Lab 1: Classification with Scikit-learn

Preliminaries

* Install software as per CSC-480 Installation Instructions
* Practice with Labideas.ipynb  notebook  : Please don’t just press tabs to run the code: read the code and make sure you understand what each line of code is doing.
* Study plot\_classifier\_comparison.ipynb : same comment as above except for the nitty gritty details of the plot representations: you should know the general ways of plotting, but the figures shown are quite sophisticated and you don’t need to waste too much time learning how to plot such nice pictures!

New work:

* Instead of outputting plots as in plot\_classifier\_comparison.ipynb, create three tables (one per data set) that lists the classifier on the y-axis/rows and the metric on the x-axis/columns. The classifiers should be all those used in the comparison. The metrics should be: accuracy, auc, precision, recall, f1-measure.  Once again, don’t waste your time “prettifying” your tables (that work is less interesting than just getting the results).
* Repeat the creation of tables using the same classifiers and metrics, but with three data sets available in scikit-learn: Digits, Wine, Breast Cancer. Instead of using a training and testing set, use 10-fold cross-validation.
  + Take 3 data sets, do the same thing that we did for Iris (lab\_ideas) and do it for all the different classifiers
  + In digits,
* Repeat using three imported data sets from OpenML:  Mushroom,   Spambase and credit approval.

Once you are done with all that, you will be ready to work on Part 1 of the assignment!