

Artificial Intelligence Assignment 3

Part1 Digit Classification

UIN: 677629729

Name: Bo-Yu, Chiang

Email: bchiang3@illinois.edu

MAP

Find highest classification accuracy of k

```
k = 0    Accuracy = 0.765000
k = 1    Accuracy = 0.771000
k = 2    Accuracy = 0.766000
k = 3    Accuracy = 0.763000
k = 4    Accuracy = 0.762000
k = 5    Accuracy = 0.758000
k = 6    Accuracy = 0.758000
k = 7    Accuracy = 0.756000
k = 8    Accuracy = 0.755000
k = 9    Accuracy = 0.755000
k = 10   Accuracy = 0.755000
k = 11   Accuracy = 0.753000
k = 12   Accuracy = 0.753000
k = 13   Accuracy = 0.752000
k = 14   Accuracy = 0.751000
k = 15   Accuracy = 0.750000
k = 16   Accuracy = 0.747000
k = 17   Accuracy = 0.746000
k = 18   Accuracy = 0.746000
k = 19   Accuracy = 0.747000
k = 20   Accuracy = 0.746000
k = 21   Accuracy = 0.747000
k = 22   Accuracy = 0.746000
k = 23   Accuracy = 0.747000
k = 24   Accuracy = 0.744000
k = 25   Accuracy = 0.742000
k = 26   Accuracy = 0.740000
k = 27   Accuracy = 0.739000
k = 28   Accuracy = 0.738000
k = 29   Accuracy = 0.734000
k = 30   Accuracy = 0.732000
k = 31   Accuracy = 0.732000
k = 32   Accuracy = 0.730000
k = 33   Accuracy = 0.730000
k = 34   Accuracy = 0.728000
k = 35   Accuracy = 0.728000
k = 36   Accuracy = 0.728000
k = 37   Accuracy = 0.726000
k = 38   Accuracy = 0.725000
k = 39   Accuracy = 0.726000
k = 40   Accuracy = 0.724000
k = 41   Accuracy = 0.724000
k = 42   Accuracy = 0.724000
k = 43   Accuracy = 0.724000
k = 44   Accuracy = 0.723000
k = 45   Accuracy = 0.723000
k = 46   Accuracy = 0.723000
k = 47   Accuracy = 0.723000
k = 48   Accuracy = 0.722000
k = 49   Accuracy = 0.721000
argMax_k = 1    Accuracy = 0.771000
```

Classification rate for each digit

```
Classification rate for 0 = 0.844444  
Classification rate for 1 = 0.962963  
Classification rate for 2 = 0.776699  
Classification rate for 3 = 0.790000  
Classification rate for 4 = 0.766355  
Classification rate for 5 = 0.673913  
Classification rate for 6 = 0.758242  
Classification rate for 7 = 0.726415  
Classification rate for 8 = 0.601942  
Classification rate for 9 = 0.800000
```

Confusion Matrix

0.84	0.00	0.01	0.00	0.01	0.06	0.03	0.00	0.04	0.00
0.00	0.96	0.01	0.00	0.00	0.02	0.01	0.00	0.00	0.00
0.01	0.03	0.78	0.04	0.01	0.00	0.06	0.01	0.05	0.02
0.00	0.02	0.00	0.79	0.00	0.03	0.02	0.06	0.02	0.06
0.00	0.01	0.00	0.00	0.77	0.00	0.03	0.01	0.02	0.17
0.02	0.02	0.01	0.13	0.03	0.67	0.01	0.01	0.02	0.07
0.01	0.07	0.04	0.00	0.04	0.05	0.76	0.00	0.02	0.00
0.00	0.06	0.03	0.00	0.03	0.00	0.00	0.73	0.03	0.13
0.02	0.01	0.03	0.14	0.02	0.06	0.00	0.01	0.60	0.12
0.01	0.01	0.01	0.03	0.09	0.02	0.00	0.02	0.01	0.80

Most prototypical instance of every digit

```
Most prototypical instance of 0:  
test data number = 723
```

```
Most prototypical instance of 1:  
test data number = 633
```

```
Most prototypical instance of 2:  
test data number = 795
```

```
Most prototypical instance of 3:  
test data number = 205
```

```
Most prototypical instance of 4:  
test data number = 111
```

```
Most prototypical instance of 5:  
test data number = 471
```

```
Most prototypical instance of 6:  
test data number = 632
```

Most prototypical instance of 7:
test data number = 784

Most prototypical instance of 8:
test data number = 560

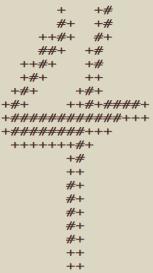
+###+
++####+
++####+
++####+
++####+ +###+
++####+ +###+
+###+ +###+
###+ +###+
###+ +###+
+###++ +###++
+####+#+
####+#+
+####+#+
+####+#+
####+#+
+####+#+
+####+#+
+####+#+

```
Most prototypical instance of 9:  
test data number = 745
```

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

Interesting example of incorrect classifications

Correct: 4, Misclassified to 9



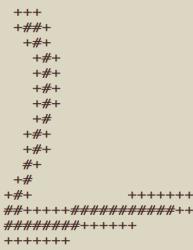
Because the upper side is so close so the classifier thinks that is much more likely to be 9

Correct: 6, Misclassified to 1



Because the hole in digit 6 is so small so the classifier thinks that is much more likely to be 1

Correct: 2, Misclassified to 6



Because the curve in digit 2 is not obvious so every digit has a low probability

Because the sample is so straight so the classifier thinks that is much more likely to be 1

The different type of writing is easy for human to recognize but is hard for computer if it is not happened very often in training sample

The 4 is so strange that all the digit has a small probability

Most Confused Pair

1.Likelihood Notation:

#: $\log(\text{likelihood}) > -0.5$

+: $\log(\text{likelihood}) > -1$

-: $\log(\text{likelihood}) > -2$

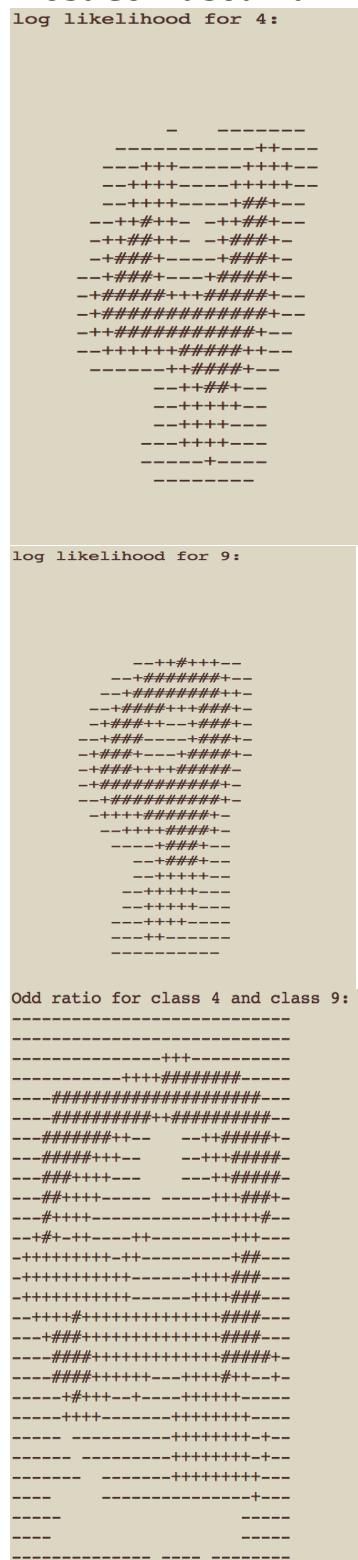
2.Odd ratio Notation:

#: $\log(\text{odd_ratio}) > 1$

+: $\log(\text{odd_ratio}) > 0$

-: $\log(\text{odd_ratio}) > -1$

Most Confused Pair1: 4 and 9



Most Confused Pair2: 8 and 3

log likelihood for 8:

log likelihood for 3:

A decorative border consisting of a repeating pattern of short horizontal dashes and vertical lines forming a diamond shape.

Odd ratio for class 8 and class 3:

Most Confused Pair3: 7 and 9

Most Confused Pair4: 5 and 3

ML

Find highest classification accuracy of k

```
k = 0 Accuracy = 0.764000
k = 1 Accuracy = 0.770000
k = 2 Accuracy = 0.766000
k = 3 Accuracy = 0.764000
k = 4 Accuracy = 0.762000
k = 5 Accuracy = 0.759000
k = 6 Accuracy = 0.757000
k = 7 Accuracy = 0.758000
k = 8 Accuracy = 0.756000
k = 9 Accuracy = 0.756000
k = 10 Accuracy = 0.757000
k = 11 Accuracy = 0.755000
k = 12 Accuracy = 0.754000
k = 13 Accuracy = 0.752000
k = 14 Accuracy = 0.752000
k = 15 Accuracy = 0.750000
k = 16 Accuracy = 0.748000
k = 17 Accuracy = 0.746000
k = 18 Accuracy = 0.746000
k = 19 Accuracy = 0.745000
k = 20 Accuracy = 0.747000
k = 21 Accuracy = 0.747000
k = 22 Accuracy = 0.747000
k = 23 Accuracy = 0.747000
k = 24 Accuracy = 0.748000
k = 25 Accuracy = 0.744000
k = 26 Accuracy = 0.742000
k = 27 Accuracy = 0.740000
k = 28 Accuracy = 0.739000
k = 29 Accuracy = 0.736000
k = 30 Accuracy = 0.733000
k = 31 Accuracy = 0.732000
k = 32 Accuracy = 0.732000
k = 33 Accuracy = 0.730000
k = 34 Accuracy = 0.729000
k = 35 Accuracy = 0.728000
k = 36 Accuracy = 0.729000
k = 37 Accuracy = 0.726000
k = 38 Accuracy = 0.726000
k = 39 Accuracy = 0.725000
k = 40 Accuracy = 0.724000
k = 41 Accuracy = 0.723000
k = 42 Accuracy = 0.724000
k = 43 Accuracy = 0.724000
k = 44 Accuracy = 0.723000
k = 45 Accuracy = 0.723000
k = 46 Accuracy = 0.723000
k = 47 Accuracy = 0.723000
k = 48 Accuracy = 0.722000
k = 49 Accuracy = 0.722000
argMax_k = 1 Accuracy = 0.770000
```

Classification rate for each digit

```
Classification rate for 0 = 0.844444
Classification rate for 1 = 0.962963
Classification rate for 2 = 0.776699
Classification rate for 3 = 0.790000
Classification rate for 4 = 0.757009
Classification rate for 5 = 0.673913
Classification rate for 6 = 0.758242
Classification rate for 7 = 0.726415
Classification rate for 8 = 0.601942
Classification rate for 9 = 0.800000
```

Confusion Matrix

0.84	0.00	0.01	0.00	0.01	0.06	0.03	0.00	0.04	0.00
0.00	0.96	0.01	0.00	0.00	0.02	0.01	0.00	0.00	0.00
0.01	0.03	0.78	0.04	0.01	0.00	0.06	0.01	0.05	0.02
0.00	0.02	0.00	0.79	0.00	0.03	0.02	0.06	0.02	0.06
0.00	0.01	0.00	0.00	0.76	0.00	0.03	0.01	0.02	0.18
0.02	0.02	0.01	0.13	0.03	0.67	0.01	0.01	0.02	0.07
0.01	0.07	0.04	0.00	0.04	0.05	0.76	0.00	0.02	0.00
0.00	0.06	0.03	0.00	0.03	0.00	0.00	0.73	0.03	0.13
0.02	0.01	0.03	0.14	0.02	0.06	0.00	0.01	0.60	0.12
0.01	0.01	0.01	0.03	0.09	0.02	0.00	0.02	0.01	0.80

The result of ML algorithm is only slight different from MAP, the rest of the report is the same as MAP algorithm.

Part1.2 Face Classification

Smoothing constant k = 1

Classification rate(0 for no face, 1 for face)

Classification rate for 0 = 0.935065

Classification rate for 1 = 0.726027

Confusion matrix

0.94	0.06
0.27	0.73

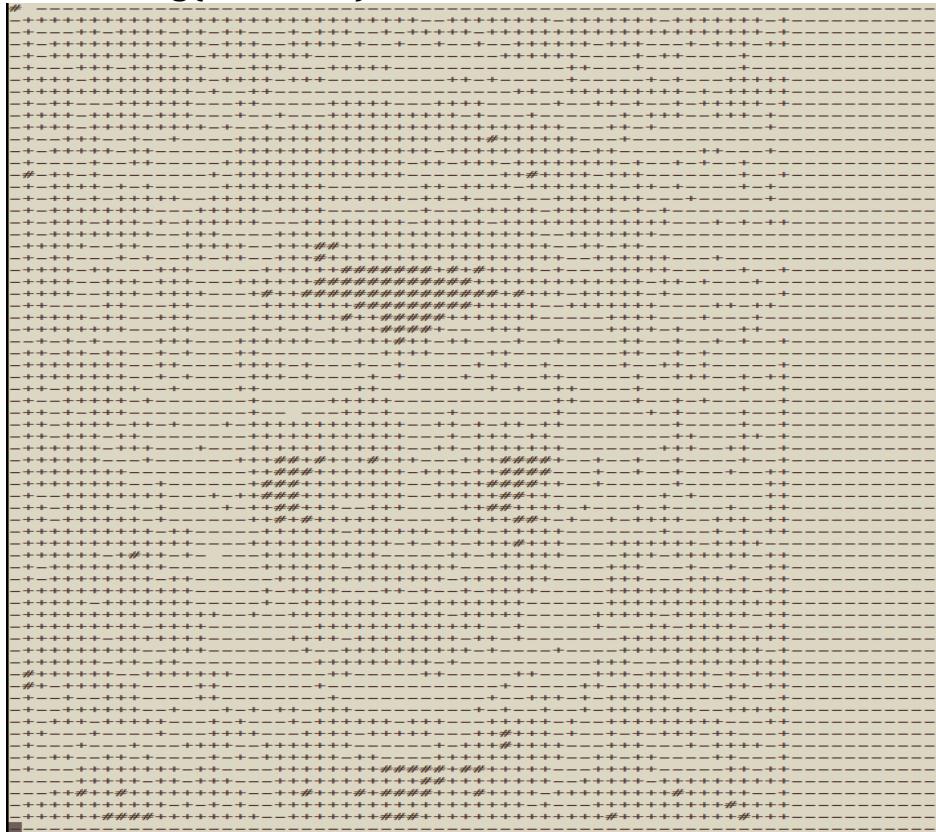
Odds ratio image for faces vs non-faces

Odd ratio Notation:

#: log(odd_ratio)>1

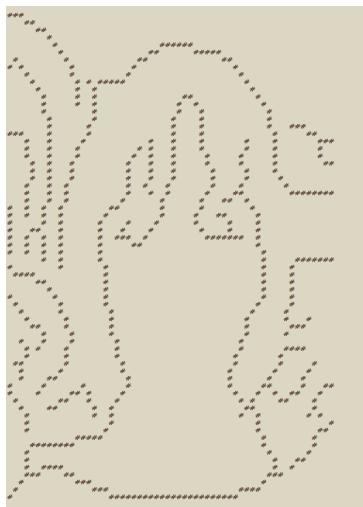
+: log(odd_ratio)>0

-: log(odd_ratio)>-1

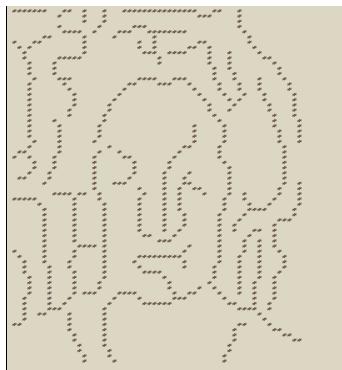


False Negative

30th



38th

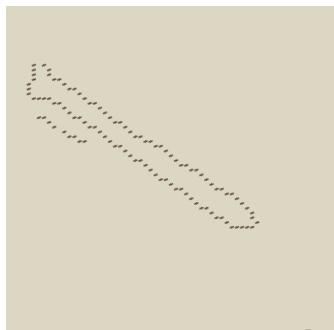


104th

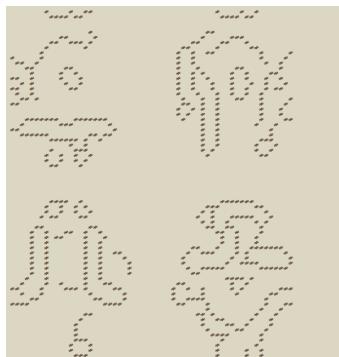


False Positive

62th



88th



93th



Part2 Text Document Classification

Classification rate for each class

```
Classification rate for class 0 = 0.941176
Classification rate for class 1 = 0.878788
Classification rate for class 2 = 0.972222
Classification rate for class 3 = 0.821429
Classification rate for class 4 = 0.914894
Classification rate for class 5 = 0.800000
Classification rate for class 6 = 0.826087
Classification rate for class 7 = 0.827586
```

Confusion Matrix

```
0.94 0.00 0.00 0.00 0.06 0.00 0.00 0.00
0.00 0.88 0.00 0.03 0.00 0.00 0.00 0.09
0.00 0.00 0.97 0.00 0.00 0.00 0.00 0.03
0.00 0.04 0.00 0.82 0.04 0.04 0.00 0.07
0.02 0.02 0.02 0.00 0.91 0.00 0.00 0.02
0.00 0.20 0.00 0.00 0.00 0.80 0.00 0.00
0.00 0.00 0.09 0.00 0.02 0.04 0.83 0.02
0.00 0.07 0.00 0.00 0.00 0.10 0.00 0.83
```

Top20 words with the highest likelihood for each class

```
Class 0
1th word: the      0.048719
2th word: of       0.022911
3th word: to       0.021289
4th word: and      0.018550
5th word: a        0.018202
6th word: in       0.013879
7th word: is       0.011548
8th word: that     0.009289
9th word: for      0.008586
10th word: _        0.008287
11th word: it      0.007650
12th word: i       0.007459
13th word: on      0.007388
14th word: be      0.005946
15th word: s       0.005760
16th word: are     0.005717
17th word: from    0.005679
18th word: space   0.005613
19th word: this    0.005396
20th word: you    0.005298

Class 1
1th word: the      0.041181
2th word: i       0.020034
3th word: to       0.019333
4th word: a        0.017975
5th word: and      0.016517
6th word: _        0.014970
7th word: is       0.012977
8th word: of       0.011575
9th word: it       0.011074
10th word: for     0.009249
11th word: in      0.008948
12th word: that    0.008882
13th word: with    0.008014
14th word: you    0.007735
15th word: on      0.006711
16th word: this    0.006633
17th word: have    0.006489
18th word: drive   0.005532
19th word: from    0.005187
20th word: are     0.005086
```

```

Class 2
1th word: the      0.039980
2th word: to       0.019450
3th word: a        0.018302
4th word: in       0.015176
5th word: and      0.015132
6th word: of       0.014828
7th word: i        0.014450
8th word: that     0.011947
9th word: is       0.011458
10th word: _        0.009080
11th word: s        0.008658
12th word: it       0.007881
13th word: he       0.007695
14th word: you     0.007125
15th word: nt       0.006940
16th word: for     0.006925
17th word: have    0.006170
18th word: was     0.005540
19th word: be       0.005333
20th word: but     0.005288

Class 3
1th word: the      0.036078
2th word: x        0.022507
3th word: to       0.019068
4th word: a        0.015997
5th word: _        0.015384
6th word: and      0.015190
7th word: of       0.013152
8th word: is       0.012726
9th word: in       0.011075
10th word: i       0.010125
11th word: for     0.009149
12th word: you     0.007154
13th word: it       0.007092
14th word: on       0.006898
15th word: that     0.006354
16th word: this     0.006160
17th word: from    0.005184
18th word: or       0.005078
19th word: are      0.004972
20th word: with    0.004953

Class 4
1th word: the      0.051114
2th word: to       0.027728
3th word: of       0.022757
4th word: a        0.018420
5th word: and      0.018420
6th word: that     0.018084
7th word: in       0.016141
8th word: i        0.013047
9th word: is       0.012880
10th word: it      0.011258
11th word: _        0.009833
12th word: you     0.009699
13th word: for     0.007348
14th word: have    0.007127
15th word: s        0.007067
16th word: not     0.006878
17th word: on       0.006814
18th word: are      0.006569
19th word: be       0.006466
20th word: they    0.006181

Class 5
1th word: _        0.019488
2th word: the     0.019039
3th word: 1        0.016509
4th word: and     0.013148
5th word: to       0.011966
6th word: for     0.011766
7th word: of       0.011633
8th word: i        0.011250
9th word: a        0.010368
10th word: 2       0.009886
11th word: in      0.007056
12th word: is      0.006541
13th word: you     0.006291
14th word: it      0.006191
15th word: 3       0.005226
16th word: with    0.004860
17th word: or      0.004061
18th word: are     0.004011
19th word: s        0.003961
20th word: from    0.003794

```

```
Class 6
1th word: the      0.039657
2th word: to       0.014896
3th word: a        0.014202
4th word: 1        0.013705
5th word: 0        0.013569
6th word: in       0.012613
7th word: and      0.012492
8th word: of       0.011875
9th word: _         0.010449
10th word: i        0.009772
11th word: 2        0.009024
12th word: that     0.007647
13th word: is       0.007074
14th word: for      0.006599
15th word: s        0.006320
16th word: was      0.006194
17th word: it       0.005959
18th word: he       0.005801
19th word: on       0.005380
20th word: 3        0.005353
Class 7
1th word: _         0.034461
2th word: the      0.032520
3th word: and      0.019745
4th word: to       0.019566
5th word: a        0.018539
6th word: of       0.017048
7th word: is       0.012920
8th word: for      0.011582
9th word: in       0.011264
10th word: it      0.009111
11th word: you     0.008786
12th word: i        0.008315
13th word: that     0.006533
14th word: on       0.006076
15th word: from     0.006043
16th word: image    0.005897
17th word: this     0.005811
18th word: with     0.005506
19th word: or       0.005407
20th word: be       0.005042
```

Most Confused Pair1: 5 and 1

```
Odd ratio for class 5 over class 1:  
1th word: wolverine 394.308003  
2th word: comics 336.231631  
3th word: hulk 320.948375  
4th word: 1500 268.985305  
5th word: annual 256.758700  
6th word: spiderman 235.362142  
7th word: ghost 223.135537  
8th word: liefeld 220.078886  
9th word: sabretooth 213.965583  
10th word: condition 207.852281  
11th word: app 186.455723  
12th word: bagged 183.399071  
13th word: hobgoblin 177.285769  
14th word: xforce 174.229118  
15th word: guida 165.059164  
16th word: punisher 155.889211  
17th word: edition 146.719257  
18th word: marvel 140.605955  
19th word: 175 137.549304  
20th word: ufl 131.436001  
  
Odd ratio for class 1 over class 5:  
1th word: bus 3064652.764335  
2th word: bios 2242035.443382  
3th word: scsi2 1532326.382168  
4th word: isa 1516196.630776  
5th word: 0 1500066.879385  
6th word: slave 1193601.602952  
7th word: scsil 1177471.851560  
8th word: vlb 1048433.840431  
9th word: diamond 838747.072344  
10th word: ati 790357.818171  
11th word: irq 774228.066779  
12th word: compos 741968.563997  
13th word: cylinders 612930.552867  
14th word: eisa 580671.050085  
15th word: settings 564541.298693  
16th word: vote 548411.547302  
17th word: led 532281.795911  
18th word: uwoca 483892.541737  
19th word: orchid 467762.790346  
20th word: aspi4dos 451633.038955
```

Most Confused Pair2: 7 and 5

```
Odd ratio for class 7 over class 5:  
1th word: jpeg 4503534.573968  
2th word: ftp 2035712.856463  
3th word: gif 1978098.341658  
4th word: sgi 1142687.876977  
5th word: 24bit 1056266.104769  
6th word: xv 931434.656023  
7th word: sites 883422.560352  
8th word: bits 816205.626412  
9th word: archive 748988.692472  
10th word: tiff 720181.435070  
11th word: visualization 710579.015935  
12th word: 0 700976.596801  
13th word: quicktime 691374.177667  
14th word: gopher 681771.758532  
15th word: archie 652964.501130  
16th word: viewer 624157.243727  
17th word: animation 585747.567190  
18th word: postscript 576145.148056  
19th word: platforms 556940.309787  
20th word: pixels 547337.890653  
  
Odd ratio for class 5 over class 7:  
1th word: wolverine 3111324.271257  
2th word: comics 2653067.208048  
3th word: hulk 2532473.244046  
4th word: spiderman 1857147.045634  
5th word: 140 1760671.874432  
6th word: liefeld 1736553.081632  
7th word: sabretooth 1688315.496031  
8th word: app 1471246.360827  
9th word: bagged 1447127.568026  
10th word: hobgoblin 1398889.982425  
11th word: xforce 1374771.189625  
12th word: punisher 1230058.432822  
13th word: cable 1157702.054421  
14th word: marvel 1109464.468820  
15th word: panther 988870.504818  
16th word: mutants 940632.919217  
17th word: pom 892395.333616  
18th word: mint 868276.540816  
19th word: bwsmith 844157.748015  
20th word: keown 820038.955215
```

Most Confused Pair3: 1 and 7

```
1th word: drives 3758232.074159
2th word: scsi2 1532326.382168
3th word: slave 1193601.602952
4th word: scsil 1177471.851560
5th word: pin 1080693.343213
6th word: vlb 1048433.840431
7th word: ati 790357.818171
8th word: irq 774228.066779
9th word: maxtor 758098.315388
10th word: gateway 741968.563997
11th word: connector 693579.309823
12th word: cylinders 612930.552867
13th word: esdi 580671.050085
14th word: vote 548411.547302
15th word: formatting 516152.044520
16th word: uwoca 483892.541737
17th word: orchid 467762.790346
18th word: asynchronous 451633.038955
19th word: mfm 419373.536172
20th word: st506st412 403243.784781

Odd ratio for class 7 over class 1:
1th word: jpeg 4503534.573968
2th word: z 1123483.038708
3th word: ray 969844.332560
4th word: xv 931434.656023
5th word: tiff 720181.435070
6th word: visualization 710579.015935
7th word: quicktime 691374.177667
8th word: gopher 681771.758532
9th word: project 672169.339398
10th word: interactive 662566.920264
11th word: viewer 624157.243727
12th word: animation 585747.567190
13th word: pixel 576145.148056
14th word: gamma 556940.309787
15th word: jfif 547337.890653
16th word: siggraph 528133.052384
17th word: den 508928.214116
18th word: rayshade 499325.794982
19th word: macintosh 451313.699310
20th word: resource 441711.280176
```

Most Confused Pair4: 6 and 2

```
Odd ratio for class 6 over class 2:
1th word: nhl 2596512.261276
2th word: pts 1741563.102075
3th word: pp 1282423.738801
4th word: leafs 1060770.253082
5th word: vancouver 1021189.273490
6th word: goals 1013273.077571
7th word: nyi 981608.293897
8th word: ahl 965775.902060
9th word: buf 934111.118386
10th word: que 902446.334712
11th word: nyr 878697.746956
12th word: winnipeg 823284.375526
13th word: penalty 759954.808178
14th word: edmonton 744122.416341
15th word: sj 720373.828586
16th word: islanders 680792.848993
17th word: penalties 657044.261237
18th word: lemieux 633295.673482
19th word: sabres 609547.085726
20th word: oilers 593714.693889
```

```
Odd ratio for class 2 over class 6:
1th word: pitcher 1330983.083279
2th word: gant 1159243.330598
3th word: pitching 1094840.923342
4th word: reds 890899.967034
5th word: pitch 815763.825236
6th word: rbi 772828.887065
7th word: mets 740627.683438
8th word: pitchers 729893.948895
9th word: hirschbeck 719160.214352
10th word: winfield 697692.745267
11th word: sabo 654757.807097
12th word: bases 633290.338012
13th word: lopez 622556.603469
14th word: bonds 611822.868927
15th word: singer 590355.399842
16th word: clemens 558154.196214
17th word: obp 547420.461671
18th word: ba 536686.727129
19th word: 0000 525952.992586
20th word: plate 515219.258043
```

Extra credit: visualize the bag-of-words representations of the documents using word cloud map

Class0:



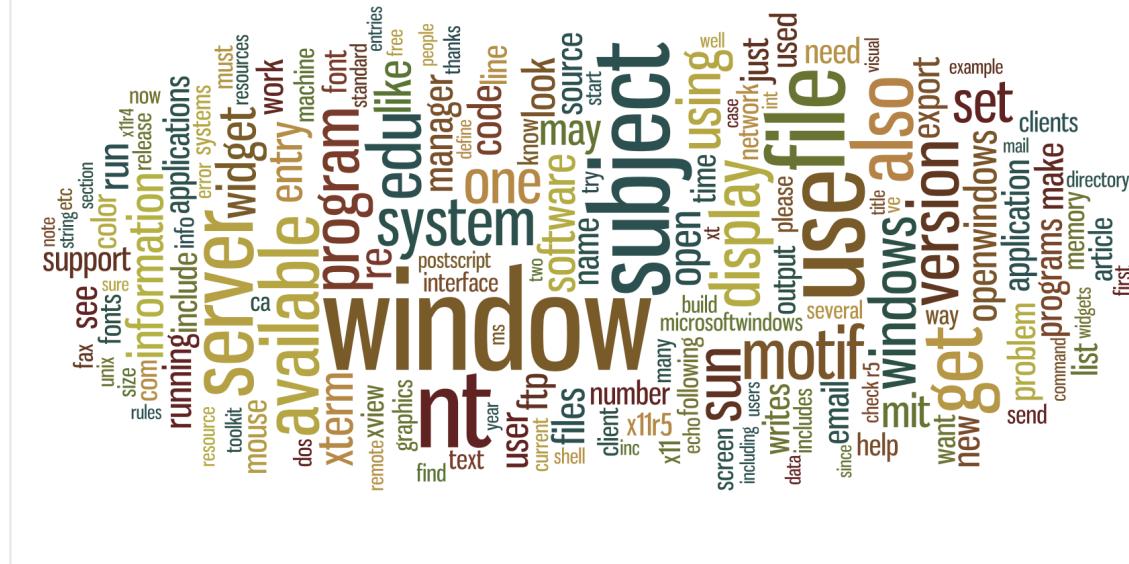
Class1:



Class2:



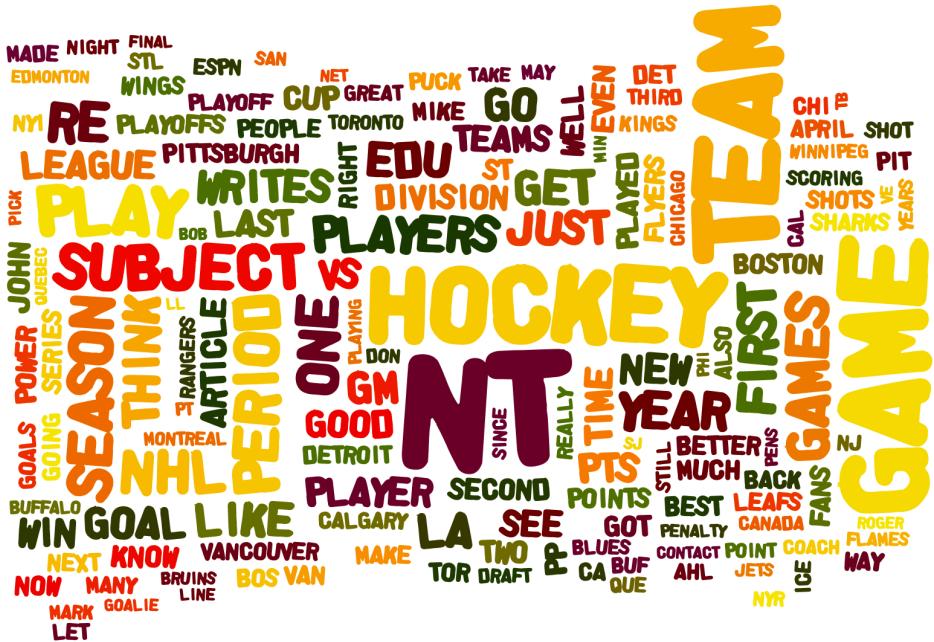
Class3:



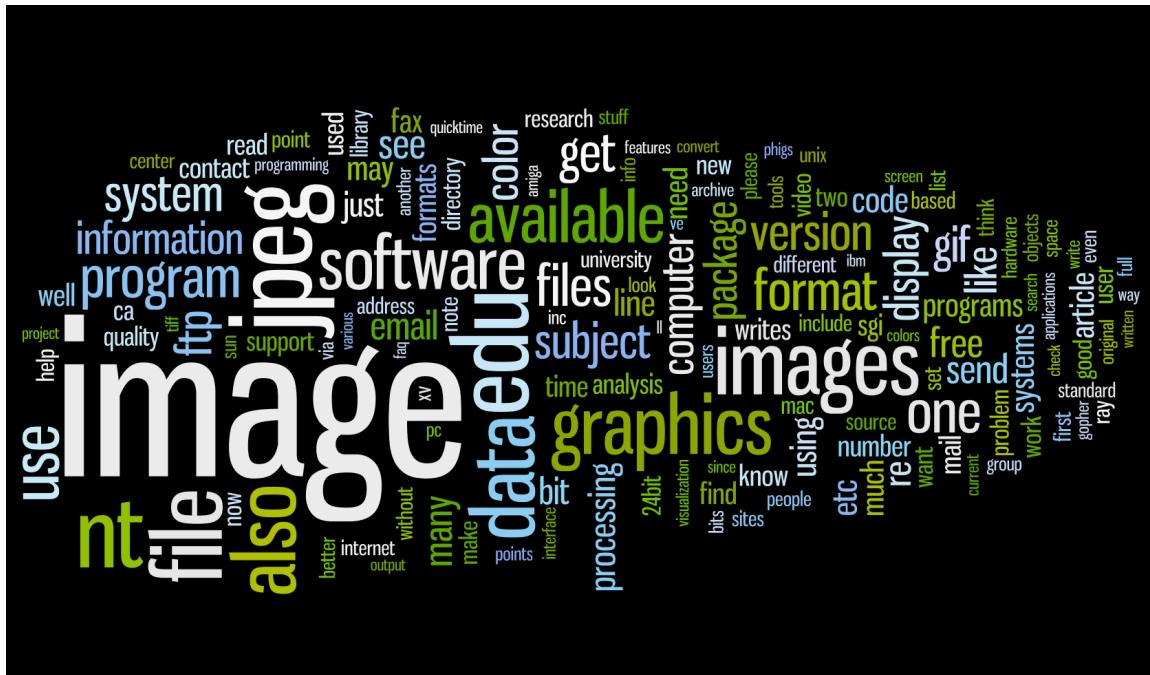
Class4:

Class5:

Class6:



Class7:



Tool link: <http://www.wordle.net/create>