ASSIGNMENT - 5

MACHINE LEARNING

Q1 to Q15 are subjective Answer

1. A smaller or lower value for the RSS is ideal in any model since it means there less variation in the data set. In other words the lower the sum of squared residuals the better the regression model is at explain the data.

2. **TSS:** This 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. So the goal of TSS is to measure the total variability of the dataset.

   ESS: 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 : The 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 itself. Instead, it estimates the Variance in the residuals.

   The equation relating these three metrics with each other:

   Total sum of squares(TSS) = explained sum of squares (ESS) + residual sum of squares (RSS).

3. Using Regularization we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

4. Gini impurity, 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.

5. Overfitting refers to the condition when the model completely fits the training data but fails to generalize the testing unseen data. Overfitting condition arises when the model memorizes the noise of the training data and fails to capture important patterns. A perfectly fit decision tree performs well for training data but performs poorly for unseen test data.

6. Ensemble methods are techniques that aim at improving the accuracy results in models by combining multiple models instead of using a single model.

7. Bagging aims to decrease variance, not bias while Boosting aims to decrease bias, not variance.

8. The 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.

9. K-fold Cross-Validation is when the dataset is split into a K number of folds and is used to evaluate The model's ability when given new data. K refers to the number of groups.

10. A hyper parameter is a parameter whose value is set before the machine learning process begins. In contrast, the values of other parameters are derived via training. Algorithm hyper parameters affect the speed and quality of the learning process.

11. 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. Logistic Regression has traditionally been used as a linear classifier, when the classes can be separated in the feature space by linear boundaries.

13. Ada Boost 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.

14. Machine learning 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. The RBF kernel function for two points computes the similarity or how close they are to each other. This kernel can be mathematically represented as follows.

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