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)
Engenharias de Computação, Telecomunicações e Teleinformática
Universidade Federal do Ceará – UFC
www.researchgate.net/profile/Guilherme_Barreto2/

Conteúdo dos Slides

Transformadas Matriciais

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

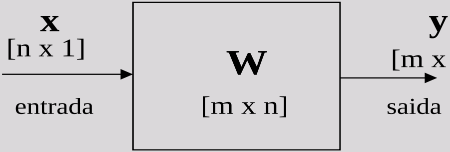
$$\mathbf{y} = \mathbf{W}\mathbf{x} \quad (\text{ou} \quad \mathbf{W}\mathbf{x} = \mathbf{y}), \tag{1}$$

em que $\mathbf{W}$ é uma matriz $m \times n$.

- Para simplificar, muitas vezes denotamos essa transformação matricial por
$$\boxed{\mathbf{x} \mapsto \mathbf{W}\mathbf{x}} \tag{2}$$
- Note que o domínio de T é o $\mathbb{R}^n$ quando $\mathbf{W}$ tem n colunas, e o contra-domínio de T é o $\mathbb{R}^m$ quando cada coluna de $\mathbf{W}$ tem m elementos.

Diagrama de Blocos

Ajuda muito no entendimento de uma transformação linear se representarmos a relações $\mathbf{y} = \mathbf{W}\mathbf{x}$ 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

Clìnicos

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
13: eosinófilos no infiltrado
14: infiltrado PNL
15: fibrose na derme papilar
16: exocitose
17: acantose
18: hiperceratose
19: paraceratose
20: dilatação em clava dos cones epiteliais
21: alongamento dos cones epiteliais da epiderme
22: estreitamento da epiderme suprapapilar
23: pústulas espongiformes
24: microabscesso de Munro
25: hipergranulose focal
26: ausência da camada granulosa
27: vacuolização e destruição da camada basal
28: espongiose
29: aspecto "dente de serra" das cristas interpapilares
30: tampões córneos foliculares
31: paraceratose perifolicular
32: infiltrado inflamatório mononuclear
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 doenças de pele descritas anteriormente, de modo semelhante ao dermatologista!
- Para isso, precisamos formular o problema de diagnóstico médico como uma transformação linear $\mathbf{y} = \mathbf{W}\mathbf{x}$.

Formalização Matemática do Problema

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

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

- Para este problema, $x_j \in \{0, 1, 2, 3, 4\}$, $j = 1, 2, ..., 33$.
- 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.

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

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

Formalização Matemática do Problema

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

$$\mathbf{y}_k = \mathbf{W}\mathbf{x}_k, \quad \forall k = 1, \ldots, N = 358 \tag{6}$$

- Note que a matriz $\mathbf{W}$, 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 $N = 358$ pacientes e os vetores-código de suas patologias nas colunas das matrizes $\mathbf{X}$ e $\mathbf{Y}$, dadas por:

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

e

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

- 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 $\mathbf{y}_i = \mathbf{W}\mathbf{x}_i$ é dada por:

$$\mathbf{Y}_{[6\times 358]} = \mathbf{W}_{[6\times 34]}\mathbf{X}_{[34\times 358]} \tag{9}$$

- Note que as matrizes $\mathbf{X}$ e $\mathbf{Y}$ são montadas a partir do banco de dados de pacientes, enquanto a matriz $\mathbf{W}$, 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 $\mathbf{W}$ na expressão acima.

Formalização Matemática do Problema

- A fim de isolar a matriz $\mathbf{W}$, vamos usar de um artifício baseado apenas nas dimensões das matrizes $\mathbf{Y}$, $\mathbf{W}$ e $\mathbf{X}$.
- Se a matriz $\mathbf{X}$ fosse quadrada, poderíamos obter sua inversa e isolar a matriz $\mathbf{W}$.
- 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:

$$\mathbf{Y}_{[6\times 358]}\mathbf{X}^T_{[358\times 34]} = \mathbf{W}_{[6\times 34]}\mathbf{X}_{[34\times 358]}\mathbf{X}^T_{[358\times 34]} \qquad (10)$$

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:

$$\mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1} = \mathbf{W}\mathbf{X}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1} \tag{11}$$

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

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

Regularização de Tikhonov

- A fim de evitar problemas causados pelo mal-condicionamento da matriz $\mathbf{X}\mathbf{X}^T$, costuma-se fazer uso da regularização de Tikhonov.
- A matriz de coeficientes $\mathbf{W}$ passa a ser estimada por meio da seguinte expressão:

$$\boxed{\mathbf{W} = \mathbf{Y}\mathbf{X}^T \left(\mathbf{X}\mathbf{X}^T + \lambda \mathbf{I}_n\right)^{-1}} \tag{13}$$

  em que $0 < \lambda \leq 1$ é chamada de constante de regularização e $\mathbf{I}_n$ é a matriz identidade de ordem n.
- Note que a constante λ é, na verdade, um **iperparâmetro**.
- Um hiperparâmetro é um parâmetro do model que deve ser pré-definido para que os parâmetros da função discriminante propriamente dita possam ser estimados.

Formalização Matemática do Problema

- De posse da matriz $\mathbf{W}$, 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:

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \tag{14}$$

em que o vetor $\mathbf{x} \in \mathbb{R}^{34}$ denota a versão "numérica" de um paciente qualquer, enquanto o vetor $\mathbf{y} \in \mathbb{R}^{6}$ simboliza o código numérico identificador da patologia.
- Cabe ao desenvolvedor do sistema, desenvolver uma interface amigável de modo a tornar a operação matemática acima transparente para o usuário.

Formalização Matemática do Problema

- Note que a expressão da Eq. (14) pode ser decomposta em m saídas individuais, ou seja

$$\boxed{y_i = \mathbf{w}_i^T \mathbf{x}} \tag{15}$$

- Esta expressão é chamada de *função discriminante linear* da i-ésima classe, $i = 1, \ldots, m$.
- No presente estudo caso, a i-ésima classe corresponde à i-ésima patologia.
- O vetor $\mathbf{w}_i$, de dimensão 34×1, é chamado de vetor de coeficientes da função discriminante da i-ésima classe.

Formalização Matemática do Problema

- O vetor $\mathbf{w}_i^T$ corresponde à i-ésima linha da matriz $\mathbf{W}$, $i = 1, \ldots, m$.
- Pondendo dessa forma ser estimado a partir das matrizes $\mathbf{X}$ e Y por meio da seguinte expressão:

$$\boxed{\mathbf{w}_i^T = \mathbf{Y}_{[i,:]}\mathbf{X}^T \left(\mathbf{X}\mathbf{X}^T\right)^{-1}}, \tag{16}$$

em que $\mathbf{Y}_{[i,:]}$ simboliza a i-ésima linha da matriz $\mathbf{Y}$.

- A expressão anterior resulta em um vetor linha. Para obter o vetor coluna $\mathbf{w}_i$, podemos utilizar a seguinte expressão:

$$\boxed{\mathbf{w}_i = \left(\mathbf{X}\mathbf{X}^T\right)^{-1} \mathbf{X}\mathbf{Y}_{[i,:]}^T} \tag{17}$$

Exemplo de Diagnóstico

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

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

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

$$\mathbf{y}_{new} = \mathbf{W}\mathbf{x}_{new} \tag{18}$$

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \\ \hat{y}_5 \\ \hat{y}_6 \end{bmatrix} = \begin{bmatrix} -0.076297 \\ 0.113172 \\ 1.061544 \\ -0.123137 \\ -0.098041 \\ 0.015406 \end{bmatrix} \tag{19}$$

Exemplo de Diagnóstico (cont.)

- Analizando as componentes do vetor $\hat{\mathbf{y}}$, percebemos que a maior delas é a terceira componente (i.e. $y_3 = 1.061544$).
- Assim, a regra de decisão é formalmente dada por

$$j^* = \text{índice da classe de } \mathbf{x}_{new} = \arg \max_{\forall j} \{\hat{y}_j\}, \qquad (20)$$

em que a função $\max$ retorna o maior valor entre todas as saídas $\hat{y}_j$ e a função $\arg$ retorna o índice (i.e. a posição) da maior saída dentro do vetor.

- Assim, o sistema computacional está sugerindo que o paciente pelo vetor de atributos clínicos e histopatológicos, $\mathbf{x}_{new}$, apresenta características da patologia **Líquen Plano**.

Implementação no Matlab/Octave

- A Eq. (12) pode ser implementada facilmente no Matlab/Octave por meio da seguinte linha de comando:

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

- Contudo, este procedimento não é recomendado devido ao seu elevado custo computacional (exige muito uso de memória) e alta susceptibilidade a erros numéricos quando a matriz $\mathbf{X}\mathbf{X}^T$ está próxima da singularidade.
- Uma forma de avaliar se esta expressão produzirá resultados confiáveis é verificando o posto da matriz $\mathbf{X}$.
- O resultado será confiável apenas se a matriz for de posto completo, i.e., posto$(\mathbf{X}) = \min(n, N)$.
- Além disso, esta expressão não escalona bem para dados de alta dimensão.

Implementação no Matlab/Octave (cont.-1)

- Para evitar problemas na inversão da matriz $\mathbf{X}\mathbf{X}^T$, costuma-se usar a regularização de Tikhonov.
- Para isso, faz-se necessário definir uma constante $0 < \lambda \leq 1$, chamada de constante de regularização.
- Assim, tem-se que usar as seguintes linhas de comando:

```
>> lam=0.01;

>> W = Y*X'*inv(X*X'+lam*eye(size(X*X')));
```

- Esta expressão favore soluções de norma mínima, porém ainda consome muita memória e não escalona bem para dados de alta dimensão.

Implementação no Matlab/Octave (cont.-1)

- Uma segunda alternativa muito usada para implementar a Eq. (12) no Matlab/Octave faz uso da seguinte linha de comando:

```
>> W = Y/X;
```

- Esta expressão usa o operador *barra invertida* "/", que usa a decomposição QR para gerar o resultado.
- Esta decomposição escalona bem para dados de alta dimensão.
- Contudo, a decomposição QR exige que a matriz $\mathbf{X}$ seja de posto completo, i.e., com posto$(\mathbf{X}) = \min(n, N)$.

Implementação no Matlab/Octave (cont.-2)

- Para evitar problemas com matrizes de posto incompleto, recomenda-se o uso do comando `pinv`:

```
>> W = Y*pinv(X);
```

- Este procedimento escalona muito bem para dados de alta dimensão, pois faz uso eficiente de memória.
- Além disso, utiliza uma versão aproximada da *Decomposição em Valores Singulares* (SVD) para calcular a inversa de $\mathbf{X}\mathbf{X}^T$ e tratar matrizes com posto($\mathbf{X}$) $< \min(n, N)$, i.e. de posto incompleto.

Regra de Aprendizado do Perceptron Simples

- Perceba que a determinação da matriz $\mathbf{W}$ via Eq. (12) exige o armazenamento em memória de todos os vetores de atributos $\mathbf{x}_i$ e seus respectivos vetores-rótulos $\mathbf{y}_i$, $i = 1, 2, \ldots, 358$.
- Uma alternativa mais econômica no uso de memória consiste em utilizar a regra de aprendizado do Perceptron, que na sua forma matricial é dada por

$$\boxed{\mathbf{W}(n+1) = \mathbf{W}(n) + \eta \mathbf{e}(n)\mathbf{x}^T(n)} \tag{21}$$

em que $0 < \eta \ll 1$ é chamada de passo de aprendizagem e n denota o instante ou iteração atual.

Regra de Aprendizado do Perceptron Simples

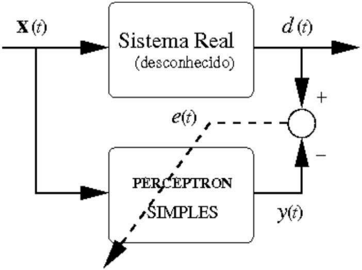
- O vetor de erros na iteração n é dado por

$$\begin{aligned} \mathbf{e}(n) &= \mathbf{y}(n) - \hat{\mathbf{y}}(n), & (22) \\ &= \mathbf{y}(n) - \text{sinal}(\mathbf{W}(n)\mathbf{x}(n)). & (23) \end{aligned}$$

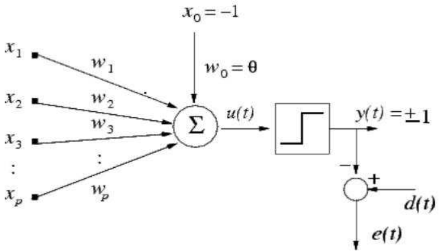
em que $\hat{\mathbf{y}}(n) = \text{sinal}(\mathbf{W}(n)\mathbf{x}(n))$ denota a saída da rede naquele instante.

Regra de Aprendizado do Perceptron Simples

O processo de aprendizagem, ou seja, de modificação dos parâmetros do neurônio M-P é guiado pelo erro (e) e pelo vetor de entrada ($\mathbf{x}$)!



Neurônio de McCulloch & Pitts + Regra de Aprendizado



$$\mathbf{w}(t+1) = \mathbf{w}(t) + \eta e(t)\mathbf{x}(t)$$

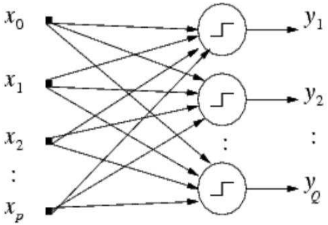
Algoritmo Perceptron Simples (1 neurônio)

1. Início (t=0)
   - 1.1 – Definir valor de η entre 0 e 1.
   - 1.2 – Iniciar $\mathbf{w}(0)$ com valores nulos ou aleatórios.
2. Funcionamento
   - 2.1 – Selecionar vetor de entrada $\mathbf{x}(t)$.
   - 2.2 – Calcular ativação $u(t)$.
   - 2.3 – Calcular saída $y(t)$.
3. Treinamento
   - 3.1 – Calcular erro: $e(t) = d(t) - y(t)$
   - 3.2 – Ajustar pesos via regra de aprendizagem.
   - 3.3 – Verificar critério de parada.
     - 3.3.1 – Se atendido, finalizar treinamento.
     - 3.3.2 – Caso contrário, fazer t=t+1 e ir para Passo 2.

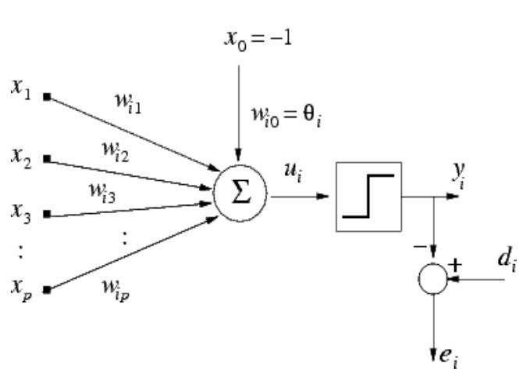
Arquitetura da Rede Perceptron Simples (Q neurônios)

Um único neurônio M-P categoriza apenas duas classes de dados.

Em problemas com múltiplas classes, deve-se utilizar vários neurônios em paralelo.



Representação do i-ésimo neurônio da rede PS.



Funcionamento do i-ésimo neurônio da rede PS.

O funcionamento de cada neurônio individualmente é o mesmo.

Assim, a ativação do i-ésimo neurônio da rede PS é dada por:

$$u_i = \mathbf{w}_i^T \mathbf{x} = w_{i1} x_1 + w_{i2} x_2 + \ldots + w_{ip} x_p$$

A saída do i-ésimo neurônio é dada por:

$$y_i = \text{sinal}(u_i) = \text{sinal}(\mathbf{w}_i^T \mathbf{x})$$

O erro do i-ésimo neurônio é dado por: $e_i = d_i - y_i$

onde d_i é a saída desejada do i-ésimo neurônio.

$i = 1, \ldots, Q$ ($Q \geq 1$ é o número de neurônios de saída).

Treinamento do i-ésimo neurônio da rede PS.

Como cada neurônio tem seu próprio vetor de pesos $\mathbf{w}_i$, $i = 1, 2, ..., Q$, então teremos agora Q regras de aprendizagem!

Ou seja, uma regra de aprendizagem para cada vetor $\mathbf{w}_i$.

Assim, a regra de aprendizagem do i-ésimo neurônio é dada por:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \eta e_i(t)\mathbf{x}(t)$$

Em que $0 < \eta << 1$ e $i=1, 2, ..., Q$.

Algoritmo Perceptron Simples (Q neurônios)

1. Início (t=0)
   - 1.1 – Definir valor de η entre 0 e 1.
   - 1.2 – Iniciar $\mathbf{w}_i(0)$ com valores aleatórios.
2. Funcionamento
   - 2.1 – Selecionar o vetor de entrada $\mathbf{x}(t)$.
   - 2.2 – Calcular as Q ativações $u_i(t)$.
   - 2.3 – Calcular as Q saídas $y_i(t)$.
3. Treinamento
   - 3.1 – Calcular os Q erros: $e_i(t) = d_i(t) - y_i(t)$
   - 3.2 – Ajustar os Q vetores de pesos $\mathbf{w}_i(t)$.
   - 3.3 – Verificar critério de parada.
     - 3.3.1 – Se atendido, finalizar treinamento.
     - 3.3.2 – Caso contrário, fazer t=t+1 e ir para Passo 2.