Machine Learning-5

1. R-squared (R2) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable. The residual sum of squares (RSS) is a statistical technique used to measure the amount of variance in a data set that is not explained by a regression model itself.
2. TSS measures how much variation there is in the observed data.  ESS is the sum of the squares of the deviations of the predicted values from the mean value of a response variable. RSS is a statistical technique used to measure the amount of variance in a data set that is not explained by a regression model itself.
3. It is needed to prevent over fitting or under fitting.
4. Gini Impurity is a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree.
5. Yes unregularized decision-trees prone to overfitting.
6. In machine learning a number of models are used for making predictions about each data point. The predictions made by different models are taken as separate votes. Subsequently, the prediction made by most models is treated as the ultimate prediction.
7. Bagging is a technique for reducing prediction variance by producing additional data for training from a dataset by combining repetitions with combinations to create multi-sets of the original data. Boosting is an iterative strategy for adjusting an observation's weight based on the previous classification.
8. The average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample.
9. It 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. Hyper parameter tuning consists of finding a set of optimal hyper parameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyper parameters maximizes the model’s performance, minimizing a predefined loss function to produce better results with fewer errors.
11. If the learning rate is too large it can cause the model to converge too quickly to a suboptimal solution,
12. No we cannot use.
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. This makes Gradient Boosting more flexible than AdaBoost.
14. 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 the estimated parameters.
15. Linear algorithm creates a line or a hyperplane which separates the data into classes. RBF Kernel is popular because of its similarity to K-Nearest Neighborhood Algorithm. It has the advantages of K-NN and overcomes the space complexity problem as RBF Kernel Support Vector Machines just needs to store the support vectors during training and not the entire dataset.

Polynomial Kernnels is a kernel function commonly used with support vector machines SVMs, that represents the similarity of training samples in a feature space over polynomials of the original variables, allowing learning of non-linear models.