CIS 419/519 Introduction to Machine Learning
Assignment 3 Submission Instructions

Check to make sure....

- For the naive Bayes implementation, when you return the class probability distribution for an instance, make certain that the classes are in-order (i.e., the 0th entry of the probability distribution vector should correspond to class 0, the 1st entry should correspond to class 1, etc.)

Files

You will be submitting only the following files, which you created or modified as part of the homework:

- `README`
- `hw3-PENNKEY.pdf` (a PDF of your homework writeup)
- `boostedDT.py`
- `predictions-BestClassifier.dat`
- `predictions-BoostedDT.dat`
- `naiveBayes.py`
- `README`

Please follow the naming conventions exactly, and do not submit additional files (since any additional or misnamed files will be automatically deleted), including the test scripts or data sets.

Your PDF writeup of Homework 3 should be named `hw3-PENNKEY.pdf`, where "PENNKEY" is your own PennKey (for example, my file would be named "hw3-eeaton.pdf").

Submission Instructions

To submit your homework, transfer all of your files into your SEAS account on eniac. Then, from a command prompt on eniac, use the `turnin` command to submit all files:

```
turnin -c cis519 -p hw3 README hw3-PENNKEY.pdf boostedDT.py naiveBayes.py
predictions-BestClassifier.dat predictions-BoostedDT.dat
```

Note that even students in CIS 419 should submit to the cis519 course, as shown in the command above (we have only one course account, cis519, for both sections of the course). Make certain that your submitted files are named correctly, and do not submit additional files, or place the files in a folder.

You can check that your submission was received by running:

```
turnin -c cis519 -p hw3 -v
```

to list the files submitted. You should see one listing for each file you submitted. The turnin command will also send you an e-mail with a subject of "cis519: submission received" to verify receipt.

Subsequent submissions will overwrite earlier submissions; if you re-submit files after the due date, your submission will be counted as late regardless of whether or not the file contents changed. The official timestamp of your submission will be the most recent (i.e., latest) timestamp of ANY file you submit. Therefore, if you submit even one file late, your entire submission will be considered late.

Automatic Testing

Once you submit your assignment, it will be picked up by the autograding engine and queued for a set of automatic tests. The autograder will run these tests and then send you a second e-mail with the results once they complete. Please be aware that this is only a limited set of tests and that we cannot provide any further detail than what is given in the report. For final grading, we will run your code on entirely new data sets with a more extensive battery of tests. We will not be releasing grading information (i.e., "10 out of 10 points") until after the homework deadline.

If the e-mailed report shows that your code failed the tests or that you were missing any files, you may modify your implementation or writeup and resubmit it. You may re-submit all files as many times as you would like; each time the autograder will run the tests and e-mail you the results. You can continue to use these comments about possible failures, crashes, or incorrect results to improve your code up until the homework deadline.

Please keep in mind that the TAs have developed the system and have tested it, but we are likely to find bugs once everyone in the class is submitting their assignment. If you find that the e-mailed reports contain indications of bugs in the testing script (e.g., your code fails a test that you're certain it should pass), please let us know via a private message on Piazza to the instructors only. Be sure to include the e-mailed testing report and describe the exact problem. Please be assured that you will receive credit for correct submissions, regardless of what the initial testing reports say. We will work to correct all bugs and make certain that the grading is accurate.