# Machine Learning worksheet 2

1. R-squared is the amount of variation as proportion of total variations.It is often used to determine how well the independent variable explain the variation in the dependent variable while RSS is the absolute amount of explained variations. It is used to know presence of errors and residuals in model. Both are used in regression analysis to measure goodness of fit. Which method is to be used depends on context that we want. They give different result for regression model.
   Still comparatively RSS should be preferred for better error free analysis. It gives the assess of the absolute fit of the model.

2. ESS(Explained Sum of Squares)- sum of squares of the deviations of predicted value from the mean value of independent variable in regression model.

   RSS(Residual Sum Of Square)- sum of squared estimate of errors- It measures the level of variance in the term of errors or residuals.

   TSS(Total Sum Of Squares)- It calculates the sum of squared differences between observed dependent variable and overall mean of dependent variables.

   TSS=RSS+ESS

   This is the relation between these three terms.

3. Regularization is a technique used in machine learning to prevent overfitting and underfitting to improve the ability of models. In regularization we fit our machine learning model in several test and reduce the errors in it. Regularization is necessary in various scenarios for several reasons. As
   Handling Multicollinearity, preventing overfitting, improve model stability, balancing the Bias variance trade off etc.

4. Gini impurity index- Gini impurity measures the probability of misclassification of a randomly chosen element in the dataset if it were randomly labelled according to the distribution of labels in the dataset. It ranges is between 0 and 1. A Gini index of 0 represents perfect (purity) equality, while an index of 100 implies perfect inequality.

5. Yes, unregularized decision trees are prone to overfitting. Overfitting occurs when a model learns the training data too well. Unregularized decision trees can keep splitting data until each training sample has its own dataset. Unregularized decision trees tend to learn complex details in the training data, which can lead to high variance and poor generalization to unseen data.

6. Ensemble techniques involve combining multiple models to create a more powerful and accurate predictive model. Ensemble methods produce more accurate strengths of models to improve overall performance and give better results than any single model alone.

7. Bagging- It combines multiple models trained on different subsets of data. To create subsets of data it uses Bootstrap. Each model has equal weight in final decision making and equal error rate that's why it has less chances to overfit and has more accuracy.
   Example- Random Forest

   Boosting- It trains models and focuses on errors made by previous models. It weights the data again based on errors. If there are more errors then there are more weight.
   Example – Adaboost, XGboost.

8. Out of bag error- The out-of-bag error is calculated in the Random Forest using the data points that were not included in out-of-bag samples in the bootstrap sample. Since these samples were not used to train the tree, they give a validation set for that tree.

9. K fold cross validation- K-fold cross-validation is a technique for evaluating predictive models. The dataset is divided into k subsets of

equal sizes called folds. Then model is trained and evaluated k times, using a different folds each time for training and validation. Performance metrics from each fold are managed to estimate the model's generalization performance. In each fold training and testing will be performed once during this entire process. It helps us to avoid overfitting. K-fold cross-validation helps to provide a more accurate estimate of the model's performance compared to a single train-test split. It utilizes the entire dataset for both training and validation.

10. When we train machine learning models, each model needs a different set of hyperparameters, which are a kind of variable. The only way to determine these is with multiple experiments where you pick a set of hyperparameters and run them through your model. This is called hyperparameter tunning. This process can be manual or can picked one of several automated hyperparameter tuning methods.

    Manually process is slow whereas automated hyperparameter is too quick.

11. Gradient descent can overfit the training data if the model has its learning rate is too high. This also overshoot the minimum point and may difficulty in convergence.
12. No, we can not use Logistic Regression for classification of Non-Linear Data because Logistic Regression is a linear algorithm primarily used for binary classification tasks. it might not perform well on non-linear data without transformation. If the relationship between features and the target variable is non-linear, Logistic Regression may struggle to fit the data accurately and leads for underfitting.

13. Adaboost - It focuses on multiple weak learners where each learner corrects the mistakes of its predecessor. It adjusts the weights of the training instances and focusing more on misclassified instances in subsequent iterations. It allows the next weak learner to pay more attention to them. It is less prone to be overfit.

    Gradient boosting - It optimizes the loss function by minimizing the errors at each step. It assigns each weak learner a coefficient that

determines its contribution to the final model. It is more prone to overfitting.

14. The bias-variance trade off is a concept in machine learning that finds the right balance between bias and variance in predictive model. It 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.

15. Linear Kernel – It's the simplest kernel represents the dot product of the input features in the original space. It is suitable for linearly separable data where classes can be separated by a straight line or hyperplane.

    RBF (Radial Basis Functions) - RBF kernel is popular for handling non-linearly separable data. measures the similarity between two samples in a higher-dimensional space using the Gaussian (radial basis) function. It measures the similarity between two samples in a higher-dimensional space and make it highly flexible.

    Polynomial kernel - It represents the similarity between samples in a feature space over polynomials of the original variables, allowing learning of non-linear models. It suitable for capturing non-linear relationships in the data.