1. What is Machine Learning?

Machine learning is a field of artificial intelligence that involves developing algorithms that allow computers to learn from and make predictions or decisions based on data. Unlike traditional programming, where explicit instructions are provided, machine learning models identify patterns in data and improve their performance over time without being explicitly programmed for specific tasks.

2. What are the different types of Machine Learning?

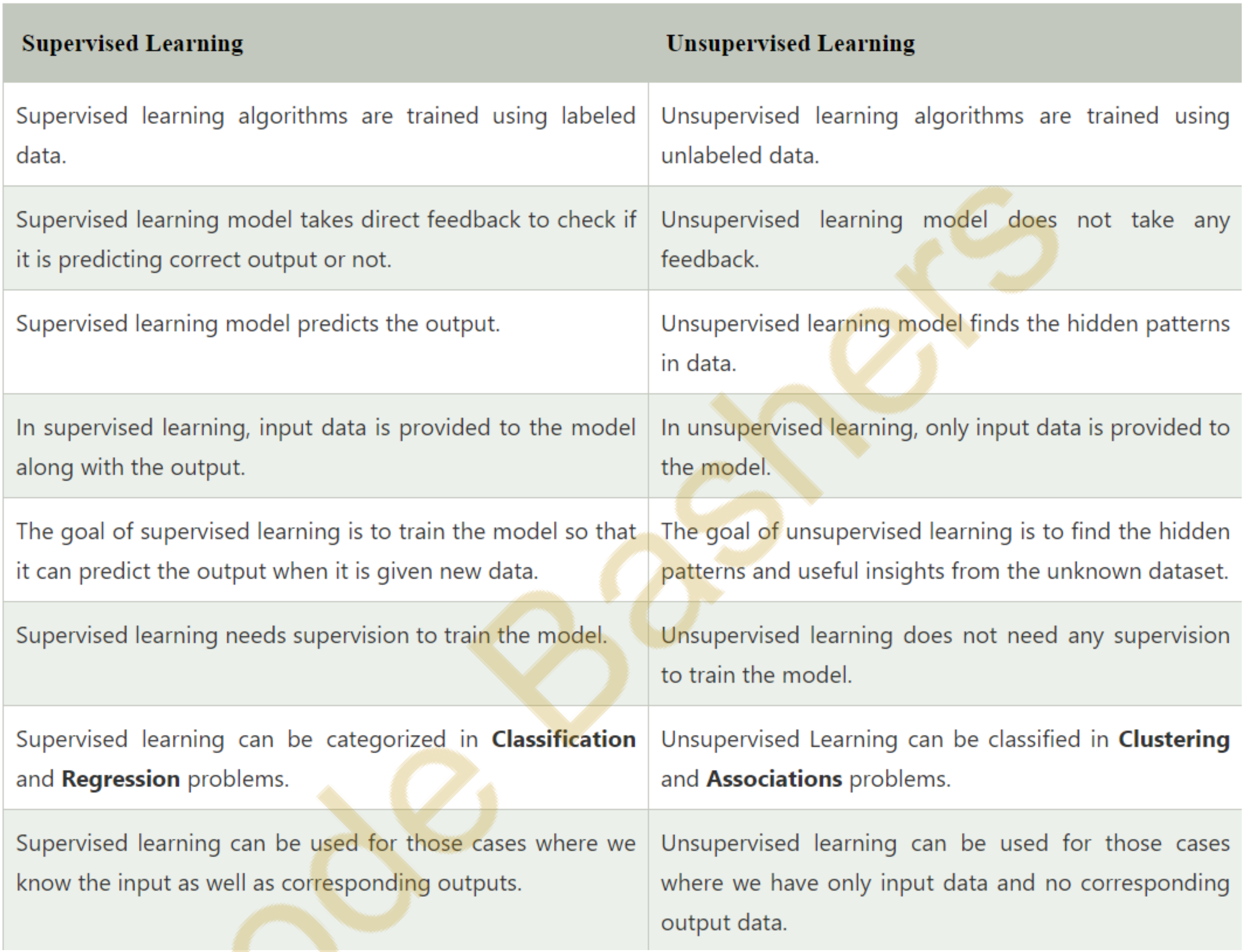
Supervised Learning: This type involves training a model on a labeled dataset, which means each training example is paired with an output label. Examples include classification and regression.

Unsupervised Learning: The model learns from unlabeled data by identifying patterns and relationships. Examples include clustering and association.

Reinforcement Learning: The model learns by interacting with an environment to maximize some notion of cumulative reward. It involves agents, actions, and rewards.

Code Bashers

3. What is the difference between supervised and unsupervised learning?



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

Code Bashers

4. Explain the bias-variance tradeoff.

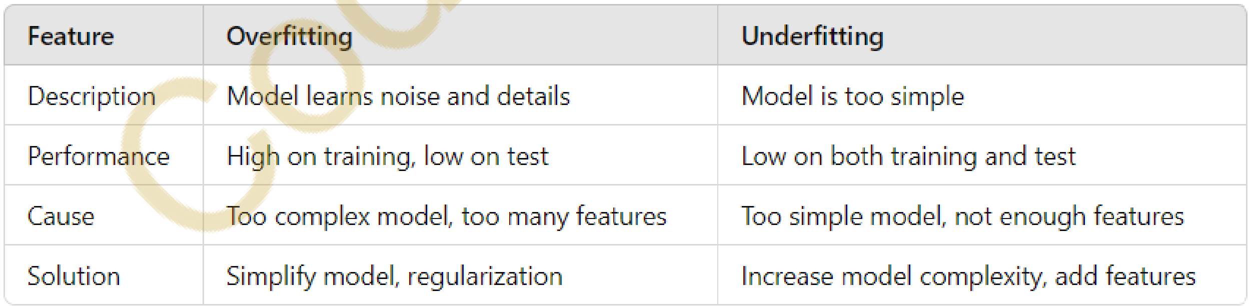
Bias: Bias refers to errors due to overly simplistic assumptions in the learning algorithm. High bias can cause underfitting, where the model fails to capture the complexity of the data.

Variance: Variance refers to errors due to excessive complexity in the learning algorithm. High variance can cause overfitting, where the model captures noise along with the underlying data pattern.

Tradeoff: The bias-variance tradeoff involves finding a balance between bias and variance to minimize the total error. This balance ensures the model generalizes well to unseen data.

Code Bashers

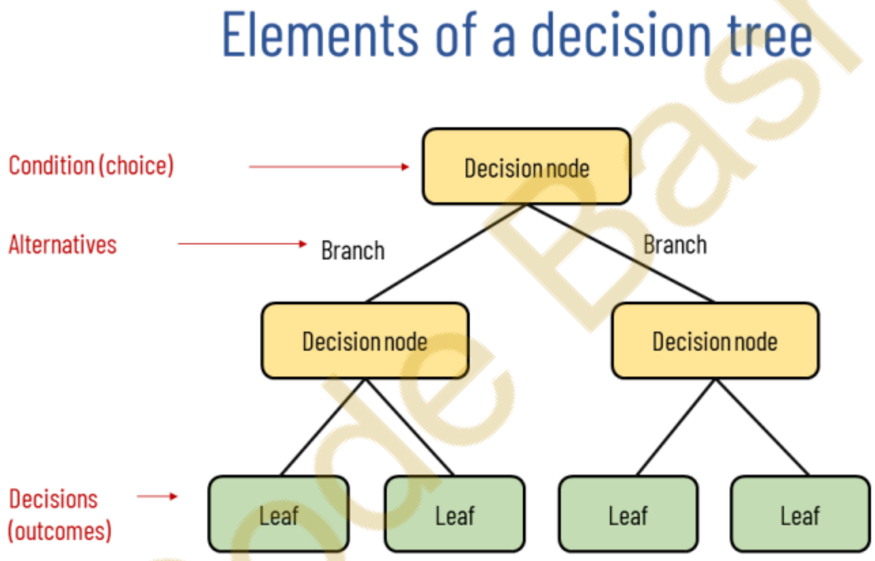
5. What is overfitting and underfitting?



Feature	Overfitting	Underfitting
Description	Model learns noise and details	Model is too simple
Performance	High on training, low on test	Low on both training and test
Cause	Too complex model, too many features	Too simple model, not enough features
Solution	Simplify model, regularization	Increase model complexity, add features

6. Explain the working of a decision tree.

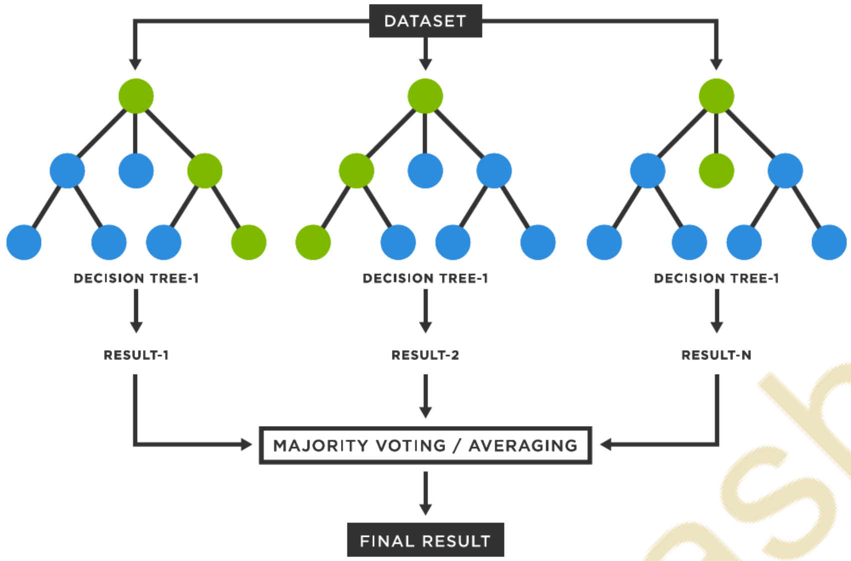
A decision tree is a model that splits data into branches based on feature values, forming a tree-like structure. Each internal node represents a decision on a feature, each branch represents an outcome of the decision, and each leaf node represents a class label or regression value. The goal is to create a tree that accurately predicts the target variable by making splits that result in the most homogeneous subsets of data.



Code Bashers

7. What is a random forest?

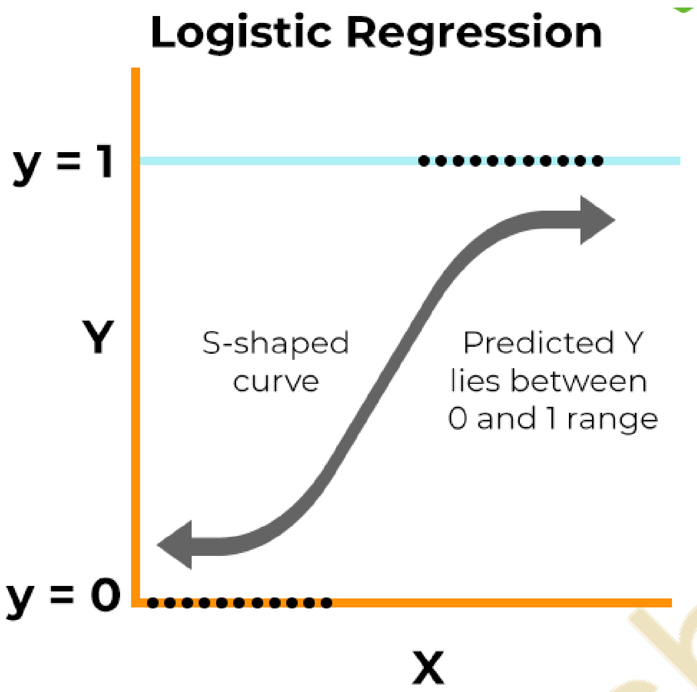
A random forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. It improves predictive accuracy and controls overfitting by averaging multiple decision trees, each trained on a different part of the same dataset with replacement (bagging).



8. What is logistic regression?

Logistic regression is a statistical method for binary classification that models the probability that an instance belongs to a particular class. It uses the logistic function (sigmoid) to map predicted values to probabilities. Unlike linear regression, which predicts continuous values, logistic regression predicts the probability of a binary outcome (0 or 1).

Code Bashers



Code Bashers

9.Why we cannot use linear regression for a classification task?

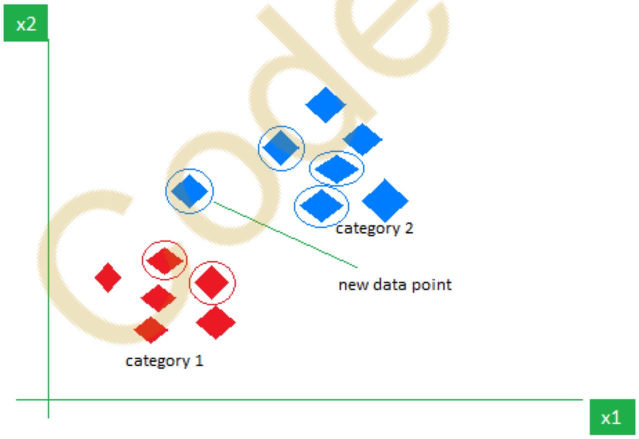
The main reason why we cannot use linear regression for a classification task is that the output of linear regression is continuous and unbounded, while classification requires discrete and bounded output values.

If we use linear regression for the classification task the error function graph will not be convex. A convex graph has only one minimum which is also known as the global minima but in the case of the non-convex graph, there are chances of our model getting stuck at some local minima which may not be the global minima. To avoid this situation of getting stuck at the local minima we do not use the linear regression algorithm for a classification task.

10. Explain the k-nearest neighbors (k-NN) algorithm.

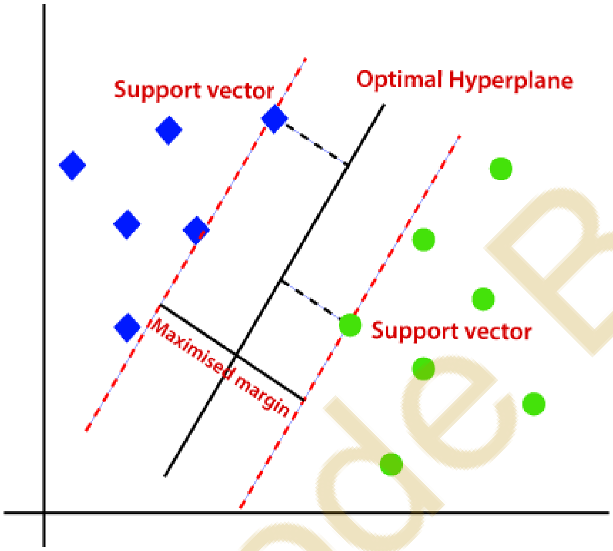
The k-NN algorithm classifies data points based on the k closest training examples in the feature space. It is a type of instance-based learning where the function is only approximated locally. To classify a new data point, the algorithm finds the k closest points to it and assigns the most common class among these neighbors. The value of k is a hyperparameter that determines the number of neighbors considered.

Code Bashers



11. What is a support vector machine (SVM)?

SVM is a supervised learning algorithm used for classification and regression. It finds the hyperplane that best separates different classes in the feature space by maximizing the margin between the closest points of the classes (support vectors). SVM can handle linear and non-linear classification using kernel functions, which transform the data into a higher-dimensional space where a linear separator can be found.



Code Bashers

12.What is the difference between k-means and the KNN algorithm?

k-means algorithm is one of the popular unsupervised machine learning algorithms which is used for clustering purposes. But the KNN is a model which is generally used for the classification task and is a supervised machine learning algorithm. The k-means algorithm helps us to label the data by forming clusters within the dataset.

13.What is the difference between L1 and L2 regularization? What is their significance?

L1 regularization: In L1 regularization also known as Lasso regularization in which we add the sum of absolute values of the weights of the model in the loss function. In L1 regularization weights for those features which are not at all important are penalized to zero so, in turn, we obtain feature selection by using the L1 regularization technique.

L2 regularization: In L2 regularization also known as Ridge regularization in which we add the square of the weights to the loss function. In both of these regularization methods, weights are penalized but there is a subtle difference between the objective they help to achieve.

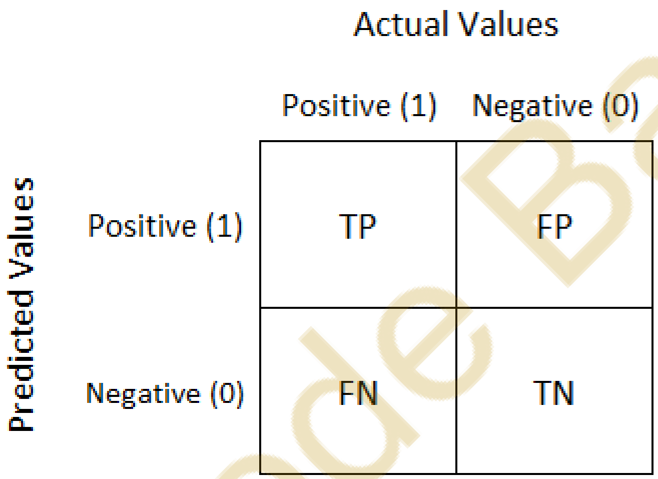
In L2 regularization the weights are not penalized to 0 but they are near zero for irrelevant features. It is often used to prevent overfitting by shrinking the weights towards zero,

Code Bashers

especially when there are many features and the data is noisy.

15. What is a confusion matrix?

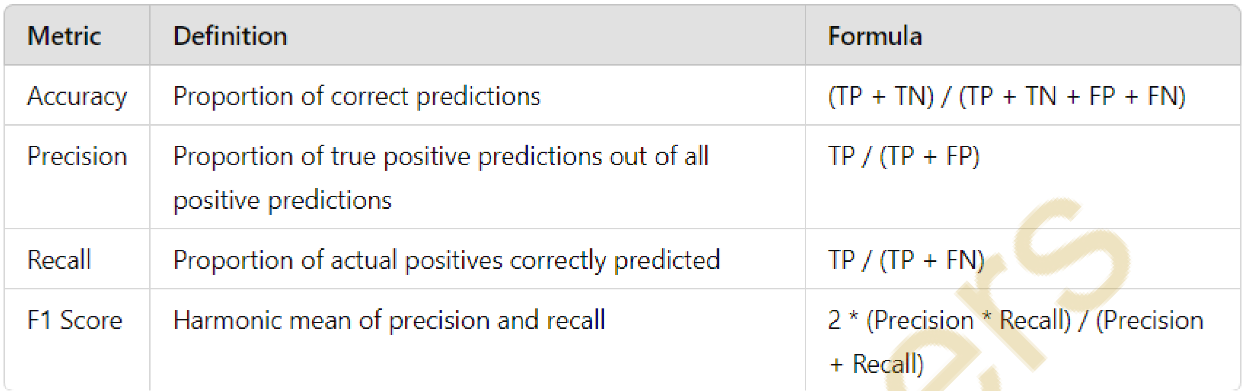
A confusion matrix is a table used to evaluate the performance of a classification model. It shows the number of true positives, true negatives, false positives, and false negatives. This matrix helps in understanding the performance of an algorithm beyond simple accuracy by providing insights into which classes are being misclassified.



		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Code Bashers

16. What is accuracy, precision, recall, and F1 score?



Metric	Definition	Formula
Accuracy	Proportion of correct predictions	(TP + TN) / (TP + TN + FP + FN)
Precision	Proportion of true positive predictions out of all positive predictions	TP / (TP + FP)
Recall	Proportion of actual positives correctly predicted	TP / (TP + FN)
F1 Score	Harmonic mean of precision and recall	2 * (Precision * Recall) / (Precision + Recall)

TP (True Positive): Correctly predicted positive cases

TN (True Negative): Correctly predicted negative cases

FP (False Positive): Incorrectly predicted positive cases

FN (False Negative): Incorrectly predicted negative cases

17. What is cross-validation?

Cross-validation is a technique used to assess the performance of a model by dividing the data into multiple subsets or folds. One subset is used as the test set while the others are used for training. This process is repeated multiple times with different subsets as the test set each time. Common techniques include k-fold cross-validation, where the data is divided into k subsets.

Code Bashers

18. Why is data preprocessing important?

Data preprocessing is crucial because real-world data is often incomplete, inconsistent, and noisy. Preprocessing involves cleaning the data (handling missing values, removing duplicates), transforming the data (normalization, encoding categorical variables), and preparing it for modeling. Proper preprocessing improves the quality of the data, leading to better model performance and more accurate predictions.

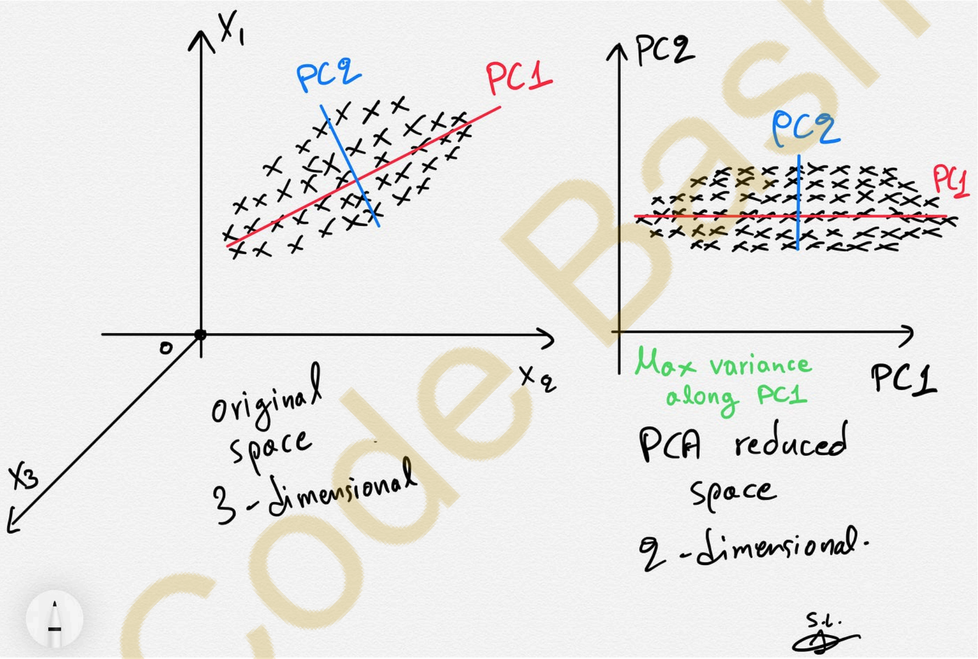
19. What is feature scaling and why is it important?

Feature scaling standardizes the range of independent variables or features in the data. It ensures that features contribute equally to the distance calculations in algorithms like k-NN and SVM, which are sensitive to the scale of data. Common techniques include normalization (scaling values between 0 and 1) and standardization (scaling to have a mean of 0 and a standard deviation of 1).

Code Bashers

20. Explain principal component analysis (PCA).

PCA is a dimensionality reduction technique that transforms the original features into a new set of uncorrelated features called principal components. These components are ordered by the amount of variance they capture from the data. PCA helps in reducing the number of features while retaining most of the variability in the data, making it easier to visualize and reducing computational cost.



Code Bashers

21. What is feature selection?

Feature selection involves selecting the most relevant features for building a model. It helps in reducing the dimensionality of the data, improving model performance, and reducing overfitting. Methods include:

Forward Selection: Starting with no features and adding one at a time based on model improvement.

Backward Elimination: Starting with all features and removing the least significant one at a time.

Regularization: Techniques like Lasso (L1) and Ridge (L2) regression that add penalties for feature coefficients, effectively shrinking less important ones to zero.

Code Bashers

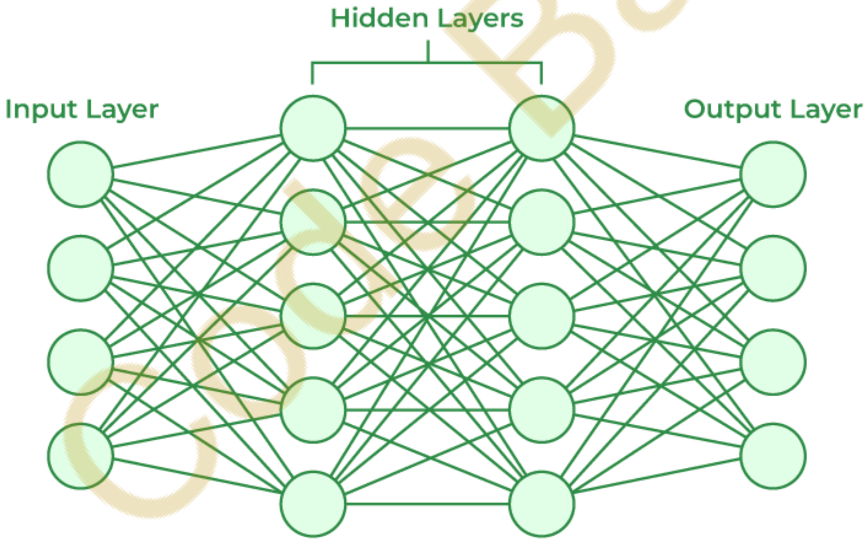
22.What is the purpose of splitting a given dataset into training and validation data?

The main purpose is to keep some data left over on which the model has not been trained so, that we can evaluate the performance of our machine learning model after training. Also, sometimes we use the validation dataset to choose among the multiple state-of-the-art machine learning models. Like we first train some models let’s say LogisticRegression, XGBoost, or any other than test their performance using validation data and choose the model which has less difference between the validation and the training accuracy.

23. What is a neural network?

A neural network is a series of algorithms that attempt to recognize relationships in a set of data through a process that mimics the way the human brain operates. It consists of layers of nodes (neurons), with each layer transforming the input data and passing it to the next layer. The network learns by adjusting the weights of the connections between nodes based on the error of the output.

Code Bashers

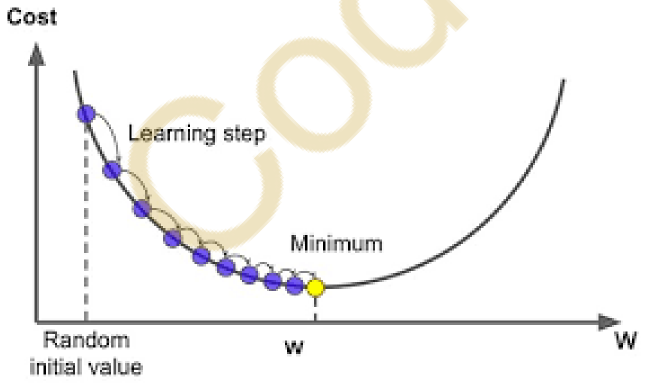


24. Explain the concept of deep learning.

Deep learning is a subset of machine learning that uses neural networks with many layers (deep networks) to model complex patterns in large amounts of data. These deep networks can automatically learn hierarchical features from the raw input, making them highly effective for tasks like image and speech recognition, natural language processing, and more.

25. What is gradient descent?

Gradient descent is an optimization algorithm used to minimize the loss function of a model by iteratively adjusting the model parameters. It calculates the gradient (partial derivatives) of the loss function with respect to the parameters and updates the parameters in the direction of the negative gradient. Variants include batch gradient descent, stochastic gradient descent, and mini-batch gradient descent.



Code Bashers

26. How would you handle an imbalanced dataset?

Handling an imbalanced dataset involves several techniques:

Resampling: Balancing the dataset by oversampling the minority class or undersampling the majority class.

Synthetic Data Generation: Creating synthetic examples for the minority class using methods like SMOTE (Synthetic Minority Over-sampling Technique).

Class Weights: Adjusting the weights of classes in algorithms to give more importance to the minority class.

Algorithm Choice: Using algorithms that are robust to class imbalance, such as decision trees and ensemble methods.

27. How do you deal with missing values in a dataset?

Techniques for handling missing values include:

Imputation: Filling missing values with mean, median, mode, or using more sophisticated methods like k-NN imputation.

Removal: Deleting rows or columns with missing values, if they are not significant or if the missing percentage is low.

Algorithms Handling Missing Data: Using algorithms that can handle missing values natively, such as decision trees.

Code Bashers

28. What is learning Rate?

The learning rate is a parameter in machine learning that determines how much the model's parameters are updated during training. It controls the size of the steps taken towards minimizing the error. A high learning rate can lead to fast but unstable learning, while a low learning rate can result in slow but stable learning. Finding the right balance is key for effective training.

Code Bashers