

1. R Squared or RSS as a measure of goodness depends purely on what the output that we require through model performance. R squared measures the variabilities in the dependant variables, its range lies between 0 and 1. R square changes on the basis of number of independent variables leads to overfitting, whereas RSS measures the overall deviation from the predicted values. Thus I think RSS is better to measure of goodness.
2. TSS determines the total deviation of observed values of dependant variables from its mean value. ESS is the deviation from predicted values of dependant variables from the mean value of dependant variables. Whereas RSS is the difference between observed and predicted value of dependant variables. The relation between all these is  $TSS = ESS + RSS$
3. Regularization is used to control or prevent overfitting in the machine learning model.
4. Gini impurity is the measure of impurity in any data set, that is commonly used in decision tree model.
5. Yes, unregularized decision trees are prone to overfitting with complex datasets or in some machine learning data sets that too have some sort of noise features, or data with different kinds of ups and downs. Therefore, regularization helps in limiting the overall maximum depth of decision tree, regulates minimum sample selection.
6. Ensemble technique is used to compile multiple individual models together to make a single model in machine learning.
7. Bagging and Boosting both are ensemble technique. They differ themselves in the approach of combining different data sets. Bagging reduces the variance whereas boosting looks onto controlling the bias of data sets.
8. Out of bag error provides a clarity on how a random forest model will be able to perform a new or unseen data model without separately working on data sets.
9. K fold technique is usually to verify the performance of machine learning algorithm and helps in generating estimate om error in unseen data sets. It involves partitioning the overall data into various folds of equal sizes and then performing the evaluation.
10. Hyper parameter tuning is used to find the optimal values in machine learning models. These are set before the learning process and controls the learning process without being learned from data.
11. If there is a large learning rate it will cause hindrance to the convergence of data optimization and therefore it results in reduced efficiency of data learning. A larger learning rate will also lead to non stability in the data convergence.
- 12.
13. Both Adaboost and gradient boosting are ensemble technique that are used to combine various weak learners. Adaboost is otherwise called as adaptive boosting, that focus on adjusting the weights of various examples and weak learners to prioritize the non-categorized models whereas gradient boosting direct minimize the error in the data models.
14. Bias and variance are related, and the bias-variance tradeoff explains how this affects the model's overall performance. In general, a model's variance rises as its bias falls (i.e., as the model grows more complicated and flexible) and vice versa. Building a model that effectively generalizes to new, untested data requires striking the correct balance between bias and variance.
15. SVM stands for support vector machine