

Regressão Linear Simples

Introdução ao tema Machine Learning e análise preditiva

Prof. Mateus Dias

Regressão Linear é uma
ferramenta da área de estatística

Definição

Regressão no sentido amplo significa estimar onde os dados são usados, suas relações.

O objetivo de maneira simples e direta é: Com “dados” históricos, é possível usar regressão para prever o futuro? (Este é o foco)

“Linear” porque os modelos que estabelecemos são lineares em termos de parâmetros. (Em suma: variáveis lineares ou não).

5 tipos muito básicos de regressão linear

- Simples
1 variável dependente e 1 variável independente
- Múltipla
1 variável dependente e 2 ou mais variáveis independentes
- Regressão logística
1 variável (dicotômica), 2+ duas independentes
- Ordinal
1 variável dependente ordinal, 1+ variáveis independentes.
- Multinomial
1 variável dependente nominal, 1+ variáveis independentes.

RLS = Regressão Linear Simples

Quando realizamos uma Análise de Regressão, estudamos a relação / comportamento entre uma variável em função de outra.

1 variável x 1 variável (1:1)

Já quando observamos o comportamento de uma variável em função de muitas outras, (1:N) ou um para muitos, dizemos que é uma RLM (Regressão linear múltipla)

Correlação

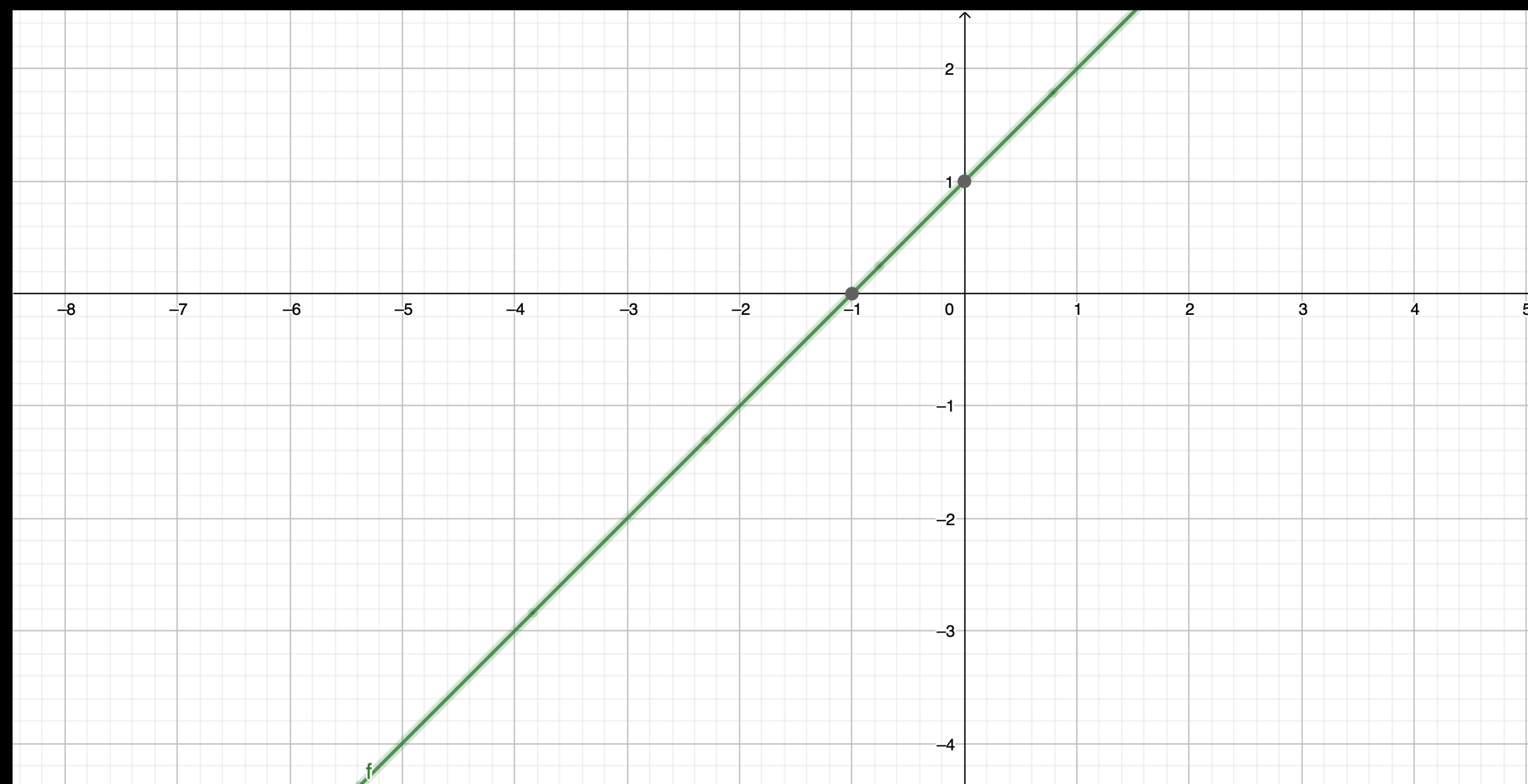
Quando estudamos física e movimento uniforme, aprendemos a correlação entre espaço e tempo e chamamos essa correlação de velocidade média, por exemplo.

Dizemos que este é um exemplo puro porque aplicado às leis da física, sempre teremos uma correlação perfeita.

Mas, e se fizéssemos uma correlação entre valor de um veículo em função do grau de dificuldade de projeto / construção? Teríamos uma correlação (ou provavelmente sim, não tão absoluta, mas possível de se observar).

RLS - Uma função de primeiro grau!

- Observe a equação da reta
- $f(x) = ax + b$ (gráfico ao lado)
- $y = ax + b$
- (Coloque as funções acima no software Geogebra para verificar o comportamento dos dois gráficos gerados)



Variável preditora e variável dependente

- Variável preditora: é aquela independente, que PODE influenciar na variável que queremos descobrir. No exemplo do valor do veículo, a variável preditora pode ser quantidade de cilindradas, consumo de combustível etc ou outra informação. Quem vai nos dizer? Os dados que dispomos. Na equação anterior, assumirá o valor de **X**.
- Variável alvo ou dependente: é a variável que queremos descobrir ou prever. No nosso exemplo, será o preço do veículo. Na equação anterior, ela terá como valor o resultado da função para cada valor de **Y**.

E o valor de A e B

- Os valores de a e b temos conhecimento quando construímos um modelo. São valores que o algoritmo deve descobrir para fazer as previsões.
- Nomeamos A e B , de coeficientes.

Ordinary Least Squares (OLS)

Como calcular os coeficientes numa função de primeiro grau? Usando o método de OLS

- O método de OLS objetiva buscar o “melhor” valor que os coeficientes possam ter, de modo que a diferença entre o valor predito pela função e o valor real, sejam os menores.

Tarefa!

Para esta aula assíncrona, leia um pouco a sós e estude este material da Profa. Márcia da USP que possui exemplos incríveis que facilitará o entendimento do assunto

<https://www.ime.usp.br/~mbranco/regressao.pdf>

Próxima aula assíncrona - Vídeo escrevendo código, explicações e exercícios de RLS em Python! Não deixem de assistir.