



Modelos de regressão

Em **estatística**, **regressão** é uma técnica que permite quantificar e inferir a relação de uma **variável dependente** (variável de resposta) com **variáveis independentes** (variáveis explicativas). A análise da regressão pode ser usada como um método descritivo da análise de dados (por exemplo, o ajustamento de curvas).



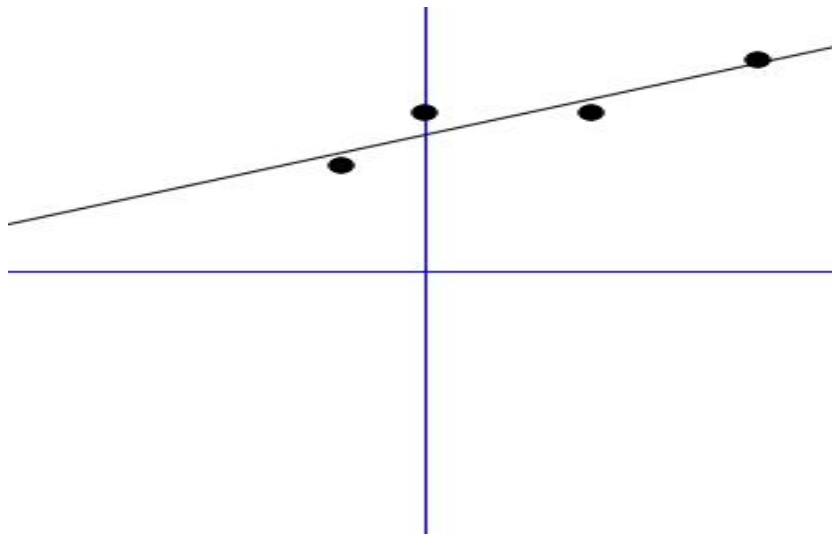
Modelos de regressão

Há vários métodos de estimação tais como **método dos mínimos quadrados**, **método dos momentos generalizado** e **logit**. A escolha do modelo depende do comportamento das variáveis e dos dados.



GPIA - Grupo de Pesquisa em Informática Aplicada
DIT - Departamento de Informática e Turismo
Me. UGO HENRIQUE PEREIRA DA SILVA

Modelos de regressão





Regressão linear

Conceito

- Regressão linear é um modelo **supervisionado** usado para prever uma variável dependente **contínua** a partir de uma ou mais variáveis independentes.
- **Exemplo:** prever o preço de um imóvel a partir de sua área e número de quartos.



GPIA - Grupo de Pesquisa em Informática Aplicada
DIT - Departamento de Informática e Turismo
Me. UGO HENRIQUE PEREIRA DA SILVA

Coeficiente de determinação (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$



◆ Conceito

- Regressão linear é um modelo **supervisionado** usado para prever uma variável dependente **contínua** a partir de uma ou mais variáveis independentes.
- **Exemplo:** prever o preço de um imóvel a partir de sua área e número de quartos.

◆ Modelo Matemático

Equação geral da regressão linear múltipla:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

- \hat{y} : valor previsto.
- β_0 : intercepto.
- β_i : coeficientes (pesos das variáveis).
- x_i : variáveis independentes.

◆ Ajuste do modelo (Método dos Mínimos Quadrados)

- Objetivo: minimizar a soma dos **erros quadráticos**:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



♦ Hipóteses da regressão linear

1. Linearidade entre X e Y.
2. Erros independentes.
3. Homocedasticidade (variância constante dos erros).
4. Normalidade dos resíduos.

Métricas de Avaliação

- **R^2 (Coeficiente de Determinação):** medida de explicação da variância.
- **MSE (Erro Quadrático Médio):** penaliza grandes erros.



Regressão Logística (Classificação Binária)

- Apesar do nome, a **regressão logística** é usada para **classificação binária**, não para regressão contínua.
- Objetivo: estimar a **probabilidade** de ocorrência de um evento $Y = 1$.

3.2 Modelo Matemático

Função logística (sigmoide):

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}$$

Essa função transforma qualquer valor real em um valor entre **0 e 1**, interpretado como **probabilidade**.



♦ Interpretação na Regressão Logística

- Em **regressão linear**, o modelo gera um valor real \hat{y} .
- Em **regressão logística**, aplicamos a função sigmoide sobre uma combinação linear das variáveis:

$$P(y = 1|X) = \sigma(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)$$

- Assim, o modelo fornece a **probabilidade de $Y = 1$** .