Linear Regression

First Linear Regression Educational Project by Saba, May 2 2025

Project Goal

- Implement Linear Regression from scratch
- Practice Fundamental ML techniques
- Predict a continuous target variable from structured data

Table of Contents

Introduction

- Linear Regression is a supervised learning method
- It models the relationship between features and continuous output
- Goal: minimize the difference between predicted and true values

Data

- Diabetes patients
- Include normalized medical features (age, BMI, blood pressure)
- Target variable represents the progression of diabetes one year after baseline

Hypothesis Functions

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n$$

Cost Function

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$
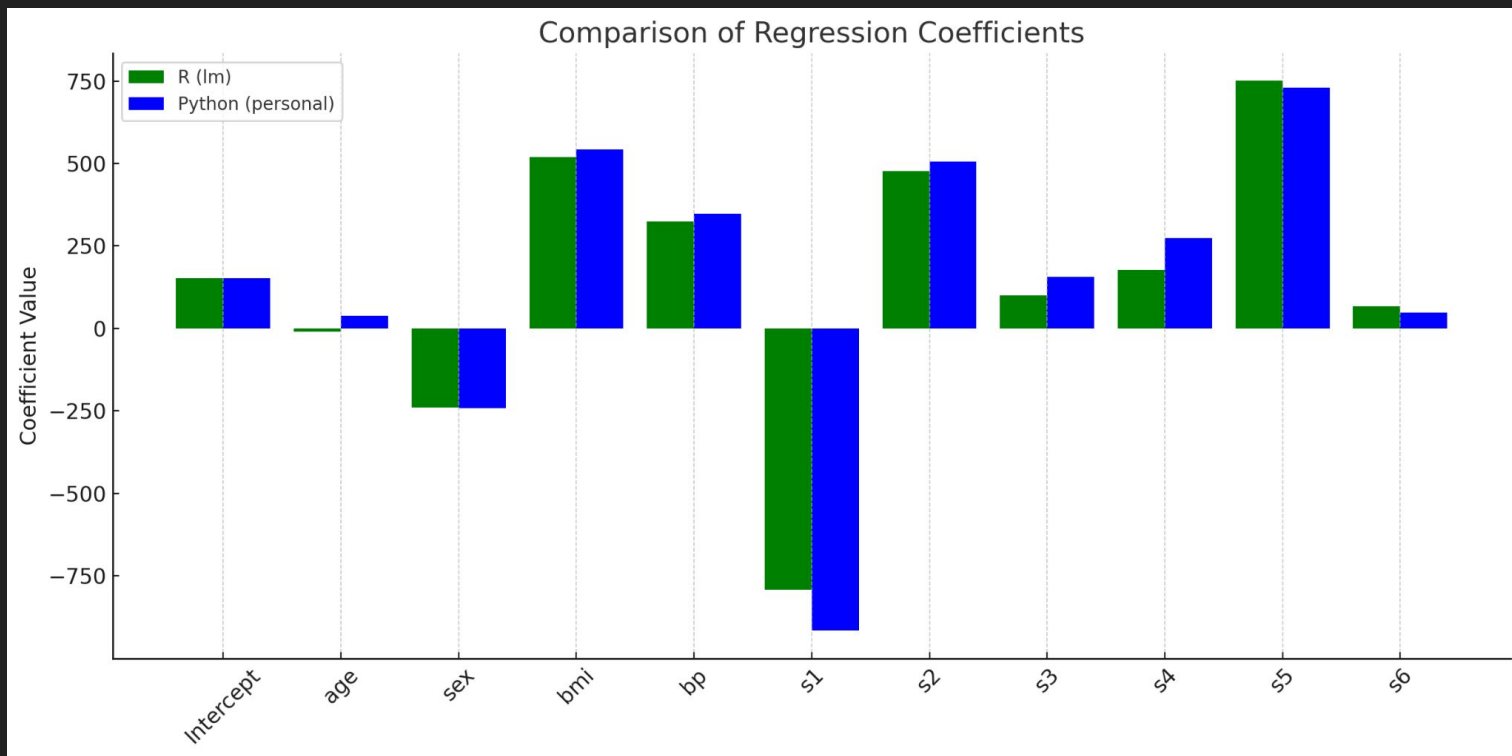
Gradient Computation

$$\frac{\partial J}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_j^{(i)}$$

Gradient Descent

$$\theta_j := \theta_j - \alpha \cdot \frac{\partial J}{\partial \theta_j}$$

Training

- Initialize weights (theta) to zeros
- Set learning rate to 0.01
- Iterate: compute gradients, update theta, compute cost
- Track cost

Comparison of Regression Coefficients

Conclusion

- Successfully implemented linear regression from scratch
- Learned key concepts: loss function, gradients, optimization
- Model predicts continuous outcomes with reasonable accuracy
- Open to feedback and improvements