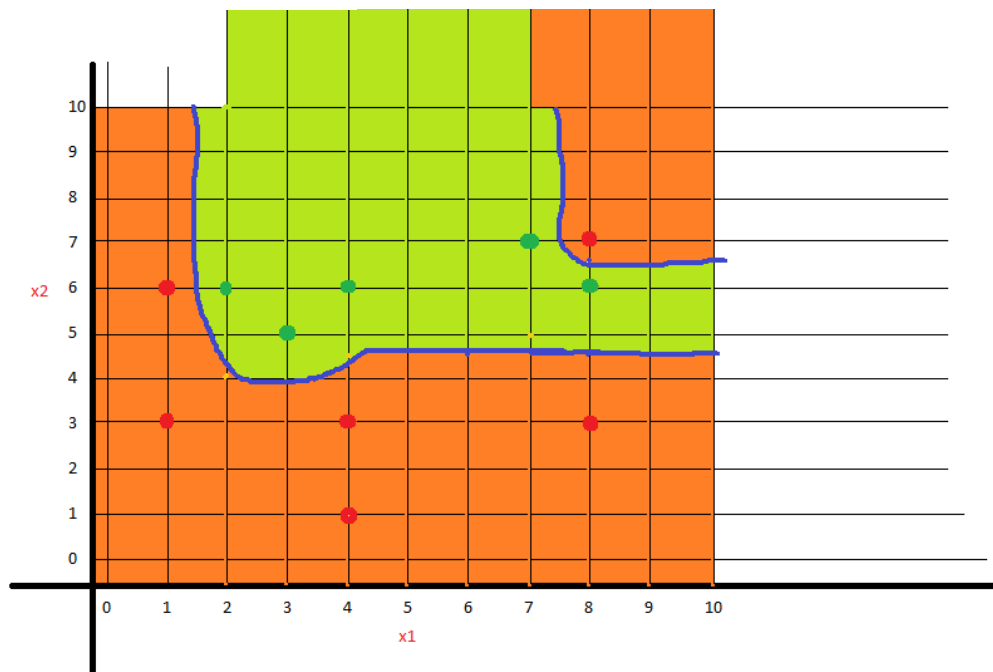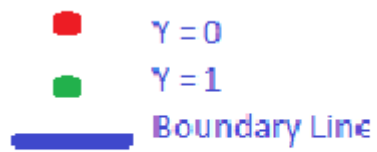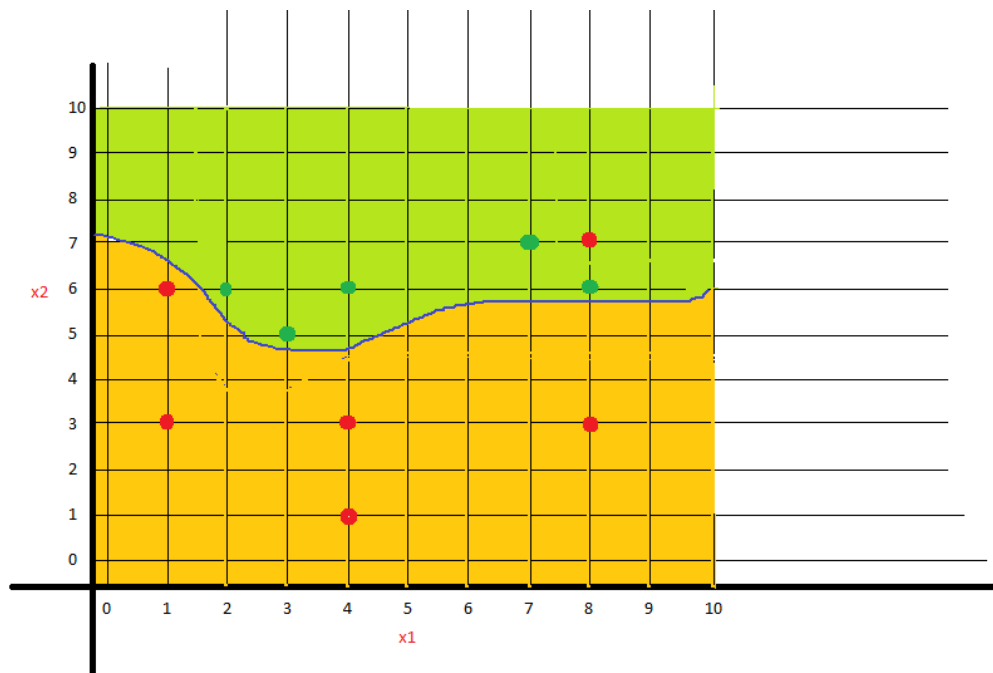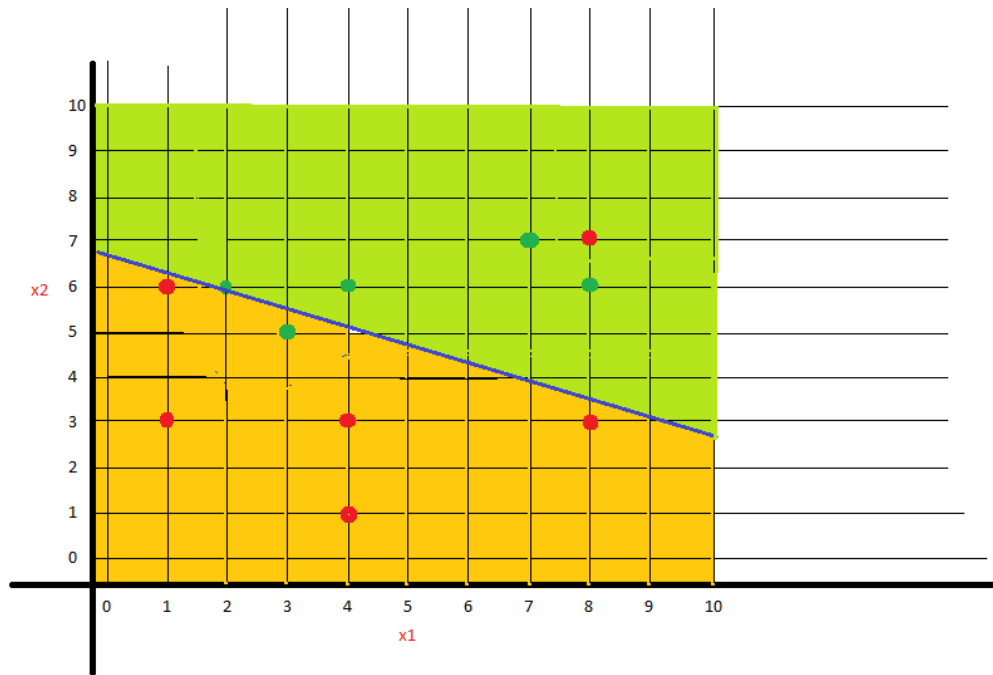Francis Mocsin – Graded assignment Part 2

A)
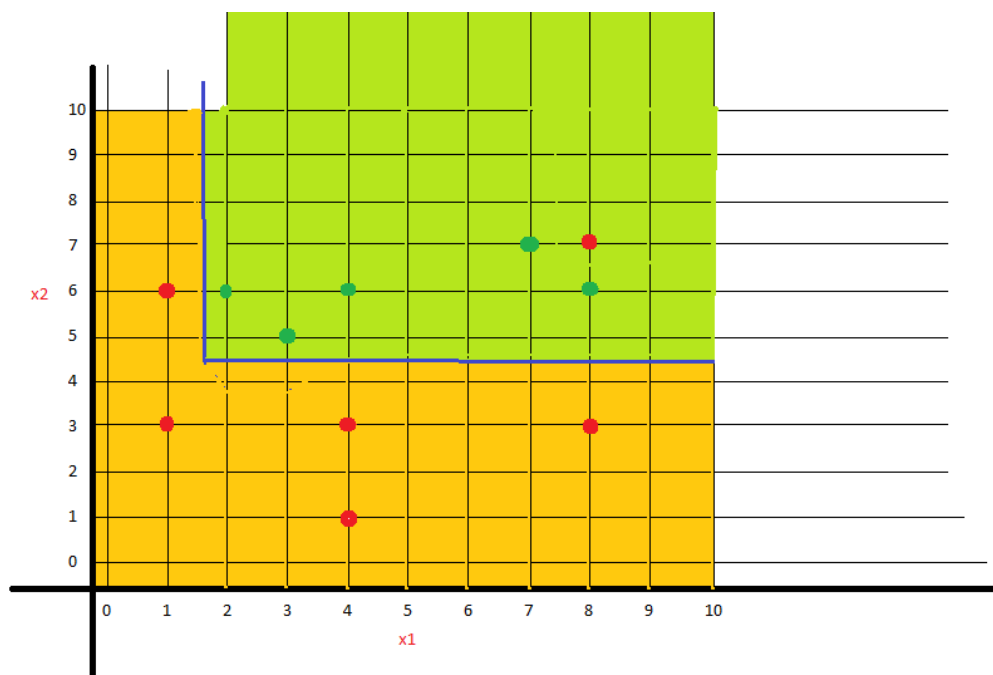
B) Intuitively speaking I would say that the K-NN learning algorithm has the best boundary line. However, as we can see in the graph it is very sensitive to outliers, changing the prediction of an entire area with just one outlier. If we manage to omit data that exhibit abnormal behaviour, and thus making the data more manageable for the K-NN learning algorithm, it is possible to create an

accurate boundary line. This could be done through the use of anomaly detection algorithms for example, which detect the points in your data that are anomalous.