Exercices SY09
Centrage et distance

1 Centrage

Exercice 1 Soit x un vecteur de taille n. On note $\overline{x}$ la moyenne empirique de l'échantillon x,

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

et on dit que x est centré lorsque $\overline{x} = 0$.

1. Montrer que x est un vecteur centré si et seulement si $\mathbb{1}_n^T x = 0$ avec $\mathbb{1}_n^T = [1, \ldots, 1]$.
2. On pose $x^\dagger = x - \overline{x}\mathbb{1}_n$. Montrer que $x^\dagger$ est centré.

On appellera $x^\dagger$ le centrage de x.

3. Montrer que le centrage est une opération linéaire :
   (a) $(x + y)^\dagger = x^\dagger + y^\dagger$
   (b) $(\lambda x)^\dagger = \lambda x^\dagger$
4. Montrer que $x^\dagger = 0$ si et seulement si $x = \lambda\mathbb{1}_n$ avec $\lambda \in \mathbb{R}$.

Autrement dit, le centrage d'un vecteur n'est jamais nul sauf si le vecteur est constant.

5. Montrer que x et y ont même centrage si et seulement si il existe $\lambda \in \mathbb{R}$ tel que $y = x + \lambda\mathbb{1}_n$.
6. Montrer que

$$\begin{pmatrix} x \\ y \end{pmatrix}^\dagger = \begin{pmatrix} x^\dagger \\ y^\dagger \end{pmatrix} \quad \text{si et seulement si} \quad \overline{x} = \overline{y}.$$

2 Distance

Exercice 2 Préciser la proximité que représente chacune des matrices suivantes.

1. $\begin{pmatrix} 0 & 1 & 2 \\ 3 & 0 & -1 \\ 1 & 2 & 0 \end{pmatrix}$ 2. $\begin{pmatrix} 1 & 3 & 2 \\ 3 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}$ 3. $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$

4. $\begin{pmatrix} 0 & 2 & 3 \\ 2 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix}$

5. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

6. $\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$

7. $\begin{pmatrix} 10 & 1 & 3 \\ 6 & 9 & 2 \\ 5 & 0 & 8 \end{pmatrix}$

8. $\begin{pmatrix} 10 & 4 & -1 \\ 4 & 10 & 5 \\ -1 & 5 & 10 \end{pmatrix}$

Exercice 3 Soit Ω est ensemble muni d'une ultramétrique d. Montrer que tout triangle est soit équilatéral soit isocèle avec une petite base.

Exercice 4 Montrer que la distance qui vaut tout le temps 1 sauf pour deux éléments identiques où elle vaut 0 est une distance ultramétrique.

Exercice 5 On suppose que $X_1, \ldots, X_n$ sont n caractéristiques binaires d'une population Ω. On note x_i la i-ième caractéristique de l'individu x et on considère la similarité entre deux individus x et y suivante,

$$s(x,y) = \frac{a+d}{a+d+b+c},$$

où

$$\begin{aligned} a &= \mathrm{card}\{i, x_i = 1, y_i = 1\} \\ d &= \mathrm{card}\{i, x_i = 0, y_i = 0\} \\ b &= \mathrm{card}\{i, x_i = 0, y_i = 1\} \\ c &= \mathrm{card}\{i, x_i = 1, y_i = 0\} \end{aligned}$$

On pose $d = 1 - s$.

1. Montrer que
$$d(x,y) = \frac{b+c}{n}$$
2. Montrer que d vérifie les propriétés de séparation et de symétrie.
3. On note
$$\begin{aligned} A &= \mathrm{card}\{i, x_i = 0, y_i = 0, z_i = 0\} \\ B &= \mathrm{card}\{i, x_i = 0, y_i = 0, z_i = 1\} \\ C &= \mathrm{card}\{i, x_i = 0, y_i = 1, z_i = 0\} \\ D &= \mathrm{card}\{i, x_i = 0, y_i = 1, z_i = 1\} \\ E &= \mathrm{card}\{i, x_i = 1, y_i = 0, z_i = 0\} \\ F &= \mathrm{card}\{i, x_i = 1, y_i = 0, z_i = 1\} \\ G &= \mathrm{card}\{i, x_i = 1, y_i = 1, z_i = 0\} \\ H &= \mathrm{card}\{i, x_i = 1, y_i = 1, z_i = 1\} \end{aligned}$$
   (a) Exprimer $d(x,y)$, $d(y,z)$ et $d(x,z)$ en fonction de A, B, C, D, E, F, G et H.
   (b) En déduire que d est une distance