SELEÇÃO E EXTRAÇÃO DE ATRIBUTOS

Prof. André Backes | @progdescomplicada

Maldição da dimensionalidade

- Maldição da dimensionalidade (ou *Curse of dimensionality*)
  - Termo que se refere a vários fenômenos que surgem na análise de dados em espaços com muitas dimensões (atributos)
    - Muitas vezes com centenas ou milhares de dimensões

Maldição da dimensionalidade

- Maldição da dimensionalidade (ou *Curse of dimensionality*)
  - Basicamente, adicionar características não significa sempre melhora no desempenho de um classificador
    - Quanto maior a dimensionalidade do seu vetor, mais dados serão necessários para a aprendizagem do modelo

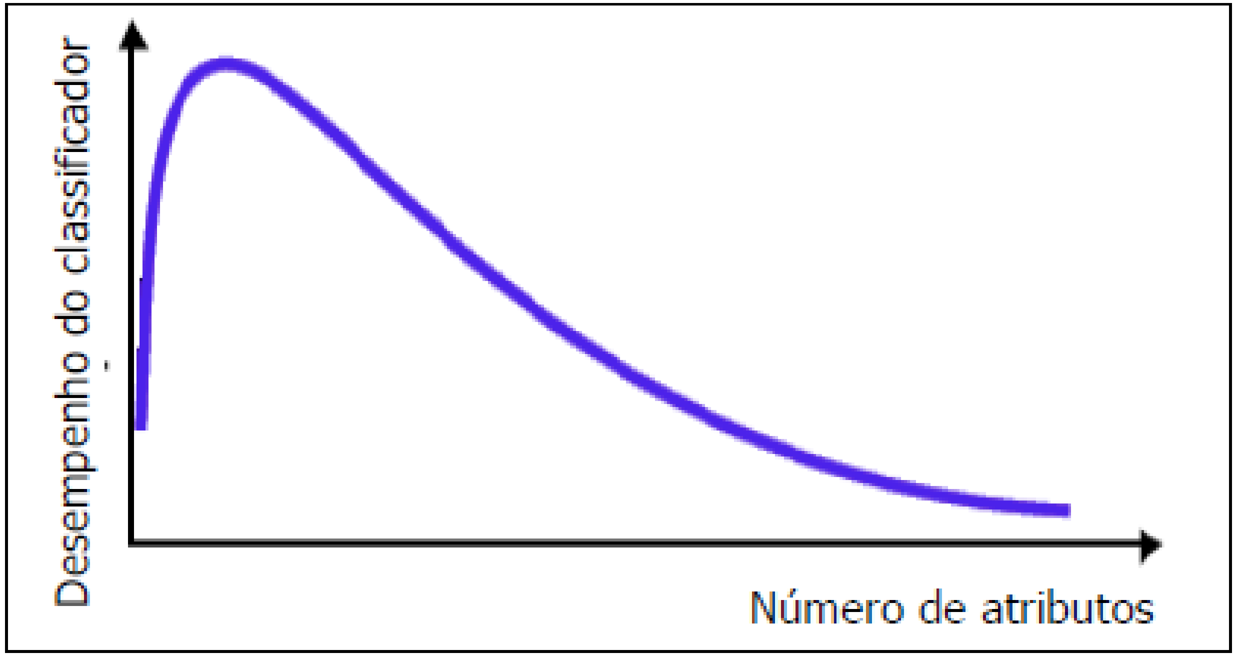
Maldição da dimensionalidade

- Suponha o seguinte problema
  - Um conjunto de dados é descrito por 20 atributos
    - Apenas 2 atributos são relevantes
    - Os demais são atributos ruins ou correlacionados

  - O resultado será um mau desempenho na classificação
    - O algoritmo K-NN é normalmente enganado quando o número de atributos é grande
    - Assim como outros classificadores também tem seu desempenho prejudicado

Maldição da dimensionalidade

- De modo geral, o desempenho de um classificador tende a se degradar a partir de um determinado nº de atributos
  - Mesmo que eles sejam atributos úteis



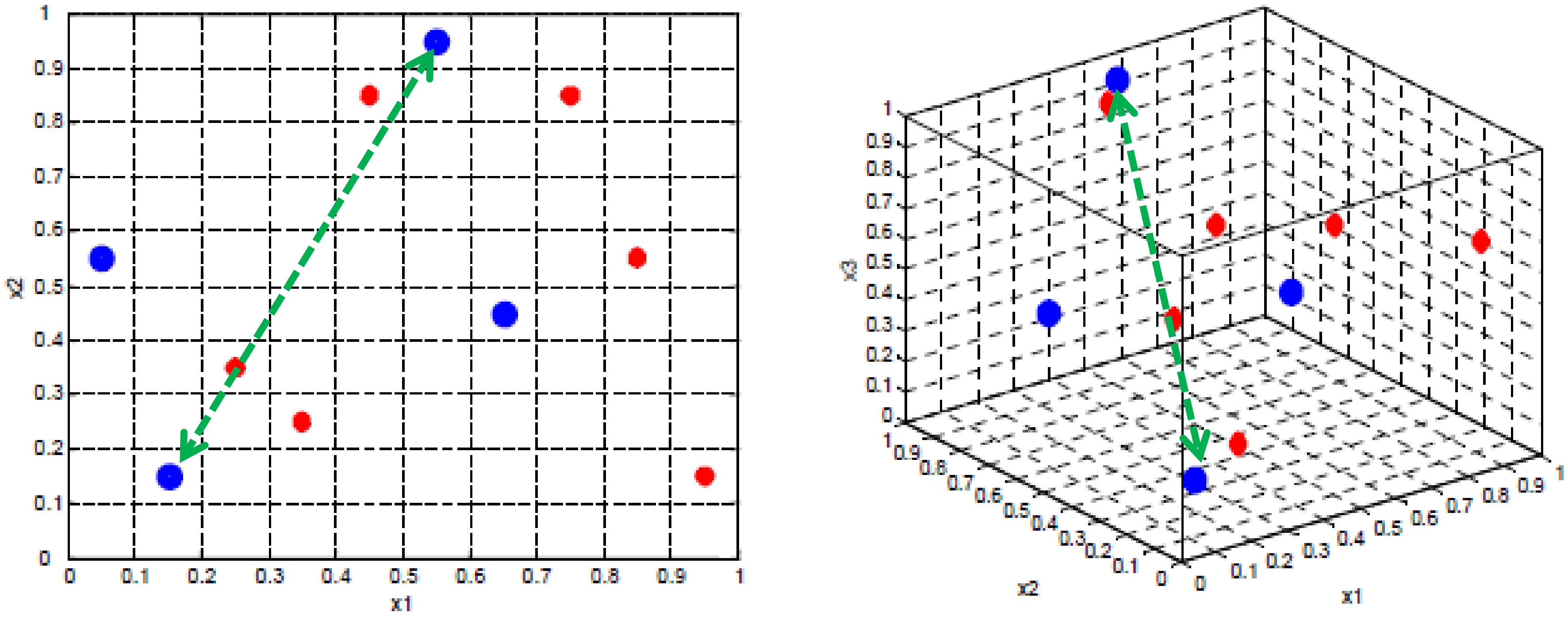
Maldição da dimensionalidade

- 1 atributo = 1 dimensão no espaço de características
  - Hiper-volume cresce exponencialmente com a adição de novos atributos
    - 1 atributo com 10 possíveis valores: 10 possíveis objetos
    - 5 atributos com 10 possíveis valores: 10^5 possíveis objetos

  - Em espaços com muitas dimensões as amostras se tornan esparsas e pouco similares
    - Objetos muito distantes uns dos outros
    - Objetos parecem eqüidistantes

Maldição da dimensionalidade

- Mais dimensões = dados mais esparsos
  - Redução de dimensionalidade pode trazer vários benefícios



Redução da dimensionalidade

- Trata-se de uma etapa importante no projeto de um sistema de classificação
  - Consiste em utilizar um número pequeno de atributos no classificador
  - Para tanto, faz-se a seleção e/ou composição de atributos mais adequados a partir dos originalmente disponíveis

Redução da dimensionalidade

- Vantagens
  - Melhora a eficácia dos classificadores
    - Elimina atributos irrelevantes ou redundantes
  - Reduz o tamanho necessário da amostra
  - Melhora a eficiência computacional dos algoritmos
    - Menos atributos envolvidos
  - Simplifica modelo gerado e facilita interpretação
  - Facilita visualização dos dados

Redução da dimensionalidade

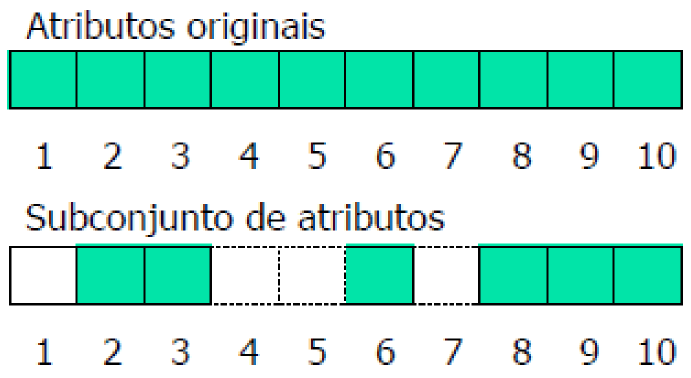
- Essencialmente, podemos reduzir a dimensionalidade de duas maneiras
  - Criação de “novos” atributos via transformação dos dados
    - Agregação de atributos
    - Extração de características
  - Seleção de atributos
    - Busca de um conjunto sub ótimo de atributos

Agregação de atributos

- Uma forma elementar de reduzir complexidade dos dados é agregar atributos
- Exemplo: dois atributos, “massa” e “volume”
  - Esses atributos podem ser agregados em um único atributo: “densidade”
    - densidade = massa / volume
  - Nesse caso, não há perda de informação relevante a um dado problema de interesse em particular

Seleção de atributos

- *Feature selection* em inglês
  - Assume que os atributos existentes já estão em uma forma apropriada.
  - No entanto
    - Alguns podem ser irrelevantes
    - Outros podem ser redundantes
  - Tais atributos podem ser descartados



Seleção de atributos

- Normalmente utiliza uma estratégia de busca que decide a maneira como as combinações de atributos são testadas de acordo com um certo critério de qualidade
  - Busca por **ordenação**
  - Seleção de **subconjunto**

Seleção de atributos

- Busca por **ordenação**
  - Ordena os atributos de acordo com sua relevância
  - Seleciona os mais relevantes segundo alguma medida
    - discriminação (para classificação)
    - prever uma saída (regressão)
  - Relevância depende da natureza do problema e dos atributos envolvidos
- Seleção de **subconjunto**
  - Seleciona um subconjunto de atributos mutuamente relevantes

Seleção de atributos

- Exemplo: busca por ordenação
  - Ordenar os atributos mais importantes para o diagnóstico de pacientes

Febre	Enjoo	Mancha	Dor	Diagnóstico
1	1	0	1	**0**
0	1	0	0	**1**
1	1	1	0	**1**
1	0	0	1	**0**
1	0	1	1	**1**
0	0	1	1	**0**

Seleção de atributos

- Exemplo: busca por ordenação
  - Atributos binários: relevância de cada atributo é estimada de acordo com o diagnóstico (exemplo apenas pedagógico)

Febre	Enjoo	Mancha	Dor
3/6	4/6	4/6	1/6

Febre	Enjoo	Mancha	Dor	Diagnóstico
1	1	0	1	**0**
0	1	0	0	**1**
1	1	1	0	**1**
1	0	0	1	**0**
1	0	1	1	**1**
0	0	1	1	**0**

Seleção de atributos

- Exemplo: busca por ordenação
  - Atributos ordenados: 2 atributos (**enjoo** e **mancha**) classificam corretamente 4/6 dos casos

Enjoo	Mancha	Febre	Dor
4/6	**4/6**	3/6	1/6

Febre	Enjoo	Mancha	Dor	Diagnóstico
1	**1**	**0**	1	**0**
0	**1**	**0**	0	**1**
1	**1**	**1**	0	**1**
1	**0**	**0**	1	**0**
1	**0**	**1**	1	**1**
0	**0**	**1**	1	**0**

Seleção de atributos

- Vantagem da busca por ordenação
  - A seleção dos atributos tem complexidade linear
    - Seleção, não a ordenação
  - Muito mais simples que combinar os atributos
    - Dado ***N*** atributos, o número de possíveis combinações de ***n*** atributos dentre ***N*** é

$$\binom{N}{n} = \frac{N!}{(N-n)!\,n!}$$

    - Para ***N = 40*** e ***n = 5***, temos ***658.008*** combinações

Seleção de atributos

- Desvantagem da busca por ordenação
  - Ordenação é deficiente: despreza correlação e redundância entre atributos
    - Atributos inúteis sozinhos porém úteis em conjunto
    - Atributos são tão úteis sozinhos quanto em conjunto
  - Nem sempre os melhores ***n*** atributos constituem o melhor subconjunto
    - Atributos devem ser não correlacionados
    - O melhor subconjunto é o mais complementar

Seleção de atributos

- Avaliar todos os subconjuntos de atributos é inviável
  - Por que não utilizar um critério de avaliação nessa busca?
- Busca heurística
  - Alguns subconjuntos são avaliados segundo algum critério até que um critério de parada seja satisfeito
  - Utiliza uma estratégia de busca para escolher os subconjuntos avaliados

Seleção de atributos

- Estratégias de Busca
  - *Backward Elimination*
    - Inicia com todos os atributos e remove um atributo por vez do conjunto
  - *Forward Selection*
    - Inicia com nenhum atributo e inclui um atributo por vez no conjunto
  - *Bidirectional Search*
    - A busca pode começar em qualquer ponto e atributos podem ser adicionados e removidos
  - *Random Search*
    - Ponto de partida da busca e atributos a serem removidos ou adicionados são decididos de forma estocástica

Seleção de atributos

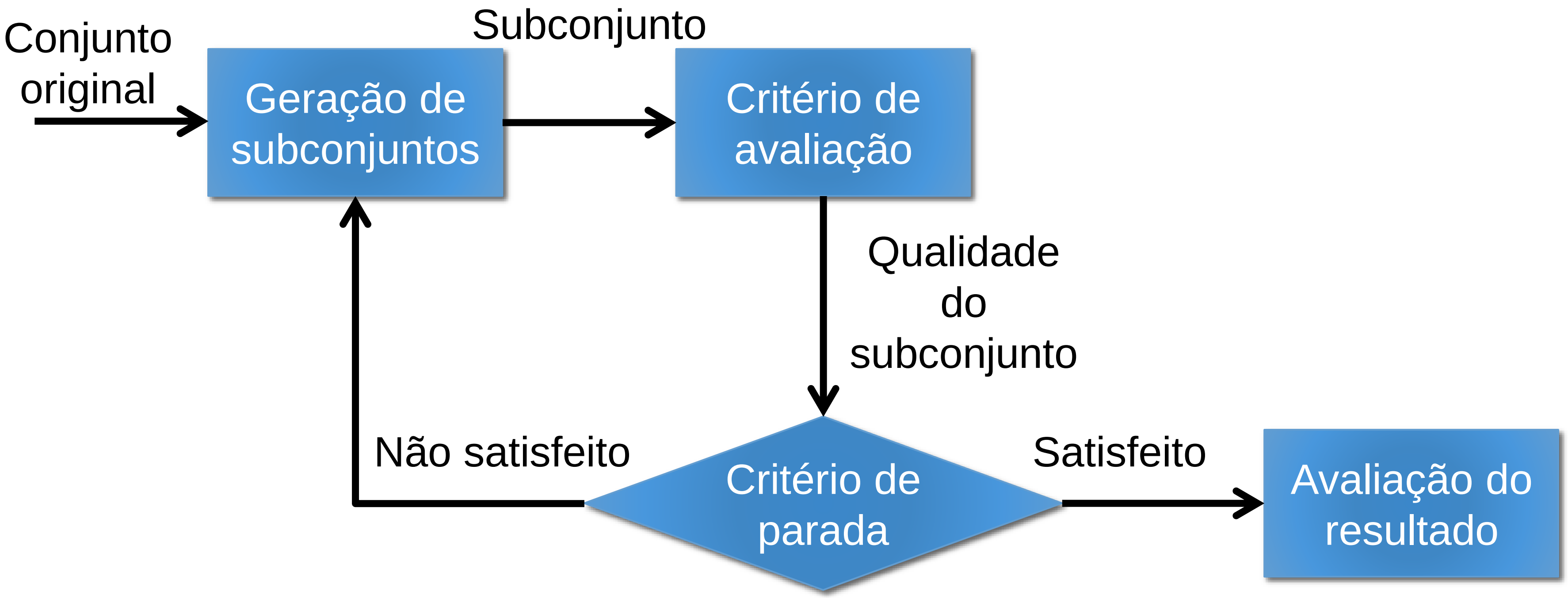
- Critérios de Avaliação
  - Inerente ao método de seleção de atributos
  - Critérios independentes
    - Medidas de correlação
    - Medidas de informação
    - Medidas de dependência
    - Medidas de consistência
  - Critérios dependentes
    - Algoritmo alvo usado para a tarefa de interesse

Seleção de atributos

- Critério de Parada
  - De modo geral, depende do método de busca utilizado
  - Algumas possibilidades
    - Número máximo de iterações
    - Valor do critério de avaliação obtido
    - Etc.

Seleção de atributos

- Visão geral do processo de seleção de subconjunto

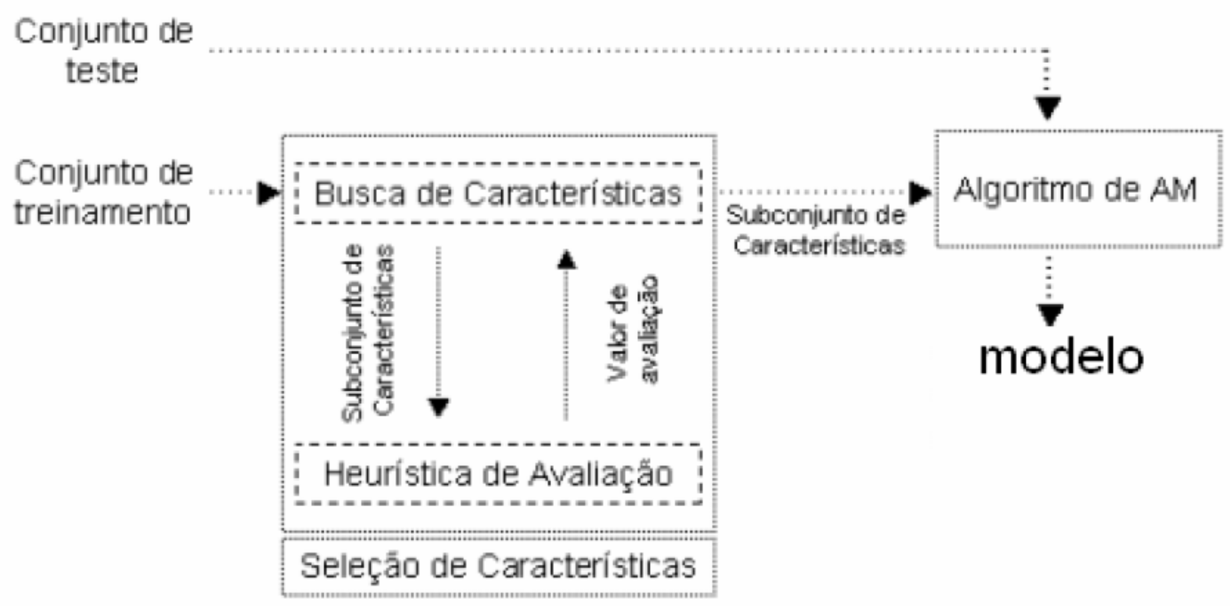


Seleção de atributos

- Existem diferentes maneiras de se fazer a seleção de atributos. Elas podem ser agrupadas em 3 categorias independentes
  - Filtros
    - Seleção de atributos é realizada a priori
  - *Wrappers*
    - O algoritmo de aprendizado é usado para guiar o processo de seleção
  - Embarcados (*Embedded*)
    - Processo de seleção faz parte do algoritmo de aprendizado

Seleção de atributos

- Filtros
  - Seleção de atributos é realizada a priori
    - Basicamente, fazem uso de alguma heurística para executar uma busca nos atributos
    - Considera apenas as propriedades intrínsecas aos próprios dados
    - Processamento mais rápido



Seleção de atributos

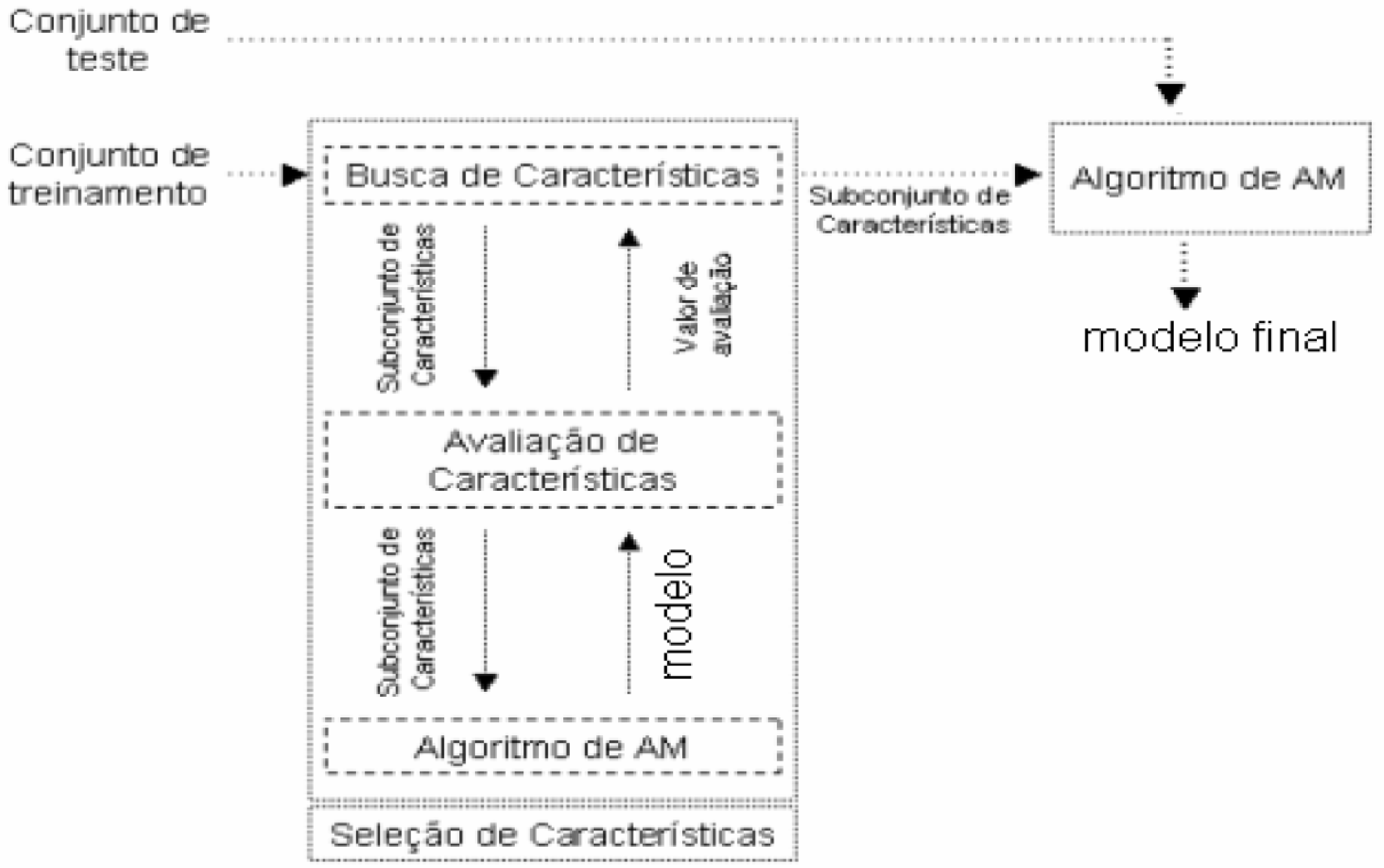
- Filtros
  - Critérios de busca
    - Medidas de correlação / informação mútua entre atributos
    - Medidas de relevância e redundância
    - Privilegiam conjuntos de atributos muito relacionados com a saída desejada e pouco relacionados entre si
  - Desvantagem
    - Seleção de forma indireta, o que pode levar a resultados inferiores

Seleção de atributos

- *Wrappers*
  - O algoritmo de aprendizado utilizado é usado para guiar o processo de seleção
    - Utilizam alguma heurística para executar uma busca
    - Uso do algoritmo de aprendizado: maximização do seu desempenho
  - Implica, em geral, em tornar o método muito custoso em termos computacionais
    - Custo pode se tornar proibitivo

Seleção de atributos

- *Wrappers*



Seleção de atributos

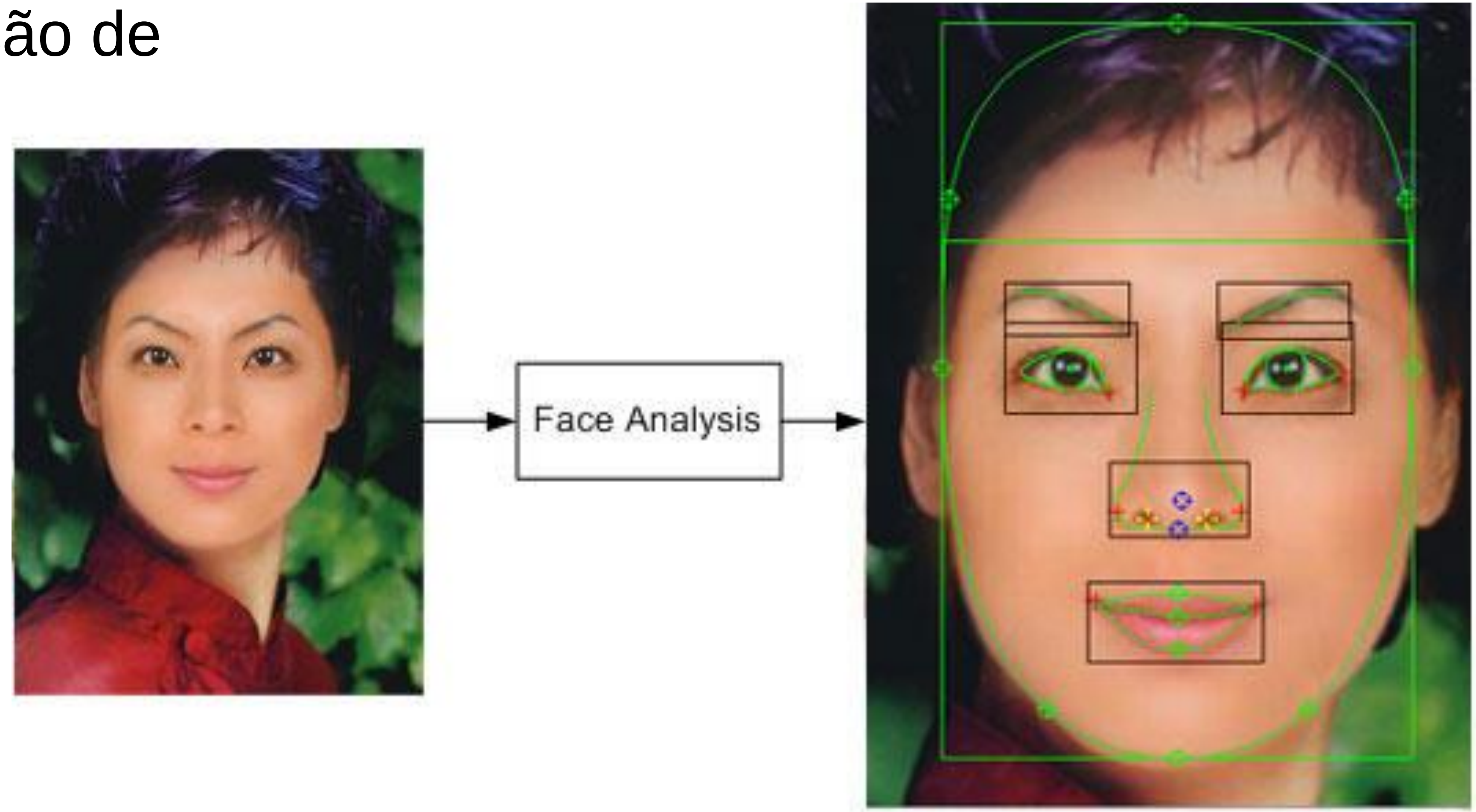
- Embarcados (*Embedded*)
  - Processo de seleção faz parte do algoritmo de aprendizado
    - Parte interna e natural do algoritmo de aprendizado
  - Exemplo
    - Classificadores baseados em Árvores de Decisão

Extração de características

- *Feature extraction* em inglês
  - Consiste em extrair, a partir dos dados brutos, características de alto nível com grande riqueza de informação relevante sobre os dados
  - Exemplo:
    - Informações sobre bordas, contornos, sombras e formas geométricas em fotografias (pixels não são bons atributos)
    - Componentes harmônicas de frequência em sinais de áudio

Extração de características

- Exemplo de extração de características



Extração de características

- Um exemplo é a *Transformação do Espaço de Atributos*
  - Gera um novo conjunto de atributos a partir da combinação de projeções dos atributos originais
  - Ex.: PCA (linear) ou Kernel PCA (não linear)

- Atributos são ortogonais (perpendiculares) e ordenados segundo a parcela de informação que conduzem
  - Podemos descartar os atributos menos representativos
  - Resultado é um espaço de dimensão menor que o original contendo a maior parcela possível da informação

Extração de características

- Vantagens
  - Simples e computacionalmente rápida em especial PCA linear
- Desvantagens
  - Técnica limitada a atributos numéricos
  - Novos atributos não podem ser interpretados como os originais
    - Atributos físicos deixam de ter um significado físico
    - Muito ruim para determinadas aplicações

Análise de Componentes Principais - PCA

- *Principal Component Analysis* em inglês
  - Forma de identificar padrões nos dados
    - Colocando em evidência suas relações, similaridades e diferenças
  - Especialmente importante para altas dimensões
    - Análise visual não é possível
  - Extrator de características
    - Uma vez encontrados os padrões, podemos comprimir os dados sem grande perda de qualidade

Análise de Componentes Principais - PCA

- Histórico
  - Pearson (1901)
    - Criou a Componente Principal (PC)
    - Procurava linhas e planos que melhor se adequavam a um conjunto de pontos em um espaço *p*-dimensional
  - Hotelling (1933)
    - Procurava encontrar um pequeno conjunto de variáveis fundamentais que expressa ***p*** variáveis
    - Hotelling procurou maximizar suas “componentes” no senso da variância das variáveis originais. Chamou de Componentes Principais

Análise de Componentes Principais - PCA

- Histórico
  - Pearson e Hotelling esbarraram no cálculo dos autovetores
    - Difícil de calcular para ordem > 4
    - PCA é mais eficiente para conjuntos de dados de alta dimensão. Sem aplicação na época
  - Retomada nos anos 60
    - Primeiros computadores capazes de resolver o problema dos autovetores de maneira rápida

Análise de Componentes Principais - PCA

- Idéia básica
  - Um número ***p*** de atributos dependentes podem ser expressas como um número ***t*** de atributos independentes
    - Sendo ***t* << *p***
  - Considere um conjunto de vetores ***x***
    - Pode-se sempre gerar uma combinação linear que mapeia o vetor ***x*** no vetor ***y***
    - Espaço definido por variáveis ortonormais (norma igual a 1)

Análise de Componentes Principais - PCA

- Combinação linear de ***x*** em ***y***
  - Transformação sem perda de informação

$$x = \sum_{j=1}^{\infty} y_j e_j$$

  - Considerando apenas ***t*** dimensões
    - Nesse caso, teremos alguma perda de informação

$$\hat{x} = \sum_{j=1}^{t} y_j e_j$$

Análise de Componentes Principais - PCA

- Definição matemática
  - Transformação linear ortogonal dos dados
  - Dados agrupados da seguinte forma
    - A maior variância por qualquer projeção dos dados fica ao longo da primeira coordenada (**primeiro componente**)
    - A segunda maior variância fica ao longo da segunda coordenada (**segundo componente**)
    - E assim por diante

Análise de Componentes Principais - PCA

- Etapas para o cálculo do PCA
  - Transformação dos dados envolve conceitos matemáticos relativamente simples
    - Subtrair a média dos dados (para cada atributo)
    - Calcular a matriz de covariâncias
    - Cálculo dos autovetores e autovalores da matriz de covariâncias
    - Ordenação dos autovetores por ordem de importância
    - Mapear os dados para o novo espaço

Análise de Componentes Principais - PCA

- Autovetores e autovalores
  - Dado um vetor ***v*** e uma matriz de transformação ***M,*** temos que ***v*** é um **autovetor** de ***M*** se
    - ***Mv*** (multiplicação da matriz ***M*** pelo vetor ***v***) resulta num múltiplo de ***v***, ou seja, em λv (multiplicação de um escalar pelo vetor)
    - Nesse caso, λ é o chamado **autovalor** de ***M*** associado ao vetor ***v***

Análise de Componentes Principais - PCA

- Autovetores e autovalores

a matriz

$$A = \begin{bmatrix} 8 & 1 & 2 \\ -1 & 5 & 1 \\ 0 & 1 & 90 \end{bmatrix},$$

tem como autovalores e respectivos autovetores,

$$\lambda_1 = 90,0115, \quad \lambda_2 = 7,6308, \quad \lambda_3 = 5,3577$$

$$x_1 = \begin{bmatrix} 0,0245 \\ 0,0115 \\ 0,9996 \end{bmatrix}, \quad x_2 = \begin{bmatrix} -0,9353 \\ 0,3539 \\ -0,0043 \end{bmatrix}, \quad x_3 = \begin{bmatrix} -0,0043 \\ -0,0111 \\ 0,9996 \end{bmatrix}$$

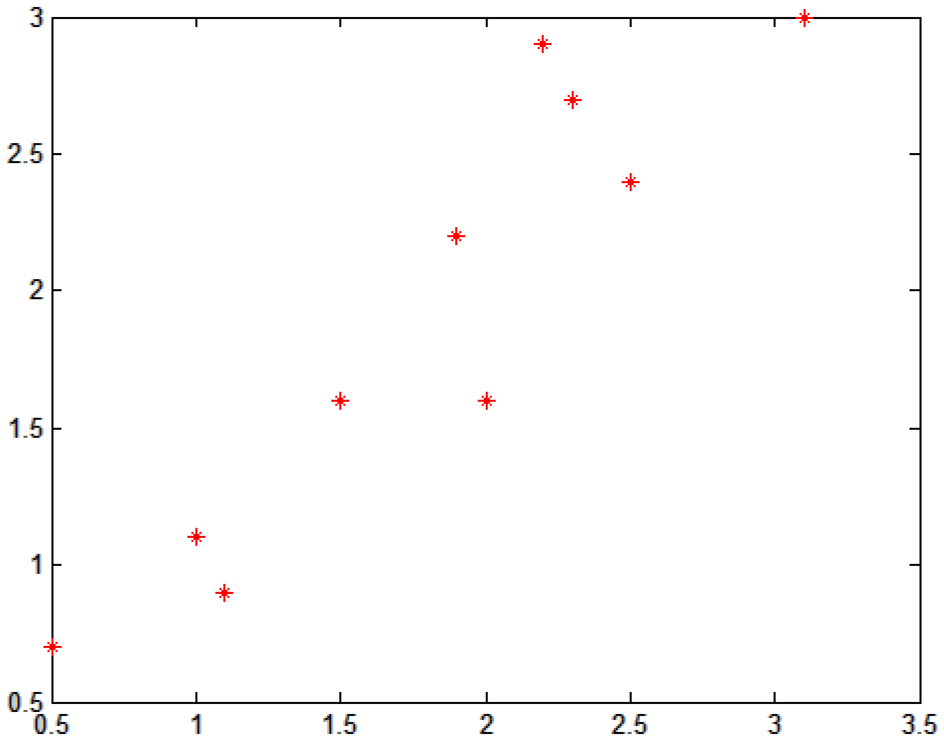
Análise de Componentes Principais - PCA

- Autovetores e autovalores
  - Propriedades
    - A matriz de transformação ***M*** deve ser quadrada
    - Nem todas as matrizes possuem autovetores
    - Para uma matriz ***n x n*** existem ***n*** autovetores

Análise de Componentes Principais - PCA

- Vamos calcular o PCA para o seguinte conjunto de dados

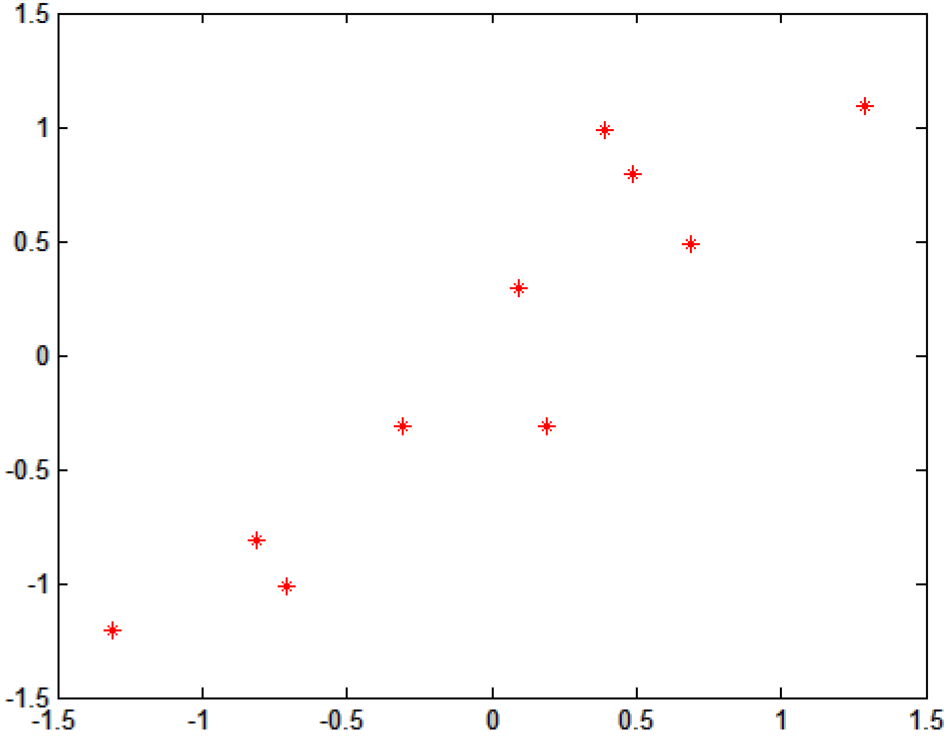
x	y
2,5	2,4
0,5	0,7
2,2	2,9
1,9	2,2
3,1	3
2,3	2,7
2	1,6
1	1,1
1,5	1,6
1,1	0,9



Análise de Componentes Principais - PCA

- O primeiro passo é subtrair a média dos dados
  - Não fazer o zscore (precisamos da variância!)

x	y
0,69	0,49
-1,31	-1,21
0,39	0,99
0,09	0,29
1,29	1,09
0,49	0,79
0,19	-0,31
-0,81	-0,81
-0,31	-0,31
-0,71	-1,01



Análise de Componentes Principais - PCA

- Na sequência, obtemos a matriz de covariância dos dados

x	y
0,69	0,49
-1,31	-1,21
0,39	0,99
0,09	0,29
1,29	1,09
0,49	0,79
0,19	-0,31
-0,81	-0,81
-0,31	-0,31
-0,71	-1,01

→

0,6166	0,6154
0,6154	0,7166

Análise de Componentes Principais - PCA

- A partir da matriz de covariância, obtemos os seus autovetores e autovalores

0,6166	0,6154
0,6154	0,7166

autovalores

0,0491	1,2840

-0,7352	0,6779
0,6779	0,7352

autovetores

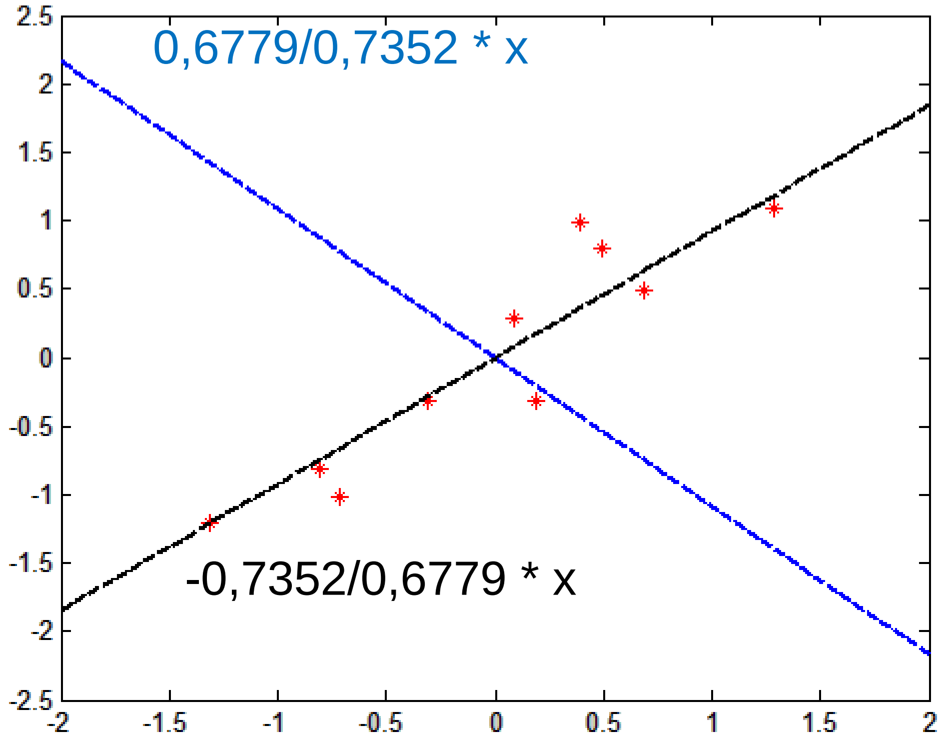
Análise de Componentes Principais - PCA

- Autovetores nos fornecem informações sobre os padrões nos dados
  - Um deles passa pelo meio dos pontos (quase uma regressão)

autovalores

0,0491	1,2840
-0,7352	0,6779
0,6779	0,7352

autovetores



Análise de Componentes Principais - PCA

- Temos que o autovetor com o maior autovalor é o componente principal do conjunto dos dados
  - Ordenar do maior para o menor

autovalores

0,0491	1,2840

-0,7352	0,6779
0,6779	0,7352

autovetores

→

autovalores

1,2840	0,0491

0,6779	-0,7352
0,7352	0,6779

autovetores

Análise de Componentes Principais - PCA

- Uma vez ordenados, podemos escolher os componentes que nos interessam
  - Podemos escolher todos
  - Podemos descartar os menos significantes
    - Reduzindo assim a dimensionalidade dos dados

autovalores

1,2840	0,0491

0,6779	-0,7352
0,7352	0,6779

autovetores

OU

autovalores

1,2840

0,6779
0,7352

autovetores

Análise de Componentes Principais - PCA

- Para obter os dados transformados pelo PCA
  - Multiplicar os dados (com a média subtraída deles) pelos autovetores escolhidos
    - Dados transformados expressam os padrões entre eles
    - Os *Componentes Principais* são combinações lineares de todo os atributos, produzindo assim novos atributos não correlacionados

Análise de Componentes Principais - PCA

- Obtendo os dados transformados

x	y
0,69	0,49
-1,31	-1,21
0,39	0,99
0,09	0,29
1,29	1,09
0,49	0,79
0,19	-0,31
-0,81	-0,81
-0,31	-0,31
-0,71	-1,01

x

0,6779	-0,7352
0,7352	0,6779

=

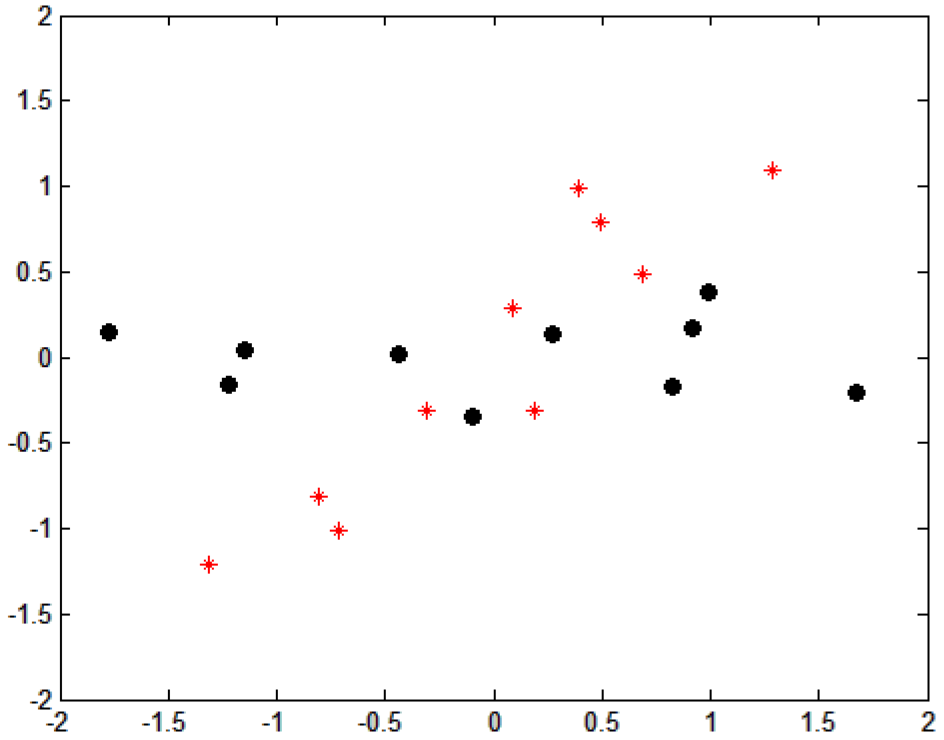
x	y
0,82	-0,17
-1,77	0,14
0,99	0,38
0,27	0,13
1,67	-0,20
0,91	0,17
-0,09	-0,34
-1,14	0,04
-0,43	0,01
-1,22	-0,16

Análise de Componentes Principais - PCA

- Obtendo os dados transformados

Dados sem média

x	y
0,69	0,49
-1,31	-1,21
0,39	0,99
0,09	0,29
1,29	1,09
0,49	0,79
0,19	-0,31
-0,81	-0,81
-0,31	-0,31
-0,71	-1,01

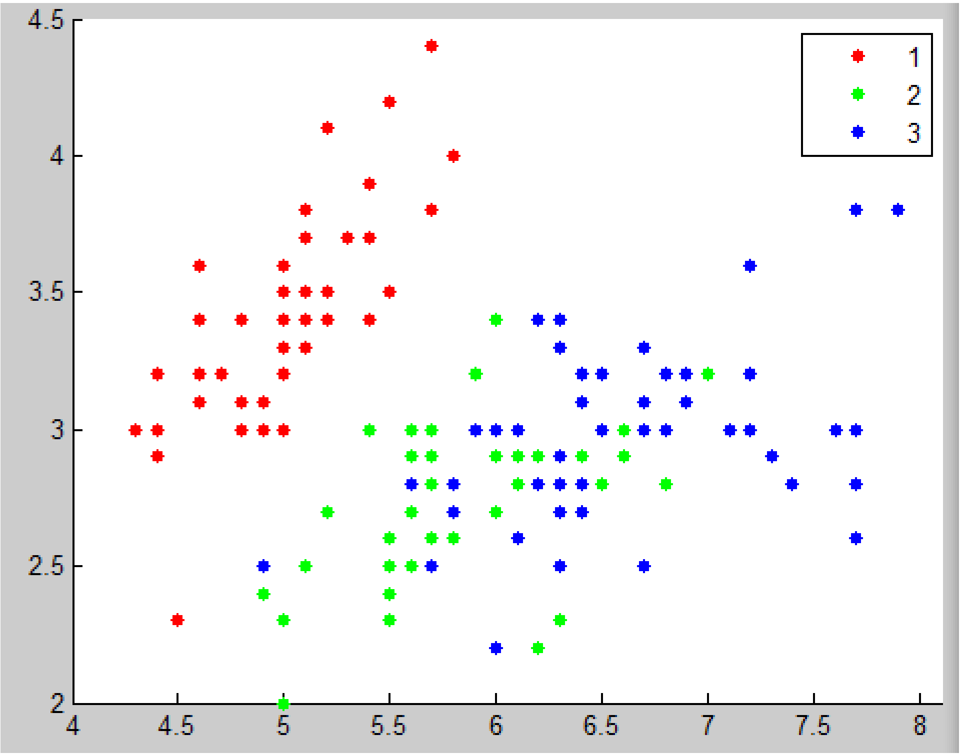


PCA

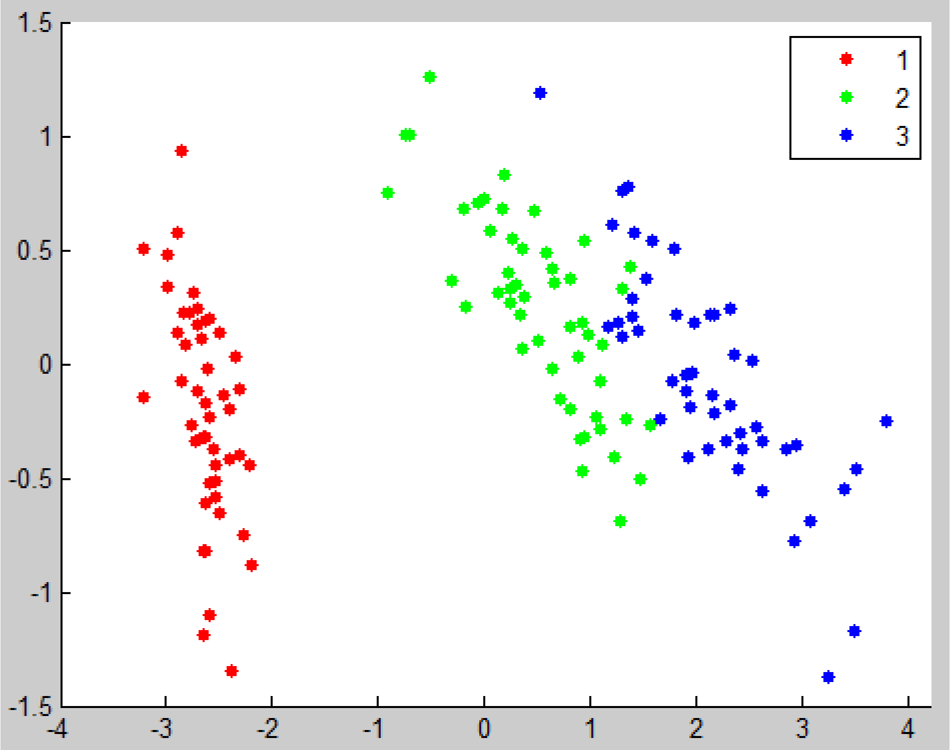
X	y
0,82	-0,17
-1,77	0,14
0,99	0,38
0,27	0,13
1,67	-0,20
0,91	0,17
-0,09	-0,34
-1,14	0,04
-0,43	0,01
-1,22	-0,16

PCA - Iris

Sem PCA



Com PCA



PCA - Iris

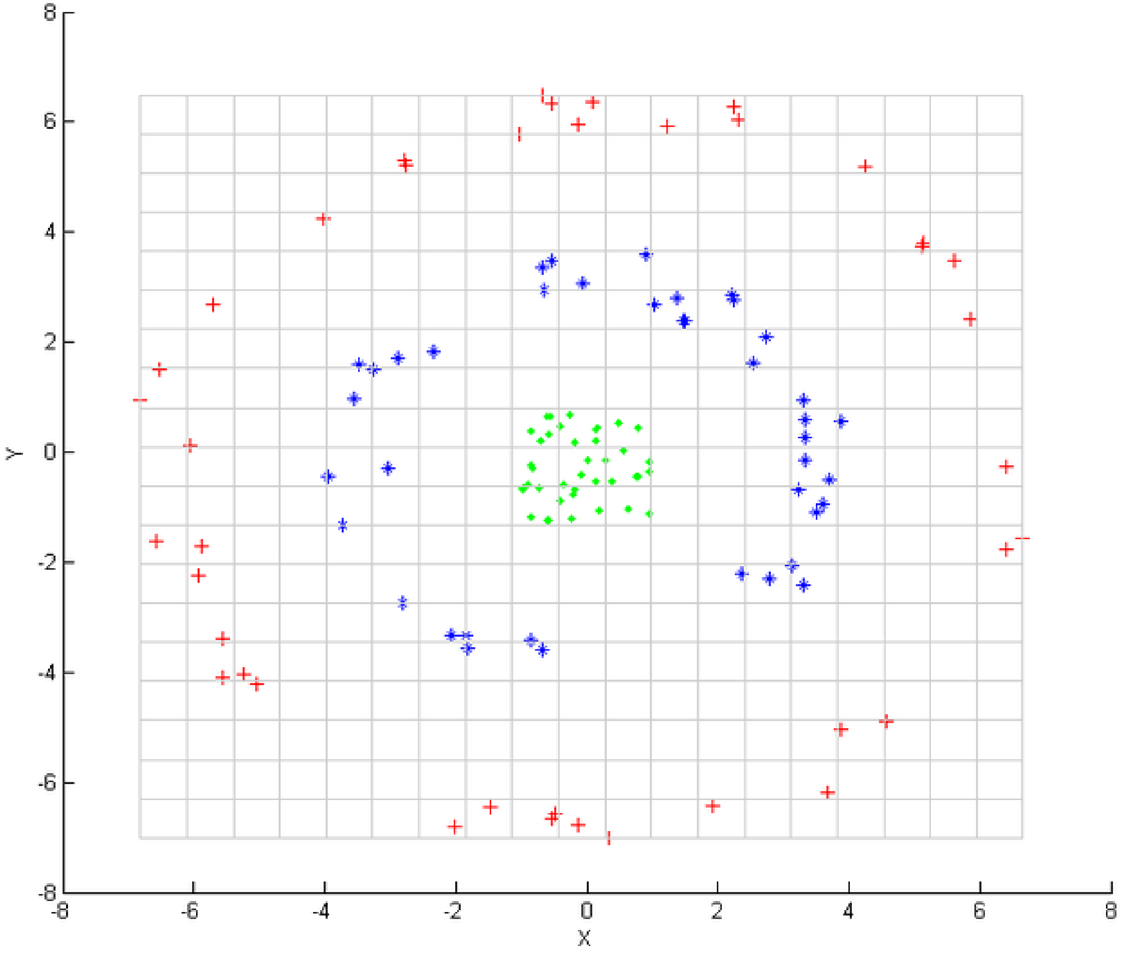
- Classificação com Knn (k = 1)
  - Sem PCA
    - 4 atributos: 94,67%
  - Com PCA
    - 1 componente: 88,67%
    - 2 componentes: 94,00%
    - 3 componentes: 90,67%
    - 4 componentes: 90,67%
  - Redução do conjunto de atributos pela **metade** com perda de apenas **0,67%**

Análise de Componentes Principais - PCA

- Problemas
  - Voltado apenas para atributos numéricos
    - Não há sentido em trabalhar com atributos discretos, mesmo depois de uma etapa de conversão
  - Caso os $\boldsymbol{p}$ atributos não tenham as mesmas unidades de medida, a combinação linear é insensata do ponto de vista “físico”
  - Só é possível extrair uma projeção linear dos dados

Análise de Componentes Principais - PCA

- PCA = projeção linear dos dados
  - Para certos conjuntos de dados isso não funciona muito bem



Kernel PCA

- Solução
  - Encontrar uma transformação não linear, isto é, um ***Kernel***
    - Essa transformação mapeia o espaço original dos padrões para um novo espaço de atributos
    - Nesse novo espaço, os padrões ***x*** passam a ser linearmente separáveis

Kernel PCA

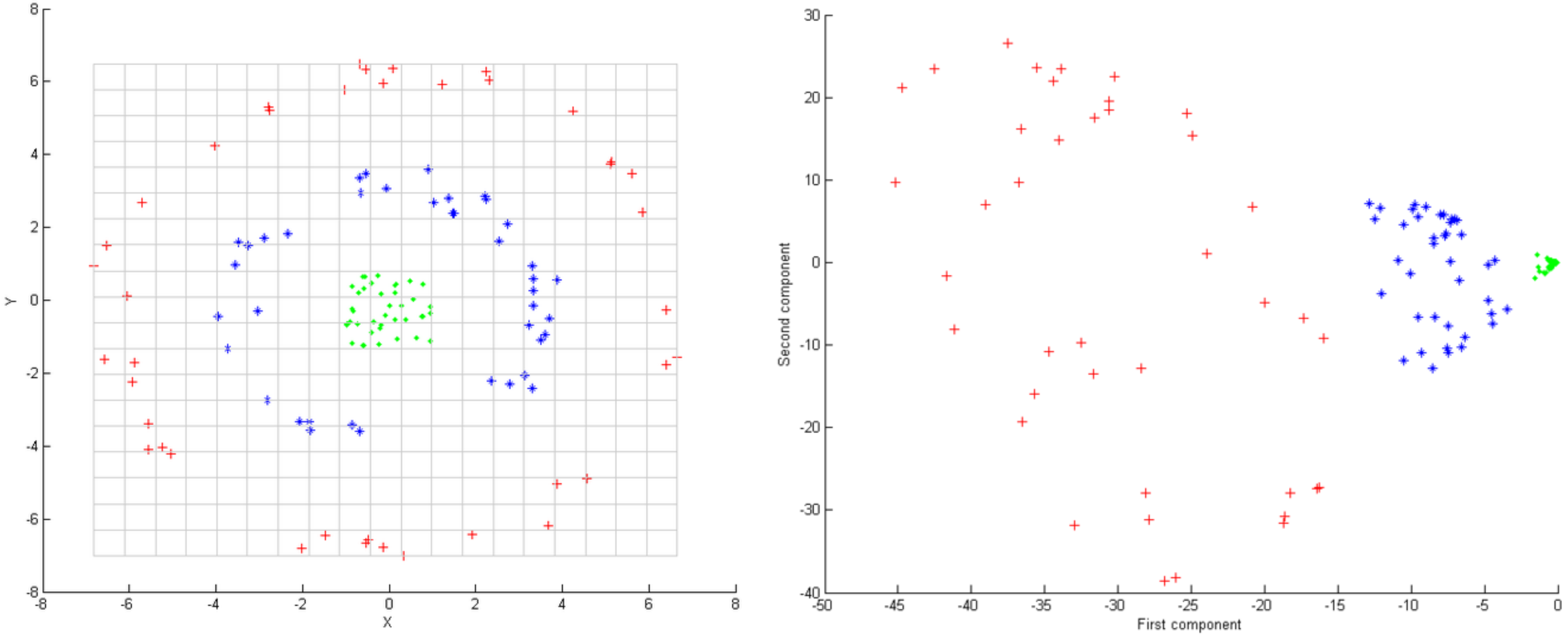
- Idéia básica
  - Utilizar uma função Kernel não linear de forma a calcular o PCA em um espaço de maior dimensão
  - Esse espaço é não linearmente relacionado ao espaço original

Kernel PCA

- Possível solução
  - Projetar os dados em um espaço de maior dimensão
  - Subtrair a média dos dados transformados (para cada atributo)
  - Calcular a matriz de covariâncias
  - Cálculo dos autovetores e autovalores da matriz de covariâncias
  - Ordenação dos autovetores por ordem de importância
  - Mapear os dados para o novo espaço

Kernel PCA

- Felizmente, Kernel PCA pode ser calculado de forma implícita
  - Sem necessidade de transformação dos dados



Análise de Componentes Independentes - ICA

- *Independent Component Analysis* em inglês
  - É uma extensão da abordagem do PCA
    - Trata-se de um método computacional para a separação de um conjunto de dados em subcomponentes aditivos
    - Supõe a independência estatística ao invés da descorrelação dos dados

Análise de Componentes Independentes - ICA

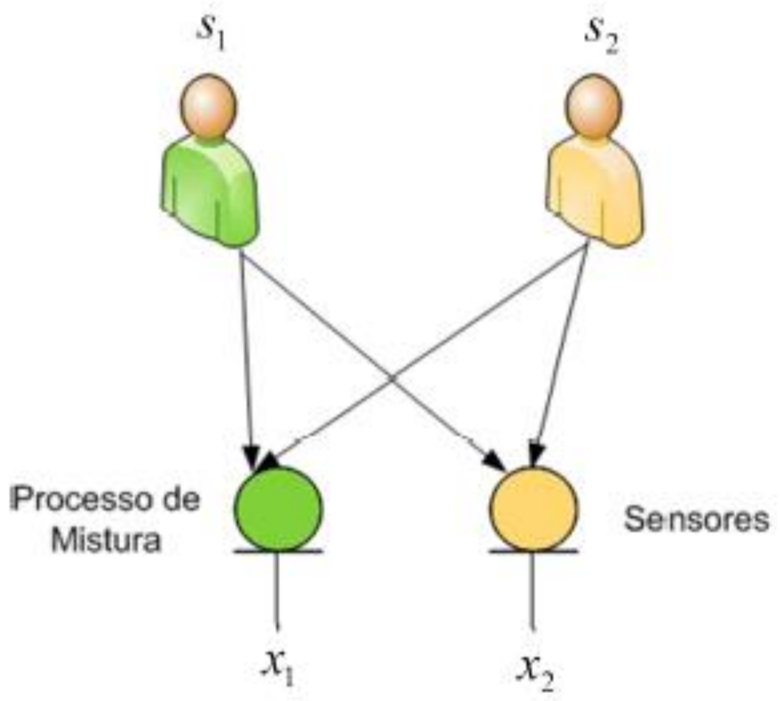
- *Descorrelação* versus *independência*
  - PCA – descorrelação dos dados
    - Se dois atributos são descorrelacionadas sua covariância é zero
    - Trabalha com média nula, o que leva a condição de ortogonalidade (perpendicularidade) da construção das direções de projeção dos componentes principais
    - Com isso, tem-se componentes de máxima variância
    - Descorrelação ***linear*** não implica na ocorrência de descorrelação ***não linear***

Análise de Componentes Independentes - ICA

- *Descorrelação* versus *independência*
  - ICA – independência dos dados
    - Independência estatística acarreta toda e qualquer descorrelação ***não linear***
    - Componentes independentes: componentes ***linear*** e ***não linearmente*** descorrelacionados
    - Preço disso tudo: para quantificar essa independência

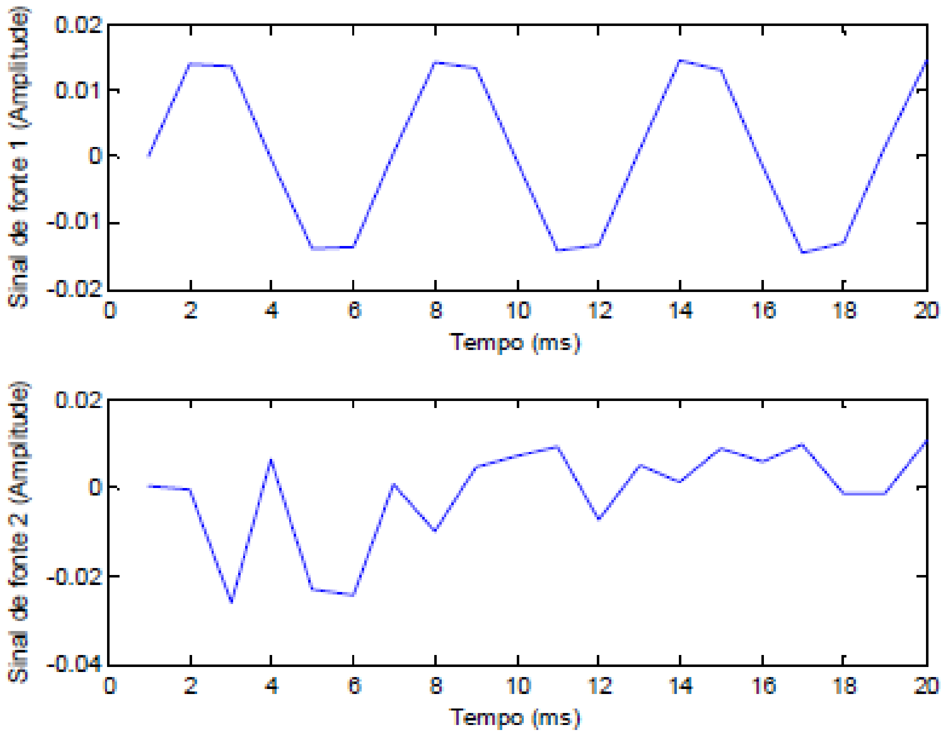
Análise de Componentes Independentes - ICA

- Motivação: separação cega de fontes
  - Problema “*cocktail party*”
    - Separação de sinais de áudio
    - Duas pessoas conversando em uma sala fechada utilizando sensores (microfones) para capturar suas vozes
    - Como separar os sinais captados pelos microfones sabendo que os sinais estão agora correlacionados?

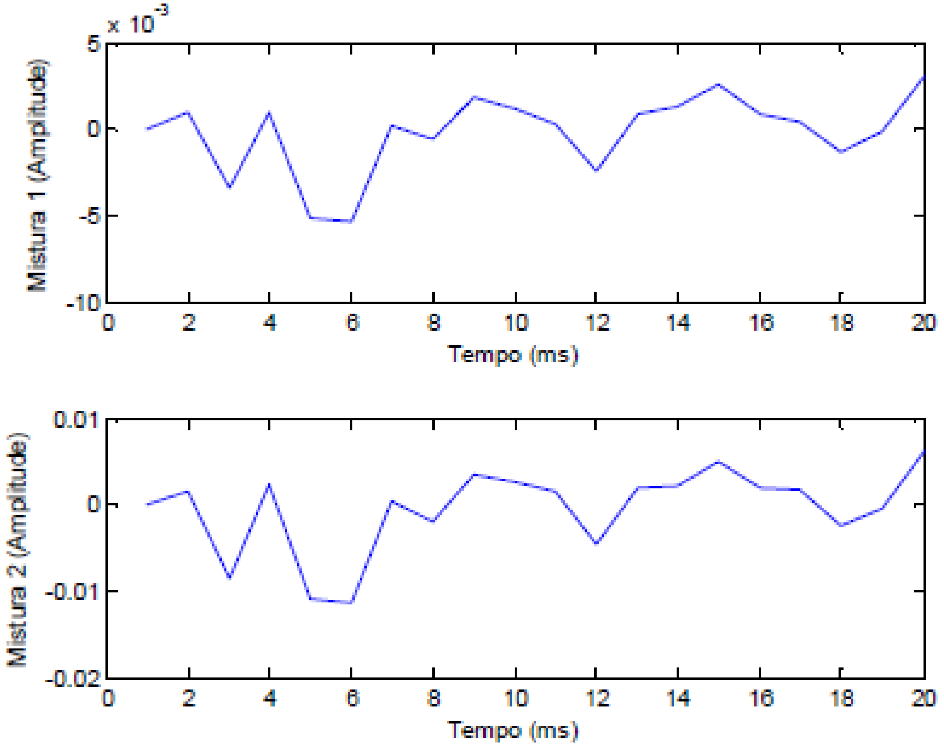


Análise de Componentes Independentes - ICA

Discursos originais



Discursos misturados



Análise de Componentes Independentes - ICA

- Modelo de mistura
  - Os dados observados ***x*** consistem de uma combinação linear de ***n*** atributos estatisticamente independentes, ***s***

$$x(i) = a_1(i)s_1(i) + a_2(i)s_2(i) + \ldots + a_n(i)s_n(i)$$

  - Em forma matricial

$$x = As$$

  - Onde ***A*** são os coeficientes de misturas

Análise de Componentes Independentes - ICA

- Modelo de mistura
  - Os componentes independentes podem ser obtidos pela inversa de ***A***, ***W***

$$s = Wx$$

  - Problema
    - A matriz ***A*** é, em geral, desconhecida
    - Porém, podemos fazer uma boa estimativa dela

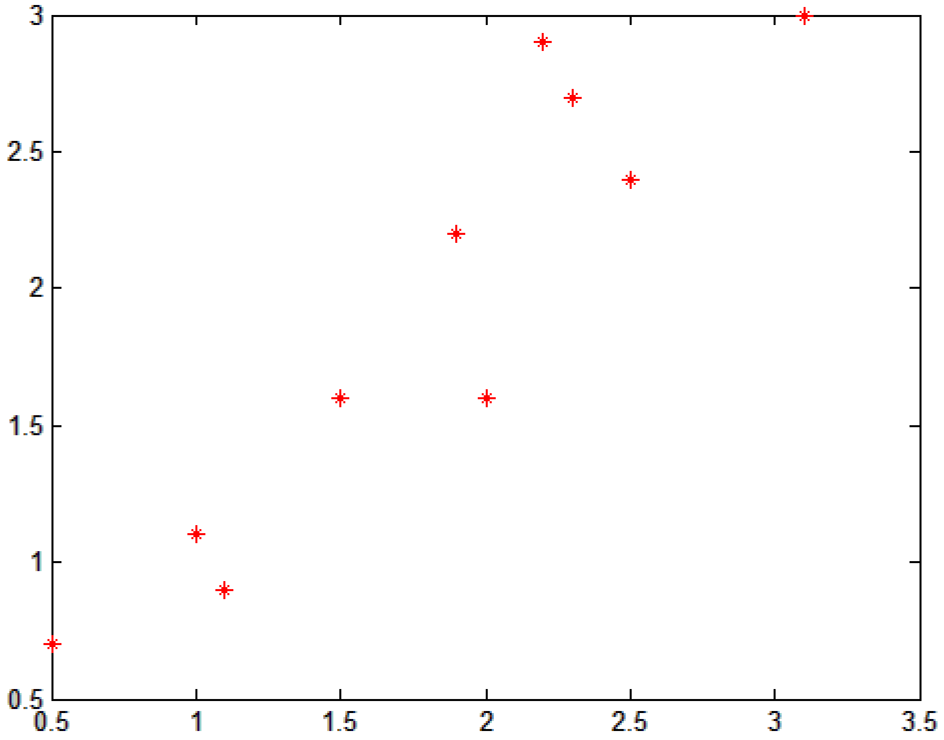
Análise de Componentes Independentes - ICA

- Etapas para o cálculo do ICA
  - Transformação dos dados envolve conceitos matemáticos relativamente simples
    - Subtrair a média dos dados (para cada atributo)
    - Branqueamento ou *whitening*
    - Cálculo da matriz de mistura ortogonal
    - Mapear os dados para o novo espaço

Análise de Componentes Independentes - ICA

- Vamos calcular o ICA para o seguinte conjunto de dados

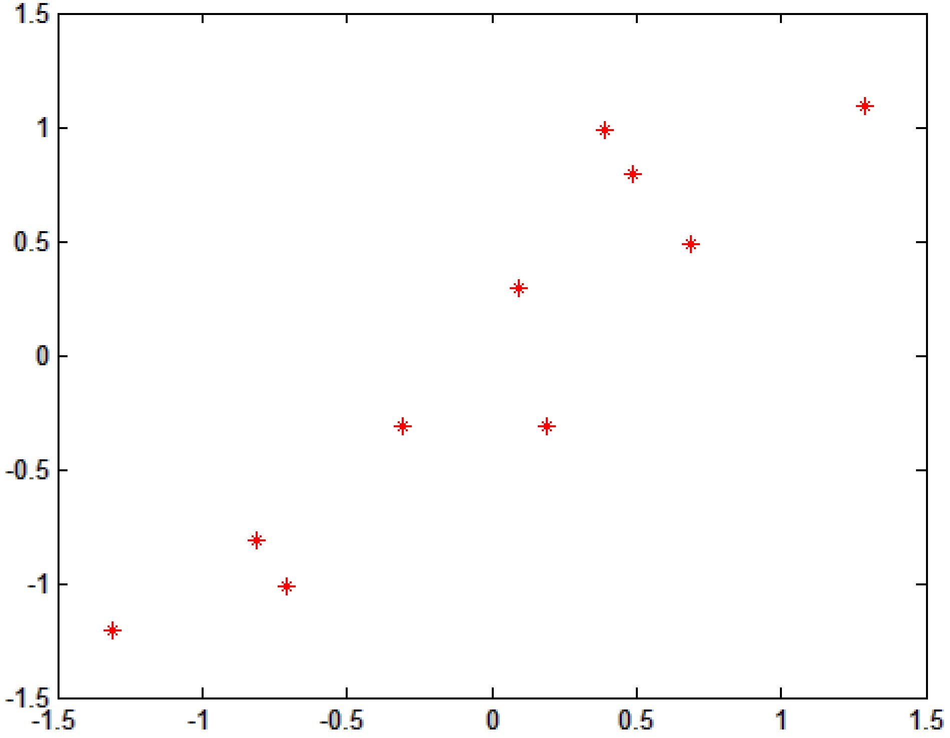
x	y
2,5	2,4
0,5	0,7
2,2	2,9
1,9	2,2
3,1	3
2,3	2,7
2	1,6
1	1,1
1,5	1,6
1,1	0,9



Análise de Componentes Independentes - ICA

- O primeiro passo é subtrair a média dos dados
  - Não fazer o zscore (precisamos da variância!)

x	y
0,69	0,49
-1,31	-1,21
0,39	0,99
0,09	0,29
1,29	1,09
0,49	0,79
0,19	-0,31
-0,81	-0,81
-0,31	-0,31
-0,71	-1,01



Análise de Componentes Independentes - ICA

- Branqueamento ou *whitening*
  - Dado uma amostra ***x*** centralizada (média zero), esse processo torna os atributos descorrelacionados e com variância igual a 1
    - Sua matriz de correlação fica igual a matriz identidade

Análise de Componentes Independentes - ICA

- Branqueamento ou *whitening*
  - Esse processo é obtido com a seguinte transformação linear
    - $z = xV$
  - Onde $V = ED^{-\frac{1}{2}}E^t$
    - ***E*** é a matriz ortogonal dos autovetores da matriz de covariância
    - ***D*** é a matriz diagonal dos autovalores da matriz de covariância

Análise de Componentes Independentes - ICA

- Branqueamento ou *whitening*
  - Obtendo os dados transformados

x	y
0,69	0,49
-1,31	-1,21
0,39	0,99
0,09	0,29
1,29	1,09
0,49	0,79
0,19	-0,31
-0,81	-0,81
-0,31	-0,31
-0,71	-1,01

x

2,8451	-1,8096
-1,8096	2,5511

=

x	y
1,0764	0,0013
-1,5374	-0,7161
-0,6819	1,8198
-0,2687	0,5769
1,6976	0,4462
-0,0355	1,1286
1,1015	-1,1346
-0,8387	-0,6005
-0,3210	-0,2298
-0,1923	-1,2917

Análise de Componentes Independentes - ICA

- Matriz de mistura ortogonal
  - A partir dos dados “branqueados” podemos obter a matriz de misturas que dá origem aos componentes independentes ***s***

$$s = Wx$$

  - Existem várias abordagens para se obter essa matriz
    - Maximização da Não Gaussianidade (kurtosis)
      - **Usando PCA: *P-ICA***
    - Estimativa da Máxima Probabilidade
    - Minimização da Informação Mútua
    - Métodos Tensoriais
    - Entre outros

Análise de Componentes Independentes - ICA

- Usando PCA: *P-ICA*
  - PCA e ICA
    - Transformação linear dos dados
    - Exploram os dados de formas diferentes
  - PCA
    - Utiliza a distribuição conjunta gaussiana para ajustar os dados
    - Busca uma transformação ortogonal que faz a distribuição conjunta gaussiana fatorável independente da verdadeira distribuição dos dados

Análise de Componentes Independentes - ICA

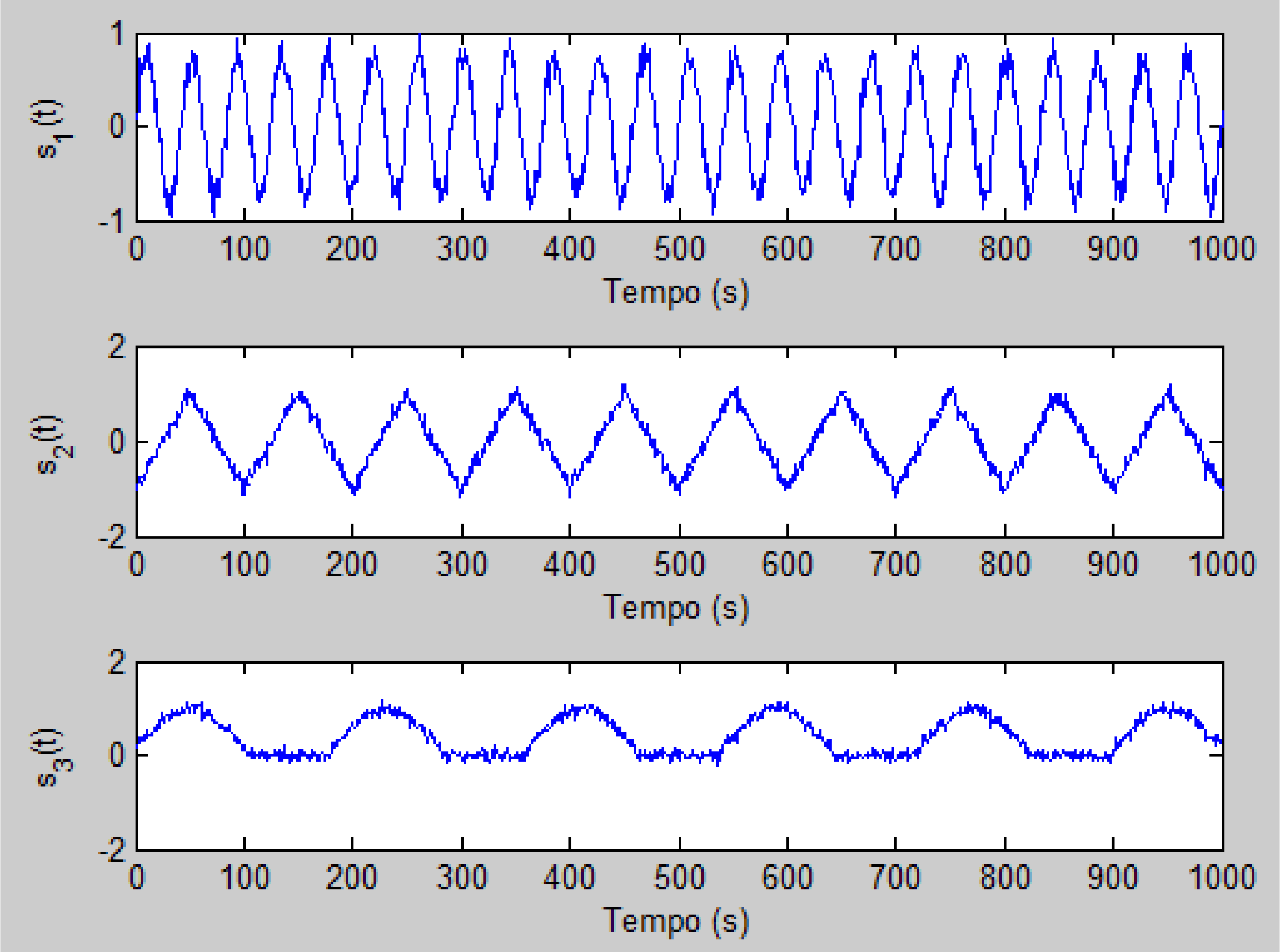
- Usando PCA: *P-ICA*
  - ICA
    - Busca uma transformação linear que faz a verdadeira distribuição conjunta dos dados transformados fatorável, de modo que as saídas são mutuamente independentes.

Análise de Componentes Independentes - ICA

- Usando PCA: *P-ICA*
  - Como fazer?
    - Branqueamento do conjunto $\boldsymbol{x}$ de dados: $\boldsymbol{v}$
    - Transformação $\boldsymbol{z} = \|\boldsymbol{v}\|\boldsymbol{v}$
    - Obter a matriz ortogonal $\boldsymbol{U}$ usando PCA em $\boldsymbol{z}$
    - A matriz de separação é dada por $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{V}$

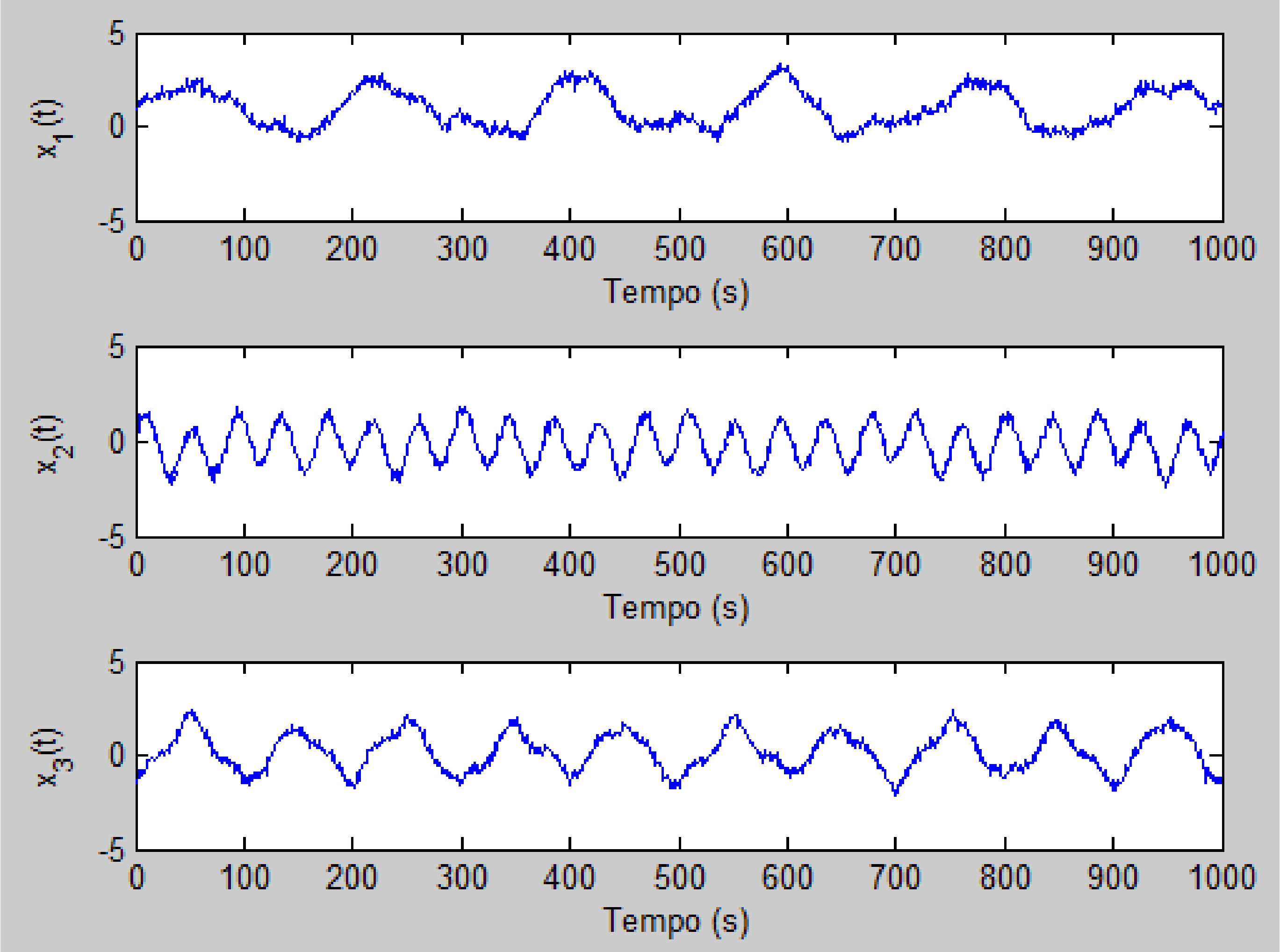
Análise de Componentes Independentes - ICA

- Exemplo
  - Dados originais



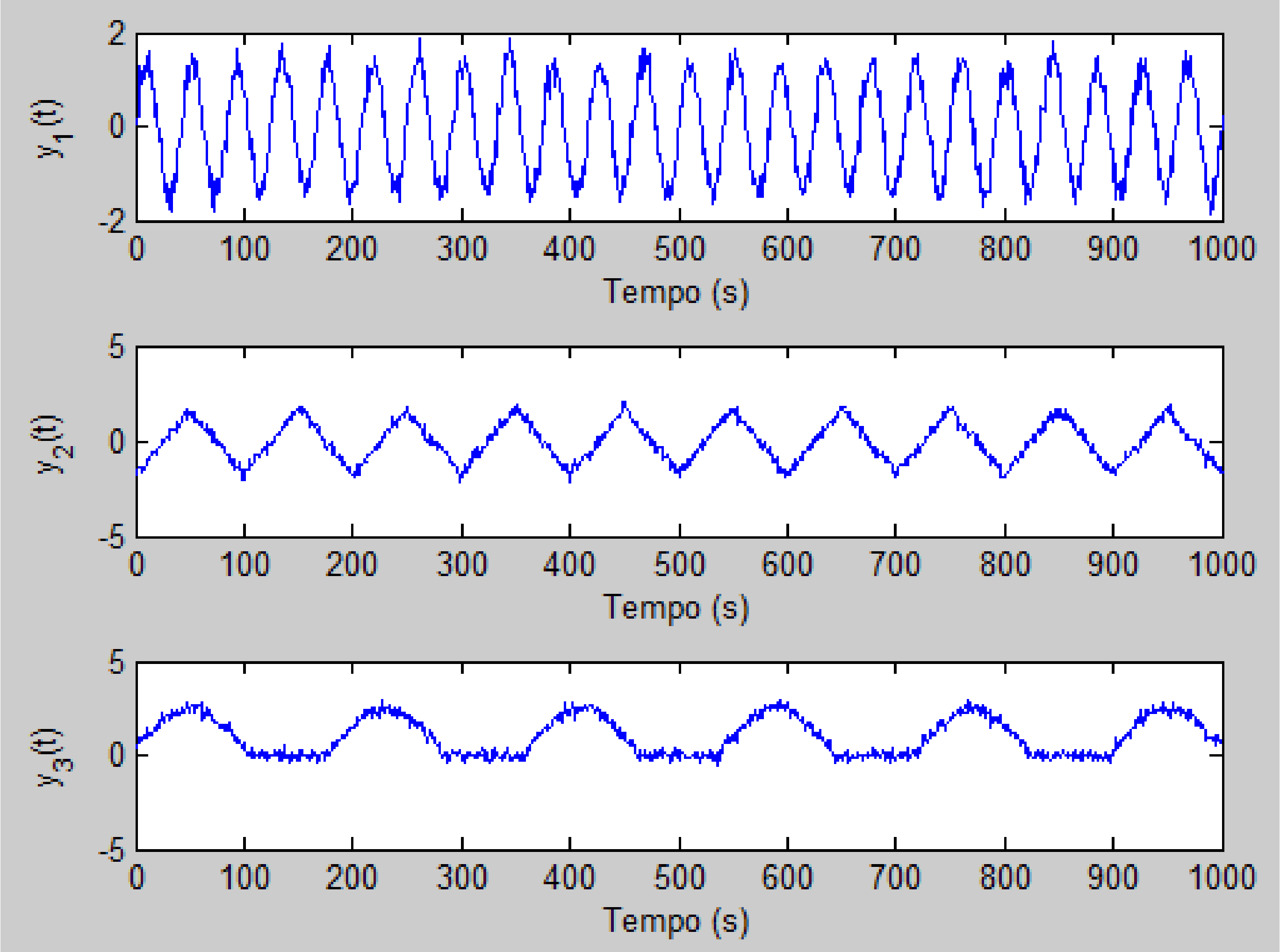
Análise de Componentes Independentes - ICA

- Exemplo
  - Dados misturados



Análise de Componentes Independentes - ICA

- Exemplo
  - Dados separados



Agradecimentos

- Agradeço ao professor
  - Prof. Ricardo J. G. B. Campello – ICMC/USP
- E ao doutorando
  - Nielsen Castelo Damasceno - UFRN
- pelo material disponibilizado