

Problem 3:

Part a:

The prediction accuracy for each value of k obtained by dividing the trace of the confusion matrix by 500 is below:

Prediction Accuracy:

$k=1$: 0.948

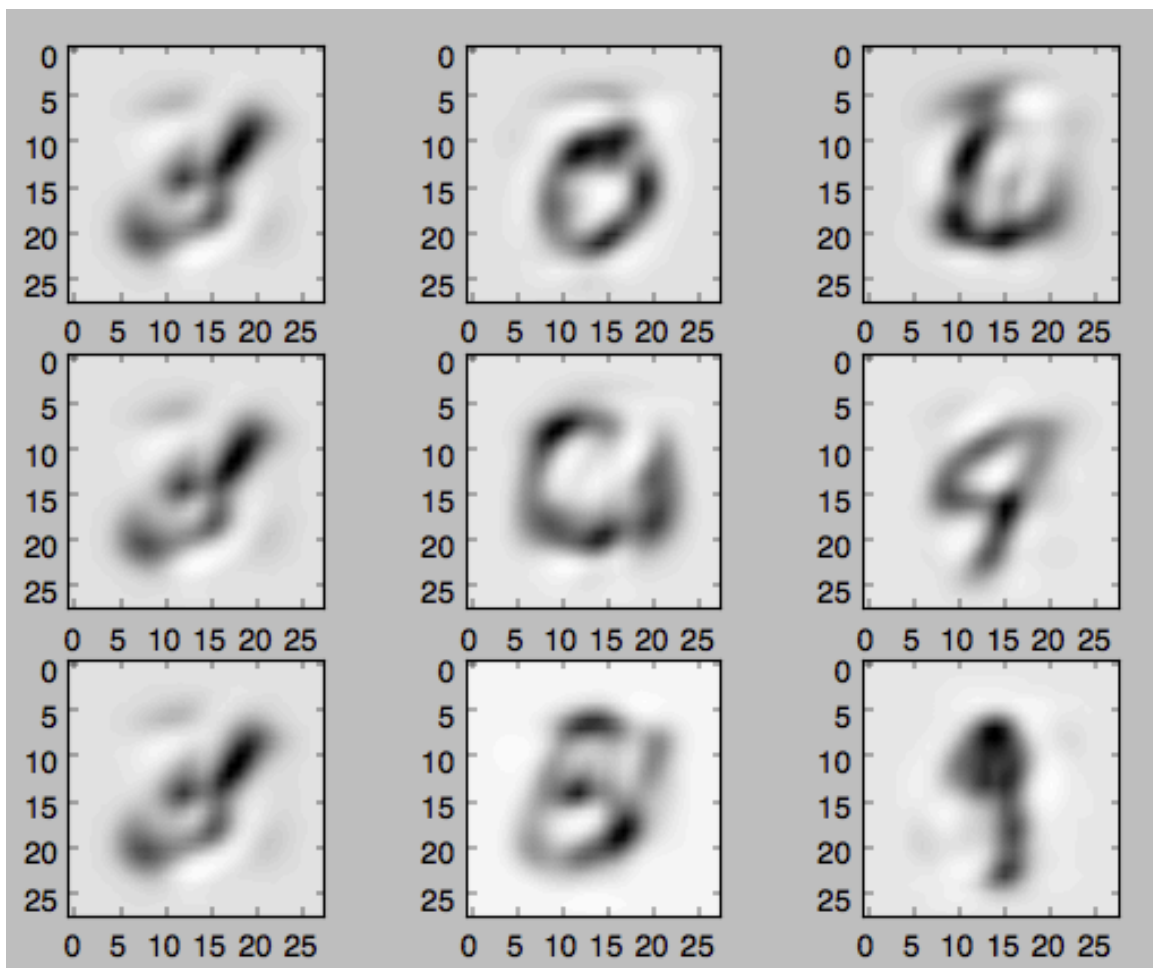
$k=2$: 0.93

$k=3$: 0.938

$k=4$: 0.946

$k=5$: 0.946

The plots of the misclassified examples are below. The first row is for $k=1$, second row is for $k=3$, and third row for $k=5$:



The true class and predicted class for each of the examples in the plots above are indicated below:

```

k=1:
True:  0.0 Predicted:  5.0
True:  0.0 Predicted:  3.0
True:  2.0 Predicted:  6.0
k=3:
True:  0.0 Predicted:  5.0
True:  9.0 Predicted:  6.0
True:  9.0 Predicted:  4.0
k=5:
True:  0.0 Predicted:  5.0
True:  8.0 Predicted:  2.0
True:  9.0 Predicted:  1.0

```

Part b:

I first split up the training data in 10 sets of size 500 each for each class. I then calculated the mean and covariance matrix for each of the 10 classes using the formulas below where the x_i 's are the 20-feature training data of an individual class and $n=500$ for each class:

$$\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\hat{\Sigma}_{ML} = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_{ML})(x_i - \hat{\mu}_{ML})^T.$$

I then calculated the Bayes Classifier for each of the 500 testing examples by using multivariate Gaussian distributions for each class. Since there are 10 classes, each $\pi_i=1/10=0.1$. One example for class '0' is below (i.e. there is a subscript of '0' on the mean and covariance matrix). I plugged in $d=20$ (since there are 20 dimensions in the data). Then for x , I would plug in each of the 500 testing examples:

$$\frac{\pi_0}{2\pi^{d/2}\sqrt{\det(\Sigma_0)}} \exp \left[-\frac{1}{2}(x - \mu_0)^T \Sigma_0^{-1}(x - \mu_0) \right]$$

For a given observation in the testing set, the Bayes Classifier is a vector of 10 values (1 value for each class). To classify the testing examples, I chose the largest value since that implies the highest likelihood.

The confusion matrix and prediction accuracy for the Bayes classification method are below:

Confusion Matrix:

```
[[ 48.  0.  0.  1.  0.  1.  0.  0.  0.  0.]
 [  0. 49.  0.  0.  0.  0.  0.  0.  1.  0.]
 [  0.  0. 48.  0.  1.  0.  1.  0.  0.  0.]
 [  0.  0.  1. 47.  0.  0.  0.  0.  2.  0.]
 [  0.  0.  0.  0. 48.  0.  0.  0.  1.  1.]
 [  0.  0.  0.  1.  0. 45.  2.  0.  1.  1.]
 [  0.  0.  0.  0.  1.  5. 43.  0.  0.  1.]
 [  0.  0.  2.  0.  2.  0.  0. 46.  0.  0.]
 [  0.  0.  1.  0.  0.  1.  0.  0. 47.  1.]
 [  1.  0.  0.  0.  2.  0.  0.  0.  0. 47.]]
```

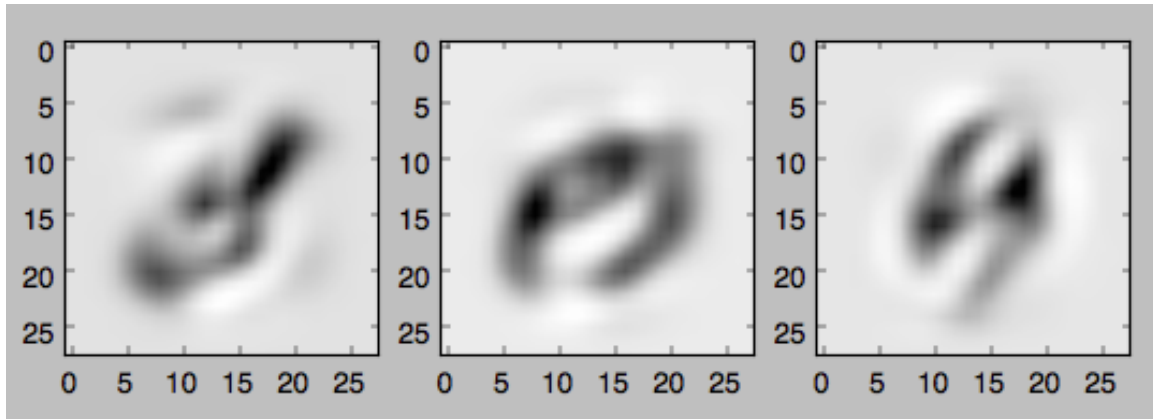
Prediction Accuracy:

Bayes classifier: 0.936

The means of each Gaussian as an image are show below:



The plots of three misclassified examples are below:

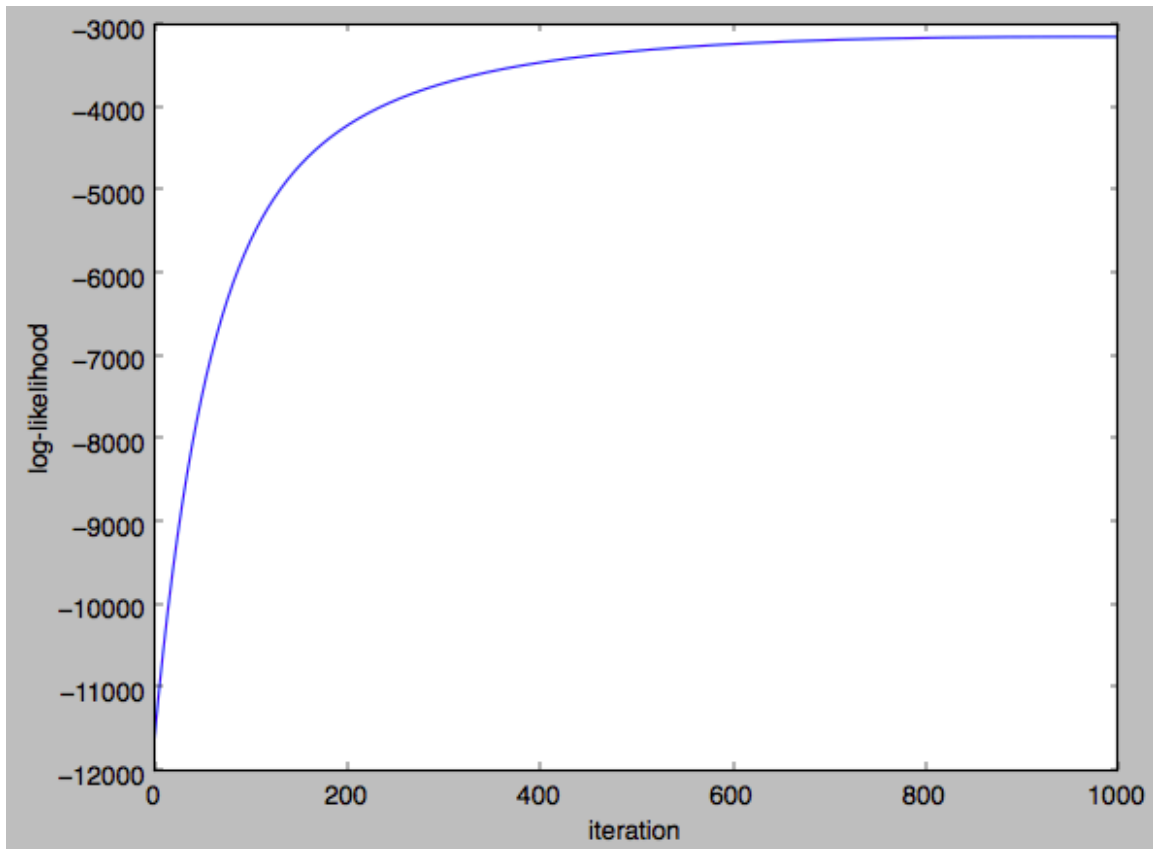


The true class and predicted class for each of the examples in the plots above as well as the probability distributions are indicated below. The testing examples were classified according to the maximum value component of the Bayes Classifier (since that implies the highest likelihood):

```
Bayes classifier:
True: 0.0 Predicted: 3.0
[ 1.38633222e-12  6.92564713e-35  2.05828743e-10  1.06512172e-08
 1.39248335e-14  1.28204537e-12  2.46494100e-12  4.61094028e-18
 8.66626849e-10  1.20554094e-16]
True: 0.0 Predicted: 5.0
[ 7.09180549e-31  0.00000000e+00  5.58282106e-31  2.43796749e-38
 1.60134371e-42  2.98180271e-26  4.61632136e-33  2.32385387e-72
 1.52278882e-27  2.57820637e-46]
True: 9.0 Predicted: 4.0
[ 2.62571528e-010  4.32811319e-126  8.42028916e-009  1.12553203e-009
 5.91153455e-007  2.11024461e-012  4.71944985e-011  2.79340367e-011
 8.65988632e-011  1.86543008e-007]
```

Part c:

The plot of the log-likelihood as a function of iteration is plotted below:



The Confusion matrix and prediction accuracy of this method are show below:

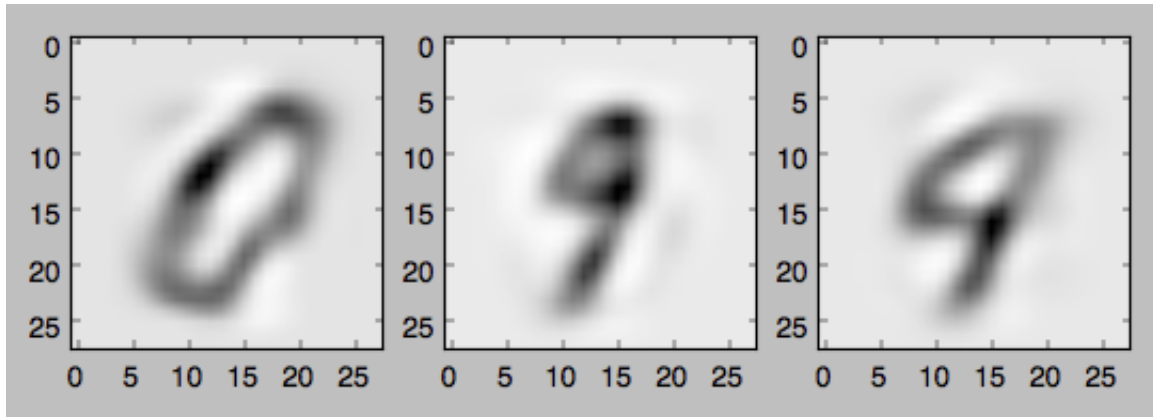
Confusion Matrix:

```
[[ 43.  0.  1.  0.  0.  5.  1.  0.  0.  0.]
 [  0. 40.  0.  0.  0.  2.  0.  0.  8.  0.]
 [  1.  0. 36.  3.  0.  0.  3.  0.  7.  0.]
 [  1.  0.  1. 38.  0.  3.  0.  0.  7.  0.]
 [  0.  0.  2.  0. 40.  1.  0.  0.  2.  5.]
 [  0.  1.  0.  6.  2. 38.  0.  0.  1.  2.]
 [  0.  0.  1.  0.  8.  4. 35.  0.  2.  0.]
 [  0.  0.  2.  0.  1.  0.  0. 42.  4.  1.]
 [  0.  0.  0.  0.  0.  3.  0.  0. 46.  1.]
 [  0.  0.  1.  0.  2.  1.  0.  0.  1. 45.]]
```

Prediction Accuracy:

Logistic regression: 0.806

The plots of three misclassified examples are below:



The true class and predicted class for each of the examples in the plots above as well as the probability distributions are indicated below. The examples were classified according to the maximum (i.e. least negative) log-likelihood value (since that implies the highest likelihood):

Logistic regression:

True: 0.0 Predicted: 5.0

```
[ -1.30339017 -20.01474182 -10.66162449 -6.81624062 -10.10648766  
  -0.32044163 -8.88818025 -13.60189305 -6.69234482 -10.70406341]
```

True: 9.0 Predicted: 8.0

```
[-14.53846085 -3.77826431 -7.91069517 -2.69593579 -6.42182671  
  -1.58180946 -10.04792203 -4.94845782 -0.60931779 -1.8891289 ]
```

True: 9.0 Predicted: 4.0

```
[-13.88890932 -21.41929589 -13.08453305 -13.83016452 -0.24477042  
  -7.5210104 -12.03997375 -5.50804167 -8.36783885 -1.54985851]
```