Programming Assignment 1: Univariate Linear Regression

Introduction

In this assignment you are required to find online a dataset suitable for the use with univariate linear regression. Or, you can generate this data using a linear equation with random noise so you do not get perfect linear data.

You have to write your own code in Python using Google Colaboratory. To do so, you have to have a Gmail account. Upload the data as csv file to your Google Drive, and then generate a public link for it that anyone can access it. Use this link to download the data into your Google Colaboratory machine.

You are asked to implement the Gradient Descent algorithm that we have explained in class to find the parameters (θ_0 & θ_1) of a univariate linear regression model ($h_\theta(x) = \theta_0 + \theta_1 x$) that fits your data.

Through solving this assignment, make sure to show in the Jupyter notebook you will submit the following:

- Plot the data points (scatter points) and the linear fit (solid line) you got in the same plot.
- Plot the iteration number (x-axis) vs the cost function $J(\theta)$ (y-axis) value for three value of the learning rate: small one leading to too many iteration for convergence, good one leading to reasonable convergence, and large one causing divergence (you will get super large values of $J(\theta)$). All these should appear on the same plot.
- Create a section within your notebook to write the values of θ_0 & θ_1, and R^2 (asquared is a statistical measure of fit that indicates how much variation of a dependent variable is explained by the independent variable(s) in a regression model)

In all of your plots, make sure to have labels for the x-axis and y-axis, a legend box, and a title. Also make sure to use text cells in the Juypter notebook to document each portion of the code and each plot. Points will be given for proper documentation within the notebook.