

机器学习原理与技术

第一章 绪论

软件与物联网工程学院

教师：张翰进

2024年3月4日

前导课程

必备基础课程：

- 线性代数
- 概率论与数理统计
- 微积分

如果有以下课程基础更好：

- 矩阵论
- 数值分析（数值计算方法）
- 最优化理论（或运筹学）
- 随机过程
- 多元统计学



什么是机器学习?

- 人们可以根据以往的经验对将要发生的事件作出判断和反应
- 机器学习是让计算机也获得这样的能力

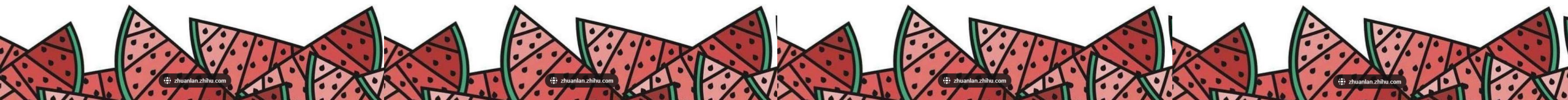
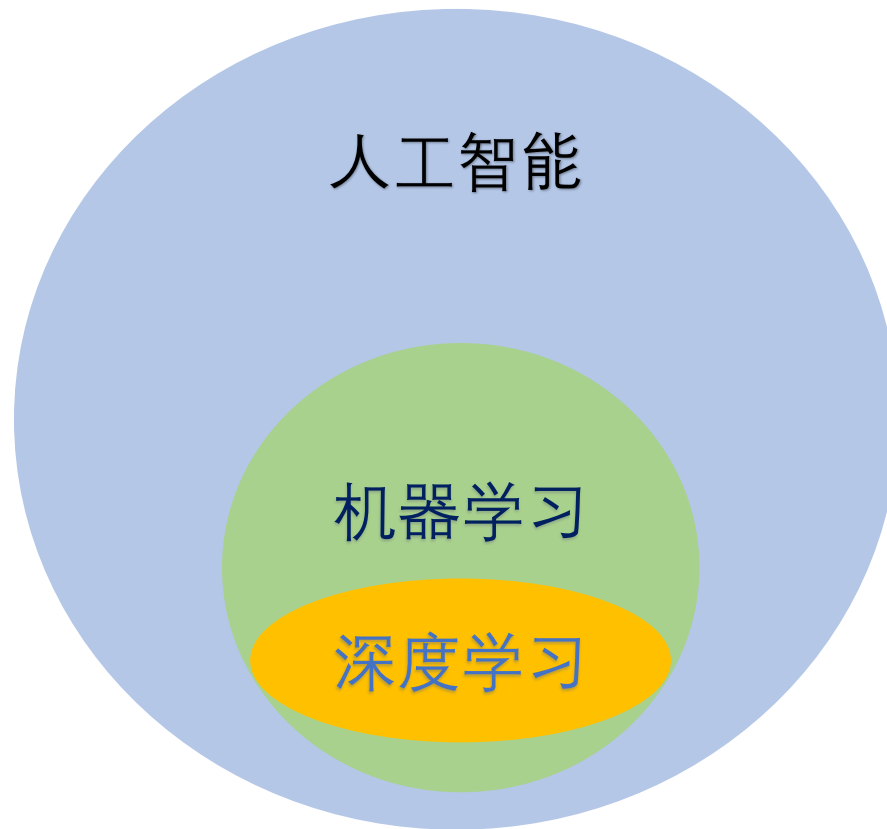
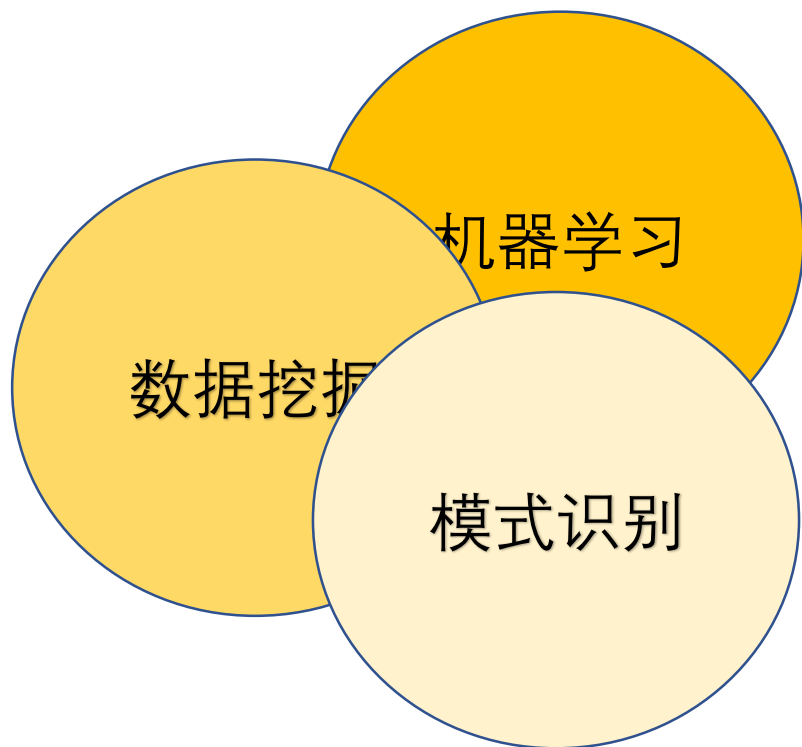
CMU Tom Mitchell对机器学习的定义:

假设用性能度量 P 来评估模型在某类任务 T 中的性能, 若该模型通过利用经验 E 在任务 T 中改善其性能 P , 那么机器通过经验 E 进行了学习。

一般认为, 机器学习是计算机从大量的数据中通过学习获得较为合理的算法模型, 并利用这种模型预测和分析新数据的过程。



关联学科



发展历程 (I)

推理期

仅有逻辑推理能力是不够的 ...

- **1956-1960s: Logic Reasoning**
- “数学家真聪明”
- 自动定理证明系统 (例如, 西蒙与纽厄尔的“Logic Theorist”系统)

知识期

要总结出知识“教”给系统, 实在太难...

- **1970s -1980s: Knowledge Engineering**
- “知识就是力量”
- 专家系统 (例如, 费根鲍姆等人的“DENDRAL”系统)

学习期

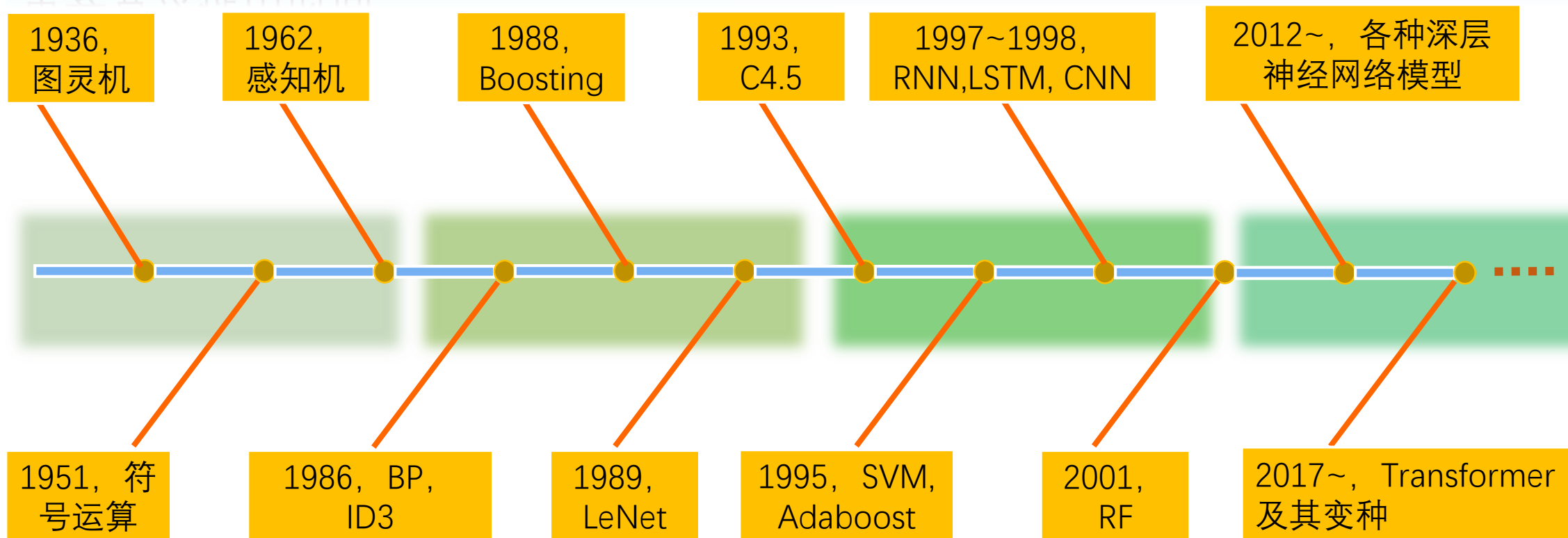
海量的数据驱动机器自动分析...

- **1990s -now: Machine Learning**
- “让系统自己学习”
- 海量的学习算法: SVM, RF, GBDT, HMM, AdaBoost, CNN, RNN.....



发展历程（II）

重要算法提出时间



应用现状和广阔前景

计算机视觉
生物学图像
遥感图像
工业探伤
安全防护
能源矿产

图像相关的应
用

序列相关的应
用

时间序列分析
语音信号处理
生物信息学

机器学习

自然语言处理
机器翻译
信息检索
文本生成

文本相关的应
用

多种数据综合
应用

智能游戏玩家
无人驾驶
影音娱乐



基本术语

- 样本（样本点）和样本空间
 - 一条带标记的数据记录 (X, y) ，或不带标记的数据记录 X ，表示样本集 S 中的一个样本（或样本点），可记为 (X_i, y_i) 或 X_i ，则样本集可以表示为 $S = \{(X_1, y_1), (X_2, y_2), \dots, (X_N, y_N)\}$ 或 $S = \{X_1, X_2, \dots, X_N\}$
 - 所有样本可能的取值构成的集合称为样本空间
- 特征和标记
 - 样本的属性，也就是样本中 X 的一个元素，称为样本 X 的一个特征，可以记为 $x^{(i)}$ ，则 $X = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$ ， n 称为特征的维度
 - 样本的类别或者取值 y 称为样本的标记
- 训练集和测试集（仅监督学习存在）
 - 用来学习得到模型的数据组成的集合称为训练集
 - 用来检验模型的泛化性能的数据组成的集合称为测试集

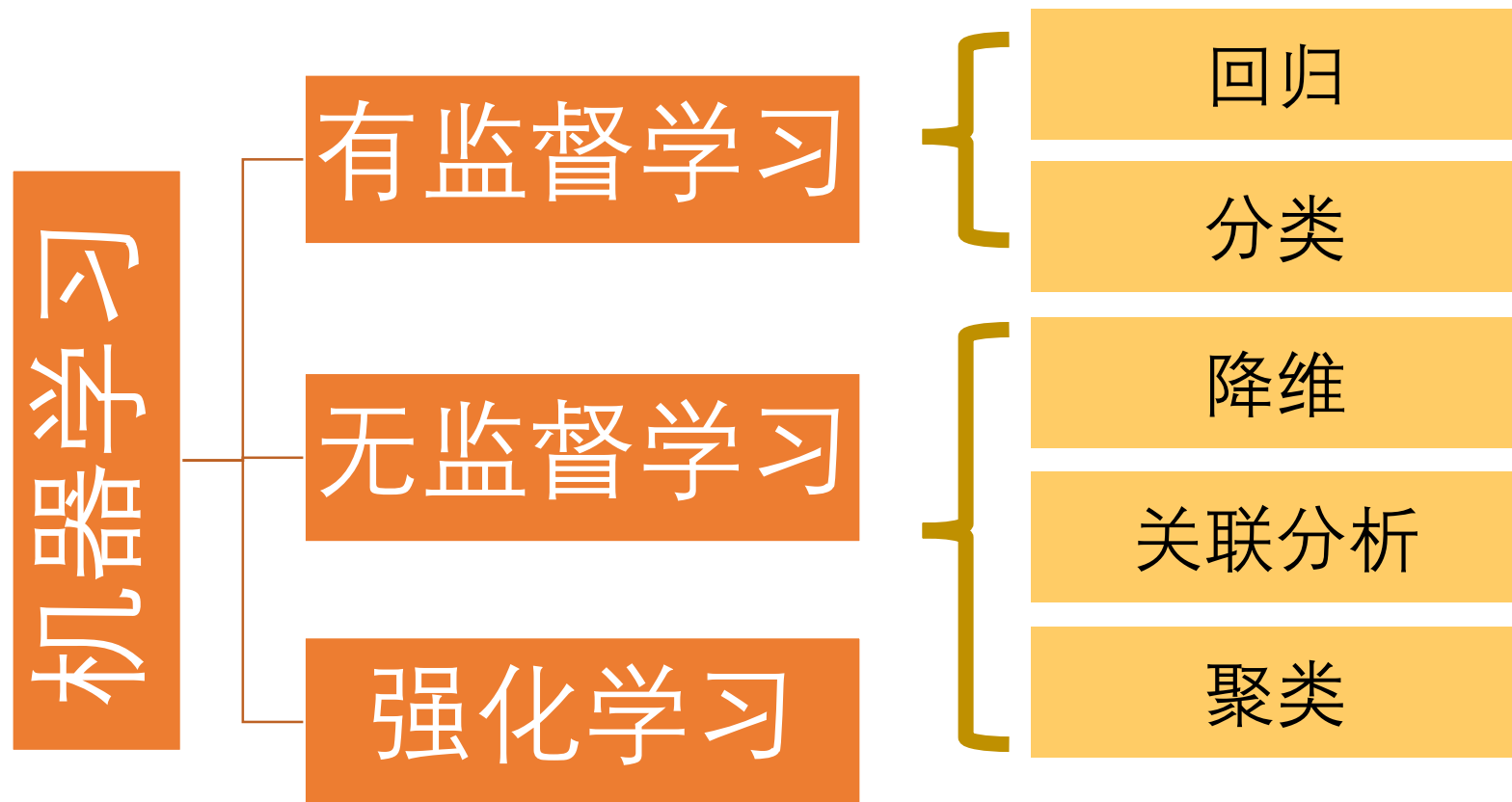


机器学习的一般流程

- 1. 得到一个有限训练数据集;
- 2. 确定包含所有可能模型的估计, 即学习模型的集合;
- 3. 选择适当的模型;
- 4. 确定求解模型的算法;
- 5. 求解最优模型参数;
- 6. 利用学习到的最优模型对新数据进行预测和分析。

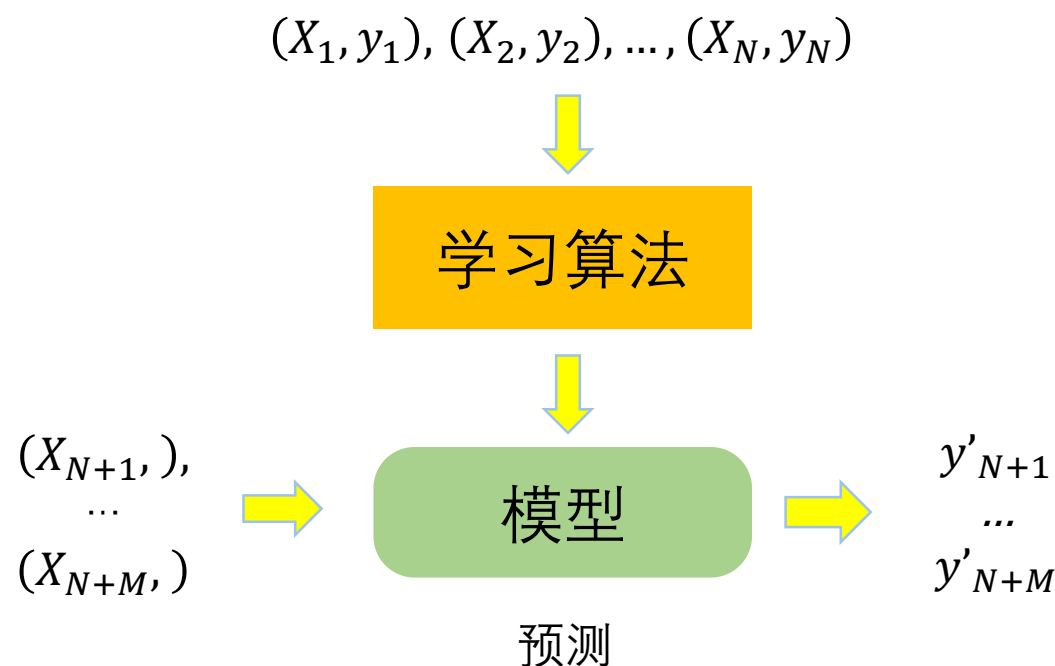


机器学习的分类



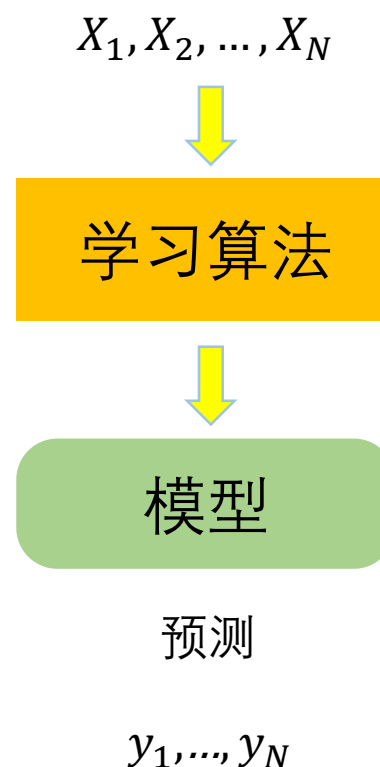
监督学习

- 从带有标记的训练数据中获得模型，并利用学习到的模型预测新数据的过程
- 有明显的训练集和测试集，部分有验证集
- 训练数据均有标记
- 根据标记 y 的取值，可以分为分类和回归问题



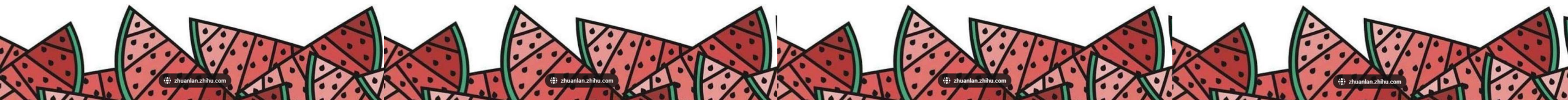
无监督学习

- 直接从无标记的数据学习得到模型
- 一般无明显的训练集和测试集
- 通常用来做聚类或者降维（降低特征的维度）
- 无监督学习通常是从数据中获取其统计规律，便于用户理解数据



强化学习

- 在行动中学习
- 不需要输入和标记，而是根据环境对智能体在不同状态 S_t 下的行为 a_t 进行奖惩 r_t （类似控制系统的反馈），逐步更新智能体，达到最优性能的目的



两个定理

- “奥卡姆剃须刀”

“当有多个假设与观察一致，则选择最简单的那个”

- “没有免费的午餐”

“一般情况下，学习算法的期望性能相同”

“算法获得的模型，在假设均匀分布情况下，误差和相同”



本章回顾：

机器学习的概念

发展历程和前景

机器学习的流程

机器学习的类别

