Hanna Syed
DAT412
Assignment 6
Neural Networks for Predicting Housing Prices

Objective

This project builds a neural network model to predict housing prices using three features: size, location score, and number of bedrooms. The model is trained, evaluated, and visualized using real housing data.

Dataset Description

- The dataset contains 500 housing records.
- Features:
    - Size (in square feet)
    - Location Score (rated 1–10)
      Number of Bedrooms
- Target Variable: House Price (USD)
- The features were normalized using StandardScaler.
- The data was split into 80% training and 20% testing.

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 (for price prediction)
  Optimizer: Adam (learning rate = 0.01)
  Loss Function: Mean Squared Error (MSE)
- Metric: Mean Absolute Error (MAE)

Training and Evaluation Results

- The model was trained for 50 epochs with a batch size of 32.
- 20% of the training set was used for validation.
- Training and validation loss decreased consistently throughout training.
- Final training loss: 1,157,646,848.0
- Final validation loss: 1,519,882,496.0
- Final training MAE: 28,320.23
- Final validation MAE: 32,793.82
- Test Loss (MSE): 1,242,353,280.0
- Test MAE: 27,912.81

- Shows that the model is effectively learning and generalizing to unseen data.

Visualizations of Predictions

- Training vs Validation Loss plot showed a consistent decrease and no divergence.
- True vs Predicted Prices scatter plot indicated strong alignment with minor deviations.
- No major outliers or systematic errors were observed.

Insights from Model Performance

- Overfitting:
    - No significant signs of overfitting.
    - Training and validation losses decreased consistently.
    - Minor fluctuations early in training stabilized over time.
- Underfitting:
    - No evidence of underfitting.
    - Both training and validation loss values steadily improved.
- Generalization:
    - The model performed well on test data, as seen in the low MAE.
    - Indicates that it learned useful patterns from the features.
- Model Improvements:
    - Additional features like neighborhood quality or crime rates could improve accuracy.
    - Dropout layers and L2 regularization can reduce risk of overfitting.
    - Tuning batch size, learning rate, or network depth could further optimize performance.

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.