Tema 2

Extração de Características

Redução de Dimensionalidade

PCA (Principal Components Analysis)

Professora:
Ariane Machado Lima

SISTEMAS DE INFORMAÇÃO

EACH

Extração de Características

- Questões envolvidas:
  1. Como organizar as características
  2. Como visualizar as características
  3. Como escolher as características
  4. Como medir as características
     - Depende da natureza do objeto (imagem, som, texto, moléculas, clientes, pacientes, etc.)
     - Não será o foco nesta disciplina

Extração de características – como organizá-las

- Tabela
  - Objetos nas linhas
  - Características nas colunas

Object	Feature 1 (area in cm^2)	Feature 2 (volume in cm^3)
1	32.67	68.48
2	28.30	63.91
3	24.99	71.95
4	26.07	59.36
5	31.92	70.33
6	31.32	68.40
7	25.14	81.00

Extração de características – como organizá-las

- Problema?
  - Para problemas de verdade precisamos que o computador lide com os dados
  - Qual formato seria mais adequado?

Extração de características – como organizá-las

- Problema?
  - Para problemas de verdade precisamos que o computador lide com os dados
  - Qual formato seria mais adequado?
- Matrizes!

$$F = \begin{bmatrix} 32.67 & 68.48 \\ 28.30 & 63.91 \\ 24.99 & 71.95 \\ 26.07 & 59.36 \\ 31.92 & 70.33 \\ 31.32 & 68.40 \\ 25.14 & 81.00 \end{bmatrix} = \begin{bmatrix} \leftarrow & \vec{f}_1^T & \rightarrow \\ \leftarrow & \vec{f}_2^T & \rightarrow \\ \leftarrow & \vec{f}_3^T & \rightarrow \\ \leftarrow & \vec{f}_4^T & \rightarrow \\ \leftarrow & \vec{f}_5^T & \rightarrow \\ \leftarrow & \vec{f}_6^T & \rightarrow \\ \leftarrow & \vec{f}_7^T & \rightarrow \end{bmatrix}$$

Extração de características – como organizá-las

- Cada linha i da matriz corresponde ao vetor de características transposto do objeto i

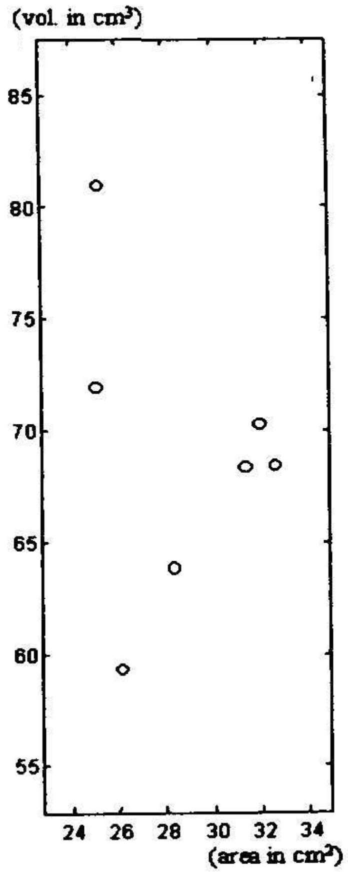
$$F = \begin{bmatrix} 32.67 & 68.48 \\ 28.30 & 63.91 \\ 24.99 & 71.95 \\ 26.07 & 59.36 \\ 31.92 & 70.33 \\ 31.32 & 68.40 \\ 25.14 & 81.00 \end{bmatrix} = \begin{bmatrix} \leftarrow & \vec{f}_1^{\,T} & \rightarrow \\ \leftarrow & \vec{f}_2^{\,T} & \rightarrow \\ \leftarrow & \vec{f}_3^{\,T} & \rightarrow \\ \leftarrow & \vec{f}_4^{\,T} & \rightarrow \\ \leftarrow & \vec{f}_5^{\,T} & \rightarrow \\ \leftarrow & \vec{f}_6^{\,T} & \rightarrow \\ \leftarrow & \vec{f}_7^{\,T} & \rightarrow \end{bmatrix}$$

Extração de características – como visualizá-las

- Sugestões?

Extração de características – como visualizá-las

- Usar as características como eixos de um gráfico, e os objetos como "pontos"



Scatterplot
ou
Gráfico de Dispersão

SISTEMAS DE INFORMAÇÃO

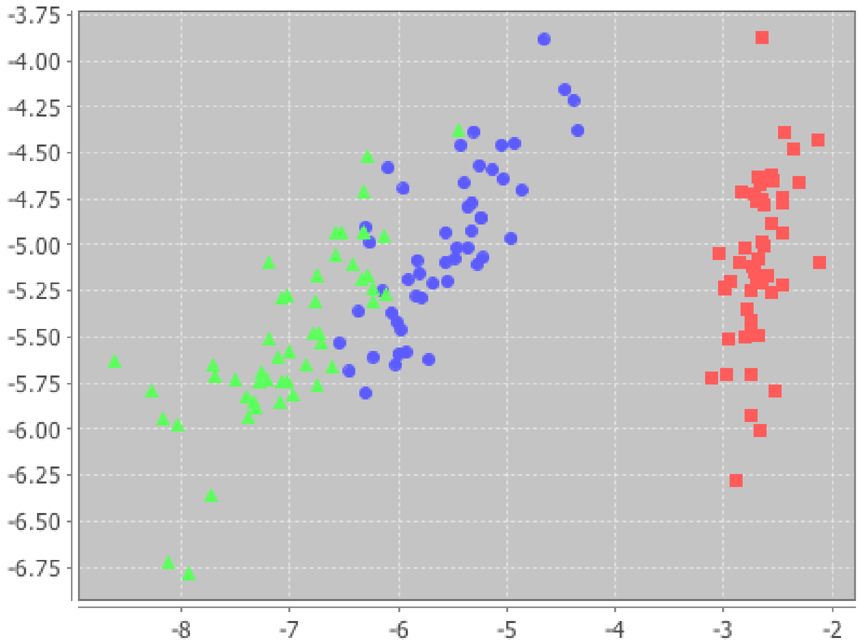
EACH

Extração de características – como visualizá-las

- E para representar objetos de classes distintas?
-

Extração de características – como visualizá-las

- E para representar objetos de classes distintas?
- Formas e / ou cores diferentes para classes diferentes

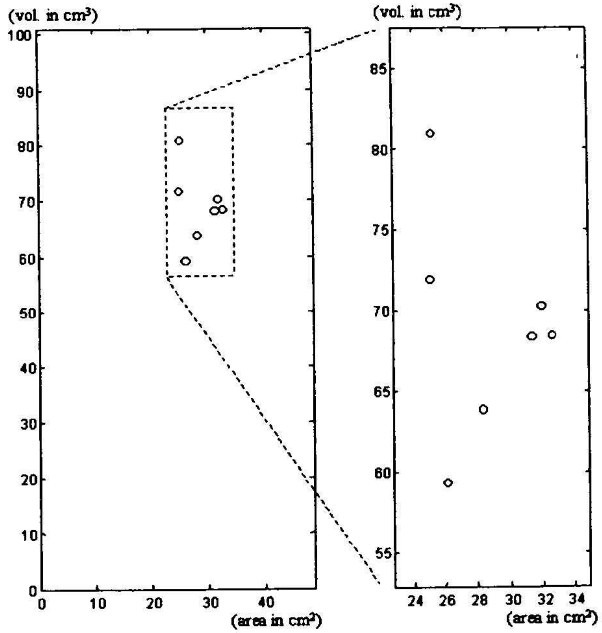


SISTEMAS DE INFORMAÇÃO

EACH

Extração de características – como visualizá-las

Visualização absoluta Visualização relativa



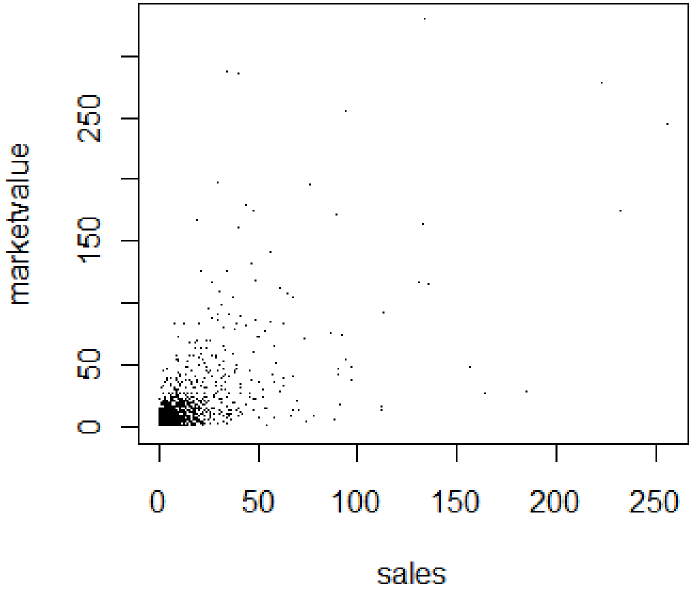
Mesma escala para evitar distorções entre as distâncias

SISTEMAS DE INFORMAÇÃO

EACH

Extração de características – como visualizá-las

Às vezes pode ser necessário mudar a escala de alguma(s) característica(s)

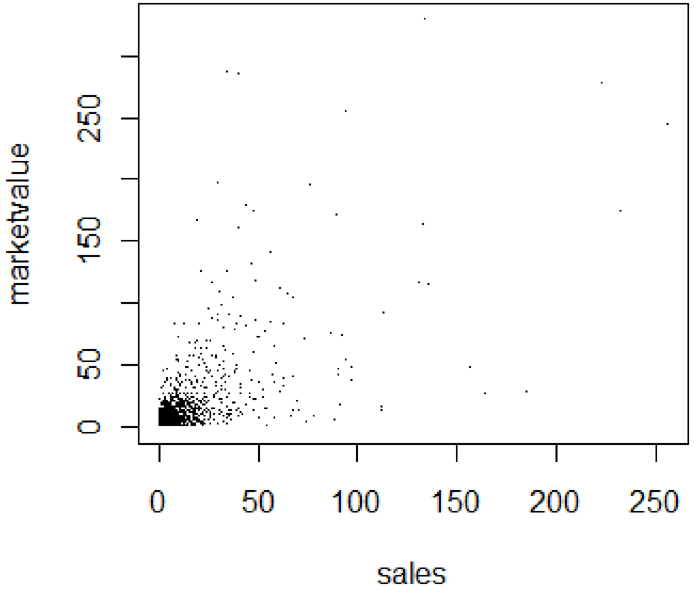


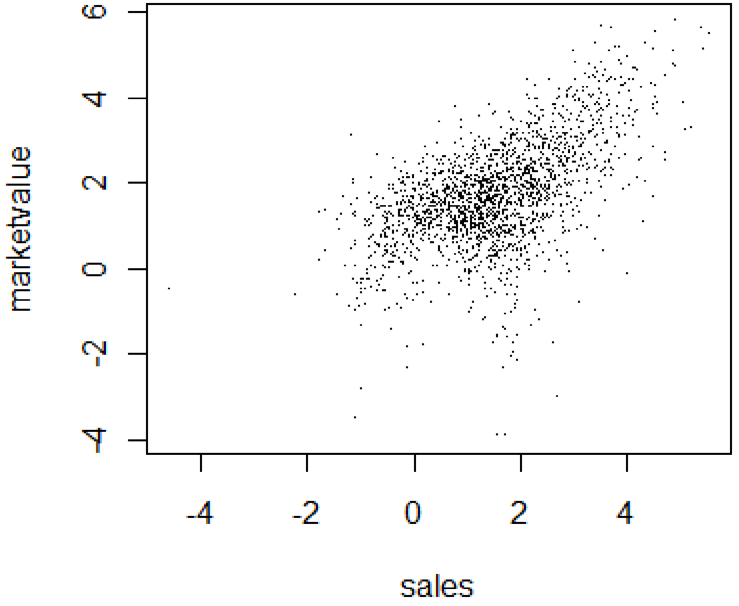
SISTEMAS DE INFORMAÇÃO

EACH

Extração de características – como visualizá-las

Às vezes pode ser necessário mudar a escala de alguma(s) característica(s) (por ex: log)



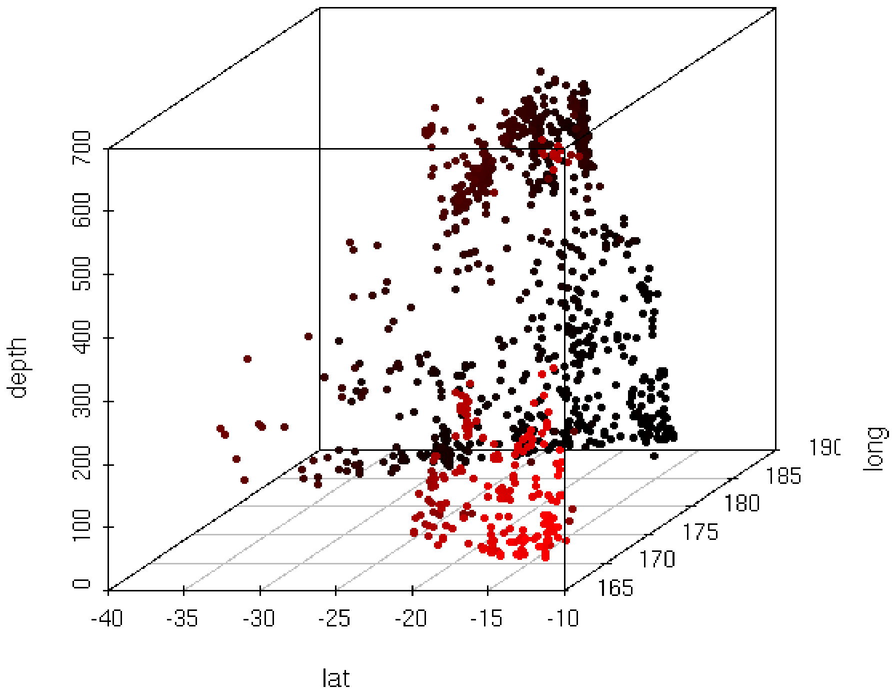


SISTEMAS DE INFORMAÇÃO

EACH

Extração de características – como visualizá-las

- 3 características



Extração de características – como visualizá-las

- Problema?
- O que fazer quando um objeto tem 4 ou mais características?

SISTEMAS DE INFORMAÇÃO

EACH

Extração de características – como visualizá-las

- Problema?
- O que fazer quando um objeto tem 4 ou mais características?
  - Não dá mais para visualizar tudo
  - Visualizar só um subconjunto
  - Aplicação de técnicas para projeção em espaço de dimensão 1, 2 ou 3

Extração de características – como escolhê-las

- ?

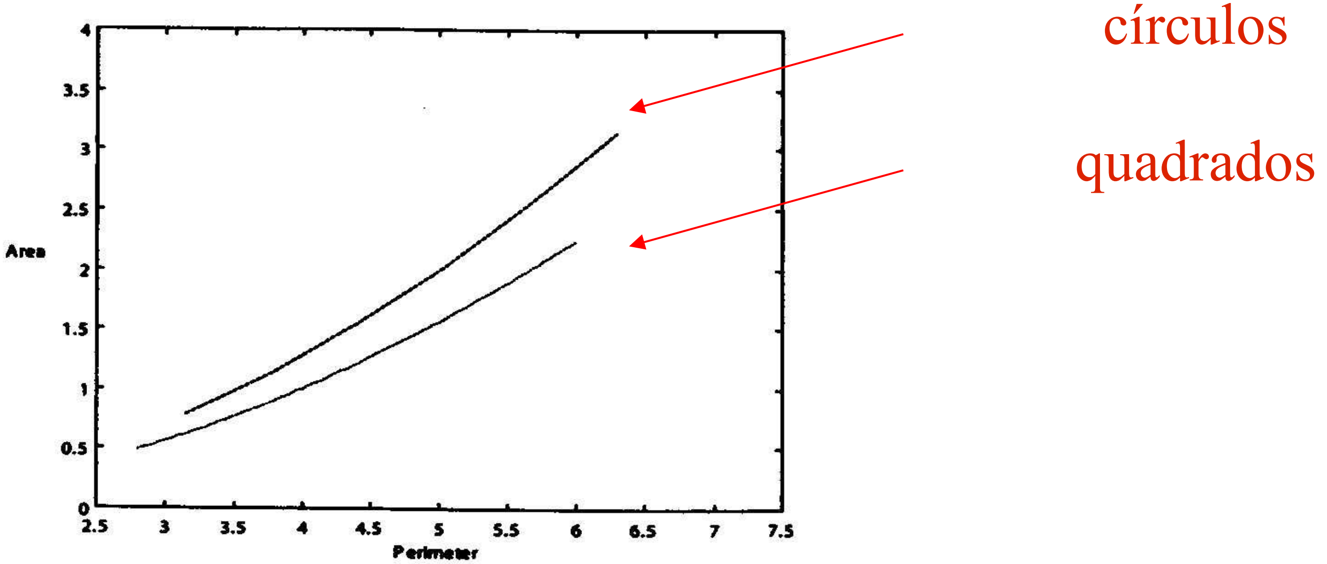
Extração de características – como escolhê-las

- Exemplo: discriminar dois tipos de queijo: quadrados e circulares (vários tamanhos)
  - Área e perímetro

Extração de características – como escolhê-las

Exemplo: discriminar dois tipos de queijo: quadrados e circulares (vários tamanhos)

- Área e perímetro

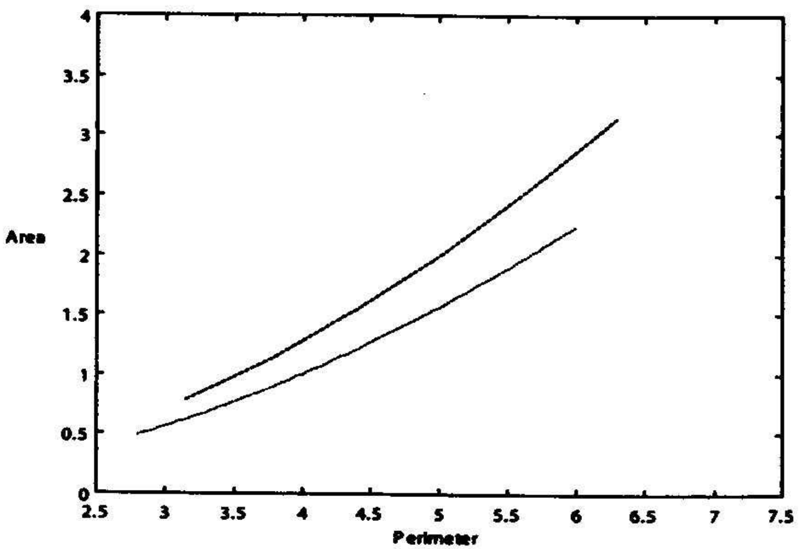


Como deveria ser...

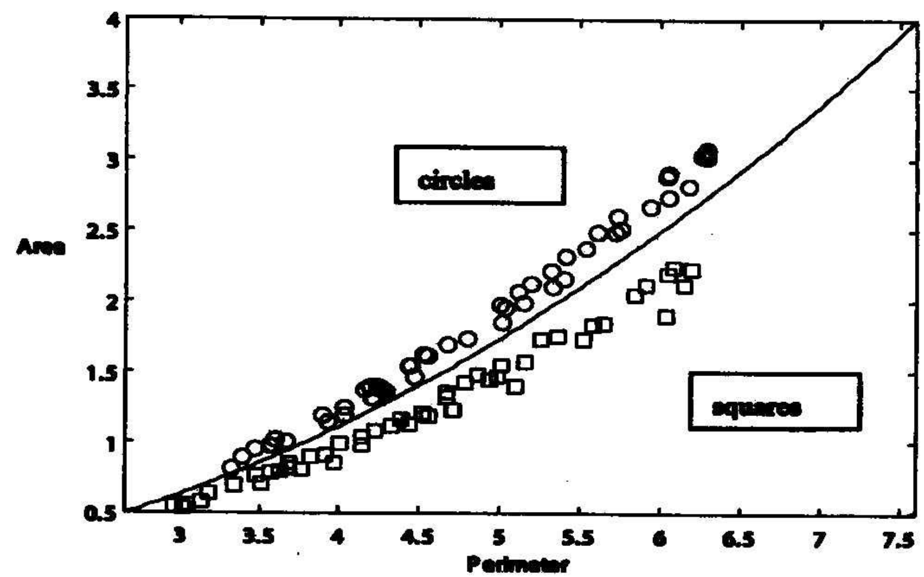
Extração de características – como escolhê-las

Exemplo: discriminar dois tipos de queijo: quadrados e circulares (vários tamanhos)

- Área e perímetro



Como deveria ser...



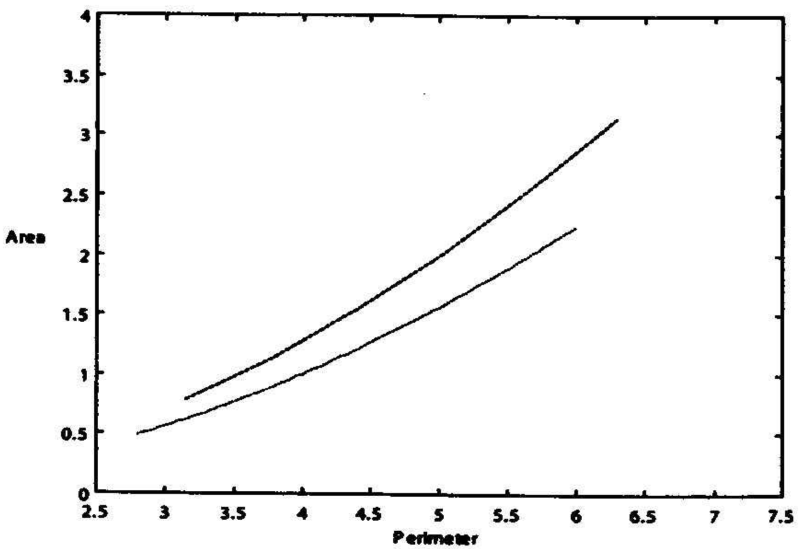
Mundo real

Extração de características – como escolhê-las

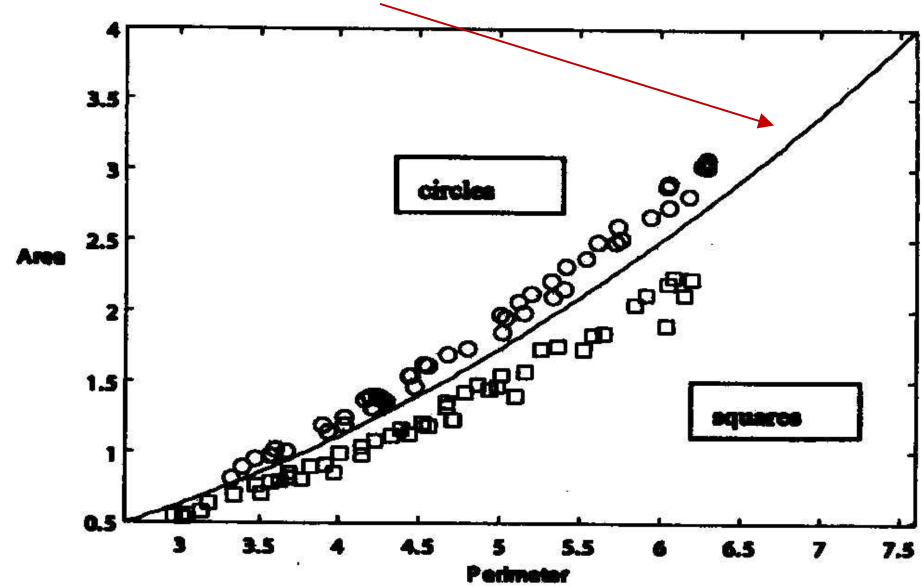
Exemplo: discriminar dois tipos de queijo: quadrados e circulares (vários tamanhos)

- Área e perímetro

Fronteira de Decisão



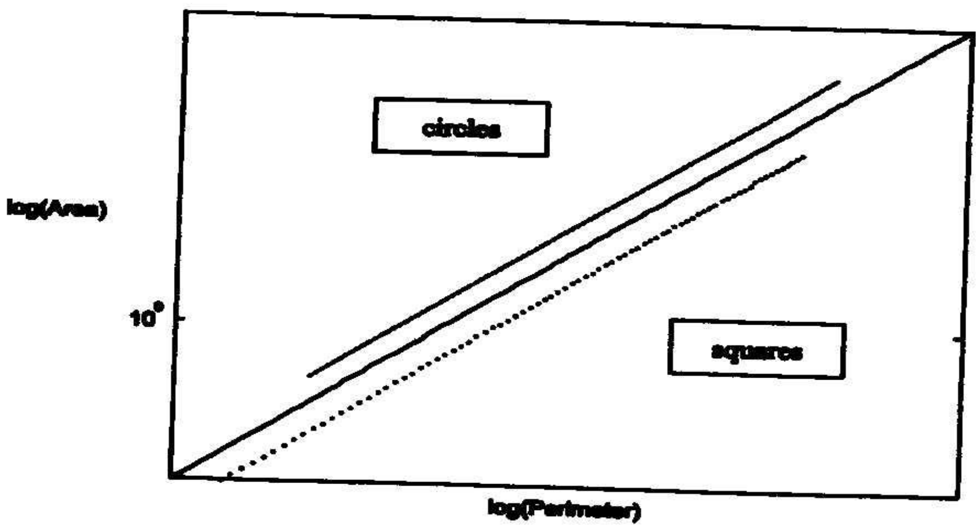
Como deveria ser...



Mundo real bom

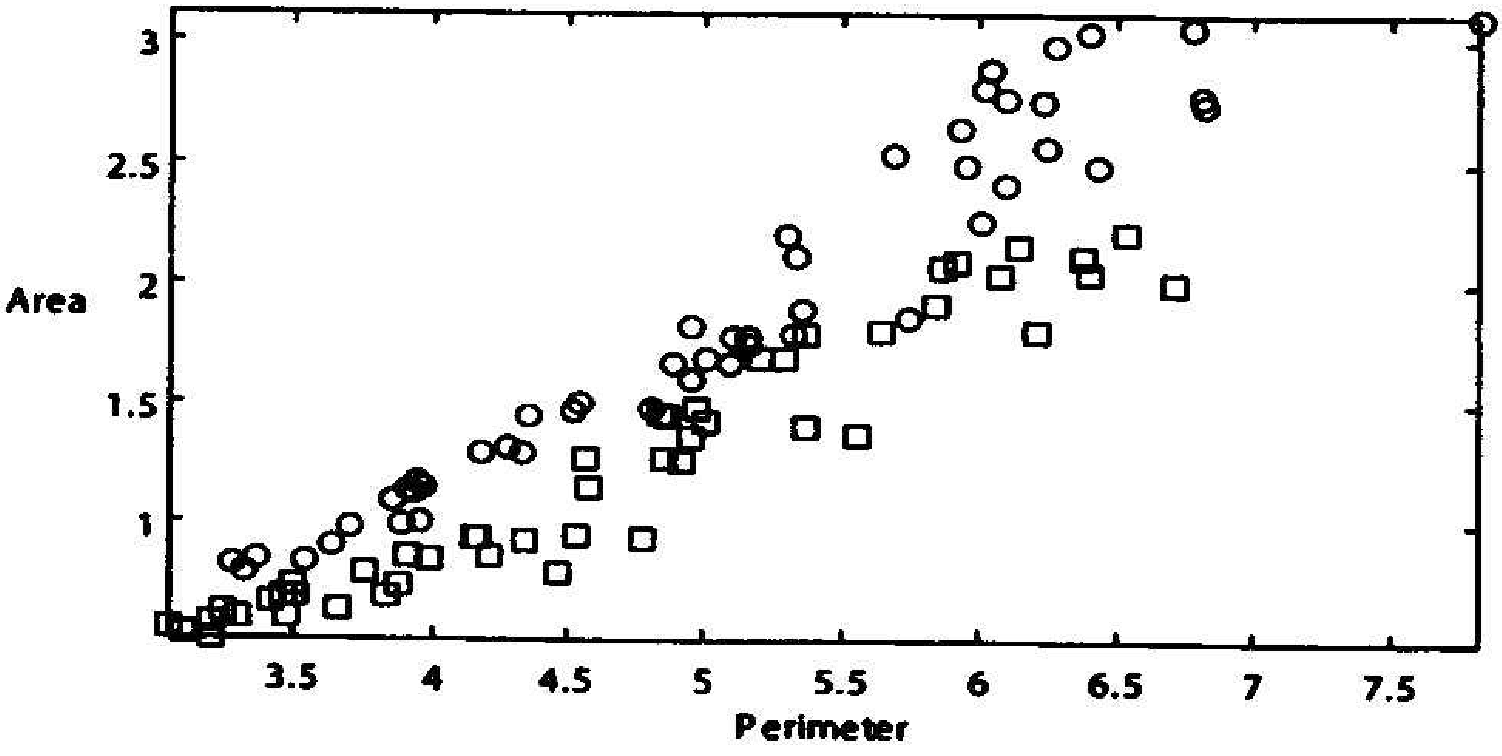
Extração de características – como escolhê-las

Transformação de variáveis pode simplificar o classificador



Neste exemplo, transformação logaritmica: uma reta como separador ao invés de uma parábola

Extração de características – como escolhê-las



Lembrando que o mundo real pode não ser tão bom...

SISTEMAS DE INFORMAÇÃO

EACH

Extração de características – como escolhê-las

- Usar um número menor de características, combinando algumas, pode fornecer resultados similares ou melhores
- *Thinness ratio*: T = Área / Perímetro^2 (vantagem de ser adimensional)
- Mundo perfeito:
  - R(círculo): $\pi r^2/(2\pi r)^2 = 1/(4\pi)$
  - R(quadrado): $l^2/(4l)^2 = 1/16$

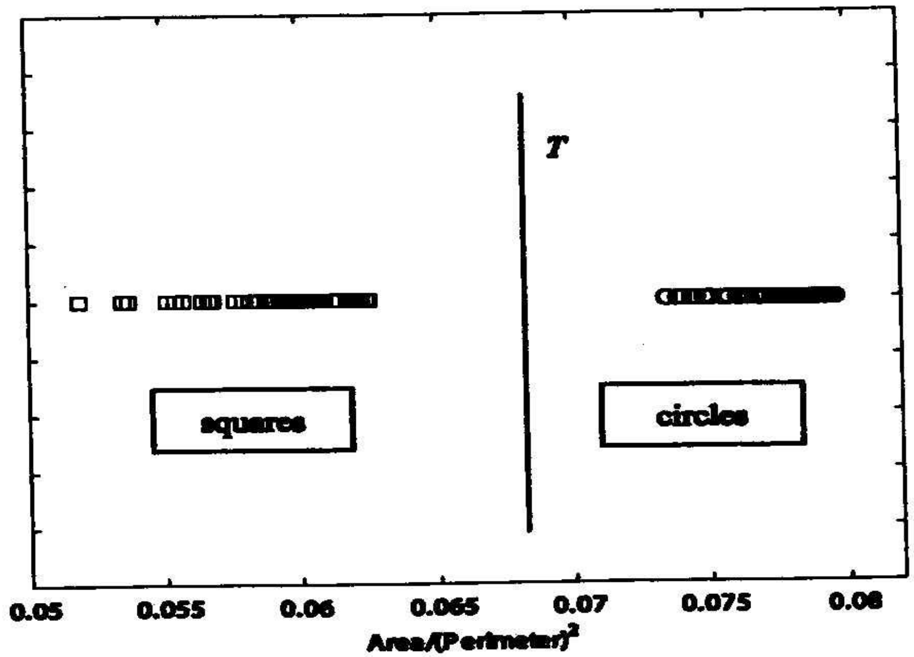
SISTEMAS DE INFORMAÇÃO

EACH

Extração de características – como escolhê-las

- *Thinness ratio*: Área / Perímetro^2 (adimensional)

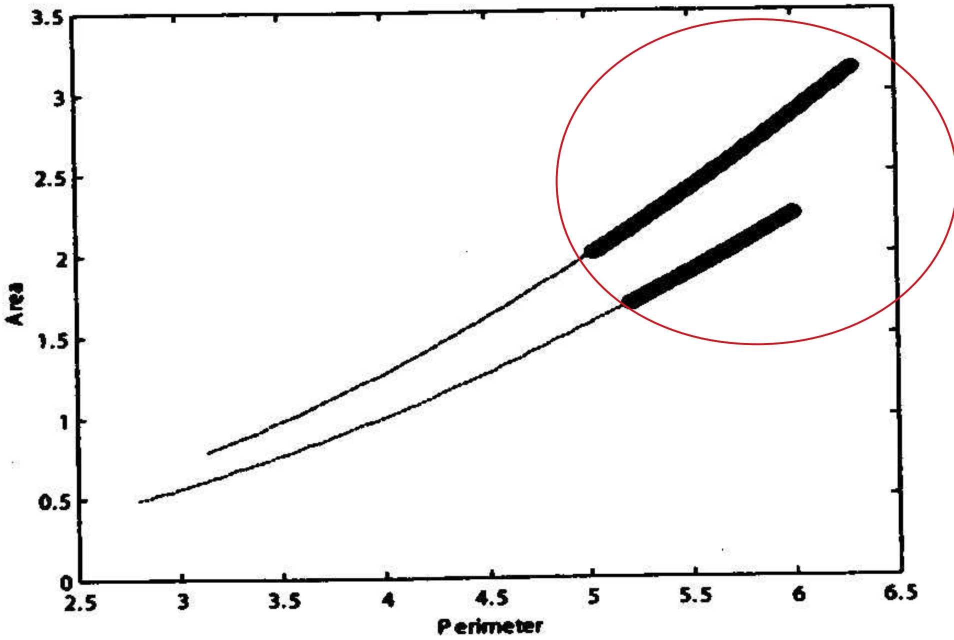
Mundo real: threshold: algo entre 1/(4π) e 1/16



Extração de características – como escolhê-las

- Complicando um pouco:
- Em épocas especiais há a produção de queijos especiais, circulares E quadrado
  - Circular: $0.8 < r \leq 1$
  - Quadrado: $1.3 < l \leq 1.5$

Extração de características – como escolhê-las



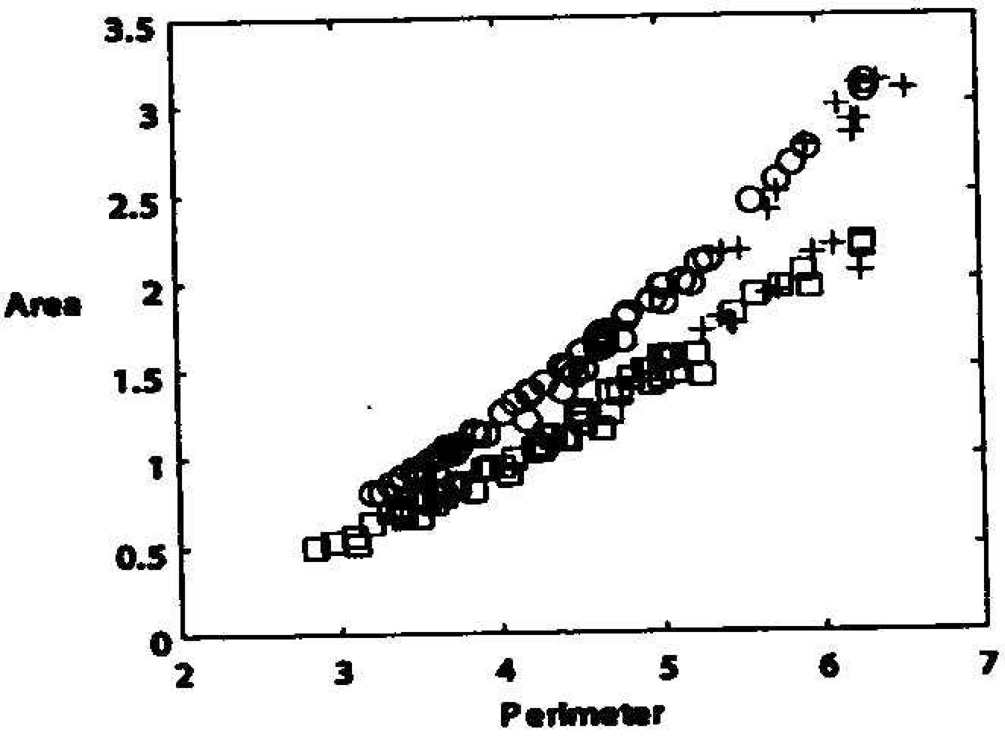
Queijos especiais

Mundo "ideal",
e mesmo assim problemático:

Regiões sobrepostas

Regiões descontínuas

Extração de características – como escolhê-las



Mundo real

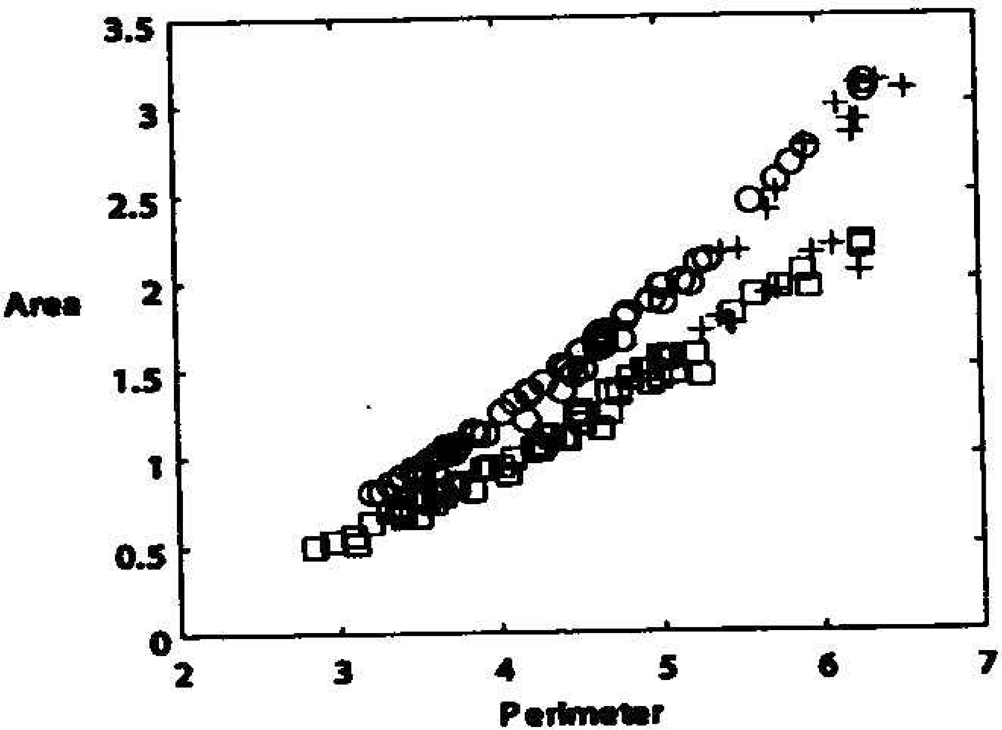
Regiões sobrepostas

Regiões descontínuas

SISTEMAS DE INFORMAÇÃO

EACH

Extração de características – como escolhê-las



Mundo real

Regiões sobrepostas

Regiões descontínuas

Como resolver?

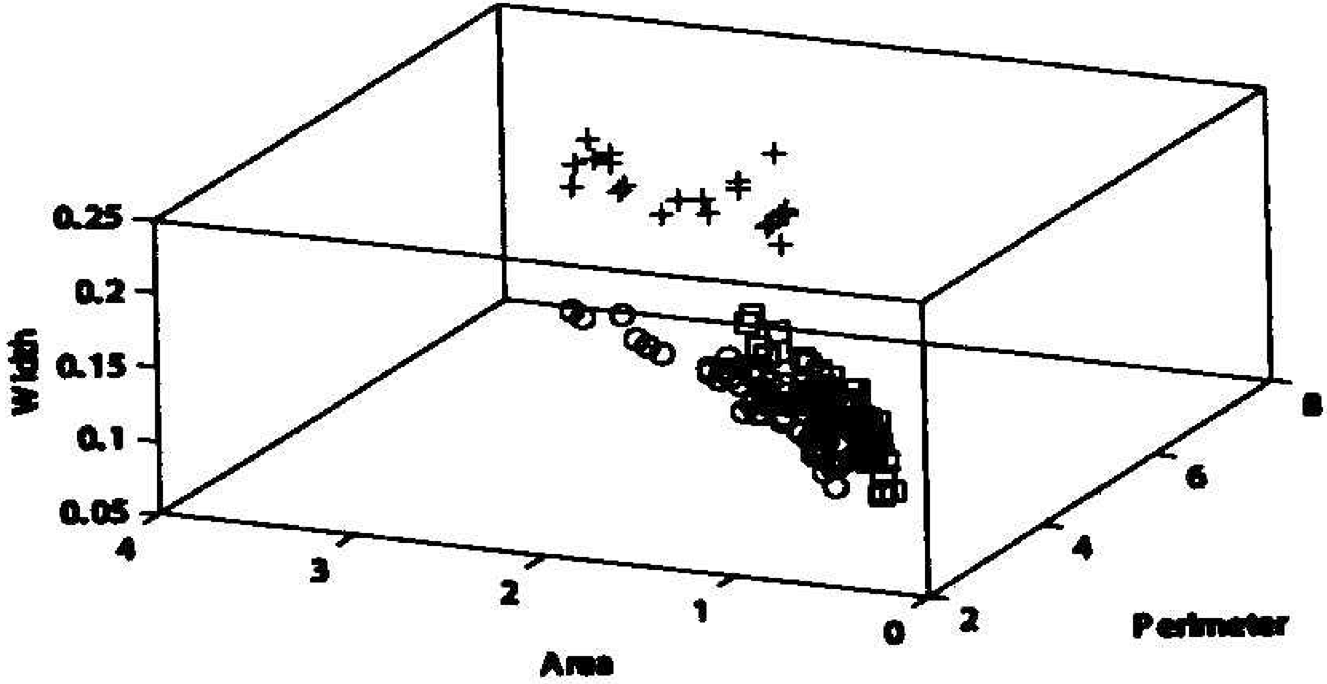
Extração de características – como escolhê-las

- Há mais características que podem ser consideradas?

Extração de características – como escolhê-las

- Se as larguras forem diferentes...

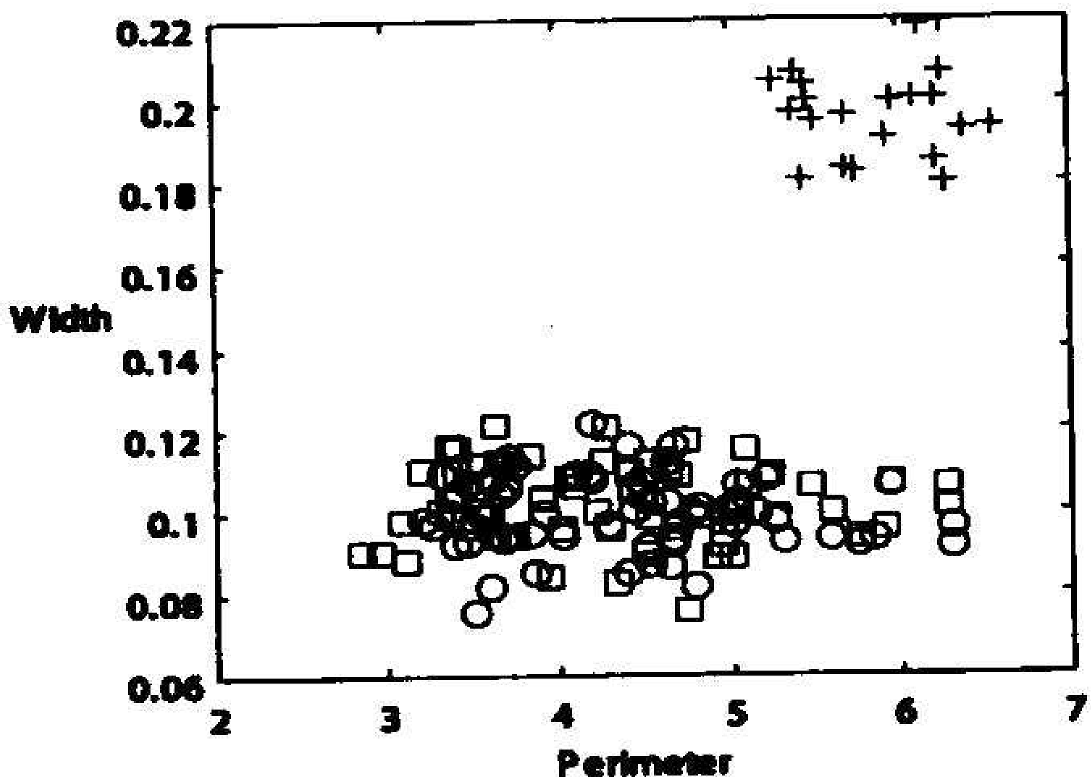
3 D



Extração de características – como escolhê-las

Se as larguras forem diferentes...

Projeção 2 D



Extração de características – como escolhê-las

E se as larguras não fossem diferentes?

Extração de características – como escolhê-las

- E se as larguras não fossem diferentes?
- É preciso descobrir características que separem as classes...
  - Talvez a data de fabricação?
- Classificar ~ descobrir (ou conhecer mais a respeito) o processo de geração dos objetos
- Modelo de cada classe
- Classificar não é fácil...

Extração de características – como escolhê-las

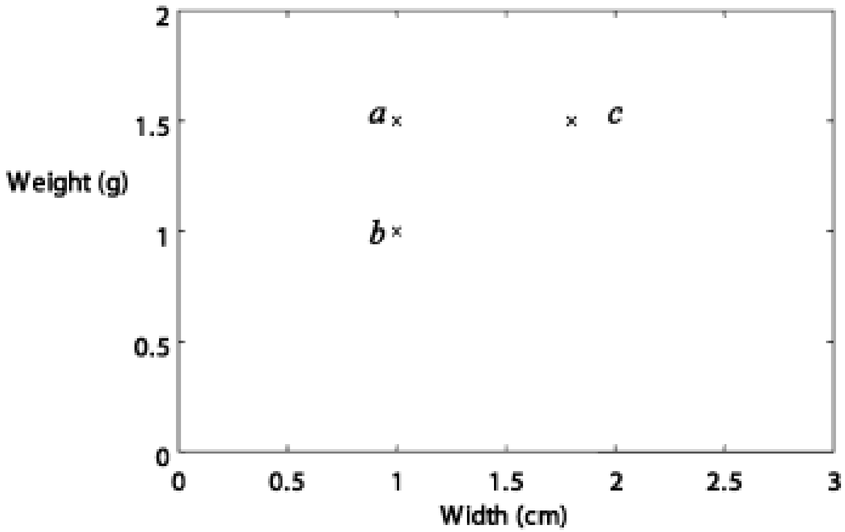
- Mapa da Mina?

Extração de características – como escolhê-las

- Mapa da Mina? Não existe, mas...
- Iremos ver (nas próximas aulas) algumas dicas e estratégias de escolha automática
- Antes, veremos um cuidado muitas vezes importante de ser tomado: normalização

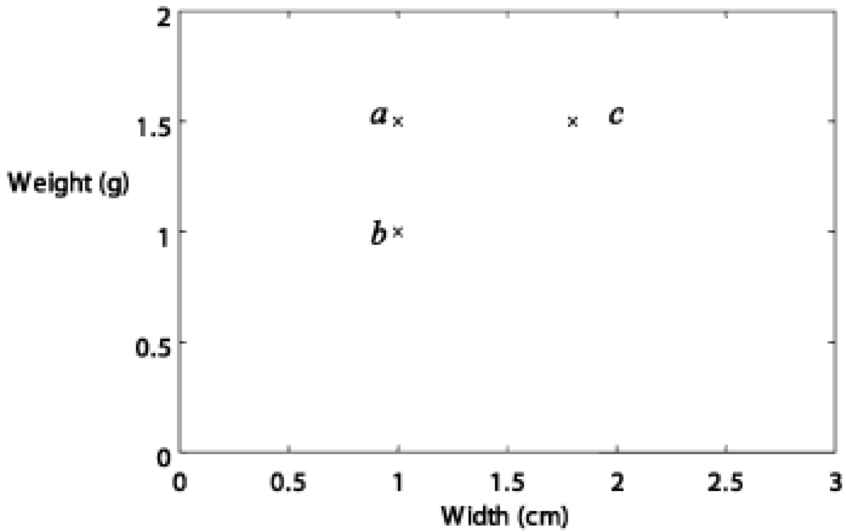
Normalização de características

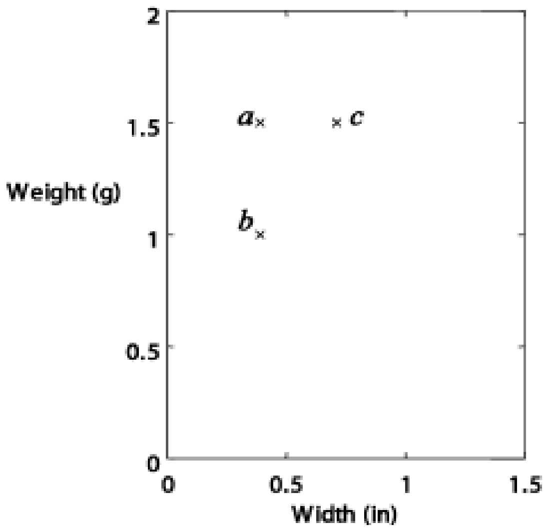
- A maioria das características são dimensionais (possuem unidade métrica)
- 3 objetos (por exemplo, folhas)



Normalização de características

- Unidades de medidas afetam as distâncias entre os objetos no espaço de características





Normalização de características

- Melhor utilizar características adimensionais
- Exemplos (o mais apropriado depende da característica e do problema):
  - Análise de Componentes Principais (novas variáveis não correlacionadas através da combinação das variáveis originais)
  - Utilizar razões entre características (ex: *thinness ratio*)
  - Normalização através da medida a partir de uma referência (ex: altura de um jogador / altura do jogador mais alto)
  - Proporção em relação à soma de todos os valores
  - Transformação normal (novas medidas com média = 0 e desvio padrão = 1

Normalização de características

- Transformação normal:
  - $x_i' = (x_i - \mu) / \sigma$ (z-score)
  - Assume-se uma distribuição normal sobre x

Normalização de características

- Transformação normal
  - $x_i' = (x_i - \mu) / \sigma$ (z-score)
  - Assume-se uma distribuição normal sobre x
  - Problema?

Normalização de características

- Transformação normal
  - $x_i' = (x_i - \mu) / \sigma$ (z-score)
  - Assume-se uma distribuição normal sobre x
  - E se não for?

Normalização de características

- Transformação normal
  - $x_i' = (x_i - \mu) / \sigma$ (z-score)
  - Assume-se uma distribuição normal sobre x
  - E se não for?
  - Checar antes (teste de normalidade)

Normalização de características

- Transformação normal
  - $x_i' = (x_i - \mu) / \sigma$ (z-score)
  - Assume-se uma distribuição normal sobre x
  - E se não for?
  - Checar antes (teste de normalidade)

Tests for Normality

ad.test	Anderson-Darling test for normality
cvm.test	Cramer-von Mises test for normality
lillie.test	Lilliefors (Kolmogorov-Smirnov) test for normality
pearson.test	Pearson chi-square test for normality
sf.test	Shapiro-Francia test for normality

Dimensionalidade

Quantas características utilizar?

Quanto mais melhor?

SISTEMAS DE INFORMAÇÃO

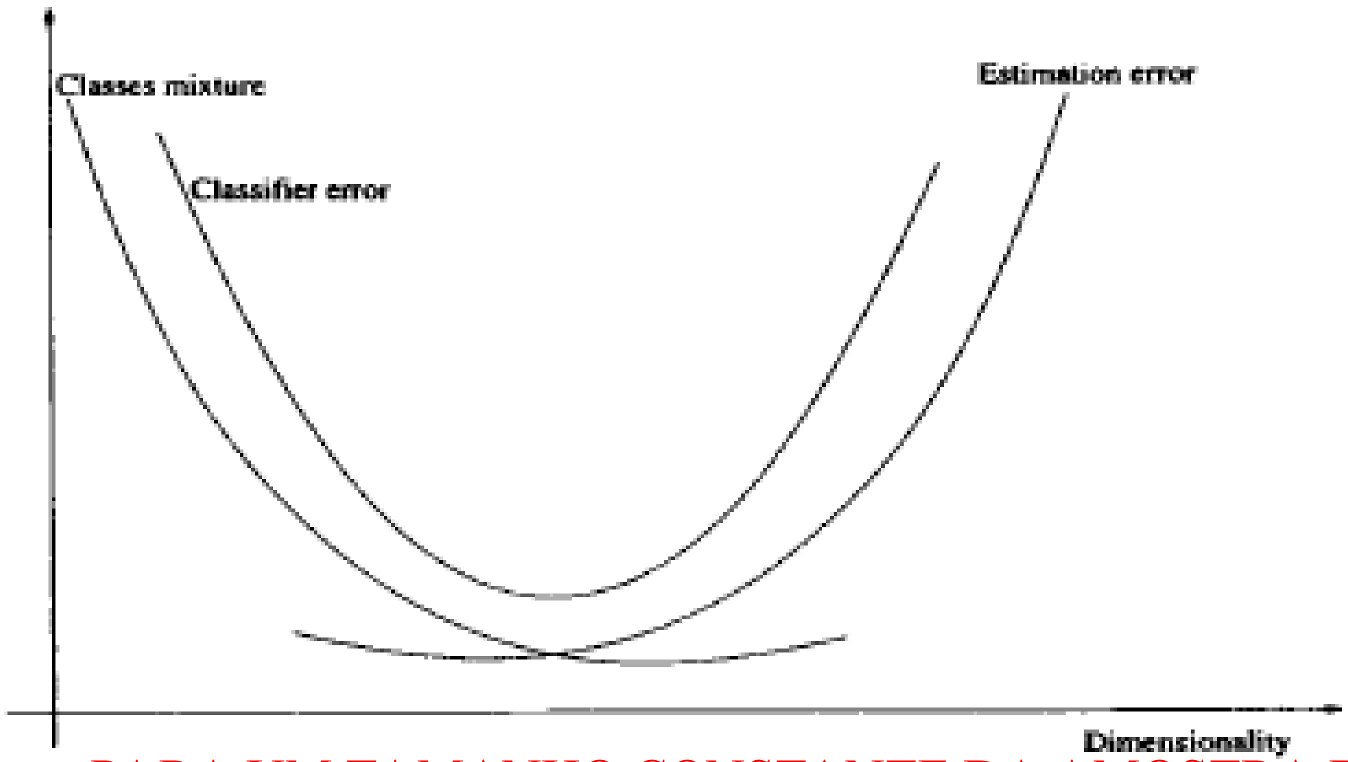
EACH

Dimensionalidade

MALDIÇÃO

Dimensionalidade

MALDIÇÃO DA DIMENSIONALIDADE



PARA UM TAMANHO CONSTANTE DA AMOSTRA DE TREINAMENTO!

SISTEMAS DE INFORMAÇÃO

EACH

Redução de Dimensionalidade

- Fusão de características

- Seleção de características

Redução de Dimensionalidade

- Fusão de características
  - PCA
  - ICA
  - Outras: análise de Fourier, wavelets e LDA (linear discriminant analysis)

- Seleção de características

Referências dessa primeira parte

- DUDA, R.; HART, P.; STORK, D**. Pattern Classification**. John Willey, 2001 (Cap. 1)

- COSTA, L. F.; CESAR, R. M. Jr. **Shape Classification and Analysis: Theory and Practice**. CRC Press, 2009 (Cap. 8.1)

SISTEMAS DE INFORMAÇÃO

EACH

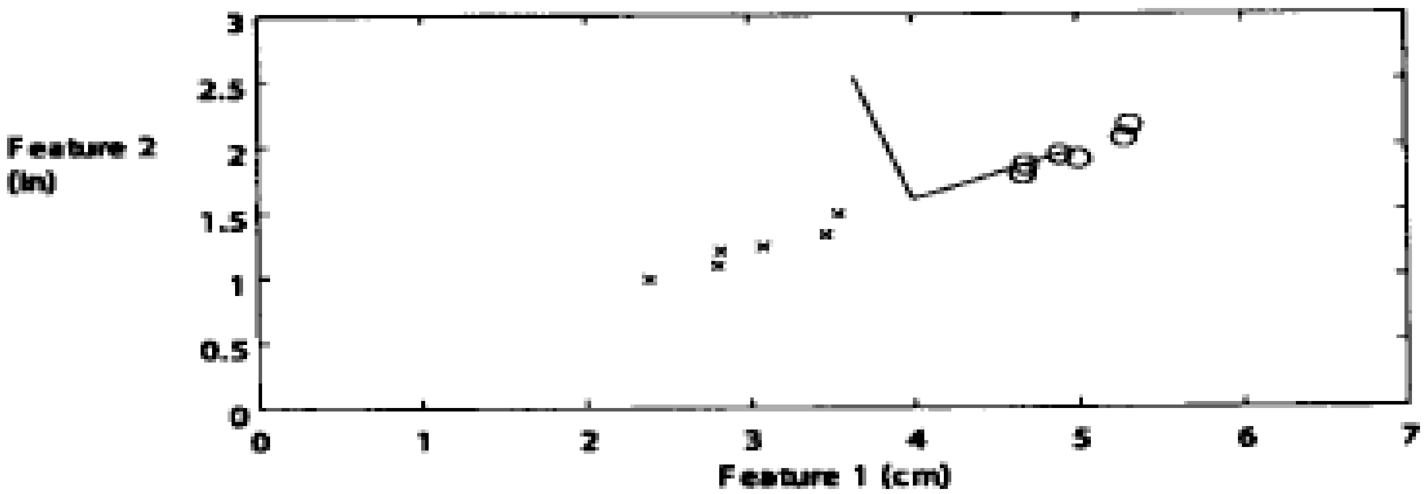
Análise de Componentes Principais

PCA – *Principal Components Analysis*

Normalização e redução de dimensionalidade (fusão de características)

Exemplo

- Duas classes de objetos
- Cada objeto é representado por 2 características
- O que se nota neste gráfico?
- Isso poderia nos ajudar a reduzir a dimensionalidade?



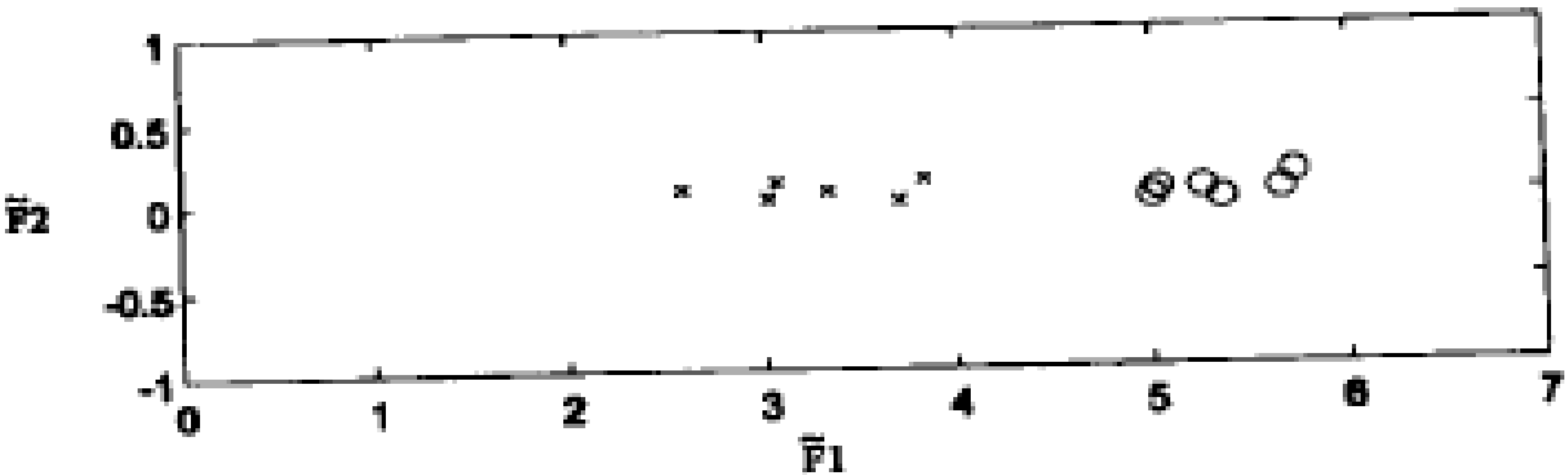
[COSTA& CESAR, 2009]

SISTEMAS DE INFORMAÇÃO

EACH

Exemplo

Transformação para 2 novas características



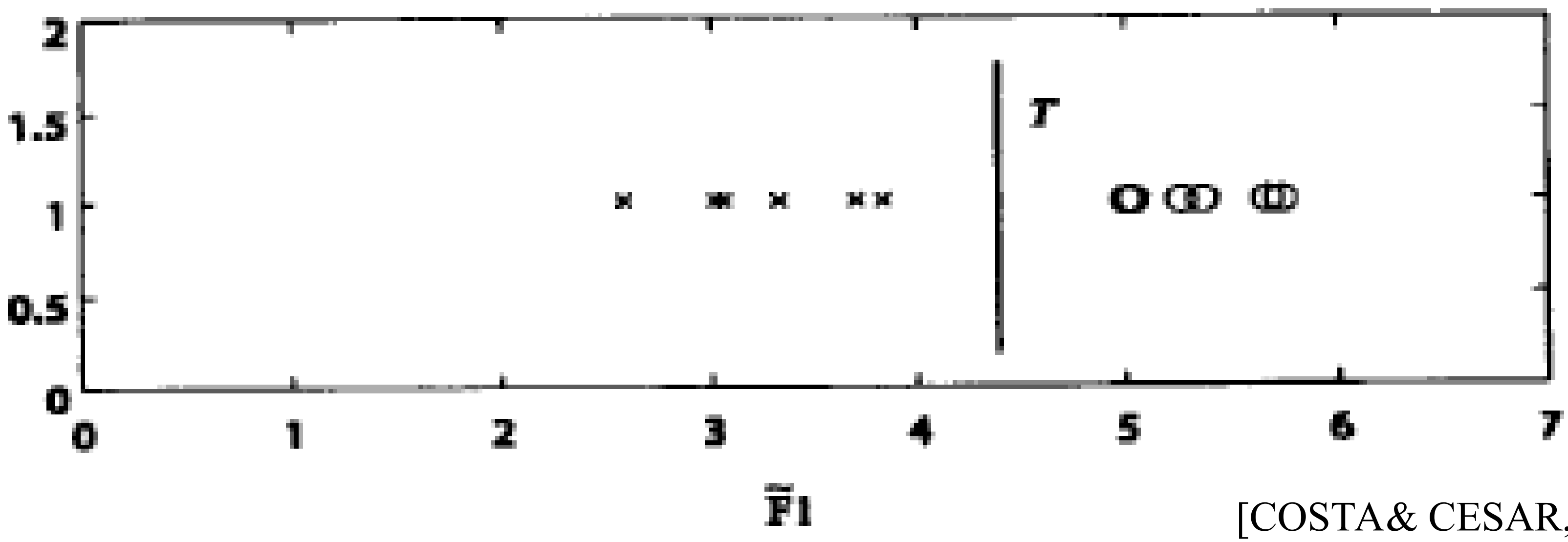
[COSTA& CESAR, 2009]

SISTEMAS DE INFORMAÇÃO

EACH

Exemplo

Transformação para apenas 1 nova característica (qual delas?)

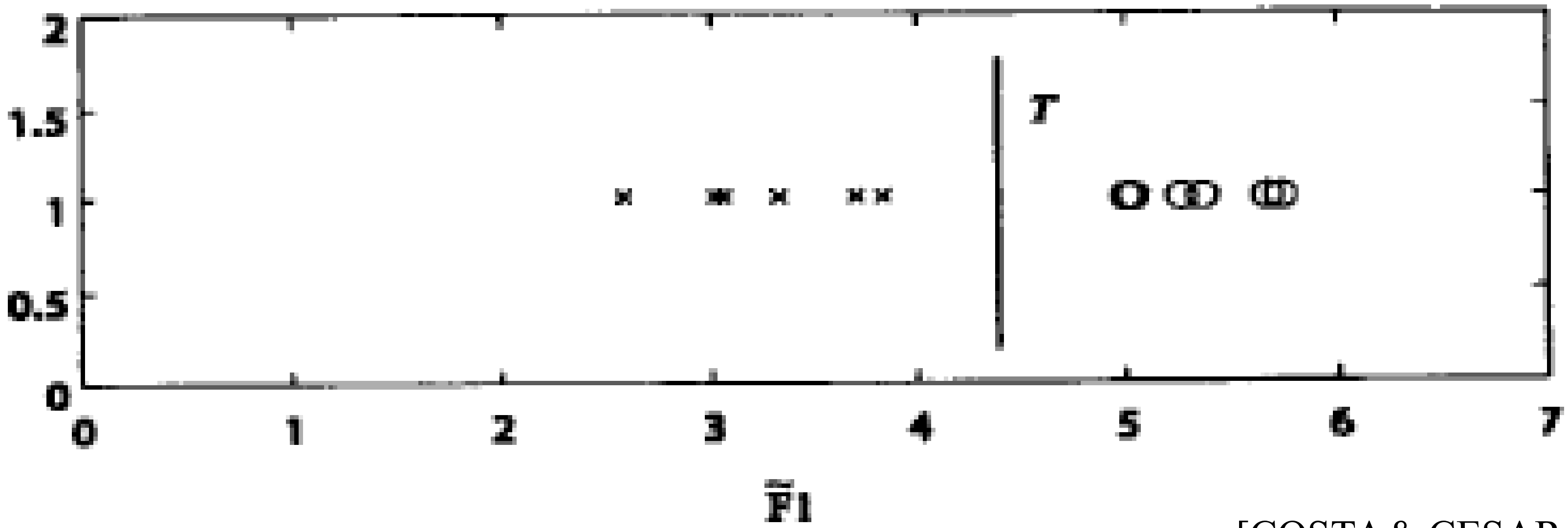


[COSTA& CESAR, 2009]

Perfeitamente separáveis por um limiar!

Exemplo

Transformação para apenas 1 nova característica (qual delas? A com maior dispersão)



[COSTA& CESAR, 2009]

Perfeitamente separáveis por um limiar!

SISTEMAS DE INFORMAÇÃO

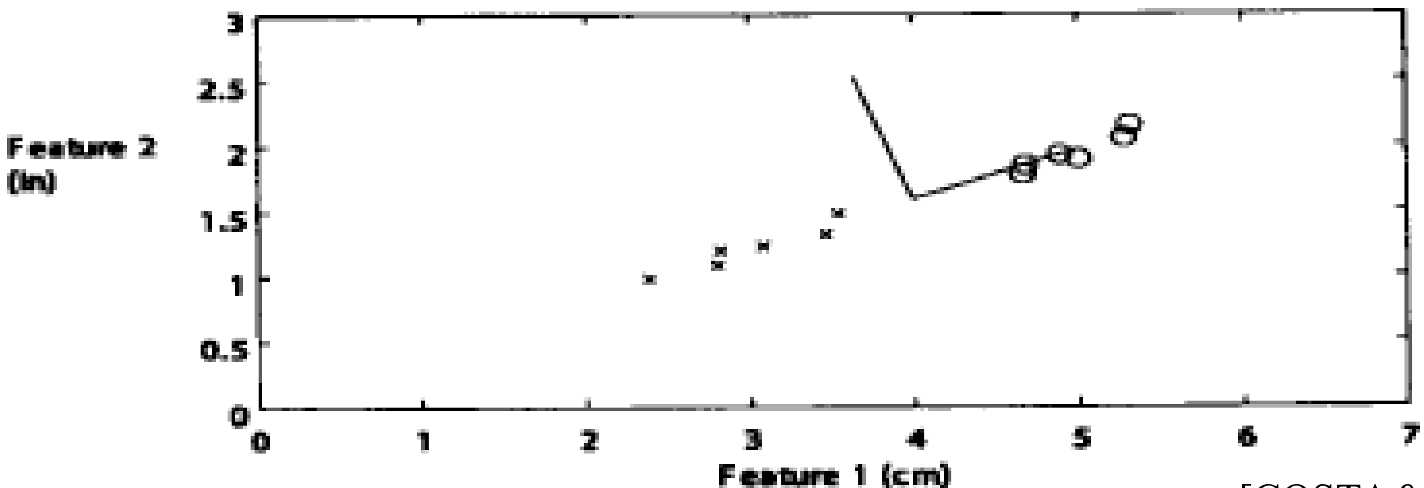
EACH

Análise de Componentes Principais

- PCA – *Principal Components Analysis*
- Assume que os dados originais estão representados por variáveis correlacionadas
- Objetivo: transformar essas variáveis em novas variáveis (mudança de base do espaço vetorial) que
  - não sejam correlacionadas
  - que as primeiras (poucas) novas variáveis retenham a maior parte da variação apresentada pelas variáveis originais (para que essa variação permita a separação das classes)

Como fazer essa transformação?

- Aplicação da transformação de Karhunen-Loève (ou *Spectral Decomposition*)
- No novo espaço, escolher as características que “governam o sinal”



[COSTA & CESAR, 2009]

Como fazer essa transformação?

- Aplicação da transformação de Karhunen-Loève (ou *Spectral Decomposition*)
- No novo espaço, escolher as características que “governam o sinal”
- Para isso, precisamos de conceitos de Álgebra Linear, Probabilidade e Estatística...
- ... vamos ver que não cursamos essas disciplinas à toa.

SISTEMAS DE INFORMAÇÃO

EACH

Álgebra linear

Álgebra Linear

- Escalares e vetores (vetores escritos em negrito ou com uma flecha em cima)

- Representação gráfica de vetores

- Espaço vetorial

- Transformação: Sejam dois espaços vetoriais R e S.
  - T: R → S
  - Exemplo: T: $\mathbb{R}^2 \rightarrow \mathbb{R}^2 \mid T(\mathbf{p}) = -\mathbf{p}$

Transformação Linear

- Transformação linear: T onde

$$T(\alpha \mathbf{p} + \beta \mathbf{q}) = \alpha T(\mathbf{p}) + \beta T(\mathbf{q})$$

  α, β escalares

- Pode ser expressa por uma matriz:

$$T(\mathbf{p}) = \mathbf{A}\mathbf{p}$$

Transformação Linear

- Exemplo:

  $T: \mathbb{R}^2 \to \mathbb{R}^2 \mid T(\mathbf{p}) = -\mathbf{p}$

- A = ?

Transformação Linear

- Exemplo:

  T: $\mathbb{R}^2 \rightarrow \mathbb{R}^2 \mid T(\mathbf{p}) = -\mathbf{p}$

- $A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$

SISTEMAS DE INFORMAÇÃO

EACH

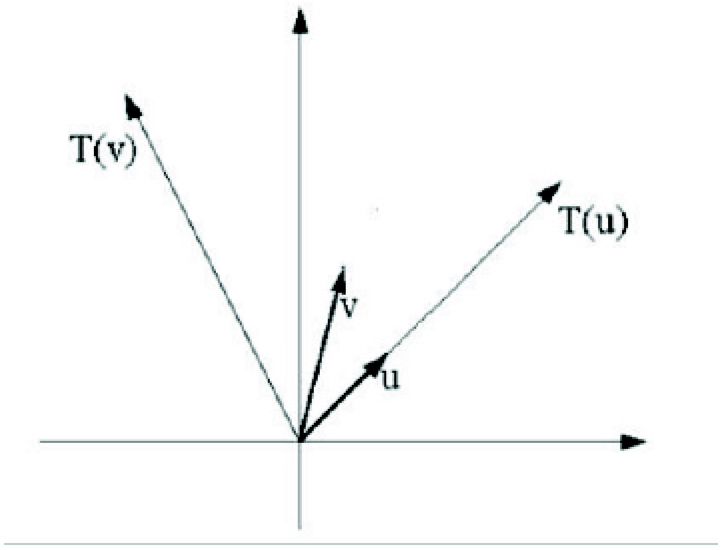
Paralelização de transformações lineares

- Transformação de vários vetores em paralelo

- Seja A uma transformação linear
  - **B** uma matriz onde a coluna i é o vetor $\mathbf{p}_i$
  - $\mathbf{C} = \mathbf{AB}$
  - sendo **C** uma matriz onde a coluna i é o vetor $\mathbf{q}_i = \mathbf{A}\mathbf{p}_i$

Autovalores e autovetores

- Dada uma matriz A_{nxn} que define uma transformação linear (não muda dimensionalidade)
- Existem vetores cuja orientação não é afetada por essa transformação (autovetores)



Ex: **u** é um autovetor de T, mas **v** não...

Autovalores e autovetores

- Dada uma matriz A_{nxn} que define uma transformação linear
- Existem vetores cuja orientação não é afetada por essa transformação (autovetores)
- Podem ser descobertos pela equação:

$$A\mathbf{x} = \lambda\mathbf{x}$$

sendo λ é um valor escalar complexo

$$A\mathbf{x} = \lambda\mathbf{x} \Rightarrow A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \Rightarrow (A - \lambda I)\mathbf{x} = \mathbf{0}$$

Autovalores e autovetores

- Dada uma matriz A_{nxn} que define uma transformação linear
- Existem vetores cuja orientação não é afetada por essa transformação (autovetores)
- Podem ser descobertos pela equação:

$$A\mathbf{x} = \lambda\mathbf{x}$$

sendo λ é um valor escalar complexo

$$A\mathbf{x} = \lambda\mathbf{x} \Rightarrow A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \Rightarrow (A- \lambda I)\mathbf{x} = \mathbf{0}$$

- só possui solução não trivial ($\mathbf{x} \neq \mathbf{0}$) se

$$\det(A- \lambda I)=0$$

Autovalores e autovetores

- Dada uma matriz A_{nxn} que define uma transformação linear
- Existem vetores cuja orientação não é afetada por essa transformação (autovetores)
- Podem ser descobertos pela equação:

$$A\mathbf{x} = \lambda\mathbf{x}$$

sendo λ é um valor escalar complexo

$$A\mathbf{x} = \lambda\mathbf{x} \quad \Rightarrow \quad A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \Rightarrow \quad (A - \lambda I)\mathbf{x} = \mathbf{0}$$

- só possui solução não trivial ($\mathbf{x} \neq \mathbf{0}$) se

$\det(A - \lambda I) = 0$ Equação característica

Soluções: autovalores

SISTEMAS DE INFORMAÇÃO

EACH

Autovalores e autovetores

Observação:
det(B) = 0 significa que B é responsável por B**x** = **0** (já que **x** ≠ **0**)

$$B\mathbf{x} = \mathbf{0} \Rightarrow \begin{cases} b_{11}x_1 + b_{12}x_2 = 0 \Rightarrow x_1 = -b_{12}x_2/b_{11} \\ b_{21}x_1 + b_{22}x_2 = 0 \Rightarrow -b_{21}b_{12}x_2/b_{11} + b_{22}x_2 = 0 \Rightarrow \\ \qquad -b_{21}b_{12}x_2 + b_{11}b_{22}x_2 = 0 \Rightarrow \end{cases}$$

Já que x2 ≠ 0, então $-b_{21}b_{12} + b_{11}b_{22} = 0$ (ie, det(B) = 0)

$$A\mathbf{x} = \lambda\mathbf{x} \Rightarrow A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \Rightarrow (A-\lambda I)\mathbf{x} = \mathbf{0}$$

- só possui solução não trivial (**x** ≠ **0**) se

det(A- λI)=0 Equação característica

Soluções: autovalores

SISTEMAS DE INFORMAÇÃO

EACH

Autovalores e autovetores

- Exemplo 1: Considere o operador linear definido no exemplo anterior: T: $R^2 \to R^2$. $T(x, y)=(4x + 5y, 2x + y)$. Encontre os autovalores de $A = \begin{bmatrix} 4 & 5 \\ 2 & 1 \end{bmatrix}$,

- Resolvemos a equação característica $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$:

$$\mathbf{A} - \lambda\mathbf{I} = \begin{bmatrix} 4 & 5 \\ 2 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4-\lambda & 5 \\ 2 & 1-\lambda \end{bmatrix}$$

$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \leftrightarrow (4 - \lambda)(1 - \lambda) - 10 = 0 \leftrightarrow \lambda^2 - 5\lambda - 6 = 0$

- $\lambda_1 = -1$ e $\lambda_2 = 6$.

SISTEMAS DE INFORMAÇÃO

EACH

Probabilidade e Estatística

Covariância

- Covariância entre duas variáveis aleatórias

$$Cov(X,Y) = E\,[(X-\mu_x)(Y-\mu_y)]$$

Covariância

- Covariância entre duas variáveis aleatórias

$$Cov(X,Y) = E\,[(X-\mu_x)(Y-\mu_y)]$$

ou

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Covariância

- Covariância entre duas variáveis aleatórias

$$Cov(X,Y) = E\,[(X-\mu_x)(Y-\mu_y)]$$

ou

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Mostra o quanto X e Y mudam juntas

Pode ser positiva (maiores valores de X correspondem aos maiores de Y e vice-versa) ou negativa (o contrário)

Magnitude: interpretação não trivial

Covariância

- Covariância entre duas variáveis aleatórias

$$Cov(X,Y) = E\left[(X-\mu_x)(Y-\mu_y)\right]$$

ou

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

- Matriz de covariância de um vetor aleatório **x**:

$$K_x = [Cov\ (x_i, x_j)]$$

Lembrando que $Cov(x_i,x_i) = \sigma^2(x_i)$ (variância)

Coeficiente de correlação linear

- Coeficiente de correlação (linear) entre duas variáveis aleatórias:

$$\rho_{X,Y} = \mathrm{Cov}(X,Y) / \sigma_x \sigma_y$$

- $|\rho_{X,Y}| \leq 1$ (quanto mais próximo de 1 "mais forte" a correlação)

- $\mathrm{Cov}(X,Y) = 0 \quad \Rightarrow \quad \rho_{X,Y} = 0 \Rightarrow$

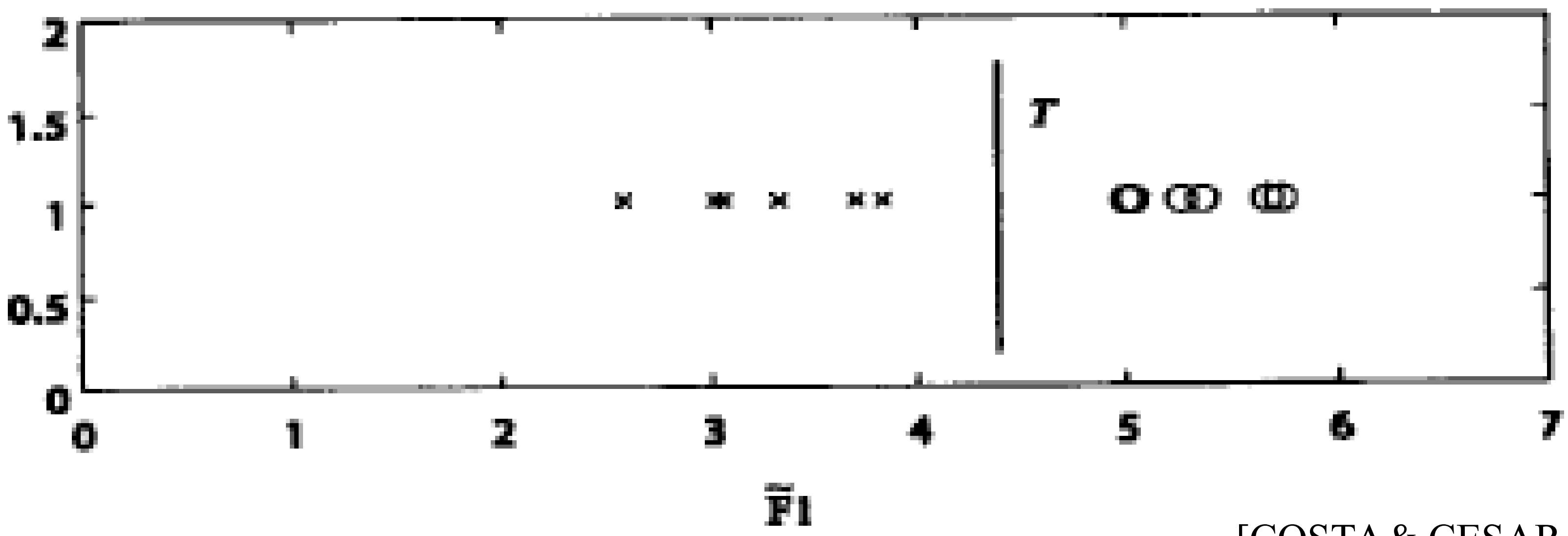
  Não há correlação linear entre X e Y

  Obs: isto NÃO quer dizer que sejam independentes

Juntando tudo

Exemplo

Transformação para apenas 1 nova característica (qual delas? A com maior dispersão)



[COSTA& CESAR, 2009]

Perfeitamente separáveis por um limiar!

Análise de Componentes Principais

- PCA – *Principal Components Analysis*
- Assume que os dados originais estão representados por variáveis correlacionadas (não independentes)
- Objetivo: transformar essas variáveis em novas variáveis (mudança de base do espaço vetorial) que
  - **não** sejam **correlacionadas**
  - que as primeiras (poucas) novas variáveis retenham a maior parte da **variação** apresentada pelas variáveis originais (para que essa variação permita a separação das classes)

PCA

as várias instâncias de **x** representam o dataset original

- Dado um vetor aleatório **x**, de n variáveis
- PCA quer achar n componentes principais por ordem decrescente de variabilidade

PCA

as várias instâncias de **x** representam o dataset original

- Dado um vetor aleatório **x**, de n variáveis
- PCA quer achar n componentes principais por ordem decrescente de variabilidade
- Primeiro componente principal y_1 (que seja uma combinação linear de **x**) é tal que a var(y_1) seja máxima, isto é quero achar um vetor $\boldsymbol{\alpha}^1$ tal que $y_1 = (\boldsymbol{\alpha}^1)^T \mathbf{x}$ e var ($y_1 = (\boldsymbol{\alpha}^1)^T \mathbf{x}$) seja máxima

PCA

as várias instâncias de **x** representam o dataset original

- Dado um vetor aleatório **x**, de n variáveis
- PCA quer achar n componentes principais por ordem decrescente de variabilidade
- Primeiro componente principal y_1 (que seja uma combinação linear de **x**) é tal que a var(y_1) seja máxima, isto é quero achar um vetor $\boldsymbol{\alpha}^1$ tal que $y_1 = (\boldsymbol{\alpha}^1)^T \mathbf{x}$ e var ($y_1 = (\boldsymbol{\alpha}^1)^T \mathbf{x}$) seja máxima
- Segundo componente principal y_2 onde a var(y_2) seja máxima (descontando a de y_1)
- E assim por diante (até n)

PCA

as várias instâncias de **x** representam o dataset original

- Dado um vetor aleatório **x**, de n variáveis
- PCA quer achar n componentes principais por ordem decrescente de variabilidade
- Primeiro componente principal y_1 (que seja uma combinação linear de **x**) é tal que a var(y_1) seja máxima, isto é quero achar um vetor $\boldsymbol{\alpha}^1$ tal que

  $y_1 = (\boldsymbol{\alpha}^1)^T \mathbf{x}$ e var ($y_1 = (\boldsymbol{\alpha}^1)^T \mathbf{x}$) seja máxima
- Segundo componente principal y_2 onde a var(y_2) seja máxima (descontando a de y_1)
- E assim por diante (até n)
- Normalmente os m primeiros c.p., m « n, representam bem a variabilidade dos dados originais

PCA – Link interessante

http://setosa.io/ev/principal-component-analysis/

PCA

- Primeiro componente principal y_1 é tal que a $var(y_1)$ seja máxima,
  isto é quero achar um vetor $\boldsymbol{\alpha}^{1}$ tal que $y_1 = (\boldsymbol{\alpha}^{1})^T \mathbf{x}$ e
  $var\,(y_1 = (\boldsymbol{\alpha}^{1})^T \mathbf{x})$ seja máxima

SISTEMAS DE INFORMAÇÃO

EACH

PCA – Exemplo em R^2 !!!

- Primeiro componente principal y_1 é tal que a var(y_1) seja máxima,

  isto é quero achar um vetor $\boldsymbol{\alpha}^{1}$ tal que $y_1 = (\boldsymbol{\alpha}^{1})^T \mathbf{x}$ e

  var ($y_1 = (\boldsymbol{\alpha}^{1})^T \mathbf{x}$) seja máxima

  - var (y_1) = var (($\boldsymbol{\alpha}^{1})^T \mathbf{x}$) = var ($\alpha^1{}_1 x_1 + \alpha^1{}_2 x_2$)

    $= (\alpha^1{}_1)^2 \text{var}(x_1) + (\alpha^1{}_2)^2 \text{var}(x_2) + 2\alpha^1{}_1\alpha^1{}_2 \text{Cov}(x_1, x_2)$

  - $= (\boldsymbol{\alpha}^{1})^T \mathbf{K}_{\mathbf{x}} \boldsymbol{\alpha}^{1}$

PCA – Exemplo em R^2 !!!

- Primeiro componente principal y_1 é tal que a $var(y_1)$ seja máxima,
  isto é quero achar um vetor $\boldsymbol{\alpha}^1$ tal que $y_1 = (\boldsymbol{\alpha}^1)^T \mathbf{x}$ e $var\ (y_1 = (\boldsymbol{\alpha}^1)^T \mathbf{x})$ seja máxima
  - $var\ (y_1) = var\ ((\boldsymbol{\alpha}^1)^T \mathbf{x}) = var\ (\alpha^1{}_1 x_1 + \alpha^1{}_2 x_2)$
    $= (\alpha^1{}_1)^2\ var(x_1) + (\alpha^1{}_2)^2\ var(x_2) + 2\alpha^1{}_1\alpha^1{}_2 Cov(x_1, x_2)$
  - $= (\boldsymbol{\alpha}^1)^T \mathbf{K}_x\ \boldsymbol{\alpha}^1$
- Precisamos achar $\boldsymbol{\alpha}^1$ que ?

PCA – Exemplo em R^2 !!!

- Primeiro componente principal y_1 é tal que a var(y_1) seja máxima,
  isto é quero achar um vetor $\boldsymbol{\alpha}^1$ tal que $y_1 = (\boldsymbol{\alpha}^1)^T \mathbf{x}$ e
  var ($y_1 = (\boldsymbol{\alpha}^1)^T \mathbf{x}$) seja máxima
  - $\text{var}(y_1) = \text{var}((\boldsymbol{\alpha}^1)^T \mathbf{x}) = \text{var}(\alpha^1{}_1 x_1 + \alpha^1{}_2 x_2)$
    $= (\alpha^1{}_1)^2 \text{var}(x_1) + (\alpha^1{}_2)^2 \text{var}(x_2) + 2\alpha^1{}_1\alpha^1{}_2 \text{Cov}(x_1, x_2)$
  - $= (\boldsymbol{\alpha}^1)^T \mathbf{K}_\mathbf{x} \boldsymbol{\alpha}^1$
- Precisamos achar $\boldsymbol{\alpha}^1$ que maximize $(\boldsymbol{\alpha}^1)^T \mathbf{K}_\mathbf{x} \boldsymbol{\alpha}^1$

PCA – Exemplo em R^2 !!!

- Primeiro componente principal y_1 é tal que a $var(y_1)$ seja máxima,

  isto é quero achar um vetor $\boldsymbol{\alpha}^1$ tal que $y_1 = (\boldsymbol{\alpha}^1)^T \mathbf{x}$ e $var\ (y_1 = (\boldsymbol{\alpha}^1)^T \mathbf{x})$ seja máxima

  - $var\ (y_1) = var\ ((\boldsymbol{\alpha}^1)^T \mathbf{x}) = var\ (\alpha^1_1 x_1 + \alpha^1_2 x_2)$

    $= (\alpha^1_1)^2\ var(x_1) + (\alpha^1_2)^2\ var(x_2) + 2\alpha^1_1\alpha^1_2 Cov(x_1, x_2)$
  - $= (\boldsymbol{\alpha}^1)^T \mathbf{K_x}\ \boldsymbol{\alpha}^1$
- Precisamos achar $\boldsymbol{\alpha}^1$ que maximize $(\boldsymbol{\alpha}^1)^T \mathbf{K_x}\ \boldsymbol{\alpha}^1$

Se adicionarmos uma restrição (por exemplo $(\boldsymbol{\alpha}^1)^T \boldsymbol{\alpha}^1 = 1$, => $(\boldsymbol{\alpha}^1)^T \boldsymbol{\alpha}^1 - 1 = 0$ (isto é, $\boldsymbol{\alpha}^1$ deve ser unitário), podemos usar a técnica de otimização de multiplicadores de Lagrange:

PCA

- Maximizar $f(\boldsymbol{\alpha}^1, \lambda) = (\boldsymbol{\alpha}^1)^T \mathbf{K}_\mathbf{x} \boldsymbol{\alpha}^1 - \lambda((\boldsymbol{\alpha}^1)^T \boldsymbol{\alpha}^1 - 1)$

PCA

- Maximizar $f(\boldsymbol{\alpha}^1, \lambda) = (\boldsymbol{\alpha}^1)^T \mathbf{K}_\mathbf{x} \boldsymbol{\alpha}^1 - \lambda((\boldsymbol{\alpha}^1)^T \boldsymbol{\alpha}^1 - 1)$
- Os pontos extremos de $f(\boldsymbol{\alpha}^1, \lambda)$ são tais que

$(\partial f/\partial \boldsymbol{\alpha}^1, \partial f/\partial \lambda) = (0,0)$ (derivadas parciais)

$\partial f/\partial \lambda$ recai na restrição $((\boldsymbol{\alpha}^1)^T \boldsymbol{\alpha}^1 - 1 = 0)$

$$\partial f/\partial \boldsymbol{\alpha}^1 = 2\mathbf{K}_\mathbf{x} \boldsymbol{\alpha}^1 - 2\lambda \boldsymbol{\alpha}^1 = 0$$

$$\mathbf{K}_\mathbf{x} \boldsymbol{\alpha}^1 - \lambda \boldsymbol{\alpha}^1 = 0$$

$$(\mathbf{K}_\mathbf{x} - \lambda I_n) \boldsymbol{\alpha}^1 = 0$$

PCA

- Maximizar $f(\boldsymbol{\alpha}^1, \lambda) = (\boldsymbol{\alpha}^1)^T \mathbf{K}_x \boldsymbol{\alpha}^1 - \lambda((\boldsymbol{\alpha}^1)^T \boldsymbol{\alpha}^1 - 1)$
- Os pontos extremos de $f(\boldsymbol{\alpha}^1, \lambda)$ são tais que

$(\partial f/\partial \boldsymbol{\alpha}^1, \partial f/\partial \lambda) = (0,0)$

$\partial f/\partial \lambda$ recai na restrição

$$\partial f/\partial \boldsymbol{\alpha}^1 = 2\mathbf{K}_x \boldsymbol{\alpha}^1 - 2\lambda\boldsymbol{\alpha}^1 = 0$$

$$\mathbf{K}_x \boldsymbol{\alpha}^1 - \lambda\boldsymbol{\alpha}^1 = 0$$

$$(\mathbf{K}_x - \lambda I_n) \boldsymbol{\alpha}^1 = 0$$

Reconhecem algo?

Autovalores e autovetores

Slide anterior da revisão de álgebra linear

- Dada uma matriz A_{nxn} que define uma transformação linear
- Existem vetores cuja orientação não é afetada por essa transformação (autovetores)
- Podem ser descobertos pela equação:

$$A\mathbf{x} = \lambda \mathbf{x}$$

onde λ é um valor escalar complexo

$$A\mathbf{x} = \lambda \mathbf{x} \;\Rightarrow\; A\mathbf{x} - \lambda \mathbf{x} = \mathbf{0} \;\Rightarrow\; (A- \lambda I)\mathbf{x} = \mathbf{0}$$

- só possui solução não trivial ($\mathbf{x} \neq \mathbf{0}$) se

$$\det(A- \lambda I)=0$$ Equação característica

Soluções: autovalores

SISTEMAS DE INFORMAÇÃO

EACH

PCA

- Maximizar $f(\boldsymbol{\alpha}^1, \lambda) = (\boldsymbol{\alpha}^1)^T \mathbf{K}_\mathbf{x} \boldsymbol{\alpha}^1 - \lambda((\boldsymbol{\alpha}^1)^T \boldsymbol{\alpha}^1 - 1)$
- Os pontos extremos de $f(\boldsymbol{\alpha}^1, \lambda)$ são tais que

  $(\partial f/\partial \boldsymbol{\alpha}^1, \partial f/\partial \lambda) = (0,0)$

  $\partial f/\partial \lambda$ recai na restrição

  $\partial f/\partial \boldsymbol{\alpha}^1 = \mathbf{K}_\mathbf{x} \boldsymbol{\alpha}^1 - \lambda \boldsymbol{\alpha}^1 = 0$

  $$(\mathbf{K}_\mathbf{x} - \lambda I_n)\, \boldsymbol{\alpha}^1 = 0$$
- Logo, λ é um autovalor de $\mathbf{K}_\mathbf{x}$ e $\boldsymbol{\alpha}^1$ é seu correspondente autovetor

SISTEMAS DE INFORMAÇÃO

EACH

PCA

- Maximizar $f(\boldsymbol{\alpha}^1, \lambda) = (\boldsymbol{\alpha}^1)^T \mathbf{K}_x \boldsymbol{\alpha}^1 - \lambda((\boldsymbol{\alpha}^1)^T \boldsymbol{\alpha}^1 - 1)$
- Os pontos extremos de $f(\boldsymbol{\alpha}^1, \lambda)$ são tais que

  $(\partial f/\partial \boldsymbol{\alpha}^1, \partial f/\partial \lambda) = (0,0)$

  $\partial f/\partial \lambda$ recai na restrição

  $\partial f/\partial \boldsymbol{\alpha}^1 = \mathbf{K}_x \boldsymbol{\alpha}^1 - \lambda \boldsymbol{\alpha}^1 = 0$

  $$(\mathbf{K}_x - \lambda I_n)\, \boldsymbol{\alpha}^1 = 0$$
- Logo, λ é um autovalor de $\mathbf{K}_x$ e $\boldsymbol{\alpha}^1$ é seu correspondente autovetor
- Quero $\boldsymbol{\alpha}^1$ que maximize $(\boldsymbol{\alpha}^1)^T \mathbf{K}_x \boldsymbol{\alpha}^1$

SISTEMAS DE INFORMAÇÃO

EACH

PCA

- Maximizar $f(\boldsymbol{\alpha}^1, \lambda) = (\boldsymbol{\alpha}^1)^T \mathbf{K_x} \boldsymbol{\alpha}^1 - \lambda((\boldsymbol{\alpha}^1)^T \boldsymbol{\alpha}^1 - 1)$
- Os pontos extremos de $f(\boldsymbol{\alpha}^1, \lambda)$ são tais que
  $(\partial f/\partial \boldsymbol{\alpha}^1, \partial f/\partial \lambda) = (0,0)$
  $\partial f/\partial \lambda$ recai na restrição
  $\partial f/\partial \boldsymbol{\alpha}^1 = \mathbf{K_x} \boldsymbol{\alpha}^1 - \lambda \boldsymbol{\alpha}^1 = 0$
  $$(\mathbf{K_x} - \lambda I_n)\, \boldsymbol{\alpha}^1 = 0$$
- Logo, λ é um autovalor de $\mathbf{K_x}$ e $\boldsymbol{\alpha}^1$ é seu correspondente autovetor
- Quero $\boldsymbol{\alpha}^1$ que maximize $(\boldsymbol{\alpha}^1)^T \mathbf{K_x} \boldsymbol{\alpha}^1$ (no caso valerá $\mathbf{K_x} \boldsymbol{\alpha}^1 = \lambda \boldsymbol{\alpha}^1$) então quero $\boldsymbol{\alpha}^1$ que maximize $(\boldsymbol{\alpha}^1)^T \lambda \boldsymbol{\alpha}^1 = \lambda (\boldsymbol{\alpha}^1)^T \boldsymbol{\alpha}^1 = \lambda$

  Portanto basta o quê ?

PCA

- Maximizar $f(\boldsymbol{\alpha}^1, \lambda) = (\boldsymbol{\alpha}^1)^T \mathbf{K}_\mathbf{x} \boldsymbol{\alpha}^1 - \lambda((\boldsymbol{\alpha}^1)^T \boldsymbol{\alpha}^1 - 1)$
- Os pontos extremos de $f(\boldsymbol{\alpha}^1, \lambda)$ são tais que

  $(\partial f/\partial \boldsymbol{\alpha}^1, \partial f/\partial \lambda) = (0,0)$

  $\partial f/\partial \lambda$ recai na restrição

  $\partial f/\partial \boldsymbol{\alpha}^1 = \mathbf{K}_\mathbf{x} \boldsymbol{\alpha}^1 - \lambda \boldsymbol{\alpha}^1 = 0$

  $$(\mathbf{K}_\mathbf{x} - \lambda I_n)\, \boldsymbol{\alpha}^1 = 0$$

- Logo, λ é um autovalor de $\mathbf{K}_\mathbf{x}$ e $\boldsymbol{\alpha}^1$ é seu correspondente autovetor
- Quero $\boldsymbol{\alpha}^1$ que maximize $(\boldsymbol{\alpha}^1)^T \mathbf{K}_\mathbf{x} \boldsymbol{\alpha}^1$ (no caso valerá $\mathbf{K}_\mathbf{x} \boldsymbol{\alpha}^1 = \lambda \boldsymbol{\alpha}^1$) então quero $\boldsymbol{\alpha}^1$ que maximize $(\boldsymbol{\alpha}^1)^T \lambda \boldsymbol{\alpha}^1 = \lambda (\boldsymbol{\alpha}^1)^T \boldsymbol{\alpha}^1 = \lambda$

  Portanto basta escolher o autovetor correspondente ao maior autovalor!

SISTEMAS DE INFORMAÇÃO

EACH

PCA

- i-ésimo componente principal: $y_i = (\boldsymbol{\alpha}^{\mathbf{i}})^T \mathbf{x}$ onde $\boldsymbol{\alpha}^{\mathbf{i}}$ é o autovetor correspondente ao i-ésimo maior autovalor de $\mathbf{K}_{\mathbf{x}}$

PCA

Seja **x** um vetor aleatório de dimensão n:

1. Calcule sua matriz de covariância K_x
2. Calcule os autovalores e autovetores de K_x
3. Monte a matriz Ω onde cada linha é um autovetor de K_x (em ordem decrescente de seus autovalores)

$$\mathbf{y} = \Omega\,\mathbf{x}$$

SISTEMAS DE INFORMAÇÃO

EACH

Transformação Karhunen-Loève

Seja **x** um vetor aleatório de dimensão n:

1. Calcule sua matriz de covariância K_x
2. Calcule os autovalores e autovetores de K_x
3. Monte a matriz Ω onde cada linha é um autovetor de K_x (em ordem decrescente de seus autovalores)

Transformação Karhunen-Loève:

$$\mathbf{y} = \Omega\, \mathbf{x}$$

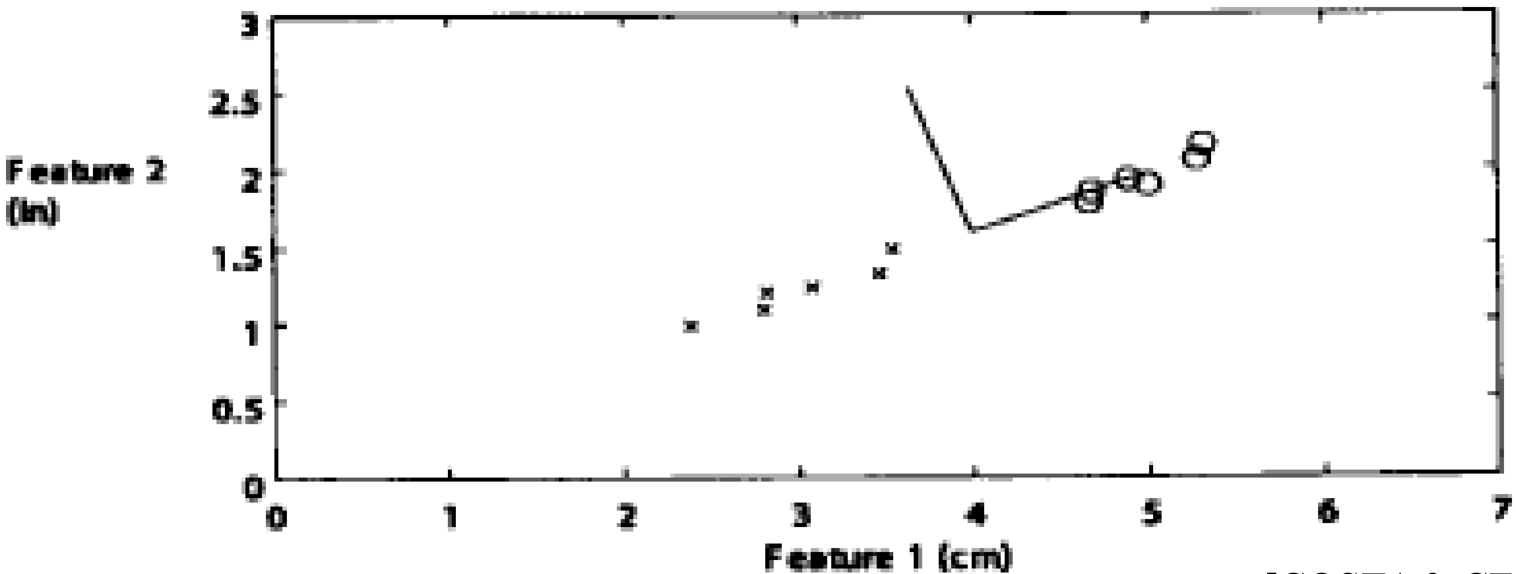
Transformação Karhunen-Loève

- Propriedades:
  - Rotaciona os antigos eixos para novas orientações no espaço n-dimensional
  - Maximiza a variância (dispersão) ao longo de cada novo eixo
  - A matriz de covariância de **y** é $K_y = \boldsymbol{\lambda} I$, onde $\boldsymbol{\lambda}$ é o vetor formados pelos autovalores de K_x e ordenados decrescentemente ($\lambda_i = var(y_i)$)
  - ⇒ novas variáveis escalares não são correlacionadas (linearmente)

SISTEMAS DE INFORMAÇÃO

EACH

Voltando ao nosso exemplo



[COSTA& CESAR, 2009]

$$K = \begin{bmatrix} 1.1547 & 0.4392 \\ 0.4392 & 0.1697 \end{bmatrix} \qquad \rho = \begin{bmatrix} 1.0 & 0.992 \\ 0.992 & 1.0 \end{bmatrix}$$

SISTEMAS DE INFORMAÇÃO

EACH

Voltando ao nosso exemplo

- Aplicando a transformação Karhunen-Loève separadamente para cada exemplo $\mathbf{x}^i$ do dataset:

$$\mathbf{y}^i = \Omega\, \mathbf{x}^i$$

SISTEMAS DE INFORMAÇÃO

EACH

Voltando ao nosso exemplo

- Aplicando a transformação Karhunen-Loève separadamente para cada exemplo $\mathbf{x}^i$ do dataset:

$$\mathbf{y}^i = \Omega\, \mathbf{x}^i$$

ou em paralelo (D seria o dataset inteiro, com os vetores de características $\mathbf{x}^i$ nas linhas, e D' o dataset todo transformado, também com os vetores de características $\mathbf{y}^i$ nas linhas):

$$D'^T = \Omega D^T$$

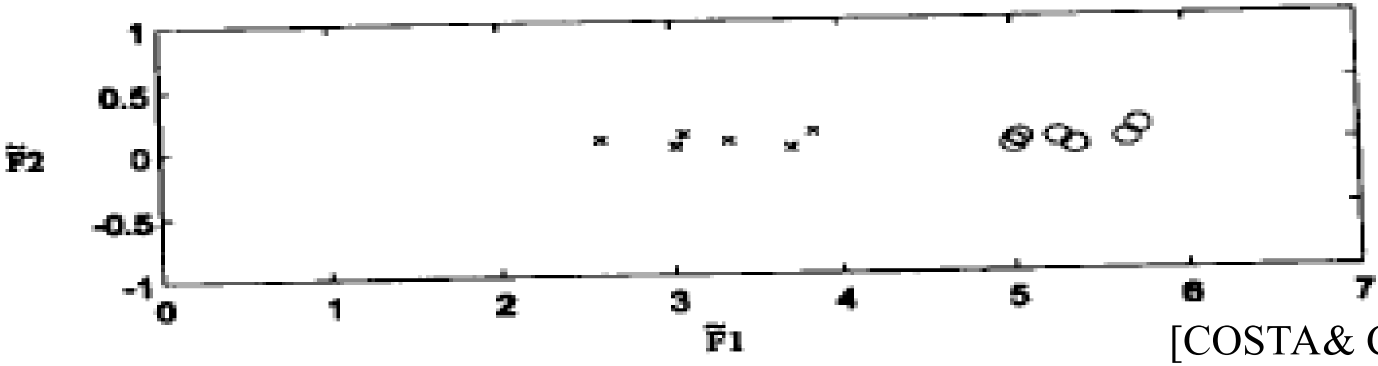
$$\mathbf{D' = D\, \Omega^T}$$

SISTEMAS DE INFORMAÇÃO

EACH

Voltando ao nosso exemplo

F'_1 varia mais que F'_2



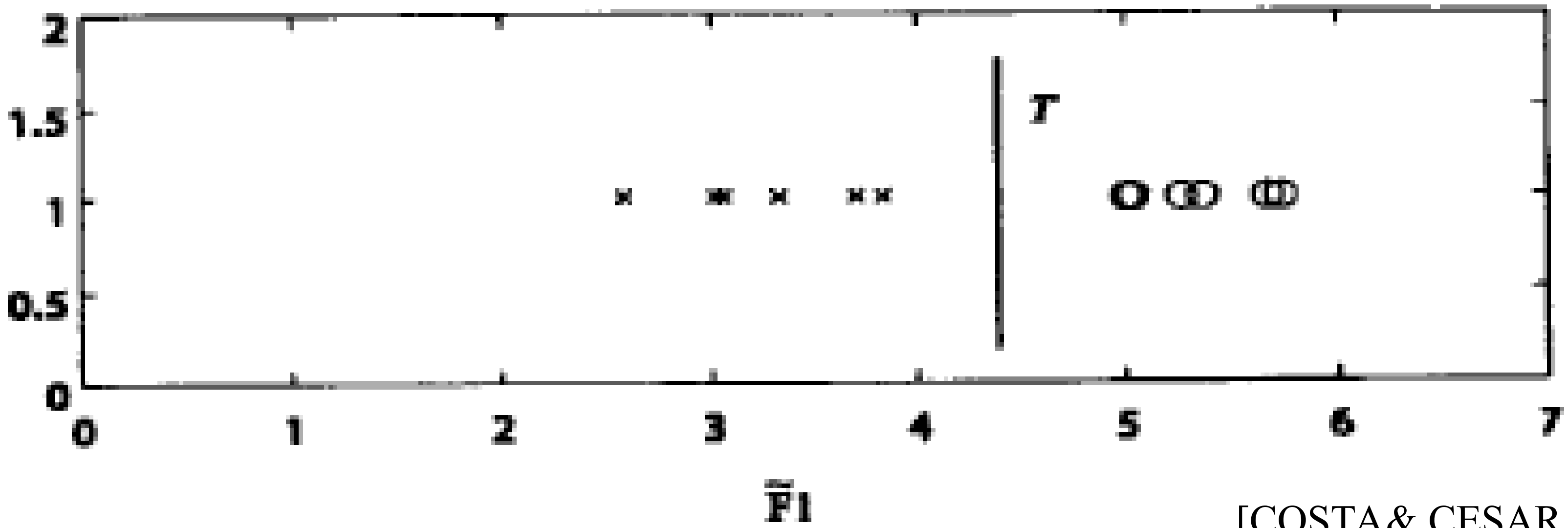
[COSTA& CESAR, 2009]

SISTEMAS DE INFORMAÇÃO

EACH

Voltando ao nosso exemplo

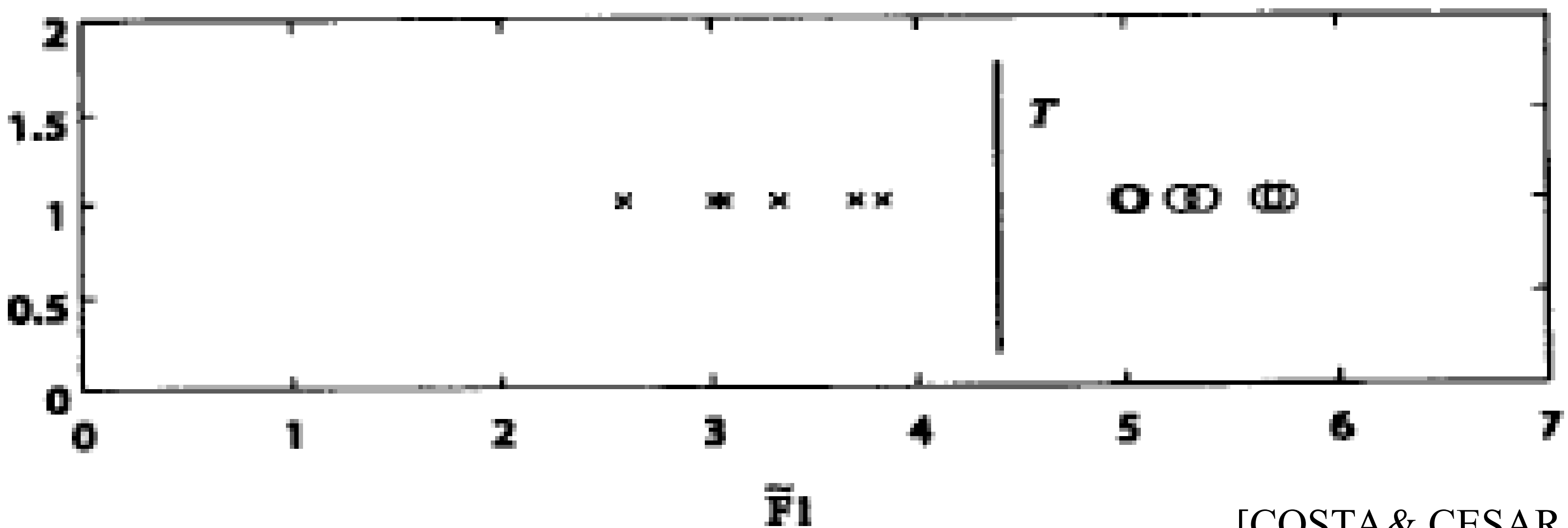
- Olhando só para F'_1



[COSTA& CESAR, 2009]

Voltando ao nosso exemplo

- Olhando só para F'_1



[COSTA& CESAR, 2009]

REDUÇÃO DE DIMENSIONALIDADE!!!

SISTEMAS DE INFORMAÇÃO

EACH

Análise de Componentes Principais

- Aplicação da transformação Karhunen-Loève
- Os maiores autovalores são os que “governam” o sinal
- Os demais são redundantes, ou contém ruídos ou bem pouco necessários (daria para descartá-los para evitar a maldição da dimensionalidade)

Variação original e componentes individuais

Qual é a proporção da variação total (dos dados originais) que um componente consegue explicar? (Lembrando que λ_i = var(y_i))

Variação original e componentes individuais

Qual é a proporção da variação total (dos dados originais) que um componente consegue explicar? (Lembrando que λ_i = var(y_i))

variação total: $\sum_j \lambda_j$

proporção da variação "explicada" pelo i-ésimo componente: $p_i = \lambda_i / \sum_j \lambda_j$

Redução de dimensionalidade

Usar, ao invés das n variáveis originais, apenas os m primeiros componentes ($m << n$)

Qual valor de m? Aquele no qual a variação acumulada ($p_1+p_2+ \ldots +p_m$) é o suficiente para "explicar" a variação dos dados

Geralmente escolhido um limiar entre 70% a 90%

- m não pode ser muito próximo de n
- não faz sentido escolher 70% e obter m muito pequeno para valores "grandes" de n
- ....

PCA – Comentários finais

- Objetivo: remover redundância (compressão), redução de dimensionalidade, normalização de variáveis
- Medida de redundância: correlação
- Se as características forem independentes,

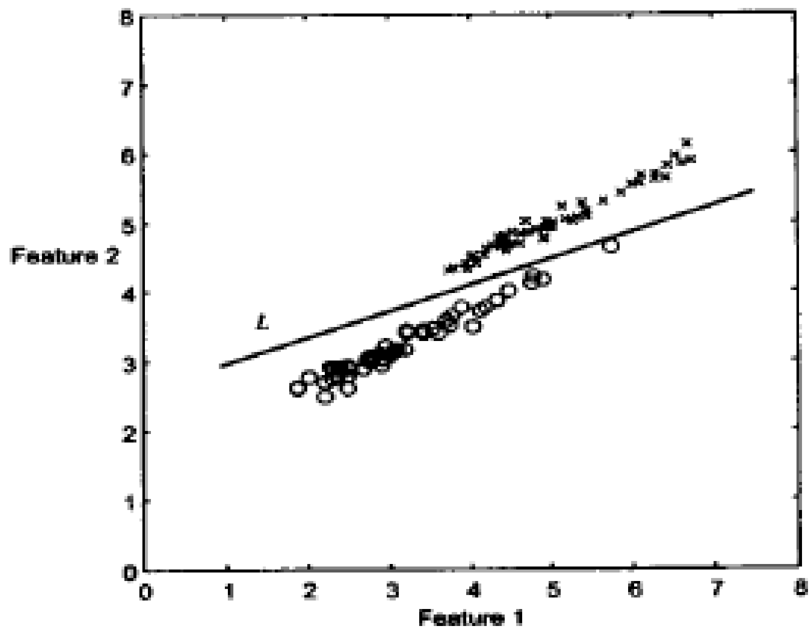
... ?

PCA – Comentários finais

- Objetivo: remover redundância (compressão), redução de dimensionalidade, normalização de variáveis
- Medida de redundância: correlação
- Se as características forem independentes, não há vantagens
- Só faz sentido se algumas das variáveis observadas são correlacionadas
- Não é feita nenhuma suposição sobre a densidade de probabilidade dos vetores

PCA – Comentários finais - CUIDADOS

- Nem sempre essa abordagem é útil (mesmo quando as características originais são correlacionadas)
  - Pode causar sobreposição de classes



[COSTA& CESAR, 2009]

SISTEMAS DE INFORMAÇÃO

EACH

PCA – Comentários finais - CUIDADOS

- Em altas dimensões não podemos visualizar os dados para decidir se aplicamos PCA ou não

- Solução?

PCA – Comentários finais - CUIDADOS

- Em altas dimensões não podemos visualizar os dados para decidir se aplicamos PCA ou não

- Solução: comparar resultados de classificação com e sem PCA

PCA – Matriz de covariância ou matriz de correlação

- Pode ser baseado em matrizes de covariância ou de correlação. Qual é melhor? (Jolliffe, 2002)
- Há uma ampla discussão sobre isso
  - Não há uma melhor solução sempre, e as duas apresentam vantagens e desvantagens
  - Mas algumas coisas parecem ser consenso...

PCA – Matriz de covariância ou matriz de correlação - Exemplo

Table 3.1. Correlations and standard deviations for eight blood chemistry variables.

	Correlation matrix ($n = 72$)							
	RBLOOD	PLATE	WBLOOD	NEUT	LYMPH	BILIR	SODIUM	POTASS
RBLOOD	1.000							
PLATE	0.290	1.000						
WBLOOD	0.202	0.415	1.000					
NEUT	−0.055	0.285	0.419	1.000				
LYMPH	−0.105	−0.376	−0.521	−0.877	1.000			
BILIR	−0.252	−0.349	−0.441	−0.076	0.206	1.000		
SODIUM	−0.229	−0.164	−0.145	0.023	0.034	0.192	1.000	
POTASS	0.058	−0.129	−0.076	−0.131	0.151	0.077	0.423	1.000
Standard deviations	0.371	41.253	1.935	0.077	0.071	4.037	2.732	0.297

PCA – Matriz de covariância ou matriz de correlação - Exemplo

Table 3.1. Correlations and standard deviations for eight blood chemistry variables.

	Correlation matrix ($n = 72$)							
	RBLOOD	PLATE	WBLOOD	NEUT	LYMPH	BILIR	SODIUM	POTASS
RBLOOD	1.000							
PLATE	0.290	1.000						
WBLOOD	0.202	0.415	1.000					
NEUT	−0.055	0.285	0.419	1.000				
LYMPH	−0.105	−0.376	−0.521	$\boxed{-0.877}$	1.000			
BILIR	−0.252	−0.349	−0.441	−0.076	0.206	1.000		
SODIUM	−0.229	−0.164	−0.145	0.023	0.034	0.192	1.000	
POTASS	0.058	−0.129	−0.076	−0.131	0.151	0.077	0.423	1.000
Standard deviations	0.371	41.253	1.935	0.077	0.071	4.037	2.732	0.297

Correlações discretas a menos de um valor

SISTEMAS DE INFORMAÇÃO

EACH

PCA – Matriz de covariância ou matriz de correlação - Exemplo

Table 3.1. Correlations and standard deviations for eight blood chemistry variables.

Correlation matrix ($n = 72$)								
	RBLOOD	PLATE	WBLOOD	NEUT	LYMPH	BILIR	SODIUM	POTASS
RBLOOD	1.000							
PLATE	0.290	1.000						
WBLOOD	0.202	0.415	1.000					
NEUT	−0.055	0.285	0.419	1.000				
LYMPH	−0.105	−0.376	−0.521	−0.877	1.000			
BILIR	−0.252	−0.349	−0.441	−0.076	0.206	1.000		
SODIUM	−0.229	−0.164	−0.145	0.023	0.034	0.192	1.000	
POTASS	0.058	−0.129	−0.076	−0.131	0.151	0.077	0.423	1.000
Standard deviations	0.371	41.253	1.935	0.077	0.071	4.037	2.732	0.297

Correlações discretas a menos de um valor

Alta variabilidade nas variâncias, o que dever ser devido às escalas muito diversas

SISTEMAS DE INFORMAÇÃO

EACH

PCA – Matriz de covariância ou matriz de correlação - Exemplo

Table 3.1. Correlations and standard deviations for eight blood chemistry variables.

	Correlation matrix ($n = 72$)							
	RBLOOD	PLATE	WBLOOD	NEUT	LYMPH	BILIR	SODIUM	POTASS
RBLOOD	1.000							
PLATE	0.290	1.000						
WBLOOD	0.202	0.415	1.000					
NEUT	−0.055	0.285	0.419	1.000				
LYMPH	−0.105	−0.376	−0.521	−0.877	1.000			
BILIR	−0.252	−0.349	−0.441	−0.076	0.206	1.000		
SODIUM	−0.229	−0.164	−0.145	0.023	0.034	0.192	1.000	
POTASS	0.058	−0.129	−0.076	−0.131	0.151	0.077	0.423	1.000
Standard deviations	0.371	$\boxed{41.253}$	1.935	0.077	0.071	$\boxed{4.037}$	2.732	0.297

Correlações discretas a menos de um valor

Alta variabilidade nas variâncias, o que dever ser devido às escalas muito diversas

SISTEMAS DE INFORMAÇÃO

EACH

PCA – Matriz de covariância ou matriz de correlação - Exemplo

Table 3.3. Principal components based on the covariance matrix for eight blood chemistry variables.

Component number	1	2	3	4
	Coefficients			
RBLOOD	0.0	0.0	0.0	0.0
PLATE	1.0	0.0	0.0	0.0
WBLOOD	0.0	−0.2	0.0	1.0
NEUT	0.0	0.0	0.0	0.0
LYMPH	0.0	0.0	0.0	0.0
BILIR	0.0	1.0	−0.2	0.2
SODIUM	0.0	0.2	1.0	0.0
POTASS	0.0	0.0	0.0	0.0
Percentage of total variation explained	98.6	0.9	0.4	0.2

PCA – Matriz de covariância ou matriz de correlação - Exemplo

Table 3.3. Principal components based on the covariance matrix for eight blood chemistry variables.

Component number	1	2	3	4
	Coefficients			
RBLOOD	0.0	0.0	0.0	0.0
PLATE	1.0	0.0	0.0	0.0
WBLOOD	0.0	−0.2	0.0	1.0
NEUT	0.0	0.0	0.0	0.0
LYMPH	0.0	0.0	0.0	0.0
BILIR	0.0	1.0	−0.2	0.2
SODIUM	0.0	0.2	1.0	0.0
POTASS	0.0	0.0	0.0	0.0
Percentage of total variation explained	98.6	0.9	0.4	0.2

PCA – Matriz de covariância ou matriz de correlação - Exemplo

Table 3.3. Principal components based on the covariance matrix for eight blood chemistry variables.

Component number	1	2	3	4
	Coefficients			
RBLOOD	0.0	0.0	0.0	0.0
PLATE	1.0	0.0	0.0	0.0
WBLOOD	0.0	−0.2	0.0	1.0
NEUT	0.0	0.0	0.0	0.0
LYMPH	0.0	0.0	0.0	0.0
BILIR	0.0	1.0	−0.2	0.2
SODIUM	0.0	0.2	1.0	0.0
POTASS	0.0	0.0	0.0	0.0
Percentage of total variation explained	98.6	0.9	0.4	0.2

Então para quê PCA?

PCA – Matriz de covariância ou matriz de correlação - Exemplo

Table 3.2. Principal components based on the correlation matrix for eight blood chemistry variables.

Component number	1	2	3	4
	Coefficients			
RBLOOD	0.2	−0.4	0.4	0.6
PLATE	0.4	−0.2	0.2	0.0
WBLOOD	0.4	0.0	0.2	−0.2
NEUT	0.4	0.4	−0.2	0.2
LYMPH	−0.4	−0.4	0.0	−0.2
BILIR	−0.4	0.4	−0.2	0.6
SODIUM	−0.2	0.6	0.4	−0.2
POTASS	−0.2	0.2	0.8	0.0
Percentage of total variation explained	34.9	19.1	15.6	9.7

PCA – Matriz de covariância ou matriz de correlação - Exemplo

Table 3.2. Principal components based on the correlation matrix for eight blood chemistry variables.

Component number	1	2	3	4
	Coefficients			
RBLOOD	0.2	−0.4	0.4	0.6
PLATE	0.4	−0.2	0.2	0.0
WBLOOD	0.4	0.0	0.2	−0.2
NEUT	0.4	0.4	−0.2	0.2
LYMPH	−0.4	−0.4	0.0	−0.2
BILIR	−0.4	0.4	−0.2	0.6
SODIUM	−0.2	0.6	0.4	−0.2
POTASS	−0.2	0.2	0.8	0.0
Percentage of total variation explained	34.9	19.1	15.6	9.7

Provavelmente detectou correlações mais sutis

SISTEMAS DE INFORMAÇÃO

EACH

PCA – Matriz de covariância ou matriz de correlação

- Matriz de correlação normaliza os dados
  - Efeito semelhante seria normalizar os dados antes de usar a matriz de covariância
- Dados com fontes muito distintas de informação (e unidades métricas distintas, intervalos de valores discrepantes) “pedem” uma normalização (prévia, ou via uso de matriz de correlação), senão algumas poucas variáveis originais vão dominar
- Há quem defenda que essa normalização é muito “brutal”, e que se seus dados “comportados” (mesma escala, mesma origem de dados, variâncias similares, etc), pode haver perda de informação valiosa
    - → preferível utilizar com matriz de covariância sobre dados não normalizados

PCA – Matriz de covariância ou matriz de correlação - CUIDADO

- PCA pode ser utilizado, por exemplo, com uma função descritiva ou para redução de dimensionalidade dentro de um contexto de aprendizado supervisionado ou não supervisionado
- Se supervisionado: após o treinamento do classificador, um NOVO dado precisa ser analisado => seu vetor de características precisa ser transformado para o novo espaço vetorial dos PCs.
- Se os dados de treinamento foram normalizados, o dado de teste precisa ser normalizado da mesma forma (por ex, usando a mesma média e desvio padrão, de cada característica original, utilizada na normalização dos dados de treinamento)

PCA – Comentários finais

- PCA, como exposto aqui, faz sentido para variáveis categóricas?

Quadro 9.2 Níveis de mensuração e estatísticas possíveis

Escala	Exemplos	Operações empíricas básicas	Estatísticas possíveis
Nominal	Sexo, cor dos olhos, partido político	Determinação de igualdade	Número de casos % moda
Ordinal	Classificação em concursos, escalas tipo Likert	Anteriores, determinação >, <	Anteriores, mediana, percentis
Intervalar	Escore de QI, temperatura medida em graus Celsius	Anteriores, determinação dos intervalos das diferenças	Anteriores, média, desvio-padrão, correlação de postos (Spearman [estatística não paramétrica, dados não normais]), correlação produto-momento (Pearson)[estatística paramétrica, dados normalizados])
Racional	Tempo, temperatura medida em graus Kelvin, número de filhos	Anteriores, determinação da igualdade de razões	Todas, por exemplo, coeficiente de variação

[Appolinário, 2012]

EACH

PCA – Comentários finais

- PCA para variáveis categóricas: necessidade de adaptações (paper no TIDIA)

PCA – Comentários finais

- O método apresentado é conhecido como *Spectral Decomposition* (analisa covariância entre as características)
- Método alternativo *Singular Value Decomposition* (analisa covariância entre os exemplos)
  - Método computacionalmente eficiente de cálculo do PCs
  - Mais apropriado quando há muito mais características do que exemplos na amostra

SISTEMAS DE INFORMAÇÃO

EACH

PCA (trabalho)

Trabalho:

- Objetivo: reduzir a dimensionalidade do dataset de vocês... (achar os componentes principais do dataset e decidir quantos componentes vai usar)

PDF de slides:

- Métodos: função utilizada e parâmetros (R, Matlab, python), se foi necessário algum pré-processamento dos dados, qual matriz utilizou (e se normalizou os dados ou não) e por quê...
  – Pode testar mais de uma alternativa para comparar os resultados
- Resultados:
  * mostrar os PCs (coeficientes) e as variâncias acumuladas
  * com base nos resultados, decidir quantos componentes vai usar, explicando o porquê dessa escolha
- Discussão: sua análise crítica dos resultados

PDF de documento: scripts utilizados (só para postar no Moodle)

PCA no R

- Função **princomp** (Spectral Decomposition, como vimos em aula, com matriz de covariância ou correlação
- Função **prcomp**: usa SVD (Single Value Decomposition)
- Além de outras, sem indícios de grandes diferenças nos resultados:
- Anderson, G. B. PRINCIPAL COMPONENT ANALYSIS IN R AN EXAMINATION OF THE DIFFERENT FUNCTIONS AND METHODS TO PERFORM PCA

PCA em Python

Links interessantes:

https://plot.ly/ipython-notebooks/principal-component-analysis/

PCA (trabalho)

Apoio:

- Manual do princomp no R
  http://stat.ethz.ch/R-manual/R-patched/library/stats/html/princomp.html

- Cap 6 do Jolliffe (ver referências)

- Diferentes formas de PCA (normalmente sem grandes diferenças nos resultados) :

  http://www.ime.usp.br/~pavan/pdf/MAE0330-PCA-R-2013

- Cap 13 do HSAUR

  http://cran.r-project.org/web/packages/HSAUR/vignettes/Ch_principal_components_analysis.pdf

- Muita coisa na internet...

Referências

- COSTA, L. F.; CESAR, R. M. Jr. **Shape Classification and Analysis: Theory and Practice**. 2. ed.CRC Press, 2009 (Cap. 8.1.6)
- EVERITT, B. ; HOTHORN, T. **A handbook of statistical analysis using R**. Ed. Chapman & Hall/CRC
- MAGALHÃES, M. N.; LIMA, A. C. P. de: **Noções de Probabilidade e Estatística**. Edusp, 2002
- JOLLIFFE, I. T. **Principal Component Analysis**. 2. ed. Springer, 2002 (Cap. 1)
  - Cap 6 discute quantos componentes escolher