

ACP cheat sheet

On considère la matrice X du nuage de point avec n individus et p variables, donc $X \in M^{n \times p}$.

Le but d'une ACP est de déterminer de nouveaux axes $a_i \in \mathbb{R}^p$ (souvent 3 ou 4) tels que l'inertie (ou la variance) de la projection du nuage de point sur ces nouveaux axes soient maximale.

On pourra alors écrire X en fonction des coordonnées des individus dans ces nouveaux axes. On nommera alors les nouvelles variables composantes principales, les axes étant eux appelés axes principaux.

Transformation de l'échantillon:

Deux possibilités : Centrer la matrice ou la centrer-réduire. On notera \bar{M} la matrice centrée (en colonne) et \hat{M} la matrice centrée-réduite (toujours en colonne). On utilisera la notation M pour désigner \bar{M} ou \hat{M} suivant les cas.

Projection

Soit $u \in \mathbb{R}^p$ un vecteur de norme 1. La projection de l'échantillon M sur u dans la base u s'écrit :

$$\pi_u(M) = M \cdot u \quad (1)$$

Preuve :

Prenons m_1 , le premier individu centrée réduit de M . m_1 peut s'écrire :

$$m_1 = \text{Vect}(u) + \text{Vect}(u)^\perp$$

$$m_1 = \alpha_1 u + \text{Vect}(u)^\perp$$

On a donc noté α_1 la coordonnée du projeté de m_1 sur u . On a :

$$\langle m_1, u \rangle = \langle \alpha_1 u + \text{Vect}(u)^\perp, u \rangle$$

$$\langle m_1, u \rangle = \langle \alpha_1 u, u \rangle$$

$$\langle m_1, u \rangle = \alpha_1 \langle u, u \rangle$$

$$\langle m_1, u \rangle = \alpha_1 \|u\|^2 \cos(0) = \alpha_1$$

On retrouve donc bien α_1 la coordonnée du projeté de m_1 sur u dans la "base" u . Or $M \cdot u$ n'est rien d'autre que le produit scalaire des individus m_i avec u .

Inertie/Variance du projeté

Soit X un nuage de points centré en colonne de n individus et p variables (X_1, \dots, X_p) . La matrice de covariance des (X_1, \dots, X_p) s'écrit :

$$Cov(X) = \frac{1}{p} X^T \cdot X$$

En partant de (1), la variance empirique de $\pi_u(M)$ s'écrit (on écrira Var car $\pi_u(M) \in \mathbb{R}^n$ et donc sa matrice de covariance n'est autre chose qu'un réel) :

$$Var(\pi_u(M)) = \frac{1}{p} u^T \cdot M^T \cdot M \cdot u$$

$$Var(\pi_u(M)) = u^T \frac{1}{p} M^T \cdot M \cdot u$$

$$Var(\pi_u(M)) = u^T \cdot cov(M) \cdot u \quad (2)$$

En vertu du théorème spectral, $C = cov(M) \in \mathcal{M}^{p \times p}$ étant une matrice carré symétrique, elle est diagonalisable dans une base de vecteurs propre orthonormée qu'on notera $v := (v_1, \dots, v_n)$.

Notons P la matrice de changement de base constituée des vecteurs propres v_i et

$\Delta = Diag(\lambda_1, \dots, \lambda_n)$ la matrice diagonale associée.

Considérons que $\Delta = Diag(\lambda_1, \dots, \lambda_n)$ ait les valeurs de sa diagonale rangées en ordre décroissant.

Alors le vecteur u qui maximise $Var(\pi_u(M))$ est v_1 le vecteur propre de $C = cov(M)$ associé à la première valeur propre λ_1 de Δ (c'est-à-dire la plus grande).

Preuve : Le but est de montrer que v_1 maximise (2). u peut s'écrire comme une combinaison linéaire des vecteurs propres v_i de $C = cov(M)$ qui forment une base orthonormée de \mathbb{R}^n . On peut donc écrire :

$$u = \alpha_1 v_1 + \dots + \alpha_n v_n$$

On a :

$$Cu = C(\alpha_1 v_1 + \dots + \alpha_n v_n)$$

$$Cu = \alpha_1 C v_1 + \dots + \alpha_n C v_n$$

Les v_i étant les vecteurs propres de C associés aux valeurs propres λ_i on a par définition :

$$Cu = \alpha_1 \lambda_1 v_1 + \dots + \alpha_n \lambda_n v_n$$

Pour continuer le développement de (2) on peut donc écrire :

$$u^T Cu = (\alpha_1 v_1 + \dots + \alpha_n v_n)^T \cdot (\alpha_1 \lambda_1 v_1 + \dots + \alpha_n \lambda_n v_n)$$

$$u^T Cu = (\alpha_1 v_1^T + \dots + \alpha_n v_n^T) \cdot (\alpha_1 \lambda_1 v_1 + \dots + \alpha_n \lambda_n v_n)$$

$$u^T C u = \sum_{i=1}^n \alpha_i^2 \lambda_i v_i^T \cdot v_i + \sum_{i \neq j}^n \alpha_i \alpha_j \lambda_j v_i^T \cdot v_j \quad (3)$$

Les vecteurs v_i formant une base orthonormée on a $v_i^T \cdot v_i = 1$ et pour $i \neq j$, $v_i^T \cdot v_j = 0$. (3) se ramène donc à :

$$u^T C u = \sum_{i=1}^n \alpha_i^2 \lambda_i \quad (4)$$

De plus, comme $u = \alpha_1 v_1 + \dots + \alpha_n v_n$, d'après pythagore (les vecteurs v_i formant une base orthonormée) on a :

$$\begin{aligned} \|\alpha_1 v_1\|^2 + \dots + \|\alpha_n v_n\|^2 &= \|\alpha_1 v_1 + \dots + \alpha_n v_n\|^2 \\ \alpha_1^2 + \dots + \alpha_n^2 &= \|u\|^2 \\ \alpha_1^2 + \dots + \alpha_n^2 &= 1 \end{aligned}$$

Et ainsi en repartant de (4) :

$$\begin{aligned} u^T C u &= \sum_{i=1}^n \alpha_i^2 \lambda_i \leq \sum_{i=1}^n \alpha_i^2 \max(\lambda_i) \\ u^T C u &\leq \lambda_1 \sum_{i=1}^n \alpha_i^2 \\ u^T C u &\leq \lambda_1 \end{aligned}$$

Donc $u^T C u$ est bornée par λ_1 et cette valeur est atteinte pour $\alpha_1 = 1$, $\alpha_2 = \dots = \alpha_n = 0$, c'est à dire $u = v_1$. CQFD. La valeur λ_1 est donc la variance empirique sur le premier axe de l'ACP.

De la même façon, on continue la recherche du deuxième axe de projection w sur le même principe en imposant qu'il soit orthogonal à u . On aboutit de la même façon à l'axe v_2 avec pour variance empirique λ_2 , etc. Ainsi la variance expliquée par le k -ième vecteur propre vaut λ_k .

A noter que si $n < p$, les n individus peuvent être résumés dans une base de n vecteurs maximum, il n'existera pas plus de n axes principaux. (la matrice $cov(M)$ aura des valeurs nulles dans sa diagonale. Par ex si $n = 1$, toutes ses valeurs propres sont nulles car la matrice est nulle (M est nulle par centrage des colonnes)).

Finalement, la question de l'ACP se ramène à un problème de diagonalisation de la matrice de covariance du nuage de point transformé (centré ou centré-réduit).

Algo:

Input :

- $X \in \mathcal{M}^{n \times p}$ matrice des observations avec n individus et p variables.
- k le nombre de composante principales qu'on souhaite calculer
- critère de transformation de X : "centré" ou "centré-réduit"

1 - On calcul $X_c = X - \bar{X}$ ou $X_{CR} = D_{1/\sigma}(X - \bar{X})$ avec D la matrice diagonale des inverses des écarts-types.

2 - On calcul la matrice de covariance (ou corrélation si X a été centré réduit) : $M = 1/p (X_c^T X_c)$ ou $M = 1/p (X_{CR}^T X_{CR})$ selon le critère de transformation choisi.

3 - On diagonalise M (par exemple avec le package simpy), i.e on détermine la matrice orthonormée des vecteurs propre P associé à Δ avec les valeurs propres ordonnées dans l'ordre décroissant.

4 - On conserve les k premiers vecteurs propres associés à Δ , qu'on notera en matrice colonne P_k . On les appelle les axes principaux de l'ACP.

5 - Le nuage X projeté et écrit en coordonnées dans la base des k premiers vecteurs propres est : $\pi(X)_{P_k} = X \cdot P_k$

6 - On pourra calculer la matrice des corrélation entre les k nouveaux axes propres et les anciennes variables (X_1, \dots, X_p) du nuage X pour voir quelle variable "contribue" à quel axe plus ou moins.

7 - On pourra aussi calculé le pourcentage d'inertie expliqué de chaque axe propre par $\lambda_j / \sum_{i=1}^p \lambda_i$

ainsi que le pourcentage de "conservation" du nouveau nuage $\pi(X)_{P_k}$ par $\sum_{i=1}^k \lambda_i / \sum_{i=1}^p \lambda_i$

Résumé : Les axes de projection de l'ACP sont les vecteurs propres de la matrice de covariance du nuage de point transformé (centré ou centré-réduit), notée $cov(M)$, associé aux valeurs propres dans l'ordre décroissant. (le premier axe de projection est le vecteur propre associé à la plus grande valeur propre, qui vaut par ailleurs l'inertie du nuage projeté sur cet axe). On a aussi vu que les nouvelles coordonnées du nuage sur un de ces axes (et dans cette nouvelle base) était :

$$\pi_{v_i}(M) = M \cdot v_i$$

Donc en notant P_k la matrice constitué en colonne de nos k premiers axes propres, on a comme nouvelle coordonnées dans la base de ces axes :

$$\pi(M) = MP_k \quad \text{avec } MP_k \in \mathcal{M}^{n \times k}$$