

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

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

Description:

1. Title of Database: wine recognition data

Updated Sept 21, 1998 by C.Blake : Added attribute information

2. Sources:

(a) Forina, M. et al, PARVUS - An Extendible Package for Data Exploration, Classification and Correlation. Institute of Pharmaceutical and Food Analysis and Technologies, Via Brigata Salerno, 16147 Genoa, Italy.

(b) Stefan Aeberhard, email: stefan@coral.cs.jcu.edu.au

(c) July 1991

3. Past Usage:

(1)

S. Aeberhard, D. Coomans and O. de Vel,
Comparison of Classifiers in High Dimensional Settings,
Tech. Rep. no. 92-02, (1992), Dept. of Computer Science and Dept. of
Mathematics and Statistics, James Cook University of North Queensland.
(Also submitted to Technometrics).

The data was used with many others for comparing various classifiers. The classes are separable, though only RDA has achieved 100% correct classification.

(RDA : 100%, QDA 99.4%, LDA 98.9%, 1NN 96.1% (z-transformed data))

(All results using the leave-one-out technique)

In a classification context, this is a well posed problem with "well behaved" class structures. A good data set for first testing of a new classifier, but not very challenging.

(2)

S. Aeberhard, D. Coomans and O. de Vel,
"THE CLASSIFICATION PERFORMANCE OF RDA"
Tech. Rep. no. 92-01, (1992), Dept. of Computer Science and Dept. of
Mathematics and Statistics, James Cook University of North Queensland.
(Also submitted to Journal of Chemometrics).

Here, the data was used to illustrate the superior performance of
the use of a new appreciation function with RDA.

4. Relevant Information:

-- These data are the results of a chemical analysis of
wines grown in the same region in Italy but derived from three
different cultivars.

The analysis determined the quantities of 13 constituents
found in each of the three types of wines.

-- I think that the initial data set had around 30 variables, but
for some reason I only have the 13 dimensional version.
I had a list of what the 30 or so variables were, but a.)
I lost it, and b.), I would not know which 13 variables
are included in the set.

-- The attributes are (donated by Riccardo Leardi, riclea@anchem.unige.it)

- 1) Alcohol
- 2) Malic acid
- 3) Ash
- 4) Alkalinity of ash
- 5) Magnesium
- 6) Total phenols
- 7) Flavanoids
- 8) Nonflavanoid phenols
- 9) Proanthocyanins
- 10) Color intensity
- 11) Hue
- 12) OD280/OD315 of diluted wines
- 13) Proline

5. Number of Instances

```
class 1 59  
class 2 71  
class 3 48
```

6. Number of Attributes

13

7. For Each Attribute:

All attributes are continuous

No statistics available, but suggest to standardise variables for certain uses (e.g. for us with classifiers which are NOT scale invariant)

NOTE: 1st attribute is class identifier (1-3)

8. Missing Attribute Values:

None

9. Class Distribution: number of instances per class

```
class 1 59  
class 2 71  
class 3 48
```

Descriptive statistics:

Dataset= wine_1vs3 : n= 107 , d= 13

Class1: n= 59

Covariance matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
[1,]	0.2136	-0.0129	-0.0156	-0.3746	0.7732	0.0659	0.0762	0.0005	0.0586	0.2337	0.0043	0.0115	36.9195
[2,]	-0.0129	0.4741	0.0041	0.1053	0.5734	-0.0195	-0.0524	-0.0043	-0.0229	-0.2197	-0.0337	0.0426	-56.8364
[3,]	-0.0156	0.0041	0.0516	0.3178	0.9124	0.0004	-0.0064	0.0074	-0.0136	-0.0350	0.0063	-0.0066	-1.4858
[4,]	-0.3746	0.1053	0.3178	6.4838	6.3716	-0.1925	-0.2906	0.0539	-0.1822	-0.6653	0.0276	-0.1070	-69.0615
[5,]	0.7732	0.5734	0.9124	6.3716	110.2279	1.0934	0.5147	0.1745	-0.2555	2.4013	-0.1362	0.4523	-344.0041
[6,]	0.0659	-0.0195	0.0004	-0.1925	1.0934	0.1149	0.1083	-0.0004	0.0522	0.2729	-0.0089	0.0064	22.1502
[7,]	0.0762	-0.0524	-0.0064	-0.2906	0.5147	0.1083	0.1580	-0.0025	0.0899	0.3651	0.0004	-0.0126	33.4995
[8,]	0.0005	-0.0043	0.0074	0.0539	0.1745	-0.0004	-0.0025	0.0049	-0.0042	-0.0132	0.0034	-0.0081	-0.2379
[9,]	0.0586	-0.0229	-0.0136	-0.1822	-0.2555	0.0522	0.0899	-0.0042	0.1698	0.2168	0.0050	0.0005	12.9784
[10,]	0.2337	-0.2197	-0.0350	-0.6653	2.4013	0.2729	0.3651	-0.0132	0.2168	1.5341	0.0041	-0.0827	161.5407
[11,]	0.0043	-0.0337	0.0063	0.0276	-0.1362	-0.0089	0.0004	0.0034	0.0050	0.0041	0.0136	-0.0129	9.1180
[12,]	0.0115	0.0426	-0.0066	-0.1070	0.4523	0.0064	-0.0126	-0.0081	0.0005	-0.0827	-0.0129	0.1275	-27.5401
[13,]	36.9195	-56.8364	-1.4858	-69.0615	-344.0041	22.1502	33.4995	-0.2379	12.9784	161.5407	9.1180	-27.5401	49071.4500

Correlation matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
[1,]	1.0000	-0.0405	-0.1486	-0.3184	0.1594	0.4207	0.4149	0.0157	0.3076	0.4083	0.0800	0.0698	0.3606
[2,]	-0.0405	1.0000	0.0262	0.0600	0.0793	-0.0835	-0.1913	-0.0894	-0.0808	-0.2576	-0.4200	0.1732	-0.3726
[3,]	-0.1486	0.0262	1.0000	0.5493	0.3825	0.0048	-0.0705	0.4659	-0.1455	-0.1242	0.2392	-0.0816	-0.0295
[4,]	-0.3184	0.0600	0.5493	1.0000	0.2383	-0.2230	-0.2871	0.3023	-0.1736	-0.2110	0.0930	-0.1177	-0.1224
[5,]	0.1594	0.0793	0.3825	0.2383	1.0000	0.3072	0.1233	0.2372	-0.0590	0.1847	-0.1114	0.1207	-0.1479
[6,]	0.4207	-0.0835	0.0048	-0.2230	0.3072	1.0000	0.8038	-0.0170	0.3736	0.6501	-0.2243	0.0532	0.2950
[7,]	0.4149	-0.1913	-0.0705	-0.2871	0.1233	0.8038	1.0000	-0.0895	0.5486	0.7416	0.0079	-0.0885	0.3804
[8,]	0.0157	-0.0894	0.4659	0.3023	0.2372	-0.0170	-0.0895	1.0000	-0.1445	-0.1525	0.4118	-0.3235	-0.0153
[9,]	0.3076	-0.0808	-0.1455	-0.1736	-0.0590	0.3736	0.5486	-0.1445	1.0000	0.4247	0.1039	0.0031	0.1422
[10,]	0.4083	-0.2576	-0.1242	-0.2110	0.1847	0.6501	0.7416	-0.1525	0.4247	1.0000	0.0282	-0.1869	0.5888
[11,]	0.0800	-0.4200	0.2392	0.0930	-0.1114	-0.2243	0.0079	0.4118	0.1039	0.0282	1.0000	-0.3107	0.3534
[12,]	0.0698	0.1732	-0.0816	-0.1177	0.1207	0.0532	-0.0885	-0.3235	0.0031	-0.1869	-0.3107	1.0000	-0.3482
[13,]	0.3606	-0.3726	-0.0295	-0.1224	-0.1479	0.2950	0.3804	-0.0153	0.1422	0.5888	0.3534	-0.3482	1.0000

Median: 13.7453 2.0431 2.4774 17.0148 105.3519 2.8009 2.9508 0.2665 1.77 5.1854 1.0248 3.1632 1092.672

Mean: 13.7447 2.0107 2.4556 17.0373 106.339 2.8402 2.9824 0.29 1.8993 5.5283 1.062 3.1578 1115.712

MCD-estimated:

MDC-0.975-Mean: 13.7227 1.738 2.4541 17.0805 106.3171 2.8412 2.9763 0.2861 1.8193 5.5912 1.0661 3.1676 1148.781

MDC-0.750-Mean: 13.7428 1.7492 2.439 16.9744 105.5385 2.8221 2.9269 0.2821 1.8418 5.3626 1.0538 3.1821 1123.718

MDC-0.500-Mean: 13.7372 1.7428 2.4607 17.1977 106.4884 2.8447 2.9588 0.2905 1.8221 5.5102 1.0614 3.157 1129.186

Class2: n= 48

Covariance matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
[1,]	0.2812	0.0637	0.0240	0.2514	-0.4859	0.0398	0.0118	0.0025	0.0816	0.4293	-0.0021	0.0191	-5.4347
[2,]	0.0637	1.1835	0.0036	0.2089	-2.0731	-0.0624	-0.0899	0.0193	-0.0994	-0.4078	0.0099	0.0021	0.3642
[3,]	0.0240	0.0036	0.0341	0.3163	0.4250	0.0310	0.0150	-0.0005	0.0146	0.0534	0.0038	0.0113	-3.1312
[4,]	0.2514	0.2089	0.3163	5.0993	3.9202	0.2938	0.1799	-0.0048	0.2432	0.8382	0.0071	0.0256	-25.6472
[5,]	-0.4859	-2.0731	0.4250	3.9202	118.6024	-0.1541	1.8180	-0.6837	0.6836	2.6223	0.0057	-0.6609	238.0120
[6,]	0.0398	-0.0624	0.0310	0.2938	-0.1541	0.1274	0.0250	0.0145	0.0905	0.2771	-0.0011	0.0195	1.7637
[7,]	0.0118	-0.0899	0.0150	0.1799	1.8180	0.0250	0.0861	-0.0231	0.0490	0.2489	-0.0098	-0.0343	-8.3254
[8,]	0.0025	0.0193	-0.0005	-0.0048	-0.6837	0.0145	-0.0231	0.0154	0.0087	0.0075	0.0022	0.0104	2.8891
[9,]	0.0816	-0.0994	0.0146	0.2432	0.6836	0.0905	0.0490	0.0087	0.1671	0.6471	-0.0197	-0.0143	9.1982
[10,]	0.4293	-0.4078	0.0534	0.8382	2.6223	0.2771	0.2489	0.0075	0.6471	5.3405	-0.1504	-0.0648	32.0230
[11,]	-0.0021	0.0099	0.0038	0.0071	0.0057	-0.0011	-0.0098	0.0022	-0.0197	-0.1504	0.0131	0.0113	0.0311
[12,]	0.0191	0.0021	0.0113	0.0256	-0.6609	0.0195	-0.0343	0.0104	-0.0143	-0.0648	0.0113	0.0740	6.1259
[13,]	-5.4347	0.3642	-3.1312	-25.6472	238.0120	1.7637	-8.3254	2.8891	9.1982	32.0230	0.0311	6.1259	13247.3293

Correlation matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
[1,]	1.0000	0.1105	0.2451	0.2099	-0.0841	0.2104	0.0761	0.0382	0.3766	0.3504	-0.0341	0.1322	-0.0891
[2,]	0.1105	1.0000	0.0181	0.0850	-0.1750	-0.1607	-0.2816	0.1428	-0.2234	-0.1622	0.0797	0.0072	0.0029
[3,]	0.2451	0.0181	1.0000	0.7585	0.2113	0.4704	0.2771	-0.0204	0.1938	0.1252	0.1805	0.2254	-0.1473
[4,]	0.2099	0.0850	0.7585	1.0000	0.1594	0.3645	0.2714	-0.0171	0.2634	0.1606	0.0276	0.0417	-0.0987
[5,]	-0.0841	-0.1750	0.2113	0.1594	1.0000	-0.0396	0.5688	-0.5057	0.1535	0.1042	0.0046	-0.2230	0.1899
[6,]	0.2104	-0.1607	0.4704	0.3645	-0.0396	1.0000	0.2389	0.3263	0.6200	0.3359	-0.0261	0.2004	0.0429
[7,]	0.0761	-0.2816	0.2771	0.2714	0.5688	0.2389	1.0000	-0.6345	0.4080	0.3669	-0.2931	-0.4297	-0.2464
[8,]	0.0382	0.1428	-0.0204	-0.0171	-0.5057	0.3263	-0.6345	1.0000	0.1718	0.0263	0.1531	0.3066	0.2022
[9,]	0.3766	-0.2234	0.1938	0.2634	0.1535	0.6200	0.4080	0.1718	1.0000	0.6849	-0.4218	-0.1286	0.1955
[10,]	0.3504	-0.1622	0.1252	0.1606	0.1042	0.3359	0.3669	0.0263	0.6849	1.0000	-0.5686	-0.1030	0.1204
[11,]	-0.0341	0.0797	0.1805	0.0276	0.0046	-0.0261	-0.2931	0.1531	-0.4218	-0.5686	1.0000	0.3643	0.0024
[12,]	0.1322	0.0072	0.2254	0.0417	-0.2230	0.2004	-0.4297	0.3066	-0.1286	-0.1030	0.3643	1.0000	0.1956
[13,]	-0.0891	0.0029	-0.1473	-0.0987	0.1899	0.0429	-0.2464	0.2022	0.1955	0.1204	0.0024	0.1956	1.0000

Median: 13.187 3.0508 2.4547 21.4506 97.5229 1.7777 0.8029 0.448 1.2204 7.7863 0.6726 1.7161 625.3559

Mean: 13.1538 3.3338 2.4371 21.4167 99.3125 1.6788 0.7815 0.4475 1.1535 7.3962 0.6827 1.6835 629.8958

MCD-estimated:

MDC-0.975-Mean: 13.2191 3.603 2.4088 21.1515 97.1212 1.6282 0.6506 0.4824 1.1158 7.4045 0.6864 1.753 648.3333

MDC-0.750-Mean: 13.2 3.2744 2.4332 21.25 100.8529 1.6526 0.8224 0.4365 1.1447 7.5238 0.6759 1.6376 621.9118

MDC-0.500-Mean: 13.1945 3.3903 2.4242 21.303 100.8788 1.6648 0.8345 0.4358 1.1585 7.6561 0.6748 1.5973 619.2424

Measures:

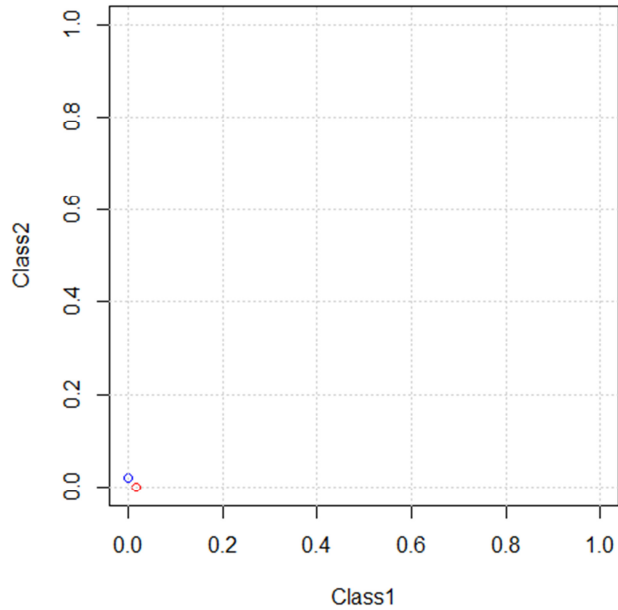
Mah. Dist: 10.78

Mah. Dist-MCD-0.975: 10.2546

Mah. Dist-MCD-0.750: 10.9787

Mah. Dist-MCD-0.500: 11.3133

DD-Plot (zonoid): wine_1vs3



DD-Plot (random Tukey) for wine_1vs3

