

MACHINE LEARNING

机器学习

LESSON 1

基本介绍（绪论）

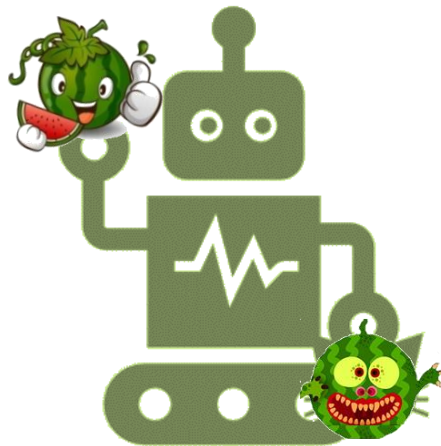


参考：
《机器学习》

Machine Learning Course
Copyright belongs to Wenting Tu.

什么是“机器学习”

- “经验”通常以“数据”存在
- 从数据中产生“模型”
- 模型对“新情况”进行“判断”



识瓜**APP**, 瓜迷用过都说好!

什么是“机器学习”

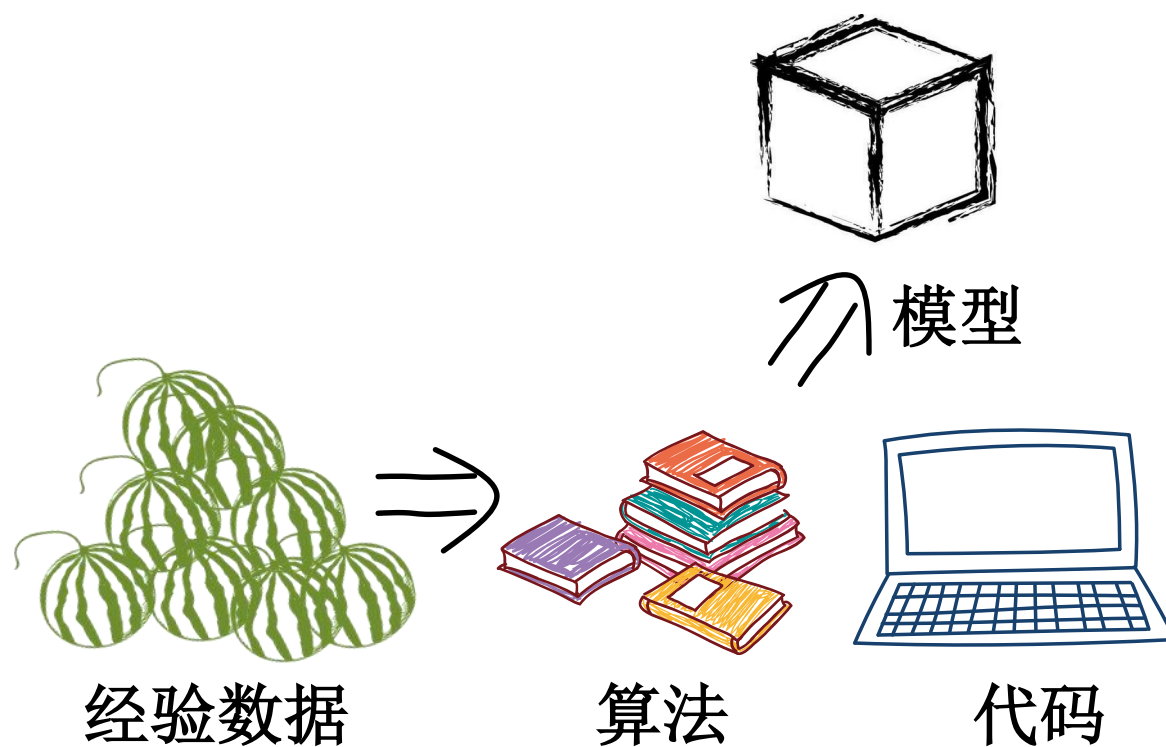
- “经验”通常以“数据”存在
- 从数据中产生“模型”
- 模型对“新情况”进行“判断”



经验数据

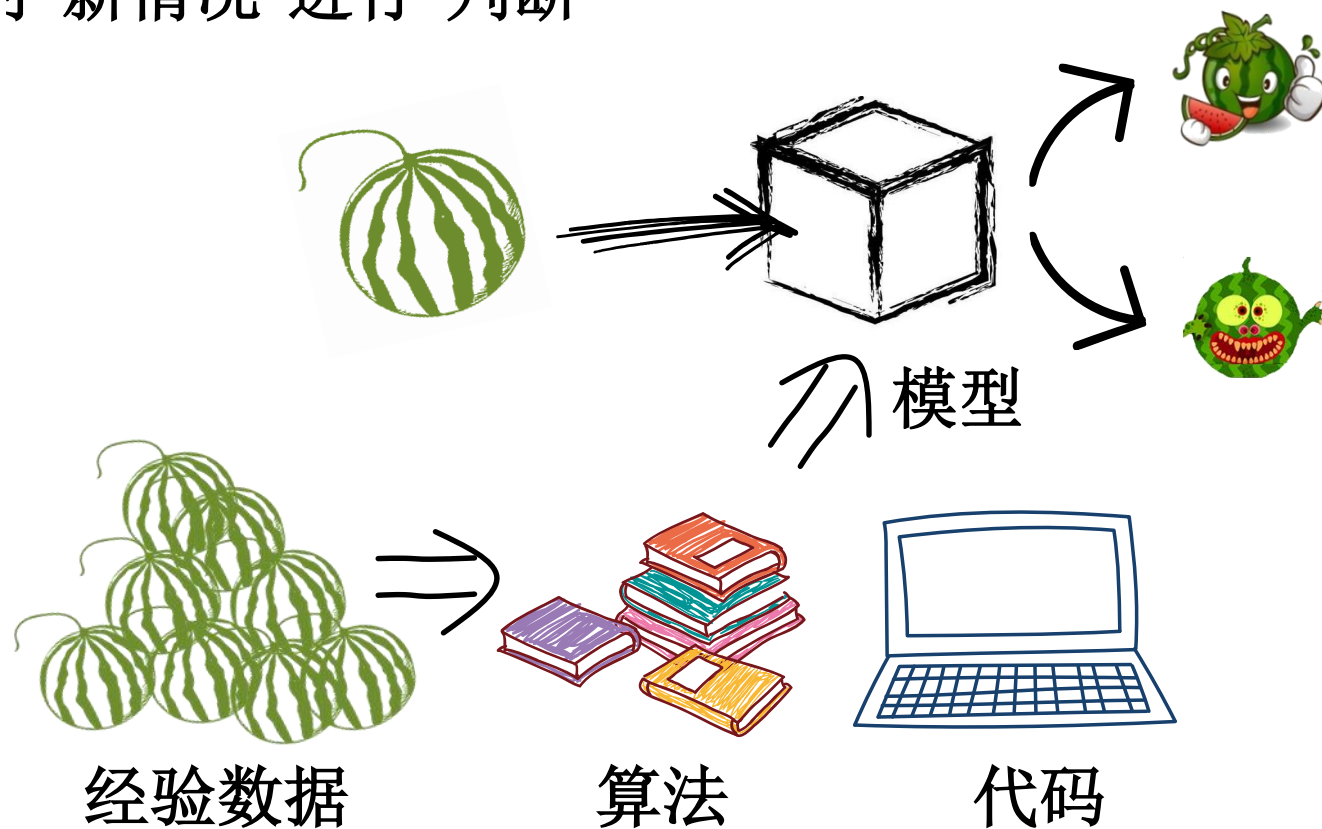
什么是“机器学习”

- “经验”通常以“数据”存在
- 从数据中产生“模型”
- 模型对“新情况”进行“判断”



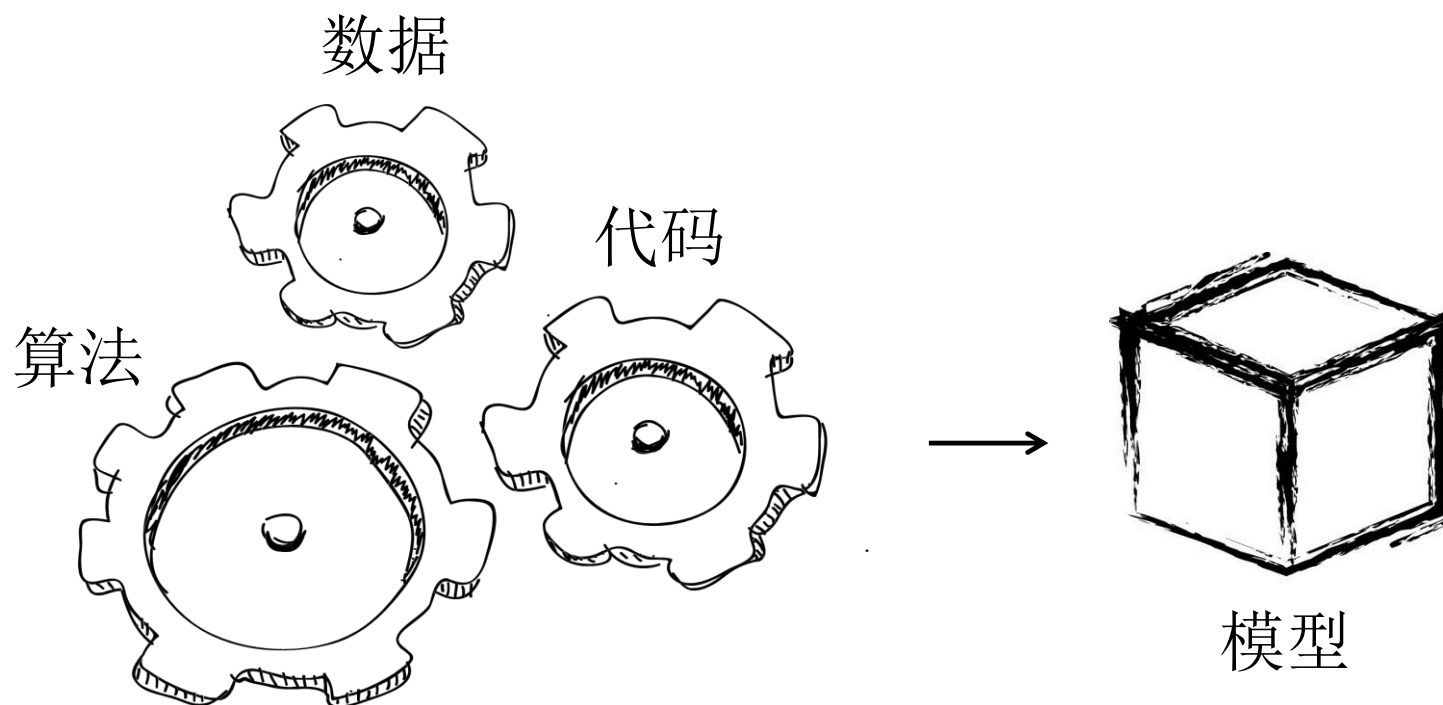
什么是“机器学习”

- “经验”通常以“数据”存在
- 从数据中产生“模型”
- 模型对“新情况”进行“判断”



什么是“机器学习”

- “经验”通常以“数据”
- 从数据中产生“模型”
- 模型对“新情况”进行“判断”



什么是“机器学习”

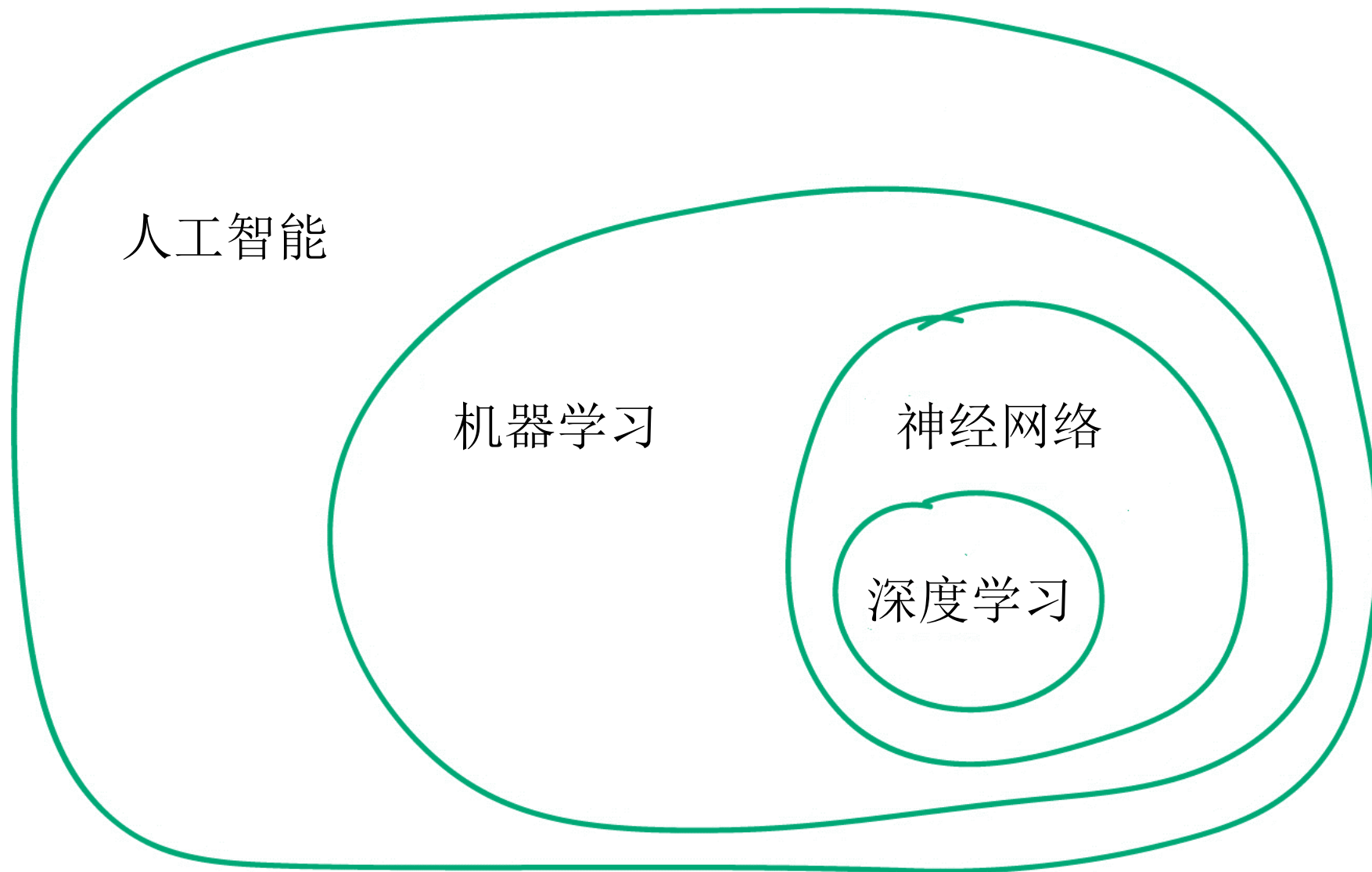
- “经验”通常以“数据”

假设用评估准则 P 来评估计算机程序在某任务 T 上的性能, 若一个程序通过利用经验 E 在任务 T 上获得了性能改善, 则我们就说关于 T 和 P , 该程序对 E 进行了(机器)学习。

A computer program is said to learn from *experience E* with respect to some class of *tasks T* and performance *measures P* , if its performance at *tasks in T* , as *measured by P* , improved with *experience E* . - (Mitchell, 1997)



什么是“机器学习”



基本术语

- “数据集”
- “示例” (instance) 或“样本”(sample)
- “属性” (attribute)或“特征”(feature)



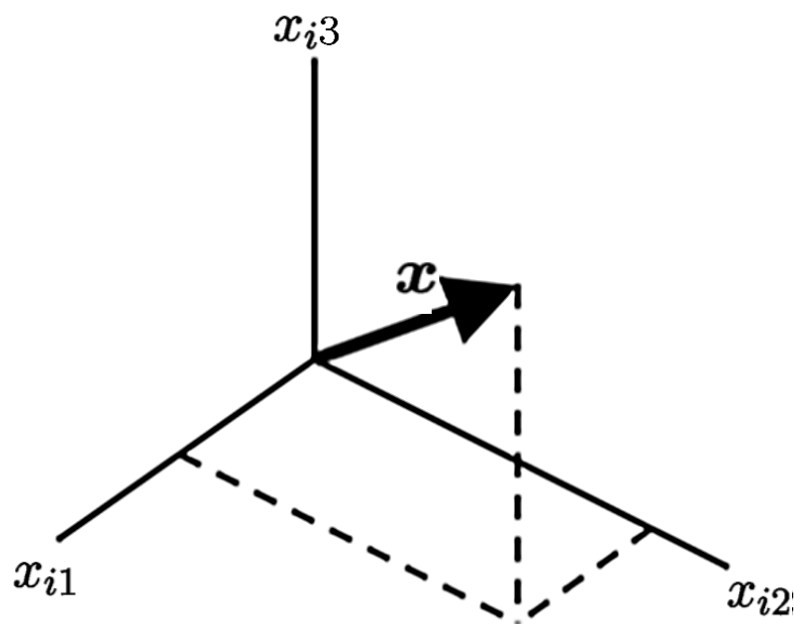
基本术语

- “特征向量”(feature vector)
- “属性空间”(attribute space)、“样本空间”(sample space)或“输入空间”

$$D = \{\mathbf{x}_1, \mathbf{x}_1, \dots, \mathbf{x}_m\}$$

$$\mathbf{x}_i = (x_{i1}; x_{i2}; \dots; x_{id})$$

$$\mathbf{x}_i \in \mathcal{X}$$



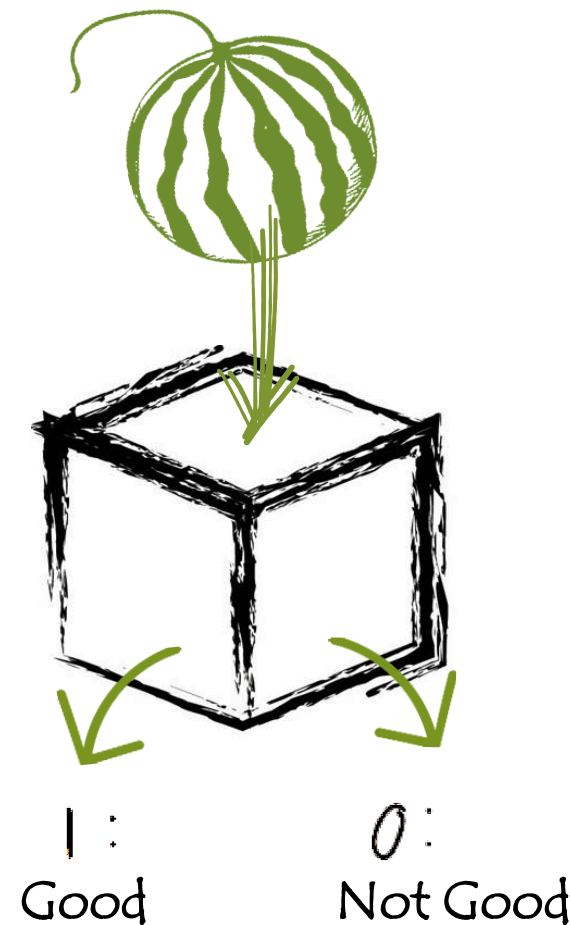
基本术语

- “学习”(learning)或“训练”(training)
- “训练样本”(training sample)
- “训练集”(training set)



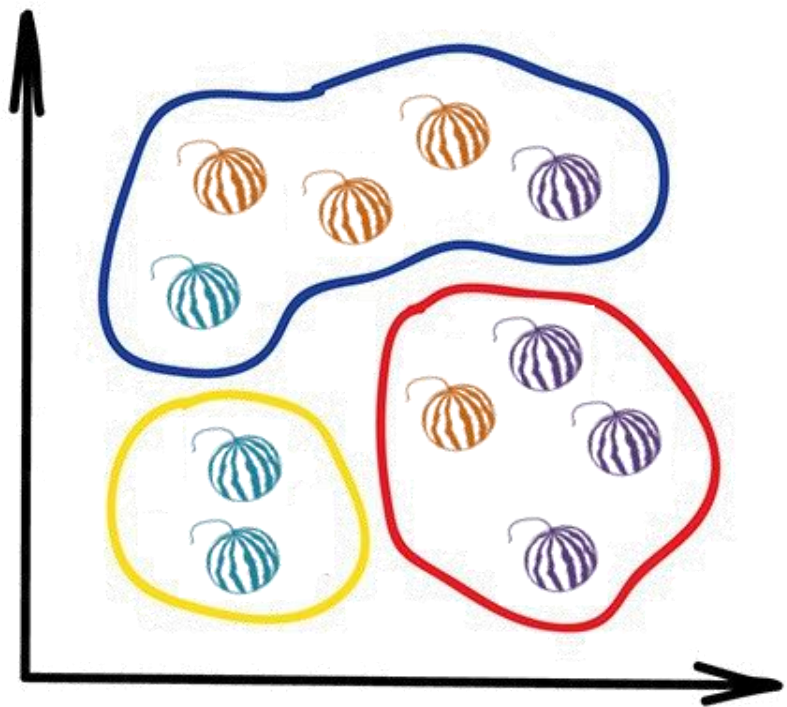
基本术语

- “标记”(label)
- “样例” (example)
- “标记空间”(label space)或“输出空间”
- “测试” (testing)
- “测试样本”(testing sample)
- “泛化” (generalization)能力



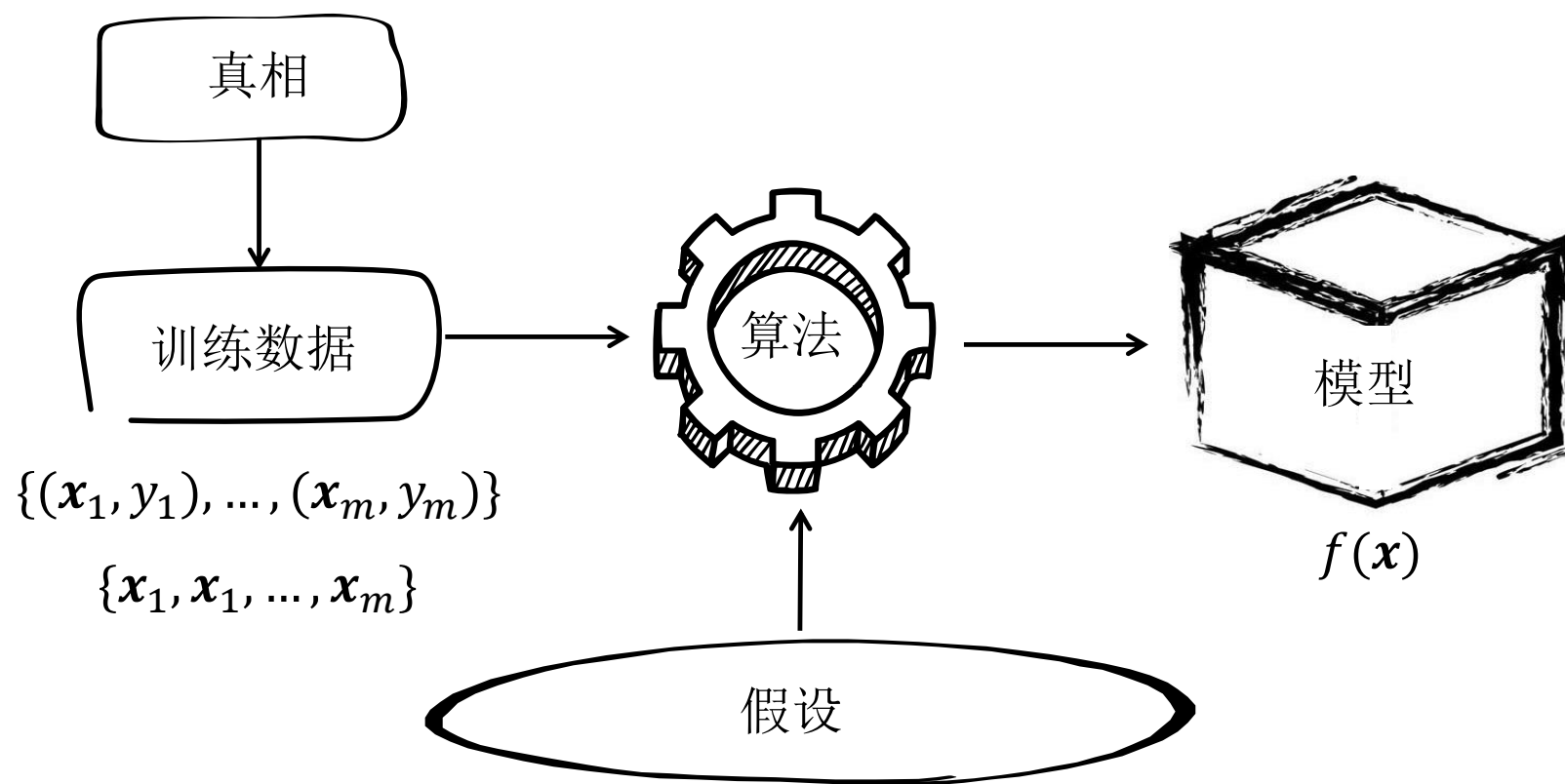
基本术语

- “分类”(classification)
- “回归”(regression)
- “聚类”(Clustering)



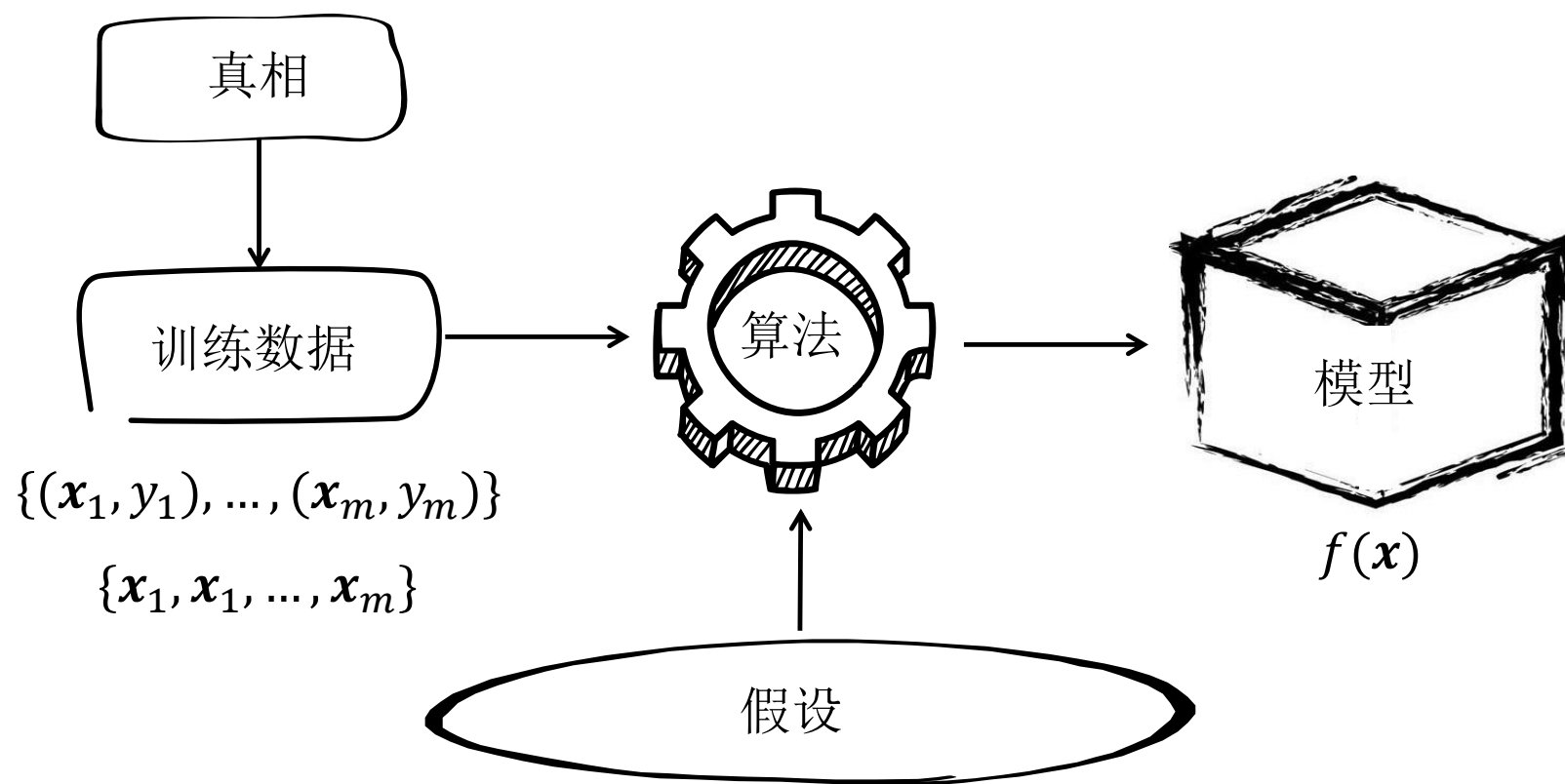
基本术语

- “假设”(hypothesis)
- “真相”或“真实”(ground - truth)



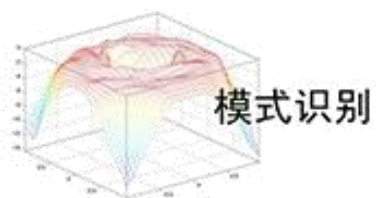
基本术语

- “监督学习”(supervised learning)
- “无监督学习”(unsupervised learning)



应用

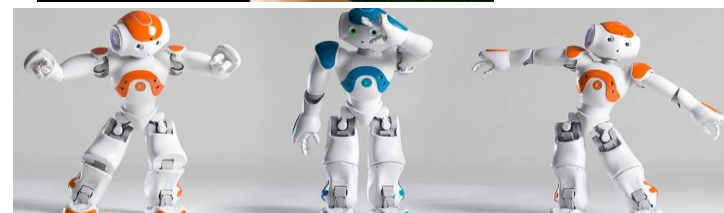
• 理论与应用



数据挖掘



语音识别



统计学习



应用

- 科技兴国



(十五) 计算机服务业45.在信息处理技术（编号：056101X）项下增加控制要点：“17.语音合成技术（包括语料库设计、录制和标注技术，语音信号特征分析和提取技术，文本特征分析和预测技术，语音特征概率统计模型构建技术等）18.人工智能交互界面技术（包括语音识别技术，麦克风阵列技术，语音唤醒技术，交互理解技术等）19.语音评测技术（包括朗读自动评分技术，口语表达自动评分技术，发音检错技术等）20.智能阅卷技术（包括印刷体扫描识别技术，手写体扫描识别技术，印刷体拍照识别技术，手写体拍照识别技术，中英文作文批改技术等）21.基于数据分析的个性化信息推送服务技术”。

应用

• 金融领域

- 信贷风控
- 精准营销
- 智能投顾
- 产品定价
- 电子支付
- ...

Credit Score



AI开发人员



专业投顾

线下服务

传统投资推荐



人工智能

云计算

数据挖掘

机器学习

Web服务

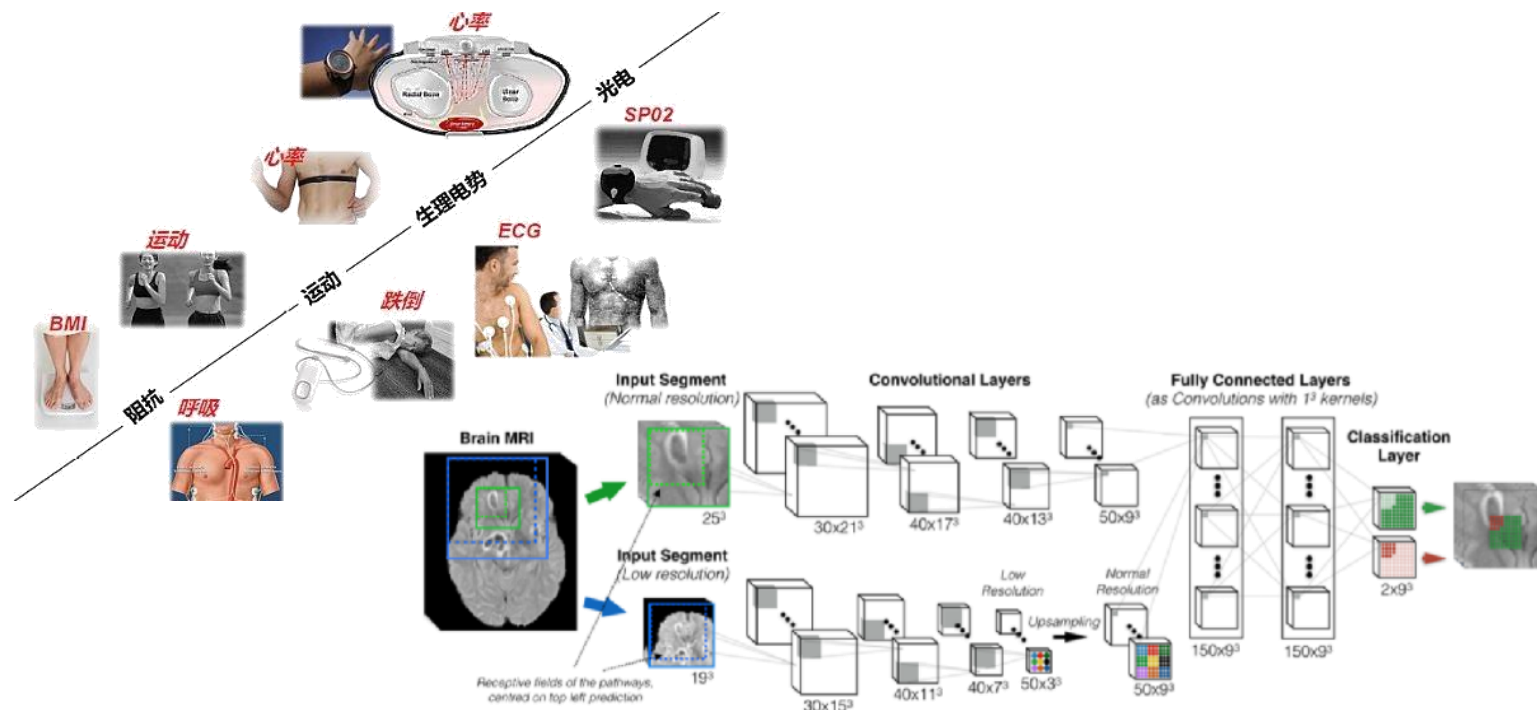
大数据
智能化
自动化
投资推荐



投资者/用户

• 医疗领域

- 医学影像识别
- 个性化治疗
- 智能医疗咨询
- 药物效果评估
- 基因编辑
- ...



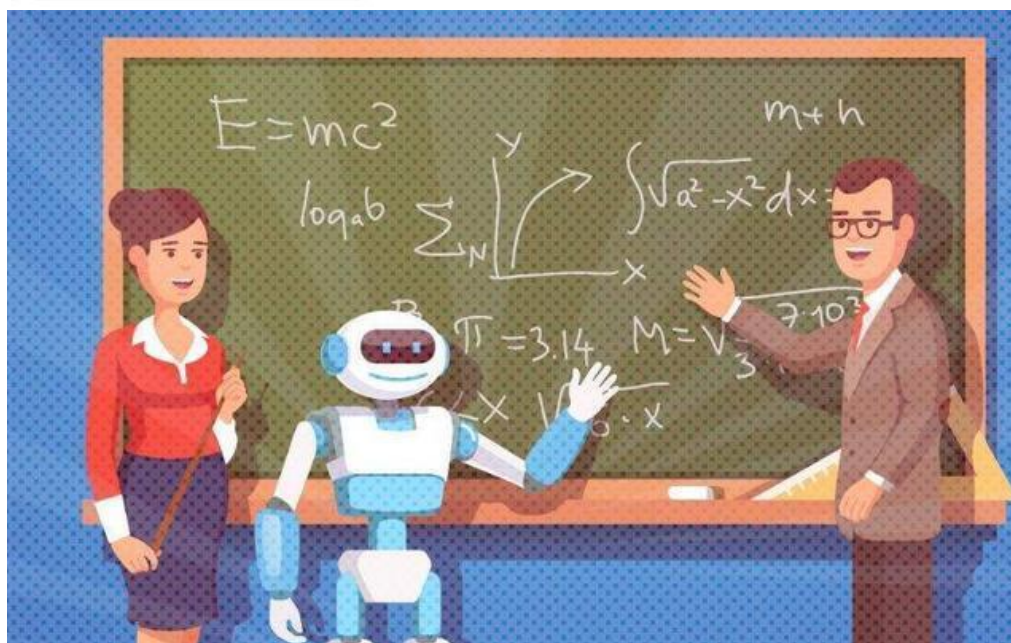
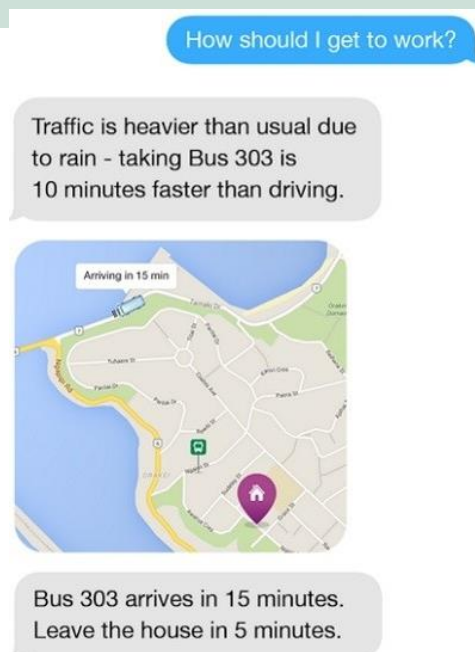
应用

• 交通领域

- 车牌号码识别
- 路径规划
- 航班晚点率预测
- 智能交通疏导
- 自动驾驶
- ...

• 教育领域

- 教学水平监测
- 学习诊断与预警
- 教育资源配置
- 学生发展预测
- 机器人辅导
- ...



MACHINE LEARNING

实践

Practice

编程平台介绍



Machine Learning Course
Copyright belongs to Wenting Tu.

安装指导

- 已安装Python

<https://scikit-learn.org/stable/install.html>

- 未安装Python

– 安装Anaconda（极力推荐）

<https://www.anaconda.com/products/individual>

课程作业编写工具

- Jupyter Notebook

<https://jupyter-notebook.readthedocs.io/en/stable/index.html>

程序示例

- 西瓜分类程序
 - preprocessing_example.ipynb
 - classification_example.ipynb
- 通过sklearn在线文档可以获得更多练习，例如
https://scikit-learn.org/stable/auto_examples/index.html#classification
(可留待后续章节讲完后再看)