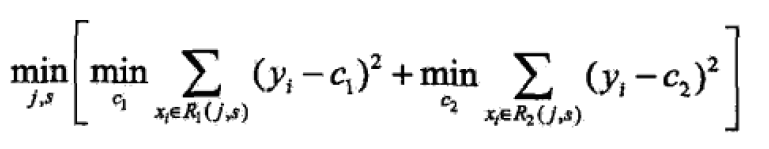
1. 自己提出的问题的理解（罗列全部）：

无

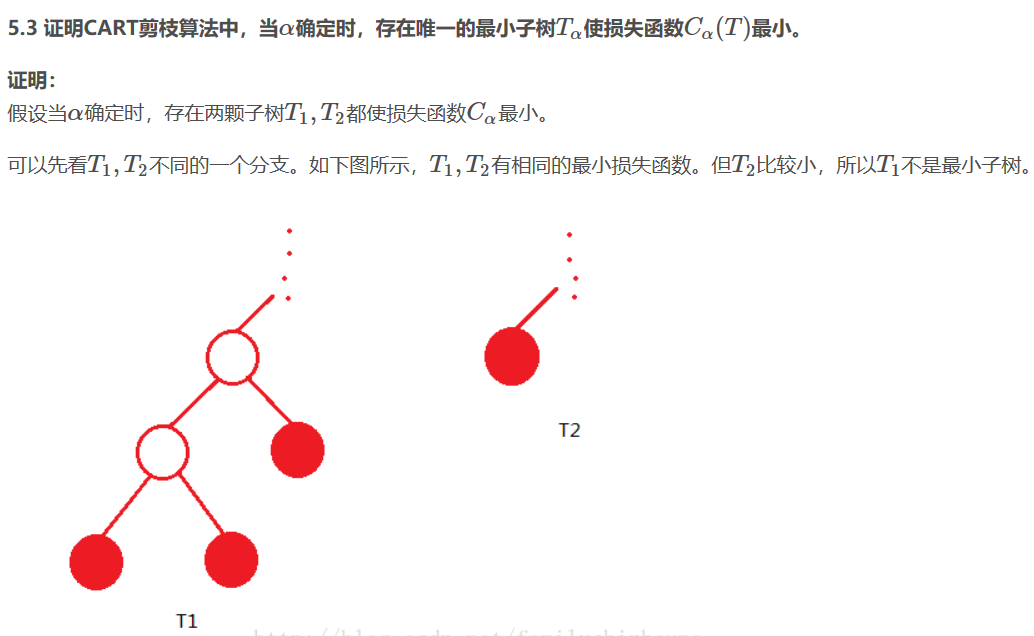
1. 别人提出的问题的理解（选择几个问题罗列，并给出理解）：
2. 问题1：CART的回归决策树生成算法中



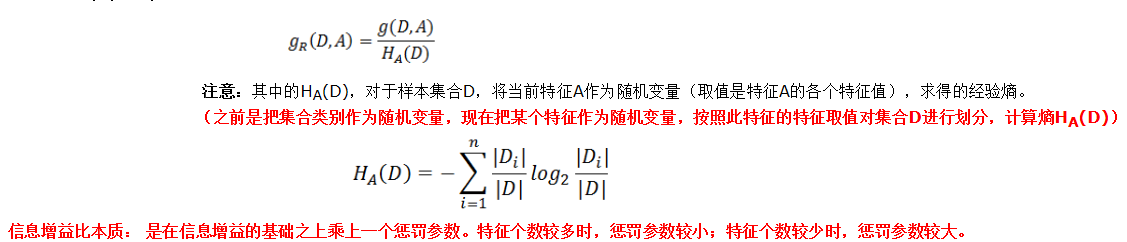
这里的求最小值中的c1和c2也是平均值吗？

讨论后的理解：最优的c值就是均值，min的部分是方差

1. 问题2：证明CART剪枝算法中，当α确定时，存在唯一的最小子树Tα使损失函数Cα(T)最小  
   讨论后的理解：（此处引用吴亦珂同学的csdn上的解答）



1. 问题3： 在5.4决策树的剪枝中，对于式子(5.14)而言，为什么较大的α促使选择较简单的模型，较小的α促使选择较复杂的模型？  
   讨论后的理解：节点数|T|越大的话，也就是复杂度越高，所以如果增加α的值的话，就会让损失函数更大，更倾向于选择复杂度低的树，α越大对复杂度高的模型的惩罚越大。
2. 问题4：ID3算法和C4.5算法的区别，我认为信息熵增益最大不是应该就意味着相对信息熵增益最大吗，毕竟信息熵是相同的  
   讨论后的理解：有区别，可以有效的去掉特征数目较多的影响



1. 问题5：CART算法在决策树生成中，基于训练数据集生成决策树，生成的决策树要尽量大是为什么？  
   讨论后的理解：为了考虑更多的情况
2. 问题6：式子5.31表示了剪枝后整体损失函数减少的程度，但为什么说是在T0中剪去g(t)最小的T，不应该剪去损失函数减少的多的吗，怎么理解？  
   讨论后的理解：对于不同的α值有不同的最优子树
3. 问题7：决策树的α值具体如何确定？  
   讨论后的理解：看具体要求，α是限制模型复杂度的。
4. （必填）读书计划
5. 本周完成的内容章节：《统计学习方法》第五章
6. 下周计划：《统计学习方法》第六章

四、读书摘要及理解或伪代码的具体实现（读书摘要、伪代码的具体实现代码等可以写到这个部分）

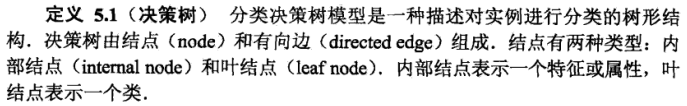
读书摘要及理解：

# 决策树

主要优点是模型具有可读性，分类速度快。包括3个步骤：特征选择、决策树的生成和决策树的裁剪。

## 5.1 决策树模型与学习

### 5.1.1 决策树模型



内部节点表示一个特征或属性，叶节点表示一个类。

### 5.1.2 决策树与if-item规则

决策树的路径或其对应的if-then规则集合具有一个重要的性质：互斥并且完备。

### 5.1.3 决策树与条件概率分布

各叶节点上的条件概率往往偏向某一个类，即属于某一类的概率较大。

### 5.1.4 决策树学习

决策树学习的算法通常是一个递归地选择最优特征，这一过程对应着对特征空间的划分，也对应着决策树的构建。

由于决策树表示一个条件概率分布，所以深浅不同的决策树对应着不同复杂度的概率模型，决策树的生成对应于模型的局部选择，决策树的剪枝对应于模型的全局选择。

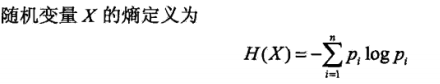
## 5.2 特征选择

### 5.2.1 特征选择问题

特征选择在于选取对训练数据具有分类能力的特征，准则通常是信息增益或者信息增益比。

### 5.2.2 信息增益

**熵**是表示随机变量不确定性的度量。



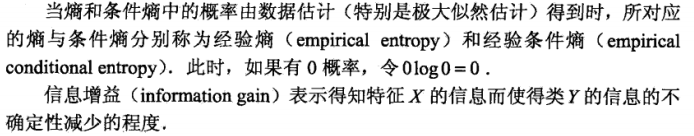
由定义可知，上只依赖于X的分布，而与X的取值无关。

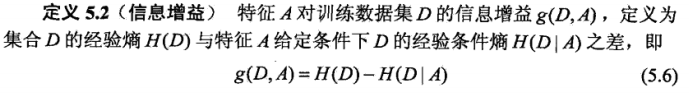
熵越大，随机变量的不确定性就越大。



**条件熵**H(X|Y)表示在一直随机变量X的条件下随机变量Y的不确定性。



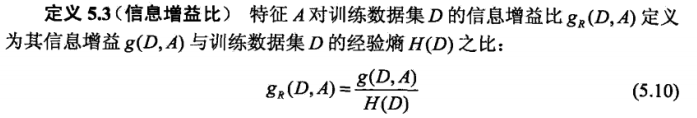






### 5.2.3 信息增益比

信息增益值的大小是相对于训练集而言的，并没有绝对意义。

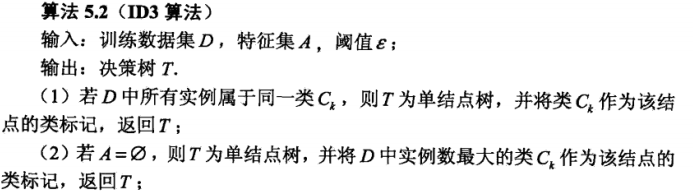


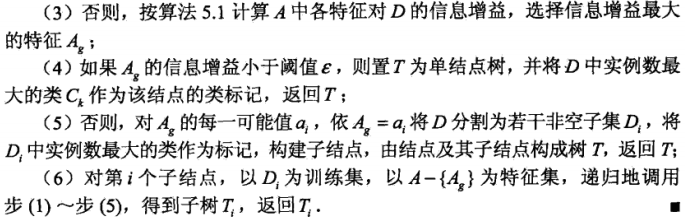
## 5.3 决策树的生成

### 5.3.1 ID3算法

直到所以特征的信息增益均很小或没有特征可以选择为止。ID3相当于用极大似然法进行概率模型的选择。

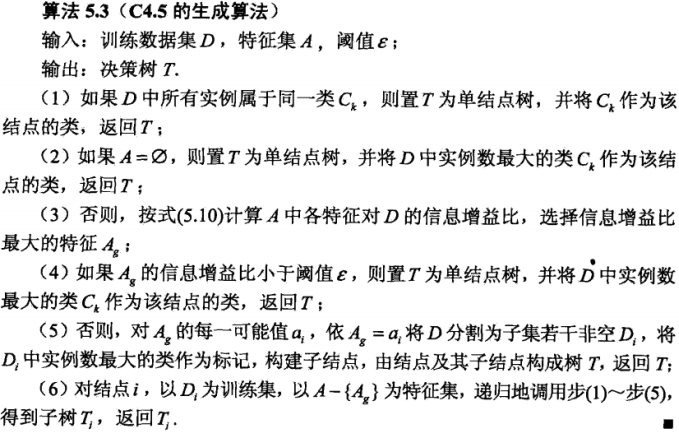
ID3 算法只有树的生成，所以该算法生成的树容易产生过拟合。





### 5.3.2 C4.5的生成算法

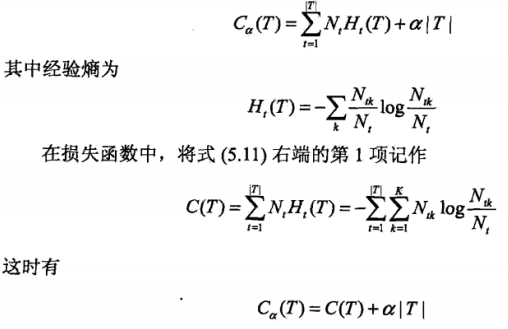
C4.5在生成的过程中，用信息增益比来选择特征。



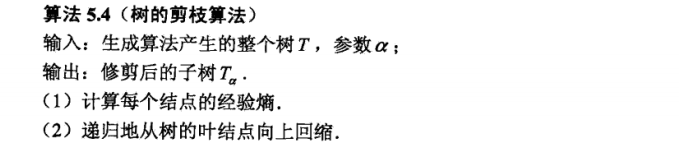
## 5.4 决策树的剪枝

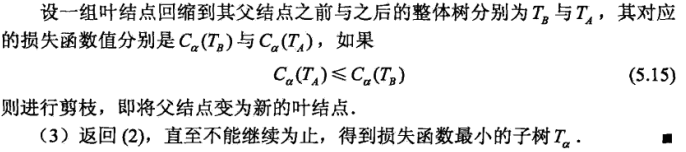
往往通过极小化决策树整体的损失函数或代价函数来实现。

**决策树的损失函数：**



当α确定是，选择损失函数最小的模型。





## 5.5 CART算法

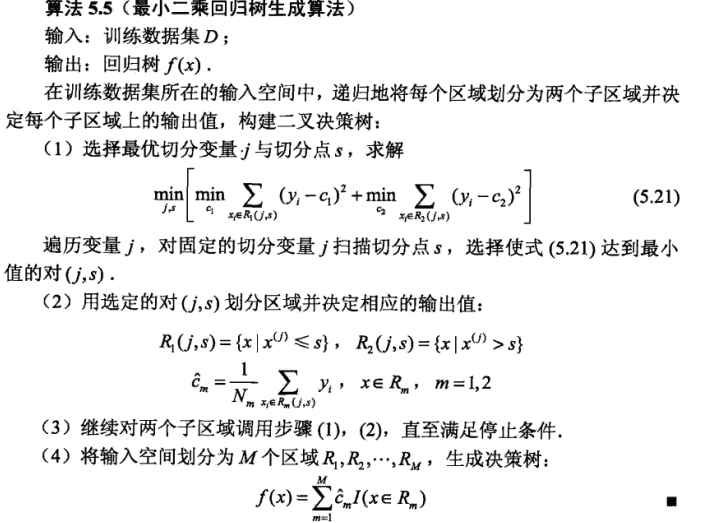
分类与回归树（CART）分为两步：

1. 决策树生成：基于训练数据集生成决策树，生成的决策树要尽量大；
2. 决策树剪枝：用验证数据集对已生成的树进行剪枝并选择最优子树，这时用损失函数最小作为剪枝的标准。

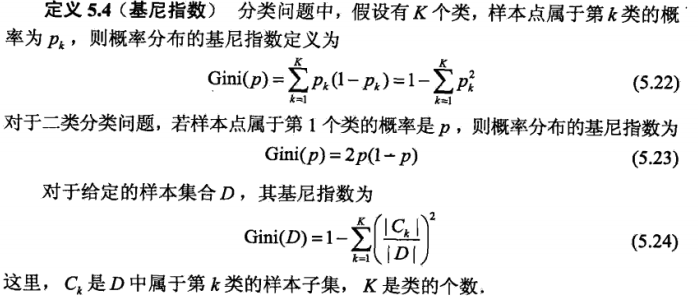
### 5.5.1 CART生成

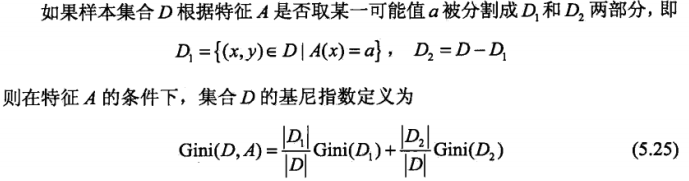
对于回归树用平方误差最小化准则，对分类树用基尼指数最小化准则。

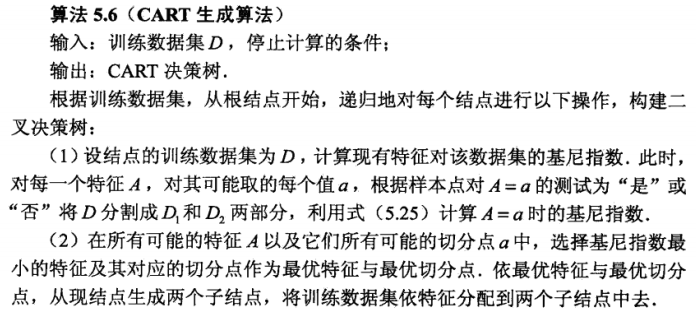
1. 回归树的生成



1. 分类树的生成









### 5.5.2 CART剪枝

