

Modelos Lineares

Otaviano da Cruz Neto

Instituto de Ciencias Exatas - ICEx / UFF

09/05/2018

Introdução(Notação)

► **Dados Individuais** ($X^{(i)}$ e $Y^{(i)}$)

$$X^{(i)} = (x_1, x_2, \dots, x_m) \quad (1)$$

$$Y^{(i)} = y^{(i)} \quad (2)$$

► **Dados da Amostra** ($X^{(i)}, Y^{(i)}$)

$$X = \begin{bmatrix} 1 & x_1^{(1)} & \dots & x_m^{(1)} \\ 1 & x_1^{(2)} & \dots & x_m^{(2)} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_1^{(N)} & \dots & x_m^{(N)} \end{bmatrix} \quad (3)$$

Introdução (Notação)

$$Y = \begin{bmatrix} y^{(0)} \\ y^{(1)} \\ \vdots \\ y^{(N)} \end{bmatrix} \quad (4)$$

Regressão

- ▶ **O que é o método de Regressão?**

O método de Regressão é uma ferramenta que leva em consideração a dependência entre as variáveis que caracterizam os dados . Ou seja, a cada valor de entrada é associado a um valor dado por uma Função Target, o padrão a ser aprendido.

- ▶ **Tipos de Regressão**

Os dois tipos de regressão que serão exibidos serão a Regressão Linear e a Regressão Logística. A regressão linear admite a dependência linear entre as variáveis e busca um hiperplano que melhor aproxima a configuração. Já a regressão logística utiliza a associação de cada $X^{(i)}$ a respostas ($Y^{(i)}$) binárias, por exemplo, 0 ou 1.

Regressão Linear

- ▶ **Vetor Normal à Hipótese**

$$W = [w_0, w_1, w_2, \dots, w_N] \quad (5)$$

- ▶ **Hipótese Linear**

$$h(w) = XW^T \quad (6)$$

- ▶ **Caracterização dos Erros dentro (E_{in}) da amostra e fora dela (E_{out})**

A preocupação com o erro que envolve o aprendizado de um padrão é evidente quando há a necessidade de aplicar em outros conjuntos fora da amostragem a qual foi implementado o método de regressão linear. Neste caso temos,

$$E_{in}(W) = \frac{1}{N} \sum_{n=1}^N \left(X^{(n)} W^T - Y^{(n)} \right)^2 = \frac{1}{N} \left(XW^T - Y \right)^2 \quad (7)$$

Regressão Linear

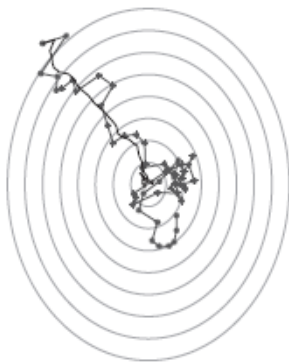
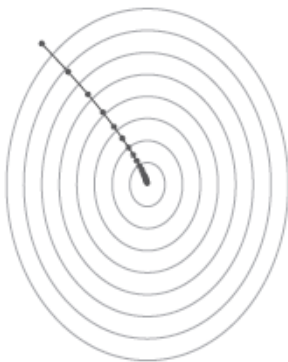
- **Gradiente de E_{in}** ($\vec{\nabla} E_{in}$)

$$\vec{\nabla} E_{in} = \frac{2}{N} X^T (XW^T - Y) \quad (8)$$

- **Gradiente decrescente (Solução Numérica)** : O método do gradiente decrescente é uma ferramenta que utiliza a propriedade do operador gradiente de apontar sempre na direção de máximo crescimento, afim de minimizar o erro dentro da amostra.

$$W_{t+1}^{\rightarrow} = W_t^{\rightarrow} - \alpha \nabla E_{in}^{\rightarrow} \quad (9)$$

Regressão Linear



Solução da Regressão Linear

► Normalização (Solução analítica)

$$\vec{\nabla} E_{in} = \frac{2}{N} X^T (XW^T - Y) = 0 \quad (10)$$

$$W^T = X^\dagger Y \quad (11)$$

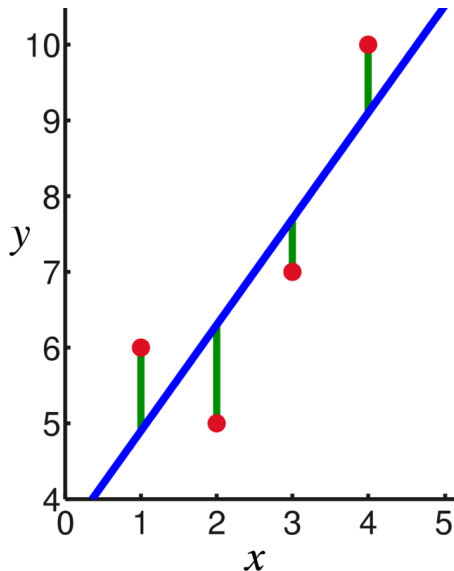
$$X^\dagger = (X^T X)^{-1} X^T \quad (12)$$

► Regressão Linear e PLA(Hipótese Inicial)

Muitas vezes na aplicação do Perceptron Learning Algorithm é inicialmente estipulado os valores do vetor normal (W) de maneira aleatória, e , por isso, pode haver uma demora em relação à convergência do algoritmo. Assim, afim de evitar tal situação, iniciar com uma hipótese que caracteriza melhor a amostragem é uma boa estratégia para diminuir o tempo de convergência. Ou seja, aplicando o algoritmo de regressão linear na amostragem e, a partir do padrão da regressão, aplicar o PLA.

Regressão Linear

► Visualização Gráfica



Interpretação Probabilística

- **Gaussiana (IID - Independente e identicamente distribuídos)**

Tomando que

$$Y_i = WX_i + \epsilon_i \quad (13)$$

Podemos expressar $\epsilon_i \approx N(0, \sigma^2)$, então a densidade de ϵ_i é dada por:

$$p(\epsilon_i) = \frac{2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Y_i - WX)^2}{2\sigma^2}\right) \quad (14)$$

Então a probabilidade(Likelihood) da Hipótese Linear é:

$$L(W) = \prod_{i=1}^N \frac{2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Y_i - WX)^2}{2\sigma^2}\right) \quad (15)$$

Interpretação Probabilística

Então derivando o $\ln L(w)$ temos:

$$l(W) = \ln L(w) \quad (16)$$

$$= \ln \prod_{i=1}^N \frac{2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Y_i - WX)^2}{2\sigma^2}\right) \quad (17)$$

$$= \sum_{i=1}^N \ln \frac{2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Y_i - WX)^2}{2\sigma^2}\right) \quad (18)$$

$$= N \ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{\sigma^2} \cdot \frac{1}{2} \sum_{i=1}^N (Y_i - WX_i)^2 \quad (19)$$

Interpretação Probabilística

Maximizando temos:

$$l(W) = \frac{1}{2} \sum_{i=1}^N (Y_i - WX_i)^2 \quad (20)$$

Aplicação

► Dados

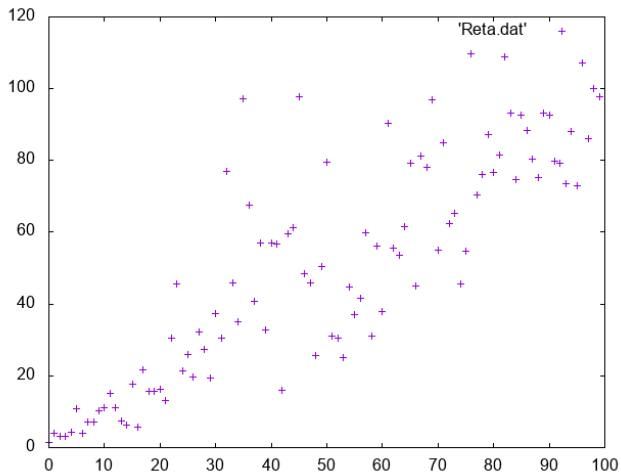


Figura 1: Dados criados a partir da reta $X=Y$.

Aplicação

► Função Custo a cada Iteração

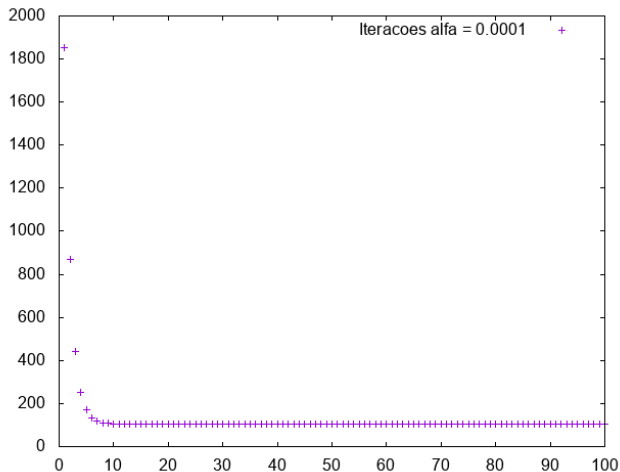


Figura 2: Gráfico de Custo por quantidade de iterações.

Aplicação

► Resultado Final

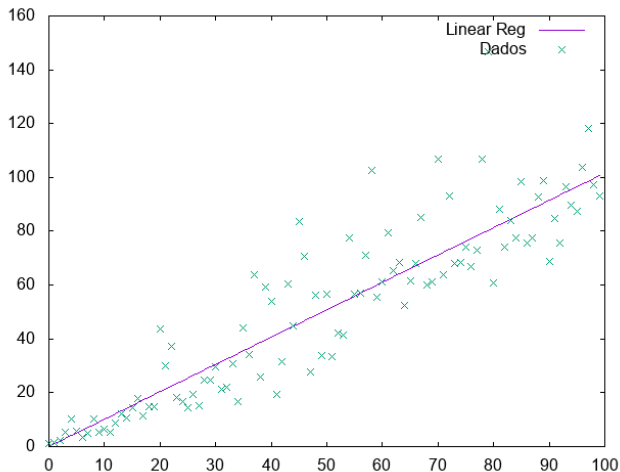


Figura 3: Gráfico da Reta de coeficiente angular 1.01 e coeficiente linear 0.01 criada a partir dos dados gerados.

Regressão Logística



Referências

[1] <http://www.portalection.com.br/analise-de-regressao>. Acesso em 04/05/2018. [2] https://www.researchgate.net/figure/A-plot-of-the-gradient-descent-algorithm-left-and-the-stochastic-gradient-descent_fig1_303257470