Lab Report

1.Title Page

List of all team members, course number and section for which each member is registered, date on which the report was written.

2.Section I

Section I: Image Classification. For both part 1.1 and part 1.2, report average classification rate, the classification rate for each class and the confusion matrix. For each class, show the test examples from that class that have the highest and lowest posterior probabilities or perceptron scores according to your classifier. Show the ten visualization plots for both feature likelihoods and perceptron weights.

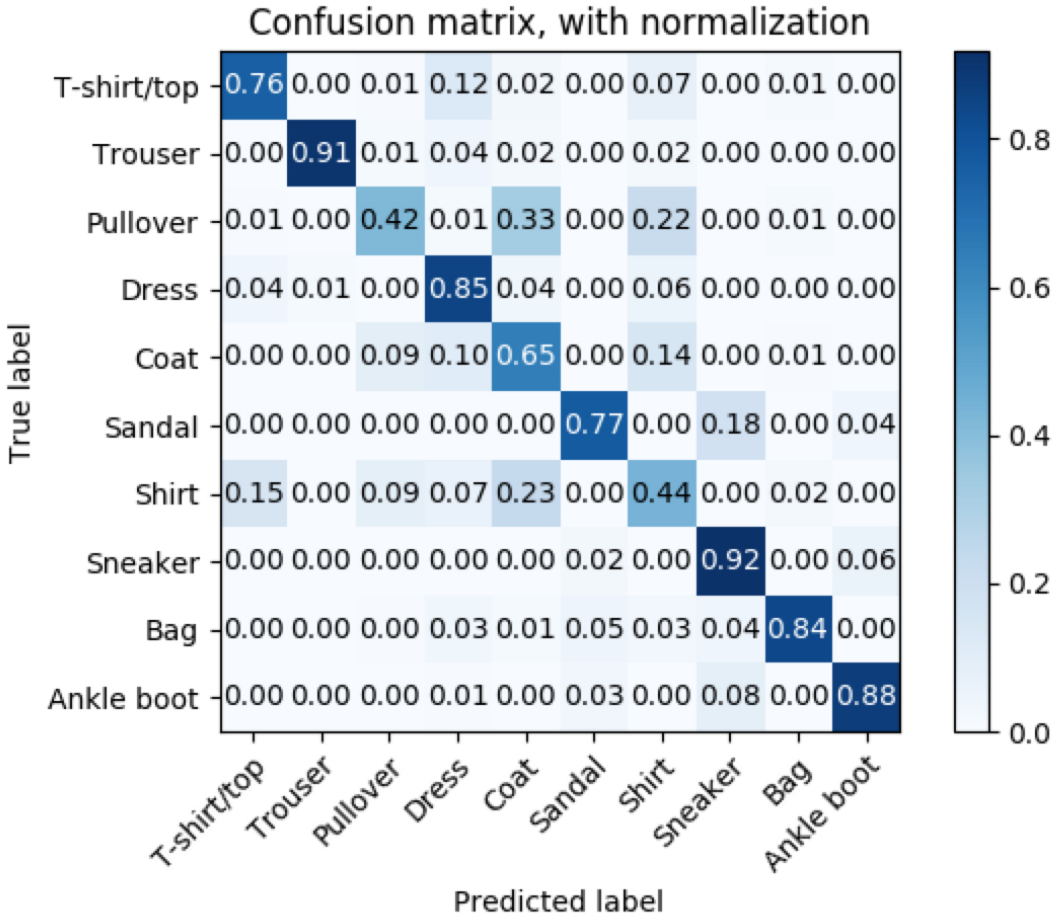
- **PART 1.1**

set laplace smoothing parameter `k=1`, the experiment result is following,

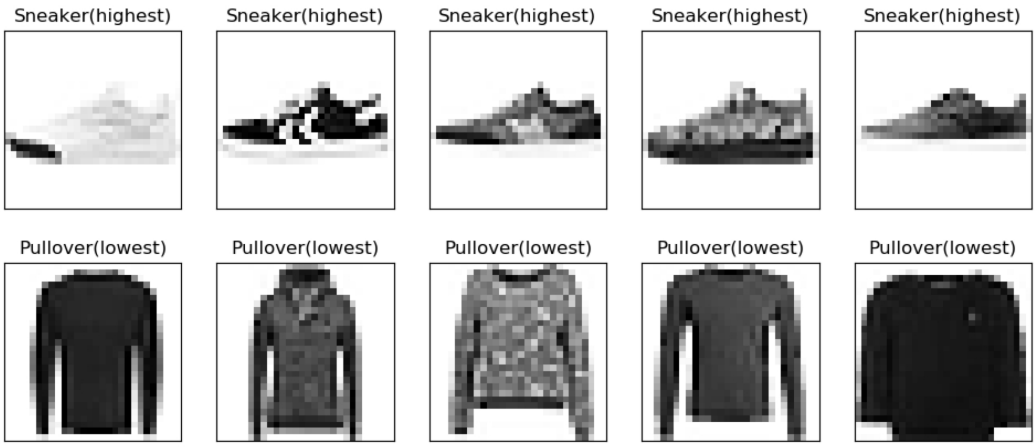
```
# 1.average classification rate
Bayes avg_acc:  0.7421

# 2.the classification rate for each class
T-shirt/top acc rate(class=0): 0.757
Trouser acc rate(class=1): 0.907
Pullover acc rate(class=2): 0.418
Dress acc rate(class=3): 0.852
Coat acc rate(class=4): 0.645
Sandal acc rate(class=5): 0.772
Shirt acc rate(class=6): 0.437
Sneaker acc rate(class=7): 0.918
Bag acc rate(class=8): 0.838
Ankle boot acc rate(class=9): 0.877

# 3.confusion matrix
[[ 0.757  0.001  0.009  0.122  0.023  0.002  0.073  0.     0.013
```

```
0.    ]
 [ 0.003  0.907  0.014  0.038  0.021  0.     0.017  0.     0.
0.    ]
 [ 0.007  0.     0.418  0.014  0.329  0.001  0.219  0.     0.012
0.    ]
 [ 0.043  0.012  0.     0.852  0.035  0.001  0.057  0.     0.
0.    ]
 [ 0.002  0.002  0.091  0.105  0.645  0.002  0.145  0.001  0.007
0.    ]
 [ 0.     0.     0.001  0.     0.     0.772  0.003  0.181  0.002
0.041]
 [ 0.153  0.001  0.087  0.071  0.23   0.001  0.437  0.001  0.019
0.    ]
 [ 0.     0.     0.     0.     0.     0.023  0.     0.918  0.
0.059]
 [ 0.     0.001  0.004  0.033  0.013  0.047  0.026  0.037  0.838
0.001]
 [ 0.     0.     0.     0.007  0.     0.031  0.     0.084  0.001
0.877]]

# 4.class have the highest/lowest posterior probabilities and test
examples

highest posterior probabilities(0.918):    Sneaker(class=7)

lowest posterior probabilities(0.418):     Pullover(class=2)
```

confusion matrix Image:



highest/lowest posterior probabilities test sample:

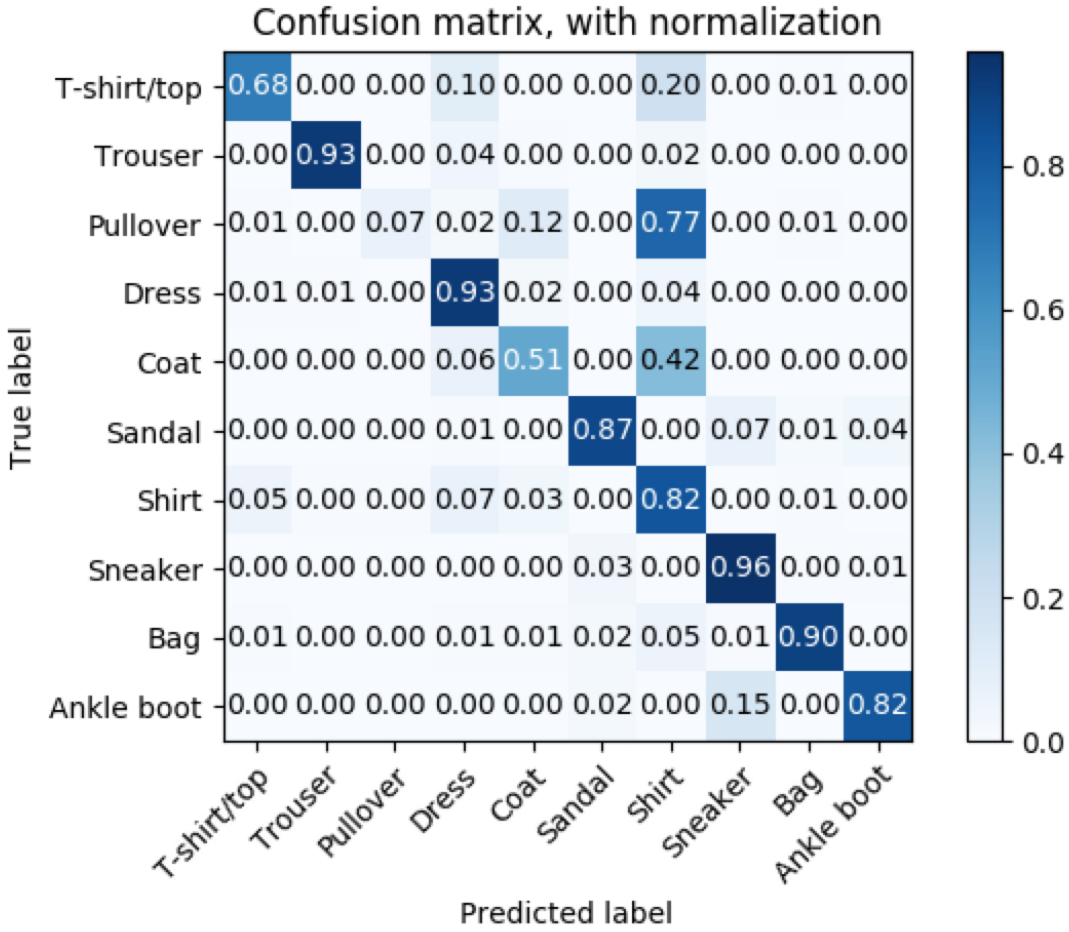


- **PART 1.2**

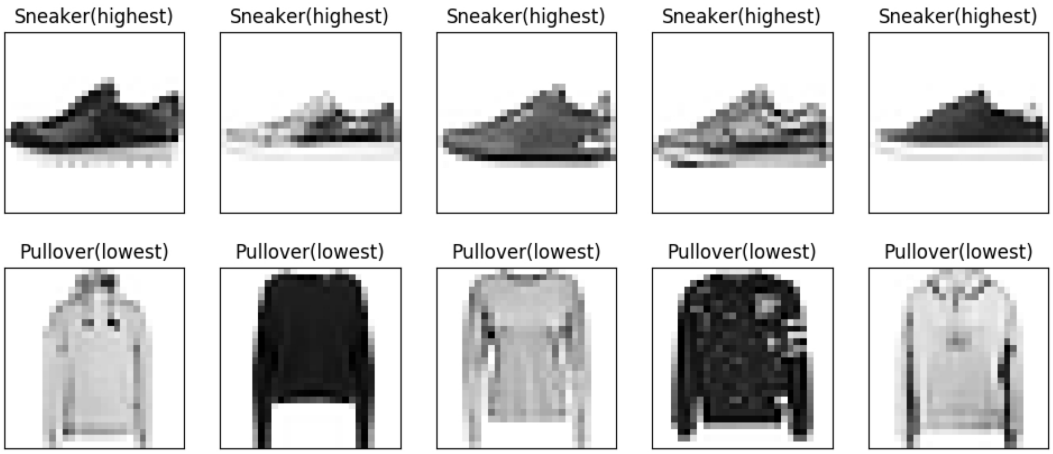
```
# 1.average classification rate
```

```
Perceptron avg_acc:  0.7488

# 2.the classification rate for each class
T-shirt/top acc rate(class=0): 0.680
Trouser acc rate(class=1): 0.927
Pullover acc rate(class=2): 0.072
Dress acc rate(class=3): 0.925
Coat acc rate(class=4): 0.512
Sandal acc rate(class=5): 0.872
Shirt acc rate(class=6): 0.825
Sneaker acc rate(class=7): 0.958
Bag acc rate(class=8): 0.895
Ankle boot acc rate(class=9): 0.822

# 3.confusion matrix
[[ 0.68   0.001  0.001  0.102  0.002  0.002  0.202  0.     0.007
0.003]
 [ 0.003  0.927  0.     0.042  0.004  0.     0.024  0.     0.
0.   ]
 [ 0.011  0.001  0.072  0.021  0.118  0.001  0.767  0.     0.009
0.   ]
 [ 0.006  0.009  0.     0.925  0.02   0.     0.038  0.     0.002
0.   ]
 [ 0.     0.001  0.     0.064  0.512  0.     0.42   0.     0.003
0.   ]
 [ 0.002  0.     0.     0.005  0.001  0.872  0.002  0.069  0.013
0.036]
 [ 0.054  0.004  0.004  0.071  0.034  0.     0.825  0.001  0.007
0.   ]
 [ 0.     0.     0.     0.     0.     0.033  0.     0.958  0.004
0.005]
 [ 0.005  0.001  0.     0.007  0.01   0.019  0.053  0.009  0.895
0.001]
 [ 0.     0.     0.     0.002  0.     0.021  0.     0.155  0.
0.822]]

# 4.class have the highest/lowest posterior probabilities and test
examples

highest posterior probabilities(0.918):    Sneaker(class=7)

lowest posterior probabilities(0.418):     Pullover(class=2)
```

confusion matrix Image:



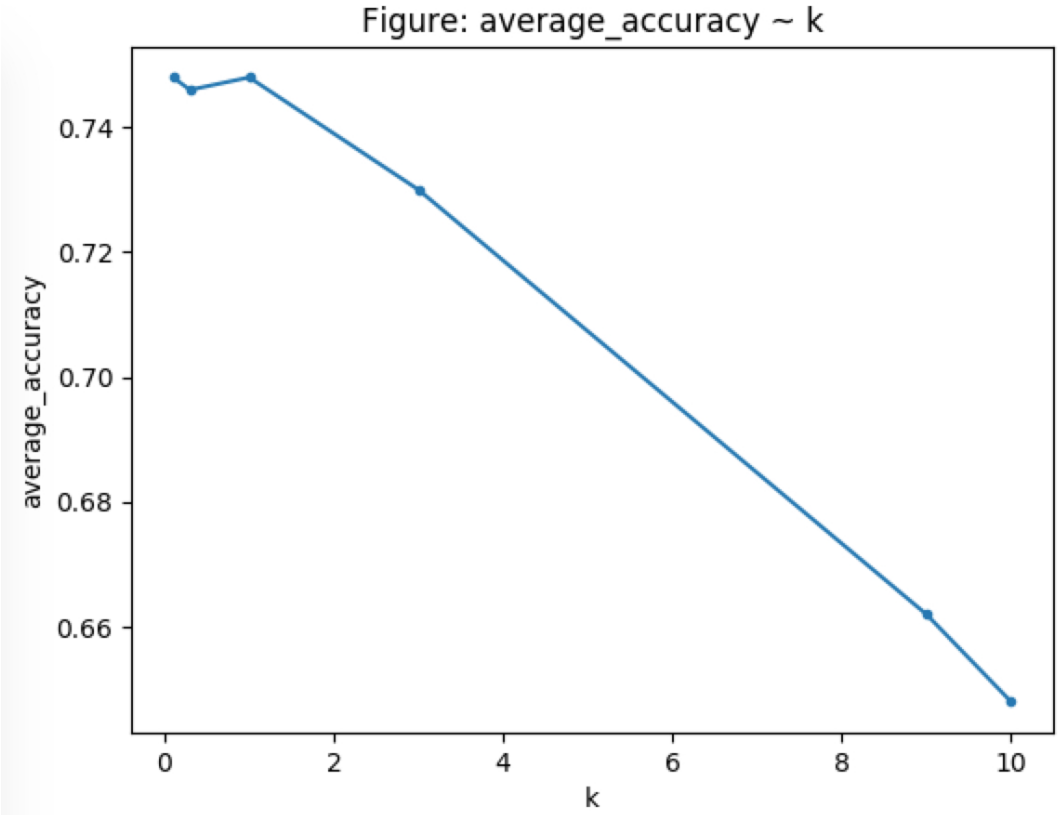
highest/lowest posterior probabilities test sample:



- **Experiment with different values of laplace smoothing parameter k**

Experiment with different `k` values `[0.1, 0.3, 1, 3, 9, 10]` and record the test results to plot the following figure.

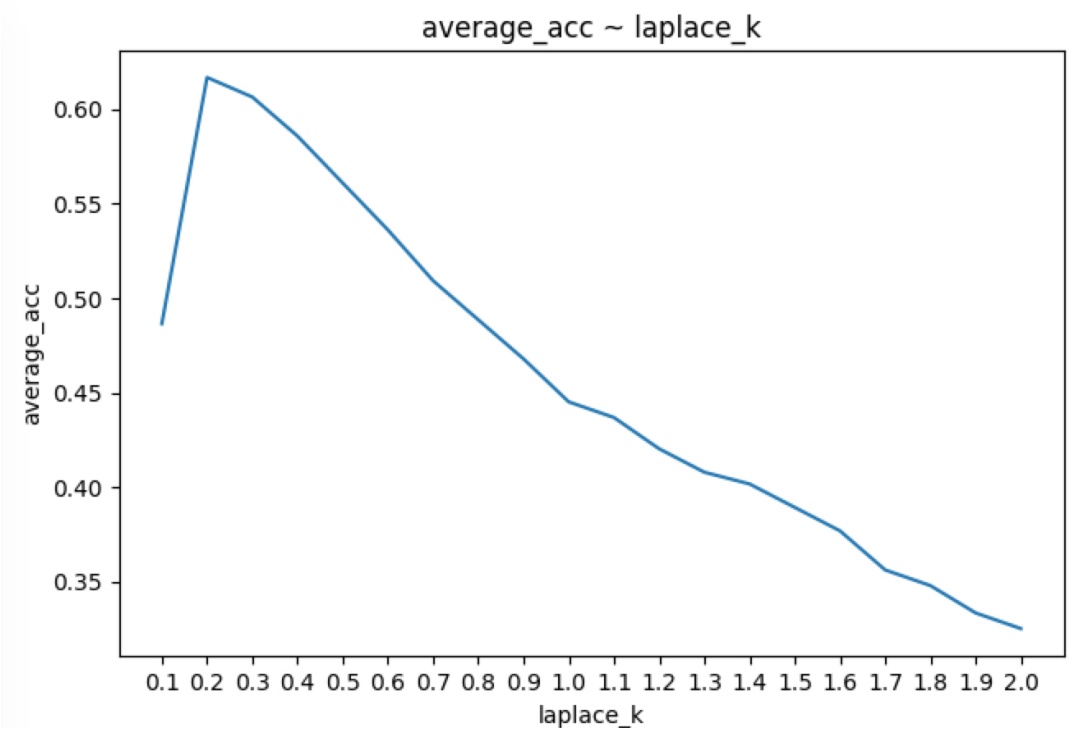
The experimental results show that when `k=1`, `average accurancy` gets the maximum value, so we take `k=1` .



3.Section II

Text Classification. Report all your results, confusion matrix ,recall ,precision, F1 score for all the 14 classes.

- fist I experiment with diffrent laplace parameter `k` to determine the best `k` for this model, the fig of average accuracy and laplace_k is following, and the best result is `lapk=0.2,average accuracy=0.617`



- set `k=0.2` ,all experiment result is following,

```
# 1.average classification rate
Perceptron avg_acc:  0.614907

# 2.recall ,precision, F1 score for all the 14 classes
---- Precision for all classes ----
Company: 0.580645
EducationalInstitution: 1.000000
Artist: 1.000000
Athlete: 1.000000
OfficeHolder: 0.428571
MeanOfTransportation: 0.972222
Building: 0.846154
NaturalPlace: 1.000000
Village: 0.091954
Animal: 0.888889
Plant: 0.545455
Album: 0.829268
Film: 0.820513
WrittenWork: 0.642857
---- Recall for all classes ----
Company: 0.878049
EducationalInstitution: 0.673913
Artist: 0.142857
```

```
Athlete: 0.434783
OfficeHolder: 0.409091
MeanOfTransportation: 0.729167
Building: 0.366667
NaturalPlace: 0.088235
Village: 1.000000
Animal: 0.320000
Plant: 0.933333
Album: 0.809524
Film: 0.842105
WrittenWork: 0.771429
---- F1 Score for all classes ----
Company: 0.699029
EducationalInstitution: 0.805195
Artist: 0.250000
Athlete: 0.606061
OfficeHolder: 0.418605
MeanOfTransportation: 0.833333
Building: 0.511628
NaturalPlace: 0.162162
Village: 0.168421
Animal: 0.470588
Plant: 0.688525
Album: 0.819277
Film: 0.831169
WrittenWork: 0.701299

# 3.confusion matrix
[[0.878  0.0   0.0   0.0   0.0  0.024  0.0   0.0  0.049  0.0   0.0
0.024  0.0  0.024 ]
 [ 0.0  0.674  0.0   0.0   0.0   0.0  0.022  0.0  0.304  0.0   0.0
0.0   0.0   0.0  ]
 [0.048  0.0  0.143  0.0  0.238  0.0   0.0   0.0  0.048  0.0   0.0
0.238 0.143 0.143 ]
 [ 0.0   0.0   0.0  0.435 0.348  0.0   0.0   0.0  0.174  0.0   0.0
0.0   0.0  0.043 ]
 [ 0.0   0.0   0.0   0.0  0.409  0.0  0.045  0.0   0.5   0.0   0.0
0.0  0.045  0.0  ]
 [0.125  0.0   0.0   0.0   0.0  0.771  0.0   0.0  0.083  0.0   0.0
0.0  0.021  0.0  ]
 [ 0.1   0.0   0.0   0.0   0.0   0.0   0.4   0.0   0.5   0.0   0.0
0.0   0.0   0.0  ]
 [ 0.0   0.0   0.0   0.0   0.0   0.0   0.0  0.088 0.735 0.059 0.088
0.0   0.0  0.029 ]
 [ 0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   1.0   0.0   0.0
0.0   0.0   0.0  ]
 [ 0.04  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.28  0.68
0.0   0.0   0.0  ]
```

```
 [0.044  0.0   0.0   0.0   0.0   0.0   0.0   0.0  0.022  0.0  0.933
0.0   0.0   0.0  ]
 [0.048  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.833 0.024 0.095 ]
 [0.132  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0  0.816 0.053 ]
 [ 0.2   0.0   0.0   0.0   0.0   0.0   0.0   0.0  0.029  0.0   0.0
0.0   0.0  0.771 ]]

# the top 20 feature words for each of the classes (each row is a
class)
[[  company      based      business    founded     records
record      bergen     systems     services     office     products
also      toronto    university    school     located     national
including  established     life    ]
 [   school       high      located    university   college
public     schools     students   education    district     county
founded       one         new         part        city       united
established independent   catholic  ]
 [    born      american     known        new         band
writer       best        rock       music      musician      work
also       singer       york       album       books       author
former    university     one     ]
 [    born      football     played      league   professional
player      plays     footballer    former     national    american
also     currently     hockey      rugby        team     australian
november     world        new     ]
 [    born       member     district   politician    state
house     democratic    senate      party       served      former
county      since    representatives republican    united
elected     american    national  representing]
 [    navy       built        war         ship        uss
united      class      aircraft     world       states     launched
service      first       named      designed     royal
commissioned  american       ii          us     ]
 [  historic     house       built      located      church
building    national    register     places      listed      county
street      united      known        also       museum      states
designed     added      hospital  ]
 [   river        lake      mountain    located      south
km        north       county       near     tributary      west
range        lies       creek        east       crater        ft
state       flows        pass    ]
 [  village     district   population   province    located
census   municipality   nepal       india       state       county
people        km        within       2010        1991      township
south      central       time    ]
 [   family     species      found       genus        moth
```

```
 gastropod      sea        known       marine    described
 tropical     snail      mollusk     endemic   subtropical   habitat
 natural     forests      snails      moist    ]
  [  species      family      plant       genus       native
 endemic    flowering     known       found       common      plants
 leaves     habitat       tree        name       grows       orchid
 south    bulbophyllum perennial  ]
  [   album      released      band      records      first
 studio     american     songs       music       second     release
 recorded      rock       debut        live       tracks      label
 albums       new          ep     ]
  [    film      directed    starring    american     stars
 released    written      based       drama       comedy
 produced      also       films       first       silent      movie
 roles        name       novel    documentary ]
  [ published      book       novel       first      journal
 written      series    newspaper    american     story       author
 new       magazine    fiction      books    peerreviewed    also
 science   publication     life    ]]
```

Also, report the change in accuracy results when the class prior changes to uniform distribution and when its removed. Provide the reasoning for these observations

- class prior is normal, `accuracy=0.617`
- class prior is uniform distribution, `accuracy=0.623`
- class prior is removed, `accuracy=0.640`

The reason why the accuracy changes is following,

the approximate value of the prior probability is

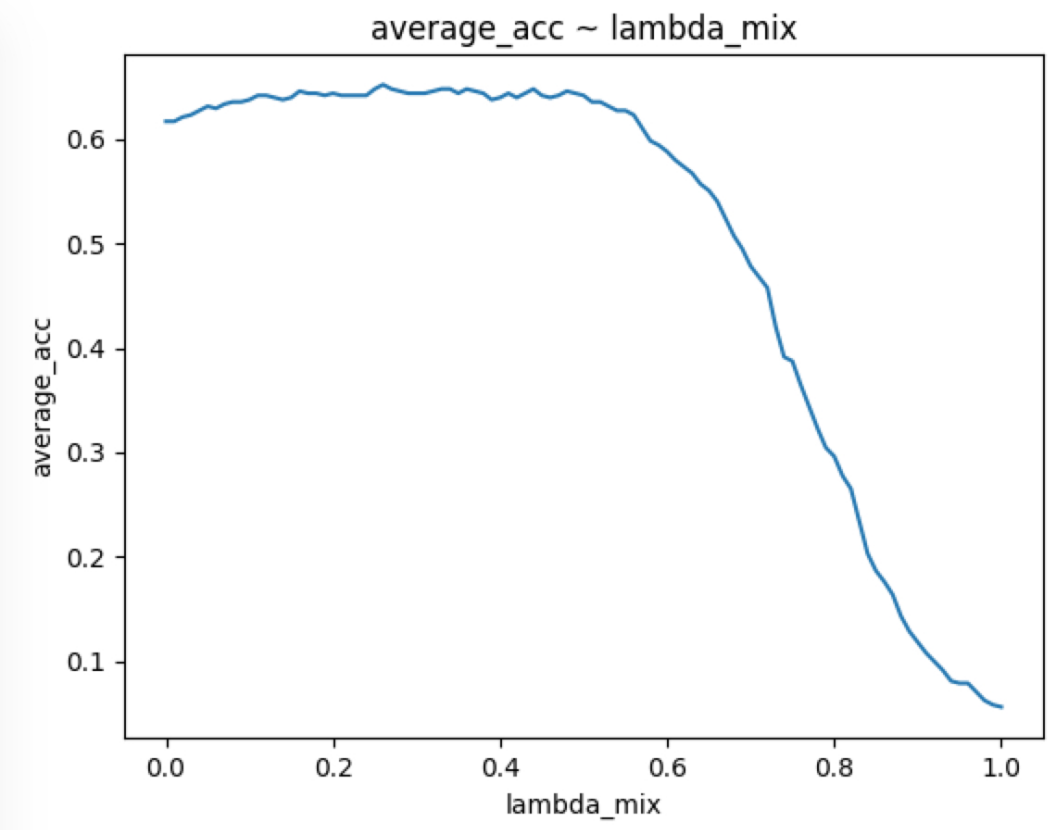
`[0.01, 0.023, 0.022, 0.035, 0.032, 0.065, 0.073, 0.062, 0.019, 0.117, 0.124, 0.146, 0.128, 0.144]`

It can be seen that the distribution of the samples is very unbalanced, so using such prior probability operations will result in a certain degree of bias in the classification. Even more, the class distribution of the test set is different from the class distribution of the training set, so this kind of Bias can cause a drop in accuracy.

4.Extra Credit

Did the bigram model help improve accuracy? Find the best parameter λ that gives the highest classification accuracy. Report the optimal parameter λ and report your results(Accuracy number) on the bigram model and optimal mixture model

experiment lambda_mix in `[round(0.01 * i, 2) for i in range(1, 101)]`, which means from 0 to 1 with step 0.01, the fig of `accuracy ~ lambda_mix` is following,



from the experiment result we get that `best parameter λ=0.26`, `highest classification accuracy=0.652174`

From the above figure, we can see that with the increase of `λ`, the value of `accuracy` starts to increase slightly. After `λ>0.6`, the value of `accuracy` drops significantly.

answer the following questions

- **Running naive Bayes on the bigram model relaxes the naive assumption of the model a bit. However, is this always a good thing? Why or why not?**

Relaxes the naive assumption means that the increase of `λ`, from the experimental results, this does not always improve the value of `accuracy`, if `λ` increases to a certain extent `(λ>0.6)`, leading that the results produced by the N-gram model dominate, and the value of `accuracy` will produce a significant drop, The reason is analyzed as follows.

`naive model` assume that the appearance of each word is independent. The constraints of this model are weak and it is easy to reach the performance bottleneck. The advantage is that it is not easy to generate sparse data during statistics, and can provide a large number of statistical samples.

`N-gram model` assumes that the appearance of a word is related to the appearance of its previous `N-1` words. The constraints of this model are relatively strong and conform to the actual situation, but it is easy to generate sparse data during statistics, resulting in results. A large number of 0 samples appear, and unless there are a large number of training samples, providing reliable data, it is less effective when applied alone.

Therefore, the appropriate `λ` means resonable combine with the `naive model` and `N-gram model`, which can obtain a better `accuracy`.

- **What would happen if we did an N-gram model where N was a really large number?**

If we did an N-gram model where N was a really large number, there will be two problems.

First, the parameter space is too large, resulting in a dimensional disaster, a large amount of statistics job,causing training speed is significantly slow. It can't be practically applied in actually.

Second, it may lead to many 0 samples in the statistics, and the statistical information becomes very unreliable, which means that the size of the training corpus is difficult to meet the needs of the model.

Statement of Contribution

Specify which team-member performed which task. You are encouraged to make this a many-to-many mapping, if applicable. e.g., You can say that "Rahul and Jason both implemented the BFS function, their results were compared for debugging and Rahul's code was submitted. Jason and Mark both implemented the DFS function, Mark's code never ran successfully, so Jason's code was submitted. Section I of the report was written by all 3 team members. Section II by Mark and Jason, Section III by Rahul and Jason.".. and so on.