CS 383 HW 1

Kevin Karnani

April 19 2021

1 Theory

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

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

   $$\theta = \begin{bmatrix} 10 & -9 \\ -9 & 169 \end{bmatrix}^{-1} \begin{bmatrix} 14 \\ -79 \end{bmatrix} = \begin{bmatrix} 0.10503418 & 0.00559354 \\ 0.00559354 & 0.00621504 \end{bmatrix} \begin{bmatrix} 14 \\ -79 \end{bmatrix} = \begin{bmatrix} 1.02858919 \\ -0.41267868 \end{bmatrix}$$

   (b) Here is a screenshot of the result given by `sklearn`:

```
regr = linear_model.LinearRegression()

regr.fit(X, Y)
print(regr.coef_, regr.intercept_)
```

```
[[ 0.          -0.41267868]] [1.02858919]
```

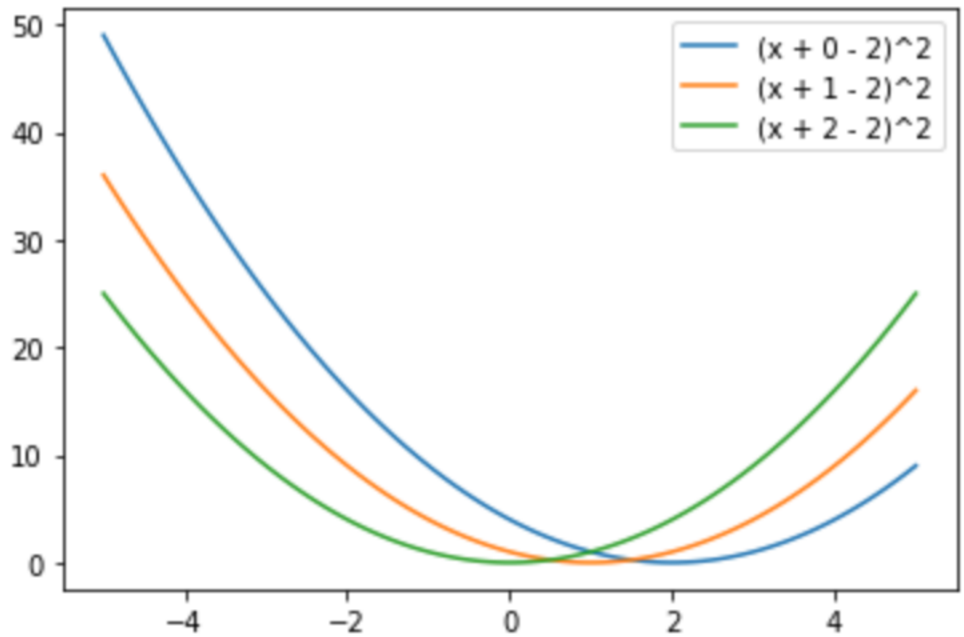
   Clearly, this matches the computations shown above in (a).

2. 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 $\frac{\partial J}{\partial x_2}$?

$$\frac{\partial J}{\partial x_1} = \frac{\partial}{\partial x_1}\Big[(x_1+x_2-2)^2\Big] = 2(x_1+x_2-2)\frac{\partial}{\partial x_1}\Big(x_1+x_2-2\Big) = 2(x_1+x_2-2).$$
$$\frac{\partial J}{\partial x_2} = \frac{\partial}{\partial x_2}\Big[(x_1+x_2-2)^2\Big] = 2(x_1+x_2-2)\frac{\partial}{\partial x_2}\Big(x_1+x_2-2\Big) = 2(x_1+x_2-2).$$

(b) Create a 2D plot of x_1 vs J `matplotlib`, for fixed values of x_2 at 0, 1, and 2.



(c) Based on your plots, what are the values of x_1 and x_2 that minimize J?
$(2, 0)$; $(1, 1)$; $(0, 2)$. In essence, J is minimized when we pick x_1 and x_2 along the plane $x_1 + x_2 = 2$.