

ASE3 Cheat sheet - Analyse en composantes principales

Rappels utiles :

1. Multiplication de matrice

- La matrice résultante doit être : $A_{n*m} * B_{m*p} \Rightarrow C_{n*p}$
- Trois propriétés importantes :
 - $AB \neq BA$
 - $I_n A = A I_n = A$
 - Si A et B sont carrées de taille n , alors
 $AB = I_n \Rightarrow BA = I_n$ et $B = A^{-1}$ (inverse de A)

2. Calcul de la trace

- Somme des termes de la diagonale d'une matrice carrée
- Notée sous la forme $Tr(A)$ ou $Trace(A)$

3. Représentation par transposition

- Échange des lignes et des colonnes d'une matrice
- Notée sous la forme A' ou A^T

Matrice des poids p_i

Nous associons à chaque individu un poids $p_i \geq 0$ qui correspond à la probabilité de choisir un individu. La somme des poids de la matrice est égale à 1.

Formule :

$$\sum_{i=1}^n p_i = p_1 + p_2 + p_3 + \dots = 1$$

$$D = \begin{pmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_n \end{pmatrix}$$

Cas uniforme : Si tous les individus ont le même poids, alors $p_i = \frac{1}{n} \Rightarrow D = \frac{1}{n} I_n$, où I_n est la matrice identité.

Moyenne des variables $\bar{X}^{(j)}$ et centre de gravité g

Définition : La moyenne d'une colonne $\bar{X}^{(j)}$ (avec j le numéro de colonne et $j \in [1, p]$) s'obtient en additionnant chaque valeur de colonne et en multipliant l'ensemble par son poids p_i :

$$\bar{X}^{(j)} = \sum_{i=1}^n P_i X_i^{(j)}$$

Définition : Le centre de gravité, représenté par le vecteur g des moyennes arithmétiques de chaque variable $X^{(j)}$, est défini par $g = (\bar{X}^{(1)}, \bar{X}^{(2)}, \dots, \bar{X}^{(j)})$.

Matrice des données centrées Y

Définition : La matrice Y s'obtient en soustrayant chaque moyenne $\bar{X}^{(j)}$ de la matrice initiale X , c'est-à-dire :

$$Y_i^{(j)} = X_i^{(j)} - \bar{X}^{(j)}, \forall j \in [1, p], \forall i \in [1, n]$$

Matrice de variance-covariance V

Définition : La matrice de variance-covariance V (ou *var-covariance*) est une matrice carrée de dimension p représentée sous la forme suivante :

$$V = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \dots & \sigma_{1,p} \\ \sigma_{2,1} & \sigma_{2,2} & \dots & \sigma_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p,1} & \sigma_{p,2} & \dots & \sigma_p^2 \end{pmatrix}$$

Formule : Cette matrice s'obtient avec la formule

$$V = Y^T * D * Y$$

Symétrie : La matrice V est symétrique, donc $V^T = V$.

Diagonalisation d'une matrice

Rappel : Soit une matrice A , diagonaliser cette matrice revient à chercher une matrice diagonale D ainsi qu'une matrice inversible P telle que :

$$A = P * D * P^{-1}$$

Dans le cours d'ASE3, diagonaliser revient à calculer les valeurs propres de la matrice afin d'en déterminer par la suite ses composantes et facteurs principaux.

Valeurs propres

Définition (rappel) : Soit une matrice A , on appelle polynôme caractéristique de A , noté en général P_A , le polynôme défini par

$$P_A(\lambda) = \det(A - \lambda I_n)$$

En calculant ce polynôme, nous pouvons trouver les **valeurs propres** de la matrice A .

Pourcentage d'inertie

Définition : L'inertie totale mesure l'étalement du nuage de points d'une matrice. L'inertie de l'axe α est calculée divisant sa valeur propre λ_α par la somme des valeurs propres des différents axes.

Formule :

$$\text{Inertie de l'axe } \alpha = \frac{\lambda_\alpha}{\sum_1^n \lambda_n}$$

⚠ La résultat doit être présenté sous forme de pourcentage.

Facteurs principaux

Définition : Les facteurs principaux sont les vecteurs propres associés aux plus grandes valeurs propres.

Formule :

Pour trouver les vecteurs propres de V , nous posons $E(\lambda_\alpha) = \text{Ker}(V - \lambda_\alpha I_n)$

Méthode :

$$\forall u = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in E(\lambda_\alpha) \Leftrightarrow (V - \lambda_\alpha I_n) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \vec{0}$$

Après calcul par intégration linéaire, on trouve :

$$E(\lambda_\alpha) = \text{Vect}\left(\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}\right)$$

(où α , β et γ sont les solutions de l'équation linéaire trouvée pour i).

Pour calculer u , on pose :

$$u^{(i)} = \frac{1}{\sqrt{\alpha^2 + \beta^2 + \gamma^2}} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$$

Remarque : $E(\lambda_\alpha)$ est une droite vectorielle et u est normé.

Composantes principales

La composante principale C est définie par :

$$C^{(i)} = Y * u^{(i)}$$

Remarque : Les composantes principales contiennent les projections d'individus sur les axes factoriels.

Coefficients de corrélation linéaire

Définition : La méthode la plus naturelle pour donner une signification à une composante principale $C^{(i)}$ est de la relier aux variables $X^{(j)}$ (variables initiales) en calculant les coefficients de corrélation linéaire :

$$p(X^{(j)}, C^{(i)})$$

Formule :

$$p(X^{(j)}, C^{(i)}) = \frac{Cov(X^{(j)}, C^{(i)})}{\sigma_{X^{(j)}} \sigma_{C^{(i)}}}$$

Remarque :

$$Cov(X^{(j)}, C^{(i)}) = \langle y^{(j)}, C^{(i)} \rangle$$

où $y^{(j)}$ est une variable centrée.