

The data set (and description) can be downloaded here:

<http://archive.ics.uci.edu/ml/machine-learning-databases/00225/>

Description:

ILPD (Indian Liver Patient Dataset) Data Set

Download: Data Folder, Data Set Description

Abstract: This data set contains 10 variables that are age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos.

Data Set Characteristics: Multivariate

Number of Instances: 583

Area: Life

Attribute Characteristics: Integer, Real

Number of Attributes: 10

Date Donated: 2012-05-21

Associated Tasks: Classification

Missing Values? N/A

Number of Web Hits: 9021

Source:

1. Bendi Venkata Ramana

ramana.bendi '@' gmail.com

Associate Professor,

Department of Information Technology,

Aditya Institute of Technology and Management,

Tekkali - 532201, Andhra Pradesh, India.

2. Prof. M. Surendra Prasad Babu

drmsprasadbabu '@' yahoo.co.in

Department of Computer Science & Systems Engineering,

Andhra University College of Engineering,

Visakhapatnam-530 003 Andhra Pradesh, India.

3. Prof. N. B. Venkateswarlu

venkat_ritch '@' yahoo.com

Department of Computer Science and Engineering,

Aditya Institute of Technology and Management,

Tekkali - 532201, Andhra Pradesh, India.

Data Set Information:

This data set contains 416 liver patient records and 167 non liver patient records. The data set was collected from north east of Andhra Pradesh, India.

Selector is a class label used to divide into groups (liver patient or not).

This data set contains 441 male patient records and 142 female patient records.

Attribute Information:

1. Age Age of the patient
2. Gender Gender of the patient
3. TB Total Bilirubin
4. DB Direct Bilirubin
5. Alkphos Alkaline Phosphotase
6. Sgpt Alamine Aminotransferase
7. Sgot Aspartate Aminotransferase
8. TP Total Protiens
9. ALB Albumin
10. A/G Ratio Albumin and Globulin Ratio
11. Selector field used to split the data into two sets (labeled by the experts)

Relevant Papers:

1. Bendi Venkata Ramana, Prof. M. S. Prasad Babu and Prof. N. B. Venkateswarlu, "A Critical Comparative Study of Liver Patients from USA and INDIA: An Exploratory Analysis", International Journal of Computer Science Issues, ISSN :1694-0784, May 2012.

2. Bendi Venkata Ramana, Prof. M. S. Prasad Babu and Prof. N. B. Venkateswarlu, "A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis", International Journal of Database Management Systems (IJDMS), Vol.3, No.2, ISSN : 0975-5705, PP 101-114, May 2011.

Citation Request:

Please refer to the repository <http://archive.ics.uci.edu/ml> (see citation policy).

See also Frank, A. & Asuncion, A. (2010). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml].

Irvine, CA: University of California, School of Information and Computer Science.

Descriptive statistics:

Dataset= indian-liver-patient_MvsF : n= 579 , d= 9

Class1: n= 140

Covariance matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	255.9175	-1.9613	-0.3324	286.2983	-253.6986	-312.0082	-2.9108	-2.7232	-0.9018
[2,]	-1.9613	23.9567	11.6929	687.7127	202.0021	321.0804	-0.0349	-0.7528	-0.3937
[3,]	-0.3324	11.6929	5.8276	345.5201	95.8396	146.9640	-0.0100	-0.3429	-0.1833
[4,]	286.2983	687.7127	345.5201	91291.8485	12676.2857	17484.3017	-3.9100	-35.2282	-20.9885
[5,]	-253.6986	202.0021	95.8396	12676.2857	8947.4189	11469.2132	-9.5502	-13.6050	-5.8988
[6,]	-312.0082	321.0804	146.9640	17484.3017	11469.2132	16529.2243	-18.1825	-24.3777	-9.4784
[7,]	-2.9108	-0.0349	-0.0100	-3.9100	-9.5502	-18.1825	1.2899	0.8195	0.1380
[8,]	-2.7232	-0.7528	-0.3429	-35.2282	-13.6050	-24.3777	0.8195	0.6906	0.1865
[9,]	-0.9018	-0.3937	-0.1833	-20.9885	-5.8988	-9.4784	0.1380	0.1865	0.0830

Correlation matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	1.0000	-0.0250	-0.0086	0.0592	-0.1677	-0.1517	-0.1602	-0.2048	-0.1957
[2,]	-0.0250	1.0000	0.9896	0.4650	0.4363	0.5102	-0.0063	-0.1851	-0.2792
[3,]	-0.0086	0.9896	1.0000	0.4737	0.4197	0.4735	-0.0036	-0.1709	-0.2636
[4,]	0.0592	0.4650	0.4737	1.0000	0.4435	0.4501	-0.0114	-0.1403	-0.2411
[5,]	-0.1677	0.4363	0.4197	0.4435	1.0000	0.9431	-0.0889	-0.1731	-0.2165
[6,]	-0.1517	0.5102	0.4735	0.4501	0.9431	1.0000	-0.1245	-0.2282	-0.2559
[7,]	-0.1602	-0.0063	-0.0036	-0.0114	-0.0889	-0.1245	1.0000	0.8683	0.4217
[8,]	-0.2048	-0.1851	-0.1709	-0.1403	-0.1731	-0.2282	0.8683	1.0000	0.7789
[9,]	-0.1957	-0.2792	-0.2636	-0.2411	-0.2165	-0.2559	0.4217	0.7789	1.0000

Median: 43.793 1.0726 0.3526 197.3145 28.1935 30.8487 6.8497 3.422 0.986

Mean: 43.1786 2.345 1 304.0214 54.7643 69.6857 6.6643 3.2729 0.949

MCD-estimated:

MDC-0.975-Mean: 43.5375 0.7888 0.1963 184.2125 23.5125 25.3125 6.705 3.3475 0.9726

MDC-0.750-Mean: 43.5375 0.7888 0.1963 184.2125 23.5125 25.3125 6.705 3.3475 0.9726

MDC-0.500-Mean: 43.7722 0.7873 0.1949 184.5316 23.5063 25.3165 6.7025 3.3342 0.9634

Class2: n= 439

Covariance matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	264.9568	1.4329	0.1804	328.9621	-281.6045	-55.7124	-3.2772	-3.5382	-1.1931
[2,]	1.4329	43.1816	16.3286	198.7516	246.3372	446.7653	0.0153	-1.1561	-0.4160
[3,]	0.1804	16.3286	8.5159	104.8222	122.8532	221.7668	0.0413	-0.5376	-0.1791
[4,]	328.9621	198.7516	104.8222	49244.7423	3464.3167	10187.4578	-9.1676	-31.2645	-17.4034
[5,]	-281.6045	246.3372	122.8532	3464.3167	41149.1582	51390.7914	-6.0655	0.3071	1.7100
[6,]	-55.7124	446.7653	221.7668	10187.4578	51390.7914	104921.6384	-1.7780	-15.7599	-5.5206
[7,]	-3.2772	0.0153	0.0413	-9.1676	-6.0655	-1.7780	1.1291	0.6201	0.0635
[8,]	-3.5382	-1.1561	-0.5376	-31.2645	0.3071	-15.7599	0.6201	0.6061	0.1718
[9,]	-1.1931	-0.4160	-0.1791	-17.4034	1.7100	-5.5206	0.0635	0.1718	0.1084

Correlation matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	1.0000	0.0134	0.0038	0.0911	-0.0853	-0.0106	-0.1895	-0.2792	-0.2226
[2,]	0.0134	1.0000	0.8515	0.1363	0.1848	0.2099	0.0022	-0.2260	-0.1922
[3,]	0.0038	0.8515	1.0000	0.1619	0.2075	0.2346	0.0133	-0.2366	-0.1864
[4,]	0.0911	0.1363	0.1619	1.0000	0.0770	0.1417	-0.0389	-0.1810	-0.2381
[5,]	-0.0853	0.1848	0.2075	0.0770	1.0000	0.7821	-0.0281	0.0019	0.0256
[6,]	-0.0106	0.2099	0.2346	0.1417	0.7821	1.0000	-0.0052	-0.0625	-0.0518
[7,]	-0.1895	0.0022	0.0133	-0.0389	-0.0281	-0.0052	1.0000	0.7496	0.1815
[8,]	-0.2792	-0.2260	-0.2366	-0.1810	0.0019	-0.0625	0.7496	1.0000	0.6700
[9,]	-0.2226	-0.1922	-0.1864	-0.2381	0.0256	-0.0518	0.1815	0.6700	1.0000

Median: 45.7294 2.2862 1.006 217.5623 42.7393 53.4339 6.4097 3.1647 0.9788

Mean: 45.2938 3.6248 1.6517 287.3303 89.533 123.4032 6.4235 3.0957 0.9464

MCD-estimated:

MDC-0.975-Mean: 44.9456 1.1075 0.387 202.8912 32.9331 36.523 6.531 3.3067 1.0134

MDC-0.750-Mean: 44.7573 1.1029 0.3854 201.9833 33.0795 36.5607 6.5301 3.3033 1.0117

MDC-0.500-Mean: 45.0126 1.0975 0.3798 201.9034 33.2941 36.4916 6.5315 3.3109 1.0155

Measures:

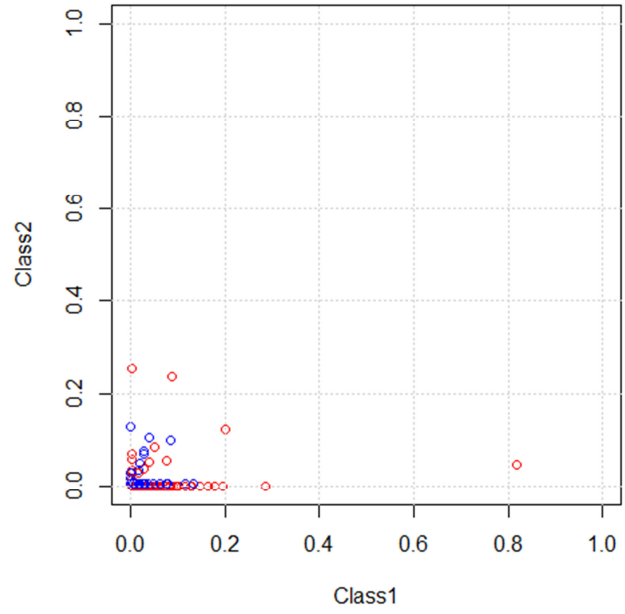
Mah.Dist: 0.4299

Mah.Dist-MCD-0.975: 0.8832

Mah.Dist-MCD-0.750: 0.8924

Mah.Dist-MCD-0.500: 0.8864

DD-Plot (zonoid): indian-liver-patient_1vs2



DD-Plot (random Tukey): indian-liver-patient_1vs2

