Problem 1:

Part 1:

(raised to power of 0)

$z^0 = 1$

$1 \cdot 1 \cdot \frac{e^{x^T w_i}}{\sum_{j=1}^{k} e^{x^T w_j}} \cdot 1 \cdot 1$

Anubha Bhargava

① $P(y|x, w_1 \dots w_K) = \prod_{i=1}^{K} \left(\frac{e^{x^T w_i}}{\sum_{j=1}^{K} e^{x^T w_j}} \right)^{\mathbb{1}(y=i)}$

(1) $P(y_1 \dots y_l | x_1 \dots x_l, w_1 \dots w_K) = \prod_{l=1}^{n} \prod_{i=1}^{K} \left(\frac{e^{x^T w_i}}{\sum_{j=1}^{K} e^{x^T w_j}} \right)^{\mathbb{1}(y_l=i)} =$

$\prod_{l=1}^{n} \left(\frac{e^{x^T w_{y_l}}}{\sum_{j=1}^{k} e^{x^T w_j}} \right) = \prod_{l=1}^{n} P(y_l | x_l, w)$

↓ take ln

$\mathcal{L} = \sum_{l=1}^{n} \ln(e^{x_l^T w_{y_l}}) - \ln\left(\sum_{j=1}^{k} e^{x_l^T w_j} \right) =$

$\mathcal{L} = \sum_{l=1}^{n} x_l^T w_{y_l} - \ln\left(\sum_{j=1}^{k} e^{x_l^T w_j} \right)$

Part 2:

(2) $\nabla_{w_i} \mathcal{L} = \sum_{l=1}^{n} \left[\mathbb{1}(y_l = i)\, x_l - \frac{e^{x_l^T w_i} x_l}{\sum_{j=1}^{k} e^{x_l^T w_j}} \right] =$

$g = \sum_{j=1}^{k} e^{x_l^T w_j}$

$f = e^{x_l^T w_i} x_l$

$f' = x_l^2 e^{x_l^T w_i}$

$g' = x_l e^{x_l^T w_j}$

$$\nabla_{w_i}^2 \mathcal{L} = -\sum_{l=1}^{n} \frac{\left(\sum_{j=1}^{k} e^{x_l^T w_j}\right)\left(x_l^2 e^{x_l^T w_i}\right) - \left(x_l e^{x_l^T w_j}\right)\left(e^{x_l^T w_i} x_l\right)}{\left(\sum_{j=1}^{k} e^{x_l^T w_j}\right)^2}$$

$$= \sum_{l=1}^{n} -x_l^2 \left(\frac{e^{x_l^T w_i} \cdot \sum_{j=1}^{k} e^{x_l^T w_j} - e^{x_l^T w_i} e^{x_l^T w_j}}{\left(\sum_{j=1}^{k} e^{x_l^T w_j}\right)^2} \right)$$

$$= \sum_{l=1}^{n} -x_l^2 e^{x_l^T w_i} \left(\frac{\sum_{j=1}^{k} e^{x_l^T w_j} - e^{x_l^T w_j}}{\left(\sum_{j=1}^{k} e^{x_l^T w_j}\right)^2} \right)$$

$$= \boxed{\sum_{l=1}^{n} -x_l^2 e^{x_l^T w_i} \left(\frac{1}{\sum_{j=1}^{k} e^{x_l^T w_j}} - \frac{e^{x_l^T w_j}}{\sum_{j=1}^{k} e^{x_l^T w_j}} \right)}$$

Problem 2:

(1) $u, v \in \mathbb{R}^d$

$\phi(u)$ - high dimensional
$\phi(v)$ mappings

$$K(u,v) = \int_{R^d} \phi_t(u)\, \phi_t(v)\, dt$$

mapping

$$\phi_t(u) = \frac{1}{(2\pi\nu)^{d/2}} e^{\left\{-\frac{\|u-t\|^2}{2\nu}\right\}}$$

Gaussian kernel

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

$$K(u,v) = \int_{R^d} \left[\frac{1}{(2\pi\nu)^{d/2}} e^{\left\{\frac{-\|u-t\|^2}{2\nu}\right\}}\right] \cdot \left[\frac{1}{(2\pi\nu)^{d/2}} e^{\left\{\frac{-\|v-t\|^2}{2\nu}\right\}}\right] dt =$$

$$K(u,v) = \int_{R^d} \left(\frac{1}{(2\pi\nu)^{d/2}}\right)^2 e^{\left\{\frac{-\|u-t\|^2}{2\nu}\right\}} e^{\left\{\frac{-\|v-t\|^2}{2\nu}\right\}}$$

$$K(u,v) = \frac{1}{(2\pi\nu)^d} \int_{R^d} e^{\left\{-\left(\frac{\|u-t\|^2+\|v-t\|^2}{2\nu}\right)\right\}} dt =$$

$\|u-t\|^2 + \|v-t\|^2 = (u^Tu - 2u^Tt + t^Tt) + (v^Tv - 2v^Tt + t^Tt) = u^Tu + v^Tv - 2u^Tt - 2v^Tt + 2t^Tt =$

$$= \|u\|^2 + \|v\|^2 + \underbrace{2\|t\|^2 - 2(u+v)^Tt}_{\left\|t - \frac{u+v}{2}\right\|^2 - \frac{\|u+v\|^2}{4}} \quad (a^2 - 2ab + b^2)$$

$$= \|u\|^2 + \|v\|^2 + 2\left(\left\|t - \frac{u+v}{2}\right\|^2\right) - 2\left(\frac{\|u+v\|^2}{4}\right)$$

$$= -\frac{\|u\|^2 + \|v\|^2}{2\nu} - \frac{\left\|t - \frac{u+v}{2}\right\|^2}{\nu} + \frac{\|u+v\|^2}{4\nu}$$

$$= \frac{1}{(2\pi\nu)^d}\left(2\pi\frac{\nu}{2}\right)^{d/2}\Bigg[\underbrace{\frac{1}{(2\pi\frac{\nu}{2})^{d/2}}\int_{R^d} e^{\left(-\frac{\|t-\frac{u+v}{2}\|^2}{\nu}\right)}}_{\text{Gaussian } \int f(t)=1}\Bigg] e^{\left(-\frac{\|u\|^2+\|v\|^2}{2\nu}+\frac{\|u+v\|^2}{4\nu}\right)}$$

$$= \frac{1}{(2\pi\nu)^d}(\pi\nu)^{d/2}\left[e^{\left(-\frac{\|u\|^2+\|v\|^2}{2\nu}+\frac{\|u+v\|^2}{4\nu}\right)}\right]$$

$$u^Tu + 2u^Tv + v^Tv$$

$$= \frac{1}{(4\pi\nu)^{d/2}}\left[e^{\left(-\frac{2(u^Tu+v^Tv)}{4\nu}+\frac{(u+v)^T(u+v)}{4\nu}\right)}\right]$$

$$= \frac{1}{(4\pi\nu)^{d/2}}\, e^{\frac{-2u^Tu-2v^Tv+u^Tu+2u^Tv+v^Tv}{4\nu}}$$

$$-(u^Tu+v^Tv-2u^Tv)$$

$$= \frac{1}{(4\pi\nu)^{d/2}}\, e^{\frac{-u^Tu-v^Tv+2u^Tv}{4\nu}}$$

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

$$\boxed{\begin{aligned}\alpha &= \frac{1}{(4\pi\nu)^{d/2}}\\ \beta &= 4\nu\end{aligned}}$$

<u>Problem 3a:</u>

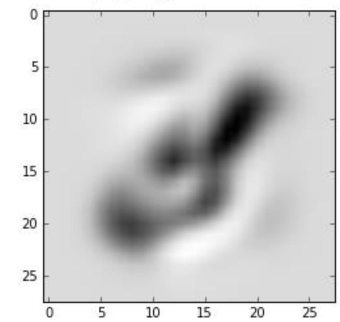
Implement the k-NN classifier for k=1,2,3,4,5.

For each k calculate the confusion matrix and show the trace of this matrix divided by 500. This is the prediction accuracy. You don't need to show the confusion matrix.

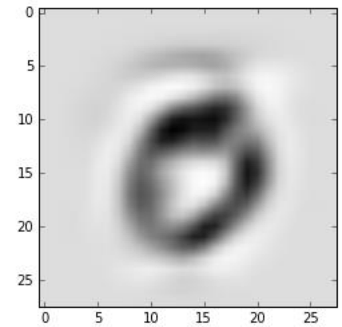
Prediction Accuracy for k=1,2,3,4,5 respectively: 94.8%, 94.8,% 94.4%, 94.6%, 94.6%

For k = 1,3,5, show three misclassified examples as images and indicate the true class and the predicted class for each one. The true class is what the correct value is. The predicted class is what was predicted.

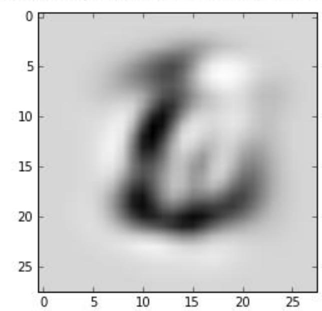
Misclassified Example (k=1), True Class: 0 Predicted Class: 5



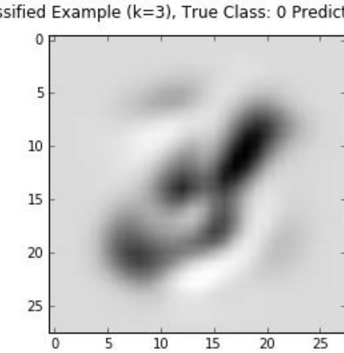
Misclassified Example (k=1), True Class: 0 Predicted Class: 3



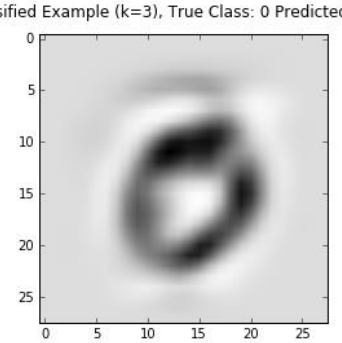
Misclassified Example (k=1), True Class: 2 Predicted Class: 6



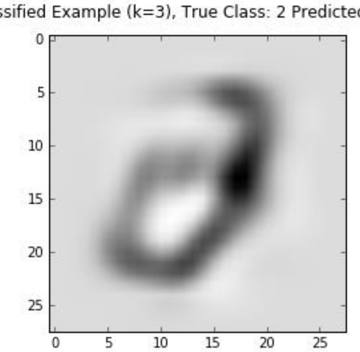
Misclassified Example (k=3), True Class: 0 Predicted Class: 5



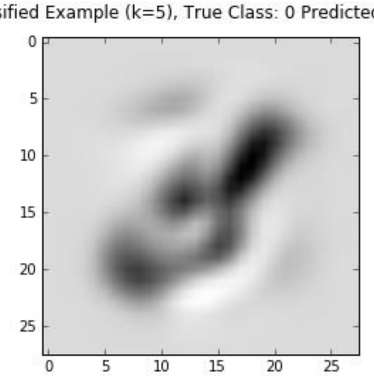
Misclassified Example (k=3), True Class: 0 Predicted Class: 3



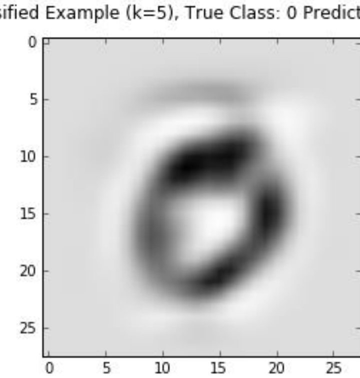
Misclassified Example (k=3), True Class: 2 Predicted Class: 3

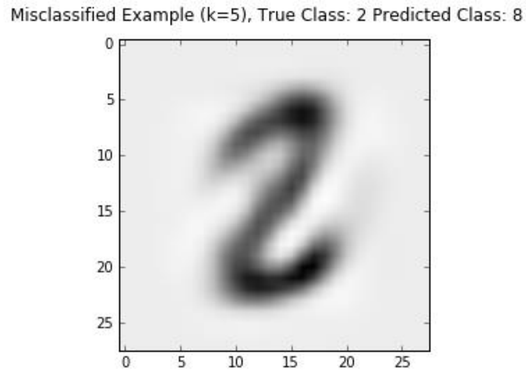


Misclassified Example (k=5), True Class: 0 Predicted Class: 5



Misclassified Example (k=5), True Class: 0 Predicted Class: 3





<u>Problem 3b:</u>

Implement the Bayes classifier using a class-specific multivariate Gaussian distribution. Derive the maximum likelihood estimate for the mean and covariance for a particular class j. Show the answer you obtain for the mean and covariance, as well as the estimate for the class prior.

I used the Bayes Classifier:

$$f(x) = \ argmax\ \pi|\Sigma|^{-1/2}e^{\{-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y)\}}$$

The following formula to calculate the mean:

$$\mu = \frac{1}{n_y}\Sigma 1(y_i = y)x_i$$

The following formula to calculate the covariance:

$$\Sigma = \frac{1}{n_y}\Sigma 1(y_i = y)(x_i - \ \mu_i)(x_i - \ \mu_i)^T$$

And lastly, the formula for prior:

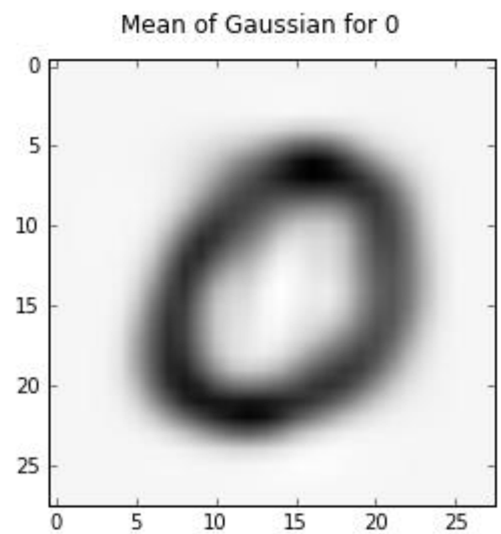
$$\pi_y = \frac{1}{n}\Sigma 1(y_i = y)$$

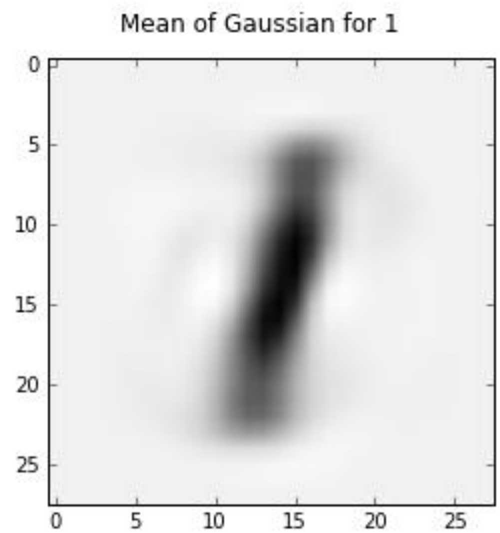
Show the confusion matrix in a table. As in problem 3a, indicate the prediction accuracy by summing along the diagonal and dividing by 500. The diagonal shows the number of matches for values 0-9 for predicted and true values.

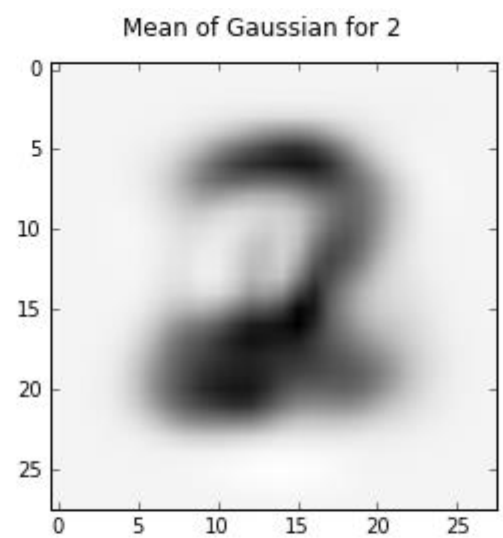
	Prediction									
True	48	0	0	1	0	1	0	0	0	0
	0	49	0	0	0	0	0	0	1	0
	0	0	47	1	1	0	1	0	0	0
	0	0	0	48	0	0	0	0	2	0
	0	0	0	0	48	0	0	1	0	1
	0	0	0	1	0	44	2	1	1	1
	0	0	0	0	1	4	44	0	1	0
	0	0	2	0	2	0	0	46	0	0
	0	0	1	0	0	1	0	0	47	1
	1	0	0	0	2	0	0	0	0	47

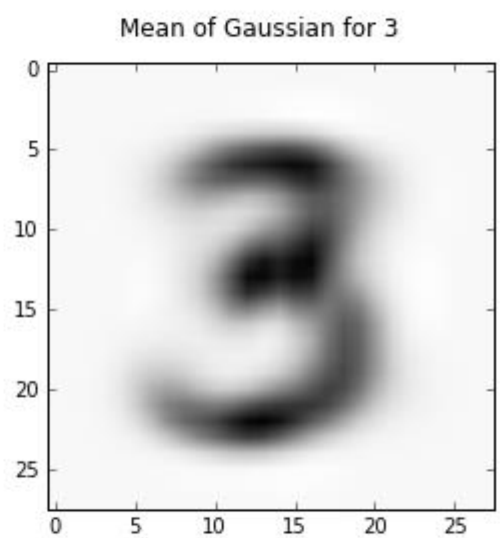
Prediction Accuracy for the Bayes Classifier: 93.6 %

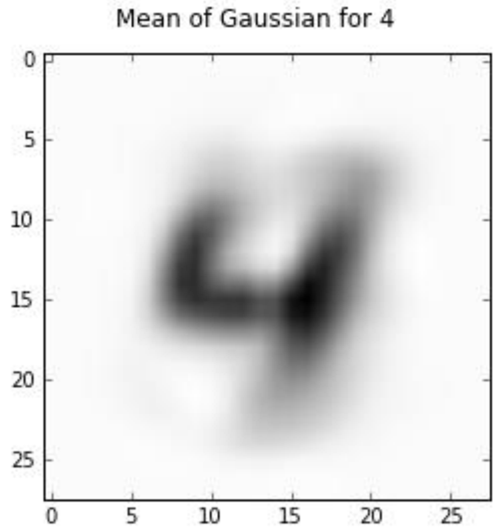
Show the mean of each Gaussian as an image using the provided Q matrix. The mean is the average of all the images.

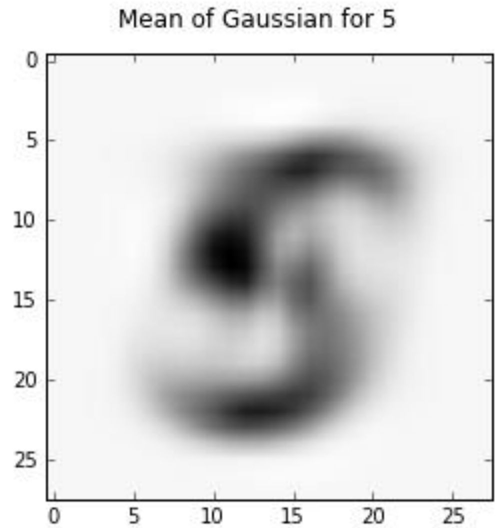


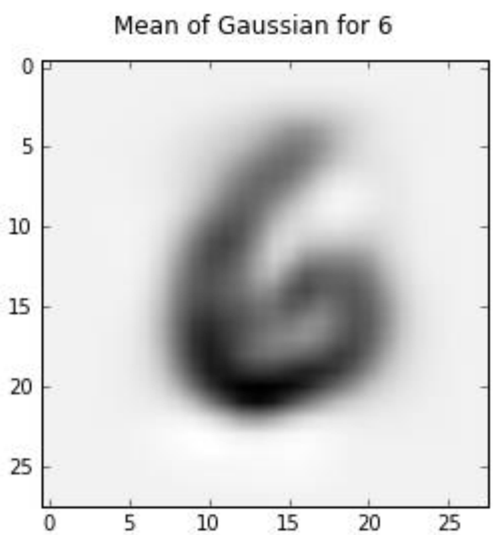


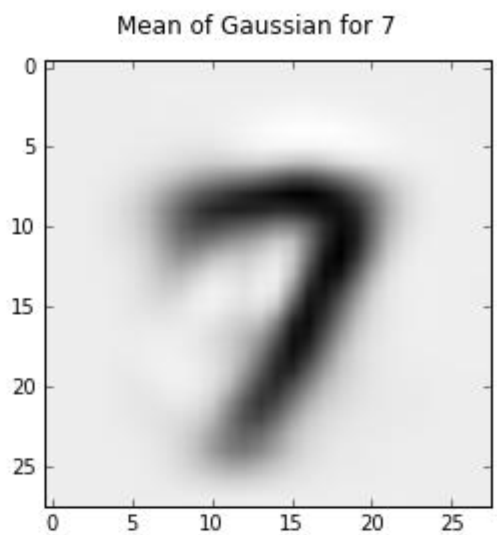


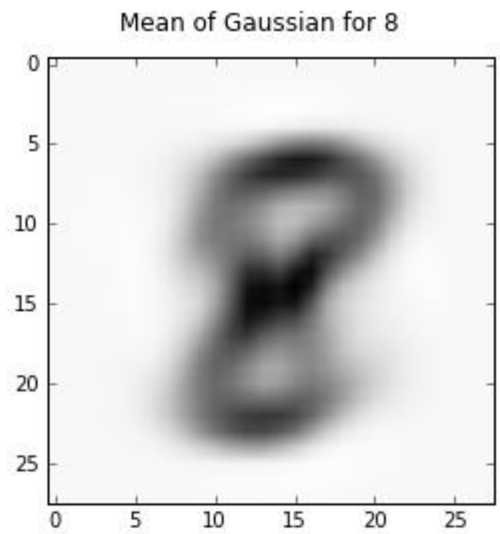


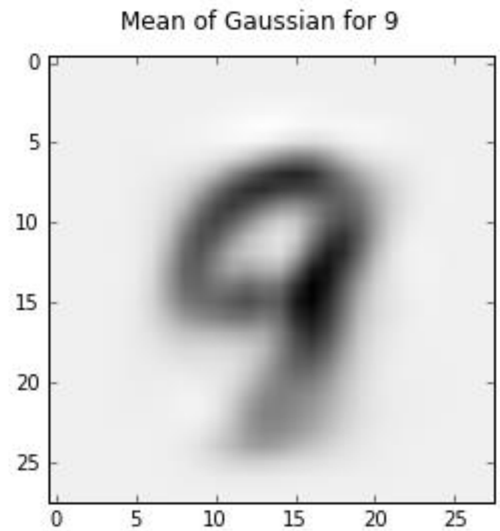

Mean of Gaussian for 4


Mean of Gaussian for 5

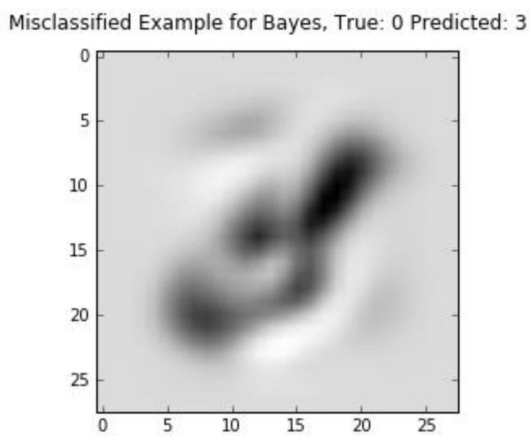

Mean of Gaussian for 6

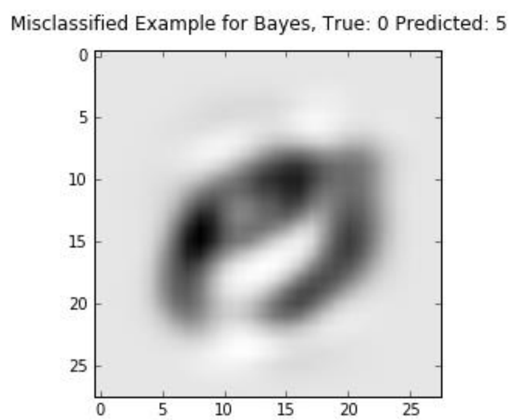

Mean of Gaussian for 7

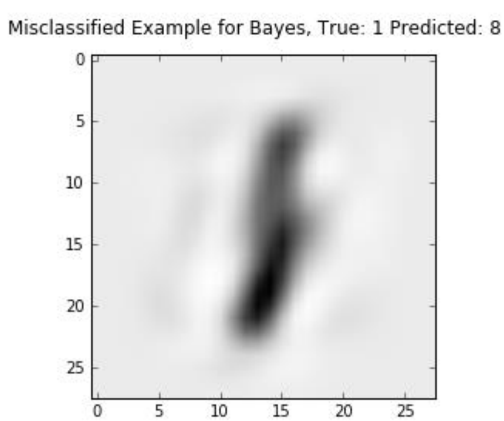




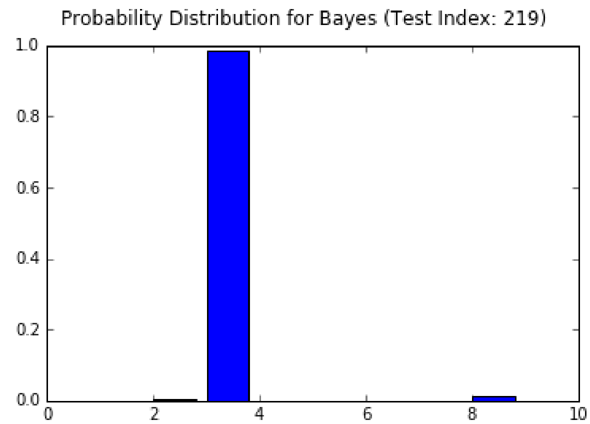
Show three misclassified examples as images and show the probability distribution on the 10 digits learned by the Bayes classifier for each one. The following plots show the misclassified instances of Bayes classifier.

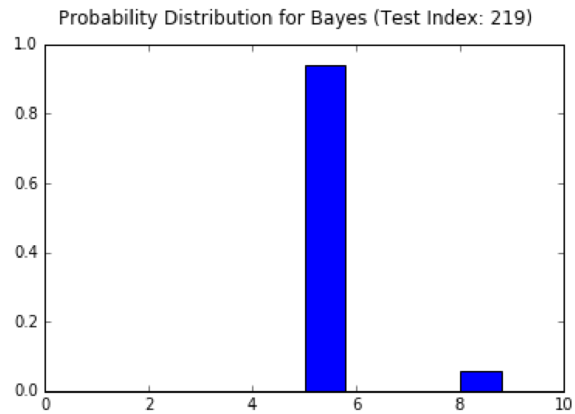


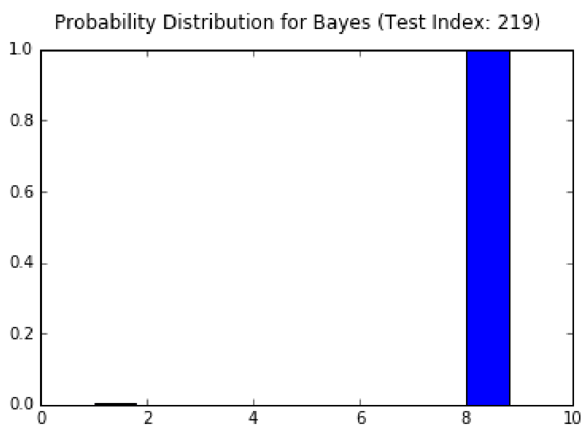




The following plots show the probability distribution for the Bayes classifier.

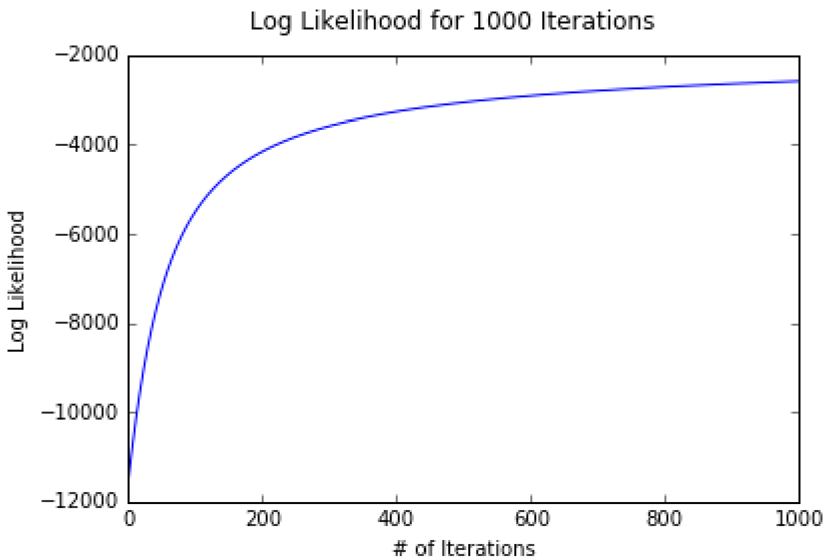






Problem 3c:

After making an update of each w_0,…w_9, calculate L and plot as a function of iteration. When getting different values of w, we get different values of log likelihood. With the more iterations, the line tends towards -2000.



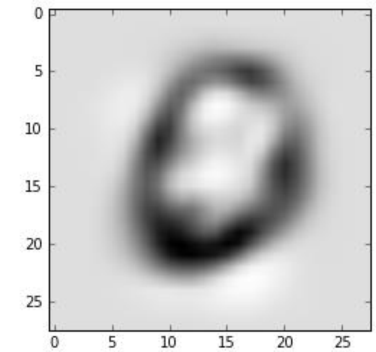
Show the confusion matrix in a table. Indicate the prediction accuracy by summing along the diagonal and dividing by 500. The diagonal shows the matches for values 0-9 from the predicted and true values.

	Prediction									
True	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

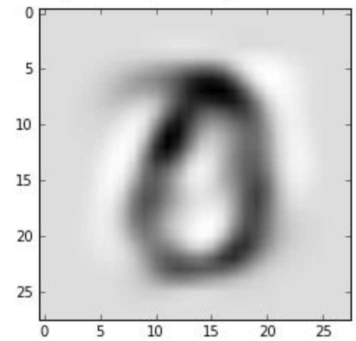
Prediction Accuracy for the Log Likelihood: 85.8 %

Show three misclassified examples as an image and show the probability distribution on the 10 digits learned by the softmax function for each one. The following plots show the misclassified cases for the multiclass logistic regression classifier.

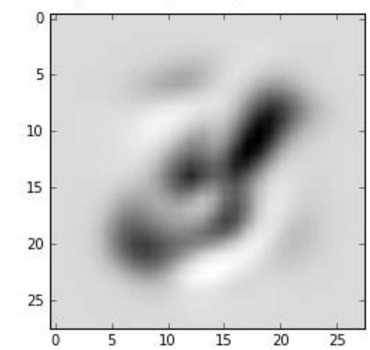
Misclassified Examples for Logistic Regression, True: 0 Predicted: 6



Misclassified Examples for Logistic Regression, True: 0 Predicted: 3

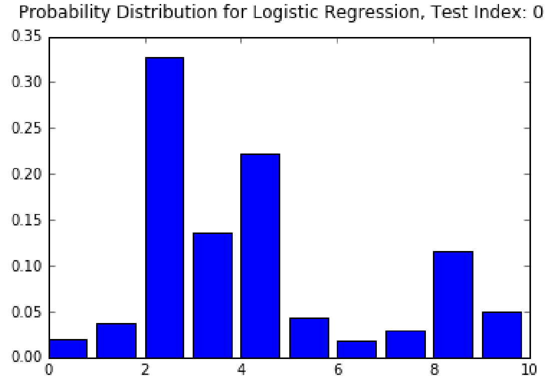


Misclassified Examples for Logistic Regression, True: 0 Predicted: 2

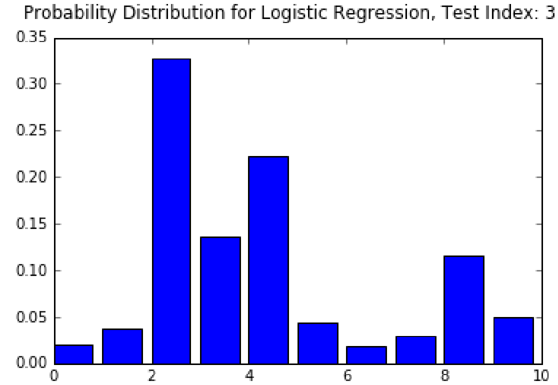


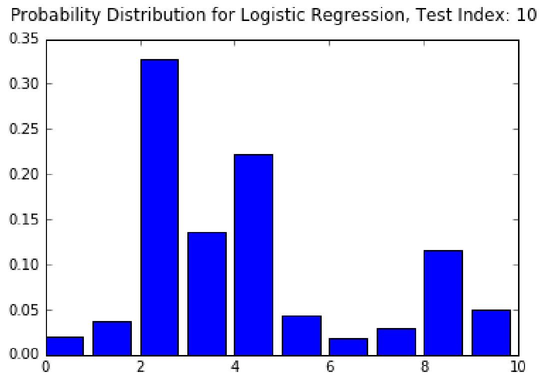
Below are the plots for the probability distribution for logistic regression. In all cases, 3 was detected the most.

Probability Distribution for Logistic Regression, Test Index: 0



Probability Distribution for Logistic Regression, Test Index: 3




Probability Distribution for Logistic Regression, Test Index: 10
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
2
4
6
8
10