Homework 2
LM2221

Problem 1)

$z_n \sim N(0, I) \rightarrow \mathbb{R}^k$

$W \sim N(0, \Upsilon^{-1}) \rightarrow \mathbb{R}^{d \times k}$

$x_n \sim N(Wz_n, \sigma^2 I) \rightarrow \mathbb{R}^d$

First we will solve the general solution. For the E-step, we are looking for the posterior distribution:

$p(z \mid x, W)$, where Υ, σ^2 are fixed & W is our parameter.

$$p(z \mid x, W) = \frac{p(x \mid W, z)\, p(z)}{\int p(x \mid W, z)\, p(z)}$$

$$= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x - Wz)^T (x - Wz)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^T z}}{\int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x - Wz)^T (x - Wz)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^T z} \, dz}$$

In this case we know $z \sim N(\mu_z, \Sigma_z)$. Given a resource $x \mid z \sim N(Wz + \mu_x, \Sigma_x)$

$z \mid x \sim N(R(x - W\mu_z - \mu_x), \Sigma_z - RW\Sigma_z^T$

$R = \Sigma_z W^T (W \Sigma_z W^T + \Sigma_x)^{-1}$

$\mu_z = 0, \Sigma_z = I$

$\mu_x = 0, \Sigma_x = \sigma^2 I$

$R = W^T (WW^T + \sigma^2 I)^{-1}$

$\mu_{z \mid x, W} = Rx$

$\Sigma_{z \mid x, W} = I - RW$

→

Now solve the M-step:

$$\underbrace{\ln p(X,Z,W)}_{Q} = \ln\left[\underbrace{p(X|Z,W)}_{\text{solved in last part}}\underbrace{p(Z)\,p(W)}_{\text{joint}}\right]$$

$$\ln\left[p(X|Z,W) + \cancel{\ln p(Z)} + \ln p(W)\right]$$

ignore does not depend on W

$$= -\frac{1}{2\sigma^2}\underbrace{\left(X^TX - X^TWZ - Z^TW^TX + Z^TW^TWZ\right)}_{(X-WZ)^T(X-WZ)} - \gamma/2\,\mathrm{tr}(W^TW)$$

$$\frac{\partial}{\partial W} = -\frac{1}{\sigma^2}\left(XZ^T + WZZ^T\right) - \gamma W$$

plugging back in to full equation

$$\frac{\partial Q}{\partial W} = \int p(Z|X,W_{old})\left[\frac{\partial}{\partial W}\ln p(X,Z,W)\right]dZ = 0$$

$$= \int p(Z|X,W_{old}\left[-\frac{1}{\sigma^2}XZ^T + \frac{1}{\sigma^2}WZZ^T + \gamma W\right]dZ = 0$$

this looks like $\int p(x)f(x)dx = E[f(x)]$

$$= -E(XZ^T) + WE(ZZ^T) + \sigma^2\gamma W\underbrace{\int p(Z|X,W^{old})dZ}_{1}$$

$$W(E(ZZ^T) + \sigma^2\gamma I) = E(XZ^T)$$

$$W = E(XZ^T)(E(ZZ^T) + \sigma^2\gamma I)^{-1}$$

$$= X\mu_{Z|X,W}\left(\Sigma_{Z|X,W} + \mu_{Z|X,W}\mu_{Z|X,W}^T + \sigma^2\gamma I\right)^{-1}$$

→

Now converting it from the general case, where $\ln p(X, Z, W) = \ln \left[\prod_{n=1}^{N} p(x_n, z_n, W) \right]$

$$W = \left[\sum_{i=1}^{n} x_i M_{z_i|x_i,w} \right] \left[\sum_{i=1}^{n} \left(\Sigma_{z_i|x_i,w} + M_{z_i|x_i,w} M_{z_i|x_i,w}^T \right) + \sigma^2 \chi \right]^{-1}$$

Sudo Code Summary:

From section 9.4 in Bishop the EM algorithm →

1. Initialize θ_{old}, where we have $p(X, Z|\theta)$ and goal is to maximize $p(X|\theta)$
2. E-step = Calculate $E[Z]$ where $p(Z|X, \theta^{old})$
3. M-step = Calculate θ_{new} through

$$\theta^{new} = \text{argmax}\, G(\theta, \theta^{old})$$

$$G(\theta, \theta_{old}) =$$

$$\sum_{Z} p(Z|X, \theta^{old}) \ln p(X, Z|\theta)$$

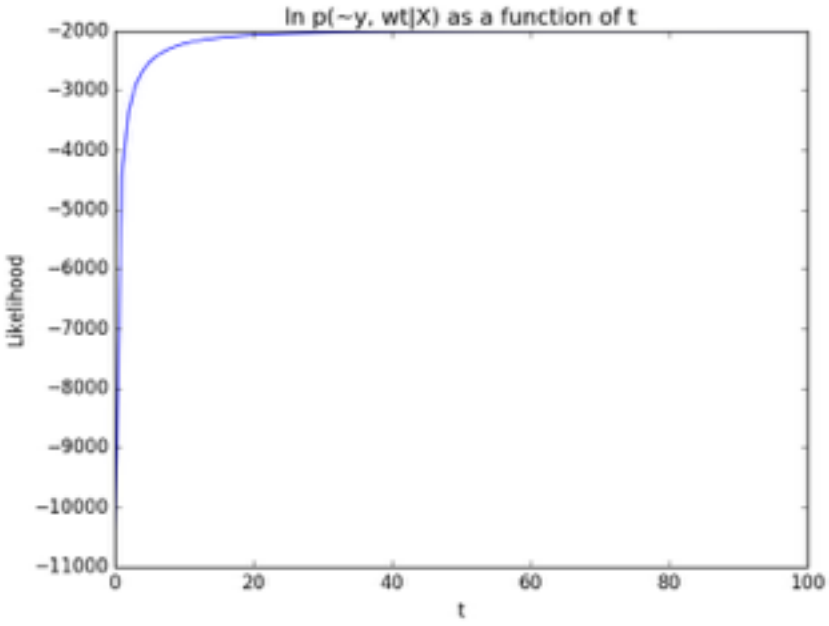
4. Do this over T interactions once the update value is smaller than ϵ → defined in problem stop.

$$\ln P_T - \ln P_{T-1} < \epsilon$$

□

Problem 2:

a) See python code
b) Plot likelihood



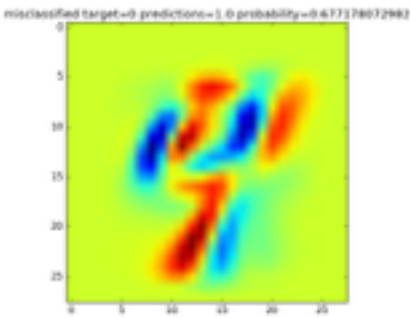
c) Confusion Matrix

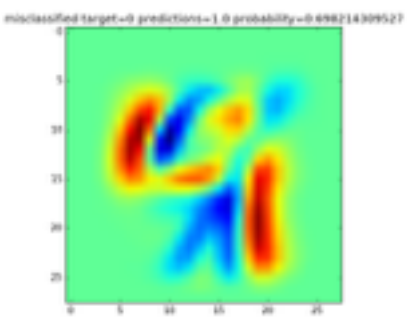
	Predict Class 4	Predict Class 9
True Class 4	930	52
True Class 9	77	932

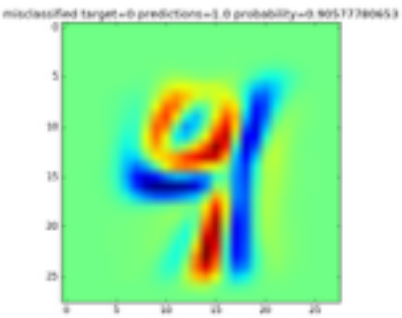
Accuracy: 93%

d) Pick three misclassified digits:

1) target=0, prediction =1 prob=.677178072982
2) target=0, prediction =1 prob=.698214309527
3) target=0, prediction =1 prob=.90577780653

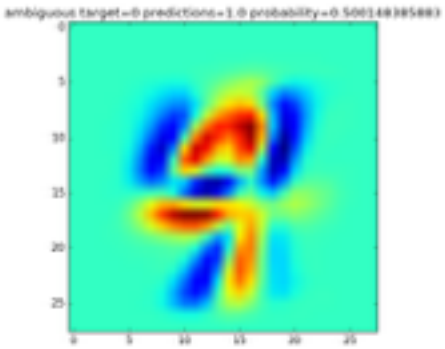


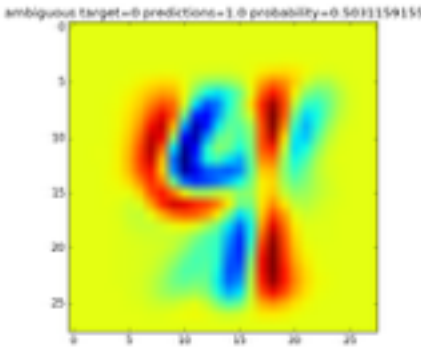


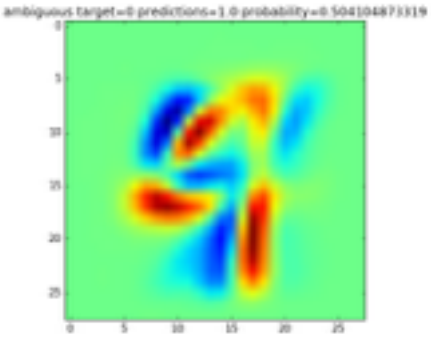


e) Pick three ambiguous digits:

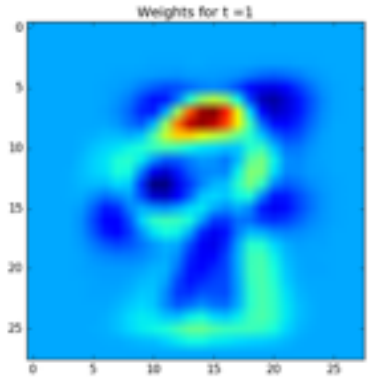
1) target=0, pred=1 prob=.500148385883
2) target=0, pred=1 prob=.5031159155
3) target=0, pred=1 prob=.504104873319

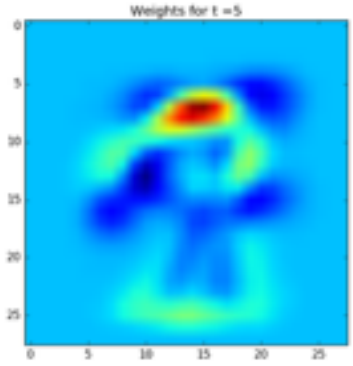


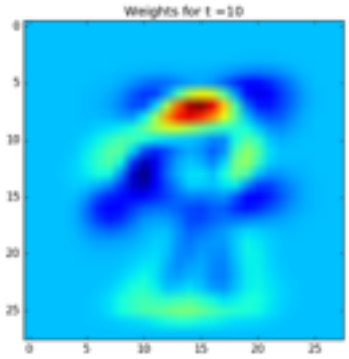


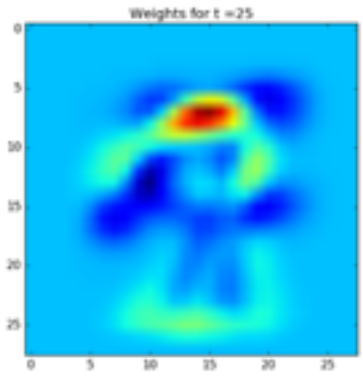


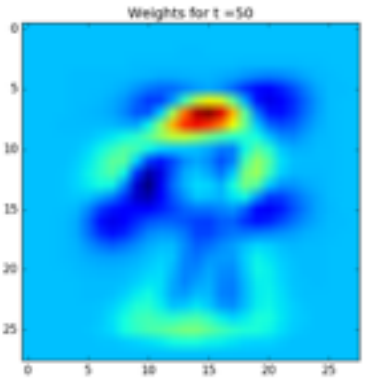
f) Plot the weights:

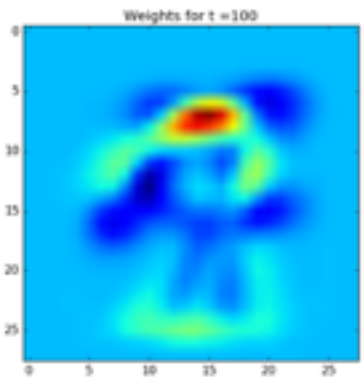












You can see the plot
change from t=1 to t=5 and then as it converges it does not change much. It is learning the weights.