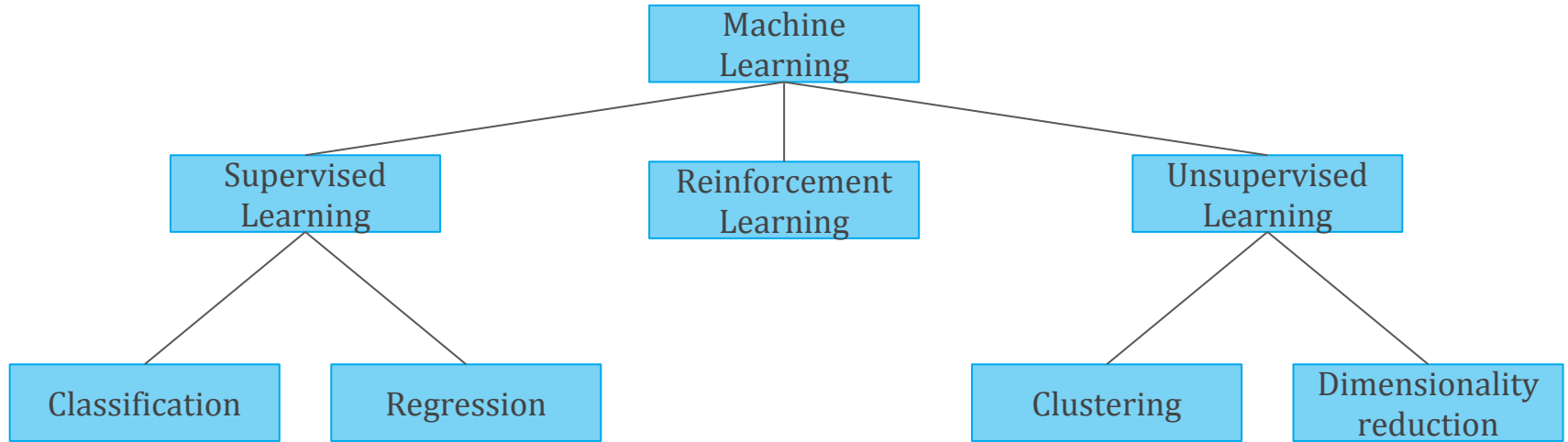


data

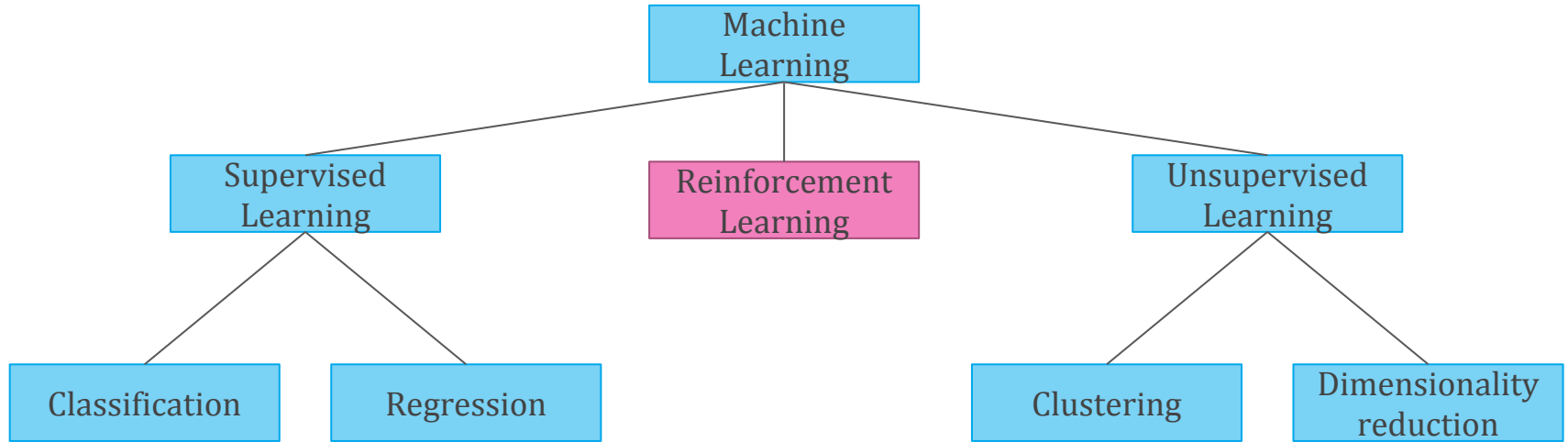
Introdução ao aprendizado de máquina

Gustavo Sutter
@suttergustavo

Modelos de machine learning



Modelos de machine learning

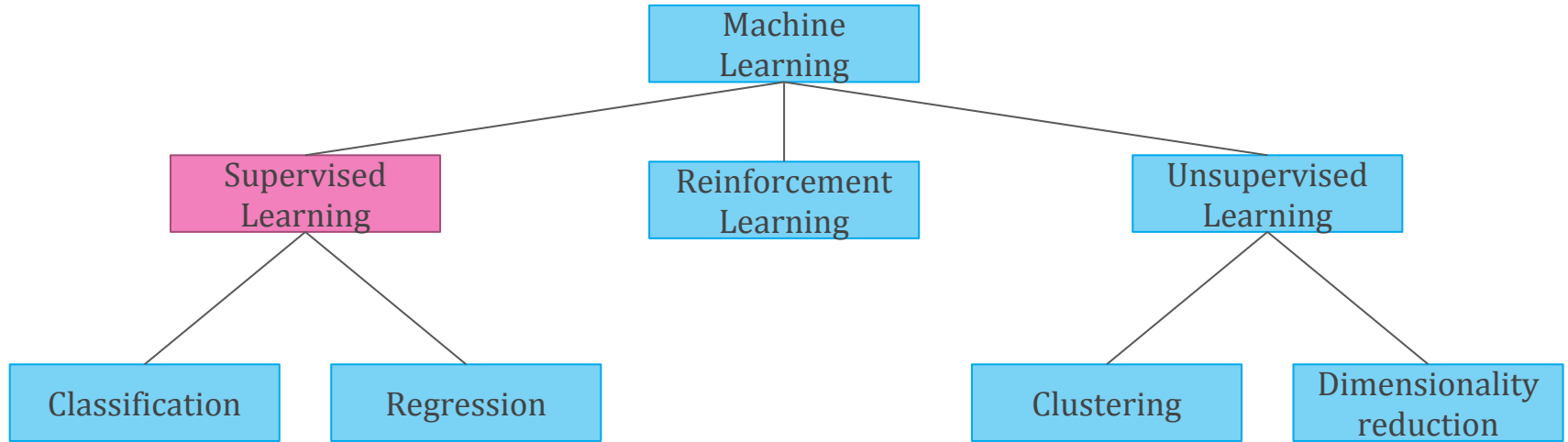


Reinforcement learning

- Agentes que interagem com um ambiente realizando uma tarefa
- Bem diferente dos outros paradigmas
- Muito em alta ultimamente:
 - AlphaGo e AlphaZero (Deepmind)
 - OpenAI Five (OpenAI)



Modelos de machine learning

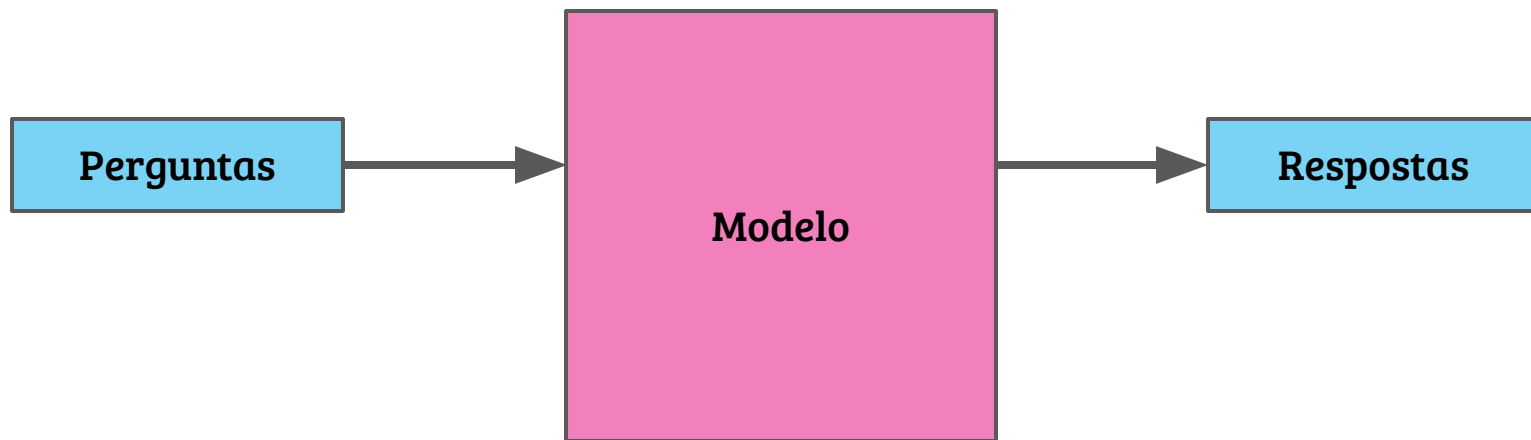


Aprendizado supervisionado

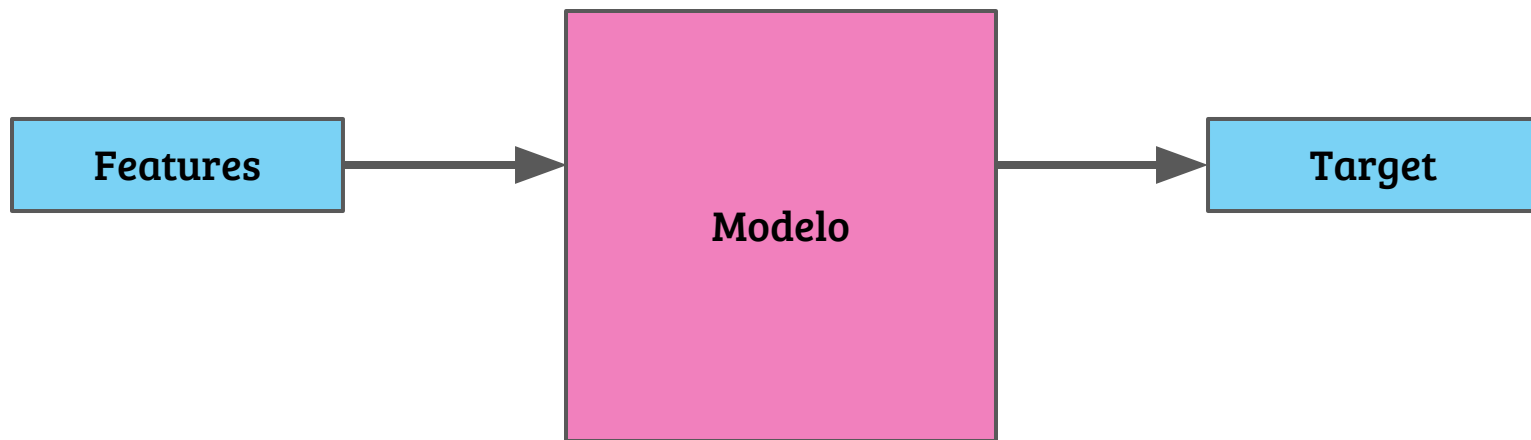
- Aprendizado supervisionado é aquele que **se conhece a resposta esperada**, tendo como objetivo aprender a mapear os atributos para essa resposta
- Essa resposta esperada é comumente chamada de *target*



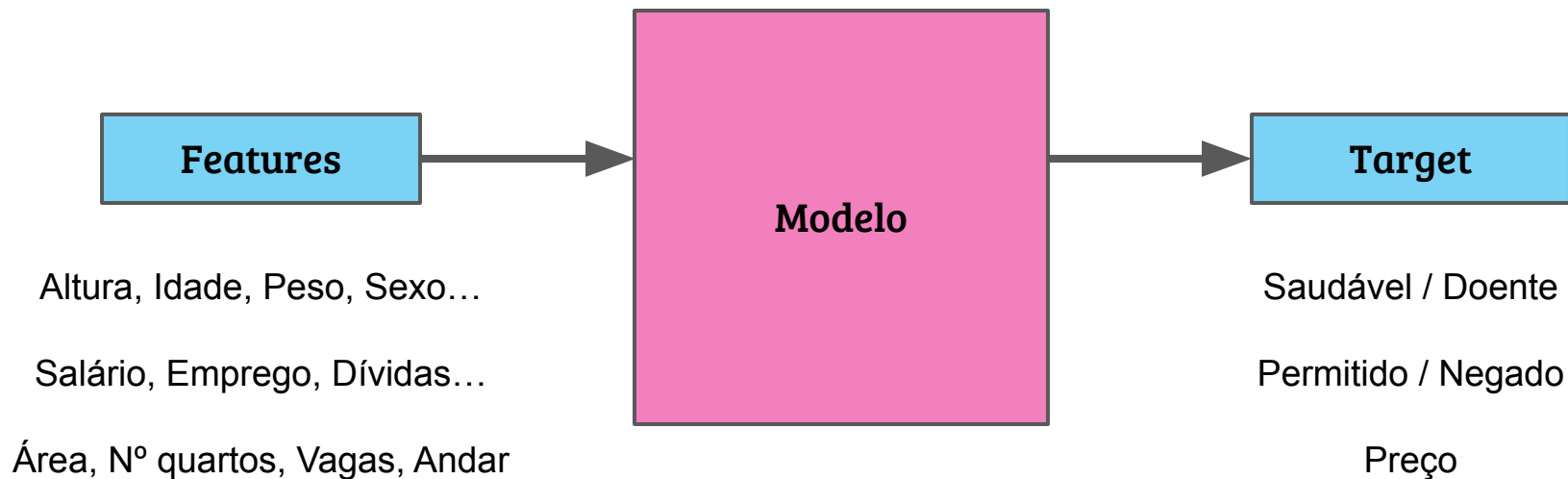
Aprendizado supervisionado



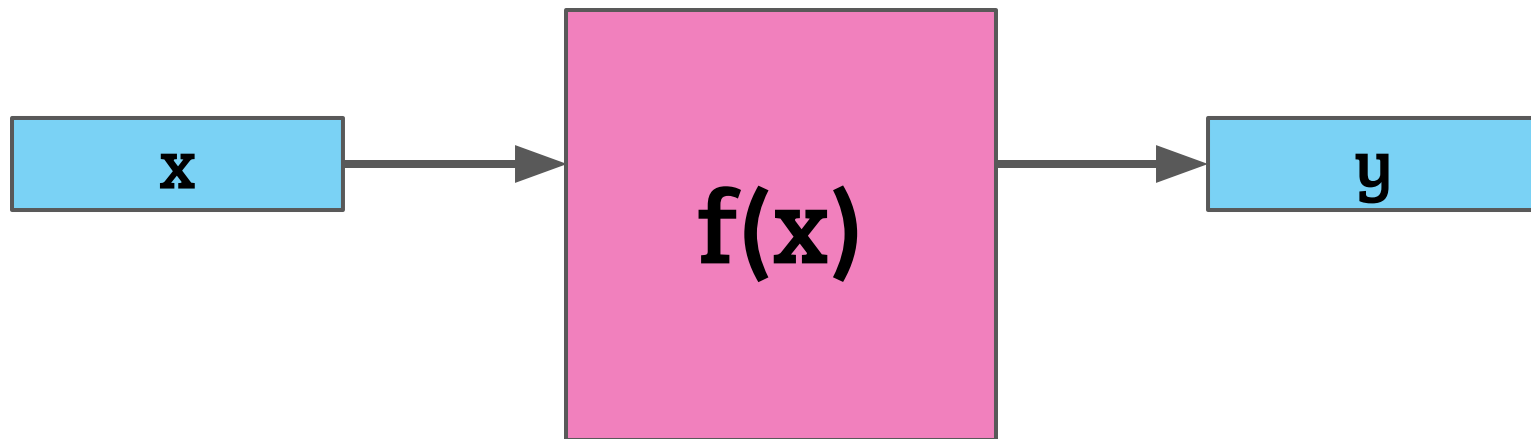
Aprendizado supervisionado



Aprendizado supervisionado



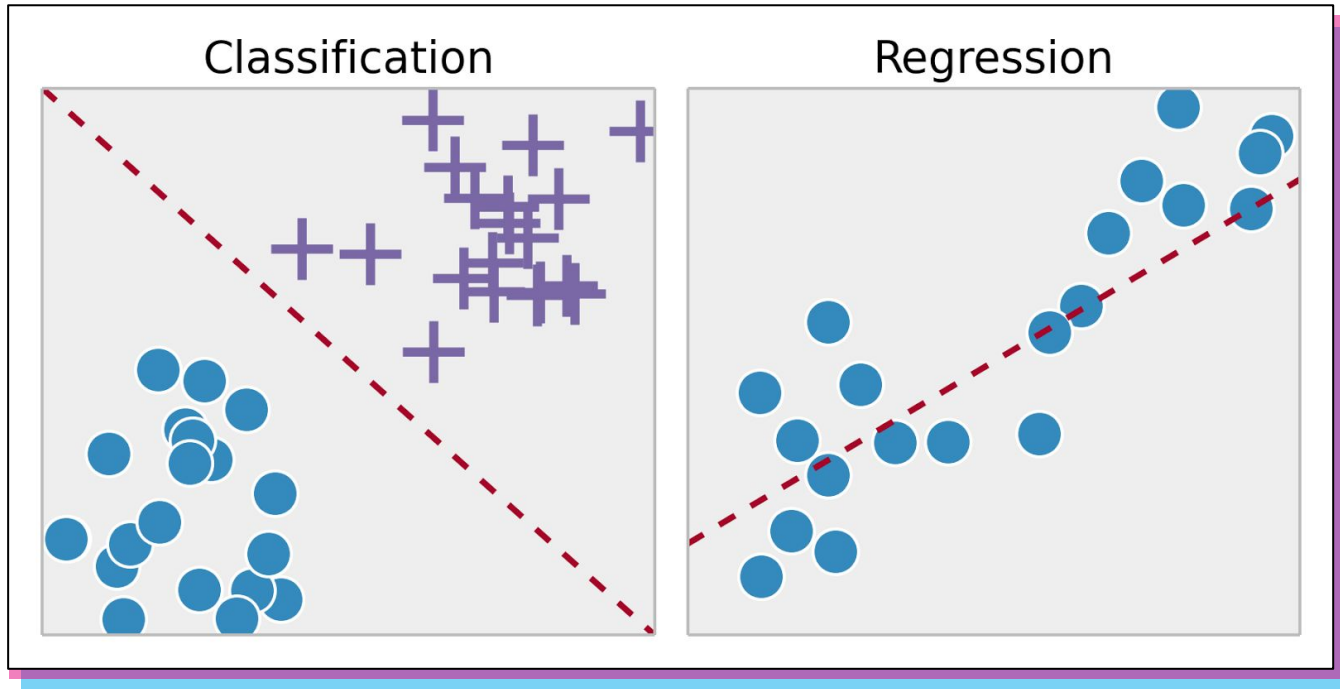
Aprendizado supervisionado



Aprendizado supervisionado

- Existem dois principais tipos de aprendizado supervisionado:
 - Classificação
 - Target é um valor discreto ou categórico (e.g., doente ou saudável)
 - Regressão
 - Target é uma valor real ou numérico (e.g., o preço de uma casa)

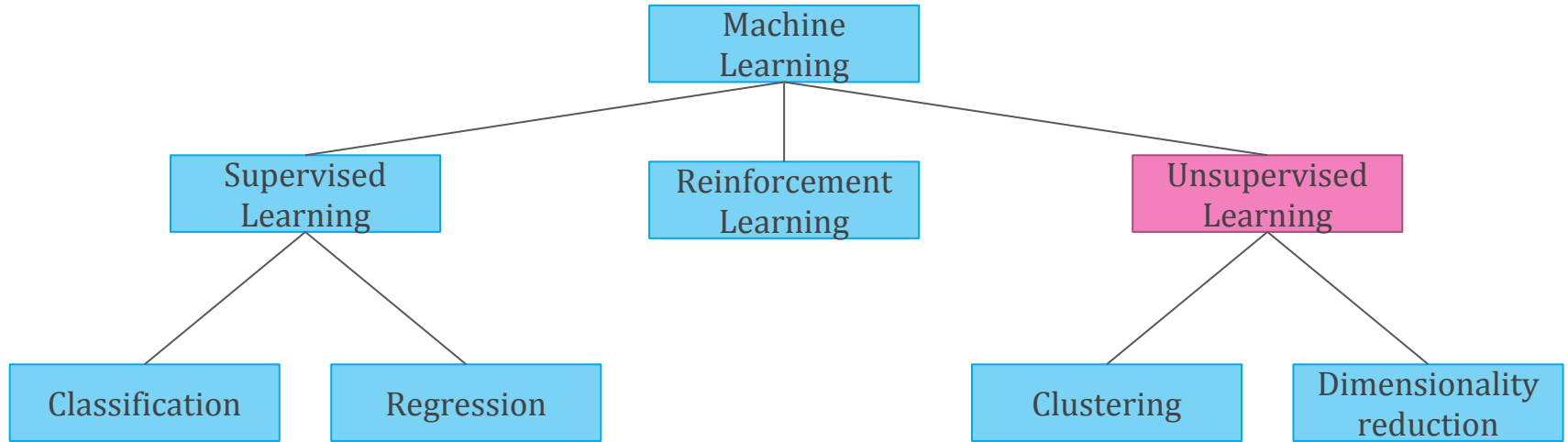




Classificação vs Regressão



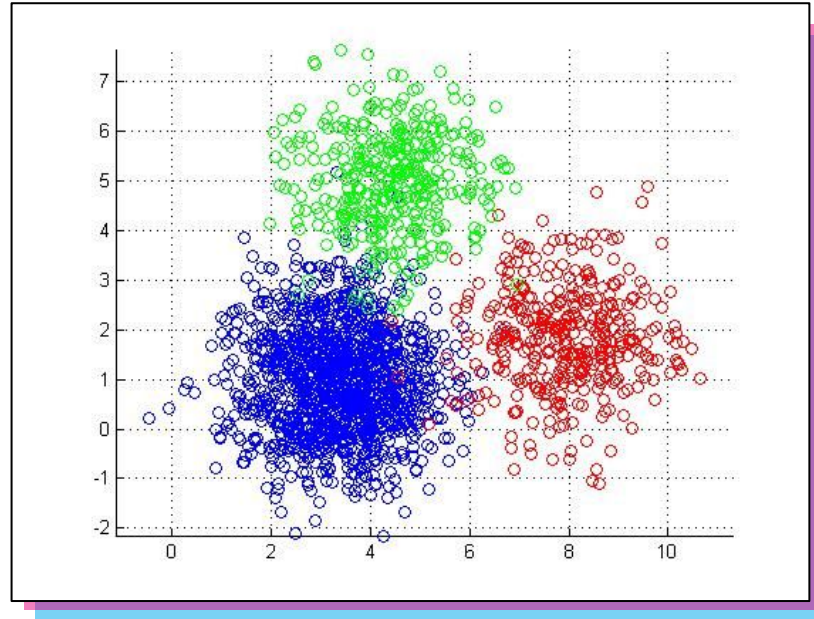
Modelos de machine learning



Aprendizado não-supervisionado

- Já o aprendizado não-supervisionado **não possui o *target***, então ele aprende com a estrutura dos dados
- É utilizado principalmente nas seguintes tarefas:
 - **Agrupamento (Clustering):** Agrupar dados em diferentes conjuntos (e.g., separar clientes em diferentes conjuntos)
 - **Redução de dimensionalidade:** Reduzir o número de dimensões que um dado é representado (e.g., transformar dados 5D em dados 3D para uma visualização)
 - **Outros:** detecção de anomalias, compressão de dados, ...

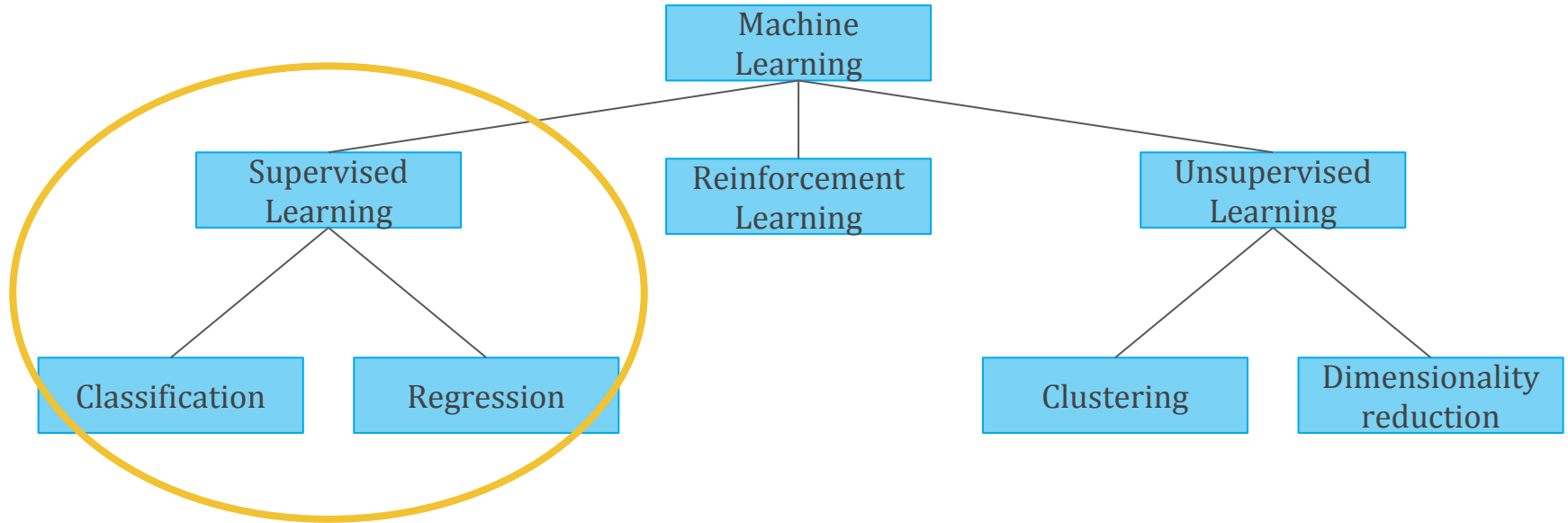




Clustering (exemplo algoritmo k-means)



Modelos de machine learning

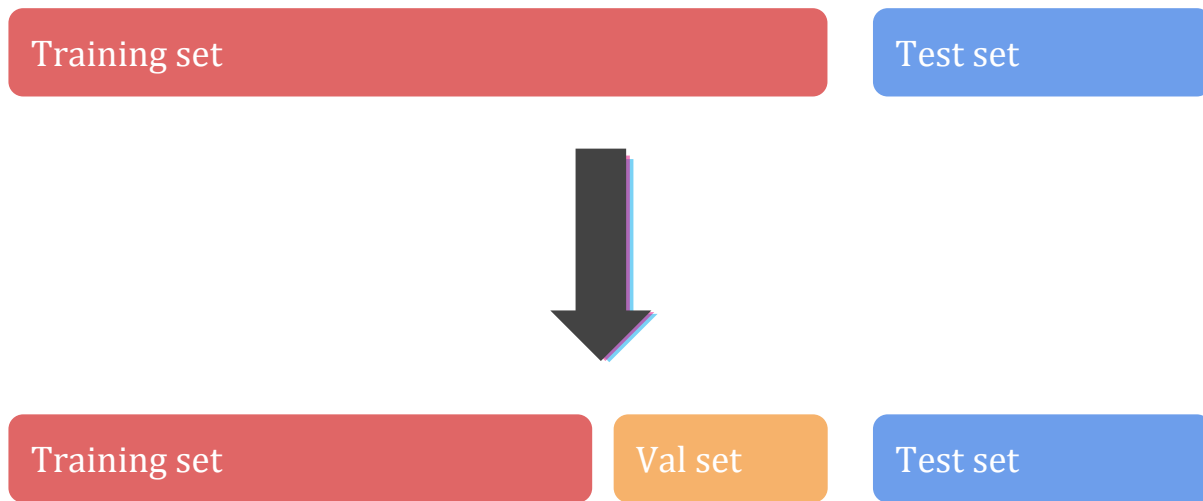


Avaliação

- A princípio um problema de ML possui dois conjuntos de dados:
 - O conjunto de treino (onde o *target* é conhecido)
 - O conjunto de teste (onde não conhecemos o *target* e queremos descobri-lo)
- Mas então como vamos avaliar nosso funcionamento?
 - Para isso é comum dividir o conjunto de treino em um **conjunto de validação** (*validation set*) e o conjunto de treino que o modelo irá receber
 - Assim temos as respostas corretas e podemos compará-las com as produzidas pelo nosso modelo



Validation Set (conjunto de validação)

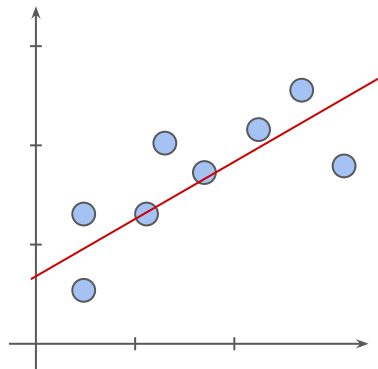


Métricas de avaliação

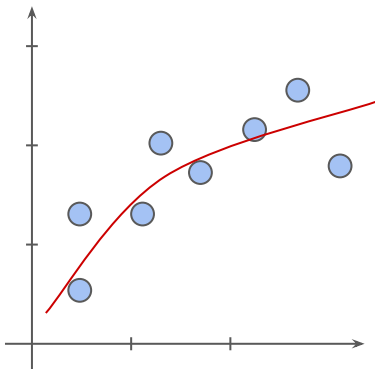
- Acurácia - classificação
 - Quantas das classificações produzidas pelo modelo estão corretas dividido pelo total de classificações realizadas
 - Valor em $[0,1]$ que indica porcentagem correta
- Mean Squared Error (Erro quadrático médio) - regressão
 - A média do quadrado das diferenças entre o valor produzido pelo modelo e o valor esperado
 - Valor em $[0,+\infty)$ que depende da escala dos *targets*
- Existem muitas outras métricas que veremos mais tarde
 - Precision, Recall, AUC, F-Score, Mean Absolute Error...



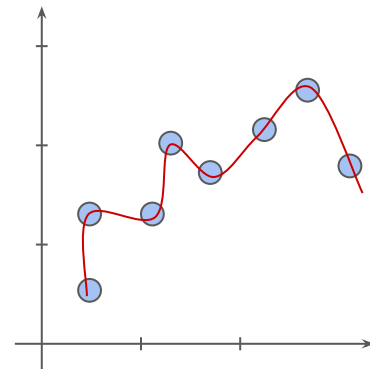
Underfitting e Overfitting



Underfitting: A função não está bem ajustada aos dados do conjunto de treino



OK: A função está ajustada aos dados de treino na quantidade correta



Overfitting: A função está mais ajustada aos dados de treino do que deveria

