MACHINE LEARNING

ASSIGNMENT - 5

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?**

   The residual sum of squares (RSS) is the absolute amount of explained variation, whereas R-squared is the absolute amount of variation as a proportion of total variation.
   R^2 is a measure of the goodness of fit of a model. In regression, the R^2 coefficient of determination is a statistical measure of how well the regression predictions approximate the real data points. An R^2 of 1 indicates that the regression predictions perfectly fit the data.So R_Squared is a better measure of goodness of fit model in regression.

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.**

   TSS :- The Total SS (TSS or SST) tells you how much variation there is in the <u>dependent variable</u>.
   Total SS = $\Sigma(Yi - \text{mean of } Y)^2$.
   RSS - The residual sum of squares tells you how much of the dependent variable's variation your model did not explain**.** It is the sum of the squared differences between the actual Y and the predicted Y:
   Residual Sum of Squares = Σ e2
   ESS - The Explained SS tells you how much of the variation in the dependent variable your model explained.
   Explained SS = $\Sigma(\text{Y-Hat} - \text{mean of } Y)^2$.

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

4. **What is Gini–impurity index?**
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.

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

6. **What is an ensemble technique in machine learning?**
Ensemble methods are **techniques that create multiple models and then combine them to produce improved results**. Ensemble methods usually produces more accurate solutions than a single model would. This has been the case in a number of machine learning competitions, where the winning solutions used ensemble methods.

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

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

9. **What is K-fold cross-validation?**
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. **What is hyper parameter tuning in machine learning and why it is done**?
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.

11. **What issues can occur if we have a large learning rate in Gradient Descent**?
This is the hyperparameter that determines the steps the gradient descent algorithm takes. Gradient Descent is too sensitive to the learning rate. If it is too big, **the algorithm may bypass the local minimum and overshoot**.

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

13. **Differentiate between Adaboost and Gradient Boosting.**
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.

14. **What is bias-variance trade off in machine learning?**
The bias–variance decomposition is a way of analyzing a learning algorithm's expected generalization error with respect to a particular problem as a sum of three terms, the bias, variance, and a quantity called the irreducible error, resulting from noise in the problem itself.

15. **Give short description each of Linear, RBF, Polynomial kernels used in SVM**.
Linear Kernel:-
 It is the most basic type of kernel, usually one dimensional in nature. It proves to be the best function when there are lots of features. The linear kernel is mostly preferred for text-classification problems as most of these kinds of classification problems can be linearly separated.

Polynomial Kenerl:-
It is a more generalized representation of the linear kernel. It **is not** as preferred as other kernel functions as it is **less efficient** and accurate.

RBF:-
It is one of the most preferred and used kernel functions in svm. It is usually chosen for non-linear data. It helps to make proper separation when there is no prior knowledge of data