MENTOR: DASHA PETROV
MENTEE: JINGYU ZHANG

THE INTRODUCTION OF DEEP LEARNING

TEXTBOOK

Drive into Deep Learning

INTRODUCTION

This project summarizes my learning progress and practical exploration in deep learning. In the first half of the quarter, I studied the fundamental concepts of neural networks through the first five chapters of Dive into Deep Learning. In the second half, I applied these ideas to a real-world regression task, predicting house prices using neural network models. This experience helped me bridge theory and practice, and deepened my understanding of how models behave with real data.

OBJECTIVE

- To gain a foundational understanding of deep learning models, especially Multilayer Perceptrons (MLPs)
- To learn the training process including forward and backward propagation, loss functions, and weight updates
- To apply these concepts to a predictive modeling task and evaluate model performance using real data

RELATED LITERATURE

This project is based on Dive into Deep Learning (Zhang et al.), a hands-on textbook that integrates mathematical theory with coding practices. Additional references include documentation from PyTorch and insights from the Kaggle data science community on handling tabular data and model regularization.

METHODOLOGY

Learning Phase:

I studied the following topics in-depth:

- **Chapter 1–2:** Linear regression vs. MLPs; why nonlinearity is needed
- **Chapter 3:** Forward & backward propagation
- **Chapter 4:** Training loop, gradient descent, loss functions
- **Chapter 5:** Numerical instability (vanishing/exploding gradients), Xavier initialization, Dropout, and L2 regularization

Project Phase:

Dataset:

Kaggle House Prices (79 features, 2000+ records)

Models:

- Linear Regression (baseline)
- Basic MLP
- MLP + ReLU + Dropout + L2

Evaluation:

5-fold cross-validation using MSE

RESULTS/FINDINGS

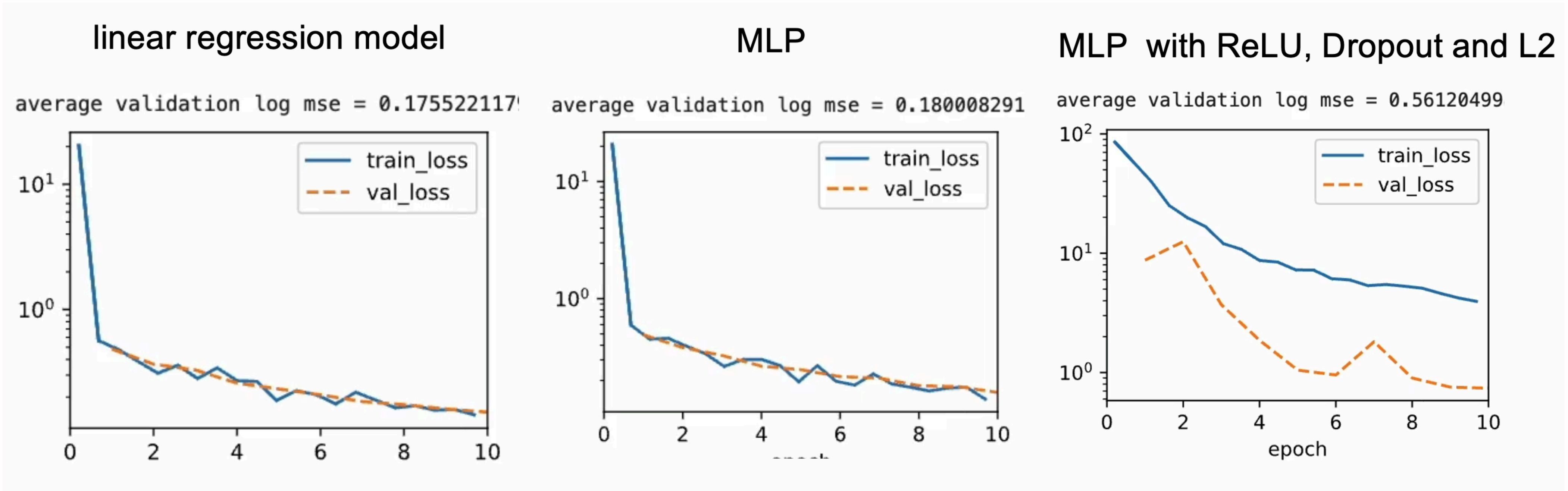
- Linear regression achieved the lowest validation MSE, outperforming the more complex MLPs
- Adding ReLU, dropout, and L2 **did not** improve model performance
- This suggests that complexity does not guarantee better performance in structured/tabular data

- Concepts like Xavier initialization and Dropout were critical to stabilizing training, but tuning remains challenging

ANALYSIS

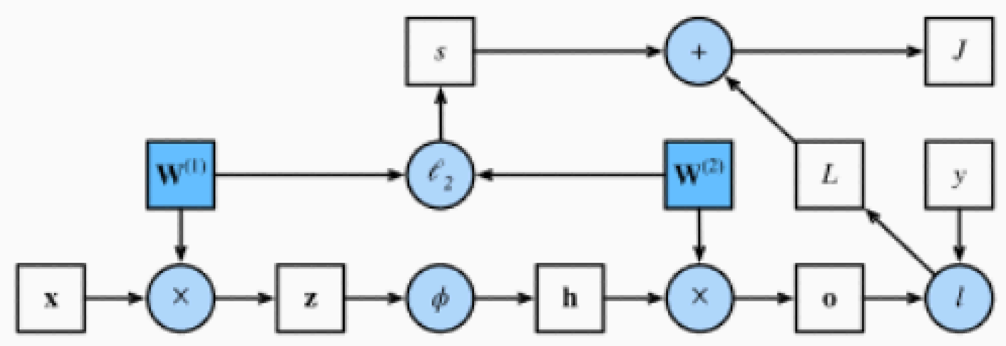
This project applies deep learning techniques to predict house prices using the Kaggle dataset. The three graphs compare a linear regression model, a basic MLP, and an MLP with ReLU, dropout, and L2 regularization.

- Linear regression achieved the lowest validation log MSE (0.1755) and performed consistently well on the test set, indicating strong generalization.
- Basic MLP had comparable validation performance (0.1800), but showed slightly higher variance between validation and test loss.
- The complex MLP performed worst (val log MSE = 0.5612), with a clear gap between training and validation, suggesting over-regularization and possible underfitting.



Forward Propagation:

Pass the input through the model to get predictions (y_hat)




```
output = activation(Wx + b)
```

Loss Function:

Compare the predicted values y_hat with the true labels y

```
loss = loss_function(y_hat, y)
```

Backward Propagation:

Compute the gradients of the loss with respect to each parameter using the chain rule.

```
loss.backward()
```

Update the weights and Zero gradients:

Use optimizer updates the weights and biases using the gradients, which must be zeroed afterward to prevent accumulation in the next iteration.

```
optimizer.step()
optimizer.zero_grad()
```

CONCLUSION

This Directed Reading Program has provided a valuable opportunity to deepen my theoretical and practical understanding of neural networks. Through a focused study of the first five chapters of Dive into Deep Learning, I developed a solid foundation in key concepts such as multilayer perceptrons, gradient-based optimization, and regularization techniques.

In parallel, regular discussions with my mentor significantly enhanced my learning experience. These interactions not only clarified complex technical ideas but also offered guidance on best practices in model implementation and evaluation.

The final project, predicting house prices using both linear models and neural networks, highlighted a critical insight: effective modeling requires not only technical proficiency but also sound judgment in choosing appropriate methods for specific data contexts. This experience has strengthened my ability to apply machine learning techniques thoughtfully, interpret results critically, and approach future work in deep learning with greater confidence and rigor.