Aprendizado de Máquina

Aula 7: Algoritmos baseados em probabilidade (parte 2)

André C. P. L. F de Carvalho
ICMC/USP
andre@icmc.usp.br

Tópicos a serem abordados nesta parte

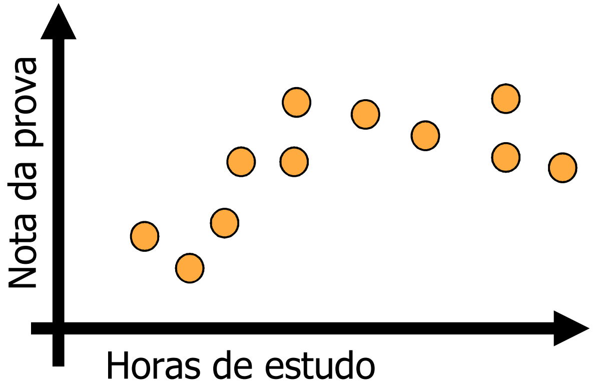
- Aproximação de funções
- Tarefas de regressão
- Regressão linear simples

Introdução

- Regressão é uma das principais tarefas preditivas
- Objetivo:
    - Aprender (aproximar) uma função que associa:
        - A descrição de um objeto, vetor de valores, conjunto de variáveis independentes (atributos preditivos), a
        - Um valor real de uma variável dependente (atributo alvo) que rotula o objeto
    - Que pode ser utilizada para prever, com boa capacidade preditiva, o rótulo de um novo objeto

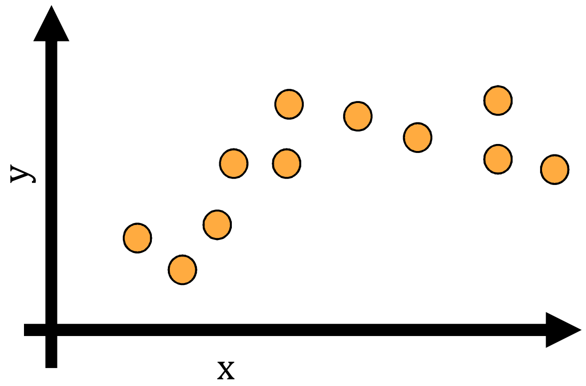
Aproximação de funções

- Procura uma função hipótese que se ajuste aos dados disponíveis
  - Permite prever o valor de saída para um novo objeto
    - Representado por um vetor de valores de entrada
  - Supor que você estudou 9 horas para a prova



Aproximação de funções

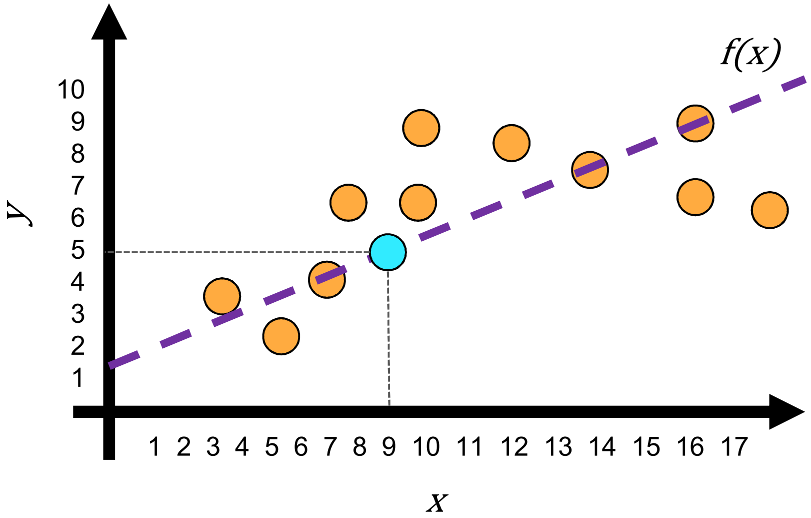
- Procura uma função hipótese que se ajuste aos dados disponíveis
  - Permite prever o valor de saída para um novo objeto
    - Representado por um vetor de valores de entrada
  - Supor que você estudou 9 horas para a prova



Aproximação de funções

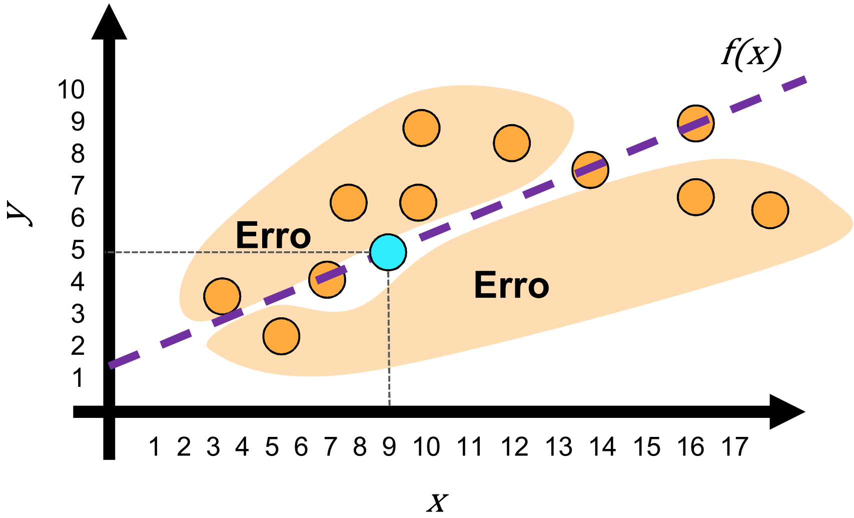
- Alternativa mais simples: aproximar por uma função linear
  - Função linear:
    - *f(x) = ax + b*

      Coeficiente angular
      Coeficiente linear
      - *Descobrir valor de a e de b*
    - *Se você estudou 9 horas (x=9)*
      - *Sua nota será 5,0 (f(x)=5)*
  - Função linear pode ser usada para prever valor da nota
    - Tarefa de regressão



Aproximação de funções

- Alternativa mais simples: aproximar por uma função linear
  - Função linear:
    - *f(x) = ax + b* (Coeficiente angular: a; Coeficiente linear: b)
      - *Descobrir valor de a e de b*
    - *Se você estudou 9 horas (x=9)*
      - *Sua nota será 5,0 (f(x)=5)*
  - Função linear pode ser usada para prever valor da nota
    - Tarefa de regressão



Tarefas de regressão

- Buscam função de regressão que melhor se aproxime aos dados
  - Minimizando o erro para todo o conjunto de dados
  - Algoritmo de **regressão linear** ajuda a encontrar esta função (hipótese, modelo)

$h_\beta(x) = \beta_0 + \beta_1 x_1 + \varepsilon \quad (\mathrm{f(x) = ax + b})$

onde:

β_j é o j$^{\text{ésimo}}$ parâmetro (j = 1,...,d)
β_0 é o termo de interceptação
x_1 é a variável independente
ε é o termo residual (ruído, erro de estimação da função verdadeira)

Tarefas de regressão

- Buscam função de regressão que melhor se aproxime aos dados
  - Minimizando o erro para todo o conjunto de dados
  - Algoritmo de **regressão linear** ajuda a encontrar esta função (hipótese, modelo)

$h_\beta(x) = \beta_0 + \beta_1 x_1$ $(\mathrm{f(x) = ax + b})$

onde:

β_j é o j$^{\text{ésimo}}$ parâmetro (j = 1,...,d)
β_0 é o termo de interceptação
x_1 é a variável independente

Tarefas de regressão

- Buscam função de regressão que melhor se aproxime aos dados
    - Minimizando o erro para todo o conjunto de dados
    - Algoritmo de **regressão linear** ajuda a encontrar esta função (hipótese, modelo)

$h_\theta(x) = \theta_0 + \theta_1 x_1$ $(f(x) = ax + b)$

onde:

θ_j é o j^ésimo^ parâmetro (j = 1,...,d)
θ_0 é o termo de interceptação
x_1 é a variável independente

Tarefas de regressão

- Buscam função de regressão que melhor se aproxime aos dados
    - Minimizando o erro para todo o conjunto de dados
    - Algoritmo de **regressão linear** ajuda a encontrar esta função (hipótese, modelo)

$$h_w(x) = w_0 + w_1 x_1$$

onde:

w_j é o j^ésimo^ parâmetro (j = 1,...,d)

w_0 é o termo de interceptação

x_1 é a variável independente

Tarefas de regressão

- Necessário encontrar valores dos parâmetros w_0 *e* w_1 que minimizem o erro da função hipótese $h_w(x) = w_0 + w_1 x$
  - $h_w(x)$: função linear
  - w_0: coeficiente linear da reta
  - w_1: coeficiente angular da reta
- Erro cometido pela função hipótese pode ser estimado por uma função de custo $J(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - h_w(x^i)\right)^2$

Tarefas de regressão

- Necessário encontrar valores dos parâmetros w_0 e w_1 que minimizem o erro da função hipótese $f(x) = w_0 + w_1 x$
  - $f(x)$: função linear
  - w_0: coeficiente linear da reta
  - w_1: coeficiente angular da reta
- Erro cometido pela função hipótese pode ser estimado por uma função de custo $J(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - f(x^i)\right)^2$

Aprendizado

- Buscam minimizar função de custo
  - Achar valores de w (w_0 *e* w_1) que gerem a função $J(w)$ de menor custo (taxa de erro)
  - Pode ser ilustrado por um gráfico que mostra a relação entre valores de parâmetros e valor do erro
  - Buscar w que minimiza $(y\text{-}f(x))^2$

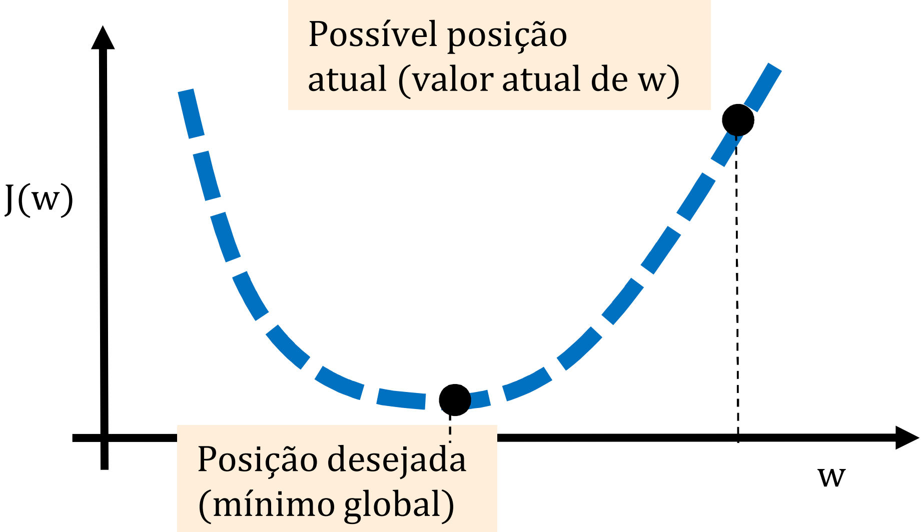
$$\min_{w_0 w_1} (y - f(x))^2$$

$$\min_{w_0 w_1} \sum_{i=1}^{n} \left(y^i - f(x^i)\right)^2 \quad \text{Erro total}$$

$$\min_{w_0 w_1} \frac{1}{n} \sum_{i=1}^{n} \left(y^i - f(x^i)\right)^2 \quad \text{Erro médio}$$

$$\min_{w_0 w_1} J(w_0, w_1)$$

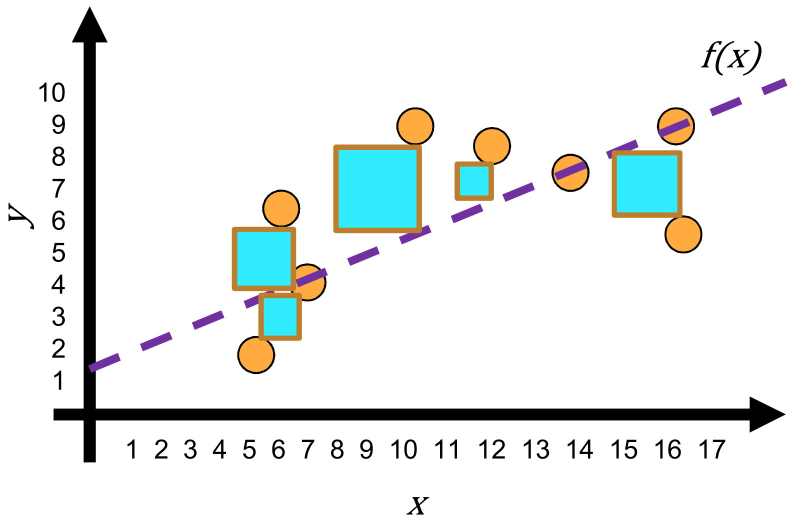
Erro quadrático médio (MSE)



Aproximação da função hipótese

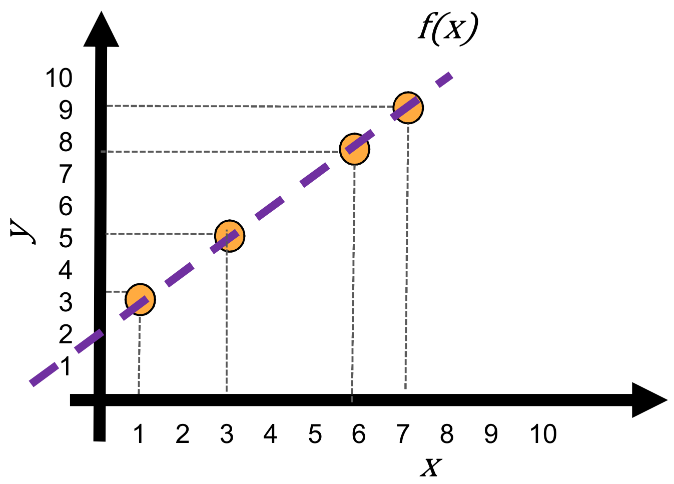
- Minimizar erro quadrático médio (MSE, Mean Squared Error)
  - Média da área dos quadrados entre saída real (y^i) e saída estimada ($f(x^i)$)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - f(x^i)\right)^2$$

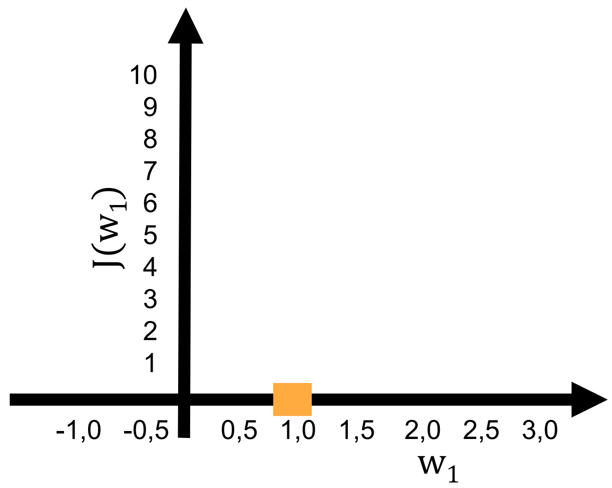


Trabalhamos com duas funções

- Função hipótese:
  - $f(x) = 2 + w_1 x_1$ (seja w_0 fixo, = 2)
  - $f(x) = 2 + 0{,}9x$

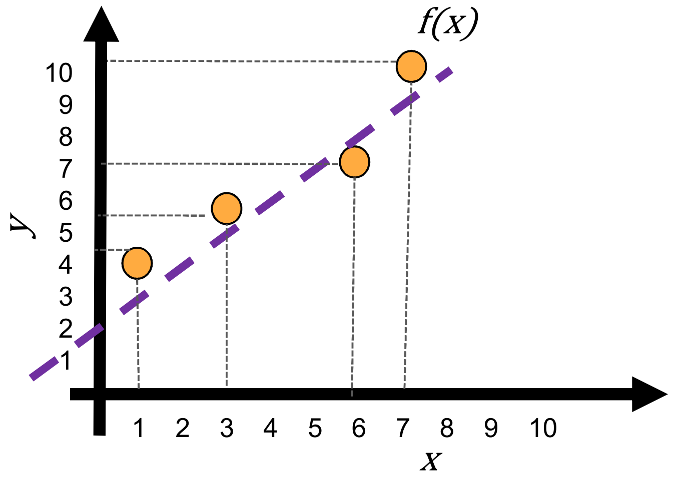


- Função de custo:
  - $J(w_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - f(x^i)\right)^2$
  - $J(0{,}9) = 0 + 0 + 0 + 0 = 0$

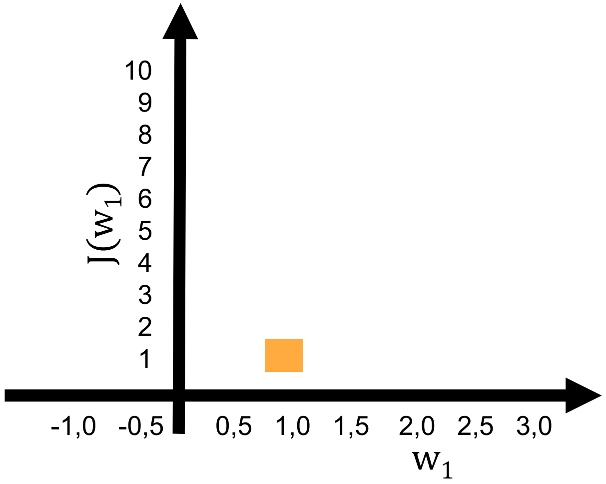


Trabalhamos com duas funções

- Função hipótese:
  - $f(x) = 2 + w_1 x_1$ (seja w_0 fixo, = 2)
  - $f(x) = 2 + 0{,}9x$

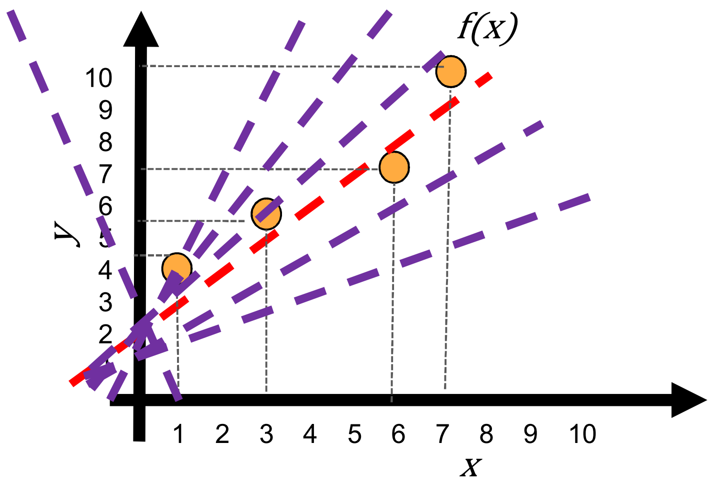


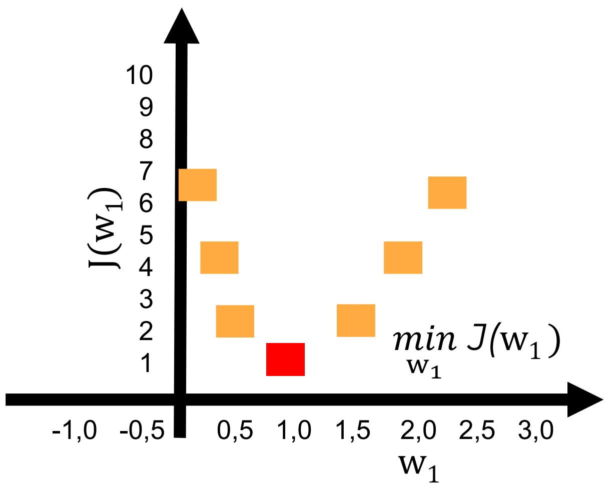
- Função de custo:
  - $J(w_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - f(x^i)\right)^2$
  - $J(0{,}9) = 1^2+1^2+1^2+1^2/4 = 1$



Avaliando diferentes valores de w_1

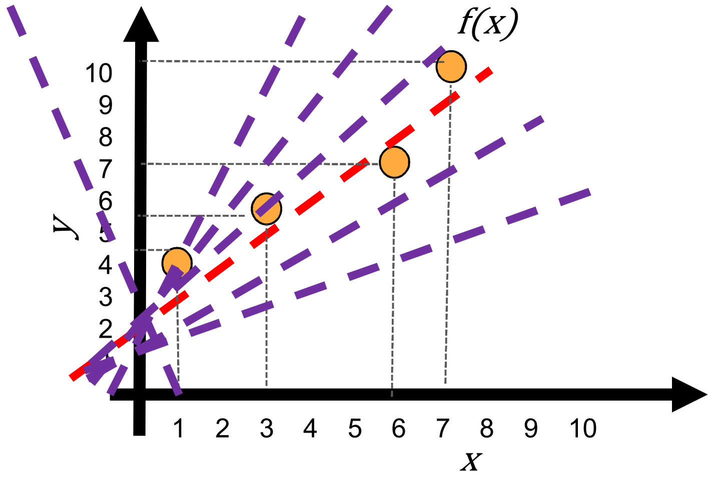
- Função hipótese:
  - $f(x) = 2 + w_1 x_1$ (seja w_0 fixo, = 2)
  - Diversos valores de w_1 (várias funções)
- Função de custo:
  - Calculada para cada valor de w_1
  - $J(w_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - f(x^i)\right)^2$



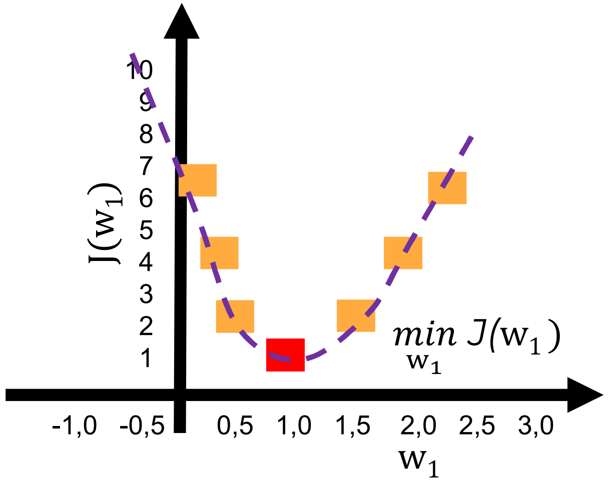


Avaliando diferentes valores de w_1

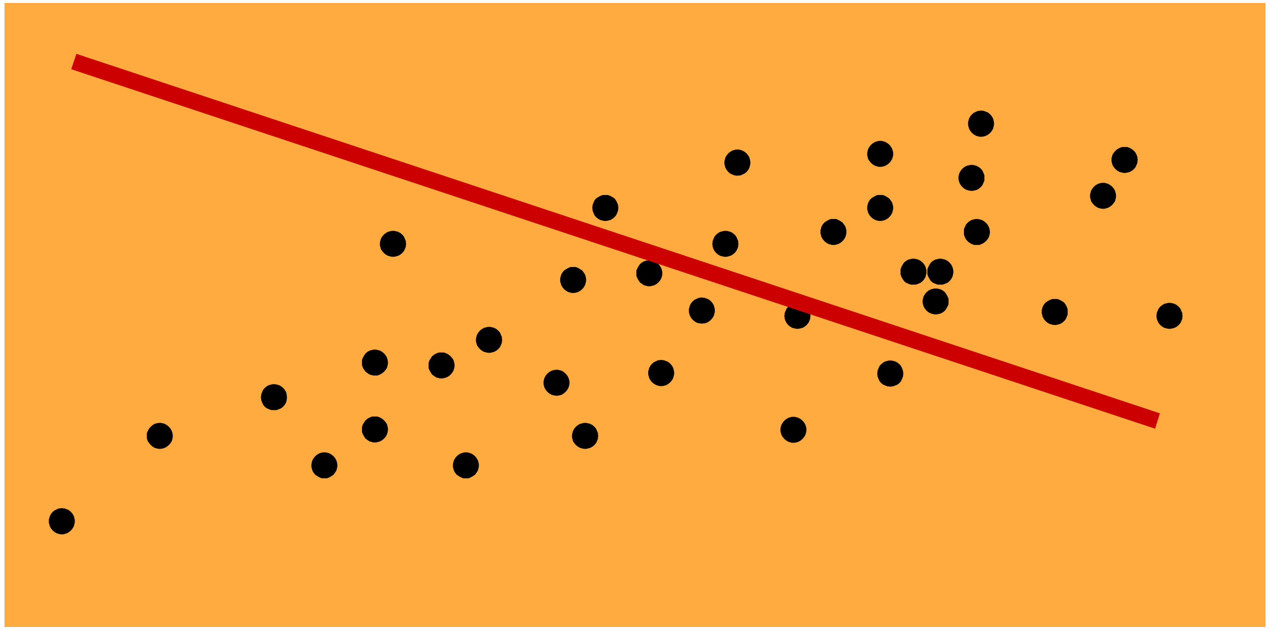
- Função hipótese:
  - $f(x) = 2 + \mathrm{w}_1 x_1$ (seja w_0 fixo, = 2)
  - Diversos valores de w_1 (várias funções)



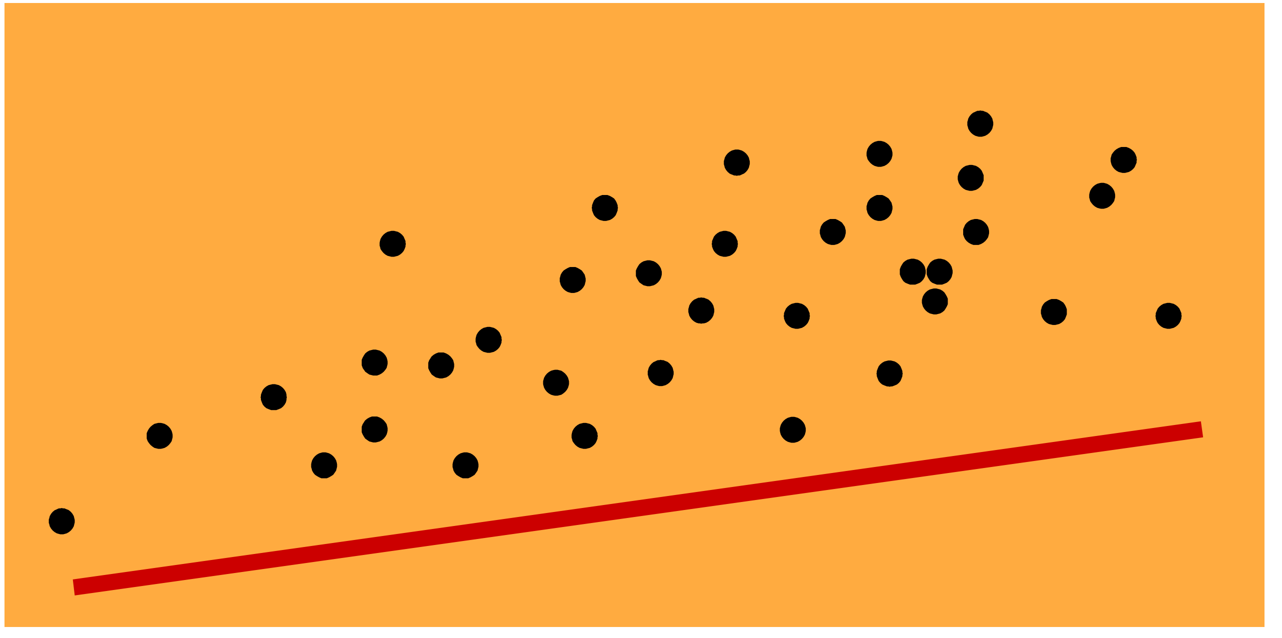
- Função de custo:
  - Calculada para cada valor de w_1
  - $J(\mathrm{w}_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - f(x^i)\right)^2$



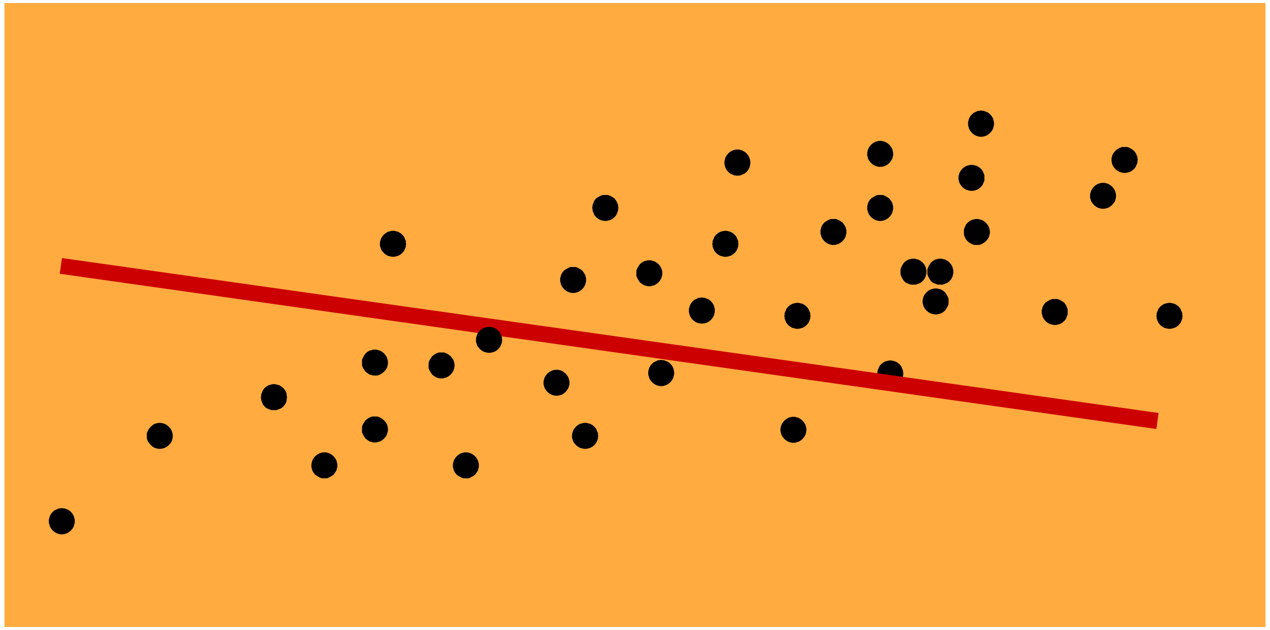
Ajuste de parâmetros



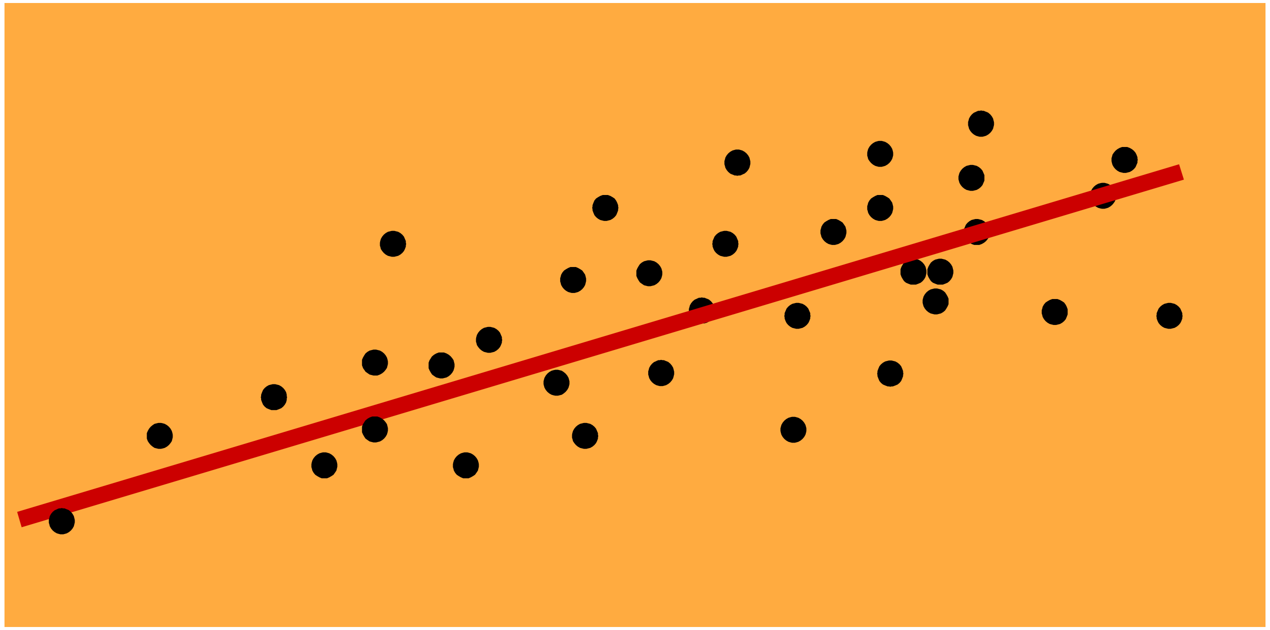
Ajuste de parâmetros



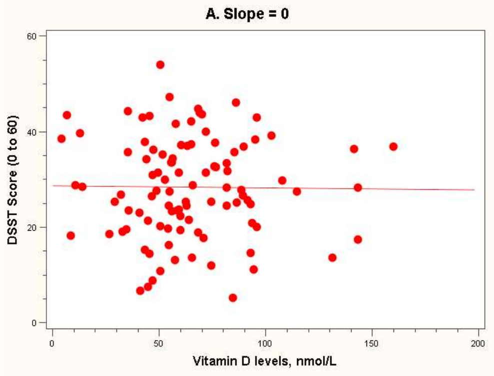
Ajuste de parâmetros



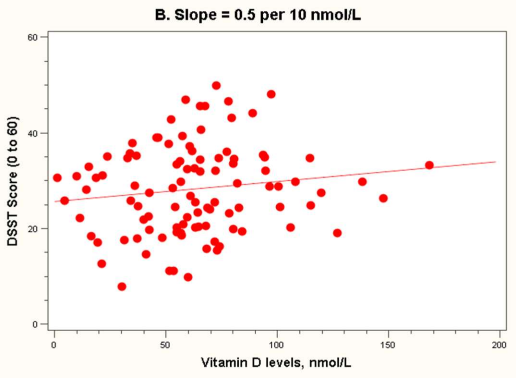
Ajuste de parâmetros



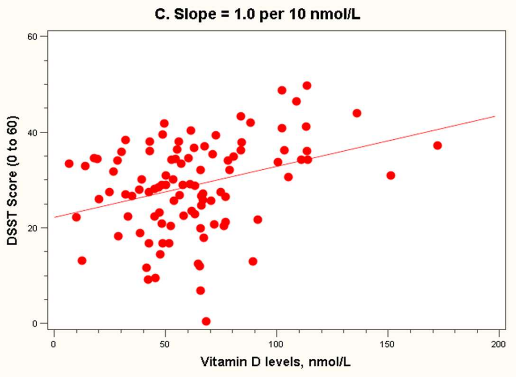
Ajuste de parâmetros



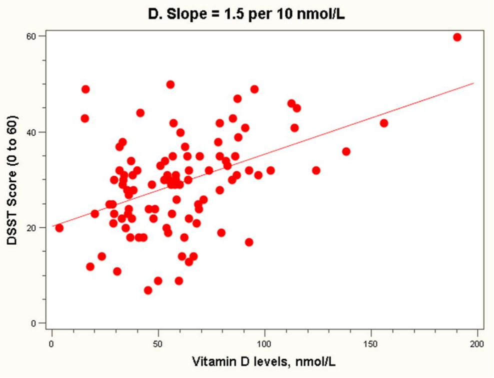
f(x) = 28



f(x) = 26 + 0,5x



f(x) = 22 + x



f(x) = 20 + 1,5x

Sumarizando

- Regressão linear simples
- Função hipótese: $h_w(x) = w_0 + w_1 x$ ou $f(x) = w_0 + w_1 x$
- Parâmetros: w_0 , w_1
- Função custo: $J(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - f(x^i)\right)^2$
  - Para simplificar cálculos, pode ser usada a função $J(w_0, w_1) = \frac{1}{2n}\sum_{i=1}^{n}\left(y^i - f(x^i)\right)^2$
- Objetivo: $\min\limits_{w_0 w_1} J(w_0, w_1)$

Continua no próximo vídeo e conjunto de slides