

Chapitre 18

Méthodes de classification

Il existe deux types de méthodes de classification :

1. les méthodes de classification non supervisées¹ ;
2. les méthodes de classification supervisées².

18.1 Les méthodes de classification non supervisées

En français, l'expression « non supervisées » est souvent omis. Les méthodes de classification ont pour objectifs de :

- produire des **groupements en classes d'objets** pour les méthodes de partitionnement, ou **en familles de classes hiérarchisées** pour les méthodes de classification hiérarchiques ;
- regrouper les éléments de la manière la moins arbitraire possible à partir de leurs vecteurs de description.

Les techniques de classification ont pour objectif d'explicitier la structure d'un ensemble de données importantes, permettant ainsi de formuler des hypothèses à vérifier dans une étape ultérieure. Il en existe cinq :

1. la classification hiérarchique³ :
 - soit avec une méthode agglomérative (ascendante) ;
 - soit avec une méthode divisive (descendante).
2. la classification par partition⁴ qui utilise un nombre de classes, fixé par l'analyste ;

1. En anglais : *clustering*

2. En anglais : *classification*

3. En anglais : *hierarchical technique*

4. En anglais : *partitioning technique*

3. la méthode des densités⁵ qui recherche des zones denses dans un nuage de points ;
4. la classification floue⁶ qui engendre des classes disjointes ;
5. la classification spectrale qui est une analyse factorielle du type Q-SORT, produisant une décomposition spectrale de la matrice $\mathbf{R} \cdot {}^t\mathbf{R}$ avec \mathbf{X} une matrice composée de n individus et p caractères.

L'objectif d'une classification est de **décrire** et d'**explorer** les données.

18.2 Les méthodes de classification supervisées

Les méthodes de classification supervisées (ou de discrimination) sont des méthodes de **classement** ou d'affectation d'individus dans des **classes pré-existantes** [Gordon, 1981].

Il existe de nombreuses techniques :

- l'analyse linéaire discriminante ;
- les analyses régularisées ;
- la régression logistique ;
- la segmentation.

L'objectif d'un classement est d'**expliquer** ou de **prévoir**.

18.3 Vocabulaire de base

Distance : mesure de la proximité entre deux objets à classer.

Ressemblance : deux objets sont d'autant plus ressemblants qu'ils sont à faible distance.

Similarité : mesure de la ressemblance entre deux objets ayant des propriétés moins fortes que la distance.

Dissimilarité : mesure de la non ressemblance entre deux objets.

Critère d'agrégation : mesure de la distance entre deux classes d'objets.

Classe (ou groupe) : regroupement d'objets.

Centre de la classe : point moyen d'une classe d'objets.

5. En anglais : *density technique*

6. En anglais : *clumping technique* ou *fuzzy clustering*

Bibliographie

[Gordon, 1981] GORDON, A. D. (1981). Classification : Methods for the Exploratory Analysis of Multivariate Data. Chapman and Hall, London.