

## 决策树模型与学习

1. 决策树学习的损失函数通常是正则化的极大似然函数。它的策略是以损失函数为目标函数的最小化。
2. 决策树定义: 节点和有向边组成。节点有两种: 内部节点与叶节点。内部节点表示一个特征或属性, 叶节点表示

### 特征选择

1. 特征选择的准则是信息增益或信息增益比
2. 熵的定义: 表示随机变量不确定性的度量。设X是一个取有限个值的离散随机变量, 其概率分布为:

$$P(X = x_i) = p_i$$

随机变量X的熵为:

$$H(X) = -\sum_{i=1}^n p_i \log_2 p_i$$

其中  $0 \log_2 0 = 0$

熵越大, 随机变量的不确定性越大。

3. 熵和条件熵的概率有数据估计得到时, 所对应的熵与条件熵分别成为经验熵和经验条件熵:

$$H(Y|X) = \sum_{i=1}^n p_i H(Y|X = x_i)$$

### 信息增益

1. 信息增益表示得知特征X的信息而使得类Y的信息的不确定性减少的程度:

$$g(D, A) = H(D) - H(D|A)$$

当  $g(D, A)$  越大, 即特征A的信息增益越大时,  $H(D|A)$  越小, 特征A对数据集D分类效果越明显。

### 信息增益比

定义: 特征A对训练数据集D的信息增益比  $g_r(D, A)$  定义为其信息增益  $g(D, A)$  与训练数据集D关于特征A的值的熵  $H_A(D)$

$$g_r(D, A) = g(D, A) / H_A(D)$$

$$H_A(D) = -\sum_{i=1}^n |D_i| / |D| \log_2 |D_i| / |D|, n \text{ 是特征A取值的个数}$$

### 常用的决策树分类算法