
算法 1 决策树学习算法

输入: 训练集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$

属性值 $A = \{a_1, a_2, \dots, a_d\}$

输出: 以 node 为根节点的一棵决策树

```

1: function TREETRAIN( $D, A$ )
2:   生成节点  $node$ 
3:   if  $D$  中样本全属于同一类别  $C_k$  then
4:     将  $node$  标记为  $C_k$  类叶节点;return
5:   end if
6:   if  $A = \phi$  then Or  $D$  中样本在  $A$  上取值相等
7:     将  $node$  标记为叶节点, 其类别为  $D$  中样本数最多的类; return
8:   end if
9:   计算  $A$  中各特征的信息增益率
10:  选择信息增益率最大的属性  $a^*$ 
11:  for  $a^*$  的每一个值  $a_*^v$  do
12:    为  $node$  生成一个分支; 令  $D_v$  表示  $D$  中在  $a^*$  上取值为  $a_*^v$  的样本子集
13:    if  $D_v$  为空 then
14:      将分支节点标记为叶节点, 其类别标记为  $D$  中样本最多的类;Return
15:    else
16:      以 TreeTrain( $D_v, A \setminus \{a^*\}$ ) 为分支节点
17:    end if
18:  end for
19: end function
  
```
