Machine Learning Exercise Lecture 4

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

27-09-2018

Solution Summary

Download full dataset, which is a Matlab format of THE MNIST DATABASE of handwritten digits: A training set of 60,000 examples, and a test set of 10,000 examples; each having a size of 28 x 28 = 784 (concatenating the pixels).

(a) From the 10-class database, choose three classes (5, 6 and 8) and then reduce dimension to 2

This is done in the attached python file. Figure 1 show the 2. dimensional plot of the training data.

(b) Perform 3-class classification based on the generated 2-dimensional data

We achieved the following percentage of correct classification:

Class 5: 41%
Class 6: 93%
Class 8: 76%

This corresponds to the plot of the data showing that class 5 and class 8 being in the same area, where class 6 is more distinct.
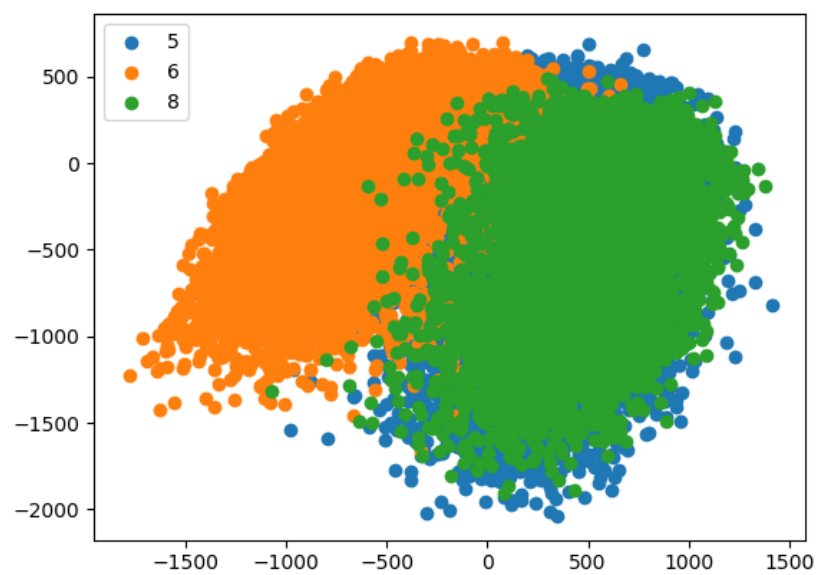
Figure 1: Plot of training set, divided into corresponding classes