

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_2vs3 : n= 119 , d= 13

Class1: 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: 12.2455 1.7611 2.2181 20.2553 88.9149 2.1877 2.0621 0.3694 1.5043 3.0557 1.0636 2.8343 484.8723

MDC-0.750-Mean: 12.2232 1.8191 2.2245 20.3085 88.6809 2.2098 2.0757 0.3602 1.5432 3.0096 1.0479 2.8394 482

MDC-0.500-Mean: 12.2226 1.6376 2.2317 20.1739 89.3261 2.2072 2.0915 0.3596 1.5387 3.0841 1.0639 2.8343 495.7391

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

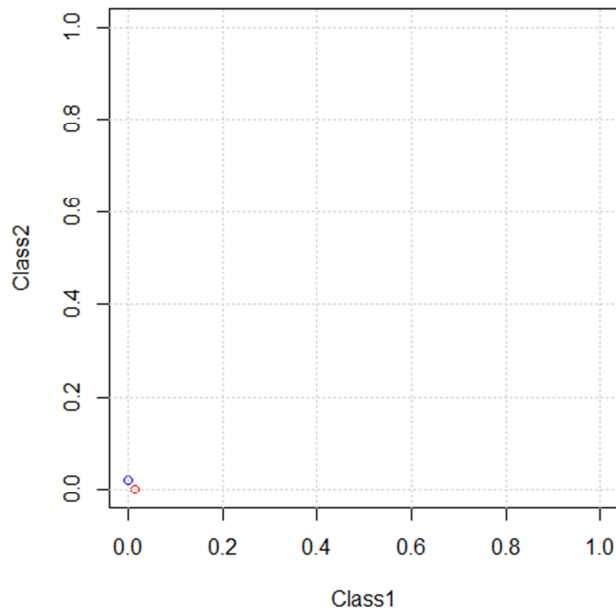
Correlation matrix:

	[,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:	13.2423	3.559	2.4335	21.2742	97.2903	1.6545	0.6816	0.4742	1.2032	7.7955	0.6632	1.7413	647.2581
MDC-0.750-Mean:	13.1514	3.2863	2.4311	21.3286	100.7429	1.6531	0.8006	0.446	1.1263	7.5031	0.6906	1.6626	637.8571
MDC-0.500-Mean:	13.2097	3.4285	2.4439	21.4848	100	1.6703	0.7639	0.4518	1.1221	7.417	0.6936	1.6903	647.7273

Measures:
 Mah. Dist: 5.2863
 Mah. Dist-MCD-0.975: 6.0036
 Mah. Dist-MCD-0.750: 6.3792
 Mah. Dist-MCD-0.500: 6.4272

DD-Plot (zonoid): wine_2vs3



DD-Plot (random Tukey): wine_2vs3

