1. **绪论**
   1. 引言

机器学习：通过经验（数据）得出模型（学习算法）

* 1. 基本术语

聚类：将训练集中的样本分组，如将西瓜分为“本地瓜”，“外地瓜”，这种样本通常不具有标记值

数据集，示例（样本，特征向量），属性（特征），属性值，维数（与属性有关），学习（训练），假设，真相（真实），学习器，预测，样例（示例加标记），标记空间（输出空间），分类（离散），回归（连续），聚类，监督学习，无监督学习，泛化

* 1. 假设空间

泛化：特殊到一般（归纳）

特化：一般到特殊（演绎）

归纳学习，概念学习，假设空间，版本空间（匹配假设空间集合）

* 1. 归纳偏好

从多个匹配假设中自己设定一个偏好，确定假设

选择一种正确的偏好，奥卡姆剃刀（选择最简单的假设），但什么叫简单？

所有算法的预测期望都是相同的（前提是所有问题同等重要），然而现实情况我们只关心自己要研究的问题，不同算法对于不同问题来说效率是不一样的

* 1. 发展历程

机器学习是人工智能发展的产物

智能发展流程：推理（推理机），知识（专家系统），自我学习

机器学习划分为：“样例中学习，”问题求解和规划中学习“，”指令中学习“，本书大部分都是从样例中学习

----

符号主义学习（有明确的概念）：决策树（常用），基于逻辑的学习

连接主义学习（黑箱）：神经网络（常用）

统计学习：支持向量机，核方法（常用），迁移学习，类比学习

------

深度学习，连接主义的一种，多层神经网络，火的原因是数据多，算力强

* 1. 应用现状

大数据时代的三大技术：机器学习，云计算，众包。

数据挖掘和机器学习间的联系：

数据挖掘对数据进行管理和分析，数据库技术提供管理，机器学习提供分析

1. **模型评估与选择**

2.1 经验误差与过拟合