
Aprendizagem de máquina

Fabrício Barth

Maio de 2020

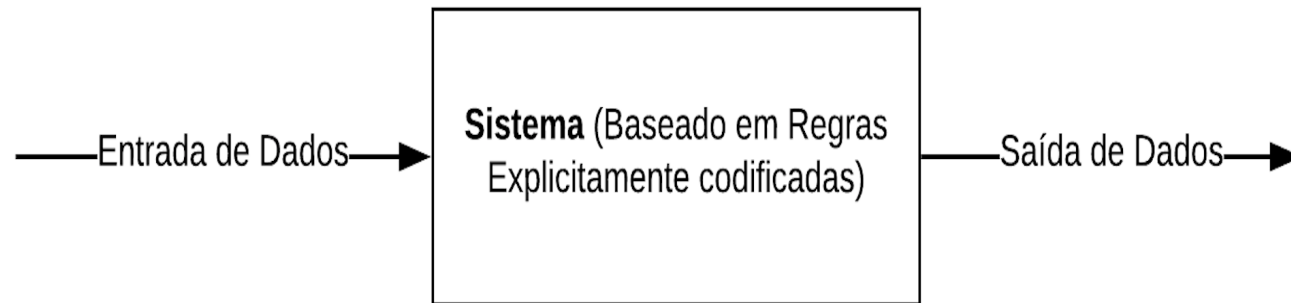
Sumário

- O que é Aprendizagem de Máquina?
- Hierarquia de aprendizado.
- Exemplos de aprendizagem supervisionada (modelos preditivos).
- Exemplos de aprendizagem não supervisionada (modelos descritivos).
- Exemplo de aprendizagem por reforço
- 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)].

Abstração para a maioria dos sistemas hoje em produção.



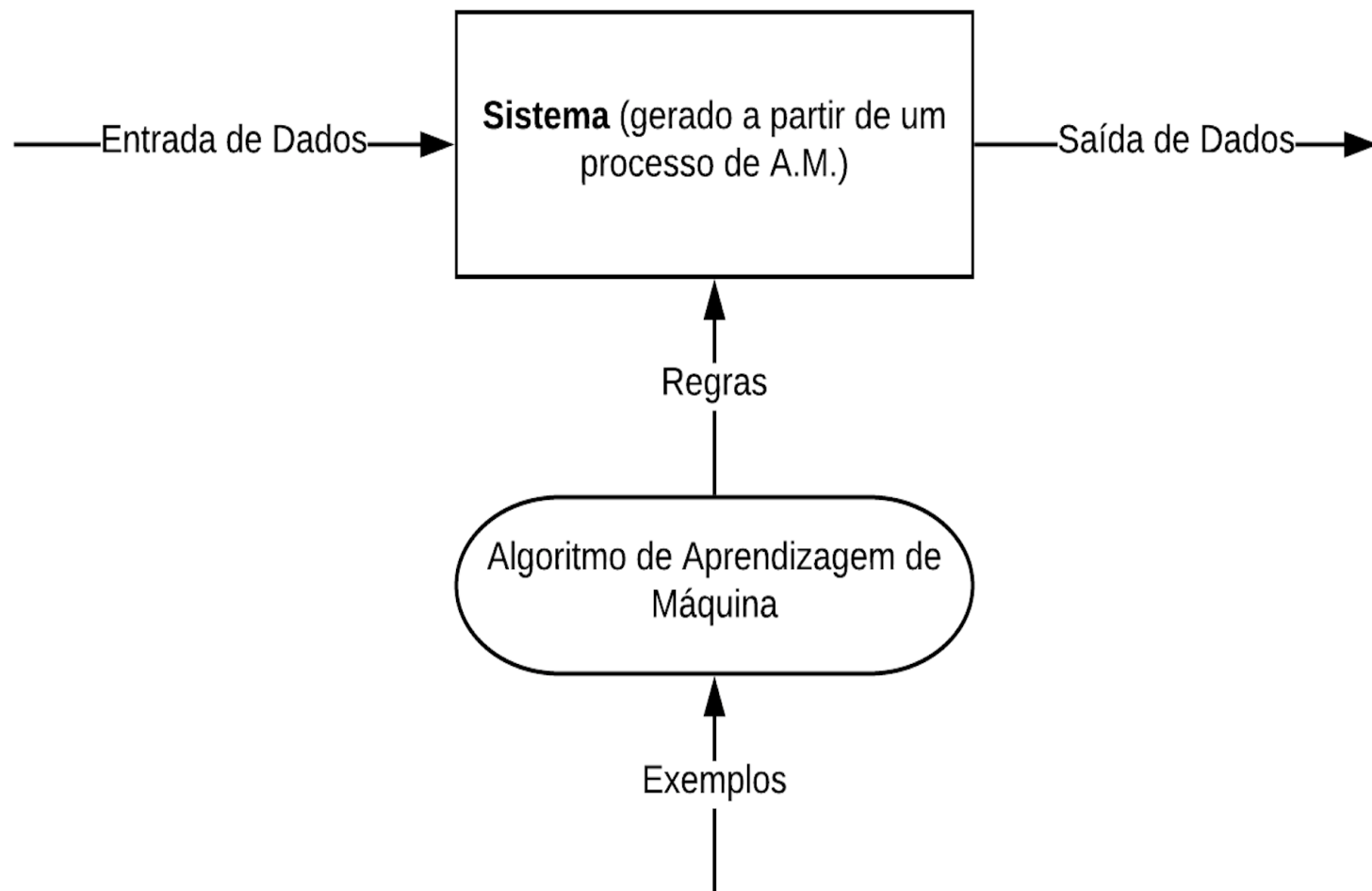
Método normalmente empregado:

- Identificação dos requisitos.
- Especificação das regras.
- Implementação das regras.
- Validação da implementação.
- *Deploy* em produção.

O que fazer quando não se sabe as regras **a priori**?

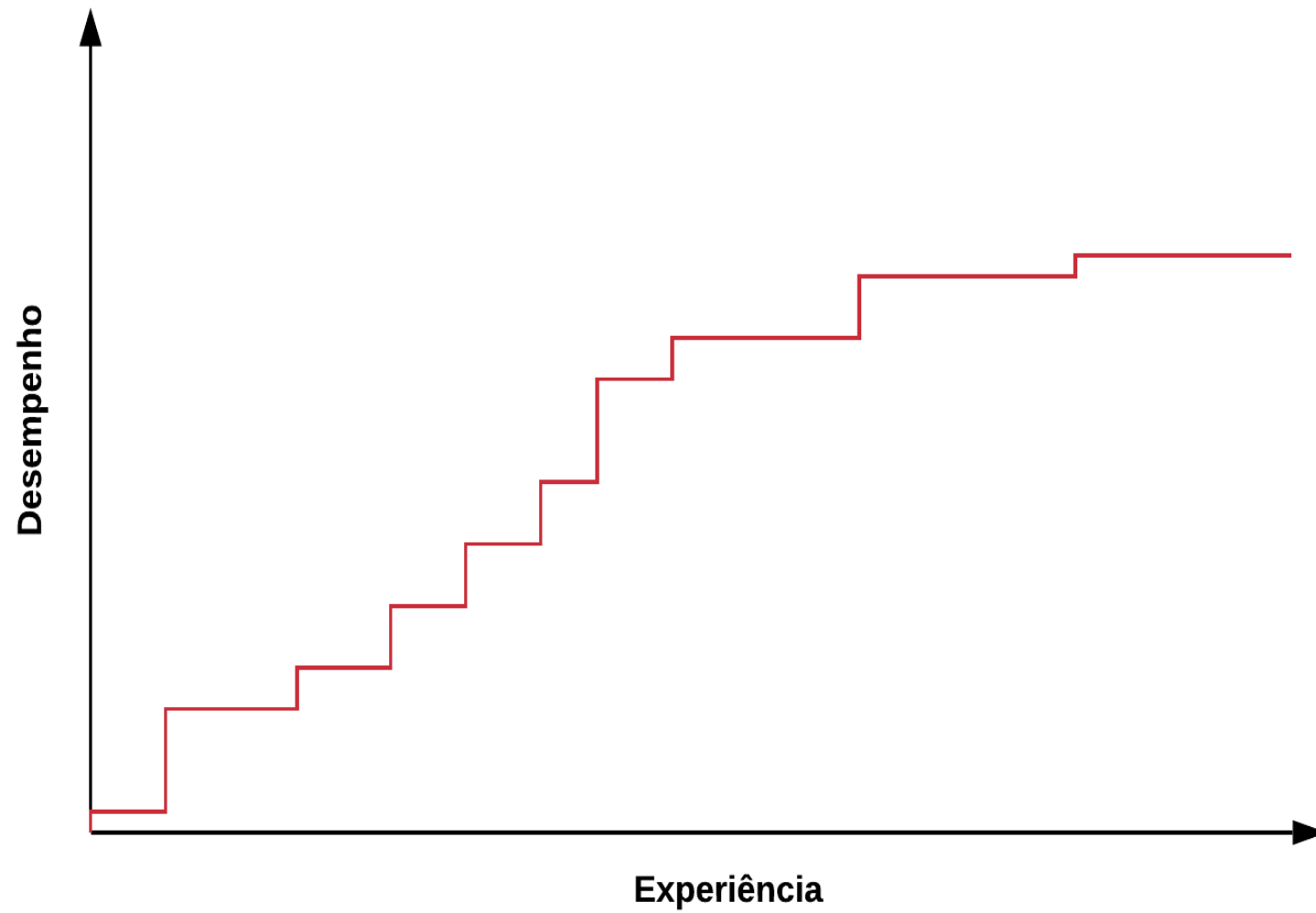
Exemplos de soluções onde é muito difícil ou impossível determinar todas as regras a priori.

- Sistemas de Recomendação.
- Sistemas para identificação de spam.
- Sistemas para identificação de fake news.
- Jogadores de Xadrez, Shogi ou Go.
- Sistema para identificação de fraude.



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)].

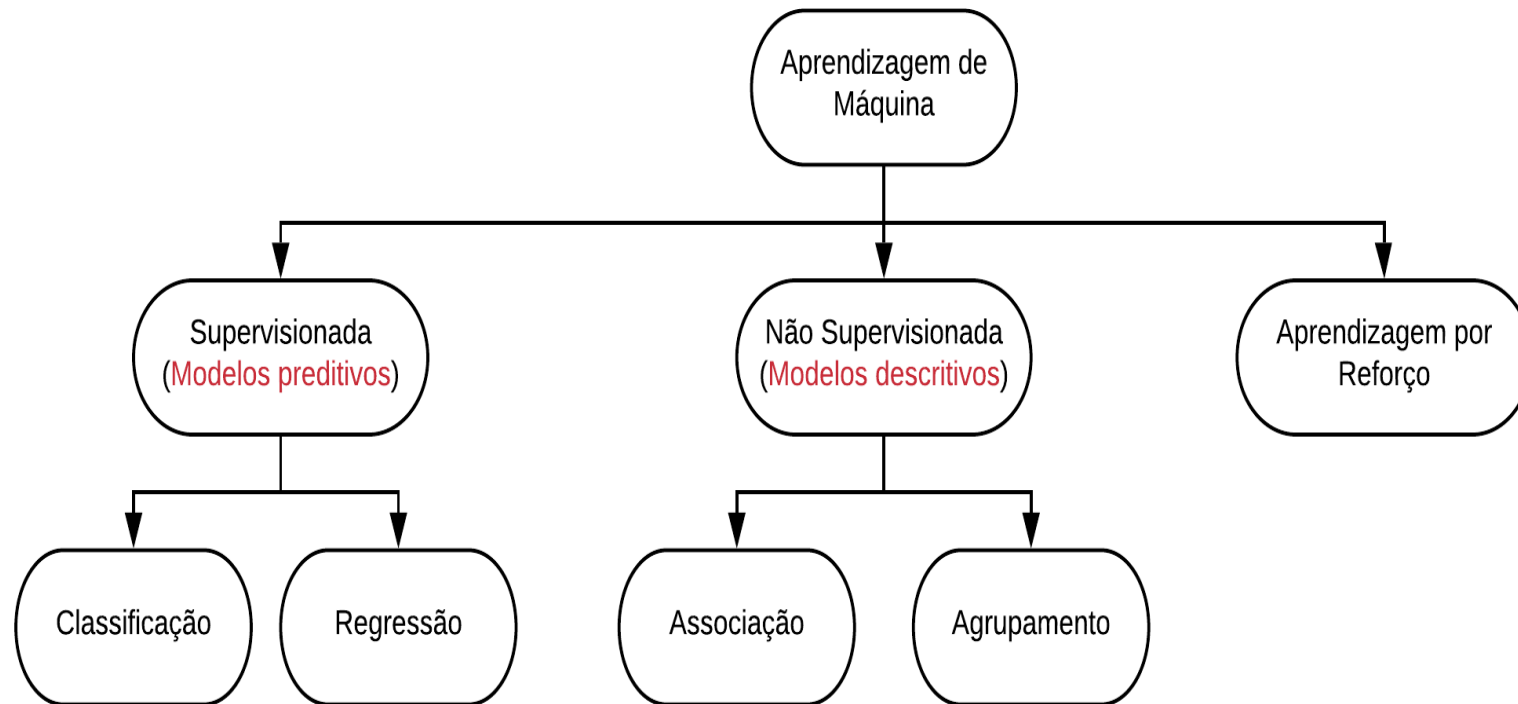


Exercício

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



Aprendizagem de Máquina Supervisionada

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.

Aprendizagem de Máquina Não Supervisionada

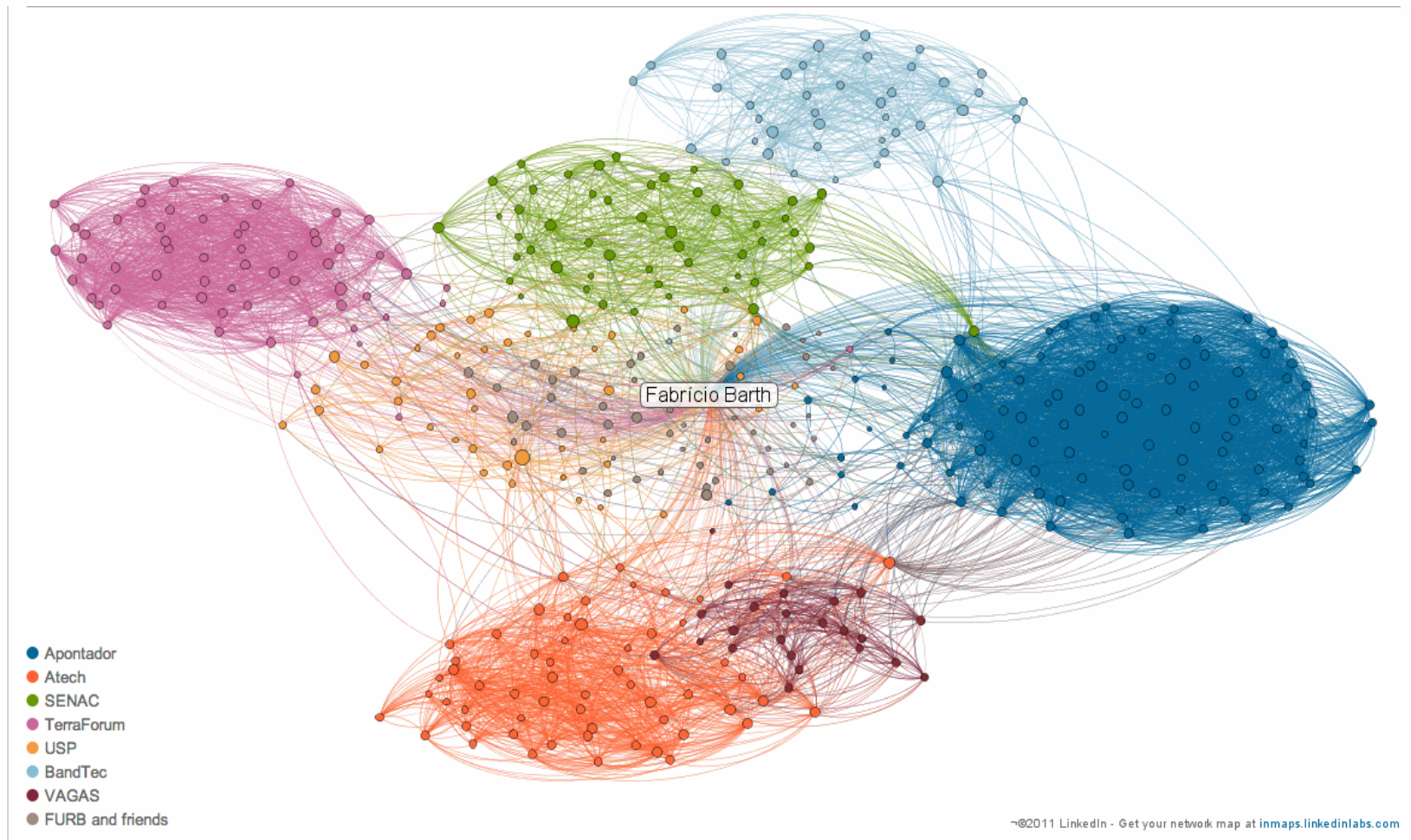
Exemplos de aprendizagem não supervisionada

- 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.

Exemplo de Regras de Associação

- **Caso do supermercado** (fralda \rightarrow cerveja)
- Quem acessa a página sobre futebol também acessa a página de volei em **90%** dos casos (futebol \rightarrow volei).
- Quem acessa a página de ofertas e a página de material de construção também finaliza a compra em **83%** dos casos (ofertas \wedge material_construção \rightarrow compra)

Exemplo de identificação de grupos em redes sociais



Exemplos de dataset utilizado em 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

Exercícios

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

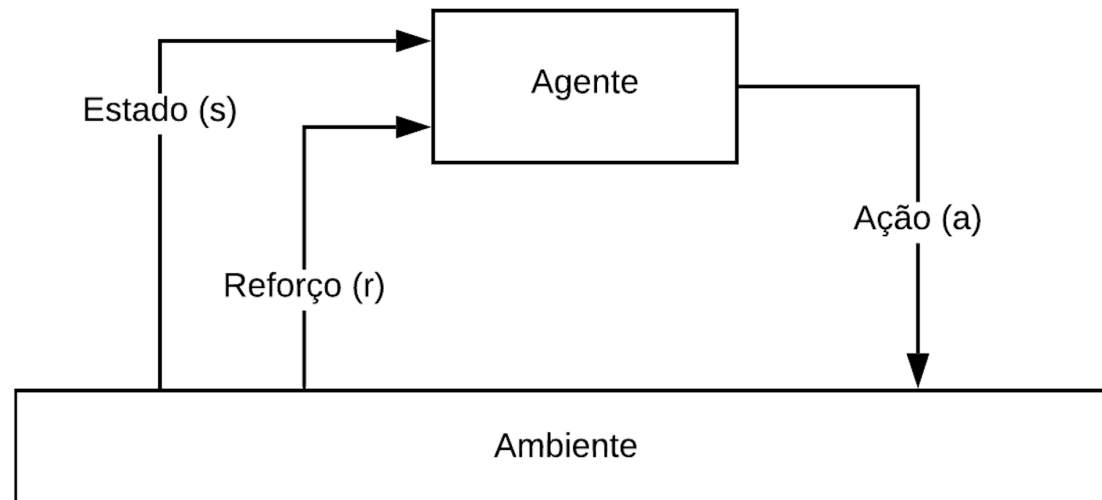
- Dado e-mails rotulados como spam e não spam, desenvolver um filtro de spam.
- Dado um conjunto de notícias encontradas na Internet, agrupá-las em conjunto de notícias que tratam do mesmo assunto.

-
- 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.

Aprendizagem por Reforço

Visão Geral

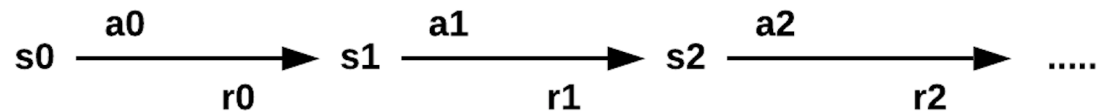
Um agente aprende a resolver uma tarefa através de repetidas interações com o ambiente, por tentativa e erro, recebendo (esporadicamente) reforços (punições ou recompensas) como retorno.



Visão Geral

- Este agente não tem conhecimento algum sobre a tarefa que precisa executar (heurísticas ou funções de utilidade específicas).
- A **tarefa** deste agente é executar uma **sequência de ações**, observar as suas **consequências** e aprender uma **política de controle**.

-
- A política de controle desejada é aquela que **maximiza** os reforços (*reward*) acumulados ao longo do tempo pelo agente: $r_0 + \gamma r_1 + \gamma^2 r_2 + \dots$ onde $0 \leq \gamma < 1$.



Material de **consulta**

- Tom Mitchell. Machine Learning, 1997.
- Iah H. Witteh 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.

-
- Richard Sutton and Andrew Barto. Reinforcement Learning: An Introduction. Second Edition, in progress. The MIT Press, 2015.
 - 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.*