



Linear Regression - Teoria

☰ Ciclo	Ciclo 03: Aprendizado supervisionado - Regressão
# Aula	18
🕒 Created	@January 23, 2023 5:47 PM
☑ Done	☑
☑ Ready	☑

Objetivo da Aula:

- ☐ O que é Linear Regression?
- ☐ O processo de treinamento
- ☐ Premissas assumidas
- ☐ Resumo
- ☐ Próxima aula

Conteúdo:

▼ 1. O que é Linear Regression?

A regressão linear é um método estatístico que permite resumir e estudar a relação entre duas variáveis contínuas (quantitativas).

▼ 1.1 Tipos de relacionamento

▼ 1.1.1 Determinístico

Em um relação determinística, há uma equação que descreve exatamente a relação entre duas variáveis.

▼ 1.1.2 Estatístico

A relação estatística entre duas variáveis não é perfeita.

▼ 1.2 Como escolher o melhor modelo?

A linha que se ajuste melhor aos dados será aquela para a qual os n erros de previsão, uma para cada ponto de dados observado, são os menores possíveis.

Uma das formas de encontrar os melhores valores dos parâmetros para produzir a reta com o menor erros geral é usar os critérios dos mínimos quadrados. Esse critério minimiza a soma dos erros de previsão ao quadrado.

▼ 1.2.1 Critério dos mínimos quadrados

▼ 1.2.1.1 Exemplo 1:

▼ Dado o seguinte modelo:

$$\hat{y} = -331.2 + 7.1x$$

$$\hat{y} = a + bx$$

▼ Temos os seguintes resultados:

Peso	Altura	Previsão Altura	Erro	Erro ao quadrado
63	127	116.1	10.9	118.81
64	121	123.2	-2.2	4.84
66	142	137.4	4.6	21.16
69	157	158.7	-1.7	2.89
69	162	158.7	3.3	10.89
71	156	172.9	-16.9	285.61
71	169	172.9	-3.9	15.21
72	165	180.0	-15.0	225.00
73	181	187.1	-6.1	37.21
75	208	201.3	6.7	44.89
				766.5

▼ 1.2.1.2 Exemplo 2:

▼ Dado o seguinte modelo:

$$\hat{y} = -266.53 + 6.1376x$$

▼ Temos os seguintes resultados:

Peso	Altura	Previsão Altura	Erro	Erro ao quadrado
63	127	120.139	6.8612	47.076
64	121	126.276	-5.2764	27.840
66	142	138.552	3.4484	11.891
69	157	156.964	0.0356	0.001
69	162	156.964	5.0356	25.357
71	156	169.240	-13.2396	175.287
71	169	169.240	-0.2396	0.057
72	165	175.377	-10.3772	107.686
73	181	181.515	-0.5148	0.265
75	208	193.790	14.2100	201.924

▼ 1.2.2 Fórmula do mínimos quadrados

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

▼ 2. O processo de treinamento

▼ 2.1 Exemplo visual do funcionamento.

1.

▼ 2.2 Os 5 passos para treinar a Regressão Linear

Passo 01: Carregue os dados

Passo 02: Aplique a fórmula para encontrar (a)

Passo 03: Como o valor de (a), use a fórmula para encontrar o valor de b

Passo 04: Use o modelo da reta para prever os valores da variável resposta, a partir do modelo da reta.

Passo 05: Definir uma métrica de performance para calcular as previsões para dados nunca vistos

▼ 3. Premissas assumidas

As 4 premissas da Regressão Linear são as seguintes:

1. A relação entre as características e a variável resposta é Linear.
2. Os erros são independentes.
3. Os erros são normalmente distribuídos.
4. Os erros para cada valor previsto tem variâncias iguais.

▼ 4. Resumo

1. Existem 2 tipos de relacionamento: Determinístico e o Estatístico
2. Usamos o critério dos mínimos quadrados para determinar o modelo
3. Os 5 passos para treinar a Regressão Linear

▼ 5. Próxima aula

Linear Regression - Prática