Calcul Scientifique et Apprentissage Automatique

Algèbre Linéaire

Elana Courtines

septembre 2022

Sandrine Mouysset (CM1 CM2): sandrine.mouysset@irit.fr
Thomas Pellegrini : thomas.pellegrini@irit.fr

1 Introduction & Définitions

L'expression **Analyse de données** recouvre les techniques ayant pour objectif la description statistique des grands tableaux n $\times$ p de données.

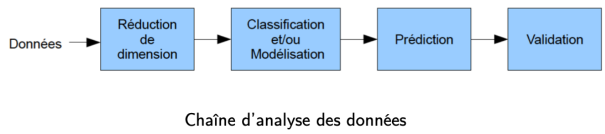


Figure 1: Récupéré sur le pdf du cours

Nature des données ?

- Qualitative :
  - ordinale : qui peut être ordonné, par exemple un ensemble de notes à un examen ;
  - nominale : qui n'est a priori pas ordonné, par exemple un ensemble de types de bois ;
- Quantitative ;
- Temporelle.

Définition **tableau de contingence** : Un tableau de contingence est une méthode de représentation de données issues d'un comptage permettant d'estimer la dépendance entre deux caractères.

Définition **individu moyen** noté g : l'individu moyen est un vecteur (de taille p, la quantité de variables présentes) contenant les moyennes de toutes les valeurs disponibles pour chaque variable. Par exemple si pour 3 caractéristiques A, B et C on a 10 échantillons, alors $g = [\overline{a}, \overline{b}, \overline{c}]$ où $\overline{a} = \frac{1}{10}\sum_{i=0}^{9} a_i$, etc.

Définition **tableau centré** noté X_C (où X est le tableau regroupant toutes les données) : est le tableau X auquel on a soustrait à chaque valeur la moyenne de la variables correspondante. Par exemple on soustrait à toutes les valeurs de la colonne de la variable A la valeur de $\overline{a}$.

Définition **matrice de (variance-)covariance** notée Σ : est une matrice carrée symétrique de taille $p{*}p$ mettant en évidence les variances de chaque variables sur la diagonal $(Var(A), Var(B), ...)$ et les covariances de chaque couple de variables $(Cov(A,B),\ Cov(A,C),\ etc...)$. Le calcul de cette matrice est le suivant : (avec 3 variables)

$$\Sigma = \frac{1}{n} * X_C^T * X_C = \begin{bmatrix} Var(A) & Cov(A,B) & Cov(A,C) \\ Cov(B,A) & Var(B) & Cov(B,C) \\ Cov(C,A) & Cov(C,B) & Var(C) \end{bmatrix}, \text{ où } X_C^T \text{ est la transposée de } X_C.$$

Note : $Cov(A,B) = Cov(B,A),\ \forall A, B \in variables.$

Définition **corrélation** : la corrélation est une valeur (sans unité) représentant un degré de correlation entre deux variables, c'est à dire "à quel point les variables sont-elles liées". Cette valeur est toujours comprise entre 0 et 1 et vaut :

$$-1 \leq Corr(A,B) = \frac{Cov(A,B)}{\sqrt{Var(A) * Var(B)}} \leq 1$$

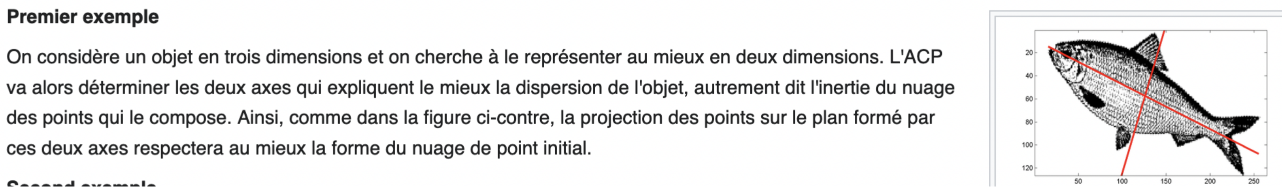
Signification de la valeur :

- $Corr(A,B) = 0$, les variables sont décorrélées, indépendantes c'est-à-dire étant donné A , on ne peut rien dire prédire sur la valeur de B ;
- $Corr(A,B) = 1$, dépendance linéaire positive de A et B ;
- $Corr(A,B) = -1$, dépendance linéaire négative de A et B.

2 Analyse en Composantes Principales (ACP)

Définition de **ACP** : l'Analyse en Composantes Principales est une méthode de réduction de dimension qui permet de transformer des variables très corrélées en variables non corrélées en changeant le plan dans lequel les variables se situent originellement. Pour ce faire, on va chercher des "composantes principales" (ou "axes principaux"), c'est à dire les nouveaux axes de notre plan (de préférence le plus orthogonaux entre eux possible = non corrélés).

Notes : le plan généré par les composantes principales peut être de une(1) dimension inférieure au plan de départ (au choix, mais cela génère plus de perte d'information). Exemple :


Premier exemple

On considère un objet en trois dimensions et on cherche à le représenter au mieux en deux dimensions. L'ACP va alors déterminer les deux axes qui expliquent le mieux la dispersion de l'objet, autrement dit l'inertie du nuage des points qui le compose. Ainsi, comme dans la figure ci-contre, la projection des points sur le plan formé par ces deux axes respectera au mieux la forme du nuage de point initial.

Figure 2: `https://fr.wikipedia.org/wiki/Analyse_en_composantes_principales`

Détermination des nouveaux axes principaux :

1^{er} axe principal : généré par un vecteur propre ($\overrightarrow{vp}$) de la matrice de variance-covariance Σ associé à la plus grande valeur propre (vp) de Σ.

2^{eme} axe principal : généré par un $\overrightarrow{vp}$ de Σ associé à la 2^{eme} plus grande vp de Σ.

3^{eme} axe principal : ...

Projection des points sur les nouveaux axes : ...

Rappel sur qu'est la trace d'une matrice carrée $A \in \mathbb{R}^{n*n}$: $Tr(A) = \sum_{i=1}^{n} a_{ii}$, *ou la somme des éléments de la diagonale.*

Calcul des valeurs propres, et détermination des vecteurs propres : les valeurs propres (notés $\lambda_1, \lambda_2, ...$) d'une matrice carrée sont des scalaires caractérisant la matrice. Ceux-ci se trouvent par calcul des racines du polynôme caractéristique de la matrice.

Pour une matrice carré A, il suffit donc de résoudre le polynôme (χ - Chi) suivant :

$$\chi_A(\lambda) = det(A - \lambda * Id)$$

On obtient alors l'ensemble des valeurs propres (racines) $\{\lambda_1, \lambda_2, ...\}$ de la matrice A.

Pour ensuite obtenir les vecteurs propres, il faut partir d'une nouveau vecteur de taille n x 1, noté X, et résoudre le système suivant :

$$A * X = \lambda_1 * X$$

Ce système a une infinité de solutions, mais il est préférable de ne conserver que le vecteur le plus simple (petit et entier). Par exemple avec une matrice 2 x 2, si le résultat est $2x = 7y$, alors l'idéal serait de prendre le vecteur $[x = 7, y = 2]$.

Ce vecteur propre constitue une composante principale (une axe principal) pour l'ACT (opération à réaliser pour autant de racines que désiré).

Contraste :

$Tr(\Sigma) = Var(h) + Var(H) = \lambda_1 + \lambda_2$
Perte d'information (contraste) de la réduction $2D \rightarrow 1D$

$C = \frac{\lambda_1}{\lambda_1 + \lambda_2}$

Application (diapo page 25/28) :

Rappel des étapes du calcul d'une corrélation :

1. Calculer le couple $g = \{\bar{\alpha}, \bar{\beta}\}$
2. Déterminer la matrice centrée X_C : X à qui on soustrait les valeurs moyennes
3. Déterminer la matrice de variance-covariance $\Sigma = \frac{1}{n} * X_C^T * X_C = \begin{bmatrix} Var(\alpha) & Cov(\alpha,\beta) \\ Cov(\alpha,\beta) & Var(\beta) \end{bmatrix}$
4. Déterminer $corr(\alpha,\beta) = \frac{Cov(\alpha,\beta)}{\sqrt{Var(\alpha)*Var(\beta)}}$

Données :

On a : $\begin{array}{c} \begin{array}{cc} \alpha & \beta \end{array} \\ \begin{bmatrix} 0 & 2 \\ \text{-2} & \text{-1} \\ 1 & 0 \\ 1 & \text{-1} \end{bmatrix} \end{array}$

D'où : $g = \{\bar{\alpha}, \bar{\beta}\} = \{0, 0\}$

On en déduit $X_C = \begin{bmatrix} 0 & 2 \\ -2 & -1 \\ 1 & 0 \\ 1 & -1 \end{bmatrix}$ et $X_C^T = \begin{bmatrix} 0 & -2 & 1 & 1 \\ 2 & -1 & 0 & -1 \end{bmatrix}$

D'où :

$$\Sigma = \frac{1}{4} * \begin{bmatrix} 0 & -2 & 1 & 1 \\ 2 & -1 & 0 & -1 \end{bmatrix} * \begin{bmatrix} 0 & 2 \\ -2 & -1 \\ 1 & 0 \\ 1 & -1 \end{bmatrix}$$

$$\Sigma = \frac{1}{4} * \begin{bmatrix} 0+4+1+1 & 0+2+0-1 \\ 0+2+0-1 & 4+1+0+1 \end{bmatrix}$$

$$\Sigma = \frac{1}{4} * \begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 6/4 & 1/4 \\ 1/4 & 6/4 \end{bmatrix}$$

On en déduit : $Corr(\alpha,\beta) = \dfrac{Cov(\alpha,\beta)}{\sqrt{Var(\alpha)*Var(\beta)}} = \dfrac{1/4}{\sqrt{6/4*6/4}} = 1/6$

Avec une corrélation aussi faible, on peut dire que ces capteurs sont indépendant l'un de l'autre.

Calcul du 1^{er} axe principal :

Rappel des étapes du calcul d'un axe principal :

1. Calculer $\chi_\Sigma(\lambda) = det(\Sigma - \lambda * I_n)$, où I_n est la matrice identité de taille n (ici $n = 2$) ;
2. Déterminer les racines de $\chi_\Sigma(\lambda)$ (autrement dit, les valeurs propres) ;
3. Déclarer une matrice X de taille $[1, n]$ pour déterminer le vecteur propre associé à la valeur propre.

Pour simplifier les fractions, nous ferons les calculs sur 4Σ plutôt que Σ (cela éviter les fractions).

On a :

$$\begin{aligned}\chi_{4\Sigma}(\lambda) = det(4\Sigma - \lambda * I_n) &= \begin{vmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{vmatrix} \\ &= (6-\lambda)^2 - 1\end{aligned}$$

On résout trivialement le polynôme pour trouver :

$\lambda_1 = 7$, le plus grand vecteur propre, donc celui à prendre pour le 1^{er} axe principal ;

$\lambda_2 = 5$.

Soit $X = \begin{bmatrix} x \\ y \end{bmatrix}$ le $\overrightarrow{vp}$ de 4Σ associé à λ_1.

On à alors :

$$\Leftrightarrow 4\Sigma X = \lambda_1 X$$

$$\Leftrightarrow \begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix} = 7 * \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\Leftrightarrow \begin{cases} 6x + y = 7x \\ x + y = 7y \end{cases}$$

$$\Leftrightarrow x = y$$

Ainsi, le 1^{er} axe principal de cette analyse est défini par le vecteur $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

On aurait par exemple pu prendre $\begin{bmatrix} 3 \\ 3 \end{bmatrix}$, puisque le vecteur est acceptable tant que $x = y$. Il est cependant préférable de prendre un vecteur ayant les valeurs les plus simples à manipuler.