Hanna Syed
DAT412
Assignment 5
Neural Networks for Predicting Housing Prices

Objective

This project implements a neural network model to predict housing prices based on size, location score, and number of bedrooms. The model is trained on synthetic data and evaluate its performance.

Dataset Description

- The dataset consists of 1,000 samples from housing1.csv
- Features: Size (sqft), Location Score (1-100), Number of Bedrooms
- Target Variable: House Price (dollars)
- The formula for generating data:
    - Y = (3) x (Size + 2) x (Location Score + 4) x (Bedrooms) + Error
- Normalized the features and split them into training and testing sets
- Split the data into 80% for training and 20% for testing to see how well the model performs

Neural Network Architecture

- Input Layer: 3 input features
- Hidden Layers:
    - Layer 1: 64 neurons, ReLU activation
    - Layer 2: 32 neurons, ReLU activation
- Output Layer: 1 neuron (regression)
- Optimizer: Adam (learning rate = 0.01)
- Loss Function: Mean Squared Error (MSE)
- Metric: Mean Absolute Error (MAE)

Training and Evaluation Results

- Training Process:
    - Model trained for 50 epochs with a size of 32
    - The validation split is set at 20%

- Training Loss:
    - Started at 104140.3047 in epoch 1 and decreased rapidly

    - Final training loss: 100.3518
    - Shows that the model is effectively learning from the training data
    - The model is successfully fitting to the data and improving its predictions on the training set

- Validation Loss:
    - Started at 104140.3047 and followed a decreasing trend.
    - Final validation loss: 105.0513
    - Some fluctuation in the validation loss, where the loss increases slightly before decreasing again
    - Shows that the model is sometimes overfitting or not generalizing correctly
    - But overall it is decreasing meaning the model is stabilizing and improving

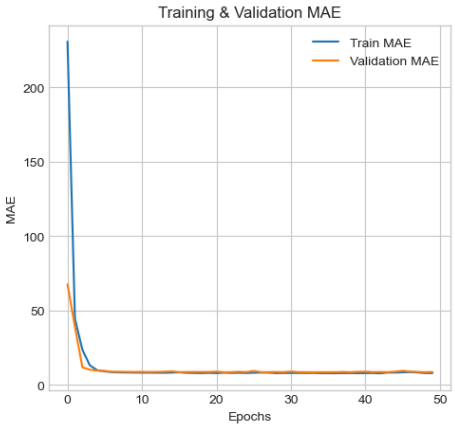
- Training MAE:
    - Started at 273.3638 and decreased to 7.9029
    - This shows the model is improving its ability to predict

- Validation MAE:
    - Started at 43.8197 and improved to 8.5977
    - This shows the model is successfully generalizing to unseen data, with relatively small error

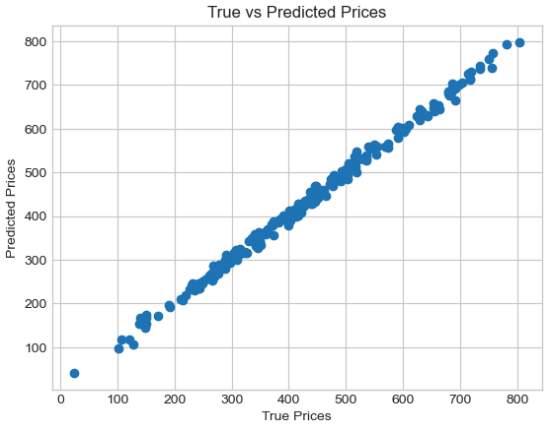
- Test Set Performance:
    - Test Loss (MSE): 113.65978240966797
    - Test MAE: 8.678581237792969
    - Shows that the model has learned from the input features and can make accurate predictions on unseen data

Visualizations

- Training & Validation Loss - Shows model convergence and overfitting signs.
- Training & Validation MAE - Shows model performance on unseen data.



- True vs Predicted Prices - Shows prediction accuracy and trends.



Observations and Insights

- Overfitting:
    - No significant signs of overfitting.
    - Overfitting happens when training loss keeps decreasing while validation loss increases, which is not the case here.
    - Training and validation losses decrease consistently, indicating the model is not memorizing the training data.
    - Small fluctuations in validation loss during early epochs could indicate slight overfitting, but the model stabilizes by the end
- Underfitting:
    - The model does not appear to be underfitting.
    - Underfitting happens when loss values remain high and do not improve, which is not observed in this case
    - Both training and validation loss steadily decrease, showing the model is learning patterns effectively
- The model generalized well, as indicated by the test loss and MAE
- Minimal divergence in loss values suggests the model did not overfit
- The loss values show steady convergence towards lower values, indicating a trained model
- Model Improvement:
    - Incorporating additional factors could make predictions more precise. For example, neighborhood quality, crime rate, walkability, etc.
    - Dropout layers and L2 regularization can help prevent the model from memorizing training data instead of learning patterns
    - Adjusting settings like learning rate, batch size, and network structure can improve accuracy and stability

Conclusion

- The model effectively predicts housing prices with low error.
- Performance on validation and test sets suggests good generalization.
- Future refinements can further enhance accuracy and robustness.