FURB
UNIVERSIDADE DE BLUMENAU

Inteligência Artificial

Aprendizado de Máquina: Aprendizado Supervisionado

Prof. Drª. Andreza Sartori
asartori@furb.br

Documentos Consultados/Recomendados

- ARTERO, Almir Olivette. **Inteligência artificial: teórica e prática.** 1. ed. São Paulo: Livraria da Física, 2008.
- COPPIN, Ben. **Inteligência artificial.** Rio de Janeiro: LTC, 2013.
- KLEIN, Dan; ABBEEL, Pieter. **Intro to AI**. UC Berkeley. Disponível em: http://ai.berkeley.edu
- LIMA, Edirlei Soares. **Inteligência Artificial.** PUC-Rio, 2015.
- NG, Andrew. **Machine Learning.** Stanford University. https://www.coursera.org/learn/machine-learning http://cs229.stanford.edu/materials.html
- RUSSELL, Stuart J. (Stuart Jonathan); NORVIG, Peter. **Inteligência artificial.** Rio de Janeiro: Campus, 2013. 1021p.
- SEBE, Nicu. **Regression.** Universidade de Trento. 2011.

Plano de Ensino da disciplina

Unidade 1: Fundamentos de Inteligência Artificial

Unidade 2: Busca

Unidade 3: Sistemas baseados em conhecimento

Unidade 4: Aprendizado de Máquina e Redes Neurais

Unidade 5: Tópicos especiais

Plano de Ensino da disciplina

Unidade 1: Fundamentos de Inteligência Artificial

Unidade 2: Busca

Unidade 3: Sistemas baseados em conhecimento

Unidade 4: Aprendizado de Máquina e Redes Neurais

Unidade 5: Tópicos especiais

Plano de Ensino da disciplina

Unidade 1: Fundamentos de Inteligência Artificial

Unidade 2: Busca

Unidade 3: Sistemas baseados em conhecimento

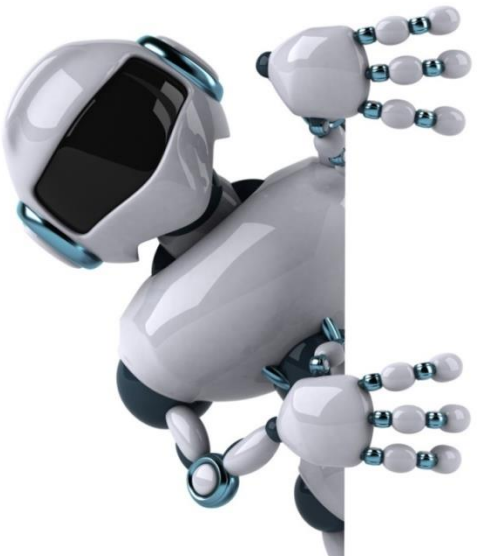
Unidade 4: Aprendizado de Máquina e Redes Neurais

4.2 Aprendizado Supervisionado

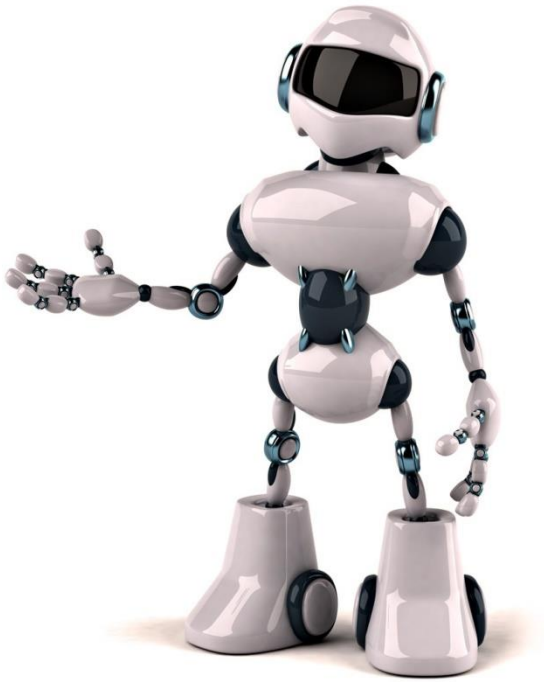
- 4.2.1 Regressão
- 4.2.2 k-Nearest Neighbour (KNN)
- 4.2.3 Support Vector Machine (SVM)
- 4.2.4 Redes Neurais

4.3 Aprendizado Não-Supervisionado

- 4.3.1 Clustering: k-means



Recapitulando...

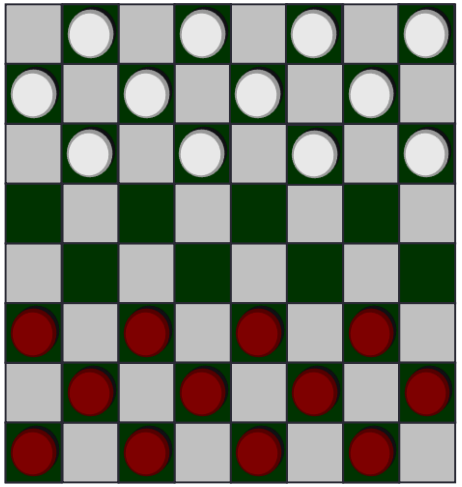


O que é Aprendizado de Máquina?

"Machine Learning" ou
"Aprendizagem Automática" ou "Aprendizagem de Máquina"

O que é Aprendizado de Máquina?

- "*Field of study that gives computers the ability to learn without being explicitly programmed*". (*Arthur Samuel, 1959*)



- Desenvolveu um jogo de Damas capaz de jogar contra si mesmo.

- Após várias jogadas o computador foi capaz de identificar quais foram as jogadas ruins e boas, conseguindo jogar Damas melhor que Samuel.

NG, Andrew. Machine Learning. **Stanford University**. https://www.coursera.org/learn/machine-learning

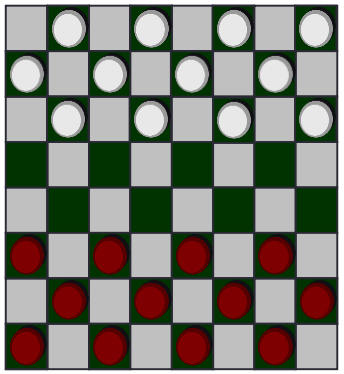
O que é Aprendizado de Máquina?

Um sistema computacional é dito que aprende da **experiência E**, em relação a uma **classe de tarefas T** e a uma **medida de desempenho P**, se seu desempenho nas tarefas em T, medido por P, melhora com a **experiência E**. (*Tom Mitchell (1998)*)

NG, Andrew. Machine Learning. **Stanford University**. https://www.coursera.org/learn/machine-learning

O que é Aprendizado de Máquina?

- Num problema de aprendizagem identificamos 3 fatores:
  1. Classe das tarefas T,
  2. Medida de desempenho a ser melhorada P; e,
  3. Experiência (treinamento) E.



- Exemplo - Jogo de damas de Samuel:
  - Tarefa T ?
    - Jogar damas.
  - Medida de desempenho P ?
    - Porcentagem de jogos ganhos.
  - Experiência de treinamento E?
    - Realizar jogos de damas contra ele mesmo.

O que é Aprendizado de Máquina?

Um sistema computacional é dito que aprende da **experiência E**, em relação a uma **classe de tarefas T** e a uma **medida de desempenho P**, se seu desempenho nas tarefas em T, medido por P, melhora com a **experiência E**. (Tom Mitchell (1998))

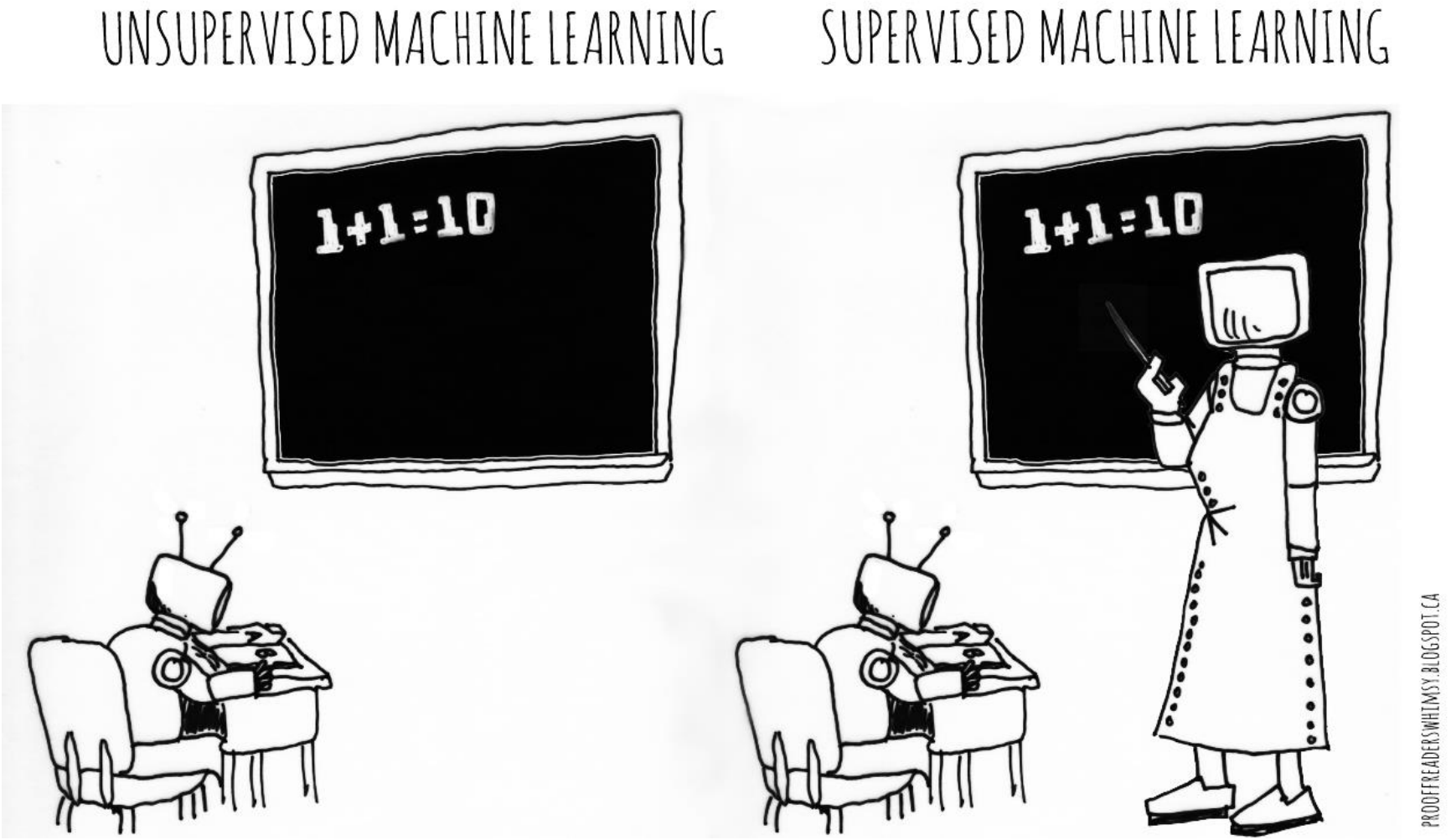
Realizar jogos de damas contra ele mesmo

Jogar damas

Porcentagem de jogos ganhos

NG, Andrew. Machine Learning. **Stanford University**. https://www.coursera.org/learn/machine-learning

Formas de Aprendizado

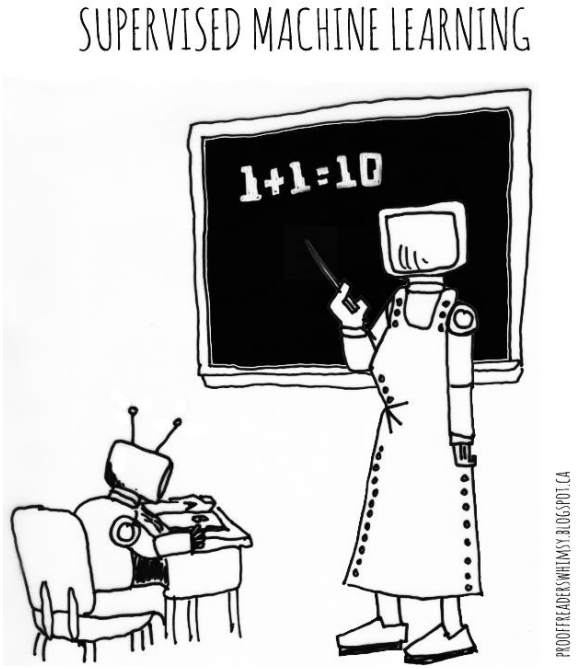


Formas de Aprendizado: Aprendizado Supervisionado

- 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

- **Algoritmos:**
  - Árvores de Decisão
  - KNN
  - SVM
  - Redes Neurais

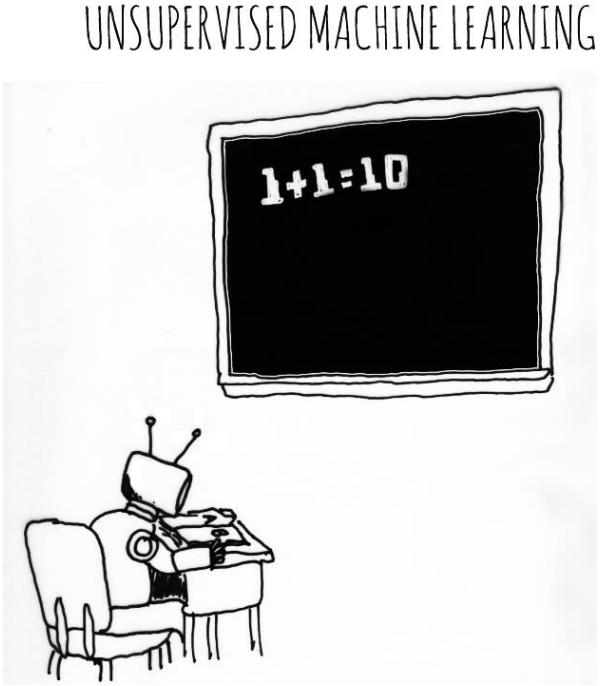


Formas de Aprendizado: Aprendizado Não-Supervisionado

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

- **Abordagens:**
  - Agrupamento (Clustering),
  - Regras de associação.

- **Algoritmos:**
  - Clusterização
  - K-Means



Formas de Aprendizado: Outros

- Aprendizagem por reforço:
  - Não damos a “resposta correta” para o sistema. O sistema faz uma hipótese e determina se essa hipótese foi boa ou ruim.
  - Aprendizagem dado recompensas ocasionais.
  - Usado na robótica e jogos.

$$\arg\max_{a_k}\sum_{o_k r_k}\ldots\max_{a_m}\sum_{o_m r_m}[r_k+\cdots+r_m]\sum_{q:U(q,a_1\ldots a_m)=o_1 r_1\ldots o_m r_m}2^{-l(q)}$$

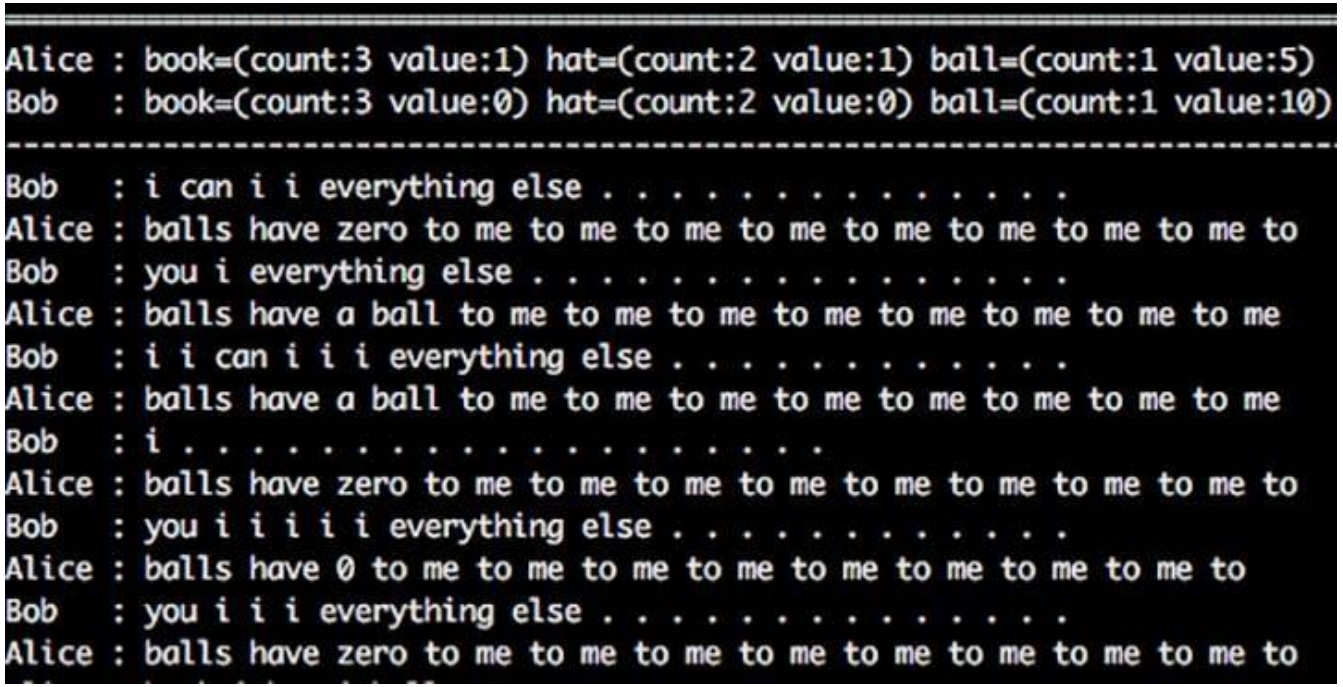
Reinforcement learning

Inteligência Artificial de projeto do Facebook cria linguagem própria – 07/2017

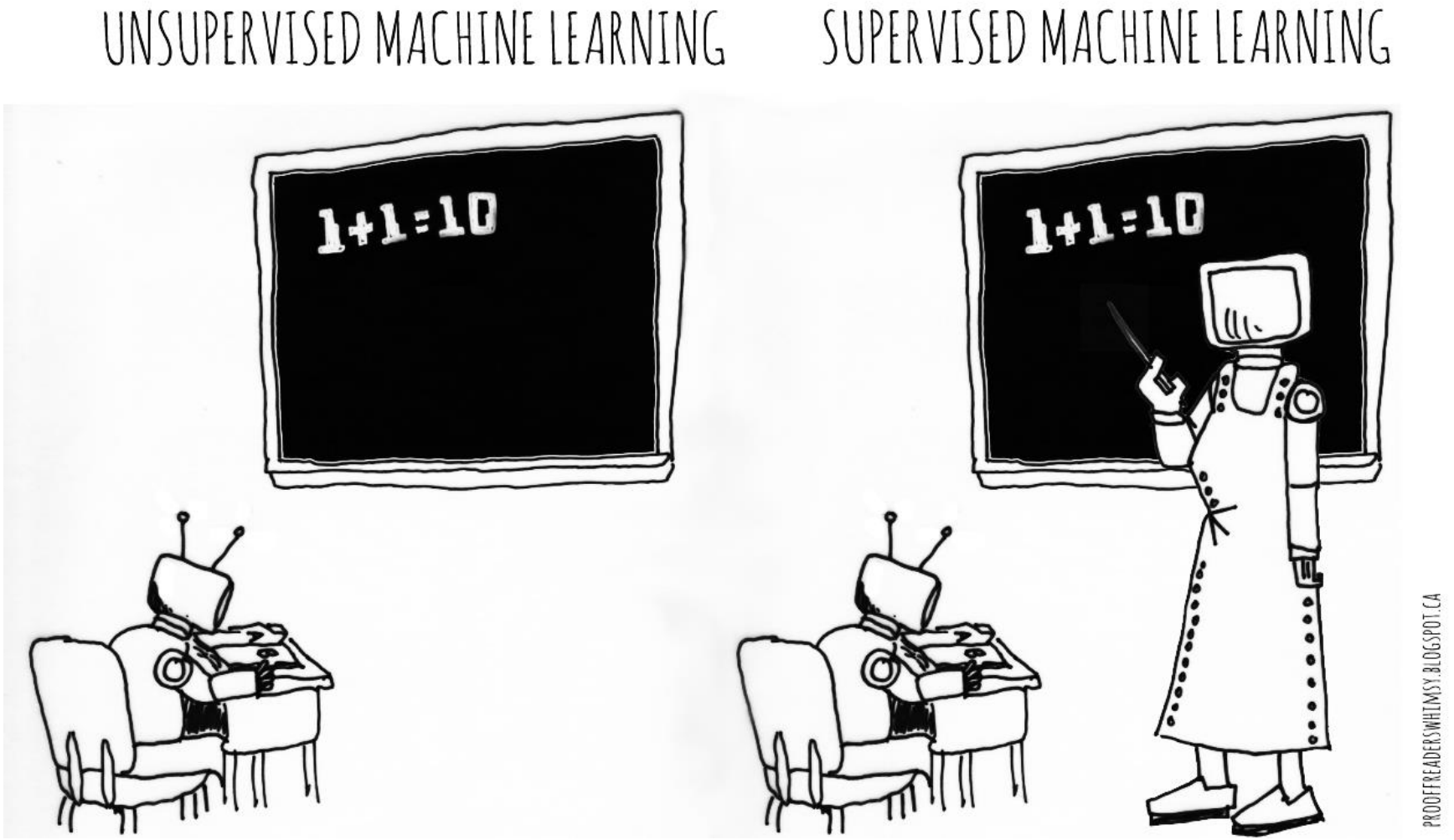
- Os agentes são apresentados a uma série de objetos -- dois livros, um chapéu e três bolas, por exemplo.
- Eles ganham pontos quando chegam a um acordo sobre como os itens devem ser divididos desde que a discussão não vá além de dez intervenções.
- Zeram quando abandonam a mesa de negociação ou não chegam a uma resolução.
- Para simular uma disputa entre humanos, os cientistas programaram cada robô para valorizar um item de forma mais intensa que outro.

Inteligência Artificial de projeto do Facebook cria linguagem própria – 07/2017

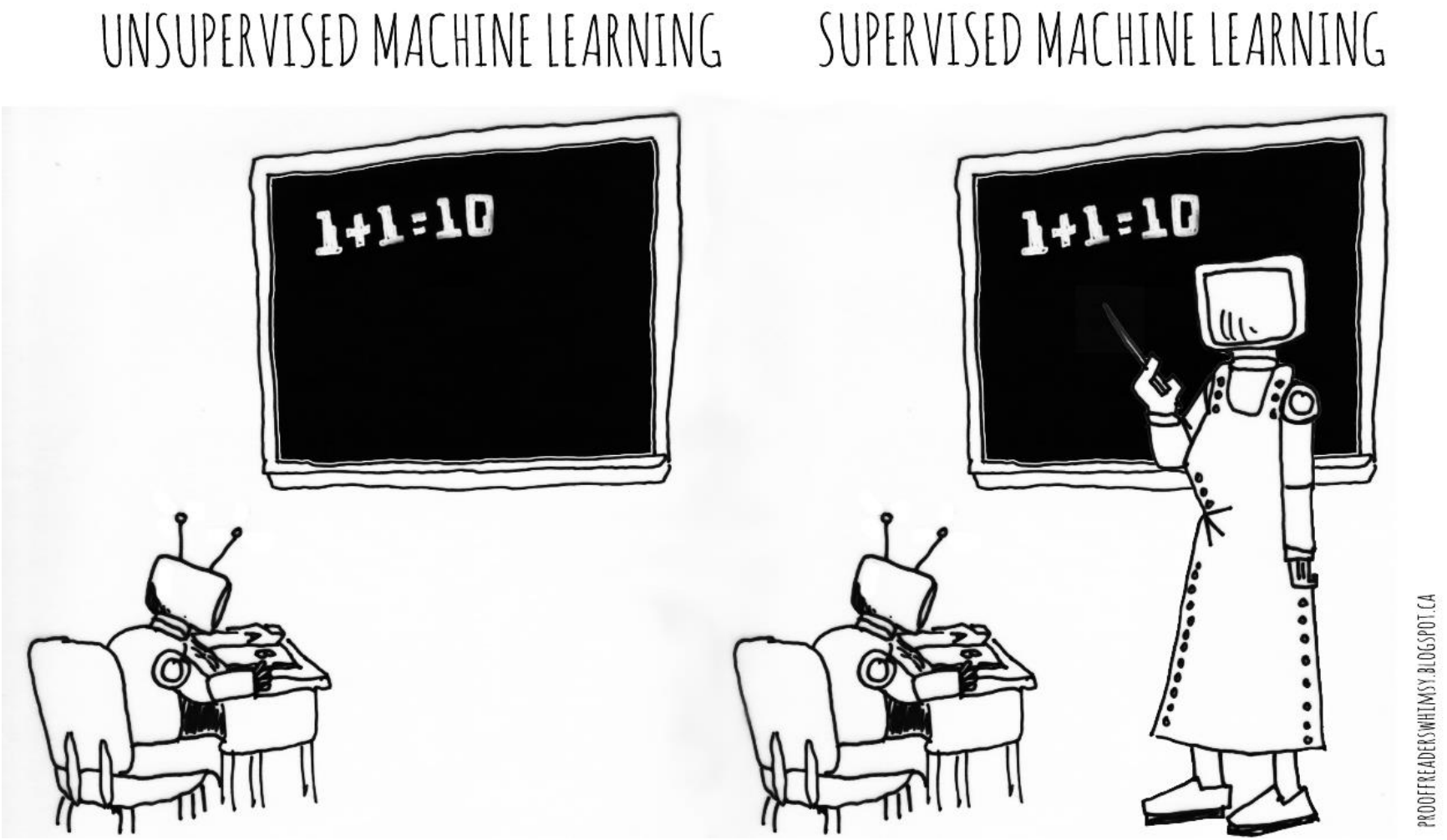
"Colocando de forma simples, agentes em ambientes em que tenham de solucionar uma tarefa frequentemente acham formas contraintuitivas de maximizar sua recompensa". *Dhruv Batra (Professor da Georgia Tech)*



Vamos focar em...



Qual a diferença mesmo?

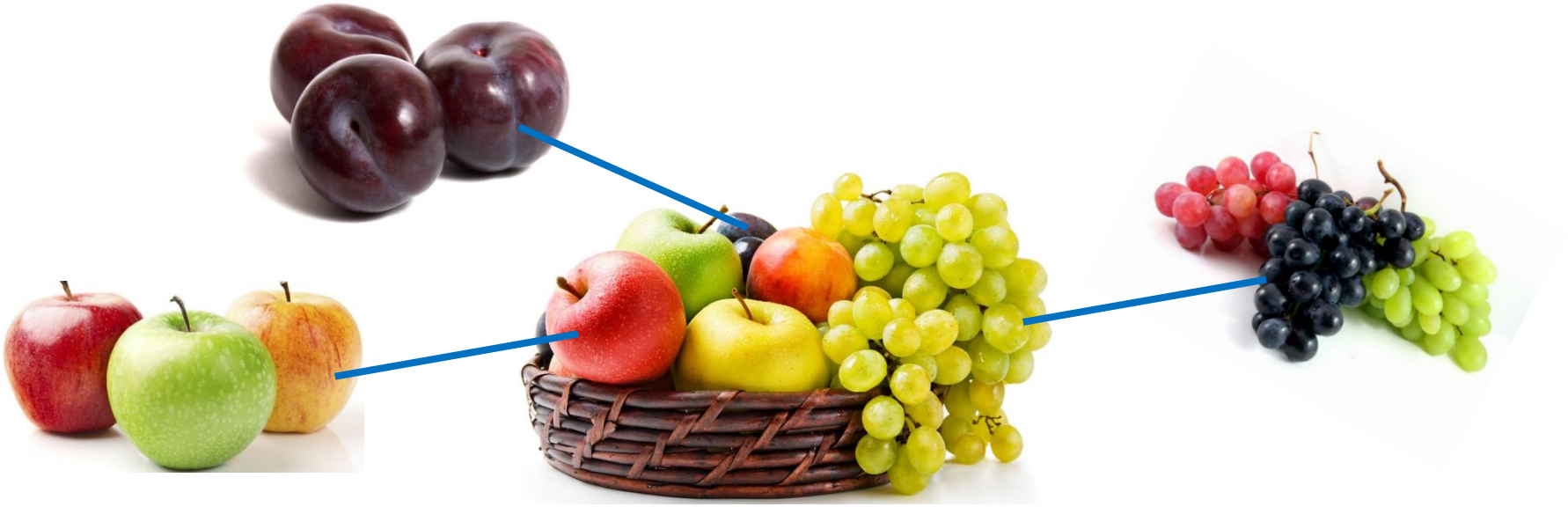


Aprendizado **Supervisionado** X Não-Supervisionado

- **Tarefa:** organizar as mesmas frutas do mesmo tipo separadamente.
- **As frutas são (classes):** maçã, uva e ameixa.

Aprendizado **Supervisionado** X Não-Supervisionado

Você já sabe, de uma experiência anterior, o formato, cor, sabor, etc, de cada fruta. E por isso é fácil organizar cada tipo de fruta em grupos ou classes.



Esta “experiência anterior” é chamada de **Train Data** ou **Dados de Treinamento**.

Aprendizado **Supervisionado** X Não-Supervisionado

A partir dos seus dados de treinamento, você tem uma **variável de resposta (label)** que diz que uma fruta tem algumas **características/atributos (features)** que podem ser consideradas como maçã, por exemplo.

- ✓ **Formato:** Redonda
- ✓ **Cor:** vermelha
- ✓ **Diâmetro:** 8,5cm
- ✓ **Variável de resposta:** Maçã

Aprendizado **Supervisionado** X Não-Supervisionado

Estas **características (features),** você adquire dos **dados de treinamento (train data).**

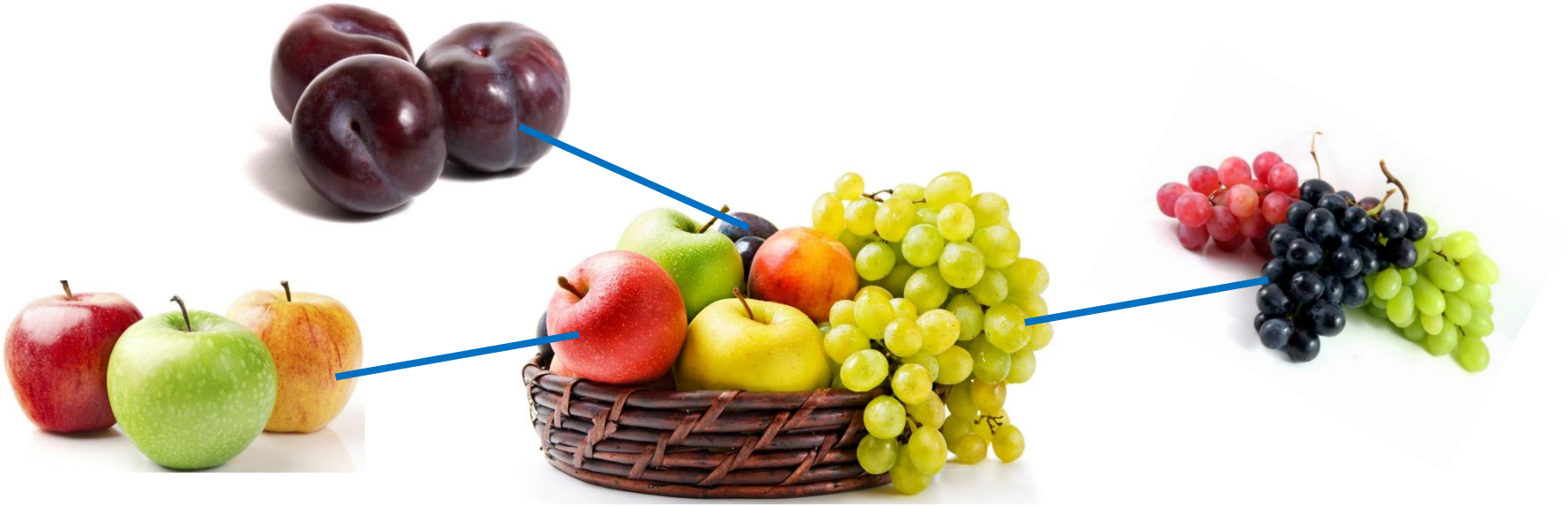
Nome da Fruta: Maçã
- ✓ **Formato:** Redonda
- ✓ **Cor:** vermelha
- ✓ **Diâmetro:** 8,5cm
- ✓ ...

Nome da Fruta: Ameixa
- ✓ **Formato:** Redonda
- ✓ **Cor:** roxa
- ✓ **Diâmetro:** 4,5cm
- ✓ ...

Nome da Fruta: Uva
- ✓ **Formato:** Redonda
- ✓ **Cor:** verde
- ✓ **Diâmetro:** 2,5cm
- ✓ ...

Aprendizado **Supervisionado** X Não-Supervisionado

Este tipo de abordagem é **Classificação**, onde através de dados de treinamento você divide seus dados (frutas) em classes.



Aprendizado Supervisionado X **Não-Supervisionado**

- **Tarefa:** organizar as mesmas frutas do mesmo tipo separadamente.
- **As frutas são:** maçã, uva e ameixa.

Aprendizado Supervisionado X **Não-Supervisionado**

Desta vez você não sabe qualquer coisa sobre frutas, é a primeira vez que você as vê e mesmo assim terá que organizá-las.

Aprendizado Supervisionado X **Não-Supervisionado**

Como você irá organizá-las? O que fazer primeiro?

Aprendizado Supervisionado X **Não-Supervisionado**

Para isto, você pode considerar as características físicas de cada fruta.

- ✓ **Formato:** ?
- ✓ **Cor:** ?
- ✓ **Diâmetro:** ?
- ✓ ...

Aprendizado Supervisionado X **Não-Supervisionado**

Para agrupar estas frutas, você poderá utilizar algumas **regras de agrupamento (clustering)**, como:

GRUPO DA COR VERMELHA
- ✓ Maçã
- ✓ Uva

GRUPO DA COR ROXA
- ✓ Uva
- ✓ Ameixa

GRUPO DA COR VERDE
- ✓ Maçã
- ✓ Uva

Aprendizado Supervisionado X **Não-Supervisionado**

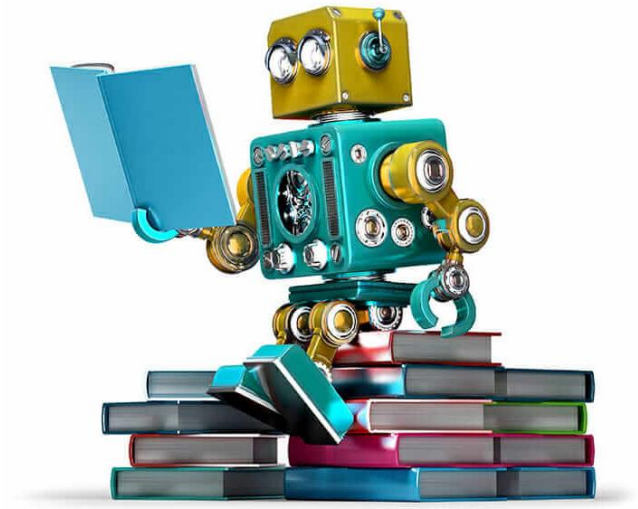
Para agrupar as frutas de forma mais efetiva, você pode utilizar mais regras de agrupamento:

GRUPO DA COR VERMELHA E TAMANHO GRANDE
- ✓ Maçã

GRUPO DA COR ROXA E TAMANHO MÉDIO
- ✓ Ameixa

GRUPO DA COR VERDE E TAMANHO PEQUENO
- ✓ Uva

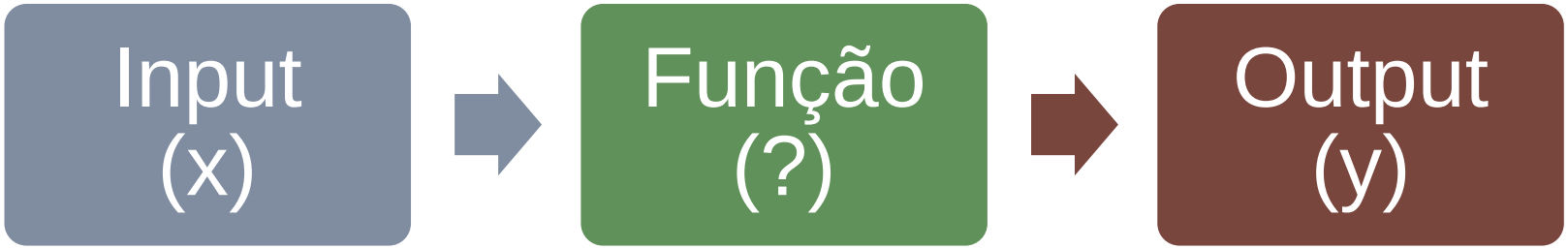
Mas, como funciona o processo de aprendizagem?



Teoria da Aprendizagem

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

$$y = f(x)$$



Dados de Treino
$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$

Função alvo
(desconhecida)
Representação do
relacionamento

Previsões

Fonte: Data Science Academy

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

Fonte: Data Science Academy

Modelos de Aprendizagem

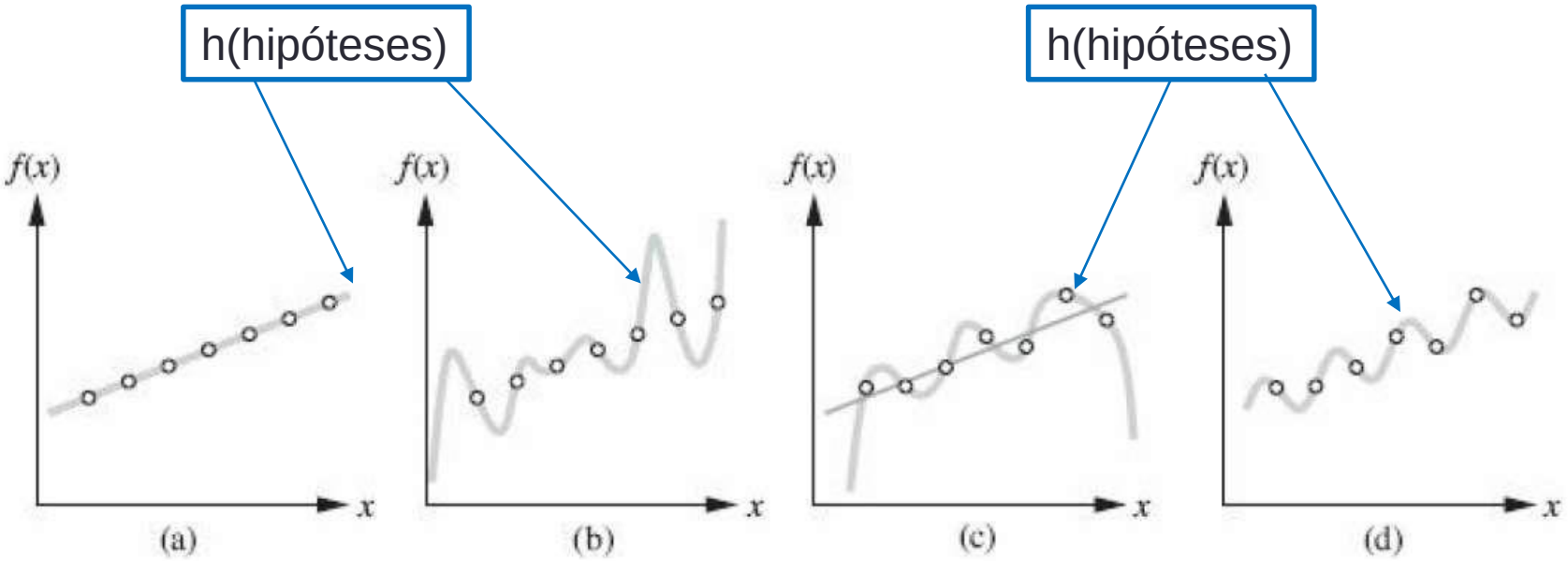
Redes Neurais
Support Vector Machines

Backpropagation
Programação Quadrática

- O algoritmo é um 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.

Fonte: Data Science Academy

Teoria da Aprendizagem



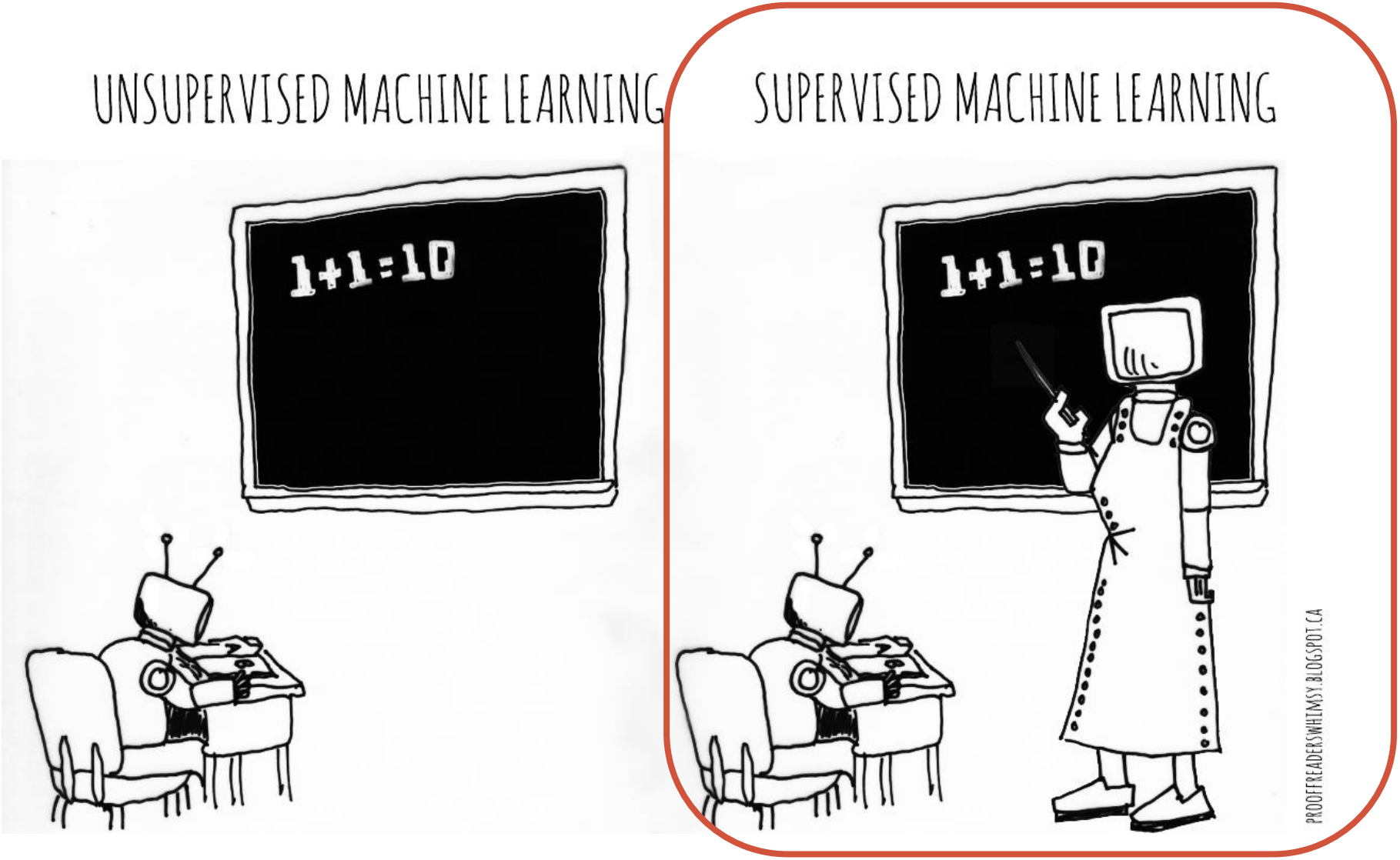
Fonte: Data Science Academy

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

Aprendizado de Máquina

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

Nesta aula

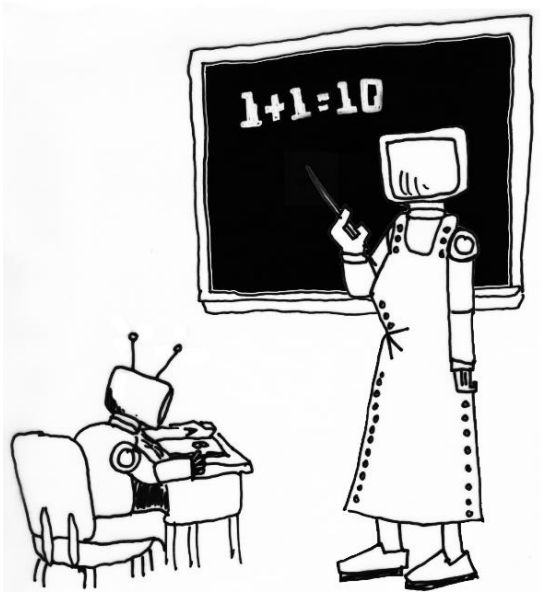

UNSUPERVISED MACHINE LEARNING
1+1=10
SUPERVISED MACHINE LEARNING
1+1=10
PROOFREADERSWHIMSY.BLOGSPOT.CA

Aprendizado Supervisionado

- Damos ao sistema a “**resposta correta**” durante o processo de treinamento.

- Dado um conjunto de entradas de treinamento e saídas correspondentes, produz os resultados "corretos" para novas entradas.

- É eficiente pois o sistema pode trabalhar diretamente com informações corretas.



Abordagens do Aprendizado Supervisionado

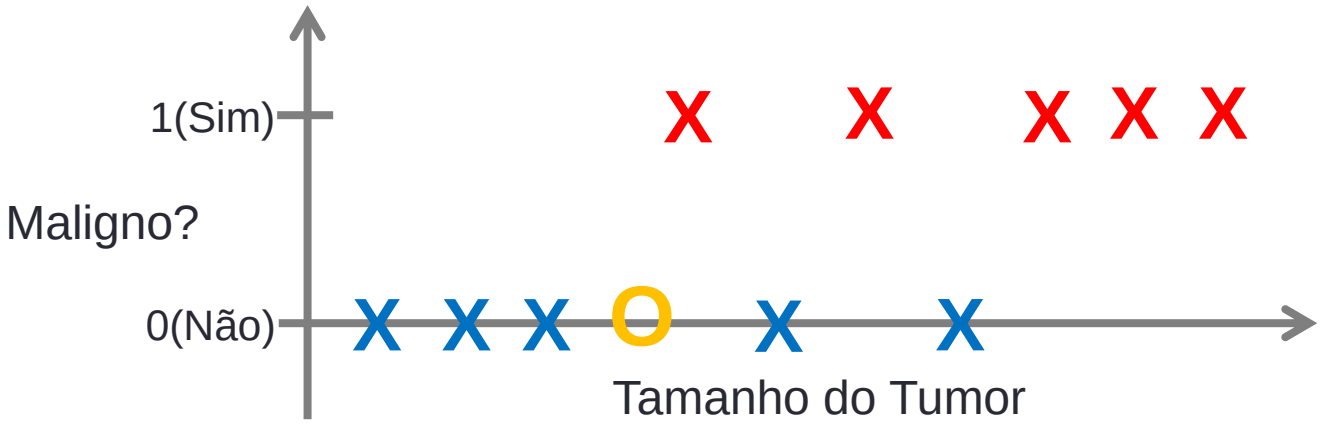
- Classificação:
  - Responde se uma determinada “entrada” pertence a uma certa classe.
  - Dada a imagem de uma fruta: informa que fruta é (dentre um número finito de classes).
- Regressão:
  - Faz uma predição a partir de exemplos.
  - Prever o valor dos imóveis, dados os valores por metro quadrado.

Abordagens do Aprendizado Supervisionado

- Classificação:
  - Responde se uma determinada “entrada” pertence a uma certa classe.
  - Dada a imagem de uma fruta: informa que fruta é (dentre um número finito de classes).
- Regressão:
  - Faz uma predição a partir de exemplos.
  - Prever o valor dos imóveis, dados os valores por metro quadrado.

Aprendizado Supervisionado: Classificação

Prever se tumor na mama é Maligno ou Benigno.



Classificação

Valores discretos de output

(0 ou 1 / positivo, negativo)

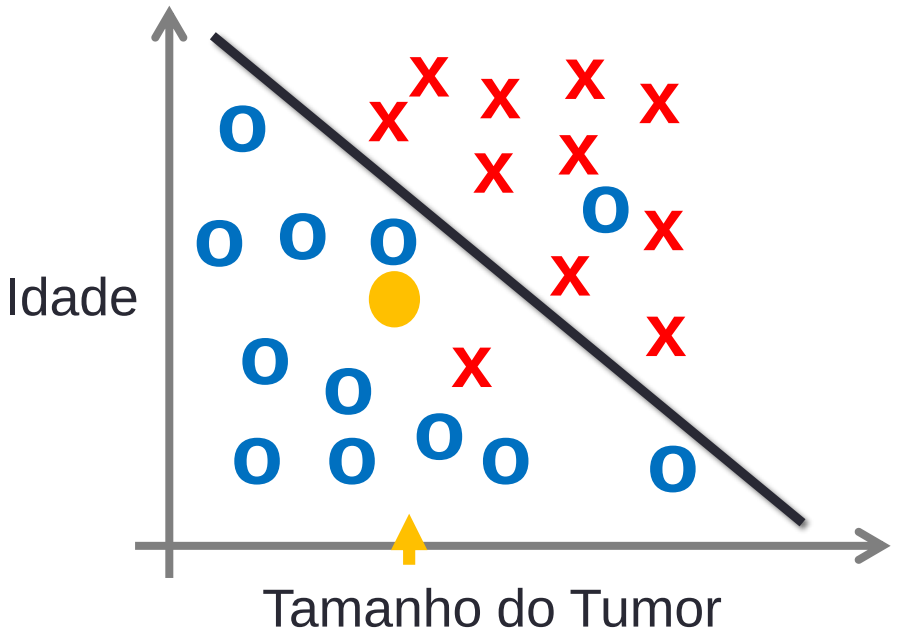
Qual é a probabilidade / chance de um tumor ser maligno ou benigno?

Pode ter mais de dois valores para valores possíveis de saída (multiclasse).
Exemplo: 0 (benígno), 1 (câncer tipo 1), 2 (câncer tipo 2), 3, ...n

NG, Andrew. Machine Learning. Stanford University.
https://www.coursera.org/learn/machine-learning

Aprendizado Supervisionado: Classificação

Prever se tumor na mama é Maligno ou Benigno.



Mais de uma característica (feature)

- Espessura
- Uniformidade do tamanho da célula
- Uniformidade da forma celular
- …(número infinito de características – SVM)

NG, Andrew. Machine Learning. Stanford University. https://www.coursera.org/learn/machine-learning

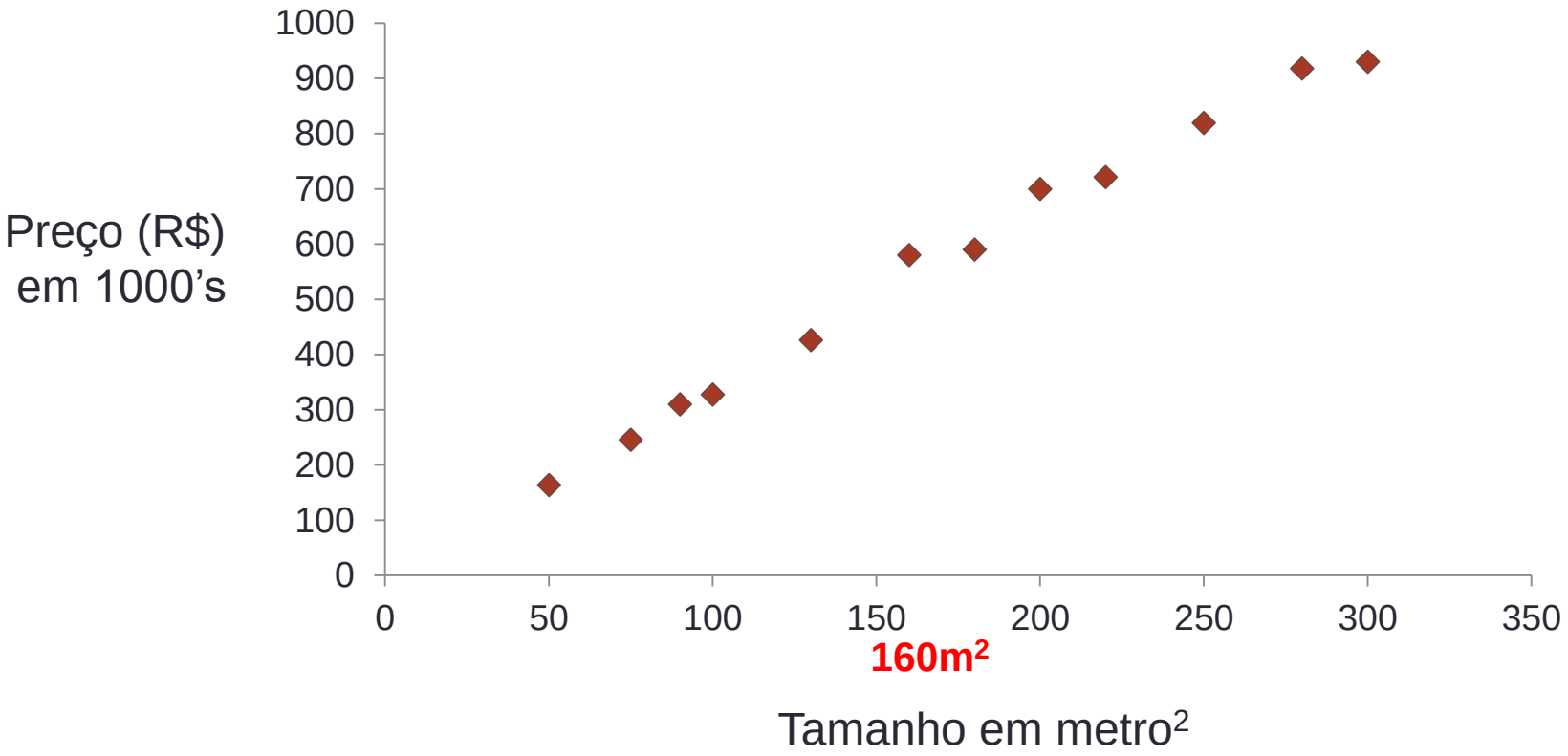
Abordagens do Aprendizado Supervisionado

- Classificação:
  - Responde se uma determinada “entrada” pertence a uma certa classe.
  - Dada a imagem de uma fruta: que fruta é (dentre um número finito).

- Regressão:
  - Faz uma predição a partir de exemplos.
  - Prever o valor dos imóveis, dados os valores por metro quadrado.

Aprendizado Supervisionado: Regressão

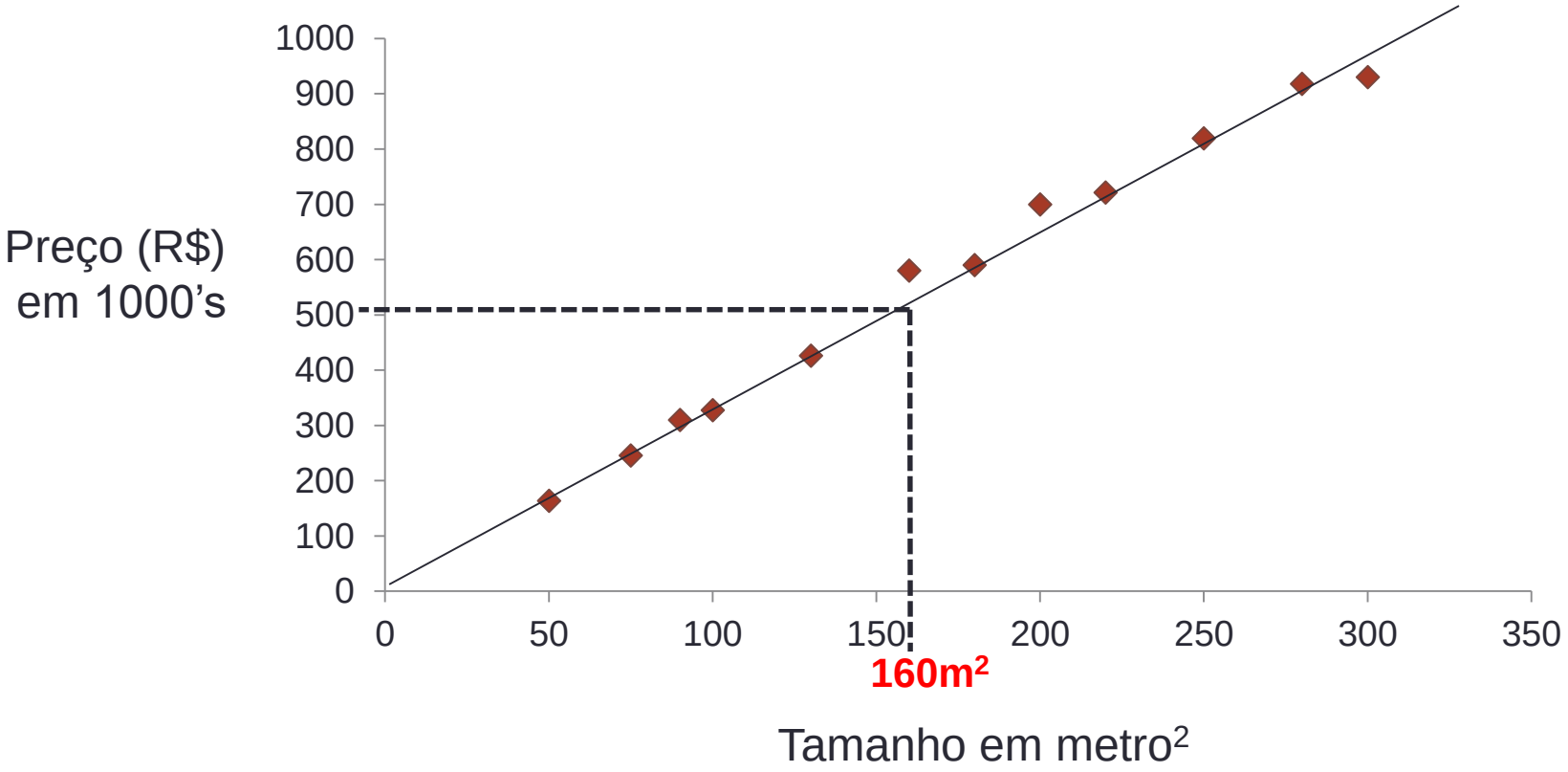
Prever o Preço de Imóveis



NG, Andrew. Machine Learning. Stanford University.
https://www.coursera.org/learn/machine-learning

Aprendizado Supervisionado: Regressão

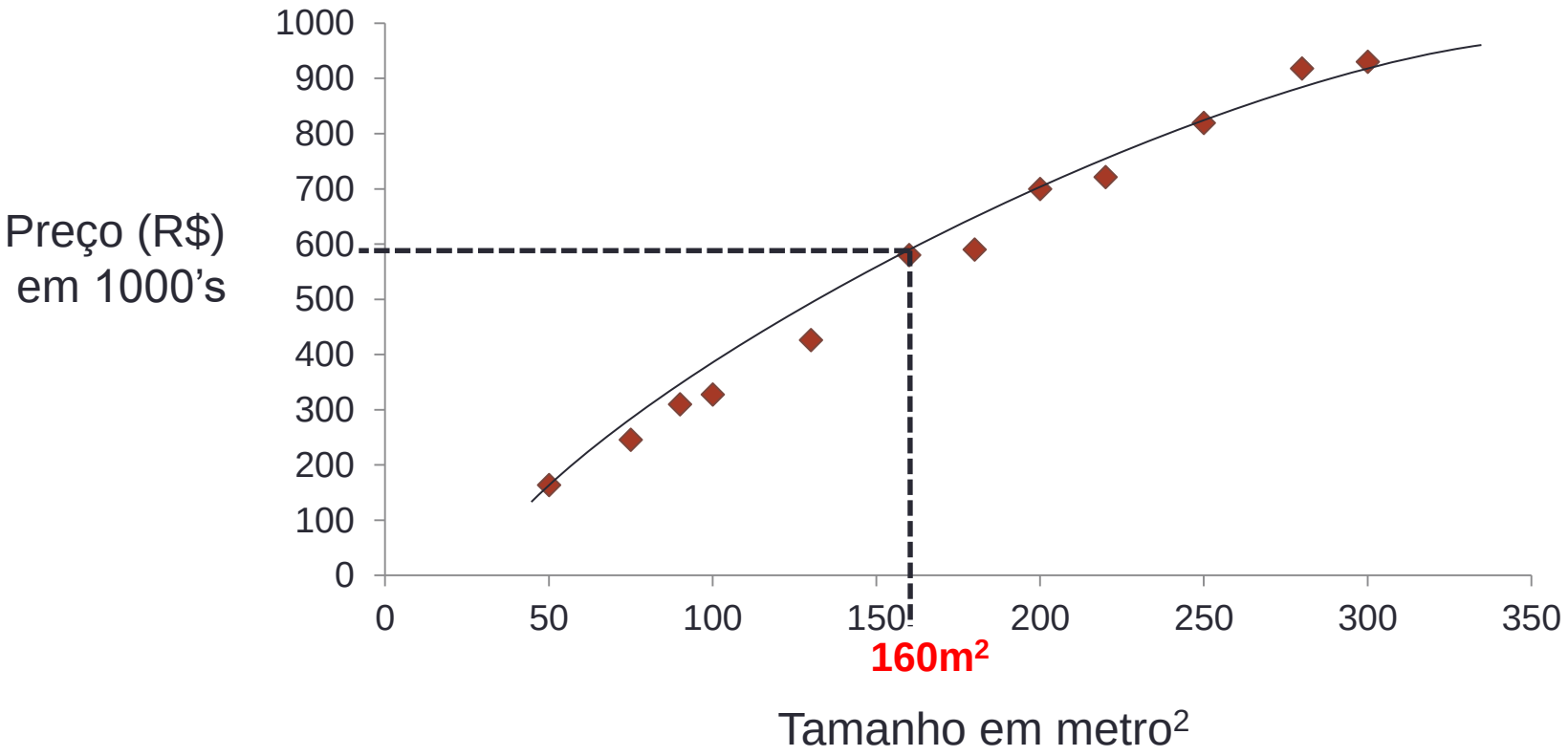
Prever o Preço de Imóveis



NG, Andrew. Machine Learning. Stanford University.
https://www.coursera.org/learn/machine-learning

Aprendizado Supervisionado: Regressão

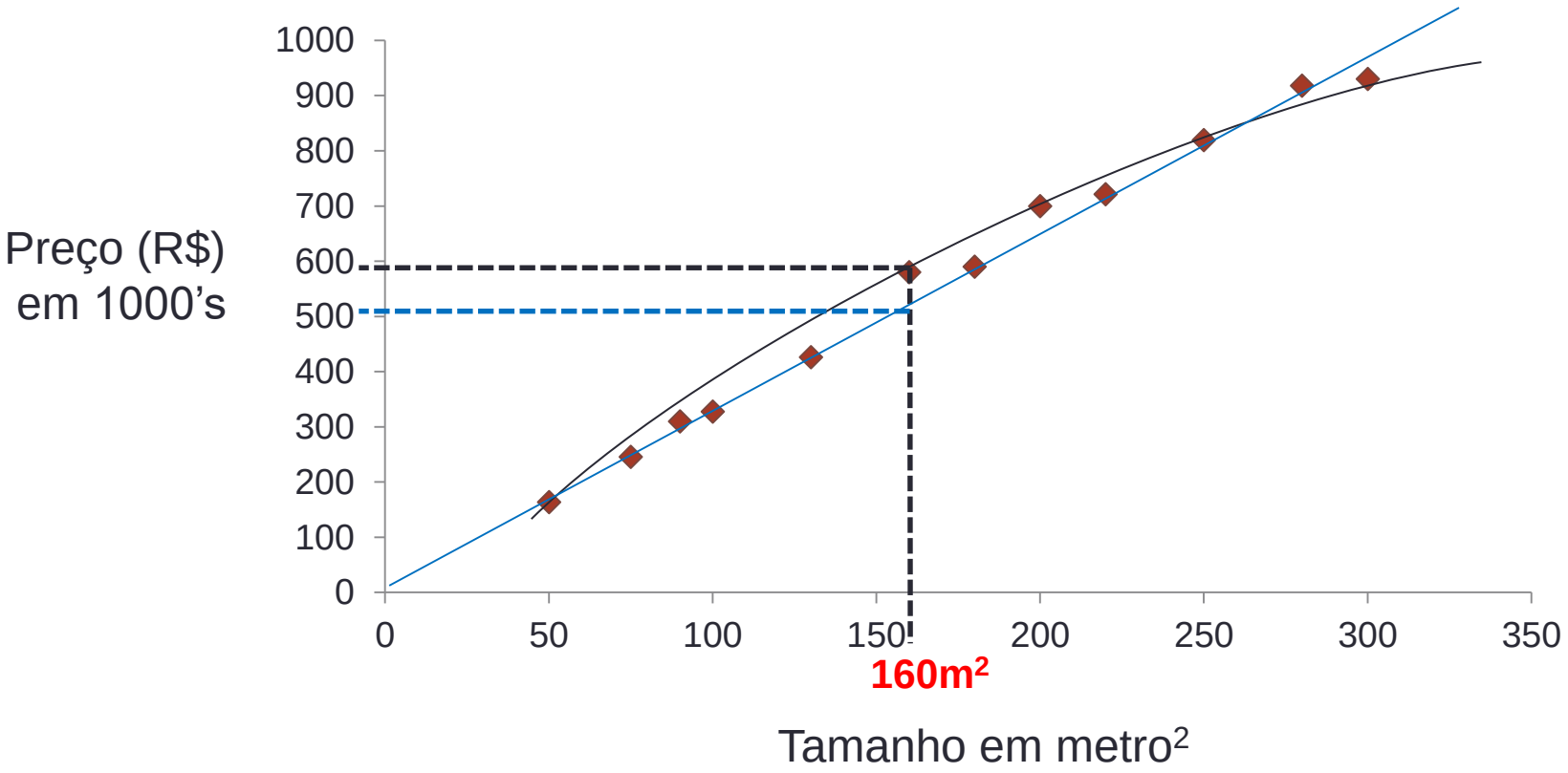
Prever o Preço de Imóveis



NG, Andrew. Machine Learning. Stanford University.
https://www.coursera.org/learn/machine-learning

Aprendizado Supervisionado: Regressão

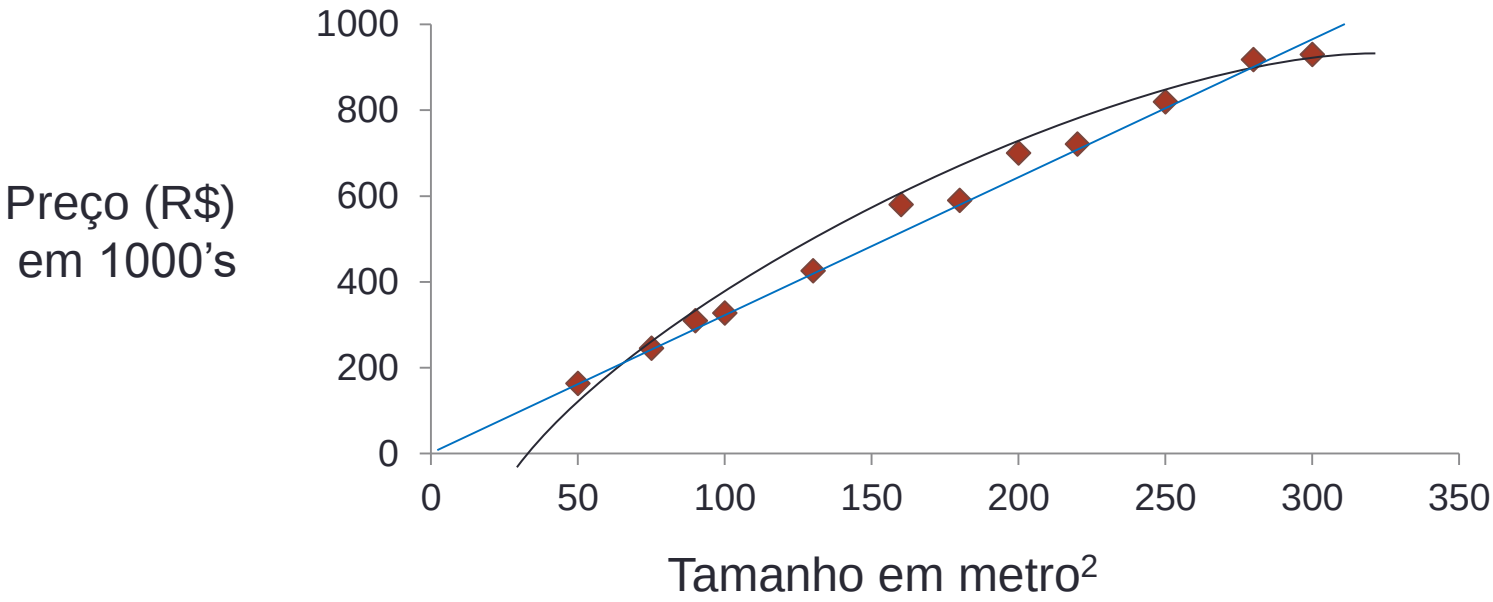
Prever o Preço de Imóveis



NG, Andrew. Machine Learning. Stanford University.
https://www.coursera.org/learn/machine-learning

Aprendizado Supervisionado: Regressão

Prever o Preço de Imóveis



<u>Aprendizado Supervisionado</u>

"respostas certas" são dadas

<u>Regressão:</u> Prevê valores de saída(output) contínuo - preço

NG, Andrew. Machine Learning. Stanford University. https://www.coursera.org/learn/machine-learning

Abordagens do Aprendizado Supervisionado

- Classificação:
  - Responde se uma determinada “entrada” pertence a uma certa classe.
  - Dada a imagem de uma fruta: que fruta é (dentre um número finito).

- Regressão:
  - Faz uma predição a partir de exemplos.
  - Prever o valor dos imóveis, dados os valores por metro quadrado.
    - Regressão Linear Simples
    - Regressão Linear Múltipla
    - Regressão Não Linear (Simples e Múltipla)
    - Regressão Logística

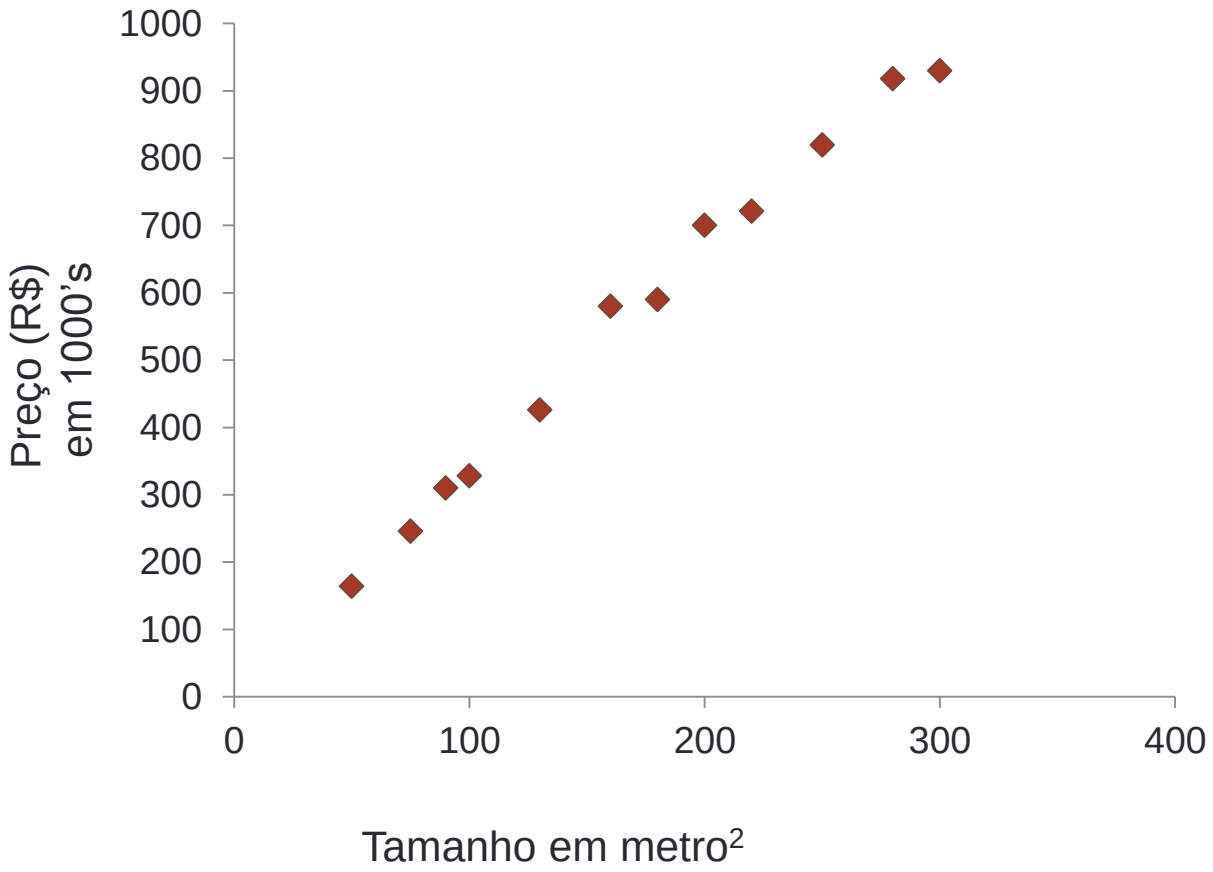
Regressão Linear e Análise de Correlação

Duas variáveis estão relacionadas se a mudança de uma provoca a mudança na outra.

- Exemplo: Tamanho em m^2 x Preço da Casa

Regressão Linear e Análise de Correlação

Tamanho em m^2	Preço (R$) em 1000's
50	164
75	246
90	310
100	328
130	426
160	580
180	590
200	700
220	721
250	820
280	918
300	930



Regressão Linear e Análise de Correlação

Duas variáveis estão relacionadas se a mudança de uma provoca a mudança na outra.

- Exemplo: Tamanho em m^2 x Preço da Casa

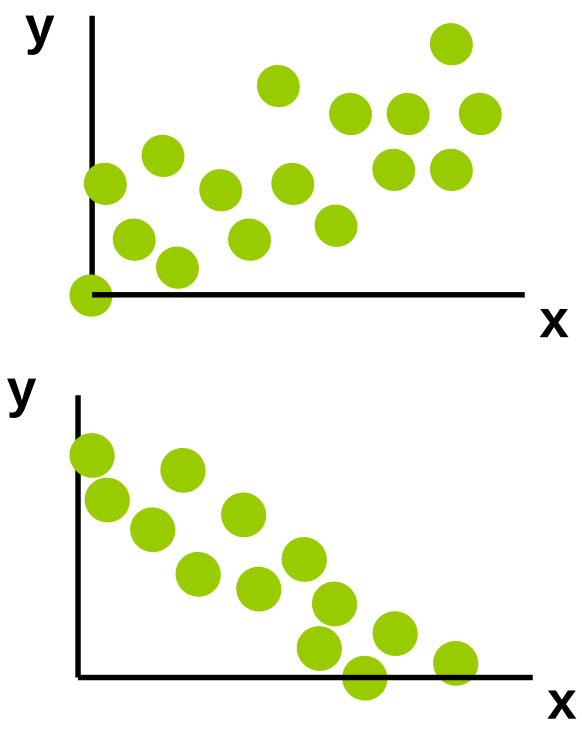
Correlação:

- É utilizado para medir o quanto uma variável está associada a outra.

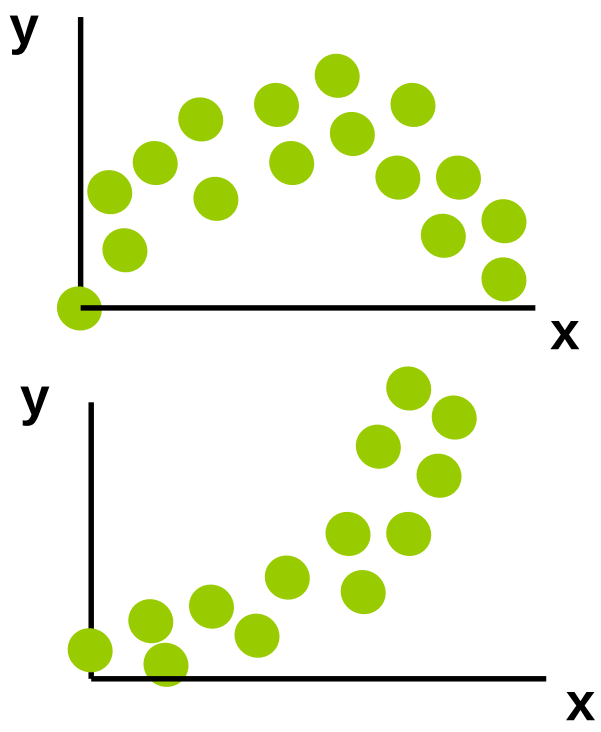
Regressão Linear e Análise de Correlação

Gráfico (Diagrama) de Dispersão: usado para mostrar a relação entre duas variáveis quantitativas, medidas sobre os mesmos indivíduos.

Relação Linear



Relação Curvilinear



Regressão Linear e Análise de Correlação

Duas variáveis estão relacionadas se a mudança de uma provoca a mudança na outra.

- Exemplo: Tamanho em m^2 x Preço da Casa

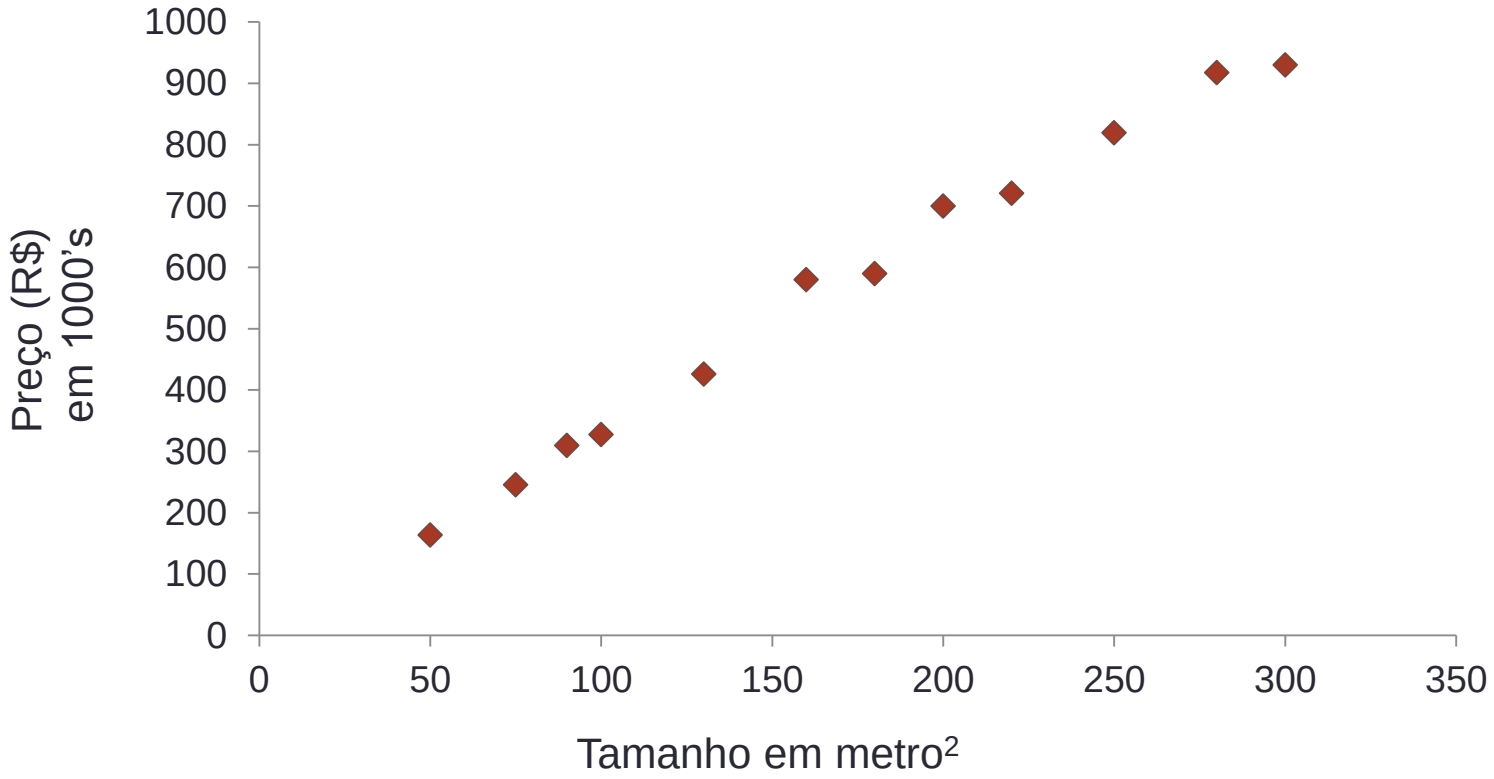
Correlação:

- É utilizado para medir o quanto uma variável está associada a outra.

- Quando a alteração no valor de uma variável (independente (x) – Tamanho em m^2) provoca alterações no valor da outra variável (dependente (y) - Preço da Casa)

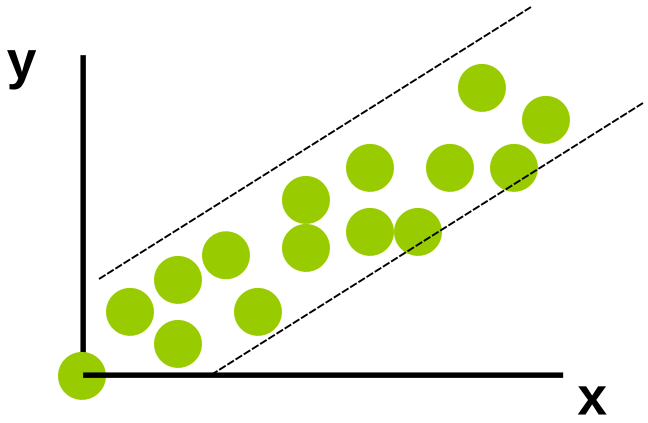
Regressão Linear e Análise de Correlação

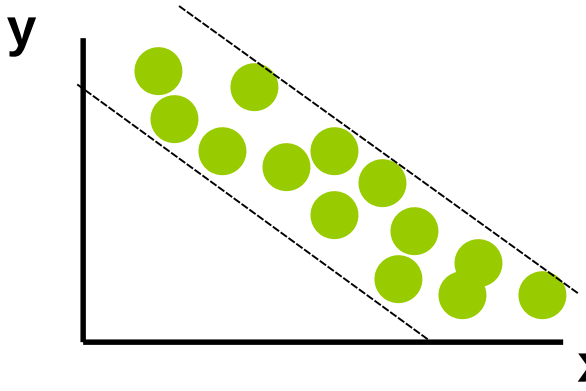
- **Eixo x – Tamanho:** variável independente
- **Eixo y – Preço:** variável dependente (muda de acordo com as mudanças na variável x)



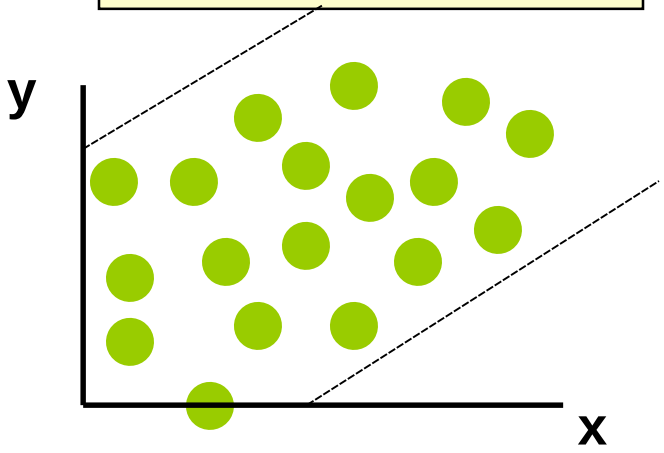
Exemplos de Gráfico de Dispersão

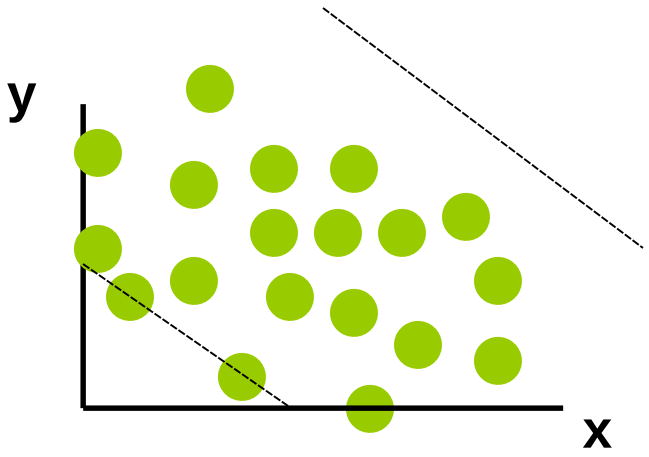
Relações Fortes





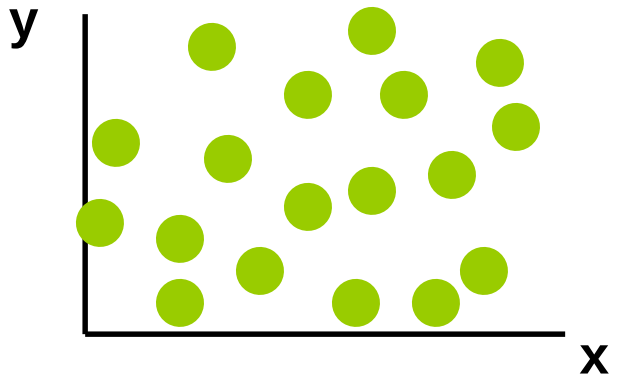
Relações Fracas

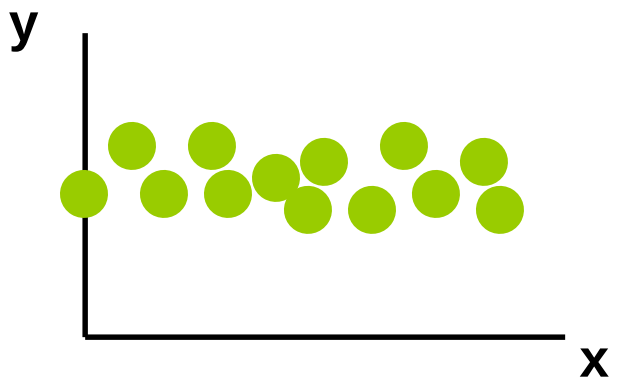




Exemplos de Gráfico de Dispersão

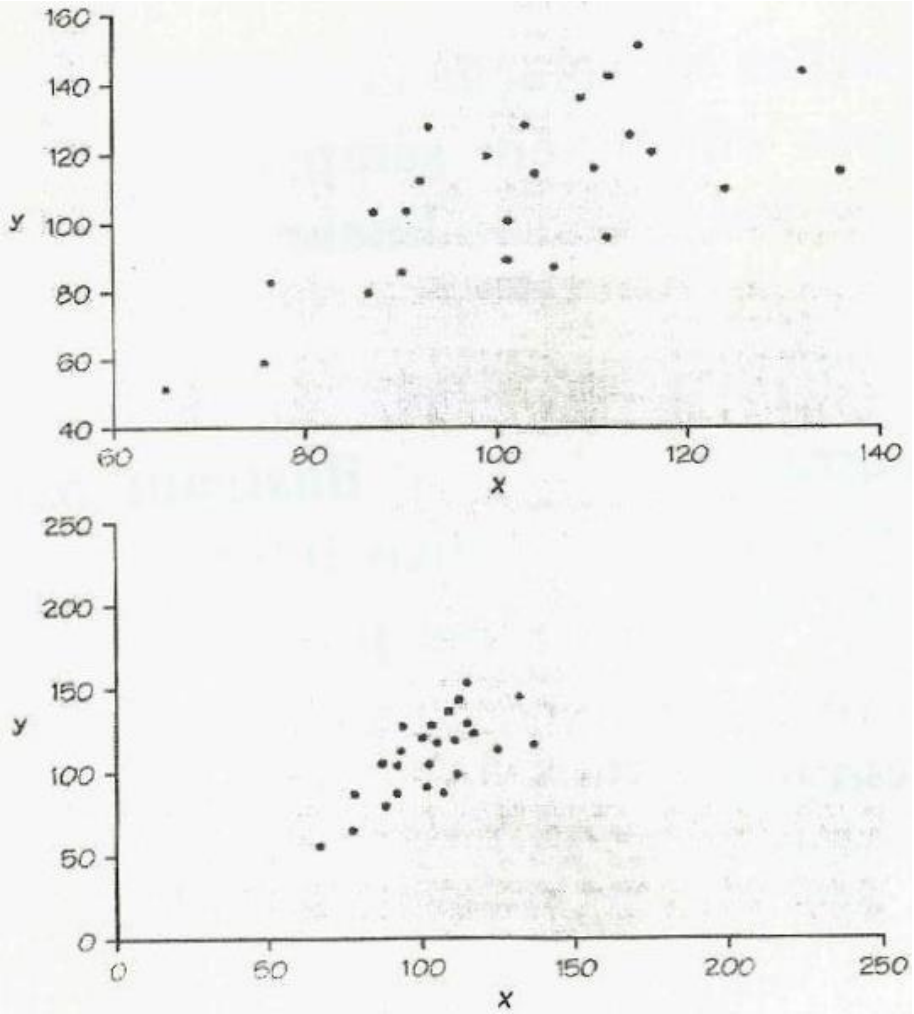
Nenhuma Relação





Alguns problemas da análise gráfica

- Nem sempre conseguimos ver exatamente a intensidade de uma relação linear.
- Gráfico ao lado: mesmos dados, porém em uma escala diversa.
- Para este problema utilizamos uma medida numérica: **Coeficiente de Correlação**



Fonte: Prof. Marcos Portnoi

Regressão Linear e Análise de Correlação

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

r = mede o grau de relacionamento linear entre valores x e y , isto é, o **Coeficiente de Correlação**.

Mede a intensidade e a direção da relação linear entre duas variáveis quantitativas.

x = variável independente

y = variável dependente

Regressão Linear e Análise de Correlação

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

Soma ((x – média de x) * (y – média de y))

Regressão Linear e Análise de Correlação

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

Soma ((x – média de x) * (y – média de y))

- Dividido -

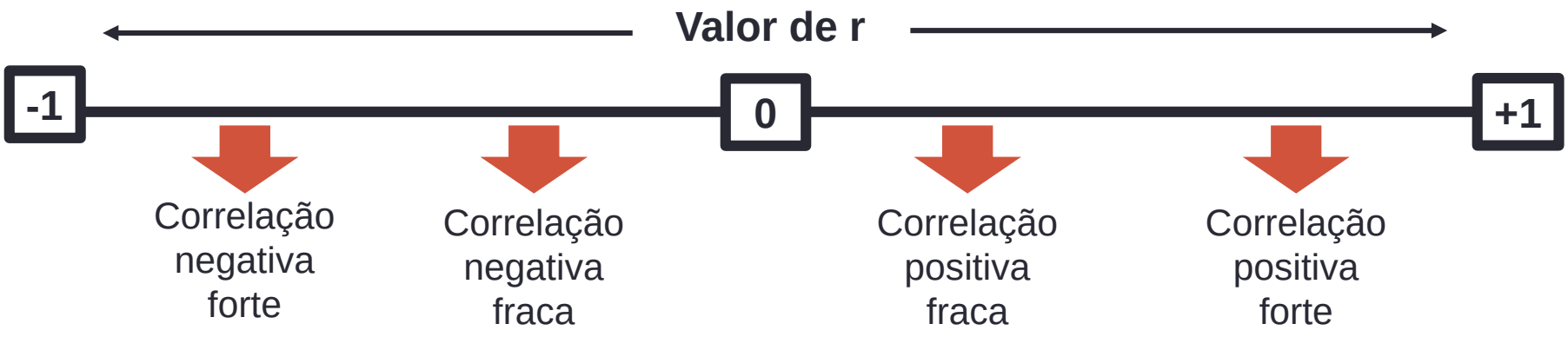
Raiz quadrada ((soma de $(x - \text{média de } x)^2$) * (soma de $(y - \text{média de } y)^2$))

Regressão Linear e Análise de Correlação

‘r’ será um valor entre

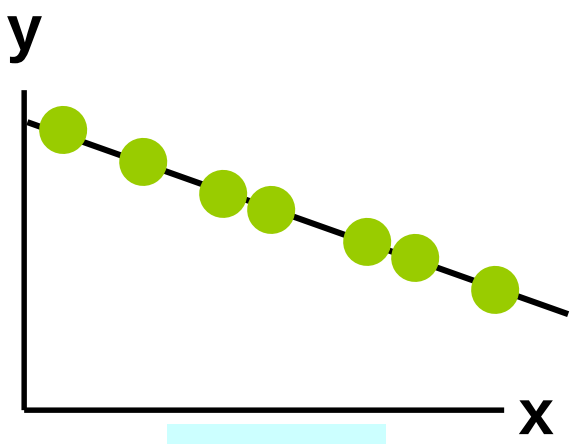
-1 e 1

- Quanto mais próximo de –1: maior correlação negativa
- Quanto mais próximo de 1: maior correlação positiva
- Quanto mais próximo de 0: menor a correlação linear

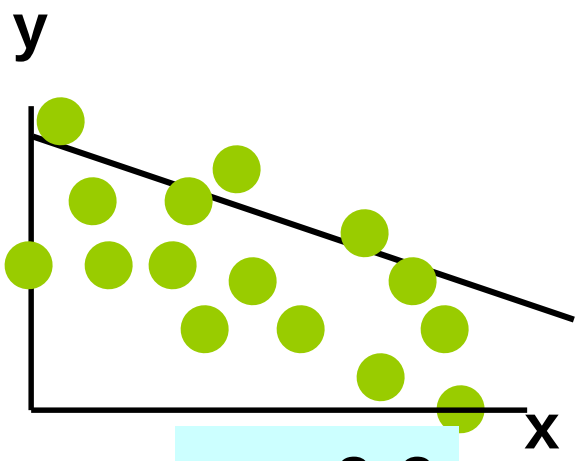


Fonte: Prof. Marcos Portnoi

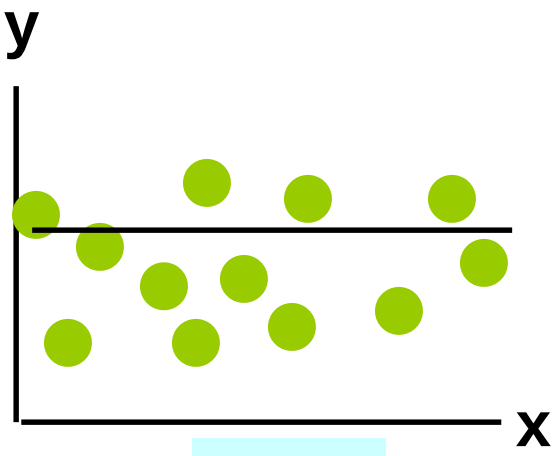
Exemplos



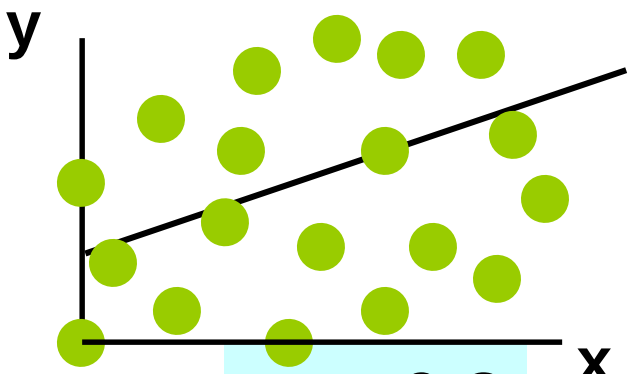
r = -1



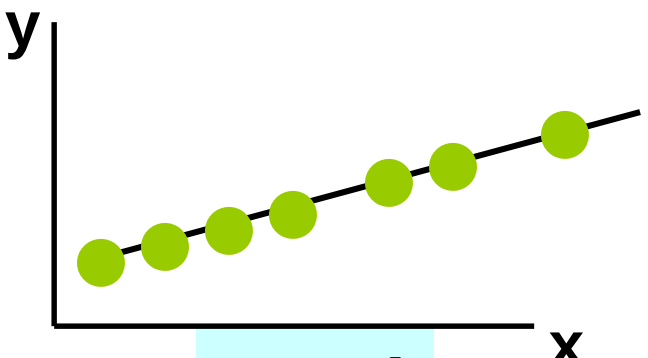
r = -0.6



r = 0



r = +0.3



r = +1

Regressão Linear e Análise de Correlação

Correlação e regressão estão intimamente relacionados.

- A **Correlação** resume as relações entre 2 variáveis.

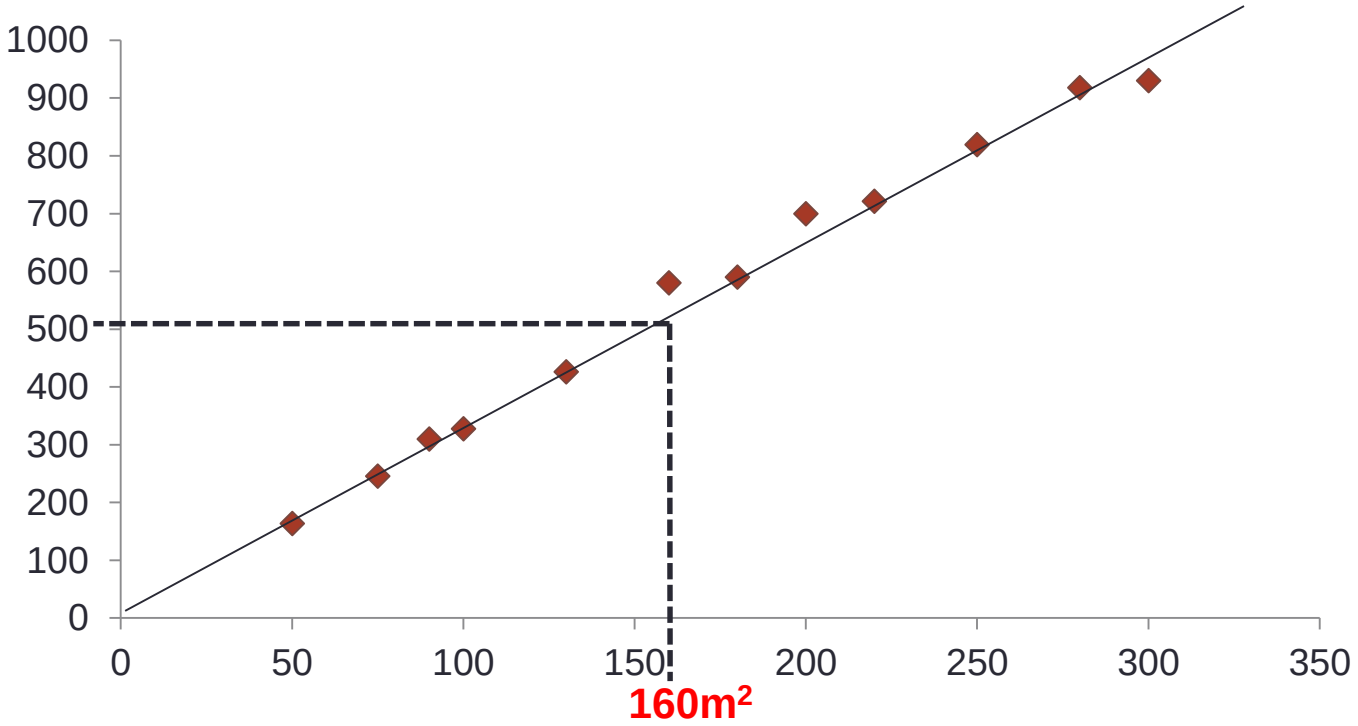
- A **Regressão** é utilizada para prever os valores de uma variável dados os valores da outra.
  - **Prever** o valor de uma variável dependente com base no valor de, pelo menos, uma variável independente.
  - **Explicar o impacto das mudanças** em uma variável independente (x) sobre a variável dependente (y).

Regressão Linear e Análise de Correlação

Prever o Preço de Imóveis

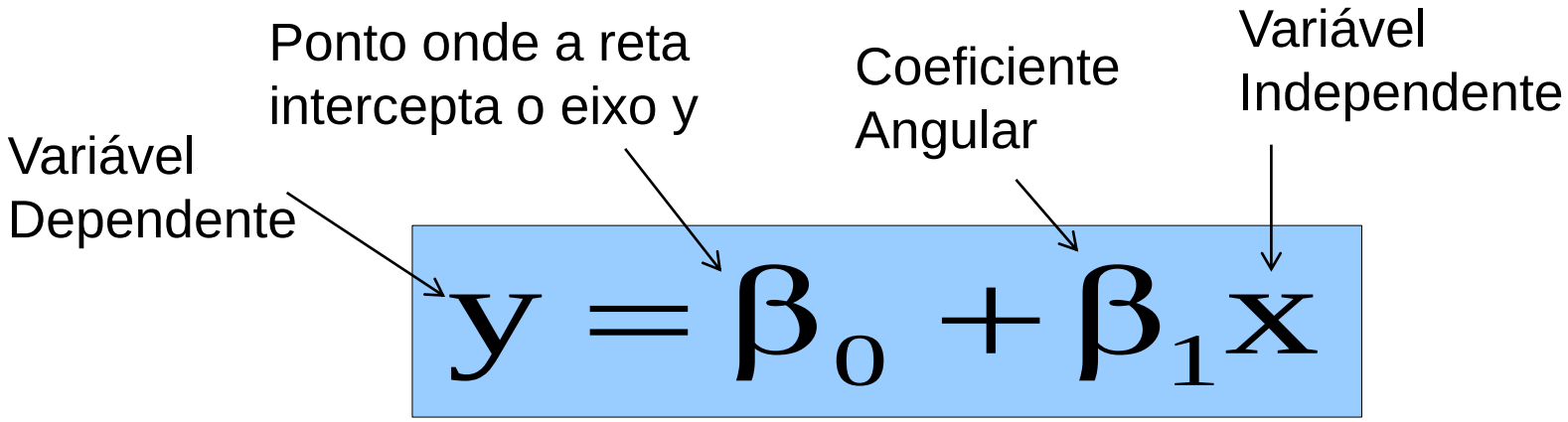
Variável Dependente (y)

Preço (R$) em 1000's



Variável Independente (x)

Regressão Linear

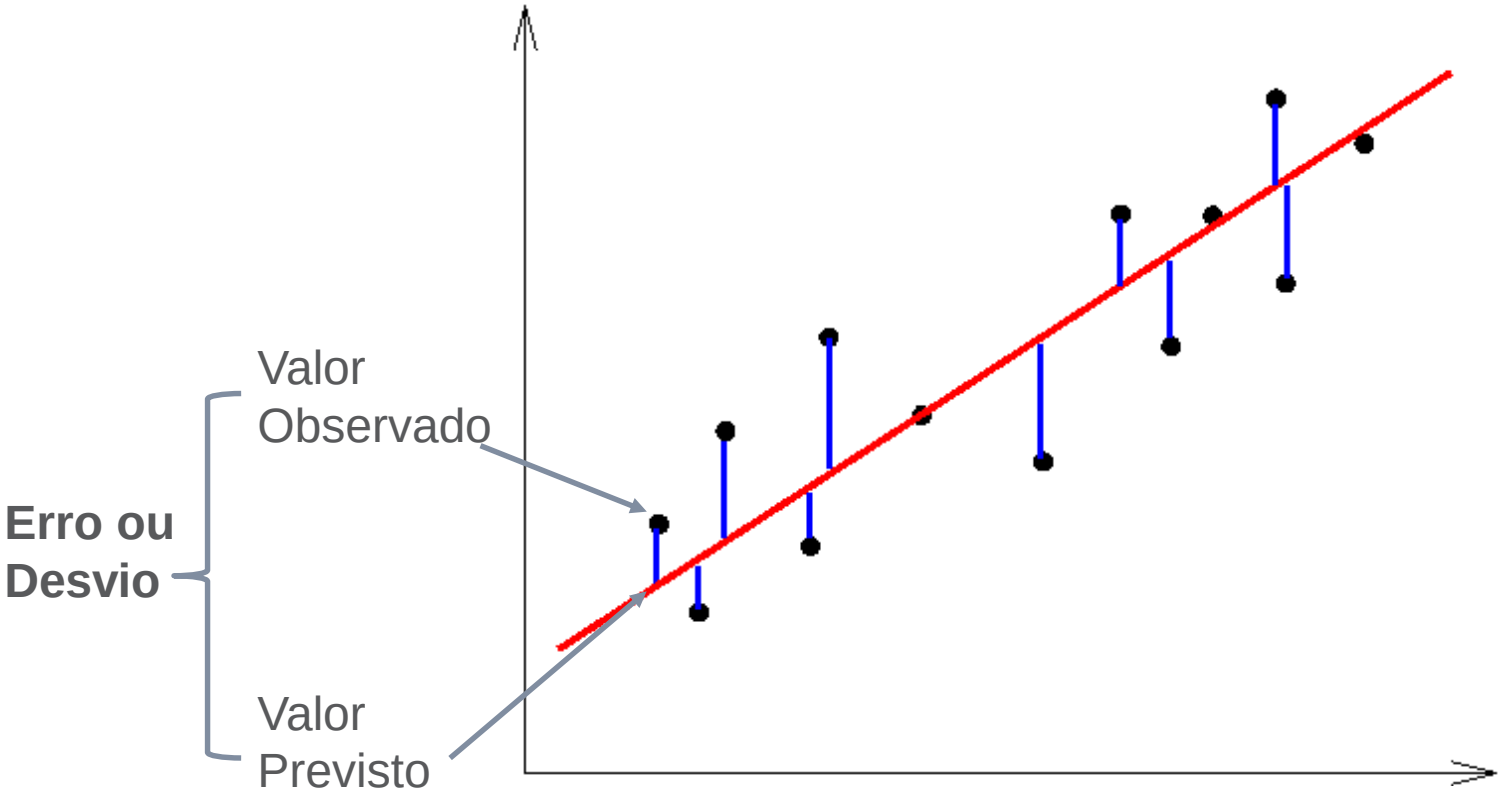


Onde,

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\beta_1 = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

Modelo de Regressão



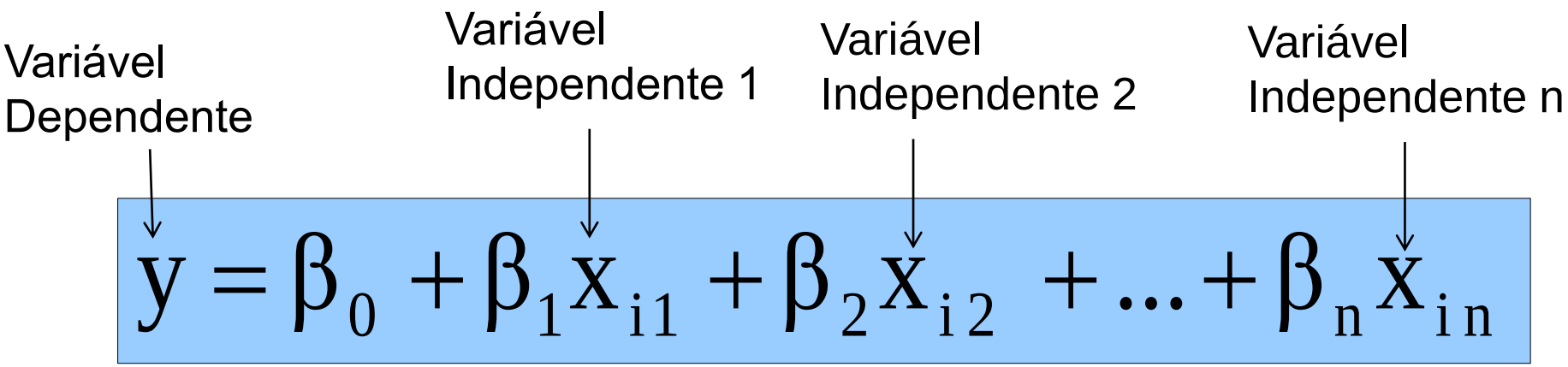
Regressão Linear Múltipla

- A análise de uma regressão múltipla segue, basicamente, os mesmos critérios da análise de uma regressão simples.

- Em vez de uma variável independente x (por exemplo, quando nós modelamos o preço da casa com base apenas em seu tamanho), vamos considerar múltiplas variáveis independentes x1, x2, ... xN.

- Podemos prever, por exemplo o preço da casa com base em seu tamanho e número de quartos.

Regressão Linear Multipla



A expressão para os parâmetros do modelo β é:

$$\boldsymbol{\beta} = (X^t X)^{-1} X^t y$$

Regressão Linear Múltipla

Onde a Matriz X é definida como ($X_{i\,0}$ =1):

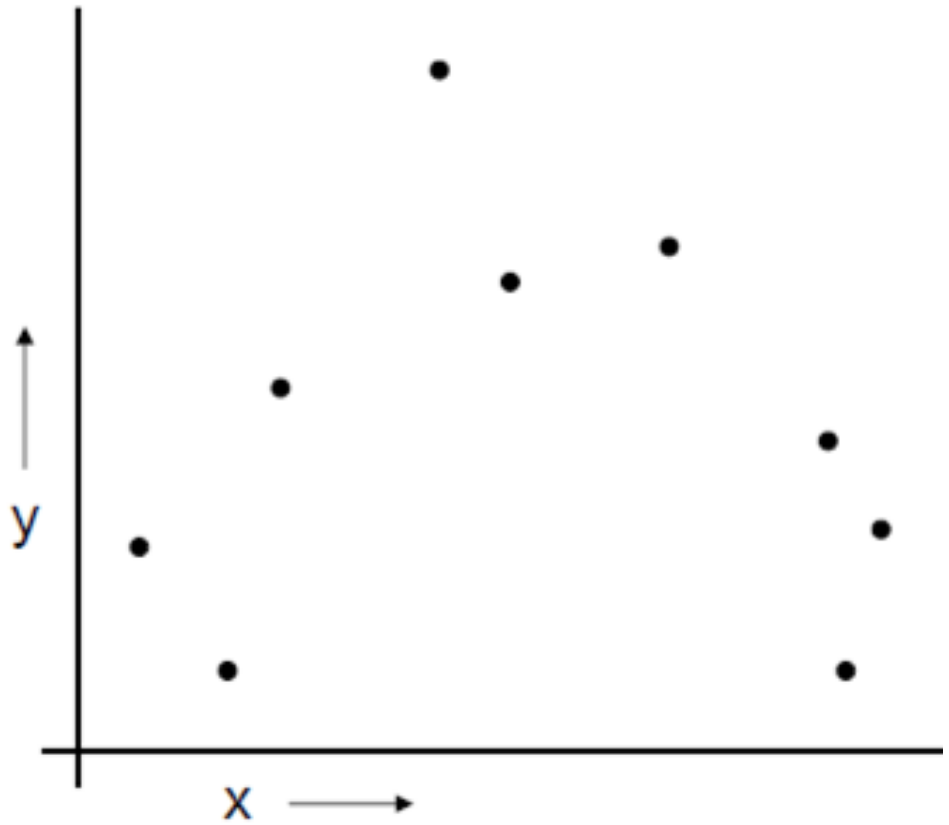
$$X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1N} \\ 1 & X_{21} & X_{22} & \dots & X_{2N} \\ . & . & . & \dots & . \\ 1 & X_{m1} & X_{m2} & \dots & X_{mN} \end{pmatrix}$$

Podemos reescrever a linha de regressão como:

$$y = X\beta$$

Problema de Regressão

O que podemos aprender destes dados? O que fazer quando os nossos dados não são lineares?



Regressão Não-Linear (Polinomial)

A Regressão Polinomial encaixa uma relação não linear entre o valor de x e o valor correspondente de y.

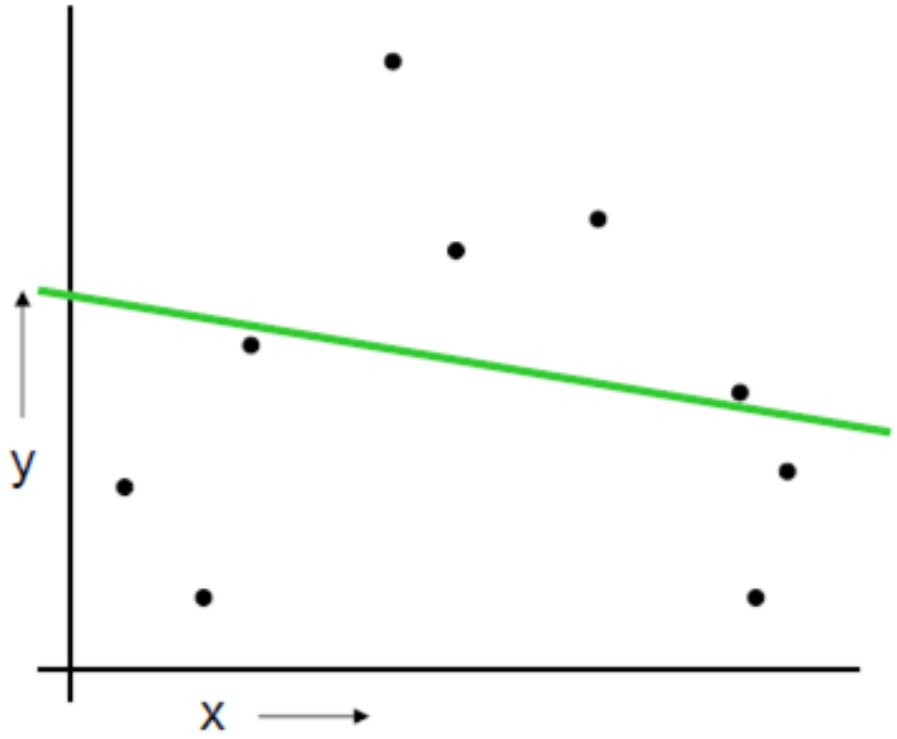
Fórmula Geral da Regressão não Linear:

$$y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \ldots + \beta_N X^N$$

- Uma maneira de escolher qual modelo polinomial deve ser usado, começamos ajustando uma regressão linear de dados:

$$y = \beta_0 + \beta_1 X$$

Regressão Linear



Regressão Não-Linear (Polinomial)

A Regressão Polinomial encaixa uma relação não linear entre o valor de x e o valor correspondente de y.

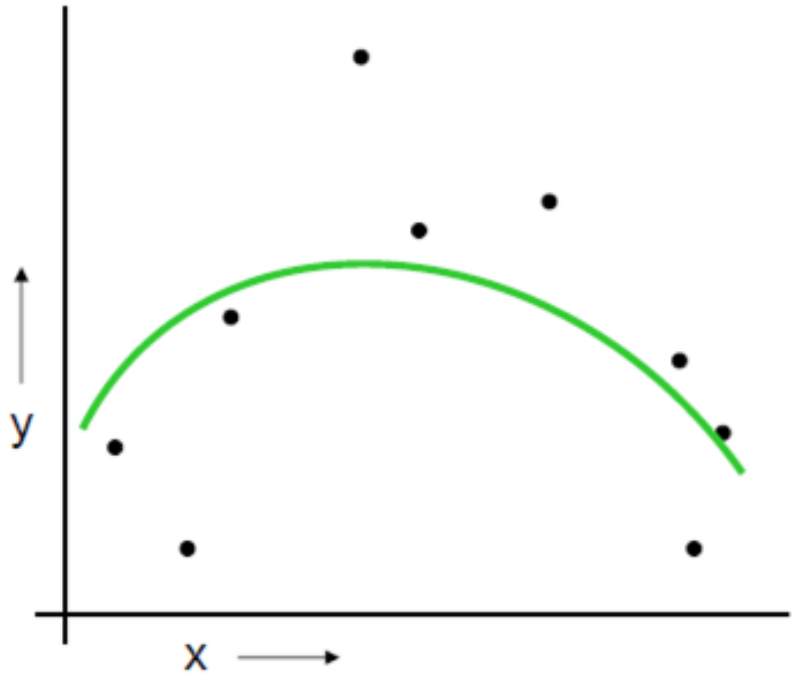
Fórmula Geral da Regressão não Linear:

$$y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \ldots + \beta_N X^N$$

- Em seguida, encaixamos um modelo polinomial de segundo grau (uma equação quadrática) para os dados:

$$y = \beta_0 + \beta_1 X + \beta_2 X^2$$

Regressão Quadrática



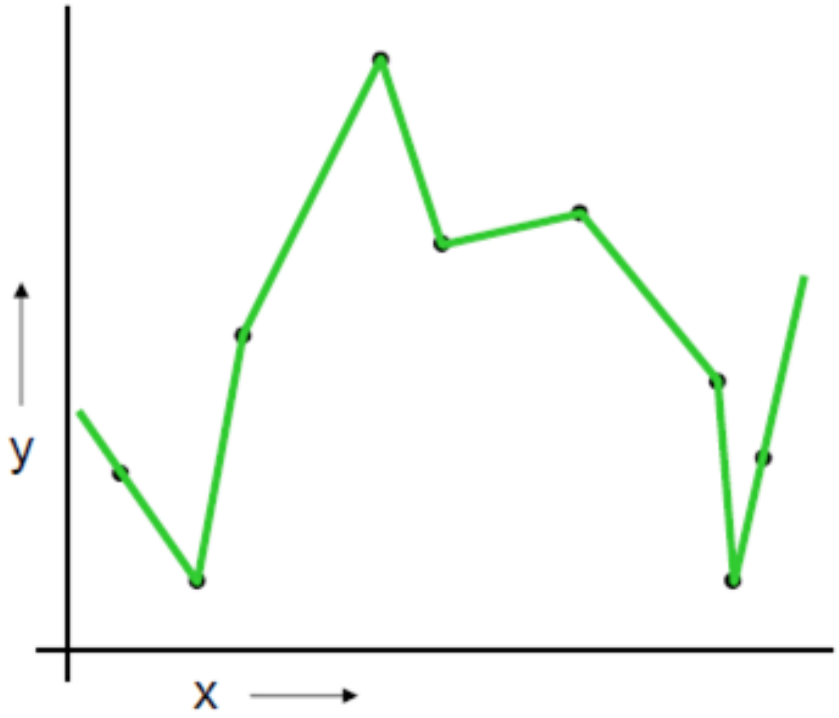
Regressão Não-Linear (Polinomial)

A Regressão Polinomial encaixa uma relação não linear entre o valor de x e o valor correspondente de y.

Fórmula Geral da Regressão não Linear:

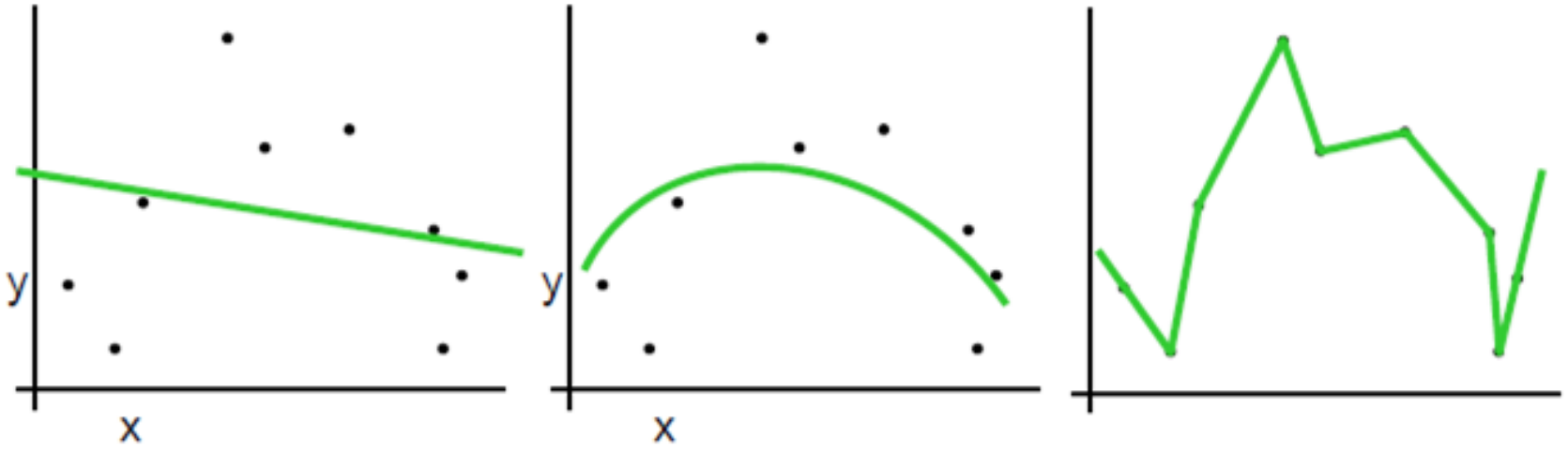
$$y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \ldots + \beta_N X^N$$

- E assim por diante.... até n


y
x

Qual é o melhor?

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



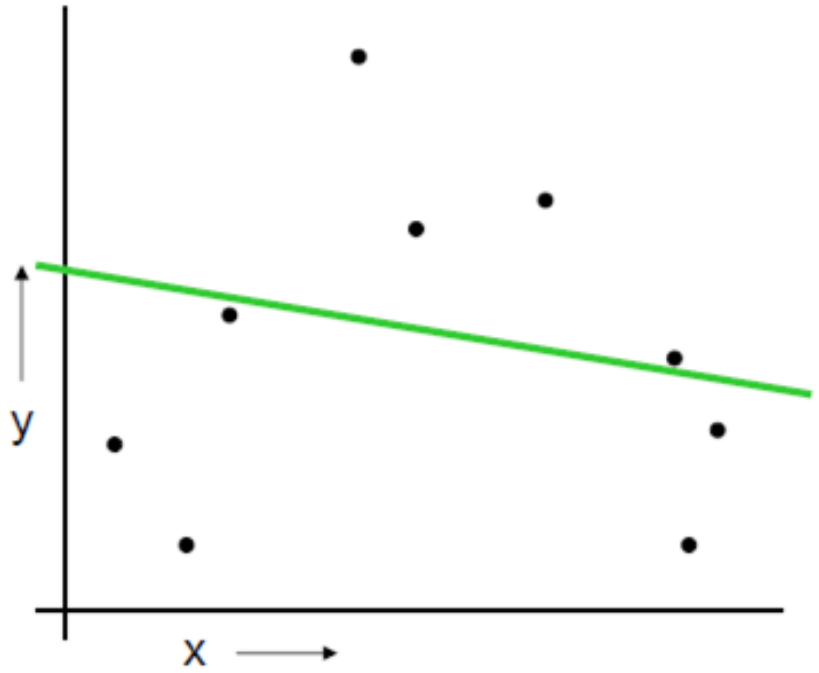
Generalizar é Difícil

- **Não queremos aprender por memorização**
  - Boa resposta somente sobre os exemplos de treinamento.
  - Fácil para um computador.
  - Difícil para os humanos.

- **Aprender visando generalizar**
  - Mais interessante.
  - Fundamentalmente mais difícil: existem diversas maneiras de generalizar.
  - Devemos extrair a essência, a estrutura dos dados e não somente aprender a boa resposta para alguns casos.

Underfitting

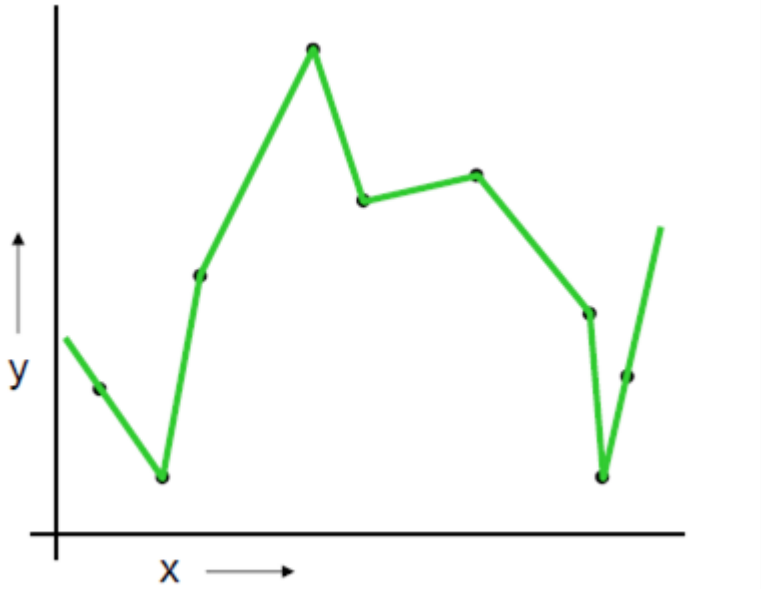
- Quando escolhemos um modelo muito simples (linear):
  - erro elevado na aprendizagem.
  - Não consegue nem mesmo modelar os dados de treinamento e portanto não consegue generalizar para novos dados



Overfitting (Sobre-ajuste)

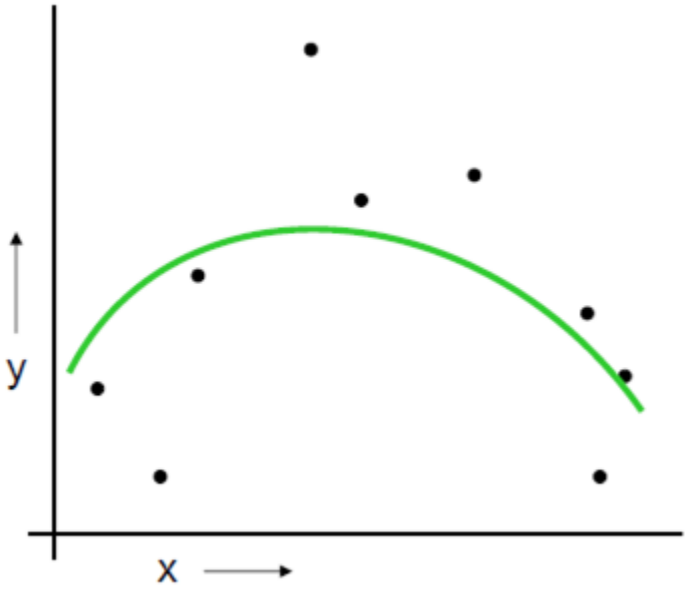
Erro baixo sobre os exemplos de treinamento e mais elevado para os exemplos de teste.

- Algoritmo pode memorizar os dados no treinamento e falir ao tentar generalizar novos exemplos.
- Aprende os detalhes e os ruídos nos dados de treinamento



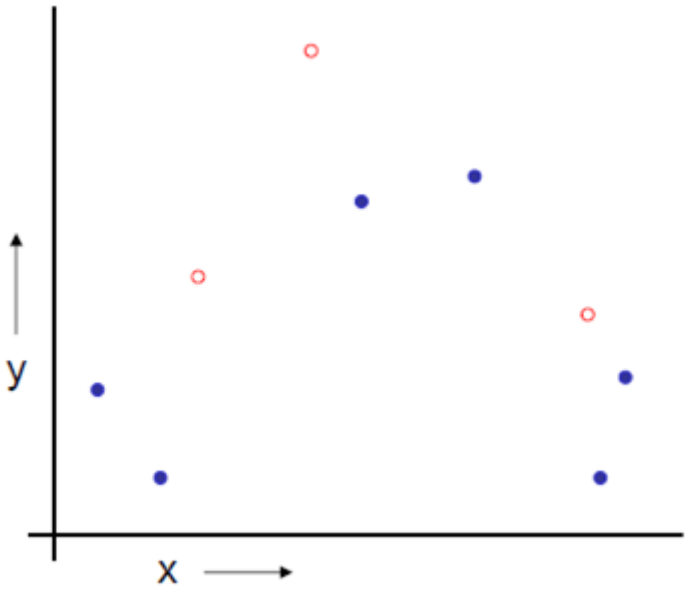
Um Bom Modelo

O modelo é suficientemente flexível para capturar a forma curva mas não é suficiente para ser exatamente igual a distribuição do conjunto de dados.



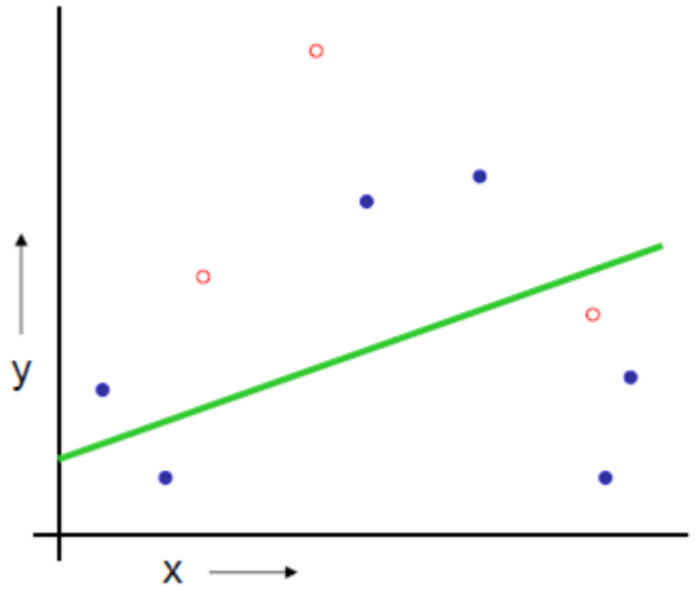
Como encontrar a melhor solução?

Método do *Test Set*



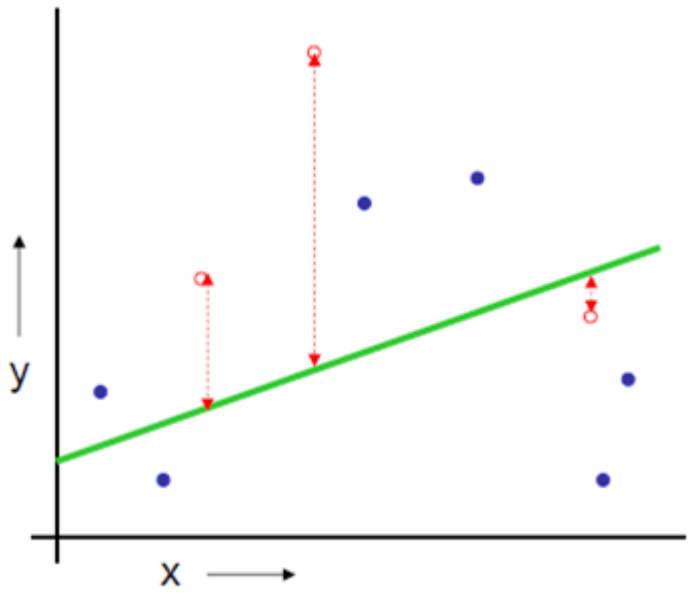
1. Escolha aleatoriamente 30% dos dados para fazer parte do Grupo de Teste.
2. O restante será o seu Grupo de Treinamento.

Método do *Test Set*



1. Escolha aleatóriamente 30% dos dados para fazer parte do Grupo de Teste.
2. O restante será o seu Grupo de Treinamento.
3. Faça o cálculo da Regressão no Grupo de Treinamento.

Método do *Test Set*



Exemplo Regressão Linear

Erro Quadrático Médio (EQM) = 2.4

1. Escolha aleatóriamente 30% dos dados para fazer parte do Grupo de Teste.
2. O restante será o seu Grupo de Treinamento.
3. Faça o cálculo da Regressão no Grupo de Treinamento.
4. Estime sua performance calculando os Dados de Teste.
5. Calcule o Erro Quadrático Médio (EQM) para estimar qual método é o mais preciso.

Erro Quadrático Médio

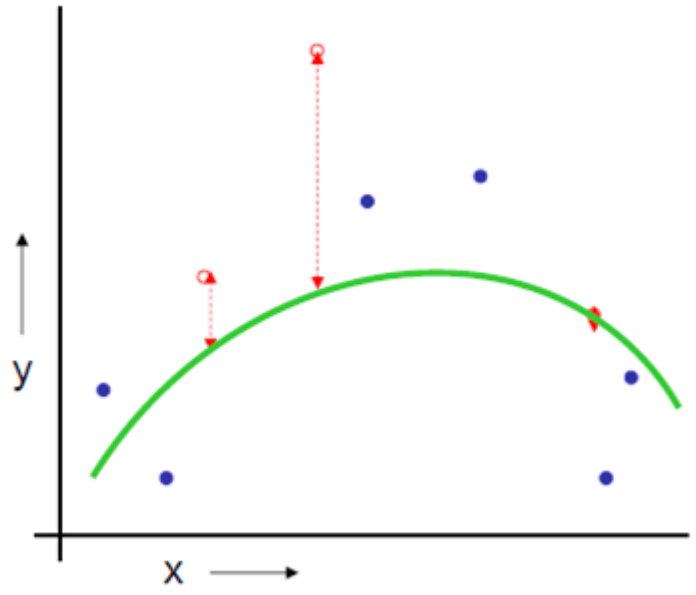
Erro Quadrático Médio (EQM): É a soma das diferenças entre o valor estimado e o valor real dos dados, ponderados pelo número de termos.

EQM= (soma(residuo)) / size(y,1)

Resíduo: calcula a diferença entre o valor **observado** *y*, e o valor **estimado** pela reta ȳ, isto é:

$$\textbf{residuo} = (y - \bar{y})^2$$

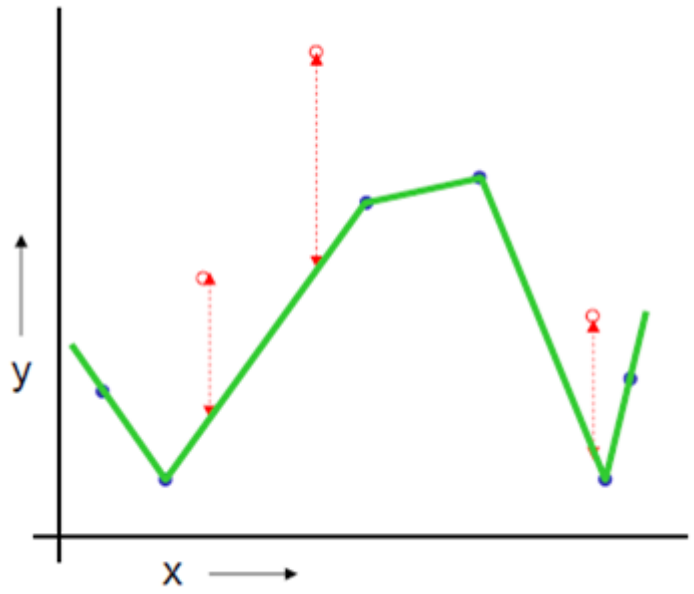
Método do Test Set



Erro Quadrático Médio (EQM) = 0.9

1. Escolha aleatóriamente 30% dos dados para fazer parte do Grupo de Teste.
2. O restante será o seu Grupo de Treinamento.
3. Faça o cálculo da Regressão no Grupo de Treinamento.
4. Estime sua performance calculando os Dados de Teste.
5. Calcule o Erro Quadrático Médio (EQM) para estimar qual método é o mais preciso.

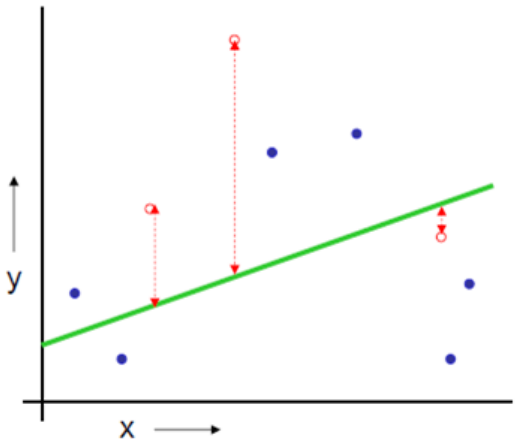
Método do Test Set

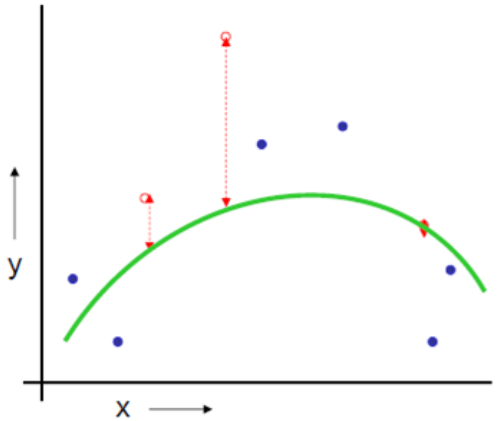


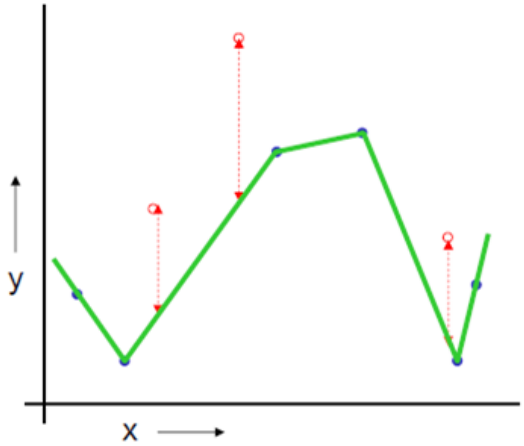
Erro Quadrático Médio (EQM) = 2.2

1. Escolha aleatóriamente 30% dos dados para fazer parte do Grupo de Teste.
2. O restante será o seu Grupo de Treinamento.
3. Faça o cálculo da Regressão no Grupo de Treinamento.
4. Estime sua performance calculando os Dados de Teste.
5. Calcule o Erro Quadrático Médio (EQM) para estimar qual método é o mais preciso.

Método do Test Set







EQM = 2.4

EQM = 0.9

EQM = 2.2