1.

Two measures are helpful in evaluating the goodness of fit in regression analysis but in different aspects. RSS is used to recognise the degree of error in the model's predictions while the other one is good at assessing the overall fit of the model and to compare different models. Higher the R squared value represent a better fit while in RSS , lower the value shows the better fit.

2.
In regression analysis:
Total Sum of Squares (TSS): It measures the total variance in the dependent variable. It's the sum of the squared differences between each observed dependent variable value and the mean of the dependent variable.
Explained Sum of Squares (ESS):Also known as Regression Sum of Squares, it quantifies the variance in the dependent variable that is explained by the regression model. It is the sum of the squared differences between the predicted values and the mean of the dependent variable.
Residual Sum of Squares (RSS):This measures the unexplained variance in the dependent variable, representing the sum of the squared differences between the observed values and the predicted values from the regression model.

The relationship among these is given by the equation: TSS = ESS + RSS. This decomposition helps evaluate how well the regression model explains the variability in the dependent variable.

3.

Regularisation techniques help to properly fit a model to an optimal model. It improves machine learning models by correcting overfitting and under-fitting by minimising needless complexity.

4.

Gini Impurity Index is a measure of how mixed or impure a dataset is. It shows the measure of how often a randomly chosen element from a dataset would be incorrectly classified if it was randomly labeled according to the distribution of labels in the dataset.

5

Unregularized decision trees are prone to overfitting, especially when the tree is allowed to grow deep and capture noise or specific details of the training data. Regularization techniques like pruning or limiting the maximum depth of the tree are often employed to mitigate this issue.

6.

It is a technique that combines multiple models into a single model to improve performance. Some common ensemble techniques are Boosting, Random Forest, Bagging, Gradient Boosting and Stacking.

7.

Bagging combines the same type of predictions, while boosting combines different types of predictions.

Bagging reduces variance, while boosting reduces bias.

In bagging, weak learners are trained in parallel. In boosting, they learn sequentially.

Bagging can be used with unstable models like decision trees. Boosting works better for stable models like linear regression.

8.

Out-of-bag error is a method of measuring the prediction error of random forests and other machine learning models. It is an estimate of the model's performance on unseen data. It is calculated by using the observations not included in the bootstrap sample to evaluate the prediction error. This provides a convenient estimate of the model's generalization error without the need for an additional validation set.

9.

K-fold cross-validation is a machine learning technique that evaluates the performance of a model by dividing a dataset into k subsets, or folds.The model is trained and evaluated k times, using a different fold as the validation set each time. Performance metrics from each fold are averaged to estimate the model's generalization performance. However, there are some limitations to k-fold cross-validation.
It increases the computational cost and time because it requires training and testing the model multiple times, and
it may not be suitable for some types of data, such as time-series or spatial data, due to the order or location of the observations.

10.

Hyper parameter tuning is the process of finding the best values for a machine learning model's hyper parameters. Hyper parameters are settings that control how a model learns. The goal of hyper parameter tuning is to find a combination of optimal hyper parameters that maximises model performance.The choice of hyper parameters can have a big impact on the performance of a machine learning model.

11.
With a large learning rate, the algorithm may take steps that are too large, causing it to overshoot the minimum of the loss function. This can prevent convergence to the optimal solution and result in instability.This instability makes it difficult to train the model effectively. A large learning rate can make it challenging to fine-tune the model parameters precisely, as the optimisation process may jump around in the parameter space without settling on a good solution.Models trained with large learning rates may not generalise well to unseen data, as they might not capture the underlying patterns effectively due to the erratic optimisation behaviour.

12.
Logistic regression is a linear classification algorithm.If the relationship between the features and the target variable is non-linear, logistic regression may not perform well because it cannot capture complex non-linear patterns in the data. In such cases, more advanced techniques like decision trees, random forests, support vector machines, or neural networks might be more appropriate.

13.
AdaBoost and Gradient Boosting are both ensemble techniques that combine weak learners to create a strong learner.

However, they have some key differences:

Base estimator: Gradient Boosting has a fixed base estimator, while AdaBoost allows the base estimator to be changed.

Loss function: AdaBoost is a special case with a particular loss function, while Gradient Boosting is a generic algorithm.

Flexibility: Gradient Boosting is more flexible than AdaBoost.

Speed: AdaBoost is generally slower than Gradient Boosting.

Missing values: Gradient Boosting and AdaBoost require explicit imputation of missing values, while XGBoost has built-in functionality to handle missing values.

Under-fitted values: AdaBoost tweaks the instance weights at every interaction, while Gradient Boosting tries to fit the new predictor to the residual errors made by the previous predictor.

14.

The bias-variance tradeoff in machine learning refers to the balance between two types of errors that a model can make: bias and variance.It describes the relationship between a model's complexity, the accuracy of its predictions, and how well it can make predictions on previously unseen data.

15.

Linear Kernel: The linear kernel is the simplest kernel function, representing a linear decision boundary between classes. It calculates the dot product between the input data points and is suitable for linearly separable datasets.

Radial Basis Function (RBF) Kernel: The RBF kernel is a popular choice for SVMs as it can handle non-linear decision boundaries. It measures the similarity between data points based on the Gaussian (radial basis) function, allowing SVMs to model complex relationships between features.

Polynomial Kernel: The polynomial kernel is another option for capturing non-linear relationships in data. It computes the similarity between data points using polynomial functions, allowing SVMs to classify data that is not linearly separable by projecting it into higher-dimensional feature space.