1. Residual Sum of Squares (RSS) is a better measure of goodness of fit model in regression because The residual sum of squares (RSS) is the absolute amount of explained variation, whereas R-squared is the absolute amount of variation as a proportion of total variation.
2. R^2 always takes on a value between 0 and 1. The closer R^2 is to 1, the better the estimated regression equation fits or explains the relationship between *X* and *Y*.

The expression

$\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ is also known as the *total sum of squares* (TSS).

This sum can be divided into the following two categories:

- Explained sum of squares (ESS**):** Also known as the *explained variation*, the ESS is the portion of total variation that measures how well the regression equation explains the relationship between *X* and *Y*.

  The Expression

$$ESS = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

- Residual sum of squares (RSS): This expression is also known as unexplained variation and is the portion of total variation that measures discrepancies (errors) between the actual values of Y and those estimated by the regression equation.

$$RSS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

3. Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.
4. Gini Index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen. But what is actually meant by 'impurity'? If all the elements belong to a single class, then it can be called pure.
5. Yes Decision trees are prone to overfitting especially when a tree is particularly deep. This is amount of specificity we look at leading to smaller sample. This is disadvantage of decision tree.
6. Ensemble learning combines the predictions from multiple models to reduce the variance of predictions and reduce generalization error. Techniques of ensemble learning can be grouped by the elements that is varied such as training data, the model, and how predictions are combined.

   Type of ensemble methods:

Bagging and boosting.

7. **Bagging**: It is a homogeneous weak learners' model that learns from each other independently in parallel and combines them for determining the model average.
   Boosting: It is also a homogeneous weak learners' model but works differently from Bagging. In this model, learners learn sequentially and adaptively to improve model predictions of a learning algorithm.

8. The out-of-bag error is the average error for each predicted outcome calculated using predictions from the trees that do not contain that data point in their respective bootstrap sample
9. k-Fold Cross-Validation Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into.
10. Hyperparameters are the knobs or settings that can be tuned before running a training job to control the behavior of an ML algorithm. They can have a big impact on model training as it relates to training time, infrastructure resource requirements (and as a result cost), model convergence and model accuracy.
11. Gradient Descent is an optimization algorithm used for minimizing the cost function in various machine learning algorithms. It is basically used for updating the parameters of the learning model. When the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error. […] When the learning rate is too small, training is not only slower, but may become permanently stuck with high training error.
12. Logistic Regression has traditionally been used as a linear classifier, i.e. when the classes can be separated in the feature space by linear boundaries. That can be remedied however if we happen to have a better idea as to the shape of the decision boundary. Logistic regression has traditionally been used to come up with a hyperplane that separates the feature space into classes. But if we suspect that the decision boundary is nonlinear we may get better results by attempting some nonlinear functional forms for the logit function.
13. AdaBoost is the first designed boosting algorithm with a particular loss function. On the other hand, Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem.
14. In statistics and machine learning, the bias–variance tradeoff is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in estimated parameter.
15. The linear, polynomial and RBF or Gaussian kernel are simply different in case of making the hyperplane decision boundary between the classes.The kernel functions are used to map the original dataset (linear/nonlinear) into a higher dimensional space with view to making it linear dataset.
    Usually linear and polynomial kernels are less time consuming and provides less accuracy than the rbf or Gaussian kernels.