Q1)
R-squared, also known as the coefficient of determination, measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It provides an indication of how well the independent variables explain the variability of the dependent variable.
RSS measures the total sum of the squared differences between the actual values and the predicted values. It gives an overall measure of the discrepancy between the data and the estimation model.
It depends on what aspect of the model's performance you are interested in. R-squared is often used to assess the overall goodness of fit of the model.

Q2)
1. Total Sum of Squares (TSS): TSS measures the total variance in the dependent variable (Y) and represents the total differences between each observed value and the mean of the dependent variable. It's a measure of the total variability in the response variable without considering the explanatory variables.

2. Explained Sum of Squares (ESS): ESS measures the variance in the dependent variable that is explained by the independent variables included in the model. In other words, it represents the variability in the dependent variable that is accounted for by the regression model.

3. Residual Sum of Squares (RSS): RSS measures the unexplained variance in the dependent variable, which is essentially the sum of the squared differences between the observed values of the dependent variable and the values predicted by the regression
TSS = ESS + RSS

Q3)
Regularization in machine learning is used to prevent overfitting, which occurs when a model learns to perform well on the training data but fails to generalize to new, unseen data. Regularization techniques, such as L1 and L2 regularization, help to control the complexity of a model by adding a penalty for large coefficients, thus discouraging overfitting and promoting better generalization to new data. This is crucial for ensuring that the model's performance extends beyond the training set.

Q4)
The Gini impurity index is a measure of how often a randomly chosen element from a set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset. In the context of decision trees, it is used to evaluate how pure a set of elements are in terms of their classification into various categories. The Gini impurity is minimized as the purity of the nodes increases during the construction of a decision tree.

Q5)

Yes, unregularized decision trees are prone to overfitting. This is because decision trees are capable of learning very complex relationships within the training data, including noise and irrelevant patterns. Without regularization, decision trees can grow deep and complex, effectively memorizing the training data, which can lead to poor generalization on unseen data. Regularization techniques such as pruning or limiting the maximum depth of the tree are often used to address this issue

Q6)
Ensemble techniques in machine learning involve constructing multiple models and combining them to produce improved results. These techniques leverage the "wisdom of the crowd" principle, where multiple models, when combined, can often outperform any individual model.

Q7)
- Bagging (Bootstrap Aggregating): In bagging, multiple instances of the same learning algorithm are trained on different subsets of the training data, usually created by bootstrapping. The final prediction is then determined by averaging or taking a vote from the predictions of all the models.

- Boosting: Boosting focuses on training a sequence of models where each model corrects the errors of its predecessor. It assigns higher weights to instances that were misclassified in the previous round, thereby emphasizing the more difficult instances.

Q8)
The out-of-bag (OOB) error is an error measure that is calculated by testing each individual tree in the ensemble on the samples that were not used in its training. Since each tree is not trained on the entire dataset, the OOB error provides an unbiased estimate of the model's performance without the need for a separate validation set. This makes it a convenient tool for assessing the Random Forest's accuracy and generalization capability.

Q9)
K-fold cross-validation is a technique used to assess the performance of a predictive model. The dataset is divided into K subsets, and the model is trained and tested K times. In each iteration, one of the K subsets is used as the test set and the remaining K-1 subsets are used as the training set. This process allows for all data points to be used for both training and validation, and the performance metric is then averaged over the K iterations to provide a more reliable estimate of the model's performance compared to a single train/test split.

Q10)
Hyperparameter tuning in machine learning involves the process of finding the optimal hyperparameters for a model. Hyperparameters are configuration settings external to the model and are not learned from the data. They control aspects such

as model complexity, learning rate, regularization strength, and so on.

Hyperparameter tuning is crucial because the right set of hyperparameters can significantly impact a model's performance. By systematically searching through different combinations of hyperparameters, machine learning practitioners can find the optimal configuration that yields the best performance on unseen data. This process is typically done using techniques such as grid search, random search, or more advanced methods like Bayesian optimization or evolutionary algorithms.

Q11)
A large learning rate in Gradient Descent can lead to overshooting the minimum, causing the algorithm to diverge rather than converge. It may result in consistently oscillating or unstable updates, hindering the model's ability to reach an optimal solution. Fine-tuning becomes challenging, and the algorithm may fail to converge or take significantly longer to do so.

Q13)
1. Loss Function:
   - Adaboost: Focuses on adjusting the weights of incorrectly classified instances. It assigns higher weights to misclassified samples, allowing subsequent weak learners to concentrate on those instances.
   - Gradient Boosting: Minimizes a loss function (e.g., mean squared error for regression problems) by fitting each weak learner to the residual errors of the combined model.

2. Weighting of Weak Learners:
   - Adaboost: Adjusts the weights of weak learners based on their performance in the previous iteration, giving more influence to those with lower error.
   - Gradient Boosting: Iteratively fits weak learners to the residuals of the combined model, with each new learner addressing the errors of the ensemble.

3. Sequential vs. Parallel:
   - Adaboost: Trains weak learners sequentially, with each new learner focusing on the mistakes of the previous ones.
   - Gradient Boosting: Can be trained sequentially or in parallel, as each weak learner is fitted to the residuals independently.

Q14)
The bias-variance tradeoff is a fundamental concept in machine learning that deals with finding the right balance between model complexity and generalization performance.

- Bias: It represents the error introduced by approximating a real-world problem, which may be complex, by a simplified model. High bias can lead to underfitting,

where the model is too simple to capture the underlying patterns in the data.

- Variance: It measures the model's sensitivity to fluctuations in the training data. A high-variance model is sensitive to noise and can capture random variations in the training data, leading to overfitting.

Finding the right tradeoff is crucial because:
- High Bias: Results in a simplistic model that may not capture the complexities of the data, leading to systematic errors.
- High Variance: Captures noise in the training data and performs well on the training set but fails to generalize to new, unseen data.

Q15)
Linear Kernel:
- Description:The linear kernel computes the inner product between data points in the original feature space. It is suitable for linearly separable data and works well when the decision boundary is expected to be a straight line.
- Use Case:Effective when the relationship between features and the target is approximately linear.

RBF (Radial Basis Function) Kernel:
- Description: The RBF kernel measures the similarity between data points based on their Euclidean distance in the transformed space. It is versatile and can handle non-linear decision boundaries, making it popular for various applications.
- Use Case: Suitable for scenarios where the decision boundary is complex and non-linear.

Polynomial Kernel:
- Description: The polynomial kernel computes the similarity between data points using a polynomial function of the original features. It allows SVMs to capture non-linear relationships, and the degree of the polynomial can be adjusted to control the complexity of the decision boundary.
- Use Case: Useful when the decision boundary is expected to have a polynomial shape.