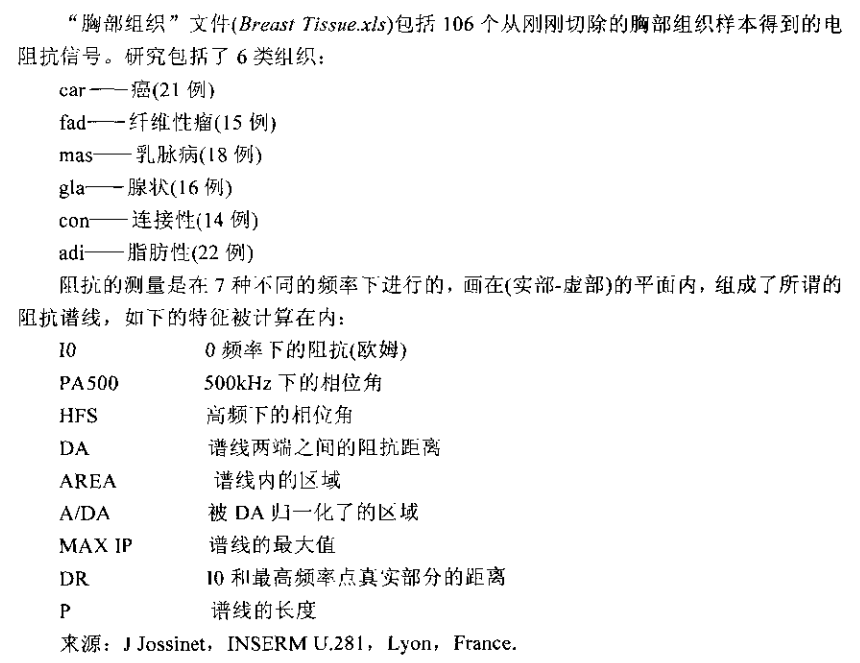
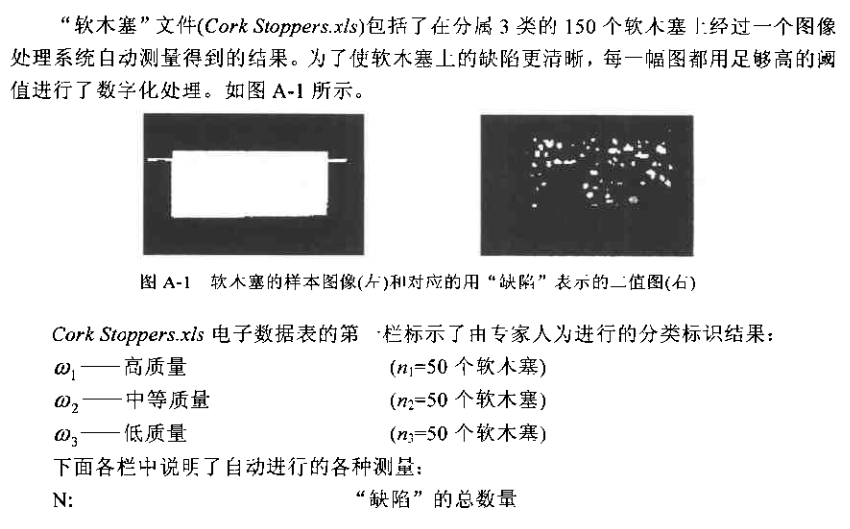
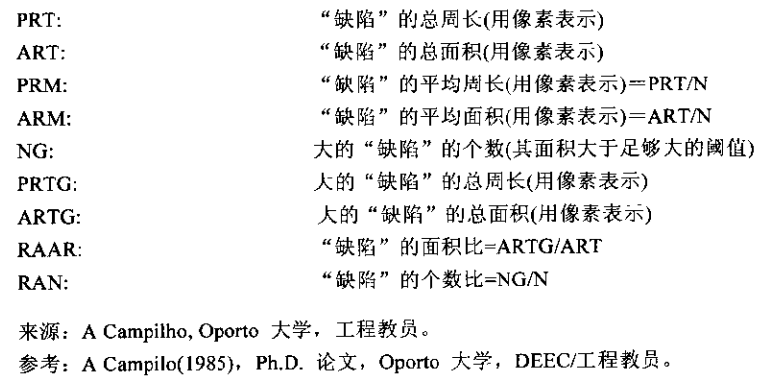
#### S1：胸部组织

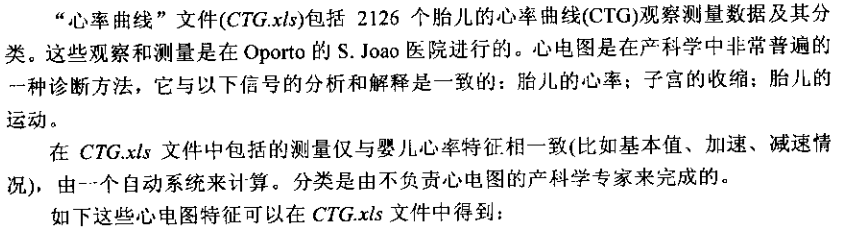


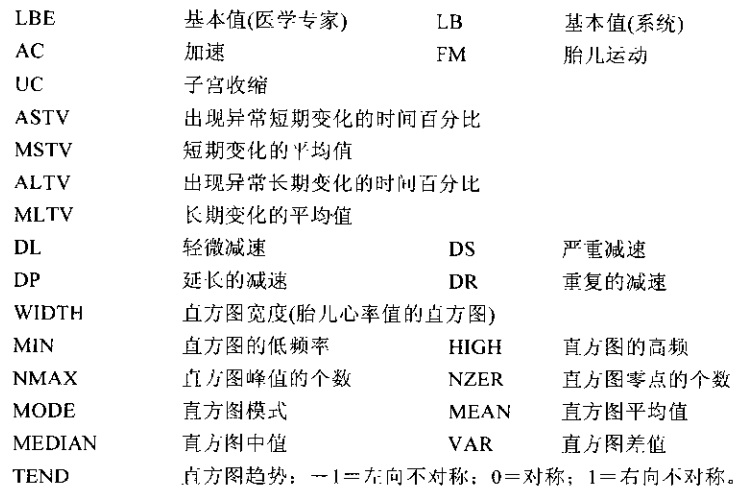
#### S2：软木塞数据

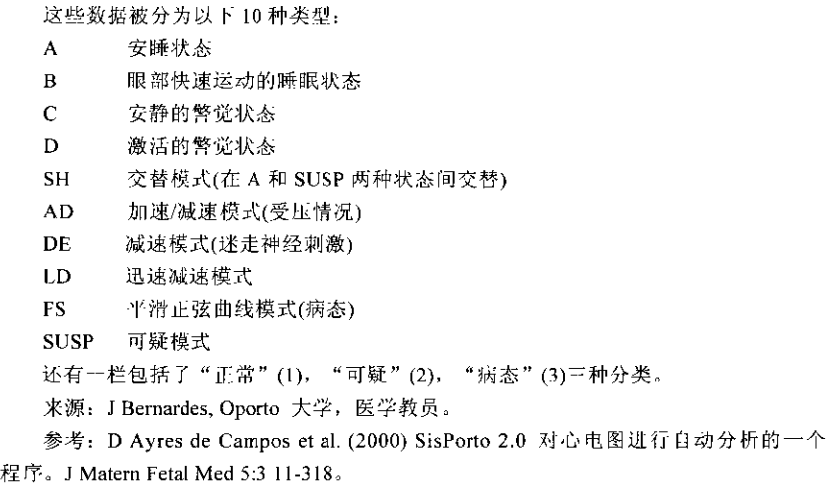




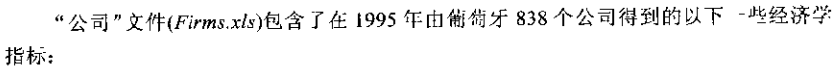
#### S3：心率曲线

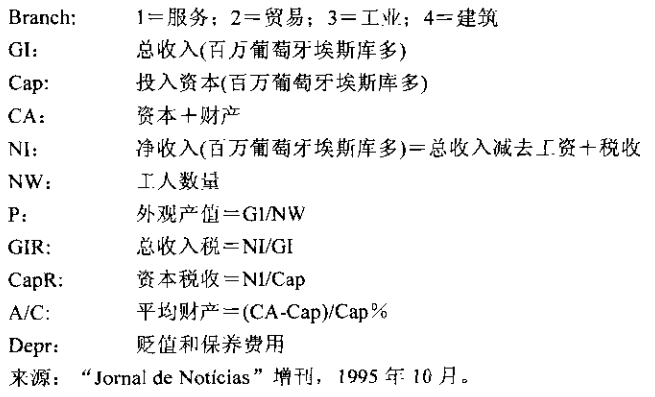




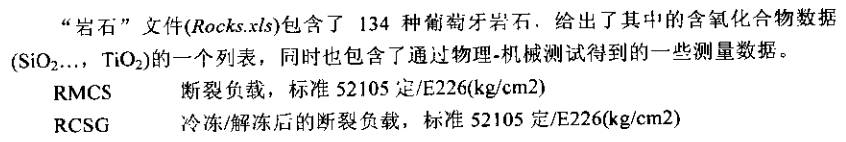


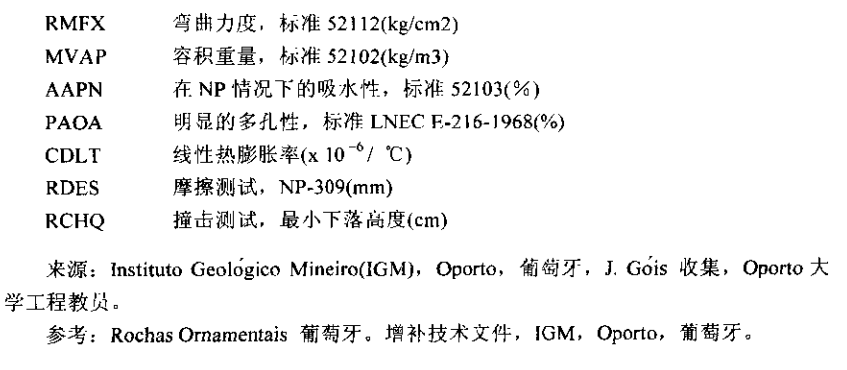
#### S4：公司数据





#### S5：岩石数据





#### **S6：鸢尾花数据集(iris.csv)**

**简介:**

Iris 数据集包括三种鸢尾花品种，每种品种有 50 个样本，以及每种花的一些特性。一种花种与其他两种可线性分离，但其他两种彼此不可线性分离。

**属性信息:**

1. sepal length in cm  
2. sepal width in cm  
3. petal length in cm  
4. petal width in cm  
5. species:  
-- Iris Setosa  
-- Iris Versicolour  
-- Iris Virginica

#### S7：酒的化学成分数据集（wine.csv）

**简介:**

这些数据是对意大利同一地区种植但来自三种不同品种的葡萄酒进行化学分析的结果。分析确定了三种类型葡萄酒中每种葡萄酒中发现的 13 种成分的含量。这些成分有：

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

每种酒的样本数目：

class 1 59

class 2 71

class 3 48

#### S8：红酒品质数据集（winequality-red.csv）

**简介:**

这数据集与葡萄牙 的红葡萄酒有关。通过检测物理化学参数变量（输入）来判定其品质（感官输出）变量。表格中的属性如下：

输入变量:

1 - fixed acidity

2 - volatile acidity

3 - citric acid

4 - residual sugar

5 - chlorides

6 - free sulfur dioxide

7 - total sulfur dioxide

8 - density

9 - pH

10 - sulphates

11 - alcohol

输出（分类变量）:

12 - quality (score between 0 and 10)

#### S9：玻璃识别数据集（glass.csv）

**简介:**

玻璃类型分类的研究是由犯罪学调查推动的。在犯罪现场，留下的玻璃如果能被正确识别，就可以作为证据。该数据集共214个样本，属性为：

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

各属性数据具有的统计特性如下：

Attribute: Min Max Mean SD Correlation with class

2. RI: 1.5112 1.5339 1.5184 0.0030 -0.1642

3. Na: 10.73 17.38 13.4079 0.8166 0.5030

4. Mg: 0 4.49 2.6845 1.4424 -0.7447

5. Al: 0.29 3.5 1.4449 0.4993 0.5988

6. Si: 69.81 75.41 72.6509 0.7745 0.1515

7. K: 0 6.21 0.4971 0.6522 -0.0100

8. Ca: 5.43 16.19 8.9570 1.4232 0.0007

9. Ba: 0 3.15 0.1750 0.4972 0.5751

10. Fe: 0 0.51 0.0570 0.0974 -0.1879

样本选择情况如下：

-- 163 Window glass (building windows and vehicle windows)

-- 87 float processed

-- 70 building windows

-- 17 vehicle windows

-- 76 non-float processed

-- 76 building windows

-- 0 vehicle windows

-- 51 Non-window glass

-- 13 containers

-- 9 tableware

-- 29 headlamps

#### S10：水质量数据集（waterQuality.csv）

**简介:**

这是一组由城市环境中水质的虚构数据创建而成的数据。根据水质参数判断水源是否安全。属性如下:

aluminium - dangerous if greater than 2.8

ammonia - dangerous if greater than 32.5

arsenic - dangerous if greater than 0.01

barium - dangerous if greater than 2

cadmium - dangerous if greater than 0.005

chloramine - dangerous if greater than 4

chromium - dangerous if greater than 0.1

copper - dangerous if greater than 1.3

flouride - dangerous if greater than 1.5

bacteria - dangerous if greater than 0

viruses - dangerous if greater than 0

lead - dangerous if greater than 0.015

nitrates - dangerous if greater than 10

nitrites - dangerous if greater than 1

mercury - dangerous if greater than 0.002

perchlorate - dangerous if greater than 56

radium - dangerous if greater than 5

selenium - dangerous if greater than 0.5

silver - dangerous if greater than 0.1

uranium - dangerous if greater than 0.3

is\_safe - class attribute {0 - not safe, 1 - safe}

#### S11：印第安人糖尿病数据集（diabetes.csv）

**简介:**

这个数据集最初来自国家糖尿病、消化和肾脏疾病研究所。该数据集的目标是根据数据集中包含的某些诊断测量结果，诊断性地预测患者是否患有糖尿病。从一个更大的数据库中选择这些实例时设置了几个限制条件。特别是，这里的所有患者都是至少 21 岁的皮马印第安裔女性。数据集中包含几个医疗变量和一个目标“结果”变量 。预测变量包括患者怀孕次数、她们的体重指数、胰岛素水平、年龄等等。

**参考文献：**

Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus

#### S12：大米类型（Cammeo和Osmancik）数据

#### （Rice\_Cammeo\_Osmancik.xlsx）

**简介：**

在土耳其种植的认证大米中，自 1997 年以来种植面积较大的奥斯曼奇克品种和自 2014 年以来种植的卡米欧品种被选作研究对象。查看奥斯曼奇克品种的一般特征时，它们外观宽长、有玻璃质感且暗淡。查看卡米欧品种的一般特征时，它们外观宽长、有玻璃质感且暗淡。对这两个品种共拍摄了 3810 张米粒图像，进行了处理并做出了特征推断。每粒大米获得了 7 个形态特征，分别为：

1. Area: Returns the number of pixels within the boundaries of the rice grain.
2. Perimeter: Calculates the circumference by calculating the distance between pixels around the boundaries of the rice grain.
3. Major Axis Length: The longest line that can be drawn on the rice grain, i.e. the main axis distance, gives.
4. Minor Axis Length: The shortest line that can be drawn on the rice grain, i.e. the small axis distance, gives.
5. Eccentricity: It measures how round the ellipse, which has the same moments as the rice grain, is.
6. Convex Area: Returns the pixel count of the smallest convex shell of the region formed by the rice grain.
7. Extent: Returns the ratio of the region formed by the rice grain to the bounding box pixels

类别Class：Commeo and Osmancik.