MACHINE LEARNING ASSIGNMENT - 5

Q1. 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?

Q2. 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: **TSS** - The sum of squares total, denoted SST, is the squared differences between the observed dependent variable and its mean. You can think of this as the dispersion of the observed variables around the mean – much like the variance in descriptive statistics.

Formula:

$$\text{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

TSS= Total sum of squares

N= number of observations

Yi = value in a sample

$\bar{y}$ = mean value of a sample

ESS:

Explained sum of square (ESS) or **Regression sum of squares** or Model sum of squares is a statistical quantity used in modeling of a process. ESS gives an estimate of how well a model explains the observed data for the process. It tells how much of the variation between observed data and predicted data is being explained by the model proposed. Mathematically, it is the sum of the squares of the difference between the predicted data and mean data.

RSS:

The residual sum of squares (RSS) **measures the level of variance in the error term, or residuals, of a regression model**. 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.

Q3. What is the need of regularization in machine learning?

Ans: 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.

Q4. What is Gini–impurity index?

Ans: Gini Index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen. The formula of the Gini Index is as follows:

Gini=1−n∑i=1(pi)2====1−∑==1=(==)2
where,
'pi' is the probability of an object being classified to a particular class.

While building the decision tree, we would prefer to choose the attribute/feature with the least Gini Index as the root node.

Q5. Are unregularized decision-trees prone to overfitting? If yes, why?

Ans: Yes, the decision trees are prone to overfitting. It is 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.

Q6. What is an ensemble technique in machine learning?

Ans: 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.

Q7. What is the difference between Bagging and Boosting techniques?

Ans: 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. It attempts to increase the weight of an observation if it was erroneously categorized. Boosting creates good predictive models in general.

Q8. What is out-of-bag error in random forests?

Ans8.) 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 Random Forest Classifier to be fit and validated whilst being trained.

Q9. What is K-fold cross-validation?

Ans.9) Cross validation is an evaluation method used in machine learning to find out how well your machine learning model can predict the outcome of unseen data. It is a method that is easy to comprehend, works well for a limited data sample and also offers an evaluation that is less biased, making it a popular choice.The data sample is split into 'k' number of smaller samples, hence the name: K-fold Cross Validation. You may also hear terms like four fold cross validation, or ten fold cross validation, which essentially means that the sample data is being split into four or ten smaller samples respectively.

Q10. What is hyper parameter tuning in machine learning and why it is done?

Ans.10) A hyperparameter is a parameter of the model whose value influences the learning process and whose value cannot be estimated from the training data. Hyperparameters are configured externally before starting the model learning/training process. Hyperparameter tuning is the process of finding the optimal hyperparameters for any given machine learning algorithm.

Q11. What issues can occur if we have a large learning rate in Gradient Descent?

Ans: It a learning rate is too large than it can cause the model to converge too quickly to a suboptimal solution, whereas if a learning rate that is too small than it can cause the process to get stuck.

Q12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

Q13. Differentiate between Adaboost and Gradient Boosting.

Gradient Boosting	Adaboost
It identifies complex observations by huge residuals calculated in prior iterations	The shift is made by up weighting the observations that are miscalculated prior
The trees with 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 weak learners should stay a weak in terms of nodes, layers, leaf nodes and splits.	The trees are called decisions stumps
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.
The gradients themselves identify the shortcomings.	Maximum weighted data points are used to identify the shortcomings
It cut downs the error components to provide clear explanations and its concepts easier to adapt and understand.	The exponential loss provides maximum weights for the samples which are fitted in worse conditions.
The gradient boosting depends on the intuition which is the next suitable possible model, when get combined with prior models that minimize the cumulative predicted errors. The crucial idea of gradient boosting is to fix the targeted outcomes for the next model to reduce the error.	Adaboost increases the performance of all the available machine learning algorithms and it is used to deal with weak learners. It gains accuracy just above the arbitrary chances of classifying the problem. The adaptable and most used algorithm in AdaBoost is decision trees with a single level.

Q14. What is bias-variance trade off in machine learning?

Ans.14) The bias-variance trade off 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.

Q15. Give short description each of Linear, RBF, Polynomial kernels used in SVM

Ans. Linear: Linear Regression is the supervised Machine Learning model in which the model finds the best fit linear line between the independent and dependent variable i.e it finds the linear relationship between the dependent and independent variable.

RBF: Radial Basis Functions (RBF) are real-valued functions that use supervised machine learning (ML) to perform as a non-linear classifier. Its value depends on the distance between the input and a certain fixed point.

Polynomial kernels used in SVM: 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.