Lab assignment 3: kNN classifier and evaluation

- o   Task 1: Build a kNN classifier
- o   Task 2: Test the kNN classifiers

Describe everything in a report.

Task 1: Build a kNN classifier

Here you have to create a program (a Matlab function for instance) that takes the following parameters:

1. a set of data, as a n x $d+1$ matrix, to be used as the training set
2. another set of data, as a m x c matrix, to be used as the test set
3. a scalar k

The program should:

1. Check that the number c of columns of the second matrix is at least the number d of columns of the first matrix - 1
2. Check that no entry in any of the two data sets is <1
3. Check that $k>0$ and $k<=$cardinality of the training set (number of rows)
4. Classify the test set according to the kNN rule, and return the classification obtained
5. If the test set has a column number d+1, use this as a target, compute and return the error rate obtained (number of errors / m)

Use the slides to implement the classifier.

Task 2: Test the kNN classifier

Use the Semeion character recognition data (see collection of datasets on the main Aulaweb page).

Compute:

- the quality indexes for binary classification: accuracy or error rate, selectivity/specificity, precision/recall, F1 measure
- on 10 tasks: each digit *vs* the remaining 9
- for several values of k, e.g., k=1,2,3,4,5,6,8,10,15,20,30,40,50