Análise Multivariada
Decomposição em Valores Singulares

Prof. George von Borries

Departamento de Estatística
Universidade de Brasília

2023

Decomposição em Valores Singulares (SVD)

A Decomposição Espectral (SD) de uma matriz quadrada $\mathbf{A}_{d\times d}$ pode ser expressa por $\mathbf{A} = \mathbf{X}\mathbf{D}\mathbf{X}^{\mathsf{T}}$ em que

- $\mathbf{X} = \left[\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(\mathrm{d})}\right]$, é a matriz de autovetores normalizados de $\mathbf{A}$;
- $\mathbf{X}\mathbf{X}^{\mathsf{T}} = \mathbf{X}^{\mathsf{T}}\mathbf{X} = \mathbf{I}$
- $\mathbf{A}\mathbf{X}_{(j)} = \lambda_j \mathbf{X}_{(j)}$, sendo λ_j o j-ésimo autovalor de $\mathbf{A}$;
- $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{\mathrm{d}})$.

A SVD é uma generalização da SD para matrizes não quadradas.

Definição (Decomposição em Valores Singulares)

Uma matriz $\mathbf{P}_{n\times d}$ *(valores reais) pode ser sempre escrita como*

$$\mathbf{P} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

em que

- $\mathbf{U}$ *é uma matriz de colunas ortonormais contendo* n *autovetores da matriz* $\mathbf{P}\mathbf{P}^T$ *("left singular vectors").*
- $\mathbf{V}$ *é uma matriz de colunas ortonormais contendo* d *autovetores da matriz* $\mathbf{P}^T\mathbf{P}$ *("right singular vectors").*
- $\mathbf{D}$ *é uma matriz* $n \times d$ *diagonal com* $\sqrt{\lambda_i}$ *em que* λ_i'*s são os* $\min\{n, d\}$ *maiores autovalores de* $\mathbf{P}\mathbf{P}^T$ *ou* $\mathbf{P}^T\mathbf{P}$*, chamados de valores singulares.*

- Por convenção, as colunas de **U**, **V** e **D** são apresentadas em ordem não crescente de valores singulares.
- A decomposição apresentada é também chamada de SVD Completa.

Note que,

$$\begin{aligned} \mathbf{P} &= \mathbf{UDV}^\top \quad \text{e} \\ \mathbf{P}^\top &= (\mathbf{UDV}^\top)^\top = (\mathbf{V}^\top)^\top\mathbf{D}^\top\mathbf{U}^\top = \mathbf{VDU}^\top \end{aligned}$$

Então,

$$\begin{aligned} \mathbf{P}^\top\mathbf{P} &= \mathbf{VDU}^\top\mathbf{UDV}^\top \\ &= \mathbf{VDDV}^\top = \mathbf{VD}^2\mathbf{V}^\top \end{aligned}$$

Note ainda que,

$$\mathbf{P}^\top\mathbf{PV} = \mathbf{VD}^2\mathbf{V}^\top\mathbf{V} = \mathbf{VD}^2$$

indicando que $\mathbf{V}$ é a matriz de autovalores de $\mathbf{P}^\top\mathbf{P}$ e $\mathbf{D}^2$ é a matriz de autovalores correspondentes.

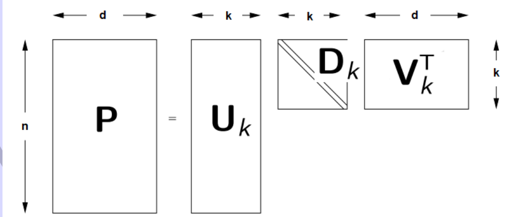
O mesmo raciocínio pode ser feito partindo de $\mathbf{PP}^\top$ e concluindo que $\mathbf{U}$ é a matriz de autovalores de $\mathbf{PP}^\top$ e $\mathbf{D}^2$ é a matriz de autovalores correspondentes.

Observações:

- A SVD truncada descarta os elementos de **D** em que $\lambda_i = 0$ ou muito pequenos. Neste caso, são descartados também os autovetores correspondentes de **U** e **V**.
- Considere que mantemos apenas os k maiores elementos da diagonal de **D**. Podemos representar a SVD_k por

$$\mathbf{P} \approx \mathbf{U}_k \mathbf{D}_k \mathbf{V}_k^T$$

ou



- Esta abordagem resulta em redução de dimensionalidade de **P**.
- $\text{tr}(\mathbf{D}_r)/\text{tr}(\mathbf{D})$ fornece a energia (ou informação) retida pela SVD truncada.

Representação Gráfica: Biplot de Gabriel

Biplots representam dados de uma matriz $\mathbf{S}$ *através de sua decomposição no produto de duas matrizes:* $\mathbf{S} = \mathbf{X}\mathbf{Y}^T$.

O nome Biplot se refere ao fato de que **dois** *grupos de pontos da matriz original (linhas e colunas) são visualzados através do produto escalar.*

Referências:

- Gabriel, K.R. (1971). "The Biplot Graphic Display of Matrices with Application to Principal Component Analysis."*Biometrika, 58, 453-467.*
- *Greenacre, M. (2010). Biplots in Practice. Disponível em* www.fbbva.es/microsite/multivariate-statistics/biplots.html

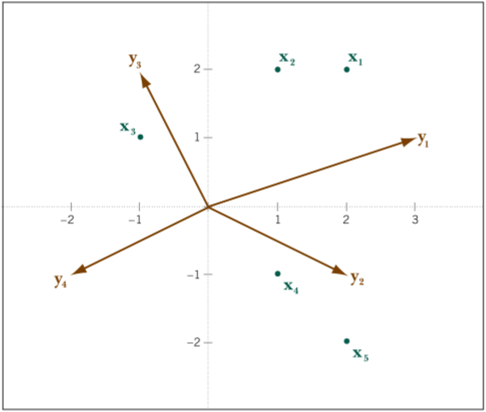
Exemplo: Matriz original e decomposição.

$$\begin{pmatrix} 8 & 2 & 2 & -6 \\ 5 & 0 & 3 & -4 \\ -2 & -3 & 3 & 1 \\ 2 & 3 & -3 & -1 \\ 4 & 6 & -6 & -2 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 1 & 2 \\ -1 & 1 \\ 1 & -1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 3 & 2 & -1 & -2 \\ 1 & -1 & 2 & -1 \end{pmatrix}$$

$$\mathbf{S} = \mathbf{X}\mathbf{Y}^{\mathsf{T}} = \begin{pmatrix} \mathbf{x}_1^{\mathsf{T}} \\ \mathbf{x}_2^{\mathsf{T}} \\ \mathbf{x}_3^{\mathsf{T}} \\ \mathbf{x}_4^{\mathsf{T}} \\ \mathbf{x}_5^{\mathsf{T}} \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \mathbf{y}_3 & \mathbf{y}_4 \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^{\mathsf{T}}\mathbf{y}_1 & \mathbf{x}_1^{\mathsf{T}}\mathbf{y}_2 & \mathbf{x}_1^{\mathsf{T}}\mathbf{y}_3 & \mathbf{x}_1^{\mathsf{T}}\mathbf{y}_4 \\ \mathbf{x}_2^{\mathsf{T}}\mathbf{y}_1 & \mathbf{x}_2^{\mathsf{T}}\mathbf{y}_2 & \mathbf{x}_2^{\mathsf{T}}\mathbf{y}_3 & \mathbf{x}_2^{\mathsf{T}}\mathbf{y}_4 \\ \mathbf{x}_3^{\mathsf{T}}\mathbf{y}_1 & \mathbf{x}_3^{\mathsf{T}}\mathbf{y}_2 & \mathbf{x}_3^{\mathsf{T}}\mathbf{y}_3 & \mathbf{x}_3^{\mathsf{T}}\mathbf{y}_4 \\ \mathbf{x}_4^{\mathsf{T}}\mathbf{y}_1 & \mathbf{x}_4^{\mathsf{T}}\mathbf{y}_2 & \mathbf{x}_4^{\mathsf{T}}\mathbf{y}_3 & \mathbf{x}_4^{\mathsf{T}}\mathbf{y}_4 \\ \mathbf{x}_5^{\mathsf{T}}\mathbf{y}_1 & \mathbf{x}_5^{\mathsf{T}}\mathbf{y}_2 & \mathbf{x}_5^{\mathsf{T}}\mathbf{y}_3 & \mathbf{x}_5^{\mathsf{T}}\mathbf{y}_4 \end{pmatrix}$$

Imagens de Greenacre, 2010.

Exemplo: Matriz original e decomposição.



$\mathbf{x}_1 = [2 \quad 2]^T$

$\mathbf{x}_2 = [1 \quad 2]^T$

$\mathbf{x}_3 = [-1 \quad 1]^T$

$\mathbf{x}_4 = [1 \quad -1]^T$

$\mathbf{x}_5 = [2 \quad -2]^T$

$\mathbf{y}_1 = [3 \quad 1]^T$

$\mathbf{y}_2 = [2 \quad -1]^T$

$\mathbf{y}_3 = [-1 \quad 2]^T$

$\mathbf{y}_4 = [-2 \quad -1]^T$

Imagem de Greenacre, 2010.

- O produto escalar entre dois vetores é o comprimento da projeção do primeiro vetor no segundo, multiplicado pelo comprimento do segundo vetor.

$$\boldsymbol{X}'\boldsymbol{Y} = ||\boldsymbol{X}||\,||\boldsymbol{Y}||\cos(\theta)$$

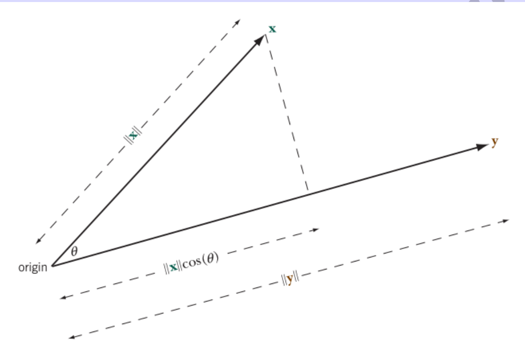


Imagem de Greenacre, 2010.

- O `R` tem a função `biplot` no pacote `stats`.

Exemplo 1: Avaliação de filmes por usuários:

Dois grupos de assinantes de uma plataforma de streaming avaliaram 3 filmes de ficção (F1,F2,F3) e dois filmes de romance (R1,R2). O grupo 1 é formado por 4 usuários jovens e o grupo 2 é formado por 3 usuários idosos.

$$
\begin{array}{c c}
 & \begin{array}{ccccc} \text{F1} & \text{F2} & \text{F3} & \text{R1} & \text{R2} \end{array} \\
\begin{array}{c} \text{J1} \\ \text{J2} \\ \text{J3} \\ \text{J4} \\ \text{I1} \\ \text{I2} \\ \text{I3} \end{array} &
\begin{pmatrix}
8 & 8 & 9 & 3 & 4 \\
7 & 6 & 7 & 3 & 5 \\
8 & 9 & 9 & 2 & 4 \\
9 & 8 & 9 & 1 & 2 \\
3 & 2 & 2 & 9 & 8 \\
2 & 1 & 2 & 8 & 8 \\
3 & 2 & 2 & 7 & 8
\end{pmatrix}
\end{array}
$$

Neste exemplo existem dois conceitos relacionados a filmes: Ficção e Romance
No `R` podemos utilizar a função `svd(M)`.

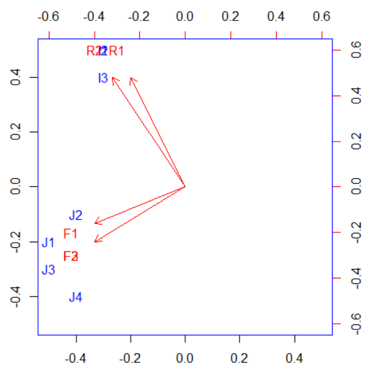
Exemplo 1: Avaliação de filmes por usuários

- Os conceitos (Ficção e Romance) envolvidos na matriz **P** são revelados por **U** e **V**. A importância relativa destes conceitos é revelada por **D**.
- **U** revela a avaliação feita pelos Grupos 1 e 2 (jovens e idosos).
- $\mathbf{V}^T$ revela a avaliação dos tipos de filmes (ficção e romance).
- Note que apesar de **P** possuir rank 5, apenas dois dos autovalores de $\mathbf{PP}^T$ possuem valores elevados, relativamente aos demais.
- Podemos então aplicar uma SVD_2 para aproximar **P**, i.e,.

$$\begin{pmatrix} 8 & 8 & 9 & 3 & 4 \\ 7 & 6 & 7 & 3 & 5 \\ 8 & 9 & 9 & 2 & 4 \\ 9 & 8 & 9 & 1 & 2 \\ 3 & 2 & 2 & 9 & 8 \\ 2 & 1 & 2 & 8 & 8 \\ 3 & 2 & 2 & 7 & 8 \end{pmatrix} \approx \begin{pmatrix} -0.5 & -0.2 \\ -0.4 & -0.1 \\ -0.5 & -0.3 \\ -0.4 & -0.4 \\ -0.3 & 0.5 \\ -0.3 & 0.5 \\ -0.3 & 0.4 \end{pmatrix} \begin{pmatrix} 32.5 & 0.00 \\ 0.0 & 15.8 \end{pmatrix} \begin{pmatrix} -0.5 & -0.5 & -0.5 & -0.3 & -0.4 \\ -0.2 & -0.3 & -0.3 & 0.6 & 0.6 \end{pmatrix}$$

Exemplo 1: Avaliação de filmes por usuários

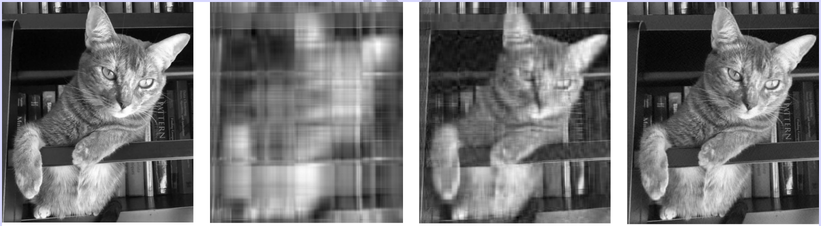
- Interpretação?



- A informação retida é de cerca de 94% da informação original.
- Observação: em estudos reais a interpretação/visualização geralmente não é trivial.

Exemplo 2: Redução de Dimensão de Imagens

- Uma imagem pode ser representada como uma matriz de píxeis. Esta matriz pode ser comprimida utilizando SVD.
- No caso de imagens coloridas, cada cor (*red, green, blue*) forma uma matriz em separado.
- O programa R `svdJudytte.R` apresenta a redução de uma imagem em matrizes de rank inferior. Alguns resultados, para diferentes reduções, são apresentados a seguir. A imagem original possui Rank 570.



Rank 570 Rank 5 Rank 20 Rank 99

Observações

- Uma importante área de aplicação é em dados de texto, como a comparação de documentos através da frequência de certas palavras. Considere, por exemplo, a seguinte matriz que apresenta a contagem de seis palavras (*lion, tiger, cheetah, jaguar, porsche, ferrari*) em seis documentos[1].

$$\mathbf{D} = \begin{array}{c} \text{Doc1} \\ \text{Doc2} \\ \text{Doc3} \\ \text{Doc4} \\ \text{Doc5} \\ \text{Doc6} \end{array} \begin{pmatrix} 2 & 2 & 1 & 2 & 0 & 0 \\ 2 & 3 & 3 & 3 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 3 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 2 & 2 & 2 \end{pmatrix}$$

  Podemos utilizar SVD para aproximar **P** e captar os conceitos escondidos (variáveis latentes) na matriz de dados. Este tipo de estudo é conhecido como Análise de Semântica Latente (*Latent Semantic Analysis*).

- SVD é uma técnica de redução dimensional mais apropriada para dados Esparsos. A análise de componentes principais (PCA) pode destruir a esparsidade dos dados.

- Remoção de ruídos também é uma aplicação de SVD.

[1] Aggarwall, 2021. Neste texto é utlizada a notação $\mathbf{D} = \mathbf{Q\Sigma P}^T$