

Aula 01- Introdução

1001524 – Aprendizado de Máquina I
2023 - I Turmas A, B e C
Prof. Dr. Murilo Naldi

naldi@ufscar.br

O que é aprendizado de máquina?

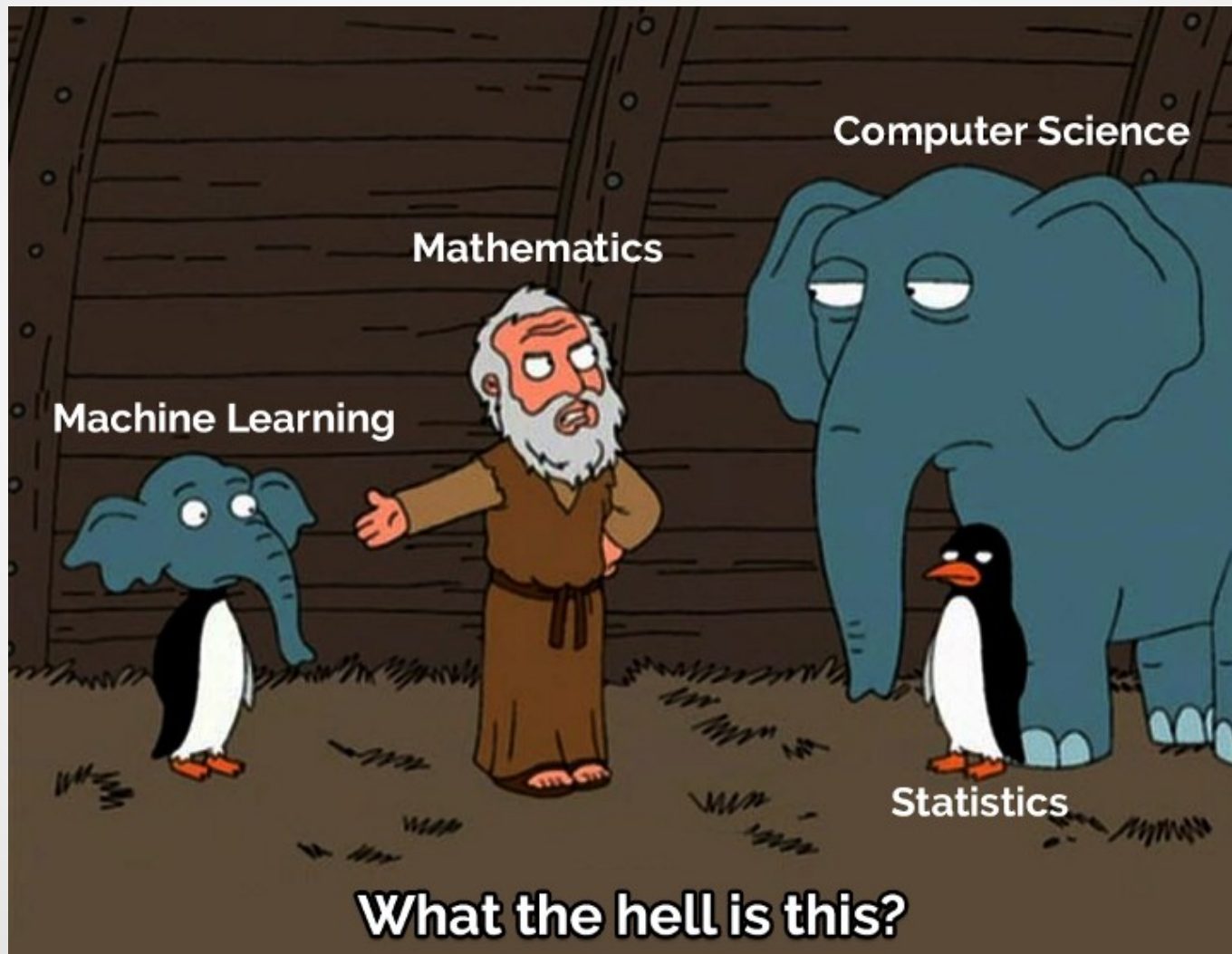
- ?

O que é aprendizado de máquina?

- “Machine Learning: the field of study that gives computer the ability to learn without being explicitly programmed.”
 - Arthur Samuel, 1959



O que é aprendizado de máquina?



O que é aprendizado de máquina?

- “A computer programming is said to learn from experience E with respect to some task T and some performance measure P if its performance on T , measured by P , improves with E .”
 - Tom Mitchell, 1998



Exemplos de aplicação

- Gerenciadores de emails sabem quando eles são marcados como *spam*. Com isso, ele aprende a filtrar melhor essas mensagens. Perguntas: o que é T, nesse caso? P? E?

“A computer programming is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, measured by P, improves with E.”

Tom Mitchell, 1998



Exemplos de aplicação

- **Marketing direcionado**
 - *Target*, uma gigantesca empresa de departamento americana com mais de 1800 lojas
 - Grandes quantidades de dados
 - Contratou estatísticos para determinar o perfil de seus clientes a partir da análise dos dados



Fonte: Charles Duhigg. O poder do hábito.
Editora Objetiva.

Exemplos de aplicação

- Pouco tempo depois, recebeu uma reclamação de um cliente, pelo fato de sua filha ter recebido cupons para compras de produtos para bebês
- Dias depois, a assessoria da *Target* ligou para o cliente para pedir desculpas
- Embaraçado, o pai não sabia o que dizer, pois descobriu que a filha estava grávida.

Exemplos de aplicação

- Recomendações para clientes:
 - Blockbuster e Netflix recomendam conteúdo aos seus clientes baseados em dados de outros perfis que associam filmes e seriados similares
 - WalMart captura transações de vendas de mais de 2900 lojas em 6 países e as transmite continuamente para seu massivo *data warehouse*.
 - Os dados obtidos são disponibilizados aos fabricantes.

Exemplos de aplicação

- Sistemas de saúde (planos)
 - Estima volumes de pacientes
 - Prepara tomada de decisão
 - Prevenção de doenças
 - Detecção de fraudes
- Exemplo:
 - Health Catalyst TM (KLAS)

Exemplos de aplicação

- Detecção de anomalias
 - Intrusão em sistemas
 - Erros
 - Fraudes
 - Análise de casos isolados
 - Como tratar?
 - Forte aplicação na área da saúde, segmentação e reconhecimento de imagens, redes e outras.

Exemplos de aplicação

- Bioinformática
 - Detecção de sequências
 - Genes
 - Análise de expressão gênica
 - Classificação simples e multi-classe
 - Diagnósticos de doenças

Exemplos de aplicação

- Mercado de ações/ativos (*Algotrading*)
 - Estudo de séries
 - Geração de modelos
 - Previsão
 - Auxílio na tomada de decisão
 - Construção do modelo



Exemplos – Análise de crédito

| I D | Proprietário | E. Civil | Renda | Investe |
|--------|--------------|------------|----------|---------|
| 1 | Sim | Solteiro | 1.500,00 | Pouco |
| 2 | Não | Casado | 812,00 | Muito |
| 3 | Não | Solteiro | 2.345,67 | Não |
| 4 | Sim | Casado | 4.768,00 | Muito |
| 5 | Não | Divorciado | 734,00 | Não |
| 6 | Não | Casado | 3.900,00 | Pouco |
| 7 | Sim | Divorciado | 2.100,00 | Muito |

- Dados de clientes, decidir se concede empréstimo ou não



Conjunto de dados

- **Objetos ou padrões:**
 - compõem o conjunto de dados
 - cada uma faz referência a uma amostra, documento, cliente, gene, sequência ...



| ID | Proprietário | E. Civil | Renda | Investe |
|----|--------------|------------|----------|---------|
| 1 | Sim | Solteiro | 1.500,00 | Sim |
| 2 | Não | Casado | 812,00 | Sim |
| 3 | Não | Solteiro | 2.345,67 | Não |
| 4 | Sim | Casado | 4.768,00 | Sim |
| 5 | Não | Divorciado | 734,00 | Não |
| 6 | Não | Casado | 3.900,00 | Sim |
| 7 | Sim | Divorciado | 2.100,00 | Sim |

Conjunto de dados

- Atributos ou características:
 - propriedade ou característica de um objeto



| ID | Proprietário | E. Civil | Renda | Investe |
|----|--------------|------------|----------|---------|
| 1 | Sim | Solteiro | 1.500,00 | Sim |
| 2 | Não | Casado | 812,00 | Sim |
| 3 | Não | Solteiro | 2.345,67 | Não |
| 4 | Sim | Casado | 4.768,00 | Sim |
| 5 | Não | Divorciado | 734,00 | Não |
| 6 | Não | Casado | 3.900,00 | Sim |
| 7 | Sim | Divorciado | 2.100,00 | Sim |

Exemplos – Análise de crédito

- O que compõe o aprendizado/decisão de ceder de crédito?
 - Entrada: x
 - Saída: y
 - Função: $f : \mathcal{X} \rightarrow \mathcal{Y}$

Exemplos – Análise de crédito

- O que compõe o aprendizado/decisão de ceder de crédito?
 - Entrada: x (aplicação / dados do cliente)
 - Saída: y (bom pagador?)
 - Função: $f : \mathcal{X} \rightarrow \mathcal{Y}$ (função de aprovação de crédito) - função objetivo

Exemplos – Análise de crédito

- O que compõe o aprendizado/decisão de ceder de crédito?
 - Entrada: x (aplicação / dados do cliente)
 - Saída: y (bom pagador?)
 - Função: $f : \mathcal{X} \rightarrow \mathcal{Y}$ (função de aprovação de crédito) - função objetivo
 - Dados exemplos:
 - $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n)$

Exemplos – Análise de crédito

- O que compõe o aprendizado/decisão de ceder de crédito?
 - Entrada: x (aplicação / dados do cliente)
 - Saída: y (bom pagador?)
 - Função: $f : \mathcal{X} \rightarrow \mathcal{Y}$ (função de aprovação de crédito) - função objetivo
 - Dados exemplos:
 - $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n)$

Hipótese: $g : \mathcal{X} \rightarrow \mathcal{Y}$

O elo faltante

Função desconhecida

$f : \rightarrow$

função de aprovação de
crédito

Exemplos de treinamento

$(x_1, y_1), \dots, (x_n, y_n)$

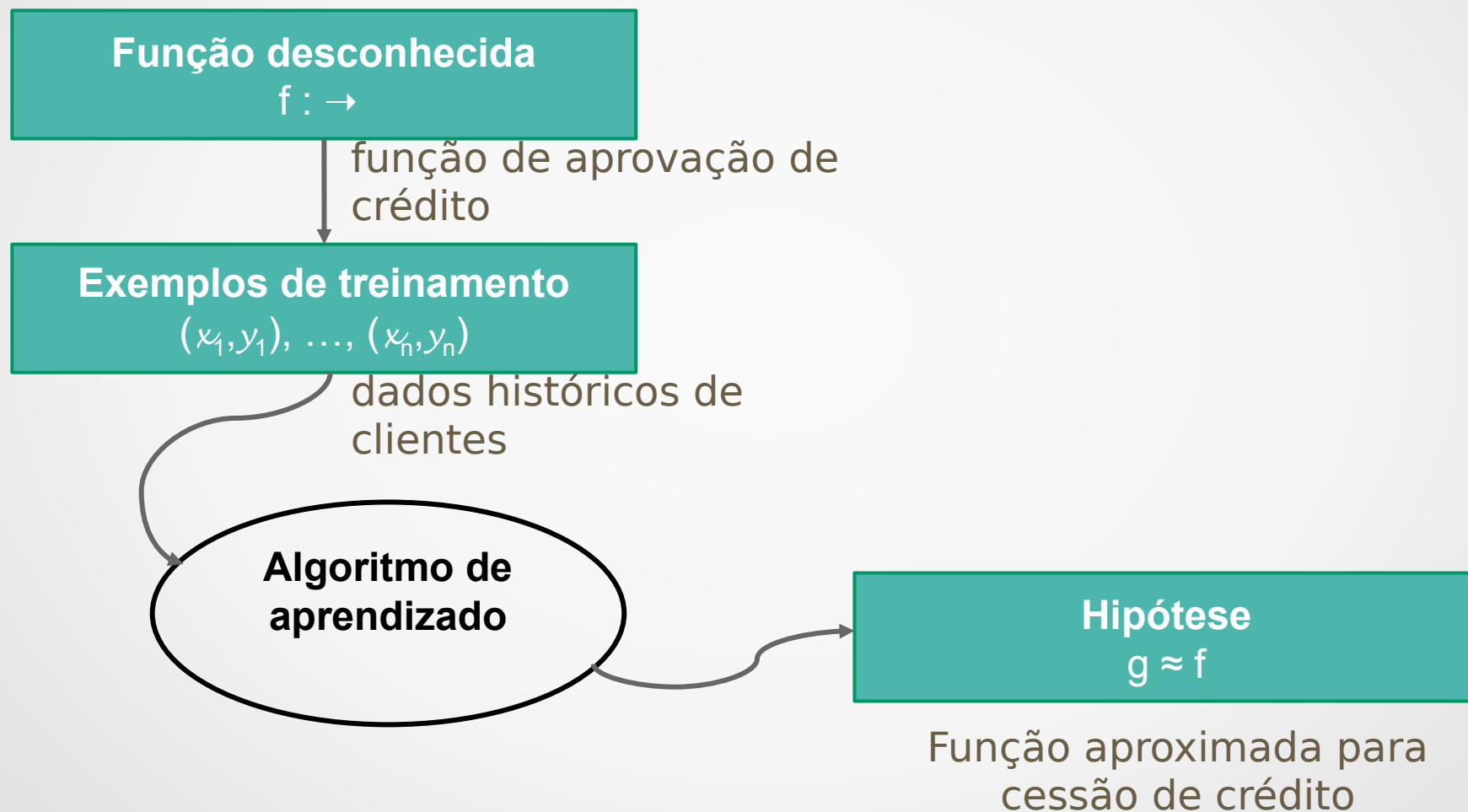
dados históricos de
clientes

Hipótese

$g \approx f$

Função aproximada para
cessão de crédito

O elo faltante



Tipos e Tarefas de AM

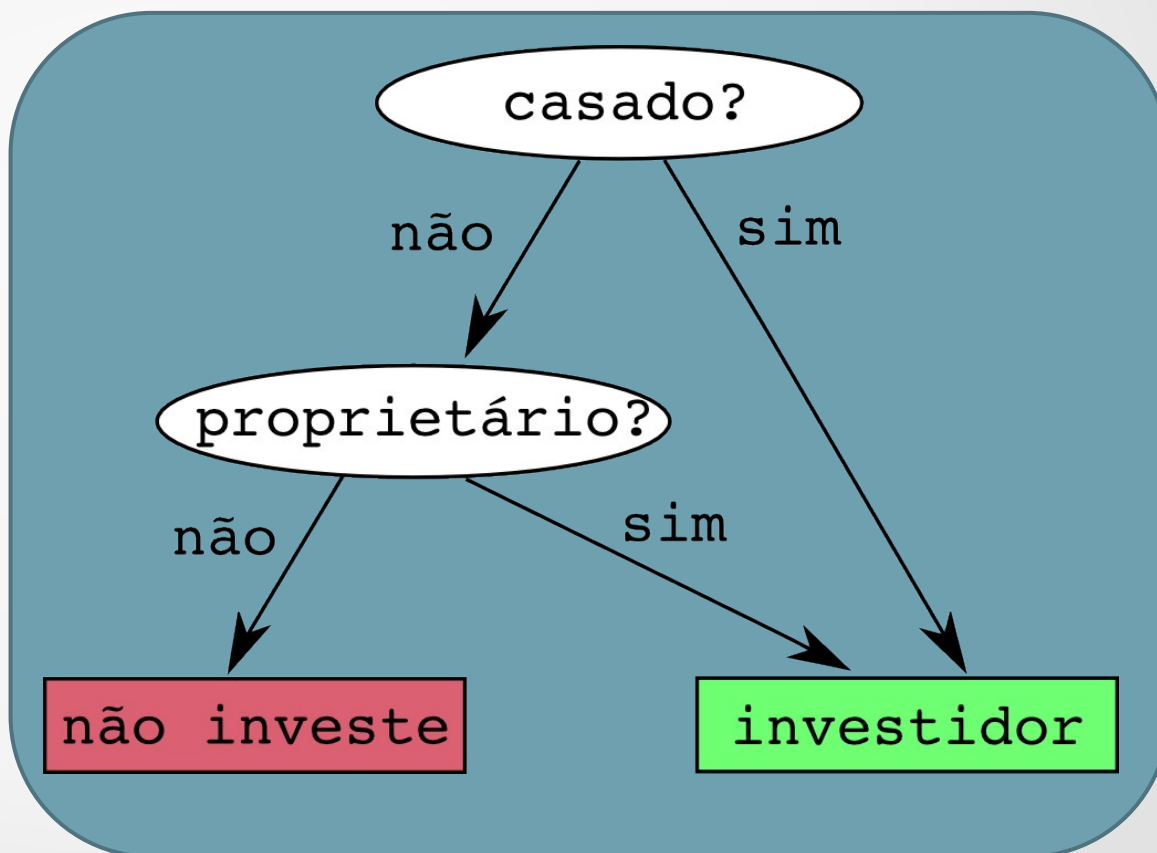
- Dependendo do objetivo, aprendizado de máquina pode ser dividido em:
 - Supervisionado
 - Não-supervisionado
 - Semi-supervisionado
- Somado à:
 - Aprendizado por reforço
 - Sistema de recomendação

Tarefas de AM

- **Tarefas de previsão:** o objetivo é desta tarefa é prever o valor de um atributo (dependente) a partir de outros (independente)
 - Associado a aprendizado supervisionado
- Exemplo:
 - Classificação
 - Regressão

Exemplo: classificação

| Índice | Proprietário | E. Civil | Renda | Investe |
|--------|--------------|------------|----------|---------|
| 1 | Sim | Solteiro | 1.500,00 | Sim |
| 2 | Não | Casado | 812,00 | Sim |
| 3 | Não | Solteiro | 2.345,67 | Não |
| 4 | Sim | Casado | 4.768,00 | Sim |
| 5 | Não | Divorciado | 734,00 | Não |
| 6 | Não | Casado | 3.900,00 | Sim |
| 7 | Sim | Divorciado | 2.100,00 | Sim |

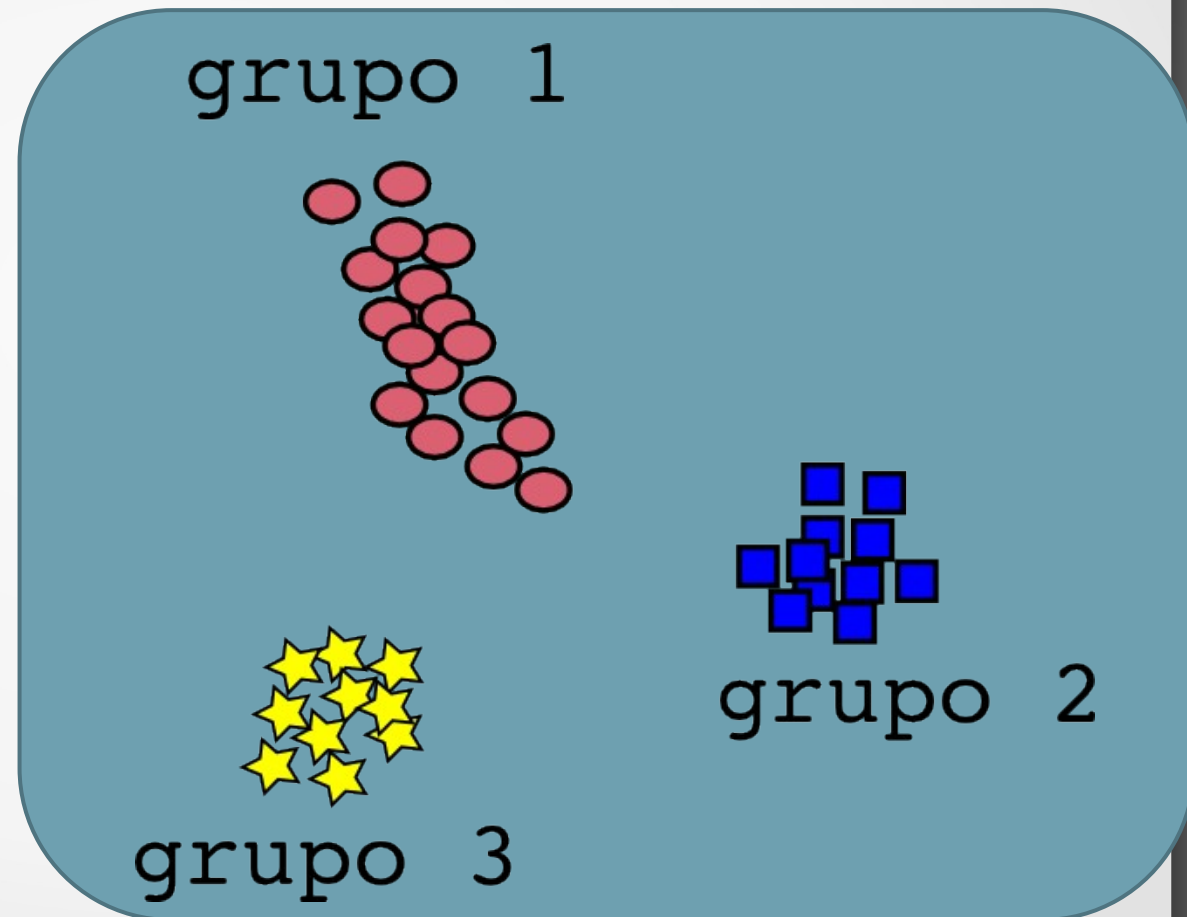


Tarefas de AM

- **Tarefas descritivas:** o objetivo é derivar padrões (correlações, tendências, grupos, trajetórias e anomalias) que resumam relacionamentos subjacentes nos dados
 - Associado a aprendizado não supervisionado
- Exemplo:
 - Agrupamento
 - Análise de associação
 - Detecção de anomalias

Exemplo: agrupamento

| Índice | Proprietário | E. Civil | Renda | Investe |
|--------|--------------|------------|----------|---------|
| 1 | Sim | Solteiro | 1.500,00 | Sim |
| 2 | Não | Casado | 812,00 | Sim |
| 3 | Não | Solteiro | 2.345,67 | Não |
| 4 | Sim | Casado | 4.768,00 | Sim |
| 5 | Não | Divorciado | 734,00 | Não |
| 6 | Não | Casado | 3.900,00 | Sim |
| 7 | Sim | Divorciado | 2.100,00 | Sim |



Exemplo: análise de associação

| Índice | Proprietário | E. Civil | Renda | Investe |
|--------|--------------|------------|----------|---------|
| 1 | Sim | Solteiro | 1.500,00 | Sim |
| 2 | Não | Casado | 812,00 | Sim |
| 3 | Não | Solteiro | 2.345,67 | Não |
| 4 | Sim | Casado | 4.768,00 | Sim |
| 5 | Não | Divorciado | 734,00 | Não |
| 6 | Não | Casado | 3.900,00 | Sim |
| 7 | Sim | Divorciado | 2.100,00 | Sim |

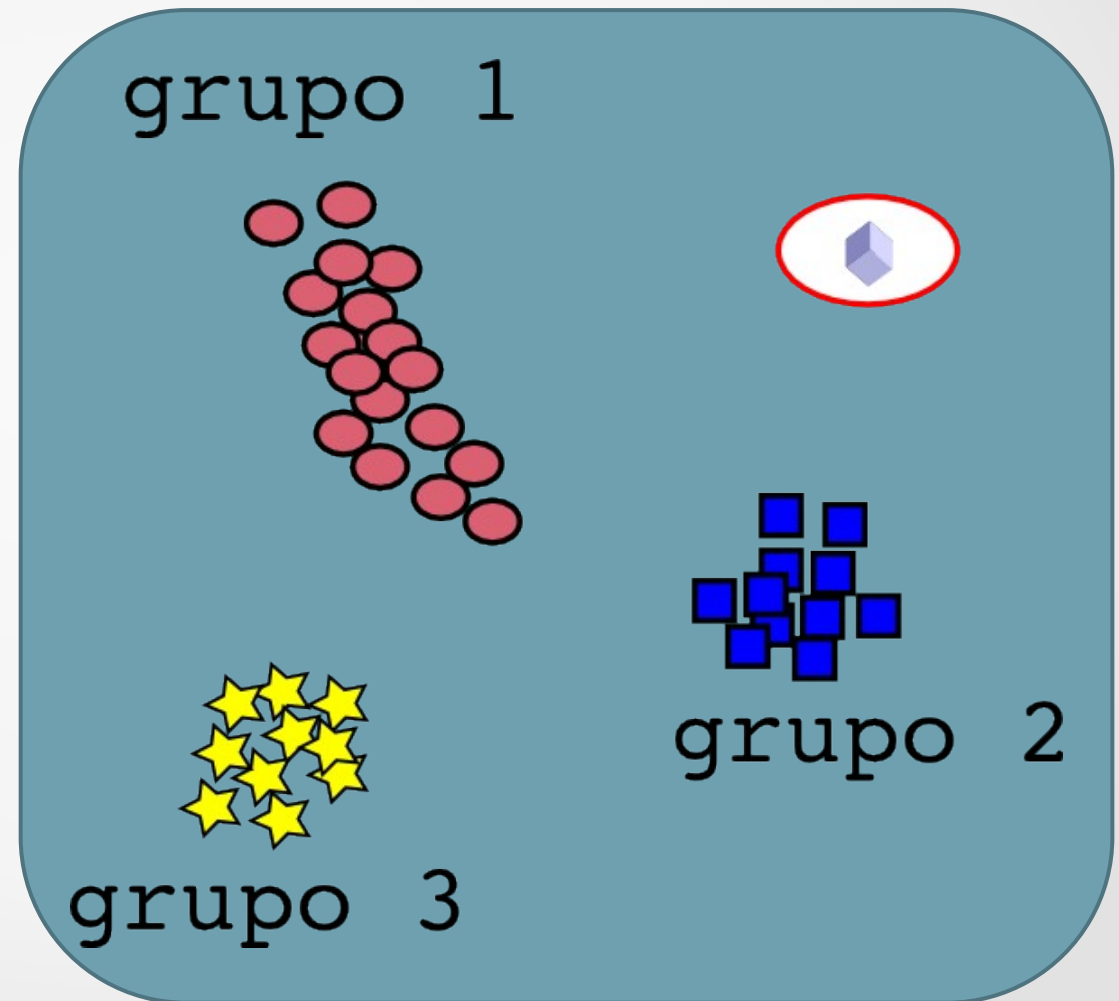


Mito: caso da relação
cerveja e fraldas

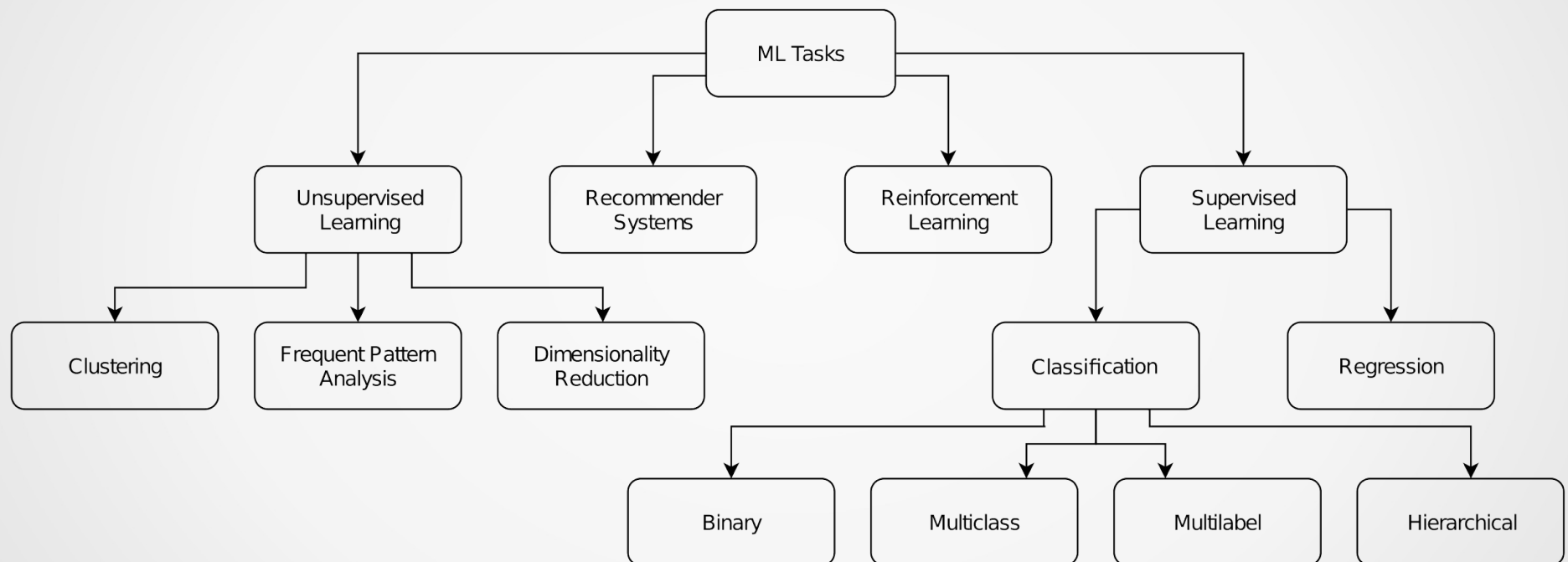


Exemplo: detecção de anomalias

| Id | Proprietário | E. Civil | Renda | Investe |
|----|--------------|------------|----------|---------|
| 1 | Sim | Solteiro | 1.500,00 | Sim |
| 2 | Não | Casado | 812,00 | Sim |
| 3 | Não | Solteiro | 2.345,67 | Não |
| 4 | Sim | Casado | 4.768,00 | Sim |
| 5 | Não | Divorciado | 734,00 | Não |
| 6 | Não | Casado | 3.900,00 | Sim |
| 7 | Sim | Divorciado | 2.100,00 | Sim |



Organização



Premissa do aprendizado

- Simplificando a premissa do aprendizado levantada por Tom Mitchell, aprendizado de máquina é a estimação de uma função de acordo com exemplos.
 - Muito amplo! Aberto a muitas possibilidades!

O que não é aprendizado de máquina?

- Inteligência artificial vs AM
- Mineração de dados vs AM
- Ciência de dados vs AM
- *Big data* vs AM
- *Deep learning* vs AM