

Homework 2

LM2221

①

Problem 1)

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

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

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

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

$P(z | X, W)$, where τ, σ^2 are fixed & W is our parameter.

$$p(z | X, W) = \frac{p(X | W, z) p(z)}{\int p(X | W, z) p(z)}$$

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

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

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

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

$$\mu_z = 0, \Sigma_z = I$$

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

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

$$\mu_{z|X, W} = RX$$

$$\Sigma_{z|X, W} = I - RW$$

→

Now solve the M-step:

(2)

$$\underbrace{\ln p(x, z, w)}_Q = \ln \left[\underbrace{p(x|z, w)}_{\text{solved in last part}} \underbrace{p(z)}_{\text{joint}} \underbrace{p(w)}_{\text{joint}} \right]$$

$$\begin{aligned} & \ln [p(x|z, w) + \ln p(z) + \ln p(w)] \\ & \quad \downarrow \quad \quad \quad \text{ignore does not depend on } w \\ & = -\frac{1}{2\sigma^2} \underbrace{(x^T x - x^T w z - z^T w^T x + z^T w^T w z)}_{(x-wz)^T(x-wz)} - \frac{\gamma}{2} + \frac{1}{2} \ln |w^T w| \end{aligned}$$

$$\frac{\partial}{\partial w} = -\frac{1}{\sigma^2} (xz^T + wz z^T) - \gamma w$$

plugging back in to full equation

$$\begin{aligned} \frac{\partial Q}{\partial w} &= \int p(z|x, w_{old}) \left[\frac{\partial}{\partial w} \ln p(x, z, w) \right] dz = 0 \\ &= \int p(z|x, w_{old}) \left[-\frac{1}{\sigma^2} xz^T + \frac{1}{\sigma^2} w z z^T + \gamma w \right] dz = 0 \end{aligned}$$

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

$$\begin{aligned} &= -E(xz^T) + w E(z z^T) + \sigma^2 \gamma w \underbrace{\int p(z|x, w_{old}) dz}_1 \\ &w (E(z z^T) + \sigma^2 \gamma I) = E(xz^T) \end{aligned}$$

$$w = E(xz^T) (E(z z^T) + \sigma^2 \gamma I)^{-1}$$

$$= X \mathcal{M}_{z|x, w} \left(\sum_{i=1}^N z_i x_i^T + \mathcal{M}_{z|x, w} \mathcal{M}_{z|x, w}^T + \sigma^2 \gamma I \right)^{-1}$$

→

Now converting it from the general case, (3)
 when $\ln p(x, z, w) = \ln \left[\prod_{i=1}^N p(x_i, z_i, w) \right]$

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

Sudo Code Summary:

from section 9.4 in Bishop the EM algorithm \rightarrow

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

$$\theta^{new} = \underset{\theta}{\operatorname{argmax}} Q(\theta, \theta^{old})$$

$$Q(\theta, \theta^{old}) =$$

$$\sum_z p(z | x, \theta^{old}) \ln p(x, z | \theta)$$

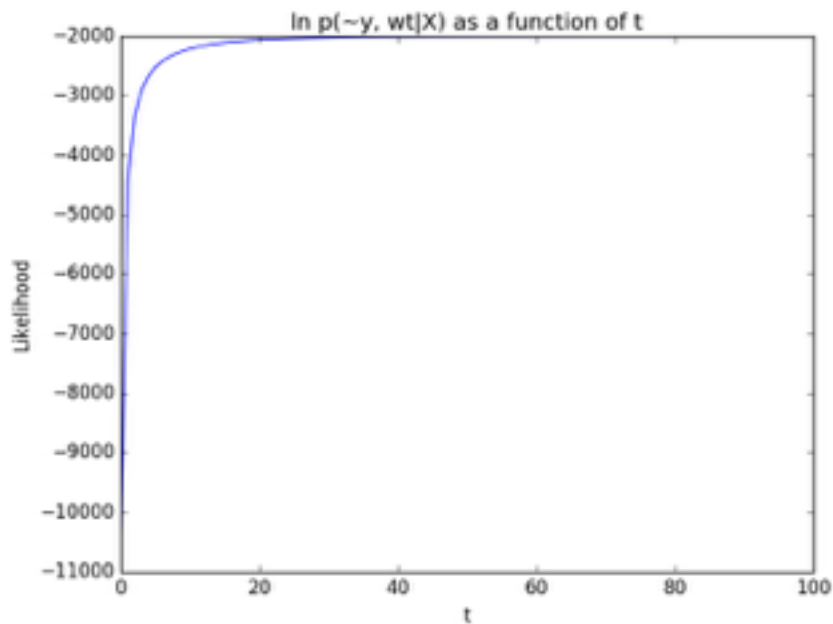
4. Do this over T iterations once the update value is smaller than $\epsilon \rightarrow$ defined in problem stop.

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

□

Problem 2:

- a) See python code
- b) Plot likelihood



c) Confusion Matrix

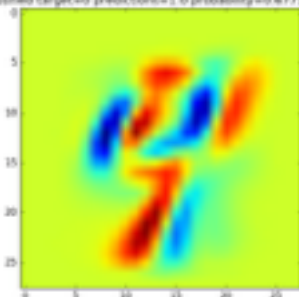
	Predict Class 4	Predict Class 9
True Class 4	930	52
True Class 9	77	932

Accuracy: 93%

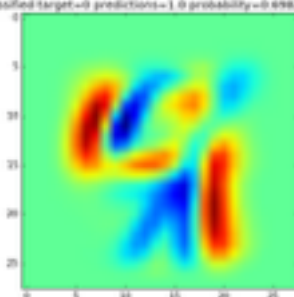
d) Pick three misclassified digits:

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

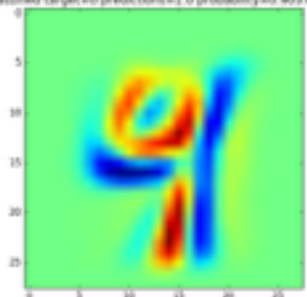
misclassified target=0 predictions=1 @ probability=0.677178072982



misclassified target=0 predictions=1 @ probability=0.698214309527



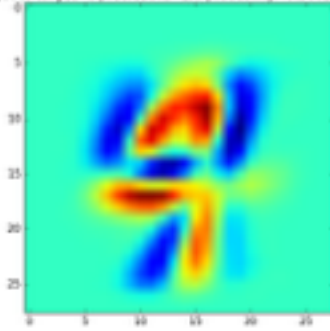
misclassified target=0 predictions=1 @ probability=0.90577780653



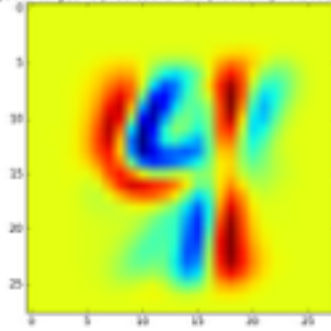
e) Pick three ambiguous digits:

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

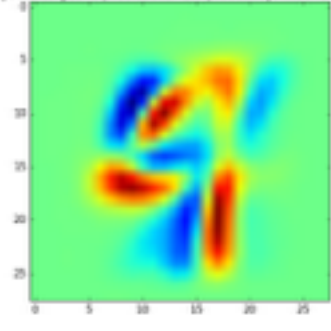
ambiguous target=0 predictions=1 @ probability=.500148385883



ambiguous target=0 predictions=1 @ probability=.5031159155

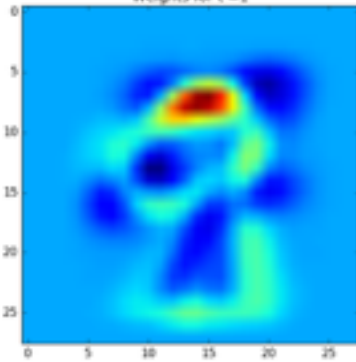


ambiguous target=0 predictions=1 @ probability=.504104873319

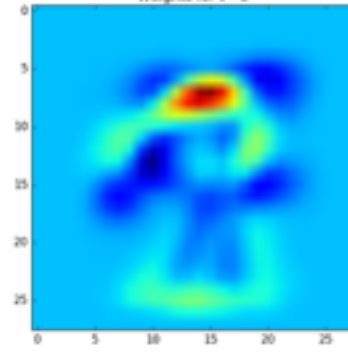


f) Plot the weights:

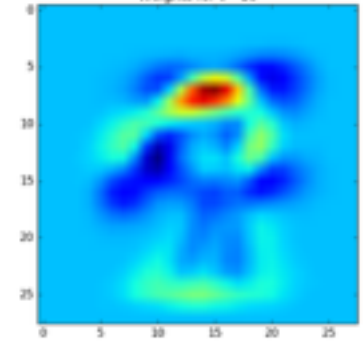
Weights for t=1



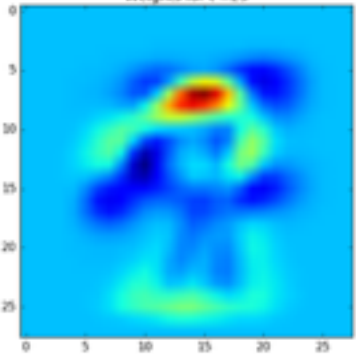
Weights for t=5



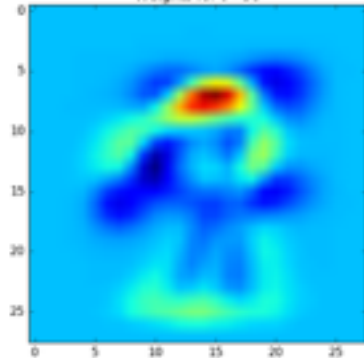
Weights for t=10



Weights for t=25



Weights for t=50



Weights for t=100

