

Regressão linear multivariada



	Tamanho (pés ²)	# Quartos	# pisos	Tempo (anos)	Preço (x \$1000)
$\mathbf{x}^{(1)}$	2104	5	1	45	460
$\mathbf{x}^{(2)}$	1416	3	2	40	232
$\mathbf{x}^{(3)}$	1534	3	2	30	315
$\mathbf{x}^{(4)}$	852	2	1	36	178
$\mathbf{x}^{(n)}$
	\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	\mathbf{x}_4	\mathbf{y}



Regressão linear multivariada



$$h_{\Theta}(\mathbf{x}) = \Theta_0 + \Theta_1 * x$$

Regressão
linear univariada

$$h_{\Theta}(\mathbf{x}) = \Theta_0 * x_0 + \Theta_1 * x_1 + \Theta_2 * x_2 + \dots + \Theta_n * x_n$$

Regressão linear
multivariada

Hipótese para **uma
única instância!**



Regressão linear multivariada



$$h_{\Theta}(\mathbf{x}) = \Theta_0 + \Theta_1 * x$$

Regressão
linear univariada

$$h_{\Theta}(\mathbf{x}) = \Theta_0 * x_0 + \Theta_1 * x_1 + \Theta_2 * x_2 + \dots + \Theta_n * x_n$$

Regressão linear
multivariada

Hipótese para **uma
única instância!**

De onde saiu
esse x_0 ?



Regressão linear multivariada



$$h_{\Theta}(\mathbf{x}) = \Theta_0 + \Theta_1 * x$$

Regressão
linear univariada

$$h_{\Theta}(\mathbf{x}) = \Theta_0 * x_0 + \Theta_1 * x_1 + \Theta_2 * x_2 + \dots + \Theta_n * x_n$$

Regressão linear
multivariada

Hipótese para **uma
única instância!**

De onde saiu
esse x_0 ?

x_0 é o bias, inserido com o valor 1 para possibilitar a multiplicação de matrizes:

$$h_{\Theta}(\mathbf{x}) = \Theta^T * \mathbf{x}$$



Exercício



Dados os pesos a seguir, realize uma predição:

Instância:

Tamanho (pés ²)	# Quartos	# pisos	Tempo (anos)	Preço (x \$1000)
2104	5	1	45	460

Pesos: $\Theta_0 = 1$, $\Theta_1 = 0.2$, $\Theta_2 = 0.5$, $\Theta_3 = 0.3$, $\Theta_4 = 0.7$

Resposta: 456.1





- Atualização dos pesos em um problema multivariado:

$$\theta_j := \theta_j - \alpha \frac{1}{N} \sum_{i=1}^N (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)}$$

Não se esqueça de armazenar os novos valores em variáveis temporárias!

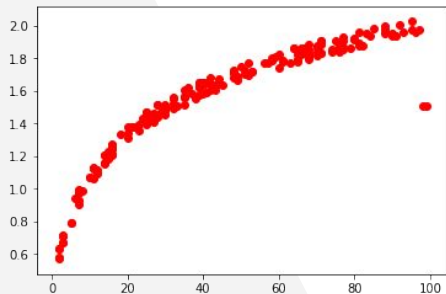
x_0 é o bias e será **sempre** 1



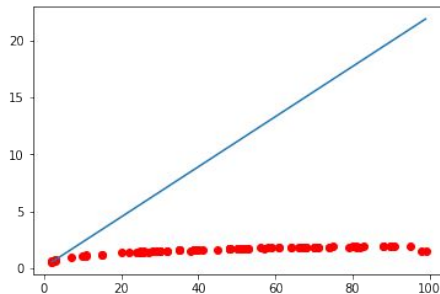
Regressão linear multivariada



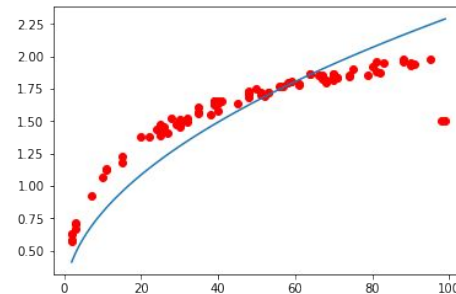
- ▶ Quando há poucos atributos ou quando os dados não são linearmente definidos, pode-se utilizar equações mais complexas com polinômios para a hipótese $h_{\theta}(x)$



X



$$h_{\theta}(x) = \theta_0 x_0 + \theta_1 x_1$$



$$h_{\theta}(x) = \theta_0 x_0 + \theta_1 \sqrt{x_1}$$

Exercício




- ▶ Exercício: Regressão Linear polinomial
- ▶ Execute o notebook e faça as tarefas listadas





Exercício: Como Outliers Afetam a Regressão Linear?

 **Objetivo:** Você recebeu um dataset real com informações sobre preços de imóveis, mas algo estranho está acontecendo... O modelo de regressão linear parece estar fazendo previsões ruins! Será que há **outliers** afetando os resultados?

◆ Passo 1: Criando o Dataset

Tarefa:

- Gere um dataset com **área do imóvel (m²)** como **variável preditora** e **preço do imóvel (R\$)** como **variável alvo**.
- Inclua um **outlier** (500m², R\$10.000) e observe como isso afeta o modelo.



Exercício



- ▶ Exercício: Regressão para um Problema do Mundo Real
- ▶ Cada grupo deve coletar um pequeno conjunto de dados reais e aplicar regressão linear. Sugestões:
 - ▶ ✓ Preço de celulares vs. Ano de lançamento.
 - ▶ ✓ Salário médio vs. Anos de experiência em uma profissão.
 - ▶ ✓ Número de seguidores em redes sociais vs. Engajamento médio (curtidas, comentários).



Exercício



- ▶ Exercício: Regressão para um Problema do Mundo Real
- ▶ Cada grupo deve coletar um pequeno conjunto de dados reais e aplicar regressão linear.
- ▶ ① Colete ao menos 5 a 10 pontos de dados reais (use sites, Google, etc.).
- ▶ ② Faça um gráfico de **dispersão** dos dados.
- ▶ ③ Use regressão linear para encontrar a equação do modelo.
- ▶ ④ Teste uma previsão: funciona bem para novos dados?
- ▶ ⑤ Escreva uma breve análise sobre o que a regressão linear pode ou não prever nesse caso.



Exercício



- ▶ Modifique o algoritmo da aula passada (para regressão linear univariada) para aceitar múltiplos atributos preditivos
- ▶ Execute o algoritmo para o dataset “Casas/house data (multidimensional).csv”
- ▶ Como o seu algoritmo se compara com a [regressão linear](#) do scikit-learn?
 - ▷ Não se esqueça de usar os mesmos parâmetros! (taxa de aprendizado, normalização, etc)





Leitura recomendada:

- ▶ Apêndice D de Introduction to Data Mining

