

## Context

To Explore more on Regression Algorithm

## Content

Each record in the database describes a Boston suburb or town. The data was drawn from the Boston Standard Metropolitan Statistical Area (SMSA) in 1970. The attributes are defined as follows (taken from the UCI Machine Learning Repository<sup>1</sup>): CRIM: per capita crime rate by town

- 1. ZN: proportion of residential land zoned for lots over 25,000 sq.ft.
- 2. INDUS: proportion of non-retail business acres per town
- 3. CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
- 4. NOX: nitric oxides concentration (parts per 10 million)  
1<https://archive.ics.uci.edu/ml/datasets/Housing>  
123  
20.2. Load the Dataset 124
- 5. RM: average number of rooms per dwelling
- 6. AGE: proportion of owner-occupied units built prior to 1940
- 7. DIS: weighted distances to five Boston employment centers
- 8. RAD: index of accessibility to radial highways
- 9. TAX: full-value property-tax rate per \$10,000
- 10. 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
- 11. MEDV: Median value of owner-occupied homes in \$1000s  
We can see that the input attributes have a mixture of units.

## Acknowledgements

Thanks to Dr.Jason