

L'analyse des composantes principales ACP

Zakaria RIDADARAJAT

02/02/2021

Synthèse :

Le document évalué dans le cadre de ce rendu a été fait par Benjamin GUIGON sur l'analyse des composantes principales.

Github: <https://github.com/benjaminguigon/PSBX/blob/main/Article%20ACP%20.pdf>

L'ACP est une méthode factorielle de réduction de dimension pour l'exploration statistique de données quantitatives complexes dont l'objectif est de revenir à un espace de dimension réduite en déformant le moins possible la réalité. Il s'agit donc d'obtenir le résumé le plus pertinent possible des données initiales.

L'auteur du rapport commence par nous présenter la méthodologie pour construire le modèle associé à l'ACP, et ensuite il expose les différentes représentations graphiques du modèle construit, qui vont nous permettre d'interpréter les données multidimensionnelles.

L'auteur a aussi explicité les différentes métriques mathématiques essentielles qui sont au cœur du processus de la réduction.

Explication des formules

La première étape de l'analyse des composantes principales consiste à représenter les données au sein du matrice de variance-covariance qui va permettre de réaliser ce résumé pertinent.

Matrice des covariances :

$$S = \sum_{i=1}^n \omega_i (x_i - \bar{x})(x_i - \bar{x})' = \bar{X}' D \bar{X}$$

De cette matrice, on va extraire, par un procédé mathématique adéquat, les facteurs que l'on recherche, idéalement 3 facteurs. Ces facteurs vont permettre de réaliser les graphiques désirés dans cet espace de petite dimension, en déformant le moins possible la configuration globale des individus selon l'ensemble des variables initiales.

Evaluation

1) lisibilité du rapport :

Le document est un peu dense et trop chargé de formule mathématique.

2) qualité du rapport :

La rédaction de ce document est de bonne qualité.

3) Aspect didactique

La lecture de ce dossier nécessite des prérequis d'algèbre et de statistiques avancées.

5) Bibliographie :

La bibliographie n'est pas mentionnée.

5) Qualité du LaTeX

Les équations sont compliquées à reproduire, mais l'auteur a pu bien les maîtriser.

Conclusion :

Il s'agit d'un document très intéressant pour faire de l'analyse exploratoire. Cependant les notions mathématiques exposées sont un peu compliquées à comprendre après la première lecture mais ils sont en globale maîtrisés.