**MachineLearning**(Worksheet-3)

## Radial basis function kernel (RBF): RBF is a Kernel method used in SVM models. RBF kernel is a function whose value depends on the distance from the origin or from some point. It is really effective in the higher dimension and effective when the number of features is more than training examples.

Polynomial: The polynomial kernel simply calculates the dot product by increasing the power of the kernel.

Linear: 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.

2. RSS used to measure the amount of [variance](https://www.investopedia.com/terms/v/variance.asp) in a data set that is not explained by a regression model. Regression is a measurement that helps to determine the strength of the relationship between a dependent variable and a series of other changing variables or independent variables whereas when evaluating the goodness-of-fit of simulated. it is not appropriate to base this on the R2 of the linear regression. The R2 quantifies the degree of any linear correlation between observed values and predicted values. while for the goodness-of-fit evaluation only one specific linear correlation should be taken into consideration.

So, RSS is a better measure for goodness of fit.

3. The TSS tells that how much variation is there in the [dependent variable](https://www.statisticshowto.com/dependent-variable-definition/).

The Explained SS tells that how much of the variation in the dependent variable our model explained.

It is the sum of the squared differences between the actual Y and the predicted Y

TSS = ESS + RSS

4. The Gini impurity measure is one of the methods used in decision tree algorithms to decide the optimal split from a root node, and subsequent splits.

5. 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.

6. 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. Ensemble learning combines the predictions from multiple neural network models to reduce the variance of predictions and reduce generalization error. Techniques for ensemble learning can be grouped by the element that is varied, such as training data, the model, and how predictions are combined.

7. Bagging is a way to decrease the variance in the prediction by generating additional data for training from dataset using combinations with repetitions to produce multi-sets of the original data. Boosting is an iterative technique which adjusts the weight of an observation based on the last classification. aging and Boosting get N learners by generating additional data in the training stage. N new training data sets are produced by random sampling with replacement from the original set. By sampling with replacement some observations may be repeated in each new training data set.

In the case of Bagging, any element has the same probability to appear in a new data set. However, for Boosting the observations are weighted and therefore some of them will take part in the new sets more often.

8. The out-of-bag (OOB) error is the average error for each zi calculated using predictions from the trees that do not contain zi in their respective bootstrap sample. This allows the Random Forest Classifier to be fit and validated whilst being trained. It can be used to obtain a more appropriate n\_estimators.

9. Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. 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. K-Folds cross-validation uses train/test indices to split data in train/test sets. Split dataset into k consecutive folds. Each fold is then used once as a validation while the k - 1 remaining folds form the training set.

10. A hyperparameter is a parameter whose value is set before the learning process begins. The benefit of hyperparameter tuning is that it is guaranteed to find the optimal combination of parameters supplied. The drawback is that it can be very time consuming and computationally expensive.

11. Stochastic gradient descent is an optimization algorithm that estimates the error gradient for the current state of the model using examples from the training dataset, then updates the weights of the model. The learning rate controls how quickly the model is adapted to the problem. larger learning rates result in rapid changes and require fewer training epochs. A learning rate that is too large can cause the model to converge too quickly to a suboptimal solution, whereas a learning rate that is too small can cause the process to get stuck.

12. Bias is the difference between the Predicted Value and the Expected Value. Variance is when the model takes into account the fluctuations in the data i.e. the noise as well. A model with a high bias error underfits data and makes very simplistic assumptions on it. A model with a high variance error overfits the data and learns too much from it. A good model is where both Bias and Variance errors are balanced. This is known as bias-variance tradeoff.

13. This is a form of regression, that constrains/ regularizes or shrinks the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.

14. For each iteration, adaptive boosting changes the sample distribution by modifying the weights attached to each of the instances. It increases the weights of the wrongly predicted instances and decreases the ones of the correctly predicted instances. The weak learner thus focuses more on the difficult instances. After being trained, the weak learner is added to the strong one according to his performance (so-called alpha weight). The higher it performs, the more it contributes to the strong learner.

On the other hand, gradient boosting doesn’t modify the sample distribution. Instead of training on a newly sample distribution, the weak learner trains on the remaining errors (so-called pseudo-residuals) of the strong learner. It is another way to give more importance to the difficult instances.

15. Logistic regression is known and used as a linear classifier. It is used to come up with a hyperplane in feature space to separate observations that belong to a class from all the other observations that do not belong to that class. The decision boundary is thus linear. Robust and efficient implementations are readily available to use logistic regression as a linear classifier. 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.