MACHINE LEARNING

Submitted by Darshik A S

1)
2) TSS = ESS + RSS
   Total sum of squares (TSS): Coefficient of determination is used as a measure of how well a regression line explains the relationship between a dependent variable and an independent variable.
   Explained sum of squares (ESS): It is a quantity used in describing how well a model, often a regression model, represents the data being modelled.
   Residual sum of squares (RSS): It is the sum of the squares of residuals. It is a measure of the discrepancy between the data and an estimation model, such as a linear regression.
3) Regularisation is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.
4) Gini impurity index measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.
5) Yes, decision trees are prone to overfitting, especially when a tree is deep. This is due to the amount of specificity we look at leading to smaller samples of events that meet the previous assumptions.
6) Ensemble techniques create multiple models and then combine them to produce improved results.
7) Bagging is the result obtained by averaging the responses on N number of estimators. Boosting assigns a second set of weights for N estimators in order to take the weighted average of the estimates.
8) The out-of-bag error is the average error for each training observation calculated using predictions from the trees that do not contain in their respective bootstrap sample.
9) Cross validation is a technique to fit a model on a data set. In cross validation the data set is divided into ‘k’ number of sets where ‘k-1’ sets are used for training and 1 set is used as validation set. It produces a model of low bias prediction.
10) hyperparameter tuning or optimization is the problem of choosing a set of optimal hyperparameters for a learning algorithm. It is important because they directly control the behaviour of the training algorithm and have a significant impact on the performance of the model being trained.
11) When the learning rate is too large, gradient descent can increase rather than decrease the training error.
12) No. logistic regression only forms a linear decision surface and that is why it cannot be used to classify non-linear data.
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.
14) Bias is the simplifying assumptions made by the model to make the target function easier to approximate. Variance is the amount that the estimate of the target function will change given different training data. Trade-off is tension between the error introduced by the bias and the variance
15) Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line.
    Radial Basis Kernel is a kernel function that is used in machine learning to find a non-linear classifier or regression line.
    Polynomial kernel represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models.