# ASSIGNMENT 5 – Machine Learning

Q1- Typically, however, a smaller or lower value for the RSS is ideal in any model since it means there's less variation in the data set. In other words, **the lower the sum of squared residuals, the better the regression model is at explaining the data**.

Q2- The total sum of squares (TSS) measures how much variation there is in the observed data, while the residual sum of squares measures the variation in the error between the observed data and modeled values.

Q3- Regularization refers to techniques that are used **to calibrate machine learning models in order 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.

Q4- Gini Impurity is **a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree**.

Q5- 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.

Q6- **Ensemble methods** is a machine learning technique that combines several base models in order to produce one optimal predictive model.

Q7- **Bagging is the simplest way of combining predictions that belong to the same type while Boosting is a way of combining predictions that belong to the different types**. Bagging aims to decrease variance, not bias while Boosting aims to decrease bias, not variance.

Q8- 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.

Q9- Cross-validation is a statistical method used to estimate the skill of machine learning models. It is commonly used in applied machine learning to compare and select a model for a given predictive modeling problem because it is easy to understand, easy to implement, and results in skill estimates that generally have a lower bias than other methods.

Q10- Hyper parameter tuning consists of **finding a set of optimal hyper parameter values for a learning algorithm while applying this optimized algorithm to any data set**. That combination of hyper parameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.

Q11- In order for Gradient Descent to work, we must set the learning rate to an appropriate value. This parameter determines how fast or slow we will move towards the optimal weights. If the learning rate is very large we will **skip the optimal solution**.

Q12- 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. That can be remedied however if we happen to have a better idea as to the shape of the decision boundary.

Q13- AdaBoost is the first designed boosting algorithm with a particular loss function. On the other hand, 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.

Q14- In statistics and machine learning, 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**.