

Machine Learning Assignment 2

Dorian Buijse

17 November 2017

Question 1

Using the supplied dataset:

X_1	1	1	2	3	4	4	4	7	8	8	8
X_2	3	6	6	5	1	3	6	7	6	7	3
Y	0	0	1	1	0	0	1	1	1	0	0

We create a scatter plot where the red circles represent $y = 0$ while the blue squares represent $y = 1$

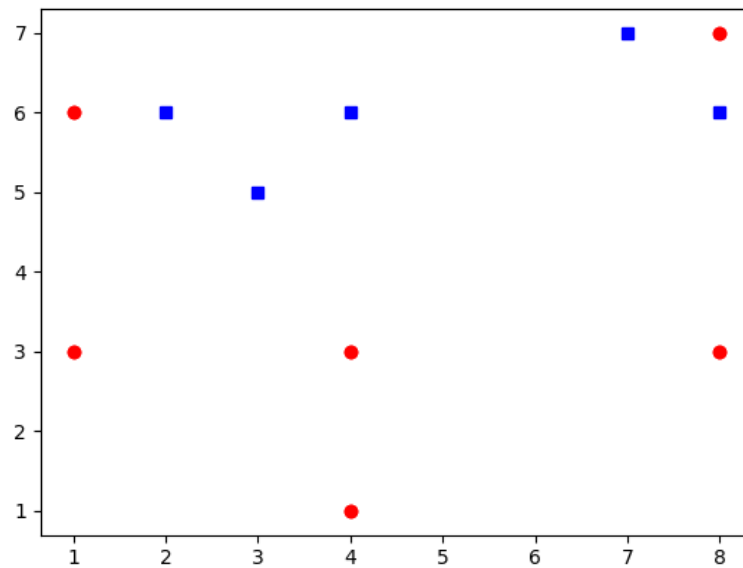


Figure 1: The Plot

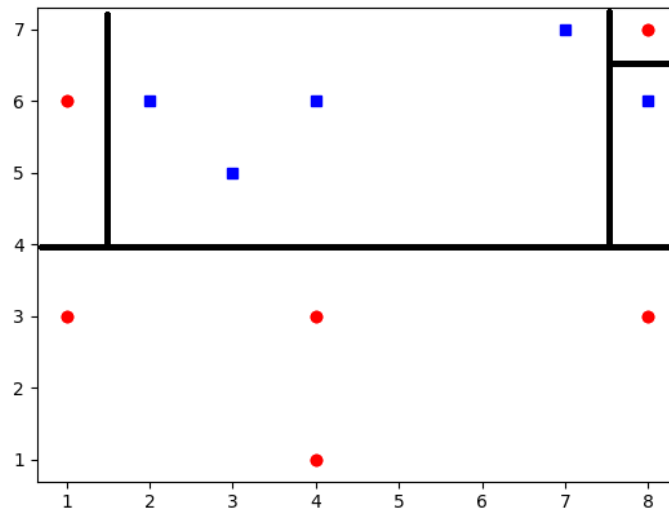


Figure 2: Decision Tree

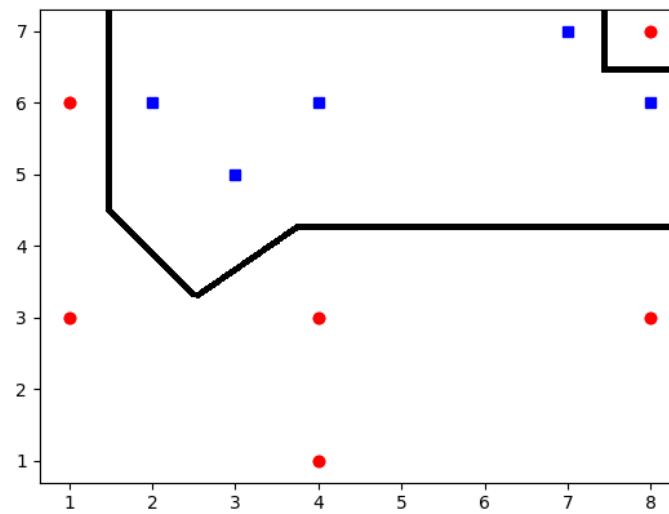


Figure 3: 1-nearest neighbour

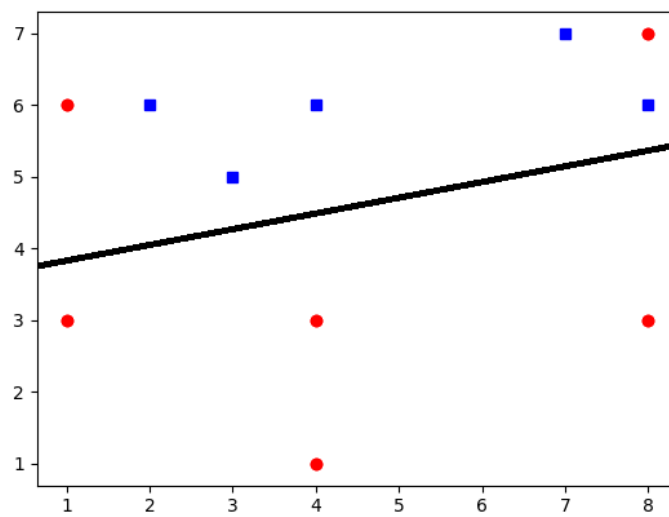


Figure 4: Plain Logistic Regression

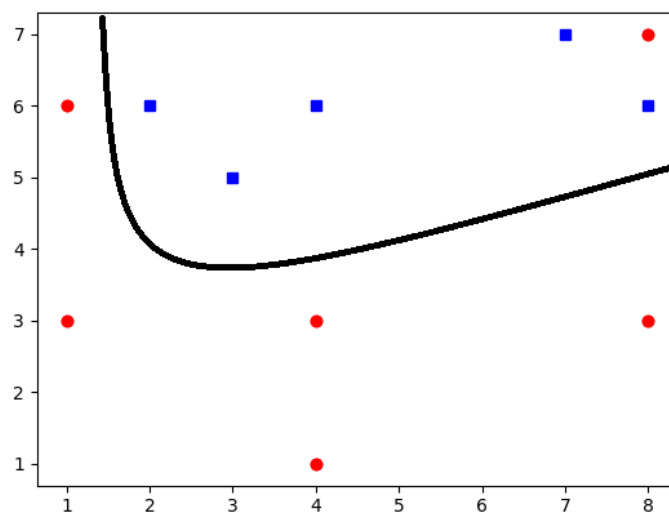


Figure 5: logistic regression with quadratic terms

Question 2

Do you intuitively think that one boundary is better than another? It may be possible to use such an intuition to invent method that uses multiple learning algorithms and combine the results, using your intuition as a prior probability.

For our data set 1-nearest neighbour seems to fit the data the best. It is however very affected by outliers. Decision tree and plain logistic regression seem to under fit the data a little where 1-nearest neighbour seems to overfit. Logistic regression with quadratic terms seems to be the least biased of the four different methods. A potential technique could be to use both logistic regression with quadratic terms and a k-nearest neighbour algorithm. That way the boundaries are fitted a bit less fluid while still retaining the quadratics overall shape.