

Analyse de données

Introduction à l'Analyse en Composantes Principales

Didier Maquin

Ecole Nationale Supérieure d'Electricité et de Mécanique

Septembre 2018



1 Analyses factorielles

1 Analyses factorielles

2 Analyse générale

- Ajustement par un sous-espace vectoriel dans \mathbb{R}^p
- Ajustement par un sous-espace vectoriel dans \mathbb{R}^n
- Lien entre les sous-espaces de \mathbb{R}^p et de \mathbb{R}^n
- Reconstruction complète et partielle de la matrice de départ

1 Analyses factorielles

2 Analyse générale

- Ajustement par un sous-espace vectoriel dans \mathbb{R}^p
- Ajustement par un sous-espace vectoriel dans \mathbb{R}^n
- Lien entre les sous-espaces de \mathbb{R}^p et de \mathbb{R}^n
- Reconstruction complète et partielle de la matrice de départ

3 Analyse en composantes principales normée

- Usage et description

1 Analyses factorielles

2 Analyse générale

- Ajustement par un sous-espace vectoriel dans \mathbb{R}^p
- Ajustement par un sous-espace vectoriel dans \mathbb{R}^n
- Lien entre les sous-espaces de \mathbb{R}^p et de \mathbb{R}^n
- Reconstruction complète et partielle de la matrice de départ

3 Analyse en composantes principales normée

- Usage et description

4 Elaboration d'un "modèle ACP"

- Choix du nombre d'axes
- Modèles ACP

1 Analyses factorielles

2 Analyse générale

- Ajustement par un sous-espace vectoriel dans \mathbb{R}^p
- Ajustement par un sous-espace vectoriel dans \mathbb{R}^n
- Lien entre les sous-espaces de \mathbb{R}^p et de \mathbb{R}^n
- Reconstruction complète et partielle de la matrice de départ

3 Analyse en composantes principales normée

- Usage et description

4 Elaboration d'un "modèle ACP"

- Choix du nombre d'axes
- Modèles ACP

Définition

Définition

- visent à établir des **représentations synthétiques** de vastes ensembles de valeurs numériques ;

Définition

- visent à établir des **représentations synthétiques** de vastes ensembles de valeurs numériques ;
- cherchent à établir un **résumé descriptif** de ces données qui soit **appréhendable** et **exploitable** par l'analyste tout en entraînant une perte d'information minimale ;

Définition

- visent à établir des **représentations synthétiques** de vastes ensembles de valeurs numériques ;
- cherchent à établir un **résumé descriptif** de ces données qui soit **appréhendable** et **exploitable** par l'analyste tout en entraînant une perte d'information minimale ;
- on **consent une perte en information afin d'obtenir un gain en signification**.

Définition

- visent à établir des **représentations synthétiques** de vastes ensembles de valeurs numériques ;
- cherchent à établir un **résumé descriptif** de ces données qui soit **appréhensible** et **exploitable** par l'analyste tout en entraînant une perte d'information minimale ;
- on **consent une perte en information afin d'obtenir un gain en signification**.

Description

Etape de réduction de dimension. Pour un nuage de points, dont chacun est muni d'une masse, dans un espace vectoriel sur lequel est défini une métrique :

Définition

- visent à établir des **représentations synthétiques** de vastes ensembles de valeurs numériques ;
- cherchent à établir un **résumé descriptif** de ces données qui soit **appréhensible** et **exploitable** par l'analyste tout en entraînant une perte d'information minimale ;
- on **consent une perte en information afin d'obtenir un gain en signification**.

Description

Etape de réduction de dimension. Pour un nuage de points, dont chacun est muni d'une masse, dans un espace vectoriel sur lequel est défini une métrique :

- calculer l'inertie totale de ce nuage ;

Définition

- visent à établir des **représentations synthétiques** de vastes ensembles de valeurs numériques ;
- cherchent à établir un **résumé descriptif** de ces données qui soit **appréhensible** et **exploitable** par l'analyste tout en entraînant une perte d'information minimale ;
- on **consent une perte en information afin d'obtenir un gain en signification**.

Description

Etape de réduction de dimension. Pour un nuage de points, dont chacun est muni d'une masse, dans un espace vectoriel sur lequel est défini une métrique :

- calculer l'inertie totale de ce nuage ;
- déterminer ses axes d'inertie ;

Définition

- visent à établir des **représentations synthétiques** de vastes ensembles de valeurs numériques ;
- cherchent à établir un **résumé descriptif** de ces données qui soit **appréhensible** et **exploitable** par l'analyste tout en entraînant une perte d'information minimale ;
- on **consent une perte en information afin d'obtenir un gain en signification**.

Description

Etape de réduction de dimension. Pour un nuage de points, dont chacun est muni d'une masse, dans un espace vectoriel sur lequel est défini une métrique :

- calculer l'inertie totale de ce nuage ;
- déterminer ses axes d'inertie ;
- repérer les points dans la base formée par ces axes d'inertie.

Définition

- visent à établir des **représentations synthétiques** de vastes ensembles de valeurs numériques ;
- cherchent à établir un **résumé descriptif** de ces données qui soit **appréhensible** et **exploitable** par l'analyste tout en entraînant une perte d'information minimale ;
- on **consent une perte en information afin d'obtenir un gain en signification**.

Description

Etape de réduction de dimension. Pour un nuage de points, dont chacun est muni d'une masse, dans un espace vectoriel sur lequel est défini une métrique :

- calculer l'inertie totale de ce nuage ;
 - déterminer ses axes d'inertie ;
 - repérer les points dans la base formée par ces axes d'inertie.
-
- **techniques descriptives permettant d'identifier les dépendances entre des observations multivariées afin d'obtenir une représentation compacte de celles-ci**

Définition

- visent à établir des **représentations synthétiques** de vastes ensembles de valeurs numériques ;
- cherchent à établir un **résumé descriptif** de ces données qui soit **appréhensible** et **exploitable** par l'analyste tout en entraînant une perte d'information minimale ;
- on **consent une perte en information afin d'obtenir un gain en signification**.

Description

Etape de réduction de dimension. Pour un nuage de points, dont chacun est muni d'une masse, dans un espace vectoriel sur lequel est défini une métrique :

- calculer l'inertie totale de ce nuage ;
 - déterminer ses axes d'inertie ;
 - repérer les points dans la base formée par ces axes d'inertie.
-
- **techniques descriptives** permettant d'identifier les **dépendances** entre des observations multivariées afin d'obtenir une **représentation compacte** de celles-ci
 - **outil de modélisation** permettant ainsi d'estimer les variables ou les paramètres du processus sur lequel les données ont été acquises

1 Analyses factorielles

2 Analyse générale

- Ajustement par un sous-espace vectoriel dans \mathbb{R}^p
- Ajustement par un sous-espace vectoriel dans \mathbb{R}^n
- Lien entre les sous-espaces de \mathbb{R}^p et de \mathbb{R}^n
- Reconstruction complète et partielle de la matrice de départ

3 Analyse en composantes principales normée

- Usage et description

4 Elaboration d'un "modèle ACP"

- Choix du nombre d'axes
- Modèles ACP

Problème posé

Etant donné un tableau rectangulaire de valeurs numériques représenté par une matrice X à n lignes et p colonnes, de terme général x_{ij} , **est-il possible de reconstituer les np valeurs x_{ij} à partir d'un plus petit nombre de valeurs numériques ?**

Problème posé

Etant donné un tableau rectangulaire de valeurs numériques représenté par une matrice X à n lignes et p colonnes, de terme général x_{ij} , **est-il possible de reconstituer les np valeurs x_{ij} à partir d'un plus petit nombre de valeurs numériques ?**

Analyses factorielles

Problème posé

Etant donné un tableau rectangulaire de valeurs numériques représenté par une matrice X à n lignes et p colonnes, de terme général x_{ij} , **est-il possible de reconstituer les np valeurs x_{ij} à partir d'un plus petit nombre de valeurs numériques ?**

Analyses factorielles

- Les analyses factorielles projettent les points sur une droite, un plan, ..., un sous-espace à q dimension (avec $q \leq p$) choisi de façon à optimiser un certain critère.

Problème posé

Etant donné un tableau rectangulaire de valeurs numériques représenté par une matrice X à n lignes et p colonnes, de terme général x_{ij} , **est-il possible de reconstituer les np valeurs x_{ij} à partir d'un plus petit nombre de valeurs numériques ?**

Analyses factorielles

- Les analyses factorielles projettent les points sur une droite, un plan, ..., un sous-espace à q dimension (avec $q \leq p$) choisi de façon à optimiser un certain critère.
- Intuitivement, on cherchera le sous-espace donnant la **meilleure visualisation** possible du nuage de points.

Problème posé

Etant donné un tableau rectangulaire de valeurs numériques représenté par une matrice X à n lignes et p colonnes, de terme général x_{ij} , **est-il possible de reconstituer les np valeurs x_{ij} à partir d'un plus petit nombre de valeurs numériques ?**

Analyses factorielles

- Les analyses factorielles projettent les points sur une droite, un plan, ..., un sous-espace à q dimension (avec $q \leq p$) choisi de façon à optimiser un certain critère.
- Intuitivement, on cherchera le sous-espace donnant la **meilleure visualisation** possible du nuage de points.
- Un bon choix consiste à rechercher la **plus grande dispersion** (le plus grand étalement) possible des projections dans le sous-espace choisi.

Approche intuitive

- Supposons qu'il existe un vecteur u_1 à n composantes et un vecteur v_1 à p composantes tels que $X = u_1 v_1^T$.

Approche intuitive

- Supposons qu'il existe un vecteur u_1 à n composantes et un vecteur v_1 à p composantes tels que $X = u_1 v_1^T$.
- On aura alors reconstitué les np valeurs de X avec $n + p$ valeurs numériques seulement (dans ce cas, le **rang de la matrice X est égal à 1 !**).

Approche intuitive

- Supposons qu'il existe un vecteur u_1 à n composantes et un vecteur v_1 à p composantes tels que $X = u_1 v_1^T$.
- On aura alors reconstitué les np valeurs de X avec $n + p$ valeurs numériques seulement (dans ce cas, le **rang de la matrice X est égal à 1 !**).
- En pratique, on cherchera donc une **approximation de rang q** pour X :

$$X = u_1 v_1^T + u_2 v_2^T + \dots + u_q v_q^T + E$$

Approche intuitive

- Supposons qu'il existe un vecteur u_1 à n composantes et un vecteur v_1 à p composantes tels que $X = u_1 v_1^T$.
- On aura alors reconstitué les np valeurs de X avec $n + p$ valeurs numériques seulement (dans ce cas, le **rang de la matrice X est égal à 1 !**).
- En pratique, on cherchera donc une **approximation de rang q** pour X :

$$X = u_1 v_1^T + u_2 v_2^T + \dots + u_q v_q^T + E$$

- E étant une matrice ($n \times p$) résiduelle dont les termes sont **suffisamment petits** pour que l'on puisse considérer que les np valeurs x_{ij} sont reconstituées de façon satisfaisante par les $q(n + p)$ valeurs des vecteurs u_i et v_i , $i = 1, \dots, q$.

Approche intuitive

- Supposons qu'il existe un vecteur u_1 à n composantes et un vecteur v_1 à p composantes tels que $X = u_1 v_1^T$.
- On aura alors reconstitué les np valeurs de X avec $n + p$ valeurs numériques seulement (dans ce cas, le **rang de la matrice X est égal à 1 !**).
- En pratique, on cherchera donc une **approximation de rang q** pour X :

$$X = u_1 v_1^T + u_2 v_2^T + \dots + u_q v_q^T + E$$

- E étant une matrice ($n \times p$) résiduelle dont les termes sont **suffisamment petits** pour que l'on puisse considérer que les np valeurs x_{ij} sont reconstituées de façon satisfaisante par les $q(n + p)$ valeurs des vecteurs u_i et v_i , $i = 1, \dots, q$.

Solution s'appuyant sur des **représentations géométriques**

La matrice X donnera lieu à deux représentations :

Approche intuitive

- Supposons qu'il existe un vecteur u_1 à n composantes et un vecteur v_1 à p composantes tels que $X = u_1 v_1^T$.
- On aura alors reconstitué les np valeurs de X avec $n + p$ valeurs numériques seulement (dans ce cas, le **rang de la matrice X est égal à 1 !**).
- En pratique, on cherchera donc une **approximation de rang q** pour X :

$$X = u_1 v_1^T + u_2 v_2^T + \dots + u_q v_q^T + E$$

- E étant une matrice ($n \times p$) résiduelle dont les termes sont **suffisamment petits** pour que l'on puisse considérer que les np valeurs x_{ij} sont reconstituées de façon satisfaisante par les $q(n + p)$ valeurs des vecteurs u_i et v_i , $i = 1, \dots, q$.

Solution s'appuyant sur des **représentations géométriques**

La matrice X donnera lieu à deux représentations :

- les n lignes de X peuvent être considérées comme les coordonnées de n points dans un espace à p dimensions \mathbb{R}^p

Approche intuitive

- Supposons qu'il existe un vecteur u_1 à n composantes et un vecteur v_1 à p composantes tels que $X = u_1 v_1^T$.
- On aura alors reconstitué les np valeurs de X avec $n + p$ valeurs numériques seulement (dans ce cas, le **rang de la matrice X est égal à 1 !**).
- En pratique, on cherchera donc une **approximation de rang q** pour X :

$$X = u_1 v_1^T + u_2 v_2^T + \dots + u_q v_q^T + E$$

- E étant une matrice ($n \times p$) résiduelle dont les termes sont **suffisamment petits** pour que l'on puisse considérer que les np valeurs x_{ij} sont reconstituées de façon satisfaisante par les $q(n + p)$ valeurs des vecteurs u_i et v_i , $i = 1, \dots, q$.

Solution s'appuyant sur des **représentations géométriques**

La matrice X donnera lieu à deux représentations :

- les n lignes de X peuvent être considérées comme les coordonnées de n points dans un espace à p dimensions \mathbb{R}^p
- les p colonnes de X peuvent représenter les coordonnées de p points dans un espace à n dimensions.

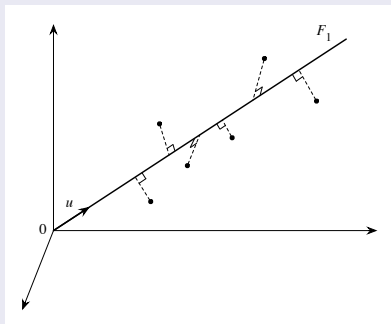
Cas idéal

Si le nuage des n points, représentant la matrice X dans cet espace, est contenu dans un sous-espace vectoriel de dimension q inférieure à p , il est alors possible de reconstituer les positions des n points (donc de reconstituer X) à partir des coordonnées sur q nouveaux axes et des composantes de ces nouveaux axes. On remplace ainsi np valeurs par $nq + pq$ autres valeurs.

Cas idéal

Si le nuage des n points, représentant la matrice X dans cet espace, est contenu dans un sous-espace vectoriel de dimension q inférieure à p , il est alors possible de reconstituer les positions des n points (donc de reconstituer X) à partir des coordonnées sur q nouveaux axes et des composantes de ces nouveaux axes. On remplace ainsi np valeurs par $nq + pq$ autres valeurs.

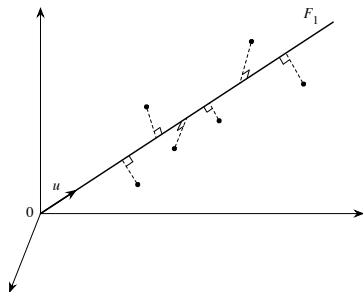
En pratique



Cas idéal

Si le nuage des n points, représentant la matrice X dans cet espace, est contenu dans un sous-espace vectoriel de dimension q inférieure à p , il est alors possible de reconstituer les positions des n points (donc de reconstituer X) à partir des coordonnées sur q nouveaux axes et des composantes de ces nouveaux axes. On remplace ainsi np valeurs par $nq + pq$ autres valeurs.

En pratique

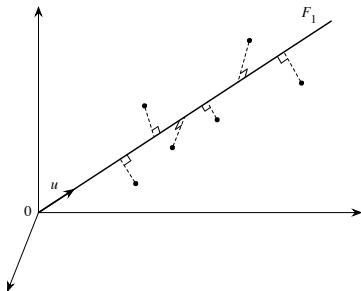


- On va donc chercher à ajuster le nuage des n points par un sous-espace vectoriel de \mathbb{R}^p , qui sera muni de la distance euclidienne usuelle.

Cas idéal

Si le nuage des n points, représentant la matrice X dans cet espace, est contenu dans un sous-espace vectoriel de dimension q inférieure à p , il est alors possible de reconstituer les positions des n points (donc de reconstituer X) à partir des coordonnées sur q nouveaux axes et des composantes de ces nouveaux axes. On remplace ainsi np valeurs par $nq + pq$ autres valeurs.

En pratique



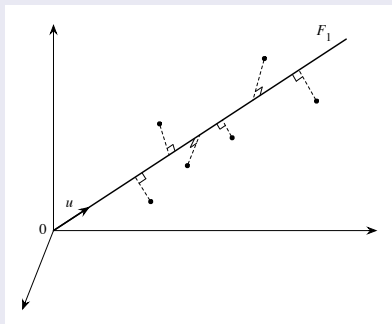
- On va donc chercher à ajuster le nuage des n points par un sous-espace vectoriel de \mathbb{R}^p , qui sera muni de la distance euclidienne usuelle.
- Commençons par chercher la droite F_1 **passant par l'origine**, qui ajuste au mieux le nuage.

Ajustement par un sous-espace vectoriel dans \mathbb{R}^p

Cas idéal

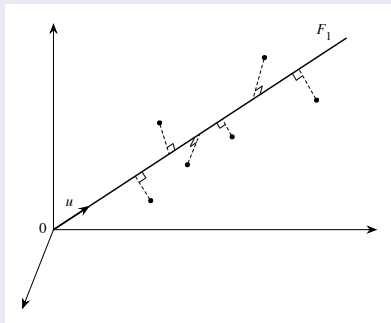
Si le nuage des n points, représentant la matrice X dans cet espace, est contenu dans un sous-espace vectoriel de dimension q inférieure à p , il est alors possible de reconstituer les positions des n points (donc de reconstituer X) à partir des coordonnées sur q nouveaux axes et des composantes de ces nouveaux axes. On remplace ainsi np valeurs par $nq + pq$ autres valeurs.

En pratique

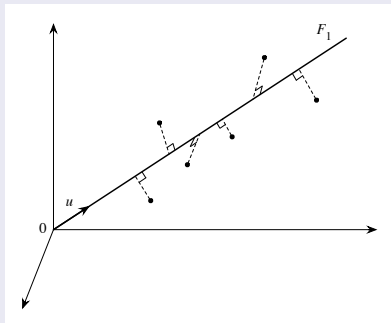


- On va donc chercher à ajuster le nuage des n points par un sous-espace vectoriel de \mathbb{R}^p , qui sera muni de la distance euclidienne usuelle.
- Commençons par chercher la droite F_1 **passant par l'origine**, qui ajuste au mieux le nuage.
- Soit u un vecteur unitaire porté par cette droite, c'est-à-dire tel que $u^T u = 1$ ou encore $\sum_{j=1}^p u_j^2 = 1$.

En pratique

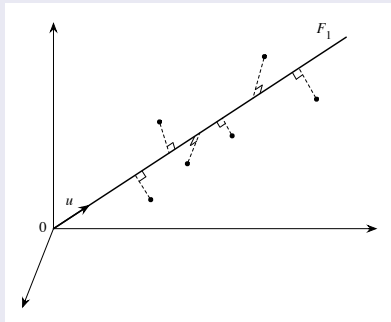


En pratique



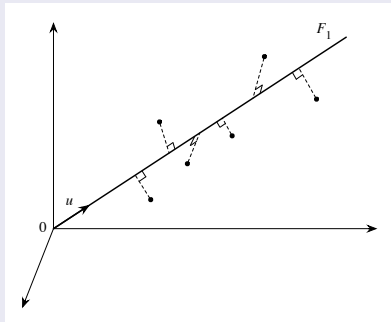
- Puisque chaque ligne de X représente un point de \mathbb{R}^p , les n lignes du vecteur Xu sont les n produits scalaires de ces points avec u et sont donc les longueurs des projections de ces n points sur F_1 .

En pratique



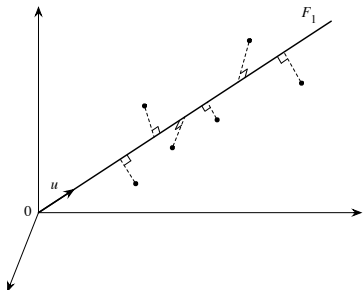
- Puisque chaque ligne de X représente un point de \mathbb{R}^p , les n lignes du vecteur Xu sont les n produits scalaires de ces points avec u et sont donc les longueurs des projections de ces n points sur F_1 .
- Pour chaque point, le carré de sa distance à l'origine se décompose en carré de sa projection sur F_1 et en carré de sa distance à F_1 .

En pratique



- Puisque chaque ligne de X représente un point de \mathbb{R}^p , les n lignes du vecteur Xu sont les n produits scalaires de ces points avec u et sont donc les longueurs des projections de ces n points sur F_1 .
- Pour chaque point, le carré de sa distance à l'origine se décompose en carré de sa projection sur F_1 et en carré de sa distance à F_1 .
- Les distances à l'origine étant données, il est équivalent de minimiser la somme des carrés des distances à F_1 ou de maximiser la somme des carrés des projections sur F_1 .

En pratique



- Puisque chaque ligne de X représente un point de \mathbb{R}^p , les n lignes du vecteur Xu sont les n produits scalaires de ces points avec u et sont donc les longueurs des projections de ces n points sur F_1 .
- Pour chaque point, le carré de sa distance à l'origine se décompose en carré de sa projection sur F_1 et en carré de sa distance à F_1 .

- Les distances à l'origine étant données, il est équivalent de minimiser la somme des carrés des distances à F_1 ou de maximiser la somme des carrés des projections sur F_1 .
- Si l'on veut que la somme des carrés des projections soit maximale, il faut chercher u qui rende maximale la quantité :

$$\phi = (Xu)^T Xu = u^T X^T Xu$$

Remarque

- La maximisation de l'expression précédente correspond à celle de l'inertie du nuage de points expliquée par la direction portée par le vecteur u .

Remarque

- La maximisation de l'expression précédente correspond à celle de l'inertie du nuage de points expliquée par la direction portée par le vecteur u .
- En effet, par définition, l'inertie d'un nuage de points par rapport à un point c correspond à la somme des carrés des distances des différents points au point c .

Remarque

- La maximisation de l'expression précédente correspond à celle de l'inertie du nuage de points expliquée par la direction portée par le vecteur u .
- En effet, par définition, l'inertie d'un nuage de points par rapport à un point c correspond à la somme des carrés des distances des différents points au point c .
- Si la mesure est effectuée par rapport à l'origine ($c=0$) les distances à prendre en compte sont les longueurs des vecteurs définis par les différents points. Dans ce cas, l'inertie (par rapport à l'origine) expliquée par la direction portée par un vecteur u correspond à la somme des carrés des projections orthogonales des différents points sur la droite portée par u .

Remarque

- La maximisation de l'expression précédente correspond à celle de l'inertie du nuage de points expliquée par la direction portée par le vecteur u .
- En effet, par définition, l'inertie d'un nuage de points par rapport à un point c correspond à la somme des carrés des distances des différents points au point c .
- Si la mesure est effectuée par rapport à l'origine ($c=0$) les distances à prendre en compte sont les longueurs des vecteurs définis par les différents points. Dans ce cas, l'inertie (par rapport à l'origine) expliquée par la direction portée par un vecteur u correspond à la somme des carrés des projections orthogonales des différents points sur la droite portée par u .
- Comme indiqué précédemment, le vecteur Xu contient les longueurs des projections des points sur la droite portée par u . La somme des carrés des différentes longueurs s'écrit donc bien :

$$\|Xu\|^2 = u^T X^T Xu$$

Solution

- Ainsi, trouver un sous-espace vectoriel à une dimension qui ajuste au mieux, au sens des moindres carrés, le nuage de n points, revient à rendre maximale la forme quadratique $u^T X^T X u$ sous la contrainte $u^T u = 1$. On désignera ce sous-espace optimal par u_1 .

Solution

- Ainsi, trouver un sous-espace vectoriel à une dimension qui ajuste au mieux, au sens des moindres carrés, le nuage de n points, revient à rendre maximale la forme quadratique $u^T X^T X u$ sous la contrainte $u^T u = 1$. On désignera ce sous-espace optimal par u_1 .
- On montre aisément que le meilleur sous-espace vectoriel à deux dimensions contient u_1 . On le trouve en cherchant le vecteur u_2 unitaire et orthogonal à u_1 tel que $u_2^T u_1 = 0$ et $u_2^T u_2 = 1$) qui rend maximale la forme quadratique $u_2^T X^T X u_2$.

Solution

- Ainsi, trouver un sous-espace vectoriel à une dimension qui ajuste au mieux, au sens des moindres carrés, le nuage de n points, revient à rendre maximale la forme quadratique $u^T X^T X u$ sous la contrainte $u^T u = 1$. On désignera ce sous-espace optimal par u_1 .
- On montre aisément que le meilleur sous-espace vectoriel à deux dimensions contient u_1 . On le trouve en cherchant le vecteur u_2 unitaire et orthogonal à u_1 tel que $u_2^T u_1 = 0$ et $u_2^T u_2 = 1$) qui rend maximale la forme quadratique $u_2^T X^T X u_2$.
- De façon analogue, à l'aide d'un raisonnement par récurrence, on verrait que le meilleur sous-espace vectoriel à q dimensions ($q \leq p$) est engendré par les vecteurs u_1, u_2, \dots, u_q où u_q est orthogonal à u_1, u_2, \dots, u_{q-1} et rend maximale la forme quadratique $u_q^T X^T X u_q$ avec $u_q^T u_q = 1$.

Solution

- Ainsi, trouver un sous-espace vectoriel à une dimension qui ajuste au mieux, au sens des moindres carrés, le nuage de n points, revient à rendre maximale la forme quadratique $u^T X^T X u$ sous la contrainte $u^T u = 1$. On désignera ce sous-espace optimal par u_1 .
- On montre aisément que le meilleur sous-espace vectoriel à deux dimensions contient u_1 . On le trouve en cherchant le vecteur u_2 unitaire et orthogonal à u_1 tel que $u_2^T u_1 = 0$ et $u_2^T u_2 = 1$) qui rend maximale la forme quadratique $u_2^T X^T X u_2$.
- De façon analogue, à l'aide d'un raisonnement par récurrence, on verrait que le meilleur sous-espace vectoriel à q dimensions ($q \leq p$) est engendré par les vecteurs u_1, u_2, \dots, u_q où u_q est orthogonal à u_1, u_2, \dots, u_{q-1} et rend maximale la forme quadratique $u_q^T X^T X u_q$ avec $u_q^T u_q = 1$.
- L'obtention des différents vecteurs u_i est aisée. Pour u_1 , le problème d'optimisation à résoudre s'écrit :

$$\begin{cases} \max_u \phi = u^T X^T X u \\ \text{sous la contrainte } u^T u = 1 \end{cases}$$

Solution

- L'obtention des différents vecteurs u_i est aisée. Pour u_1 , le problème d'optimisation à résoudre s'écrit :

$$\begin{cases} \max_u \phi = u^T X^T X u \\ \text{sous la contrainte } u^T u = 1 \end{cases}$$

Solution

- L'obtention des différents vecteurs u_i est aisée. Pour u_1 , le problème d'optimisation à résoudre s'écrit :

$$\begin{cases} \max_u \phi = u^T X^T X u \\ \text{sous la contrainte } u^T u = 1 \end{cases}$$

- Le lagrangien associé à ce problème s'écrit :

$$\mathcal{L} = u^T X^T X u - \lambda(u^T u - 1)$$

où λ est un multiplicateur de Lagrange.

Solution

- L'obtention des différents vecteurs u_i est aisée. Pour u_1 , le problème d'optimisation à résoudre s'écrit :

$$\begin{cases} \max_u \phi = u^T X^T X u \\ \text{sous la contrainte } u^T u = 1 \end{cases}$$

- Le lagrangien associé à ce problème s'écrit :

$$\mathcal{L} = u^T X^T X u - \lambda(u^T u - 1)$$

où λ est un multiplicateur de Lagrange.

- La dérivation du lagrangien par rapport aux différentes composantes de u , puis l'annulation de ces dérivées conduisent à la relation matricielle :

$$2X^T X u - 2\lambda u = 0,$$

c'est-à-dire :

$$X^T X u = \lambda u.$$

Solution

- L'obtention des différents vecteurs u_i est aisée. Pour u_1 , le problème d'optimisation à résoudre s'écrit :

$$\begin{cases} \max_u \phi = u^T X^T X u \\ \text{sous la contrainte } u^T u = 1 \end{cases}$$

- Le lagrangien associé à ce problème s'écrit :

$$\mathcal{L} = u^T X^T X u - \lambda(u^T u - 1)$$

où λ est un multiplicateur de Lagrange.

- La dérivation du lagrangien par rapport aux différentes composantes de u , puis l'annulation de ces dérivées conduisent à la relation matricielle :

$$2X^T X u - 2\lambda u = 0,$$

c'est-à-dire :

$$X^T X u = \lambda u.$$

- Ceci montre que u est vecteur propre de la matrice $X^T X$. Remarquons également que $u^T X^T X u = \lambda u^T u = \lambda$.

Solution

- Le maximum cherché est donc une valeur propre de $X^T X$. Ainsi, u sera le vecteur propre u_1 correspondant à la **plus grande** valeur propre λ_1 de la matrice $X^T X$.

Solution

- Le maximum cherché est donc une valeur propre de $X^T X$. Ainsi, u sera le vecteur propre u_1 correspondant à la **plus grande** valeur propre λ_1 de la matrice $X^T X$.
- Notons encore que cette plus grande valeur propre λ_1 représente la part d'inertie expliquée par la droite F_1

Solution

- Le maximum cherché est donc une valeur propre de $X^T X$. Ainsi, u sera le vecteur propre u_1 correspondant à la **plus grande** valeur propre λ_1 de la matrice $X^T X$.
- Notons encore que cette plus grande valeur propre λ_1 représente la part d'inertie expliquée par la droite F_1
- Si l'on recherche l'espace à deux dimensions qui s'ajuste au mieux au nuage, nous devons trouver une deuxième droite, portée par le vecteur unitaire t , passant par l'origine et qui maximise $t^T X^T X t$ en étant orthogonal à u_1 (c'est-à-dire $t^T u_1 = 0$ et $t^T t = 1$).

Solution

- Le maximum cherché est donc une valeur propre de $X^T X$. Ainsi, u sera le vecteur propre u_1 correspondant à la **plus grande** valeur propre λ_1 de la matrice $X^T X$.
- Notons encore que cette plus grande valeur propre λ_1 représente la part d'inertie expliquée par la droite F_1
- Si l'on recherche l'espace à deux dimensions qui s'ajuste au mieux au nuage, nous devons trouver une deuxième droite, portée par le vecteur unitaire t , passant par l'origine et qui maximise $t^T X^T X t$ en étant orthogonal à u_1 (c'est-à-dire $t^T u_1 = 0$ et $t^T t = 1$).
- Le problème d'optimisation correspondant s'écrit :

$$\left\{ \begin{array}{l} \max_t \phi = t^T X^T X t \\ \text{sous les contraintes } t^T t = 1 \\ \phantom{\text{sous les contraintes }} t^T u_1 = 0 \end{array} \right.$$

Solution

- Le maximum cherché est donc une valeur propre de $X^T X$. Ainsi, u sera le vecteur propre u_1 correspondant à la **plus grande** valeur propre λ_1 de la matrice $X^T X$.
- Notons encore que cette plus grande valeur propre λ_1 représente la part d'inertie expliquée par la droite F_1
- Si l'on recherche l'espace à deux dimensions qui s'ajuste au mieux au nuage, nous devons trouver une deuxième droite, portée par le vecteur unitaire t , passant par l'origine et qui maximise $t^T X^T X t$ en étant orthogonal à u_1 (c'est-à-dire $t^T u_1 = 0$ et $t^T t = 1$).
- Le problème d'optimisation correspondant s'écrit :

$$\begin{cases} \max_t \phi = t^T X^T X t \\ \text{sous les contraintes } t^T t = 1 \\ t^T u_1 = 0 \end{cases}$$

- Le lagrangien associé au problème d'optimisation s'exprime sous la forme suivante :

$$\mathcal{L} = t^T X^T X t - \lambda(t^T t - 1) - \mu t^T u_1$$

où λ et μ sont les multiplicateurs de Lagrange associés aux contraintes.

Solution

- Le lagrangien associé au problème d'optimisation s'exprime sous la forme suivante :

$$\mathcal{L} = t^T X^T X t - \lambda(t^T t - 1) - \mu t^T u_1$$

où λ et μ sont les multiplicateurs de Lagrange associés aux contraintes.

Solution

- Le lagrangien associé au problème d'optimisation s'exprime sous la forme suivante :

$$\mathcal{L} = t^T X^T X t - \lambda(t^T t - 1) - \mu t^T u_1$$

où λ et μ sont les multiplicateurs de Lagrange associés aux contraintes.

- L'annulation de ses dérivées partielles par rapport aux composantes de t conduit à la relation matricielle :

$$2X^T X t - 2\lambda t - \mu u_1 = 0$$

Solution

- Le lagrangien associé au problème d'optimisation s'exprime sous la forme suivante :

$$\mathcal{L} = t^T X^T X t - \lambda(t^T t - 1) - \mu t^T u_1$$

où λ et μ sont les multiplicateurs de Lagrange associés aux contraintes.

- L'annulation de ses dérivées partielles par rapport aux composantes de t conduit à la relation matricielle :

$$2X^T X t - 2\lambda t - \mu u_1 = 0$$

- En prémultipliant cette relation par u_1^T et en remarquant que $u_1^T X^T X = \lambda_1 u_1^T$, on obtient :

$$\underbrace{2 u_1^T X^T X t}_{=\lambda_1 u_1^T t} - 2\lambda u_1^T t - \underbrace{\mu u_1^T u_1}_{=1} = 0$$

Solution

- Le lagrangien associé au problème d'optimisation s'exprime sous la forme suivante :

$$\mathcal{L} = t^T X^T X t - \lambda(t^T t - 1) - \mu t^T u_1$$

où λ et μ sont les multiplicateurs de Lagrange associés aux contraintes.

- L'annulation de ses dérivées partielles par rapport aux composantes de t conduit à la relation matricielle :

$$2X^T X t - 2\lambda t - \mu u_1 = 0$$

- En prémultipliant cette relation par u_1^T et en remarquant que $u_1^T X^T X = \lambda_1 u_1^T$, on obtient :

$$\underbrace{2 u_1^T X^T X t}_{= \lambda_1 u_1^T t} - 2\lambda u_1^T t - \mu \underbrace{u_1^T u_1}_{=1} = 0$$

- Comme $u_1^T t = 0$, il vient $\mu = 0$; on déduit donc de l'équation initiale que :

$$X^T X t = \lambda t$$

Solution

- Le lagrangien associé au problème d'optimisation s'exprime sous la forme suivante :

$$\mathcal{L} = t^T X^T X t - \lambda(t^T t - 1) - \mu t^T u_1$$

où λ et μ sont les multiplicateurs de Lagrange associés aux contraintes.

- L'annulation de ses dérivées partielles par rapport aux composantes de t conduit à la relation matricielle :

$$2X^T X t - 2\lambda t - \mu u_1 = 0$$

- En prémultipliant cette relation par u_1^T et en remarquant que $u_1^T X^T X = \lambda_1 u_1^T$, on obtient :

$$\underbrace{2 u_1^T X^T X t}_{= \lambda_1 u_1^T t} - 2\lambda u_1^T t - \underbrace{\mu u_1^T u_1}_{= 1} = 0$$

- Comme $u_1^T t = 0$, il vient $\mu = 0$; on déduit donc de l'équation initiale que :

$$X^T X t = \lambda t$$

- Ainsi t est le second vecteur propre associé à la seconde plus grande valeur propre de $X^T X$. Le résultat s'étend aux éléments propres $\alpha = 1, 2, \dots, r$ où r est le rang de $X^T X$.

Solution

- Le lagrangien associé au problème d'optimisation s'exprime sous la forme suivante :

$$\mathcal{L} = t^T X^T X t - \lambda(t^T t - 1) - \mu t^T u_1$$

où λ et μ sont les multiplicateurs de Lagrange associés aux contraintes.

- L'annulation de ses dérivées partielles par rapport aux composantes de t conduit à la relation matricielle :

$$2X^T X t - 2\lambda t - \mu u_1 = 0$$

- En prémultipliant cette relation par u_1^T et en remarquant que $u_1^T X^T X = \lambda_1 u_1^T$, on obtient :

$$\underbrace{2 u_1^T X^T X t}_{= \lambda_1 u_1^T t} - 2\lambda u_1^T t - \underbrace{\mu u_1^T u_1}_{= 1} = 0$$

- Comme $u_1^T t = 0$, il vient $\mu = 0$; on déduit donc de l'équation initiale que :

$$X^T X t = \lambda t$$

- Ainsi t est le second vecteur propre associé à la seconde plus grande valeur propre de $X^T X$. Le résultat s'étend aux éléments propres $\alpha = 1, 2, \dots, r$ où r est le rang de $X^T X$.

On s'est intéressé, jusqu'à maintenant, au nuage de n points dans l'espace de dimension p . On peut également considérer le nuage de p points dans l'espace de dimension n . On cherche donc le sous-espace de dimension q ($q \leq p$) pour lequel la somme des carrés des projections est maximale.

On s'est intéressé, jusqu'à maintenant, au nuage de n points dans l'espace de dimension p . On peut également considérer le nuage de p points dans l'espace de dimension n . On cherche donc le sous-espace de dimension q ($q \leq p$) pour lequel la somme des carrés des projections est maximale.

Solution

- On applique la même technique que précédemment.

On s'est intéressé, jusqu'à maintenant, au nuage de n points dans l'espace de dimension p . On peut également considérer le nuage de p points dans l'espace de dimension n . On cherche donc le sous-espace de dimension q ($q \leq p$) pour lequel la somme des carrés des projections est maximale.

Solution

- On applique la même technique que précédemment.
- Le premier axe factoriel de \mathbb{R}^n est porté par un vecteur unitaire v et l'on cherche à rendre maximale la somme des carrés des projections $X^T v$.

$$\begin{cases} \max_v \phi = v^T X X^T v \\ \text{sous la contrainte } v^T v = 1 \end{cases}$$

On s'est intéressé, jusqu'à maintenant, au nuage de n points dans l'espace de dimension p . On peut également considérer le nuage de p points dans l'espace de dimension n . On cherche donc le sous-espace de dimension q ($q \leq p$) pour lequel la somme des carrés des projections est maximale.

Solution

- On applique la même technique que précédemment.
- Le premier axe factoriel de \mathbb{R}^n est porté par un vecteur unitaire v et l'on cherche à rendre maximale la somme des carrés des projections $X^T v$.

$$\begin{cases} \max_v \phi = v^T X X^T v \\ \text{sous la contrainte } v^T v = 1 \end{cases}$$

- On nomme v_1 le vecteur satisfaisant le problème précédent.

On s'est intéressé, jusqu'à maintenant, au nuage de n points dans l'espace de dimension p . On peut également considérer le nuage de p points dans l'espace de dimension n . On cherche donc le sous-espace de dimension q ($q \leq p$) pour lequel la somme des carrés des projections est maximale.

Solution

- On applique la même technique que précédemment.
- Le premier axe factoriel de \mathbb{R}^n est porté par un vecteur unitaire v et l'on cherche à rendre maximale la somme des carrés des projections $XX^T v$.

$$\begin{cases} \max_v \phi = v^T XX^T v \\ \text{sous la contrainte } v^T v = 1 \end{cases}$$

- On nomme v_1 le vecteur satisfaisant le problème précédent.
- En suivant ensuite une démarche analogue à la précédente, on peut établir que le sous-espace vectoriel de dimension q qui assure une dispersion maximale des p points est défini par une **base orthonormée formée des q vecteurs propres v_1, v_2, \dots, v_q** correspondant aux q plus grandes valeurs propres de la matrice XX^T .

Relations entre les vecteurs u_i et v_i

Par définition de v_i , on a :

$$XX^T v_i = \mu_i v_i$$

où v_i et μ_i désignent respectivement le $i^{\text{ème}}$ vecteur propre et la $i^{\text{ème}}$ valeur propre de XX^T .

Relations entre les vecteurs u_i et v_i

Par définition de v_i , on a :

$$XX^T v_i = \mu_i v_i$$

où v_i et μ_i désignent respectivement le $i^{\text{ème}}$ vecteur propre et la $i^{\text{ème}}$ valeur propre de XX^T .

Notons r le rang de la matrice X , donc le rang de la matrice XX^T .

Relations entre les vecteurs u_i et v_i

Par définition de v_i , on a :

$$XX^T v_i = \mu_i v_i$$

où v_i et μ_i désignent respectivement le $i^{\text{ème}}$ vecteur propre et la $i^{\text{ème}}$ valeur propre de XX^T .

Notons r le rang de la matrice X , donc le rang de la matrice XX^T .

Il y a alors r valeurs propres non nulles avec $r \leq \min(n, p)$.

Relations entre les vecteurs u_i et v_i

Par définition de v_i , on a :

$$XX^T v_i = \mu_i v_i$$

où v_i et μ_i désignent respectivement le $i^{\text{ème}}$ vecteur propre et la $i^{\text{ème}}$ valeur propre de XX^T .

Notons r le rang de la matrice X , donc le rang de la matrice XX^T .

Il y a alors r valeurs propres non nulles avec $r \leq \min(n, p)$.

En prémultipliant les deux membres de cette relation par la matrice X^T , on obtient :

$$X^T X (X^T v_i) = \mu_i (X^T v_i)$$

Relations entre les vecteurs u_i et v_i

Par définition de v_i , on a :

$$XX^T v_i = \mu_i v_i$$

où v_i et μ_i désignent respectivement le $i^{\text{ème}}$ vecteur propre et la $i^{\text{ème}}$ valeur propre de XX^T .

Notons r le rang de la matrice X , donc le rang de la matrice XX^T .

Il y a alors r valeurs propres non nulles avec $r \leq \min(n, p)$.

En prémultipliant les deux membres de cette relation par la matrice X^T , on obtient :

$$X^T X (X^T v_i) = \mu_i (X^T v_i)$$

A chaque vecteur propre v_i , ($i \leq r$) de XX^T correspond donc un vecteur propre $u_i = X^T v_i$ de $X^T X$ relatif à la même valeur propre μ_i .

Relations entre les vecteurs u_i et v_i

Par définition de v_i , on a :

$$XX^T v_i = \mu_i v_i$$

où v_i et μ_i désignent respectivement le $i^{\text{ème}}$ vecteur propre et la $i^{\text{ème}}$ valeur propre de XX^T .

Notons r le rang de la matrice X , donc le rang de la matrice XX^T .

Il y a alors r valeurs propres non nulles avec $r \leq \min(n, p)$.

En prémultipliant les deux membres de cette relation par la matrice X^T , on obtient :

$$X^T X (X^T v_i) = \mu_i (X^T v_i)$$

A chaque vecteur propre v_i , ($i \leq r$) de XX^T correspond donc un vecteur propre $u_i = X^T v_i$ de $X^T X$ relatif à la même valeur propre μ_i .

Toute valeur propre non nulle de XX^T est donc valeur propre de $X^T X$ et les vecteurs propres correspondants sont liés par les relations :

$$u_i = k_i X^T v_i$$

où k_i sont des constantes.

Relations entre les vecteurs u_i et v_i

En prémultipliant de façon analogue par X les deux membres de la relation $X^T Xu_i = \lambda_i u_i$, on obtient :

$$XX^T(Xu_i) = \lambda_i(Xu_i)$$

Relations entre les vecteurs u_i et v_i

En prémultipliant de façon analogue par X les deux membres de la relation $X^T Xu_i = \lambda_i u_i$, on obtient :

$$XX^T(Xu_i) = \lambda_i(Xu_i)$$

Ainsi, à tout vecteur propre u_i , ($i \leq r$) de $X^T X$ correspond un vecteur propre Xu_i de XX^T relatif à la même valeur propre λ_i . On a donc finalement, pour tout $i \leq r$:

$$\lambda_i = \mu_i, \quad \text{et} \quad v_i = k'_i Xu_i$$

où les coefficients k'_i sont constants.

Relations entre les vecteurs u_i et v_i

En prémultipliant de façon analogue par X les deux membres de la relation $X^T Xu_i = \lambda_i u_i$, on obtient :

$$XX^T(Xu_i) = \lambda_i(Xu_i)$$

Ainsi, à tout vecteur propre u_i , ($i \leq r$) de $X^T X$ correspond un vecteur propre Xu_i de XX^T relatif à la même valeur propre λ_i . On a donc finalement, pour tout $i \leq r$:

$$\lambda_i = \mu_i, \quad \text{et} \quad v_i = k'_i Xu_i$$

où les coefficients k'_i sont constants.

En écrivant que tous les vecteurs u_i et v_i sont normés, $u_i^T u_i = v_i^T v_i = 1$, on trouve $k_i = k'_i = 1/\sqrt{\lambda_i}$.

Relations entre les vecteurs u_i et v_i

En prémultipliant de façon analogue par X les deux membres de la relation $X^T X u_i = \lambda_i u_i$, on obtient :

$$X X^T (X u_i) = \lambda_i (X u_i)$$

Ainsi, à tout vecteur propre u_i , ($i \leq r$) de $X^T X$ correspond un vecteur propre $X u_i$ de $X X^T$ relatif à la même valeur propre λ_i . On a donc finalement, pour tout $i \leq r$:

$$\lambda_i = \mu_i, \quad \text{et} \quad v_i = k'_i X u_i$$

où les coefficients k'_i sont constants.

En écrivant que tous les vecteurs u_i et v_i sont normés, $u_i^T u_i = v_i^T v_i = 1$, on trouve $k_i = k'_i = 1/\sqrt{\lambda_i}$.

On peut alors écrire, pour $i = 1, \dots, r$, le système de relations suivant souvent appelées formules de transition :

$$u_i = \frac{1}{\sqrt{\lambda_i}} X^T v_i, \qquad v_i = \frac{1}{\sqrt{\lambda_i}} X u_i$$

Relations entre les vecteurs u_i et v_i

L'axe F_i qui porte le vecteur unitaire u_i est appelé $i^{\text{ème}}$ **axe factoriel de \mathbb{R}^p** et l'axe G_i qui porte le vecteur unitaire v_i est appelé $i^{\text{ème}}$ axe factoriel de \mathbb{R}^n .

Relations entre les vecteurs u_i et v_i

L'axe F_i qui porte le vecteur unitaire u_i est appelé $i^{\text{ème}}$ **axe factoriel de \mathbb{R}^p** et l'axe G_i qui porte le vecteur unitaire v_i est appelé $i^{\text{ème}}$ axe factoriel de \mathbb{R}^n .

Les coordonnées des points du nuage sur le $i^{\text{ème}}$ axe dans \mathbb{R}^p (respectivement dans \mathbb{R}^n) sont, par construction, les composantes du vecteur Xu_i (respectivement de $X^T v_i$).

Relations entre les vecteurs u_i et v_i

L'axe F_i qui porte le vecteur unitaire u_i est appelé $i^{\text{ème}}$ **axe factoriel de \mathbb{R}^p** et l'axe G_i qui porte le vecteur unitaire v_i est appelé $i^{\text{ème}}$ axe factoriel de \mathbb{R}^n .

Les coordonnées des points du nuage sur le $i^{\text{ème}}$ axe dans \mathbb{R}^p (respectivement dans \mathbb{R}^n) sont, par construction, les composantes du vecteur Xu_i (respectivement de $X^T v_i$).

Il y a donc proportionnalité entre coordonnées des points sur le $i^{\text{ème}}$ axe dans un espace et composantes unitaires (cosinus directeur) du $i^{\text{ème}}$ axe de l'autre espace.

Reconstruction complète

Les formules de transition précédentes peuvent s'écrire de manière plus compacte en définissant les matrices de vecteurs propres $U = (u_1 \ u_2 \ \dots \ u_r)$ (de dimension $p \times r$) et $V = (v_1 \ v_2 \ \dots \ v_r)$ (de dimension $n \times r$) et la matrice diagonale des valeurs propres $\Lambda = \text{diag}(\lambda_1 \ \lambda_2 \ \dots \ \lambda_r)$.

Reconstruction complète

Les formules de transition précédentes peuvent s'écrire de manière plus compacte en définissant les matrices de vecteurs propres $U = (u_1 \ u_2 \ \dots \ u_r)$ (de dimension $p \times r$) et $V = (v_1 \ v_2 \ \dots \ v_r)$ (de dimension $n \times r$) et la matrice diagonale des valeurs propres $\Lambda = \text{diag}(\lambda_1 \ \lambda_2 \ \dots \ \lambda_r)$.

On a alors :

$$U = X^T V \Lambda^{-1/2}$$

et

$$V = X U \Lambda^{-1/2}$$

Reconstruction complète

Les formules de transition précédentes peuvent s'écrire de manière plus compacte en définissant les matrices de vecteurs propres $U = (u_1 \ u_2 \ \dots \ u_r)$ (de dimension $p \times r$) et $V = (v_1 \ v_2 \ \dots \ v_r)$ (de dimension $n \times r$) et la matrice diagonale des valeurs propres $\Lambda = \text{diag}(\lambda_1 \ \lambda_2 \ \dots \ \lambda_r)$.

On a alors :

$$U = X^T V \Lambda^{-1/2} \quad \text{et} \quad V = X U \Lambda^{-1/2}$$

En multipliant, par la droite, la seconde égalité par la matrice régulière $\Lambda^{1/2}$, on obtient :

$$XU = V \Lambda^{1/2}$$

Reconstruction complète

Les formules de transition précédentes peuvent s'écrire de manière plus compacte en définissant les matrices de vecteurs propres $U = (u_1 \ u_2 \ \dots \ u_r)$ (de dimension $p \times r$) et $V = (v_1 \ v_2 \ \dots \ v_r)$ (de dimension $n \times r$) et la matrice diagonale des valeurs propres $\Lambda = \text{diag}(\lambda_1 \ \lambda_2 \ \dots \ \lambda_r)$.

On a alors :

$$U = X^T V \Lambda^{-1/2} \quad \text{et} \quad V = X U \Lambda^{-1/2}$$

En multipliant, par la droite, la seconde égalité par la matrice régulière $\Lambda^{1/2}$, on obtient :

$$XU = V \Lambda^{1/2}$$

ou encore en post-multipliant par la matrice U^T et en remarquant que $UU^T = I$ (les vecteurs propres d'une matrice symétrique sont orthogonaux), on a également :

$$XUU^T = V \Lambda^{1/2} U^T$$

Reconstruction complète

Les formules de transition précédentes peuvent s'écrire de manière plus compacte en définissant les matrices de vecteurs propres $U = (u_1 \ u_2 \ \dots \ u_r)$ (de dimension $p \times r$) et $V = (v_1 \ v_2 \ \dots \ v_r)$ (de dimension $n \times r$) et la matrice diagonale des valeurs propres $\Lambda = \text{diag}(\lambda_1 \ \lambda_2 \ \dots \ \lambda_r)$.

On a alors :

$$U = X^T V \Lambda^{-1/2} \quad \text{et} \quad V = X U \Lambda^{-1/2}$$

En multipliant, par la droite, la seconde égalité par la matrice régulière $\Lambda^{1/2}$, on obtient :

$$XU = V\Lambda^{1/2}$$

ou encore en post-multipliant par la matrice U^T et en remarquant que $UU^T = I$ (les vecteurs propres d'une matrice symétrique sont orthogonaux), on a également :

$$XUU^T = V\Lambda^{1/2}U^T$$

c'est-à-dire :

$$X = V\Lambda^{1/2}U^T$$

Reconstruction complète

Cette dernière expression indique clairement que l'on peut reconstruire (complètement) la matrice X si l'on connaît les valeurs propres et les vecteurs propres des matrices symétriques XX^T et X^TX .

Reconstruction complète

Cette dernière expression indique clairement que l'on peut reconstruire (complètement) la matrice X si l'on connaît les valeurs propres et les vecteurs propres des matrices symétriques XX^T et $X^T X$.

Cette décomposition est connue sous le nom de **décomposition en valeurs singulières** (SVD for Singular Value Decomposition).

Reconstruction complète

Cette dernière expression indique clairement que l'on peut reconstruire (complètement) la matrice X si l'on connaît les valeurs propres et les vecteurs propres des matrices symétriques XX^T et $X^T X$.

Cette décomposition est connue sous le nom de **décomposition en valeurs singulières** (SVD for Singular Value Decomposition).

Par définition, on appelle $i^{\text{ème}}$ composante principale le vecteur dont les composantes sont les coordonnées des points du nuage sur le $i^{\text{ème}}$ axe factoriel (cette analyse peut être effectuée dans \mathbb{R}^p ou dans \mathbb{R}^n).

Composantes principales dans \mathbb{R}^p (projections des observations sur les vecteurs propres)	$C_o = XU = V\Lambda^{1/2}$
Composantes principales dans \mathbb{R}^n (projections des variables sur les vecteurs propres)	$C_v = X^T V = U\Lambda^{1/2}$

Reconstruction partielle

La reconstruction peut également être partielle ou approchée, on parlera alors d'**estimation**. En effet, la décomposition précédente peut s'écrire sous la forme :

$$X = \sum_{i=1}^r \sqrt{\lambda_i} v_i u_i^T$$

Reconstruction partielle

La reconstruction peut également être partielle ou approchée, on parlera alors d'**estimation**. En effet, la décomposition précédente peut s'écrire sous la forme :

$$X = \sum_{i=1}^r \sqrt{\lambda_i} v_i u_i^T$$

Si les valeurs propres $\lambda_i, i = 1, \dots, r$ sont rangées par ordre décroissant, une estimation (reconstruction partielle) peut consister à ne tenir compte que des **q premières valeurs propres** (les q plus grandes) :

$$X \simeq \hat{X} = \sum_{i=1}^q \sqrt{\lambda_i} v_i u_i^T$$

Reconstruction partielle

La reconstruction peut également être partielle ou approchée, on parlera alors d'**estimation**. En effet, la décomposition précédente peut s'écrire sous la forme :

$$X = \sum_{i=1}^r \sqrt{\lambda_i} v_i u_i^T$$

Si les valeurs propres $\lambda_i, i = 1, \dots, r$ sont rangées par ordre décroissant, une estimation (reconstruction partielle) peut consister à ne tenir compte que des **q premières valeurs propres** (les q plus grandes) :

$$X \simeq \hat{X} = \sum_{i=1}^q \sqrt{\lambda_i} v_i u_i^T$$

On remplace ainsi les np éléments de X par un ensemble de $q(n+p)$ nombres constitué des q vecteurs $\sqrt{\lambda_i} v_i$ à n composantes et q vecteurs u_i à p composantes.

Reconstruction partielle

Comme chaque valeur propre λ_i mesure la somme des carrés des projections sur l'axe F_i , la part d'inertie expliquée par l'axe F_i s'explique sous la forme :

$$I(k) = \frac{\lambda_k}{\sum_{i=1}^r \lambda_i}$$

Reconstruction partielle

Comme chaque valeur propre λ_i mesure la somme des carrés des projections sur l'axe F_i , la part d'inertie expliquée par l'axe F_i s'explique sous la forme :

$$I(k) = \frac{\lambda_k}{\sum_{i=1}^r \lambda_i}$$

La qualité de la reconstruction s'appuyant sur les seules q premières valeurs propres peut donc être évaluée par la quantité :

$$\tau_q = \frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^r \lambda_i}$$

Reconstruction partielle

Comme chaque valeur propre λ_i mesure la somme des carrés des projections sur l'axe F_i , la part d'inertie expliquée par l'axe F_i s'exprime sous la forme :

$$I(k) = \frac{\lambda_k}{\sum_{i=1}^r \lambda_i}$$

La qualité de la reconstruction s'appuyant sur les seules q premières valeurs propres peut donc être évaluée par la quantité :

$$\tau_q = \frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^r \lambda_i}$$

Ce quotient, appelé **taux d'inertie**, mesure ainsi la part de la dispersion du nuage imputable au sous-espace à q dimensions retenu.

1 Analyses factorielles

2 Analyse générale

- Ajustement par un sous-espace vectoriel dans \mathbb{R}^p
- Ajustement par un sous-espace vectoriel dans \mathbb{R}^n
- Lien entre les sous-espaces de \mathbb{R}^p et de \mathbb{R}^n
- Reconstruction complète et partielle de la matrice de départ

3 Analyse en composantes principales normée

- Usage et description

4 Elaboration d'un "modèle ACP"

- Choix du nombre d'axes
- Modèles ACP

Usage et description

- L'analyse en composantes principales s'utilise essentiellement lorsque la matrice de données initiale T représente des valeurs numériques issues de n observations de p variables.

Usage et description

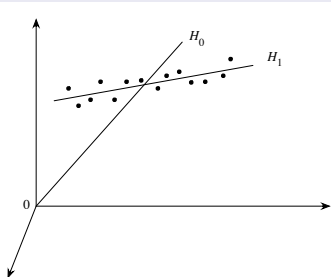
- L'analyse en composantes principales s'utilise essentiellement lorsque la matrice de données initiale T représente des valeurs numériques issues de n observations de p variables.
- Lorsqu'on effectue l'analyse dans \mathbb{R}^p , les n points dans cet espace sont les observations. On souhaite obtenir une représentation de la proximité de ces observations dans un espace de faible dimension.

Usage et description

- L'analyse en composantes principales s'utilise essentiellement lorsque la matrice de données initiale T représente des valeurs numériques issues de n observations de p variables.
- Lorsqu'on effectue l'analyse dans \mathbb{R}^p , les n points dans cet espace sont les observations. On souhaite obtenir une représentation de la proximité de ces observations dans un espace de faible dimension.
- Or, le sous-espace à une dimension qui assure une déformation minimale des proximités entre observations n'a pas à passer par l'origine du repère.

Usage et description

- L'analyse en composantes principales s'utilise essentiellement lorsque la matrice de données initiale T représente des valeurs numériques issues de n observations de p variables.
- Lorsqu'on effectue l'analyse dans \mathbb{R}^p , les n points dans cet espace sont les observations. On souhaite obtenir une représentation de la proximité de ces observations dans un espace de faible dimension.
- Or, le sous-espace à une dimension qui assure une déformation minimale des proximités entre observations n'a pas à passer par l'origine du repère.



- Il est clair que le sous-espace affine H_1 rend mieux compte des proximités entre points que le sous-espace vectoriel H_0 (droite passant par l'origine).

Usage et description

- Cette remarque conduit à prendre comme nouvelle origine le centre de gravité du nuage dont les p composantes sont les p moyennes arithmétiques \bar{t}_j .

Usage et description

- Cette remarque conduit à prendre comme nouvelle origine le centre de gravité du nuage dont les p composantes sont les p moyennes arithmétiques \bar{t}_j .
- On effectuera donc l'analyse générale sur la matrice de terme général $x_{ij} = (t_{ij} - \bar{t}_j)/\sqrt{n}$.

Usage et description

- Cette remarque conduit à prendre comme nouvelle origine le centre de gravité du nuage dont les p composantes sont les p moyennes arithmétiques \bar{t}_j .
- On effectuera donc l'analyse générale sur la matrice de terme général $x_{ij} = (t_{ij} - \bar{t}_j)/\sqrt{n}$.
- L'influence du niveau de chacune des variables sera ainsi éliminée. Le coefficient $1/\sqrt{n}$ n'a pour objet que de faire coïncider la matrice à diagonaliser $X^T X$ avec la **matrice des covariances expérimentales** conformément à un usage répandu.

Usage et description

- Cette remarque conduit à prendre comme nouvelle origine le centre de gravité du nuage dont les p composantes sont les p moyennes arithmétiques \bar{t}_j .
- On effectuera donc l'analyse générale sur la matrice de terme général $x_{ij} = (t_{ij} - \bar{t}_j)/\sqrt{n}$.
- L'influence du niveau de chacune des variables sera ainsi éliminée. Le coefficient $1/\sqrt{n}$ n'a pour objet que de faire coïncider la matrice à diagonaliser $X^T X$ avec la **matrice des covariances expérimentales** conformément à un usage répandu.
- Une modification supplémentaire de la matrice initiale peut également être nécessaire si les dispersions des variables sont très différentes ; elle conduira à l'analyse en composantes principales normée, c'est-à-dire l'analyse de la matrice de terme général :

$$x_{ij} = (t_{ij} - \bar{t}_j)/s_j\sqrt{n} \quad \text{avec} \quad s_j^2 = \frac{1}{n} \sum_{i=1}^n (t_{ij} - \bar{t}_j)^2$$

Usage et description

- En analyse en composantes principales normée, la matrice à diagonaliser $X^T X$ coïncide alors avec la **matrice des corrélations** entre les variables.

Usage et description

- En analyse en composantes principales normée, la matrice à diagonaliser $X^T X$ coïncide alors avec la **matrice des corrélations** entre les variables.

Usage et description

- En analyse en composantes principales normée, la matrice à diagonaliser $X^T X$ coïncide alors avec la **matrice des corrélations** entre les variables.
- En résumé, l'analyse du nuage de points dans \mathbb{R}^p conduit à effectuer une **translation d'origine au centre de gravité** du nuage et à **changer les échelles** des différents axes. L'analyse générale s'effectue alors sur la matrice des corrélations.

Usage et description

- En analyse en composantes principales normée, la matrice à diagonaliser $X^T X$ coïncide alors avec la **matrice des corrélations** entre les variables.
- En résumé, l'analyse du nuage de points dans \mathbb{R}^p conduit à effectuer une **translation d'origine au centre de gravité** du nuage et à **changer les échelles** des différents axes. L'analyse générale s'effectue alors sur la matrice des corrélations.
- Si l'on considère maintenant l'analyse dans \mathbb{R}^n , les p points du nuage sont maintenant les points-variables. L'analyse de la matrice X dans \mathbb{R}^p induit une analyse dans \mathbb{R}^n . Cependant, les indices i et j ne jouent pas des rôles symétriques dans la transformation initiale. Les interprétations géométriques associées à cette transformation seront donc différentes.

Usage et description

- En analyse en composantes principales normée, la matrice à diagonaliser $X^T X$ coïncide alors avec la **matrice des corrélations** entre les variables.
- En résumé, l'analyse du nuage de points dans \mathbb{R}^p conduit à effectuer une **translation d'origine au centre de gravité** du nuage et à **changer les échelles** des différents axes. L'analyse générale s'effectue alors sur la matrice des corrélations.
- Si l'on considère maintenant l'analyse dans \mathbb{R}^n , les p points du nuage sont maintenant les points-variables. L'analyse de la matrice X dans \mathbb{R}^p induit une analyse dans \mathbb{R}^n . Cependant, les indices i et j ne jouent pas des rôles symétriques dans la transformation initiale. Les interprétations géométriques associées à cette transformation seront donc différentes.
- Les composantes principales peuvent être considérées comme de nouvelles variables, combinaisons linéaires des variables initiales, non corrélées entre elles et d'inertie (ou de variance) maximale.

1 Analyses factorielles

2 Analyse générale

- Ajustement par un sous-espace vectoriel dans \mathbb{R}^p
- Ajustement par un sous-espace vectoriel dans \mathbb{R}^n
- Lien entre les sous-espaces de \mathbb{R}^p et de \mathbb{R}^n
- Reconstruction complète et partielle de la matrice de départ

3 Analyse en composantes principales normée

- Usage et description

4 Elaboration d'un "modèle ACP"

- Choix du nombre d'axes
- Modèles ACP

Introduction

- En règle générale, l'élaboration d'un modèle passe par deux étapes. Au cours de la première, on choisit une classe de modèles imposant ainsi la **structure** de celui-ci. On utilise ensuite les données expérimentales pour **identifier les paramètres** décrivant la structure choisie.

Introduction

- En règle générale, l'élaboration d'un modèle passe par deux étapes. Au cours de la première, on choisit une classe de modèles imposant ainsi la **structure** de celui-ci. On utilise ensuite les données expérimentales pour **identifier les paramètres** décrivant la structure choisie.
- Lorsqu'on effectue une analyse en composantes principales, par construction, on se restreint à la classe des **modèles linéaires**. La structure du modèle dépend du **nombre de composantes principales retenues** pour l'estimation de la matrice initiale.

Introduction

- En règle générale, l'élaboration d'un modèle passe par deux étapes. Au cours de la première, on choisit une classe de modèles imposant ainsi la **structure** de celui-ci. On utilise ensuite les données expérimentales pour **identifier les paramètres** décrivant la structure choisie.
- Lorsqu'on effectue une analyse en composantes principales, par construction, on se restreint à la classe des **modèles linéaires**. La structure du modèle dépend du **nombre de composantes principales retenues** pour l'estimation de la matrice initiale.
- L'estimation des paramètres du "modèle ACP" se résume au choix des valeurs et vecteurs propres de la matrice de corrélation des données.

Choix du nombre d'axes

- L'analyse en composantes principales cherche donc une approximation de la matrice initiale des données X au moyen d'une matrice de rang inférieur issue d'une décomposition en valeurs singulières.

Choix du nombre d'axes

- L'analyse en composantes principales cherche donc une approximation de la matrice initiale des données X au moyen d'une matrice de rang inférieur issue d'une décomposition en valeurs singulières.
- La question qui se pose alors concerne le choix du nombre de composantes principales (nombre d'axes) qui doit être retenu pour "capter" l'information pertinente contenue dans X .

Choix du nombre d'axes

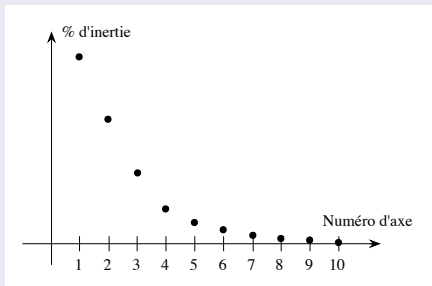
- L'analyse en composantes principales cherche donc une approximation de la matrice initiale des données X au moyen d'une matrice de rang inférieur issue d'une décomposition en valeurs singulières.
- La question qui se pose alors concerne le choix du nombre de composantes principales (nombre d'axes) qui doit être retenu pour "capter" l'information pertinente contenue dans X .
- Règle heuristique élémentaire – Approche inertielle
La part d'inertie du nuage de points, expliquée par l'axe F_i , est proportionnelle à la valeur propre correspondante. Le nombre d'axes retenu peut alors être obtenu en fixant un seuil correspondant au pourcentage minimum d'inertie que l'on veut restituer. Pour un seuil I_{min} , le nombre de composantes à retenir est le plus petit entier q vérifiant la relation :

$$\tau_q = \frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^r \lambda_i} \geq I_{min}$$

Si les données initiales sont bruitées, la capacité à fournir le nombre "correct" de composantes principales dépendra fortement du rapport signal sur bruit.

Choix du nombre d'axes

- Règles heuristiques élémentaires – Approche "décroissance des valeurs propres"
Comme l'inertie expliquée par chaque axe va en décroissant, on peut construire la courbe de décroissance des axes en portant en abscisse les numéros des axes (ou des valeurs propres correspondantes) et en ordonnée les pourcentages d'inertie restitués par chaque axe (l'amplitude des valeurs propres).



Si l'on observe un "coude" dans cette courbe, on ne retiendra que les axes dont le numéro d'ordre est situé avant ce coude. Pour l'exemple de la figure précédente, on retiendrait 5 composantes.

Modèles ACP

- Revenons sur la décomposition en valeurs singulières d'une matrice de données X de dimension $n \times p$

$$X = VSU^T$$

où S est la matrice diagonale des valeurs singulières correspondant aux racines carrées des valeurs propres de la matrice $X^T X$ et $V \in \mathbb{R}^{n \times p}$, $S \in \mathbb{R}^{p \times p}$ et $U \in \mathbb{R}^{p \times p}$

Modèles ACP

- Revenons sur la décomposition en valeurs singulières d'une matrice de données X de dimension $n \times p$

$$X = VSU^T$$

où S est la matrice diagonale des valeurs singulières correspondant aux racines carrées des valeurs propres de la matrice $X^T X$ et $V \in \mathbb{R}^{n \times p}$, $S \in \mathbb{R}^{p \times p}$ et $U \in \mathbb{R}^{p \times p}$

- On fera l'hypothèse que la décomposition est telle que les valeurs singulières sont ordonnées par ordre décroissant d'amplitude. Lorsqu'on effectue une estimation de X , en ne s'appuyant que sur les q premières composantes principales, on n'utilise que les q premières colonnes de V et de U .

Modèles ACP

- Revenons sur la décomposition en valeurs singulières d'une matrice de données X de dimension $n \times p$

$$X = VSU^T$$

où S est la matrice diagonale des valeurs singulières correspondant aux racines carrées des valeurs propres de la matrice $X^T X$ et $V \in \mathbb{R}^{n \times p}$, $S \in \mathbb{R}^{p \times p}$ et $U \in \mathbb{R}^{p \times p}$

- On fera l'hypothèse que la décomposition est telle que les valeurs singulières sont ordonnées par ordre décroissant d'amplitude. Lorsqu'on effectue une estimation de X , en ne s'appuyant que sur les q premières composantes principales, on n'utilise que les q premières colonnes de V et de U .
- La décomposition peut alors être partitionnée de la manière suivante :

$$X = (V_1 \quad V_2) \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} U_1^T \\ U_2^T \end{pmatrix} = V_1 S_1 U_1^T + V_2 S_2 U_2^T$$

où $V_1 \in \mathbb{R}^{n \times q}$, $S_1 \in \mathbb{R}^{q \times q}$ et $U_1^T \in \mathbb{R}^{q \times p}$

Modèles ACP

- Revenons sur la décomposition en valeurs singulières d'une matrice de données X de dimension $n \times p$

$$X = VSU^T$$

où S est la matrice diagonale des valeurs singulières correspondant aux racines carrées des valeurs propres de la matrice $X^T X$ et $V \in \mathbb{R}^{n \times p}$, $S \in \mathbb{R}^{p \times p}$ et $U \in \mathbb{R}^{p \times p}$

- On fera l'hypothèse que la décomposition est telle que les valeurs singulières sont ordonnées par ordre décroissant d'amplitude. Lorsqu'on effectue une estimation de X , en ne s'appuyant que sur les q premières composantes principales, on n'utilise que les q premières colonnes de V et de U .
- La décomposition peut alors être partitionnée de la manière suivante :

$$X = (V_1 \quad V_2) \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} U_1^T \\ U_2^T \end{pmatrix} = V_1 S_1 U_1^T + V_2 S_2 U_2^T$$

où $V_1 \in \mathbb{R}^{n \times q}$, $S_1 \in \mathbb{R}^{q \times q}$ et $U_1^T \in \mathbb{R}^{q \times p}$

- L'estimation de X s'écrit alors :

$$\hat{X} = V_1 S_1 U_1^T$$

Modèles ACP

- Cette estimation peut également s'exprimer en fonction de X . En effet, on a :

$$X = V_1 S_1 U_1^T + V_2 S_2 U_2^T$$

en post-multipliant l'expression de X par $U_1 U_1^T$, on obtient :

$$X U_1 U_1^T = V_1 S_1 U_1^T U_1 U_1^T + V_2 S_2 U_2^T U_1 U_1^T$$

Modèles ACP

- Cette estimation peut également s'exprimer en fonction de X . En effet, on a :

$$X = V_1 S_1 U_1^T + V_2 S_2 U_2^T$$

en post-multipliant l'expression de X par $U_1 U_1^T$, on obtient :

$$X U_1 U_1^T = V_1 S_1 U_1^T U_1 U_1^T + V_2 S_2 U_2^T U_1 U_1^T$$

- En notant que les vecteurs propres de la matrice U forment une base orthogonale, on a $U_1^T U_1 = I$ et $U_2^T U_1 = 0$ d'où :

$$\hat{X} = X U_1 U_1^T = X C$$

Modèles ACP

- Cette estimation peut également s'exprimer en fonction de X . En effet, on a :

$$X = V_1 S_1 U_1^T + V_2 S_2 U_2^T$$

en post-multipliant l'expression de X par $U_1 U_1^T$, on obtient :

$$X U_1 U_1^T = V_1 S_1 U_1^T U_1 U_1^T + V_2 S_2 U_2^T U_1 U_1^T$$

- En notant que les vecteurs propres de la matrice U forment une base orthogonale, on a $U_1^T U_1 = I$ et $U_2^T U_1 = 0$ d'où :

$$\hat{X} = X U_1 U_1^T = X C$$

La matrice C représente alors le "modèle ACP" des données initiales

Modèles ACP

- Cette estimation peut également s'exprimer en fonction de X . En effet, on a :

$$X = V_1 S_1 U_1^T + V_2 S_2 U_2^T$$

en post-multipliant l'expression de X par $U_1 U_1^T$, on obtient :

$$XU_1 U_1^T = V_1 S_1 U_1^T U_1 U_1^T + V_2 S_2 U_2^T U_1 U_1^T$$

- En notant que les vecteurs propres de la matrice U forment une base orthogonale, on a $U_1^T U_1 = I$ et $U_2^T U_1 = 0$ d'où :

$$\hat{X} = XU_1 U_1^T = XC$$

La matrice C représente alors le "modèle ACP" des données initiales

- Différents modèles ACP peuvent être engendrés selon le nombre de composantes principales retenues ; ces modèles sont implicitement tous déterminés une fois que l'on a calculé les valeurs propres et vecteurs propres de la matrice de corrélation.

Modèles ACP

- Cette estimation peut également s'exprimer en fonction de X . En effet, on a :

$$X = V_1 S_1 U_1^T + V_2 S_2 U_2^T$$

en post-multipliant l'expression de X par $U_1 U_1^T$, on obtient :

$$X U_1 U_1^T = V_1 S_1 U_1^T U_1 U_1^T + V_2 S_2 U_2^T U_1 U_1^T$$

- En notant que les vecteurs propres de la matrice U forment une base orthogonale, on a $U_1^T U_1 = I$ et $U_2^T U_1 = 0$ d'où :

$$\hat{X} = X U_1 U_1^T = X C$$

La matrice C représente alors le "modèle ACP" des données initiales

- Différents modèles ACP peuvent être engendrés selon le nombre de composantes principales retenues ; ces modèles sont implicitement tous déterminés une fois que l'on a calculé les valeurs propres et vecteurs propres de la matrice de corrélation.
- Contrairement à des techniques de régression linéaire multivariées où l'incorporation d'une variable dans un modèle remet en cause l'identification paramétrique, il suffit ici simplement de sélectionner des vecteurs propres.

Remarque

- Considérons une matrice X , de dimension $n \times p$ (avec $n > p$), de données non bruitées, dont le rang est égal à r .

Remarque

- Considérons une matrice X , de dimension $n \times p$ (avec $n > p$), de données non bruitées, dont le rang est égal à r .
- Cela signifie qu'il existe implicitement $p - r$ relations linéaires indépendantes entre les colonnes de cette matrice. Il y a alors $p - r$ valeurs singulières nulles dans la décomposition en valeurs singulières de X (ou $p - r$ valeurs propres nulles dans la décomposition en valeurs propres de $X^T X$).

Remarque

- Considérons une matrice X , de dimension $n \times p$ (avec $n > p$), de données non bruitées, dont le rang est égal à r .
- Cela signifie qu'il existe implicitement $p - r$ relations linéaires indépendantes entre les colonnes de cette matrice. Il y a alors $p - r$ valeurs singulières nulles dans la décomposition en valeurs singulières de X (ou $p - r$ valeurs propres nulles dans la décomposition en valeurs propres de $X^T X$).
- Lorsque les données considérées sont bruitées, ou lorsque les relations de dépendance ne sont pas rigoureusement linéaires, les valeurs propres précédentes ne sont plus rigoureusement nulles.

Remarque

- Considérons une matrice X , de dimension $n \times p$ (avec $n > p$), de données non bruitées, dont le rang est égal à r .
- Cela signifie qu'il existe implicitement $p - r$ relations linéaires indépendantes entre les colonnes de cette matrice. Il y a alors $p - r$ valeurs singulières nulles dans la décomposition en valeurs singulières de X (ou $p - r$ valeurs propres nulles dans la décomposition en valeurs propres de $X^T X$).
- Lorsque les données considérées sont bruitées, ou lorsque les relations de dépendance ne sont pas rigoureusement linéaires, les valeurs propres précédentes ne sont plus rigoureusement nulles.
- On s'intéresse alors aux seules valeurs propres "dominantes" au sens de celles qui expliquent le plus l'inertie ou la variance des points-observations, d'où l'intérêt des critères de choix du nombre d'axes



Didier Maquin

Professeur d'Automatique
Université de Lorraine

Ecole Nationale Supérieure d'Electricité et de Mécanique

Ecole Nationale Supérieure des Mines de Nancy

Centre de Recherche en Automatique de Nancy

Contact : didier.maquin@univ-lorraine.fr

Localisation : bureau 116 jaune

Plus de détails ?

Site personnel : <http://www.cran.univ-lorraine.fr/didier.maquin>

Site du laboratoire : <http://www.cran.univ-lorraine.fr>