

## 十五、规则学习

主讲教师：俞扬

# 大纲

---

- 基本概念
- 序贯覆盖
- 剪枝优化
- 一阶规则学习
- 归纳逻辑程序设计

# 基本概念

---

## □ 机器学习里的规则:

若....., 则.....

- 回归

若  $\hat{h}(\mathbf{x}) = \hat{y}$        $y = \hat{y}$ , 则

- 分类

若  $h(\mathbf{x}) > 0$        $class, \Rightarrow 1$     若  $h(\mathbf{x}) \leq 0$        $class, \Rightarrow -1$

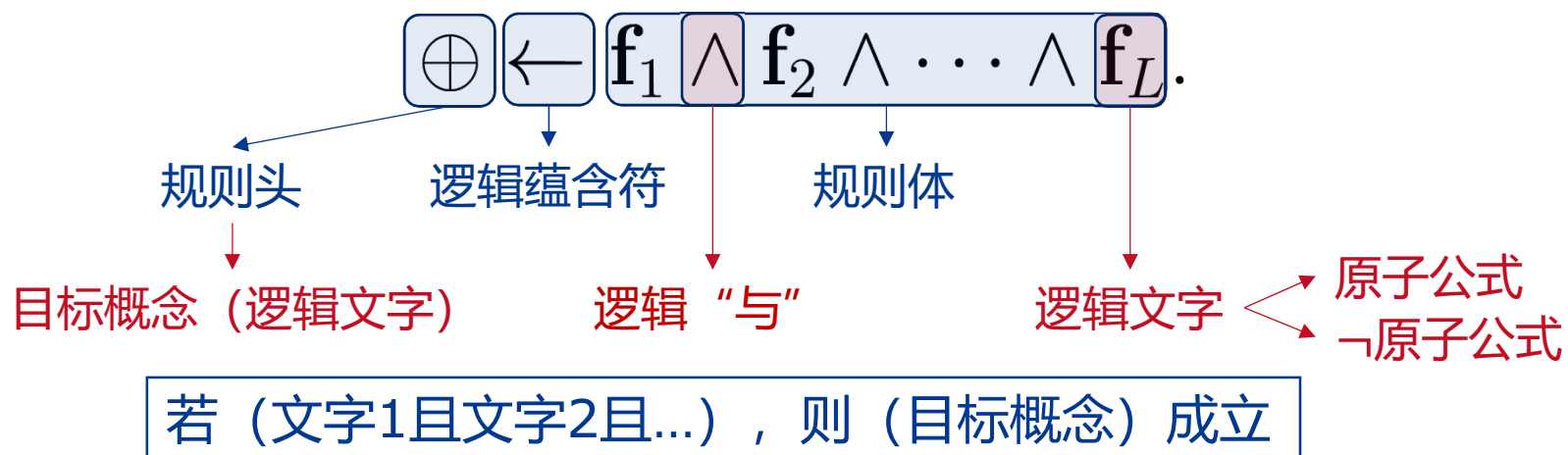
- 聚类

若  $\forall j \neq i, dist(\mathbf{x}, c_i) \leq dist(\mathbf{x}, c_j)$        $C(\mathbf{x}) = c_i$

- .....

# 基本概念

## □ 逻辑规则



## □ 规则集

规则 1: 好瓜( $X$ )  $\leftarrow$  根蒂( $X$ , 蜷缩)  $\wedge$  脐部( $X$ , 凹陷).

规则 2:  $\neg$ 好瓜( $X$ )  $\leftarrow$  纹理( $X$ , 模糊).

- 充分性与必要性
- 冲突消解: 顺序规则、缺省规则、元规则

# 基本概念

## □ 命题逻辑 $\rightarrow$ 命题规则

- 原子命题:  $A, B, C, \dots$
- 逻辑连词:  $\leftarrow, \rightarrow, \leftrightarrow, \wedge, \vee, \neg, \dots$

$R \leftarrow P \wedge Q.$

好瓜  $\leftarrow$  (根蒂=蜷缩)  $\wedge$  (脐部=凹陷).

## □ 一阶逻辑 $\rightarrow$ 一阶规则

- 常量:  $a, b, c, \dots, 1, 2, 3, \dots$
- 变量:  $A, B, C, \dots$
- ( $n$ 元) 谓词/函数:  $p/n, f/n, \dots$
- 项: 常量|变量|函数/谓词(项<sub>1</sub>, 项<sub>2</sub>, ...)
- 原子公式: 函数/谓词(项<sub>1</sub>, 项<sub>2</sub>, ...)  
父亲( $X, Y$ ), 自然数(39), 偶数(后继(1)), ...
- 逻辑连词
- 逻辑量词:  $\exists, \forall$

$\forall X (p(f(X)) \leftarrow f(X)).$

$\forall X (\text{自然数}(\text{后继}(X)) \leftarrow \text{自然数}(X)).$

# 大纲

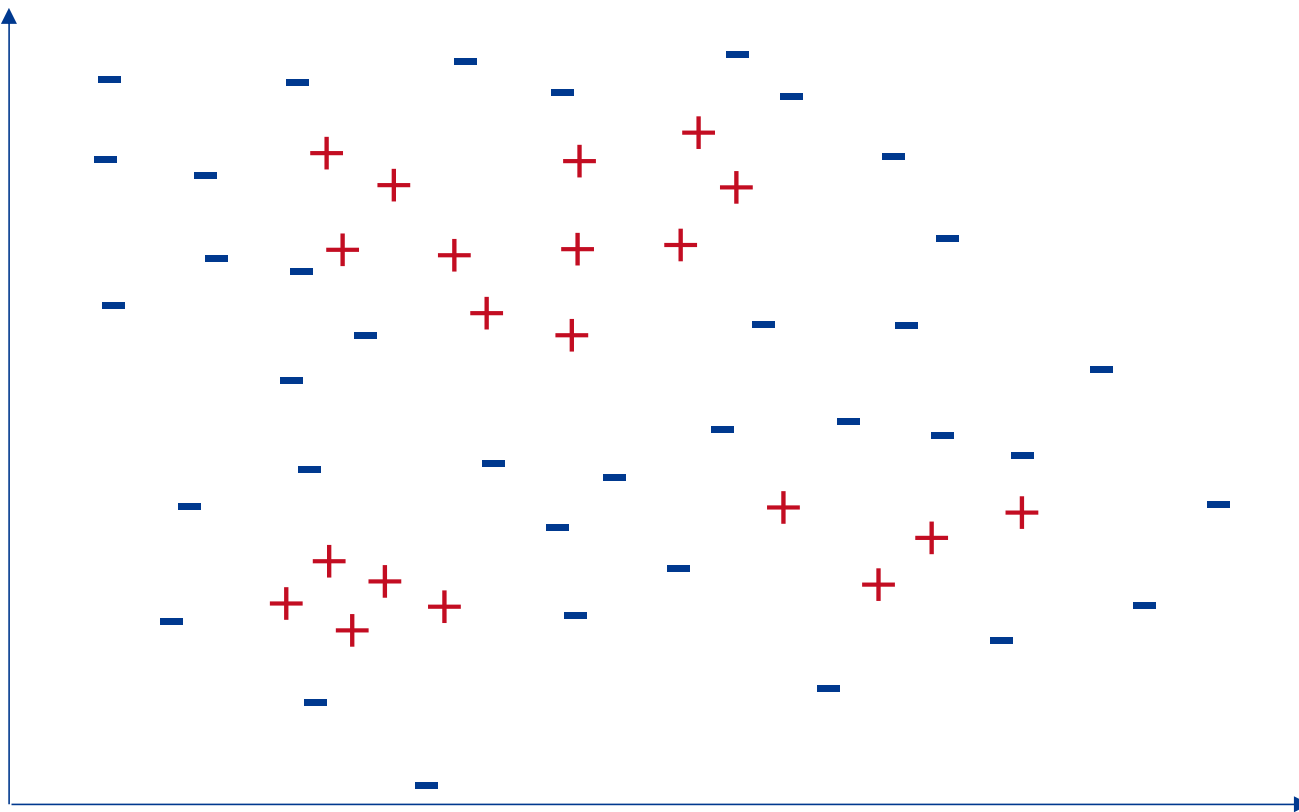
---

- 基本概念
- 序贯覆盖
- 剪枝优化
- 一阶规则学习
- 归纳逻辑程序设计

# 序贯覆盖

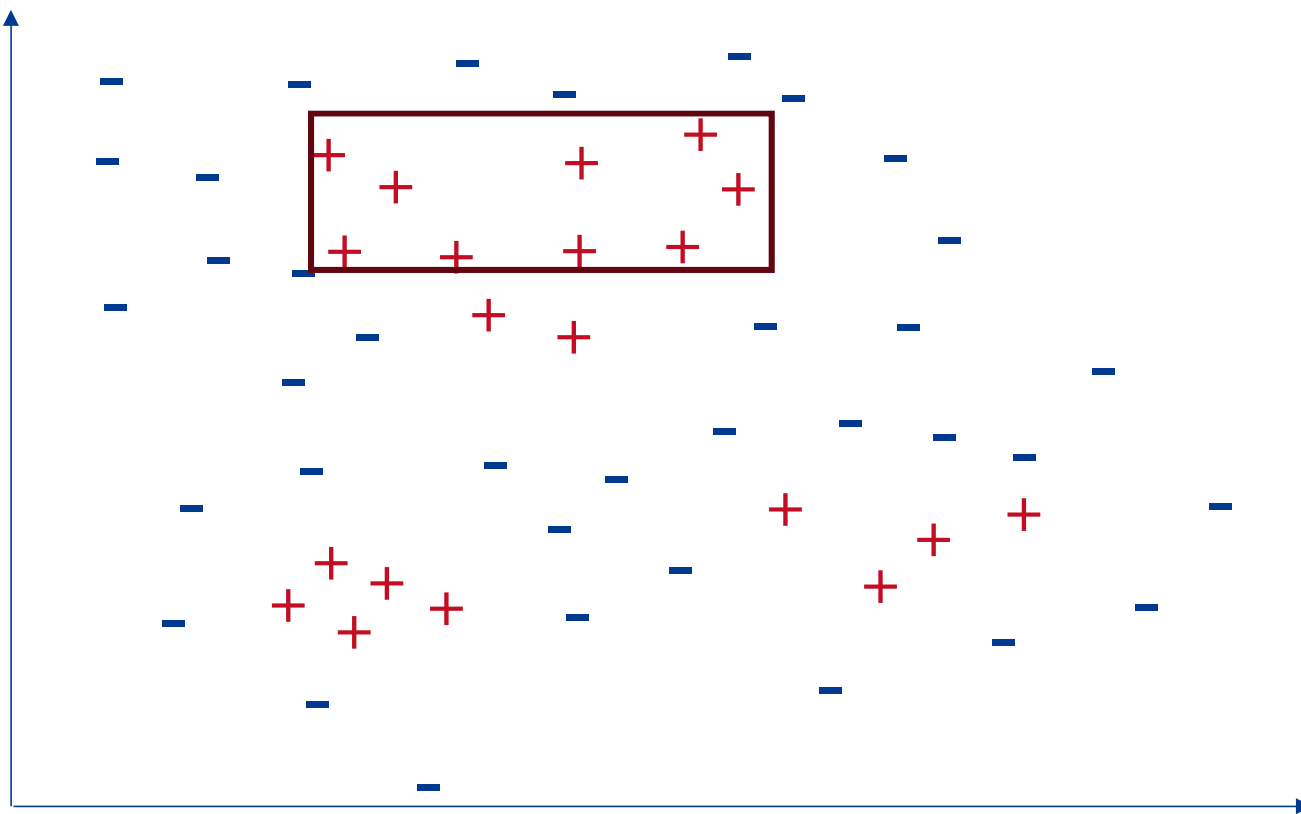
---

## □ 分治策略



# 序贯覆盖

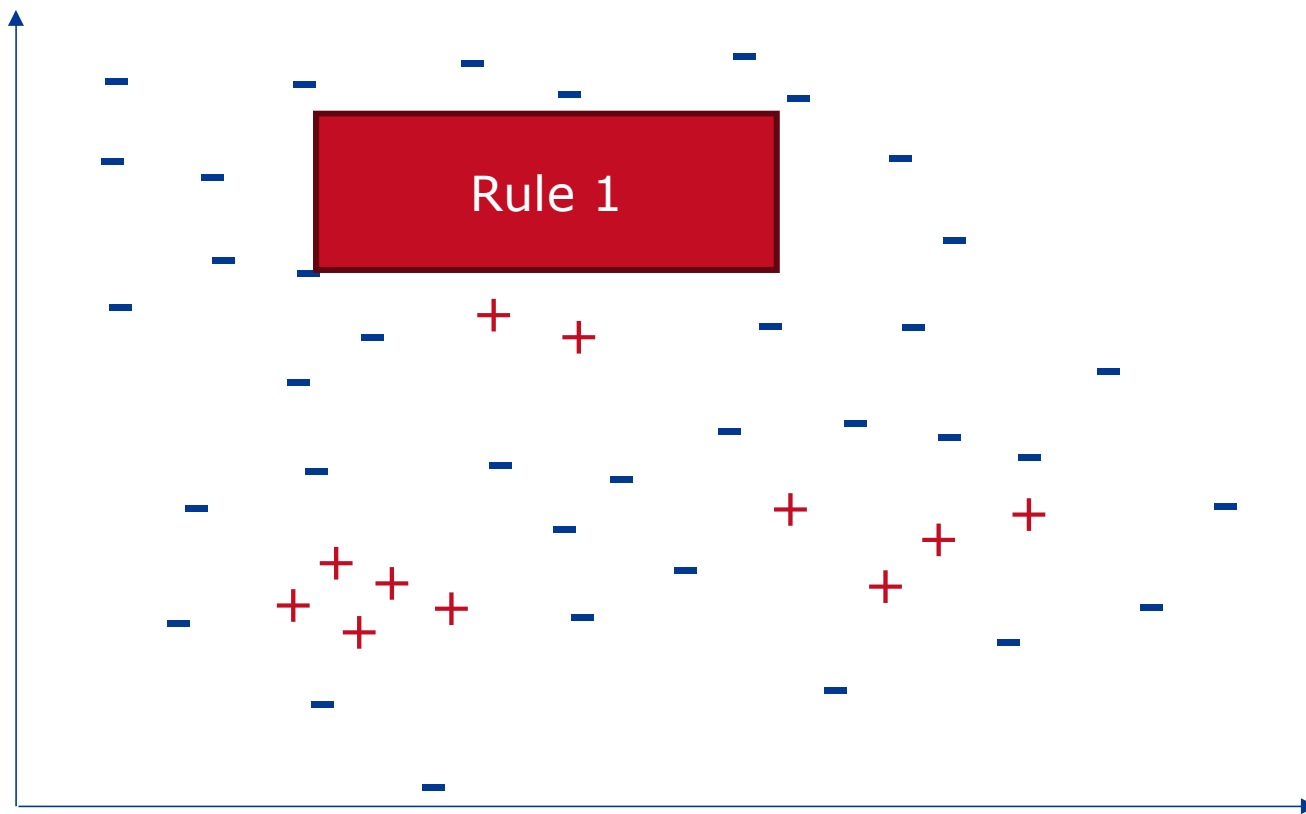
---





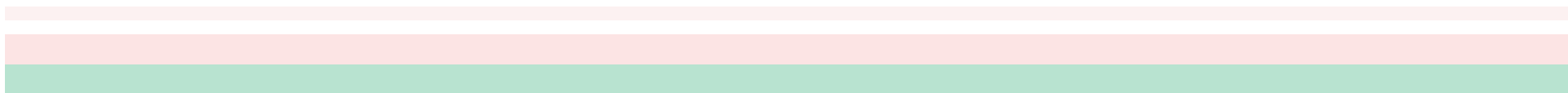
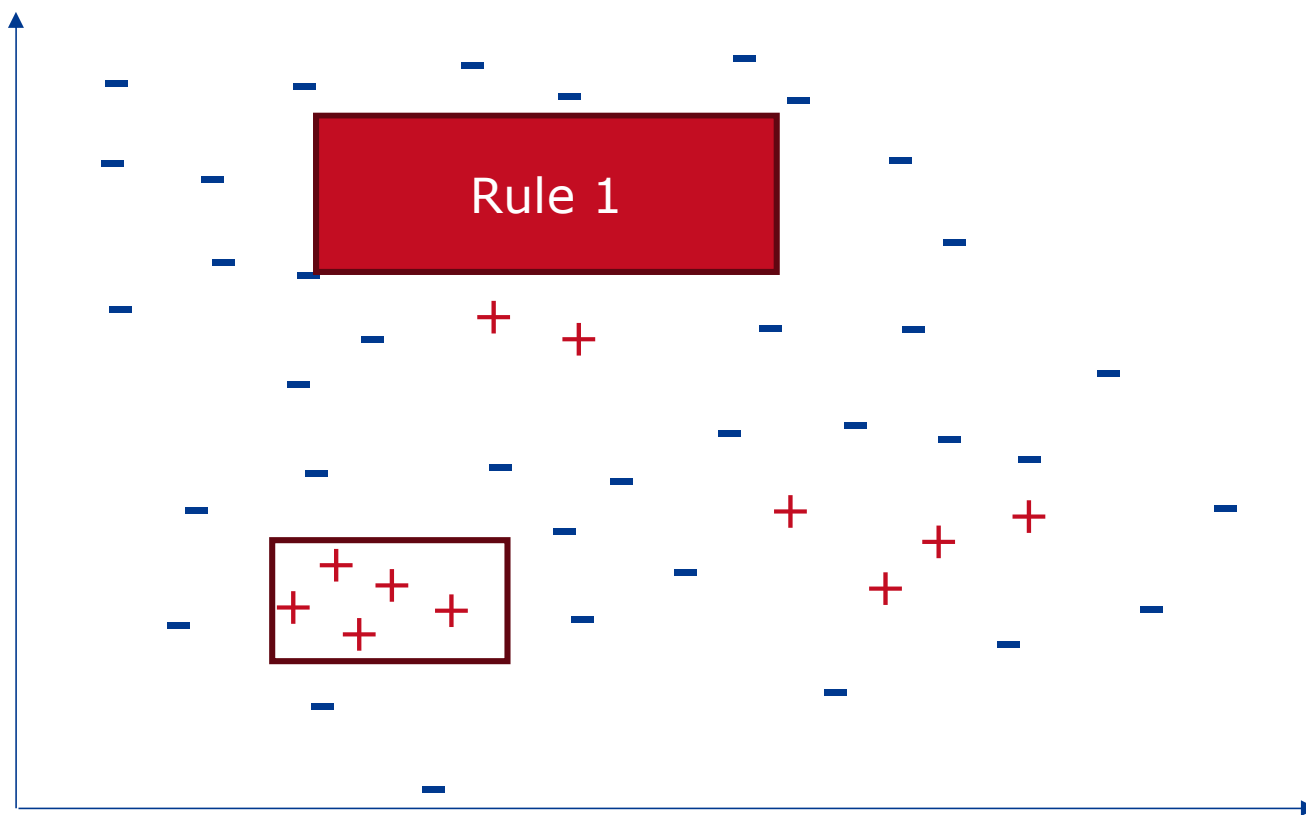
# 序贯覆盖

---



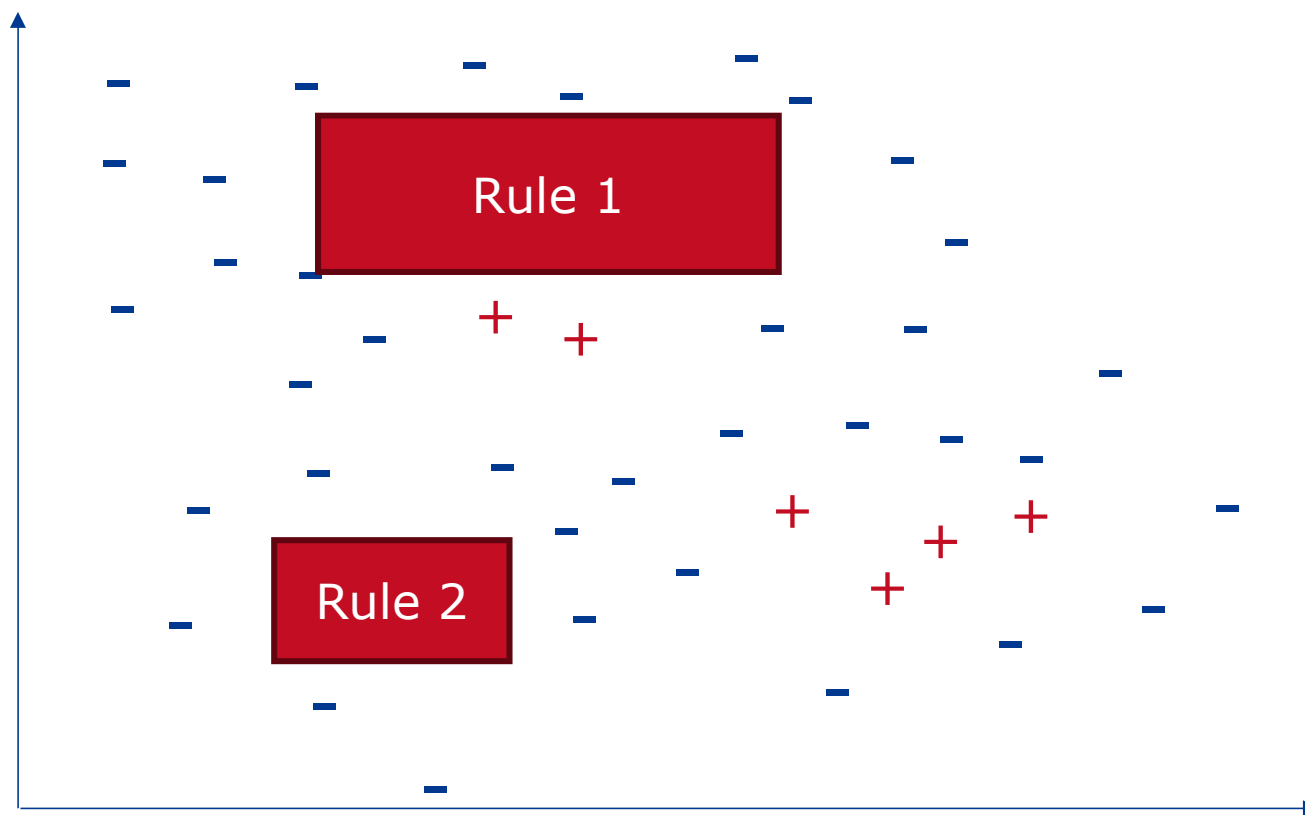
# 序贯覆盖

---



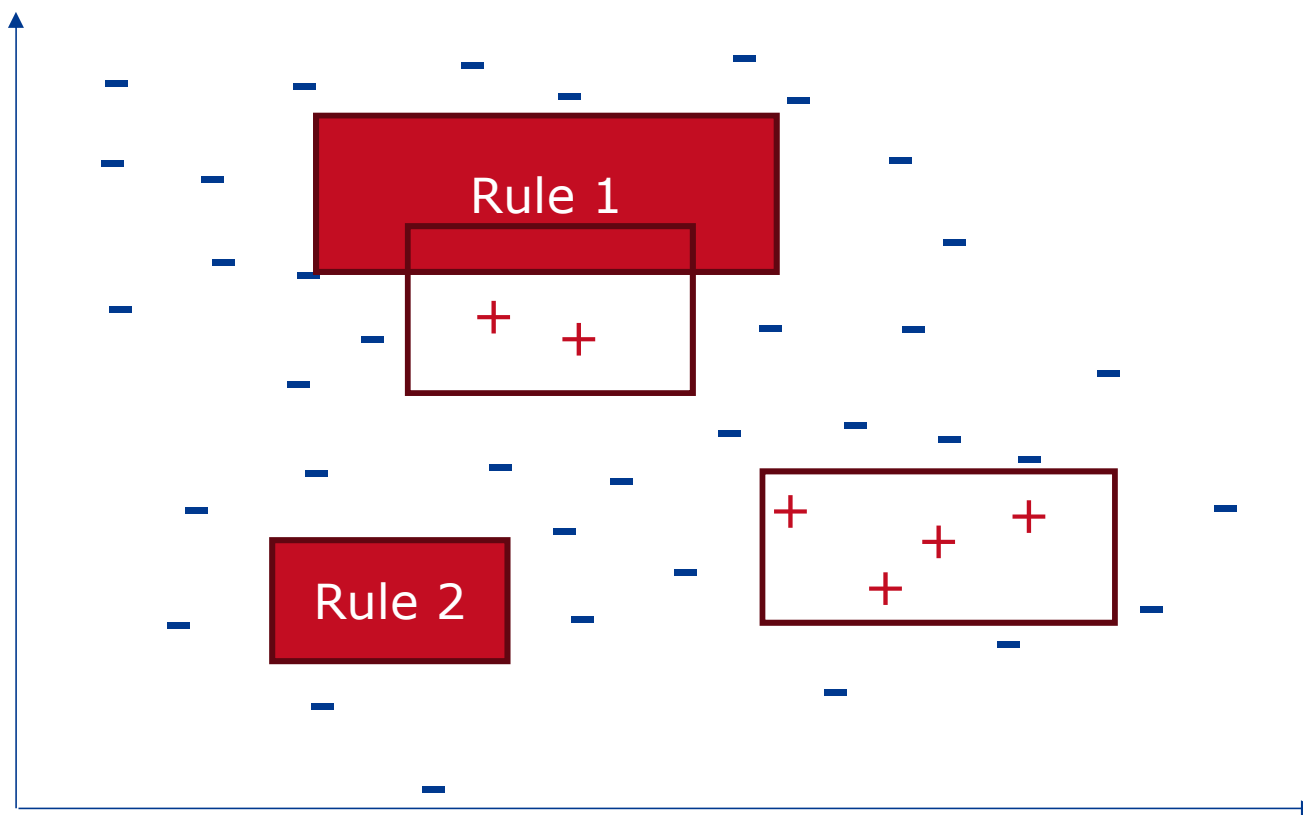
# 序贯覆盖

---



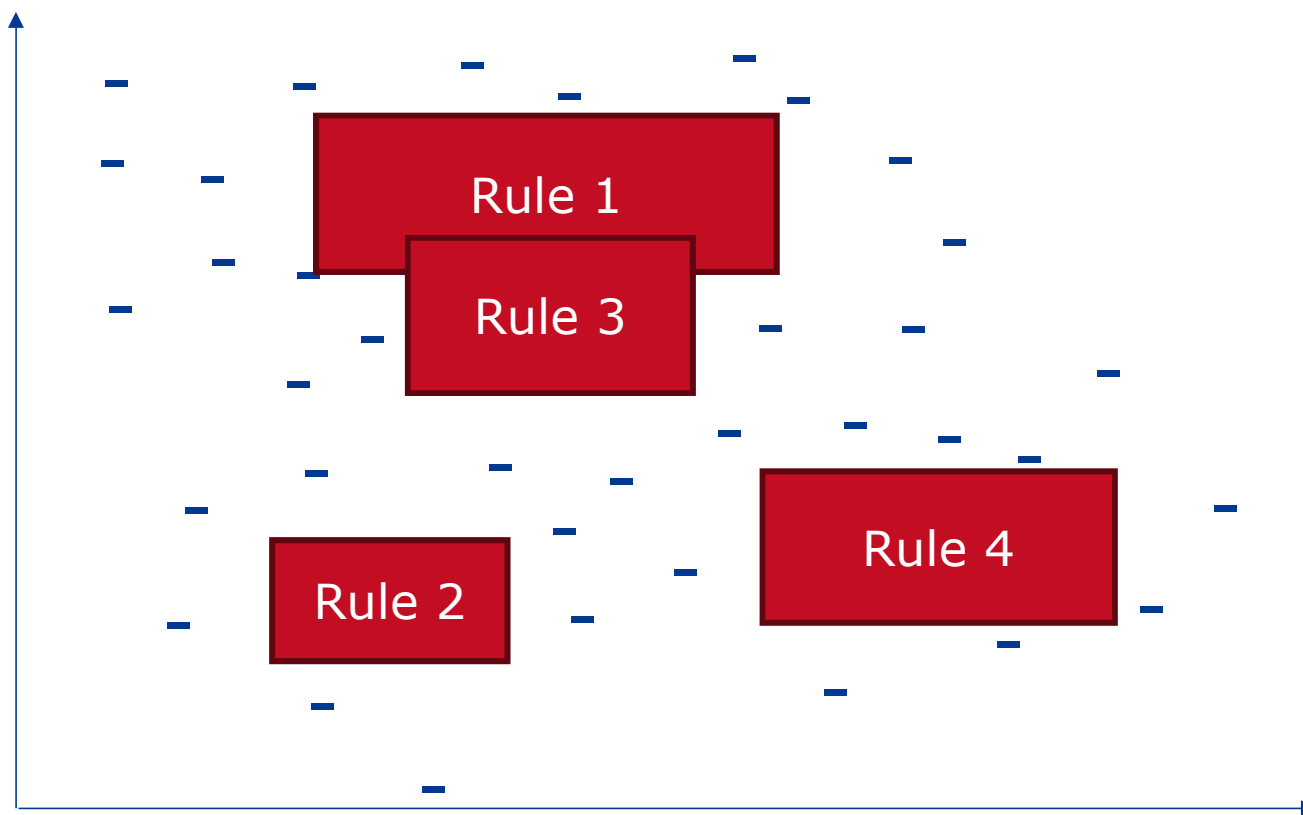
# 序贯覆盖

---



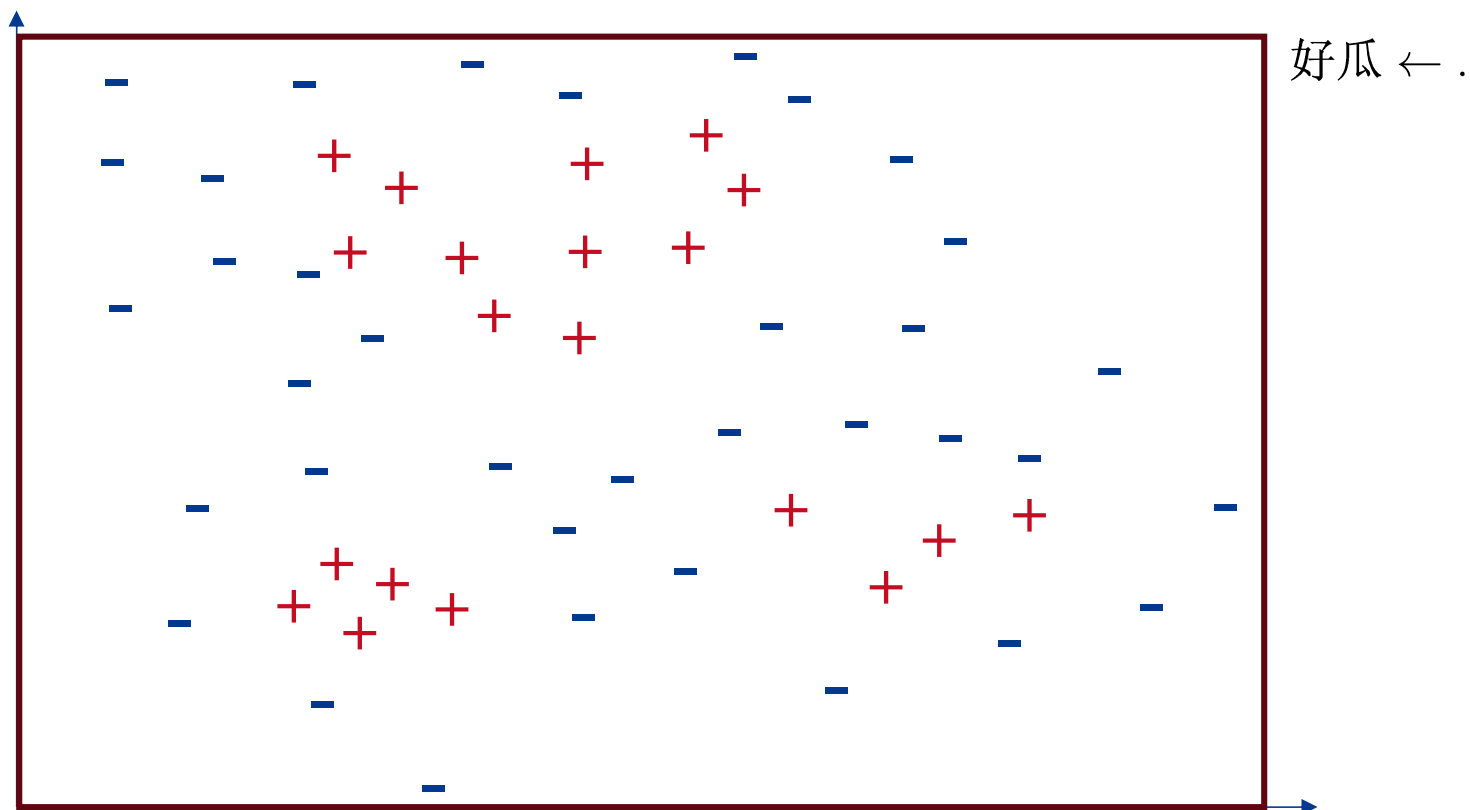
# 序贯覆盖

---



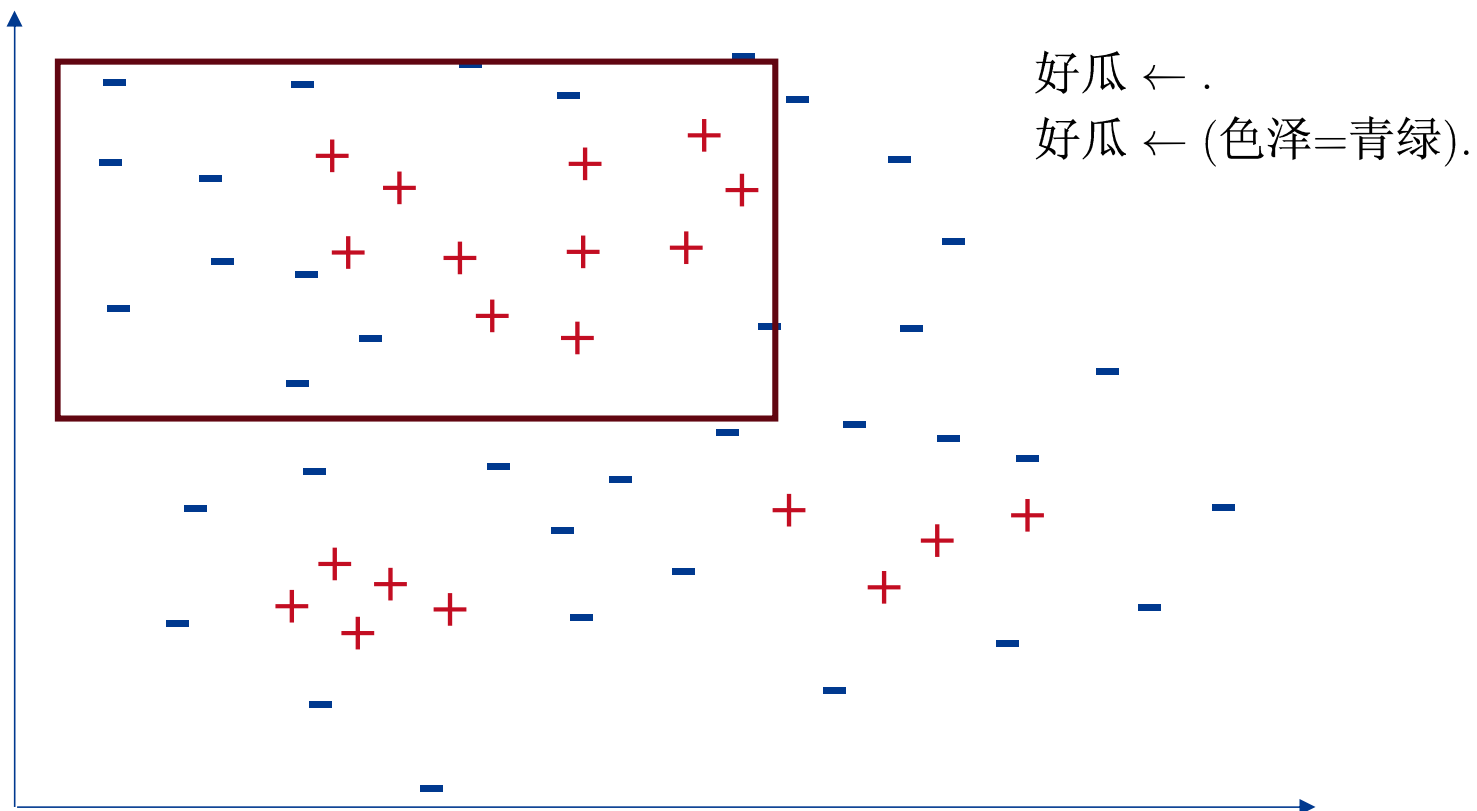
# 单条规则学习

- 自顶向下策略：一般到特殊（特化）



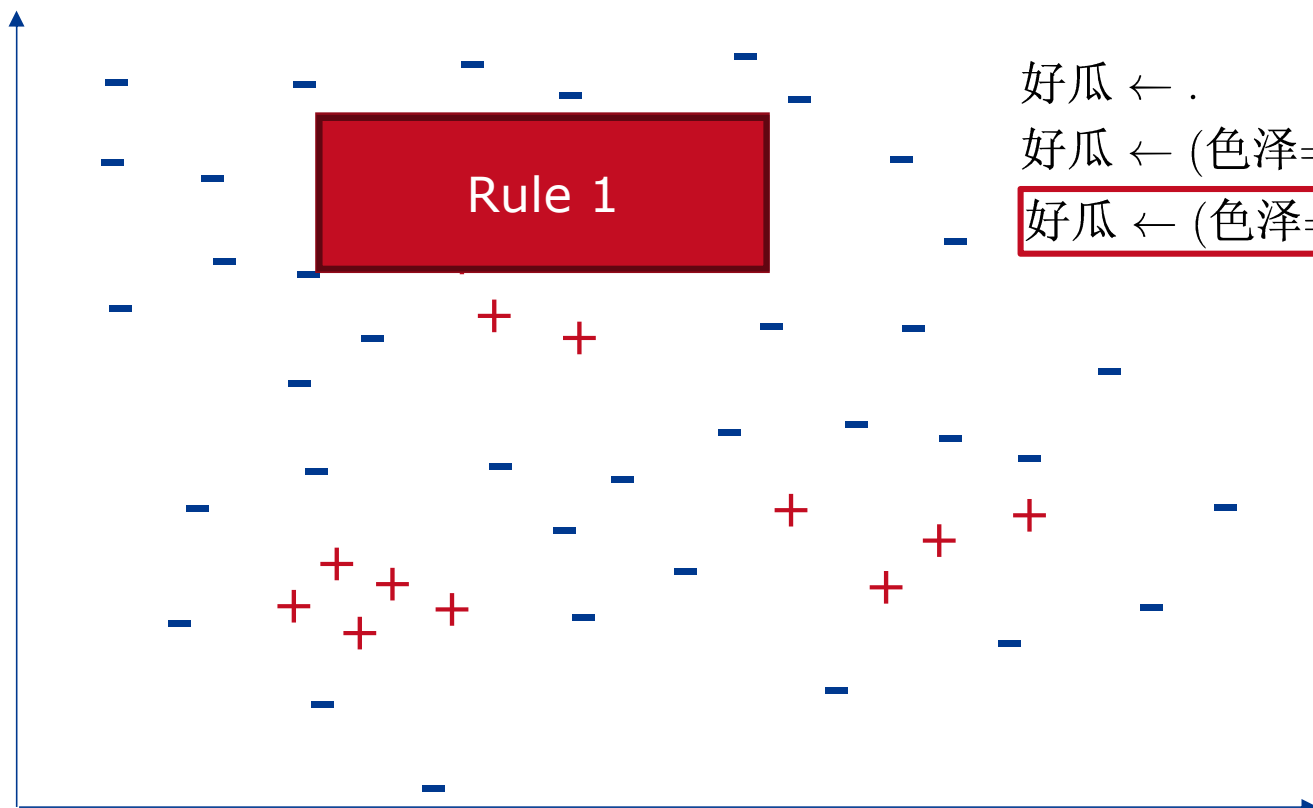
# 单条规则学习

□ 自顶向下策略：一般到特殊（特化）



# 单条规则学习

□ 自顶向下策略：一般到特殊（特化）



好瓜  $\leftarrow$  .

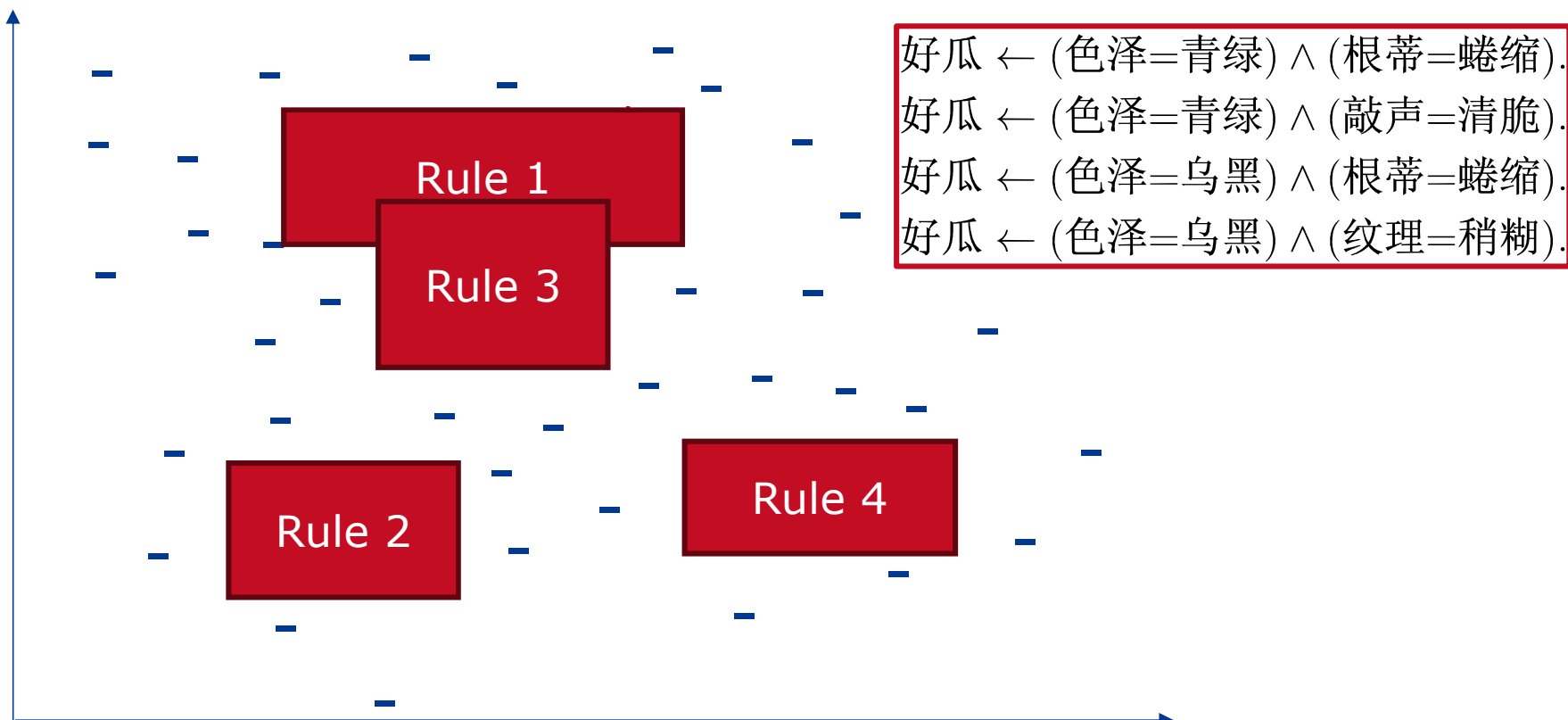
好瓜  $\leftarrow$  (色泽=青绿).

好瓜  $\leftarrow$  (色泽=青绿)  $\wedge$  (根蒂=蜷缩).



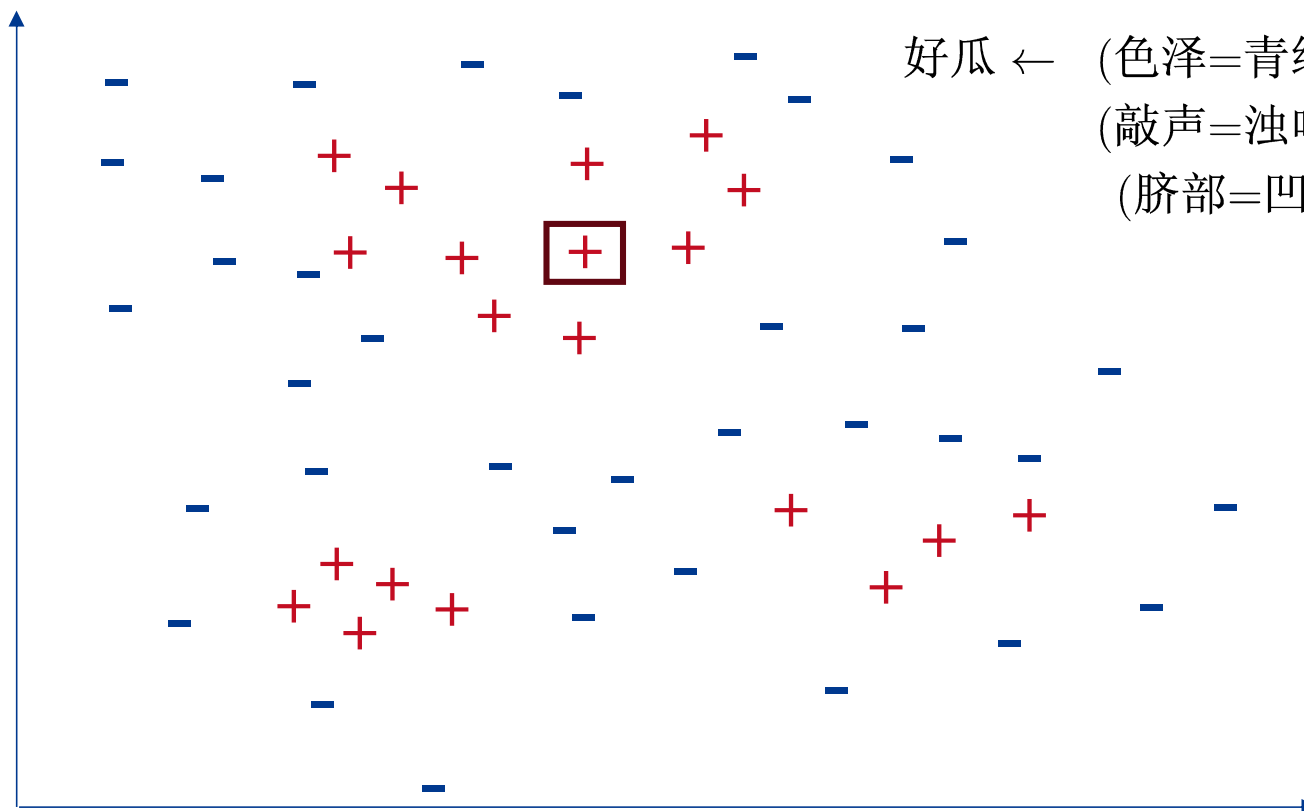
# 单条规则学习

□ 自顶向下策略：一般到特殊（特化）



# 单条规则学习

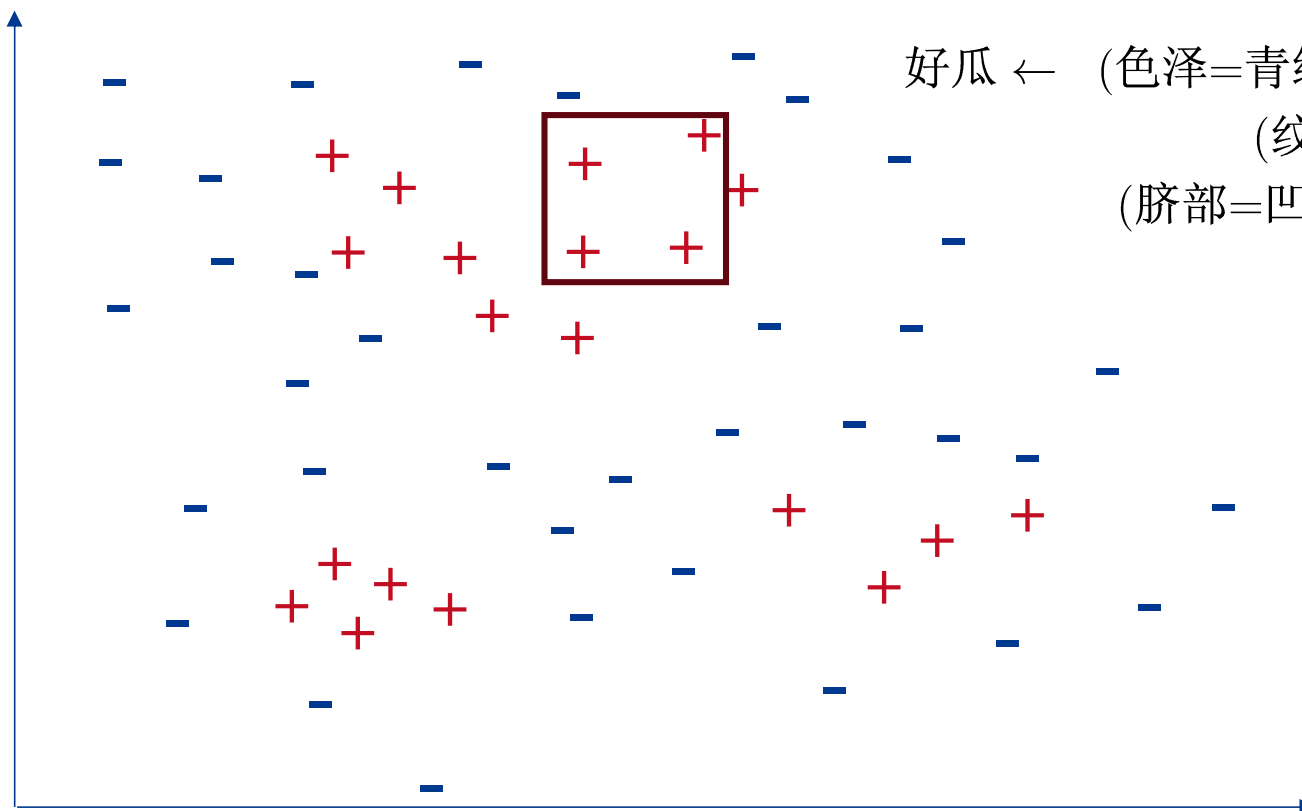
## □ 自底向上策略：特殊到一般（泛化）



好瓜  $\leftarrow$  (色泽=青绿)  $\wedge$  (根蒂=蜷缩)  $\wedge$   
(敲声=浊响)  $\wedge$  (纹理=清晰)  $\wedge$   
(脐部=凹陷)  $\wedge$  (触感=硬滑).

# 单条规则学习

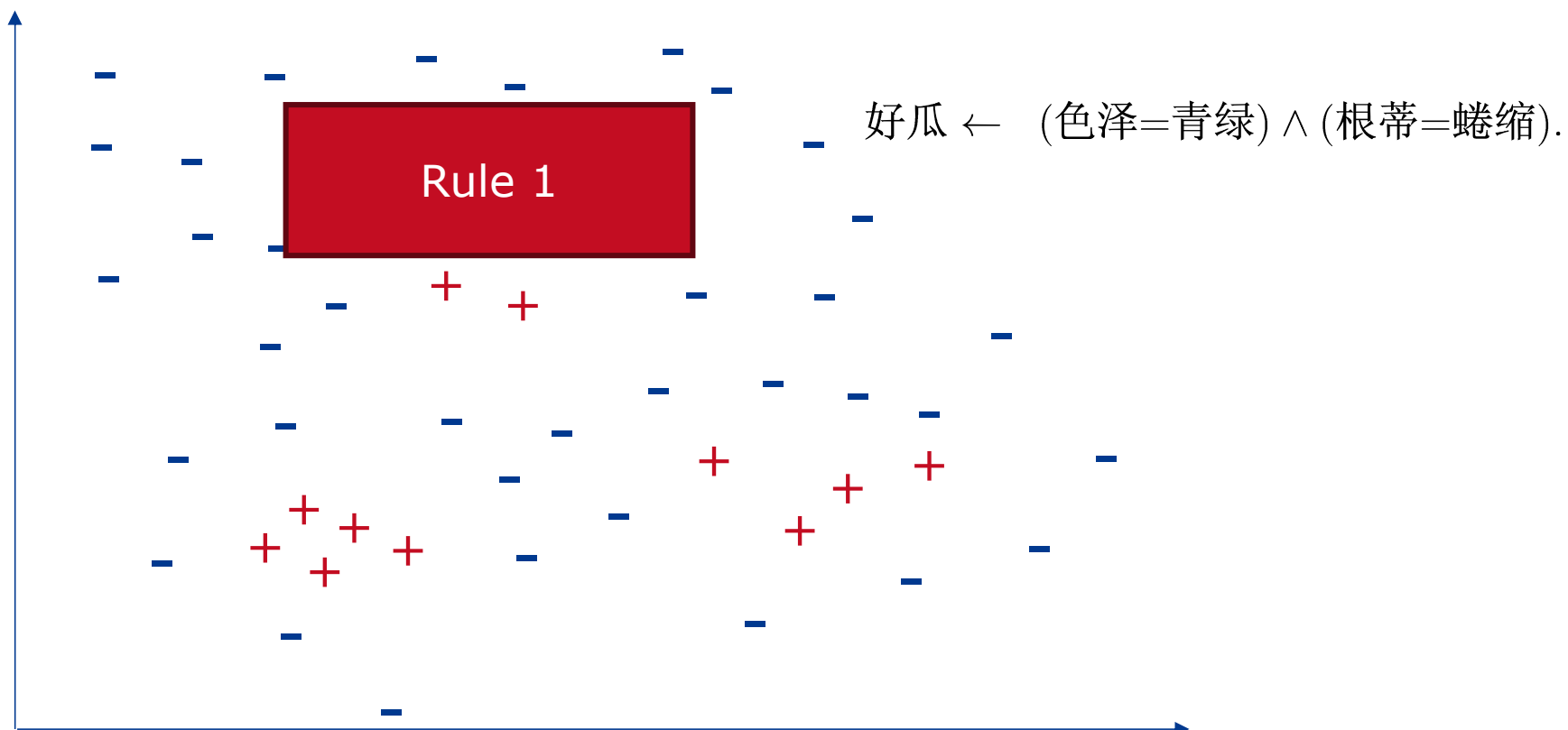
## □ 自底向上策略：特殊到一般（泛化）



好瓜  $\leftarrow$  (色泽=青绿)  $\wedge$  (根蒂=蜷缩)  $\wedge$   
(纹理=清晰)  $\wedge$   
(脐部=凹陷)  $\wedge$  (触感=硬滑).

# 单条规则学习

- 自底向上策略：特殊到一般（泛化）



# 单条规则学习

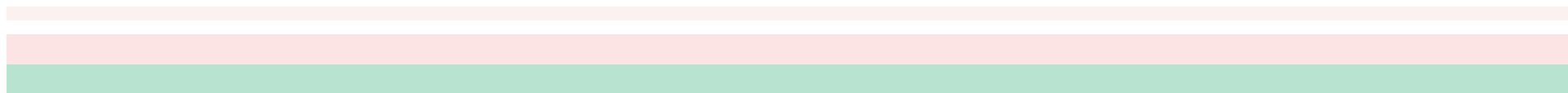
---

## □ 规则评判：增加/删除哪一个候选文字

- 准确率
- 信息熵增益（率）
- 基尼系数
- .....

## □ 规避局部最优

- 集束搜索：每次保留最优的多个候选规则
- .....



# 大纲

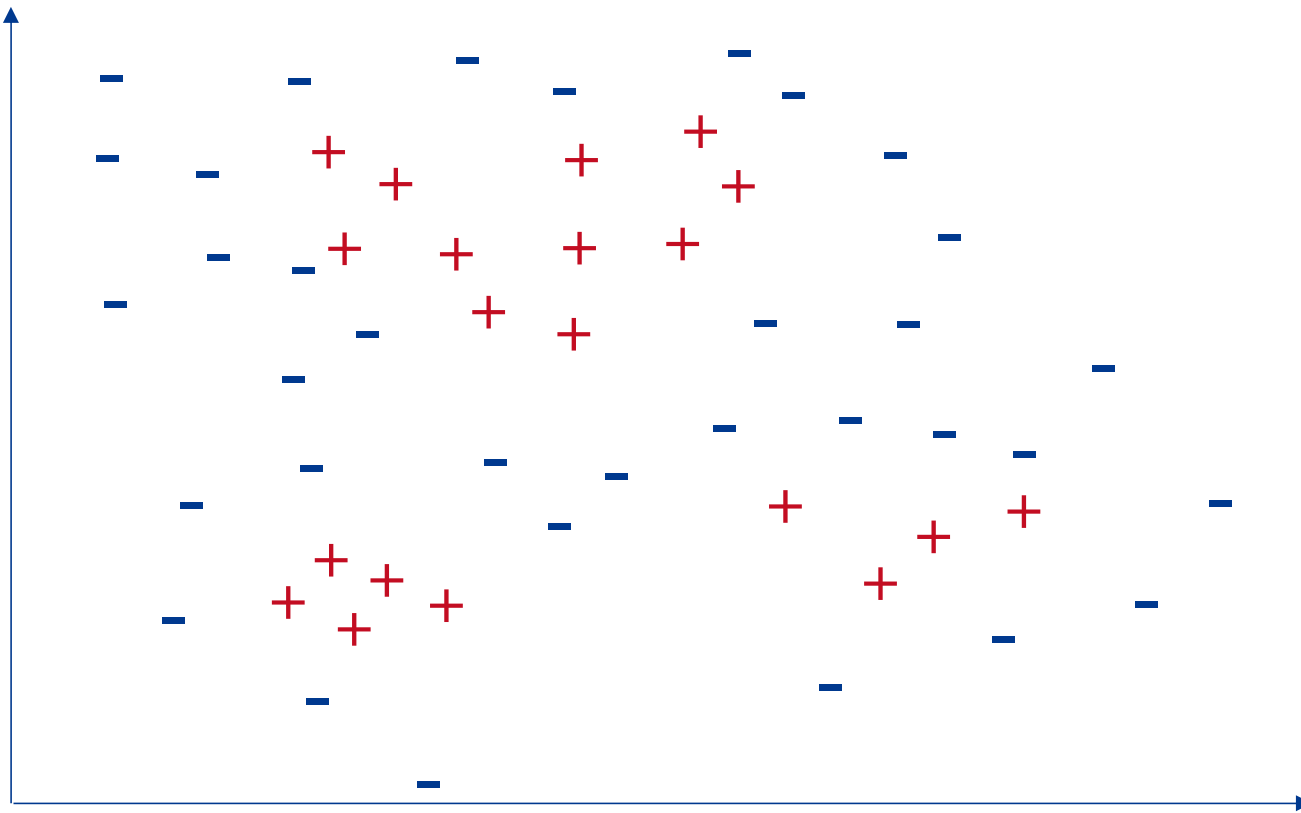
---

- 基本概念
- 序贯覆盖
- 剪枝优化
- 一阶规则学习
- 归纳逻辑程序设计

# 序贯覆盖（非最优）

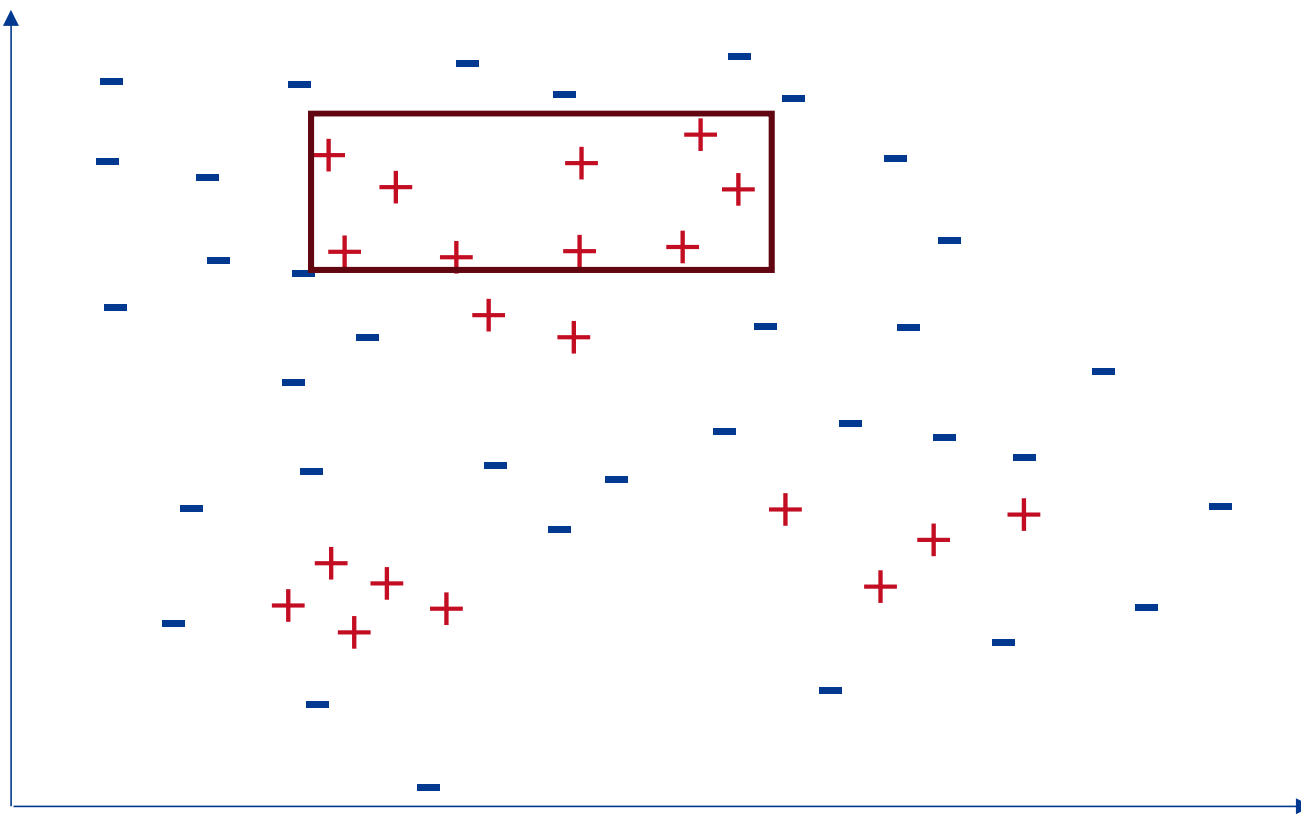
---

□ 贪心算法导致的非最优情况



## 序贯覆盖 (非最优)

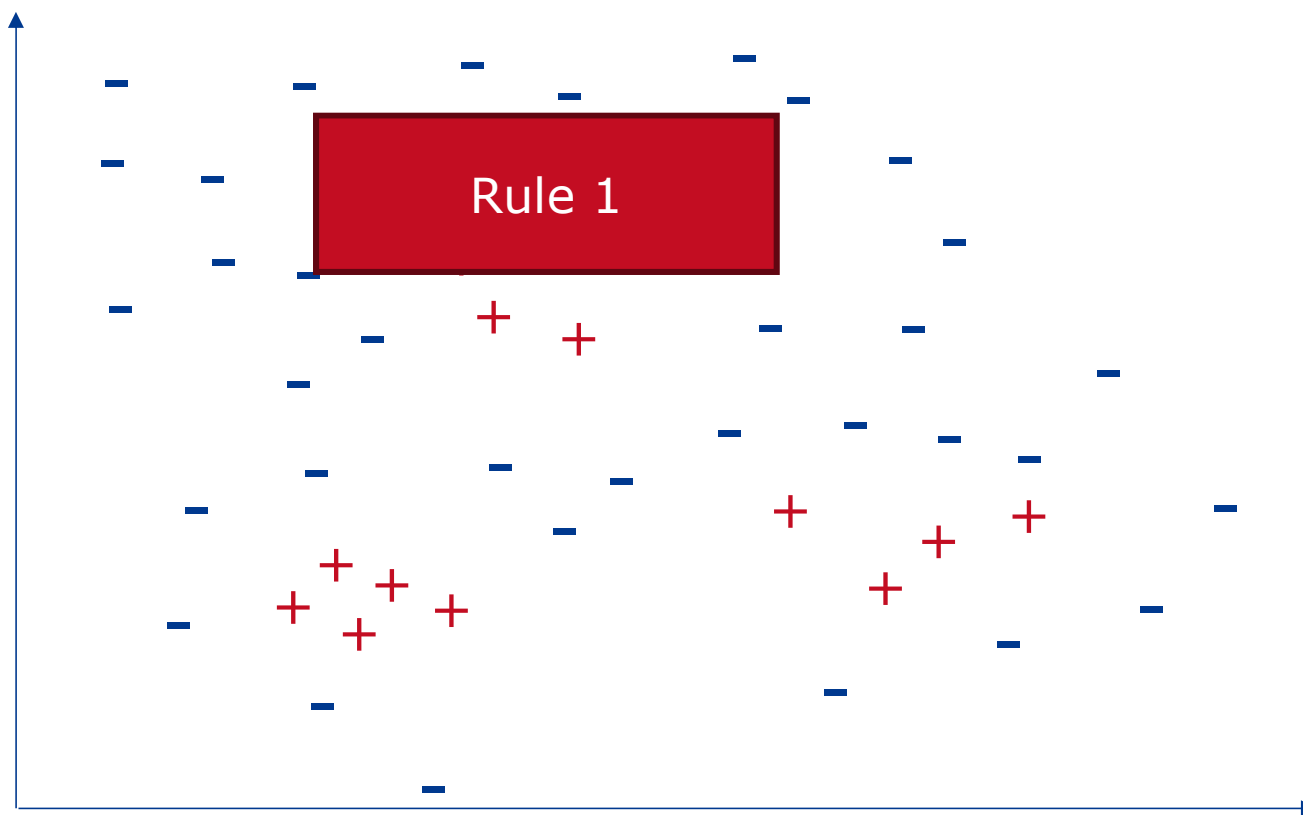
---





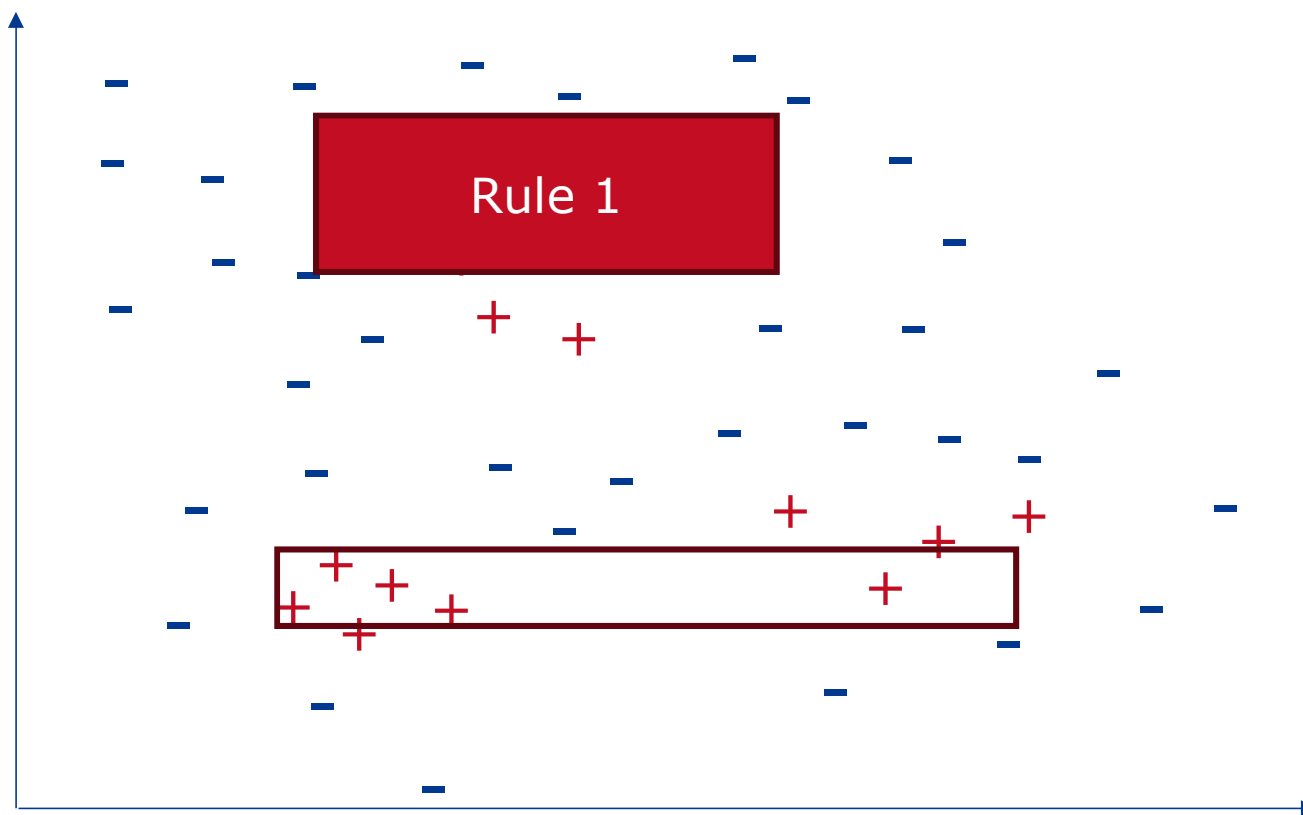
# 序贯覆盖 (非最优)

---



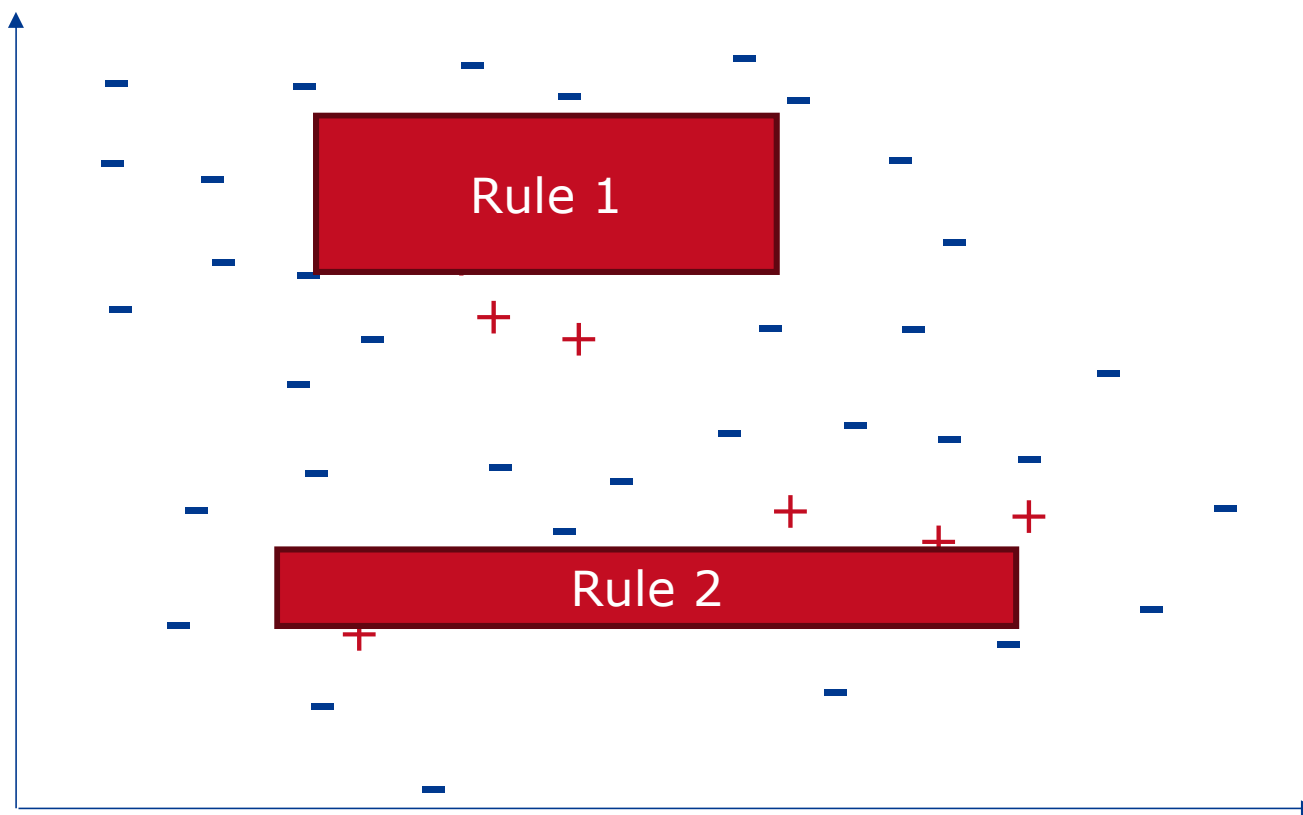
# 序贯覆盖 (非最优)

---



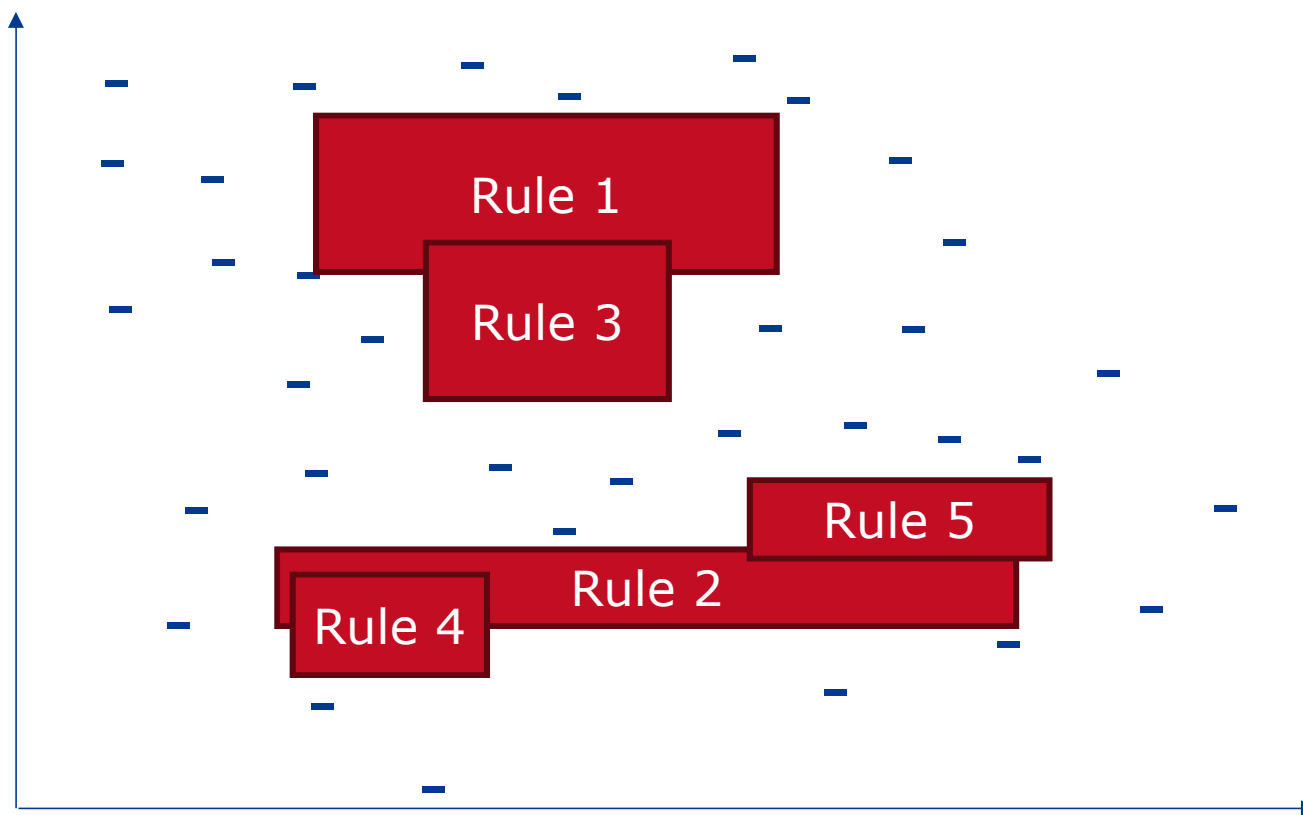
## 序贯覆盖 (非最优)

---



## 序贯覆盖（非最优）

---



# 剪枝优化

---

## □ 预剪枝

- 似然率统计量[Clark and Niblett, 1989]

$$LRS = 2 \cdot \left( \hat{m}_+ \log_2 \frac{\frac{\hat{m}_+}{\hat{m}_+ + \hat{m}_-}}{\frac{m_+}{m_+ + m_-}} + \hat{m}_- \log_2 \frac{\frac{\hat{m}_-}{\hat{m}_+ + \hat{m}_-}}{\frac{m_-}{m_+ + m_-}} \right)$$

## □ 后剪枝

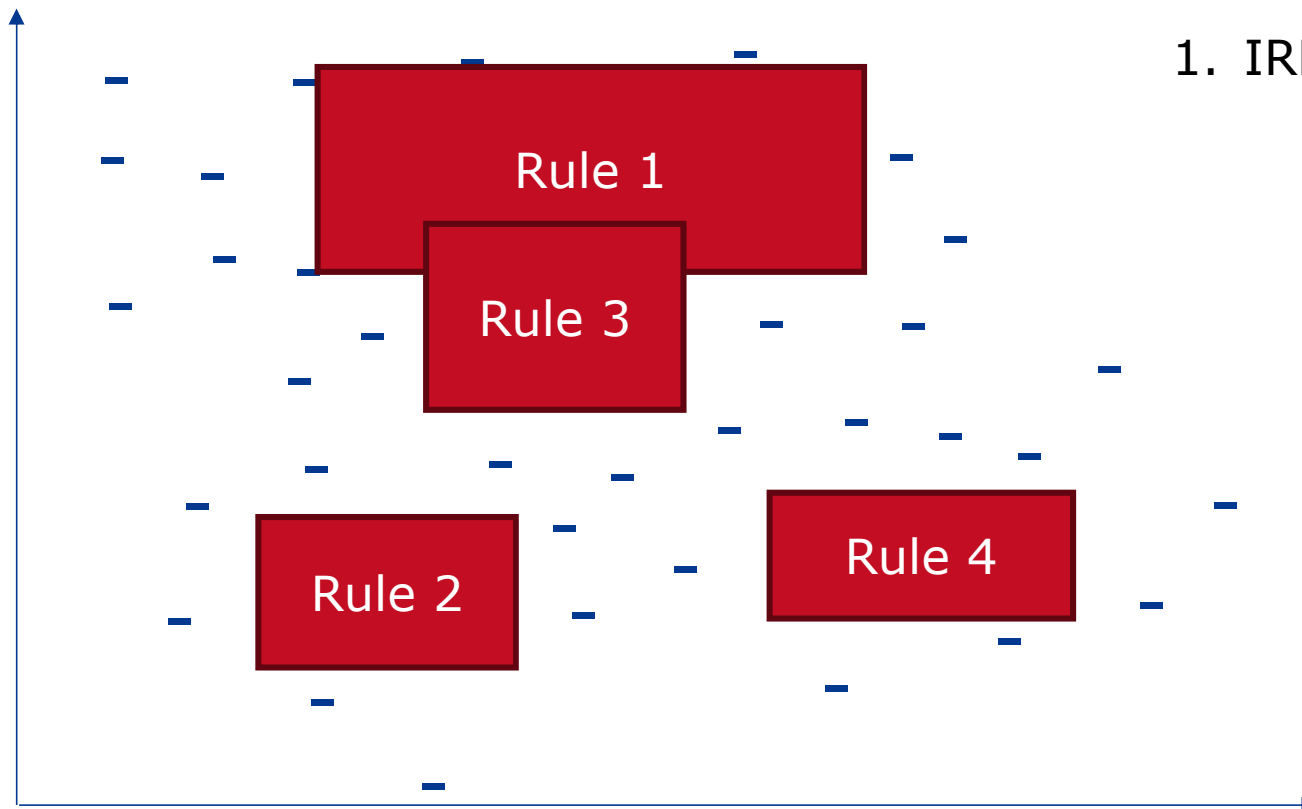
- 减错剪枝 (REP) [Brunk and Pazzani, 1991]
  - 穷举所有可能的剪枝操作 (删除文字, 删除规则)
  - 用验证集反复剪枝直到准确率无法提高

## □ 二者结合

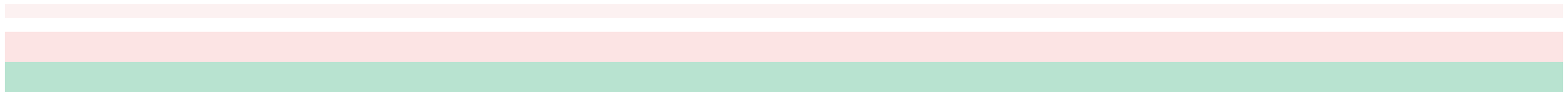
- IREP[Fürnkranz and Widmer, 1994]
  - 每生成一条新规则即对其进行REP剪枝
- IREP\*[Cohen, 1995]
  - 对IREP的改进

# RIPPER [Cohen, 1995]

---

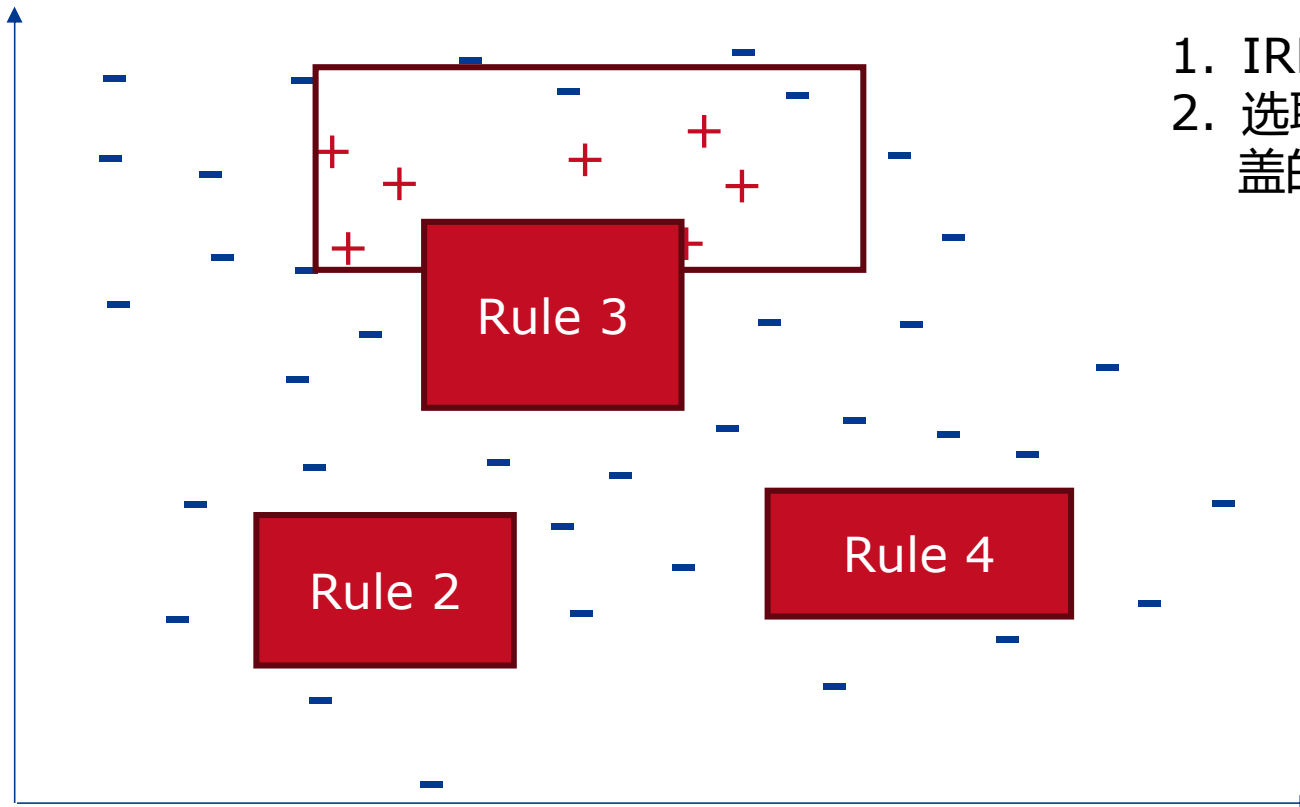


1. IREP\*生成规则集



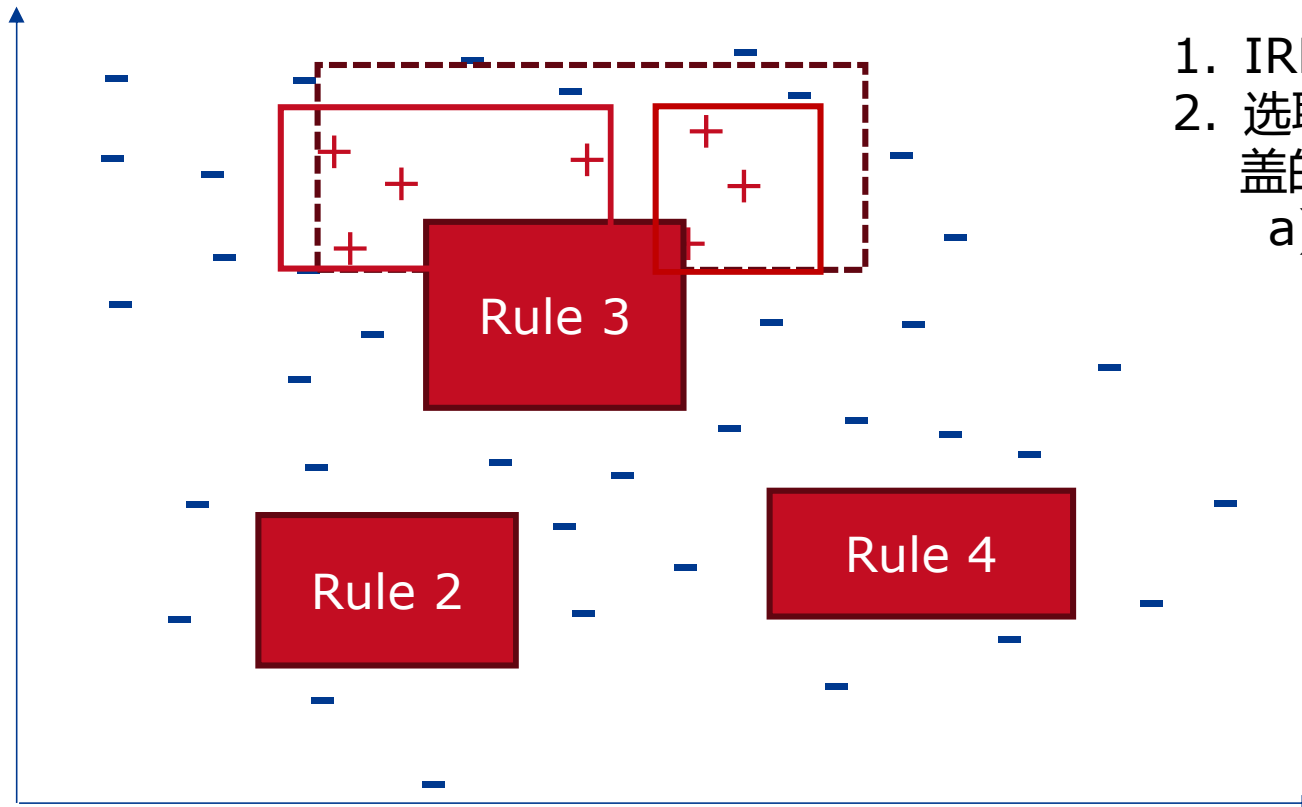
# RIPPER

---



1. IREP\*生成规则集
2. 选取一条规则，找到其覆盖的样例

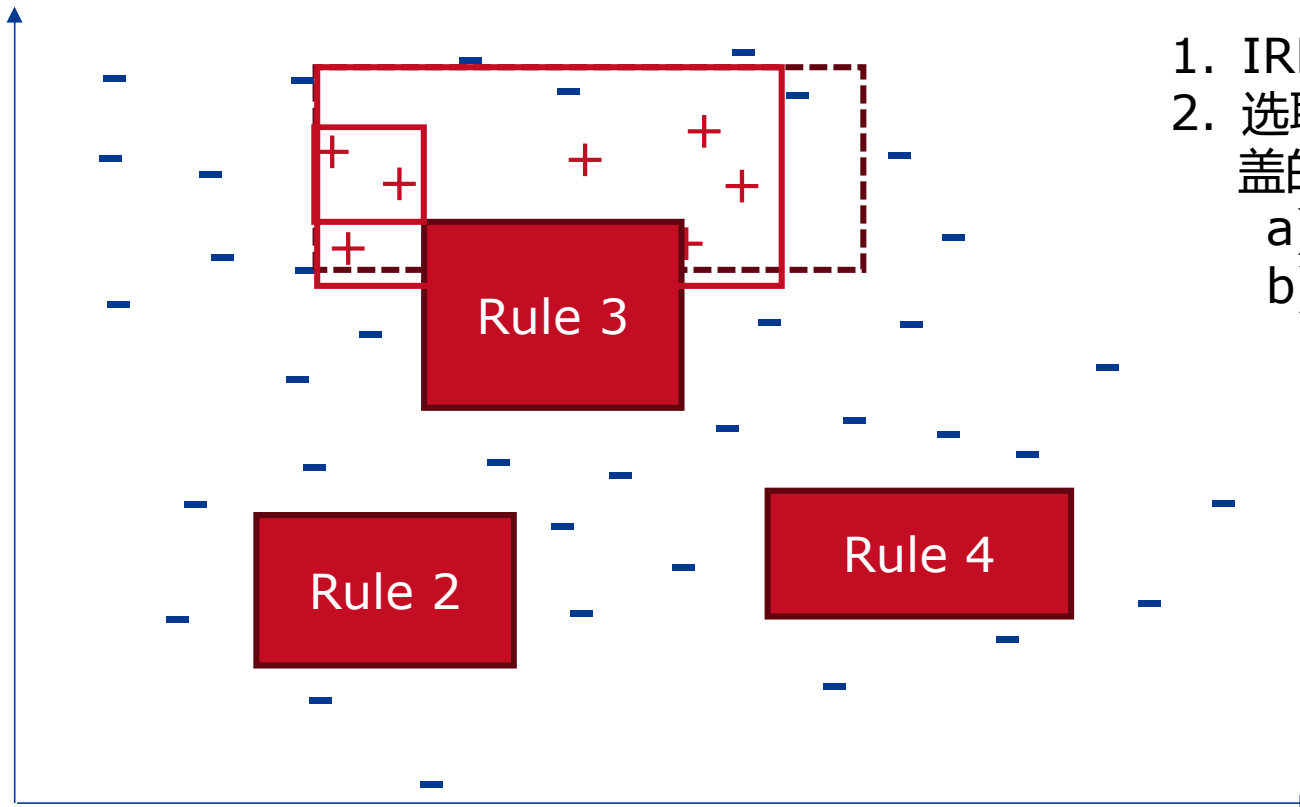
# RIPPER



1. IREP\*生成规则集
2. 选取一条规则，找到其覆盖的样例
  - a) 重新生成规则

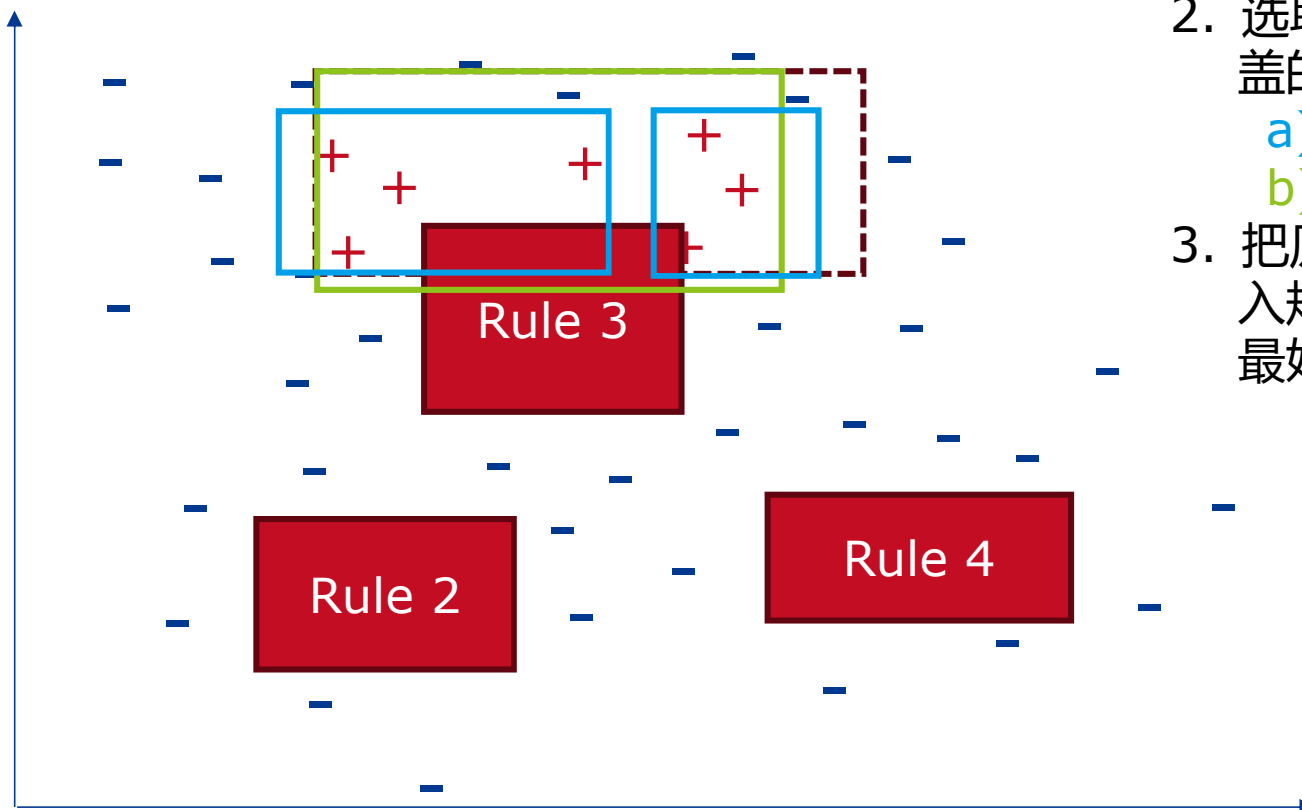


# RIPPER



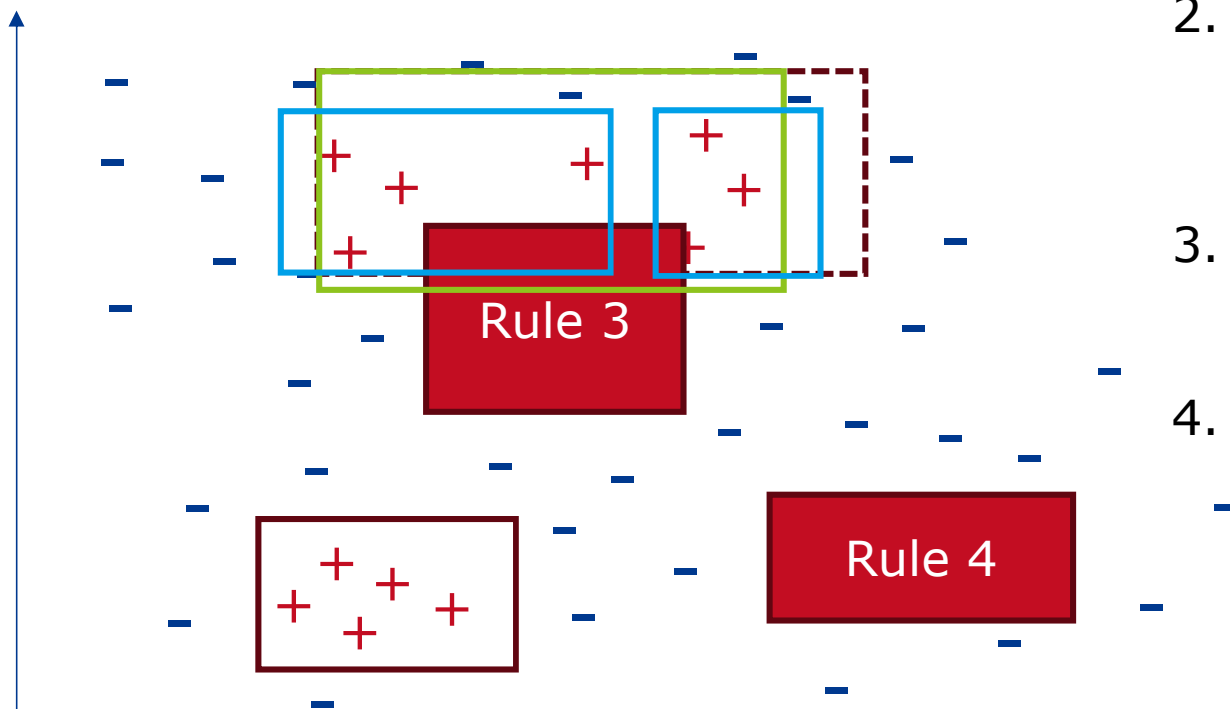
1. IREP\*生成规则集
2. 选取一条规则，找到其覆盖的样例
  - a) 重新生成规则
  - b) 特化原规则再泛化

# RIPPER



1. 生成规则集
2. 选取一条规则，找到其覆盖的样例
  - a) 重新生成规则
  - b) 特化原规则再泛化
3. 把原规则和新规则分别置入规则集进行评价，留下最好的

# RIPPER



1. 生成规则集
2. 选取一条规则，找到其覆盖的样例
  - a) 重新生成规则
  - b) 特化原规则再泛化
3. 把原规则和新规则分别置入规则集进行评价，留下最好的。
4. 反复优化直到无法进步

RIPPER将所有规则放在一起优化，通过全局的考虑来缓解序贯覆盖的局部性

# 大纲

---

- 基本概念
- 序贯覆盖
- 剪枝优化
- 一阶规则学习
- 归纳逻辑程序设计

# 一阶规则学习

---

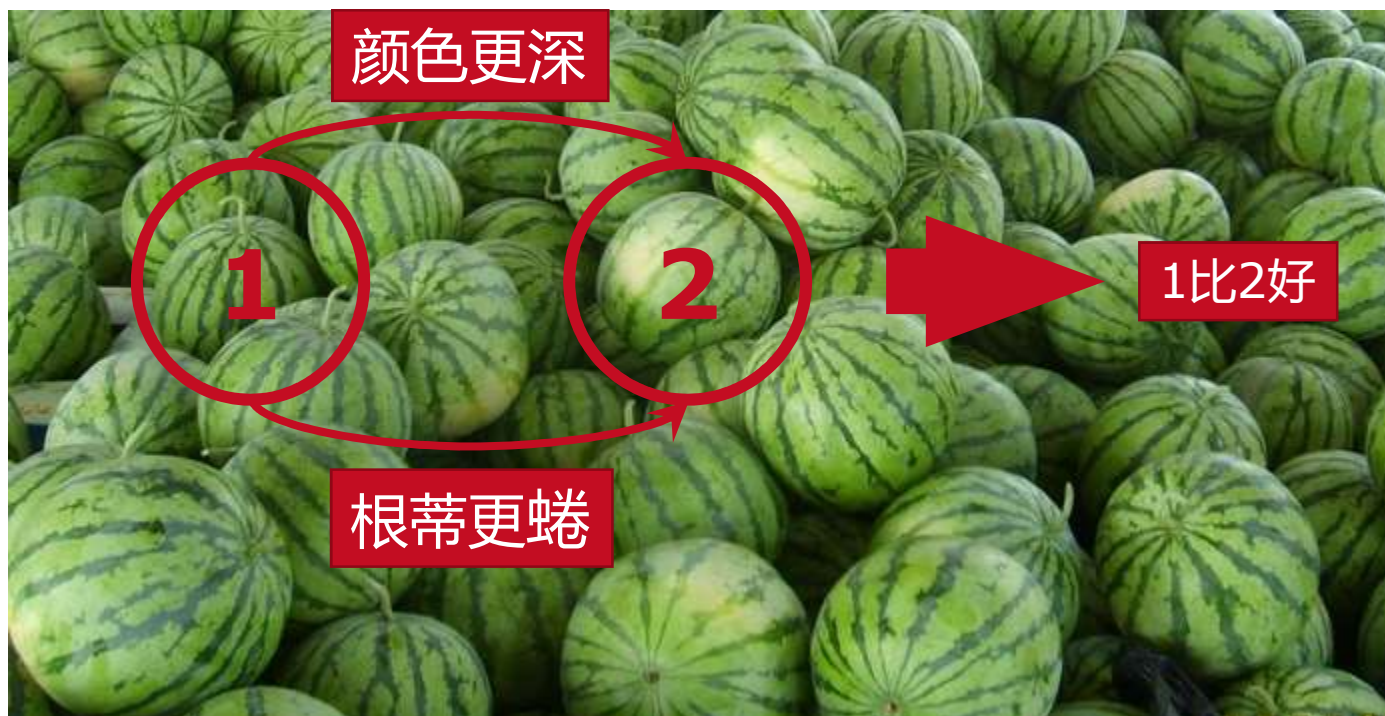
- “一阶”的目的：描述一类物体的性质、相互关系



如何量化颜色、根蒂、脐部、敲声的属性值？

# 一阶规则学习

- 利用一阶关系来挑“更好的”瓜



$\forall X \forall Y (\text{更好}(X, Y) \leftarrow \text{颜色更深}(X, Y) \wedge \text{根蒂更弯}(X, Y)).$   
 $\text{更好}(X, Y) \leftarrow \text{颜色更深}(X, Y) \wedge \text{根蒂更弯}(X, Y).$

# 一阶规则学习

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
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
10	青绿	硬挺	清脆	清晰	平坦	软粘	否
14	浅白	稍蜷	沉闷	稍糊	凹陷	硬滑	否
15	乌黑	稍蜷	浊清	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

## 属性-值(Attribute-value)数据 【命题逻辑】

- “色泽深度”：乌黑 > 青绿 > 浅白；
- “根蒂弯度”：蜷缩 > 稍蜷 > 硬挺；
- ...
- “更好”：好瓜 > 坏瓜

## 背景知识

## 关系型(Relational)数据 【一阶逻辑】

## 样例

色泽更深(2,1)	色泽更深(2,6)	色泽更深(2,10)	色泽更深(2,14)
色泽更深(2,16)	色泽更深(2,17)	色泽更深(3,1)	色泽更深(3,6)
...			
色泽更深(15,16)	色泽更深(15,17)	色泽更深(17,14)	色泽更深(17,16)
根蒂更蜷(1,6)	根蒂更蜷(1,7)	根蒂更蜷(1,10)	根蒂更蜷(1,14)
...			
更好(1,10)	更好(1,14)	更好(1,15)	更好(1,16)
...			
更好(7,14)	更好(7,15)	更好(7,16)	更好(7,17)
¬更好(10,1)	¬更好(10,2)	¬更好(10,3)	¬更好(10,6)
...			



# FOIL[Quinlan, 1990]

## □ 序贯覆盖生成规则集

## □ 自顶向下学习单条规则

- 候选文字需考虑所有可能的选项

更好(X, Y) ← {  
    色泽更深(X, Y)    色泽更深(Y, X)    色泽更深(X, Z)    色泽更深(Z, Y)  
    色泽更深(Y, Z)    色泽更深(Z, X)    色泽更深(X, X)    色泽更深(Y, Y)  
    根蒂更蜷(X, Y)    ...  
    敲声更沉(X, Y)    ...  
    ...

- 是否引入新变量?
- 是否使用否定文字?
- 是否允许递归?
- 是否引入函数嵌套? :  $p(f(f...(f(X))))$

- 规则生长的评判标准为**FOIL**增益

$$F\_GAIN = \hat{m}_+ (\log_2 \frac{\hat{m}_+}{\hat{m}_+ + \hat{m}_-}) - \log_2 \frac{m_+}{m_+ + m_-}$$

## □ 后剪枝优化规则集



# 大纲

---

- 基本概念
- 序贯覆盖
- 剪枝优化
- 一阶规则学习
- 归纳逻辑程序设计

# 归纳逻辑程序[Muggleton, 1991]

---

- 目标：完备地学习一阶规则（Horn子句）
- 仍然以序贯覆盖方法学习规则集
- 一般采用自底向上策略学习单条规则
  - 不需要列举所有可能的候选规则
  - 对目标概念的搜索维持在样例附近的局部区域
  - 自顶向下策略的搜索空间对于规则长度呈指数级增长

# 最小一般泛化(LGG)[Plotkin, 1970]

---

- “泛化”：将覆盖率低的规则变换为覆盖率高的规则
- “一般”：覆盖率尽可能高
- “最小”：变换时对原规则的改动尽可能小
  
- 寻找两条规则**LGG**的步骤：
  - 找出两条规则中涉及相同谓词的文字
  - 考察谓词后括号里的项：
    - $lgg(t, t) = t$
    - $lgg(s, t), s \neq t$ 
      - $s, t$ 不是谓词相同的项, 则  $lgg(s, t) = V$ ,  $V$ 为任意未出现过的变量
      - $s, t$ 为谓词相同的项, 递归考察其括号内的项
  - 删除没有相同谓词出现的文字

# 最小一般泛化(LGG)[Plotkin, 1970]

更好(1, 10)  $\leftarrow$  根蒂更蜷(1, 10)  $\wedge$  声音更沉(1, 10)  $\wedge$  脐部更凹(1, 10)  $\wedge$  触感更硬(1, 10).

更好(1, 15)  $\leftarrow$  根蒂更蜷(1, 15)  $\wedge$  脐部更凹(1, 15)  $\wedge$  触感更硬(1, 15).

$lgg(10, 15) = Y$



更好(1, Y)  $\leftarrow$  根蒂更蜷(1, Y)  $\wedge$  声音更沉(1, Y)  $\wedge$  脐部更凹(1, Y)  $\wedge$  触感更硬(1, Y).

更好(1, Y)  $\leftarrow$  根蒂更蜷(1, Y)  $\wedge$  脐部更凹(1, Y)  $\wedge$  触感更硬(1, Y).



更好(1, Y)  $\leftarrow$  根蒂更蜷(1, Y)  $\wedge$  脐部更凹(1, Y)  $\wedge$  触感更硬(1, Y).

更好(2, 10)  $\leftarrow$  颜色更深(2, 10)  $\wedge$  根蒂更蜷(2, 10)  $\wedge$  敲声更沉(2, 10)

$\wedge$  脐部更凹(2, 10)  $\wedge$  触感更硬(2, 10).

$lgg(1, 2) = X$

$lgg(Y, 10) = Y_2$



更好(X, Y<sub>2</sub>)  $\leftarrow$  根蒂更蜷(X, Y<sub>2</sub>)  $\wedge$  脐部更凹(X, Y<sub>2</sub>)  $\wedge$  触感更硬(X, Y<sub>2</sub>).

# 最小一般泛化(LGG)[Plotkin, 1970]

---

## □ 其他基于LGG的ILP算法

- 考虑否定文字
- 不同的初始化选择
  - 多条特殊规则
  - 考虑所有背景知识 (RLGG) [Plotkin, 1971]
- ...

## □ 能不能在学习过程中自动归纳新知识？

瓜瓢更沙( $X, Y$ )  $\leftarrow$  敲声更响( $X, Y$ ).

瓜皮更薄( $X, Y$ )  $\leftarrow$  脐部更小( $X, Y$ ).

更好( $X, Y$ )  $\leftarrow$  瓜瓢更沙( $X, Y$ )  $\wedge$  瓜皮更薄( $X, Y$ ).

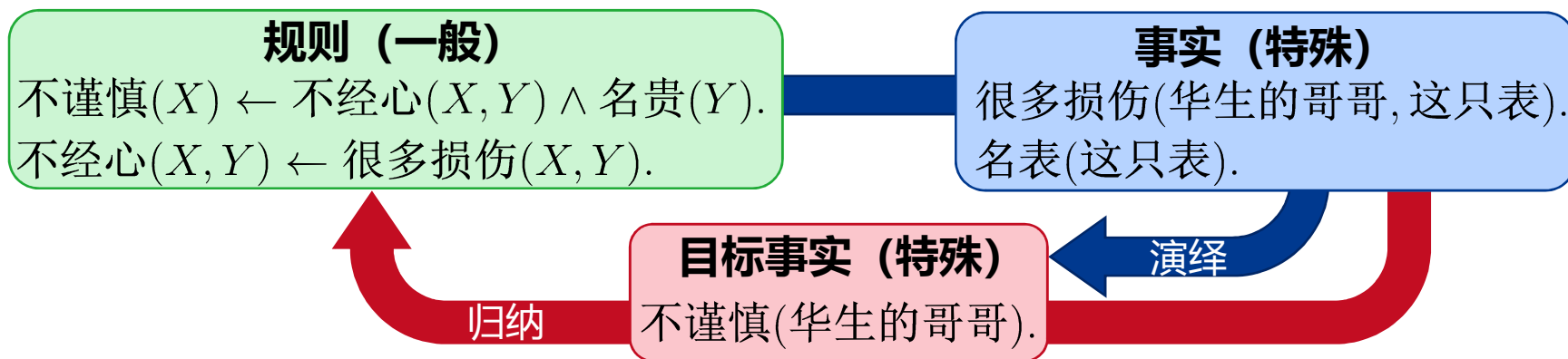
# 逆归结[Muggleton and Buntine, 1988]

## □ 演绎(deduction) VS 归纳(induction)

“.....猜想是很不好的习惯，它有害于作逻辑的推理。你所以觉得奇怪，是因为你没有了解我的思路，没有注意到往往能推断出大事来的那些细小问题 (the **small facts** upon which **large inferences** may depend)。举例来说吧，我开始时曾说你哥哥的行为很不谨慎。请看这只表，不仅下面边缘上有凹痕两处，整个表的上面还有无数的伤痕，这是因为惯于把表放在有钱币、钥匙一类硬东西的衣袋里的缘故。对一只价值五十多镑的表这样不经心，说他生活不检点，总不算是过分吧！.....。”

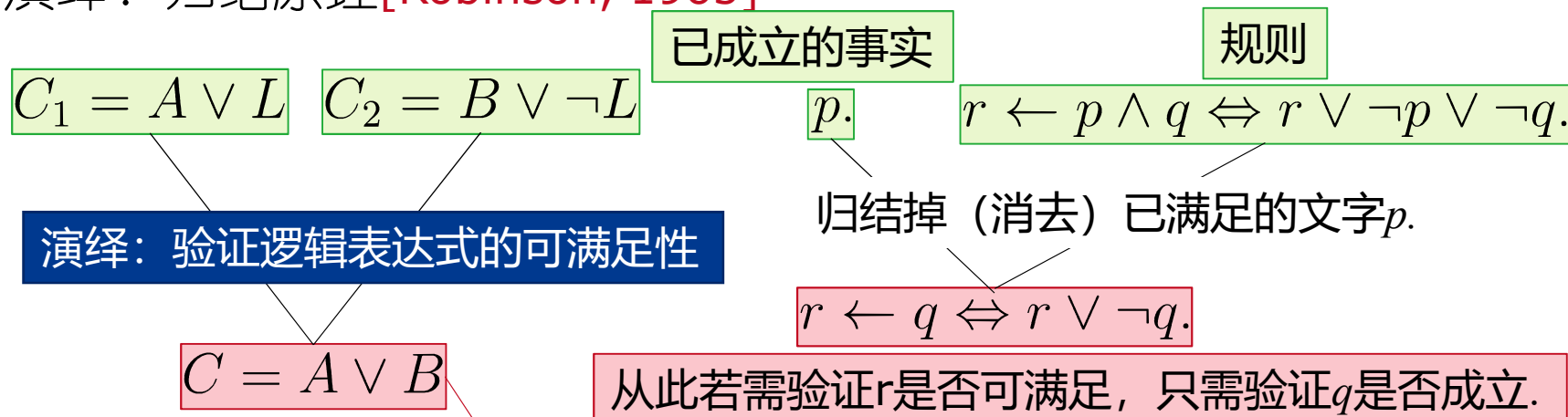
歇洛克·福尔摩斯

演绎法研究 (The science of deduction) —— 《四签名》

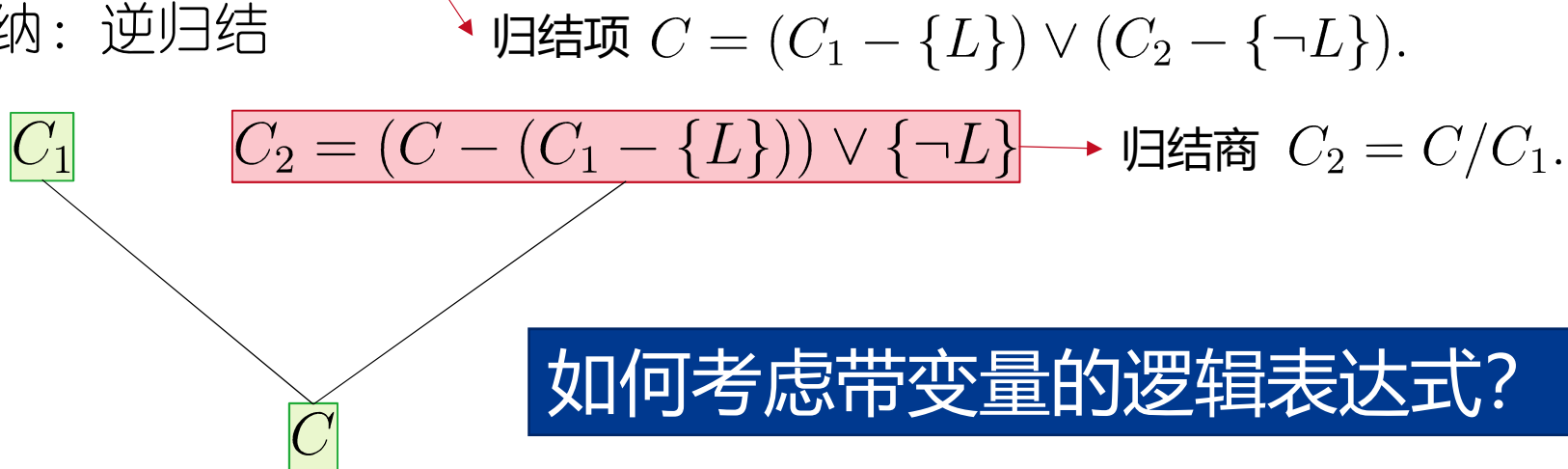


# 逆归结[Muggleton and Buntine, 1988]

## □ 演绎：归结原理[Robinson, 1965]



## □ 归纳：逆归结



# Prolog

f.pl:

male(di).

male(jianbo).

female(xin).

female(yuan).

female(yuqing).

father(jianbo,di).

father(di,yuqing).

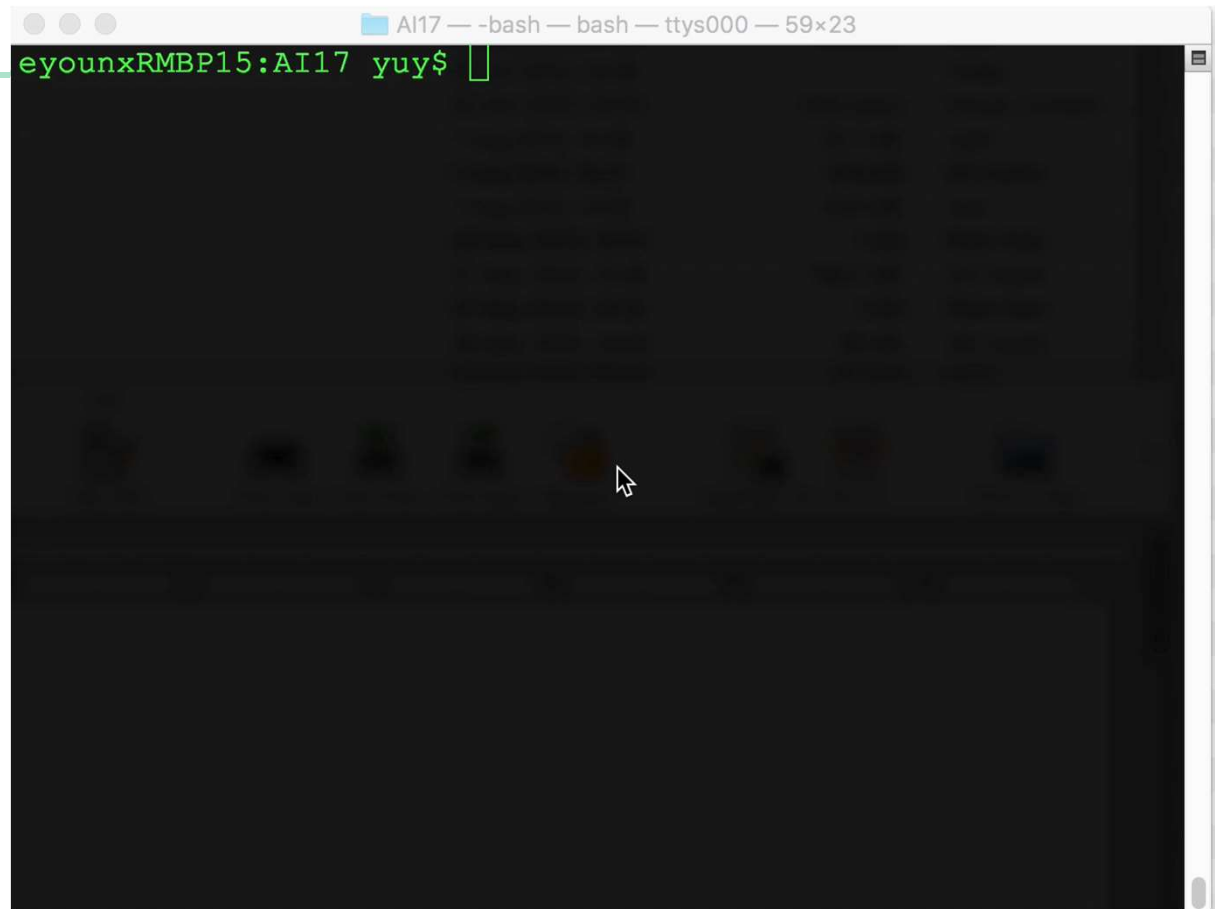
mother(xin,di).

mother(yuan,yuqing).

grandfather(X,Y):-father(X,Z),father(Z,Y).

grandmother(X,Y):-mother(X,Z),father(Z,Y).

daughter(X,Y):-father(X,Y),female(Y).





# 归纳逻辑程序学习

---

- 完备地学习一阶规则
- 谓词发明：发现领域中隐含的结构
- 得到的规则可直接作为逻辑程序进行应用
- 其他归纳逻辑程序学习方法：
  - 命题化学习[Lavrac and Dzeroski, 1993]
  - 逆蕴含[Muggleton, 1995]
  - Meta-Interpretive Learning[Muggleton and Lin, 2013]
  - .....

# 总结

---

- 逻辑规则
  - 命题规则、一阶规则
- 规则集学习：序贯覆盖
  - 单条规则学习：自顶向下、自底向上
  - 剪枝与后处理
- 一阶规则学习
- 归纳逻辑程序学习
  - 最小一般泛化
  - 逆归结

# 参考工具

---

- WEKA中的JRIP  
(<http://weka.sourceforge.net/doc.dev/weka/classifiers/rules/JRip.html>)
  - ILP的集成工具ALEPH  
(<http://www.cs.ox.ac.uk/activities/machlearn/Aleph/aleph.html>)
  - 相对最小一般泛化: GOLEM  
(<http://www.doc.ic.ac.uk/~shm/golem.html>)
  - 逆演绎: Progol  
(<http://www.doc.ic.ac.uk/~shm/progol.html>)
  - Prolog环境: YAP (<http://www.dcc.fc.up.pt/~vsc/Yap/>)
- 