Machine Learning Assignment-5

Answers:

1. R-squared explain the proportion of variance in our data for example the proportion of variance in the observed data which the model explains or the reduction in error over the null model. The value of R-squared is lies between 0 to 1. The closer to 1 and better to fit.

   The residual sum of square (RSS) is sum of squared distance between actual vs predicted values.

   RSS = Σi=1 to n [(yi – y'i)^2]

   In RSS where yi is a give data and y'i is the fitted value for yi.
   Its mean the actual number we get depends largely on the scale of our response variable.
   So R2 is a better goodness of fit models in regression.

2. **RSS** = The residual sum of square (RSS) is sum of squared distance between actual vs predicted values.

   RSS = Σi=1 to n [(yi – y'i)^2]

   ESS = The explained sum of squares is the sum of squares of the difference between the predicted data and the mean data.

   ESS = total sum of squares – residual sum of squares

   ESS=Σi=1 to n [(y'i – ymean)^2]

   TSS: Total sum of square (TSS) = Explained sum of squares (ESS) + residual sum of squares(RSS).

   TSS = Σi=1 to n [(y'i – ymean)^2] + Σi=1 to n [(yi – y'i)^2]

   The relation between the all 3 is:

   TSS = RSS + ESS

3. **Need of Regularization:** Regularization is on of the key concepts in machine learning as it helps choose a simple model rather than a complex model one. We want out model to perform both on the train and the new unseen data, meaning the model must have the ability to be generalized. Regularization refers to the modification that can be made to a learning algorithm that help to reduce error not the training error. It's also prevent over-fitting and decrease the complexity of the model.

4. **Gini-Impurity Index:** Gini Index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly**.** If all the elements belong to a single class, then it can be called pure. The degree of Gini Index varies between 0 and 1. Gini Index, also known as Gini impurity, calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly.

5. **Are unregularized decision trees prone to overfitting:** Yes, unregularized decision tress prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specify we look at leading to smaller sample the previous assumptions.

6. Ensemble techniques are one of the machine learning approaches for improving the model by combining several models. The main advantage of ensemble learning is to reduce the variance and bias factors. It helps to increase the accuracy of the model and reduces the variability of prediction. Bagging and Boosting are two most used techniques in Machine learning.

7. Difference b/w the Bagging and Boosting:

    - Bagging - Bagging is the simplest way of combining predictions that belong to the same type. Each model receives equal weight.

    - Boosting – Boosting is a way of combining predictions that belong to different types. Models are weighted according to their performance.

8. Out of beg error in random forests: Out of beg error is also called the out of beg estimate, is a method of measuring the prediction error of random forests, boosted decision tree and other machine learning models utilizing bootstrap aggregating.

9. K-fold 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.

10. In machine learning, tuning or hyper parameter optimization is the difficulty of picking a collection of optimal parameters for a model learning algorithm. A hyper Parameter is also called a model predictor, since its value is used as a starting point for the model learning algorithm. Unlike the normal parameters, the numerical values of hyper parameters are learned as the training progresses. The choice of the hyper Parameter should be made carefully in order not to make model errors at the end of the training.

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.

12. No, logistic regression only forms linear decision surface. 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. 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. The bias–variance trade-off 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.

15. Kernels in SVM:

    - Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. It is mostly used when there are many Features in a particular Data Set.
    - 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. Gaussian Kernel is of the following format.
    - In the polynomial kernel, we simply calculate the dot product by increasing the power of the kernel.