Intro ML Homework 2

Jaskin Kabir
Student Id: 801186717
GitHub:
https://github.com/jaskinkabir/Intro_ML/tree/main/HM2

September 2024

1 Training With Raw Data

This assignment utilized the housing price dataset. The features that were listed as yes/no values were first mapped to 1 and 0 respectively. Then, the data was split into training and validation sets with an 80/20 split.

a *Five-Dimensional Model*
Using the following 5 variables as inputs for training: [area, bedrooms, bathrooms, stories, parking], the linear regression produced the following validation and training loss curves:

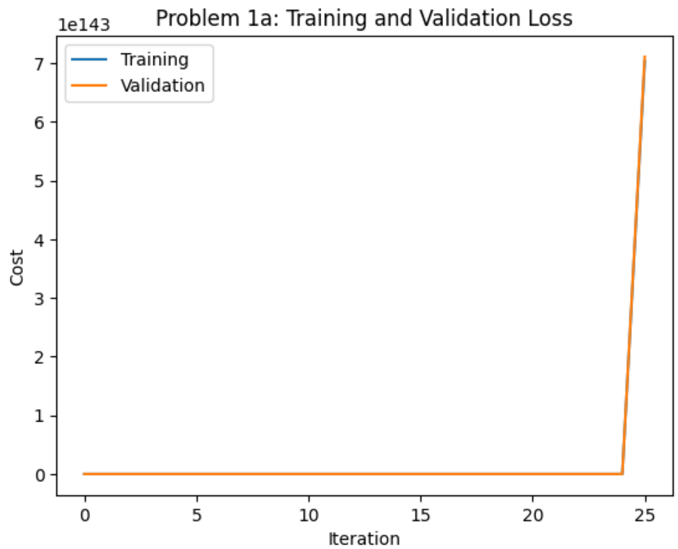

Figure 1: Five-Dimensional Regression

Without any feature scaling or regularization, the model does not converge. With the different input variables occupying such wildly different ranges, the regression cannot adjust the model appropriately. The validation and training losses are almost identical, which is why only the validation loss curve can be seen in Figure 1. Because the model's loss diverges to infinity, the learning rate was chosen to be 0.01 to prevent overflow. With this value of α, a maximum of 25 training iterations could be performed before overflow occured.

b *11-Dimensional Model*
Using the following eleven variables as inputs for training: ['area', 'bedrooms', 'bathrooms', 'stories', 'mainroad', 'guestroom', 'basement', 'hotwaterheating', 'airconditioning', 'parking', 'prefarea'], the linear regression produced the following validation and training loss curves:

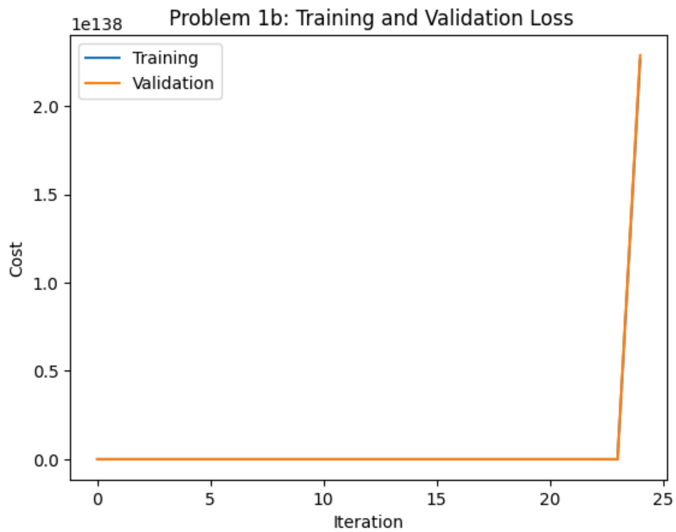

Figure 2: 11-Dimensional Regression

This regression resulted in an almost identical final loss value as the 5-dimensional regression. Clearly, scaling and/or regularization techniques must be employed in order for this model to converge.

2 Feature Scaling

Before linear regressions were performed, the parameters for the normalization and standardization techniques were fit to the full housing dataset. Then, it was split into the 5-variable and 11-variable training and validation sets.
All graphs from Figure 3 and onwards show the first 50 training iterations of a 1000 iteration training process in order to better visualize the models' progress. Since these models converge, the learning rate α was chosen to be 0.1 to generate better models.

a *5-Dimensional Regression With Feature Scaling*

By applying a min-max normalization to the inputs, the five-dimensional regression was able to converge to a final loss value of 8.803×10^5, which is a significant improvement over the practically infinite loss achieved by the model trained on this same data without normalization. Applying standardization to this data also resulted in a regression that converged, but to a 1.478% larger final loss value of 8.935×10^5.

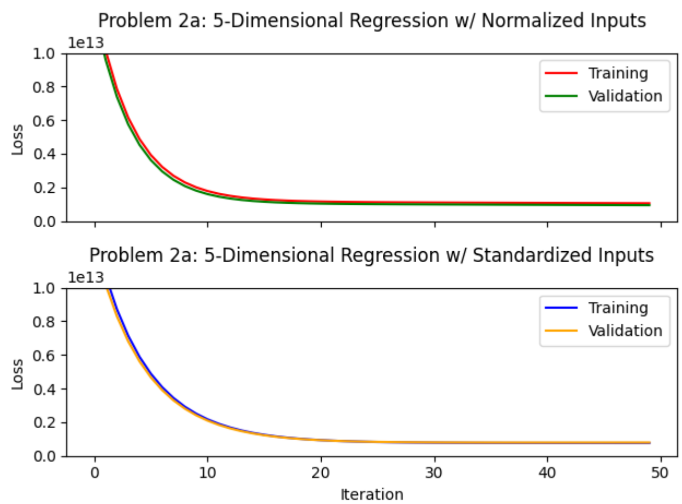

Figure 3: 5-Dimensional Regression With Feature Scaling

As shown in Figure 3, the model trained on normalized data had a training loss that was slightly higher than the validation loss. The opposite was true for the standardized data, whose model showed very slight overfiting. For this data, input normalization achieved the better training.

b *11-Dimensional Regression With Feature Scaling*

Applying the same min-max normalization and standardization to the 11-Dimensional input dataset resulted in much lower loss values than the 5-Dimensional dataset. The normalized regression reached a final loss value of 7.523×10^5, which marks a 17.01% improvement over the normalized 5-Dimensional regression. The standardized regression reached a loss of 7.6023×10^5, which is an improvement of 17.52%.

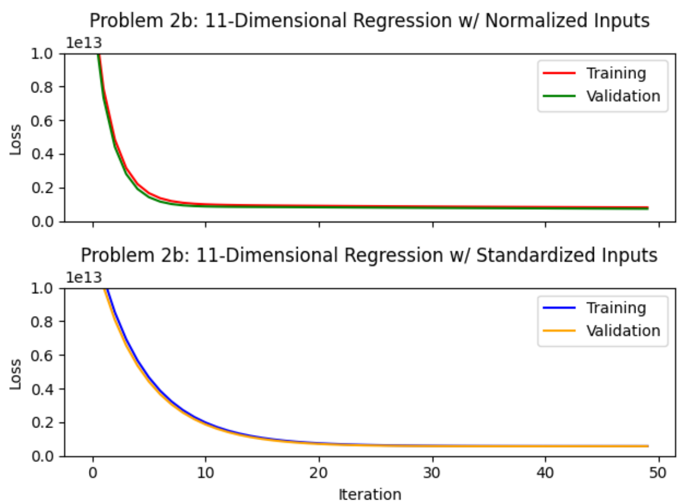

Figure 4: 11-Dimensional Regression With Feature Scaling

Much like in Figure 3, Figure 4 shows a normalized regression whose validation loss is slightly lower than its training loss, and a standardized regression whose training loss is slightly lower than its validation loss.

While standardization showed a greater improvement when increasing the dimensionality of the model, normalization resulted in the lower loss value of the two techniques. Therefore, with this dataset, normalization was again the better feature scaling method.

3 Regularization

By adding a penalty parameter λ to the gradient descent algorithm, a further improvement in loss was found with respect to that of the unregularized, feature scaled data shown in Figures 3 and 4. However, this improvement is impossible to achieve with any value of λ when the number of training iterations is reduced from 1000 to 50. This is likely due to the role of regularization in alleviating overfit. Since the increased iteration count can lead to overfitted models, adding a penalty parameter should have an increased improvement in model performance when training occurs over more iterations.

a *5-Dimensional Regression With Regularization*
Choosing a parameter value of $\lambda = 0.008$ resulted in an improvement in loss of 0.834% for the normalized data when compared to the unregularized regression model shown in Figure 3. For the standardized data, choosing $\lambda = 0.015$ was able to achieve an improvement in cost of only 0.155%, showing the normalized data to be the better dataset to use training regularization on in this case.
These λ values were found to result in the greatest improvement over the unregularized training.

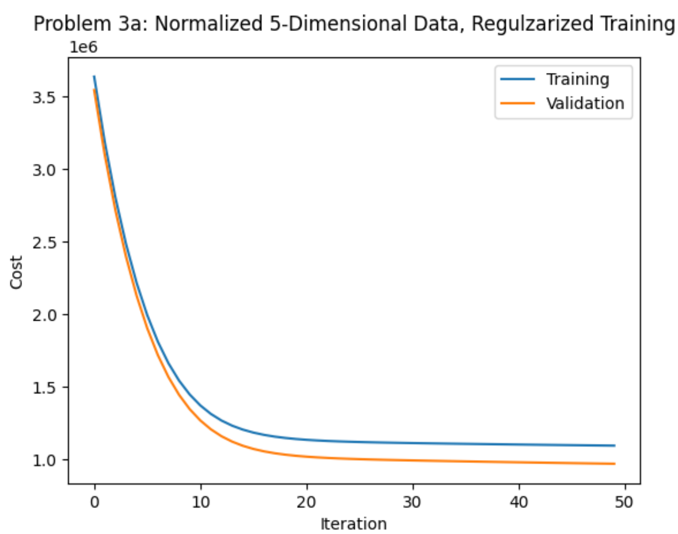

Figure 5: 5-Dimensional Regression With Normalized Data and Regularized Training

b *11-Dimensional Regression With Regularization*
For the normalized, 11-dimensional data, the best value of λ was found to be 0.022m which resulted in an improvement in final loss of 0.323%. The best value of λ for the standardized data of 0.10 resulted in a loss reduction of 0.506%. Interestingly, while the normalized data was better able to take advantage of the regularization for the 5-dimensional data, the standardized data saw the greater improvement for the 11-dimensional data.

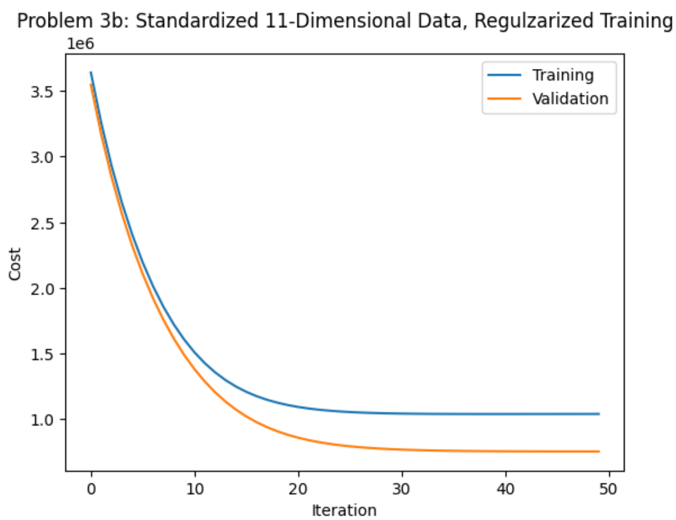

Figure 6: 5-Dimensional Regression With Normalized Data and Regularized Training