**MACHINE LEARNING FROM DATA**

**Fall 2018**

**Report: Lab Session 3 – K-Nearest Neighbors and Parzen windows**

**Names:**

**Group:**

Instructions

* Answer the questions
* Save the report and upload the file

Questions

Q1. Complete a table with training and test errors obtained for kNN and discuss the results. What is the value of k? 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. Complete a table with training and test errors. 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. Edit the script and modify the code to find the optimal value of k by K-fold cross validation on the training set. Use K=10 folds. Plot the average train and validation errors (average over the folds) as a function of k. Use the optimal value of k to compute the error on the test set.

Q5. Copy the new code

Q6. Use K-fold cross validation (K=10) to select the best parameter h (Parzen windows), where possible values of h are 1, 10, 20 and 100.

Plot the average train and validation errors (average over the folds) as a function of k. Use the optimal value of h to compute the error on the test set.

Q7 Copy the new code

Q8. Use the Microarray dataset. Complete a table with the training and test errors for different values of k=1, 2, 3, 4. Discuss the results.