Clase 3: Técnicas Multivariadas

Justo Andrés Manrique Urbina

7 de septiembre de 2019

1. Muestreo aleatorio

Sea $X = (x_1, x_2, \ldots, x_p)^T$ un vector aleatorio dónde cada X_i es una variable aleatoria. Una muestra de tamaño n para X es entonces $(X_1, X_2, \ldots, X_n)$ (cada X_1 es un vector como X). Asumamos que la distribución de cada X_i es la misma que la X y además son independientes. Así, se obtiene la siguiente base de datos:

$$x = \begin{pmatrix} X_{11} & X_{12} & \ldots & X_{1p} \\ X_{21} & X_{22} & \ldots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \ldots & X_{np} \end{pmatrix}$$

En dónde cada $(x_1, x_2, \ldots, x_p)^T$ es una observación de X_i.

1.1. Estadísticas

La media muestral, $\bar{X}$ para $\bar{X}$ es $\bar{X} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n)^T$, en dónde:

$$\bar{X}_i = \frac{(x_{1i} + x_{2i} + \ldots + x_{ni})}{n}.$$

Sí $y_i^{'} = (X_{1i}, X_{2i}, \ldots, X_{ni})$. Y se tiene que $\theta = (1, 1, \ldots, 1)^T$. Entonces su longitud es unitaria puesto que:

$$(\frac{1}{\sqrt{n}}\sqrt{1 + \ldots + 1}) = 1.$$

Así, la proyección de $y_i^{'}$ sobre $\frac{1}{\sqrt{n}}\theta$ es

$$y_i^T * (\frac{1}{\sqrt{n}}\theta)\frac{1}{\sqrt{n}}\theta.$$

$$\frac{x_{1i} + \ldots + x_{ni}}{n} = \bar{X}_i.$$

2. Propiedades de $\bar{X}$

Para el vector $\bar{X}_n$=$(\mu_1, \mu_2, \ldots, \mu_p)^T$, se tiene que:

$$E(\bar{X}) = E(\mu_1, \mu_2, \ldots, \mu_p) = \mu, \text{vector.}$$

$$\Sigma_{\bar{X}} = \frac{1}{n}\Sigma_X.$$

en dónde Σ_X es la matriz de varianza y covarianza de X.

Proof 1.

$$E(\bar{X}) = \frac{(X_1 + \ldots + X_n}{n}.$$

$$\frac{(E(X_1), E(X_2), \ldots, E(X_n))}{n}.$$

$$\frac{n\mu}{n} = \mu.$$

Proof 2.

$$\Sigma_X = E((\bar{X} - \mu)(\bar{X} - \mu)^T).$$

3. Varianza Generalizada

$$|\Sigma| = det(\Sigma).$$

o la traza de la matriz. En componentes principales se utiliza la traza.

4. Fórmulas

$$E(S_n) = \frac{n-1}{n}\Sigma_X.$$

en dónde

$S_n$varianza - covarianza muestral de tamaño n.

5. Distribución normal multivariada

Para un vector $X \in \mathbb{R}^p$:

$$\frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|}e^{-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu)}.$$

$$\mu = E(X) \in \mathbb{R}^n.$$

$$\Sigma = \Sigma_X \text{ matriz } p * p.$$

El término $(\bar{X}-\mu)^T\Sigma^{-1}(\bar{X}-\mu)=c^2$ es un elipsoide. Se puede demostrar que los autovalores y autovectores de Σ^{-1} son:

$$\frac{1}{\lambda_1},\frac{1}{\lambda_2},\ldots,\frac{1}{\lambda_p}.$$

$$e_i=\Sigma^{-1}\Sigma e_i=\Sigma^{-1}(\lambda_i e_i)=\lambda_i\Sigma^{-1}e_i.$$

$$\rightarrow \frac{1}{\lambda_i}e_i=\Sigma_{-1}e_i\rightarrow e_i \text{ es autovector de } \Sigma^{-1} \text{ y } \frac{1}{\lambda_i} \text{ es autovalor de } \Sigma^{-1}.$$

Para ello se utilizó la siguiente propiedad: Si A es definida positiva $\rightarrow$ sus autovalores son mayores que 0.

Proof 3. *Si A es definida positiva, entonces*

$$0<x^TAx,\forall x\neq 0.$$

En particular, para Σ y $x\neq 0$ se tiene que $0<x^T\Sigma x$, entonces para e, autovector con autovalor λ, se tiene que:

$$0<e^T\Sigma e=e^T\lambda e=\lambda e^Te=\lambda.$$

Propiedad: También Σ^{-1} es definida positiva:

$$x^T\Sigma^{-1}x=x^T(\sum_{i=1}^{p}(\frac{1}{\lambda_i})e_ie_t^T)x.$$

$$\sum_{i=1}^{p}\frac{1}{\lambda_i}(x^Te_i)^2,x\neq 0.$$

Σ^{-1}es definida positiva.

Demostremos por qué indicamos que es una elipsoide:

$$(x-\mu)^T\sum_{i=1}^{p}\frac{1}{\lambda_1}e_ie_i^T(X-\mu).$$

$$\sum_{i=1}^{p}\frac{1}{\lambda_i}((X-\mu)^Te_i)^2.$$

$$\sum_{i=1}^{p}\frac{1}{(\sqrt{\lambda_i})^2}((X-\mu)^Te_i)^2.$$

La elipsoide se define como:

$$\frac{((X_1-\mu_1)e_i)^2}{\frac{1}{c^2}(\sqrt{\lambda_1})^2}+\frac{((X_2-\mu_2)e_2)^2}{\frac{1}{c^2}(\sqrt{\lambda_2})^2}=1.$$

6. Componentes principales

Los componenes principales, el análisis de conglomerados, escalamiento multidimensional no requiere que las variables tengan distribuciones. Se tienen las siguientes variables univariadas $(X_1, X_2, \ldots, X_p)$. Cada X_i es un vector $(x_1, x_2, \ldots, x_p)^T$ con una matriz de varianza y covarianza Σ. Con estas variables se forman las siguientes combinaciones lineales:

$$Y_1 = \alpha_1^{'} X = a_{11}X_1 + \ldots + a_{1p}X_p.$$

$$\vdots$$

$$Y_p = \alpha_p^T X = a_{p1}X_1 + \ldots + a_{pp}X_p.$$

Cada a_i es un vector. La varianza de Y_i es definida por:

$$var(Y_i) = var(a_i^T X) = a_i^T \Sigma_X a_i.$$

$$cor(Y_i Y_j) = a_i^T \Sigma a_j.$$

Entonces se define que Y_1 es el primer componente principal CP_1. Si la varianza de Y es la mayor, con la condicion adicional de que $a_i^T a_i = 1$. ¿Cuàl es la combinación?

Resultado: El vector a que satisface el criterio es el que satisface: $max_a = \frac{a^T \Sigma a}{a^T a}$ por la última propiedad (Clase 2). Se tiene entonces que:

$$max_a \frac{a^T \Sigma a}{a^T a} = \lambda_1.$$

dónde λ_1 es el mayor autovalor de Σ y el máximo se alcanza cuando $a = e_1$ dónde e_1 es el autovector de Σ correspondiente a λ_1. Para Y_2 se busca el a_2 de tal modo que explique la mayor varianza no explicada por Y_2 pero que además $a_2^T a_2 = 1$ y $a_2^T a_1 = 0$. Para Y_k se busca a_k de tal manera que explique en mayor grado la varianza no explicada por $Y_1, \ldots, Y_{k-1}$ pero además $a_k^T a = 1$ y $a_k^T a_j = 0$ para $j = 1, 2, \ldots, k-1$.

6.1. Propiedades

$$var(Y_i) = var(e_1^T X) = var(e_1^T \Sigma_X e_i) = e_i^T \lambda_i e_i = \lambda_i.$$

Siempre y cuando se establezca que $0 \leq \lambda_p \leq \ldots \leq \lambda_1$

$$Cov(Y_i, Y_j) = 0.$$

$$\sigma_{11} + \ldots + \sigma_{pp} = \sum_{i=1}^{p} \sigma_{ii} = tr(\Sigma).$$

$$\Sigma_X = P \Lambda P^T.$$

$$tr(\Sigma) = tr(P \Lambda P^T$$

Entonces $tr(\Sigma) = tr(\Lambda) = \lambda_1 + \ldots + \lambda_p$