Machine Learning Assignment 5

QUES 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?

Answer. Residual Sum of Squares is a better measure of goodness of fit model in regression because 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, lower the sum of squared residuals, the better the regression model is at explaining the data.

QUES 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.

Answer. TSS(Total Sum of Squares**):** **Actual Value-Mean Value and square it.** The sum of squares is a statistical measure of deviation from the mean. It is also known as variation. It is calculated by adding together the squared differences of each data point. To determine the sum of squares, square the distance between each data point and the line of best fit, then add them together.

RSS(Residual Sum Of Squares): **RSS is calculated= Actual Value-Prediction value and square it.** It 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, or error term.

ESS(Explained Sum Of Squares): 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.

In some cases: Total sum of squares (TSS) = Explained sum of squares (ESS) + Residual sum of squares (RSS).

QUES 3. What is the need of regularization in machine learning?

Answer. Regularization is used in Machine Learning to minimize the adjusted loss function and prevent overfitting or underfitting. Regularization techniques help reduce the possibility of overfitting and help us obtain an optimal model.

QUES 4. What is Gini–impurity index?

Answer. 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.

QUES 5. Are unregularized decision-trees prone to overfitting? If yes, why?

Answer. 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.

QUES 6. What is an ensemble technique in machine learning?

Answer. 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.

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

Answer. **Bagging** is a technique for reducing prediction variance by producing additional data for training from a dataset by combining repetitions with combinations to create multi-sets of the original data.

Boosting is an iterative strategy for adjusting an observation's weight based on the previous classification.

QUES 8. What is out-of-bag error in random forests?

Answer. 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.

QUES 9. What is K-fold cross-validation?

Answer. 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.

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

Answer. 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.

QUES 11. What issues can occur if we have a large learning rate in Gradient Descent?

Answer. 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.

QUES 12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

Answer. 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. 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. That can be remedied however if we happen to have a better idea as to the shape of the decision boundary.

QUES 13. Differentiate between Adaboost and Gradient Boosting.

Answer. **Adaboost:** AdaBoost or Adaptive Boosting is the first Boosting ensemble model. The method automatically adjusts its parameters to the data based on the actual performance in the current iteration. Meaning, both the weights for re-weighting the data and the weights for the final aggregation are re-computed iteratively. In practice, this boosting technique is used with simple classification trees or stumps as base-learners, which resulted in improved performance compared to the classification by one tree or other single base-learner.

Gradient Boosting: Gradient Boost is a robust machine learning algorithm made up of Gradient descent and Boosting. The word ‘gradient’ implies that you can have two or more derivatives of the same function. Gradient Boosting has three main components: additive model, loss function and a weak learner. The technique yields a direct interpretation of boosting methods from the perspective of numerical optimization in a function space and generalizes them by allowing optimization of an arbitrary loss function.

QUES 14. What is bias-variance trade off in machine learning?

Answer. 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.

QUES 15. Give short description each of Linear, RBF, Polynomial kernels used in SVM.

Answer. **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: In machine learning, 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.

Polynomial Kernel: In machine learning, the polynomial kernel is a kernel function 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.