GitHub Code Repository: https://github.com/pharmon0/Harmon_ECGR5105

Problem 1)

Develop a code that run linear regression with gradient decent algorithm for each of the explanatory variables in isolation. In this case, you assume that in each iteration, only one explanatory variable (either X1, or X2, or X3) is explaining the output. Basically, you need to do three different training, one per each explanatory variable. For the learning rate, explore different values between 0.1 and 0.01 (your choice). Initialize your parameters to zero (theta to zero).

1. **Report the linear model you found for each explanatory variable.**

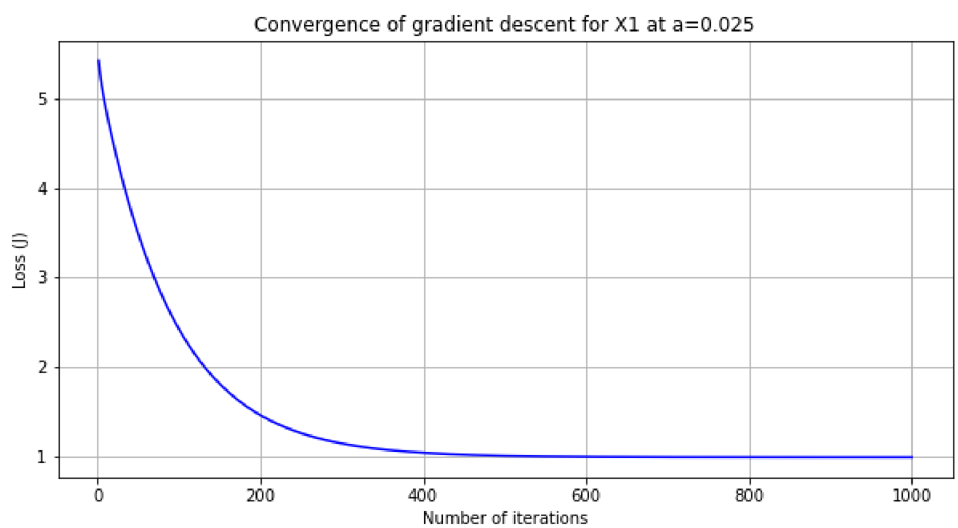
   The gradient-descent learned model for each variable can be found below, with values rounded to three decimal places.

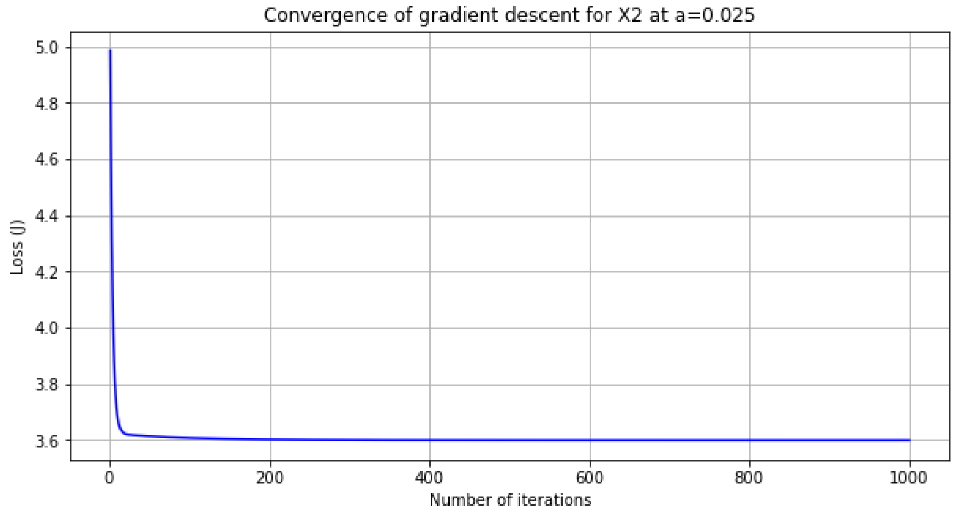
$$h(x_1) = -2.038\, x_1 + 5.928$$
$$h(x_2) = 0.558\, x_2 + 0.736$$
$$h(x_3) = -0.520\, x_3 + 2.871$$

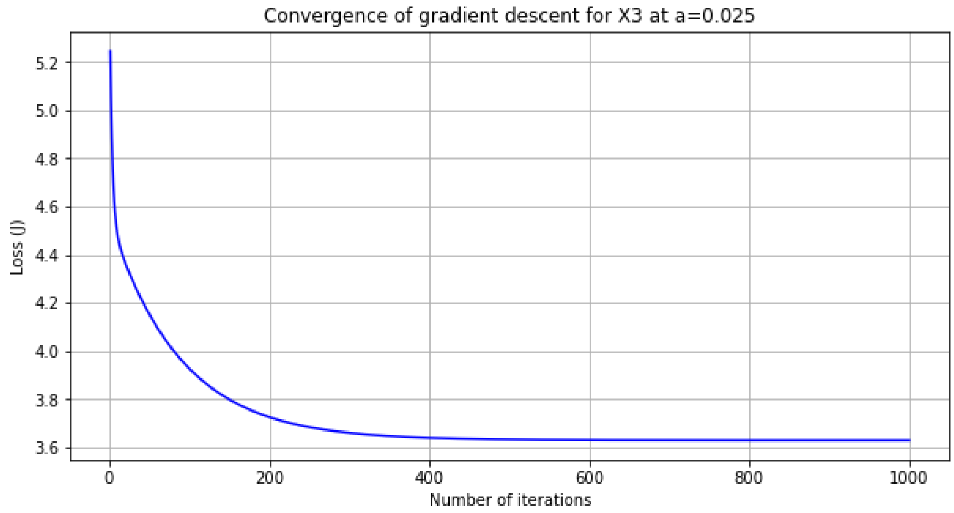
2. **Plot the final regression model and loss over the iterations per each explanatory variable.**

   The final convergent losses for each variable were found to be as shown below, rounded to three decimal places.

$$J_{1\,final} = 0.985$$
$$J_{2\,final} = 3.599$$
$$J_{3\,final} = 3.629$$







3. **Which explanatory variable has the lowest loss for explaining the output?**

   It seems that X1 shows the lowest convergent loss, at 0.985. This is significantly lower than the loss of either X2 or X3, each being above 3.5 for their convergent losses.

4. **Based on your training observations, describe the impact of different learning rates on the final loss and number of training iterations.**

   It seems that changing the learning rate affects how many iterations it takes to reach the convergent loss, but if given enough iterations, even a slower learning rate will come to the same convergent loss as the fast ones. The key part of this statement being that it takes more iterations to converge for a lower learning rate.

Problem 2)

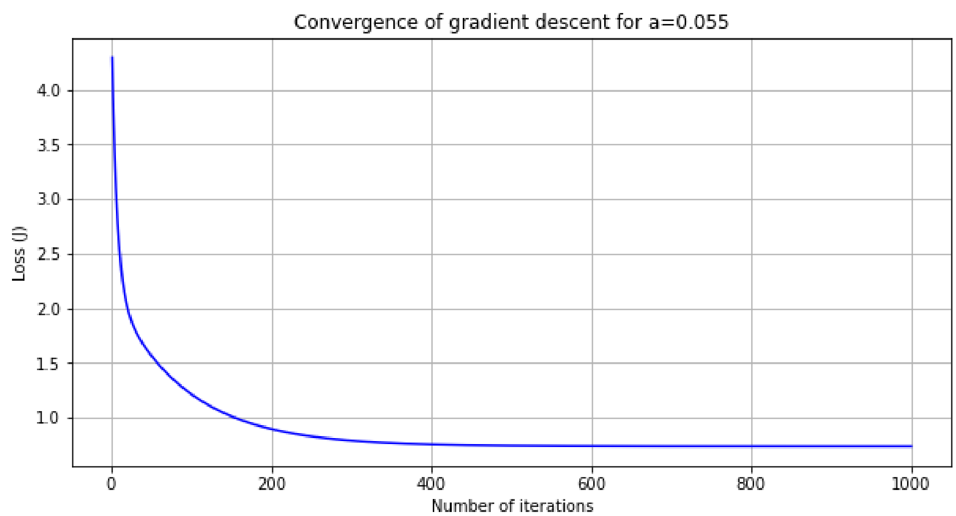
This time, run linear regression with gradient descent algorithm using all three explanatory variables. For the learning rate, explore different values between 0.1 and 0.01 (your choice). Initialize your parameters (theta to zero).

1. **Report the final linear model you found the best.**

   The lowest convergent loss was found to be 0.738, and the corresponding model is shown below. All values are rounded to three decimal places.

$$h(x) = -2.001\ x_1 + 0.536\ x_2 - 0.262\ x_3 + 5.293$$

2. **Plot loss over iterations.**



3. **Based on your training observations, describe the impact of different learning rates on the final loss and number of training iterations.**

   Much like the single-variant versions in part A, it was found that, given sufficient iterations, the learning rate did not have a noticeable effect on the convergent loss. It primarily effects the speed at which the loss converges. At a learning rate of 0.055 or higher, the convergence occurred within 1000 iterations, but when reduced to 0.025, it took more than 2000 iterations to converge to the same loss value.

4. **Predict the output for new inputs (X1, X2, X3) = (1,1,1), (2,0,4), and (3,2,1).**

   Using the generated model from Part B2, the results were calculated below.

$$h(1,1,1) = -2.001(1) + 0.536(1) - 0.262(1) + 5.293 = 3.566$$
$$h(2,0,4) = -2.001(2) + 0.536(0) - 0.262(4) + 5.293 = 0.242$$
$$h(3,2,1) = -2.001(3) + 0.536(2) - 0.262(1) + 5.293 = 0.100$$

GitHub Code Repository: https://github.com/pharmon0/Harmon_ECGR5105