Homework 1

Joshua Foster

801268118

Repo: https://github.com/jfoste81/Homework1

Problem 1:

1. The linear model that I found for each model at a .01 learning rate can be found in the notebook, however for easier viewing here it is again:

```
Model for X1:
Final Model: Y = 5.2929 + -1.7912 * X1
Final Loss: 1.0364

Model for X2:
Final Model: Y = 0.6870 + 0.5767 * X2
Final Loss: 3.5997

Model for X3:
Final Model: Y = 2.5912 + -0.4097 * X3
Final Loss: 3.6397
```

   Briefly, X1 has the best predictive power with its strong negative relationship between it and Y, which we derive from the -1.79 coefficient. We can see from it having the lowest loss among all 3 of the models, at ~1.03, that X1 is the best predictor among the 3. X2 is a significantly weaker predictor of Y, which we can derive from its ~3.6 loss. We can see from the .5767 coefficient that it still has a moderately positive relationship but is still much weaker than X1. Finally, X3 is the weakest predictor of the 3 models with the highest loss, ~3.64. We can see X3 has a weak negative relationship with Y from its -.4097 coefficient, which also happens to have the weakest predictive power among the 3 models.
2. The regression model and loss plots are also in the notebook, but the results of the plots visualize the analysis I presented in question 1.
3. We see from the loss values in the previous questions that X1 provides the lowest loss and is the best predictor of the Y output. Interestingly, X1's model also has the largest absolute coefficient values, which implies a larger influence on the output value.
4. The impact that different learning rates is highly significant as a higher learning rate can help you converge to a low loss value in a lot less iterations but can cause overshooting and make the model miss the absolute minimum loss. With a lower learning rate, the model needs many more iterations to reach a minimal loss rate but is more consistent in finding an absolute minimum. However, with a lower learning rate you run the risk of

getting stuck in a local minimum as you don't have a high enough learning rate to get past this and find the absolute minimum.

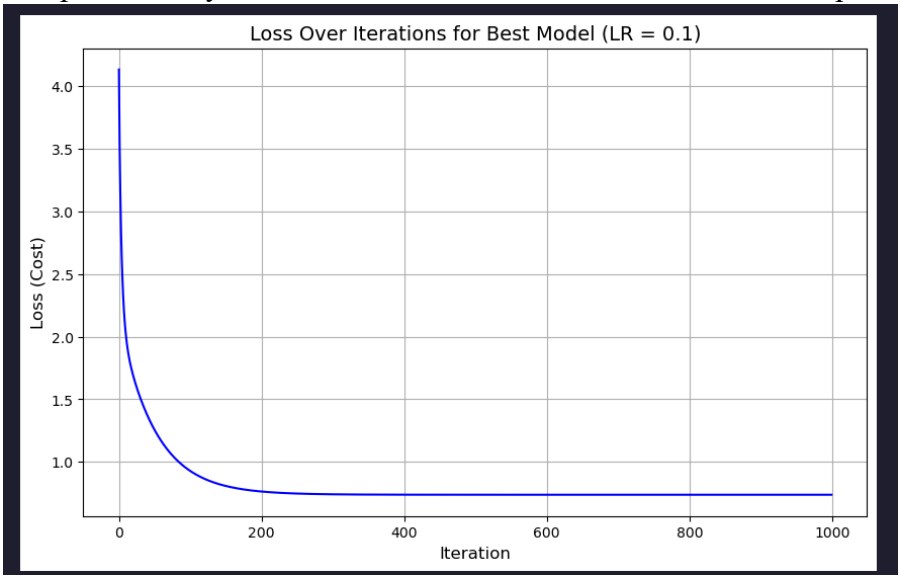
Problem 2:

1. The best linear model I found was:

```
Best Final Linear Model:
Model Equation: Y = 5.3139 + -2.0037*X1 + 0.5326*X2 + -0.2656*X3
```

   And I found the best learning rate to be 0.1.

2. The plot is in my notebook, however for convenience here is the plot again:



3. There is a plot below the plot I mentioned in question 2 that showcases the impact that that each learning rate has on the model, and we can see that the higher learning rates (.1 and .05) outclass the lower learning rate (.01) entirely. In 1000 iterations, not only did the higher learning rates reach a lower loss value than the lower learning rate but they also accomplished it much quicker. The loss values for .1 and .05 ended the same, but the .1 learning rate achieved this minimal loss rate in many less iterations. The .1 learning rate achieved this minimal loss rate in ~200 iterations, while it took around 400-500 iterations to reach the same value. We can see from this that the .1 learning rate performed the best.
4. The predictions for the new values are in my notebook, but they are as follows:

```
Predictions for New Data:
Input 1: [1 1 1] → Predicted Y: 3.5773
Input 2: [2 0 4] → Predicted Y: 0.2443
Input 3: [3 2 1] → Predicted Y: 0.1025
```