Assignment 1
COMP8600

March 24, 2022

Name: Xuecheng Zhang (**u6284513**),
Yichao Jiang (**u7236045**)

07/03/2022

1 Question 1

Answer to Question 1.1

The class belongs to program A.

Answer to Question 1.2

Assume that X = {Program}, Y = {The class contains X% of boys}

$$P(X=A|Y=55\%) = \frac{P(Y=55\%|X=A)P(X=A)}{\sum_X P(Y=55\%|X=A)P(X=A)} \tag{1}$$

$$P(X=B|Y=55\%) = \frac{P(Y=55\%|X=B)P(X=B)}{\sum_X P(Y=55\%|X=B)P(X=B)} \tag{2}$$

Assume that we have N students in this class and $P(Y=55\%|X=A)$, $P(Y=55\%|X=B)$ satisfies binomial distribution and thus we obtain

$$(1) = \frac{C_n^{0.55n}0.65^{0.55n}(1-0.65)^{0.45n} \times \frac{1}{2}}{C_n^{0.55n}0.65^{0.55n}(1-0.65)^{0.45n} \times \frac{1}{2} + C_n^{0.55n}0.45^{0.55n}(1-0.45)^{0.45n} \times \frac{1}{2}} \tag{3}$$

$$(2) = \frac{C_n^{0.55n}0.45^{0.55n}(1-0.45)^{0.45n} \times \frac{1}{2}}{C_n^{0.55n}0.65^{0.55n}(1-0.65)^{0.45n} \times \frac{1}{2} + C_n^{0.55n}0.45^{0.55n}(1-0.45)^{0.45n} \times \frac{1}{2}} \tag{4}$$

Thus, we calculate $\frac{P(X=A|Y=55\%)}{P(X=B|Y=55\%)}$

$$\frac{P(X=A|Y=55\%)}{P(X=B|Y=55\%)} = \frac{0.65^{0.55n}0.35^{0.45n}}{0.45^{0.55n}0.55^{0.45n}} \tag{5}$$

$$= \frac{13}{9}^{0.55n} \frac{7}{11}^{0.45n} \tag{6}$$

We apply the logarithm to (6), and we get

$$0.55ln(\frac{13}{9})n + 0.45ln(\frac{7}{11})n \approx -0.0011n < 0 \tag{7}$$

Thus, we can derive that $P(X=B|Y=55\%) > P(X=A|Y=55\%)$, thus the class is more likely belong to **program B**.

Answer to Question 1.3

The variance of program A is $\sigma = n \times 0.65 \times 0.35 = 0.2275n$, while the variance of program B is $\sigma = n \times 0.45 \times 0.55 = 0.2475n$. The variance of B is larger than the variance of A. Also, since the difference between 45%,55% and 55%, 65% are the same, the program which obtains larger variance is more likely.

2 Question 2

Answer to Question 2.1

$$q(x|\eta) = exp(\eta^T u(x) - \psi(\eta)) = \frac{exp(\eta^T u(x))}{exp(\psi(\eta))} \tag{8}$$

$$\int q(x|\eta)dx = \int \frac{exp(\eta^T u(x))}{exp(\psi(\eta))}dx = \frac{\int exp(\eta^T u(x))dx}{exp(\psi(\eta))} \tag{9}$$

Since $\psi(\eta) = log \int(\eta^T u(x))dx$, we have

$$\int q(x|\eta)dx = \frac{\int exp(\eta^T u(x))dx}{exp(\log\int(\eta^T u(x))dx)} = \frac{\int exp(\eta^T u(x))dx}{\int exp(\eta^T u(x))dx} = 1 \tag{10}$$

In addition, since exp function for x $\in \mathbb{R}$, we have $exp(x) > 0$, i.e. $q(x|\eta) > 0$. Thus, combined with (10), we prove that $q(x|\eta)$ is a valid probability density function.

Answer to Question 2.2

$$P(\mu|x) = \frac{P(x|\mu)P(\mu)}{P(x)} = \frac{P(x|\mu)P(\mu)}{\int P(x|\mu)P(\mu)d\mu} \tag{11}$$

Since $\int P(x|\mu)P(\mu)d\mu$ does not relate to μ (also is not related to η and u) and can be removed. i.e. $P(\mu|x) \propto P(x|\mu)P(\mu)$

$$\begin{aligned} P(\mu|x) &\propto P(x|\mu)P(\mu) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}(x-\mu)^2 + \eta^T u(\mu) - \log\int exp(\eta^T u(\mu))d\mu\} \end{aligned} \tag{12}$$

Since $\log\int exp(\eta^T u(\mu))d\mu$ does not relate to μ and we can rewrite (12) as below:

$$\begin{aligned} &\frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}(x-\mu)^2 + \eta^T u(\mu)\} \\ =&\frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}x^2 - \frac{1}{2\sigma^2}\mu^2 + \frac{x}{2\sigma^2}2\mu + \eta^T u(\mu)\} \end{aligned} \tag{13}$$

Thus, we obtain that

$$\begin{cases} \hat{\eta} = \begin{bmatrix} \frac{x}{\sigma^2} \\ -\frac{1}{2\sigma^2} \\ \eta^T \end{bmatrix} \\ \hat{u} = \begin{bmatrix} \mu \\ \mu^2 \\ u(\mu) \end{bmatrix} \end{cases} \tag{14}$$

And we have shown that $\mu|x \backsim EXP(\hat{u}, \hat{\eta})$

Answer to Question 2.3

$$\begin{aligned}P(x|\mu) &= \frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}x^2 - \frac{1}{2\sigma^2}\mu^2 + \frac{\mu}{\sigma^2}x\}\end{aligned} \tag{15}$$

Let $u = (x, x^2)$, we can obtain $\eta = (\frac{\mu_0}{\sigma_0^2}, -\frac{1}{2\sigma_0^2})$
Thus, we have

$$u = (x, x^2), \eta = (\frac{\mu_0}{\sigma_0^2}, -\frac{1}{2\sigma_0^2}) \tag{16}$$

We apply the result of 2.2 into (16) and we obtain below:

$$\begin{cases} \hat{\eta} = \begin{bmatrix} \frac{x}{\sigma^2} \\ -\frac{1}{2\sigma^2} \\ \frac{\mu_0}{\sigma_0^2} \\ -\frac{1}{2\sigma_0^2} \end{bmatrix} \\ \hat{u} = \begin{bmatrix} \mu \\ \mu^2 \\ x \\ x^2 \end{bmatrix} \end{cases} \tag{17}$$

Answer to Question 2.4

1. show that the minimisation of $D_{KL}[\overline{q}(x) : q(x|\eta)]$ is equivalent to the MLE of η using all data points x_i.

Minimise $D_{KL}[\overline{q}(x) : q(x|\eta)]$**:**

$$\begin{aligned}&D_{KL}[\overline{q}(x) : q(x|\eta)] \\ &= \int \frac{1}{N}\sum_{i=1}^{N}\sigma(x-x_i)log\frac{\frac{1}{N}\sum_{i=1}^{N}\sigma(x-x_i)}{q(x|\eta)}dx \\ &= \int \frac{1}{N}\sum_{i=1}^{N}\sigma(x-x_i)\left(log\frac{1}{N}\sum_{i=1}^{N}\sigma(x-x_i) - log(q(x|\eta))\right)dx \\ &= \int \frac{1}{N}\sum_{i=1}^{N}\sigma(x-x_i)\left(log\frac{1}{N}\sum_{i=1}^{N}\sigma(x-x_i) - \eta^T u(x) + \psi(\eta)\right)dx\end{aligned} \tag{18}$$

Next, we compute the derivative of (18). Since $\int \frac{1}{N}\sum_{i=1}^{N}\sigma(x-x_i)(log\frac{1}{N}\sum_{i=1}^{N}$ $\sigma(x-x_i))dx$ does not have η parameter and thus its derivative with respect to η is 0. Thus, we have

$$\frac{\partial}{\partial\eta}\left(-\int\frac{1}{N}\sum_{i=1}^{N}\sigma(x-x_i)\eta^T u(x)dx + \int\frac{1}{N}\sum_{i=1}^{N}\sigma(x-x_i)\psi(\eta)\mathrm{d}x\right)$$

$$
\begin{aligned}
&= \frac{\partial \left(-\frac{\eta^T}{N} \sum_{i=1}^{N} u(x_i) + \frac{1}{N} \sum_{i=1}^{N} \psi(\eta) \right)}{\partial \eta} \\
&= \frac{\partial \left(-\frac{\eta^T}{N} \sum_{i=1}^{N} u(x_i) + \frac{1}{N} \sum_{i=1}^{N} \psi(\eta) \right)}{\partial \eta} \\
&= \frac{\partial \left(-\frac{\eta^T}{N} \sum_{i=1}^{N} u(x_i) + \psi(\eta) \right)}{\partial \eta} \\
&= -\frac{\sum_{i=1}^{N} u(x_i)}{N} + \frac{\partial \psi(\eta)}{\partial \eta}
\end{aligned} \tag{19}
$$

Let $-\frac{\sum\limits_{i=1}^{N} u(x_i)}{N} + \frac{\partial \psi(\eta)}{\partial \eta} = 0$, and we obtain $\frac{\partial \psi(\eta)}{\partial \eta} = \frac{\sum\limits_{i=1}^{N} u(x_i)}{N}$

Calculate the MLE of η

The log-likelihood for a single data point x is $\log q(x|\eta) = \eta^T u(x) - \psi(\eta)$. For $\{x_i\}_{i=1}^{N}$ a set of independently distributed points, the log-likelihood function is

$$
\log \prod_{i=1}^{N} q(x_i|\eta) = \sum_{i=1}^{N} \log q(x_i|\eta) = \eta^T \sum_{i=1}^{N} u(x_i) - \psi(\eta) \tag{20}
$$

We calculate the partial derivative of (20) with respect to η and we obtain that

$$
\frac{\partial \eta^T \sum\limits_{i=1}^{N} u(x_i) - \psi(\eta)}{\partial \eta} = \frac{\sum\limits_{i=1}^{N} u(x_i)}{N} - \frac{\partial \psi(\eta)}{\partial \eta} \tag{21}
$$

Let $\frac{\sum\limits_{i=1}^{N} u(x_i)}{N} - \frac{\partial \psi(\eta)}{\partial \eta} = 0$, and we obtain $\frac{\partial \psi(\eta)}{\partial \eta} = \frac{\sum\limits_{i=1}^{N} u(x_i)}{N}$

Thus, the minimisation of $D_{KL}[\overline{q}(x) : q(x|\eta)]$ is equivalent to the MLE of η using all data points x_i.

2. Show that the minimisation of $D_{KL}[\overline{q}(x) : q(x|\eta)]$ is equivalent to MLE using a single data point consisting of the mean of all data points x_i.

Since u is a linear map, we can rewrite

$$
\frac{\sum_{i=1}^{N} u(x_i)}{N} = u(\frac{\sum_{i=1}^{N} (x_i)}{N}) = u(\overline{x}) \tag{22}
$$

i.e. the both equations have $\nabla \psi(\eta) = u(\overline{x})$. Thus, we have proved that the minimisation of $D_{KL}[\overline{q}(x) : q(x|\eta)]$ is equivalent to MLE using a single data point consisting of the mean of all data points x_i.

Answer to Question 2.5

$$
\begin{aligned}
&D_{KL}[\eta_1 : \eta_2] = D_{KL}[q(x|\eta_1) : q(x|\eta_2)] \\
&= \int q(x|\eta_1)\log\left(\frac{q(x|\eta_1)}{q(x|\eta_2)}\right)dx \\
&= \int q(x|\eta_1)\log q(x|\eta_1)dx - \int q(x|\eta_1)\log q(x|\eta_2)dx \\
&= \int q(x|\eta_1)(\eta_1^T u(x) - \psi(\eta_1))dx - \int q(x|\eta_1)(\eta_2^T u(x) - \psi(\eta_2))dx
\end{aligned} \tag{23}
$$

$$
\begin{aligned}
LH &= \int q(x|\eta_1)(\eta_1^T u(x) - \psi(\eta_1))dx \\
&= \eta_1^T \int u(x)q(x|\eta_1)dx - \psi(\eta_1)\int q(x|\eta_1)dx
\end{aligned} \tag{24}
$$

Since $q(x|\eta_1) \backsim EXP(u, \eta)$ and $x \backsim EXP(u, \eta)$ and $\int q(x|\eta_1)dx = 1$ we have:

$$
\eta_1^T \int u(x)q(x|\eta_1)dx - \psi(\eta_1)\int q(x|\eta_1)dx = \eta_1^T \mathbb{E}[u(x)] - \psi(\eta_1) = \lambda_1^T \eta_1 - \psi(\eta_1)
$$

$$
\begin{aligned}
RH &= \int q(x|\eta_1)(\eta_2^T u(x) - \psi(\eta_2))dx \\
&= \eta_2^T \int u(x)q(x|\eta_1)dx - \psi(\eta_2)\int q(x|\eta_2)dx
\end{aligned} \tag{25}
$$

Since $q(x|\eta_1) \backsim EXP(u, \eta)$ and $x \backsim EXP(u, \eta)$ and $\int q(x|\eta_2)dx = 1$ we have:

$$
\eta_2^T \int u(x)q(x|\eta_1)dx - \psi(\eta_2)\int q(x|\eta_2)dx = \eta_2^T \mathbb{E}[u(x)] - \psi(\eta_2) = \lambda_1^T \eta_2 - \psi(\eta_2) \tag{26}
$$

Thus, we have proved that:

$$
D_{KL}[\eta_1 : \eta_2] = LH - RH = \psi(\eta_2) - \psi(\eta_1) - \lambda_1^T(\eta_2 - \eta_1)
$$

Answer to Question 2.6

From **question 2.5**, we obtain that

$$
\begin{aligned}
a^2 &= D_{KL}[\eta_1 : \eta_2] = \psi(\eta_2) - \psi(\eta_1) - \lambda_1^T(\eta_2 - \eta_1) \\
b^2 &= D_{KL}[\eta_2 : \eta_3] = \psi(\eta_3) - \psi(\eta_2) - \lambda_2^T(\eta_3 - \eta_2) \\
a^2 + b^2 &= \psi(\eta_3) - \psi(\eta_1) - \lambda_1^T(\eta_2 - \eta_1) - \lambda_2^T(\eta_3 - \eta_2) \\
c^2 &= D_{KL}[\eta_1 : \eta_3] = \psi(\eta_3) - \psi(\eta_1) - \lambda_1^T(\eta_3 - \eta_1)
\end{aligned}
$$

Thus, if we compare $\lambda_1^T(\eta_2 - \eta_1) + \lambda_2^T(\eta_3 - \eta_2)$ with $\lambda_1^T(\eta_3 - \eta_1)$, we can compare $a^2 + b^2$ with c^2.

First, we assume that $a^2 + b^2 = c^2$ and we need to show that $(\eta_2 - \eta_3) \perp (\lambda_1 - \lambda_2)$

$$
\begin{aligned}
&a^2 + b^2 = c^2 \\
&\lambda_1^T(\eta_2 - \eta_1) + \lambda_2^T(\eta_3 - \eta_2) = \lambda_1^T(\eta_3 - \eta_1)
\end{aligned}
$$

$$
\begin{aligned}
&\lambda_1^T\eta_2 - \lambda_1^T\eta_1 + \lambda_2^T\eta_3 - \lambda_2^T\eta_2 = \lambda_1^T\eta_3 - \lambda_1\eta_1 \\
&\lambda_1^T(\eta_2 - \eta_3) = \lambda_2^T(\eta_2 - \eta_3) \\
&(\lambda_1 - \lambda_2)^T(\eta_2 - \eta_3) = 0 \qquad (27)
\end{aligned}
$$

We have proved that $a^2 + b^2 = c^2 \implies (\eta_2 - \eta_3) \perp (\lambda_1 - \lambda_2)$

Next, we assume that $(\eta_2 - \eta_3) \perp (\lambda_1 - \lambda_2)$ and we need to show that $\lambda_1^T(\eta_2 - \eta_1) + \lambda_2^T(\eta_3 - \eta_2) = \lambda_1^T(\eta_3 - \eta_1)$ and thus $a^2 + b^2 = c^2$

$$
\begin{aligned}
&(\lambda_1 - \lambda_2)^T(\eta_2 - \eta_3) = 0 \\
&\lambda_1^T\eta_2 - \lambda_1^T\eta_3 - \lambda_2^T\eta_2 + \lambda_2^T\eta_3 = 0 \\
&\lambda_1^T\eta_2 - \lambda_2^T\eta_2 + \lambda_2^T\eta_3 = \lambda_1^T\eta_3 \\
&\lambda_1^T\eta_2 - \lambda_1^T\eta_1 + \lambda_2^T\eta_3 - \lambda_2^T\eta_2 = \lambda_1^T\eta_3 - \lambda_1^T\eta_1 \\
&\lambda_1^T(\eta_2 - \eta_1) + \lambda_2^T(\eta_3 - \eta_2) = \lambda_1^T(\eta_3 - \eta_1) \implies a^2 + b^2 = c^2 \qquad (28)
\end{aligned}
$$

We have proved that $(\eta_2 - \eta_3) \perp (\lambda_1 - \lambda_2) \implies a^2 + b^2 = c^2$.

In conclusion, we have shown that iff $a^2 + b^2 = c^2$ the difference in natural parameters $(\eta_2 - \eta_3)$ is perpendicular to the difference in expectation parameters in expectation parameters $(\lambda_1 - \lambda_2)$.

3 Question 3

Answer to Question 3.1

Since X, Z are independent and identically distributed.

$$
\begin{aligned}
&\sum_Z p(Z|X, \vartheta^{old}) \text{log } p(X, Z|\vartheta) \qquad (29) \\
&= \sum_Z P(Z|X, \vartheta^{old}) \text{log} \prod_{n=1}^N p(x_n, z_n|\vartheta) \qquad (30)
\end{aligned}
$$

$$
\begin{aligned}
\text{log} \prod_{n=1}^N p(x_n, z_n|\vartheta) &= \sum_{n=1}^N \text{log } p(x_n, z_n|\vartheta) \\
&= \sum_{n=1}^N \text{log } p(x_n|z_n, \vartheta) + \sum_{n=1}^N \text{log } p(z_n|\vartheta) \\
&= \sum_{n=1}^N \text{log} \prod_{k=1}^K \ (q(x_n|\eta_k))^{z_{nk}} + \sum_{n=1}^N \text{log} \prod_{k=1}^K \pi_k^{z_{nk}} \\
&= \sum_{n=1}^N \sum_{k=1}^K z_{nk} \text{log } q(x_n|\eta_k) + \sum_{n=1}^N \sum_{k=1}^K z_{nk} \text{log} \pi_k \\
&= \sum_{n=1}^N \sum_{k=1}^K z_{nk} (\text{log} q(x_n|\eta_k) + \text{log}\pi_k) \qquad (31)
\end{aligned}
$$

Then combine (30) with (31) and we obtain:

$$
\sum_Z p(Z|X, \vartheta^{old}) \text{log } p(X, Z|\vartheta) \qquad (32)
$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} P(z_{nk}|x_n, \vartheta^{old}) z_{nk}(\log\ q(x_n|\eta_k) + \log\ \pi_k) \tag{33}$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} p(z_{nk} = 1|x_n, \vartheta^{old})(\log\ q(x_n|\eta_k) + \log\ \pi_k) \tag{34}$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma^{old}(z_{nk})(\log\pi_k + \log\ q(x_n|\eta_k)) \tag{35}$$

Answer to Question 3.2

Firstly, we show that $\pi_k = \frac{N_k}{N}$. Since, in M step of EM, we need to maximise the expectation $\sum_Z\ p(Z|X, \vartheta^{old})log\ p(X, Z|\vartheta)$ in which we need to maximise $\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma^{old}(z_{nk})(\log\pi_k + \log\ q(x_n|\eta_k))$ under $\sum_k \pi_k = 1$ restriction.

Then, we introduce Lagrange multiplier λ.

$$\text{minimise} \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma^{old}(z_{nk})(\log\pi_k + \log\ q(x_n|\eta_k))$$
$$\text{subject to:} \sum_k \pi_k = 1$$

$$\mathcal{L}(\pi, \lambda) = \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma^{old}(z_{nk})(\log\pi_k + \log\ q(x_n|\eta_k)) + \lambda(\sum_k \pi_k - 1) \tag{36}$$

Next, calculate the partial derivative of $\frac{\partial \mathcal{L}}{\partial \pi}$.

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \pi_k} &= \frac{\partial}{\partial \pi_k}\sum_{n=1}^{N}\sum_{k=1}^{K} \gamma^{old}(z_{nk})\log\pi_k + \lambda(\sum_k \pi_k - 1) \\
&= \sum_{n=1}^{N}\frac{\partial}{\partial \pi_k}\left(\sum_{k=1}^{K} \gamma^{old}(z_{nk})\log\pi_k + \lambda(\sum_k \pi_k - 1)\right) \\
&= \sum_{n=1}^{N}\frac{\partial}{\partial \pi_k}\left(\gamma^{old}(z_{nk})\log\pi_k + \lambda\pi_k\right) \\
&= \sum_{n=1}^{N}\frac{\gamma^{old}(z_{nk})}{\pi_k} + \lambda = 0
\end{aligned} \tag{37}$$

Next, we need to solve the above equation (37). At first, we multiply both sides of the equation by π_k at the same time.

$$\sum_{n=1}^{N}\gamma^{old}(z_{nk}) + \pi_k\lambda = 0 \tag{38}$$

$$\sum_{k=1}^{K}\sum_{n=1}^{N}\gamma^{old}(z_{nk}) + \sum_{k=1}^{K}\pi_k\lambda = 0 \tag{39}$$

Since $\sum_{k=1}^{K}\pi_k = 1$ and thus we obtain:

$$\sum_{k=1}^{K}\sum_{n=1}^{N}\gamma^{old}(z_{nk}) + \lambda = 0 \tag{40}$$

$$N + \lambda = 0 \tag{41}$$

Thus, we obtain that $N = -\lambda$ and substitute it into (38).

$$\sum_{n=1}^{N} \gamma^{old}(z_{nk}) - N\pi_k = 0 \tag{42}$$

$$\pi_k = \frac{\sum_{n=1}^{N} \gamma^{old}(z_{nk})}{N} = \frac{N_k}{N} \tag{43}$$

Next, we need to show that $\lambda_k = \frac{1}{N_k}\sum_{n=1}^{N}(u(x_n) \cdot \gamma^{old}(z_{nk}))$. Firstly, from **Eq.(2.7)**, we know that $\lambda \overset{\text{def}}{=} \mathbb{E}_{x \backsim EXP(u,\eta)} = \nabla\psi(\eta)$ and $\lambda_k \overset{\text{def}}{=} \mathbb{E}_{x \backsim EXP(u_k,\eta_k)}$

To maximise the Eq. (3.5), we calculate the derivative of it w.r.t. η and let it equal to 0.

$$\begin{aligned}
&\frac{\partial}{\partial \eta_j} \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma^{old}(z_{nk})(\log \pi_k + \log\, q(x_n|\eta_k)) \\
=&\sum_{n=1}^{N} \frac{\partial}{\partial \eta_j} \sum_{k=1}^{K} \gamma^{old}(z_{nk}) \log\, q(x_n|\eta_k) \\
=&\sum_{n=1}^{N} \gamma^{old}(z_{nj}) \frac{\partial}{\partial \eta_j} \log\, q(x_n|\eta_j) \\
=&\sum_{n=1}^{N} \gamma^{old}(z_{nj}) \frac{\partial}{\partial \eta_j} \left(\eta_j^T u(x_n) - \psi(\eta_j)\right) \text{ From Eq. (2.4)} \\
=&\sum_{n=1}^{N} \gamma^{old}(z_{nj}) \left(u(x_n) - \lambda_j\right) \text{ From Eq. (2.7)} \\
=&\sum_{n=1}^{N} \gamma^{old}(z_{nj}) u(x_n) - \sum_{n=1}^{N} \gamma^{old}(z_{nj})\lambda_j = 0
\end{aligned} \tag{44}$$

And we solve the above equation and we obtain that:

$$\lambda_k = \frac{\sum_{n=1}^{N} \gamma^{old}(z_{nk}) u(x_n)}{\sum_{n=1}^{N} \gamma^{old}(z_{nk})} = \frac{1}{N_k} \sum_{n=1}^{N} u(x_n)\gamma^{old}(z_{nk}) \tag{45}$$

Thus, we have shown that $\lambda_k = \frac{1}{N_k}\sum_{n=1}^{N}(u(x_n) \cdot \gamma^{old}(z_{nk}))$ and thus $\eta_k = \nabla_{\varphi}(\frac{1}{N_k}\sum_{n=1}^{N}(u(x_n) \cdot \gamma^{old}(z_{nk})))$

Answer to Question 3.3

See code.

Answer to Question 3.4

$$\log\, p(t, \omega|\alpha, \beta) = \log\, p(\omega|\alpha) + \log\, p(t|\omega, \beta) \tag{46}$$

For $\log\, p(\omega|\alpha)$, the dimension of ω is R^M

$$\log\, p(\omega|\alpha) = -\log(\sqrt{(2\pi)^M \alpha^{-1}}) + \log(exp(-\frac{1}{2}\omega^T \alpha I \omega))$$

$$= -M\log(\sqrt{\frac{2\pi}{\alpha}}) - \frac{1}{2}\omega^T\alpha I\omega$$
$$= \frac{M}{2}\log(\frac{\alpha}{2\pi}) - \frac{\alpha}{2}\omega^T\omega \tag{47}$$

For log $p(t|\omega,\beta)$:

$$\begin{aligned}
\log\, p(t|\omega,\beta) &= \log\, \prod_{n=1}^{M} p(t_n|\omega,\beta) \\
&= \sum_{n=1}^{N} \log\, p(t_n|\phi_n,\omega,\beta) \\
&= \sum_{n=1}^{N} \log\, \mathcal{N}(t_n|y(\phi_n,\omega),\beta^{-1}) \\
&= \sum_{n=1}^{N} \log\, \frac{\beta}{\sqrt{2\pi}} exp(-\frac{(t_n-\omega^T\phi_n)^2\beta}{2}) \\
&= \sum_{n=1}^{N} \log\, \frac{\beta}{\sqrt{2\pi}} - \sum_{n=1}^{N} \frac{(t_n-\omega^T\phi_n)^2\beta}{2} \\
&= N\log\frac{\beta}{\sqrt{2\pi}} - \frac{\beta}{2}\sum_{n=1}^{N}(t_n-\omega^T\phi_n)^2
\end{aligned} \tag{48}$$

Thus, combined with (47) and (48), we are able to show that:

$$\begin{aligned}
\mathop{\mathbb{E}}_{Z|X,\vartheta^{old}}[\log\, p(t,\omega|\alpha,\beta)] &= \mathbb{E}_\omega\left(N\log\frac{\beta}{\sqrt{2\pi}} - \frac{\beta}{2}\sum_{n=1}^{N}(t_n-\omega^T\phi_n)^2 + \frac{M}{2}\log(\frac{\alpha}{2\pi}) - \frac{\alpha}{2}\omega^T\omega\right) \\
&= \frac{N}{2}\log\frac{\beta}{2\pi} - \frac{\beta}{2}\sum_{n=1}^{N}\mathbb{E}[(t_n-\omega^T\phi_n)^2] + \frac{M}{2}\log\frac{\alpha}{2\pi} - \frac{\alpha}{2}\mathbb{E}[\omega^T\omega]
\end{aligned} \tag{49}$$

Answer to Question 3.5

We can directly find derivation of equation (3.15) in the text:

$$\frac{\partial \mathop{\mathbb{E}}_{Z|X,\vartheta^{old}}[\log\, p(t,\omega|\alpha,\beta)]}{\partial\alpha} = 0 \tag{50}$$
$$\frac{\partial \mathop{\mathbb{E}}_{Z|X,\vartheta^{old}}[\log\, p(t,\omega|\alpha,\beta)]}{\partial\beta} = 0 \tag{51}$$

Then we calculate their second order derivation:

$$\frac{\partial^2 \mathop{\mathbb{E}}_{Z|X,\vartheta^{old}}[\log\, p(t,\omega|\alpha,\beta)]}{\partial\alpha^2} = -\frac{M\alpha^2}{2} < 0 \tag{52}$$
$$\frac{\partial^2 \mathop{\mathbb{E}}_{Z|X,\vartheta^{old}}[\log\, p(t,\omega|\alpha,\beta)]}{\partial\beta^2} = -\frac{N\beta^2}{2} < 0 \tag{53}$$

So when the first order derivation equal to zero,they get the maximum.With simple mathematic operation:

$$\frac{M}{2\alpha} - \frac{\mathbb{E}[\omega^T\omega]}{2} = 0 \tag{54}$$

$$\frac{N}{2\beta} - \frac{\sum_{n=1}^{N} \mathbb{E}[(t_n - \omega^T \phi_n)^2]}{2} = 0 \tag{55}$$

So we can get:

$$\alpha = \frac{M}{\mathbb{E}[\omega^T \omega]} \tag{56}$$

$$\beta = \frac{N}{\sum_{n=1}^{N} \mathbb{E}[(t_n - \omega^T \phi_n)^2]} \tag{57}$$

Answer to Question 3.6

For the first equation, since the posterior probability of the Gaussian distribution is still the Gaussian distribution. The μ of the ω is m_N and the Σ is S_N. Thus we directly use hint(2):

$$\mathbb{E}[\omega^T \omega] = tr(\mu^T \mu) + tr(\Sigma) \tag{58}$$

$$= m_N^T m_N + tr(S_N) \tag{59}$$

For the second equation:

$$\begin{aligned}
&\sum_{n=1}^{N} \mathbb{E}[(t_n - \omega^T \phi_n)^2] \\
&= \mathbb{E}[(t - \Phi\omega)^T (t - \Phi\omega)] \\
&= \mathbb{E}[t^T t - \Phi^T \omega t - t\omega^T \Phi + \omega^T \Phi^T \Phi \omega]
\end{aligned}$$

Using (3.20) and (3.21), we can obtain by simplification:

$$\begin{aligned}
&t^T t - \omega^T \Phi^T t - t^T \Phi \omega + \omega^T \Phi^T \Phi \omega \\
&= t^T t - m_N^T \Phi^T t - t^T \Phi m_N + tr(m_N^T \Phi^T \Phi m_N) + tr(\Phi^T \Phi S_N)
\end{aligned} \tag{60}$$

Since, $m_N \in R^M$, $\Phi \in R^{N \times M}$, thus $m_N^T \Phi^T \Phi m_N \in R$ and thus $tr(m_N^T \Phi^T \Phi m_N) = m_N^T \Phi^T \Phi m_N$. And we can then deduce that:

$$= t^T t - m_N^T \Phi^T t - t^T \Phi m_N + m_N^T \Phi^T \Phi m_N + tr(\Phi^T \Phi S_N) \tag{61}$$

$$= (t - \Phi m_N)^T (t - \Phi m_N) + tr(\Phi^T \Phi S_N) \tag{62}$$

$$= ||t - \Phi m_N||_2^2 + tr(\Phi^T \Phi S_N) \tag{63}$$

Thus, we have shown the equations hold.

Answer to Question 3.7

See code.