**MACHINE LEARNING FROM DATA**

**Report: Lab Session 2 – Feature selection - PCA**

**Names:**

Instructions

Getting the material:

* Download and uncompress the file **Mlearn\_Lab2.zip**

Handling your work:

* Answer the questions in the document **Mlearn\_Lab2\_report\_surnames.doc**
* Provide complete and concise answers, **maximum 5 pages**.
* Save the report, convert to pdf
* Write the new code in a Colab Notebook **Mlearn\_lab2\_3\_surnames.ipynb**.
* Zip and upload to Atenea the pdf report and the notebook in **a single file**.

Questions

Q1. Include the plots of the phoneme spectra.

Q2. Include the error probabilities for the training and test sets obtained with the linear classifier (LC) and the quadratic classifier (QC), using all the features. Discuss the results.

Q3. Include the confusion matrices for the test set obtained with the linear classifier (LC) and the quadratic classifier (QC), using all the features. Discuss the results.

Q4. Which features would you choose? Show the error probabilities for the training and test sets obtained with the linear and the quadratic classifier. Compare with the previous case (using all features) and discuss the results.

Q5. Include the scatter plot and decision boundaries obtained. Discuss the results.

Q6. Show the error curves for the linear and the quadratic classifier on the training and on the test set.

Q7. Discuss which dimension is the most adequate for the linear classifier and which is the best one for the quadratic classifier. Remember that it is important not to overfit on the training data (the test error should not be much larger than the training error).