Homework 2 BMML E6720 Ve)

1) $(x_1 \dots x_N)$ $x \in \mathbb{R}^d$ $x_n \sim N(Wz_n, \sigma^2 I)$ $W \in \mathbb{R}^{d \times k}$

$z_n \sim N(0, I)$

$$p(W) = (2\pi)^{-\frac{k}{2}} \exp\left[-\frac{\lambda}{2} \mathrm{trace}(W^T W)\right],$$

$$p(X_N \mid W, z_n) = (2\pi\sigma^2 I)^{-d/2} \exp\left[\frac{-1}{2\sigma^2 I}(x_n - Wz_n)^T (x_n - Wz_n)\right]$$

$$p(Z_N) = (2\pi)^{-\frac{k}{2}} \exp\left[-\frac{1}{2} z_n^T z_N\right]$$

goal $\rightarrow$ $W' = \underset{W}{\arg\max} \ln p(x_1 \dots x_n, W)$

now, from class

$$\ln p(x_1 \dots x_n, W) \geq \int q(z_n) \ln \frac{p(x_1 \dots x_n, W, z_n)}{q(z_n)} \, dz_n$$

$$q(z) = p(z \mid W, x) \propto \prod_{i=1}^{n} p(x_n \mid W, z_n) p(z_n)$$

$$q(z) \propto \prod_{i=1}^{n} (2\pi\sigma^2 I)^{-\frac{1}{2}} \exp\left[\frac{-1}{2\sigma^2 I}(x_n - Wz_n)^T (x_n - Wz_n)\right] (2\pi)^{-\frac{k}{2}} \exp\left[-\frac{1}{2} z_N^T z_N\right]$$

constants drop out

$$\propto \exp\left[-\frac{1}{2} \sum_{n=1}^{N} \left[\frac{1}{\sigma^2} z_N^T W^T W z_n - \frac{2}{\sigma^2} x_n^T W z_n + z_n^T z_n\right]\right]$$

$$\propto \exp\left[\frac{1}{2\sigma^2} \sum_{n=1}^{N} \left[-2 x_n^T W z_n + z_n^T (\sigma^2 I + W^T W) z_n\right]\right]$$

$$\mu^* = \left(\frac{\sigma^2 I}{\sigma^2 I + W^T W}\right) W^T x_{all}$$

$\sigma^2 z_n^T z_n + z_n^T W^T W z_n$

$$\Sigma^* = \left(\frac{\sigma^2 I}{\sigma^2 I + W^T W}\right) \quad \dots \text{ I think that's it.}$$

$z_n^T (\sigma^2 I + W^T W) z_n$

This is a $N(\mu^*, \Sigma^*)$

expectation of joint likelihood

$\mathcal{L}(W) = E_q[\ln p(z, x, W)]$

YW

$p(z_n, x_1 \cdots x_n, W) \propto p(W) \prod_{i=1}^{N} p(x_i | z_i W) p(z_i)$

$\ln p(z_n, x_n, W) \propto \ln p(w) + \sum_{n=1}^{N} \ln p(x_n | z_n, W)$ + some constant

$\propto E_q \ln p(w) + E_q \sum_{i=1}^{N} \left[\ln \left[2\pi\sigma^2 I \right]^{-\frac{1}{2}} \exp \left[\frac{1}{2\sigma^2 I} \sum_{i=1}^{N} (x_i - Wz_i)^T (x_i - Wz_i) \right] \right]$

$\propto \ln p(w) + E_q \left[\sum_{i=1}^{N} \ln (2\pi\sigma^2 I) + \frac{1}{2\sigma^2 I} \sum_{i=1}^{N} (x_i - Wz_i)^T (x_i - Wz_i) \right]$

$\propto \ln p(w) + E_q \left[\frac{1}{2\sigma^2 I} \sum_{i=1}^{N} (x_i - Wz_i)^T (x_i - Wz_i) \right)$

$\propto \ln p(w) + \frac{1}{2\sigma^2 I} \sum_{i=1}^{N} E_q \left[x_i^T x_i - 2 x_i^T W z_i + W^T z_i^T z_i W \right]$

$\propto \ln p(w) + \frac{1}{2\sigma^2 I} \sum_{i=1}^{N} \left[E_q(x_i^T x_i) + E_q[z_n^T W^T W z_n] - 2 E_q[x_i^T W z_i] \right]$

$\propto \ln p(w) + \frac{1}{2\sigma^2 I} \sum_{i=1}^{N} \underbrace{\cancel{x_i^T x_i}}_{\text{don't care}} - 2 x_i^T W \mu^* + E_q[z_n^T W^T W z_n]$

$= E[(z_i W)^T (z_i W)] + \text{trace}(W^T W \operatorname{Cov}(z_i))$

$\propto \frac{-\lambda}{2} \text{trace}(W^T W) + \frac{1}{2\sigma^2} \sum (\mu^* W)^T (\mu^* W) + \text{trace}(W^T W \Sigma^*) - 2 x_i^T W \mu^*$

M Step

$\nabla \mathcal{L}_t(W) = -\lambda W + \frac{1}{2\sigma^2} \sum \left[W (\mu^{*T} \mu^* + \Sigma^*) \right] - \frac{2}{2\sigma^2} \sum x_i^T \mu^*$

$= 0 = -\lambda W + \frac{1}{2\sigma^2} \sum W (\mu^{*T} \mu^* + \Sigma^*) - \frac{1}{\sigma^2} \sum x_i^T \mu^*$

$-\lambda W + \frac{W}{2\sigma^2} \cancel{\sum} (\mu^{*T} \mu^* + \Sigma^*) = \frac{1}{\sigma^2} \sum x_i^T \mu^*$

$$\boxed{W_t = \frac{\sum_{i=1}^{N} x_i^T \mu^*}{-\lambda I \sigma^2 + \frac{1}{2} (\mu^{*2} + \Sigma^*)}}$$

EM algorithm

1) Initialize W with dimensions

2) for iteration t,

a) calculate $E_{q_t}[z]$, which is $\frac{W^T x_n}{(\sigma^2 I + W^T W)}$ E step

b) calculate $W_t = \frac{\sum_{i=1}^{N} x_i \mu^*}{-\lambda\sigma^2 + \frac{1}{2}(\mu^{*2} + \Sigma^*)}$

3) once you have this,

calculate $\ln p_t(x_1 \ldots x_N, W)$ from $\ln p(x_1, \ldots x_n, W, z_n)$ until converges → so difference is sufficiently small

Scanned by CamScanner

Problem 2)

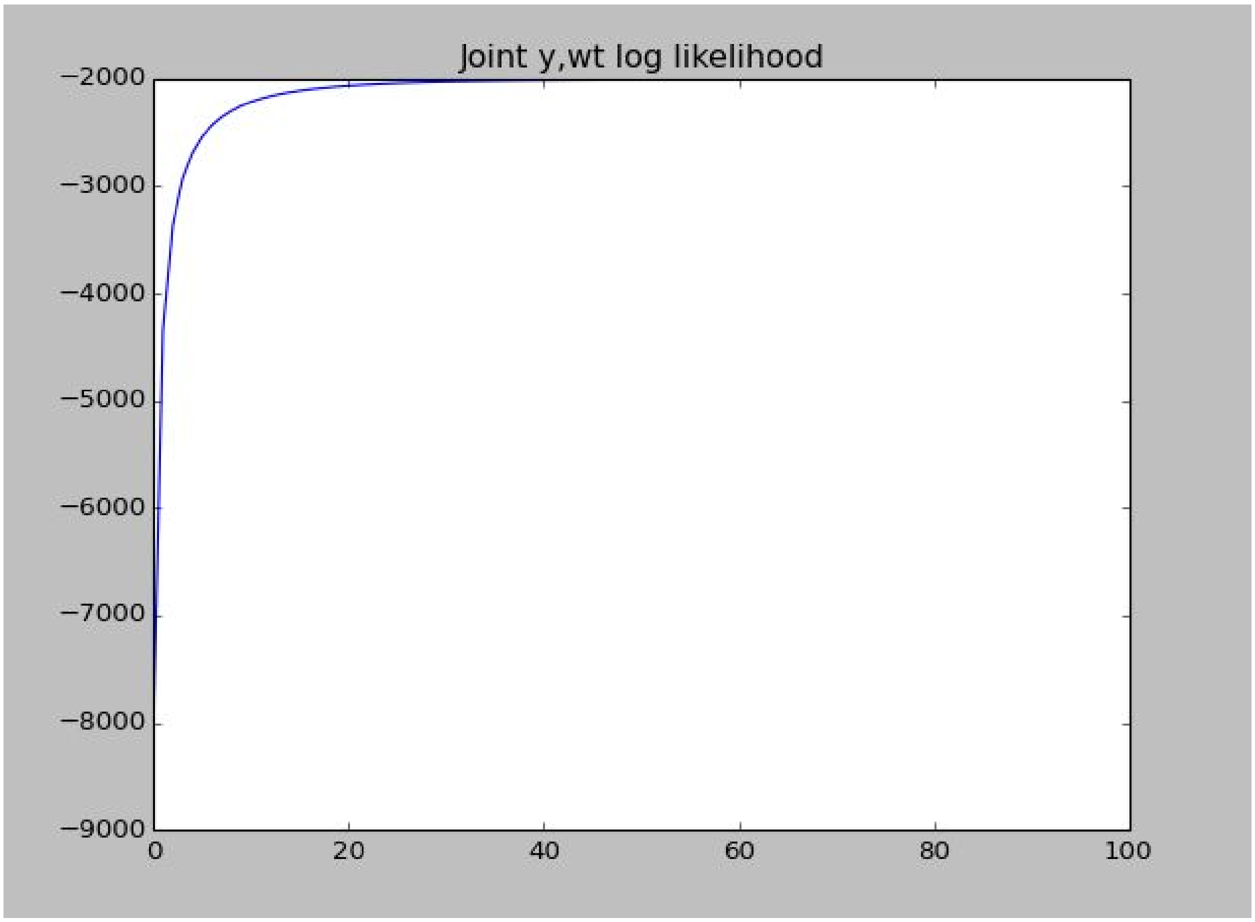
a)# Estep

p[y_1] = xw[y_1] + (sigma * norm.pdf(-xw_sigma[y_1]) / (1.0-norm.cdf(- xw_sigma[y_1])))

p[y_0] = xw[y_0] + (sigma * -norm.pdf(-xw_sigma[y_0]) / (norm.cdf(- xw_sigma[y_0])))

Main loop for E step

b)

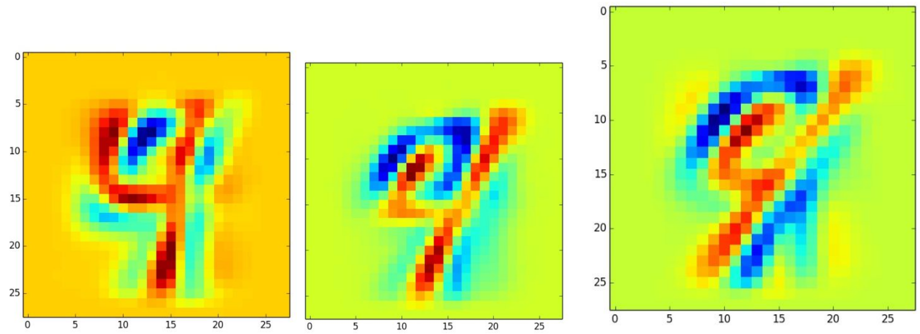


c)

```
Accuracy : 0.935208437971
1862/1991
Actual       0    1
Predicted
0          930   77
1           52  932
```

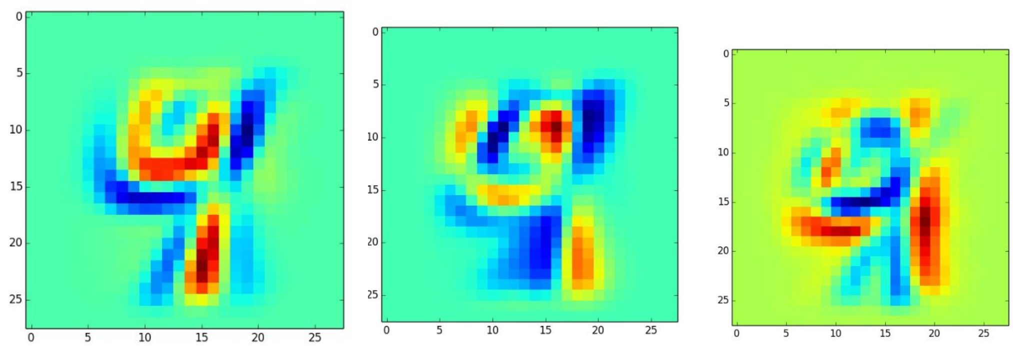
Where the 0 is a 4 and the 1 is a 9.

d)Printing Misclassified digits,
40,45, 64



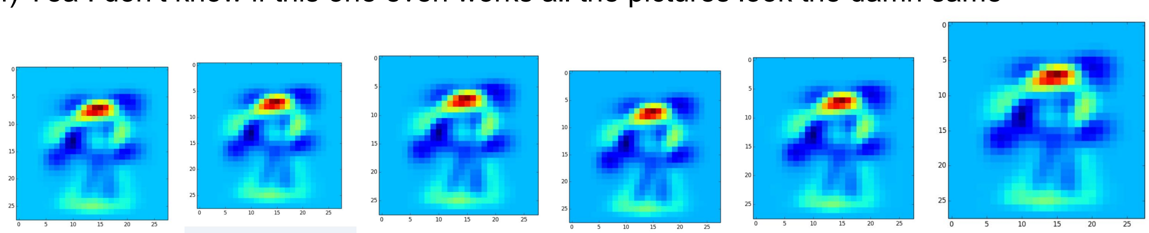
Predictive probability that
Is a 9 = 0.64 is a 9 = 0.78 is a 9 = 0.82
Actual =4 actual =4 actual =4

e) Most ambiguous predictions
1293, 676, 586



Probability = 0.5024 0.5064 0.50026

f) Yea I don't know if this one even works all the pictures look the damn same



W at 1 5 10 25 50 100
It's settling into something? Looks like the first few ones have a value of w that has some fluctuation. Anyways, I can't actually tell, because I'm pretty colorblind… they all look pretty much exactly the same