
Aprendizagem de máquina

IBM Data & AI Expert Services

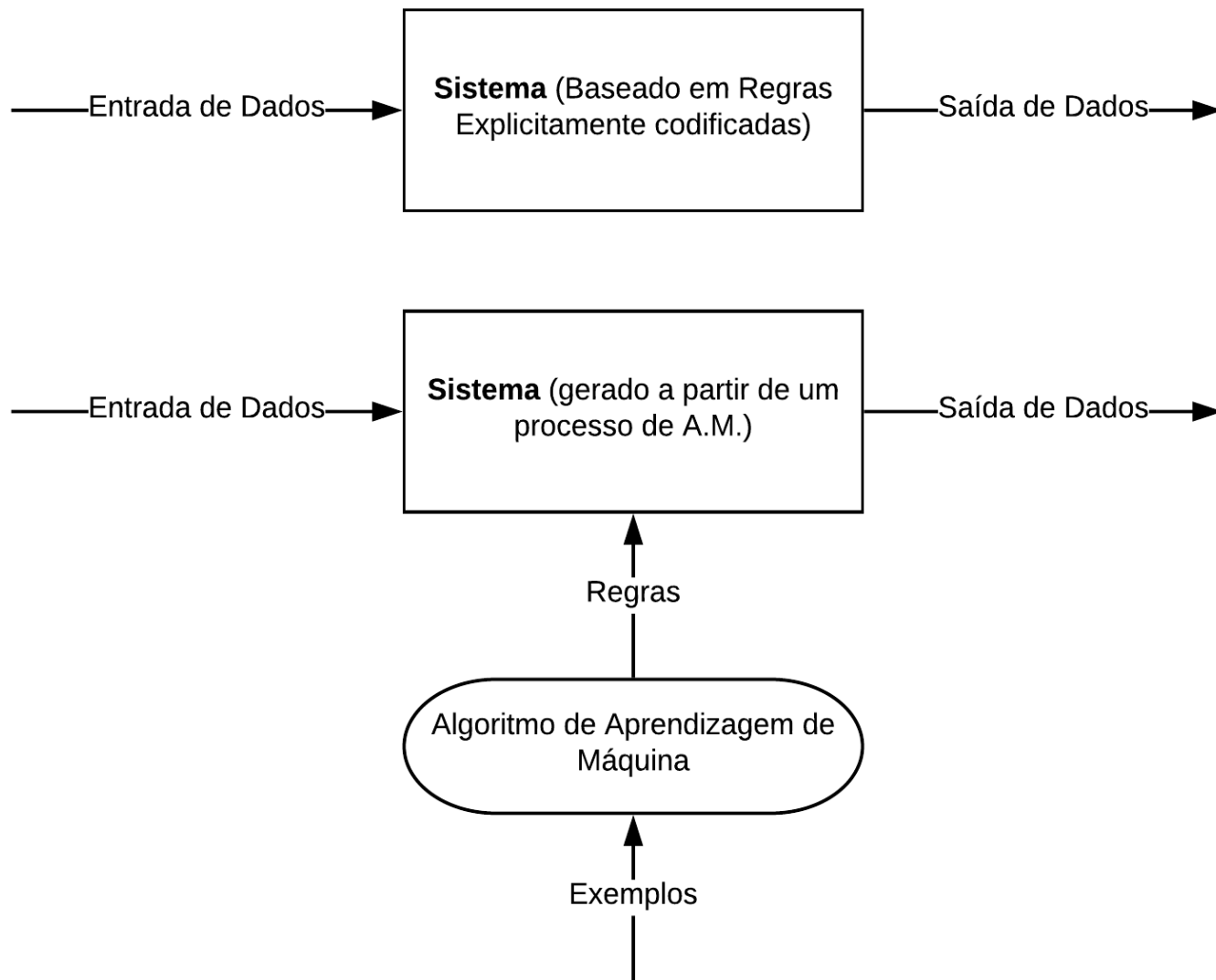
Outubro de 2019

Sumário

- O que é Aprendizagem de Máquina (A.M.)?
- Hierarquia de aprendizado.
- Exemplos de aprendizagem supervisionada (modelos preditivos).
- Exemplos de aprendizagem não supervisionada (modelos descritivos).
- Tarefas mais adequadas para A.M.
- Referências e exercícios.

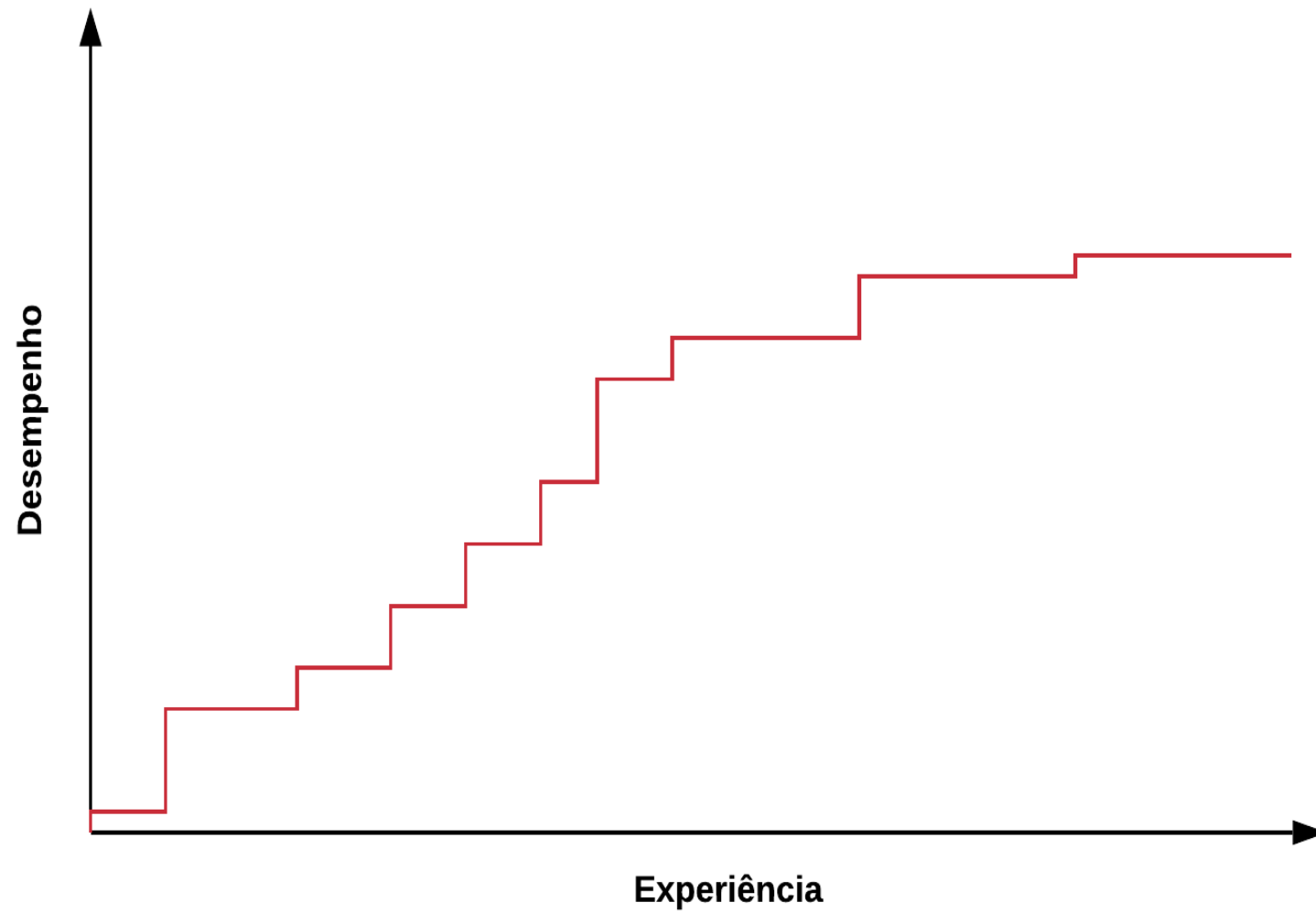
O que é Aprendizagem de Máquina?

- Área de estudo que fornece aos computadores a habilidade de aprender sem serem explicitamente programados [Arthur Samuel (1959)].
- Área de estudo que permite desenvolver sistemas computacionais sem explicitamente programá-los.



O que é Aprendizagem de Máquina?

- Um programa de computador aprende a partir de uma experiência E a realizar uma tarefa T . A medida que a sua experiência aumenta deve-se esperar que o desempenho D para realizar a tarefa também aumente. [Tom Mitchell (1998)].

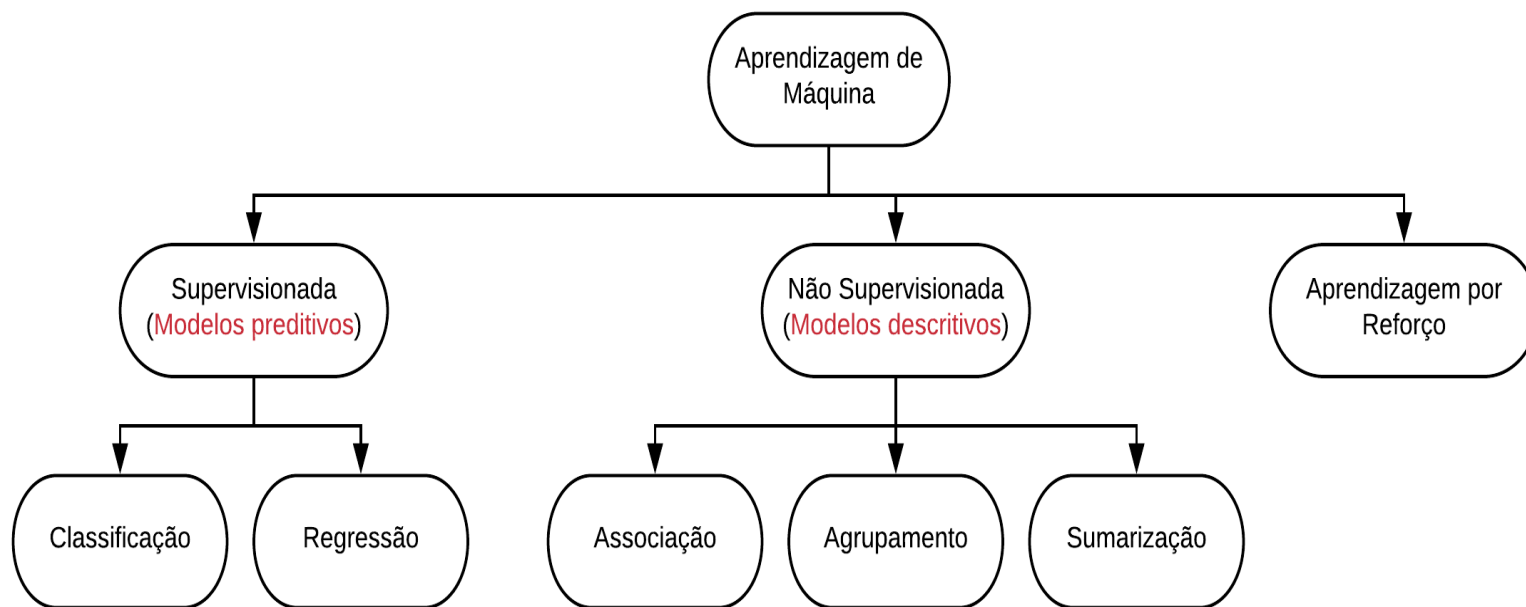


Exercício

O que cada uma das sentenças abaixo descreve segundo a definição do Tom Mitchell?

- Classificar e-mails como spam ou não spam.
- Verificar quais e-mails o usuário classifica como spam.
- O número (ou fração) de e-mails corretamente classificados como spam ou não spam.

Hierarquia de aprendizado



Exemplos de aprendizagem supervisionada

- Estimar o preço de uma casa.
 - ★ atributos: tamanho, posição geográfica, material.
 - ★ classe: preço (regressão).
- Determinar se uma pessoa tem câncer benigno ou maligno.
 - ★ atributos: tamanho do tumor, formato do tumor, idade do paciente.
 - ★ classe: tumor benigno ou tumor maligno (classificação).

-
- Determinar se é um texto publicado em uma rede social é inadequado ou não.
 - ★ atributos: quantidade de palavras encontradas no texto, quantidade de palavras proibidas encontradas no texto, quantidade de textos já criados pelo usuário, idade do usuário no sistema, quantidade de textos criados pelo usuário e moderados, ...
 - ★ classe: texto adequado ou não (classificação).
 - ★ **classes**: texto adequado, texto inadequado, texto com propaganda (classificação com múltiplas classes).

Exemplo de dataset com **classe**

Idade	Miopia	Astigmat.	Lacrimej.	Lentes
jovem	míope	não	reduzido	nenhuma
jovem	míope	não	normal	fraca
jovem	míope	sim	reduzido	nenhuma
jovem	míope	sim	normal	forte
...
adulto	míope	não	reduzido	nenhuma

Exercícios

Que problema deve ser tratado como problema de regressão e que problema deve ser tratado como problema de classificação?

- A sua empresa possui 1.000 itens idênticos em estoque. Você quer prever quantos destes itens serão vendidos nos próximos três meses.
- Você quer examinar clientes seus e para cada um decidir se ele irá pagar todo o financiamento ou não.

Exemplos de aprendizagem não supervisionada

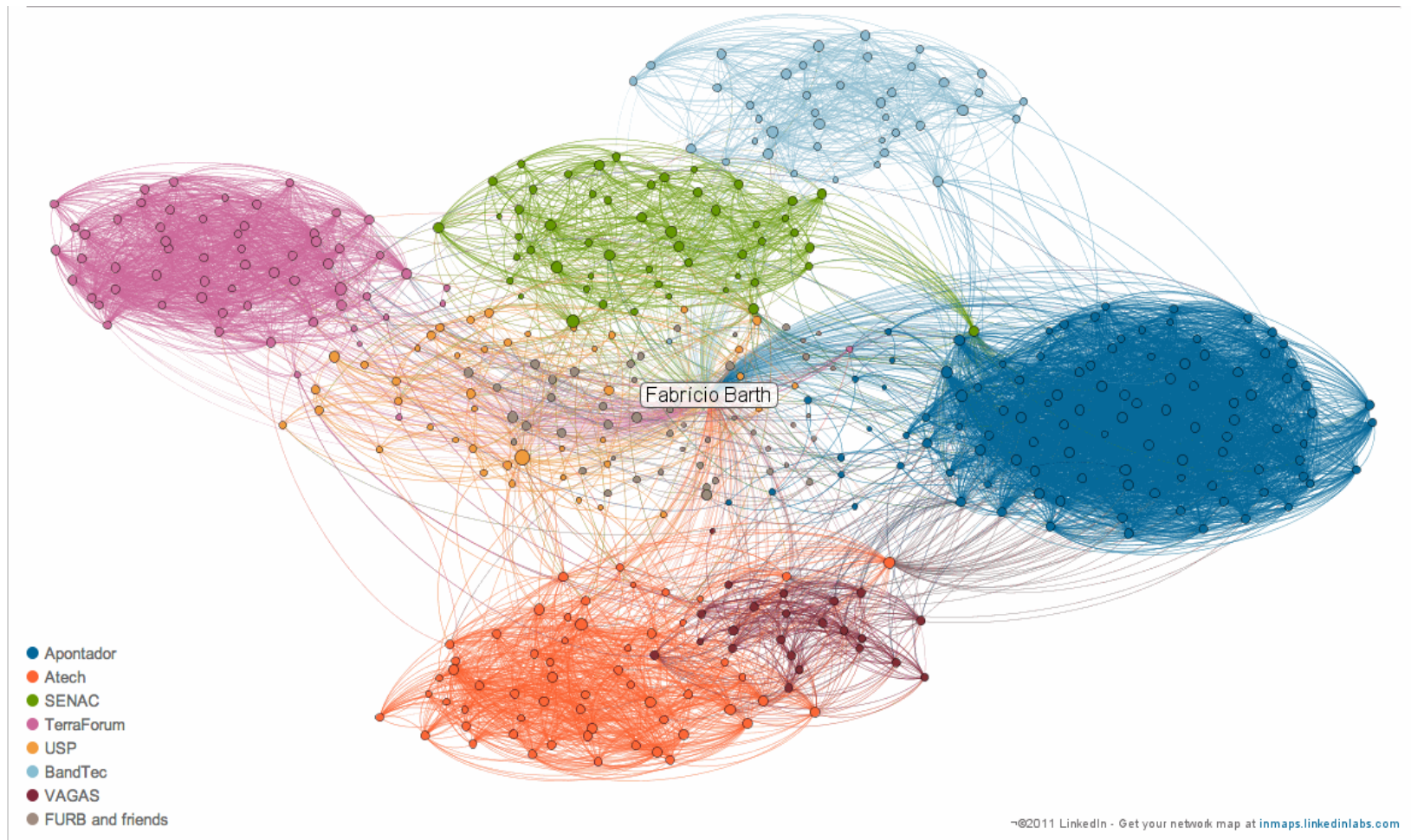
- Dado conjuntos de itens adquiridos na mesma compra, identificar padrões de compra.
- Identificar padrões de navegação em sites.
- Agrupar notícias semelhantes publicadas por várias fontes de informação.
- Em uma rede social, identificar sub-grupo de pessoas.

Exemplos de aprendizagem não supervisionada

Table 1: Exemplo de tabela com as transações dos usuários

usuário	$categoria_1$	$categoria_2$	$categoria_3$	\dots	$categoria_m$
$user_1$	0	2	0	\dots	1
$user_2$	1	1	0	\dots	0
$user_3$	2	0	1	\dots	0
$user_4$	0	1	0	\dots	0
\dots	\dots	\dots	\dots	\dots	\dots
$user_n$	1	1	0	\dots	1

Exemplo de identificação de grupos em redes sociais



Exercícios

Quais dos problemas abaixo você iria resolver com uma abordagem não supervisionada de aprendizagem?

- Dado e-mails rotulados como spam e não spam, desenvolver um filtro de spam.
- Dado um conjunto de notícias encontradas na Internet, agrupá-las em conjunto de notícias que tratam do mesmo assunto.

-
- Dado uma base de clientes, descobrir segmentos de clientes.
 - Dado uma base de pacientes diagnosticados com diabetes ou não, aprender a classificar novos pacientes com diabetes ou não.

Tarefas mais adequadas para AM

1. Aprende uma função que mapeia entradas e saídas muito bem definidas;
2. É necessário existir [grandes] datasets digitais ou deve ser fácil criar estes pares de entrada e saída;
3. O objetivo e as métricas para avaliação da tarefa devem ser claros;

-
4. Para a execução da tarefa não deve existir longas cadeias de decisão ou raciocínio que depende de conhecimento de fundo ou senso comum.
- Planejamento da reforma de uma casa **vs** Sistema que identifica por imagens quando uma casa precisa de reforma.
 - Waze **vs** Assistente que planeja todas as férias do usuário.

-
5. Não é necessário explicar como a decisão foi feita.
 - Existem diversos trabalhos acadêmicos e algumas iniciativas comerciais sobre este assunto. No entanto, as soluções disponíveis ainda são muito fracas.
 6. Não se espera uma solução ótima e o caso de uso deve ser tolerante ao erro.
 7. O conceito ou a função a ser aprendida não pode mudar rapidamente ao longo do tempo.
 8. A tarefa não deve ter necessidade de habilidades manuais ou relacionadas com mobilidade.

Material de **consulta**

- Tom Mitchell. Machine Learning, 1997.
- Ian H. Witten and Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques (Third Edition), 2011.
- Capítulo 1 e 2 do livro Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An Introduction to Statistical Learning with Applications in R. Springer, 4th edition, 2014.

-
- Andrew Ng. Machine Learning. Coursera.
<https://www.coursera.org/learn/machine-learning>
 - *Faceli, Lorena, Gama, Carvalho. Inteligência Artificial: uma abordagem de aprendizado de máquina, 2011.*