Homework IV – Group 017

(ist1106992, ist1106154)

I. Pen-and-paper

1.

$$P(K=1|x_1) = \frac{0.015}{0.015+0.031} = 0.326$$

$$P(K=2|x_1) = \frac{0.031}{0.015+0.031} = 0.674$$

x_2:

$$\text{Posterior}(K=1|x_2) = P(x_2|K=1)P(K=1) =$$

$$= 0.5 \times N(x_2|\mu_1, \Sigma_1) = 0.5 \times N\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} \middle| \begin{bmatrix} 2 \\ -1 \end{bmatrix} ; \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}\right) =$$

$$= \frac{1}{2} \times \frac{1}{2\pi\sqrt{15}} e^{-\frac{1}{2}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \end{bmatrix}\right)^T \cdot \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \end{bmatrix}\right)} =$$

$$= \frac{1}{4\pi\sqrt{15}} e^{-\frac{1}{2}[-2 \ \ 3] \cdot \frac{1}{15}\begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 3 \end{bmatrix}} =$$

$$= \frac{1}{4\pi\sqrt{15}} e^{-\frac{32}{15}} = 0.0024$$

$$\text{Posterior}(K=2|x_2) = P(x_2|K=2)P(K=2) =$$

$$= 0.5 \times N(x_2|\mu_2, \Sigma_2) = 0.5 \times N\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} \middle| \begin{bmatrix} 1 \\ 1 \end{bmatrix} ; \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)$$

$$= \frac{1}{8\pi} \times e^{-\frac{1}{2}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)^T \cdot \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)} =$$

$$= \frac{1}{8\pi} \times e^{-\frac{1}{2}[-1 \ \ 1] \cdot \frac{1}{4}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix}} = \frac{1}{8\pi} \times e^{-\frac{1}{2}}$$

$$= 0.0241$$

$$P(K=1|x_2) = \frac{0.0024}{0.0024+0.0241} = 0.091$$

$$P(K=2|x_2) = \frac{0.0241}{0.0024+0.0241} = 0.909$$

$\underline{x_3:}$

$$\text{Posterior}(K=1 \mid x_3) = P(x_3 \mid K=1) P(K=1) =$$

$$= 0.5 \times N(x_3 \mid \mu_1, \Sigma_1) = 0.5 \times N\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} \middle| \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}\right)$$

$$= \frac{1}{2} \times \frac{1}{2\pi\sqrt{15}} \times e^{-\frac{1}{2}\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \end{bmatrix}\right)^T \cdot \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \end{bmatrix}\right)}$$

$$= \frac{1}{4\pi\sqrt{15}} \times e^{-\frac{1}{2}[1 \; 0] \cdot \frac{1}{15}\begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}} = \frac{1}{4\pi\sqrt{15}} \times e^{-\frac{2}{15}}$$

$$= 0.018$$

$$\text{Posterior}(K=2 \mid x_3) = P(x_3 \mid K=2) P(K=2) =$$

$$= 0.5 \times N(x_3 \mid \mu_2, \Sigma_2) = 0.5 \times N\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} \middle| \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right) =$$

$$= \frac{1}{8\pi} e^{-\frac{1}{2}\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)^T \cdot \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)} =$$

$$= \frac{1}{8\pi} e^{-\frac{1}{2}[2 \; -2] \cdot \frac{1}{4}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -2 \end{bmatrix}} = \frac{1}{8\pi} e^{-2} =$$

$$= 0.0054$$

$$P(K=1 \mid x_3) = \frac{0.018}{0.018 + 0.0054} = 0.769$$

$$P(K=2 \mid x_3) = \frac{0.0054}{0.018 + 0.0054} = 0.231$$

$\underline{\text{M-step}}$

$$P(K=1) = \frac{\sum_{i=1}^{3} P(K=1 \mid x_i)}{3} = \frac{0.326 + 0.091 + 0.769}{3}$$

$$= 0.395$$

$$P(K=2) = \frac{\sum_{i=1}^{3} P(K=2|x_i)}{3} = \frac{0.674 + 0.909 + 0.231}{3} =$$

$$= 0.6051$$

$$\mu_1 = \frac{\sum_{i=1}^{3} P(K=1|x_i) \cdot x_i}{\sum_{i=1}^{3} P(K=1|x_i)} = \frac{0.326 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.091 \begin{bmatrix} 0 \\ 2 \end{bmatrix} + 0.769 \begin{bmatrix} 3 \\ -1 \end{bmatrix}}{0.326 + 0.091 + 0.769}$$

$$= \begin{bmatrix} 2.22 \\ -0.495 \end{bmatrix}$$

$$\mu_2 = \frac{\sum_{i=1}^{3} P(K=2|x_i) \cdot x_i}{\sum_{i=1}^{3} P(K=2|x_i)} = \frac{0.674 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.909 \begin{bmatrix} 0 \\ 2 \end{bmatrix} + 0.231 \begin{bmatrix} 3 \\ -1 \end{bmatrix}}{0.674 + 0.909 + 0.231} =$$

$$= \begin{bmatrix} 0.754 \\ 0.875 \end{bmatrix} \quad P(K=1|x_i)$$

$$\Sigma_1 = \frac{1}{N_1} \sum_{i=1}^{3} \sqrt{} \cdot (x_i - \mu_1) \cdot (x_i - \mu_1)^T =$$

$$= \frac{0.326 \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2.22 \\ -0.495 \end{bmatrix} \right) \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2.22 \\ -0.495 \end{bmatrix} \right)^T + 0.091 \left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 2.22 \\ -0.48 \end{bmatrix} \right)}{\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 2.22 \\ -0.495 \end{bmatrix} \right)^T + 0.769 \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 2.22 \\ -0.495 \end{bmatrix} \right) \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 2.22 \\ -0.495 \end{bmatrix} \right)^T}{0.326 + 0.091 + 0.769}$$

$$= \frac{0.326 \begin{bmatrix} -1.22 \\ 0.495 \end{bmatrix} \begin{bmatrix} -1.22 & 0.495 \end{bmatrix} + 0.091 \begin{bmatrix} -2.22 \\ 2.495 \end{bmatrix} \begin{bmatrix} -2.22 & 2.495 \end{bmatrix} +}{}$$
$$\frac{+ 0.769 \begin{bmatrix} 0.78 \\ -0.505 \end{bmatrix} \begin{bmatrix} 0.78 & -0.505 \end{bmatrix}}{0.326 + 0.091 + 0.769}$$

$$= \begin{bmatrix} 1.182 & -0.846 \\ -0.846 & 0.710 \end{bmatrix}$$

$$\Sigma_2 = \frac{1}{N_2} \sum_{i=1}^{\cdot} P(K=2|x_i) \cdot (x_i - \mu_2) \cdot (x_i - \mu_2)^T =$$

$$0.674 \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.754 \\ 0.875 \end{bmatrix} \right) \cdot \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.754 \\ 0.875 \end{bmatrix} \right)^T + 0.909$$

$$\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.754 \\ 0.875 \end{bmatrix} \right) \left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.754 \\ 0.875 \end{bmatrix} \right)^T + 0.231 \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0.754 \\ 0.875 \end{bmatrix} \right)$$

$$\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0.754 \\ 0.875 \end{bmatrix} \right)^T$$

$$= \frac{}{0.674 + 0.909 + 0.231} =$$

$$= \frac{0.674 \begin{bmatrix} 0.246 \\ -0.875 \end{bmatrix} [0.246 \ -0.875] + 0.909 \begin{bmatrix} -0.754 \\ 1.125 \end{bmatrix} [-0.754 \ 1.125]}{0.674 + 0.909 + 0.231}$$

$$\frac{+ 0.231 \begin{bmatrix} 2.246 \\ -1.875 \end{bmatrix} [2.246 \ -1.875]}{}$$

$$= \begin{bmatrix} 0.949 & -1.041 \\ -1.041 & 1.366 \end{bmatrix}$$

2ª etapa:

E-step

x_1:

$$Posterior(K=1|x_1) = P(x_1|K=1) \, P(K=1) =$$

$$= 0.395 \, N(x_1|\mu_1, \Sigma_1) =$$

$$= 0.395 \, N \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \Big| \begin{bmatrix} 2.22 \\ -0.445 \end{bmatrix} \cdot \begin{bmatrix} 1.182 & -0.846 \\ -0.846 & 0.710 \end{bmatrix} \right) =$$

$$= 0.395 \times \frac{1}{2\pi \sqrt{0.124}} \times e^{-\frac{1}{2} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2.22 \\ -0.445 \end{bmatrix} \right)^T \cdot \begin{bmatrix} 1.182 & -0.846 \\ -0.846 & 0.710 \end{bmatrix}^{-1} }$$

$$\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2.22 \\ -0.445 \end{bmatrix} \right) =$$

$$= 0{,}395 \times \frac{1}{2\pi\sqrt{0{,}124}} \times e^{-\frac{1}{2}\left[-1{,}22 \;\; 0{,}495\right] \cdot \begin{bmatrix} 5{,}749 & 6{,}85 \\ 6{,}85 & 9{,}571 \end{bmatrix} \cdot \begin{bmatrix} -1{,}22 \\ 0{,}495 \end{bmatrix}}$$

$$= 0{,}395 \times \frac{1}{2\pi\sqrt{0{,}124}} \times e^{-1{,}314} = 0{,}048$$

$$\text{Posterior}(K=2\,|\,x_1) = P(x_1\,|\,K=2)\,P(K=2) =$$

$$= 0{,}605 \times N(x_1\,|\,\mu_2, \xi_2) =$$

$$= 0{,}605 \times N\left(\begin{bmatrix}1\\0\end{bmatrix}\Big|\begin{bmatrix}0{,}754\\0{,}875\end{bmatrix} \cdot \begin{bmatrix}0{,}449 & -1{,}041\\-1{,}041 & 1{,}366\end{bmatrix}\right) =$$

$$= 0{,}605 \times \frac{1}{2\pi\sqrt{0{,}213}} \times e^{-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}0{,}754\\0{,}875\end{bmatrix}\right)^T \cdot \begin{bmatrix}0{,}449 & -1{,}041\\-1{,}041 & 1{,}366\end{bmatrix}^{-1}} \cdot \; \text{④}$$

$$= \frac{0{,}605}{2\pi\sqrt{0{,}213}} \times e^{-\frac{1}{2}\left[0{,}246 \;\; -0{,}875\right] \cdot \begin{bmatrix}6{,}424 & 4{,}893\\4{,}895 & 4{,}463\end{bmatrix} \cdot \begin{bmatrix}0{,}246\\-0{,}875\end{bmatrix}} =$$

$$= \frac{0{,}605}{2\pi\sqrt{0{,}213}} \times e^{-0{,}899} = 0{,}089$$

$$P(K=1\,|\,x_1) = \frac{0{,}048}{0{,}048+0{,}089} = 0{,}35$$

$$P(K=2\,|\,x_1) = \frac{0{,}089}{0{,}048+0{,}089} = 0{,}65$$

$$\underline{x_2:}$$

$$\text{Posterior}(K=1\,|\,x_2) = P(x_2\,|\,K=1)\,P(K=1) =$$

$$= 0{,}395\,N(x_2\,|\,\mu_1, \xi_1) = 0{,}395\,N\left(\begin{bmatrix}0\\2\end{bmatrix}\Big|\begin{bmatrix}2{,}22\\0{,}495\end{bmatrix} \cdot \begin{bmatrix}1{,}182 & -0{,}846\\-0{,}846 & 0{,}71\end{bmatrix}\right) =$$

$$= \frac{0{,}395}{2\pi\sqrt{0{,}124}} \times e^{-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}2{,}22\\0{,}495\end{bmatrix}\right)^T \cdot \begin{bmatrix}1{,}182 & -0{,}846\\-0{,}846 & 0{,}71\end{bmatrix}^{-1} \cdot \left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}2{,}22\\-0{,}495\end{bmatrix}\right)} =$$

$$= \frac{0.395}{2\pi\sqrt{0.124}} \times \ell^{-\frac{1}{2}[-2.22 \; 2.495] \cdot \begin{bmatrix} 5.744 & 6.85 \\ 6.85 & 4.571 \end{bmatrix} \cdot \begin{bmatrix} -2.22 \\ 2.495 \end{bmatrix}} =$$

$$= \frac{0.395}{2\pi\sqrt{0.124}} \times \ell^{-6.015} = 4.359 \times 10^{-4}$$

Posterior $(K=2|x_2) = P(x_2|K=2) P(K=2) =$

$$= 0.605 \, N(x_2|\mu_2, \Sigma_2) = 0.605 \, N\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} \Big| \begin{bmatrix} 0.754 \\ 0.875 \end{bmatrix}; \begin{bmatrix} 0.944 & -1.041 \\ -1.041 & 1.366 \end{bmatrix}\right)$$

$$= \frac{0.605}{2\pi\sqrt{0.213}} \times \ell^{-\frac{1}{2}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.754 \\ 0.875 \end{bmatrix}\right)^T \cdot \begin{bmatrix} 0.944 & -1.041 \\ -1.041 & 1.366 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.754 \\ 0.875 \end{bmatrix}\right)}$$

$$= \frac{0.605}{2\pi\sqrt{0.213}} \times \ell^{-\frac{1}{2}[-0.754 \; 1.125] \cdot \begin{bmatrix} 6.424 & 4.895 \\ 4.895 & 4.463 \end{bmatrix} \cdot \begin{bmatrix} -0.754 \\ 1.125 \end{bmatrix}} =$$

$$= \frac{0.605}{2\pi\sqrt{0.213}} \times \ell^{-0.498} = 0.127$$

$$P(K=1|x_2) = \frac{4.359\times10^{-4}}{4.359\times10^{-4}+0.127} = 0.0034$$

$$P(K=2|x_2) = \frac{0.127}{0.127+4.359\times10^{-4}} = 0.9966$$

$\underline{x_3:}$

Posterior $(K=1|x_3) = P(x_3|K=1) P(K=1)$

$$= 0.395 \, N(x_3|\mu_1, \Sigma_1) = 0.395 \, N\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} \Big| \begin{bmatrix} 2.22 \\ -0.495 \end{bmatrix}; \begin{bmatrix} 1.182 & -0.846 \\ -0.846 & 0.770 \end{bmatrix}\right)$$

$$= \frac{0.395}{2\pi\sqrt{0.124}} \times \ell^{-\frac{1}{2}\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 2.22 \\ -0.495 \end{bmatrix}\right)^T \cdot \begin{bmatrix} 1.182 & -0.846 \\ -0.846 & 0.71 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 2.22 \\ -0.495 \end{bmatrix}\right)}$$

$$= \frac{0.395}{2\pi\sqrt{0.124}} \times \ell^{-\frac{1}{2}[0.78 \; -0.505] \cdot \begin{bmatrix} 5.744 & 6.85 \\ 6.85 & 4.571 \end{bmatrix} \cdot \begin{bmatrix} 0.78 \\ -0.505 \end{bmatrix}}$$

$$= \frac{0,395}{2\pi\sqrt{0,124}} \times e^{-0,271} = 0,136$$

Posterior $(K=2|x_3) = P(x_3|K=2)\,P(K=2) =$

$$= 0,605\,N(x_3|\mu_2,\xi_2) = 0,605\,N\left(\begin{bmatrix}3\\-1\end{bmatrix}\middle|\begin{bmatrix}0,754\\0,875\end{bmatrix}, \begin{bmatrix}0,949 & -1,041\\-1,041 & 1,366\end{bmatrix}\right)$$

$$= \frac{0,605}{2\pi\sqrt{0,213}} \times e^{-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}0,754\\0,875\end{bmatrix}\right)^{T} \cdot \begin{bmatrix}0,949 & -1,041\\-1,041 & 1,366\end{bmatrix}^{-1} \cdot \left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}0,7\\4,87\end{bmatrix}\right)}$$

$$= \frac{0,605}{2\pi\sqrt{0,213}} \times e^{-\frac{1}{2}[2,246\; -1,875]\cdot\begin{bmatrix}6,424 & 4,895\\4,895 & 4,463\end{bmatrix}\cdot\begin{bmatrix}2,246\\-1,875\end{bmatrix}} =$$

$$= \frac{0,605}{2\pi\sqrt{0,213}} \times e^{-3,434} = 0,0067$$

$$P(K=1|x_3) = \frac{0,136}{0,136+0,0067} = 0,953$$

$$P(K=2|x_3) = \frac{0,0067}{0,136+0,0067} = 0,047$$

M-step

$$P(K=1) = \frac{\sum\limits_{i=1}^{3} P(K=1|x_i)}{3} = \frac{0,953+0,0034+0,35}{3} = 0,4355$$

$$P(K=2) = \frac{\sum\limits_{i=1}^{3} P(K=1|x_i)}{3} = \frac{0,047+0,9966+0,65}{3} = 0,5645$$

$$\mu_1 = \frac{\sum_{i=1}^{3} P(K=1|x_i) \cdot x_i}{\sum_{i=1}^{3} P(K=1|x_i)} = \frac{0,953\begin{bmatrix}3\\-1\end{bmatrix} + 0,0034\begin{bmatrix}0\\2\end{bmatrix} + 0,35\begin{bmatrix}1\\0\end{bmatrix}}{0,953 + 0,0034 + 0,35}$$

$$= \begin{bmatrix}2,456\\-0,724\end{bmatrix}$$

$$\mu_2 = \frac{\sum_{i=1}^{3} P(K=2|x_i) \cdot x_i}{\sum_{i=1}^{3} P(K=2|x_i)} = \frac{0,047\begin{bmatrix}3\\-1\end{bmatrix} + 0,9966\begin{bmatrix}0\\2\end{bmatrix} + 0,65\begin{bmatrix}1\\0\end{bmatrix}}{0,047 + 0,9966 + 0,65}$$

$$= \begin{bmatrix}0,467\\1,149\end{bmatrix}$$

$$\Sigma_1 = \frac{1}{N_1}\sum_{i=1}^{3} P(K=1|x_i) \cdot (x_i - \mu_1) \cdot (x_i - \mu_1)^T =$$

$$0,953 \cdot \left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2,456\\-0,724\end{bmatrix}\right) \cdot \left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2,456\\-0,724\end{bmatrix}\right)^T + 0,0034$$

$$\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}3,456\\-0,724\end{bmatrix}\right) \cdot \left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2,456\\-6,724\end{bmatrix}\right)^T + 0,35 \cdot \left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2,456\\-6,724\end{bmatrix}\right)$$

$$\cdot \left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2,456\\-6,724\end{bmatrix}\right)^T$$

$$= \frac{0,953 + 0,0034 + 0,35}{}$$

$$= \frac{0,953 \cdot \begin{bmatrix}0,544\\-0,276\end{bmatrix} \cdot \begin{bmatrix}0,544 & -0,276\end{bmatrix} + 0,0034\begin{bmatrix}-2,456\\2,724\end{bmatrix}\begin{bmatrix}-2,456 & 2,724\end{bmatrix}}{1,3064}$$

$$\frac{+ 0,35\begin{bmatrix}-1,456\\0,724\end{bmatrix}\begin{bmatrix}-1,456 & 0,724\end{bmatrix}}{}$$

$$= \begin{bmatrix}0,8 & -0,409\\-0,409 & 0,215\end{bmatrix}$$

$$\Sigma_2 = \frac{1}{N_2} \sum_{i=1}^{3} P(K=2|x_i) \cdot (x_i - \mu_2) \cdot (x_i - \mu_2)^T =$$

$$0.65 \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.467 \\ 1.149 \end{bmatrix} \right) \cdot \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.467 \\ 1.149 \end{bmatrix} \right)^T + 0.9966 \left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.46 \\ 1.14 \end{bmatrix} \right)$$

$$\cdot \left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.467 \\ 1.149 \end{bmatrix} \right)^T + 0.047 \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0.467 \\ 1.149 \end{bmatrix} \right) \cdot \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0.467 \\ 1.149 \end{bmatrix} \right)$$

$$= \frac{0.65 \cdot \begin{bmatrix} 0.533 \\ -1.149 \end{bmatrix} \cdot \begin{bmatrix} 0.533 & -1.149 \end{bmatrix} + 0.9966 \begin{bmatrix} -0.467 \\ 0.851 \end{bmatrix} \cdot \begin{bmatrix} -0.467 & 0.85 \end{bmatrix}}{0.65 + 0.9966 + 0.047}$$

$$\frac{+0.047 \begin{bmatrix} 2.533 \\ -2.149 \end{bmatrix} \begin{bmatrix} 2.533 & -2.149 \end{bmatrix}}{1.6936} =$$

$$= \begin{bmatrix} 0.415 & -0.619 \\ -0.619 & 1.06 \end{bmatrix}$$

2.
a)

a)
$$x_1: \text{Posterior}(K=1|x_1) = P(x_1|K=1) P(K=1) =$$

$$= 0.4355 \, N(x_1|\mu_1, \Sigma_1) = 0.4355 \, N \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \Big| \begin{bmatrix} 2.456 \\ -0.724 \end{bmatrix}, \begin{bmatrix} 0.8 & -0.409 \\ -0.409 & 0.215 \end{bmatrix} \right)$$

$$= \frac{0.4355}{2\pi \sqrt{0.0047}} \, e^{-\frac{1}{2} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2.456 \\ -0.724 \end{bmatrix} \right)^T \cdot \begin{bmatrix} 0.8 & -0.409 \\ -0.409 & 0.215 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2.456 \\ -0.724 \end{bmatrix} \right)}$$

$$= \frac{0.4355}{2\pi \sqrt{0.0047}} \, e^{-\frac{1}{2} \begin{bmatrix} 1.456 & 0.724 \end{bmatrix} \cdot \begin{bmatrix} 45.561 & 86.671 \\ 86.671 & 169.577 \end{bmatrix} \cdot \begin{bmatrix} 1.456 \\ 0.724 \end{bmatrix}}$$

$$= \frac{0.4355}{2\pi \sqrt{0.0047}} \, e^{-184.087} = 1.14 \times 10^{-80}$$

Posterior$(K=2|x_1) = P(x_1|K=2)\,P(K=2) =$

$= 0,5645\, N(x_1|\mu_2, \xi_2) =$

$= 0,5645\, N\left(\begin{bmatrix}1\\0\end{bmatrix} \Big| \begin{bmatrix}0,467\\1,149\end{bmatrix}, \begin{bmatrix}0,415 & -0,619\\-0,619 & 1,06\end{bmatrix}\right) =$

$= \dfrac{0,5645}{2\pi\sqrt{0,057}}\, e^{-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0,467\\1,149\end{bmatrix}\right)^T \cdot \begin{bmatrix}0,415 & -0,619\\-0,619 & 1,06\end{bmatrix}^{-1} \cdot \left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0,\\1,1\end{bmatrix}\right)}$

$= \dfrac{0,5645}{2\pi\sqrt{0,057}}\, e^{-\frac{1}{2}[0,533\;\; -1,149] \cdot \begin{bmatrix}18,682 & 10,91\\10,91 & 7,314\end{bmatrix} \cdot \begin{bmatrix}0,533\\-1,149\end{bmatrix}}$

$= \dfrac{0,5645}{2\pi\sqrt{0,057}}\, e^{-0,8} = 0,169$

Logo, como Posterior$(K=2|x_1) >$ Posterior$(K=1|x_1)$, então $x_1 \in$ Cluster 2

$\underline{x_2:}$

Posterior$(K=1|x_2) = P(x_1|K=1)\,P(K=1)$

$= 0,4355\, N(x_2|\mu_1, \xi) =$

$= 0,4355\, N\left(\begin{bmatrix}0\\2\end{bmatrix} \Big| \begin{bmatrix}2,456\\-0,724\end{bmatrix}, \begin{bmatrix}0,8 & -0,409\\-0,409 & 0,215\end{bmatrix}\right) =$

$= \dfrac{0,4355}{2\pi\sqrt{0,0047}}\, e^{-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2,456\\-0,724\end{bmatrix}\right)^T \cdot \begin{bmatrix}0,8 & -0,409\\-0,409 & 0,215\end{bmatrix}^{-1} \cdot \left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2,456\\-0,724\end{bmatrix}\right)}$

$= \dfrac{0,4355}{2\pi\sqrt{0,0047}}\, e^{-\frac{1}{2}[-2,456\;\; 2,724] \cdot \begin{bmatrix}45,561 & 86,671\\86,671 & 169,522\end{bmatrix} \cdot \begin{bmatrix}-2,456\\2,724\end{bmatrix}} =$

$= \dfrac{0,4355}{2\pi\sqrt{0,0047}}\, e^{-186,529} = 9,474\times 10^{-82}$

$$\text{Posterior}(K=2 \mid x_2) = P(x_2 \mid K=2)\, P(K=2) =$$

$$= 0.5645 \, N(x_2 \mid \mu_2, \Sigma_2) =$$

$$= 0.5645 \, N\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} \Big| \begin{bmatrix} 0.467 \\ 1.149 \end{bmatrix}, \begin{bmatrix} 0.415 & -0.619 \\ -0.619 & 1.06 \end{bmatrix}\right) = \begin{bmatrix} 0.467 \\ 1.149 \end{bmatrix} =$$

$$= \frac{0.5645}{2\pi\sqrt{0.057}} \, e^{-\frac{1}{2}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.467 \\ 1.149 \end{bmatrix}\right)^T \begin{bmatrix} 0.415 & -0.619 \\ -0.619 & 1.06 \end{bmatrix}^{-1} \left(\begin{bmatrix} 0 \\ 2 \end{bmatrix}\right)}$$

$$= \frac{0.5645}{2\pi\sqrt{0.057}} \, e^{-\frac{1}{2}[-0.467 \;\; 0.851]\cdot\begin{bmatrix} 18.682 & 10.91 \\ 10.91 & 7.314 \end{bmatrix}\cdot\begin{bmatrix} -0.467 \\ 0.851 \end{bmatrix}}$$

$$= \frac{0.5645}{2\pi\sqrt{0.057}} \, e^{-0.35} = 0.265$$

Logo, como Posterior$(K=2 \mid x_2)$ > Posterior$(K=1 \mid x_2)$, então $x_2 \in$ Cluster 2

x_3: Posterior$(k=1|x_3) = P(x_3|k=1)P(k=1) =$

$= 0,4355 \, N(x_3|\mu_1, \Sigma_1) = 0,4355 \, N\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix}\begin{bmatrix} 2,456 \\ 4,704 \end{bmatrix}, \begin{bmatrix} 0,8 & -0,409 \\ -0,409 & \end{bmatrix}\right)$

$= \dfrac{0,4355}{2\pi\sqrt{0,0047}} \, e^{-\frac{1}{2}\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 2,456 \\ 4,704 \end{bmatrix}\right)^T \begin{bmatrix} 0,8 & -0,409 \\ -0,409 & 0,715 \end{bmatrix}^{-1} \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 2,456 \\ -0,12 \end{bmatrix}\right)}$

$= \dfrac{0,4355}{2\pi\sqrt{0,0047}} \, e^{-\frac{1}{2}\begin{bmatrix} 0,544 & -0,12 \end{bmatrix} \cdot \begin{bmatrix} 45,561 & -86,611 \\ 86,611 & 169,517 \end{bmatrix} \cdot \begin{bmatrix} 0,544 \\ -0,716 \end{bmatrix}}$

$= \dfrac{0,4355}{2\pi\sqrt{0,0047}} \, e^{-0,185} = 0,84$

Posterior$(k=2|x_3) = P(x_3|k=2)P(k=2) =$

$= 0,5645 \, N(x_3|\mu_2, \Sigma_2) =$

$= 0,5645 \, N\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix}\begin{bmatrix} 0,467 \\ 1,149 \end{bmatrix}, \begin{bmatrix} 0,415 & -0,619 \\ -0,619 & 1,06 \end{bmatrix}\right) =$

$= \dfrac{0,5645}{2\pi\sqrt{0,057}} \, e^{-\frac{1}{2}\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0,467 \\ 1,149 \end{bmatrix}\right)^T \begin{bmatrix} 0,415 & -0,619 \\ -0,619 & 1,06 \end{bmatrix}^{-1} \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0,467 \\ 1,149 \end{bmatrix}\right)}$

$= \dfrac{0,5645}{2\pi\sqrt{0,057}} \, e^{-\frac{1}{2}\begin{bmatrix} 2,533 & -2,149 \end{bmatrix} \cdot \begin{bmatrix} 18,682 & 10,91 \\ 10,91 & 7,314 \end{bmatrix} \cdot \begin{bmatrix} 2,533 \\ -2,149 \end{bmatrix}}$

$= \dfrac{0,5645}{2\pi\sqrt{0,057}} \, e^{-17,437} = 1,006\times 10^{-8}$

Logo, como Posterior$(k=1|x_3)$ > Posterior$(k=2|x_3)$

então $x_3 \in$ Cluster 1

Logo, Cluster 1 = $\{x_3\}$ e Cluster 2 = $\{x_1, x_2\}$

b)

b)

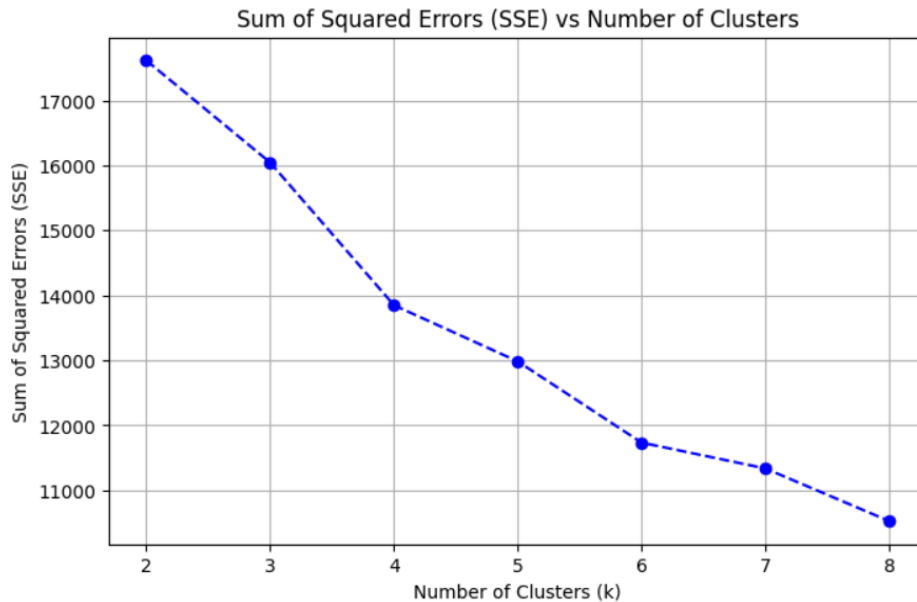$$S(\text{Cluster } 2) = \frac{S(x_1) + S(x_2)}{2}$$

x_1:

$a = d(x_1, x_2) = \sqrt{(1-0)^2 + (0-2)^2} = \sqrt{5}$

$b = d(x_1, x_3) = \sqrt{(1-3)^2 + (0+1)^2} = \sqrt{5}$

$S(x_1) = 1 - \frac{a}{b} = 1 - \frac{\sqrt{5}}{\sqrt{5}} = 0$

x_2:

$a = d(x_2, x_1) = \sqrt{5}$

$b = d(x_2, x_3) = \sqrt{(0-3)^2 + (2+1)^2} = 3\sqrt{2}$

$S(x_2) = 1 - \frac{\sqrt{5}}{3\sqrt{2}} = 0,473$

Logo, $S(\text{cluster } 2) = \frac{0,473}{2} = 0,2365$

II. Programming and critical analysis

1.
a)

b)

    The plot generated above, shows the SSE or inertia as a function of the number of clusters (k) in a K-Means clustering model.

    Generally, and as seen in the plot, the SSE decreases as the number of clusters increases because the points are closer to their respective centroids resulting in smaller and more exact clusters.

According to the knee/elbow method, there is a point where SSE reductions slow down, meaning that adding more cluster doesn't benefit the model significantly. For this example, that happens for a k value of 4, because for smaller k values the function's slope is around -2000 and for larger values is around -1000. This increase in the slope's value means that the difference in the SSE gets less significant as doesn't justify adding more clusters to the model.

    Given this, 4 is the optimal number of clusters for this model because it provides both a low value for SSE and a well fitted model (a lower number of clusters means that the model doesn't categorize its data well enough leading to underfitting and a higher number of clusters leads to clusters that are too specific and not meaningful which can cause overfitting).

c)

    The k-means and k-modes models are used in almost the same way but differ in the most appropriate type of data for each one. The k-means operates better for numerical features while k-modes works better for categorical data.

    This happens because k-means optimizes clustering by minimizing the variance between data points using Euclidean distance, while k-modes uses the most frequent category to determine cluster centroids.

    Given this, after analyzing the accounts.csv dataset, we noticed that many of its features (such as job, marital, education, housing, loan, etc.) are categorical which matches better with the k-modes model.

Having used pd.get_dummies() in exercise 1, also means that the type of features in our dataset doesn't match the k-means model appropriately (because this function is used to change categorical variables to numerical at the expense of the model's simplicity as it gets more complex and the relations between features may get distorted and give worst performance to the overall model).

    In conclusion, the model must be chosen with the dataset's features types in mind and in this case k-modes makes more sense.

2.

a)

 When applying Principal Component Analysis (PCA) the data is transformed into principal components that capture the data's variance.

 With the results from the previous exercise, the top 2 components only capture variances of 11.679% and 11.076% meaning that only 22.76% of the global variance is captured in total. Given this, we conclude that the dataset has high dimensionality which means that important patterns and structures could only be captured by adding more components to the analysis. Although many information is lost in this process the data is simplified and easier to interpret as we'll see in the next exercise.

 In conclusion, 2 components is not enough to get a comprehensive view of the dataset (for this to occur the total variability should be around 80% to 90%) but, to get results that can be easily converted into plots and interpreted 2 or 3 is the maximum number of components that can be analyzed.

b)
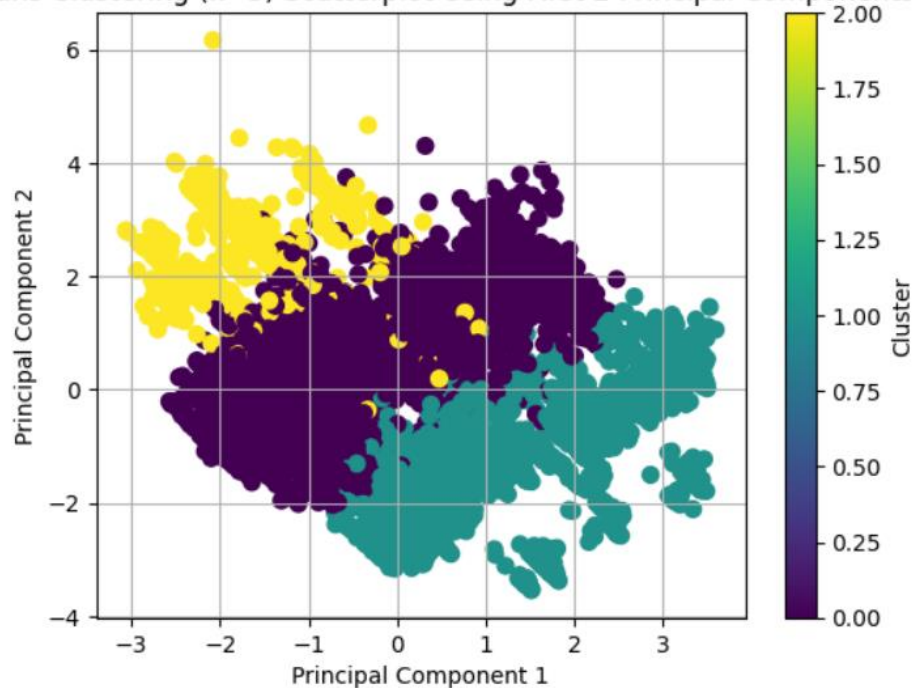
 The scatterplot above shows the clusters generated when analysing the 2 components mentioned in the previous exercise.

 Upon looking at the plot, we can see that the points from each cluster are grouped fairly well inside the cluster that they belong to. However, we see that there is some overlap in the clusters especially with the purple one (overlaps both the yellow and teal clusters). This means that there isn't a distinct separation between them, due to only having 2 principal components to provide variance to the analysis.

 Given this, we assessed that the 22.76% of total variance achieved with the first 2 principal components is not enough to fully separate them from eachother but it's still possible to distinguish the cluster that they belong to so the clusters are considered to be separated enough.

c)
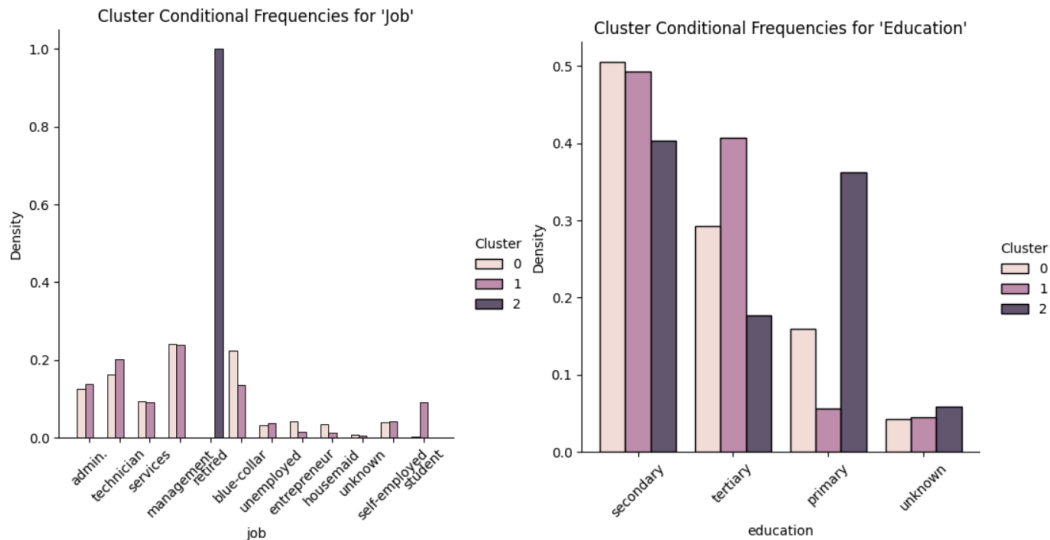
When looking at the Job plot, we can split our analysis into the 3 clusters. The most significant bar in the plot belongs to cluster 2 and it means that the retired individuals dominate this cluster.

The other two clusters are distributed more equally between the remaining categories, especially "admin", "technician", "blue-collar", "services" and "management". Cluster 1, has a majority of the observations for the first two categories and for the "student" part of the sample. Meanwhile, cluster 0 represent better the other categories as its conditional frequency is higher than the other clusters'.

The categories "unemployed", "entrepreneur", "housemaid", "unknown" and "self-employed" have lower conditional frequencies for all clusters which means that these categories are less likely to subscribe a term deposit.

When looking at the Education plot and analysing the cluster 0, we notice that it has the highest proportion of individuals with "secondary" education and a significant part of people with "tertiary" education and the conditional frequency is lower for the other two categories. Similarly, cluster 1 also has big relevance for "secondary" and "tertiary" educations with emphasis on the "tertiary" (where its smaller than "secondary" but the largest inside the category by a large difference). Meanwhile, cluster 2 has the lower conditional frequency of the three for the first two categories (that are dominated by the other two clusters) but contains most of the observations of individuals with "primary" education.

The "unknown" category has a small number of observations for all the clusters and its less meaningful for this analysis.

Given this, cluster 2 represents mainly retired individuals with primary or unknown education, cluster 0 captures a higher proportion of working professionals ("blue-collar", "services" and "management") with secondary or tertiary education while cluster 1 is more diverse, with the majority of "admin" and "services" jobs and individuals with secondary and tertiary education.

END