1. R-squared is better measure of goodness of fit model in regression. Because it measures how close the observed Y values are to the predicted values from the model.

2. The total sum of squares(TSS) is a variation of the values of a dependent variable from the sample mean of the dependent variable.

The explained sum of squares (ESS) is the sum of the squares of the deviations of the predicted values from the mean value of a response variable in a regression model

Residual sum of square or error is the difference between the observation of actual and predicted value.

TSS=ESS+RSS

3. Regularization is needed in machine learning to overcome the underfitting and overfitting of the data.

4. Gini Index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.

5. Yes, decision trees are prone to overfitting, especially when a tree is deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. This small sample could lead to unsound conclusions.

6. A huge dataset can't learn by the single model so we use ensemble technique.

Ensemble technique is a collection of week learners or many decision tree model.

7. Bagging is learning the data parlelly and Boosting is learning the data sequentially.

Bagging	Boosting
Random Forest	Ada Boost
Extra trees	Gradient Boosting

8. Out-of-bag (OOB) error or out-of-bag estimate, is a method of measuring the prediction error of random forests.

9. K-fold cross validation is used majorly to check any underfitting and overfitting happened in the model by its score by changing the K value.

10. Hyper parameters are the knobs or settings that can be tuned before running a training job to control the behavior of an ML algorithm.

11. If the step size is too large, it can jump over, the minima we are trying to reach. This can lead to osculations around the minimum or in some cases to outright divergence.

12. No, we can’t use Logistic Regression for classification of Non-Linear Data. Because that data is not in a ordered and has more than two value. Logistic regression can be used if we have only two values like binary value(0,1), True or False.

13. AdaBoost is the first designed boosting algorithm with a particular loss function. Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem.

14. Bias-variance trade off is an error, underfitting and overfitting can happen in regression and classification.

15. Linear Kernel is used when the data is separated using a single Line. It is one of the most common kernels to be used.

When the data set is non-linear, it is recommended to use kernel functions such as Radial Basis Function.

The polynomial kernel is commonly used with support vector machines (SVMs) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models.