

1. 表1中若只给出1、4两个样例，请给出版本空间

编号	色泽	根蒂	敲声	好瓜
1	青绿	卷缩	浊响	是
4	乌黑	稍卷	沉闷	否

“

解答：

- 计算假设空间

$$\text{space_size} = 3(\text{青绿} + \text{乌黑} + *) * 3(\text{卷缩} + \text{稍卷} + *) * 3(\text{浊响} + \text{沉闷} + *) + 1(\text{空集合}) = 28$$

空间如下

1. 青绿, 卷缩, 浊响
2. 青绿, 卷缩, 沉闷
3. 青绿, 卷缩, *
4. 青绿, 稍卷, 浊响
5. 青绿, 稍卷, 沉闷
6. 青绿, 稍卷, *
7. 青绿, *, 浊响
8. 青绿, *, 沉闷
9. 青绿, *, *
10. 乌黑, 卷缩, 浊响
11. 乌黑, 卷缩, 沉闷
12. 乌黑, 卷缩, *
13. 乌黑, 稍卷, 浊响
14. 乌黑, 稍卷, 沉闷
15. 乌黑, 稍卷, *
16. 乌黑, *, 浊响
17. 乌黑, *, 沉闷
18. 乌黑, *, *
19. *, 卷缩, 浊响
20. *, 卷缩, 沉闷
21. *, 卷缩, *
22. *, 稍卷, 浊响
23. *, 稍卷, 沉闷
24. *, 稍卷, *
25. , , 浊响
26. , , 沉闷
27. , , *
28. NULL

- 搜索空间

1. 删除与正例不一致

“

根据样例1，删除 2, 4, 5, 6, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 22, 23, 24, 26, 28 剩下 1,3,7,9,19,21,25,27。

2. 删除和负例一致的

“

根据样例4，删除27 剩下 1,3,7,9,19,21,25

- 得到版本空间

“

1,3,7,9,19,21,25

-
1. 与使用单个合取式来进行假设表示相比，使用“析合范式”将使得假设空间具有更强的表示能力。若使用最多包含k个合取式的析合范式来表达1.1的西瓜分类问题的假设空间，试估算有多少种可能的假设

“

解答:

- 计算假设空间

$$\text{space_size} = 3^4 + 1 = 49$$

我们已经知道版本空间

1. *, 卷缩, *
2. *, *, 清脆
3. *, 卷缩, 清脆

从版本空间覆盖样本项(泛化到具体,演绎的过程)

1. 的演绎有 $3^4 = 12$,
2. 的演绎有 $3^4 = 12$
3. 的演绎有 3

由于1 完全覆盖3, 并且 1, 2之间存在 3个完全相同项, 因此好瓜的项总计有

$$\text{good_space_size} = 12 + 12 - 3 = 21$$

并且完全由版本空间中的1, 2 两项决定

- 不考虑冗余情况 [组合拆分数的问题]

不考虑冗余, K最多为21个因此表达的项为

$$\sum_{i=0}^K$$

- 考虑冗余情况 [树的遍历组合问题]

-
1. 若数据包含噪声, 则假设空间中可能不存在与所有训练样本都一致的假设。在此情形下, 试设计一种归纳偏好用于假设选择

“

解答:

- 噪声去除, 将有噪声的项去除掉。
- 降低置信, 如果发现有冲突, 将训练数据的label(分类)去掉。同时计算最近距离最小的k个项, 将k个项的类别赋予该项。

-
1. 本章1.4节在论述“没有免费的午餐”定理时, 默认使用了“分类错误率”作为性能度量来对分类器进行评估。若换用其他性能度量, 试证明没有免费的午餐”定理仍成立

“

解答:

略

-
1. 试述机器学习在互联网搜索的哪些环节起什么作用

“

解答:

- 搜索引擎里面个性化展示

- 页面排序的rank
- 广告精准推荐