ACH2016 - Inteligência Artificial
Aula 04 - Regressão Linear

Valdinei Freire da Silva

valdinei.freire@usp.br - Bloco A1 100-O

Russell e Norvig, Capítulo 18

Tarefa de Aprendizado Supervisionado

Dado um conjunto de treinamento com N exemplos de pares entrada-saída

$$(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N),$$

onde cada y_i foi gerado por uma função f desconhecida, isto é, $y_i = f(x_i)$.

Descubra uma função h que aproxima a verdadeira função f.

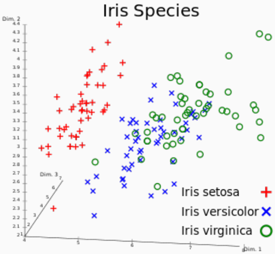
x é a entrada e y é a saída.

x e y podem ser qualquer valor, números ou categorias, x usualmente é um vetor de valores (atributos).

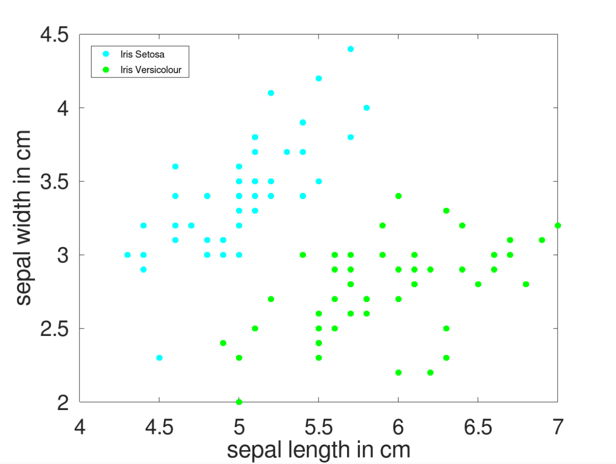
Banco de Dados Iris

Atributos:

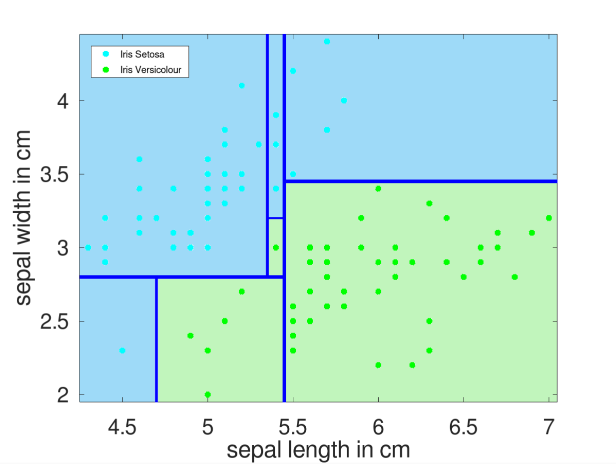
- comprimento da sépala
- largura da sépala
- comprimento da pépala
- largura da pépala



Banco de Dados Iris



Banco de Dados Iris - Árvore de Decisão



Classificação Linear

Classificação Binária Linear (d atributos)

$$h_{\mathrm{w}}(\mathrm{x}) = \begin{cases} 1, & \text{se } \mathrm{w}^{\top}\mathrm{x} > 0 \\ 0, & \text{caso contrário} \end{cases}$$

x é o vetor de entradas (d atributos) adicionado de um atributo "dummy" com valor fixo 1.

w é o vetor de pesos ($d + 1$ dimensões).

$\mathrm{w}^{\top}$ é o transposto do vetor w.

$\mathrm{w}^{\top}\mathrm{x}$ é a multiplicação matricial.

Como encontrar o vetor w que classifique adequadamente os dados da amostras?

Classificação Linear

A seguinte regra pode ser utilizada para tentar melhorar um vetor w_t:

- sorteie um vetor de pesos iniciais w_0 e repita os passos abaixo
- sorteie aleatoriamente um exemplo $e_i = (\mathrm{x}_i, y_i)$
- atualize o vetor w_t

  $$\mathrm{w}_{t+1} \leftarrow \mathrm{w}_t + \alpha_t(y_i - h_{\mathrm{w}_t}(\mathrm{x}_i))x_i$$

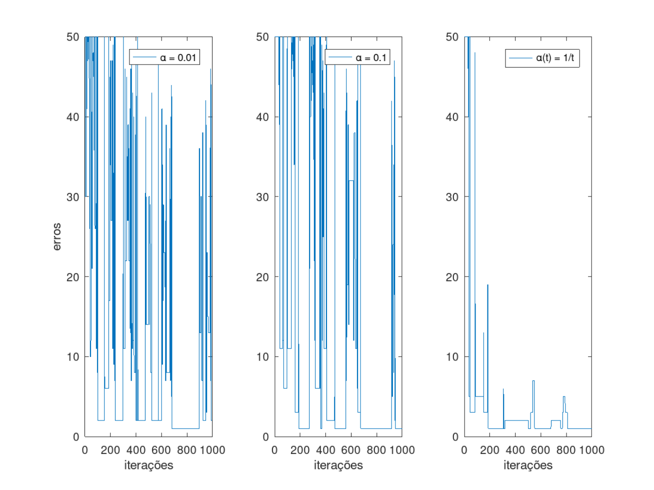
  onde α_t é a taxa de aprendizado no tempo t.

Classificação Linear

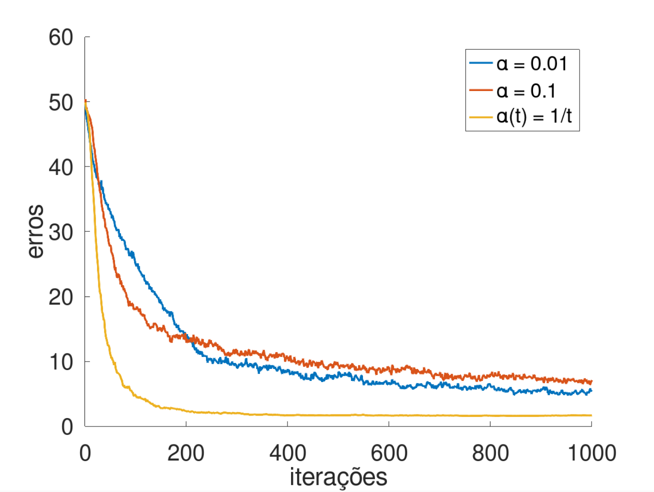
- se a saída $h_{\mathrm{w}}(\mathrm{x})$ está correta nenhuma alteração é realizada
- se $y = 1$, mas $h_{\mathrm{w}}(\mathrm{x}) = 0$, então w_i é aumentado se x_i é positivo, e diminuído quando x_i é negativo, isso faz com que $\mathrm{w}^\top \mathrm{x}$ aumente.
- se $y = 0$, mas $h_{\mathrm{w}}(\mathrm{x}) = 1$, então w_i é aumentado se x_i é negativo, e diminuído quando x_i é positivo, isso faz com que $\mathrm{w}^\top \mathrm{x}$ diminua.
- converge se

$$\sum_{t=1}^{\infty} \alpha_t = \infty \text{ e } \sum_{t=1}^{\infty} (\alpha_t)^2 < \infty$$

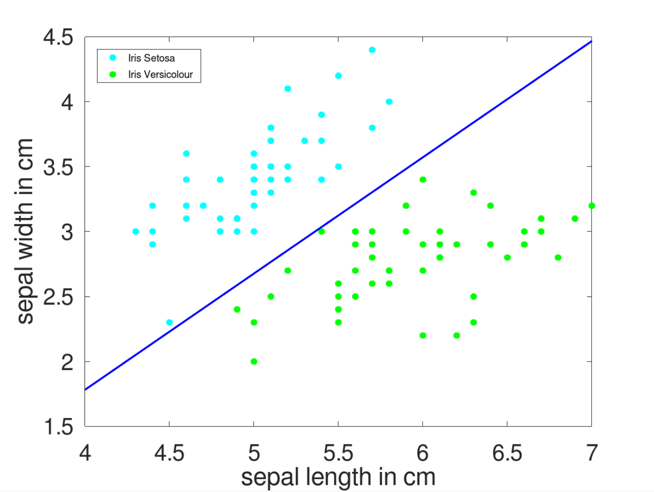
Classificação Linear (1 repetição)



Classificação Linear (1000 repetições)



Classificação Linear



Classificação Linear com Regressão Logística

Classificação Binária Linear de Múltiplas Variáveis (d variáveis)

$$h_{\mathrm{w}}(\mathrm{x}) = \mathrm{Logistic}(\mathrm{w}^\top \mathrm{x}) = \frac{1}{1 + e^{-\mathrm{w}^\top \mathrm{x}}}$$

x é o vetor de entrada adicionado de um atributo "dummy" com valor fixo 1.

Classifica como 1 se $h_{\mathrm{w}}(\mathrm{x}) > \beta$, e 0 caso contrário.

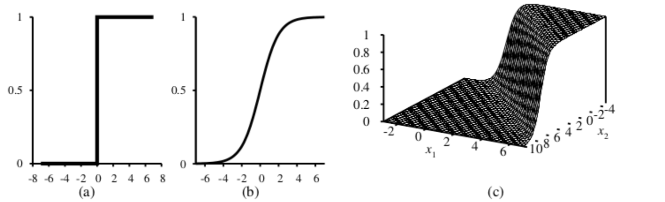
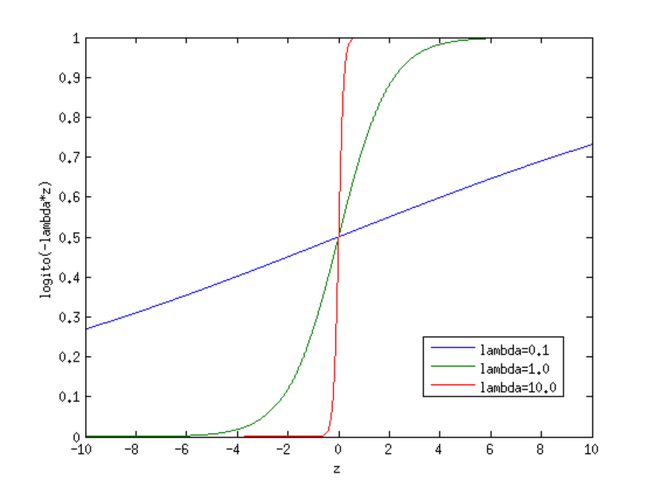


Figure 18.17 **FILES: .** (a) The hard threshold function $Threshold(z)$ with 0/1 output. Note that the function is nondifferentiable at $z = 0$. (b) The logistic function, $Logistic(z) = \frac{1}{1+e^{-z}}$, also known as the sigmoid function. (c) Plot of a logistic regression hypothesis $h_{\mathbf{w}}(\mathbf{x}) = Logistic(\mathbf{w} \cdot \mathbf{x})$ for the data shown in Figure 18.14(b).

Classificação Linear com Regressão Logística



Classificação Linear com Regressão Logística

Considere $h_{\mathrm{w}}(x_i)$ como sendo a probabilidade da entrada x_i ser classifica como 1 e definida a função perda (*loss function*) como sendo o *negative log-likelihood*:

$$Loss(h_{\mathrm{w}}) = -\sum_{i=1}^{n} y_i h_{\mathrm{w}}(x_i) + (1 - y_i)(1 - h_{\mathrm{w}}(x_i))$$

Solução:

$$\min_{\mathrm{w} \in \mathbb{R}^{d+1}} Loss(h_{\mathrm{w}}) \Rightarrow \nabla_{\mathrm{w}} Loss(h_{\mathrm{w}}) = 0$$

Pode-se encontrar iterativamente:

$$\mathrm{w}_{t+1} \leftarrow \mathrm{w}_t - \alpha_t \nabla_{\mathrm{w}_t} Loss(h_{\mathrm{w}_t})$$

$$\nabla_{\mathrm{w}_t} Loss(h_{\mathrm{w}_t}) = -\sum_{i=1}^{n} (y_i - h_{\mathrm{w}_t}(\mathrm{x}_i))\mathrm{x}_i$$

Classificação Linear com Regressão Logística

Versão Batch (batelada ou em lote):

- sorteie um vetor de pesos w_0 e repita o passo abaixo
- atualize o vetor w_t.

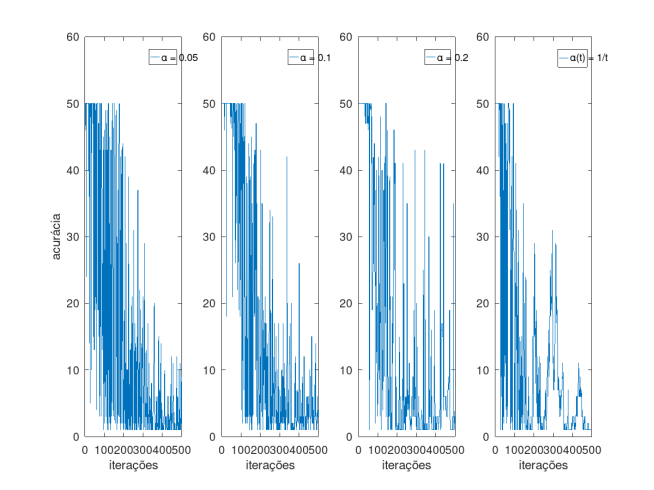
$$\mathrm{w}_{t+1} \leftarrow \mathrm{w}_t + \alpha_t \sum_{i=1}^{n} (y_i - h_{\mathrm{w}_t}(\mathrm{x}_i))\mathrm{x}_i$$

Versão Estocástica:

- sorteie um vetor de pesos w_0 e repita os passos abaixo
- sorteie aleatoriamente um exemplo $e_i = (\mathrm{x}_i, y_i)$
- atualize o vetor w_t.

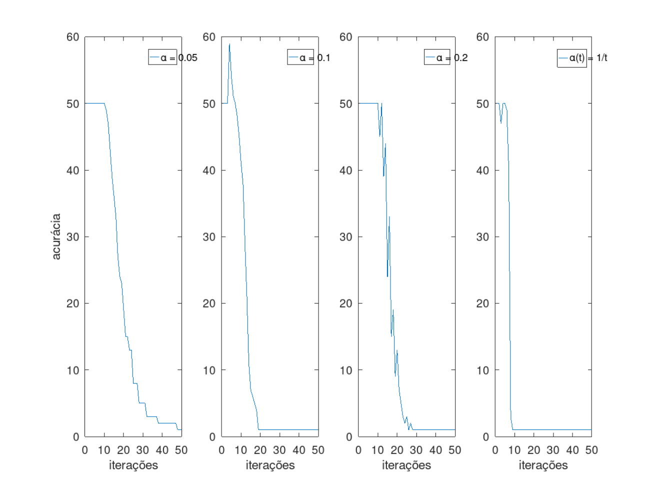
$$\mathrm{w}_{t+1} \leftarrow \mathrm{w}_t + \alpha_t (y_i - h_{\mathrm{w}_t}(\mathrm{x}_i))\mathrm{x}_i$$

Regressão Logística (1 repetição - estocástica)

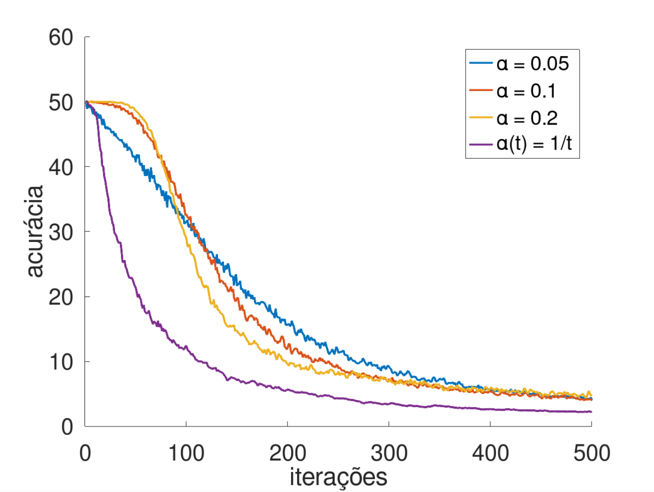


Regressão Logística (1 repetição - batelada)

α' inversamente proporcional a quantidade de exemplos ($\alpha/100$)



Regressão Logística (1000 repetições - estocástica)



Regressão Logística (1000 repetições - batelada)

α' inversamente proporcional a quantidade de exemplos ($\alpha/100$)