Human-Oriented Robotics Winter term 2013/2014
Prof. Dr. Kai Arras, Social Robotics Lab University of Freiburg
Lab instructors: Timm Linder, Luigi Palmieri, Billy Okal Department of Computer Science

Exercise 6:
k-NN, Cross Validation, Performance Measures

Submission: Send your solution to `palmieri@informatik.uni-freiburg.de` until December 16, 2013 with subject line "[`exercises`] `Sheet 6`". All files (Matlab scripts, exported figures, handwritten notes in pdf/jpg format) should be put into a single zip file named `lastname_sheet6.zip`.

For this exercise, you will need to download a dataset file from the course website. This exercise continues on supervised learning with emphasis on k-Nearest Neighbors classification, cross validation and performance measures.

Exercise 6.1: Classification with k-NN

Load the provided dataset `circular.txt` to Matlab. Implement a k-NN classifier for the dataset given using an initially chosen value of $k = 5$ and using Euclidean distance metric, then perform the following steps.

a) Compute the performance measures TP, FP, TN, FN, precision, recall, F-Measure and accuracy all for $k = 5$.

b) Implement additional distance metrics namely 1-norm, ∞-norm and repeat part (a) above.

Exercise 6.2: Cross-Validation

In this exercise we want to find the best value for k (the number of neighbors) and the best distance metric.

a) Use 5-fold cross validation for $k = \{1, 5, 15\}$ and the three distance metrics from above. Make a 3x3 table showing the nine accuracy measure averaged over the cross-validation runs and highlight the best combination of k and distance metric.

Exercise 6.3: Classification Comparison

Finally, we want to find the best classifier for detecting people in 2D range data as presented in the course. The comparison includes k-NN classification and Support Vector Machines. Download the dataset `peopleofficedata.txt` from Exercise 4 and the starter code of Exercise 5 that contains the SVM classification code.

a) Use 5-fold cross-validation and vary the SVM kernel $\{\text{linear}, \text{RBF}\}$, the stiffness parameter $C = \{50, 350\}$ and σ^2 of the RBF kernel $\sigma^2 = \{1.0, 2.0\}$ as well as $k = \{1, 5, 15\}$. Use the Euclidean distance metric.

b) Find the best classifier in terms of the accuracy measure and show the nine results in a table

c) (Optional) Include the Bayes classifier into the comparison