MACHINE LEARNING – ASSIGNMENT 5

Ans 1. R-Squared and Residual sum of Squares (RSS) are commonly used to assess the goodness of fit of model in Regression. They both capture different direction of performance and the evaluation of the particular context.

R-squared-

1) R-Squared: - It is a measurement of the proportion of the variance in the dependant variable which is explained or summarized by the independent variable in the regression model.
2) R-Squared: - It ranges from zero to one. The higher value indicates the best fit for the model. Therefore, a value of 1 means that the model is perfectly explains the variance with the dependant variable, whereas 0 means that the model is giving no explanation.
3) It can be useful to compare different models.

Residual sum of Squares-

1) This measures the total squared difference between the observed values of the dependant variable and the predicted values from the regression model.
2) It measures the errors or residuals in regression model.
3) A lesser RSS indicates a better fit as it indicates the model's predictions that are closer to the actual values that are being observed.

Which measure to use-

RSS is more focussing on the magnitude of the prediction of errors and residuals. It helps in minimizing the predicted errors which is a primary concern in the regression model. And therefore, RSS is more appropriate choice.

Ans2. Sum of squares

It also indicates as SS, is a tool in statistics which is used to identify the dispersion of data and also how well the data can fit the model in the regression analysis. It is calculated by finding the sum of the squared differences in the analysis of regression. The most important outputs in regression analysis are the sum of squares. Sum of squares is the general rule that indicates a better model that shows a less variation in the data.

Type of sum of squares

There are three main types of sums of squares

a. Total sum of squares.
b. Explained sum of squares.
c. Residual sum of squares.

TSS, RSS, ESS are commonly used to calculate the performance and to evaluate the performance of a regression model.

For example- We have a data set of 5O individuals with their annual salary.

TSS (Total sum of squares)

It is the sum of squared differences between the actual salary and the mean salary. It therefore shows the total variation in the target variable.

RSS (Residual sum of squares)

It is the sum of the squared differences between the predicted salary and the actual salary of the data set. It therefore, shows the unexplained variation in the target variable which is not explained by the regression model.

ESS (Explained sum of squares)

It is the sum of the squared differences between the predicted salary and the mean salary. It shows the variation that is explained by the independent variables.

The relationship between TSS, RSS and ESS are

$$TSS = ESS + RSS$$

This shows, the total variation in the target variable can be explained by the independent variables ESS and RSS.

Ans3. Regularization

Regularization is known as one of the most important concepts of Machine Learning. This technique is used to prevent the model from overfitting and underfitting. It is used to prevent the model by adding extra information to it. Regularization gave the accuracy and generalization to the model.

Need of Regularization in Machine Learning

1) We use regularization in Machine Learning to properly fit the model to our set of tests.
2) It helps to reduce the possibility of overfitting.
3) It helps us to obtain an optical model.
4) It used to calibrate to minimize the loss function.
5) It is one of the key concepts in Machine Learning that helps to choose a simple model rather than a complex one.

Ans4. (a) Gini-Impurity index is a tool for measurement used to build Decision Trees to find how the features of a set of data must be split the nodes to form a tree.

(B) Gini-Impurity of a set of data is a number between 0-0.5, that shows the likelihood of the new and random data is being misclassified.

(c) It measures the impurity of the split and therefore the feature with the lowest impurity would determine the best feature for splitting the current node.

(d) The decision tree informed for accurate and robust predictions.

(e) The decision tree selects the optimal split by identifying the features that results in homogeneous subsets of data.

Ans 5. Decision tree is an approach for the classification and regression tasks in machine learning. Overfitting is one of the problems which describes if your model is no longer generalizes well. Overfitting indicated to the test error condition where the model completely fits the training data but fails to generalise the unseen data. This condition arises when the model fails to capture the important patterns. A perfectly fit decision tree performs well but performs poorly for unseen test data. The decision tree is allowed to train to its model.

In case of unregularized decision tree, they are not able to learn a training set to a point of high granularity. And therefore, this makes them easily overfitting. Allowing a decision tree to split, it indicates the behaviour of the model that makes it prone to the point of perfection.

Ans 6. Ensemble techniques

Ensemble techniques are the methods that are used to create different multiple models and then add them to make some improved results.

(a) Ensemble techniques in machine learning can produce more relevant solutions than that of a single model can give.
(b) In the Ensemble technique, machine learning professionals use a different number of models to make predictions about each data format.
(c) The predictions in each data format are made by different models and these predictions are treated as the ultimate predictions.

Therefore, Ensemble techniques are used to improve the performance of the machine learning models. Ensemble technique also results into a more stable model among the data set.

Ans 7. Bagging and Boosting are the two different types of ensemble techniques. These different techniques are used to decrease the amount of error and to optimize the model efficiently.

(a) Bagging technique adds multiple models of different subsets of data whereas Boosting technique focussed on the error or mistake made by the previous model.
(b) The Bagging technique is used to decrease the variance by averaging out individual error from the model. Whereas Boosting technique reduces both bias and variance by correcting the previous model.
(c) Bagging technique used boots trap to create the subsets of data. Whereas Boosting technique revised the error from the previous model and misclassified the next model.
(d) Bagging technique serves as equal weight in the most final decision whereas Boosting technique serves on the basis of accuracy, better accuracy and higher accuracy.
(e) Bagging technique has an equal error rate in each model, whereas in Boosting technique, it weights like higher error, less error.

Ans 8. The out-of-bag (OOB) error is also known as out-of-bag estimate. It is a procedure to measure the predicted error of the Machine Learning models. It is a method to measure the prediction error of random forests, decision trees and other Machine Learning models.

Random Forests

Random forests are an ensemble technique in machine learning. It is one of the best models used to supervised learning in decision trees, in which the model step makes decisions and predict the values. Therefore, Random Forests are also an ensemble technique learning method for classification and the regression problems. By constructing many decision trees in a random forest algorithm helps in generalizing the data pattern therefore reduce overfitting.

Therefore, out-of-bag (OOS) is a way of validating the Random Forest model.

Ans 9. K-fold cross validation is a method of technique to evaluate or to analyse the predictive model. The data set should be divided or separated into K-subsets or K number of folds. This model is trained and then evaluated K times, to use different folds for the validation of set in each time. Performance criteria for each subset or fold are averaged each time to calculate the generalization of the performance of the model. Thus, this method results in model assessment, the selection and giving a more accurate measure for the effectiveness of model. In this technique, test would be performed each time during the entire process. This thus helps us in avoiding overfitting. We have to split the data into 3 sets- Training, Testing and Validation. And thus, the model has been validating multiple times based on the value which is called K.

Ans 10. A machine learning model is expressed as a mathematical model who do have several different parameters that are need to be learned from the data. By training the model with the existing data, we can fit the model with the parameter.

So, Hyperparameter is another kind od parameter that cannot be easily learned from the regular training process.

Hyperparameter Tuning

It is the process of selecting the optimal values for a machine learning process model. These are the settings that are used to control the learning process of the model. The ultimate goal is to find the values that are leading to the best performance for the best optimal results. Hyperparameter turning has some advantages that helps in its existence are-

1. Hyperparameter turning improves the model performance.
2. It helps in reducing the overfitting and underfitting.
3. It helps in enhancing the generalization of model.
4. It helps in full optimization of resource utilization.
5. It helps in improving model interpretability.

Ans 11. The learning rate is an essential hyperparameter that efficiently affects the performance of Gradient Descent. It helps in determining how Fastly or how slowly our model learns and adapts and therefore plays an important role in managing and controlling both the convergence and divergence of the algorithms.

When it is said that the learning rate is too large, the gradient descent can suffer a lot from divergence in the algorithm. This reverts that when the weights increase exponentially, this in results into exploding gradients which can cause problems by having a high loss of values.

And therefore, in order to avoid these unbalanced issues with various learning rates, we can use the adaptive techniques which helps in adjusting their own learning rates through the real time observations of parameters. These measures make sure to have better results than the standard results of Gradient Descent by avoiding potential pitfalls in terms of high gains or slow losses.

Ans 12. Logistic Regression

Logistics Regression is a classification of algorithm that is used to find the probability of the event of success and the events of failure. It is used when the dependant variable is binary in nature i.e. 0 or 1, true or false. It helps in supporting by categorizing the data into different classes by studying the relationship for the labelled data set.

It is also known as Binomial Logistics regression. It is easier to be implement and efficient to be trained. It can easily extend into multiple classes and measure how appropriate a prediction is. Therefore, Logistics Regression used fixed variables to predict the results. But in Non-Linear data, it allows us to model the relationship between variables that do not have a clear linear relationship and therefore regression would not work effectively on these types of data.

Non-Linear Regression technique is used as a model to show the relationship between a dependant and independent variables. And therefore, it allows for more complex relationships.

Ans13. **Ada Boost and Gradient Boosting** are two different ensemble learning techniques. They add or combine multiple weak learners to form a strong model. Although, they are different in their personal approach. Therefore, the difference between Ada Boost and Gradient Boosting are as follows

(1) In Ada Boost, the weights of different samples are adjusted at each level of iteration. Whereas in Gradient Boosting, no reweighting of the samples takes place.
(2) In Ada Boost, shift is made by the observation that are miscalculated price. Whereas in Gradient Boosting, complex observations by large residuals calculated in price.
(3) In Ada Boost, the trees are known a decision stumps. Whereas, in Gradient Boosting the trees are constructed using algorithm based on split points and purity scores.
(4) Ada Boost gives values by observing variance with data. Whereas, Gradient Boosting depends totally on the maximum vote of the week learners by its accuracy.

Ans 14. Making a balance between accuracy and the ability to make predictions after the training data in a machine learning model is called bias-variance trade off.

Bias

The difference between the predictions of the values by the machine learning model and the accurate value is known as the bias. High in biasing can indicate to a large erroe in training as well as in data set. Therefore, it is recommended that an algorithm should be low-biased to avoid the issue

of underfitting. In case of high biasing, the data predicted does not fitted accurately in the data set. Therefore, such fitting is known as the underfitting of data in dataset.

Variance

The variability of prediction in a given data indicates the spread of our data is known as the variance of the model. With high variance in the model, shows a very complex fit and therefore not able to fit accurately. Therefore, such models show high error rates on the test data. High variance is known as the overfitting of data.

Bias Variance Trade off

It is said that if the algorithm is very simple, then it is on high bias and low variance and therefore it is error-prone condition. If the algorithm fit too complex then it is on high variance and low bias. Therefore, the entries will not perform wee. And therefore, there is something between these condition that is known as trade off or Bias Variance Trade-off.

Ans 15. SVM is known as Support Vector Machine. It is a famous supervised machine learning algorithm that is used to classify regression algorithms. The goal of SVM algorithm is to make the best decision so that we can easily put the new data point in the correct category in the future.

Linear SVM

Linear SVM is made for linearly separable data that indicates if a dataset can be classified into two sets by using a single straight line, then such data is known as linearly separable data. And this classifier is called as Linear SVM classifier.

RBF

Also known as Radial basis function. It is a famous kernel function used to kernelized various different learning algorithms. It is used to support vector machine classification.

Polynomial Kernel

Polynomial kernel is a kernel function in machine learning. It is commonly used with SVMs. This shows the similarity of samples over polynomials of the original variables. The polynomial kernel not only focus on the given features but also the combination of these.