Analyse factorielle generale

Introduction

AF = une famille de methodes geometrique dont l'un des objectifs est de resumer l'information du tableau de depart X par un tableau de plus faible dimension

$$X_{n,p} = \begin{array}{c|ccc} & 1 & & p \\ 1 & & & \\ i & & x_i^j & \\ n & & & \end{array}$$

x_i^j = valeur prise par l'individu i pour le caractere j

$$\underset{(n,p)}{X} \longrightarrow \underset{(n,q)}{Y} \quad \text{avec } (p > q)$$

- visualisez cette information (Tableau chiffrés)

L'AF permet de repondre à ce type de preoccupation en éliminant les redondance (repetitions) d'information contennues ds le tableau de depart.

De AF plusieurs variants seront developpés

ACP - AFC - AFCM - Analyse discriminante.

I Notations et problemes posés

Soit $X(n,p)$ un tableau de données quantitatif

$$X = \begin{array}{c|cc} & 1 & p \\ 1 & & \\ n & & \end{array}$$

→ variables. (paramètrés) ou caractè

$\forall i, j \quad x_i^j \in \mathbb{R}$.

individus, objets

$$"j" \quad I \longrightarrow \mathbb{R} \qquad I \text{ population de taille } n = \mathrm{Card}\, I)$$

X est un tableau a deux entrées.

$$X \nearrow X = [X^1, \ldots, X^p] \quad \text{avec } X^j \in \mathbb{R}^n$$

$$X \searrow X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \quad \text{avec } X_i \in \mathbb{R}^p$$

$(\mathbb{R}^n, \mathcal{B})$ $\mathcal{B}$ base canonique $\{e_1, \ldots, e_n\}$.

$\forall j$ $\mathbb{R}^n \xrightarrow{X^j} \mathbb{R}$.

$X^j \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}) = (\mathbb{R}^n)^*$

$\forall i$ $X^j(e_i) = x_i^j$

Le tableau X de depart : $X : \mathbb{R}^p \longrightarrow \mathbb{R}^n$

$X \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^n)$

$\mathbb{R}^n$ est $\mathbb{R}^p$ sont respectivement munis de leurs base canonique

$\mathcal{B}' = \{e'_1, \ldots, e'_p\} \longrightarrow \mathbb{R}^p$

$\mathcal{B} = \{e_1, \ldots, e_n\} \longrightarrow \mathbb{R}^n$

$\forall j$ $X(e'_j) = X^j = \sum_{i=1}^{n} x_i^j e_i$

Remarque

$X \longrightarrow \begin{cases} \mathbb{R}^n \\ \mathbb{R}^p \end{cases}$

$X = \{n$ pts de $\mathbb{R}^p\}$ ou $X = \{p$ pts de $\mathbb{R}^n\}$

Pb posé

On se place sur $\mathbb{R}^p$ $X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ avec $X_i^j \in \mathbb{R}$ ${}^tX_i \in \mathbb{R}^p$

Determiner les sous espaces vectoriels de dim $1, \ldots, r$ qui ajuste au mieux ce nuage de n points de $\mathbb{R}^p$

Hypothese : ajusté mieux $\Leftrightarrow$ l'ajustement se fait avec le minimum de perte d'information

$\Leftrightarrow$ satisfait au critere de Ω c.

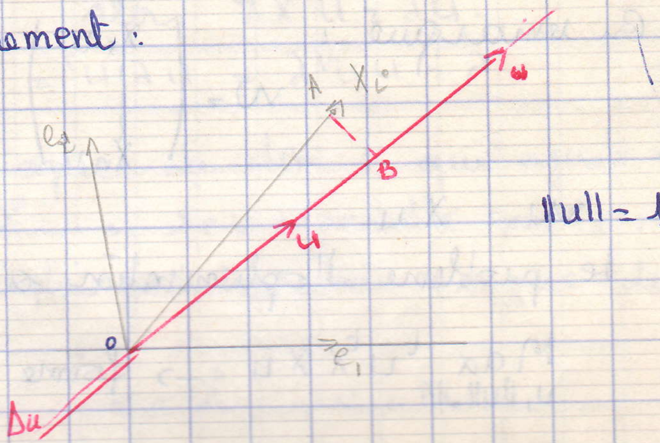
a – Recherche du s e v de dim 1.

Ce qui equivalent a la recherche du vecteur u (u base de E_1) tel que la norme de u soit egale a 1. et tq Δ_u ajuste au mieux le nuage de n pts de $\mathbb{R}^p$

$\Leftrightarrow$ | La somme des ecarts de pts a leur projeté sur ~~la~~ l'axe soit minimale.

Schematiquement :

$(\mathbb{R}^p, I_p)$



$\|u\| = 1$, $u \in \mathbb{R}^p$

$u = ?$ $\|u\| = 1$ tq $\sum_{i=1} [A_i B_i]^2$ soit minimum

On a : (theor de pythagore)

$\forall i = 1, \ldots n$ $[A_i B_i]^2 = \|{}^t X_i\|^2 - (OB_i)^2$

Pb 1 : $\min\limits_{\substack{u \\ \|u\|=1}} \sum_{i=1}^n [A_i B_i]^2 = \underbrace{\sum_{i=1}^n \|{}^t X_i\|^2}_{C^{te}} - [OB_i]^2$

donc $\max\limits_{u, \|u\|=1} \sum_{i=1}^n (OB_i)^2$

$\overline{OB_i} = \|{}^t X_i\| \cos \alpha$

$\langle u, {}^t X_i \rangle = \|u\| \, \|{}^t X_i\| \cos \alpha = \|{}^t X_i\| \cos \alpha = \overline{OB_i}$

${}^t({}^t X_i)\, u = X_i u = \overline{OB_i}$

Le probleme d'optimisation lie a la recherche de Δu (u) s'ecrit :

$$\max_{u, \|u\|} \sum_{i=1}^n \| X_i u\|^2$$

Soit $w_i = \overline{OP_{oi}} = X_i u$, $w_i \longrightarrow$ la representation de l'objet i sur l'axe

$$W = \begin{pmatrix} w_1 \\ | \\ w_n \end{pmatrix} \in \mathbb{R}^n \longrightarrow \text{les n individus sur l'axe } \Delta$$

$$\underset{u,\|u\|=1}{Max} \sum_{i=1}^{n} \|X_i u\|^2 = \sum_{i=1}^{n} w_i^2 = \|W\|^2 = {}^tW W$$

On remarque :

$$W = \begin{pmatrix} X_1 u \\ | \\ X_n u \end{pmatrix} = \begin{pmatrix} X_1 \\ | \\ X_n \end{pmatrix} u$$

$W = Xu$

le probleme d'optimisation a resoudre est :

$$\underset{u,\|u\|=1}{Max}\ {}^tu\,{}^tX X u \longrightarrow \text{forme quadratique}$$

Le lagrangien

$$\mathcal{L}(u) = {}^tu\,{}^tX X u - \lambda({}^tu u - 1)$$

$$\frac{\partial \mathcal{L}}{\partial u} = 0 \quad \text{(condition necessaire)}$$

Rappels : derivation d'une fonction quadratique.

$$f : \mathbb{R}^M \longrightarrow \mathbb{R}.$$

$$u \longrightarrow \|u\|_M^2 = \langle u,u\rangle_M = M(\langle u,u\rangle) \quad (\text{M metrique}$$

$df \in \mathcal{L}(\mathbb{R}^M, \mathbb{R})$ tq.

$$f(u+H) - f(u) = df(H) + \varepsilon(H) \quad \varepsilon(H) \longrightarrow 0 \text{ qd } \|H\| \longrightarrow 0$$

On a :

$$f(u+H) = \langle u+H, u+H\rangle = \|u\|_M^2 + \|H\|_M^2 + \underset{\in\mathbb{R}}{\langle u,H\rangle_M} + \underset{\in\mathbb{R}}{\langle H,u\rangle_M}$$

$$f(u+H) - f(u) = 2\,{}^tuMH + \|H\|^2$$

$$\frac{\partial f}{\partial u} = 2\,{}^tuM.$$

$$\frac{\partial \mathcal{L}}{\partial u} = 2\,{}^t u\,{}^t X X - 2\lambda\,{}^t u = 0$$

$${}^t u\,{}^t X X = \lambda\,{}^t u \iff {}^t X X u = \lambda u$$

u vecteur propre de la matrice ${}^t X X$ associé (sauf) à la valeur propre λ.

On cherche $\underset{u,\ \|u\|=1}{\mathrm{Max}}\ {}^t u\,{}^t X X u$.

Si u existe alors ${}^t X X u = \lambda u$.

$\Rightarrow {}^t u\,{}^t X X = {}^t u \lambda u = \lambda\,{}^t u u = \lambda$.

E_1 le sous espace de dim 1 qui ajuste au mieux les n points de $\mathbb{R}^p$ = sev engendré par u_1, le vecteur propre associé à la plus grande valeur propre de la matrice ${}^t X X$

$W = X u \qquad W \in \mathbb{R}^n \quad , W_i \in \mathbb{R}$

$\|W\| = {}^t W . W = \langle W, W \rangle_{I_n}$

Proposition

En notant par E_2 le sev de dim 2 de $\mathbb{R}^p$

(E_2 = sev qui ajuste au mieux le nuage des n pt de $\mathbb{R}^p$)

Alors

$$E_1 \subset E_2.$$

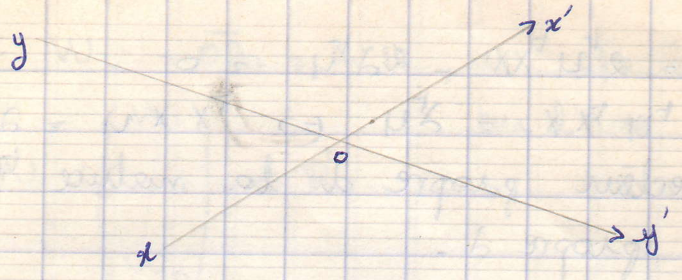
Remarque

Si $E_2 \supset E_1 \Rightarrow$ la recherche de E_2 se limite à la recherche d'un autre vecteur u_2 ($\|u_2\| = 1$ et $u_2 \perp u_2$)

Preuve

Dim $E_2 = 2 \Rightarrow E_2$ engendré par deux axe distincts

Soit donc (x, x') et (y, y') les deux axes.



Si $E_1 \not\subset E_2$

$E_1 = \{ x \in \mathbb{R}^2 , \ x = \lambda u_1 \} = \Delta u_1$

$\left.\begin{array}{l} \Delta u_1 \neq xx' \\ \Delta u_2 \neq yy' \end{array}\right\}$ et

Si on considère le sev $\tilde{E}_2$ engendré par xx' et Δu

La représentation sur Δu_1 et meilleur sur yy'

$\Rightarrow$ $\hat{E}_2$ ajuste au mieux le nuage que E_2.

$\Rightarrow$ E_2 n'est pas meilleur $\longrightarrow$ contradiction

Conclusion $E_2 \supset E_1$

déterminer $E_2 \Leftrightarrow$ recherche Δu_2 ?

Le pb d'optimisation s'écrit :

$$\underset{\substack{\|u_2\| = 1 \\ \langle u_2 , u_1 \rangle = 0}}{\mathrm{Max}} \| W^{(2)} \|^2 \qquad (5)$$

$$W^{(2)} = \begin{pmatrix} W_1^{(2)} \\ \vdots \\ W_n^{(2)} \end{pmatrix} \in \mathbb{R}^n$$

$W_i^{(2)}$ = mesure algébrique de la projection du vecteur ${}^tX_i$ sur Δ

$W_i^{(2)} = X_i \cdot u_2$.

$W^{(2)} = X u_2$.

$\| W^{(2)} \|^2 = {}^t u_2 \, {}^tX \, X u_2 = \langle W^{(2)} , W^{(2)} \rangle$

Resolution du pb (5)

Extremum lié à deux contraintes

Le lagrangien s'écrit

$$\mathcal{L}(u_1, u_2) = {}^tu_2 {}^tX X u_2 - \lambda({}^tu_2 u_1 - 1) - \mu\, {}^tu_2 u_1$$

$\left|\dfrac{\partial \mathcal{L}}{\partial u_1} = 0 \quad \Rightarrow \mu\, {}^tu_2 = 0 \quad \Rightarrow \mu = 0\right.$

$\left|\dfrac{\partial \mathcal{L}}{\partial u_2} = 0 \quad \Rightarrow 2\, {}^tu_2 {}^tX X - 2\lambda {}^tu_2 - \mu u_1 = 0\right.$

$\Rightarrow {}^tu_2 {}^tX X = \lambda {}^tu_2$. (Transposés des 2 membres ⇒)

$\Rightarrow {}^tX X u_2 = \lambda u_2$

u_2 vecteur propre de ${}^tXX$.

$\|\omega^{(2)}\| = {}^tu_2 {}^tX X u_2 = {}^tu_2 \lambda u_2 = \lambda \|u_2\|^2$

Le $\max \|\omega^{(2)}\|^2$ est obtenu en choisissant λ max

E_2 sev engendré par u_1 et u_2 où :

u_1 vect propre associé à $\lambda_1 = \max\limits_{i=1,p} \lambda_i$

u_2 vect propre associé à $\lambda_2 = \max\limits_{\lambda_i \neq \lambda_1} \lambda_i$

Generalisation :

Le sev E_r de dim. r, contient $E_1, E_2, \ldots, E_{r-1}$.

$(E_1 \subset E_2 \subset \ldots \subset E_{r-1} \subset E_r)$

E_r sev engendré par les axes orthogonaux $\Delta u_1, \ldots, \Delta u_r$

où u_j vecteur propre de la matrice ${}^tXX$ associé à λ_j

où $(\lambda_1 > \lambda_2 \ldots > \lambda_{j-1} > \lambda_j > \ldots > \lambda_r)$

Remarque

En pratique on se restreint en general à deux ou trois axes.

Pour les raisons :

- La representation graphique est possible.

sur les deux premiers axes qu'on recueille le max d'information, car les valeurs propre de la matrice ${}^tXX$ vi très rapidement.

Definition 1

Les axes engendrés par les vect propre normés, $\{\Delta u_r,\ r\geq 1\}$ associés resp au val propre λ_r de ${}^tXX$ sont dits axes factoriels.

Δu_1 = 1er axe factoriel

$\vdots$

Δu_j = $j^{ème}$ axe fact

Remarque.

Parfois la projections des individus sur le 1ère plan factoriel $(\Delta u_1, \Delta u_2)$ il y a beaucoup p^{ts} qui ne sont pas representés, au lieu de passer au sev à dim on fait des rotations d'axes.

$(\Delta u_1, \Delta u_3)$ 2ème plan fact

$(\Delta u_2, \Delta u_3)$.

Definition 2.

les operateurs de projections des p^{ts} ${}^tX_i$ sur $(\Delta u_r)_{r\geq 1}$ sont dits facteurs

L'operateur de proj $\mathbb{R}^p \longrightarrow \mathbb{R}$. l'app $\varphi_{(r)}$ = facteur.

${}^tX_i \longrightarrow X_i u_r = W_i$

axe factoriel Δu_r / ${}^tXX u_r = \lambda_r u_r$.

Proposition

1 - Les vp $\lambda_1 > \lambda_2 > \text{—} \lambda_r \text{—}$ sont non negatives

2 - Si λ_m est la $1^{ère}$ vp nulle de ce rangement

$\lambda_1 > \text{——} \lambda_{m-1} > 0$

$rg\ ({}^tX \cdot X) = m-1$

et $\forall i = 1, \text{—} m \quad {}^tX_i \in E_{m-1}^{\perp} = \text{orthog } E_{m-1}$

où E_{m-1} sev engendré par $u_1, \text{—}, u_{m-1}$

Preuve.

Soit λ vp de ${}^tX X$ et vect propre associé

${}^tX X u = \lambda u.$

${}^tu\,({}^tX X u) = {}^tu\,(\lambda u)$

${}^tu\,{}^tX X u = \lambda\,{}^tu u \Leftrightarrow \|Xu\|^2 = \lambda \|u\| = \lambda \geq 0.$

Posons $A = {}^tX X$

. A est une matrice symétrique ${}^tA = {}^t({}^tX X) = A$

A est définie symétique.

$\forall u \in \mathbb{R}^? \quad A(u, u) = \langle u . u \rangle_A = {}^tu A u \geq 0$?

On a ${}^tu A u = {}^tu {}^tX X u = {}^t(Xu)(Xu) = \langle Xu, Xu \rangle_I = \|Xu\|_I^2 \geq 0$

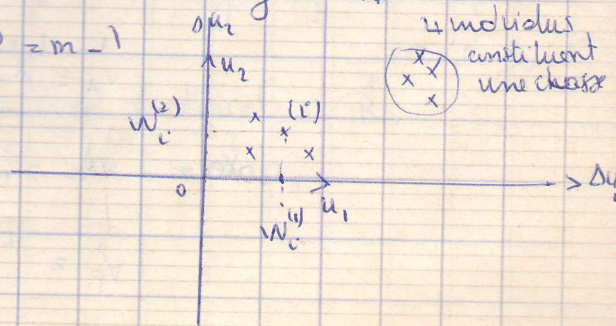
A étant symétrique définie positive $\Rightarrow$ A diagonalisable

$\Rightarrow \exists P$ inversible telle que $A = P^{-1} D P$ où Diagonale

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_{m-1} \end{pmatrix} \qquad P = (u_1, \text{—}, u_{m-1})$$

$rg\ D = m-1$ et $rg\ P = m-1 \Rightarrow rg\ P^{-1} = m-1$

$A = P^{-1} D P \Rightarrow rg\ A = rg\ D = m-1$



Remarque

$\|u_1\| = \|u_2\| = 1$

$u_1 \perp u_2$

Pbs

? Paramètres qui ont contribués à la construction de la classe.

$\Rightarrow$ Necessité d'une representation simultanée, individus caractère, sur les axes et les plans factoriels

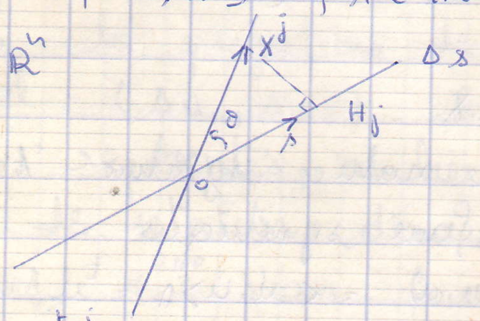
Proposition

Dans le cas où on se place sur $(\mathbb{R}^n, I_n)$ les axes factoriels Δ_s sont les axes passant par $O_{\mathbb{R}^n}$ et engendrés par les vecteurs propres de la matrice X^tX

Demonstration

On se place sur $(\mathbb{R}^n, I_n)$

$X = [X^1, \_\_, X^p]$, $X^j \in \mathbb{R}^n$



$\overline{OH_j} = {}^tX^j s$

$\langle s, X^j \rangle = \underbrace{\|s\|}_{=1} \cdot \|X^j\| \cos(\theta) = \|X^j\| \cos\theta$

$\cos\theta = \dfrac{\overline{OH_j}}{\|X^j\|} \iff \overline{OH_j} = \|X^j\| \cos\theta = \langle s, X^j \rangle = {}^tX^j s$

En notant par :

$$V^j = \begin{pmatrix} V_1^1 \\ | \\ | \\ V_p^p \end{pmatrix} \in \mathbb{R}^p$$

avec
$$\left| \begin{array}{l} V_1^1 = {}^tX^1 s \\ | \\ V_j^j = {}^tX^j s \longrightarrow \text{proj du paramètre 1 sur } \Delta s \\ | \\ V_p^p = {}^tX^p s \end{array} \right.$$

$v^j = {}^t X s$ En effet.

$$V = \begin{pmatrix} {}^t X^1 s \\ \vdots \\ {}^t X^p s \end{pmatrix} = \begin{pmatrix} {}^t X^1 \\ \vdots \\ {}^t X^p \end{pmatrix} s = {}^t X s$$

Le pb d'optimisation : $\underset{\|s\|=1}{\underset{s}{\max}} \|V\|^2$ critère des nc.

$\|V\|^2 = {}^t V V = {}^t s X {}^t X s$.

$\underset{\|s\|=1}{\underset{s}{\max}} \; {}^t s X {}^t X s$.

$\mathcal{L}(s) = {}^t s X {}^t X s - \lambda({}^t s s - 1)$

$\dfrac{\partial \mathcal{L}(s)}{\partial s} = 0 = 2\, {}^t s X {}^t X - 2\lambda\, {}^t s = 0$

${}^t X$ $\quad X {}^t X s = \lambda s \Rightarrow s$ vect prop associé à la vp λ

s est le $\vec{vp}$ de la matrice $X {}^t X$ $\quad X {}^t X s = \lambda s$.

$\Rightarrow {}^t s (X {}^t X s) = \lambda\, {}^t s s = \lambda \qquad \|s\| = 1$

Pour avoir le max, on choisit le $\vec{vp}$, associé à la plus grande valeur propre de la matrice $X {}^t X$

$s_1 \longrightarrow \vec{vp}$ de $X {}^t X \longrightarrow \lambda_1$ + grande vp $\longrightarrow \Delta s_1$ 1ère axe fa

$s_2 \longrightarrow \vec{vp}$ de $X {}^t X \longrightarrow \lambda_2$ 2ème + grd vp $\longrightarrow \Delta s_2$ 2ème ax

$\vdots$

$s_r \longrightarrow \vec{vp}$ de $X {}^t X \longrightarrow \lambda_r$ rième + grd vp $\longrightarrow \Delta s_r$ rième axe fact

<u>représentation des ind sur les axes factoriel</u> $\rightarrow \mathbb{R}^p$

individus $\longrightarrow \Delta u_r \longrightarrow w^{(r)} = X u_r$

representation des parametres sur les axe fact $\longrightarrow \mathbb{R}^n$

variables: $\longrightarrow \Delta_{s_r} \longrightarrow V^{(r)} = {}^tX\, s_r$

Proposition

Les matrices ${}^tX\,X$ et $X\,{}^tX$ ont les m̂ valeurs propres.

Demonstration

Soit λ une valeur propre de la matrice ${}^tXX$ et $\vec{vp}$ correspondant

$${}^tX\,X\,v = \lambda v$$

$\Rightarrow$ En multipliant par la matrice X.

$$X({}^tXXv) = X\lambda v$$

$$= (X{}^tX)(Xv) = \lambda Xv$$

λ valeur propre de $X{}^tX$ et Xv le vect prop associe

Proposition

En notant par u_r le $\vec{vp}$ de la matrice ${}^tX{}^tX$

s_r " " " " ${}^tXX$

On a $\begin{cases} s_r = Xu_r \cdot \dfrac{1}{\sqrt{\lambda_r}} & (1) \\ u_r = {}^tX\,s_r \cdot \dfrac{1}{\sqrt{\lambda_r}} & (2) \end{cases}$

Demonstraction

$$X{}^tX\left(\frac{1}{\sqrt{\lambda_r}}Xu_r\right) = \lambda_r\left(\frac{1}{\sqrt{\lambda_r}}Xu_r\right) ? \text{ et } \left\|\frac{1}{\sqrt{\lambda_r}}Xu_r\right\| \overset{?}{=}$$

$$\frac{1}{\sqrt{\lambda_r}}X{}^tXXu = \frac{1}{\sqrt{\lambda_r}}X\lambda_r u_r = \lambda_r\left(\frac{1}{\sqrt{\lambda_r}}Xu_r\right)$$

$$\left\|\frac{1}{\sqrt{\lambda_r}}Xu_r\right\| = \left\langle\frac{1}{\sqrt{\lambda_r}}Xu_r, \frac{1}{\sqrt{\lambda_r}}Xu_r\right\rangle = \frac{1}{\lambda_r}\langle Xu_r, Xu_r\rangle = \frac{1}{\lambda_r}{}^t(Xu_r)(X$$

$$= \frac{1}{\lambda_r}({}^tu_r\underbrace{{}^tXXu_r}_{\lambda_r u_r}) = {}^tu_ru_r = \|u_r\| = 1$$