TD : Dimension

1 Exercice 1 : Compréhension de l'ACP

Soit le vecteur aléatoire $X = (X_1, X_2, X_3, X_4)^\top$ dont l'espérance et la variance sont, respectivement, données par

$$\mu = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \text{ et } \Sigma = \begin{pmatrix} 9 & 1 & -1 & 2 \\ 1 & 4 & -1 & 1 \\ -1 & -1 & 16 & 0 \\ 2 & 1 & 0 & 9 \end{pmatrix}.$$

Les valeurs et vecteurs propres (normés) correspondants de Σ sont

$$\lambda_1 = 16.27, \qquad \alpha_1 = (0.165, 0.098, -0.980, 0.059)^\top, \tag{1}$$
$$\lambda_2 = 11.12, \qquad \alpha_1 = (0.665, 0.169, 0.171, 0.707)^\top, \tag{2}$$
$$\lambda_3 = 6.95, \qquad \alpha_1 = (0.718, -0.017, 0.077, -0.691)^\top, \tag{3}$$
$$\lambda_4 = 3.67, \qquad \alpha_1 = (0.118, -0.981, -0.070, 0.139)^\top. \tag{4}$$
$$\tag{5}$$

1. Trouvez une combinaison linéaire de X_1, X_2, X_3 et X_4 dont la variance est 6.95.
2. Soit $Z = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4$. Trouvez les valeurs de a_0, a_1, a_2, a_3, a_4 telles que Z est de moyenne 0 et de variance maximale. Quelle est la valeur de cette variance ? Quelle est la covariance entre cette combinaison linéaire et celle trouvée dans la question 1.
3. Donnez une matrice diagonale M telle que $M\Sigma M = R$, où R est la matrice des corrélations de X.

2 Exercice 2 : Matrice de covariance et de corrélation

Démontrer que la matrice de variance-covariance échantillonnnale des données centrées et réduites est égale à la matrice de corrélation échantillonnale des données initiales.

3 Exercice 3 : Pratique de l'ACP

Soit un vecteur de trois variables aléatoires $X = (X_1, X_2, X_3)^\top$ dont la matrice de variance est donnée par $\mathrm{Var}(X) = \Sigma$. On vous dit que les valeurs propres de Σ sont $3, 2$ et 1 et que les vecteurs propres normés correspondants sont, respectivement, $v_1 = (0, -1/\sqrt{2}, 1/\sqrt{2})^\top$, $v_2 = (1, 0, 0)^\top$ et $v_3 = (0, 1/\sqrt{2}, 1/\sqrt{2})^\top$.

1. En fait, X_1, X_2 et X_3 représentent, respectivement, la circonférence du poignet droit, un score de capacité pulmonaire et l'indice de masse corporelle des individus d'une certaine population. Les médecins aimeraient résumer ces trois mesures en un seul score, noté C, qui est une combinaison linéaire des trois variables originales, i.e. $C = c_1X_1 + c_2X_2 + c_3X_3$ pour des nombres réels c_1, c_2 et c_3) et qui capture le plus de variabilité possible. Trouvez les valeurs de c_1, c_2 et c_3 telles que $\sum_{j=1}^{3} c_j^2 = 1$ et que la variance de C sera maximale. Quelle est la valeur de cette variance maximale ?
2. Quelle proportion de la variabilité totale de X_1, X_2 et X_3 est capturée par C ?
3. Retrouvez la matrice Σ à partir des informations données dans l'énoncé de la question.