

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) Alcalinity 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_1vs2 : n= 130 , 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: 13.7255 1.7568 2.4585 17.175 106 2.8172 2.9208 0.2868 1.8438 5.318 1.0557 3.1825 1116.375

MDC-0.750-Mean: 13.7478 1.7351 2.4393 17.0244 105.5854 2.829 2.9439 0.2834 1.839 5.438 1.0546 3.1761 1121.951

MDC-0.500-Mean: 13.7569 1.7387 2.4254 16.7846 105.5641 2.8362 2.9713 0.291 1.8049 5.6015 1.0644 3.1656 1128.205

Class2: n= 71

Covariance matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
[1,]	0.2894	-0.0117	-0.0365	-0.1014	-0.2696	-0.0136	-0.0145	-0.0045	-0.0614	0.1342	-0.0002	-0.0348	3.6514
[2,]	-0.0117	1.0314	0.0476	0.8094	-1.3065	0.0218	0.0802	0.0161	0.1287	-0.1909	-0.0841	0.0796	-35.7978
[3,]	-0.0365	0.0476	0.0995	0.7347	0.6825	0.0193	0.0701	0.0117	0.0082	0.0176	-0.0020	0.0252	2.0810
[4,]	-0.1014	0.8094	0.7347	11.2210	0.1831	0.2337	0.7360	0.0758	0.2195	-0.2660	-0.0522	0.6356	-7.6396
[5,]	-0.2696	-1.3065	0.6825	0.1831	280.6797	0.6403	0.0200	-0.4032	3.0037	0.6807	0.4244	-0.6338	1315.8461
[6,]	-0.0136	0.0218	0.0193	0.2337	0.6403	0.2974	0.2967	-0.0287	0.1256	0.0853	0.0044	0.1313	1.4513
[7,]	-0.0145	0.0802	0.0701	0.7360	0.0200	0.2967	0.4980	-0.0206	0.2121	0.2471	-0.0042	0.2031	-13.6026
[8,]	-0.0045	0.0161	0.0117	0.0758	-0.4032	-0.0287	-0.0206	0.0154	-0.0240	0.0021	-0.0008	-0.0254	-2.9756
[9,]	-0.0614	0.1287	0.0082	0.2195	3.0037	0.1256	0.2121	-0.0240	0.3625	-0.0411	-0.0066	0.1153	11.7763
[10,]	0.1342	-0.1909	0.0176	-0.2660	0.6807	0.0853	0.2471	0.0021	-0.0411	0.8555	-0.0049	-0.0538	14.8850
[11,]	-0.0002	-0.0841	-0.0020	-0.0522	0.4244	0.0044	-0.0042	-0.0008	-0.0066	-0.0049	0.0412	-0.0053	3.6517
[12,]	-0.0348	0.0796	0.0252	0.6356	-0.6338	0.1313	0.2031	-0.0254	0.1153	-0.0538	-0.0053	0.2466	-8.6508
[13,]	3.6514	-35.7978	2.0810	-7.6396	1315.8461	1.4513	-13.6026	-2.9756	11.7763	14.8850	3.6517	-8.6508	24715.3678

Correlation matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
[1,]	1.0000	-0.0214	-0.2149	-0.0563	-0.0299	-0.0463	-0.0382	-0.0682	-0.1896	0.2698	-0.0020	-0.1303	0.0432
[2,]	-0.0214	1.0000	0.1487	0.2379	-0.0768	0.0394	0.1119	0.1276	0.2105	-0.2033	-0.4080	0.1578	-0.2242
[3,]	-0.2149	0.1487	1.0000	0.6953	0.1291	0.1121	0.3149	0.2998	0.0430	0.0602	-0.0312	0.1606	0.0420
[4,]	-0.0563	0.2379	0.6953	1.0000	0.0033	0.1279	0.3114	0.1826	0.1088	-0.0859	-0.0768	0.3821	-0.0145
[5,]	-0.0299	-0.0768	0.1291	0.0033	1.0000	0.0701	0.0017	-0.1941	0.2978	0.0439	0.1248	-0.0762	0.4996
[6,]	-0.0463	0.0394	0.1121	0.1279	0.0701	1.0000	0.7710	-0.4247	0.3826	0.1691	0.0397	0.4847	0.0169
[7,]	-0.0382	0.1119	0.3149	0.3114	0.0017	0.7710	1.0000	-0.2353	0.4993	0.3786	-0.0294	0.5796	-0.1226
[8,]	-0.0682	0.1276	0.2998	0.1826	-0.1941	-0.4247	-0.2353	1.0000	-0.3216	0.0185	-0.0337	-0.4131	-0.1527
[9,]	-0.1896	0.2105	0.0430	0.1088	0.2978	0.3826	0.4993	-0.3216	1.0000	-0.0738	-0.0544	0.3858	0.1244
[10,]	0.2698	-0.2033	0.0602	-0.0859	0.0439	0.1691	0.3786	0.0185	-0.0738	1.0000	-0.0261	-0.1171	0.1024
[11,]	-0.0020	-0.4080	-0.0312	-0.0768	0.1248	0.0397	-0.0294	-0.0337	-0.0544	-0.0261	1.0000	-0.0524	0.1145
[12,]	-0.1303	0.1578	0.1606	0.3821	-0.0762	0.4847	0.5796	-0.4131	0.3858	-0.1171	-0.0524	1.0000	-0.1108
[13,]	0.0432	-0.2242	0.0420	-0.0145	0.4996	0.0169	-0.1226	-0.1527	0.1244	0.1024	0.1145	-0.1108	1.0000

Median: 12.3121 1.8089 2.2183 20.7242 88.7897 1.9897 1.9028 0.3976 1.5782 3.0512 1.0454 2.6943 490.7916

Mean: 12.2787 1.9327 2.2448 20.238 94.5493 2.2589 2.0808 0.3637 1.6303 3.0866 1.0563 2.7854 519.507

MCD-estimated:

MDC-0.975-Mean: 12.2268 1.657 2.2326 20.2128 89.3404 2.196 2.0862 0.366 1.5362 3.076 1.0596 2.8443 491.8936

MDC-0.750-Mean: 12.1954 1.6659 2.2393 20.2609 89.2391 2.198 2.0926 0.367 1.5537 3.0602 1.0559 2.8526 488.8913

MDC-0.500-Mean: 12.2135 1.7979 2.2294 20.2083 88.7292 2.1894 2.0687 0.37 1.5269 3.039 1.0504 2.836 483.5625

Measures:

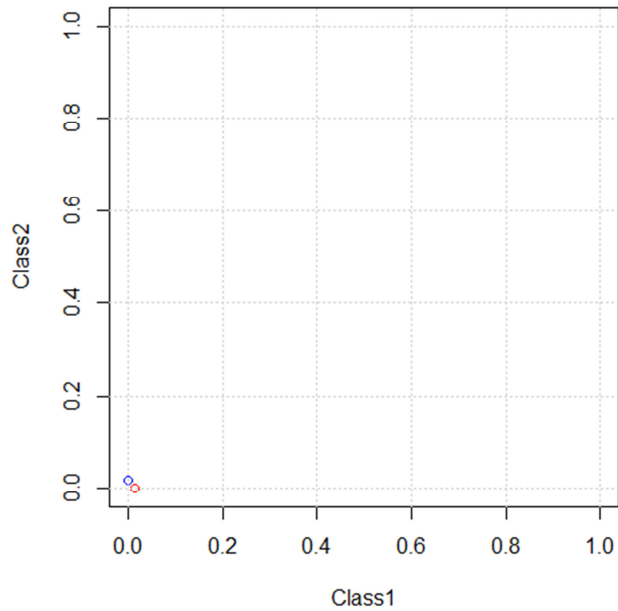
Mah. Dist: 4.9816

Mah. Dist-MCD-0.975: 4.9854

Mah. Dist-MCD-0.750: 4.9469

Mah. Dist-MCD-0.500: 5.3041

DD-Plot (zonoid): wine_1vs2



DD-Plot (random Tukey) for wine_1vs2

