Análisis factorial

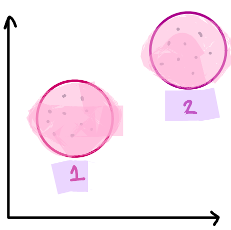
Wednesday, April 20, 2022 7:40 PM

Similar a PCA

Obj: Describir la estructura de cov. de los datos de muchas vars en términos de unas pocas.

Busca factores comunes (relaciones) entre las vars. para juntarlas

Voy a suponer q' las vars se pueden agrupar maximizando la correlación intragrupal y minimizando la correlación intergrupal



Maximizar la diferencia entre 1 y 2 y minimizar las internas de los grupos

C/ grupo representa un único aspecto o factor q' justifica correlaciones.

Ej: Puntaje exámenes

$\underbrace{\text{Phys, Math, Chem}}$

Los puntajes en estas suelen parecerse

"inteligencia"

$\underbrace{\text{Agilidad, Fuerza}}$

Puntajes similares

"Desempeño físico"

Formulación

Sea $\mathbb{X}$ un vec. aleatorio observable con p componentes

Sea μ la media y Σ la cov. El modelo factorial asume q' $\mathbb{X}$ es lin dep. de V.a's $F_1, .., F_m$ (factores comunes) y p fuentes de variación $\varepsilon_1, .., \varepsilon_p$ (errores o factores específicos)

$$X_1 - \mu_1 = l_{11} F_1 + l_{12} F_2 + .. + l_{1m} F_m + \varepsilon_1$$

$$\vdots$$

$$X_p - \mu_p = l_{p1} F_1 + l_{p2} F_2 + .. + l_{pm} F_m + \varepsilon_p$$

$$\underset{p\times 1}{\mathbb{X} - \mu} = \underset{p\times m}{\mathbb{L}}\ \underset{m\times 1}{\mathbb{F}} + \underset{p\times 1}{\varepsilon}$$

l_{ij} "Loading": Es el peso de la var i en el factor j.

¿Cuánto aporta i en j?

Estos m factores comunes son similares a los m primeros C.P

Suposiciones sobre $\mathbb{F}$ y $\mathbb{L}$:

1. $E[\mathbb{F}] = \underset{m\times 1}{0}$ $\qquad$ • $\operatorname{cov}(\mathbb{F}) = E[\mathbb{F}\mathbb{F}'] = \underset{m\times m}{\mathbb{I}}$

2. $E[\mathcal{E}] = \underset{m\times 1}{0}$ $\qquad$ • $\operatorname{cov}(\mathcal{E}) = E[\mathcal{E}\mathcal{E}']$

$$= \psi = \begin{bmatrix} \psi_1 & & 0 \\ & \ddots & \\ 0 & & \psi_p \end{bmatrix}$$

ψ_i: i-ésimo error

3. $\mathcal{E}, \mathbb{F}$ indep $\rightarrow$ $\operatorname{cov}(\mathcal{E}, \mathbb{F}) = 0$

$$\hookrightarrow \cancel{E[\mathcal{E}\mathbb{F}]} - \cancel{E[\mathcal{E}]}\,\cancel{E[\mathbb{F}]} = 0$$

Tarea: Demostrar

$\Sigma = \operatorname{cov}(\mathbb{X}) = \mathbb{L}\mathbb{L}' + \psi$ (i)

Pista: $(\mathbb{X} - \mu)(\mathbb{X} - \mu)' = \mathbb{L}\mathbb{F} + \mathcal{E}$

$\operatorname{cov}(\mathbb{X}, \mathbb{F}) = \mathbb{L}$ (ii)

$$\mathbb{X} - \mu = \mathbb{L}\mathbb{F} + \mathcal{E}$$

$$\begin{aligned}
\mathbb{L}\mathbb{F} + \mathcal{E} &= (\mathbb{X} - \mu)(\mathbb{X} - \mu)' \\
&= (\mathbb{L}\mathbb{F} + \mathcal{E})(\mathbb{L}\mathbb{F} + \mathcal{E})' \\
&= (\mathbb{L}\mathbb{F} + \mathcal{E})((\mathbb{L}\mathbb{F})' + \mathcal{E}') \\
&= \mathbb{L}\mathbb{F}(\mathbb{L}\mathbb{F})' + \mathbb{L}\mathbb{F}\mathcal{E}' + \mathcal{E}(\mathbb{L}\mathbb{F})' + \mathcal{E}\mathcal{E}'
\end{aligned}$$

$$\begin{aligned}
\Sigma = \operatorname{cov}(\mathbb{X}) &= E[(\mathbb{X} - \mu)(\mathbb{X} - \mu)'] \\
&= E[\mathbb{L}\mathbb{F}(\mathbb{L}\mathbb{F})' + \mathbb{L}\mathbb{F}\mathcal{E}' + \mathcal{E}(\mathbb{L}\mathbb{F})' + \mathcal{E}\mathcal{E}'] \\
&= E[\mathbb{L}\mathbb{F}(\mathbb{L}\mathbb{F})'] + E[\mathbb{L}\mathbb{F}\mathcal{E}'] + E[\mathcal{E}(\mathbb{L}\mathbb{F})'] + E[\mathcal{E}\mathcal{E}'] \\
&= \mathbb{L}E[\mathbb{F}\mathbb{F}']\mathbb{L}' + \mathbb{L}E[\mathbb{F}\mathcal{E}'] + \mathbb{L}'[\mathcal{E}\mathbb{F}'] + E[\mathcal{E}\mathcal{E}'] \\
&= \mathbb{L}\mathbb{I}\mathbb{L}' + \mathbb{L}\cancel{E[\mathbb{F}\mathcal{E}']} + \mathbb{L}'\cancel{E[\mathbb{F}'\mathcal{E}]} + \psi \\
&= \mathbb{L}\mathbb{L}' + \psi
\end{aligned}$$

De ① se sigue q: ($\mathbb{L}_i$ es fila)

- $var(x_i) = l_{i1}^2 + \cdots + l_{im}^2 + \psi_i = \mathbb{L}_i \mathbb{L}_i' + \psi_{ii}$
- $cov(x_i, x_k) = l_{i1} l_{k1} + \cdots + l_{im} l_{km} = \mathbb{L}_i \mathbb{L}_k' + \psi_{ik}$

$\hookrightarrow \psi_{ik} = 0$

De ⑪:

- $Cov(x_i, F_j) = l_{ij}$

Se asume relaº lin entre $\mathbb{X}$ y $\mathbb{F}$, si no, no sirve...