MACHINE LEARNING – WORKSHEET 5

1. R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit model in regression and why?

   R-SQUARE- This measures how successful the fit is in explaining the variation of the data. R-square is the square of the correlation between the response values and the predicted response values.

   The residual sum of squares (RSS) on the other and is the sum of the squared distances between the actual and predicted values:

2. What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sum of Squares) in regression. Also mention the equation relating these three metrics with each other.

   Total SS = Explained SS + Residual Sum of Squares.

   The Total SS (TSS or SST) quantifies the variation there is in the dependent variable.
   Total SS = Σ(Yi – mean of Y)2.

   The Explained SS explains how much of the variation in the dependent variable your model explained.
   Explained SS = Σ(Y-Hat – mean of Y)2.

   The residual sum of squares is the sum of the squared differences between the actual Y and the predicted Y:
   Residual Sum of Squares = Σ e2

3. What is the need of regularization in machine learning?

   Regularization is a technique that are used to minimize the adjusted loss function and prevent overfitting or underfitting. Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

4. What is Gini–impurity index?

   Gini Impurity is a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree.

5. Are unregularized decision-trees prone to overfitting? If yes, why?

   Decision Trees are a non-parametric supervised machine learning approach for classification and regression tasks. Overfitting is a common problem, a data scientist needs to handle while training decision tree models. Comparing to other machine learning algorithms, decision trees can easily overfit.

6. What is an ensemble technique in machine learning?

   Ensemble methods are techniques that create multiple models and then combine them to produce improved results. Ensemble methods usually produces more accurate solutions than a single model would.

7. What is the difference between Bagging and Boosting techniques?

   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 observation's weight based on the previous classification. Bagging works on homogenous data set where as boosting works on heterogenous data set

8. What is out-of-bag error in random forests?

   Out-of-bag error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating(bagging). Bagging uses subsampling with replacement to create training samples for the model to learn from.

9. What is K-fold cross-validation?

   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 data sample is split into ‘k’ number of smaller samples, hence the name: K-fold Cross Validation

10. What is hyper parameter tuning in machine learning and why it is done?

    In machine learning, hyperparameter optimization or tuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm. A hyperparameter is a parameter whose value is used to control the learning process. By contrast, the values of other parameters are learned.

11. What issues can occur if we have a large learning rate in Gradient Descent?

    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.

12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

    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. Differentiate between Adaboost and Gradient Boosting.

    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. What is bias-variance trade off in machine learning?

    The bias–variance decomposition is a way of analyzing a learning algorithm's expected generalization error with respect to a particular problem as a sum of three terms, the bias, variance, and a quantity called the irreducible error, resulting from noise in the problem itself.

15. Give short description each of Linear, RBF, Polynomial kernels used in SVM.

    RBF: 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.

    POLY: 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

    LINEAR: 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. Linear kernel functions are faster than other functions.