



机器学习

第一讲 绪论

魏凤凤 助理教授

fengfeng_scut@163.com

计算机科学与工程学院

课程要求



- 平时成绩：20%
- 大作业：20%
- 期末考试（闭卷）：60%



华南理工大学
South China University of Technology

什么是机器学习?

What is Machine Learning (ML)?



Machine Learning

什么是机器学习

- 机器学习 (Machine Learning, ML) 是**人工智能**的一个分支。
- **机器学习理论**主要是设计和分析一些**让计算机可以自动“学习”**的算法。
- **机器学习算法**是一类**从数据中自动分析获得规律，并利用规律对未知数据进行预测**的算法。

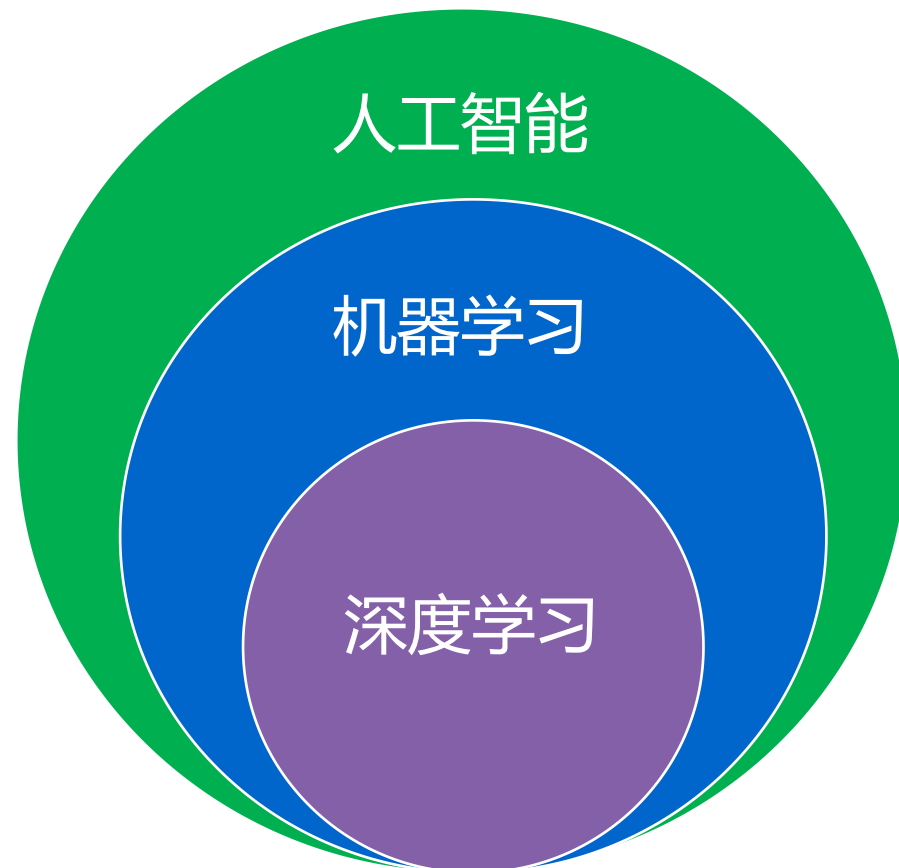
*[机器学习 - 维基百科，自由的百科全书](#)

机器学习与人工智能、深度学习的关系

人工智能：机器展现的人类智能

机器学习：计算机利用已有的数据(经验)，得出某种模型，并利用此模型预测未来的一种方法

深度学习：实现机器学习的一种技术



机器学习的发展

机器学习源自人工智能

1956年夏 美国达特茅斯学院

J. McCarthy, M. Minsky, N. Lochester, C. E. Shannon,
H.A. Simon, A. Newell, A. L. Samuel 等10余人

达特茅斯会议标志着人工智能这一学科的诞生

John McCarthy (1927 - 2011):

1971年获图灵奖, 1985年获IJCAI终身成就奖。人工智能之父。他提出了“人工智能”的概念, 设计出函数型程序设计语言Lisp, 发展了递归的概念, 提出常识推理和情境演算。出生于共产党家庭, 从小阅读《10万个为什么》, 中学时自修CalTech的数学课程, 17岁进入CalTech时免修两年数学, 22岁在Princeton获博士学位, 37岁担任Stanford大学AI实验室主任。



John McCarthy
(1927-2011)
“人工智能之父”
1971年图灵奖

机器学习的发展

第一阶段：推理期

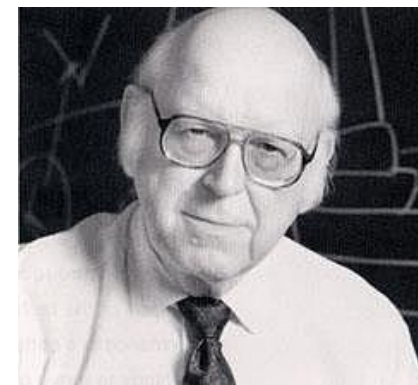
1956-1960s: Logic Reasoning

- ◆ 出发点: “数学家真聪明!”
- ◆ 主要成就: 自动定理证明系统
(例如, 西蒙与纽厄尔的 “Logic Theorist” 系统)

渐渐地, 研究者们意识到, 仅有逻辑推理能力是不够的 ...



赫伯特 西蒙
(1916-2001)
1975年图灵奖



阿伦 纽厄尔
(1927-1992)
1975年图灵奖

机器学习的发展

第二阶段：知识期

1970-1980s: Knowledge Engineering

- ◆ 出发点: “知识就是力量!”
- ◆ 主要成就: 专家系统
(例如, 费根鲍姆等人的 “DENDRAL” 系统)



爱德华 费根鲍姆
(1936-?)
1994年图灵奖

渐渐地, 研究者们发现, 要总结出知识再 “教” 给系统, 实在太难了 ...

机器学习的发展

第三阶段：学习期

1990s-now: Machine Learning

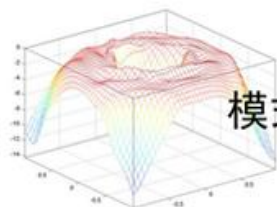
- ◆ 出发点: “让系统自己学!”
- ◆ 主要成就:

机器学习是作为“突破知识工程瓶颈”之利器而出现的

恰好在20世纪90年代中后期，人类发现自己淹没在数据的汪洋中，对自动数据分析技术——机器学习的需求日益迫切



机器学习已“无处不在”



模式识别

计算机视觉



数据挖掘



机器学习



语音识别



统计学习



自然语言处理

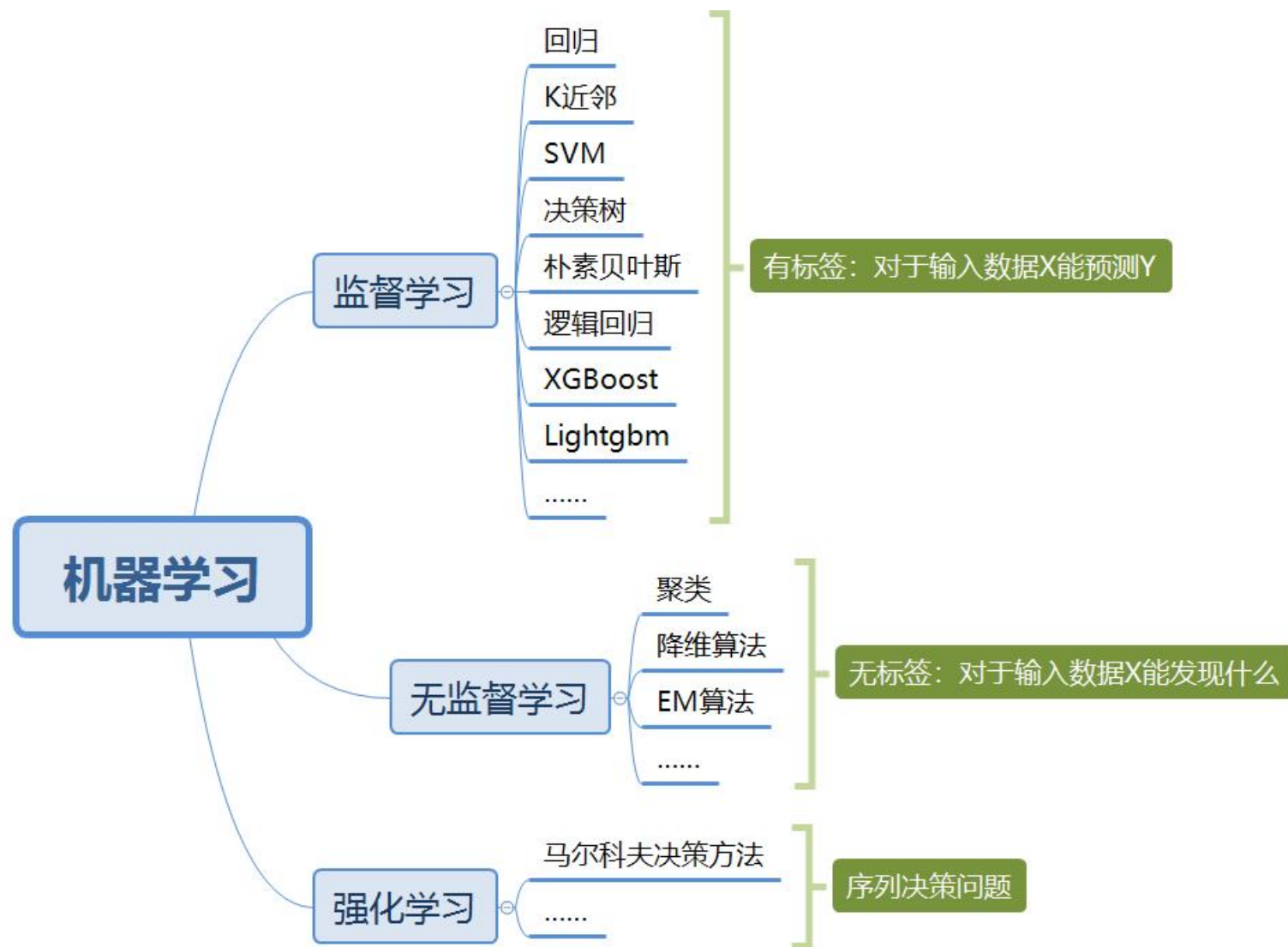


机器学习的应用



机器学习的范围

机器学习的类型



机器学习的类型

监督学习

✓ 分类 (Classification)

- ✓ 身高1.80m, 体重100kg的男人肥胖吗?
- ✓ 如何根据肿瘤的体积、患者的年龄来判断良性或恶性?

✓ 回归 (Regression)

- ✓ 如何预测广州大学城的房价?
- ✓ 如何预测未来的股票市场走向?

机器学习的类型

无监督学习

- ✓ **聚类 (Clustering)**
 - ✓ 如何将教室里的学生按兴趣划分为5类?
- ✓ **降维 (Dimension Reduction)**
 - ✓ 如何将原高维空间中的数据点映射到低维度的空间中?

强化学习

- ✓ **用于描述和解决智能体 (agent) 在与环境的交互过程中通过学习策略以达成回报最大化或实现特定目标的问题**

机器学习的概念

1. 模型

- 机器学习首先要考虑使用什么样的模型。
- 模型的类别大致有两种：一是**概率模型(Probabilistic Model)**和**非概率模型(Non-Probabilistic Model)**。

➤ 在监督学习中，概率模型可被表示为 $P(y|x)$ ，非概率模型则为 $y = f(x)$ 。

其中， x 是输入， y 是输出。

➤ 在无监督学习中，概率模型可被表示为 $P(z|x)$ ，非概率模型则为 $z = f(x)$ 。

其中， x 是输入， z 是输出。

机器学习的概念

1. 模型

- 决策树、朴素贝叶斯、隐马尔科夫模型、高斯混合模型属于**概率模型**。
- 感知机、支持向量机、KNN、AdaBoost、K-means以及神经网络均属于**非概率模型**。
 - 对于非概率模型，可按照**判别函数线性与否**分成**线性模型**与**非线性模型**。
 - 感知机、线性支持向量机、KNN、K-means是线性模型。
 - 核支持向量机、AdaBoost、神经网络属于非线性模型。

2. 损失函数

1) 0-1损失函数(0-1 Loss Function)

$$L(Y, f(X)) = \begin{cases} 1, & Y \neq f(X) \\ 0, & Y = f(X) \end{cases}$$

2) 平方损失函数(Quadratic Loss Function)

$$L(Y, f(X)) = (Y - f(X))^2$$

3) 绝对损失函数(Absolute Loss Function)

$$L(Y, f(X)) = |Y - f(X)|$$

4) 对数损失函数(Logarithmic Loss Function)

$$L(Y, P(Y|X)) = -\log P(Y|X)$$

机器学习的概念

2. 损失函数

- 损失函数值越小，模型性能越好。
- 给定一个数据集，我们将训练数据集的平均损失称为**经验风险**。
- 基于经验风险最小化原则，可构建**全局损失函数**求解最优化问题：

$$\min_f \frac{1}{N} \sum_{n=1}^N L(y_n, f(x_n))$$

- 当**样本数量足够大**时，根据大数定理，**经验风险会近似于模型的期望风险**。此时，经验风险最小化能确保有好的学习性能。
- 当**样本数量不足**时，单单利用经验风险最小化可能会导致“**过拟合**”的问题。

机器学习的概念

2. 损失函数

- 在原有基础上加上用于控制模型复杂度的**正则项(Regularizer)**，得到结构最小化准则：

$$\min_f \frac{1}{N} \sum_{n=1}^N L(y_n, f(x_n)) + \lambda J(f)$$

- 其中， **$J(f)$** 代表对模型复杂度的惩罚。模型越复杂， **$J(f)$** 越大，模型越简单， **$J(f)$** 就越小。
- **λ** 是一个正的常数，也叫**正则化系数**，用于平衡经验风险和模型复杂度。
- 一般来说，结构风险小的模型需要经验风险和模型复杂度同时小，因此对训练数据和测试数据都能有较好的拟合。

3. 优化算法

- 算法指的是模型学习中的具体计算方法。
- 一般来说，基于参数模型构建的统计学习问题都为最优化问题，它们都具有显式解析解。
- 现有的优化方法主要有：**梯度下降法、牛顿法、拟牛顿法、ADAM**等等。
- 具体的算法，我们会在各自章节中介绍。本课程中，用梯度下降法作为主要的优化算法。

4. 模型评估

- 当损失函数给定时，我们将基于模型**训练数据的误差(Training Error)**和**测试数据的误差(Testing Error)**作为模型评估的标准。
- 测试误差的具体定义为：
$$E_{test} = \frac{1}{N'} \sum_{n=1}^{N'} L(y_n, \hat{f}(x_n))$$
- 其中 N' 为测试数据数量， $L(y_n, \hat{f}(x_n))$ 是损失函数， y_n 代表真实标签， $\hat{f}(x_n)$ 代表预测标签。
- 一般来说，若我们模型学习的效果好，则训练误差和测试误差接近一致。

机器学习的背景知识

1. 数学基础

高等数学

导数、微分、泰勒公式.....

线性代数

向量、矩阵、行列式、秩、线性方程组、特征值和特征向量.....

概率论与数理统计

随机事件和概率、概率的基本性质和公式、常见分布、期望、协方差.....

2. Python

环境安装: Anaconda, Pycharm.....

模块安装: Numpy, Pandas, Scipy, Matplotlib, Scikit-learn.....

机器学习的开发流程



机器学习的开发流程



数据搜集



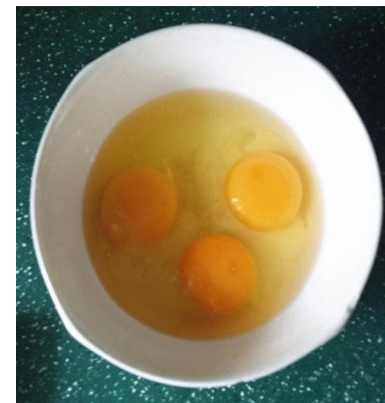
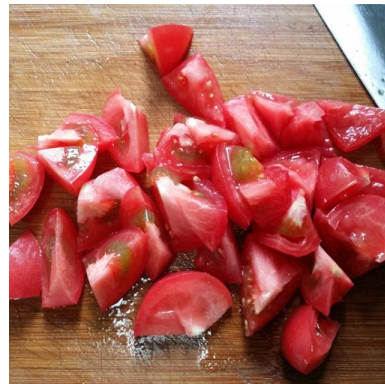
数据清洗



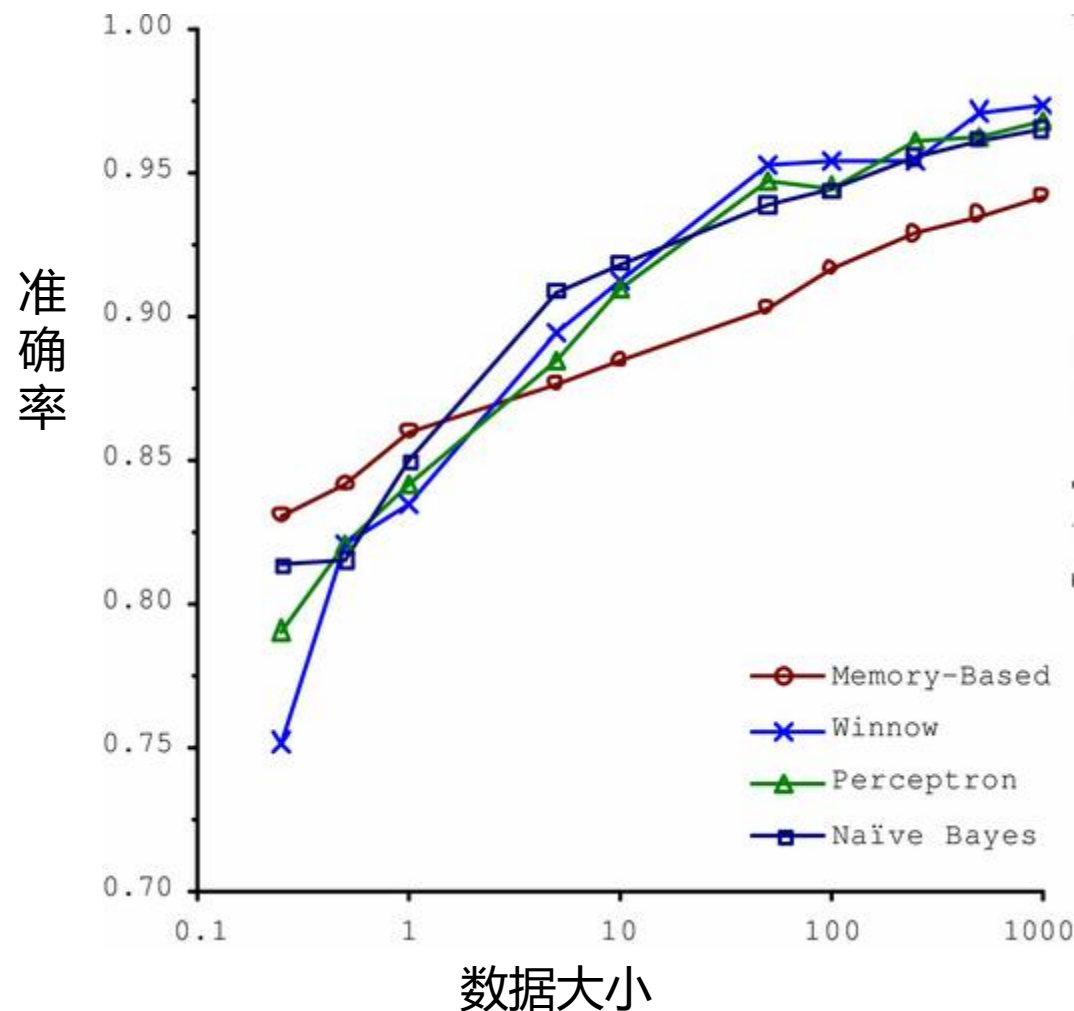
特征工程



数据建模



数据决定一切



通过这张图可以看出，各种不同算法在输入的数据量达到一定级数后，都有相近的高准确度。于是诞生了机器学习界的名言：

成功的机器学习应用不是拥有最好的算法，而是拥有最多的数据！

参考文献



- [1] Andrew Ng. Machine Learning[EB/OL]. StanfordUniversity,2014.<https://www.coursera.org/course/ml>
- [2] 李航. 统计学习方法[M]. 北京: 清华大学出版社,2019.
- [3] 周志华. 机器学习[M]. 北京: 清华大学出版社,2016.
- [4] Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning[M]. New York: Springer,2001.
- [5] CHRISTOPHER M. BISHOP. Pattern Recognition and Machine Learning[M]. New York: Springer,2006.
- [6] Stephen Boyd, Lieven Vandenberghe, Convex Optimization[M]. Cambridge: Cambridge University Press, 2004.
- [7] TOM M MICHELLE. Machine Learning[M]. New York: McGraw-Hill Companies,Inc,1997.



谢谢

魏凤凤 助理教授
fengfeng_scut@163.com
计算机科学与工程学院

