人工智能 I

杨杰

yangjie@seu.edu.cn

2025/11/24，五四楼-303，9:50~11:25

本周四实验课安排

周次	10	
日期	2025/11/27	
星期	星期四	
时间	16:40-18:15	19:00-20:35
机房地点	中心楼实验室 D区	五五楼435 软实七
实验任务	决策树与随机森林	

注意两个时间段是不同机房！请勿跑错！
注意18:15机房电脑自动关机！请及时保存！

本节课安排

- 监督学习：
  - 线性回归
  - 对率回归
  - 支持向量机
  - 决策树
  - 随机森林
  - 贝叶斯分类器
  - 感知机与神经网络

本节课安排

- 监督学习：
  - 线性回归
  - 对率回归
  - 支持向量机
  - 决策树 Decision Tree
  - 随机森林
  - 贝叶斯分类器
  - 感知机与神经网络

决策树（Decision Tree）

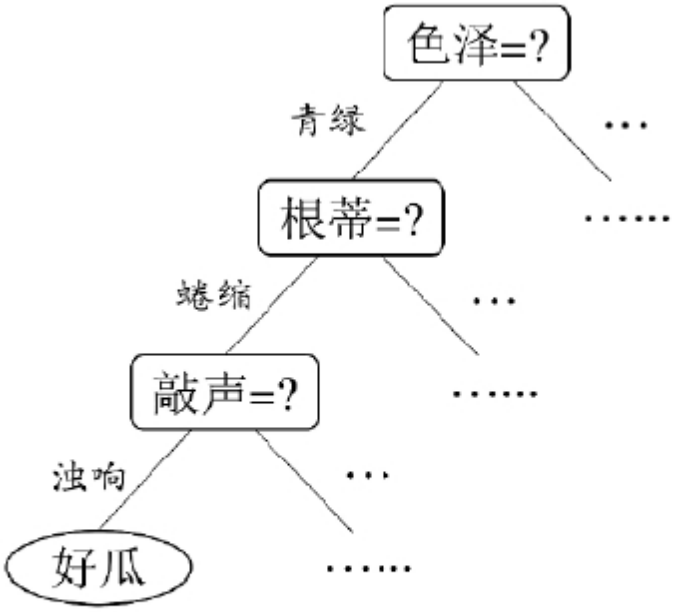


图 4.1 西瓜问题的一棵决策树

表 1.1 西瓜数据集

编号	色泽	根蒂	敲声	好瓜
1	青绿	蜷缩	浊响	是
2	乌黑	蜷缩	浊响	是
3	青绿	硬挺	清脆	否
4	乌黑	稍蜷	沉闷	否

决策树基于树结构进行决策（与人的决策过程类似）：

- 每个“内部节点”对应于某个属性上的测试
- 每个分支对应于该测试的一种可能结果（即该属性的某个取值）
- 每个“叶节点”对应于一个预测结果

学习过程：寻找一颗能（尽量）正确划分训练集的决策树，寻求最优泛化性能

预测过程：将测试样本从根节点开始，沿着划分属性所构成的“判定测试序列”下行，直到叶节点

基本思路

在每个中间节点寻找一个“最优”划分属性，试图用一颗**“尽量小”的决策树**获得**“尽量低”的训练误差**

基本流程：分而治之，自根至叶的递归过程

TreeGenerate(D, A)　　数据集D，属性集A

- 选择划分属性a（含V个属性值），把D划分为 $D^1 \cup D^2 \cup \cdots \cup D^V$，　　$D^v = \{\mathbf{x} | \mathbf{x} \in D, x_a = a^v\}$
- TreeGenerate($D^v, A \backslash \{a\}$), $v = 1, \cdots, V$

三种停止条件：

（1）当前节点包含的样本全属于同一类别，无需划分；

（2）当前属性集为空，或是所有样本在所有属性上取值相同，无法划分；

（3）当前节点包含的样本集合为空，不能划分。

基本算法

输入: 训练集 $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$;
属性集 $A = \{a_1, a_2, \ldots, a_d\}$.

过程: 函数 TreeGenerate(D, A)

1: 生成结点 node;
2: **if** D 中样本全属于同一类别 C **then**
3: 将 node 标记为 C 类叶结点; **return**
4: **end if**
5: **if** $A = \varnothing$ **OR** D 中样本在 A 上取值相同 **then**
6: 将 node 标记为叶结点, 其类别标记为 D 中样本数最多的类; **return**
7: **end if**
8: 从 A 中选择最优划分属性 a_*;
9: **for** a_* 的每一个值 a_*^v **do**
10: 为 node 生成一个分支; 令 D_v 表示 D 中在 a_* 上取值为 a_*^v 的样本子集;
11: **if** D_v 为空 **then**
12: 将分支结点标记为叶结点, 其类别标记为 D 中样本最多的类; **return**
13: **else**
14: 以 TreeGenerate(D_v, $A \setminus \{a_*\}$)为分支结点
15: **end if**
16: **end for**

输出: 以 node 为根结点的一棵决策树

递归返回，情形(1)

递归返回，情形(2)

当前结点的类后验分布

递归返回，情形(3)

采用父结点的类后验分布

决策树算法的核心

信息熵

- 信息熵 (entropy) 描述的是随机变量的信息量（不确定性），是度量样本集合纯度最常用的一种指标
  - 大多数样本属于同一类，则纯度高；反之，纯度低
- 假定当前样本集合D中第k类样本所占的比例为$p_k, k = 1, \cdots, N$，我们把D的信息熵定义为：

$$H(D) = -\sum_{k=1}^{N} p_k \log_2 p_k$$

  - 约定$0 \log_2 0 = 0$
  - $0 \leq H(D) \leq \log_2 N$
  - **$H(D)$越小，纯度越高**

熵（Entropy）

- 熵(Entropy)：随机变量 X 的概率分布为 P，其信息量（不确定量）为：

$$H(X) = \mathbf{E}[I(x)] = -\mathbf{E}[\log P(x)] = -\sum P(x) \log P(x)$$

  - 也记为$H(P)$，描述概率分布的性质
  - 熵是信息中不确定性的平均度量
  - **不确定性越高，熵越大**
  - 例如，抛一枚均匀的硬币，结果（正面或反面）的熵很高，因为你完全无法预测。

一个例子

表 4.1 西瓜数据集 2.0

编号	色泽	根蒂	敲声	纹理	脐部	触感	好瓜
1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	是
2	乌黑	蜷缩	沉闷	清晰	凹陷	硬滑	是
3	乌黑	蜷缩	浊响	清晰	凹陷	硬滑	是
4	青绿	蜷缩	沉闷	清晰	凹陷	硬滑	是
5	浅白	蜷缩	浊响	清晰	凹陷	硬滑	是
6	青绿	稍蜷	浊响	清晰	稍凹	软粘	是
7	乌黑	稍蜷	浊响	稍糊	稍凹	软粘	是
8	乌黑	稍蜷	浊响	清晰	稍凹	硬滑	是
9	乌黑	稍蜷	沉闷	稍糊	稍凹	硬滑	否
10	青绿	硬挺	清脆	清晰	平坦	软粘	否
11	浅白	硬挺	清脆	模糊	平坦	硬滑	否
12	浅白	蜷缩	浊响	模糊	平坦	软粘	否
13	青绿	稍蜷	浊响	稍糊	凹陷	硬滑	否
14	浅白	稍蜷	沉闷	稍糊	凹陷	硬滑	否
15	乌黑	稍蜷	浊响	清晰	稍凹	软粘	否
16	浅白	蜷缩	浊响	模糊	平坦	硬滑	否
17	青绿	蜷缩	沉闷	稍糊	稍凹	硬滑	否

17个样本

正例占$p_1 = \frac{8}{17}$

反例占$p_2 = \frac{9}{17}$

划分前的信息熵：

$$H(D) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 = 0.998$$

一个例子

表 4.1 西瓜数据集 2.0

编号	色泽	根蒂	敲声	纹理	脐部	触感	好瓜
1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	是
2	乌黑	蜷缩	沉闷	清晰	凹陷	硬滑	是
3	乌黑	蜷缩	浊响	清晰	凹陷	硬滑	是
4	青绿	蜷缩	沉闷	清晰	凹陷	硬滑	是
5	浅白	蜷缩	浊响	清晰	凹陷	硬滑	是
6	青绿	稍蜷	浊响	清晰	稍凹	软粘	是
7	乌黑	稍蜷	浊响	稍糊	稍凹	软粘	是
8	乌黑	稍蜷	浊响	清晰	稍凹	硬滑	是
9	乌黑	稍蜷	沉闷	稍糊	稍凹	硬滑	否
10	青绿	硬挺	清脆	清晰	平坦	软粘	否
11	浅白	硬挺	清脆	模糊	平坦	硬滑	否
12	浅白	蜷缩	浊响	模糊	平坦	软粘	否
13	青绿	稍蜷	浊响	稍糊	凹陷	硬滑	否
14	浅白	稍蜷	沉闷	稍糊	凹陷	硬滑	否
15	乌黑	稍蜷	浊响	清晰	稍凹	软粘	否
16	浅白	蜷缩	浊响	模糊	平坦	硬滑	否
17	青绿	蜷缩	沉闷	稍糊	稍凹	硬滑	否

若以色泽为划分属性：

- D^1（色泽=青绿），6个样本正例占$\frac{3}{6}$，反例占$\frac{3}{6}$，其信息熵为：

$$H(D^1) = 1$$

- D^2（色泽=乌黑）与D^3（色泽=浅白）的信息熵为：

$$H(D^2) = 0.918$$
$$H(D^3) = 0.722$$

$$\mathrm{Ent}(D^1) = -\left(\frac{3}{6}\log_2\frac{3}{6} + \frac{3}{6}\log_2\frac{3}{6}\right) = 1.000$$
$$\mathrm{Ent}(D^2) = -\left(\frac{4}{6}\log_2\frac{4}{6} + \frac{2}{6}\log_2\frac{2}{6}\right) = 0.918$$
$$\mathrm{Ent}(D^3) = -\left(\frac{1}{5}\log_2\frac{1}{5} + \frac{4}{5}\log_2\frac{4}{5}\right) = 0.722$$

信息增益

- 信息增益（information gain）以信息熵为基础，计算某个划分对信息熵所造成的变化
  - 设属性a的值域为$\{a^1, a^2, \cdots, a^V\}$
  - D^v:数据集D中在属性a上取值为a^v的样本集合
  - 以属性a对数据集D进行划分，所取得的信息增益为

$$\text{Gain}(D,a) = H(D) - \sum_{v=1}^{V} \frac{|D^v|}{|D|} H(D^v)$$

$H(D)$：划分前的信息熵

$H(D^v)$：划分后的信息熵

$\frac{|D^v|}{|D|}$：第v个分支的权重，样本越多权重越高

一个例子

表 4.1 西瓜数据集 2.0

编号	色泽	根蒂	敲声	纹理	脐部	触感	好瓜
1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	是
2	乌黑	蜷缩	沉闷	清晰	凹陷	硬滑	是
3	乌黑	蜷缩	浊响	清晰	凹陷	硬滑	是
4	青绿	蜷缩	沉闷	清晰	凹陷	硬滑	是
5	浅白	蜷缩	浊响	清晰	凹陷	硬滑	是
6	青绿	稍蜷	浊响	清晰	稍凹	软粘	是
7	乌黑	稍蜷	浊响	稍糊	稍凹	软粘	是
8	乌黑	稍蜷	浊响	清晰	稍凹	硬滑	是
9	乌黑	稍蜷	沉闷	稍糊	稍凹	硬滑	否
10	青绿	硬挺	清脆	清晰	平坦	软粘	否
11	浅白	硬挺	清脆	模糊	平坦	硬滑	否
12	浅白	蜷缩	浊响	模糊	平坦	软粘	否
13	青绿	稍蜷	浊响	稍糊	凹陷	硬滑	否
14	浅白	稍蜷	沉闷	稍糊	凹陷	硬滑	否
15	乌黑	稍蜷	浊响	清晰	稍凹	软粘	否
16	浅白	蜷缩	浊响	模糊	平坦	硬滑	否
17	青绿	蜷缩	沉闷	稍糊	稍凹	硬滑	否

若以色泽为划分属性：

- D^1（色泽=青绿），6个样本正例占$\frac{3}{6}$，反例占$\frac{3}{6}$，其信息熵为：

$$H(D^1) = 1$$

- 同理，D^2（色泽=乌黑）与D^3（色泽=浅白）的信息熵为：

$$H(D^2) = 0.918$$
$$H(D^3) = 0.722$$

划分后的信息熵：

$$\sum_{v=1}^{3} \frac{|D^v|}{|D|} H(D^v) = 0.889$$

属性“色泽”的信息增益为：

$$\text{Gain}(D, \text{色泽}) = H(D) - \sum_{v=1}^{3} \frac{|D^v|}{|D|} H(D^v) = 0.109$$

一个例子

$$\mathrm{Ent}(D^1) = -(\tfrac{3}{6}\log_2\tfrac{3}{6} + \tfrac{3}{6}\log_2\tfrac{3}{6}) = 1.000$$
$$\mathrm{Ent}(D^2) = -(\tfrac{4}{6}\log_2\tfrac{4}{6} + \tfrac{2}{6}\log_2\tfrac{2}{6}) = 0.918$$
$$\mathrm{Ent}(D^3) = -(\tfrac{1}{5}\log_2\tfrac{1}{5} + \tfrac{4}{5}\log_2\tfrac{4}{5}) = 0.722$$

$$\begin{aligned}\mathrm{Gain}(D, \text{色泽}) &= \mathrm{Ent}(D) - \sum_{v=1}^{3}\frac{|D^v|}{|D|}\mathrm{Ent}(D^v) \\ &= 0.998 - (\tfrac{6}{17}\times 1.000 + \tfrac{6}{17}\times 0.918 + \tfrac{5}{17}\times 0.722) \\ &= 0.109\end{aligned}$$

课堂练习

表 4.1　西瓜数据集 2.0

编号	色泽	根蒂	敲声	纹理	脐部	触感	好瓜
1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	是
2	乌黑	蜷缩	沉闷	清晰	凹陷	硬滑	是
3	乌黑	蜷缩	浊响	清晰	凹陷	硬滑	是
4	青绿	蜷缩	沉闷	清晰	凹陷	硬滑	是
5	浅白	蜷缩	浊响	清晰	凹陷	硬滑	是
6	青绿	稍蜷	浊响	清晰	稍凹	软粘	是
7	乌黑	稍蜷	浊响	稍糊	稍凹	软粘	是
8	乌黑	稍蜷	浊响	清晰	稍凹	硬滑	是
9	乌黑	稍蜷	沉闷	稍糊	稍凹	硬滑	否
10	青绿	硬挺	清脆	清晰	平坦	软粘	否
11	浅白	硬挺	清脆	模糊	平坦	硬滑	否
12	浅白	蜷缩	浊响	模糊	平坦	软粘	否
13	青绿	稍蜷	浊响	稍糊	凹陷	硬滑	否
14	浅白	稍蜷	沉闷	稍糊	凹陷	硬滑	否
15	乌黑	稍蜷	浊响	清晰	稍凹	软粘	否
16	浅白	蜷缩	浊响	模糊	平坦	硬滑	否
17	青绿	蜷缩	沉闷	稍糊	稍凹	硬滑	否

练习：计算

- Gain(D，根蒂)
- Gain(D，敲声)
- Gain(D，纹理)
- Gain(D，脐部)
- Gain(D，触感)

一个例子

表 4.1 西瓜数据集 2.0

编号	色泽	根蒂	敲声	纹理	脐部	触感	好瓜
1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	是
2	乌黑	蜷缩	沉闷	清晰	凹陷	硬滑	是
3	乌黑	蜷缩	浊响	清晰	凹陷	硬滑	是
4	青绿	蜷缩	沉闷	清晰	凹陷	硬滑	是
5	浅白	蜷缩	浊响	清晰	凹陷	硬滑	是
6	青绿	稍蜷	浊响	清晰	稍凹	软粘	是
7	乌黑	稍蜷	浊响	稍糊	稍凹	软粘	是
8	乌黑	稍蜷	浊响	清晰	稍凹	硬滑	是
9	乌黑	稍蜷	沉闷	稍糊	稍凹	硬滑	否
10	青绿	硬挺	清脆	清晰	平坦	软粘	否
11	浅白	硬挺	清脆	模糊	平坦	硬滑	否
12	浅白	蜷缩	浊响	模糊	平坦	软粘	否
13	青绿	稍蜷	浊响	稍糊	凹陷	硬滑	否
14	浅白	稍蜷	沉闷	稍糊	凹陷	硬滑	否
15	乌黑	稍蜷	浊响	清晰	稍凹	软粘	否
16	浅白	蜷缩	浊响	模糊	平坦	硬滑	否
17	青绿	蜷缩	沉闷	稍糊	稍凹	硬滑	否

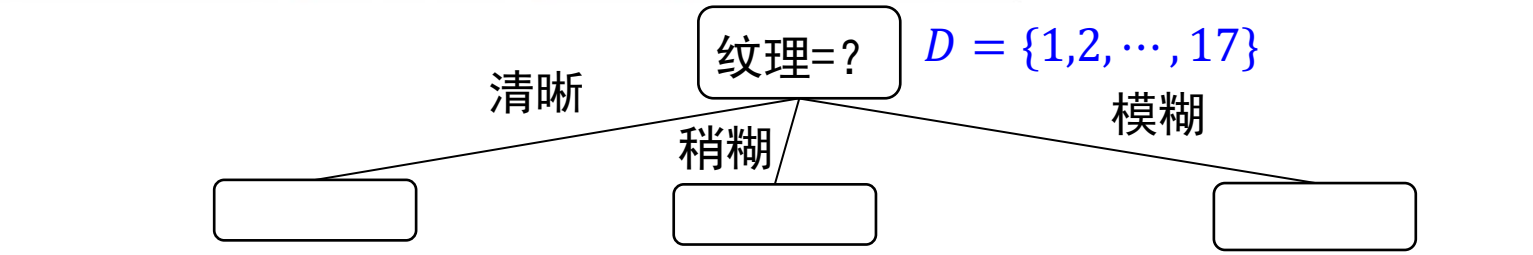
$\text{Gain}(D, \text{色泽}) = 0.109$
$\text{Gain}(D, \text{根蒂}) = 0.143$
$\text{Gain}(D, \text{敲声}) = 0.141$
$\text{Gain}(D, \text{纹理}) = 0.381$
$\text{Gain}(D, \text{脐部}) = 0.289$
$\text{Gain}(D, \text{触感}) = 0.006$

一个例子

表 4.1 西瓜数据集 2.0

编号	色泽	根蒂	敲声	纹理	脐部	触感	好瓜
1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	是
2	乌黑	蜷缩	沉闷	清晰	凹陷	硬滑	是
3	乌黑	蜷缩	浊响	清晰	凹陷	硬滑	是
4	青绿	蜷缩	沉闷	清晰	凹陷	硬滑	是
5	浅白	蜷缩	浊响	清晰	凹陷	硬滑	是
6	青绿	稍蜷	浊响	清晰	稍凹	软粘	是
7	乌黑	稍蜷	浊响	稍糊	稍凹	软粘	是
8	乌黑	稍蜷	浊响	清晰	稍凹	硬滑	是
9	乌黑	稍蜷	沉闷	稍糊	稍凹	硬滑	否
10	青绿	硬挺	清脆	清晰	平坦	软粘	否
11	浅白	硬挺	清脆	模糊	平坦	硬滑	否
12	浅白	蜷缩	浊响	模糊	平坦	软粘	否
13	青绿	稍蜷	浊响	稍糊	凹陷	硬滑	否
14	浅白	稍蜷	沉闷	稍糊	凹陷	硬滑	否
15	乌黑	稍蜷	浊响	清晰	稍凹	软粘	否
16	浅白	蜷缩	浊响	模糊	平坦	硬滑	否
17	青绿	蜷缩	沉闷	稍糊	稍凹	硬滑	否

$\text{Gain}(D, \text{色泽}) = 0.109$
$\text{Gain}(D, \text{根蒂}) = 0.143$
$\text{Gain}(D, \text{敲声}) = 0.141$
$\text{Gain}(D, \text{纹理}) = 0.381$
$\text{Gain}(D, \text{脐部}) = 0.289$
$\text{Gain}(D, \text{触感}) = 0.006$



选择划分属性准则

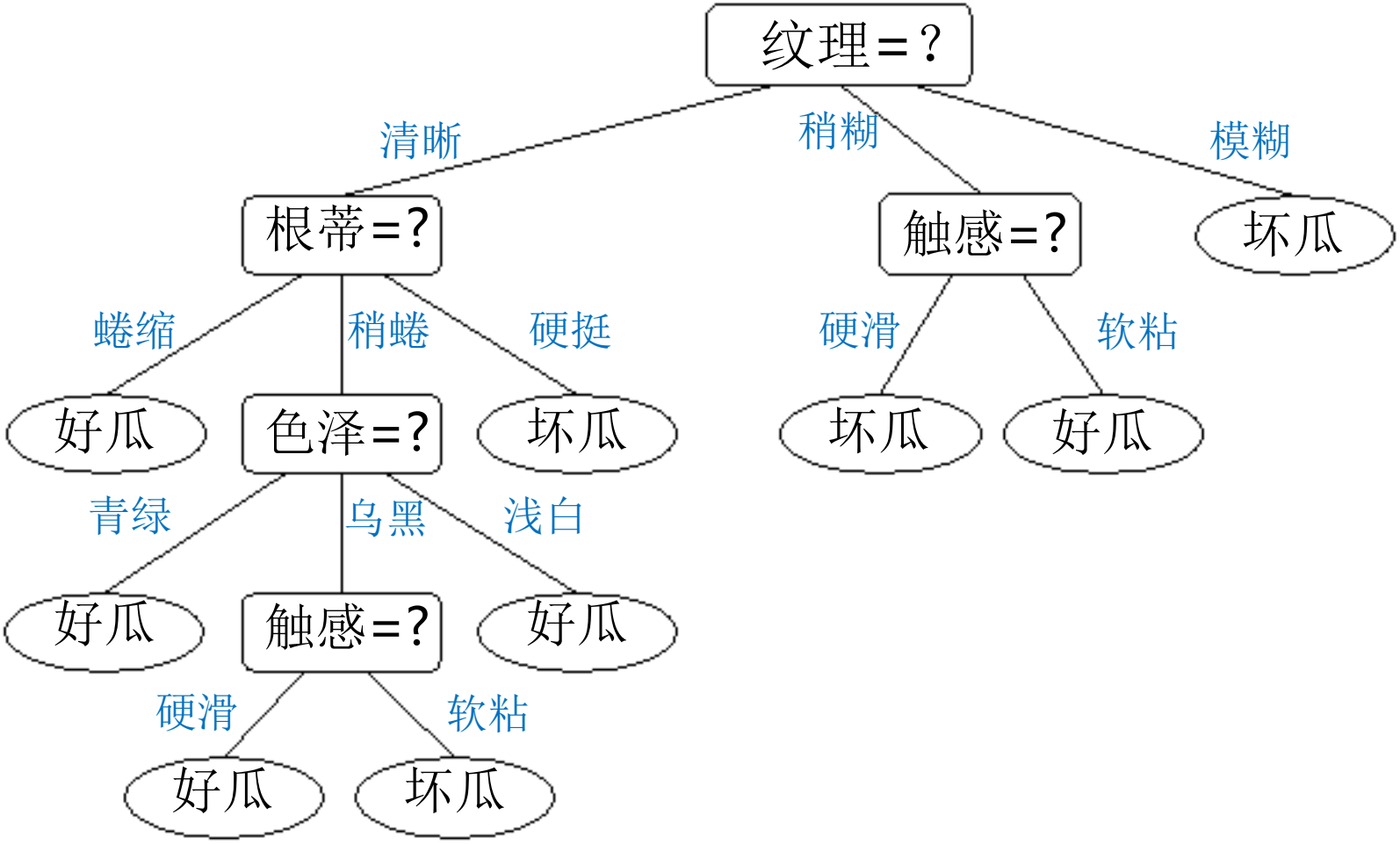
- 信息增益越大意味着用该属性来划分所获得的“纯度提升”越大
- 因此可以用信息增益来进行决策树的划分属性选择

$$\mathrm{Gain}(D,a) = H(D) - \sum_{v=1}^{V} \frac{|D^v|}{|D|} H(D^v)$$

- 著名的ID3决策树学习算法就是以信息增益为准则来选择划分属性

决策树

对每个分支结点做进一步划分，最终得到决策树



一个例子

D^1

编号	色泽	根蒂	敲声	纹理	脐部	触感	好瓜
11	浅白	硬挺	清脆	模糊	平坦	硬滑	否
12	浅白	蜷缩	浊响	模糊	平坦	软粘	否
16	浅白	蜷缩	浊响	模糊	平坦	硬滑	否
1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	是
2	乌黑	蜷缩	沉闷	清晰	凹陷	硬滑	是
3	乌黑	蜷缩	浊响	清晰	凹陷	硬滑	是
4	青绿	蜷缩	沉闷	清晰	凹陷	硬滑	是
5	浅白	蜷缩	浊响	清晰	凹陷	硬滑	是
6	青绿	稍蜷	浊响	清晰	稍凹	软粘	是
8	乌黑	稍蜷	浊响	清晰	稍凹	硬滑	是
15	乌黑	稍蜷	浊响	清晰	稍凹	软粘	否
10	青绿	硬挺	清脆	清晰	平坦	软粘	否
7	乌黑	稍蜷	浊响	稍糊	稍凹	软粘	是
9	乌黑	稍蜷	沉闷	稍糊	稍凹	硬滑	否
13	青绿	稍蜷	浊响	稍糊	凹陷	硬滑	否
14	浅白	稍蜷	沉闷	稍糊	凹陷	硬滑	否
17	青绿	蜷缩	沉闷	稍糊	稍凹	硬滑	否

增益率

- 信息增益的缺点：偏向于选择取值多的属性
  - 可能导致过拟合，比如把“编号”当做一个属性
- 增益率(gain ratio)

$$\text{Gain\_Ratio}(D,a)=\frac{\text{Gain}(D,a)}{\text{IV}(a)}$$

  - $\text{IV}(a)$表示属性a的固有值(intrinsic value)

$$\text{IV}(a)=-\sum_{v=1}^{V}\frac{|D^v|}{|D|}\log_2\left(\frac{|D^v|}{|D|}\right)$$

- 问题彻底解决了？

NO! 增益率偏向于选择取值少的属性

启发式：先从候选划分属性中找出信息增益高于平均水平的，再从中选取增益率最高的

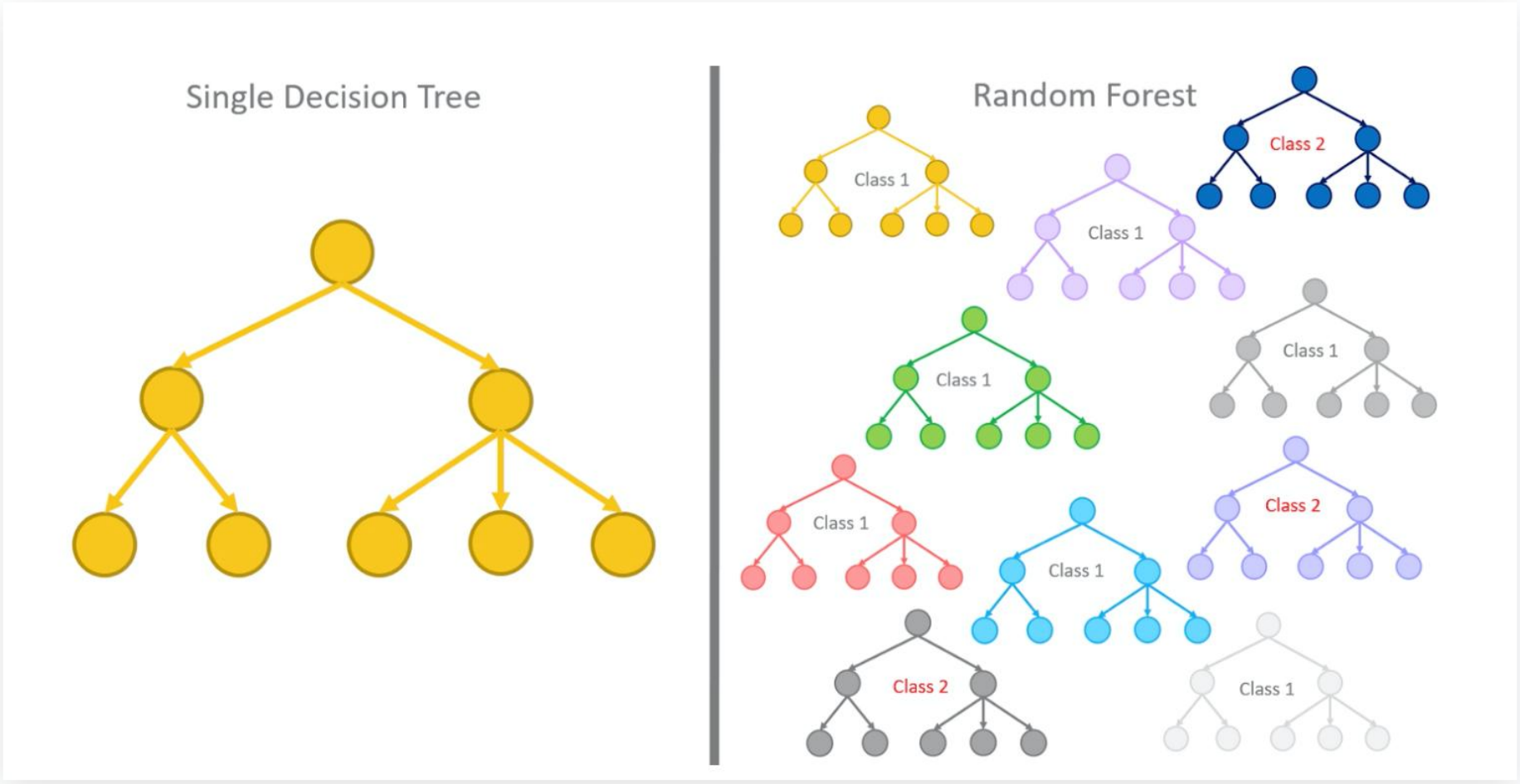
关于决策树

- 解决多分类问题的通用方法
  - 适用于属性值离散的数据
  - 也适用于属性值连续的数据（需对数据进行离散化）
  - 特点：对划分属性的选择很敏感！
- 优点：学习能力强（拟合能力强）、可应对缺失值
- 缺点：容易导致过拟合
  - 剪枝
    - ✓ 预剪枝
    - ✓ 后剪枝
- 一种强大的改进：随机森林
  - 下一小节……

本节课安排

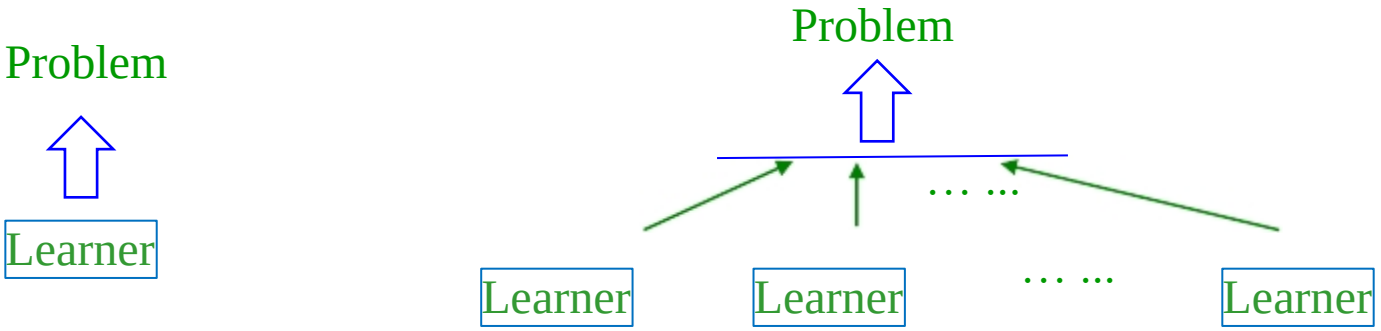
- 监督学习：
  - 线性回归
  - 对率回归
  - 支持向量机
  - 决策树
  - **随机森林**
  - 贝叶斯分类器
  - 感知机与神经网络

随机森林


Single Decision Tree
Random Forest
Class 1
Class 2
Class 1
Class 1
Class 1
Class 1
Class 2
Class 1
Class 2
Class 1

集成学习

集成学习(Ensemble Learning)：通过构建并结合多个学习器来完成学习任务



- 同质(homogeneous)集成：集成中只包含同种类型的“个体学习器”，相应的学习算法称为“基学习算法” (base learning algorithm)，个体学习器亦称“基学习器” (base learner)
- 异质(heterogeneous)集成：个体学习器由不同的学习算法生成，一般不叫“基学习算法”

常用结合方法

- 分类问题
  - h_i表示第i个基学习器, $i = 1, \cdots, T$
  - h_i^j表示h_i在类别标记c_j上的预测输出, $j = 1, \cdots, N$
    - $h_i^j(\mathbf{x})$表示h_i预测$\mathbf{x}$属于c_j的概率（或评分）
  - 投票法：
    - 绝对多数投票法

$$H(\mathbf{x}) = \begin{cases} c_j, & \text{if } \sum_{i=1}^{T} h_i^j(\mathbf{x}) > 0.5 \sum_{k=1}^{N} \sum_{i=1}^{T} h_i^k(\mathbf{x}) \\ \text{reject}, & \text{else.} \end{cases}$$

T

    - 相对多数投票法

$$H(\mathbf{x}) = c_{j^*}, \ j^* = \arg\max_j \sum_{i=1}^{T} h_i^j(\mathbf{x})$$

    - 加权投票法

$$H(\mathbf{x}) = c_{j^*}, \ j^* = \arg\max_j \sum_{i=1}^{T} w_i h_i^j(\mathbf{x})$$

- 回归问题
  - 简单平均法
  - 加权平均法

如何得到好的集成?

令个体学习器好而不同!

	测试例1	测试例2	测试例3
h_1	√	√	×
h_2	×	√	√
h_3	√	×	√
集成	√	√	√

(a) 集成提升性能

	测试例1	测试例2	测试例3
h_1	√	√	×
h_2	√	√	×
h_3	√	√	×
集成	√	√	×

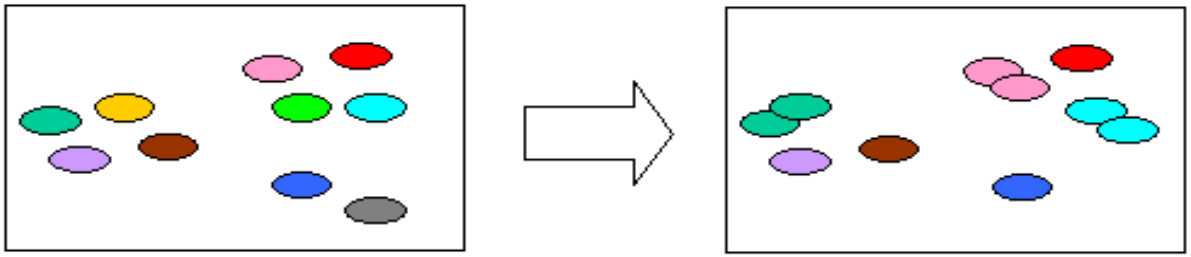
(b) 集成不起作用

	测试例1	测试例2	测试例3
h_1	√	×	×
h_2	×	√	×
h_3	×	×	√
集成	×	×	×

(c) 集成起负作用

Bagging

- Bagging：基于一种学习算法、一个数据集的集成方法
  - 有回放采样(bootstrap sampling)，也称可重复采样



  - ✓ 生成和原数据集一样大的新数据集
  - ✓ 约有36.8% 的样本不出现
  - 算法框架：通过数据扰动，获得具有一定差异性的基学习器

训练阶段：训练集$D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{m}$；基学习算法$\mathcal{L}$；
训练轮数目T
For t = 1, ……,T
对D进行有回放采样，得到D_t；
$h_t = \mathcal{L}(D_t)$；
End For
测试阶段（分类）：相对多数投票法

- 决策树

随机森林

- 随机森林(Random Forest，简称RF)是bagging的一个扩展变种
- 采样的随机性
- 属性选择的随机性

随机森林

- 随机森林(Random Forest)
  - 采样Bagging算法框架，基学习器为决策树
  - 增加基学习器的差异性：
    - 设当前节点的属性集合为$A = \{a_1, \cdots, a_d\}$。从A中随机选取包含k个属性的子集，记为A'
      - $k = \log_2 d$
    - 从A'中选择最优划分属性：
      - 推荐：先从候选划分属性中找出信息增益高于平均水平的，再从中选取增益率最高的

- 随机森林被誉为代表集成学习技术水平的方法

练习

小张喜欢吃西瓜，但时不时会买到不好吃的坏瓜。经过《人工智能导论》的学习后，小张决定使用机器学习方法改善自己买西瓜的策略。小张收集了 8 个西瓜数据样本：

编号	色泽青绿	根蒂蜷缩	敲声浊响	好瓜？
1	1	1	1	1
2	0	1	0	1
3	0	1	1	1
4	1	1	0	1
5	0	0	0	-1
6	1	0	0	-1
7	0	0	1	-1
8	1	0	1	-1

如果使用决策树对以上数据进行判别训练，请计算把每个属性作为划分属性的信息增益(Information Gain)。