

Classification générative des données de grande dimension

Cristian PREDA

Polytech'Lille & Laboratoire Paul Painlevé, Université Lille 1
cristian.preda@polytech-lille.fr

Résumé

Le progrès continu de l'outil informatique tant en capacité de mémoire qu'en puissance de calcul permet d'enregistrer des données de plus en plus volumineuses. Ainsi, un très grand nombre de variables peuvent être observées pour l'étude d'un même phénomène. C'est le cas lorsqu'on observe une famille de variables $X = \{X_t\}_{t \in \mathcal{T}}$ indexée par un paramètre t variant dans un espace continu $\mathcal{T} \subseteq \mathbb{R}$. Dans certaines situations il est possible d'observer X_t pour tout instant $t \in \mathcal{T}$. C'est le cas des phénomènes dont les changements de comportement se produisent à des instants déterminés. Dans d'autres cas, il est techniquement impossible d'observer X_t pour tout $t \in \mathcal{T}$. On dispose alors d'une discrétisation suffisamment fine $\{t_i\}_{i=1, \dots, N}$ de \mathcal{T} pour pouvoir considérer que le comportement de $\{X_t\}_{t \in \mathcal{T}}$ peut être approché par celui de $\{X_{t_i}\}_{i=1, \dots, N}$. Des outils comme l'interpolation ou le lissage (lorsque les données sont observées en présence d'erreurs) sont employés pour rendre le caractère "fonctionnel" à ce type de données.

Nous allons présenter les principaux développements réalisés dans l'analyse des données fonctionnelles en évoquant notamment les analyses factorielles et les modèles de régression pour ce type de données. Des applications en économie et médecine sont présentées.

Références

- [1] Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statistics and Probability Letters*, 45, 11-22.
- [2] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and practice*. Springer-Verlag, New York.
- [3] Preda, C. and Saporta, G. (2005). PLS regression on a stochastic process. *Computational Statistics and Data Analysis*, 48, 149-158.
- [4] Preda, C., Saporta, G. and Lévêder, C. (2007). PLS classification for functional data. *Computational Statistics*, 22, 223-235.
- [5] Ramsay, J.O. and Silverman, B. W. (2002). *Applied Functional Data Analysis*. Springer-Verlag, New York.
- [6] Ramsay, J.O. and Silverman, B. W. (2005). *Functional Data Analysis, Second edition*. Springer-Verlag, New York.
- [7] Saporta G. (1981) Méthodes exploratoires d'analyse de données temporelles. *Cahiers du B.U.R.O., Université Pierre et Marie Curie, Paris*, 37-38.