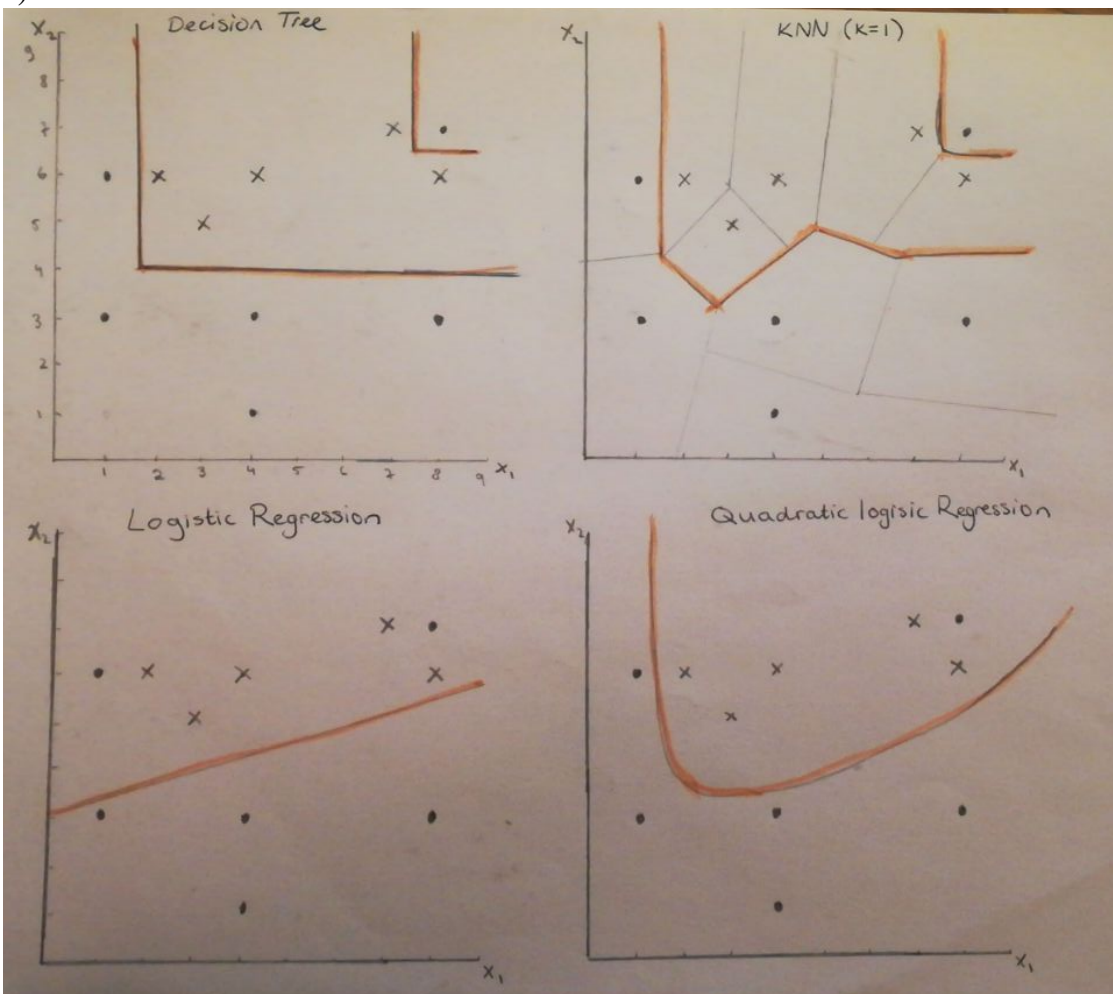
Machine Learning – Assignment 2 – Vera Grosskop

1)



2)

From the graphs above it is clear that Decision Trees and KNN fit this data best. There is no misclassified data. However, both these methods are computationally expensive and have a higher tendency to over fit. The Quadratic Logistic Regression is computationally less expensive end might be better generalizable to larger data. In conclusion, for this specific data the DT and KNN are best, but if we want to generalize then probably the QLR would perform better.

Another option is to combine the two methods. One could for example use Decision Trees or KNN for anomaly detection within a certain boundary that has earlier been drawn by QLR. By doing this it would be possible to pick out the anomalies from a smaller data set, which would therefore also take less time.