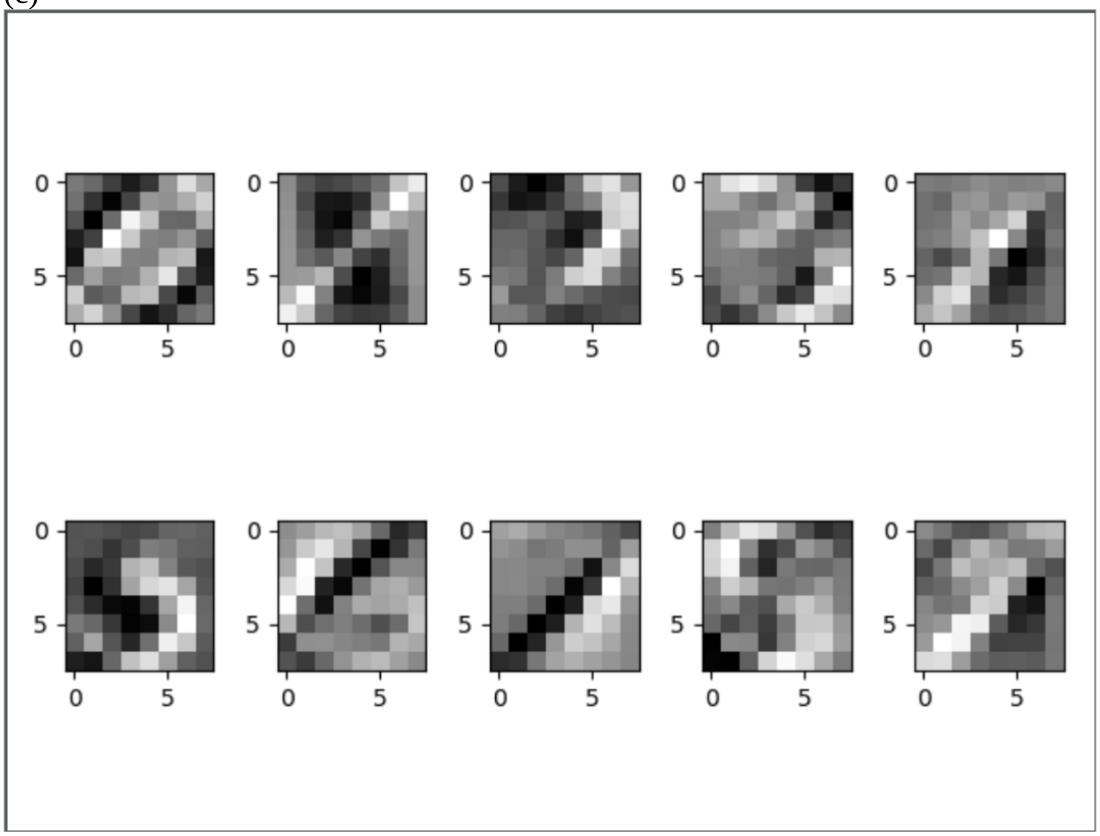
Name: Yuhan Shao
Student Number: 1002281327
UTORid: shaoyuha

**CSC411 Assignment 5**

1. (a)(b)

```
Train average likelihood:  -0.12462443666863032
Test average likelihood:  -0.1966732032552558
Train accuracy:  0.9814285714285714
Test accuracy:  0.97275
```

(c)



2.(a) • By the problem: There are $N$ samples and $K$ classes:
Therefore, we can get the likelihood:

$$p(D|\vec{\theta}) = \prod_{i=1}^{N}\prod_{k=1}^{K}\theta_k^{x_k^{(i)}} = \prod_{i=1}^{N}\prod_{k=1}^{K}\theta_k^{\mathbb{I}\{x_i=k\}} = \prod_{k=1}^{K}\theta_k^{\sum_{i=1}^{N}\mathbb{I}\{x_i=k\}}$$

$$= \prod_{k=1}^{K}\theta_k^{N_k} \quad \# N_k := \text{the count for outcome } k$$

• Since $p(\vec{\theta}|D) = \frac{P(\vec{\theta})\cdot P(D|\vec{\theta})}{P(D)}$, we have:

$$p(\vec{\theta}|D) \propto P(D|\vec{\theta})P(\vec{\theta})$$

$$\propto \left(\prod_{k=1}^{K}\theta_k^{N_k}\right)\left(\prod_{k=1}^{K}\theta_k^{\alpha_k-1}\right) = \prod_{k=1}^{K}\theta_k^{(N_k+\alpha_k)-1}$$

• Therefore, $p(\vec{\theta}|D) \propto \text{Dirichlet}(N_1+\alpha_1, N_2+\alpha_2, \ldots, N_K+\alpha_K)$

• Hence, $p(x=k|D) = \int p(x=k|\vec{\theta})\, p(\vec{\theta}|D)$

$$= \int \theta_k\, p(\vec{\theta}|D)$$

$$= E[\theta_k|D]$$

$$= \frac{N_k+\alpha_k}{\sum_{k=1}^{K}(N_k+\alpha_k)} = \frac{N_k+\alpha_k}{(\sum_{k=1}^{K}\alpha_k)+N}$$

(b) • Denotes: $\hat{\theta} = \text{argmax}\, p(\vec{\theta}|D) = \text{argmax}\log(p(\vec{\theta}|D))$

$$f = \log(p(\vec{\theta}|D)) = \log\left(\prod_{k=1}^{K}\theta_k^{N_k+\alpha_k-1}\right) = \sum_{k=1}^{K}(N_k+\alpha_k-1)\log(\theta_k)$$

• Consider using Lagrange's Theorem with constraint: $g(\vec{\theta}) = \sum_{k=1}^{K}\theta_k = 1$ to find $\theta_k$ that makes maximum value of $f$:

$$\begin{cases} \frac{\partial f}{\partial \theta_k} = \frac{N_k+\alpha_k-1}{\theta_k} \\ \frac{\partial g}{\partial \theta_k} = 1 \end{cases}, \text{ where } 1 \le k \le K$$

$$\Rightarrow \begin{cases} \frac{\partial f}{\partial \theta_k} = \lambda \frac{\partial g}{\partial \theta_k} \\ g(\theta) = \sum_{i=1}^{k}\theta_i = 1 \end{cases}$$

$$\Rightarrow \frac{N_1+\alpha_1-1}{\theta_1} = \frac{N_2+\alpha_2-1}{\theta_2} = \cdots = \frac{N_K+\alpha_K-1}{\theta_K} = \lambda$$

$$\Rightarrow \theta_k = \frac{N_k+\alpha_k-1}{\lambda}$$

• Therefore, to compute $\lambda$:
From $g(\vec{\theta})$: $\sum_{k=1}^{K}\theta_k = \sum_{k=1}^{K}\frac{N_k+\alpha_k-1}{\lambda}$

$$= \frac{N+(\sum_{k=1}^{K}\alpha_k)-K}{\lambda} = 1$$

$\Rightarrow$ So, get $\lambda = N + (\sum_{k=1}^{K}\alpha_k) - K$

$\Rightarrow$ Therefore, $\hat{\theta}_k = \frac{N_k+\alpha_k-1}{N+(\sum_{k=1}^{K}\alpha_k)-K}$

3. (a) • For scalar-valued $z$ with the probabilistic:

$z \sim \mathcal{N}(0,1)$ and $\vec{x}|z \sim \mathcal{N}(z\vec{u}, \Sigma)$ where $\Sigma = \mathrm{diag}(\sigma_1^2, \dots, \sigma_D^2)$

• Then, with the Appendix, plug in: $b=0$, $\mu=0$, $\Lambda=1$, we have:

$$\begin{cases} p(z) = \mathcal{N}(0,1) \\ p(\vec{x}) = \mathcal{N}(0, \Sigma + \vec{u}\cdot\vec{u}) \\ p(\vec{x}|z) = \mathcal{N}(z \mid z\vec{u}, \Sigma) \\ p(z|\vec{x}) = \mathcal{N}(z \mid C\cdot(\vec{u}^T\Sigma^{-1}\vec{u}), C) \quad \text{where } C = \frac{1}{1+\vec{u}^T\Sigma^{-1}\vec{u}} \end{cases}$$

• Therefore,

mean: $E[z|\vec{x}] = C\cdot(\vec{u}^T\Sigma^{-1}\vec{x}) = \frac{\vec{u}^T\Sigma^{-1}\vec{x}}{1+\vec{u}^T\Sigma^{-1}\vec{u}}$

var: $E[z^2|\vec{x}] - (E[z|\vec{x}])^2 = C$

$$E[z^2|\vec{x}] = (E[z|\vec{x}])^2 + C = \left(\frac{\vec{u}^T\Sigma^{-1}\vec{x}}{1+\vec{u}^T\Sigma^{-1}\vec{u}}\right)^2 + \frac{1}{1+\vec{u}^T\Sigma^{-1}\vec{u}}$$

(b) • $\vec{u}_{new} \leftarrow \mathrm{argmax}_{\vec{u}} \left(\frac{1}{N}\sum_{i=1}^N E_{q(z^{(i)})}[\log p(z^{(i)}, \vec{x}^{(i)})]\right)$

$$= \mathrm{argmax}_{\vec{u}} \left(\frac{1}{N}\sum_{i=1}^N E_{q(z^{(i)})}[\log(p(\vec{x}^{(i)}|z^{(i)})\cdot p(z^{(i)}))]\right)$$

$$= \mathrm{argmax}_{\vec{u}} \left(\frac{1}{N}\sum_{i=1}^N E_{q(z^{(i)})}[\log p(\vec{x}^{(i)}|z^{(i)}) + \log p(z^{(i)})]\right)$$

$$= \mathrm{argmax}_{\vec{u}} \left(\frac{1}{N}\sum_{i=1}^N E_{q(z^{(i)})}[\log p(\vec{x}^{(i)}|z^{(i)})] + E_{q(z^{(i)})}[\log p(z^{(i)})]\right)$$

$\mathcal{N}(z, 0, 1)$

$$= \mathrm{argmax}_{\vec{u}} \left(\frac{1}{N}\sum_{i=1}^N E_{q(z^{(i)})}[\log p(\vec{x}^{(i)}|z^{(i)})] + E_{q(z^{(i)})}\left[\log\left(\frac{\exp(-\frac{(z^{(i)})^2}{2})}{\sqrt{2\pi}}\right)\right]\right)$$

$$= \mathrm{argmax}_{\vec{u}} \left(\frac{1}{N}\sum_{i=1}^N E_{q(z^{(i)})}[\log p(\vec{x}^{(i)}|z^{(i)})] + E_{q(z^{(i)})}[-\tfrac{1}{2}((z^{(i)})^2 + \log 2\pi)]\right) \quad (*)$$

• Then, by $\vec{x}|z \sim \mathcal{N}(z\vec{u}, \Sigma)$, consider:

$$E_{q(z^{(i)})}[\log p(\vec{x}^{(i)}|z^{(i)})]$$

$$= E_{q(z^{(i)})}\left[\log\left(\frac{\exp(-\frac{1}{2}\cdot\frac{(x-z\vec{u})^2}{\Sigma})}{\sqrt{2\pi\Sigma}}\right)\right]$$

$$= E_{q(z^{(i)})}\left[-\tfrac{1}{2}\left(\log\Sigma + \log 2\pi + \frac{(x^{(i)}-z\vec{u})^2}{\Sigma}\right)\right] \quad (**)$$

• Then, plug $(**)$ into $(*)$:

$$\vec{u}_{new} \leftarrow \mathrm{argmax}_{\vec{u}}\left(\frac{1}{N}\sum_{i=1}^N E_{q(z^{(i)})}\left[-\tfrac{1}{2}\left(\log\Sigma + \log 2\pi + \frac{(x^{(i)}-z\vec{u})^2}{\Sigma}\right)\right] + E_{q(z^{(i)})}[-\tfrac{1}{2}((z^{(i)})^2 + \log 2\pi)]\right)$$

$$\vec{u}_{new} \leftarrow \mathrm{argmax}_{\vec{u}}\left(\frac{1}{N}\sum_{i=1}^N E_{q(z^{(i)})}\left[-\tfrac{1}{2}\left(\log\Sigma + \frac{(x^{(i)}-z\vec{u})^2}{\Sigma} + (z^{(i)})^2\right)\right]\right)$$

# ignore constants NOT influence the argmax.

• Take partial derivative w.r.t $\vec{u}$:

Let $f(\vec{u}) = \sum_{i=1}^N E_{q(z^{(i)})}\left[-\tfrac{1}{2}\left(\log\Sigma + \frac{(x^{(i)}-z\vec{u})^2}{\Sigma} + (z^{(i)})^2\right)\right]$

$$\frac{\partial f(\vec{u})}{\partial \vec{u}} = \sum_{i=1}^N E_{q(z^{(i)})}\left[\frac{(x^{(i)}-z\vec{u})}{\Sigma}\cdot z\right]$$

$$= -\frac{1}{\Sigma}\left(\sum_{i=1}^N E_{q(z^{(i)})}[x^{(i)}\cdot z] - \sum_{i=1}^N E_{q(z^{(i)})}[z\vec{u}z]\right)$$

• Set $\frac{\partial f(\vec{u})}{\partial \vec{u}} = 0$, then we have:

$$\sum_{i=1}^N E_{q(z^{(i)})}[x^{(i)}\cdot z] = \sum_{i=1}^N E_{q(z^{(i)})}[z\vec{u}z]$$

• Therefore, $\vec{u}_{new} = \frac{\sum_{i=1}^N E_{q(z^{(i)})}[x^{(i)}\cdot z]}{\sum_{i=1}^N E_{q(z^{(i)})}[z^2]}$

$$= \frac{\sum_{i=1}^N (E_{q(z^{(i)})}[z|\vec{x}]\cdot x^{(i)})}{\sum_{i=1}^N E_{q(z^{(i)})}[z^2|\vec{x}]}$$

# given $\vec{x}$, not change the expected value of $z$ and $z^2$.

$$= \frac{\sum_{i=1}^N (m^{(i)}\cdot x^{(i)})}{\sum_{i=1}^N s^{(i)}}$$

# $m^{(i)} = E[z|\vec{x}]$

# $s^{(i)} = E[z^2|\vec{x}]$