机器学习课程第一次作业报告

* 所选用的机器学习算法:
  + 决策树与支持向量机
* 数据集:
  + UCI网站数据集: ”Car Evaluation”
  + 链接: <http://archive.ics.uci.edu/ml/datasets/Car+Evaluation>
* 原始数据选择与分析:
  + 我们小组经过筛选比较,选定了”Car Evaluation”作为数据集.该数据集数据量适中,特征数目和特征值类型,类别都比较多.适合进行机器学习的训练和测试.
  + 数据集整体情况如下:
    - 类别(4类):
      * unacc, acc, good, vgood
    - 特征(5种):
      * buying: vhigh, high, med, low.
      * maint: vhigh, high, med, low.
      * doors: 2, 3, 4, 5more.
      * persons: 2, 4, more.
      * lug\_boot: small, med, big.
      * safety: low, med, high.
  + 可见各维特征均为离散型变量, 不过取值类型不统一,需要对数据进行预处理.
  + 我们通过代码分析了原始数据类别的分布情况,见下图: \*\*\*\*[这里是图]\*\*\*