

Machine Learning 2017/2018: Graded Assignment 2 Part 2

Deadline: 17 November 2017, 23:59.

Consider the dataset:

$x_1 = \{1, 1, 2, 3, 4, 4, 4, 7, 8, 8, 8\}$

$x_2 = \{3, 6, 6, 5, 1, 3, 6, 7, 6, 7, 3\}$

$y = \{0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0\}$

- (a) **(3 points)** Draw the data points in 2D (the class with color or form). Draw the class boundaries that would be found by (1) Decision Trees, (2) 1-nearest neighbor, (3) plain logistic regression and (4) logistic regression with quadratic terms. The drawing should illustrate the differences but does not need to be correct by the millimeter. You are most welcome to use your own or other programs to calculate the class boundaries but is also OK if you make a reasonable approximation without calculating it precisely. Also, you can either have one plot with four boundaries or four plots, one per classifier. Make sure that you clearly indicate which boundary belongs to which classifier.
- (b) **(2 points)** 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. Explore this line of thought. This is a difficult question, but try to be creative!