

MACHINE LEARNING

- A smaller or lower value for the RSS is ideal in any model since it means there's less variation in the data set. In other words, the lower the sum of squared residuals, the better the regression model is at explaining the data.
- **TSS** represents the total sum of squares. It is the squared values of the dependent variable to the sample mean. In other words, the total sum of squares measures the variation in a sample. **(ESS)** The explained sum of squares (ESS) is the sum of the squares of the deviations of the predicted values from the mean value of a response variable, in a standard regression model. Residual Sum of Squares **(RSS)** is a statistical method that helps identify the level of discrepancy in a dataset not predicted by a regression model.

Total sum of squares (TSS) = explained sum of squares (ESS) + residual sum of squares (RSS).

- 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.
- Gini Index, also known as Gini impurity, calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly. If all the elements are linked with a single class, then it can be called pure.
- 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.
- 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.
- **Bagging** also known as Bootstrap aggregation, is an ensemble learning method that looks for different ensemble learners by varying the training dataset. Unlike a single model trained on the entire dataset, bagging creates multiple weak learners or base models trained on a subset of the original dataset. The number of models to use and the size of the subsets is decided by the data scientist building the model. **Boosting** is an ensemble technique that looks to change the training data and adjust the weight of the observations based on the previous classification. Unlike the bagging approach, boosting involves dependence on weak learners. The weak learners take the results of the previous weak learner into account and adjust the weights of the data points, which converts the weak learner into a strong learner.
- 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.
- 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.
- 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.
- If the learning rate is very large, we will skip the optimal solution.
- Logistic regression is neither linear nor is it a classifier.
- 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.

- 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.
- 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.