UCI

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Search

○ Repository ○ Web

Google

View ALL Data Sets

Statlog (Shuttle) Data Set

Download: Data Folder, Data Set Description

Abstract: The shuttle dataset contains 9 attributes all of which are numerical. Approximately 80% of the data belongs to class 1

Data Set Characteristics:	Multivariate	Number of Instances:	58000	Area:	Physical
Attribute Characteristics:	Integer	Number of Attributes:	9	Date Donated	N/A
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	130281

Source:

Donor:

Jason Catlett
Basser Department of Computer Science,
University of Sydney, N.S.W., Australia

Data Set Information:

Approximately 80% of the data belongs to class 1. Therefore the default accuracy is about 80%. The aim here is to obtain an accuracy of 99 - 99.9%.

The examples in the original dataset were in time order, and this time order could presumably be relevant in classification. However, this was not deemed relevant for StatLog purposes, so the order of the examples in the original dataset was randomised, and a portion of the original dataset removed for validation purposes.

Attribute Information:

The shuttle dataset contains 9 attributes all of which are numerical. The first one being time. The last column is the class which has been coded as follows :
1 Rad Flow
2 Fpv Close
3 Fpv Open
4 High
5 Bypass
6 Bpv Close
7 Bpv Open

Relevant Papers:

N/A

Papers That Cite This Data Set[1]:

Ira Cohen and Fabio Gagliardi Cozman and Nicu Sebe and Marcelo Cesar Cirelo and Thomas S. Huang. Semisupervised Learning of Classifiers: Theory, Algorithms, and Their Application to Human-Computer Interaction. IEEE Trans. Pattern Anal. Mach. Intell, 26. 2004. [View Context].

Jun Wang and Bin Yu and Les Gasser. Concept Tree Based Clustering Visualization with Shaded Similarity Matrices. ICDM. 2002. [View Context].

Richard Nock. Inducing Interpretable Voting Classifiers without Trading Accuracy for Simplicity: Theoretical Results, Approximation Algorithms, and Experiments. J. Artif. Intell. Res. (JAIR, 17. 2002. [View Context].

Grigorios Tsoumakas and Ioannis P. Vlahavas. Effective Stacking of Distributed Classifiers. ECAI. 2002. [View Context].

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

Stephen D. Bay. Multivariate Discretization for Set Mining. Knowl. Inf. Syst, 3. 2001. [View Context].

Haixun Wang and Carlo Zaniolo. CMP: A Fast Decision Tree Classifier Using Multivariate Predictions. ICDE. 2000. [View Context].

Khaled A. Alsabti and Sanjay Ranka and Vineet Singh. CLOUDS: A Decision Tree Classifier for Large Datasets. KDD. 1998. [View Context].

Nir Friedman and Moisés Goldszmidt. Discretizing Continuous Attributes While Learning Bayesian Networks. ICML. 1996. [View Context].

Ron Kohavi. Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid. KDD. 1996. [View Context].

Pedro Domingos. Linear-Time Rule Induction. KDD. 1996. [View Context].

Ron Kohavi. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. IJCAI. 1995. [View Context].

Ron Kohavi. The Power of Decision Tables. ECML. 1995. [View Context].

Ron Kohavi and George H. John. Automatic Parameter Selection by Minimizing Estimated Error. Computer Science Dept. Stanford University. [View Context].

Wl odzisl/aw Duch and Rudy Setiono and Jacek M. Zurada. Computational intelligence methods for rule-based data understanding. [View Context].

Chris Giannella and Bassem Sayrafi. An Information Theoretic Histogram for Single Dimensional Selectivity Estimation. Department of Computer Science, Indiana University Bloomington. [View Context].

Christophe Giraud and Tony Martinez. ADYNAMIC INCREMENTAL NETWORK THAT LEARNS BY DISCRIMINATION. AA. [View Context].

Wl odzisl and Rafal Adamczak and Krzysztof Grabczewski. Optimization of Logical Rules Derived by Neural Procedures. Department of Computer Methods, Nicholas Copernicus University. [View Context].

Chih-Wei Hsu and Cheng-Ru Lin. A Comparison of Methods for Multi-class Support Vector Machines. Department of Computer Science and Information Engineering National Taiwan University. [View Context].

Jeffrey P. Bradford and Clayton Kunz and Ron Kohavi and Clifford Brunk and Carla Brodley. Appears in ECML-98 as a research note Pruning Decision Trees with Misclassification Costs. School of Electrical Engineering. [View Context].

Jun Wang. Classification Visualization with Shaded Similarity Matrix. Bei Yu Les Gasser Graduate School of Library and Information Science University of Illinois at Urbana-Champaign. [View Context].

Krzysztof Grabczewski and Wl/odzisl/aw Duch. THE SEPARABILITY OF SPLIT VALUE CRITERION. Department of Computer Methods, Nicolaus Copernicus University. [View Context].

Mohammed Waleed Kadous and Claude Sammut. The University of New South Wales School of Computer Science and Engineering Temporal Classification: Extending the Classification Paradigm to Multivariate Time Series. [View Context].

Adil M. Bagirov and Julien Ugon. An algorithm for computation of piecewise linear function separating two sets. CIAO, School of Information Technology and Mathematical Sciences, The University of Ballarat. [View Context].

Citation Request:

Thanks to NASA for allowing us to use the shuttle datasets.

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

Supported By: In Collaboration With:

About || Citation Policy || Donation Policy || Contact || CML