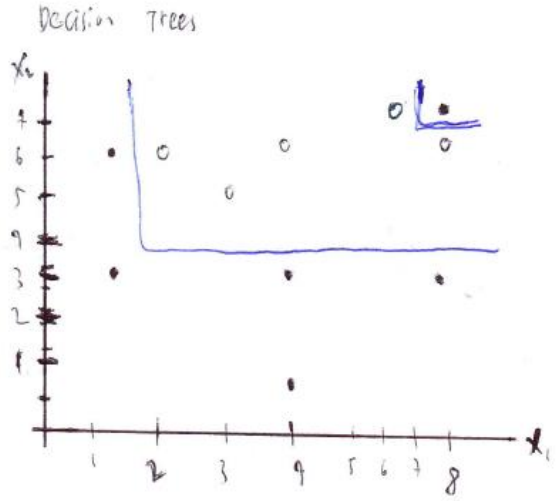
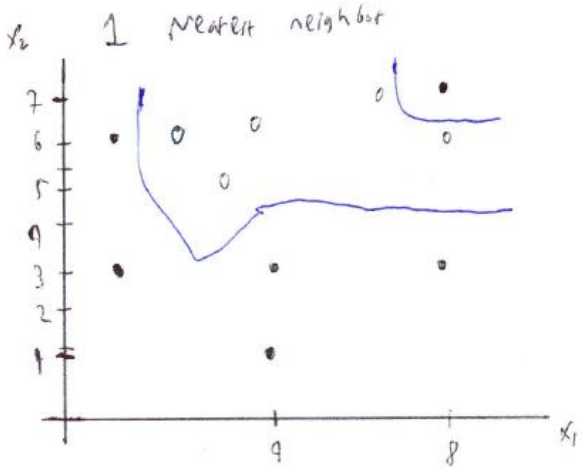
Mahdiazhari Austian

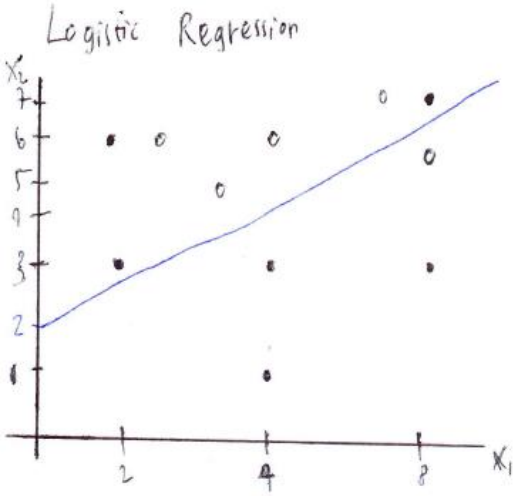
11206004

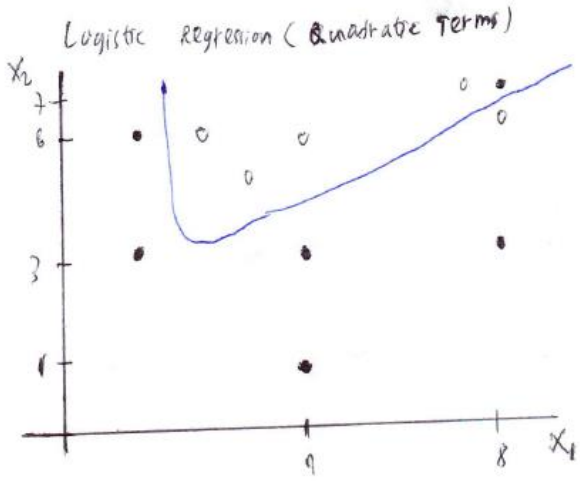
Graded Assignment 2 Part 2

1.









Notes:

For the Decision tree and the 1 Nearest Neighbor, the area under the big line indicates y=0. This is also true for the area above the small boundary (the one with a single black dot bounded by a small line). For both Logistic Regressions, the area under the blue line is y=0.

2. From what I can see here, the Decision tree and the nearest neighbor seem to try to fit the data perfectly, especially with that single dot at the top corner of the graphs. While the Decision tree and nearest neighbor might be the most accurate, it might be that these two classifiers overfit. In real life it might be that this single dot is just an anomaly or something from a measurement error. The normal

logistic regression looks too loose for the data, therefore the quadratic logistic regression works best. It seems to achieve a balance between overfitting and underfitting the given data. My intuition tells me that the use of classifiers depends on the job that we want to accomplish (depends on the data that we have). Before classifying our data, it would be good to get a general idea of the shape of the data, and then applying the algorithms to it. On the other hand, it is also possible to just use algorithms on the data, and then compare the results. The one with the seemingly best fit, is the best algorithm for the dataset.