Stanly Gomes
Student ID: 801118166

Homework 1 Report

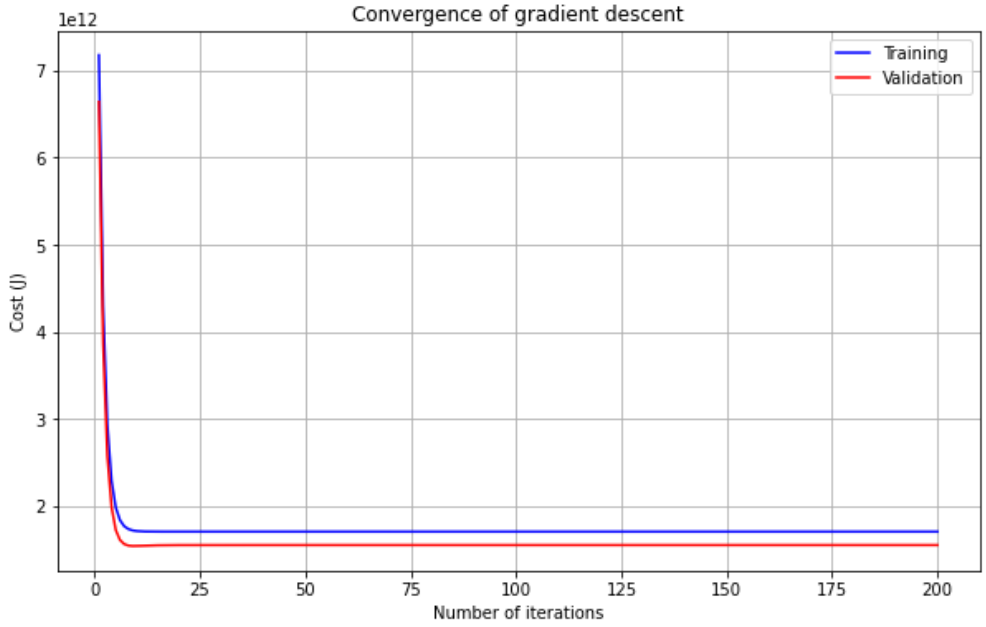
GitHub Repository Link: https://github.com/NaraPvP/IntroToML

Problem 1a:

The best parameters/theta values for each input in this scenario are the following:

a. X0 Parameter: 8.26427521e-01
b. Area: 8.61034777e+02
c. Bedrooms: 3.09868836e+00
d. Bathrooms: 1.68933722e+00
e. Stories: 2.58304257e+00
f. Parking: 7.84212996e-01

I know this based on the training and validation losses shown below. The learning rate is set at 0.00000001 and iterations are set at 200.



Problem 1b:

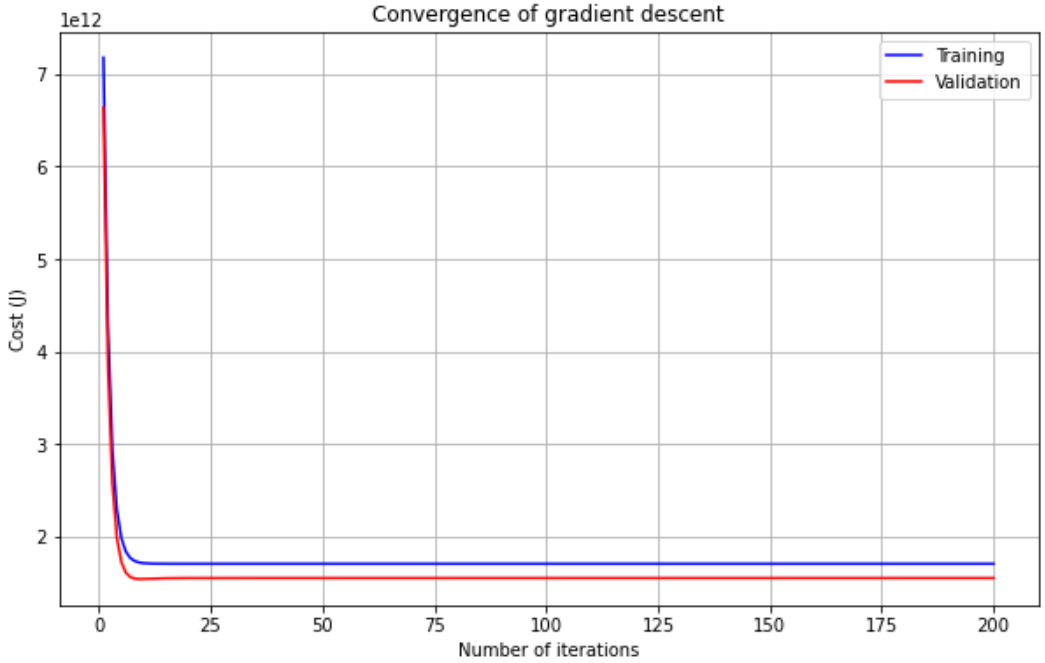
The best parameters/theta values for each input in this scenario are the following:

a. X0 Parameter: 8.26427373e-01
b. Area: 8.61034564e+02
c. Bedrooms: 3.09868789e+00
d. Bathrooms: 1.68933700e+00
e. Stories: 2.58304225e+00
f. Mainroad: 7.49235626e-01
g. Guestroom: 3.48746484e-01
h. Basement: 5.71781420e-01
i. Hotwaterheating: 1.45670881e-01

Stanly Gomes
Student ID: 801118166

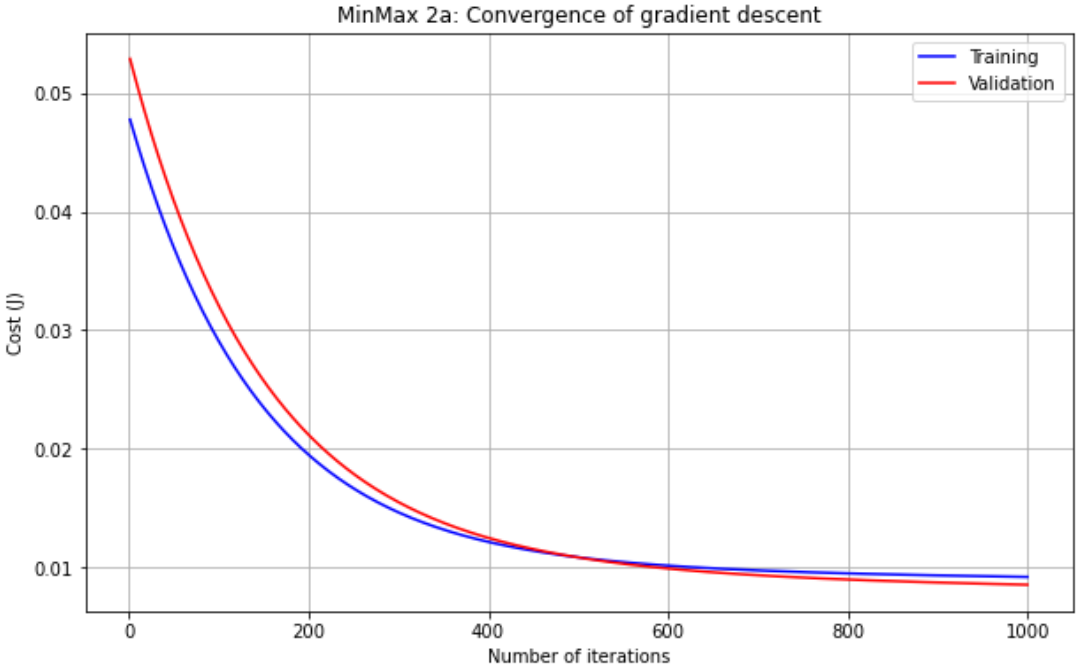
j. Airconditioning: 6.66373716e-01
k. Parking: 7.84212884e-01
l. Prefarea: 3.66320406e-01

I know this based on the training and validation losses shown below. The learning rate is set at 0.00000001 and iterations are set at 200.



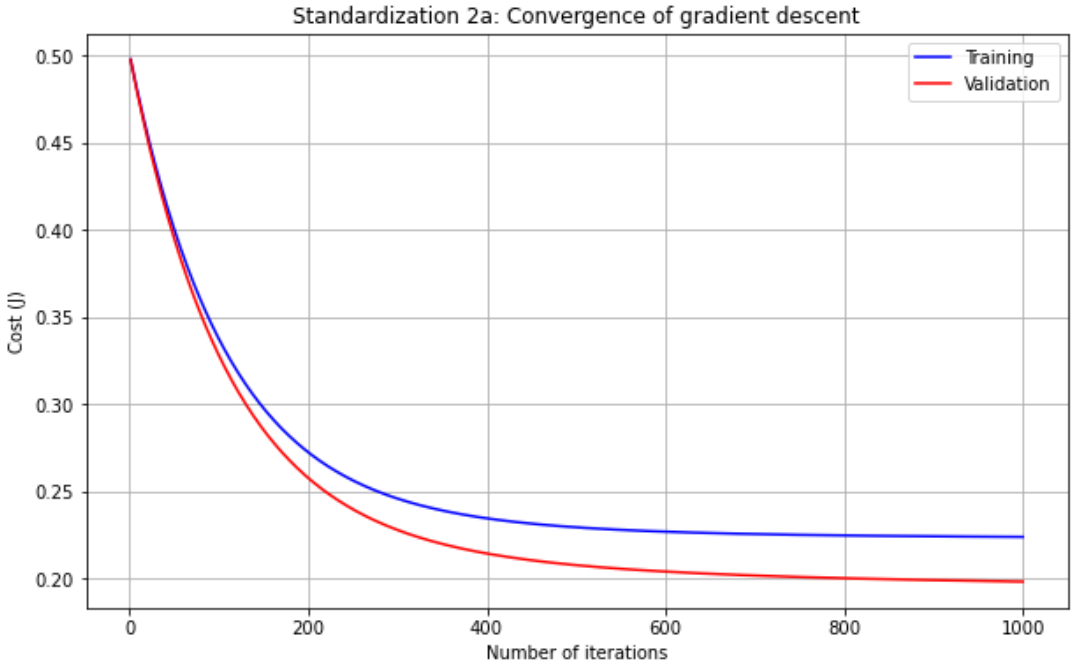
Problem 2a:

Here is the training and validation loss history for the Min/Max Normalization scaling approach:



Stanly Gomes
Student ID: 801118166

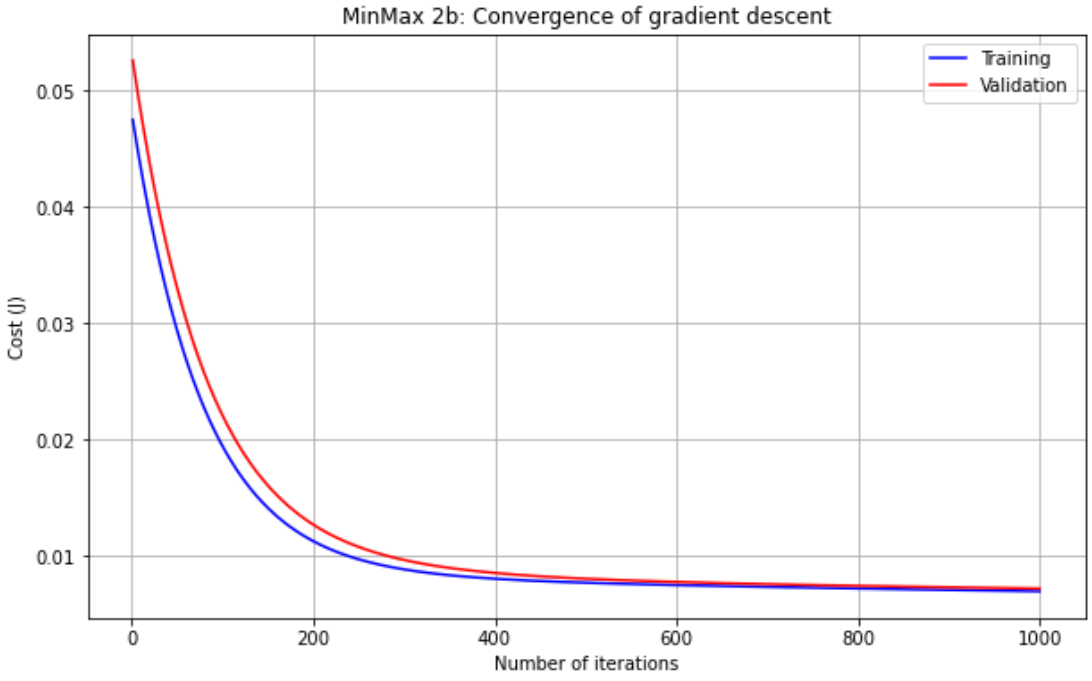
Next, the training and validation loss history for the Standardization scaling approach is shown below:



Based on the loss histories shown above, scaling greatly reduces the loss of our model compared to no feature scaling. Between normalization and standardization, normalization is able to yield smaller loss (greater convergence) than standardization for this dataset.

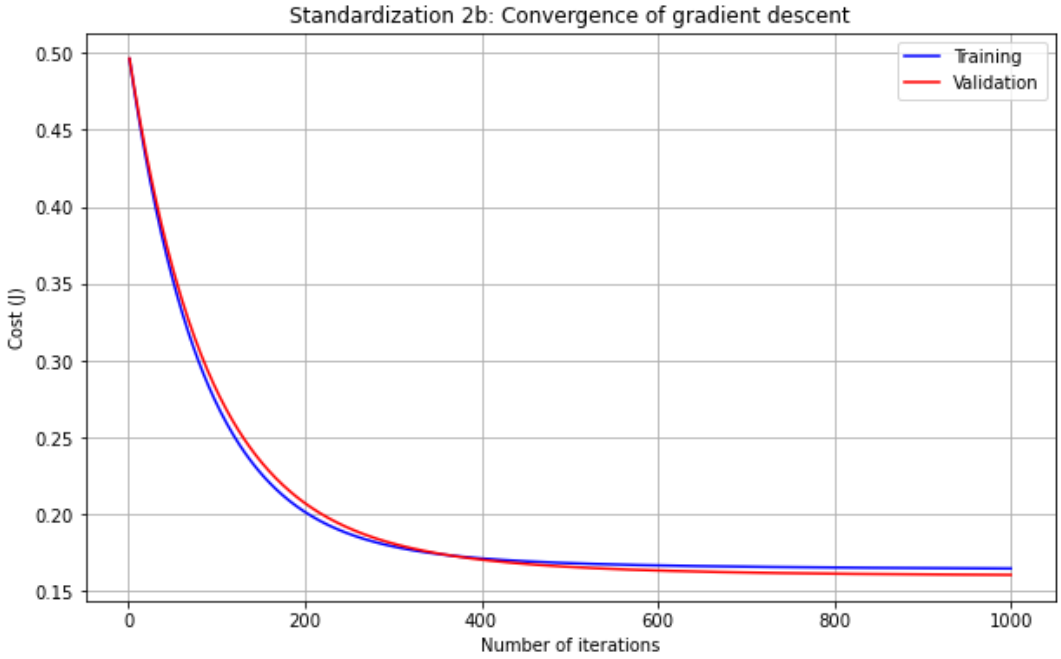
Problem 2b:

Here is the training and validation loss history for the Min/Max Normalization scaling approach:



Stanly Gomes
Student ID: 801118166

Next, the training and validation loss history for the Standardization scaling approach is shown below:



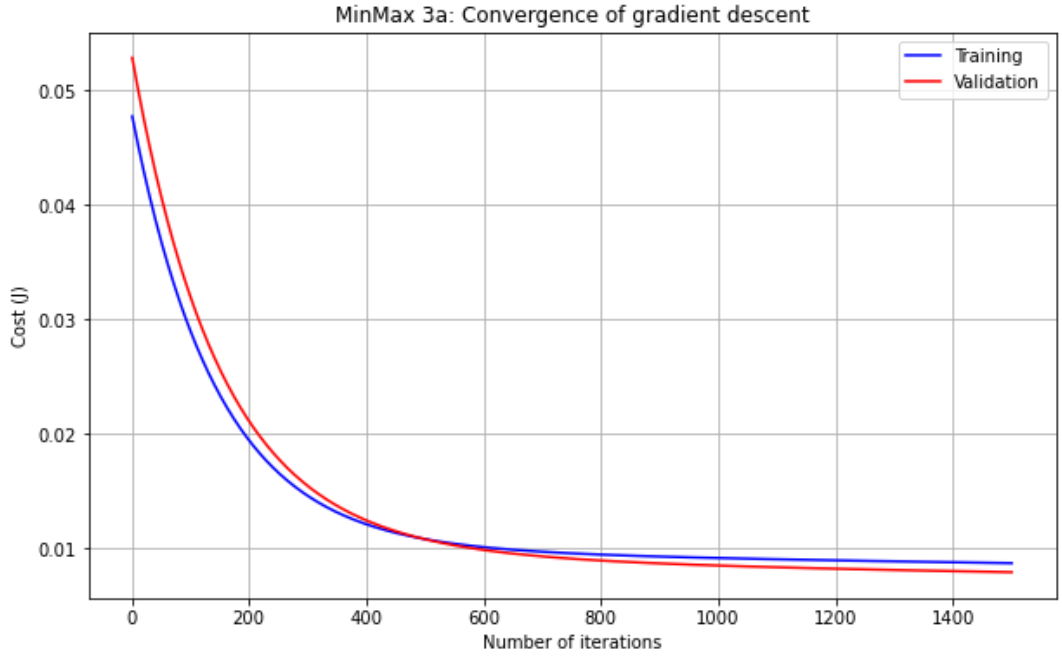
Similarly to Problem 2a, the convergence of gradient descent is much greater compared to no feature scaling being used. Along with this, the increased number of parameters has allowed for less loss in our model. As previously, normalization performs better than standardization for this dataset due to smaller cost values shown in the plots.

Problem 3a

The training and validation loss histories for training done with parameter penalties are shown below. Min/Max normalization scaling was used due to it have better convergence in the gradient descent compared to standardization. The loss has decreased even more than what was done in Problem 2a. This makes sense due to it only improving by lessening the effect of

Stanly Gomes
Student ID: 801118166

abnormaly large/small data values.



Problem 3b

The training and validation loss histories for training done with parameter penalties are shown below. Min/Max normalization scaling was used due to it have better convergence in the gradient descent compared to standardization. The loss has decreased even more than what was done in Problem 2a. This makes sense due to it only improving by lessening the effect of abnormaly large/small data values. Along with this, the loss decreased even more due to the increased number of parameters. With all of the optimizations used in these scenarios, it has resulted in the most accurate model out of the ones explored.