Programming Challenge

Please submit the following:

1. Jupyter notebook file and related source files.
2. A README file where you mention what programming language you used, operating system, computer architecture and step-by-step compile and/or execute instructions.
3. A one-page document explaining your approaches and assumptions.

You can use either Python or R language for this challenge. This programming challenge must be submitted before 03/17/2020.

Dataset:

1. Please download the CSV file located at https://dpr.box.com/s/noj2pt36x2divmw9nezlx0s17q90qfif
2. For this programming assignment you will be required to perform regression on the target attributes.

Problems:

1. Load the data into memory. Briefly explain the dataset e.g. How many samples are there in the dataset? How many attributes per sample did you see?
2. Assume we have samples and attributes per sample. Then make an appropriate matrix and vector.
3. Prepare 2 datasets from the data you loaded in memory:
    a. Split the data at random into set A containing 80% of the samples which will be used for training.
    b. Split the data at random into set B containing 20% of the samples which will be used for testing.
4. For each of the A, B train datasets above, solve for the linear regression hypothesis (without a regularizer), predict the target values and plot the regression.
5. Repeat problem 4 with L2 regularization.
6. For each of the training datasets A, B perform nørmalization on each of the attributes, and repeat problem 4 and 5. For normalization please adapt the “standardization” policy.