

Aprendizado de Máquina

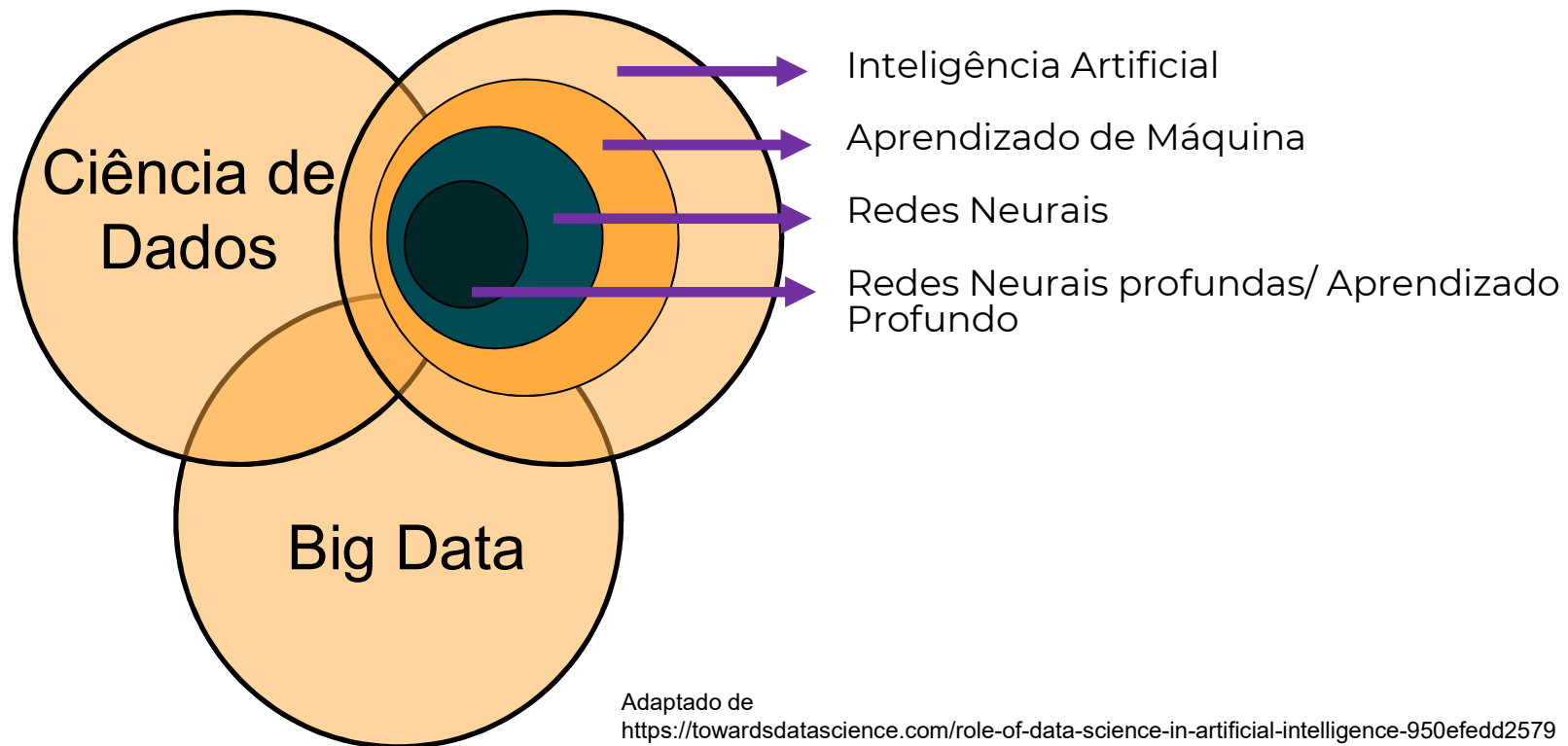
Revisão

André C. P. L. F de Carvalho
ICMC/USP

andre@icmc.usp.br



Começando com o ABC

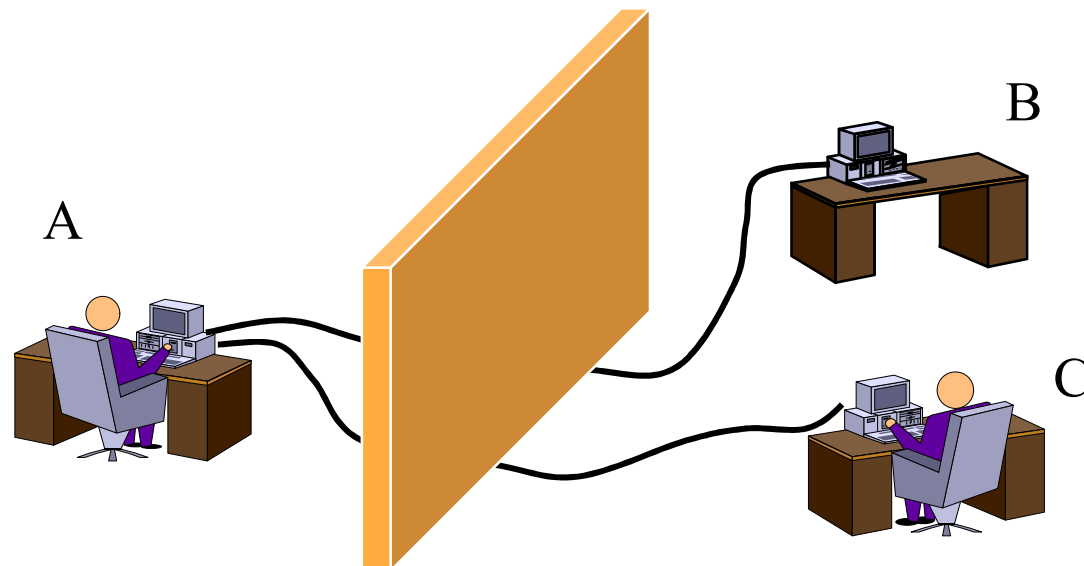


Adaptado de
<https://towardsdatascience.com/role-of-data-science-in-artificial-intelligence-950efedd2579>

Teste de Turing

- Realização do teste
 - Selecionar 3 indivíduos: A, B, C
 - A: Interrogador humano
 - B: Máquina
 - C: Humano
 - Supor que não existe contato físico entre A, B e C
 - A comunica-se com B e C indiretamente
 - Se A for incapaz de descobrir quem, entre B e C, é a máquina, a máquina é considerada inteligente

Teste de Turing

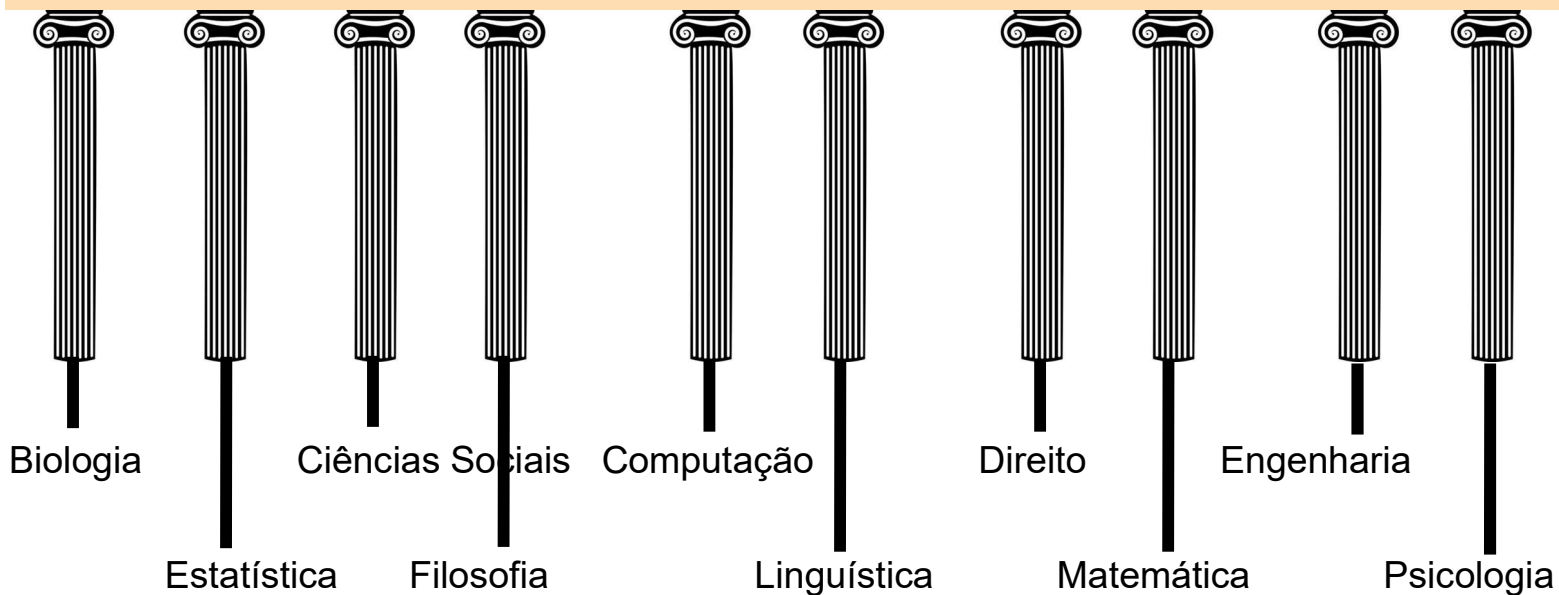


<http://www.youtube.com/watch?v=WnzlbyTZsQY>

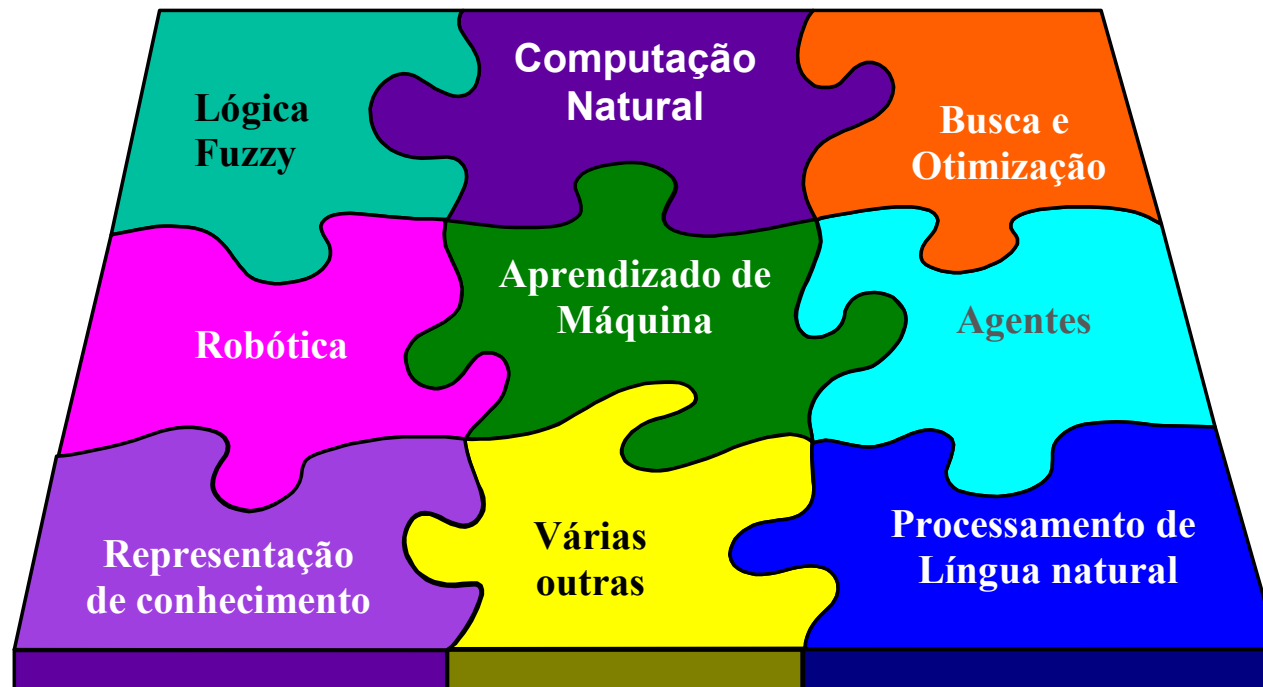
<http://www.youtube.com/watch?v=nAun0e8MuoE>

Áreas de apoio para IA

Inteligência Artificial



Sub-áreas da IA



Inteligência Artificial responsável

- É Inclusiva e está em consonância com os objetivos de desenvolvimento sustentável da ONU
- É reproduzível
 - Disponibilização e curadoria de dados e códigos
- Respeita a privacidade
 - Aplicando AM a 10 (300) likes, é possível conhece melhor sua personalidade que colega de trabalho (cônjuge)
 - Segue práticas justas para lidar com informação
- Presta contas (*accountability*)
 - Alguém deve responder pelas consequências de seu uso

Inteligência Artificial responsável

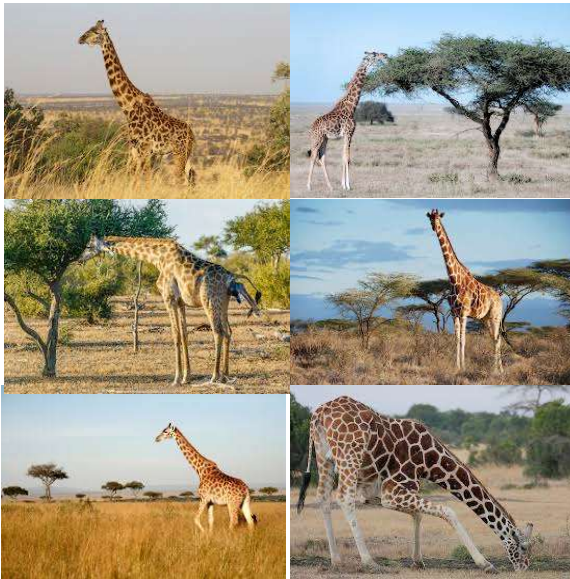
- É justa
 - Tomada de decisão não deve embutir preconceito
- Transparência
 - Segue a Lei Geral de Proteção aos Dados (LGPD)
 - Baseada na General Data Protection Regulation (GDPR-UE)
 - Direito a informação

Dados não estruturados

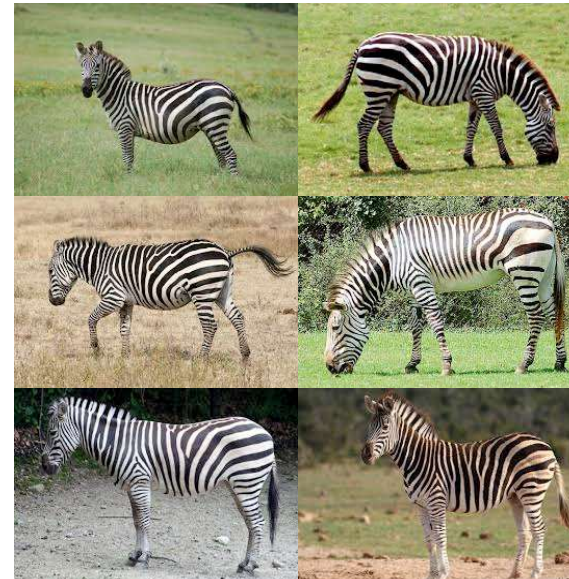
Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt. Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt.



Dados não estruturados



Girafa

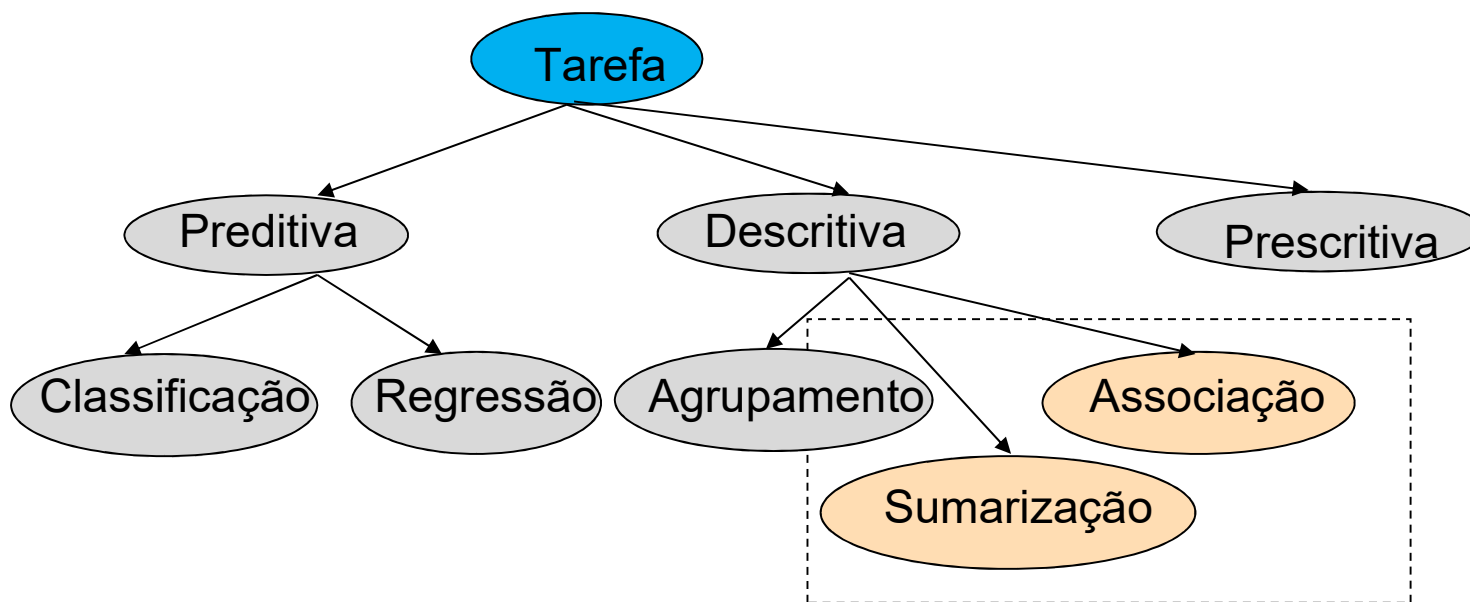


Zebra

Dados estruturados rotulados

Atributos de entrada (preditivos)				
Altura	TamanhoRabo	Textura	Classe	
500	110	Manchas	Girafa	Exemplos (objetos, instâncias)
440	90	Manchas	Girafa	
240	45	Listas	Zebra	
520	115	Manchas	Girafa	
260	50	Listas	Zebra	
230	50	Listas	Zebra	
				Atributo alvo

Tarefas de aprendizado



Tarefas preditivas

- Tarefas que precisam de um modelo capaz de prever o rótulo (atributo alvo) de seus exemplos
 - A partir do valor de cada um dos atributo preditivos do exemplo
 - Modelo preditivo
- Para a induzir o modelo pode ser usado um algoritmo de aprendizado de máquina
 - Algoritmo ensina (treina) o modelo a desempenhar bem sua tarefa por meio de um processo de aprendizado
 - Para isso, usa um conjunto de dados de treinamento
 - Um outro conjunto de dados, conjunto de dados de teste, avalia o quão bem o modelo aprendeu a realizar a tarefa

Tarefas descritivas

- Também buscam por modelos em um processo de treinamento
 - Descrevem ou sumarizam dados de uma tarefa
 - Treinamento utiliza todo o conjunto de dados
 - Ex.: Agrupamento de dados
- Algumas tarefas descritivas não possuem uma fase de treinamento
 - Ex.: Tarefas de sumarização e de associação de itens frequentes

Sumarização

- Objetivo: encontrar descrição simples e resumida para um conjunto de dados
- Frequentemente utilizada para:
 - Exploração interativa de dados
 - Geração automática de relatórios
 - Exemplo:
 - Definir perfis de pacientes com comorbidade

Sumarização

Escolaridade	Pressão alta	Sexo	Idade	Comorbidade
Médio	Sim	M	34	Não
Superior	Não	F	40	Não
Superior	Não	F	31	Não
Fundamental	Sim	F	18	Não
Médio	Não	M	76	Sim
Superior	Não	F	35	Não
Fundamental	Sim	M	20	Não
Superior	Não	M	76	Sim
Fundamental	Não	M	43	Não
Médio	Não	F	27	Sim

Escolaridade mais comum: Superior
 Frequência de comorbidade: 30%
 Idade média: 40
 Igualdade de sexo: S
 Menor idade: 18

Associação de itens frequentes

- Frequent itemsets
- Objetivo: dado um conjunto de itens e uma base de dados de transações
 - Encontrar conjunto de regras que, nas várias transações realizadas, associem a presença de um item à presença de outros itens
 - Conjunto de regras de associação
- Exemplo:
 - Procurar por itens que são frequentemente comprados juntos em um supermercado
 - Problema das cestas de compras

Associação de itens frequentes

- Problema das cestas de compras
 - Conjunto de transações, em que cada transação é uma compra feita em um dado supermercado

Transação	Itens comprados
1	pão, queijo, manteiga, massa
2	pão, geleia, suco
3	queijo, arroz, massa
4	queijo, vinho
5	massa, queijo, pão

66% dos clientes que compraram pão também compraram queijo

75% dos clientes que compraram queijo também compraram massa

Algoritmos de aprendizado de máquina

- Para cada tarefa de aprendizado de máquina, podemos pensar em um grande número de aplicações
 - Cada aplicação é representada por um conjunto de dados
- Para encontrar uma boa solução para uma aplicação precisamos encontrar um modelo que se ajuste bem ao seu conjunto de dados
 - Em aplicações preditivas, deve produzir a saída desejada para cada entrada
- A forma de procurar por esses modelos é descrita por um algoritmo de aprendizado de máquina
 - Sequência de passos que definem como procurar por um bom modelo para um conjunto de dados
 - Conjunto de regras que definem como ajustar os parâmetros do modelo

Algoritmos de aprendizado de máquina

- Podem seguir diferentes paradigmas
 - Supervisionado
 - Podem ser usados em tarefas preditivas (mais comum) ou descritivas
 - Algoritmo de aprendizado supervisionado usa atributos preditivos e alvo
 - Não supervisionado
 - Podem ser usados em tarefas descritivas (mais comum) ou preditivas
 - Algoritmo de aprendizado não supervisionado usa apenas atributos preditivos
 - Semi-supervisionado
 - Caso especial: aprendizado ativo
 - Por reforço

Algoritmos de AM

- Induzem modelos (funções, hipóteses) a partir de um conjunto de dados
 - Idealmente, dados devem ser:
 - Estruturados
 - Representativos
 - De boa qualidade
- Possuem um viés
 - Tendência a privilegiar uma ou mais hipóteses que atendam a um dado critério

Viés indutivo

- Algoritmos de AM precisam ter um viés indutivo
 - Necessário para restringir o espaço de busca
 - Sem viés não há generalização
 - Sem generalização não ocorre aprendizado
 - Modelos (regras / equações) seriam especializados para os dados usados para sua indução

Viés indutivo

- Viés de preferência ou busca
 - Define como as hipóteses são pesquisadas no espaço de possíveis hipóteses
 - Preferência de algumas hipóteses sobre outras
 - Ex.: preferência por hipóteses simples (curtas)
- Viés de representação ou linguagem
 - Define o espaço de busca de hipóteses
 - Restringe as hipóteses que podem ser geradas
 - Ex.: hipóteses no formato de árvores de decisão

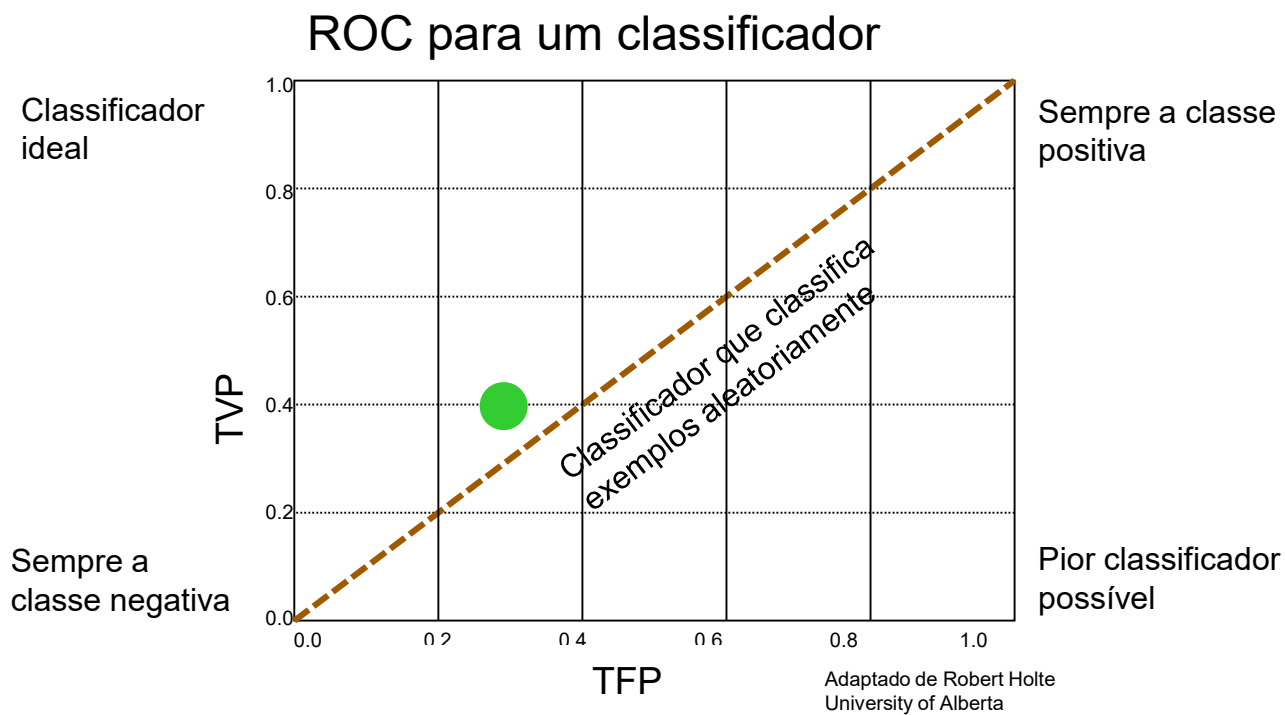
Algoritmos preditivos

- Podem ser agrupados por diferentes critérios
 - Baseados em proximidade
 - K-vizinhos mais próximos
 - Baseados em otimização (conexionistas)
 - Redes Neurais
 - Baseados em probabilidade
 - Naive Bayes
 - Baseados em procura (lógicos)
 - Indução de árvores de decisão

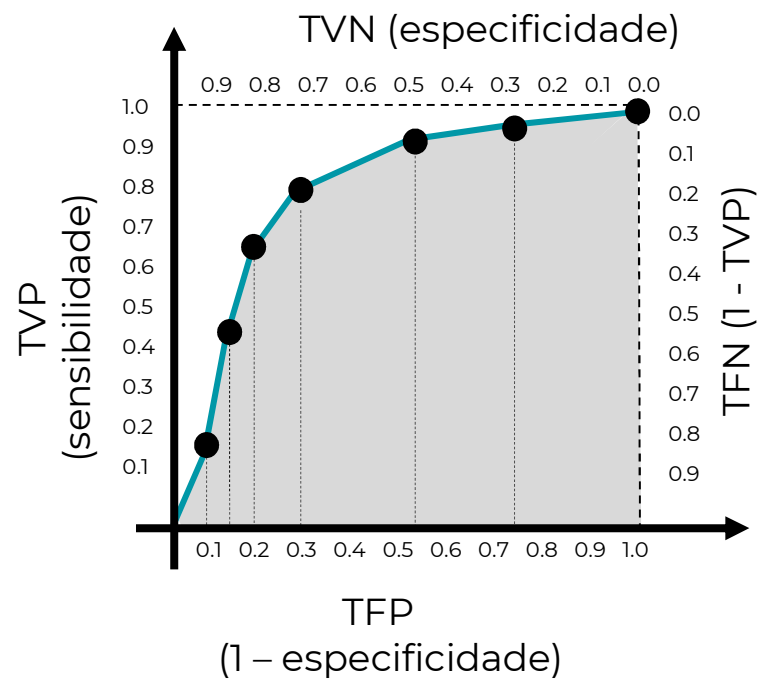
Ajuste de (hiper-)parâmetros

- Algoritmos de AM ajustam valores de um conjunto de parâmetros
 - Parâmetros do modelo que está sendo induzido (gerado)
 - Cada conjunto de valores de parâmetros pode gerar um modelo com comportamento diferente
 - Ajustados pelo algoritmo
- Algoritmos de AM possuem hiper-parâmetros
 - Definem que modelos o algoritmo pode gerar
 - Ajustados por quem está usando o algoritmo
 - Buscando induzir modelos com o melhor desempenho (preditivo) possível

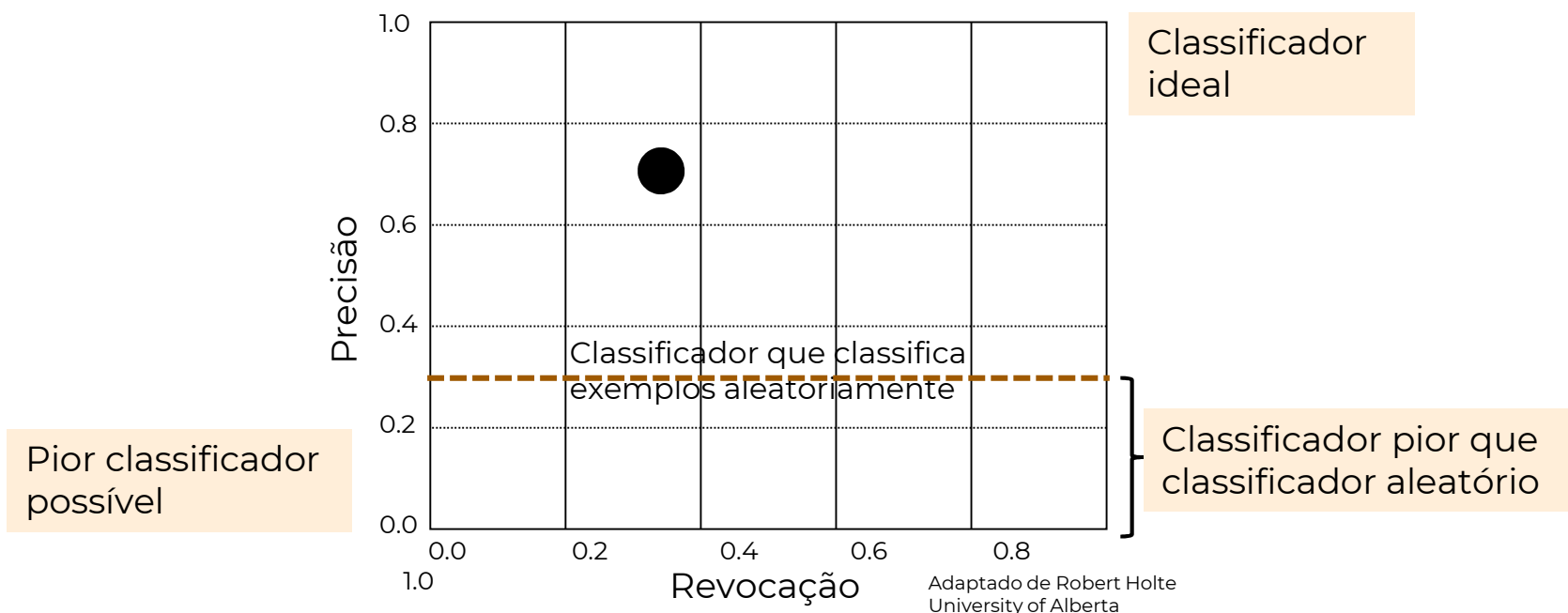
Gráficos ROC



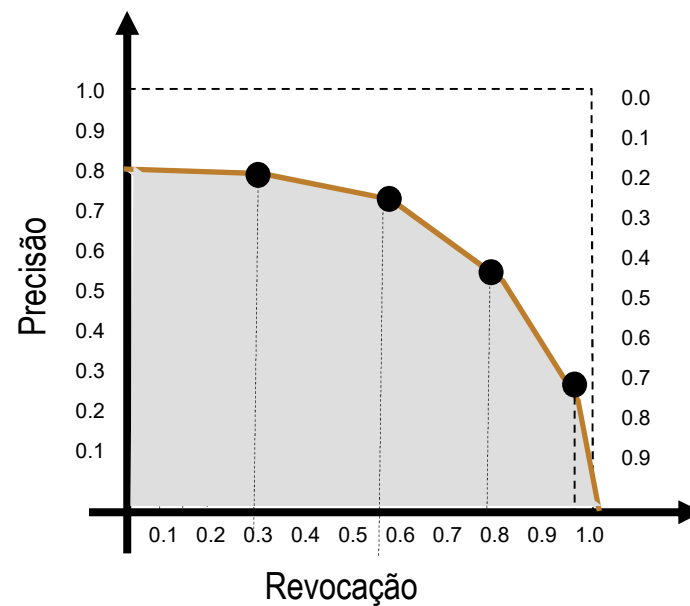
Área sob a Curva ROC



Gráficos precisão-revocação



Área sob a Curva Precisão-Revocação



Raciocínio baseado em casos

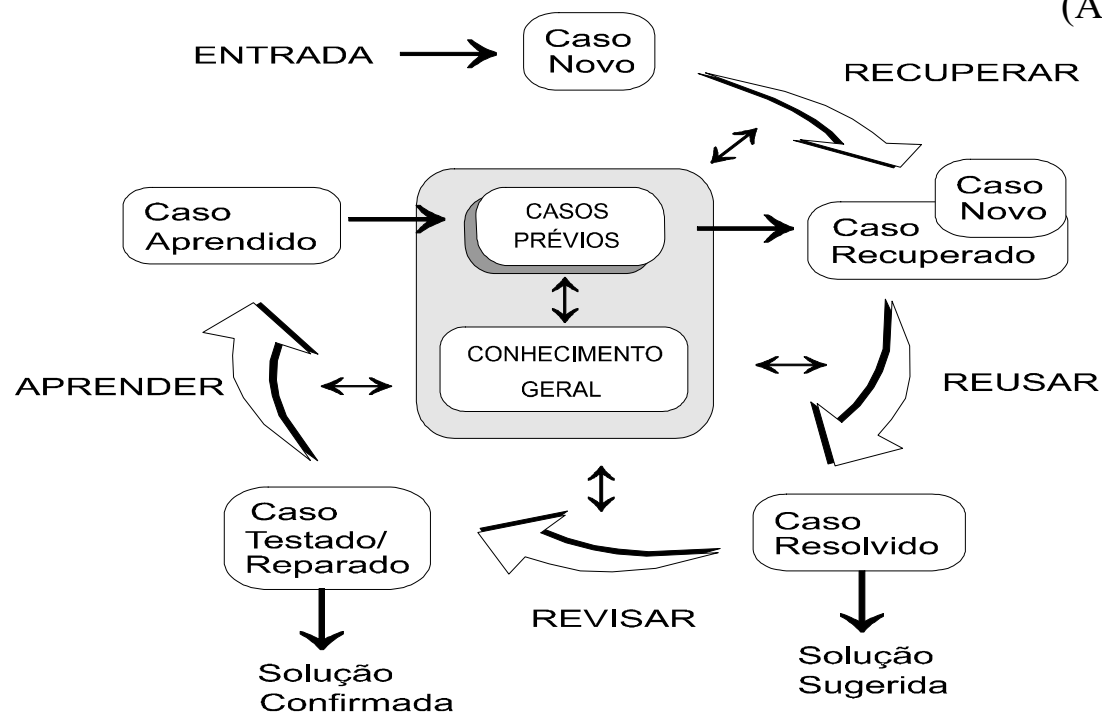
- Case-based reasoning (CBR)
- Mais que um sistema de gerenciamento de banco de dados
 - Sistemas de CBR procuram casos anteriores semelhantes ao problema (caso) atual
 - Buscam por eles em uma base de casos indexada
 - Indexação e representação de casos facilitam
 - Recuperação de casos relevantes e
 - Comparação desses casos com o problema atual
 - Permitem ainda a adaptação de casos recuperados

O que é um caso?

- Situação problemática, estrutura de dados com duas partes
 - Uma parte problema
 - Descrição de características (situações) de problemas específicos
 - Usada para identificar um caso
 - Indexação e recuperação
 - Uma parte solução
 - Explica como o caso foi resolvido anteriormente
 - De forma bem (ou mal) sucedida
 - Pode ser adaptada quando o caso é recuperado

Ciclo de um sistema de CBR

(Adaptado de Aamodt, 1993)



Sistemas de recomendação

- Recommender systems (RSs)
- Permitem lidar com sobrecarga de informação
 - Quando existe um grande número de opções para que um usuário ativo escolha uma delas
 - Usuário ativo: usuário que usa o RS para receber uma recomendação
 - Ajudando as pessoas a fazerem escolhas melhores
- Assim como algoritmos de aprendizado de máquina (AM), RSs utilizam álgebra linear
 - Manipulação de matrizes
 - Matrizes de avaliações (rating matrices)

Principais abordagens de Sistemas de recomendação

- Filtragem baseada em conteúdo
 - Usa informações sobre o próprio usuário ativo, ou item desejado, para propor itens
- Sistemas baseados em conhecimento
 - Usados para itens que não são frequentemente adquiridos
 - Ex.: imóveis, veículos e artigos de luxo
- Filtragem colaborativa (mais comum)
 - Usa escolhas de itens feitas por outro(s) usuário(s) para propor itens ao usuário ativo
- Aspectos dessas abordagens podem ser combinados, gerando uma abordagem híbrida

Colaborativa X Conteúdo

Colaborativa



Recomendar itens



Faz recomendações que levam em conta os gostos de usuários parecidos



Conteúdo



Recomendar itens

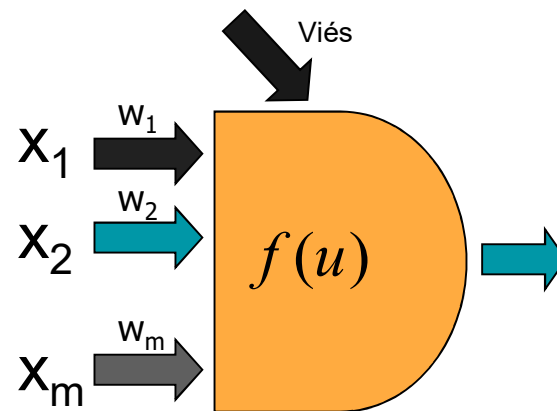


Faz recomendações que levam em conta o perfil do usuário

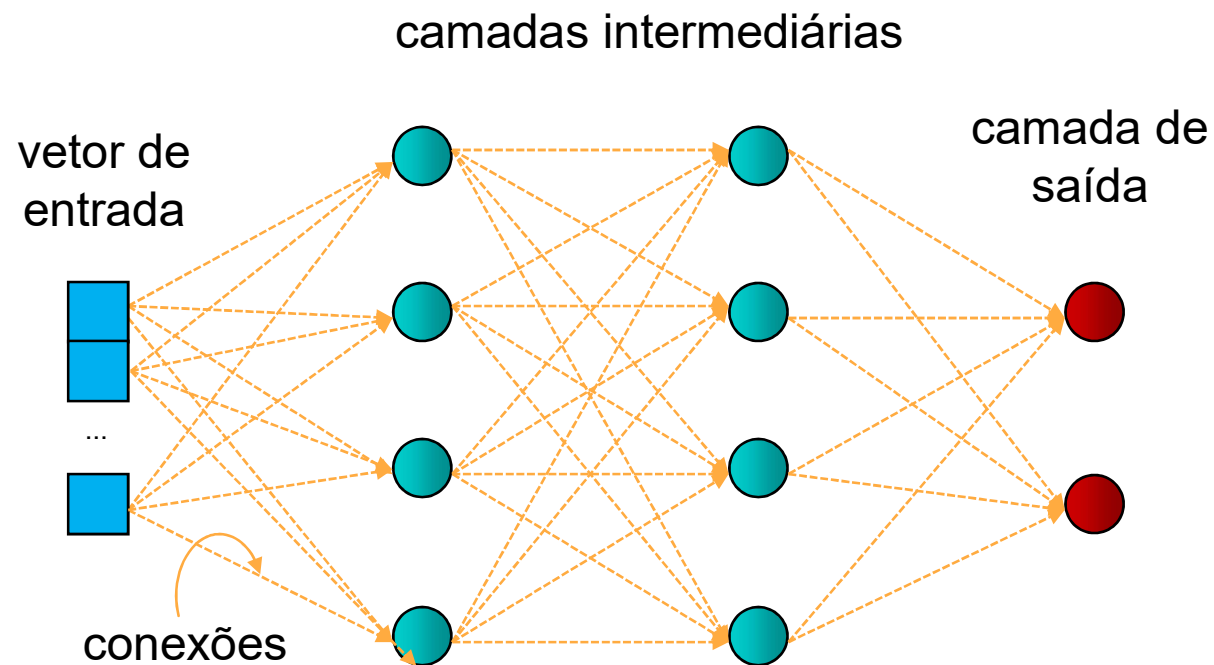


Redes Neurais

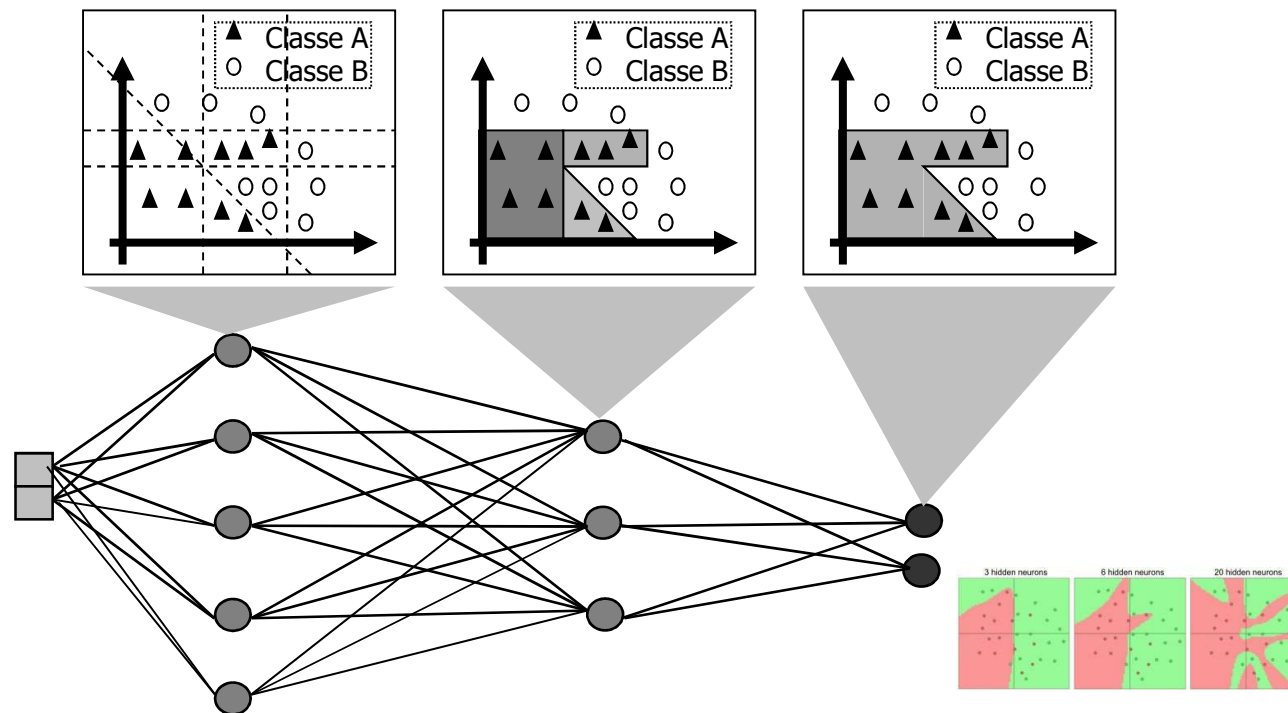
- Sistemas distribuídos inspirados no cérebro humano
 - Redes são compostas por várias unidades de processamento (“neurônios”)
 - Interligadas por um grande número de conexões (“sinapses”)
- Bom desempenho preditivo em várias aplicações



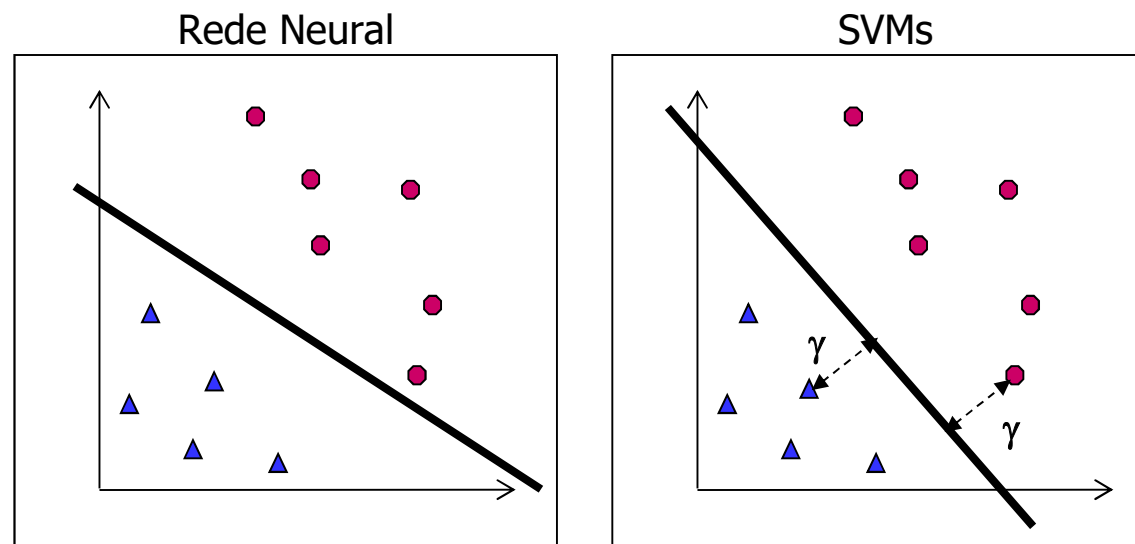
MLP e backpropagation



Redes MLP como classificadores

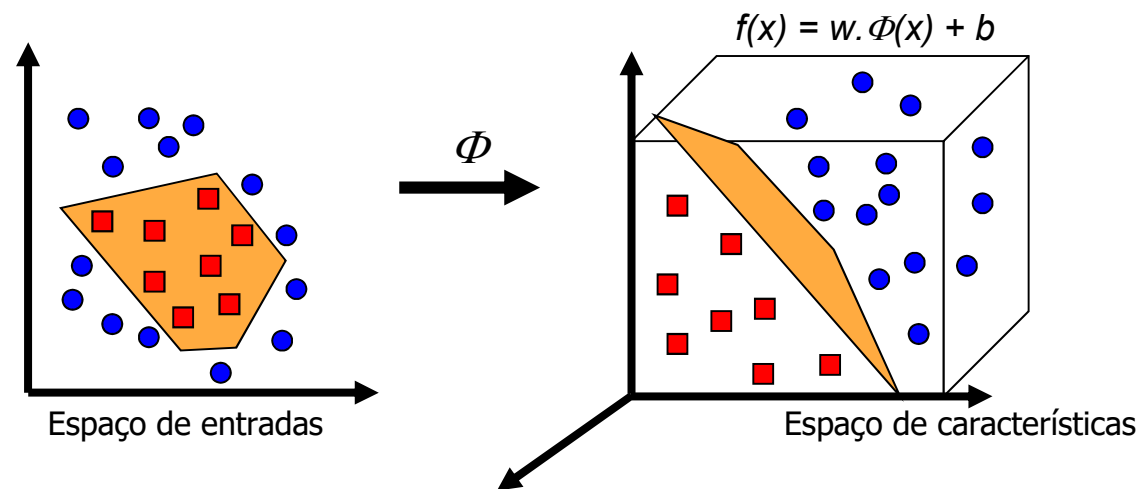


Máquinas de Vetores de Suporte (SVMs)

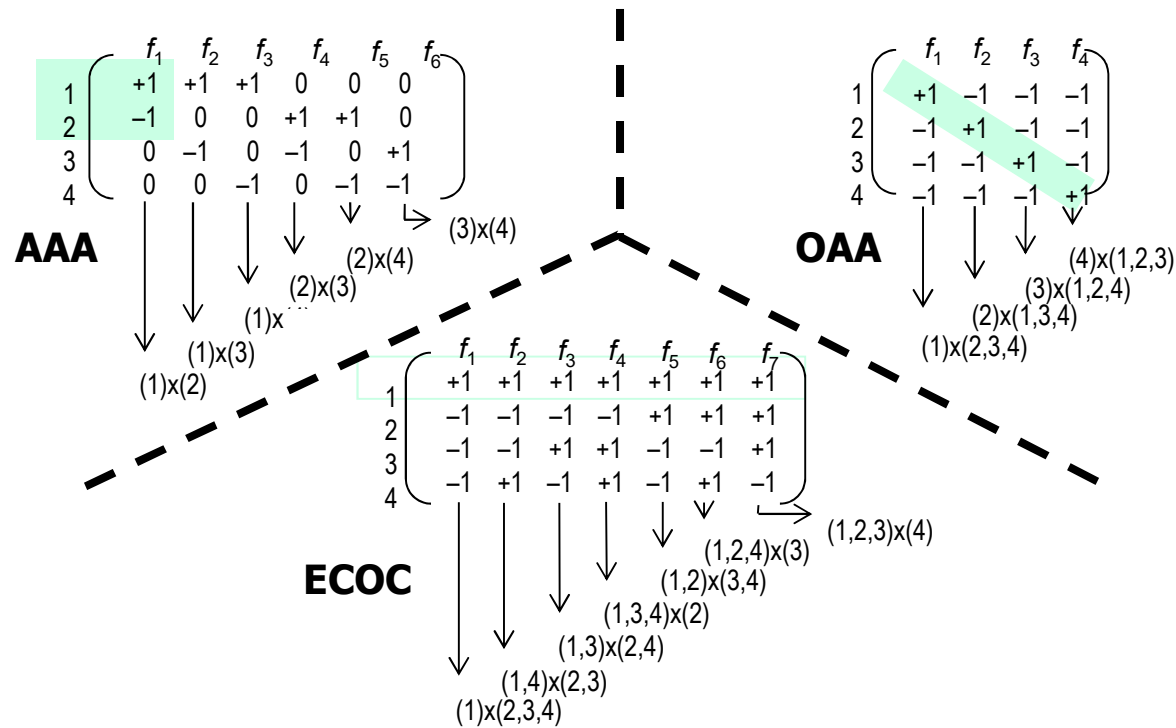


Problemas não linearmente separáveis

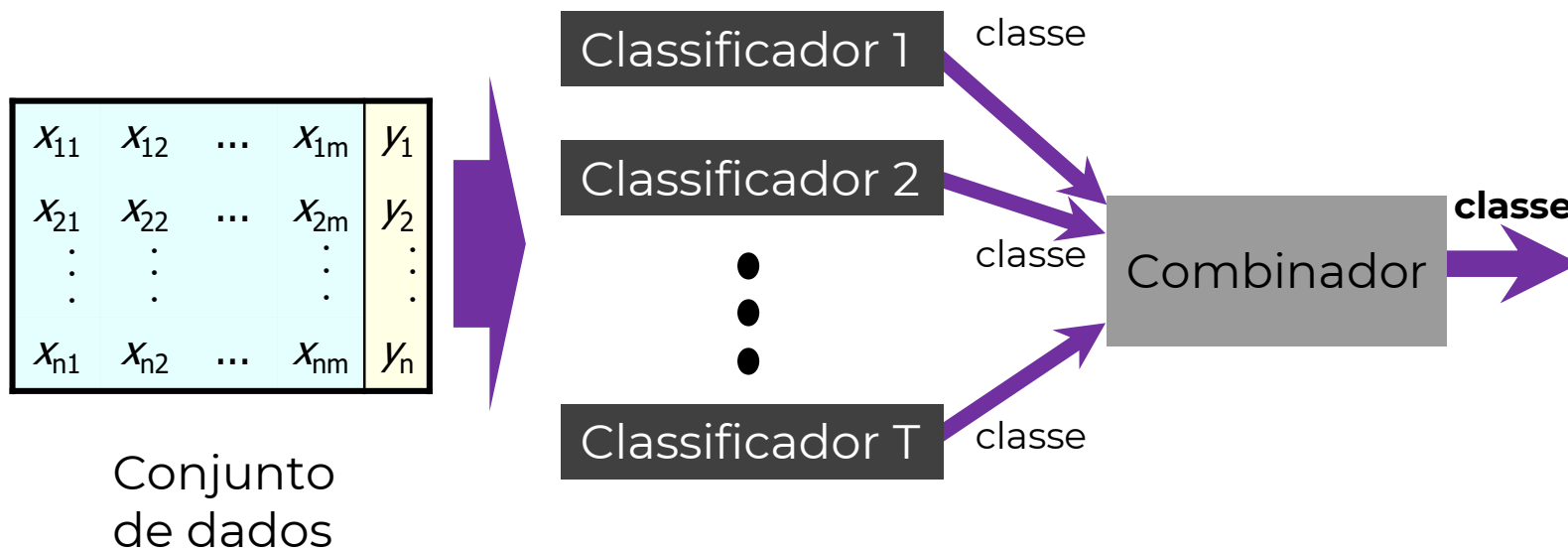
- Generalização para problemas não lineares
 - Mapeamento de dados de entrada para um espaço de maior dimensão



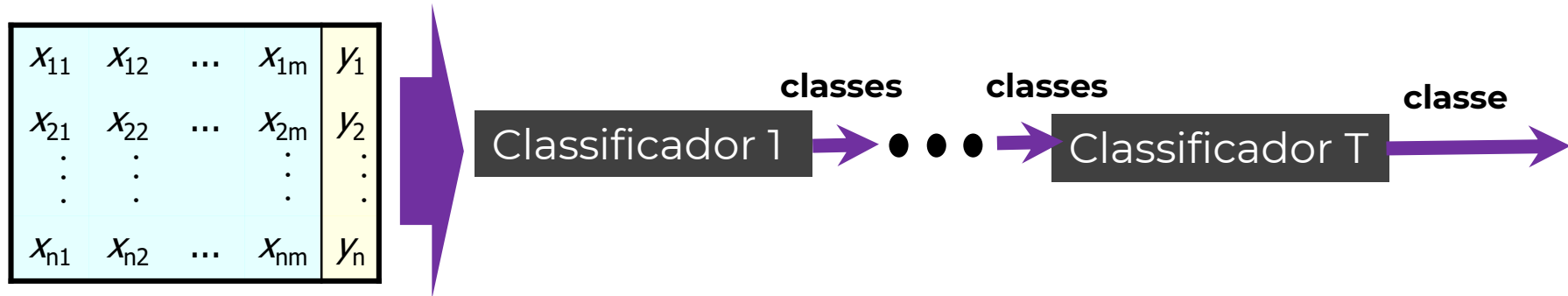
Matrizes de códigos



Combinação paralela

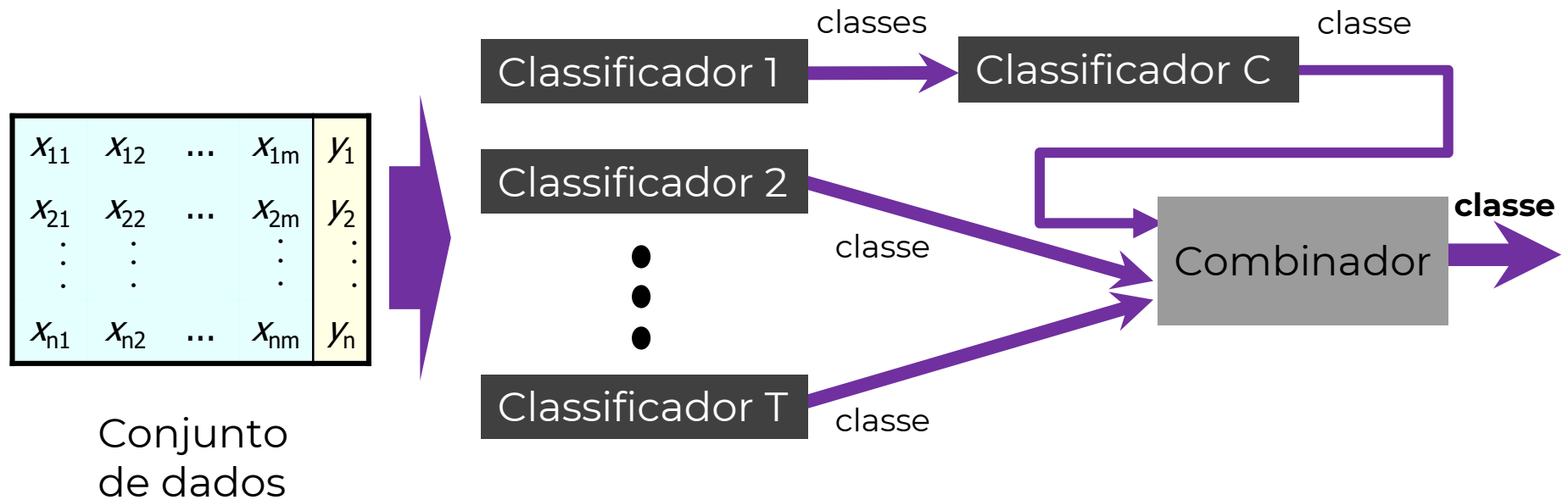


Combinação sequencial (em cascata, pipeline)

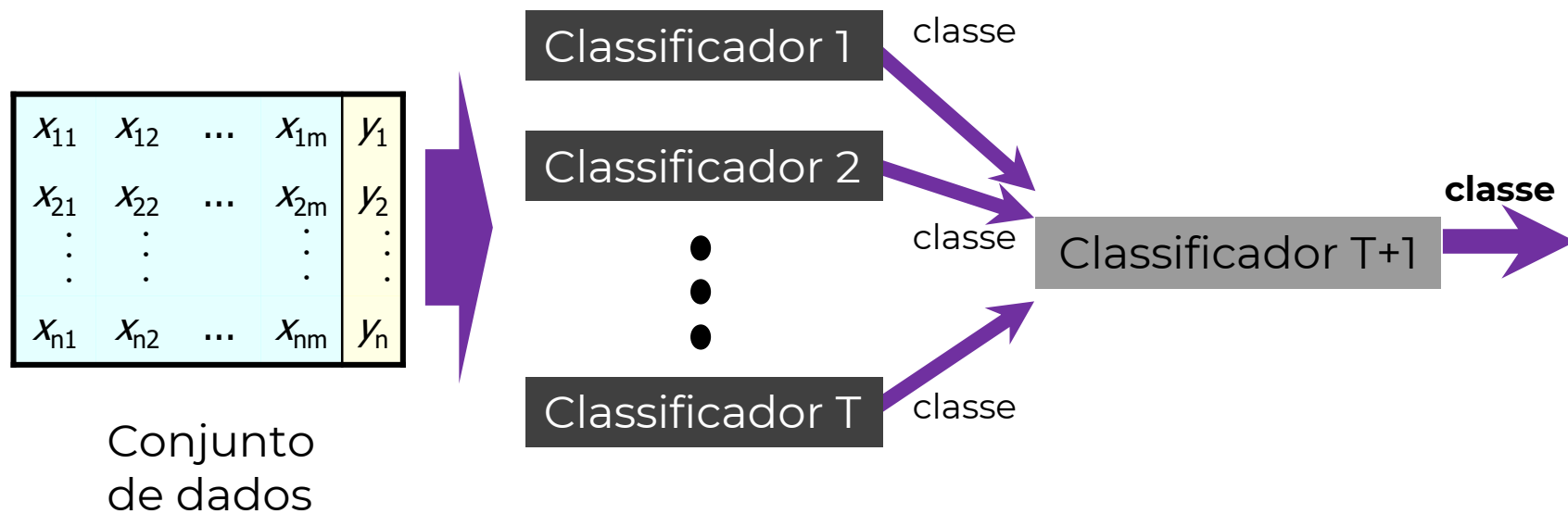


Conjunto
de dados

Combinação híbrida (hierárquica)



Stacking



Bagging

Amostra: $x_8, x_6, x_3, x_5, x_3, x_{10}$

Amostra: $x_3, x_7, x_1, x_5, x_5, x_1$

⋮

Amostra: $x_6, x_2, x_4, x_9, x_6, x_4$

Classificador 1

Classificador 2

⋮

Classificador T

classe

classe

classe

Combinador

classe

Boosting

- Seja um conjunto de dados treinamento formado por 5 exemplos: $\{x_1, x_2, x_3, x_4, x_5\}$

Exemplos	x_1	x_2	x_3	x_4	x_5
Pesos atuais	0,20	0,20	0,20	0,20	0,20
Classificação	Correta	Incorreta	Correta	Correta	Incorreta
Novos pesos	0,10	0,35	0,10	0,10	0,35

Exemplos	x_1	x_2	x_3	x_4	x_5
Pesos atuais	0,10	0,35	0,10	0,10	0,35
Classificação	Correta	Incorreta	Correta	Incorreta	Correta
Novos pesos	0,00	0,60	0,00	0,25	0,15

Algoritmos para comitês de árvores de decisão

- Combinam a predição de várias árvores de decisão (ADs), usando:
 - Algoritmos baseados em Bagging
 - Random forests (1995)
 - Combina as árvores no final do processo de treinamento
 - Algoritmos baseados em Boosting
 - Extreme gradient boosting (2014)
 - Começa a combinar as árvores no início do processo de treinamento
 - LightGBM (2017)
 - CatBoost (2017)

Aprendizado de máquina de ponta-a-ponta

Inclui vários aspectos

Lidar com
valores ausentes

Lidar com dados
desbalanceados

Extrair atributos

Selecionar
atributos



Escolher/Modificar
algoritmo de AM

Ajustar
hiperparâmetros

Verificar overfitting

Descobrir bugs

Aprendizado de máquina de ponta-a-ponta

Inclui vários aspectos interdependentes

Lidar com valores ausentes

Lidar com dados desbalanceados

Extrair atributos

Selecionar atributos



Adaptado de Rick Caruana, Research opportunities in AutoML Microsoft Research

Escolher/Modificar algoritmo de AM

Ajustar hiperparâmetros

Verificar overfitting

Descobrir bugs

A solução mais apropriada

- Nova hipótese:
 - É possível selecionar não apenas o algoritmo mais adequado para uma nova tarefa, mas também
 - Técnicas de pré-processamento (pós-processamento)
 - Valores para os hiperparâmetros
 - ...
- Aprendizado de máquina automático (automatizado) – AutoML

AutoML

- Automatiza aplicação de aprendizado de máquina a problemas reais
 - Apoia tanto leigo quanto especialista
- Engloba vários tópicos:
 - Otimização Bayesiana
 - Otimização combinatória
 - Aprendizado de máquina
 - Meta-aprendizado
 - Transferência de aprendizado

Principais abordagens de AutoML

- Otimização
 - Algoritmos e/ou hiperparâmetros
 - Propõe o que pode não existir
- Meta-aprendizado
 - Algoritmos e/ou hiperparâmetros
 - Seleciona entre o que já existe
- Híbrido
 - Combina abordagens anteriores

Fim da
apresentação