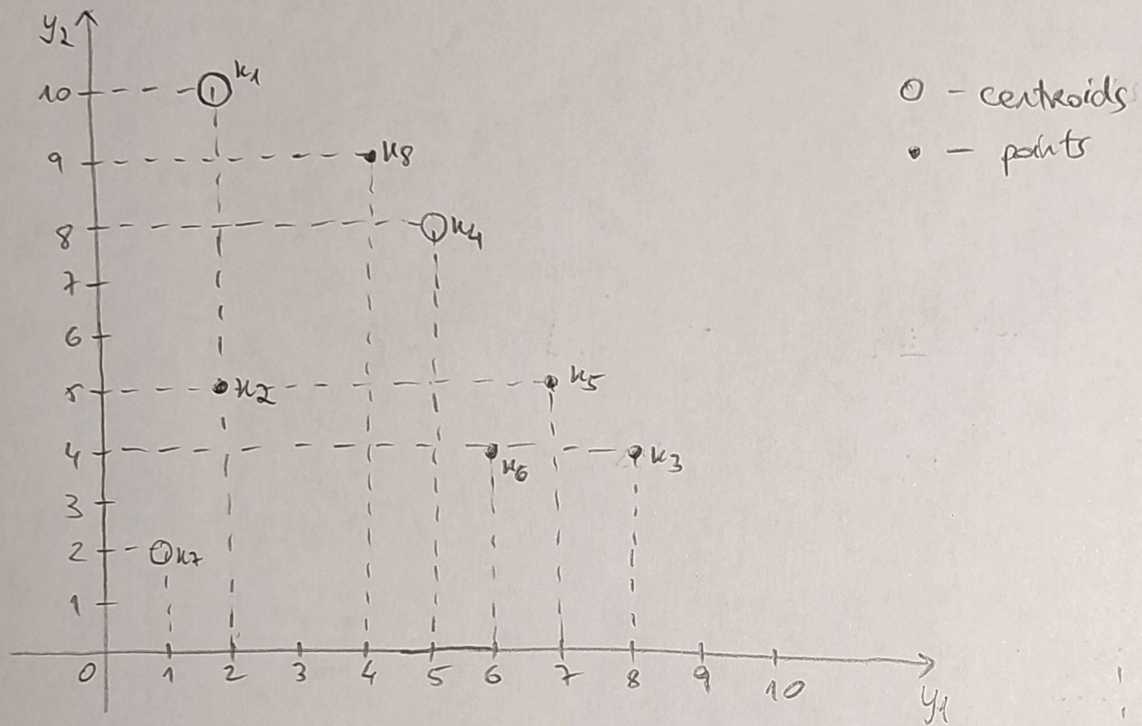
Madalena Yang 110206
Madalena Mota 110355

Homework IV

Part A - Clustering

x	y_1	y_2
x_1	2	10
x_2	2	5
x_3	8	4
x_4	5	8
x_5	7	5
x_6	6	4
x_7	1	2
x_8	4	9

a.



K-means

1) Initialize centroids
2) Assign points to clusters
3) Adjust centroids → mean
4) Re-assign points

b. 1) $\mu_1 = [2 \quad 10]^T \qquad \mu_2 = [5 \quad 8]^T \qquad \mu_3 = [1 \quad 2]^T$

2) $\underline{\underline{x_2}}$: $(2,5)$

$\|x_2 - \mu_1\|_2 = \sqrt{(2-2)^2 + (5-10)^2} = \sqrt{25} = 5$

$\|x_2 - \mu_2\|_2 = \sqrt{(2-5)^2 + (5-8)^2} = \sqrt{18} = 3\sqrt{2}$

$\|x_2 - \mu_3\|_2 = \sqrt{(2-1)^2 + (5-2)^2} = \sqrt{10}$

$c_{x_2} = \underset{c \in \{1,2,3\}}{\operatorname{argmin}} \|x_2 - \mu_c\|_2 = c_3 \qquad \boxed{x_2 \longrightarrow c_3}$

$\underline{\underline{x_3}}$: $(8,4)$

$\|x_3 - \mu_1\|_2 = \sqrt{(8-2)^2 + (4-10)^2} = \sqrt{36+36} = \sqrt{72}$

$\|x_3 - \mu_2\|_2 = \sqrt{(8-5)^2 + (4-8)^2} = \sqrt{25}$

$\|x_3 - \mu_3\|_2 = \sqrt{(8-1)^2 + (4-2)^2} = \sqrt{49+4} = \sqrt{53}$

$c_{x_3} = \underset{c \in \{1,2,3\}}{\operatorname{argmin}} \|x_3 - \mu_c\|_2 = c_2 \qquad \boxed{x_3 \longrightarrow c_2}$

$\underline{\underline{x_5}}$: $(7,5)$

$\|x_5 - \mu_1\|_2 = \sqrt{(7-2)^2 + (5-10)^2} = \sqrt{50}$

$\|x_5 - \mu_2\|_2 = \sqrt{(7-5)^2 + (5-8)^2} = \sqrt{13}$

$\|x_5 - \mu_3\|_2 = \sqrt{(7-1)^2 + (2-5)^2} = \sqrt{45}$

$c_{x_5} = \underset{c \in \{1,2,3\}}{\operatorname{argmin}} \|x_5 - \mu_c\|_2 = c_2 \qquad \boxed{x_5 \longrightarrow c_2}$

x_6 : $\|x_6-\mu_1\|_2 = \sqrt{(6-2)^2+(4-10)^2} = \sqrt{16+36} = \sqrt{52}$

(6,4)

$\|x_6-\mu_2\|_2 = \sqrt{(6-5)^2+(4-8)^2} = \sqrt{1+16} = \sqrt{17}$

$\|x_6-\mu_3\|_2 = \sqrt{(6-1)^2+(4-5)^2} = \sqrt{25+1} = \sqrt{26}$

$C_{x_6} = \underset{c\in\{1,2,3\}}{\arg\min} \|x_6-\mu_c\|_2 = c_2$ $\quad \boxed{x_6 \to c_2}$

x_8 : $\|x_8-\mu_1\|_2 = \sqrt{(4-2)^2+(9-10)^2} = \sqrt{4+1} = \sqrt{5}$

(4,9)

$\|x_8-\mu_2\|_2 = \sqrt{(4-5)^2+(9-8)^2} = \sqrt{2}$

$\|x_8-\mu_3\|_2 = \sqrt{(4-1)^2+(9-5)^2} = \sqrt{25}$

$C_{x_8} = \underset{c\in\{1,2,3\}}{\arg\min} \|x_8-\mu_c\|_2 = c_2$ $\quad \boxed{x_8 \to c_2}$

$C_{x_1} = c_1$ $\quad C_{x_4} = c_2$ $\quad C_{x_7} = c_3$

$\boxed{x_1 \to c_1}$ $\quad \boxed{x_4 \to c_2}$ $\quad \boxed{x_7 \to c_3}$

3) $\mu_1 = \frac{x_1}{1} = (2,10)$

$\mu_2 = \frac{x_3+x_4+x_5+x_6+x_8}{5} = \left(\frac{8+5+7+6+4}{5}, \frac{4+8+5+4+9}{5}\right) = (6,6)$

$\mu_3 = \frac{x_2+x_7}{2} = \left(\frac{2+1}{2}, \frac{5+2}{2}\right) = \left(\frac{3}{2}, \frac{7}{2}\right) = (1,5 ; 3,5)$

4) $\boxed{x_1 \to c_1}$

x_2 : $\|x_2-\mu_1\|_2 = \sqrt{(2-2)^2+(5-10)^2} = \sqrt{25}$

(2,5)

$\|x_2-\mu_2\|_2 = \sqrt{(2-6)^2+(5-6)^2} = \sqrt{17}$

$\|x_2-\mu_3\|_2 = \sqrt{(2-1,5)^2+(5-3,5)^2} = \sqrt{2,5}$

$C_{x_2} = \underset{c\in\{1,2,3\}}{\arg\min} \|x_2-\mu_c\|_2 = c_3$ $\quad \boxed{x_2 \to c_3}$

x_3 : $\|x_3-\mu_1\|_2 = \sqrt{(8-2)^2+(4-10)^2} = \sqrt{72}$

(8,4)

$\|x_3-\mu_2\|_2 = \sqrt{(8-6)^2+(4-6)^2} = \sqrt{8}$

$\|x_3-\mu_3\|_2 = \sqrt{(8-1,5)^2+(4-3,5)^2} = \sqrt{6,5^2+0,5^2} = \sqrt{42,5}$

$C_{x_3} = \underset{c\in\{1,2,3\}}{\arg\min} \|x_3-\mu_c\|_2 = c_2$ $\quad \boxed{x_3 \to c_3}$

x_4 : $\|x_4-\mu_1\|_2 = \sqrt{(5-2)^2+(8-10)^2} = \sqrt{13}$

(5,8)

$\|x_4-\mu_2\|_2 = \sqrt{(5-6)^2+(8-6)^2} = \sqrt{5}$

$\|x_4-\mu_3\|_2 = \sqrt{(5-1,5)^2+(8-3,5)^2} = \sqrt{3,5^2+4,5^2} = \sqrt{32,5}$

$C_{x_4} = \underset{c\in\{1,2,3\}}{\arg\min} \|x_4-\mu_c\|_2 = c_2$ $\quad \boxed{x_4 \to c_2}$

x_5: $\|x_5 - \mu_1\|_2 = \sqrt{(7-2)^2 + (5-10)^2} = \sqrt{50}$
(7,5)
$\|x_5 - \mu_2\|_2 = \sqrt{(7-6)^2 + (5-6)^2} = \sqrt{2}$

$\|x_5 - \mu_3\|_2 = \sqrt{(7-1{,}5)^2 + (5-3{,}5)^2} = \sqrt{5{,}5^2 + 1{,}5^2}$

$C_{x_5} = \underset{c \in \{1,2,3\}}{\operatorname{argmin}} \|x_5 - \mu_c\|_2 = c_2$ $\boxed{x_5 \to c_2}$

x_6: $\|x_6 - \mu_1\|_2 = \sqrt{(6-2)^2 + (4-10)^2} = \sqrt{16+36} = \sqrt{52}$
(6,4)
$\|x_6 - \mu_2\|_2 = \sqrt{(6-6)^2 + (4-6)^2} = \sqrt{4}$

$\|x_6 - \mu_3\|_2 = \sqrt{(6-1{,}5)^2 + (4-3{,}5)^2} = \sqrt{4{,}5^2 + 0{,}5^2} = \sqrt{20{,}5}$

$C_{x_6} = \underset{c \in \{1,2,3\}}{\operatorname{argmin}} \|x_6 - \mu_c\|_2 = c_2$ $\boxed{x_6 \to c_2}$

x_7: $\|x_7 - \mu_1\|_2 = \sqrt{(1-2)^2 + (2-10)^2} = \sqrt{1+64} = \sqrt{65}$
(1,2)
$\|x_7 - \mu_2\|_2 = \sqrt{(1-6)^2 + (2-6)^2} = \sqrt{25+16} = \sqrt{41}$

$\|x_7 - \mu_3\|_2 = \sqrt{(1-1{,}5)^2 + (2-3{,}5)^2} = \sqrt{0{,}5^2 + 1{,}5^2} = \sqrt{2{,}5}$

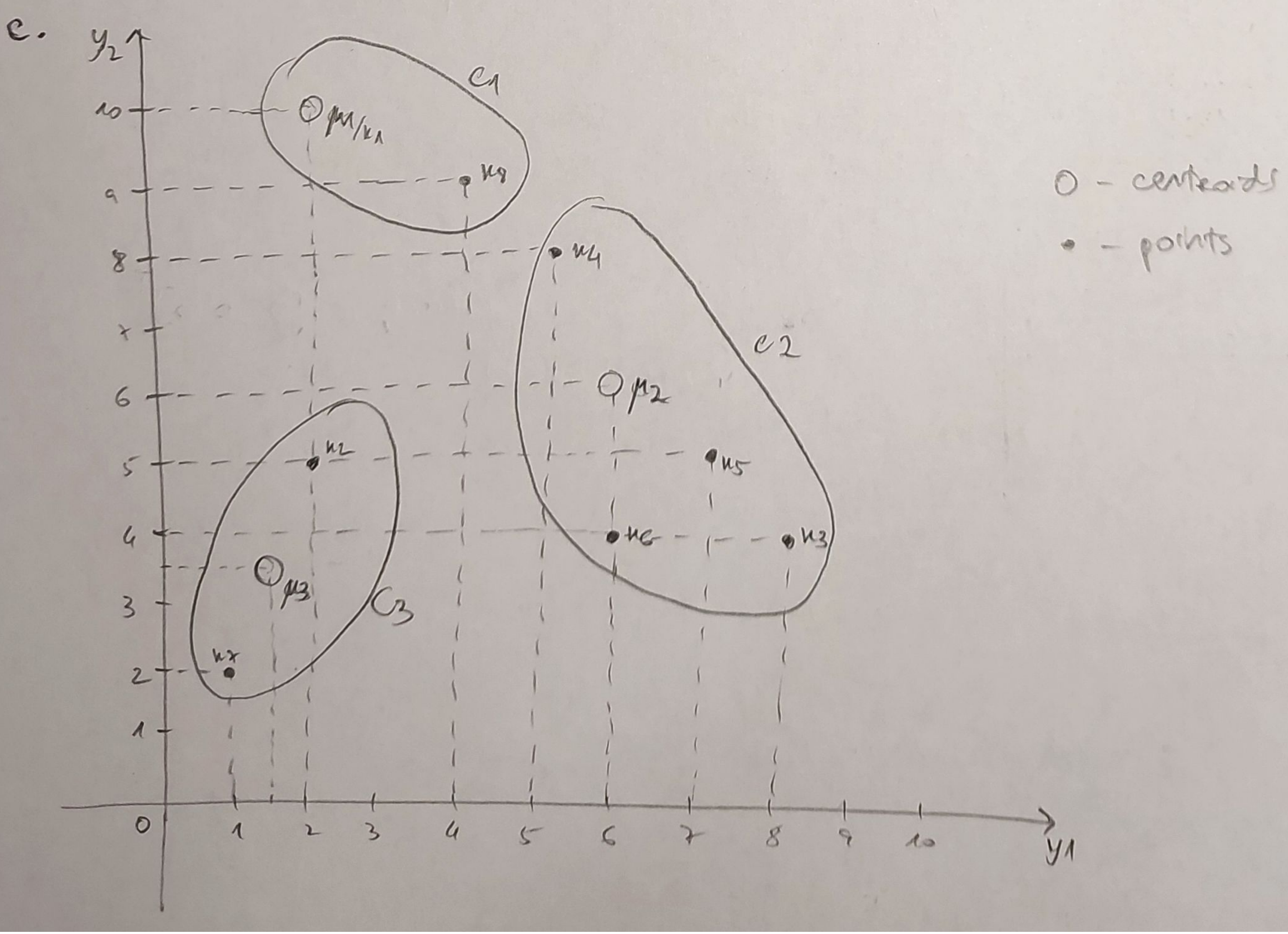
$C_{x_7} = \underset{c \in \{1,2,3\}}{\operatorname{argmin}} \|x_7 - \mu_c\|_2 = c_3$ $\boxed{x_7 \to c_3}$

x_8: $\|x_8 - \mu_1\|_2 = \sqrt{(4-2)^2 + (9-10)^2} = \sqrt{5}$
(4,9)
$\|x_8 - \mu_2\|_2 = \sqrt{(4-6)^2 + (9-6)^2} = \sqrt{13}$

$\|x_8 - \mu_3\|_2 = \sqrt{(4-1{,}5)^2 + (9-3{,}5)^2} = \sqrt{36{,}5}$

$C_{x_8} = \underset{c \in \{1,2,3\}}{\operatorname{argmin}} \|x_8 - \mu_c\|_2 = c_1$ $\boxed{x_8 \to c_1}$

c.



d. Different centroid initializations greatly impact the convergence speed, cluster quality (how well data is grouped) and stability of the K-means algorithm. This algorithm mostly assumes that clusters are well-separated, not overlapping and spherical. However this is not true for real datasets.

If the centroids are too close or too far, it can lead to poor convergence and lead to converging into a local minimum instead of a global one. Moreover, rerunning the algorithm is a challenge, as different centroids lead to different clustering results.

We can counteract these difficulties by running the algorithm multiple times with different seeds and select the best, removing outliers, or using methods like PCA or LDA to normalize data perform dimensionality reduction, or optimize class separation.

Part B – PCA

D	y_1	y_2	y_3	y_4
x_1	5	0	1	+
x_2	0	5	0	+
x_3	1	0	-1	–

PCA

→ apply K-L Transform

1) center data
2) compute covariance matrix
3) eigenvalues and eigenvectors
4) $x' = \mu^T x$

→ n dimensions of higher variance (n higher λ)

$$\text{cov}(x,y) = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{N-1}$$

for samples

a. 1) center data $\quad x_i' = (x_i - \mu)$ $\quad \boxed{N=3}$

$$\mu = \frac{1}{3}\left(\begin{bmatrix}5\\0\\1\end{bmatrix} + \begin{bmatrix}0\\5\\0\end{bmatrix} + \begin{bmatrix}1\\0\\-1\end{bmatrix}\right) = \begin{bmatrix}2\\5/3\\0\end{bmatrix}$$

$$x_1' = \begin{bmatrix}5\\0\\1\end{bmatrix} - \begin{bmatrix}2\\5/3\\0\end{bmatrix} = \begin{bmatrix}3\\-5/3\\1\end{bmatrix}$$

$$x_2' = \begin{bmatrix}0\\5\\0\end{bmatrix} - \begin{bmatrix}2\\5/3\\0\end{bmatrix} = \begin{bmatrix}-2\\10/3\\0\end{bmatrix} \qquad x_3' = \begin{bmatrix}1\\0\\-1\end{bmatrix} - \begin{bmatrix}2\\5/3\\0\end{bmatrix} = \begin{bmatrix}-1\\-5/3\\-1\end{bmatrix}$$

2) Compute covariance matrix

Centered Dataset

D'	y_1	y_2	y_3	y_4
x_1'	3	-5/3	1	+
x_2'	-2	10/3	0	+
x_3'	-1	-5/3	-1	–

Since D' is centered dataset, $\mu = [0 \ 0 \ 0]^T$ (removed mean)

$$\Sigma = \begin{bmatrix} \underbrace{\text{cov}(y_1,y_1)}_{\sigma_1^2} & \text{cov}(y_1,y_2) & \text{cov}(y_1,y_3) \\ \text{cov}(y_2,y_1) & \underbrace{\text{cov}(y_2,y_2)}_{\sigma_2^2} & \text{cov}(y_2,y_3) \\ \text{cov}(y_3,y_1) & \text{cov}(y_3,y_2) & \underbrace{\text{cov}(y_3,y_3)}_{\sigma_3^2} \end{bmatrix} = \begin{bmatrix} 7 & -5 & 2 \\ -5 & \frac{25}{3} & 0 \\ 2 & 0 & 1 \end{bmatrix}$$

- $\sigma_1^2 = cov(y_1, y_1) = \frac{1}{2}\left(3^2 + (-2)^2 + (-1)^2\right) = \frac{14}{2} = 7$

- $cov(y_1, y_2) = cov(y_2, y_1) = \frac{1}{2}\left[3 \cdot \left(-\frac{5}{3}\right) + (-2) \cdot \frac{10}{3} + (-1) \cdot \left(-\frac{5}{3}\right)\right] = -\frac{10}{2} = -5$

- $cov(y_1, y_3) = cov(y_3, y_1) = \frac{1}{2}\left[3 \cdot 1 + (-2) \cdot 0 + (-1) \cdot (-1)\right] = 2$

- $\sigma_2^2 = cov(y_2, y_2) = \frac{1}{2}\left[\left(-\frac{5}{3}\right)^2 + \left(\frac{10}{3}\right)^2 + \left(-\frac{5}{3}\right)^2\right] = \frac{25}{3}$

- $cov(y_2, y_3) = cov(y_3, y_2) = \frac{1}{2}\left[\left(-\frac{5}{3}\right) \cdot 1 + \frac{10}{3} \cdot 0 + \left(-\frac{5}{3}\right) \cdot (-1)\right] = 0$

- $\sigma_3^2 = cov(y_3, y_3) = \frac{1}{2}\left[1^2 + 0^2 + (-1)^2\right] = 1$

b. 3) calculate eigenvalues and eigenvectors

$$\det(\Sigma - \lambda I) = 0 \Leftrightarrow \det\left(\begin{bmatrix} 7-\lambda & -5 & 2 \\ -5 & \frac{25}{3}-\lambda & 0 \\ 2 & 0 & 1-\lambda \end{bmatrix}\right) = 0 \Leftrightarrow$$

$$\Leftrightarrow (7-\lambda)\left(\frac{25}{3}-\lambda\right)(1-\lambda) - \left[2 \cdot 2 \cdot \left(\frac{25}{3}-\lambda\right)\right] - \left[(-5)(-5)(1-\lambda)\right] = 0 \Leftrightarrow$$

$$\Leftrightarrow \left(\frac{175}{3} - 7\lambda - \frac{25}{3}\lambda + \lambda^2\right)(1-\lambda) - \left(\frac{100}{3} - 4\lambda\right) - (25 - 25\lambda) = 0 \Leftrightarrow$$

$$\Leftrightarrow \frac{175}{3} - 7\lambda - \frac{25}{3} + \lambda^2 - \frac{175}{3}\lambda + 7\lambda^2 + \frac{25}{3}\lambda - \lambda^3 - \frac{100}{3} + 4\lambda - 25 + 25\lambda = 0 \Leftrightarrow$$

$$\Leftrightarrow -\lambda^3 + \frac{49}{3}\lambda^2 - \frac{134}{3}\lambda = 0 \Leftrightarrow \lambda_1 = 0 \lor \lambda_2 = \frac{49 + \sqrt{793}}{6} \lor \lambda_3 = \frac{49 - \sqrt{793}}{6}$$

- $\Sigma v_1 = \lambda_1 v_1 \Leftrightarrow \Sigma v_1 - \lambda_1 v_1 = 0 \Leftrightarrow (\Sigma - \lambda_1 I) v_1 = 0 \Leftrightarrow$

$$\Leftrightarrow \begin{bmatrix} 7 & -5 & 2 \\ -5 & \frac{25}{3} & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \Leftrightarrow \begin{cases} 7x - 5y + 2z = 0 \\ 2x + z = 0 \end{cases} \Leftrightarrow \begin{cases} y = \frac{3}{5}x \\ z = -2x \end{cases}$$

Assuming $x = 5$: $v_1 = \begin{bmatrix} 5 \\ 3 \\ -10 \end{bmatrix}$

Normalize: $\quad \|v_1\|_2 = \sqrt{5^2 + 3^2 + (-10)^2} = \sqrt{134}$

$$v_1' = \frac{v_1}{\|v_1\|_2} = \begin{bmatrix} \frac{5}{\sqrt{134}} \\ \frac{3}{\sqrt{134}} \\ -\frac{10}{\sqrt{134}} \end{bmatrix} \simeq \begin{bmatrix} 0,4319 \\ 0,2592 \\ -0,8639 \end{bmatrix}$$

$\bullet\ \Sigma v_2 = \lambda_2 v_2 \Leftrightarrow \Sigma v_2 - \lambda_2 v_2 = 0 \Leftrightarrow (\Sigma - \lambda_2 I) v_2 = 0 \Leftrightarrow$

$$\Leftrightarrow \begin{bmatrix} 7 - \frac{49+\sqrt{793}}{6} & -5 & 2 \\ -5 & \frac{25}{3} - \frac{49+\sqrt{793}}{6} & 0 \\ 2 & 0 & 1 - \frac{49+\sqrt{793}}{6} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} -5x + \left(\frac{25}{3} - \frac{49+\sqrt{793}}{6}\right) y = 0 \\ 2x + \left(1 - \frac{49+\sqrt{793}}{6}\right) z = 0 \end{cases} \Leftrightarrow \begin{cases} y = \dfrac{5x}{\frac{25}{3} - \frac{49+\sqrt{793}}{6}} \\ z = \dfrac{-2x}{1 - \frac{49+\sqrt{793}}{6}} \end{cases}$$

Assuming $x=1$:

$$v_2 = \begin{bmatrix} 1 \\ \dfrac{5}{\frac{25}{3} - \frac{49+\sqrt{793}}{6}} \\ \dfrac{-2}{1 - \frac{49+\sqrt{793}}{6}} \end{bmatrix}$$

Normalize:

$$\|v_2\|_2 = \sqrt{1^2 + \left(\frac{5}{\frac{25}{3} - \frac{49+\sqrt{793}}{6}}\right)^2 + \left(\frac{-2}{1 - \frac{49+\sqrt{793}}{6}}\right)^2} \simeq$$

$$\simeq 1{,}4995$$

$$v_2' = \frac{v_2}{\|v_2\|_2} \simeq \begin{bmatrix} 0{,}6669 \\ -0{,}7366 \\ 0{,}1125 \end{bmatrix}$$

$\bullet\ \Sigma v_3 = \lambda_3 v_3 \Leftrightarrow \Sigma v_3 - \lambda_3 v_3 = 0 \Leftrightarrow (\Sigma - \lambda_3 I) v_3 = 0 \Leftrightarrow$

$$\Leftrightarrow \begin{bmatrix} 7 - \frac{49-\sqrt{793}}{6} & -5 & 2 \\ -5 & \frac{25}{3} - \frac{49-\sqrt{793}}{6} & 0 \\ 2 & 0 & 1 - \frac{49-\sqrt{793}}{6} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} -5x + \left(\frac{25}{3} - \frac{49-\sqrt{793}}{6}\right) y = 0 \\ 2x + \left(1 - \frac{49-\sqrt{793}}{6}\right) z = 0 \end{cases} \Leftrightarrow \begin{cases} y = \dfrac{5x}{\frac{25}{3} - \frac{49-\sqrt{793}}{6}} \\ z = \dfrac{-2x}{1 - \frac{49-\sqrt{793}}{6}} \end{cases}$$

Assuming $k=1$:

$$V_3 = \begin{bmatrix} 1 \\ \dfrac{5}{\frac{25}{3} - \frac{49-\sqrt{793}}{6}} \\ \dfrac{-2}{1 - \frac{49-\sqrt{793}}{6}} \end{bmatrix}$$

Normalize:

$$\|v_3\|_2 = \sqrt{1^2 + \left(\dfrac{5}{\frac{25}{3} - \frac{49-\sqrt{793}}{6}}\right)^2 + \left(\dfrac{-2}{1-\frac{49-\sqrt{793}}{6}}\right)^2} \approx 1,6469$$

$$v_3' = \frac{v_3}{\|v_3\|_2} \approx \begin{bmatrix} 0,6072 \\ 0,6247 \\ 0,4910 \end{bmatrix}$$

$\boxed{K=2}$ choose 2 eigenvectors with higher eigenvalues

$$\lambda_2 > \lambda_3 > \lambda_1$$

$$v_2' = \begin{bmatrix} 0,6669 \\ -0,7366 \\ 0,1125 \end{bmatrix} \qquad v_3' = \begin{bmatrix} 0,6072 \\ 0,6247 \\ 0,4910 \end{bmatrix}$$

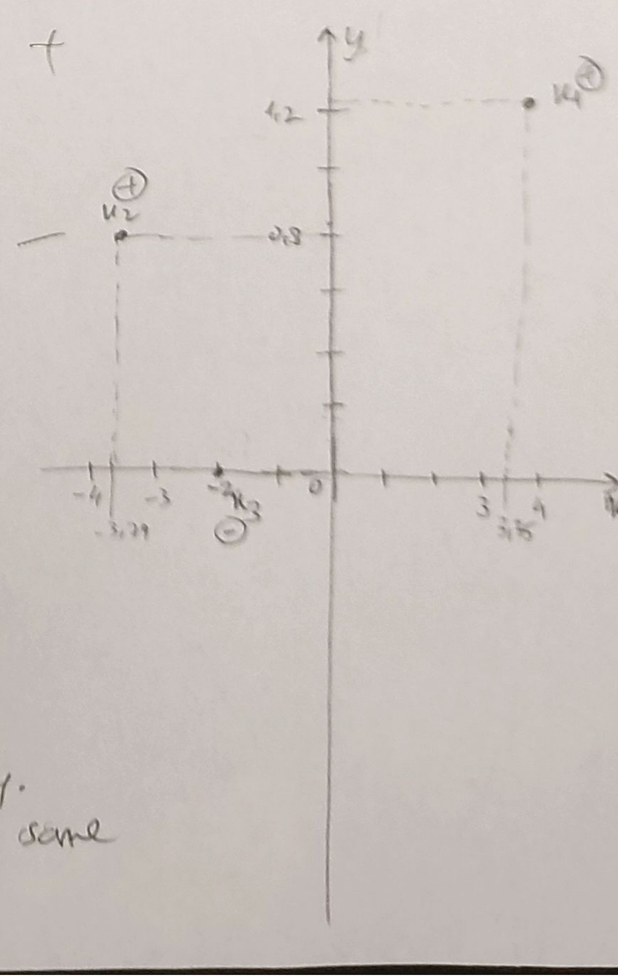
$$U = \text{Feature vector } (k=2) = \begin{array}{cc} \begin{bmatrix} 0,6669 & 0,6072 \\ -0,7366 & 0,6247 \\ 0,1125 & 0,4910 \end{bmatrix} \\ \quad v_2' \qquad\quad v_3' \end{array}$$ defines the PCA plane

c. $x_i' = U^T x_{centered}$ class (y_i)

$$x_1' = \begin{bmatrix} 0,6669 & -0,7366 & 0,1125 \\ 0,6072 & 0,6247 & 0,4910 \end{bmatrix} \begin{bmatrix} 3 \\ -5/3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3,3409 \\ 1,2714 \end{bmatrix} \quad +$$

$$x_2' = \begin{bmatrix} 0,6669 & -0,7366 & 0,1125 \\ 0,6072 & 0,6247 & 0,4910 \end{bmatrix} \begin{bmatrix} -2 \\ 10/3 \\ 0 \end{bmatrix} = \begin{bmatrix} -3,7891 \\ 0,8679 \end{bmatrix} \quad +$$

$$x_3' = \begin{bmatrix} 0,6669 & -0,7366 & 0,1125 \\ 0,6072 & 0,6247 & 0,4910 \end{bmatrix} \begin{bmatrix} -1 \\ 5/3 \\ -1 \end{bmatrix} = \begin{bmatrix} -2,0071 \\ -0,0570 \end{bmatrix} \quad -$$



Yes, this projection plane discriminates the 2 classes for the given points. From the plot, points x_1 and x_2 (class +) have $y>0$ while x_3 (class −) has $y<0$, so the horizontal line $y=0$ perfectly separates the 3 samples. Therefore a linear decision boundary ($y=0$) exists that classifies all three points correctly.

Note: this conclusion holds for the provided dataset only. With additional points the classes could overlap and the same projection might no longer separate them.