Dallas Delaney - ddelane2
Doug Zhu - dzhu10
Erik Delanois – delanos2
3 Credits

A.I. MP3 Report

Problem 1

IMPLEMENTATION

In this problem we first read in all of the training data in order to build our $P(F_{ij} \mid \text{class})$. Once we were able to build this probability for every pixel in every numeric character, we were able to calculate our P(class) based on the frequency of occurrence of each character. Once all of this was stored in a database we were able to calculate $P(\text{class}) \cdot P(f_{1,1} \mid \text{class}) \cdot P(f_{1,2} \mid \text{class}) \cdot \ldots \cdot P(f_{28,28} \mid \text{class})$ with the logarithmic adjustment for each image in the test data for each class. The class that returned the highest value would be our estimate. We then saved the most and least prototypical instances of each class by selecting the highest and lowest $P(\text{class}) \cdot P(f_{1,1} \mid \text{class}) \cdot P(f_{1,2} \mid \text{class}) \cdot \ldots \cdot P(f_{28,28} \mid \text{class})$ values. After this we were able to calculate the percentage of the test images that were classified properly along with the confusion matrix. We were able to achieve 77.1% correct classification of the images so we are confident in our design. An interesting fact about our approach is that it is very dynamic and flexible. You can make this program run on images with any dimensions you want, with as many characters as you want, and with as many training/test cases as you want and it will calculate all the same output data. This is because we defined everything in terms of constants located at the beginning of the DigitClassification.h file so to change any of the attributes mentioned above you simply need to change the corresponding constant to the appropriate value. This made doing the facial classification much simpler as we only had to edit the constant values.

```
Smoothing Value = 1

Character      Correct               Attempts              Percentage

0              76                    90                    84.4444%
1              104                   108                   96.2963%
2              80                    103                   77.6699%
3              79                    100                   79%
4              83                    107                   77.5701%
5              62                    92                    67.3913%
6              69                    91                    75.8242%
7              77                    106                   72.6415%
8              61                    103                   59.2233%
9              80                    100                   80%

Total          771                   1000                  77.1%


Confusion Matrix:
      0        1      2       3      4       5      6       7      8      9

0     84.444   0.000  1.111   0.000  1.111   5.556  3.333   0.000  4.444  0.000

1     0.000    96.296 0.926   0.000  0.000   1.852  0.926   0.000  0.000  0.000

2     0.971    2.913  77.670  3.883  0.971   0.000  5.825   0.971  4.854  1.942

3     0.000    2.000  0.000   79.000 0.000   3.000  2.000   6.000  2.000  6.000

4     0.000    0.935  0.000   0.000  77.570  0.000  2.804   0.935  1.869  15.888

5     2.174    2.174  1.087   13.043 3.261   67.391 1.087   1.087  2.174  6.522

6     1.099    5.495  4.396   0.000  4.396   6.593  75.824  0.000  2.198  0.000

7     0.000    5.660  2.830   0.000  2.830   0.000  0.000   72.642 2.830  13.208

8     0.971    0.971  2.913   13.592 1.942   7.767  0.000   0.971  59.223 11.650

9     1.000    1.000  1.000   3.000  9.000   2.000  0.000   2.000  1.000  80.000

For Character '0' this is the most prototypical image:



              +#++
            +#####+
           +########+
          +##########+
         +######+ +##+
        +####+++  +###+
        ####+      +##+
       +###+        ##+
       ####+        ##+
       ###+         +##
       ###+         ##+
       ###+         ##+
       ###+        +##+
       ###+       +##+
       ####+     +###+
       +###++   +####+
        #####++#####+
        +##########+
         +########+
           ++++++
For Character '0' this is the least prototypical image:
```

```
    ++++++#+++
   +##########+
 +##############
+###############+
########++  ++###+
#######+      ####
  +###+       +###+
  +###         +###+
  ####          ###+
  ####          ####
  ###+          ####
  ####          ####
  ####+         ####
  +####+        ####
   +####        ####
    +###++      ####
     +#####++++#####
      +############+
       +###########+
        +++#######
```

```
For Character '1' this is the most prototypical image:
```

```
    +#+
    +##+
    +##+
    ###+
    ###+
    ###
   +##+
   +##+
   +##+
   ###+
   ###
  +##+
  +##+
  ###+
 +###
 +##+
+###+
+####
+###+
 +#+
For Character '1' this is the least prototypical image:
```

```
               +#
               ##+
              +#++
              ++#+
             +#+#+
            +#+++
            ##+#+
           +#++#+
          +#+ ++
         +#+ +#+
        +#+  ++     +
         +##+#########++
          ++##+++++++++#+
            #+
           +#+
           +#
           +#
           #+
           #+
           #+
```

```
For Character '2' this is the most prototypical image:
```

```
          ++++++
          ######++
         +########+
         +#++  +###+
         ++     +##+
                 +##
                 +##
                  +#
                  +#
                  +#
                 +#+
                 +#+
           +++++ ##+
         +######+##
        +#########+
        +##++++#####+
        ##+  +###+####++
        ##++####+  +##+
        +#####++
         +##++
For Character '2' this is the least prototypical image:

          +++#+
        +#####+
       +##+++
       ##+
      +##
      +#+
      +#+
      ##+
      +#+
      +#+
       ##
       ##+         +++
       +##       +#####+
        +#+     +###++##+
         ##+    ###   +##
         +##+  +##+    +#+
          ++##+ ##+    +#+
            ++#####+  +##
              +##########
                 +++++++
```

```
For Character '3' this is the most prototypical image:


         +#####++
         +########+
            +++####+
                ++##
                 +##+
                +##+
              ++##+
             +###+
            +###+
           +#####+
           ###+###+
           ++  +##+
                +#+
                +##
                +##
               +##+
              ++##+
     +++++++++###+
     +##########+
      +++#####++
For Character '3' this is the least prototypical image:



          +++++++++#+
       ++############
      +############++
      +##++++++++++
      ##+
      +#+
      +##+
      +##++++++++++++
      +##############+
       +##############+
        +++++++ +++ +##+
                    +##+
                    +##+
                    +##+
       ++           +##+
       #+           +##
       ##+         +##+
       +##+++++++++###+
         +###########+
          ++#######+


For Character '4' this is the most prototypical image:




           +       +#
          +#+      +#
          ##+      +#
         +##+    +++#
        +###     +##+
        +##+     +##+
        ###      +##+
        ##+     +###+
        #+    ++####+
      ++####+######
       +++#######++
          +++++##+
              +#+
              +#+
              ##+
             +##+
             +#+
             +#+
             +#+
             +#+
For Character '4' this is the least prototypical image:



              +##++++
          ++#########+
        ++############+
     ++###########+####
    +#######++#+++ ++##
   +####+           +#+
   ####+            ##+
   ###+           ++##+
   ###+        +++####+
   #####++++##########+
   +###################
    +##########++  +###
      ++++++++     +###
                   +###
                   +##+
                   +##+
                   +###
                   +##+
                   +##+
                   +##+

For Character '5' this is the most prototypical image:



                   ++#+
            ++++++####+
            #########+
            ###+++++
           +#++
           +#+
          +##
          +#++++
         +######+
         +##+++##
          +    +#+
               +#+
               +#+
                ##
       +#+      ##+
       +#+     +##
       ##+     ###
       +##+   +##+
        ##+++###+
           ###++
For Character '5' this is the least prototypical image:




           ++############
       ++++#############+
      +##+######++++++
     +#####++
     +###++
     +###++++++
      #########++
      +++##########++
         +++++#######++
              +++######+
     ++           ++####
     ##+             ###+
     ###+           +###+
     +###++        +####+
      +####++   +++####+
       ++#############+
         ++#########++
           +++###+++



For Character '6' this is the most prototypical image:


                +#+
              +###+
             +###+
            +###+
            +###+
           +###+
          +###+
          +##+
          +##+
         +###+
         +##+
         +##+   ++##+
        +###+  +####+
        +###++######+
        +##+ +######+
        +###+#######+
        +#########+
         +#######+
          +#####+
            +##+
For Character '6' this is the least prototypical image:


       #+
      +##+
      +##+
      +##+
      +##+
      +###
      +###
      +###
      +###
      +##+      +++++++
      +##+     +#######+
       +#+    +####+###++
       +##    +##++  +##+
       +##+  +###+   +##+
        ##+  +##+    +##+
        +###++##+   +###+
          #######+ ++###+
          ++######+####+
             ++#######++
                +++++




For Character '7' this is the most prototypical image:





        +#++ ++++++
        ###########+
        ++###+++++#+
          +++    +#+
                +##+
                +##+
               +##+
               +##+
               +##+
              +##+
              +##+
              ##+
             +##+
             ##+
            +##+
            +#+
           +##
           +##
           +##
           +++
For Character '7' this is the least prototypical image:




    ####++##+
   ###########+
   ############+
   +############+
    +++#########+
       ++++++####++
            ++#####+
             ++#####
              ++####
               ++###+
                +####+
                +#####
                +#####
                +####+
                +####+
                +####+
                 +###
                 +###+
                 +####
                 +#####



For Character '8' this is the most prototypical image:


                +##+
              ++####+
            ++######+
           +####++##+
          ++###+ +##+
          +##+   +##+
          ###     +#++
          ###    ++###
          +##++++###++
          +#######++
           #####++
          +#####
         +######+
         ###++##+
        +##+  ###+
        +##+  ###+
        +##+ +###
        +#######+
         +######+
          ++##++
For Character '8' this is the least prototypical image:
```

```
     +++#####+
   +####+++###++++
 +####++    ++###+
+###+         ###+
+##           ###
+##           ###
+###    +++++###+
 +###+++########+
  +######++++++++
   +####       +
  +#####+
  +#+++##++
  ##+  +###+
  ##+   +###+
  +#++   +###+
   ###++  ++###+
   +####+   +###+
     +####+++###+
       ++#######+
         +++###++
```

```
For Character '9' this is the most prototypical image:
```

```
     ++##++
    +######+
  +####++++#+
 +###+   ++##
 +##+   +####
+##+    +####
+##+    +###+
+##+  ++####
 +#########+
  +#######+
   +++++##+
      +##+
      +##+
      +#+
     +##+
     +##+
     +##
     +#+
     +#+
     +#+
For Character '9' this is the least prototypical image:




       ++++####++
     ++###########++
    +###++++++++####+
    ##+          +###+
   +##+         +####
   +###++     +#####+
    +#####++++##++++
     ++########+
        ++######++
          +#######+
          +#+  +###+
          ##+   +###+
         +##+    +###+
          ##+      ##+
          +##      ###+
          +##+     +##
           +##+    +#+
           +###++++##+
            +#######+
              +##+++

This is the odds ratio for 0/5 :



          +++++++++ -----
        --++++++++++ -----
      ----++++++++++ ------
    + --- ++++++++++ -------
    +--- ++++++++++++ ------
     -  ++++++++++++++ -----
    -++++++++  +++++++++----
    -+++++++   - ++++++++--
    -+++++  ----- +++++++-
    ++++++ -------+++++++++
  + ++++++---------++++++++
    ++++++---------+++++++
   -++++++---------++++++
   -+++++++------- ++++++
   -+++++++-------+++++++-
  - ++++++++ ----++++++++
  --++++++++    +++++++++
  --++++++++++++++++++++-++
    +++++++++++++++++++--++
   +++++++++++++++  + ---++
     ++++++++++  --------
     -------------------
     ---------------
      -------------
```

```
This is the odds ratio for 5/3 :
++++++++++++++++++++++++++++
++++++++++++++++++++++++++++
++++++++++++++++++++++++++++
+++++--------------+++++++++
+++----------------+++++++++
+++--------------- +++++++++
+++--------------- +++++++++
+-+---------- ----- ++++++++
+++------- +++++ --- +++++++
++------+++++++ -----+++++++
++---- +++++++ -------++++++
+++-+++++++++ ---------+++++
++++++++++++ ----------+++++
+++-++++++++ -------  ++++++
++++++++++++ --------+ +++++
++++++++++++   -------++++-+
++--++++++++++++------ +++++
++--- ++++++++++ ------+++++
++--- ++++++++++  -----+++++
++---- +++++++++ ------+++++
++---- +++++++  -------+++++
++-----       ---------+++++
++-------- ----------- +++++
++------------------ +++++++
+++-----------   +++++++++++
+++------ +++ ++++ ----+++++
++++++++++++++--+++---++++++
++++++++++++++++++++++++++++


This is the odds ratio for 8/3 :



     -----------------
   -----------------+++++
   ------------ +++++++++++
   ----------     ++++++++++
 - --------  +      ++++++++
   ------++++++ --  ++++++++
  ----- ++++++ ----++++++++
  -- +++++++++---- ++++++++
   -+++++++++ --- ++++++++++
    ++++++++ --- +++++++++++
   -+++++ -         ++++++
    -+++--- +++++------+++
   --- -- +++++++------ + -
  ----- +++++++++-------+
  -----++++++++++-------+
  -----++++++++++-------+
  ----- ++++++++ -------+
  ------+++++  ---------+
  -------     ---------++
  --------     --- ---+++
  --------       ++  ++++
   ------ ++++++++++++ -
   -----------++++++ --
          ----------- ++
This is the odds ratio for 8/9 :



          +++++++++++
     + ++++++++++++++++++
      +++++++++++++++++++++
    + ++++++++++++++++++++++
    ++++++++  --- ++++++++++
     +++++ -------- ++++++++
    + +  ---------- +++++++
     ------- +++ -  +++++++
    ------- ++++    ++++++++
   --------+++++  --- ++++++
   --------+++++ -----++++-
   -------- ++++ ----- ++ -
   -------- +++ -------++--
   ------- ++++ ---- ++++--
  -- -- ++++++ ---- +++++
  --++++++++++---  ++++++
   -+++++++++ ---++++++++
   -+++++++++   ++++++++
    ++++++++++++++++++++-
    +++++++++++++++++ ++-
    -+++++++++++++  -----
    -++++++++++  --------
    ---------------------
   --------------------
            ----- ---
```

Problem 2

IMPLEMENTATION

This part of the MP was done in Python.
In order to implement problem 2.1, we first had make sure we had all of our word likelihoods saved somewhere. In order to do this, we saved the likelihoods of each word from the training data into a dictionary based on whichever label it had. Then we went through each document of the test data and classified it into one label. We had to go through each line and find the sum of the likelihood that the document would fit a certain label. For multinomial naive Bayes we calculated the sum of every word's likelihood and frequency for each label. The email was then classified as whichever likelihood was greater, and the confusion matrix was updated. To implement the Bernoulli Naive Bayes problem, we commented out the line where the likelihood accounted for the word frequency. The classification rate for both models and both datasheets were the same (email - 97%, movie - 75%). To have our code work for the movie datasheet, we just had to change the way we searched for labels by changing a number in our conditional statement (labels were -1 and 1 instead of 0 and 1). To find the classification rates we divided the number of documents we labeled correctly by the total number of documents.

CONFUSION MATRIX

Bernoulli:

	Not Spam	Spam
Not Spam	126	4
Spam	2	128

Classification Rate: 97.6%

	Good Review	Bad Review
Good Review	382	118
Bad Review	126	374

Classification Rate: 75.6%

Multinomial:

	Not Spam	Spam
Not Spam	126	4
Spam	2	128

Classification Rate: 97.6%

	Good Review	Bad Review
Good Review	383	117
Bad Review	124	376

Classification Rate: 75.9%

The top 20 for words for the emails are listed below. They are in the format word_(# of occurrences)

TOP 20 WORDS FOR EMAILS

TOP SPAM WORDS
email_1380
s_1207
order_1159
report_1053
our_965
address_954
mail_923
program_828
send_800
free_744
money_722
list_713
receive_662
name_627
business_608
one_553
d_541
work_528
com_524

TOP NORMAL WORDS
language_1130
university_906
s_661
linguistic_477
de_445
information_444
conference_378
workshop_360
email_321
paper_320
e_314

english_312
one_280
please_278
include_277
edu_271
http_264
research_259
abstract_253

TOP 20 WORDS FOR MOVIES

The top 20 for words for the movies are listed below. They are in the format word_(# of occurrences)

TOP NEGATIVE REVIEW WORDS
movie_290
film_227
like_163
one_143
--_114
bad_87
story_85
much_83
time_75
even_70
characters_64
good_64
little_62
would_58
comedy_57
never_53
nothing_52
makes_51
plot_51

TOP POSITIVE REVIEW WORDS
film_285
movie_187
--_136
one_111
like_99
story_94
good_84
comedy_83
way_80
even_76
time_73
best_72

much_66
performances_62
funny_60
make_60
life_58
us_58
makes_58

<u>EXTRA CREDIT</u>

Problem 1 extra credit: Facial Recognition:

*see description of problem 1 for details as it has the same implementation.

```
Smoothing Value = 10

Character      Correct               Attempts              Percentage

0              54                    77                    70.1299%
1              72                    73                    98.6301%

Total          126                   150                   84%
```

```
Confusion Matrix:
     0        1

0    70.130   29.870

1    1.370    98.630
```

```
For Character '0' this is the most prototypical image:

  ####    #
       ### ###########

                                            ####
  ##                                       #    ##
 #  #                                     #   #   #
  ##                                       ###     #
                                                    ##
                                             #        #
                                             #   #####
                                        #    # ##
                                      ## #    #
                                     #    #  #
                                      #    #
      ##                  ###  ##
 #####  #                #    #  #     ####
      #  #               #    # #    ##    #
 ###   ##                 ###  ##   #  #    #
                                 # #   #     #
                                   #   ######
                                    ###        ##
                                                 #
                              ####          #     #
                             #    ##         #####
                              #     #
                               #   #
                                #  #
   ##                           ###
  #  #                    ##  ##   ##
   ## #                  #  ## #
                         #   # #
     ###               ##   #  #   ###
 ####   #            ##  # #   #  #   #
   #    #  #      # #     #     ##    #
   #    #  #     #       #        ####
 #######    #    #  #####
          ##      #         #
                           # #
                          #   #
               ####  ##  #   #
              #     #  #  # #
              #  #  #  #   #
      #        ##   #  #        #
     # #             ##        # #
     #  ###                   #   #
     #                  #     #  #
      ###            #  #      ##
         #           #  #
                     #  #
                     #  #
  ##                  ##
 #  #
  # #
   #

              ##
             #  #                             ##
        ##   #  #                            #  #
       #  #   # #             ####          #    #
       #  #   #  #    ##     #    #          #   #
        ##    #  #   #  #   #   ##           # #####
              #  #   #  #   #  #              #     #
                 #    ##     ##              #      #
              ###                       ##    #  ###     ##
             #  #                      #  #    ##       #
              ##            ###         ##               ##
                           #   #####
                           #        #

For Character '0' this is the least prototypical image:

 ########  ################################################
         ##
     #
     #  #                #  #             #  #            #
     # #   ##            #  #     ####    #   #           #
      ##  #  #            ##     #    # # #    ########   #
 #    # ##           #    #      #   #   #           # #  #
 #   #    #         # #           ###         #    ##   # #
 #  #      #   ##  #   #    ##   #  #   ##  ## #### #     #
  #       #   #   #   #    #  ## #  #  #  ##         #    #
 #        #  #   #   #     #       #  #  #            ### #
 #         ###   #   #     #      #  ####            #    #
 #         #  #  #  #       ######                  #     #
 #         #   # #  #                              #      #
 #         #    #   #     #                        #      #
 #         #        #    #           ###  ###      #      #
 #          #  #     ## #     #######   #    # #  # #     #
 #            # #      ##    #   #   #  #   #   ##  #     #
 #         # #      #  # #    # #       #   #        #    #
 #        #  #      #  #  #    # #######   #   #      #   #
 #         ##  #####  ####        #    # ##     ####   ## #
 #           ##                  #  #  #            #     #
 #                             ##   ###  #   #      #  #  #
 #              ##         ####    #   ##  ## #      ##   #
 #             #  #       #        #   # #  # #     #     #
 #             #  #      # #  ###  # ##  #   #     #  #   #
 #            #    #    #   ##      #  ##   #     #    ###
 #        ##            ##   #      #  #   #    # #    #
 #       #  ###        #  # #       # #    #  ##  ###  #
 #       #     ########   #         #      #    ##   ##   #
 #       #   #        #    #         #      #       #     #
 #        #  ### #     #  #           #   #  ##  ###      #
 #         ##   #         #        #  #   #    #          #
 #               #       #          # #  #    #     ######
 #               # ####  #         #  #  #    #    #      #
 #               #     ##           ##   #   #    #     # #
 #               #  #    #          #     ###     #  # #  #
 #               #   #   #                        #   #   #
 #         ##   #     #  #              #         #   #   #
 #        #  #  #      ##                #     ### #  #
 #       #  #    ####                 #   ### #  #    # ##
 #       #  #        #                #        # #   ##   #
 #        #    #  #     #              ###       # ##  #  #
 #             #   #   # ##               ### #         ##
 #              #  #  #    #                 ## ########  #
 #               ##    ####                    #          #
 # #             #           #                         #
    ##           #  #      ##                          ###
      #          # #      #  #                  #######   #
  ####            #       #   #                #     #    #
                           #  #               #      #
         #                  ##               #   ######
          # #            #                  #   #
          #  ###         #   ###     ###     ###         #
          #     #      # #  #   ## ##   ##              #
          #      #    #  #   #    # #     ##    ########
           #    #     #   #   ##    #  # #    ##
            ####      #    ##       #  #   ###    #  ## # #
            #         #   #         # #   #   #   #     #
            #         #  #         ###   #    #   #
            #          #              #  #   # ###       #
            #              #            #   #     #     #
            #            ## #       ####   #       #  ##
         ######       ###    #######       #        #   ##
        #                   #              #        #     #
        #                                           #
       #
 ### ######        ########################################

For Character '1' this is the most prototypical image:










     # #
     #  #
     #  #
     #   ##
    #      #
    #       ##
    #         #
     ####      #
         ##     ##
           #      #
            #      #
      ##     ##     #
        #      #     ##
         #      ##     #
           #      #     #
            ##     #     ##
              ##    #      ##
                     ##      ##
                       #       #
                        #       #
                         ##      #
                           #      ##
                            #       ##
                             #        #
                              ##       #
                                #       #
                                 ##      ##
                                   #       ##
                                    #        #
                                     #        #
                                      ##       #
                                        #       #
                                         ##     #
                                           #     #
                                            #####
```

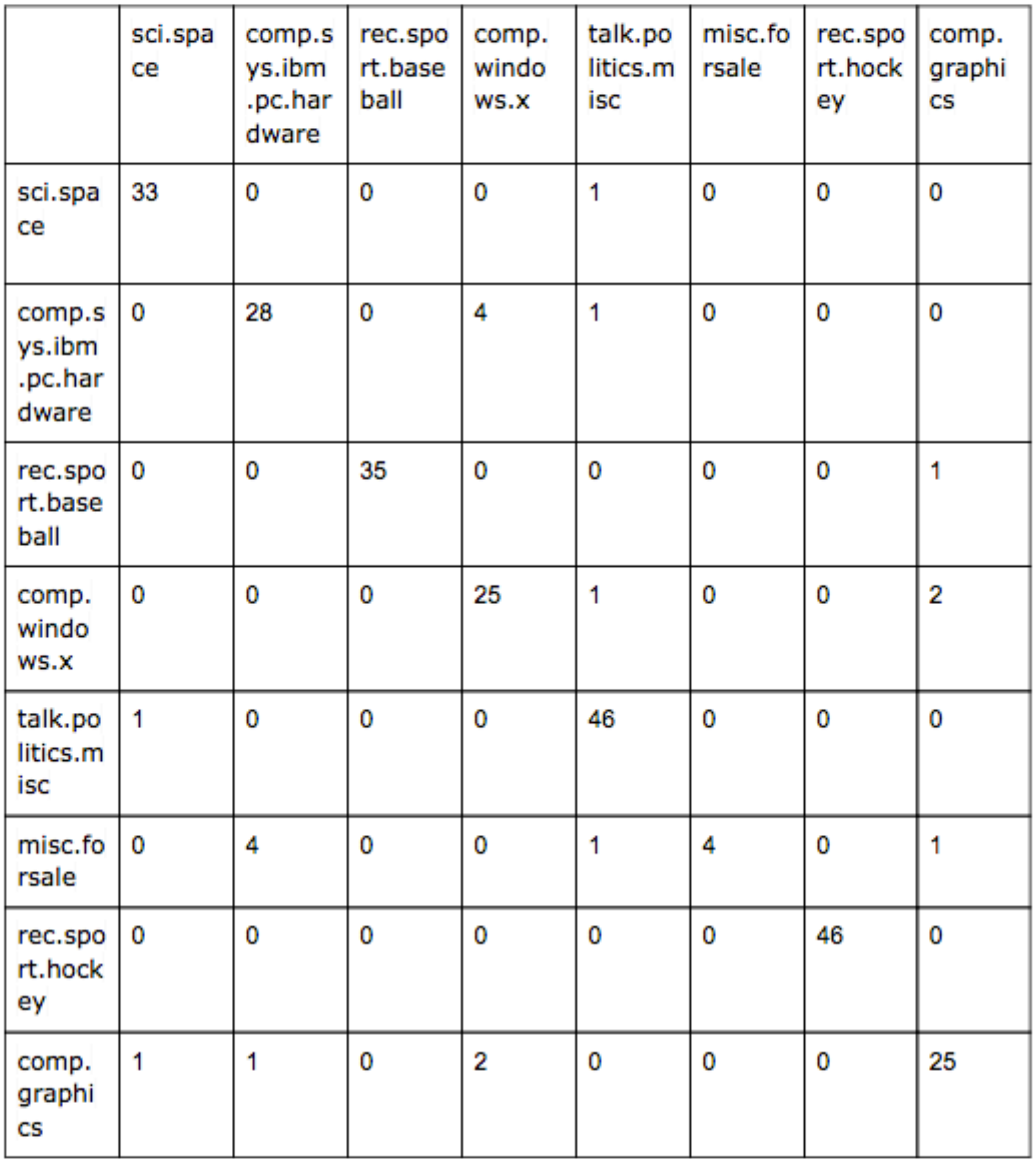
For Character '1' this is the least prototypical image:

```
                            ####       ####         ###
   #########################               #########   ##
  #                         #                            #
  #                         #                            #
  #                         #                            #
  #                        #              #######        #
  #                   ##    #     ##     #       #       #
  #                 ##  ##   #   #  #    #        #      #
  #                #      #   #  #   #  #    #   #       #
  #                #       #  #       ##   #  #          #
  #                #       #  #   #  #     #   #         #
  #               #        # #   #  #    # #    #        #
  #               #     #   #    #      #   #   #        #
  #             ##    ##    #    #     #    #  #         #
  #            #     #    ##      #####      #  #        #
  #           #       ####                   #   #       #
  #           #                               #  #       #
  #          #                                #  #       #
  #         #                                 #  #       #
  #         #                             ### #  #       #
  #        #                            ##    #  #       #
  #        #                           #  #   #  #       #
  #        #                              #  ##          #
  #       #                               #    #         #
  #       #                         ######      #        #
  #        #                       #     #   #           #
  #        #                       #     #  # #          #
  #         # ####           ##    #     #  # #  #       #
  #          #    #       ###  ##   #    #  # #  #       #
  #                   ####       #  #    #  #  ##        #
  #               ####           #  #    #  #            #
  #            ###               #  #   #  #   #         #
  #           #                #   #   #   #  #          #
  #          #      #######     #   #  #     ###         #
  #          #    ##       ##   #  #  #         #        #
  #         #    #           #   #     #         #  ###  #
  #         #    #               #                ##     #
  #      #   #    ########       #######   #######       #
  #     #   #             #######       ###         #### #
   #####    #                                      #     #
  #        #                                      #      #
  #     ####                                      #      #
  #    #    #                                     #      #
  #     #                                         #      #
  #      ###     ##                    ######     #      #
  #         #   #  #     ##############      ###  #      #
  #         #   #   #   #                         #      #
  #         #   #    #  #                     #   #      #
  #         #   #    #  #                     #   #      #
  #         #   #    #  #  ###################    #      #
  #         #   #    #  #                         #      #
  #          #  #     #                           #      #
  #          #  #      ######                     #      #
  #          #  #            #                    #      #
  #          #   #           #                    #      #
  #          #   #            ##                  #      #
  #          #   #              ####        #     #      #
  #          #   #                  ######## #    #      #
  #           ###                             #   #      #
              #                               #   #      #
             #                                 #  # #    #
  ############                                       ####
 #            #                                          #
 #                      ############################     #
 # ############   ######                            #    #
 #  #          ###     #                            #   #
 #  #                   #                           #   #
```

Problem 2 extra credit:

We have implemented problem 2.2. The only thing different about this problem was that there were many more labels. We added more dictionaries to store word frequencies and likelihoods for words in each label, but the way we updated the confusion matrixes and found the classification rates were the same.

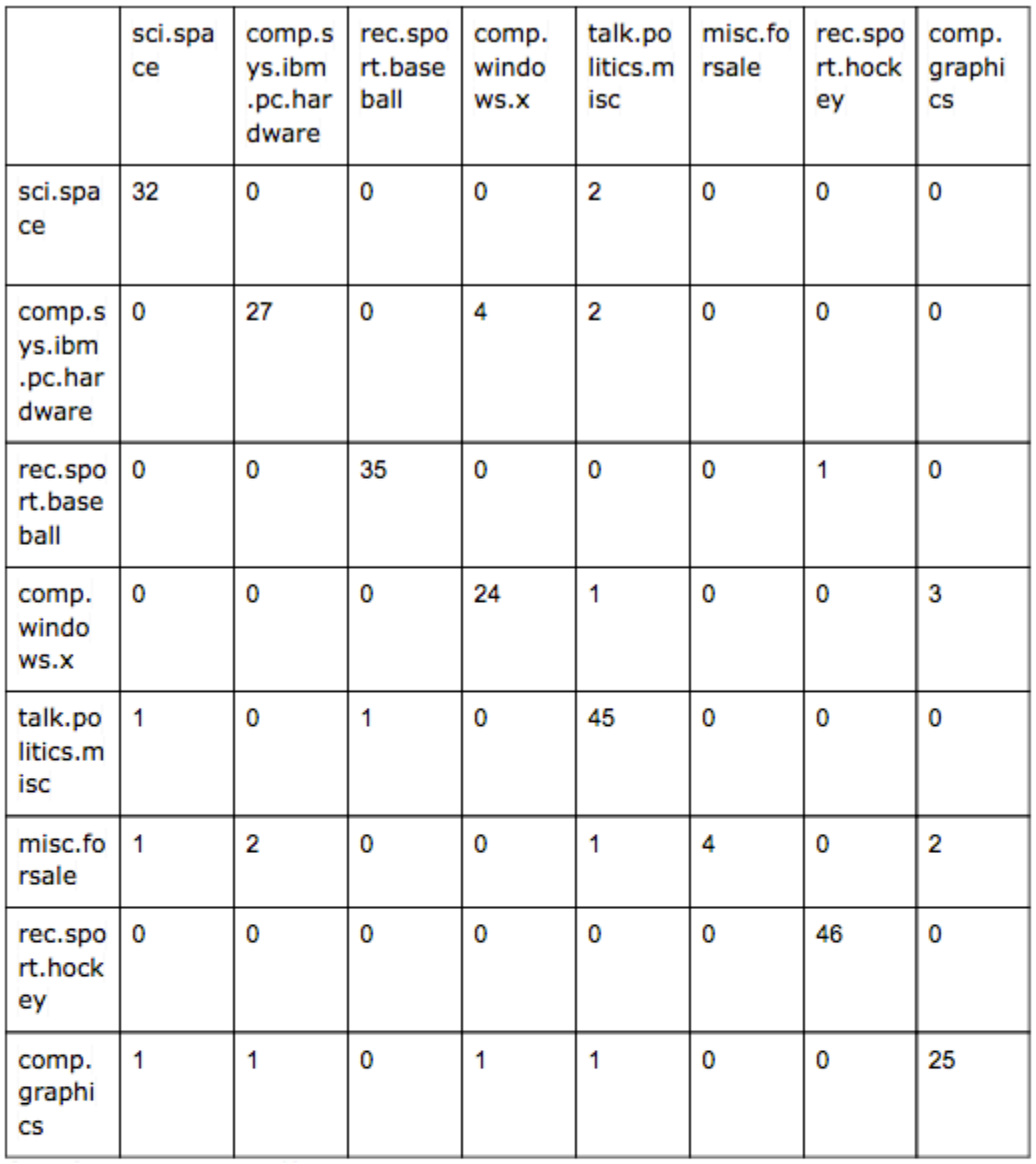
Bernoulli Confusion Matrix



	sci.space	comp.sys.ibm.pc.hardware	rec.sport.baseball	comp.windows.x	talk.politics.misc	misc.forsale	rec.sport.hockey	comp.graphics
sci.space	33	0	0	0	1	0	0	0
comp.sys.ibm.pc.hardware	0	28	0	4	1	0	0	0
rec.sport.baseball	0	0	35	0	0	0	0	1
comp.windows.x	0	0	0	25	1	0	0	2
talk.politics.misc	1	0	0	0	46	0	0	0
misc.forsale	0	4	0	0	1	4	0	1
rec.sport.hockey	0	0	0	0	0	0	46	0
comp.graphics	1	1	0	2	0	0	0	25

Classification Rate: 92.015%

Multinomial Confusion Matrix



	sci.space	comp.sys.ibm.pc.hardware	rec.sport.baseball	comp.windows.x	talk.politics.misc	misc.forsale	rec.sport.hockey	comp.graphics
sci.space	32	0	0	0	2	0	0	0
comp.sys.ibm.pc.hardware	0	27	0	4	2	0	0	0
rec.sport.baseball	0	0	35	0	0	0	1	0
comp.windows.x	0	0	0	24	1	0	0	3
talk.politics.misc	1	0	1	0	45	0	0	0
misc.forsale	1	2	0	0	1	4	0	2
rec.sport.hockey	0	0	0	0	0	0	46	0
comp.graphics	1	1	0	1	1	0	0	25

Classification Rate: 90.49%

The top 20 for words for the newsgroups are listed below. They are in the format word_(# of occurrences)

TOP 20 WORDS FOR EACH CATEGORY (2.2)

```
TOP sci_space WORDS
space_1030
nt_593
would_560
one_384
launch_352
nasa_345
earth_332
subject_328
like_304
us_280
system_278
also_277
writes_271
could_263
first_253
data_253
time_253
orbit_251
edu_251
TOP comp_sys_ibm_pc_hardware WORDS
drive_496
scsi_416
nt_392
ide_306
one_262
card_253
drives_232
controller_229
system_216
disk_216
subject_205
use_204
would_203
edu_198
hard_191
bus_189
get_177
m_176
data_164
TOP rec_sport_baseball WORDS
nt_936
would_454
year_427
edu_416
```

```
writes_355
one_316
game_316
good_299
team_294
subject_293
last_288
article_287
think_287
players_275
like_267
baseball_255
games_242
better_240
well_222
TOP comp_windows_x WORDS
x_3598
window_522
use_455
nt_433
subject_426
file_396
server_363
also_323
get_312
available_312
edu_286
motif_284
version_277
system_270
sun_256
program_256
c_254
one_252
m_248
TOP talk_politics_misc WORDS
nt_1400
would_951
people_831
q_692
one_601
mr_599
think_571
writes_552
president_552
article_509
```

```
government_497
stephanopoulos_452
know_448
us_417
edu_413
like_404
subject_378
going_372
get_331
TOP misc_forsale WORDS
new_218
edu_189
dos_160
sale_145
appears_144
art_135
subject_132
wolverine_128
shipping_117
cover_115
price_115
one_112
list_112
comics_109
drive_108
nt_105
hulk_104
good_101
vs_98
TOP rec_sport_hockey WORDS
nt_838
game_653
team_635
hockey_564
would_402
play_374
subject_339
period_338
season_333
nhl_327
games_322
one_312
first_290
year_283
think_276
players_271
```

```
get_262
la_262
edu_259
TOP comp_graphics WORDS
image_889
jpeg_468
edu_438
nt_419
file_406
images_389
data_369
also_365
graphics_346
software_315
available_301
use_289
one_259
program_252
files_242
format_240
get_235
version_221
system_219
```

Contributions

Erik Delanois - Problem 1
Doug Zhu, Dallas Delaney - Problem 2