Data Driven Optimization - Tutorial 3

A.K. Cherukuri

April 25, 2023

1 Gradient Descent

We will use the same dataset as we did for linear regression in the first tutorial, called 'startup data.csv'. We will again perform a linear regression exercise, but we will now go into detail regarding how we can use gradient descent to solve this problem iteratively. Recall the linear regression equation:

$$h(x, \theta) = \theta_0 + \theta_1 x. \tag{1.1}$$

Before you start coding, please be sure about the following:

- We will use the mean squared error as our cost function, so express the cost function C for this problem.
- Compute the partial derivatives:

$$\frac{\partial C}{\partial \theta_0} \text{ and } \frac{\partial C}{\partial \theta_1}.$$

- What do we need, further, to update the parameter θ at each iteration?

Now, in Matlab:

1. Load and normalize the data, set your x_i and y_i variables. We'll use the RD spent as our independent variable and the profit as our dependent variable.
2. Visualize the data by creating a scatter plot.
3. Define the functions you'll need for your update step. You can use function handles for this, which you can later call for the given variables.
4. Create the update step with a for loop that runs for a fixed amount of iterations.
5. Create 2 plots:
   (a) Plot $h(x, \theta)$ in the scatter plot.
   (b) Plot the evolution of the cost function over iterations.
6. Now, change the for loop of your code to a while loop and define a fixed tolerance you want to reach before the loop can stop. Could it be more efficient to code it like this?