

Análisis Multivariado: Tarea 3

Técnicas de reducción de dimensión

Fecha de entrega: 12 de abril.

PCA

1. (.5 puntos) Sea \mathbf{x} un vector aleatorio p -dimensional, calcula la correlación entre el j -ésimo elemento de la variable, x_j , y el i -ésimo componente principal. ¿Qué interpretación le puedes dar?
2. (.5 puntos) Si estandarizamos los datos, i.e., $\mathbf{Z} = \Sigma^{-\frac{1}{2}}\mathbf{X}$. ¿Tiene sentido realizar un análisis de componentes principales? Justifica tu respuesta.
3. (2.5 puntos) Considerar el archivo *EPL_19_20.txt* que contiene la tabla de posiciones de la Liga Premier Inglesa de la temporada 2019-2020. Realiza un análisis (detallado) por componentes principales y proporciona un reporte completo del procedimiento, consideraciones y análisis de resultados.

FA

4. (1 punto) Bajo el supuesto de normalidad considere el estimador máximo verosímil de Σ dado por $\hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi}$, mostrar que $\text{tr}(S\hat{\Sigma}^{-1}) = p$. ¿Qué interpretación le puedes dar?
5. (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 y reportando cada uno de los pasos realizados con su justificación.

MDS

5. (.5 puntos) Mostrar que para la matriz doblemente centrada \mathbf{B} se tiene que

$$b_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}$$

5. (1.5 puntos) El archivo *mx.txt* contiene la distancia en kilómetros de 15 ciudades de México. Realiza un escalamiento multidimensional y trata de identificarlas justificando cada uno de los pasos de tu análisis.
6. (1.5 puntos) En este ejercicio se hará un análisis de una figura conocida como el rollo suizo. Para esto realizar lo siguiente:

- i. Construir el famoso rollo suizo dado por las coordenadas

$$\begin{aligned} x_i &= \frac{1}{2}v_i \sin(4\pi v_i) \\ y_i &= u_i - \frac{1}{2} \\ z_i &= \frac{1}{2}v_i \cos(4\pi v_i), \end{aligned}$$

donde u_i y v_i son variables aleatorias uniformes en el intervalo $(0, 1)$ (considerar $n = 1000$) y graficar la figura resultante coloreando los puntos usando un mapeo apropiado (e.g., utilizando la tercer coordenada).

- ii. Obtener la matriz de distancias Euclidianas.
- iii. Realizar un escalamiento multidimensional métrico con $k = 2$ y graficar los resultados coloreando los puntos utilizando el mismo mapeo que en el primer inciso.

¿Es el resultado esperado? Comenta los resultados haciendo énfasis en los supuestos del escalamiento multidimensional métrico.