Mohammad Bagher Soroush - 400130273 - ML 3 Report

1 .Explanation of Process Steps

The notebook involves the following:

Data Preparation:

- The Boston Housing dataset is loaded.
- Outliers are removed using the Z-score method with a threshold (e.g., |Z| < 3).

Normalization:

- Features are split into groups and processed using:
- Z-score normalization for features sensitive to scale and requiring standardization.
- Min-Max normalization for features requiring bounded values (e.g., between 0 and 1).

Model Training:

Two Linear Regression models are trained:

- On normalized data.
- On non-normalized data.
- Data is split into training and testing sets using train_test_split.

Model Evaluation:

- Metrics such as MAE, MSE, RMSE, and R2 are computed and compared.

Visualization:

- Evaluation metrics are visualized through bar plots to compare normalized and non-normalized models..

2 .Why Apply Min-Max or Z-score Normalization?

- Z-score Normalization: Applied when data has outliers or needs centering around zero (e.g., lstat, age). This ensures features have mean
- 0 and standard deviation 1 ,improving regression stability.
- Min-Max Normalization: Used when values need to fall in a fixed range (e.g., indus, rm), essential for models sensitive to varying scales.

3 .Model Comparison

- Metrics comparison (e.g., MAE, MSE) shows marginal differences between normalized and non-normalized data.
- Normalization improves stability and can handle poorly scaled data better.

4 .Outlier Removal with Z-score

- Each data point is converted to a Z-score.
- Data points with threshold |Z| > threshold are treated as outliers and removed.
- Helps minimize skewness and ensures the model isn't biased by extreme values.

5 .Graph Interpretation

Bar plots compare models across metrics:

- Normalized Model: Slightly higher R2 and better error metrics due to improved feature scaling.
- Non-Normalized Model: Performs similarly but might face instability with different datasets.