Algorithmes : Analyse Factorielle (AF) et

Analyse en Composantes Principales (ACP)

Analyse Factorielle

Input :

- $X_1, X_2, ..., X_n$ € IR^p
- $m_1, m_2, ..., m_n$ € IR

Output :

- $X_1, X_2, ..., X_n$ € IR^q (q < p)
- $m_1, m_2, ..., m_n$ € IR

Input → Output

n points dans un s.e.v ***S*** de dimension p → n points dans un s.e.v ***P*** de dimension q

➔ Perte d'information : à Minimiser

Distance$_S$ (Xi, Xj) < Distance$_P$ (Xi, Xj)

L'analyse factorielle consiste à trouver le s.e.v *P* qui minimise :

$$\sum_{i=1}^{n}\sum_{j=1}^{n} m_i \cdot m_j \cdot \left(\mathrm{d}^2(X_i, X_j) - \mathrm{d}^2(\mathrm{Proj}_P(X_i), \mathrm{Proj}_P(X_j))\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} m_i \cdot m_j \cdot \mathrm{d}^2(X_i, X_j) - \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n} m_i \cdot m_j \cdot \mathrm{d}^2(\mathrm{Proj}_P(X_i), \mathrm{Proj}_P(X_j))}_{\textbf{à maximiser}}$$

Pour trouver P, on doit chercher q axes Δ_i avec :

Δ_1 tel que Δ_1 passe par O et $I_E(\Delta_1)$ est maximale

Δ_2 tel que Δ_2 passe par O, $\Delta_1 \perp \Delta_2$ et $I_E(\Delta_2)$ est maximale

.

Δ_q tel que Δ_q passe par O, $\Delta_1 \perp \Delta_q$, $\Delta_2 \perp \Delta_q$ etc. et $I_E(\Delta_q)$ est maximale

Si la dimension de P n'est pas fixée, on fixe un seuil (c.à.d un pourcentage de l'inertie du nuage) et on cherche la valeur de q minimale tel que $I_E(P) \geq seuil$ (exemple seuil= 80% de $I(O)$).

➔ Les axes factoriels $\Delta_1 \dots \Delta_q$ sont les axes passant par l'origine et dont les vecteurs directeurs sont les q premiers vecteurs propres de la matrice V=X'.M.X

$$X = \begin{pmatrix} X_{11} & \dots & X_{1p} \\ \vdots & & \vdots \\ X_{i1} & \dots & X_{ip} \\ \vdots & & \vdots \\ X_{n1} & \dots & X_{np} \end{pmatrix} \begin{matrix} \leftarrow X_1 \\ \vdots \\ \leftarrow X_i \\ \vdots \\ \leftarrow X_n \end{matrix} \text{ de taille } (n,p) \quad \text{et} \quad M = \begin{pmatrix} m_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & m_n \end{pmatrix} \text{ de taille } (n,n)$$

➔ C'est-à-dire, V = A D A' (A'=transposé(A)) avec A la matrice des vecteurs propres (U_1, … , U_p) et D la matrice des valeurs propres ($\lambda_1, \dots, \lambda_p$) tel que $\lambda_1 > \lambda_2 > \dots > \lambda_p$

L'inertie expliquée par chaque axe factoriel Δ_t est égale à la valeur propre λ_t correspondante à son vecteur directeur (i.e. vecteur propre U_t)

void XN[n][q] ← AF_Cas1(void X[n][p], float M[n][n], int q)

BEGIN

V ← X'.M.X

//X'=transposé(X)

[v1…vp, λ1… λp] ← Diagonalise(V)

// V=P.A.P' avec P est la matrice des vecteurs propres v1 … vp et A la matrice des valeurs
// propres λ1 … λp (λ1> λ2> …> λp)

Axe-Fact1 ← Droite passant par l'origine et dont le vecteur directeur est v1 et son inertie expliqué est égal à λ1

Axe-Fact2 ← Droite passant par l'origine et dont le vecteur directeur est v2 et son inertie expliqué est égal à λ2

…

Axe-Factq ← Droite passant par l'origine et dont le vecteur directeur est vq et son inertie expliqué est égal à λq

Pour i=1 …n

    Pour k=1…q

    XN[i][k]=produit_scalaire(X[i][],vk)

    //le coordonnée de l'individu Xi sur Axe_Factk est égal à son produit scalaire avec le

    // vecteur directeur de cet axe (=kème vecteur propre vk)

    Finpour

Finpour

END

void XN[n][q] ← AF_Cas2(void X[n][p], float M[n][n], float seuil)

BEGIN

V ← X'.M.X

//X'=transposé(X)

[v1…vp, λ1… λp] ← Diagonalise(V)

// V=P.A.P' avec P est la matrice des vecteurs propres v1 … vp et A la matrice des valeurs
// propres λ1 … λp (λ1> λ2> …> λp)

Axe-Fact1 ← Droite passant par l'origine et dont le vecteur directeur est v1 et son inertie expliqué est égal à λ1

Si λ1/somme(λ1… λp) >= seuil) alors q=1

    Sinon Axe-Fact2 ← Droite passant par l'origine et dont le vecteur directeur est v2 et son inertie expliqué est égal à λ2

    Si (λ1+ λ2)/somme(λ1… λp) >= seuil) alors q=2

        Sinon Axe-Fact3 ← Droite passant par l'origine et dont le vecteur directeur est v3 et son inertie expliqué est égal à λ3

….

Pour i=1 …n

    Pour k=1…q

    XN[i][k]=produit_scalaire(X[i][],vk)

    //le coordonnée de l'individu Xi sur Axe_Factk est égal à son produit scalaire avec le

    // vecteur directeur de cet axe (=kème vecteur propre vk)

    Finpour

Finpour

END

Analyse en Composantes Principales

Input : $X_1, X_2, \ldots, X_n \in IR^p$ $(m_1=m_2= m_n = 1/n)$

Output : $X_1, X_2, \ldots, X_n \in IR^q$ $(q < p)$

Step 0 : Centrer et Réduire la population

$$Xij \rightarrow (Xij\text{-}moy(Att\_j))/ecart\_type(Att\_j))$$

Ainsi, moy(Att_j)=0

et ecart_type(Att_j)=1

- ➔ La matrice à diagonaliser V n'est autre que la matrice des coefficients des corrélations linéaires des p attributs (qui est symétrique avec des valeurs égales à 1 sur le diagonal)

En plus, puisque V = A D A', on a trace(V)=trace(D) $\rightarrow$ $p= \lambda_1+ \ldots + \lambda_p$

void XN[n][q] ← ACP_Cas1(float X[n][p], float M[n][n]=(1/n).1I$_{nxn}$, int q)

BEGIN

V ← Matrice des Coefficients des Corrélations Linéaires des p attributs

[v1…vp, λ1… λp] ← Diagonalise(V)

// V=P.A.P' avec P est la matrice des vecteurs propres v1 … vp et A la matrice des valeurs
// propres λ1 … λp (λ1> λ2> …> λp), sachant que somme(λ1,…, λp)=p

Axe-Fact1 ← Droite passant par l'origine et dont le vecteur directeur est v1 et son inertie expliqué est égal à λ1

Axe-Fact2 ← Droite passant par l'origine et dont le vecteur directeur est v2 et son inertie expliqué est égal à λ2

…

Axe-Factq ← Droite passant par l'origine et dont le vecteur directeur est vq et son inertie expliqué est égal à λq

Pour i=1 …n

    Pour k=1…q

    XN[i][k]=produit_scalaire(X[i][],vk)

    //le coordonnée de l'individu Xi sur Axe_Factk est égal à son produit scalaire avec le

    // vecteur directeur de cet axe (=kème vecteur propre vk)

    Finpour

Finpour

END

void XN[n][q] ← ACP_Cas2(float X[n][p], float M[n][n]=(1/n).1I$_{nxn}$, float seuil)

BEGIN

V ← Matrice des Coefficients des Corrélations Linéaires des p attributs

[v1…vp, λ1… λp] ← Diagonalise(V)

// V=P.A.P' avec P est la matrice des vecteurs propres v1 … vp et A la matrice des valeurs
// propres λ1 … λp (λ1> λ2> …> λp), sachant que somme(λ1,…, λp)=p

Axe-Fact1 ← Droite passant par l'origine et dont le vecteur directeur est v1 et son inertie expliqué est égal à λ1

Si λ1/p >= seuil) alors q=1

Sinon Axe-Fact2 ← Droite passant par l'origine et dont le vecteur directeur est v2 et son inertie expliqué est égal à λ2

Si (λ1+ λ2)/p >= seuil) alors q=2

Sinon Axe-Fact3 ← Droite passant par l'origine et dont le vecteur directeur est v3 et son inertie expliqué est égal à λ3

….

Pour i=1 …n

Pour k=1…q

XN[i][k]=produit_scalaire(X[i][],vk)

//le coordonnée de l'individu Xi sur Axe_Factk est égal à son produit scalaire avec le

// vecteur directeur de cet axe (=kème vecteur propre vk)

Finpour

Finpour

END