**107.Wine**

1. 数据库网址

2. 数据库描述

【1.[数据集名称]数据集由[机构名或人名]采集；】The data used in our experiments were collected by E. Alpaydin, C. Kaynak, from Department of Computer Engineering,Bogazici University at July,1998.【2.用于[什么实验目的]】We used preprocessing programs made available by NIST to extract normalized bitmaps of handwritten digits from a preprinted form.【3】

【4】The database has 5620 samples, respectively belong to optdigits.tra with 3823 samples and optidigits.tes with 1797 samples. The categories of network system include seven categories, as shown in Table 1.

Table 1 Category Distribution of Network System [根据数据库绘制]

|  |  |  |  |
| --- | --- | --- | --- |
| Invasion Categories | optdigits.tra | optdigits.tes | Total Number of Samples |
|  |  |  |  |
|  |  |  |  |
| Total number of samples in total |  |  | 178 |

|  |  |
| --- | --- |
| **Abstract**: Using chemical analysis determine the origin of wines |  |

**Source:**

Original Owners:   
  
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. 

**Data Set 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 (dontated 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   
  
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.

**Attribute Information:**

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)   
  
  
 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.

-- The attributes are (dontated 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