MACHINE LEARNING FROM DATA

Lab Session 3 – Feature selection: MDA

1. Goal

The goal of this session is to

- Use and compare two methods for dimensionality reduction: multiple discriminant analysis (MDA or Fisher discriminant analysis) and principal component analysis (PCA)
- Test the two methods using synthetic and real datasets

2. Instructions

Getting the material:

- Download and uncompress the file **Mlearn_Lab3.zip**

Handling your work:

- Answer the questions in the document **Mlearn_Lab3_report_surnames.doc**
- Provide complete and concise answers, **maximum 6 pages**.
- Save the report, convert to pdf
- Write the new code in a Colab Notebook **Mlearn_lab3_3_surnames.ipynb**.
- Zip and upload to Atenea the pdf report and the notebook in **a single file**.

3. Introduction and previous study

Read the slides corresponding to lecture 3: feature selection by dimensionality reduction using PCA and MDA.

4. Feature selection on a synthetic Gaussian dataset

Now we will use Colab Notebook **Mlearn_lab3_1_Synthetic_PCA_MDA.ipynb**
This script generates training and test sets with 3-dimensional vectors (d=3) corresponding to three classes (c=3). The classes follow Gaussian distributions, with mean and covariance matrices specified in the code.
A linear and a quadratic classifier are used to classify the data, and error probabilities are computed. Next, the classification is repeated using two features and then using only one feature. The goal is to compare the performance of the classifiers as we reduce the number of features using PCA or MDA.

Run the code.
Note that different values of the seed parameter will produce different eigenvectors for the class covariance matrices.
Use SNR = 10dB.

Q1: Complete the table with the training and test errors for the linear (LC) and the quadratic (QC) classifiers when using three, two and one feature, and SNR=10dB. In this case PCA is used for feature selection. Discuss the results. Analyse the scatter plots in two dimensions and in one dimension.

Repeat the analysis for MDA, using the same seed to generate exactly the same dataset.
Q2: Complete the table with the training and test errors for the linear (LC) and the quadratic (QC) classifiers when using three, two and one feature, and SNR=10dB. In this case MDA is used for feature selection. Discuss the results. Analyse the scatter plots in two dimensions and in one dimension.

Repeat the previous analysis using a different SNR value, SNR= 0 dB:

Q3: Use PCA for feature selection. Complete the table with the training and test errors for the linear (LC) and the quadratic (QC) classifiers when using three, two and one feature, and SNR= 0 dB.
Discuss the results. Analyse the scatter plots in two dimensions and in one dimension.

Q4: Use MDA for feature selection. Complete the table with the training and test errors for the linear (LC) and the quadratic (QC) classifiers when using three, two and one feature, and SNR= 0 dB. Discuss the results. Analyse the scatter plots in two dimensions and in one dimension.

From now on, and to check the effectivity of the previous techniques, we will use a Gaussian dataset where the three class centroids (i.e. the means) are colinear and the three classes have identical covariance matrices. Therefore, the alignment direction of the classes corresponds to the direction of minimum variance.

Now we will use Colab Notebook **Mlearn_lab3_2_Synthetic_collinear_PCA_MDA.ipynb**

Q5. Find and write the three vectors corresponding to the class means. Give also the value of the seed used in your experiments if you change it. How many features can we use with MDA?

Repeat the previous analysis for PCA and MDA and SNR = -5 dB:

Q6. Complete a table with the training and test errors for the linear (LC) and the quadratic (QC) classifiers when using three, two and one feature, and SNR= -5 dB. Use PCA and MDA for feature selection. Discuss the results. In which cases is MDA clearly better than PCA?

5. Feature selection on the Phoneme dataset

5.1. Dimensionality reduction using MDA

In this section we will repeat the analysis done in the previous lab with the Phoneme dataset, but using MDA instead of PCA for dimensionality reduction.

Write the code to use MDA for a number of features d’ ranging from 1 to *dmax* .

Q7. Which is the maximum number of features *dmax*? Show the error curves for the linear and the quadratic classifier on the training and on the test set.

Q8. Compare results and discuss the use of PCA and MDA for the Phoneme dataset