complex matrix, then

$$\text{tr}(A) = \sum_{i=1}^{n} \lambda_i \text{ ; where } \lambda_1, \lambda_2, \ldots, \lambda_n$$

are the eigenvalues of A (counted with multiplicity).

(4) The trace of $n \times n$ identity matrix is the dimension of the space ; namely n.

$\text{tr}(I_n) = n$

(5) The trace of a Hermitian matrix is real, because the elements on the diagonal are real.

2.10 INTRODUCTION TO EIGEN VALUE AND EIGEN VECTOR

The concept of eigen value and eigen vector have a considerable theoretical interest and wide-ranging application. This concept is crucial in solving systems of differential equations, analyzing population growth models, and calculating powers of matrices (in order to define the exponential matrix). Other areas such as physics, sociology, biology, economics and statistics have focused considerable attention on "eigenvalues" and "eigenvectors" their applications and computations. Eigen values and eigen vectors also plays an important role in the study of vibration of beams, probability (Markov process), quantum mechanism etc.

Eigen values are certain numbers associated with square matrices. These are fundamental in applications as diverse as population dynamics, electrical networks, chemical reactions etc.

2.11 PREREQUISITE

[I] Cramer's Rule : Consider two different equations,

$$a_{11}x + a_{12}y + a_{13}z = 0$$

$$a_{21}x + a_{22}y + a_{23}z = 0$$

then by Cramer's rule,

$$\frac{x}{\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}} = \frac{-y}{\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}} = \frac{z}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} = k$$

[II] Synthetic Division :

In the process of finding roots of polynomial equation f(D) = 0 by synthetic division method, the first root is obtained by trial and error method.

e.g. $D^4 - 2D^3 - 3D^2 + 4D + 4 = 0$

$$\begin{array}{c|ccccc} 1 & 1 & -2 & -3 & 4 & 4 \\ & \times & 1 & -1 & -4 & 0 \\ \hline & 1 & -1 & -4 & 0 & \boxed{4} \end{array}$$

← These are the coefficients. Write down the coefficient of the power of D in order. [Adding the missing power of D by zero]

If we get here zero then D = 1 is one of the root of equation and if not zero then try for next.

$$\begin{array}{c|ccccc} -1 & 1 & -2 & -3 & 4 & 4 \\ & \times & -1 & 3 & 0 & -4 \\ \hline & 1 & -3 & 0 & 4 & \boxed{0} \end{array} \longrightarrow$$

$\therefore$ D = – 1 is one of the root and it gives

$D^3 - 3D^2 + 0D + 4 = 0$ i.e. $D^3 - 3D^2 + 4 = 0$

Again do the same procedure

$$\begin{array}{c|cccc} -1 & 1 & -3 & 0 & 4 \\ & \times & -1 & 4 & -4 \\ \hline & 1 & -4 & 4 & \boxed{0} \end{array}$$

gives again D = – 1

And $D^2 - 4D + 4 = 0$ and so on.

2.12 EIGEN VALUES AND EIGEN VECTORS

If A is a square matrix, then a vector X is said to be Eigen vector (or characteristics vector) of the matrix A if there exist a number λ such that,

$$AX = \lambda X$$

If A is a square matrix of order n then X is a column matrix of order n. The number λ is known as **Eigen values or characteristic roots or latent roots.**

$$AX = \lambda X$$

$$AX = \lambda IX$$

$$AX - \lambda IX = 0$$

$$[A - \lambda I]X = 0$$

If A is a square matrix then the characteristic polynomial of matrix A is $|A - \lambda I|$. If A is a square matrix then the characteristic equation is $|A - \lambda I| = 0$.

e.g. If
$$A = \begin{bmatrix} a_{11} & a_{12} & -- & -- & a_{1n} \\ a_{21} & a_{22} & -- & -- & a_{2n} \\ -- & -- & -- & -- & -- \\ -- & -- & -- & -- & -- \\ -- & -- & -- & -- & -- \\ a_{n1} & a_{n2} & -- & -- & a_{nn} \end{bmatrix}$$

then $|A-\lambda I| = \begin{vmatrix} a_{11}-\lambda & a_{12} & -- & -- & a_{1n} \\ a_{21} & a_{22}-\lambda & -- & -- & a_{2n} \\ -- & -- & -- & -- & -- \\ -- & -- & -- & -- & -- \\ -- & -- & -- & -- & -- \\ a_{n1} & a_{n2} & -- & -- & a_{nn}-\lambda \end{vmatrix}$

is a characteristic polynomial of matrix A.

$$|A-\lambda I| = \begin{vmatrix} a_{11}-\lambda & a_{12} & -- & -- & a_{1n} \\ a_{21} & a_{22}-\lambda & -- & -- & a_{2n} \\ -- & -- & -- & -- & -- \\ -- & -- & -- & -- & -- \\ -- & -- & -- & -- & -- \\ a_{n1} & a_{n2} & -- & -- & a_{nn}-\lambda \end{vmatrix} = 0$$

Gives characteristic equation of A.

The roots of characteristics equation gives Eigen Values or Characteristic Roots or Latent Roots

Note : (i) If A is a square matrix of order 2 then $|A-\lambda I| = 0$ is $\lambda^2 - S_1\lambda + |A| = 0$

Where S_1 = sum of the principal diagonal elements i.e. $[a_{11} + a_{22} + a_{33}]$

(ii) If A is a square matrix of order 3 then $|A-\lambda I| = 0$ is $\lambda^3 - S_1\lambda^2 + S_2\lambda - |A| = 0$

Where S_1 = sum of the principal diagonal elements i.e. $[a_{11} + a_{22} + a_{33}]$

S_2 = sum of minors of the principal diagonal elements. $= \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$

(iii) Spectrum of the matrix is a set of Eigen values

(iv) The trace of a square matrix is the sum of its diagonal elements i.e. Trace of a matrix

$A = a_{11} + a_{22} + a_{33}$

(v) The largest of the absolute values of the Eigen values of A is called the spectral radius of A.

(vi) If λ is a Eigen value of the matrix A then matrix $[A - \lambda I]$ is singular.

(vii) Degree of characteristics equation of matrix A is equal to order of matrix A.

(viii) Eigen values of diagonal matrix, upper triangular matrix, and lower triangular matrix are the diagonal elements.

e.g. Diagonal matrix

$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$ $\therefore$ Eigen values are $\lambda = 2, 1, 5$

Upper triangular matrix :

$\begin{bmatrix} 1 & 2 & 5 \\ 0 & 3 & -2 \\ 0 & 0 & 2 \end{bmatrix}$ $\therefore$ Eigen values are $\lambda = 1, 3, 2$

Lower triangular matrix

$\begin{bmatrix} 2 & 0 & 0 \\ -3 & 5 & 0 \\ 0 & 7 & 4 \end{bmatrix}$ $\therefore$ Eigen values are $\lambda = 2, 5, 4$

2.13 NON-REPEATED EIGEN VALUES FOR UNSYMMETRIC MATRIX

Ex. 2.13.1

Find the Eigen values and Eigen vectors of the matrix.

$A = \begin{bmatrix} 14 & -10 \\ 5 & -1 \end{bmatrix}$

☑ **Soln. :**

Step I : Given matrix is,

$$A = \begin{bmatrix} 14 & -10 \\ 5 & -1 \end{bmatrix}$$

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

Since,

$$A - \lambda I = \begin{bmatrix} 14 & -10 \\ 5 & -1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

→ **Using Scalar multiplication**

$$= \begin{bmatrix} 14 & -10 \\ 5 & -1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

→ **Using subtraction of Matrices**

$$[A - \lambda I] = \begin{bmatrix} 14-\lambda & -10 \\ 5 & -1-\lambda \end{bmatrix}$$

$\therefore$ Characteristic equations is,

$$\therefore \begin{vmatrix} 14-\lambda & -10 \\ 5 & -1-\lambda \end{vmatrix} = 0$$

$$\begin{vmatrix} \boxed{14-\lambda} & \boxed{-10} \\ \boxed{5} & \boxed{-1-\lambda} \end{vmatrix} = 0$$

By solving determinant

$$(14-\lambda)(-1-\lambda)+(-5)(-10) = 0$$

$$\therefore (1+\lambda)(\lambda-14)+50=0 \quad [\text{or } \lambda^3 - S_1\lambda^2 + S_2\lambda - |A| = 0]$$

$$\lambda^2 - 13\lambda - 14 + 50 = 0 \Rightarrow \lambda^2 - 13\lambda + 36 = 0$$

Which is a quadratic equation,

$$\therefore (\lambda-4)(\lambda-9) = 0 \quad \Big| \quad \text{Factors} : (-4)+(-9) = -13$$

$$\lambda - 4 = \text{and } \lambda - 9 = 0 \quad \Big| \quad (-4)\times(-9) = 36$$

$$\therefore \quad \lambda = 4, 9$$

Hence **Eigen values are $\lambda_1 = 4, \lambda_2 = 9$** ✓ **...Ans.**

$$\left\{\begin{array}{l} \textbf{Verification} : \lambda_1+\lambda_2 = a_{11}+a_{22} \text{ and } |A| = \lambda_1 \cdot \lambda_2 \\ 4+9 = 14-1 = 13 \text{ and } |A| = 4\times 9 = 36 \end{array}\right\}$$

Step II : Eigen Vectors

To find Eigen vector $X = \begin{bmatrix} x \\ y \end{bmatrix}$ corresponding to Eigen values $\lambda_1 = 4, \lambda_2 = 9$

Consider $[A-\lambda I]\ X = 0$

$$\begin{bmatrix} 14-\lambda & -10 \\ 5 & -1-\lambda \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{0} \qquad ...(1)$$

$$\because \mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

For $\lambda_1 = 4$, Put $\lambda = 4$ in Equation (1), it gives,

$$\therefore \begin{bmatrix} 14-4 & -10 \\ 5 & -1-4 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{0}$$

$$\begin{array}{c} \begin{bmatrix} 10 & -10 \\ 5 & -5 \end{bmatrix} \begin{array}{l} \leftarrow R_1 \\ \leftarrow R_2 \end{array} \end{array} \quad \begin{array}{c} \begin{bmatrix} x \\ y \end{bmatrix} \\ \uparrow \\ C_1 \end{array} = \mathbf{0}$$

➔ **Using Matrix Multiplication**

$$\therefore \begin{bmatrix} R_1C_1 \\ R_2C_1 \end{bmatrix} = \mathbf{0}$$

$$R_1C_1 = (10)(x) - (-10)(y) = 10x - 10y$$

$$R_1C_1 = (5)(x) - (-5)(y) = 5x + 5y$$

$$\begin{bmatrix} 10x - 10y \\ 5x - 5y \end{bmatrix} = \mathbf{0}$$

By Equating corresponding elements of both side matrices, we get,

$$10x - 10y = 0$$

$$5x - 5y = 0$$

$$\Rightarrow x = y \textbf{ ...(choose } \mathbf{x = 1} \textbf{ to get smallest integers)}$$

$$\therefore \text{For Eigen value } \lambda_1 = 4, \text{ Eigen vector } X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

For $\lambda_2 = 9$, Put $\lambda = 9$ in Equation (1), it gives,

$$\begin{bmatrix} 14-9 & -10 \\ 5 & -1-9 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{0}$$

$$\begin{bmatrix} 5 & -10 \\ 5 & -10 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{0}$$

$$\begin{bmatrix} 5x - 10y \\ 5x - 10y \end{bmatrix} = \mathbf{0}$$

$$5x - 10y = 0$$

$$5x - 10y = 0 \Rightarrow 5x = 10y \Rightarrow x = \frac{10}{5}y$$

$$x = 2y \textbf{ ...(choose xy = 1 to get smallest integers)}$$

$$\therefore \text{For Eigen value } \lambda_2 = 9 \text{ Eigen vector } X_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Hence Eigen vectors are

$$\mathbf{X_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\lambda_1 = 4} ; \mathbf{X_2} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}_{\lambda_2 = 9} ✓ \qquad \textbf{...Ans.}$$

2.14 EXAMPLES ON NON-REPEATED EIGEN VALUES FOR SYMMETRIC MATRIX

Ex. 2.14.1

Find the Eigen value and Eigen vectors for the following Matrix : $\begin{bmatrix} 3 & -1 & 1 \\ -1 & 5 & -1 \\ 1 & -1 & 3 \end{bmatrix}$

☑ **Soln. :**

Step I : Given matrix is,

$$A = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 5 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

The characteristic equation of the given matrix A is,

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

$$\text{Since, } A - \lambda I = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 5 & -1 \\ 1 & -1 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

➔ **Using Scalar Multiplication**

$$A - \lambda I = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 5 & -1 \\ 1 & -1 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$$

➔ **Using Subtraction of Matrices**

$$[A - \lambda I] = \begin{bmatrix} 3-\lambda & -1 & 1 \\ -1 & 5-\lambda & -1 \\ 1 & -1 & 3-\lambda \end{bmatrix}$$

∴ Characteristic equations is,

$$\begin{vmatrix} 3-\lambda & -1 & 1 \\ -1 & 5-\lambda & -1 \\ 1 & -1 & 3-\lambda \end{vmatrix} = 0$$

➔ **Using Standard formula**

$$\lambda^3 - S_1 \lambda^2 + S_2 \lambda - |A| = 0 \qquad \text{...(1)}$$

Where S_1 and S_2 are sum of the minors of order 1 and 2 along the principal diagonal respectively.

S_1

$$= \begin{bmatrix} 3 & -1 & 1 \\ -1 & 5 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

S_1 = sum of diagonal elements

$= \quad 3 + 5 + 3 = 11$ (Addition of diagonal Elements)

$S_2 = \quad$ [Minor of 3] + [Minor of 5] + [Minor of 3]

...(Sum of Minors of diagonal elements)

$$S_2 = \begin{vmatrix} \cancel{3} & \cancel{-1} & \cancel{1} \\ \cancel{-1} & 5 & -1 \\ \cancel{1} & -1 & 3 \end{vmatrix} + \begin{vmatrix} 3 & \cancel{-1} & 1 \\ \cancel{-1} & \cancel{5} & \cancel{-1} \\ 1 & \cancel{-1} & 3 \end{vmatrix} + \begin{vmatrix} 3 & -1 & \cancel{1} \\ -1 & 5 & \cancel{-1} \\ \cancel{1} & \cancel{-1} & \cancel{3} \end{vmatrix}$$

$$= \begin{vmatrix} 5 & -1 \\ -1 & 3 \end{vmatrix} + \begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix} + \begin{vmatrix} 3 & -1 \\ -1 & 5 \end{vmatrix}$$

$$= [\underbrace{(5)(3)}_{15} - \underbrace{(-1)(-1)}_{1}] + [\underbrace{(3)(3)}_{9} - \underbrace{(1)(1)}_{1}] + [\underbrace{(3)(5)}_{15} - \underbrace{(-1)(-1)}_{1}]$$

$= [15 - 1] + [9 - 1] + [15 - 1]$

$S_2 \quad = 14 + 8 + 14 = 36$

$|A| =$

$$\begin{vmatrix} 3 & -1 & 1 \\ -1 & 5 & -1 \\ 1 & -1 & 3 \end{vmatrix}$$

$= 3 \times$ [Minor of 3] $- (-1) \times$ [Minor of (-1)] $+ (1) \times$ [Minor of 1]

$$= 3 \times \begin{vmatrix} \cancel{3} & \cancel{-1} & \cancel{1} \\ \cancel{-1} & 5 & -1 \\ \cancel{1} & -1 & 3 \end{vmatrix} + 1 \times \begin{vmatrix} \cancel{3} & \cancel{-1} & \cancel{1} \\ -1 & \cancel{5} & -1 \\ 1 & \cancel{-1} & 3 \end{vmatrix} + 1 \times \begin{vmatrix} \cancel{3} & \cancel{-1} & \cancel{1} \\ -1 & 5 & \cancel{-1} \\ 1 & -1 & \cancel{3} \end{vmatrix}$$

$$= 3 \times \begin{vmatrix} 5 & -1 \\ -1 & 3 \end{vmatrix} + 1 \times \begin{vmatrix} -1 & -1 \\ 1 & 3 \end{vmatrix} + 1 \begin{vmatrix} -1 & 5 \\ 1 & -1 \end{vmatrix}$$

$$= 3 \times [\underbrace{(5)(3)}_{15} - \underbrace{(-1)(-1)}_{1}] + 1 \times [\underbrace{(-1)(3)}_{-3} - \underbrace{(1)(-1)}_{-1}] + 1 \times [\underbrace{(-1)(-1)}_{1} - \underbrace{(1)(5)}_{5}]$$

$= 3 \times [15 - 1] + 1 \times [-3 + 1] + 1 \times [1 - 5]$

$= 3(14) + 1(-2) + 1(-4)$

$|A| = 42 - 2 - 4 = 36$

Substitute the values of S_1, S_2 and $|A|$ in Equation (1), we get,

$$\lambda^3 - 11\lambda^2 + 36\lambda - 36 = 0$$

➔ **Using Synthetic division**

$$\begin{array}{c|cccc} 2 & 1 & -11 & 36 & -36 \\ & - & 2 & -18 & 36 \\ \hline & 1 & -9 & 18 & \boxed{0} \end{array}$$

$\lambda = 2, \quad (\lambda^2 - 9\lambda + 18)$

$(\lambda - 6)(\lambda - 3) = 0$

$\therefore \qquad \lambda = 2, 3, 6$

Hence the Eigen values are $\lambda_1 = 2, \lambda_2 = 3, \lambda_3 = 6$ ✓ ...Ans.

{Verification , $S_1 = \lambda_1 + \lambda_2 + \lambda_3$; $|A| = \lambda_1 \lambda_2 \lambda_3$**}**

Step II: Eigen Vectors :

To find Eigen vector $X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ corresponding to Eigen values $\lambda_1 = 2$, $\lambda_2 = 3$; $\lambda_3 = 6$

Consider, $[A - \lambda I] X = 0$

$$\begin{bmatrix} 3-\lambda & -1 & 1 \\ -1 & 5-\lambda & -1 \\ 1 & -1 & 3-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \qquad \ldots(2)$$

$$\because 0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

For $\lambda_1 = 2$, Put $\lambda = 2$ in Equation (2), it gives,

$$\begin{bmatrix} 3-2 & -1 & 1 \\ -1 & 5-2 & -1 \\ 1 & -1 & 3-2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

→ **Using Matrix Multiplication**

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

By Equating corresponding elements of both side matrices, we get,

$$\left.\begin{aligned} x - y + z &= 0 \\ -x + 3y - z &= 0 \\ x - y + z &= 0 \end{aligned}\right\} \text{Consider these}$$

→ **Using Cramer's Rule**

$$\frac{x}{\begin{vmatrix} 3 & -1 \\ -1 & 1 \end{vmatrix}} = \frac{y}{\begin{vmatrix} -1 & -1 \\ 1 & 1 \end{vmatrix}} = \frac{z}{\begin{vmatrix} -1 & 3 \\ 1 & -1 \end{vmatrix}} = k$$

$$\frac{x}{3-1} = \frac{-y}{-1+1} = \frac{z}{1-3} = k$$

$$\frac{x}{2} = \frac{-y}{0} = \frac{z}{-2} = k$$

$$\therefore \quad x = 2k, \quad y = 0, \quad z = -2k$$

$$\ldots\left(\text{Choose } k = \frac{1}{2}, \text{ to get smallest integer}\right)$$

$\therefore$ For eigen values $\lambda_2 = 2$, Eigen vector is $X_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$

For $\lambda_2 = 3$, Put $\lambda = 3$ in Equation (2), it gives,

$$\begin{bmatrix} 3-3 & -1 & 1 \\ -1 & 5-3 & -1 \\ 1 & -1 & 3-3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

→ **Using matrix multiplication**

$$\begin{bmatrix} 0x - y + z \\ -x + 2y - 2 \\ x - y + 0z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

By Equating corresponding elements of both side matrices, we get,

$$\left.\begin{aligned} 0x - y + z &= 0 \\ -x + 2y - z &= 0 \\ x - y + 0z &= 0 \end{aligned}\right\} \text{Consider these}$$

→ **Using Cramer's Rule**

$$\frac{x}{\begin{vmatrix} 2 & -1 \\ -1 & 0 \end{vmatrix}} = \frac{-y}{\begin{vmatrix} -1 & -1 \\ 1 & 0 \end{vmatrix}} = \frac{z}{\begin{vmatrix} -1 & 2 \\ 1 & -1 \end{vmatrix}} = k$$

$$\frac{x}{0-1} = \frac{-y}{0+1} = \frac{z}{1-2} = k$$

$$\frac{x}{-1} = \frac{-y}{1} = \frac{z}{-1} = k$$

$$\Rightarrow x = -k, \; y = -k, \; z = -k$$

$$\ldots(\text{Choose } k = -1, \text{ to get smallest integer})$$

$\therefore$ For eigen values $\lambda_2 = 3$, Eigen vector is $X_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

For $\lambda_3 = 6$, Put $\lambda = 6$ in Equation (2), it gives,

$$\begin{bmatrix} 3-6 & -1 & 1 \\ -1 & 5-6 & -1 \\ 1 & -1 & 3-6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

→ **Using Matrix Multiplication**

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

By Equating corresponding elements of both side matrices, we get,

$$\left.\begin{aligned} -3x - y + z &= 0; \\ -x - y - z &= 0; \\ x - y - 3z &= 0 \end{aligned}\right\} \text{Consider these}$$

➔ **Using Cramer's Rule**

$$\frac{x}{\begin{vmatrix} -1 & -1 \\ -1 & -3 \end{vmatrix}} = \frac{-y}{\begin{vmatrix} -1 & -1 \\ 1 & -3 \end{vmatrix}} = \frac{z}{\begin{vmatrix} -1 & -1 \\ 1 & -1 \end{vmatrix}} = k$$

$$\frac{x}{3-(1)} = \frac{-y}{3-(-1)} = \frac{z}{(1-(-1))} = k$$

$$\frac{x}{2} = \frac{-y}{4} = \frac{z}{2} = k \Rightarrow \quad x = 2k, \quad y = -4k, \quad z = 2k$$

$$\ldots\left(\text{Choose } k = \frac{1}{2} \text{ to get lowest integers}\right)$$

$\therefore$ For Eigen values $\lambda_3 = 6$, Eigen vector is $X_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$

Hence Eigen vectors corresponding to Eigen values are ;

$$\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}_{\lambda=2}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{\lambda=3}, \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}_{\lambda=6}$$ ✓...Ans.

2.15 EXAMPLE ON REPEATED EIGEN VALUES FOR NON-SYMMETRIC MATRIX

Ex. 2.15.1

Find Eigen values and Eigen vectors for the following matrix : $A = \begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix}$

☑ **Soln. :**

Step I: Given matrix is,

$$A = \begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix}$$

The characteristic equation is,

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

Since, $A - \lambda I = \begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

➔ **Using Scalar Multiplication**

$$= \begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$$

➔ **Using Subtraction of Matrices**

$$[A - \lambda I] = \begin{bmatrix} -9-\lambda & 4 & 4 \\ -8 & 3-\lambda & 4 \\ -16 & 8 & 7-\lambda \end{bmatrix}$$

$\therefore$ Characteristic equations is,

$$\begin{vmatrix} -9-\lambda & 4 & 4 \\ -8 & 3-\lambda & 4 \\ -16 & 8 & 7-\lambda \end{vmatrix} = 0$$

➔ **Using Standard formula**

$$\lambda^3 - S_1\lambda^2 + S_2\lambda - |A| = 0$$

Where S_1 and S_2 are sum of the minors of order 1 and 2 along the principal diagonal respectively.

$$S_1 = \begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix}$$ ← S_1 = Sum of Diagonal elements

$S_1 = -9 + 3 + 7 = 1$ (Addition of diagonal Elements)

S_2 = [Minor of (– 9)] + [Minor of 3] + [Minor of 7]

$$= \begin{vmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{vmatrix} + \begin{vmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{vmatrix} + \begin{vmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{vmatrix}$$

(Sum of Minors of diagonal elements)

$$= \begin{vmatrix} 3 & 4 \\ 8 & 7 \end{vmatrix} + \begin{vmatrix} -9 & 4 \\ -16 & 7 \end{vmatrix} + \begin{vmatrix} -9 & 4 \\ -8 & 3 \end{vmatrix}$$

$$= [\underbrace{(3)(7)}_{21} - \underbrace{(8)(4)}_{32}] + [\underbrace{(-9)(7)}_{-63} - \underbrace{(-16)(4)}_{-64}] + [\underbrace{(-9)(3)}_{-27} - \underbrace{(-8)(4)}_{-32}]$$

$$= [21 - 32] + [-63 + 64] + [-27 + 32]$$

$$S_2 = -11 + 1 + 5 = -5$$

$$|A| = \begin{vmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{vmatrix}$$

$= (-9) \times$ [Minor of (– 9)] – 4 × [Minor of 4] + 4 × [Minor of 4]

$$= (-9) \times \begin{vmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{vmatrix} - 4 \times \begin{vmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{vmatrix} + 4 \times \begin{vmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{vmatrix}$$

$$= (-9) \times \begin{vmatrix} 3 & 4 \\ 8 & 7 \end{vmatrix} - 4 \times \begin{vmatrix} -8 & 4 \\ -16 & 7 \end{vmatrix} + 4 \times \begin{vmatrix} -8 & 3 \\ -16 & 8 \end{vmatrix}$$

$= (-9) \times [(3)(7) - (8)(4)] - 4 \times [(-8)(7) - (16)(4)]$
$\quad + 4 \times [(-8)(8) - (-16)(3)]$
$= (-9)[21 - 32] - 4 \times [-56 + 64] + 4 \times [-64 + 48]$
$= -9(-11) - 4(8) + 4(-16) = 99 - 32 - 64$

$|A| = 3$

Substitute the values of S_1, S_2 and $|A|$ in Equation (1), we get

$\therefore \lambda^3 - \lambda^2 - 5\lambda - 3 = 0$

➔ Using Synthetic Division

$$\begin{array}{r|rrrr} -1 & 1 & -1 & -5 & -3 \\ & - & -1 & 2 & 3 \\ \hline & 1 & -2 & -3 & \boxed{0} \end{array}$$

$\lambda = -1, \lambda^2 - 2\lambda - 3 = 0$

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

$\therefore \quad \lambda = -1, -1, 3$

Hence Eigen value are $\lambda_1 = 3, \lambda_2 = -1, \lambda_3 = -1$ ✓...Ans.

{Verification : $S_1 = \lambda_1 + \lambda_2 + \lambda_3$ and $|A| = \lambda_1 . \lambda_2 . \lambda_3$}

Step II : Eigen vectors :

To find Eigen vector $X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ corresponding to Eigen values, $\lambda_1 = 3, \quad \lambda_2 = \lambda_3 = 1$

Consider, $[A - \lambda I] X = 0$

$$\begin{bmatrix} -9-\lambda & 4 & 4 \\ -8 & 3-\lambda & 4 \\ -16 & 8 & 7-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0}$$

$$\because \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad ...(1)$$

For $\lambda_1 = 3$, Put $\lambda = 3$ in Equation (2), it gives,

$$\begin{bmatrix} -9-3 & 4 & 4 \\ -8 & 3-3 & 4 \\ -10 & 8 & 7-7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -12 & 4 & 4 \\ -8 & 0 & 4 \\ -16 & 8 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

➔ Using matrix multiplication

$$\begin{bmatrix} -12x + 4y + 4z \\ -8x + 0y + 4z \\ -16x + 4y + 4z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

By Equating corresponding elements of both side matrices, we get,

$$\left.\begin{aligned} -12x + 4y + 4z &= 0; \\ -8x + 0y + 4z &= 0; \\ -16x + 8y + 4z &= 0 \end{aligned}\right\} \text{Consider these}$$

➔ Using Cramer's Rule

$$\frac{x}{\begin{vmatrix} 0 & 4 \\ 8 & 4 \end{vmatrix}} = \frac{-y}{\begin{vmatrix} -8 & 4 \\ -16 & 4 \end{vmatrix}} = \frac{z}{\begin{vmatrix} -8 & 0 \\ -16 & 8 \end{vmatrix}} = k$$

$$\frac{x}{0-32} = \frac{-y}{-32-(-64)} = \frac{z}{-64-0} = k$$

$$\frac{x}{-32} = \frac{-y}{32} = \frac{z}{-64} = k$$

$\therefore \quad x = -32k, \quad y = -32k, \quad z = -64k$

...(Choose $k = -\frac{1}{32}$ to get smallest integers)

$\therefore$ For Eigen value $\lambda_1 = 3$, Eigen vector is, $X_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$

For $\lambda_2 = \lambda_3 = -1$, Put $\lambda = -1$ in Equation (2), it gives,

$$\begin{bmatrix} -9+1 & 4 & 4 \\ -8 & 3+1 & 4 \\ -16 & 8 & 7+1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -8 & 4 & 4 \\ -8 & 4 & 4 \\ -16 & 8 & 8 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

For repeated Eigen value, find rank of matrix $[A - \lambda I]$ at $\lambda_2 = \lambda_3 = -1$

$$[A - \lambda I] = \begin{bmatrix} -8 & 4 & 4 \\ -8 & 4 & 4 \\ -16 & 8 & 8 \end{bmatrix}$$

operate $R_2 - R_1$; $R_3 - 2R_1$ $\sim \begin{bmatrix} -8 & 4 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Rank = r = 1 and n = number of unknowns = 3

$\therefore$ n − r = 3 − 1 = 2 i.e. Two linearly independent vectors are possible.

Here matrix A is unsymmetric.

$\therefore \quad -8x + 4y + 4z = 0$ (From matrix)

Put $z = 0; -8x + 4y = 0 \Rightarrow 2x = y \quad \therefore X_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$

Put $y = 0; \quad -8x + 4z = 0 \Rightarrow 2x = z \quad \therefore X_3 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$

Hence Eigen vectors are,

$$X_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}_{\lambda_1 = 3} ; X_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}_{\lambda_2 = -1} ; X_3 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}_{\lambda_3 = -1}$$ ✓...Ans.

2.16 EXAMPLE ON REPEATED EIGEN VALUES FOR SYMMETRIC MATRIX

Ex. 2.16.1

Find Eigen values and Eigen vectors for the following matrix : $A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$

☑ **Soln. :**

Step I : Given matrix is,

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

The characteristic equation of the given matrix A is,

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

Since,

$$[A - \lambda I] = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

➔ **Using Scalar Multiplication**

$$= \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$$

➔ **Using Subtraction of matrices**

$$[A - \lambda I] = \begin{bmatrix} 3-\lambda & 1 & 1 \\ 1 & 3-\lambda & -1 \\ 1 & -1 & 3-\lambda \end{bmatrix}$$

∴ Characteristic equations is,

$$\begin{vmatrix} 3-\lambda & 1 & 1 \\ 1 & 3-\lambda & -1 \\ 1 & -1 & 3-\lambda \end{vmatrix} = 0$$

➔ **Using Standard formula**

$$\lambda^3 - S_1 \lambda^2 + S_2 \lambda - |A| = 0 \qquad ...(1)$$

Where S_1 and S_2 are sum of the minors of order 1 and 2 along principal diagonal respectively.

S_1

$$= \begin{vmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{vmatrix}$$

S_1 = sum of diagonal elements

$S_1 = 3 + 3 + 3 = 9$ (Addition of diagonal Elements)

S_2 = [Minor of 3] + [Minor of 3] + [Minor of 3]

(Sum of Minors of diagonal elements)

$$= \begin{vmatrix} \cancel{3} & \cancel{1} & \cancel{1} \\ \cancel{1} & 3 & -1 \\ \cancel{1} & -1 & 3 \end{vmatrix} + \begin{vmatrix} 3 & \cancel{1} & 1 \\ \cancel{1} & \cancel{3} & \cancel{-1} \\ 1 & \cancel{-1} & 3 \end{vmatrix} + \begin{vmatrix} 3 & 1 & \cancel{1} \\ 1 & 3 & \cancel{-1} \\ \cancel{1} & \cancel{-1} & \cancel{3} \end{vmatrix}$$

$$= \begin{vmatrix} 3 & -1 \\ -1 & 3 \end{vmatrix} + \begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix} + \begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix}$$

$$= [\underbrace{(3)(3)}_{9} - \underbrace{(-1)(-1)}_{1}] + [\underbrace{(3)(3)}_{9} - \underbrace{(1)(1)}_{1}]$$

$$- [\underbrace{(3)(3)}_{9} - \underbrace{(1)(1)}_{1}]$$

$$S_2 = [9 - 1] + [9 - 1] + [9 - 1] = 8 + 8 + 8 = 24$$

$|A|$

$$= \begin{vmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{vmatrix}$$

$= 3 \times$ [Minor of 3] $- 1 \times$ [Minor of 1] $+ 1 \times$ [Minor of 1]

$$= 3 \times \begin{vmatrix} \cancel{3} & \cancel{1} & \cancel{1} \\ \cancel{1} & 3 & -1 \\ \cancel{1} & -1 & 3 \end{vmatrix} - 1 \times \begin{vmatrix} \cancel{3} & \cancel{1} & \cancel{1} \\ 1 & \cancel{3} & -1 \\ 1 & \cancel{-1} & 3 \end{vmatrix} + 1 \times \begin{vmatrix} \cancel{3} & \cancel{1} & \cancel{1} \\ 1 & 3 & \cancel{-1} \\ 1 & -1 & \cancel{3} \end{vmatrix}$$

$$= 3 \times \begin{vmatrix} 3 & -1 \\ -1 & 3 \end{vmatrix} - 1 \times \begin{vmatrix} 1 & -1 \\ 1 & 3 \end{vmatrix} + 1 \times \begin{vmatrix} 1 & 3 \\ 1 & -1 \end{vmatrix}$$

$$= 3 \times [\underbrace{(3)(3)}_{9} - \underbrace{(-1)(-1)}_{1}] - 1 \times [\underbrace{(1)(3)}_{3} - \underbrace{(1)(-1)}_{-1}]$$

$$- 1[\underbrace{(1)(-1)}_{-1} - \underbrace{(1)(3)}_{3}]$$

$$= 3 \times [9 - 1] - 1 \times [3 + 1] + 1 \times [-1 - 3]$$

$$|A| = 3(8) - 1(4) + 1(-4) = 24 - 4 - 4 = 16$$

Substitute values of S_1, S_2 and |A| in Equation (1), it gives

$$\lambda^3 - 9\lambda^2 + 24\lambda - 16 = 0$$

➔ **Using Synthetic division**

$$\begin{array}{l|rrrr} 1 & 1 & -9 & 24 & -16 \\ & - & 1 & -8 & 16 \\ \hline & 1 & -8 & 16 & \boxed{0} \end{array}$$

$\lambda = 1,\ \lambda^2 - 8\lambda + 16 = 0$

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

$\therefore \quad \lambda = 1, 4, 4$

Eigen value are $\lambda_1 = 1, \lambda_2 = 4, \lambda_3 = 4$ ✓...**Ans.**

{**Verification :** $S_1 = \lambda_1 + \lambda_2 + \lambda_3$ and $|A| = \lambda_1 \lambda_2 \lambda_3$ }

Step II : Eigen vectors :

To find Eigen vector $X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ corresponding to

Eigen values, $\lambda_1 = 1,\ \lambda_2 = \lambda_3 = 4$

Consider, $[A - \lambda I]\, X = 0$

$$\begin{bmatrix} 3-\lambda & 1 & 1 \\ 1 & 3-\lambda & -1 \\ 1 & -1 & 3-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \qquad ...(2)$$

$$\because \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

For $\lambda_1 = 1$, Put $\lambda = 1$ in Equation (2), it gives,

$$\begin{bmatrix} 3-1 & 1 & 1 \\ 1 & 3-1 & -1 \\ 1 & -1 & 3-1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0$$

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0$$

By equating corresponding elements of both side matrices, we get,

$$\left.\begin{aligned} 2x + y + z &= 0 \\ x + 2y - z &= 0 \\ x - y + 2z &= 0 \end{aligned}\right\} \text{Consider these}$$

➔ **Using Cramer's Rule**

$$\frac{x}{\begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix}} = \frac{-y}{\begin{vmatrix} 1 & -1 \\ 1 & 2 \end{vmatrix}} = \frac{z}{\begin{vmatrix} 1 & 2 \\ 1 & -1 \end{vmatrix}} = k$$

$$\frac{x}{4-1} = \frac{-y}{2+1} = \frac{z}{-1-2} = k$$

$$\frac{x}{3} = \frac{y}{-3} = \frac{z}{-3} = k$$

$\therefore \quad x = 3k,\ y = -3k,\quad z = -3k$

$\therefore$ For Eigen value $\lambda_1 = 1$, Eigen vector is, $X_1 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$

For $\lambda_2 = \lambda_3 = 4$, Put $\lambda = 4$ in Equation (2), it gives,

$$\begin{bmatrix} 3-4 & 1 & 1 \\ 1 & 3-4 & -1 \\ 1 & -1 & 3-4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0$$

$$\begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0$$

Here Eigen value are repeated. First find rank of $[A - \lambda I]$ at $\lambda_2 = 4$

$$\therefore [A - \lambda I] = \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix}$$

Operate $R_2 + R_1$; $R_3 + R_1 \sim \begin{bmatrix} -1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$\therefore$ Rank = r = 1 and n = number of unknowns = 3

$\therefore$ n − r = 3 − 1 = 2, i.e. Two Eigen vectors are possible.

Here, matrix A is symmetric, therefore Eigen vectors are orthogonal.

By above matrix, $-x + y + z = 0$

Put $z = 0$; $x = y$;

$\therefore \quad X_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and consider $X_3 = \begin{bmatrix} l \\ m \\ n \end{bmatrix}$

$\therefore \quad X_1 X_3^T = 0$ and $X_2 X_3^T = 0$

Hence Eigen vectors are,

$$\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} [l\ m\ n] = 0 \quad \text{and} \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} [l\ m\ n] = 0$$

$-l + m + n = 0$ and $l + m = 0 \Rightarrow l = -m$, Put $m = k$

$\therefore \quad l = -k$ and $n = l - m = -k - k = -2k$

$\therefore \ X_3 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$

Hence Eigen vectors are,

$$X_1 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}_{\lambda_1 = 1};$$

$$X_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}_{\lambda_2 = 4};$$

$$X_3 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}_{\lambda_3 = 4}$$

✓ ...**Ans.**

2.17 ORTHOGONAL PROJECTIONS

Notation

- Let W be a subspace of R^n and let **X** be a vector in $\mathbf{R}^n$. We denote the closest vector to X on W by X_W.
- Note that, X_W is the closest vector to X on W implies that difference $X - X_W$ is **Orthogonal** to the vectors in W.

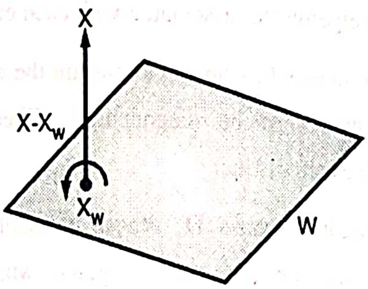


Fig. 2.17.1

It $X_W^{\perp} = X - X_W$, then we have

$$X = X_W + X_W^{\perp},$$

where X_W is in W and $X_W^{\perp}$ is in $W^{\perp}$.

2.17.1 Orthogonal Decomposition

- Let W be a subspace of R^n and let X be a vector in R^n. The expression

$$X = X_W + X_W^{\perp},$$

- For X_W in W and $X_W^{\perp}$ in $W^{\perp}$ is called orthogonal decomposition of X with respect to W, and the closet vector X_W is the **orthogonal projection** of X on W.
- Since X_W is the **closest** vector on W to X ; the distance from X to the subspace W is the length of the vector X_W to X, i.e., the length of $X_W^{\perp}$.

2.17.2 Closest Vector and Distance

We restate :

Let W be a subspace of R^n and let X be a vector in R^n.

(i) The orthogonal projection X_W is the closest vector to X in W,

(ii) The distance from X to W is $\| X_W^{\perp} \|$.

2.17.3 Example (Orthogonal Decomposition with respect to the XY - plane)

- Let W be XY – plane, in R^3 so WT is Z – axis. We compute the orthogonal decomposition of a vector with respect to W :

$$X = \begin{pmatrix}1\\2\\3\end{pmatrix} \Rightarrow X_W = \begin{pmatrix}1\\2\\0\end{pmatrix}, X_W^{\perp} = \begin{pmatrix}0\\0\\3\end{pmatrix}$$

$$X = \begin{pmatrix}a\\b\\c\end{pmatrix} \Rightarrow X_W = \begin{pmatrix}a\\b\\0\end{pmatrix}, X_W^{\perp} = \begin{pmatrix}0\\0\\c\end{pmatrix}$$

2.17.4 Orthogonal Projection onto a Line

Let L be a line, then

$$X_L = \frac{u \cdot x}{u \cdot u} u ;$$

$$X_{LT} = X - X_L, \text{ for any vector X}$$

Ex. 2.17.1 : **Compute the orthogonal projection of $X = \begin{pmatrix}-6\\4\end{pmatrix}$ onto the line L spanned by $u = \begin{pmatrix}3\\2\end{pmatrix}$, and find the distance from X to L.**

Soln. :

We have,

$$X_L = \frac{x \cdot u}{u \cdot u} u ;$$

$$= \frac{\begin{pmatrix}-6\\4\end{pmatrix} \cdot \begin{pmatrix}3\\2\end{pmatrix}}{\begin{pmatrix}3\\2\end{pmatrix} \cdot \begin{pmatrix}3\\2\end{pmatrix}} \begin{pmatrix}3\\2\end{pmatrix}$$

$$= \frac{-18+8}{9+4} \begin{pmatrix}3\\2\end{pmatrix}$$

$$X_L = \frac{-10}{13} \begin{pmatrix}3\\2\end{pmatrix} \quad \text{...(i)}$$

$$X_{LT} = X - X_L = \begin{pmatrix}-6\\4\end{pmatrix} + \frac{10}{13}\begin{pmatrix}3\\2\end{pmatrix}$$

$$= \frac{1}{13}\begin{pmatrix}-48\\32\end{pmatrix}$$

2.18 DIAGONALISATION

- A diagonalisation matrix is a square matrix that can be transformed into a diagonal matrix, that is a matrix with non-diagonal elements as zeros.

- The mathematical relation between a matrix and its diagonalised matrix is :

$$A = PDP^{-1}$$

or equivalently,

$$D = P^{-1}AP,$$

- Where A is the matrix to be diagonalised P is a matrix whose columns are the eigenvectors of A, P^{-1} is its inverse matrix, and D is the diagonal matrix composed by the eigenvalues of A.
- Matrix A and matrix D are similar matrices. And B is an invertible matrix.

2.18.1 When a Matrix can be Diagonalised ?

All matrices are not diagnolisable. There are three methods to know whether a matrix is diagnolisable or not.

(1) A square matrix of order n is diagnolisable, if it has n **linearly independent eigenvectors.** In other words, if these vectors form a basis.

To check whether the eigenvectors meet this condition, it is enough to show that the determinant of the matrix P is non-zero.

$\therefore$ If det (P) $\neq 0$ then matrix is diagnolisable.

(2) Eigenvectors of different eigenvalues are always linearly independent. Therefore, **if all eigenvalues of the matrix are unique,** then the matrix is diagnolisable.

(3) Another method is by using algebraic and geometric multiplicities.

The algebraic multiplicity is the number of times an eigen value is repeated, and the geometric multiplicity is the dimension of the null space of matrix $(A - \lambda I)$.

If the algebraic multiplicity is equal to the geometric multiplicity for each eigen value, the matrix is diagnolisable.

α_λ = algebraic multiplicity

= multiplicity of eigen value λ

m_λ = geometric multiplicity

$$\alpha_\lambda \geq m_\lambda \geq 1$$

If $\alpha_\lambda = m_\lambda,\ \forall_\lambda,$ then the matrix is diagnolisable.

2.18.2 Method of Diagonalising a Matrix

- The process of diagonalising a matrix is based on finding the eigenvalues and eigenvecors of a matrix. The steps to diagonalise a matrix are :

1. Find the eigenvalues of the matrix.
2. Find the eigenvector associated with each eigonvalue.
3. Form the matrix P, whole columns are the eigenvectors of the matrix to be diagonalised (Check whether eigenvalues are distinct.)
4. Form diagonal matrix D, Whose elements are all 0 except those on the main diagonal which are the eigenvalues found in step 1.

Note : The eigenvectors of matrix P can be placed in any order, but the eigen values of diagonal matrix D must be placed in that same order.

- For example, the first eigenvalue of diagonal matrix D must correspond to the eigenvector of the first column of matrix P.

2.18.3 Solved Examples

Ex. 2.18.1 : Diagonalise the matrix

$$A = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix}$$

Soln. :

Step 1 : To find eigenvalue of A

Characteristic equation of A in λ is

$$|A - \lambda I| = 0;$$

$$\text{i.e. } |A - \lambda I| = \begin{bmatrix} 2-\lambda & 2 \\ 1 & 3-\lambda \end{bmatrix} = \lambda^2 - 5\lambda + 4 = 0$$

$$\therefore (\lambda - 1)(\lambda - 4) = 0$$

$$\therefore \lambda_1 = 1,\ \lambda_2 = 4$$

Step 2 :

(i) To find eigenvectors for $\lambda_1 = 1$;

Matrix equation is $(A - \lambda I) X = 0$

$$\therefore \begin{bmatrix} 2-1 & 2 \\ 1 & 3-1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

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

$\therefore x + 2y = 0 \rightarrow$ by R_1

Let $x = -2, y = 1$

$$\therefore X_1 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

(ii) For $\lambda_2 = 4$, matrix equation is

$$\begin{bmatrix} 2-4 & 2 \\ 1 & 3-4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

$$\therefore \begin{bmatrix} -2 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\therefore x - y = 0 \rightarrow$ by R_2

$\therefore x = y$

Let $x = 1, \therefore y = 1$

$$\therefore X_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\therefore P = \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix}$$

▶ **Step 3 :**

Since both eigenvalues are distinct,

$\therefore$ Matrix A is diagonalisable.

$$\therefore \mathbf{D} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{4} \end{bmatrix}$$

Note : The matrix P is called the change of basis matrix

$\therefore$ The change of basis matrix and the diagonalisable matrix are.

$$P = \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix}$$

and $D = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$

Ex. 2.18.2 : Diagonalise the following matrix

$$A = \begin{bmatrix} 3 & 4 \\ -1 & -2 \end{bmatrix}$$

☑ **Soln. :**

▶ **Step I :**

To find eigenvalue of A :

Characteristic equation of A in λ is $|A - \lambda I| = 0$;

i.e. $\begin{vmatrix} 3-\lambda & 4 \\ -1 & -2-\lambda \end{vmatrix} = 0$

i.e. $\lambda^2 - \lambda - 2 = 0$

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

$\therefore \lambda_1 = -1, \lambda_2 = 2$

▶ **Step (2) : To find eigen vectors :**

Matrix equation of A in λ is $(A - \lambda I) X = 0$

i.e. $\begin{bmatrix} 3-\lambda & 4 \\ -1 & -2-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$

(i) For $\lambda_1 = -1$

$$\therefore \begin{bmatrix} 4 & 4 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

by R_2, $x + y = 0$

Let $x = -1, y = 1$

$$\therefore X_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

(ii) For $\lambda_2 = 2$; Matrix equation is

$$\begin{bmatrix} 1 & 4 \\ -1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

by R_1, $x + 4y = 0$

Let $x = -4, y = 1$

$$\therefore X_2 = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$$

Now $P = [X_1 \, X_2] = \begin{bmatrix} -1 & -4 \\ 1 & 1 \end{bmatrix}$

▶ **Step 3 :**

Since eiqen values are distinct, $\therefore$ A is diagonalisable.

$$\therefore D = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix},$$

$$P = \begin{bmatrix} -1 & -4 \\ 1 & 1 \end{bmatrix}$$

Ex. 2.18.3 : Diagonalise the following matrix

$$A = \begin{bmatrix} 2 & 0 & 2 \\ -1 & 2 & 1 \\ 0 & 1 & 4 \end{bmatrix}$$

☑ **Soln. :**

▶ **Step (1) : To find eigenvalues of A.**

Characteristic equation of A in λ is $|A - \lambda I| = 0$;

i.e. $$\begin{bmatrix} 2-\lambda & 0 & 2 \\ -1 & 2-\lambda & 1 \\ 0 & 1 & 4-\lambda \end{bmatrix} = 0$$

$\therefore -\lambda^3 + 8\lambda^2 - 19\lambda + 12 = 0$

$\therefore \lambda_1 = 1, \lambda_2 = 3, \lambda_3 = 4$

▶ **Step (2) : To find eigenvector :**

Matrix equation of A in λ is $(A - \lambda I) X = 0$

i.e. $$\begin{bmatrix} 2-\lambda & 0 & 2 \\ -1 & 2-\lambda & 1 \\ 0 & 1 & 4-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

(i) For $\lambda_1 = -1$

$$\begin{bmatrix} 1 & 0 & 2 \\ -1 & 1 & 1 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$\therefore$ by R_1, $x + 2z = 0$

by R_2, $-x + y + z = 0$

by R_3, $y + 3z = 0$,

$\therefore x = -2z, \quad y = -3z$

Let $z = 1$, $\therefore x = -2, y = -3$

$$\therefore X_1 = \begin{bmatrix} -2 \\ -3 \\ 1 \end{bmatrix}$$

(ii) For $\lambda_2 = 3$; Matrix equation is

$$\begin{bmatrix} -1 & 0 & 2 \\ -1 & -1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Expanding,

$-x + 2z = 0$

$-x - y + z = 0$

$y + z = 0$

$\therefore x = 2z$

$y = -z$

Let $z = 1$, $\therefore x = 2, y = -1$

$$\therefore X_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

(ii) For $\lambda_3 = 4$; Matrix equation is

$$\begin{bmatrix} -2 & 0 & 2 \\ -1 & -2 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Expanding,

$-2x + 2z = 0$

$-x - 2y + z = 0$

$y = 0$

$\therefore x = z, \quad y = 0$

Let $z = 1, \quad \therefore x = 1$

$$\therefore X_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

▶ **Step (III)**

We construct matrix P, formed by the eigenvecors of the matrix :

$$P = \begin{bmatrix} -2 & 2 & 1 \\ -3 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Since all eigenvalues are distinct, hence the matrix A is diagonalisable.

And diagonal matrix is

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

$\therefore$ Matrices P and D are :

$$P = \begin{bmatrix} -2 & 2 & 1 \\ -3 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$\text{And } D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Ex. 2.18.4 : Diagonalise if possible, the following square matrix.

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

☑ **Soln. :**

▶ **Step (I) : To find eigenvalues of A.**

Characteristic equation of A in λ is $|A - \lambda I| = 0$;

$$\text{i.e.} \begin{bmatrix} -1-\lambda & 3 & 1 \\ 0 & 2-\lambda & 0 \\ 3 & -1 & 1-\lambda \end{bmatrix} = 0$$

$$\text{i.e.} -\lambda^3 + 2\lambda^2 + 4\lambda - 8 = 0$$

$$\therefore \text{Roots are } \lambda_1 = 2, \lambda_2 = 2, \lambda_3 = -2$$

The eigenvalue -2 has simple algebraic multiplicity on the other hand, the eigenvalue 2 has double multiplicity.

▶ **Step (II)**

We calculate eigenvector associated with each eigenvalue.

(i) For $\lambda_3 = -2$; matrix equation is

$$\begin{bmatrix} 1 & 3 & 1 \\ 0 & 4 & 0 \\ 3 & -1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Expanding ,

$$x + 3y + z = 0 \qquad \therefore y = 0 \text{ and}$$

$$4y = 0 \qquad x = -z$$

$$3x - y + 3z = z,$$

$$\text{Let } x = 1, \quad z = -1$$

$$\therefore X_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

(ii) For $\lambda_1 = 2$; matrix equation is $(A - \lambda I) X = 0$;

$$\begin{bmatrix} -3 & 3 & 1 \\ 0 & 0 & 0 \\ 3 & -1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Expanding ,

$$-3x + 3y + z = 0$$

$$3x - y - z = 0$$

$$\text{Let } y = 0, \qquad z = 3x$$

$$\text{Let } x = 1, \qquad z = 3,$$

$$\therefore X_2 = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$$

(iii) For $\lambda_2 = 2$;

Since the eigenvalve 2 is repeated, we have to calculate another eigenvector,

$$\text{Let } X_3 = \begin{bmatrix} -1 \\ 0 \\ -3 \end{bmatrix}$$

$$\therefore P = [X_1 + X_2 + X_3]$$

$$= \begin{bmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ -1 & 3 & -3 \end{bmatrix}$$

▶ **Step (III)**

We observe that the three vectors are not linearly independent, since the two eigenvectors of the eigenvalue 2 are a linear combination of each other.

Also, note that $|P| = 0$

$$\text{i.e.} \begin{vmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ -1 & 3 & -3 \end{vmatrix} = 0$$

$\therefore$ Matrix A is not diagonalisable.

Ex. 2.18.5 : Diagonalise if possible, the following 3×3 matrix.

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

☑ **Soln. :**

Step (I) : To find eigenvalues of A.

Characteristic equation of A in λ is $|A - \lambda I| = 0$;

$$\therefore \begin{vmatrix} 3-\lambda & 0 & 0 \\ 0 & 2-\lambda & 1 \\ 0 & 1 & 2-\lambda \end{vmatrix} = 0$$

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

$\therefore$ Roots are, $\lambda_1 = 1, \lambda_2 = 3, \lambda_3 = 3$

Step (II) : To find eigenvector :

(i) For $\lambda_1 = 1$, the matrix equation is $(A - \lambda I) X = 6$

$$\text{i.e.} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Expanding ,

$$2x = 0, \qquad \therefore x = 0$$
$$y + z = 0, \qquad y = -z$$
$$\text{Let } z = 1, \qquad \therefore y = -1,$$

$$\therefore X_1 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

(iii) For $\lambda_2 = 3$; Matrix equation is

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Expanding ,

$$\text{i.e. } 0 \cdot x = 0,$$
$$-y + z = 0, \qquad \therefore y = z$$
$$\text{Let } x = 0, \quad y = 1, \quad z = 1,$$

$$\therefore X_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

(iii) For $\lambda_3 = 3$;

Since the eigenvalue 3 is repeated twice we calculate another eigenvector,

$$\text{Let } x = 1, \quad y = 0, \quad z = 0,$$

$$\therefore X_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Step (III) : Now, the matrix P of eigenvectors is

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

We observe that, in this case there are 3 linearly independent vectors, even though 3 has double algebraic multiplicity.

Also note that

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

$\therefore$ A is diagonalisable and

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix};$$

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

Are the two matrices.

Ex. 2.18.6 : Diagonalise if possible, the matrix :

$$A = \begin{bmatrix} 2 & 1 & 2 & 0 \\ 1 & -3 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

☑ **Soln. :**

Step (I) : We find eigenvalues of A.

Characteristic equation of A in λ is $|A - \lambda I| = 0$;

$$\text{i.e.} \begin{vmatrix} 2-\lambda & 1 & 2 & 0 \\ 1 & -3-\lambda & 1 & 0 \\ 0 & -1 & -\lambda & 0 \\ 0 & 0 & 0 & 5-\lambda \end{vmatrix} = 0$$

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

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

$\therefore$ Roots are :

$$\lambda_1 = 0, \qquad \lambda_2 = -3,$$

$$\lambda_3 = 2 \quad \text{and} \quad \lambda_4 = 5$$

Step (2) : To find eigenvectors :

Matrix equation of A in λ is $(A - \lambda I)X = 0$

(i) For $\lambda_1 = 0$; matrix equation is

$$\begin{bmatrix} 2 & 1 & 2 & 0 \\ 1 & -3 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Solving :

$$2w + x + 2y = 0$$
$$w - 3x + y = 0$$
$$-x = 0$$
$$5z = 0$$
$$\therefore x = 0, \qquad z = 0,$$
$$w = -y$$
$$\text{Let } y = 1, \qquad \therefore w = -1$$

$$\therefore X_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

(ii) For $\lambda_2 = -3$; matrix equation is

$$\begin{bmatrix} 5 & 1 & 2 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Expanding ,

$$5w + x + 2y = 0$$
$$w + y = 0$$
$$-x + 3y = 0$$
$$8z = 0$$
$$\therefore w = -y, \qquad x = 3y,$$
$$z = 0$$
$$\text{Let } y = 1, \qquad \therefore w = -1$$
$$x = 3, \qquad z = 0$$

$$\therefore X_2 = \begin{bmatrix} -1 \\ 3 \\ 1 \\ 0 \end{bmatrix}$$

(iii) For $\lambda_3 = 2$; the matrix equation is

$$\begin{bmatrix} 0 & 1 & 2 & 0 \\ 1 & -5 & 1 & 0 \\ 0 & -1 & -2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Expanding ,

$$x + 2y = 0, \qquad w - 5x + y = 0$$
$$-x - 2y = 0, \qquad 3z = 0$$
$$\therefore x = -2y, \qquad w = -11y,$$
$$z = 0$$
$$\text{Let } y = 1, \qquad \therefore x = -2$$
$$y = -11, \qquad z = 0$$

$$\therefore X_3 = \begin{bmatrix} -11 \\ -2 \\ 1 \\ 0 \end{bmatrix}$$

(iv) For $\lambda_4 = 5$; the matrix equation becomes

$$\begin{bmatrix} -3 & 1 & 2 & 0 \\ 1 & -8 & 1 & 0 \\ 0 & -1 & -5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Expanding ,

$$-3w + x + 2y = 0, \qquad w - 8x + y = 0$$
$$-x - 5y = 0, \qquad \therefore w = x = y$$
$$0z = 0,$$
$$\text{Let } z = 1, \qquad w = x = y = 0$$

$$\therefore X_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$\therefore$ P = matrix of eigenvectors

$$P = \begin{bmatrix} -1 & -1 & -11 & 0 \\ 0 & 3 & -2 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

And diagonal matrix

$$D = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

Since all the eigen values are distinct A is diagonalisable.

2.19 SVD AND ITS APPLICATIONS

- The singular value decomposition (SVD) of a matrix is a factorisation of that matrix into three matrices.
- It has interesting algebraic properties and conveys important geometrical and theoretical insights about linear transformations.
- It has important applications in Data Science.

2.19.1 Mathematics behind SVD

- Let A be $m \times n$ matrix, then SVD of A is given by

$$A = uWV^T$$

- Where :

  u : $m \times n$ matrix of the orthonormal eigenvectors of AA^T

  V^T : transpose of a $n \times n$ matrix containing Orthonormal eigenvectors of A^TA

  W : a diagonal matrix of the singular values which are the square roots of the eigenvalues of A.

2.19.2 Solved Example

Ex. 2.19.1 : Find SVD for the matrix.

$$A = \begin{bmatrix} 3 & 3 & 2 \\ 2 & 3 & -2 \end{bmatrix}$$

Soln. :

Step (I) : First we find AA^T and its eigenvalues

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

$$A \cdot A^T = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$$

The characteristic equation of AA^T is given by

$$|A \cdot A^T - \lambda I| = 0;$$

$$\text{i.e.} \begin{bmatrix} 17-\lambda & -8 \\ 8 & 17-\lambda \end{bmatrix} = 0$$

$$\therefore (17-\lambda)^2 - 64 = 0$$

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

$$\therefore (\lambda - 25)(\lambda - 9) = 0$$

$$\therefore \lambda_1 = 25$$

$$\lambda_2 = 9$$

Step (II) : Singular values are given by

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

$$\sigma_2 = \sqrt{\lambda_2} = \sqrt{9} = 3 \quad \text{...(i}$$

Step (III) : To find orthonormal set of eigenvecto of $A \cdot A^T$.

Matrix equation of $A \cdot A^T$ is given by

$$|A \cdot A^T - \lambda I| X = 0$$

$$\text{i.e.} \begin{bmatrix} 17-\lambda & 8 \\ 8 & 17-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

(i) For $\lambda_1 = 25$

$$\begin{bmatrix} -8 & 8 \\ 8 & -8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\therefore \text{ by } R_1, -8x_1 + 8x_2 = 0$$

$$\therefore x_1 = x_2$$

$\therefore$ Normalised eigenvector is $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

(ii) For $\lambda_2 = 9$;

In similar fashion, normalised eigenvector is

$$\therefore u = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \quad \text{...(ii)}$$

Step (IV) : We find eigenvectors of A^TA.

The eigenvalues of A^TA are 25, 9, 0 and since A^TA is symmetric, eigenvectors are orthogonal.

Matrix equation of A^TA is

$(A^TA - \lambda I) X = 0$

(i) For $\lambda_1 = 25$; matrix equation is

i.e. $$\begin{bmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Expanding ,

$$\text{by } R_1, -12x + 12y + 2z = 0$$
$$\text{by } R_3, \; 2x - 2y - 17z = 0,$$
$$\text{Let } z = 0,$$
$$\text{Then } x = y$$

Let $x = y = 1$

$\therefore$ Normalised eigenvector is

$$X_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$$

(ii) For $\lambda_2 = 9$;

In the similar fashion, normalised eigenvector is

$$X_2 = \begin{bmatrix} \frac{1}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \\ \frac{4}{\sqrt{18}} \end{bmatrix}$$

(iii) For $\lambda_3 = 0$;

For the third eigenvector, we use the property that eigenvectors are orthogonal ;

i.e. $X_1 \perp X_3$ and $X_2 \perp X_3$

Let $$X_3 = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

$\because X_1^T X_3 = 0$

$$\therefore \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0$$

$$\therefore \frac{a}{\sqrt{2}} + \frac{b}{\sqrt{2}} = 0 \qquad \therefore a + b = 0 \qquad ...(iii)$$

and $X_2^T X_3 = 0$

$$\therefore \begin{bmatrix} \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0$$

$$\therefore a - b + 4c = 0 \qquad ...(iv)$$

From (iii) $b = -a$,

From (iv) $c = -\frac{a}{2}$

$$\therefore X_3 = \begin{bmatrix} a \\ -a \\ -\frac{a}{2} \end{bmatrix}$$

$$\therefore \text{Normalised eigenvector} = X_3 = \begin{bmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \end{bmatrix}$$

Step (V) :

From (1), a diagonal matrix

$$W = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}$$

and V = orthonormal eigenvector of A^TA

$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} & -\frac{2}{3} \\ 0 & \frac{4}{\sqrt{18}} & \frac{1}{3} \end{bmatrix}$$

From (i)

$$u = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$\therefore$ Final SVD equation is

$$A = uWV^T$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

2.19.3 Applications of SVD

(1) Calculation of pseudo inverse

- Moore - Penrose inverse is the generalisation of the matrix inverse that may not be invertible (such as low - rank matrices).
- If the matrix is invertible, then its inverse will be equal to pseudo – inverse but pseudo - inverse exists for the matrix that is not invertible. It is denoted by A^+.
- Suppose, we calculate the pseudo - **inverse of a matrix M :**

(i) SVD of M is

$$M = uWV^T$$

Multiplying both sides by M^{-1}

$$\therefore M^{-1}M = M^{-1}uWV^T$$

$$\therefore I = M^{-1}uWV^T$$

Multiplying both sides by V

$$V = M^{-1}uWV^TV$$

$$\therefore V = M^{-1}uW$$

Since W is singular matrix, multiplying by W^{-1},

$$\therefore VW^{-1} = M^{-1}uWW^{-1}$$

$$\therefore VW^{-1} = M^{-1}u$$

Multiplying by u^T

$$\therefore VW^{-1}u^T = M^{-1}uu^T$$

$$\therefore VW^{-1}u^T = M^{-1}$$

$$\text{i.e. } M^+ = VW^{-1}u^T$$

This equation gives pseudo – inverse.

(2) Solving a set of Homogeneous Linear Equation

The equation is $MX = b$

If $MX = b$

Multiply by M^{-1}

$$M^{-1}MX = M^{-1}b$$

$$\therefore X = M^{-1}b$$

From the pseudo – inverse,

$$M^{-1} = VW^{-1}u^T$$

$$\therefore X = VW^{-1}u^Tb$$

(3) Rank, Range and Null Space

(i) The rank of matrix M can be calculated from SVD by the number of non - zero singular values.

(ii) The range of matrix M is the **left** singular vectors of u corresponding to the non - zero singular values.

(iii) The null space of matrix M is the Right singular vectors of V corresponding to the zeroed singular values.

$$M = uWV^T$$

(4) Curve - Fitting Problem

(i) Singular value decomposition can be used to minimise the least square error. It uses the Pseudo inverse to approximate it.

(ii) Besides this applications, singular value decomposition and Pseudo - inverse can also be used in Digital signal processing and image processing.

Chapter Ends...

□□□