Analyse de données

Individu moyen : $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} X^T . 1_n$ $\quad X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} = \begin{pmatrix} v_1 & \cdots & v_p \end{pmatrix}$

Données centrées : $x_i^c = x_i - \bar{x} \Rightarrow X^c = X - 1_n \bar{x}^T$ (toujours)

calculer $\bar{v}_1, \dots, \bar{v}_p$.

Réduire les données $z_i = \frac{x_i - \bar{x}}{\sigma_x}$ • Si unités ≠ ou dynamiques ≠.

Matrice de covariance $\Sigma = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^T (x_i - \bar{x}) = \frac{1}{n} X_c^T X_c$

- $Cov(x_i, x_j) \approx 0 \Rightarrow$ pas liées
- $Cov(x_i, x_j) \simeq \pm 1 \Rightarrow$ positivement / négativement liées.

Distance : $d(x_i, x_j) = (x_i - x_j)^T \Sigma^{-1} (x_i - x_j)$

Inertie totale : $I(x) = trace(\Sigma) = \sum_{i=1}^{n} \sigma_i^2$.

ACP : $n \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}_p \longrightarrow n \begin{pmatrix} & \vdots & \\ \cdots & c_{ij} & \cdots \\ & \vdots & \end{pmatrix}_q$ $\quad c_{ij} = \langle x_i, u_j \rangle$.

$\Pi = u u^T$ (projecté) $\quad \tilde{x}_i = \Pi x_i$.

Fonction de coût $J(\beta) = \frac{1}{2} \sum_{i=1}^{n} \| y_i - f(x_i, \beta) \|^2$

cas linéaire : $J(\beta) = \frac{1}{2} \| A\beta - b \|^2$.

Inertie par rapport à un axe $I(j) = \frac{\lambda_j}{\sum \lambda_i}$

Corrélation : Il y a corrélation si $\exists j \,/\, I(j) \gg I(i)$.

Axe principal : v_1 le vect propre associé à λ_1 vp $|\lambda|$ grande.

Composante principale : $C_1 = X v_1$.