Inteligência Artificial
Aula 10 - vídeo 1 - Regressão Linear e Logística

João C. P. da Silva

Dept. Ciência da Computação - UFRJ

28 de outubro de 2020

Regressão Linear

- Modelo estatístico que examina a relação linear entre duas ou mais variáveis.
- Relação linear: quando uma (ou mais) variável independente cresce (ou decresce), a variável dependente cresce (ou decresce).
  - Peso x Pressão Sanguínea
  - Vendas x Anúncios
  - Horas de Estudo x Nota

Regressão Linear

- Aprendizado Supervisionado

Regressão Linear

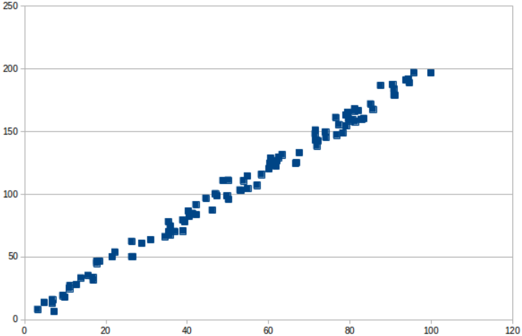
- Aprendizado Supervisionado
- Dado um conjunto de pontos $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, encontre a hipótese (reta) que "casa" com os dados.

Regressão Linear

- Aprendizado Supervisionado
- Dado um conjunto de pontos $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, encontre a hipótese (reta) que "casa" com os dados.
- Número de exemplos de treinamento: n

Regressão Linear

- Aprendizado Supervisionado
- Dado um conjunto de pontos $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, encontre a hipótese (reta) que "casa" com os dados.
- Número de exemplos de treinamento: n
- Os valores $x_1, x_2, \cdots, x_n$ são os valores de *entrada (features)* e os valores $y_1, y_2, \cdots, y_n$ são os valores de *saída (target)*.



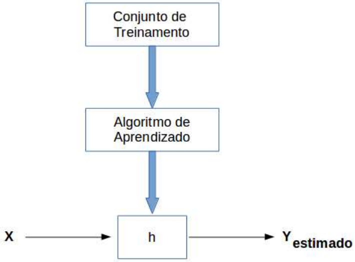
Regressão Linear

- Aprendizado Supervisionado
- Dado um conjunto de pontos $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, encontre a hipótese (reta) que "casa" com os dados.
- Número de exemplos de treinamento: n
- Os valores $x_1, x_2, \cdots, x_n$ são os valores de *entrada (features)* e os valores $y_1, y_2, \cdots, y_n$ são os valores de *saída (target)*.

X	Y
95.724162408	197.179636092
35.7576189281	67.5906695414
28.8168474238	60.8541328206
99.9584813087	196.907396981
66.8097483121	125.311128524
58.2156926413	115.785784589
53.8210763379	110.762772705
81.2960821704	157.98528569
80.6486970595	159.61941373
78.2528136925	149.003865539

Regressão Linear

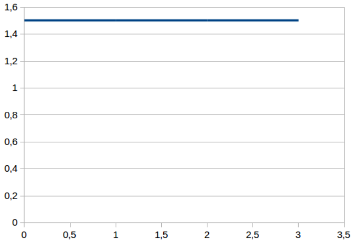
- Aprendizado Supervisionado
- Dado um conjunto de pontos $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, encontre a hipótese (reta) que "casa" com os dados.
- Número de exemplos de treinamento: n
- Os valores $x_1, x_2, \cdots, x_n$ são os valores de *entrada (features)* e os valores $y_1, y_2, \cdots, y_n$ são os valores de *saída (target)*.



Regressão Linear

- A reta que procuramos é aquela que satisfaz a equação $Y = \theta_1 * X + \theta_0$ que *melhor se aproxima dos dados*.
- Parâmetros: θ_0, θ_1

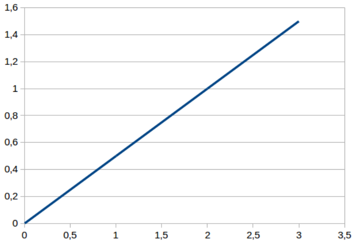
Figura: $\theta_0 = 1.5, \theta_1 = 0$



Regressão Linear

- A reta que procuramos é aquela que satisfaz a equação $Y = \theta_1 * X + \theta_0$ que *melhor se aproxima dos dados.*
- Parâmetros: θ_0, θ_1

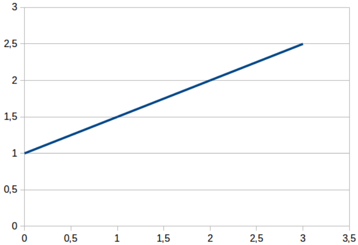
Figura: $\theta_0 = 0, \theta_1 = 0.5$



Regressão Linear

- A reta que procuramos é aquela que satisfaz a equação $Y = \theta_1 * X + \theta_0$ que *melhor se aproxima dos dados*.
- Parâmetros: θ_0, θ_1

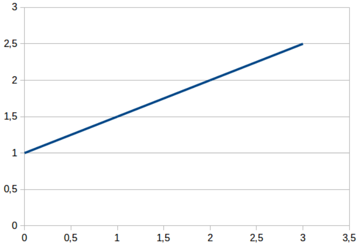
Figura: $\theta_0 = 1, \theta_1 = 0.5$



Regressão Linear

- A reta que procuramos é aquela que satisfaz a equação $Y = \theta_1 * X + \theta_0$ que *melhor se aproxima dos dados*.
- Parâmetros: θ_0, θ_1

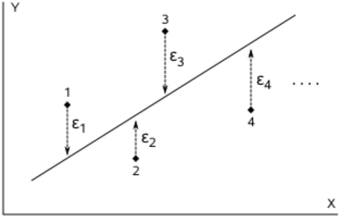
Figura: $\theta_0 = 1, \theta_1 = 0.5$



Como escolher estes parâmetros ?

Regressão Linear

- A reta que procuramos é aquela que satisfaz a equação $Y = \theta_1 * X + \theta_0$ que *melhor se aproxima dos dados*.
- Parâmetros: θ_0, θ_1



- Queremos **minimizar** o erro entre os valores reais e os valores previstos:

$$\epsilon_i = (y_{i_{real}} - y_{i_{previsto}}) \Rightarrow \text{Erro} = \sum_{i=1}^{n} \epsilon_i^2 \rightarrow 0$$

Note que $y_{i_{real}}$ vem dos dados e $y_{i_{previsto}}$ vem da equação $Y = \theta_1 * X + \theta_0$

Regressão Linear

- Queremos **minimizar** o erro entre os valores reais e os valores previstos:

$$\epsilon_i = (y_{i_{real}} - y_{i_{previsto}}) \Rightarrow \text{Erro} = \sum_{i=1}^{n} \epsilon_i^2 \to 0$$

Note que $y_{i_{real}}$ vem dos dados e $y_{i_{previsto}}$ vem da equação $Y = \theta_1 * X + \theta_0$.

- $\text{Erro} = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_{i_{real}} - y_{i_{previsto}})^2 = \sum_{i=1}^{n} (y_i - (\theta_1 * x_i + \theta_0))^2 = J(\theta_0, \theta_1)$

- Queremos *aprender* quem são θ_1 e θ_0.

$$\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_1} = \sum_{i=1}^{n} 2 * (y_i - (\theta_1 * x_i + \theta_0)) * (-x_i)$$

$$\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_0} = \sum_{i=1}^{n} 2 * (y_i - (\theta_1 * x_i + \theta_0)) * (-1)$$

Regressão Linear

Queremos *aprender* quem são θ_1 e θ_0.

$$\frac{\partial J(\theta_0,\theta_1)}{\partial\theta_1} = \sum_{i=1}^{n} 2*(y_i - (\theta_1 * x_i + \theta_0))*(-x_i) = 0 \therefore \theta_1 * \sum_{i=1}^{n} x_i^2 + \theta_0 * \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i * x_i$$

$$\frac{\partial J(\theta_0,\theta_1)}{\partial\theta_0} = \sum_{i=1}^{n} 2*(y_i - (\theta_1 * x_i + \theta_0))*(-1) = 0 \therefore \theta_1 * \sum_{i=1}^{n} x_i + \theta_0 * n = \sum_{i=1}^{n} y_i$$

$$\theta_1 * \sum_{i=1}^{n} x_i^2 + \theta_0 * \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i * x_i$$

$$\theta_1 * \sum_{i=1}^{n} x_i + \theta_0 * n = \sum_{i=1}^{n} y_i$$

Regressão Linear

Queremos *aprender* quem são θ_1 e θ_0.

$$\theta_1 * \sum_{i=1}^{n} x_i^2 + \theta_0 * \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i * x_i$$

$$\theta_1 * \sum_{i=1}^{n} x_i + \theta_0 * n = \sum_{i=1}^{n} y_i$$

Resolvendo:

$$\theta_1 = \frac{n * \sum_{i=1}^{n} y_i * x_i - \sum_{i=1}^{n} x_i * \sum_{i=1}^{n} y_i}{n * \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \qquad \theta_0 = \frac{\sum_{i=1}^{n} y_i * \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i * \sum_{i=1}^{n} y_i * x_i}{n * \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

Regressão Linear

Queremos *aprender* quem são θ_1 e θ_0.

$$\theta_1 = \frac{n * \sum_{i=1}^{n} y_i * x_i - \sum_{i=1}^{n} x_i * \sum_{i=1}^{n} y_i}{n * \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \qquad \theta_0 = \frac{\sum_{i=1}^{n} y_i * \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i * \sum_{i=1}^{n} y_i * x_i}{n * \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

ou simplificando

$$\theta_1 = \frac{\overline{xy} - \overline{x} * \overline{y}}{\overline{x^2} - \overline{x}^2} \qquad \theta_0 = \frac{\overline{y} * \overline{x^2} - \overline{x} * \overline{xy}}{\overline{x^2} - \overline{x}^2}$$

onde

$$\overline{x} = \frac{1}{n} * \sum_{i=1}^{n} x_i \qquad \overline{xy} = \frac{1}{n} * \sum_{i=1}^{n} x_i * y_i$$

Inteligência Artificial
Aula 10 - vídeo 1 - Regressão Linear e Logística

João C. P. da Silva

Dept. Ciência da Computação - UFRJ

28 de outubro de 2020