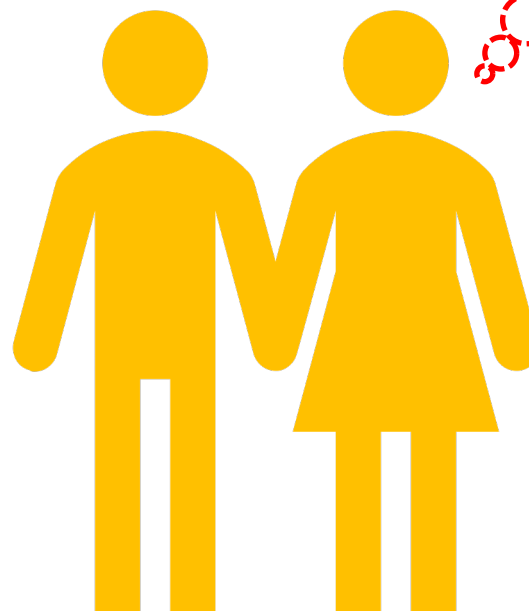
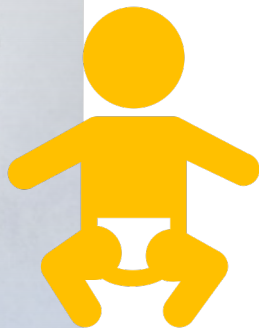
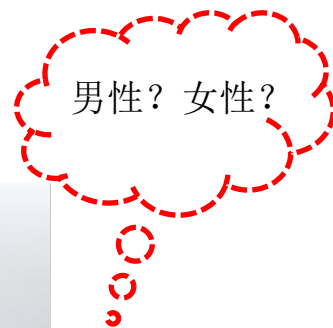


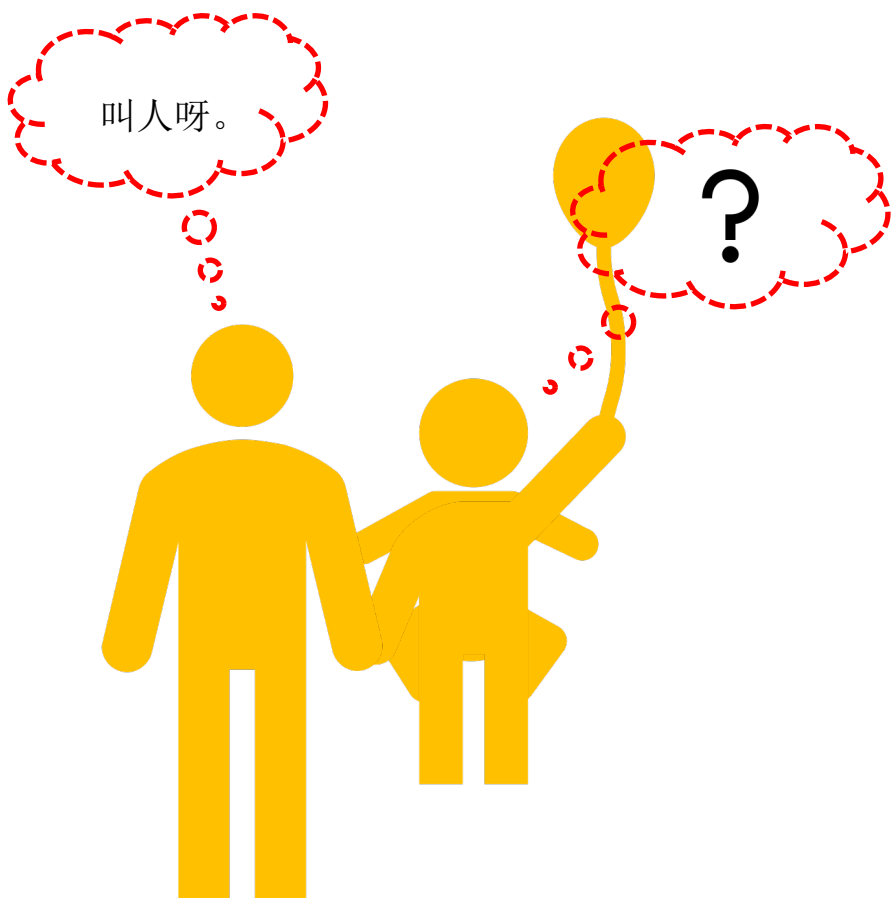
机器学习绪论



“学习”



“测试”



引言

微湿路面

感到和风

看到晚霞

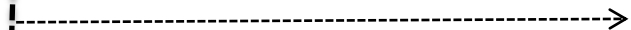
色泽青绿

根蒂蜷缩

敲声混响

-----> 生活经验 -----> 明天是好天

-----> 生活经验 -----> 好瓜



人的体验学习过程

“学习”



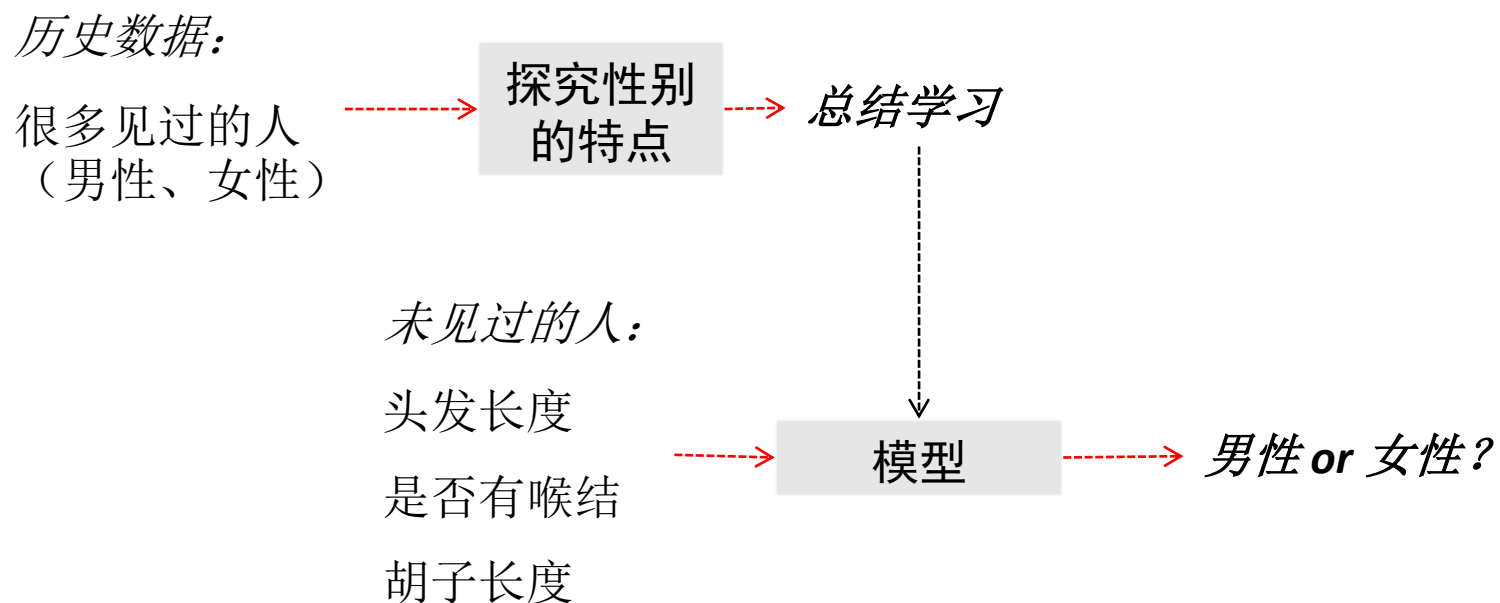
“测试”



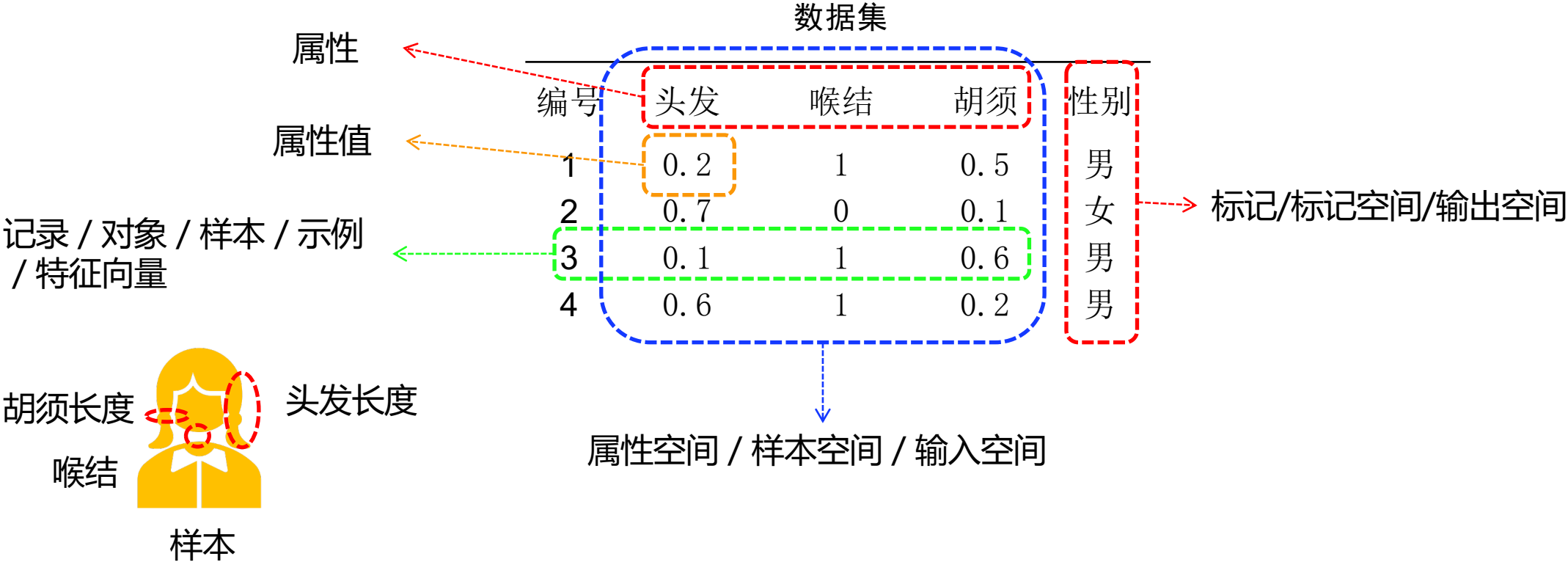
引言

机器学习的概念

- 致力于研究如何通过计算的手段，利用经验（历史数据）来改善系统自身的性能 [机器学习]。
- 从数据中产生模型的算法，即“学习算法”。

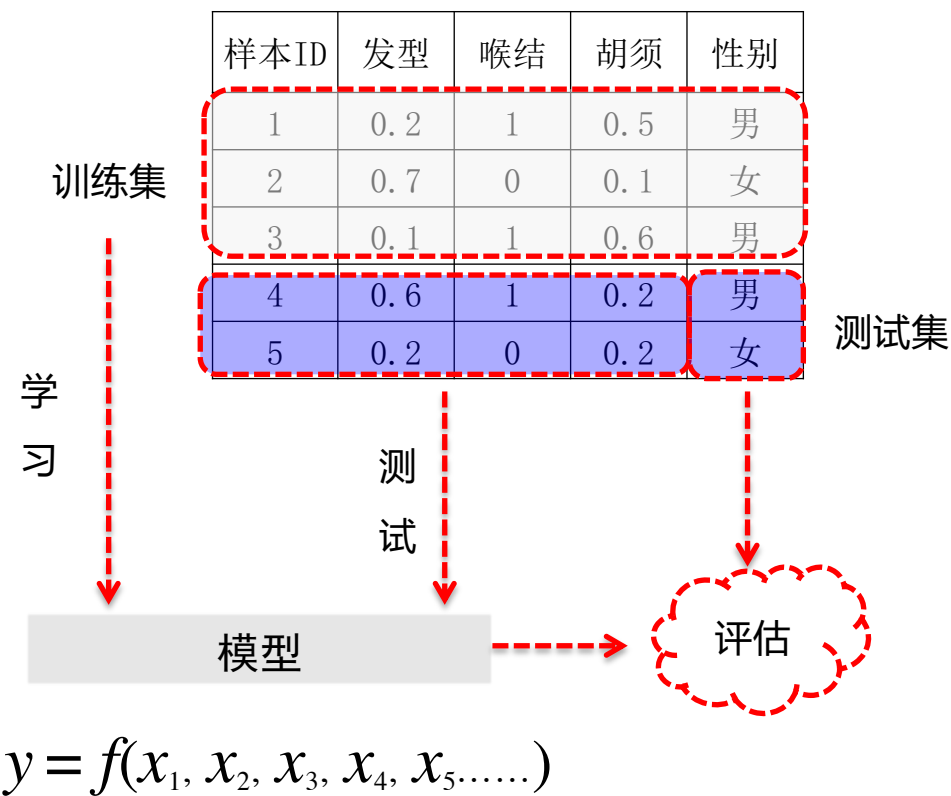


基本术语

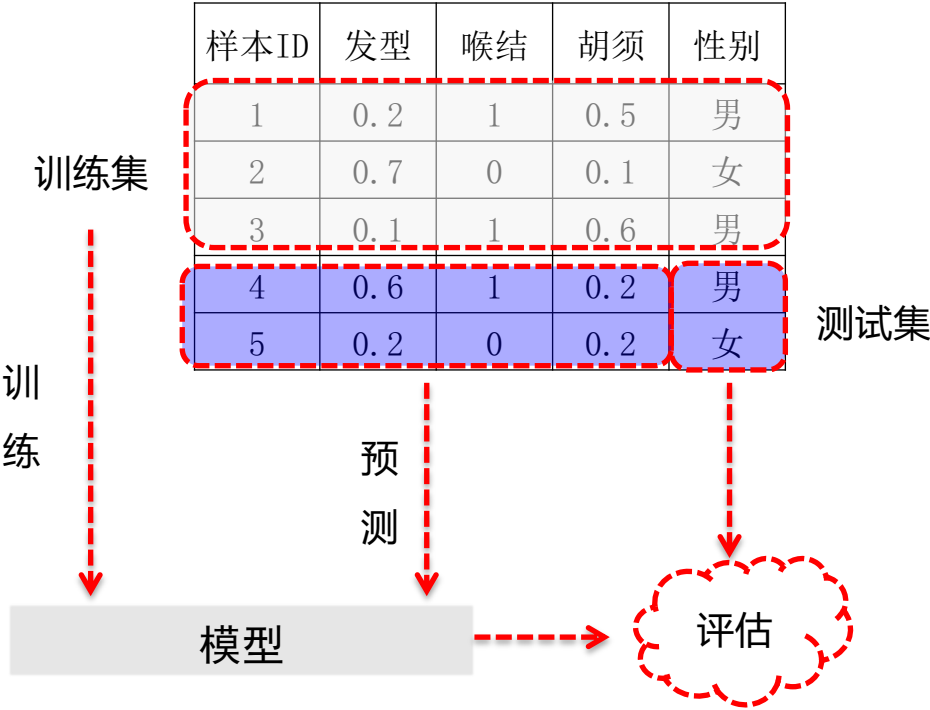
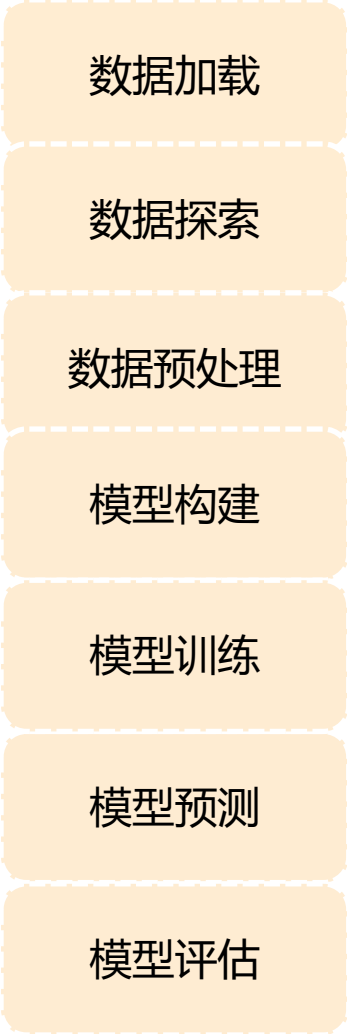


基本术语

- 学习（训练）：从数据中学得模型的过程
- 训练集：参与模型训练的样本集合
- 测试：学得模型后，使用其样本进行预测的过程
- 测试集：被预测的样本集合
- “假设”：学得模型对应的关于数据的某种潜在规律
- 泛化能力：学得模型适用于新样本的能力
- 独立同分布：样本空间的全体样本都服从一个未知的分布，且相互独立

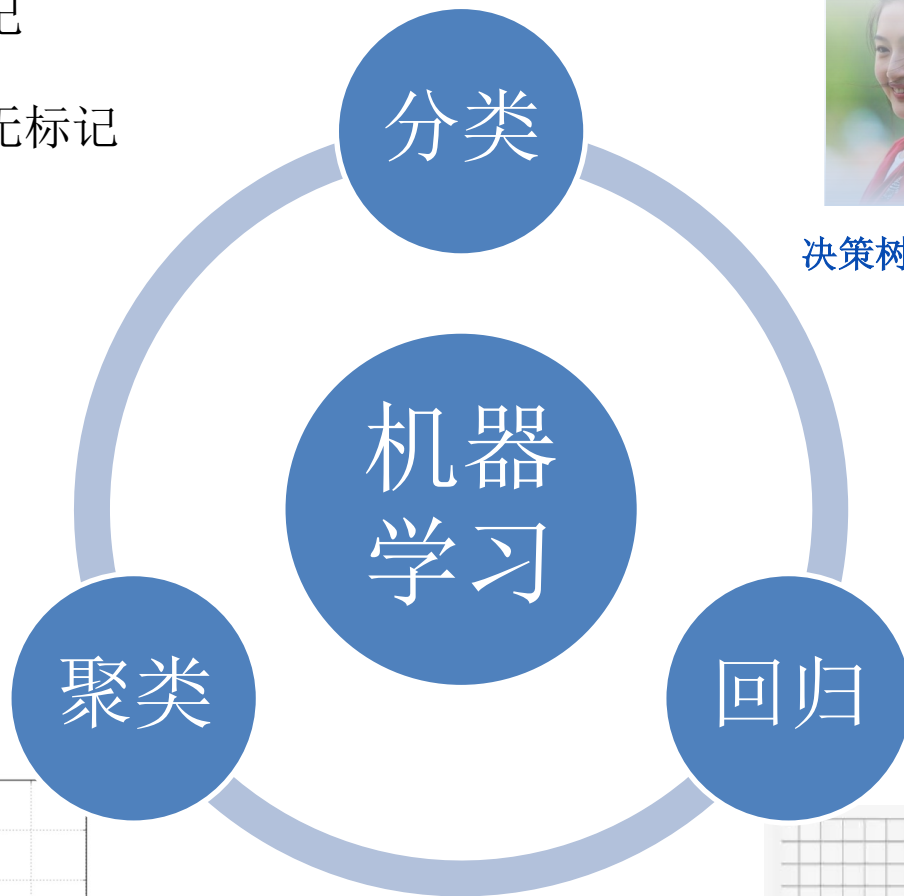


学习过程



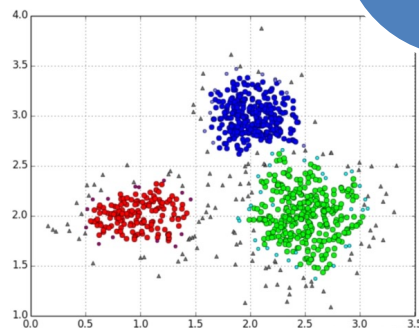
模型选择

- 监督学习 (Supervised learning) : 训练样本有标记
- 无监督学习 (Unsupervised learning) : 训练样本无标记
- 分类 (Classification) : 输出结果是离散值
- 回归 (Regression) : 输出结果是连续值
- 聚类 (Clustering) : 无标签的归类

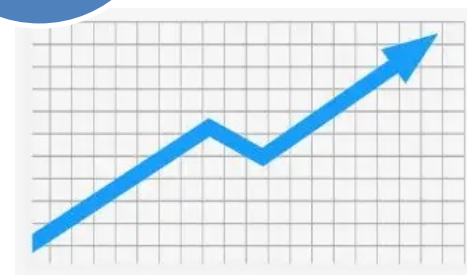


决策树、KNN、朴素贝叶斯

K-Means
K-Medoids
层次聚类

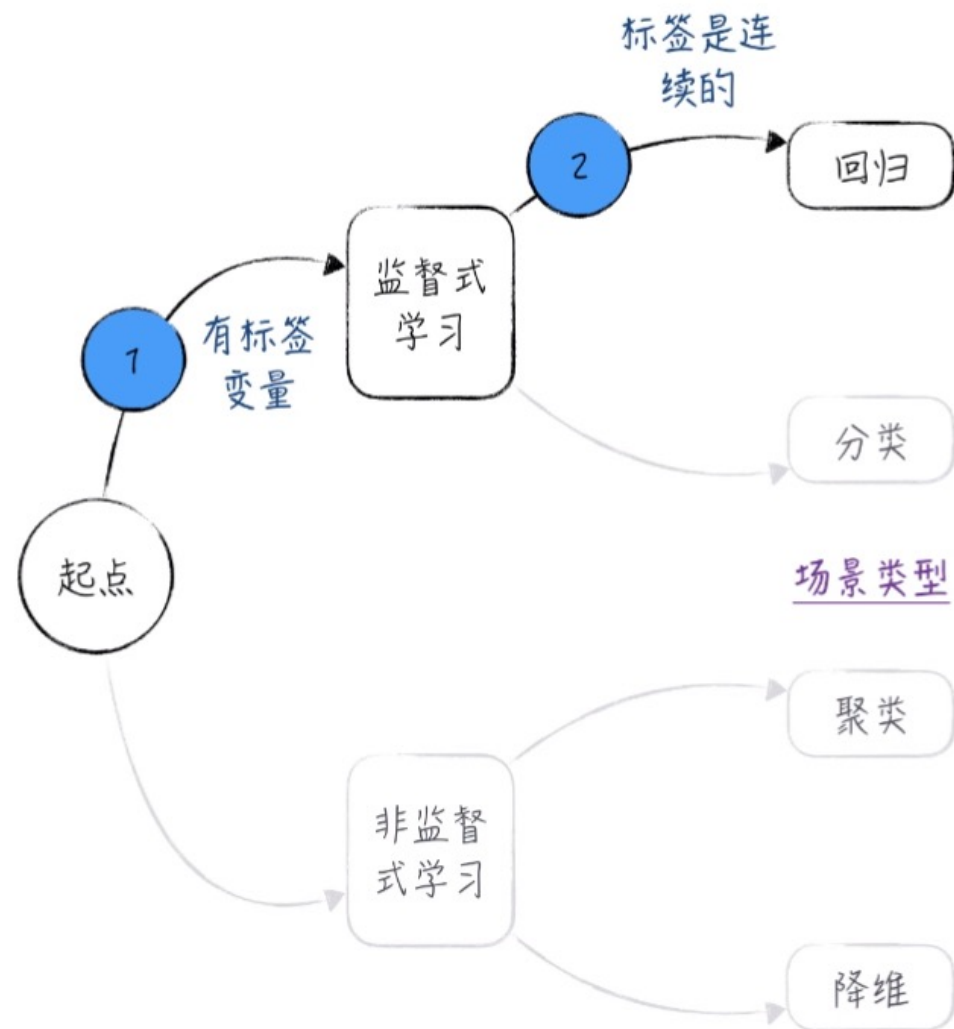


线性回归



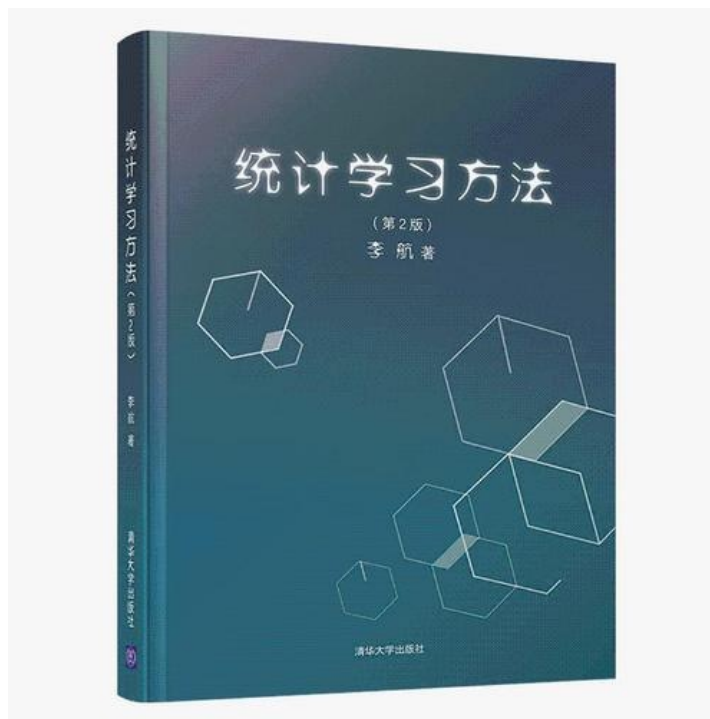
模型选择

- 监督学习 (Supervised learning) : 训练样本有标记
- 无监督学习 (Unsupervised learning) : 训练样本无标记
- 分类 (Classification) : 输出结果是离散值
- 回归 (Regression) : 输出结果是连续值
- 聚类 (Clustering) : 无标签的归类



推荐图书

- 吴恩达: [机器学习视频](#)





Thank you!