

# 南大周志华《机器学习》课程笔记

Introduction: 最近自学机器学习课程, 注意到了南京大学周志华老师的课程。我是在学堂在线平台观看的, 注意到b站上也有相应视频, 但b站上并未获得授权, 随时有消失的可能。

周志华老师的网络教学视频中, 与其西瓜书相比确实少了一些内容。但幸运的是, 缺失的内容实际上对于初学者来说并不会产生太大影响。目前这一笔记也遵循视频内容, 相比西瓜书中也会有一些缺失, 敬请谅解。可能以后如果有机会和时间, 我会再阅读周志华老师的书籍将缺失内容补全。

一切内容敬请关注我的个人Page页面。

全系列笔记请见: [click here](#)

About Me: [点击进入我的Personal Page](#)

## 第一章 绪论

机器学习的定义: 利用经验改善系统自身的性能

计算学习理论中最重要的理论模型是概率近似正确(PAC)理论, 其解释为: 以很高概率得到很好的模型。

### 基本术语

术语一:

- 数据集: 训练、测试
- 示例(instance)、样例(example): 示例不给结果, 样例给结果
- 样本(sample)
- 属性(attribute)、特征(feature)、属性值
- 属性空间、样本空间、输入空间: 将每个属性想象成坐标上的一个轴, 这样可以形成一个向量
- 特征向量(feature vector)
- 标记空间、输出空间

术语二:

- 假设(hypothesis): 模型所学习到的
- 真相(ground-truth): 对的假设
- 学习器(learner)

术语三:

- 分类(输出离散)、回归(输出连续)
- 二分类、多分类
- 正类、反类 (二分类中, 通常假设正、负类是可交换的)

术语四:

- 监督学习(supervised learning): 有导师的学习, 拿到数据中有期望结果——分类、回归
- 无监督学习(unsupervised learning): 无导师的学习, 拿到数据中无期望结果——聚类、密度估计 (便于进行进一步处理)

术语五

- 未见样本(unseen instance)
- 未知"分布"
- 独立同分布(i.i.d)
- 泛化(generalization)

## 归纳偏好

归纳偏好指机器学习算法在学习过程中，对某种类型假设的偏好。

一般原则：奥卡姆剃刀(Occam's razor)——若非必要、勿增实体

任何一个有效的机器学习算法必然有其偏好，学习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！

利用“奥卡姆剃刀”原则时，确定哪个假设更“简单”，这个问题不是简单的。

例如：

$$y_1 = ax^2 + bx + c$$

$$y_2 = ax^3 + c$$

我们可以说 $y_1$ 最高只有两次，因此其简单；也可以说 $y_2$ 只有两项，因此其简单。

## NFL定理

No Free Lunch：一个算法 $L_a$ 若在某些问题上比另一个算法 $L_b$ 好，那么必然存在另一些问题上 $L_b$ 比 $L_a$ 好。

NFL定理的重要前提：所有“问题”出现的机会相同，或所有问题同等重要。但实际情况并非如此，我们通常只关注自己正在试图解决的问题。

因此，脱离具体问题，空泛地说“什么学习算法更好”是没有意义的。