FLIP ROBO

ASSIGNMENT – 5

Machine Learning

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-Squared (R^2 or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, r-squared shows how well the data fit the regression model (the goodness of fit).
   R-squared can take any value between 0 to 1. Whereas The residual sum of squares (RSS) is the sum of the squared distances between your actual versus your predicted values. A residual sum of squares (RSS), also known as the sum of squared residuals (SSR), 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 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. 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.
   RSS : RSS or Residual sum of squares is given by the summation of squares of error values i.e., ground value – predicted value.
   ESS: ESS of Explained sum of squares is given by the summation of squares of the deiation of the predicted value from the mean of the variable.
   TSS: TSS or Total sum of squared is given by the summation of deviation of ground truth from the mean of the variable.

   The relation between the above 3 could be linearly expressed as :

   TSS = RSS + ESS

3. What is the need of regularization in machine learning?
   Regularization is a penalty faced by in case of regressions. 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. What is Gini–impurity index?

   Gini index or Gini impurity 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. Are unregularized decision-trees prone to overfitting? If yes, why?

   Yes, decision trees are prone to overfitting. 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. What is an ensemble technique in machine learning?

   Ensemble techniques are the algorithms created combining multiple weak learners to a strong learning model. Random Forest, XG Boosts, Gradient Boosting are some examples of ensemble learning techniques. These are 2 types of Ensemble techniques, Bagging and Boosting.

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

   Bagging, which is also known as bootstrap aggregating sits on top of the majority voting principle. Boosting is another 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. What is out-of-bag error in random forests?

   Out of sample is a technique to verify the performance of a boostrapping model with out having to use a validations set. This is an advantage if:
   Your data set is to small to split in to training,validation and test.
   Gives a second validation on the model allowing.

9. What is K-fold cross-validation?

   K Fold cross validation means training and testing with different subset of the training and testing data so that the model wont be biased over some cords in the dataset. The K in K fold is the integer defining how any times does the subset should be created and trained and tested. For example a 5 Fold cros validation will create 5 subsets in both training ad testing dataset, train and predict are output 5 accuracy values. Averagin those values would gives us a grater idea of how good the model is.

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

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

11. What issues can occur if we have a large learning rate in Gradient Descent?
    When the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error.

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.
    Gradient boosting defies boosting as a numerical optimisation problem where the objective is to minimise 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. What is bias-variance trade off in machine learning?
    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 knwn as bias-variance tradeoff.
15. Give short description each of Linear, RBF, Polynomial kernels used in SVM.
    SVM also known as Support Vector Machine is a supervised machine learning algorithm which can be used for both classification or regression challenges.SVM uses different kernels for different types of questions. A linear kernel allows you to use linear functions, which are really impoverished. As you increase the order of the polynomial kernel, the size of the function class increases. In the polynomial kernel, we simply calculate the dot product by increasing the power of the kernel. Gaussian RBF(Radial Basis Function) is another popular Kernel method used in SVM models for more. RBF kernel is a function whose value depends on the distance from the origin or from some point.