Homework 2

Jian Huang

2023/10/02

Task 1: Fitting a regression line to the student debt data

My assumption:

$$y = w_0 + w_1 * x$$

x is year (input); y is the student debt (label).
Run gradient descent to find the minima of the least square cost function:

$$g(w) = \frac{1}{P}\sum_{p=1}^{P}(\dot{x}_p^T - y_p)^2$$

1. A single fingure of the original data and fitted line.

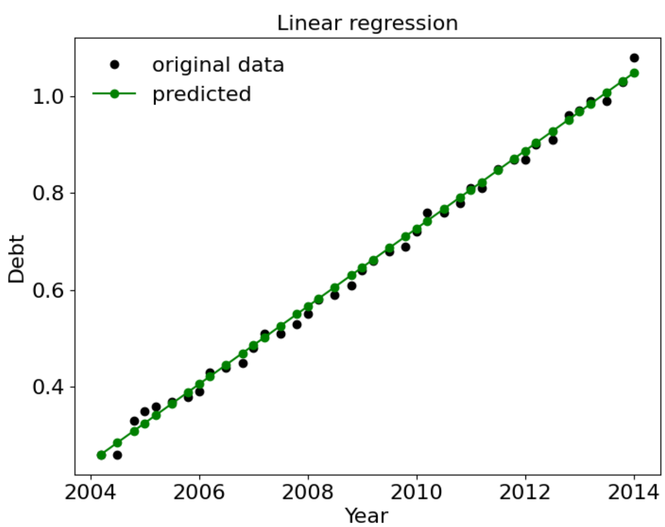

Figure 1: Linear regression of the student debt data

2. The equation of the fitted line.

params	My training	The closed-form solution
w0	-160.7253	-160.729045
w1	0.08032442	0.08032442

$$y = -160.7253 + 0.08032442 * x$$

3. The predicted debt in 2030.

$$y = -160.7253 + 0.08032442 * 2030 = 2.333$$

Task 2: Compare the Least Squares and Least Absolute Deviation costs

Using the same machine learning-based linear regression procedure in the above, the new dataset (with an outlier) was trained based on the least squares and least absolutes cost functions. Below is the least absolutes cost function:

$$g(w) = \frac{1}{P}\sum_{p=1}^{P} |(\dot{x}_p^T - y_p)|$$

1. Single figure containing the two fitted lines and the original data.

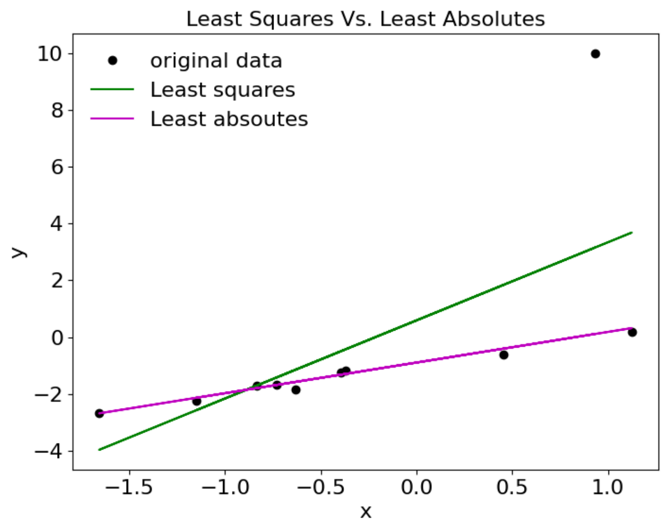

Figure 2: Comparison of two cost functions

2. Equations for the least squares and least absolutes regression lines.

params	Least Squares	Least Absolutes
w0	0.59016388	-0.89472246
w1	2.75164103	1.07923030

Therefore, the equation for the least squares regression:

$$y = 0.59016388 + 2.75164103 * x$$

the equation for the least absoutes regression:

$$y = -0.89472246 + 1.07923030 * x$$