Machine Learning Exercise Lecture 4

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

27-09-2018

Solution Summary

Perform classification for the entire MNIST dataset based on the algorithms introduced

Use LDA for dimensionality reduction to 2 or 9 dimensions, classify the dimension-reduced data and compare this classification performance with that of using PCA.

To do the dimensional reduction the scikit-learn package was importet together with a class for Bayesian classification for ease of classification.
Furthermore the PCA and LDA methods were analysed by seeing how well classification were possible using the different methods. This was further more done for 2 and 9 dimensions.

The results are as following:
LDA with a dimensional reduction to 9 Accuracy for:
All classes is 87.30%
Class 0 is 95.92%
Class 1 is 96.56%
Class 2 is 79.07%
Class 3 is 87.43%
Class 4 is 90.43%
Class 5 is 82.40%
Class 6 is 89.46%
Class 7 is 84.05%
Class 8 is 81.11%
Class 9 is 85.33%

Equivelant PCA dimensional reduction Accuracy for:
All classes is 11.35% Class 0 is 0.00%
Class 1 is 100.00%
Class 2 is 0.00%
Class 3 is 0.00%
Class 4 is 0.00%
Class 5 is 0.00%
Class 6 is 0.00%
Class 7 is 0.00%
Class 8 is 0.00%

Class 9 is 0.00%

LDA with a dimensional reduction to 2 Accuracy for:
All classes is 87.30%
Class 0 is 95.92%
Class 1 is 96.56%
Class 2 is 79.07%
Class 3 is 87.43%
Class 4 is 90.43%
Class 5 is 82.40%
Class 6 is 89.46%
Class 7 is 84.05%
Class 8 is 81.11%
Class 9 is 85.33%

Equivelant PCA dimensional reduction Accuracy for:
All classes is 46.07%
Class 0 is 77.96%
Class 1 is 95.68%
Class 2 is 10.56%
Class 3 is 65.05%
Class 4 is 42.36%
Class 5 is 5.94%
Class 6 is 48.12%
Class 7 is 71.60%
Class 8 is 33.06%
Class 9 is 0.30%

For a visualisation of the differences between dimensional reduction a plot consisting of the 2D dimensionally reduced space for the PCA and LDA is created and shown in figure 1 and 2

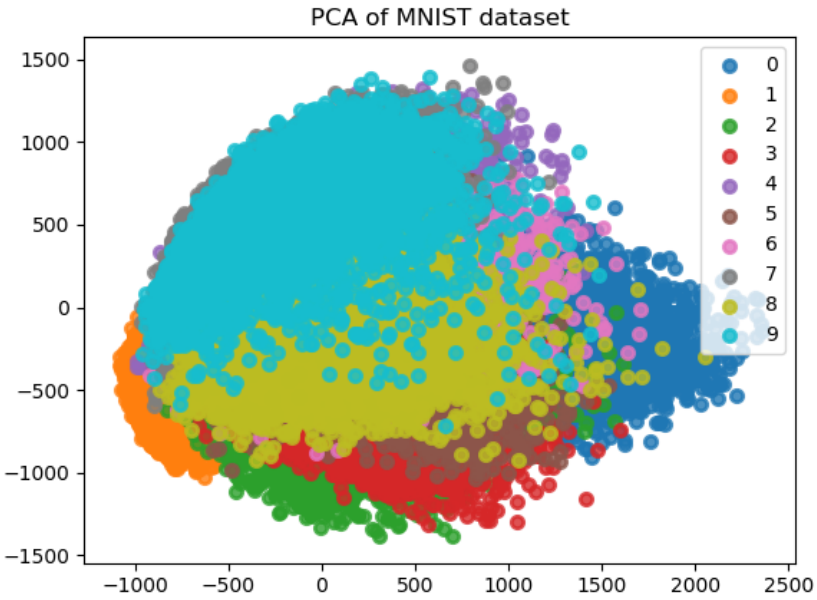

Figure 1: 2D representation of the MNIST dataset using PCA

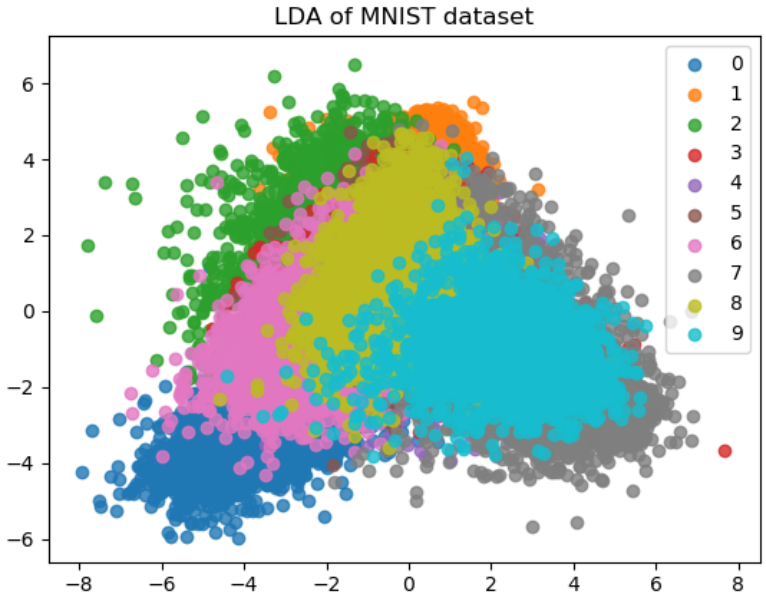

Figure 2: 2D representation of the MNIST dataset using LDA