Analyse canonique

Jacques Bailhache (jacques.bailhache@gmail.com)

July 30, 2020

1 Présentation

L'analyse canonique consiste, à partir de données avec deux groupes de variables, à trouver une combinaison linéaire des variables du premier groupe et une combinaison linéaire des variables du deuxième groupe qui soient les plus corrélées possibles.

2 Formulation mathématique

Les données sont représentées par deux tableaux X et Y correspondant aux deux groupes de variables. Les deux tableaux ont n lignes (nombre d'individus). Le tableau X a p colonnes (nombre de variables du premier groupe). Le tableau Y a q colonnes (nombre de variables du deuxième groupe).
Le problème consiste à trouver a et b qui rendent maximal ${}^ta^tXYb$ avec les contraintes ${}^ta^tXXa = 1$ et ${}^tb^tYYb = 1$.
Il s'agit de rendre maximal le lagrangien associé à ce problème :

$${}^ta^tXYb - \lambda({}^ta^tXXa - 1) - \mu({}^tb^tYYb - 1)$$

En écrivant que les dérivées partielles de ce lagrangien par rapport à a et b sont égales à 0, on obtient :

- ${}^tXYb - 2\lambda^tXXa = 0$
- ${}^tYXa - 2\mu^tYYb = 0$

En multipliant respectivement par ${}^ta$ et ${}^tb$ on obtient :

- ${}^ta^tXYb = 2\lambda$
- ${}^tb^tYXa = 2\mu$

On a donc $\lambda = \mu$. On pose $\beta = 2\lambda = 2\mu$. On a alors :

- ${}^tXYb = \beta^tXXa$
- ${}^tYXa = \beta^tYYb$

donc

$${}^tYX({}^tXX)^-1^tXYb = \beta^2\ {}^tYYb$$

Donc b est vecteur propre de

$$M = ({}^tYY)^{-1}\ {}^tYX({}^tXX)^{-1}\ {}^tXY$$

De même, a est vecteur propre de

$$N = ({}^tXX)^{-1}\ {}^tXY({}^tYY)^{-1}\ {}^tYX$$

et

$$a = \frac{1}{\beta}({}^tXX)^{-1}\ {}^tXYb\ ;\ b = \frac{1}{\beta}({}^tYY)^{-1}\ {}^tYXa$$

En multipliant respectivement par X ey Y on obtient :

- $Xa = \frac{1}{\beta} X ({}^t XX)^{-1}\, {}^t XYb$
- $Yb = \frac{1}{\beta} Y ({}^t YY)^{-1}\, {}^t YXa$

Dans ces formules apparaissent les opérateurs de projection orthogonale $P_X = X({}^t XX)^{-1}\, {}^t X$ et $P_Y = Y({}^t YY)^{-1}\, {}^t Y$. Donc chacun des vecteurs Xa et Yb est colinéaire à la projection de l'autre.

3 Référence

Statistique exploratoire multidimensionnelle, DUNOD, 2.1 Analyse canonique