

机器学习（西瓜书）

绪论

基本术语

引言

问题

- 通过计算的手段，利用经验来改善系统自身的性能
- 有了数据
- 通过某种学习算法
- 得到模型
- 进行预测

有了数据

- 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
- 维数

- 属性甜度

通过某种学习算法

- 学习
- 训练

有监督学习

分类

二元类

Y正负

瓜农眼中：这个瓜该不该摘，这个瓜熟没熟，我要不要摘



Y大于2

多分类

黑美人



小地雷



特小凤



市场上要买哪种西瓜

回归

Y=R实数集

某段时间内西瓜的价格，啥时间卖西瓜最合适



无监督学习

聚类

- 我们不知道要分为几类，机器自己分
- 每个组称为"簇", cluster

测试

进行预测

- 测试样本
- 泛化能力

假设空间

分支主题 2

分支主题 3