

ACP Parcimonieuse

Stéphane Caron/Sofia Harrouch

2018-03-13

Contents

Motivation derrière la méthodologie	1
Description de la méthodologie	1
Justification de la méthodologie	1
Application de la méthodologie	1
Autres éléments pertinents	1
Bibliographie	1

Motivation derrière la méthodologie

Les méthodes statistiques de réduction de la dimensionnalité ont généralement comme objectifs de réduire la dimension d'un jeu de données dans le but de simplifier l'interprétation des données, de permettre la visualisation des données ou même de d'améliorer la performance de certaines méthodes appliquées sur ces données réduites. En termes simples, réduire la dimensionnalité revient à réduire le nombre de variables (p) mesurées.

L'analyse en composante principale est une méthode classique de réduction de la dimensionnalité. Cette méthode permet de créer des combinaisons linéaires des différentes variables du jeu de données tout en conservant le plus de variabilité possible. Chacune des nouvelles composantes principales créées possèdent un vecteur de coefficients de saturation (loadings) de dimension $p \times 1$, correspondant en quelque sorte à l'importance attribuée à chacune des différentes variables originales du jeu de données. Il est donc possible d'interpréter ces coefficients de saturation et d'obtenir une interprétation plus généralisée de certaines composantes principales calculées.

Cette interprétation peut se révéler assez complexe dans le cas où une composante principale est expliquée (coefficients de saturation élevés) par plusieurs variables originales du jeu de données. De plus, il peut être difficile de définir à partir de quelle valeur exactement un coefficient de saturation est considéré comme étant "non important" pour une composante principale.

Description de la méthodologie

Justification de la méthodologie

Application de la méthodologie

Autres éléments pertinents

Bibliographie