Decomposições Matriciais

Prof. Caio Azevedo

- Reescrever uma matriz (que atenda certos requisitos) como produto (em geral) de outras matrizes com certas propriedades de interesse.
- Utilidade:
  - Demonstraçãoes (obtenções) de resultados teóricos (propriedades).
  - Melhor desempenho de outros algoritmos computacionais (integração, maximização, simulação).
- Decomposições importantes:
  - Cholesky.
  - QR.
  - Valor singular (e espectral).

- Seja $\boldsymbol{\Sigma}_{p \times p}$, simétrica e positiva definida (todos os seus auto-valores são positivos).
- Exemplo: Matriz de covariâncias e matriz de correlações.
- Definição: Dada uma matriz $\boldsymbol{\Sigma}$ existe uma matriz $\mathbf{L}$, diagonal inferior com os valores da diagonal estritamente positivos, $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}'$.
- Tal decomposição é única. A demonstração pode ser feita via decomposição $\mathbf{LU}$.
- Existem alguns algoritmos que possibilitam a obtanção de $\mathbf{L}$.

Estrutura

- De um modo geral

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \dots & \sigma_p^2 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \dots & l_{pp} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & \dots & l_{2p} \\ 0 & l_{22} & \dots & l_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

Exemplo

- Exemplo:

$$\mathbf{\Sigma} = \begin{bmatrix} 3 & 1 & 0.5 \\ 1 & 2 & 0.8 \\ 0.5 & 0.8 & 4 \end{bmatrix}$$

$$\mathbf{L} \approx \begin{bmatrix} 1,732 & 0 & 0 \\ 0,577 & 1,291 & 0 \\ 0,289 & 0,491 & 1,917 \end{bmatrix}$$

Aplicações

- Métodos (algoritmos) de estimação:
  - Em geral, sempre que o método envolver inversão de matrizes positivas definidas. A inversão de $\mathbf{LL}'$ é mais estável.
  - Mínimos quadrados generalizados (resolução de sistemas lineares).
  - Mínimos quadrados reponderados.
  - Máxima verossimilhança (utilização de algoritmos de maximização: Newton-Raphson, Escore de Fisher, BFGS).
- Análise de componentes principais.
- Simulação Estocástica.

Sistemas lineares

- Considere o sistema $\mathbf{Ax} = \mathbf{b}$.
- Caso a seja inversível a solução é dada por $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.
- Inverter a matriz $\mathbf{A}$ pode ser muito despendioso computacionalmente.

  Por outro lado, se a matriz $\mathbf{A}$ for diagonal superior ou inferior, a solução torna-se mais simples.

A diagonal inferior

- Considere

$$\begin{bmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$$

- Solução $x_i = \dfrac{b_i - \sum_{j=1}^{i-1} a_{ij}x_j}{a_{ii}}$

A diagonal superior

- Considere

$$\begin{bmatrix} a_{11} & a_{21} & \dots & a_{2p} \\ 0 & a_{22} & \dots & a_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$$

- Solução $x_i = \dfrac{b_i - \sum_{j=i+1}^{p} a_{ij}x_j}{a_{ii}}$

Aplicação da decomposição de Cholesky

- Ao invés de calcular $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ resolve-se o sistema em duas etapas.
- Primeiro $\mathbf{L}\mathbf{y} = \mathbf{b}$, depois $\mathbf{L}'\mathbf{x} = \mathbf{y}$. Mutiplicando-se esta última por $\mathbf{L}$ recupera-se $\mathbf{A}\mathbf{x} = \mathbf{b}$.
- Algoritmo
  1. Resolver $\mathbf{L}\mathbf{y} = \mathbf{b}$.
  2. Depois, resolver $\mathbf{L}'\mathbf{x} = \mathbf{y}$.

Aplicação no modelo de regressão normal linear

- Modelo

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\xi}$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & X_{21} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{np} \end{bmatrix}, \boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}$$

- Suposição $\boldsymbol{\xi} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.
- Estimador de mínimos quadrados de $\boldsymbol{\beta}$, minimizar $(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$, equivale a resolver $(\mathbf{X}'\mathbf{X})\beta = \mathbf{X}'\mathbf{Y}$.

Cont.

- Solução natural $\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Em geral $\mathbf{X}'\mathbf{X}$ é positiva definida.
- Seja $\mathbf{L}_x$ a decomposição de Cholesky de $\mathbf{X}'\mathbf{X}$. Então, pode-se fazer:
  1. Resolver $\mathbf{L}_x\boldsymbol{\gamma} = \mathbf{XY}$
  2. Depois, resolver $\mathbf{L}_x'\boldsymbol{\beta} = \boldsymbol{\gamma}$
- Pode ser bastante útil quando a dimensão de $\mathbf{X}'\mathbf{X}$ for grande e/ou for muito instável sua inversão.

Estimação por máxima verossimilhança

- Seja $X_1, ..., X_n \overset{iid}{\sim} X$, $X \sim F_X(, \boldsymbol{\theta})$.
- Verossimilhança $L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_X(x_i; \boldsymbol{\theta})$.
- Logverossimilhança $l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln f_X(x_i; \boldsymbol{\theta})$.
- Exemplo, $X \sim \text{gama}(r, \lambda), \mathcal{E}(X) = r\lambda$.
- Verossimilhança

$$L(\boldsymbol{\theta}) = \frac{1}{\Gamma(r)^n} \lambda^{nr} e^{-n\overline{x}/\lambda} \prod_{i=1}^{n} x_i^{r-1}$$

.

Cont.

- Maximizar $l(\theta)$ em relação à θ.
- Resolver $\mathbf{S}(\widehat{\theta}) = \mathbf{0}$.
- Em geral, nãp há solução analítica explícita.
- Algoritmo Escore de Fisher.
- Fazer, m= 0,1,2,... até que algum critério de convergência seja satisfeito.

$$\theta^{(m+1)} = \boldsymbol{\theta}^{(m)} + \mathbf{I}(\boldsymbol{\theta}^{(m)})\mathbf{S}(\boldsymbol{\theta}^{(m)})$$

- $\mathbf{I}(\boldsymbol{\theta}) = -\mathcal{E}(\mathbf{H}(\boldsymbol{\theta}))$, informação de Fisher.

Cont.

- Seja $\mathbf{L}$ a decomposição de Cholesky de $\mathbf{I}(\theta)$.
- Temos que $\mathbf{I}(\theta)^{-1} = (\mathbf{L}')^{-1}\mathbf{L}^{-1}$.
- A inversão da informação de Fisher pode ser instável, principalmente em relação à $\mathbf{L}$
- A inversão de matrizes diagonais inferiores/superiores é menos dispendiosa computacionalmente.

Aplicação em modelos lineares generalizados

- Seja

$$
\begin{aligned}
y_i &\sim \text{FE}(\mu_i, \phi) \\
\mu_i &= g^{-1}(\mathbf{X}_i \boldsymbol{\beta})
\end{aligned}
$$

  $\mathbf{X}_i$ é uma determinada linha de $\mathbf{X}$.

- Estimação por máxima verossimilhança: mínimos quadrados reponderados.

$$
\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{(m)}\mathbf{z}^{(m)}
$$

  m=1,2,.... (até obter-se convergência).

Cont.

- Analogamente ao caso anterior, pode-se considerar a decomposição de Cholesky ($\mathbf{L}$) de ($\mathbf{X}'\mathbf{W}\mathbf{X}$), em cada passo do processo iterativo.
- A inversão de $\mathbf{L}$, em cada passo do algoritmo.
- Mais rapidez e estabilidade.
- Quando a matriz $\mathbf{X}$ é composta de variáveis correlacionadas (Pearson), as matrizes ($\mathbf{X}'\mathbf{X}$) e ($\mathbf{X}'\mathbf{W}\mathbf{X}$) pode não ser inversíveis.

Simulação de uma Normal multivariada

- $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\mu} \in \mathcal{R}^p, \boldsymbol{\Sigma} > 0.$
- Sabemos que $\mathbf{X} = \boldsymbol{\Psi}\mathbf{Z} + \boldsymbol{\mu}$, $\boldsymbol{\Psi} = \text{Cholesky}(\boldsymbol{\Sigma})$, $\mathbf{Z} = (Z_1, ..., Z_n), Z_i \overset{i.i.d.}{\sim} N(0,1), i = 1, ..., p.$
- $\mathbf{x} = \boldsymbol{\Psi} F_Z^{-1}(\mathbf{u}) + \boldsymbol{\mu}, \mathbf{u} = (u_1, ..., u_p).$

Definição

- Seja $\mathbf{X}_{n\times p}$, com as colunas linearmente independentes.
- Exemplo: matriz de planejamento num modelo de regressão (sem multicolinariedade).
- Definição: dada uma matriz $\mathbf{X}$ existem duas matrizes, $\mathbf{Q}$ e $\mathbf{R}$, tais que as colunas de $\mathbf{Q}$ são vetores ortonormais e $\mathbf{R}$ é uma matriz triangular superior.
- Algoritmo básico:
  1. Gera-se $\mathbf{Q}$ através das colunas de $\mathbf{X}$ através do processo de ortnormalização de Gram-Schimdt.
  2. Calcula-se $\mathbf{R} = \mathbf{Q}'\mathbf{X} \Rightarrow \mathbf{Q}'\mathbf{X} = \mathbf{Q}'\mathbf{Q}\mathbf{R} \Rightarrow \mathbf{R} = \mathbf{Q}'\mathbf{X}$.

Cont.

- Seja

$$\mathbf{X} = \begin{bmatrix} 1 & -1 \\ 3 & 4 \\ 4 & 5 \end{bmatrix}$$

$$\mathbf{Q} \approx \begin{bmatrix} -0,196 & 0,977 \\ -0,588 & -0,188 \\ -0,784 & -0,102 \end{bmatrix} ; \mathbf{R} \approx \begin{bmatrix} -5,100 & -6,079 \\ 0 & -2,224 \end{bmatrix}$$

Aplicações

- Métodos (algoritmos) de estimação:
    - Mínimos quadrados generalizados (resolução de sistemas lineares).
    - Mínimos quadrados reponderados.
    - Máxima verossimilhança (utilização de algoritmos de maximização: Newton-Raphson, Escore de Fisher, BFGS).
- Análise de componentes principais.
- Simulação MCMC.

Mínimos quadrados

- Do modelo linear normal, temos que o EMQ pode ser reescrito como

$$
\begin{aligned}
(\mathbf{X}'\mathbf{X})\beta &= \mathbf{X}'\mathbf{Y} \Rightarrow ((\mathbf{QR})'\mathbf{QR})\beta = \mathbf{X}'\mathbf{Y} \Rightarrow (\mathbf{R}'\mathbf{R})\beta = \mathbf{R}'\mathbf{Q}'\mathbf{Y} \\
\beta &= \mathbf{R}^{-1}\mathbf{Q}'\mathbf{Y}
\end{aligned}
$$

Mínimos quadrados reponderados

- A expressão, dado $\mathbf{W} = \mathbf{L}\mathbf{L}'$ é pode ser reescrita como

$$
\begin{aligned}
\widehat{\beta} &= (\mathbf{X}'\mathbf{L}\mathbf{L}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{L}\mathbf{L}'\mathbf{z} = ((\mathbf{X}^*)'\mathbf{X}^*)^{-1}(\mathbf{X}^*)'\mathbf{L}'\mathbf{z} \\
&= (\mathbf{R}'\mathbf{Q}'\mathbf{Q}\mathbf{R})^{-1}\mathbf{R}'\mathbf{Q}'\mathbf{L}'\mathbf{z} = (\mathbf{R}')^{-1}\mathbf{Q}'\mathbf{L}'\mathbf{z}
\end{aligned}
$$

Simulação de Monte Carlo via Cadeias de Markov

- Simular iterativamente

1. $Z_i \sim N_{[a,b]}(\boldsymbol{\beta}'\boldsymbol{\beta}, \sigma_i^2)$.
2. $\boldsymbol{\beta} \sim N_p((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}, \boldsymbol{\Sigma})$.

Simulação de Monte Carlo via Cadeias de Markov

- Simular iterativamente

1. $Z_i \sim N_{[a,b]}(\boldsymbol{\beta}'\boldsymbol{\beta}, \sigma_i^2)$.
2. $\boldsymbol{\beta} \sim N_p((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}, \boldsymbol{\Sigma})$.

Definição

- Seja $\mathbf{A}_{(I \times J)}$, então podemos escrever :

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}'$$

em que

$\mathbf{U}$: colunas formadas pelos autovetores (ortonormalizados) de $\mathbf{AA}'$.

$\mathbf{V}$: coluna formadas pelos autovetores (ortonormalizados) de $\mathbf{A}'\mathbf{A}$.

$\boldsymbol{\Lambda}$: matriz diagonal com sua diagonal dada por $(\lambda_1, \lambda_2, ...\lambda_{min(I,J)})'$, que são as raízes quadradas positivas dos autovalores maiores que zero, obtidos a partir das matrizes $\mathbf{AA}'$ ou $\mathbf{A}'\mathbf{A}$ (os autovalores maiores que zero são iguais).

Aplicações

- Redução de dados (vaiáveis e observações) (ACP, AF).
- Mínimos quadrados.
- Mínimos quadrados reponderados.
- Análise de correspondência.
- Processamento e compressão de imagens
- Detecção de outliers.