

Conclusion

Gradient Descent Methods

SGD: Initially, the loss is very large, but it decreases rapidly following an exponential trend. SGD achieves the smallest final loss value.

MBGD (batch size = 10): The initial loss for MBGD is smaller compared to SGD. The loss decreases following a linear trend. MBGD falls somewhere between SGD and BGD.

BGD: The initial loss for BGD is the smallest. The change in loss becomes very small, as epochs increasing. BGD ends with the largest final loss.

Normalization

Without normalization: The convergence was much slower. This is because the data had a larger range of values, causing the gradients to vary significantly, which slowed down the training process.

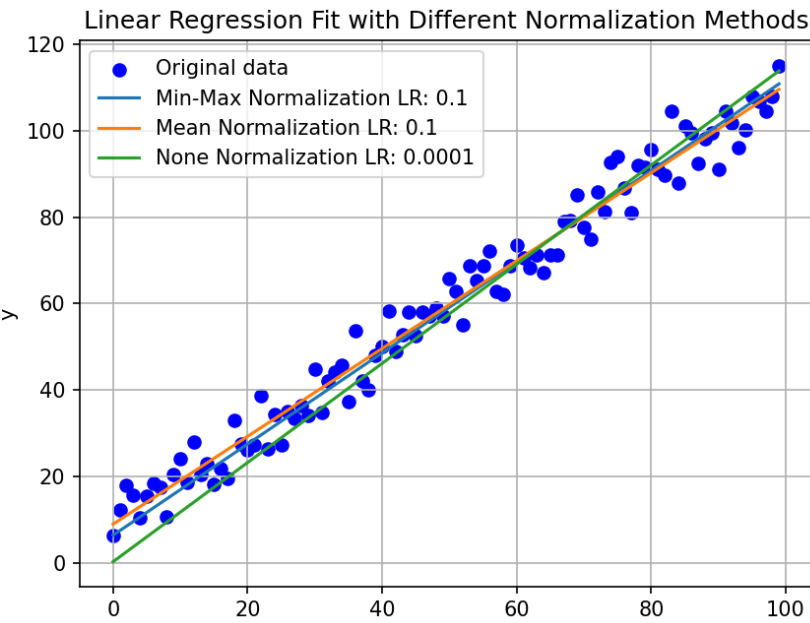
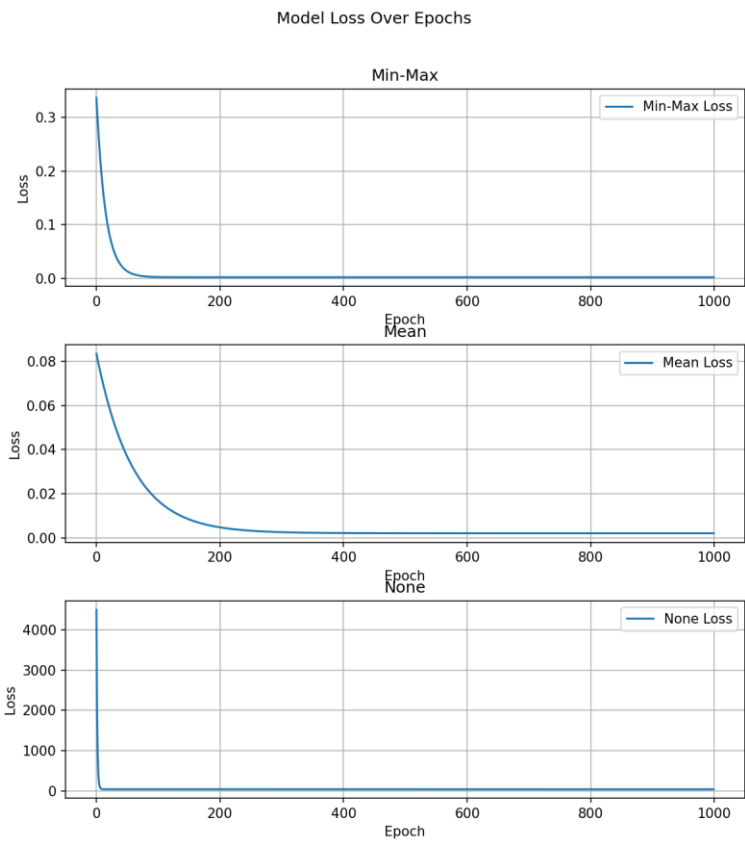
With normalization (both min-max and mean normalization): The convergence speed improved significantly compared to the case without normalization. Because the data was scaled to a smaller range.

Learning rate adjustment: After applying normalization, I had to increase the learning rate by a much larger factor. Since the values of the features were now smaller, a larger learning rate was needed to make meaningful updates to the model parameters.

Better fit: After normalization, the model achieved better fitting performance,

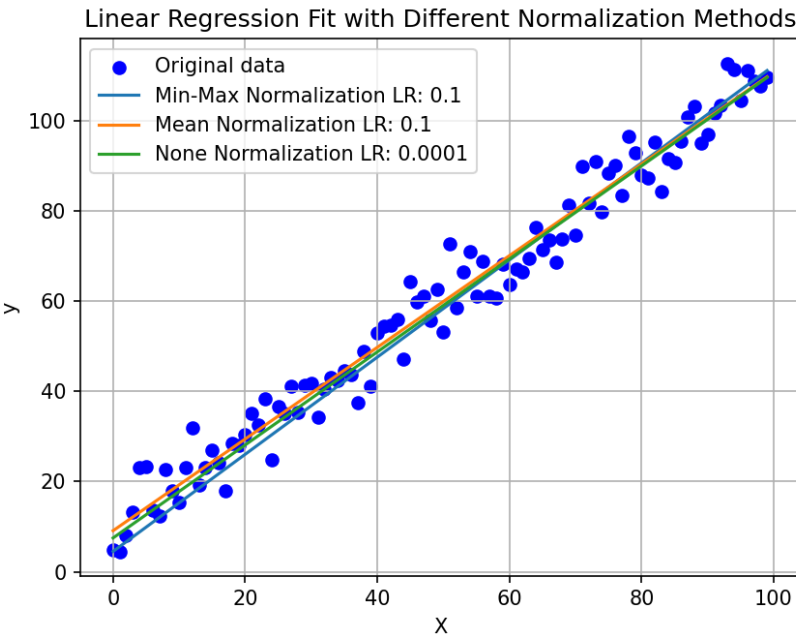
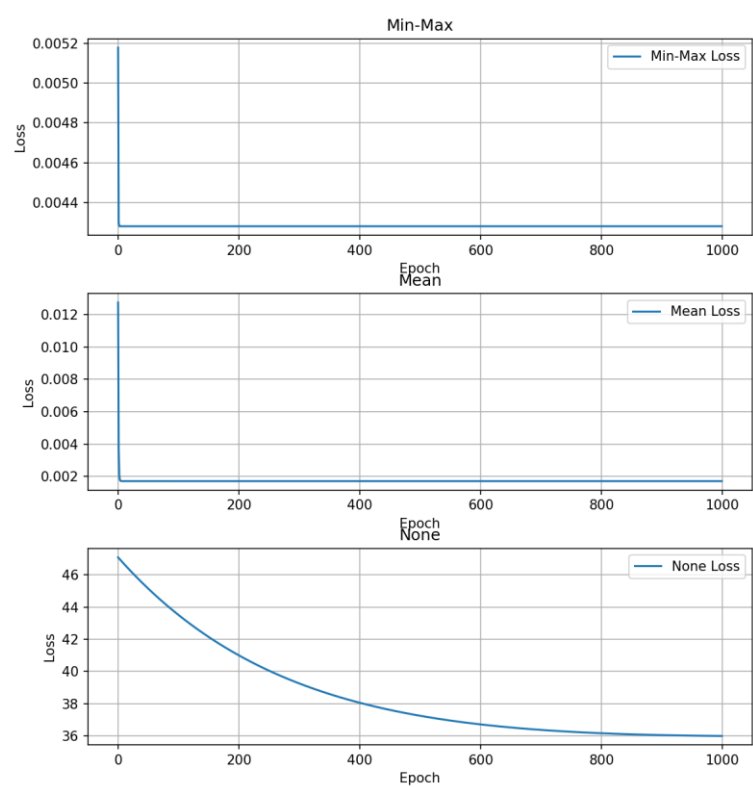
Experimental Results

BGD



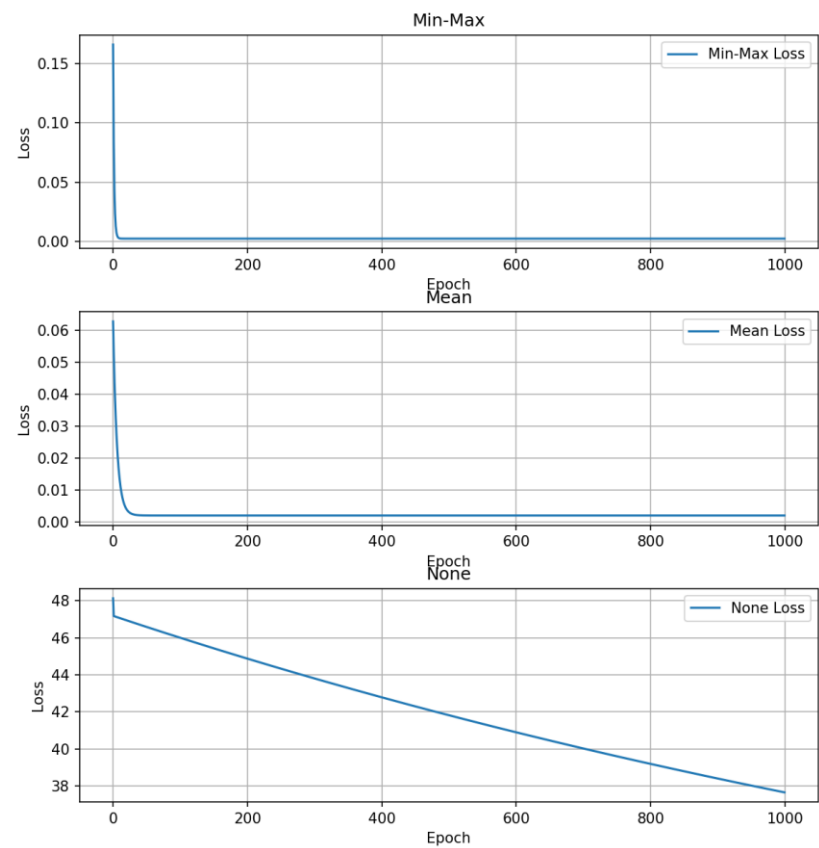
SGD

Model Loss Over Epochs



MBGD

Model Loss Over Epochs



Linear Regression Fit with Different Normalization Methods

