1 Espaces métriques

Définition 1.1. *Un **espace métrique** est un couple (E, d) où E est un ensemble non vide et d est une application, appelée* distance *ou* métrique*,*

$$\mathcal{D} : E \times E \to \mathbb{R}^+$$

telle que, pour tous $x, y, z \in E$, les propriétés suivantes sont vérifiées :

1. ***Séparation*** *: $\mathcal{D}(x, y) = 0 \iff x = y$;*
2. ***Symétrie*** *: $\mathcal{D}(x, y) = \mathcal{D}(y, x)$;*
3. ***Inégalité triangulaire*** *: $\mathcal{D}(x, z) \leq \mathcal{D}(x, y) + \mathcal{D}(y, z)$.*

Note : la positivité $\mathcal{D}(x, y) \geq 0$ est une conséquence des autres axiomes[1]

2 Espaces vectoriels normés

Définition 2.1. *Soit E un espace vectoriel sur le corps $\mathbb{K}$ ($\mathbb{R}$ ou $\mathbb{C}$). Une **norme** sur E est une application*

$$\|\cdot\| : E \to \mathbb{R}^+$$

telle que, pour tous vecteurs $\mathbf{x}, \mathbf{y} \in E$ et tout $\lambda \in \mathbb{K}$, on a :

1. ***Séparation*** *: $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$;*
2. ***Homogénéité*** *: $\|\lambda\mathbf{x}\| = |\lambda| \, \|\mathbf{x}\|$;*
3. ***Inégalité triangulaire*** *(ou sous-additivité) : $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.*

*Un espace vectoriel muni d'une norme est appelé un **espace vectoriel normé**.*

Proposition 2.1. *Soit $(E, \|\cdot\|)$ un espace vectoriel normé. $\mathcal{D} : E \times E \to \mathbb{R}^+$ définie par*

$$\mathcal{D}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$$

est une distance sur E. Ainsi, tout espace vectoriel normé est un espace métrique.

Donnons un exemple. Soit l'espace vectoriel $E = \mathbb{R}^d$. Pour un vecteur $\mathbf{x} = (x_1, \ldots, x_d)$, la **norme euclidienne** (ou norme L2) est définie par :

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{d} x_i^2}$$

La distance induite par cette norme est la **distance euclidienne**, définie pour deux points $\mathbf{x}$ et $\mathbf{y}$ par :

$$\mathcal{D}_E(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}$$

On vérifie facilement qu'elle satisfait aux axiomes d'une métrique.

1. En effet, $0 = \mathcal{D}(x, x) \leq \mathcal{D}(x, y) + \mathcal{D}(y, x) = 2\mathcal{D}(x, y)$, donc $\mathcal{D}(x, y) \geq 0$.

3 Mise en pratique computationnelle

Dans de nombreuses applications (classification, clustering), on doit résoudre :

$$\arg\min_{i\in\{1,\dots,k\}} \mathcal{D}_E(x, c_i)$$

où $x \in^d$ est un point à classifier et $\{c_1, \dots, c_k\}$ sont des points.

On peut utiliser la distance euclidienne au carré, $\mathcal{D}_E^2$ car cela permet d'éviter l'opération `sqrt`. Le résultat est le même dans les deux cas : pour tout $a, b \in^+$ avec $a < b$, on a $a^2 < b^2$, on a :

$$\arg\min_{i\in\{1,\dots,k\}} \mathcal{D}_E(x, c_i) = \arg\min_{i\in\{1,\dots,k\}} \mathcal{D}_E^2(x, c_i)$$

En pratique, pour calculer $\mathcal{D}_E^2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$, on utilise l'identité remarquable issue du produit scalaire :

$$\begin{aligned}
\|\mathbf{x} - \mathbf{y}\|^2 &= (\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y}) \\
&= \mathbf{x} \cdot \mathbf{x} - 2(\mathbf{x} \cdot \mathbf{y}) + \mathbf{y} \cdot \mathbf{y} \\
&= \|\mathbf{x}\|^2 - 2(\mathbf{x} \cdot \mathbf{y}) + \|\mathbf{y}\|^2
\end{aligned}$$

En pratique, lorsqu'on fait un clustering ou un algorithme de classification en python, on peut alors calculer une seule fois $\|\mathbf{x}\|^2$ et $\|\mathbf{y}\|^2$ en stockant leurs valeurs, et faire une boucle `for` qui itère sur $\mathbf{x} \cdot \mathbf{p_i}$