ACTU PS5841 Data Science in Finance & Insurance - Autumn 2021 (Y. Wang)
Assignment - 4
Assigned 10/19/21, Due 10/26/21

Problem 1. Gradient Descent
The file data.csv contains a small sample from the unobserved true data generating model,

$$y = 10 + 20x_1 + 30x_2 + \epsilon, \quad \epsilon \sim N(\mu = 0, \sigma^2 = 4)$$

You decide to use the data to fit a linear regression model, assuming homoscedasticity:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

A. Use code to calculate and report the theretical LSE value of $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ as well as the sum of squared residuals $\sum_i (y_i - \hat{y}_i)^2$

B. Use a simple neural network with no hidden layers to perform the same regression involving the loss function

$$R = \sum_i (y_i - \hat{y}_i)^2$$

Please write your own code from scratch and refrain from using existing deep learning packages.

Use gradient descent or stochastic gradient descent, with the initial weights $\boldsymbol{\beta}^{(0)} = (0.01, 0.01, 0.01)^T$ as well as a fixed learning rate of your choice, to estimate the weights $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ of the neural network.

Report the following at convergence:
b1. The smallest loss your network is able to achieve
b2. The corresponding learning rate
b2. The corresponding estimated "optimal" weights
b3. The corresponding number of updates when your network reaches the "optimal" estimates
b4. A plot of the final 50 (or a reasonable number of) updates of the loss values vs the update number, in the following format