Machine Learning

2. TSS = ESS+RSS
The total sum of squares is a variation of the values of a dependent variable from the mean of dependent variable.
The residual sum of squares is used to decide if a statistical model is a good fit for your data.
The Explained sum of squares is the sum of squares of the deviation of the predicted values from the mean value of a response variable.

3. Regularization techniques is used to reduce the chances of overfitting in the regression models. Lasso and Ridge regression are the two algorithms used in regularization.

4. Gini Impurity index is a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree.

5. Decision trees are prone to overfitting, especially when a tree is particularly deep. This is because the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions.

6. Ensemble techniques creates multiples models and then combine them to produce improved results.
Bagging and boosting are the two techniques involved in ensemble techniques.

7. The bagging technique combines multiple models trained on different subsets od data, whereas boosting trains the model sequentially, focusing on the error made by the previous model. Bagging decreases variance but not bias, whereas boosting decreases bias not variance. Bagging tries to solve the overfitting problem, whereas boosting tries to reduce bias.

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 a technique for evaluating predictive models. The dataset is divided into k subsets or folds. The model is trained and evaluated k times.

10. When we are training the model, each dataset and model needs a different set of hyperparameters, which are a kind of variable. The only way to determine these is through multiple experiments, where you pick a set of hyperparameters and run them through the model. This is called hyperparameter tuning.

11. If the learning rate is too large, gradient descent can suffer from divergence. This means that weights increase exponentially, resulting in exploding gradients which can cause problem such as instabilities and overly high loss values.

13. Gradient boosting trains the learners and reduces the weak learners loss functions by training the model's residues. Adaboost focuses on training the prior miscalculated observations and alters the data distribution to improve sample weight values.

14. Variance-bias trade-off deals with the balance with the model bias and variance. It is the model's ability to accurately represents the underlying data patterns(low bias) and its susceptibility to fluctuations with the changes in the training data.