ECGR 4105 – HW # 1

Nahush D. Tambe - 801060297

https://github.com/Ntambe25

Problem 1a:

Given below, Figure 1, shows the plot for problem 1, part a. With the pre-defined functions for computing cost and gradient descent, the values of loss(J) and theta, cost history, and validation cost were returned. With the training complete for the training set, the validation and the training losses were plotted on the same plot. For generating this plot, the pre-processing part with normalization and standardization was skipped and as a result the learning rate to achieve convergence was 1e-10 and the # of iterations used were 4,000.

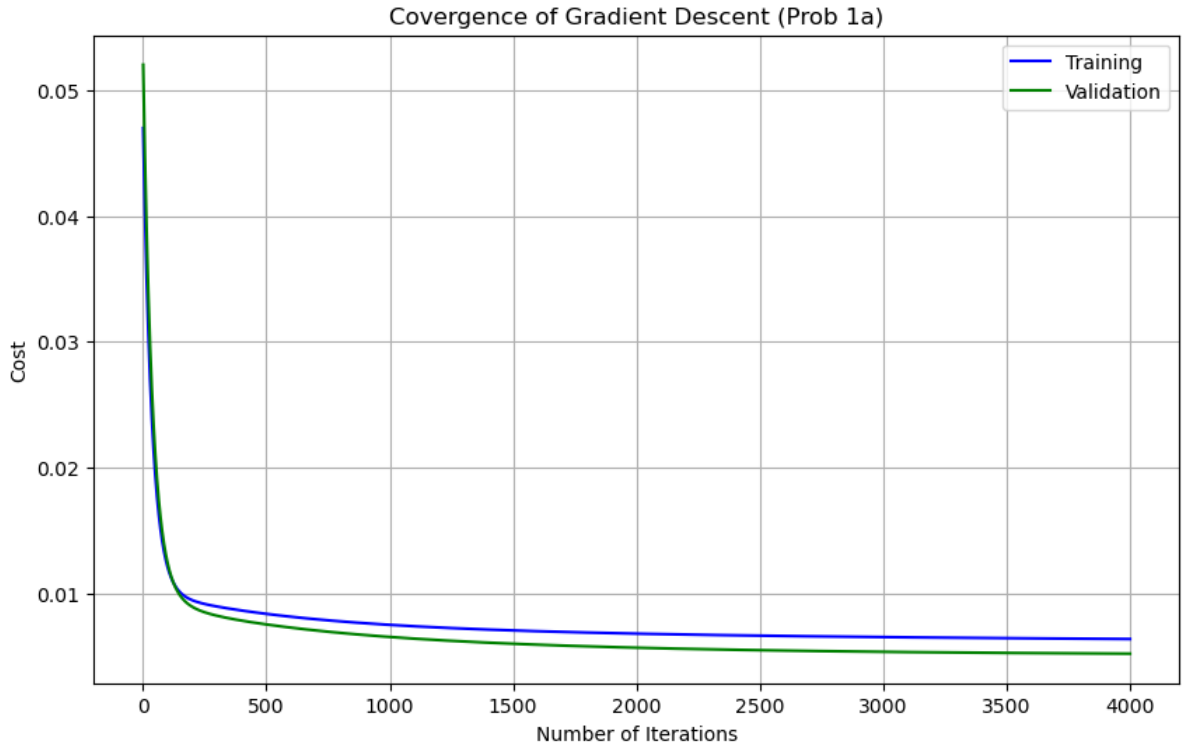

Figure 1: Gradient Descent Plot for Problem 1a

Problem 1b:

Given below, Figure 2, shows the plot for problem 1, part b. Similar to the earlier part, the process for generating this plot was the same, except for the number of input variables was increased. With the absence of normalization and standardization as a part of pre-processing the input data, the learning rate that was used to achieve the convergence was 1e-10 and the # of iterations used were 12,000. The difference between the number of iterations was just to experiment with the values and as it can be seen when comparing the plots for part a and b, that the convergence happens more properly when the number of iterations is 12,000.

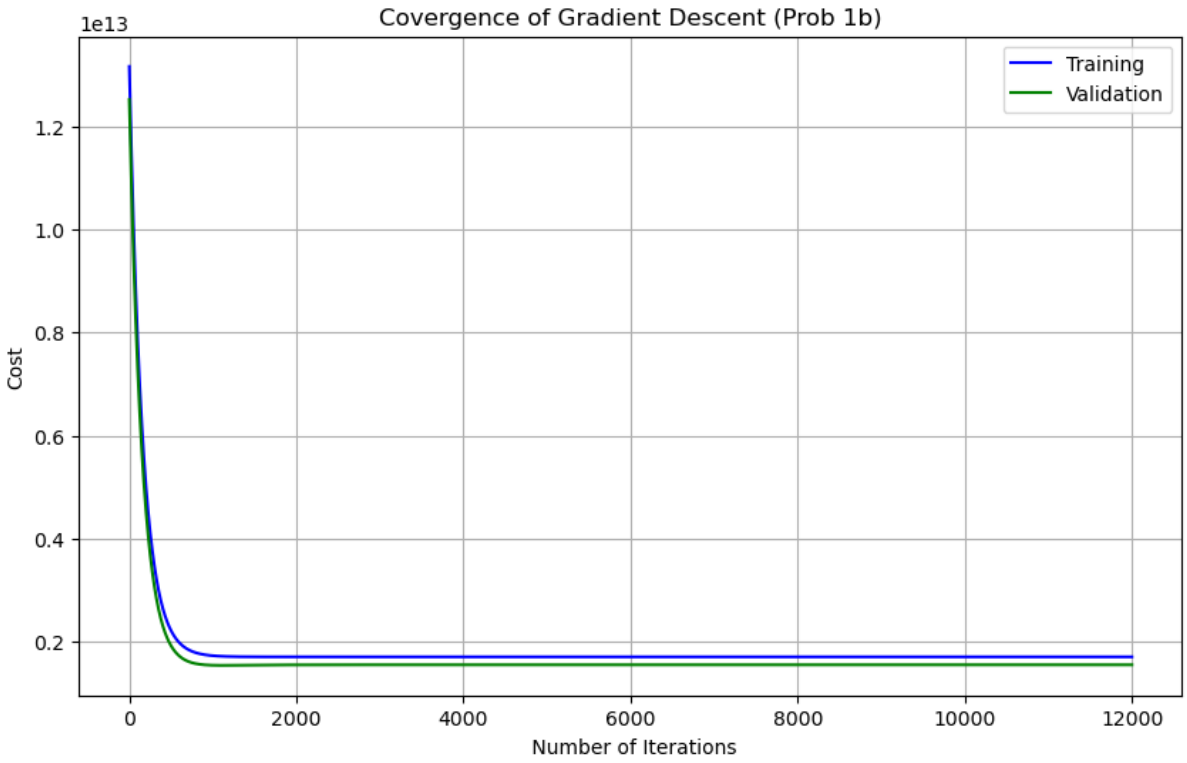

Figure 2: Gradient Descent Plot for Problem 1b

Problem 2a:

Given below, Figure 3, shows the plot for problem 2, part a. Similar to the earlier problem, the process for generating this plot was the same, except for a change, which was, as a part of the input data pre-processing, normalization technique was used. The MinMax Scalar was used to scale down the input data. With the presence of normalization as a part of pre-processing the input data, the learning rate that was used to achieve the convergence was 0.01 and the # of iterations used were 10,000.

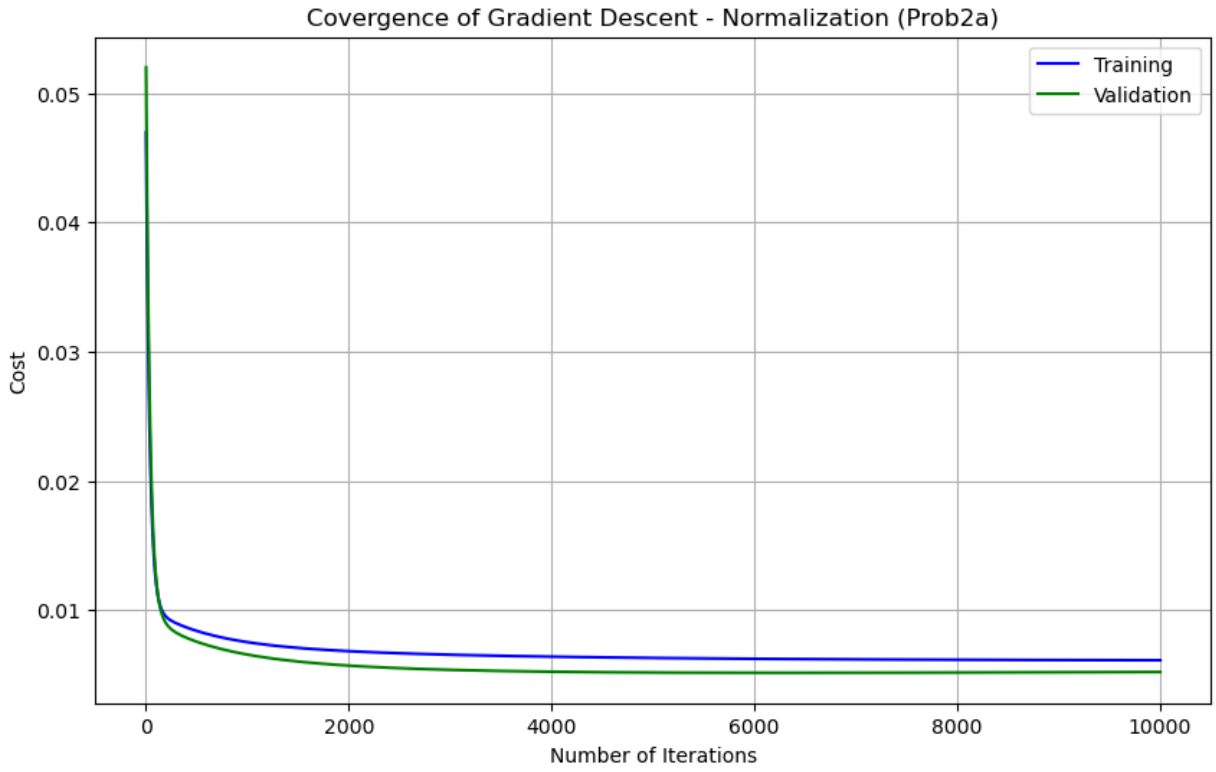

Figure 3: Gradient Descent Plot for Problem 2a with Normalization

Given below, Figure 4, shows the plot for problem 2, part a. Similar to the earlier part, the process for generating this plot was the same, except for a change, which was, as a part of the input data pre-processing, standardization technique was used. The Standard Scalar was used to scale down the input data. With the presence of standardization as a part of pre-processing the input data, the learning rate that was used to achieve the convergence was 0.01 and the # of iterations used were 10,000. With the learning rate and the # of iterations value kept the same, the validation loss was smaller for the normalization than the standardization. With the comparison for the models, both the plots looked the same.

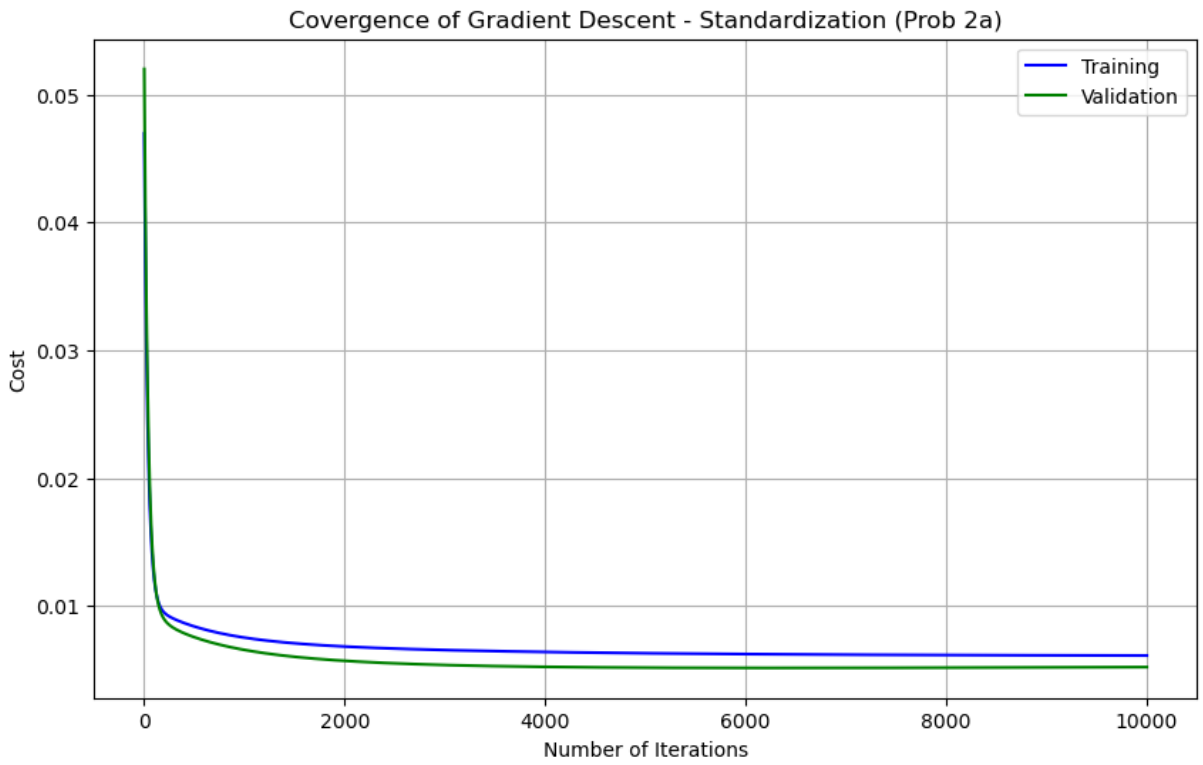

Figure 4: Gradient Descent Plot for Problem 2a with Standardization

Given below are two figures showing the values for loss history and validation loss. It can clearly be seen that the normalization technique yields a smaller value of loss.

```
Final value of theta = [0.04616923 0.3719979  0.10007242 0.20951034 0.13738962 0.10242389]
cost_history_Norm = [0.04697501 0.04596297 0.04497855 ... 0.00612886 0.00612885 0.00612884]
Validation_loss_Norm = [0.05200566 0.05088612 0.04979644 ... 0.00522422 0.00522425 0.00522427]

Final value of theta = [2.53813280e-16 3.88481823e-01 8.90638389e-02 3.04158794e-01
 2.40439813e-01 1.60108157e-01]
cost_history_Stand = [0.49011053 0.48058679 0.47141506 ... 0.22309632 0.22309632 0.22309632]
validation_loss_Stand = [0.48960607 0.47959092 0.46994044 ... 0.19448616 0.19448616 0.19448616]
```

When both the plots for problem 2a are compared with the plot in problem 1a, the convergence was obtained at a learning rate at a higher amount when the input data was pre-processed using the normalization and standardization techniques. This was the result of the input data being scaled at a similar scale across all the input variables.

Problem 2b:

Given below, Figure 5, shows the plot for problem 2, part b. Similar to the earlier problem, the process for generating this plot was the same, except for a change, which was, as a part of the input data pre-processing, normalization technique was used. The MinMax Scalar was used to scale down the input

data. With the presence of normalization as a part of pre-processing the input data, the learning rate that was used to achieve the convergence was 0.01 and the # of iterations used were 12,000.

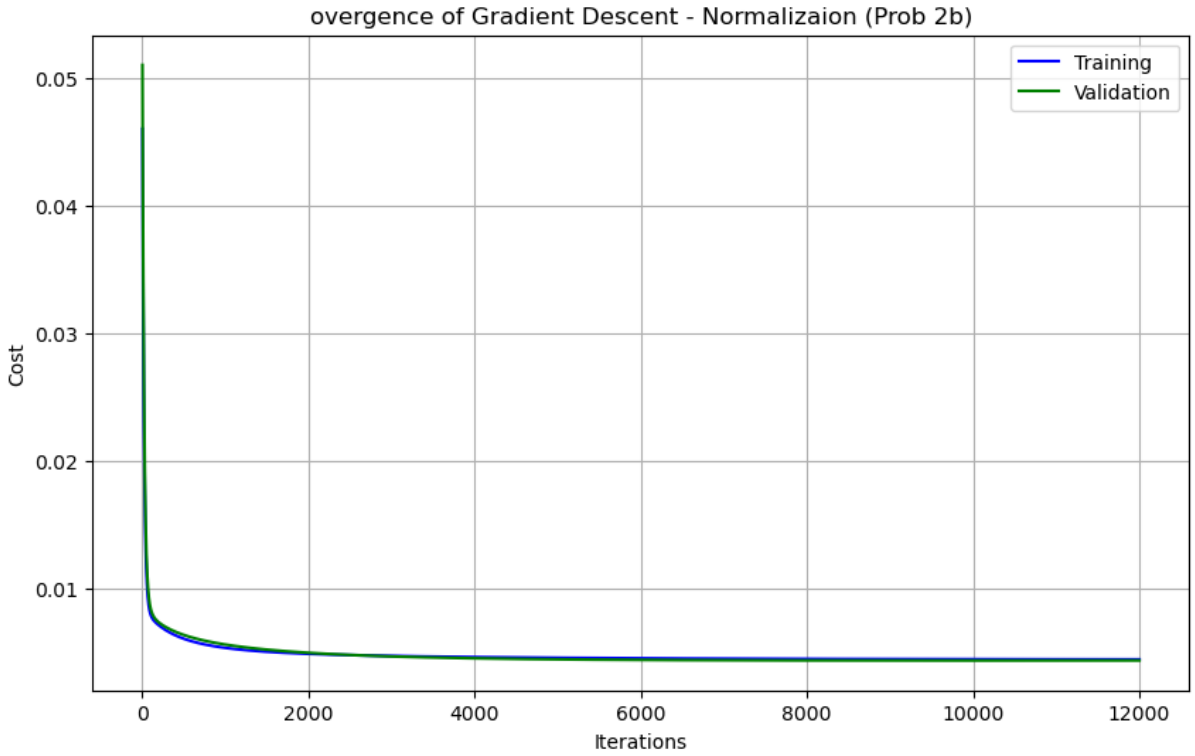

Figure 5: Gradient Descent Plot for Problem 2b with Normalization

Given below, Figure 6, shows the plot for problem 2, part b. Similar to the earlier problem, the process for generating this plot was the same, except for a change, which was, as a part of the input data pre-processing, Standardization technique was used. The Standard Scalar was used to scale down the input data. With the presence of normalization as a part of pre-processing the input data, the learning rate that was used to achieve the convergence was 0.01 and the # of iterations used were 12,000. With the learning rate and the # of iterations value kept the same, the validation loss was smaller for the normalization than the standardization. With the comparison for the models, both the plots looked the similar with a change in the standardization plot which showed a smoother convergence than the normalization plot.

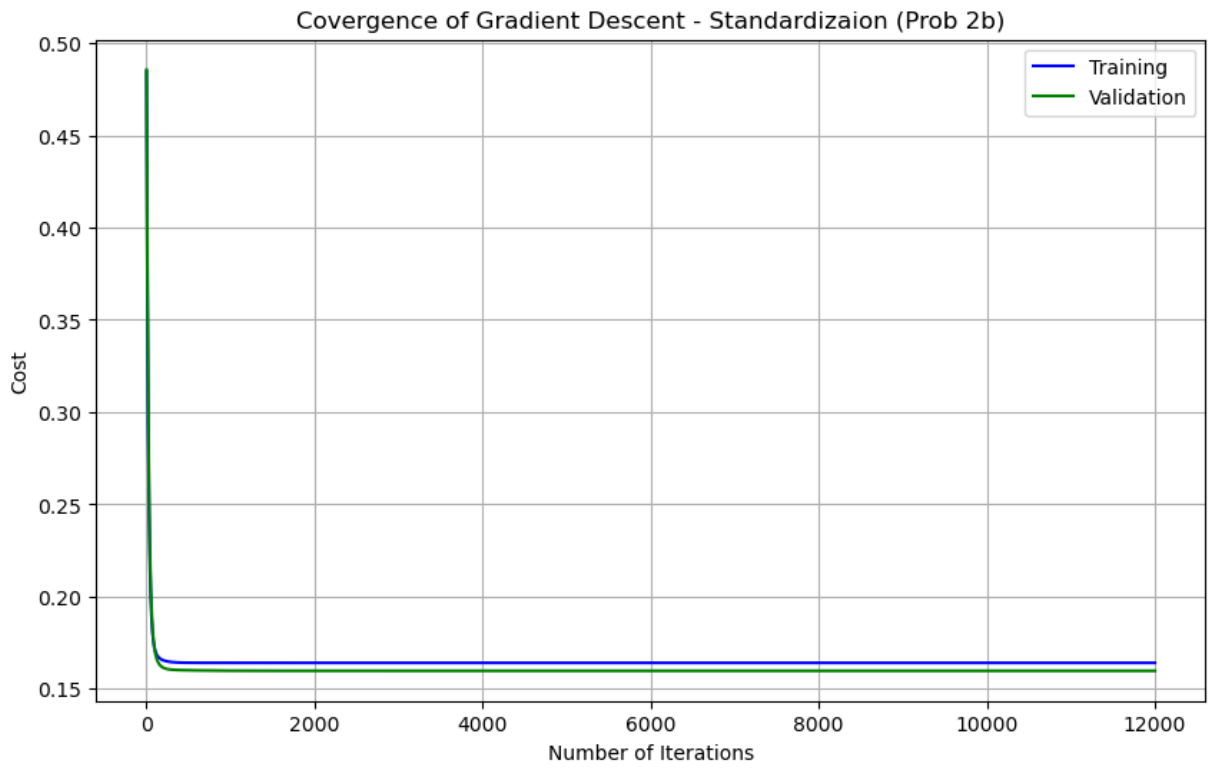

Figure 6: Gradient Descent Plot for Problem 2b with Standardization

Given below are two figures showing the values for loss history and validation loss. It can clearly be seen that the normalization technique yields a smaller value of loss.

```
Final value of theta = [-0.00045882  0.2770333   0.06753607  0.17874094  0.1114527   0.04416715
  0.03976096  0.02884317  0.10441519  0.07924342  0.07003713  0.06477103]
cost_history_Norm = [0.04602978 0.04414316 0.04235062 ... 0.00449891 0.00449891 0.00449891]
validation_loss_Norm = [0.05102886 0.04900518 0.04707974 ... 0.00441321 0.00441321 0.00441322]
```

```
Final value of theta = [2.97977748e-16 2.86802300e-01 5.66977841e-02 2.60175343e-01
 2.00694996e-01 8.59288751e-02 8.77004231e-02 8.75035985e-02
 1.26523800e-01 2.19592259e-01 1.11768608e-01 1.60270222e-01]
cost_history_Stand = [0.48459181 0.4699109  0.45592266 ... 0.16415713 0.16415713 0.16415713]
validation_loss_Stand = [0.48553152 0.47171589 0.45852282 ... 0.15981494 0.15981494 0.15981494]
```

When both the plots for problem 2b are compared with the plot in problem 1b, the convergence was obtained at a learning rate at a higher amount when the input data was pre-processed using the normalization and standardization techniques. This was the result of the input data being scaled at a similar scale across all the input variables.

Problem 3a:

Given below, Figure 7, shows the plot for problem 3, part a . Similar to the earlier problems, the process for generating this plot was the same, except for a change, which was, as a part of the input data pre-processing, normalization technique was used and some of the input parameters were penalized as required by the instructions. The MinMax Scalar was used to scale down the input data and the reason for using this technique over the standardization was because in Problem 2, the normalization technique yielded more better results. With the presence of normalization as a part of pre-processing the input data, the learning rate that was used to achieve the convergence was 0.01 and the # of iterations used were only 1000. This change was significant as in the normalization plot for problem 2a, the number of iteration needed for the plot to smoothly converge were 10,000, which were 10 times as much needed with the plot below that made use of the regularization technique.

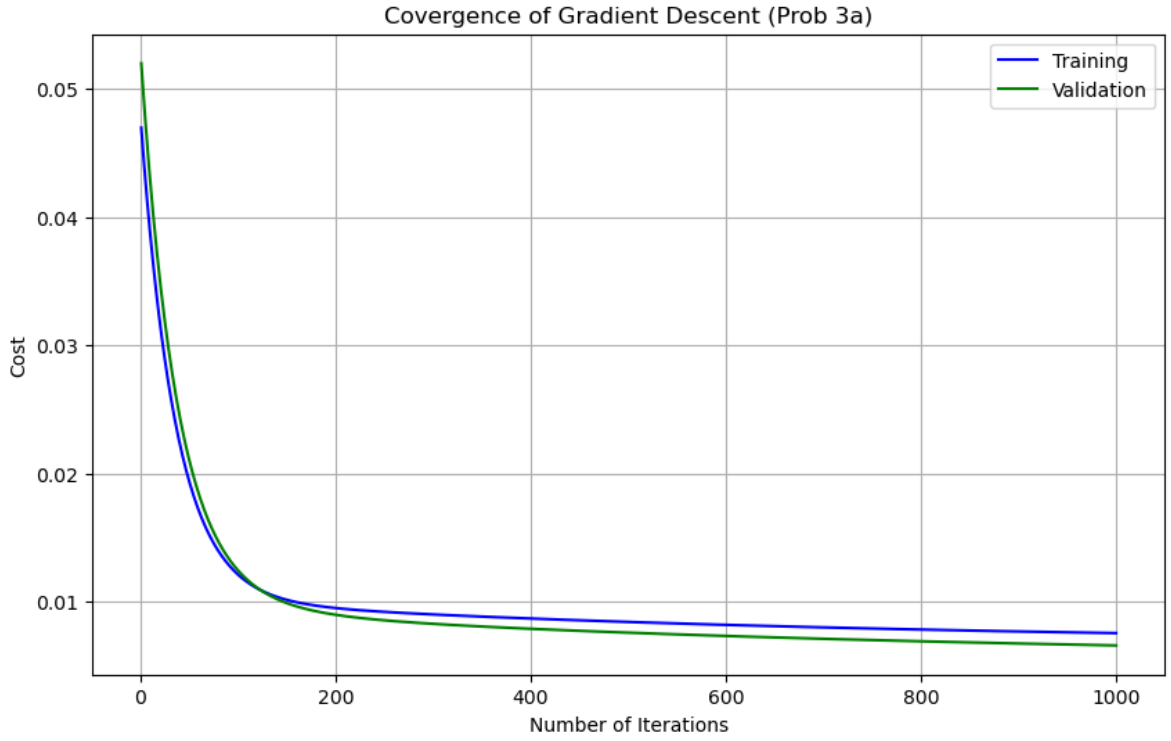

Figure 7: Gradient Descent Plot for Problem 3a

Given below, Figure 8, shows the plot for problem 3, part b. Similar to the earlier problems, the process for generating this plot was the same, except for a change, which was, as a part of the input data pre-processing, normalization technique was used and some of the input parameters were penalized as required by the instructions. The MinMax Scalar was used to scale down the input data and the reason for using this technique over the standardization was because in Problem 2, the normalization technique yielded more better results. With the presence of normalization as a part of pre-processing the input data, the learning rate that was used to achieve the convergence was 0.01 and the # of iterations used were only 1000. This change was significant as in the normalization plot for problem 2b, the number of iteration needed for the plot to smoothly converge were 10,000, which were 10 times as much needed with the plot below that made use of the regularization technique.

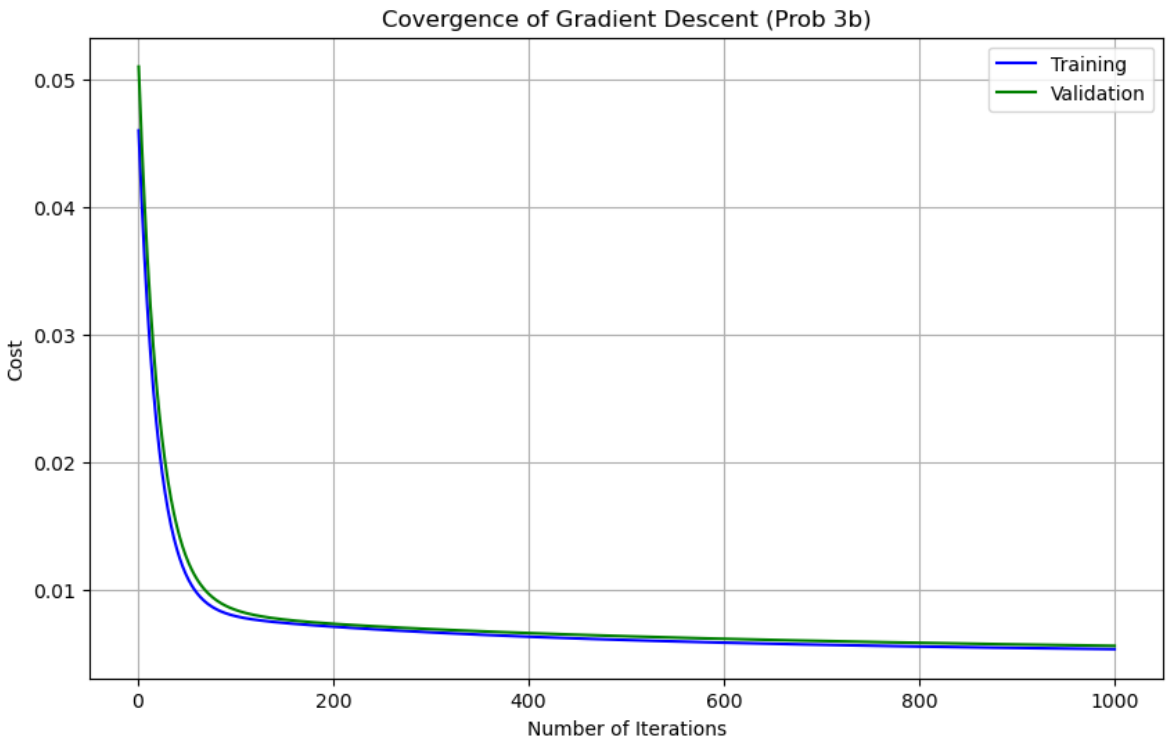

Figure 8: Gradient Descent Plot for Problem 3b
