Analyse des Données

Analyse Factorielle : Fondements 1

Notes du cours

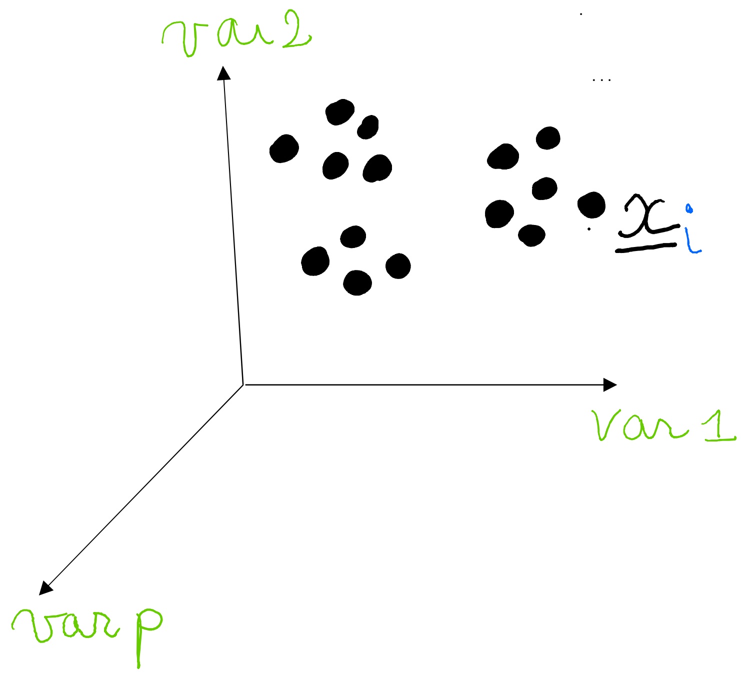
Données

$$
\begin{array}{cc}
 & \text{variables } p \\
\text{individus } n & \begin{pmatrix} x_i^1 & x_i^2 & \cdots & x_i^p \\ & & & \end{pmatrix}
\end{array}
$$

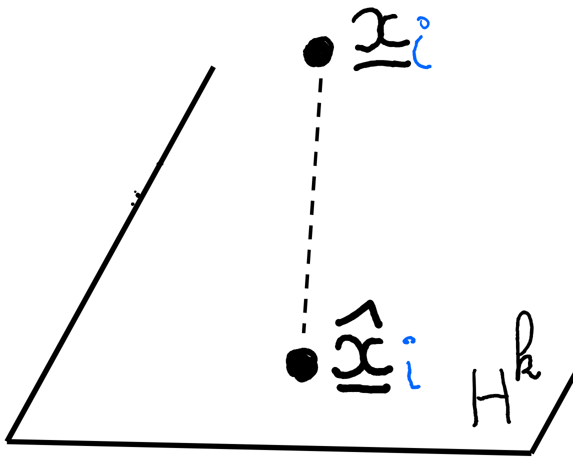
individu

$$\underline{x}_i = (x_i^1, \dots x_i^p) \in \mathbb{R}^p$$

Nuage des individus $\mathcal{N}(I)$



Problème : Représenter le nuage dans un espace de dimension faible k sans perdre trop d'information

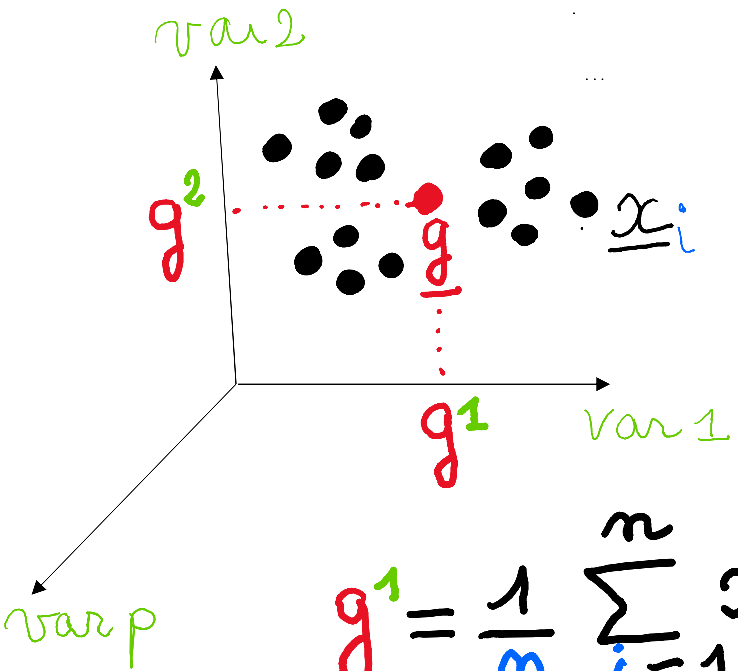


$k \ll p$

Un problème d'optimisation

$$\underset{H^k}{\mathrm{Min}} \; M(H^k) = \frac{1}{n} \sum_{i=1}^{n} d^2(\underline{x}_i, \hat{\underline{x}}_i)$$

Les trois prochains cours sont consacrés à la résolution de ce problème !
Il est nécessaire d'introduire différents ingrédients.

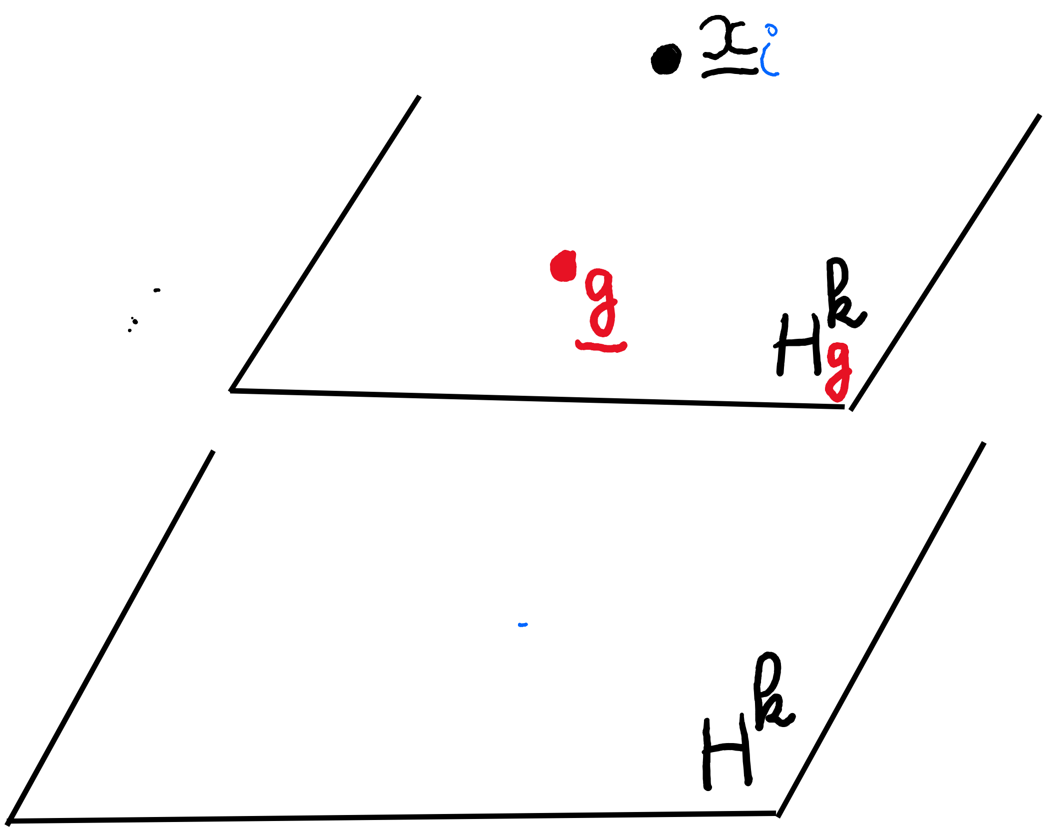


$$g^1 = \frac{1}{n}\sum_{i=1}^{n} x_i^1$$

$$\underline{g} = (g^1, \dots, g^p) \in \mathbb{R}^p$$

Généralisation de la moyenne au cas multi-dimensionnel

Barycentre du nuage $\mathcal{N}(I)$



H^k_g : sous-espace affine // à H^k passant par g

$$M(H^k) = M(H^k_g) + d^2(H^k, H^k_g)$$

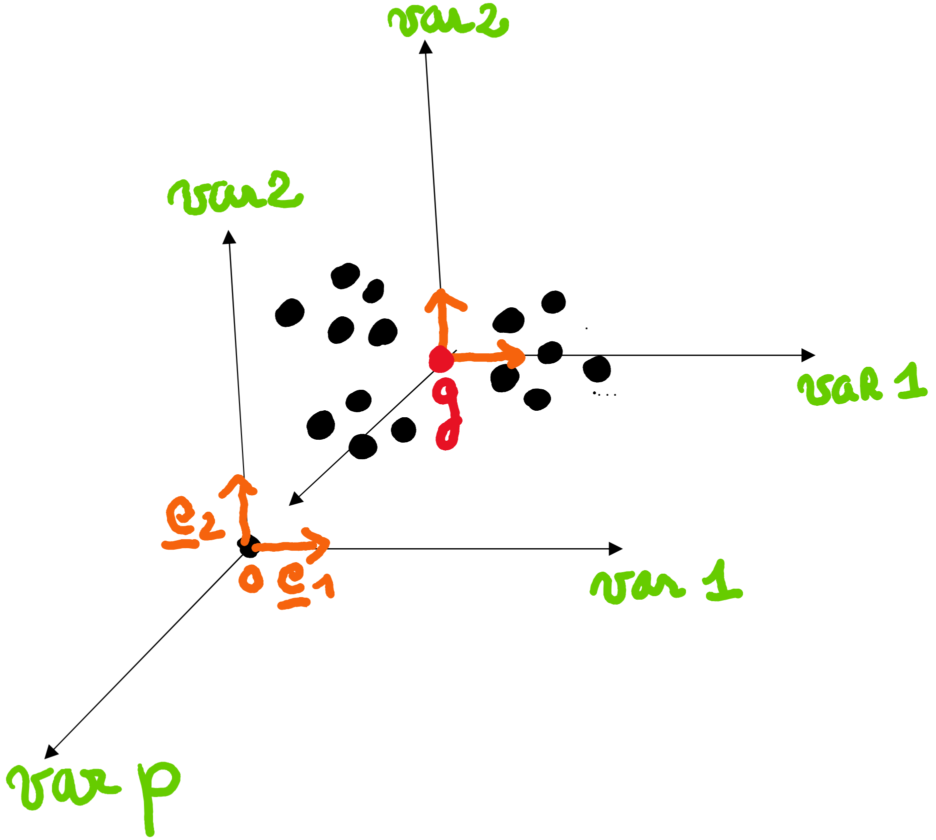
Conséquence. Se limiter dans la recherche aux sous espaces passant par le barycentre.

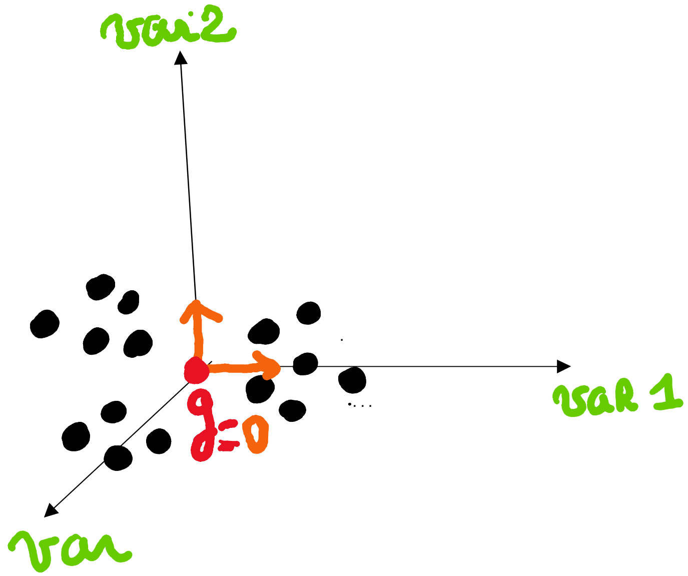
Problème :

Rechercher H^k, $k \ll p$, qui passe par le barycentre g du nuage $N(I)$ et qui minimise $M(H^k)$

Astuce

Le barycentre devient l'origine du repère



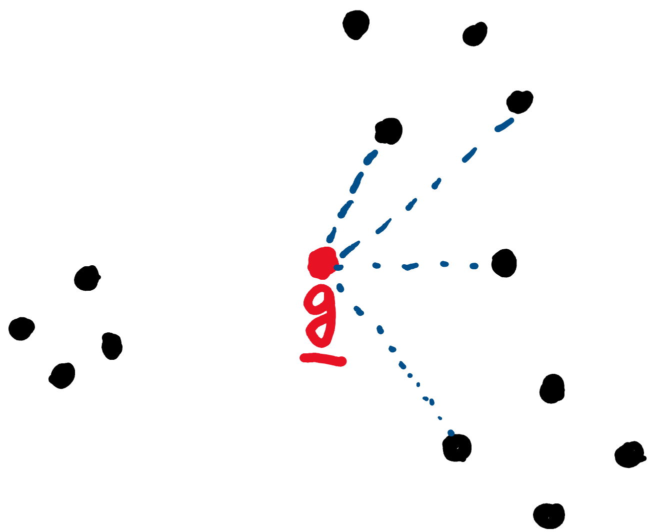


Problème

Pour $k \ll p$ fixé, chercher le sous-espace $H^k_{g=0}$ passant par g (origine du repère) minimisant $M(H^k_{g=0})$

$$\min_{H^k_g} M(H^k_g) = \frac{1}{n} \sum_{i=1} d^2(\underline{x}_i, \hat{\underline{x}}_i)$$

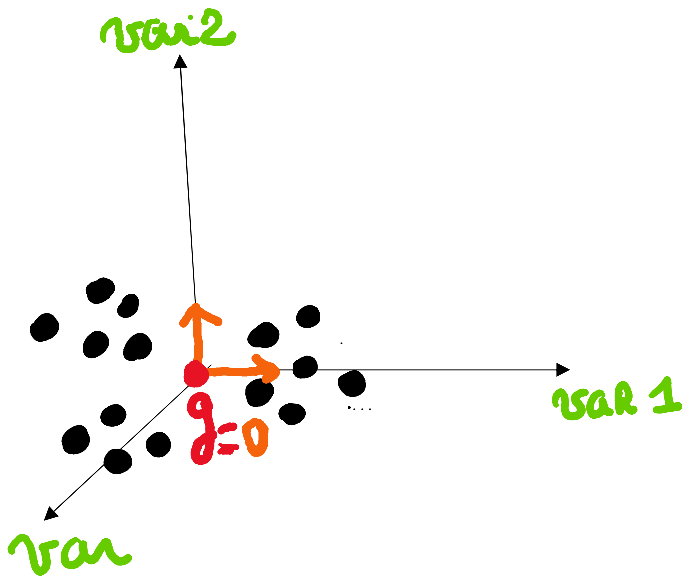
Réécriture du problème



Inertie du nuage $\mathcal{N}(I)$

$$I_g = \frac{1}{n}\sum_{i=1}^{n} d^2(\underline{x}_i, \underline{g})$$

Généralisation de la variance au cas multidimensionnel

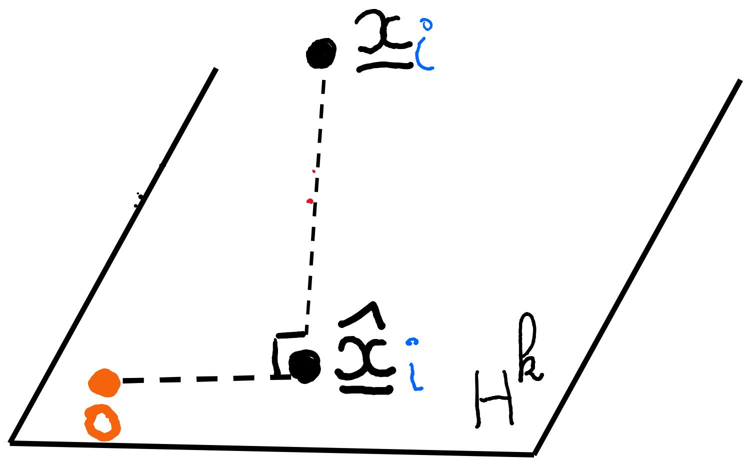


Réécriture de l'inertie
en centrant le repère sur g

$$I_g = \frac{1}{n}\sum_{i=1}^{n} d^2(\underline{x}_i, \underline{g})$$

$$I_{g=0} = \frac{1}{n}\sum_{i=1}^{n} d^2(\underline{x}_i, \underline{g})$$

$$= \frac{1}{n}\sum_{i=1}^{n} \|\underline{x}_i\|^2$$



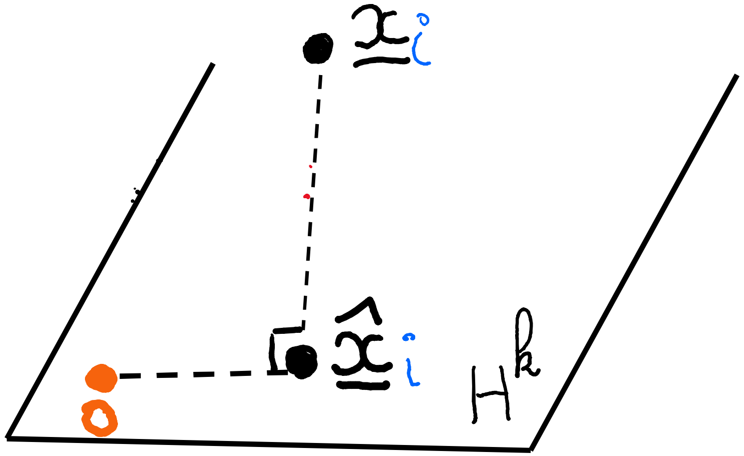
Pythagore

$$I_O = \frac{1}{n}\sum_{i=1}^{n} \|\hat{\underline{x}}_i\|^2 + \frac{1}{n}\sum_{i=1}^{n} \|\underline{x}_i - \hat{\underline{x}}_i\|^2$$

inertie totale

inertie expliquée

inertie résiduelle



$$I_O = \frac{1}{n}\sum_{i=1}^{n}\|\hat{\underline{x}}_i\|^2 + \frac{1}{n}\sum_{i=1}^{n}\|\underline{x}_i - \hat{\underline{x}}_i\|^2$$

dépend du
nuage et pas
de la projection

Maximiser $\Longleftrightarrow$ Minimiser

Réécriture du problème

$$I_0 = \frac{1}{n}\sum_{i=1}^{n}\|\hat{x}_i\|^2 + \frac{1}{n}\sum_{i=1}^{n}\|\underline{x}_i - \hat{x}_i\|^2$$

Maximiser Minimiser

Problème : Pour k fixé, $k \ll p$, chercher le sous-espace H^k passant par le barycentre tel que

$$\underset{\hat{x}_i \in H^k}{\text{Max}} \; \frac{1}{n}\sum_{i=1}^{n}\|\hat{x}_i\|^2$$