Name: Jingyi Yuan
UNI: jy2736
Class: EECS E6892

Homework 02

Problem 1

Step 1. Calculate the posterior:

$$q_t(z) = p(z|x, w_{t-1}) = \frac{p(x, z, w_{t-1})}{p(x, w_{t-1})}$$

$$= \prod_{n=1}^{N} \frac{p(x_n|z_n, w_{t-1})\, p(z_n)\, p(w_{t-1})}{p(w_{t-1}) \int p(x_n|z_n, w)\, p(z_n)\, dz_n}$$

$$\propto p(x_n|z_n, w_{t-1}) \cdot p(z_n)$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x_n - w^T z_n)^2\right\} \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} z_n z_n^T\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\left[(w_{t-1}^T w_{t-1} + \sigma^2 I) z_n z_n^T - 2w^T x_n z_n + x_n^T x_n\right]\right\}$$

$$\sim N\left((w_{t-1}^T w_{t-1} + \sigma^2 I) w^T x_n,\ (w_{t-1}^T w_{t-1} + \sigma^2 I)^{-1} \sigma^2\right\}$$

Step 2: Calculate $\mathcal{L}(w)$.

$$p(x, z, w) = \int q_t(z) \ln \frac{\ln p(x, z, w)}{q_t(z)}\, dz$$

$$= E_{q_t(z)}[\ln p(x, z, w)] - E_{q_t(z)} \ln q_t(z)$$

$$\therefore \mathcal{L}(w) = E_{q_t(z)}[\ln p(x, z, w)] + \text{constant}$$

$$= E_{q_t(z)}\left[\ln\left(p(w) \cdot \prod_{n=1}^{N} p(z_n)\, p(x_n|z_n, w)\right)\right] + \text{constant}$$

$$= E_{q_t(z_n)}[\ln p(w)] + \sum_{n=1}^{N} E_{q_t(z_n)}\left[\ln p(x_n|z_n, w)\right] + \text{constant}$$

$$= \ln p(w) + \sum_{n=1}^{N} E_{q_t(z_n)} \left\{\frac{1}{2\sigma^2}\left[(w^T w + \sigma^2 I) z_n z_n^T - 2w^T x_n z_n + x_n^T x_n\right]\right\}$$

Step 3: $\nabla \mathcal{L}(w) = 0$.

$$\nabla \mathcal{L}(w) = -\lambda w - \sum_{n=1}^{N} \frac{1}{2\sigma^2} E_{q_t(z_n)}\left[2w \cdot z_n z_n^T - 2 x_n z_n\right]$$

$$= -\lambda w - \frac{1}{\sigma^2} \sum_{n=1}^{N}\left[w E_{q_t(z_n)}(z_n z_n^T) - x_n E_{q_t(z_n)}(z_n)\right] = 0$$

$$\therefore \frac{1}{\sigma^2}\sum_{n=1}^{N} x_n E_{q_t(z_n)}(z_n) = \lambda w + w \sum_{n=1}^{N} E_{q_t(z_n)}[z_n z_n^T] \cdot \frac{1}{\sigma^2}$$

$$\therefore w = \left(\lambda I + \frac{1}{\sigma^2}\sum_{n=1}^{N} E_{q_t(z_n)}[z_n z_n^T]\right)^{-1} \cdot \frac{1}{\sigma^2} \cdot \left[\sum_{n=1}^{N} x_n E_{q_t(z_n)}(z_n)\right]$$

$$E_{q_t(z_n)}(z_n) = (w_{t-1}^T w_{t-1} + \sigma^2 I)^{-1} w_{t-1}^T x_n.$$

$$E_{q_t(z_n)}(z_n z_n^T) = D(z_n) + E_{q_t(z_n)}(z_n)\left[E_{q_t(z_n)}(z_n)\right]^T = (w_{t-1}^T w_{t-1} + \sigma^2 I)^{-1}\sigma^2 + (w_{t-1}^T w_{t-1} + \sigma^2 I)^{-1} w_{t-1}^T x_n \cdot \left[(w_{t-1}^T w_{t-1} + \sigma^2 I)^{-1} w_{t-1}^T x_n\right]^T.$$

$\therefore$ Summary (pseudo-code):

1. $w_0 = \text{zeros}(d, 1)$
2. for $t = 1:T$ //E-Step
3. $E_{q_t(z_n)}(z_n) = (w_{t-1}^T w_{t-1} + \sigma^2 I)^{-1} w_{t-1}^T x_n$
4. $E_{q_t(z_n)}(z_n z_n^T) = (w_{t-1}^T w_{t-1} + \sigma^2 I)^{-1}\sigma^2 + (w_{t-1}^T w_{t-1} + \sigma^2 I)^{-1} w_{t-1}^T x_n \left[(w_{t-1}^T w_{t-1} + \sigma^2 I)^{-1} w_{t-1}^T x_n\right]^T$
5. end
6. $w_t = \left(\lambda I + \frac{1}{\sigma^2}\sum_{n=1}^{N} E_{q_t(z_n)}[z_n z_n^T]\right)^{-1} \frac{1}{\sigma^2}\left[\sum_{n=1}^{N} x_n E_{q_t(z_n)}(z_n)\right]$ //M-Step

Problem 2

a)

Run the algorithm on the data set provided for T = 100 iterations (the following is only part of the code)

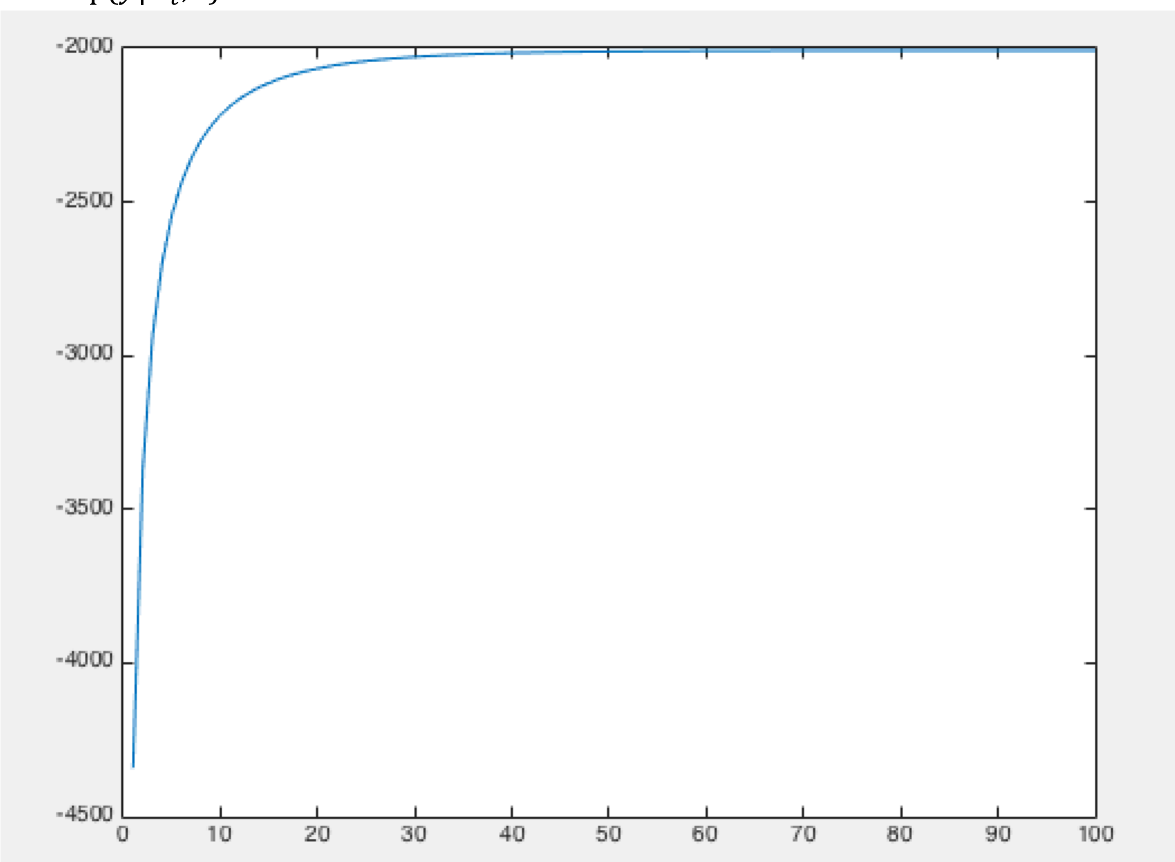
```
for t = 1:T
    s0 = -Sample0.' * w ./ sigma;%5842*1
    Eqt_yequals_0 = Sample0.' * w + sigma .* (-normpdf(s0)./normcdf(s0));%5842*1
    s1 = -Sample1.' * w ./ sigma;%5949*1
    Eqt_yequals_1 = Sample1.' * w + sigma .* (normpdf(s1))./(One-normcdf(s1));%5949*1

    Eqt_fi = [Eqt_yequals_0' Eqt_yequals_1'];
    xiEqt = Xtrain * (Eqt_fi.');
    w = pinv(lambda .* I + x ./ (sigma^2)) * (xiEqt./(sigma^2));%1*1
    W(:,t) = w;
end
```

After this, we get w:

```
w =

   -0.3316
    0.7914
    0.4201
   -1.0141
    1.4106
    1.3368
    0.2828
   -0.6027
    1.4279
   -0.2909
   -0.5306
    0.5726
   -0.3308
   -0.1713
    0.4084
```

b)

Plot $\ln p(\vec{y}|w_t,\mathbf{X})$ as a function of t:



c)

The confusion matrix is shown below. As it is suggested in the comments:

confsMatrix(1,1) = 930 is the total number of 4's classified as 4's.

confsMatrix(1,2) = 52 is the total number of 4's classified as 9's.

confsMatrix(2,1) = 77 is the total number of 9's classified as 4's.

confsMatrix(2,2) = 932 is the total number of 9's classified as 9's.

```
confsMatrix =

   930    52
    77   932
```

The accuracy of the classifier is:

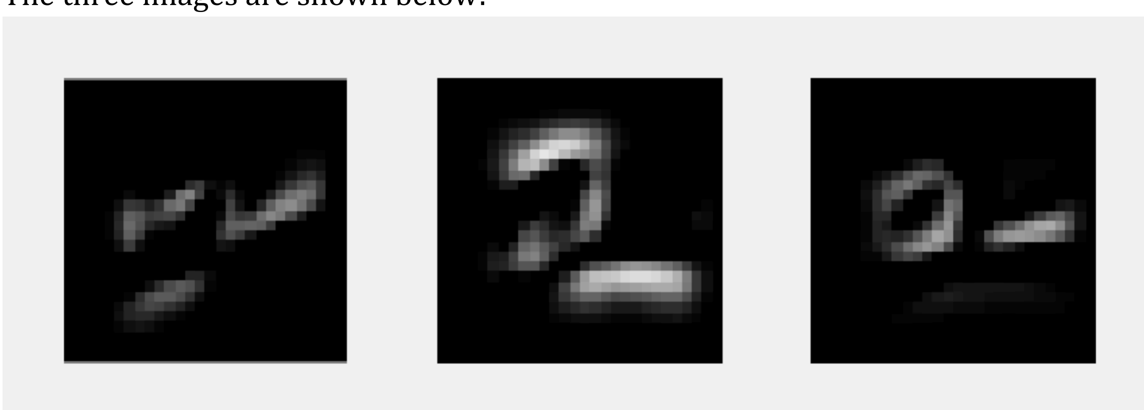
```
correctness =

    0.9352
```

d)

The predictive probabilities of the three misclassified digits are shown below:

```
The predictive probility equals 0 of the 1 misclassified is = 3.228109e-01
The predictive probility equals 1 of the 1 misclassified is = 6.771891e-01
The predictive probility equals 0 of the 2 misclassified is = 3.017857e-01
The predictive probility equals 1 of the 2 misclassified is = 6.982143e-01
The predictive probility equals 0 of the 3 misclassified is = 9.421276e-02
The predictive probility equals 1 of the 3 misclassified is = 9.057872e-01
```

The three images are shown below:

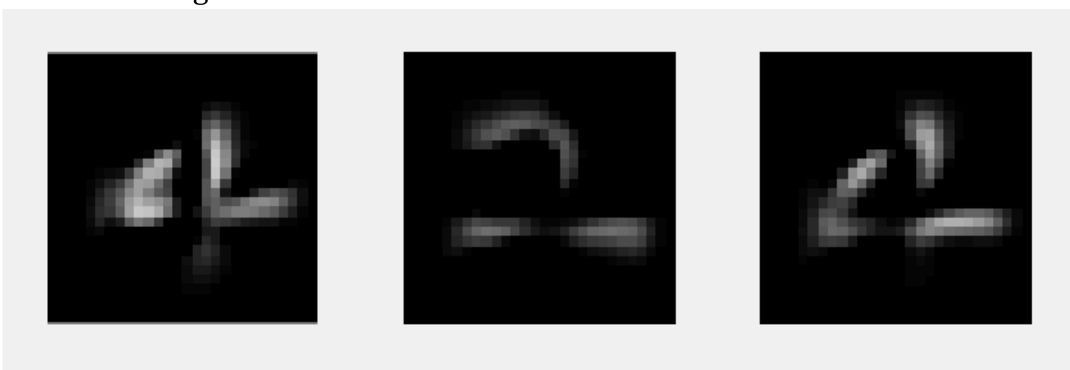


e)

The predictive probabilities of the three most ambiguous predictions are shown below:

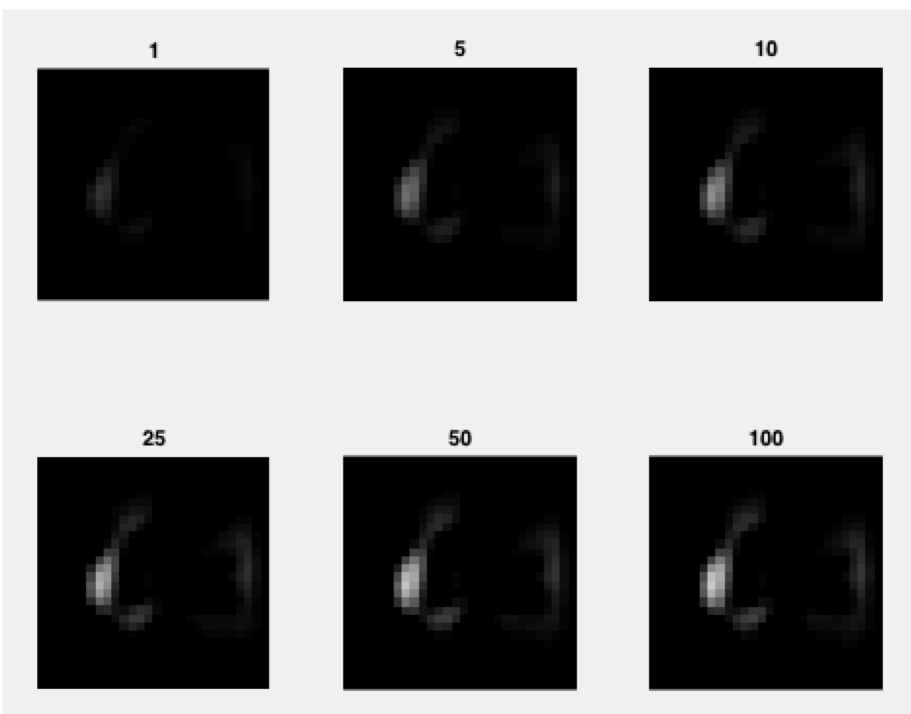
```
The predictive probility equals 0 of the 1 ambiguous number is = 4.998317e-01
The predictive probility equals 1 of the 1 ambiguous number is = 5.001683e-01
The predictive probility equals 0 of the 2 ambiguous number is = 4.969085e-01
The predictive probility equals 1 of the 2 ambiguous number is = 5.030915e-01
The predictive probility equals 0 of the 3 ambiguous number is = 4.959111e-01
The predictive probility equals 1 of the 3 ambiguous number is = 5.040889e-01
```

The three images are shown below:



f)

Treat the vector w_t as if it were a digit and reconstruct it as an image for t = 1, 5, 10, 25, 50, 100. These images are shown below:



It has a tendency to be clear and the bright parts become more obvious.