

西瓜书笔记

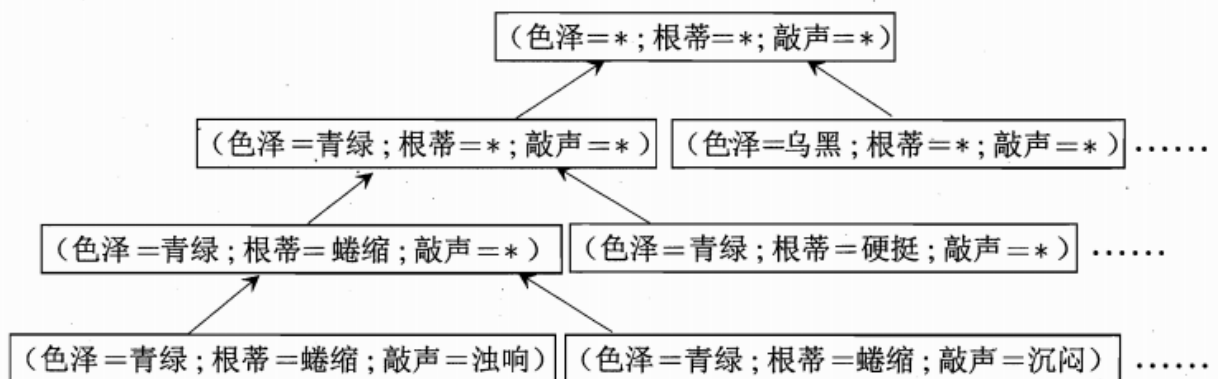
- [1.1 引言](#)
- [1.3 假设空间](#)
- [1.4 归纳偏好](#)

1.1 引言

- 机器学习致力于研究如何通过计算的手段，利用经验来改善系统自身的性能。
- 在计算机中，“经验”通常以数据形式存在，因此，机器学习主要研究在计算机上从数据中产生模型的算法，即“学习算法”
- 有了学习算法，我们把经验数据提供给它，它就能基于这些数据产生模型
- 在面对新的情况时，模型会给我们提供相应的判断

1.3 假设空间

- 归纳和演绎是科学推理的两大基本手段
- 我们可以把学习过程看作一个在所有假设组成的空间中进行搜索的过程，搜索目标是找到与训练集匹配的假设。
- 假设的表示一旦确定，假设空间及其规模大小就确定了
- 拿西瓜举例：若色泽、根蒂、敲声分别有3、2、2种可能取值，则我们面临的假设空间规模大小为 $4 \times 3 \times 3 + 1 = 37$ 中，其中1代表空集



- 可以有許多策略对这个假设空间进行搜索，例如自顶向下，从一般到特殊，或是自底向上、从特殊到一般，搜索过程中可以不断删除与正例不一致的假设，和与反例一致的假设，最终将会获得与训练集一致的假设

1.4 归纳偏好

- 任何一个有效的机器学习算法本身必带有归纳偏好，否则它将被假设空间看似在训练集上“等效”的假设所迷惑，而无法产生确定的学习效果。
- 那么，有没有一般性的原则来引导算法确立的“正确的”偏好呢？
“奥卡姆剃刀”：若有多个假设与观察一致，则选最简单的那个

- 在具体的现实问题中，算法的归纳偏好是否与问题本身匹配，大多数时候直接决定了算法能否取得好的性能
- “没有免费的午餐”定理:式 (1.2) 显示出，总误差竟然与学习算法无关！对于任意两个算法，我们都有：

$$\sum_f E_{ote}(\mathcal{L}_a|X, f) = \sum_f E_{ote}(\mathcal{L}_b|X, f) ,$$

- 所以，学习算法自身的归纳偏好与问题是否相配，往往会起到决定性的作用