1. 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.If strength and direction of a linear relationship should be presented, then r is the correct statistic. If the proportion of explained variance should be presented, then r^2 is the correct statistic.
2. The total sum of squares (TSS) measures how much variation there is in the observed data, The sum of squares measures the deviation of data points away from the mean value. A higher sum of squares indicates higher variability while a lower result indicates low variability from the mean. residual sum of squares measures the variation in the error between the observed data and modeled values.R2 = MSS/TSS = (TSS − RSS)/TSS
3. Regularization techniques in machine learning is done to minimize overfitting by nullifying the excessive data points that are arisen in the data
4. Gini Index, also known as Gini impurity, calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly
5. Decision trees are very prone to overfitting. This usually happens when the decision trees go very deep and then start intermixing the assumptions that were already made previously. This thereby harms the analysis and leads to overfitting. There are many techniques that can be used to avoid overfitting.
6. Ensemble techniques in machine learning are nothing but techniques that improve the results of the outcome by analyzing it with multiple methods and not just a single one.
7. Bagging technique merges the same type of predictions,whereas Boosting is a method of merging different types of predictions. Bagging decreases variance, not bias, and solves over-fitting issues in a model. Boosting decreases bias, not variance.
8. The out-of-bag (OOB) error is the average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample.
9. K-fold Cross-Validation is when the dataset is split into a K number of folds and is used to evaluate the model's ability when given new data.
10. Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set.
11. In order for Gradient Descent to work, we must set the learning rate to an appropriate value. This parameter determines how fast or slow we will move towards the optimal weights. If the learning rate is very large we will skip the optimal solution.
12. Logistic regression is used for linear data because it is better at analyzing the data that is along a single plane, non linear data is harder to analyze for logistic regression.
13.

AdaBoost	GradientBoost
Both AdaBoost and Gradient Boost use a base weak learner and they try to boost the performance of a weak learner by iteratively shifting the focus towards problematic observations that were difficult to predict. At the end, a strong learner is formed by addition (or weighted addition) of the weak learners.	
In AdaBoost, shift is done by up-weighting observations that were misclassified before.	Gradient boost identifies difficult observations by large residuals computed in the previous iterations.
In AdaBoost "shortcomings" are identified by high-weight data points.	In Gradientboost "shortcomings" are identified by gradients.
Exponential loss of AdaBoost gives more weights for those samples fitted worse.	Gradient boost further dissect error components to bring in more explanation.
AdaBoost is considered as a special case of Gradient boost in terms of loss function, in which exponential losses.	Concepts of gradients are more general in nature.

14. 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 the estimated parameters.
15. The linear kernel is typically used on data sets with large amounts of features as increasing the dimensionality on these data set does not necessarily improve separability
    RBF kernel is a popular kernel function used in various kernelized learning algorithms.
    the polynomial kernel is a kernel function 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.