1. Theory Questions

1.

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

a. Compute the coefficients for closed-form linear regression.

$$J = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y_i}\right)^2 \quad , \hat{Y_i} = X_i * w + b$$

$$J = \frac{1}{N}\sum_{i=1}^{N}(Y_i - (X_i w + b))^2$$

$$J = \frac{1}{N}\sum_{i=1}^{N}(Y_i - (X_i w))^2 \ , condense\ bias\ term\ into\ Xi, X = \begin{bmatrix} 1 & -2 \\ 1 & -5 \\ 1 & -3 \\ 1 & 0 \\ 1 & -8 \\ 1 & -2 \\ 1 & 1 \\ 1 & 5 \\ 1 & -1 \\ 1 & 6 \end{bmatrix}$$

$$J = \frac{1}{N}\left(Y_i - (X_i w)\right)^T * \left(Y_i - (X_i w)\right), using\ linear\ algreba$$

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

$$\frac{dJ}{dW} = \frac{1}{N}\left(0 - (Y^T X)^T - X^T Y + 2X^T X w\right) = \frac{1}{N}\left(2X^T X w - 2X^T Y\right)$$

$$W = (X^T X)^{-1} X^T Y$$

$$(X^T X)^{-1} = \begin{bmatrix} 0.10503418 & 0.00559354 \\ 0.00559354 & 0.00621504 \end{bmatrix}$$

$$(X^T X)^{-1} X^T = \begin{bmatrix} 0.0938 & 0.0771 & 0.0883 & 0.105 & 0.0603 & 0.0938 & 0.1106 & 0.133 & 0.0994 & 0.1386 \\ -0.0068 & -0.0255 & -0.0131 & 0.0056 & -0.0441 & -0.0068 & 0.0118 & 0.0367 & -0.0006 & 0.0429 \end{bmatrix}$$

$$(X^T X)^{-1} X^T Y = \begin{bmatrix} 1.0286 \\ -0.4127 \end{bmatrix}$$

$$\boldsymbol{b = 1.02858919}, \qquad \boldsymbol{w = -0.41267868}$$

b. What Is $\hat{Y}$?

$$\hat{Y}_i = X_i * w + b$$

$$\hat{Y}_i = \begin{bmatrix} 1.85394655 \\ 3.0919826 \\ 2.26662523 \\ 1.02858919 \\ 4.33001865 \\ 1.85394655 \\ 0.6159105 \\ -1.03480423 \\ 1.44126787 \\ -1.44748291 \end{bmatrix}$$

c. What is RMSE?

$$\boldsymbol{RMSE} = \sqrt[2]{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2} = \mathbf{3.70132592}$$

2. For the function J, where w = [w1, w2] are our weights to learn:

$$J = (x_1 w_1 - 5x_2 w_2 - 2)^2$$

$$J = (x_1 w_1 - 5x_2 w_2 - 2) * (x_1 w_1 - 5x_2 w_2 - 2)$$

$$J = w_1^2 x_1^2 - 10w_1 x_1 w_2 x_2 - 4w_1 x_1 + 25w_2^2 x_2^2 + 20x_2 w_2 + 4$$

a. What are the partial gradients $\frac{dJ}{dw_1}, \frac{dJ}{dw_2}$?

$$\frac{dJ}{dw_1} = 2w_1 x_1^2 - 10x_1 w_2 x_2 - 4x_1 + 0 + 0 + 0$$

$$\frac{dJ}{dw_1} = 2w_1 x_1^2 - 10x_1 w_2 x_2 - 4x_1$$

$$\frac{dJ}{dw_2} = 0 - 10x_1 w_1 x_2 - 0 + 50w_2 x_2^2 + 20x_2 + 0$$

$$\frac{dJ}{dw_2} = 50w_2 x_2^2 - 10x_1 w_1 x_2 + 20x_2$$

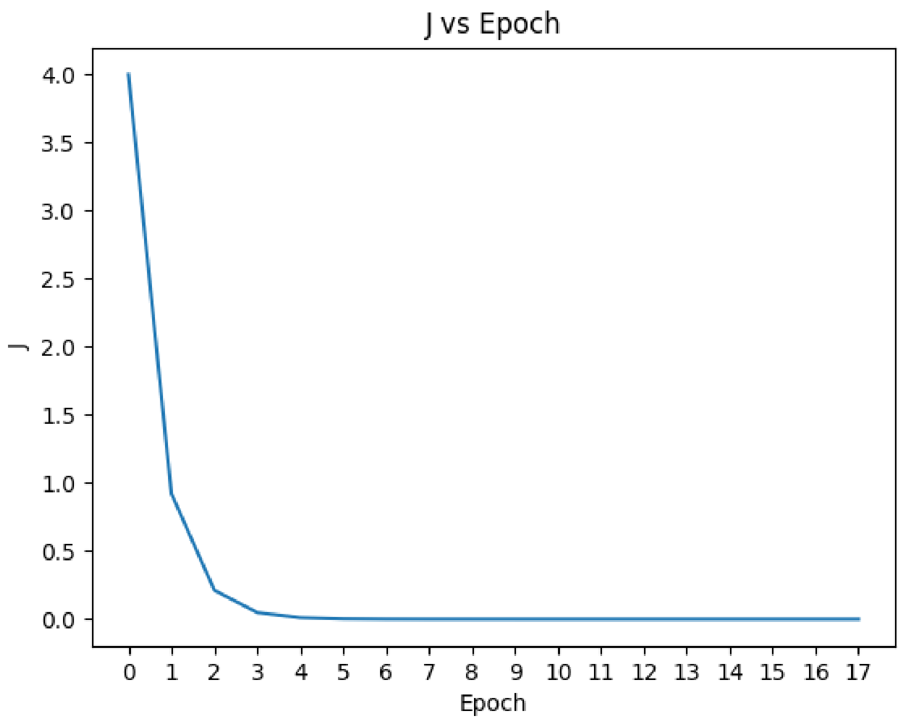
b. What are the values of the partial gradients when w = [0,0] and x = [1,1]?

$$\frac{dJ}{dw_1} = 2 * 0 * 1 - 10 * 1 * 0 - 4 * 1 = -4$$

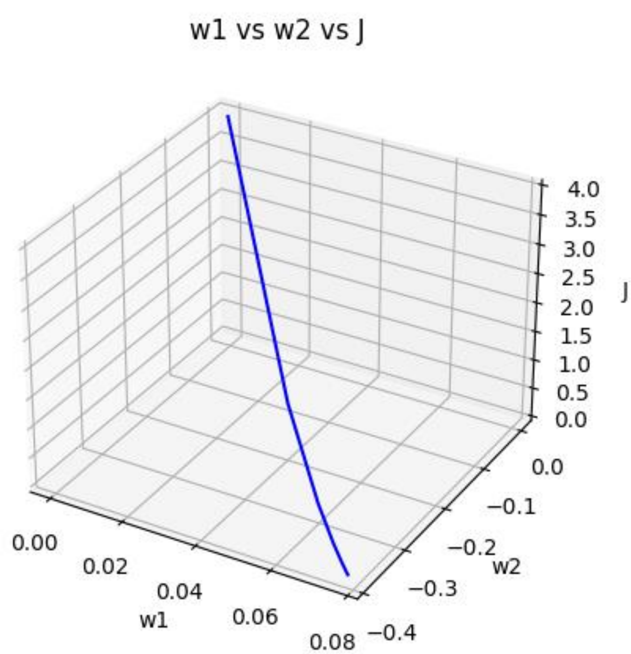
$$\frac{dJ}{dw_2} = 50 * 0 * 1 - 10 * 0 * 1 + 20 * 1 = 20$$

2. **Gradient Descent**

a. Epoch Vs J



b. w_1 vs w_2 vs J



c. Final w_1: 0.07692278372731857
Final w_2: -0.38461391863659294
Final J: 5.8111496830407836e-11
Epochs: 17

3. Closed Form Linear Regression

Model 1 RMSE For Training Set: 6080.390950336484
Model 1 RMSE For Testing Set: 7024.280057720885

Model 2 RMSE For Training Set: 5757.954440690525
Model 2 RMSE For Testing Set: 6606.030095968515

Model 3 RMSE For Training Set: 5757.888992248822
Model 3 RMSE For Testing Set: 6604.31622177858

Model 4 RMSE For Training Set: 5757.888992248821
Model 4 RMSE For Testing Set: 6604.316221778578