

# Introdução ao aprendizado de máquina

Prof. Wesin Ribeiro Alves

# Agenda

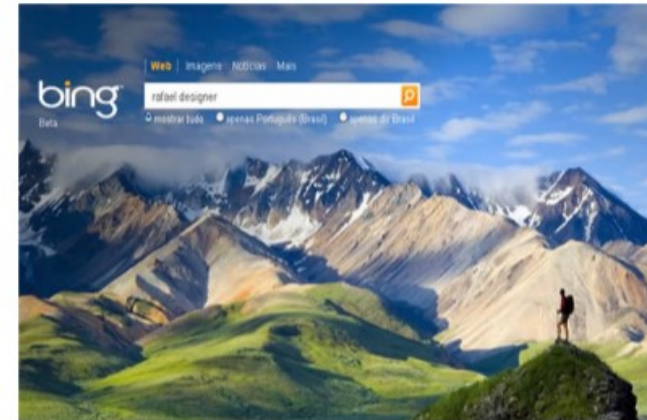
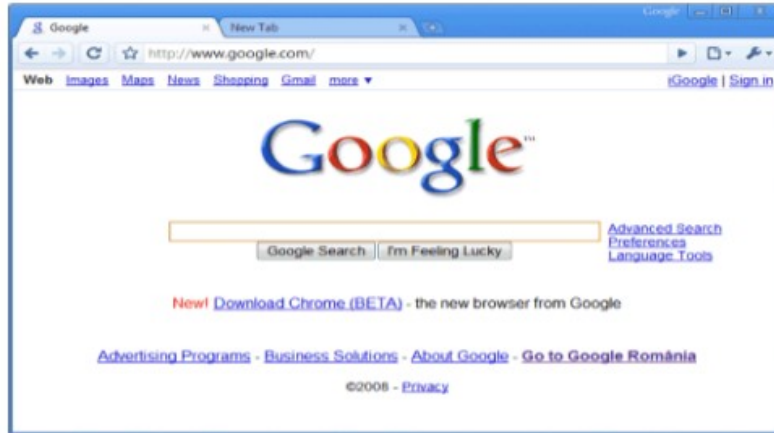
- Introdução
- Contexto histórico
- O que é aprendizado de máquina
- Tipos de aprendizado de máquina
- Principais desafios
- Teste e validação

extinct species ?

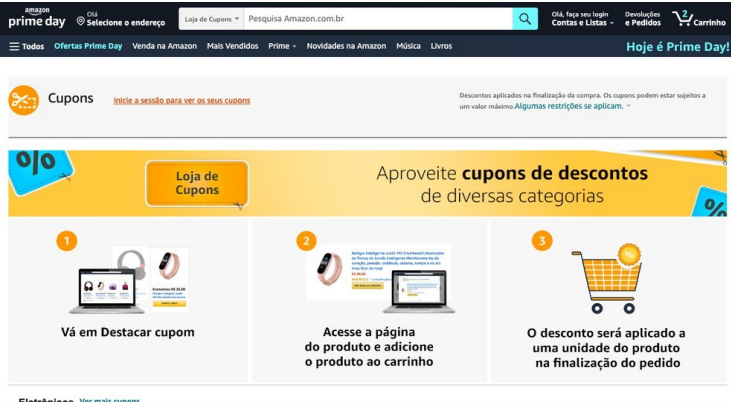
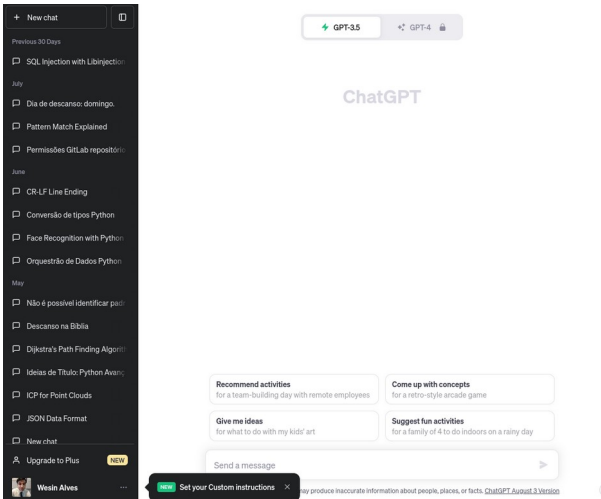
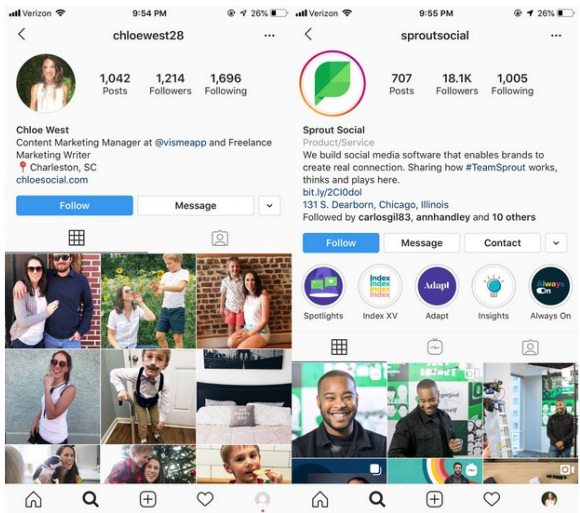


Se eu baixar um cópia do  
wikipedia, meu computador vai  
ficar mais inteligente?

# Primeiras aplicações de AM



# Aplicações modernas



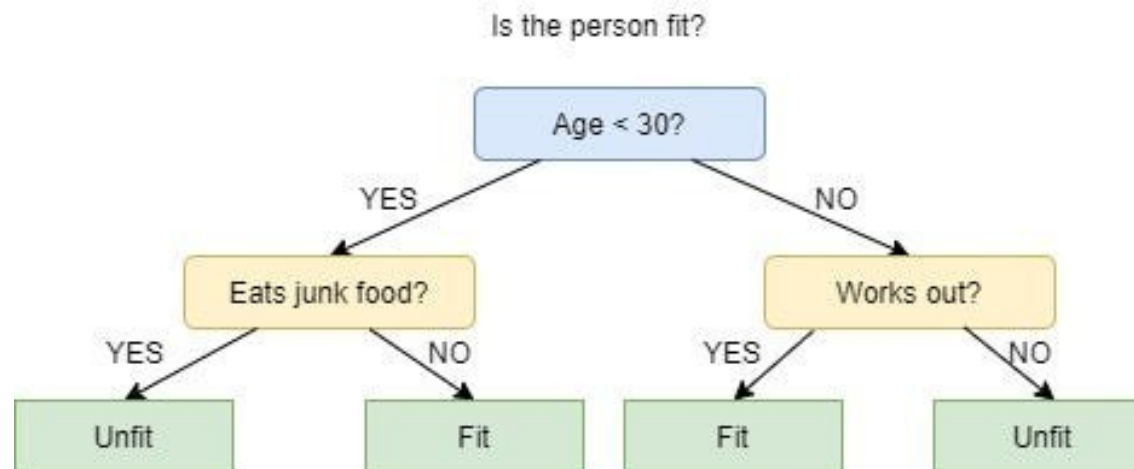
# Contexto histórico

- 1940s – 1960s
  - McCulloch and Pitts apresentaram o artigo “A logical calculus of the ideas immanent in nervous activity” em 1943.
  - Arthur Samuel (IBM) desenvolveu o programa jogador de damas em 1955.
  - McCarthy organizou um conferencia em Dartmouth College, onde o termo *Inteligência Artificial* foi introduzido pela primeira vez em 1956.
  - Rosenblatt desenvolveu o *perceptron* como o primeiro modelo para aprendizado supervisionado em 1958.
  - Minsky e Papert escreveram um artigo (1969) provando as limitações do *perceptron*, que provocou o primeiro inverno na área por 20 anos.



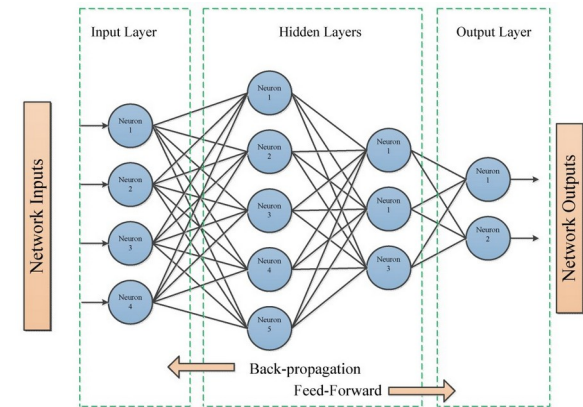
# Contexto histórico

- 1970s
  - Desenvolvimento de algoritmos mais práticos, muitas vezes através de técnicas simbólicas
  - Sistemas especialistas e o gargalo de aquisição de conhecimento
  - Algoritmo ID3 (Iterative Dichotomiser 3) de Quinlan



# Contexto histórico

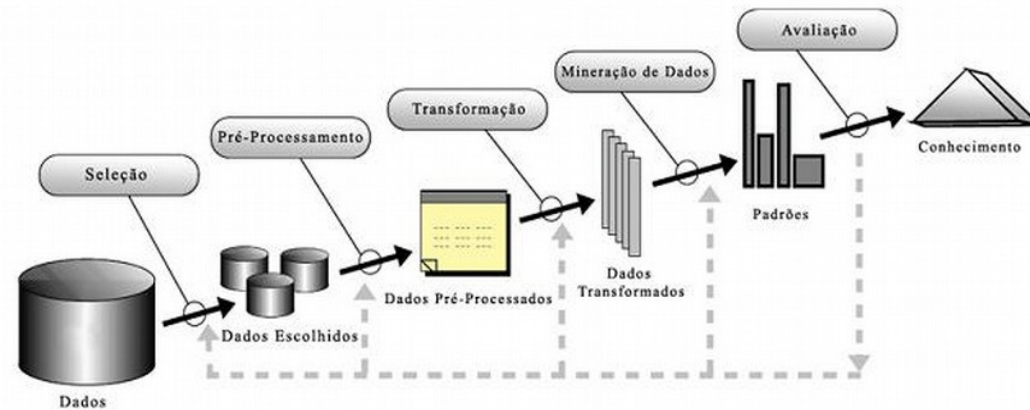
- 1980s
  - Surgimento de novas técnicas e maior ênfase em metodologias de avaliação
  - Rumelhart (1986) propôs o algoritmo *backpropagation* que superava as limitações do *perceptron*, interrompendo o inverno na área.





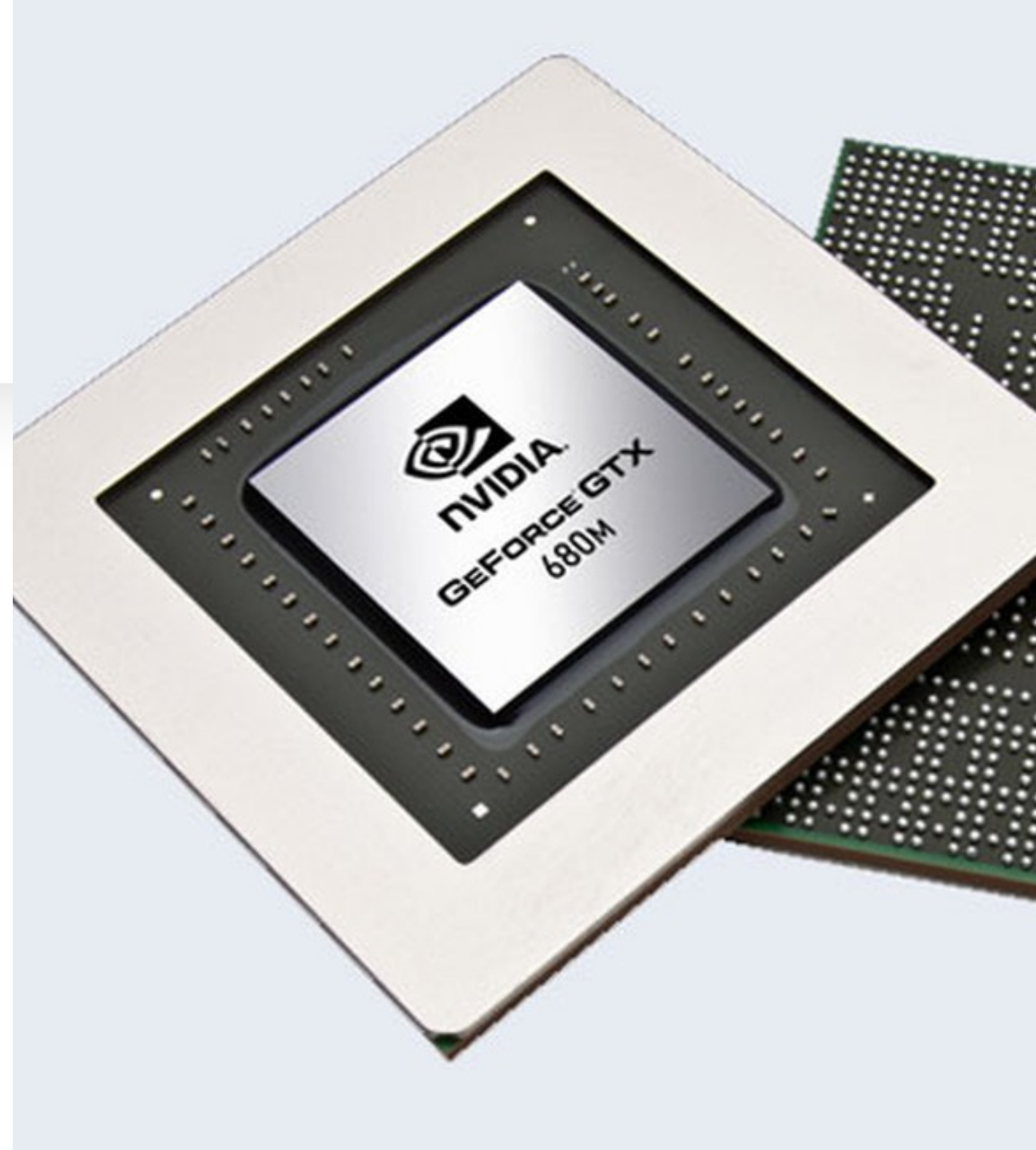
# Contexto histórico

- 1990s
  - Sistemas inteligentes híbridos, aprendizado por reforço, aprendizado por redes bayesianas. Primerias aplicações práticas.
  - Mineração de dados (1996) emerge como aplicação importante para muitas atividades comerciais



# Contexto histórico

- 2000s
  - Máquina de vetores de suporte (SVM).
  - Modelos gráficos probabilísticos.
- 2010s
  - Aumento do poder computacional (GPUs)
  - Carros autônomos da Google
  - Visão computacional
  - Aprendizado profundo
  - Fim do segundo inverno nas pesquisas em redes neurais artificiais



# O que é aprendizado de máquina?

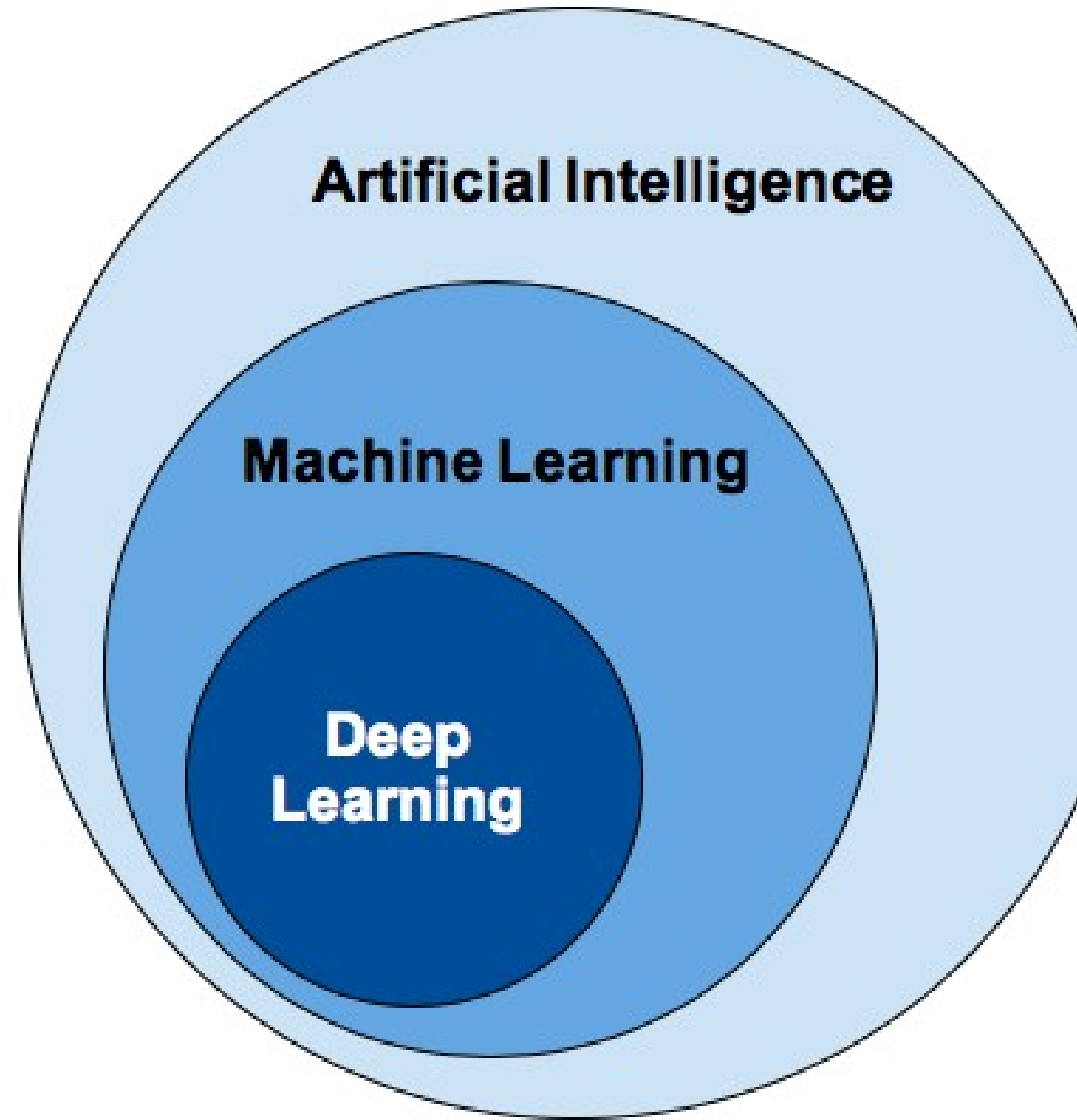
- Arthur Samuel (1959): é o campo de estudo que fornece aos computadores a habilidade de apreender sem ser explicitamente programado para isso.
- Tom Mitchell (1998): Um programa de computador aprende a partir da **experiência**  $E$ , em relação a uma classe de **tarefas**  $T$ , com medida de desempenho  $D$ , se seu **desempenho** em  $T$ , medido por  $D$ , melhora com  $E$ .

# O que é aprendizado de máquina?

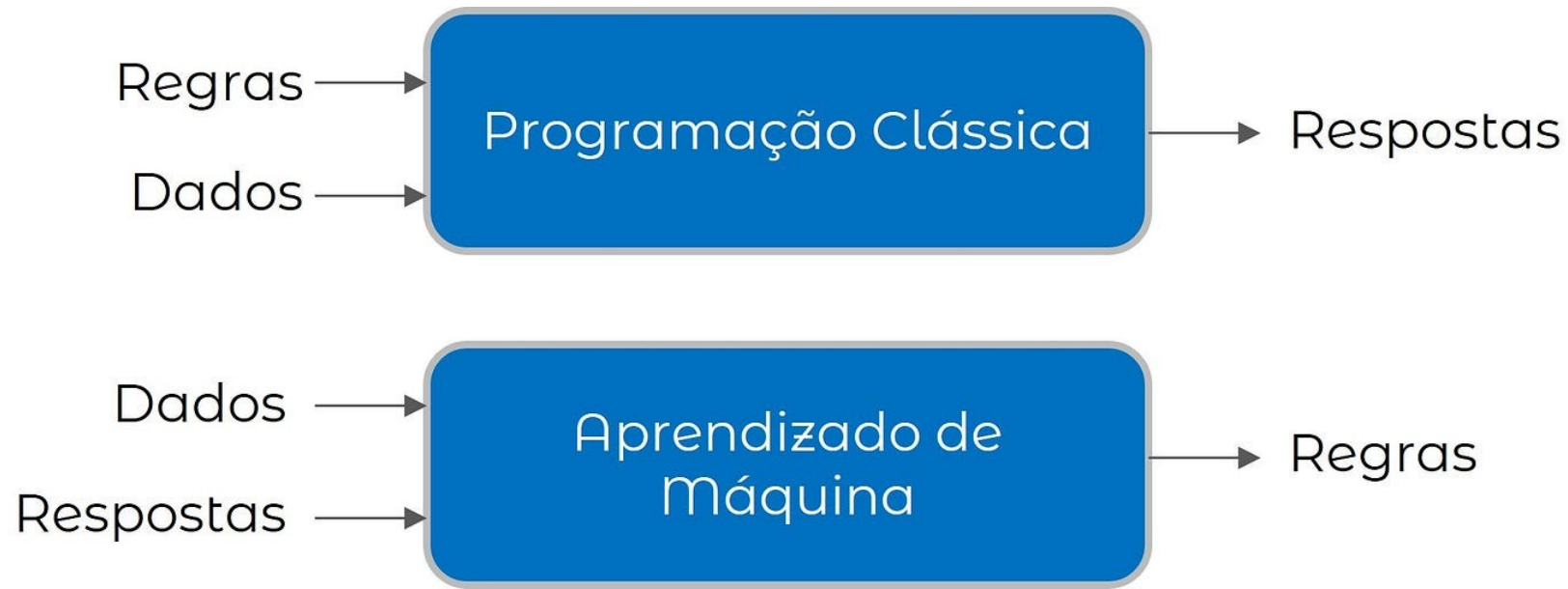
- Aurélio Géron (2019): é a ciência (e arte) de programar computadores de modo que eles possam aprender a partir dos dados.
- Andreas Muller (2016): O aprendizado de máquina é sobre extrair conhecimento de dados. É um campo de pesquisa na interseção de estatística, inteligência artificial e ciência da computação e também é conhecido como análise preditiva ou aprendizagem estatística.
- Lorena, et al. (2011): é o processo de indução de uma hipótese (ou aproximação de função) a partir da experiência passada.

# O que é aprendizado de máquina?

- É uma subárea da inteligência artificial que tem por objetivo ensinar máquinas a aprenderem **tarefas inteligentes** por meio de um **processo de aprendizagem** baseado em **experiência**.

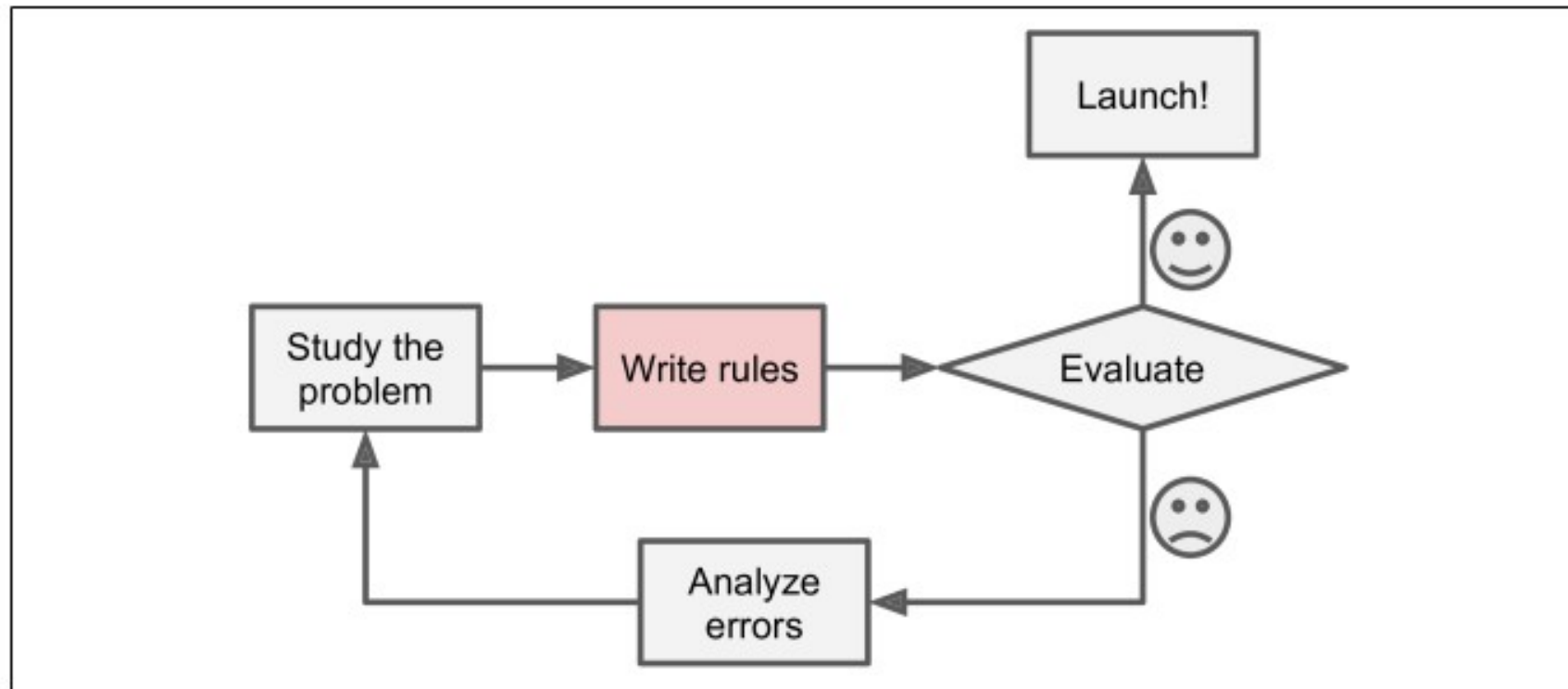


# O que não é aprendizado de máquina?



# Por que usar aprendizado de máquina?

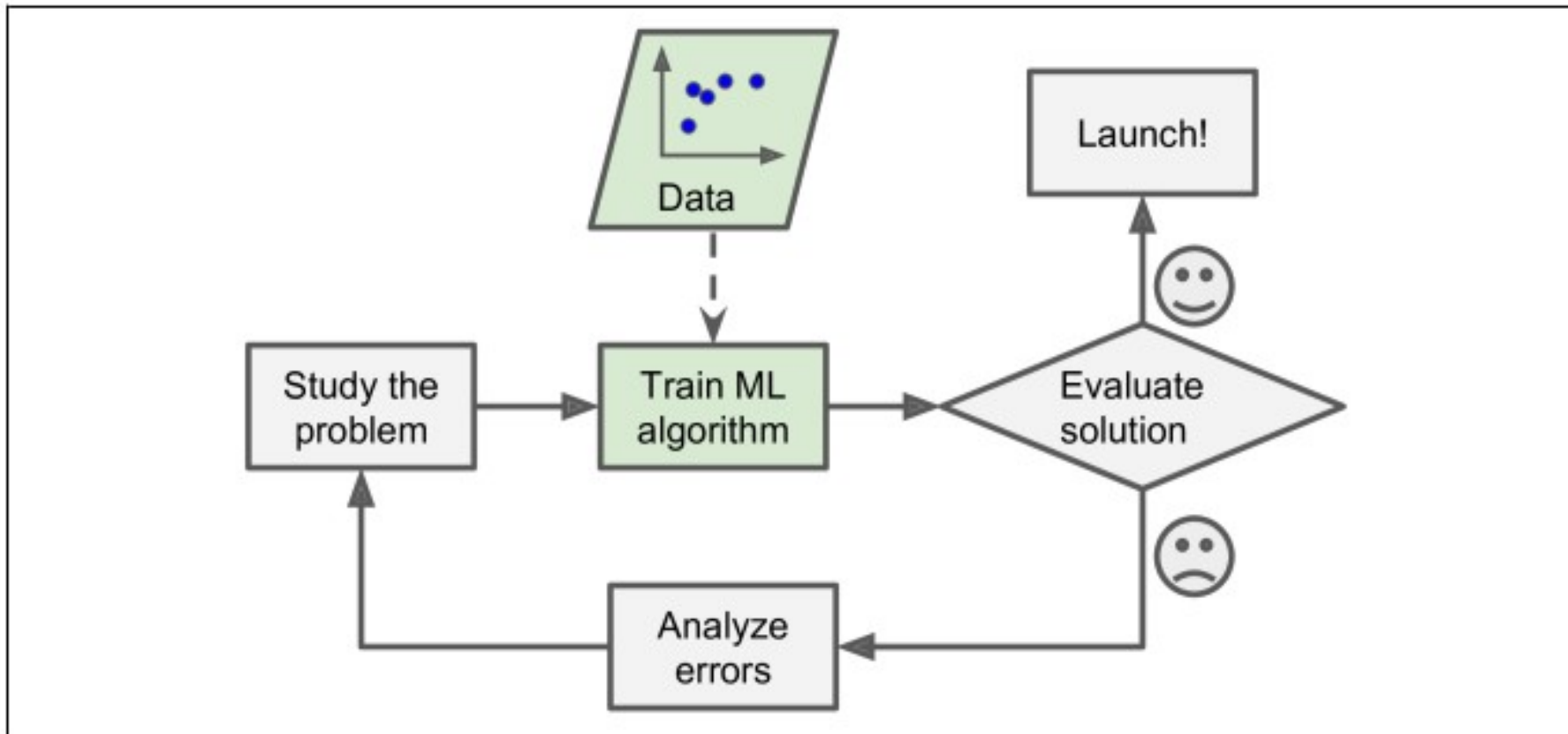
- Abordagem tradicional (ex: filtro de spam)





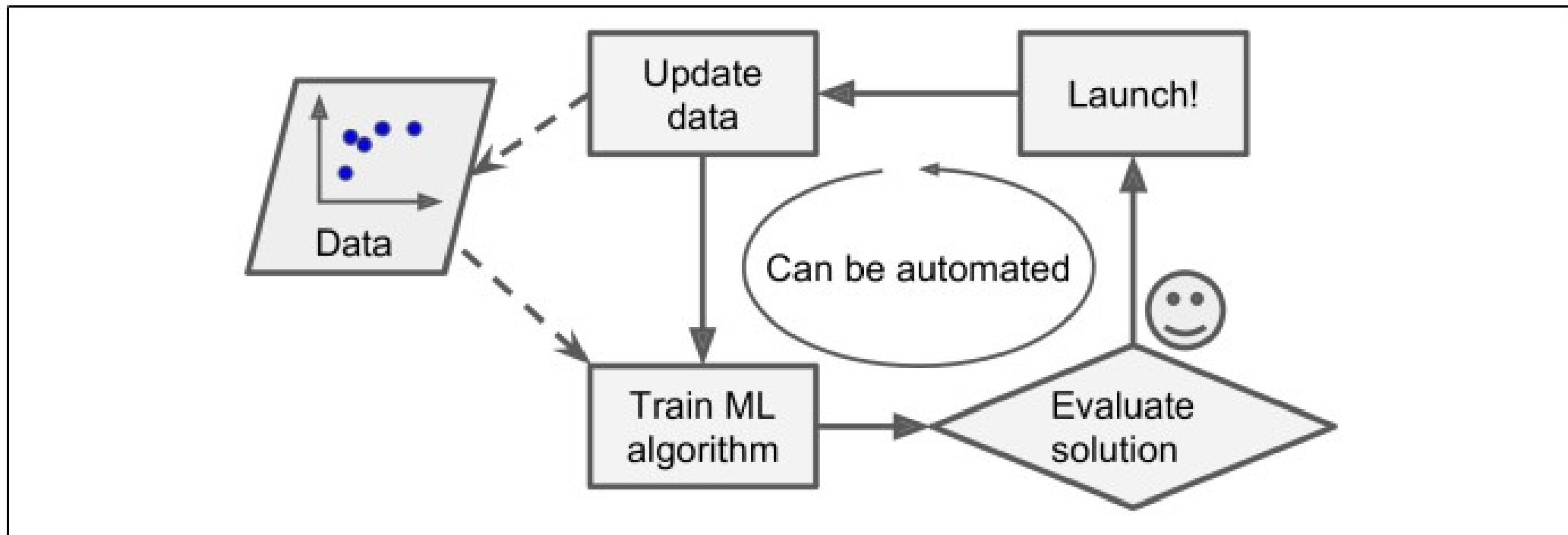
# Por que usar aprendizado de máquina?

- Abordagem com aprendizado de máquina



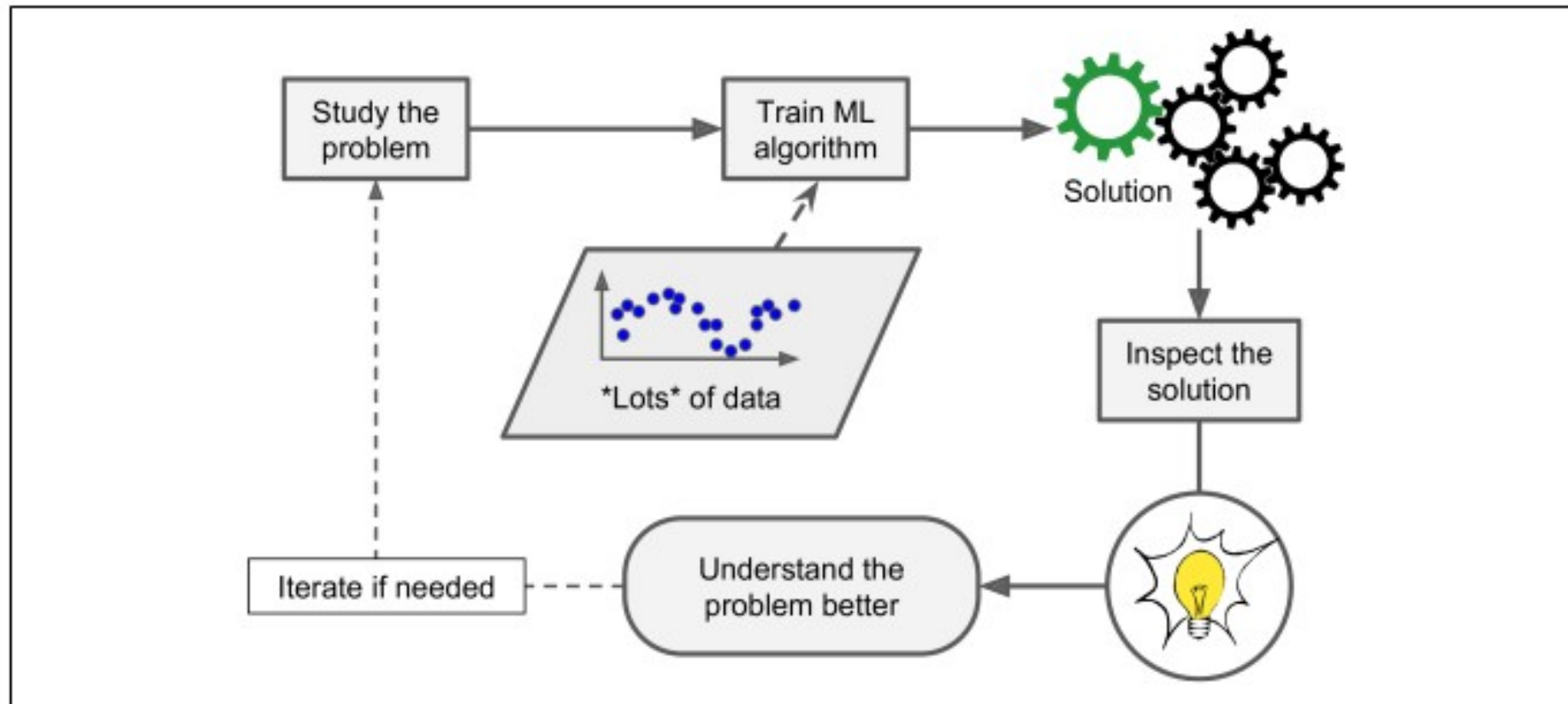
# Por que usar aprendizado de máquina?

- Adaptação automática a mudanças



# Por que usar aprendizado de máquina?

- Aprendizado de máquina pode ajudar humanos a aprender (copiloto)



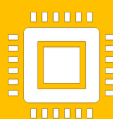
# Aprendizado de máquina é excelente para:



Problemas para qual soluções existentes exigem grande quantidade de ajustes manuais ou uma longa lista de regras.



Problemas complexos para qual não existe boa solução usando abordagem tradicional.



Ambientes flutuantes, pois um sistema de aprendizado de máquina pode se adaptar a dados novos.



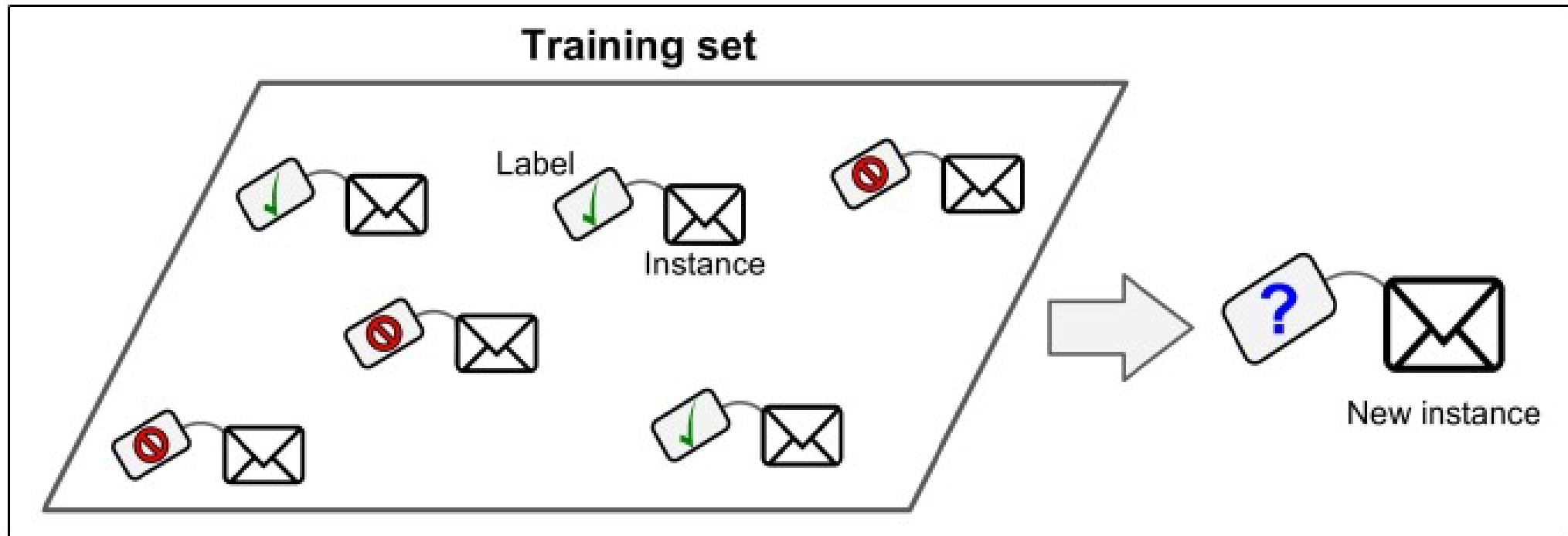
Obter ideias sobre problemas complexos e grande quantidade de dados.

# Tipos de aprendizado de máquina

- Há várias formas de classificação que é interessantes classificá-los em categorias abrangentes baseadas em:
  - Dependem ou não de supervisão humana (supervisionado, semi-supervisionado, não-supervisionado, aprendizado por reforço);
  - Podem ou não aprender incrementalmente em tempo real (online ou aprendizado em lote);
  - Detectam padrões em dados para construir um modelo preditivo ou simplesmente comparam novos dados com os dados conhecidos (aprendizado baseado em modelo ou baseado em instâncias)

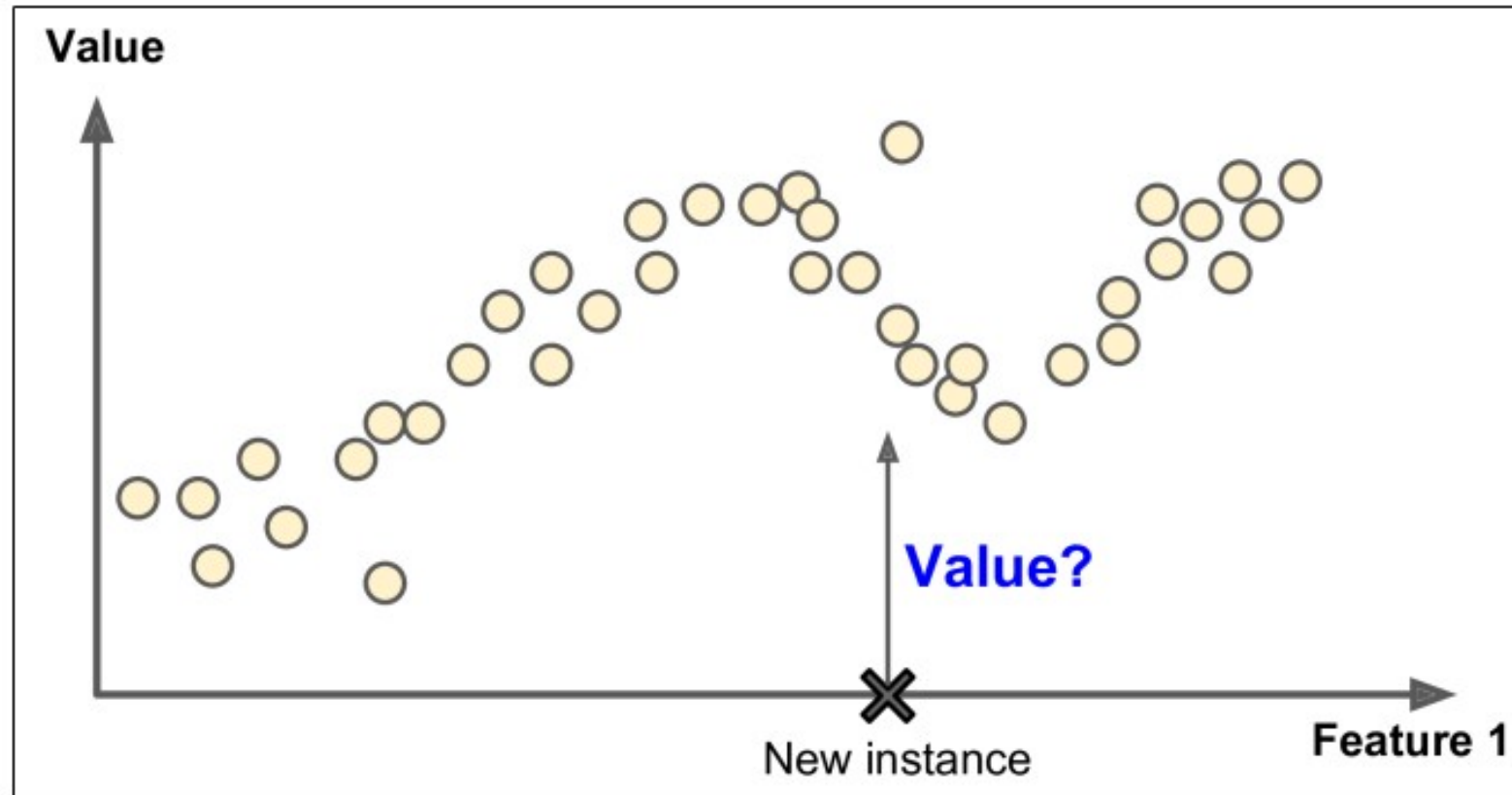
# Supervisão humana

- Aprendizado supervisionado (classificação)



# Supervisão humana

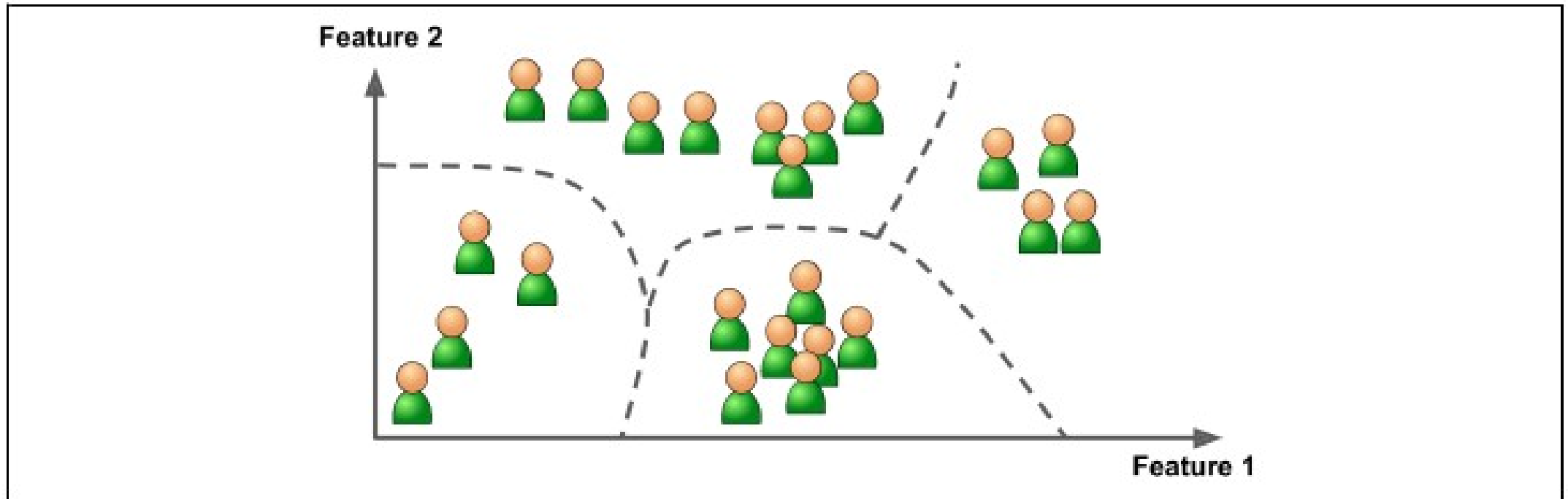
- Aprendizado supervisionado (regressão)





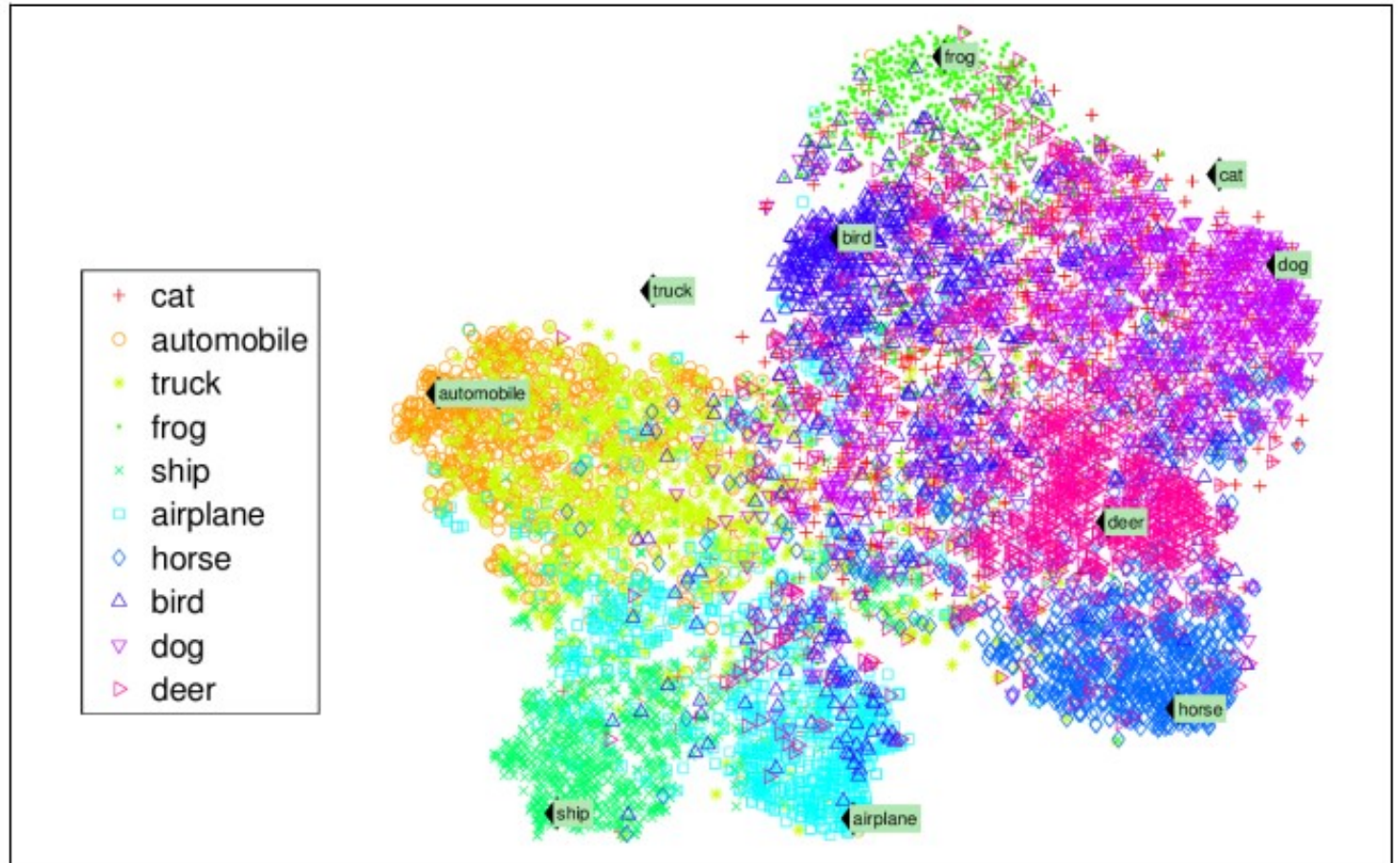
# Supervisão humana

- Aprendizado não-supervisionado (agrupamento)



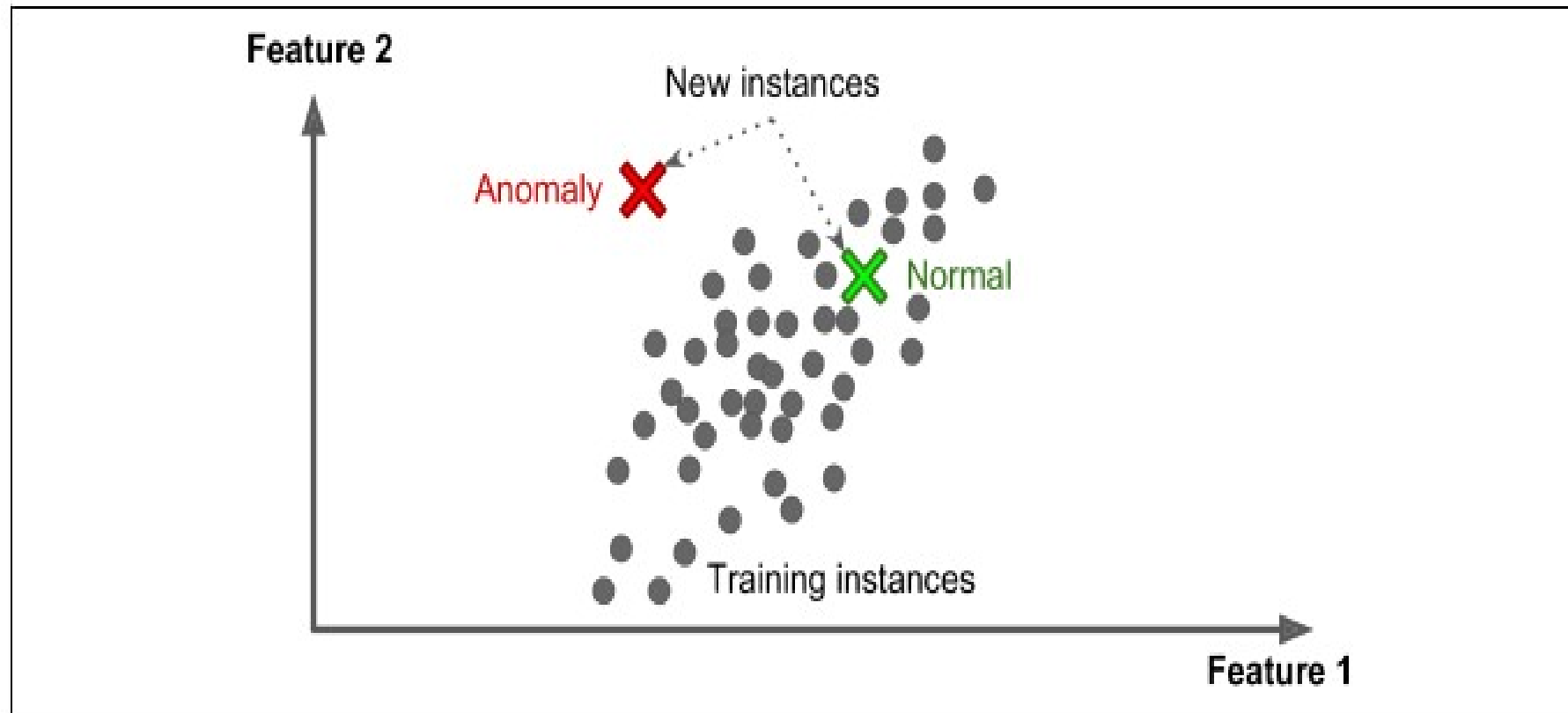
# Supervisão humana

- Aprendizado não-supervisionado (redução de dimensionalidade)



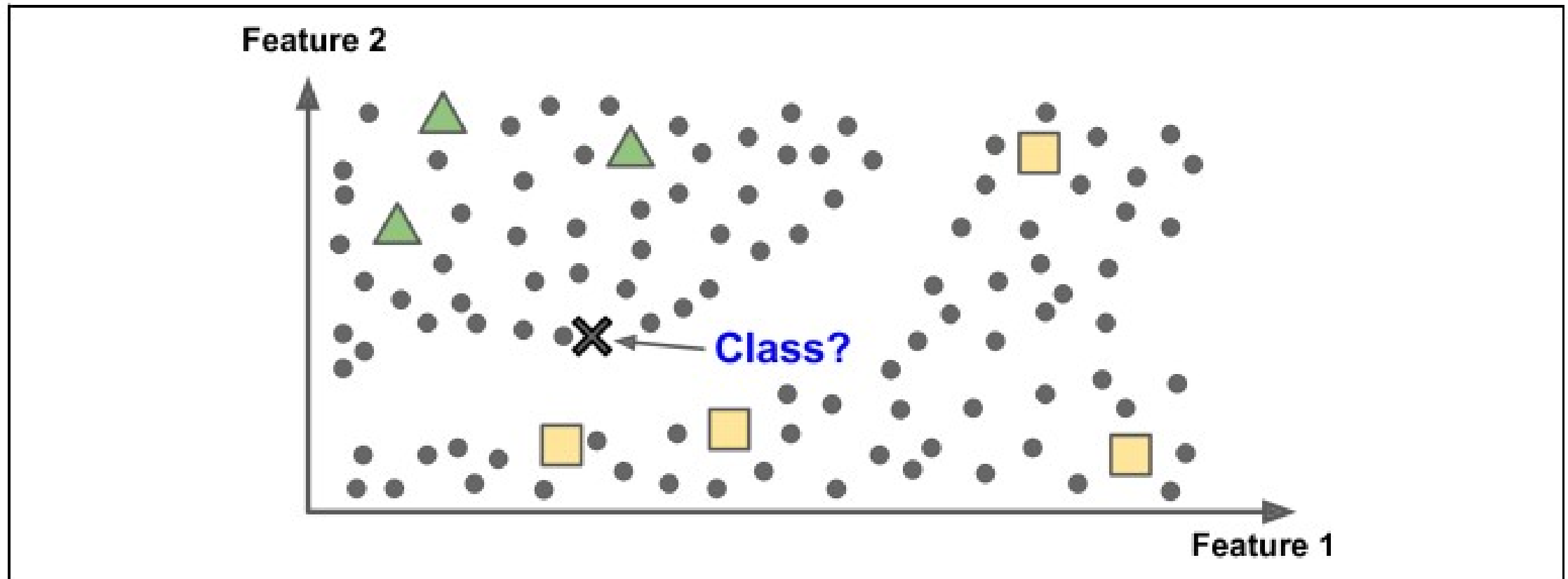
# Supervisão humana

- Aprendizado não-supervisionado (detecção de anomalia)



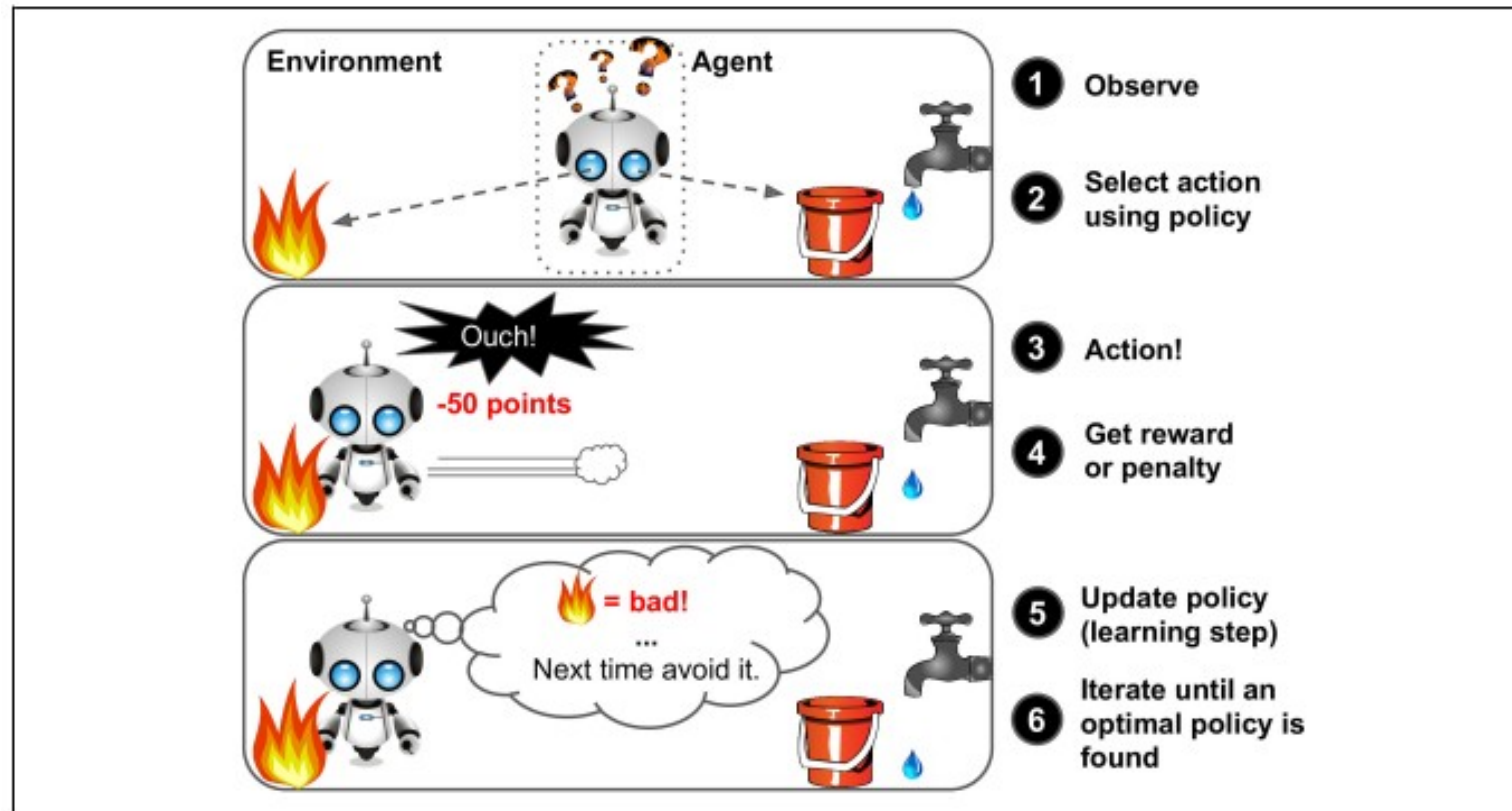
# Supervisão humana

- Aprendizado semi-supervisionado (ex: Google Photos)



# Supervisão humana

- Aprendizado por reforço



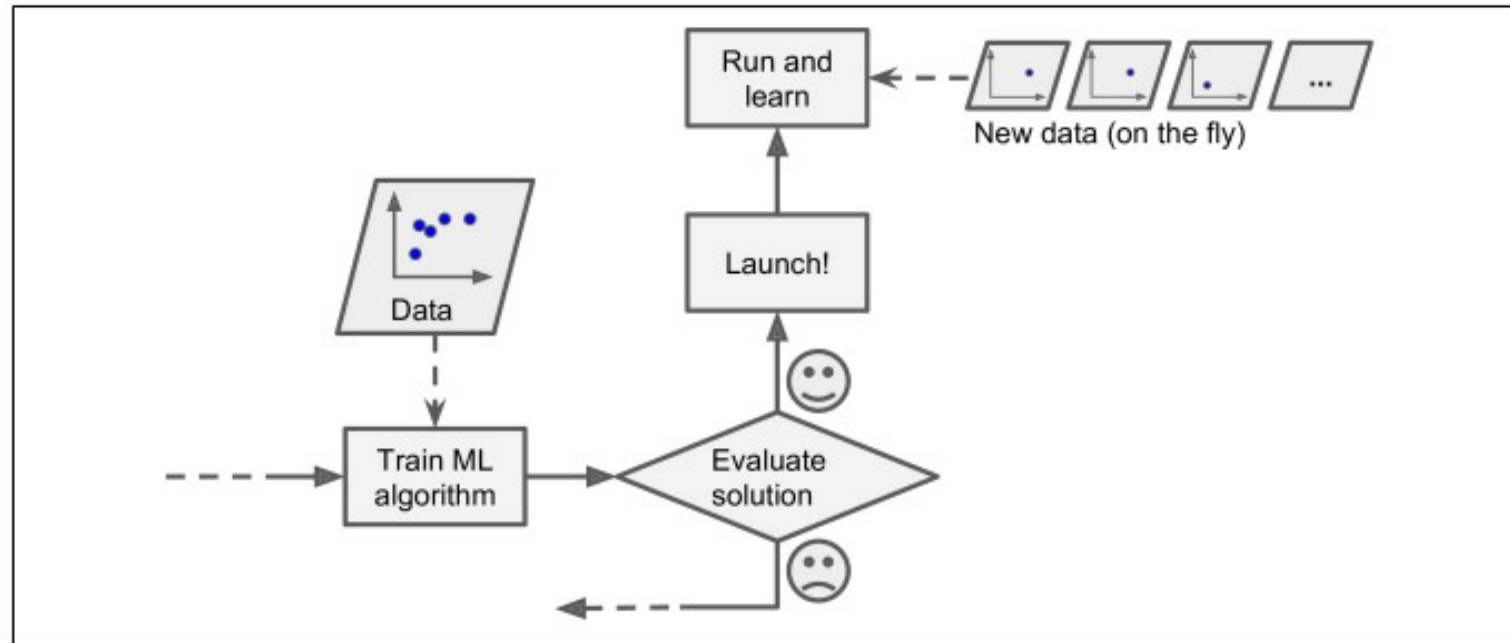
An illustration on the left side of the slide. It features a row of seven light bulbs hanging from above. The first six bulbs are white with black outlines. The seventh bulb, which is slightly lower than the others, is yellow and glowing, with several short lines radiating from it to represent light. A white hand is shown reaching up from the right, with its index finger pointing towards the glowing yellow bulb. The background behind the bulbs is dark, and the hand is white.

# Aprendizado Incremental

- Aprendizado em lote:
  - O sistema é incapaz de aprender incrementalmente
  - Precisa ser treinado usando todos os dados disponíveis
  - Depende de muitos dados e recursos computacionais
  - Uma vez treinado e lançado em produção, roda sem aprender mais nada, a não ser que refaça todo o processo de aprendizagem

# Aprendizado incremental

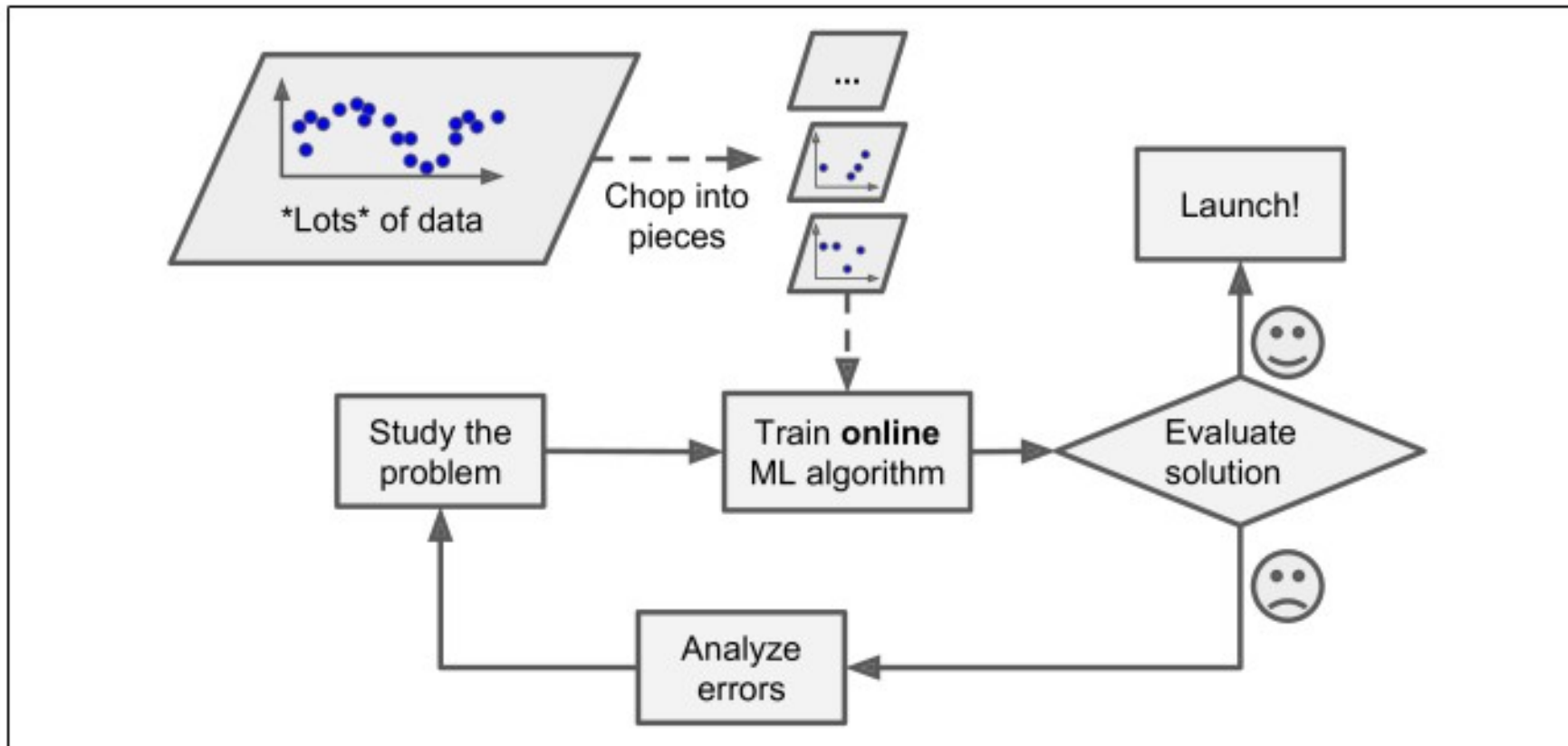
- Aprendizado online:
  - Fluxo contínuo de dados





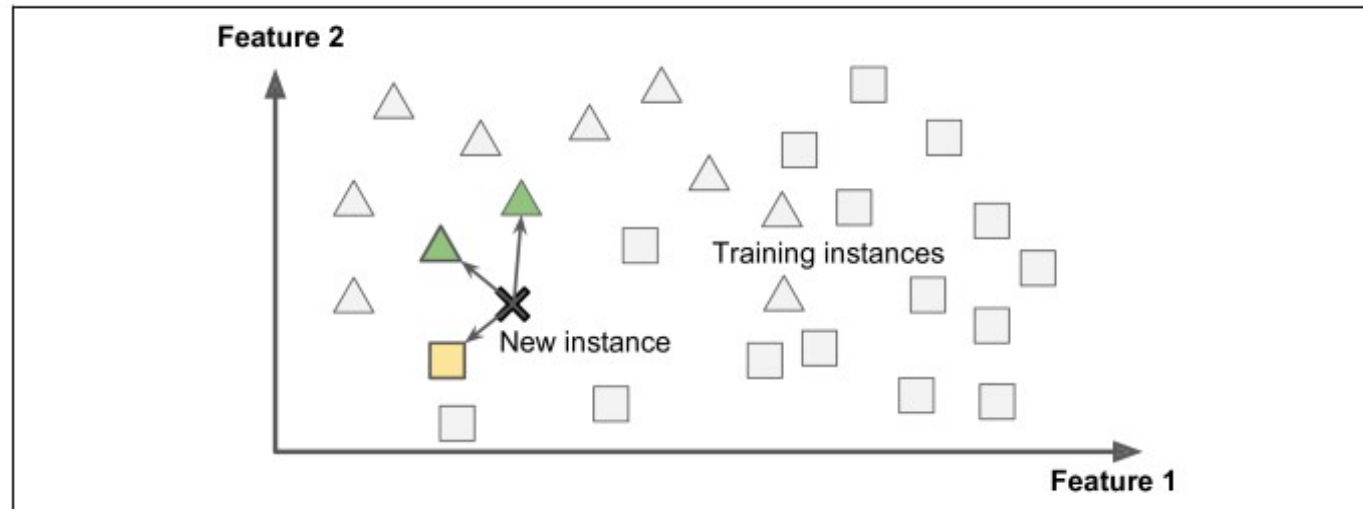
# Aprendizado incremental

- Aprendizado online:
  - Grande quantidade de dados que não cabe na memória do computador



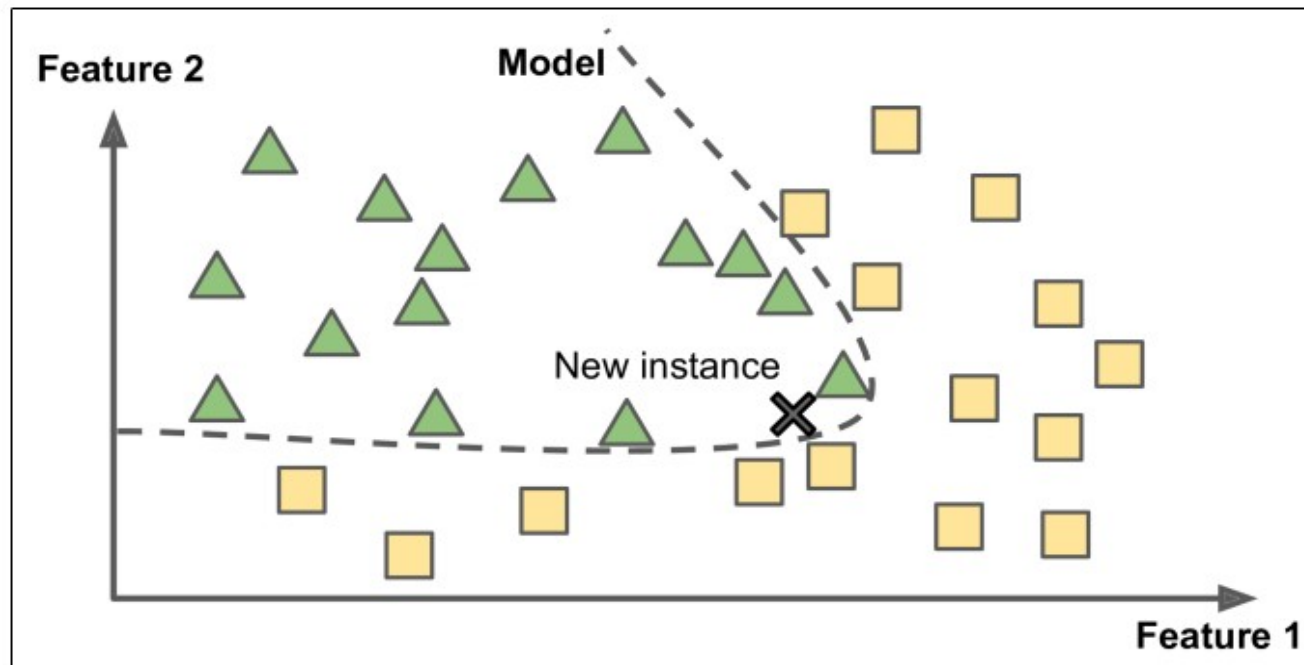
# Aprendizado baseado em **instâncias** x modelo

- Necessitam de uma medida de similaridade
- Aprende por memorização
- Depois generaliza para novos casos



# Aprendizado baseado em instâncias x modelo

- Necessitam definir valores de parâmetros
- Encontrar o melhor modelo que divide os dados
- Minimizar a função custo



# Principais desafios em aprendizado de máquina

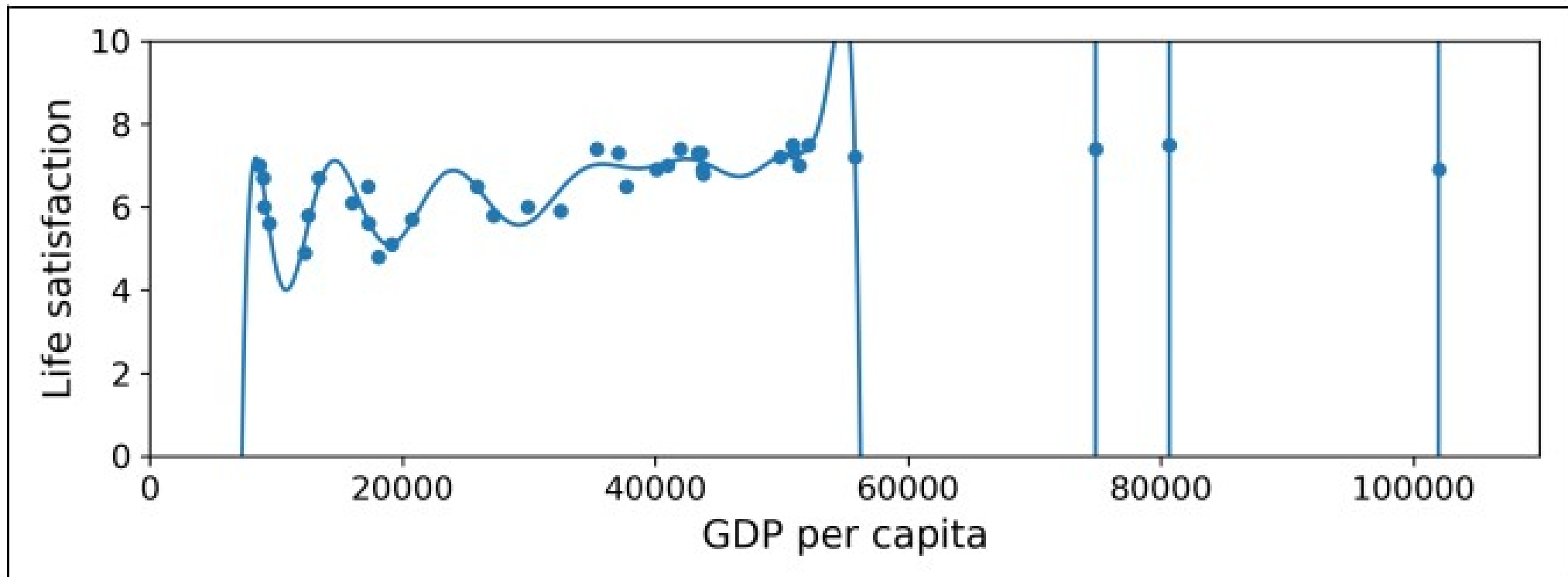
- Quantidade insuficiente de dados de treinamento
  - Os algoritmos dependem de dados
  - Problemas simples requerem milhares de exemplos
  - Problemas complexos requerem milhões de exemplos
- Dados de treinamento não representativo
  - Para o algoritmo generalizar bem, é crucial que os dados de treinamento sejam representativos para novos casos

# Principais desafios em aprendizado de máquina

- Dados de baixa qualidade
  - Erros, anomalias e ruídos nos dados são os piores inimigos dos algoritmos de aprendizado de máquina
  - Dados ausentes precisam ser tratados
- Atributos irrelevantes
  - Seu sistema só irá aprender se conter atributos relevantes suficientes e não muitos atributos irrelevantes.
  - **Seleção de atributos:** selecionar os atributos mais relevantes
  - **Extração de atributos:** combinar atributos existentes para produzir atributos melhores

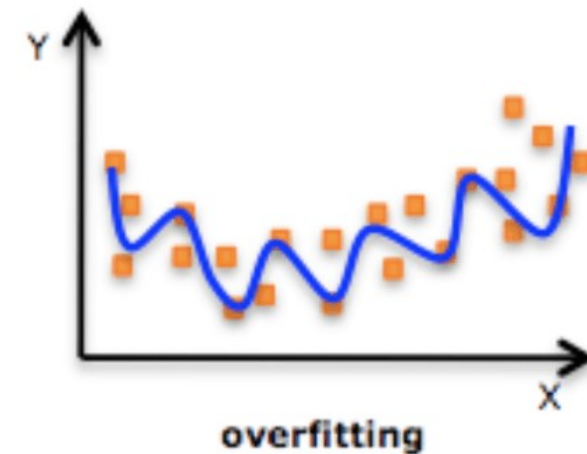
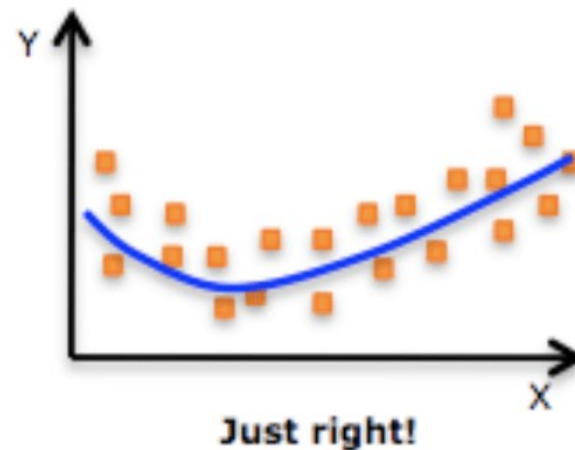
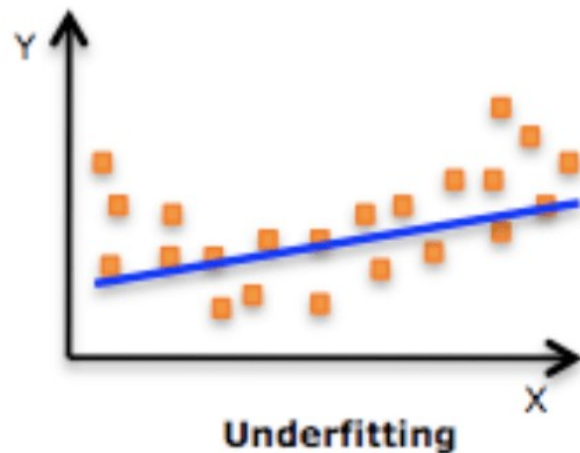
# Principais desafios em aprendizado de máquina

- Sobreajuste acontece quando o modelo é muito complexo para a quantidade de dados e ruído existente nos dados de treinamento



# Principais desafios em aprendizado de máquina

- Subajuste é o oposto do sobreajuste
- O subajuste ocorre quando seu modelo é tão simples para aprender a estrutura fundamental dos dados
- Selecione um modelo mais poderoso ou melhore os atributos de entrada





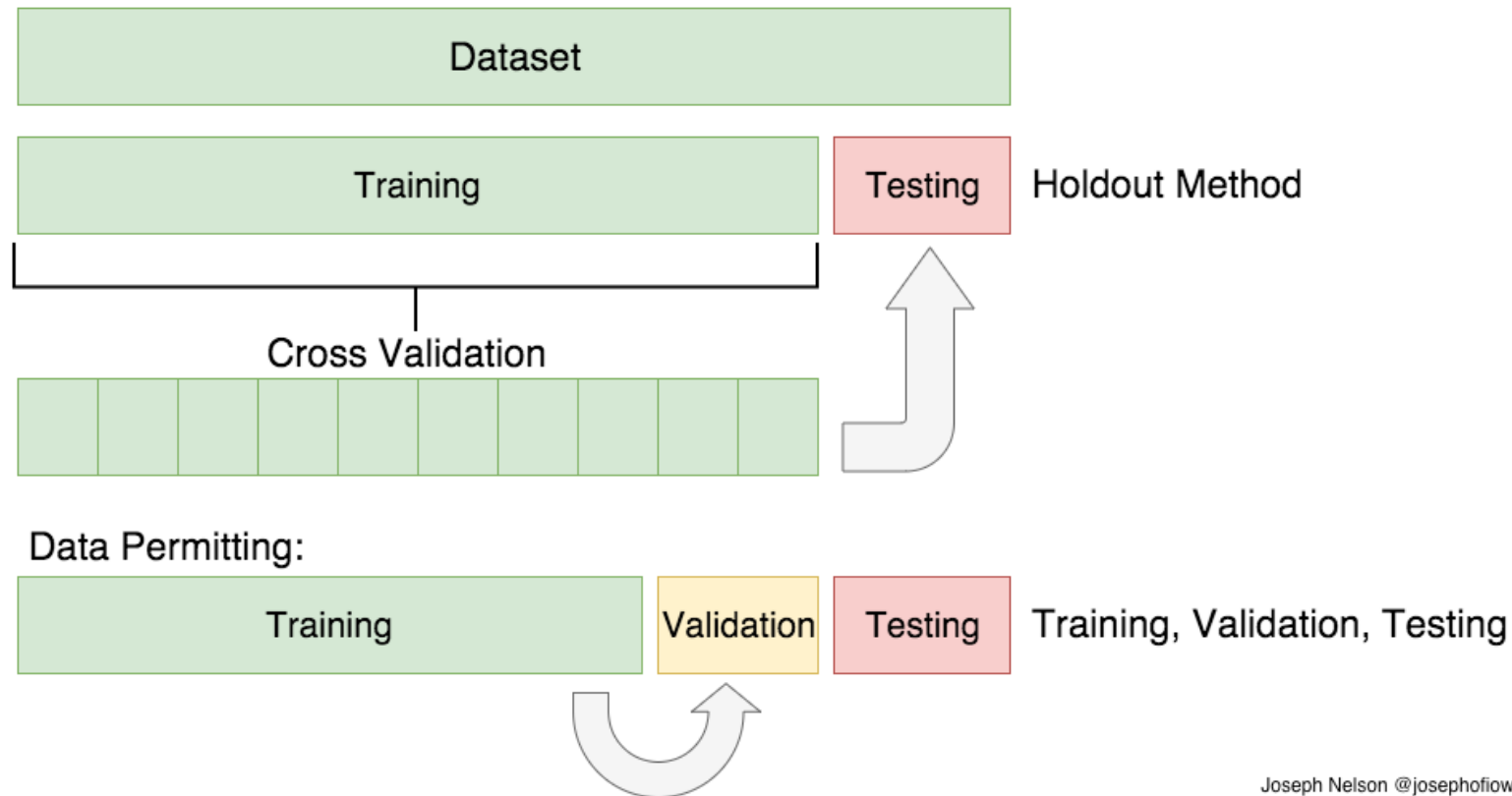
# Teste e validação

- Nunca teste seu modelo com os mesmos dados de treinamento.
- É comum dividir os dados em conjunto de treino e teste (80/20)
- Se o erro de treino é baixo, mas o erro de teste é alto, significa que seu modelo está com sobreajuste nos dados de treino.



# Conjunto de validação e validação cruzada

- Recurso usando quando se quer fazer ajuste de hiper-parâmetros ou seleção de modelos



# Atividade 1

- Pesquisar e fazer resumo de um artigo na área que você deseja fazer seu projeto transversal