HW02

EECS 6770

Nan Zhao

nz2250

Problem 1:

Setup: $X = \{x_1, x_2 \cdots x_n\}$, $Z = \{z_1 \cdots\cdots z_N\}$ $\quad p(W) = \left(\frac{\lambda}{2\pi}\right)^{dk/2} \exp\left\{-\frac{\lambda}{2}\,\text{trace}(W^TW)\right\}$

$x_n \sim N(Wz_n, \sigma^2 I) \quad z_n \sim N(0,1)$

Find: $W' = \text{argmax} \ln p(x_1, \cdots, x_N, W)$.

Steps:

① use z as $q(\phi)$ in class example to set up "E" step.

$$\ln p(X,W) = \underbrace{\int q(z) \ln \frac{p(W,X,z)}{q(z)}\,dz}_{=\mathcal{L}(W)} + \underbrace{\int q(z) \ln \frac{q(z)}{p(z|W,X)}\,dz}_{=KL(q\|p)}$$

E-Step at iteration t.

set $q_t(z) = p(z \mid X, W_{t-1}) = \prod_{i=1}^{n} p(z_i \mid x_i, W_{t-1})$

$\mathcal{L}_t(W) = \int q_t(z) \ln p(X,W,z)\,dz - \int q_t(z) \ln q_t(z)\,dz$

for every i:

$$p(z_i|x_i, W_{t-1}) = \frac{p(x_i|z_i, W_{t-1})\,p(z_i|W_{t-1})\,p(W_{t-1})}{p(x_i, W_{t-1})}$$

since z doesn't depend on W and we dont want depend on $p(W_{t-1})$)

we have:

$$= \frac{p(x_i|z_i, W_{t-1})\,p(z_i)}{\int p(x_i|W, z_i)\,p(z_i)\,dz_i}$$

we look at $p(x_i|z_i, W_{t-1})\,p(z_i) = N(z_i; \mu_i, \Sigma_i)$

$p(x_i | W, z_i) p(z_i)$

$$= (2\pi)^{-\frac{d}{2}} (|\sigma^2 I|)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(x_i - Wz_i)^T(\sigma^2 I)^{-1}(x_i - Wz_i)\} \cdot (2\pi)^{-\frac{k}{2}} (|I|)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(z_i)^T (I)^{-1} z_i\}$$

$$= 2\pi^{-\frac{d+k}{2}} (|\sigma^2 I|)^{-\frac{1}{2}} \exp\{-\frac{1}{2}[(x_i - Wz_i)^T (\sigma^2 I)^{-1}(x_i - Wz_i) + z_i^T z_i]\}$$

$$\Sigma_i = (I + \frac{1}{\sigma^2} W_{t-1}^T W_{t-1})^{-1}, \quad \mu_i = \Sigma_i \cdot W_{t-1}^T x_n / \sigma^2$$

<u>M - Step:</u>

$$\mathcal{L}_t(W) = \int q_t(z) \ln p(X, W_{t-1}, Z) dz - \int q_t(z) \ln q_t(z) dz$$

where $q_t(z) = p(z | X, W_{t-1})$

sub it in, we have

$$\mathcal{L}_t(W) = \sum_{i=1}^{n} \int p(z_i | x_i, W_{t-1}) \ln p(x_i | W_{t-1}, z_i) p(W_{t-1}) p(z_i) dz_i$$

$$- \sum_{i=1}^{n} \int p(z_i | x_i, W) \ln p(z_i | x_i, W) dz_i$$

$$\Rightarrow \sum_{i=1}^{n} \int p(z_i | x_i, W_{t-1}) \ln p(x_i | W_{t-1}, z_i) p(W_{t-1}) p(z_i) dz_i + \text{constant}$$

$$= \sum_{i=1}^{n} \mathbb{E}_{q_t} \ln p(x_i | W_{t-1}, z_i) p(W_t) p(z_i) + \text{constant}$$

$$= \sum_{i=1}^{n} \mathbb{E}_{q_t}[\ln p(x_i | W_t, z_i)] + \mathbb{E}_{q_t}[\ln p(W_t)] + \mathbb{E}_{q_t}[\ln p(z_i) + \text{constant}$$

$$= \sum_{i=1}^{n} \mathbb{E}_{q_t}[\ln p(x_i | W_{t-1}, z_i)] + \mathbb{E}_{q_t}[\ln p(W_t)] + \text{constant}$$

for simplification, W now denotes W_{t-1}

$$= \sum_{i=1}^{n} \mathbb{E}_{q_t}[-\frac{1}{2\sigma^2}(x_i - Wz_i)^T (x_i - Wz_i)] + (-\frac{\lambda}{2} tr(W^T W)) + \text{constant}$$

$$= \sum_{i=1}^{n} -\frac{1}{2\sigma^2}\{-2tr(\mu_i x_i^T W) + tr(W^T W(\mu_i \mu_i^T + \Sigma))\} + (-\frac{\lambda}{2} tr(W^T W)) + \text{constant}$$

exclude every term not involving W and absorb into constant, use linear expectation, take derivative of $\mathcal{L}(W)$:

$$\mathcal{L}\nabla W = 0 = \sum_{i}^{n} \frac{1}{\sigma^2} x_i \mu_i^T - \sum_{i}^{n} \frac{1}{2\sigma^2} \mathbb{E}_{q}[2(Wz_i)z_i^T] - \frac{\lambda}{2} \cdot 2W$$

$$\sum_{i=1}^{n} x_i \mu_i^T = W\left[\sum_{i=1}^{n} (\Sigma_i + \mu_i \mu_i^T) + \sigma^2 \lambda \cdot I\right]$$

$$W = \left(\sum_{i=1}^{n} x_i \mu_i^T\right)\left(\sum_{i=}^{n} \mu_i \mu_i^T + \Sigma \cdot n + \sigma^2 \lambda \cdot I\right)^{-1}$$

Algorithm Steps:

First, we set $q(z_i) = p(z_i | x_i, W_t)$ and get $L(w)$

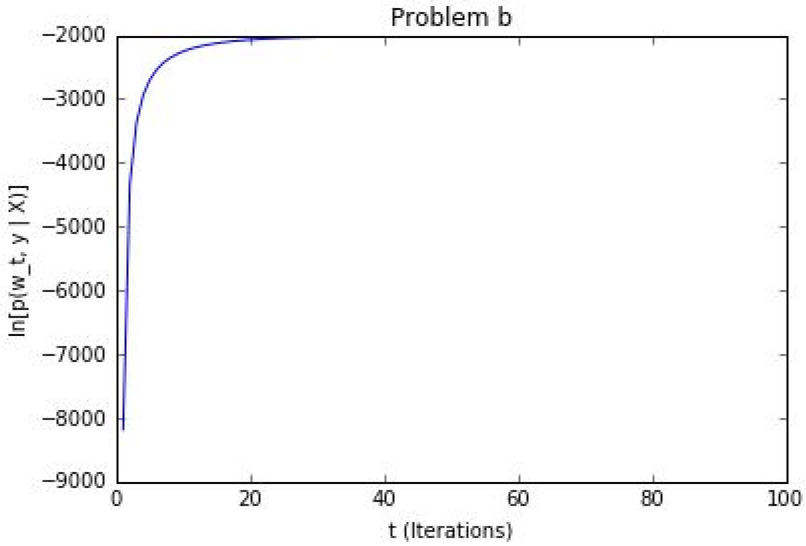
and then we take derivative of $L(w)$ and get W_{t+1} which maximize $L_{W_t}(W)$

After we get new W, we get the new $L(w)$.

Repeat the process to find better and better W.

Problem 2

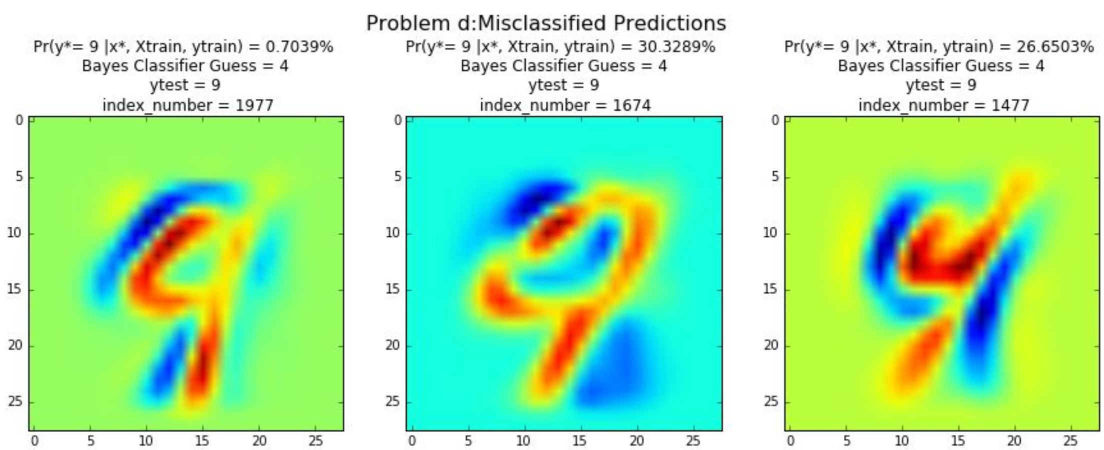
a) Just run iteration for 100 times



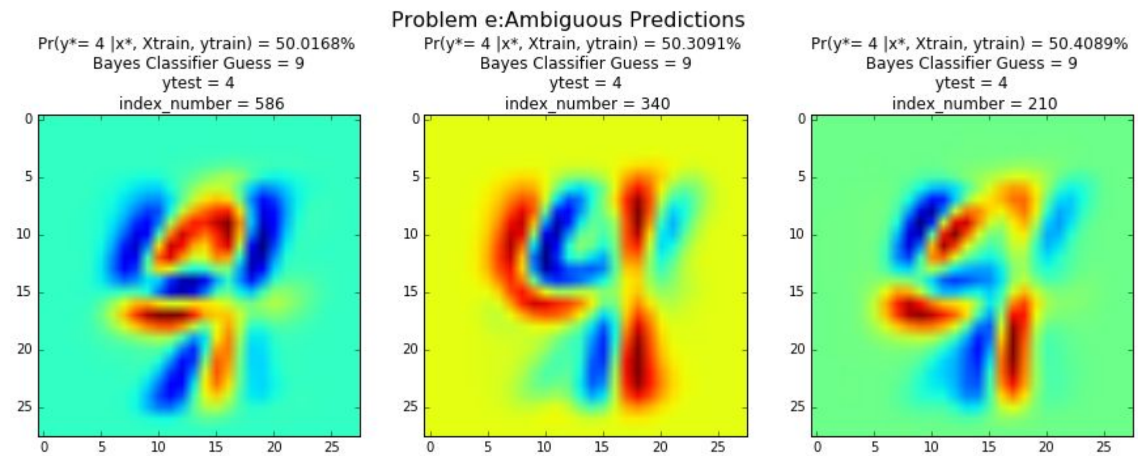
b)

	Classified as 4	Classified as 9
ytest= 4	930	52
ytest= 9	77	932

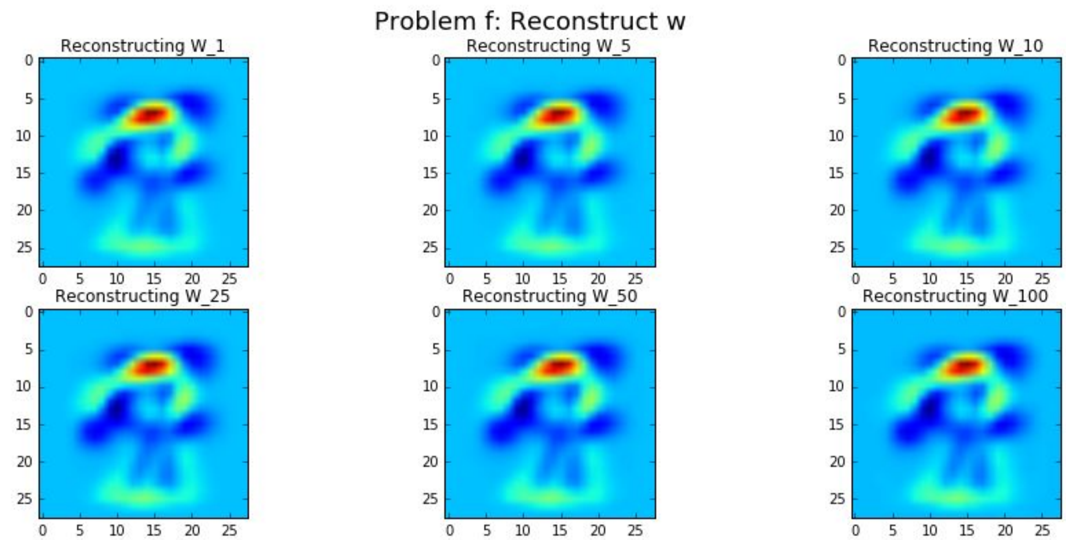
c)



d)


Problem e:Ambiguous Predictions
Pr(y*= 4 |x*, Xtrain, ytrain) = 50.0168%
Bayes Classifier Guess = 9
ytest = 4
index_number = 586
Pr(y*= 4 |x*, Xtrain, ytrain) = 50.3091%
Bayes Classifier Guess = 9
ytest = 4
index_number = 340
Pr(y*= 4 |x*, Xtrain, ytrain) = 50.4089%
Bayes Classifier Guess = 9
ytest = 4
index_number = 210

e)


Problem f: Reconstruct w
Reconstructing W_1
Reconstructing W_5
Reconstructing W_10
Reconstructing W_25
Reconstructing W_50
Reconstructing W_100

f)