 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?

ANSWER -- The general equation for R-squared is RSE= sqrt(RSS/N-P-1)where n is degree of freedom

and p refer to number of predictors.

So in the case linear regression we have number of predictors(p) = 1 that leads to n-2 in the dominator and our equation becomes

RSE= sqrt(RSS/n-p-1).

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.

Answer--

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

Answer—Regularization refers to technique that are used to calibrate machine learning models in order to minimize

the adjusted loss function and prevent overfitting or underfitting.

1. . What is Gini–impurity index? .

Answer—Gini impurity is a measurement used to build Decision Trees to determine how the feature of a dataset should split nodes to from the tree . More precisely, the impurity of a dataset is a number between 0-0.5 . which indicates the likelihood of new, random data being misclassified if it were given a random classs label according to

The class disturbation in the dataset

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

Answer—Decision trees are a powerful method for data mining .as they can handel both numerical and categorical data. And can easily interpret the results. However,decision trees can also suffer from overfitting. Which means that they learn too much from the training data and fail to generaluize well to new data

6.What is an ensemble technique in machine learning?

Answer—Ensemble method are technique that aim to improving accuracy of result in models instead of using a single model. The combined model increase the accuracy of the reading significantly.

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

Answer-- IN bagging , models are trained independently parrakllel on different random subsets of the data.whereas in boosting models are trained sequentially with each models learning from the errors of the previous one.

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

Anmswer— Out –of-bag score is a performance metric for a mnachine learning models,specificically for ensemble models such as random forests. It is calculated using the samples that are not used in the training of the model,it is called out-of – bag sample.

9.What is K-fold cross-validation?

Answer—K- fold cross-validation is a technique for evalucating predictive models.The dataset is divided into k subset

Or folds as the validation set each times.performace metrics from each fold are average to estimate the models generalization performance.

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

Answer— Hyper parameter tuning is a critical to achieving optimal performance in machine learning models.

A poorly turned model can result in underfitting or overfitting ,which can lead to inaccurate predictons and poor performance on new data.

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

Answer— Gradient Descent can find the global minimum with ease but as ninconvex problems emerge gradient desent can struggle to the find the global minimum ,where the model achieve the best results.Recall that when the slope of the coist function is at or close to Zero ,the model stoips learning.

,

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

Answer-- NO, Difference Between Linear regression and Logistic Regression are linera regression handel Regression problem and Logistic regression is used to handle the classification problem ,Linear Regression provide a continoious output but Logistic regression provide discrete output.

13.Differentiate between Adaboost and Gradient Boosting.

Answer—Difference between adaboost and gradient boosting

Then Ada boosting bulid a new stumps based on the error that the previous stump made . Ada Boost continue to makes stumps we’ve asked for.IN contrast Gradient boost start by making a single leaf instead of tree or stump.

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

Answer-- .The bias- varience describes the relationship between a model”s complexity, the accuracy of its predictions,and how well it can make prediction on previsiously unseen data that were not used to train the models

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

Answer—IIn 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.

