**MACHINE LEARNING FROM DATA**

**Report: Lab Session 3 – Feature selection: PCA and MDA**

**Names:**

Instructions

Getting the material:

* Download and uncompress the file **Mlearn\_Lab3.zip**

Handling your work:

* Answer the questions in the document **Mlearn\_Lab3\_report\_surname.doc**
* Provide complete and concise answers, **maximum 5 pages**.
* Save the report, convert to pdf
* Write the new code in a Colab Notebook **Mlearn\_lab3\_3\_surname.ipynb**.
* Zip and upload to Atenea the pdf report and the notebook in **a single file**.

Questions

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. Analyze the scatter plots in two dimensions and in one dimension.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 3 features | | 2 features | | 1 feature | |
| Test | Train | Test | Train | Test | Train |
| LC |  |  |  |  |  |  |
| QC |  |  |  |  |  |  |

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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 3 features | | 2 features | | 1 feature | |
| Test | Train | Test | Train | Test | Train |
| LC |  |  |  |  |  |  |
| QC |  |  |  |  |  |  |

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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 3 features | | 2 features | | 1 feature | |
| Test | Train | Test | Train | Test | Train |
| LC |  |  |  |  |  |  |
| QC |  |  |  |  |  |  |

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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 3 features | | 2 features | | 1 feature | |
| Test | Train | Test | Train | Test | Train |
| LC |  |  |  |  |  |  |
| QC |  |  |  |  |  |  |

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 changed it). How many features can we use with MDA?

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?

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