Analyse en composantes principales

Jacques Bailhache (jacques.bailhache@gmail.com)

July 30, 2020

On a un tableau de données X à n lignes et p colonnes. Chaque ligne représente un individu et chaque colonne une variable. Ce tableau représente un ensemble de points dans un espace de dimension p.
L'analyse en composantes principales consiste à projeter ces points sur un sous-espace de dimension $q < p$ tel que les projections des points soient le plus écartés possible.

1 Cas q = 1 : projection sur une droite

Soit u_1 un vecteur unitaire (${}^t u_1 u_1 = 1$) de cette droite.
Les projections de ces n points sur cette droits sont :

$$C = X u_1$$

L'écartement de ces projections est mesuré par une valeur appelée inertie que l'on calcule par la formule :

$${}^t C C = {}^t u_1 {}^t X X u_1$$

Le problème consiste donc à maximiser ${}^t u_1 {}^t X X u_1$ sous la contrainte ${}^t u_1 u_1 = 1$.
On définit le lagrangien :

$$L(u_1, \lambda) = {}^t u_1 {}^t X X u_1 - \lambda_1 ({}^t u_1 u_1 - 1)$$

On a alors :

- $\frac{\partial L(u_1,\lambda_1)}{\partial u_1} = 0 \Rightarrow 2({}^t X X u_1 - \lambda_1 u_1) = 0 \Rightarrow {}^t X X u_1 = \lambda_1 u_1$
- $\frac{\partial L(u_1,\lambda_1)}{\partial \lambda_1} = 0 \Rightarrow {}^t u_1 u_1 - 1 = 0 \Rightarrow {}^t u_1 u_1 = 1$

u_1 est le premier vecteur propre de ${}^t X X$ et λ_1 la valeur propre correspondante.

2 Cas q = 2 : projection sur un plan

Le plan de projection optimal, c'est-à-dire celui qui maximise l'écartement des projections, contient nécessairement la droite optimale considérée dans la section précédente. Ce plan est donc défini par le vecteur u_1 déterminé précédemment, et par un autre vecteur u_2 que nous devons trouver.
On cherche u_2 tel que ${}^t u_2 {}^t X X u_2$ soit maximal, sous les contraintes ${}^t u_2 u_2 = 1$ et ${}^t u_2 u_1 = 0$.
Le lagrangien correspondant est :

$$L(u_2, \lambda_2, \mu) = {}^t u_2 {}^t X X u_2 - \lambda_2 ({}^t u_2 u_2 - 1) - \mu ({}^t u_2 u_1)$$

En écrivant que la dérivée partielle de ce lagrangien respectivement par rapport à u_2, λ_2, μ est égale à 0, on obtient :

- ${}^t X X u_2 = \lambda_2 u_2$
- ${}^t u_2 u_2 = 1$
- ${}^t u_1 u_2 = 0$

u_2 est le deuxième vecteur propre de ${}^t X X$ associé à la deuxième valeur propre λ_2.

3 Cas général

On peut montrer que le sous-espace de dimension q pour lequel les projections des points représentés par X sont le plus écartés est défini par les q vecteurs propres de ${}^tXX$ associés aux q plus grandes valeurs propres.

4 Références

- `http://cours.polymtl.ca/geo/marcotte/glq3402/chapitre3.pdf`
  ou `http://log.chez.com/text/math/chapitre3.pdf`
- `http://www.arnaud.martin.free.fr/Doc/polyAD.pdf`
  ou `http://log.chez.com/text/math/polyAD.pdf`
- `http://www2.agroparistech.fr/IMG/pdf/AnalyseComposantesPrincipales-AgroParisTech.pdf`
  ou `http://log.chez.com/text/math/AnalyseComposantesPrincipales-AgroParisTech.pdf`
- `http://asi.insa-rouen.fr/enseignants/~gasso/public/Courses/DM/pca.pdf`
  ou `http://log.chez.com/text/math/pca.pdf`
- `https://www.lamsade.dauphine.fr/~atif/lib/exe/fetch.php?media=teaching:coursad_acp.pdf`
  ou `http://log.chez.com/text/math/coursad_acp.pdf`
- `http://hamrita.e-monsite.com/medias/files/chap2ad.pdf`
  ou `http://log.chez.com/text/math/chap2ad.pdf`
- `http://www.foad-mooc.auf.org/IMG/pdf/M03-5.pdf`
  ou `http://log.chez.com/text/math/M03-5.pdf`
- `https://www.iro.umontreal.ca/~vincentp/ift3395/cours/continuous_latent_variables_print.pdf`
  ou `http://log.chez.com/text/math/continuous_latent_variables_print.pdf`
- `http://www.math.tu-dresden.de/~gournay/SMV.pdf`
  ou `http://log.chez.com/text/math/SMV.pdf`
- `http://www.ressources-actuarielles.net/EXT/ISFA/1226-02.nsf/d512ad5b22d73cc1c1257052003f1aed/da8b20974b2986`
  `$FILE/Me%CC%81moire%20TRIEU%20Thi%20Diep.pdf`
  ou `http://log.chez.com/text/math/Me%CC%81moire%20TRIEU%20Thi%20Diep.pdf`