Analyse des données

Chapitre 2

Analyse en Composantes Principales

5. Analyse d'un nuage de points

L'analyse se base essentiellement sur la notion d'inertie.

5.1 Ajustement du nuage des individus

Considérons le nuage des individus défini par :

$$\mathcal{N}(I) = \left\{ \left(X_i, \frac{1}{m} \right) ; X_i \in \Re^n, i = 1, 2, \ldots, m \right\}.$$

Il s'agit de déterminer un meilleur ajustement du nuage des individus par un sous espace de dimension q strictement inférieure à n.

Si : $\boldsymbol{q} = \mathbf{1}$, il s'agit d'ajuster par **un axe** c'est à dire une droite.
Si : $\boldsymbol{q} = \mathbf{2}$, il s'agit d'ajuster par **un plan**.
Si : $\boldsymbol{q} = \mathbf{3}$, il s'agit de déterminer **un hyperplan**.
Sinon, **un sous espace** de dimension $\boldsymbol{q} <<<< \boldsymbol{n}$.

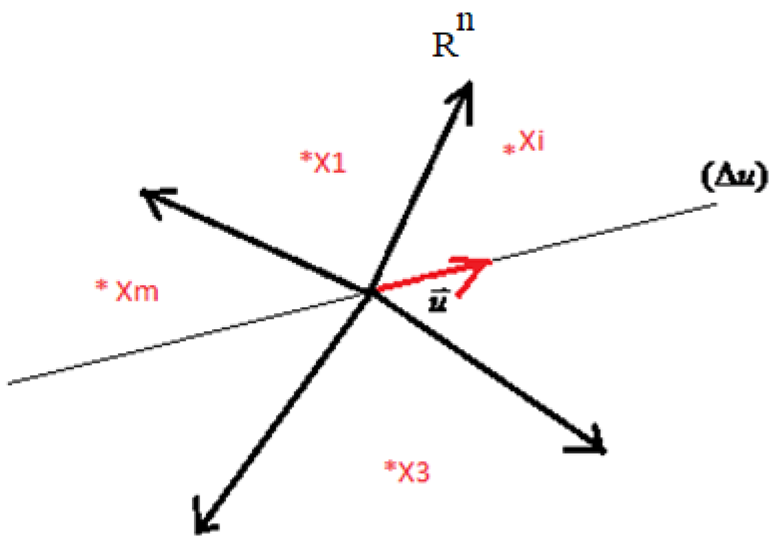
Pour réaliser cet ajustement, il suffit de faire de la **projection**.

Autrement dit :

Pour déterminer ce sous espace de représentation il faut **projeter** tous les points (**individus**) sur cet **sous espace**.

i) Ajustement du nuage par une droite

Il s'agit de remplacer l'espace $\Re^n$ par un **seul axe**. Donc, tout revient à déterminer une droite ($\boldsymbol{\Delta u}$) passant par **l'origine** et engendrée par le vecteur unitaire $\vec{\boldsymbol{u}}$ i.e $\|\vec{\boldsymbol{u}}\| = \mathbf{1}$.



Présentation de l'axe de projection et des individus dans l'espace $\Re^n$.

Analyse des données

La **projection** d'un individu quelconque $\boldsymbol{X_i}$ sur la droite (Δu) est présentée par la figure suivante :

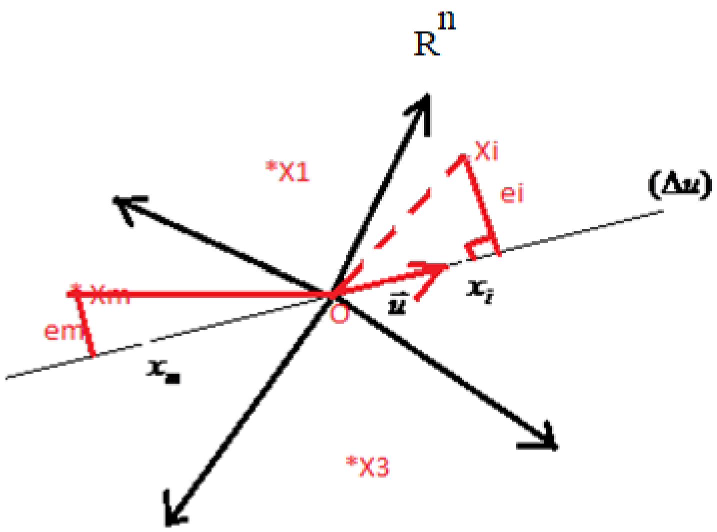


Illustration des projections des individus sur l'axe (Δu).

Si on note par $\boldsymbol{x_i}$ la grandeur algébrique de cette projection, nous avons :

$$\boldsymbol{x_i = X_i \,.\, \vec{u}.}$$

Bien-sûr, ou encore :

$$\boldsymbol{x_i = X_i \,.\, \vec{u}^t}$$

Ceci n'est que le produit scalaire suivant :

$$\langle X_i, \vec{u} \rangle = \langle \vec{u}, X_i \rangle.$$

De plus, de la figure précédente, nous avons l'égalité suivante :

$$\boldsymbol{x_i^2 + e_i^2 = \|X_i\|^2.}$$

Ce qui donne :

$$\boldsymbol{x_i^2 = \|X_i\|^2 - e_i^2.}$$

Interprétation :

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.

Par conséquent, pour **réduire la dimension** de l'espace initial, il faut pour chaque individu X_i, $i = 1, \dots, m$:

maximiser l'information x_i et minimiser la perte d'information $\mathrm{e_i}$.

Donc, pour avoir l'information totale maximale, il suffit de :

Analyse des données

- ***Projeter*** tous les individus sur l'axe (Δu).
- ***Déterminer*** $\vec{u}$ tel que la somme des carrées de ces projections soit maximale c'est-à-dire ***maximiser*** $\sum_{i=1}^{m} x_i^2$.

Or :

$$x_i^2 = \langle \overrightarrow{0X_i}, \vec{u} \rangle^2.$$

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

Alors :

$$\sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} (X_i\vec{u})^t \times (X_i\vec{u}) = \sum_{i=1}^{m} \vec{u}^t \cdot X_i^t \cdot X_i \cdot \vec{u}.$$

Ce qui donne :

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

D'où tout revient à :

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

Ce problème se présente sous la forme :

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

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 :

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

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.

Alors :

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

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

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

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

La solution est donnée par :

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

En s'appuyant sur :

$$***** \quad \frac{\partial (X^t \cdot M \cdot X)}{\partial X} = MX + M^t.X = (M + M^t)X.*****$$

Où, X est un vecteur et M est une matrice.

Nous avons :

$$\frac{\partial}{\partial \vec{u}}(\vec{\boldsymbol{u}}^{\boldsymbol{t}} \times (\boldsymbol{X^t} \cdot \boldsymbol{X}) \times \vec{\boldsymbol{u}}) = \boldsymbol{2 \cdot (X^t.X) \cdot \vec{u}}.$$

Il suffit de remplacer : M par la matrice $X^t.X$ et comme $X^t.X$ est symétrique, alors : $M + M^t = 2(X^t.X)$.

D'où,

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

L'équation (1) est équivalente à :

$$\boldsymbol{(X^t.X) \cdot \vec{u} = \lambda \cdot \vec{u}.} \qquad (2).$$

C'est-à-dire :

$\boldsymbol{\vec{u}}$ est un **vecteur propre** de la matrice $\boldsymbol{X^t.X}$ associé à la **valeur propre** $\boldsymbol{\lambda}$.

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

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

C'est-à-dire :

$$\boldsymbol{\vec{u}^t \times (X^t \cdot X) \times \vec{u}} \ = \lambda.$$

Donc,

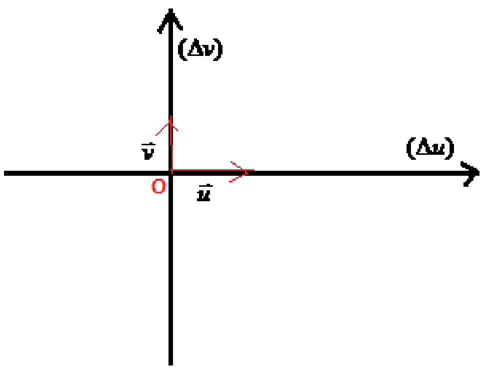
le **maximum** de $\vec{u}^t \times (X^t \cdot X) \times \vec{u}$ sous la **contrainte** $\|\vec{u}\|^2 = 1$

correspond à la **plus grande valeur propre** de $X^t \cdot X$.

> **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 $X^t \cdot X$.***

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 $\langle \vec{u}, \vec{v} \rangle = 0$.



Présentation du deuxième axe.

Donc,

tout revient à maximiser ${}^t\vec{v} \cdot ({}^tX \cdot X) \cdot \vec{v}$ sous les contraintes :

$${}^t\vec{v} \cdot \vec{v} = 1 \quad \text{et} \quad {}^t\vec{v} \cdot \vec{u} = 0 = {}^t\vec{u} \cdot \vec{v}.$$

Ainsi,

le Lagrangien cette fois ci s'écrit comme suit :

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

Où λ et μ sont les multiplicateurs de Lagrange.

La solution est donnée par :

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

Analyse des données

C'est-à-dire :

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

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

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

C'est-à-dire :

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

Or :

$(X^t \cdot X) \cdot \vec{u} = \lambda\vec{u}$ c'est-à-dire $\vec{u}^t \cdot (X^t \cdot X) = \lambda \cdot \vec{u}^t$, alors :

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

Donc,

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

.

C'est-à-dire :
$\vec{v}$ est un vecteur propre de la matrice $X^t \cdot X$ associé à la valeur propre λ.
Donc,
le ***maximum*** de $\vec{v}^t \cdot (X^t \cdot X) \cdot \vec{v}$ correspond à la ***deuxième grande valeur propre*** de $X^t \cdot X$.
Ainsi,

le ***meilleur plan*** est constitué du :

1er 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.