

Introdução à Ciência de Dados e Aprendizado de Máquina

Introdução ao
Aprendizado de Máquina



Robson Fernandes

Acadêmico

Mestrando em Matemática, Estatística e Computação Aplicadas (Data Science & Machine Learning) - USP

Especialização em Reconhecimento de Padrões e Análise de Imagens - UNICAMP

Pós-Graduado em Arquitetura de Software Distribuído - PUC-MG

MBA em Engenharia de Software Orientada a Serviços – SOA – METROCAMP

Certificado – JavaScript e HTML5 Developer – W3C INTERNACIONAL

Autor do Livro Gestão da Tecnologia da Informação: Teoria e Prática

Profissional

Cientista de Dados Sênior – Finch Soluções

Docente Pós-Graduação - MBA em Data Science & Machine Learning - UNIP

Docente Pós-Graduação em Engenharia de Software - USC

Docente Graduação em Ciência da Computação - UNIP

Site

<http://robsonfernandes.net>

e-mail

robson.fernandes@usp.br / robs.fernandes@outlook.com

Aprendizado de Máquina

Machine Learning (aprendizado de máquina) é uma área que representa uma evolução nos campos de Ciência da Computação, Análise de Dados, Engenharia de Software e Inteligência Artificial.

Computação Tradicional **vs** AM

- **Computação Tradicional**

- **Escreve algoritmo detalhando como resolver um problema.**

- **Aprendizado de Máquina**

- **Aprende a resolver um problema observando seus dados.**

Computação Tradicional **vs** AM

- **Computação Tradicional**

- **Pessoas programam.**

- **Aprendizado de Máquina**

- **Computadores programam.**

Computação Tradicional **vs** AM

- **Computação Tradicional**

- Programa que **funciona 90%** das vezes pode ser considerado ruim.

- **Aprendizado de Máquina**

- Modelo que **acerta 90%** das vezes pode ser considerado muito bom e eficiente.

Paradigmas de Aprendizado de Máquina

- **Simbólico**

- Buscam aprender construindo representações simbólicas
(Expressão lógica, Árvores de Decisão Regras)

- **Estatístico**

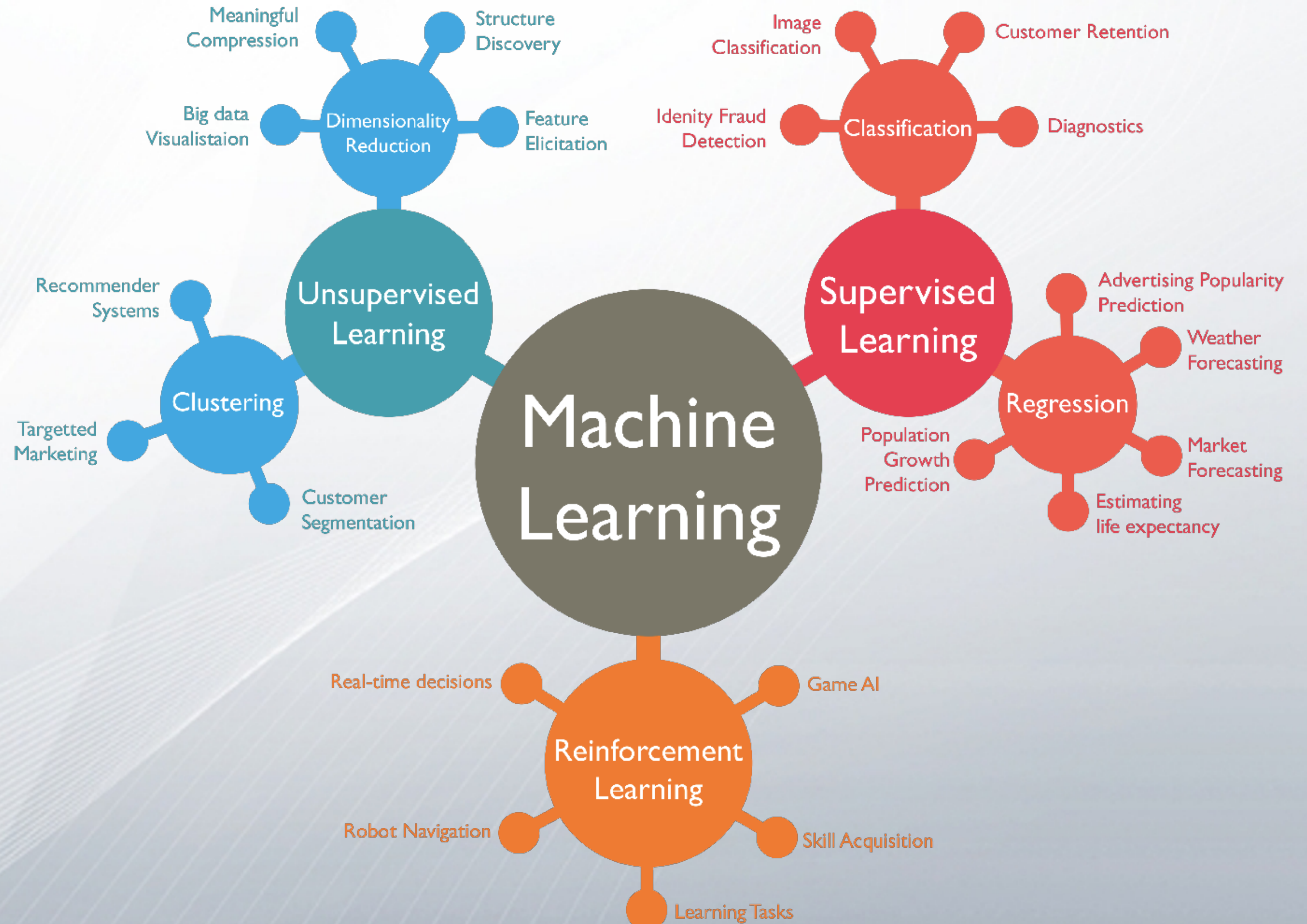
- Buscam métodos estatísticos
(Aprendizado Bayesiano, Séries Temporais, Regressão Linear)

Paradigmas de Aprendizado de Máquina

- **Conexionista**
- Modelos inspirados no modelo biológico do sistema nervoso
(**Redes Neurais**)
- **Evolutivo**
- Teoria de Darwin
(**Algoritmos Genéticos**)

Aprendizado de Máquina - Sub-Áreas

- **Aprendizado Supervisionado**
 - **Regressão**
 - **Classificação**
- **Aprendizado Não Supervisionado**
 - **Clustering**
 - **Redução de Dimensão**
- **Aprendizagem por Reforço**
 - **Decisão em Tempo real**
 - **Navegação de robôs**



Aprendizado de Máquina - Aprendizado Supervisionado

- **Aprendizado Supervisionado**

- É o termo usado sempre que o programa é “**treinado**” sobre um **conjunto de dados pré-definido**. Baseado no treinamento com os dados pré-definidos, o programa pode tomar decisões precisas quando recebe novos dados.



Aprendizado de Máquina - Aprendizado Não Supervisionado

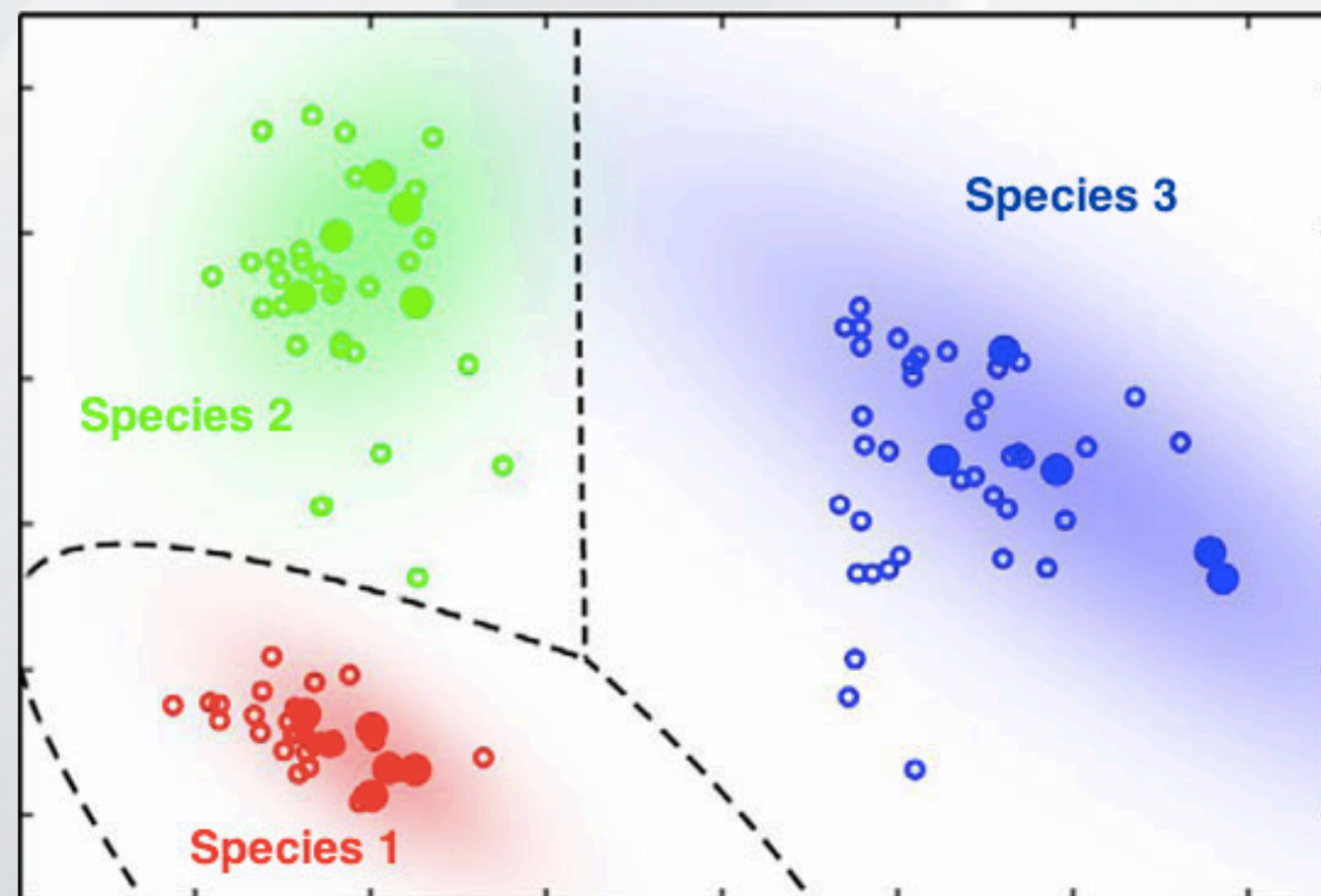
- **Aprendizado Não Supervisionado**
 - Termo usado quando um programa pode **automaticamente encontrar padrões e relações** em um conjunto de dados.



Aprendizado de Máquina - Classificação

- **Classificação**

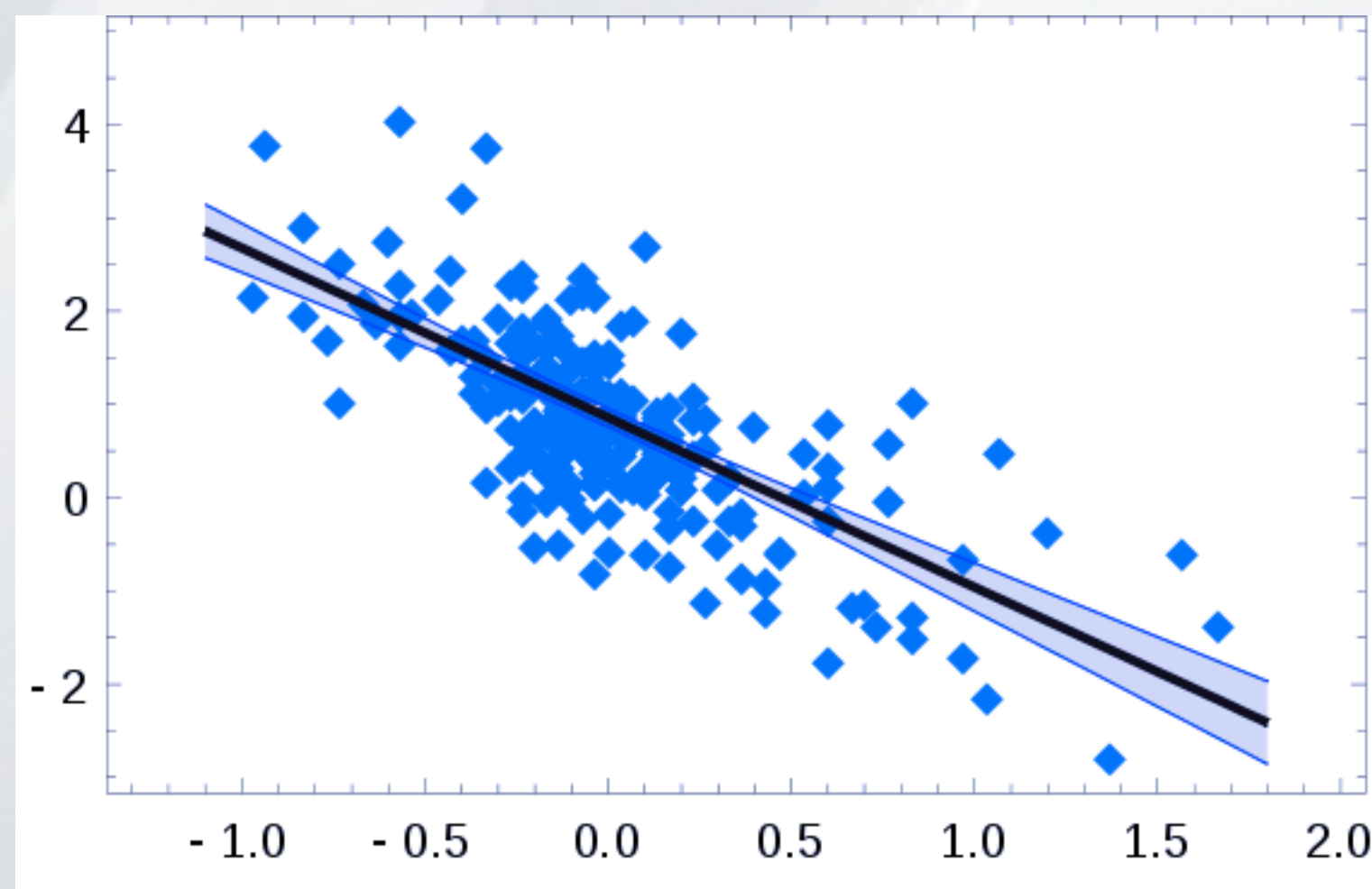
- A classificação é uma sub-categoria de aprendizagem supervisionada. É o processo de tomar algum tipo de entrada e atribuir um rótulo a ela.
- Sistemas de classificação são usados geralmente quando as previsões são de natureza distinta, ou seja, um simples “sim” ou “não”.
- **Exemplo: Mapeamento de uma imagem de uma pessoa e classifica-la como masculino ou feminino.**



Aprendizado de Máquina - Regressão

- **Regressão**

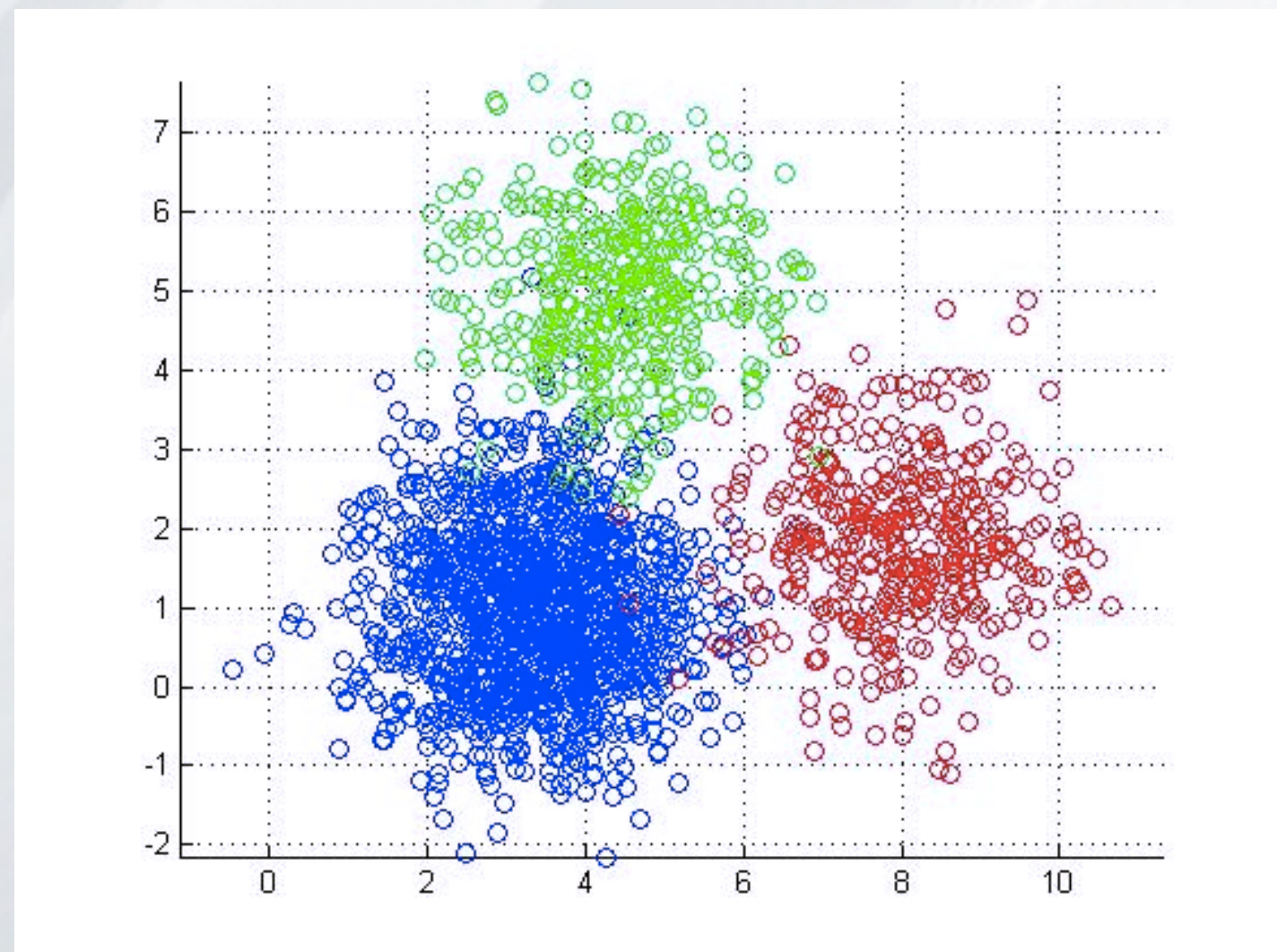
- A regressão é uma sub-categoria de aprendizagem supervisionada usada quando o valor que está sendo previsto **difere de um “sim” ou “não”** e que **siga um espectro contínuo**.
- **Exemplo: Sistemas de regressão poderiam ser usados, por exemplo, para responder às perguntas: “Quanto custa?” ou “Quantos existem?”.**



Aprendizado de Máquina - Clustering

- **Clustering**

- A clustering é uma sub-categoria de aprendizagem não supervisionada.
- O interesse principal é desvendar a organização dos padrões em clusters (agrupamentos) consistentes, os quais permitirão descobrir similaridades e diferenças entre padrões bem como derivar conclusões úteis a respeito deles.



Aprendizado de Máquina - Redução de Dimensão

- **Redução de Dimensão**
 - Em muitos casos, datasets possuem um número elevado de atributos e de observações (objetos).
 - Análise de dados complexa (muitos atributos):
 - Pode ficar muito cara computacionalmente se todo o conjunto de dados for considerado;
 - Dependendo do tamanho do dataset, os algoritmos podem não rodar satisfatoriamente.
 - **Solução** \Rightarrow Redução de Dimensão.
 - **Redução de atributos e/ou objetos.**



Bibliografia

- SILVA. L. A.; PERES. S. M; BOSCARIOLI C. **Introdução à Mineração de Dados**. Elsevier. 2016
- FACELI, Katti; Lorena, Ana Carolina; Gama, João ; de Carvalho, A. C. P. L. F. (2011). **Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina**. 1. ed. Rio de Janeiro: LTC.
- PROVOST, F.; Fawcett, T. **Data Science for Business: What you need to know about data mining and data-analytic thinking** by O'Reilly Media, 2013.
- FLACH, P. (2012). **Machine Learning: The Art and Science of Algorithms that Make Sense of Data**. Cambridge University Press.
- ALPAYDIN, E. (2004). **Introduction to Machine Learning**. MIT Press.