SVM is a classifying algorithm that separates a class of data points using a separating hyperplane. This hyperplane is not limited to two dimensional it can be multidimensional. When mapping from multidimensional to 2 dimensional we get transformations called kernels. When we have overlapping data such that some data points of a class clearly overlap unto another class’ space it is difficult to achieve a perfect partition. Using a regularization parameter and gamma however non linear classification can be achieved. By tuning the parameters mentioned above and SVM can achieve higher accuracy. The kernel plays the role of using linear algebra to predict new inputs polynomial and exponential kernels serve the role of calculating hyperplanes in higher dimensions. Regularization is a parameter used to avoid misclassifying training examples. Lastly the gamma parameter determines how much consideration is given to data points with low proximity from the hyperplane.

Logistic regression is a regression analysis used when the output is dichotomous, this is a predictive analysis. This type of regression requires one or more predictors ad calculates the probability of the binary response. The model of logistic regression models the probability of an output, doesn’t actually classify but can be used to classify by using a threshold value. There is a key difference between linear and logistic regression. This is that the conditional distribution is a Bernoulli distribution rather than a Gaussian distribution because the dependent variable is binary and second is that logistic regression is limited to binary outputs since the output of logistic regression is a probability.

Random Forests are a learning method in machine learning that is used for classification and regresssion. It functions by creating several different decision trees and then calculating the mode or mean of those tree so to produce as its output. Decision trees often overfit their training set and so random forests correct this problem. The issue with decision trees is that they often have very high variance and so random forests are a method of averaging the output of decision trees and reducing variance. It does however increase the bias in some cases. Random trees are also used to rank the importance of certain variables in the regression or classification problem.