SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY:: PUTTUR
Siddharth Nagar, Narayanavanam Road – 517583

QUESTION BANK (DESSCRIPTIVE)

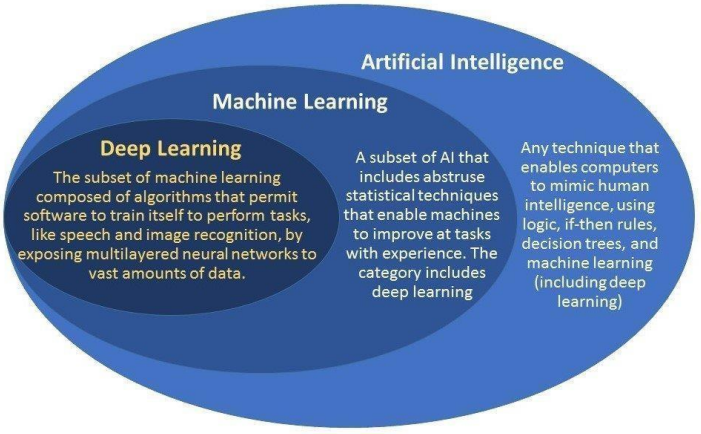
Subject with Code: INTRODUCTION TO MACHINE LEARNING (20CS0904)

Course & Branch: B. Tech – CSM

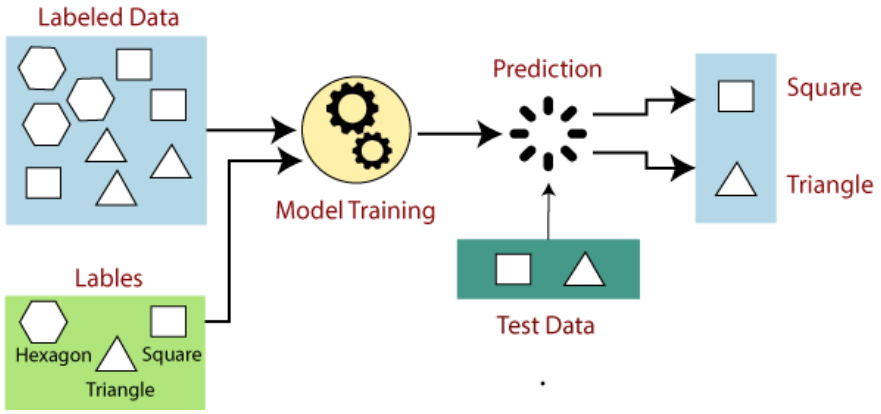
Year & Sem: III B.Tech & I-Sem

Regulation: R20

UNIT –I
INTRODUCTION MACHINE LEARNING

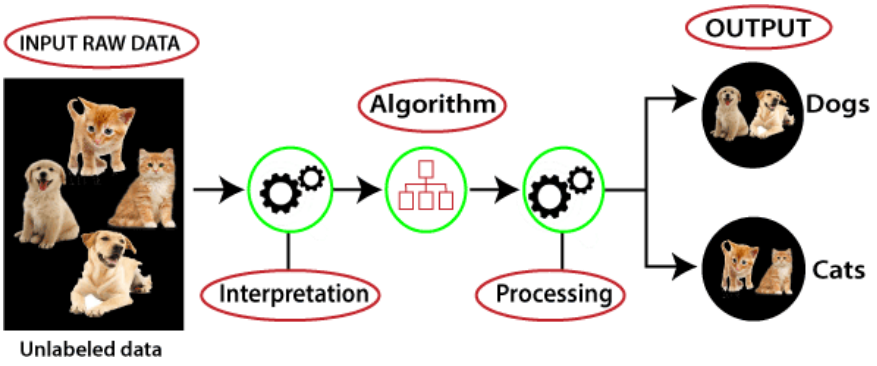
1	Describe about Machine Learning algorithms with their predictions.	**[L2][CO1]**	**[12M]**
	Machine Learning is a subset of artificial intelligence (AI) that involves the development of algorithms and statistical models that enable computers to learn from data and make predictions or decisions without being explicitly programmed. In other words, Machine Learning is a process of teaching computers to learn from data and improve their performance over time.  Machine Learning is used in a variety of applications such as image recognition, speech recognition, natural language processing, recommendersystems, and predictive analytics. These applications involve the use of large amounts of data and sophisticated algorithms to identify patterns and relationships between variables. Predictions are a key aspect of Machine Learning. Machine Learning models use statistical algorithms to analyse data and make predictions about future outcomes. For example, a Machine Learning model can be used to predict whether a customer is likely to buy a product, based on their past purchasing behaviour and other data points. In the healthcare industry, Machine Learning models can be used to predict the likelihood of a patient developing a particular disease based on their medical history and other risk factors. The accuracy of Machine Learning predictions depends on the quality and quantity of data used to train the model, as well as the sophistication of the algorithms used. Machine Learning models can be further improved by adjusting the parameters of the algorithms or by adding more data to the training set. As Machine Learning continues to advance, it has the potential to		

	revolutionize many industries by providing accurate predictions and insights that can help businesses make better decisions and improve outcomes forindividuals.		
2	Define basic concepts in Machine Learning.	**[L1][CO1]**	**[12M]**
	1. Data: Machine Learning algorithms learn from data. The data can be in the form of structured data (e.g., rows and columns in a spreadsheet) or unstructured data (e.g., text, images, audio). Data is typically split into a training set (used to train the model) and a test set (used to evaluate the model's performance). 2. Feature: A feature is a measurable attribute of the data that can be used to make predictions. For example, if we're trying to predict whether a loan will be approved, the features might include the applicant's income, credit score, and employment status. 3. Model: A model is a mathematical representation of the relationships between the features and the target variable (the variable we're trying to predict). The goal of Machine Learning is to find the best model that accurately predicts the target variable. 4. Training: During the training process, the Machine Learning algorithm adjusts the parameters of the model to minimize the error between the predicted output and the actual output. This is done by feeding the algorithm the training set of data and updating the model's parameters after each iteration. 5. Testing: After the model has been trained, it is tested on a separate test set of data to evaluate its performance. The performance is measured using various metrics such as accuracy, precision, recall, and F1 score. 6. Supervised learning: In supervised learning, the training data includes both the features and the target variable. The goal is to learn a mapping between the features and the target variable. Examples of supervised learning include classification (predicting a categorical variable) and regression (predicting a continuous variable). 7. Unsupervised learning: In unsupervised learning, the training data only includes the features. The goal is to discover patterns or relationships in the data. Examples of unsupervised learning include clustering (grouping similar data points together) and dimensionality reduction (reducing the number of features while preserving the most important information). 8. Reinforcement learning: In reinforcement learning, the algorithm learns through trial and error. The algorithm receives feedback in the form of rewards or penalties based on its actions. The goal is to learn a policy that maximizes the expected reward over time. Reinforcement learning is often used in robotics and game playing.		
3	Discuss the Machine Learning techniques with neat diagrams	**[L2][CO2]**	**[12M]**
	1. Regression: Regression is a technique used to predict continuous numerical values. Linear regression is a popular method where the model learns a linear relationship between the input features and the output. 2. Classification: Classification is a technique used to predict categorical values. The most popular algorithm for classification is logistic regression. Other popular algorithms include decision trees, random forests, and support vector machines.		

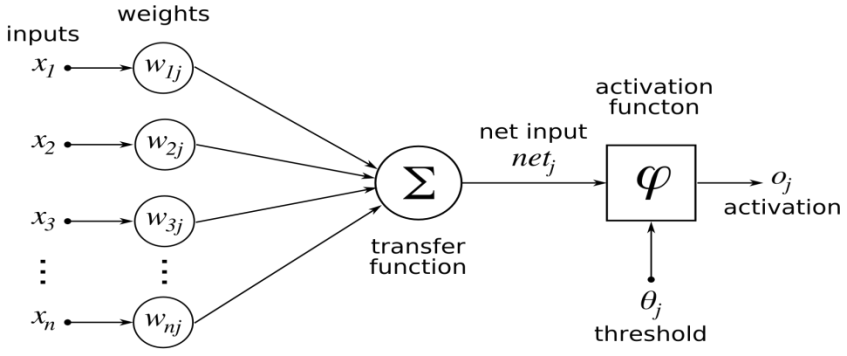


3. Clustering: Clustering is a techni que used to group similar data points together based on their features. Clustering algorithms include k-means, hierarchical clustering, and DBSCAN.

4. Dimensionality reduction: Dimensionality reduction is a technique used to reduce the number of input features while preserving the most important information. Principal component analysis (PCA) is a popular method for dimensionality reduction.

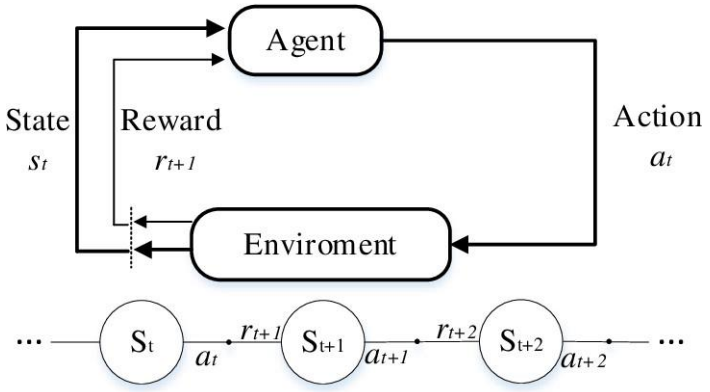


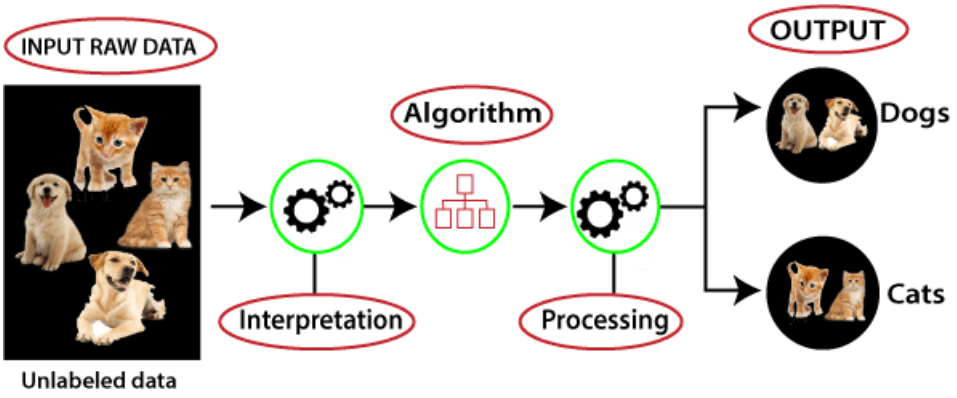
5. Neural networks: Neural networks are a set of algorithms modeled after the human brain. They consist of layers of interconnected nodes and can be used for a variety of tasks such as image recognition, natural language processing, and speech recognition.



6. Deep learning: Deep learning is a subset of neural networks that uses multiple layers of nodes to learn increasingly complex representations of the data. Deep learning has been used to achieve state-of-the-art performance in tasks such as image recognition and natural language processing.

7. Reinforcement learning: Reinforcement learning is a technique used to teach machines to learn through trial and error. The algorithm receives

	feedback in the form of rewards or penalties based on its actions, and its goal is to learn a policy that maximizes the expected reward over time.  8. Ensemble learning: Ensemble learning is a technique that combines multiple models to improve the accuracy of predictions. Examples of ensemble methods include bagging, boosting, and stacking.		
4	Explain about Supervised Learning techniques.	**[L2][CO3]**	**[12M]**
	Supervised Learning is a type of Machine Learning where the algorithm learns to predict an output variable (also known as a target variable or dependent variable) from input variables (also known as features or independent variables) based on labeled data. In other words, the algorithm learns from examples where both the input and output variables are known. Here are some common techniques used in Supervised Learning: 1. Linear Regression: Linear regression is a technique used to predict a continuous output variable based on one or more input variables. The goal is to find a linear relationship between the input variables and the output variable. Linear regression can be used for both simple and multiple linear regression problems. 2. Logistic Regression: Logistic regression is a technique used to predict a categorical output variable based on one or more input variables. The goal is to find a relationship between the input variables and the probability of the output variable being in a certain category. 3. Decision Trees: Decision trees are a technique used to predict a categorical or continuous output variable based on one or more input variables. Decision trees are built by recursively splitting the data into subsets based on the values of the input variables until a stopping criterion is met. 4. Random Forests: Random forests are a type of ensemble learning technique that combines multiple decision trees to improve the accuracy of predictions. Random forests are built by constructing multiple decision trees on randomly selected subsets of the data and then averaging their predictions. 5. Support Vector Machines (SVMs): SVMs are a technique used to predict a categorical or continuous output variable based on one or more input variables. The goal is to find a hyperplane that separates the data into different classes with the largest possible margin. 6. Naive Bayes: Naive Bayes is a probabilistic technique used to predict a categorical output variable based on one or more input variables. Naive		

	Bayes assumes that the input variables are independent of each other and calculates the probability of the output variable based on the probabilities of the input variables. These are just some of the techniques used in Supervised Learning. Each technique has its own strengths and weaknesses, and the choice of technique depends on the specific task and the characteristics of the data.		
5	Explain the Un-Supervised Learning techniques.	**[L2][CO2]**	**[12M]**
	Unsupervised Learning is a type of Machine Learning where the algorithm learns to identify patterns and structures in the data without being explicitly trained on labelled data. In other words, the algorithm learns to find relationships and groupings in the data on its own.  Here are some common techniques used in Unsupervised Learning: 1. Clustering: Clustering is a technique used to group similar data points together based on their features. The goal is to identify natural groupings in the data without prior knowledge of the group labels. Common clustering algorithms include k-means, hierarchical clustering, and DBSCAN. 2. Dimensionality Reduction: Dimensionality reduction is a technique used to reduce the number of input features while preserving the most important information. The goal is to simplify the data and make it easier to analyse. Common dimensionality reduction techniques include principal component analysis (PCA), t-distributed stochastic neighbour embedding (t-SNE), and autoencoders. 3. Anomaly Detection: Anomaly detection is a technique used to identify data points that are significantly different from the rest of the data. The goal is to detect unusual patterns or outliers in the data. Common anomaly detection techniques include density-based methods, distance- based methods, and clustering-based methods. 4. Association Rule Learning: Association rule learning is a technique used to identify relationships between variables in the data. The goal is to find patterns in the data such as frequent itemsets and association rules. Common association rule learning algorithms include Apriori and FP-Growth. 5. Generative Models: Generative models are a type of Unsupervised Learning technique that learns to generate new data samples that are similar to the input data. The goal is to learn the underlying structure of		

		the data and use it to generate new samples. Common generative models include variational autoencoders, generative adversarial networks (GANs), and Boltzmann machines. These are just some of the techniques used in Unsupervised Learning. Each technique has its own strengths and weaknesses, and the choice of technique depends on the specific task and the characteristics of the data.		
6	a)	What is the role of pre-processing of data in machine learning? Why it is needed?	**[L3][CO1]**	**[6M]**
		Major steps of Data Preprocessing are: 1. Data Acquisition 2. Data Normalization/Cleaning 3. Data Formatting 4. Data Sampling 5. Data Scaling Manipulating data is often the most time-consuming part of data science. So much so that in many enterprises, data analysts spend much of their valuable time preparing the data rather than drawing insights from it,which is the main task. Data preprocessing is where you start to ―prepare‖ the data for the machine learning algorithm. There are a few different types of preprocessing that you can do. you can, for example, filter the data to remove any invalid entries. You can also reduce the size of the dataset to make it easier to process. You can also normalize the data to make it more consistent. Here are the some of the major steps in Data preprocessing : **Step 1: Data Acquisition** This is probably the most important step in the preprocessing process. The data you will be working with will almost certainly come from somewhere. In the case of machine learning, it's usually a spreadsheet application (Excel, Google Sheets, Etc.) that is manipulated by someone else. In the best case, it's a tool like R or Python that you can use to grab the data and perform some basic manipulations easily. There are a few things to note here. First, the data you'll be working with might be in a format that is not directly usable by the machine learning algorithm. For example, if you're trying to load data from an SPSS file, you'll need to do some cleaning to get the data into a valid format. Second, the tools you mentioned can also do quite a bit of cleaning, but sometimes it's more explicit data processing that you're looking for. Before you take the next step, you will need to import all the libraries like Python for the preprocessing tasks. You may also use the Python programming language and its built-in data library to perform more sophisticated data processing.		

The three core Python libraries for this purpose are Pandas, NumPy, and Matplotlib to easily manipulate your data in several ways.

Step 2: Data Normalization/Cleaning

Here, you delete the unwanted data and fix the instances of missing data by removing them. The term ―data cleaning‖ is a little misleading because it makes it sound like you're just trying to fix the data. In reality, you're trying to eliminate errors and inconsistencies so that our data is as consistent as possible.

This means removing any invalid or erroneous values. There are a number of things you can do here. You can make sure that each data item is unique and standardize various properties of the data, such as their unit of measurement. Ensure that each data point has a uniquely determined value.

This means no duplicates and no missing values.

Step 3: Data Formatting

Data formatting begins once you have clean data. It helps convert the data into a more usable format by machine learning algorithms. Data can be available in different forms, including proprietary and Parquet formats, among others. Learning models can work effectively with data when its formatting is appropriate.

You can use several different formats, and each has its own benefits.

One popular option these days goes under the brand name Tensor Flow or TFRecords, enabling us to establish one unified set of labeled training records across different models within MLflow for flexible model auditing.

Step 4: Data Sampling

You need to ensure that the data samples represent the population from which they came, for this is where bias and variance can come into play. Bias is the tendency for data to exhibit patterns that are not representative of the population from which it came.

One of the most important things you can do when working with data is to ensure you're sampling it properly. This means that you're taking a representative sample of the data rather than just grabbing whatever data is available. Instead of picking the entire dataset, you can use a smaller sample of the whole, thus saving time and memory space.

This is also important because it ensures you get a fair data representation. You'll get questionable results with biases if you're sampling too heavily in one direction.

Also, you need to split the dataset into two – for training and test purposes. Training sets are subsets of datasets used for training machine learning models. The output is something you already know. In contrast, a test set is a subset of the dataset useful for testing the machine learning model. To predict outcomes, the ML model uses the test set.

A 70:30 or 80:20 ratio is usually used for the dataset, i.e. you take either 70% or 80% of the data for training the model while leaving out the rest 30% or 20% for testing. What guides this decision is the form and size of the dataset in question.

Step 5: Data Scaling

Data Scaling is the standardization of independent variables within a range. To put it another way, feature scaling limits the range of variables so that their comparison is fair.

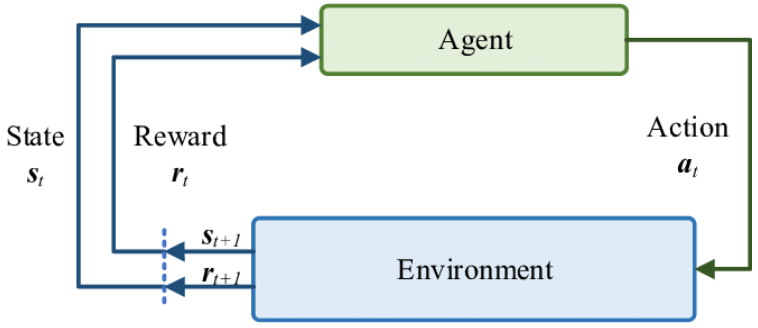
Standardizing the features of a dataset reduces the variability within a dataset so that its comparison and analysis become easier. Like 0-100 or 0-1. It helps ensure that the data you've received has similar properties.

There are several different ways that you can standardize the data. For example, you can use standard deviation to reduce the variance within a dataset.

Once the preprocessing steps are done, you need to undertake the rest of the data processing steps like data transformation before loading the data into the machine learning algorithm and training the algorithm. This is, essentially, a process of ―teaching‖ the machine learning algorithm how to recognize and understand the patterns in our data.

	b)	Analyze Reinforcement Learning with neat diagram.	**[L4][CO1]**	**[6M]**

- 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 labelled data, unlike supervised learning.
- Since there is no labelled data, so the agent is bound to learn by its experience only.
- "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."



- **Agent** – is the sole decision-maker and learner
- **Environment** – a physical world where an agent learns and decides the actions to be performed
- **Action** – a list of action which an agent can perform
- **State** – the current situation of the agent in the environment
- **Reward** – For each selected action by agent, the environment gives a reward. It's usually a scalar value and nothing but feedback from the environment
- **Policy** – the agent prepares strategy (decision-making) to map situations to actions.

Here are some common techniques used in Reinforcement Learning:

		1. Q-Learning: Q-Learning is a technique used to learn the optimal policy for a given environment by estimating the Q-values of each state-action pair. The Q-value represents the expected reward for taking a certain action at a certain state. 2. Deep Reinforcement Learning: Deep Reinforcement Learning is a technique that uses deep neural networks to estimate the Q-values of each state-action pair. This allows the agent to handle high-dimensional input and learn more complex policies. 3. Actor-Critic: Actor-Critic is a technique that combines elements of both value-based and policy-based methods. It uses a value function to estimate the Q-value and an actor function to select the action. 4. Policy Gradient: Policy Gradient is a technique that directly optimizes the policy by updating the parameters of the policy function based on the gradient of the expected reward. These are just some of the techniques used in Reinforcement Learning. Each technique has its own strengths and weaknesses, and the choice of technique depends on the specific task and the characteristics of the environment.		
7	a)	Explain data processing and techniques used for data preprocessing.	**[L2][CO1]**	**[6M]**
		As same as Answer 6a)		
	b)	Analyze the real world applications of ML.	**[L4][CO1]**	**[6M]**
		The following are applications of Machine Learning 1. Image Recognition: Image recognition is one of the most common applications of machine learning. It is used to identify objects, persons, places, digital images, etc. The popular use case of image recognition and face detection is, **Automatic friend tagging suggestion**: 2. Speech Recognition: While using Google, we get an option of "**Search by voice**," it comes under speech recognition, and it's a popular application of machine learning. Speech recognition is a process of converting voice instructions into text, and it is also known as "**Speech to text**", or "**Computer speech recognition**." At present, machine learning algorithms are widely used by various applications of speech recognition. **Google assistant**, **Siri**, **Cortana**, and **Alexa** are using speech recognition technology to follow the voice instructions. 3. Traffic prediction: If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions. It predicts the traffic conditions such aswhether traffic is cleared, slow-moving, or heavily congested with the help of two ways: • **Real Time location** of the vehicle form Google Map app and sensors		

- **Average time has taken** on past days at the same time.

Everyone who is using Google Map is helping this app to make it better. It takes information from the user and sends back to its database to improve the performance.

5. Product recommendations: Machine learning is widely used by various e-commerce and entertainment companies such as **Amazon**, **Netflix**, etc., for product recommendation to the user. Whenever we search for some product on Amazon, then we started getting an advertisement for the same product while internet surfing on the same browser and this is because of machine learning.

Google understands the user interest using various machine learning algorithms and suggests the product as per customer interest.

As similar, when we use Netflix, we find some recommendations for entertainment series, movies, etc., and this is also done with the help of machine learning.

6. Self-driving cars: One of the most exciting applications of machine learning is self-driving cars. Machine learning plays a significant role in self-driving cars. Tesla, the most popular car manufacturing company is working on self-driving car. It is using unsupervised learning method to train the car models to detect people and objects while driving.

7. Email Spam and Malware Filtering:

Whenever we receive a new email, it is filtered automatically as important, normal, and spam. We always receive an important mail in our inbox with the important symbol and spam emails in our spam box, and the technology behind this is Machine learning. Below are some spam filters used byGmail:

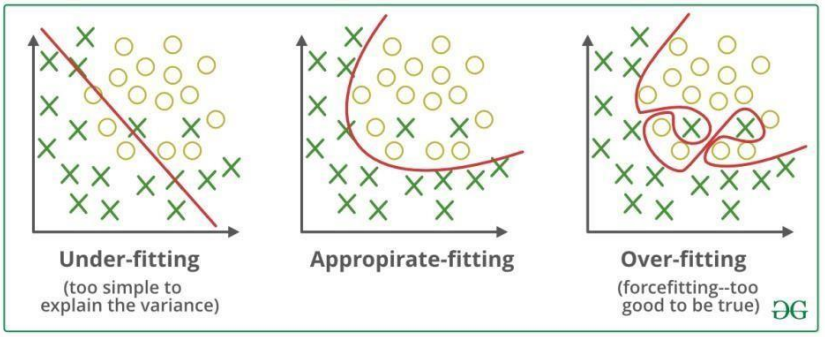
- Content Filter
- Header filter
- General blacklists filter
- Rules-based filters
- Permission filters

Some machine learning algorithms such as **Multi-Layer Perceptron**, **Decision tree**, and **Naïve Bayes classifier** are used for email spam filtering and malware detection.

8. Virtual Personal Assistant: We have various virtual personal assistants such as **Google assistant**, **Alexa**, **Cortana**, **Siri**. As the name suggests, they help us in finding the information using our voice instruction. These assistants can help us in various ways just by our voice instructions such as Play music, call someone, Open an email, Scheduling an appointment, etc.

These virtual assistants use machine learning algorithms as an important

		part. These assistant record our voice instructions, send it over the server on a cloud, and decode it using ML algorithms and act accordingly. 9. Online Fraud Detection: Machine learning is making our online transaction safe and secure by detecting fraud transaction. Whenever we perform some online transaction, there may be various ways that a fraudulent transaction can take place such as **fake accounts**, **fake ids**, and **steal money** in the middle of a transaction. So to detect this, **Feed Forward Neural network** helps us by checking whether it is a genuine transaction or a fraud transaction. For each genuine transaction, the output is converted into some hash values, and these values become the input for the next round. For each genuine transaction, there is a specific pattern which gets change for the fraud transaction hence, it detects it and makes our online transactions moresecure. 10. Stock Market trading: Machine learning is widely used in stock market trading. In the stock market, there is always a risk of up and downs in shares, so for this machine learning's **long short term memory neural network** is used for the prediction of stock market trends. 11. Medical Diagnosis:In medical science, machine learning is used for diseases diagnoses. With this, medical technology is growing very fast and able to build 3D models that can predict the exact position of lesions in the brain. It helps in finding brain tumors and other brain-related diseaseseasily. 12. Automatic Language Translation:Nowadays, if we visit a new place and we are not aware of the language then it is not a problem at all, as for this also machine learning helps us by converting the text into our known languages. Google's GNMT (Google Neural Machine Translation) provide this feature, which is a Neural Machine Learning that translates the text into our familiar language, and it called as automatic translation.The technology behind the automatic translation is a sequence to sequence learning algorithm, which is used with image recognition and translates the text from one language to another language.		
8	Write about brief explanation for Probability theory		**[L3][CO1]**	**[12M]**
	Probability theory is a branch of mathematics that deals with the study of random events and their likelihood of occurrence. It provides a framework for analysing uncertain situations and making predictions based on the available information. The fundamental concept of probability theory is the probability of an event, which is a measure of the likelihood of that event occurring. Probability is usually expressed as a number between 0 and 1, where 0 represents an impossible event and 1 represents a certain event. For example, the probability of flipping a fair coin and getting heads is 0.5, or 50%.			

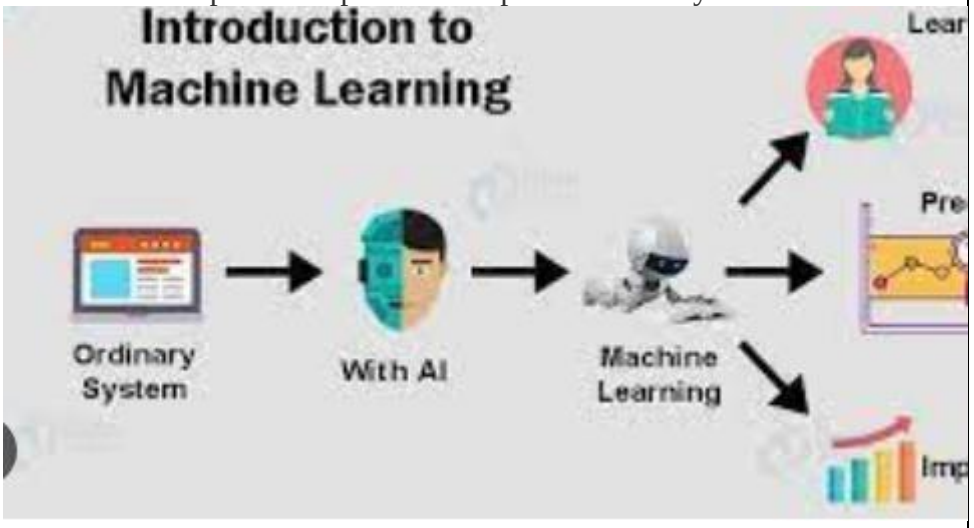
		Probability theory also includes concepts such as random variables, which are variables that can take on different values in a random manner, and probability distributions, which describe the probabilities of different outcomes of a random variable. There are many different types of probability distributions, including the normal distribution, the binomial distribution, and the Poisson distribution, each of which has its own properties and uses. In addition to these basic concepts, probability theory also includes tools and techniques for analysing and manipulating probabilities, such as Bayes' theorem, which allows for the updating of probabilities based on new information, and hypothesis testing, which is used to test the validity of statistical claims based on sample data. Probability theory has a wide range of applications in many fields, including statistics, physics, engineering, economics, and finance. It is also an essential foundation for many machine learning and artificial intelligence techniques, such as Bayesian networks and probabilistic graphical models, which rely on probabilistic reasoning to make predictions and decisions		
9	a)	Differentiate the Bias and Variance tradeoff in Machine Learning.	**[L4][CO1]**	**[6M]**
		Bias and variance are two important concepts in Machine Learning that are related to the performance of a model. The tradeoff between bias and variance is a key consideration when designing and training machine learning models. Bias refers to the error that is introduced by approximating a real-world problem with a simpler model. Models with high bias tend to be too simple and may underfit the data, resulting in poor performance on both the training and test sets. In other words, the model is too rigid and cannot capture the complexity of the underlying data.  Variance, on the other hand, refers to the error that is introduced by the model being too sensitive to the noise in the training data. Models with high variance tend to be too complex and may overfit the data, resulting in good performance on the training set but poor performance on the test set. In other words, the model is too flexible and captures not only the underlying patterns but also the noise in the data. The bias-variance tradeoff refers to the tradeoff between the two types of errors. As the complexity of the model increases, the variance tends to increase while the bias tends to decrease. Conversely, as the complexity of the model decreases, the bias tends to increase while the variance tends to decrease. The goal is to find a balance between bias and variance that results in a model that performs well on both the training and test sets. In practice, this means choosing a model that is complex enough to capture the underlying patterns in the data but not so complex that it		

overfits the data. Techniques such as regularization, cross-validation, and ensemble methods can help achieve this balance and improve the performance of machine learning models.

b) Compare Machine Learning and Artificial Intelligence. **[L4][CO1]** **[6M]**

Sl.No.	ARTIFICIAL INTELLIGENCE	MACHINE LEARNING
1.	1956 The terminology —Artificial Intelligence‖ was originally used by John McCarthy, who also hosted the first AI conference.	The terminology —Machine Learning‖ was first used in 1952 by IBM computer scientist Arthur Samuel, a pioneer in artificial intelligence and computer games.
2.	AI stands for Artificial intelligence, where intelligence is defined as the ability to acquire and apply knowledge.	ML stands for Machine Learning which is defined as the acquisition of knowledge or skill
3.	AI is the broader family consisting of ML and DL as its components.	Machine Learning is the subset of Artificial Intelligence.
4.	The aim is to increase the chance of success and not accuracy.	The aim is to increase accuracy, but it does not care about; the success
5.	AI is aiming to develop an intelligent system capable of performing a variety of complex jobs. decision-making	Machine learning is attempting to construct machines that can only accomplish the jobs for which they have been trained.
6.	It works as a computer program that does smart work.	Here, the tasks systems machine takes data and learns from data.
7.	The goal is to simulate natural intelligence to solve complex problems.	The goal is to learn from data on certain tasks to maximize the performance on that task.
8.	AI has a very broad variety of applications.	The scope of machine learning is constrained.

9.	AI is decision-making.	ML allows systems to learn new things from data.
10.	It is developing a system that mimics humans to solve problems.	It involves creating self-learning algorithms.
11.	AI will go for finding the optimal solution.	ML will go for a solution whether it is optimal or not.
12.	AI leads to intelligence or wisdom.	ML leads to knowledge.
13.	AI is a broader family consisting of ML and DL as its components.	ML is a subset of AI.
14.	Three broad categories of AI are : 1. Artificial Narrow Intelligence (ANI) 2. Artificial General Intelligence (AGI) 3. Artificial Super Intelligence (ASI)	Three broad categories of ML are : 1. Supervised Learning 2. Unsupervised Learning 3. Reinforcement Learning
15.	AI can work with structured, semi-structured, and unstructured data.	ML can work with only structured and semi-structured data.
16.	AI's key uses include- • Siri, customer service via chatbots • Expert Systems • Machine Translation like Google Translate • Intelligent humanoid robots such as Sophia, and so on.	The most common uses of machine learning- • Facebook's automatic friend suggestions • Google's search algorithms • Banking fraud analysis • Stock price forecast • Online recommender systems, and so on.
17.	AI refers to the broad field of creating machines that can simulate human intelligence and perform tasks such as	ML is a subset of AI that involves training algorithms on data to make predictions, decisions, and recommendations.

		language, recognizing images and sounds, making decisions, and solving complex problems.		
10	a)	What is Machine learning? Explain the need of it.	**[L2][CO1]**	**[6M]**
		Machine Learning is said as a subset of **artificial intelligence** that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own. The term machine learning was first introduced by **Arthur Samuel** in **1959. A machine has the ability to learn if it can improve its performance by gaining more data.** The need for machine learning is increasing day by day. The reason behind the need for machine learning is that it is capable of doing tasks that are too complex for a person to implement directly.  **Following are some key points which show the importance of Machine Learning:** • Rapid increment in the production of data • Solving complex problems, which are difficult for a human • Decision making in various sector including finance Finding hidden patterns and extracting useful information from data.		
	b)	List out applications and some popular algorithms used in Machine Learning. Explain it. Applications of Machine Learning: 1. Image Recognition: Image recognition is one of the most common applications of machine learning. It is used to identify objects, persons, places, digital images, etc. The popular use case of image recognition and face detection is, **Automatic friend tagging suggestion**:	**[L1][CO1]**	**[6M]**

2. Speech Recognition

While using Google, we get an option of "**Search by voice**," it comes under speech recognition, and it's a popular application of machine learning.

Speech recognition is a process of converting voice instructions into text, and it is also known as "**Speech to text**", or "**Computer speech recognition**." At present, machine learning algorithms are widely used by various applications of speech recognition. **Google assistant**, **Siri**, **Cortana**, and **Alexa** are using speech recognition technology to follow the voice instructions.

3. Traffic prediction:

If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions.

It predicts the traffic conditions such as whether traffic is cleared, slow-moving, or heavily congested with the help of two ways:

- **Real Time location** of the vehicle form Google Map app and sensors
- **Average time has taken** on past days at the same time.

Everyone who is using Google Map is helping this app to make it better. It takes information from the user and sends back to its database to improve the performance.

4. Product recommendations:

Machine learning is widely used by various e-commerce and entertainment companies such as **Amazon**, **Netflix**, etc., for product recommendation to the user. Whenever we search for some product on Amazon, then we started getting an advertisement for the same product while internet surfing on the same browser and this is because of machine learning.

Google understands the user interest using various machine learning algorithms and suggests the product as per customer interest.

As similar, when we use Netflix, we find some recommendations for entertainment series, movies, etc., and this is also done with the help of machine learning.

5. Self-driving cars:

One of the most exciting applications of machine learning is self-driving cars. Machine learning plays a significant role in self-driving cars. Tesla,the most popular car manufacturing company is working on self-drivingcar. It is using unsupervised learning method to train the car models to detect people and objects while driving.

6. Email Spam and Malware Filtering:

Whenever we receive a new email, it is filtered automatically as important,

normal, and spam. We always receive an important mail in our inbox with the important symbol and spam emails in our spam box, and the technology behind this is Machine learning. Below are some spam filters used by Gmail:

- Content Filter
- Header filter
- General blacklists filter
- Rules-based filters
- Permission filters

Some machine learning algorithms such as **Multi-Layer Perceptron**, **Decision tree**, and **Naïve Bayes classifier** are used for email spam filtering and malware detection.

7. Virtual Personal Assistant:

We have various virtual personal assistants such as **Google assistant**, **Alexa**, **Cortana**, **Siri**. As the name suggests, they help us in finding the information using our voice instruction. These assistants canhelp us in various ways just by our voice instructions such as Play music, call someone, Open an email, Scheduling an appointment, etc.

These virtual assistants use machine learning algorithms as an important part.

These assistant record our voice instructions, send it over the server on a cloud, and decode it using ML algorithms and act accordingly.

8. Online Fraud Detection:

Machine learning is making our online transaction safe and secure by detecting fraud transaction. Whenever we perform some online transaction, there may be various ways that a fraudulent transaction can take place such as **fake accounts**, **fake ids**, and **steal money** in the middle of a transaction. So to detect this, **Feed Forward Neural network** helps us by checking whether it is a genuine transaction or a fraud transaction.

For each genuine transaction, the output is converted into some hash values, and these values become the input for the next round. For each genuine transaction, there is a specific pattern which gets change for the fraud transaction hence, it detects it and makes our online transactions moresecure.

9. Stock Market trading:

Machine learning is widely used in stock market trading. In the stockmarket, there is always a risk of up and downs in shares, so for this machinelearning's **long short term memory neural network** is used for the prediction of stock market trends.

10. Medical Diagnosis:

In medical science, machine learning is used for diseases diagnoses. With

this, medical technology is growing very fast and able to build 3D models that can predict the exact position of lesions in the brain. It helps in finding brain tumors and other brain-related diseases easily.

11. Automatic Language Translation:

Nowadays, if we visit a new place and we are not aware of the language then it is not a problem at all, as for this also machine learning helps us by converting the text into our known languages. Google's GNMT (Google Neural Machine Translation) provide this feature, which is a Neural Machine Learning that translates the text into our familiar language, and it called as automatic translation.

The technology behind the automatic translation is a sequence to sequence learning algorithm, which is used with image recognition and translates the text from one language to another language.

Popular Algorithms used in machine learning:

There are numerous machine learning algorithms available, each with its strengths and weaknesses. The choice of algorithm depends on the nature of the problem, the type and size of the data, and the desired outcome. Here are some popular machine learning algorithms:

1. Linear Regression: A supervised learning algorithm used for regression tasks. It models the relationship between the dependent variable and one or more independent variables by fitting a linear equation to the data.

2. Logistic Regression: A supervised learning algorithm used for classification tasks. It models the relationship between the independent variables and the probability of a binary outcome using the logistic function.

3. Decision Trees: Supervised learning algorithms that build a tree-like model of decisions and their possible consequences. They split the data based on feature values to make predictions.

4. Random Forests: An ensemble learning method that combines multiple decision trees to make predictions. It improves generalization and reduces overfitting compared to individual decision trees.

5. Support Vector Machines (SVM): A supervised learning algorithm used for both classification and regression tasks. SVM finds the best hyperplane that separates data points of different classes or predicts a continuous target variable.

6. Naive Bayes: A probabilistic supervised learning algorithm based on Bayes' theorem. It assumes independence among features and is particularly efficient for text classification and spam filtering tasks.

7. k-Nearest Neighbors (k-NN): A lazy learning algorithm that classifies new instances based on their similarity to existing labeled instances. It assigns the most frequent class label among the k nearest neighbors in the feature space.

		8. Neural Networks: Deep learning algorithms that consist of interconnected layers of artificial neurons. They can learn complex patterns and relationships in data and are widely used for image recognition, natural language processing, and other tasks. **9. Gradient Boosting Methods:** Ensemble learning techniques that combine weak learners, such as decision trees, in a sequential manner to create a strong predictive model. Examples include AdaBoost, Gradient Boosting Machines (GBM), and XGBoost. **10. Clustering Algorithms:** Unsupervised learning algorithms used to identify groups or clusters within data. Examples include k-meansclustering, hierarchical clustering, and DBSCAN. **11. Dimensionality Reduction Algorithms:** Techniques used to reduce the number of features in a dataset while preserving essential information. Principal Component Analysis (PCA) and t-SNE (t-Distributed Stochastic Neighbor Embedding) are commonly used for dimensionality reduction. **12. Reinforcement Learning Algorithms:** Algorithms that learn through interaction with an environment and receive rewards or penalties based on their actions. Reinforcement learning is often used in robotics, game playing, and control systems.		

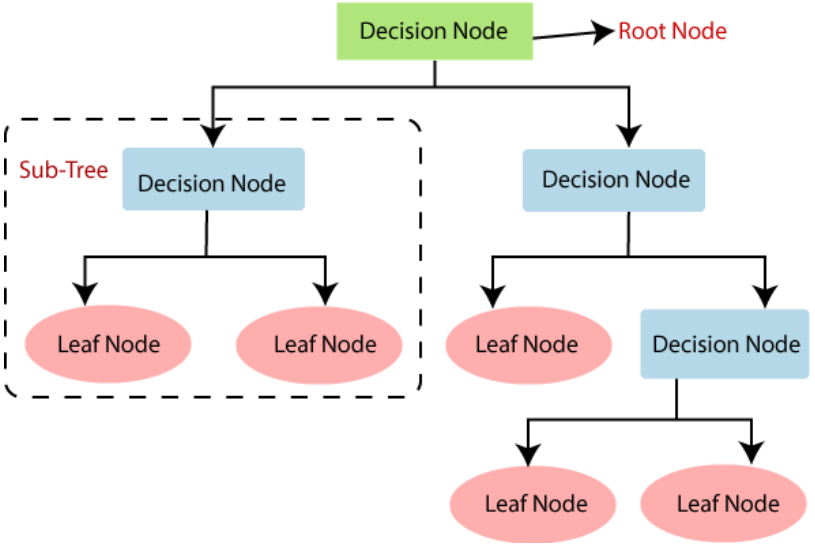
Course Code: 20CS0904 **R20**

CLASSIFICATION AND REGRESSION
UNIT-II

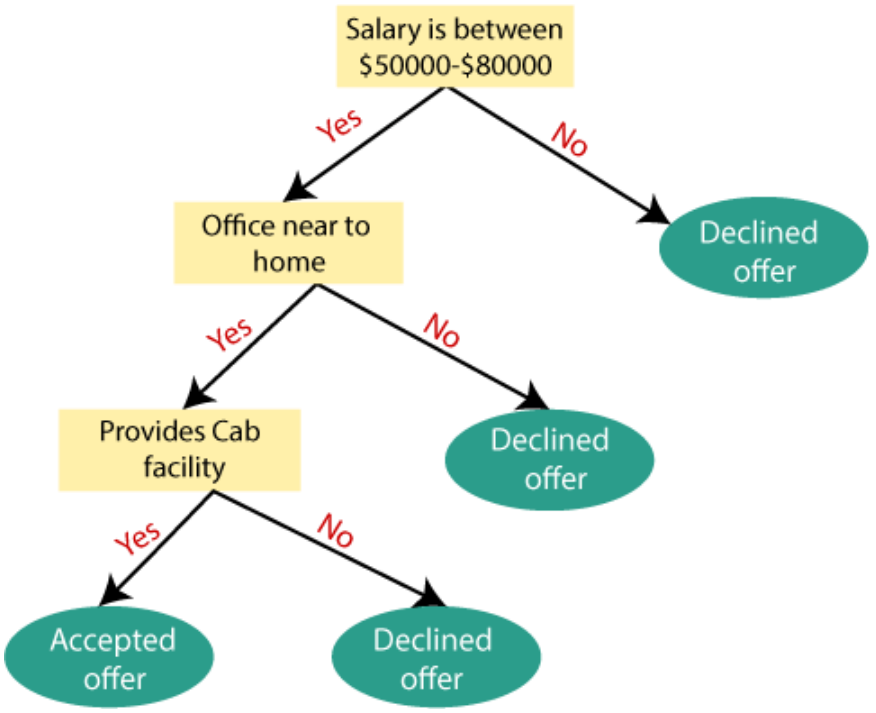
1	Explain about machine learning classification and its usage.	**[L2] [CO1]**	**[12M]**
	Machine learning classification is a supervised learning technique used to classify input data into one of several predefined classes or categories. Classification is a fundamental problem in machine learning and has many applications in areas such as image recognition, speech recognition, natural language processing, fraud detection, and customer segmentation. There are several types of machine learning classification algorithms, including: 1. Logistic Regression - a linear model that predicts the probability of an example belonging to a certain class. 2. Decision Trees - a non-parametric model that learns a hierarchical series of decisions based on the input features to predict the class label. 3. Random Forests - an ensemble learning method that uses multiple decision trees to improve classification accuracy. 4. Support Vector Machines (SVMs) - a powerful and flexible model that separates data points using a hyperplane in a high-dimensional space. 5. Neural Networks - a powerful and flexible model that can learn complex non-linear relationships between input features and class labels. Machine learning classification has many applications, including: 1. Image Recognition - Classifying images based on their contents, such as identifying whether an image contains a cat or a dog. 2. Fraud Detection - Identifying fraudulent transactions based on transaction data and user behaviour patterns. 3. Sentiment Analysis - Analysing text data to determine the sentiment or emotion expressed, such as identifying whether a movie review is positive or negative. 4. Customer Segmentation - Dividing customers into groups based on their behaviour.		
2	Explain Decision Tree Classification technique with an example.	**[L2] [CO1]**	**[12M]**
	o Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. o In a Decision tree, there are two nodes, which are the DecisionNode and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of		

those decisions and do not contain any further branches.

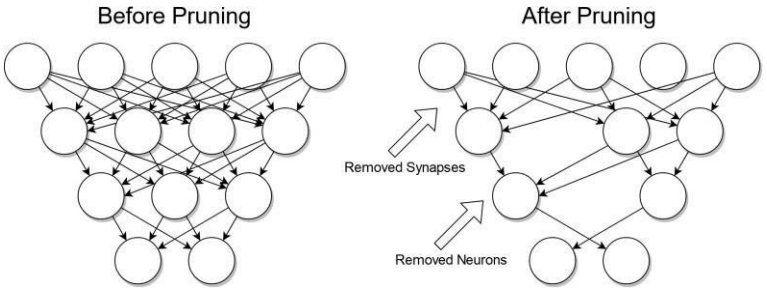
- The decisions or the test are performed on the basis of features of the given dataset.
- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
- It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
- In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.

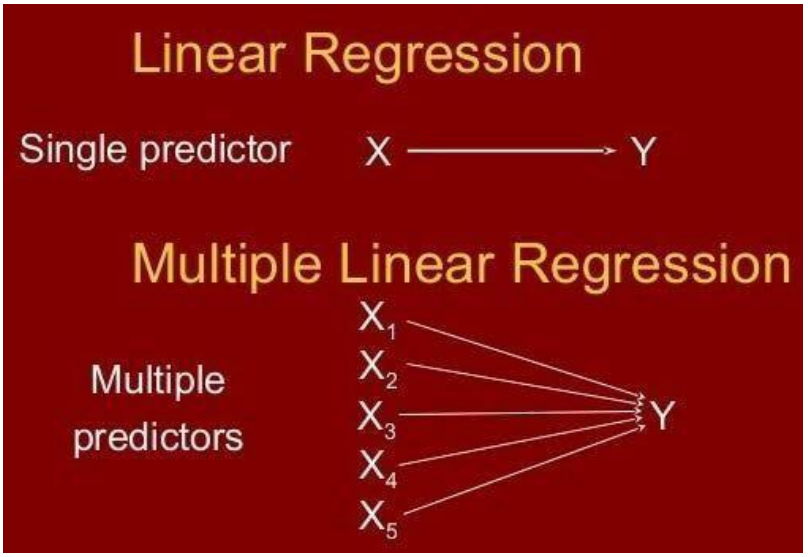


- **Step-1:** Begin the tree with the root node, says S, which contains the complete dataset.
- **Step-2:** Find the best attribute in the dataset using **Attribute Selection Measure (ASM).**
- **Step-3:** Divide the S into subsets that contains possible values for the best attributes.
- **Step-4:** Generate the decision tree node, which contains the best attribute.
- **Step-5:** Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

				
3	a)	Describe about Multivariate Tree prediction.	**[L1] [CO1]**	**[6M]**
		Multivariate tree prediction is a type of machine learning algorithm that is used for regression problems with multiple input variables. The goal of multivariate tree prediction is to predict a continuous output variable based on two or more input variables. In other words, it is a type of regression analysis that focuses on the relationship between multiple independent variables and one dependent variable. The algorithm works by constructing a decision tree based on the input data. The decision tree is a hierarchical structure that consists of a root node, internal nodes, and leaf nodes. The root node represents the entire dataset, and the internal nodes represent decisions based on the input variables. The leaf nodes represent the predicted output values. To construct the decision tree, the algorithm recursively partitions the input data based on the values of the input variables. At each node, the algorithm selects the split that minimizes the variance of the output values within each split. This process continues until a stopping criterion is met, such as reaching a maximum depth or minimum number of data points in each leaf node.		
	b)	Describe about Univariate Tree prediction.	**[L1] [CO1]**	**[6M]**
		Univariate tree prediction is a type of machine learning algorithm that is used for regression problems. The goal of Univariate tree prediction is to predict a continuous output variable based on a single input variable. In other words, it is a type of regression analysis that focuses on the relationship between one independent variable and one dependent variable. The algorithm works by constructing a decision tree based on the input data. The decision tree is a hierarchical structure that consists of a root node, internal nodes, and leaf nodes. The root node represents the entire dataset, and the internal nodes represent decisions based on the input variable. The leaf nodes represent the predicted output values.		

		To construct the decision tree, the algorithm recursively partitions the input data based on the value of the input variable. At each node, the algorithm selects the split that minimizes the variance of the output values within each split. This process continues until a stopping criterion is met, such as reaching a maximum depth or minimum number of data points in each leaf node. Once the decision tree is constructed, it can be used to make predictions for new input data. To make a prediction, the algorithm follows the decision path in the tree based on the value of the input variable and returns the predicted output value at the corresponding leaf node.		
4		Explain the role of Pruning in machine learning.	**[L1][CO1]**	**[12M]**
		Pruning is a technique used in decision trees to reduce overfitting and improve the generalization performance of the model. Overfitting occurs when a decision tree becomes too complex and captures noise or irrelevant details from the training data, which can lead to poor performance on unseen data. Pruning involves the process of removing branches or nodes from a decision tree to simplify its structure and make it more general. This is typically done by setting certain conditions or criteria that determine when and how to prune the tree. There are two main types of pruning techniques: **Pre-Pruning (Early Stopping):** Pre-pruning involves stopping the growth of the tree before it becomes fully expanded. This is usually done by setting stopping criteria based on various measures such as maximum tree depth, minimum number of samples required at a node, minimum improvement in impurity measures (e.g., information gain or Gini index), or other statistical significance tests. If a node does not meet these criteria, it is considered a leaf node and no further splitting is performed. **Post-Pruning (Cost Complexity Pruning):** Post-pruning involves growing the tree to its full size and then selectively removing branches or nodes based on their estimated predictive ability. This is done by assigning a cost or penalty to each node based on measures like impurity or error rate. A complexity parameter, such as the cost complexity parameter or pruning parameter, is used to control the trade-off between simplicity and accuracy. By iteratively removing nodes with the highest cost, the tree is pruned to a more optimal size that balances complexity and performance. The goal of pruning is to find the right balance between complexity and generalization. By reducing the complexity of the decision tree, pruning helps to avoid overfitting and improves the model's ability to generalize well to unseen data. Pruning is an essential step in decision tree construction, especially when dealing with complex datasets or when the decision tree grows too large.		

			
5	Explain in detail about a) Lasso Regression b) Ridge Regression	**[L2][CO1]**	**[12M]**
	Regularization methods for linear regression models are ridge and lasso regression. They help to solve the overfitting issue, which arises when a model is overly complicated and fits the training data too well, leading to worse performance on unknown data. **Ridge Regression** • Ridge regression is one of the most robust versions of linear regression in which a small amount of bias is introduced so that we can get better long term predictions. • The amount of bias added to the model is known as **Ridge Regression penalty**. We can compute this penalty term by multiplying with the lambda to the squared weight of each individual features. • The equation for ridge regression will be: $L(x, y) = \text{Min}(\sum_{i=1}^{n}(y_i - w_i x_i)^2 + \lambda \sum_{i=1}^{n}(w_i)^2)$ • y is the actual value, • wixi denotes the predicted value, • wi denotes the feature coefficient. • A general linear or polynomial regression will fail if there is high collinearity between the independent variables, so to solve such problems, Ridge regression can be used. • Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as **L2 regularization**. • It helps to solve the problems if we have more parameters than samples. **Lasso Regression**		

	○ Lasso regression is another regularization technique to reduce the complexity of the model. ○ It is similar to the Ridge Regression except that penalty term contains only the absolute weights instead of a square of weights. ○ Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0. ○ It is also called as **L1 regularization**. The equation for Lasso regression will be: $L(x, y)= \text{Min}(\sum_{i=1}^{n}(y_i - w_i x_i)^2 + \lambda \sum_{i=1}^{n} \|w_i\|)$ where y is the actual value • wixi denotes the predicted value • wi denotes the feature coefficient. Lasso regression can reduce certain coefficients to zero, conducting feature selection in effect. With high-dimensional datasets where many characteristics could be unnecessary or redundant, this is very helpful.		
6	Explain about Linear Regression and its types.	**[L2][CO3]**	**[12M]**
	 **Linear Regression is generally classified into two types:** 1. Simple Linear Regression 2. Multiple Linear Regression 1. Simple In Simple Linear Regression, we try to find the relationship between a single independent variable (input) and a corresponding dependent variable (output). This can be expressed in the form of a straight line. The same equation of a line can be re-written as:		

$$Y = \beta_0 + \beta_1 X + \epsilon$$

- Y represents the output or dependent variable.
- β0 and β1 are two unknown constants that represent the intercept and coefficient (slope) respectively.
- ε (Epsilon) is the error term.

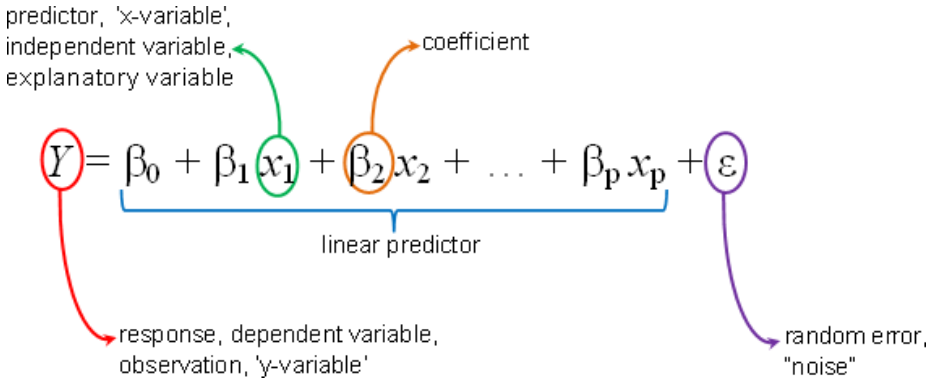
The following is a sample graph of a Simple Linear Regression Model :

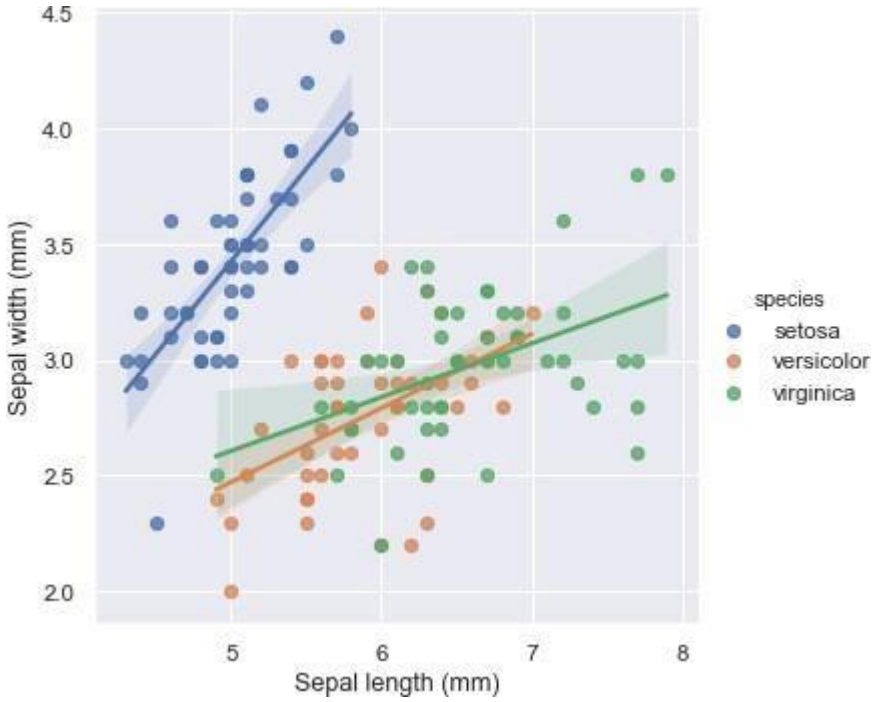
Multiple Linear Regression

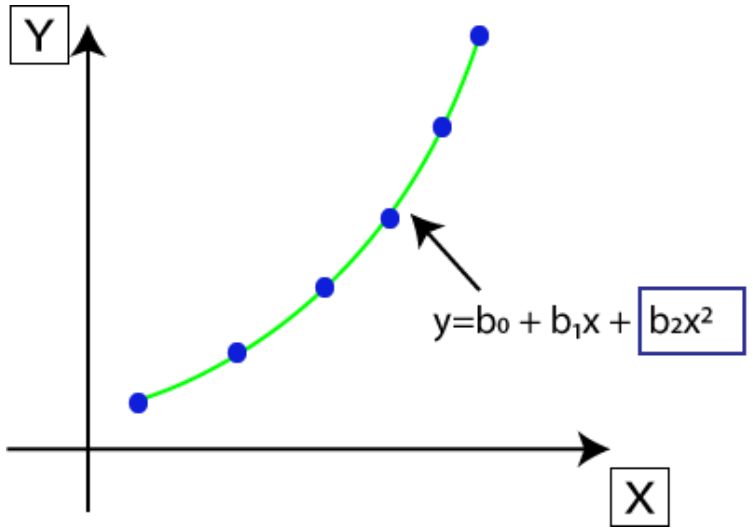
In Multiple Linear Regression, we try to find the relationship between **2 or more independent variables (inputs)** and the corresponding dependent variable (output). The independent variables can be continuous or categorical.

The equation that describes how the predicted values of y is related to **p independent variables** is called as

Multiple Linear Regression equation :



			
7	a) Explain in detail about polynomial regression technique	**[L2] [CO2]**	**[6M]**
	Polynomial Regression is a type of regression which models the **non-linear dataset** using a linear model. • It is similar to multiple linear regression, but it fits a non- linear curve between the value of x and corresponding conditional values of y. • Suppose there is a dataset which consists of datapoints which are present in a non-linear fashion, so for such case, linear regression will not best fit to those data points. To cover such data points, we need Polynomial regression. • I**n Polynomial regression, the original features are transformed into polynomial features of given degree and then modelled using a linear model.** Which means the data points are best fitted using a polynomial line.		



- The equation for polynomial regression also derived from linear regression equation that means Linear regression equation $Y= b_0+ b_1x$, is transformed into Polynomial regression equation $Y= b_0+b_1x+ b_2x^2+ b_3x^3+ + b_nx^n$.
- Here Y is the **predicted/target output, b_0, b_1,... b_n are theregression coefficients**. x is our **independent/input variable**.
- The model is still linear as the coefficients are still linear with quadratic

b) Differentiate between classification and regression. **[L4] [CO2]** **[6M]**

Classification	Regression
Classification gives out discrete values.	Regression gives continuous values.
Given a group of data, this method helps group the data into different groups.	It uses the mapping function to map values to continuous output.
In classification, the nature of the predicted data is unordered.	Regression has ordered predicted data.
The mapping function is used to map values to pre−defined classes.	It attempts to find a best fit line. It tries to extrapolate the graph to find/predict the values.
Example include Decision tree, logistic regression.	Examples include Regression tree (Random forest), Linear regression

		Classification is done by measuring the accuracy.	Regression is done using the root mean square error method.			
8	Describe about Multiple linear regression and MLR equations				**[L1][CO2]**	**[12M]**

Multiple Linear Regression

Multiple Linear Regression is an extension of Simple Linear regression as it takes more than one predictor variable to predict the response variable.

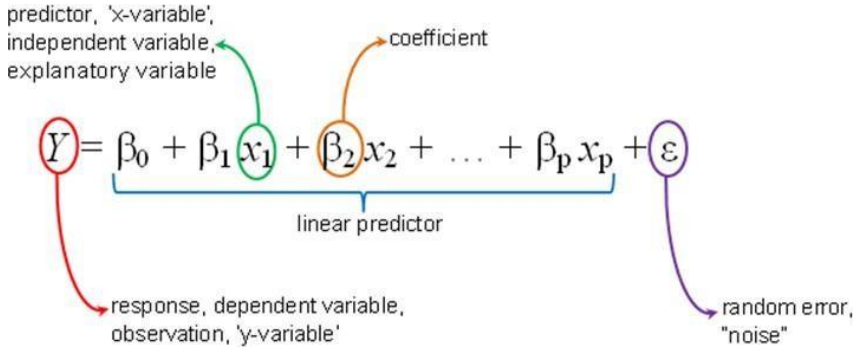
Multiple Linear Regression is one of the important regression algorithms which models the linear relationship between a single dependent continuous variable and more than one independent variable.

In Multiple Linear Regression, we try to find the relationship between **2 or more independent variables (inputs)** and the corresponding dependent variable (output). The independent variables can be continuous or categorical.

- For MLR, the dependent or target variable(Y) must be the continuous/real, but the predictor or independent variable may be of continuous or categorical form.
- Each feature variable must model the linear relationship with the dependent variable.
- MLR tries to fit a regression line through a multidimensional space of data-points.

The equation that describes how the predicted values of y is related to **p independent variables** is called as

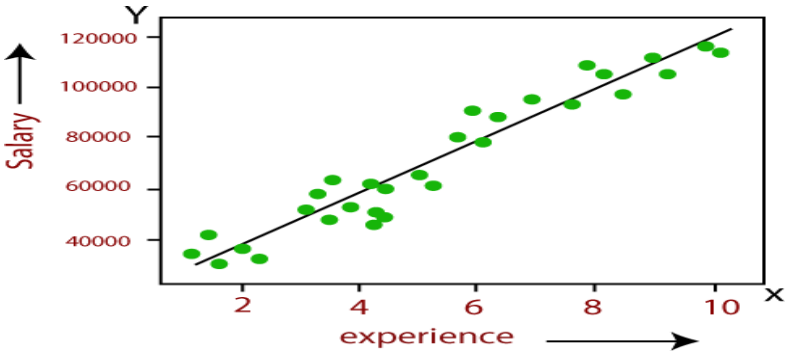
Multiple Linear Regression equation :



Assumptions for Multiple Linear Regression:

- A **linear relationship** should exist between the Target and predictor variables.
- The regression residuals must be **normally distributed**.
- MLR assumes little or **no multicollinearity** (correlation between the independent variable) in data.

Implementation of Multiple Linear Regression

	1. Data Pre-processing Steps 2. Fitting the MLR model to the training set 3. Predicting the result of the test set		
9	Explain in details of types of Regression model in ML.	**[L2] [CO2]**	**[12M]**
	Regression Analysis is a statistical process for estimating the relationships between the dependent variables or criterion variables and one or more independent variables or predictors. Regression analysis is generally used when we deal with a dataset that has the target variable in the form of continuous data. Regression analysis explains the changes in criteria in relation to changes in select predictors. **Explain about linear, logistic, polynomial , lasso, ridge regression** Linear Regression: ○ Linear Regression is one of the most simple Machine learning algorithm that comes under Supervised Learning technique and used for solving regression problems. ○ It is used for predicting the continuous dependent variable with the help of independent variables. ○ The goal of the Linear regression is to find the best fit line that can accurately predict the output for the continuous dependent variable. ○ If single independent variable is used for prediction then it is called Simple Linear Regression and if there are more than two independent variables then such regression is called as Multiple Linear Regression. ○ By finding the best fit line, algorithm establish the relationship between dependent variable and independent variable. And the relationship should be of linear nature. ○ The output for Linear regression should only be the continuous values such as price, age, salary, etc. The relationship between the dependent variable and independent variable can be shown in below image:  In above image the dependent variable is on Y-axis (salary) and independent		

variable is on x-axis(experience). The regression line can be written as:

$y= a_0+a_1x+ \varepsilon$

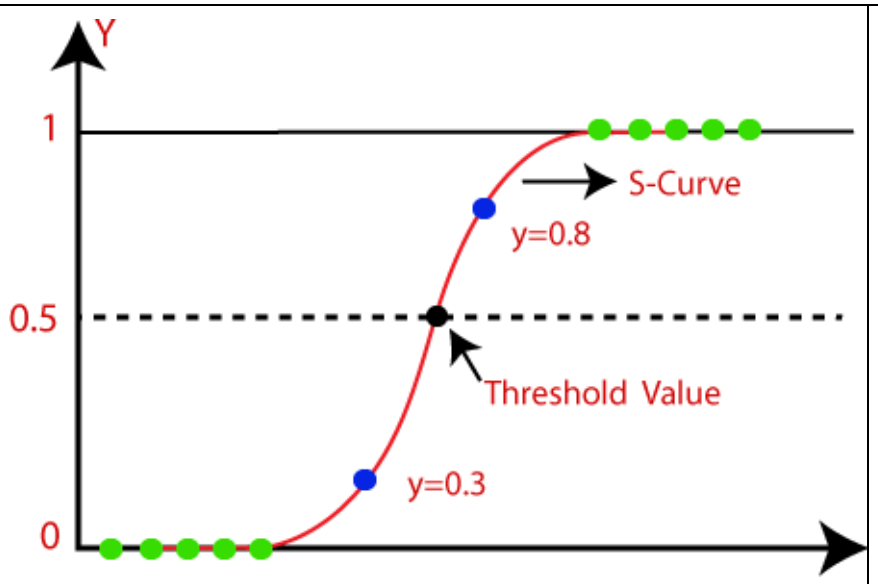
Where, a_0 and a_1 are the coefficients and ε is the error term.

Logistic Regression:

- Logistic regression is one of the most popular Machine learning algorithm that comes under Supervised Learning techniques.
- It can be used for Classification as well as for Regression problems, but mainly used for Classification problems.
- Logistic regression is used to predict the categorical dependent variable with the help of independent variables.
- The output of Logistic Regression problem can be only between the 0 and 1.
- Logistic regression can be used where the probabilities between two classes is required. Such as whether it will rain today or not, either 0 or 1, true or false etc.
- Logistic regression is based on the concept of Maximum Likelihood estimation. According to this estimation, the observed data should be most probable.
- In logistic regression, we pass the weighted sum of inputs through an activation function that can map values in between 0 and 1. Such activation function is known as **sigmoid function** and the curve obtained is called as sigmoid curve or S-curve. This sigmoid function is used to model the data in logistic regression. The function can be represented as:

$$f(x)= \frac{1}{1+e^{-x}}$$

- f(x)= Output between the 0 and 1 value.
- x= input to the function
- e= base of natural logarithm.



- The equation for logistic regression is:

$$\log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

here are three types of logistic regression:

- **Binary(0/1, pass/fail)**
- **Multi(cats, dogs, lions)**
- **Ordinal(low, medium, high)**

10	Explain about real world Applications of regression in machine learning.	**[L2] [CO1]**	**[12M]**

1. **Predictive Modeling and Forecasting:**
   - Sales forecasting
   - Demand forecasting
   - Stock price prediction
   - Weather forecasting
2. **Marketing and Customer Analysis:**
   - Market response modeling
   - Customer lifetime value prediction
   - Market share analysis
3. **Financial Analysis:**
   - Credit scoring and risk assessment
   - Portfolio management
   - Financial market analysis
   - Fraud detection

4. Healthcare and Medical Research:

- Disease prediction and diagnosis
- Patient outcome prediction
- Drug effectiveness analysis
- Health risk assessment

5. Quality Control and Process Optimization:

- Manufacturing process optimization
- Product quality control
- Supply chain optimization
- Anomaly detection

6. Social Sciences and Behavioral Analysis:

- Social and economic impact analysis
- Opinion mining and sentiment analysis
- Education research
- Demographic analysis

7. Energy and Utilities:

- Energy consumption prediction
- Load forecasting
- Energy price modeling
- Renewable energy optimization

8. Sports Analytics:

- Player performance analysis
- Outcome prediction
- Team composition optimization
- In-game decision-making

9. Environmental Analysis:

- Climate modeling
- Pollution prediction
- Environmental impact assessment
- Natural resource management

10. Time Series Analysis:

- Trend analysis
- Seasonal pattern identification
- Economic indicator forecasting
- Stock market analysis

Course Code: 20CS0904 **R20**

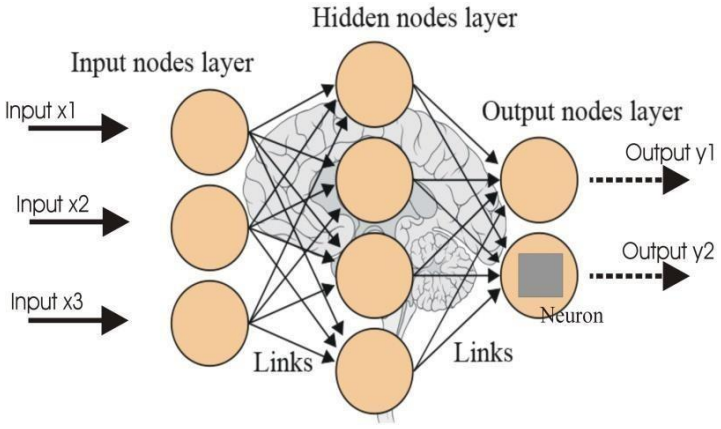
UNIT –III
Learning Models and Decision Theory

1	A	Describe Artificial Neural Networks	[L1][CO3]	**[4M]**

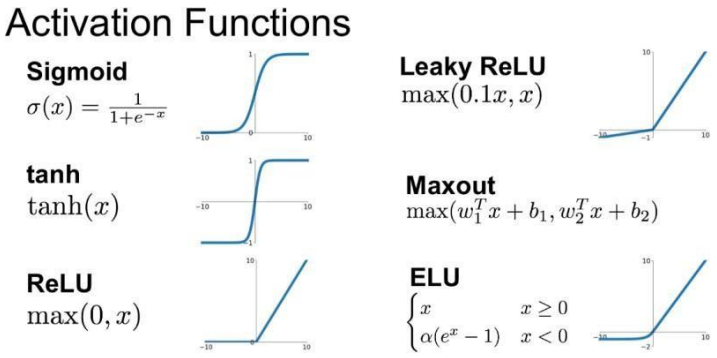
An Artificial Neural Network (ANN) is a computational model inspired by the human brain's neural structure. It consists of interconnected nodes (neurons) organized into layers. Information flows through these nodes, and the network adjusts the connection strengths (weights) during training to learn from data, enabling it to recognize patterns, make predictions, and solve various tasks in machine learning and artificial intelligence.

1. There are three layers in the network architecture: the input layer, the hidden layer (more than one), and the output layer. Because of the numerous layers are sometimes referred to as the MLP (Multi-Layer

   Perceptron).



2. It is possible to think of the hidden layer as a ―distillation layer,‖ which extracts some of the most relevant patterns from the inputs and sends them on to the next layer for further analysis. It accelerates and improves the efficiency of the network by recognizing just the most important informationfrom the inputs and discarding the redundant information.
3. The activation function is important for two reasons: first, it allows you to turn on your computer. This model captures the presence of non-linear relationships between the inputs.It contributes to the conversion of the input into a more usable output.



4. Finding the ―optimal values of W ― weights‖ that minimize prediction error is critical to building a successful model. The ―backpropagation algorithm‖ does this by converting ANN into

a learning algorithm by learning from mistakes.

5. The optimization approach uses a ―gradient descent‖ technique to quantify prediction errors. To find the optimum value for W, small adjustments in W are tried, and the impact on prediction errors is examined. Finally, those W values are chosen as ideal since further W changes do not reduce mistakes.

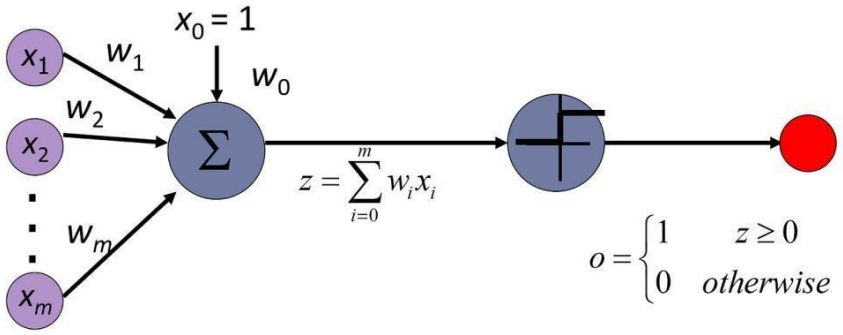
	B	Sketch the types of architectures of neural networks	[L2][CO3]	**[8M]**

Three important types of neural networks:

1. Artificial Neural Networks (ANN)
2. Convolution Neural Networks (CNN)
3. Recurrent Neural Networks (RNN)

Perceptron

The perceptron is a fundamental type of neural network used for binary classification tasks. It consists of a single layer of artificial neurons (also known as perceptrons) that take input values, apply weights, and generate an output. The perceptron is typically used for linearly separable data, where it learns to classify inputs into two categories based on a decision boundary.

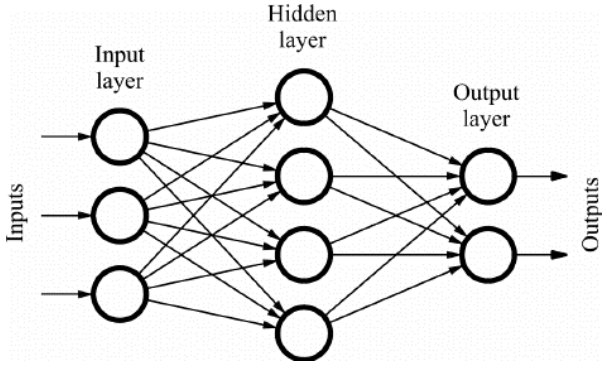


1. Artificial Neural Networks (ANN)

Feed Forward Network

The Feed Forward (FF) networks consist of multiple neurons and hidden layers which are connected to each other. These are called ―feed-forward‖ because the data flow in the forward direction only, and there is no backward propagation. Hidden layers might not be necessarily present in the network depending upon the application.

More the number of layers more can be the customization of the weights. And hence, more will be the ability of the network to learn. Weights are notupdated as there is no back propagation. The output of multiplication ofweights with the inputs is fed to the activation function which acts as athreshold value.



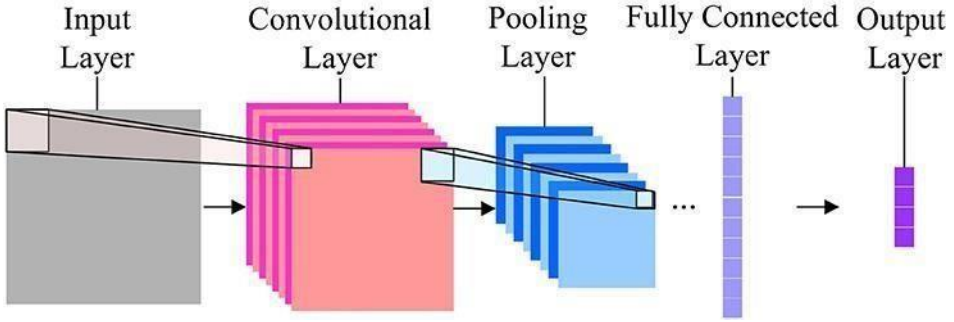
2.Convolutional Neural Networks

When it comes to image classification, the most used neural networks are Convolution Neural Networks (CNN). CNN contain multiple convolution layers which are responsible for the extraction of important features from the image. The earlier layers are responsible for low-level details and the later layers are responsible for more high-level features.

The Convolution operation uses a custom matrix, also called as filters, to convolute over the input image and produce maps. These filters are initialized randomly and then are updated via back propagation. One example of such a filter is the Canny Edge Detector, which is used to find the edges in any image.

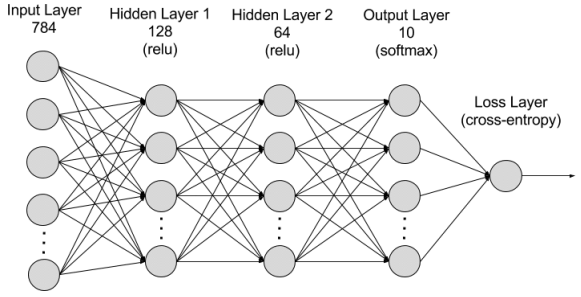
After the convolution layer, there is a pooling layer which is responsible for the aggregation of the maps produced from the convolutional layer. It can be Max Pooling, Min Pooling, etc. For regularization, CNNs also include an option for adding dropout layers which drop or make certain neurons inactive to reduce overfitting and quicker convergence.

CNNs use ReLU (Rectified Linear Unit) as activation functions in the hidden layers. As the last layer, the CNNs have a fully connected dense layer and the activation function mostly as Softmax for classification, and mostly ReLU for regression.



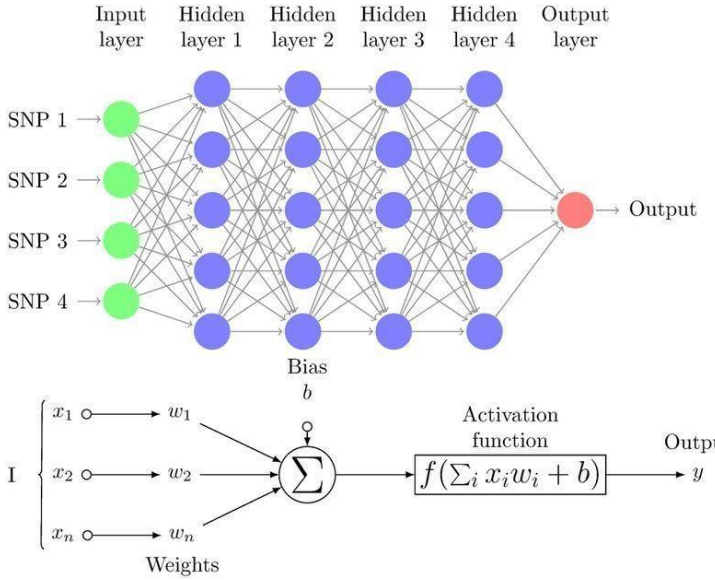
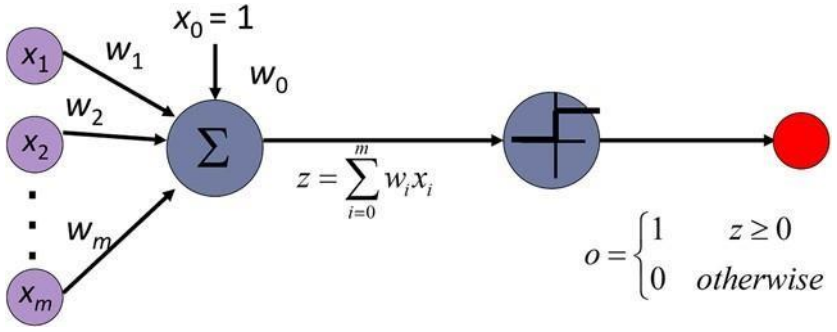
b) Recurrent Neural Network (RNN)?

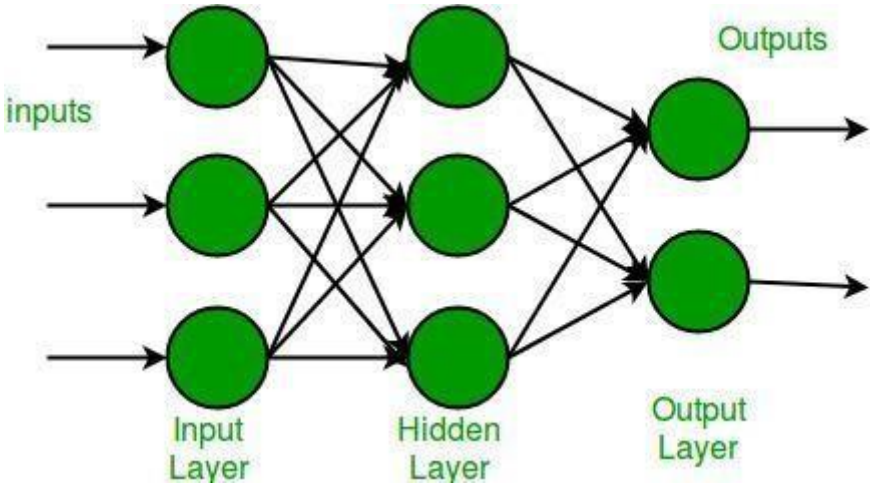
Recurrent Neural Network(RNN) is a type of Neural Network where the output from the previous step is fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other. Still, in cases when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is its Hidden state, which remembers some information about a sequence. The state is also referred to as Memory State since it remembers the previous input to the network. It uses the same parameters for each input as it performs



2	What is multilayer perceptron? Explain in detail.	[L2][CO4]	**[12M]**

Cou	se Co	**e:** 20CS0904		
		A multilayer perceptron (MLP) Neural network belongs to the feedforward neural network. It is an Artificial Neural Network in which all nodes are interconnected with nodes of different layers. The word Perceptron was first defined by Frank Rosenblatt in his perceptron program. Perceptron is a basic unit of an artificial neural network that defines the artificial neuron in the neural network. It is a supervised learning algorithm that contains nodes' values, activation functions, inputs, and node weights to calculate the output. The Multilayer Perceptron (MLP) Neural Network works only in the forward direction. All nodes are fully connected to the network. Each node passes its value to the coming node only in the forward direction. The MLP neural network uses a Back propagation algorithm to increase the accuracy of the training model. **Working of Multilayer Perceptron Neural Network** • The input node represents the feature of the dataset. • Each input node passes the vector input value to the hidden layer. • In the hidden layer, each edge has some weight multiplied by the input variable. All the production values from the hidden nodes are summed together. To generate the output • The activation function is used in the hidden layer to identify the active nodes. • The output is passed to the output layer.		

		• Calculate the difference between predicted and actual output at the output layer. • The model uses back propagation after calculating the predicted output. 		
3	A	Explain single layer perceptron in detail	[L2][CO3]	**[6M]**
		Perceptron The perceptron is a fundamental type of neural network used for binary classification tasks. It consists of a single layer of artificial neurons (also known as perceptrons) that take input values, apply weights, and generate an output. The perceptron is typically used for linearly separable data, where it learns to classify inputs into two categories based on a decision boundary. 		
	B	**Explain multi-layer perceptron in detail** **Multi-layer Perceptron** Multi-layer perception is also known as MLP. It is fully connected dense layers, which transform any input dimension to the desired dimension. A multi-layer perception is a neural network that has multiple layers. To create a neural network we combine neurons together so that the outputs of some neurons are inputs of other neurons. A gentle introduction to neural networks and TensorFlow can be found here: Neural Networks	[L2][CO3]	**[6M]**

		Introduction to TensorFlow A multi-layer perceptron has one input layer and for each input, there is one neuron(or node), it has one output layer with a single node for each output and it can have any number of hidden layers and each hidden layer can have any number of nodes. A schematic diagram of a Multi-Layer Perceptron (MLP) is depicted below.  In the multi-layer perceptron diagram above, we can see that there are three inputs and thus three input nodes and the hidden layer has three nodes. The output layer gives two outputs, therefore there are two output nodes. The nodes in the input layer take input and forward it for further process, in the diagram above the nodes in the input layer forwards their output to each of the three nodes in the hidden layer, and in the same way, the hidden layer processes the information and passes it to the output layer. Every node in the multi-layer perception uses a sigmoid activation function. The sigmoid activation function takes real values as input and converts them to numbers between 0 and 1 using the sigmoid formula. $\alpha(x) = 1/(1 + \exp(-x))$ Now that we are done with the theory part of multi-layer perception, let's go ahead and implement some code in python using the TensorFlow library.		
		Refer 2 question		
4		Describe a) Feed Forward Neural Networks b) Recurrent Neural Networks c) Convolutional Neural Networks a) **A feedforward neural network** It is one of the simplest types of artificial neural networks devised. In this network, the information moves in only one direction—forward—from the input nodes, through the hidden nodes (if any), and to the output nodes. There are no cycles or loops in the network. Feedforward neural networks were the first type of artificial neural network invented and are simpler than their	[L1][CO3]	**[12M]**

counterparts like recurrent neural networks and convolutional neural networks.

Architecture of Feedforward Neural Networks

The architecture of a feedforward neural network consists of three types of layers: the input layer, hidden layers, and the output layer. Each layer is made up of units known as neurons, and the layers are interconnected by weights.

Input Layer: This layer consists of neurons that receive inputs and pass them on to the next layer. The number of neurons in the input layer is determined by the dimensions of the input data.

Hidden Layers:

These layers are not exposed to the input or output and can be considered as the computational engine of the neural network. Each hidden layer's neurons take the weighted sum of the outputs from the previous layer, apply an activation function, and pass the result to the next layer. The network can have zero or more hidden layers.

Output Layer: The final layer that produces the output for the given inputs. The number of neurons in the output layer depends on the number of possible outputs the network is designed to produce.

Each neuron in one layer is connected to every neuron in the next layer, making this a fully connected network. The strength of the connection between neurons is represented by weights, and learning in a neural network involves updating these weights based on the error of the output.

Feedforward Neural Networks Work

The working of a feedforward neural network involves two phases: the feedforward phase and the backpropagation phase.

Feedforward Phase:

In this phase, the input data is fed into the network, and it propagates forward through the network. At each hidden layer, the weighted sum of the inputs is calculated and passed through an activation function, which introduces non-linearity into the model. This process continues until the output layer is reached, and a prediction is made.

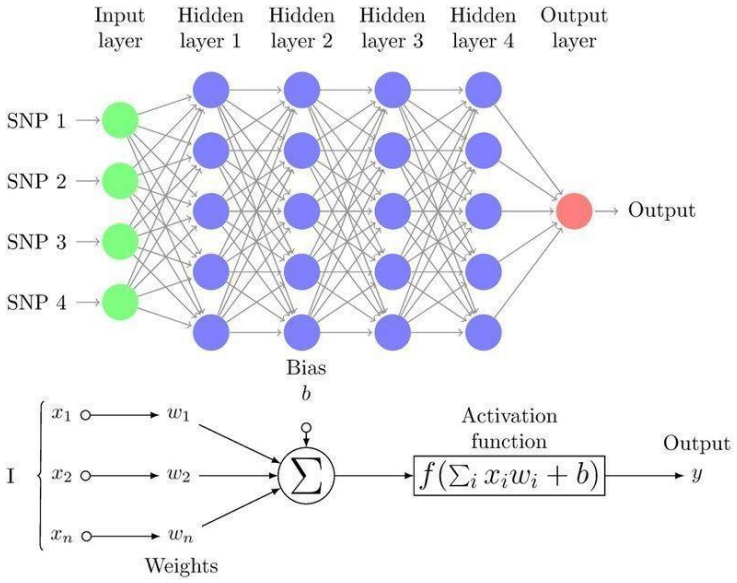
Backpropagation Phase:

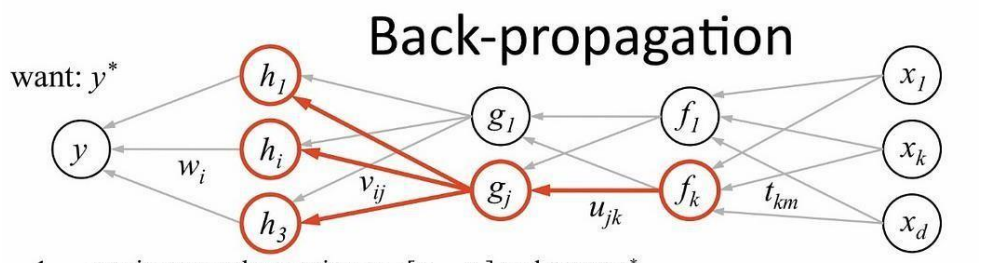
Once a prediction is made, the error (difference between the predicted output and the actual output) is calculated. This error is then propagated back through the network, and the weights are adjusted to minimize this error. The process of adjusting weights is typically done using a gradient descent optimization algorithm.

b) Recurrent Neural Network (RNN)?

Recurrent Neural Network(RNN) is a type of Neural Network where the output from the previous step is fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other. Still, in cases when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is its Hidden state, which remembers some information about a sequence. The state is also referred to as Memory State since it remembers the previous input to the network. It uses the same parameters for each input as it performs

		the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks. C) **Convolutional Neural Networks** When it comes to image classification, the most used neural networks are Convolution Neural Networks (CNN). CNN contain multiple convolution layers which are responsible for the extraction of important features from the image. The earlier layers are responsible for low-level details and the later layers are responsible for more high-level features. The Convolution operation uses a custom matrix, also called as filters, to convolute over the input image and produce maps. These filters are initialized randomly and then are updated via back propagation. One example of such a filter is the Canny Edge Detector, which is used to find the edges in any image. After the convolution layer, there is a pooling layer which is responsible for the aggregation of the maps produced from the convolutional layer. It can be Max Pooling, Min Pooling, etc. For regularization, CNNs also include an option for adding dropout layers which drop or make certain neurons inactive to reduce overfitting and quicker convergence. CNNs use ReLU (Rectified Linear Unit) as activation functions in the hidden layers. As the last layer, the CNNs have a fully connected dense layer and the activation function mostly as Softmax for classification, and mostly ReLU for regression.		
5	A	State and explain implementation of multilayer perceptron. A multilayer perceptron (MLP) Neural network belongs to the feedforward neural network. It is an Artificial Neural Network in which all nodes are interconnected with nodes of different layers. The word Perceptron was first defined by Frank Rosenblatt in his perceptron program. Perceptron is a basic unit of an artificial neural network that defines the artificial neuron in the neural network. It is a supervised learning algorithm that contains nodes' values, activation functions, inputs, and node weights to calculate the output. The Multilayer Perceptron (MLP) Neural Network works only in the forward direction. All nodes are fully connected to the network. Each node passes its value to the coming node only in the forward direction. The MLP neural network uses a Back propagation algorithm to increase the accuracy of the training model. **Working of Multilayer Perceptron Neural Network** • The input node represents the feature of the dataset. • Each input node passes the vector input value to the hidden layer. • In the hidden layer, each edge has some weight multiplied by the input variable. All the production values from the hidden nodes are summed together. To generate the output • The activation function is used in the hidden layer to identify the active nodes. • The output is passed to the output layer. • Calculate the difference between predicted and actual output at the output layer.	[L1][CO4]	**[8M]**

		The model uses back propagation after calculating the predicted output. 		
	B	What are the advantages of multilayer perceptron?	[L1][CO4]	**[4M]**
		• It can be used to solve complex nonlinear problems. • It handles large amounts of input data well. • Makes quick predictions after training. • The same accuracy ratio can be achieved even with smaller samples.		
6	Explain back propagation algorithm with example?		[L2][CO4]	**[6M]**
	Backpropagation is the essence of neural network training. It is the method of fine-tuning the weights of a neural network based on the error rate obtained in the previous epoch (i.e., iteration). Proper tuning of the weights allows you to reduce error rates and make the model reliable by increasing its generalization. Backpropagation in neural network is a short form for ―backward propagation of errors.‖ It is a standard method of training artificial neural networks. This method helps calculate the gradient of a loss function with respect to all the weights in the network. The main features of Backpropagation are the iterative, recursive and efficient method through which it calculates the updated weight to improve the network until it is not able to perform the task for which it is being trained. Derivatives of the activation function to be known at network design time are required to Backpropagation			



1. receive new observation $\boldsymbol{x} = [x_1 \ldots x_d]$ and target y^*
2. **feed forward:** for each unit g_j in each layer 1…L
   compute g_j based on units f_k from previous layer: $g_j = \sigma\left(u_{j0} + \sum_k u_{jk} f_k\right)$
3. get prediction y and error (y-y^*)
4. **back-propagate error:** for each unit g_j in each layer L…1

(a) compute error on g_j

$$\underbrace{\frac{\partial E}{\partial g_j}} = \sum_i \underbrace{\sigma'(h_i) v_{ij}} \underbrace{\frac{\partial E}{\partial h_i}}$$

should g_j be higher or lower? | how h_i will change as g_j changes | was h_i too high or too low?

(b) for each u_{jk} that affects g_j

(i) compute error on u_{jk}

$$\underbrace{\frac{\partial E}{\partial u_{jk}}} = \underbrace{\frac{\partial E}{\partial g_j}} \underbrace{\sigma'(g_j) f_k}$$

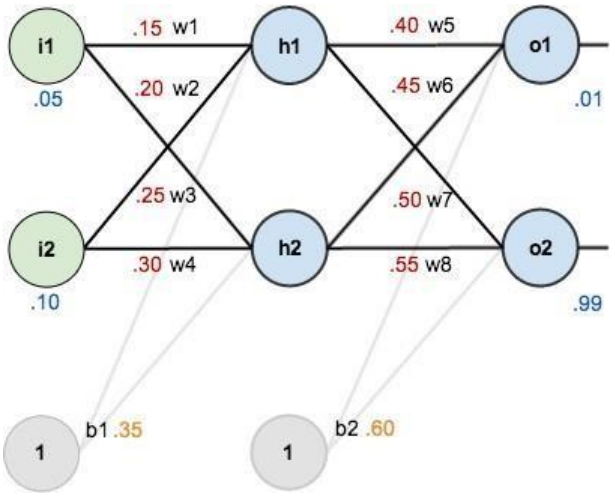
do we want g_j to be higher/lower | how g_j will change if u_{jk} is higher/lower

(ii) update the weight

$$u_{jk} \leftarrow u_{jk} - \eta \frac{\partial E}{\partial u_{jk}}$$

.

Example



The goal of Backpropagation is to optimize the weights so that the neural network can learn how to correctly map arbitrary inputs to outputs.

Given inputs 0.05 and 0.10, we want the neural network to output 0.01 and 0.99.

The Forward Pass

To begin, let's see what the neural network currently predicts given the weights and biases above and inputs of 0.05 and 0.10. To do this we'll feed those inputs forward though the network.

We figure out the *total net input* to each hidden layer neuron, *squash* the total net input using an *activation function* (here we use the *logistic function*), then repeat the process with the output layer neurons.

Here's how we calculate the total net input for h_1:

$$net_{h1} = w_1 * i_1 + w_2 * i_2 + b_1 * 1$$

$$net_{h1} = 0.15 * 0.05 + 0.2 * 0.1 + 0.35 * 1 = 0.3775$$

$$out_{h1} = \frac{1}{1+e^{-net_{h1}}} = \frac{1}{1+e^{-0.3775}} = 0.593269992$$

$$net_{o1} = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1$$

We repeat this process for the output layer neurons, using the output from the hidden layer neurons as inputs.

Here's the output for o_1:

$net_{o1} = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1$

$net_{o1} = 0.4 * 0.593269992 + 0.45 * 0.596884378 + 0.6 * 1 = 1.105905967$

$out_{o1} = \frac{1}{1+e^{-net_{o1}}} = \frac{1}{1+e^{-1.105905967}} = 0.75136507$

And carrying out the same process for o_2 we get:

$out_{o2} = 0.772928465$

We can now calculate the error for each output neuron using the **squared error function** and sum them to get the total error:

$E_{total} = \sum \frac{1}{2}(target - output)^2$

Some sources refer to the target as the *ideal* and the output as the *actual*.

The $\frac{1}{2}$ is included so that exponent is cancelled when we differentiate later on. The result is eventually multiplied by a learning rate anyway so it doesn't matter that we introduce a constant here [1].

For example, the target output for o_1 is 0.01 but the neural network output 0.75136507, therefore its error is:

Repeating this process for o_2 (remembering that the target is 0.99) we get:

The total error for the neural network is the sum of these errors:

$E_{total} = E_{o1} + E_{o2} = 0.274811083 + 0.023560026 = 0.298371109$

The Backwards Pass

Our goal with backpropagation is to update each of the weights in the network so that they cause the actual output to be closer the target output, thereby minimizing the error for each output neuron and the network as a whole.

Output Layer

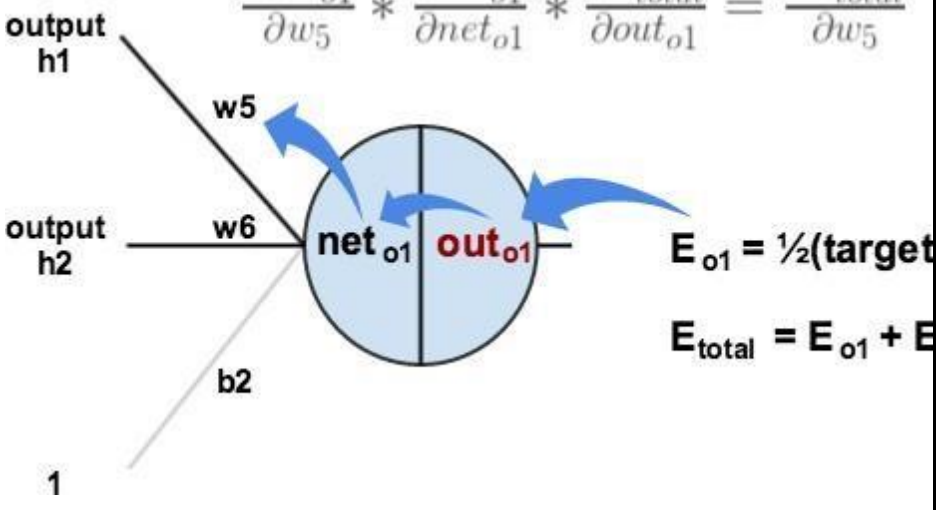
Consider w_5. We want to know how much a change in w_5 affects the total error, aka $\frac{\partial E_{total}}{\partial w_5}$.

$\frac{\partial E_{total}}{\partial w_5}$ is read as —the partial derivative of with respect to w_5—. You can also say —the gradient with respect to w_5—.

By applying the **chain rule** we know that:

Visually, here's what we're doing:

$$\frac{\partial net_{o1}}{\partial w_5} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial E_{total}}{\partial out_{o1}} = \frac{\partial E_{total}}{\partial w_5}$$



We need to figure out each piece in this equation.

First, how much does the total error change with respect to the output?

$E_{total} = \frac{1}{2}(target_{o1} - out_{o1})^2 + \frac{1}{2}(target_{o2} - out_{o2})^2$

$\frac{\partial E_{total}}{\partial out_{o1}} = 2 * \frac{1}{2}(target_{o1} - out_{o1})^{2-1} * -1 + 0$

$\frac{\partial E_{total}}{\partial out_{o1}} = -(target_{o1} - out_{o1}) = -(0.01 - 0.75136507) = 0.74136507$

$-(target - out)$ is sometimes expressed as $out - target$

When we take the partial derivative of the total error with respect to out_{o1}, the quantity $\frac{1}{2}(target_{o2} - out_{o2})^2$ becomes zero because does not affect it which means we're taking the derivative of a constant which is zero.

Next, how much does the output of o_1 change with respect to its total net input?

The partial **derivative of the logistic function** is the output multiplied by 1 minus the output:

$out_{o1} = \frac{1}{1+e^{-net_{o1}}}$

$\frac{\partial out_{o1}}{\partial net_{o1}} = out_{o1}(1 - out_{o1}) = 0.75136507(1 - 0.75136507) = 0.186815602$

Finally, how much does the total net input of $o1$ change with respect to w_5?

$net_{o1} = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1$

$\frac{\partial net_{o1}}{\partial w_5} = 1 * out_{h1} * w_5^{(1-1)} + 0 + 0 = out_{h1} = 0.593269992$

Putting it all together:

You'll often see this calculation combined in the form of the **delta rule**:

$\frac{\partial E_{total}}{\partial w_5} = -(target_{o1} - out_{o1}) * out_{o1}(1 - out_{o1}) * out_{h1}$

Alternatively, we have $\frac{\partial E_{total}}{\partial out_{o1}}$ and $\frac{\partial out_{o1}}{\partial net_{o1}}$ which can be written as $\frac{\partial E_{total}}{\partial net_{o1}}$, aka δ_{o1} (the Greek letter delta) aka the *node delta*. We can use this to rewrite the calculation above:

$\delta_{o1} = -(target_{o1} - out_{o1}) * out_{o1}(1 - out_{o1})$

Therefore:

$\frac{\partial E_{total}}{\partial w_5} = \delta_{o1} out_{h1}$

Some sources extract the negative sign from δ so it would be written as:

$\frac{\partial E_{total}}{\partial w_5} = -\delta_{o1} out_{h1}$

To decrease the error, we then subtract this value from the current weight (optionally multiplied by some learning rate, eta, which we'll set to 0.5):

$w_5^+ = w_5 - \eta * \frac{\partial E_{total}}{\partial w_5} = 0.4 - 0.5 * 0.082167041 = 0.35891648$

Some sources use α (alpha) to represent the learning rate, **others use** η (eta), and **others** even use ϵ (epsilon).

We can repeat this process to get the new weights w_6, w_7, and w_8:

$w_6^+ = 0.408666186$

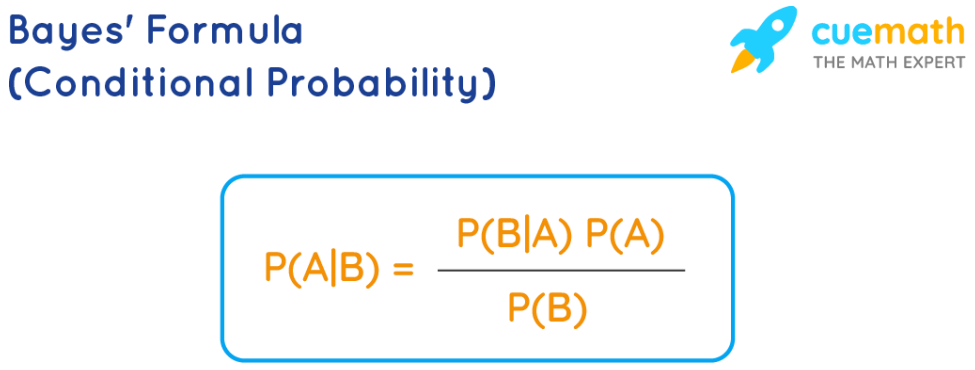
$w_7^+ = 0.511301270$

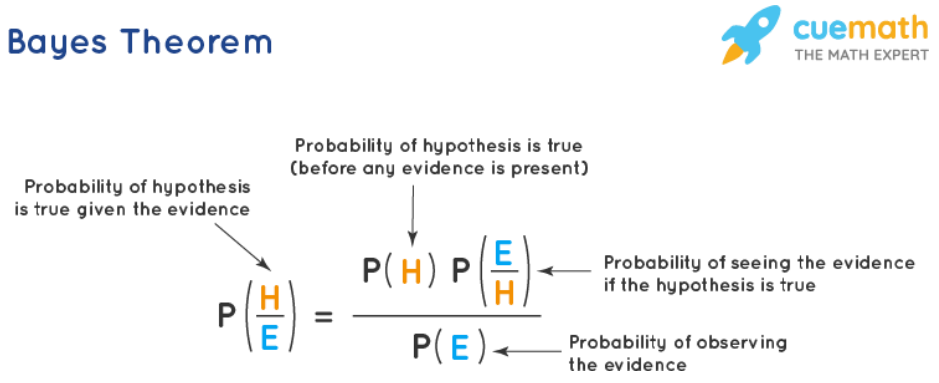
$w_8^+ = 0.561370121$

We perform the actual updates in the neural network *after* we have the new weights leading into the hidden layer neurons .

7	A	Describe Bayesian decision classifier.	[L2][CO4]	**[6M]**

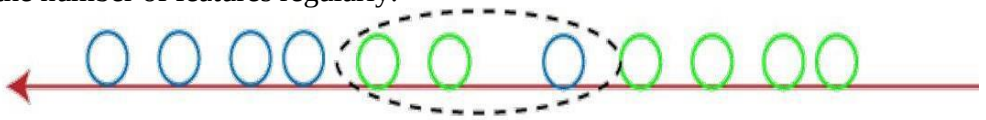
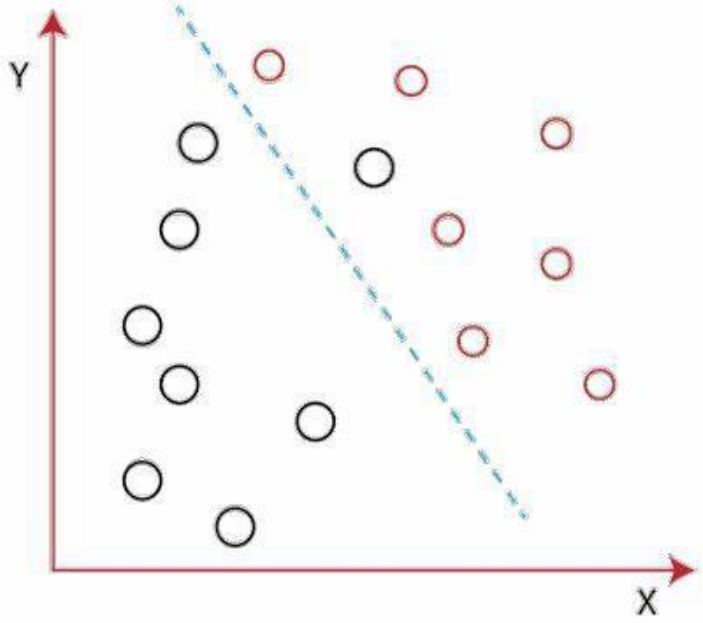
Bayesian Decision Theory is a fundamental statistical approach to the problem of pattern classification. It is considered as the ideal pattern classifier and often used as the benchmark for other algorithms because its decision rule automatically minimizes its loss function.





	B	Explain linear discriminant analysis	[L1][CO4]	**[6M]**

1. Linear Discriminant Analysis (LDA) is a supervised learning algorithm used for classification tasks in machine learning. It is a technique used to find a linear combination of features that best separates the classes in a dataset.
2. LDA works by projecting the data onto a lower-dimensional space that maximizes the separation between the classes. It does this by finding a set of linear discriminants that maximize the ratio of between-class variance to within-class variance. In other words, it finds the directions in the feature space that best separate the different classes of data.
3. LDA assumes that the data has a Gaussian distribution and that

		the covariance matrices of the different classes are equal. It also assumes that the data is linearly separable, meaning that a linear decision boundary can accurately classify the different classes.		
8		Explain linear discriminant analysis with an example? Linear Discriminant analysis is one of the most popular dimensionality reduction techniques used for supervised classification problems in machine learning. It is also considered a pre- processing step for modeling differences in ML and applications of pattern classification. Whenever there is a requirement to separate two or more classes having multiple features efficiently, the Linear Discriminant Analysis model is considered the most common technique to solve such classification problems. For e.g., if we have two classes with multiple features and need to separate them efficiently. When we classify them using a single feature, then it may show overlapping. To overcome the overlapping issue in the classification process, we must increase the number of features regularly.  Overlapping To overcome the overlapping issue in the classification process, we must increase the number of features regularly. Example: Let's assume we have to classify two different classes having two sets of data points in a 2- dimensional plane as shown below image:  However, it is impossible to draw a straight line in a 2-d plane that can separate these data points efficiently but using linear Discriminant analysis; we can dimensionally reduce the 2-D plane into the 1-D plane. Using this technique, we can also maximize the separability between multiple classes.	[L2][CO4]	**[12M]**
9		Distinguish logistic regression and Bayesian logistic regression.	[L4][CO3]	**[12M]**

Logistic regression and Bayesian logistic regression are both methods used for binary classification problems. However, there are several key differences between the two methods:

1. Approach: Logistic regression is a frequentist approach to modelling, while Bayesian logistic regression is a probabilistic approach.
2. Parameter estimation: In logistic regression, the parameters of the model are estimated using maximum likelihood estimation (MLE). In Bayesian logistic regression, the parameters are estimated using Bayesian inference, which takes into account prior knowledge about the parameters.
3. Prior knowledge: Logistic regression does not incorporate prior knowledge about the parameters of the model, while Bayesian logistic regression allows for the incorporation of prior knowledge.
4. Uncertainty estimation: Logistic regression does not provide a measure of uncertainty for the parameter estimates or predictions, while Bayesian logistic regression does.
5. Computational complexity: Bayesian logistic regression is more computationally complex than logistic regression, as it involves the estimation of the posterior distribution of the parameters.

In summary, while both logistic regression and Bayesian logistic regression can be used for binary classification problems, Bayesian logistic regression allows for the incorporation of prior knowledge, provides a measure of uncertainty, and can handle small datasets and large numbers of predictors, at the cost of increased computational complexity. Logistic regression is a simpler method that is commonly used in practice, but it may not be appropriate when prior knowledge is available or when uncertainty estimates are needed.

10	A	State and explain discriminant functions	[L2][CO4]	**[6M]**

Discriminant Analysis refers to a statistical technique that may determine group membership based on a collection of metric predictors that are independent variables. The primary function of this technique is to assign each observation to a particular group or category according to the data's independent characteristics.

Discriminant analysis (DA) is a multivariate technique which is utilized to divide two or more groups of observations (individuals) premised on variables measured on each experimental unit (sample) and to discover the impact of each parameter in dividing the groups.

In addition, the prediction or allocation of newly defined observations to previously specified groups may be examined using a linear or quadratic function for assigning each individual to existing groups. This can be done by determining which group each individual belongs to.

Types

- **Linear discriminant analysis**
- **Quadratic discriminant analysis** .

Linear Discriminant Analysis

Often known as LDA, is a supervised approach that attempts to predict the class of the Dependent Variable by utilizing the linear combination ofthe Independent Variables. It is predicated on the hypothesis that the **independent variables** have a normal distribution (continuous and numerical) and that each class has the same variance and covariance. Both classification and conditionality reduction may be accomplished with the assistance of this method.

Quadratic Discriminant Analysis

It is a subtype of Linear Discriminant Analysis (LDA) that uses quadratic combinations of independent variables to predict the class ofthe **dependent variable**. The assumption of the normal distribution is maintained. Even if it does not presume that the classes have an equal covariance. The QDA produces a quadratic decision boundary.

B | Differentiate between linear and nonlinear discriminant functions | [L1][CO4] | **[6M]**

No	Linear Classification	Non-Linear Classification
1.	Linear Classification refers to categorizing a set of data points into a discrete class based on a linear combination of its explanatory variables.	Non-Linear Classification refers to categorizing those instances that are not linearly separable.
2	It is possible to classify data with a straight line.	It is not easy to classify data with a straight line.
3	Data is classified with the help of a hyperplane.	The utilization of kernels is made to transform non-separable data into separable data.

Course Code: 20CS0904 **R20**

UNIT –IV
BAYESIAN DECISION THEORY AND PARAMETRIC METHODS

1	Explain Bayesian decision theory in detail.	[L2][CO4]	**[12M]**

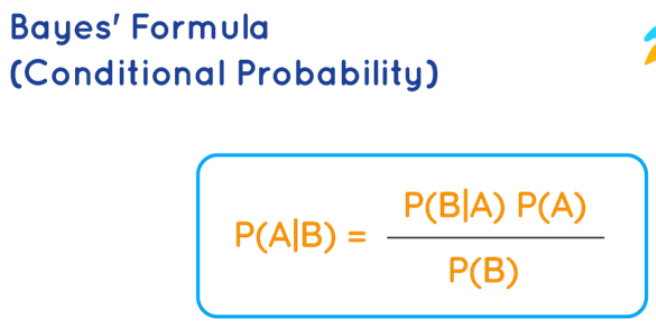
Bayesian decision theory is a statistical framework for decision-making under uncertainty. It is based on the principles of Bayesian statistics, which involves updating prior beliefs with new data to obtain posterior beliefs. In the context of decision theory, Bayesian methods can be used to determine the optimal decision based on the available information.

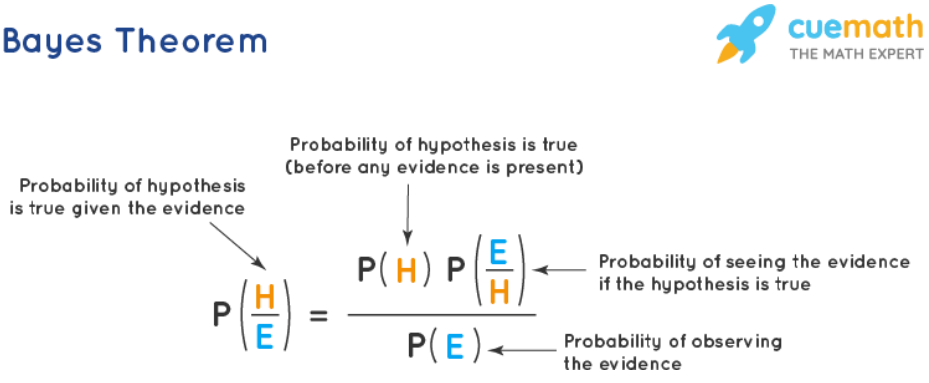
The basic framework of Bayesian decision theory involves three components:

1. A set of possible decisions or actions that can be taken.
2. A set of possible states of the world, which are not directly observable but can be inferred from the available data.
3. A set of consequences or outcomes that result from each decision and state of the world.

The goal of Bayesian decision theory is to choose the decision that maximizes the expected utility, which is the sum of the utility of each outcome weighted by its probability.

To apply Bayesian decision theory, we need to specify a prior distribution over the possible states of the world and update this distribution with new data using Bayes' theorem. The resulting posterior distribution can be used to calculate the expected utility of each decision, which can then be compared to determine the optimal decision.



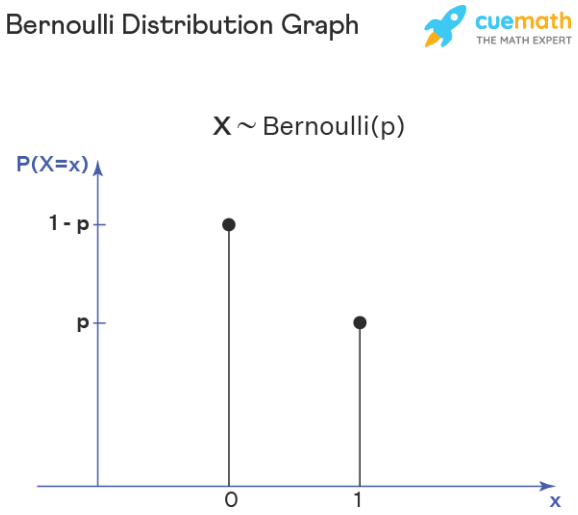


	For example, consider a medical diagnosis problem where a doctor needs to decide whether a patient has a particular disease or not. The doctor can order a diagnostic test, but the test is not perfect and can produce false positive or false negative results. The doctor can also choose to treat the patient or not treat the patient based on the test result. Using Bayesian decision theory, the doctor can specify a prior distribution over the probability of the patient having the disease based on prior knowledge and experience. The doctor can then update this distribution with the test results using Bayes' theorem to obtain the posterior distribution over the probability of the patient having the disease. The doctor can then calculate the expected utility of each decision (e.g., treat the patient if the posterior probability is above a certain threshold) based on the posterior distribution and the utility of each outcome (e.g., curing the patient or causing harm). The decision with the highest expected utility is then chosen as the optimal decision.		
2	Write are the classifications in Bayesian decision theory? State with example?	[L3][CO4]	**[12M]**
	Bayesian decision theory, there are two main classifications of decisions: 1. Deterministic decisions: In deterministic decisions, the decision is based solely on the observed data and the decision rule is fixed. The decision rule does not incorporate uncertainty about the true state of the world. For example, consider a factory that produces electronic components. The quality control department checks each component for defects and either accepts orrejects it based on a fixed threshold. The decision to accept or reject a component is deterministic and does not depend on any uncertainty about the true quality of the component. 2. Stochastic decisions: In stochastic decisions, the decision rule incorporates uncertainty about the true state of the world. The decision is based on the posterior distribution over the possible states of the world, which is updated with new data. For example, consider a financial portfolio manager who needs to decide which stocks to buy or sell. The manager can use Bayesian decision theory to estimate the expected return and risk of each stock based on historical data and other relevant information. The manager can then update the posterior distribution ofthe expected return and risk with new data and use this distribution to make a decision about which stocks to buy or sell. In general, stochastic decisions are more flexible and robust than deterministic decisions, as they can incorporate uncertainty and adapt to changing conditions. However, they can also be more computationally intensive and require more data and prior information to make reliable decisions.		
3	Describe the losses obtained in Bayesian decision theory?	[L1][CO5]	**[12M]**
	In Bayesian decision theory, the loss function is a way to measure the cost or penalty associated with making a particular decision in a particular state of the world. The loss function can take different forms depending on the context of the decision problem. Here are some common types of loss functions used in Bayesian decision theory: 1. 0-1 loss: This is a binary loss function that assigns a loss of 1 if the decision is incorrect and 0 if it is correct. For example, in a medical diagnosis problem, the loss of misdiagnosing a patient with a disease could be set to 1, while the loss of correctly diagnosing a healthy patient could be set to 0.		

	2. Quadratic loss: This loss function penalizes decisions that are far from the true value more than those that are close. The loss function takes the form of a squared difference between the decision and the true value. Quadratic loss is commonly used in regression problems, where the goal is to predict a continuous variable. 3. Absolute loss: This loss function penalizes decisions that are far from the true value, but not as strongly as quadratic loss. The loss function takes the form of an absolute difference between the decision and the true value. Absolute loss is commonly used in regression problems where outliers are present. 4. Logarithmic loss: This loss function is commonly used in binary classification problems where the goal is to predict a probability. The loss function takes the form of the negative logarithm of the predicted probability for the true class. Logarithmic loss is useful for penalizing low confidence predictions. In Bayesian decision theory, the goal is to choose the decision that minimizes the expected loss, which is the sum of the losses weighted by their respective probabilities. The optimal decision is the one that minimizes the expected loss. By incorporating the loss function into the decision-making process, Bayesian decision theory provides a principled and flexible approach to decision-making under uncertainty.		
4	Explain discriminant functions?	[L2][CO4]	**[12M]**
	Discriminant functions are a set of mathematical functions that are used to classify observations into different classes based on their measured features or variables. In pattern recognition and machine learning, discriminant functions areused to identify patterns or relationships in data and to make predictions aboutthe class or category of a new observation. In discriminant analysis, the discriminant function is derived from the features of the training data and is used to classify new observations into one of the pre-defined classes. There are different types of discriminant functions, depending on the nature of the data and the problem at hand. Some common types of discriminant functions include: 1. **Linear discriminant function**: In linear discriminant function, the decision boundary is a linear function of the input features. The goal of linear discriminant analysis is to find a linear combination of the features that maximally separates the classes. Linear discriminant analysis is commonly used in problems with two or more classes. 2. **Quadratic discriminant function**: In quadratic discriminant function, the decision boundary is a quadratic function of the input features. The goal of quadratic discriminant analysis is to find a quadratic function that maximally separates the classes. Quadratic discriminant analysis is commonly used when the classes have different covariance matrices. 3. **Non-parametric discriminant function:** In non-parametric discriminant function, the decision boundary is a non-linear function of the input features. The goal of non-parametric discriminant analysis is to find a non- linear function that maximally separates the classes. Non-parametric discriminant analysis is commonly used when the data distribution is not well-defined or when the classes have complex boundaries.		

	Discriminant functions can be used in a variety of applications, including image recognition, speech recognition, natural language processing, and bioinformatics. They provide a powerful and flexible tool for identifying patterns and making predictions in complex data sets.		
5	Define parametric methods? Explain maximum likelihood estimation.	[L1][CO4]	**[12M]**
	Parametric methods are statistical techniques that assume that the data being analyzed follows a certain probability distribution with a set of parameters. The parameters of the distribution are estimated from the data and used to make inferences and predictions. Maximum likelihood estimation (MLE) is a method of estimating the parameters of a parametric distribution by maximizing the likelihood function. The likelihood function is the probability of the observed data given the parameters of the distribution. The goal of MLE is to find the set of parameters that maximizes the likelihood function. To illustrate the concept of MLE, consider the example of estimating the mean and standard deviation of a normal distribution given a set of data points. The probability density function of a normal distribution is defined by two parameters: the mean (μ) and the standard deviation (σ). The likelihood function of the normal distribution is given by: $L(\mu,\sigma|X) = \prod i=1n\ f(xi|\mu,\sigma)$ where X = {x1, x2, ..., xn} is the set of n observed data points, and $f(xi|\mu,\sigma)$ is the probability density function of the normal distribution. The goal of MLE is to find the values of μ and σ that maximize the likelihood function. This can be done by taking the partial derivatives of the likelihood function with respect to μ and σ, setting them equal to zero, and solving for μ and σ. The resulting estimates of μ and σ are the maximum likelihood estimates. One advantage of MLE is that it is a consistent estimator, which means that as the sample size increases, the estimated parameters converge to the true values. MLE also has good properties in terms of efficiency and asymptotic normality. However, one potential drawback of MLE is that it can be sensitive to outliers and deviations from the assumed distribution. In addition, MLE can be computationally intensive and may require numerical optimization methods to find the maximum likelihood estimates.		
6	State and explain the following a. Bernoulli density b. Multinomial density c. Gaussian density	[L1][CO4]	**[12M]**
	a) The Bernoulli density is a probability distribution that models a binary outcome, where there are only two possible outcomes, typically represented as 0 and 1. The Bernoulli distribution is characterized by a single parameter p, which represents the probability of the outcome being 1. The probability mass function of the Bernoulli distribution is given by: $P(X=x) = p\text{^}x * (1-p)\text{^}(1-x)$ where X is the binary random variable, x can take the values of 0 or 1, and p is the probability of X being 1. For example, if we have a coin that is flipped and we are interested in the		

probability of getting heads, we can model this situation with a Bernoulli distribution, where p represents the probability of getting heads on a single flip. If p=0.5, then the probability of getting heads is equal to the probability of getting tails, and the distribution is symmetric. If p is not equal to 0.5, then the distribution is skewed towards the more likely outcome.



b)

The multinomial density is a probability distribution that models outcomes with more than two possible categories. The multinomial distribution is characterized by a vector of probabilities p1, p2, ..., pk, where pi represents the probability of observing category i. The multinomial distribution is used when we have a fixed number of independent trials, and each trial can result in one of k possible outcomes. The probability mass function of the multinomial distribution is given by:

P(X = (x1, x2, ..., xk)) = n! / (x1! x2! ... xk!) * p1^x1 * p2^x2 * ... * pk^xk

where X is a vector of random variables representing the frequencies of each category, n is the total number of trials, and xi represents the number of times category i was observed.

The multinomial distribution satisfies the properties of a probability distribution, meaning that the sum of the probabilities over all possible outcomes equals 1.

For example, if we have a six-sided die that is rolled 10 times and we are interested in the probability of each outcome, we can model this situation with a multinomial distribution, where p1 = p2 = p3 = p4 = p5 = p6 = 1/6, and n=10. The probability of observing 3 ones, 2 twos, 1 three, 2 fours, 1 five, and 1 six is given by the multinomial distribution.

The expected value and variance of a multinomial distribution are given by:

E(Xi) = n*pi

Var(Xi) = n*pi*(1-pi)

C) The Gaussian density is a continuous probability distribution that is also known as the normal distribution. It is widely used in statistical analysis and machine learning due to its convenient mathematical properties and its ability to model a

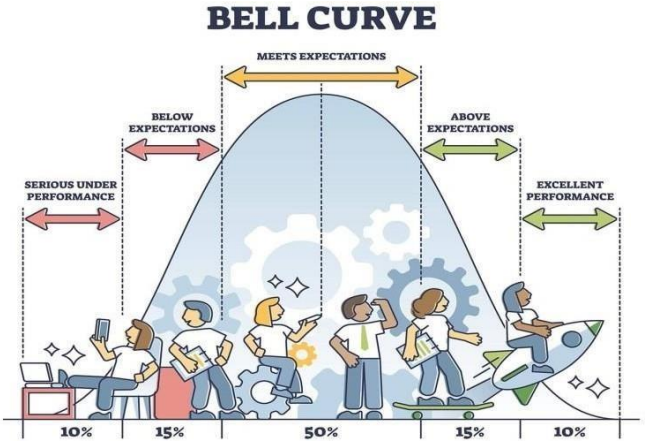
wide variety of real-world phenomena.

The Gaussian distribution is characterized by two parameters: the mean (μ) and the variance (σ^2). The probability density function (PDF) of the Gaussian distribution is given by:

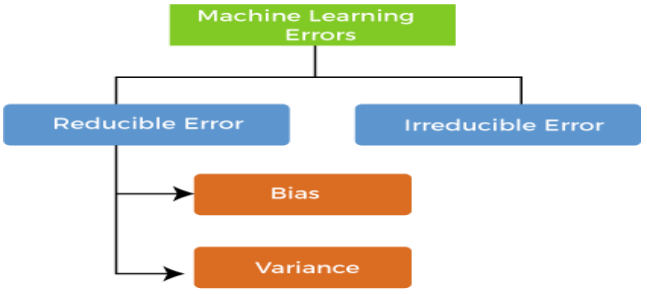
f(x) = (1 / (σ*sqrt(2π))) * exp(-(x-μ)^2 / (2σ^2))

where x is the random variable, μ is the mean, σ is the standard deviation, and π is the mathematical constant pi.

The Gaussian distribution has a bell-shaped curve, with the peak of the curve at the mean μ. The standard deviation σ controls the spread of the curve. Larger values of σ result in a flatter and wider curve, while smaller values of σ result in a taller and narrower curve.



7	a	Write about bias and variance?	[L3][CO4]	**[6M]**



Bias is simply defined as the differences between actual or expected values and the predicted values are known as error or bias error or error due to bias.

$$\text{Bias}(\hat{Y}) = E(\hat{Y}) - Y$$

- **Low Bias:** Low bias value means fewer assumptions are taken to build the target function. In this case, the model will closely match the training dataset.
- **High Bias:** High bias value means more assumptions are taken to build the target function. In this case, the model will not match the training dataset closely.

Variance tells that how much a random variable is different from its expected value.

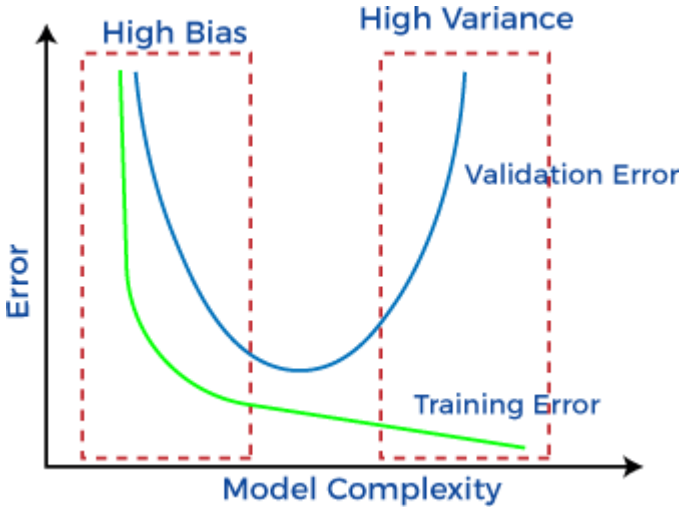
Variance refers to the changes in the model when using different portions of the training data set.

Simply stated, variance is the variability in the model prediction—how much the ML function can adjust depending on the given data set.

- Models with high bias will have low variance.
- Models with high variance will have a low bias.

Characteristics of a high variance model include:

- Noise in the data set
- Potential towards overfitting
- Complex models
- Trying to put all data points as close as possible



b Describe the Bernoulli density? Give an example? [L1][CO3] **[6M]**

Bernoulli distribution is a **discrete probability distribution**, meaning it's concerned with discrete random variables. A discrete random variable is one that has a finite or countable number of possible values—the number of heads you get when tossing three coins at once, or the number of students in a class. So: A discrete probability distribution describes the probability that each possible value of a discrete random variable will occur—for example, the probability of getting a six when rolling a die. When dealing with discrete variables, the probability of each value falls between 0 and 1, and the sum of all the probabilities is equal to 1. So, in the die example, assuming we're using a standard die, the probability of rolling a six is 0.167, or 16.7%. This is based on dividing 1 (the sum of all probabilities) by 6 (the number of possible outcomes).

Bernoulli distribution example: Tossing a coin The coin toss example is perhaps the easiest way to explain Bernoulli distribution. Let's say that the outcome of "heads" is a "success," while an outcome of "tails" is a "failure." In this instance:

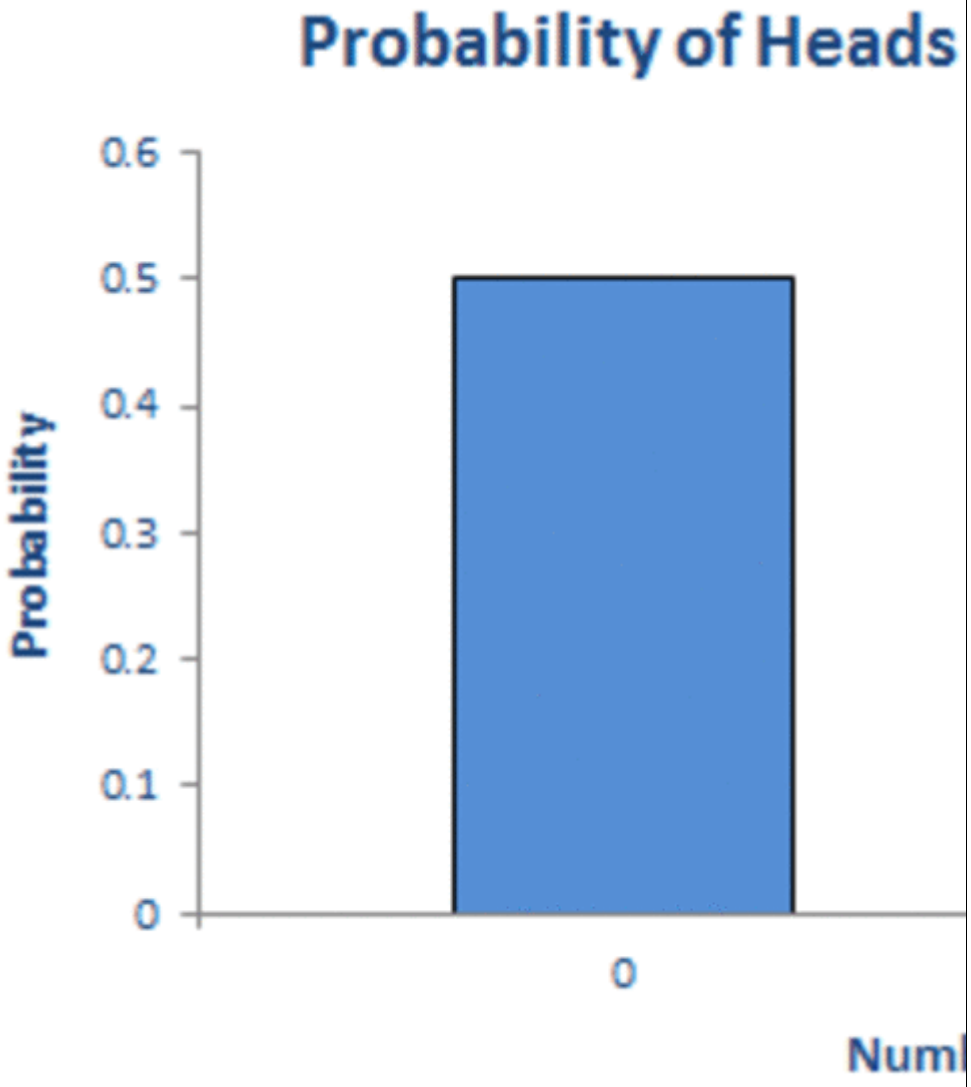
- The probability of a successful outcome (landing on heads) is written as ***p***
- The probability of a failure (landing on tails), written as ***q***, is calculated as ***1 **–** p***

With a standard coin, we know that there's a 50/50 chance of landing on either

So, in this case:

- *p = 0.5*
- *q = 1– **0.5*

So, in our coin toss example, both *p* and *q = 0.5*. On a graph, you'd represent the probability of a failure as "0" and the probability of success as "1," both on the y-axis.



Further examples of Bernoulli distribution

The coin-toss example is a very simple one, but there are actually many scenarios in life that have a yes-no outcome. For example:

- Will you pass or fail a test?
- Will your favorite sports team win or lose their next match?
- Will you be accepted or rejected for that job you applied for?

		Will you roll a six in the opening round of your favorite board game? • Will you win or lose the lottery?		
8		Explain with example of bias and variance?	[L3][CO5]	**[12M]**

Bias is simply defined as the inability of the model because of that there is some difference or error occurring between the model's predicted value and the actual value. These differences between actual or expected values and the predicted values are known as error or bias error or error due to bias. Bias is a systematic error that occurs due to wrong assumptions in the machine learning process.
Let *Y* be the true value of a parameter, and let *Y*^ be an estimator of *Y* based on a sample of data. Then, the bias of the estimator *Y*^ is given by:

Bias(y^)=E(Y^)-Y

where E(Y^) $E(\hat{Y})$ is the expected value of the estimator *Y*^. It is the measurement of the model that how well it fits the data.

- **Low Bias:** Low bias value means fewer assumptions are taken to build the target function. In this case, the model will closely match the training dataset.
- **High Bias:** High bias value means more assumptions are taken to build the target function. In this case, the model will not match the training dataset closely.

The high-bias model will not be able to capture the dataset trend. It is considered as the underfitting model which has a high error rate. It is due to a very simplified algorithm.
For example, a linear regression model may have a high bias if the data has a non-linear relationship.

Ways to reduce high bias in Machine Learning:

- **Use a more complex model:** One of the main reasons for high bias is the very simplified model. it will not be able to capture the complexity of the data. In such cases, we can make our mode more complex by increasing the number of hidden layers in the case of a deep neural network. Or we can use a more complex model like Polynomial regression for non-linear datasets, CNN for image processing, and RNN for sequence learning.
- **Increase the number of features:** By adding more features to train the dataset will increase the complexity of the model. And improve its ability to capture the underlying patterns in the data.
- **Reduce Regularization of the model:** Regularization techniques such as L1 or L2 regularization can help to prevent overfitting and improve the generalization ability of the model. if the model has a high bias, reducing the strength of regularization or removing it altogether can help to improve its performance.
- **Increase the size of the training data:** Increasing the size of the training data can help to reduce bias by providing the model with more examples to learn from the dataset.

What is Variance?

Variance is the measure of spread in data from its mean position. In machine learning variance is the amount by which the performance of a predictive model changes when it is trained on different subsets of the training data. More specifically, variance is the variability of the model that how much it is sensitive to another subset of the training dataset. i.e. how much it can adjust on the new subset of the training dataset.
Let Y be the actual values of the target variable, and *Y*^ be the predicted values of the target variable. Then the variance of a model can be measured as the expected value of the square of the difference between predicted values and the expected value of the predicted values.

Variance=E[*Y*^-E(*Y*^)2]

where $E[\bar{Y}]$ is the expected value of the predicted values. Here expected value is

averaged over all the training data.
Variance errors are either low or high-variance errors.

- **Low variance:** Low variance means that the model is less sensitive to changes in the training data and can produce consistent estimates of the target function with different subsets of data from the same distribution. This is the case of underfitting when the model fails to generalize on both training and test data.
- **High variance:** High variance means that the model is very sensitive to changes in the training data and can result in significant changes in the estimate of the target function when trained on different subsets of data from the same distribution. This is the case of overfitting when the model performs well on the training data but poorly on new, unseen test data. It fits the training data too closely that it fails on the new training dataset.

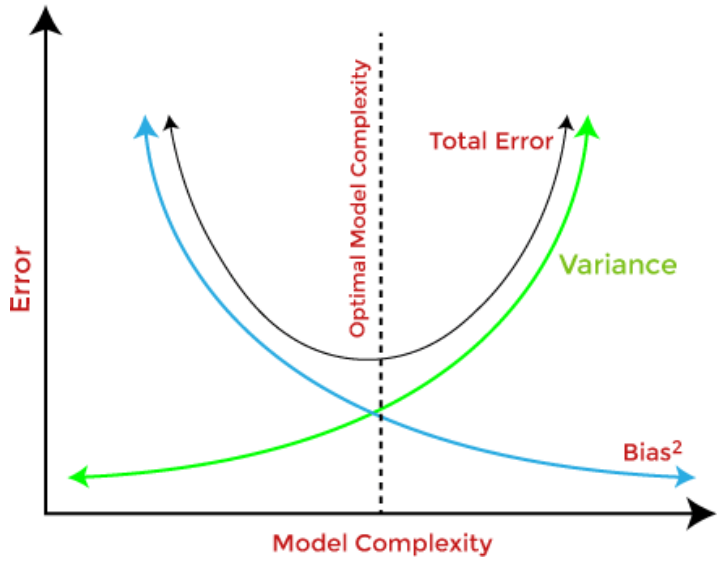
Ways to Reduce the reduce Variance in Machine Learning:

- **Cross-validation:** By splitting the data into training and testing sets multiple times, cross-validation can help identify if a model is overfitting or underfitting and can be used to tune hyperparameters to reduce variance.
- **Feature selection:** By choosing the only relevant feature will decrease the model's complexity. and it can reduce the variance error.
- **Regularization:** We can use L1 or L2 regularization to reduce variance in machine learning models
- **Ensemble methods:** It will combine multiple models to improve generalization performance. Bagging, boosting, and stacking are common ensemble methods that can help reduce variance and improve generalization performance.
- **Simplifying the model:** Reducing the complexity of the model, such as decreasing the number of parameters or layers in a neural network, can also help reduce variance and improve generalization performance.
- **Early stopping:** Early stopping is a technique used to prevent overfitting by stopping the training of the deep learning model when the performance on the validation set stops improving.

9	a	What is bias/variance dilemma? Explain in detail?	[L1][CO3]	**[6M]**

Bias-Variance Trade-Off

While building the machine learning model, it is really important to take care of bias and variance in order to avoid overfitting and underfitting in the model. If the model is very simple with fewer parameters, it may have low variance and high bias. Whereas, if the model has a large number of parameters, it will have high variance and low bias. So, it is required to make a balance between bias and variance errors, and this balance between the bias error and variance error is known as **the Bias-Variance trade-off.**



b What is estimator? explain briefly [L1][CO4] **[6M]**

An estimator is a function of the sample data that provides an estimate of a population parameter. For example, if we are interested in estimating the average height of all students in a university, we can use the mean height of a sample of students as an estimator of the population mean. The estimator itself is a random variable because it is a function of the random sample data. Its value varies from sample to sample due to the randomness inherent in the sampling process.

Properties of Good Estimators

For an estimator to be considered good, it should possess certain desirable properties:

Unbiasedness

An estimator is said to be unbiased if its expected value is equal to the true value of the parameter being estimated. In other words, on average, it hits the target parameter. An estimator that systematically overestimates or underestimates the parameter is considered biased.

Consistency

A consistent estimator is one where the estimates become closer to the true parameter value as the sample size increases. Consistency ensures that by collecting more data, we can improve the accuracy of our estimate.

Efficiency

Efficiency refers to the variance of the estimator. Among all unbiased estimators, an efficient estimator has the smallest variance. An efficient estimator provides more precision and is less spread out around the true parameter value.

Sufficiency

A sufficient estimator captures all the information in the sample that is needed to estimate the parameter. No other statistic calculated from the same sample

Course C

can provide any additional information about the parameter.

Types of Estimators

There are various types of estimators used in statistics, and they are chosen based on the situation and the properties they possess. Some common types are:

Point Estimators

A point estimator provides a single value as the estimate of the parameter. For instance, the sample mean is a point estimator for the population mean.

Interval Estimators

An interval estimator, or an interval estimate, provides a range of values within which the parameter is expected to lie. Confidence intervals are a common example of interval estimation.

Methods of Estimation

There are several methods used to derive estimators, each with its own set of principles and criteria:

Method of Moments

The method of moments involves equating the population moments (like mean, variance) to the sample moments and solving these equations to estimate the parameters.

Maximum Likelihood Estimation (MLE)

MLE finds the parameter values that maximize the likelihood function, which measures how likely it is to observe the given sample data for different parameter values.

Least Squares Estimation

In least squares estimation, the parameters are estimated by minimizing the sum of squared differences between the observed values and the values predicted by the model.

Bayesian Estimation

Bayesian estimation incorporates prior knowledge or beliefs about the parameter in the form of a prior distribution, which is updated with sample information to obtain a posterior distribution of the parameter.

10	Explain various model selection procedures?	[L2][CO4]	**[12M]**

Model selection is an essential phase in the development of powerful and precise predictive models in the field of machine learning. Model selection is the process of deciding which algorithm and model architecture is best suited for a particular task or dataset.

In machine learning, the process of selecting the top model or algorithm from a list of potential models to address a certain issue is referred to as model selection. It entails assessing and contrasting various models according to how well they function and choosing the one that reaches the highest level of accuracy or prediction power.

- **Problem formulation:** Clearly express the issue at hand, including the kind of predictions or task that you'd like the model to carry out (for example, classification, regression, or clustering).
- **Candidate model selection:** Pick a group of models that are appropriate for the issue at hand. These models can include straightforward methods like decision trees or linear regression as well as more sophisticated ones like deep neural networks, random forests, or support vector machines.
- **Performance evaluation:** Establish measures for measuring how well each model performs. Common measurements include area under the receiver's operating characteristic curve (AUC-ROC), recall, F1-score, mean squared error, and accuracy, precision, and recall. The type of problem and the particular requirements will determine which metrics are used.
- **Training and evaluation:** Each candidate model should be trained using a subset of the available data (the training set), and its performance should be assessed using a different subset (the validation set or via cross-validation). The established evaluation measures are used to gauge the model's effectiveness.
- **Model comparison:** Evaluate the performance of various models and determine which one performs best on the validation set. Take into account elements like data handling capabilities, interpretability, computational difficulty, and accuracy.

R20

- **Hyperparameter tuning:** Before training, many models require that certain hyperparameters, such as the learning rate, regularization strength, or the number of layers that are hidden in a neural network, be configured. Use methods like grid search, random search, and Bayesian optimization to identify these hyperparameters' ideal values.
- **Final model selection:** After the models have been analyzed and fine-tuned, pick the model that performs the best. Then, this model can be used to make predictions based on fresh, unforeseen data

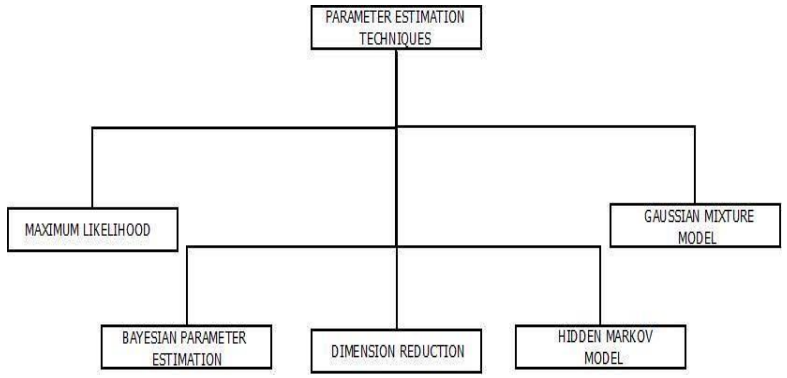
UNIT –V
MULTIVARIATE METHODS

1	Write about multivariate methods?	[L3][CO5]	**[12M]**
	Multivariate methods are statistical techniques used to analyze and model relationships between multiple variables simultaneously. These methods are used when there are multiple dependent variables and independent variables, and the goal is to understand the relationships between them. Some of the commonly used multivariate methods are: 1. Principal Component Analysis (PCA): PCA is a technique used to identify patterns in high-dimensional data. It involves finding a set of orthogonal variables (principal components) that capture the maximum amount of variance in the data. PCA can be used to reduce the dimensionality of the data, to visualize the data in a lower-dimensional space, and to identify the most important variables. 2. Factor Analysis: Factor analysis is a technique used to identify underlying factors that explain the covariance between a set of observed variables. It involves extracting a set of latent variables (factors) that account for the observed correlations between the variables. Factor analysis can be used to reduce the dimensionality of the data, to identify the most important variables, and to identify the underlying structure of the data. 3. Cluster Analysis: Cluster analysis is a technique used to identify groups (clusters) of similar objects based on their similarity or distance. It involves grouping the objects into clusters such that the objectswithin each cluster are more similar to each other than to objects in other clusters. Cluster analysis can be used for exploratory data analysis, data visualization, and data mining. 4. Discriminant Analysis: Discriminant analysis is a technique used to classify objects into predefined categories based on their measurements on multiple variables. It involves finding a set of discriminant functions that can separate the objects into different groups. Discriminant analysis can be used for classification, prediction, and feature selection. 5. Canonical Correlation Analysis: Canonical correlation analysis is a technique used to identify the correlations between two sets of variables. It involves finding a set of canonical variables (linear combinations of the original variables) that maximize the correlation between the two sets of variables. Canonical correlation analysis can beused to identify the relationships between different variables, to identify the most important variables, and to reduce the dimensionality of the data. These multivariate methods can be used alone or in combination with each other, depending on the specific problem and dataset. The choice of the method depends on the goals of the analysis, the nature of the variables, and the computational resources available		
2	What is parameter estimation? Explain in detail?	[L1][CO5]	**[12M]**
	Parameter estimation is the process of determining the values of unknown parameters that define the underlying statistical model from observed data. In other words, it involves finding the best estimate of the parameters that would		

	best explain the observed data. There are two types of parameter estimation techniques: point estimation and interval estimation. 1. **Point Estimation:** Point estimation is a technique used to estimate the value of an unknown parameter from a sample of data. The goal of point estimation is to find a single value that is the best estimate of the true value of the parameter. This estimate is called the point estimate. Point estimates are usually obtained using a statistical model that describes the relationship between the parameters and the data. Examples of point estimation include the maximum likelihood estimator (MLE) and the method of moments estimator (MME). 2. **Interval Estimation:** Interval estimation is a technique used to estimate the range of values that a parameter is likely to fall within. The range of values is called the confidence interval, and it is defined by two values: the lower bound and the upper bound. The confidence interval provides a range of values that is likely to contain the true value of the parameter with a certain level of confidence. Examples of interval estimation include the t-test, the F-test, and the chi- square test. The choice of the parameter estimation technique depends on the nature of the problem and the type of data available. Point estimation is useful when the goal is to estimate the value of a parameter with a single value, whereas interval estimation is useful when the goal is to estimate the range of values within which the true value of the parameter is likely to fall.		
3	Explain multivariate normal distribution in detail?	[L2][CO4]	**[12M]**
	Multivariate normal distribution, also known as multivariate Gaussian distribution, is a probability distribution that describes the joint distribution of a set of random variables that are correlated with each other. It is a generalization of the univariate normal distribution to higher dimensions. The multivariate normal distribution is defined by two parameters: the mean vector μ and the covariance matrix Σ. The mean vector μ is a p-dimensional vector that represents the expected value of each of the p random variables. The covariance matrix Σ is a $p \times p$ matrix that represents the degree of correlation between each pair of the p random variables. The probability density function (pdf) of the multivariate normal distribution is given by: $f(x) = (1/((2\pi)^{p/2} |\Sigma|^{1/2})) * \exp(-1/2(x-\mu)^T\Sigma^{-1}(x-\mu))$ where x is a p-dimensional vector, $|\Sigma|$ is the determinant of the covariance matrix Σ, and $(\cdot)^T$ denotes the transpose of a matrix or vector. The pdf of the multivariate normal distribution has several important properties: 1. It is symmetric around the mean vector μ. 2. It has a peak at the mean vector μ. 3. It has an elliptical shape, with the shape determined by the covariance matrix Σ. 4. The parameters μ and Σ uniquely determine the distribution.		

		One of the most important applications of the multivariate normal distribution is in statistical inference and data analysis. It is commonly used to model the joint distribution of a set of continuous random variables		
4	a	List the features of multivariate normal distribution?	[L1][CO6]	**[6M]**
		The features of the multivariate normal distribution are: 1. Symmetry: The distribution is symmetric around its mean vector. 2. Elliptical shape: The shape of the distribution is determined by the covariance matrix, which reflects the degree of correlation among the random variables. 3. Peak at the mean: The distribution has a maximum at the mean vector, indicating that it is most likely that the random variables take values near the mean. 4. Uniquely determined by mean and covariance: The multivariate normal distribution is completely determined by its mean vector and covariance matrix. 5. Different values of covariance matrix produce different shapes: When the covariance matrix is diagonal, the distribution is a product of independent univariate normal distributions. When the covariance matrix is non-diagonal, the distribution is elliptical and the correlation between variables must be taken into account. 6. Conditional and marginal distributions are also normal: If a subset of the variables is fixed, the conditional distribution of the remaining variables is also normal. Similarly, the marginal distribution of any subset of the variables is also normal. Central Limit Theorem: The multivariate normal distribution is a key component in the central limit theorem, which states that the sum of a large number of independent random variables with finite means and variances approaches a normal distribution.		
	b	Write the applications of multivariate normal distribution?	[L3][CO4]	**[6M]**
		The multivariate normal distribution has a wide range of applications in various fields, some of which are: 1. Finance: In finance, the multivariate normal distribution is used to model asset returns and portfolio optimization. It is also used in risk management and asset pricing models. 2. Engineering: In engineering, the multivariate normal distribution is used to model the behavior of complex systems, such as electronic circuits, control systems, and manufacturing processes. 3. Social sciences: In the social sciences, the multivariate normal distribution is used to analyze data from surveys and experiments, and to model relationships between variables in fields such as psychology and sociology. 4. Medical research: In medical research, the multivariate normal distribution is used to model the distribution of patient characteristics, such as age, sex, and disease severity, and to analyze the relationships between these characteristics and health outcomes. 5. Image processing: In image processing, the multivariate normal distribution is used to model the distribution of pixel values in		

		images, and to segment and classify images based on their features. 6. Machine learning: In machine learning, the multivariate normal distribution is used as a building block for many models, such as Gaussian mixture models, Bayesian networks, and Hidden Markov models. Quality control: In quality control, the multivariate normal distribution is used to model the distribution of measurements from production processes, and to detect deviations from expected values.		
5		State and explain tuning complexity?	[L1][CO5]	**[12M]**
		Tuning complexity, also known as model complexity tuning, is a process of selecting the appropriate level of complexity for a machine learning model. It is a crucial step in the machine learning pipeline as an overly complex or overly simple model can lead to poor performance. Tuning complexity involves finding the best values for the hyperparameters of a model. Hyperparameters are the parameters of a model that are not learned during training, but are set before training. Examples of hyperparameters include the number of hidden layers in a neural network, thelearning rate of the optimizer, and the regularization strength. Tuning complexity involves trying out different values of hyperparameters and evaluating the model's performance on a validation set. The goal is tofind the set of hyperparameters that maximizes the model's performancewithout overfitting or underfitting the data. Overfitting occurs when the model is too complex for the data, leading to poor generalization performance. Underfitting occurs when the model is too simple for the data, leading to poor performance on both training and testsets. The process of tuning complexity is iterative, and involves training and evaluating the model with different hyperparameter values until the best set of hyperparameters is found. A common approach is to use a grid search or a randomized search to explore the hyperparameter space.		
6	a	Write some features of multivariate normal distribution?	[L3][CO5]	**[6M]**
		The multivariate normal distribution is a probability distribution over a vector of random variables, where each variable may be correlated with one another. Some features of the multivariate normal distribution are: 1. It is a generalization of the univariate normal distribution to multiple dimensions. 2. The distribution is characterized by a mean vector and a covariance matrix. 3. The mean vector specifies the expected value of each random variable in the vector. 4. The covariance matrix describes the linear relationship between pairs of random variables. 5. The multivariate normal distribution is fully specified by its mean vector and covariance matrix. 6. The distribution is symmetric around the mean vector and has a bell-shaped curve. 7. The distribution is parameterized by the mean and covariance matrix, which can be estimated from data. 8. The marginal distributions of each variable in the vector are		

		univariate normal distributions. The multivariate normal distribution is widely used in multivariate statistical analysis, machine learning, and data science.		
	b	List few parameter estimation techniques?	[L1][CO3]	**[6M]**
		 Here are some commonly used parameter estimation techniques: 1. Maximum Likelihood Estimation (MLE): It is a widely used method for estimating the parameters of a statistical model. The maximum likelihood estimate is the value of the parameter that maximizes the likelihood function. 2. Bayesian Estimation: It is a method of estimating the parameters of a statistical model by incorporating prior knowledge about the parameters. $$P(x|X) = \int P(x|\Theta)\, P(\Theta/X)\, d\Theta$$ 3. Least Squares Estim ation (LSE): It is a method for estimating the parameters of a model by minimizing the sum of the squared residuals. 4. Maximum A Posteriori (MAP) Estimation: This method estimates the parameters of a model by maximizing the posterior probability of the parameters given the data and prior information. 5. Expectation-Maximization (EM) Algorithm: Expectation maximization the process that is used for clustering the data sample. It works on the concept of, starting with the random theory and randomly classified data along with the execution of below mentioned steps. **Step-1("E")**: In this step, Classification of current data using the theory that is currently being used is done. **Step-2("M")**: In this step, With the help of current classification of data, theory for that is generated. Thus EM means, Expected classification for each sample is generated used step-1 and theory is generated using step-2. 6. Kernel Density Estimation (KDE): It is a non-parametric method for estimating the probability density function of a random variable. Generalized Method of Moments (GMM): It is a method for estimating the parameters of a model by matching the theoretical moments of the		

	model with the empirical moments of the data.		
7	Explain in detail about clustering and types of clustering?	[L2][CO5]	**[12M]**

Clustering is the task of dividing the unlabeled data or data points intodifferent clusters such that similar data points fall in the same cluster than those which differ from the others

Types of Clustering

Broadly speaking, clustering can be divided into two subgroups:

- **Hard Clustering:** In this, each input data point either belongs to a cluster completely or not.
- **Soft Clustering**: In this, instead of putting each input data point into a separate cluster, a probability or likelihood of that data point being in those clusters is assigned.

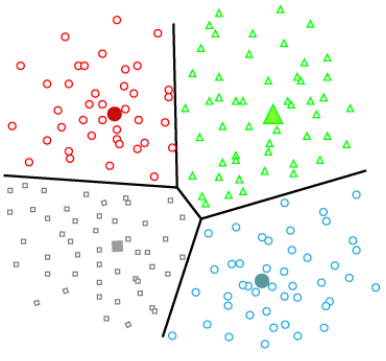
Main clustering methods used in Machine learning:

1. **Partitioning Clustering**
2. **Density-Based Clustering**
3. **Distribution Model-Based Clustering**
4. **Hierarchical Clustering**
5. **Fuzzy Clustering**

Partitioning Clustering

It is a type of clustering that divides the data into non-hierarchical groups. It is also known as the **centroid-based method**. The most common example of partitioning clustering is the **K-Means Clustering algorithm**.

In this type, the dataset is divided into a set of k groups, where K is used to define the number of pre-defined groups. The cluster centre is created in such a way that the distance between the data points of one cluster is minimum as compared to another cluster centroid.

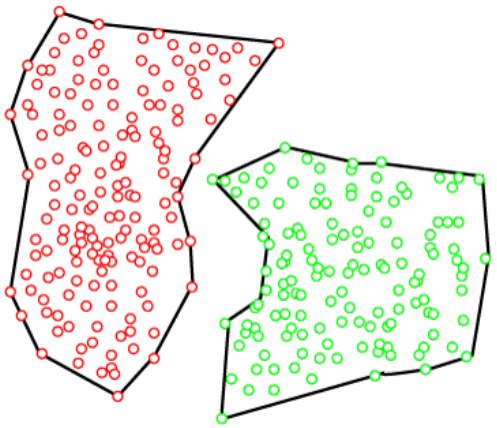


Density-Based Clustering

The density-based clustering method connects the highly-dense areas into clusters, and the arbitrarily shaped distributions are formed as long as the dense region can be connected. This algorithm does it by identifying different clusters in the dataset and connects the areas of high densities into clusters.The dense areas in data space are divided from each other by sparser areas.

These algorithms can face difficulty in clustering the data points if the dataset

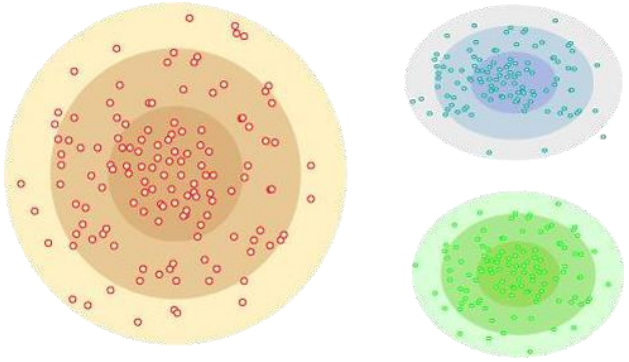
has varying densities and high dimensions.



Distribution Model-Based Clustering

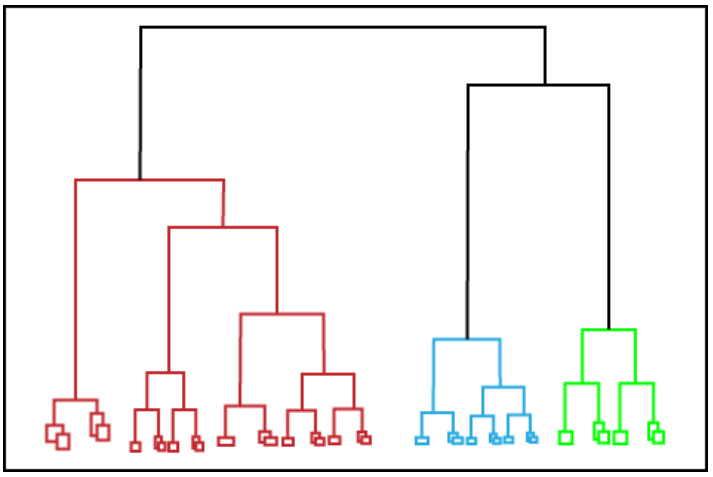
In the distribution model-based clustering method, the data is divided based on the probability of how a dataset belongs to a particular distribution. The grouping is done by assuming some distributions commonly **Gaussian Distribution**.

The example of this type is the **Expectation-Maximization Clustering algorithm** that uses Gaussian Mixture Models (GMM).



Hierarchical Clustering

Hierarchical clustering can be used as an alternative for the partitioned clustering as there is no requirement of pre-specifying the number of clusters to be created. In this technique, the dataset is divided into clusters to create a tree-like structure, which is also called a **dendrogram**. The observations or any number of clusters can be selected by cutting the tree at the correct level. The most common example of this method is the **Agglomerative Hierarchical algorithm**.



Fuzzy Clustering

Fuzzy clustering is a type of soft method in which a data object may belong to more than one group or cluster. Each dataset has a set of membership coefficients, which depend on the degree of membership to be in a cluster. **Fuzzy C-means algorithm** is the example of this type of clustering; it is sometimes also known as the Fuzzy k-means algorithm.

8	a	Explain how multivariate regression is implemented?	[L3][CO5]	**[6M]**

Regression analysis is an important statistical method that allows us to examine the relationship between two or more variables in the dataset.

Multivariate regression is a technique used to measure the degree towhich the various independent variable and various dependent variables are linearly related to each other. The relation is said to be linear due tothe correlation between the variables. Once the multivariate regression is applied to the dataset, this method is then used to predict the behavior of the response variable based on its corresponding predictor variables.

Multivariate regression is commonly used as a supervised algorithm in machine learning, a model to predict the behavior of dependent variables and multiple independent variables.

Steps to achieve multivariate regression

Step 1: Select the features

First, you need to select that one feature that drives the multivariate regression. This is the feature that is highly responsible for the change in your dependent variable.

Step 2: Normalize the feature

Now that we have our selected features, it is time to scale them in a certain range (preferably 0-1) so that analysing them gets a bit easy.

To change the value of each feature, we can use:

Step 3: Select loss function and formulate a hypothesis

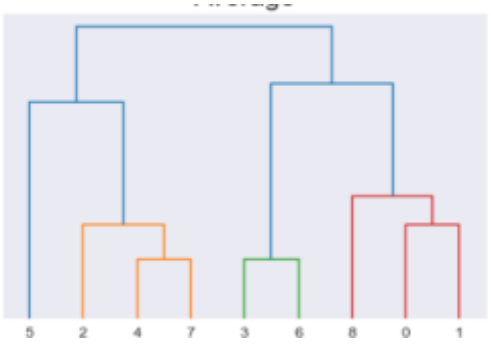
A formulated hypothesis is nothing but a predicted value of the response variable and is denoted by h(x).

A loss function is a calculated loss when the hypothesis predicts a wrong value. A cost function is a cost handled for those wrongly predicting

		hypotheses. **Step 4: Minimize the cost and loss function** Both cost function and loss function are dependent on each other. Hence, in order to minimize both of them, minimization algorithms can be run over the datasets. These algorithms then adjust the parameters of the hypothesis. One of the minimization algorithms that can be used is the gradient descent algorithm. **Step 5: Test the hypothesis** The formulated hypothesis is then tested with a test set to check its accuracy and correctness.		
	b	Describe the uses of multivariate regression?	[L1][CO4]	**[6M]**
		Multivariate regression is a statistical technique used to examine the relationship between multiple independent variables and a dependent variable. It extends the principles of simple linear regression, which involves only one independent variable. Here are some common uses of multivariate regression: **1. Economics and Finance:** • **Macroeconomic Modeling:** Multivariate regression is frequently used in economics to model the relationships between various economic indicators such as GDP, inflation, unemployment, and interest rates. • **Financial Analysis:** In finance, multivariate regression can be applied to analyze the factors influencing stock prices, bond yields, or other financial instruments. **2. Marketing and Business:** • **Market Research:** Companies use multivariate regression to understand the impact of multiple marketing variables (e.g., advertising expenditure, product features, pricing) on sales or market share. • **Customer Behavior Analysis:** It helps analyze the factors affecting customer behavior, such as purchasing decisions, satisfaction, and loyalty. **3. Environmental Science:** • **Climate Modeling:** Multivariate regression is used to model the relationships between various climate variables, such as temperature, precipitation, and atmospheric pressure. • **Environmental Impact Assessments:** Researchers use multivariate regression to assess the impact of multiple factors on environmental outcomes, such as air or water quality. **4. Medicine and Healthcare:** • **Clinical Research:** In medical studies, multivariate regression is employed to analyze the impact of multiple variables on health outcomes, taking into account factors		

		like age, gender, and treatment regimens. • **Epidemiology:** Researchers use multivariate regression to study the relationships between multiple risk factors and the occurrence of diseases. **5. Social Sciences:** • **Education Research:** Multivariate regression is used to analyze the impact of various factors (e.g., teachingmethods, socio-economic status) on academic achievement. • **Psychology Studies:** Psychologists use multivariate regression to examine the relationships between multiple variables, such as personality traits, environmental factors, and mental health outcomes. **6. Quality Control and Manufacturing:** • **Process Optimization:** In manufacturing, multivariate regression can be applied to optimize production processes by analyzing the impact of multiple variables on product quality. • **Quality Assurance:** It is used to assess the influence of various factors on the quality of manufactured goods. **7. Sports Analytics:** • **Performance Analysis:** Multivariate regression can be used in sports analytics to understand the factors influencing team or individual performance, considering variables like player skills, coaching strategies, and environmental conditions. **8. Public Policy and Government:** • **Policy Evaluation:** Governments use multivariateregression to evaluate the impact of policies by considering various factors affecting social and economic outcomes. • **Social Welfare Programs:** It helps analyze the effectiveness of social programs by examining the influenceof multiple variables on outcomes like poverty rates or educational attainment.		
9	Explain in detail about a) Agglomerative Clustering b) Hierarchical Clustering	[L2][CO5]	**[12M]**
	Hierarchical clustering is a popular method for grouping objects. It creates groups so that objects within a group are similar to each other and different from objects in other groups. Clusters are visually represented in a hierarchical tree called a dendrogram. **Hierarchical clustering types** There are two main types of hierarchical clustering: 1. **Agglomerative**: Initially, each object is considered to be its own cluster. According to a particular procedure, the clusters are then merged step by step until a single cluster remains. At the end of the cluster merging process, a cluster containing all the elements will be formed. 2. **Divisive**: The Divisive method is the opposite of the Agglomerative method. Initially, all objects are considered in a single cluster. Then the division process is performed step by step until each object forms a different cluster. The cluster division or splitting procedure is carried		

out according to some principles that maximum distance between neighboring objects in the cluster.



Steps to perform Agglomerative Clustering :

1. Find the two features that are ―closest‖ in multivariate space
2. Replace them with a single feature at their mean
3. Repeat with the next two closest features, and continue until all the features are subsumed into one cluster

10	a	Explain maximum likelihood estimation in detail?	[L2][CO5]	**[6M]**

Maximum likelihood estimation is a method that determines values for the parameters of a model. The parameter values are found such that they maximise the likelihood that the process described by the model produced the data that were actually observed.

In Maximum Likelihood Estimation, we maximize the conditional probability of observing the data **(X)** given a specific probability distribution and its parameters **(theta – ɵ)**

- P(X,ɵ) where X is the joint probability distribution of all observations from 1 to n.

P(X1, X2, X3….Xn; ɵ)

- The resulting conditional probability is known as the likelihood of observing the data with the given model parameters and denoted as **(L)**
- L(X, ɵ)

The joint probability can also be defined as the multiplication of the conditional probability for each observation given the distribution parameters

- Sum i to n log [(P(xi, ɵ)]

As log is used mostly in the likelihood function, it is known as log-likelihood function. It is common in optimization problems to prefer to minimize the cost function.

Therefore, the negative of the log-likelihood function is used and known as Negative Log-Likelihood function.

- **Minimize**: Sum i to n log [(P(xi, ɵ)]
- The Maximum Likelihood Estimation framework can be used as a basis for estimating the parameters of many different machine learning models for regression and classification predictive modeling. This includes the logistic regression model.

		Let X1, X2, X3, ……, Xn be a random sample from a distribution with a parameter θ. Suppose that we have observed X1=x1, X2=x2, ……, Xn=xn. 1. If Xi's are discrete, then the likelihood function is defined as L (x1, x2, ⋯, xn; θ) = Px1x2⋯xn(x1, x2,⋯,xn;θ). 1. If Xi's are jointly continuous, then the likelihood function is defined as L (x1, x2, ⋯, xn; θ) = fx1x2⋯xn(x1, x2,⋯,xn;θ).		
	b	What is minimum mean square error estimation?	[L1][CO4]	**[6M]**
		Minimum mean square error (MMSE) estimation is a method used to estimate a parameter or a signal in the presence of noise. It is a popular technique used in signal processing, communication systems, and other fields where the goal is to extract a signal from a noisy environment. In MMSE estimation, the goal is to minimize the expected value of the squared error between the estimate and the true value of the parameter or signal. This is done by using the conditional probability distribution of the parameter or signal given the observed data. The MMSE estimate is the value of the parameter or signal that minimizes this expected value of the squared error. One important application of MMSE estimation is in linear regression, where the goal is to estimate the parameters of a linear model given a set of observations. In this case, the MMSE estimate of the parameters is the solution to the least squares problem, which involves minimizing the sum of the squared errors between the observed data and the predictions of the linear model. MMSE estimation can also be used in other contexts, such as in signal processing applications where the goal is to estimate a signal that has been corrupted by noise. In this case, the MMSE estimate of the signal can be computed using a Wiener filter, which is a linear filter that minimizes the expected value of the squared error between the estimated signal and the true signal.		