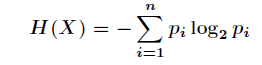
1.决策树

决策树是依托决策而建立起来的树。是一种预测模型，代表的是一种对象属性与对象值之间的映射关系，每一个节点代表某个对象，树中每个分叉路径代表某个可能的属性值，而每一个叶子节点则对应从根节点到该叶子节点所经历路径所表示的属性值。决策树仅有单一输出。

2.信息熵

熵是对不确定性的度量。离散随机事件出现的概率。一个系统越是有序，信息熵就越低，反之一个系统越是混乱它的信息熵就越高。所以信息熵是系统有序化程度的一个度量。

加入一个随机变量X的取值为X={x1,x2,x3,.......,xn}，每一种取到的概率分别是{p1,p2,p3,.......,pn}，那么X的熵定义为：



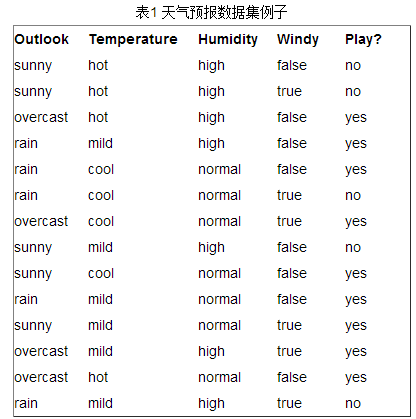
意思是一个变量的变化情况越多，那么它携带的信息量就越大。

3.信息增益

信息增益是针对一个一个特征而言的，就是看一个特征t，系统有它和没有它时的信息量各是多少，两者的差值就是这个特征给系统带来的信息量，即信息增益。

举例：

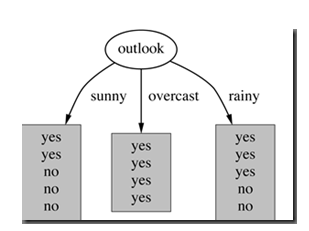
学习目标是play 或 not play:



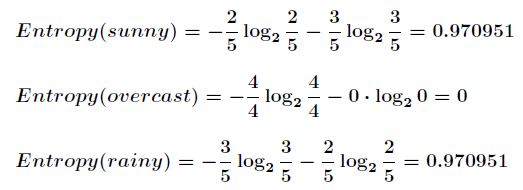
可以看出，总共14个样本，9个yes的，5个no的。 那么信息熵是：

clipboard.png

在决策树分类问题中，某个属性的信息增益就是决策树在进行该属性划分前和划分后信息的差值，以Outlook为例：



划分后，数据被分为三部分了（三种Outlook值到最终结果的映射），那么各个分支的信息熵计算如下：



那么划分后的信息熵为：

clipboard.png

Entropy(S|T)代表在特征属性T的条件下样本的条件熵。 那么最终得到的该特征属性T（这里就是Outlook属性）带来的信息增益为：

clipboard.png

ID3核心思想： 在决策树的每个非叶子节点划分之前，先计算每一个属性所带来的信息增益，选择最大信息增益的属性来划分，因为信息增益越大，区分样本的能力就越强，越具有代表性，这是一种自顶向下的贪心策略。

公式：

clipboard.png