CARTES TOPOLOGIQUES

1ère séance

Introduction

- Modèles à apprentissage non supervisé

  → Regrouper les patients qui semblent avoir des analyses médicales semblables

  → Identifier les différents types de consommateurs (comportement d'achats)

- **But** : analyser des données d'observation par leur structures.

- Kohonen : Représentation de données multidimensionnelles de grande taille.

  ◇ Projection de partitions

  → selon une structure de voisinage en dimension 1, 2 ou 3

  ◇ **Ordre topologique**

  → les distances entre observations sont directement visibles sur la carte

Quantification Vectorielle

$\mathcal{D}$: espace des données d'observation (notées z) de dimension n.

$\mathcal{A}$: ensemble d'apprentissage $\quad \mathcal{A} = \{\mathbf{z}_i, i = 1, \ldots, N\}$

On a : $\mathcal{A} \subset \mathcal{D} \subset \mathcal{R}^n$

Réduire l'information de $\mathcal{D}$:

- En la <u>résumant</u> par un ensemble de **p** **<u>référents</u>**

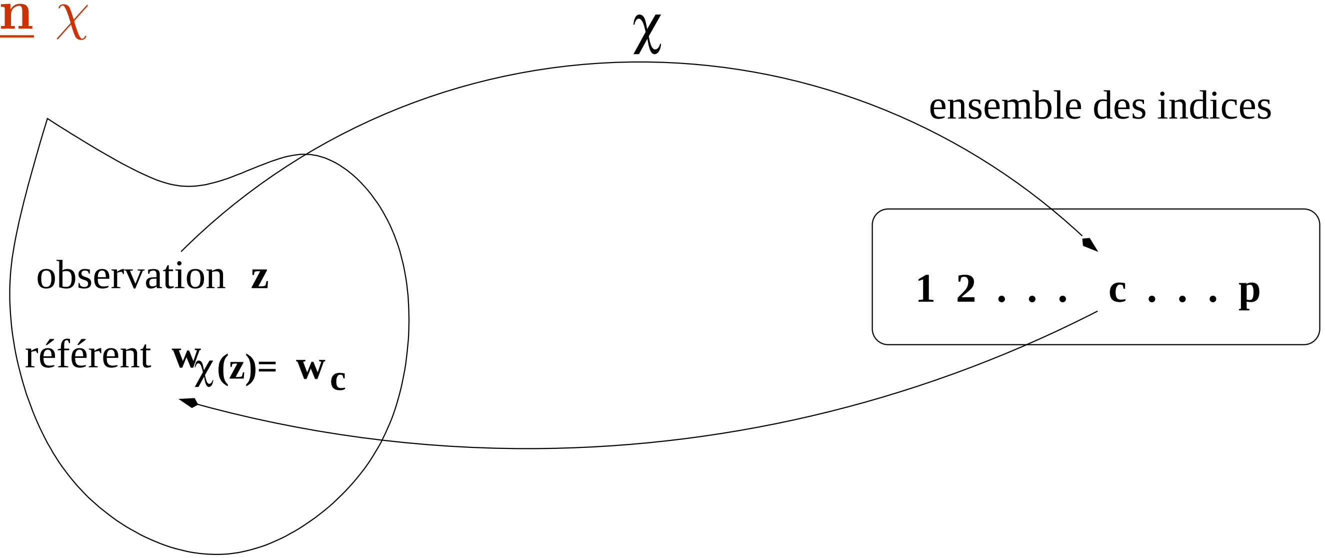
$$\mathcal{W} = \{\mathbf{w}_c; c = 1, ..., p\}$$

- En <u>réalisant</u> une partition de $\mathcal{D}$ en **p** sous-ensembles par l'intermédiaire d'une **<u>fonction d'affectation</u>** χ

$$\chi : \mathcal{D} \rightarrow \{1, 2, \ldots, p\}$$

$$P = \{P_1, P_2, \ldots, P_p\}$$

$$P_c = \{\mathbf{z} \in \mathcal{D} / \chi(\mathbf{z}) = \mathbf{c}\}$$



espace des observations et des référents

Différentes Méthode de Quantification Vectorielle

$$\Downarrow$$

Différentes détermination de W et χ

- Méthode des K-moyennes
- Cartes topologiques auto-organisatrice de Kohonen (SOM)
- Cartes topologiques probabilistes (PRSOM)
- ...

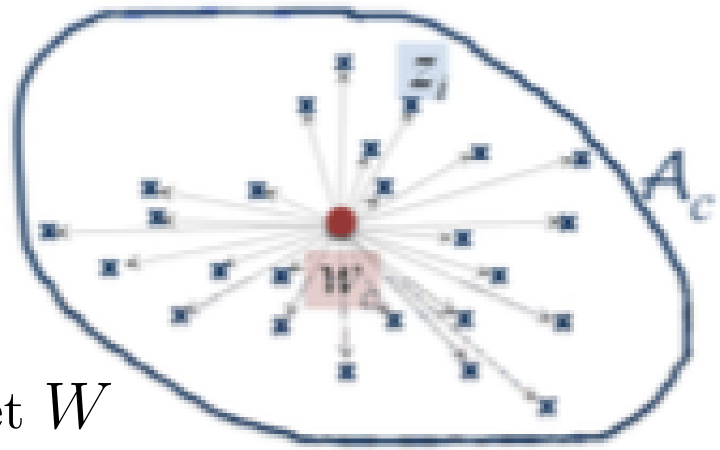
☞ **Chaque méthode correspond à la minimisation d'une fonction de coût spécifique**

☞ **Algorithme en 2 étapes :**

- Affectation $\longrightarrow$ fonction χ

- Minimisation $\longrightarrow$ référents W

Méthode des k-moyennes



- Inertie intra-groupes : $\mathcal{I}_c = \sum_{\mathbf{z_i} \in \mathcal{A}} \left\| \mathbf{z_i} - \mathbf{w_c} \right\|^2$

- Minimiser la somme des inerties locales par rapport à χ et W

$$\mathcal{I}(\mathcal{W}, \chi) = \sum_c \mathcal{I}_c$$

Soit, pour faire ressortir χ :

$$\mathcal{I}(\mathcal{W}, \chi) = \sum_c \mathcal{I}_c = \sum_c \sum_{\substack{\mathbf{z_i} \in \mathcal{A} \\ \chi(\mathbf{z_i}) = \mathbf{c}}} \left\| \mathbf{z_i} - \mathbf{w_c} \right\|^2$$

- L'inertie $\mathcal{I}_c$ représente l'erreur de quantification obtenue si l'on remplace chaque observation de P_c par son référent $\mathbf{w_c}$

- Minimisation itérative de l'inertie qui fixe alternativement la partition (χ) puis recalcule $\mathcal{W}$

Minimisation de l'inertie $\mathcal{I}(\mathcal{W}, \chi)$

- **Phase d'affectation**

  Pour un ensemble $\mathbf{W}$ de référents fixe, la minimisation de $\mathbf{I}$ par rapport à χ s'obtient en affectant chaque observation $\mathbf{z}$ au référent $\mathbf{w_c}$ selon la nouvelle fonction d'affectation χ

$$\boxed{\chi(\mathbf{z}) = \arg\min_{\mathbf{r}} \|\mathbf{z} - \mathbf{w_r}\|^{\mathbf{2}}} \qquad \mathbf{(1)}$$

- **Phase de minimisation**

  La fonction χ est fixée. La fonction $\mathbf{I(W,}\chi\mathbf{)}$ est quadratique et convexe par rapport à $\mathbf{W}$. Le minimum global est atteint pour :

$$\frac{\partial I}{\partial \mathbf{W}} = [\frac{\partial I}{\partial \mathbf{w_1}}, \frac{\partial I}{\partial \mathbf{w_2}}, \cdots, \frac{\partial I}{\partial \mathbf{w_p}}]^T = 0 \qquad \Rightarrow \qquad \forall c, \sum_{\mathbf{z_i} \in \mathcal{P}_\mathbf{c}} (\mathbf{z_i} - \mathbf{w_c}) = \mathbf{0}$$

$$\Downarrow$$

$$\text{Les } p \text{ nouveaux référents} \longrightarrow \boxed{\mathbf{w_c} = \frac{\displaystyle\sum_{\mathbf{z_i} \in \mathcal{A}_\mathbf{c}} \mathbf{z_i}}{\mathbf{n_c}}} \qquad \mathbf{(2)}$$

Algorithme des k-moyennes

1. **Initialisation**

   - t=0 : indice d'itération
   - fixer le nombre maximum d'itérations : N_{iter}
   - choisir les p référents initiaux (en général d'une manière aléatoire) : $\mathcal{W}^t$

2. **Etape de minimisation itérative t** (à partir de 1),

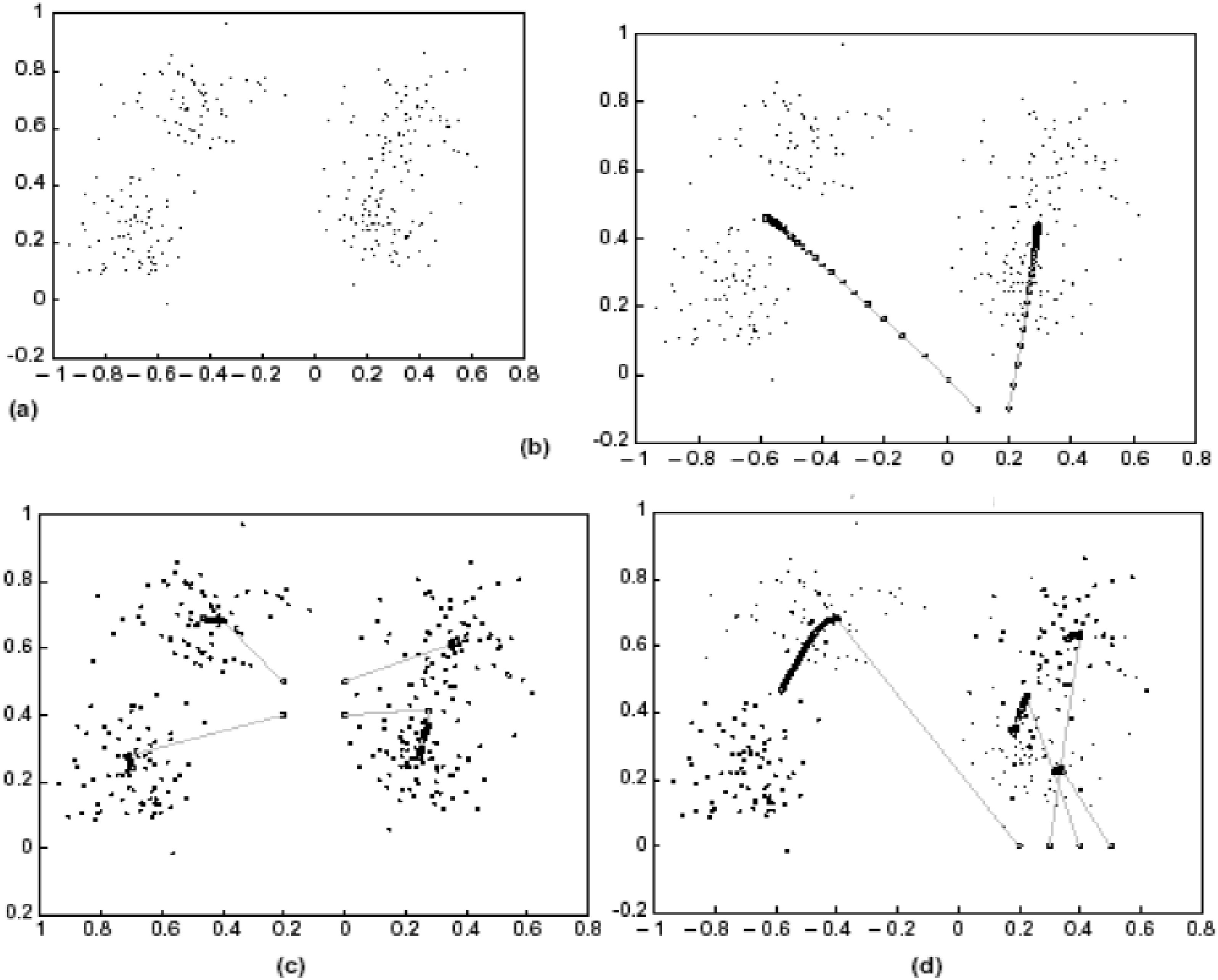
   (à l'itération t, l'ensemble des référents $\mathcal{W}^{t-1}$ de l'étape précédente est connu)

   - **Phase d'affectation** : mise à jour de la fonction d'affectation χ^t associée à $\mathcal{W}^{t-1}$: on affecte chaque observation $\mathbf{z}$ au référent le plus proche (1).

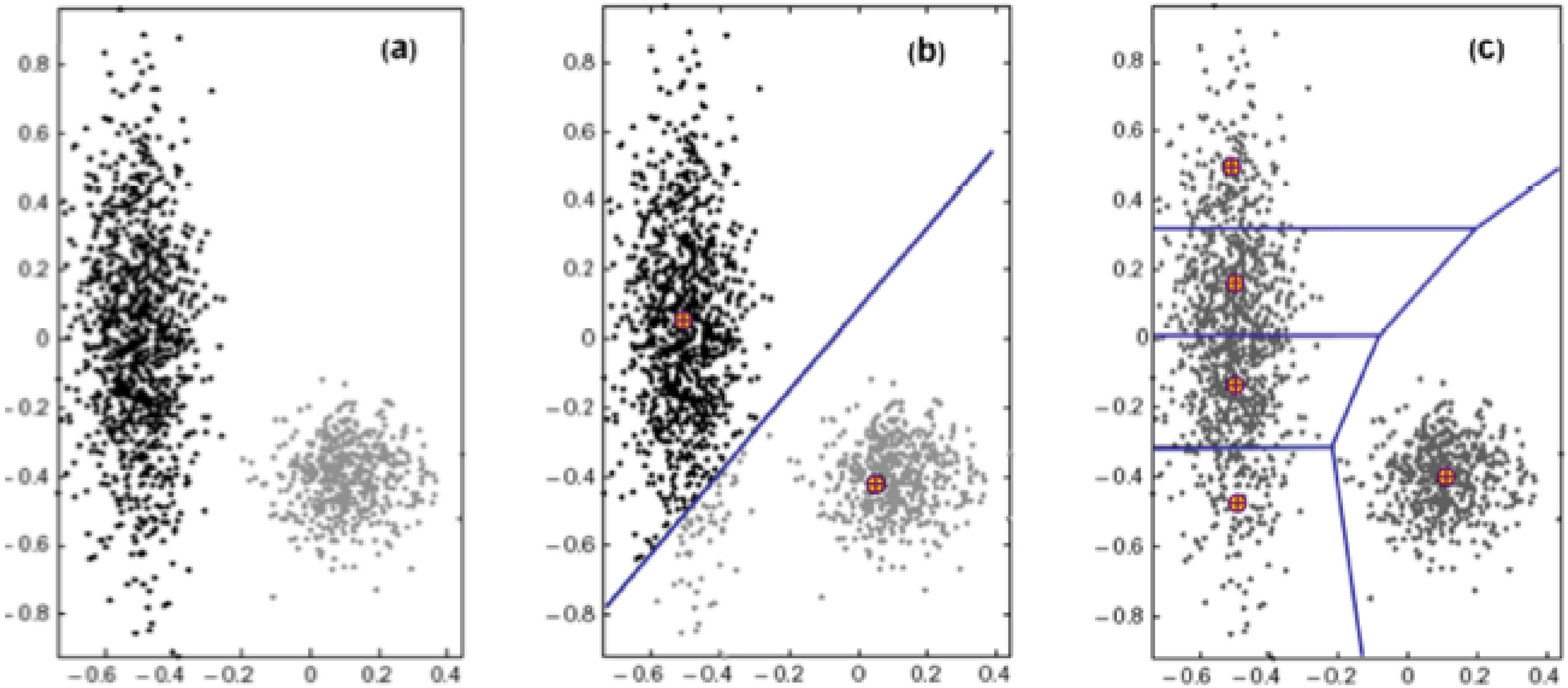
   - **Phase de minimisation** : calcul des nouveaux référents W^t en appliquant l'équation (2).

3. **Répéter** l'étape itérative jusqu'à atteindre $t > N_{iter}$ itérations ou une stabilisation de $\mathbf{I}$.

Sensibilité aux conditions initiales



Comportement de l'algorithme des k-moyennes en fonctions des densités sous-jacentes

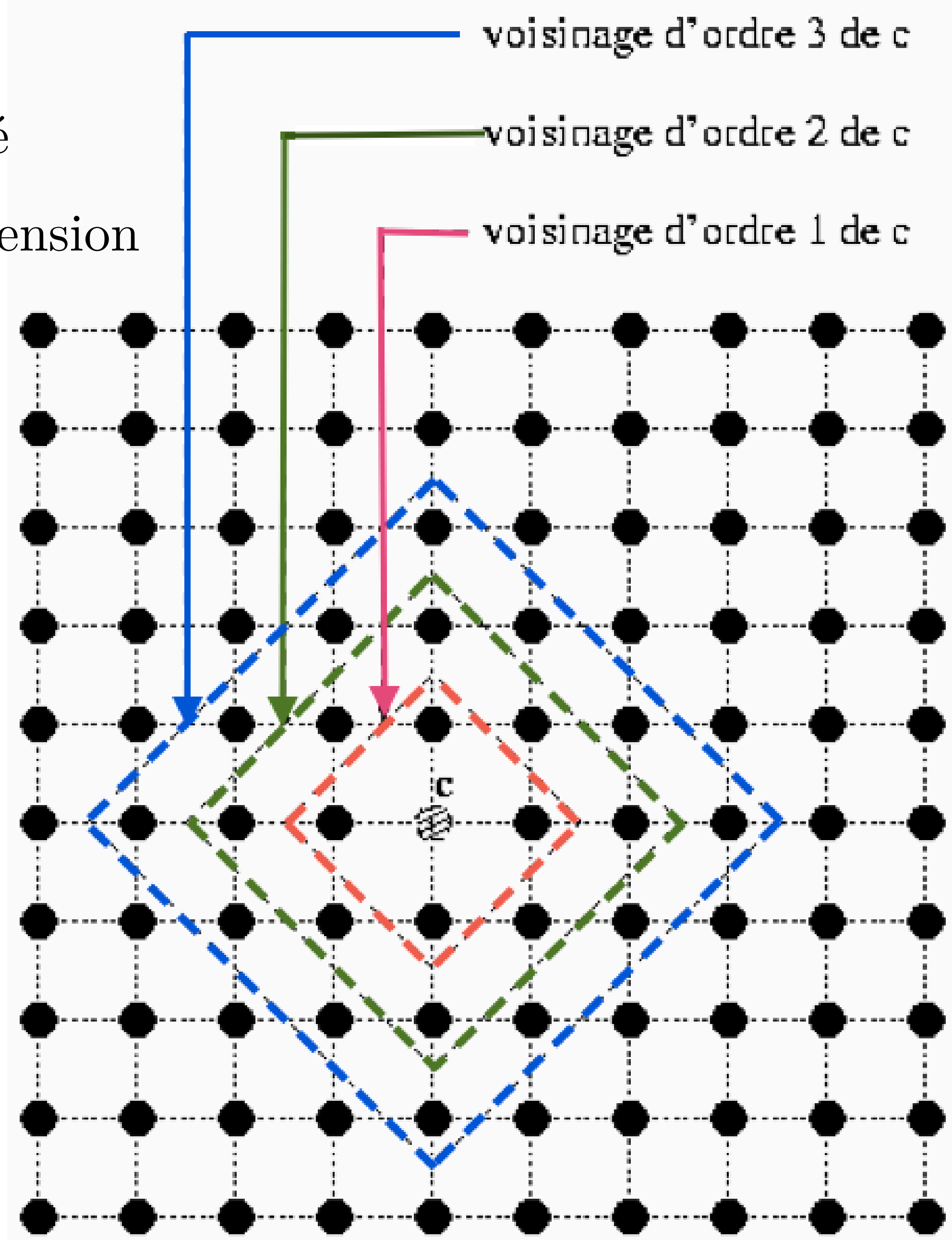


(a) Données simulées selon deux distributions gaussiennes de matrice de variance-covariance différentes.

(b) Référents et partition obtenue à la convergence avec 2 référents

(c) Avec 5 référents

Carte topologique



- L'ensemble des indices est maintenant ordonné
- Il s'agit d'un espace discret (**C**) de faible dimension à des fins de visualisation (**1-D, 2-D**).
- $\mathcal{C}$ ensemble de **neurones** connectés par une structure de graphe non-orienté muni d'une distance discrète δ sur $\mathcal{C}$ et d'une **structure de voisinage** :

$$V_c(d) = \{r \in \mathcal{C}, \delta(c,r) \leq d\}$$

$$\downarrow$$

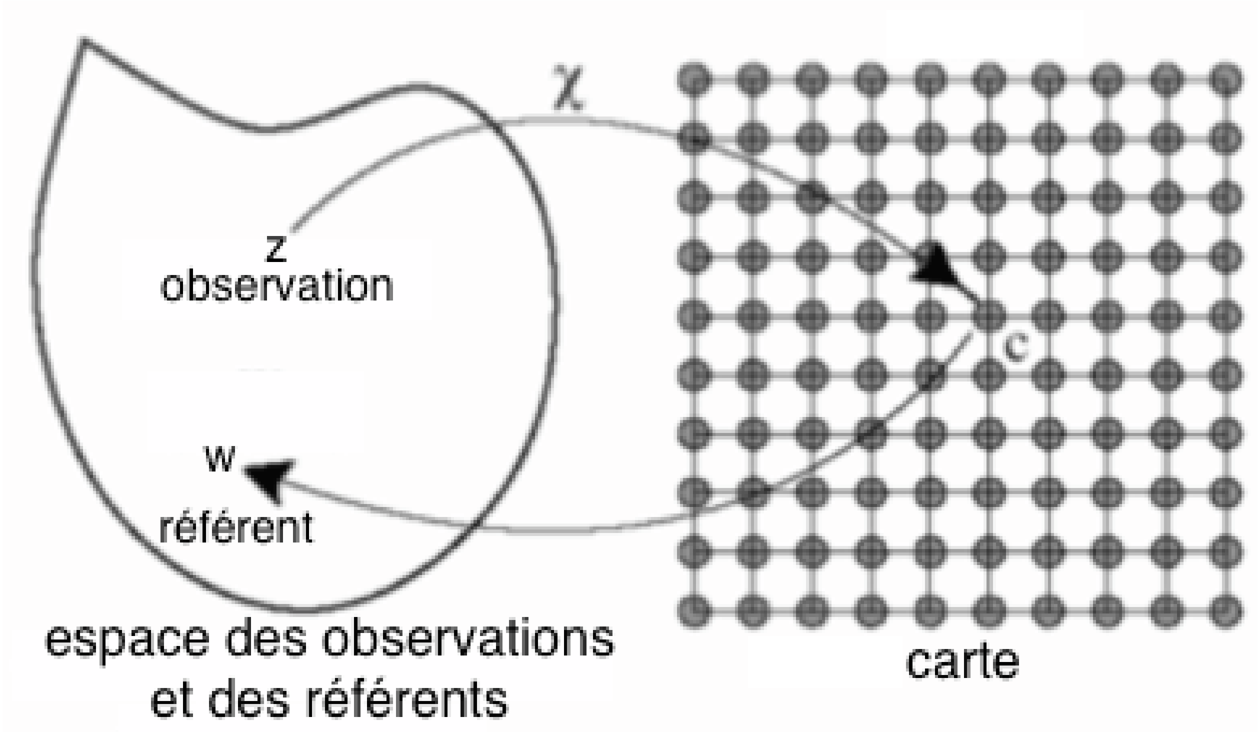
Voisinage d'ordre d

$\delta(c,r)$ = **longueur du plus court chemin entre c et r sur le graphe**

Quantification par la carte

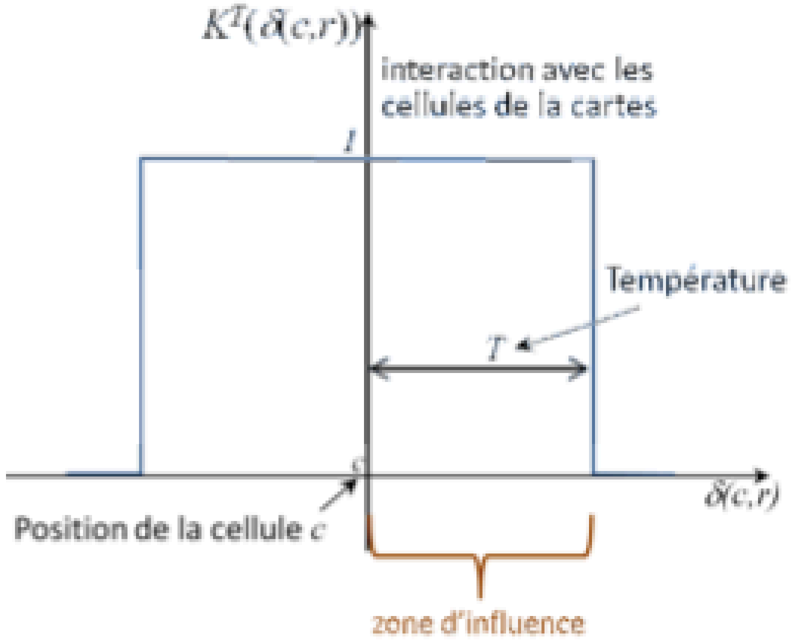
- Chaque neurone c de $\mathcal{C}$ est associé à un vecteur référent $\mathbf{w_c}$ de l'espace des données $\mathcal{D}$
- L'apprentissage approxime la densité sous-jacente des données tout en cherchant à respecter une contrainte de conservation de la topologie de la carte $\mathcal{C}$
- Deux neurones c et r "voisins" par rapport à la topologie discrète de la carte $\mathcal{C}$ sont associés à deux vecteurs référents $\mathbf{w_c}$ et $\mathbf{w_r}$ "proches" dans l'espace des données $\mathcal{D}$.

☞ **NOTION DE CONSERVATION DE LA TOPOLOGIE**

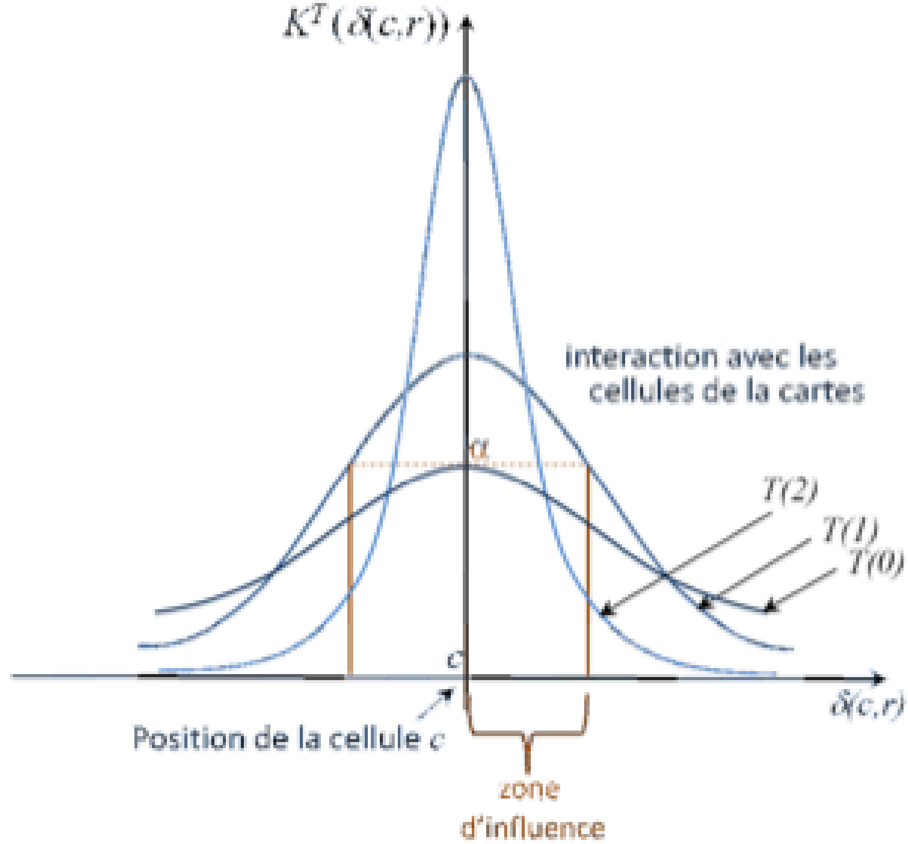


Les indices des référents sont ordonée par la topologie

Fonction de Voisinage



- $K^T(\delta) = \begin{cases} 1 & \text{si } \delta \le T \\ 0 & \text{sinon} \end{cases}$



- $K^T(\delta) = exp(-|\delta| / T)$
- $K^T(\delta) = exp(\delta^2 / -T^2)$

Taille du voisinage - zone d'influence - voisinage significatif

$$V_c^T = \{ r \in C \mid \delta(c,r) \le T \}$$

$$V_c^T = \{ r \in C \mid K^T(\delta(c,r)) > \alpha \}$$

Apprentissage

Apprentissage de la carte = minimisation d'une fonction de coût

$$\boxed{J^T_{som}(\chi, \mathcal{W}) = \sum_{\mathbf{z_i} \in \mathcal{A}} \sum_{c \in C} K^T(\delta(c, \chi(\mathbf{z_i}))\|\mathbf{z_i} - \mathbf{w_c}\|^{\mathbf{2}}}$$

La valeur de T détermine la taille du voisinage

$$V_c^T = \{r \in \mathcal{C} / K^T(\delta(c, r)) > \alpha\}$$

<u>Remarque</u> : T=0 $\Rightarrow$ fonction de coût des K-moyennes

Le seuil α gère l'ordre des valeurs significatives prises en compte par le calcul.

Chaque observation $\mathbf{z}$ calcul une <u>**distance généralisée**</u> à l'ensemble des neurones de la carte par l'intermédiaire de la fonction K^T

$$\boxed{d^T(\mathbf{z_i}, \mathbf{w}_{\chi(\mathbf{z_i})}) = \sum_{\mathbf{c} \in \mathbf{C}} \mathbf{K^T}(\delta(\mathbf{c}, \chi(\mathbf{z_i}))\|\mathbf{z_i} - \mathbf{w_c}\|^{\mathbf{2}}}$$