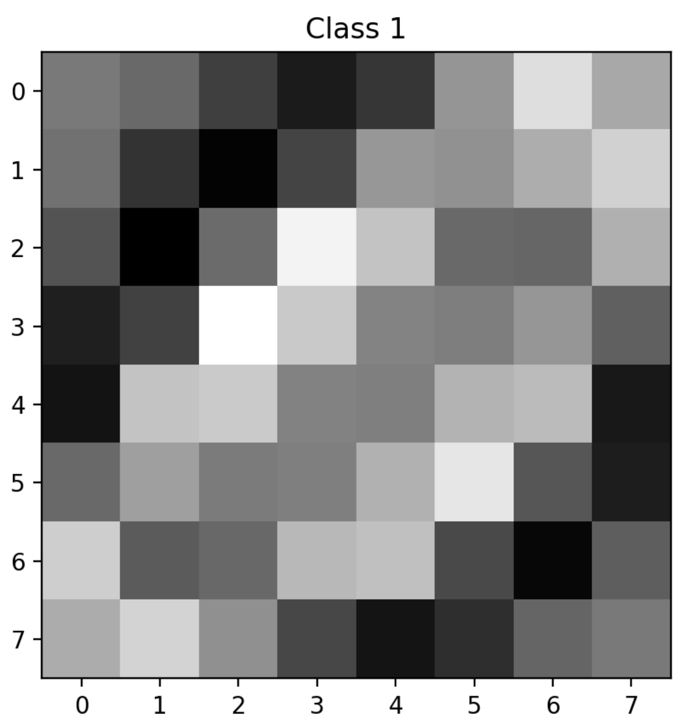
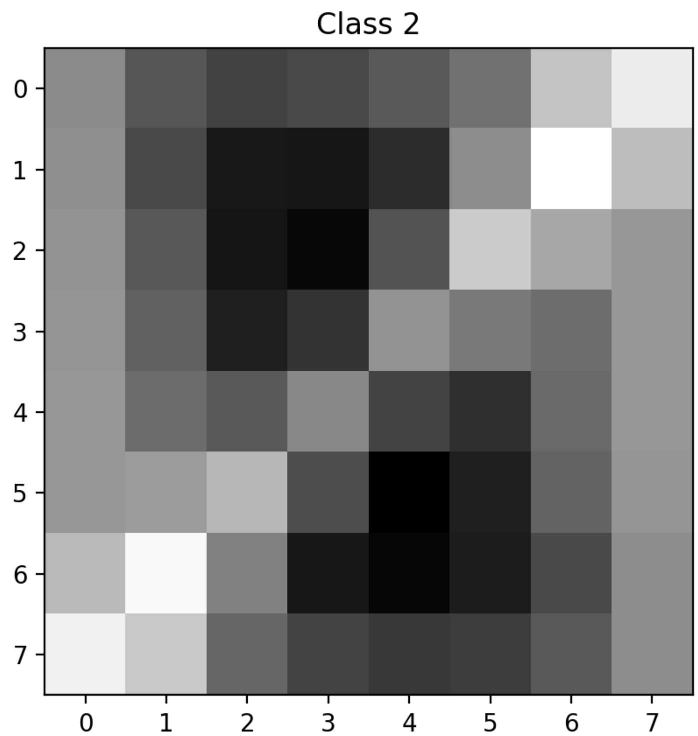
Q1

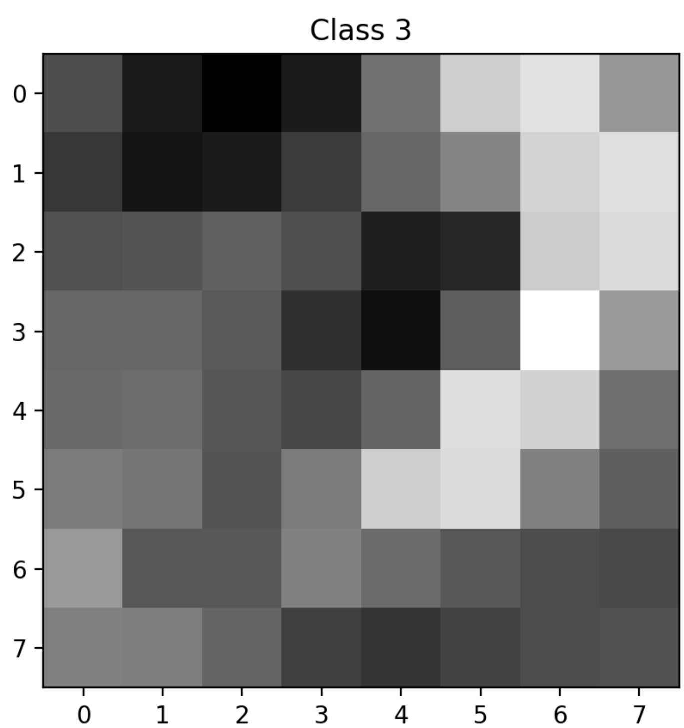
(a) and (b)

```
The average conditional log-likehood for train set: -2.427209529662601
The average conditional log-likehood for test set: -2.4992582962493053
The accuracy of train set prediction is 0.9814285714285714
The accuracy of test set prediction is 0.97275
```

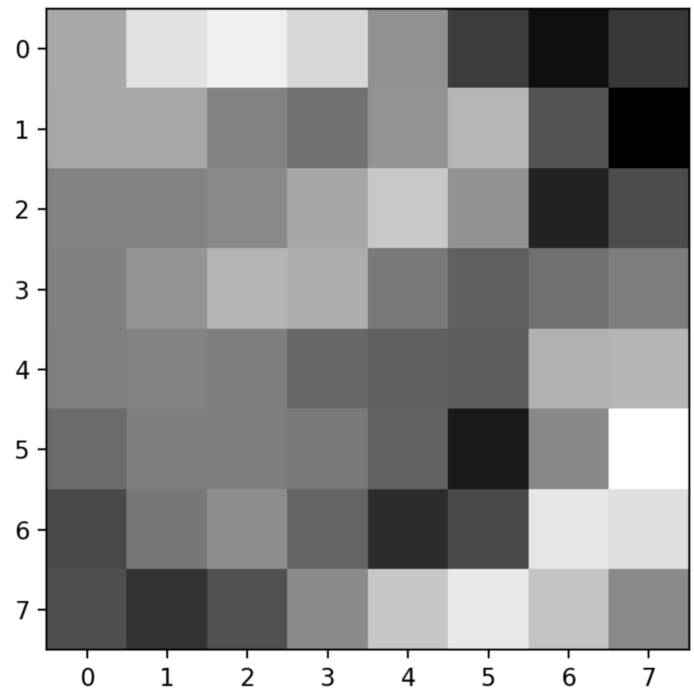
(c)



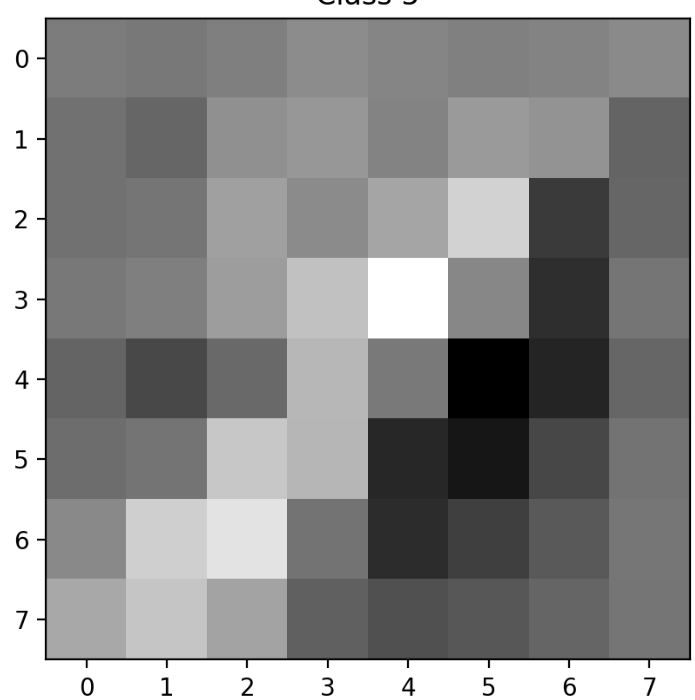

Class 2
0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7


Class 3
0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7

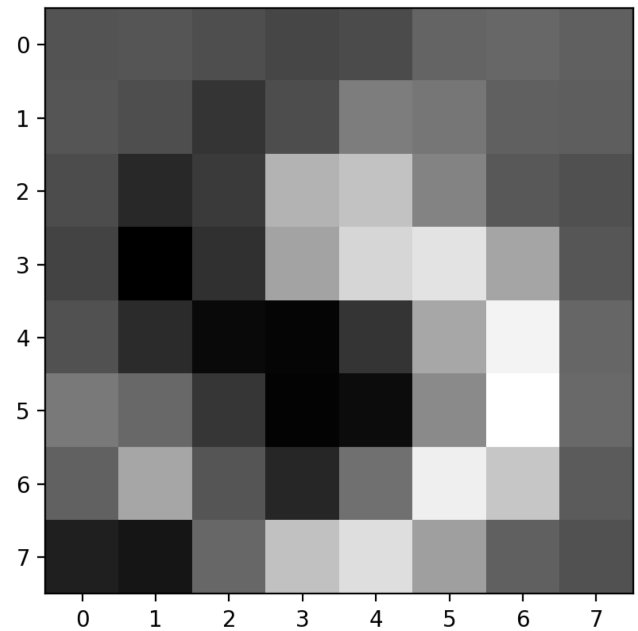
Class 4



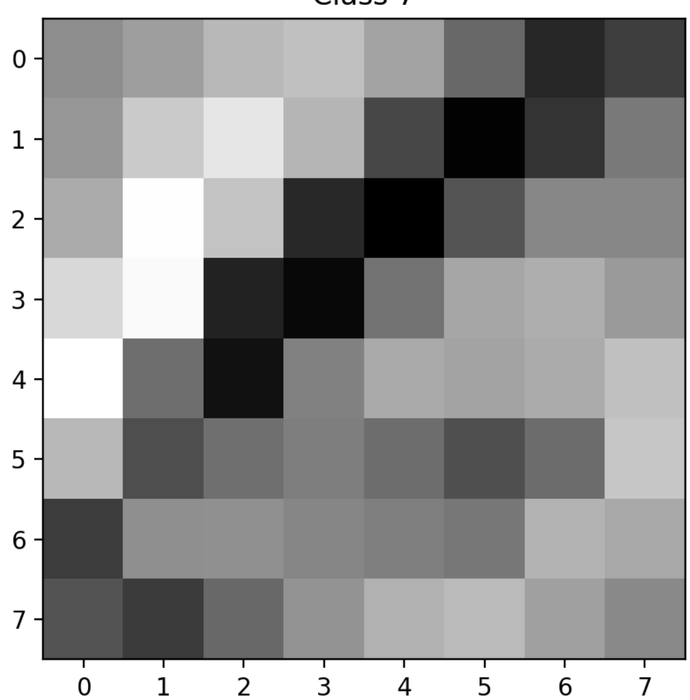
Class 5

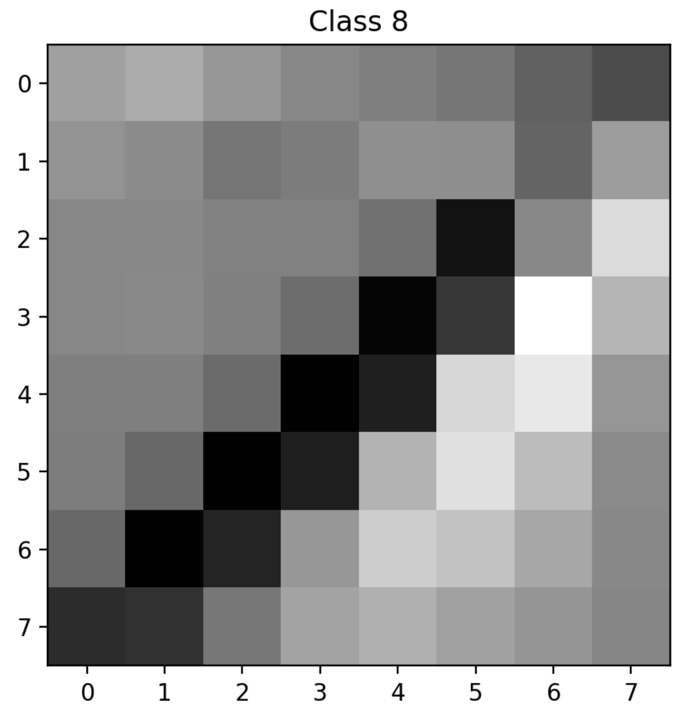


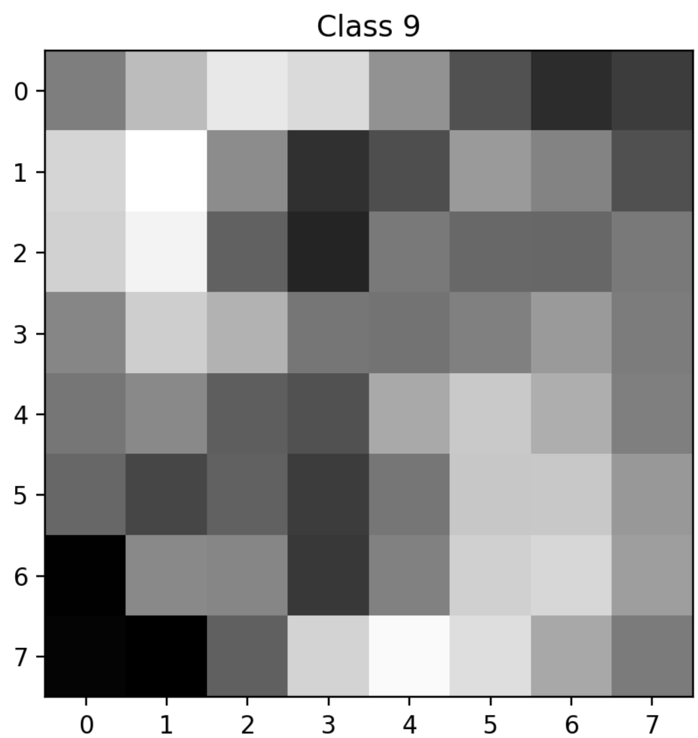
Class 6



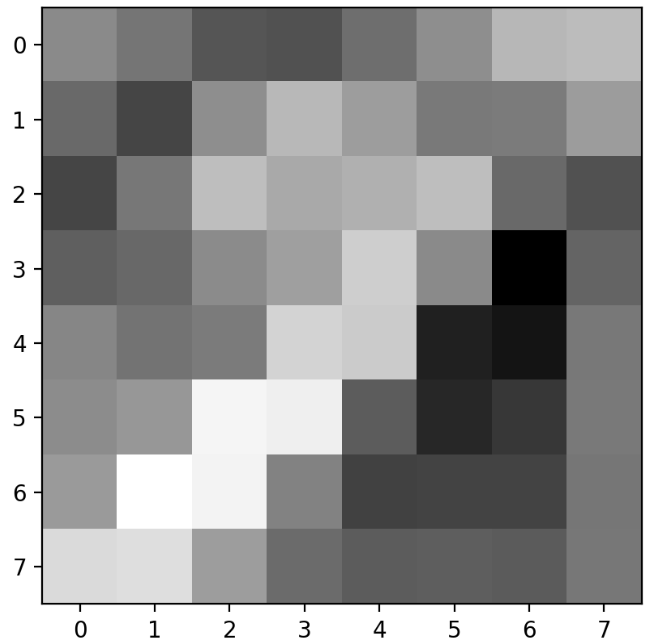
Class 7




Class 8
0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7


Class 9
0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7

Class 10



Q2.

$$P(X;\theta) = \prod_{k=1}^{K} \theta_k^{x_k}, \text{ where } X = \langle \underbrace{0 \cdots 1 \cdots 0}_{k} \rangle$$

$$\hat{\theta}_k = \frac{N_k}{N} \quad \text{for } \theta_1, \theta_2 \cdots \theta_K$$

$$\theta \sim \text{Dirichlet}(a_1, \dots, a_K).$$

$$\Rightarrow ① \; P(\theta) \propto \theta_1^{a_1 - 1} \cdots \theta_K^{a_K - 1}$$

$$② \; E[\theta_k] = \frac{a_k}{\sum_{k'} a_{k'}}$$

$$D = \{X^{(n)}\}_{n=1}^{N}$$

(a)
$$P(\theta|D) \propto P(\theta) \cdot P(D|\theta)$$

$$\propto \theta_1^{a_1-1} \cdots \theta_K^{a_K-1} \cdot P(X^{(1)}, \cdots, X^{(n)} | \theta)$$

$$= \theta_1^{a_1-1} \cdots \theta_K^{a_K-1} \cdot P(X^{(1)}|\theta) \cdot P(X^{(2)}|\theta) \cdots P(X^{(n)}|\theta)$$

$$= \theta_1^{a_1-1} \cdots \theta_K^{a_K-1} \cdot \prod_{k=1}^{K} \theta_k^{x_k^{(1)}} \cdot \prod_{k=1}^{K} \theta_k^{x_k^{(2)}} \cdots \prod_{k=1}^{K} \theta_k^{x_k^{(n)}}$$

$$= \prod_{k=1}^{K} \theta_k^{a_k-1} \prod_{k=1}^{K} \theta_k^{x_k^{(1)}} \cdot \prod_{k=1}^{K} \theta_k^{x_k^{(2)}} \cdots \prod_{k=1}^{K} \theta_k^{x_k^{(n)}}$$

$$= \prod_{k=1}^{K} \theta_k^{a_k-1} \cdot \theta_k^{x_k^{(1)}} \cdot \theta_k^{x_k^{(2)}} \cdots \theta_k^{x_k^{(n)}}$$

$$= \prod_{k=1}^{K} \theta_k^{\alpha_k - 1 + \sum_{n=1}^{N} x_k^n}$$

$$= \prod_{k=1}^{K} \theta_k^{\alpha_k - 1 + N_k}$$

$$= \prod_{k=1}^{K} \theta_k^{\alpha_k + N_k - 1}$$

which $\sim$ Dirichlet $(\alpha_1 + N_1, \dots, \alpha_K + N_K)$

Therefore,

$$E[\theta_k | D] = \frac{\alpha_k + N_k}{\sum_{k'} \alpha_{k'} + N_{k'}}$$

(b) $P(\theta|D) \propto \prod_{k=1}^{K} \theta_k^{\alpha_k + N_k - 1}$

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\ P(\theta|D)$$

$$\equiv \underset{\theta}{\operatorname{argmax}}\ \log P(\theta|D)$$

(Since log function keep same optimization direction)

$$= \underset{\theta}{\operatorname{argmax}}\ \log\left(\prod_{k=1}^{K} \theta_k^{\alpha_k + N_k - 1}\right)$$

$$= \underset{\theta}{\operatorname{argmax}} \sum_{k=1}^{K} \log\left(\theta_k^{\alpha_k + N_k - 1}\right)$$

$$= \underset{\theta}{\operatorname{argmax}} \sum_{k=1}^{K} (\alpha_k + N_k - 1) \cdot \log(\theta_k)$$

Also, we know $\sum_{k=1}^{K} \theta_k = 1$

$$\Rightarrow \quad f(\theta) = \sum_{k=1}^{K} (\alpha_k + N_k - 1) \log(\theta_k)$$

$$g(\theta) = \sum_{k=1}^{K} \theta_k - 1$$

$$\begin{aligned} L(\theta, \lambda) &= f(\theta) + \lambda g(\theta) \\ &= \sum_{k=1}^{K} (\alpha_k + N_k - 1) \log(\theta_k) + \\ &\qquad \lambda \left(\sum_{k=1}^{K} \theta_k - 1 \right) \end{aligned}$$

- take arbitrary $k \in [1, \cdots K]$

$$\frac{\partial L}{\partial \theta_k} = (\alpha_k + N_k - 1) \cdot \frac{1}{\theta_k} + \lambda$$

$$\frac{\partial L}{\partial \lambda} = \sum_{k=1}^{K} \theta_k - 1$$

$$\text{Let } \frac{\partial L}{\partial A_k} = 0 \Rightarrow A_k = -\frac{a_k + N_k - 1}{\lambda}$$

$$\frac{\partial L}{\partial \lambda} = 0 \Rightarrow \sum_{k=1}^{K} A_k = 1$$

$$\Rightarrow -\frac{a_1 + N_1 - 1}{\lambda} - \frac{a_2 + N_2 - 1}{\lambda} - \cdots - \frac{a_K + N_K - 1}{\lambda} = 1$$

$$\Rightarrow -\lambda = \sum_{k=1}^{K} a_k + N_k - K$$

$$\Rightarrow A_k = \frac{a_k + N_k - 1}{\sum_{k=1}^{K} a_k + N_k - K}$$

$\square$

3.

(a) From appendix, we have

$$p(z) = \mathcal{N}(z \mid \mu, \Lambda^{-1})$$

$$p(x \mid z) = \mathcal{N}(x \mid Az + b, L^{-1})$$

$$p(x) = \mathcal{N}(x \mid A\mu + b, L^{-1} + A\Lambda^{-1}A^T)$$

$$p(z \mid x) = \mathcal{N}(z \mid C(A^T L(x-b) + \Lambda\mu), C)$$

where $C = (\Lambda + A^T L A)^{-1}$

Since z is drawn from $\mathcal{N}(0, 1)$

$\Rightarrow \mu = 0, \ \Lambda = 1, \ b = 0$

Also, we know $x \mid z \sim \mathcal{N}(zu, \Sigma)$

$$\Rightarrow \quad p(z) = \mathcal{N}(0, 1)$$

$$p(x \mid z) = \mathcal{N}(z \mid zu, \Sigma)$$

$$p(x) = \mathcal{N}(x \mid 0, \Sigma + u \cdot u)$$

$$P(z|x;\theta) = N(z|C\cdot(u^T\cdot\Sigma^{-1}x),\ C)$$

where $C=(1+u^T\cdot\Sigma^{-1}\cdot u)^{-1}$

$$m = E(z|x;\theta) = C\cdot(u^T\cdot\Sigma^{-1}\cdot x)$$

$$= \frac{u^T\cdot\Sigma^{-1}\cdot x}{1+u^T\cdot\Sigma^{-1}u}$$

$$C = E(z^2|x;\theta) - [E(z|x;\theta)]^2$$

$$\Rightarrow E(z^2|x;\theta) = C + [E(z|x;\theta)]^2$$

(b)

Goal

Find μ_{new}, such that

$$\mu_{new} = \underset{\mu}{\arg\max} \frac{1}{N}\sum_{n=1}^{N} E_{q(z^{(n)})}\left[\log P(z^{(n)}, x^{(n)})\right]$$

$$= \underset{\mu}{\arg\max} \frac{1}{N}\sum_{n=1}^{N} E_{q(z^{(n)})}\left[\log P(x^{(n)} \mid z^{(n)}) + \log P(z^{(n)})\right]$$

$$= \underset{\mu}{\arg\max} \frac{1}{N}\sum_{n=1}^{N} E_{q(z^{(n)})}\left[\log P(x^{(n)} \mid z^{(n)})\right] + E_{q(z^{(n)})}\left[\log P(z^{(n)})\right]$$

Since $z \sim N(0,1)$

$$\Rightarrow N(z, 0, 1) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2} z^2\right)$$

$$= \underset{\mu}{\arg\max} \frac{1}{N}\sum_{n=1}^{N} E_{q(z^{(n)})}\left[\log P(x^{(n)} \mid z^{(n)})\right] +$$

$$E_{q(z^{(n)})}\left[-\frac{1}{2}\left(\log(2\pi) + (z^{(n)})^2\right)\right]$$

$$= \underset{\mu}{\arg\max} \frac{1}{N}\sum_{n=1}^{N} E_{q(z^{(n)})}\left[-\frac{1}{2}\log\Sigma - \frac{1}{2}\log 2\pi - \frac{1}{2}\cdot\frac{(x^{(n)} - z\mu)^2}{\Sigma}\right] +$$

$$E_{q(z^{(n)})}\left[-\frac{1}{2}\cdot\log 2\pi + (z^{(n)})^2\right]$$

$$\left(\text{Since } \log p(x|z) = \frac{1}{\sqrt{\Sigma}\cdot\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}\cdot\frac{(x-z\cdot u)^2}{\Sigma}\right)\right)$$

$$\Rightarrow \frac{\partial}{\partial u} \cdot f(u) = \frac{1}{N}\sum_{n=1}^{N} E_{q(z^{(n)})}\left[0 - 0 + \frac{1}{2}\cdot(x^{(n)} - zu)^T \cdot \Sigma^{-1} \cdot 2 \cdot z\right]$$

$$= \frac{1}{N}\sum_{n=1}^{N} E_{q(z^{(n)})}\left[(x^{(n)} - zu)^T \cdot \Sigma^{-1} \cdot z\right]$$

$$\text{Let } \frac{1}{N}\sum_{n=1}^{N} E_{q(z^{(n)})}\left[(x^{(n)} - zu)^T \cdot \Sigma^{-1} \cdot z\right] = 0$$

$$\Rightarrow \sum_{n=1}^{N} E_{q(z^{(n)})}\left[(x^{(n)} - zu)^T \cdot \Sigma^{-1} \cdot z\right] = 0$$

$$\Rightarrow \sum_{n=1}^{N} E_{q(z^{(n)})}\left[x^{(n)} \cdot \Sigma^{-1} \cdot z - zu\Sigma^{-1} z\right] = 0$$

$$\Rightarrow \sum_{n=1}^{N} E_{q(z^{(n)})}\left[x^{(n)} \cdot z\right] = \sum_{n=1}^{N} E_{q(z^{(n)})}\left[zuz\right]$$

$$\Rightarrow \sum_{n=1}^{N} E_{q(z^{(n)})}\left[z\right] x^{(n)} = \sum_{n=1}^{N} E_{q(z^{(n)})}\left[zz\right] \cdot u$$

$$\Rightarrow u = \frac{\sum_{n=1}^{N} E_{q(z^{(n)})}[z]\, x^{(n)}}{\sum_{n=1}^{N} E_{q(z^{(n)})}[zz]}$$

Since $m = E(z | x ; \theta)$

$$S = E(z^2 | x ; \theta)$$

$$\Rightarrow \mu = \frac{\sum_{n=1}^{N} m^{(n)} x^{(n)}}{\sum_{n=1}^{N} S^{(n)}}$$

$\square$