Homework #4: Classifiers and their Decision Boundaries

Consider the following classifiers:

1. classification tree of with depth of 1 (so-called a "stump", set the parameter of the tree learner "Limit the depth to" to 1)

2. classification tree of with depth of 3

3. logistic regression

4. random forest with ten trees

5. support vector machines with radial basis function (RBF) kernel (widget called SVM in Orange, check that kernel is set to RBF)

For each of the classifiers above paint:

A. a data set where the classifier finds the "right" decision boundary

B. a data set where the classifier fails to find the "right" decision boundary, but where the decision boundary exists in the sense that you, human, would be able to find it and draw it on the scatterplot

Demonstrate A and B using scatter plots. A minimal workflow that you could use contains the Paint Data, Predictions, and Scatter Plot, plus a learner (say, Classification Tree, receiving the data and passing a classifier to the Predictions). In the scatter plot, you can color the dots by the predicted class and set the shape to represent the true class value. For instance, for the classification tree, the Scatter Plot widget could look something like the following:
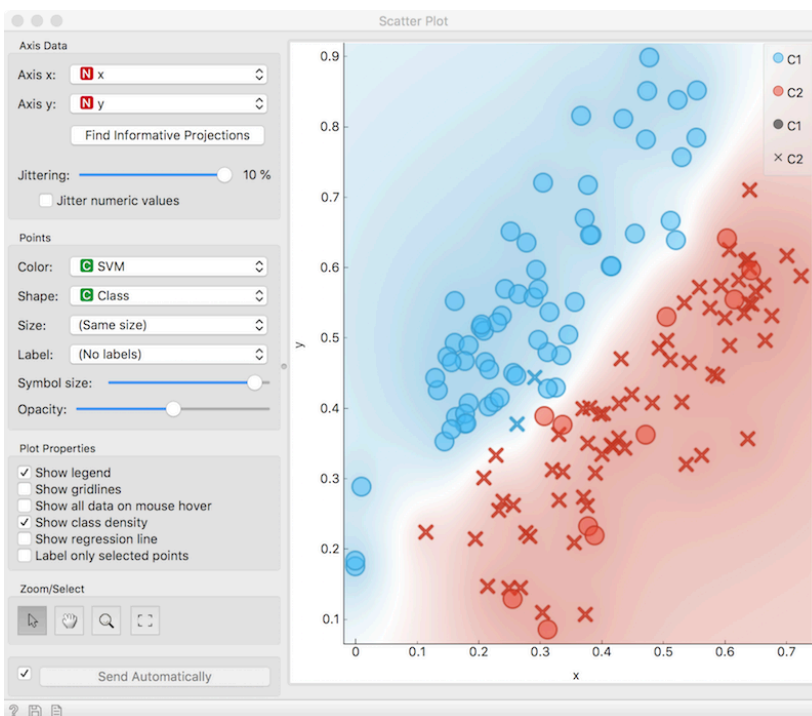
In the homework, show only the graphs (press Ctrl-C or Cmd-C in the widget to copy the image to the clipboard, or save the visualisation using the diskette icon at the bottom of the widget), not the entire widget.

Submit the homework as a short report in PDF. The report should include a title of the homework, your name and email, and scatter plots of successes/failures (A and B) for each of the classifiers. With each graph, you may report also on AUC scores you get using your data and 10-fold cross-validation (use Test &

Score widget). For now, it is sufficient to know that AUC scores close to 1.0 are excellent, and scores around 0.5 are poor.

It may happen that for some of the classifiers you won't be able to paint a data set matching A and B. If this is the case, please provide your intuition why.

The report should not exceed one page. The limit on page length and the limits in the number of paragraphs and sentences are strict. Use 11 pt Calibre or Arial or similar sans-serif font, and 1.2 spacing between lines. Use 6 pt separation between paragraphs.

Submit your report as a PDF document (not Word). Name this document as lastname-firstname-4.pdf (like smith-mary-4.pdf; notice there are no spaces in the name, all letters are lowercase, and the dash is used to separate the first and last name) where last name is your last name and first name your first name.

Submit your homework as a PDF document attachment in the email to bzupan@gmail.com with subject "DM-HW4" (copy the subject title and then paste it into the email title field; notice there are no spaces in the subject title). The deadline is 9:00 am this coming Tuesday, February 15.

Happy painting!