ACP cheat sheet

On considère la matrice X du nuage de point avec n individus et p variables, donc $X \in M^{n*p}$.
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 $\overline{X} = (\overline{X_1}, ..., \overline{X_p})$ la matrice X centrée et $\widehat{X}$la matrice X centrée-réduite (toujours en colonne). On utilisera la notation M pour désigner $\overline{X}$ ou $\widehat{X}$ suivant le cas choisi. Si les variables ne sont pas à la même échelle (par ex des grammes avec des distances en mètre), il faut obligatoirement réduire le nuage de point.

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 :

$$\pi_u(M) = M . u \tag{1}$$

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

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

On a donc noté $\alpha_1$la coordonnée du projeté de $m_1$sur u. On a :

$$\begin{aligned} < m_1, u > &= < \alpha_1 u + Vect(u)^{\perp}, u > \\ < m_1, u > &= < \alpha_1 u , u > \\ < m_1, u > &= \alpha_1 < u , u > \end{aligned}$$

$$< m_1, u > \ = \alpha_1 ||u||^2 cos(0) \ = \ \alpha_1$$

On retrouve donc bien α_1 la coordonnée du projeté de m_1 sur u dans la "base" u. Or $M . u$ 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 $(X_1, ..., X_p)$. La matrice de covariance des $(X_1, ..., X_p)$ s'écrit :

$$Cov(X) \ = \frac{1}{n} X^T . X$$

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) :

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

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

$$Var(\pi_u(M)) \ = \ u^T . cov(M) \ . u \qquad (2)$$

En vertu du théorème spectral, $C \ = \ cov(M) \ \in \mathcal{M}^{p*p}$ étant une matrice carré symétrique, elle est diagonalisable dans une base de vecteurs propre orthonormée qu'on notera $v := (v_1, ..., v_p)$. Notons P la matrice de changement de base constituée des vecteurs propres v_i et $\Delta \ = \ Diag(\lambda_1, ..., \lambda_n)$ la matrice diagonale associée. On a donc :

$$C \ = \ cov(M) \ = \ P\Delta P^{-1} \qquad (3)$$

Considérons que $\Delta \ = \ Diag(\lambda_1, ..., \lambda_p)$ 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 :

$$u = \alpha_1 v_1 + ... + \alpha_p v_p$$

On a :

$$Cu = C(\alpha_1 v_1 + ... + \alpha_p v_p)$$
$$Cu = \alpha_1 C v_1 + ... + \alpha_p C v_p$$

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

$$Cu = \alpha_1 \lambda_1 v_1 + ... + \alpha_p \lambda_p v_p$$

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

$$u^T C u = (\alpha_1 v_1 + ... + \alpha_p v_p)^T . (\alpha_1 \lambda_1 v_1 + ... + \alpha_p \lambda_p v_p)$$
$$u^T C u = (\alpha_1 v_1^T + ... + \alpha_p v_p^T) . (\alpha_1 \lambda_1 v_1 + ... + \alpha_p \lambda_p v_p)$$

$$u^T C u = \sum_{i=1}^{p} \alpha_i^2 \lambda_i v_i^T . v_i + \sum_{i \neq j}^{p} \alpha_i \alpha_j \lambda_j v_i^T . v_j \quad (4)$$

Les vecteurs v_i formant une base orthonormée on a $v_i^T . v_i = 1$ et pour $i \neq j$, $v_i^T . v_j = 0$. (4) se ramène donc à :

$$u^T C u = \sum_{i=1}^{p} \alpha_i^2 \lambda_i \quad (5)$$

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

$$||\alpha_1 v_1||^2 + ... + ||\alpha_p v_p||^2 = ||\alpha_1 v_1 + ... + \alpha_p v_p||^2$$
$$\alpha_1^2 + ... + \alpha_p^2 = || u||^2$$
$$\alpha_1^2 + ... + \alpha_p^2 = 1$$

Et ainsi en repartant de (5) :

$$u^T C u = \sum_{i=1}^{p} \alpha_i^2 \lambda_i \leq \sum_{i=1}^{p} \alpha_i^2 max(\lambda_i)$$
$$u^T C u \leq \lambda_1 \sum_{i=1}^{p} \alpha_i^2$$
$$u^T C u \leq \lambda_1$$

Donc $u^T Cu$ est bornée par λ_1 et cette valeur est atteinte pour $\alpha_1 = 1,\ \alpha_2 =... = \alpha_p = 0$, 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 :

- $X \in \mathcal{M}^{n*p}$ 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 $X_C = X - \overline{X}$ ou $X_{CR} = D_{1/\sigma}(X - \overline{X})$ 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 : $\pi(X)_{P_k} = X.P_k$

6 - On pourra calculer la matrice des corrélation entre les k nouveaux axes propres et les anciennes variables $(X_1, ..., X_p)$ du nuage Xpour 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 :

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

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

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

P étant une matrice de passage dans la base des axes propres orthonormés, on sait que les propriétés de changement de base nous disent que :

Soit x' exprimé dans la nouvelle base exprimé par P et x exprimé dans la base canonique, alors on a :

$$Px' = x$$
$$x'^T P^T = x^T$$
$$x'^T P^{-1} = x^T$$
$$x'^T = x^T P$$

Or MP n’est autre que cela. On aurait aussi pu dire que cela découle de l’extension de la démo de (1).

Représentation et interprétation:

On représentera couramment les individus projetés sur les axes propres 1, 2 et 3, cad en notant le nuage des individus écrit dans sa nouvelle base des axes propres :

$$X_{pca} = MP = (F_1, F_2, \ldots)$$

Avec F_i la i-ème composante principale. On représentera généralement (F_1, F_2) ainsi que (F_1, F_3) dans deux plans séparés.

Pour ce qui est de l'interprétation, on aimerait ensuite dire quelles variables participent aux composantes principales. Par exemple, pour la première composante principale, on a en développant $F_1 = MP_1$ et en notant $P_1 = u_1 = (u_{1,1}, \ldots, u_{1,p})$:

$$F_1 = u_{1,1}M_1 + \ldots + u_{1,p}M_p$$

M étant notre nuage de point initial X centré ou centré-réduit, cela nous permet de voir quelles variables interviennent dans nos composantes principales. On peut représenter ces coordonnées dans un cercle de rayon 1 (puisque les coordonnées de u_1 sont comprises entre 0 et 1. Cette représentation est souvent noté **factor map** en anglais.

On pourra également représenter le **cercle des corrélation** entre les variables d'origine et les composantes principales.

Lien entre les axes propres et les corrélation des composantes principales avec les variables d'origine :

On notera $\overline{X}$ la matrice X centrée et $\widehat{X}$ la matrice centrée-réduite.

En notant D_{1/σ_X} la matrice diagonale avec pour valeur en diagonale $(1/\sigma_{X_1}, \ldots, 1/\sigma_{X_p})$ les inverses des écarts type des colonnes $(X_1, \ldots, X_p)$, on a :

$$\widehat{X} = \overline{X} D_{1/\sigma_X}$$

Cas 1 où le nuage M est centré-réduit, i.e $M = \widehat{X}$:

On a :

$$corr(F, X) = 1/n\ \widehat{F}^T \widehat{X}$$
$$corr(F, X) = 1/n\ (\overline{F}D_{1/\sigma_F})^T \widehat{X} \qquad (5)$$

Prop : $\overline{F} = F$ car $F = MP$et M est centrée. cela se vérifie en calculant le centre de gravité de F qu'on écrit $g = F^T D_{1/n} I$, I étant le vecteur unitaire dans R^n.

On a $g = P^T M^T D_{1/n} I$, or comme M est centrée on a $M^T D_{1/n} I = 0$ et donc $g = 0$. En repartant de (5) on a donc :

$$corr(F, X) = 1/n \, (F D_{1/\sigma_F})^T \widehat{X}$$

$$corr(F, X) = 1/n \, (MPD_{1/\sigma_F})^T \widehat{X}$$

$$corr(F, X) = 1/n \, D_{1/\sigma_F} P^T \widehat{X}^T \widehat{X}$$

$$corr(F, X) = D_{1/\sigma_F} P^T \widehat{X}^T \, 1/n \, \widehat{X}$$

$$corr(F, X) = D_{1/\sigma_F} P^T cov(M)$$

Or on rappel (voir (3)) que $cov(M) = P \Delta P^{-1}$. On a donc :

$$corr(F, X) = D_{1/\sigma_F} P^T P \Delta P^{-1}$$

$$corr(F, X) = D_{1/\sigma_F} \Delta P^{-1}$$

Or on rappel que $\Delta = diag(\lambda_1, ..., \lambda_p)$ avec $\lambda_i = Var(\pi_{u_i}(M)) = Var(MP_i) = Var(F_i) = \sigma_i^2$. Ainsi, $\Delta = diag(\sigma_1^2, ..., \sigma_p^2) = D_{\sigma_F^2}$, et donc :

$$corr(F, X) = D_{1/\sigma_F} D_{\sigma_F^2} P^{-1}$$

$$corr(F, X) = D_{\sigma_F} P^{-1}$$

Ainsi,

$$corr(F_i, X_j) = \sigma_{F_i} u_{j,i} \quad \text{avec } P = mat(u_{i,j})$$

Dans le cas 2 où le nuage M est centré, i.e $M = \overline{X}$:
On trouve de la même façon :

$$corr(F, X) = D_{\sigma_F} P^{-1} D_{1/\sigma_X}$$

et donc

$$corr(F_i, X_j) = \sigma_{F_i} u_{j,i} \, 1/\sigma_{X_j}$$