

60-473/574 Pattern Recognition - Fall 2017
Assignment 2

Deadline: November 3, 2017, at 11:59pm

This assignment must be done **individually**. The main goal of this assignment is that students obtain hands-on experience in using tools for support vector machines and apply them to real problems. For this assignment, use your favorite software package or programming language. Implementations of SVM in Matlab, Octave, Python, C/C++ and Java are available. Note that you **do not** have to implement the SVM. You can use an existing implementation of SVM such as LibSVM in Weka, Scikit, Matlab or Octave.

You will work with the 4 datasets given in Resources (clusterincluster, halfkernel, twogaussians and twosprials).

1. Download a SVM toolbox from any of the sources cited in class (LibSVM or Scikit's SVC is recommended) and install it on your system.
2. Using the SVM tool you installed, run three different classifiers with the following kernels and their parameters: (a) SVM-L: linear kernel; (b) SVM-P: polynomial kernel – degree 2; (c) SVM-R: RBF. Note: you do not have to submit anything for this item, though you have to make sure you use all these classifiers in item #3.
3. Run the three classifiers with *default parameters* on the 4 datasets using 10-fold cross validation, obtaining, for each classifier, the *averages* of the five measures of efficiency seen in class: PPV, NPV, specificity, sensitivity, accuracy, where class 1 corresponds to “positive” and class 2 to “negative”.
4. Compare and comment on the performances you obtained for the three classifiers on the 4 datasets. Provide valid reasons for justifying why the classification is better for some kernels than others on each particular dataset.
5. For SVM-R, plot the ROC curve and find the AUC for each dataset. Note: for constructing the ROC curve, you can run SVM-R with different parameters. Eventually, you can apply grid search to obtain the best parameters (not required for this assignment).
6. Compare and comment on the performance (both accuracy and AUC) of the classifiers.

Submit:

- 1) A report in PDF that includes all the items as required:
 - a) The source from which the SVM tool was obtained, a brief description of the tool and how it was used in your classifiers.
 - b) All measures of efficiency for each kernel (classifier) and each dataset.
 - c) Comparison and comments on measures of efficiency obtained for the different classifiers, kernels, and parameter optimization.
 - d) The plots of the ROC curves.
 - e) Comparison and comments against the classification results.
- 2) The source code or screenshots that show how the classifiers were run.

Note: Missing explanations about your implementations will imply marks deducted!

Upload *a single* Zip file that includes all the files to the Blackboard system no later than the date/time specified as the deadline. Late submissions will be penalized with 10% per day for up to 3 days – after the 3rd day, the mark will be **zero**.