Trabalho Prático

Estudo Comparativo de Modelos de Regressão

1 Objetivo

Realizar um estudo comparativo de quatro tipos de regressão (linear, polinomial, logarítmica e exponencial) aplicados a um dataset de seguros de saúde, avaliando qual modelo apresenta melhor desempenho na predição de custos.

2 Dataset Fornecido

Você receberá um dataset contendo informações sobre segurados e seus respectivos custos de seguro de saúde.

2.1 Descrição das Variáveis

Variável	**Tipo**	**Descrição**
age	Numérica	Idade do segurado
sex	Categórica	Sexo (male/female)
bmi	Numérica	Índice de Massa Corporal
children	Numérica	Número de filhos/dependentes
smoker	Categórica	Fumante (yes/no)
region	Categórica	Região geográfica
charges	Numérica	Custos do seguro (variável alvo)

Tabela 1: Descrição das variáveis do dataset

2.2 Identificação das Variáveis

Defina claramente:

- **Variável Dependente (Y):** Qual variável você deseja prever?
- **Variáveis Independentes (X):** Quais variáveis serão usadas como preditores?

3 Atividades a Serem Realizadas

3.1 1. Regressão Linear Múltipla

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \epsilon \quad (1)$$

3.2 2. Regressão Polinomial

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2^2 + \beta_5 X_3 + ... + \epsilon \tag{2}$$

3.3 3. Regressão Logarítmica

$$\log(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \epsilon \tag{3}$$

3.4 4. Regressão Exponencial

$$Y = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6)} \cdot \epsilon \tag{4}$$

Métricas de Avaliação:

- **R² (Coeficiente de Determinação):**

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{5}$$

- **R² Ajustado:**

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1} \tag{6}$$

- **RMSE (Root Mean Squared Error):**

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \tag{7}$$

- **MAE (Mean Absolute Error):**

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \tag{8}$$