

**决策树**

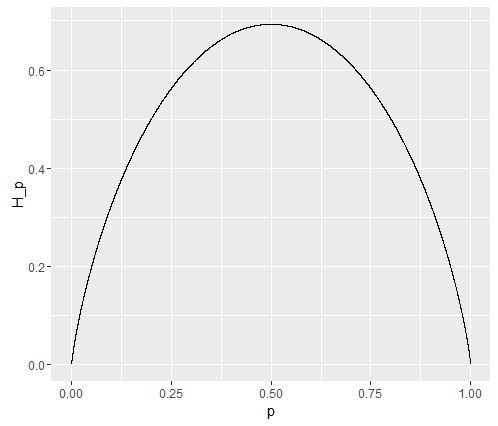
**不确定性：**

不确定性函数满足两个条件：1、不确定性函数是概率的单调递降函数；2、两个独立符号所产生的不确定性应等于各自不确定性之和，即这称为可加性。同时满足这两个条件的函数是对数函数，即。

**熵：**

熵是不确定性的期望值，在信源中，考虑的不是某一单个符号发生的不确定性，而是要考虑这个信源所有可能发生情况的平均不确定性。若信源符号有种取值：，对应概率为：，且各种符号的出现彼此独立。这时，信源的平均不确定性应当为单个符号不确定性的统计平均值可称为信息熵，即。

当随机变量只有0和1两种取值时, , 期望为，求导得到：



**信息增益（ID3）：**

假设样本集为，包含属性，现在求属性的信息增益，可以用属性将划分为v个子集

基尼系数

举例说明：

