About Citation Policy Donate a Data Set Contact

Search

◉ Repository ◯ Web

Google™

View ALL Data Sets

Center for Machine Learning and Intelligent Systems

Auto MPG Data Set

Download: Data Folder, Data Set Description

Abstract: Revised from CMU StatLib library, data concerns city-cycle fuel consumption

Data Set Characteristics:	Multivariate	Number of Instances:	398	Area:	N/A
Attribute Characteristics:	Categorical, Real	Number of Attributes:	8	Date Donated	1993-07-07
Associated Tasks:	Regression	Missing Values?	Yes	Number of Web Hits:	406994

Source:

This dataset was taken from the StatLib library which is maintained at Carnegie Mellon University. The dataset was used in the 1983 American Statistical Association Exposition.

Data Set Information:

This dataset is a slightly modified version of the dataset provided in the StatLib library. In line with the use by Ross Quinlan (1993) in predicting the attribute "mpg", 8 of the original instances were removed because they had unknown values for the "mpg" attribute. The original dataset is available in the file "auto-mpg.data-original".

"The data concerns city-cycle fuel consumption in miles per gallon, to be predicted in terms of 3 multivalued discrete and 5 continuous attributes." (Quinlan, 1993)

Attribute Information:

1. mpg: continuous
2. cylinders: multi-valued discrete
3. displacement: continuous
4. horsepower: continuous
5. weight: continuous
6. acceleration: continuous
7. model year: multi-valued discrete
8. origin: multi-valued discrete

9. car name: string (unique for each instance)

Relevant Papers:

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]:

Dan Pelleg. Scalable and Practical Probability Density Estimators for Scientific Anomaly Detection. School of Computer Science Carnegie Mellon University. 2004. [View Context].

Qingping Tao Ph. D. MAKING EFFICIENT LEARNING ALGORITHMS WITH EXPONENTIALLY MANY FEATURES. Qingping Tao A DISSERTATION Faculty of The Graduate College University of Nebraska In Partial Fulfillment of Requirements. 2004. [View Context].

Christopher R. Palmer and Christos Faloutsos. Electricity Based External Similarity of Categorical Attributes. PAKDD. 2003. [View Context].

Wai Lam and Kin Keung and Charles X. Ling. PR 1527. Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong. 2001. [View Context].

Dan Pelleg and Andrew W. Moore. Mixtures of Rectangles: Interpretable Soft Clustering. ICML. 2001. [View Context].

Jinyan Li and Kotagiri Ramamohanarao and Guozhu Dong. Combining the Strength of Pattern Frequency and Distance for Classification. PAKDD. 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].

Zhi-Hua Zhou and Shifu Chen and Zhaoqian Chen. A Statistics Based Approach for Extracting Priority Rules from Trained Neural Networks. IJCNN (3). 2000. [View Context].

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

D. Greig and Hava T. Siegelmann and Michael Zibulevsky. A New Class of Sigmoid Activation Functions That Don't Saturate. 1997. [View Context].

Johannes Furnkranz. Pairwise Classification as an Ensemble Technique. Austrian Research Institute for Artificial Intelligence. [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].

Citation Request:

Please refer to the Machine Learning Repository's <u>citation policy</u>

Supported By: In Collaboration With: