Cuestiones Temas 4,5, 6 y 8

1 Análisis de Componentes Principales

Sea $\mathbf{X}$ una matriz de datos $n \times p$ y $\widehat{\Sigma}$ la matriz de varianzas asociada.

1. Describe cual es el objetivo fundamental del ACP.
2. Define las puntuaciones de los n puntos sobre las p CP.
3. Demuestra que $d(x_i, x_j) = d(y_i, y_j)$ con y_i el vector de puntuaciones sobre las componentes principales correspondiente al individuo i.
4. Demuestra que la varianza de la $j-$ésima CP es el $j-$ésimo mayor autovalor de $\widehat{\Sigma}$
5. Determina el coeficiente de correlación lineal entre $y_{(i)}$ ($i-$ésima CP) y $x_{(j)}$.($j-$ésima variable).
6. Demuestra que la primera CP es la combinación lineal (normalizada) de máxima varianza.
7. Demuestra que la suma de las varianzas de las variables originales es igual a la suma de las varianzas de las CP.
8. Demuestra que las CP están incorreladas
9. Considérese el conjunto de datos muestrales
$$\mathbf{X} = \begin{pmatrix} 5 & 3 & ? & ? & ? \\ 4 & a & ? & ? & ? \\ 4 & a & ? & ? & ? \\ 1 & a & ? & ? & ? \end{pmatrix}.$$
   (a) Determina el mayor autovalor de $\widehat{\mathbf{\Sigma}}$ sabiendo que el autovector asociado es $(1, 0, 0, 0, 0)$.
   (b) Calcula el valor de a sabiendo que el segundo autovector es $(0, 1, 0, 0, 0)$

2 Análisis Factorial

1. Describe cual es el objetivo básico del Análisis Factorial.
2. Describe el Modelo Factorial Ortogonal y las hipótesis del mismo.
3. Demuestra que bajo las hipótesis del modelo:
$$\mathbf{\Sigma} = \mathbf{LL}' + \mathbf{\Psi} \Leftrightarrow \begin{cases} \sigma_{ii} = l_{i1}^2 + ... + l_{im}^2 + \psi_i \\ \sigma_{ik} = \sum_{j=1}^{m} l_{ij} \cdot l_{kj} \end{cases} \tag{1}$$
$$Cov(\underline{X}, \underline{F}) = \mathbf{L} \Leftrightarrow l_{ij} = Cov(X_i, F_j) \tag{2}$$

4. Demuestra que si las variables originales están normalizadas

$$\rho\left(X_i, F_j\right) = l_{ij}.$$

5. Describe las (3) etapas básicas, y el objetivo de cada una de ellas, cuando se realiza un análisis factorial.

6. Describe el Método de las Componentes Principales para estimar la matriz de cargas factoriales y la matriz de varianzas específicas.

7. Demuestra que en el Método de las Componentes Principales, las cargas factoriales estimadas no cambian cuando se incrementa el número de factores.

8. Describe

   (a) En qué consiste la rotación de factores.

   (b) En que consiste la rotación varimax y qué pretende.

3 Análisis de Correspondencia

1. (a) Plantea los problemas de independencia de caracteres y de homogeneidad de poblaciones.

   (b) Demuestra que los estadísticos χ^2 asociados a ambos problemas y sus distribuciones coinciden.

2. Sean $r'_1, ..., r'_n$ las distribuciones condicionadas por filas asociadas a un tabla de contingencia

$$\begin{array}{ccccccc}
 & B_1 & ... & B_j & ... & B_p & \\
A_1 & N_{11} & ... & N_{1j} & ... & N_{1p} & N_{1.} \\
\vdots & \vdots & & \vdots & & \vdots & \vdots \\
A_i & N_{i1} & ... & N_{ij} & ... & N_{ip} & N_{i.} = \sum_{j=1}^{p} N_{ij} \\
\vdots & \vdots & & \vdots & & \vdots & \vdots \\
A_n & N_{n1} & ... & N_{nj} & ... & N_{np} & N_{n.} \\
 & & & & & & \\
 & N_{.1} & ... & N_{.j} = \sum_{i=1}^{n} N_{ij} & ... & N_{.p} & N
\end{array}$$

   (a) Define el centro de gravedad de las filas: m'_r

   (b) Define la distancia ji-cuadrado entre dos distribuciones.

   (c) Demuestra que $\chi^2 = N \sum_{i=1}^{n} f_{i.} d_{\chi^2}\left(r_i, m_r\right)$

3. Describe el objetivo del análisis de correspondencias (por filas).

4. Analogías y diferencias entre el ACP y el AC.

5. Plantea y resuelve el problema de optimización al que conduce el AC.
6. Demuestra que el cuadrado de la distancia euclídea entre las proyecciones de dos filas P_{r_i} y $P_{r_{i'}}$ coincide con la distancia $ji-cuadrado$ entre las filas r_i y $r_{i'}$.
7. Concepto de vértice en el problema del AC por filas. Interpretación.
8. Determina las coordenadas de m_r' en la nueva base.
9. Demuestra que $(0, m_r)$ es un autovalor-vector de la matriz

$$S_r = \sum_{i=1}^{n} f_{i.}\,(r_i - m_r)\,(r_i - m_r)^t\, D_p^{-1}.$$

10. Determina las coordenadas de las filas correspondientes al autovector m_r. ¿Qué evidencian?
11. Demuestra que la varianza ponderada de las coordenadas correspondientes al autovector u_k se puede expresar como

$$\sum_{i=1}^{n} f_{i.}\left(r_i^t D_p^{-1} u_k\right)^2.$$

12. Define la inercia total y demuestra que es igual a la suma de los autovalores de S_r.

4 Análisis Discriminante

1. Discriminación en dos poblaciones con distribuciones conocidas
    (a) Define la probabilidad total de error de clasificación de una regla discriminante.
    (b) Determina la regla discriminante que minimiza la probabilidad total de error de clasificación
    (c) Plantea y resuelve el problema original de Fisher.
    (d) Demuestra que supuesto que las poblaciones se distribuyen según $N(\mu_i, \Sigma)$, $i = 1, 2$, la regla discriminante que minimiza la probabilidad total de error coincide con la resultante de la propuesta de Fisher.
2. Coordenadas discriminantes canónicas
    (a) Planteamiento y solución.
    (b) Describe el criterio de clasificación basado en las coordenadas canónicas.