1. R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit model in regression and why?

Ans- R-squared is generally a better measure of the goodness of fit for a regression model than the residual sum of squares (RSS). , is a statistical measure that represents the proportion of the variance for the dependent variable that's explained by the independent variables in the model

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 (Total Sum of Squares)- **TSS** (total sum of squares) is equal to **ESS** (explained sum of squares) plus **RSS** (residual sum of squares), we need to start with the definitions of these terms and then use some algebraic manipulations to arrive at the desired result.

ESS (Explained Sum of Squares) – T he residual sum of squares essentially measures the variation of modeling errors. In other words, it depicts how the variation in the dependent variable in a regression model cannot be explained by the model. Generally, a lower residual sum of squares indicates that the regression model can better explain the data, while a higher residual sum of squares indicates that the model poorly explains the data.

SSE = $\sum(\hat{Y}i - \bar{Y})^2$

RSS (Residual Sum of Squares)- The regression sum of squares describes how well a regression model represents the modeled data. A higher regression sum of squares indicates that the model does not fit the data well.

The formula for calculating the regression sum of squares is

RSS = $\sum ni=1\ (yi - f(xi))2$

3 . What is the need of regularization in machine learning?

Ans- Regularization in machine learning serves as a method to forestall a model from overfitting. Overfitting transpires when a model not only discerns the inherent pattern within the training data but also incorporates the noise, potentially leading to subpar performance on fresh, unobserved data

4. What is Gini–impurity index?

Ans- Gini Impurity measures how well does a node splits the data set between the two outcomes. It aims to reduce the impurity score from the root node of the tree to the leaf node.

5. Are unregularized decision-trees prone to overfitting? If yes, why

Ans -Yes, Overfitting happens when any learning processing overly optimizes training set error at the cost test error. While it’s possible for training and testing to perform equality well in cross validation, it could be as the result of the data being very close in characteristics, which may not be a huge problem. In the case of decision tree’s they can learn a training set to a point of high granularity that makes them easily overfit. Allowing a decision tree to split to a granular degree, is the behavior of this model that makes it prone to learning every point extremely well — to the point of perfect classification — ie: overfitting.

6. What is an ensemble technique in machine learning?

Ans- Ensemble learning refers to a machine learning approach where several models are trained to address a common problem, and their predictions are combined to enhance the overall performance.

7. What is the difference between Bagging and Boosting techniques?

Ans- Bagging and boosting are ensemble learning techniques. Bagging (Bootstrap Aggregating) reduces variance by averaging multiple models, while boosting reduces bias by combining weak learners sequentially to form a strong learner.

8. What is out-of-bag error in random forests?
Ans- Out-of-bag (OOB) error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging).

9. What is K-fold cross-validation?
Ans- ross-validation is a statistical method used to estimate the skill of machine learning models.

It is commonly used in applied machine learning to compare and select a model for a given predictive modeling problem because it is easy to understand, easy to implement, and results in skill estimates that generally have a lower bias than other methods.

In this tutorial, you will discover a gentle introduction to the k-fold cross-validation procedure for estimating the skill of machine learning models.

After completing this tutorial, you will know:

- That k-fold cross validation is a procedure used to estimate the skill of the model on new data.
- There are common tactics that you can use to select the value of k for your dataset.
- There are commonly used variations on cross-validation such as stratified and repeated that are available in scikit-learn.

10. What is hyper parameter tuning in machine learning and why it is done?

Ans- Hyperparameters directly control model structure, function, and performance. Hyperparameter tuning allows data scientists to tweak model performance for optimal results. This process is an essential part of machine learning, and choosing appropriate hyperparameter values is crucial for success.

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

Ans- If the learning rate is too high, the algorithm may overshoot the minimum, and if it is too low, the algorithm may take too long to converge. Overfitting: Gradient descent can overfit the training data if the model is too complex or the learning rate is too high.

12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?
Ans-Not, Logistic regression is simple and easy to implement, but it also has some drawbacks. One of them is that it assumes a linear relationship between the input features and the output. This means that it cannot capture the complexity and non-linearity of the data.

13. Differentiate between Adaboost and Gradient Boosting.
Ans- The most significant difference is that gradient boosting minimizes a loss function like MSE or log loss while AdaBoost focuses on instances with high error by adjusting their sample weights adaptively.

14. What is bias-variance trade off in machine learning?
Ans-T he bias-variance tradeoff is a fundamental concept in machine learning that describes the relationship between a model's accuracy, complexity, and how well it can predict unseen data. It's a tradeoff between a model's ability to represent data patterns accurately (low bias) and its susceptibility to changes in the training data (high variance).

15. Give short description each of Linear, RBF, Polynomial kernels used in SVM.
Ans- The most common SVM kernels are linear, good for straight-line data, polynomial, and useful for curves. Radial basis function (RBF), is great for complex patterns. Also, sigmoid can handle different kinds of data changes