MACHINE LEARNING ASSIGNMENT

ANS. NO.1

R-Squared

R-Squared (R^2 or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable.
R-squared shows how well the data fit the regression model (the goodness of fit)

R-squared will give you an estimate of the relationship between movements of a dependent variable based on an independent variable's movements. It doesn't tell you whether you chosen model is good or bad, nor will it tell you whether the data and predictions are biased.

Residual Sum of Squares (RSS)

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.

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.

Conclusion: Thus Residual Sum of Squares (RSS) is best measure of goodness of fit model in regression.

ANS.NO.2

TSS: This gives you the distance from the linear line drawn to each particular variable. You could also describe TSS as the dispersion of observed variables around the mean, or the variance. So, the goal of TSS is to measure the total variability of the dataset.

ESS: 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.

RSS: 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.

The equation relating these three metrics with each other:

Total sum of squares (TSS) = explained sum of squares (ESS) + residual sum of squares (RSS).

ANS.NO.3

Regularization is most important of machine learning. It is a technique to prevent the model from over fitting 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 over fitted. 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. L1 regularization
2. L2 regularization
3. Dropout regularization

ANS.NO.4

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. More precisely, the Gini Impurity of a dataset is a number between 0-0.5, which indicates the likelihood of new, random data being misclassified if it were given a random class label according to the class distribution in the dataset.

ANS.NO.5

Decision trees are prone to overfitting, 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.

Only regularized decision trees are prone to overfitting and Un regularized decision-trees not prone to overfitting

ANS.NO.6

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.

ANS.NO.7

(Bagging) is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It decreases the variance and helps to avoid over fitting. It is usually applied to decision tree method. Bagging is a special case of the model averaging approach.

(Boosting) is an ensemble modelling technique that attempts to build a strong classifier from the number of weak classifiers. It is done by building a model by using weak models in series. Firstly, a model is built from the training data. Then the second model is built which tries to correct the errors present in the first model. This procedure is continued and models are added until either the complete training data set is predicted correctly or the maximum number of models is added.

ANS.NO.8

An error estimate is made for cases that were not used when constructing the tree. This is called an out-of-bag(OOB) error estimate mentioned as a percentage. The decision trees are prone to over fitting and this is the main drawback of it.

ANS.NO.9

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.

ANS.NO.10

In machine learning, a hyper parameter is a parameter whose value is used to control the learning process. By contrast, the values of other parameters (typically node weights) are derived via training.

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

Hyper parameters in Machine learning are those parameters that are explicitly defined by the user to control the learning process. These hyper parameters are used to improve the learning of the model, and their values are set before starting the learning process of the model.

ANS.NO.11

In order for Gradient Descent to work, we must set the learning rate to an appropriate value. This parameter determines how fast or slow we will move towards the optimal weights. **If the learning rate is very large we will skip the optimal solution**.

ISSUES:

Vanishing Gradient. If the range of the initial values for the weights is not carefully chosen, and the range of the values of the weights during training is not controlled, a vanishing gradient would occur which is the main hurdle to learning in deep networks.

ANS.NO.12

Logistic regression is neither linear nor is it a classifier. The idea of a "decision boundary" has little to do with logistic regression, which is instead a direct probability estimation method that separates predictions from decision.

ANS.NO.13

Difference between Gradient boosting vs AdaBoost

Adaboost and gradient boosting are types of ensemble techniques applied in machine learning to enhance the efficacy of week learners. The concept of boosting algorithm is to crack predictors successively, where every subsequent model tries to fix the flaws of its predecessor. Boosting combines many simple models into a single composite one. By attempting many simple techniques, the entire model becomes a strong one, and the combined simple models are called week learners. So the adaptive boosting and gradient boosting increases the efficacies of these simple model to bring out a massive performance in the machine learning algorithm.

ANS.NO.14

In statistics and machine learning, 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.

ANS.NO.15

Linear SVM: 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.

Liner SVM: 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.

Polynominal Kernel: In machine learning, 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