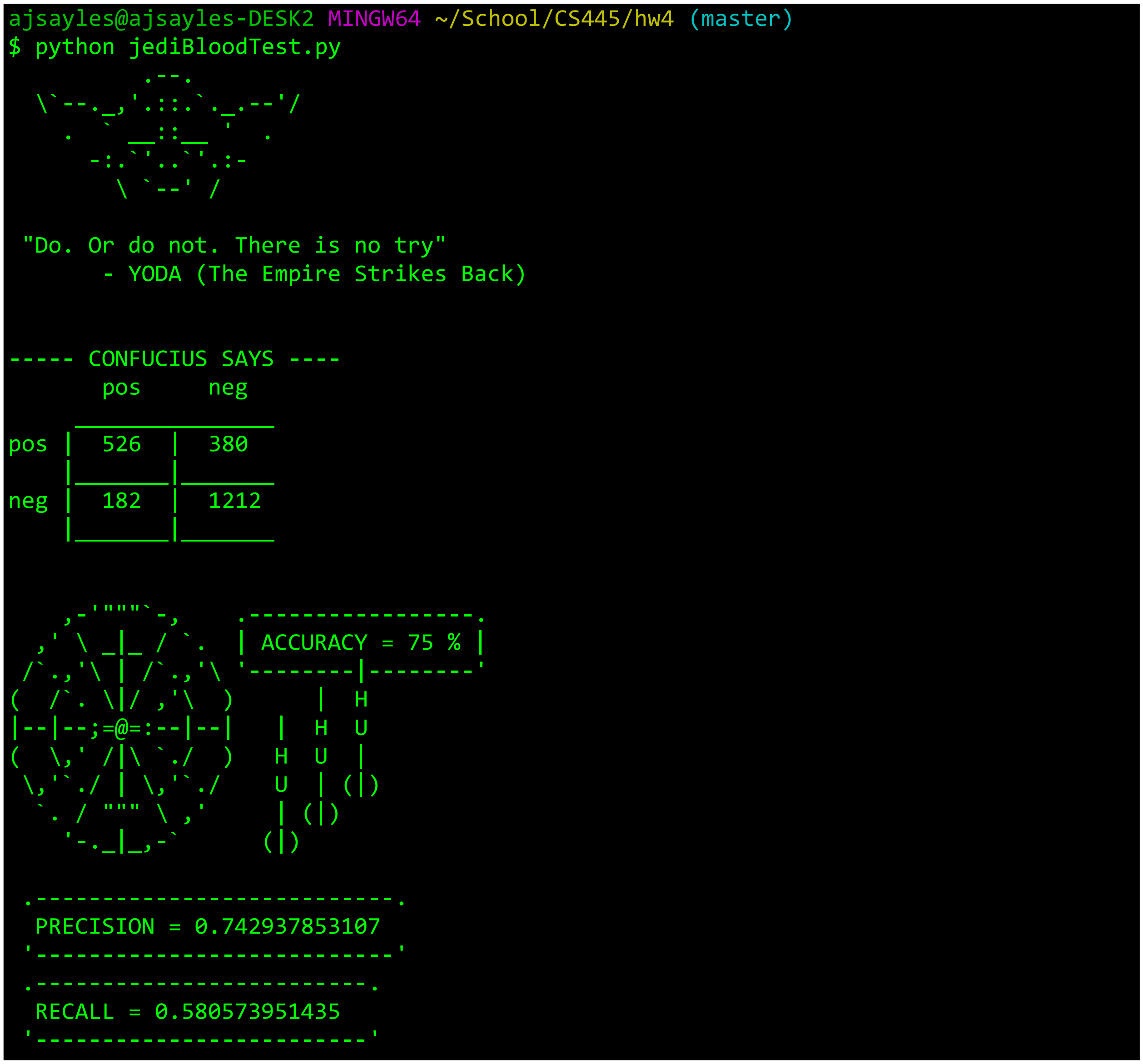
Armondo Sayles
Portland State University
CS445, M. Mitchell
23 February 2016

Homework 4: Naive Bayes Classification

This assignment required making Gaussian naïve Bayes classifying program. A model can be built based on the assumption that the continuous data has a Gaussian distribution. First the data is segregated into two sets based on classification. After that it is aggregated into two similar sets for training and testing Then some statistical analysis is done to compute the mean and standard deviation for each feature in the training set. Finally the probability distribution is performed to decide the class of each instance in the test set.



```
ajsayles@ajsayles-DESK2 MINGW64 ~/School/CS445/hw4 (master)
$ python jediBloodTest.py
          .--.
  \`--._,'.::.`._.--'/
    .  ` __::__ '  .
      -:.`'..`'.:-
        \ `--' /

 "Do. Or do not. There is no try"
       - YODA (The Empire Strikes Back)


----- CONFUCIUS SAYS ----
       pos     neg
     _______________
pos |  526   |  380
    |_______|_______
neg |  182   |  1212
    |_______|_______


    ,-'"""`-,     .-----------------.
  ,' \ _|_ / `.   | ACCURACY = 75 % |
 /`.,'\ | /`.,'\  '--------|--------'
(  /`. \|/ ,'\  )       |  H
|--|--;=@=:--|--|    |  H  U
(  \,' /|\ `./  )    H  U  |
 \,'`./ | \,'`./     U  | (|)
  `. / """ \ ,'      | (|)
    '-._|_,-`       (|)

 .---------------------------.
  PRECISION = 0.742937853107
 '---------------------------'
 .-------------------------.
  RECALL = 0.580573951435
 '-------------------------'
```

Armondo Sayles
Portland State University
CS445, M. Mitchell
23 February 2016

Do you think the attributes here are independent, as assumed by Naïve Bayes?

No, I would expect a higher accuracy as the attributes become more independent.

Does Naïve Bayes do well on this problem in spite of the independence assumption?

Sure. A 75% accuracy could be considered doing well. With the accuracy going as low as 60% I would begin to question the “wellness” of this classifier.

Speculate on other reasons Naïve Bayes might do well or poorly on this problem.

There is another assumption with the method we implemented, in that we assumed a Gaussian distribution of the continuous data. If the data isn’t distributed as such, this model would be invalid.