

EXTREME DATA MINING: THE KILLER APP FOR META-LEARNING?

Carlos Soares
csoares@liacc.up.pt



DI, U. Minho

Jan. 07, Braga

Plano

- Motivação
 - *Extreme Data Mining*
 - necessidade de novas metodologias de data mining
 - problema concreto
- Revisões sobre aprendizagem automática
- Meta-aprendizagem
- Desafios

Data Mining

- Utilização “universal” de bases de dados
- ⇒ Extracção de conhecimento de dados
 - detecção de fraude: transacções de cartão de crédito, chamadas telefónicas, impostos
 - marketing directo: seleccionar clientes a quem enviar catálogo, anúncio a colocar em página web
 - medicina: diagnóstico, monitorização da condição de pacientes
 - indústria: identificação de falhas
 - ecologia: previsão de níveis de poluição
 - ciência: classificação de objectos
 - informática: detecção de spam, software adaptativo
 - ...

Extreme Data Mining

Muitos dados

+

Muito detalhe

+

Muitos problemas

+

Muitos modelos

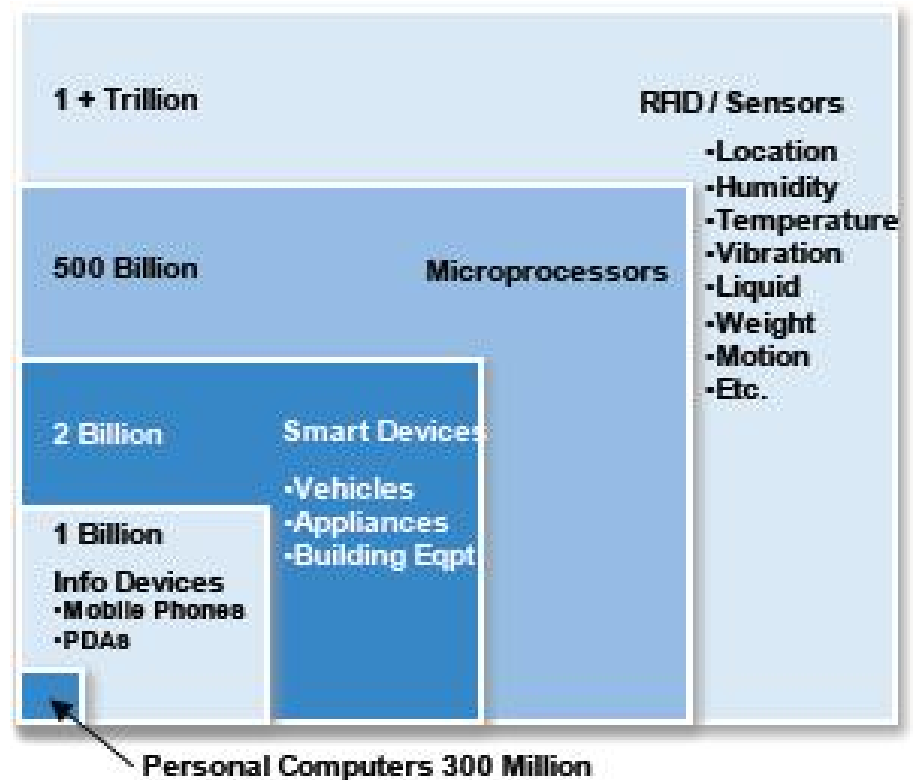
=

Extreme Data Mining

(adaptado da
definição de
Fogelman)

Muitos Dados

- Data mining tipicamente motivado por **grande** quantidade de dados
- mas **grande** está ficando cada vez **maior**



retirado de H. Tirri, *What Do People Really Do – On the power of Reality Mining for E-learning*, EDEN 2005 Annual Conference

+ Muito Detalhe

- É possível recolher dados mais detalhados
 - incluindo tempo e espaço
- sobre pessoas e seu comportamento
 - ex. cartões, call center, web e telemóveis
- ... produtos, sua produção, distribuição e utilização
 - ex. produção, RFID, sensores em automóveis e roupa
- ... processos
 - ex. email e ambientes colaborativos

+ Muitos Problemas

- Novas aplicações para as empresas
 - de marketing para controle de produção, desenvolvimento de produtos, supply chain management, etc.
- para o governo
 - ex. detecção de fraude fiscal, gestão de situações de catástrofe, monitorização ambiental
- ... e para as pessoas
 - ex. procura de informação, monitorização da saúde

+ Muitos Modelos

- Conhecimento mais específico
- ... ou seja, modelos para conjuntos mais pequenos
 - ex. [Fogelman 06]
 - “broadband communications company moved from 5 cross-sell models per year to 1600;
 - A wireless communications company that produces 700 CRM models per year;”
- ... no extremo, para objectos individuais
 - ex. um modelo de recomendação para cada cliente

Extreme Data Mining

Muitos dados

+

Muito detalhe

+

Muitos problemas

+

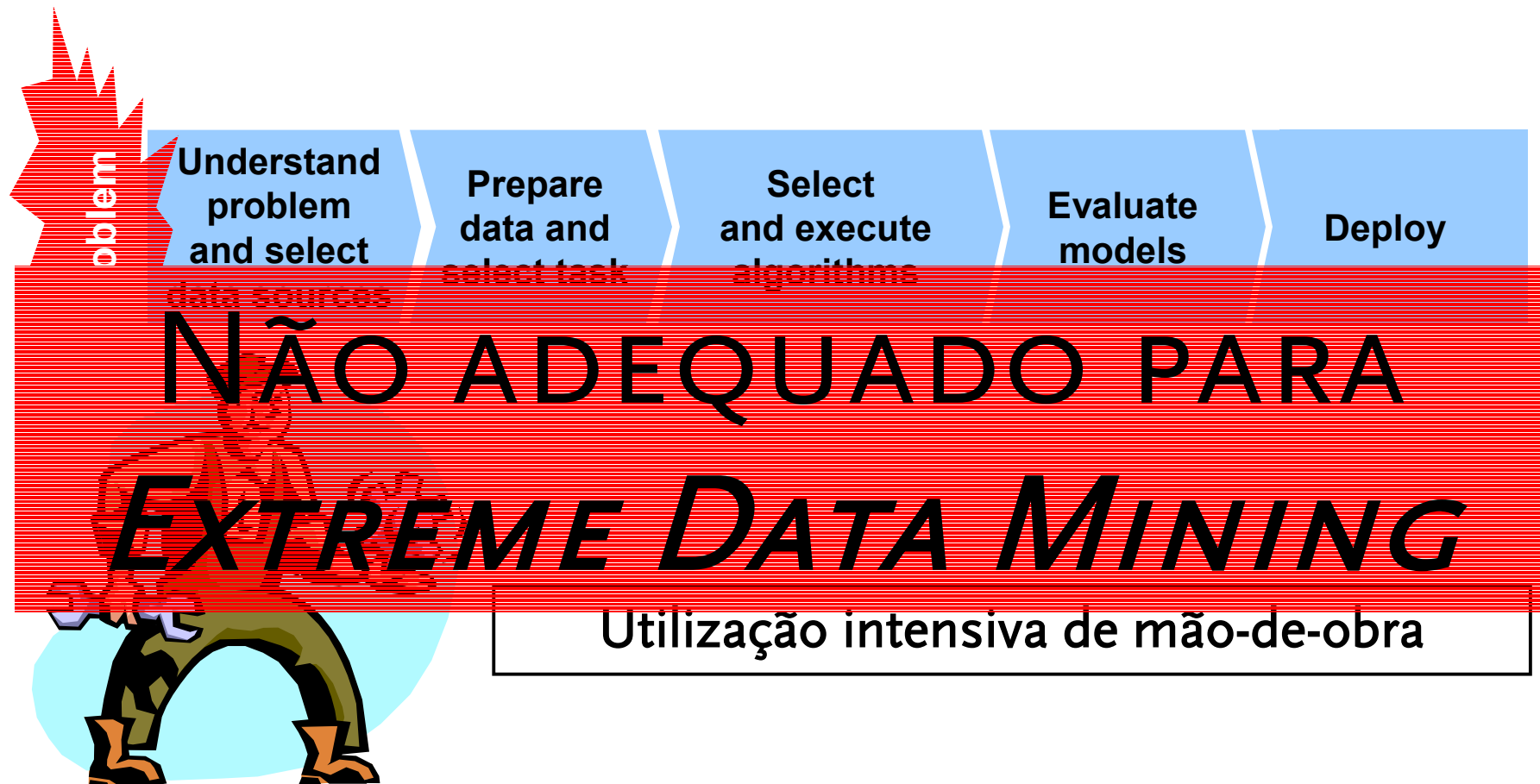
Muitos modelos

=

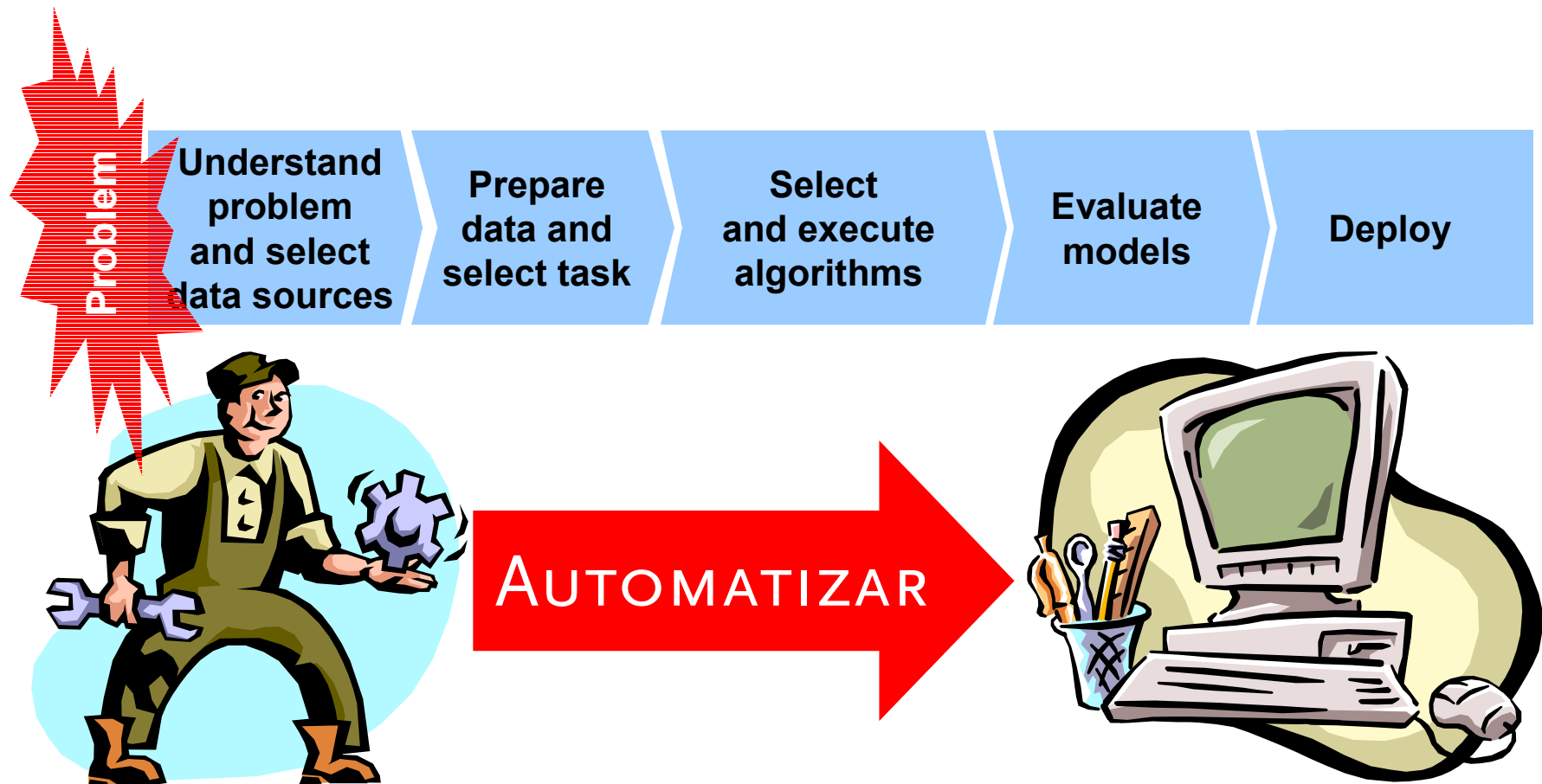
Muito Data Mining

(adaptado da
definição de
Fogelman)

Processo de Data Mining



Metodologias para Extreme Data Mining



Abordagens Possíveis

- *Templates*
 - ex. SPSS Clementine
 - pouco flexíveis
- Meta-aprendizagem
 - adaptação automática de processos de data mining para os problemas com base em resultados anteriores

Um Problema Concreto

- Prever a duração das viagens dos autocarros STCP
 - ex. para gestão das escalas
 - e informação aos clientes
- Muitas linhas
 - ... muitos modelos

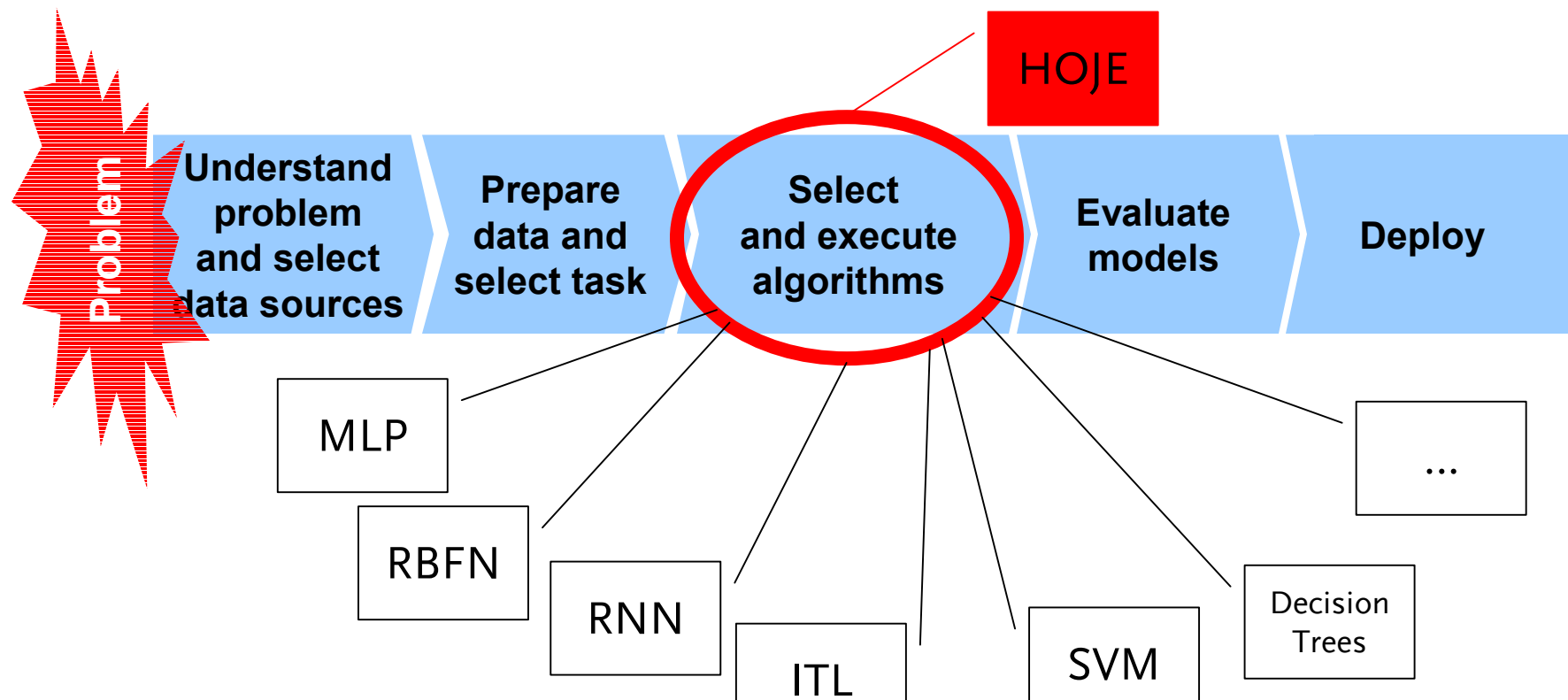


Plano

- Motivação
- Revisões sobre aprendizagem automática
 - aproximação de funções com aprendizagem automática
 - viés e escolha
 - algumas soluções
- Meta-aprendizagem
- Desafios

Foco da Palestra

- Objectivo: automatizar tanto quanto possível processo de data mining

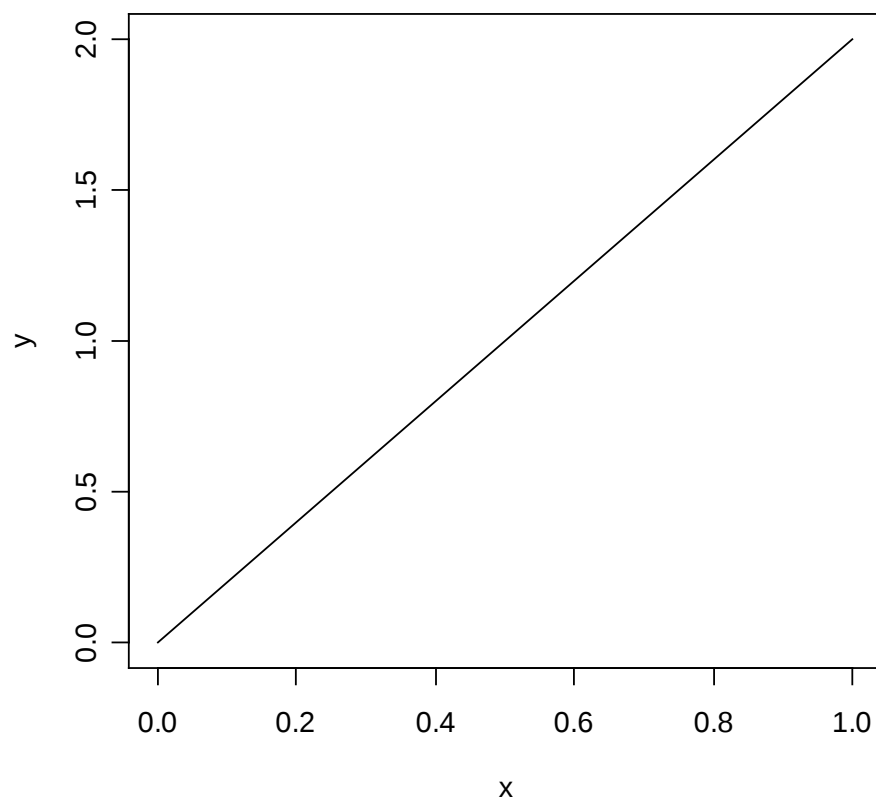


QUAL USAR?

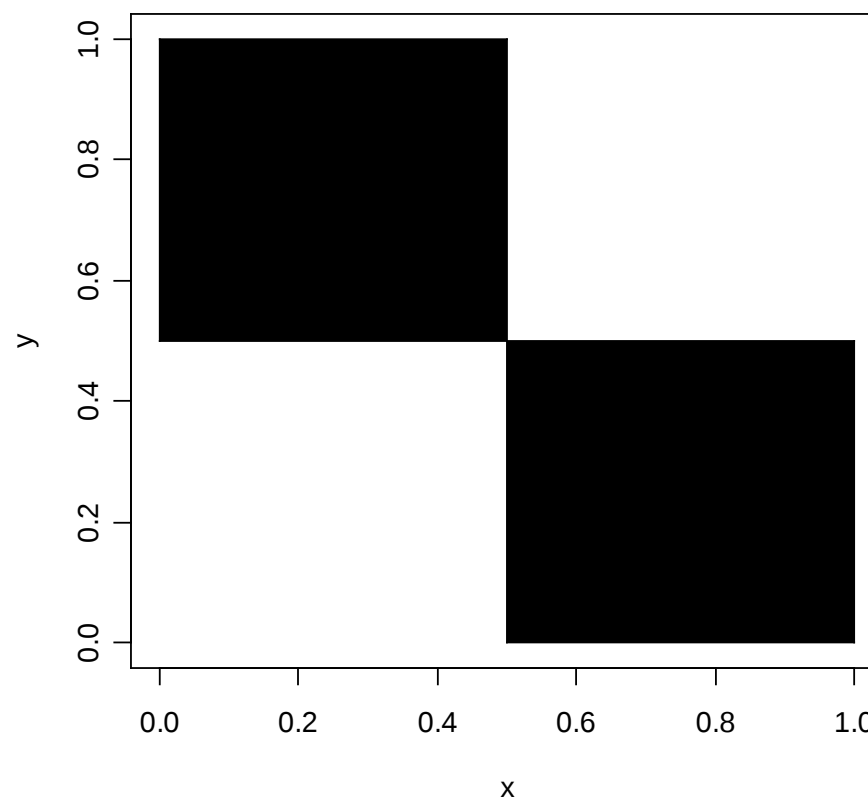
Problema (1/2)

- Fenómenos quantificáveis...

$$y=2*x$$



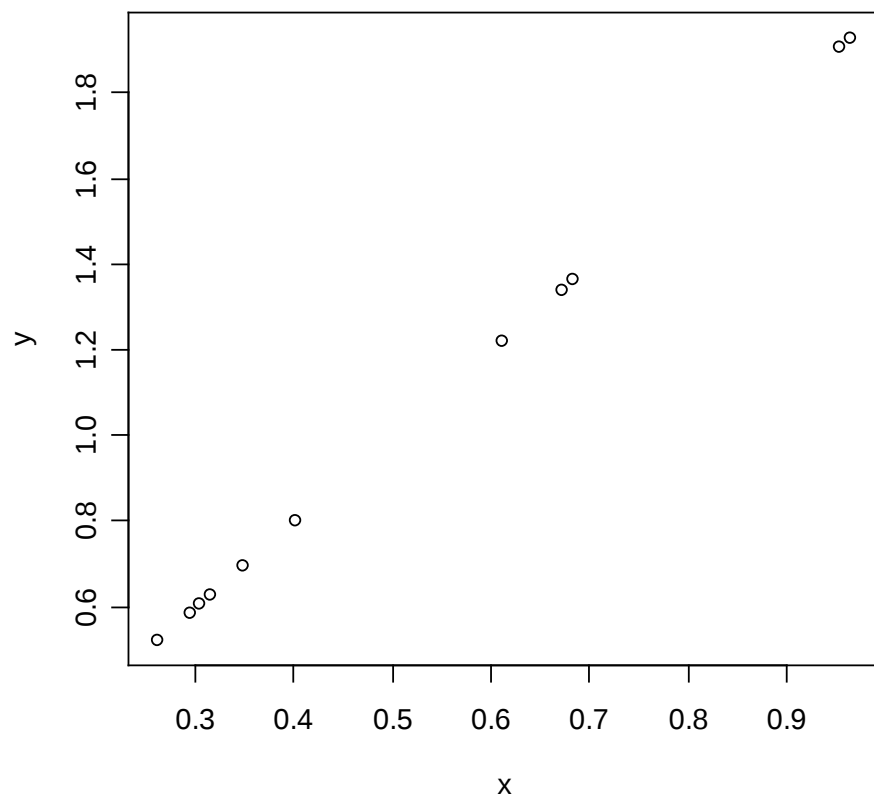
$$\text{preto: } (x > 0.5) \text{ XOR } (y > 0.5)$$



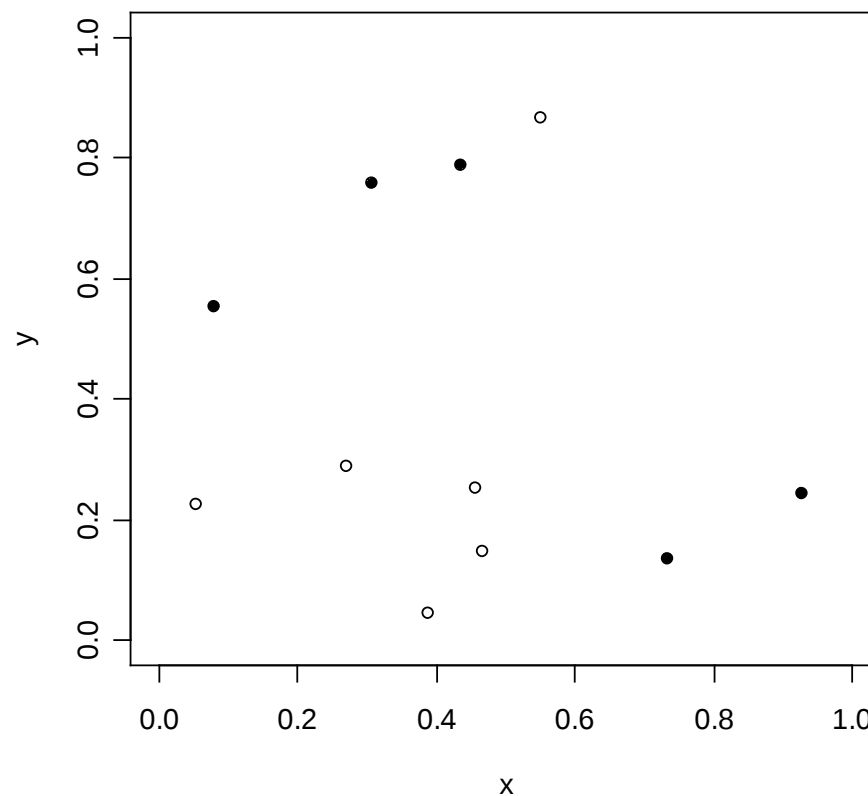
Problema (2/2)

- ... conhecidos através de amostras

$y=?(x)$

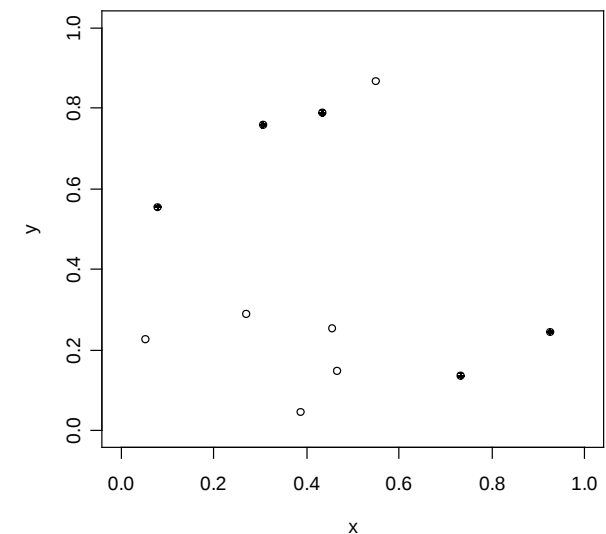
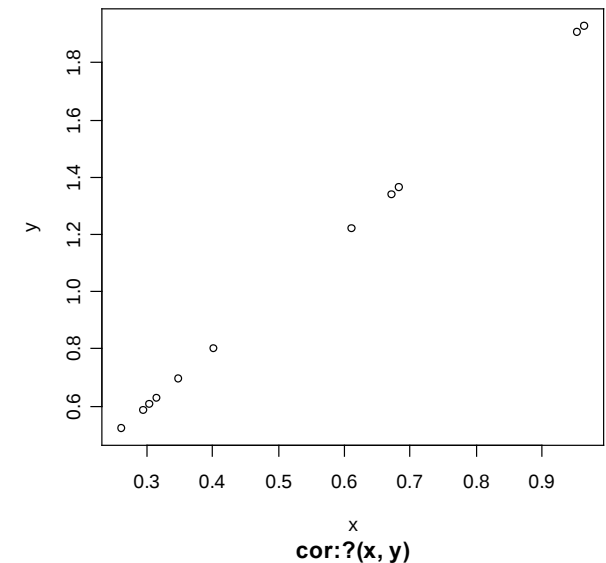


$cor:?(x, y)$



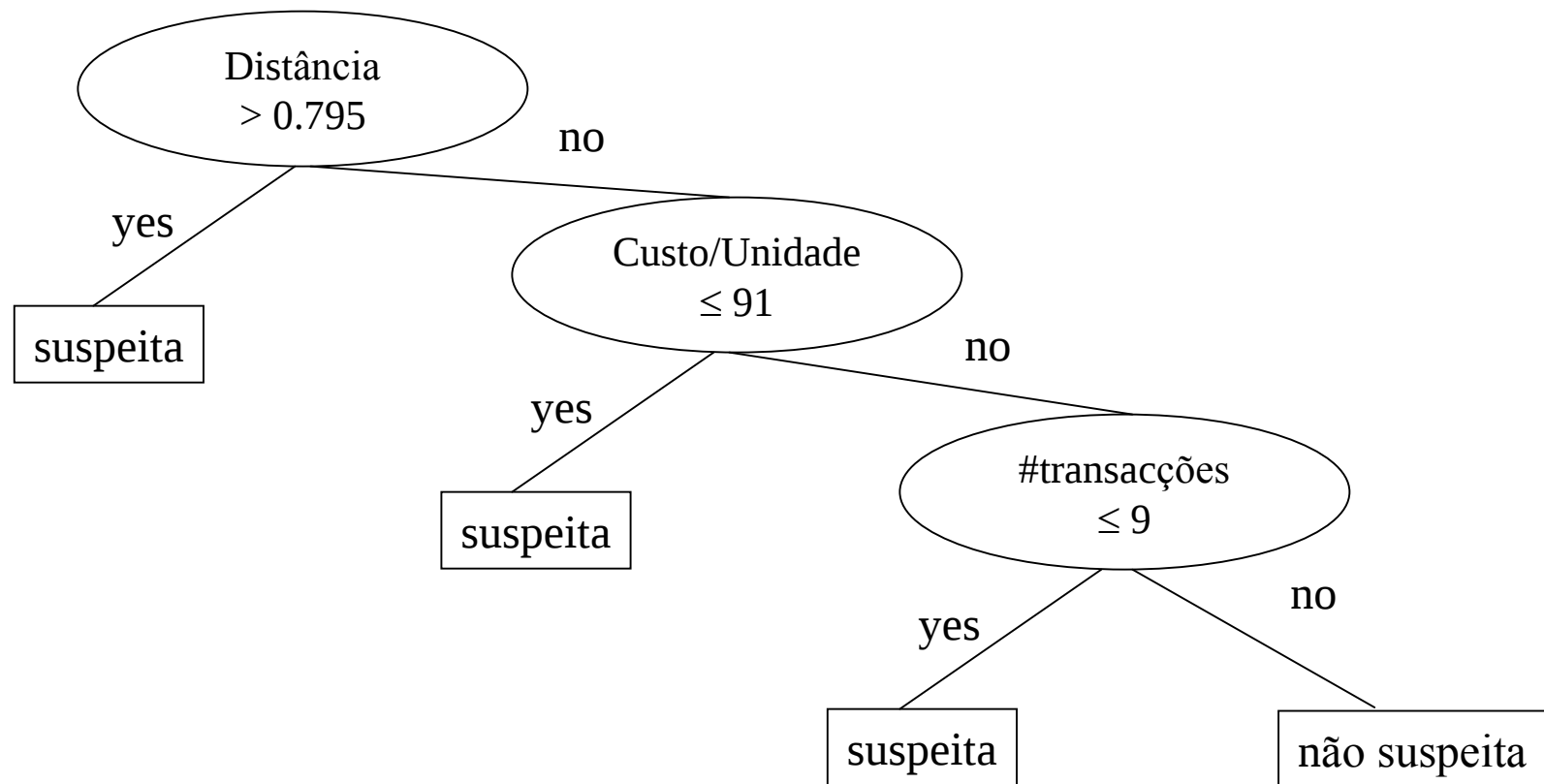
Aplicações da Aprendizagem Automática (*Machine Learning*)

- Com base em amostras de dados
- ... Modelação
 - qual a função que descreve a relação entre x e y ?
 - como descrever a área a preto como função de x e y ?
- ... Previsão
 - qual o valor de y dado $x=0.35$?
 - qual a cor do ponto ($x=0.3$, $y=0.7$)?



Exemplo: Árvore de Decisão

- Detecção de erros em transacções de comércio externo
 - aplicação do INE

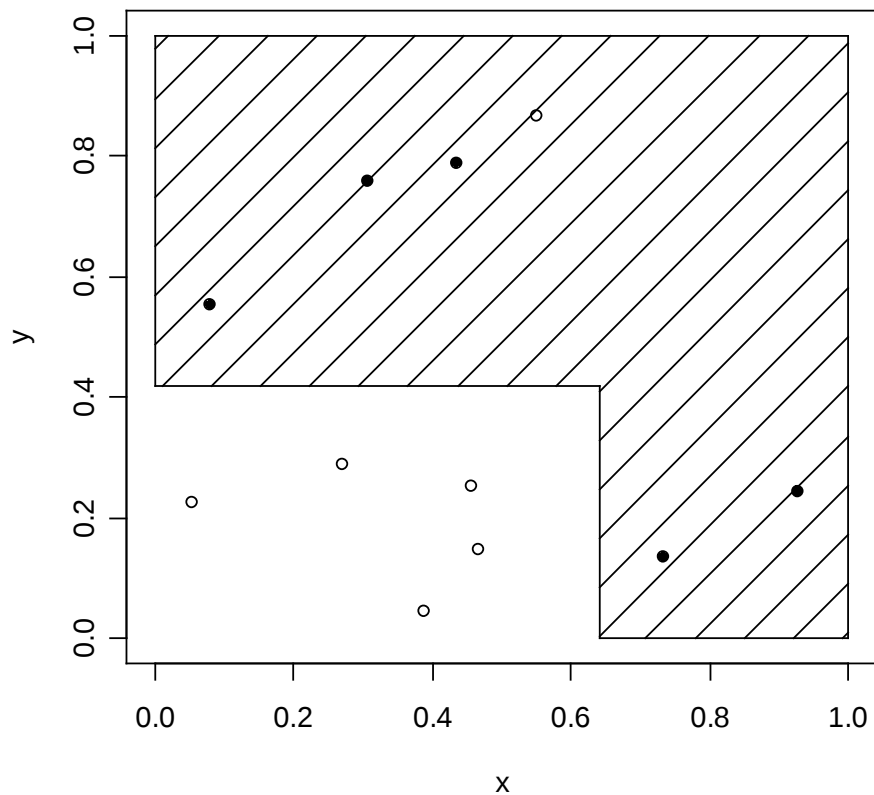


(versão simplificada)

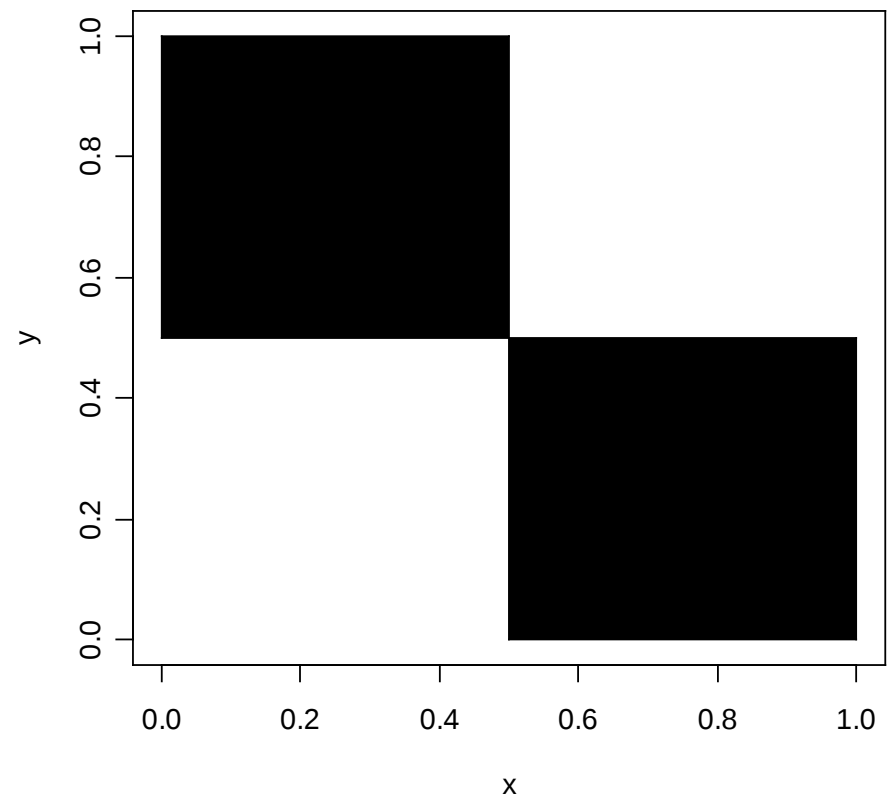
Algumas Questões (1/2)

- Representatividade da amostra

árvore de decisão
preto: $(x \geq 0.64) \text{ OR } (x < 0.64, y > 0.42)$

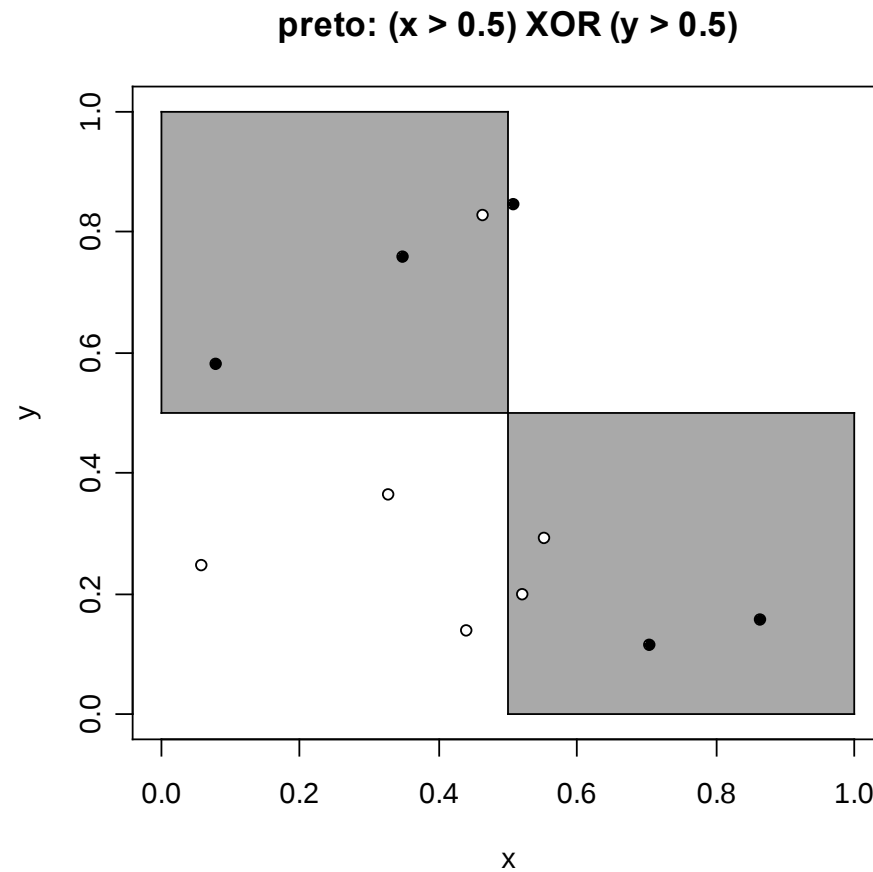


preto: $(x > 0.5) \text{ XOR } (y > 0.5)$



Algumas Questões (2/2)

- Ruído



Algoritmos de Aprendizagem Automática: Exemplos

- Árvores de decisão
- Indução de regras
- Redes neurais/*Support Vector Machines*
- Discriminantes lineares
- Naive Bayes
- *k-Nearest Neighbors*

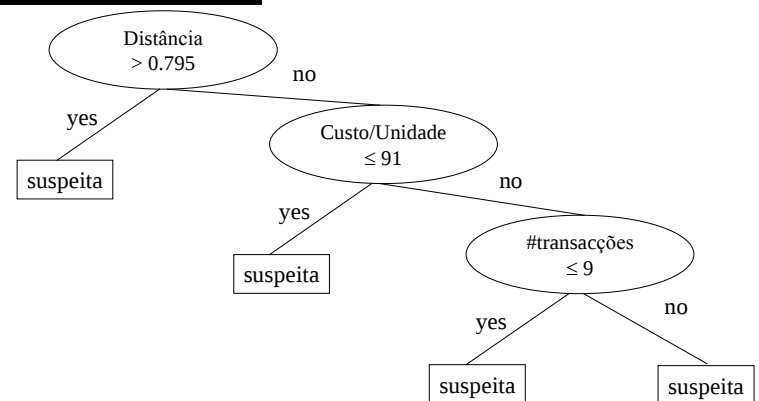
QUAL A DIFERENÇA?

Viés (*Bias*)

- Critério para preferir um modelo relativamente a outro
 - adaptação da definição des Jardins e Gordon (95)
 - extensão de Mitchell (80, 90), que excluía o erro
- Corolário
 - dada uma amostra de dados e um algoritmo de aprendizagem
 - ... nem todos os modelos podem ser obtidos
 - ... isto é, apenas algumas funções podem ser aproximadas
- Tipos de viés
 - representação: espaço de modelos
 - procedimental: algoritmo de procura

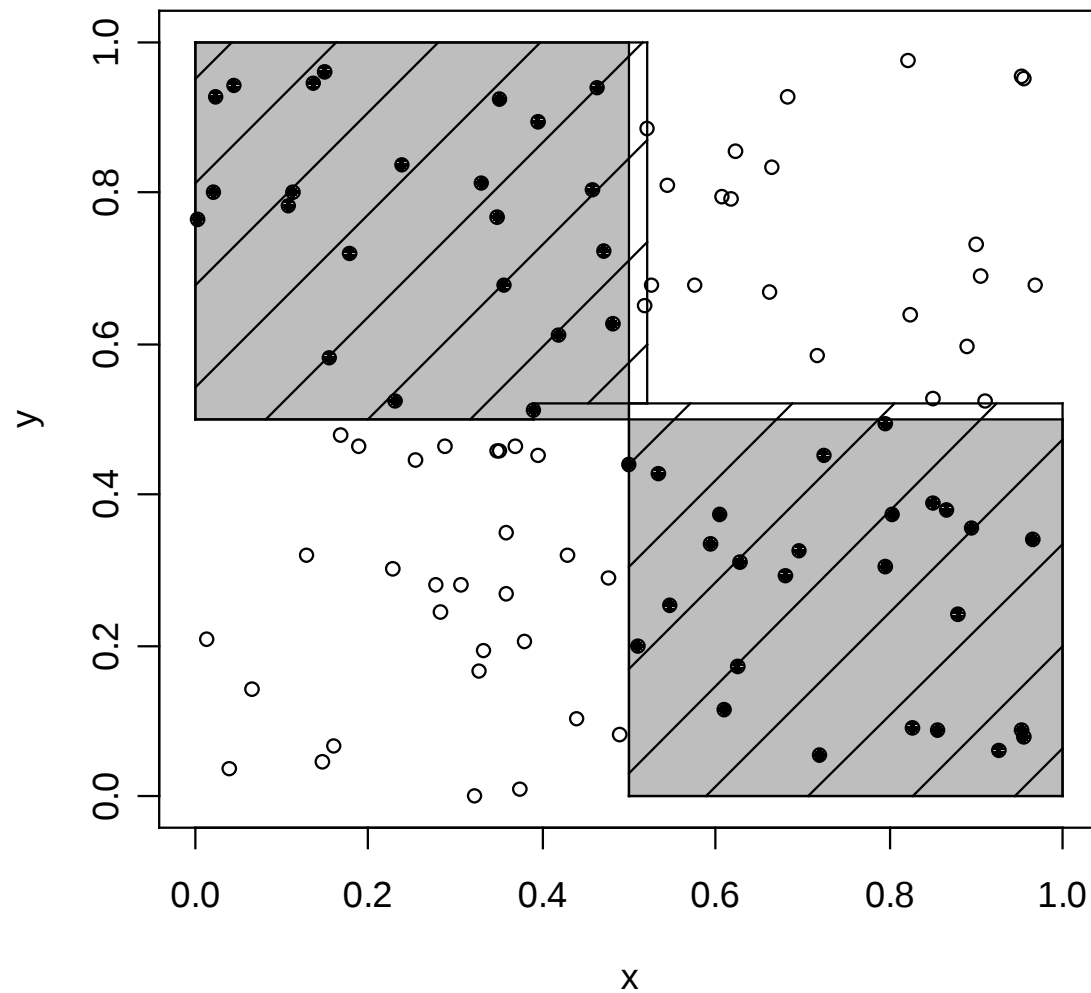
Exemplo: Viés do Algoritmo ID₃

- *Top-Down Induction of Decision Trees*
 - Quinlan (86)
- Espaço de modelos: expressões DNF
 - disjunção de conjunções
 - ... que definem hiperplanos
- Algoritmo de Procura: *Top-Down Induction*
 - começar por modelos simples (grandes hiperplanos)
 - gradualmente aumentar complexidade (mais pequenos)



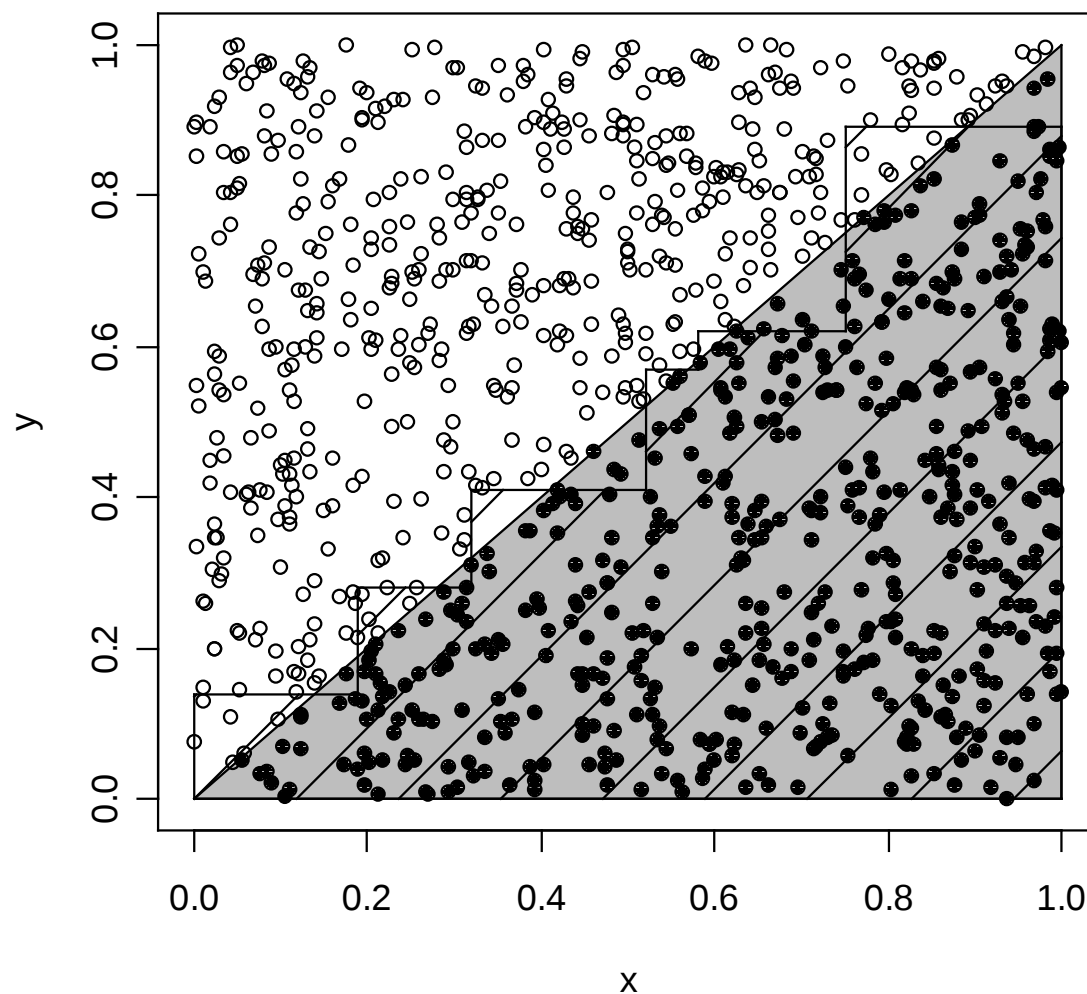
Exemplo: ID₃ Adequado

amostra: 100 exemplos



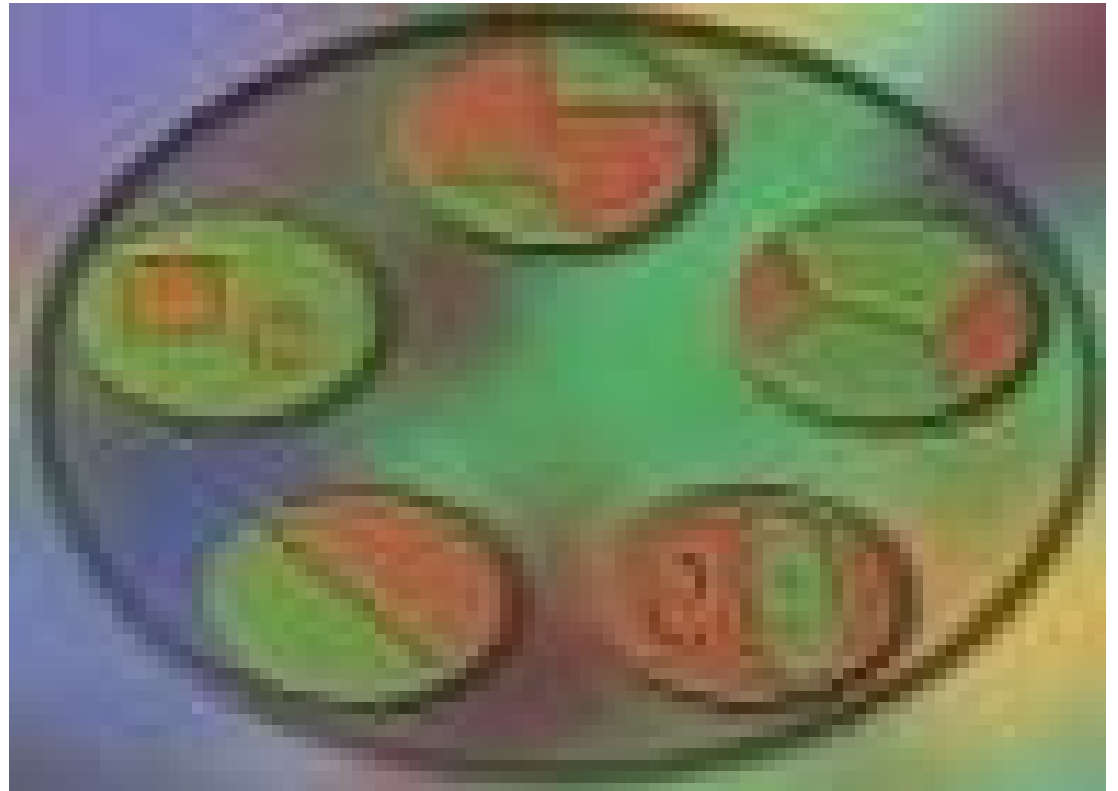
Exemplo: ID₃ Desadequado

amostra: 1000 exemplos



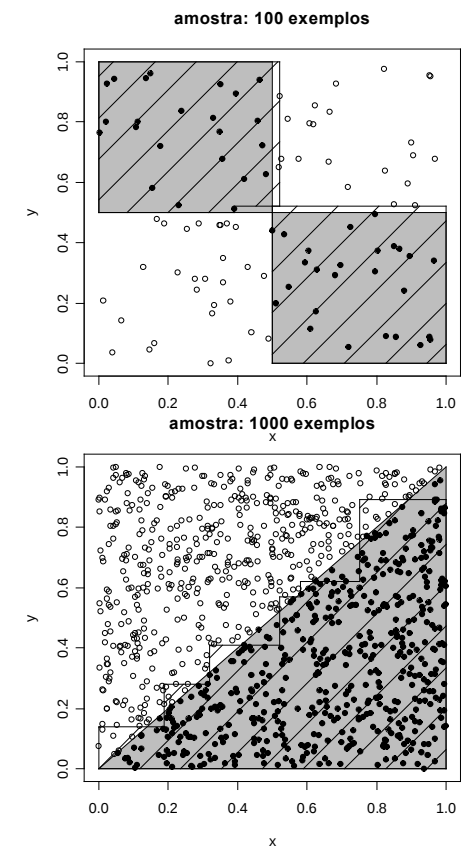
Tipos de Espaços de Modelos (Segundo Langley - 2000)

- Árvores de decisão
- Regras lógicas
- Casos
- Redes neuronais
- Descrições probabilísticas



Aprendizagem Automática: Resumo

- Encontrar função (modelo) que melhor se adequa à amostra de dados
- Escolha de algoritmo limita os modelos que se pode obter a partir de uma amostra de dados
 - pode correr bem
 - ... ou nem por isso



Solução I: Algoritmo Sem Viés

- Aprendizagem sem viés é inútil (Mitchell 97, Cap. 2)
 - um algoritmo que não assume nada relativamente à função que está a tentar aprender não tem nenhuma base racional para classificar casos desconhecidos
- Voltando à definição de viés
 - critério para preferir um modelo relativamente a outro
- ... e ao objectivo da aprendizagem
 - encontrar função (modelo) que melhor se adequa à amostra de dados
- ... como escolher o melhor modelo se todos os modelos são considerados igualmente adequados?

Solução II: À Bruta

- Testar todos os algoritmos em cada problema
 - escolher o melhor
- Computacionalmente impossível
 - muitos algoritmos
 - ... muitos deles com parâmetros
 - ... recursos limitados

Solução III: O Eleito

- Estudo experimental
 - alguns algoritmos (alguns valores dos parâmetros)
 - ... alguns problemas
 - ... escolher o melhor
 - ... usar sempre
- Teorema *No-Free Lunch*
 - Wolpert (96)
 - no limite (e sem mais informação), o erro médio de todos os algoritmos é o mesmo
- ... de valor prático limitado
- ... mas corolário confirmado pela prática
 - estudos empíricos mostram que “o eleito” não existe

Solução IV: Meta-Aprendizagem

Aprender por experiência quando
diferentes viés são adequados para
problemas particulares

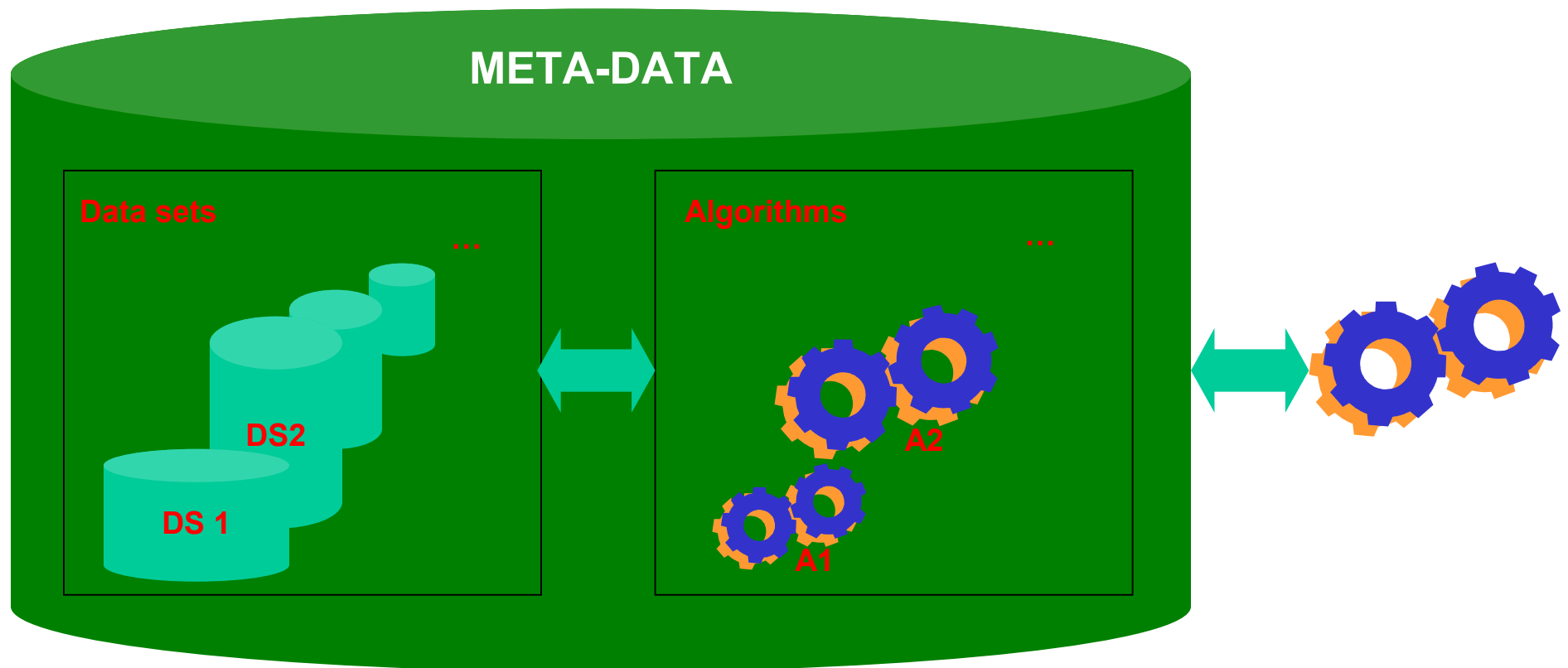
Rendell, Seshu e Tcheng (1987)

- Viés, viés... não
 - ... implementações de viés
 - ... ou algoritmos
- Há outras definições

Plano

- Motivação
- Revisões sobre aprendizagem automática
- Meta-aprendizagem
 - meta-aprendizagem por imagens
 - diferentes abordagens
- Desafios

Meta-Aprendizagem por Imagens



Base vs. Meta: Exemplos (Casos)

i	$x_{i,1}$	$x_{i,2}$	$target$
1	.	.	.
2	.	.	.
3	.	.	.

i	$x_{i,1}$	$x_{i,2}$	$target$
1	.	.	?
2	.	.	?
3	.	.	?

Base

Treinar

$x_{i,1}$	$x_{i,2}$	$target$
.	.	.
.	.	.
.	.	.

i	$x_{i,1}$	$x_{i,2}$	$target$
1	.	.	.
2	.	.	.
3	.	.	.

$$target = 1.04 \times x_1 + 0.38 \times x_2 + \dots$$

Aplicar

i	$x_{i,1}$	$x_{i,2}$	$target$
1	.	.	.
2	.	.	.
3	.	.	.

Meta

- objectos de interesse no domínio
 - ex. pacientes; clientes

- problemas de aprendizagem
 - ex. diagnóstico da doença x no hospital y ; direct marketing para empresa z

Base vs. Meta: Variável-objectivo

i	$x_{i,1}$	$x_{i,2}$	$target$
1	.	.	.
2	.	.	.
3	.	.	.

.	.	?
.	.	?
.	.	?

Treinar

$$target = 1.04 \times x_1 + 0.38 \times x_2 + \dots$$

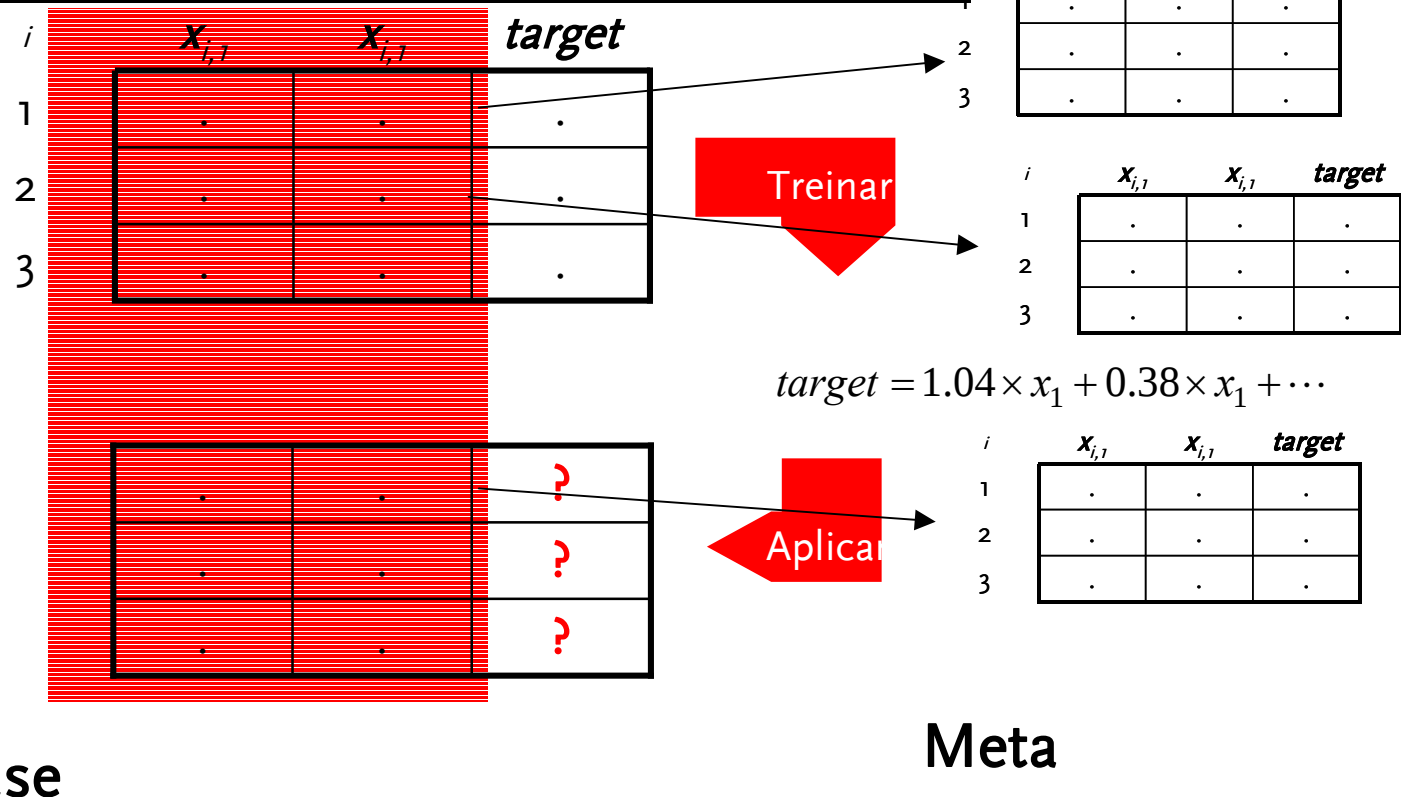
Aplicar

Base

Meta

- decisão específica de um problema
 - ex. diagnóstico; enviar promoção a cliente ou não
- algoritmo a usar
 - ex. árvores de decisão, MLP

Base vs. Meta: Variáveis Independentes



- características dos objectos de interesse
 - ex. emprego, idade e renda da pessoa; preço e tipo de produto

- características do problema de aprendizagem e medidas de qualidade das soluções
 - ex. número de variáveis, número de classes e taxa de erro

Base vs. Meta: Problema de Aprendizagem

i	$x_{i,1}$	$x_{i,2}$	$target$
1	.	.	.
2	.	.	.
3	.	.	.

Treinar

$$target = 1.04 \times x_1 + 0.38 \times x_2 + \dots$$

.	.	?
.	.	?
.	.	?

Aplicar

Base

Meta

- relacionamento entre variáveis do domínio de interesse
 - e.g. perfil da pessoa e rendimento; sintomas e diagnóstico
- relacionamento entre características do problema e adequação dos viés
 - e.g. morfologia da amostra e melhor algoritmo

Abordagens

- Alteração de viés durante a resolução de um problema e entre problemas diferentes
- Construção de um modelo feito de partes com viés diferentes
- Modelo constituído por vários modelos combinados
- Selecção de algoritmo(s) para um dado problema
- outras classificações são possíveis

Adaptação Contínua de Modelos/Algoritmos

- Alteração de viés durante a resolução de um problema e entre problemas diferentes
 - também conhecida por *knowledge transfer* ou *learning to learn*
- Exemplo: *Self-Modifying Policies*
 - Schmidhuber, Zhao e Schraudolph (97)
 - partes do algoritmo são capazes de alterar outras partes do algoritmo
 - ... e até de se alterar a si mesmas
 - probabilidade de utilização de uma parte depende do seu mérito
 - implementação com algoritmo de *reinforcement learning*
 - ... assume semelhança entre os problemas

Construção de Modelos

- Construção de um modelo feito de partes com viés diferentes
- Exemplo: *Model Class Selection*
 - Brodley (93)
 - algoritmo partição recursiva (*à la* árvores de decisão)
 - escolha de melhor viés em cada nó
 - implementação com 3 viés diferentes
 - ... sem aprendizagem: escolha feita com base em regras obtidas de peritos

Combinação de Modelos

- Modelo constituído por vários modelos combinados
- Exemplos: *Stacked Generalization/Cascade Generalization*
 - Wolpert (92)/Gama e Brazdil (00)
 - meta-conjunto de dados constituído pelas predições dos modelos de base
 - meta-modelo obtido por aprendizagem com meta-conjunto de dados
 - possivelmente recursivo

Seleccção de Algoritmos

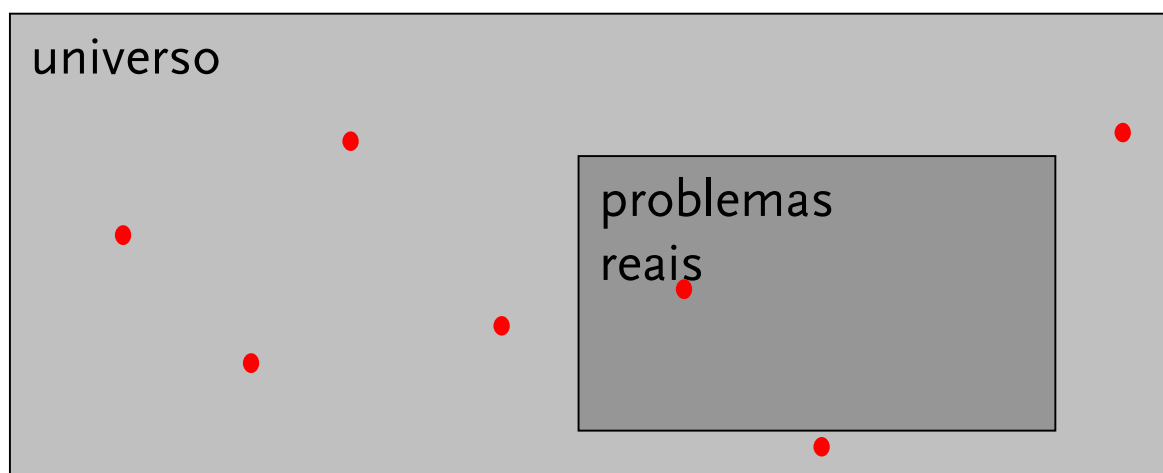
- Seleccção de algoritmo(s) para um dado problema
- Exemplos: recomendação de algoritmos
 - Rendell, Seshu e Tcheng (87)
 - Variable Bias Management System
 - Aha (92)
 - generalization from case studies
 - Brazdil, Gama e Henery (94)
 - abordagem sistemática
 - parte do projecto StatLog
 - projecto METAL (02 – várias publicação por vários autores)
 - follow-up to the work on the StatLog project
 - Data Mining Advisor website
 - www.metal-kdd.org (parcialmente funcional)

Plano

- Motivação
- Revisões sobre aprendizagem automática
- Meta-aprendizagem
- **Desafios**
 - obtenção de problemas de aprendizagem
 - obtenção de meta-dados
 - caracterização de problemas
 - outras aplicações

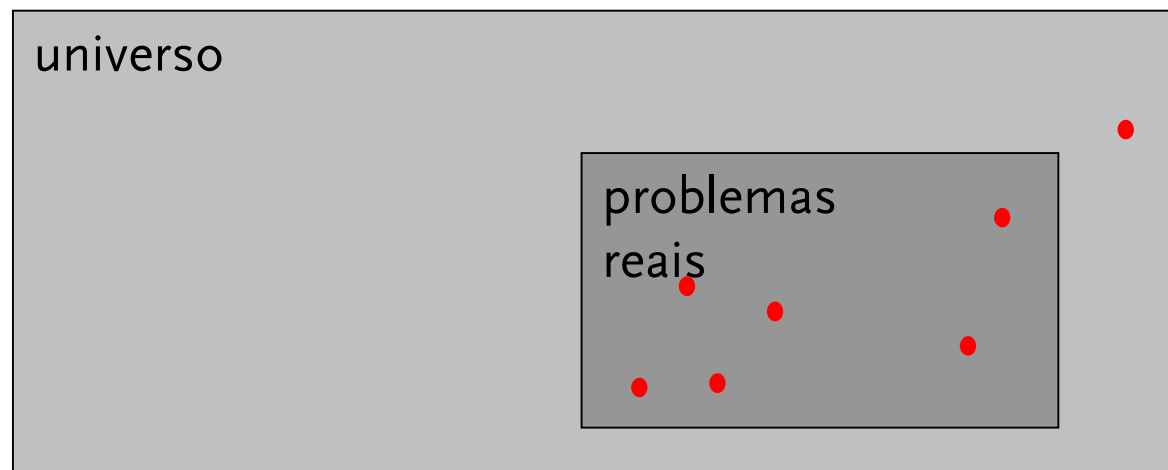
Quantidade/Qualidade dos Meta-dados

- Meta-aprendizagem usando dezenas de problemas
 - amostra pequena
 - resultados positivos mas serão estáveis?
- Como gerar mais meta-dados?
 - métodos **aleatórios** não são adequados



Aplicações Simuladas com Dados Reais

- Obter dados reais
 - ex. transacções numa empresa, intervalo entre passagem de carros
- Simular aplicações
 - escolher variável dependente independentemente de problema fazer sentido ou não
- Objectivo:



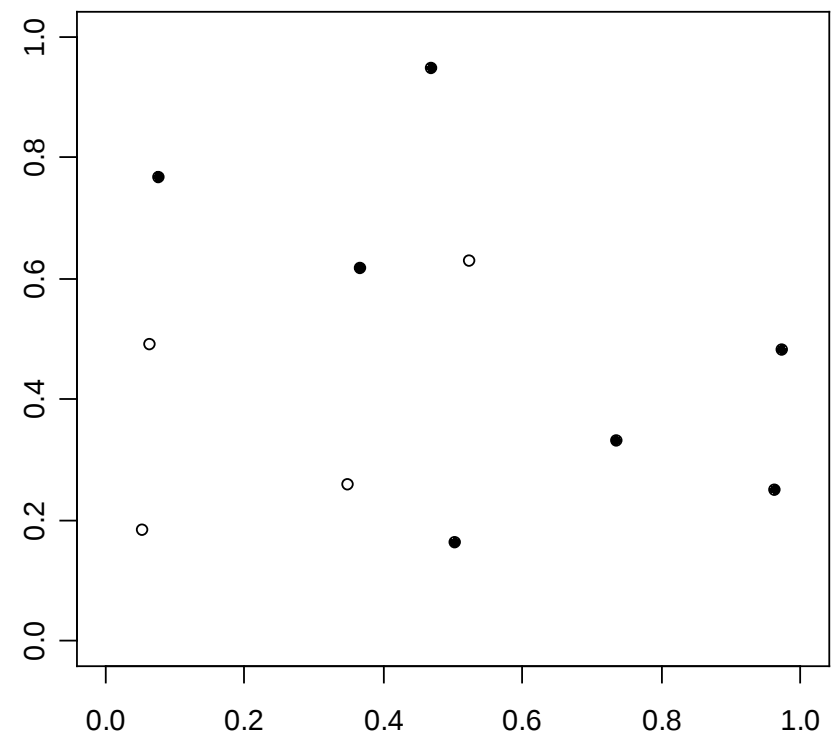
Obtenção de Meta-dados

- Pré-selecção de alternativas
 - que algoritmos/métodos de pré-processamento?
 - existentes nas ferramentas disponíveis
 - restrições aos modelos (ex. legibilidade)
 - preferências do analista (... ou ignorância relativamente a outros)
 - que parâmetros?
 - domínios frequentemente infinitos (parâmetros contínuos)
 - como saber se conjunto é adequado?
- Executar todas as alternativas é computacionalmente muito pesado
 - meta-active learning?

Caracterização dos Problemas

- Meta-variáveis: medidas que contêm informação sobre o desempenho **relativo** dos algoritmos
- ... mas que sejam computacionalmente menos pesadas que os algoritmos

adequação do viés: $?(x, y)$



Tipos de Meta-variáveis

- Gerais, Estatísticas e de Teoria da Informação
 - *à la* StatLog
 - ex. nr. variáveis, proporção de atributos numéricos com *outliers*, entropia de classes
- Landmarkers
 - Bensusan and Giraud-Carrier (2000)
 - versões simplificadas dos algoritmos contêm informação sobre os resultados das versões mais complexas
 - subsampling landmarks: resultados dos algoritmos em sub-amostras
 - Furnkranz and Petrak (01), Soares, Petrak and Brazdil (01)
- Baseadas em modelos
 - Bensusan, Giraud-Carrier and Kennedy (2000)
 - características de um modelo obtido com algoritmo (ou versão simplificada)

Outras Aplicações

- Testadas
 - regressão
 - recomendação de parâmetros (SVM)
 - recomendação de parâmetros de métodos de pré-processamento
- Em desenvolvimento
 - séries temporais
 - métodos de optimização
- A testar
 - métodos de pré-processamento + algoritmo + parâmetros
 - detecção de *outliers*
 - problemas específicos
- Extreme Data Mining

Conclusões

- Extreme Data Mining
 - variante importante de DM
 - importantes desafios de investigação
- Meta-aprendizagem
 - tecnologia importante para Extreme DM
 - ... ainda numa fase inicial do seu desenvolvimento
- ... que representa importantes desafios de investigação
 - e interessantes!

Alguém quer participar?

(csoares@liacc.up.pt)

Referências

algo enviesadas também...

- L. Rendell, R. Seshu and D. Tcheng (87), More Robust Concept Learning Using Dynamically-Variable Bias, in Proc. of the Fourth Int. Workshop on Machine Learning, pp 66-78
- P. Brazdil, C. Soares and J. Costa (03), Ranking Learning Algorithms: Using {IBL} and Meta-Learning on Accuracy and Time Results, Machine Learning, 50 (3), pp 251-277
- C. Soares, P. Brazdil and P. Kuba (04), A Meta-Learning Method to Select the Kernel Width in Support Vector Regression, Machine Learning, 54, pp 195-20
- C. Soares (03), Is the UCI Repository Useful for Data Mining?, Proceedings of the 11th Portuguese Conference on Artificial Intelligence (EPIA2003), LNAI 2902, pp 209-223
- C. Soares, P. Brazdil (06), Selecting Parameters of SVM Using Meta Learning and Kernel Matrix based Meta Features, Proceedings of the 21st Annual ACM Symposium on Applied Computing, 2006