**2.1什么是熵，熵的计算公式？（课堂上已做同学，不用再做了）**

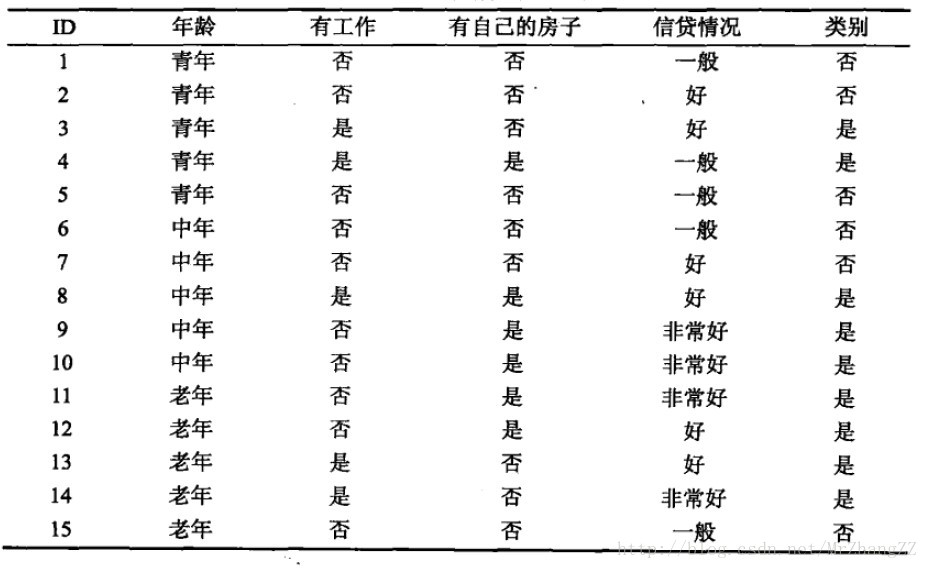
**2.2决策树如何进行“剪枝”处理？**

**2.3试析使用“最小训练误差”作为决策树划分选择的缺陷。**

**2.4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。**

**2.5对表2-1的训练数据集，根据信息增益准则选择最优特征，列出计算步骤； 建立决策树，画出该决策树。**

表格 2-1：训练数据集



**2.6 考虑表2-2中二元分类问题的训练样本。**

**(a)计算整个训练样本集的Gini指标值。**

**(b)计算属性顾客ID的Gini指标值。**

**(c)计算属性性别的Gini指标值。**

**(d)计算使用多路划分属性车型的Gini指标值。**

**(e)计算使用多路划分属性衬衣尺码的Gini指标值。**

**(f)下面哪个属性更好，性别、车型还是衬衣尺码？**

**(g)解释为什么属性顾客ID的Gini值最低，但是不能作为属性测试条件。**

