 ASSIGNMENT - 5

MACHINE LEARNING

**Q1 to Q15 are subjective answer type questions, Answer them briefly.**

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?

Ans. Typically, however, a smaller or lower value for the RSS is ideal in any model since it means there's less variation in the data set. In other words, the lower the sum of squared residuals, the better the regression model is at explaining the data.

1. 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.

Ans. The total sum of squares (TSS) measures how much variation there is in the observed data, while the residual sum of squares measures the variation in the error between the observed data and modeled values.

RSS is one of the types of the Sum of Squares (SS) – the rest two being the Total Sum of Squares (TSS) and Sum of Squares due to Regression (SSR) or Explained Sum of Squares (ESS). Sum of squares is a statistical measure through which the data dispersion is assessed to determine how well the data would fit the model in regression analysis.

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

Ans. Regularization refers to techniques that are used to calibrate machine learning models in order 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.

1. What is Gini–impurity index?

Ans. Gini Index, also known as 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.

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

Ans. Decision trees are prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. This small sample could lead to unsound conclusions.

1. What is an ensemble technique in machine learning?

Ans. Ensemble methods are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly. This has boosted the popularity of ensemble methods in machine learning.

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

Ans. Bagging is a method of merging the same type of predictions. Boosting is a method of merging different types of predictions. Bagging decreases variance, not bias, and solves over-fitting issues in a model. Boosting decreases bias, not variance.

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

Ans. The out-of-bag (OOB) 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 RandomForestClassifier to be fit and validated whilst being trained.

1. What is K-fold cross-validation?

Ans. 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 the data sample is split into. For example, if you see that the k-value is 5, we can call this a 5-fold cross-validation.

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

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.

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

Ans. 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.

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

Ans. Logistic regression is neither linear nor is it a classifier. The idea of a "decision boundary" has little to do with logistic regression, which is instead a direct probability estimation method that separates predictions from decision.

1. Differentiate between Adaboost and Gradient Boosting.

Ans. Adaboost and gradient boosting are types of ensemble techniques applied in machine learning to enhance the efficacy of week learners. The concept of boosting algorithm is to crack predictors successively, where every subsequent model tries to fix the flaws of its predecessor. Boosting combines many simple models into a single composite one. By attempting many simple techniques, the entire model becomes a strong one, and the combined simple models are called week learners. So the adaptive boosting and gradient boosting increases the efficacies of these simple model to bring out a massive performance in the machine learning algorithm.

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

Ans. In statistics and machine learning, the bias–variance tradeoff 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.

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

Ans. 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 a Large number of Features in a particular Data Set.

RBF Kernel is popular because of its similarity to K-Nearest Neighborhood Algorithm. It has the advantages of K-NN and overcomes the space complexity problem as RBF Kernel Support Vector Machines just needs to store the support vectors during training and not the entire dataset.

The polynomial kernel is often used in SVM classification problems where the data is not linearly separable. By mapping the data into a higher-dimensional space, the polynomial kernel can sometimes find a hyperplane that separates the classes.

