CS 383 Assignment 3

Kevin Tayah (kst46@drexel.edu)

February 27, 2021

1. Theory Questions

1. a) To calculate the coefficients using the Least Square Method, we must average up the following equations for all the observations:

$$w = 2x(\hat{y} - y)$$
$$b = 2(\hat{y} - y)$$

Here are all the weights and bias pair for each observation (x, y) pair where since we not have gone through any iterations yet, will assume $\hat{Y}$ is 0 for all observations:

$$\begin{pmatrix} w & b \\ 2 & -2 \\ -20 & 8 \\ 3 & -2 \\ 0 & -6 \\ 88 & -22 \\ 10 & -10 \\ 0 & 0 \\ 5 & 2 \\ -3 & 6 \\ -6 & -2 \end{pmatrix}$$

We can sum these up to get an average w and average b, $w = 7.9$ and $b = -2.8$.

b) Using the model above, we can create a function $J = w * x + b$ to predict a Y value. If we were to compute the predicted values based on this function, we get:

$$\begin{pmatrix} -18.6 \\ -42.3 \\ -26.5 \\ -2.8 \\ -66 \\ -18.6 \\ 5.1 \\ 36.7 \\ -10.7 \\ 44.6 \end{pmatrix}$$

c) The RMSE for this training set can be calculated via the following formulate:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(\hat{Y} - Y)^2}$$

The answer we get after using this formulate is, 35.3917222

2. a) Given the least squares equation $J = (x_1 w_1 - 5x_2 w_2 - 2)^2$, the partial derivatives of $\frac{\partial J}{\partial w_1}$ and $\frac{\partial J}{\partial w_2}$ are:

$$\frac{\partial J}{\partial w_1} = 2(x_1 w_1 - 5x_2 w_2 - 2)\frac{\partial}{\partial w_1}(x_1 w_1 - 5x_2 w_2 - 2) \text{ Chain rule}$$
$$\frac{\partial J}{\partial w_1} = 2(x_1 w_1 - 5x_2 w_2 - 2)(x_1)$$
$$\frac{\partial J}{\partial w_1} = 2x_1(x_1 w_1 - 5x_2 w_2 - 2)$$

$$\frac{\partial J}{\partial w_2} = 2(x_1 w_1 - 5x_2 w_2 - 2)\frac{\partial}{\partial w_2}(x_1 w_1 - 5x_2 w_2 - 2) \text{ Chain rule}$$
$$\frac{\partial J}{\partial w_2} = 2(x_1 w_1 - 5x_2 w_2 - 2)(-5x_2)$$
$$\frac{\partial J}{\partial w_2} = -10x_2(x_1 w_1 - 5x_2 w_2 - 2)$$

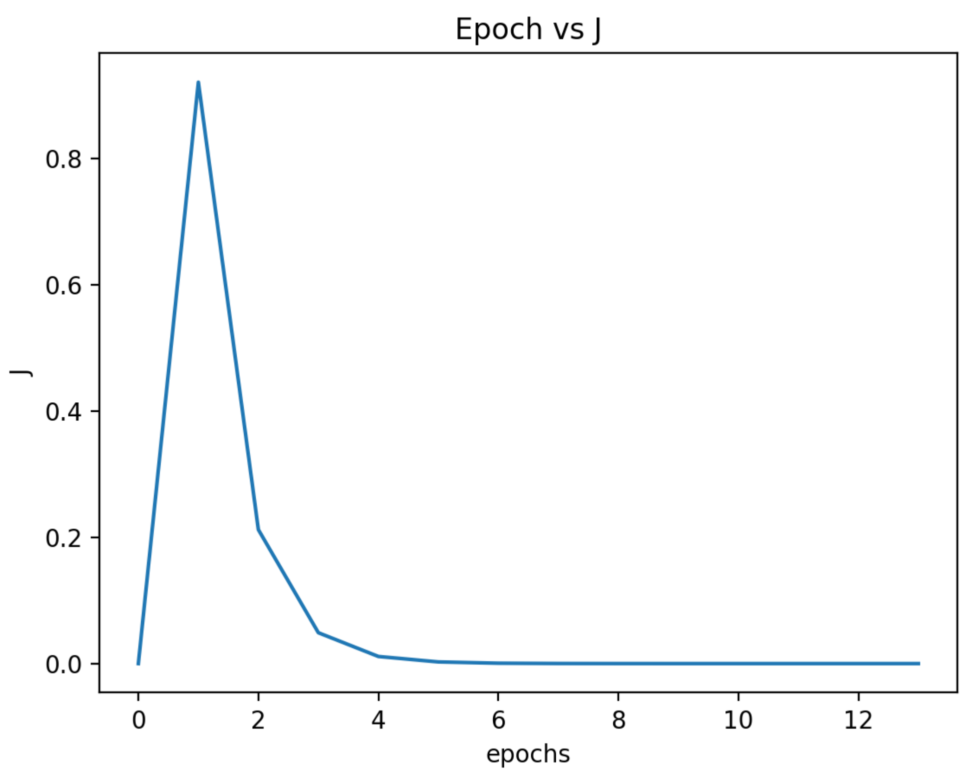
b) Given $w = [0, 0]$ and $x = [1, 1]$ the values of $\frac{\partial J}{\partial w_1}$ and $\frac{\partial J}{\partial w_2}$ are:

$$\frac{\partial J}{\partial w_1} = 2(1)(1(0) - 5(1)(0) - 2)$$
$$\frac{\partial J}{\partial w_1} = -4$$

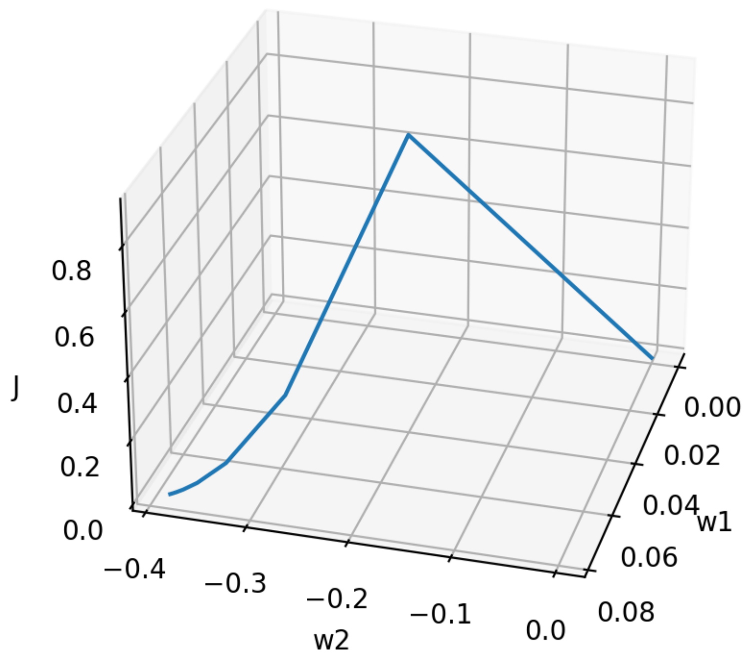
$$\frac{\partial J}{\partial w_2} = -10(1)(1(0) - 5(1)(0) - 2)$$
$$\frac{\partial J}{\partial w_2} = 20$$

2. Gradient Descent

1. Epoch vs J



2. J vs w_1 vs w_2



3. It took 13 epochs to reach a final value of $J = 2.0622065555048642{*}10^{-8}$ and $w = [0.07691755, -0.38458777]$

3. Closed Form Linear Regression

Below you will find a list of the information given from the 4 models using preprocessing methods 1-4:

- Preprocessing method 1:
  Training information
  weights: [218.13832263, -57.41278231, 36.90871322, 330.47728465, 20059.24358413, -441.31754793]
  bias: 0
  RMSE: 6243.984107190172
  Validation info
  RMSE: 7083.475724339989

- Preprocessing method 2:
  Training information
  weights: [232.989874, -10.24718970, 108.464153, 381.427337, 2054.04450, -408.453093]
  bias: -3080.058789522787

RMSE: 6065.34049514213

Validation info
RMSE: 6902.128449993521

- Preprocessing method 3:
  Training information
  weights: [222.98189153, -52.71680395, 32.77075058, 360.80795316, 20203.87547882, -729.2321168, -1098.2996296, -325.72647028, -1195.82469336]
  bias: 0
  RMSE: 6199.074047504609
  Validation info
  RMSE: 7059.13695361966

- Preprocessing method 4:
  Training information
  weights: [236.268979, 1.84364657, 102.512374, 404.385673, 20567.4969, -556.140568, -929.900220, -355.069324, -1128.28129]
  bias: -2969.391404282822
  RMSE: 6049.845528829296
  Validation info
  RMSE: 6901.06854334101