

SCC0270/SCC5809 - Redes Neurais

Aula 12 - PCA e Rede PCA

Profa. Dra. Roseli Aparecida Francelin Romero
SCC - ICMC - USP

2017

Análise de Componentes Principais (PCA)

- Dado um conjunto de amostras, cada qual com um conjunto finito de variáveis.
- Derivar novas componentes que produzam uma descrição mais simples do sistema.
- Reduzir as variáveis originais a um número menor de variáveis ortogonais (não correlacionadas).
- Mudança de espaço de variáveis.

Análise de Componentes Principais (PCA)

Objetivo

Dadas p variáveis, deseja-se achar combinações lineares dessas para produzir índices que **não** sejam correlacionados, de tal forma que:

- Índices Z : componentes principais.

Sumário

1 PCA clássica

- Caracterização
- Exemplo
- Redução de dimensionalidade
- Algoritmo

2 Rede PCA adaptativa

Análise de Componentes Principais (PCA)

- i -ésima componente principal.

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$$

- Com restrição:

$$a_{i1}^2 + a_{i2}^2 + \dots + a_{ip}^2 = 1$$

- E com $Z_1, Z_2, \dots, Z_{i-1}, Z_i$ não correlacionados.

Análise de Componentes Principais (PCA)

- **PCA:** resume-se em encontrar os autovalores e autovetores da matriz C de covariância dos dados.

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix}$$

Análise de Componentes Principais (PCA)

- Supondo que os autovalores da matriz C estejam ordenados da seguinte forma:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_j \geq \dots \geq \lambda_p$$

- Os autovetores associados são:

$$a_1, a_2, \dots, a_j, \dots, a_p$$

Análise de Componentes Principais (PCA)

- **Propriedades:**

$$a_i^T a_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

- **Para:**

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p$$

- $(a_{i1}, a_{i2}, \dots, a_{ip})$ são os elementos do i-ésimo autovetor correspondente.

Análise de Componentes Principais (PCA)

- A soma dos autovalores corresponde ao traço da matriz covariância C :

$$\lambda_1 + \lambda_2 + \dots + \lambda_p = c_{11} + c_{22} + \dots + c_{pp}$$

- $\text{var}(Z_1) \geq \text{var}(Z_2) \geq \dots \geq \text{var}(Z_p)$

$$\text{var}(Z_i) = \lambda_i$$

Sumário

1 PCA clássica

- Caracterização
- Exemplo
- Redução de dimensionalidade
- Algoritmo

2 Rede PCA adaptativa

Exemplo: conjunto Iris.dat

Tabela 1: Aplicando PCA na base de dados Iris.dat

		Autovetores (coeficientes)			
Componente	Autovalor	X ₁	X ₂	X ₃	X ₄
1	2.91082	0.522371	-0.263356	0.581254	0.565611
2	0.92122	0.372320	0.925556	0.021094	0.065417
3	0.14735	-0.721015	0.242033	0.140889	0.633804
4	0.02061	-0.261998	0.124137	0.801155	-0.523543

- Z_1 é responsável por 72.77% do total da variância.
- Z_2 é responsável por 23.03% do total da variância.
- Z_3 é responsável por 3.68% do total da variância.
- Z_4 é responsável por 0.52% do total da variância.

Reconstrução dos dados originais

$$\begin{aligned} Z &= [Z_1, Z_2, \dots, Z_p]^T \\ &= [X^T a_1, X^T a_2, \dots, X^T a_{p-1}]^T \\ &= A^T X \end{aligned}$$

matriz ortogonal $\rightarrow A^T = A^{-1}$

$$X = A \cdot Z = \sum_{i=1}^p Z_i a_i$$

Sumário

1 PCA clássica

- Caracterização
- Exemplo
- Redução de dimensionalidade
- Algoritmo

2 Rede PCA adaptativa

Redução de dimensionalidade

- Sejam $\lambda_1, \lambda_2, \dots, \lambda_m$ os m autovalores da matriz C .
- Então, $X' \sim X$, onde:

$$X' = \sum_{i=1}^m Z_i a_i \quad m < p$$

- Erro: $e = X - X'$, de modo que:

$$\mathbf{e} = \sum_{i=m+1}^p \mathbf{Z}_i \mathbf{a}_i$$

Redução de dimensionalidade

- O vetor de erro e é ortogonal ao vetor X' , que aproxima X .
- **Princípio da ortogonalidade:** $e^T X' = 0$

Aplicação

Tabela 2: Taxas de compressão para blocos 8x8.

Número de Componentes Principais	Dimensão do Bloco Compactado	Taxa de Compressão
1	8x1	$1 - \left(\frac{1}{8}\right) = 87,5\%$
2	8x2	$1 - \left(\frac{2}{8}\right) = 75\%$
3	8x3	$1 - \left(\frac{3}{8}\right) = 62,5\%$
4	8x4	$1 - \left(\frac{4}{8}\right) = 50\%$

Aplicação

Tabela 3: Taxas de compressão para blocos 16x16.

Número de Componentes Principais	Dimensão do Bloco Compactado	Taxa de Compressão
1	16x1	$1 - \left(\frac{1}{16} \right) = 93,75\%$
2	16x2	$1 - \left(\frac{2}{16} \right) = 87,5\%$

Sumário

1 PCA clássica

- Caracterização
- Exemplo
- Redução de dimensionalidade
- Algoritmo

2 Rede PCA adaptativa

-

Normalizando os dados (passo 1 do algoritmo)

- Dados centrados na média e variância 1:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$x_i \leftarrow \frac{(x_i - \bar{x})}{\text{var}(x_i)}$$

- Escalonamento pela variância (quando há variáveis muito dominantes em relação às demais):

$$var(\bar{x}_j) = \frac{1}{n-1} \sum_{i=1}^n (\bar{x}_{ji} - \bar{\bar{x}}_j)^2$$

$$\frac{x_i}{\sqrt{\text{var}(x_i)}}$$

Sumário

1 PCA clássica

2 Rede PCA adaptativa

- RNA para compressão de imagens
- Regra de Hebb
- Convergência
- Algoritmo
- Experimentos e resultados
- Conclusões

Sumário

- 1 PCA clássica
- 2 Rede PCA adaptativa
 - RNA para compressão de imagens
 - Regra de Hebb
 - Convergência
 - Algoritmo
 - Experimentos e resultados
 - Conclusões

RNA para compressão de imagens

- Padrão JPEG: mais utilizado.
- PCA Clássica: método estatístico multivariado.
- Rede PCA Adaptativa: arquitetura de Redes Neurais Artificiais.

Motivação

- Redução da quantidade de dados armazenadas em sistemas computacionais.
- Redução da dimensionalidade de imagens que ocupem grande quantidade de memória.
- Obtenção de métodos que:
 - Atinjam altas taxas de compressão.
 - Não prejudiquem a qualidade visual.

Sumário

1 PCA clássica

2 Rede PCA adaptativa

- RNA para compressão de imagens
- Regra de Hebb
- Convergência
- Algoritmo
- Experimentos e resultados
- Conclusões

Regra de Hebb

Postulado de Hebb

“Quando um axônio da célula A está suficientemente próximo para excitar uma célula B e repetidamente tenta excitá-la, algum processo crescente ou mudanças metabólicas ocorrem em ambas as células.”

- Transformando em regras:
 - Se dois neurônios ligados por uma sinapse são simultaneamente ativados, a intensidade dessa sinapse(conexão) é aumentada.
 - Se dois neurônios ligados por uma sinapse são ativados assincronamente, a intensidade dessa sinapse é diminuída ou até mesmo eliminada.

Regra de Hebb

- Regra de Hebb:

$$\Delta w_{kj}(n) = \eta y_k(n) x_j(n)$$

- Onde:
 - η é uma constante positiva \rightarrow velocidade de aprendizado.
 - $y_k(n)$ é a saída do neurônio k no tempo n .
 - x_j é o j -ésimo elemento do vetor de entrada no tempo n .
- Regra anti-hebbiana:

$$\Delta w_{kj}(n) = -\eta y_k(n) x_j(n)$$

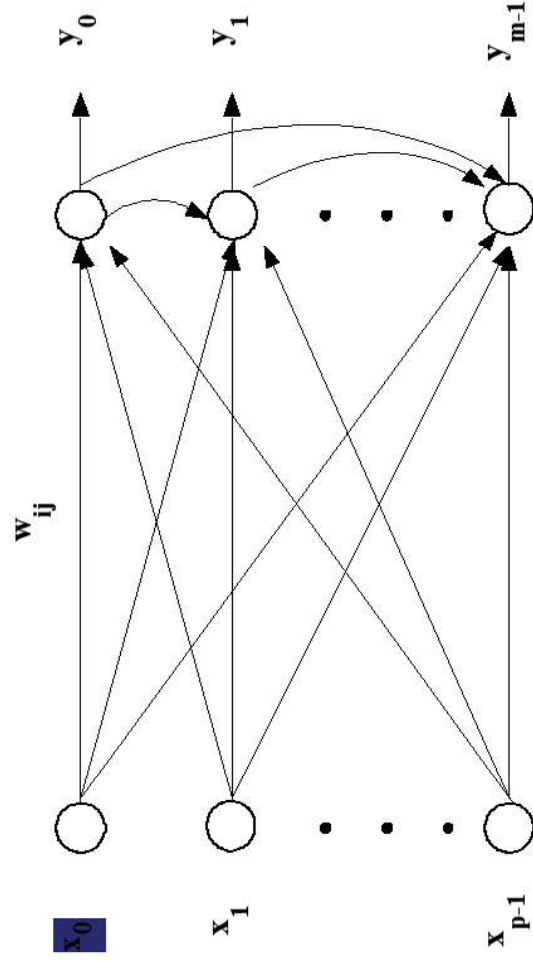
Sumário

1 PCA clássica

2 Rede PCA adaptativa

- RNA para compressão de imagens
- Regra de Hebb
- **Convergência**
- Algoritmo
- Experimentos e resultados
- Conclusões

Rede PCA adaptativa



$$\text{Saída : } y_j(n) = \sum_{i=0}^{p-1} w_{ij}(n) x_i(n) + \sum_{l < j} u_{lj}(n) y_l(n),$$

Ajuste dos pesos: $\Delta w_{ij}(n) = \eta x_i(n) y_j(n)$

Ajuste dos pesos laterais: $\Delta u_{lj}(n) = -\mu y_l(n) y_j(n)$

Teorema de convergência

Teorema

Se a matriz de pesos sinápticos $W(n)$ for associada a valores aleatórios no tempo $n = 0$, então, com probabilidade 1, a regra generalizada de Hebb irá convergir na média e, no limite, irá se aproximar de uma matriz cujas colunas serão os primeiros m autovetores da matriz C de covariância dos vetores de entrada $x(n)$, ordenados por ordem decrescente de autovalor.

Teorema de convergência

- Portanto, no limite, pode-se escrever:

$$\Delta w_j(n) \rightarrow 0, \quad w_j \rightarrow a_j \quad j = 0, 1, \dots, m - 1$$

- Tal que $\|w_j(n)\| = 1$ para todo j .
- Os valores representam os autovetores normalizados associados aos m maiores autovalores da matriz C de covariância dos vetores de entrada $x(n)$, estando esses autovalores ordenados em ordem decrescente.

Teorema de convergência

- Pode-se acelerar a convergência da rede introduzindo um termo *momentum* β e deixando que os parâmetros de aprendizagem e o *momentum* diminuam com o tempo.

$$\Delta w_{ij}(n+1) = \eta(n)x_i y_j + \beta(n)\Delta w_{ij}(n) \quad (1)$$

$$\Delta u_{lj}(n+1) = -\mu(n)y_l y_j + \beta(n)\Delta u_{lj}(n) \quad (2)$$

- Onde:
 - $\eta(n+1) = \max\{\alpha\eta(n), 0.0001\}$
 - $\mu(n+1) = \max\{\alpha\eta(n), 0.0002\}$
 - $\beta(n+1) = \max\{\alpha\eta(n), 0.0001\}$
 - α é o fator de limitação.

Sumário

1 PCA clássica

2 Rede PCA adaptativa

- RNA para compressão de imagens
- Regra de Hebb
- Convergência
- **Algoritmo**
- Experimentos e resultados
- Conclusões

Algoritmo

- 1 Inicialize todos os pesos de conexões com pequenos valores aleatórios, escolha os valores para os parâmetros de aprendizagem e normalize-os em $[0, 1]$.
 - Se normalizar em $[-1, 1]$, pode mudar os sinais dos autovetores.
- 2 Repita:
 - 1 Selecione aleatoriamente um padrão p -dimensional e apresente-o à rede.
 - 2 Ajuste os pesos das conexões entre a camada de entrada e a camada de saída, de acordo com a equação 1.
 - 3 Normalize os vetores-peso (em colunas).
 - 4 Atualize os pesos laterais, de acordo com a equação 2 (não precisa normalizar).
 - 5 Modifique os parâmetros β , η e μ .

Até que todos os pesos laterais sejam suficientemente pequenos **OU** um número de iterações máximo seja atingido.

Sumário

1 PCA clàssica

2 Rede PCA adaptativa

- RNA para compressão de imagens
- Regra de Hebb
- Convergência
- Algoritmo
- Experimentos e resultados
- Conclusões

- Conjunto composto por 208 imagens médicas.
- Dimensão: 480x640 pixels (valor em nível de cinza).

- Conjunto composto por 208 imagens médicas.
- Dimensão: 480x640 pixels (valor em nível de cinza).
- Representação de cortes de um fígado humano
- aquisição a partir de um microscópio laser, pelo *Departamento de Patologia da Fiocruz*.

-

Taxas de Compressão para Blocos 32x32

Número de Componentes Principais	Taxa de Compressão
1	$1 - \left(1/_{32}\right) = 96,875\%$
2	$1 - \left(2/_{32}\right) = 93,75\%$
3	$1 - \left(3/_{32}\right) = 90,625\%$

Taxa de Compressão

Resultados

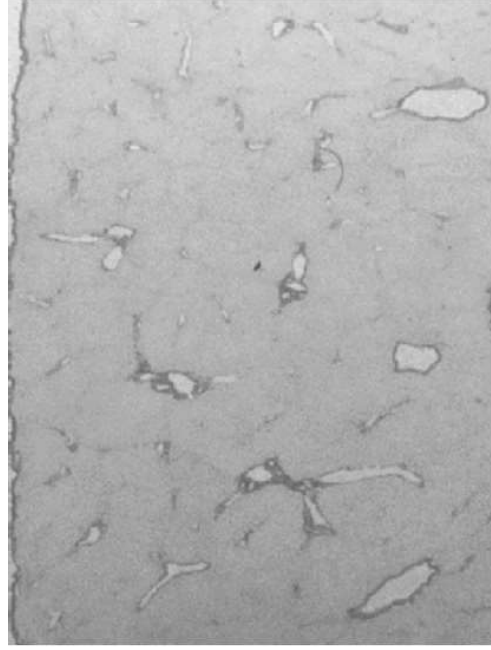
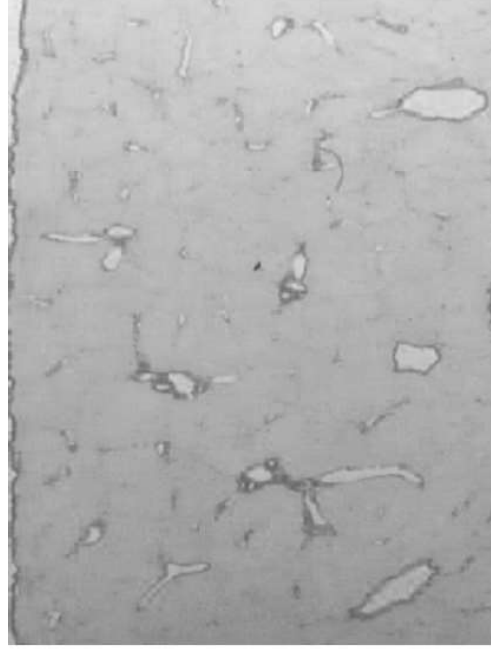
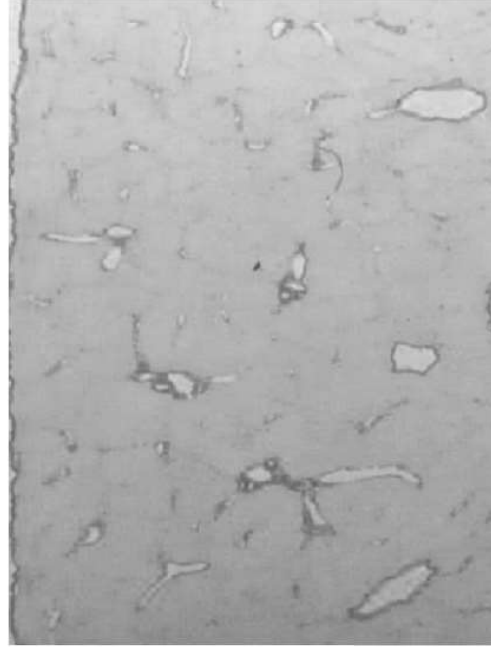


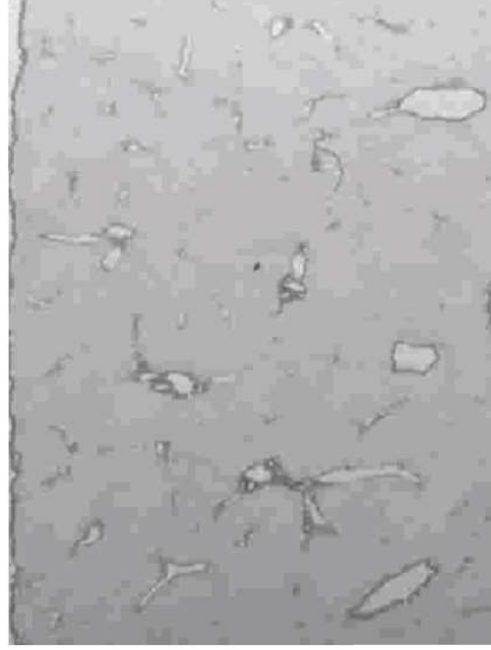
Imagem Original



Compressão PCA Clássica



Compressão Rede PCA



Compressão JPEG

Sumário

- 1 PCA clássica
- 2 Rede PCA adaptativa
 - RNA para compressão de imagens
 - Regra de Hebb
 - Convergência
 - Algoritmo
 - Experimentos e resultados
 - Conclusões

Conclusões

- Três técnicas de compressão foram apresentadas e aplicadas em uma sequência de imagens médicas.
- Analisando os MSE obtidos pelas técnicas:
 - Os resultados da Rede PCA foram bem similares aos obtidos pela PCA clássica.
 - O desempenho do padrão JPEG foi inferior aos desempenhos obtidos pelas duas outras técnicas.

- 45/45