ASSIGNMENT - 5

MACHINE LEARNING

1. The residual sum of squares (RSS) is the absolute amount of explained variation, whereas R-squared is the absolute amount of variation as a proportion of total variation.
   And both are commonly used measures to assess the goodness of fit of a regression model, but they capture different aspects of model performance, and the choice between them depends on the context and what you want to evaluate.
   The RSS is useful for assessing the goodness-of-fit of the regression model***, as well as for comparing different models to determine which one provides the best fit.***

2. ***The sum of squares total (SST) or the total sum of squares (TSS)** is the sum of squared differences between the observed dependent variables and the overall mean.
   * **The explained sum of squares (ESS) is** the sum of the squares of the deviations of the predicted values from the mean value of a response variable, in a standard regression model .

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

   THE EQUATION RELATING THESE THREE METRICS WITH EACH OTHER:-

   $$TSS = ESS + RSS$$

3. Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.Sometimes the 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.

The commonly used regularization techniques are :-

1. Lasso Regularization – L1 Regularization
2. Ridge Regularization – L2 Regularization
3. Elastic Net Regularization – L1 and L2 Regularization

4. The Gini Index is the additional approach to dividing a decision tree. Purity and impurity in a junction are the primary focus of the Entropy and Information Gain framework. The Gini Index, also known as Impurity,

calculates the likelihood that somehow a randomly picked instance would be erroneously cataloged.

5.

6. **Ensemble learning combines multiple machine learning** models into a single model. The aim is to increase the performance of the model. Bagging aims to decrease variance, boosting aims to decrease bias, and stacking aims to improve prediction accuracy. Bagging and boosting combine homogenous weak learners.
7. **Bagging is a learning approach that aids in enhancing the performance, execution, and precision of machine learning algorithms. It is the easiest method of merging predictions that belong to the same type. In bagging, each model is assembled independently.**

Boosting is an approach that iteratively modifies the weight of observation based on the last classification. It is a method of merging predictions that belong to different types. In boosting, the new models are impacted by the implementation of earlier built models.

8. **OOB (out-of-bag)** score is a performance metric for a **machine learning** model, specifically for ensemble models such as random forests. It is calculated using the samples that are not used in the training of the model, which is called out-of-bag samples. These samples are used to provide an unbiased estimate of the model's performance, which is known as the OOB score.

9. **K-fold cross-validation** is a technique for evaluating predictive models. The dataset is divided into k subsets or folds. The model is trained and evaluated k times, using a different fold as the validation set each time. Performance metrics from each fold are averaged to estimate the model's generalization performance. This method aids in model assessment, selection, and hyperparameter tuning, providing a more reliable measure of a model's effectiveness.

10. **Hyperparameters**,are a type of Hyperparameters cannot be directly learned from the regular training process. They are usually fixed before the actual training process begins. These parameters express important properties of the model such as its complexity or how fast it should learn. This article aims to explore various strategies to tune hyperparameters for Machine learning models.

Purpose:-

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.*

11. If the learning rate is too high, the algorithm may overshoot the minimum, and if it is too low, the algorithm may take too long to converge. Overfitting: *Gradient descent can overfit the training data if the model is too complex or the learning rate is too high.*

12.

13. *In **Adaboost,** the weights of the sample are adjust at each interaction.*

The final model is formed by prediction from individual tress through a weighted sum. Training process starts with a decision stump. At every step the weights of the training samples are misclassified Are increased for the next interation. The next tree is built sequentially on the same training data but using the newly weighted training samples.

*In **Gradient Boosting**, no reweighting of the samples takes place in GBM.the final model is the equal weighted sum of all the individual trees. It uses gradient descent to iteratively fit new weak learners to the residuals of the previous ones, minimizing a loss function. There are several loss function to choose from, mean squared error being most common for regression and cross entropy for classification.*

14. In statistics and machine learning, the **bias–variance tradeoff** describes the relationship between a model's complexity, the accuracy of its predictions, and how well it can make predictions on previously unseen data that were not used to train the model. In general, as we increase the number of tunable parameters in a model, it becomes more flexible, and can better fit a training data set. It is said to have lower error, or bias. However, for more flexible models, there will tend to be greater **variance** to the model fit each time we take a set of samples to create a new training data set. It is said that there is greater variance in the model's estimated parameters.

15. *Linear SVM* is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.A simple linear SVM classifier works by making a straight line between two classes. That means all of the data points on one side of the line will represent a category and the data points on the other side of the line will be put into a different category.

Radial Basis Function Support Vector Machine (RBF SVM) is a powerful machine learning algorithm that can be used for classification and regression tasks. It is a non-parametric model that works well with non-linear and high-dimensional data.

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.

END

STATISTICS:-

1ans. B
2ans. D
3ans. C
4ans. B
5ans. C
6ans. A
7ans. A
8ans. B
9ans. B
10ans. A

END