MachineLearning-Worksheet5

1. **RSS**:
   A residual sum of squares (RSS) measures the level of variance in the error term, or residuals, of a regression model.
   Ideally, the sum of squared residuals should be a smaller or lower value than the sum of squares from the regression model's inputs.
   R2:
   R-Squared is a statistical measure of fit that indicates how much variation of a dependent variable is explained by the independent variable(s) in a regression model.

   The RSS is just the absolute amount of explained variation, the R squared is the absolute amount of variation as a proportion of total variation.Rsquared is better.

2. Regularisation is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting. The commonly used regularisation techniques are : L1 regularisation. L2 regularisation3.

3. Gini index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.

4. 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.

5. Ensemble methods are techniques that create multiple models and then combine them to produce improved results. Ensemble methods usually produces more accurate solutions than a single model would.

6. Bagging is a method of merging the same type of predictions. Boosting is a method of merging different types of predictions.
   Bagging decreases variance, not bias, and solves over-fitting issues in a model.
   Boosting decreases bias, not variance.

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

8. Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample.

9. 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. As such, the procedure is often called k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation.

10. Hyper parameter optimization or tuning is the problem of choosing a set of optimal hyper parameters for a learning algorithm. A hyperparameter is a parameter whose value is used to control the learning process. Controlling the behaviour of the training algorithm and has a significant impact on the performance of the model .

11. When the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error.

12. Logistic regression is indeed non linear in terms of Odds and Probability, however it is linear in terms of Log Odds.

13. AdaBoost is designed for boosting algorithm with a particular loss function. Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem. This makes Gradient Boosting more flexible than AdaBoost.

14. Bias is the simplifying assumptions made by the model to make the target function easier to approximate. Variance is the amount that the estimate of the target function will change given different training data. Trade-off is tension between the error introduced by the bias and the variance.

15. Polynomial: It is popular in image processing.

    RBF: It is a general-purpose kernel; used when there is no prior knowledge about the data.
    Linear: It is useful when dealing with large sparse data vectors. It is often used in text categorization.