UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

Tópicos Especiais em Fundamentos de Computação: Decomposição de matrizes

Alunos: André Riccioppo, Fábio Pereira, Gabriel M. Almeida e Hudson Romualdo

Goiânia-GO
2024

Sumário

1 Introdução

Este relatório é um entregável parte da avaliação da disciplina de Tópicos Especiais em Fundamentos de Computação, do programa de pós-graduação do Instituto de Informática da Faculdade Federal de Goiás (INF/UFG). Os tópicos apresentados nas próximas seções são baseadas no livro texto (DEISENROTH; FAISAL; ONG, 2020), base da disciplina, além de referências destacadas e listadas no corpo e no fim deste documento.

O restante deste trabalho está distribuído da seguinte forma. Capítulo 2 apresenta definições e aplicações de Determinante e Traço de matrizes. Capítulo 3 apresenta informações sobre Autovalores e Autovetores de matrizes. Capítulo 4 define Decomposição de Cholesky; Capítulo 5 apresenta os conceitos de Decomposição de Autovetores, Autovalores e Diagonalização de matrizes. Capítulo 6 define a Decomposição de Valores Singulares de matrizes. Capítulo 7 apresenta conceitos e aplicações práticas de Aproximação de Matrizes. Por fim, a Capítulo 8 apresentamos as considerações finais.

2 Determinante e Traço de matrizes

Um determinante é um objeto matemático na análise e solução de sistemas de equações lineares. Os determinantes são definidos apenas para matrizes quadradas $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, ou seja, matrizes com o mesmo número de linhas e colunas. O determinante de uma matriz quadrada $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ é uma função que mapeia $\boldsymbol{A}$ em um número real sendo representada por *det*($\boldsymbol{A}$) ou $|\boldsymbol{A}|$.

$$\det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

1 Teorema (Matriz Invertível)

Só é possível inverter qualquer matriz $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ se e somente se *det*($\boldsymbol{A}$) $\neq 0$.

Se $\boldsymbol{A}$ é uma matriz 1 x 1, ou seja, é um número escalar, temos: $\boldsymbol{A} = a \iff \boldsymbol{A}^{-1} = \frac{1}{a}$. Então $a \cdot \frac{1}{a} = 1$ se mantém, se e somente se $a \neq 0$.

Para matrizes 2×2 sabemos que $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$. Então, inverso de $\boldsymbol{A}$ é:

$$\boldsymbol{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

Portanto, $\boldsymbol{A}$ é inversível se e somente se: $a_{11}a_{22} - a_{12}a_{21} \neq 0$. O resultado dessa expressão é o determinante de uma matriz $\boldsymbol{A} \in \mathbb{R}^{n \times n}$.

$$\det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

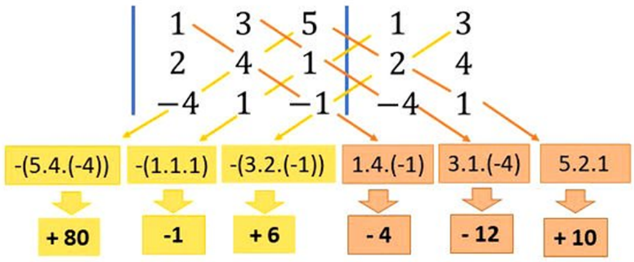
2 Teorema (Regra de Sarrus)

Para matrizes 3×3 utilizamos a Regra de Sarrus para obter o determinante, conforme abaixo:

Em uma das estratégias comumente utilizadas para a aplicação dessa regra, as duas primeiras colunas são duplicadas e multiplicam-se os elementos diagonais principais da matriz e

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23}$$
$$- a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}\,.$$

os soma aos produtos dos elementos formando linhas diagonais em direção oposta, subtraindo-os dessa soma.



Se uma matriz $\boldsymbol{T}$ é triangular, o determinante é o produto dos elementos da diagonal principal, ou seja:

$$\det(\boldsymbol{T}) = \prod_{i=1}^{n} T_{ii}$$

3 Teorema (Expansão de Laplace)

Para matrizes cuja ordem é superior a 3 o cálculo do determinante requer um algoritmo geral, a **Expansão de Laplace**. O uso do teorema reduz o problema de calcular o determinante de uma matriz de ordem superior a 3 aplicando a expansão recursivamente até o cálculo de determinantes de matrizes 2×2 no final.

Considere uma matriz $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. Então, para todo $j = 1, ..., n$:

- Expansão ao longo da coluna j

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j})$$

- Expansão ao longo da linha j

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(\boldsymbol{A}_{j,k})$$

Aqui $\boldsymbol{A}_{k,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ é a submatriz de $\boldsymbol{A}$ que obtemos ao excluir a linha k e a coluna j.

$$|\boldsymbol{A}_{3x3}| = \begin{vmatrix} \cancel{a_{11}} & \cancel{a_{12}} & \cancel{a_{13}} \\ \cancel{a_{21}} & a_{22} & a_{23} \\ \cancel{a_{31}} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$$

Uma demonstração do teorema pode ser observado no exemplo a seguir onde $det(\boldsymbol{A}) = 186$:

$$\boldsymbol{A} = [\boldsymbol{r},\ \boldsymbol{g},\ \boldsymbol{b}] = \begin{bmatrix} \cancel{2} & \cancel{6} & \cancel{1} \\ 0 & 1 & 4 \\ -8 & 0 & -1 \end{bmatrix}$$

$$\left.\begin{array}{l} (-1)^{1+1} \cdot 2 \cdot det\left(\begin{vmatrix} 1 & 4 \\ 0 & -1 \end{vmatrix}\right) = 1 \cdot 2 \cdot -1 = -2 \\[2ex] (-1)^{1+2} \cdot 6 \cdot det\left(\begin{vmatrix} 0 & 4 \\ -8 & -1 \end{vmatrix}\right) = -1 \cdot 6 \cdot 32 = -192 \\[2ex] (-1)^{1+3} \cdot 1 \cdot det\left(\begin{vmatrix} 0 & 1 \\ -8 & 0 \end{vmatrix}\right) = 1 \cdot 1 \cdot 8 = 8 \end{array}\right\} |-2 - 192 + 8| = 186$$

4 Propriedades do Determinante

Para $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, o determinante exibe as seguintes propriedades:

- O determinante de um produto de matrizes é o produto dos determinantes correspondentes, $det(\boldsymbol{AB}) = det(\boldsymbol{A})det(\boldsymbol{B})$.
- Os determinantes são invariantes à transposição, ou seja, $det(\boldsymbol{A}) = det(\boldsymbol{A}^T)$.
- Se $\boldsymbol{A}$ é regular (invertível), então $det(\boldsymbol{A}^{-1}) = \frac{1}{det(\boldsymbol{A})}$
- Matrizes semelhantes possuem o mesmo determinante. Portanto, para um mapeamento linear $\Phi : V \to V$ todas as matrizes de transformação $\boldsymbol{A}_\Phi$ tem o mesmo determinante. Assim, o determinante é invariante à escolha da base de um mapeamento linear.
- Adicionar um múltiplo de uma coluna/linha a outra não altera $det(\boldsymbol{A})$.

- A multiplicação de uma coluna/linha com $\lambda \in \mathbb{R}$ escala $det(\boldsymbol{A})$ por λ. Em particular, $det(\lambda \boldsymbol{A}) = \lambda^n det(\boldsymbol{A})$.
- Trocar duas linhas/colunas altera o sinal de $det(\boldsymbol{A})$.

5 Teorema (Posto da Matriz)

Uma matriz quadrada $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ tem $det(\boldsymbol{A}) \neq 0$ se e somente se $rk(\boldsymbol{A}) = n$, ou seja, se a matriz $\boldsymbol{A}$ tem **posto completo**. O posto é o número máximo de colunas/linhas linearmente independentes de uma matriz. Se a matriz for quadrada, e o posto for igual ao número de linhas, então a matriz tem posto completo.

6 Traço

O traço de uma matriz quadrada $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ é definido como:

$$\mathrm{tr}(\boldsymbol{A}) := \sum_{i=1}^{n} a_{ii} \,,$$

isto é, o traço é a soma dos elementos da diagonal principal de $\boldsymbol{A}$.

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

$$\mathrm{tr}\, A = a_{11} + a_{22} + a_{33} + a_{44}$$

O Traço satisfaz as seguintes propriedades:

- $tr(\boldsymbol{A} + \boldsymbol{B}) = tr(\boldsymbol{A}) + tr(\boldsymbol{B})$ para $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$
- $tr(\alpha\boldsymbol{A}) = \alpha \cdot tr(\boldsymbol{A})$, para $\alpha \in \mathbb{R}$ e $\boldsymbol{A} \in \mathbb{R}^{n\times n}$
- $tr(\boldsymbol{I}_n) = n$
- $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$ para $\boldsymbol{A} \in \mathbb{R}^{n\times k}, \boldsymbol{B} \in \mathbb{R}^{k\times n}$

As propriedades do traço dos produtos de matrizes são mais gerais. Especificamente, o traço é invariante sob permutações cíclicas, ou seja, $tr(\boldsymbol{AKL}) = tr(\boldsymbol{KLA})$ para matrizes $\boldsymbol{A} \in \mathbb{R}^{a\times n}, \boldsymbol{K} \in \mathbb{R}^{k\times l}, \boldsymbol{L} \in \mathbb{R}^{l\times a}$.

Esta propriedade generaliza para produtos de um número arbitrário de matrizes.

7 Polinômio Característico

Para $\lambda \in \mathbb{R}$ e uma matriz quadrada $\boldsymbol{A} \in \mathbb{R}^{n\times n}$

$$\begin{aligned} p\boldsymbol{A}(\lambda) &:= det(\boldsymbol{A} - \lambda\boldsymbol{I}) \\ &= c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_{n-1}\lambda^{n-1} + (-1)^n\lambda^n , \\ c_0, \ldots, c_{n-1} &\in \mathbb{R} \end{aligned}$$

é o polinômio característico de $\boldsymbol{A}$.

Adicionalmente:

$$c_0 = det(\boldsymbol{A})$$
$$c_{n-1} = (-1)^{n-1} tr(\boldsymbol{A})$$

Este polinômio característico nos permitirá calcular autovalores e autovetores.

3 Autovalores e Autovetores

Assim como o Determinante e o Traço, Autovalores e Autovetores são características de Matrizes Quadradas.

- Autovalores são valores numéricos que descrevem a magnitude da transformação efetuada por uma matriz em uma determinada direção.
- Autovetores são vetores que fornecem a direção da transformação na qual a matriz age apenas esticando ou comprimindo o vetor, sem alterar sua orientação.

Seja $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ uma matriz quadrada. Então $\lambda \in \mathbb{R}$ é um autovalor de $\boldsymbol{A}$ e $x \in \mathbb{R}^n \setminus \{0\}$ é o autovetor correspondente de $\boldsymbol{A}$ se: $\boldsymbol{A}x = \lambda\mathbf{x}$

As seguintes afirmações são equivalentes:

- λ é um autovalor de $\boldsymbol{A} \in \mathbb{R}^{n \times n}$
- Existe um $\boldsymbol{x} \in \mathbb{R}^n \setminus \{0\}$ tal que $\boldsymbol{A}x = \lambda\mathbf{x}$, ou de forma equivalente, $(\boldsymbol{A} - \lambda\boldsymbol{I}_n)x = 0$ pode ser resolvido de forma não trivial, ou seja, $x \neq 0$.
- *rk*$(\boldsymbol{A} - \lambda\boldsymbol{I}_n) < n$
- *det*$(\boldsymbol{A} - \lambda\boldsymbol{I}_n) = 0$

1 Colinearidade e Codireção

Dois vetores que apontam na mesma direção são chamados de codirecionados.

Dois vetores são colineares se e somente se esses dois vetores estiverem ao longo da mesma linha ou se forem paralelos entre si na mesma direção ou em direção oposta.

Se $\boldsymbol{x}$ é um autovetor de $\boldsymbol{A}$ associado ao autovalor λ, então para qualquer $c \in \mathbb{R}$ temos que $c\boldsymbol{x}$ é um autovetor de $\boldsymbol{A}$ com o mesmo autovalor, pois: $\boldsymbol{A}(c\boldsymbol{x}) = c\boldsymbol{A}\boldsymbol{x} = c\lambda\boldsymbol{x} = \lambda(c\boldsymbol{x})$

Assim, todos os vetores que são colineares a $\boldsymbol{x}$ também são autovetores de $\boldsymbol{A}$.

2 Teorema (Autoespaço e Espectro)

$\lambda \in \mathbb{R}$ é um autovalor de $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ se e somente se λ é uma raiz do polinômio característico $p_A(\lambda)$ de $\boldsymbol{A}$.

Considerando que uma matriz quadrada $\boldsymbol{A}$ tenha um autovalor λ_i. A multiplicidade algébrica de λ_i é o número de vezes que a raiz aparece no polinômio característico.

O conjunto de todos os autovetores de $\boldsymbol{A}$ associados a um autovalor λ abrange um subespaço de $\mathbb{R}^n$, que é chamado de **autoespaço** de $\boldsymbol{A}$ com relação λ e é representado por E_λ. O conjunto de todos os autovalores de $\boldsymbol{A}$ é chamado de **autoespectro**, ou apenas *espectro*, de $\boldsymbol{A}$.

Se λ é um autovalor de $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ então o autoespaço correspondente E_λ é o espaço de soluções do sistema homogêneo de equações lineares $(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = 0$. Geometricamente, o autovetor correspondente a um autovalor não nulo aponta em uma direção esticada pela transformação linear. O autovalor é o fator pelo qual ela é esticada. Se o autovalor for negativo, a direção é invertida.

3 Propriedades úteis

- Uma matriz $\boldsymbol{A}$ e sua transposta $\boldsymbol{A}^\top$ possuem os mesmos autovalores, mas não necessariamente os mesmos autovetores.
- O autoespaço E_λ é o espaço nulo de $\boldsymbol{A} - \lambda \boldsymbol{I}$, uma vez que:

$$
\begin{aligned}
\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x} &\iff \boldsymbol{A}\boldsymbol{x} - \lambda \boldsymbol{x} = 0 \\
&\iff (\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = 0 \iff \boldsymbol{x} \in \ker(\boldsymbol{A} - \lambda \boldsymbol{I}).
\end{aligned}
$$

- Matrizes semelhantes possuem os mesmos autovalores. Portanto, um mapeamento linear Φ possui autovalores independentes da escolha da base de sua matriz de transformação. Isso torna os autovalores, juntamente com o determinante e o traço, parâmetros característicos principais de uma aplicação linear, pois são todos invariantes em relação à mudança de base.
- Matrizes simétricas e definidas positivas sempre têm autovalores reais e positivos.

4 Intuição Gráfica em Duas Dimensões

Visão geral de cinco mapeamentos lineares e suas matrizes de transformação associadas $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ projetando 400 pontos codificados por cores $x \in \mathbb{R}^2$ (coluna da esquerda) em pontos de destino $\boldsymbol{A}_i\boldsymbol{x}$ (coluna da direita).

A coluna central retrata o primeiro vetor próprio, esticado por seu valor próprio associado λ_1, e o segundo vetor próprio esticado pelo seu valor próprio λ_1. Cada linha retrata o efeito de uma das cinco matrizes de transformação $\boldsymbol{A}_i$ com respeito à base padrão.

- $\boldsymbol{A}_1$: A direção dos dois autovetores corresponde aos vetores da base canônica em $\mathbb{R}^2$, ou seja, a dois eixos cardinais. O eixo vertical é ampliado por um fator de 2 (autovalor $\lambda_1 = 2$), e o eixo horizontal é comprimido por um fator de $\frac{1}{2}$ (autovalor $\lambda_2 = \frac{1}{2}$). O mapeamento preserva a área ($det(\lambda \boldsymbol{A}_1) = 1 = 2 \times \frac{1}{2}$).

- $\boldsymbol{A}_2$: corresponde a um mapeamento de cisalhamento, ou seja, ele cisalha os pontos ao longo do eixo horizontal para a direita se eles estiverem metade positiva do eixo vertical e para a esquerda, e vice-versa. Este mapeamento preserva a área ($det(\lambda \boldsymbol{A}_1) = 2$). O autovalor $\lambda_1 = \lambda_1 = 1$ é repetido e os autovetores são colineares (desenhados aqui para enfatizar em duas direções opostas). Isso indica que o mapeamento age apenas ao longo de uma direção (o eixo horizontal).

- $\boldsymbol{A}_3$: roda os pontos em $\frac{\pi}{6} rad = 30°$ no sentido anti-horário e possui apenas autovalores complexos, refletindo que o mapeamento é uma rotação (portanto, nenhum autovetor é traçado). Uma rotação precisa preservar o volume, portanto o determinante é igual a 1.

- $\boldsymbol{A}_4$: representa uma aplicação na base padrão que reduz um domínio bidimensional para uma dimensão. Uma vez que um autovalor é 0, o espaço na direção do vetor próprio (azul) correspondente a $\lambda_1 = 0$ colapsa, enquanto o vetor próprio ortogonal (vermelho) estica o espaço por um fator $\lambda_2 = 2$. Portanto, a área da imagem é 0.

- $\boldsymbol{A}_5$: é um mapeamento de cisalhamento e alongamento que redimensiona o espaço em 75% uma vez que $|det(\boldsymbol{A}_5)| = \frac{3}{4}$. Ele alonga o espaço ao longo do autovetor (vermelho) λ_2 por um fator de 1,5 e comprime-o ao longo do eigenvetor ortogonal (azul) por um fator de 0,5.

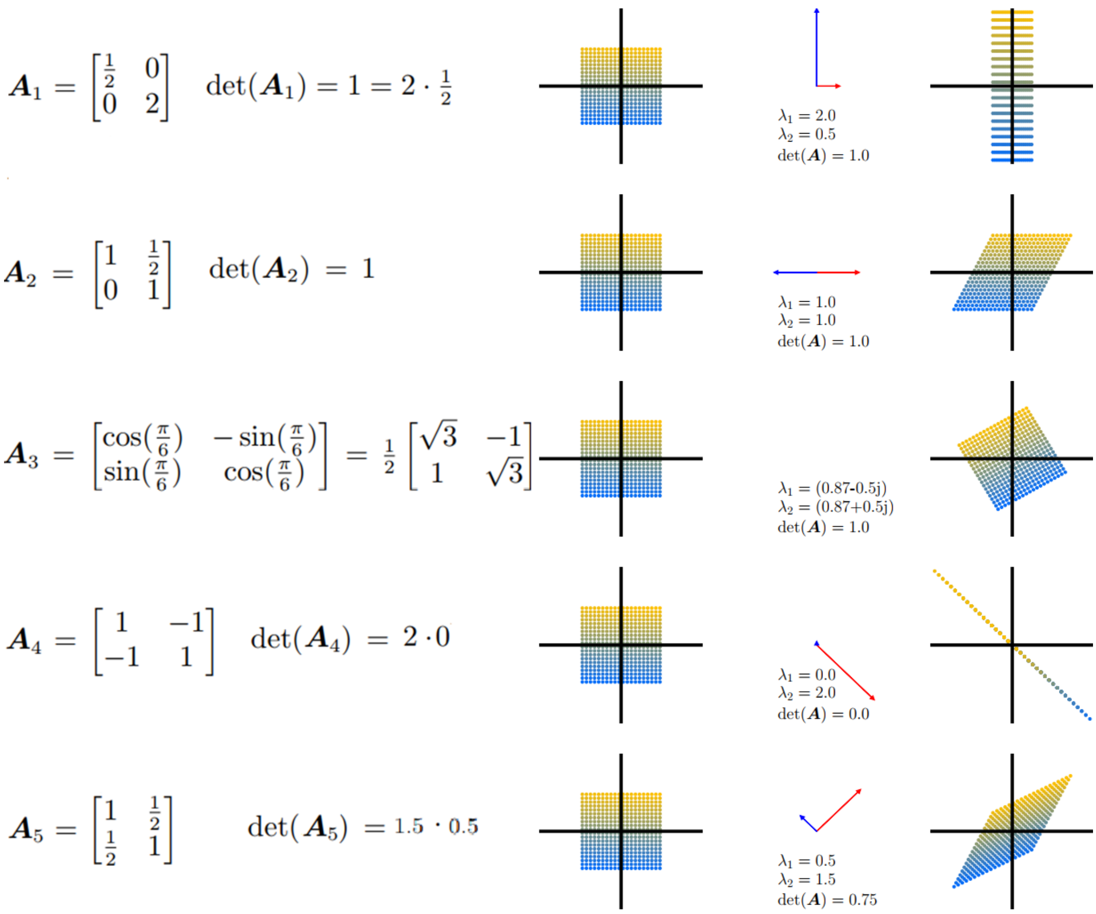
5 Teorema (Autovalores Independentes)

Os autovetores $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ de uma matriz $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ com *n* autovalores distintos $\lambda_1, ..., \lambda_n$ são linearmente independentes.

Um matriz $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ é defeituosa se ela possui menos de *n* autovetores linearmente independentes, não necessariamente com autovalores distintos, mas que ainda assim formam bases de $\mathbb{R}^n$.

6 Teorema (Matriz Simétrica)

Dada uma matriz $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, sempre podemos obter uma matriz simétrica, semidefinida positiva $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ definindo $\boldsymbol{S} := \boldsymbol{A}^\top \boldsymbol{A}$.



7 Teorema Espectral

Se $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ for simétrica, existe uma base ortonormal do espaço vetorial correspondente V que consiste de autovetores de $\boldsymbol{A}$, e cada autovalor é real. Uma implicação direta do teorema espectral é que a decomposição de autovalores de uma matriz simétrica $\boldsymbol{A}$ existe (com autovalores reais) e que podemos encontrar uma base ortogonal de autovetores de modo que $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top$, sendo que:

- $\boldsymbol{P}$ é a matriz cujas colunas são os autovetores de $\boldsymbol{A}$.
- $\boldsymbol{D}$ é a matriz diagonal formada pelos autovalores de $\boldsymbol{A}$.

8 Teorema (Determinante e Autovalores)

O determinante de uma matriz $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ é o produto de seus autovalores, ou seja, $det(\boldsymbol{A}) = \prod_{i=1}^{n} \lambda_i$ onde $\lambda_i \in \mathbb{C}$ são (possivelmente repetidos) autovalores de $\boldsymbol{A}$.

9 Teorema (Traço e Autovalores)

O traço de uma matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ é a soma de seus autovalores, ou seja, $tr(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i$, onde $\lambda_i \in \mathbb{C}$ são (possivelmente repetidos) autovalores de $\boldsymbol{A}$.

10 Conclusões

O conteúdo sobre as características de matrizes quadradas revela a essência intrínseca desses objetos matemáticos fundamentais.

Primeiramente, exploramos o determinante, uma medida que expressa como a matriz transforma o volume no espaço. O determinante é crucial para entender a inversibilidade da matriz e sua importância em sistemas lineares. A Regra de Sarrus e a Expansão de Laplace são ferramentas valiosas para calcular determinantes de matrizes 3x3 e maiores, respectivamente. Elas nos permitem encontrar determinantes de forma sistemática, mesmo para matrizes de tamanho considerável.

Em seguida, o traço, a soma dos elementos da diagonal principal da matriz, fornece informações sobre a soma dos autovalores, desempenhando um papel essencial na análise e manipulação de dados em uma variedade de aplicações.

Os autovalores e autovetores desempenham um papel central na teoria das transformações lineares, representando as direções e escalas nas quais uma matriz transforma o espaço. Os autoespaços são os espaços gerados pelos autovetores correspondentes aos mesmos autovalores, fornecendo uma decomposição útil das transformações lineares.

O teorema espectral revela uma conexão profunda entre as propriedades de uma matriz e suas características espectrais, afirmando que toda matriz simétrica pode ser diagonalizada por uma matriz ortogonal, destacando a importância dos autovalores e autovetores em sua forma mais geral.

Portanto, ao compreender e aplicar essas características e ferramentas associadas, somos capazes de explorar e analisar uma variedade de problemas de ciências de dados, ciência da computação, engenharia e outras disciplinas.

4 Decomposição de Cholesky

A decomposição de Cholesky ou fatoração de Cholesky é um processo de decomposição matricial aplicável a matrizes que possuem as seguintes características: São matrizes **positivas**, **definidas** e **simétricas**. A decomposição transforma a matriz original no produto de duas matrizes triangulares, sendo uma delas **triangular inferior** e a outra **triangular superior**, conforme mostra a figura abaixo.

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & \cdots & l_{n1} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & l_{nn} \end{bmatrix}$$

Figura 1 – Exemplo de uma matriz decomposta

1 Teorema (Decomposição de Cholesky)

Uma matriz simétrica, positiva, definida A pode ser fatorada no produto $A = L * L^T$, onde L é a matriz triangular inferior, com elementos da diagonal positivos. L é chamada de fator de Cholesky de A, e L é única.

Uma forma de observar o teorema de decomposição de Cholesky é visualizando a decomposição em si usando como exemplo uma matriz definida, positiva e simétrica 3×3 como no exemplo abaixo.

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \boldsymbol{LL}^{\top} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

Ao multiplicarmos as duas matrizes $(L * L^T)$ obtemos:

$$\boldsymbol{A} = \begin{bmatrix} l_{11}^2 & l_{21}l_{11} & l_{31}l_{11} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{31}l_{21} + l_{32}l_{22} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

Se compararmos o lado esquerdo com o lado direito, podemos ver que existe um padrão para encontrar os elementos da diagonal principal.

$$l_{11} = \sqrt{a_{11}}\,, \quad l_{22} = \sqrt{a_{22} - l_{21}^2}\,, \quad l_{33} = \sqrt{a_{33} - (l_{31}^2 + l_{32}^2)}$$

E o mesmo vale para os elementos abaixo da diagonal principal, que também seguem um padrão.

$$l_{21} = \frac{1}{l_{11}} a_{21}\,, \quad l_{31} = \frac{1}{l_{11}} a_{31}\,, \quad l_{32} = \frac{1}{l_{22}} (a_{32} - l_{31} l_{21})\,.$$

Com isso, podemos concluir que é possível calcular retroativamente quais são os componentes l_{ij} para L, se tivermos os valores a_{ij} para A, junto aos valores previamente calculados de l_{ij}, sendo assim um processo iterativo.

OBS: A decomposição de Cholesky também nos permite calcular determinantes de forma muito eficiente. Dada a decomposição de Cholesky $A = LL^\top$, sabemos que $\det(A) = \det(L)\det(L^\top) = (\det(L))^2$. Como L é uma matriz triangular, o determinante é simplesmente o produto das entradas de sua diagonal, de modo que $\det(A) = \prod_i l_{ii}^2$. Sendo assim, diversas bibliotecas de cálculo numérico utilizam a decomposição de Cholesky para realizarem a computação de forma mais eficiente.

2 Conclusões

Como visto acima a decomposição de Cholesky nos traz uma forma bastante interessante de decompor uma matriz em duas outras matrizes triangulares, uma delas triangular superior, e a outro triangular inferior.

Ao primeiro ver, pode parecer que decompor uma matriz já bem estabelecida em outras duas não faria muito sentido, no entanto, quando pensamos no exemplo dado na observação acima, poder calcular o determinante da matriz de forma rápida já mais do que traz benefícios para a utilização do método.

Em contrapartida, também vemos que o método apenas é aplicado com sucesso a matrizes que possuem características bem específicas, que são elas: serem positivas, definidas e simétricas. Nos casos da vida real isto nem sempre será verdade, e portanto será necessário utilizar outros métodos quando tratarmos de matrizes que possuam características diferentes das elencadas acima. Tais métodos serão vistos nas sessões mais abaixo.

5 Decomposição em autovalores, autovetores e diagonalização

1 Definição de matriz diagonalizável

Uma matriz $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ é diagonalizável se ela é similar a uma matriz diagonal, isto é, se existe uma matriz invertível $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ tal que $D = P^{-1}AP$.

Intuitivamente, a matriz diagonal é aquela cujos valores que não pertencem à diagonal principal são todos iguais a 0 (zero), assim como exemplificado na figura abaixo.

$$\boldsymbol{D} = \begin{bmatrix} c_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c_n \end{bmatrix}.$$

Figura 2 – Exemplo de uma matriz diagonal

Tais matrizes são importantes, pois por meio delas é possível calcular rapidamente determinantes, potências de matrizes, e matrizes inversas. O determinante pode ser visto como o produto dos elementos da diagonal ($Det(D) = c1 * c2 * ... * cn$), a potência D^k pode ser calculada elevando cada elemento diagonal à potência k, e a matriz inversa D^{-1} é o inverso de seus elementos diagonais, se todos eles forem diferentes de zero.

Um entendimento importante sobre diagonalização de matrizes é que o ato de diagonalizar é uma forma de expressar a mesma transformação linear que seria aplicada pela matriz A, porém em outra base, e esta base no final é formada pelos autovetores de A.

2 Teorema: Decomposição em autovalores e autovetores

Uma matriz quadrada $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ pode ser fatorada na forma

$$A = PDP^{-1}$$

onde $\boldsymbol{P} \in \mathbb{R}^{n \times n}$, e D é uma matriz diagonal, cujos elementos da diagonal são os autovalores de A, se e somente se os autovetores de A formarem uma base de $\mathbb{R}^n$

O teorema acima reforça o que falamos ao final da subseção anterior, que a diagonalização realiza uma transformação linear utilizando como base os autovetores da matriz A. Isso nos traz duas implicações: Apenas matrizes não defeituosas podem ser diagonalizadas, e que as colunas de P são os n autovetores da matriz A (**OBS:** Uma matriz é dita "não defeituosa"se ela possui um conjunto completo de autovalores e autovetores linearmente independentes).

Para matrizes simétricas podemos obter resultados ainda mais fortes para a decomposição em autovalores e autovetores, conforme mostra o teorema seguinte.

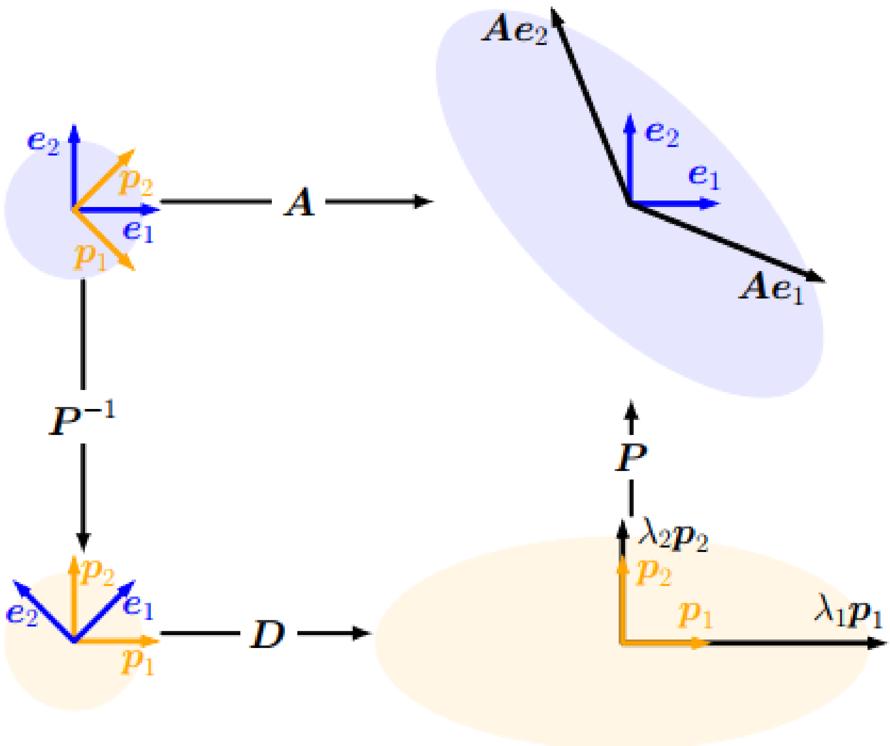
3 Teorema: Matrizes simétricas e diagonalização

Uma matriz simétrica $S \in \mathbb{R}^{n \times n}$ sempre pode ser diagonalizada.

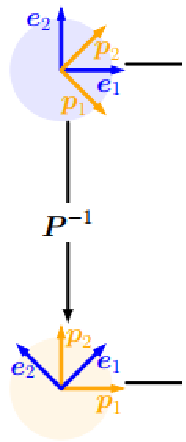
O teorema acima é uma consequência direta do teorema espectral. Ademais, o teorema espectral postula que podemos encontrar uma base ortonormal de autovetores de $\mathbb{R}^n$. Isto torna P uma matriz ortogonal, de forma que: $D = P^T AP$.

4 Intuição geométrica da decomposição em autovalores e autovetores

A imagem abaixo, mostra a sequência de passos das multiplicações de matrizes que precisamos operar para chegar ao resultado final.

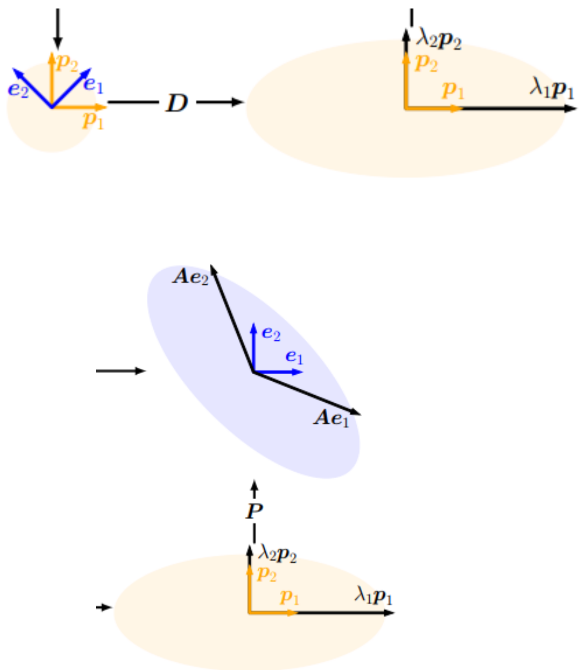


Podemos enxergar os passos por trás da decomposição em autovalores e autovetores como uma sequência de transformações. Do canto superior esquerdo para o canto inferior esquerdo: P^{-1} executa uma mudança de base (aqui desenhada em $\mathbb{R}^2$, e representada como uma operação de rotação) da base padrão para a base de autovetores.



Da parte inferior esquerda para a parte inferior direita: *D* realiza uma operação de escala ao longo dos autovetores ortogonais, representados aqui por um círculo esticado em uma elipse.

Por fim, da parte inferior direita para a parte superior direita: *P* desfaz a mudança de base (representada como uma rotação reversa) voltando ao sistema de coordenadas originais.

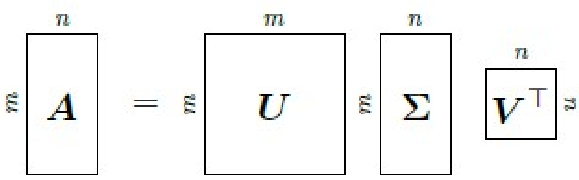


6 Decomposição de Valores Singulares

A decomposição em valores singulares (*Singular Value Decomposition* – SVD) é usada para permitir a análise de uma matriz usando partes mais simples e significativas. Ela pode ser aplicada a todos os tipos de matrizes, não apenas às matrizes quadradas, e sempre existe. Por esse motivo já foi chamada de teorema fundamental da álgebra linear.

1 Teorema

Dada uma matriz retangular $A \in \mathbf{R}^{m\,x\,n}$ de posto $r \in$ [0, min$(m, n]$, a SVD de A é uma decomposição na forma



em que

$U \in \mathbf{R}^{m\,x\,m}$ é uma matriz ortogonal com colunas $u_i, i = 1, \ldots, m$ e

$V \in \mathbf{R}^{n\,x\,n}$ é uma matriz ortogonal com colunas $v_j, i = 1, \ldots, n$.

Os vetores u_i são chamados vetores singulares à esquerda, e os vetores v_i são chamados vetores singulares à direita.

Σ é uma matriz diagonal $m\,x\,n$ com $\Sigma_{ii} = \sigma_i \geq 0$ e $\Sigma_{ij} = 0, i \neq j$, onde as entradas diagonais $\sigma_i, i = 1, \ldots, r$ de Σ são chamadas de valores singulares. Por convenção, os valores singulares são ordenados $\sigma_1 \geq \sigma_2 \geq \sigma_r \geq 0$.

$\Sigma \in \mathbf{R}^{m\,x\,n}$ é retangular e do mesmo tamanho que A. Isso significa que Σ possui uma submatriz diagonal que contém os valores singulares e pode precisar ser preenchida com zeros, se m $\neq$ n. Especificamente, se m > n, então a matriz Σ possui estrutura diagonal até a linha n e linhas formadas por 0 de n + 1 até m de forma que

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & 0 \end{bmatrix} \tag{6.1}$$

Se m < n, a matriz Σ possui uma estrutura diagonal até a coluna m e colunas formadas por 0 de m+1 até n:

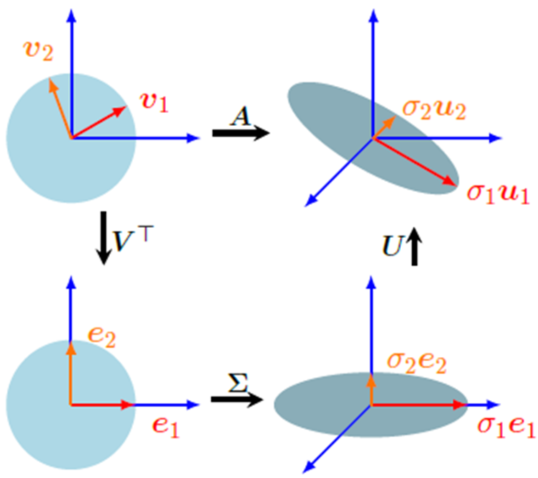
$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & \vdots & & \vdots \\ 0 & 0 & \sigma_m & 0 & \dots & 0 \end{bmatrix} \tag{6.2}$$

Então, para uma matriz A de posto r, apenas os r valores singulares são não nulos, enquanto os demais são zero. Isso significa que a matriz Σ é essencialmente uma matriz m x n, onde m é o número de linhas de A e n o número de colunas, mas apenas os r valores singulares são distintos de zero.

2 Interpretação geométrica da SVD

A SVD de uma matriz pode ser interpretado como a decomposição de uma transformação linear correspondente $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ em três operações. A SVD realiza uma mudança de base via $V^\top$ (rotação), seguido por uma escala e dilatação (ou contração) de sua dimensionalidade através da matriz de valor singular Σ. Ao final, realiza uma segunda mudança de base via U (nova rotação).

Então, dada uma matriz de transformação de uma transformação linear $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ relacionado às bases padrão B e C de $\mathbf{R}^n$ e $\mathbf{R}^m$, respectivamente. Sejam dadas, ainda, bases $\tilde{B}$ de $\mathbf{R}^n$ e $\tilde{C}$ de $\mathbf{R}^m$:



1. A matriz V faz a mudança de base no domínio $\mathbf{R}^n$ para $\tilde{B}$ (representado pelos vetores v_1 vermelho e v_2 laranja no gráfico de cima à esquerda para a base padrão B. $V^\top = V^{-1}$ realiza uma mudança de base de B para $\tilde{B}$. Os vetores vermelho e laranja se tornam alinhados à base canônica no gráfico de baixo à esquerda.

2. Tendo sido o sistema de coordenadas alterado para $\tilde{B}$, Σ escala as novas coordenadas pelos valores singulares σ_i (adicionando ou removendo dimensões).No exemplo, Σ é a matriz de

transformação de Φ com respeito a $\tilde{B}$ e $\tilde{C}$, representados pelos vetores vermelho e laranja sendo esticados e repousando no plano $e_1 - e_2$, mostrado no gráfico de baixo à direita.

3. U realiza a mudança de base no codomínio $\mathbf{R}^m$, representado pela rotação dos vetores vermelho e laranja do plano $e_1 - e_2$, conforme o gráfico de cima à direita.

A SVD expressa uma mudança de base tanto do domínio $\mathbf{R}^n$ como no codomínio $\mathbf{R}^m$. Isso é diferente da autodecomposição (que opera no mesmo espaço vetorial), onde a mesma mudança de base é aplicada e depois desfeita. O que torna a SVD especial é que essas duas bases diferentes estão simultaneamente ligadas pela matriz de valor singular Σ.

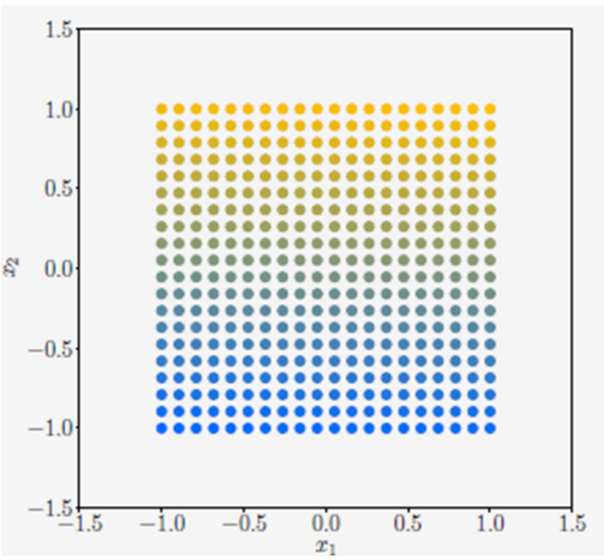
3 Exemplo de uma decomposição em valores singulares

Para melhor entendimento, apresenta-se um exemplo, onde serão aplicadas matrizes de transformação da SVD a um conjunto de vetores do $\mathbf{R}^2$ para visualizar o efeito de cada transformação de maneira mais clara.

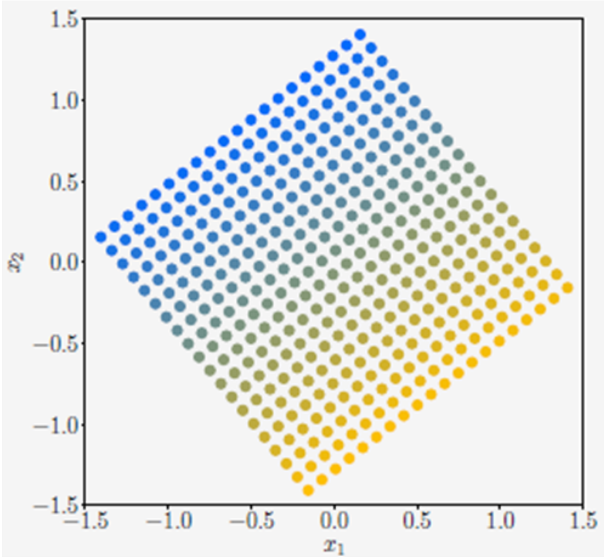
Considere um mapeamento de um grid quadrado de vetores $X \in \mathbf{R}^2$ que se encaixa em um quadrado tamanho 2 x 2 centralizado na origem. Usando a base padrão, os vetores são mapeados usando:

$$A = \begin{bmatrix} 1 & -0.8 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = U\Sigma V^\top = \begin{bmatrix} -0,79 & 0 & -0,62 \\ 0,38 & -0,78 & -0,49 \\ -0,48 & -0,62 & 0,62 \end{bmatrix} \begin{bmatrix} 1,62 & 0 \\ 0 & 1,0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0,78 & 0,62 \\ -0,62 & -0,78 \end{bmatrix} \tag{6.3}$$

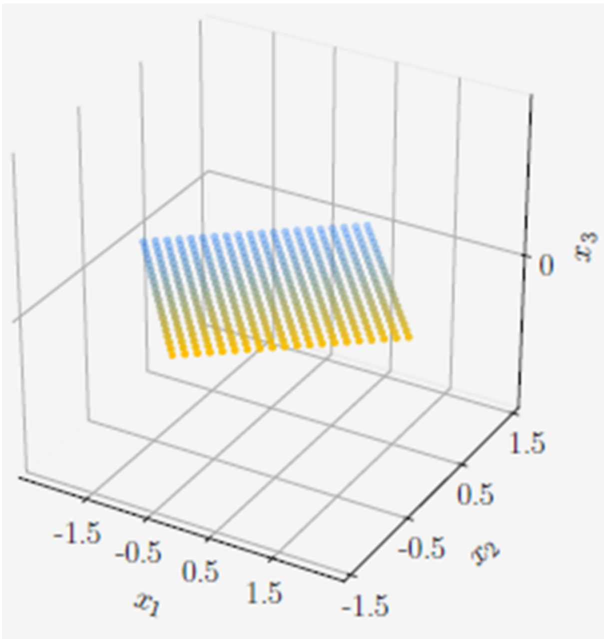
Início: conjunto de vetores X



Então aplica $V^\top \in \mathbf{R}^{2x2}$, que rotaciona X



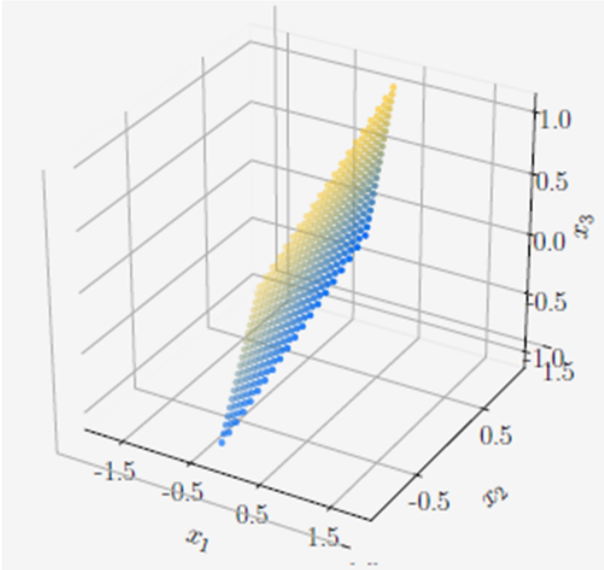
Então esses vetores são mapeados usando a matriz de valor singular Σ para o codomínio $\mathbf{R}^3$. Observe que neste momento todos os vetores continuam no plano $x_1 - x_2$ e a terceira coordenada sempre é 0. Os vetores no plano $x_1 - x_2$ foram esticados pelos valores singulares.



O mapeamento direto dos vetores X por A ao codomínio $\mathbf{R}^3$ corresponde à transformação de X por $U\Sigma V^\top$, onde U faz a rotação no codomínio $\mathbf{R}^3$, de forma que os vetores mapeados não fiquem mais restritos ao plano $x_1 - x_2$, mas se mantenham em um plano

4 Construção da SVD

A SVD de uma matriz geral guarda algumas similaridades com a autodecomposição de uma matriz quadrada.



Comparando a autodecomposição de uma matriz simétrica, definida positiva, definida por $S = S^\top = PDP^\top$, com a SVD correspondente $S = U\Sigma V^\top$, podemos definir que $U = P = V$ e $D = \Sigma$. Nesse caso, verificamos que a SVD de uma matriz simétrica, definida positiva corresponde à sua autodecomposição.

Calcular a SVD de $A \in \mathbf{R}^{m x n}$ é o equivalente a encontrar dois conjuntos de bases ortonormais $U = (u_1, \ldots, u_m)$ e $V = (v_1, \ldots, v_n)$ do codomínio $\mathbf{R}^m$ e do domínio $\mathbf{R}^n$ respectivamente. A partir dessas bases ordenadas, podem ser construídas as matrizes U e V.

O cálculo inicia-se com a construção o conjunto ortonormal de vetores singulares à direita $v_1, \ldots, v_n \in \mathbf{R}^n$. Depois, constrói-se o conjunto ortonormal de vetores singulares à esquerda $u_1, \ldots, u_m \in \mathbf{R}^m$. Então, liga-se os dois e exige-se que a ortogonalidade do v_i seja preservada sob a transformação de A. Isso é importante, pois se sabe que as imagens Av_i formam um conjunto de vetores ortogonais. Essas imagens são normalizadas por fatores escalares, que correspondem aos seus valores singulares.

O teorema espectral indica que os autovetores de uma matriz simétrica formam uma base ortonormal, o que significa que podem ser diagonalizados. Podemos construir uma matriz simétrica, semidefinida positiva $A^\top A \in \mathbf{R}^{n x n}$ a partir de qualquer matriz retangular $A \in \mathbf{R}^{m x n}$. Então, podemos diagonalizar $A^\top A$ e obter:

$$A^\top A = PDP^\top = P \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{bmatrix} P^\top \tag{6.4}$$

onde P é uma matriz ortogonal, composta por uma autobase ortonormal. $\lambda_i \geq 0$ são os autovalores de $A^\top A$.

Assumindo que a SVD de A existe:

$$A^\top A = (U\Sigma V^\top)^\top (U\Sigma V^\top) = V\Sigma^\top U^\top U\Sigma V^\top \tag{6.5}$$

onde U, V são matrizes ortogonais. Então, sendo $U^\top U = I$, obtém-se:

$$A^\top A = V\Sigma^\top \Sigma V^\top = V \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix} V^\top \tag{6.6}$$

Comparando, podemos identificar que

$$V^\top = P^\top \tag{6.7}$$

$$\sigma_i^2 = \lambda_i \tag{6.8}$$

Então, os autovetores de $A^\top A$ que compõe P são os vetores singulares à direita (V). Os autovalores da matriz $A^\top A$ são os valores singulares de Σ.

A obtenção dos vetores singulares à esquerda (U) pode acontecer usando-se um procedimento similar. Começa-se calculando a SVD da matriz simétrica $AA^\top \in \mathbf{R}^{m x m}$

$$AA^\top = (U\Sigma V^\top)(U\Sigma V^\top)^\top = U\Sigma V^\top V\Sigma^\top U^\top = U \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m^2 \end{bmatrix} U^\top \tag{6.9}$$

O teorema espectral diz que $AA^\top = SDS^\top$ pode ser diagonalizado e podemos encontrar uma base ortonormal de autovetores de $AA^\top$, coletados em S. Os autovetores ortonormais de $AA^\top$ são os vetores singulares à esquerda (U) e formam uma base ortonormal no codomínio da SVD.

Então, como $AA^\top$ e $A^\top A$ possuem os mesmos autovalores diferentes de zero, as entradas não zero das matrizes Σ na SVD devem ser iguais para ambos os casos.

Por fim, devemos unir todos os passos. Temos um conjunto ortonormal de vetores singulares à direita (V). Para terminar a construção da SVD, conectamos esses vetores com os vetores ortonormais (U). Para chegar nesse ponto, usamos o fato que imagens do v_i em A também devem ser ortogonais. É exigido que o produto interno entre Av_i e Av_j deve ser 0 para $i \neq j$. Para quaisquer dois autovetores ortogonais v_i, v_j, $i \neq j$, temos que:

$$(Av_i)^\top (Av_j) = v_i^\top (A^\top A)v_j = v_i^\top (\lambda_j v_j) = \lambda_j v_i^\top v_j = 0 \tag{6.10}$$

Para o caso m $\geq$ r, $\{Av_1, \ldots, Av_r\}$ é a base de um subespaço de dimensão r de $\mathbf{R}^m$.

Para completar a construção da SVD, são necessários vetores singulares à esquerda que sejam ortonormais. Normalizam-se as imagens dos vetores singulares à direita Av_i, obtendo

$$u_i := \frac{Avi_i}{\left\|Av_i\right\|} = \frac{1}{\sqrt{\lambda_i}} Av_i = \frac{1}{\sigma_i} Av_i \tag{6.11}$$

Onde a última parte é obtida da equação que demonstra que os autovalores de $AA^\top$ são tais que $\sigma_i^2 = \lambda_i$.

Então, os autovetores de $A^\top A$, que sabemos ser os vetores singulares à direita v_i, e suas imagens normalizadas sob A, os vetores singulares à esquerda u_i formam duas bases ortonormais auto consistentes conectados através da matriz de valor singular Σ.

Dessa forma a equação pode ser rearranjada para obter a equação de valor singular $Av_i = \sigma_i u_i, i = 1, \ldots, r$.

Essa equação lembra a equação de autovalores, mas os vetores nos lados esquerdos e direitos não são os mesmos.

Para n < m, a equação se sustenta apenas para i $\leq$ n, mas não diz nada sobre u_i para i > n. Porém, sabemos que eles são ortonormais. Por outro lado, para m < n a equação se sustenta apenas para i $\leq$ m. Para i > m, temos que $Av_i = 0$, e sabemos também que v_i formam um conjunto ortonormal. Isso significa que a SVD também fornece uma base ortonormal do kernel (espaço nulo) de A, o conjunto de vetores x com $Ax = 0$.

Concatenando v_i como as colunas de V e u_i como as colunas de U, temos

$$AV = U\Sigma \tag{6.12}$$

onde Σ tem as mesmas dimensões que A, e uma estrutura diagonal de linhas 1, ..., r. Desta forma, temos que $A = U\Sigma V^\top$, que é a SVD de A.

5 Exemplo – Cálculo da SVD de uma matriz

$$A = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} \tag{6.13}$$

a Calcular os vetores singulares à direita a partir da base de autovetores de $A^\top A$

Primeiro serão calculados os vetores singulares à direita de $A^\top A$. Inicia-se com o cálculo de $A^\top A$.

$$A^\top A = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \tag{6.14}$$

Através da decomposição em autovalores de $A^\top A$, obtemos a matriz P formada pelos autovetores de A e matriz diagonal D com os autovalores de A.

$$A^\top A = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} = PDP^\top \tag{6.15}$$

Os vetores singulares à direita são obtidos pelas colunas de P, portanto

$$[V = P = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix} \tag{6.16}$$

b Calcular a matriz de valores singulares

Os valores singulares σ_i são as raízes quadradas dos autovalores de $A^\top A$, obtidos da matriz D calculada no passo anterior. Como rk(A) = 2, existem apenas dois valores singulares diferentes de zero: $\sigma_1 = \sqrt{6}$ e $\sigma_2 = 1$. A matriz de valor singular deve ser do mesmo tamanho que A, então, obtemos

$$\Sigma = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{6.17}$$

c Calcular os vetores singulares à esquerda como imagem normalizada dos vetores singulares à direita

Os vetores singulares à esquerda são obtidos pelo cálculo da imagem dos vetores singulares à direita em A, normalizados pela divisão por seu valor singular correspondente.

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} \\ \frac{-2}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix} \tag{6.18}$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{1} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{bmatrix} \tag{6.19}$$

$$U = \begin{bmatrix} u_1, u_2 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \tag{6.20}$$

6 Comparação de decomposição de autovalores com decomposição de valores singulares

Sendo a autodecomposição de $A = PDP^{-1}$ e a SVD de $A = U\Sigma V^\top$

- A SVD sempre existe para qualquer matriz $\mathbf{R}^{m\,x\,n}$. A autodecomposição é definida apenas para matrizes quadradas $\mathbf{R}^{n\,x\,n}$ e existem apenas se for possível encontrar uma base de autovetores de $\mathbf{R}^n$

- Os vetores na matriz de autodecomposição P não são necessariamente ortogonais (por exemplo, a mudança de base não é uma simples rotação ou escala). Por outro lado, os vetores nas matrizes U e V na SVD são ortonormais, como representado pelas rotações.

- Tanto a autodecomposição como a SVD são composições de três transformações lineares:

  1. Mudanças de bases no domínio
  2. Escala independente de cada novo vetor base e transformação do domínio para codomínio
  3. Mudança de base no codomínio

  Uma diferença chave entre a autodecomposição e a SVD é que na SVD, domínio e codomínio podem ser espaços vetoriais de dimensões diferentes.

- Na SVD as matrizes de vetor singular esquerda e direita U e V são geralmente não inversas entre si (elas fazem a mudança de base em espaços vetoriais diferentes). Na autodecomposição, as matrizes de mudança de base P e P^{-1} são inversas entre si.

- Na SVD, as entradas da matriz diagonal Σ são todas reais e não negativas, o que não é necessariamente verdade para a matriz diagonal da autodecomposição.

- A SVD e a autodecomposição são fortemente relacionadas pelas suas projeções:

  – Os vetores singulares à esquerda de A são autovetores de $AA^\top$
  – Os vetores singulares à direita de A são autovetores de $A^\top A$
  – Os valores singulares não-zero de A são as raízes quadradas dos autovalores não zero tanto de $AA^\top$ como de $A^\top A$

- Para matrizes simétricas $A \in \mathbf{R}^{n\,x\,n}$ a decomposição de autovalores e a SVD são únicos, conforme o teorema espectral.

7 Conclusões

A Decomposição em Valores Singulares é usada em vários tipos de aplicações de *machine learning*, desde problemas de mínimos quadrados em ajuste de curvas até a resolução de sistemas de equações lineares. Essas aplicações aproveitam várias propriedades importantes da SVD, sua relação com o posto de uma matriz e sua capacidade de aproximar matrizes de certo posto com matrizes de posto menor.

A substituição de uma matriz por sua SVD permite que os cálculos sejam mais resistentes a erros numéricos de arredondamento. Ela é utilizada em aplicações desde a redução de dimensionalidade e modelagem de tópicos até a compressão e clusterização de dados.

7 Aproximação de matrizes

Aproximação de matrizes é uma ferramenta matemática essencial para simplificação e análise de dados complexos. As matrizes de aproximação oferecem uma representação compacta de sistemas lineares e não lineares, permitindo a captura de informações essenciais enquanto reduzem a complexidade computacional. Um exemplo de aplicação, que iremos discutir no decorrer desta seção, é o uso de matrizes de aproximação em sistemas de recomendação e na compressão de imagens.

A Decomposição de Valores Singulares (SVD), que apresentamos na Seção 6, tem papel fundamental para a aplicação de aproximação de matrizes de forma otimizada. Por definição, a decomposição em valores singulares de uma matriz $A_{m\times n}$ é da forma

$$A_{m\times n} = U\Sigma V^{\top}, \tag{7.1}$$

onde, Σ é a matriz diagonal dos valores singulares, e U e V são as matrizes de valores singulares à esquerda e à direita, respectivamente.

A relação entre a quantidade de valores singulares em Σ e a quantidade de vetores singulares em U e V, decorrentes da SVD de uma matriz $A_{m\times n}$, é diretamente associado ao posto da matriz original, i.e. $A_{m\times n}$. O posto de uma matriz $A_{m\times n}$, denotado por $posto(A)$, é definido como o máximo de vetores, linhas ou colunas, linearmente independentes em A.

Seja $A_{m\times n}$ uma matriz qualquer, com r vetores, do tipo coluna, linearmente independentes. Desta forma, o $posto(A_{m\times n}) = r$. Pela definição de SVD, a matriz Σ tem r valores singulares distribuídos em sua diagonal principal, e dimensão $r \times r$. Por outro lado, as matrizes U e V possuem r vetores singulares e dimensão $m \times r$ e $n \times r$, respectivamente. Desta forma, se $r = m = n$, ambas matrizes U e V serão matrizes quadradas. Nesse caso, dizemos que o SVD completo é aplicado. Entretanto, se $r < m$ e $r < n$, então as matrizes U e V não são quadradas, com número diferente de linhas e colunas, neste caso dizemos que o SVD reduzido é aplicado.

1 Aplicações de aproximação de matrizes

Enquanto matrizes quadradas são fáceis de trabalhar, por nos permitirem facilmente identificá-las. Em problemas práticos reais, a representação de dados em matrizes quadradas não são tão comuns, apresentando uma ordem de complexidade maior. Um exemplo de matrizes que encontramos na prática são matrizes altas e magras, que contém maior quantidade de linhas do que colunas. Tais matrizes são bem comuns em bases de dados, com diversas linhas (entradas) e uma quantidade menor de parâmetros (variáveis).

Usuários	Matrix	Alien	Serenity	Casablanca	Amelie
Ana	1	1	1	0	0
Bruno	3	3	3	0	0
Camila	4	4	4	0	0
Danilo	5	5	5	0	0
Emilia	0	0	0	4	4
Fabio	0	0	0	5	5
Gabriela	0	0	0	2	2

Tabela 1 – Exemplo de base de dados com avaliações de filmes.

A Tabela 1 representa um exemplo prático de base de dados onde é possível utilizar matrizes do tipo altas e magras para representação. A base de dados apresenta classificações, de 1 a 5 estrelas, de diferentes usuários de um serviço de *streaming*. Assim, a cada posição da tabela, é apresentada a nota atribuída por um usuário, representado pela linha, a um filme, representado por uma coluna. Usuários que não assistiram um determinado filme, apresentam valor 0.

Podemos representar a base de dados de avaliações de filmes descrita com uma matriz $A_{m \times n}$, de $posto(A) = 2$, que pode ser ilustrada como:

$$A_{7\times5} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix}, \tag{7.2}$$

onde linhas representam os usuários e colunas representam os filmes. Assim, ao calcular o SVD para esta matriz obtemos o seguinte resultado:

$$\begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 \\ \mathbf{4} & \mathbf{4} & \mathbf{4} & 0 & 0 \\ \mathbf{5} & \mathbf{5} & \mathbf{5} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{4} & \mathbf{4} \\ 0 & 0 & 0 & \mathbf{5} & \mathbf{5} \\ 0 & 0 & 0 & \mathbf{2} & \mathbf{2} \end{bmatrix} = \begin{bmatrix} \mathbf{0.14} & 0 \\ \mathbf{0.42} & 0 \\ \mathbf{0.56} & 0 \\ \mathbf{0.70} & 0 \\ 0 & \mathbf{0.60} \\ 0 & \mathbf{0.75} \\ 0 & \mathbf{0.30} \end{bmatrix} \times \begin{bmatrix} \mathbf{12.4} & 0 \\ 0 & \mathbf{9.5} \end{bmatrix} \times \begin{bmatrix} \mathbf{0.58} & \mathbf{0.58} & \mathbf{0.58} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{0.71} & \mathbf{0.71} \end{bmatrix}, \tag{7.3}$$

representando a decomposição SVD do tipo $A = U\Sigma V^\top$.

Note que na base de dados, é possível identificar dois blocos que relacionam usuário e filmes. O primeiro bloco, mais à esquerda, representa usuários que assistem filmes de ficção científica (*Matrix*, *Alien* e *Serenity*), enquanto o bloco mais à esquerda representam usuários que assistem filmes de romance (*Casablanca* e *Amelie*). Com isso, observamos que a matriz A original possui $posto(A) = 2$, que é exatamente o número de conceitos de filmes identificados.

Dessa forma, a matriz U relaciona em cada linha os usuários com cada um dos conceitos, onde os três primeiros usuários, nas três primeiras linhas, apresentam valores diferentes de zero em U, justamente por assistirem filmes de ficção científica. Por outro lado, a matriz V relaciona filmes a conceitos, ou seja, os filmes *Matrix*, *Alien* e *Serenity*, estão relacionados ao primeiro conceito (ficção científica), enquanto *Casablanca* e *Amelie*, estão relacionados ao segundo conceito (romance). Desta forma, a solução do SVD relaciona as linhas e colunas da matriz de entrada A com os conceitos identificados nos dados, que são da mesma ordem que o $posto(A)$.

Entretanto, a base de dados descrita contendo avaliações de filmes feitas por usuários não é um exemplo prático tão interessante, pois os usuários na base de dados assistem filmes de apenas um conceito, ou seja, não existe um usuário que assiste filmes de ficção científica e romance. Dessa forma, modificamos a matriz anterior inserindo dois usuários que assistem filmes dos dois conceitos. O SVD da matriz modificada é apresentado a seguir:

$$\begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 \\ \mathbf{4} & \mathbf{4} & \mathbf{4} & 0 & 0 \\ \mathbf{5} & \mathbf{5} & \mathbf{5} & 0 & 0 \\ 0 & \color{red}{\mathbf{2}} & 0 & \mathbf{4} & \mathbf{4} \\ 0 & 0 & 0 & \mathbf{5} & \mathbf{5} \\ 0 & \color{red}{\mathbf{1}} & 0 & \mathbf{2} & \mathbf{2} \end{bmatrix} = \begin{bmatrix} \mathbf{0.14} & -0.02 & -0.01 \\ \mathbf{0.42} & -0.07 & -0.03 \\ \mathbf{0.56} & -0.09 & -0.04 \\ \mathbf{0.70} & -0.11 & -0.05 \\ 0.15 & \mathbf{0.59} & \mathbf{0.65} \\ 0.07 & \mathbf{0.73} & \mathbf{-0.67} \\ 0.07 & \mathbf{0.29} & \mathbf{0.32} \end{bmatrix} \times \begin{bmatrix} \mathbf{12.4} & 0 & 0 \\ 0 & \mathbf{9.5} & 0 \\ 0 & 0 & \mathbf{1.3} \end{bmatrix} \times \begin{bmatrix} \mathbf{0.56} & \mathbf{0.59} & \mathbf{0.56} & 0.09 & 0.09 \\ -0.12 & 0.02 & -0.12 & \mathbf{0.69} & \mathbf{0.69} \\ 0.40 & \mathbf{-0.80} & 0.40 & 0.09 & 0.09 \end{bmatrix}, \tag{7.4}$$

onde a matriz A tem $posto(A) = 3$, uma vez que o vetor coluna da segunda coluna não pode ser escrito como combinação linear dos demais vetores.

Desta forma, para aplicar o conceito de matriz de aproximação, basta desconsiderar um ou mais valores/vetores singulares de A. Assim, ao resolver o sistema, obtemos uma matriz aproximada de A, com valores diferentes em suas posições, entretanto, que mantém o comportamento dos dados de A original. A seguir apresentamos como utilizar aproximação de matrizes para recomendar filmes a um usuário.

a Sistemas de recomendação

Considere a matriz A a seguir e sua decomposição SVD:

$$\begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & 0 & 0 \\ \color{red}{\mathbf{0}} & \mathbf{3} & \mathbf{3} & 0 & 0 \\ \mathbf{4} & \mathbf{4} & \color{red}{\mathbf{0}} & 0 & 0 \\ \mathbf{5} & \mathbf{5} & \mathbf{5} & 0 & 0 \\ 0 & \mathbf{2} & 0 & \mathbf{4} & \mathbf{4} \\ 0 & 0 & 0 & \mathbf{5} & \mathbf{5} \\ 0 & \mathbf{1} & 0 & \mathbf{2} & \mathbf{2} \end{bmatrix} = \begin{bmatrix} \mathbf{0.14} & -0.02 & -0.01 \\ \mathbf{0.42} & -0.07 & -0.03 \\ \mathbf{0.56} & -0.09 & -0.04 \\ \mathbf{0.70} & -0.11 & -0.05 \\ 0.15 & \mathbf{0.59} & \mathbf{0.65} \\ 0.07 & \mathbf{0.73} & \mathbf{-0.67} \\ 0.07 & \mathbf{0.29} & \mathbf{0.32} \end{bmatrix} \times \begin{bmatrix} \mathbf{12.4} & 0 & 0 \\ 0 & \mathbf{9.5} & 0 \\ 0 & 0 & \mathbf{1.3} \end{bmatrix} \times \begin{bmatrix} \mathbf{0.56} & \mathbf{0.59} & \mathbf{0.56} & 0.09 & 0.09 \\ -0.12 & 0.02 & -0.12 & \mathbf{0.69} & \mathbf{0.69} \\ 0.40 & \mathbf{-0.80} & 0.40 & 0.09 & 0.09 \end{bmatrix}, \tag{7.5}$$

observe que os usuários da segunda e terceira linha assistiram apenas dois filmes. Com isso, uma forma de calcular uma previsão de nota que ambos usuários poderiam atribuir aos filmes destacados, aos quais os usuários não assistiram, é utilizando a aproximação de matrizes.

Para calcular uma matriz de aproximação, podemos utilizar um subconjunto dos valores singulares e vetores singulares descritos, a fim de obter uma matriz próxima à matriz A original. Desta forma, vamos considerar apenas os dois primeiros valores singulares maiores, ou seja, 12.4 e 9.5, bem como os respectivos vetores singulares em U e V. Abaixo apresentamos a matriz de aproximação resultante dessa operação:

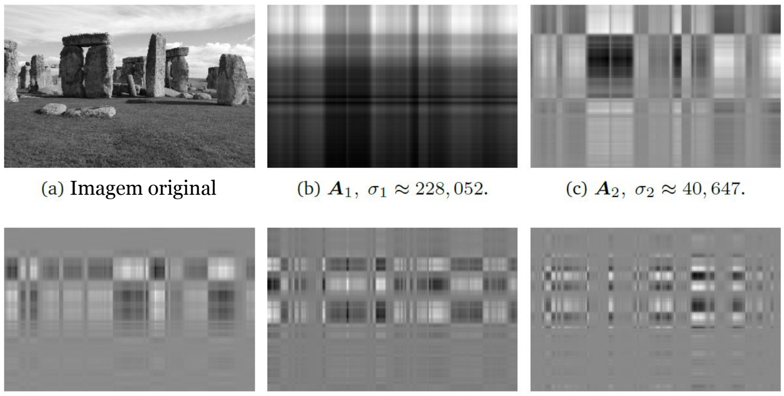
$$\begin{bmatrix} \mathbf{0.14} & -0.02 \\ \mathbf{0.42} & -0.07 \\ \mathbf{0.56} & -0.09 \\ \mathbf{0.70} & -0.11 \\ 0.15 & \mathbf{0.59} \\ 0.07 & \mathbf{0.73} \\ 0.07 & \mathbf{0.29} \end{bmatrix} \times \begin{bmatrix} \mathbf{12.4} & 0 \\ 0 & \mathbf{9.5} \end{bmatrix} \times \begin{bmatrix} \mathbf{0.56} & \mathbf{0.59} & \mathbf{0.56} & 0.09 & 0.09 \\ -0.12 & 0.02 & -0.12 & \mathbf{0.69} & \mathbf{0.69} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{1.1} & \mathbf{0.8} & -0.1 & -0.1 \\ \mathbf{1.9} & \mathbf{2.3} & \mathbf{1.6} & 0.1 & 0.1 \\ \mathbf{2.7} & \mathbf{3.3} & \mathbf{2.3} & 0.1 & 0.1 \\ \mathbf{4.8} & \mathbf{5.7} & \mathbf{4.1} & -0.1 & 0.1 \\ 0.4 & \mathbf{1.4} & 0.3 & \mathbf{4.1} & \mathbf{4.1} \\ -0.4 & 0.6 & -0.4 & \mathbf{4.9} & \mathbf{4.9} \\ 0.2 & \mathbf{0.7} & 0.2 & \mathbf{2} & \mathbf{2} \end{bmatrix}. \tag{7.6}$$

Note que a matriz obtida ao final da operação, é uma aproximação da matriz de entrada na Equação (7.5). Assim, a matriz aproximada resultante, preenche os valores anteriormente zerados na matriz original com os valores 1.9 e 2.3, o que pode representar uma previsão da nota que os respectivos usuários atribuiriam aos filmes não assistidos. Portanto, essa informação aproximada pode ser utilizada por algoritmos de recomendação para sugerir aos usuários os filmes baseados nos valores da matriz de aproximação.

b Compressão de imagens

Outro exemplo de aplicação de aproximação de matrizes é na compressão de imagens. Uma vez que as imagens podem ser representadas em uma estrutura de dados do tipo grid, podemos então utilizar a representação de matrizes para representar imagens. Desta forma, a matriz original da decomposição SVD representa uma imagem original, que desejamos comprimir. Ao calcular-se a decomposição SVD, obtém-se a lista de valores singulares, onde cada valor singular representa uma representação da imagem original. Figura 3 apresenta um exemplo de representação de imagens obtida por cada valor singular.

Desta forma, para obter a compressão da imagem original, basta selecionar, em sequência crescente, os valores singulares e seus respectivos vetores singulares. Desta forma, a cada novo valor singular considerado na compressão, maior é a imagem aproximada, entretanto melhor é a qualidade da imagem obtida na aproximação. Figura 4 apresenta um exemplo de compressão de imagem a partir da combinação de valores singulares, observe que a qualidade da imagem vai melhorando a cada novo valor singular considerado.



(a) Imagem original (b) $\boldsymbol{A}_1,\ \sigma_1 \approx 228,052.$ (c) $\boldsymbol{A}_2,\ \sigma_2 \approx 40,647.$

(d) $\boldsymbol{A}_3,\ \sigma_3 \approx 26,125.$ (e) $\boldsymbol{A}_4,\ \sigma_4 \approx 20,232.$ (f) $\boldsymbol{A}_5,\ \sigma_5 \approx 15,436.$

Figura 3 – Exemplo de representação de imagem a partir de valores singulares (DEISENROTH; FAISAL; ONG, 2020).

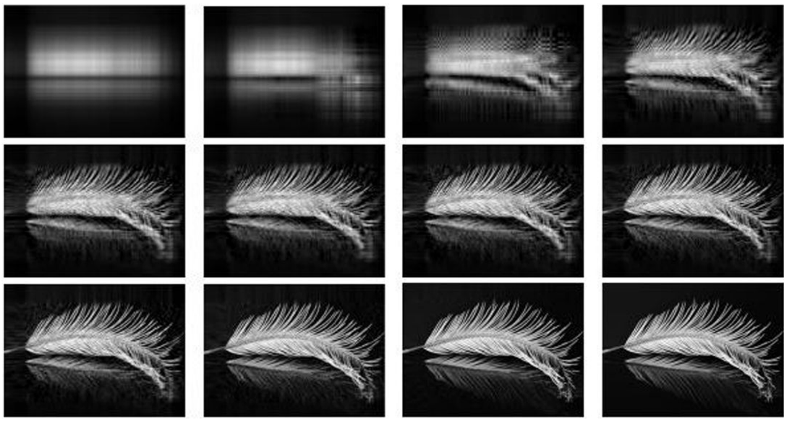


Figura 4 – Compressão de imagem para valores singulares sequencialmente selecionados (COMPTON; ERNSTBERGER, 2020).

Por fim, para verificar se uma compressão é boa o suficiente podemos utilizar o método *Root Mean Square Error (RMSE)* (HODSON, 2022), uma métrica para avaliar a precisão de modelos de conversão. Ele calcula a raiz quadrada da média dos quadrados das diferenças entre os valores da imagem original com os valores da imagem comprimida, i.e. matriz original e matriz de aproximação. Além disso, o Teorema de Eckart-Young (CHIPMAN, 2020), mostra que a matriz de aproximação A_k, de posto k, obtida através dos k maiores valores singulares $\{\sigma_1, \sigma_2, ..., \sigma_k\}$, é a melhor matriz de aproximação, de posto k, para a matriz original.

8 Conclusão

Neste trabalho, nós apresentamos conceitos sobre o conteúdo do Capítulo 4 do livro texto (DEISENROTH; FAISAL; ONG, 2020). Apresentamos os conceitos sobre determinante e traço de matrizes, autovalores, autovetores, decomposição de Cholesky e decomposição de autovalores, autovetores e diagonalização de matrizes. Por fim, apresentamos conceitos sobre Decomposição de Valores Singulares (SVD) e aproximação de matrizes e suas aplicações.

Referências

CHIPMAN, J. S. “Proofs” and Proofs of the Eckart–Young Theorem. In: *Stochastic processes and functional analysis*. [S.l.]: CRC Press, 2020. p. 71–83. Citado na página 33.

COMPTON, E. A.; ERNSTBERGER, S. L. Singular Value Decomposition: Applications to Image Processing. *Citations Journal of Undergraduate Research*, v. 17, 2020. Citado na página 33.

DEISENROTH, M. P.; FAISAL, A. A.; ONG, C. S. *Mathematics for machine learning*. [S.l.]: Cambridge University Press, 2020. Citado 3 vezes nas páginas 3, 33 e 34.

HODSON, T. O. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, Göttingen, Germany, v. 2022, p. 1–10, 2022. Citado na página 33.