Machine Learning Assignment-5 Answers

Q.1

Ans: - R-squared is a better measure of goodness of fit model in regression because it tells us how much of the variance in the dependent variable is explained by the independent variables. So, higher the r-squared value, better is the model fit.

Q.2

Ans: -

A) The TSS tells you how much variation there is in the dependent variable. TSS = Σ(Yi – mean of Y)2.

B) Explained sum of squares (ESS) is the sum of the differences between the predicted value and the mean of the dependent variable. Tt describes how well our line fits the data.
SSE=n∑i=1ε2i

C) The residual sum of squares (RSS) measures the level of variance in the error term, or residuals, of a regression model. The smaller the residual sum of squares, the better your model fits your data; the greater the residual sum of squares, the poorer your model fits your data.
RSS = $\sum_{i=1}^{n} (y_i - f(x_i))^2$

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

Q.3

Ans: - Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.
While training a machine learning model, the model can easily be overfitted or under fitted. To avoid this, we use regularization in machine learning to properly fit a model onto our test set. Regularization techniques help reduce the chance of overfitting and help us get an optimal model.

Q.4

Ans: - The Gini- Impurity Index measures the probability for a random instance being misclassified when chosen randomly. The lower the Gini Index, the better and lower the likelihood of misclassification.

Q.5

Ans: - Unregularized decision trees are prone to overfitting when they capture noise in the data. Pruning and setting appropriate stopping criteria are used to address this assumption.

1. Easy to interpret and make for straightforward visualizations.
2. The internal workings are capable of being observed and thus make it possible to reproduce work.
3. Can handle both numerical and categorical data.

4. Perform well on large datasets
5. Are extremely fast

Q.6

Ans: - Ensemble methods are techniques that create multiple models and then combine them to produce improved results. Ensemble methods in machine learning usually produce more accurate solutions than a single model would.

Q.7

Ans: - Bagging and boosting are different ensemble techniques that use multiple models to reduce error and optimize the model. The bagging technique combines multiple models trained on different subsets of data, whereas boosting trains the model sequentially, focusing on the error made by the previous model.

Q.8

Ans: - The out-of-bag error in random forest 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.

Q.9

Ans: - K-fold cross-validation is a technique for evaluating predictive models. The dataset is divided into k subsets or folds. The model is trained and evaluated k times, using a different fold as the validation set each time. Performance metrics from each fold are averaged to estimate the model's generalization performance.

Q.10

Ans: - Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.

Hyperparameters in Machine learning are those parameters that are explicitly defined by the user to control the learning process. These hyperparameters are used to improve the learning of the model, and their values are set before starting the learning process of the model.

Q.11

Ans: - If the learning rate is too high, the algorithm may overshoot the minimum, and if it is too low, the algorithm may take too long to converge. Overfitting: Gradient descent can overfit the training data if the model is too complex or the learning rate is too high.

Q.12

Ans: - No, we don’t use Logistic Regression for classification of Non-Linear Data, because Non-linear data could imply elements not associated with class separation, suggesting different sources of non-linearity.

Q.13

Ans: - Adaboost is computed with a specific loss function and becomes more rigid when comes to few iterations. But in gradient boosting, it assists in finding the proper solution to additional iteration modelling problem as it is built with some generic features.
And gradient boosting has the best model performance (Accuracy 0.839) when learning rate is 0.2, which is higher than the best performance of AdaBoost (Accuracy 0.825)

Q.14

Ans: - the bias–variance tradeoff describes the relationship between a model's complexity, the accuracy of its predictions, and how well it can make predictions on previously unseen data that were not used to train the model.

Q.15

Ans: - 1) RBF -The radial basis function kernel, or RBF kernel, is a popular kernel function used in various kernelized learning algorithms. In particular, it is commonly used in support vector machine classification.

2) Linear Kernel - It 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 a large number of Features in a particular Data Set.

3) Polynomial kernel - It is commonly used with support vector machines (SVMs) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models.