Housing Data Set

Download: Data Folder, Data Set Description

Abstract: Taken from StatLib library

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:**	186182

Source:

Origin:

This dataset was taken from the StatLib library which is maintained at Carnegie Mellon University.

Creator:

Harrison, D. and Rubinfeld, D.L.
'Hedonic prices and the demand for clean air', J. Environ. Economics & Management, vol.5, 81-102, 1978.

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(Bk - 0.63)^2 where Bk 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

Relevant Papers:

Belsley, Kuh & Welsch, 'Regression diagnostics: Identifying Influential Data and Sources of Collinearity', Wiley, 1980. 244-261.
[Web Link]

Quinlan,R. (1993). Combining Instance-Based and Model-Based Learning. In Proceedings on the Tenth International Conference of Machine Learning, 236-243, University of Massachusetts, Amherst. Morgan Kaufmann.
[Web Link]

Papers That Cite This Data Set[1]:

Manuel Oliveira. Library Release Form Name of Author: Stanley Robson de Medeiros Oliveira Title of Thesis: Data Transformation For Privacy-Preserving Data Mining Degree: Doctor of Philosophy Year this Degree Granted. University of Alberta Library. 2005. [View Context].

Gavin Brown. Diversity in Neural Network Ensembles. The University of Birmingham. 2004. [View Context].

Predrag Radivojac and Zoran Obradovic and A. Keith Dunker and Slobodan Vucetic. Feature Selection Filters Based on the Permutation Test. ECML. 2004. [View Context].

Glenn Fung and M. Murat Dundar and Jinbo Bi and Bharat Rao. A fast iterative algorithm for fisher discriminant using heterogeneous kernels. ICML. 2004. [View Context].

Kristiaan Pelckmans and Jos De Brabanter and J. A. K Suykens and Bart De Moor and K. U. Leuven - ESAT. The Differogram: Non-parametric Noise Variance Estimation and its Use for Model Selection. SCDSISTA. 2004. [View Context].

Bart Hamers and J. A. K Suykens. Coupled Transductive Ensemble Learning of Kernel Models. Bart De Moor. 2003. [View Context].

Christopher K I Williams and Carl Edward Rasmussen and Anton Schwaighofer and Volker Tresp. Observations on the Nystrom Method for Gaussian Process Prediction. Division of Informatics Gatsby Computational Neuroscience Unit University of Edinburgh University College London. 2002. [View Context].

David Hershberger and Hillol Kargupta. Distributed Multivariate Regression Using Wavelet-Based Collective Data Mining. J. Parallel Distrib. Comput, 61. 2001. [View Context].

Thomas Melluish and Craig Saunders and Ilia Nouretdinov and Volodya Vovk and Carol S. Saunders and I. Nouretdinov V.. The typicalness framework: a comparison with the Bayesian approach. Department of Computer Science. 2001. [View Context].

Martin H C Law and James T. Kwok. Applying the Bayesian Evidence Framework to u -Support Vector Regression. ECML. 2001. [View Context].

Peter L. Hammer and Alexander Kogan and Bruno Simeone and Sandor Szedm'ak. R u t c o r Research R e p o r t. Rutgers Center for Operations Research Rutgers University. 2001. [View Context].

Zhi-Hua Zhou and Jianping Wu and Weiyu Tang and Zen Chen. Combining Regression Estimators: GA-Based Selective Neural Network Ensemble. International Journal of Computational Intelligence and Applications, 1. 2001. [View Context].

Nir Friedman and Iftach Nachman. Gaussian Process Networks. UAI. 2000. [View Context].

Endre Boros and Peter Hammer and Toshihide Ibaraki and Alexander Kogan and Eddy Mayoraz and Ilya B. Muchnik. An Implementation of Logical Analysis of Data. IEEE Trans. Knowl. Data Eng, 12. 2000. [View Context].

Rudy Setiono and Huan Liu. A connectionist approach to generating oblique decision trees. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 29. 1999. [View Context].

Jinyan Li and Xiuzhen Zhang and Guozhu Dong and Kotagiri Ramamohanarao and Qun Sun. Efficient Mining of High Confidience Association Rules without Support Thresholds. PKDD. 1999. [View Context].

Christopher J. Merz and Michael J. Pazzani. A Principal Components Approach to Combining Regression Estimates. Machine Learning, 36. 1999. [View Context].

H. Altay Guvenir and Ilhan Uysal. Regression on feature projections. a Department of Computer Engineering, Bilkent University. 1999. [View Context].

Ayhan Demiriz and Kristin P. Bennett and Mark J. Embrechts. Semi-Supervised Clustering Using Genetic Algorithms. Dept. 1999. [View Context].

Mauro Birattari and Gianluca Bontempi and Hugues Bersini. Lazy Learning Meets the Recursive Least Squares Algorithm. NIPS. 1998. [View Context].

Huan Liu and Rudy Setiono. Feature Transformation and Multivariate Decision Tree Induction. Discovery Science. 1998. [View Context].

Sreerama K. Murthy and Simon Kasif and Steven Salzberg. A System for Induction of Oblique Decision Trees. Department of Computer Science Johns Hopkins University. 1994. [View Context].

Yin Zhang and W. Nick Street. Bagging with Adaptive Costs. Management Sciences Department University of Iowa Iowa City. [View Context].

Tapani Raiko and Harri Valpola. MISSING VALUES IN NONLINEAR FACTOR ANALYSIS. Helsinki University of Technology, Neural Networks Research Centre. [View Context].

Dorian Suc and Ivan Bratko. Combining Learning Constraints and Numerical Regression. National ICT Australia, Sydney Laboratory at UNSW. [View Context].

Ayhan Demiriz and Kristin P. Bennett. Chapter 1 OPTIMIZATIONAPPROACHESTOSEMI-SUPERVISED LEARNING. Department of Decision Sciences and Engineering Systems & Department of Mathematical Sciences, Rensselaer Polytechnic Institute. [View Context].

Luc Hoegaerts and J. A. K Suykens and J. Vandewalle and Bart De Moor. Subset Based Least Squares Subspace Regression in RKHS. Katholieke Universiteit Leuven Department of Electrical Engineering, ESAT-SCD-SISTA. [View Context].

S. Sathiya Keerthi. Improvements to SMO Algorithm for SVM Regression. Author for Correspondence: Prof. [View

Context].

Jarkko Tikka. AB HELSINKI UNIVERSITY OF TECHNOLOGY Department of Automation and Systems Technology Jarkko Tikka Learning linear dependency trees from multivariate data. Helsinki University of Technology Abstract of Master's thesis Department of Automation and Systems Technology Author Date. [View Context].

David R. Musicant. DATA MINING VIA MATHEMATICAL PROGRAMMING AND MACHINE LEARNING. Doctor of Philosophy (Computer Sciences) UNIVERSITY. [View Context].

Ayhan Demiriz and Kristin P. Bennett and John Shawe and I. Nouretdinov V.. Linear Programming Boosting via Column Generation. Dept. of Decision Sciences and Eng. Systems, Rensselaer Polytechnic Institute. [View Context].

Jianping Wu and Zhi-Hua Zhou and Cheng-The Chen. Ensemble of GA based Selective Neural Network Ensembles. National Laboratory for Novel Software Technology Nanjing University. [View Context].

C. Titus Brown and Harry W. Bullen and Sean P. Kelly and Robert K. Xiao and Steven G. Satterfield and John G. Hagedorn and Judith E. Devaney. Visualization and Data Mining in an 3D Immersive Environment: Summer Project 2003. [View Context].

David R. Musicant and Alexander Feinberg. Active Set Support Vector Regression. [View Context].

Nir Friedman and Daphne Koller (koller@cs. stanford. edu. A Bayesian Approach to Structure Discovery in Bayesian Networks. School of Computer Science & Engineering Hebrew University. [View Context].

Citation Request:

Please refer to the Machine Learning Repository's citation policy

---

[1] Papers were automatically harvested and associated with this data set, in collaboration with Rexa.info