

Exemple Soit la matrice des données

$$X = \begin{pmatrix} 1 & 6 \\ 3 & 2 \end{pmatrix}.$$

Traité en cours.

4. Analyse d'un nuage de points

Puisque notre but est de **réduire la dimension** de l'espace des données,

L'analyse des données en composantes principales consiste à **étudier les projections** des points du nuage sur :

- i) un **axe**, dimension = 1
- ii) un **plan**, dimension = 2
- iii) un **hyperplan**, de dimension = 3

i) Cas d'une droite (un axe)

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

Alors la projection de X_i sur (Δu) notée x_i est présentée par :

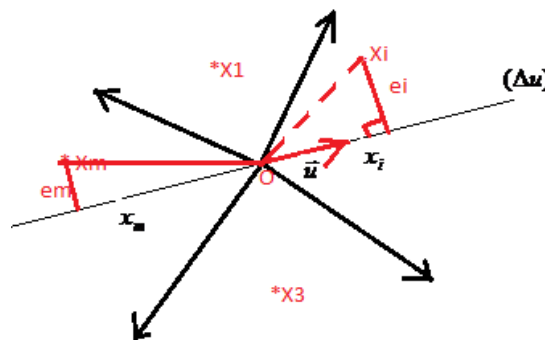


Illustration des projections des individus sur l'axe (Δu) .

Elle est définie par :

$$x_i = X_i \times \vec{u}$$

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**, nous avons :

$$x_i^2 + e_i^2 = \|X_i\|^2.$$

Ce qui donne

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

Donc, ça peut être interprété comme suit :

x_i **Représente l'information projetée sur la l'axe i.e la droite**

e_i **n'est que l'erreur ou bien l'information perdue.**

Pour réduire la dimension de l'espace, il faut donc :

maximiser l'information x_i , et **minimiser** la perte d'information e_i .

pour chaque individu X_i , $i = 1, 2, \dots, m$.

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

$$\text{maximiser : } \sum_{i=1}^m x_i^2.$$

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 :

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

D'où tout revient à maximiser :

$$\vec{u}^t \times (X^t \cdot X) \times \vec{u} \quad \text{sous la contrainte : } \|\vec{u}\| = 1.$$

C'est-à-dire :

$$\text{Maximiser : } \begin{cases} {}^t \vec{u} \cdot (X^t \cdot X) \cdot \vec{u} \\ \|\vec{u}\|^2 = {}^t \vec{u} \cdot \vec{u} = 1 \end{cases} \quad (P).$$

Il s'agit d'un problème d'optimisation avec contraintes.

Pour que nous puissions résoudre ce problème, nous donnons le rappel suivant :

Rappels

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 :

$$L(X, \lambda) = f(x) + \lambda \cdot g(x).$$

Avec :

- X : les variables figurant dans la fonction à optimiser.
- $f(x)$: la fonction à optimiser.
- λ : 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 :

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

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

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

et la solution est donnée par :

$$Sol = \begin{cases} \frac{\partial L}{\partial \vec{u}} = 0 \\ \frac{\partial L}{\partial \lambda} = 0 \end{cases}.$$

C'est-à-dire

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

Ce qui est équivalent à :

$$({}^tX \cdot X) \cdot \vec{u} = \lambda \cdot \vec{u} \quad (2).$$

C'est-à-dire :

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

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

$$\begin{aligned} {}^t\vec{u} \cdot ({}^tX \cdot X) \cdot \vec{u} &= {}^t\vec{u} \cdot \lambda \cdot \vec{u} \\ &= \lambda \cdot ({}^t\vec{u} \cdot \vec{u}) = \lambda \cdot \|\vec{u}\|^2 = \lambda. \end{aligned}$$

C'est-à-dire

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

Donc le **maximum** de ${}^t\vec{u} \cdot ({}^tX \cdot X) \cdot \vec{u}$ sous la contrainte $\|\vec{u}\|^2 = 1$ correspond à la **plus grande valeur propre** de ${}^tX \cdot X$. Par conséquent la meilleure droite (Δu) (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) Cas d'un plan

Pour déterminer le meilleur plan de visualisation du 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 $\langle \vec{u}, \vec{v} \rangle = 0$.

Donc, maximiser ${}^t\vec{v} \cdot ({}^tX \cdot X) \cdot \vec{v}$ sous les contraintes ${}^t\vec{u} \cdot \vec{u} = 1$ et ${}^t\vec{v} \cdot \vec{u} = 0 = {}^t\vec{u} \cdot \vec{v}$.

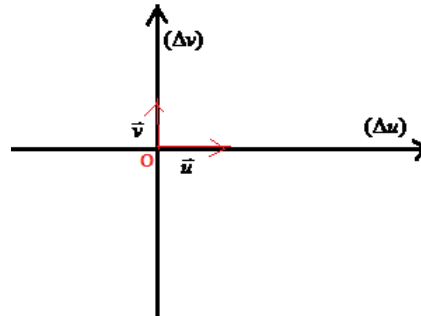


Illustration du plan ajustant.

Le Lagrangien cette fois ci s'écrit comme suit :

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

Où λ et μ sont les multiplicateurs de Lagrange.

La solution est donnée par :

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

c'est-à-dire

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

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

$$2{}^t\vec{u} \cdot ({}^tX \cdot X) \cdot \vec{v} - 2\lambda{}^t\vec{u} \cdot \vec{v} - \mu{}^t\vec{u} \cdot \vec{u} = 0.$$

Avec les contraintes imposées, nous obtenons :

$$2{}^t\vec{u} \cdot ({}^tX \cdot X) \cdot \vec{v} = \mu \quad (3)$$

Or, $({}^tX \cdot X) \cdot \vec{u} = \lambda \cdot \vec{u}$ c'est-à-dire ${}^t\vec{u} \cdot ({}^tX \cdot X) = \lambda \cdot {}^t\vec{u}$, alors l'équation (3) donne :

$$\mu = 2\lambda \cdot {}^t\vec{u} \cdot \vec{v} = 0.$$

Donc la solution est donnée par :

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

$$\Leftrightarrow ({}^tX \cdot X) \cdot \vec{v} = \lambda \cdot \vec{v}.$$

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 ${}^t\vec{v} \cdot ({}^tX \cdot X) \cdot \vec{v}$ correspond à la **deuxième grande valeur propre** de ${}^tX \cdot X$. Ainsi, le meilleur plan est constitué du 1^{er} axe factoriel (Δu) et le second axe factoriel (Δv).

iii) Cas d'un sous espace vectoriel

Les procédures précédentes peuvent être généralisées à 3, 4, ... axes factoriels.

6. Analyse en composantes principales

Le principe de l'ACP est de trouver des sous espaces de dimension $q \leq n$, issus de combinaisons linéaires des variables X^i , $i = 1, 2, \dots, n$ 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

Pour centrer le nuage de points, nous centrons les variables X^j , $j = 1, 2, \dots, n$

Ainsi, le tableau (matrice) X transformé 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 :

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

Ce qui donne

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

Remarque La matrice des variances-covariances des données centrées V est une matrice carrée de taille (n, n) . V est une matrice symétrique c'est-à-dire ${}^tV = V$.

D'autre part sur la diagonale, nous avons les $\text{var}(X^i)$, $i = 1, 2, \dots, n$ car pour : $i = j$,

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

ii) Centrage et Réduction des données

La question qui se pose est la suivante !!

Les données centrées réduites sont définies par :

$$z_i^j = \frac{y_i^j}{s_j} \text{ pour : } i = 1, 2, \dots, m \text{ et } j = 1, 2, \dots, n. \text{ Où } s_j \text{ est l'écart type de } X^j.$$

$$Z = \begin{pmatrix} z_1^1 & \dots & z_1^j & \dots & z_1^n \\ \vdots & & \vdots & & \vdots \\ z_i^1 & \dots & z_i^j & \dots & z_i^n \\ \vdots & & \vdots & & \vdots \\ z_m^1 & \dots & z_m^j & \dots & z_m^n \end{pmatrix}.$$

Cette matrice centrée réduite nous permet de définir la matrice de corrélation R :

$$R = \begin{pmatrix} 1 & r_1^2 & \dots & \dots & r_1^n \\ r_2^1 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & r_{m-1}^n \\ r_n^1 & \dots & \dots & r_n^{n-1} & 1 \end{pmatrix}.$$

Le terme général est défini par :

$$r_i^j = r_{ij} = \rho_{X^i X^j} = \frac{\text{cov}(X^i, X^j)}{s_i \times s_j}, \forall i, j.$$

Remarque La matrice de corrélation R n'est que la matrice de variances-covariances des données centrées-réduites c'est-à-dire $R = \frac{1}{m} ({}^t Z \cdot Z)$. Cette matrice nous permet de résumer la structure des dépendances linéaires entre les n 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 : $\lambda_1, \lambda_2, \dots, \lambda_q$ 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, \dots, F_q associés aux valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_q$. Ces vecteurs propres correspondent aux **axes principaux** (**factoriels**) de l'ACP. Ces axes sont orthogonaux deux à deux et ils ne sont pas corrélés.

Exemple 1 Soit la matrice de données suivante : $X = \begin{pmatrix} 4 & 5 \\ 6 & 7 \\ 8 & 0 \end{pmatrix}.$

Traité en détails en cours.