1. As compared to RSS, R2 is supposed to give good fit for the model. But R2 is also having its own cons as it doesn't change with the change in variables, so adjusted R2 is used for better prediction or fitting the model. RSS do not gives a scaled value so mostly R2 is taken into consideration.
2. TSS is the sum of squares of the difference between the actual values and their mean.
   ESS measures how much variation is there in the modelled values and this value is compared with TSS.
   RSS gives us the total square of the distance of actual points from the regression line. In short RSS is the difference between the sum of squares of actual and predicted value.
   (TSS=ESS+RSS)
3. Regularization can be defined as the techniques used to minimize he adjusted loss function and prevent the model from over fitting or under fitting. We can fit our models appropriately on a given set and hence minimize the errors by properly using the regularization techniques.
4. The Gini impurity is a number between 0-1, which indicates the likelihood of new, random data being misclassified if it were given a random class label according to the class distribution of the set. In simple words it calculates the probability of a specific feature that is classified incorrectly while selecting randomly.
5. Un-regularized decision trees are prone to over-fitting mostly when a tree is deep i.e., it will take into consideration probability of each feature and if we provide the tree with a max-depth then it will lead to biased model as it will take into consideration till the depth provided which may exclude certain important feature.
6. Ensemble methods in a machine learning as its name suggests ensembles or combines several base models in order to produce one optimal predictive model.
7. Bagging is basically a method of merging the same type of predictions or homogenous predictions whereas boosting is a method of merging different types of predictions or heterogeneous predictions. In bagging each model receives an equal weight, but in boosting models are weighed as per their performance.
8. The OOB error is the average error of each training observations is calculated using the predictions from the trees that do not contain training observations in their respective sample. This allows the random forest classifier to be fit and validated whilst being trained.
9. K-fold cross validation is a technique where we have multiple train_test instead of 1. Suppose if k-value is 3, we can simply call this a 3-fold cross validation where k-value represents the number of groups' dataset is split into.
10. Hyper-parameters are the parameters that are defined to control the learning process before applying a machine-learning algorithm to a dataset. These are used to specify the learning capacity and complexity of the models.
11. If the learning rate for the Gradient Descent will be too high there is a probability of missing the optimal solution.
12. We cannot use the Logistic Regression of classification of non-linear data as the target label will not be having any linear correlation with the features as a result it can't predict targets with good accuracy.

13. Adaboost is an ensemble technique where a number of weak learners are combined together to form a strong learner. Adaboost increases the predictive accuracy by assigning weights to both observations at the end of every tree and weights to every classifier.
    Gradient Boosting just like AdaBoost is a technique which combines the weak learners to form a strong learner, but here the residual of the current classifier becomes the input for the next consecutive classifier on which trees are built.
14. It is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters. To build a good model, we need to find a good balance between the bias and variance such that it minimizes the errors.
    a. **Linear** – Linear Kernel is used when the data is linearly separable i.e., it can be separated using a single line. It is mostly used when there are large numbers of features in a particular dataset.
    b. **Polynomial** – In this technique the data is mapped into a high-dimensional space using a polynomial function. It is mostly used when the data is not linearly separable.
    c. **Radial Bias Function(RBF)** – It places a radial bias function centered at each point, then perform the linear manipulations to map points to higher-dimensional spaces that are easier to separate.