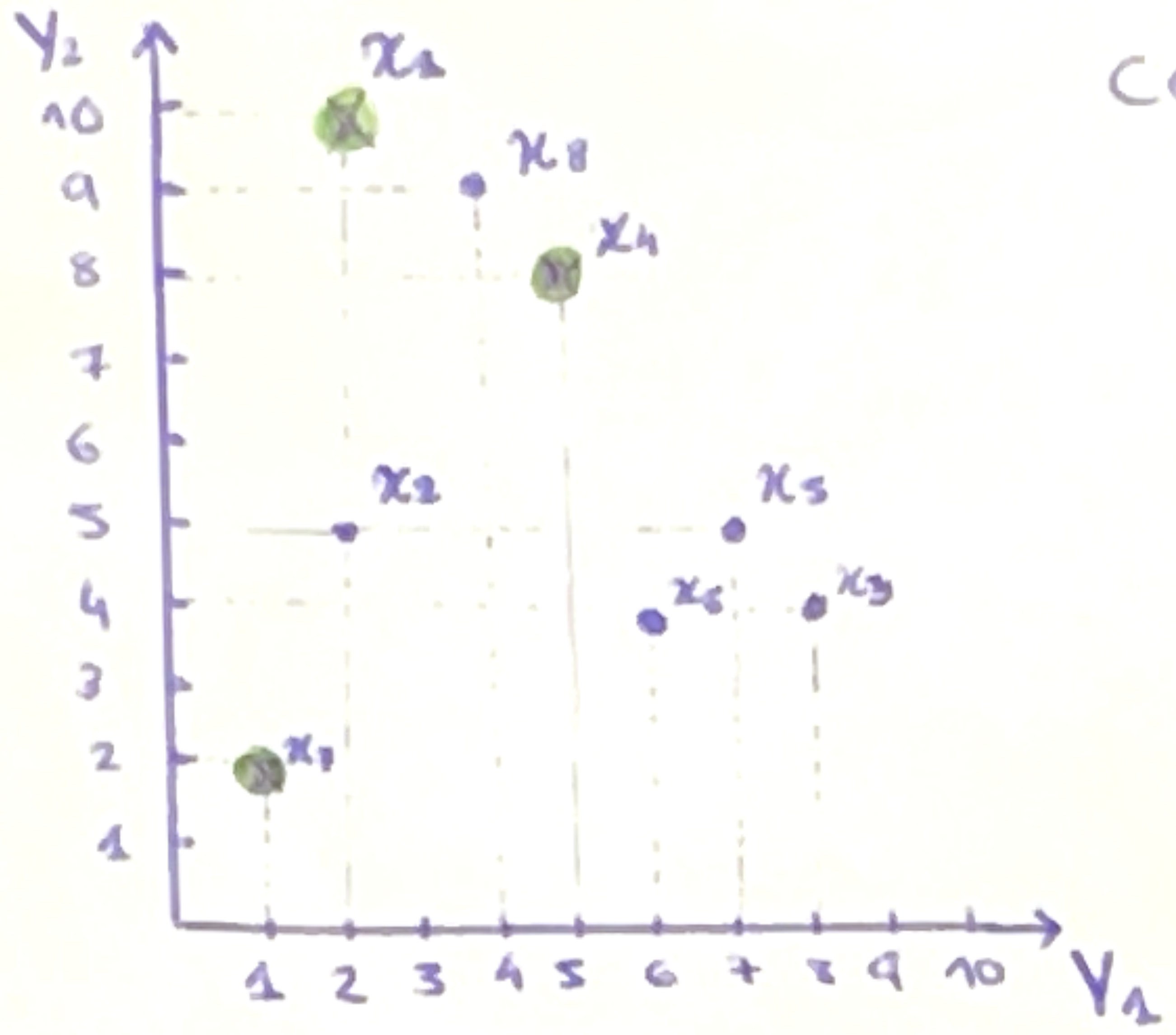
Homework IV Francisco Silva ist1110409 Joana Cardoso ist1109864

Pen and Paper

Part A - Clustering

a.



centróides:
$x_1(2,10)$
$x_4(5,8)$
$x_7(1,2)$

b. $K=3$; 3 centroides => 3 clusters

1) "Assign points to clusters"

$x_1 \rightarrow$ Cluster 1 (C_1)
$x_4 \rightarrow$ cluster 2 (C_2)
$x_7 \rightarrow$ cluster 3 (C_3)

$\boxed{x_1 \rightarrow C_1}$

$x_2(2,5)$:

$d(x_1,x_2)^2 = (2-2)^2+(10-5)^2 = 5^2 = 25$ $d(x_4,x_2)^2=(2-5)^2+(5-8)^2=18$

$d(x_7,x_2)^2=(1-2)^2+(2-5)^2=\underline{10}$ logo $\boxed{x_2 \rightarrow C_3}$

$x_3(8,4)$:

$d(x_1,x_3)^2=(2-8)^2+(10-4)^2=72$ $d(x_4,x_3)^2=(5-8)^2+(8-4)^2=\underline{25}$

$d(x_7,x_3)^2=(1-8)^2+(2-4)^2=53$ logo $\boxed{x_3 \rightarrow C_2}$

$\boxed{x_4 \rightarrow C_2}$

$x_5(7,5)$:

$d(x_1,x_5)^2=(2-7)^2+(10-5)^2=50$ $d(x_4,x_5)^2=(5-7)^2+(8-5)^2=\underline{13}$

$d(x_7,x_5)^2=(1-7)^2+(2-5)^2=45$ logo $\boxed{x_5 \rightarrow C_2}$

$x_6(6,4)$:

$d(x_1,x_6)^2=(2-6)^2+(10-4)^2=52$ $d(x_4,x_6)^2=(5-6)^2+(8-4)^2=\underline{17}$

$d(x_7,x_6)^2=(1-6)^2+(2-4)^2=29$ logo $\boxed{x_6 \rightarrow C_2}$

~~$x_7(1,2)$~~

$\boxed{x_7 \rightarrow C_3}$

$x_8(4,9)$:

$d(x_1,x_8)^2=(2-4)^2+(10-9)^2=5$ $d(x_4,x_8)^2=(5-4)^2+(8-9)^2=\underline{2}$

$d(x_7,x_8)^2=(1-4)^2+(2-9)^2=58$ logo $\boxed{x_8 \rightarrow C_2}$

Assim temos:
$C_1: x_1$; $C_2: x_3, x_4, x_5, x_6, x_8$; $C_3: x_2; x_7$

3

2) "Adjust centroids" → ci a média dos pontos atribuídos a cada ~~centróide~~ cluster, sejam μ_1, μ_2, μ_3 os novos centróides dos clusters

C_1:

só tem um ponto, logo o centróide mantém-se. (pq o ponto é o centróide)

$\mu_1 = x_1 = (2, 10)$

C_2: $\{x_3, x_4, x_5, x_6, x_8\}$

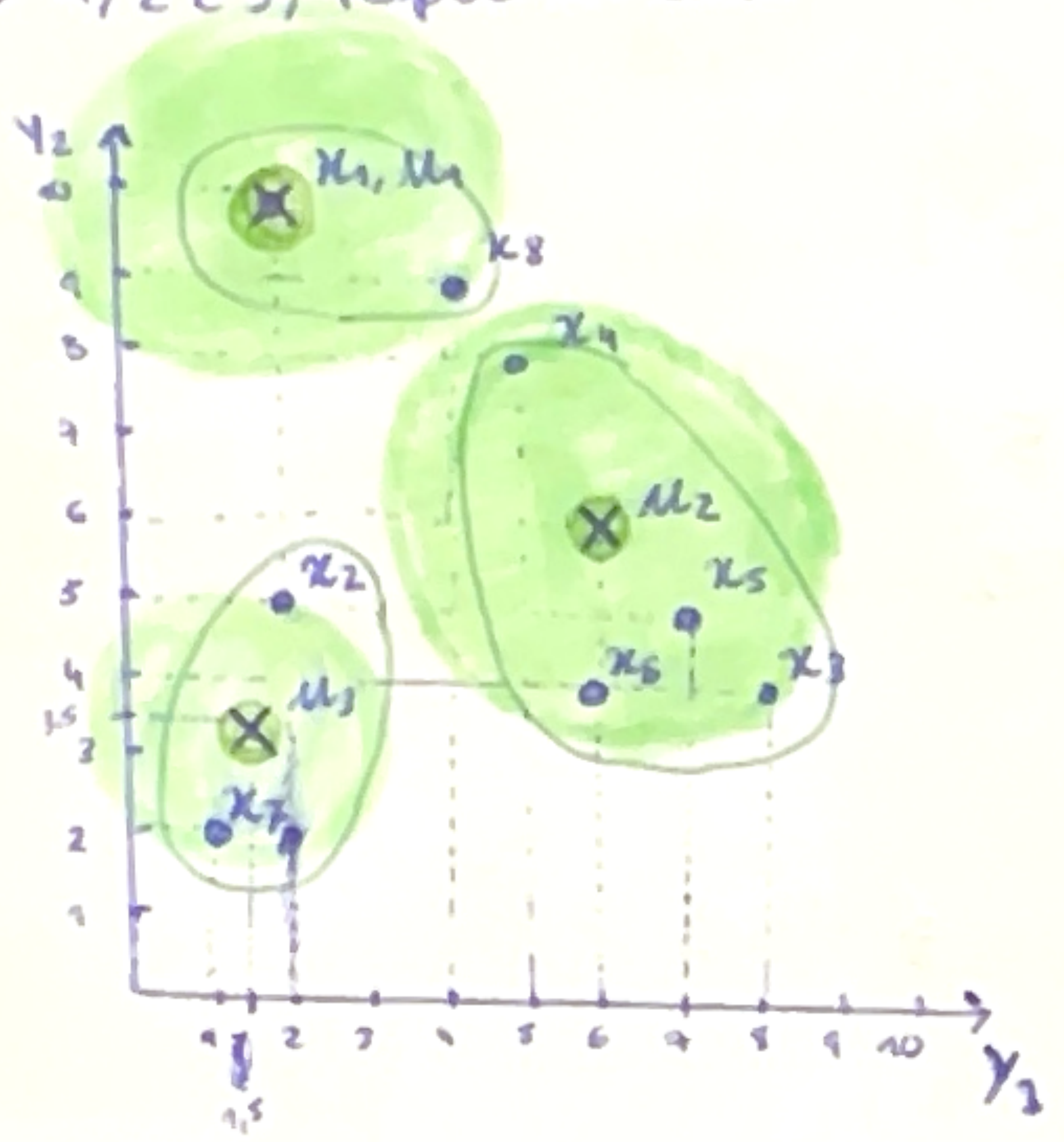
$$\mu_2 = \begin{bmatrix} \frac{8+5+7+6+4}{5} \\ \frac{4+8+5+4+9}{5} \end{bmatrix}^t = (6,6)$$

C_3: $\{x_2, x_7\}$

$$\mu_3 = \begin{bmatrix} \frac{2+1}{2} \\ \frac{5+2}{2} \end{bmatrix}^t = (1{,}5; 3{,}5)$$

R: Os novos centroids são $\mu_1 = (2,10)$; $\mu_2 = (6,6)$; $\mu_3 = (1{,}5; 3{,}5)$ dos clusters 1, 2 e 3, respetivamente

c)



$C_1 = \{x_1, x_8\}$

$C_2 = \{x_3, x_4, x_5, x_6\}$

$C_3 = \{x_2, x_7\}$

Os clusters são círculos, centrados nos centroids e raio de, respetivamente c1, c2, c3:

: $\sqrt{5}$; $\sqrt{8}$; $\sqrt{2{,}5}$

Representados por

(ñ a linha verde)

Para os novos centróides calculados, temos de voltar a atribuir cada observação aos clusters

$x_1 = \mu_1$ logo $x_1 \to C_1$

x_2:

$d(\mu_1, x_2)^2 = 25$; $d(\mu_2, x_2)^2 = (2-6)^2 + (5-6)^2 = 17$

$d(\mu_3, x_2)^2 = (2-1{,}5)^2 + (5-3{,}5)^2 = 2{,}5$; $x_2 \to C_3$

x_3:

$d(\mu_1, x_3)^2 = 72$; $d(\mu_2, x_3)^2 = (8-6)^2 + (4-6)^2 = 8$

$d(\mu_3, x_3)^2 = (8-1{,}5)^2 + (4-3{,}5)^2 = 42{,}5$; $x_3 \to C_2$

x_4:

$d(\mu_1, x_4)^2 = 13$; $d(\mu_2, x_4)^2 = (5-6)^2 + (8-6)^2 = 5$

$d(\mu_3, x_4)^2 = (5-1{,}5)^2 + (8-3{,}5)^2 = 32{,}5$; $x_4 \to C_2$

x_5:

$d(\mu_1, x_5)^2 = 50$; $d(\mu_2, x_5)^2 = (7-6)^2 + (5-6)^2 = 2$

$d(\mu_3, x_5)^2 = (7-1{,}5)^2 + (5-3{,}5)^2 = 32{,}5$; $x_5 \to C_2$

x_6:

$d(\mu_1, x_6)^2 = 52$; $d(\mu_2, x_6)^2 = (6-6)^2 + (4-6)^2 = 2$

$d(\mu_3, x_6)^2 = (6-1{,}5)^2 + (4-3{,}5)^2 = 20{,}5$ $x_6 \to C_2$

x_7:

$d(\mu_1, x_7)^2 = 65$; $d(\mu_2, x_7)^2 = (1-6)^2 + (2-6)^2 = 41$

$d(\mu_3, x_7)^2 = (1-1{,}5)^2 + (2-3{,}5)^2 = 2{,}5$; $x_7 \to C_3$

x_8:

$d(\mu_1, x_8)^2 = 5$; $d(\mu_2, x_8)^2 = (4-6)^2 + (9-6)^2 = 13$

$d(\mu_3, x_8)^2 = (4-1{,}5)^2 + (9-3{,}5)^2 = 36{,}5$ $x_8 \to C_1$

d) O algoritmo K-means está muito dependente da inicialização dos centróides, afetando diretamente a convergência e a performance do modelo

- Tempo de convergência (tempo até terminar o algoritmo e não ~~ser~~ serem necessárias mais épocas pois os centroides não alteraram)

Uma inicialização mais próxima da solução final ótima requer menos iterações/épocas para os centróides estabilizarem as suas posições.

Pelo contrário, uma inicialização pobre pode exigir um número muito maior de épocas para que os centróides se desloquem através do espaço de dados até encontrarem as suas posições finais, tornando o processo computacionalmente mais lento

- Performance (desempenho/qualidade do algoritmo, ou seja, se chegou ou não à distribuição das observações por clusters mais acertada)

Há algumas circunstâncias em que, devido à inicialização dos centróides, a performance do modelo fica comprometida:

→ Se forem inicializados muito próximos uns dos outros, o algoritmo pode ter dificuldade a separar clusters que são naturalmente distintos.

→ Se forem inicializados equidistantemente dos ~~k~~ k clusters, também a performance pode ficar comprometida

→ Uma inicialização de um centróide que nenhum ponto de dados lhe seja o mais próximo, pode levar à formação de cluster vazios, pois o centróide fica "preso" e incapaz de se atualizar (já que só faz média c/ ele mesmo)

Para além disso, como o k-means é um algoritmo greedy, há certas inicializações de centróides que podem fazer o modelo convergir para mínimos locais, não atingindo a solução ótima e separação natural dos dados.

Part B - PCA

a) 1) centrar os dados

$$\bar{x} = \frac{1}{3}\left[\begin{bmatrix}5\\0\\1\end{bmatrix}+\begin{bmatrix}0\\5\\0\end{bmatrix}+\begin{bmatrix}1\\0\\-1\end{bmatrix}\right] = \begin{bmatrix}6/3\\5/3\\0/3\end{bmatrix} = \begin{bmatrix}2\\5/3\\0\end{bmatrix}\begin{matrix}\to y_1\\ \to y_2\\ \to y_3\end{matrix}$$

$$x_i' = x_i - \bar{x} \quad x_1' = \begin{bmatrix}5\\0\\1\end{bmatrix} - \begin{bmatrix}2\\5/3\\0\end{bmatrix} = \begin{bmatrix}3\\-5/3\\1\end{bmatrix} \quad x_2' = \begin{bmatrix}0\\5\\0\end{bmatrix} - \begin{bmatrix}2\\5/3\\0\end{bmatrix} = \begin{bmatrix}-2\\10/3\\0\end{bmatrix} \quad x_3' = \begin{bmatrix}1\\0\\-1\end{bmatrix} - \begin{bmatrix}2\\5/3\\0\end{bmatrix} = \begin{bmatrix}-1\\-5/3\\-1\end{bmatrix}$$

matriz das covariâncias.

$$\begin{bmatrix}var(y_1) & cov(y_1,y_2) & cov(y_1,y_3)\\ cov(y_2,y_1) & var(y_2) & cov(y_2,y_3)\\ cov(y_3,y_1) & cov(y_3,y_1) & var(y_3)\end{bmatrix}$$

$$var(y_1) = cov(y_1,y_1) = \frac{(5-2)^2+(0-2)^2+(1-2)^2}{3-1} = \frac{14}{2} = 7$$

$$var(y_2) = cov(y_2,y_2) = \frac{(0-5/3)^2+(5-5/3)^2+(0-5/3)^2}{3-1} = \frac{50/3}{2} = 8,(3)$$

$$var(y_3) = cov(y_3,y_3) = \frac{(1-0)^2+(0)^2+(-1-0)^2}{3-1} = 1$$

$$cov(y_1,y_2) = cov(y_2,y_1) = \frac{(5-2)(-5/3)+(-2)(5-5/3)+(-5/3)(1-2)}{3-1} = \frac{-10}{2} = -5$$

$$cov(y_1,y_3) = cov(y_3,y_1) = \frac{(5-2)(1)+(-2)\times 0+(1-2)\times(-1)}{3-1} = \frac{4}{2} = 2$$

$$cov(y_2,y_3) = cov(y_3,y_2) = \frac{(-5/3)(1)+(5-5/3)(0)+(-5/3)(-1)}{3-1} = \frac{0}{2} = 0$$

Assim, temos a matriz das covariâncias:

$$\Sigma = \begin{bmatrix}7 & -5 & 2\\ -5 & 8,(3) & 0\\ 2 & 0 & 1\end{bmatrix}$$

b) calcular valores e vetores próprios de Σ

valores próprios

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

$$\Leftrightarrow 2\left(-\left(\frac{50}{3}-2\lambda\right)\right) + (1-\lambda)\left((7-\lambda)\left(\frac{25}{3}-\lambda\right)-25\right) = 0 \Leftrightarrow -\frac{100}{3} + 4\lambda + \frac{100}{3} - \frac{146}{3}\lambda + \frac{49}{3}\lambda^2 - \lambda^3 = 0 \Leftrightarrow$$

$$\Leftrightarrow -\lambda^3 + \frac{49}{3}\lambda^2 - \frac{134}{3}\lambda = 0 \Leftrightarrow -\lambda\left(\lambda^2 - \frac{49}{3}\lambda + \frac{134}{3}\right) = 0 \Leftrightarrow \lambda_3 = 0 \vee \lambda^2 - \frac{49}{3}\lambda + \frac{134}{3} = 0 \Leftrightarrow$$

$$\Leftrightarrow \lambda_3 = 0 \vee \lambda_1 = 3,47 \vee \lambda_2 = 12,86$$

Quanto maior o valor próprio, maior a variância dos dados, explicada pelo vetor próprio correspondente.

Como queremos o plano (2 dim) de projeção, escolhemos os 2 maiores valores próprios $\lambda_1 \wedge \lambda_2$ escolhidos, vamos calcular os correspondentes vetores próprios, u_1 e u_2

Para $\lambda_1 = 3,47$, vetor próprio u_1

$$(\Sigma - \lambda_1 I)u_1 = 0 \Leftrightarrow \begin{bmatrix}7-3,47 & -5 & 2\\ -5 & 25/3-3,47 & 0\\ 2 & 0 & 1-3,47\end{bmatrix}\begin{bmatrix}u_{11}\\u_{12}\\u_{13}\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix} \Leftrightarrow$$

$$\begin{cases}3,53u_{11} - 5u_{12} + 2u_{13} = 0\\ -5u_{11} + 5,83u_{12} = 0\\ 2u_{11} - 2,47u_{13} = 0\end{cases} \Leftrightarrow \begin{cases}\text{---}\\ u_{12} = 1,028\,u_{11}\\ u_{13} = 0,8097\,u_{11}\end{cases}$$

se $u_{11} = 1$; $u_{12} = 1,028$; $u_{13} = 0,8097$

$$u_1 = (1;\ 1,0282;\ 0,8097) \quad \|u_1\| = \sqrt{1+1,0572+0,6556} = 1,647$$

$$u_{1\,norm} \approx \frac{u_1}{\|u_1\|} \approx (0,607;\ 0,624;\ 0,492)$$

Para $\lambda_2 = 12,86$, vetor próprio u_2

$$(\Sigma - \lambda_2 I)u_2 = 0 \iff \begin{bmatrix} 7-12,86 & -5 & 2 \\ -5 & 23/3-12,86 & 0 \\ 2 & 0 & 1-12,86 \end{bmatrix}\begin{bmatrix} u_{21} \\ u_{22} \\ u_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \iff \begin{cases} -5,86u_{21} - 5u_{22} + 2u_{23} = 0 \\ -5u_{21} - 4,527u_{22} = 0 \iff \\ 2u_{21} - 11,86u_{23} = 0 \end{cases}$$

$$\iff \begin{cases} u_{22} = -1,105\,u_{21} \\ u_{23} = 0,1686\,u_{21} \end{cases}$$

se $u_{21} = 1$; $u_{22} = -1,105$; $u_{23} = 0,1686$

$u_2 = (1; -1,105; 0,1686)$ $\quad \|u_2\| = \sqrt{1+1,22+0,0284} \approx 1,5$

$$u_{2,norm} = \frac{u_2}{\|u_2\|} = (0,667; -0,737; 0,1124)$$

u_1 e u_2 são vetores linearmente independentes (perpendiculares entre si) $u_1 \cdot u_2 = 0$ e juntos formam um plano, ou seja, constituem a base ortonormal ideal para projetar, em que qualquer combinação linear $a u_1 + b u_2$ $(a, b \in \mathbb{R}^2)$ dá um ponto desse plano.

$\text{Span}\{(0,607; 0,624; 0,492), (0,667; -0,737; 0,1124)\}$, → plano ótimo 2D

c)

$x' = U^T x$, $\quad U = \begin{bmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{bmatrix}$, seja $u_3(a, b, c)$ (irrelevante)

$$= \begin{bmatrix} 0,607 & 0,667 & a \\ 0,627 & -0,737 & b \\ 0,492 & 0,1124 & c \end{bmatrix} \quad U^T = \begin{bmatrix} 0,607 & 0,627 & 0,492 \\ 0,667 & -0,737 & 0,1124 \\ a & b & c \end{bmatrix}$$

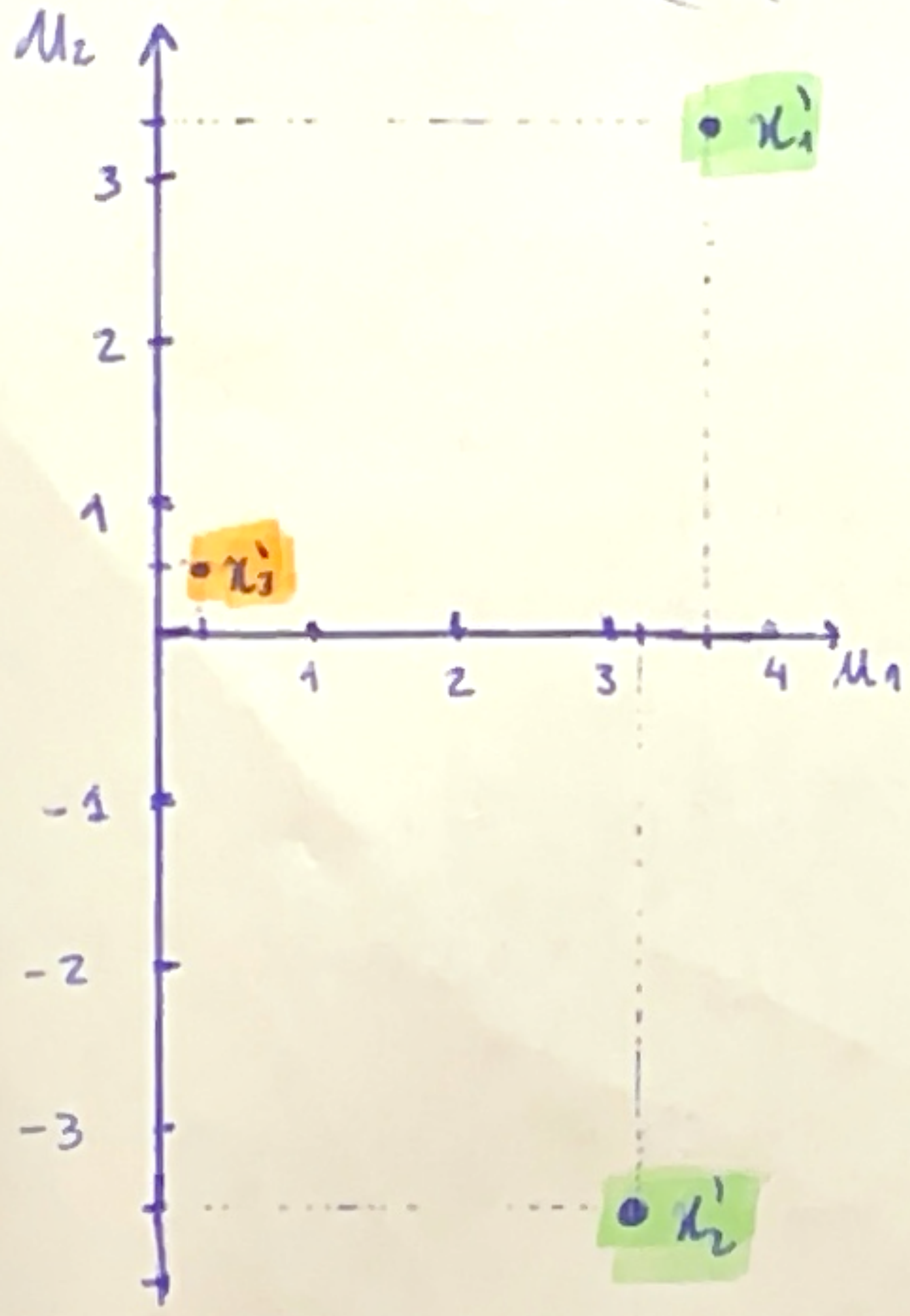
$$x_1' = U^T\begin{bmatrix} 5 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 5\times 0,607 + 0,492 \\ 5\times 0,667 + 0,1124 \\ 5a + c \end{bmatrix} = \begin{bmatrix} 3,527 \\ 3,4474 \\ 5a+c \end{bmatrix} \longrightarrow \begin{bmatrix} 3,527 \\ 3,447 \end{bmatrix} \; (3,527; 3,447) \to x_1'$$

→ relativo a λ_3, descartado

$$x_2' = U^T\begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 5\times 0,627 \\ 5\times(-0,737) \\ 5b \end{bmatrix} = \begin{bmatrix} 3,135 \\ -3,685 \\ \cancel{5b} \end{bmatrix} \longrightarrow \begin{bmatrix} 3,135 \\ -3,685 \end{bmatrix} \; (3,135; -3,685) \to x_2'$$

$$x_3' = U^T\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0,607 - 0,492 \\ 0,667 - 0,1124 \\ a - c \end{bmatrix} = \begin{bmatrix} 0,115 \\ 0,5546 \\ \cancel{a-c} \end{bmatrix} \longrightarrow \begin{bmatrix} 0,115 \\ 0,5546 \end{bmatrix} \; (\cancel{3,135}) \; (0,115; 0,5546) \to x_3'$$

a primeira coordenada é relativa a u_1 e a segunda a u_2



Pos (+)

Neg (−)

Podemos analisar os pontos no plano de projeção, no entanto apenas 3 pontos é limitado para garantir um padrão discriminatório.

Podemos interpretar que os pontos positivos estão mais longe da origem enquanto o ponto de classe negativa x_3' está mais perto da origem.