



中国科学院自动化研究所  
INSTITUTE OF AUTOMATION  
CHINESE ACADEMY OF SCIENCES

# 机器学习

中国科学院自动化研究所

中科院大学人工智能学院

张文生

2019-12-05

# 机器学习与人工智能

机器学习来源于人工智能。

机器学习(Machine Learning)

Tom

如果一个系统能够通过执行某种过程而改进它的性能，这就是学习。

Herbert Simon

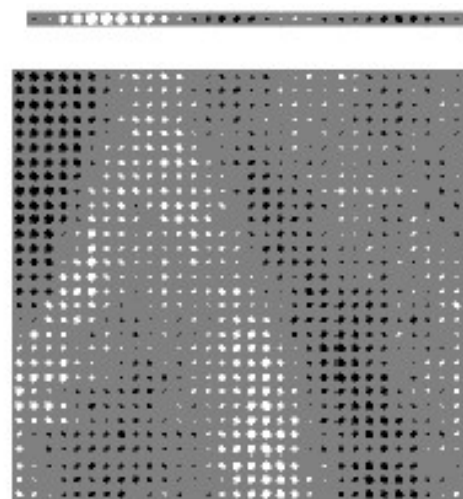
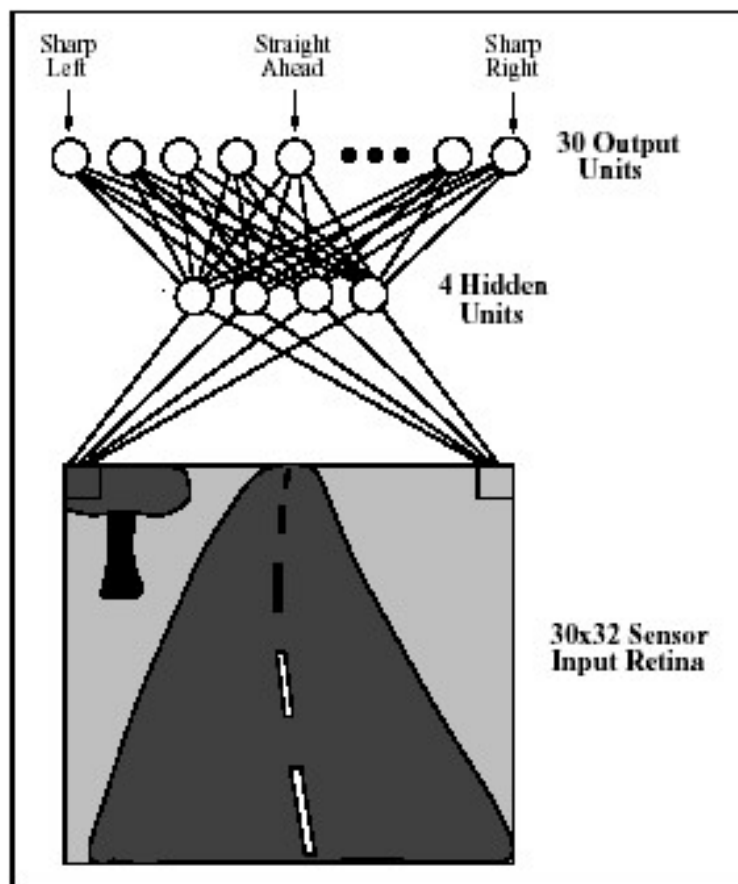
- **Common notions:**
  - **Input: data**
  - **variation**
  - **Improved performance (vague)**
    - **Measurable**

[illegible]

属性1	属性2	属性3	属性4	属性5	属性6	属性7	属性8	属性9	决策
1	1	1		1			1		1
1			1	1	1	1			1
	1			1		1	1	1	1
		1	1	1	1			1	1
1			1			1	1	1	0
		1			1	1	1	1	0
1	1	1			1			1	0
1	1	1	1			1			0

# 汽车自动驾驶问题





## 一些基本概念

- 问题空间
- 样本集
  - 数据收集
- 泛化
  - 把从一些例子中得到的经验应用到更广泛的情况中去。
  - 样本集的合理性
- 测试集



## 机器学习的研究趋势

- 尽管“学习机制”还是研究的动力，然而，“烦恼网络”的危机，使得更为重要的推动力来自“有效利用”信息
- 传统领域借用机器学习提高研究水平
- 应用驱动的机器学习方法层出不穷
- 基于机器学习的数据分析成为解决复杂问题的关键之一

## 传统领域借用机器学习提高研究水平

- 文本与图像占信息的多数，近些年来，模式识别最引人注目的是机器学习在这个领域扮演日益重要的角色
- 在文本分析与自然语言理解上，随着数据资源建设逐渐完善，关注的焦点是机器学习

## 符号机器学习

- **Solomonoff**关于文法归纳的研究是最早的符号机器学习(1959)
  - 1967年**Gold**证明，这是不可能的实现的
- **Samuel**将学习限制在结构化数据，学习演变为约简算法，这是现代符号机器学习的基础
  - 如果将每条规则理解为一个分类器，符号机器学习是**Hebb**路线

# 归纳学习与分析学习

- 归纳学习
  - 领导报告
  - 特殊—→一般
  - 归纳—演绎
  - **Bacon(1561-1626)**
  - 应用:
    - 数字识别/汉字识别

- 归纳学习方法

- 覆盖法

- 1980年, Michalski提出基于逻辑的**AQ11**, **AQ15**等算法, 称为**AQ**系列算法。

- 分治法

- 1966, **CLS**
    - 1986, **ID3**

- 分析学习

- 特点是应用领域知识驱动学习进程, 即学习进程是基于对领域知识的演绎

- 案例学习
    - 类比学习
    - 解释学习

## 要解决的问题

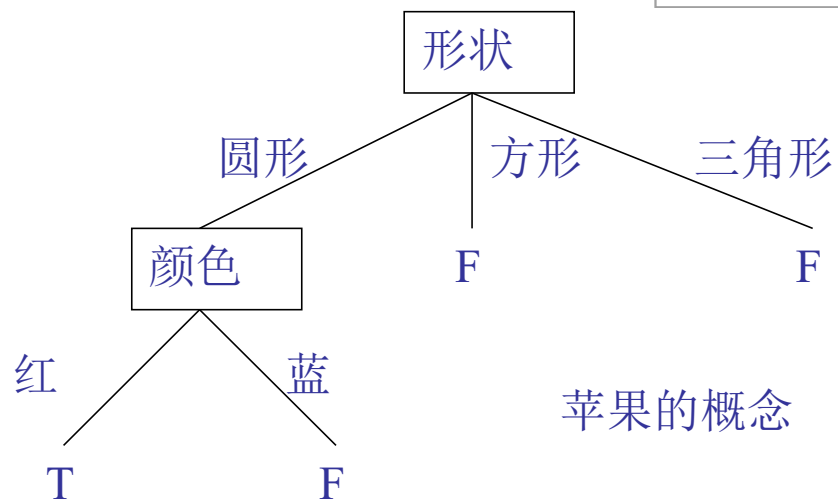
形状	颜色	尺寸	类别
圆	红	大	Y
圆	红	小	Y
圆	蓝	小	N
方	红	大	N
三角	红	小	N

正例

反例

# 决策树

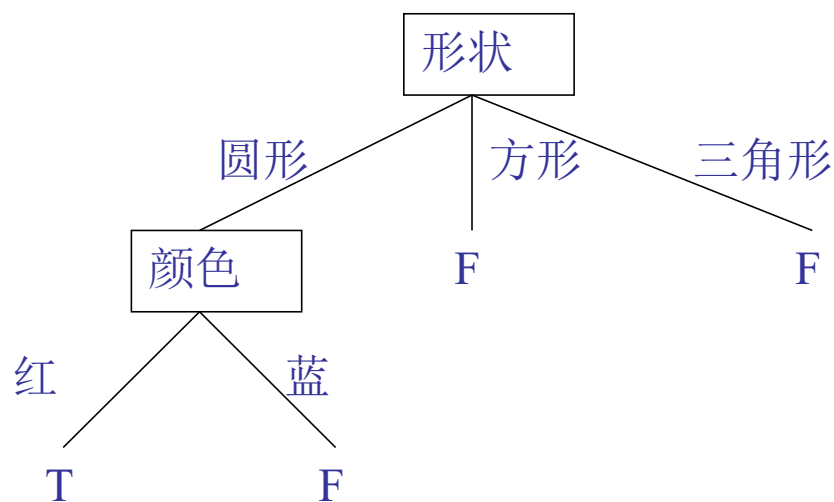
形状	颜色	尺寸	类别
圆	红	大	Y
圆	红	小	Y
圆	蓝	小	N
方	红	大	N
三角	红	小	N



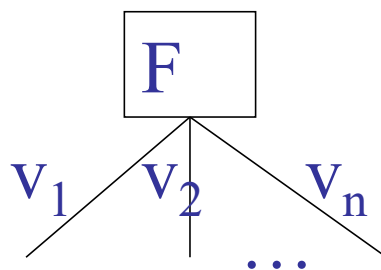


- 使用决策树:

- 决策树形成以后，对任何一个新的实例，可以应用决策树进行分类。
- 例: 蓝气球（大/圆/蓝）



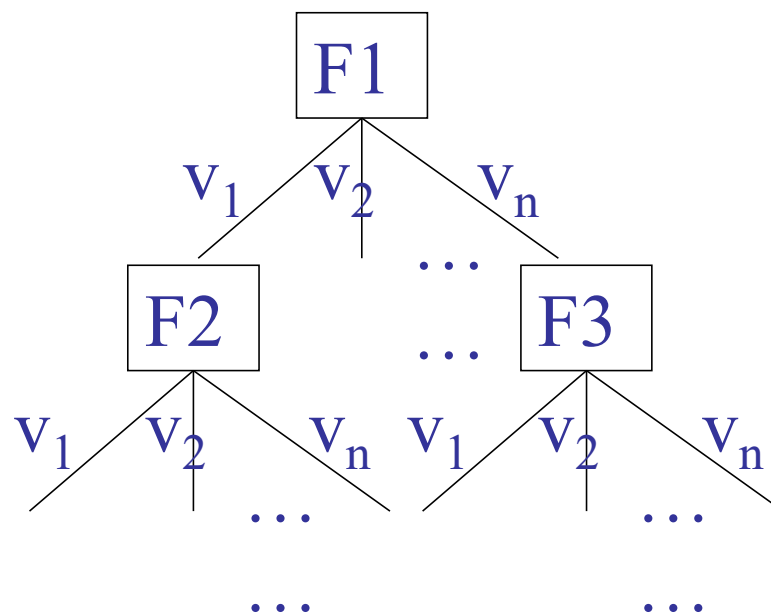
- 决策节点 (decision node) 的一般表示



$F$ : 属性

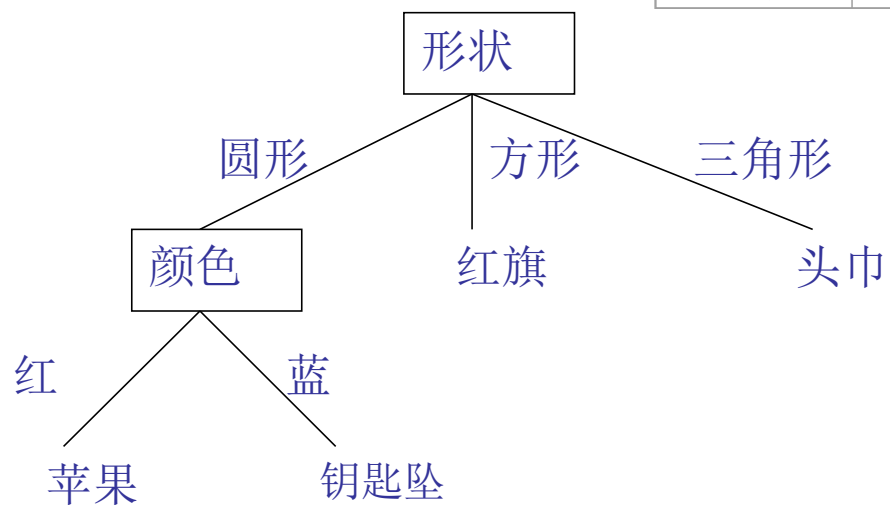
$\{v_1, v_2, \dots, v_n\}$ : 属性值

— 每一层的节点不一定完全一样



## — 一棵决策树表示多个概念

形状	颜色	尺寸	类别
圆	红	大	苹果
圆	红	小	苹果
圆	蓝	小	钥匙坠
方	红	大	红旗
三角	红	小	头巾

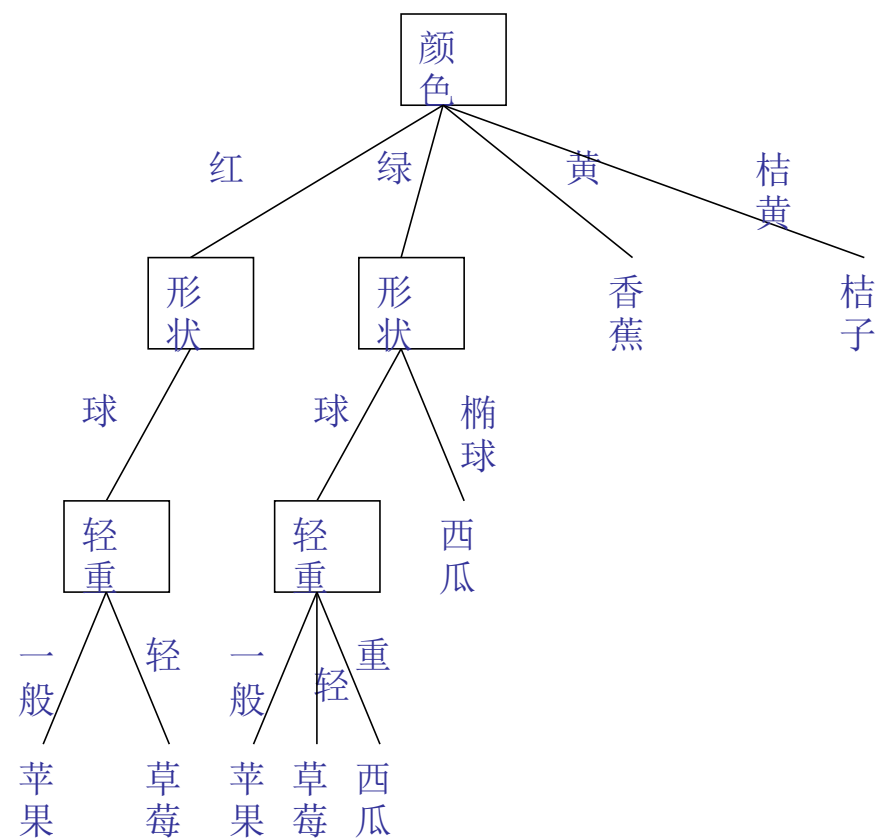


- 多个概念的学习算法  
  令 $C$ 为样本集

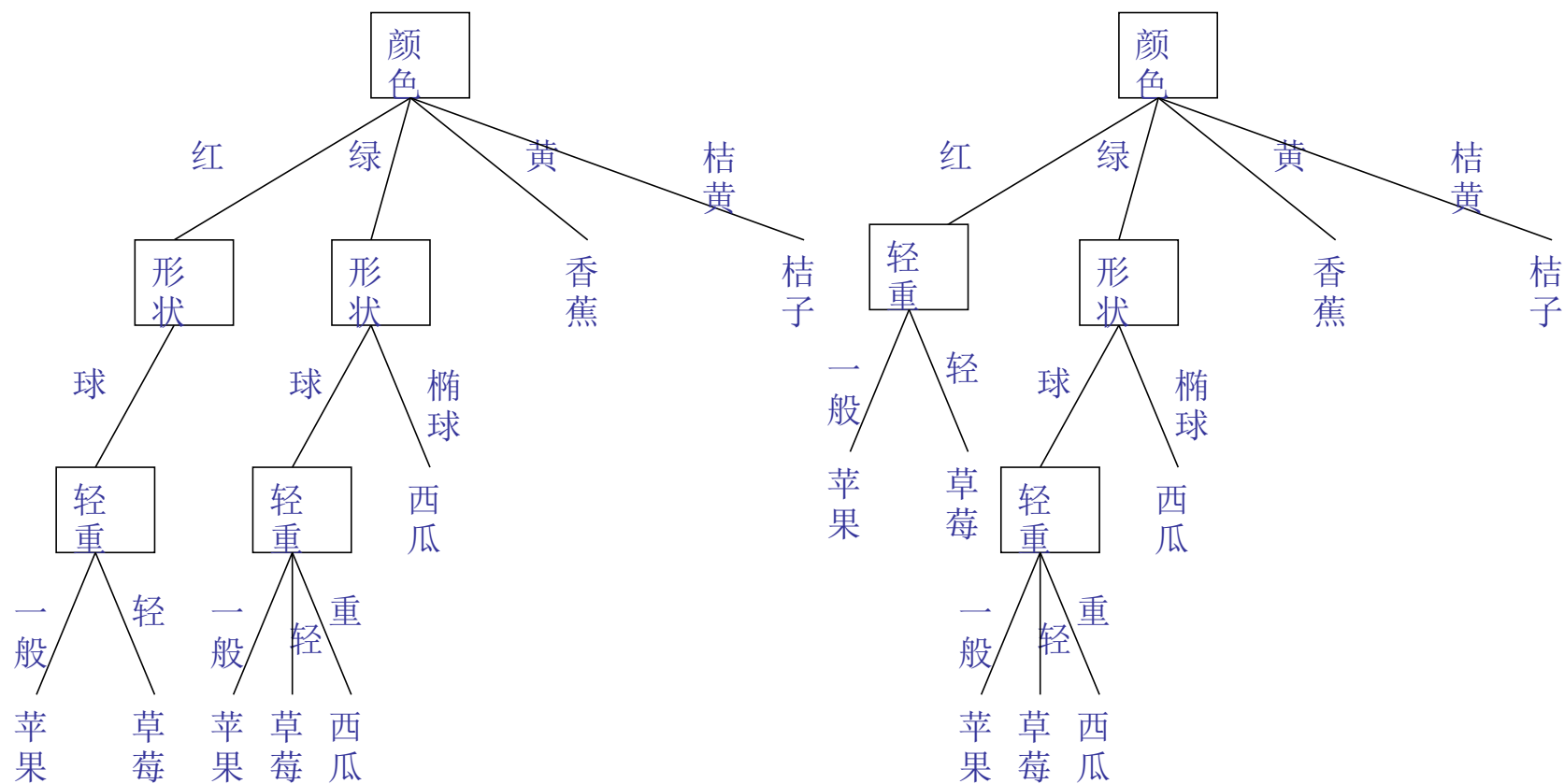
1. 如果所有在 $C$ 中的样本都为同一类，则建立一个终止节点，标记为此类，返回；
2. 否则，选择一个属性 $F$ ，建立决策节点；
3. 按照 $F$ 的值，将 $C$ 中的样本集分为一些子集， $C_1, C_2, \dots, C_n$ ；
4. 对每个 $C_i$ 递归调用以上过程。

# 例子

颜色	形状	轻重	概念
红	球	一般	苹果
绿	球	一般	苹果
黄	弯月	一般	香蕉
红	球	轻	草莓
绿	球	轻	草莓
绿	椭球	重	西瓜
绿	球	重	西瓜
桔黄	椭球	轻	桔子



# 决策树不唯一



## – ID3

- **Quinlan**

- Iterative Dichotomizer (version) 3

- **J.R. Quinlan. "*Induction of Decision Trees*". Machine Learning 1, pp. 81-106. Kluwer Academic Publishers. Boston, MA, USA, 1986.**

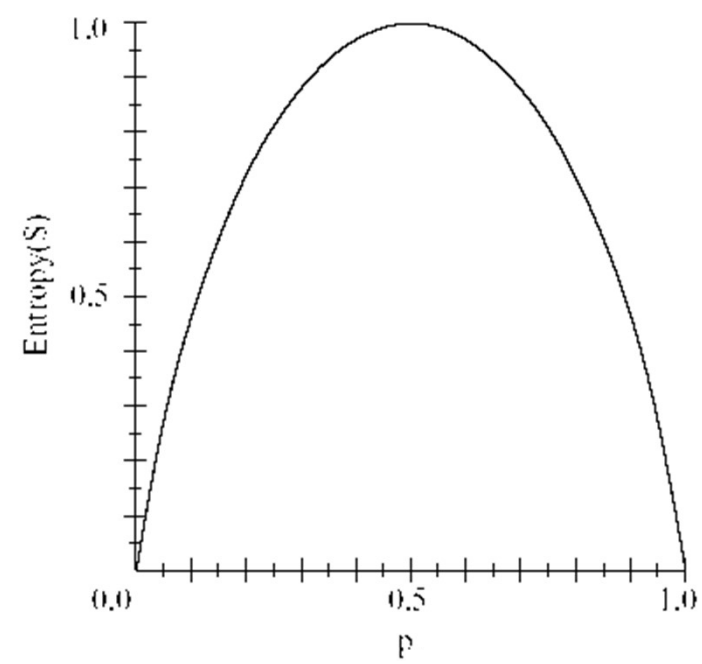
- 选择熵值最小的属性作为当前决策节点



# 熵(entropy)

- 物理系统的无序性
  - 法国物理学家克劳修斯用熵描述系统的无序性。
  - 无序性越高，则熵越大。
- 在信息论中，信息的不确定性对应于物理系统的无序性。
  - 通信

- $H = -\sum p_i \log_2 p_i \quad 1 \leq i \leq k$
- 例1:  $p_1 = p_2 = 1/2$   
 $H1 = -(1/2) * \log_2(1/2) - (1/2) * \log_2(1/2) = 1$
- 例2:  $p_1 = 1/4 \quad p_2 = 3/4$   
 $H2 = -(1/4) * \log_2(1/4) - (3/4) * \log_2(3/4) = 0.81$
- 例3:  $p_1 = 1 \quad p_2 = 0$   
 $H3 = -1 * \log_2 1 = 0$



## ID3 中的熵方法

对每一个特征，计算熵：

$$G(a, X) = \sum p(a(X,i), X) * H(a(X,i), \omega)$$

(对*i*求和，*i*是属性值)

$$H(X, \omega) = -\sum p(\omega_i, X) * \log_2 p(\omega_i, X)$$

(对*i*求和，*i*是类别数)

其中，**X**：样本集，**a**：属性。**ω**：类别。

**a(X,i)**：样本集**X**中，属性**a**取第*i*类值的样本构成的样本子集。

**p(a(X,i),X)**：样本子集**a(X,i)**的大小/样本集**X**的大小

**p(ω<sub>i</sub>,X)**：样本集**X**中，第*i*类样本的数目/样本集**X**的大小

- 思路：
  - 考察任一个属性，计算其熵值；
    - 对这个特定属性，考察根据其不同的属性值分成的若干子集；
    - 对任意子集，考察不同的类别，计算其熵值并求和；
  - 从所有属性中，选择熵值最小的属性作为当前决策节点。

颜色	形状	轻重	概念
红	球	一般	苹果
绿	球	一般	苹果
黄	弯月	一般	香蕉
红	球	轻	草莓
绿	球	轻	草莓
绿	椭球	重	西瓜
绿	球	重	西瓜
桔黄	椭球	轻	桔子

## 例子

1. “颜色” 属性:

$$p(\text{color}(X, \text{红}), X) = 2/8$$

$$H(\text{color}(X, \text{红}), \omega)$$

$$= -(1/2) * \log(1/2) - (1/2) * \log(1/2) = 1$$

$$p(\text{color}(X, \text{绿}), X) = 4/8$$

$$H(\text{color}(X, \text{绿}), \omega)$$

$$= -(1/4) * \log(1/4) - (1/4) * \log(1/4) - (1/2) * \log(1/2) = 3/2$$

$$p(\text{color}(X, \text{黄}), X) = 1/8$$

$$H(\text{color}(X, \text{黄}), \omega) = -1 * \log 1 = 0$$

$$p(\text{color}(X, \text{桔黄}), X) = 1/8$$

$$H(\text{color}(X, \text{桔黄}), \omega) = -1 * \log 1 = 0$$

$$G(\text{color}, X)$$

$$=(2/8)*1+(4/8)*(3/2)+(1/8)*0+(1/8)*0=1$$



## 2. “形状” 属性:

$$p(\text{sphere (X,球),X}) = 5/8$$

$$H(\text{sphere (X,球),}\omega)$$

$$=-(2/5)*\log(2/5)-(2/5)*\log(2/5)-(1/5)*\log(1/5)$$

$$=0.5931$$

$$p(\text{sphere (X,弯月),X}) = 1/8$$

$$H(\text{sphere (X,弯月),}\omega) = -1* \log 1 = 0$$

$$p(\text{sphere (X, 椭圆)}, X) = 2/8$$

$$H(\text{sphere (X, 椭圆)}, \omega)$$

$$= -(1/2) * \log(1/2) - (1/2) * \log(1/2) = 1$$

$$G(\text{sphere}, X)$$

$$= (5/8) * 0.5931 + (1/8) * 0 + (2/8) * 1$$

$$= 0.6207$$

3. “轻重” 属性:

$$p(\text{weight}(X, \text{一般}), X) = 3/8$$

$$H(\text{weight}(X, \text{一般}), \omega)$$

$$= -(2/3) * \log(2/3) - (1/3) * \log(1/3) = 0.9183$$

$$p(\text{weight}(X, \text{轻}), X) = 3/8$$

$$H(\text{weight}(X, \text{轻}), \omega)$$

$$= -(2/3) * \log(2/3) - (1/3) * \log(1/3) = 0.9183$$

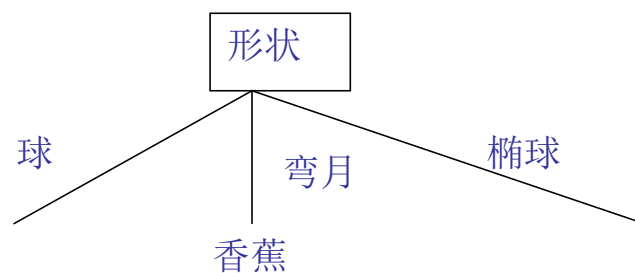
$$p(\text{weight}(X, \text{重}), X) = 2/8$$

$$H(\text{weight}(X, \text{重}), \omega) = -1 * \log 1 = 0$$

$$G(\text{weight}, X) = (3/8) * 0.9183 + (3/8) * 0.9183 + (2/8) * 0 = 0.6887$$

- $G(\text{color}, X) = 1$
- $G(\text{sphere}, X) = 0.6207$
- $G(\text{weight}, X) = 0.6887$
- 选择 “sphere” 为第一个决策节点。

颜色	形状	轻重	概念
红	球	一般	苹果
绿	球	一般	苹果
黄	弯月	一般	香蕉
红	球	轻	草莓
绿	球	轻	草莓
绿	椭球	重	西瓜
绿	球	重	西瓜
桔黄	椭球	轻	桔子



颜色	轻重	概念
红	一般	苹果
绿	一般	苹果
红	轻	草莓
绿	轻	草莓
绿	重	西瓜

1. “颜色” 属性:

$$p(\text{color}(X, \text{红}), X) = 2/5$$

$$H(\text{color}(X, \text{红}), \omega)$$

$$= -(1/2) * \log(1/2) - (1/2) * \log(1/2) = 1$$

$$p(\text{color}(X, \text{绿}), X) = 3/5$$

$$H(\text{color}(X, \text{绿}), \omega)$$

$$= -(1/3) * \log(1/3) - (1/3) * \log(1/3) - (1/3) * \log(1/3)$$

$$= 1.5851$$

$$G(\text{color}, X) = (2/5) * 1 + (3/5) * 1.5851 = 1.3511$$

2. “轻重” 属性:

$$p(\text{color}(X, \text{一般}), X) = 2/5$$

$$H(\text{color}(X, \text{一般}), \omega) = 0$$

$$p(\text{color}(X, \text{轻}), X) = 2/5$$

$$H(\text{color}(X, \text{轻}), \omega) = 0$$

$$p(\text{color}(X, \text{重}), X) = 1/5$$

$$H(\text{color}(X, \text{重}), \omega) = 0$$

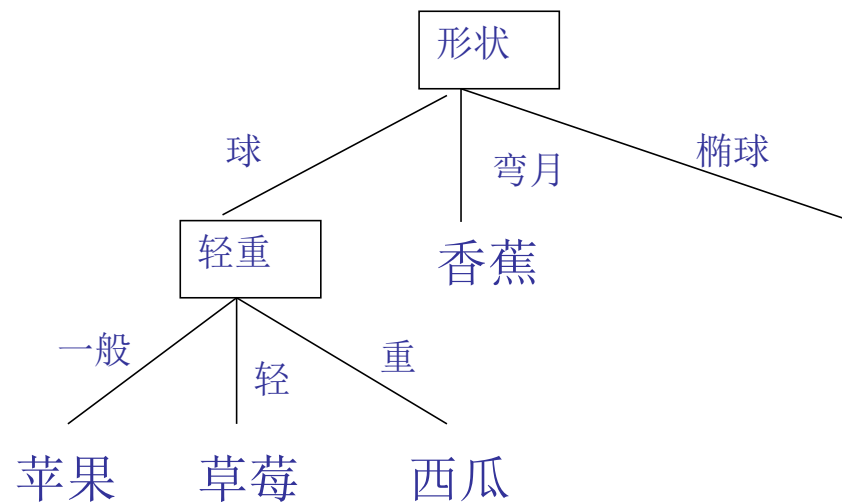
$$G(\text{weight}, X) = 0$$



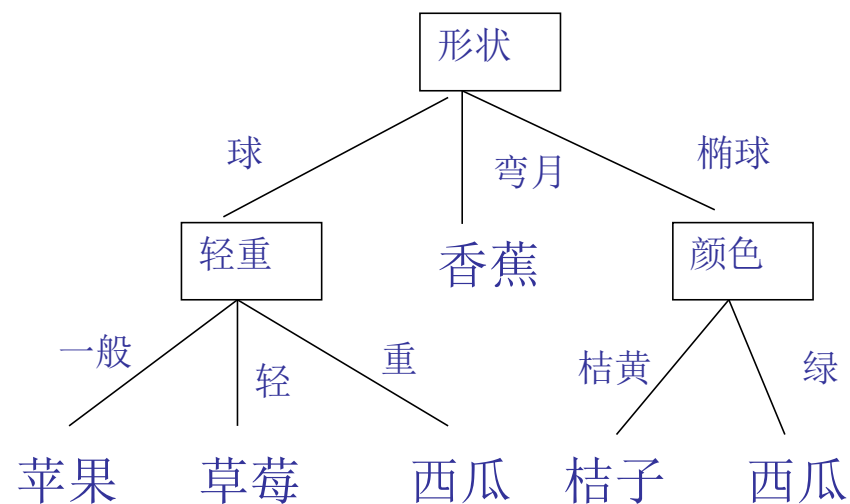
$G(\text{color}, X) = 1.3511$

$G(\text{weight}, X) = 0$

选择weight.



颜色	形状	轻重	概念
红	球	一般	苹果
绿	球	一般	苹果
黄	弯月	一般	香蕉
红	球	轻	草莓
绿	球	轻	草莓
绿	椭球	重	西瓜
绿	球	重	西瓜
桔黄	椭球	轻	桔子



- **1994年Quinlan提出C4.5， 解决了**
  - (1)将决策树转换为等价的规则，
  - (2)解决了连续量的数据的学习问题.
- **ID3被整合在SQL Server2000中；**
- 速度较快
- 倾向于选择取值多的属性.
- 增量算法

# 统计机器学习

- 1943, McCulloch, Pitts, MP模型
- 1959, Rosenblatt, Perceptron
  - 多层Perceptron
- 1969, Minsky, << Perceptron >>
- 1986, Rumelhart, Hinton, PDP, BP
- 1971, Vapnik, Chervonenkis, VC维
- 1995, 统计学习理论

# 解释学习

- **EBL (Explanation-based learning)**
- **1972年，Nilsson设计了机器人规划系统，叫STRIPS。**
  - 形成宏操作技术。
  - 一旦一个机器人进行了一个规划，到达了目标，就通过概括把这个规划转换成一个宏操作保留起来，以备以后遇到类似问题，可以直接使用。
  - 但这种概括不是简单的以变量代替常量，而是分析规划中那些步骤是必要的，根据这些必要的步骤形成宏操作。

- 目的：将不可用或不易用的知识转换为可用或易用的知识。
- 可用或易用的标准：可操作性。
- 可操作性
  - 树立长套
  - “树立长套”就是在每次上手时，都打自己最长的那门花色，直到对方不再有这门花色。
  - “上手”就是轮到自己出牌。
  - 在每次轮到自己出牌时，都打自己最长的那门花色，直到对方不再有这门花色。

- 几个概念：

- 目标概念 (**Goal Concept**)：某概念的高级形式的描述，通常是不可操作的
- 训练例子 (**Training example**)：以低级形式描述的某概念的特例，通常是可操作的
- 领域知识 (**Domain Theory**)：推理规则和事实集，是学习的依据，用以证明训练例子符合目标概念的描述

- 解释学习的一般算法：

1. 给定一个知识库；
2. 给定一个目标状态 $\mathbf{G}$ ；
3. 输入一个实例 $\mathbf{e}$ ；
4. 使用知识库中的知识，证明 $\mathbf{e}$ 是 $\mathbf{G}$ 的一个实例；（解释）
5. 对上一步获得的解释进行推广，得到一条更一般的知识；（泛化）
6. 将泛化后的知识存入数据库；



- 解释有两种方法：
  - Nilsson的目标回归法(goal regression)，曾用于描述机器人规划程序STRIPS的操作原理；
  - 基于解释的特化（EBS），用于PRODIGY系统；PRODIGY是个解题系统，能从专家解题的过程中学到专家解题的技巧。从本质上看，与目标回归法是一致的，都是把待解目标一步步回归为知识库中的基本单元。

## 解释的目标回归法

设目标状态为 $G_1 \wedge G_2 \wedge \dots \wedge G_n$ ;

1 建立目标集合 $G=\{G_1, G_2, \dots, G_n\}$ ;

2 将实例送入知识库;

3 执行以下循环:

3.1 如果知识库中的信息使 $G$ 中的某些元素满足, 则删除这些元素;

3.2 如果 $G$ 为空, 成功跳出循环; 否则:

3.3 取 $G$ 中的第一个目标 $G_i$ , 用知识库中的产生式规则 $R_j$ 实行一步回归 (即向后推理);

3.4 从 $G$ 中删除 $G_i$ , 将 $R_j$ 的条件部分 $G_{i1}, G_{i2}, \dots, G_{im}$ 放入 $G$ ;

3.5 转3.1。

- 学习 “杯子” 的概念

领域知识:

$\text{stable}(x) \wedge \text{liftable}(x) \wedge \text{drinkfrom}(x) \rightarrow \text{cup}(x)$

$\text{bottom}(x,y) \wedge \text{flat}(y) \rightarrow \text{stable}(x)$

$\text{light-weight}(x) \wedge \text{graspable}(x) \rightarrow \text{liftable}(x)$

$\text{has}(x, y) \wedge \text{concavity}(y) \wedge \text{upward-pointing}(y) \rightarrow \text{drinkfrom}(x)$

$\text{small}(x) \wedge \text{madefrom}(x, \text{plastic}) \rightarrow \text{light-weight}(x)$

$\text{big}(x) \wedge \text{full-of}(x, \text{air}) \rightarrow \text{light-weight}(x)$

$\text{has}(x, y) \wedge \text{handle}(y) \rightarrow \text{graspable}(x)$

- 训练例子:

`small(obj), madefrom(obj, plastic),`

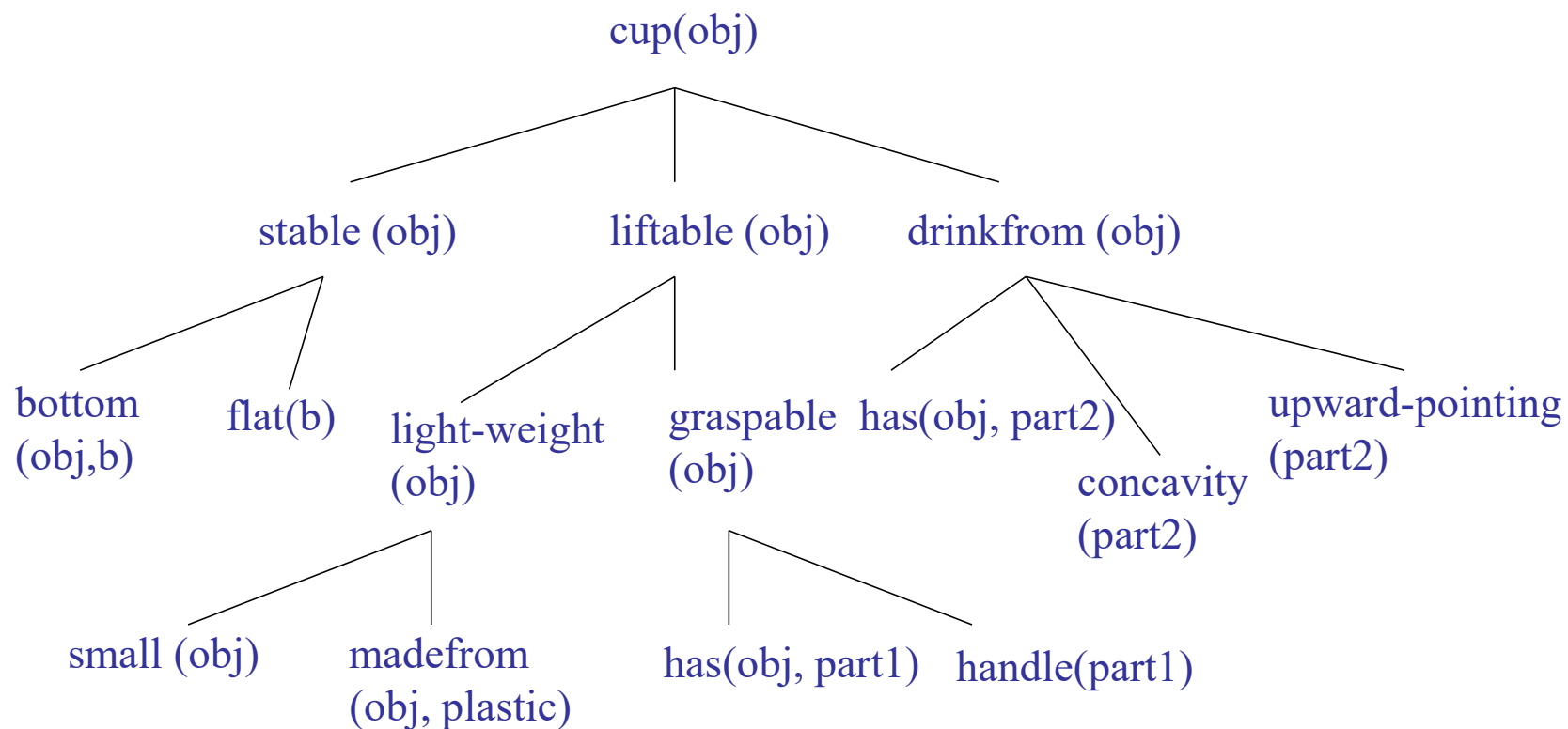
`has(obj, part1), handle(part1),`

`has(obj, part2), concavity(part2),`

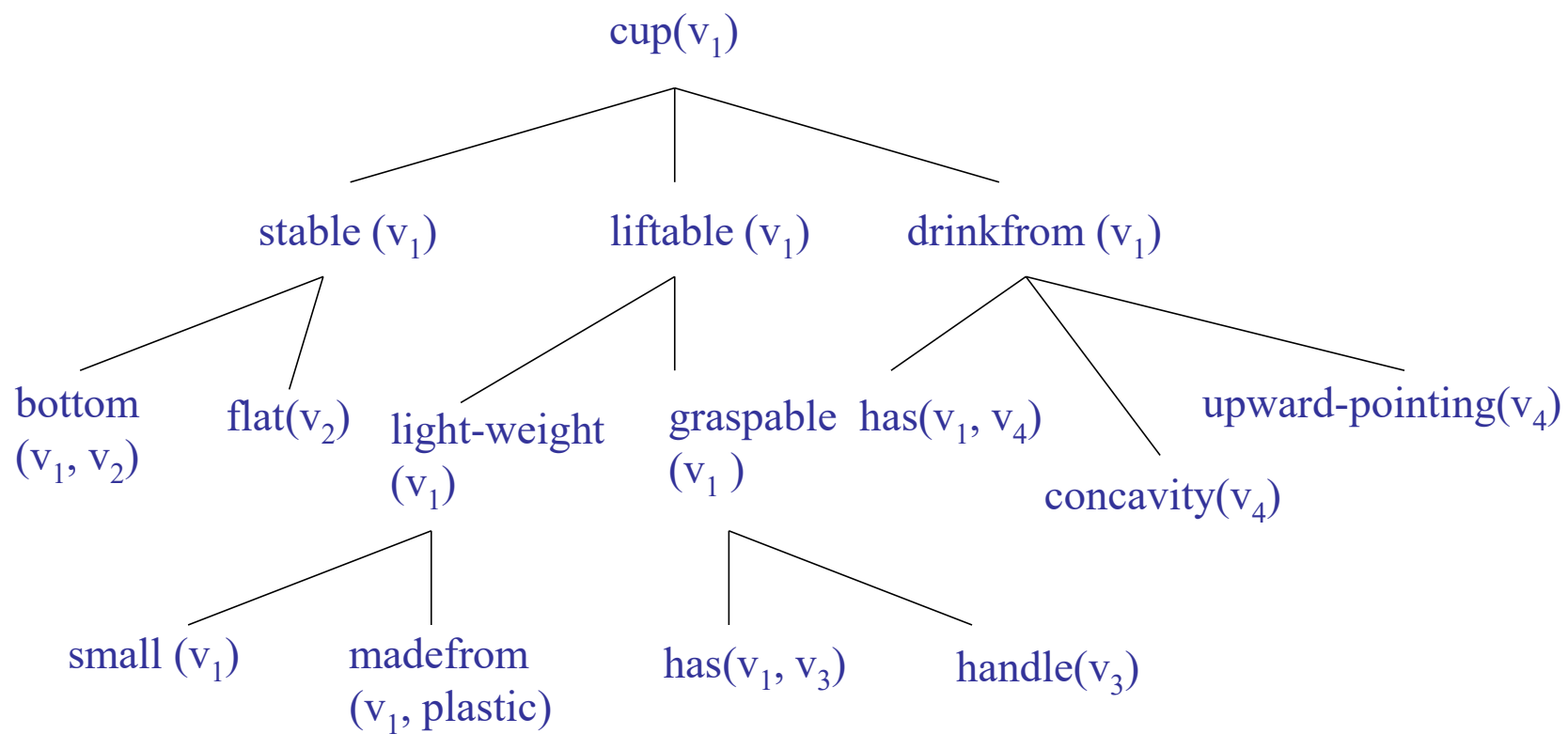
`upward-pointing(part2), bottom(obj, b),`

`flat(b), color(obj, red)`

## 训练例子是cup的一个实例



# 泛化



- 总结为规则:

**if**  $\text{small}(v_1) \wedge \text{madefrom}(v_1, \text{plastic}) \wedge \text{has}(v_1, v_3) \wedge \text{handle}(v_3) \wedge \text{has}(v_1, v_4) \wedge$   
 $\text{concavity}(v_4) \wedge \text{upward-pointing}(v_4) \wedge \text{bottom}(v_1, v_2) \wedge \text{flat}(v_2)$

**then**  $\text{cup}(v_1)$



**中国科学院自动化研究所**  
**INSTITUTE OF AUTOMATION**  
**CHINESE ACADEMY OF SCIENCES**

**感谢同学们光临！**

**敬请交流和指正**