

1. 根据下表 1 所给的训练数据集，利用信息增益比（C4.5 算法）生成决策树。

答：特征(年龄)的信息增益比为 0.052

特征(有工作)的信息增益比为 0.352

特征(有自己的房子)的信息增益比为 0.433

特征(信贷情况)的信息增益比为 0.232

特征(有自己的房子)的信息增益比最大，选择为根节点划分特征

当有自己的房子为是

此时类别均为是，纯净，停止划分

当有自己的房子为否

特征(年龄)的信息增益比为 0.164

特征(有工作)的信息增益比为 1.000

特征(信贷情况)的信息增益比为 0.340

特征(有工作)的信息增益比最大，选择为当前节点特征

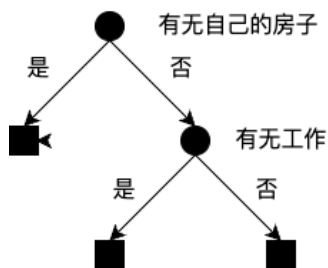
当有工作为否

此时类别均为否，纯净，停止划分

当有工作为是

此时类别均为是，纯净，停止划分

决策树如下：



2. 对下表 1 所给的训练数据集，根据信息增益准则选择最优特征。

答：特征(年龄)的信息增益为 0.083

特征(有工作)的信息增益为 0.324

特征(有自己的房子)的信息增益为 0.420

特征(信贷情况)的信息增益为 0.363

特征(有自己的房子)的信息增益最大，选择为根节点划分特征

当有自己的房子为是

此时类别均为是，纯净，停止划分

当有自己的房子为否

特征(年龄)的信息增益为 0.252

特征(有工作)的信息增益为 0.918