

4.1 试证明对于不含冲突数据 (即特征向量完全相同但标记不同) 的训练集, 必存在与训练集一致 (即训练误差为 0) 的决策树

假设我们有训练集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 其中 x_i 是第 i 个样本的特征向量, 并且特征向量是唯一的, y_i 是第 i 个样本的标记

要证明存在一个决策树 T , 对于所有的 i , 有 $T(x_i) = y_i$

考虑数学归纳法对于训练集中的每个样本 x_i , 由于特征向量是唯一的, 因此我们可以选择一个特征 f , 使得 $f(x_i)$ 是唯一的. 将决策树的根节点设置为根据特征 f 分为两个子集的决策, 即 $f(x_i) \leq c$ 和 $f(x_i) > c$. 对于每个子集, 重复这个过程, 由于特征是唯一的, 我们可以得到一个符合要求的决策树