Machine Learning
Homework Assignment 2

Yuntong Wang
UNI: yw2768

February 26, 2016

1 Problem 1

1.1 Part 1

The log likelihood $\mathcal{L}$ of data $(x_1, y_1), ..., (x_n, y_n)$ using an i.i.d. assumption is:

$$\begin{aligned}\mathcal{L} &= \ln p(y_1, ..., y_n | x_1, ..., x_n, w_1, ..., w_k) \\ &= \ln \prod_{i=1}^{k} \prod_{t=1}^{n} \left(\frac{e^{x_t^T w_i}}{\sum_{j=1}^{k} e^{x_t^T w_j}}\right)^{1(y_t = i)}\end{aligned} \tag{1}$$

For each data point (x_t, y_t) where t is in the range of 1 to n, there is only single one class i it belongs to where $i = y_t$. So when $i = y_t$, we have

$$\prod_{t=1}^{n} \left(\frac{e^{x_t^T w_{y_t}}}{\sum_{j=1}^{k} e^{x_t^T w_j}}\right)^{1(y_t = y_t)} = \prod_{t=1}^{n} \left(\frac{e^{x_t^T w_{y_t}}}{\sum_{j=1}^{k} e^{x_t^T w_j}}\right)$$

And for the rest of classes i where $i \neq y_t$, we have:

$$\prod_{t=1}^{n} \left(\frac{e^{x_t^T w_i}}{\sum_{j=1}^{k} e^{x_t^T w_j}}\right)^{0} = 1$$

So now we have

$$\begin{aligned}\mathcal{L} &= \ln \prod_{t=1}^{n} \frac{e^{x_t^T w_{y_t}}}{\sum_{j=1}^{k} e^{x_t^T w_j}} \\ &= \sum_{t=1}^{n} \left(\ln e^{x_t^T w_{y_t}} - \ln \sum_{j=1}^{k} e^{x_t^T w_j} \right) \\ &= \sum_{t=1}^{n} \left(x_t^T w_{y_t} - \ln \sum_{j=1}^{k} e^{x_t^T w_j} \right)\end{aligned} \tag{2}$$

1.2 Part 2

Now we need to calculate $\nabla_{w_i}\mathcal{L}$ and $\nabla^2_{w_i}\mathcal{L}$.

$$\nabla_{w_i}\mathcal{L} = \sum_{t=1}^{n}\left(x_t^T \cdot 1\{y_t = i\} - \frac{x_t^T e^{x_t^T w_i}}{\sum_{j=1}^{k} e^{x_t^T w_j}}\right) \tag{3}$$

$$\nabla^2_{w_i}\mathcal{L} = -\sum_{t=1}^{n}\frac{1}{\sum_{j=1}^{k} e^{x_t^T w_j}}\left(x_t^T\right)^2 e^{x_t^T w_i} \tag{4}$$

2 Problem 2

$$\begin{aligned} k(u,v) &= \int \phi_t(u)\phi_t(v)dt \\ &= \int \frac{1}{(2\pi\nu)^d}\exp\left(-\frac{||u-t||^2 + ||v-t||^2}{2\nu}\right)dt \end{aligned} \tag{5}$$

$$\begin{aligned} ||u-t||^2 + ||v-t||^2 &= u^Tu + t^Tt - 2u^Tt + v^Tv + t^Tt - 2v^Tt \\ &= 2||t||^2 - 2(u+v)^Tt + ||u||^2 + ||v||^2 \\ &= 2\left(||t||^2 - (u+v)^Tt + ||u||^2 + ||v||^2\right) \\ &= 2\left(||t||^2 - (u+v)^Tt + ||\frac{u+v}{2}||^2 - ||\frac{u+v}{2}||^2 + ||u||^2 + ||v||^2\right) \\ &= 2\left(||t - \frac{u+v}{2}||^2 - ||\frac{u+v}{2}||^2 + ||u||^2 + ||v||^2\right) \end{aligned} \tag{6}$$

So, there is:

$$\begin{aligned} k(u,v) &= \int \frac{1}{(2\pi\nu)^d}\exp\left(-\frac{||t-\frac{u+v}{2}||^2 - ||\frac{u+v}{2}||^2 + ||u||^2 + ||v||^2}{\nu}\right)dt \\ &= \frac{1}{(2\pi\nu)^{d/2}}\exp\left(-\frac{-||\frac{u+v}{2}||^2 + ||u||^2 + ||v||^2}{\nu}\right)\int \frac{1}{(2\pi\nu)^{d/2}}\exp\left(-\frac{1}{\nu}\left(||t - \frac{u+v}{2}||^2\right)\right)dt \end{aligned} \tag{7}$$

Since there is

$$\int \frac{1}{(2\pi\nu)^{d/2}}\exp\left(-\frac{1}{\nu}\left(||t - \frac{u+v}{2}||^2\right)\right)dt = \frac{1}{2^{d/2}}$$

and

$$\begin{aligned} -||\frac{u+v}{2}||^2 + ||u||^2 + ||v||^2 &= \frac{1}{4}\left(2||u||^2 - (||u||^2 + 2u^Tv + ||v||^2) + 2||v||^2\right) \\ &= \frac{||u-v||^2}{4} \end{aligned} \tag{8}$$

Therefore, we can get:

$$k(u,v) = \frac{1}{2^d(\pi\nu)^{d/2}} \exp\left(-\frac{||u-v||^2}{4\nu}\right)$$

And we can set:

$$\alpha = \frac{1}{2^d(\pi\nu)^{d/2}}$$

$$\beta = 4\nu$$

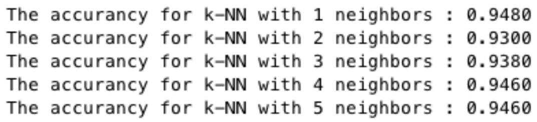
to get the Gaussian kernel:

$$k(u,v) = \alpha \exp\left(-\frac{||u-v||^2}{\beta}\right)$$

3 Problem 3

3.1 Problem 3a

2. Show the prediction accuracy of k-NN classifier for $k = 1,2,3,4,5$:



```
The accurancy for k-NN with 1 neighbors : 0.9480
The accurancy for k-NN with 2 neighbors : 0.9300
The accurancy for k-NN with 3 neighbors : 0.9380
The accurancy for k-NN with 4 neighbors : 0.9460
The accurancy for k-NN with 5 neighbors : 0.9460
```

Figure 1: prediction accuracy of k-NN classifier for $k = 1,2,3,4,5$

Misclassified examples:

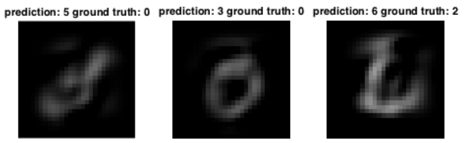

prediction: 5 ground truth: 0
prediction: 3 ground truth: 0
prediction: 6 ground truth: 2

Figure 2: missclassified images of k-NN classifier for $k = 1$

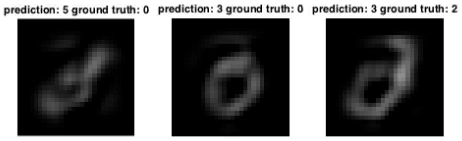

prediction: 5 ground truth: 0
prediction: 3 ground truth: 0
prediction: 3 ground truth: 2

Figure 3: missclassified images of k-NN classifier for $k = 3$

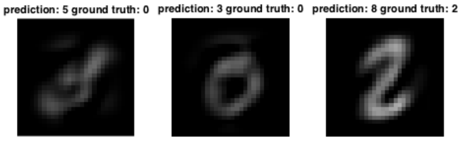

prediction: 5 ground truth: 0
prediction: 3 ground truth: 0
prediction: 8 ground truth: 2

Figure 4: missclassified images of k-NN classifier for $k = 5$

3.2 Problem 3b

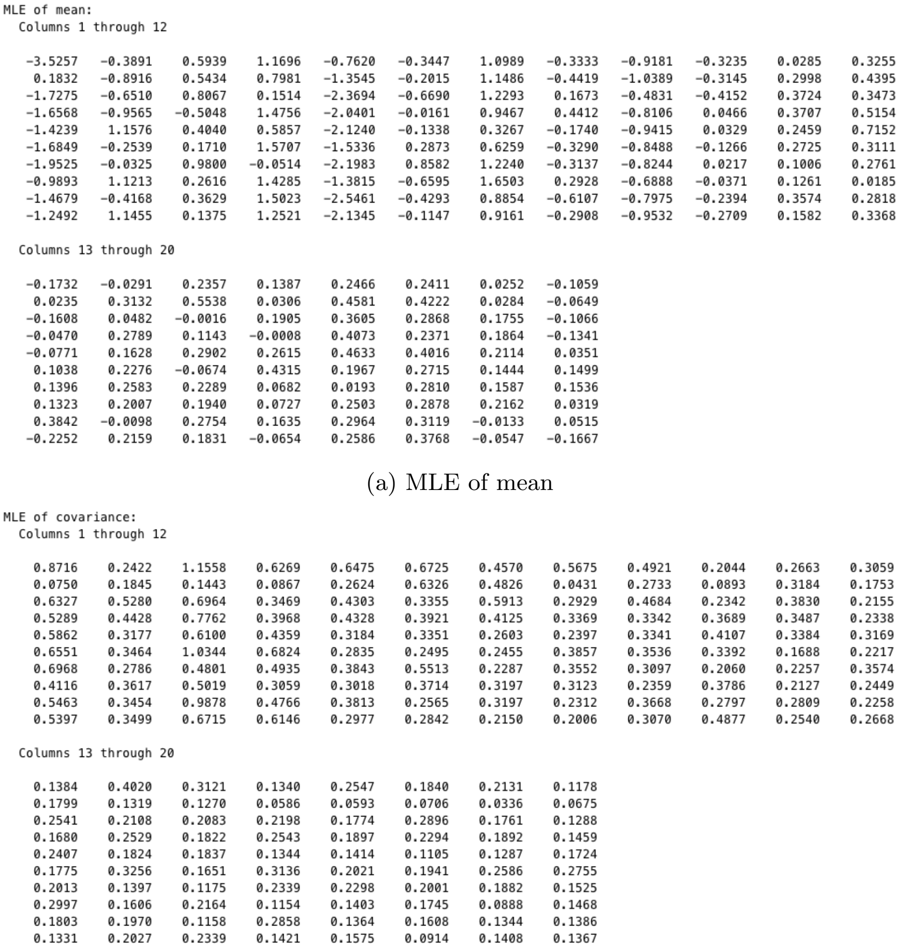
1. For a particular class j, where j is in the range 0 to 9 in this data set, the MLE for the mean and covariance are:

$$\hat{\mu}_j = \frac{1}{n_j}\sum_{i=1}^{n} 1(y_i = j)x_i$$

$$\hat{\Sigma}_j = \frac{1}{n_j}\sum_{i=1}^{n} 1(y_i = j)(x_i - \hat{\mu}_j)^2$$

where n_j denotes the number of training examples that belong to class j. $\hat{\mu}_j$ and $\hat{\Sigma}_j$ are all 1 x 20 vector for this data set.

The answer I obtained from the dataset are:



```
MLE of mean:
  Columns 1 through 12

   -3.5257   -0.3891    0.5939    1.1696   -0.7620   -0.3447    1.0989   -0.3333   -0.9181   -0.3235    0.0285    0.3255
    0.1832   -0.8916    0.5434    0.7981   -1.3545   -0.2015    1.1486   -0.4419   -1.0389   -0.3145    0.2998    0.4395
   -1.7275   -0.6510    0.8067    0.1514   -2.3694   -0.6690    1.2293    0.1673   -0.4831   -0.4152    0.3724    0.3473
   -1.6568   -0.9565   -0.5048    1.4756   -2.0401   -0.0161    0.9467    0.4412   -0.8106    0.0466    0.3707    0.5154
   -1.4239    1.1576    0.4040    0.5857   -2.1240   -0.1338    0.3267   -0.1740   -0.9415    0.0329    0.2459    0.7152
   -1.6849   -0.2539    0.1710    1.5707   -1.5336    0.2873    0.6259   -0.3290   -0.8488   -0.1266    0.2725    0.3111
   -1.9525   -0.0325    0.9800   -0.0514   -2.1983    0.8582    1.2240   -0.3137   -0.8244    0.0217    0.1006    0.2761
   -0.9893    1.1213    0.2616    1.4285   -1.3815   -0.6595    1.6503    0.2928   -0.6888   -0.0371    0.1261    0.0185
   -1.4679   -0.4168    0.3629    1.5023   -2.5461   -0.4293    0.8854   -0.6107   -0.7975   -0.2394    0.3574    0.2818
   -1.2492    1.1455    0.1375    1.2521   -2.1345   -0.1147    0.9161   -0.2908   -0.9532   -0.2709    0.1582    0.3368

  Columns 13 through 20

   -0.1732   -0.0291    0.2357    0.1387    0.2466    0.2411    0.0252   -0.1059
    0.0235    0.3132    0.5538    0.0306    0.4581    0.4222    0.0284   -0.0649
   -0.1608    0.0482   -0.0016    0.1905    0.3605    0.2868    0.1755   -0.1066
   -0.0470    0.2789    0.1143   -0.0008    0.4073    0.2371    0.1864   -0.1341
   -0.0771    0.1628    0.2902    0.2615    0.4633    0.4016    0.2114    0.0351
    0.1038    0.2276   -0.0674    0.4315    0.1967    0.2715    0.1444    0.1499
    0.1396    0.2583    0.2289    0.0682    0.0193    0.2810    0.1587    0.1536
    0.1323    0.2007    0.1940    0.0727    0.2503    0.2878    0.2162    0.0319
    0.3842   -0.0098    0.2754    0.1635    0.2964    0.3119   -0.0133    0.0515
   -0.2252    0.2159    0.1831   -0.0654    0.2586    0.3768   -0.0547   -0.1667
```

(a) MLE of mean

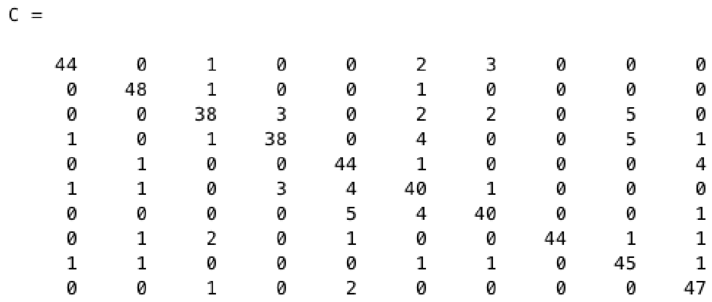
```
MLE of covariance:
  Columns 1 through 12

    0.8716    0.2422    1.1558    0.6269    0.6475    0.6725    0.4570    0.5675    0.4921    0.2044    0.2663    0.3059
    0.0750    0.1845    0.1443    0.0867    0.2624    0.6326    0.4826    0.0431    0.2733    0.0893    0.3184    0.1753
    0.6327    0.5280    0.6964    0.3469    0.4303    0.3355    0.5913    0.2929    0.4684    0.2342    0.3830    0.2155
    0.5289    0.4428    0.7762    0.3968    0.4328    0.3921    0.4125    0.3369    0.3342    0.3689    0.3487    0.2338
    0.5862    0.3177    0.6100    0.4359    0.3184    0.3351    0.2603    0.2397    0.3341    0.4107    0.3384    0.3169
    0.6551    0.3464    1.0344    0.6824    0.2835    0.2495    0.2455    0.3857    0.3536    0.3392    0.1688    0.2217
    0.6968    0.2786    0.4801    0.4935    0.3843    0.5513    0.2287    0.3552    0.3097    0.2060    0.2257    0.3574
    0.4116    0.3617    0.5019    0.3059    0.3018    0.3714    0.3197    0.3123    0.2359    0.3786    0.2127    0.2449
    0.5463    0.3454    0.9878    0.4766    0.3813    0.2565    0.3197    0.2312    0.3668    0.2797    0.2809    0.2258
    0.5397    0.3499    0.6715    0.6146    0.2977    0.2842    0.2150    0.2006    0.3070    0.4877    0.2540    0.2668

  Columns 13 through 20

    0.1384    0.4020    0.3121    0.1340    0.2547    0.1840    0.2131    0.1178
    0.1799    0.1319    0.1270    0.0586    0.0593    0.0706    0.0336    0.0675
    0.2541    0.2108    0.2083    0.2198    0.1774    0.2896    0.1761    0.1288
    0.1680    0.2529    0.1822    0.2543    0.1897    0.2294    0.1892    0.1459
    0.2407    0.1824    0.1837    0.1344    0.1414    0.1105    0.1287    0.1724
    0.1775    0.3256    0.1651    0.3136    0.2021    0.1941    0.2586    0.2755
    0.2013    0.1397    0.1175    0.2339    0.2298    0.2001    0.1882    0.1525
    0.2997    0.1606    0.2164    0.1154    0.1403    0.1745    0.0888    0.1468
    0.1803    0.1970    0.1158    0.2858    0.1364    0.1608    0.1344    0.1386
    0.1331    0.2027    0.2339    0.1421    0.1575    0.0914    0.1408    0.1367
```

(b) MLE of covariance

Figure 5: MLE for mean and covaiance

And the estimate of class prior for each class is all 0.1.

2. The confusion matrix is showed below:



```
C =

    44     0     1     0     0     2     3     0     0     0
     0    48     1     0     0     1     0     0     0     0
     0     0    38     3     0     2     2     0     5     0
     1     0     1    38     0     4     0     0     5     1
     0     1     0     0    44     1     0     0     0     4
     1     1     0     3     4    40     1     0     0     0
     0     0     0     0     5     4    40     0     0     1
     0     1     2     0     1     0     0    44     1     1
     1     1     0     0     0     1     1     0    45     1
     0     0     1     0     2     0     0     0     0    47
```

Figure 6: confusion matrix for problem 3b

And the prediction accuracy is 0.8560.

3. Show the mean of each Gaussian as an image:

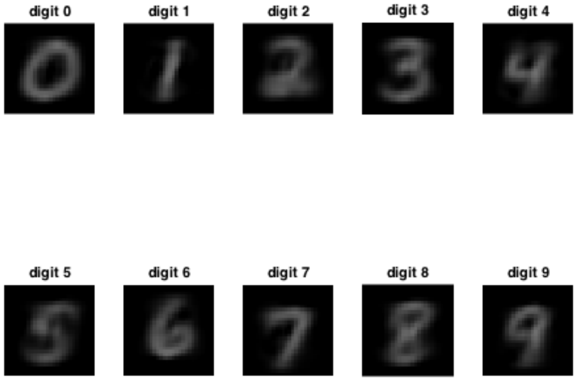


Figure 7: mean image for each class

4. Show three misclassified images and the probability distribution on the 10 digits leaned by the Bayes classifier:

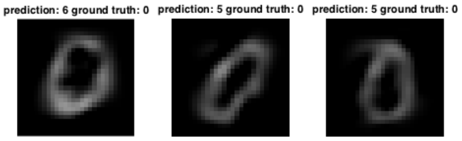
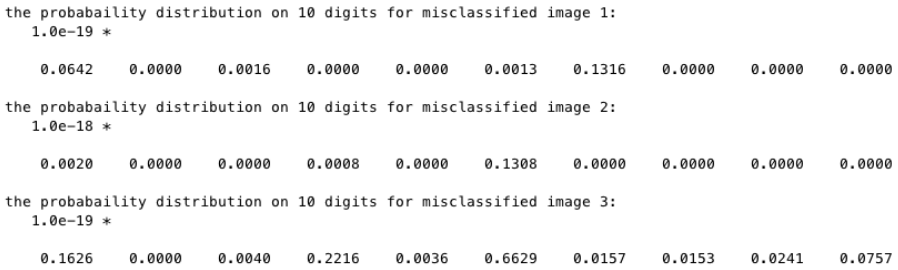

prediction: 6 ground truth: 0
prediction: 5 ground truth: 0
prediction: 5 ground truth: 0

Figure 8: three misclassified images examples



```
the probabaility distribution on 10 digits for misclassified image 1:
   1.0e-19 *

    0.0642    0.0000    0.0016    0.0000    0.0000    0.0013    0.1316    0.0000    0.0000    0.0000

the probabaility distribution on 10 digits for misclassified image 2:
   1.0e-18 *

    0.0020    0.0000    0.0000    0.0008    0.0000    0.1308    0.0000    0.0000    0.0000    0.0000

the probabaility distribution on 10 digits for misclassified image 3:
   1.0e-19 *

    0.1626    0.0000    0.0040    0.2216    0.0036    0.6629    0.0157    0.0153    0.0241    0.0757
```

Figure 9: probability distribution for three misclassified images

3.3 Problem 3c

2. Calculate $\mathcal{L}$ and plot as a function of iteration:

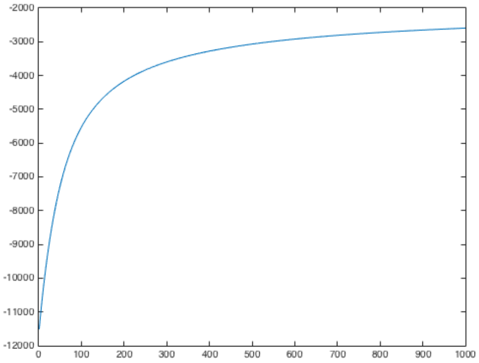
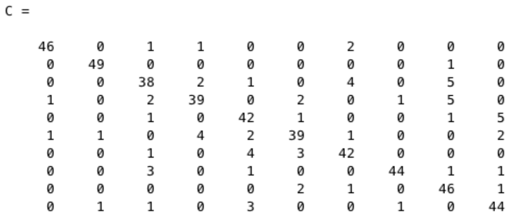


Figure 10: $\mathcal{L}$ as a function of iteration

3. The prediction accuracy is 0.8580. And confusion matrix is showed below:



```
C =

    46     0     1     1     0     0     2     0     0     0
     0    49     0     0     0     0     0     0     1     0
     0     0    38     2     1     0     4     0     5     0
     1     0     2    39     0     2     0     1     5     0
     0     0     1     0    42     1     0     0     1     5
     1     1     0     4     2    39     1     0     0     2
     0     0     1     0     4     3    42     0     0     0
     0     0     3     0     1     0     0    44     1     1
     0     0     0     0     0     2     1     0    46     1
     0     1     1     0     3     0     0     1     0    44
```

Figure 11: confusion matrix of problem 3c

4. Show three misclassified images and the probability distribution on the 10 digits leaned by the softmax function:

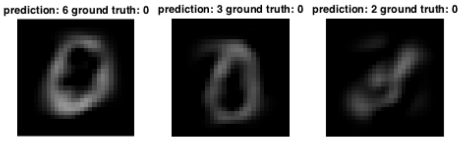
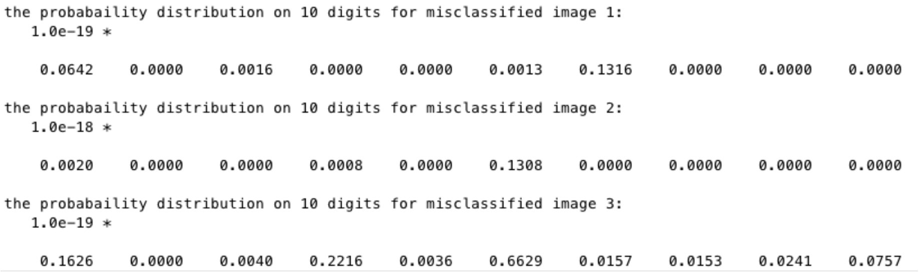

prediction: 6 ground truth: 0
prediction: 3 ground truth: 0
prediction: 2 ground truth: 0

Figure 12: three misclassified images examples



```
the probabaility distribution on 10 digits for misclassified image 1:
   1.0e-19 *

    0.0642    0.0000    0.0016    0.0000    0.0000    0.0013    0.1316    0.0000    0.0000    0.0000

the probabaility distribution on 10 digits for misclassified image 2:
   1.0e-18 *

    0.0020    0.0000    0.0000    0.0008    0.0000    0.1308    0.0000    0.0000    0.0000    0.0000

the probabaility distribution on 10 digits for misclassified image 3:
   1.0e-19 *

    0.1626    0.0000    0.0040    0.2216    0.0036    0.6629    0.0157    0.0153    0.0241    0.0757
```

Figure 13: probability distribution for three misclassified images