ACH2053 – Introdução à Estatística

Aula 09c: Regressão Linear e Logística

Valdinei Freire

valdinei.freire@usp.br

http://www.each.usp.br/valdinei

Escola de Artes, Ciências e Humanidades - USP

2025

Problema de Regressão

Definição Geral

Modelo estatístico que relaciona variáveis explicativas $\mathbf{X}$ com uma resposta Y:

$$Y = f(\mathbf{X}) + \epsilon$$

Regressão Linear

- $Y \in \mathbb{R}$
- Ex.: Preços, temperaturas

Regressão Logística

- $Y \in \{0, 1\}$
- Ex.: Classificação binária

Regressão Linear

Modelo com erro gaussiano:

$$Y = \mathbf{X}\boldsymbol{\beta} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

Solução com MLE:

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2}{2\sigma^2}}$$

$$\ell(\boldsymbol{\beta}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2$$

$$\arg\max_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} \mathsf{MSE}$$

Minimizar o Erro Quadrático Médio (MSE):

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n}(y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2 = \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2$$

Teorema de Bayes

$$p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) \propto \underbrace{p(\mathbf{y}, \mathbf{X}|\boldsymbol{\beta})}_{\text{Likelihood}} \underbrace{p(\boldsymbol{\beta})}_{\text{Prior}}$$

Prior Gaussiano (L2)

$$\beta_j \sim \mathcal{N}(0, \tau^2)$$

$$\Rightarrow \|\boldsymbol{\beta}\|_2^2$$

Prior Laplaciano (L1)

$$p(\beta_j) \propto e^{-\lambda|\beta_j|}$$

$$\Rightarrow \|\boldsymbol{\beta}\|_1$$

Modelo para $P(Y = 1|\mathbf{x})$:

$$\sigma(\mathbf{x}^T\boldsymbol{\beta}) = \frac{1}{1 + e^{-\mathbf{x}^T\boldsymbol{\beta}}}$$

Verossimilhança Bernoulli

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \sigma(\mathbf{x}_i^T\boldsymbol{\beta})^{y_i}(1 - \sigma(\mathbf{x}_i^T\boldsymbol{\beta}))^{1-y_i}$$

Comparação dos Modelos

Regressão Linear	**Regressão Logística**
MSE (Normal)	Cross-Entropy (Bernoulli)
$\|\boldsymbol{\beta}\|_2^2$ (Ridge)	$\|\boldsymbol{\beta}\|_2^2$ (L2)
$\|\boldsymbol{\beta}\|_1$ (Lasso)	$\|\boldsymbol{\beta}\|_1$ (L1)

Unificação

Ambos são casos especiais de Modelos Lineares Generalizados (GLMs)