ECGR 5105 Homework 2: Regularization

Lauren Bourque (Student ID: 801264851)

September 26, 2023

GitHub Link

Click here to view the code

Problem 1a.

1. Identify the Best Parameters:

The input variables and their respective theta values are listed in the table below:

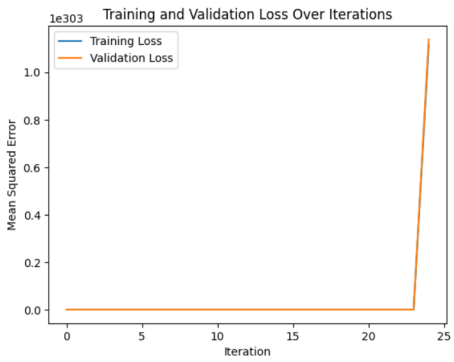
Input Variable	Theta Value
Area	5.98484218e+147
Bedrooms	2.95329852e+144
Bathrooms	1.31790186e+144
Stories	1.82502423e+144
Parking	8.45329210e+143

According to the values, area, bedrooms, and stories are the highest weighted values, making them the best parameters for the linear regression model.

2. Plotting the Losses:

The training and validation losses are plotted below:

Figure 1: Training and Validation Losses for 1a



I chose a learning rate of 0.01 and 25 iterations because the graph grew exponentially around 25. This prevented an overflow and also allowed me to print the theta values to find the best parameters.

Problem 1b.

1. Identify the Best Parameters:

The input variables and their respective theta values are listed in the table below:

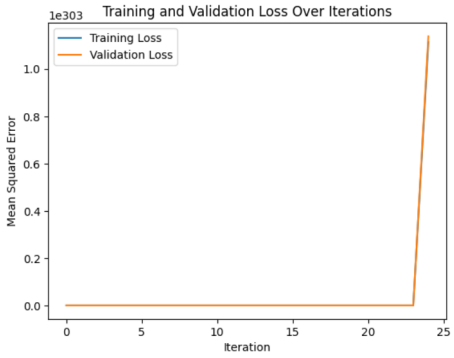
Input Variable	Theta Value
Area	5.98484218e+147
Bedrooms	2.95329852e+144
Bathrooms	1.31790186e+144
Stories	1.82502423e+144
Mainroad	8.97890683e+143
Guestroom	1.94855510e+143
Basement	3.65250060e+143
Hotwaterheating	5.49811100e+142
Airconditioning	3.36640742e+143
Parking	8.45329874e+143
Prefarea	2.67467546e+143

According to the values, area, bedrooms, bathrooms, and stories are the highest weighted values, making them the best parameters for the linear regression model.

2. Plotting the Losses:

The training and validation losses are plotted below:

Figure 2: Training and Validation Losses for 1b



I chose a learning rate of 0.01 and 25 iterations because the graph grew exponentially around 25. This prevented an overflow and also allowed me to print the theta values to find the best parameters.

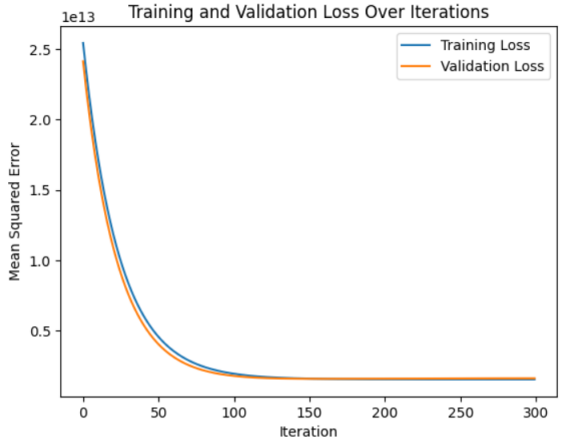
Problem 2a.

1. Normalization:

Figure 3: Normalization Training and Validation Losses for 2a

2. Standardization:

Figure 4: Standardization Training and Validation Losses for 2a

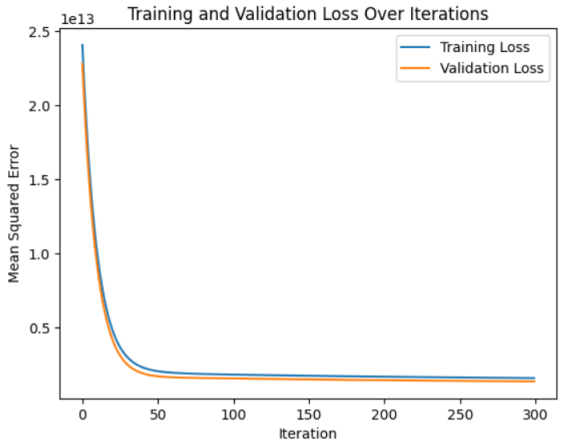


Standardization achieves the best training because the training and validation loss plots converge the most. This means that the difference between the losses is the fewest, making them more accurate.

Problem 2b.

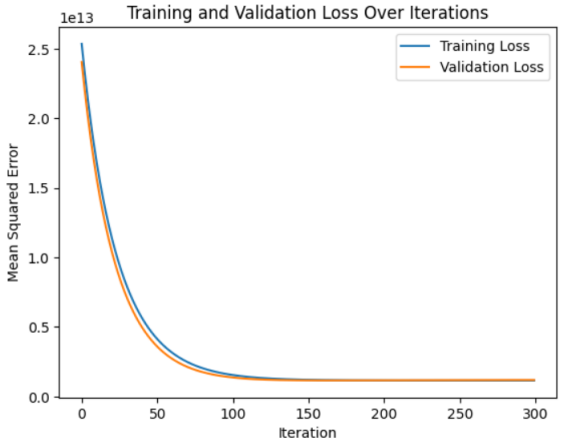
1. Normalization:

Figure 5: Normalization Training and Validation Losses for 2b



2. Standardization:

Figure 6: Standardization Training and Validation Losses for 2b

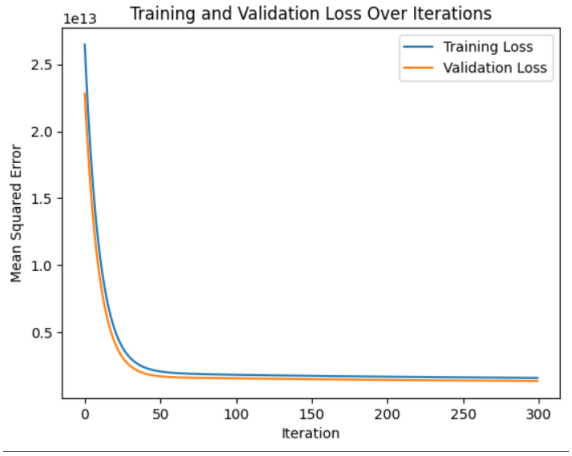


Standardization achieves the best training because the training and validation loss plots converge the most. This means that the difference between the losses is the fewest, making them more accurate.

Problem 3a.

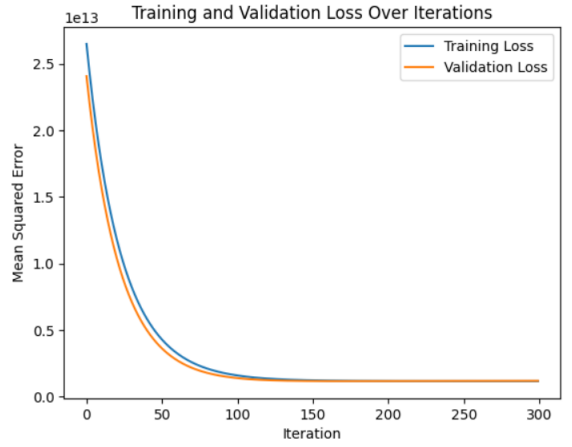
1. Normalization:

Figure 7: Normalization Training and Validation Losses for 3a



2. Standardization:

Figure 8: Standardization Training and Validation Losses for 3a

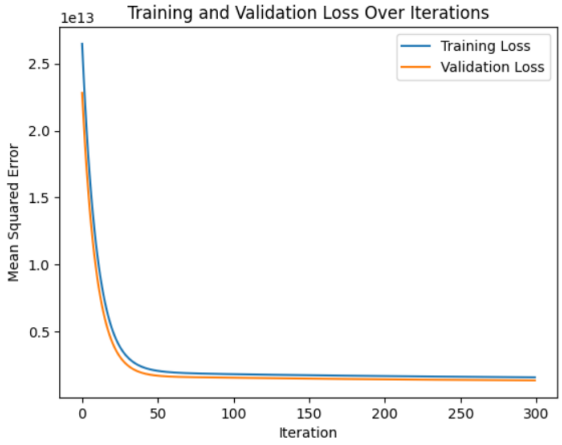


Standardization achieves the best training because the training and validation loss plots converge the most. This means that the difference between the losses is the fewest, making them more accurate. The parameters penalty made the difference slightly smaller, making the predictions more accurate.

Problem 3b.

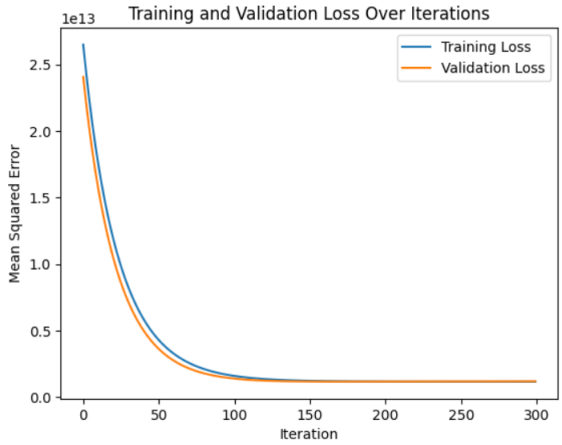
1. Normalization:

Figure 9: Normalization Training and Validation Losses for 3b



2. Standardization:

Figure 10: Standardization Training and Validation Losses for 3b



Standardization achieves the best training because the training and validation loss plots converge the most. This means that the difference between the losses is the fewest, making them more accurate. The parameters penalty made the difference slightly smaller, making the predictions more accurate.