

Apprentissage des cartes topologiques

☞ Minimisation d'une fonction de coût

$$J_{som}^T(\chi, \mathcal{W}) = \sum_{\mathbf{z}_i \in \mathcal{A}} \sum_{c \in \mathcal{C}} K^T(\delta(c, \chi(\mathbf{z}_i)) \|\mathbf{z}_i - \mathbf{w}_c\|^2$$

- \mathcal{A} : ensemble d'apprentissage $\mathcal{A} = \{\mathbf{z}_i, i = 1, \dots, N\}$
- χ : une fonction d'affectation
- \mathcal{W} : l'ensemble des p vecteurs référents qui forment la carte
- $\chi(\mathbf{z}_i)$: le neurone particulier de la carte \mathcal{C} qui est affecté à l'observation \mathbf{z}_i
- $\delta(c, \chi(\mathbf{z}_i))$: la distance sur la carte \mathcal{C} entre un neurone c quelconque et le neurone $\chi(\mathbf{z}_i)$ affecté à l'observation \mathbf{z}_i (= le plus court chemin entre c et $\chi(\mathbf{z}_i)$)
- K : une fonction de voisinage, paramétrée par un terme de température T
 - La valeur de T détermine la taille du voisinage
 - Chaque observation \mathbf{z} calcul une distance généralisée à l'ensemble des neurones de la carte par l'intermédiaire de la fonction K^T

$$d^T(\mathbf{z}_i, \mathbf{w}_{\chi(\mathbf{z}_i)}) = \sum_{c \in \mathcal{C}} K^T(\delta(c, \chi(\mathbf{z}_i)) \|\mathbf{z}_i - \mathbf{w}_c\|^2$$

- $T=0 \Rightarrow$ fonction de coût des K-moyennes

Optimisation non adaptative des cartes topologiques

Minimisation itérative de $J_{som}(\chi, W)$ pour un paramètre T fixé

- **Phase d'affectation**

L'ensemble W des référents est fixé, la minimisation s'obtient en affectant chaque observation z au référent w_c selon la nouvelle fonction d'affectation χ^T

$$\chi^T(z) = \arg \min_{r \in C} \sum_{c \in C} \mathbf{K}^T(\delta(c, r)) \|z - w_c\|^2 = \arg \min_{r \in C} d^T(z, w_r) \quad (3)$$

- **Phase de minimisation**

La partition χ^T est fixée. La fonction $J_{som}(\chi, W)$ est minimisée par rapport à l'ensemble des référents W .

La fonction étant convexe par rapport aux paramètres, le minimum global est atteint pour

$$w_c^T = \frac{\sum_{r \in C} \mathbf{K}^T(\delta(c, r)) Z_r}{\sum_{r \in C} \mathbf{K}^T(\delta(c, r)) n_r} \quad (4)$$

Z_r représente la somme de toutes les observations affectées au neurone r

$$= \sum_{z_i \in \mathcal{A}; \chi(z_i)=r} z_i$$

n_r = nombre des observations affectées au neurone r

Algorithme «nuées dynamiques» des cartes topologiques

T fixé

1. Phase d'initialisation

$t=0$: indice d'itération

Choisir la structure et la taille p de la carte

Choisir les p référents initiaux (en général d'une manière aléatoire)

N_{iter} : nombre maximum d'itérations.

2. Etape itérative 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} .

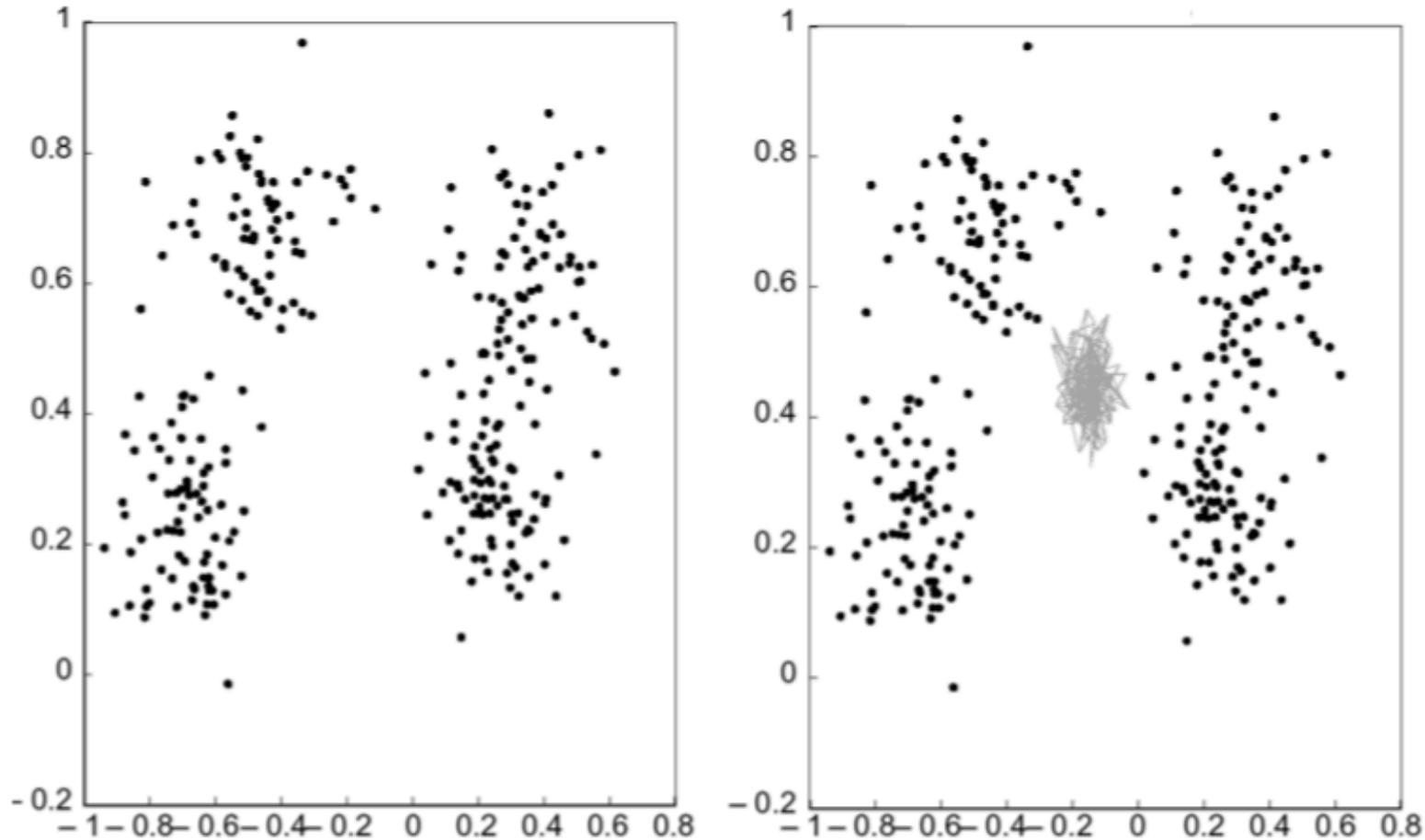
On affecte chaque observation \mathbf{z} au référent défini à partir de l'expression (3).

- Phase de minimisation

Calcul des nouveaux référents \mathcal{W}^t , associés à χ^t , en appliquant l'équation (4).

3. Répéter l'étape itérative jusqu'à atteindre N_{iter} itérations ou une stabilisation de la fonction.

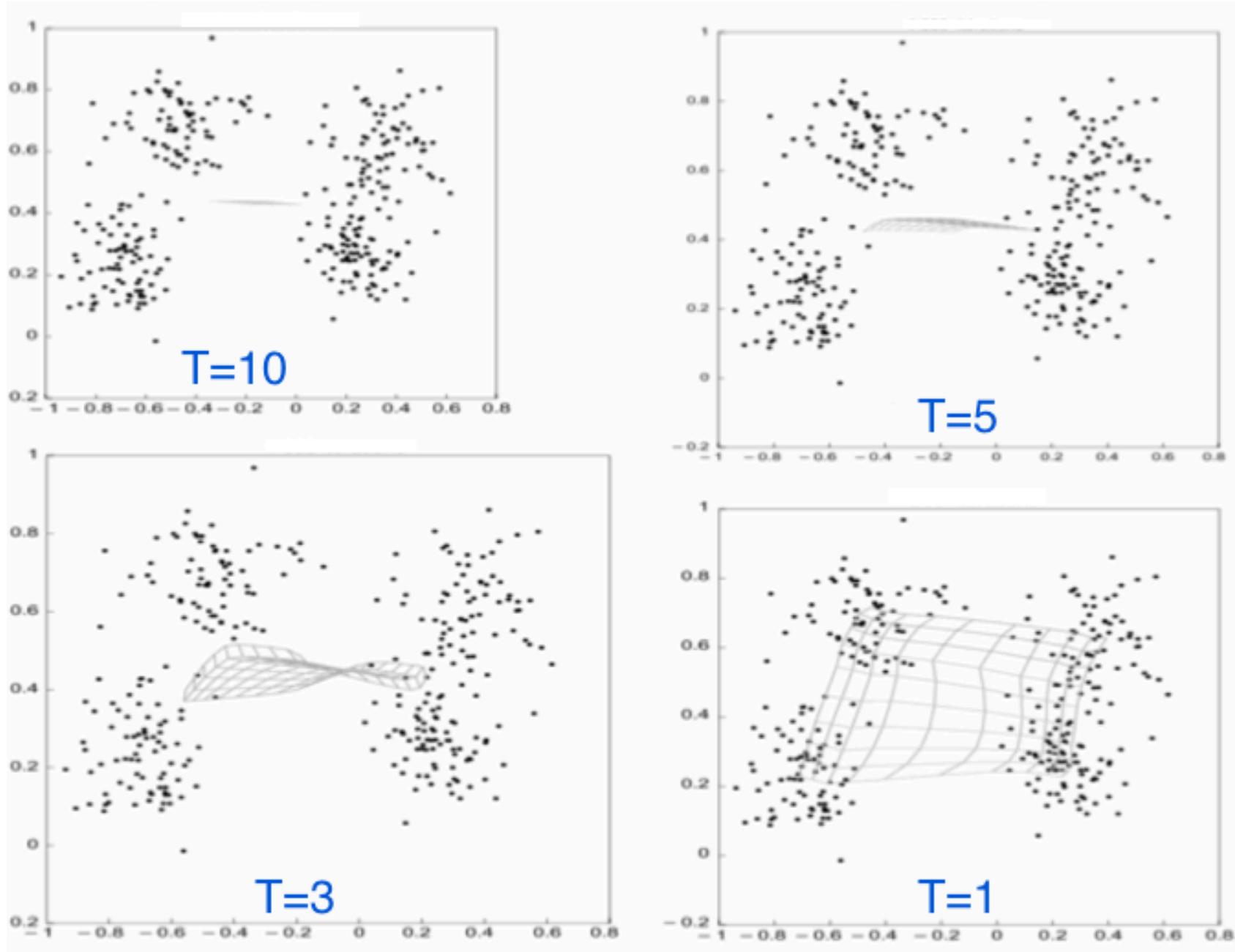
Exemple : données simulées



Données simulées selon quatre gaussiennes qui se recouvrent partiellement

- **A gauche** : les données
- **A droite** : représentation simultanée des données et des référents au moment de la phase d'initialisation de l'algorithme.

Carte topologique obtenue à la convergence pour T fixé



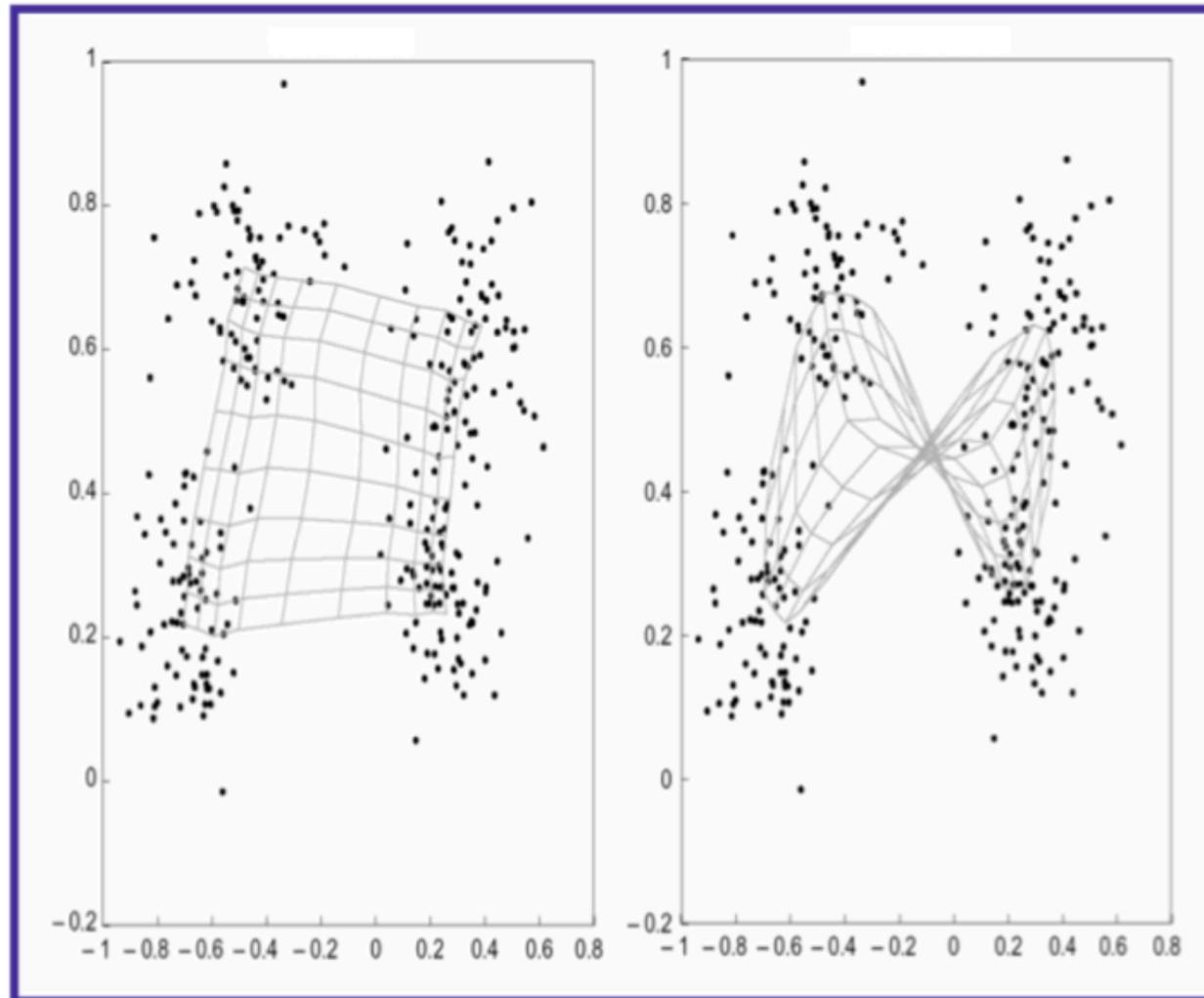
Algorithme des cartes topologiques

- La minimisation à T fixé est répétée un certain nombre de fois en faisant décroître la valeur de T .
- L'ordre apparaît pour les grandes valeurs de T , la carte se déploie alors pour recouvrir les données et faire diminuer la variance intra.

Les paramètres déterminants de la minimisation sont :

- l'intervalle de variation de T , la valeur initiale T^{max} et la valeur finale T^{min}
- Le nombre de fois où l'étape itérative est effectuée.
- La manière dont le paramètre T décroît dans l'intervalle $[T^{max}, T^{min}]$.

Une mauvaise gestion de la décroissance de T peut introduire des anomalies dans l'ordre

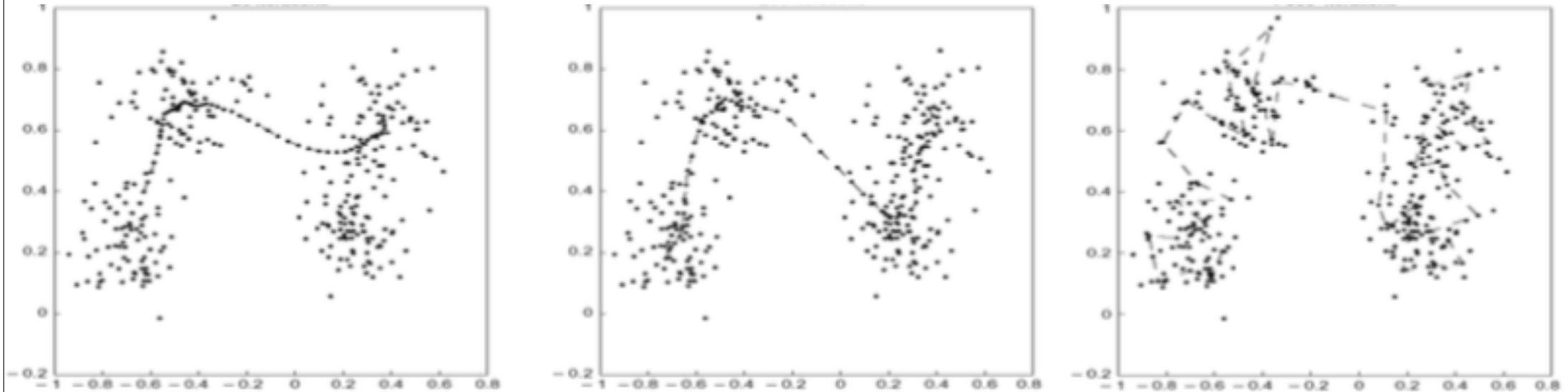


Qualité de la carte : erreur de quantification et erreur topologique

Evolution de l'apprentissage

Données simulées : 4 gaussiennes

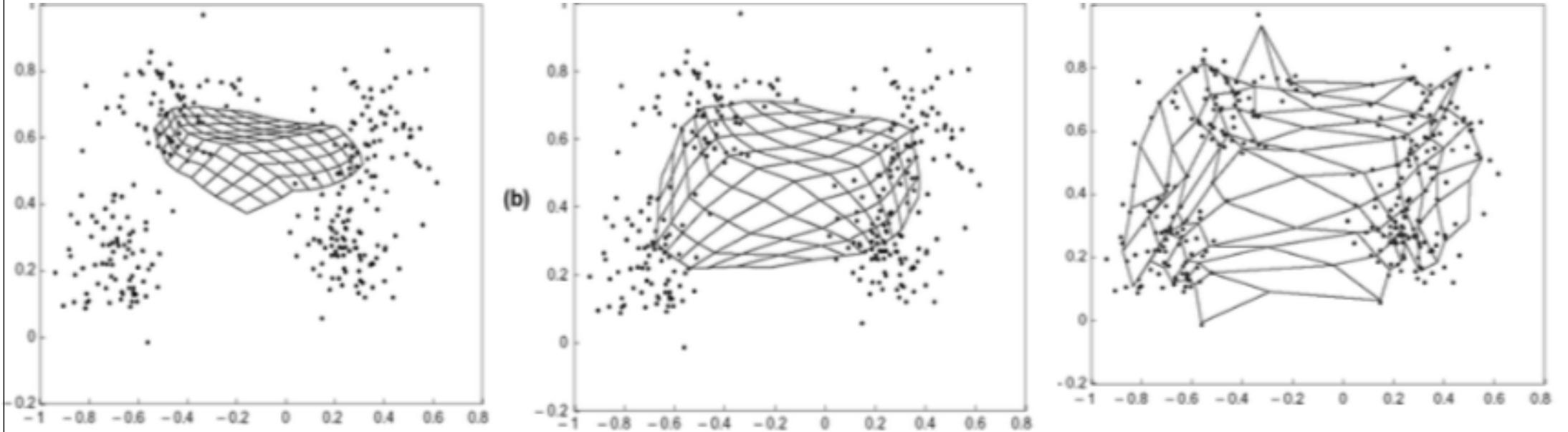
Carte 1-D de 50 neurones



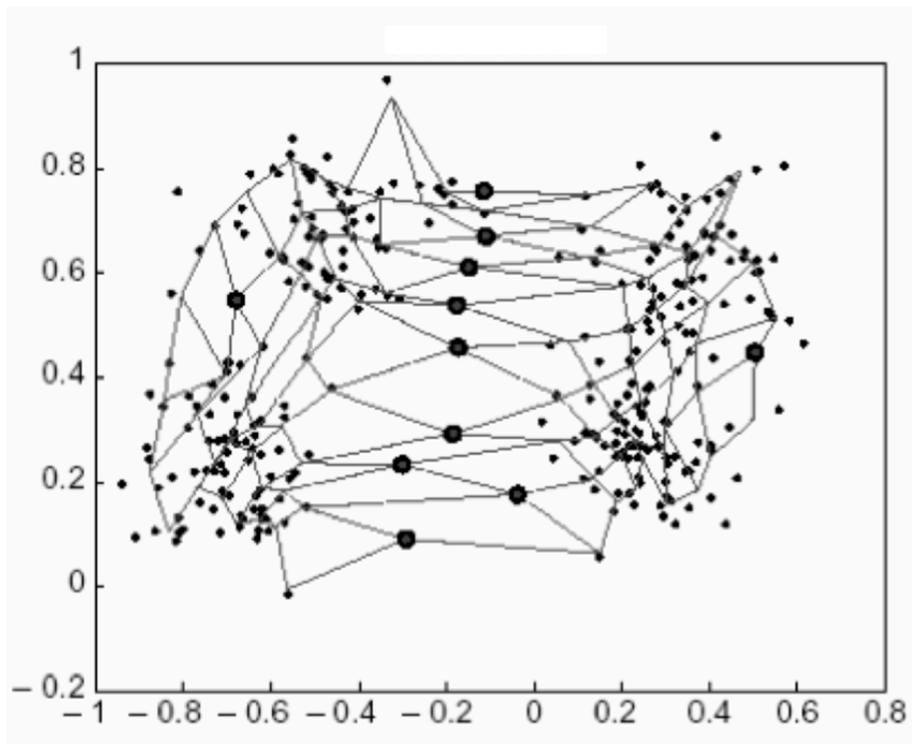
Evolution de l'apprentissage

Données simulées : 4 gaussiennes

Carte 2-D de 10*10 neurones



Visualisation des différents sous-ensembles



Les frontières dessinées par les neurones n'ayant rien captés permettent de distinguer grossièrement les différents groupes.

Dans l'exemple les observations ont été découpées en deux sous-ensembles regroupant à chaque fois deux gaussiennes.

Ordre Topologique

$$J_{som}^T = \sum_r \sum_{\mathbf{z}_i \in P_r \cap \mathcal{A}} \sum_c K^T(\delta(c, r)) \|\mathbf{z}_i - \mathbf{w}_c\|^2 = \sum_c \sum_r \sum_{\mathbf{z}_i \in P_r} K^T(\delta(c, r)) \|\mathbf{z}_i - \mathbf{w}_c\|^2$$

$$J_{som}^T = \left[\sum_c \sum_{r \neq c} \sum_{\mathbf{z}_i \in P_r} K^T(\delta(c, r)) \|\mathbf{z}_i - \mathbf{w}_c\|^2 \right] + \left[K^T(\delta(c, c)) \sum_c \sum_{\mathbf{z}_i \in P_c} \|\mathbf{z}_i - \mathbf{w}_c\|^2 \right]$$
$$= \frac{1}{2} \sum_c \sum_{r \neq c} K^T(\delta(c, r)) \left[\sum_{\mathbf{z}_i \in P_r} \|\mathbf{z}_i - \mathbf{w}_c\|^2 + \sum_{\mathbf{z}_i \in P_c} \|\mathbf{z}_i - \mathbf{w}_r\|^2 \right]$$
$$+ K^T(\delta(c, c)) \sum_c \sum_{\mathbf{z}_i \in P_c} \|\mathbf{z}_i - \mathbf{w}_c\|^2.$$

Induit l'ordre sur la carte, deux neurones proches doivent avoir des référents proches

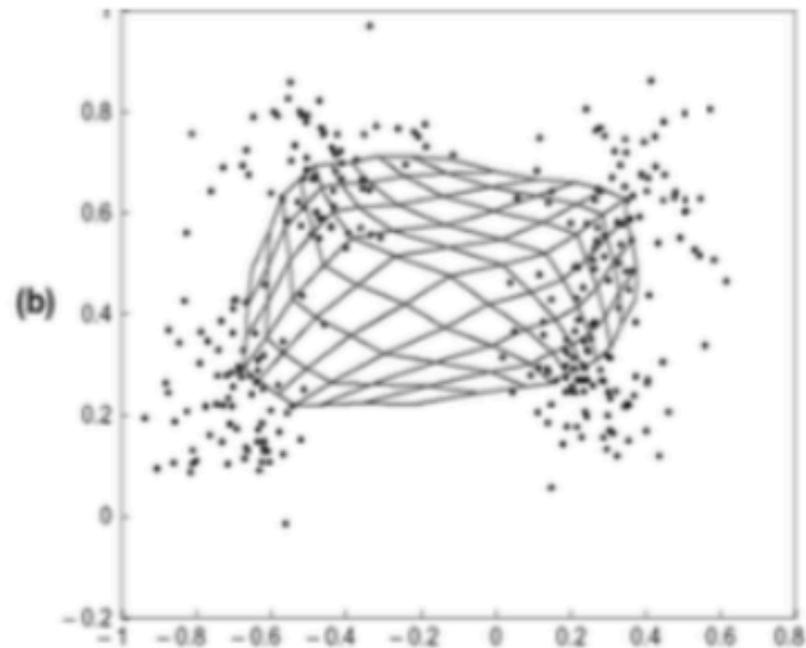
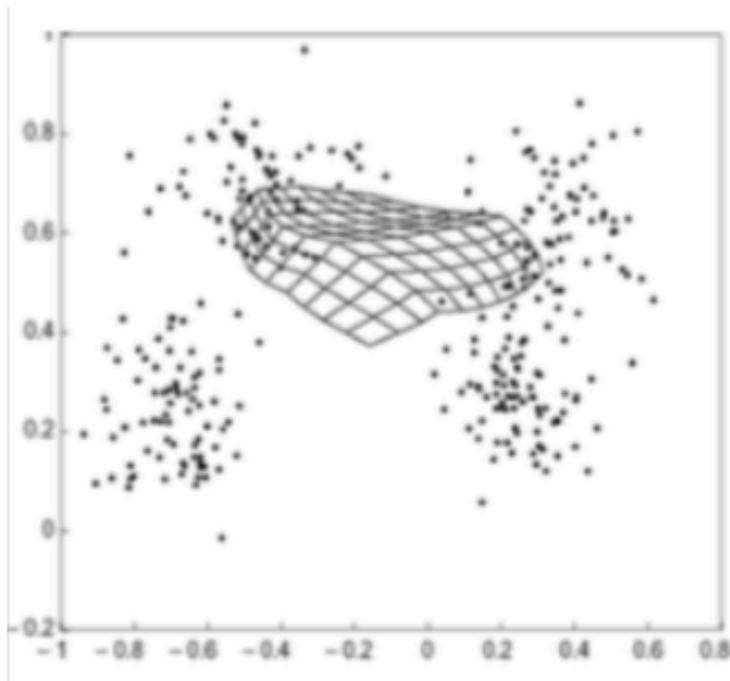
Fonction de coût des k-moyennes pondéré par $K^T(0)$

T petit: carte topologique et k-moyennes coincident

Déroulement de l'algorithme : 2 phases

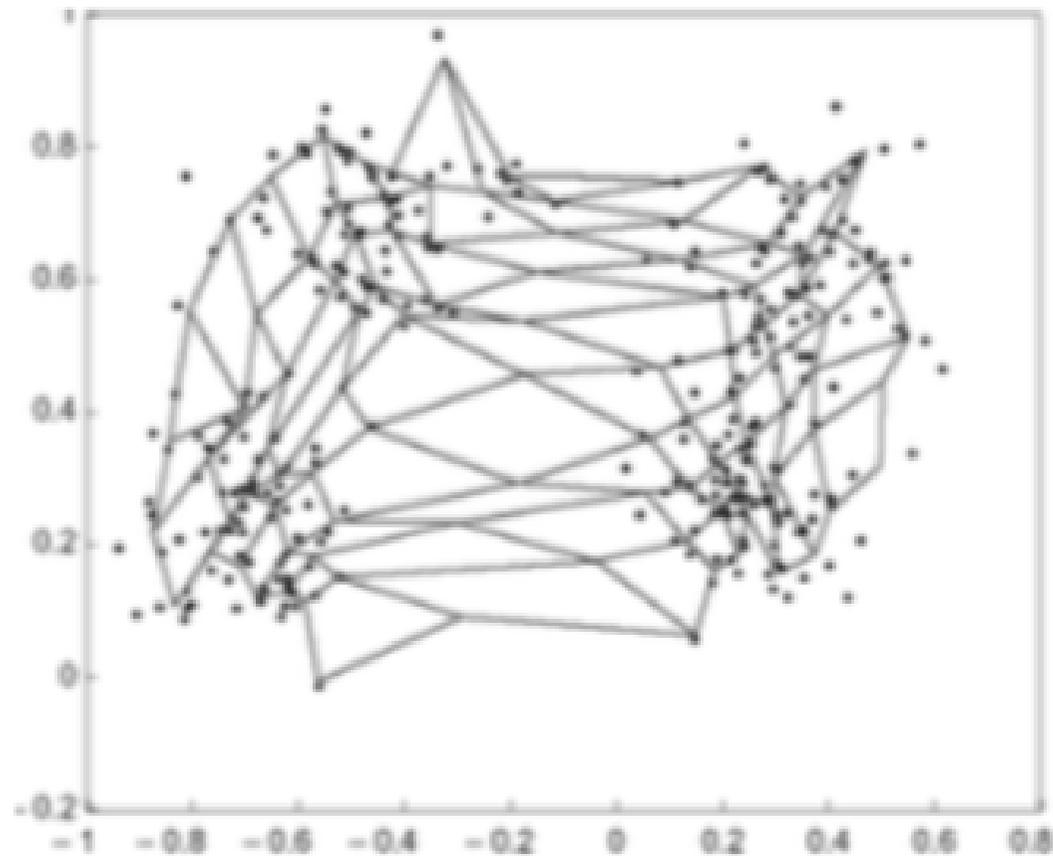
La convergence vers la solution peut se décomposer en 2 phases

- **Phase 1 (Auto organisation)** : correspond aux grandes valeurs de T
 - Minimisation du premier terme de J_{som}
 - L'ordre se met en place

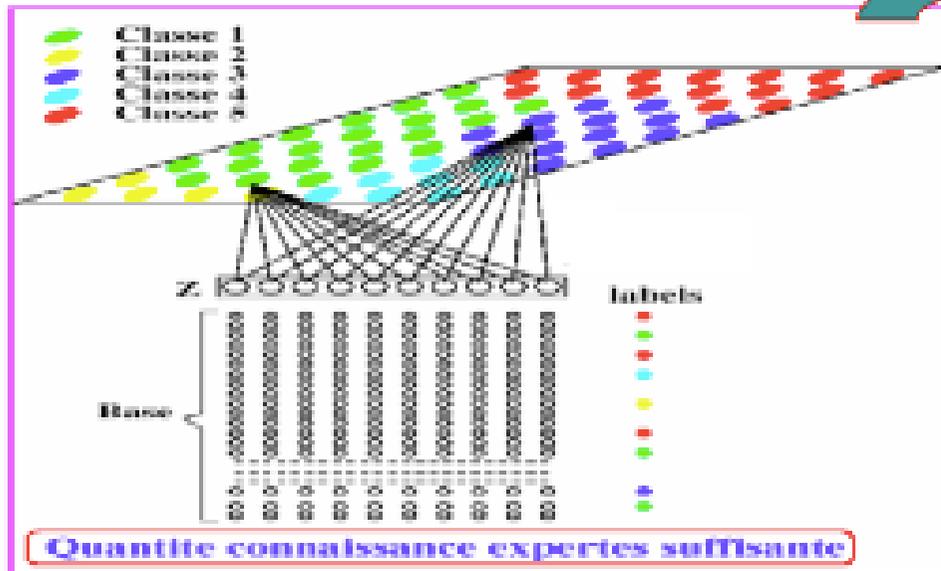


Déroulement de l'algorithme : 2 phases

- **Phase 2 (Convergence)** : correspond aux petites valeurs de T
 - Minimisation du second terme de J_{som}
 - La variance intra est minimisée,
les référents deviennent les centres de gravité de la partition



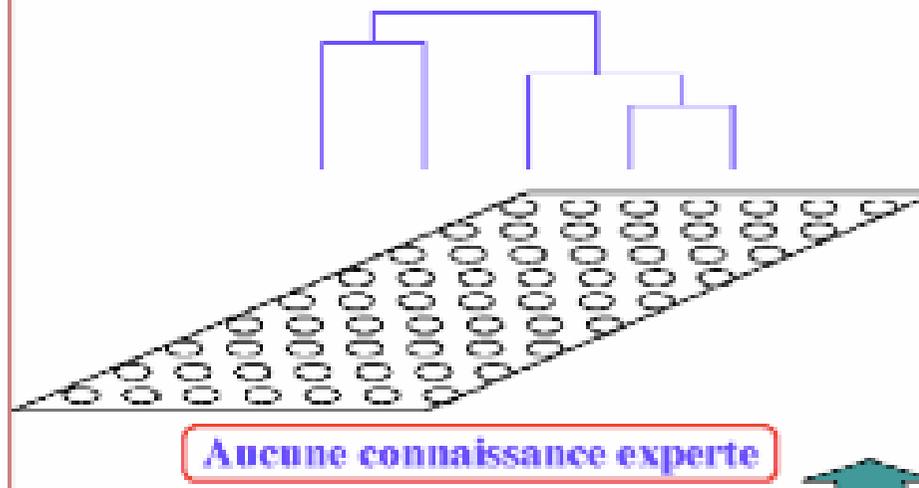
Classification et carte topologique



**Base étiquetée statistiquement
représentative:**

1. Affectation à l'aide de la fonction χ
2. Utilisation du vote majoritaire sur les différents sous-ensemble de la partition
3. Etiquetage du neurone

CAH sur les neurones + interprétation



Pas de connaissance experte

Sous l'hypothèse d'un bon ordre topologique

*Il est très probable que deux neurones voisins
représentent des données de même classe.*

*Une Classification Ascendante hiérarchique
Permet de les retrouver*