

## **MACHINE LEARNING ASSIGNMENT- 5**

Q1) Both R-squared or Residual Sum of Squares (RSS) is a better measure of goodness of fit model in regression because if the residual sum of squares results in a lower figure, it signifies that the regression model explains the data better than when the result is higher. R-squared is a statistical measure that represents the goodness of fit of a regression model. The ideal value for r-square is 1. The closer the value of r-square to 1, the better is the model fitted. R-square is a comparison of the residual sum of squares (SSres) with the total sum of squares (SStot).

Q2) TSS is the sum of the squared deviations of the predicted values from the observed mean of y, ESS is the explained sum of squares, and RSS is the residual sum of squares. EQUATION:  $TSS = ESS + RSS + 2 \sum_i (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$

Q3) Regularization is a technique to prevent the model from overfitting by adding extra information to it. Sometimes a machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be deal with the help of a regularization technique.

This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model.

It mainly regularizes or reduces the coefficient of features toward zero. In simple words, *"In regularization technique, we reduce the magnitude of the features by keeping the same number of features."*

Q4) Gini-impurity or Gini-index in machine learning is a metric to measure the randomness in a feature. It measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen. But what is actually meant by 'impurity'? If all the elements belong to a single class, then it can be called pure. The degree of Gini Index varies between 0 and 1.

Q5) In the case of decision tree's they can learn a training set to a point of high granularity that makes them easily overfit. Allowing a decision tree to split to a granular degree, is the behavior of this model that makes it prone to learning every point extremely well — to the point of perfect classification i.e., overfitting.

Q6) Ensemble techniques are one of the machine learning approaches for improving the model by combining several models. The main advantage of ensemble learning is to reduce the variance and bias factors. It helps to increase the accuracy of the model and reduces the variability of prediction.

Q7) Bagging mainly attempts to tackle the over-fitting issue. Boosting tries to reduce bias. If the classifier is unstable (high variance), then we need to apply bagging. If the classifier is steady and straightforward (high bias), then we need to apply boosting. Every model receives an equal weight. Models are weighted by their performance.

Q8) The out-of-bag error is the average error for each predicted outcome calculated using predictions from the trees that do not contain that data point in their respective bootstrap sample.

Q9) K-fold cross-validation is a method for estimating the performance of a model on unseen data. It 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.

Q10) 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. Parameters are the knobs or settings that can be tuned before running a training job to control the behavior of an ML algorithm because it needs to tune hyperparameters in machine learning to minimize the error.

Q11) If a large learning rate in Gradient Descent occur as seen as step size,  $\eta$ . As such, gradient descent is taking successive steps in the direction of the minimum. If the step size  $\eta$  is too large, it can (plausibly) "jump over" the minima trying to reach, i.e., it overshoots that. This can lead to oscillations around the minimum or in some cases to outright divergence.

Q12) Yes, it uses Logistic Regression for classification of Non-Linear Data. It is a statistical analysis method used to determine relationships between continuous variables and categorical variables.

Q13) 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.

Q14) The bias-variance tradeoff 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) SHORT DESCRIPTION: Linear Kernel: They are used when data is linearly separable. Radial Basis Function [RBF] Kernel: It uses trick actually refers to using efficient and less expensive ways to transform data into higher dimensions. Polynomial Kernel: It is a kernel function commonly used with support vector machines (SVMs) and other kernelized models.