



INTELIGÊNCIA COMPUTACIONAL

PROF. JOSENALDE OLIVEIRA

josenalde.oliveira@ufrn.br

<https://github.com/josenalde/ic>

ANÁLISE E DESENVOLVIMENTO DE SISTEMAS - UFRN

UM POUCO SOBRE **APRENDIZAGEM DE MÁQUINA**

Machine Learning - ML



Ser humano estabelece **conexões** para lidar com coisas novas

Similaridade pode ser óbvio para o humano, mas não para computadores

Máquinas operam sobre **tarefas frequentes**, com alto volume e velocidade

Desafio: máquinas serem ensinadas e depois aprenderem 'sozinhas'

UM POUCO SOBRE **APRENDIZAGEM DE MÁQUINA**

Machine Learning - ML

Como uma criança aprende que ambos são dinossauros?



E aqui?



APRENDIZAGEM DE MÁQUINA

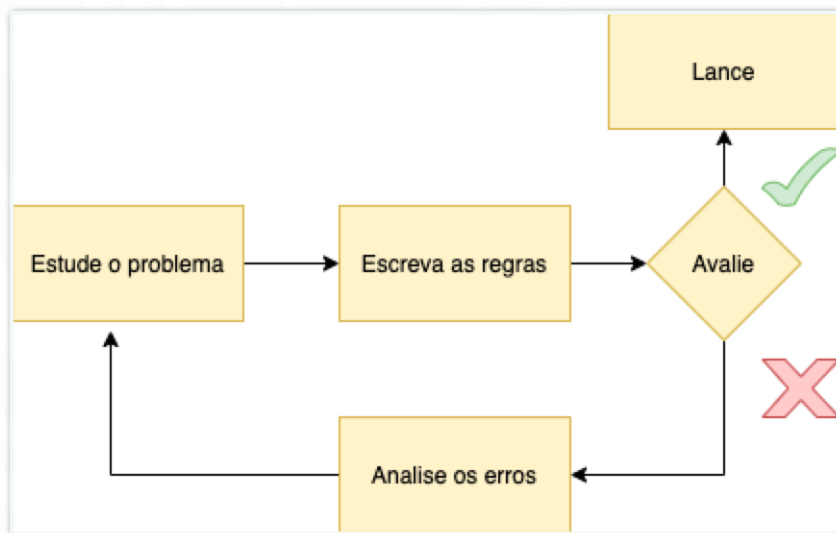


<https://www.serpro.gov.br/menu/noticias/noticias-2019/democratizando-a-inteligencia-artificial>

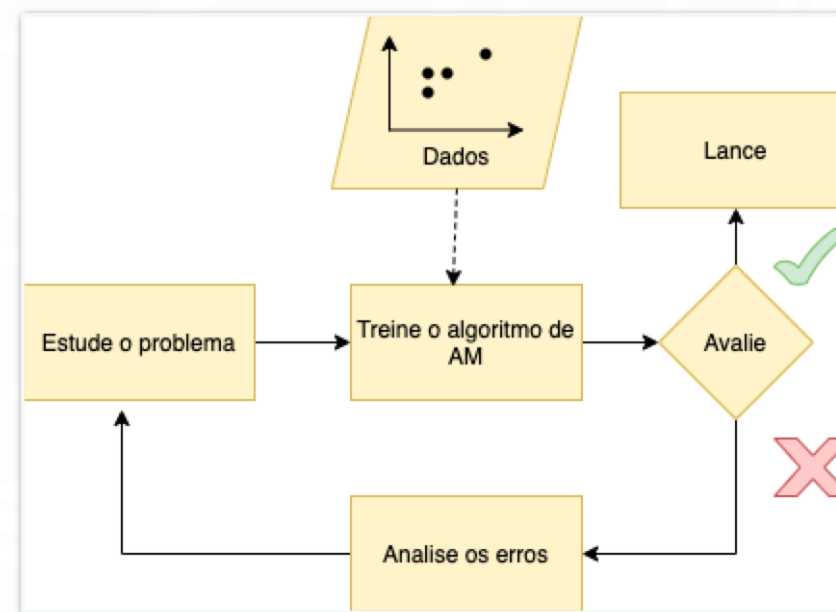
APRENDIZAGEM DE MÁQUINA

Área da Inteligência Artificial que investiga **o desenvolvimento de algoritmos que são capazes de aprender a partir dos dados**, adquirindo conhecimento de forma automática

Sem AM



Com AM

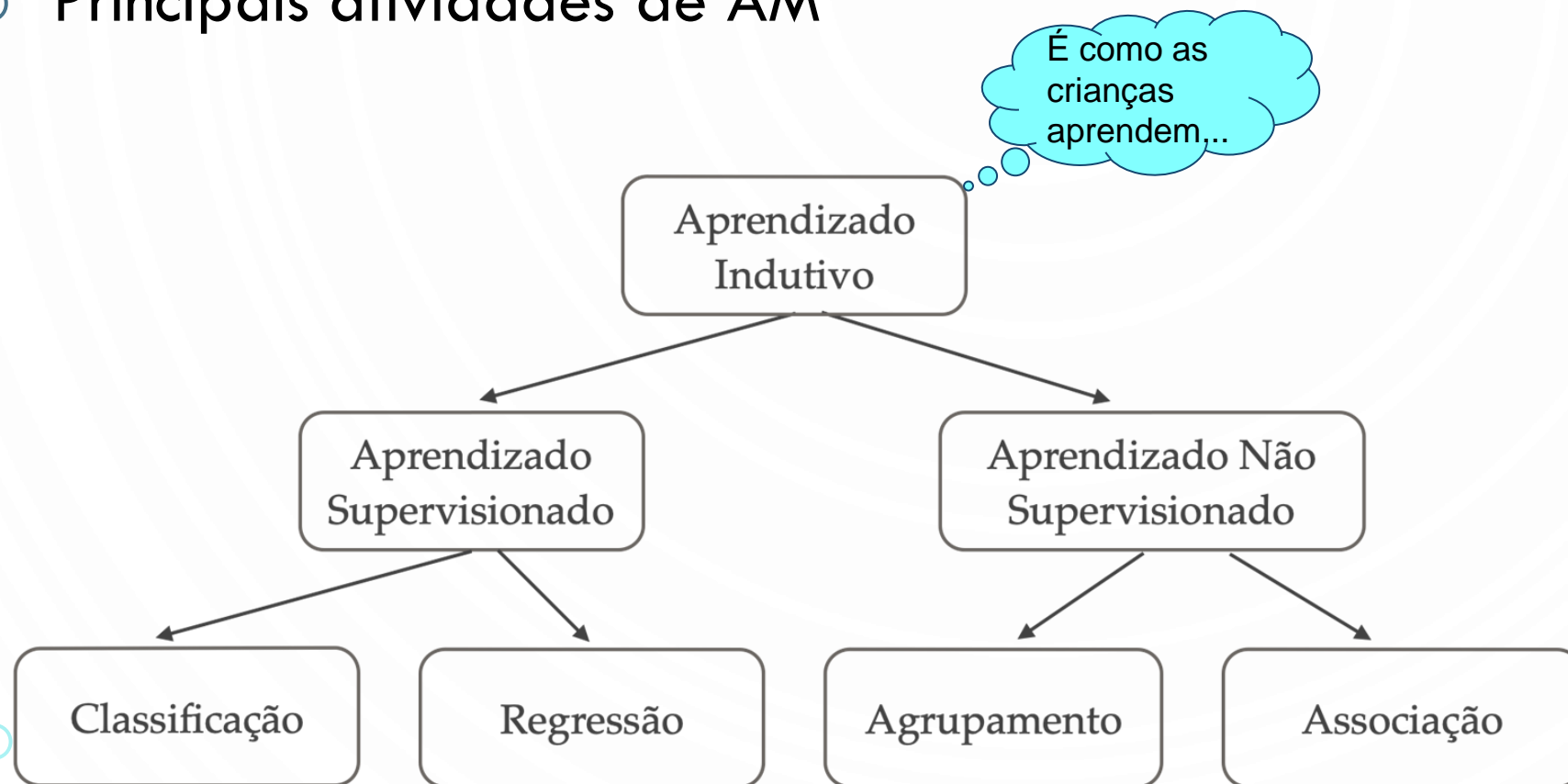


APRENDIZAGEM DE MÁQUINA



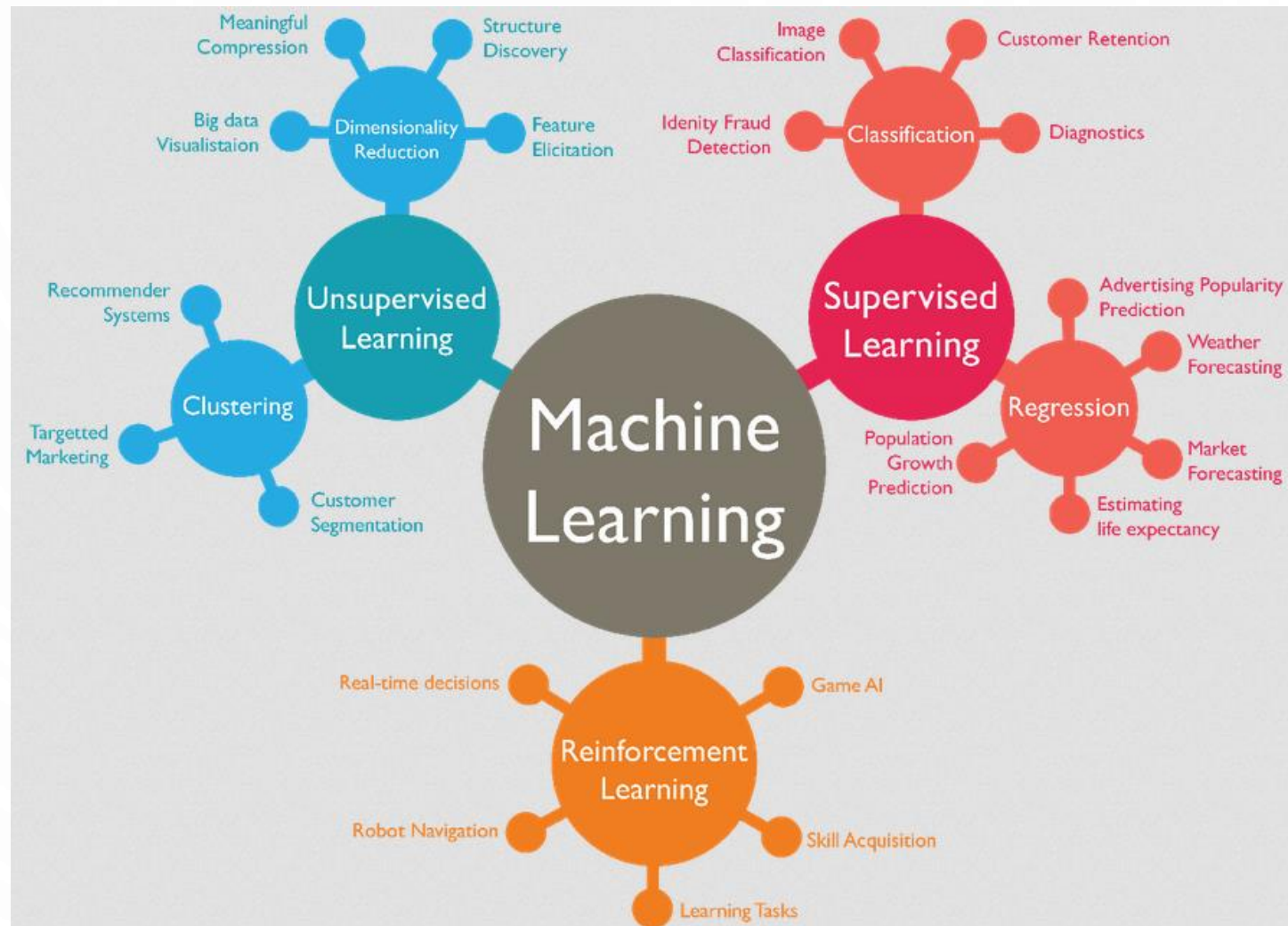
APRENDIZAGEM DE MÁQUINA

• Principais atividades de AM



APRENDIZAGEM DE MÁQUINA

- Principais atividades de AM



DATASETS

- Bases de dados para algoritmos de aprendizado de máquina são formadas por amostras do domínio que se deseja aprender
- Variáveis numéricas, categóricas (uni, bi, poli)
- Variáveis categóricas normalmente codificadas para AM (encoder)

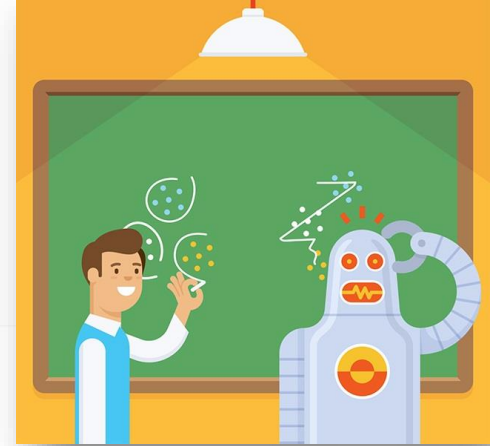
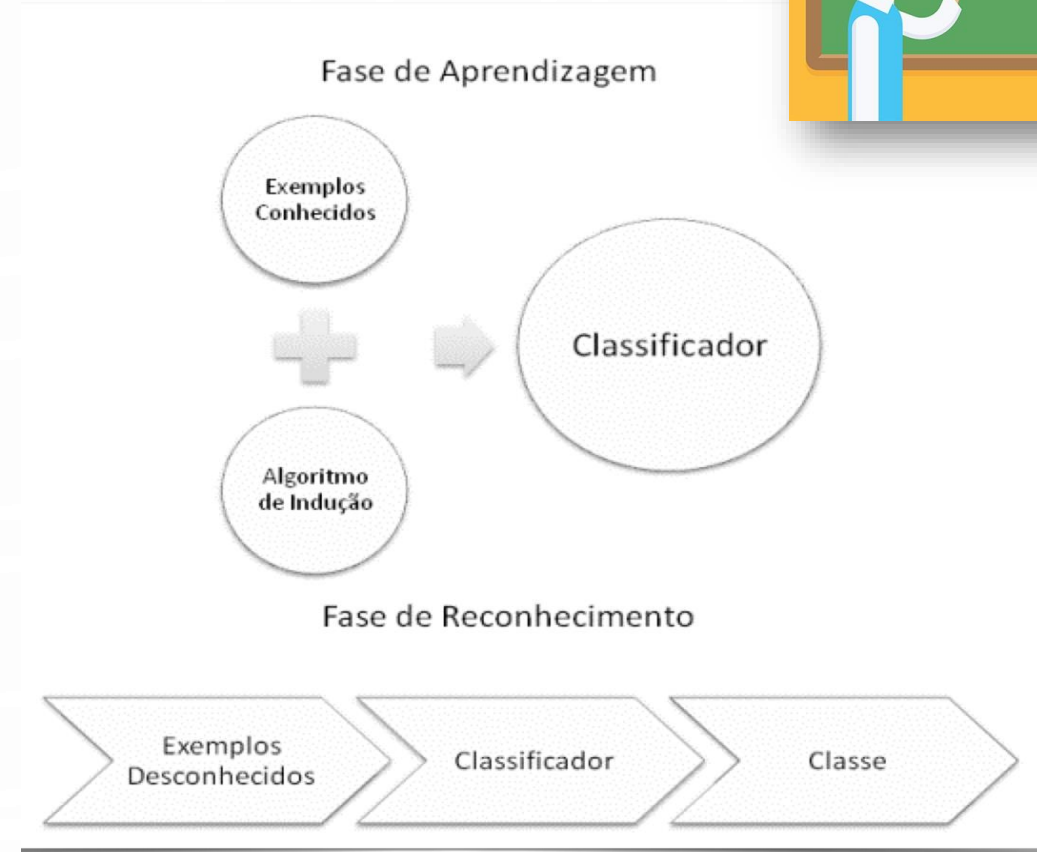
| Id. | Nome | Idade | Sexo | Peso | Manchas | Temp. | # Int. | Est. | Diagnóstico |
|------|---------|-------|------|------|--------------|-------|--------|------|-------------|
| 4201 | João | 28 | M | 79 | Concentradas | 38,0 | 2 | SP | Doente |
| 3217 | Maria | 18 | F | 67 | Inexistentes | 39,5 | 4 | MG | Doente |
| 4039 | Luiz | 49 | M | 92 | Espalhadas | 38,0 | 2 | RS | Saudável |
| 1920 | José | 18 | M | 43 | Inexistentes | 38,5 | 8 | MG | Doente |
| 4340 | Cláudia | 21 | F | 52 | Uniformes | 37,6 | 1 | PE | Saudável |
| 2301 | Ana | 22 | F | 72 | Inexistentes | 38,0 | 3 | RJ | Doente |
| 1322 | Marta | 19 | F | 87 | Espalhadas | 39,0 | 6 | AM | Doente |
| 3027 | Paulo | 34 | M | 67 | Uniformes | 38,4 | 2 | GO | Saudável |

DATASETS

- As amostras são comumente chamadas de exemplos ou padrões
- Em geral, eles são representados por um vetor de características que os descreve, também denominados de atributos, campos, features (ou variáveis de decisão)
 - Cada amostra corresponde a uma ocorrência/observação/registro na base de dados
 - Cada atributo está associado a uma propriedade da amostra
 - Quais atributos são relevantes/significativos? (significância estatística, multicolinearidade, podem ser agrupados/reduzidos? (pca)?
 - Feature selection

CLASSIFICAÇÃO DE PADRÕES

- Classificação de padrões é o **processo de atribuição de rótulos discretos, também chamados de classes, a amostras de um domínio**
- A classificação de padrões dá-se em duas fases distintas: fase de aprendizagem e fase de reconhecimento



CLASSIFICAÇÃO DE PADRÕES



Orquídea



Orquídea



Orquídea



Orquídea



Orquídea



Orquídea



Flor do Deserto



Flor do Deserto



Flor do Deserto



Flor do Deserto



Flor do Deserto



Flor do Deserto

CLASSIFICAÇÃO DE PADRÕES

Orquídea ou Flor do deserto?



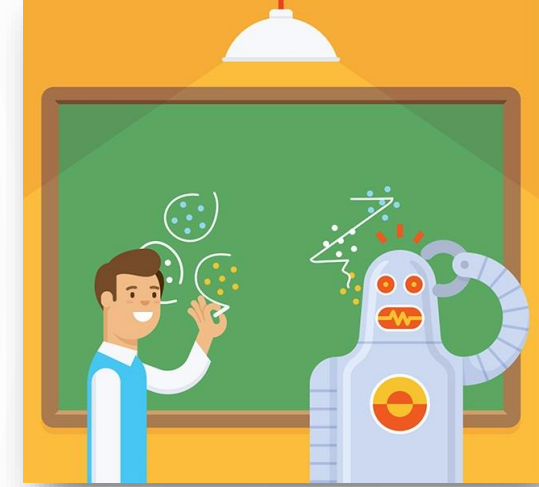
CLASSIFICAÇÃO DE PADRÕES

Orquídea ou Flor do deserto?

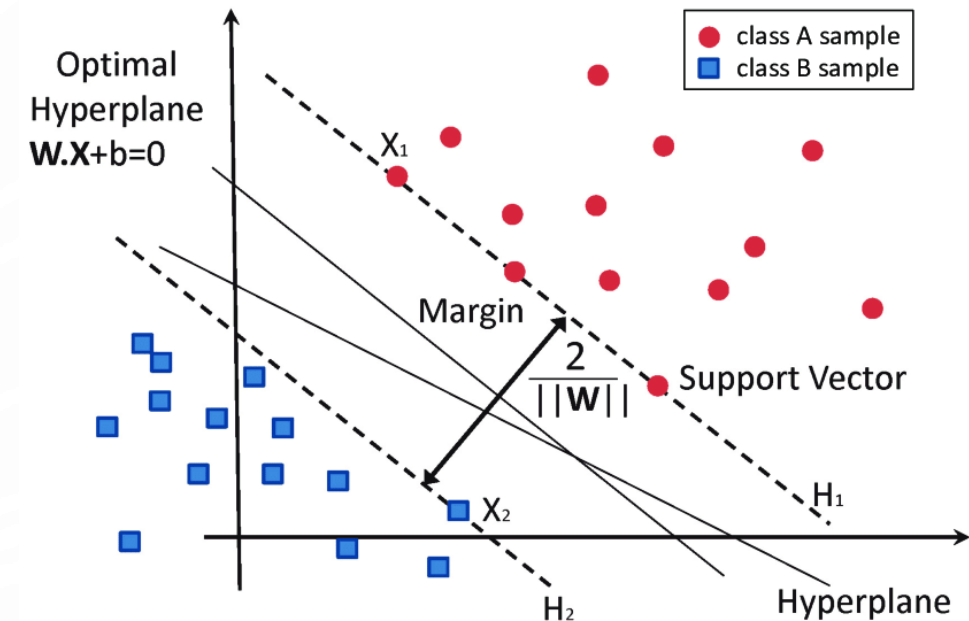
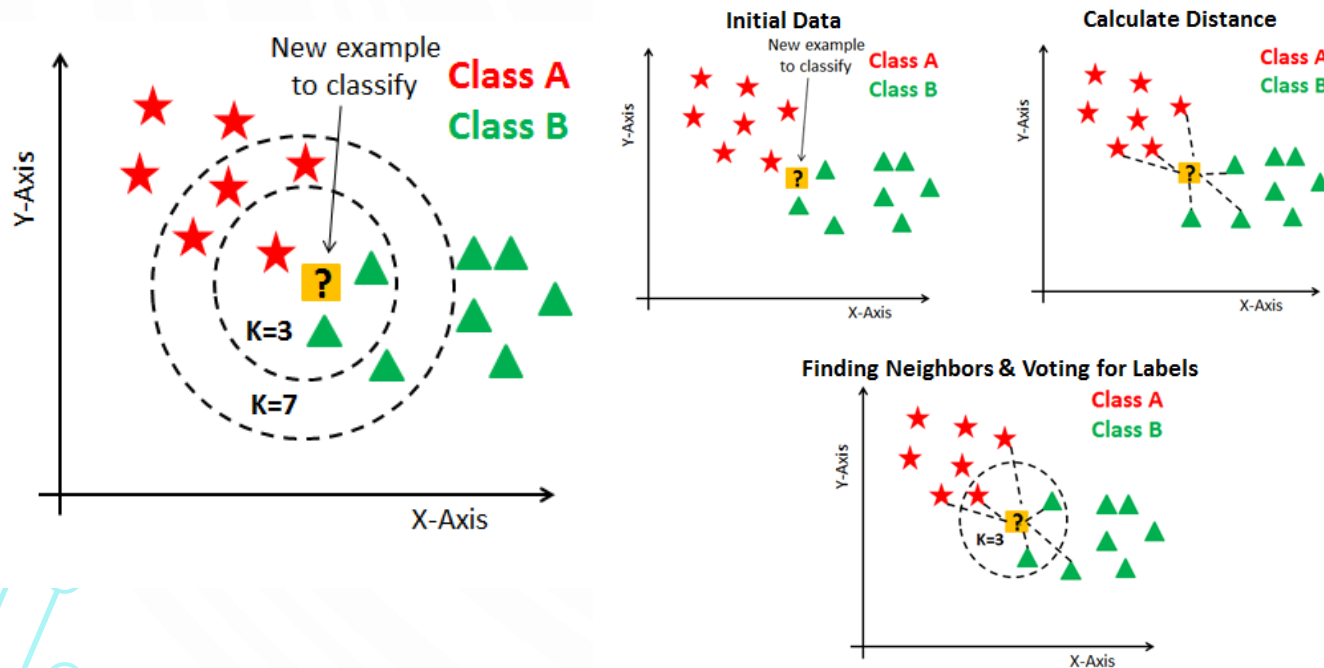


ASPECTOS DE MACHINE LEARNING

- **Classificação: supervisionado** – conjuntos de treino / teste, métricas de avaliação mais bem definidas, por comparar com o *ground truth*, matriz de confusão etc.



Métodos mais comuns: KNN, SVM, Árvore de Decisão, Redes Neurais



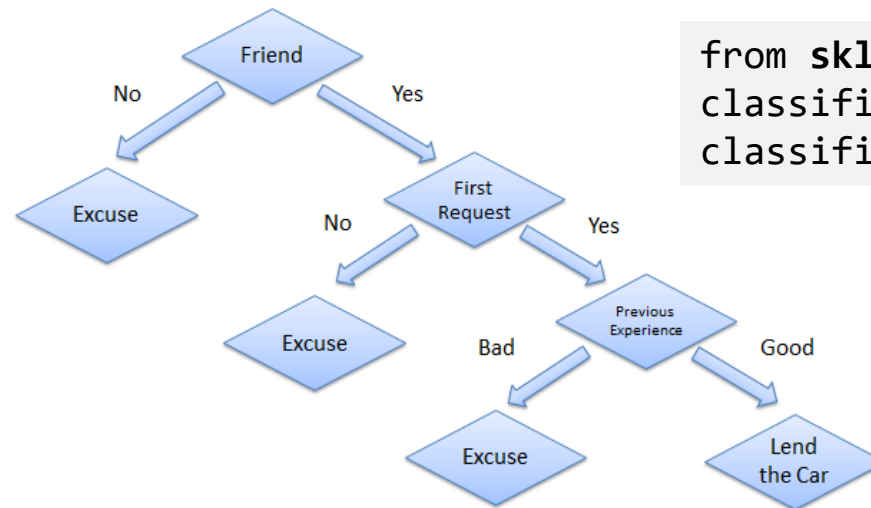
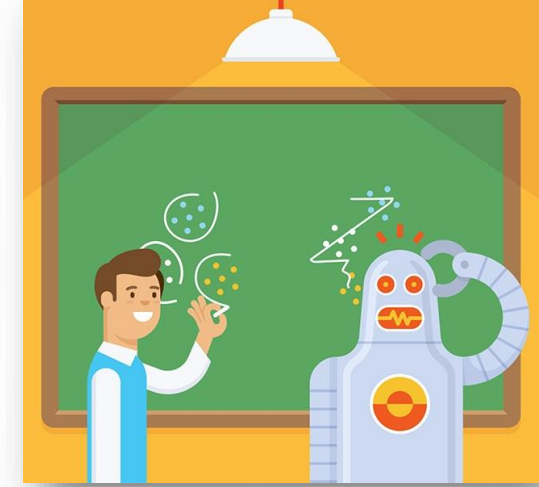
https://github.com/josenalde/datascience/blob/main/notebooks/nb_knn1.ipynb

https://github.com/josenalde/datascience/blob/main/notebooks/nb_svm1.ipynb

ASPECTOS DE MACHINE LEARNING

- **Classificação: supervisionado** – conjuntos de treino / teste, métricas de avaliação mais bem definidas, por comparar com o *ground truth*, matriz de confusão etc.

Métodos mais comuns: KNN, SVM, Árvore de Decisão, Redes Neurais



```
from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier()
classifier.fit(X_train, y_train)
```

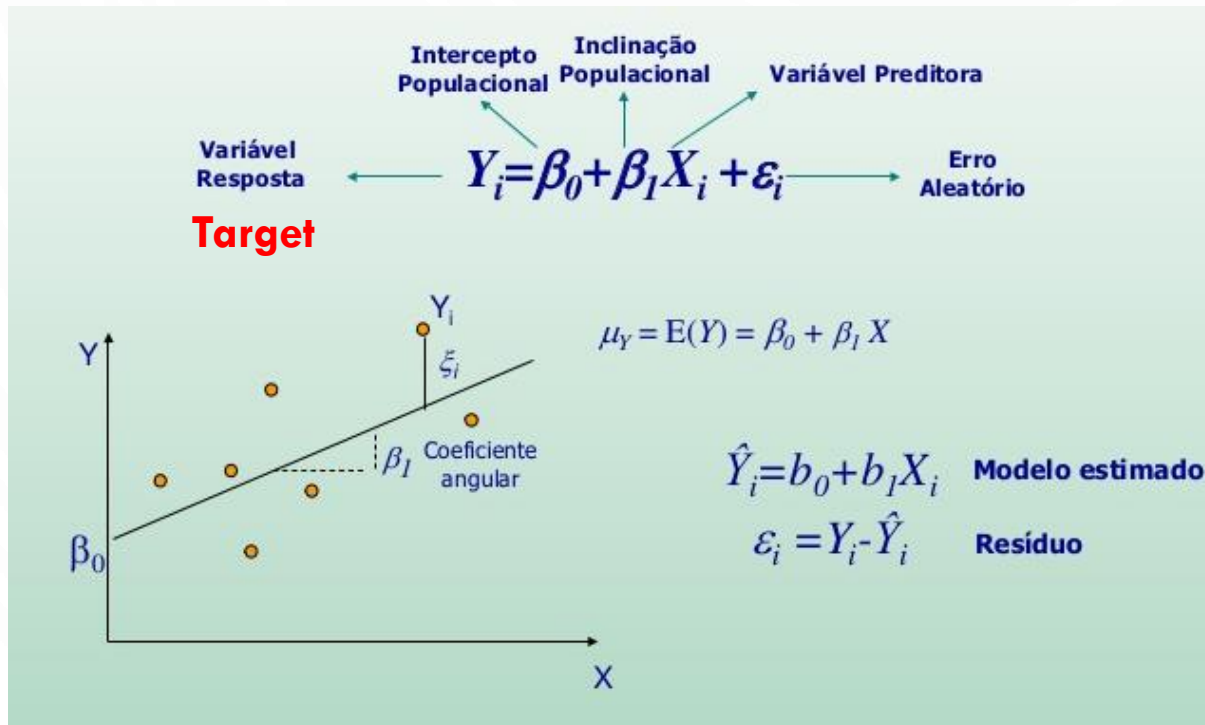
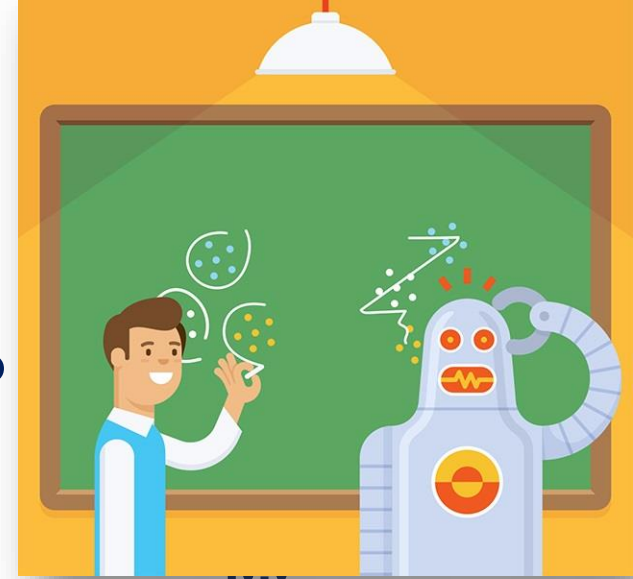
<https://www.datacamp.com/community/tutorials/decision-tree-classification-python>

REGRESSÃO

- A tarefa de regressão, também chamada de predição ou estimação, ocorre quando o atributo alvo da base de dados que se deseja aprender possui valor contínuo, como o preço de uma casa ou o lucro de uma empresa
- Nesse caso, o algoritmo de aprendizado de máquina deve encontrar um modelo matemático (equação) capaz de mapear as entradas numa saída esperada

ASPECTOS DE MACHINE LEARNING

- Um bom projeto de ML tem boa capacidade preditiva
- Acurácia nas decisões (acertos)
- Performance preditiva! Nem sempre a interpretação do processo é simples (pois decisões envolvem processos complexos)
- Em Inferência (regressão por exemplo), a relação entre as variáveis é melhor interpretável, mas usualmente pior performance preditiva



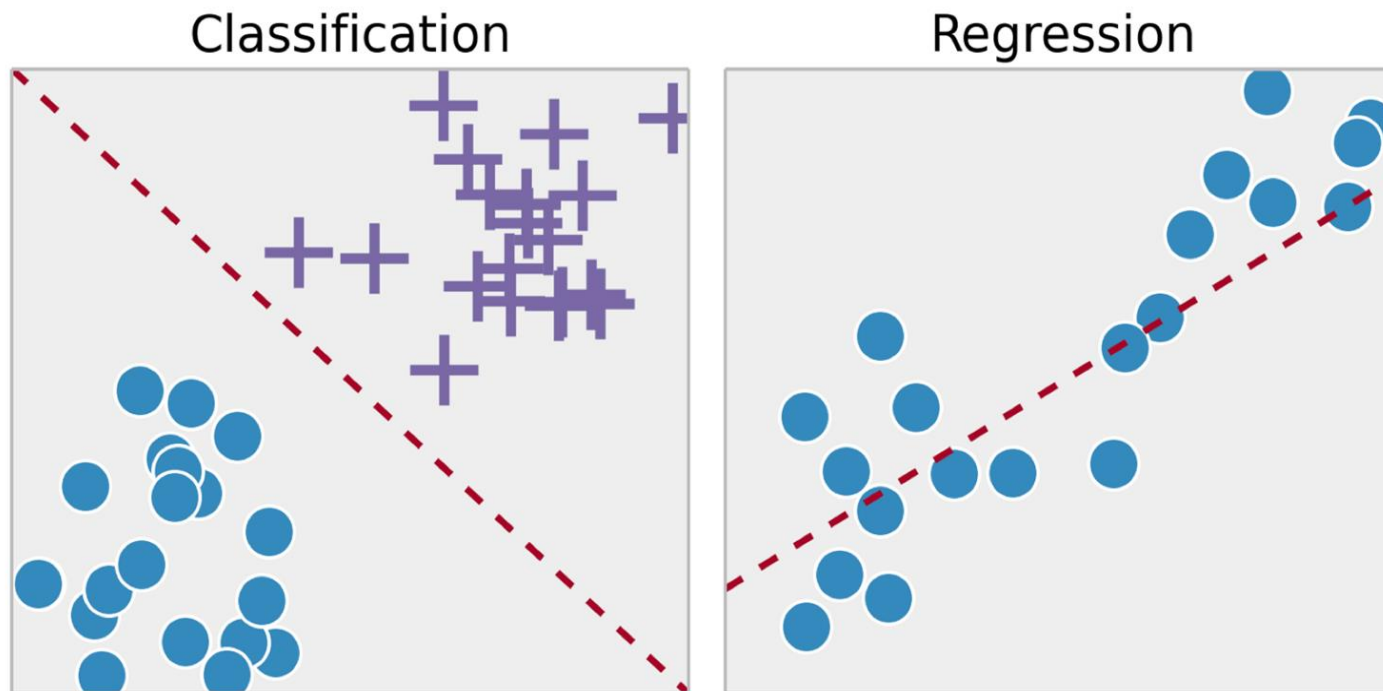
Regressão/Inferência: target quantitativo
Classificação: target qualitativo
Regressão logística: target quantitativo é probabilidade de um evento ocorrer como função de outros fatores.

REGRESSÃO

| Bedrooms | Sq. feet | Neighborhood | Sale price |
|----------|----------|--------------|------------|
| 3 | 2000 | Normaltown | \$250,000 |
| 2 | 800 | Hipsterton | \$300,000 |
| 2 | 850 | Normaltown | \$150,000 |
| 1 | 550 | Normaltown | \$78,000 |
| 4 | 2000 | Skid Row | \$150,000 |

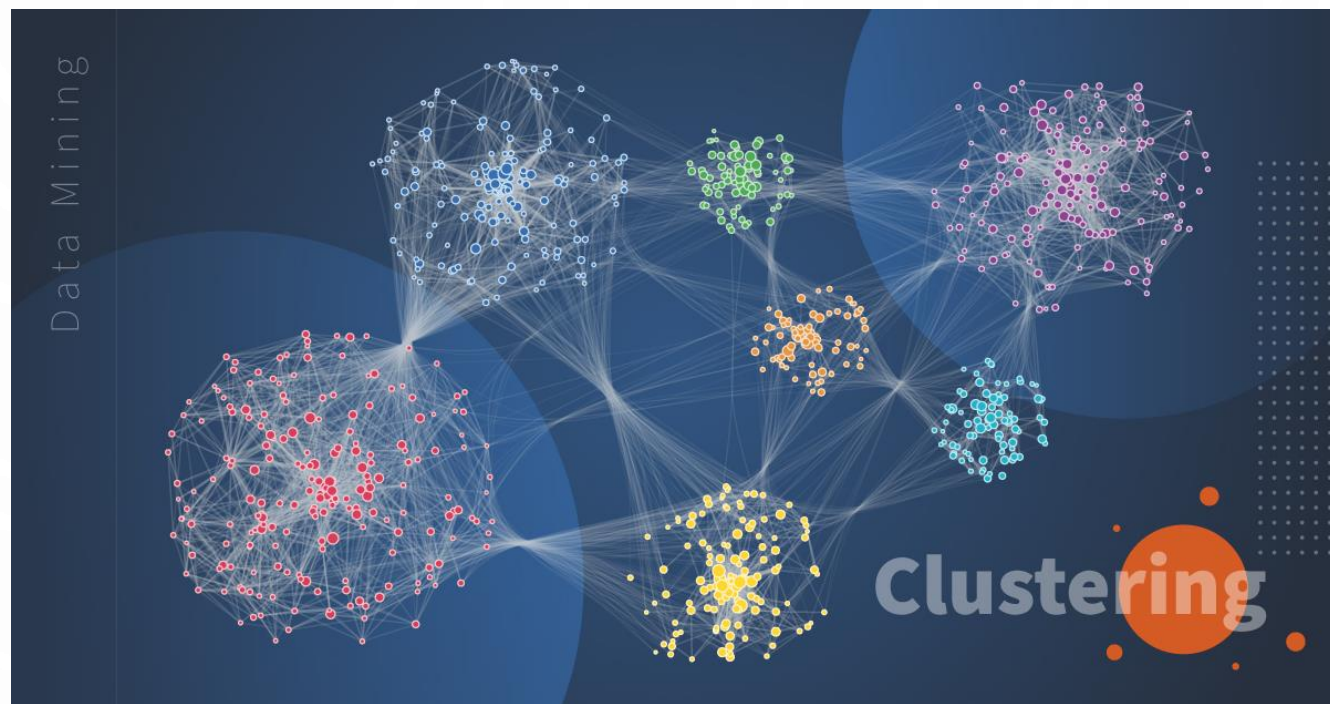
| Bedrooms | Sq. feet | Neighborhood | Sale price |
|----------|----------|--------------|------------|
| 3 | 2000 | Hipsterton | ??? |

REGRESSÃO E CLASSIFICAÇÃO



AGRUPAMENTO

- A tarefa de agrupamento busca reunir os exemplos por similaridade, criando grupos que serão posteriormente rotulados pelo cientista de dados
- No exemplo a seguir espera-se que o algoritmo consiga criar N grupos de notícias por similaridade de conteúdo

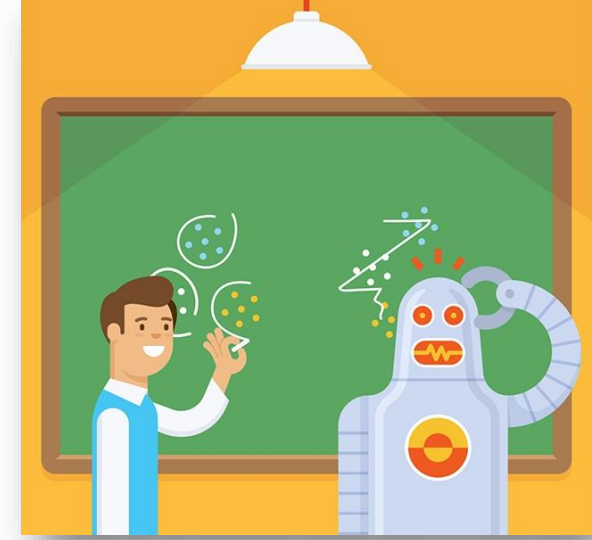
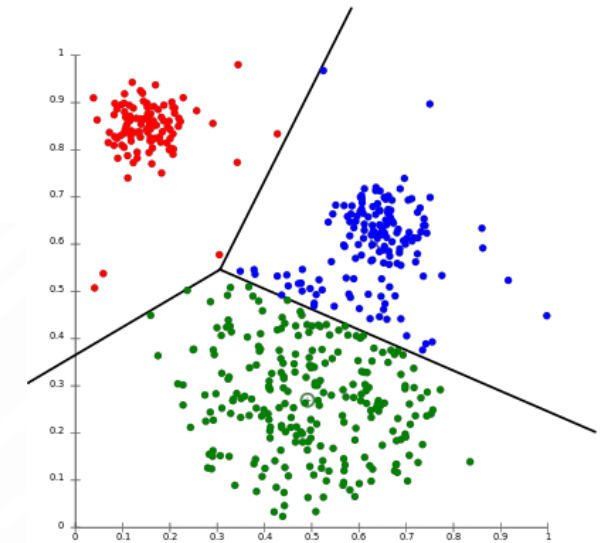
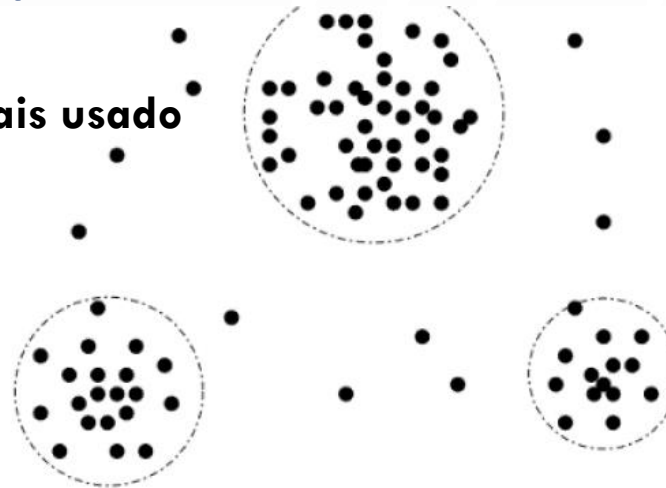
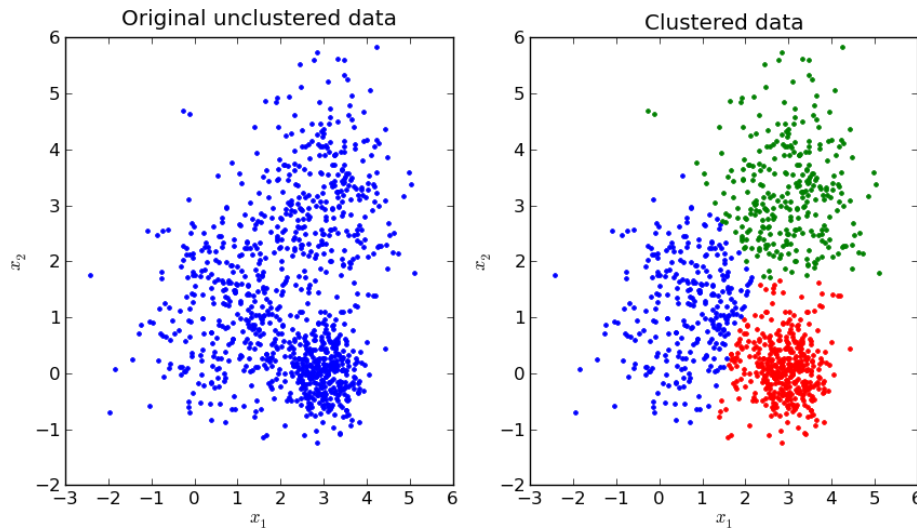


ASPECTOS DE MACHINE LEARNING

- **Agrupamentos: não supervisionado – maximizar semelhanças (minimizar distâncias) dentro do cluster e maximizar diferenças (maximizar distâncias) entre clusters**

Método K-means (incluindo Fuzzy c-means) é o mais usado

**Mas existem vários outros métodos e variantes:
Hierarquicos, aglomerativos, incremental etc.**



https://github.com/josenalde/datascience/blob/main/notebooks/nb_kmeans1.ipynb

ASSOCIAÇÃO

transações

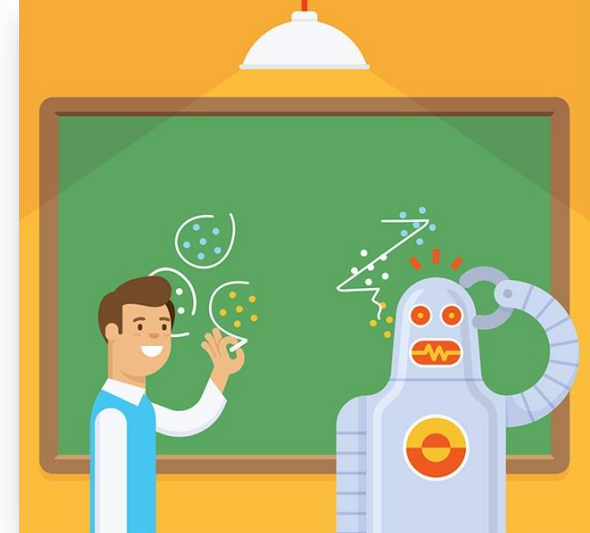
| TID | Items |
|-----|-------------------------------|
| 1 | pão, leite |
| 2 | pão, fralda, cerveja, ovos |
| 3 | leite, fraldas, cerveja, coca |
| 4 | pão, leite, fraldas, cerveja |
| 5 | pão, leite, fraldas, coca |

Exemplos de regras de associação

$\{\text{fraldas}\} \rightarrow \{\text{cerveja}\},$
 $\{\text{leite, pão}\} \rightarrow \{\text{ovos, coca}\},$
 $\{\text{cerveja, pão}\} \rightarrow \{\text{leite}\},$

Implicação significa co-ocorrência, e não causa!!!

- Também chamado de mineração de regras de associação
- Essa tarefa busca identificar padrões nos dados analisados, como ocorrências de valores juntos em um mesmo exemplo da base
- No exemplo a seguir, espera-se que o algoritmo encontre os itens que normalmente são adquiridos juntos



EXEMPLOS SUPERVISIONADOS

- Estimar o preço de uma casa
 - Atributos: tamanho, posição geográfica, material
 - Classe: Preço (**regressão**)
- Determinar se uma pessoa tem câncer benigno ou maligno
 - Atributos: Tamanho do tumor, formato do tumor, idade do paciente
 - Classe: Benigno ou maligno (**classificação**)

EXEMPLOS NÃO SUPERVISIONADOS

- Identificar padrões de compras dos clientes de um supermercado (**Associação**)
- Identificar padrões de navegação em sites (**Associação**)
- Agrupar notícias semelhantes publicadas em várias fontes (**Agrupamento**)
- Numa rede social, identificar subgrupos de pessoas (**Agrupamento**)
- Identificar aves macho e fêmea em lotes mistos de frangos de corte em aviários, a partir de medidas de peso coletados por balança automática (**Agrupamento**)