

Exemplo de Aplicação de Transformações Lineares: Análise das Componentes Principais

Guilherme de Alencar Barreto

`gbarreto@ufc.br`

Departamento de Engenharia de Teleinformática (DETI)
Engenharias de Computação, Telecomunicações e Teleinformática
Universidade Federal do Ceará – UFC

www.researchgate.net/profile/Guilherme_Barreto2/

Conteúdo dos Slides

- 1 Transformadas Matriciais
- 2 Descrição do Problema
- 3 Algoritmo PCA: Passo-a-Passo
- 4 Diagonalização da Matriz de Covariância
- 5 Interpretação Geométrica
- 6 Redução de Dimensionalidade
- 7 Exemplos Teórico-Computacionais

Transformadas Matriciais

Para cada $\mathbf{x} \in \mathbb{R}^p$, uma transformada matricial é definida por

$$\mathbf{y} = \mathbf{W}\mathbf{x} \quad (\text{ou} \quad \mathbf{W}\mathbf{x} = \mathbf{y}), \quad (1)$$

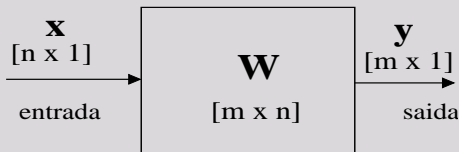
em que \mathbf{W} é uma matriz $q \times p$.

- Para simplificar, muitas vezes denotamos essa transformação matricial por

$$\boxed{\mathbf{x} \mapsto \mathbf{W}\mathbf{x}} \quad (2)$$

Diagrama de Blocos

Ajuda muito no entendimento de uma transformação linear se representarmos a relações $y = \mathbf{W}x$ na forma de um diagrama de blocos do tipo entrada-saída.



Formalização Matemática do Problema

- Considere um conjunto de dados formado por N vetores de atributos \mathbf{x}_k , $k = 1, \dots, N$, que estão organizados ao longo das colunas da matriz \mathbf{X} :

$$\mathbf{X} = [\mathbf{x}_1 \mid \mathbf{x}_2 \mid \cdots \mid \mathbf{x}_N], \quad \dim(\mathbf{X}) = p \times N \quad (3)$$

- Cada vetor de atributo \mathbf{x}_k tem dimensão $p \times 1$, ou seja

$$\mathbf{x}_k = \begin{bmatrix} x_{k,1} \\ x_{k,2} \\ \vdots \\ x_{k,p} \end{bmatrix} \quad (4)$$

- Assume-se que p é muito grande (i.e. $p \rightarrow \infty$).

Formalização Matemática do Problema

- Deseja-se transformar cada vetor \mathbf{x}_k no conjunto de dados em um outro vetor \mathbf{y}_k de dimensão q , ou seja

$$\mathbf{y}_k = \begin{bmatrix} y_{k,1} \\ y_{k,2} \\ \vdots \\ y_{k,q} \end{bmatrix}, \quad \text{tal que } q \leq p. \quad (5)$$

- Isto deve ser feito por meio de uma transformação linear:

$$\mathbf{y}_k = \mathbf{Q}\mathbf{x}_k, \quad \forall k = 1, \dots, N. \quad (6)$$

- Além disso, esta transformação deve preservar a informação **relevante** constante no conjunto \mathbf{X} .

Algoritmo PCA: Passo-a-Passo

- **Passo 1** - Determinar o vetor-médio dos dados em \mathbf{X} :

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{k=1}^N \mathbf{x}_k. \quad (7)$$

- **Passo 2** - Centralizar os dados: $\mathbf{x}_k = \mathbf{x}_k - \bar{\mathbf{x}}$.
- **Passo 3** - Estimar a matriz de covariância dos dados em \mathbf{X} :

$$\mathbf{C}_{\mathbf{x}} = E[\mathbf{x}\mathbf{x}^T], \quad (8)$$

$$\approx \frac{1}{N} \sum_{k=1}^N \mathbf{x}_k \mathbf{x}_k^T, \quad (9)$$

em que $E[\cdot]$ é o operador valor esperado.

Algoritmo PCA: Passo-a-Passo

- **Passo 4** - Determinar os p autovalores da matriz \mathbf{C}_x e os p autovetores correspondentes. Em outras palavras, resolver o seguinte sistema de equações:

$$\mathbf{C}_x \mathbf{v} = \lambda \mathbf{v}, \quad (10)$$

em que λ e \mathbf{v} denotam, respectivamente, o autovalor e o autovetor associado.

- Os autovalores são as raízes do polinômio em λ , de ordem p , obtido a partir da seguinte expressão:

$$\det(\mathbf{C}_x - \lambda \mathbf{I}_p) = 0. \quad (11)$$

Algoritmo PCA: Passo-a-Passo

- Como a matriz \mathbf{C}_x é positiva definida, todos os seus p autovalores são positivos e distintos.
- Os autovetores devem ser ordenados em modo decrescente:

$$\lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_p \quad (12)$$

- Os autovetores são determinados resolvendo-se o sistema na Eq. (10) p vezes, uma para cada autovalor:

$$\mathbf{C}_x \mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i = 1, \dots, p. \quad (13)$$

em que λ_i e \mathbf{v}_i são, respectivamente, o i -ésimo autovalor e o autovetor associado.

Algoritmo PCA: Passo-a-Passo

- Os autovetores \mathbf{v}_i , $i = 1, \dots, p$, formam um conjunto de vetores **ortonormais**:

$$\mathbf{v}_i^T \mathbf{v}_j = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } i \neq j \end{cases} \quad (14)$$

- Dispor os autovetores ao longo das colunas da matriz \mathbf{V} :

$$\mathbf{V} = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_p]. \quad (15)$$

- A matriz \mathbf{V} é quadrada e de dimensões $p \times p$.
- Passo 5** - Definir a matriz de transformação \mathbf{Q} como segue.

$$\mathbf{Q} = \mathbf{V}^T. \quad (16)$$

Algoritmo PCA: Passo-a-Passo

- **Passo 6** - Aplicar a matriz \mathbf{Q} sobre os vetores de atributos originais.
- Isto pode ser feito vetor a vetor:

$$\mathbf{y}_k = \mathbf{V}^T \mathbf{x}_k, \quad (17)$$

para $k = 1, \dots, N$.

- Ou de uma vez só:

$$\mathbf{Y} = \mathbf{V}^T \mathbf{X}. \quad (18)$$

Diagonalização da Matriz C_x

- Ao Aplicar a matriz Q sobre os vetores de atributos originais, da forma como está definida na Eq. (17), percebemos que as dimensões dos vetores x_k e y_k são iguais (i.e. $p = q$).
- Portanto, não temos aqui uma redução de dimensionalidade.
- O que, então, acontece com os vetores de dados?
- Que tipo de transformação eles sofreram?
- Para responder estas questões, vamos precisar calcular a matriz de covariância dos dados transformados, ou seja, precisamos determinar

$$C_y = E[yy^T]. \quad (19)$$

Diagonalização da Matriz \mathbf{C}_x

- A partir da Eq. (19) e da Eq. (18), obtemos:

$$\mathbf{C}_y = E[\mathbf{y}\mathbf{y}^T] \quad (20)$$

$$= E[(\mathbf{V}^T \mathbf{x}) (\mathbf{V}^T \mathbf{x})^T] \quad (21)$$

$$= E[(\mathbf{V}^T \mathbf{x}) (\mathbf{x}^T \mathbf{V})] \quad (22)$$

$$= E[\mathbf{V}^T (\mathbf{x}\mathbf{x}^T) \mathbf{V}] \quad (23)$$

$$= \mathbf{V}^T E[\mathbf{x}\mathbf{x}^T] \mathbf{V} \quad (24)$$

$$= \mathbf{V}^T \mathbf{C}_x \mathbf{V} \quad (25)$$

- Este resultado mostra que dada a matriz de covariância dos dados originais \mathbf{C}_x e a matriz de autovetores \mathbf{V} , facilmente obtemos a matriz de covariância dos dados transformados \mathbf{C}_y .

Diagonalização da Matriz C_x

- Contudo, não diz muita coisa sobre a forma da matriz de covariância dos dados transformados.
- Para isso, vamos expandir os produtos matriz-vetor da Eq. (25).
- Primeiro, vamos expandir o produto $C_x V$:

$$C_x V = [C_x v_1 \mid C_x v_2 \mid \cdots \mid C_x v_p] \quad (26)$$

- Este produto resulta em uma matriz $p \times p$.
- Usando a Eq. (13), chegamos ao seguinte resultado:

$$C_x V = [\lambda_1 v_1 \mid \lambda_2 v_2 \mid \cdots \mid \lambda_p v_p] \quad (27)$$

Diagonalização da Matriz \mathbf{C}_x

- Agora, lembrando que \mathbf{V}^T também é uma matriz $p \times p$, podemos realizar a segunda parte do produto:

$$\mathbf{V}^T \mathbf{C}_x \mathbf{V} = \begin{bmatrix} \mathbf{v}_1^T \\ \dots \\ \mathbf{v}_2^T \\ \dots \\ \vdots \\ \dots \\ \mathbf{v}_p^T \end{bmatrix} [\lambda_1 \mathbf{v}_1 \mid \lambda_2 \mathbf{v}_2 \mid \dots \mid \lambda_p \mathbf{v}_p] \quad (28)$$

Diagonalização da Matriz C_x

- Agora, lembrando que V^T também é uma matriz $p \times p$, podemos realizar a segunda parte do produto:

$$C_y = V^T C_x V \quad (29)$$

$$= \begin{bmatrix} \lambda_1 \mathbf{v}_1^T \mathbf{v}_1 & \lambda_2 \mathbf{v}_1^T \mathbf{v}_2 & \cdots & \lambda_p \mathbf{v}_1^T \mathbf{v}_p \\ \lambda_1 \mathbf{v}_2^T \mathbf{v}_1 & \lambda_2 \mathbf{v}_2^T \mathbf{v}_2 & \cdots & \lambda_p \mathbf{v}_2^T \mathbf{v}_p \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 \mathbf{v}_p^T \mathbf{v}_1 & \lambda_2 \mathbf{v}_p^T \mathbf{v}_2 & \cdots & \lambda_p \mathbf{v}_p^T \mathbf{v}_p \end{bmatrix} \quad (30)$$

$$= \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 \cdots & 0 & \lambda_p \end{bmatrix} \quad (31)$$

Diagonalização da Matriz C_x

- O produto $V^T C_x V$ pode também ser desenvolvido em função dos elementos de C_x .

$$C_y = V^T C_x V \quad (32)$$

$$= \begin{bmatrix} \sigma_1^2 \mathbf{v}_1^T \mathbf{v}_1 & \sigma_{12} \mathbf{v}_1^T \mathbf{v}_2 & \cdots & \sigma_{1n} \mathbf{v}_1^T \mathbf{v}_p \\ \sigma_{21} \mathbf{v}_2^T \mathbf{v}_1 & \sigma_2^2 \mathbf{v}_2^T \mathbf{v}_2 & \cdots & \sigma_{2n} \mathbf{v}_2^T \mathbf{v}_p \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} \mathbf{v}_p^T \mathbf{v}_1 & \sigma_{n2} \mathbf{v}_p^T \mathbf{v}_2 & \cdots & \sigma_p^2 \mathbf{v}_p^T \mathbf{v}_p \end{bmatrix} \quad (33)$$

$$= \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 \cdots & 0 & \sigma_p^2 \end{bmatrix} \quad (34)$$

Diagonalização da Matriz C_x

- Resumindo, temos que a matriz de covariância dos dados transformados $C_y = V^T C_x V$ tem a seguinte forma:

$$C_y = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 \cdots & 0 & \lambda_p \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 \cdots & 0 & \sigma_p^2 \end{bmatrix} \quad (35)$$

Diagonalização da Matriz C_x

- Do exposto na Eq. (35), tiramos as seguintes conclusões:
 - 1 A matriz de covariância dos dados transformados é diagonal, ou seja, não há correlação entre as componentes do vetor y .
 - 2 Ou seja, PCA atua sobre os dados em X para gerar um novo conjunto de dados Y , cuja matriz de covariância é diagonal.
 - 3 As variâncias das variáveis transformadas são iguais aos autovalores de C_x .
 - 4 Os autovalores, por sua vez, são iguais às variâncias das variáveis originais.
- Uma consequência imediata desses resultados é que não é necessário calcular os autovalores da matriz x , já que eles são iguais às variâncias das variáveis originais!!

Interpretação Geométrica

- A transformação linear implementada por PCA pode ser vista como uma mudança de base.
- Uma base no espaço \mathbb{R}^n é qualquer conjunto de n vetores linearmente independentes (LI), a partir dos quais os vetores daquele espaço são escritos ou representados.
- Por exemplo, a base canônica do \mathbb{R}^2 é formada pelos vetores

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (36)$$

- Assim, qualquer vetor $\begin{bmatrix} a & b \end{bmatrix}^T$ no \mathbb{R}^2 pode ser escrito como

$$\begin{bmatrix} a \\ b \end{bmatrix} = a\mathbf{e}_1 + b\mathbf{e}_2 = a \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (37)$$

Interpretação Geométrica

- Uma outra possível base do \mathbb{R}^2 é formada pelos vetores

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{e} \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \quad (38)$$

- Neste caso, o vetor $[a \ b]^T$, originalmente escrito com relação à base $\mathcal{B}_1 = \{\mathbf{e}_1, \mathbf{e}_2\}$, com relação à base $\mathcal{B}_2 = \{\mathbf{v}_1, \mathbf{v}_2\}$ passa ser representado como

$$\begin{bmatrix} a \\ b \end{bmatrix} = a\mathbf{v}_1 + (b - a)\mathbf{v}_2 = a \begin{bmatrix} 0 \\ 1 \end{bmatrix} + (b - a) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (39)$$

$$\begin{bmatrix} a \\ b \end{bmatrix}_{\mathcal{B}_1} \equiv \begin{bmatrix} a \\ (b - a) \end{bmatrix}_{\mathcal{B}_2}. \quad (40)$$

Interpretação Geométrica

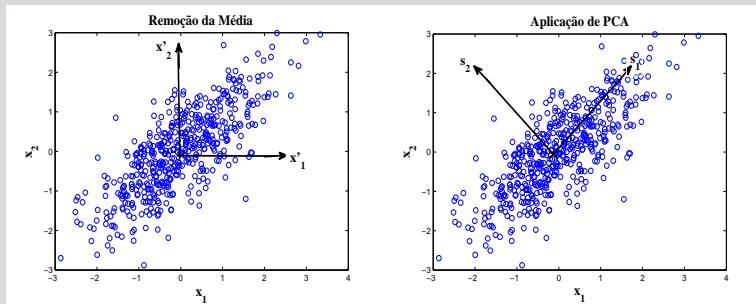
- Existem muitas outras possibilidades de se formar uma base no \mathbb{R}^2 , bastando para isso que os vetores sejam LI.
- A base \mathcal{B}_1 é chamada de base ortonormal $\{\mathbf{e}_1, \mathbf{e}_2\}$, porque os vetores \mathbf{e}_1 e \mathbf{e}_2 são ortogonais (ou perpendiculares) entre si, pois seu produto interno é nulo.
- Além disso, ambos tem norma unitária (i.e. $\|\mathbf{e}_1\| = \|\mathbf{e}_2\| = 1$).
- Com relação à PCA, os vetores de dados originais \mathbf{x}_k são escritos em relação à base canônica do \mathbb{R}^p .
- Enquanto os vetores transformados \mathbf{y}_k são escritos em relação à base formada pelos autovetores da matriz de covariância $\mathbf{C}_{\mathbf{x}}$.

Interpretação Geométrica

- Lembre-se que no espaço original as componentes do vetor \mathbf{x}_k estão correlacionadas, enquanto no espaço transformado as componentes do vetor \mathbf{y}_k não.
- Dito de outra forma, no sistema de coordenadas perpendiculares associado à nova base formada pelos autovetores de \mathbf{C}_x , as componentes de \mathbf{y}_k são descorrelacionadas.
- Do ponto de vista geométrico, o processo de descorrelação levado a cabo via PCA corresponde a uma rotação do sistema de coordenadas no qual os dados são representados.

Interpretação Geométrica

- Graficamente, o processo de diagonalização da matriz de covariância de um conjunto de dados, ou equivalente, de descorrelação dos atributos de um conjunto de dados, está mostrado na figura abaixo.



PCA para Redução de Dimensão

- Se montarmos a matriz \mathbf{V} apenas com os autovetores associados aos q ($q < p$) primeiros autovalores, ou seja

$$\mathbf{V}_q = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_q], \quad (41)$$

o vetor $\mathbf{y}_k = \mathbf{V}^T \mathbf{x}_k$ terá dimensão $q \times 1$. Note que a matriz \mathbf{V}_q tem dimensão $p \times q$.

- Consequentemente, a matriz de covariância dos dados transformados $\mathbf{C}_{\mathbf{y}^{(q)}}$ agora tem dimensão $q \times q$:

$$\mathbf{C}_{\mathbf{y}^{(q)}} = \mathbf{V}_q^T \mathbf{C}_{\mathbf{x}} \mathbf{V}_q = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 \cdots & 0 & \lambda_q \end{bmatrix} \quad (42)$$

Uma Medida da Informação Contida em \mathbf{X}

- Lembrando que nosso objetivo inicial era encontrar uma transformação linear que preservasse a informação relevante contida nos dados originais. Mas, como quantificar a informação relevante em um conjunto de dados?
- Podemos definir a Variância Total (VT) como uma medida da quantidade de informação contida nos dados originais:

$$VT = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_p^2 \quad (43)$$

$$= \lambda_1 + \lambda_2 + \cdots + \lambda_p \quad (44)$$

- Como $\lambda_i = \sigma_i^2$, podemos criar uma medida de quanto da informação (i.e. variância) do conjunto original está sendo representada no autovalor λ_i .

Uma Medida da Informação Contida em \mathbf{X}

- Chamaremos esta medida de variância explicada pelo i -ésimo autovalor (VE_i):

$$VE_i = 100 \times \frac{\lambda_i}{VT} = 100 \times \left(\frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \right) \quad (45)$$

- Consequentemente, a porcentagem da variância total dos dados explicada pelos primeiros q autovalores é dada por:

$$VE(q) = 100 \times \frac{\sum_{i=1}^q \lambda_i}{VT} \quad (46)$$

$$= 100 \times \left(\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_q}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \right) \quad (47)$$

Implementação em Matlab/Octave

- Assumiremos que os dados estão dispostos como na matriz \mathbf{X} definida na Eq. (3).
- Assim, primeiro passo consiste na estimação da matriz de covariância dos dados (\mathbf{C}_x).

```
>> Cx=cov(X') ;
```

- Atenção:** A matriz \mathbf{X} entra transposta no comando COV porque, por convenção, o Matlab considera que os dados estão dispostos ao longo das linhas de \mathbf{X} , e não ao longo das colunas.

Implementação em Matlab/Octave (cont.-1)

- O segundo passo consiste em determinar os autovalores e autovetores da matriz C_X .

```
>> [V L]=eig(Cx);    % matrizes de autovetores/valores  
>> L=diag(L);        % vetor de autovalores nao-ordenados  
>> [L I]=sort(L,'descend');    % autovalores ordenados  
>> V=V(:,I);         % ordena autovetores associados
```

- **Atenção:** A função EIG retorna uma matriz “L” cujo os autovalores estão na diagonal principal. Daí a necessidade de se usar em seguida o comando DIAG, para extrair os autovalores e colocá-los em um vetor.

Implementação em Matlab/Octave (cont.-2)

- O terceiro passo consiste em determinar os Q maiores autovalores responsáveis por explicar, pelo menos, $tol\%$ da informação contida nos dados originais.

```
>> tol=0.95;    % variancia a ser mantida
>> VQ=cumsum(L)/sum(L);    % variancia explicada
>> Q=length(find(VE<=tol));    % Num. compon. principais
>> Vq=V(:,1:Q);    % matriz com Q primeiros autovetores
>> Y=Vq'*X;    % dados transformados
```

- **Atenção:** O número de componentes principais vai variar em função do valor de tol . Quanto maior (menor) o valor de tol , maior (menor) será o valor de Q .