MACHINE LEARNING

Assignment No.05

Answer No.01:- Well a smaller or lower value for the RSS is ideal in any model since it means there's less variation in the dataset.

In other words,the lower the sum of squared residuals,the better the regression model is at explaining the data.

Now let me tell you why RSS is a better measure of goodness of fit model in regression because it measures the level of variance in the error term,or residuals,of a regression model. The smaller the residual sum of squares,the better your model fits your data.the greater the residual sum of squares,the poorer your model fits your data.

Answer No.02:-TSS gives you the distance from the linear line drawn to each particular variable.You could also describe TSS as the dispersion of observed variables around the mean or the variance.

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.

(RSS) Residual sum of squares measures the level of variance in the error term,or residuals,of a regression model.

this equation relating these three metrics with each other

TSS = ESS + RSS

Answer No.03:- 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 apprppriately on a given test set and hence reduce the errors in it.

Answer No.04:- Gini index also known as Gini impurity index, it 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.

Answer No.05:- 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.

This small sample could lead to unsound conclusions.

Answer No.06:- 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.This has boosted the popularity of ensemble methods in machine learning.

Answer No.07:- Bagging is a technique for reducing prediction variance by producing additional data for training from a dataset by combining repetitions with combinations to create multi sets of the original data.

Boosting is an iterative strategy for adjusting an observations weight based on the previous classification. It attempts to increase the weight of an observation if it was erroneously categorized.Boosting creates good predictive models in general.

Answer No.08:- 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 Random Forest Classifier to be fit and validated whilst being trained.

Answer No.09:- 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.

Answer No.10:- Hyper parameter tuning consists of finding a set of optimal hyper parameter values for a learning algorithm while applying this optimized algorithm to any data set.

Hyper parameter tuning is an essential part of controlling the behaviour of a machine learning model.If we don't correctly tune our hyper parameters,our estimated model parameters produce suboptimal results,as they don't minimize the loss function.This means our model makes more errors.

Answer No.11:- In order for Gradient Descent to work,we must set the learning rate to an appropriate value.This parameter determines how fast or slow we will move towards the optimal weights.If the learning rate is very large we will skip the optimal solution.

A learning rate that is too large can cause the model to converge too quickly to a suboptimal solution,whereas a learning rate that is too small can cause the process to get stuck.

Answer No.12:- Logistic Regression has traditionally been used as a linear classifier,i.e. when the classes can be separated in the feature space by linear boundaries.

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.

Answer No.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.

AdaBoost or Adaptive Boosting is the first Boosting ensemble model.

Gradient Boost is a robust machine learning algorithm made up of Gradient descent and Boosting.

Answer No.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.

Answer No.15:- Linear SVM is used for linearly separable data,which means if a dataset can be classified into two classes by using a single straight line,then such data is termed as linearly separable data and classifier is used called as Linear SVM classifier.

In Machine learning the radial basis function kernel (RBF) is a popular kernel function used in various kernelized learning algorithms.In particular,it is commonly used in support vector machine classification.

In machine learning ,the polynomial kernel is a kernel function commonly used with support vector machine and other kernelized models,that represents the similarity of vectors in a feature space over polynomials of the original variables,allowing learning of non linear models.