

MÓDULO 1

# INTRODUÇÃO IA E CD, Técnicas e Algoritmos de aprendizado de máquina,

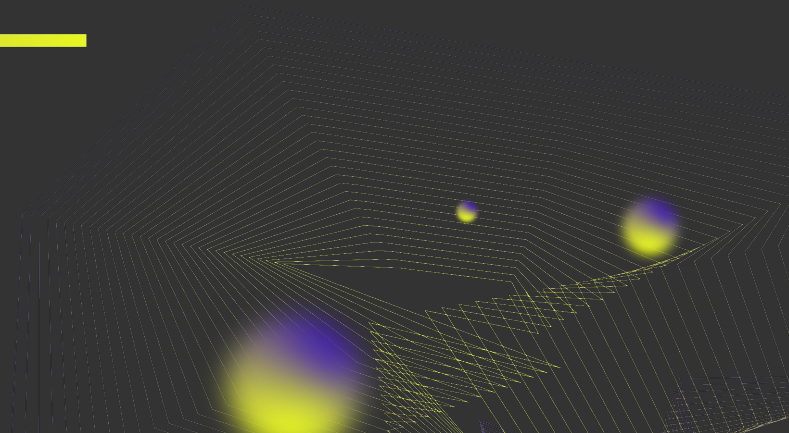
---

ESPECIALIZAÇÃO  
INTELIGÊNCIA ARTIFICIAL  
& CIÊNCIA DE DADOS

**SEAD**  
**UFES** | Superintendência de  
Educação a Distância

**Vinícius Fernandes Soares Mota**

Doutor/Professor Departamento de Informática



# ÍNDICE



1. Tipos de aprendizado
2. Aprendizado Supervisionado
3. Aprendizado Não supervisionado
4. Aprendizado por reforço



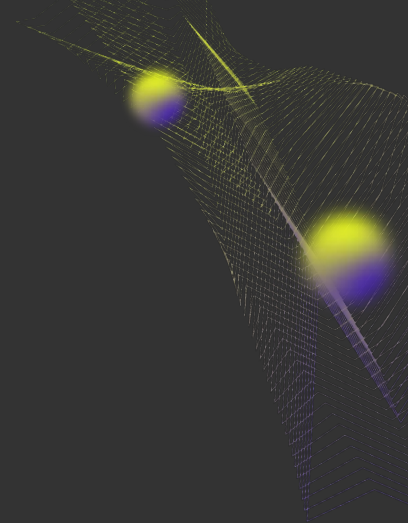
# 1. Tipos de Aprendizado

Em busca de uma definição



# Aprendizado de máquina

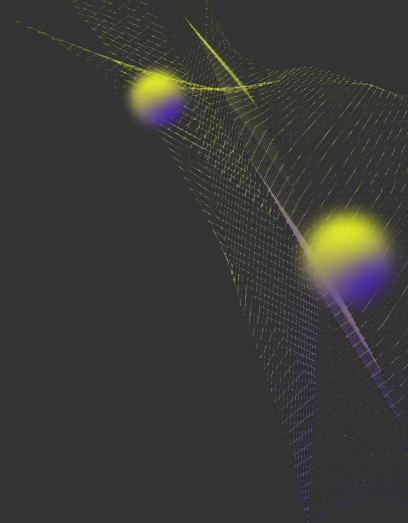
- Consiste em deixar um computador encontrar um caminho para resolver um problema usando dados.
- Esta aula fornece uma visão geral das classes de problemas e tipos de algoritmos de aprendizado.
- Tais algoritmos visam criar e treinar os modelos que
  - Prever valores futuros
  - Classificar os dados em grupos ou categorias.





# Tipos de aprendizado de máquina

- **Aprendizado supervisionado**
  - Dia: ensolarado/nublado/chuvoso
  - Email é um spam ou não
- **Aprendizado não supervisionado**
  - Detectar fraude bancária
  - Detecção de anomalias nos dados
- **Aprendizado semi-supervisionado**
- **Aprendizado por reforço**
  - Aprender um jogo
  - Aprender a dirigir





# Problemas Classificação VS Regressão

## Classificação:

- Quando a saída for um valor de um conjunto finito de valores
  - Dia: ensolarado/nublado/chuvoso
  - Email é um spam ou não

## Regressão:

- Quando a saída for um valor contínuo
  - Prever a temperatura amanhã
  - Número de casos COVID-19

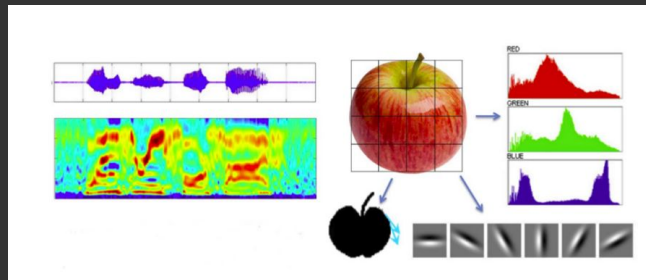


## 2. Aprendizado supervisionado

Para problemas cujo valor esperado é conhecido

# Aprendizado supervisionado

- O algoritmo recebe um conjunto de **atributos X** e uma **classe Y** (X, Y)
  - A classe é chamada de rótulos (*labels*) ou alvo (*target*)
- **Atributo (feature)** é uma descrição de um objeto.
  - Qualquer descrição de um objeto pode ser tratado como um atributo, seja um número, um texto, uma data, um booleano etc.

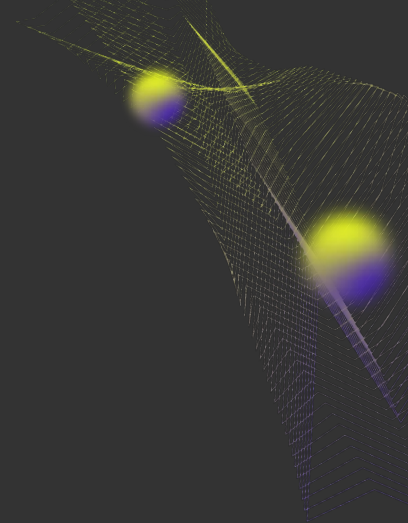






# Aprendizado supervisionado

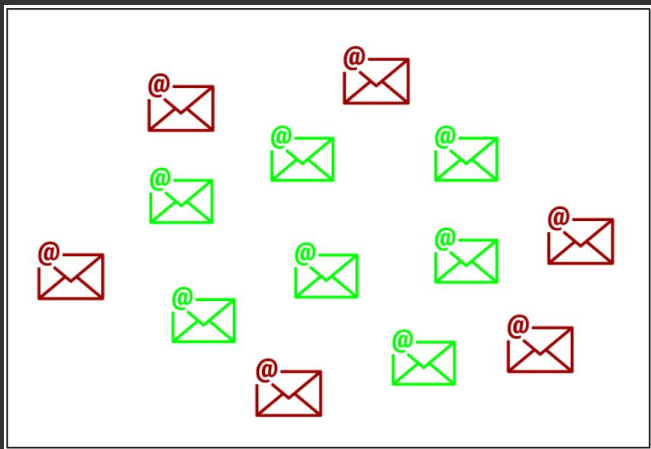
- O algoritmo recebe um conjunto de **atributos X** e uma **classe Y**  $(X, Y)$ 
  - A classe é chamada de rótulos (*labels*) ou alvo (*target*)
- O termo **supervisionado surge porque existe um supervisor** indicando qual é a resposta certa
- O **detector de spam é um bom exemplo de problema supervisionado**



# Aprendizado supervisionado

Rótulos:

- Spam
- Não Spam



Novo email



?



# Aprendizado supervisionado

- **Cada e-mail é definido como uma amostra ou instância do dataset**
  - Terminologia: uma amostra é representada por um conjunto de atributos **X**
  - A tupla  $(X, Y)$  indica que os atributos **X** são da classe **Y**
- **Uma amostra pode conter diversos atributos ou *features***
  - Exemplo de atributos: número de palavras, quantidade de vogais, existência de uma palavra específica, etc
  - Terminologia: os atributos são representados por  $(x_1, x_2, x_3, \dots, x_d)$
  - **Obs**: caso exista múltiplas saídas, **Y** é representado da mesma forma:  $(y_1, y_2, \dots, y_z)$



# Aprendizado supervisionado

## Definindo formalmente

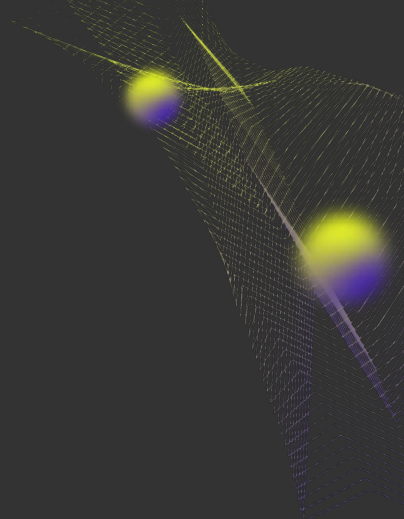
Dado um conjunto de treino de  $N$  pares de exemplos de entrada e saída

$(x_1, y_1), (x_2, y_2), \dots (x_N, y_N),$

em que cada par foi gerado por uma função desconhecida

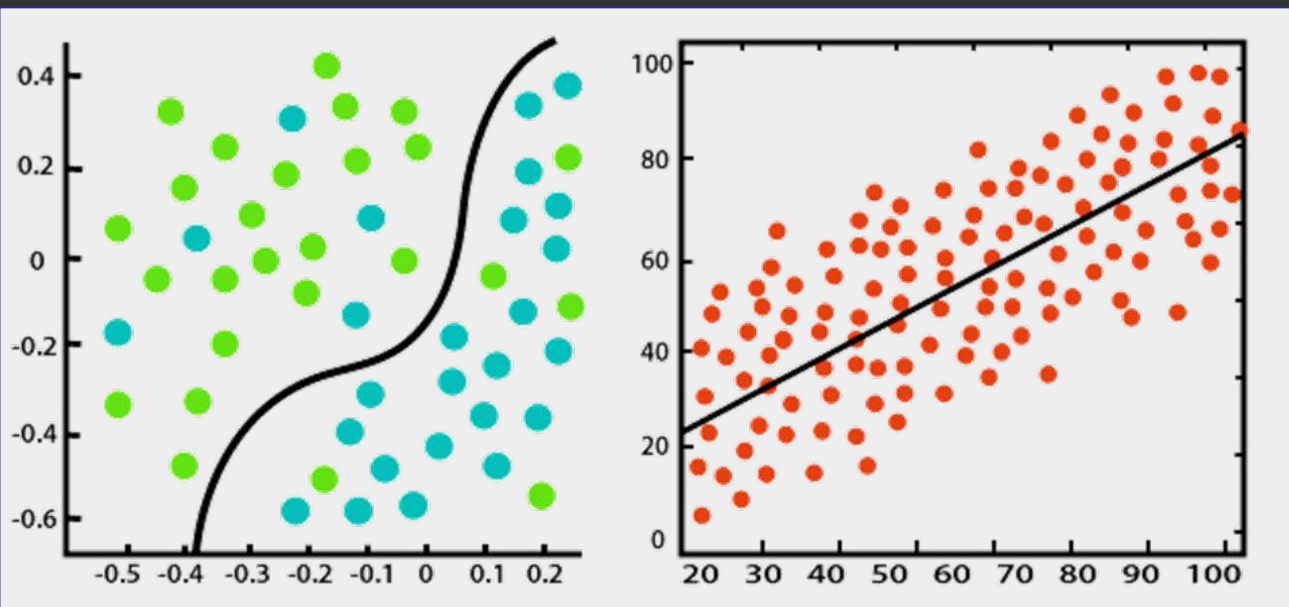
$y = f(x),$

descobrir uma função  $h$  que se aproxime da função verdadeira  $f$ .



# Aprendizado supervisionado

## Classificação vs Regressão



**CLASSIFICAÇÃO**

**REGRESSÃO**



# Aprendizado supervisionado

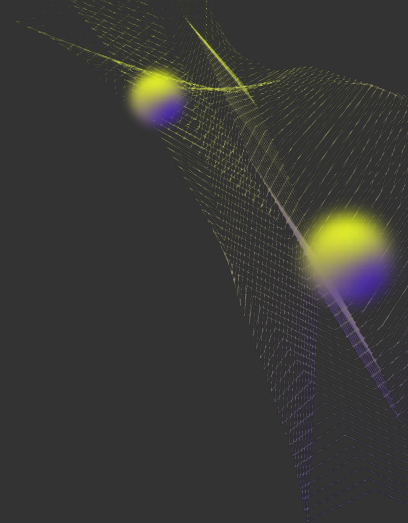
- **São exemplo de problemas supervisionados:**
  - Classificação de dados
    - Detector de spam
    - Detector de faces
    - Detector de câncer
  - Regressão
    - Determinar o preço de um carro
    - Determinar nível de saúde de uma pessoa
  -





# Aprendizado supervisionado

- **São exemplos de algoritmos supervisionados**
  - KNN
  - árvore de decisão
  - naive bayes
  - regressão linear,
  - etc





# 3. Aprendizado não supervisionado

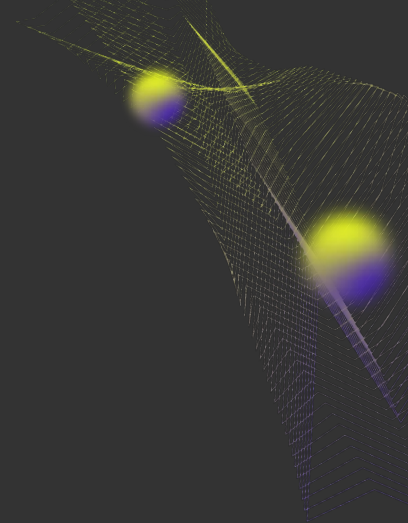
Para problemas sem rótulos conhecidos





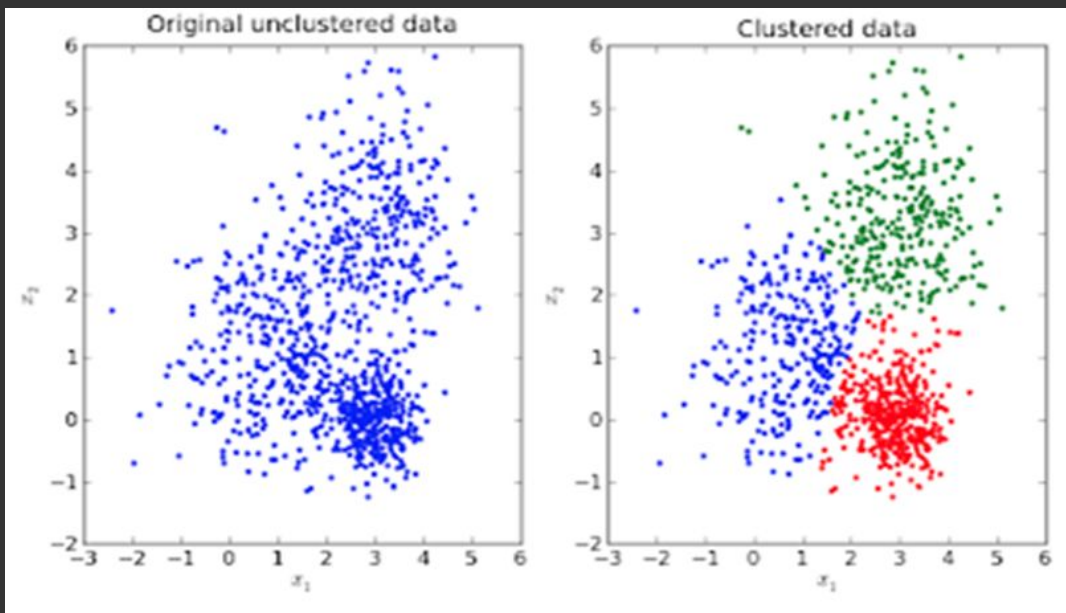
# Aprendizado não supervisionado

- **Em problemas não supervisionado, a solução não é fornecida**
  - Em outras palavras, só temos acesso a  $X$
  - O objetivo do modelo é obter  $Y$  **sem** supervisão
- **Não existe a “figura” do instrutor/supervisor**
  - O modelo deve entender o relacionamento entre os dados por ele mesmo



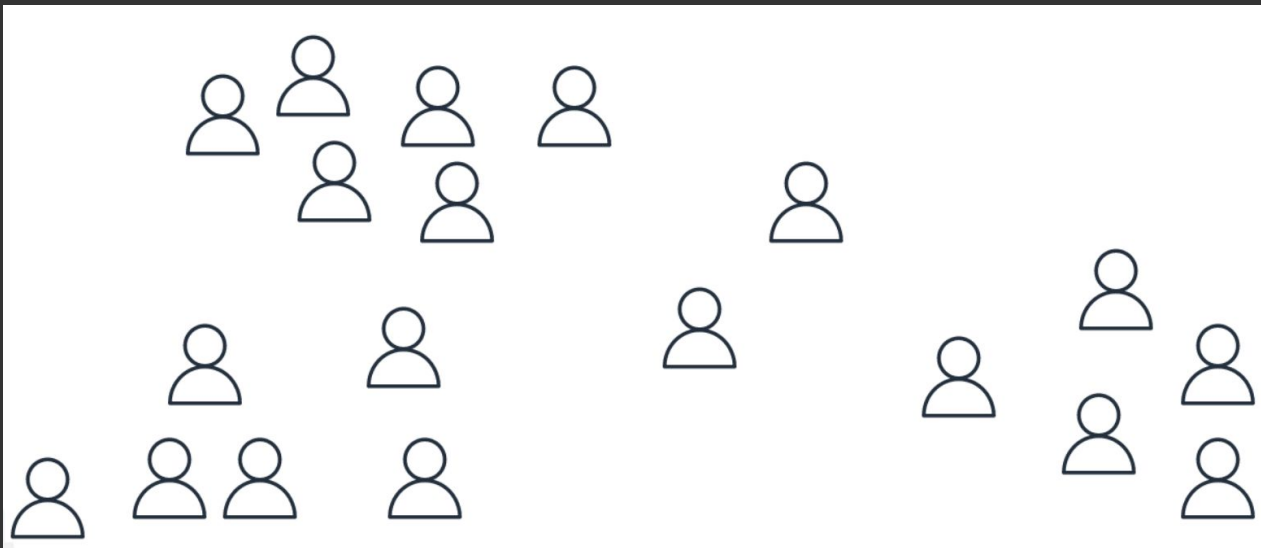
# Aprendizado não supervisionado

Consiste em mapear a entrada para a saída com base em exemplos apenas de amostras de entrada



# Aprendizado não supervisionado

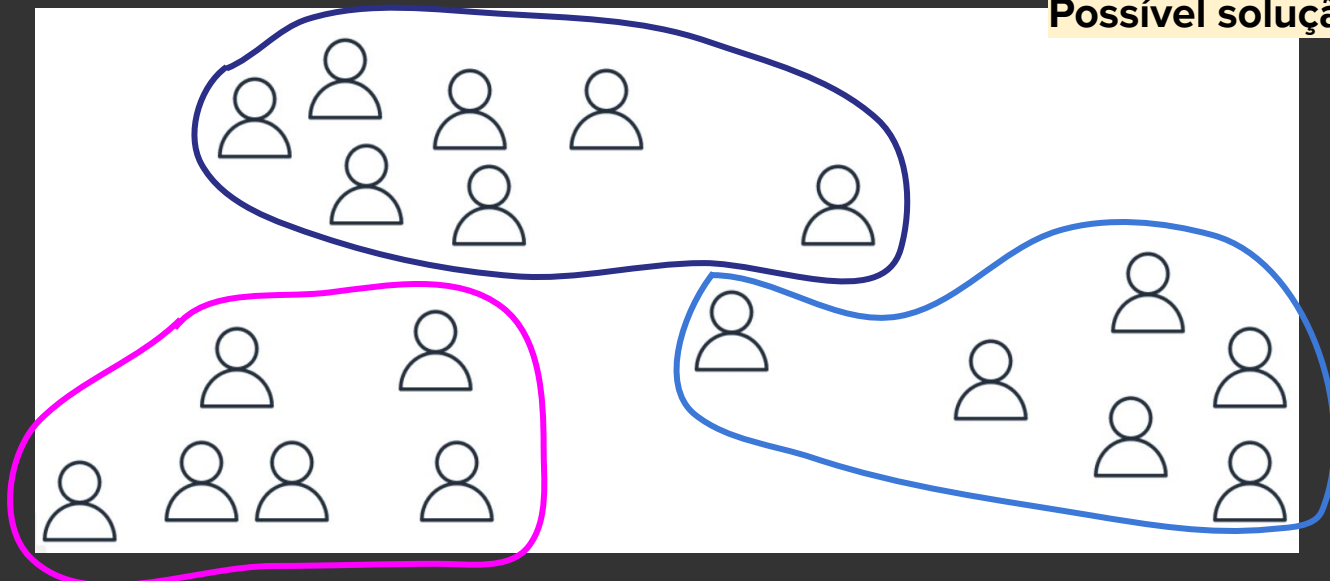
**Por exemplo, selecionar usuários com perfis similares em um banco de dados**



# Aprendizado não supervisionado

Por exemplo, selecionar usuários com perfis similares em um banco de dados

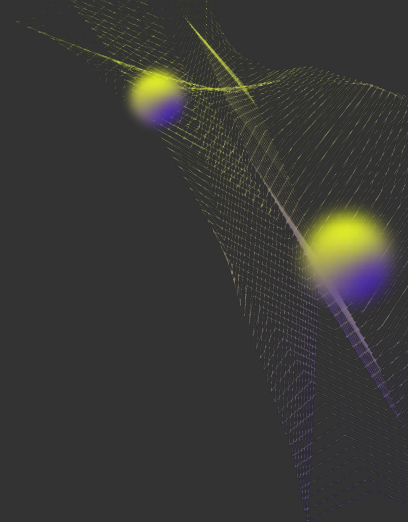
Possível solução





# Aprendizado não supervisionado

- **São exemplo de problemas não supervisionados:**
  - Clusterização / agrupamento
    - Identificar perfis similares de usuários
    - Agrupar casas semelhantes
  - Estimação de densidade
    - Estimar a distribuição de uma base dados
  - Visualização de redução dimensionalidade
    - Simplificar os dados sem perder muita informação
  - Detecção de anomalia
    - Identificar fraudes em cartão de crédito





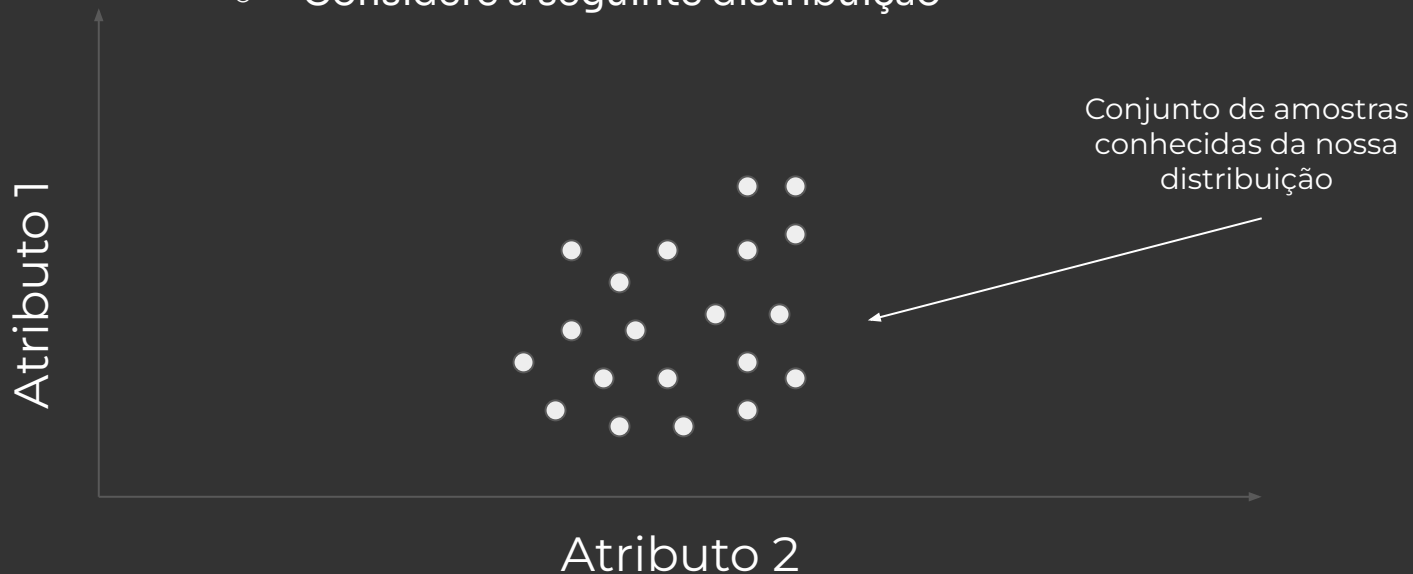
# Aprendizado não supervisionado

- **São exemplos de algoritmos supervisionados:**
  - K-Means
  - PCA
  - t-SNE
  - etc



# Aprendizado não supervisionado

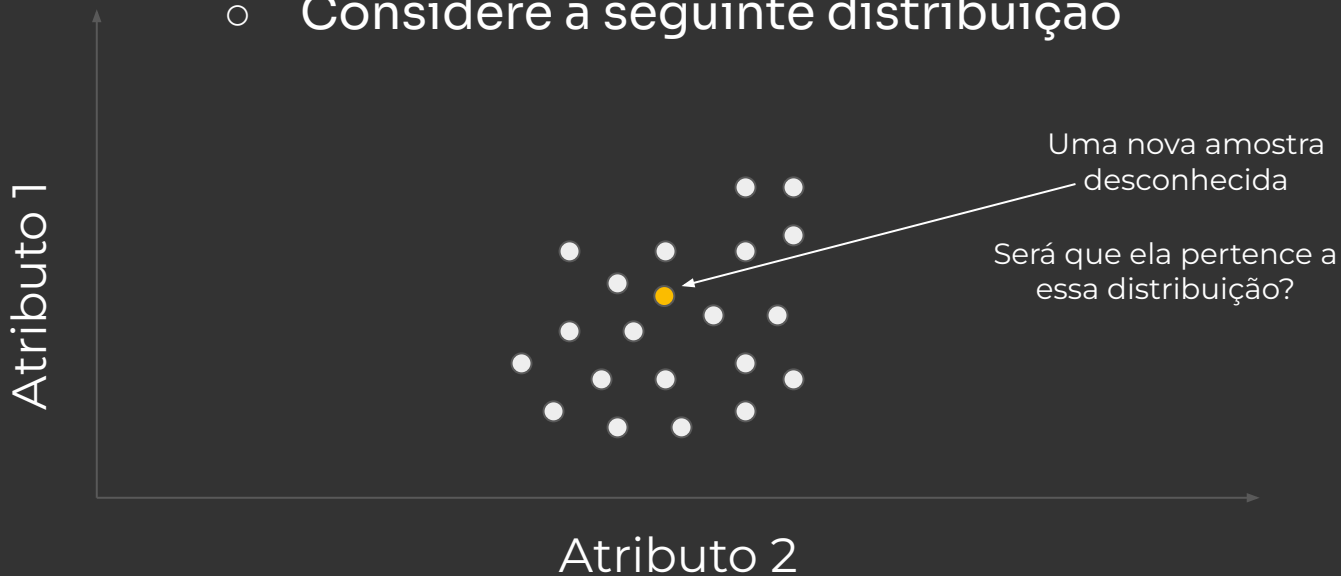
- Também utilizados para Detecção de anomalia:
  - Considere a seguinte distribuição



# Aprendizado não supervisionado

- **Deteção de anomalia:**

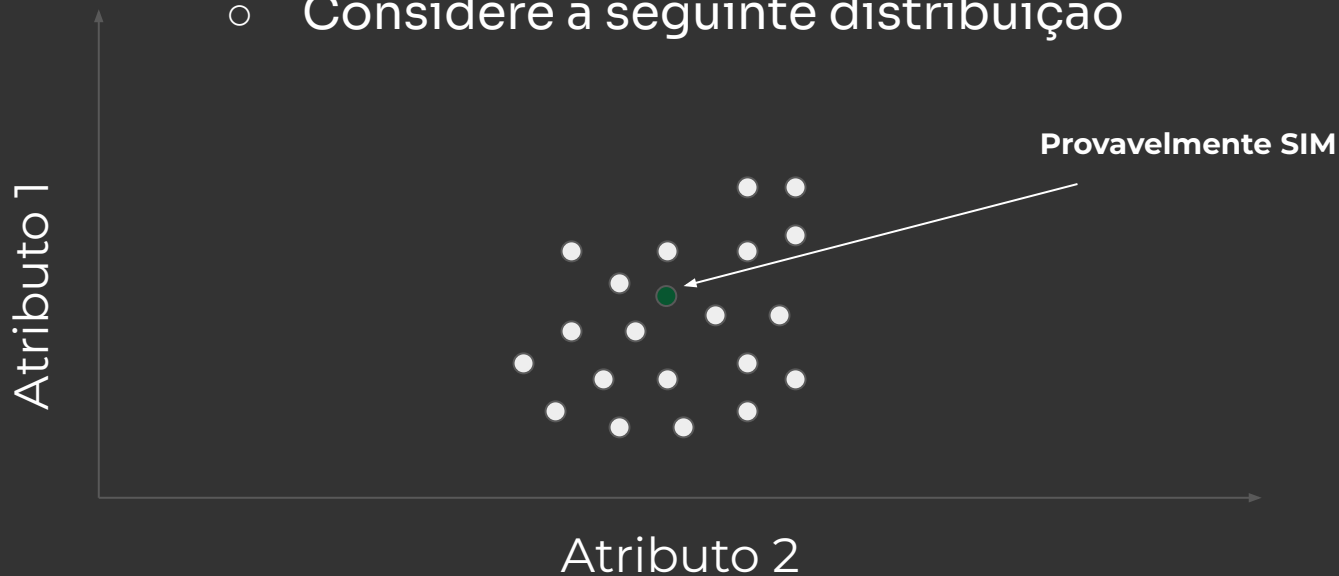
- Considere a seguinte distribuição





# Aprendizado não supervisionado

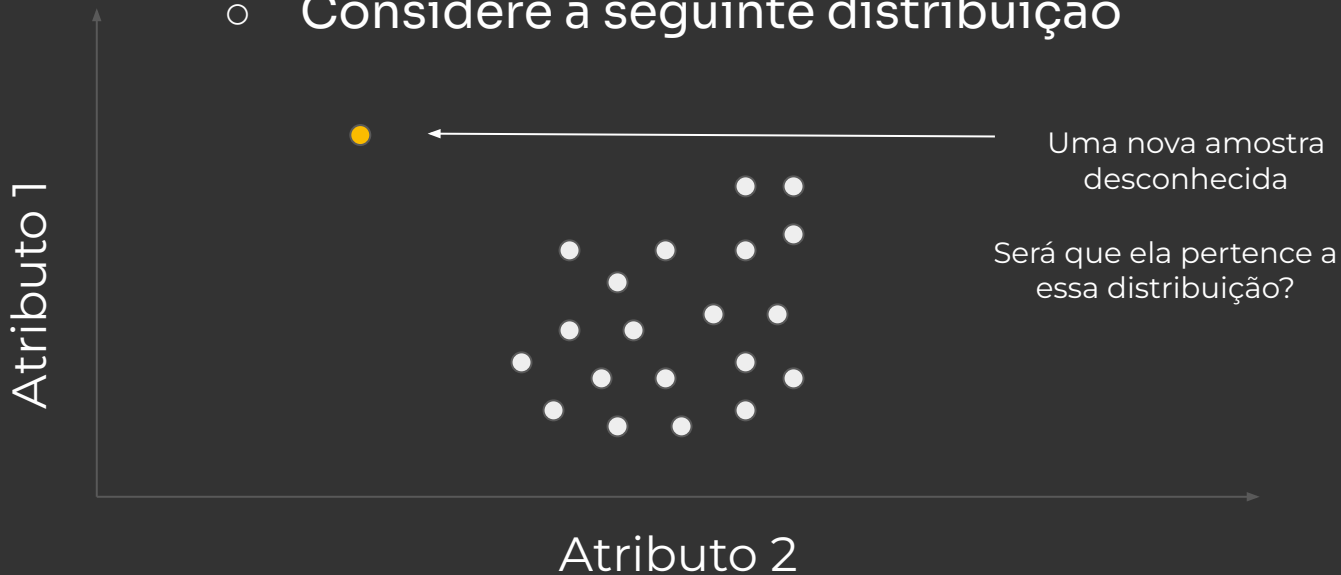
- **Deteccção de anomalia:**
  - Considere a seguinte distribuição



# Aprendizado não supervisionado

- **Deteccção de anomalia:**

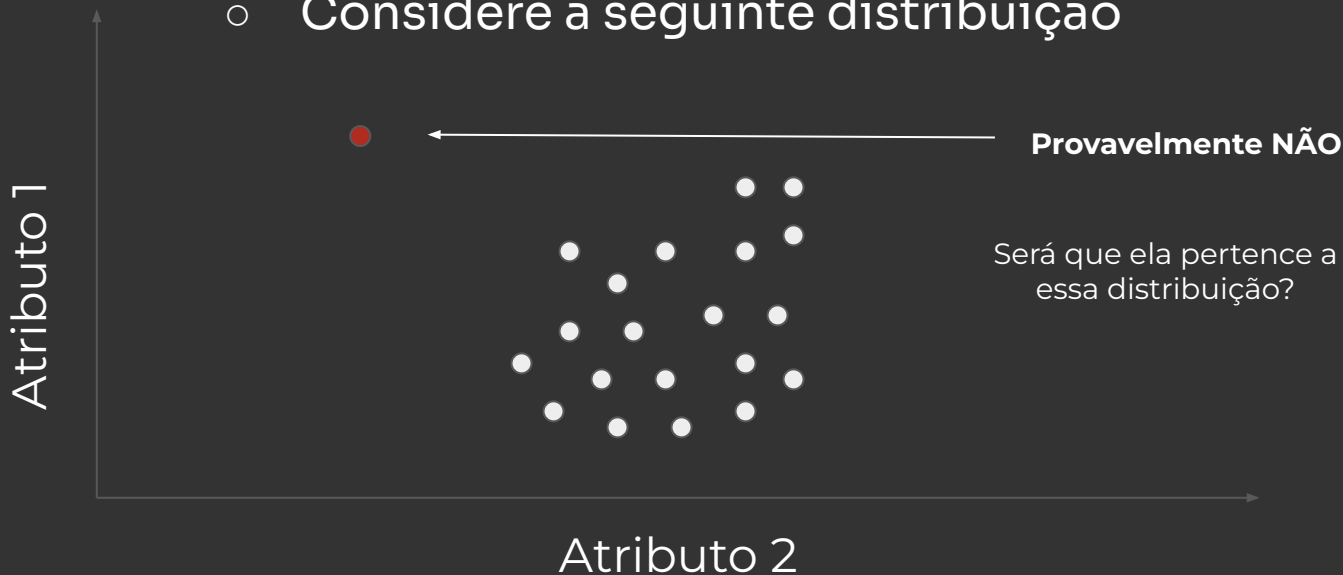
- Considere a seguinte distribuição



# Aprendizado não supervisionado

- **Deteccção de anomalia:**

- Considere a seguinte distribuição



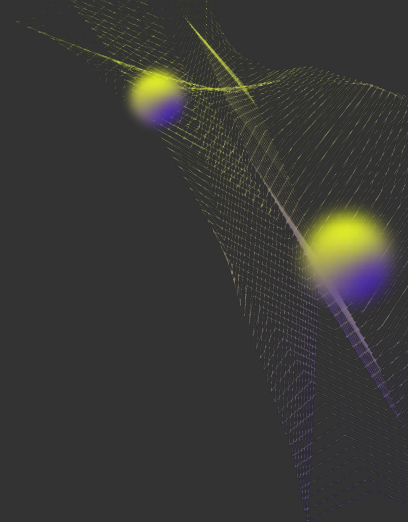


## 4. Aprendizado por reforço



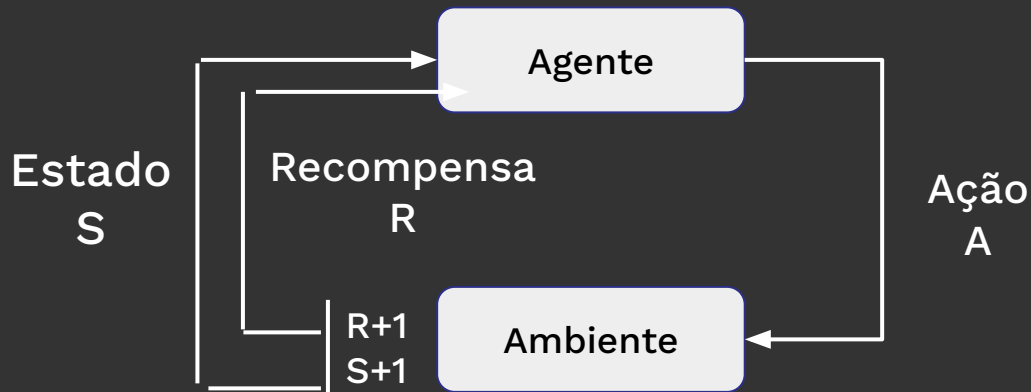
# Aprendizado por reforço

- **O objetivo é aprender como realizar uma sequência de ações em um ambiente para maximizar recompensas cumulativas.**
  - Um agente no aprendizado por reforço é o componente que toma decisões sobre quais ações devem ser tomadas.
  - Para alcançá-lo, o agente precisa ter a capacidade de interagir com o ambiente para obter dados (estado, ação e recompensa).



# Aprendizado por reforço

- O objetivo é aprender como realizar uma sequência de ações em um ambiente para maximizar recompensas cumulativas.



# Aprendizado por reforço





## 5. Quando usar cada tipo



# Diversas técnicas de aprendizado



# Algoritmos de aprendizado



# Algoritmos de aprendizagem



Estes algoritmos serão estudados em detalhes nas disciplinas:

- **Classificação de dados**
- **Clusterização e Redução de Dimensionalidade**
- **Regressão e Séries Temporais**

INTELIGÊNCIA  
ARTIFICIAL &  
CIÊNCIA DE DADOS

**Vinícius Fernandes Soares Mota**  
Professor/Departamento de Informática

*vinicius.mota@inf.ufes.br*