MACHINE LEARNING WORKSHEET -5

1. R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit model in regression and why?
   Ans: **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. In other words, r-squared shows how well the data fit the regression model (the goodness of fit).

   Residual Sum of Squares: 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, or error term.

   R-squared or Residual Sum of Squares (RSS) both are a better measure of goodness of fit model in regression.

2. What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sum of Squares) in regression. Also mention the equation relating these three metrics with each other.
   Ans: **Total Sum of Squares:** The Total SS (TSS or SST) tells you how much variation there is in the dependent variable.

   $$\text{Total SS} = \Sigma(Yi - \text{mean of } Y)^2.$$

   Explained Sum of Squares: The Explained SS tells you how much of the variation in the dependent variable your model explained.

   $$\text{Explained SS} = \Sigma(Y\text{-Hat} - \text{mean of } Y)^2.$$

   Residual Sum of Squares: The residual sum of squares tells you how much of the dependent variable's variation your model did not explain. It is the sum of the squared differences between the actual Y and the predicted Y:

   $$\text{Residual Sum of Squares} = \Sigma\, e2$$

Equation relating these three metrics is given by :

TSS = ESS + RSS

where, TSS = Total Sum of Squares, ESS = Explained Sum of Squares, RSS = Residual Sum of Squares

3. What is the need of regularization in machine learning?
   Ans:
   - To minimize the adjusted loss function and prevent overfitting or underfitting.

   - Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.
   - There are two main types of regularization techniques: Ridge Regularization and Lasso Regularization.

4. What is Gini–impurity index?
   Ans: Gini Index, also known as Gini impurity, calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly. If all the elements are linked with a single class then it can be called pure.

5. Are unregularized decision-trees prone to overfitting? If yes, why?
   Ans: Yes. They are prone to over-fitting. Decision trees are prone to over-fitting, 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.

6. What is an ensemble technique in machine learning?
   Ans: An ensemble method is a technique which uses multiple independent similar or different models/weak learners to derive an output or make some predictions.

7. What is the difference between Bagging and Boosting techniques?

Bagging	**Boosting**
Objective to decrease variance, not bias.	Objective to decrease bias, not variance.

Each model is built independently.	New models are affected by the implementation of the formerly developed model.
It is the simplest way of connecting predictions that belong to a similar type.	It is a method of connecting predictions that belong to multiple types.
Several training data subsets are randomly drawn with replacement from the whole training dataset.	Each new subset includes the components that were misclassified by previous models.
Bagging can solve the over-fitting problem.	Boosting can boost the over-fitting problem.

8. What is out-of-bag error in random forests?
   Ans: 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. This allows the RandomForestClassifier to be fit and validated whilst being trained.

9. What is K-fold cross-validation?
   Ans: 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. Each fold is used as a testing set at one point in the process.

10. What is hyper parameter tuning in machine learning and why it is done?
    Ans: Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.

11. What issues can occur if we have a large learning rate in Gradient Descent?
    Ans: A learning rate that is too large can cause the model to converge too quickly to a suboptimal solution.

12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?
    Ans: No, logistic regression only forms linear decision surface. Logistic Regression has traditionally been used as a linear classifier, i.e. when the classes can be separated in the feature space by linear boundaries.

13. Differentiate between Adaboost and Gradient Boosting.
    Ans: Gradient Boosting is a generic algorithm to find approximate solutions to the additive modelling problem, while AdaBoost be a special case with a particular loss function. Hence, Gradient Boosting is much more flexible.

    On the other hand, AdaBoost can be interpreted from a much more intuitive perspective and can be implemented without the reference to gradients by reweighting the training samples based on classifications from previous learners

14. What is bias-variance trade off in machine learning?
    Ans: 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.

    If our model is too simple and has very few parameters then it may have high bias and low variance. On the other hand if our model has large number of parameters then it's going to have high variance and low bias. So we need to find the right/good balance without overfitting and under fitting the data.

    This trade off in complexity is why there is a trade off between bias and variance. An algorithm can't be more complex and less complex at the same time.

15. Give short description each of Linear, RBF, Polynomial kernels used in SVM.
    Ans: **Linear Kernel** is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. It is mostly used when there are many Features in a particular Data Set.

Gaussian RBF(Radial Basis Function) is another popular Kernel method used in SVM models for more. RBF kernel is a function whose value depends on the distance from the origin or from some point. Gaussian Kernel is of the following format.

In the **polynomial kernel**, we simply calculate the dot product by increasing the power of the kernel.