Submission: Send your solution via email to `palmieri@informatik.uni-freiburg.de` until December 16, 2014 with subject "[`exercises`] `Sheet 6`". All files (Matlab scripts, exported figures, hand-written notes in pdf/jpg format) should be compressed into a single zip file named `lastname_sheet6.zip`.

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

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 given dataset with $k = 5$ and using an 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 the 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 3×3 table showing the nine accuracy measures averaged over the cross-validation runs. 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 motivated in the lecture). The comparison will cover k-NN classification and Support Vector Machines. Use the provided dataset `peopleofficedata.txt` (taken from Exercise 4) and the given SVM classification starter code (from Exercise 5).

  a) Use 5-fold cross-validation and vary the SVM kernel {linear, 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 display the nine results in a table.

  c) (Optional) Include the Naïve Bayes classifier in the comparison.