**ECE 448 FALL 2020**

**Assignment 3:**

**Naive Bayes/Perceptron/Logistic Regression Classification**

**Nov. 16, 2020**

**Yucheng Jin**

**Yiqing Xie**

**Hangtao Jin**

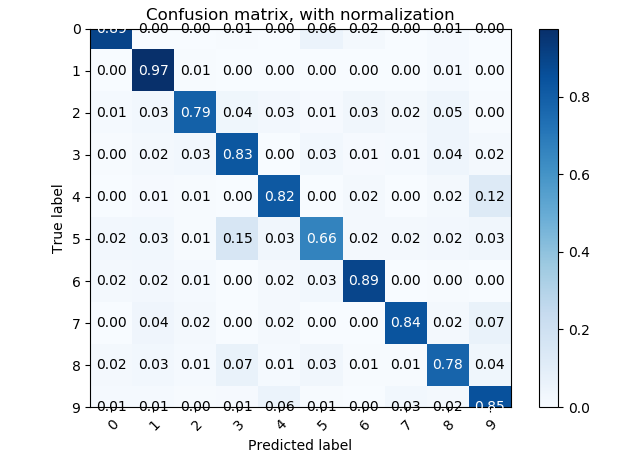
**Section I**

Average classification rate: for k = 1.0, the rate is 0.836

The classification rate for each class:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 0.89 | 0.97 | 0.79 | 0.83 | 0.82 | 0.66 | 0.89 | 0.84 | 0.78 | 0.85 |

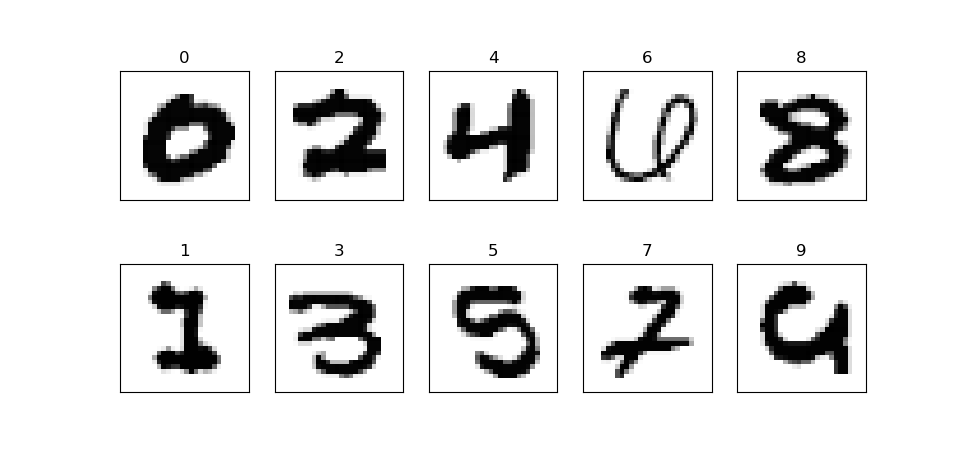
The confusion matrix:



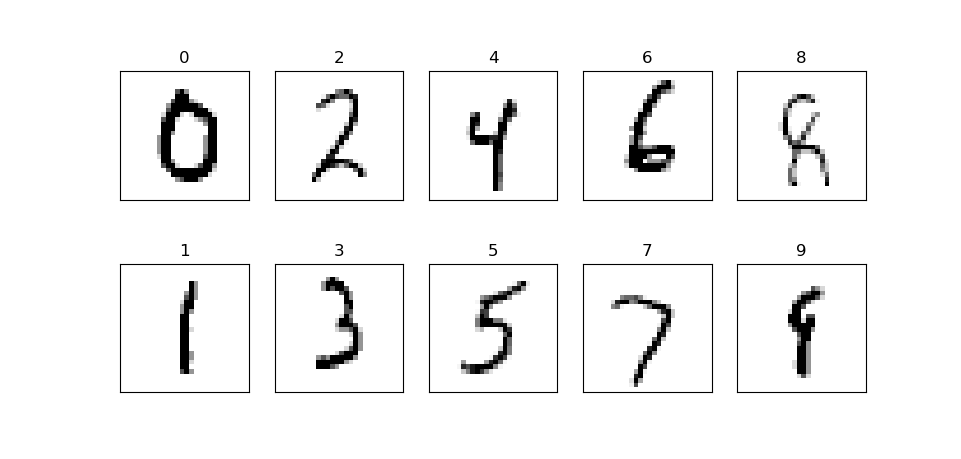
Plot(k=1.0):

A picture containing looking, sitting, photo, empty

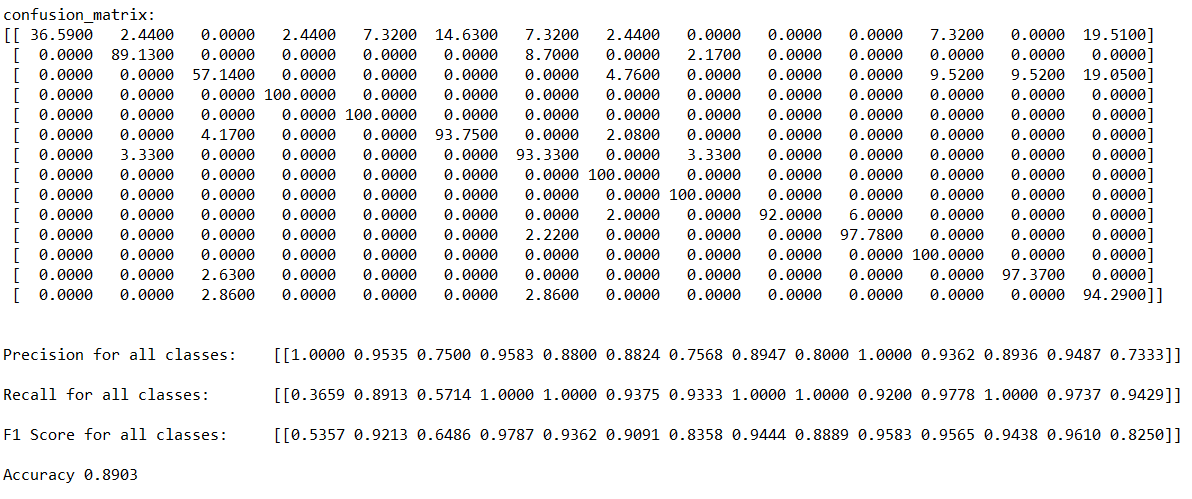
Description automatically generated

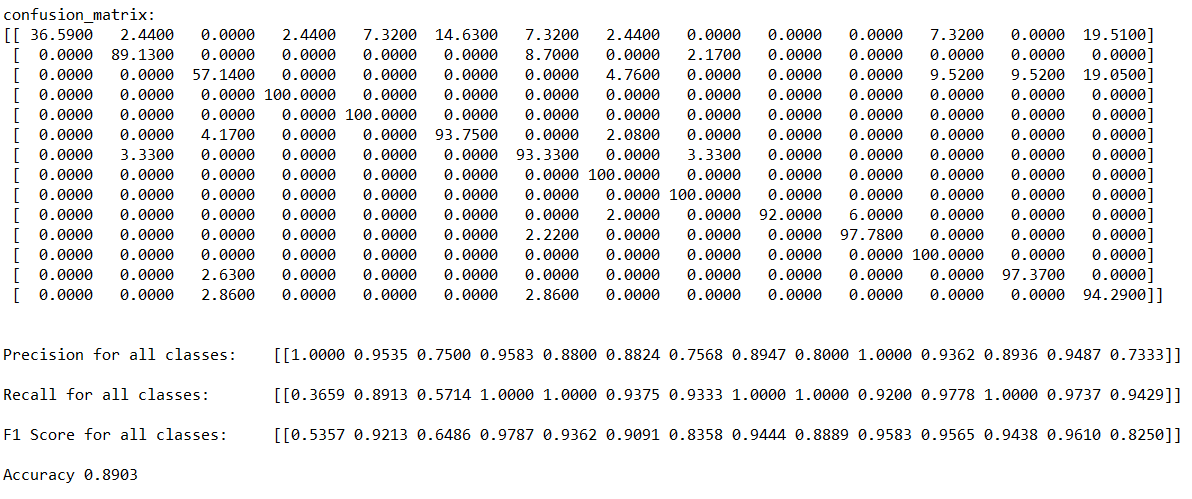
Example with the lowest posterior probability:

Example with the highest posterior probability:



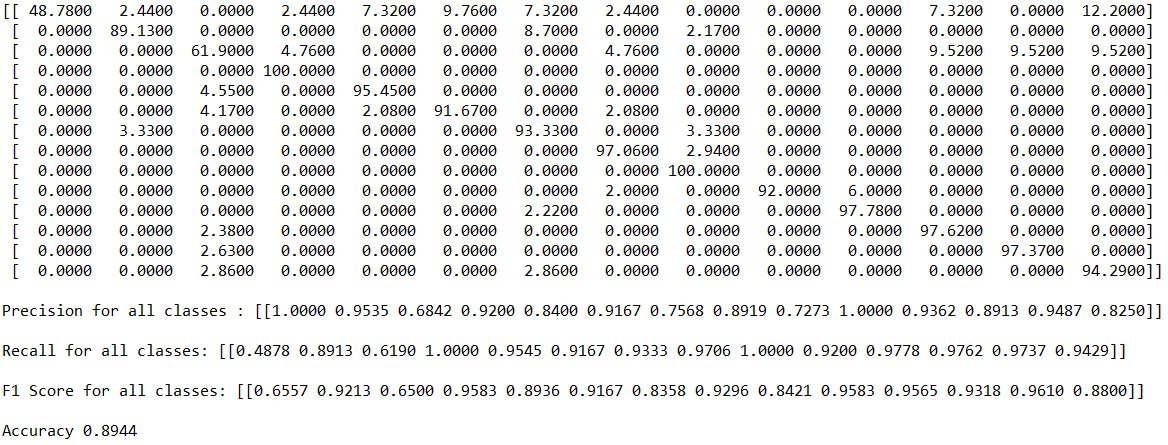
**Section II**

MAP confusion matrix:

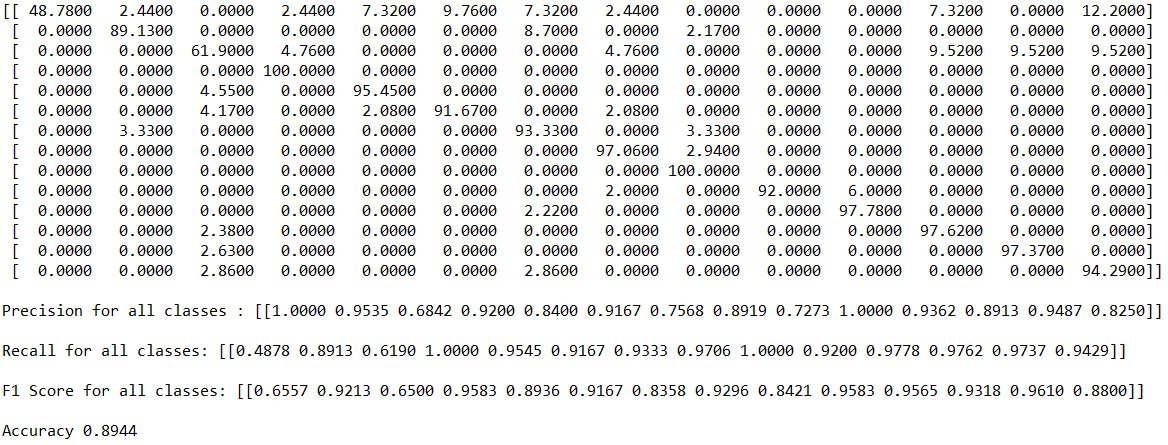
MAP output:

Top 20 feature words:

ML accuracy:



Uniform distribution accuracy:



Explanation for different accuracy:

If we ignore class prior or apply a uniform distribution, the accuracy of the result increase. Including the class prior doesn’t always beneficial. That’s because the class distribution between training and test sets may be different. What’s more, the result ignoring the class prior and that using uniform distribution share the same result.

**Section III**

Part 3.1: Perceptron model

Perceptron:

E\_train is 0.0, E\_test is 0.10804179999999984

Average number of iterations is 3.7419999999999636

Chart, scatter chart

Description automatically generated

Part 3.2: Logistic regression model

Logistic:

E\_train is 0.05280000000000003, E\_test is 0.05781600000000003

Chart, scatter chart

Description automatically generated

**Extra Credit for Section II**

Bigram model:



Optimal mixture model ():



Question 1:

Relaxing the naïve assumption isn’t always a good thing. As we can see, when using , which means only consider bigram model and relax the assumption the most, the accuracy decreases. This may because such strict test cases cause a lot of documents share similarly low probability. They have small probability in all classes, which let them be randomly classified.

Question 2:

When N is a really large number, we may find that almost all test documents have minimal probability and because they all don’t fit any test case. Thus such model is useless.

**Contribution**

Part I: Yucheng Jin

Part II: Yiqing Xie

Part III: Hangtao Jin