

1. R-squared is a better measure of goodness of fit model in regression because it provides a single number that summarizes the proportion of variance in the dependent variable that is explained by the model, which is more interpretable and easier to compare models.
2. TSS is the sum of squared differences between the observed dependent variables and the overall mean. ESS is the sum of differences between the predicted value and the mean of the dependent variable. RSS is used to measure the amount of variance in a data set that is not explained by a regression model. It is the sum of the squared values of the residuals.
 $TSS = ESS + RSS$.
3. While training a machine learning model, the model can easily be overfitted or under fitted. To avoid this, we use regularization in machine learning to properly fit a model onto our test set. Regularization techniques help reduce the chance of overfitting and help us get an optimal model.
4. Gini-impurity index measures how often a randomly chosen element of a set would be incorrectly labelled if it were labelled randomly and independently according to the distribution of labels in the set. It reaches its minimum (zero) when all cases in the node fall into a single target category.
5. Decision trees are a popular and powerful method for data mining, as they can handle both numerical and categorical data, and can easily interpret the results. However, decision trees can also suffer from overfitting, which means that they learn too much from the training data and fail to generalize well to new data.
6. Ensemble methods are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly.
7. Bagging is the simplest way of combining predictions that belong to the same type whereas Boosting is a way of combining predictions that belong to different types.
8. Out-of-bag error is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging). Bagging uses subsampling with replacement to create training samples for the model to learn from.
9. k-fold is a statistical method used to estimate the skill of machine learning models. It is commonly used in applied machine learning because it is easy to implement and skill estimates that have a lower bias than other methods.
10. Hyperparameters tuning is the process of selecting the optimal values for a machine learning model's hyperparameters. It is done because it allows to tweak model performance for optimal results. This process is essential part of machine learning.
11. Having large learning rate can lead to exploding or oscillating performance over the training epochs and to a lower final performance.

12. Logistics 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 computed with a specific loss function and becomes more rigid when comes to few iterations. But Gradient boosting assists in finding the proper solution to additional iteration modelling problems.
14. Bias-Variance Trade-off is a fundamental concept in machine learning that deals with the balance between model bias and variance. In simpler terms, it refers to the trade-off between a model's ability to accurately represent the underlying data patterns (low bias) and its susceptibility to fluctuations with changes in the training data (high variance).