Homework IV: I) Clustering:

$X = \{x_1 = \begin{pmatrix}2\\0\end{pmatrix}, x_2 = \begin{pmatrix}1\\0\end{pmatrix}, x_3 = \begin{pmatrix}0\\1\end{pmatrix}\}$; $K=2$; $\mu_1 = \begin{pmatrix}1\\0\end{pmatrix}$; $\mu_2 = \begin{pmatrix}0\\1\end{pmatrix}$

$\Sigma_1 = \begin{pmatrix}1&0\\0&1\end{pmatrix}$; $\Sigma_2 = \begin{pmatrix}1&0\\0&1\end{pmatrix}$; $\pi_1 = p(c_1=1) = 0.5$; $\pi_2 = p(c_2=1) = 0.5$

$\boxed{\text{Step 1:}}$ Expectation step: $\rightarrow$ For x_1: Cluster $C=1$:

Prior: $p(C=1) = 0.5$; Likelihood: $p(x_1|C=1) = \mathcal{N}\left(\mu_1 = \begin{pmatrix}1\\0\end{pmatrix}, \Sigma_1 = \begin{pmatrix}1&0\\0&1\end{pmatrix}\right) =$

$$= \frac{\exp\left(-\frac{1}{2}(x_1-\mu_1)^T(\Sigma_1)^{-1}(x_1-\mu_1)\right)}{\sqrt{(2\pi)^K \det(\Sigma_1)}} = \frac{\exp\left(-\frac{1}{2}\left[\begin{pmatrix}2\\0\end{pmatrix}-\begin{pmatrix}1\\0\end{pmatrix}\right]^T\begin{pmatrix}1&0\\0&1\end{pmatrix}^{-1}\left[\begin{pmatrix}2\\0\end{pmatrix}-\begin{pmatrix}1\\0\end{pmatrix}\right]\right)}{\sqrt{(2\pi)^2\det\begin{pmatrix}1&0\\0&1\end{pmatrix}}} =$$

$$= \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}1\\0\end{pmatrix}^T\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\right)}{2\pi} = \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}1&0\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\right)}{2\pi} = \frac{\exp\left(\begin{pmatrix}-\frac{1}{2}&0\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\right)}{2\pi} =$$

$$= \frac{\exp\left(\begin{pmatrix}-\frac{1}{2}&0\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\right)}{2\pi} = \frac{\exp\left(-\frac{1}{2}\right)}{2\pi} \approx 0.0965$$

Joint probability: $p(C=1, x_1) = p(C=1)\,p(x_1|C=1) = 0.5 \times 0.0965 =$
$= 0.04825$

Cluster $C=2$: Prior: $p(C=2) = 0.5$

Likelihood: $p(x_1|C=2) = \mathcal{N}\left(\mu_2 = \begin{pmatrix}0\\1\end{pmatrix}, \Sigma_2 = \begin{pmatrix}1&0\\0&1\end{pmatrix}\right) =$

$$= \frac{\exp\left(-\frac{1}{2}(x_1-\mu_2)^T(\Sigma_2)^{-1}(x_1-\mu_2)\right)}{\sqrt{(2\pi)^K\det(\Sigma_2)}} = \frac{\exp\left(-\frac{1}{2}\left[\begin{pmatrix}2\\0\end{pmatrix}-\begin{pmatrix}0\\1\end{pmatrix}\right]^T\begin{pmatrix}1&0\\0&1\end{pmatrix}^{-1}\left[\begin{pmatrix}2\\0\end{pmatrix}-\begin{pmatrix}0\\1\end{pmatrix}\right]\right)}{\sqrt{(2\pi)^2\det\begin{pmatrix}1&0\\0&1\end{pmatrix}}} =$$

$$= \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}2\\-1\end{pmatrix}^T\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}2\\-1\end{pmatrix}\right)}{2\pi} = \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}2&-1\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}2\\-1\end{pmatrix}\right)}{2\pi} =$$

$$= \frac{\exp\left(\begin{pmatrix}-1&\frac{1}{2}\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}2\\-1\end{pmatrix}\right)}{2\pi} = \frac{\exp\left(\begin{pmatrix}-1&\frac{1}{2}\end{pmatrix}\begin{pmatrix}2\\-1\end{pmatrix}\right)}{2\pi} = \frac{\exp\left(-\frac{5}{2}\right)}{2\pi} \approx 0.0131$$

Joint probability: $p(C=2, x_1) = p(C=2)\,p(x_1|C=2) = 0.5 \times 0.0131 =$
$= 0.00655$

$C=1$: $p(C=1|x_1) = \frac{p(C=1,x_1)}{p(C=1,x_1)+p(C=2,x_1)} = \frac{0.04825}{0.04825+0.00655} \approx$
≈ 0.8805

$C=2$: $p(C=2|x_1) = \frac{p(C=2,x_1)}{p(C=1,x_1)+p(C=2,x_1)} = \frac{0.00655}{0.04825+0.00655} \approx 0.1195$

$\rightarrow$ For x_2: Cluster $C=1$: Prior: $p(C=1) = 0.5$

Likelihood: $p(x_2|C=1) = \mathcal{N}\left(\mu_1 = \begin{pmatrix}1\\0\end{pmatrix}, \Sigma_1 = \begin{pmatrix}1&0\\0&1\end{pmatrix}\right) = \frac{\exp\left(-\frac{1}{2}(x_2-\mu_1)^T(\Sigma_1)^{-1}(x_2-\mu_1)\right)}{\sqrt{(2\pi)^K\det(\Sigma_1)}} =$

$$= \frac{\exp\left(-\frac{1}{2}\left[\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}1\\0\end{pmatrix}\right]^T \begin{pmatrix}1&0\\0&1\end{pmatrix}^{-1}\left[\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}1\\0\end{pmatrix}\right]\right)}{\sqrt{(2\pi)^2 \det\begin{pmatrix}1&0\\0&1\end{pmatrix}}} = \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}0\\0\end{pmatrix}^T \begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}0\\0\end{pmatrix}\right)}{2\pi} =$$

$$= \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}0&0\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}0\\0\end{pmatrix}\right)}{2\pi} = \frac{\exp\left(\begin{pmatrix}0&0\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}0\\0\end{pmatrix}\right)}{2\pi} = \frac{\exp\left(\begin{pmatrix}0&0\end{pmatrix}\begin{pmatrix}0\\0\end{pmatrix}\right)}{2\pi} =$$

$$= \frac{\exp(0)}{2\pi} \approx 0,1592$$

Joint probability: $p(C=1, x_2) = p(C=1)\,p(x_2 | C=1) = 0,5 \times 0,1592 \approx$
$\approx 0,0796$

Cluster $C=2$: Prior: $p(C=2) = 0,5$

Likelihood: $p(x_2 | C=2) = \mathcal{N}\left(\mu_2 = \begin{pmatrix}0\\1\end{pmatrix}, \Sigma_2 = \begin{pmatrix}1&0\\0&1\end{pmatrix}\right) = \frac{\exp\left(-\frac{1}{2}(x_2-\mu_2)^T(\Sigma_2)^{-1}(x_2-\mu_2)\right)}{\sqrt{(2\pi)^k \det(\Sigma_2)}} =$

$$= \frac{\exp\left(-\frac{1}{2}\left[\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}0\\1\end{pmatrix}\right]^T \begin{pmatrix}1&0\\0&1\end{pmatrix}^{-1}\left[\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}0\\1\end{pmatrix}\right]\right)}{\sqrt{(2\pi)^2 \det\begin{pmatrix}1&0\\0&1\end{pmatrix}}} = \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}1\\-1\end{pmatrix}^T \begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix}\right)}{2\pi} =$$

$$= \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}1&-1\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix}\right)}{2\pi} = \frac{\exp\left(\begin{pmatrix}-\frac{1}{2}&\frac{1}{2}\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix}\right)}{2\pi} = \frac{\exp\left(\begin{pmatrix}-\frac{1}{2}&\frac{1}{2}\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix}\right)}{2\pi} =$$

$$= \frac{\exp(-1)}{2\pi} \approx 0,0585$$

Joint probability: $p(C=2, x_2) = p(C=2)\,p(x_2 | C=2) = 0,5 \times 0,0585 =$
$= 0,02925$

$C=1$: $p(C=1 | x_2) = \frac{p(C=1, x_2)}{p(C=1, x_2) + p(C=2, x_2)} = \frac{0,0796}{0,0796 + 0,02925} \approx 0,7313$

$C=2$: $p(C=2 | x_2) = \frac{p(C=2, x_2)}{p(C=1, x_2) + p(C=2, x_2)} = \frac{0,02925}{0,0796 + 0,02925} \approx 0,2687$

For x_3: Cluster $C=1$: Prior: $p(C=1) = 0,5$

Likelihood: $p(x_3 | C=1) = \mathcal{N}\left(\mu_1 = \begin{pmatrix}1\\0\end{pmatrix}, \Sigma_1 = \begin{pmatrix}1&0\\0&1\end{pmatrix}\right) =$

$$= \frac{\exp\left(-\frac{1}{2}(x_3-\mu_1)^T(\Sigma_1)^{-1}(x_3-\mu_1)\right)}{\sqrt{(2\pi)^k \det(\Sigma_1)}} =$$

$$= \frac{\exp\left(-\frac{1}{2}\left[\begin{pmatrix}0\\1\end{pmatrix}-\begin{pmatrix}1\\0\end{pmatrix}\right]^T \begin{pmatrix}1&0\\0&1\end{pmatrix}^{-1}\left[\begin{pmatrix}0\\1\end{pmatrix}-\begin{pmatrix}1\\0\end{pmatrix}\right]\right)}{\sqrt{(2\pi)^2 \det\begin{pmatrix}1&0\\0&1\end{pmatrix}}} = \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}-1\\1\end{pmatrix}^T \begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}-1\\1\end{pmatrix}\right)}{2\pi} =$$

$$= \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}-1&1\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}-1\\1\end{pmatrix}\right)}{2\pi} = \frac{\exp\left(\begin{pmatrix}\frac{1}{2}&-\frac{1}{2}\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}-1\\1\end{pmatrix}\right)}{2\pi} = \frac{\exp\left(\begin{pmatrix}\frac{1}{2}&-\frac{1}{2}\end{pmatrix}\begin{pmatrix}-1\\1\end{pmatrix}\right)}{2\pi} =$$

$$= \frac{\exp(-1)}{2\pi} \approx 0,0585$$

Joint probability: $p(C=1, x_3) = p(C=1)\,p(x_3 | C=1) = 0,5 \times 0,0585 = 0,02925$

Cluster $C=2$: Prior: $p(C=2) = 0,5$

Likelihood: $p(x_3 | C=2) = \mathcal{N}\left(\mu_2 = \begin{pmatrix}0\\1\end{pmatrix}, \Sigma_2 = \begin{pmatrix}1&0\\0&1\end{pmatrix}\right) =$

Digitalizada com CamScanner

$$= \frac{\exp\left(-\frac{1}{2}(x_3-\mu_2)^T(\Sigma_2)^{-1}(x_3-\mu_2)\right)}{\sqrt{(2\pi)^K \det(\Sigma_2)}} = \frac{\exp\left(-\frac{1}{2}\left[\binom{0}{1}-\binom{0}{1}\right]^T\begin{pmatrix}1&0\\0&1\end{pmatrix}^{-1}\left[\binom{0}{1}-\binom{0}{1}\right]\right)}{\sqrt{(2\pi)^2\det\begin{pmatrix}1&0\\0&1\end{pmatrix}}} =$$

$$= \frac{\exp\left(-\frac{1}{2}\binom{0}{0}^T\begin{pmatrix}1&0\\0&1\end{pmatrix}\binom{0}{0}\right)}{2\pi} = \frac{\exp\left(-\frac{1}{2}(0\ 0)\begin{pmatrix}1&0\\0&1\end{pmatrix}\binom{0}{0}\right)}{2\pi} = \frac{\exp\left((0\ 0)\begin{pmatrix}1&0\\0&1\end{pmatrix}\binom{0}{0}\right)}{2\pi} =$$

$$= \frac{\exp\left((0\ 0)\binom{0}{0}\right)}{2\pi} = \frac{\exp(0)}{2\pi} \approx 0.1592$$

Joint probability: $p(C=2, x_3) = p(C=2)p(x_3|C=2) = 0.5 \times 0.1592 = 0.0796$

$C=1$: $p(C=1|x_3) = \frac{p(C=1,x_3)}{p(C=1,x_3)+p(C=2,x_3)} = \frac{0.02925}{0.02925+0.0796} \approx 0.2687$

$C=2$: $p(C=2|x_3) = \frac{p(C=2,x_3)}{p(C=1,x_3)+p(C=2,x_3)} = \frac{0.0796}{0.02925+0.0796} \approx 0.7313$

Maximization Step: → For $C=1$: $\mu_1 = \frac{\sum_{n=1}^{3} p(C=1|x_n)x_n}{\sum_{n=1}^{3} p(C=1|x_n)} =$

$$= \frac{p(C=1|x_1)x_1 + p(C=1|x_2)x_2 + p(C=1|x_3)x_3}{p(C=1|x_1)+p(C=1|x_2)+p(C=1|x_3)} =$$

$$= \frac{0.8805\binom{2}{0}+0.7313\binom{1}{0}+0.2687\binom{0}{1}}{0.8805+0.7313+0.2687} = \frac{\binom{1.7610}{0}+\binom{0.7313}{0}+\binom{0}{0.2687}}{1.8805} =$$

$$= \frac{\binom{2.4923}{0.2687}}{1.8805} = \frac{1}{1.8805}\binom{2.4923}{0.2687} \approx \binom{1.3253}{0.1429}$$

$$\Sigma_{1ij} = \frac{\sum_{n=1}^{3} p(C=1|x_n)(x_{ni}-\mu_{1i})(x_{nj}-\mu_{1j})}{\sum_{n=1}^{3} p(C=1|x_n)}$$

$$\Sigma_{1_{11}} = \frac{p(C=1|x_1)(x_{1_1}-\mu_{1_1})^2 + p(C=1|x_2)(x_{2_1}-\mu_{1_1})^2 + p(C=1|x_3)(x_{3_1}-\mu_{1_1})^2}{p(C=1|x_1)+p(C=1|x_2)+p(C=1|x_3)} =$$

$$= \frac{0.8805(2-1.3253)^2+0.7313(1-1.3253)^2+0.2687(0-1.3253)^2}{0.8805+0.7313+0.2687} \approx$$

$$\approx \frac{0.4008+0.0774+0.4720}{1.8805} = \frac{0.9502}{1.8805} \approx 0.5053$$

$$\Sigma_{1_{12}} = \Sigma_{1_{21}} = \frac{1}{p(C=1|x_1)+p(C=1|x_2)+p(C=1|x_3)}\Big[p(C=1|x_1)(x_{1_1}-\mu_{1_1})(x_{1_2}-\mu_{1_2}) +$$

$$+ p(C=1|x_2)(x_{2_1}-\mu_{1_1})(x_{2_2}-\mu_{1_2}) + p(C=1|x_3)(x_{3_1}-\mu_{1_1})(x_{3_2}-\mu_{1_2})\Big] =$$

$$= \Big[0.8805(2-1.3253)(0-0.1429)+0.7313(1-1.3253)(0-0.1429)+$$

$$+0.2687(0-1.3253)(1-0.1429)\Big]\frac{1}{0.8805+0.7313+0.2687} = \frac{-0.0849+0.0340+0.3053}{1.8805} =$$

$$= \frac{-0{,}3561}{1{,}8805} \approx -0{,}1894$$

$$\Sigma_{1_{22}} = \frac{\gamma(c=1|x_1)(x_{1_2}-\mu_{1_2})^2 + \gamma(c=1|x_2)(x_{2_2}-\mu_{1_2})^2 + \gamma(c=1|x_3)(x_{3_2}-\mu_{1_2})^2}{\gamma(c=1|x_1)+\gamma(c=1|x_2)+\gamma(c=1|x_3)} =$$

$$= \frac{0{,}8805(0-0{,}1429)^2 + 0{,}7313(0-0{,}1429)^2 + 0{,}2687(1-0{,}1429)^2}{0{,}8805+0{,}7313+0{,}2687} \approx \frac{0{,}0180+0{,}0149+0{,}1974}{1{,}8805} =$$

$$= \frac{0{,}2303}{1{,}8805} \approx 0{,}1225$$

$$\Sigma_1 = \begin{pmatrix} 0{,}5053 & -0{,}1894 \\ -0{,}1894 & 0{,}1225 \end{pmatrix}$$

$$\gamma(c=1) = \frac{\sum_{n=1}^{3}\gamma(c=1|x_n)}{\sum_{l=1}^{2}\sum_{n=1}^{3}\gamma(c=l|x_n)} = \frac{\gamma(c=1|x_1)+\gamma(c=1|x_2)+\gamma(c=1|x_3)}{\gamma(c=1|x_1)+\gamma(c=1|x_2)+\gamma(c=1|x_3)+\gamma(c=2|x_1)+\gamma(c=2|x_2)+\gamma(c=2|x_3)} =$$

$$= \frac{0{,}8805+0{,}7313+0{,}2687}{0{,}8805+0{,}7313+0{,}2687+0{,}1195+0{,}2687+0{,}7313} = \frac{1{,}8805}{3} \approx 0{,}6268$$

→ For $c=2$: $\mu_2 = \frac{\sum_{n=1}^{3}\gamma(c=2|x_n)x_n}{\sum_{n=1}^{3}\gamma(c=2|x_n)} = \frac{\gamma(c=2|x_1)x_1+\gamma(c=2|x_2)x_2+\gamma(c=3|x_3)}{\gamma(c=2|x_1)+\gamma(c=2|x_2)+\gamma(c=2|x_3)} =$

$$= \frac{0{,}1195\binom{2}{0}+0{,}2687\binom{1}{0}+0{,}7313\binom{0}{1}}{0{,}1195+0{,}2687+0{,}7313} = \frac{\binom{0{,}2390}{0}+\binom{0{,}2687}{0}+\binom{0}{0{,}7313}}{1{,}1195} =$$

$$= \frac{1}{1{,}1195}\begin{pmatrix} 0{,}5077 \\ 0{,}7313 \end{pmatrix} = \begin{pmatrix} 0{,}4535 \\ 0{,}6532 \end{pmatrix}$$

$$\Sigma_{2_{ij}} = \frac{\sum_{n=1}^{3}\gamma(c=2|x_n)(x_{n_i}-\mu_{2_i})(x_{n_j}-\mu_{2_j})}{\sum_{n=1}^{3}\gamma(c=2|x_n)}$$

$$\Sigma_{2_{11}} = \frac{\gamma(c=2|x_1)(x_{1_1}-\mu_{2_1})^2+\gamma(c=2|x_2)(x_{2_1}-\mu_{2_1})^2+\gamma(c=2|x_3)(x_{3_1}-\mu_{2_1})^2}{\gamma(c=2|x_1)+\gamma(c=2|x_2)+\gamma(c=2|x_3)} =$$

$$= \frac{0{,}1195(2-0{,}4535)^2+0{,}2687(1-0{,}4535)^2+0{,}7313(0-0{,}4535)^2}{0{,}1195+0{,}2687+0{,}7313} \approx$$

$$\approx \frac{0{,}5164}{1{,}1195} \approx 0{,}4613$$

$$\Sigma_{2_{12}} = \Sigma_{2_{21}} = \left[\gamma(c=2|x_1)(x_{1_1}-\mu_{2_1})(x_{1_2}-\mu_{2_2})+\gamma(c=2|x_2)(x_{2_1}-\mu_{2_1})(x_{2_2}-\mu_{2_2}) + \gamma(c=2|x_3)(x_{3_1}-\mu_{2_1})(x_{3_2}-\mu_{2_2})\right]\frac{1}{\gamma(c=2|x_1)+\gamma(c=2|x_2)+\gamma(c=2|x_3)} =$$

$$= \frac{0{,}1195(2-0{,}4535)(0-0{,}6532)+0{,}2687(1-0{,}4535)(0-0{,}6532)+0{,}7313(0-0{,}4535)(1-0{,}6532)}{0{,}1195+0{,}2687+0{,}7313} \approx$$

$$\approx \frac{-0{,}3316}{1{,}1195} \approx -0{,}2962$$

$$\Sigma_{2_{22}} = \frac{p(c=2|x_1)(x_{1_2}-\mu_{2_2})^2 + p(c=2|x_2)(x_{2_2}-\mu_{2_2})^2 + p(c=2|x_3)(x_{3_2}-\mu_{2_2})^2}{p(c=2|x_1)+p(c=2|x_2)+p(c=2|x_3)} =$$

$$= \frac{0.1195(0-0.6532)^2 + 0.2687(0-0.6532)^2 + 0.7313(1-0.6532)^2}{0.1195+0.2687+0.7313} \approx \frac{0.2536}{1.1195} \approx$$

$$\approx 0.2265$$

$$\Sigma_2 = \begin{pmatrix} 0.4613 & -0.2962 \\ -0.2962 & 0.2265 \end{pmatrix}$$

$$p(c=2) = \frac{\sum_{n=1}^{3} p(c=2|x_n)}{\sum_{l=1}^{2}\sum_{n=1}^{3} p(c=l|x_n)} = \frac{p(c=2|x_1)+p(c=2|x_2)+p(c=2|x_3)}{p(c=1|x_1)+p(c=1|x_2)+p(c=1|x_3)+p(c=2|x_1)+p(c=2|x_2)+p(c=2|x_3)}$$

$$= \frac{0.1195+0.2687+0.7313}{0.8805+0.7313+0.2687+0.1195+0.2687+0.7313} = \frac{1.1195}{3} \approx 0.3732$$

$\boxed{\text{Step 2:}}$ Expectation step: → For $x=1$: Cluster $c=1$:

Prior: $p(c=1) = 0.6268$

Likelihood: $p(x_1|c=1) = \mathcal{N}\left(x_1 \middle| \mu_1 = \begin{pmatrix} 1.3253 \\ 0.1429 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 0.5053 & -0.1894 \\ -0.1894 & 0.1225 \end{pmatrix}\right) =$

$$= \frac{\exp\left(-\frac{1}{2}(x_1-\mu_1)^T(\Sigma_1)^{-1}(x_1-\mu_1)\right)}{\sqrt{(2\pi)^k \det(\Sigma_1)}} = \frac{\exp\left(-\frac{1}{2}\left[\begin{pmatrix}2\\0\end{pmatrix}-\begin{pmatrix}1.3253\\0.1429\end{pmatrix}\right]^T \begin{pmatrix} 0.5053 & -0.1894 \\ -0.1894 & 0.1225 \end{pmatrix}^{-1}\left[\begin{pmatrix}2\\0\end{pmatrix}-\begin{pmatrix}1.3253\\0.1429\end{pmatrix}\right]\right)}{\sqrt{(2\pi)^2 \det\begin{pmatrix} 0.5053 & -0.1894 \\ -0.1894 & 0.1225 \end{pmatrix}}} =$$

$$= \frac{\exp\left(-\frac{1}{2}(0.6747 \quad -0.1429)\begin{pmatrix} 4.7067 & 7.2771 \\ 7.2771 & 19.4145 \end{pmatrix}\begin{pmatrix} 0.6747 \\ -0.1429 \end{pmatrix}\right)}{2\pi\sqrt{0.0260}} =$$

$$= \frac{\exp\left((-0.3374 \quad 0.0714)\begin{pmatrix} 4.7067 & 7.2771 \\ 7.2771 & 19.4145 \end{pmatrix}\begin{pmatrix} 0.6747 \\ -0.1429 \end{pmatrix}\right)}{1.0131} =$$

$$= \frac{\exp\left((-1.0684 \quad -1.0691)\begin{pmatrix} 0.6747 \\ -0.1429 \end{pmatrix}\right)}{1.0131} \approx \frac{\exp(-0.5681)}{1.0131} \approx 0.5593$$

Joint Probability: $p(c=1, x_1) = p(c=1)\,p(x_1|c=1) = 0.6268 \times 0.5593 \approx 0.3506$

Cluster $c=2$: Prior: $p(c=2) = 0.3732$

Likelihood: $p(x_1|c=2) = \mathcal{N}\left(x_1 \middle| \mu_2 = \begin{pmatrix} 0.4535 \\ 0.6532 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 0.4613 & -0.2962 \\ -0.2962 & 0.2265 \end{pmatrix}\right) =$

$$= \frac{\exp\left(-\frac{1}{2}(x_1-\mu_2)^T(\Sigma_2)^{-1}(x_1-\mu_2)\right)}{\sqrt{(2\pi)^k \det(\Sigma_2)}} = \frac{\exp\left(-\frac{1}{2}\left[\begin{pmatrix}2\\0\end{pmatrix}-\begin{pmatrix}0.4535\\0.6532\end{pmatrix}\right]^T \begin{pmatrix} 0.4613 & -0.2962 \\ -0.2962 & 0.2265 \end{pmatrix}^{-1}\left[\begin{pmatrix}2\\0\end{pmatrix}-\begin{pmatrix}0.4535\\0.6532\end{pmatrix}\right]\right)}{\sqrt{(2\pi)^2 \det\begin{pmatrix} 0.4613 & -0.2962 \\ -0.2962 & 0.2265 \end{pmatrix}}} =$$

$$= \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}1,5465 & -0,6532\end{pmatrix}\begin{pmatrix}13,5224 & 17,6836\\ 17,6836 & 27,5403\end{pmatrix}\begin{pmatrix}1,5465\\ -0,6532\end{pmatrix}\right)}{2\pi\sqrt{0,0168}} \approx$$

$$\approx \frac{\exp\left(\begin{pmatrix}-0,7732 & 0,3266\end{pmatrix}\begin{pmatrix}13,5224 & 17,6836\\ 17,6836 & 27,5403\end{pmatrix}\begin{pmatrix}1,5465\\ -0,6532\end{pmatrix}\right)}{0,8144} \approx$$

$$\approx \frac{\exp\left(\begin{pmatrix}-4,6800 & -4,6783\end{pmatrix}\begin{pmatrix}1,5465\\ -0,6532\end{pmatrix}\right)}{0,8144} \approx \frac{\exp(-6,9320)}{0,8144} \approx 0,0012$$

Joint Probability: $p(C=2, x_1) = p(C=2)\,p(x_1|C=2) = 0,3732 \times 0,0012 \approx 0,0004$

$C=1$: $p(C=1|x_1) = \frac{p(C=1,x_1)}{p(C=1,x_1)+p(C=2,x_1)} = \frac{0,3506}{0,3506+0,0004} \approx 0,9989$

$C=2$: $p(C=2|x_1) = \frac{p(C=2,x_1)}{p(C=1,x_1)+p(C=2,x_1)} = \frac{0,0004}{0,3506+0,0004} \approx 0,0011$

→ For $x=2$: Cluster $C=1$: Prior: $p(C=1) = 0,6268$

Likelihood: $p(x_2|C=1) = \mathcal{N}\left(x_2 \middle| \mu_1 = \begin{pmatrix}1,3253\\ 0,1429\end{pmatrix}, \Sigma_1 = \begin{pmatrix}0,5053 & -0,1894\\ -0,1894 & 0,1225\end{pmatrix}\right) =$

$$= \frac{\exp\left(-\frac{1}{2}(x_2-\mu_1)^T(\Sigma_1)^{-1}(x_2-\mu_1)\right)}{\sqrt{(2\pi)^k \det(\Sigma_1)}} = \frac{\exp\left(-\frac{1}{2}\left[\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}1,3253\\0,1429\end{pmatrix}\right]^T\begin{pmatrix}0,5053 & -0,1894\\ -0,1894 & 0,1225\end{pmatrix}\left[\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}1,3253\\0,1429\end{pmatrix}\right]\right)}{\sqrt{(2\pi)^2 \det\begin{pmatrix}0,5053 & -0,1894\\ -0,1894 & 0,1225\end{pmatrix}}} \approx$$

$$\approx \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}-0,3253 & -0,1429\end{pmatrix}\begin{pmatrix}4,7067 & 7,2771\\ 7,2771 & 19,4145\end{pmatrix}\begin{pmatrix}-0,3253\\ -0,1429\end{pmatrix}\right)}{2\pi\sqrt{0,0260}} \approx$$

$$\approx \frac{\exp\left(\begin{pmatrix}0,1626 & 0,0714\end{pmatrix}\begin{pmatrix}4,7067 & 7,2771\\ 7,2771 & 19,4145\end{pmatrix}\begin{pmatrix}-0,3253\\ -0,1429\end{pmatrix}\right)}{1,0131} \approx \frac{\exp\left(\begin{pmatrix}1,2849 & 2,5694\end{pmatrix}\begin{pmatrix}-0,3253\\ -0,1429\end{pmatrix}\right)}{1,0131} \approx$$

$$\approx \frac{\exp(-0,7852)}{1,0131} \approx 0,4501$$

Joint Probability: $p(C=1, x_2) = p(C=1)\,p(x_2|C=1) = 0,6268 \times 0,4501 \approx 0,2821$

Cluster $C=2$: Prior: $p(C=2) = 0,3732$

Likelihood: $p(x_2|C=2) = \mathcal{N}\left(x_2 \middle| \mu_2 = \begin{pmatrix}0,4535\\ 0,6532\end{pmatrix}, \Sigma_2 = \begin{pmatrix}0,4613 & -0,2962\\ -0,2962 & 0,2265\end{pmatrix}\right) =$

$$= \frac{\exp\left(-\frac{1}{2}(x_2-\mu_2)^T(\Sigma_2)^{-1}(x_2-\mu_2)\right)}{\sqrt{(2\pi)^k \det(\Sigma_2)}} = \frac{\exp\left(-\frac{1}{2}\left[\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}0,4535\\0,6532\end{pmatrix}\right]^T\begin{pmatrix}0,4613 & -0,2962\\ -0,2962 & 0,2265\end{pmatrix}^{-1}\left[\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}0,4535\\0,6532\end{pmatrix}\right]\right)}{\sqrt{(2\pi)^2 \det\begin{pmatrix}0,4613 & -0,2962\\ -0,2962 & 0,2265\end{pmatrix}}} \approx$$

$$\approx \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}0,5465 & -0,6532\end{pmatrix}\begin{pmatrix}13,5224 & 17,6836\\ 17,6836 & 27,5403\end{pmatrix}\begin{pmatrix}0,5465\\ -0,6532\end{pmatrix}\right)}{2\pi\sqrt{0,0168}} \approx$$

$$\approx \frac{\exp\left(\begin{pmatrix}-0,2732 & 0,3266\end{pmatrix}\begin{pmatrix}13,5224 & 17,6836\\ 17,6836 & 27,5403\end{pmatrix}\begin{pmatrix}0,5465\\ -0,6532\end{pmatrix}\right)}{0,8144} \approx$$

$$\approx \frac{\exp\left(\begin{pmatrix}2{,}0811 & 4{,}1635\end{pmatrix}\begin{pmatrix}0{,}5465\\-0{,}6532\end{pmatrix}\right)}{0{,}8144} \approx \frac{\exp(-1{,}5822)}{0{,}8144} \approx 0{,}2523$$

Joint Probability: $p(C=2, x_2) = p(C=2)\,p(x_2|C=2) = 0{,}3732 \times 0{,}2523 \approx 0{,}0942$

C=1: $p(C=1|x_2) = \dfrac{p(C=1,x_2)}{p(C=1,x_2)+p(C=2,x_2)} = \dfrac{0{,}2821}{0{,}2821+0{,}0942} \approx 0{,}7497$

C=2: $p(C=2|x_2) = \dfrac{p(C=2,x_2)}{p(C=1,x_2)+p(C=2,x_2)} = \dfrac{0{,}0942}{0{,}2821+0{,}0942} \approx 0{,}2503$

→ For x=3: Cluster C=1: Prior: $p(C=1) = 0{,}6268$

Likelihood: $p(x_3|C=1) = \mathcal{N}\left(x_3 \middle| \mu_1 = \begin{pmatrix}1{,}3253\\0{,}1429\end{pmatrix}, \Sigma_1 = \begin{pmatrix}0{,}5053 & -0{,}1894\\-0{,}1894 & 0{,}1225\end{pmatrix}\right) =$

$$= \frac{\exp\left(-\frac{1}{2}(x_3-\mu_1)^T(\Sigma_1)^{-1}(x_3-\mu_1)\right)}{\sqrt{(2\pi)^K \det(\Sigma_1)}} = \frac{\exp\left(-\frac{1}{2}\left[\begin{pmatrix}0\\1\end{pmatrix}-\begin{pmatrix}1{,}3253\\0{,}1429\end{pmatrix}\right]^T\begin{pmatrix}0{,}5053 & -0{,}1894\\-0{,}1894 & 0{,}1225\end{pmatrix}^{-1}\left[\begin{pmatrix}0\\1\end{pmatrix}-\begin{pmatrix}1{,}3253\\0{,}1429\end{pmatrix}\right]\right)}{\sqrt{(2\pi)^2\det\begin{pmatrix}0{,}5053 & -0{,}1894\\-0{,}1894 & 0{,}1225\end{pmatrix}}} \approx$$

$$\approx \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}-1{,}3253 & 0{,}8571\end{pmatrix}\begin{pmatrix}4{,}7067 & 7{,}2771\\7{,}2771 & 19{,}4145\end{pmatrix}\begin{pmatrix}-1{,}3253\\0{,}8571\end{pmatrix}\right)}{2\pi\sqrt{0{,}0260}} \approx$$

$$\approx \frac{\exp\left(\begin{pmatrix}0{,}6626 & -0{,}4286\end{pmatrix}\begin{pmatrix}4{,}7067 & 7{,}2771\\7{,}2771 & 19{,}4145\end{pmatrix}\begin{pmatrix}-1{,}3253\\0{,}8571\end{pmatrix}\right)}{1{,}0131} \approx$$

$$\approx \frac{\exp\left(\begin{pmatrix}-0{,}0003 & -3{,}4992\end{pmatrix}\begin{pmatrix}-1{,}3253\\0{,}8571\end{pmatrix}\right)}{1{,}0131} \approx \frac{\exp(-2{,}9988)}{1{,}0131} \approx 0{,}0492$$

Joint Probability: $p(C=1, x_3) = p(C=1)\,p(x_3|C=1) = 0{,}6268 \times 0{,}0492 \approx 0{,}0308$

Cluster C=2: Prior: $p(C=2) = 0{,}3732$

Likelihood: $p(x_3|C=1) = \mathcal{N}\left(x_3 \middle| \mu_2 = \begin{pmatrix}0{,}4535\\0{,}6532\end{pmatrix}, \Sigma_2 = \begin{pmatrix}0{,}4613 & -0{,}2962\\-0{,}2962 & 0{,}2265\end{pmatrix}\right) =$

$$= \frac{\exp\left(-\frac{1}{2}(x_3-\mu_2)^T(\Sigma_2)^{-1}(x_3-\mu_2)\right)}{\sqrt{(2\pi)^K \det(\Sigma_2)}} = \frac{\exp\left(-\frac{1}{2}\left[\begin{pmatrix}0\\1\end{pmatrix}-\begin{pmatrix}0{,}4535\\0{,}6532\end{pmatrix}\right]^T\begin{pmatrix}0{,}4613 & -0{,}2962\\-0{,}2962 & 0{,}2265\end{pmatrix}^{-1}\left[\begin{pmatrix}0\\1\end{pmatrix}-\begin{pmatrix}0{,}4535\\0{,}6532\end{pmatrix}\right]\right)}{\sqrt{(2\pi)^2\det\begin{pmatrix}0{,}4613 & -0{,}2962\\-0{,}2962 & 0{,}2265\end{pmatrix}}} \approx$$

$$\approx \frac{\exp\left(-\frac{1}{2}\begin{pmatrix}-0{,}4535 & 0{,}3468\end{pmatrix}\begin{pmatrix}13{,}5224 & 17{,}6836\\17{,}6836 & 27{,}5403\end{pmatrix}\begin{pmatrix}-0{,}4535\\0{,}3468\end{pmatrix}\right)}{2\pi\sqrt{0{,}0168}} \approx$$

$$\approx \frac{\exp\left(\begin{pmatrix}0{,}2268 & -0{,}1734\end{pmatrix}\begin{pmatrix}13{,}5224 & 17{,}6836\\17{,}6836 & 27{,}5403\end{pmatrix}\begin{pmatrix}-0{,}4535\\0{,}3468\end{pmatrix}\right)}{0{,}8144} \approx$$

$$\approx \frac{\exp\left(\begin{pmatrix}0{,}0005 & -0{,}7648\end{pmatrix}\begin{pmatrix}-0{,}4535\\0{,}3468\end{pmatrix}\right)}{0{,}8144} \approx \frac{\exp(-0{,}2655)}{0{,}8144} \approx 0{,}9416$$

Joint Probability: $p(C=2, x_3) = p(C=2)\,p(x_3|C=2) = 0{,}3732 \times 0{,}9416 \approx$
$\approx 0{,}3514$

C=1: $p(c=1|x_3) = \frac{p(c=1,x_3)}{p(c=1,x_3)+p(c=2,x_3)} = \frac{0{,}0308}{0{,}0308+0{,}3514} \approx 0{,}0806$

C=2: $p(c=2|x_3) = \frac{p(c=2,x_3)}{p(c=1,x_3)+p(c=2,x_3)} = \frac{0{,}3514}{0{,}0308+0{,}3514} \approx 0{,}9194$

Maximization step: → For C=1:

$$\mu_1 = \frac{\sum_{n=1}^{3} p(c=1|x_n)x_n}{\sum_{n=1}^{3} p(c=1|x_n)} = \frac{p(c=1|x_1)x_1 + p(c=1|x_2)x_2 + p(c=1|x_3)x_3}{p(c=1|x_1)+p(c=1|x_2)+p(c=1|x_3)} \approx$$

$$\approx \frac{0{,}9989\binom{2}{0} + 0{,}7497\binom{1}{0} + 0{,}0806\binom{0}{1}}{0{,}9989+0{,}7497+0{,}0806} \approx \frac{\binom{1{,}9978}{0} + \binom{0{,}7497}{0} + \binom{0}{0{,}0806}}{1{,}8292} \approx$$

$$\approx \frac{1}{1{,}8292}\binom{2{,}7475}{0{,}0806} \approx \binom{1{,}5020}{0{,}0441}$$

$$\Sigma_{1_{11}} = \frac{p(c=1|x_1)(x_{1_1}-\mu_{1_1})^2 + p(c=1|x_2)(x_{2_1}-\mu_{1_1})^2 + p(c=1|x_3)(x_{3_1}-\mu_{1_1})^2}{p(c=1|x_1)+p(c=1|x_2)+p(c=1|x_3)} =$$

$$= \frac{0{,}9989(2-1{,}5020)^2 + 0{,}7497(1-1{,}5020)^2 + 0{,}0806(0-1{,}5020)^2}{0{,}9989+0{,}7497+0{,}0806} \approx \frac{0{,}6185}{1{,}8292} \approx 0{,}3381$$

$$\Sigma_{1_{12}} = \Sigma_{1_{21}} = \frac{1}{p(c=1|x_1)+p(c=1|x_2)+p(c=1|x_3)}\Big[p(c=1|x_1)(x_{1_1}-\mu_{1_1})(x_{1_2}-\mu_{1_2}) +$$

$$+ p(c=1|x_2)(x_{2_1}-\mu_{1_1})(x_{2_2}-\mu_{1_2}) + p(c=1|x_3)(x_{3_1}-\mu_{1_1})(x_{3_2}-\mu_{1_2})\Big] =$$

$$= \frac{0{,}9989(2-1{,}5020)(0-0{,}0441) + 0{,}7497(1-1{,}5020)(0-0{,}0441) + 0{,}0806(0-1{,}5020)(1-0{,}0441)}{0{,}9989+0{,}7497+0{,}0806} \approx$$

$$\approx \frac{-0{,}1211}{1{,}8292} \approx -0{,}0662$$

$$\Sigma_{1_{22}} = \frac{p(c=1|x_1)(x_{1_2}-\mu_{1_2})^2 + p(c=1|x_2)(x_{2_2}-\mu_{1_2})^2 + p(c=1|x_3)(x_{3_2}-\mu_{1_2})^2}{p(c=1|x_1)+p(c=1|x_2)+p(c=1|x_3)} =$$

$$= \frac{0{,}9989(0-0{,}0441)^2 + 0{,}7497(0-0{,}0441)^2 + 0{,}0806(1-0{,}0441)^2}{0{,}9989+0{,}7497+0{,}0806} \approx$$

$$\approx \frac{0{,}0770}{1{,}8292} \approx 0{,}0421 \qquad \Sigma_1 = \begin{bmatrix} 0{,}3381 & -0{,}0662 \\ -0{,}0662 & 0{,}0421 \end{bmatrix}$$

$$p(c=1) = \frac{\sum_{n=1}^{3} p(c=1|x_n)}{\sum_{l=1}^{2}\sum_{n=1}^{3} p(c=l|x_n)} = \frac{p(c=1|x_1)+p(c=1|x_2)+p(c=1|x_3)}{p(c=1|x_1)+p(c=1|x_2)+p(c=1|x_3)+p(c=2|x_1)+p(c=2|x_2)+p(c=2|x_3)} =$$

$$= \frac{0{,}9989+0{,}7497+0{,}0806}{0{,}9989+0{,}7497+0{,}0806+0{,}0011+0{,}2503+0{,}9194} = \frac{1{,}8292}{3} \approx 0{,}6097$$

→ For C=2: $\mu_2 = \frac{\sum_{n=1}^{3} p(c=2|x_n)x_n}{\sum_{n=1}^{3} p(c=2|x_n)} = \frac{p(c=2|x_1)x_1 + p(c=2|x_2)x_2 + p(c=2|x_3)x_3}{p(c=2|x_1)+p(c=2|x_2)+p(c=2|x_3)} =$

$$= \frac{0,0011\begin{pmatrix}2\\0\end{pmatrix} + 0,2503\begin{pmatrix}1\\0\end{pmatrix} + 0,9194\begin{pmatrix}0\\1\end{pmatrix}}{0,0011 + 0,2503 + 0,9194} = \frac{\begin{pmatrix}0,0022\\0\end{pmatrix} + \begin{pmatrix}0,2503\\0\end{pmatrix} + \begin{pmatrix}0\\0,9194\end{pmatrix}}{1,1708} =$$

$$= \frac{1}{1,1708}\begin{pmatrix}0,2525\\0,9194\end{pmatrix} \approx \begin{pmatrix}0,2157\\0,7853\end{pmatrix}$$

$$\Sigma_{2_{11}} = \frac{p(c=2|x_1)(x_{1_1}-\mu_{2_1})^2 + p(c=2|x_2)(x_{2_1}-\mu_{2_1})^2 + p(c=2|x_3)(x_{3_1}-\mu_{2_1})^2}{p(c=2|x_1) + p(c=2|x_2) + p(c=2|x_3)} =$$

$$= \frac{0,0011(2-0,2157)^2 + 0,2503(1-0,2157)^2 + 0,9194(0-0,2157)^2}{0,0011+0,2503+0,9194} \approx$$

$$\approx \frac{0,2002}{1,1708} \approx 0,1710$$

$$\Sigma_{2_{12}} = \Sigma_{2_{21}} = \frac{1}{p(c=2|x_1)+p(c=2|x_2)+p(c=2|x_3)}\Big[p(c=2|x_1)(x_{1_1}-\mu_{2_1})(x_{1_2}-\mu_{2_2}) +$$

$$+ p(c=2|x_2)(x_{2_1}-\mu_{2_1})(x_{2_2}-\mu_{2_2}) + p(c=2|x_3)(x_{3_1}-\mu_{2_1})(x_{3_2}-\mu_{2_2})\Big] =$$

$$= \frac{0,0011(2-0,2157)(0-0,7853) + 0,2503(1-0,2157)(0-0,7853) + 0,9194(0-0,2157)(1-0,7853)}{0,0011+0,2503+0,9194} \approx$$

$$\approx \frac{-0,1983}{1,1708} \approx -0,1694$$

$$\Sigma_{2_{22}} = \frac{p(c=2|x_1)(x_{1_2}-\mu_{2_2})^2 + p(c=2|x_2)(x_{2_2}-\mu_{2_2})^2 + p(c=2|x_3)(x_{3_2}-\mu_{2_2})^2}{p(c=2|x_1)+p(c=2|x_2)+p(c=2|x_3)} =$$

$$= \frac{0,0011(0-0,7853)^2 + 0,2503(0-0,7853)^2 + 0,9194(1-0,7853)^2}{0,0011+0,2503+0,9194} \approx \frac{0,1974}{1,1708} \approx$$

$$\approx 0,1686$$

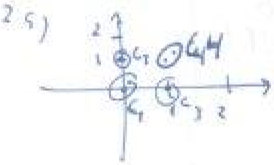
$$\Sigma_2 = \begin{bmatrix}0,1710 & -0,1694\\ -0,1694 & 0,1686\end{bmatrix}$$

$$p(c=2) = \frac{\sum_{n=1}^{3} p(c=2|x_n)}{\sum_{l=1}^{2}\sum_{n=1}^{3} p(c=l|x_n)} =$$

$$= \frac{p(c=2|x_1)+p(c=2|x_2)+p(c=2|x_3)}{p(c=1|x_1)+p(c=1|x_2)+p(c=1|x_3)+p(c=2|x_1)+p(c=2|x_2)+p(c=2|x_3)} =$$

$$= \frac{0,0011+0,2503+0,9194}{0,9989+0,7497+0,0806+0,0011+0,2503+0,9194} = \frac{1,1708}{3} \approx 0,3903$$

2 c)



$$c_1 = \frac{1}{2}\begin{pmatrix}0.2\\0.2\end{pmatrix} = \begin{pmatrix}0.1\\0.1\end{pmatrix} \qquad c_2 = 1\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}0\\1\end{pmatrix}$$

$$c_3 = 1\begin{pmatrix}1\\0\end{pmatrix} \qquad c_4 = \frac{1}{2}\begin{pmatrix}2.2\\2.2\end{pmatrix} = \begin{pmatrix}1.1\\1.1\end{pmatrix}$$

$$u^{[1]} = \begin{bmatrix}F(u_1)\\F(u_2)\\F(u_3)\\F(u_4)\\F(u_5)\\F(u_6)\end{bmatrix} = \begin{bmatrix}1.8\\1.871\\1.871\\1.91\\1.538\\2.191\end{bmatrix}$$

$$z^{[2]} = W^{[2]} u^{[1]} + b^{[2]} = \begin{bmatrix}1&1&1&1&1&1\\1&1&1&1&1&1\\1&1&1&1&1&1\\1&1&1&1&1&1\\1&1&1&1&1&1\\1&1&1&1&1&1\\1&1&1&1&1&1\end{bmatrix}\begin{bmatrix}1.8\\1.871\\1.871\\1.91\\1.538\\2.191\end{bmatrix} + \begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix}$$

$$= \begin{bmatrix}12.181\\12.181\\12.181\\12.181\\12.181\\12.181\end{bmatrix} \qquad u^{[2]} = \text{sigmoid}\left[z^{[2]}\right] = \begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix}$$

$$\delta^{[2]} = u^{[2]} - t = \begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix} - \begin{bmatrix}1\\0\\0\\1\\1\\1\end{bmatrix} = \begin{bmatrix}0\\1\\1\\0\\0\\0\end{bmatrix}$$

$$\frac{\partial C}{\partial W^{[2]}} = \delta^{[2]} \frac{\partial z^{[2]^T}}{\partial W^{[2]}} = \delta^{[2]} \left(u^{[1]}\right)^T =$$

$$\begin{pmatrix}0\\1\\1\\0\\1\end{pmatrix}\begin{bmatrix}1.8 & 1.871 & 1.871 & 1.91 & 1.538 & 2.191\end{bmatrix} =$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1.8 & 1.871 & 1.91 & 1.538 & 2.151 \\ 1.8 & 1.871 & 1.91 & 1.538 & 2.181 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$w^{[2]} = w^{[2]} - \eta \frac{\partial E}{\partial w^{[2]}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -0.8 & -0.871 & -0.871 & -0.91 & -0.538 & -1.181 \\ -0.8 & -0.871 & -0.871 & -0.91 & -0.538 & -1.181 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[2]}} = \delta^{[2]} \frac{\partial z^{[2]}}{\partial b^{[2]}} = \delta^{[2]} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$b^{[2]} = b^{[2]} - \eta\, b^{[2]} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - 1 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

2 epoch $z^{[2]} = w^{[2]} a^{[1]} + b^{[2]} =$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -0.8 & -0.871 & -0.871 & -0.91 & -0.538 & -1.151 \\ -0.8 & -0.871 & -0.871 & -0.91 & -0.538 & -1.151 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1.8 \\ 1.871 \\ 1.871 \\ 1.91 \\ 1.538 \\ 2.191 \end{bmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$= \begin{bmatrix} 12.181 \\ -11.87 \\ -11.87 \\ 12.181 \\ 12.181 \\ 12.181 \end{bmatrix} \qquad a^{[2]} = \mathrm{sigmoid}(z^{[2]}) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad \delta^{[2]} = a^{[2]} - t = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[2]}} = \delta^{[2]} \frac{\partial d^{[2]T}}{\partial w^{[2]}} = \delta^{[2]} \left[u^{[1]}\right]^{T} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1.8 & 1.871 & 1.871 & 1.91 & 1.528 & 2.191 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1.8 & 1.871 & 1.871 & 1.91 & 1.528 & 2.191 \\ 1.8 & 1.871 & 1.871 & 1.91 & 1.528 & 2.191 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$w^{[2]} = w^{[2]} - \eta \frac{\partial E}{\partial w^{[2]}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -2.6 & -2,742 & -2,742 & 2,82 & -1,076 & -3,782 \\ -2,6 & -2,742 & -2,742 & 2,82 & -1,076 & -3,782 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[2]}} = \delta^{[2]} \frac{\partial d^{[2]T}}{\partial b^{[2]}} = \delta^{[2]} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad b^{[2]} = b^{[2]} - \eta b^{[2]} =$$

$$= \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

2 b)

$$u^{[1]} = F\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1,91$$

The Output of the RBF Network to querry point $\begin{bmatrix} 1 \\ 1 \end{bmatrix}^T$ is ~~1.91~~ 1.91

III) PCA: $X = \left(x_1 = \begin{pmatrix}1\\1\end{pmatrix}; x_2 = \begin{pmatrix}2\\2\end{pmatrix}; x_3 = \begin{pmatrix}3\\3\end{pmatrix}; x_4 = \begin{pmatrix}5\\4\end{pmatrix}\right)$

a) $\mu = \frac{1}{4}\left(\begin{pmatrix}1\\1\end{pmatrix} + \begin{pmatrix}2\\2\end{pmatrix} + \begin{pmatrix}3\\3\end{pmatrix} + \begin{pmatrix}5\\4\end{pmatrix}\right) = \frac{1}{4}\begin{pmatrix}11\\10\end{pmatrix} = \begin{pmatrix}2{,}75\\2{,}50\end{pmatrix}$

$X_1 = \left(\begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}2{,}75\\2{,}50\end{pmatrix}\right)\left(\begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}2{,}75\\2{,}50\end{pmatrix}\right)^T = \begin{pmatrix}-1{,}75\\-1{,}50\end{pmatrix}(-1{,}75 \ \ -1{,}50) = \begin{pmatrix}3{,}0625 & 2{,}6250\\2{,}6250 & 2{,}2500\end{pmatrix}$

$X_2 = \left(\begin{pmatrix}2\\2\end{pmatrix} - \begin{pmatrix}2{,}75\\2{,}50\end{pmatrix}\right)\left(\begin{pmatrix}2\\2\end{pmatrix} - \begin{pmatrix}2{,}75\\2{,}50\end{pmatrix}\right)^T = \begin{pmatrix}-0{,}75\\-0{,}50\end{pmatrix}(-0{,}75 \ \ -0{,}50) = \begin{pmatrix}0{,}5625 & 0{,}3750\\0{,}3750 & 0{,}2500\end{pmatrix}$

$X_3 = \left(\begin{pmatrix}3\\3\end{pmatrix} - \begin{pmatrix}2{,}75\\2{,}50\end{pmatrix}\right)\left(\begin{pmatrix}3\\3\end{pmatrix} - \begin{pmatrix}2{,}75\\2{,}50\end{pmatrix}\right)^T = \begin{pmatrix}0{,}25\\0{,}50\end{pmatrix}(0{,}25 \ \ 0{,}50) = \begin{pmatrix}0{,}0625 & 0{,}1250\\0{,}1250 & 0{,}2500\end{pmatrix}$

$X_4 = \left(\begin{pmatrix}5\\4\end{pmatrix} - \begin{pmatrix}2{,}75\\2{,}50\end{pmatrix}\right)\left(\begin{pmatrix}5\\4\end{pmatrix} - \begin{pmatrix}2{,}75\\2{,}50\end{pmatrix}\right)^T = \begin{pmatrix}2{,}25\\1{,}50\end{pmatrix}(2{,}25 \ \ 1{,}50) = \begin{pmatrix}5{,}0625 & 3{,}3750\\3{,}3750 & 2{,}2500\end{pmatrix}$

$C = \frac{1}{4-1}\left(\begin{pmatrix}3{,}0625 & 2{,}6250\\2{,}6250 & 2{,}2500\end{pmatrix} + \begin{pmatrix}0{,}5625 & 0{,}3750\\0{,}3750 & 0{,}2500\end{pmatrix} + \begin{pmatrix}0{,}0625 & 0{,}1250\\0{,}1250 & 0{,}2500\end{pmatrix} + \right.$

$\left. + \begin{pmatrix}5{,}0625 & 3{,}3750\\3{,}3750 & 2{,}2500\end{pmatrix}\right) = \frac{1}{3}\begin{pmatrix}8{,}7500 & 6{,}5000\\6{,}5000 & 5{,}0000\end{pmatrix} \approx \begin{pmatrix}2{,}9167 & 2{,}1667\\2{,}1667 & 1{,}6667\end{pmatrix}$

$|C - \lambda I| = 0 \Leftrightarrow \left|\begin{pmatrix}2{,}9167 & 2{,}1667\\2{,}1667 & 1{,}6667\end{pmatrix} - \lambda\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right| = 0 \Leftrightarrow$

$\Leftrightarrow \left|\begin{pmatrix}2{,}9167 & 2{,}1667\\2{,}1667 & 1{,}6667\end{pmatrix} - \begin{pmatrix}\lambda & 0\\0 & \lambda\end{pmatrix}\right| = 0 \Leftrightarrow \left|\begin{pmatrix}2{,}9167-\lambda & 2{,}1667\\2{,}1667 & 1{,}6667-\lambda\end{pmatrix}\right| = 0 \Leftrightarrow$

$\Leftrightarrow (2{,}9167-\lambda)(1{,}6667-\lambda) - (2{,}1667)(2{,}1667) = 0 \Leftrightarrow$

$\Leftrightarrow 4{,}8613 - 2{,}9167\lambda - 1{,}6667\lambda + \lambda^2 - 4{,}6946 = 0 \Leftrightarrow$

$\Leftrightarrow \lambda^2 - 4{,}5834\lambda + 0{,}1667 = 0 \Leftrightarrow \lambda = \frac{4{,}5834 \pm \sqrt{(-4{,}5834)^2 - 4(1)(0{,}1667)}}{2(1)} \Leftrightarrow$

$\Leftrightarrow \lambda = \frac{4{,}5834 \pm \sqrt{20{,}3408}}{2} \Leftrightarrow \lambda_1 \approx 0{,}0367 \vee \lambda_2 \approx 4{,}5467$

$C u_1 = \lambda_1 u_1 \Leftrightarrow (C - \lambda_1 I)u_1 = 0 \Leftrightarrow$

$\Leftrightarrow \left[\begin{pmatrix}2{,}9167 & 2{,}1667\\2{,}1667 & 1{,}6667\end{pmatrix} - 0{,}0367\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right]\begin{pmatrix}u_1\\u_2\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \Leftrightarrow$

$\Leftrightarrow \left[\begin{pmatrix}2{,}9167 & 2{,}1667\\2{,}1667 & 1{,}6667\end{pmatrix} - \begin{pmatrix}0{,}0367 & 0\\0 & 0{,}0367\end{pmatrix}\right]\begin{pmatrix}u_{11}\\u_{12}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \Leftrightarrow$

$\Leftrightarrow \begin{pmatrix}2{,}8800 & 2{,}1667\\2{,}1667 & 1{,}6300\end{pmatrix}\begin{pmatrix}u_{11}\\u_{12}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \Leftrightarrow \begin{pmatrix}2{,}8800u_{11} + 2{,}1667u_{12}\\2{,}1667u_{11} + 1{,}6300u_{12}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$

$2{,}8800u_{11} + 2{,}1667u_{12} = 0 \Leftrightarrow u_{12} = \frac{-2{,}8800u_{11}}{2{,}1667} \Leftrightarrow u_{12} \approx -1{,}3292u_{11}$

Choosing $u_{11} = 1$, we have: $u_1 = \begin{pmatrix}u_{11}\\-1{,}3292u_{11}\end{pmatrix} \Leftrightarrow u_1 = \begin{pmatrix}1\\-1{,}3292\end{pmatrix}$

$u_1 = \frac{u_1}{\|u_1\|_2} = \frac{1}{\sqrt{1^2+(-1,3292)^2}}\begin{pmatrix}1\\-1,3292\end{pmatrix} \approx \frac{1}{1,6634}\begin{pmatrix}1\\-1,3292\end{pmatrix} \approx$

$\approx \begin{pmatrix}0,6012\\-0,7991\end{pmatrix}$

$Cu_2 = \lambda_2 u_2 \Leftrightarrow (C-\lambda_2 I)u_2 = 0 \Leftrightarrow$

$\Leftrightarrow \left[\begin{pmatrix}2,9167 & 2,1667\\2,1667 & 1,6667\end{pmatrix} - 4,5467\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right]\begin{pmatrix}u_{21}\\u_{22}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \Leftrightarrow$

$\Leftrightarrow \left[\begin{pmatrix}2,9167 & 2,1667\\2,1667 & 1,6667\end{pmatrix} - \begin{pmatrix}4,5467 & 0\\0 & 4,5467\end{pmatrix}\right]\begin{pmatrix}u_{21}\\u_{22}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \Leftrightarrow$

$\Leftrightarrow \begin{pmatrix}-1,6300 & 2,1667\\2,1667 & -2,8800\end{pmatrix}\begin{pmatrix}u_{21}\\u_{22}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \Leftrightarrow \begin{pmatrix}-1,6300u_{21}+2,1667u_{22}\\2,1667u_{21}-2,8800u_{22}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$

$-1,6300u_{21}+2,1667u_{22} = 0 \Leftrightarrow u_{22} = \frac{1,6300u_{21}}{2,1667} \Leftrightarrow u_{22} \approx 0,7523u_{21}$

$u_2 = \begin{pmatrix}u_{21}\\0,7523u_{21}\end{pmatrix}$; choosing $u_{21}=1$, we have: $u_2 = \begin{pmatrix}1\\0,7523\end{pmatrix}$

$u_2 = \frac{u_2}{\|u_2\|_2} = \frac{1}{\sqrt{1^2+0,7523^2}}\begin{pmatrix}1\\0,7523\end{pmatrix} \approx \frac{1}{1,2514}\begin{pmatrix}1\\0,7523\end{pmatrix} \approx \begin{pmatrix}0,7991\\0,6012\end{pmatrix}$

$U_{K-L} = \begin{pmatrix}0,6012 & 0,7991\\-0,7991 & 0,6012\end{pmatrix}$

b) $\lambda_1 = 0,0367$; $\lambda_2 = 4,5467 \Rightarrow \lambda_2 > \lambda_1$, so the projected points are associated to λ_2

$y_1 = u_2^T x_1 = (0,7991\ \ 0,6012)\begin{pmatrix}1\\1\end{pmatrix} = (1,4003)$

$y_2 = u_2^T x_2 = (0,7991\ \ 0,6012)\begin{pmatrix}2\\2\end{pmatrix} = (2,8006)$

$y_3 = u_2^T x_3 = (0,7991\ \ 0,6012)\begin{pmatrix}3\\3\end{pmatrix} = (4,2009)$

$y_4 = u_2^T x_4 = (0,7991\ \ 0,6012)\begin{pmatrix}5\\4\end{pmatrix} = (6,4003)$

IV) VC Dimension:

a) 1) Between the input layer and the hidden layer, we will have:

4×4 weights matrix, so $4\times4=16$ parameters

4×1 bias vector, so $4\times1=4$ parameters

Between the hidden layer and the output layer, we will have:

1×4 weights matrix, so $1\times4=4$ parameters

1×1 bias vector, so $1\times1=1$ parameter

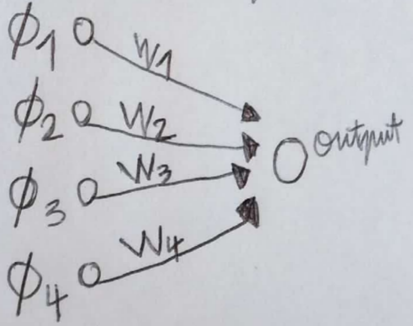
So, in total there are $16+4+4+1=$ 25 free parameters

2) Hidden layer has 4 units, and so we have 4 cluster centers.

$$\phi_n=\phi(\|x-c_k\|)=\exp\left(-\frac{\|x-c_k\|^2}{2\sigma^2}\right),\ K=1,2,3,4$$

And so we have $\sigma+c_1+c_2+c_3+c_4$, and so, we will have $1+1+1+1+1=5$ parameters

So, between the hidden layer and the output layer, we have:



We have 4 weights, w_1, w_2, w_3 and w_4, so we have 4 parameters

Since we are performing a logistic regression, we also have the threshold as a free parameter, and so we have 1 more parameter for the threshold.

So, in total there are $5+4+1=$ 10 free parameters

So, by computing both scenarios, since the scenario 2 has less free parameters, we can conclude that scenario 2 has the smallest VC dimension

b) Regularization is used to reduce variance.
VC dimension being high implies on high variance.
So, we can say that Regularization and VC dimension have an inverse relation.