```
"""
Created on 2017/8/11

machine-learning-course

@author: DSG
"""
```

● 香农熵

符号 x_i 的信息定义为：

$$l(x_i) = -\log_2^{p(x_i)}$$

$p(x_i)$ 是数据集中每个分类的概率。在这里就是数据集中每个类别的数据量占总数据量的比份。

熵定义为信息的期望值：

$$H = -\sum_{i=1}^{n} p(x_i)\log_2^{p(x_i)}$$

N 表示数据中类别总数，熵越大表明数据越混乱。

● 决策树

每次按照最大信息增益作为数据分类的依据，信息增益定义为，如果安照数据的某个特征的不同取值把数据划分为不同的子集，然后分别计算每个子集的香农熵，最后对所有子集的香农熵求期望（期望就是某个子集数据量占总数据集量的比份），最后得到的期望就是划分后的数据熵，这个熵和原数据集的熵的差异就是信息增益。这个差异为原始熵减去新熵。

算法思路：

1. 数据划分依据选择 :对数据的集中的所有特征计算信息增益,选择最大的增益作为本次数据化分的依据。
2. 决策树构建思路：首先选择出数据集中本次最佳分类特征，利用这个分类的标签作为 root 节点创建子树，找到这个特征下所有的非重复取值，以每个取值为节点创建 root 节点的孩子节点，同时以这个取值切分数据集，对每个子集递归使用以上的方法，值得注意的是每次消耗掉一个特征。递归结束的的两个条件：1、接受到的数据集中所有的类别标签完全一样的时候，已经归一，返回；2、只剩下一个特征时，但是标签不唯一，采用投票是选举法结束。