

Exemplo de Aplicação de Transformações Lineares: Sistema de Auxílio ao Diagnóstico Médico

Guilherme de Alencar Barreto

`gbarreto@ufc.br`

Departamento de Engenharia de Teleinformática (DETI)
Curso de Graduação em Engenharia de Teleinformática (CGETI)
Universidade Federal do Ceará – UFC
<http://www.deti.ufc.br/~guilherme>

Conteúdo dos Slides

- 1 Transformadas Matriciais
- 2 Descrição do Problema
- 3 Diagnóstico Médico via Software
- 4 Exemplo Teórico-Computacional

Transformadas Matriciais

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

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

em que A é uma matriz $m \times n$.

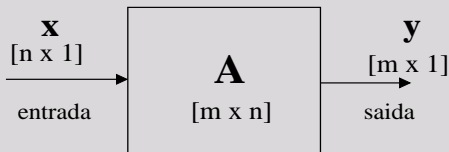
- Para simplificar, muitas vezes denotamos essa transformação matricial por

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

- Note que o domínio de T é o \mathbb{R}^n quando A tem n colunas, e o contra-domínio de T é o \mathbb{R}^m quando cada coluna de A tem m elementos.

Diagrama de Blocos

Ajuda muito no entendimento de uma transformação linear se representarmos a relações $y = Ax$ (ou $b = Ax$) na forma de um diagrama de blocos do tipo entrada-saída.



Definição do Problema

- Considere que um médico tem que diagnosticar a doença de pele de um certo paciente com base em
 - **Informações clínicas:** informações coletadas pelo médico durante a anamnese e inspeção visual da pele no consultório.
 - **Informações histopatológicas:** normalmente resultam de uma biópsia, ou seja, da análise do tecido em um laboratório de patologia.

Doenças de Pele Envolvidas no Problema

Após um período, tal médico coletou tais informações sobre 358 pacientes e suas respectivas patologias.

Doença (número de pacientes)
Psoríase(111)
Dermatite seborréica(60)
Líquen plano(71)
Pitiríase rósea(48)
Dermatite crônica(48)
Pitiríase rubra pilar(20)

Informações de Natureza Clínica

Clínicos

- 1: eritema
- 2: escala
- 3: bordas definidas
- 4: coceira
- 5: fenômeno de Koebner
- 6: pápulas poligonais
- 7: pápulas foliculares
- 8: envolvimento da mucosa oral
- 9: envolvimento do joelho e do cotovelo
- 10: envolvimento do escalpo
- 11: histórico familiar
- 34: idade

Informações de Natureza Histopatológica

Histopatológicos

- | | |
|---|--|
| 12: incontinência de melanina | 23: pústulas espongiformes |
| 13: eosinófilos no infiltrado | 24: microabscesso de Munro |
| 14: infiltrado PNL | 25: hipergranulose focal |
| 15: fibrose na derme papilar | 26: ausência da camada granulosa |
| 16: exocitose | 27: vacuolização e destruição da
camada basal |
| 17: acantose | 28: espongiase |
| 18: hiperqueratose | 29: aspecto “dente de serra” das
cristas interpapilares |
| 19: parakeratose | 30: tampões cárneos foliculares |
| 20: dilatação em clava dos
cones epiteliais | 31: parakeratose perifolicular |
| 21: alongamento dos cones
epiteliais da epiderme | 32: infiltrado inflamatório mononuclear |
| 22: estreitamento da epiderme
suprapapilar | 33: infiltrado em banda |

Banco de Dados dos Pacientes

- Cada medida clínica ou histopatológica pode ser entendida como uma variável que o médico usa para guiar sua decisão (diagnóstico).
- O médico organiza em um fichário (ou computador) as informações de cada um dos 358 pacientes e o valor numérico correspondente de cada medida clínica ou histopatológica.
- De posse deste banco de dados, usando a Álgebra Linear é possível desenvolver um sistema computacional capaz de “diagnosticar” as seis doenças de pele descritas anteriormente, de modo semelhante ao dermatologista!
- Para isso, precisamos formular o problema de diagnóstico médico como uma transformação linear $y = Ax$.

Formalização Matemática do Problema

- Cada paciente vai ser representado por um vetor de dimensão $n = 34$:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{32} \\ x_{33} \\ x_{34} \end{bmatrix} = \begin{bmatrix} \text{eritema} \\ \text{escala} \\ \text{bordas definidas} \\ \vdots \\ \text{infiltrado inflamatório mononuclear} \\ \text{infiltrado em banda} \\ \text{idade} \end{bmatrix} \quad (3)$$

- Para este problema, $x_j \in \{0, 1, 2, 3, 4\}$, $j = 1, 2, \dots, 33$.
- Somente a variável x_{34} assume valores maiores que 4.

Formalização Matemática do Problema

- A cada paciente é também associado um vetor-código de dimensão ($m = 6$) para a sua patologia, isto é, um identificador (ID) para a sua patologia.

$$\text{Psoríase: } \mathbf{y} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \text{Derm. Seborréica: } \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \text{Líquen Plano: } \mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (4)$$

$$\text{Pitiríase rósea: } \mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \text{Derm. crônica: } \mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{Pitiríase rubra pilar: } \mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (5)$$

Formalização Matemática do Problema

- Note que no banco de dados teremos 358 vetores $\mathbf{x}_i \in \mathbb{R}^{34}$ e 358 vetores $\mathbf{y}_i \in \mathbb{R}^6$, $i = 1, \dots, 358$, representando 358 pacientes e suas respectivas patologias.
- O índice i denota o i -ésimo paciente no banco de dados.
- Note que o objetivo é determinar uma matriz \mathbf{A} que para um dado vetor de entrada (paciente) \mathbf{x}_i forneça o vetor-código (diagnóstico) da patologia correspondente:

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i, \quad \forall i = 1, \dots, 358 \quad (6)$$

- Note que a matriz \mathbf{A} , de dimensões 6×34 , atua como se fosse uma versão matemática do médico especialista em questão.

Formalização Matemática do Problema

- Para facilitar, podemos organizar os 358 pacientes e os vetores-código de suas patologias nas colunas das matrizes \mathbf{X} e \mathbf{Y} , dadas por:

$$\mathbf{X} = [\mathbf{x}_1 \mid \mathbf{x}_2 \mid \cdots \mid \mathbf{x}_{358}] \quad (7)$$

e

$$\mathbf{Y} = [\mathbf{y}_1 \mid \mathbf{y}_2 \mid \cdots \mid \mathbf{y}_{358}] \quad (8)$$

- Note que a matriz \mathbf{X} tem dimensões 34×358 e a matriz \mathbf{Y} tem dimensões 6×358 .

Formalização Matemática do Problema

- A versão matricial da transformação $\mathbf{y}_i = \mathbf{A}\mathbf{x}_i$ é dada por:

$$\mathbf{Y}_{[6 \times 358]} = \mathbf{A}_{[6 \times 34]} \mathbf{X}_{[34 \times 358]} \quad (9)$$

- Note que as matrizes \mathbf{X} e \mathbf{Y} são montadas a partir do banco de dados de pacientes, enquanto a matriz \mathbf{A} , que define a transformação, é desconhecida.
- Note também que, como a matriz $\mathbf{X}_{[34 \times 358]}$ é retangular (i.e. não-quadrada), não podemos obter sua inversa a fim de isolar a matriz \mathbf{A} na expressão acima.

Formalização Matemática do Problema

- A fim de isolar a matriz \mathbf{A} , vamos usar de um artifício baseado apenas nas dimensões das matrizes \mathbf{Y} , \mathbf{A} e \mathbf{X} .
- Se a matriz \mathbf{X} fosse quadrada, poderíamos obter sua inversa e isolar a matriz \mathbf{A} .
- A grande “sacada” do artifício está em manipular (ou atuar sobre) a matriz \mathbf{X} a fim de obter uma matriz quadrada.
- Para isso, vamos multiplicar (pela direita) ambos os lados da equação acima pela matriz \mathbf{X}^T , que é a matriz transposta de \mathbf{X} , obtendo a seguinte expressão:

$$\mathbf{Y}_{[6 \times 358]} \mathbf{X}_{[358 \times 34]}^T = \mathbf{A}_{[6 \times 34]} \mathbf{X}_{[34 \times 358]} \mathbf{X}_{[358 \times 34]}^T \quad (10)$$

Formalização Matemática do Problema

- Com isso, percebemos que a matriz $\mathbf{X}\mathbf{X}^T$ é quadrada, de dimensão 34×34 , podendo assim ser invertida.
- Multiplicando ambos os lados da equação (pela direita) por $(\mathbf{X}\mathbf{X}^T)^{-1}$, obtemos:

$$\mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1} = \mathbf{A}\mathbf{X}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1} \quad (11)$$

- De onde resulta a seguinte expressão para cálculo da matriz de transformação \mathbf{A} :

$$\boxed{\mathbf{A} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}} \quad (12)$$

Formalização Matemática do Problema

- De posse da matriz \mathbf{A} , podemos usá-la como componente de um software de tomada de decisão voltado para o auxílio ao diagnóstico médico.
- Matematicamente, isto pode ser feito por meio da seguinte transformação matricial:

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i, \quad (13)$$

em que o vetor $\mathbf{x}_i \in \mathbb{R}^{34}$ é a versão “numérica” do paciente, enquanto o vetor $\mathbf{y}_i \in \mathbb{R}^6$ é a versão numérica (código) da patologia.

- Cabe ao desenvolvedor do sistema, desenvolver uma interface amigável de modo a tornar a operação matemática acima transparente para o usuário.

Exemplo de Diagnóstico

- Um certo paciente do médico usuário do sistema computacional de auxílio ao diagnóstico foi codificado pelo seguinte vetor de atributos:

$$\mathbf{x}_i = [2 \ 1 \ 2 \ 3 \ 1 \ 3 \ 0 \ 3 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 2 \ 0 \ 2 \ 0 \ 0 \ 0 \ 0 \ 0 \ 2 \ 0 \ 2 \ 3 \ 2 \ 0 \ 0 \ 2 \ 3 \ 26]^T$$

- Ao multiplicarmos este vetor pela matriz \mathbf{A} calculada na Equação (12), obtemos o seguinte vetor de saída \mathbf{y}_i :

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i = \begin{bmatrix} -0.076297 \\ 0.113172 \\ 1.061544 \\ -0.123137 \\ -0.098041 \\ 0.015406 \end{bmatrix} \quad (14)$$

Exemplo de Diagnóstico (cont.)

- Analizando as componentes do vetor y_i , percebemos que a maior delas é a terceira componente (i.e. $y_{i3} = 1.061544$).
- Isto deve ser interpretado da seguinte forma:

O sistema computacional está sugerindo que o paciente, cujos os valores numéricos dos 12 atributos clínicos e os 22 histopatológicos constam no vetor x_i , apresenta características da patologia **Líquen Plano**.

Implementação no Matlab/Octave

De posse das matrizes de vetores de atributos dos pacientes \mathbf{X} e da matriz de códigos de suas patologias \mathbf{Y} , a matriz de transformação \mathbf{A} pode ser calculada facilmente no Matlab/Octave através do seguinte comando:

```
>> A = Y*X'*inv(X*X');
```

ou usando o operador *barra invertida* “/”:

```
>> A = Y/X;
```

Deve-se dar preferência à segunda opção por ser menos susceptível a erros numéricos. A primeira opção também não é recomendada para problemas de alta dimensionalidade. Faça o teste!