

绪论

引言：问题

通过计算的手段，利用經驗來改善自身的性能
有了數據
通過某种学习算法
得到模型
進行预测

有了數據

数据集：100个西瓜

样本：1个西瓜

特征向量

样本空间

颜色、大小、敲起来的振幅（目前是3维向量）

维度

属性：颜色(特征向量的其中一个维度)

通過某种学习算法

学习

训练

有监督学习

二分类

$Y \in \{-1, 1\}$

瓜没熟，不摘；熟了才摘

多分类

Y 的可选值数量 > 2

市场上有各种瓜

$Y \in R$

回归

某段时间内，什么时候卖瓜最合适

聚类

我们并不知道要分几类，让机器自己分
每个组称为“簇”

无监督学习

测试

测试样本

泛化能力

進行预测

假设空间

科学的推理手段

归纳：特殊→一般

演绎：一般→特殊

狭义：从数据集得到概念

广义：从样本中学习

归纳偏好：

同一个数据集训练出了不同的模型，要如何选择？

原则：奥卡姆剃刀——选择最简单的

基本術語：

归纳偏好： 当多个假设都能解释已有的数据时，又该如何选择？
原则：奥卡姆剃刀——选择最简单的

發展歷程： 一种程序，拥有自我改善的能力，人为干预越少越好