Faculté d'Informatique
Département d'IA et SD
M1/ MIV
Module : Analyse des données
Enseignant : N. Laiche

1 Mars 2022

Analyse en Composantes Principales

4. Mesures de liaison

4.1 Rappels

4.2 Mesures de liaison

1) ***Pour les individus***

Soient X_i et X_j deux individus quelconques, donc deux vecteurs de $\Re^n$

On utilise ***la distance Euclidienne,*** elle est définie par :

$$d\left(X_i,\, X_j\right) = \left(\sum_{k=1}^{n}(x_i^k - x_j^k)^2\right)^{1/2}.$$

2) ***Pour les variables***

Soient X et Y deux variables de $\Re^m$, alors nous appliquons les deux mesures suivantes :

- ***Covariance :*** Elle permet de préciser le sens de la liaison entre les variables,

$$\operatorname{cov}(X,Y) = \frac{1}{m} \times \sum_{j=1}^{m}\left(x_j - \overline{X}\right)\times\left(y_j - \overline{Y}\right).$$

Remarque La covariance n'est que $\left\langle \widetilde{X}, \widetilde{Y} \right\rangle_M$ pour $\widetilde{X}$ et $\widetilde{Y}$ les deux Variables centrées de X et Y respectivement et M est la matrice des poids.

- ***Coefficient de corrélation :*** Ce n'est qu'une normalisation de la covariance Il mesure l'intensité de la liaison linéaire entre les variables,

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)}\cdot\sqrt{\operatorname{var}(Y)}}.$$

Remarque

Le coefficient de corrélation ***prend ses valeurs entre -1 et 1.***

Le coefficient de corrélation mesure le ***cosinus de l'angle inscrit*** entre les variables. En effet,

$$\frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)}\cdot\sqrt{\operatorname{var}(Y)}} = \frac{\left\langle \widetilde{X}, \widetilde{Y} \right\rangle_M}{\left\|\widetilde{X}\right\|_M \cdot \left\|\widetilde{Y}\right\|_M} = COS\,(\theta_{\widetilde{X},\widetilde{Y}}).$$

D'où,

$$\rho_{X,Y} = Cos(X,Y).$$

Définition

Deux variables sont dites :

- ***décorrélées*** si leur coefficient est nul, et
- ***linéairement liées*** si leur coefficient de corrélation est de ***module égal à 1***.

Exemple Soit la matrice des données

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

Alors : les variables sont $X^1 = {}^t(1,3)$ **et** $X^2 = {}^t(6,2)$.

Comme :

$$\text{cov}(X,Y) = \frac{1}{m} \times \sum_{j=1}^{m} \left(x_j - \overline{X}\right) \times \left(y_j - \overline{Y}\right), \quad \rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)} \cdot \sqrt{\text{var}(Y)}}$$

Alors :

- Le ***centre de gravité*** du nuage est donné par :

$$g = (\overline{X}^1, \overline{X}^2) = \left(\frac{1+3}{2}, \frac{6+2}{2}\right) = (2,4).$$

La matrice ***centrée*** est la matrice :

$$Y = \begin{pmatrix} 1-2 & 6-4 \\ 3-2 & 2-4 \end{pmatrix}, \quad Y = \begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix}.$$

- Les ***variances*** sont données par :

$$Var(X^1) = \frac{1}{2} \cdot \sum_{j=1}^{2} \left(x_j^k - \overline{X^1}\right)^2 = (½)*(1+1) = 1$$

$$Var(X^2) = \frac{1}{2} \cdot \sum_{j=1}^{2} \left(x_j^k - \overline{X^2}\right)^2 = (½)*(4+4) = 4.$$

D'où,

$$\text{cov}(X^1, X^2) = \frac{1}{2} \times (-2-2) = -2 < 0$$

(Produit scalaire relativement à la matrice des poids)

Sens de la liaison entre les deux variables

Donc,

Les deux variables sont de sens contraire.

$$\rho_{X^1,X^2} = \frac{\mathrm{cov}(X^1,X^2)}{\sigma_{X^1}\cdot\sigma_{X^2}} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}\,(X)}\cdot\sqrt{\mathrm{var}\,(Y)}} = \frac{-2}{1*2} = -1$$

Donc,

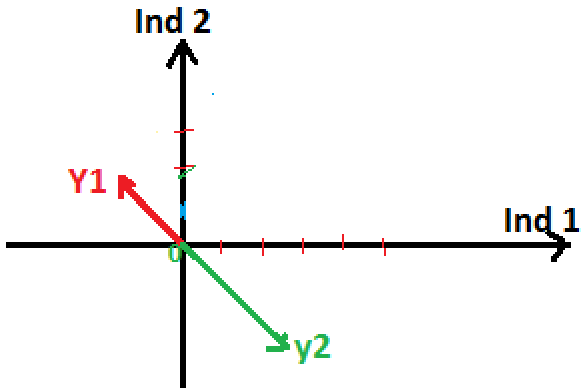
L'angle entre les deux variables est égal à pi.

C'est à dire

Les deux variables sont colinéaires mais de sens contraire

Remarque

Ces résultats peuvent être interprétés graphiquement :



Presentation des variables dans l'espace des individus.

5. Ana lyse des points

Puisque notre but est de ***réduire la dimension*** de l'espace des données,
Ça ne peut être que par ***projections*** !
Donc,

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

Ce qui revient à déterminer le ***meilleur ajustement du nuage*** par

un sous espace de $\Re^n$ ***pour*** $n \geq 1$.

Faculté d'Informatique **1 Mars 2022**
Département d'IA et SD
M1/ MIV
Module : Analyse des données
Enseignant : N. Laiche

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 $\|\vec{u}\| = 1$

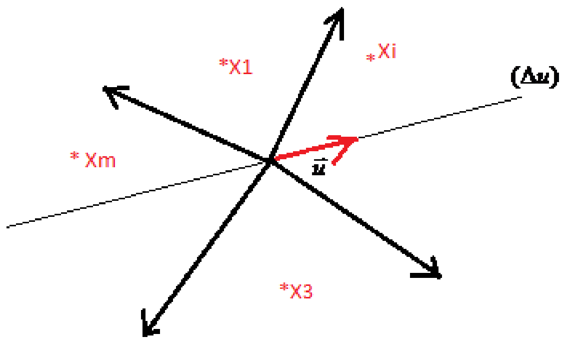


Figure 1. Présentation de l'axe de projection et des individus dans l'espace $\Re^n$.

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

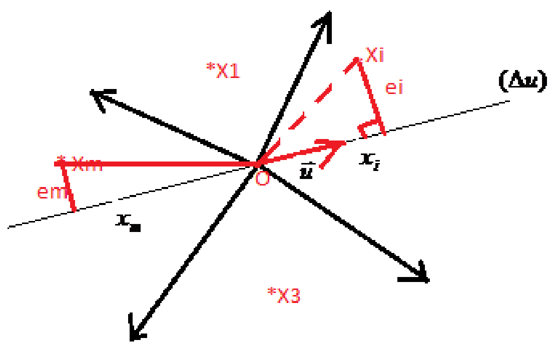


Figure 2. Illustration des projections des individus sur l'axe (Δu) .

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*** (voir figure 2), 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***, il suffit de minimiser la perte d'information*** e_i**.**

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

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

D'où tout revient à maximiser : ${}^t\vec{u}\times({}^tX\cdot X)\times\vec{u}$ sous la contrainte : $\|\vec{u}\| = 1$.

Que nous pouvons présenter comme suit :

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 :

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 :

Faculté d'Informatique ***1 Mars 2022***
Département d'IA et SD
M1/ MIV
Module : Analyse des données
Enseignant : N. Laiche

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

Avec :

$-X$: les variables figurant dans la fonction à optimiser.

$-f(x)$: la fonction à optimiser.

$-\lambda$: 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 (notons bien qu' il s'agit d'une condition nécessaire d'existence de solution) :

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