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?

A) R-Squared is standard goodness of fit for linear models and its formula incorporates RSS. R-squared is better as it avoids some Residual sum of squares (RSS) limitations, making it easier to interpret.

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.

A) Total Sum of Squares (TSS)- It is the sum of squared difference between the observed dependent variable and the overall mean.
TSS=$\sum$**(yi-mean of Y)**^{2}

B) Explained Sum of Squares (ESS)- It is sum of squares of the deviation of predicted values from the mean values of a response variable in standard regression model.
ESS=$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$

C) Residual Sum of Squares (RSS) – It is the sum of squares of residuals It measures the variation in error between observed data and modelled values .

TSS = ESS + RSS

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

A) It is a set of methods for reducing overfitting.

4. What is Gini–impurity index?

A) It assists the CART (Classification and Regression tree) algorithm in identifying the most suitable feature for node splitting during the construction of a decision tree classifier.

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

A) Unlike other regression models Decision-trees won't use regularization to overfitting, it uses others like tree pruning, limiting tree depth, minimum samples per leaf node, Feature selection and engineering, Ensemble methods, cross-validation and increasing training data.

6. What is an ensemble technique in machine learning?

A) Ensemble means collection of things, it combines several individual models to produce more accurate predictions than a single model alone.

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

A) Bagging and Boosting are two types of Ensemble learning, these two decreases the variance of a single estimate as they combine several estimates from different models which results to higher stability and avoid overfitting. Boosting creates a strong classifier by combining multiple weak

classifiers, this process involves building models sequentially, where it aims to correct errors made by previous ones.

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

A) The out of bag error is the method of measuring prediction error of random forests and other machine learning models utilizing bootstrap aggregating. Bagging uses subsampling with replacement to create training samples for the model to learn from.

9. What is K-fold cross-validation?

A) It splits the data into K subsets, each is held out in turn as a validation set.

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

A) It is a process of selecting the optimal values for a machine learning models hyperparameters. These are settings that control the learning process of the model such as learning rate. The goal of hyperparameter tuning is to find the values that lead to the best performance on a given task.

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

A) Choosing an appropriate learning rate is crucial for efficient and effective model training. Large learning rate can cause over shooting while too small can result in slow convergence. Learning rate influences the size of steps taken towards the minimum.

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

A) No. because logistic regression is a statistical model that models the log-odds of an event as a linear combination of one or more independent variables.

13. Differentiate between Adaboost and Gradient Boosting.

A) They are part of ensemble techniques applied in machine learning to enhance the efficacy of week learners.

Adaboost	Gradient Boosting
1. Learning happens by weights updation. Increase of misclassification, decrease of right classification.	Learning happens by optimizing the loss function.
2. Identifies complex observations by huge residual calculated in prior iterations.	The shift is made by up-weighting the observations that are miscalculated.
3. The trees with week learners are constructed using greedy algorithm based on split points and purity scores. The trees grow deeper with 8-30 two terminal nodes. The week learners should stay a week in terms of nodes, layers, leaf nodes and splits.	The trees are called decision stumps.
4. The classifiers are weighted precisely and their prediction capacity is constrained to learning rate and increasing accuracy.	Every classifier has different weight assumptions to its final prediction that depend on the performance.
5. It develops a tree with help of previous classifier residuals by capturing	It gives values to classifiers by observing determined variance with data. Here all the

	variances in data. The final prediction depends on the maximum vote of the week learners and is weighted by accuracy.	week learners possess equal weight and it is usually fixed as the rate for learning which is too minimum in magnitude.
6.	Here the gradiants themselves identify the shortcomings.	Maximum weighted data points are used to identify the shortcomings.
7.	Gradiant boosing cut down the error components to provide clear explanations and its concepts are easier to adapt and understand.	The exponential loss provides maximum weights for the samples which are fitted in worse conditions.
8.	This method trains the learners and depends on reducing loss functions of that week learner by training the residues of the model.	It focuses on training the prior miscalculated observations and it alters the distribution of the dataset to enhance the weight on sample values which are hard for classification.

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

A) It 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 model. It balances the bias error and variance error of a model.

15. Give short description on each of Linear, RBF, Polynomial kernels used in SVM (Support Vector Machine)

A) Linear kernel- It is used when data is linearly seperable.

RBF- Radial basis Function, it is more complex and efficient and can combine multiple polynomial kernels multiple times of different degrees to project the non-linearly separable data into higher dimensional space so that it can be separated using hyperplane.

Polynomial Kernel- It represents the similarity of vectors in training set of data in a feature space over polynomials of the original variables used in the kernel.