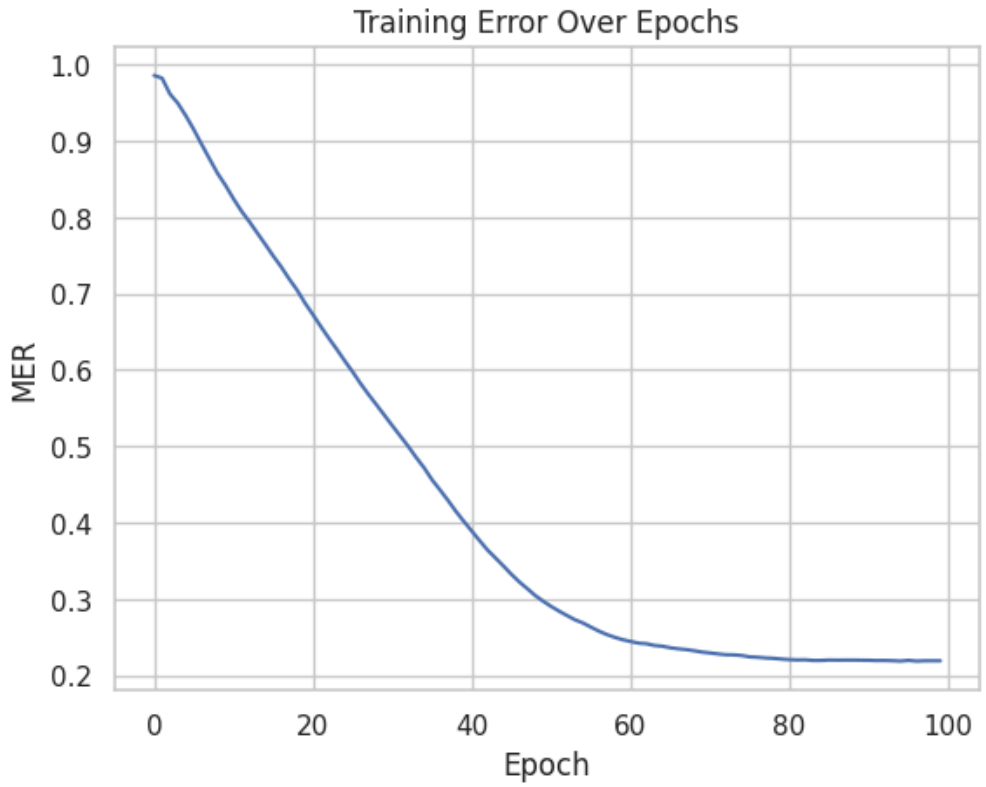
Section C, Team 56

Boqing (Niko) Zheng, Genna Barge, Kainat Nazir, Yiqiao (Kevin) Wang, Rajat Bajaj

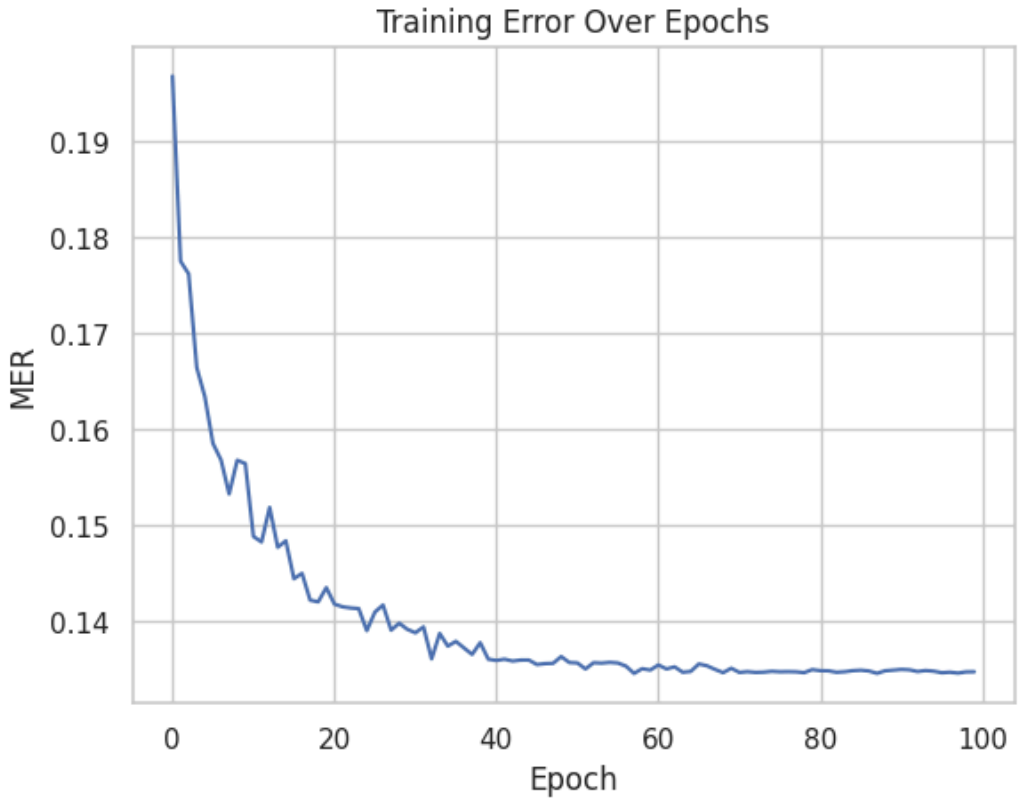
Analysis of Training Errors in Linear Regression and Multi-Layer Perceptron Models for House Price Prediction

This report explores the performance of two models for house price prediction: a linear regression model and a multi-layer perceptron (MLP) model. It includes training error plots for each model along with explanations and business implications of these error trends.



- **Training Error Analysis of the Linear Regression Model**

Plot Description: In the training error plot for the linear regression model, we observe that the median error rate (MER) decreases rapidly in the early epochs and stabilizes at around 0.22 by epoch 50. This trend suggests that while the linear regression model can quickly approximate basic relationships in the data, it lacks the complexity needed to capture more subtle nonlinear patterns that may exist in home price data. The high level of final error indicates that the model may have oversimplified the relationship between features and home prices, resulting in inaccurate predictions. For Zillow management, where accurate pricing directly impacts customer trust and revenue, relying solely on linear models may not provide the accuracy needed to make optimal pricing decisions.



- **Training Error Analysis of the Multi-Layer Perceptron (MLP) Model**

Plot Interpretation: As can be seen from the figure, the error decreases rapidly within the first 20 epochs, from about 0.19 to less than 0.14, indicating that the model quickly learns the important features in the data at the initial stage, and has a certain degree of accuracy in predicting house prices. Subsequently, the decline of the error gradually slows down and levels off between 20 and 40 epochs, showing that the model is gradually approaching the optimal solution. After 40 epochs, the error basically converges and stays at the level of about 0.14, and further training fails to significantly reduce the error, which indicates that the model may have reached the optimal state under the current architecture and parameters. The decreasing oscillations in the figure also reflect the effectiveness of the dynamic learning rate adjustment method StepLR. The learning rate decreases by 5% every 10 epochs, which allows for more accurate convergence to the optimal parameter combination. For Zillow management, the MLP has a smaller error compared to a linear model, which means the model can be used as a home price forecasting tool, but there is room for improvement in accuracy.

- **Q1: What are your final training errors of the multilayer perceptron model and the linear regression model?**

After training the linear regression model and the MLP model for 100 epochs, their final MERs are 0.21941586 vs. 0.13469726, respectively.

- **Q2: What is the test error shown on Kaggle? How does it compare with the train error?**

After uploading the prediction results of the MLP model in the test set to Kaggle, the MER of the model is 0.14033. This is slightly higher than our MER of 0.13469726 in the training set, which may mean that the model has some deficiencies in its generalization ability in the test set, or it has an overfitting problem in the training set, and it may also be due to the fluctuation of the data from the training set and the test set.

- **Insights for Senior Management at Zillow**

For Zillow's senior management, the analysis shows that the multi-layer perceptron (MLP) model offers a notable advantage over the linear regression model in terms of prediction accuracy. Implementing the MLP model can lead to several strategic benefits for Zillow:

1. **Enhanced Pricing Precision**: The MLP model's ability to capture complex relationships results in more accurate house price predictions, which can help Zillow set competitive and realistic pricing for properties listed on the platform. This precision can lead to higher customer satisfaction by providing pricing that aligns more closely with market realities.
2. **Operational Efficiency**: By reducing errors in predictions, the MLP model minimizes the need for frequent price adjustments, potentially reducing manual interventions and associated costs. This improvement in operational efficiency aligns with Zillow’s goal of providing accurate and timely information to its users.
3. **Competitive Advantage**: Leveraging a model that offers better accuracy in predictions positions Zillow as a trusted source in the real estate market. Accurate and reliable pricing can enhance the platform’s reputation, attract more users, and increase the volume of transactions, strengthening Zillow's competitive edge in the industry.