Institución Universitaria
Acreditada en Alta Calidad

Innovación Tecnológica con Sentido Humano

Análisis de Factores

Cristian Guarnizo-Lemus
cristianguarnizo@itm.edu.co

Maestria en Automatización y Control Industrial

Contenido

Contenido

Introducción

- En FA representamos las variables $x_1, x_2, \ldots, x_p$ como una combinación lineal de unas pocas variables $z_1, z_2, \ldots, z_m$ $(m < p)$ llamados factores.
- Los factores son variables latentes que generan los x's.
- Si las variables originales $x_1, x_2, \ldots, x_p$ están moderadamente correlacionadas, la dimensionalidad básica del sistema es menor que p.
- El objetivo del análisis de factores es reducir la redundancia entre las variables usando un número pequeño de factores.

Introducción

Ejemplo matriz de correlación

$$\mathbf{R} = \begin{bmatrix} 1.00 & .90 & .05 & .05 & .05 \\ .90 & 1.00 & .05 & .05 & .05 \\ .05 & .05 & 1.00 & .90 & .90 \\ .05 & .05 & .90 & 1.00 & .90 \\ .05 & .05 & .90 & .90 & 1.00 \end{bmatrix}$$

Diferencias con PCA

A pesar que PCA es también una técnica de reducción de dimensionalidad, existen 2 diferencias con FA

- Las componentes principales están definidas como una combinación lineal de la variables originales. En FA, las variables originales son expresadas como una combinación lineal de factores.
- En PCA, se explica un gran parte de la varianza total de la variables. En FA, buscamos cuantificar las convarianzas o correlaciones entre las variables.

Diferencias con PCA

A pesar que PCA es también una técnica de reducción de dimensionalidad, existen 2 diferencias con FA

- Las componentes principales están definidas como una combinación lineal de la variables originales. En FA, las variables originales son expresadas como una combinación lineal de factores.
- En PCA, se explica un gran parte de la varianza total de la variables. En FA, buscamos cuantificar las convarianzas o correlaciones entre las variables.

Contenido

Factores Ortogonales

Asumimos una muestra aleatoria $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ con media $\boldsymbol{\mu}$ y matriz de covarianza $\boldsymbol{\Sigma}$. Expresamos cada variable como una combinación lineal de factores comunes $z_1, z_2, \ldots, z_m$, acompañadas con un termino de error para la parte de las variables que es única.

$$
\begin{aligned}
x_1 - \mu_1 &= w_{11}z_1 + w_{12}z_2 + \cdots + w_{1m}z_m + \epsilon_1 \\
x_2 - \mu_2 &= w_{21}z_1 + w_{22}z_2 + \cdots + w_{2m}z_m + \epsilon_2 \\
&\vdots \\
x_p - \mu_p &= w_{p1}z_1 + w_{p2}z_2 + \cdots + w_{pm}z_m + \epsilon_p
\end{aligned}
$$

$$\mathbf{x} - \boldsymbol{\mu} = \mathbf{Wz} + \boldsymbol{\epsilon}$$

Factores Ortogonales

$$\mathbf{x} - \boldsymbol{\mu} = \mathbf{W}\mathbf{z} + \boldsymbol{\epsilon}$$

Adicionalmente, asumimos

$$\mathbb{E}[\mathbf{z}] = \mathbf{0},\ \operatorname{cov}(\mathbf{z}) = \mathbf{I},\ \mathbb{E}[\boldsymbol{\epsilon}] = \mathbf{0},$$

y

$$\operatorname{cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \psi_p, \end{bmatrix}$$

y $\operatorname{cov}(\mathbf{z}, \boldsymbol{\epsilon}) = \mathbf{0}$. Prácticamente

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}),\ \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

Factores Ortogonales

Recordando que

$$\mathbf{x} = \mathbf{W}\mathbf{z} + \boldsymbol{\mu} + \boldsymbol{\epsilon}, \text{ y } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}), \ \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

Calculamos el modelo probabilístico de $\mathbf{x}$ así

$$\mathbb{E}[\mathbf{x}] = \mathbb{E}[\mathbf{W}\mathbf{z}] + \mathbb{E}[\boldsymbol{\mu}] + \mathbb{E}[\boldsymbol{\epsilon}],$$

$$\begin{aligned}
\mathrm{cov}(\mathbf{x}) &= \mathbb{E}[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top] \\
&= \mathbb{E}[(\mathbf{W}\mathbf{z} + \boldsymbol{\epsilon})(\mathbf{W}\mathbf{z} + \boldsymbol{\epsilon})^\top] \\
&= \mathbb{E}[\mathbf{W}\mathbf{z}\mathbf{z}^\top\mathbf{W}^\top] + \mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top] \\
&= \mathbf{W}\mathbb{E}[\mathbf{z}\mathbf{z}^\top]\mathbf{W}^\top + \mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top] \\
&= \mathbf{W}\mathbf{I}\mathbf{W}^\top + \boldsymbol{\Psi} \\
\boldsymbol{\Sigma} &= \mathbf{W}\mathbf{W}^\top + \boldsymbol{\Psi}
\end{aligned}$$

Entonces $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Análisis de términos

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{W}\mathbf{W}^\top + \boldsymbol{\Psi}), \text{ y } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}), \ \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

$$\operatorname{cov}(x_i, x_i) = \mathbf{w}_{i:}\mathbf{w}_{i:}^\top + \psi_i = \sum_{k=1}^{m} w_{ik}^2 + \psi_i$$

$$\operatorname{cov}(x_i, x_j) = \mathbf{w}_{i:}\mathbf{w}_{j:}^\top = \sum_{k=1}^{m} w_{ik} w_{jk}$$

Innovación Tecnológica con Sentido Humano

Contenido

Calculo de matriz de pesos- EM

No se puede determinar **W** de manera cerrada usando máxima verosimilitud. Pero la podemos calcular de manera iterativa empleando el algoritmo EM, del cual obtenemos los siguientes paso

- Paso E:

$$\mathbb{E}[\mathbf{z_n}] = \mathbf{G}\mathbf{W}^\top \boldsymbol{\psi}^{-1}(\mathbf{x_n} - \boldsymbol{\mu})$$
$$\mathbb{E}[\mathbf{z_n}\mathbf{z_n}^\top] = \mathbf{G} + \mathbb{E}[\mathbf{z_n}]\mathbb{E}[\mathbf{z_n}]^\top$$

  con $\mathbf{G} = (\mathbf{I} + \mathbf{W}^\top \boldsymbol{\psi}^{-1}\mathbf{W})^{-1}$.

- Paso M:

$$\mathbf{W}^{\text{new}} = \left[\sum_{n=1}^{N} (\mathbf{x}_n - \boldsymbol{\mu})\,\mathbb{E}\left[\mathbf{z}_n\right]^\top\right]\left[\sum_{n=1}^{N} \mathbb{E}\left[\mathbf{z}_n\mathbf{z}_n^\top\right]\right]^{-1}$$
$$\boldsymbol{\psi}^{\text{new}} = \operatorname{diag}\left\{\mathbf{S} - \mathbf{W}^{\text{new}}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left[\mathbf{z}_n\right](\mathbf{x}_n - \boldsymbol{\mu})^\top\right\}$$

Calculo de matriz de pesos- PC

- De la muestra aleatoria $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ se obtiene la matriz de covarianza muestral $\mathbf{S}$ e intentamos encontrar un estimador de $\hat{\mathbf{W}}$ que aproxime

$$\mathbf{S} \approx \hat{\mathbf{W}}\hat{\mathbf{W}}^\top + \hat{\boldsymbol{\Psi}}.$$

- En este método, olvidamos $\hat{\boldsymbol{\Psi}}$, entonces $\mathbf{S} = \hat{\mathbf{W}}\hat{\mathbf{W}}^\top$. Para factorizar $\mathbf{S}$ empleamos SVD

$$\mathbf{S} = \mathbf{U}\mathbf{D}\mathbf{U}^\top,$$

  donde $\mathbf{U}$ es una matriz ortogonal construida con los vectores propios normalizados y $\mathbf{D}$ es una matriz diagonal con valores propios $\lambda_1, \lambda_2, \ldots, \lambda_p$ de $\mathbf{S}$.
- Se aproxima la matriz $\hat{\mathbf{W}} = \mathbf{U}\mathbf{D}^{1/2}$.

Calculo de matriz de pesos- PC

$$\begin{bmatrix} \hat{w}_{11} & \hat{w}_{12} \\ \hat{w}_{21} & \hat{w}_{22} \\ \hat{w}_{31} & \hat{w}_{32} \\ \hat{w}_{41} & \hat{w}_{42} \\ \hat{w}_{51} & \hat{w}_{52} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \\ u_{41} & u_{42} \\ u_{51} & u_{52} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} == \begin{bmatrix} \sqrt{\lambda_1}u_{11} & \sqrt{\lambda_2}u_{12} \\ \sqrt{\lambda_1}u_{21} & \sqrt{\lambda_2}u_{22} \\ \sqrt{\lambda_1}u_{31} & \sqrt{\lambda_2}u_{32} \\ \sqrt{\lambda_1}u_{41} & \sqrt{\lambda_2}u_{42} \\ \sqrt{\lambda_1}u_{51} & \sqrt{\lambda_2}u_{52} \end{bmatrix}$$

Calculo de matriz de pesos- PC PF - Ejemplo

	Principal Component Loadings		Principal Factor Loadings		
Variables	z_1	z_2	z_1	z_2	Communalities
Kind	.969	-.231	.981	-.210	.995
Intelligent	.519	.807	.487	.774	.837
Happy	.785	-.587	.771	-.544	.881
Likeable	.971	-.210	.982	-.188	.995
Just	.704	.667	.667	.648	.837
Variance accounted for	3.263	1.538	3.202	1.395	
Proportion of total variance	.653	.308	.704	.307	
Cumulative proportion	.653	.960	.704	1.01	

Contenido

Seleccionando el numero de factores

- Seleccione m igual al numero de factores necesarios para describir una varianza predeterminada (80%) del total de la varianza $\mathrm{tr}(\mathbf{S})$.
- Seleccione m igual al numero de valores propios mayor que su promedio.
- Usa el test scree o “scree plot” (L curva) basado en el gráfico de los valores propios de **S** o **R**. Si la gráfica baja abruptamente, seguida de una linea recta con una pendiente pequeña, seleccione m igual al numero de valores propios antes de comenzar la linea recta.
- Pruebe la hipótesis que m es el numero correcto de factores (sobre EM), $H_0 : \mathbf{\Sigma} = \mathbf{W}\mathbf{W}^\top + \mathbf{\Psi}$.

$$\left(n - \frac{2p + 4m + 11}{6}\right) \ln\left(\frac{\left|\hat{\mathbf{W}}\hat{\mathbf{W}}^\top + \hat{\mathbf{\Psi}}\right|}{|\mathbf{S}|}\right) \sim \chi^2_v, \; v = \frac{1}{2}[(p-m)^2 - p - m].$$

Contenido

Validación del Análisis de Factores

- Para muchos estadistas, FA es controversial y no pertenece a una técnica multivariada legitima.
- Las razones son: la dificultad para seleccionar *m*, la variedad de métodos para extraer los factores, muchas técnicas rotaciones, y la subjetividada en interpretación.

Referencias

- Alvin C. Rencher. “Methods of Multivariate Analysis”, 2002. Capitulo 13.
- C. Bishop. “Pattern Recognition and Machine Learning”, 2006. Sección 12.2.4.
- W. K. Härdle and L. Simar. “Applied Multivariate Statistical Analysis”, 2019. Capitulo 12.