Machine Learning Assignment 2 part 2
Katja Bouman

Decision boundaries

Below you can see the decision boundaries of Decision Trees (figure 2), 1-nearest neighbor (figure 3), plain logistic regression (figure 4), and logistic regression with quadratic terms figure 5), on the data set indicated below and in figure 1.

The dataset:

x1 = {1, 1, 2, 3, 4, 4, 4, 7, 8, 8, 8}
x2 = {3, 6, 6, 5, 1, 3, 6, 7, 6, 7, 3}
y = {0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0}

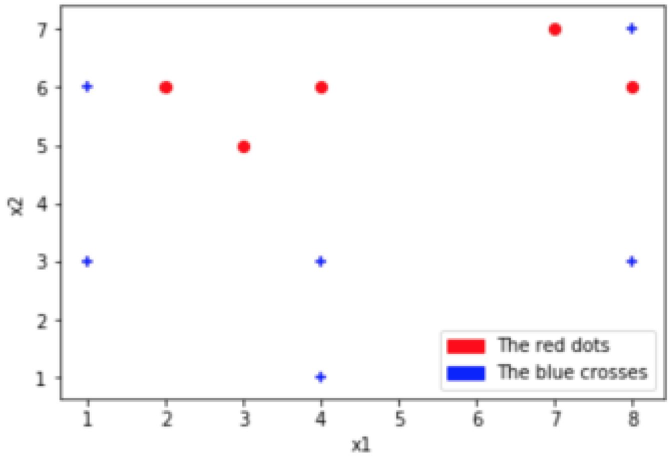

Figure 1: The dataset

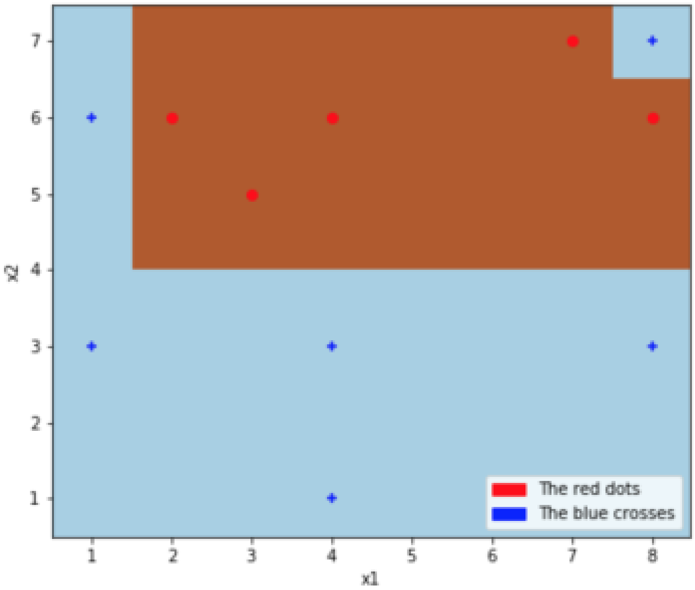

Figure 2: Decision Tree decision boundary

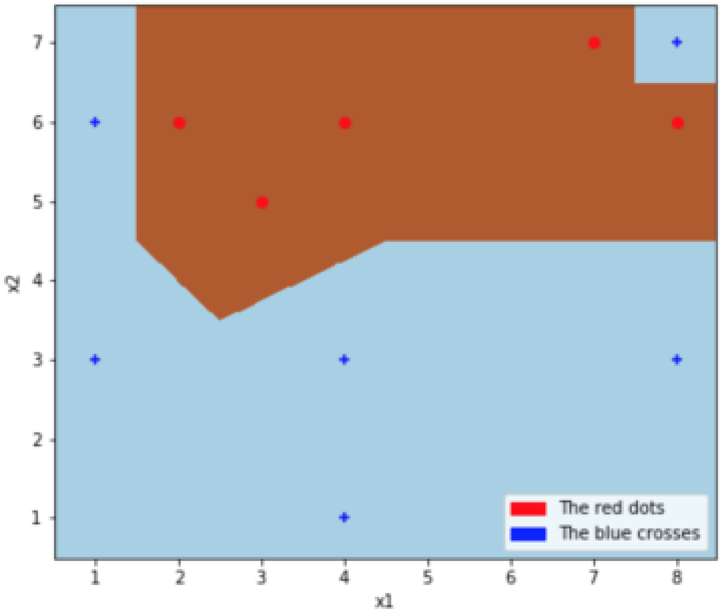

Figure 3: 1-nearest neighbour's decision boundary

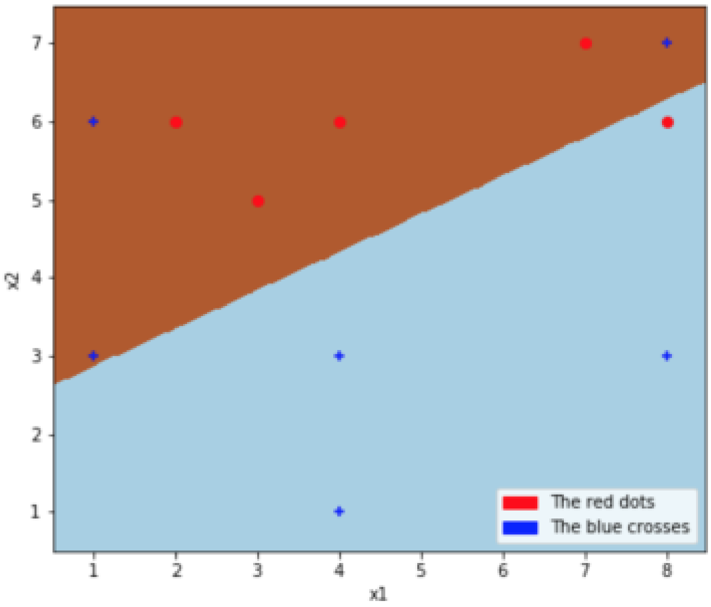

Figure 4: Logistic Regression decision boundary

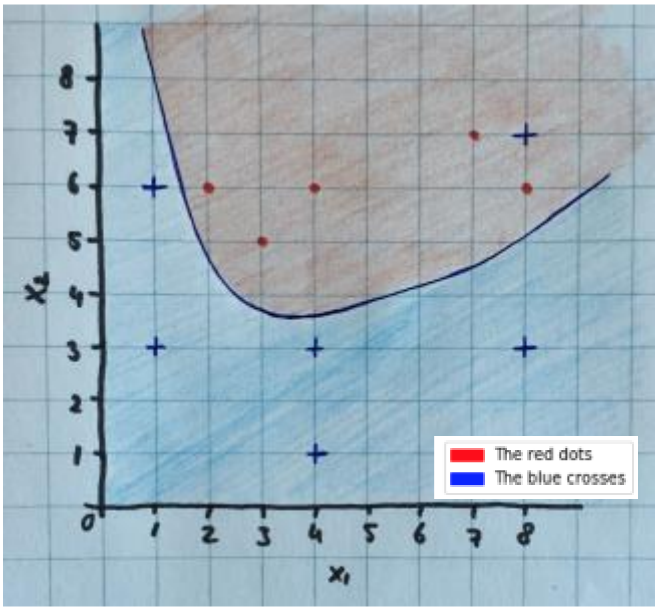

Figure 5: Logistic Regression with quadratic terms decision boundary

Analyses

Judging a model without a test set is difficult, because are not able to fully judge how the model will perform on examples outside of the training set. It is clear to see that both the decision tree and 1-nearest neighbour classifier do well on the given dataset. However, when you look at the quadratic logistic regression in figure 5, you can see that it performs well except from the top left corner where a blue cross is on the wrong side of the decision boundary. However, I argue that this blue data point might be an outlier, which would mean that the quadratic logistic regression preforms best of all the models, because it does not over fit the data. Eventhough I believe k-nearest neighbourhood fits this data set best, I would suggest the quadratic logistic regression as the best decision boundary as it generalizes best and will thus also fit other data.