数据集：<http://archive.ics.uci.edu/ml/datasets/Wine+Quality>

简介：数据集包括两份数据，分别是来自葡萄牙北部的红、白青葡萄酒的样本。目的是根据物理特性对葡萄酒的质量进行建模。此样本数据，仅包含物理化学特性以及人工评估质量信息，不包含葡萄的类型、酒的品牌、售价等信息。

注意点：

（1）类别不均衡，普通质量的葡萄酒的数量远远多于极好和极差的葡萄酒的数量。

（2）所给的11个特征不完全是无关的。

备注：

输入特征（物理化学等客观特征）：

1 - fixed acidity（非挥发性酸）

2 - volatile acidity（挥发性酸度）

3 - citric acid（柠檬酸）

4 - residual sugar（残糖）

5 - chlorides（氯化物）

6 - free sulfur dioxide（游离二氧化硫）

7 - total sulfur dioxide（总二氧化硫量）

8 - density（稠密）

9 - pH

10 - sulphates（硫酸盐）

11 - alcohol（酒精）

输出变量（人工评估数据）：

12 - quality (分数在0~10之间)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 4898 | **Area:** | Business |
| **Attribute Characteristics:** | Real | **Number of Attributes:** | 12 | **Date Donated** | 2009-10-07 |
| **Associated Tasks:** | Classification, Regression | **Missing Values?** | N/A | **Number of Web Hits:** | 358104 |