

- 第6次作业
  - 1.机器学习: 朴素贝叶斯法与K近邻法分类

## 第6次作业

---

### 1.机器学习: 朴素贝叶斯法与K近邻法分类

给定乳腺癌诊断的训练集 `breast-cancer-wisconsin-train.data` 与测试集 `breast-cancer-wisconsin-test.data`. 数据集说明如下:

- 来源于[UCI Machine Learning Repository: Breast Cancer Wisconsin \(Diagnostic\) Data Set](#), 并进行了一定的处理: 删除了有数据缺失的数据并将原文件分成了训练集和测试集两部分.
- 数据集的每一行是一个患者的诊断数据, 每个数据用逗号隔开.
- 第1列是患者的id编号.
- 第2-10列是患者的9个特征, 包括肿瘤细胞大小, 形状等特征; 这些特征已经过预处理, 每个特征的取值为1-10之间(整数).
- 第11列是患者的诊断结果, 2代表良性, 4代表恶性.

要求: 在训练集上分别用朴素贝叶斯法和K近邻法完成分类器的训练. 最后计算测试集的分类准确率来评估两种算法的性能.