**MACHINE LEARNING FROM DATA**

**Report: Lab Session 4 – K-Nearest Neighbors and Parzen windows**

**Names:**

**Group:**

Instructions

* Download and uncompress the file **Mlearn\_Lab4.zip**
* Answer the questions in the document **Mlearn\_Lab4\_report\_surname.docx,** save and convert to pdf
* Write the new code in the same Colab Notebook **Mlearn\_lab4\_knn\_parzen.ipynb**

Questions

Q1. Copy the results obtained with kNN on the train and test sets, and discuss the results. What is the value of *k* (default)? Analyze the confusion matrices and identify the two most challenging classes.

Q2. Run again the script, using PCA to reduce the dimensionality of the feature space, selecting d’=64 features. Observe the eigenvectors and the images reconstructed using only the first d’ eigenvectors (those with the highest eigenvalues). Discuss. Copy train and test results. Discuss the results and compare with the previous case (no PCA).

Q3. Repeat the previous analysis using PCA with d’=9 features, and MDA with d’=9 features. Discuss which method is the best for image reconstruction and which one is preferable for classification.

Q4. Find the optimal value of *k* on the training set. Use at least 10 values for *k*. Plot the train and validation errors as a function of *k*. Use the optimal value of *k* to compute the error on the test set. Discuss the results.

Q5. Use the same gridsearchcv strategy (with a single stratified split) to select the best parameter *h*. Try different values of *h* (for example 1, 10, 20, 30…. try more values). Plot the train and validation errors as a function of *h*. Use the optimal value of *h* to compute the error and other metrics on the test set. Discuss the results.