Machine Learning Exercise Lecture 4

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

27-09-2018

Solution Summary

This assignment is based on the previously generated 2-dimensional data of the three classes (5, 6 and 8) from the MNIST database of handwritten digits.

a) First, mix the 2-dimensional data (training data only) by removing the labels and then use one Gaussian mixture model to model them.

The matlab data available was firstly downloaded and hstacked in python to make sure that other users have access to the appropriate data. Further more the mean and covariance matrices found for the classification problem in the last exercise ("using PCA for dimensional reduction") was hard-coded into the script for comparison with the values for the unsupervised grouping.

The results for the Gaussian mixture model is shown in figure 1

Figure 1: The data after the unsupervised grouping

For a further visualisation each of the groups are plotted side by side in figure 2

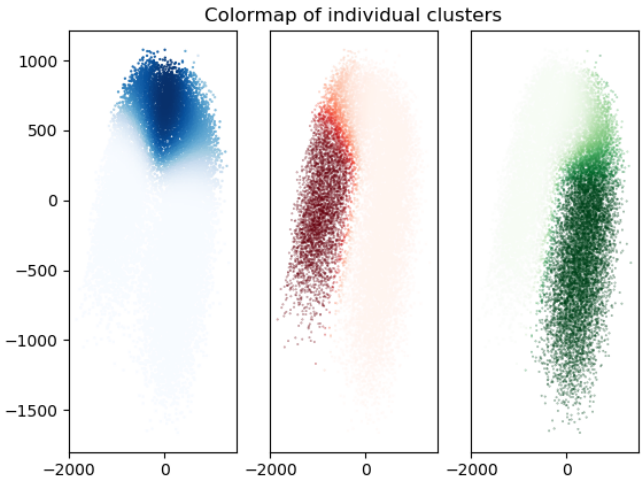

Figure 2: The data visualised with the sepuratue clusters

b) Secondly, compare the Gaussian mixture model with the Gaussian models trained in the previous assignment, in terms of mean and variance values as well as through visualisation.

For a visual comparison between the supervised and unsupervised Gaussian mixture models, the different Gaussian mean values together with an ellipse characterising the 95% confidence interval for these were plotted (Figure 3.

These can give further insight into the behaviour of the different ways the to methods do the classification.

As is seen with the unsupervised grouping the different Gaussian aren't overlapping and group the data set into three sets (A top left, right set and bottom right set). Where the supervised makes two of the Gaussian close to each other.

For data where information isn't as overlapping (as the 5 and 8 is) the unsupervised learning would give better results.

A quick table for summarising the mean

	S_Z1	S_Z2	U_Z1	U_Z2
Group 1	344.77	-466.28	347.86	-300.59
Group 2	-599.03	-157.48	-853.01	27.43
Group 3	478.06	-516.13	103.07	612.08

If wanted an empirical values with the euclidean distance between the mean values could be calculated and the covariances could be further analysed but its deemed that the best representation of the differences for the models is shown in figure 3.

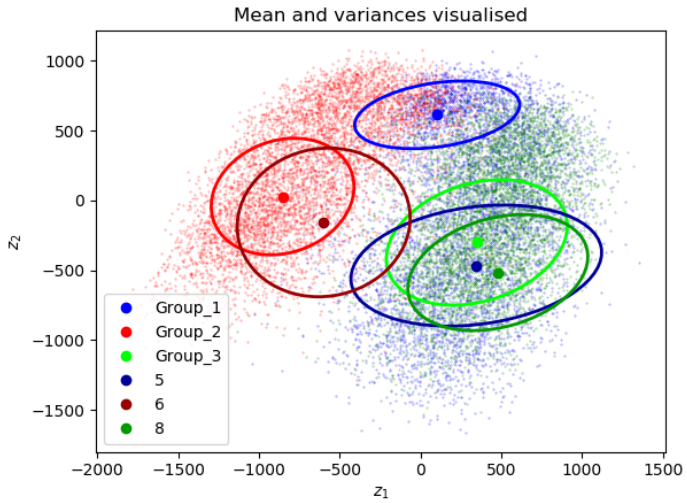

Figure 3: The different mean values together with variance for visualisation. The brighter colours represent the Gaussian with parameters given by the unsupervised grouping, where the darker colours are the Gaussian from the last exercise.