Machine Learning Exercise Lecture 3

Christian Hilligsøe, Laura Nyrup and Trine Jensen, grp. 234

12-09-2018

Solution Summary

You are given, as the train data, trn_x and trn_y along with their class labels trn_x_class and trn_y_class. The task is to classify the following TEST data.

(a) Classify instances in tst_xy, and use the corresponding label file tst_xy_class to calculate the accuracy

The achieved accuracy is 70.9%

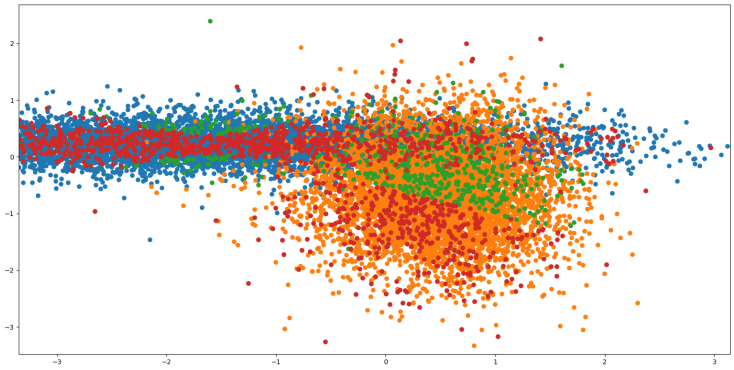

Figure 1: Blue: class 1(trn_x.txt), orange: class 2(trn_y.txt), red: correct classified test points, green: misclassified test points.

The result corresponds to the two classes overlapping, as such point i the overlap is more likely to be miss-classified. Hence a accuracy close to 100% might not be realistic. This theory is supported by the plot illustrating that the density of the green miss-classified point seems higher around the mutual area of the two classes.

(b) Classify instances in tst_xy_126 by assuming a uniform prior over the space of hypotheses, and use the corresponding label file tst_xy_126_class to calculate the accuracy

The achieved accuracy is $83,7\%$

(c) Classify instances in tst_xy_126 by assuming a prior probability of 0.9 for Class x and 0.1 for Class y, and use the corresponding label file tst_xy_126_class to calculate the accuracy; compare the results with those of (b).

The achieved accuracy is $95,7\%$
The better accuracy here correspond to the test data visually having a higher density in the area of class 1.