Universidade Federal da Paraíba

Coordenação do Curso de Ciência de Dados e Inteligência Artificial

Modelo de Aprendizagem Linear II
Regressão Logística

Prof. Gilberto Farias

Roteiro

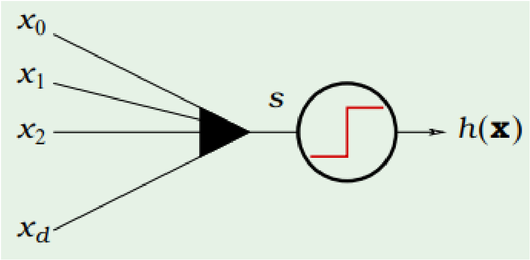
- O modelo de Regressão Logística
- Medida de Erro
- Algoritmo de Aprendizagem

Terceiro modelo linear – Regressão Logística

$$s = \sum_{i=0}^{d} w_i x_i$$

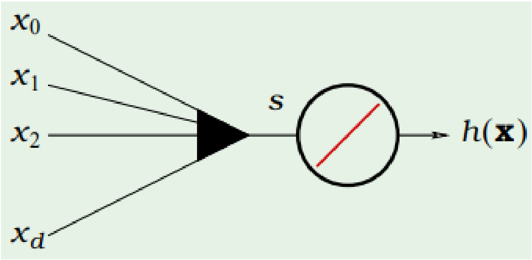
classificação linear (*perceptron*)

$$h(x) = sign(s)$$



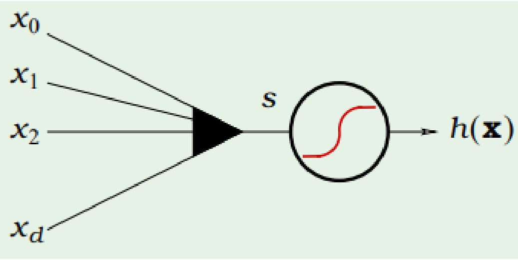
regressão linear

$$h(x) = s$$



regressão logística

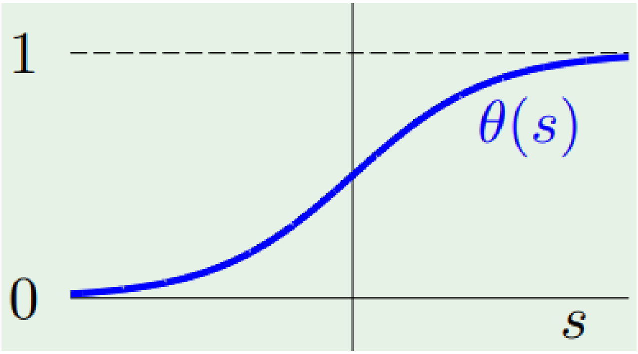
$$h(x) = \theta(s)$$



Função logística θ

A fórmula

$$\theta(s) = \frac{e^s}{1 + e^s}$$



limiar suave : incerteza

Interpretação probabilística

$h(x) = \theta(s)$ é interpretado como uma probabilidade

Exemplo. Predição de ataque cardíaco

Entrada $\boldsymbol{x}$: nível de colesterol, idade, peso, .. Etc.

$\theta(s)$: probabilidade de ataque cardíaco

O valor $\boldsymbol{s} = \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x}$ “pontuação de risco”

Probabilidade genuína

Os dados $(\boldsymbol{x}, y)$ com y binário (± 1) são gerados por uma função de ruído alvo:

$$P(y|x) = \begin{cases} f(x) & \text{para } y = +1; \\ 1 - f(x) & \text{para } y = -1. \end{cases}$$

O alvo $f: \mathbb{R}^d \to [0,1]$ é a probabilidade

Aprender $g(x) = \theta(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x}) \approx f(x)$

Medida de erro para a Regressão Logística

Uso da verossimilhança

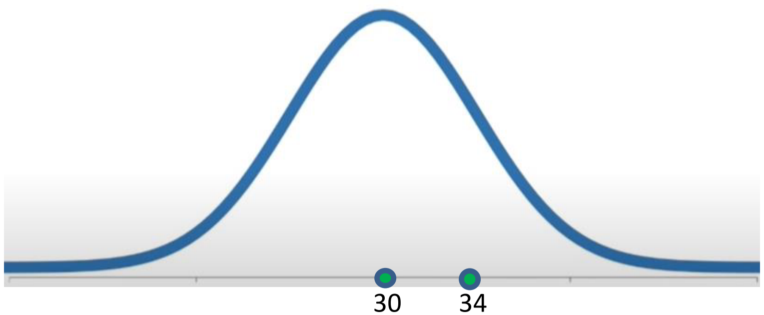
Para cada $(\boldsymbol{x}, y)$, y é gerado pela probabilidade $f(x)$

Medida de erro plausível baseada em verossimilhança:

Se $h = f$ quão provável se obtém y de x

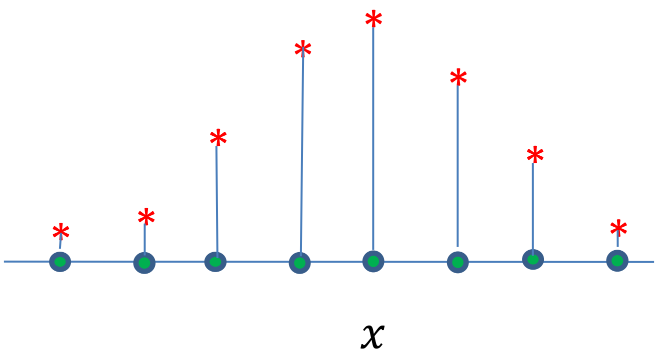
Probabilidade x Verossimilhança

θ=(σ, μ)



$P(30 \leq x \leq 34|\theta)$ =??

θ=??



$L(\theta|x)$: estimadores da distribuição de probabilidades

$$P(x|\theta) = L(\theta|x)$$

Uso da verossimilhança

Para cada $(\boldsymbol{x}, y)$, y é gerado pela probabilidade $f(x)$

Medida de erro plausível baseada em verossimilhança:

Se $h = f$ quão provável se obtém y de x

$$P(y|x) = \begin{cases} f(x) & \text{para } y = +1; \\ 1 - f(x) & \text{para } y = -1. \end{cases}$$

Uso da verossimilhança

Para cada $(\boldsymbol{x}, y)$, y é gerado pela probabilidade $f(x)$

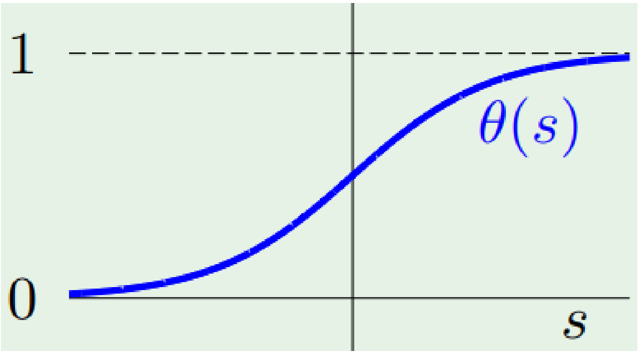
Medida de erro plausível baseada em verossimilhança:

Se $h = f$ quão provável se obtém y de x

$$P(y|x) = \begin{cases} h(x) & \text{para } y = +1; \\ 1 - h(x) & \text{para } y = -1. \end{cases}$$

Fórmula para a verossimilhança

$$P(y|x) = \begin{cases} h(x) & \text{para } y = +1; \\ 1 - h(x) & \text{para } y = -1. \end{cases}$$



notando que $\theta(-s) = 1 - \theta(s)$

Substituindo h$(x) = \theta(w^T x)$

Simplificando temos: $P(y|x) = \theta(y w^T x)$

Verossimilhança de $D = (x_1, y_1), .., (x_N, y_N)$, gerados independentemente, é:

$$\prod_{n=1}^{N} P(y_n|x_n) = \prod_{n=1}^{N} \theta(y_n w^T x_n)$$

Maximizando a verossimilhança

Maximize $$\prod_{n=1}^{N} \theta(y_n w^T x_n)$$

Maximizando a verossimilhança

Maximize $\ln\left(\prod_{n=1}^{N} \theta(y_n w^T x_n)\right)$ ln : função monótona

Maximizando a verossimilhança

Maximize

$$\frac{1}{N}\ln\left(\prod_{n=1}^{N}\theta(y_n w^T x_n)\right)$$

Maximizando a verossimilhança

Minimize

$$-\frac{1}{N}\ln\left(\prod_{n=1}^{N}\theta(y_n w^T x_n)\right)$$

$$=\frac{1}{N}\sum_{n=1}^{N}\ln\left(\frac{1}{\theta(y_n w^T x_n)}\right)$$

$$\left[\theta(s)=\frac{1}{1+e^{-s}}\right]$$

$$E_{in}=\frac{1}{N}\sum_{n=1}^{N}\ln\left(1+e^{-y_n w^T x_n}\right)$$

Maximizando a verossimilhança

Minimize $$-\frac{1}{N}\ln\left(\prod_{n=1}^{N}\theta(y_n w^T x_n)\right)$$

$$=\frac{1}{N}\sum_{n=1}^{N}\ln\left(\frac{1}{\theta(y_n w^T x_n)}\right)$$

$$\left[\theta(s)=\frac{1}{1+e^{-s}}\right]$$

$$E_{in}=\frac{1}{N}\sum_{n=1}^{N}\underbrace{\ln\left(1+e^{-y_n w^T x_n}\right)}_{e(h(x_n),y_n)}$$

erro de “entropia cruzada”

Como minimizar E_{in}

Para regressão logística,

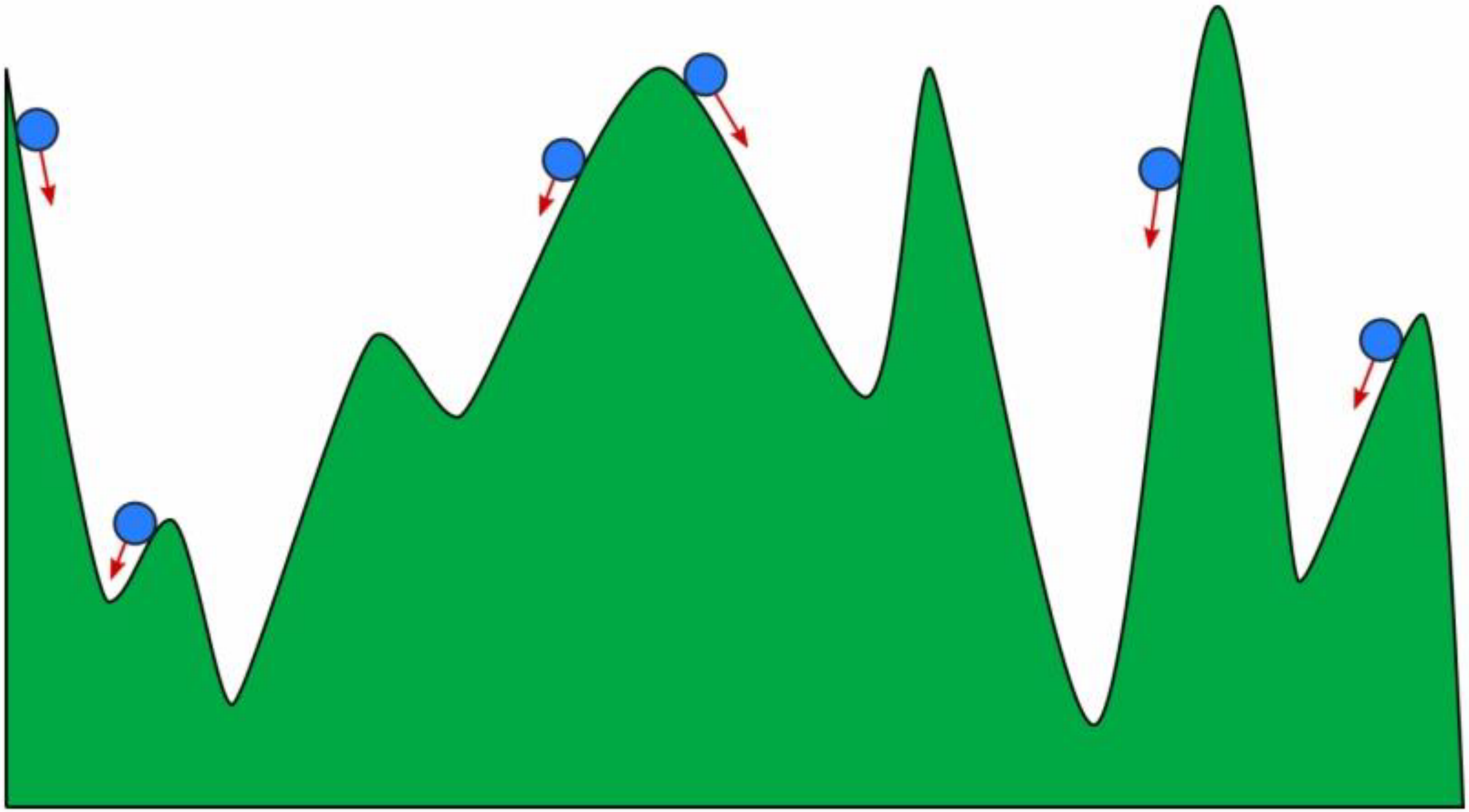
$$E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} \ln\left(1 + e^{-y_n w^T x_n}\right) \longleftarrow \text{solução } \textbf{iterativa}$$

Para regressão linear,

$$E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} (w^T x_n - y_n)^2 \longleftarrow \text{solução fechada}$$

Método do Gradiente Descendente

Método geral de otimização não linear



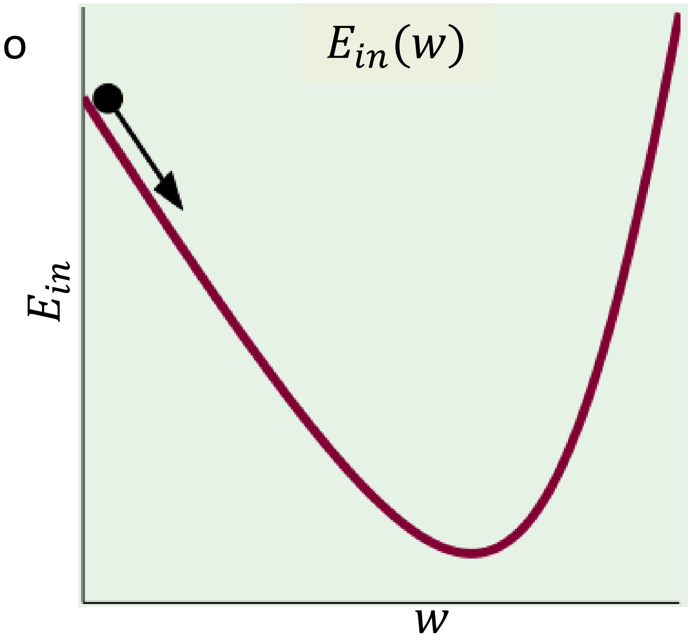
Método iterativo – Gradiente descendente

Inicia com $w(0)$; avança na direção de maior inclinação

Passos de tamanho fixo: $w(1) = w(0) - \eta\,\hat{\mathrm{v}}$

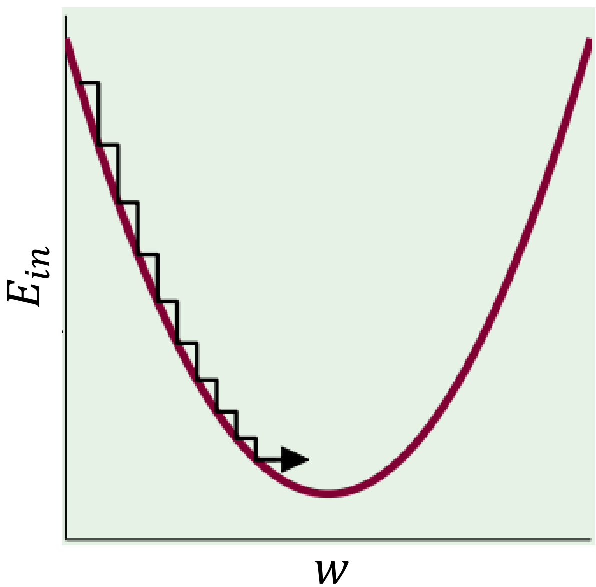
Qual a direção de $\hat{\mathrm{v}}$?

$$\hat{\mathrm{v}} = \nabla E_{in}(w(0))$$

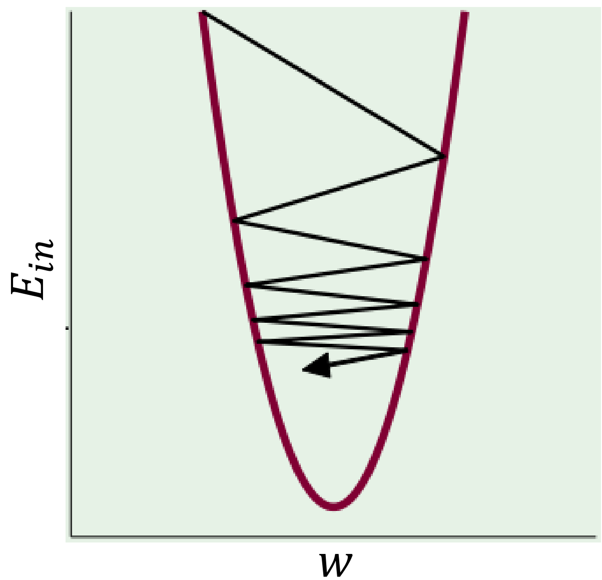


Qual o tamanho do passo?

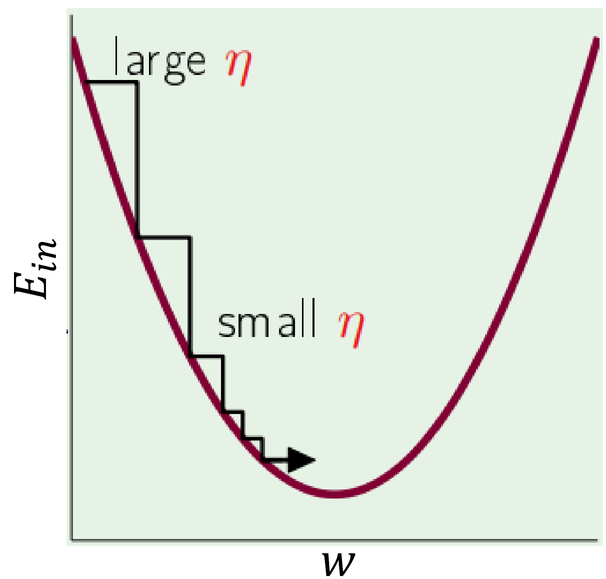
Como η afeta o algoritmo:



η pequeno



η grande



η na medida

η deve variar junto com a inclinação da superfície → $\eta \nabla E_{in}(w(0))$

comportamento da derivada

Algoritmo da Regressão Logística

1. Inicialize os pesos no tempo $t = 0$ para $w(0)$;
2. *Para* $t = 0,1,2, \ldots$ *faça*
3. *Compute o gradiente*

$$g_t = -\frac{1}{N}\sum_{n=1}^{N} \frac{y_n x_n}{1 + e^{y_n w^{\mathrm{T}}(t) x_n}}$$

4. *Atualize os pesos* $w(t+1) = w(t) - \eta g_t$*;*
5. *Itere para o próximo passo enquanto não atingir a condição de parada;*
6. *Retorne o peso final* ***w.***

Configurações do Algoritmo

- Inicialização das variáveis:
  - $\eta = 0.1$ (bons resultados na prática)
  - $w(0) = 0$ ou inicia com valor aleatório

- Condição de parada:
  - O ideal seria $\|g_t\| = 0$ (ponto mínimo)
  - Limitar a quantidade de iterações;
  - Definir um limiar para o valor de $\|g_t\|$.

Resumo dos Modelos Lineares