MACHINE LEARNING WORKSHEET-5

Ques1:

R-squared is a better measure of goodness of fit model in regression because it represent smaller difference between the observed and fitted data.

Ques2:

The Total Sum of Squared regression is the sum of the differences between the predicted value and the mean of the dependent variable.

$$TSS=\sum (y_i-\overline{y})^2 \quad ; i=1, 2,\ldots,N$$

The Explained Sum of Squares is the deviation of the predicted values from the mean value of a response variable, in a standard regression model.

$$ESS=\sum(\hat{y}_i-\overline{y})^2 \quad ; i= 1, 2, \ldots,N$$

The Residual Sum of Squares measures the level of variance in the error term of a regression model.

$$RSS=\sum(y_i-\hat{y}_i)^2 \quad ; i=1, 2, \ldots,N$$

Ques3:

Regularization is used to minimized the error and prevent the overfitting or underfitting in the machine learning models.

Ques4:

Gini impurity is going to find which feature has low impurity. The value impurity lies between 0 and 1, where 0 being no impurity and 1 denotes random distribution.

Ques5:

Yes, decision tree is prone to overfitting. A perfect fit decision tree perform well for train data but poor for unseen test data.

Ques6:

An ensemble technique method is a technique which use multiple learning algorithms to obtain better predictive performance.

Ques7:

Bagging combine prediction belongs to the same types and boosting combine prediction belongs to the different types.

Bagging decrease variance, not bias while boosting decrease bias, not variance.

Ques8:

The out of bag error is the average for each calculated using predictions from the trees that do not contain in their respective bootstrap sample.

Ques9:

K-fold cross validation is when the dataset is split into k number of folds and used to evaluate the model's ability when given new data.

Ques10:

In hyper parameter tuning we find the optimal parameter for the optimized algorithm to any data set. We perform tuning to improve our result.

Ques11:

If we have large learning rate in gradient descent it cause the model to converge too quickly to suboptimal solution.

Ques12:

No, we cannot use Logistic Regression for classification of non-linear data because it is linear model.

Ques13:

AdaBoost is first algorithm with a particular loss of function whereas gradient boosting search approximate solution of the problem.

Gradient boosting is more flexible than AdaBoost.

Ques14:

It is the property of a model that the variance of the parameter estimated across sample can be reduced by increasing the bias in the estimated parameters.

Ques15:

Linear kernel is used to separate the data linearly.

Radial basis function perform linear manipulations to map points to higher-dimensional spaces that are easier to separate.

Polynomial kernel works on original variables, allowing learning of non-linear models.