Machine Learning Worksheet -5

1. R-squared is better measure of goodness of fit as it provides the relative measure in the form of proportion/percentage of the variance of the dependent variable that can be explained by the model. Therefore, R-squared closer to 1 gives us the indication of a good fit. Whereas, **RSS** (Residual Sum of Squares) is just the sum of squares of the residuals/errors for the model's predictions, which is an absolute value, therefore, it may not be reliable to measure the goodness of fit.

2. TSS (Total Sum of Squares): It measures total variations in the observed data.

RSS (Residual Sum of Squares): It measures total variations in the errors between the observed and the predicted or modelled values.

ESS (Explained Sum of Squares): It measures how well a regression model represents the modelled data.

They all satisfy the following equation:

TSS = ESS + RSS

3. Regularization in Machine Learning helps to reduce the model's overall complexities, which may reduce the risk of overfitting of models that have high variance.

4. In a decision tree algorithm, weighted **Gini-impurity index** for a particular column gives the value of weighted sum of impurities of sub-nodes or

categories of that particular column. Therefore, based on the lowest weighted Gini-impurity index, the algorithm selects a column as a decision node for splitting the data for constructing decision tree.

Mathematically,

Gini-impurity index = 1 – $\sum_{i=1}^{n}(\boldsymbol{pi})^2$

Here, ***pi*** is the probability of an object or category being classified to a particular class.

Also, it is to be noted that the Gini-impurity index calculated for each sub-node are added by the weighted sum to get final weighted Gini-impurity index.

5. Un-regularized decision trees are usually prone to overfitting as the constructed tree does not apply any pruning technique to reduce the complexities and hence dept of a tree increases to overfit the training set.

6. Ensemble techniques uses weak learners or base models that are trained together using various methods like **bagging, boosting** or **stacking** to form a single powerful and robust model. This technique is usually used to reduce **bias** or **variance** of a model to improve overall accuracy of the model.

7.

S. No.	**Bagging**	**Boosting**
1	Each base model is built independently.	New base models are dependent on the performance of the previous models

2	Base models are built parallelly.	Base models are built sequentially.
3	Each base model has equal weights.	Base models are weighted according to their performance.
4	This technique mainly aims to reduce variance of a model.	This technique mainly aims to reduce bias of a model.
5	Example: - Random Forest	Example: - Gradient Boosting

8. Out of bag error is the average prediction error from the random forest model for that samples which are not included in the bootstrap samples for building the model.

9. K-fold cross-validation is a model validation technique in which an entire training set is divided into **K-folds.** Here, **1**-fold is treated as a testing set and the rest **K-1 folds** as training set. And, the model is evaluated for all **K** iterations in which the data in training and testing set varies based on folds for each iteration. And the final score is evaluated as the average score over all **K** iterations. This technique is often used when the amount of data is not sufficient to split into training, validation and testing set.

10. Hyper-parameter tuning involves in testing and experimenting with different hyper-parameters of the model that are not learned during training process. These hyper-parameters are manually set or set by default before the

training process and they control the overall behaviour and performance of a Machine Learning model.

11. Large learning rate in Gradient Descent increases the risk of over-shooting the minimum of the cost function, in which the learning algorithm may never converge to the minimum or can even become unstable and diverge from the minimum.

12. Logistic regression cannot be used for classification of Non-linear data points as logistic regression is a linear classification model that learns linear decision boundary for classifying data.

13.

S No.	**Adaboost**	**Gradient Boosting**
1	The base model usually consists of a single level decision trees, also known as decision stumps.	Base models are decision trees with greater depth.
2	The misclassifications or errors are accounted by the updated weights associated with data points.	The misclassifications or errors are accounted by the gradient optimization steps involved.

3	All base models can have different importance's based on their performance in training set for the final prediction.	All base models are weighted equally in which their predictive capacity is restricted with fixed learning rate to increase overall accuracy.

14. The assumptions made by the model to make the mapping or target function easier and considerably simple for the problem is called **bias** and error associated with the model due to this property is called **bias error**. Whereas, Variance is a property of model that is sensitive to small fluctuations in training data and thus is an indication of how much the estimate of the target function changes by giving different training data and the error associated due to this property is called **variance error**. Since, decreasing only bias error increases variance error and vice-versa, therefore, there is a **trade-off** between variance and bias of the model.

15. Kernels in SVM are basically mathematical functions that transform the input data points into required form. The most common SVM kernels are as follows:

Linear: This kernel is used to classify linearly separable data points with maximum margin.

Polynomial kernels: This kernel represents the similarity of vectors that are training samples in a feature space over polynomials of the original variables

by dot product operations. This technique allows to map the input data points into high dimensional space which further allows classification of non-linear data points.

RBF (Radial Basis Function): The output of this kernel mathematically represents the relationship between two vectors or data points in infinite dimensions by taking dot product that includes the coordinates for an infinite number of dimensions. The higher the output of this kernel the more similar are the data points and vice-versa, to whether consider them in a same or different class. This kernel is robust with respect to almost all kinds of data points and is most computationally demanding compared to other kernels.