

Introdução ao Aprendizado de Máquina

Prof. Erneson A. Oliveira

MBA em Ciência de Dados
Universidade de Fortaleza

1 de Fevereiro de 2020



Aula 1 - Introdução ao Aprendizado de Máquina



O que são Dados?

Dados são unidades de informação.

O que é Ciência?

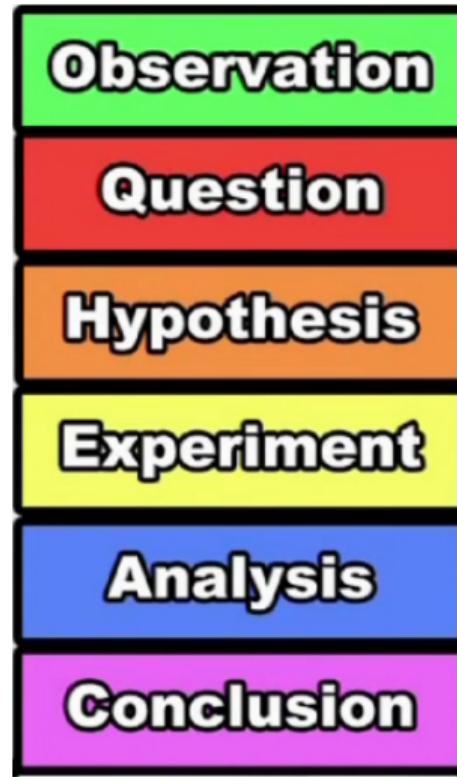
Ciência é aquisição de conhecimento através do Método Científico.

O que é Método Científico?

Método Científico é um conjunto de regras que produzem o conhecimento científico.



Ibn Al-Haytham (965-1040)

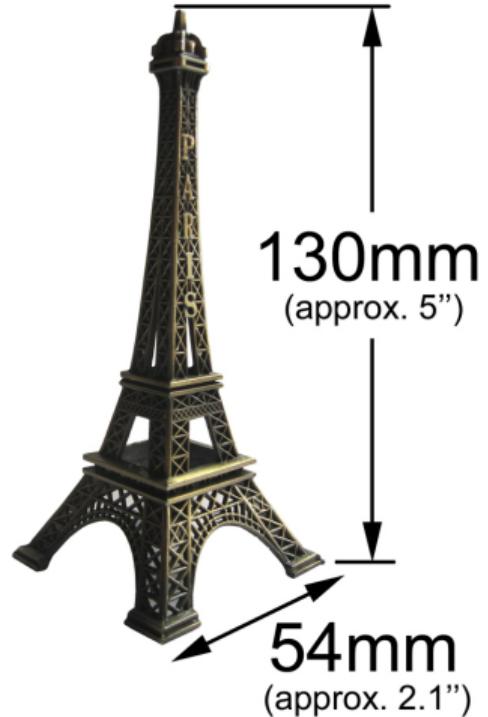


O que é Modelo?

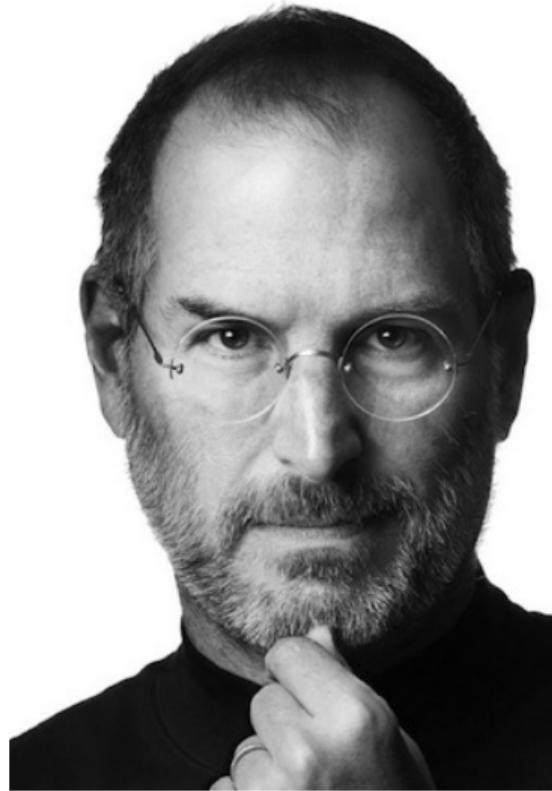
Modelo é uma representação simplificada da realidade.

Modelagem descritiva



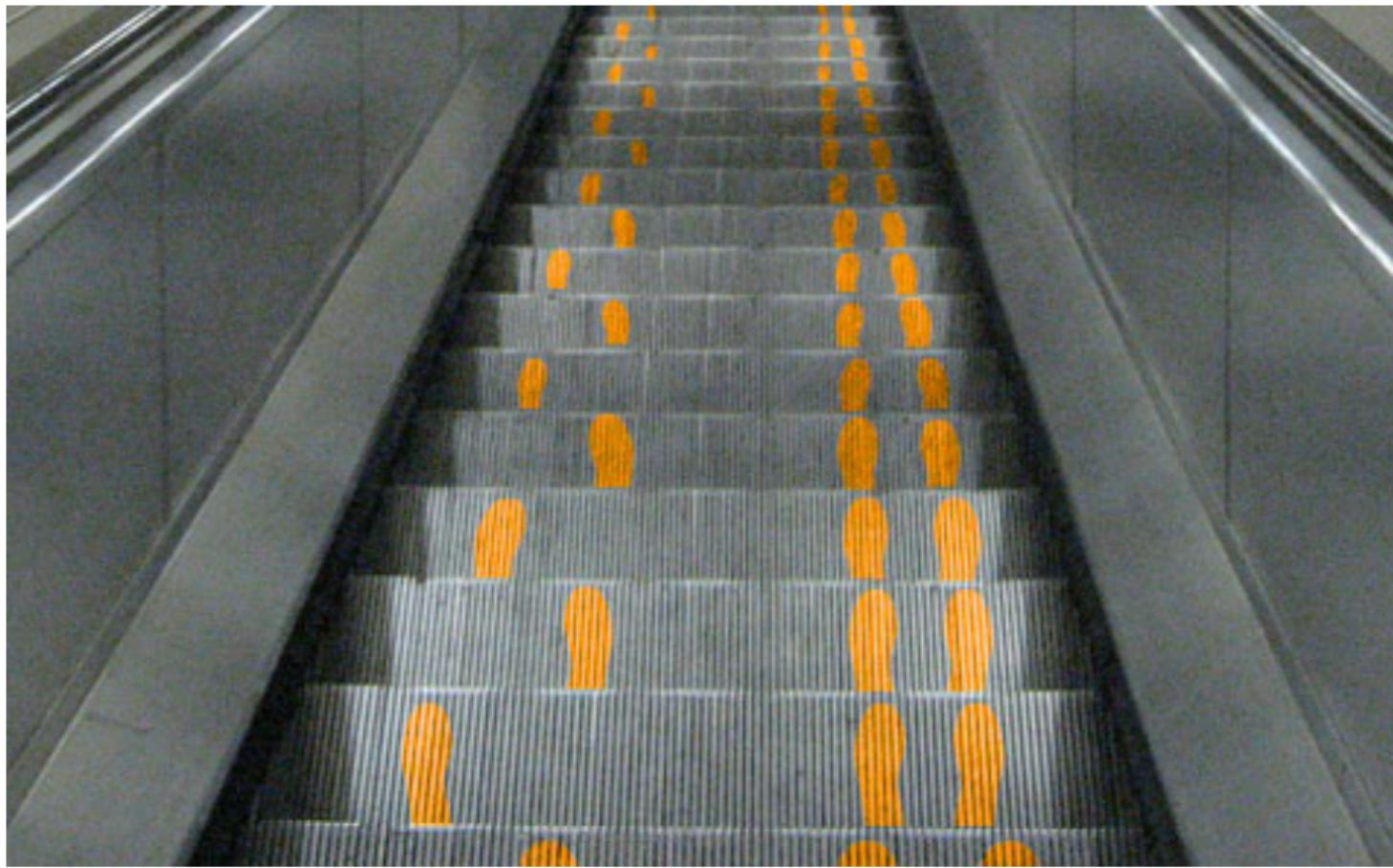


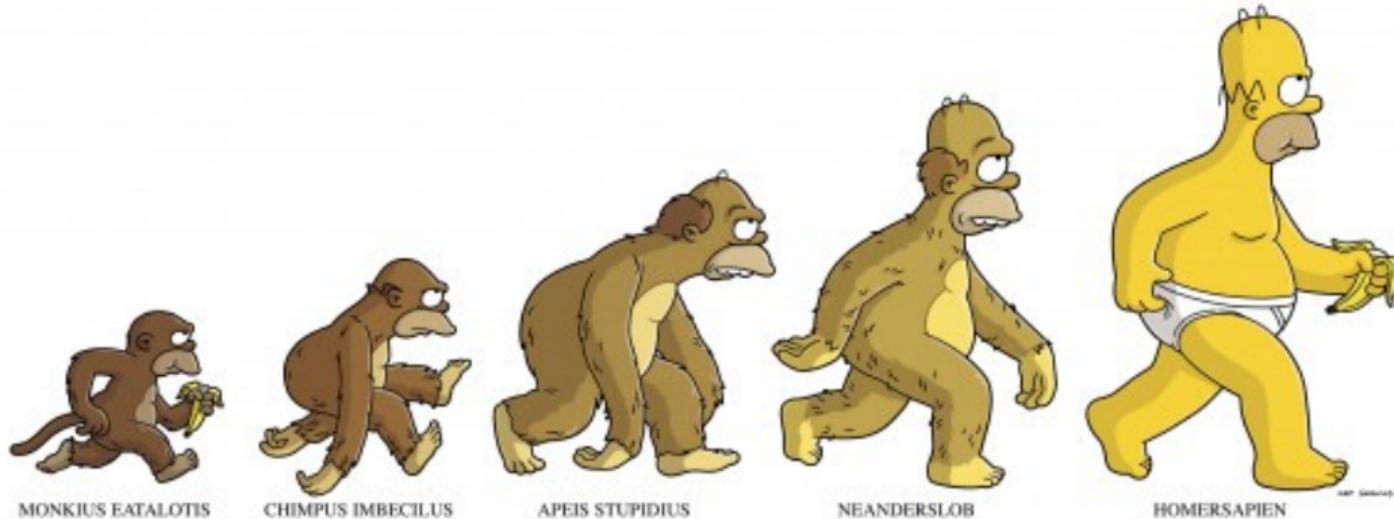
Steve Jobs by Walter Isaacson



Modelagem baseada em regras

$$\vec{F} = m\vec{a}$$





HOMERSAPIEN

Como construir um modelo?

1. Observar o sistema do seu interesse;

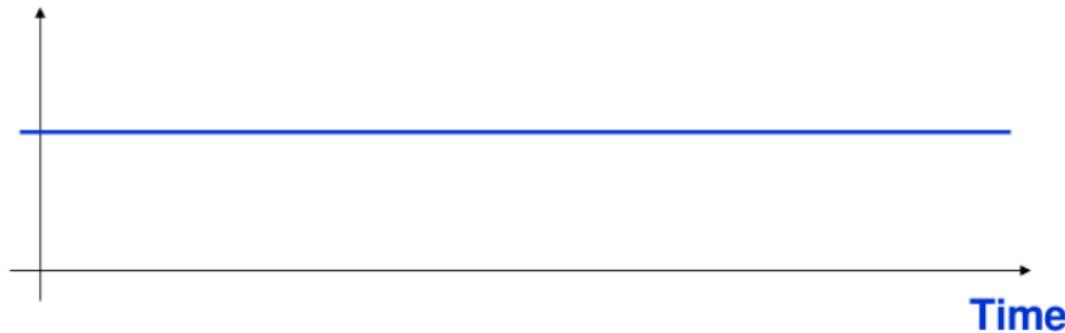
1. Observar o sistema do seu interesse;
2. Refletir sobre as possíveis regras que poderiam causar as características do sistema que foram vistas na observação;

1. Observar o sistema do seu interesse;
2. Refletir sobre as possíveis regras que poderiam causar as características do sistema que foram vistas na observação;
3. Derivar predições daquelas regras e compará-las com a realidade;

1. Observar o sistema do seu interesse;
2. Refletir sobre as possíveis regras que poderiam causar as características do sistema que foram vistas na observação;
3. Derivar predições daquelas regras e compará-las com a realidade;
4. Repetir os passos acima para modificar as regras até que você esteja satisfeito com o modelo.

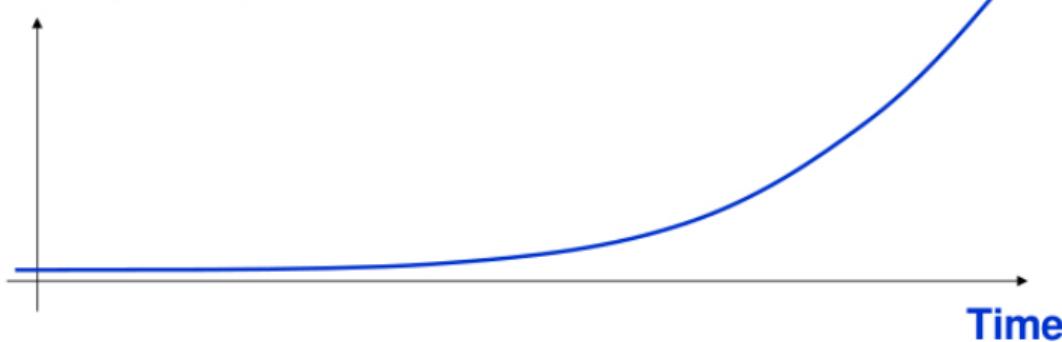
Exemplos

Some quantity



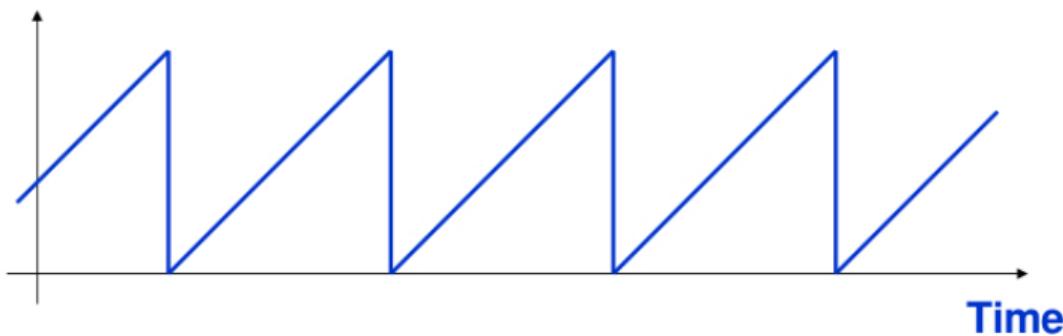
$$x(t) = C$$

Some quantity



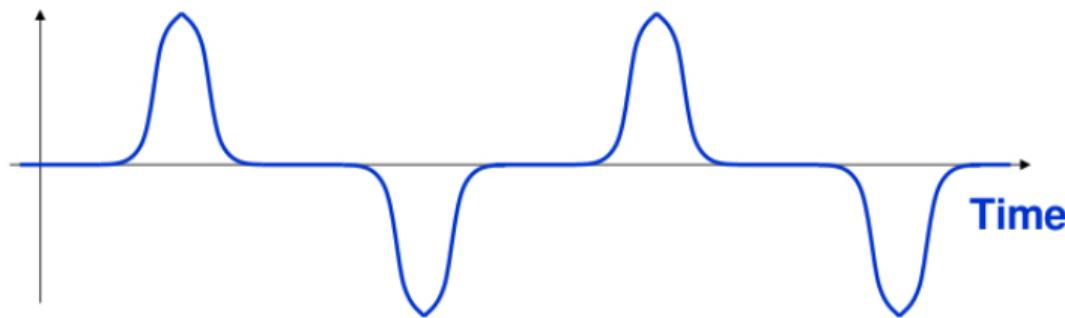
$$x(t) = ae^{bt}$$

Some quantity



$$x(t) = a \arctan[\cot(bt)]$$

Some quantity



$$x(t) = ?$$

O que é um bom modelo?

Um **bom modelo** é simples, válido e robusto.

O que é Ciência de Dados?

Ciência de Dados é extração de conhecimento a partir de dados.

O que é Aprendizado de Máquina (AM)?

O que é AM?

“Campo de estudo que dá ao computador a habilidade de aprender sem ser explicitamente programado.”

–Arthur Samuel, 1959



O que é AM?

“Diz-se que um programa de computador aprende com a experiência E em relação a alguma tarefa T e alguma medida de desempenho P, se seu desempenho em T, medido por P, melhora com a experiência E.”

–Tom Mitchell, 1997

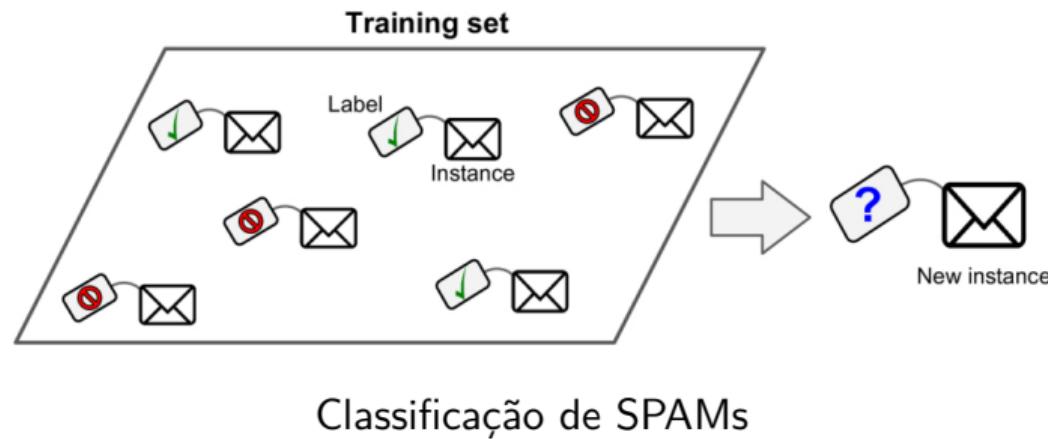


O que é AM?

Filtro de SPAM

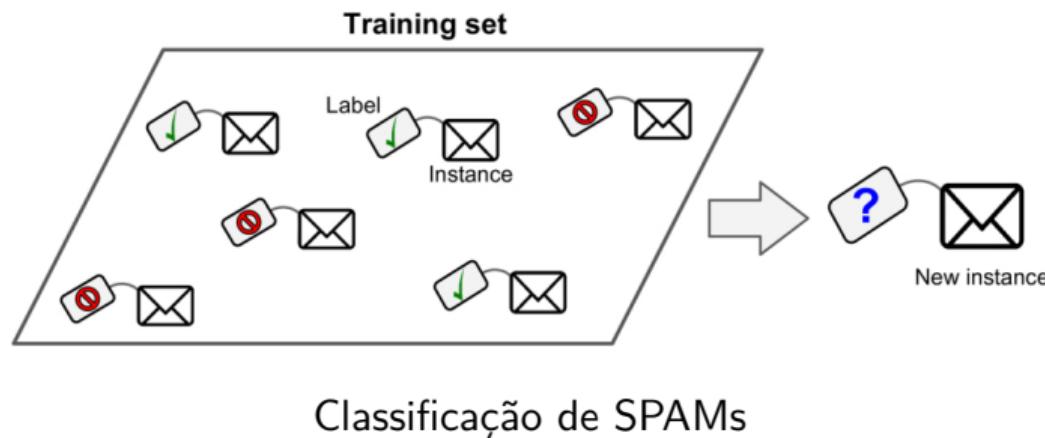
O que é AM?

- ▶ Conjunto de treinamento: Rótulos e instâncias.



O que é AM?

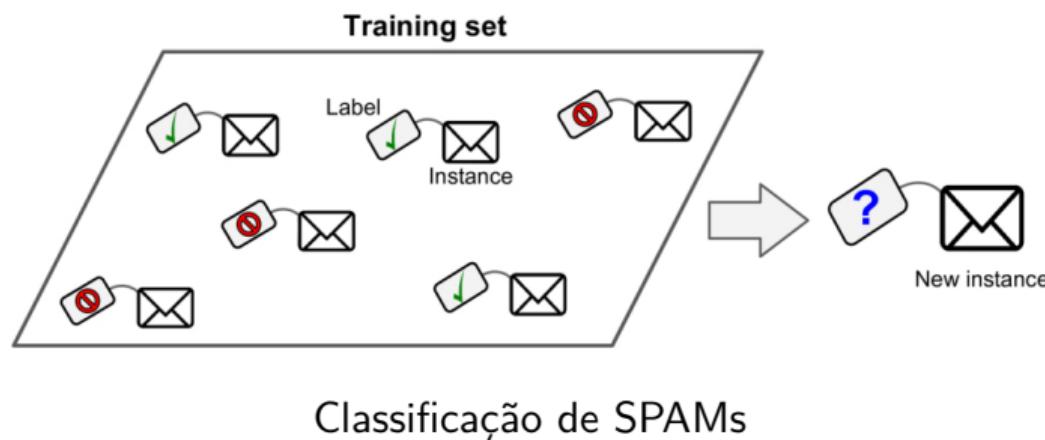
- ▶ Conjunto de treinamento: Rótulos e instâncias.



- ▶ Experiência (E): É o conjunto de treinamento;

O que é AM?

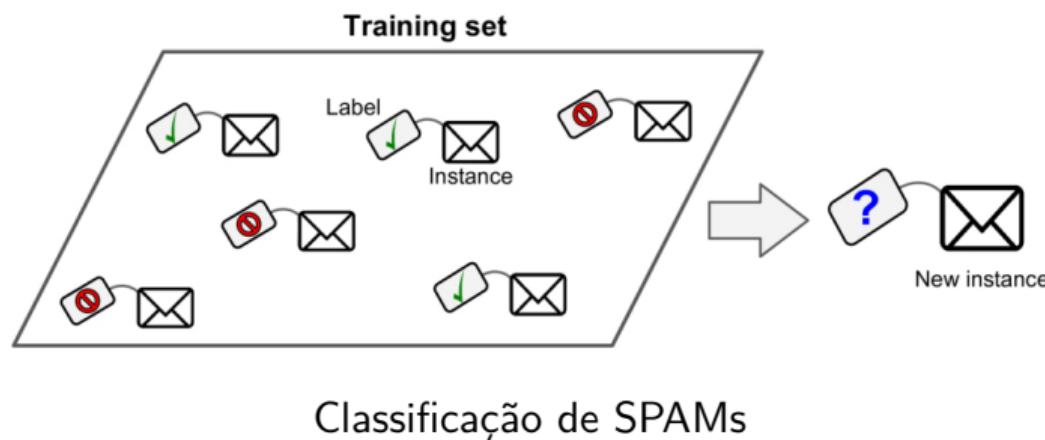
- ▶ Conjunto de treinamento: Rótulos e instâncias.



- ▶ Experiência (E): É o conjunto de treinamento;
- ▶ Tarefa (T): Ato de classificar novos e-mails em SPAM ou não-SPAM;

O que é AM?

- ▶ Conjunto de treinamento: Rótulos e instâncias.



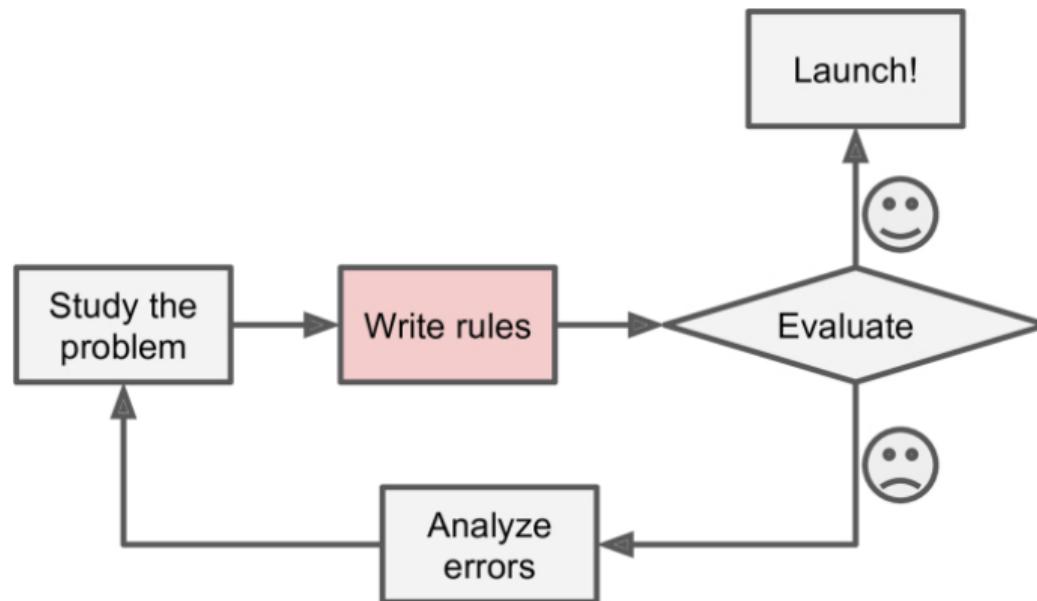
- ▶ Experiência (E): É o conjunto de treinamento;
- ▶ Tarefa (T): Ato de classificar novos e-mails em SPAM ou não-SPAM;
- ▶ Medida de desempenho (P): Medida para avaliar as classificações (e.g. Acurácia).

Qual é a vantagem de usar AM?

Qual é a vantagem de usar AM?

Abordagem tradicional de programação

Qual é a vantagem de usar AM?

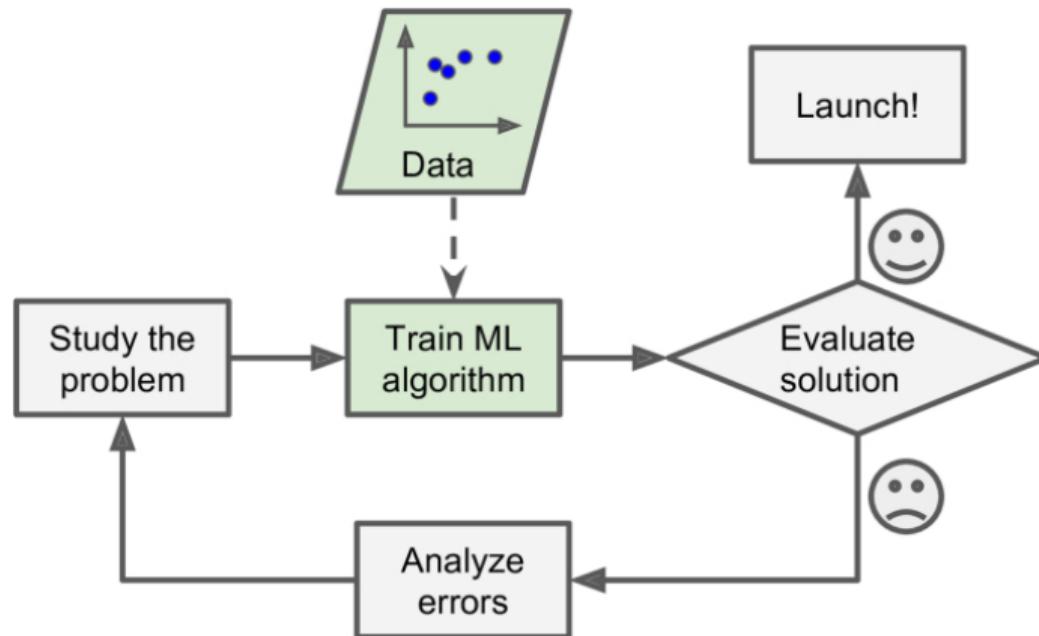


- ▶ O programa seria longo, complexo e difícil de manter.

Qual é a vantagem de usar AM?

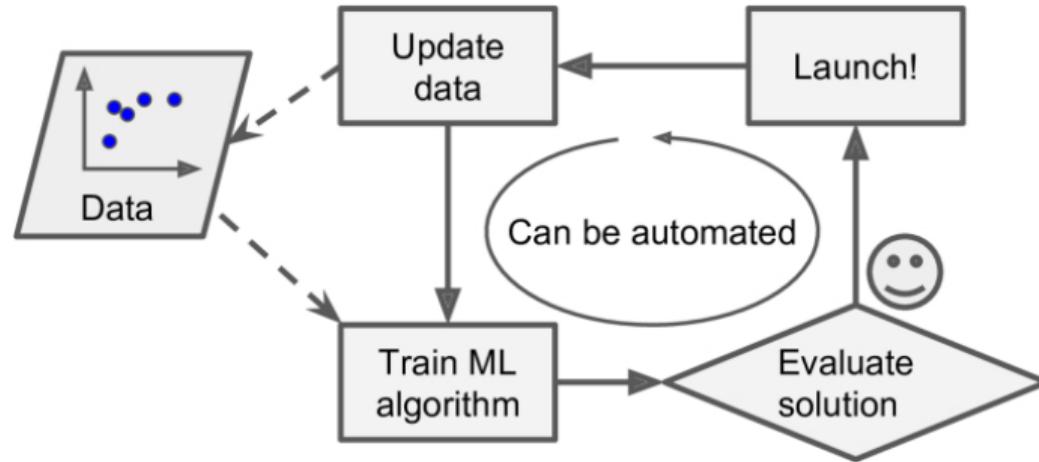
Abordagem de AM

Qual é a vantagem de usar AM?



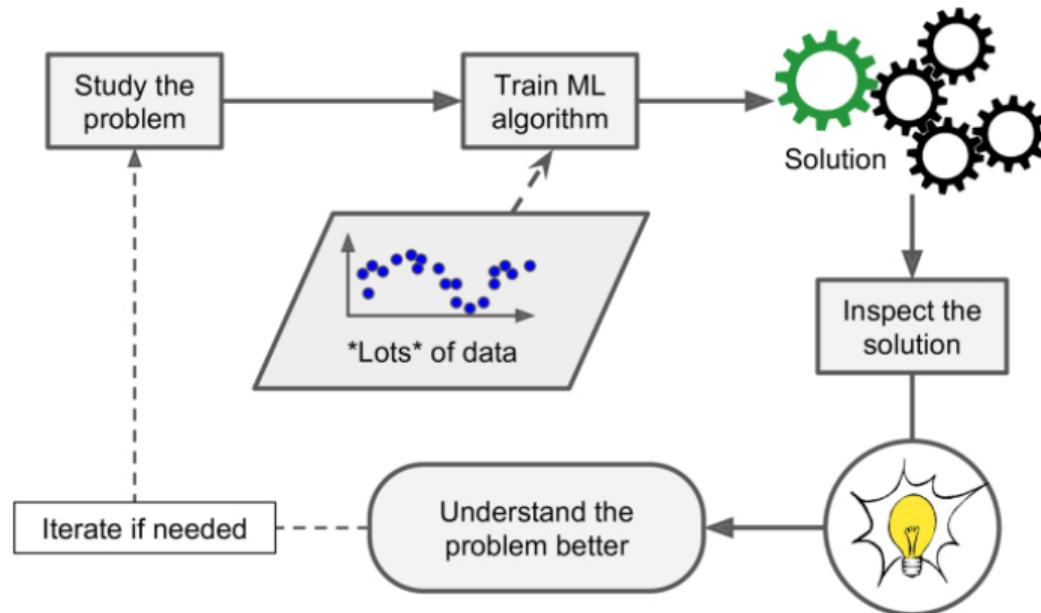
- ▶ O programa é mais curto, mais fácil de manter e, provavelmente, mais preciso.

Qual é a vantagem de usar AM?



- ▶ O programa se adapta automaticamente à mudanças (e.g. Reconhecimento de fala).

Qual é a vantagem de usar AM?



- ▶ AM podem fazer humanos aprenderem (e.g. *mineração de dados*).

AM é bom pra que?

AM é bom pra que?

AM é bom pra que?

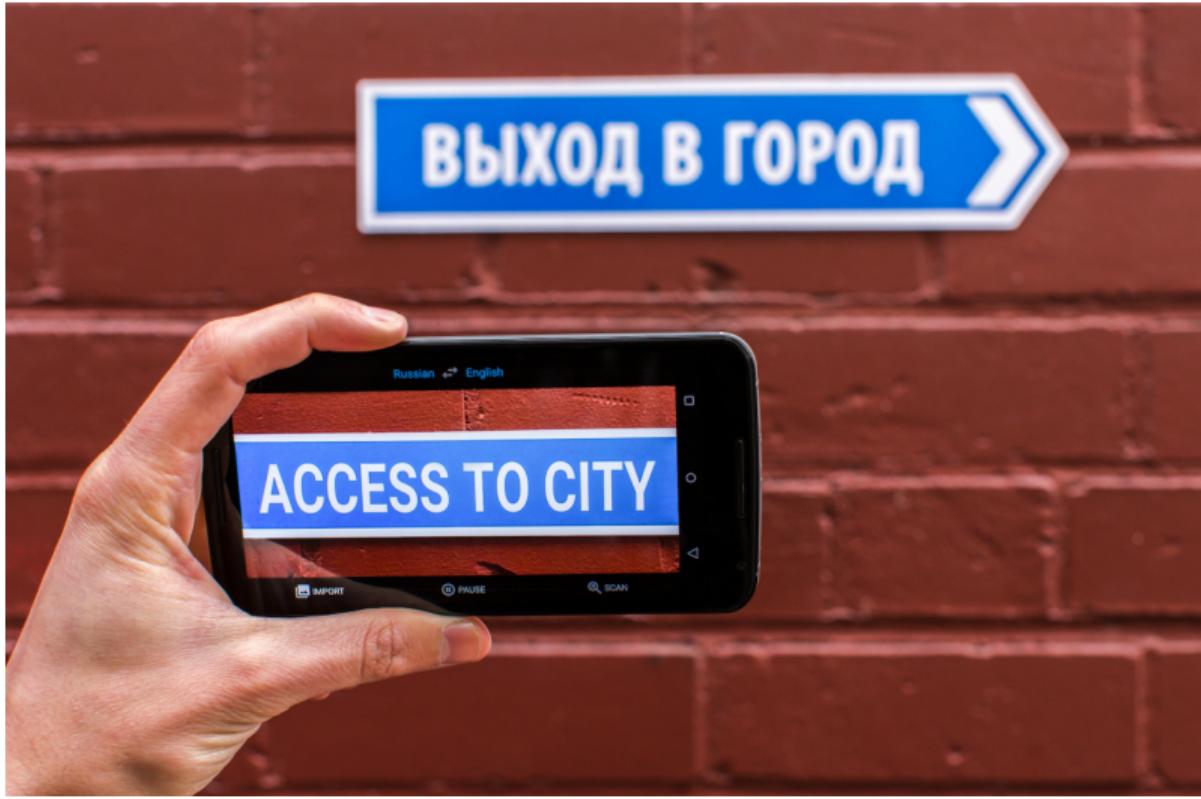
- ▶ Problemas que são muito complexos para abordagens tradicionais;

AM é bom pra que?

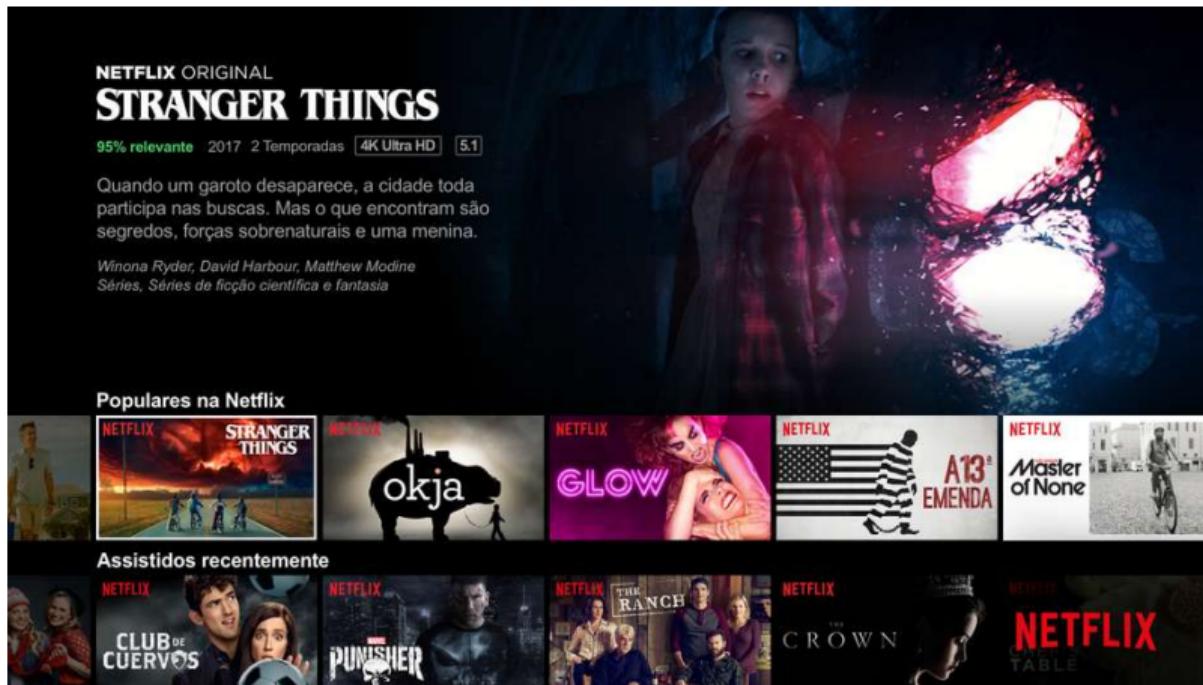
- ▶ Problemas que são muito complexos para abordagens tradicionais;
- ▶ Quando não existe algoritmo.

Aplicações

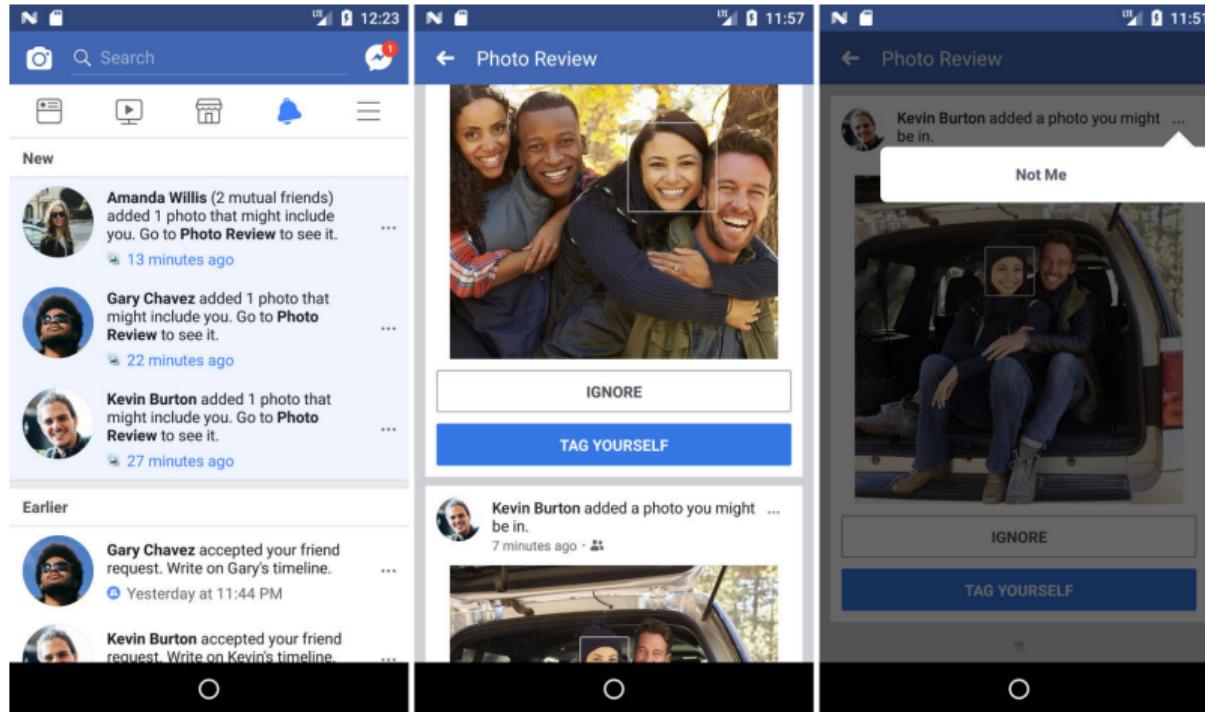
Aplicações



Aplicações



Aplicações



Aplicações



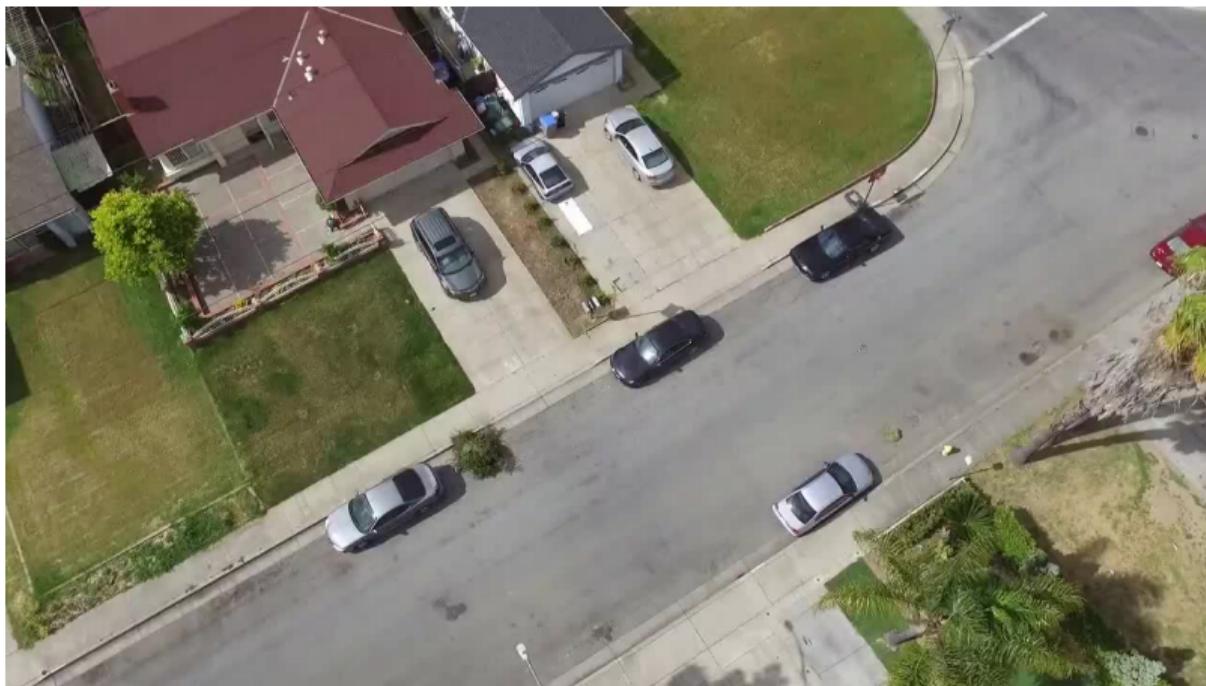
Aplicações



Aplicações



Aplicações



Aplicações



Aplicações

Few-Shot Adversarial Learning of Realistic Neural Talking Head models

Egor Zakharov^{1,2} Aliaksandra Shysheya^{1,2} Egor Burkov^{1,2} Victor Lempitsky^{1,2}

¹Samsung Research

²Skolkovo Institute of Science and Technology



Tipos de sistemas de AM (SAM)

Tipos de Sistemas de AM (SAM)

- ▶ SAM treinados ou não com supervisão humana: Aprendizado supervisionado, aprendizado não-supervisionado, aprendizado semi-supervisionado e aprendizado por reforço;

Tipos de Sistemas de AM (SAM)

- ▶ SAM treinados ou não com supervisão humana: Aprendizado supervisionado, aprendizado não-supervisionado, aprendizado semi-supervisionado e aprendizado por reforço;
- ▶ SAM que aprendem ou não em tempo real: Aprendizado em lote (*offline*) e aprendizado *online*;

Tipos de Sistemas de AM (SAM)

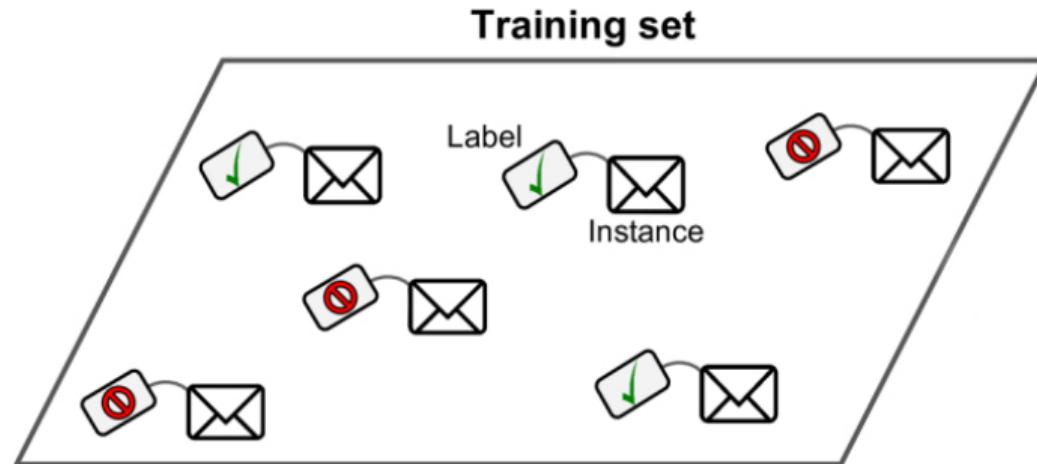
- ▶ SAM treinados ou não com supervisão humana: Aprendizado supervisionado, aprendizado não-supervisionado, aprendizado semi-supervisionado e aprendizado por reforço;
- ▶ SAM que aprendem ou não em tempo real: Aprendizado em lote (*offline*) e aprendizado *online*;
- ▶ SAM que funcionam comparando novos dados a dados conhecidos ou detectando padrões no conjunto de treinamento e construindo um modelo preditivo: Aprendizado baseado em instância e aprendizado baseado em modelo.

SAM treinados ou não com supervisão humana

Aprendizado supervisionado

Aprendizado supervisionado

- ▶ Conjunto de treinamento rotulado!

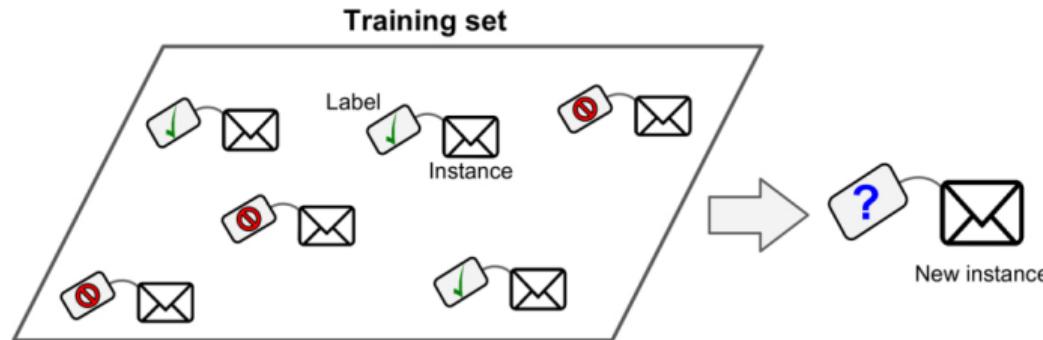


- ▶ O SAM é treinado para aprender com a supervisão humana.

Classificação

Aprendizado supervisionado

- ▶ Conjunto de treinamento: Rótulos e instâncias.



- ▶ O SAM é treinado para classificar novas instâncias como pertencentes a uma determinada classe.

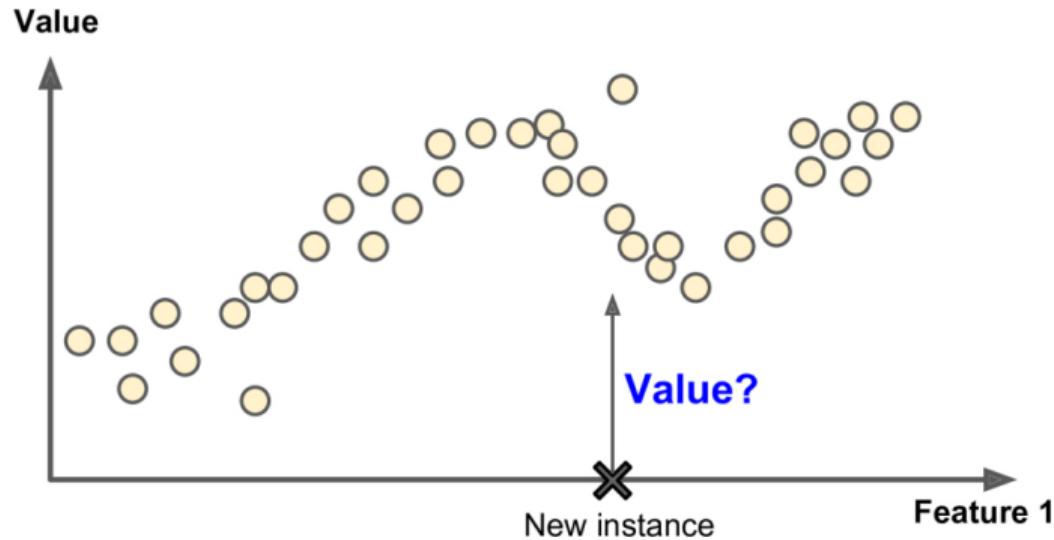
Régressão

Aprendizado supervisionado

- ▶ Conjunto de treinamento: rótulos e características (ou atributos).

Modelo	Quilometragem (km)	Ano	Marca	...	Valor (R\$)
Corolla	1 000	2019	Toyota	...	60 000
Fusca	100 000	1984	Volkswagen	...	2 000
Onix	15 000	2017	Chevrolet	...	30 000
:	:	:	:	:	:
Duster	10 000	2018	Renault	...	45 000

Aprendizado supervisionado



- ▶ O SAM é treinado para prever um valor numérico alvo (rótulo) de uma nova instância dado um conjunto de características.

Aprendizado supervisionado

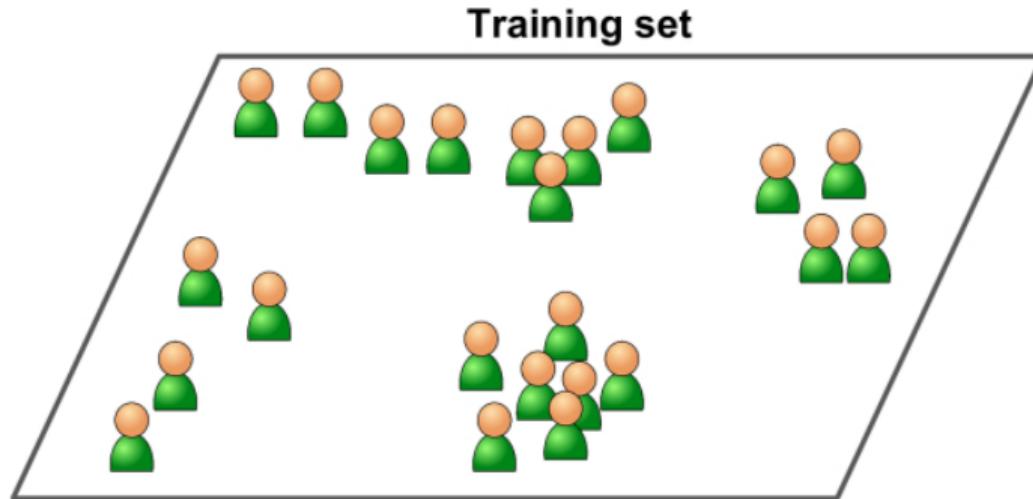
Alguns dos mais importantes algoritmos de aprendizado supervisionado:

- ▶ k-Vizinhos Mais Próximos;
- ▶ Regressão Linear;
- ▶ Regressão Logística;
- ▶ Máquinas de Vetor de Suporte;
- ▶ Árvores de Decisão e Florestas Aleatórias;
- ▶ Redes Neurais.

Aprendizado não-supervisionado

Aprendizado não-supervisionado

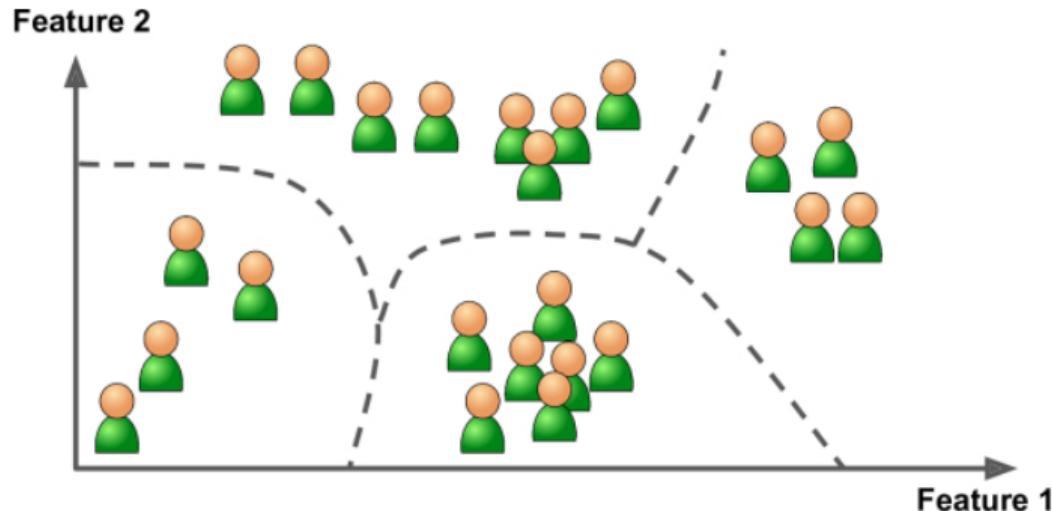
- ▶ Conjunto de treinamento não-rotulado!



- ▶ O SAM é treinado para aprender sem a supervisão humana.

Agrupamento

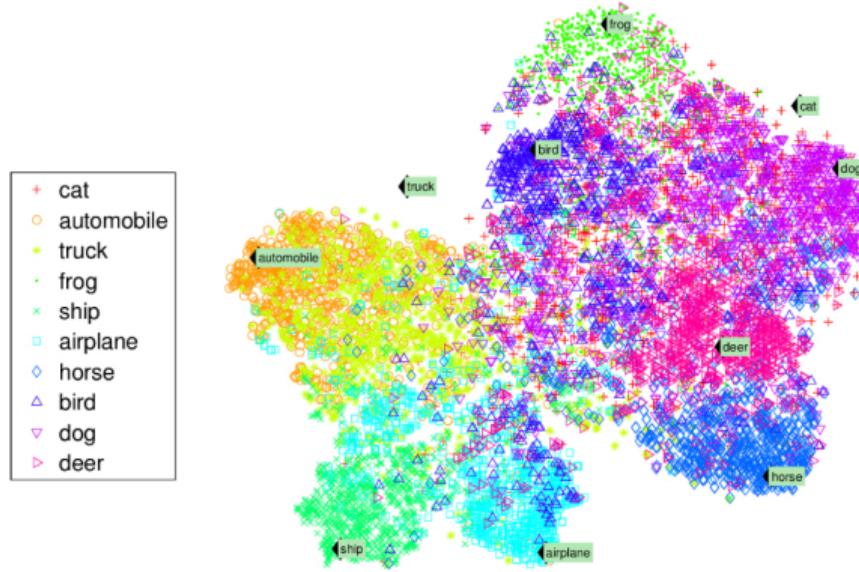
Aprendizado não-supervisionado



- ▶ O SAM é treinado para detectar grupos de elementos parecidos (e.g. Seguidores no Twitter).

Visualização

Aprendizado não-supervisionado



- ▶ O SAM é treinado para representar o seu dado graficamente de forma mais intuitiva (e.g. Aglomerados semânticos).

Redução de Dimensionalidade

Aprendizado não-supervisionado

Modelo	Quilometragem (km)	Ano	Marca	...	Valor (R\$)
Corolla	1 000	2019	Toyota	...	60 000
Fusca	100 000	1984	Volkswagen	...	2 000
Onix	15 000	2017	Chevrolet	...	30 000
:	:	:	:	:	:
Duster	10 000	2018	Renault	...	45 000

- ▶ O SAM é treinado para simplificar o dado sem perder tanta informação (Extração de característica).

Detecção de Anomalia (ou *Outliers*)

Aprendizado não-supervisionado



- ▶ O SAM é treinado para identificar comportamentos anormais (e.g. Operações bancárias suspeitas).

Regra de Associação

Aprendizado não-supervisionado



- ▶ O SAM é treinado para identificar relações de interesse (e.g. Cestas de compras).

Aprendizado não-supervisionado

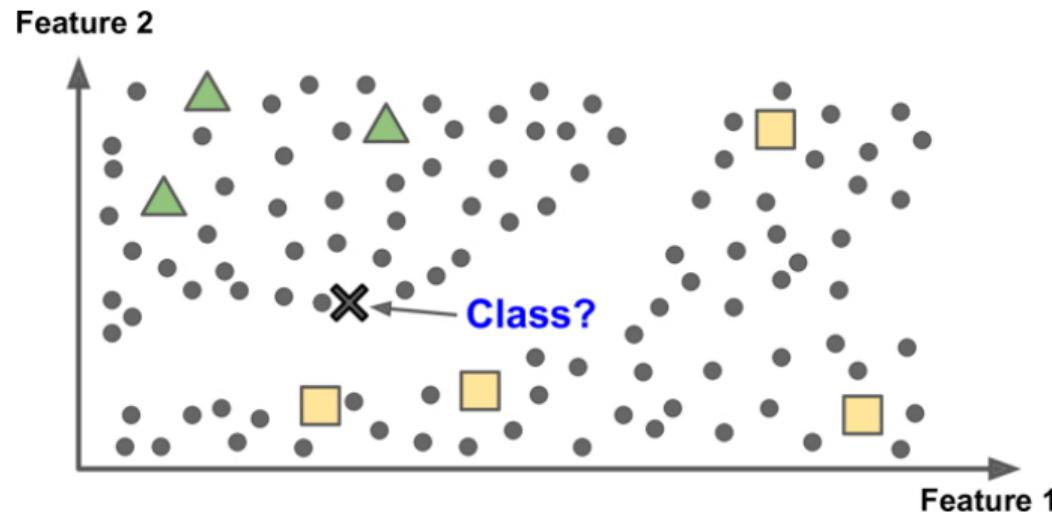
Alguns dos mais importantes algoritmos de aprendizado não-supervisionado:

- ▶ *k-Means*;
- ▶ Análise de agrupamento hierárquico;
- ▶ Maximização da expectativa;
- ▶ Análise de Componentes Principais;
- ▶ *Locally-linear embedding* (LLE);
- ▶ *t-distributed Stochastic Neighbor Embedding* (t-SNE);
- ▶ Apriori;
- ▶ Eclat.

Aprendizado semi-supervisionado

Aprendizado semi-supervisionado

- ▶ Conjunto de treinamento parcialmente rotulado!

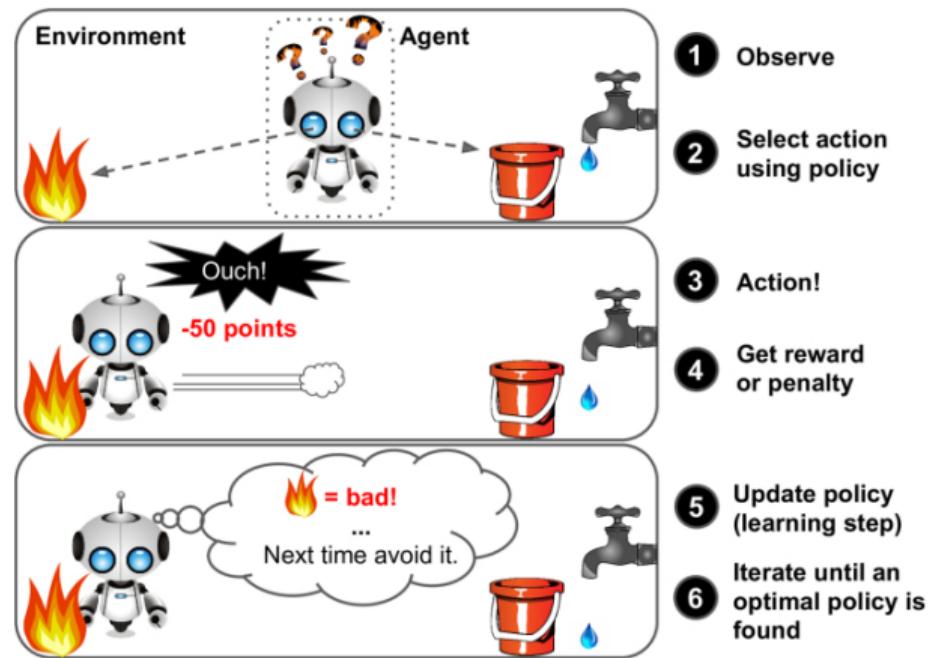


- ▶ O SAM é treinado para aprender através de uma combinação de aprendizado supervisionado e não-supervisionado (e.g. Fotos no Facebook).

Aprendizado por reforço

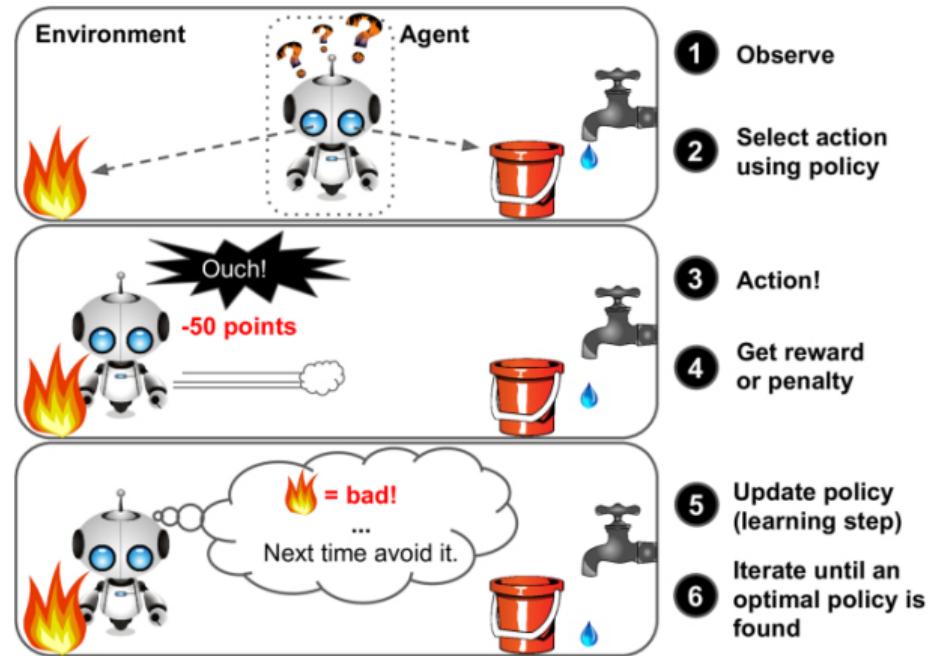
Aprendizado por reforço

- ▶ Ambiente (conjunto de treinamento) e Agente (SAM);



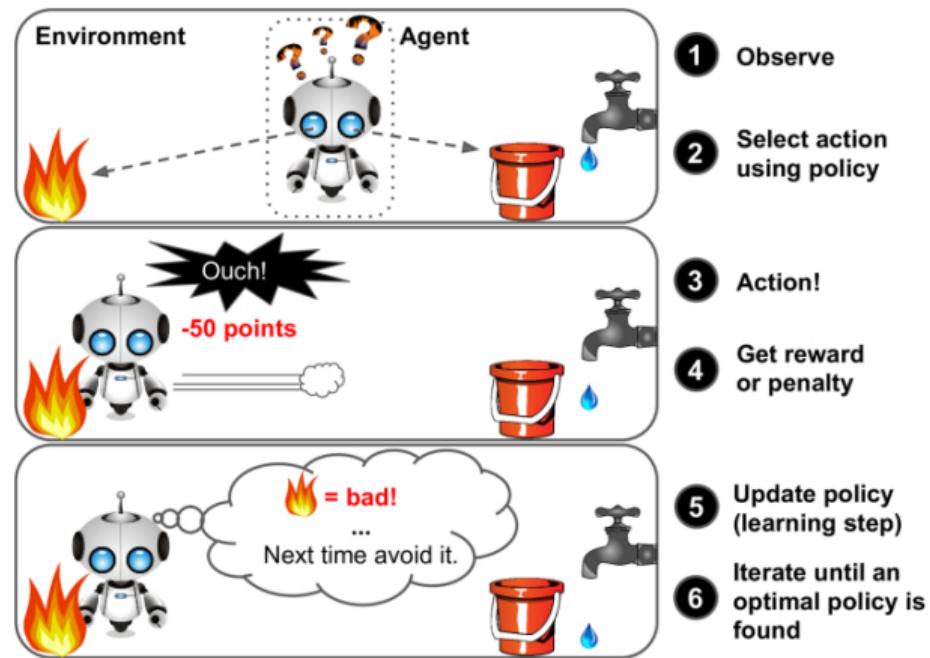
Aprendizado por reforço

- ▶ Ambiente (conjunto de treinamento) e Agente (SAM);
- ▶ O agente pode observar o ambiente, selecionar e realizar ações ganhando recompensas (ou sofrendo penalidades);

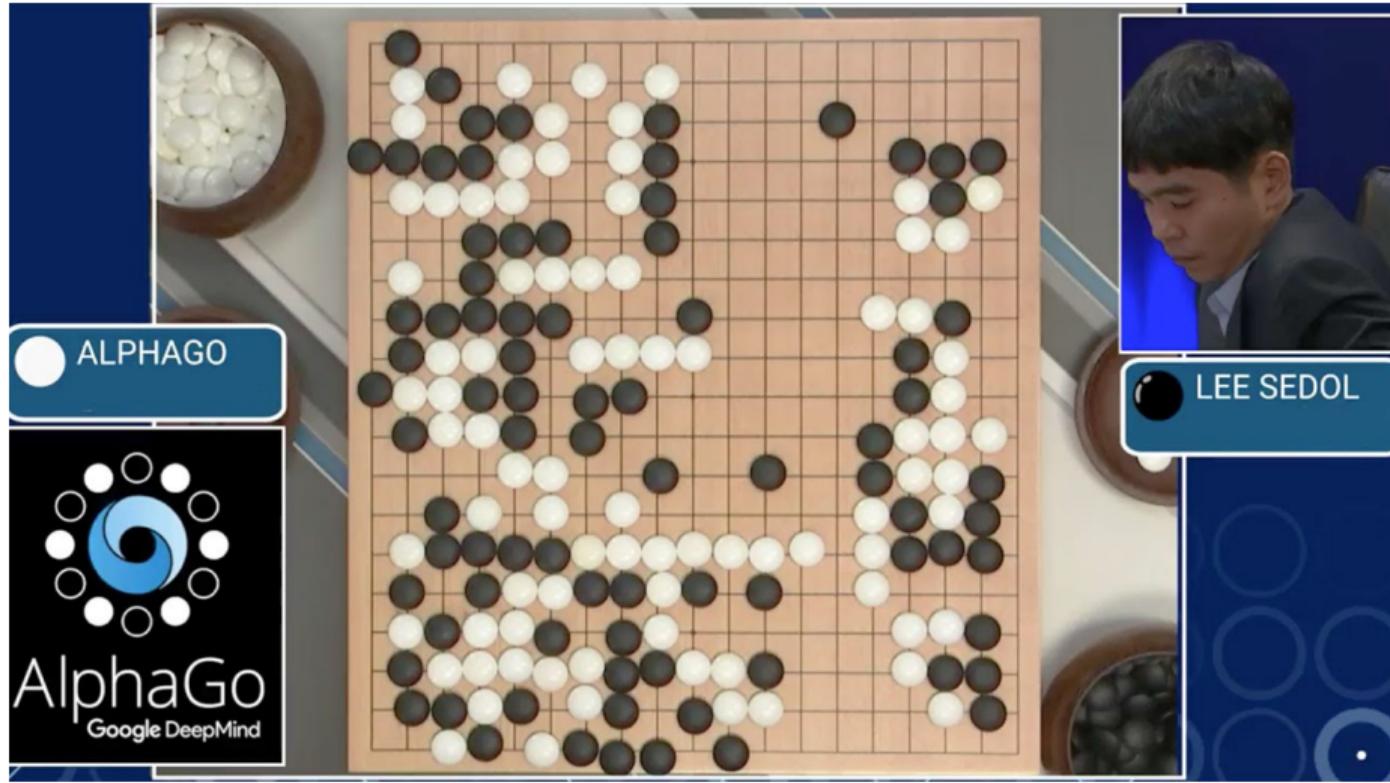


Aprendizado por reforço

- ▶ Ambiente (conjunto de treinamento) e Agente (SAM);
- ▶ O agente pode observar o ambiente, selecionar e realizar ações ganhando recompensas (ou sofrendo penalidades);
- ▶ O agente é treinado para aprender a melhor estratégia (política).



Aprendizado por reforço



SAM que aprendem ou não em tempo real

Aprendizado em Lote (*offline*)

Aprendizado em Lote (offline)



- ▶ O SAM é treinado com todos os dados disponíveis e lançado em produção.

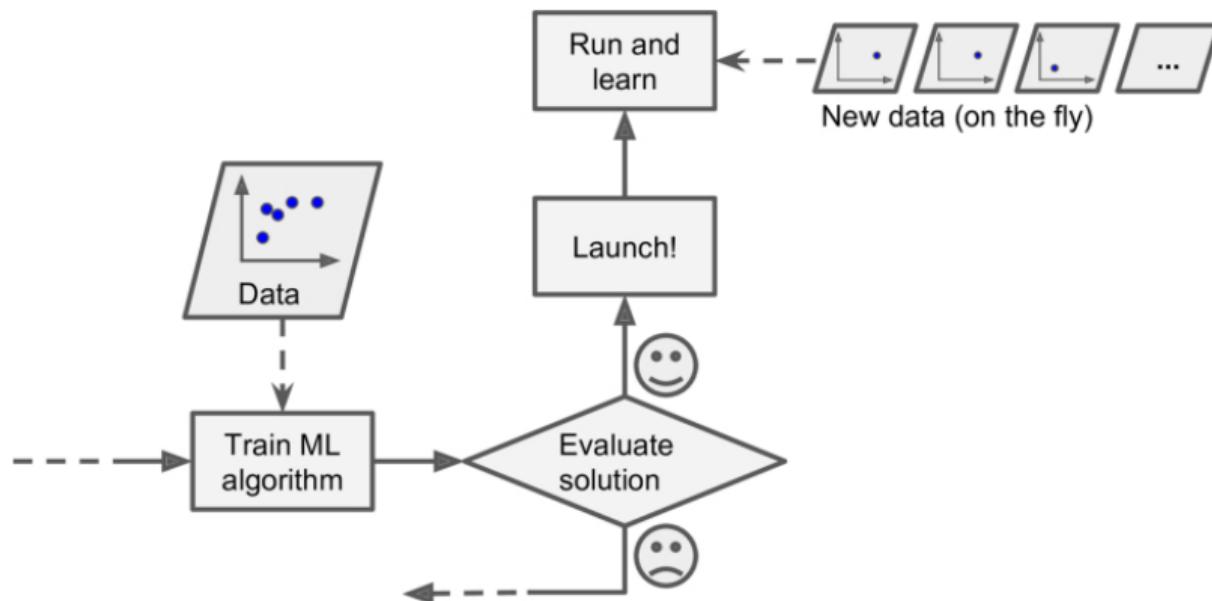
Aprendizado *online*

Aprendizado online



- ▶ O SAM é lançado em produção e treinado continuamente por novos dados.

Aprendizado online

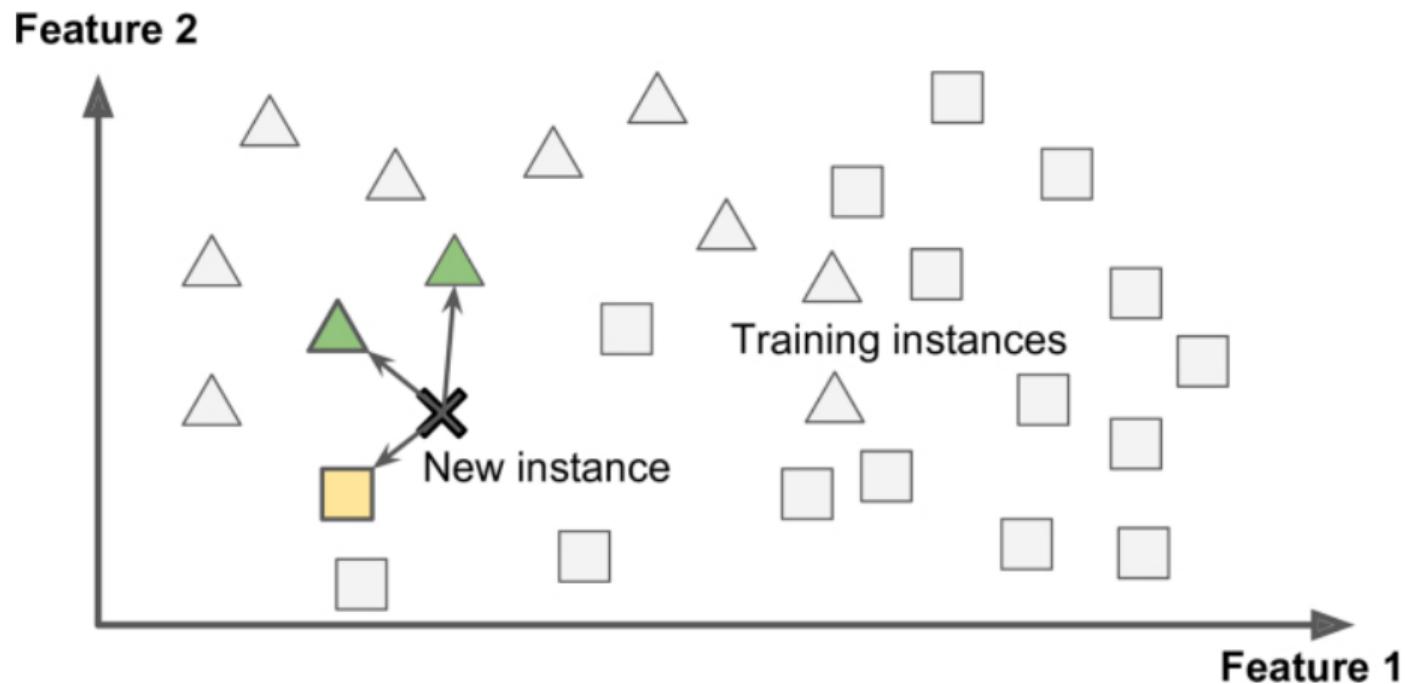


- ▶ Taxa de Aprendizado: O quanto veloz o SAM se adapta à mudanças nos dados.

SAM que usam comparações de dados ou
modelo preditivo

Aprendizado baseado em instância

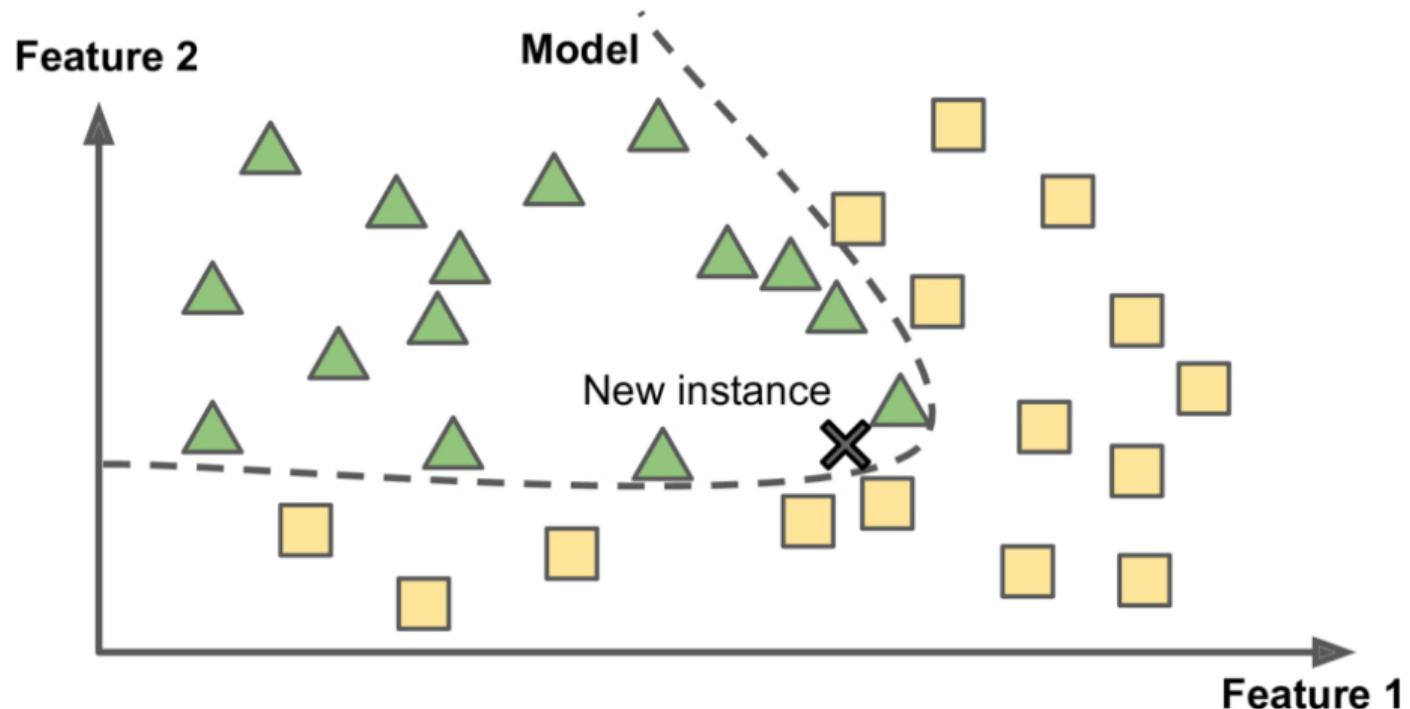
Aprendizado baseado em instância

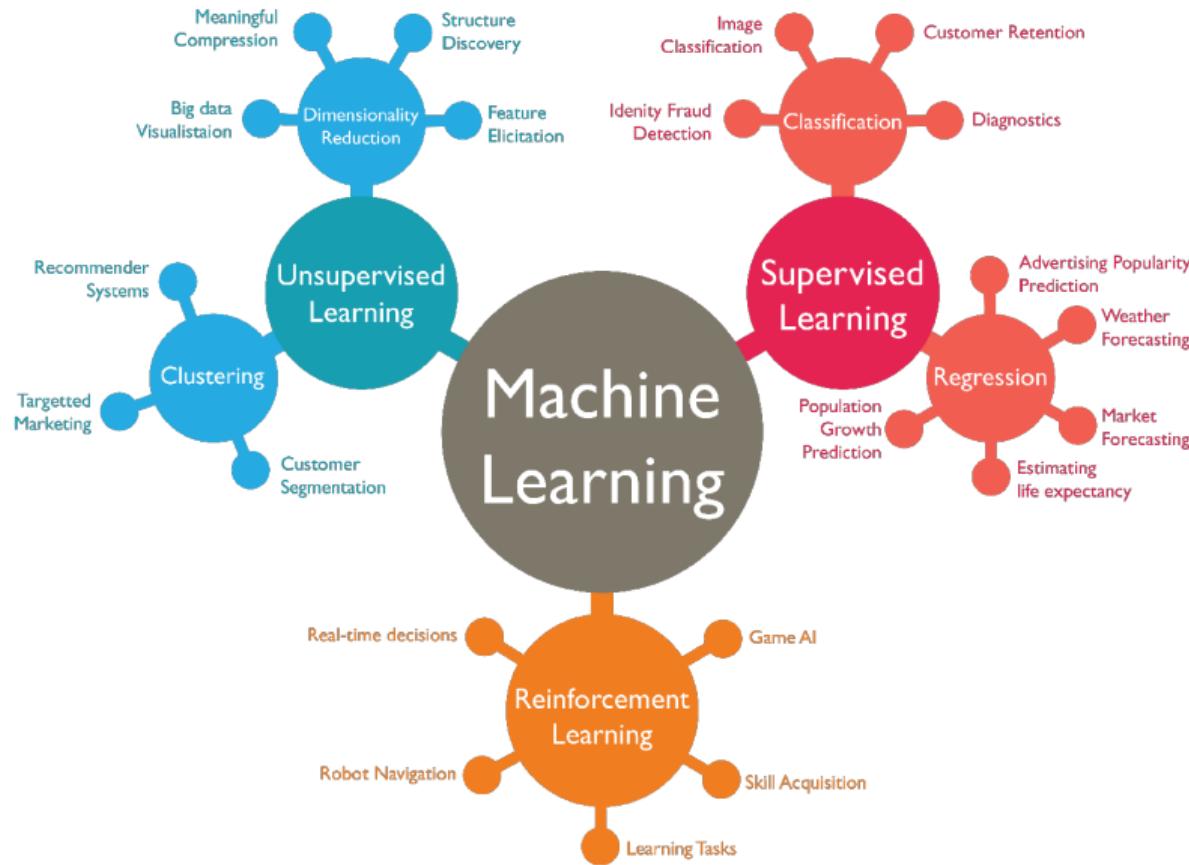


- ▶ Medida de Similaridade: O quanto similar, dado um critério, são duas instâncias.

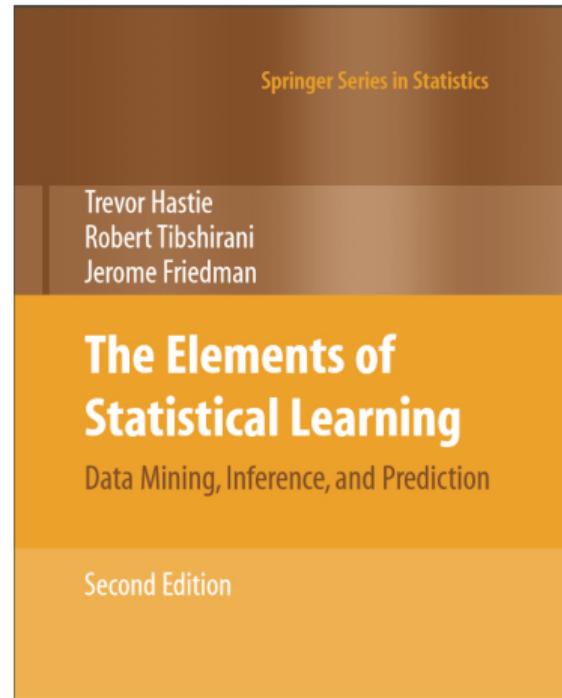
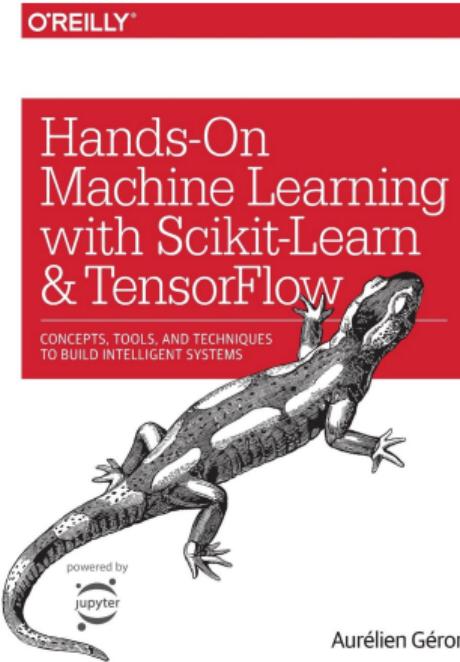
Aprendizado baseado em modelo

Aprendizado baseado em modelo





Bibliografia





Hasta la vista, baby!