Using `class_tool`– The Assignment 5 Harness

April 12, 2010

The `class_tool` program acts as a harness for assignment 5. `class_tool` reads digits data (provided in cs730/public/digits.data on agate) on standard input and uses this data to train and test your classifier. `class_tool` will run your classifier program (which is specified on the command-line, or by default the program "`class`" will be used) giving it input as specified in the assignment 5 handout. The output of your program will be collected and statistics and information will be available on output.

1 Output

`class_tool` has two forms of output:

1. A confusion matrix is printed to standard output, along with the accuracy of the classifier. The confusion matrix shows the percentage of times that an item of a given label (the row) is classified as a certain label (the column) by the classifier.

2. A (optional) HTML output shows information about the most confident and least confident correct matches along with the most confident false positives and false negatives for each digit. If used on agate, the output will go to a website which you can browser to. The URL for this site is based on your username on agate and it has the form:

   `http://cs.unh.edu/~cs730/asn5/<username>`

2 Arguments

`class_tool` has a number of arguments:

seed Specify the seed for the random number generator (the default is to use a seed value of 1).

maxsize Specify the max size of training data for the trials. This option is useful for debugging when you do not want to wait for all of the trials to complete.

quiet Do not print output for each trial to standard output.

html Output HTML.

htmldir Specify the root directory for the HTML output. This option cannot be used with the version in cs730/public on agate. The default is to output to the cs730 website on agate in a directory with your username. If you use `class_tool`on your own machine, you should specify a different "htmldir".

3 Using `class_tool`

To use `class_tool` on agate.cs.unh.edu, use the following command:

```
~cs730/public/class_tool [-<class_tool_arg> ...] [<agent_arg> ...] < ~cs730/public/digits.data
```

i.e.

```
~cs730/public/class_tool -seed ‘date +%M%S‘ -maxsize 500 \
                -agent ./myagent nb < ~cs730/public/digits.data
```

will run the program `myagent` with a seed based on the current minute and second for all trials with less than or equal to 500 instances of training data using naive Bayes.