

Data Set Information:

Concerns housing values in suburbs of Boston.

Attribute Information:

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 oxides concentration (parts per 10 million)
6. RM: average number of rooms per dwelling
7. AGE: proportion of owner-occupied units built prior to 1940
8. DIS: weighted distances to five Boston employment centres
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 blacks by town
13. LSTAT: % lower status of the population
14. MEDV: Median value of owner-occupied homes in \$1000's

Data Set Characteristics:	Multivariate	Number of Instances:	506	Area:	N/A
Attribute Characteristics:	Categorical, Integer, Real	Number of Attributes:	14	Date Donated	1993-07-07
Associated Tasks:	Regression	Missing Values?	No	Number of Web Hits:	210506

Notebook 上面的:

As a reminder, we are using three features from the Boston housing dataset: "RM", "LSTAT", and "PTRATIO". For each data point (neighborhood):

- "RM" is the average number of rooms among homes in the neighborhood.
- "LSTAT" is the percentage of all Boston homeowners who have a greater net worth than homeowners in the neighborhood.
- "PTRATIO" is the ratio of students to teachers in primary and secondary schools in the neighborhood.