Q:2 What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sum of Squares) in regression? Also mention the equation relating these three metrics with each other.
Ans :

TSS is the total variation in the target variable, and can be calculated as the sum of the squared differences between the actual target values and the mean target value:

$TSS = \sum(y_i - \bar{y})^2$

where y_i is the actual target value, $\bar{y}$ is the mean target value, and the sum is taken over all the data points.

ESS is the explained variation in the target variable, and can be calculated as the sum of the squared differences between the predicted target values and the mean target value:

$ESS = \sum(\hat{y}_i - \bar{y})^2$

where $\hat{y}_i$ is the predicted target value, and the sum is taken over all the data points.

RSS is the residual variation in the target variable, and can be calculated as the sum of the squared differences between the actual target values and the predicted target values:

$RSS = \sum(y_i - \hat{y}_i)^2$

where y_i is the actual target value, and $\hat{y}_i$ is the predicted target value, and the sum is taken over all the data points.

The relationship between these three metrics can be expressed as follows:

TSS = ESS + RSS

Q:3 What is the need of regularization in machine learning?
Ans:- Regularization helps to balance the bias-variance tradeoff in machine learning.
By adding a penalty to the models loss function, Regularization can help to reduce the variance of the model,
which is the tendency to overfit to the training data.

Q:4 What is Gini–impurity index?
Ans :- The Gini impurity index is a measure of the heterogeneity of a set of categorical data, commonly used in decision trees and random forests to determine the best splitting criteria for the data.

Q:5 Are unregularized decision-trees prone to overfitting? If yes, why?
Ans: Yes, unregularized decision trees are prone to overfitting, especially when the tree is deep.

The reason for this is that decision trees are capable of fitting the training data perfectly, by creating branches that split the data until each leaf contains only a single data point.
This is known as memorizing the training data, and it can lead to overfitting because the model becomes too complex and captures the noise in the training data rather than the underlying patterns.

Q:6 What is an ensemble technique in machine learning?
Ans:- An ensemble technique in machine learning is a method of combining multiple models to improve the overall accuracy and robustness of a predictive model.

Q:7. What is the difference between Bagging and Boosting techniques?
Ans : Bagging (Bootstrap Aggregating) is a technique in which multiple models are trained on different subsets of the training data,
where each subset is selected randomly with replacement.

Boosting is a technique in which multiple models are trained sequentially,
where each subsequent model focuses more on the samples that were misclassified by the previous model

Q:8 What is out-of-bag error in random forests?
Ans : The OOB error in Random Forest is a useful metric that provides an estimate of the test error of the model without the need for a separate validation set.
It is calculated by evaluating the OOB samples using only the trees that were not trained on those samples
and then averaging the prediction errors.

Q:9 What is K-fold cross-validation?
Ans : K-fold cross-validation is a technique used in machine learning to evaluate
the performance of a model on a dataset.

Q:11. What issues can occur if we have a large learning rate in Gradient Descent?
Ans:

If the learning rate is too large, it can cause several issues that can negatively affect the performance of the model,
including:
1)Divergence,
2}Slow convergence
3)Unstable behavior

Q:13 Differentiate between Adaboost and Gradient Boosting.

Ans : Adaboost and Gradient Boosting are both boosting algorithms that aim to improve the performance of a model by combining weak learners to create a stronger model. However, they differ in the method of weighted instances, model complexity,
learning rate, parallelism, and robustness to noisy data.

Q:14 What is the bias-variance trade off in machine learning?
Ans : The bias-variance tradeoff is a crucial concept in machine learning that
helps us understand the relationship between a models
ability to fit the training data and its ability to generalize to new, unseen data, and
guides us in selecting appropriate models that strike a balance between these two properties.