Machine Learning

1. R squared method is better than Residual Sum of Squares this is because in R square we are squaring the residuals. 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. Total Sum of Square (TSS)- The Total SS (TSS or SST) tells how much variation there is in the dependent variable. Total SS = Σ(Yi – mean of Y)2.

Residual Sum of Squares (RSS)- The residual sum of squares is used to help to decide if a statistical model is a good fit for the data. It measures the overall difference between the data and the values predicted by the estimation model (a "residual" is a measure of the distance from a data point to a regression line). Residual SS = Σ(Yi – f(Xi))2.

Explained Sum of Squares (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 the actual data in a standard regression model. Explained SS = Σ(Y-Hat – mean of Y)2.

3. Regularization helps in the following stages of machine learning: i) Providing a good accuracy in the model. ii) Helps in the problem of overfitting- A high-dimensional dataset having too many features can sometimes lead to overfitting. iii) Helps in underfitting- It prevents the loss of important data due to underfitting.

4. Gini Index is one of the most popular algorithm which is used by Decision Tree for selecting the best split. Gini Impurity measures the impurity of the nodes in a Decision Tree, hence lower the Gini impurity we can safely infer the purity will be more and a higher chance of the homogeneity of the nodes. Gini Impurity is always between 0 to 1.

   Gini Impurity=1- Gini

5. Yes, if decision trees are left unregularized will be prone overfitting. This is because it will grow which means each leaf node will represent one data point. In order to overcome this issue of overfitting, we should prune the tree. We can prune decision Tree by setting Max-depth of the tree or by setting minimum data points in each node

6. Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model.For example: Bagging or Bootstrap Aggregation and Random Forest Model.

7. Bagging attempts to tackle the over-fitting issue. Each model in bagging is trained parallelly and independently where in a final prediction is created from the prediction of every models.

Boosting tries to reduce bias. Boosting is an interative process which trains all the models together and gets a certain prediction, a second model is then built which tries to correct the errors present in the first model. This procedure is continued and models are added until either the complete training data set is predicted correctly or the maximum number of models is added.

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. This allows the RandomForestClassifier to be fit and validated whilst being trained

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.

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.In machine learning, a hyperparameter is a parameter whose value is used to control the learning process That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.
    Hyperparameter is required as it can give the optimized values for hyperparameters, which maximizes the model's predictive accuracy.

11. If the learning rate is very large then it will skip the optimal solution. A higher rate could result in a model that might not be able to predict anything accurately

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. Bias Variance Trade off is a tradeoff between a model’s ability to minimize bias and variance which is referred to as the best solution for selecting a value of regularization constant. Proper understanding of these errors would help to avoid the overfitting and underfitting of a data set while training the algorithm.

15. **Linear Kernel**: It is the most basic type of kernel, usually one dimensional in nature. It proves to be the best function when there are lots of features. The linear kernel is mostly preferred for

text-classification problems as most of these kinds of classification problems can be linearly separated.

Polynomial Kernel: It is a more generalized representation of the linear kernel. It is not as preferred as other kernel functions as it is less efficient and accurate.

Gaussian Radial Basis Formula- It is one of the most preferred and used kernel functions in svm. It is usually chosen for non-linear data. It helps to make proper separation when there is no prior knowledge of data.

Statistics

1. Mean
2. All of these
3. 6
4. Chisqared distribution
5. Binomial Distribution
6. Hypothesis
7. Null Hypothesis
8. Two tailed
9. Research Hypothesis
10. np.