Machine Learning Assignment

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:1)R-squared or Residual Sum of Squares (RSS) both are a better measure of goodness of fit model in regression.

Residual Sum of Squares(RSS):

1) The residual sum of squares (RSS) 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.
2)The smaller the residual sum of squares, the better your model fits your data; the greater the residual sum of squares, the poorer your model fits your data.
3)A value of zero means your model is a perfect fit.
4) RSS in short will determines how well the model explains or represents the data.

R-squared:

1) A statistical measure that determines the proportion of variance in the dependent variable that can be explained by the independent variable.
2)How well the data fit the regression model is explained by R-squared.
3)Formula : R-squared = Sum of squares due to regression / Total sum of squares.

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.

Ans: Total Sum of Squares (TSS):

->How much variance is there in the dependent variable.

TSS = Σ(Yi – mean of Y)2.

Explained Sum of Squares(ESS):

->The Explained SS tells you how much of the variation in the dependent variable your model explained.

Explained SS = Σ(Y-Hat – mean of Y)2.

Residual Sum of Squares(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:

RSS = (Actual y – Predicted Y)2

Equation relating these three metrics is given by :

TSS = ESS + RSS

where, TSS = Total Sum of Squares
ESS = Explained Sum of Squares
RSS = Residual Sum of Squares

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

Ans: 1)Regularization techniques help reduce the chance of overfitting and help us get an optimal model.

2) Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent over fitting or under fitting.

3) Using regularization, we can fit machine learning model appropriately on a given test set and hence reduce the errors in it.

4)Two Types of Regularization are there

i)L1 Regularization(LASSO)

ii)L2 Regularization(Ridge)

4) What is gini–impurity index?

Ans: Gini Impurity is a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree.

5) Are unregularized decision-trees prone to over-fitting? If yes, why?

Ans: Yes,they are prone to over-fitting. Decision trees are prone to over-fitting, 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?

Ans: 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.

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

Ans:

	Bagging	Boosting
1.	The simplest way of combining predictions that belong to the same type.	A way of combining predictions that belong to the different types.
2.	Aim to decrease variance, not bias.	Aim to decrease bias, not variance.
3.	Each model receives equal weight.	Models are weighted according to their performance.
4.	Each model is built independently.	New models are influenced by the performance of previously built models.
5.	Different training data subsets are selected using row sampling with replacement and random sampling methods from the entire training dataset.	Every new subset contains the elements that were misclassified by previous models.
6.	Bagging tries to solve the over-fitting problem.	Boosting tries to reduce bias.

7. If the classifier is unstable (high variance), then apply bagging.	If the classifier is stable and simple (high bias) the apply boosting.
In this base classifiers are trainedparallelly.	In this base classifiers are trained sequentially.
9 Example: The Random forest model uses Bagging.	Example: The AdaBoost uses Boosting techniques

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

Ans: The out-of-bag(OOB) error is the average error for each[zi]calculated using predictions from the trees that do not contain[zi]in their respective bootstrap sample. This allows the [Random Forest classifier]to be fit and validated whilst being trained.

Out-of-bag error is a method of measuring the prediction error of random forests.

9) 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. Each fold is used as a testing set at one point in the process.

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

Ans: 1)Hyper-parameter tuning is choosing a set of optimal hyperparameters for a learning algorithm.

2) A hyper-parameter is a model argument whose value is set before the learning process begins.

3) The key to machine learning algorithms is hyperparameter tuning.

4) Hyper-parameter tuning is done to increase the accuracy of the model.

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

Ans: A learning rate that is too large can cause the model to converge too quickly to a suboptimal solution.

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

Ans:

13. Differentiate between Adaboost and Gradient Boosting.

Ans:

Gradient boosting vs AdaBoost

Gradient boosting It identifies complex observations by huge residuals calculated in prior iterations.	Adaboosting The shift is made by upweighting they observation that all miss calculated prior
The trees with week learners are constructed using a greedy algorithm based on split points and purity scores. The trees are grown deeper with eight to thirty-two terminal nodes. The week learners should stay a week in terms of nodes, layers, leaf nodes, and splits. Gradient boosting The trees with week learners are constructed using a greedy algorithm based on split points and purity scores. The trees are grown deeper with eight to thirty-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
#1. Model #3. Classifier The classifiers are weighted precisely and their prediction capacity is constrained to learning rate and increasing accuracy.	AdaBoost Every classifier has different weight assumptions to its final prediction that depend on the performance. AdaBoost determined variance with equal weight the week learners and it is usually fixed as the rate for learning which is too minimum in magnitude.
Gradient boosting #2. Trees It develops a tree with help of previous classifier residuals by capturing variances in data. The final prediction depends on the maximum vote of the week learners and is weighted by its accuracy. Gradient boosting	#7. Applications AdaBoost It gives values to classifiers by observing . Here all AdaBoost Maximum weighted data points are used to identify the shortcomings. www.educba.com
#4. Prediction Here, the gradients themselves identify	AdaBoost The exponential loss provides maximum

shortcomings. Gradient boosting The shift is made by up-weighting the observations that are miscalculated prior.	weights for the samples which are fitted in worse conditions.
The trees are called decision stumps. This method trains the learners and depends on reducing the loss functions of that week learner by training the residues of the model. #6. Loss Value	Its focus 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?

Ans:1) If the algorithm is too simple (hypothesis with linear eg:) then it may be on high bias and low variance condition and thus is error-prone.
2)If algorithms fit too complex (hypothesis with high degree eg:.) then it may be on high variance and low bias.
3)In the latter condition, the new entries will not perform well.
4)Well, there is something between both of these conditions, known as Trade-off or Bias Variance Trade-off.
5)This trade-off in complexity is why there is a trade-off between bias and variance.
6)An algorithm can’t be more complex and less complex at the same time.

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

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

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. One of the examples where there are a lot of features, is Text Classification, as each alphabet is a new feature. So we mostly use Linear Kernel in Text Classification.

RBF is the default kernel used within the sklearn's SVM classification algorithm and can be described with the following formula: where gamma can be set manually and has to be >0.