UCI Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Liver Disorders Data Set

Download: Data Folder, Data Set Description

Abstract: BUPA Medical Research Ltd. database donated by Richard S. Forsyth

Data Set Characteristics:	Multivariate	**Number of Instances:**	345	**Area:**	Life
Attribute Characteristics:	Categorical, Integer, Real	**Number of Attributes:**	7	**Date Donated**	1990-05-15
Associated Tasks:	N/A	**Missing Values?**	No	**Number of Web Hits:**	47750

Source:

Creators:

BUPA Medical Research Ltd.

Donor:

Richard S. Forsyth
8 Grosvenor Avenue
Mapperley Park
Nottingham NG3 5DX
0602-621676

Data Set Information:

The first 5 variables are all blood tests which are thought to be sensitive to liver disorders that might arise from excessive alcohol consumption. Each line in the bupa.data file constitutes the record of a single male individual.

It appears that drinks>5 is some sort of a selector on this database. See the PC/BEAGLE User's Guide for more information.

Attribute Information:

1. mcv mean corpuscular volume
2. alkphos alkaline phosphotase
3. sgpt alamine aminotransferase
4. sgot aspartate aminotransferase
5. gammagt gamma-glutamyl transpeptidase
6. drinks number of half-pint equivalents of alcoholic beverages drunk per day
7. selector field used to split data into two sets

Relevant Papers:

PC/BEAGLE User's Guide (written by Richard S. Forsyth).

Papers That Cite This Data Set[1]:

Zhi-Hua Zhou and Yuan Jiang. NeC4.5: Neural Ensemble Based C4.5. IEEE Trans. Knowl. Data Eng, 16. 2004. [View Context].

Yuan Jiang and Zhi-Hua Zhou. Editing Training Data for kNN Classifiers with Neural Network Ensemble. ISNN (1). 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].

Jochen Garcke and Michael Griebel. Classification with sparse grids using simplicial basis functions. Intell. Data Anal, 6. 2002. [View Context].

Michail Vlachos and Carlotta Domeniconi and Dimitrios Gunopulos and George Kollios and Nick Koudas. Non-linear dimensionality reduction techniques for classification and visualization. KDD. 2002. [View Context].

Xavier Llor and David E. Goldberg and Ivan Traus and Ester Bernad i Mansilla. Accuracy, Parsimony, and Generality in Evolutionary Learning Systems via Multiobjective Selection. IWLCS. 2002. [View Context].

Jochen Garcke and Michael Griebel. Data mining with sparse grids using simplicial basis functions. KDD. 2001. [View Context].

Jochen Garcke and Michael Griebel and Michael Thess. Data Mining with Sparse Grids. Computing, 67. 2001. [View Context].

Petri Kontkanen and Jussi Lahtinen and Petri Myllymäki and Henry Tirri. Unsupervised Bayesian visualization of high-dimensional data. KDD. 2000. [View Context].

Carlotta Domeniconi and Jing Peng and Dimitrios Gunopulos. An Adaptive Metric Machine for Pattern Classification. NIPS. 2000. [View Context].

Iñaki Inza and Pedro Larrañaga and Basilio Sierra and Ramon Etxeberria and Jose Antonio Lozano and Jos Manuel Peña. Representing the behaviour of supervised classification learning algorithms by Bayesian networks. Pattern Recognition Letters, 20. 1999. [View Context].

Guido Lindner and Rudi Studer. AST: Support for Algorithm Selection with a CBR Approach. PKDD. 1999. [View Context].

Kristin P. Bennett and Erin J. Bredensteiner. A Parametric Optimization Method for Machine Learning. INFORMS Journal on Computing, 9. 1997. [View Context].

Jennifer A. Blue and Kristin P. Bennett. Hybrid Extreme Point Tabu Search. Department of Mathematical Sciences Rensselaer Polytechnic Institute. 1996. [View Context].

Peter D. Turney. Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm. CoRR, csAI/9503102. 1995. [View Context].

Gabor Melli. A Lazy Model-Based Approach to On-Line Classification. University of British Columbia. 1989. [View Context].

Aynur Akku and H. Altay Guvenir. Weighting Features in k Nearest Neighbor Classification on Feature Projections. Department of Computer Engineering and Information Science Bilkent University. [View Context].

Greg Ridgeway. The State of Boosting. Department of Statistics University of Washington. [View Context].

Adil M. Bagirov and Alex Rubinov and A. N. Soukhojak and John Yearwood. Unsupervised and supervised data classification via nonsmooth and global optimization. School of Information Technology and Mathematical Sciences, The University of Ballarat. [View Context].

Adil M. Bagirov and John Yearwood. A new nonsmooth optimization algorithm for clustering. Centre for Informatics and Applied Optimization, School of Information Technology and Mathematical Sciences, University of Ballarat. [View Context].

H. Altay G uvenir and Aynur Akkus. WEIGHTED K NEAREST NEIGHBOR CLASSIFICATION ON FEATURE PROJECTIONS. Department of Computer Engineering and Information Science Bilkent 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. DATA MINING VIA MATHEMATICAL PROGRAMMING AND MACHINE LEARNING. Doctor of Philosophy (Computer Sciences) 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