

Brief Report

Findings

1. **Data Loading and Exploration:**
 - The dataset was successfully loaded and inspected.
 - Initial exploration shows that there are 506 entries and 14 columns in the dataset.
 - No missing values were found in the dataset.
2. **Data Normalization and Standardization:**
 - Specific columns ('CRIM', 'ZN', 'TAX', 'B') were normalized to the range [0, 1].
 - The same columns were also standardized using `StandardScaler`.
3. **Modeling:**
 - Linear regression was performed using the features excluding 'RAD' and the target variable 'MEDV'.
 - The model was trained on 70% of the data and tested on the remaining 30%.
4. **Model Performance:**
 - The scatter plot of true vs. predicted values shows the relationship between actual house prices and the predicted prices from the model.
5. **Coefficients:**
 - The coefficients for each feature were extracted, indicating the weight each feature has in predicting the target variable 'MEDV'.

Model's Performance

- The model's performance can be assessed by further metrics such as R-squared, Mean Squared Error (MSE), etc., which were not included in the provided code but can be calculated if needed.

Challenges Faced

- The initial issue with loading the data correctly due to the delimiter was resolved by specifying `delim_whitespace=True`.
- Ensuring the data was correctly formatted and clean was crucial for accurate model training.

Overall, the linear regression model provides insights into the relationships between the features and the target variable 'MEDV'. Further analysis and improvement can be done by exploring more advanced models or techniques.

Dataset and Its Features

Dataset: Boston Housing Dataset

The Boston Housing dataset contains information collected by the U.S Census Service concerning housing in the area of Boston, Massachusetts. The dataset contains 506 observations of 14 different features.

Features Description

1. **CRIM**: Per capita crime rate by town.
2. **ZN**: Proportion of residential land zoned for lots over 25,000 sq. ft.
3. **INDUS**: Proportion of non-retail business acres per town.
4. **CHAS**: Charles River dummy variable (1 if tract bounds river; 0 otherwise).
5. **NOX**: Nitric oxide concentration (parts per 10 million).
6. **RM**: Average number of rooms per dwelling.
7. **AGE**: Proportion of owner-occupied units built before 1940.
8. **DIS**: Weighted distances to five Boston employment centers.
9. **RAD**: Index of accessibility to radial highways.
10. **TAX**: Full-value property tax rate per \$10,000.
11. **PTRATIO**: Pupil-teacher ratio by town.
12. **B**: $1000(B_k - 0.63)^2$ where B_k is the proportion of Black residents by town.
13. **LSTAT**: Percentage of lower status of the population.
14. **MEDV**: Median value of owner-occupied homes in \$1000s.