$$x_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{bmatrix} \qquad x_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{bmatrix}$$

θ - ângulo formado por x_1 e x_2

$$\cos\theta = \frac{x_1' x_2}{L_{x_1} L_{x_2}} = \frac{x_1' x_2}{\sqrt{x_1' x_1}\,\sqrt{x_2' x_2}}$$

Sejam $e_1 = \begin{bmatrix} x_{11} - \bar{x}_1 \\ x_{12} - \bar{x}_1 \\ \vdots \\ x_{1n} - \bar{x}_1 \end{bmatrix}$ $\qquad e_2 = \begin{bmatrix} x_{21} - \bar{x}_2 \\ x_{22} - \bar{x}_2 \\ \vdots \\ x_{2n} - \bar{x}_2 \end{bmatrix}$

$$L^2_{e_1} = e_1' e_1 = \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2 \qquad L^2_{e_2} = \sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2$$

$$e_1' e_2 = \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

$\therefore\ L^2_{e_j} = n s_j \qquad s_j$: variância amostral de X_j

$j = 1, 2.$

$e_1' e_2 = n s_{12}$ $\quad s_{12}$: covariância amostral entre X_1 e X_2

Seja θ_{12}: ângulo formado por e_1 e e_2.

$$\cos(\theta_{12}) = \frac{e_1' e_2}{L_{e_1} L_{e_2}} =$$

$$= \frac{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2} \sqrt{\sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2}}$$

$= r_{12}$ Coeficiente de Correlação amostral entre X_1 e X_2.

Se x_1 e x_2 forem centralizadas ou já tiverem média amostral igual a zero, então

$$\cos(\theta_{12}) = \frac{\sum_{i=1}^{n} x_{1i} x_{2i}}{\sqrt{\sum_{i=1}^{n} x_{1i}^2} \sqrt{\sum_{i=1}^{n} x_{2i}^2}} = \frac{x_1' x_2}{L_{x_1} L_{x_2}}$$

θ_{12}: ângulo formado por x_1 e x_2.

Diferenciação de funções de vetores

$x' = [x_1, \ldots, x_p]$

Seja $f(x)$ uma função contínua de x tal que

$f: \mathbb{R}^p \to \mathbb{R}$ e $\dfrac{\partial f(x)}{\partial x_i}$ e $\dfrac{\partial^2 f(x)}{\partial x_i \partial x_j}$ existam.

O vetor de derivadas parciais de $f(x)$ com relação a cada x_i é definido como

$$\frac{\partial f(x)}{\partial x} = \begin{bmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \dfrac{\partial f(x)}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_p} \end{bmatrix}$$

i) Se $f(x) = a'x$ $\quad a' = [a_1 \; a_2 \ldots a_p]$

$$f(x) = a_1x_1 + a_2x_2 + \ldots + a_px_p \Rightarrow \frac{\partial f(x)}{\partial x} = a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

ii) Se $f(x) = x' A x$

A matriz quadrada $p \times p$, simétrica $\Rightarrow \dfrac{\partial f(x)}{\partial x} = 2 A x$

Ex: $x' = \begin{bmatrix} x_1 & x_2 \end{bmatrix}$ $\quad A = \begin{bmatrix} a_{11} & a \\ a & a_{22} \end{bmatrix}$ $\quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$x' A x = \begin{bmatrix} x_1 a_{11} + x_2 a & x_1 a + x_2 a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} =$$

$$= a_{11} x_1^2 + a x_1 x_2 + a x_1 x_2 + a_{22} x_2^2 = a_{11} x_1^2 + 2 a x_1 x_2 + a_{22} x_2^2$$

$$\frac{\partial f(x)}{\partial x} = \begin{bmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \dfrac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2 a_{11} x_1 + 2 a x_2 \\ 2 a x_1 + 2 a_{22} x_2 \end{bmatrix}$$

$$2 A x = \begin{bmatrix} 2 a_{11} x_1 + 2 a x_2 \\ 2 a x_1 + 2 a_{22} x_2 \end{bmatrix}$$

A matriz de segundas derivadas parciais de $f(x)$ é definida como

$$H(x) = \frac{\partial^2 f(x)}{\partial x' \partial x} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_p} \\ \vdots & & \\ \frac{\partial^2 f(x)}{\partial x_p \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_p^2} \end{bmatrix}$$

$H(x)$ - matriz Hessiana

Máximos e Mínimos de $f(x)$

Se $x = x_0$ é ponto de máximo ou de mínimo de $f(x)$ então $\left.\frac{\partial f(x)}{\partial x}\right|_{x=x_0} = 0$ ↳ vetor nulo

a) Se $\left.\frac{\partial f(x)}{\partial x}\right|_{x=x_0} = 0$ e $H(x_0)$ é positiva definida

então x_0 é ponto de mínimo de $f(x)$.

b) Se $\left.\frac{\partial f(x)}{\partial x}\right|_{x=x_0} = 0$ e $H(x_0)$ é negativa definida

$[x' H(x) x < 0, \ \forall x \neq 0]$, então x_0 é ponto de máximo de $f(x)$.

Exemplo

$x_{p \times 1}$ $\quad C_{n \times p}$ $\quad K_{n \times n}$ simétrica $\quad a_{n \times 1}$

$$f(x) = (a - Cx)' K (a - Cx)$$
$$= a' K a - a' K C x - x' C' K a + x' C' K C x$$

Obs: $a' K C x = x' C' K a$ porque $a' K C x = (x' C' K a)'$
e ambos são escalares

$$\therefore f(x) = a' K a - 2 a' K C x + x' C' K C x$$

$$\frac{\partial f(x)}{\partial x} = -2 \underbrace{\underset{p\times n}{C'} \, \underset{n\times n}{K} \, \underset{n\times 1}{a}}_{p\times 1} + 2 C' K C x$$

$$\frac{\partial f(x)}{\partial x} = \underset{p\times 1}{0} \Rightarrow \underbrace{\underset{p\times n}{C'} \, \underset{n\times n}{K} \, \underset{n\times p}{C}}_{p\times p} x = C' K a$$

Se $C'KC$ é de posto completo ($\Rightarrow$ inversível)

$x_0 = (C'KC)^{-1} C'K a$ é solução de $\frac{\partial f(x)}{\partial x} = 0$.

Verifica-se que se K é positiva definida e C é de posto p então $\frac{\partial^2 f}{\partial x^2} = H(x) = 2C'KC$ é positiva definida e x_0 é ponto de mínimo de $f(x)$.

Maximização sujeita a constantes

Maximizar ou minimizar $f(x)$ sujeito à restrição $g(x) = c \;\forall x$.

Constrói-se a nova função

$h(x, \lambda) = f(x) - \lambda\,[g(x) - c]$

λ - multiplicador de Lagrange (λ escalar).

Obs: Maximizar ou minimizar $f(x)$ é equivalente a " " " $h(x,\lambda)$ pois, válida a restrição, $g(x) = c$ e $h(x,\lambda) = f(x)$.

$$\frac{\partial h(x,\lambda)}{\partial x} = \frac{\partial f(x)}{\partial x} - \lambda \frac{\partial g(x)}{\partial x} = \underset{p \times 1}{0}.$$

$$\frac{\partial h(x,\lambda)}{\partial \lambda} = -g(x) + c = 0 \left[\Rightarrow g(x) = c \text{, restrição} \right].$$

Resolve-se em x a equação $\frac{\partial h(x,\lambda)}{\partial x} = 0$, após eliminação de λ através de manipulações algébricas.

Ex: Obter o máximo da forma quadrática

$f(x) = x' A x$, onde $\underset{p \times p}{A}$ é simétrica, positiva definida, sujeito à restrição $x'x = 1$.

$$h(x, \lambda) = x' A x - \lambda \left[x'x - 1 \right]$$

$$\frac{\partial h(x,\lambda)}{\partial x} = 2 \underset{p\times p}{A} \underset{p\times 1}{x} - 2\lambda \underset{p\times 1}{x} = \underset{p\times 1}{0} .$$

$$\left[A - \lambda I_p \right] x = 0_{p\times 1} .$$

As soluções em x dessa equação são os vetores característicos de A. (λ é a raiz característica) Pre-multiplicando por x' temos

$$x' A x - \lambda \underbrace{x'x}_{1} = 0_{1\times 1} \Rightarrow$$

$$\Rightarrow \lambda = x' A x .$$

Como $x' A x$ é a quantidade a ser maximizada, tomar λ como a maior raiz característica de A e x_0, ponto de máximo de $f(x)$ é o vetor característico associado.

Para minimizar $f(x)$, x_0 seria o vetor característico associado à menor raiz característica de A.

Decomposição de Matrizes

1) Decomposição Espectral

Qualquer matriz simétrica A, $p \times p$, pode ser escrita como

$$A = \Gamma D \Gamma'$$

$$D = \text{diag}(\lambda_1, \lambda_2 \ldots \lambda_p)$$

$$\Gamma = [x_1 \;\; x_2 \ldots x_p]$$

$[A - \lambda_i I]\, x_i = 0$ tem infinitas soluções,
x_i solução $\Rightarrow K x_i$ solução

onde $\lambda_1, \lambda_2 \ldots \lambda_p$ são as raízes características de A e $x_1, x_2 \ldots x_p$ são os correspondentes vetores característicos "normalizados" $[\, x_i' x_i = 1, \; i = 1, 2 \ldots p \,]$

$$\Gamma = \begin{array}{c} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{p1} & x_{p2} & & x_{pp} \end{bmatrix} \\ \begin{array}{cccc} x_1 \quad & x_2 & \quad\quad & x_p \end{array} \end{array}$$

Verifica-se que

$$\Gamma D \Gamma' = \lambda_1 \underbrace{\underset{p\times 1}{x_1}\,\underset{1\times p}{x_1'}}_{p\times p} + \lambda_2 x_2 x_2' + \dots + \lambda_p x_p x_p'$$

Obs: a) Se x_i e x_j são dois vetores característicos associados a raízes características diferentes de uma matriz simétrica, então

$$\underset{1\times p}{x_i'}\ \underset{p\times 1}{x_j} = \underset{1\times 1}{0}$$

Dizemos que x_i e x_j são ortogonais.

b) A matriz Γ é ortogonal.

Γ é matriz ortogonal $\Longleftrightarrow$ $x_i' x_i = 1$

$x_i' x_j = 0 \quad i \neq j.$

Propriedades da Matriz ortogonal

$P_{n\times n}$ ortogonal $\Longleftrightarrow P^{-1} = P'$

P é ortogonal $\Longleftrightarrow P'P = I$

P é ortogonal $\Longrightarrow \det(P) = \pm 1$

2) Decomposição de Cholesky

Toda matriz positiva definida A pode ser escrita como

$$A = WW'$$

onde W é não singular.

Essa decomposição é a base do procedimento de mínimos Quadrados Generalizados da Regressão.

3) Decomposição em Valores Singulares

Toda matriz X nxp pode ser decomposta como

$$\underset{n\times p}{X} = \underset{n\times p}{U}\,\underset{p\times p}{D}\,\underset{p\times p}{T'}$$

em que U é $n \times p$, T é $p \times p$, $U'U = I$, $T'T = I$

e D é uma matriz diagonal $p \times p$ com elementos da diagonal principal $\mu_j \geq 0$ $j = 1, 2, \ldots p$.

μ_j são denominados valores singulares de X.

T é a matriz cujas colunas são os vetores característicos de $X'X$

U é a matriz cujas colunas são os vetores característicos de XX'.

Além disso, verifica-se que o quadrado dos valores singulares de X (μ_j^2) são as raízes características de $X'X$.

$$X'X = (UDT')'\,UDT' = TD'\underbrace{U'U}_{I}DT' = T\underbrace{D'D}_{D\text{diag}}T' =$$

$$= \underbrace{TD^2T'}_{\text{Decomposição espectral de } X'X}$$

Decomposições singulares são úteis na análise de multicolinearidade em modelos de regressão.

Exercício

Determine uma decomposição singular da matriz

$$A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}.$$