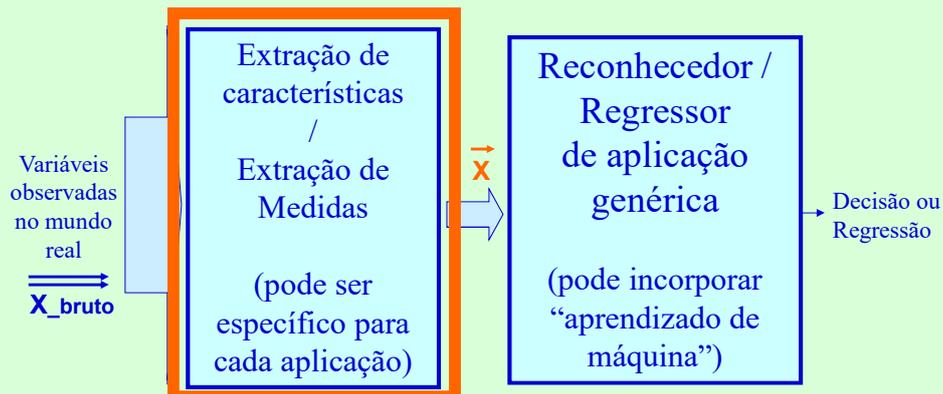


Uma técnica estatística clássica de redução de dimensionalidade do vetor de entradas X e de extração de características com grande número de aplicações: Principal Component Analysis – PCA – Análise de Componentes Principais

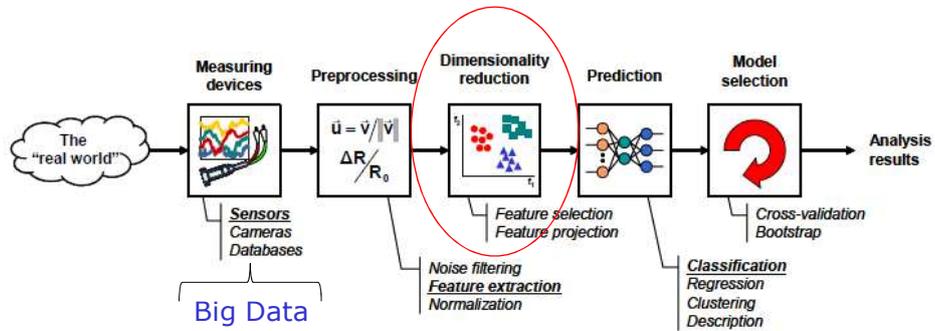
© Prof. Emilio Del Moral – EPUSP

... O 1o estágio gera um Vetor de Medidas, \vec{X}
(o segundo estágio operará sobre tal vetor)



© Prof. Emilio Del Moral Hernandez

PATTERN RECOGNITION and MACHINE INTELLIGENCE



Ricardo Osuna, 2002

Prof. Edson Kitani - 04/2015

56

PCA – Principal Components Analysis

Teoria e Aplicações

Prof. Dr. Edson C. Kitani

Prof. Dr. Emilio Del Moral Hernandez

57

© Prof. Emilio Del Moral – EPUSP



GRUPO DE INTELIGÊNCIA COMPUTACIONAL, MODELAGEM E NEUROCOMPUTAÇÃO ELETRÔNICA – ICONE-EPUSP

LABORATÓRIO DE SISTEMAS INTEGRADOS –LSI
DEPARTAMENTO DE ENGENHARIA DE SISTEMAS ELETRÔNICOS – PSI
ESCOLA POLITÉCNICA DA USP

58

© Prof. Emilio Del Mo



Prof. Dr. Edson C. Kitani

<http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4269320J9>

Formação:

Tecnico em Eletrônica (ETEP – Santo André)
Tecnólogo em Automação Industrial (Centro Universitário de Santo André)
Especialista em Mecânica Fina (Universidade São Judas)
Mestre em Eng. Elétrica (Centro Universitário da FEI)
Doutor em Ciências (POLI – USP)

Atividade profissional:

- Professor Associado na FATEC- Santo André (Eletrônica Automotiva)
- Chefe do Departamento de Projetos de Máquinas na Mahle Anéis

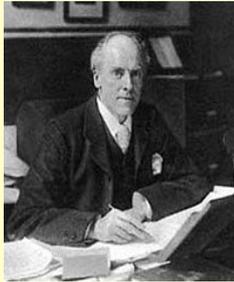
ekitani@lsi.usp.br www.lsi.usp.br/~edson

59

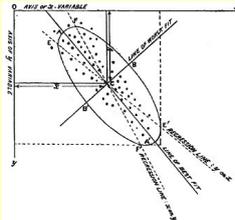
© Prof. Emilio Del Moral⁵⁹ EPUSP

HISTÓRICO DO PCA

Descrito inicialmente por Karl Pearson no artigo, “*On lines and planes of closest fit to systems of points in space*”, *Philosophical Magazine*, 1901.



1857 - 1936

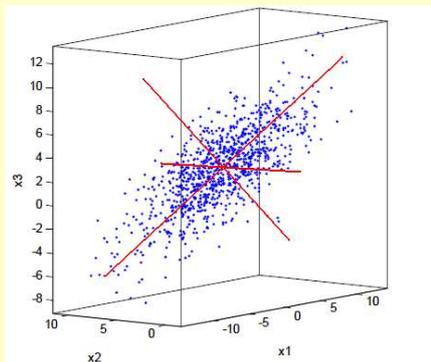


“In many physical, statistical, and biological investigations it is desirable to represent a system of points in plane, three or higher dimensioned space by the best-fitting straight line or plane” (Pearson, 1901).

60

Prof. Edson Kitani - 04/2015 © Prof. Emilio Del Moral EPUSP

Para 3 Dimensões



$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$$

$$\Sigma_{\mathbf{x}} = \frac{1}{(N-1)} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

$$\bar{\mathbf{x}} = [0 \quad 5 \quad 2]^T$$

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} 25 & -1 & 7 \\ -1 & 4 & -4 \\ 7 & -4 & 10 \end{bmatrix}$$

61

Extraído de Osuna 2002

Prof. Edson Kitani - 04/2015 © Prof. Emilio Del Moral EPUSP



Exemplo Numérico com Matlab

```
x = [1, 2; 3, 3; 3, 3; 5, 5; 4, 5; 6, 5; 6, 5; 8, 7; 9, 8]
```

```
figure(1)
```

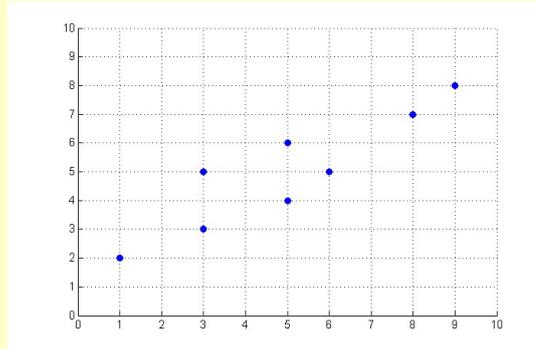
```
scatter(x(:,1),x(:,2))
```

```
Media = mean(x)
```

$$\bar{x} = [5 \quad 5]^T$$

```
Sigma = cov(x,1)Nota
```

$$\Sigma_x = \begin{bmatrix} 6,25 & 4,25 \\ 4,25 & 3,5 \end{bmatrix}$$



A normalização é realizado para N.

62

Prof. Edson Kitani - 04/2015 © Prof. Emilio Del Moral EPUSP



Autovalores (λ) e Autovetores (v)

Associado a cada matriz quadrada $A = ((a_{ij}))$ de ordem n , temos uma função

$$f(\lambda) = |A - \lambda I| = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{vmatrix}$$

chamada de Função Característica da matriz A. E a função

$$f(\lambda) = |A - \lambda I| = 0,$$

que pode ser expressa na forma polinomial

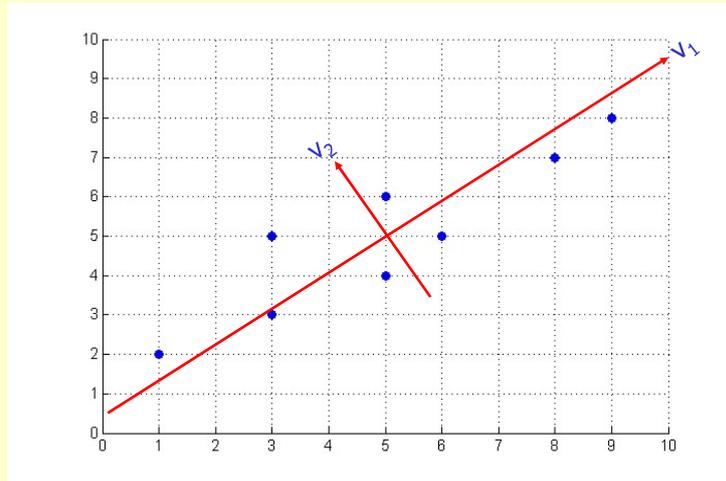
$$c_0 \lambda^n + c_1 \lambda^{n-1} + \dots + c_{n-1} \lambda + c_n = 0$$

é chamada de Equação Característica da matriz A.

63

Prof. Edson Kitani - 04/2015 © Prof. Emilio Del Moral EPUSP

Exemplo Numérico

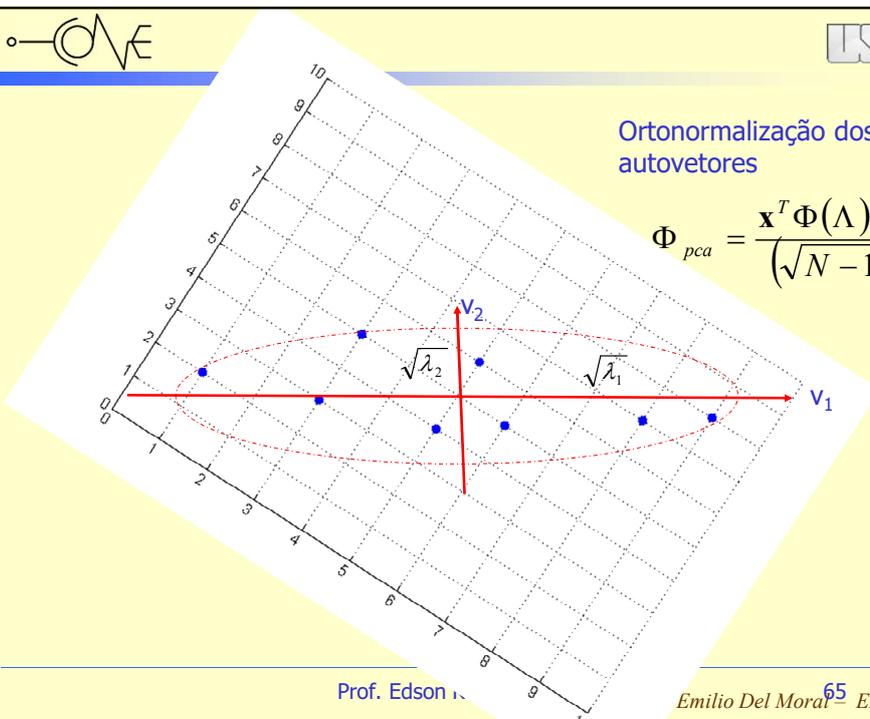


64

Prof. Edson Kitani - 04/2015 © Prof. Emilio Del Moral - EPUSP

Ortonormalização dos autovetores

$$\Phi_{pca} = \frac{\mathbf{x}^T \Phi(\Lambda)^{-\frac{1}{2}}}{(\sqrt{N-1})}$$



65

Prof. Edson Kitani - 04/2015 © Prof. Emilio Del Moral - EPUSP

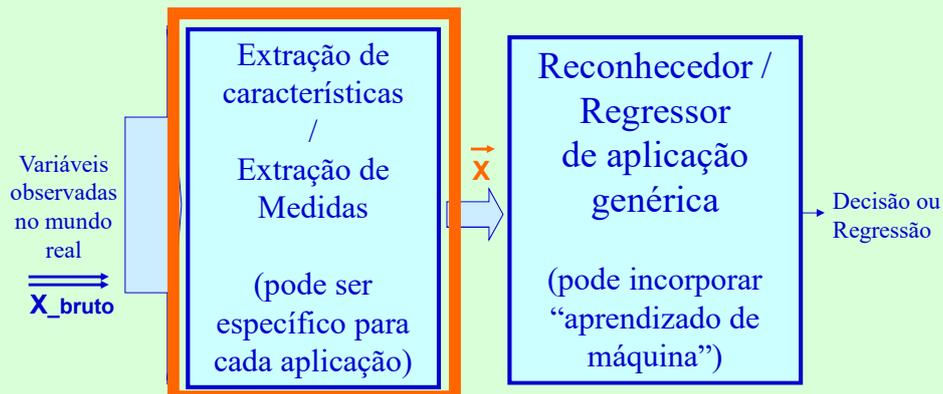
Mais detalhes matemáticos e algumas referências em PCA podem ser encontradas nos materiais gerados pelo Dr. Edson Kitani em complemento a estes slides aqui destacados:

http://www.lsi.usp.br/icone/psi2672/2011/monitor/PCA_Junho_2011_ECK_EdsonKitani_PSI2672.pdf

*Uma pincelada rápida em uma técnica **neural** recente de redução de dimensionalidade do vetor de entradas X e de extração de características sendo bastante estudada atualmente no contexto de Deep Learning: Autoencoders (auto-codificadores) e Stacked Autoencoders (vários auto-codificadores encadeados) –*

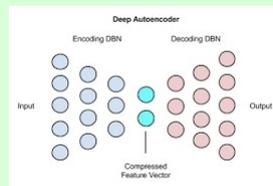
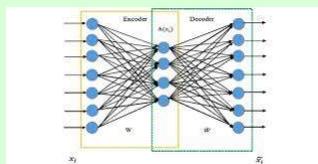
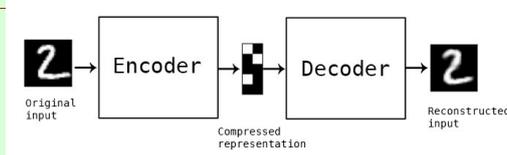
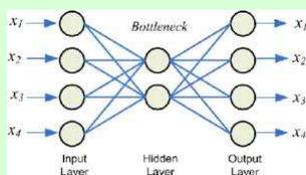
© Prof. Emilio Del Moral – EPUSP

... O 1o estágio gera um Vetor de Medidas, \vec{X}
(o segundo estágio operará sobre tal vetor)



© Prof. Emilio Del Moral Hernandez

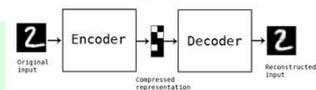
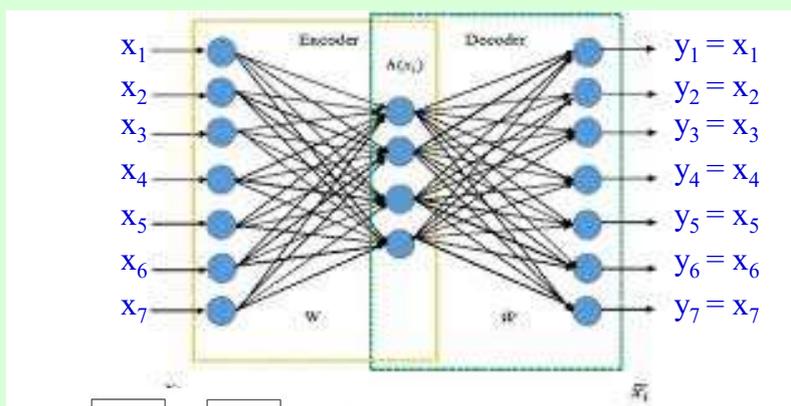
Autoencoders e Stacked Auto-encoders (imagens da internet)



Um pouquinho de Deep Learning

© Prof. Emilio Del Moral Hernandez

Um autoencoder detalhado (imagem da internet, adaptada)

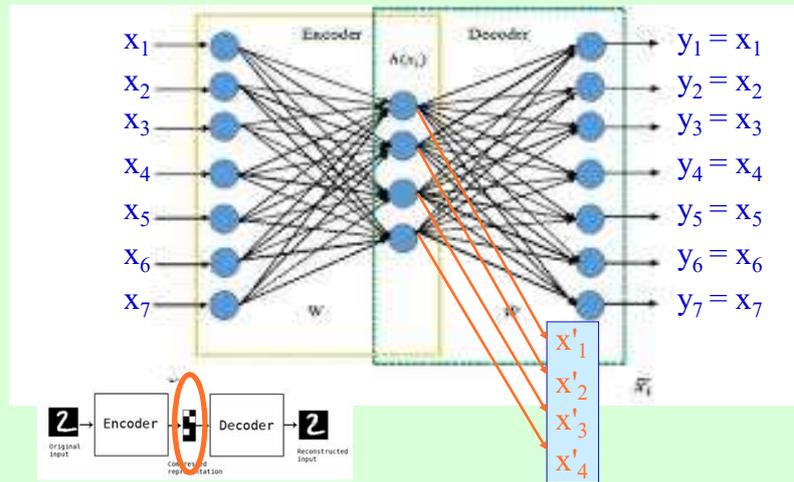


Um pouquinho de Deep Learning

© Prof. Emilio Del Moral Hernandez

Um autoencoder detalhado (imagem da internet, adaptada)

72



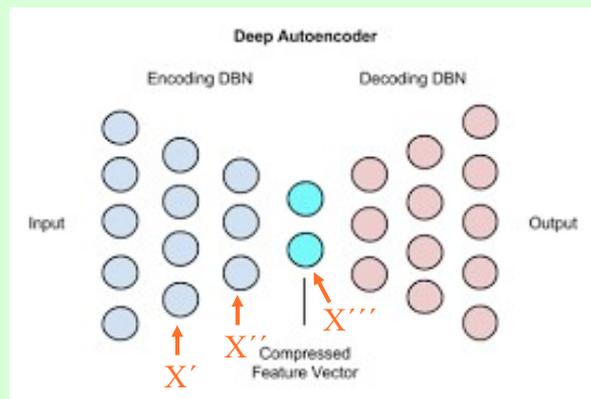
Um pouquinho de Deep Learning

© Prof. Emilio Del Moral Hernandez

72

Autoencoders e Stacked Auto-encoders (imagens da internet)

73



Um pouquinho de Deep Learning

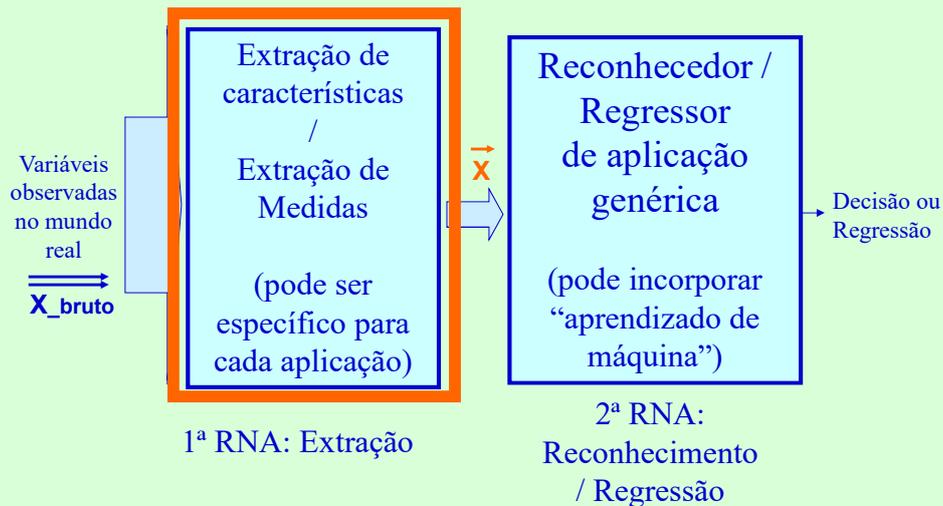
© Prof. Emilio Del Moral Hernandez

73

Note que nesta técnica, o primeiro estágio da “solução em dois estágios” também é uma rede neural (não só o segundo estágio é uma RNA), mas esse primeiro estágio é uma RNA específica, desenhada apenas para a codificação compacta de variáveis; ela não realiza a regressão ou o reconhecimento, que são feitos pela segunda rede neural.

© Prof. Emilio Del Moral – EPUSP

... O 1o estágio gera um Vetor de Medidas, \vec{X}
(o segundo estágio operará sobre tal vetor)



© Prof. Emilio Del Moral Hernandez

Ciclos de refinamento num projeto de modelagem para reconhecimento e/ou regressão, com duração de médio / longo prazo e com interesse de otimização de desempenho máxima

© Prof. Emilio Del Moral – EPUSP

... Aspectos de refinamentos nos projetos ...

Diferentes aspectos estudados no treinamento / otimização / caracterização da RNA ...

- Vários Delta W sequenciados (gradiente descendente)
- Re-sorteios de pesos iniciais (fugindo de mínimos locais)
- k-fold cross validation (avaliando sensibilidade aos dados empíricos)
- Diferentes graus de complexidade do modelo neural (evitando sobreaprendizado)
- Ensaios com vários Pré-Processamentos alternativos (aumentando desempenho)
- Descarte de algumas variáveis de menor relevância, p/ melhora do desempenho
- Aumento de $M+M'c$ / as mesmas variáveis (nesse aumento pode haver custo com novas coletas X;y)
- Acréscimo de variáveis x incluídas no modelo (há custo extra com novas coletas)
- ... Outros ...

139

© Prof. Emilio Del Moral – EPUSP

Resumindo, vários dos tópicos vistos nas diversas aulas podem ser aplicados no refinamento de um protótipo regressor / reconhecedor, num processo de vários ciclos de ensaios numéricos com redes neurais distintas / com dados de treino e teste distintos / com extrações de características distintas / etc, explorando várias alternativas e escolhendo as melhores soluções ...

... e note que, diferentemente do ambiente de prototipação interativa simples disponível no MBP, em ambientes com possibilidade de programação, como o Matlab, o Tensor Flow e outros, esta exploração ampla de várias configurações de modelos e ensaios computacionais pode ser parcialmente automatizada, através de loops aninhados que controlam a realização dos vários ensaios com características distintas das RNAs e dos dados.

© Prof. Emilio Del Moral – EPUSP

O Ciclo completo da modelagem:

0) **Formalização do problema, mapeamento quantitativo em um modelo neural inicial e ... 0b) coleta de pares empíricos (X,y)**

1) Fase de TREINO da RNA (MLP): com conhecimento dos X e dos y, que são ambos usados na calibração do modelo

2) Fase de TESTE / Caracterização da qualidade da RNA para generalizar: temos novos pares X e y, com y guardado “na gaveta”, usado apenas para avaliação, não para re-calibração. É como um ensaio de uso final do modelo, com possibilidade de medir a sua qualidade com o y que foi guardado na gaveta.

[Fase de refinamentos sucessivos da RNA e/ou dos dados e/ou do modelo, em ciclos diversos, recomeçando desde o passo 0 ou do passo 1]

3) Fase de USO FINAL da RNA, com y efetivamente não conhecido, e estimado com conhecimento dos X + uso do modelo calibrado.

.... **Diferenças e semelhanças entre 1, 2 e 3**

141

© Prof. Emilio Del Moral – EPUSP

*E sempre devemos usar
redes neurais para
implementar regressões e
reconhecimento de
padrões??*

© Prof. Emilio Del Moral – EPUSP

**A áreas mais nobres de uso dos nossos MLPs –
... Mapa em construção, junto com vocês**

*A rigor não
precisariamos de um
“full MLP” nestes
casos*

*O MLP é muito útil e um
diferencial importante na
modelagem*

x univariado linear

x é multivariado (X) mas
temos fenômenos $y(X)$
basicamente lineares

X e y Binários
categóricos sem
ambiguidades de $y(X)$

x univariado
mas não linear

X e y Binários categóricos com
algumas ambiguidades em $y(X)$

X multivariado não
linear e com não
linearidade/s
desconhecidas

143

© Prof. Emilio Del Moral – EPUSP