1. What Is Machine Learning? What is the need for machine learning? Explain the learning system for machine learning.

Machine Learning (ML) is a promising and flourishing field. It enables top management of an organization to extract knowledge from data stored in various archives. This knowledge facilitates decision-making, which helps in development of organization.

Another pioneer of AI, Tom Mitchell's definition of machine learning states that, "*A computer program is said to learn from experience E, with respect to task T and some performance measure P, if its performance on T measured by P improves with experience E.*"

Need for machine learning:

Business organizations use huge amount of data for their daily activities. Earlier, the full potential of this data was not utilized due to two reasons. One reason was not being able to integrate these sources fully. Second is lack of awareness on existence of such software tools.

Machine learning has become so popular because of three reasons:

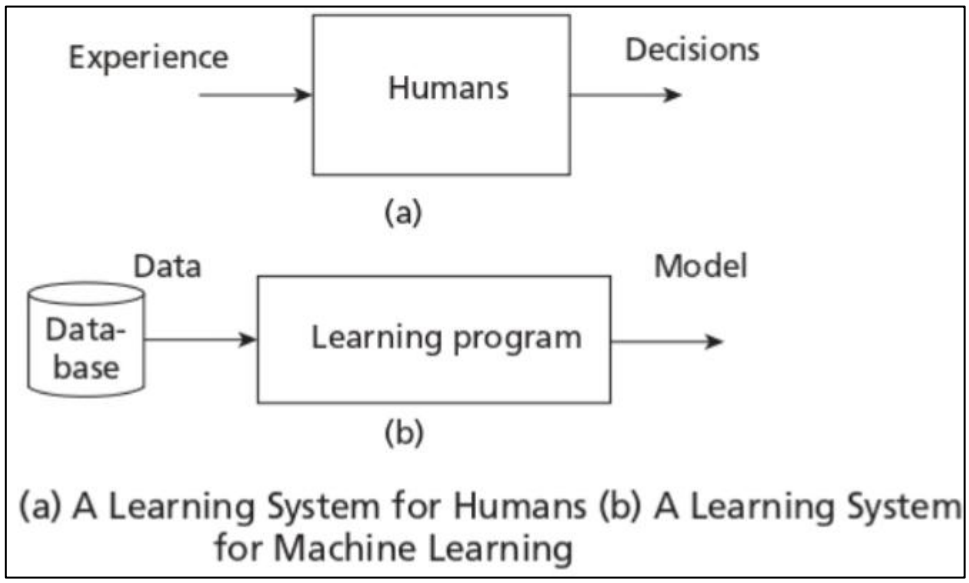
1. **High volume of available data** to manage: Big companies such as Facebook, Twitter, and YouTube generate huge amount of data that grows at a phenomenal rate.

2. Second reason is that the **cost of storage has reduced**. The hardware cost has also dropped. Therefore, it is easier now to capture, process, store, distribute, and transmit the digital information.

3. Third reason is the **availability of complex algorithms** now. Especially with the advent of deep learning, many algorithms are available for machine learning.

Learning system of Machine learning

The aim of machine learning is to learn set of rules from the given dataset automatically so that it can predict the unknown data correctly. As humans take decisions based on an experience, computers make models based on extracted patterns in the input data and then use these data-filled models for prediction. For computers, the learnt model is equivalent to human experience. This is explained in below figure.



(a) A Learning System for Humans (b) A Learning System for Machine Learning

Here the output model can be any or all of the following forms:

1. Mathematical equation

2. Relational diagrams like trees/graphs

3. Logical if/else rules

4. Groupings called clusters

Similar to how human gains experience by seeing, observing, reading books and from teachers, A system gains experience using the following steps:

1. Data Collection – The first step is gathering data, which serves as the foundation for learning.
2. Abstraction & Concept Formation – Data is processed into abstract concepts, similar to how humans recognize objects.
3. Generalization & Heuristics – Machines rank concepts, infer patterns, and develop heuristics (educated guesses) for decision-making.
4. Evaluation & Course Correction – Heuristics may sometimes fail, requiring evaluation and adjustments to improve learning.

2. Explain the knowledge pyramid with neat diagram.

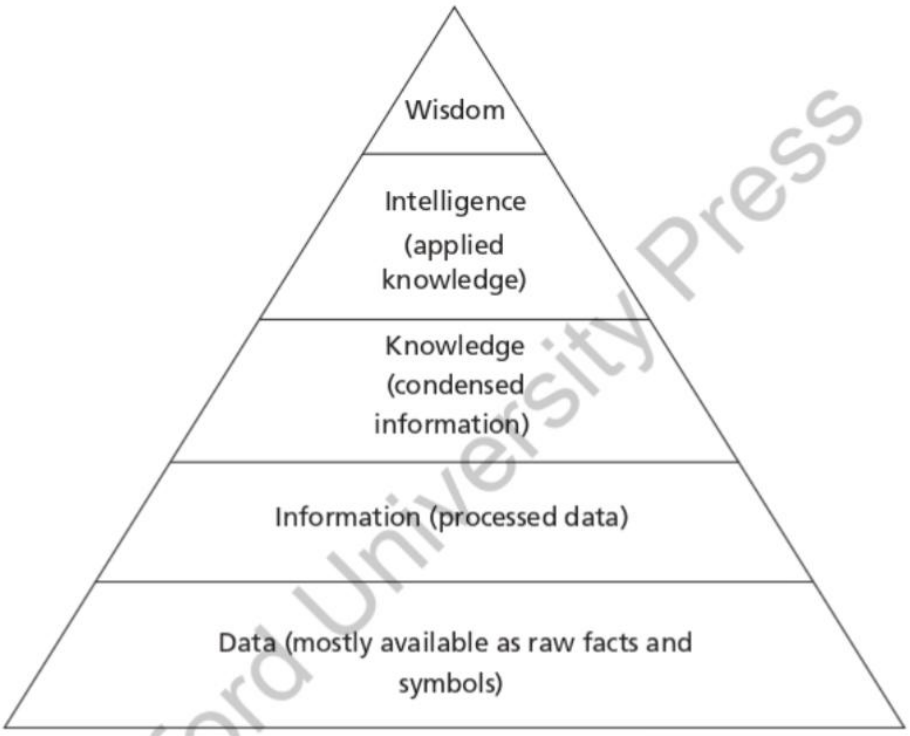


Figure 1.1: The Knowledge Pyramid

Data: All facts are data. Data can be numbers, text, anything that can be processed by a computer. Today, organizations are accumulating vast amount of data.

Processed data: It is also called as information and it includes patterns, associations, relationships among data. Ex: Sales Data can be analysed to show which is fast selling product.

Knowledge: Condensed Information is called knowledge. Ex: Historical data and future trends obtained from sales data can be called knowledge.

Intelligence: Knowledge is useless unless put into action. Intelligence is the applied knowledge for actions. Computers have been successful till this stage.

Wisdom: It represents maturity of mind, this so far is only exhibited by humans.

3. Explain the types of Machine Learning

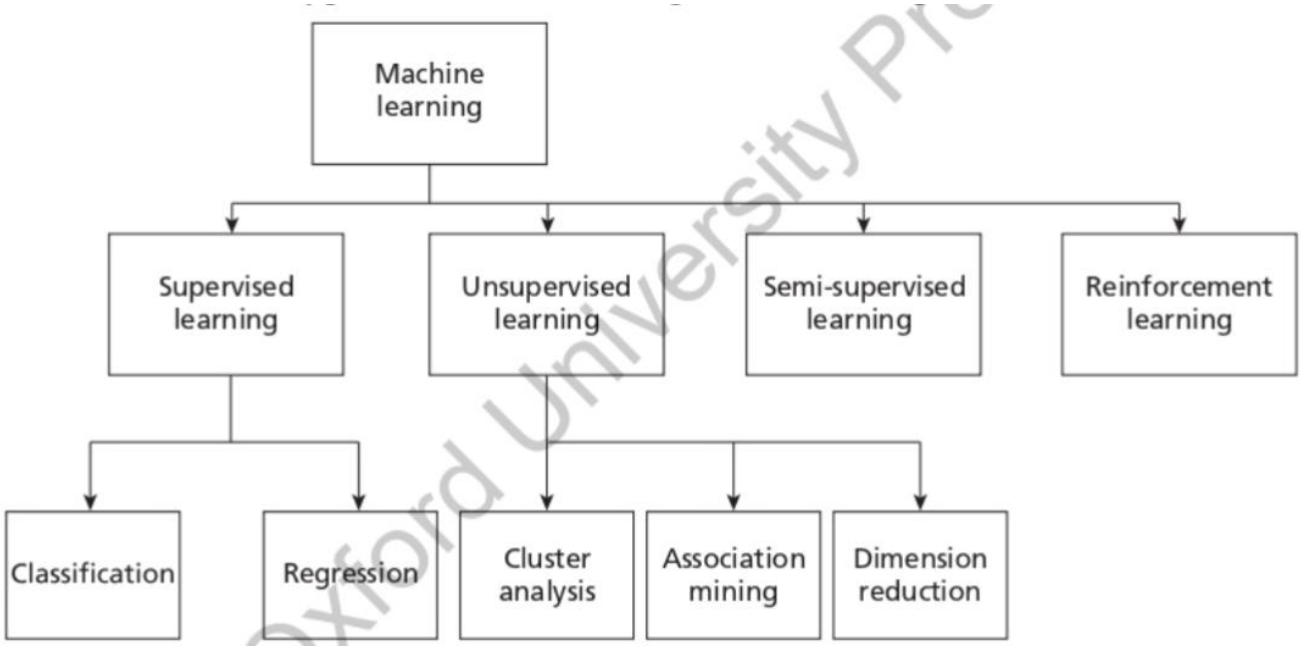


Figure 1.5: Types of Machine Learning

Machine learning is divided into 4 major categories:

1. Supervised Learning:

In this learning, supervised algorithms use labelled dataset. Working of supervised learning is as follows:

Stage 1: A "Teacher" communicates the information to student, student is supposed to master it and understand it. (Training Phase)

Stage 2: "Teacher" asks students a set of questions to find out how much information has been grasped by student. (Testing Phase)

This again has 2 methods:

i. Classification: Deals with discrete labelled data

ii. Regression: Deals with continuous labelled data.

2. Unsupervised learning:

In this method the algorithm is fed data without any labels. It observes the examples and recognize patterns based on the principles of grouping.

Examples: Cluster Analysis, Dimensionality Reduction, Association Mining.

3. Semi-Supervised Learning:

There are circumstances where the dataset has a huge collection of unlabelled data and some labelled data. Labelling is a costly process and difficult to perform by the humans. Semi-supervised algorithms use unlabelled data by assigning a pseudo-label. Then, the labelled and pseudo-labelled dataset can be combined.

4. Reinforcement Learning:

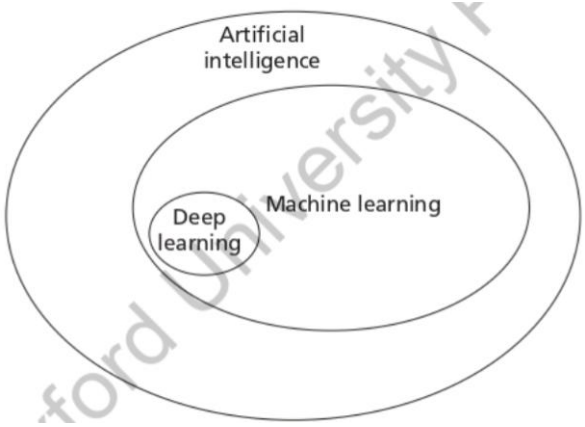
This learning method mimics human behaviour. It allows the agent to interact with environment to get rewards. A positive reward reinforces the behaviour while a negative reward makes the agent avoid the action. Hence the learning becomes possible.

4. Explain the relationship of Machine Learning with other major fields.

1. Machine learning and Artificial intelligence:

Machine learning is a subfield of AI focused on extracting patterns from data for prediction. It emerged from AI's evolution and includes techniques like reinforcement learning to generate results from unknown instances.

Relation between AI and ML is shown in below figure:



2. Machine learning, Data Science, Data Mining, and Data Analytics.

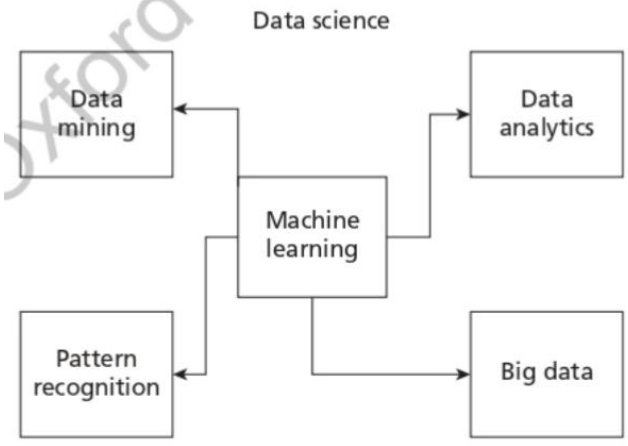
Data Science deals with gathering of data for analysis. It includes:

i. Big Data: It is a field of Data Science that deals with data with High Volume, High variety and high velocity, i.e., speed at which data is generated.

ii. Data Mining: It is another branch of data science which aims to extract hidden patterns that are present in the data, whereas machine learning aims to use it for prediction.

iii. Data Analytics: It is another branch of data science which aims to extract useful knowledge from crude data. This is also closely related to ML

iv. Pattern recognition: It is an engineering field which use ML algorithms to extract patterns. We can view this as one of application of ML.



3. Machine learning and statistics:

Statistics is branch of maths that has strong theoretical foundation. It works by making assumptions and sets a hypothesis and performs experiments to verify and validate the hypothesis, it requires high knowledge of statistical procedures. But ML comparatively has less assumptions and requires less statistical knowledge, it often requires interaction with various tools to automate the process of learning. Many say that ML is just latest version of 'old Statistics' hence there is a good relationship between Statistics and ML

5. Explain Supervised Learning in detail.

Supervised Learning:

As the name suggests, there is a supervisor or teacher component in supervised learning. A supervisor provides labelled data so that the model is constructed and generates test data.

Working of supervised learning is as follows:

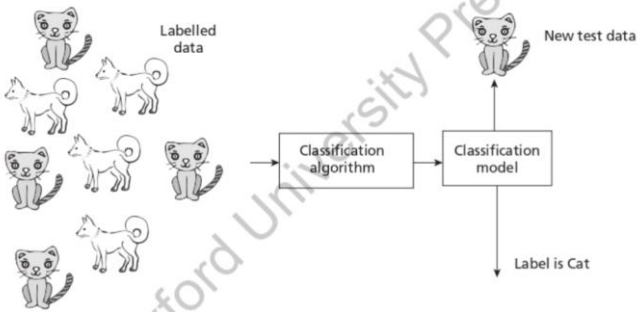
> Stage 1: A "Teacher" communicates the information to student; student is supposed to master it and understand it. (Training Phase)
>
> After First stage a model is generated.
>
> Stage 2: "Teacher" asks students a set of questions to find out how much information has been grasped by student. (Testing Phase)

This kind of learning is typically called supervised learning.

Supervised learning has 2 methods:

i. Classification:

- Classification is a supervised learning method. The input attributes of the classification algorithms are called independent variables. The target attribute is called label or dependent variable.
- The relationship between the input and target variable is represented in the form of a structure which is called a classification **model**.
- The aim of classification is to predict the 'label' of new independent variable.
- An example is shown in Figure below where a classification algorithm takes a set of labelled data images of dogs and cats to construct a model that can later be used to classify an unknown test image data.

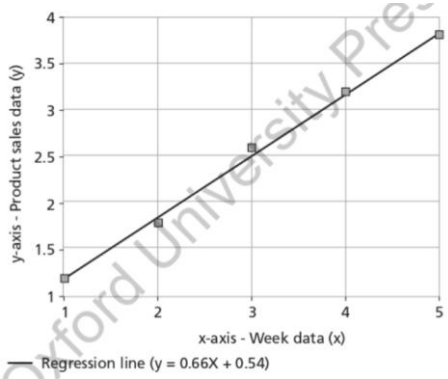


Some of the key algorithms of classification are:

- Decision Tree
- Random Forest
- Support Vector Machines
- Naïve Bayes
- Artificial Neural Network and Deep Learning networks like CNN

ii. Regression:

Regression models, unlike classification algorithms, predict continuous variables like price. In other words, it is a number. A regression model is shown in figure for a dataset that represent weeks input x and product sales y.



The regression model takes input x and generates a model in the form of a fitted line of the form y = f(x). Here, x is the independent variable that may be one or more attributes, and y is the dependent variable.

The advantage of this model is that prediction for product sales (y) can be made for any week data (x). For example, the prediction for eighth week can be made by substituting x as 8 in that regression formula and get y.

Both regression and classification models are supervised algorithms. The main difference is that regression models predict continuous values such as product price, while classification concentrates on discrete values.

6. Explain Unsupervised Learning in detail.

In the absence of a supervisor or teacher, self-instruction is the most common kind of learning process. This process of self-instruction is based on the concept of trial and error. In this method the algorithm is fed data without any labels. It observes the examples and recognize patterns based on the principles of grouping.

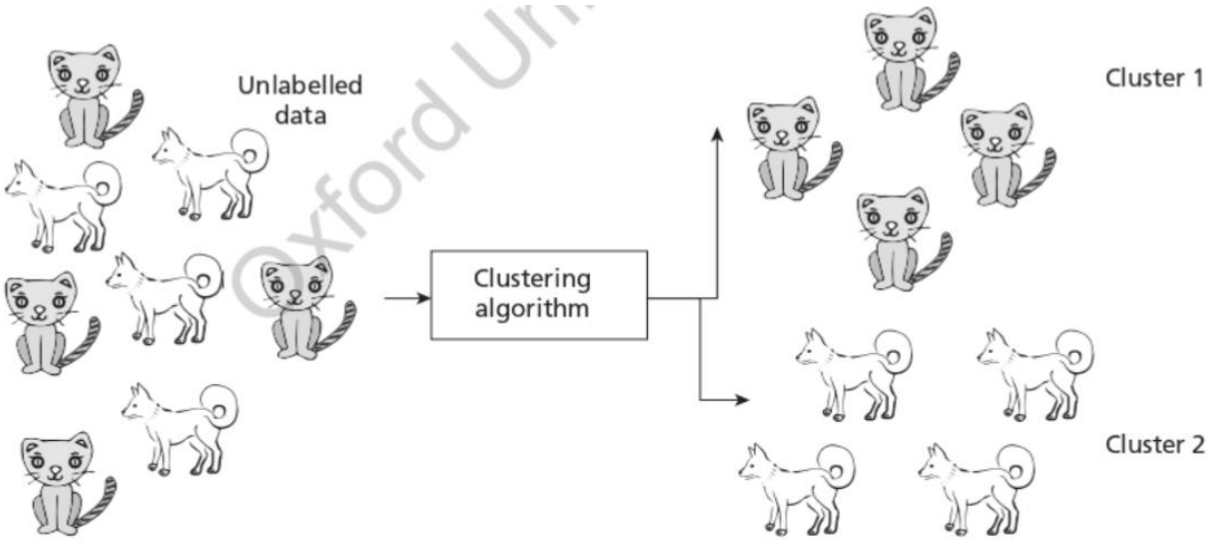
Some examples for unsupervised learning include:

i. Cluster analysis:

Cluster analysis is an example of unsupervised learning. It aims to group objects into disjoint clusters or groups.

Cluster analysis groups objects based on its attributes. All the data objects of the partitions are similar in some aspect and differ from the data objects in the other groups significantly.

An example of clustering scheme is shown in figure below



Some key clustering algorithms are:

k-means algorithm

Hierarchical algorithms

ii. Dimensionality Reduction

Dimensionality reduction algorithms are examples of unsupervised algorithms. It takes a higher dimension data as input and outputs the data in lower dimension by taking advantage of the variance of the data. It aims to retain maximum variance and remove some dimensions.

iii. Association Mining:

This unsupervised algorithm works like a simple if then statement, i.e., If an event 'B' is prone to occur after event 'A', they are both said to be associated. Determining such associations is called association mining.

7. What are the challenges in Machine Learning?

Some challenges of machine learning are

1. **Problems** - Machine learning can deal with the 'well-posed' problems where specifications are complete and available. Computers cannot solve 'ill-posed' problems.

2. **Huge data** - This is a primary requirement of machine learning. Availability of a quality data is a challenge. A quality data means it should be large and should not have data problems such as missing data or incorrect data.

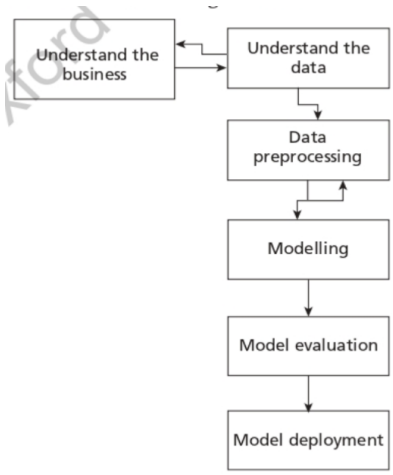
3. **High computation power** - With the availability of Big Data, the computational resource requirement has also increased. Systems with GPU or even TPU are required to execute machine learning algorithms.

4. **Complexity of the algorithms** - Algorithms have become a big topic of discussion and it is a challenge for machine learning professionals to design, select, and evaluate optimal algorithms.

5. **Bias/Variance** - Variance is the error of the model. This leads to a problem called bias/variance tradeoff. A model that fits the training data correctly but fails for test data, in general lacks generalization, is called overfitting. The reverse problem is called underfitting where the model fails for training data but has good generalization. Overfitting and underfitting are great challenges for machine learning algorithms.

8. Explain the Machine Learning process with neat diagram.

A typical ML process is as follows:



A Machine Learning/Data Mining Process

1. **Understanding the business** - This step involves understanding the objectives and requirements of the business organization. Data Scientist should emphasise the working and dependencies of organization before making a solid Data analysis.

2. **Understanding the data** - It involves the steps like data collection, study of the characteristics of the data, formulation of hypothesis, and matching of patterns to the selected hypothesis.

3. **Preparation of data** - This step involves producing the final dataset by cleaning the raw data and preparation of data for the data mining process, suitable strategies should be adopted to handle the missing data else the model generates inaccurate results.

4. **Modelling** – In this step a model is obtained from all the data collected and processed.

5. **Evaluate** – In this step the model is tested with testing datasets, the model is tweaked if the required performance is not met. Choosing a good performance metric is also very important in this step.

6. **Deployment** - This step involves the deployment of model to improve the existing process or for a new situation

9. Explain the applications of Machine Learning in detail.

Machine Learning technologies are used widely now in different domains. Machine learning applications are everywhere! One encounters many machine learning applications in the day-to-day life. Some applications are listed below:

1. **Sentiment analysis** - This is an application of natural language processing (NLP) where the words of documents are converted to sentiments like happy, sad, and angry.

2. **Recommendation systems** - These are systems that make personalized purchases possible. For example, Amazon recommends users to find related books or books bought by people who have the same taste like you, and Netflix suggests shows or related movies of your taste. The recommendation systems are based on machine learning.

3. **Voice assistants** - Products like Amazon Alexa, Microsoft Cortana, Apple Siri, and Google Assistant are all examples of voice assistants. They take speech commands and perform tasks. These chatbots are the result of machine learning technologies.

4. **Technologies like Google Maps** and those used by Uber are all examples of machine learning which offer to locate and navigate shortest paths to reduce time.

The machine learning applications are enormous. Some other applications of Machine Learning are:

S.No.	Problem Domain	Applications
1.	Business	Predicting the bankruptcy of a business firm
2.	Banking	Prediction of bank loan defaulters and detecting credit card frauds
3.	Image Processing	Image search engines, object identification, image classification, and generating synthetic images
4.	Audio/Voice	Chatbots like Alexa, Microsoft Cortana. Developing chatbots for customer support, speech to text, and text to voice
5.	Telecommuni-cation	Trend analysis and identification of bogus calls, fraudulent calls and its callers, churn analysis
6.	Marketing	Retail sales analysis, market basket analysis, product performance analysis, market segmentation analysis, and study of travel patterns of customers for marketing tours
7.	Games	Game programs for Chess, GO, and Atari video games
8.	Natural Language Translation	Google Translate, Text summarization, and sentiment analysis
9.	Web Analysis and Services	Identification of access patterns, detection of e-mail spams, viruses, personalized web services, search engines like Google, detection of promotion of user websites, and finding loyalty of users after web page layout modification
10.	Medicine	Prediction of diseases, given disease symptoms as cancer or diabetes. Prediction of effectiveness of the treatment using patient history and Chatbots to interact with patients like IBM Watson uses machine learning technologies.
11.	Multimedia and Security	Face recognition/identification, biometric projects like identification of a person from a large image or video database, and applications involving multimedia retrieval
12.	Scientific Domain	Discovery of new galaxies, identification of groups of houses based on house type/geographical location, identification of earthquake epicenters, and identification of similar land use

10. What is data? What are the elements of Big Data? Explain

All facts are data. Data is available in different data sources like flat files, databases, or data warehouses. Data by itself is meaningless. It must be processed to generate any information.

Elements of Big data:

Data whose volume is less and can be stored and processed by a small-scale computer is called 'small data'. These data are collected from several sources, and integrated and processed by a small-scale computer. Big data, on the other hand, is a larger data whose volume is much larger than 'small data' and is characterized as follows:

1. Volume: Small traditional data is measured in terms of gigabytes (GB) and terabytes (TB), but Big Data is measured in terms of petabytes (PB) and exabytes (EB). One exabyte is 1 million terabytes.

2. Velocity: The speed at which new data is collected and generated is called velocity, and it helps understand relative growth of big data.

3. Variety: Bigdata can have data of different:

i. Form: text, graphical, audio, video, etc..

ii. Function: It could be conversation, transaction etc...

iii. Source of Data: It could be public data, social media data multimodal data etc...

4. Veracity of data - Veracity of data deals with aspects like conformity to the facts, truthfulness, believability, and confidence in data. There may be many sources of error. So, veracity is one of the most important aspects of data.

5. Validity - Validity is the accuracy of the data for taking decisions or for any other goals that are needed by the given problem.

6. Value - Value is the characteristic of big data that indicates the value of the information that is extracted from the data and its influence on the decisions that are taken

Thus these 6 Vs are helpful to characterize the big data.

11. What are the types of data. What are the types of Analytics?

In Big Data there are primarily 3 kinds of data. Structured, unstructured and semi-structured.

Structured Data Summary

Structured data is data stored in an organized manner, typically in tables or databases, allowing easy retrieval using tools like SQL. It is commonly used in machine learning and categorized into the following types:

1. Record Data – A dataset where each object consists of multiple attributes. It is stored in a matrix format, with rows representing objects and columns representing attributes.
2. Data Matrix – A variation of record data where all attributes are numeric. Standard matrix operations can be applied, and each attribute represents a dimension in a multidimensional space.
3. Graph Data – Data that represents relationships among objects, such as web pages connected by hyperlinks. It is structured as a graph where nodes represent objects, and edges define relationships.
4. Ordered Data – Data with attributes that follow a specific order. It includes:
    - Temporal Data (associated with time, e.g., time series data).
    - Sequence Data (ordered sequences without timestamps, e.g., DNA sequences).
    - Spatial Data (data related to locations, e.g., maps).

Structured data ensures efficient storage, retrieval, and analysis for various applications.

Unstructured Data

Unstructured data includes video, image, and audio. It also includes textual documents, programs, and blog data. It is estimated that 80% of the data are unstructured data.

Semi-Structured Data

Semi-structured data are partially structured and partially unstructured. These include data like XML/JSON data, RSS feeds, and hierarchical data.

Types of analytics

There are primarily 4 types of data analytics, they are:

1. **Descriptive Analytics:**

    It is about describing the main features of the data. After data collection is done, descriptive analytics deals with the collected data and quantifies it. There are two aspects of statistics - Descriptive and Inference. Descriptive analytics only focuses on the description part of the data and not the inference part.

2. **Diagnostic Analytics:**

    It deals with the question - 'Why?". This is also known as causal analysis, as it aims to find out the cause and effect of the events.

3. **Predictive Analytics:**

    It deals with the question - 'What will happen in future given this data?'. This involves applying algorithms to identify the patterns to predict the future.

4. **Prescriptive Analytics:**

    It is about the finding the best course of action for the business organizations. Prescriptive analytics goes beyond prediction and helps in decision making by giving a set of actions.

12. Explain Big data Analysis framework.

Big data framework is a 4 layered architecture which has the following layers:

1. **Data connection layer:**

   It has data ingestion mechanisms and data connectors. Data ingestion means taking raw data and importing it into appropriate data structures. It performs the tasks of ETL process i.e., extract, transform and load operations.

2. **Data management layer**

   It performs preprocessing of data. The purpose of this layer is to allow parallel execution of queries, and do read, write and data management tasks. There may be many schemes that can be implemented by this layer such as data-in-place or constructing data repositories and more.

3. **Data analytics later**

   It has many functionalities such as statistical tests, machine learning algorithms to understand, and construction of machine learning models. This layer implements many model validation mechanisms too. The processing is done by following means:

   i. Cloud computing:

   Cloud computing is an emerging technology which is basically a business service model or simply called as pay-per-usage model. The term 'Cloud' refers to the Internet that provides sharing of resources. It offers different kinds of services such as Infrastructure as a Service, Platform as a Service, and Software as a Service.

   The cloud can be of any of the following types:

   a. Public cloud: Accessible to anyone

   b. Private cloud: Only accessible to a specific organization.

   c. Community Cloud: Accessible to a group of organizations.

   d. Hybrid Cloud: Combination of 2 or more cloud types.

   ii. Grid Computing:

   Grid Computing is a parallel and distributed computing framework consisting of a network of computers offering a super computing service as a single virtual supercomputer.

   iii. H-Computing (High Performance Computing or HPC):

   It enables to perform complex tasks at high speed. It aggregates computing power in such a way that provides much higher performance to solve complex problems in science, engineering, research or business. It leverages parallel processing techniques for solving complex computational problems.

4. **Presentation layer**

   It has mechanisms such as dashboards, and applications that display the results of analytical engines and machine learning algorithms.

13. Explain Data Preprocessing.

In real world, the available data is 'dirty'. By this word 'dirty', it means:

- Incomplete data
- Inaccurate data
- Outlier data
- Data with missing values
- Data with inconsistent values
- Duplicate data

- Data preprocessing improves the quality of the data, The raw data must be pre-processed to give accurate results.
- The process of detection and removal of errors in data is called data cleaning.
- Data wrangling means making the data processable for machine learning algorithms.
- 'Bad data can't be used for Machine Learning. It need to be processed to be usable.

Table 2.1: Illustration of 'Bad' Data

Patient ID	Name	Age	Date of Birth (DoB)	Fever	Salary
1.	John	21		Low	–1500
2.	Andre	36		High	Yes
3.	David	5	10/10/1980	Low	" "
4.	Raju	136		High	Yes

Data preprocessing typically include the following steps:

Handling missing values:

Missing values can be handled by one of the following means:

i. Ignore the tuple: Delete the row with missing values from dataset.

ii. Fill in the values manually: A domain expert can analyse and fill the values manually, but it is time consuming and may not be feasible for large datasets.

iii. A global constant can be used to fill the missing values: Values like null and infinity can be used, but the model generated may not be accurate.

iv. The attribute value may be filled by the attribute value. Say, the average income can replace a missing value.

v. Use the attribute mean for all samples belonging to the same class. Here, the average value replaces the missing values of all tuples that fall in this group.

vi. Use the most probable value to fill in the missing value. The most probable value can be obtained from other methods like classification and decision tree prediction.

Removal of noisy or Outlier Data:

Noise is random error in a value. It can be removed by using **binning**. Binning techniques commonly used are 'smoothing by means' where the mean of the bin replaces the values of the bins, 'smoothing by bin medians' where the bin median replaces the bin values, and 'smoothing by bin boundaries' where the bin value is replaced by the closest bin boundary.

Data Integration and Transformation:

Data integration involves routines that merge data from multiple sources into a single data source. So, this may lead to redundant data. The main goal of data integration is to detect and remove redundancies that arise from integration. Data transformation routines perform operations like normalization to improve the performance of the data mining algorithms. They include algorithms like:

1. Min-Max

2. z-Score

14. Consider the following set: S= {11,13,16,20,23,25,27,30,35}. Apply various binning techniques and show the result.

Let us assume the bin size is 3. The data distributed across bins become:

Bin 1: 11,13,16

Bin 2: 20,23,25

Bin 3: 27,30,35

Using 'smoothing by means' method we get:

Bin 1: 13.3, 13.3, 13.3

Bin 2: 22.6, 22.6, 22.6

Bin 3: 30.6, 30.6, 30.6

Using 'smoothing by bin boundaries' we get:

Bin 1: 11,11,16

Bin 2: 20,25,25

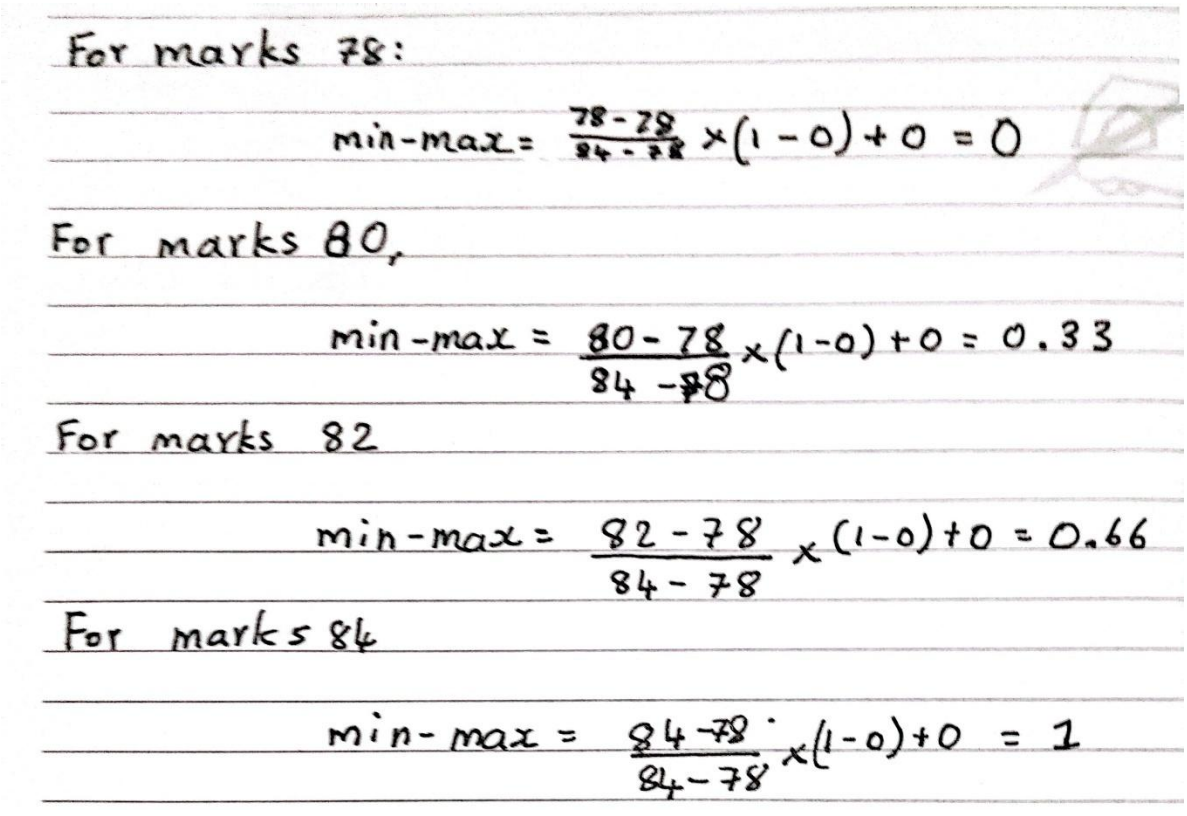
Bin 3: 27,27,35

15. Consider the following set: V= {78,80,82,84}. Apply Min-Max procedure and map the marks to a new range 0-1

To find min-max we use the formula:

$$min\text{-}max = \frac{V - min}{max - min} \times (new\ max - new\ min) + new\ min$$

Soln: Minimum of set V is 78, max is 84, *newmin* is 0 *newmax* is 1



Thus, min-max normalization gives a new set ={0, 0.33, 0.66, 1}

16. Consider the following list: V= {20,30,40}, convert the marks to z-score.

Given $V = \{20, 30, 40\}$

$$V^* = \frac{V - \mu}{\sigma}$$

$\mu \rightarrow$ mean

$\sigma \rightarrow$ Standard deviation

$$\mu = \frac{20+30+40}{3} = 30$$

$$\sigma = \sqrt{\frac{\Sigma(x_i - \mu)}{N}}$$

$$\sigma = \sqrt{\frac{(10)^2+0^2+(10)^2}{2}} = \sqrt{\frac{200}{2}} = \sqrt{100} = 10$$

$\Rightarrow \sigma = 10$

$$\text{z-score for } 20 = \frac{20-30}{10} = -1$$

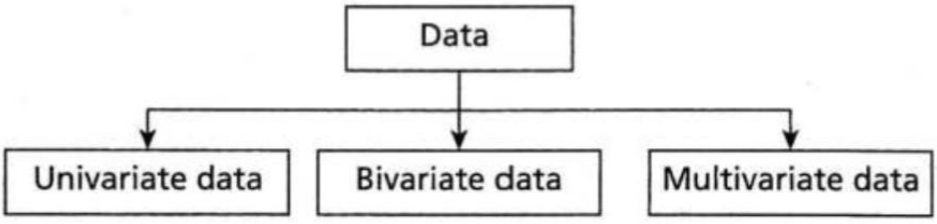
$$\text{z-score for } 30 = \frac{30-30}{10} = 0$$

$$\text{z-score for } 40 = \frac{40-30}{10} = 1$$

$\therefore V^* = \{-1, 0, 1\}$

18. Explain types of data based on variables.

Types of data based on variables can be classified into 3 groups as shown in below figure:



In case of univariate data, the dataset has only one variable. Remaining all attributes are independent values. A variable is also called as category.

Bivariate data indicates that the number of variables used are two.

Multivariate data uses three or more variables

17. What is descriptive statistics. Explain the classification of Data.

Descriptive Statistics:

Descriptive statistics is a branch of statistics that does dataset summarization. It is used to summarize and describe data. Descriptive statistics are just descriptive and do not go beyond that.

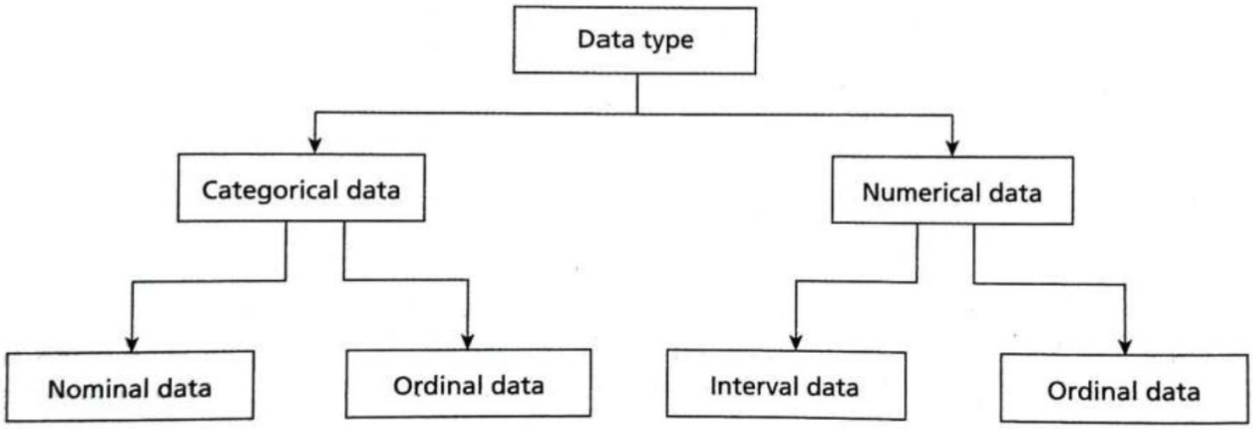
Data visualization is a branch of study that is useful for investigating the given data.

Descriptive analytics and data visualization techniques help to understand the nature of the data,

This step is often known as Exploratory Data Analysis (EDA). The focus of EDA is to understand the given data and to prepare it for ML algorithms. EDA includes descriptive statistics and data visualization.

Classification of Data

Types of data is summarized by figure below:



Broadly, data can be classified into two types:

1. **Categorical or qualitative data:**

The categorical data can be divided into two types. They are nominal type and ordinal type.

Nominal Data - Nominal data are symbols and cannot be processed like a number. Nominal data type provides only information but has no ordering among data. Only operations like (=, ≠) are meaningful for these data.

Ordinal Data - It provides enough information and has natural order. For example, Fever = (Low, Medium, High) is an ordinal data.

2. **Numerical or quantitative data**

It can be divided into two categories. They are interval type and ratio type.

Interval Data - Interval data is a numeric data for which the differences between values are meaningful. For example, there is a difference between 30 degree and 40 degree.

Ratio Data - For ratio data, both differences and ratio are meaningful. The difference between the ratio and interval data is the position of zero in the scale. For example, 0^* Centigrade ≠ 0^* Fahrenheit. Therefore, the data is not ratio data, it is interval data.

Another way of classifying the data is to classify it as:

1. **Discrete Data:** This kind of data is recorded as integers. For example, the responses of the survey can be discrete data. Items on a restaurant menu is discrete data.

2. **Continuous Data:** It can be fitted into a range and includes decimal point. For example, age is a continuous data.

19. Explain univariate data analysis and visualization.

Univariate analysis is the simplest form of statistical analysis. As the name indicates, the dataset has only one variable. The aim of univariate analysis is to describe data and find patterns.

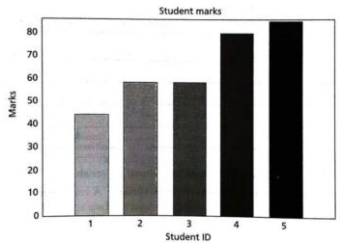
Univariate data description involves finding the frequency distributions, central tendency measures, dispersion or variation, and shape of the data.

Visualization of Univariate data:

There are many ways to visualize/present univariate data some of popular methods are:

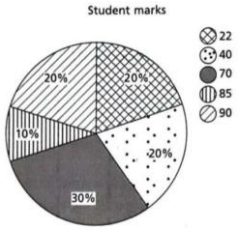
i. **Bar Graph:** A Bar chart (or Bar graph) is used to display the frequency distribution for variables. Bar charts are used to illustrate discrete data.

Example:



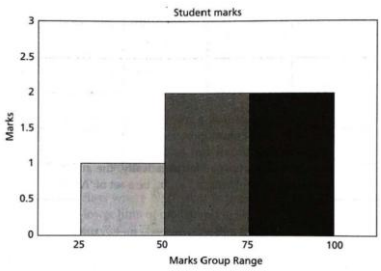
ii. **Pie Chart:** These are also useful in illustrating univariate data. It deals with percentage of data having a single value as opposed to frequency of the value.

Example:



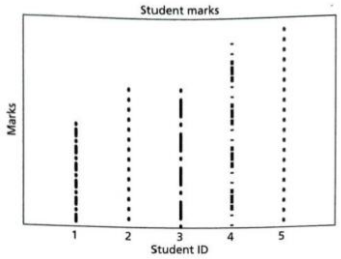
iii. **Histogram:** These are used to illustrate the frequency of continuous variables.

Example:



iv. **Dot Plots:** These are similar to bar charts. They are less clustered as compared to bar charts, as they illustrate the bars only with single points.

Example:



20. Explain Central tendency.

One cannot remember all the data. Therefore, a condensation or summary of the data is necessary. This makes the data analysis easy and simple. One such summary is called **central tendency**.

Mass data have tendency to concentrate at certain values, normally in the central location. It is called measure of central tendency. Popular measures are mean, median and mode.

1. Mean

i. Arithmetic Mean - Arithmetic average (or mean) is a measure of central tendency that represents the 'centre' of the dataset. It can be found by adding all the data and dividing the sum by the number of observations. It is denoted by $\bar{x}$ and is given by:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

ii. Weighted mean - Unlike arithmetic mean that gives the weightage of all items equally, weighted mean gives different importance to different items as the item importance varies. Hence, different weightage can be given to items.

iii. Geometric mean - Geometric mean is the N^{th} root of the product of N items. The formula for computing geometric mean is given as follows:

$$\text{Geometric mean} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{N}} = \sqrt[N]{x_1 \times x_2 \times \cdots \times x_N}$$

2. Median:

This is the middle most value in a sorted distribution. If total number of items are odd middle value is median, if number of items are even, average of 2 items in middle is median.

In continuous data Median is calculated as follows:

$$\text{Median} = L_1 + \frac{\frac{N}{2} - cf}{f} \times i$$

3. Mode:

The value that has the highest frequency is called mode. Mode is only for discrete data and is not applicable for continuous data. Normally, the dataset is classified as unimodal, bimodal and trimodal with modes 1, 2 and 3, respectively.

23. Plot Stem and Leaf plot for the History subject marks {55, 70, 70, 90, 95, 97}

stem	Leaf
5	5
6	
7	0 0
8	
9	0 5 7

21. What is Dispersion? Explain the measures of Dispersion.

The spreadout of a set of data around the central tendency is called dispersion. Dispersion is represented by various ways such as range, variance, standard deviation, and standard error.

Measures of Dispersion:

i. Range:

Range is the difference between the maximum and minimum of values of the given list of data.

ii. Standard Deviation

Standard deviation is the average distance from the mean of the dataset to each point. The formula for sample standard deviation is given by:

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

Here, N is the size of the population, x_i is observation or value from the population and μ is the population mean. Often, N - 1 is used instead of N in the denominator as N - 1 gives an answer closer to the actual value.

ii. **Quartile and Inter Quartile Range:**

Percentiles are about data that are less than the coordinates by some percentage of the total value.

k^{th} percentile is the property that the k% of the data lies at or below X_i

For example, median is 50th percentile and can be denoted as $Q_{0.50}$. The 25th percentile is called first quartile (Q_1) and the 75th percentile is called third quartile (Q_3). Another measure that is useful to measure dispersion is Inter Quartile Range (IQR). The IQR is the difference between Q, and Q_1

$$\text{Interquartile percentile} = Q_3 - Q_1$$

Outliers are normally the values falling apart at least by the amount 1.5 × IQR above the third quartile or below the first quartile.

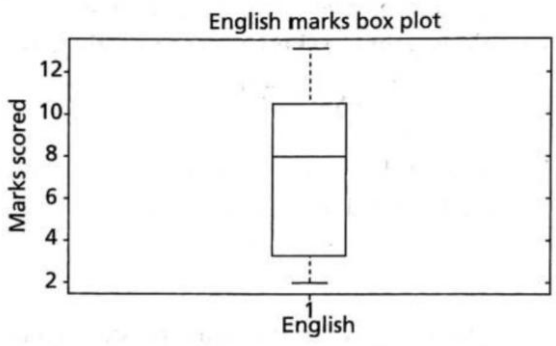
iii. **Five-point Summary and Box Plots:**

"<Minimum, Q_1, Median, Q_2, Maximum>" is called five-point summary.

Box plots can be used to illustrate data distributions and summary of data. It is the popular way for plotting five number summaries. A Box plot is also known as a Box and whisker plot.

The box contains bulk of the data that are between first and third quartiles. The line inside the box indicates median of the data. If the median is not equidistant, then the data is skewed. The lines that project from the ends of the box indicate the spread of the maximum and minimum of the data value.

Example:

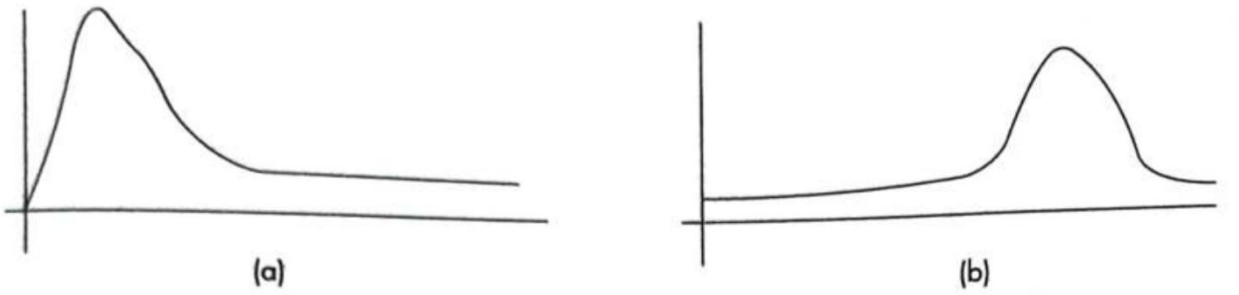


22. Explain Skewness and Kurtosis w.r.t Shape of data.

Skewness and Kurtosis indicate the symmetry/asymmetry and peak location of the dataset.

Skewness

Ideally, skewness should be zero as in ideal normal distribution. More often, the given dataset may not have perfect symmetry.



(a) Positive Skewed and (b) Negative Skewed Data

A data set having far higher values is said to be skewed to the right, also called as positive skewed Data

On other hand, a data having far more lower values is said to be left skewed or negative skewed data.

If the data is skewed there is a higher chance of outliers in the dataset.

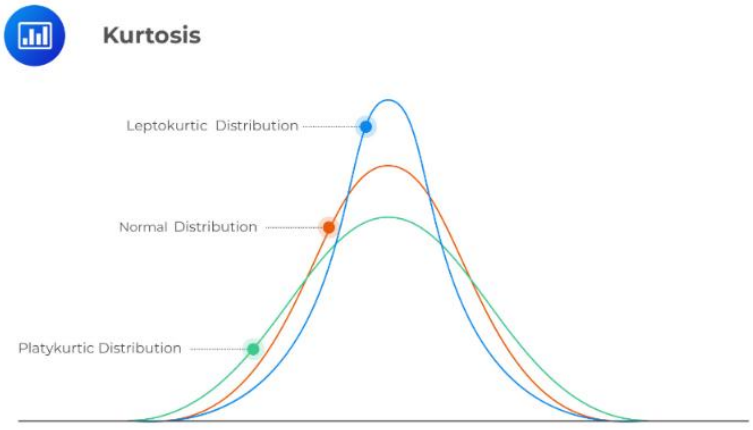
Skewness of data can be given as:

$$\frac{1}{N} \times \sum_{i=1}^{N} \frac{(x_i - \mu)^3}{\sigma^3}$$

Kurtosis

Kurtosis also indicates the peaks of data. If the data is high peak, then it indicates higher kurtosis and vice versa.

Kurtosis is the measure of whether the data is heavy tailed, or light tailed relative to normal distribution.



Leptokurthic distribution means it has higher kurtosis, Platykurtic distribution indicates low kurtosis.

Kurthosis is given by the following formula:

$$\frac{\sum_{i=1}^{N}(x_i - \bar{x})^4 / N}{\sigma^4}$$

Module 2

24. Explain Bivariate and multivariate data

Bivariate Data

Bivariate Data involves two variables. Bivariate data deals with causes of relationships. The aim is to find relationships among data.

Example:

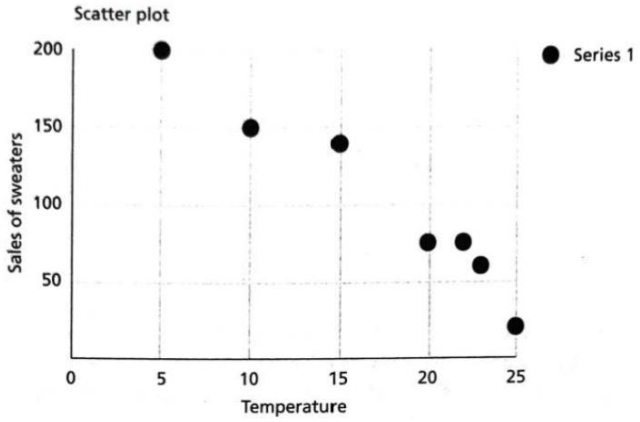
Temperature (in centigrade)	Sales of Sweaters (in thousands)
5	200
10	150
15	140
20	75
22	60
23	55
25	20

The relationships can then be used in comparisons, finding causes, and in further explorations. To do that, graphical display of the data is necessary. One such graph method is called scatter plot.

Scatter plot:

Scatter plot is used to visualize bivariate data. It is useful to plot two variables with or without nominal variables, to illustrate the trends, and also to show differences.

Example:



Multivariate Data:

The multivariate data is like bivariate data but may have more than two dependant variables. Some of the multivariate analysis are regression analysis, principal component analysis, and path analysis. In machine learning, almost all datasets are multivariable.

Visualization for multivariate data include Heatmap, Pair plots etc and the aim of multivariate analysis is much more than finding a relationship.

25. Explain Covariance and Correlation. Find the covariance and correlation of data X= {2, 3, 4, 6, 7} and Y = { 2, 5, 8, 15, 24}

Covariance:

Covariance is a measure of joint probability of random variables, say X and Y. It is defined as covariance(X, Y) or COV(X, Y), is used to measure the variance between two dimensions and is given by:

$$\text{cov}(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(X))(y_i - E(Y))$$

Correlation:

The Pearson correlation coefficient is the most common test for determining any association between two phenomena. It measures the strength and direction of a linear relationship between the x and y variables.

The correlation indicates the relationship between dimensions using its sign.

1. If the value is positive, it indicates that the dimensions increase together.
2. If the value is negative, it indicates that while one-dimension increases, the other dimension decreases.
3. If the value is zero, then it indicates that both the dimensions are independent of each other.

It is given by:

$$r = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$$

Problem:

Given $X = \{2,3,4,6,7\}$ and $Y = \{2,5,8,15,24\}$

$$E(X) = \frac{2+3+4+6+7}{5} = 4.4$$

$$E(Y) = \frac{2+5+8+15+24}{5} = 10.8$$

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

$$= \frac{(21.12 + 8.12 + 1.12 + 6.72 + 34.32)}{5}$$

$$\boxed{cov(x,y) = 14.28}$$

$$\sigma_x = \sqrt{\frac{(2-4.4)^2+(3-4.4)^2+(4-4.4)^2+(6-4.4)^2+(7-4.4)^2}{5}}$$

$$= \sqrt{3.44} = 1.854$$

$$\sigma_y = \sqrt{\frac{(2-10.8)^2+(5-10.8)^2+(8-10.8)^2+(15-10.8)^2+(24-10.8)^2}{5}}$$

Appointments

$$= \sqrt{62.16} = 7.884$$

$$r = \frac{14.28}{1.854 \times 7.884} \Rightarrow \boxed{r = 0.976}$$

26. Explain the steps involved for applying Gaussian Elimination. Solve the following set of equations using Gaussian Elimination method:

$2x_1 + 5x_2 = 7$

$6x_1 + 12x_2 = 18$

The procedure for applying Gaussian elimination is given as follows:

1. Write the given matrix.

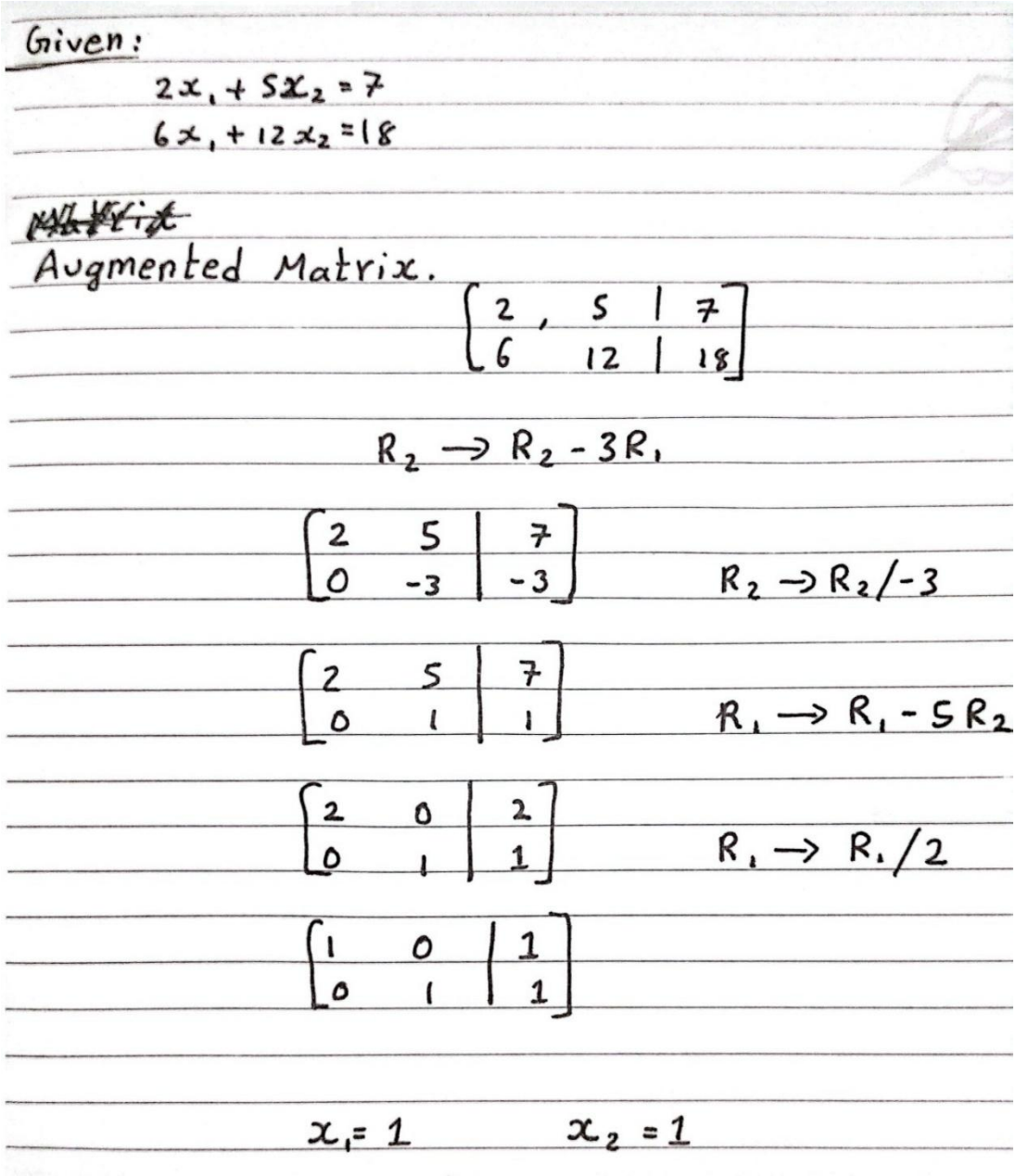
2. Append vector y to the matrix A. This matrix is called augmentation matrix.

3. Keep the element a_{11} as pivot and eliminate all a_{11} in second row using the matrix operation, $R_2 - \left(\frac{a_{21}}{a_{11}}\right)$, here R_2 is the second row and $\left(\frac{a_{21}}{a_{11}}\right)$ is called the multiplier. The same logic can be used to remove a_{11} in all other equations.

4. Repeat the same logic and reduce it to reduced echelon form. Find X from this form.

5. Substitute the x in any of the equation to get the value of y.

Problem:



27. Find LU decomposition of the given matrix:

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

Answer:

Given:

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

To find LU Decomposition

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 \\ 2 & 3 & 2 \\ 1 & 4 & 2 \end{bmatrix} \qquad R_2 \rightarrow R_2 - R_1$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 \\ 0 & 2 & -1 \\ 1 & 4 & 2 \end{bmatrix} \qquad R_3 \rightarrow R_3 - \frac{1}{2} R_1$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1/2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 \\ 0 & 2 & -1 \\ 0 & \frac{7}{2} & \frac{1}{2} \end{bmatrix} \qquad R_3 \rightarrow R_3 - \frac{7}{4} R_2$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ \frac{1}{2} & 7/4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & 2.25 \end{bmatrix}$$

Appointments

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0.5 & 1.75 & 1 \end{bmatrix} \qquad U = \begin{bmatrix} 2 & 1 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & 2.25 \end{bmatrix}$$

28. What are the two types of probability distributions. Explain the various probability distributions under each category.

The two types of probability distributions are:

1. Discrete probability distribution:

Binomial, Poisson, Bernoulli distribution fall under this category and these distributions deal with discrete values.

i. Binomial distribution:

Binomial distribution has only two outcomes: success or failure. This is also called Bernoulli trial.

The objective of this distribution is to find probability of getting success k out of n trials and is given as:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Its PDF is given as

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Its mean variance and standard deviation is given by:

$$\mu = n \times p \qquad \sigma^2 = np(1-p) \qquad \sigma = \sqrt{np(1-p)}$$

ii. Poisson Distribution:

It is another important distribution that is quite useful. Given an interval of time, this distribution is used to model the probability of k number of events happening. For Example, number of emails received in 1 hour etc.... Its PDF is gives by:

$$f(X = x; \lambda) = Pr[X = x] = \frac{e^{-\lambda}\lambda^x}{x!}$$

Mean is λ and Standard deviation is $\sqrt{\lambda}$

iii. Bernoulli Distribution:

This distribution models an experiment whose outcome is binary. The outcome is positive with *p* and negative with *1-p*. The PMF of this distribution is given as:

$$f(k;p) = \begin{cases} q = 1-p & \text{if } k = 0 \\ p & \text{if } k = 1. \end{cases}$$

Mean is p and variance is p(1-p) or pq.

2. **Continuous Probability Distributions:** Normal, Rectangular, and Exponential distributions fall under this category.

i. Normal Distribution:

Normal distribution is a continuous probability distribution. This is also known as gaussian distribution or bell-shaped curve distribution. It is the most common distribution function. The shape of this distribution is a typical bell-shaped curve. In normal distribution, data tends to be around a central value with no bias on left or right. The heights of the students, blood pressure of a population, and marks scored in a class can be approximated using normal distribution.

PDF of the normal distribution is given as:

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

ii. Rectangular Distribution

This is also known as uniform distribution. It has equal probabilities for all values in the range a, b. The uniform distribution is given as follows:

$$P(X = x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{Otherwise} \end{cases}$$

iii. Exponential Distribution

This is a continuous uniform distribution. This probability distribution is used to describe the time between events in a Poisson process. The PDF is gives as:

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases} \qquad (\lambda > 0)$$

29. Discuss density estimation in detail. Not there for IA-1

30. Explain various dimensionality reduction techniques. Not there for IA-1

31. Explain the process of obtaining principal components and its relevance in feature reduction.

Not there for IA-1

32. Apply PCA for the following matrix and prove that it works. Not there for IA-1

33. Apply SVD for the following matrix. Perform matrix decompositions and prove that SVD

Works Not there for IA-1

34. Explain the steps involved in design of a learning system

The design of learning systems focuses on these steps:

1. Choosing a training experience

- Let us consider designing of a chess game.
- In direct experience, individual board states and correct moves of the chess game are given directly. In indirect system, the move sequences and results are only given, the game agent plays against itself and learns the good moves.
- If the training samples and testing samples have the same distribution, the results would be good, hence training sample should cover maximum possible moves.

2. Choosing a target function

- In this step, the type of knowledge that needs to be learnt is determined.
- In direct experience, a board move is selected and is determined whether it is a good move or not against all other moves. If it is the best move, then it is chosen.
- In indirect experience, all legal moves are accepted, and a score is generated for each. The move with largest score is then chosen and executed.

3. Representation of a target function

The representation of knowledge may be a table, collection of rules or a neural network. The linear combination of these factors can be coined as:

$$V = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3$$

where, x_1, x_2 and x_3 represent different board features and w_0, w_1, w_2 and w_3 represent weights.

4. Function approximation

The focus is to choose weights and fit the given training samples effectively. The aim is to reduce the error given as:

$$E \equiv \sum_{\substack{\text{Training} \\ \text{Samples}}} \left[V_{train}(b) - \hat{V}(b) \right]^2$$

Here b is sample and $\hat{V}(b)$ is predicted hypothesis

The approximation is carried out in these 2 steps:

i. Error is computed. That is difference between trained and expected hypothesis.

ii. For every x_i the weight is updated using the formula:

$$w_i = w_i + \mu \times error(b) \times x_i$$

Here μ is constant which moderates change in weights.

Thus, the learning system has the following components:

- A Performance system to allow the game to play against itself.
- A Critic system to generate the samples.
- A Generalizer system to generate a hypothesis based on samples.
- An Experimenter system to generate a new system based on the currently learnt function. This is sent as input to the performance system.

35. Explain Concept Learning

Concept learning helps to classify an object that has a set of common, relevant features. Thus, it helps a learner compare and contrast categories based on the similarity and association of positive and negative instances in the training data to classify an object.

For example, humans can identify different kinds of animals based on common relevant features and categorize all animals based on specific sets of features. The special features that distinguish one animal from another can be called as a concept

This way of learning categories for object and to recognize new instances of those categories is called as concept learning

Concept learning requires three things:

1. Input - Training dataset which is a set of training instances, each labeled with the name of a concept or category to which it belongs.

2. Output - Target concept or Target function. It is a mapping function f(x) from input x to output y. It is to determine the specific features or common features to identify an object.

3. Test – New Instance to test the learned model.

Formally, Concept learning is defined as-"Given a set of hypotheses, the learner searches through the hypothesis space to identify the best hypothesis that matches the target concept".

36. Explain with example how to represent hypothesis.

A hypothesis 'h' approximates a target function 'f' to represent the relationship between the independent attributes and the dependent attribute of the training instances.

Each hypothesis is represented as a conjunction of attribute conditions.

Example: (Tail=Short)^(Color=Black).....

The set of hypothesis in the search space is called as hypotheses. Generally 'H' is used to represent the hypotheses and 'h' is used to represent a candidate hypothesis.

In a hypothesis, each attribute can take value as either'?' or 'Φ' or can hold a single value. Their meanings are as follows:

- "?" denotes that the attribute can take any value [e.g., Color = ?]
- " Φ" denotes that the attribute cannot take any value, i.e., it represents a null value [e.g., Horns = ϕ]
- Single value denotes a specific single value from acceptable values of the attribute, i.e., the attribute 'Tail' can take a value as 'short' [e.g., Tail = Short]

For example, a hypothesis '*h*' will look like,

	Horns	Tail	Tusks	Paws	Fur	Color	Hooves	Size
h =	<No	?	Yes	?	?	Black	No	Medium>

37. Define Generalization and Specialization hypothesis. Explain Specific to General Learning and General to Specific Learning with example.

Generalization – Specific to General Learning This learning methodology will search through the hypothesis space for an approximate hypothesis by generalizing the most specific hypothesis.

Consider the following table:

S.No.	Horns	Tail	Tusks	Paws	Fur	Color	Hooves	Size	Elephant
1.	No	Short	Yes	No	No	Black	No	Big	Yes
2.	Yes	Short	No	No	No	Brown	Yes	Medium	No
3.	No	Short	Yes	No	No	Black	No	Medium	Yes
4.	No	Long	No	Yes	Yes	White	No	Medium	No
5.	No	Short	Yes	Yes	Yes	Black	No	Big	Yes

In generalization we start with most specific hypothesis i.e.,

$h = \langle \varphi \quad \varphi \quad \varphi \quad \varphi \quad \varphi \quad \varphi \quad \varphi \quad \varphi \rangle$

And only consider positive instances and generalize the most specific hypothesis, ignoring negative instances.

Iteration 1:

I1: No Short Yes No No Black No Big **Yes (Positive instance)**
h1 = <No Short Yes No No Black No Big>

Iteration 2:

2nd instance is negative so we can safely ignore it. Therefore h2=h1

Iteration 3:

I3: No Short Yes No No Black No Medium **Yes (Positive instance)**
h3 = <No Short Yes No No Black No ?>

Iteration 4:

4nd instance is negative so we can safely ignore it. Therefore h4=h3

Iteration 5:

I5: No Short Yes Yes Yes Black No Big **Yes (Positive instance)**
h5 = <No Short Yes ? ? Black No ?>

This hypothesis is generated using generalisation method. And can classify any positive instances to true.

Specialization - General to Specific Learning

This learning methodology will search through the hypothesis space for an approximate hypothesis by specializing the most general hypothesis

Consider the same table. We start with most general hypothesis, that is all animals are elephants.

We check that positive instances are properly classified and negative instances are removed and add any hypothesis if needed.

Iteration 1:

*I*1:	No	Short	Yes	No	No	Black	No	Big	**Yes (Positive instance)**
*h*1 = <?	?	?	?	?	?	?	?>		

Iteration 2:

Since the instance is negative hypothesis is generated to reject the instance

*I*2 :	Yes	Short	No	No	No	Brown	Yes	Medium	**No (Negative instance)**
*h*2 = <No	?	?	?	?	?	?	?>		
	<?	?	Yes	?	?	?	?	?>	
	<?	?	?	?	?	Black	?	?>	
	<?	?	?	?	?	?	No	?>	
	<?	?	?	?	?	?	?	Big>	

Iteration 3:

*I*3:	No	Short	Yes	No	No	Black	No	Medium	**Yes (Positive instance)**
*h*3 = <No	?	?	?	?	?	?	?>		
	<?	?	Yes	?	?	?	?	?>	
	<?	?	?	?	?	Black	?	?>	
	<?	?	?	?	?	?	No	?>	
	<?	?	?	?	?	?	?	Big>	

Iteration 4:

In this iteration we remove any hypothesis inconsistent with negative hypothesis.

*I*4:	No	Long	No	Yes	Yes	White	No	Medium	**No (Negative instance)**
*h*4 = <?	?	Yes	?	?	?	?	?>		
	<?	?	?	?	?	Black	?	?>	
	<?	?	?	?	?	?	?	Big>	

Iteration 5:

*I*5:	No	Short	Yes	Yes	Yes	Black	No	Big	**Yes (Positive instance)**
*h*5 = <?	?	Yes	?	?	?	?	?>		
	<?	?	?	?	?	Black	?	?>	
	<?	?	?	?	?	?	?	Big>	

h5 is hypothesis space generated which classify both positive and negative instances correctly.

38. Write Find-S algorithm. What are the limitations?

Find-S Algorithm:

Input: Positive instances in the Training dataset
Output: Hypothesis 'h'

1. Initialize 'h' to the most specific hypothesis.
   $h = <\varphi \quad \varphi \quad \varphi \quad \varphi \quad \varphi \quad .....>$
2. Generalize the initial hypothesis for the first positive instance [Since 'h' is more specific].
3. For each subsequent instances:
   If it is a positive instance,
   Check for each attribute value in the instance with the hypothesis 'h'.
   If the attribute value is the same as the hypothesis value, then do nothing,
   Else if the attribute value is different than the hypothesis value, change it to '?' in 'h'.
   Else if it is a negative instance,
   Ignore it.

Limitations of Find-S Algorithm

1. Find-S algorithm tries to find a hypothesis that is consistent with positive instances, ignoring all negative instances.
2. The algorithm finds only one unique hypothesis, wherein there may be many other hypotheses that are consistent with the training dataset
3. Many times, the training dataset may contain some errors; hence such inconsistent data instances can mislead this algorithm in determining the consistent hypothesis since it ignores negative instances

39. Apply Find-S algorithm for the given dataset:

CGPA	Interactiveness	Practical Knowledge	Communication Skills	Logical Thinking	Interest	Job Offer
≥9	Yes	Excellent	Good	Fast	Yes	Yes
≥9	Yes	Good	Good	Fast	Yes	Yes
≥8	No	Good	Good	Fast	No	No
≥9	Yes	Good	Good	Slow	No	Yes

Step 1: Initialize h to most specific hypothesis. i.e.,

$h = <\varphi \quad \varphi \quad \varphi \quad \varphi \quad \varphi \quad \varphi>$

Step2: Generalize the hypothesis with the first positive instance of dataset:

I1: ≥9 Yes Excellent Good Fast Yes **Positive instance**
h = <≥9 Yes Excellent Good Fast Yes>

Step 3: Repeat until all instances are over, if instance is negative ignore it, if instance is positive and attribute value matches with hypothesis don't change it else replace it with '?'

Iteration 1:

I2: ≥9 Yes Good Good Fast Yes **Positive instance**
h = <≥9 Yes ? Good Fast Yes>

Iteration 2:

Instance 3 is negative so we can ignore it.

Iteration 3:

I4: ≥9 Yes Good Good Slow No **Positive instance**
h = <≥9 Yes ? Good ? ?>

So the final hypothesis with Find-S algorithm is:

$h = <\geq 9 \quad Yes \quad ? \quad Good \quad ? \quad ?>$

40. What is a Version Space? Write List-then-Eliminate algorithm.

Version space

The version space contains the subset of hypotheses from the hypothesis space that is consistent with all training instances in the training dataset.

List-Then-Eliminate algorithm.

It is used to find version spaces of datasets. The idea is list hypothesis space and eliminate any hypothesis inconsistent with dataset.

Algorithm:

Input: Version Space – a list of all hypotheses

Output: Set of consistent hypotheses

1. Initialize the version space with a list of hypotheses.
2. For each training instance,
   - remove from version space any hypothesis that is inconsistent.

41. Write Candidate Elimination algorithm.

Input: Set of instances in the Training dataset

Output: Hypothesis *G* and *S*

1. Initialize *G*, to the maximally general hypotheses.
2. Initialize *S*, to the maximally specific hypotheses.
   - Generalize the initial hypothesis for the first positive instance.
3. For each subsequent new training instance,
   - If the instance is **positive**,
     - Generalize *S* to include the positive instance,
       - Check the attribute value of the positive instance and *S*,
         - If the attribute value of positive instance and *S* are different, fill that field value with '?'.
         - If the attribute value of positive instance and *S* are same, then do no change.
     - Prune *G* to exclude all inconsistent hypotheses in *G* with the positive instance.
   - If the instance is **negative**,
     - Specialize *G* to exclude the negative instance,
       - Add to *G* all minimal specializations to exclude the negative example and be consistent with *S*.
         - If the attribute value of *S* and the negative instance are different, then fill that attribute value with *S* value.
         - If the attribute value of *S* and negative instance are same, no need to update '*G*' and fill that attribute value with '?'.
     - Remove from *S* all inconsistent hypotheses with the negative instance.

42. Apply Candidate Elimination algorithm for the given dataset

CGPA	Interactiveness	Practical Knowledge	Communication Skills	Logical Thinking	Interest	Job Offer
≥9	Yes	Excellent	Good	Fast	Yes	Yes
≥9	Yes	Good	Good	Fast	Yes	Yes
≥8	No	Good	Good	Fast	No	No
≥9	Yes	Good	Good	Slow	No	Yes

Step 1:

Initialize most general and most specific hypothesis.

$G = <?$? ? ? ? ?>

$S = <\varphi$ φ φ φ φ φ>

Step 2:

Generalize the initial hypothesis using first instance.

$I1$:	≥9	Yes	Excellent	Good	Fast	Yes	**Positive instance**
$S1 = <$	≥9	Yes	Excellent	Good	Fast	Yes>	
$G1 = <?$		?	?	?	?	?>	

Based on next instance if it is positive generalize S to include the instance and prune G of any inconsistent hypothesis. Repeat this until all instances are over. If it is negative Update G to exclude all features of negative instance and modify S if it is inconsistent.

Iteration 1:

$I2$:	≥9	Yes	Good	Good	Fast	Yes	**Positive instance**
$S2 = <$	≥9	Yes	?	Good	Fast	Yes>	

Iteration 2:

$I3$:	≥8	No	Good	Good	Fast	No	**Negative instance**
$G3 = <$	≥9	?	?	?	?	?>	
	<?	Yes	?	?	?	?>	
	<?	?	?	?	?	Yes>	
$S3 = <$	≥9	Yes	?	Good	Fast	Yes>	

Iteration 3:

$I4$:	≥9	Yes	Good	Good	Slow	No	**Positive instance**
$S4 = <$	≥9	Yes	?	Good	?	?>	
$G4 = <$	≥9	?	?	?	?	?>	
	<?	Yes	?	?	?	?>	

The final version space is:

<≥9	Yes	?	?	?	?>
<≥9	?	?	Good	?	?>
<?	Yes	?	Good	?	?>

Module-3

43. Explain the difference between Instance based and model-based learning.

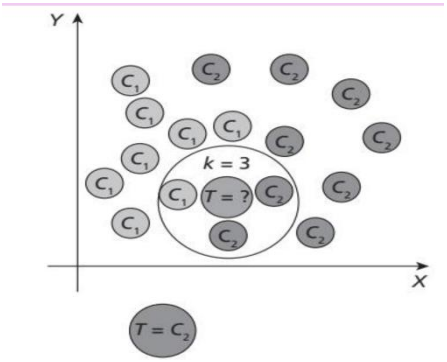
An instance is an entity or an example in the training dataset. Instance-based methods learn or predict the class label of a test instance only when a new instance is given for classification and until then it delays the processing of the training dataset.
It is also referred to as lazy learning methods since it does not generalize any model from the training dataset but just keeps the training dataset as a knowledge base until a new instance is given.
In contrast, model-based learning, generally referred to as *eager* learning, tries to generalize the training data to a model before receiving test instances. These algorithms basically learn in two phases, called training phase and testing phase. In training phase, a model is built from the training dataset and is used to classify a test instance during the testing phase. Some examples of models constructed are decision trees, neural networks and Support Vector Machines. But Instance based learning models delays processing of data till testing phase.

Instance-based Learning	Model-based Learning
Lazy Learners	Eager Learners
Processing of training instances is done only during testing phase	Processing of training instances is done during training phase
No model is built with the training instances before it receives a test instance	Generalizes a model with the training instances before it receives a test instance
Predicts the class of the test instance directly from the training data	Predicts the class of the test instance from the model built
Slow in testing phase	Fast in testing phase
Learns by making many local approximations	Learns by creating global approximation

44. Explain Nearest neighbor learning. Write K- Nearest neighbor learning algorithm.

A natural approach to similarity-based classification is k-Nearest-Neighbors (k-NN), which is a method used for both classification and regression problems. It is a simple and powerful algorithm that predicts the category of the test instance based on any of distance like Euclidian, Manhattan distance etc and picks “k” nearest neighbours and classifies based on probability of neighbouring nodes.

Visual representation of 3-NN is shown below:



k-NN Algorithm:

Inputs: Training dataset T, distance metric d, Test instance t, the number of nearest neighbors k
Output: Predicted class or category
Prediction: For test instance t,

1. For each instance i in T, compute the distance between the test instance t and every other instance i in the training dataset using a distance metric (Euclidean distance).
   [Continuous attributes - Euclidean distance between two points in the plane with coordinates (x_1, y_1) and (x_2, y_2) is given as dist $((x_1, y_1), (x_2, y_2)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$]
   [Categorical attributes (Binary) - Hamming Distance: If the value of the two instances is same, the distance d will be equal to 0 otherwise $d = 1$.]
2. Sort the distances in an ascending order and select the first k nearest training data instances to the test instance.
3. Predict the class of the test instance by majority voting (if target attribute is discrete valued) or mean (if target attribute is continuous valued) of the k selected nearest instances.

45. Consider the student performance training dataset of 8 data instances shown in the following table, classify whether a student will pass or fail in that course for a test instance (6.1, 40, 5)

S.No.	CGPA	Assessment	Project Submitted	Result
1.	9.2	85	8	Pass
2.	8	80	7	Pass
3.	8.5	81	8	Pass
4.	6	45	5	Fail
5.	6.5	50	4	Fail
6.	8.2	72	7	Pass
7.	5.8	38	5	Fail
8.	8.9	91	9	Pass

Soln:

Step 1: Calculate Euclidian distance from test node to every node of dataset:

S.No.	CGPA	Assessment	Project Submitted	Result	Euclidean Distance
1.	9.2	85	8	Pass	$\sqrt{(9.2-6.1)^2+(85-40)^2+(8-5)^2}$ $= 45.2063$
2.	8	80	7	Pass	$\sqrt{(8-6.1)^2+(80-40)^2+(7-5)^2}$ $= 40.09501$
3.	8.5	81	8	Pass	$\sqrt{(8.5-6.1)^2+(81-40)^2+(8-5)^2}$ $= 41.17961$
4.	6	45	5	Fail	$\sqrt{(6-6.1)^2+(45-40)^2+(5-5)^2}$ $= 5.001$
5.	6.5	50	4	Fail	$\sqrt{(6.5-6.1)^2+(50-40)^2+(4-5)^2}$ $= 10.05783$
6.	8.2	72	7	Pass	$\sqrt{(8.2-6.1)^2+(72-40)^2+(7-5)^2}$ $= 32.13114$
7.	5.8	38	5	Fail	$\sqrt{(5.8-6.1)^2+(38-40)^2+(5-5)^2}$ $= 2.022375$
8.	8.9	91	9	Pass	$\sqrt{(8.9-6.1)^2+(91-40)^2+(9-5)^2}$ $= 51.23319$

Step 2: Pick 3 lowest distances.

Instance	Euclidean Distance	Class
4	5.001	Fail
5	10.05783	Fail
7	2.022375	Fail

Step 3: Predict the class of text instance:

Since all of the nearest neighbours are “Fail”, The test instance will also be classified as “Fail”