

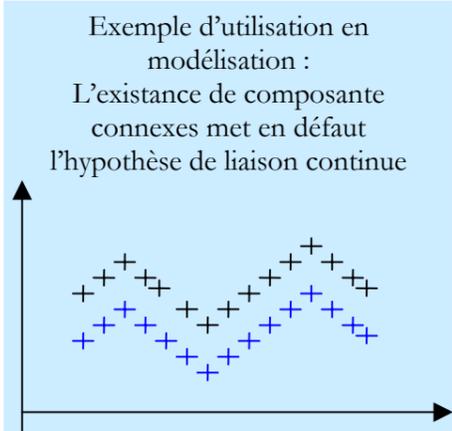
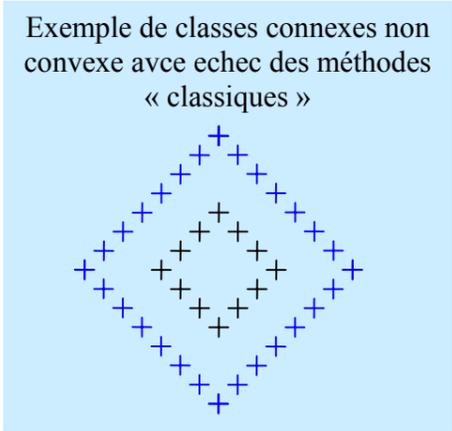
CLASSIFICATION CONNEXE

CATHERINE AARON
SAMOS-MATISSE

L'hypothèse de connexité est l'hypothèse la plus générale que l'on puisse faire sur les classes issues d'une segmentation. En effet, sans imposer d'hypothèse a priori sur la forme des classes, elle impose l'existence d'un chemin continue liant, deux à deux, tous les points d'une même classe.

Sous la seule hypothèse de connexité des classes les méthodes reposant sur une minimisation de variance intra classe, sur des regroupements autour de barycentre ou sur une séparation des classes par des hyper plans sont inopérantes.

Les méthodes pouvant mener à une classification connexe telles que les méthodes bayésiennes ou neuronales sont efficaces mais présupposent une connaissance a priori du nombre de classes.



I- approche hiérarchique

L'algorithme hiérarchique de constitution de classes connexes correspond ce fait par la distance min entre les groupes : algorithme « classique »
En revanche les indicateurs de variances intra-classes ne sont pas compatibles avec la notion de connexité :

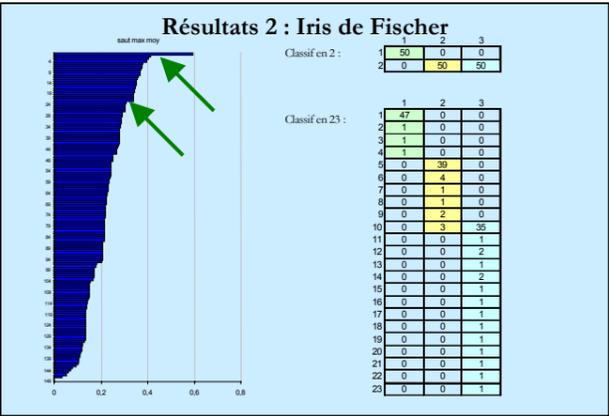
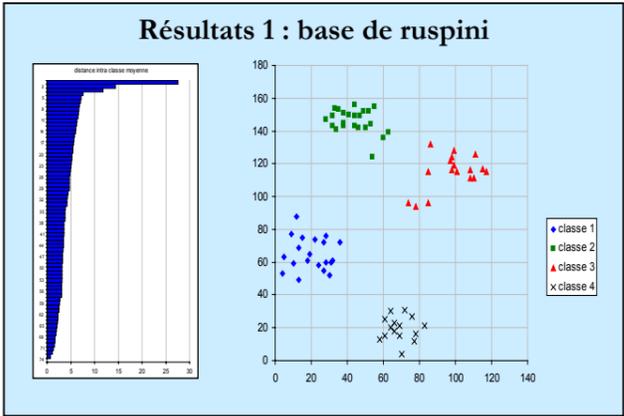
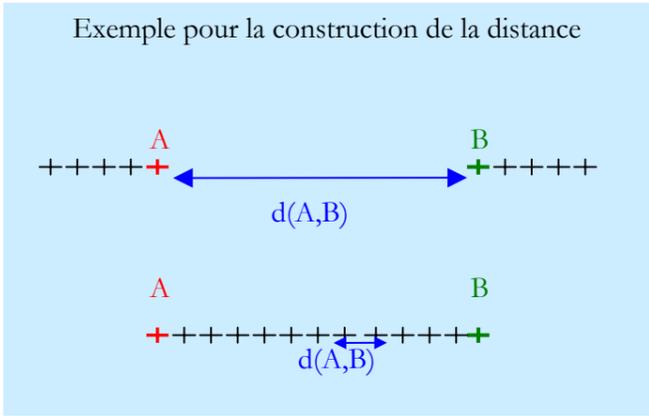
Construction d'une distance intra classe comatible avec la notion de connexité :

Les distances usuelles de type distances euclidienne ne sont pas compatibles : dans les exemples suivants les distances entre les points A et B ne doivent pas être identiques :

On peut alors choisir, comme distance entre deux points ; **le minimum sur les chemins liant les points du plus grand pas sur le chemin.**

Soit $E(x, y)$ l'ensemble des chemins liant x a y c'est a dire l'union sur $p \in \{2, \dots, N\}$ des p -uplets $(x_{i_1}, \dots, x_{i_p})$ avec $x_{i_1} = x, x_{i_p} = y$ alors : $d_c(x, y) = \min_{E(x,y)} (\max_j (d(x_{i_j}, x_{i_{j+1}})))$

On définit la distance intra classe comme moyenne au sein des classes de cette distance



- Temps de calcul en long (en N^3)
- Traitement des classes fortement in-homogène en distance problématique
- Construction d'un teste de significativité pour les variation de distance intra

II- approche Bayésienne

Sous hypothèse de densité connue

L'étude des composantes connexes en coupant par des lignes de niveaux ne résout pas le problème des classes non homogènes

Solution proposée : affecter a chaque point le maximum local obtenu par « montée » de gradient.

- Paramétrages des algorithme pour assurer la convergence de la recherche des maxima locaux
- Construction d'un critère du choix de nombre de classes

Densité inconnue

On tentera de construire, avec plusieurs méthodes d'estimation de densité (noyaux, plus proches voisins, ondelettes...), des algorithmes menant à la construction de dendrogrames. Puis on établira une regle de choix du nombre de classes

Exemple d'estimation de densité sur la base de ruspini en fonction de la taille du noyau