Wine.data

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.

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

1st attribute is class identifier (1-3)

Dim1 : 找到两个数据点之间的边（k邻域之间的相互关系）。

Dim2：k近邻的平均距离。

KNN算法经过归一化之后的数据呈现为：

data[0][0] data[1][0] data[2][0] data[3][0] data[4][0] data[5][0]

data[0][1] data[0][1] data[0][1] data[0][1] data[0][1] data[0][1]

data[0][1] data[0][1] data[0][1] data[0][1] data[0][1] data[0][1]

data[0][1] data[0][1] data[0][1] data[0][1] data[0][1] data[0][1]

data[0][1] data[0][1] data[0][1] data[0][1] data[0][1] data[0][1]

data[0][1] data[0][1] data[0][1] data[0][1] data[0][1] data[0][1]

housing Data Set

1. CRIM: per capita crime rate by town

2. ZN: proportion of residential land zoned for lots over 25,000 sq.ft.

3. INDUS: proportion of non-retail business acres per town

4. CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)

5. NOX: nitric oxides concentration (parts per 10 million)

6. RM: average number of rooms per dwelling

7. AGE: proportion of owner-occupied units built prior to 1940

8. DIS: weighted distances to five Boston employment centres

9. RAD: index of accessibility to radial highways

10. TAX: full-value property-tax rate per $10,000

11. PTRATIO: pupil-teacher ratio by town

12. B: 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town

13. LSTAT: % lower status of the population

14. MEDV: Median value of owner-occupied homes in $1000's

Concerns housing values in suburbs of Boston.

**Arrhythmia Data Set**   
*Download*: [Data Folder](https://archive.ics.uci.edu/ml/machine-learning-databases/arrhythmia/), [Data Set Description](https://archive.ics.uci.edu/ml/machine-learning-databases/arrhythmia/arrhythmia.names)

**Abstract**: Distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups.

Data Set Information:

This database contains 279 attributes, 206 of which are linear valued and the rest are nominal.

Concerning the study of H. Altay Guvenir: "The aim is to distinguish between the presence and absence of cardiac arrhythmia and to classify it in one of the 16 groups. Class 01 refers to 'normal' ECG classes 02 to 15 refers to different classes of arrhythmia and class 16 refers to the rest of unclassified ones. For the time being, there exists a computer program that makes such a classification. However there are differences between the cardiolog's and the programs classification. Taking the cardiolog's as a gold standard we aim to minimise this difference by means of machine learning tools."

The names and id numbers of the patients were recently removed from the database.

Attribute Information:

-- Complete attribute documentation:

1 Age: Age in years , linear

2 Sex: Sex (0 = male; 1 = female) , nominal

3 Height: Height in centimeters , linear

4 Weight: Weight in kilograms , linear

5 QRS duration: Average of QRS duration in msec., linear

6 P-R interval: Average duration between onset of P and Q waves in msec., linear

7 Q-T interval: Average duration between onset of Q and offset of T waves in msec., linear

8 T interval: Average duration of T wave in msec., linear

9 P interval: Average duration of P wave in msec., linear

Vector angles in degrees on front plane of:, linear

10 QRS

11 T

12 P

13 QRST

14 J

15 Heart rate: Number of heart beats per minute ,linear

Of channel DI:

Average width, in msec., of: linear

16 Q wave

17 R wave

18 S wave

19 R' wave, small peak just after R

20 S' wave

21 Number of intrinsic deflections, linear

22 Existence of ragged R wave, nominal

23 Existence of diphasic derivation of R wave, nominal

24 Existence of ragged P wave, nominal

25 Existence of diphasic derivation of P wave, nominal

26 Existence of ragged T wave, nominal

27 Existence of diphasic derivation of T wave, nominal

Of channel DII:

28 .. 39 (similar to 16 .. 27 of channel DI)

Of channels DIII:

40 .. 51

Of channel AVR:

52 .. 63

Of channel AVL:

64 .. 75

Of channel AVF:

76 .. 87

Of channel V1:

88 .. 99

Of channel V2:

100 .. 111

Of channel V3:

112 .. 123

Of channel V4:

124 .. 135

Of channel V5:

136 .. 147

Of channel V6:

148 .. 159

Of channel DI:

Amplitude , \* 0.1 milivolt, of

160 JJ wave, linear

161 Q wave, linear

162 R wave, linear

163 S wave, linear

164 R' wave, linear

165 S' wave, linear

166 P wave, linear

167 T wave, linear

168 QRSA , Sum of areas of all segments divided by 10, ( Area= width \* height / 2 ), linear

169 QRSTA = QRSA + 0.5 \* width of T wave \* 0.1 \* height of T wave. (If T is diphasic then the bigger segment is considered), linear

Of channel DII:

170 .. 179

Of channel DIII:

180 .. 189

Of channel AVR:

190 .. 199

Of channel AVL:

200 .. 209

Of channel AVF:

210 .. 219

Of channel V1:

220 .. 229

Of channel V2:

230 .. 239

Of channel V3:

240 .. 249

Of channel V4:

250 .. 259

Of channel V5:

260 .. 269

Of channel V6:

270 .. 279

**Glass Identification Data Set**   
*Download*: [Data Folder](https://archive.ics.uci.edu/ml/machine-learning-databases/glass/), [Data Set Description](https://archive.ics.uci.edu/ml/machine-learning-databases/glass/glass.names)

**Abstract**: From USA Forensic Science Service; 6 types of glass; defined in terms of their oxide content (i.e. Na, Fe, K, etc)

Source:

Creator:

B. German

Central Research Establishment

Home Office Forensic Science Service

Aldermaston, Reading, Berkshire RG7 4PN

Donor:

Vina Spiehler, Ph.D., DABFT

Diagnostic Products Corporation

(213) 776-0180 (ext 3014)

Data Set Information:

Vina conducted a comparison test of her rule-based system, BEAGLE, the nearest-neighbor algorithm, and discriminant analysis. BEAGLE is a product available through VRS Consulting, Inc.; 4676 Admiralty Way, Suite 206; Marina Del Ray, CA 90292 (213) 827-7890 and FAX: -3189. In determining whether the glass was a type of "float" glass or not, the following results were obtained (# incorrect answers):

Type of Sample -- Beagle -- NN -- DA

Windows that were float processed (87) -- 10 -- 12 -- 21

Windows that were not: (76) -- 19 -- 16 -- 22

The study of classification of types of glass was motivated by criminological investigation. At the scene of the crime, the glass left can be used as evidence...if it is correctly identified!

Attribute Information:

1. Id number: 1 to 214

2. RI: refractive index

3. Na: Sodium (unit measurement: weight percent in corresponding oxide, as are attributes 4-10)

4. Mg: Magnesium

5. Al: Aluminum

6. Si: Silicon

7. K: Potassium

8. Ca: Calcium

9. Ba: Barium

10. Fe: Iron

11. Type of glass: (class attribute)

-- 1 building\_windows\_float\_processed

-- 2 building\_windows\_non\_float\_processed

-- 3 vehicle\_windows\_float\_processed

-- 4 vehicle\_windows\_non\_float\_processed (none in this database)

-- 5 containers

-- 6 tableware

-- 7 headlamps

**Ionosphere Data Set**   
*Download*: [Data Folder](https://archive.ics.uci.edu/ml/machine-learning-databases/ionosphere/), [Data Set Description](https://archive.ics.uci.edu/ml/machine-learning-databases/ionosphere/ionosphere.names)

**Abstract**: Classification of radar returns from the ionosphere

Data Set Information:

This radar data was collected by a system in Goose Bay, Labrador. This system consists of a phased array of 16 high-frequency antennas with a total transmitted power on the order of 6.4 kilowatts. See the paper for more details. The targets were free electrons in the ionosphere. "Good" radar returns are those showing evidence of some type of structure in the ionosphere. "Bad" returns are those that do not; their signals pass through the ionosphere.

Received signals were processed using an autocorrelation function whose arguments are the time of a pulse and the pulse number. There were 17 pulse numbers for the Goose Bay system. Instances in this databse are described by 2 attributes per pulse number, corresponding to the complex values returned by the function resulting from the complex electromagnetic signal.

Attribute Information:

-- All 34 are continuous

-- The 35th attribute is either "good" or "bad" according to the definition summarized above. This is a binary classification task.