EECS 6892 BMML HW#2

Name: Sung-Yen Liu
UNI: sl3763

Problem 1.

$$x_n \sim N(x_n \mid Wz_n, \sigma^2 I), \quad W \sim \left(\frac{\lambda}{2\pi}\right)^{\frac{dk}{2}} \exp\left\{-\frac{\lambda}{2}\operatorname{trace}(W^TW)\right\}, \quad z_n \sim N(z_n \mid 0, I)$$

Problem: Solve $W^* = \underset{W}{\operatorname{argmax}} \ln p(X, W)$

We start from W_0

for $t = 1, \ldots, T$

$$\ln p(X,W) = \underbrace{\int q_t(z) \ln \frac{p(X,W,Z)}{q_t(Z)} dZ}_{\mathcal{L}} + \underbrace{\int q_t(Z) \ln \frac{q_t(Z)}{p(Z \mid W, X)} dZ}_{KL(q_t(Z) \| p(Z \mid W,X))}$$

① In E-step

we assume $q_t(Z) = p(Z \mid W, X)$, which leads to $KL = 0$

By Bayes rule, and the conjugate characteristic of Gaussian distribution

$$p(z \mid W, X) \propto \underbrace{p(Z)}_{N} \cdot \underbrace{p(X \mid Z, W)}_{N}$$

$N(0, I)$ and $N(WZ_n, \sigma^2 I)$ respectively

$$\Rightarrow p(z \mid W, X) = N(z \mid \mu, \Sigma) \quad \text{where} \begin{cases} \mu_n = \Sigma W^T(\sigma^2 I)^{-1} X_n = \sigma^{-2} \Sigma_n W^T X_n \\ \Sigma_n = (I + W^T(\sigma^2 I)^{-1} W)^{-1} = (I + \sigma^{-2} W^T W)^{-1} \end{cases}$$

② In M-step

we are solving $W_t = \underset{W}{\operatorname{argmax}} \underbrace{\mathbb{E}_{q_t(Z)}\left[\ln \frac{p(X,W,Z)}{q_t(Z)}\right]}_{= \int q_t(Z) \ln \frac{p(X,W,Z)}{q_t(Z)} dZ}$, maximizing the $\mathcal{L}$

$$W_t = \underset{W}{\operatorname{argmax}} \mathbb{E}_{q_t(Z)}\left[\frac{\ln p(Z,W,X)}{q_t(Z)}\right] = \underset{W}{\operatorname{argmax}}\left(\mathbb{E}_{q_t(Z)}[\ln p(Z,W,X)] - \underbrace{\mathbb{E}_{q_t(Z)}[\ln q_t(Z)]}\right)$$

nothing to do with W
we can ignore this term

$$\ln p(Z,W,X) = \ln\left(p(X \mid W, Z) \cdot p(W) \cdot p(Z)\right)$$

$$= \sum_{n=1}^{N} \ln N(x_n \mid WZ_n, \sigma^2 I) + \ln p(W) + \underbrace{\sum_{n=1}^{N} \ln N(Z_n \mid 0, I)}$$

$$W_t = \underset{W}{\operatorname{argmax}}\left(\ln p(W) + \sum_{n=1}^{N} \mathbb{E}_{q_t(Z)}[\ln p(X_n \mid W, Z_n)]\right)$$

$$\ln p(W) = \frac{dk}{2} \ln\left(\frac{\lambda}{2\pi}\right) - \frac{\lambda}{2} \operatorname{trace}(W^T W)$$

$$\ln p(X_n \mid W, Z_n) = C - \frac{1}{2\sigma^2}(X_n - WZ_n)^T(X_n - WZ_n) = C - \frac{1}{2\sigma^2}(X_n^T - Z_n^T W^T)(X_n - WZ_n)$$

we only consider the terms related to W

$$W_t = \underset{W}{\operatorname{argmax}} \Big(\underbrace{-\frac{\lambda}{2} \operatorname{trace}(W^T W) + \sum_{n=1}^{N} \mathbb{E}_{q_t(Z)} \left[-\frac{1}{2\sigma^2} (-Z_n^T W^T X_n - X_n^T W Z_n + Z_n^T W^T W Z_n) \right]}_{\mathcal{L}} \Big)$$

To maximize $\mathcal{L}$, we set W_t s.t. $\frac{\partial \mathcal{L}}{\partial W} = 0$

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial W} = -\lambda W - \frac{1}{2\sigma^2} \sum_{n=1}^{N} \mathbb{E}_{q_t(Z)} \left[-Z_n X_n^T - X_n^T Z_n + 2 W Z_n Z_n^T \right] = 0$$

same

$$\Rightarrow \lambda W = \frac{1}{2\sigma^2} \sum_{n=1}^{N} \mathbb{E}_{q_t(Z)} \left[2 X_n^T Z_n \right] - \frac{1}{2\sigma^2} \sum_{n=1}^{N} \mathbb{E}_{q_t(Z)} \left[2 W Z_n Z_n^T \right]$$

$$\Rightarrow W \left(\lambda + \sigma^{-2} \sum_{n=1}^{N} \mathbb{E}_{q_t(Z)} [Z_n Z_n^T] \right) = \sigma^{-2} \sum_{n=1}^{N} \mathbb{E}_{q_t(Z)} \left[X_n^T Z_n \right]$$

$$\Rightarrow W = \left(\sigma^{-2} \sum_{n=1}^{N} X_n \mathbb{E}_{q_t(Z)} [Z_n^T] \right) \left(\lambda + \sigma^{-2} \sum_{n=1}^{N} \mathbb{E}_{q_t(Z)} [Z_n Z_n^T] \right)^{-1}$$

Pseudo code:

1. Initialize W_0 to a vector of all zeros
2. For iteration $t = 1, \dots, T$

Since $q_t(Z) = \mathcal{N}(Z_n \mid \mu_n', \Sigma_n')$

E-step:

we update $q_t(Z)$ by
$$\begin{cases} \mu'_{n(t)} = \sigma^{-2} \Sigma'_{n(t)} W_{t-1}^T X_n & \cdots\cdots (\mathbb{E}_{q_t(n)}[Z_n^T]) \\ \Sigma'_{n(t)} = (I + \sigma^{-2} W_{t-1}^T W_{t-1})^{-1} & \cdots\cdots (\mathbb{E}_{q_t(n)}[Z_n Z_n^T]) \end{cases}$$

M-Step:

$$W_t = \left(\sigma^{-2} \sum_{n=1}^{N} X_n \mathbb{E}_{q_t(Z)} [Z_n^T] \right) \left(\lambda + \sigma^{-2} \sum_{n=1}^{N} \mathbb{E}_{q_t(Z)} [Z_n Z_n^T] \right)^{-1}$$ for next iteration

Calculate
$$\ln p(X, W) = \ln p(W) + \sum_{i=1}^{N} \ln p(X_i)$$
$$= \frac{dK}{2} \ln\left(\frac{\lambda}{2\pi}\right) - \frac{\lambda}{2} \operatorname{trace}(W^T W) + \sum_{n=1}^{N} \left[-\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (X_n - W Z_n)^T (X_n - W Z_n) \right.$$

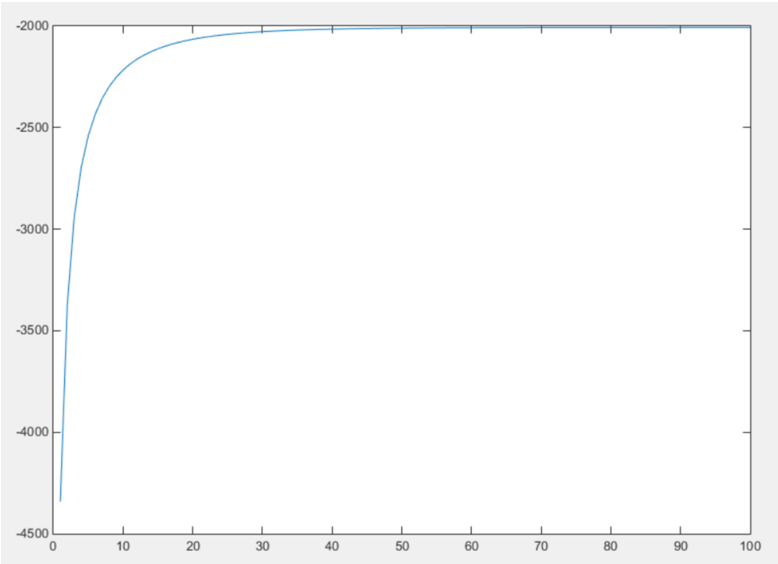
until the criterion is satisfied.

Problem 2

a) Attached please find the code of implementation.

Accuracy after 100 iterations: 93.57% (1863/1991)

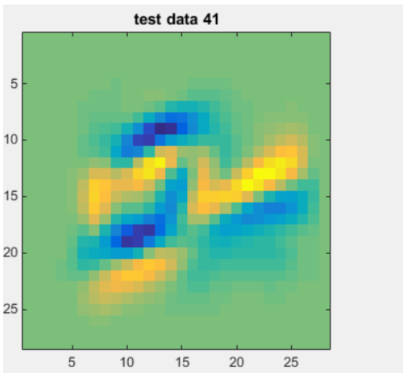
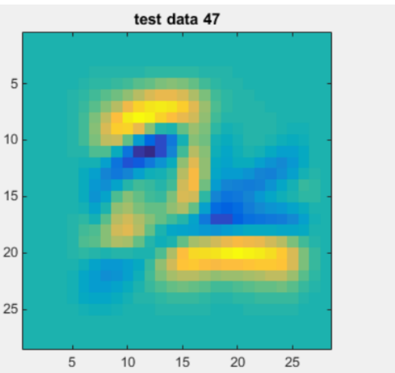
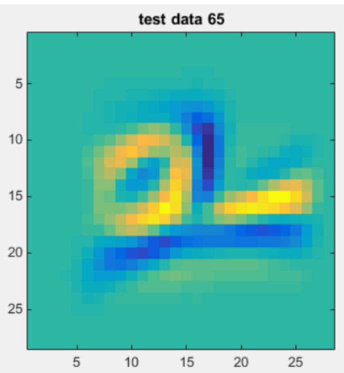
b) Figure:



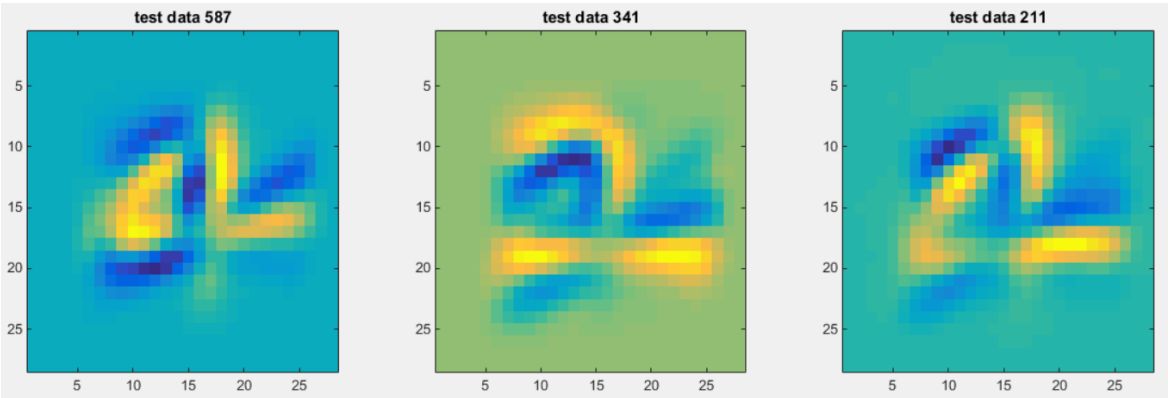
c) Confusion matrix

	Classified as 0 (4)	Classified as 1 (9)
Label 0 (4)	931	51
Label 1 (9)	77	932

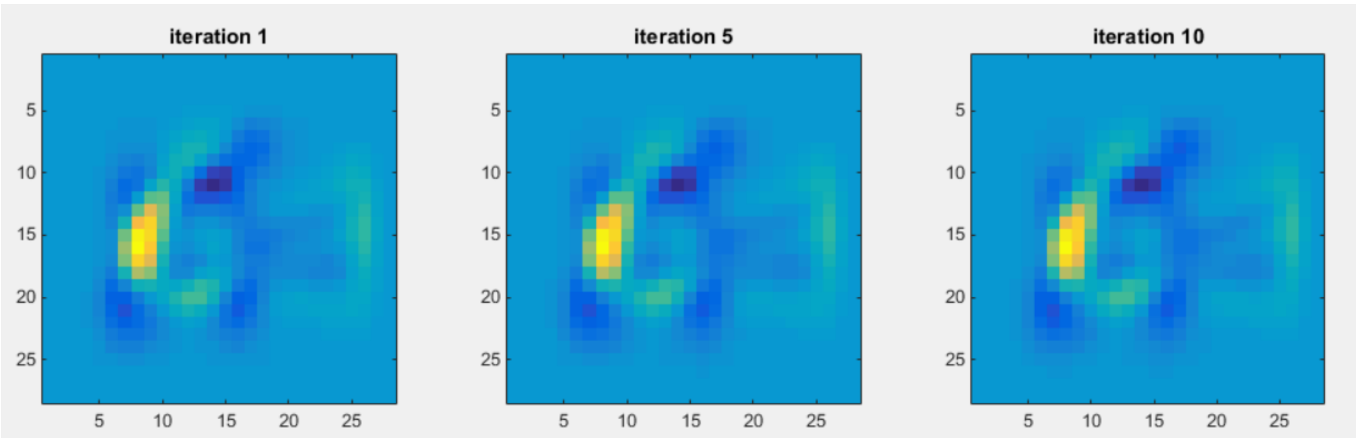
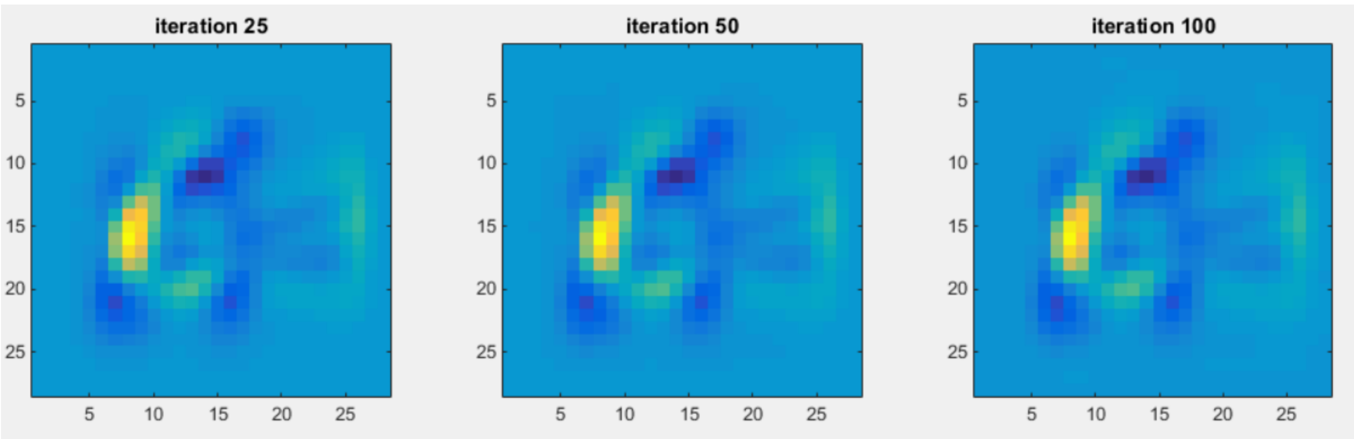
d) Three misclassified digits: (labeled as 4 but misclassified as 9)

Test data index	41	47	65
Images			
Predictive prob.	0.6775	0.6994	0.9062

e) Three most ambiguous digits:

Test data index	587	341	211
Images			
Predictive prob.	0.4995	0.5038	0.5038

f) Vector w_t for different t

Iteration 1	Iteration 5	Iteration 10
		
Iteration 25	Iteration 50	Iteration 100
		

Through the figure in b), it can be shown that the vector w_t is converging after 20~30 iterations, thus the last three figures above are almost the same (both visually and numerically). Compare w_1 to w_{100}, although they are visually the same, there are slight difference on vector value, and the color depth are also distinct. In conclusion, using EM algorithm for optimizing gives an acceptable approximate solution in a few iterations.