

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. Because we feel that large negative residuals (i.e., points far below the line) are as bad as large positive ones (i.e., points that are high above the line). By squaring the residual values, we treat positive and negative discrepancies in the same way
2. The residual sum of squares is used to help you decide if a statistical model is a good fit for your data. It measures the overall difference between your data and the values predicted by your estimation model (a “residual” is a measure of the distance from a data point to a regression line). In ANOVA, Total SS is related to the total sum and explained sum with the following formula:
$$\text{Total SS} = \text{Explained SS} + \text{Residual Sum of Squares}.$$
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 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. Decision trees are prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions.
6. Ensemble Methods, what are they? Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model.
7. Bagging is a method of merging the same type of predictions. 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. Out-of-bag (OOB) error, also called out-of-bag estimate, 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. OOB error is the mean prediction error on each training sample x_i , using only the trees that did not have x_i in their bootstrap sample
9. 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. As such, the procedure is often called k -fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as $k=10$ becoming 10-fold cross-validation.
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. That

combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.

11. Gradient descent is an algorithm that numerically estimates where a function outputs its lowest values. That means it finds local minima, but not by setting $\nabla f = 0$ like we've seen before. Instead of finding minima by manipulating symbols, gradient descent approximates the solution with numbers. Furthermore, all it needs in order to run is a function's numerical output, no formula required.
12. Logistic regression is neither linear nor is it a classifier. The idea of a "decision boundary" has little to do with logistic regression, which is instead a direct probability estimation method that separates predictions from decision
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. 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 the estimated parameters.
15. Unlike linear or polynomial kernels, RBF is more complex and efficient at the same time that it can combine multiple polynomial kernels multiple times of different degrees to project the non-linearly separable data into higher dimensional space so that it can be separable using a hyperplane.