Iris Data Set

Download: Data Folder, Data Set Description

Abstract: Famous database; from Fisher, 1936

Data Set Characteristics:	Multivariate	**Number of Instances:**	150	**Area:**	Life
Attribute Characteristics:	Real	**Number of Attributes:**	4	**Date Donated**	1988-07-01
Associated Tasks:	Classification	**Missing Values?**	No	**Number of Web Hits:**	1324564

Source:

Creator:

R.A. Fisher

Donor:

Michael Marshall (MARSHALL%PLU '@' io.arc.nasa.gov)

Data Set Information:

This is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. (See Duda & Hart, for example.) The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other.

Predicted attribute: class of iris plant.

This is an exceedingly simple domain.

This data differs from the data presented in Fishers article (identified by Steve Chadwick, spchadwick '@' espeedaz.net). The 35th

sample should be: 4.9,3.1,1.5,0.2,"Iris-setosa" where the error is in the fourth feature. The 38th sample: 4.9,3.6,1.4,0.1,"Iris-setosa" where the errors are in the second and third features.

Attribute Information:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
-- Iris Setosa
-- Iris Versicolour
-- Iris Virginica

Relevant Papers:

Fisher,R.A. "The use of multiple measurements in taxonomic problems" Annual Eugenics, 7, Part II, 179-188 (1936); also in "Contributions to Mathematical Statistics" (John Wiley, NY, 1950).
[Web Link]

Duda,R.O., & Hart,P.E. (1973) Pattern Classification and Scene Analysis. (Q327.D83) John Wiley & Sons. ISBN 0-471-22361-1. See page 218.
[Web Link]

Dasarathy, B.V. (1980) "Nosing Around the Neighborhood: A New System Structure and Classification Rule for Recognition in Partially Exposed Environments". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, No. 1, 67-71.
[Web Link]

Gates, G.W. (1972) "The Reduced Nearest Neighbor Rule". IEEE Transactions on Information Theory, May 1972, 431-433.
[Web Link]

See also: 1988 MLC Proceedings, 54-64.

Papers That Cite This Data Set[1]:

Ping Zhong and Masao Fukushima. A Regularized Nonsmooth Newton Method for Multi-class Support Vector Machines. 2005. [View Context].

Anthony K H Tung and Xin Xu and Beng Chin Ooi. CURLER: Finding and Visualizing Nonlinear Correlated Clusters. SIGMOD Conference. 2005. [View Context].

Igor Fischer and Jan Poland. Amplifying the Block Matrix Structure for Spectral Clustering. Telecommunications Lab. 2005. [View Context].

Sotiris B. Kotsiantis and Panayiotis E. Pintelas. Logitboost of Simple Bayesian Classifier. Informatica. 2005. [View Context].

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].

Jennifer G. Dy and Carla Brodley. Feature Selection for Unsupervised Learning. Journal of Machine Learning Research, 5. 2004. [View Context].

Jeroen Eggermont and Joost N. Kok and Walter A. Kosters. Genetic Programming for data classification: partitioning the search space. SAC. 2004. [View Context].

Remco R. Bouckaert and Eibe Frank. Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. PAKDD. 2004. [View Context].

Mikhail Bilenko and Sugato Basu and Raymond J. Mooney. Integrating constraints and metric learning in semi-supervised clustering. ICML. 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].

Yuan Jiang and Zhi-Hua Zhou. Editing Training Data for kNN Classifiers with Neural Network Ensemble. ISNN (1). 2004. [View Context].

Sugato Basu. Semi-Supervised Clustering with Limited Background Knowledge. AAAI. 2004. [View Context].

Judith E. Devaney and Steven G. Satterfield and John G. Hagedorn and John T. Kelso and Adele P. Peskin and William George and Terence J. Griffin and Howard K. Hung and Ronald D. Kriz. Science at the Speed of Thought. Ambient Intelligence for Scientific Discovery. 2004. [View Context].

Eibe Frank and Mark Hall. Visualizing Class Probability Estimators. PKDD. 2003. [View Context].

Ross J. Micheals and Patrick Grother and P. Jonathon Phillips. The NIST HumanID Evaluation Framework. AVBPA. 2003. [View Context].

Sugato Basu. Also Appears as Technical Report, UT-AI. PhD Proposal. 2003. [View Context].

Dick de Ridder and Olga Kouropteva and Oleg Okun and Matti Pietikäinen and Robert P W Duin. Supervised Locally Linear Embedding. ICANN. 2003. [View Context].

Aristidis Likas and Nikos A. Vlassis and Jakob J. Verbeek. The global k-means clustering algorithm. Pattern Recognition, 36. 2003. [View Context].

Zhi-Hua Zhou and Yuan Jiang and Shifu Chen. Extracting symbolic rules from trained neural network ensembles. AI Commun, 16. 2003. [View Context].

Jeremy Kubica and Andrew Moore. Probabilistic Noise Identification and Data Cleaning. ICDM. 2003. [View Context].

Julie Greensmith. New Frontiers For An Artificial Immune System. Digital Media Systems Laboratory HP Laboratories Bristol. 2003. [View Context].

Manoranjan Dash and Huan Liu and Peter Scheuermann and Kian-Lee Tan. Fast hierarchical clustering and its validation. Data Knowl. Eng, 44. 2003. [View Context].

Bob Ricks and Dan Ventura. Training a Quantum Neural Network. NIPS. 2003. [View Context].

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

Geoffrey Holmes and Bernhard Pfahringer and Richard Kirkby and Eibe Frank and Mark A. Hall. Multiclass Alternating Decision Trees. ECML. 2002. [View Context].

Inderjit S. Dhillon and Dharmendra S. Modha and W. Scott Spangler. Class visualization of high-dimensional data with applications. Department of Computer Sciences, University of Texas. 2002. [View Context].

Manoranjan Dash and Kiseok Choi and Peter Scheuermann and Huan Liu. Feature Selection for Clustering - A Filter Solution. ICDM. 2002. [View Context].

Ayhan Demiriz and Kristin P. Bennett and Mark J. Embrechts. A Genetic Algorithm Approach for Semi-Supervised Clustering. E-Business Department, Verizon Inc.. 2002. [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].

Jinyan Li and Guozhu Dong and Kotagiri Ramamohanarao and Limsoon Wong. DeEPs: A New Instance-based Discovery and Classification System. Proceedings of the Fourth European Conference on Principles and Practice of Knowledge Discovery in Databases. 2001. [View Context].

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

David Horn and A. Gottlieb. The Method of Quantum Clustering. NIPS. 2001. [View Context].

Asa Ben-Hur and David Horn and Hava T. Siegelmann and Vladimir Vapnik. A Support Vector Method for Clustering. NIPS. 2000. [View Context].

Neil Davey and Rod Adams and Mary J. George. The Architecture and Performance of a Stochastic Competitive Evolutionary Neural Tree Network. Appl. Intell, 12. 2000. [View Context].

Edgar Acuna and Alex Rojas. Ensembles of classifiers based on Kernel density estimators. Department of Mathematics University of Puerto Rico. 2000. [View Context].

Manoranjan Dash and Huan Liu. Feature Selection for Clustering. PAKDD. 2000. [View Context].

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

David M J Tax and Robert P W Duin. Support vector domain description. Pattern Recognition Letters, 20. 1999. [View Context].

Ismail Taha and Joydeep Ghosh. Symbolic Interpretation of Artificial Neural Networks. IEEE Trans. Knowl. Data Eng, 11. 1999. [View Context].

Stephen D. Bay. Combining Nearest Neighbor Classifiers Through Multiple Feature Subsets. ICML. 1998. [View Context].

Wojciech Kwedlo and Marek Kretowski. Discovery of Decision Rules from Databases: An Evolutionary Approach. PKDD. 1998. [View Context].

Foster J. Provost and Tom Fawcett and Ron Kohavi. The Case against Accuracy Estimation for Comparing Induction Algorithms. ICML. 1998. [View Context].

Igor Kononenko and Edvard Simec and Marko Robnik-Sikonja. Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF. Appl. Intell, 7. 1997. [View Context].

. Prototype Selection for Composite Nearest Neighbor Classifiers. Department of Computer Science University of Massachusetts. 1997. [View Context].

Ke Wang and Han Chong Goh. Minimum Splits Based Discretization for Continuous Features. IJCAI (2). 1997. [View Context].

Ethem Alpaydin. Voting over Multiple Condensed Nearest Neighbors. Artif. Intell. Rev, 11. 1997. [View Context].

Tapio Elomaa and Juho Rousu. Finding Optimal Multi-Splits for Numerical Attributes in Decision Tree Learning. ESPRIT Working Group in Neural and Computational Learning. 1996. [View Context].

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

Daniel C. St and Ralph W. Wilkerson and Cihan H. Dagli. RULE SET QUALITY MEASURES FOR INDUCTIVE LEARNING ALGORITHMS. proceedings of the Artificial Neural Networks In Engineering Conference 1996 (ANNIE. 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].

George H. John and Ron Kohavi and Karl Pfleger. Irrelevant Features and the Subset Selection Problem. ICML. 1994. [View Context].

Zoubin Ghahramani and Michael I. Jordan. Learning from incomplete data. MASSACHUSETTS INSTITUTE OF TECHNOLOGY ARTIFICIAL INTELLIGENCE LABORATORY and CENTER FOR BIOLOGICAL AND COMPUTATIONAL LEARNING DEPARTMENT OF BRAIN AND COGNITIVE SCIENCES. 1994. [View Context].

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

Wl odzisl/aw Duch and Karol Grudzinski. Prototype based rules - a new way to understand the data. Department of Computer Methods, Nicholas Copernicus University. [View Context].

H. Altay Guvenir. A Classification Learning Algorithm Robust to Irrelevant Features. Bilkent University, Department of Computer Engineering and Information Science. [View Context].

Ron Kohavi and Brian Frasca. Useful Feature Subsets and Rough Set Reducts. the Third International Workshop on Rough Sets and Soft Computing. [View Context].

Enes Makalic and Lloyd Allison and David L. Dowe. MML INFERENCE OF SINGLE-LAYER NEURAL NETWORKS. School of Computer Science and Software Engineering Monash University. [View Context].

G. Ratsch and B. Scholkopf and Alex Smola and Sebastian Mika and T. Onoda and K. -R Muller. Robust Ensemble Learning for Data Mining. GMD FIRST, Kekul#estr. [View Context].

YongSeog Kim and W. Nick Street and Filippo Menczer. Optimal Ensemble Construction via Meta-Evolutionary Ensembles. Business Information Systems, Utah State University. [View Context].

Maria Salamo and Elisabet Golobardes. Analysing Rough Sets weighting methods for Case-Based Reasoning Systems. Enginyeria i Arquitectura La Salle. [View Context].

Lawrence O. Hall and Nitesh V. Chawla and Kevin W. Bowyer. Combining Decision Trees Learned in Parallel. Department of Computer Science and Engineering, ENB 118 University of South Florida. [View Context].

Anthony Robins and Marcus Frean. Learning and generalisation in a stable network. Computer Science, The University of Otago. [View Context].

Geoffrey Holmes and Leonard E. Trigg. A Diagnostic Tool for Tree Based Supervised Classification Learning Algorithms. Department of Computer Science University of Waikato Hamilton New Zealand. [View Context].

Shlomo Dubnov and Ran El and Yaniv Technion and Yoram Gdalyahu and Elad Schneidman and Naftali Tishby and Golan Yona. Clustering By Friends : A New Nonparametric Pairwise Distance Based Clustering Algorithm. Ben Gurion University. [View Context].

Michael R. Berthold and Klaus--Peter Huber. From Radial to Rectangular Basis Functions: A new Approach for Rule Learning from Large Datasets. Institut fur Rechnerentwurf und Fehlertoleranz (Prof. D. Schmid) Universitat Karlsruhe. [View Context].

Norbert Jankowski. Survey of Neural Transfer Functions. Department of Computer Methods, Nicholas Copernicus University. [View Context].

Karthik Ramakrishnan. UNIVERSITY OF MINNESOTA. [View Context].

Wl/odzisl/aw Duch and Rafal Adamczak and Geerd H. F Diercksen. Neural Networks from Similarity Based Perspective. Department

of Computer Methods, Nicholas Copernicus University. [View Context].

Fernando Fern#andez and Pedro Isasi. Designing Nearest Neighbour Classifiers by the Evolution of a Population of Prototypes. Universidad Carlos III de Madrid. [View Context].

Asa Ben-Hur and David Horn and Hava T. Siegelmann and Vladimir Vapnik. A Support Vector Method for Hierarchical Clustering. Faculty of IE and Management Technion. [View Context].

Lawrence O. Hall and Nitesh V. Chawla and Kevin W. Bowyer. Decision Tree Learning on Very Large Data Sets. Department of Computer Science and Engineering, ENB 118 University of South Florida. [View Context].

G. Ratsch and B. Scholkopf and Alex Smola and K. -R Muller and T. Onoda and Sebastian Mika. Arc: Ensemble Learning in the Presence of Outliers. GMD FIRST. [View Context].

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

Huan Liu. A Family of Efficient Rule Generators. Department of Information Systems and Computer Science National University of Singapore. [View Context].

Rudy Setiono and Huan Liu. Fragmentation Problem and Automated Feature Construction. School of Computing National University of Singapore. [View Context].

Takao Mohri and Hidehiko Tanaka. An Optimal Weighting Criterion of Case Indexing for Both Numeric and Symbolic Attributes. Information Engineering Course, Faculty of Engineering The University of Tokyo. [View Context].

Fran ois Poulet. Cooperation between automatic algorithms, interactive algorithms and visualization tools for Visual Data Mining. ESIEA Recherche. [View Context].

Huan Li and Wenbin Chen. Supervised Local Tangent Space Alignment for Classification. I-Fan Shen. [View Context].

Adam H. Cannon and Lenore J. Cowen and Carey E. Priebe. Approximate Distance Classification. Department of Mathematical Sciences The Johns Hopkins University. [View Context].

A. da Valls and Vicen Torra. Explaining the consensus of opinions with the vocabulary of the experts. Dept. d'Enginyeria Informtica i Matemtiques Universitat Rovira i Virgili. [View Context].

Wl/odzisl/aw Duch and Rafal Adamczak and Krzysztof Grabczewski. Extraction of crisp logical rules using constrained backpropagation networks. Department of Computer Methods, Nicholas Copernicus University. [View Context].

Eric P. Kasten and Philip K. McKinley. MESO: Perceptual Memory to Support Online Learning in Adaptive Software. Proceedings of the Third International Conference on Development and Learning (ICDL. [View Context].

Karol Grudzi nski and Wl/odzisl/aw Duch. SBL-PM: A Simple Algorithm for Selection of Reference Instances in Similarity Based Methods. 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].

Alexander K. Seewald. Dissertation Towards Understanding Stacking Studies of a General Ensemble Learning Scheme ausgefuhrt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Naturwissenschaften. [View Context].

Wl odzisl and Rafal Adamczak and Krzysztof Grabczewski and Grzegorz Zal. A hybrid method for extraction of logical rules from data. Department of Computer Methods, Nicholas Copernicus University. [View Context].

Wl/odzisl/aw Duch and Rafal Adamczak and Geerd H. F Diercksen. Classification, Association and Pattern Completion using Neural Similarity Based Methods. Department of Computer Methods, Nicholas Copernicus University. [View Context].

Stefan Aeberhard and Danny Coomans and De Vel. THE PERFORMANCE OF STATISTICAL PATTERN RECOGNITION METHODS IN HIGH DIMENSIONAL SETTINGS. James Cook University. [View Context].

Michael P. Cummings and Daniel S. Myers and Marci Mangelson. Applying Permuation Tests to Tree-Based Statistical Models: Extending the R Package rpart. Center for Bioinformatics and Computational Biology, Institute for Advanced Computer Studies, University of Maryland. [View Context].

Ping Zhong and Masao Fukushima. Second Order Cone Programming Formulations for Robust Multi-class Classification. [View Context].

Wl odzisl/aw Duch and Rafal Adamczak and Norbert Jankowski. Initialization of adaptive parameters in density networks. Department of Computer Methods, Nicholas Copernicus University. [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].

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].

Andrew Watkins and Jon Timmis and Lois C. Boggess. Artificial Immune Recognition System (AIRS): An ImmuneInspired Supervised Learning Algorithm. (abw5,jt6@kent.ac.uk) Computing Laboratory, University of Kent. [View Context].

Gaurav Marwah and Lois C. Boggess. Artificial Immune Systems for Classification : Some Issues. Department of Computer Science Mississippi State University. [View Context].

Igor Kononenko and Edvard Simec. Induction of decision trees using RELIEFF. University of Ljubljana, Faculty of electrical engineering & computer science. [View Context].

Daichi Mochihashi and Gen-ichiro Kikui and Kenji Kita. Learning Nonstructural Distance Metric by Minimum Cluster Distortions. ATR Spoken Language Translation research laboratories. [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