

Análisis Multivariado: Tarea 3

Técnicas de reducción de dimensión I

Fecha de entrega: 22 de octubre

PCA

1. (.5 puntos) Considerar la transformación dada por $\mathbf{Z} = \Sigma^{-\frac{1}{2}}\mathbf{X}$. ¿Tiene sentido realizar un análisis de componentes principales para \mathbf{Z} ?
2. (.5 puntos) Sean U_1 y U_2 dos variables aleatorias independientes con distribución uniforme $(0,1)$ y $X = (X_1, X_2, X_3, X_4)$, con $X_1 = U_1$, $X_2 = U_2$, $X_3 = U_1 + U_2$ y $X_4 = U_1 - U_2$. Calcula la matriz de correlación de X . ¿Cuántos componentes principales son de interés?
3. (2 puntos) Considerar el archivo *oecd.txt*, el cual contiene una serie de mediciones sobre 11 tópicos que la OECD ha identificado como esenciales con respecto a las condiciones y la calidad de vida en distintos países. Realiza un análisis de componentes principales, proporcionando un reporte completo del procedimiento, consideraciones y análisis de resultados.
4. (2 puntos) Considera la base de datos *heptathlon_2016.txt*, la cual contiene los resultados de los competidores de la prueba del heptatlón en los juegos olímpicos de Río 2016. Realiza un análisis de componentes principales, proporcionando un reporte completo del procedimiento, consideraciones y análisis de resultados.

FA

5. (.5 puntos) Sea \mathbf{x} un vector aleatorio p -dimensional y \mathbf{f} el conjunto de m factores latentes. Obtener $\text{Cov}(\mathbf{x}, \mathbf{f})$.
6. (.5 puntos) Si \mathbf{x} es un vector aleatorio cuya matriz de covarianza se puede escribir como $\Sigma = \Lambda\Lambda^T + \Psi$, muestra que existen factores \mathbf{f}, \mathbf{u} tal que el modelo de k factores se cumple para \mathbf{x} . ¿Si \mathbf{x} se distribuye normal multivariado que se puede decir de la distribución de \mathbf{f} y de \mathbf{u} ? *Hint: Considera $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I} + \Lambda^T\Psi^{-1}\Lambda)$ independiente de \mathbf{x} y define*

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{f} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_p & \Lambda \\ -\Lambda^T\Psi^{-1} & \mathbf{I}_k \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{x} - \mu \\ \mathbf{y} \end{pmatrix}.$$

7. (2 puntos) Considera la matriz de correlación dada en el archivo *R_FA.txt*, entre ratings de 9 enunciados acerca del dolor hechos por 123 personas sufriendo de dolor extremo. A cada uno de los 9 enunciados se le dio un valor del 1 al 6 (completamente de acuerdo, muy de acuerdo, de acuerdo, en desacuerdo, muy en desacuerdo, completamente en desacuerdo), siendo los enunciados los siguientes:
- a. Si en el futuro tengo o no dolor depende de la habilidad y conocimiento de mis médicos.
 - b. Cuando tengo dolor, es usualmente a causa de algo que hice o que dejé de hacer.
 - c. Si tengo o no dolor depende de lo que los médicos hagan por mi.
 - d. No puedo obtener ayuda para mi dolor a menos de que busque atención médica.
 - e. Cuando tengo dolor sé que es a causa de que no he hecho el ejercicio adecuado ni he comido de forma adecuada.
 - f. El dolor de las personas resulta de su imprudencia.
 - g. Soy directamente responsable de mi dolor.
 - h. Alivio del dolor es controlado principalmente por los médicos.
 - i. Personas que no tienen dolor son meramente afortunadas.

Asumiendo normalidad en las variables, realiza un análisis de factores, proporcionando un reporte completo del procedimiento e interpretación de los resultados.

8. (2 puntos) Para la base de datos del heptatlón realiza un análisis de factores, proporcionando un reporte completo del procedimiento e interpretación de los resultados. ¿Hay alguna diferencia con respecto al análisis por componentes principales realizado en el ejercicio 4?