**所有数据均取自UCI**

**Leaf叶数据集 （340行16列）**

共有340个样本，该数据集包含从从总共40种不同植物物种中提取的叶片标本的数字图像中提取的形状和纹理特征的集合。

属性信息：（列）

1.类别  
2.试样号  
3.偏心 率  
4.纵横比  
5.伸长率  
6. 坚固性  
7.随机凸度  
8.等压因子  
9.最大压痕深度  
10.松散度  
11.平均强度  
12.平均对比度  
13.平滑度  
14.第三时刻  
15.均匀性  
16.熵

**Glass玻璃识别数据集（214行11列）**

214个样本，一共7类玻璃，用其中化学成分定义数据

属性信息：（列）

1.类别

2. ID号：1至214  
3. RI：折射率  
4. Na：钠（单位度量：相应氧化物的重量百分比，属性4-10）  
5. Mg：镁  
6. Al：铝  
7. Si：硅  
8. K：钾  
9.的Ca：钙  
10.巴：钡  
11. Fe为铁

**Wine葡萄酒数据集（178行14列）**

178个样本，对来自意大利同一地区但来自三个不同品种的葡萄酒进行化学分析的结果。分析确定了三种葡萄酒中每种所含13种成分的数量。

属性信息：（列）

1.类别

2.酒精  
3.苹果酸  
4.灰分  
5.灰分的碱度  
6.镁  
7总酚  
8.黄酮类 化合物  
9.非类黄酮酚  
10.原花青素  
11.颜色强度  
12.色调  
13.稀释酒的OD280 / OD315  
14.脯氨酸

**Spambase垃圾邮件数据集（4601列58行）**

共4601个样本，通过57个属性将电子邮件分为垃圾邮件或非垃圾邮件，类别在最后一列

属性信息：（列）

1.类别（1为垃圾邮件，0不是）

2-49：某些特定单词的出现频率

50-55：某些特定符号出现的频率

56. 不间断的大写字母序列的平均长度

57. 最长的不间断的大写字母序列的长度

58. 不间断的大写字母序列的长度之和

**wdbc乳腺癌威斯康星州（诊断）数据集(569行31列)**

威斯康星州诊断性乳腺癌数据库，将肿瘤分为良性和恶性两类

属性信息：（列）

1.类别 恶性=1 良性=0

2-31.从三个分离平面上提取细胞核的10个特征，共30维，具体特征如下

a）半径（中心到周长上的点的距离的平均值）  
b）纹理（灰度值的标准偏差）  
c）周边  
d）面积  
e）光滑度（半径长度的局部变化）  
f）紧密度（周长^ 2 /面积-1.0）  
g）凹度（轮廓凹部的严重程度）  
h ）凹点（轮廓的凹入部分的数量）  
i）对称性  
j）分形维数（“海岸线近似”-1）