TD1 - Tableaux et Espaces (détaillé : produits scalaires matriciels)

Données initiales

Matrice des données :

$$X = \begin{pmatrix} 3 & -1 \\ -1 & 2 \\ 4 & 2 \end{pmatrix}$$

Donc

$$x_1 = (3, -1), \quad x_2 = (-1, 2), \quad x_3 = (4, 2)$$

1. Caractéristiques des données

1-1) Dimensions :

$$n = 3 \text{ (3 lignes)}$$

$$p = 2 \text{ (2 colonnes)}$$

1-2) Nuage de points:

$$x_1(3, -1)$$

$$x_2(-1, 2)$$

$$x_3(4, 2)$$

2. Espace des individus

2-1) Produit scalaire : Produit scalaire euclidien usuel dans

$$\mathbb{R}^2 : \langle x, y \rangle = x^T y$$

2-2) Norme du 3ème individu x_3 = (4, 2) :

Par produit matriciel :

$$|x_3|^2 = x_3 x_3^T = \begin{pmatrix} 4 & 2 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \end{pmatrix} = 16 + 4 = 20$$

Donc :

$$|x_3| = \sqrt{20} = 2\sqrt{5} \approx 4,4721$$

2-3) Généralisation matricielle : Pour la matrice des produits scalaires entre individus est :

$$G = XX^T = \begin{pmatrix} 3 & -1 \\ -1 & 2 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} 3 & -1 & 4 \\ -1 & 2 & 2 \end{pmatrix}$$

$$G = \begin{pmatrix} 9+1 & -3-2 & 12-2 \\ -3-2 & 1+4 & -4+4 \\ 12-2 & -4+4 & 16+4 \end{pmatrix} = \begin{pmatrix} 10 & -5 & 10 \\ -5 & 5 & 0 \\ 10 & 0 & 20 \end{pmatrix}$$

Les normes sont

$$|x_1|^2 = \begin{pmatrix} 3 & -1 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \end{pmatrix} = 10 \quad |x_1| = \sqrt{10}$$

$$|x_2|^2 = \begin{pmatrix} -1 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \end{pmatrix} = 5 \quad |x_2| = \sqrt{5}$$

$$|x_3|^2 = \begin{pmatrix} 4 & 2 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \end{pmatrix} = 20 \quad |x_3| = 2\sqrt{5}$$

2-4) Vecteur unitaire colinéaire à x_3 :

$$u = \frac{x_3}{|x_3|} = \frac{1}{2\sqrt{5}} \begin{pmatrix} 4 & 2 \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

2-5) Projection de x_2 = (-1, 2)sur l'axe dirigé par u :

$$\text{proj}_u(x_2) = (x_2 u^T), u$$

Calcul du produit scalaire :

$$x_2 u^T = \begin{pmatrix} -1 & 2 \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} = \frac{-2+2}{\sqrt{5}} = 0$$

Donc :

$$\text{proj}_u(x_2) = 0 \cdot u = \begin{pmatrix} 0 \; 0 \end{pmatrix}$$

2-6) Angle entre x_2 et x_3 :

Par produit matriciel :

Formule

$$x_2^T x_3 = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \begin{pmatrix} 4 & 2 \end{pmatrix} = -4 + 4 = 0$$

Les vecteurs sont orthogonaux ⇒ angle = 90°.

2-7) Métrique utilisée : La métrique euclidienne standard

$$M = I_2.$$

2-8) Distance euclidienne entre x_1 = (3, -1) et x_2 = (-1, 2) :

$$x_1 - x_2 = \begin{pmatrix}3 & -1\end{pmatrix} - \begin{pmatrix}-1 & 2\end{pmatrix} = \begin{pmatrix}4 & -3\end{pmatrix}$$

$$|x_1 - x_2|^2 = (x_1 - x_2)^T (x_1 - x_2) = \begin{pmatrix}4 & -3\end{pmatrix} \begin{pmatrix}4 & -3\end{pmatrix} = 16 + 9 = 25$$

Donc :

$$|x_1 - x_2| = \sqrt{25} = 5$$

2-9) Centre (moyenne) du nuage $$\bar{x}$$:

$$\bar{x} = \frac{1}{3} X^T \mathbf{1} = \frac{1}{3} \begin{pmatrix}3 & -1 & 4 \\ -1 & 2 & 2\end{pmatrix} \begin{pmatrix}1 \\ 1 \\ 1\end{pmatrix} = \frac{1}{3} \begin{pmatrix}6 \\ 3\end{pmatrix} = \begin{pmatrix}2 \\ 1\end{pmatrix}$$

Homework : Métrique diagonale

Pour une métrique diagonale orthogonaux :

Matrice de centrage :

$$H = I_3 - \frac{1}{3} \mathbf{1}\mathbf{1}^T$$

$$H = \begin{pmatrix}1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1\end{pmatrix} - \frac{1}{3} \begin{pmatrix}1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1\end{pmatrix} = \begin{pmatrix}\frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3}\end{pmatrix}$$

Données centrées :

$$X' = HX$$

$$X' = \begin{pmatrix}\frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3}\end{pmatrix} \begin{pmatrix}3 & -1 \\ -1 & 2 \\ 4 & 2\end{pmatrix} = \begin{pmatrix}1 & -2 \\ -3 & 1 \\ 2 & 1\end{pmatrix}$$

Donc

$$x'_1 = (1, -2), x'_2 = (-3, 1), x'_3 = (2, 1)$$

Condition d'orthogonalité :

$$(x'_2)^T M x_3\text{'} = 0 \quad avec \quad M = \begin{pmatrix}a & 0 \\ 0 & b\end{pmatrix}$$

$$(x'_2)^T M x'_3 = \begin{pmatrix}-3 & 1\end{pmatrix} \begin{pmatrix}a & 0 \\ 0 & b\end{pmatrix} \begin{pmatrix}2 \\ 1\end{pmatrix} = \begin{pmatrix}-3a & b\end{pmatrix} \begin{pmatrix}2 \\ 1\end{pmatrix} = -6a + b = 0 \quad , M = a \begin{pmatrix}1 & 0 \\ 0 & 6\end{pmatrix}$$

3. Espace des variables

3-1) Centrage : Pour centrer X, utiliser la matrice de centrage H

3-2) Tableau centré X_0 = HX :

$$X_0 = \begin{pmatrix} 1 & -2 \\ -3 & 1 \\ 2 & 1 \end{pmatrix}$$

3-3) Norme D de la 2ème variable centrée :

$$< X_2', X_2^T > |v_2|_D^2 = v_2^T D v_2 = \begin{pmatrix} -2 & 1 & 1 \end{pmatrix} \frac{1}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$$

$$= \frac{1}{3} \begin{pmatrix} -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{3}(4+1+1) = \frac{6}{3} = 2$$

Donc :

3-4) Produit scalaire D entre les variables centrées :

Colonnes

$$v_1 = \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix}$$

et

$$v_2 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$$

$$\langle v_1, v_2 \rangle_D = v_1^T D v_2 = \begin{pmatrix} 1 & -3 & 2 \end{pmatrix} \frac{1}{3} I_3 \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$$

$$= \frac{1}{3} \begin{pmatrix} 1 & -3 & 2 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{3}(-2-3+2) = \frac{-3}{3} = -1$$

3-5) Généralisation matricielle : La matrice des produits scalaires entre variables centrées est :

$$S = X_0^T D X_0 = X_0^T \frac{1}{3} I_3 X_0 = \frac{1}{3} X_0^T X_0$$

$$X_0^T X_0 = \begin{pmatrix} 1 & -3 & 2 \\ -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ -3 & 1 \\ 2 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1+9+4 & -2-3+2 \\ -2-3+2 & 4+1+1 \end{pmatrix} = \begin{pmatrix} 14 & -3 \\ -3 & 6 \end{pmatrix}$$

Donc :

$$S = \frac{1}{3} \begin{pmatrix} 14 & -3 \\ -3 & 6 \end{pmatrix} = \begin{pmatrix} \frac{14}{3} & -1 \\ -1 & 2 \end{pmatrix}$$

3-6) Réduction : Diviser chaque colonne de X_0 par sa norme D.

3-7) Calcul des normes D :

$$|v_1|D^2 = S11 = \frac{14}{3}$$

$$|v_2|D^2 = S22 = 2$$

Matrice des normes :

$$\Sigma = \mathrm{diag}(\sqrt{S_{11}}, \sqrt{S_{22}}) = \mathrm{diag}(\sqrt{\frac{14}{3}}, \sqrt{2})$$

Tableau centré-réduit :

$$X_{00} = X_0 \Sigma^{-1} = \begin{pmatrix} 1 & -2 \\ -3 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{14/3}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$X_{00} \approx \begin{pmatrix} 0,46291 & -1,41421 \\ -1,38873 & 0,70711 \\ 0,92582 & 0,70711 \end{pmatrix}$$

3-8) Matrice de corrélation :

$$R = X_{00}^T D X_{00} = \Sigma^{-1} S \Sigma^{-1}$$

$$R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

avec :

$$\rho = \frac{S_{12}}{\sqrt{S_{11}}\sqrt{S_{22}}} = \frac{-1}{\sqrt{14/3}\sqrt{2}} \approx -0,32733$$