Les transformations de Schoenberg : quelques applications en Analyse de Données

François Bavaud, Université de Lausanne

Soit D_{ij} le carré d'une distance euclidienne (quelconque) entre paires (i, j) de n points. La classe de toutes les fonctions $\varphi(D)$ transformant ces composantes en une autre distance euclidienne carrée $\tilde{D}_{ij} = \varphi(D_{ij})$ a été déterminée par Isaac J.Schoenberg en 1938. La classe, fort riche, est constituée des transformations $\varphi(D)$ nulles à l'origine, et dont la dérivée $\varphi'(D)$ est complètement monotone (i.e. $\varphi'(D) \geq 0$, $\varphi''(D) \leq 0$, $\varphi'''(D) \geq 0$, etc.), ou, de façon équivalente, telle que $\varphi'(D)$ soit la transformée de Laplace d'une mesure non négative.

Toute transformation de Schoenberg induit donc (via le *multidimensional scaling*) un plongement de l'espace euclidien original dans un nouvel espace euclidien, comparable aux plongements de haute dimensionalité utilisés en *Machine Learning*.

La plupart des méthodes classiques en Analyse de Données font intervenir explicitement ou implicitement des distances euclidiennes entre observations. Ainsi, toute transformation de Schoenberg induit une variante spécifique de ces méthodes. Des exemples en analyse en composantes principales, analyse des correspondances, analyse discriminante, partitionnement et en estimation robuste du centroïde illustreront ces possibilités.