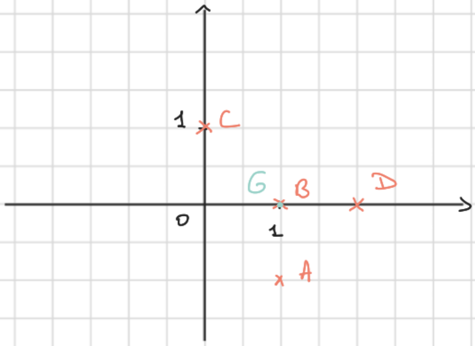
Exercice 1 Inertie suivant un axe, vers l'ACP

Soit un nuage de points de matrice X :

$$\text{individus}\quad X = \begin{array}{c} A \\ B \\ C \\ D \end{array}\begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{pmatrix}.$$

1. Construire le nuage.
2. Calculer le centre d'inertie et l'inertie totale.
3. Centrer le nuage de points.
4. Calculer l'opérateur de projection sur un axe de vecteur unitaire $u(a, b)$.
5. En déduire les coordonnées des projetés suivant les axes de vecteur directeur $(1, 0)$, $(0, 1)$, $(1, 1)$ et leur affixe suivant u (il y a un piège).
6. Calculer l'inertie projetée suivant les axes de vecteur directeur $(1, 0)$, $(0, 1)$, $(\sqrt{2}/2, \sqrt{2}/2)$.
7. Trouver l'axe suivant lequel l'inertie projetée est maximale.
8. Retrouver les résultats avec R et/ou Python.

1.



2. $g = (1, 0)$

$$\frac{1+1+0+2}{4} = 1$$

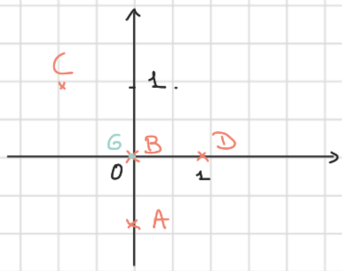
$$\frac{-1+0+1+0}{4} = 0$$

moyenne des variables (colonnes)

$$I_T = \sum_{i=1}^{4} p_i \, d^2(g, i) = \frac{1}{4}(1+0+2+1) = 1.$$

3.

$$X_c = \begin{pmatrix} 0 & -1 \\ 0 & 0 \\ -1 & 1 \\ 1 & 0 \end{pmatrix}$$



$\rightarrow$ nuage centré.

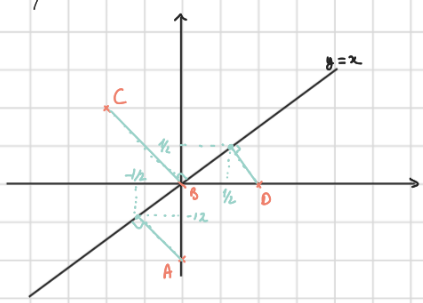
4. $\Pi_u = u u^T = \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} a & b \end{pmatrix} = \begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix}$

$$\vec{OM'} = \Pi_u \vec{OM}$$

$u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $\quad \Pi_{u_1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ $\quad X_{u_1} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ -1 & 0 \\ 1 & 0 \end{pmatrix}$ et avec $u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ $\quad X_{u_2} = \begin{pmatrix} 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$

$u_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\rightarrow$ on le norme.

$\Pi_{u_3} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ $\quad X_{u_3} = \begin{pmatrix} -1/2 & -1/2 \\ 0 & 0 \\ 0 & 0 \\ 1/2 & 1/2 \end{pmatrix}$



6. $u_1 = \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \end{pmatrix}$ $\rightarrow$ $I_{u_1} = \frac{1}{4} \times 2 = \frac{1}{2}$

$u_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ $\rightarrow$ $I_{u_2} = \frac{1}{2}$

$u_3 = \begin{pmatrix} -\sqrt{2}/2 \\ 0 \\ 0 \\ \sqrt{2}/2 \end{pmatrix}$ $\rightarrow$ $I_{u_3} = \frac{1}{4}\left(\frac{1}{2} + 0 + 0 + \frac{1}{2}\right) = \frac{1}{4}$

7 - Affixe suivant O : X_u

Inertie I_u = variance de X_u : $\frac{1}{4} u^T X^T X u = u^T \left(\frac{1}{4} X^T X\right) u.$

Max de $\underbrace{u^T\left(\frac{1}{4} X^T X\right) u}_{f(u)}$ sous la contrainte $\underbrace{\|u\|^2}_{h(u)} = 1$

On cherche λ et u tq : $\begin{cases} \nabla f(u) + \lambda \nabla h(u) = 0. \\ u^T u = 1 \end{cases}$

Exercice 2 Inertie, vers la classification

On définit 6 points $M_1(1,0), M_2(1,2), M_3(3,6), M_4(3,0), M_5(4,5)$ et $M_6(6,5)$.

1. Placer les points dans un repère. Calculer l'inertie du nuage.
2. Calculer le tableau des distances entre points.
3. Méthode des centres mobiles

   On définit deux groupes en affectant le point au groupe 1 si il est plus proche de M_1 que de M_4 et au groupe 2 sinon.

   (a) Déterminer le centre de gravité C_1^1 et C_2^2 et l'inertie intra de chacun des groupes obtenus.

   (b) Continuer jusqu'à stabilité des groupes. Cet algorithme est appelée méthode des centres mobiles ou k-means.

   (c) Construire un tableau indiquant l'inertie inter et intra à chaque étape. Qu'observe-t-on ?

4. (a) CAH : classification ascendante hiérarchique

   On regroupe à partir du tableau des distances les deux points les plus proches en une classe c_7. On calcule alors la distance des autres points à cette classe en prenant le critère d'agrégation correspondant à la plus petite distance obtenue avec chacun des points.

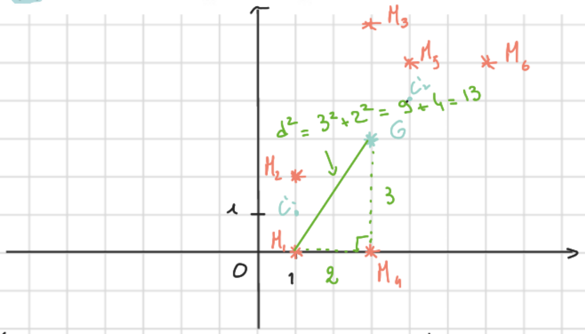
   (b) Déterminer le nouveau tableau de distance obtenu après ce premier regroupement.

   (c) Réitérer l'opération jusqu'à n'obtenir qu'une classe.

   (d) Construire le dendogramme obtenu en construisant l'arbre représentant les regroupements successifs avec comme ordonnée la distance correspondant au regroupement.

   (e) En déduire par coupure de l'arbre une classification en deux classes.

1.



$X = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ 3 & 6 \\ 3 & 0 \\ 4 & 5 \\ 6 & 5 \end{pmatrix}$ $g = \frac{1}{6}(18, 18) = (3, 3)$

$I_T = \frac{1}{6}\left(13 + 9 + 5 + 5 + 13 + 9\right) = 9$

p_i

2 -

d^2	M_1	M_2	M_3	M_4	M_5	M_6
M_1	0	4	40	4	34	50
M_2	/	0	20	8	18	34
M_3	/	/	0	36	2	10
M_4	/	/	/	0	26	34
M_5	/	/	/	/	0	4
M_6	/	/	/	/	/	0

→ distance au carré.

→ question 4 (a) : la plus petite distance donc 1er groupe $M_3 M_5$

3. <u>Etape 0</u> : groupe 1 = M_1, M_2 (étape 0 : initialisation)
groupe 2 = M_3, M_4, M_5, M_6

<u>Etape 1</u> : $C_1^1 = (1,1)$: M_1, M_2, M_4 (nouveaux groupes, M_4 plus proche de C_1 que de C_2^1)
$C_2^1 = (4,4)$: M_3, M_5, M_6

<u>Etape 2</u> : $C_1^2 = (5/3, 2/3)$ Les groupes ne changent pas : → stabilisation des groupes
$C_2^2 = (13/3, 16/3)$

c) <u>étape 0</u> : $I_{intra}^{(0)} = \frac{1}{6}(1+1) + \frac{1}{6}(17+5+5+1) = \frac{30}{6} = 5$
poids des individus

$I_{inter}^{(0)} = I_{total} - I_{intra}^{(0)} = 9 - 5 = 4.$

<u>étape 1</u> : $I_{intra}^{(1)} \searrow$ $I_{inter}^{(0)} \nearrow$

4 (a) groupe : $M_3 - M_5$

d^2	M_2	M_4	M_6	(M_3M_5)
M_1	4	4	50	34
M_2	/	8	34	18
M_4	/	/	34	26
M_6	/	/	/	4

→ pour la dernière colonne on prend la plus petite distance entre $M_3 - M_i$ et $M_5 - M_i$

<u>Etape 2</u> : groupe : $M_1 - M_2$

d^2	M_6	(M_3M_5)	(M_1M_2)
M_4	34	26	4
M_6	/	4	34
(M_3M_5)	/	/	18

→ <u>prochain étape</u> : groupe : $M_3M_5 - M_6$

d^2	(M_1M_2)	$(M_3M_5M_6)$
M_4	4	26
(M_1M_2)	/	18

d^2	$M_1M_2M_4$
$M_3M_5M_6$	18

→ <u>prochain étape</u> : $(M_1M_2M_4 - M_3M_5M_6)$ fin

(d)

→ on en déduit 2 groupes.