Machine Learning Assignment - 5

1.

A residual sum of squares (RSS) is a statistical technique used to measure the amount of variance in a data set that is not explained by a regression model.
R Squared is statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable in a regression model.
R-Squared is the better measure of goodness of fit compared to RSS. R-squared explains to what extent the variance of one variable explains the variance of the second variable. So, if the R2 of a model is 0.50, then approximately half of the observed variation can be explained by the model's inputs.

2.

TSS (Total sum of squares): TSS is given by the summation of deviation of ground truth from the mean of the variable.

RSS (Residual sum of squares): RSS is given by the summation of squares of error values i.e., ground value – predicted value.

ESS (Explained sum of squares): ESS is given by the summation of squares of the deviation of the predicted value from the mean of the variable.

The equation between the above three is linearly expressed as :

TSS = RSS + ESS

3.

Regularization in the context of machine learning refers to a collection of strategies that help the machine learn more than solely memorize. Regularization constraints or shrinks the coefficient towards zero. This means that this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.

4.

Gini impurity index measures the probability of a particular variable to be wrongly classified when chosen randomly. This measures is calculated where the modeling contains Tree Algorithms like Decision Tress or random forest.

5.

Yes, decision tress are prone to overfitting especially when a tree is particularly deep. But unlike other algorithms decision tree does not use regularization to fight against overfitting. Instead it uses pruning. There are mainly to types of pruning performed:

Pre-pruning that stop growing the tree earlier, before it perfectly classifies the training set.

Post-pruning that allows the tree to perfectly classify the training set, and then post prune the tree.

6.

Ensemble techniques are the algorithms created combining multiple weak learners to a strong learning model.

7.

Bagging, which is also known as bootstrap aggregating sits on top of the majority voting principle. It increases the accuracy of models through decision trees, which reduces variance to a large extent. Boosting is an ensemble procedure to make a collection of predictors. In other words, we fit consecutive trees, usually random samples, and at each step, the objective is to solve net error from the prior trees.

8.

Out-of-bag 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

9.

K Fold cross validation is an evaluation method used in machine learning to find out how well your machine learning model can predict the outcome of unseen data. It is a method that is easy to comprehend, works well for a limited data sample and also offers an evaluation that is less biased, making it a popular choice. The K in K fold is the integer defining how many times does the subset should be created and trained and tested. For example a 5 Fold cross validation will create 5 subsets in both training and testing dataset, train and predicts output 5 accuracy values. Averaging those values would give us a greater idea of how good the model is.

10.

Hyper parameters are the parameters of the model algorithms which are to be tuned in order to get maximum accuracy from the machine learning model.

11.

When the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error.

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.

13.

Gradient boosting defies boosting as a numerical optimization problem where the objective is to minimize the loss function of the model by adding weak learners using gradient descent. Whereas, method focuses on training upon misclassified observations. Alters the distribution of the training dataset to increase weights on sample observations that are difficult to classify.

14.

There is a tradeoff between a model's ability to minimize bias and variance. Understanding these two types of error can help us diagnose model results and avoid the mistake of over- or under-fitting. This is known as bias-variance tradeoff.

15.

SVM also known as Support Vector Machine is a supervised machine learning algorithm which can be used for both classification and regression challenges. SVM uses different kernels for different types of questions.

A linear kernel allows to use linear functions, which are really impoverished. As the order of the polynomial kernel is increased, the size of the function class increases. In the polynomial kernel, we simply calculate the dot product by increasing the power of the kernel. RBF (Radial Basis Function) is another popular Kernel method used in SVM models whose value depends on the distance from the origin or from some point.