# **一.信息熵**

表示一个随机变量的复杂性或者不确定性。

当前样本集D中第 i 类样本所占的比例 pi ,则D的信息熵为

IMG_256  
信息熵越小,代表事件越确定。  
换到决策树中可以表示某类样本所占总样本数的比例很大。

# **二.条件熵**

条件熵：表示在直到某一条件后，某一随机变量的复杂性或不确定性。

假定数据集D中离散属性a有V个可能的取值,若使用a对数据进行划分,则会产生V个分支节点,其中第v个分支节点包含了D中所有在属性a上取值为av的样本,记为Dv,  
IMG_257

# **三.信息增益**

表示在知道某一条件后，某一随机变量的不确定性的减少量。

假定数据集D中离散属性a有V个可能的取值,若使用a对数据进行划分,则会产生V个分支节点,其中第v个分支节点包含了D中所有在属性a上取值为av的样本,记为Dv,可以用属性a对样本集D进行划分所得的信息增益(informatongan)  
IMG_258  
描述了在知道a之后数据集D不确定性减少的程度。

# **四.条件熵和信息增益的关系**

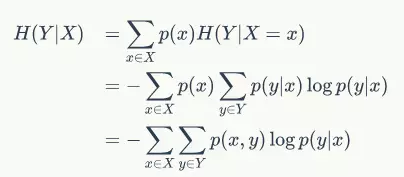
由公式和例子可是看出 : ****信息增益 = 信息熵 - 条件熵****

# **五.怎么理解条件熵**

## **1.条件熵**

定义为X给定条件下，Y的条件概率分布的熵对X的数学期望  
设有随机变量（X,Y），其联合概率分布为  
IMG_259  
条件熵 H（Y|X）表示在已知随机变量X的条件下随机变量Y的不确定性。  
随机变量X给定的条件下随机变量Y的条件熵H(Y|X)

## **2.推导公式**



# **六.举例**

1. 信息熵  
   假如双十一我要剁手买一件衣服，但是我一直犹豫着要不要买，我决定买这件事的不确定性（熵）为2.6。
2. 条件熵  
   我在看了这件衣服的评价后，我决定买衣服这件事的不确定性是1.2。  
   我在线下实体店试穿衣服后，我决定买衣服这件事的不确定性是0.9。
3. 信息增益  
   上面条件熵给出了两个：  
   一个是看了网上的评价，此时的信息增益是Gain1=2.6−1.2=1.4。  
   另一个是线下试穿了衣服，此时的信息增益Gain2=2.6−0.9=1.7。

很显然我在线下试穿衣服之后对于决定买这件衣服的不确定度下降更多，更通俗的说就是我试穿衣服之后买这件衣服的可能性更大了。所以如果有看买家评价和线下试穿两个属性，首先应该选择线下试穿来构建内部节点。