USTHB - Faculté d'Informatique **06 Mars 2022**
Département d'IA et SD

M1 : MIV
Enseignant : N. Laiche

Analyse en composantes principales

5. Ana lyse des points

L'analyse des données en composantes principales consiste à étudier les projections des points du nuage sur un axe, un plan ou bien un hyperplan bien choisi. Autrement dit tout revient à déterminer le meilleur ajustement du nuage par un sous espace de $\Re^n$ pour $n \geq 1$.

i) Ajustement du nuage par une droite (un axe)

Soit la droite (Δu) passant par l'origine, et engendré par le vecteur unitaire $\vec{u}$ i.e $\|\vec{u}\| = 1$

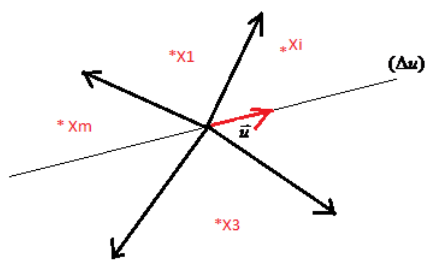

Figure 1. Présentation de l'axe de projection.

Alors la projection de X_i sur (Δu) notée x_i est définie par : $x_i = X_i.\vec{u}$ et ce n'est que le produit scalaire $\langle X_i, \vec{u} \rangle = \langle \vec{u}, X_i \rangle$.

Partant du théorème de Pythagore (voir figure 2), nous avons : $x_i^2 + e_i^2 = \|X_i\|^2$ i.e

$$x_i^2 = \|X_i\|^2 - e_i^2$$

où x_i représente l'information projetée sur la l'axe i.e la droite et e_i n'est que l'erreur ou bien l'information perdue.

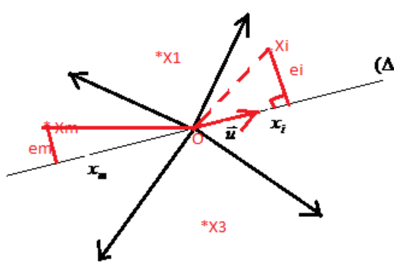

Figure 2. Projection des individus sur (Δu) .

Ainsi, pour maximiser l'information, il suffit de minimiser la perte d'information e_i .

Analyse en composantes principales

Ceci est vrai pour chaque individu X_i, $i = 1, 2, \ldots, m$.

Donc pour avoir l'information totale maximale, il faut :

1) Projeter tous les points (individus) sur la droite (Δu).
2) Choisir $\vec{u}$ tel que la somme des carrées de ces projections soit maximale.

C'est à dire **maximiser :** $\sum_{i=1}^{m} x_i^2$ **.**

Or $x_i^2 = \langle X_i\vec{u}, X_i\vec{u} \rangle = {}^t(X_i\vec{u}) \times (X_i\vec{u})$, alors :

$$\begin{aligned}\sum_{i=1}^{m} x_i^2 &= \sum_{i=1}^{m} {}^t(X_i\vec{u}) \times (X_i\vec{u}) = \sum_{i=1}^{m} {}^t\vec{u} \cdot {}^tX_i \cdot X_i \cdot \vec{u} \\ &= {}^t\vec{u} \times \left(\sum_{i=1}^{m} {}^tX_i \cdot X_i \right) \times \vec{u} = {}^t\vec{u} \times ({}^tX \cdot X) \times \vec{u}\end{aligned}.$$

D'où tout revient à maximiser $\begin{cases} {}^t\vec{u} \cdot ({}^tX \cdot X) \cdot \vec{u} \\ \|\vec{u}\|^2 = {}^t\vec{u} \cdot \vec{u} = 1 \end{cases}$. (P)

Il s'agit d'un problème d'optimisation avec contraintes. Pour que nous puissions résoudre ce problème, nous donnons le rappel suivant :

<u>Rappels</u>

Multiplicateur de Lagrange Le multiplicateur de Lagrange est une méthode d'optimisation permettant de trouver les points stationnaires d'une fonction dérivable sous-contraintes. Formellement, l'écriture du Lagrangien est donnée par :

$$L(X, \lambda) = f(x) + \lambda \cdot g(x). \text{ Avec}$$

$-X$: les variables figurant dans la fonction à optimiser.
$-f(x)$: la fonction à optimiser.
$-\lambda$: le multiplicateur de Lagrange = inconnu à déterminer.
$-g(x)$: la contrainte à imposer dans le problème à résoudre.

A lors la solution est obtenue en résolvant le système des dérivées partielles suivant (il s'agit d'une condition nécessaire d'existence de solution) :

$$Sol = \begin{cases} \dfrac{\partial L}{\partial x_i} = \dfrac{\partial f}{\partial x_i} + \lambda \dfrac{\partial g}{\partial x_i} = 0, \quad \forall i \;\; et \;\; X = (x_i)_i \\ \dfrac{\partial L}{\partial \lambda} = g(x) = 0 \end{cases}.$$

Analyse en composantes principales

Donc en appliquant le multiplicateur de Lagrange à notre problème (P), nous définissons la fonction de Lagrange comme suit :

$$L(\vec{u},\lambda)= {}^t\vec{u}\cdot\left({}^tX\cdot X\right)\cdot\vec{u}-\lambda\cdot\left({}^t\vec{u}\cdot\vec{u}-1\right),$$

et la solution est donnée par : $Sol=\begin{cases}\dfrac{\partial L}{\partial \vec{u}}=0\\ \dfrac{\partial L}{\partial \lambda}=0\end{cases}$

C'est-à-dire :

$$Sol=\begin{cases}2\cdot({}^tX\cdot X)\cdot\vec{u}-2\lambda\cdot\vec{u}=0 \;\Leftrightarrow\; ({}^tX\cdot X)\cdot\vec{u}-\lambda\cdot\vec{u}=0 \qquad (1)\\ et\\ {}^t\vec{u}\cdot\vec{u}=1 \;\; i.e\;\; \|\vec{u}\|^2=1\end{cases}.$$

L'équation (1) est équivalente à $({}^tX\cdot X)\cdot\vec{u}=\lambda\cdot\vec{u}$

C'est-à-dire :

$\vec{u}$ est un **vecteur propre** de la matrice ${}^tX\cdot X$ associé à la **valeur propre** λ.

Par suite :

$${}^t\vec{u}\cdot({}^tX\cdot X)\cdot\vec{u} = {}^t\vec{u}\cdot\lambda\cdot\vec{u} = \lambda\cdot({}^t\vec{u}\cdot\vec{u}) = \lambda\cdot\|\vec{u}\|^2 = \lambda$$

D'où le **maximum** de ${}^t\vec{u}\cdot\left({}^tX\cdot X\right)\cdot\vec{u}$ sous la contrainte $\|\vec{u}\|^2=1$ correspond à la **plus grande valeur propre** de ${}^tX\cdot X$. Par conséquent la meilleure droite (Δu) qui n'est que le meilleur axe qui ajuste le nuage des points est celle engendrée par le vecteur propre normé $\vec{u}$ associé à la plus grande valeur propre λ de ${}^tX\cdot X$.

ii) Ajustement du nuage par un plan

Pour déterminer le meilleur plan ajustant le nuage, il suffit de déterminer la deuxième droite (Δv) de vecteur unitaire $\vec{v}$ i.e $\|\vec{v}\|^2=1$ passant par l'origine et perpendiculaire à (Δu), la droite déterminée précédemment c'est à dire $\langle\vec{u},\vec{v}\rangle=0$.

Analyse en composantes principales

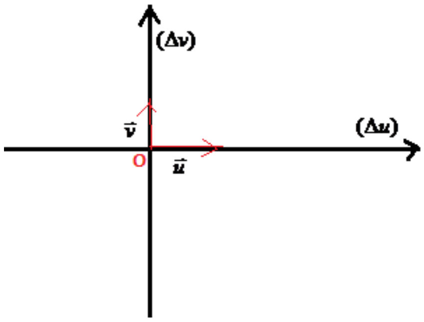

Figure 3. Présentation du deuxième axe.

Donc,
tout revient à maximiser ${}^t\vec{v} \cdot \left({}^tX \cdot X\right) \cdot \vec{v}$ sous les contraintes ${}^t\vec{v} \cdot \vec{v} = 1$ et ${}^t\vec{v} \cdot \vec{u} = 0 = {}^t\vec{u} \cdot \vec{v}$.
Ainsi,
le Lagrangien cette fois ci s'écrit comme suit

$$L(\vec{v}, \lambda) = {}^t\vec{v} \cdot \left({}^tX \cdot X\right) \cdot \vec{v} - \lambda \cdot \left({}^t\vec{v} \cdot \vec{v} - 1\right) - \mu \cdot \left({}^t\vec{v} \cdot \vec{u} - 0\right).$$

Où λ et μ sont les multiplicateurs de Lagrange.

La solution est donnée par :

$$Sol = \begin{cases} \dfrac{\partial L}{\partial \vec{v}} = 0 \\ {}^t\vec{v} \cdot \vec{v} = 1 \\ {}^t\vec{v} \cdot \vec{u} = 0 \\ \end{cases}$$

C'est-à-dire :

$$Sol = \begin{cases} 2 \cdot ({}^tX \cdot X) \cdot \vec{v} - 2\lambda \cdot \vec{v} - \mu.\vec{u} = 0 & (2) \\ et \\ {}^t\vec{v} \cdot \vec{v} = 1 \;\; i.e \;\; \|\vec{v}\|^2 = 1, \;\; {}^t\vec{v} \cdot \vec{u} = 0 \end{cases}.$$

En multipliant l'équation (2) par ${}^t\vec{u}$, nous obtenons :

$$2 \cdot {}^t\vec{u} \cdot ({}^tX \cdot X) \cdot v - 2 \cdot \lambda {}^t\vec{u} \cdot \vec{v} - \mu \cdot {}^t\vec{u} \cdot \vec{u} = 0 \quad \text{c'est-à-dire} \quad 2 \cdot {}^t\vec{u} \cdot ({}^tX \cdot X) \cdot v = \mu .$$

Or $({}^tX \cdot X) \cdot \vec{u} = \lambda \vec{u}$ c'est-à-dire ${}^t\vec{u} \cdot ({}^tX \cdot X) = \lambda \cdot {}^t\vec{u}$, alors : $\mu = 2\lambda \cdot {}^t\vec{u} . \vec{v} = 0$

Analyse en composantes principales

Donc,

$$\frac{\partial L}{\partial \vec{v}} = 0 \Leftrightarrow 2 \cdot ({}^tX \cdot X) \cdot \vec{v} - 2\lambda \cdot \vec{v} = 0 \Leftrightarrow ({}^tX \cdot X) \cdot \vec{v} = \lambda \cdot \vec{v} .$$

C'est-à-dire $\vec{v}$ est un vecteur propre de la matrice ${}^tX \cdot X$ associé à la valeur propre λ, et donc le maximum de ${}^t\vec{v} \cdot \left({}^tX \cdot X\right) \cdot \vec{v}$ correspond à la deuxième grande valeur propre de ${}^tX \cdot X$.
Ainsi, le meilleur plan est constitué du 1$^{\text{er}}$ axe factoriel (Δu) et le second axe factoriel (Δv) .

iii) Ajustement du nuage par un sous espace vectoriel

Les procédures précédentes peuvent être généralisées à 3, 4, … axes factoriels.
A chaque valeur propre de ${}^tX \cdot X$, nous pouvons associer un vecteur propre correspondant à un axe principal (factoriel). Les vecteurs propres calculés constituent ainsi une base orthonormée du sous espace vectoriel correspondant qui représente le meilleur ajustement.

6. Ana lyse en composantes principales

Le principe de l'ACP est de trouver des sous espaces de dimension $q \le n$, issus de combinaisons linéaires des variables X^i, $i = 1, 2, \ldots, n$ telles que la variance du nuage autour de ces espaces soit maximale. Dans la mise en œuvre de cette méthode, il est souhaitable de centrer et de réduire les données du tableau (matrice initiale) X . Pour ce faire nous appliquons les deux transformations suivantes :

i) Centrage des données Soit $g \in \mathfrak{R}^n$, le centre de gravité du nuage des points :

$$g = (\bar{X}^1, \bar{X}^2, \ldots, \bar{X}^n) \text{ où } \bar{X}^j = \frac{1}{m}\sum_{i=1}^{m} x_i^j \text{ pour } j = 1, 2, \ldots, n .$$

Pour centrer le nuage de points, nous centrons les variables X^j, $j = 1, 2, \ldots, n$ par la transformation suivante : $X^j \to Y^j = X^j - \bar{X}^j \times \mathrm{I}$ où $I = {}^t(1, 1, \ldots, 1) \in \mathfrak{R}^m$.

Ainsi, le tableau (matrice) X sera transformée en Y :

$$Y = \begin{pmatrix} y_1^1 & \cdots & y_1^j & \cdots & y_1^n \\ \vdots & & \vdots & & \vdots \\ y_i^1 & \cdots & y_i^j & \cdots & y_i^n \\ \vdots & & \vdots & & \vdots \\ y_m^1 & \cdots & y_m^j & \cdots & y_m^n \end{pmatrix} \text{ où } y_i^j = x_i^j - \bar{X}^j .$$

Cette dernière matrice nous permet de définir la matrice V de covariance des variables X^j comme suit :

USTHB - Faculté d'Informatique **06 Mars 2022**
Département d'IA et SD

M1 : MIV
Enseignant : N. Laiche

Analyse en composantes principales

$$V = \begin{pmatrix} \mathrm{cov}(X^1, X^1) & \cdots & \mathrm{cov}(X^1, X^j) & \cdots & \mathrm{cov}(X^1, X^n) \\ \vdots & & \vdots & & \vdots \\ \mathrm{cov}(X^i, X^1) & \cdots & \mathrm{cov}(X^i, X^j) & \cdots & \mathrm{cov}(X^i, X^n) \\ \vdots & & \vdots & & \vdots \\ \mathrm{cov}(X^n, X^1) & \cdots & \mathrm{cov}(X^n, X^j) & \cdots & \mathrm{cov}(X^n, X^n \end{pmatrix}.$$

Avec :

$$\mathrm{cov}(X^i,\ X^j) = \frac{1}{m}\sum_{k=1}^{m}(x_k^i - \overline{X}^i)\cdot(x_k^j - \overline{X}^j) \quad \text{pour tout } i, j = 1, 2, \ldots, n$$

$$= \frac{1}{m}\sum_{k=1}^{m} y_k^i \cdot y_k^j = \frac{1}{m}({}^tY^i \cdot Y^j).$$

Ce qui donne :

$$V = \frac{1}{m}({}^tY \cdot Y).$$

Remarque La matrice des variances-covariances des données centrées V est une matrice symétrique c'est-à-dire ${}^tV = V$. Sur la diagonale, nous avons les $\mathrm{var}\,(X^i)$, $i = 1, 2, \ldots, n$.
Car pour : $i = j$,

$$\mathrm{cov}(X^i,\ X^i) = \frac{1}{m}\sum_{k=1}^{m}(x_k^i - \overline{X}^i)^2 = \mathrm{var}\,(X^i).$$

ii) Centrage et Réduction des données Les données centrées réduites sont définies par : $z_i^j = \dfrac{y_i^j}{s_j}$ pour tout $i, j = 1, 2, \ldots, n$. Où s_j est l'écart type de la variable X^j .

$$\text{D'où} \quad Z = \begin{pmatrix} z_1^1 & \cdots & z_1^j & \cdots & z_1^n \\ \vdots & & \vdots & & \vdots \\ z_i^1 & \cdots & z_i^j & \cdots & z_i^n \\ \vdots & & \vdots & & \vdots \\ z_m^1 & \cdots & z_m^j & \cdots & z_m^n \end{pmatrix}.$$

Cette matrice nous permet de définir la matrice de corrélation R qui est donnée par :

Analyse en composantes principales

$$R=\begin{pmatrix} 1 & r_1^2 & \cdots & \cdots & r_1^n \\ r_2^1 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & r_{m-1}^n \\ r_m^1 & \cdots & \cdots & r_m^{n-1} & 1 \end{pmatrix} \; ; \; r_i^j = r_{ij} = \rho_{X^i X^j} = \frac{\mathrm{cov}(X^i, X^j)}{s_i \times s_j}, \; \forall i, j .$$

Si : $i = j,\ r_i^i = \dfrac{\mathrm{cov}(X^i, X^i)}{\mathrm{cov}(X^i, X^i)} = 1$.

Remarque

N'oubliez pas que si : $i = j$, $\mathrm{cov}(X^i, X^i) = \mathrm{var}\,(X^i)$ et $s_i = \sqrt{\mathrm{var}\,(X^i)}$.

On note $s_i = \sqrt{\mathrm{var}\,(X^i)}$ ou bien $\sigma_i = \sqrt{\mathrm{var}\,(X^i)}$.

Question Pourquoi réduire ?.

Comme le calcul des composantes principales se fait en fonction des variables, alors les variables d'écarts types importants auront plus de poids que des variables d'écarts types faibles.

Revenons maintenant à notre matrice de corrélation. R n'est que la matrice de variances-covariances des données centrées-réduites c'est-à-dire $R = \dfrac{1}{m}({}^tZ \cdot Z)$. Cette matrice nous permet de résumer la structure des dépendances linéaires entre les n variables. Autrement dit, elle reflète les relations entre les variables.

Ainsi, après avoir centré et réduit les données, la procédure à suivre pour réaliser une analyse en composantes principales est la suivante :

1- Centrer et normer (réduire) les variables : $X \rightarrow Z$ (matrice centrée-réduite).

2- Déterminer la matrice de corrélation R.

3- Déterminer les valeurs propres les plus grandes de la matrice R : $\lambda_1, \lambda_2, ..., \lambda_q$ où q n'est que la dimension du sous espace factoriel retenu, c'est à dire le nombre de nouvelles variables i.e les composantes principales.

4- Déterminer les vecteurs propres $F_1, F_2, ..., F_q$ associés aux valeurs propres $\lambda_1, \lambda_2, ..., \lambda_q$ .Ces vecteurs propres correspondent aux axes principaux (factoriels) de l'ACP. Ces axes sont orthogonaux deux à deux et ils ne sont pas corrélés.

USTHB - Faculté d'Informatique **06 Mars 2022**
Département d'IA et SD

M1 : MIV
Enseignant : N. Laiche

Analyse en composantes principales

Remarque Dans le cas où le nuage de points est simplement centré mais non réduit, l'ACP est dite canonique ou simple et dans ce cas tout revient à diagonaliser la matrice de covariance i.e tout revient à faire l'étude sur la matrice V .

Exemple 1 Soit la matrice de données suivante : $X = \begin{pmatrix} 3 & 1 \\ 1 & 5 \end{pmatrix}$.

X est la matrice d'observations de deux variables sur deux individus. Les variables sont

$X^1 = {}^t(3,1)$ et $X^2 = {}^t(1,5)$, et donc le centre de gravité est donné par $g = {}^t(\overline{X}^1, \overline{X}^2)$ i.e

$g = {}^t\left(\frac{3+1}{2}, \frac{1+5}{2}\right) = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$. Ceci nous conduit vers la matrice centrée Y définie par :

$Y = \begin{pmatrix} 3-2 & 1-3 \\ 1-2 & 5-3 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ -1 & 2 \end{pmatrix}$, et la matrice centrée réduite Z est définie par :

$Z = \begin{pmatrix} 1/\sigma_1 & -2/\sigma_2 \\ -1/\sigma_1 & 2/\sigma_2 \end{pmatrix}$ où σ_1, σ_2 sont les écarts types de X^1, X^2 respectivement.

$\sigma_1 = \sqrt{\operatorname{var}(Y^1)} = \sqrt{\frac{1}{2}\sum_{i=1}^{2}(y_i^1)^2} = \sqrt{\frac{1}{2}(1+1)} = 1$ et $\sigma_2 = \sqrt{\operatorname{var}(Y^2)} = \sqrt{\frac{1}{2}\sum_{i=1}^{2}(y_i^2)^2} = \sqrt{\frac{1}{2}(4+4)} = 2$,

alors : $Z = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Nous pouvons remarquer que $\operatorname{var}(Z^1) = \operatorname{var}(Z^2) = 1$.

Par suite, la matrice d'informations des variables qu'est la matrice de corrélation est donnée par $R = \frac{1}{2}({}^tZ \cdot Z)$: $R = \frac{1}{2}\begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$.

De cette matrice, nous avons : $r_1^2 = r_2^2 = -1$ ce qui veut dire que Y^1 et Y^2 sont colinéaires et de sens contraire. i.e l'angle formé entre les deux vecteurs est égale à π .

Analyse en composantes principales

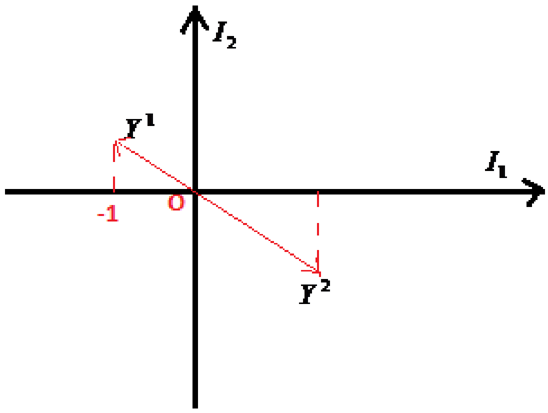

Figure 4. Représentation graphique des variables centrées.

Exemple 2

Soit le tableau de données suivant : $X = \begin{pmatrix} 4 & 5 \\ 6 & 7 \\ 8 & 0 \end{pmatrix}$.

A traiter prochainement in chaa Allah.