Linear Regression

Ahnaf An Nafee

February 2021

1 Theory

1. LR Supervised Training Dataset:

   (a) Coefficients for Closed-form Linear Regression:

$$X = \begin{bmatrix} 1 & -2 \\ 1 & -5 \\ 1 & -3 \\ 1 & 0 \\ 1 & -8 \\ 1 & -2 \\ 1 & 1 \\ 1 & 5 \\ 1 & -1 \\ 1 & 6 \end{bmatrix}, Y = \begin{bmatrix} 1 \\ -4 \\ 1 \\ 3 \\ 11 \\ 5 \\ 0 \\ -1 \\ -3 \\ 1 \end{bmatrix}$$

$$w = \frac{1}{N}(X^T X)^{-1} X^T Y \tag{1}$$

$$w = \frac{1}{10}(\begin{bmatrix} 1 & -2 \\ 1 & -5 \\ 1 & -3 \\ 1 & 0 \\ 1 & -8 \\ 1 & -2 \\ 1 & 1 \\ 1 & 5 \\ 1 & -1 \\ 1 & 6 \end{bmatrix}^T \begin{bmatrix} 1 & -2 \\ 1 & -5 \\ 1 & -3 \\ 1 & 0 \\ 1 & -8 \\ 1 & -2 \\ 1 & 1 \\ 1 & 5 \\ 1 & -1 \\ 1 & 6 \end{bmatrix})^{-1} \begin{bmatrix} 1 & -2 \\ 1 & -5 \\ 1 & -3 \\ 1 & 0 \\ 1 & -8 \\ 1 & -2 \\ 1 & 1 \\ 1 & 5 \\ 1 & -1 \\ 1 & 6 \end{bmatrix}^T \begin{bmatrix} 1 \\ -4 \\ 1 \\ 3 \\ 11 \\ 5 \\ 0 \\ -1 \\ -3 \\ 1 \end{bmatrix} \tag{2}$$

$$w = \frac{1}{10}(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -2 & -5 & -3 & 0 & -8 & -2 & 1 & 5 & -1 & 6 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 1 & -5 \\ 1 & -3 \\ 1 & 0 \\ 1 & -8 \\ 1 & -2 \\ 1 & 1 \\ 1 & 5 \\ 1 & -1 \\ 1 & 6 \end{bmatrix})^{-1} \begin{bmatrix} 1 & -2 \\ 1 & -5 \\ 1 & -3 \\ 1 & 0 \\ 1 & -8 \\ 1 & -2 \\ 1 & 1 \\ 1 & 5 \\ 1 & -1 \\ 1 & 6 \end{bmatrix}^T \begin{bmatrix} 1 \\ -4 \\ 1 \\ 3 \\ 11 \\ 5 \\ 0 \\ -1 \\ -3 \\ 1 \end{bmatrix} \tag{3}$$

$$w = \frac{1}{10}(\begin{bmatrix} 10 & -9 \\ -9 & 169 \end{bmatrix})^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -2 & -5 & -3 & 0 & -8 & -2 & 1 & 5 & -1 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ -4 \\ 1 \\ 3 \\ 11 \\ 5 \\ 0 \\ -1 \\ -3 \\ 1 \end{bmatrix} \tag{4}$$

$$w = \frac{1}{10}(\begin{bmatrix} 0.10503418 & 0.00559354 \\ 0.00559354 & 0.00621504 \end{bmatrix}) \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -2 & -5 & -3 & 0 & -8 & -2 & 1 & 5 & -1 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ -4 \\ 1 \\ 3 \\ 11 \\ 5 \\ 0 \\ -1 \\ -3 \\ 1 \end{bmatrix} \tag{5}$$

$$w = \begin{bmatrix} 0.010503418 & 0.000559354 \\ 0.000559354 & 0.000621504 \end{bmatrix} \begin{bmatrix} 14 \\ -79 \end{bmatrix} \tag{6}$$

$$w = \begin{bmatrix} 0.102858919 \\ -0.041267868 \end{bmatrix} \tag{7}$$

(b) Predictions, Y , for the Training Data:

$$Y_{pred} = \begin{bmatrix} 0.18539466 \\ 0.30919826 \\ 0.22666252 \\ 0.10285892 \\ 0.43300186 \\ 0.18539466 \\ 0.06159105 \\ -0.10348042 \\ 0.14412679 \\ -0.14474829 \end{bmatrix}$$

(c) RMSE:

RMSE: -0.7071067811865475

2. Function Weights:

(a) Partial Gradients:

$J = (x_1w_1 - 5x_2w_2 - 2)^2 = (x_1w_1 - 5x_2w_2 - 2)(x_1w_1 - 5x_2w_2 - 2)$
$J = x_1^2w_1^2 - 5x_1x_2w_1w_2 - 2x_1w_1 - 5x_1x_2w_1w_2 + 25x_2^2w_2^2 + 10x_2 - 2x_1w_1 + 10x_2w_2 + 4$
$J = x_1^2w_1^2 - 10x_1x_2w_1w_2 - 4x_1w_1 + 25x_2^2w_2^2 + 10x_2 + 10x_2w_2 + 4$

$$\frac{\partial J}{\partial w_1} = 2x_1^2w_1 - 10x_1x_2w_2 - 4x_1 \tag{8}$$

$$\frac{\partial J}{\partial w_2} = -10x_1x_2w_1 + 50x_2^2w_2 + 10 + 10x_2 \tag{9}$$

(b) Value of the Partial Gradients:

$w_1 = 0 \quad x_1 = 1$
$w_2 = 0 \quad x_2 = 1$

$\frac{\partial J}{\partial w_1} = 2 \times 0^2 \times 1 - 10 \times 1 \times 1 \times 0 - 4 \times 1 = -4$
$\frac{\partial J}{\partial w_2} = -10 \times 1 \times 1 \times 0 + 50 \times 1^2 \times 0 + 10 + 10 \times 1 = 20$

2 Gradient Descent

1. Plot of epoch vs J:

2. 3D plot of w_1 vs w_2, vs J:

3. Final learned values of w_1, w_2, and J, and the number of epochs required to get there:

3 Closed Form Linear Regression