CS 383 - Machine Learning

Assignment 3 - Linear Regression

1 Theory

1. (10pts) Consider the following supervised dataset:

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

Compute the coefficients for the linear regression using least squares estimate (LSE). Show your work and remember to add a bias feature. Compute this model using **all** of the data (don't worry about separating into training and testing sets).

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

$$(X^T X) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -2 & -5 & -3 & 0 & -8 & -2 & 1 & 5 & -1 & 6 \end{bmatrix} \begin{bmatrix} -2 \\ -5 \\ -3 \\ 0 \\ -8 \\ -2 \\ 1 \\ 5 \\ -1 \\ 6 \end{bmatrix} = \begin{bmatrix} 10 & -9 \\ -9 & 169 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{(10 * 169) - ((-9) * (-9))} = \frac{1}{1609} \begin{bmatrix} 169 & 9 \\ 9 & 10 \end{bmatrix} = \begin{bmatrix} \frac{169}{1609} & \frac{9}{1609} \\ \frac{9}{1609} & \frac{10}{1609} \end{bmatrix}$$

$$(X^TY) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -2 & -5 & -3 & 0 & -8 & -2 & 1 & 5 & -1 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ -4 \\ 1 \\ 3 \\ 11 \\ 5 \\ 0 \\ -1 \\ -3 \\ 1 \end{bmatrix} = \begin{bmatrix} 14 \\ -79 \end{bmatrix}$$

$$(X^TX)^{-1}(X^TY) = \begin{bmatrix} \frac{169}{1609} & \frac{9}{1609} \\ \frac{9}{1609} & \frac{10}{1609} \end{bmatrix} \begin{bmatrix} 14 \\ -79 \end{bmatrix} = \begin{bmatrix} 1.028 \\ -0.400 \end{bmatrix}$$

For the function $J = (x_1 + x_2 - 2)^2$, where x_1 and x_2 are a single valued variables (not vectors):

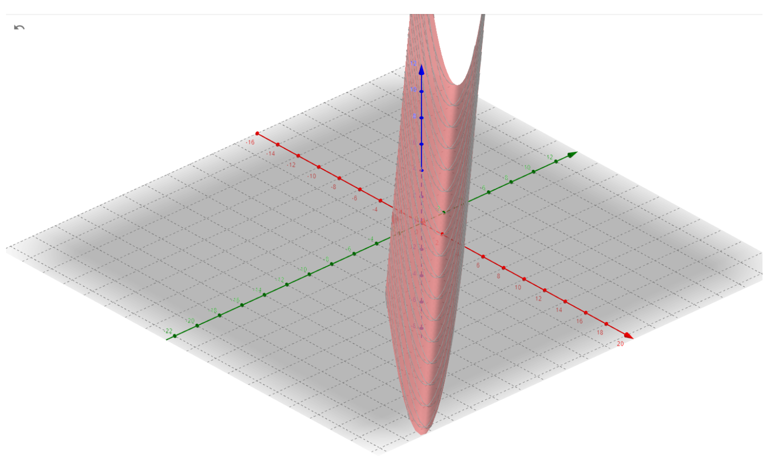
(a) What are the partial gradients, $\frac{\partial J}{\partial x_1}$ and $\frac{\partial J}{\partial x_2}$? Show work to support your answer. (4pts)

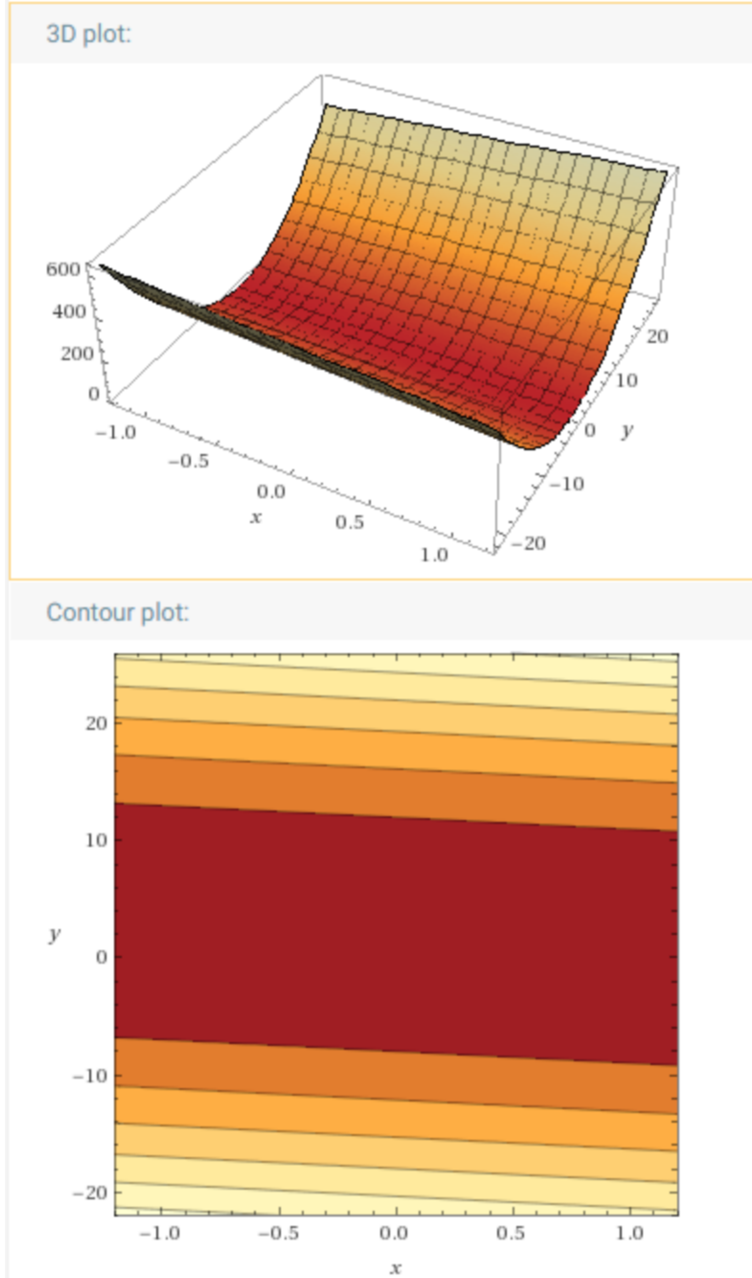
The square rule for partial derivatives state that for a function raised to the power of two, its derivative will be that function times 2. Ex: $\mathrm{x}^2 = 2x$

$$\frac{\partial J}{\partial x_1} = 2(\mathrm{x}_1 + x_2 - 2)$$

$$\frac{\partial J}{\partial x_2} = 2(\mathrm{x}_1 + x_2 - 2)$$

(b) Create a 3D plot of x_1 vs x_2, vs J use a software package of your choosing. (4pts)





(c) Based on your plot, what are the values of x_1 and x_2 that minimize J? (2pts)

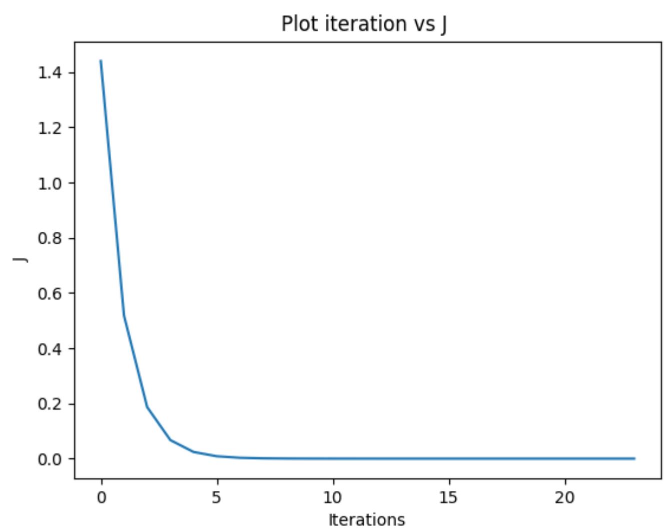
in the form [x_1, x_2] the values that minimize J are: [1,1] [2,0] [0,2]

2 Gradient Descent

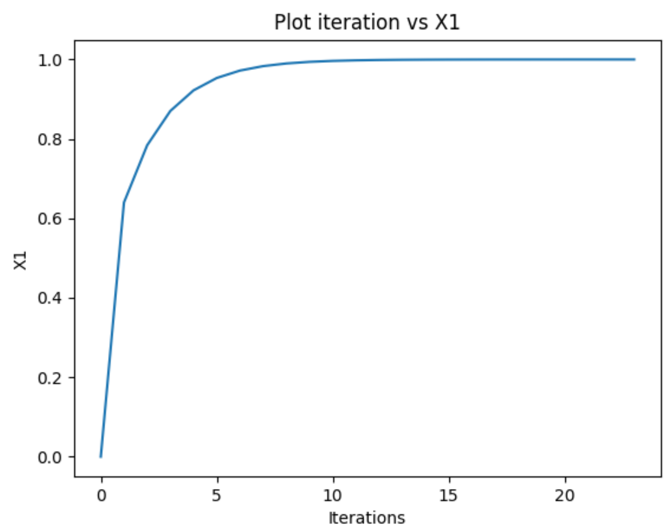
In this section we want to visualize the gradient descent process on the function $J = (x_1 + x_2 - 2)^2$. You should have already derived (pun?) the gradient of this function in the theory section. To bootstrap the process, initialize $x_1 = 0, x_2 = 0$ and terminate the process when the change in the J from one iteration to another is less that 2^{-32}. In addition, we'll use a learning rate of $\eta = 0.1$.

In your report you will need

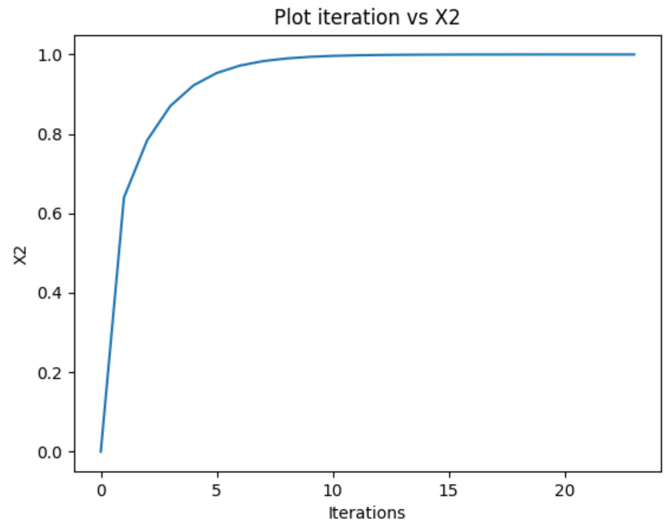
(a) Plot iteration vs J



(b) Plot iteration vs x_1



(c) Plot iteration vs x_2


Plot iteration vs X2
1.0
0.8
0.6
0.4
0.2
0.0
X2
0
5
10
15
20
Iterations

3 Closed Form Linear Regression

(a) The final model in the form $y = \theta_0 + \theta_1 x_1 + ...$

g(x) = [-142.97479460381078] + [4.6105534994738235]x_1 + [0.031787731706128336]x_2

(b) The root mean squared error.

Root Mean Squared Error: 20.19215923804834

4 S-Folds Cross-Validation

(a) The average and standard deviation of the root mean squared error for $S = 2$ over the 20 different seed values..

When S = 2
Mean RMSE = 21.698287881816135
Standard Deviation = 1.505806466802261

(b) The average and standard deviation of the root mean squared error for $S = 4$ over the 20 different seed values.

When S = 4
Mean RMSE = 21.30202763614809
Standard Deviation = 0.8171386807576563

(c) The average and standard deviation of the root mean squared error for $S = 22$ over 20 different seed values.

When S = 22
Mean RMSE = 21.01623896621227
Standard Deviation = 0.21324491972491713

(d) The average and standard deviation of the root mean squared error for $S = N$ (where N is the number of samples) over 20 different seed values. This is basically *leave-one-out* cross-validation.

When S = 44
Mean RMSE = 21.004621083284924
Standard Deviation is = 2.7519201823675253e-15