

Aprendizado de Máquina

Principais Conceitos

Prof. Dr^a. Andreza Sartori

asartori@furb.br

Documentos Consultados/Recomendados

- RUSSEL, Stuart; NORVIG, Peter. **Inteligência artificial**. Rio de Janeiro : GEN LTC, 2013. 1 recurso online. Disponível em: <https://integrada.minhabiblioteca.com.br/books/9788595156104> . Acesso em: 26 jul. 2021.
- NG, Andrew; Guestrin, Carlos; Charikar, Moses. **Machine Learning**. Stanford University. Disponível em: <http://cs229.stanford.edu/materials.html>
- MALIK, Jitendra. **Computer Vision**. UC Berkeley. Disponível em: <https://www-inst.eecs.berkeley.edu/~cs280/sp15/index.html>
- IA Expert Academy. Plataforma de Cursos sobre Inteligência Artificial. Disponível em: <https://www.youtube.com/channel/UCaGrIWpwjVXT6OIQh9W4Riw>
- Caruso, Tiago. Introdução ao Aprendizado de Máquina (Machine Learning). Udemy, 2022
- Louppe, Gilles. Deep Learning, ULiège, 2022. Disponível em: <https://github.com/glouppe/info8010-deep-learning>

Conteúdo Programático:

Unidade 1: Fundamentos de Aprendizado de Máquina

Unidade 2: Aprendizado Supervisionado

Unidade 3: Aprendizado Não Supervisionado

Unidade 4: Redes Neurais Artificiais

Unidade 5: Aplicações de Aprendizado de Máquina



Conteúdo Programático:

Unidade 1: Fundamentos de Aprendizado de Máquina

Unidade 2: Aprendizado Supervisionado

Unidade 3: Aprendizado Não Supervisionado

Unidade 4: Redes Neurais Artificiais

Unidade 5: Aplicações de Aprendizado de Máquina



Conteúdo Programático:

Unidade 1: Fundamentos de Aprendizado de Máquina

1.1. Definições

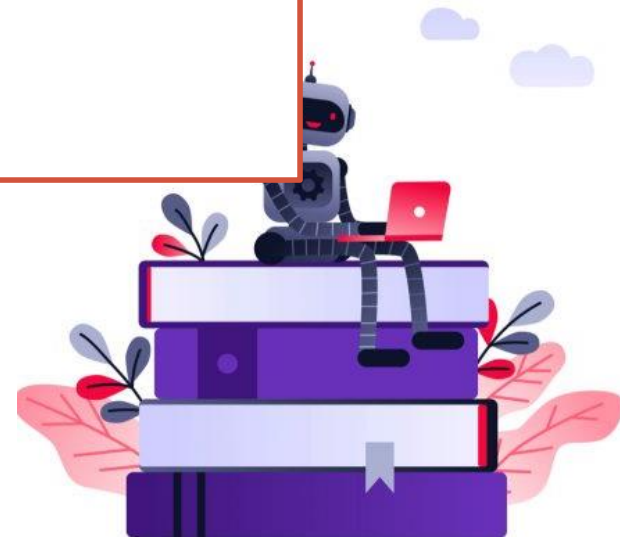
1.2. Tipos de Aprendizado de Máquina

1.2.1. Aprendizado Supervisionado

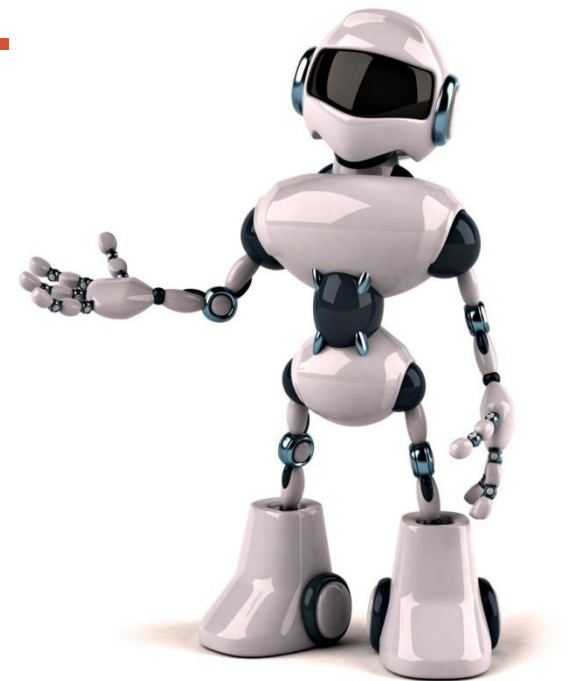
1.2.2. Aprendizado Não Supervisionado

1.2.3. Aprendizado por Reforço

1.3. Principais Conceitos



Recapitulando...



Como podemos aprender?

- Que estratégias utilizamos para aprender?



Tipos de Aprendizado de Máquina



Aprendizado
Supervisionado

The diagram consists of three red-outlined circles arranged in a triangular pattern. The top-left circle contains the text 'Aprendizado Supervisionado'. The top-right circle contains the text 'Aprendizado Não Supervisionado'. The bottom-center circle contains the text 'Aprendizado Por Reforço'.

Aprendizado
Não
Supervisionado

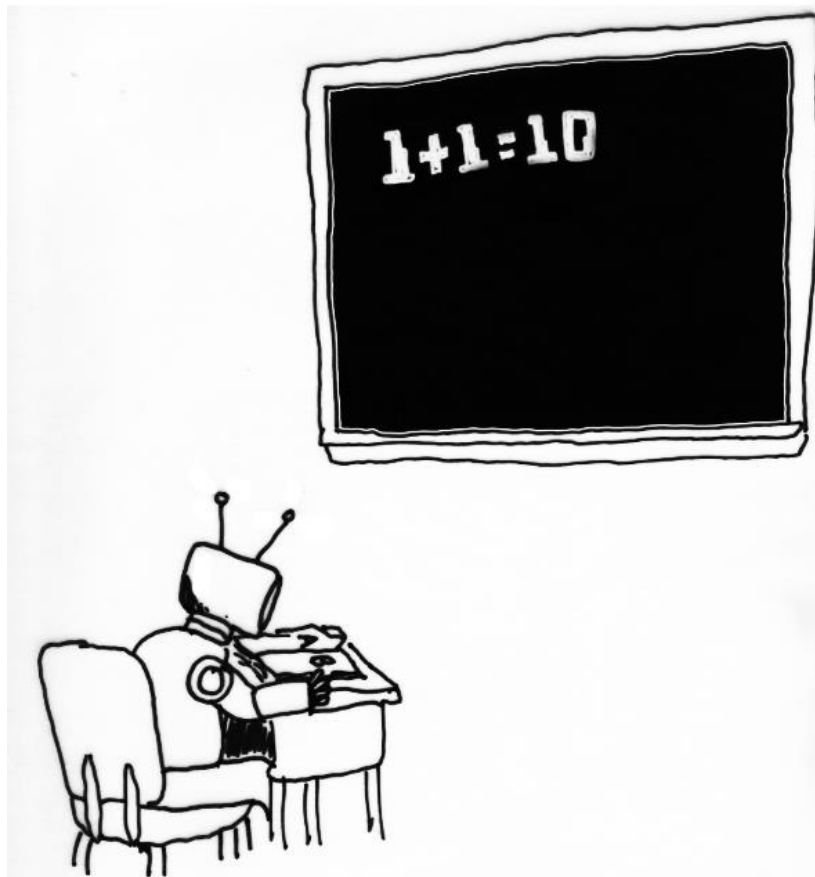
Aprendizado
Por Reforço

Tipos de Aprendizado de Máquina

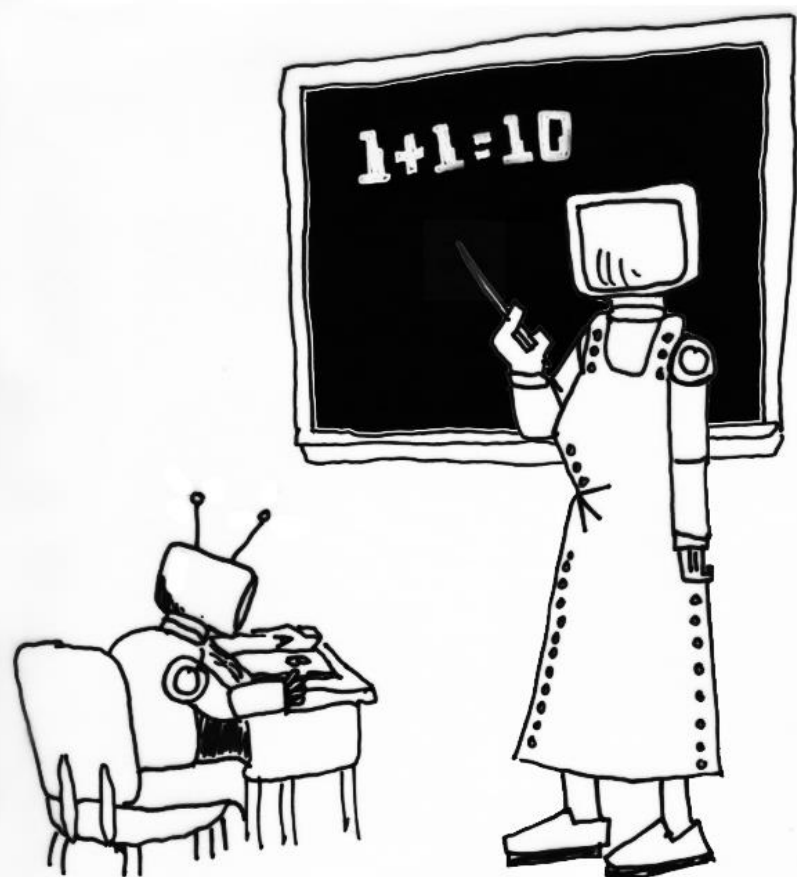


Tipos de Aprendizado de Máquina

UNSUPERVISED MACHINE LEARNING



SUPERVISED MACHINE LEARNING

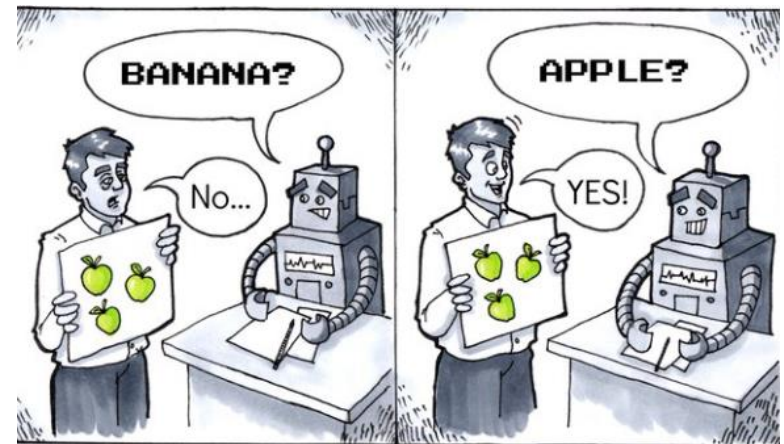


Aprendizado Supervisionado

- Vamos ensinar o computador como e/ou o que ele deve fazer.
- Aprendizagem de uma função a partir de exemplos de entrada e saída.
- Damos respostas corretas para cada exemplo.

- **Abordagens:**

- Classificação
- Regressão



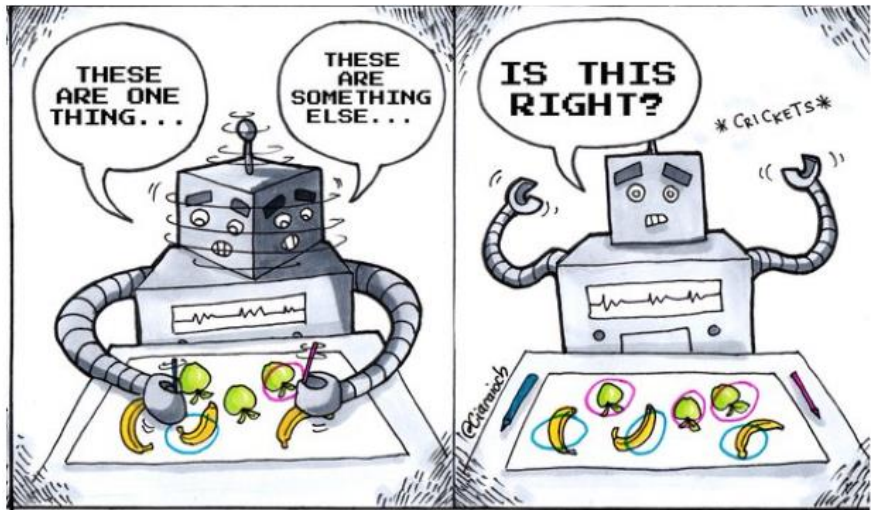
Supervised Learning

Aprendizado Não Supervisionado

- Deixamos o computador aprender sozinho.
- Quando não há valores de saída específicos.
- Respostas corretas não são dadas.

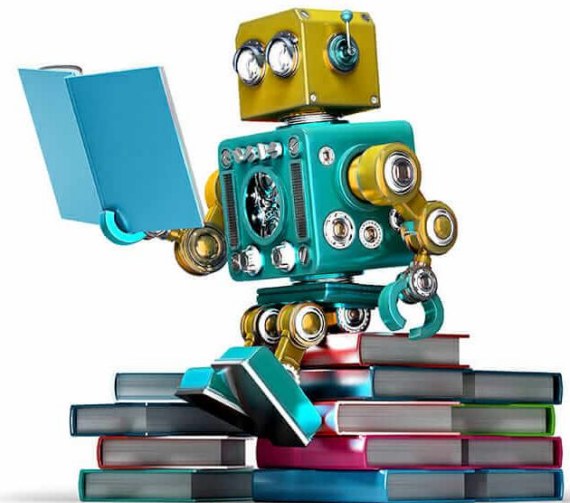
- **Abordagens:**

- Agrupamento (Clustering),
- Regras de associação.
- Detecção de desvios
- Padrões sequenciais
- Sumarização



Unsupervised Learning

Mas, como funciona o processo de aprendizagem?



O que é um modelo?

- Simplificação/representação da realidade
- Mas não é cópia da realidade



Marc Chagall
The Green Donkey, 1911



René Magritte
A Traição das Imagens do artista, 1929

O que é um modelo?

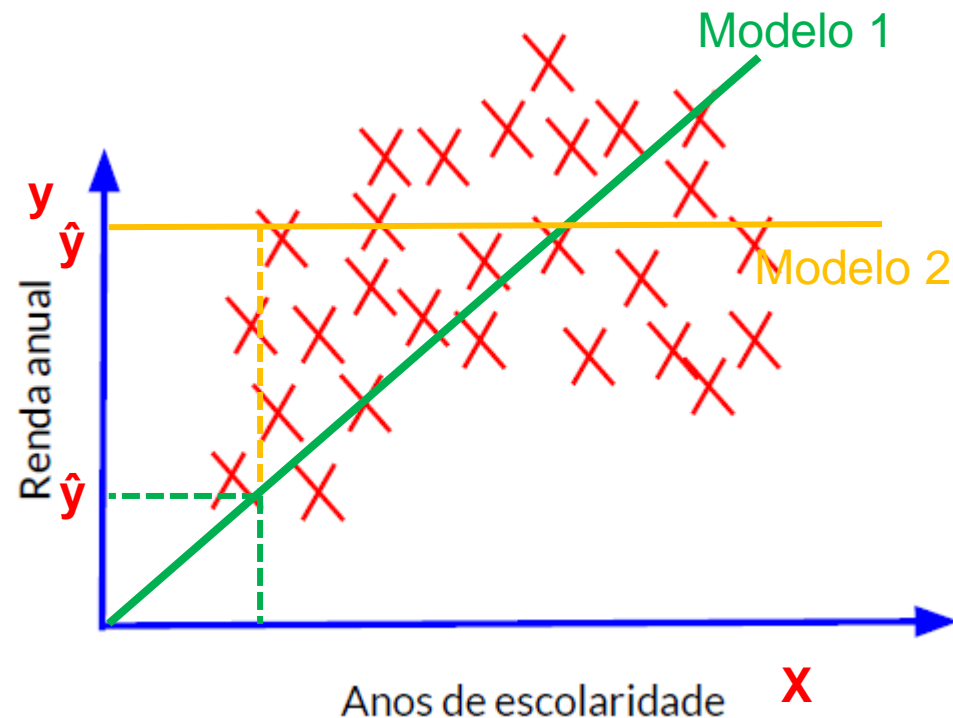
- Um modelo é a **representação de um sistema** através de conceitos matemáticos.
- Geralmente, especifica a **relação entre variáveis**.
- **Modelo** = conjunto de relações lógico-matemáticas entre as variáveis que descrevemos **dados observados e as respostas desejadas**.
- Modelo = algoritmo + parâmetros ajustados para máximo desempenho com base nos dados observados.

Representação de um Modelo

- **Variáveis X (independente):** são as características ou input (o que vamos usar para explicar)
- **Variável y (dependente):** variável objetivo ou rótulo (label) ou output (é o que queremos prever).
- **n** exemplos (pares de variáveis $x^{(i)}, y^{(i)}$)
- **\hat{y}** : Valores estimados

Observação	Anos de escolaridade (x)	Renda anual (y)
1	8	60,000
2	16	116,000
...
N	12	97,000

$n \{x, y\}$



Teoria da Aprendizagem

Dado um conjunto de treinamento de N pares de exemplos de entrada e saída

$$(x_1, y_1), (x_2, y_2), \dots (x_n, y_n),$$

onde cada valor de y pode ser encontrado por uma função desconhecida:

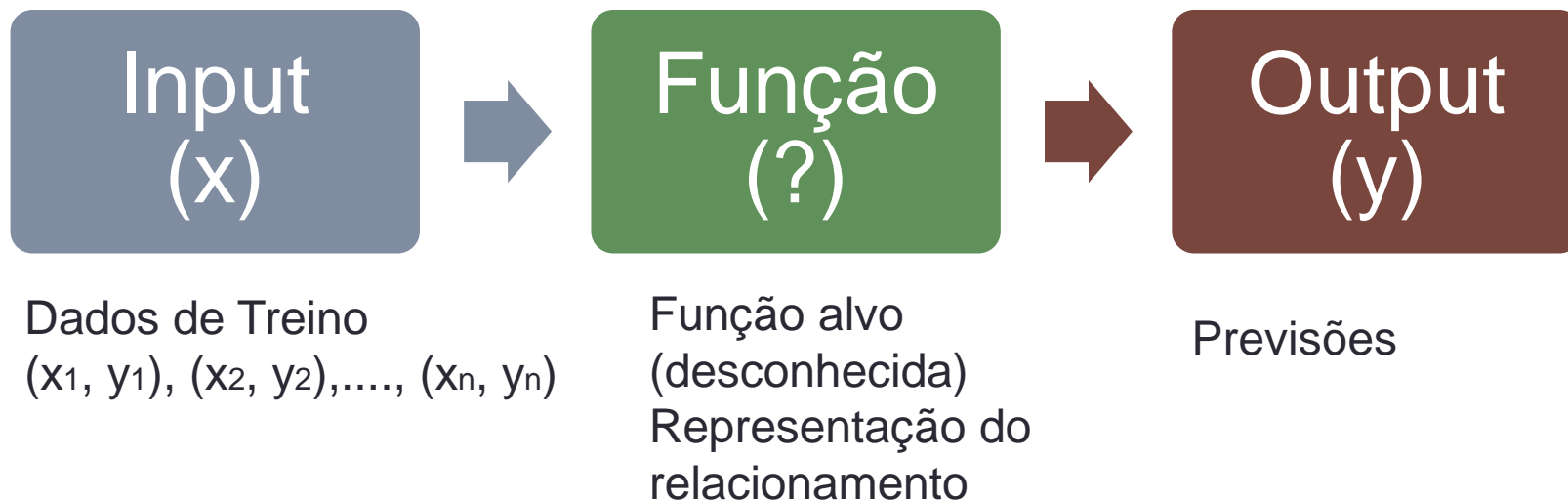
$$y = f(x),$$

o objetivo da aprendizagem é descobrir uma função **h (hipótese)** que se aproxime da função verdadeira f

Teoria da Aprendizagem

O objetivo da aprendizagem é descobrir uma função **h** (**hipótese**) que se aproxime da função verdadeira **f**

$$y = f(x)$$



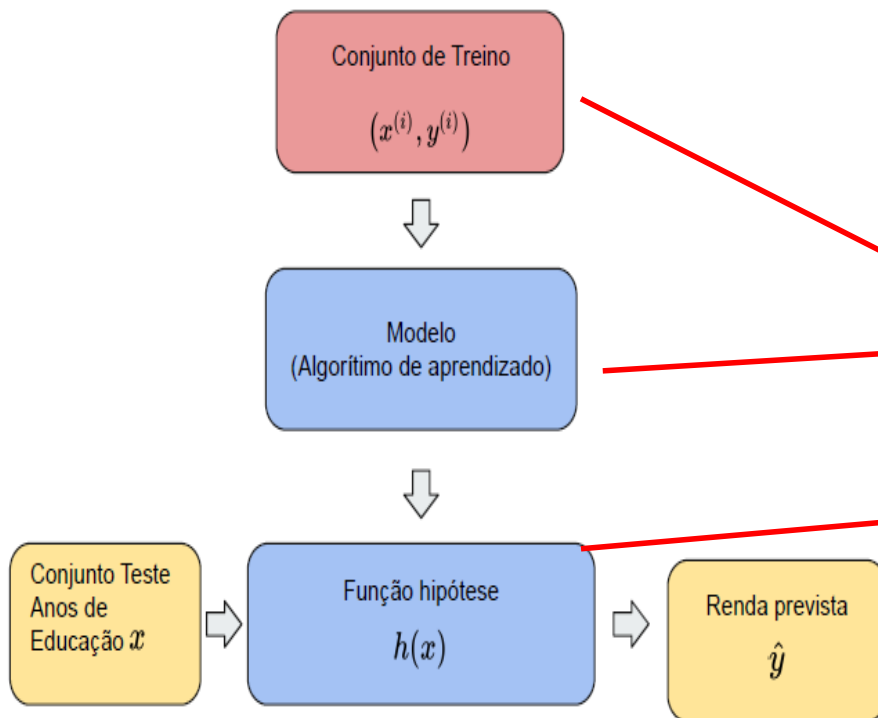
Exemplo de Modelo

Exemplo em Regressão Linear:

$$y = \alpha + \beta x$$

x = anos de escolaridade

y = renda



$$\bar{\alpha} = 10.000, \bar{\beta} = 20.000$$

$$h(x) = \bar{\alpha} + \bar{\beta} * x$$

$$h(x) = 10.000 + 20.000 * x$$

$$\hat{y} = 10.000 + 20.000 * 8$$

$$\hat{y} = 170.000,00$$

$$X=8$$

Modelos de Aprendizagem

Espaço de Hipóteses

Contém os recursos com os quais podemos trabalhar.

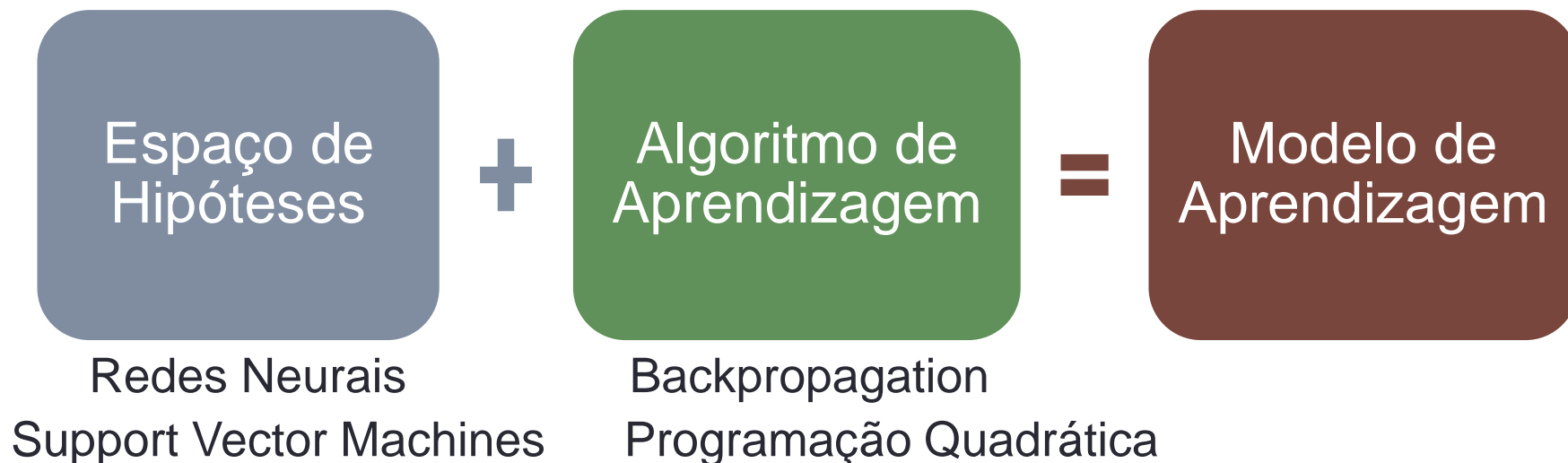
Exemplo: Redes Neurais Artificiais, Support Vector Machines

Algoritmo de Aprendizagem

Recebe os dados e navega pelo Espaço de Hipóteses a fim de encontrar a melhor hipótese que gera o resultado desejado.

Exemplo: Backpropagation, Programação Quadrática

Modelos de Aprendizagem



- O algoritmo é uma peça de código escrito que permite buscar dentro do espaço de hipóteses uma solução.
- A combinação entre espaço de hipóteses e o algoritmo de aprendizagem é que gera o modelo de aprendizagem.
- É possível usar mais de 1 algoritmo no mesmo espaço de hipóteses.

O que é um Bom Modelo?

Campos de estudo diferentes tem noções diferentes do que é um bom modelo.



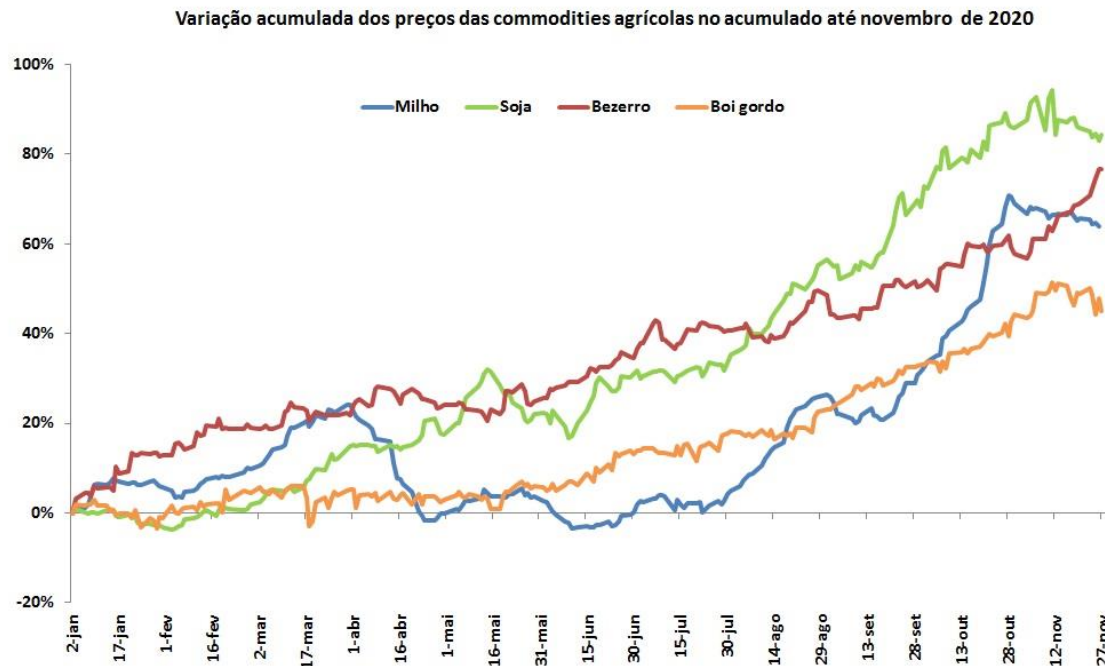
O que é um Bom Modelo?

- Algumas características de bons modelos:
 1. Poder explicativo - mapa do metro



O que é um Bom Modelo?

- Algumas características de bons modelos:
 2. Poder preditivo - quão boas são as previsões do modelo

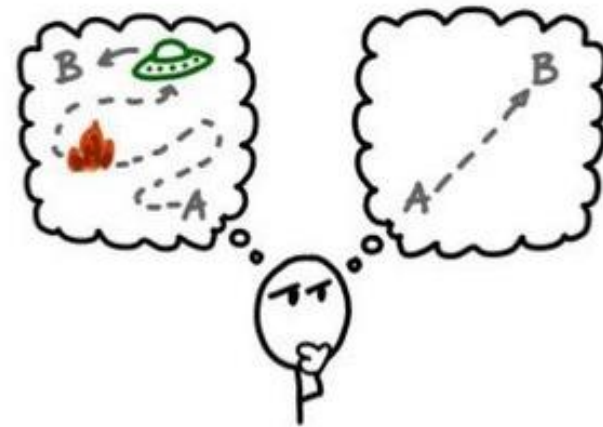


O que é um Bom Modelo?

- Algumas características de bons modelos:
 3. Falseabilidade - sabemos se o modelo está errado
 4. Simplicidade - Navalha de Occam



Occam's Razor



"When faced with two equally good hypotheses, always choose the simpler."

Vamos supor que tenhamos dois desafios em vista:

1. Salvar as nossas vidas de um ataque de um dinossauro feroz:

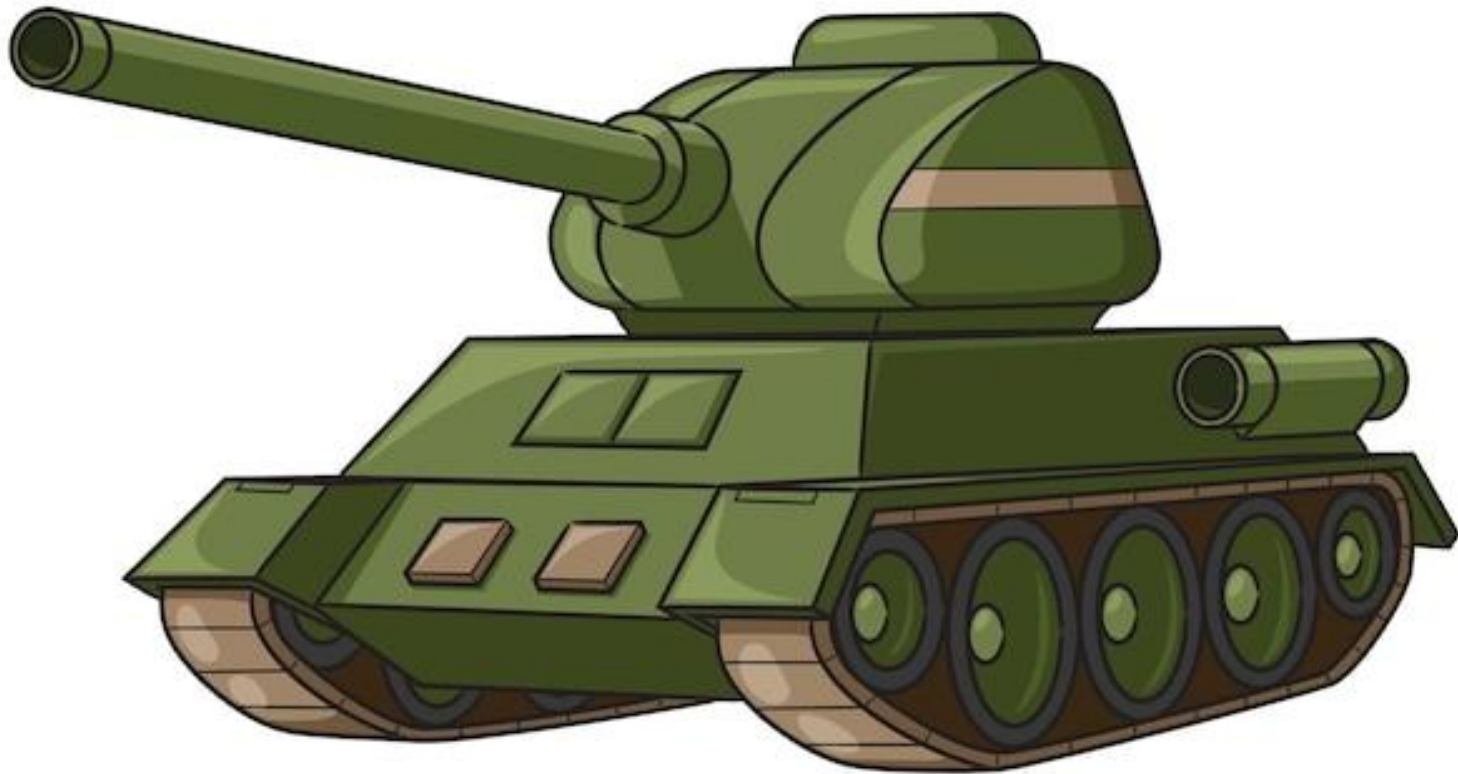


Vamos supor que tenhamos dois desafios em vista:

2. Enfrentar um único mosquito que está desafiando o nosso sono:



Para executar tais tarefas temos disponível:



Para executar tais tarefas temos disponível:



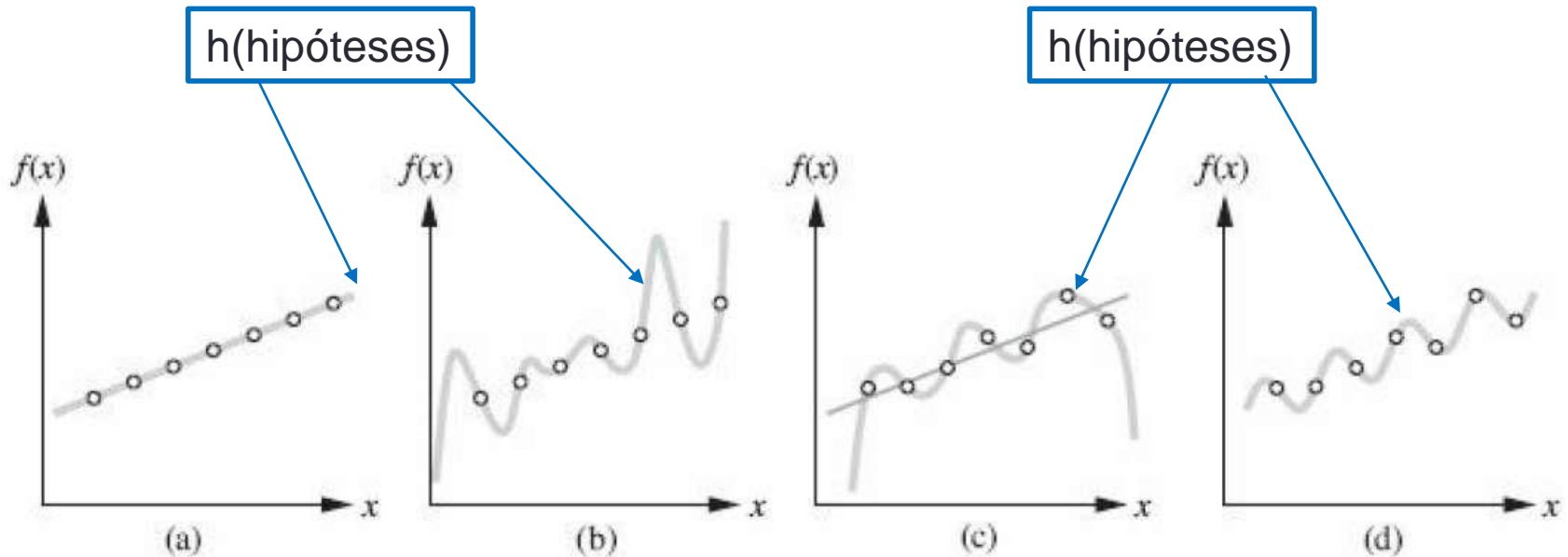
Qual deles você escolheria para cada tarefa?



Overfitting X Underfitting

- A pergunta parece retórica, mas um projeto de aprendizado de máquina mal executado pode escolher a ferramenta errada!
 - Escolhendo o tanque para abater o mosquito: Overfitting;
 - Escolher o chinelo para neutralizar o tiranossauro: Underfitting.

Teoria da Aprendizagem



Fonte: Data Science Academy

Os exemplos são pontos no plano (x, y) , onde $y = f(x)$.

O que é um Bom Modelo?

- Algumas características de bons modelos:
 1. Poder explicativo - mapa do metro
 2. Poder preditivo - quão boas são as previsões do modelo
 3. Falseabilidade - sabemos se o model está errado
 4. Simplicidade - navalha de Occam
 5. Generalizável - se aplica a situações diferentes

Em Aprendizado de Máquina um bom modelo é aquele que faz boas previsões em uma nova base de dados (***out of sample***).

É necessário também dividir a base de dados



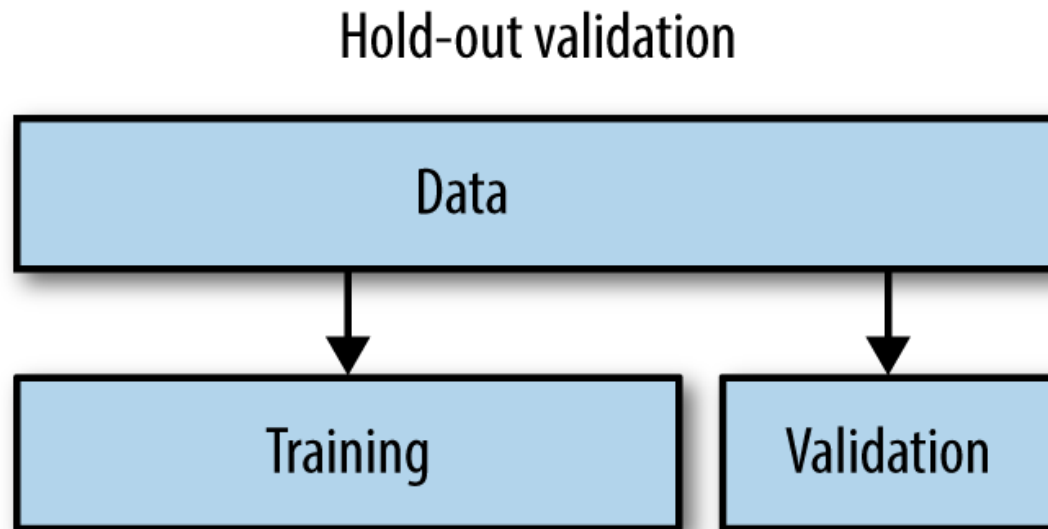
Por que dividir os conjuntos de dados?

- O objetivo da modelagem preditiva é encontrar um algoritmo capaz de gerar modelos que generalizem além do conjunto de dados para o qual foram treinados.
- Para fazer isso é comum criar um experimento para avaliar diferentes algoritmos, com suas respectivas parametrizações, aplicados a um mesmo conjunto de dados.
- Existem diferentes abordagens, que podem variar de acordo com a complexidade do problema e o volume de dados disponível.

Validações Cruzadas (Cross-Validations)

1. Holdout:

- Divide o grupo de dados em 2
 - uma parte para treinamento e outra para teste
- É comum considerar $2/3$ dos dados para treinamento e o $1/3$ para teste

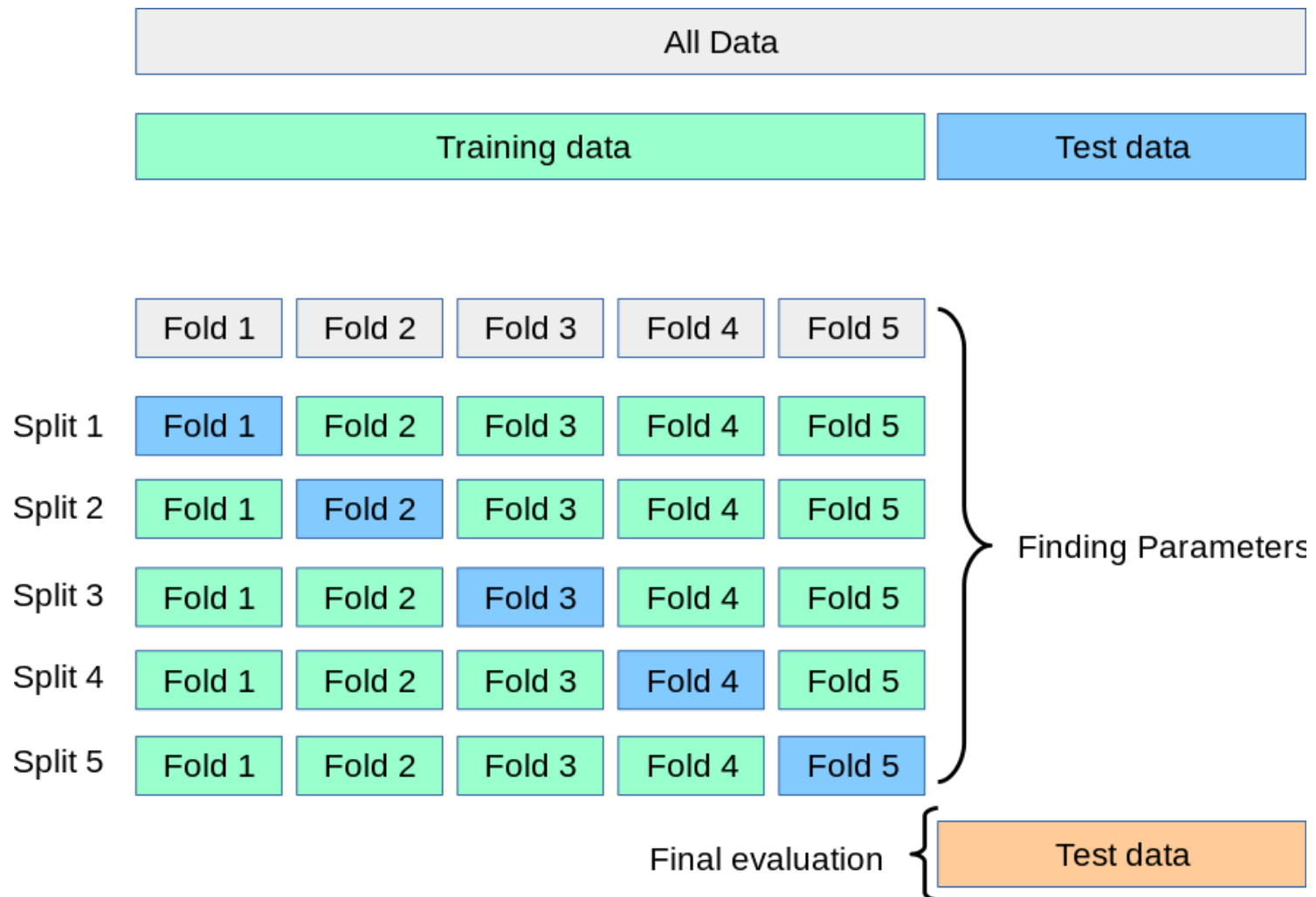


Validações Cruzadas (Cross-Validations)

2. **K-fold: (Mais utilizado!)**

- Divide o DataSet em k subconjuntos do mesmo tamanho
- Então, um subconjunto é utilizado para teste e os k restantes são utilizados para treinamento.
- Este processo é realizado k vezes alternando de forma circular o subconjunto de teste.

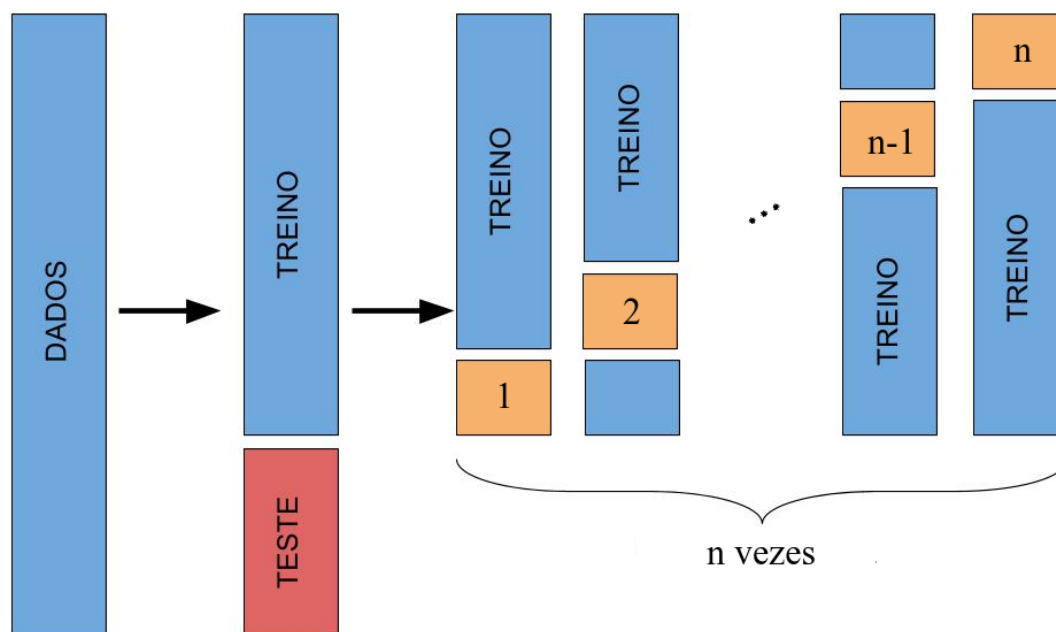
Validações Cruzadas: K-fold



Validações Cruzadas (Cross-Validations)

3. **Leave-one-out (LOOCV):** alto custo computacional!

- Caso específico do k-fold



Em cada etapa do processo, apenas uma observação é deixada de fora do treino. Repetimos esse procedimento n vezes, excluindo em cada momento uma observação diferente.

Conjunto de Treino, Validação e Teste

Educação (x)	Renda (y)
8	60,000
16	116,000
8	80,000
12	146,000
10	125,000
15	146,000
12	136,000
10	125,000
15	146,000
15	146,000

Treino

Validação

Teste

- O **conjunto de treino** é usado para estimar os parâmetros do seu modelo.
- O **conjunto de validação** serve escolhemos os hiperparâmetros e decidir qual modelo vamos usar.
- O **conjunto teste** funciona como a medida final da performance do modelo.

Conjunto de Treino, Validação e Teste

Educação (x)	Renda (y)
8	60,000
16	116,000
8	80,000
12	146,000
10	125,000
15	146,000
12	136,000
10	125,000
15	146,000
15	146,000

Treino

Validação

Teste

- Modelo 1:

$$y = \beta x$$

- Modelo 2:

$$y = \alpha + \beta x$$

- A base de dados de **treino** serve para estimar os valores nos modelos.
- Para estimar quão bom é a performance de cada modelo se usa o conjunto de **validação**.

$$f(x) = \text{modelo 1}$$

$$f(x) = \text{modelo 2}$$

Função
Custo

- Por fim se calcula a Função Custo no conjunto de **teste** para demonstrar quão bom é seu modelo.

Qual a proporção dos dados para cada Conjunto?

Educação (x)	Renda (y)
8	60,000
16	116,000
8	80,000
12	146,000
10	125,000
15	146,000
12	136,000
10	125,000
15	146,000
15	146,000

Treino

Validação

Teste

- Se conjunto de **treino for pequeno**, a variância dos parâmetros será grande.
- Se o conjunto de **validação for pequeno** a variância da estatística de seleção de modelos será grande.
- Se o conjunto **teste for pequeno** a variância do seu teste do modelo será grande.
- Em prática se recomenda dividir (70,15,15) .
- Se você tiver muitos dados, a proporção de dados no conjunto de treino tende a aumentar.
 - Se estiver estimando por exemplo 1 milhão de parâmetros (observações) 98%,1%,1%

Como dividir os dados?

Se as observações forem independentes, se divide a base de dados aleatoriamente

Educação (x)	Renda (y)
8	60,000
16	116,000
8	80,000
12	146,000
10	125,000
15	146,000
12	136,000
10	125,000
15	146,000
15	146,000

Treino
Validação
Treino
Treino
Teste
Validação
Treino
Teste
Treino
Treino

Em caso de Séries Temporais (LSTM), não se separam aleatoriamente

Tempo (x)	Preço (y)
1	60
2	62
3	65
4	64
5	68
6	66
7	69
8	71

Treino
70%
Validação
15%
Teste
15%

Como Avaliar um Modelo?*

N = número de observações

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

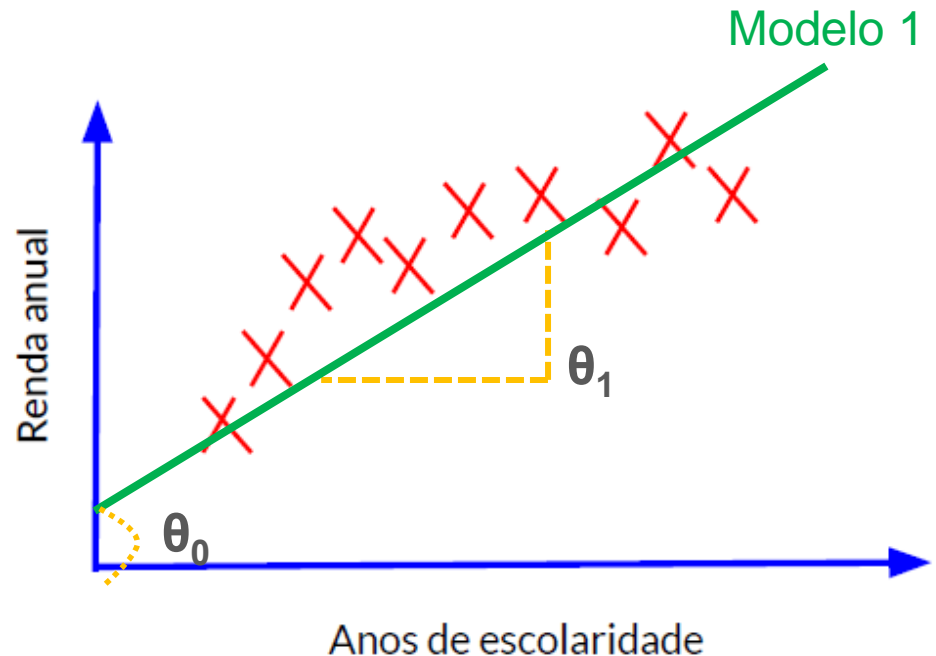
θ'_i s : parâmetros

Como parâmetros diferentes geram modelos diferentes?

Como escolher parâmetros?

θ_0 = interceptação

θ_1 = é a inclinação da reta, que mostra a relação entre escolaridade e a renda anual



Como Avaliar um Modelo?*

θ_0 = interceptação
 θ_1 = inclinação da reta
(mostra a relação entre escolaridade e a renda anual)

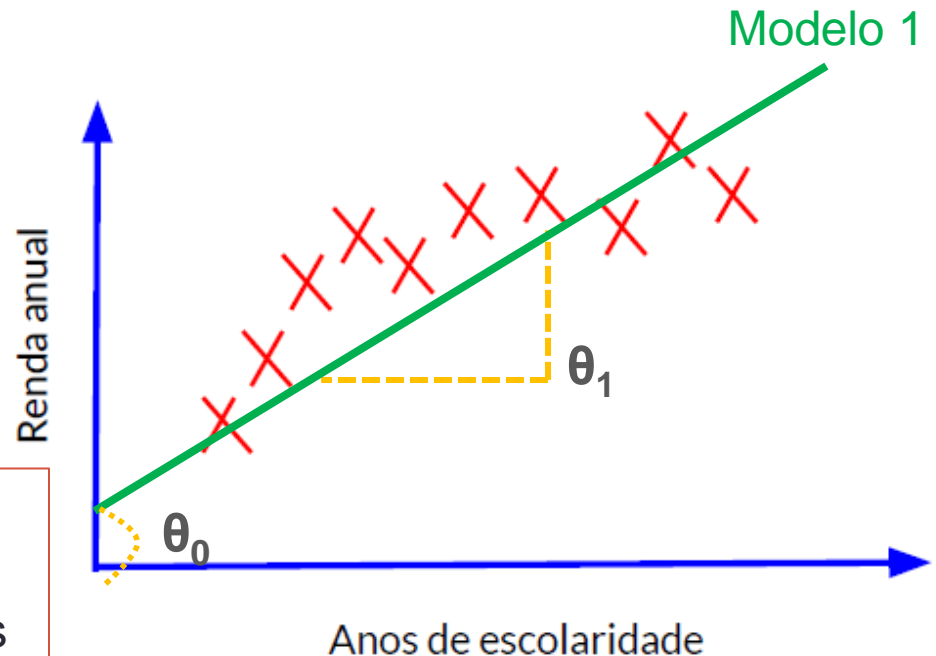
N = número de observações

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

θ'_i 's : parâmetros

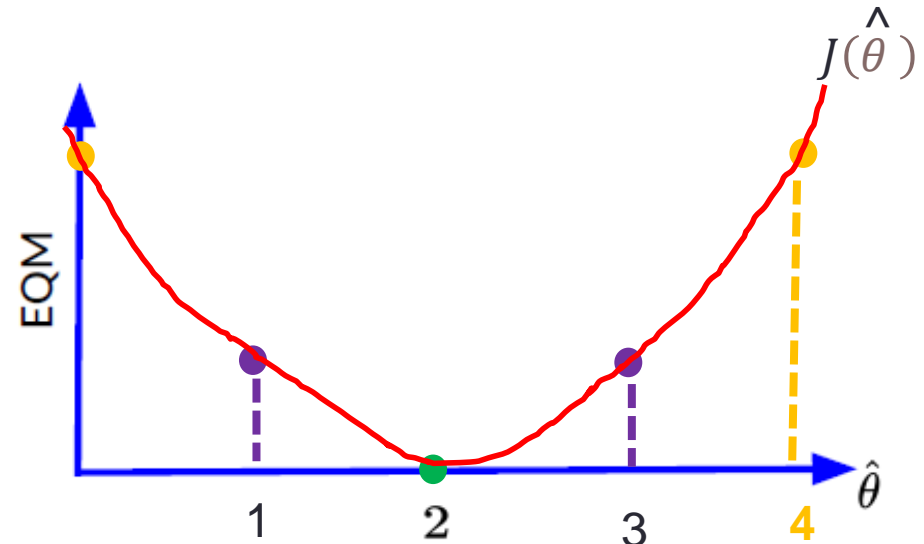
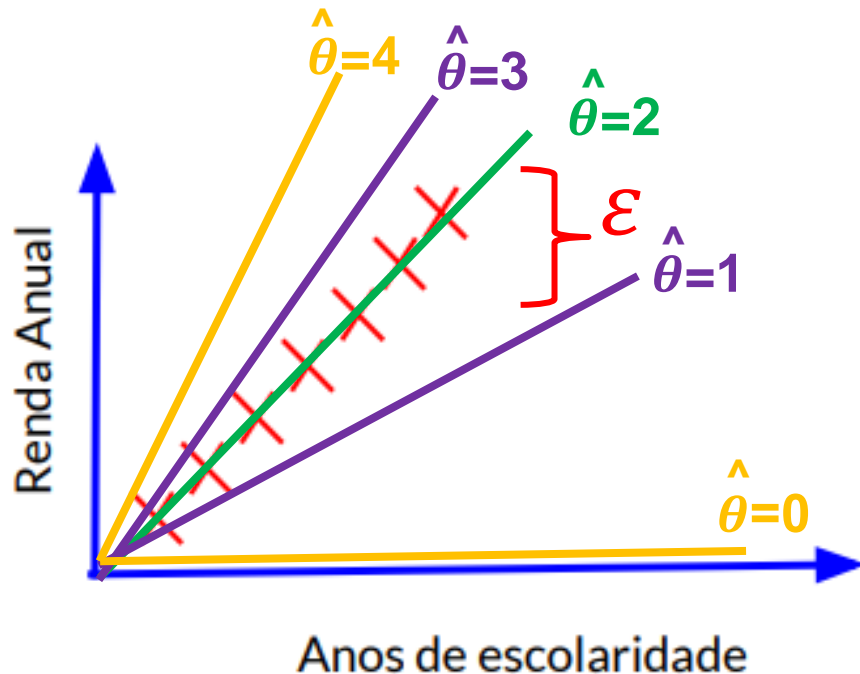
Função de Custo:

- É uma medida de dispersão dos valores verdadeiros em relação aos valores previstos
- Mede quão ruim é o modelo
 - Por isso é necessário minimizar a função.



$$\text{Min}_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

Como Avaliar um Modelo?*



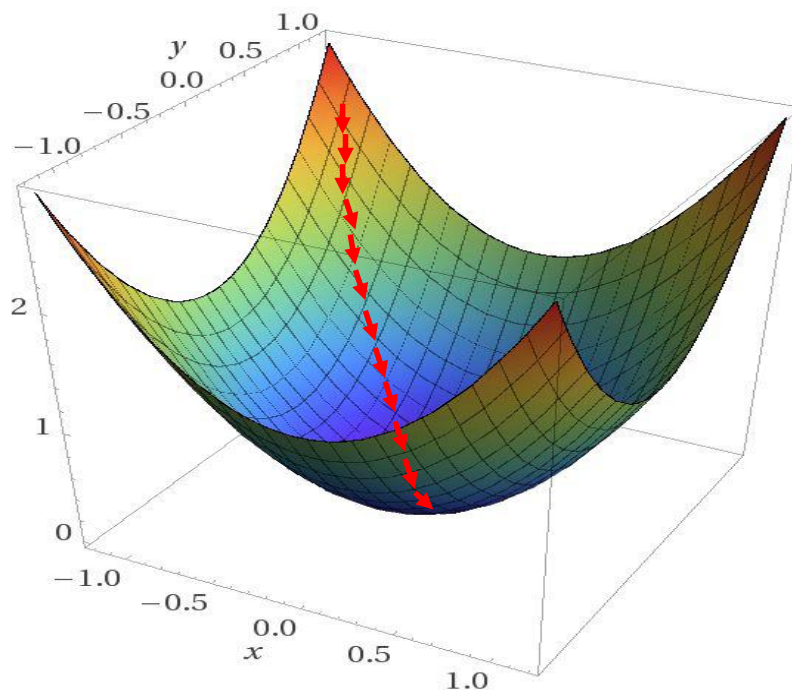
$$\text{Suponha } y = 2 * x$$

Só há um parâmetro, a inclinação da curva, que nesse caso é igual a 2.

Como a função custo (EQM) muda de acordo com o nosso parâmetro estimado θ ?

Método do Gradiente para Minimizar a Função de Custo

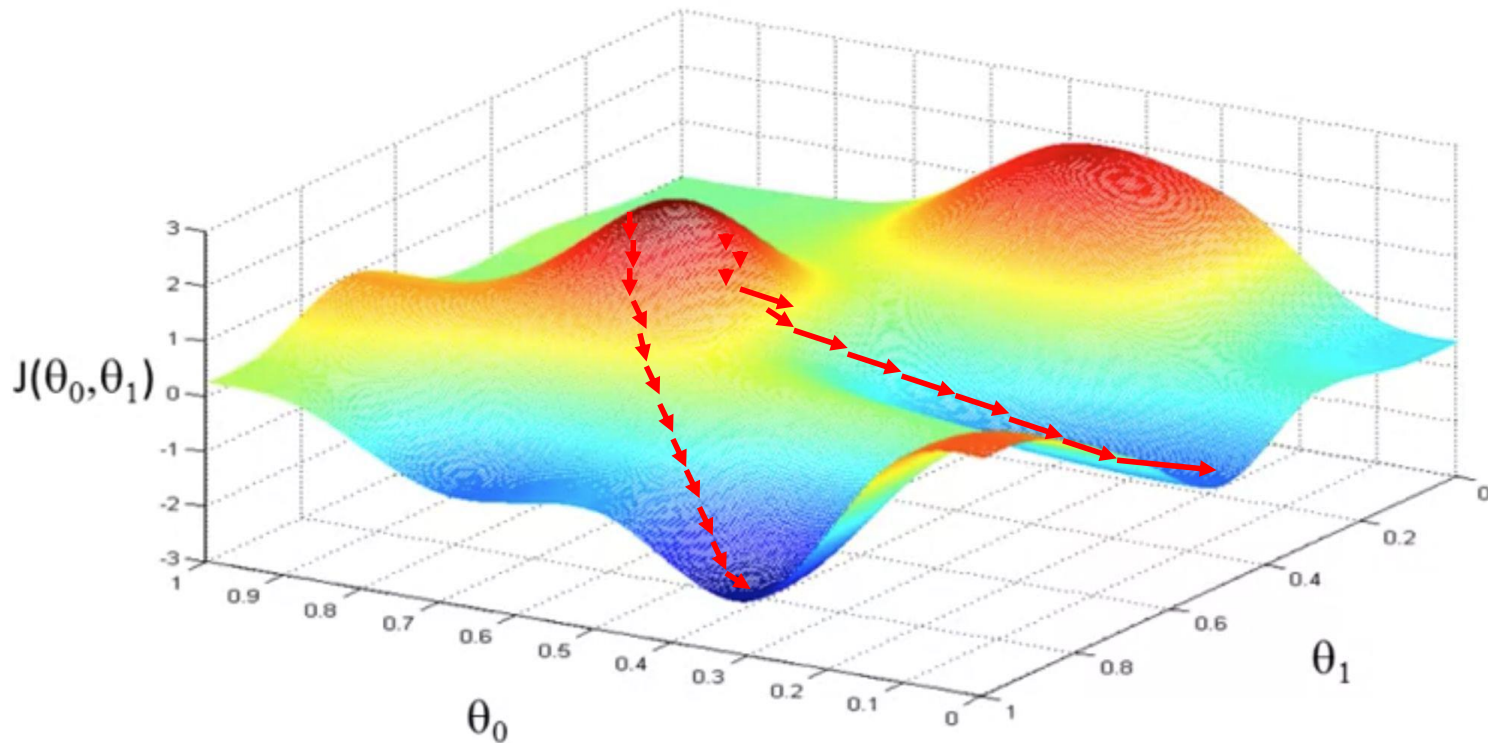
- Podemos minimizar a função custo usando o método do gradiente.
- **Gradiente** é um vetor com a derivada da função com respeito a cada uma das variáveis.
- O gradiente sempre aponta na direção de maior aumento da função (logo devemos sempre mover na direção oposta)



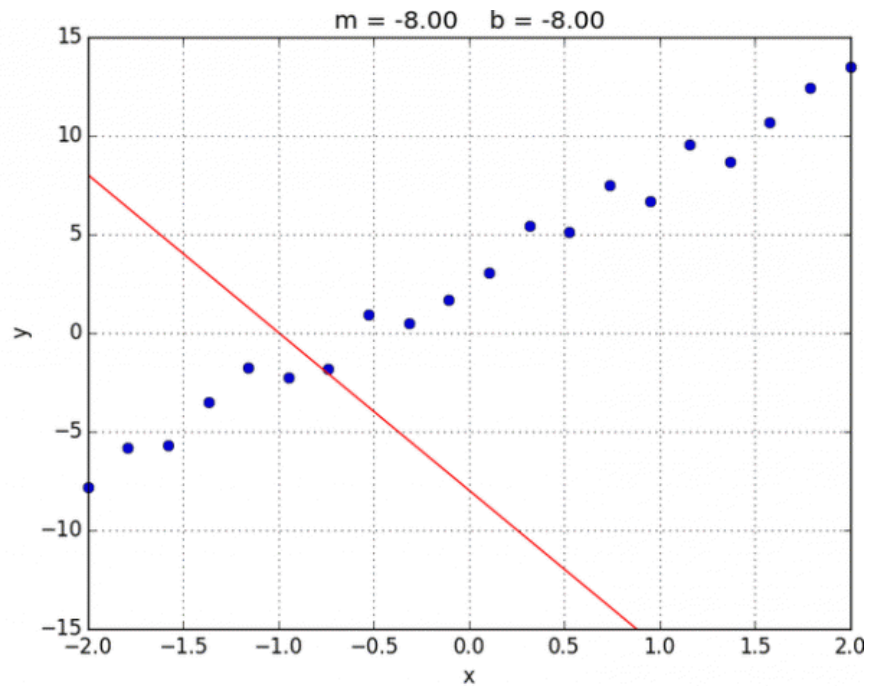
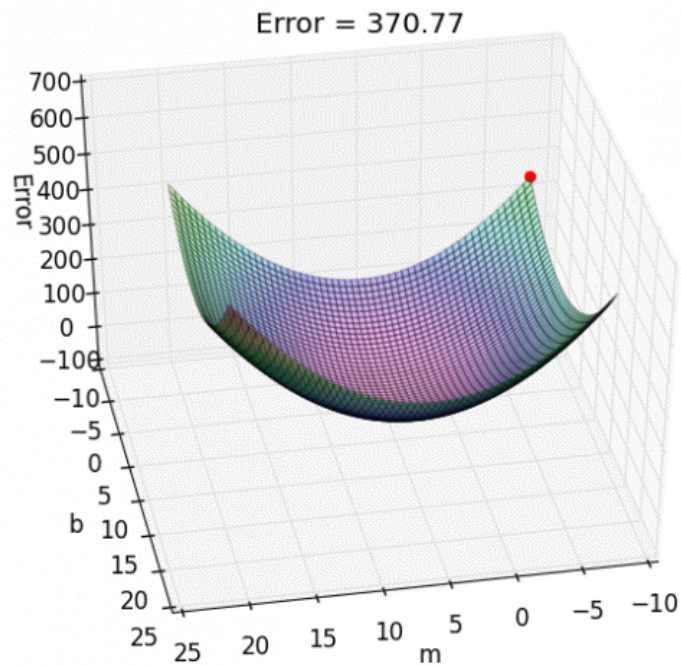
Computed by Wolfram|Alpha

Método do Gradiente para Minimizar a Função de Custo

- Garante que você irá chegar num ótimo local, mas não garante que você irá chegar num ótimo global.
- O que se faz na prática é iniciar o método gradiente com vários valores distintos e computar o mínimo com todos os valores iniciais.



Método do Gradiente



Método gradiente pseudo código

Objetivo: $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

Comece com valores aleatórios: θ_0, θ_1

Enquanto não estiver em um mínimo:

 calcule o gradiente de $J(\theta_0, \theta_1)$

 atualize θ_0, θ_1 conforme sua derivada

Você sabe que vai estar no mínimo quando todas as derivadas forem aproximadamente iguais a zero.

Matemática do método do gradiente

- Atualize os parâmetros simultaneamente até convergir:

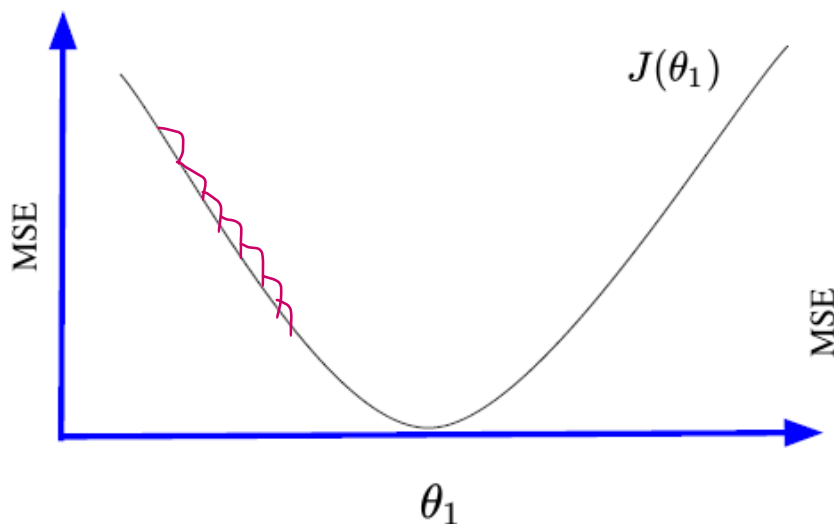
$$\begin{aligned}\theta_0 &:= \theta_0 - \alpha \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_0} \\ \theta_1 &:= \theta_1 - \alpha \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_1}\end{aligned}$$

1. Calcula a derivada de teta 0,
2. Calcula a derivada de teta 1
3. Então dá um passo alfa na direção oposta do gradiente
4. Se sabe que chegou ao mínimo local se cada um dos valores forem aproximadamente zero.

- O que é alpha?
 - Hiperparâmetro = velocidade do aprendizado

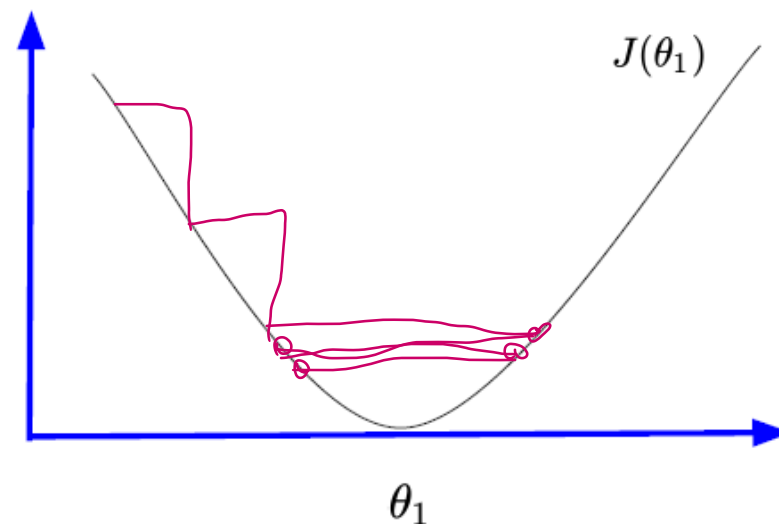
Hyperparâmetro: velocidade do aprendizado

Velocidade de aprendizado pequena demais



gradiente é muito lento
 $\alpha = 0.0001$

Velocidade de aprendizado muito grande



gradiente pode saltar o mínimo
 $\alpha = 100$

Normalmente se inicia com $\alpha = 0.1$, mas varia dependendo do modelo

Método do gradiente na Regressão Linear*

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

$$J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^N (h_{\theta}(x^i) - y^i)^2$$

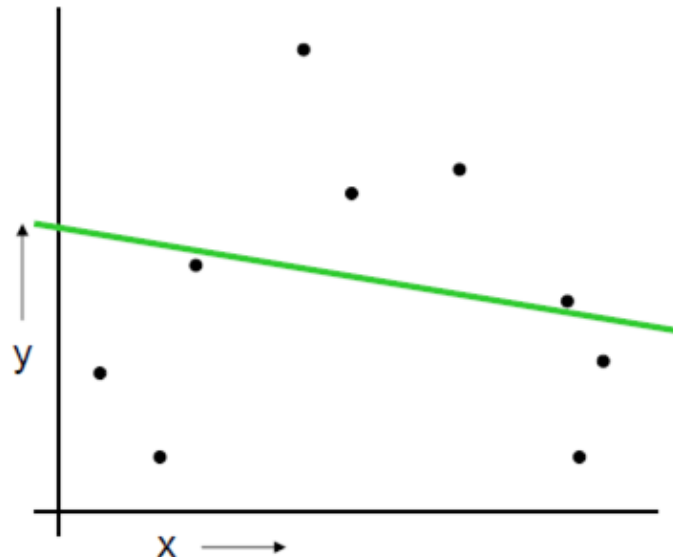
Aprendizado de Máquina

- Nenhum algoritmo único ou uma combinação de algoritmos é 100% preciso o tempo todo.
- Pelo menos não ainda!!



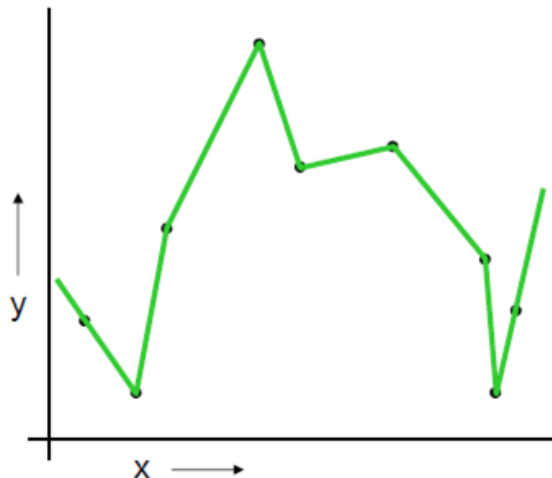
Como usar a função custo para escolher o modelo

- Sub-adequação (underfitting - **viés**) - Quando escolhemos um modelo muito simples:
 - erro elevado na aprendizagem.
 - Não consegue nem mesmo modelar os dados de treinamento e portanto não consegue generalizar para novos dados



Como usar a função custo para escolher o modelo

- Sobre-adequação (overfitting - **variância**) ocorre quando o modelo é muito complexo.
 - Erro baixo sobre os exemplos de treinamento e mais elevado para os exemplos de teste.
 - Algoritmo pode memorizar os dados no treinamento e falir ao tentar generalizar novos exemplos.
 - Aprende os detalhes e os ruídos nos dados de treinamento



Usando a função custo para escolher o modelo

$$\text{Treino EQM} = \frac{1}{2N_{\text{treino}}} \sum_{i=1}^{N_{\text{treino}}} (\hat{y}_i - y_i)^2$$

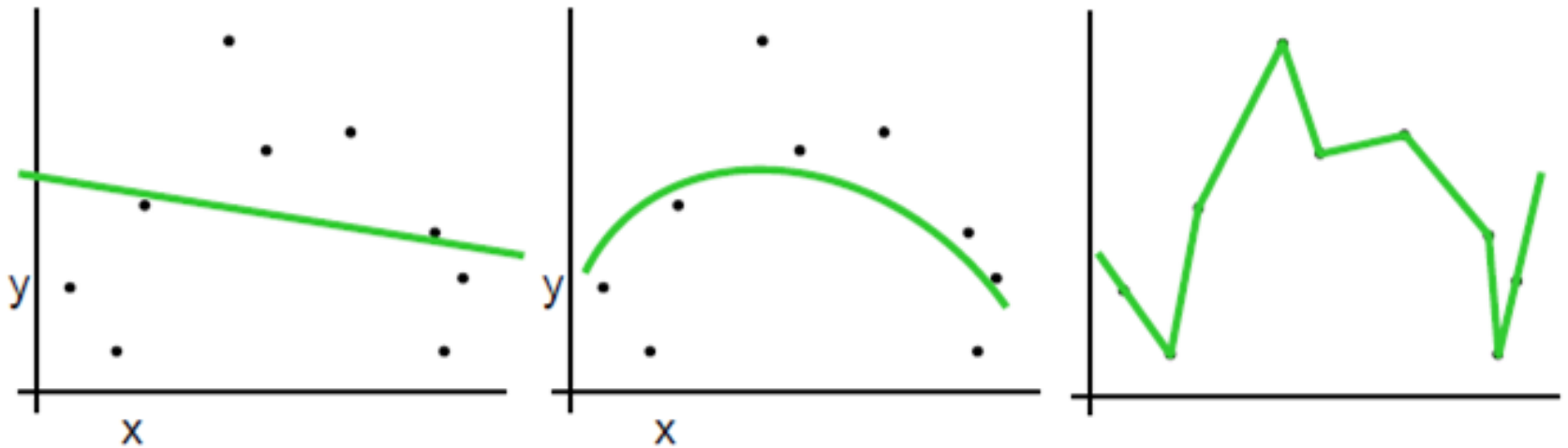
$$\text{Val EQM} = \frac{1}{2N_{\text{val}}} \sum_{j=1}^{N_{\text{val}}} (\hat{y}_j - y_j)^2$$

$$\text{Teste EQM} = \frac{1}{2N_{\text{teste}}} \sum_{k=1}^{N_{\text{teste}}} (\hat{y}_k - y_k)^2$$

- Podemos usar o custo no conjunto de treino e de validação para julgar se o **modelo está sobre ou sub adequado**.
- Se o modelo tem uma performance ruim no conjunto de treino, ele provavelmente está sub-adequado. **EQM alto $\gg \gg 0$**
- Se a diferença entre o custo no conjunto treino e validação for grande, o modelo está sobre-adequado. **Diferença EQM = ValEQM – TreinoEQM $\gg \gg 0$**
- Para escolher os modelos você deve verificar aquele que tem o menor EQM na validação.

Qual é o melhor?

Como saber se um algoritmo de aprendizado produziu uma teoria que irá fazer uma previsão corretamente?



Responda:

Qual é a principal característica de um bom modelo em aprendizado de máquina?

- a) Um bom modelo é simples
- b) Um bom modelo é possível identificar se está errado
- c) Um bom modelo é intuitivo, na qual sabemos explicar porque ele faz cada previsão
- d) Um bom modelo faz boas previsões “out of sample” – base de dados onde não foi treinado

Resposta: Resposta

Qual é a principal característica de um bom modelo em aprendizado de máquina?

- a) Um bom modelo é simples
- b) Um bom modelo é possível identificar se está errado
- c) Um bom modelo é intuitivo, na qual sabemos explicar porque ele faz cada previsão
- d) **Um bom modelo faz boas previsões “out of sample” – base de dados onde não foi treinado**

Responda:

Qual é a diferença entre conjunto de teste e validação?

- a) Conjunto de teste serve para escolher os hiperparâmetros do modelo e o conjunto de validação será a medida final da validade do modelo.
- b) Conjunto de validação serve para escolher os hiperparâmetros do modelo e o conjunto de teste será a medida final de performance do modelo.
- c) Conjunto de teste serve para treinar os parâmetros do modelo e o conjunto de validação será a medida final da validade do modelo.
- d) Conjunto de validação serve para treinar os parâmetros do modelo e o conjunto de teste será a medida final de performance do modelo.

Resposta: Resposta

Qual é a diferença entre conjunto de teste e validação?

- a) Conjunto de teste serve para escolher os hiperparâmetros do modelo e o conjunto de validação será a medida final da validade do modelo.
- b) **Conjunto de validação serve para escolher os hiperparâmetros do modelo e o conjunto de teste será a medida final de performance do modelo.**
- c) Conjunto de teste serve para treinar os parâmetros do modelo e o conjunto de validação será a medida final da validade do modelo.
- d) Conjunto de validação serve para treinar os parâmetros do modelo e o conjunto de teste será a medida final de performance do modelo.

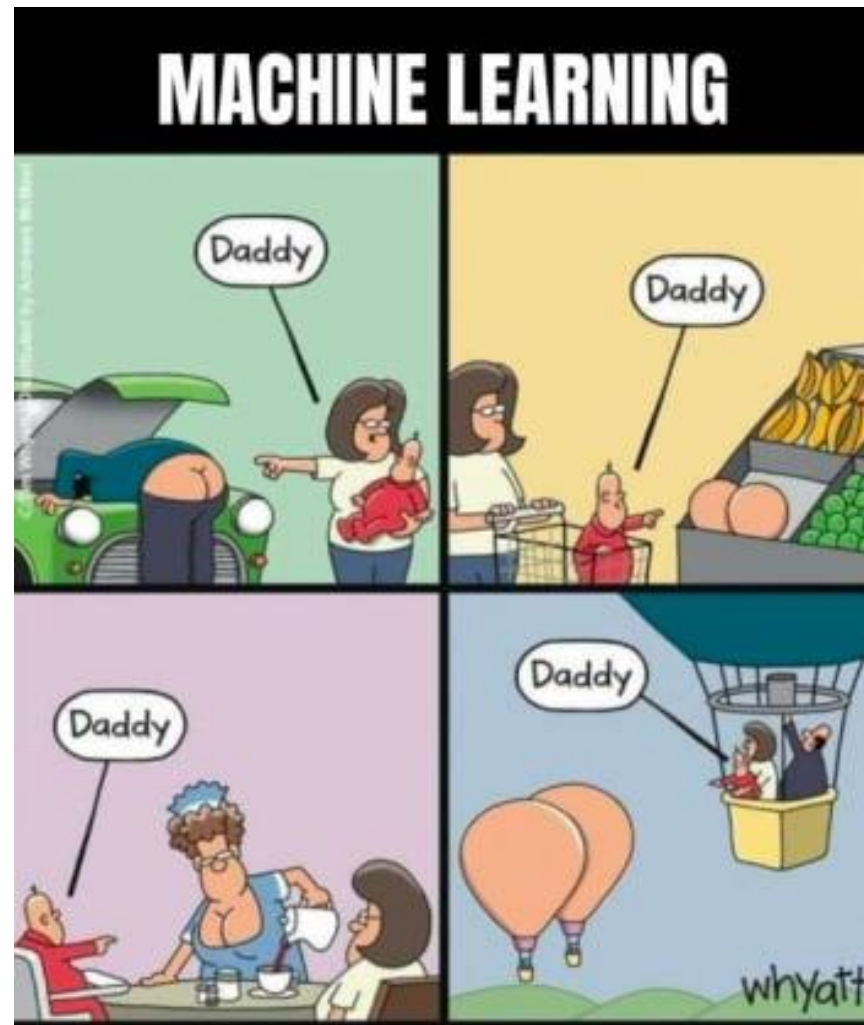
Não existe almoço grátis!

THERE'S
NO
FREE
LUNCH



- Teorema da economia, que se refere ao custo de oportunidade.
- Em ML: não existe um modelo de aprendizado que seja universalmente melhor em todos os casos.
 - Para melhorar o desempenho de um modelo em uma determinada tarefa, é necessário pagar um preço em termos de tempo, recursos e/ou complexidade.
- [“The Lack of A Priori Distinctions Between Learning Algorithms”](#).

Como podemos avaliar um modelo de AM?



O que são medidas de avaliação?

- Uma medida de avaliação quantifica a performance de um modelo preditivo, o que envolve comparar o resultado esperado e predito.
- Cada tarefa de aprendizado de máquina possui suas próprias medidas de avaliação.