

机器学习导引

第一章 绪论（讲义）

主讲：何新卫

一、课程简介

《机器学习导引》是一门面向研究生的引导性课程，共 32 学时（16 次课），旨在帮助大家构建机器学习的基础知识体系。先修课程：微积分、概率论、矩阵论、Python/C，助教：陈佳恒。

考核方式：平时成绩 40%（课堂表现与课堂小测验），期末项目报告 60%（按模板撰写，提交代码与 readme.md，应用型或算法型二选一）。

二、什么是机器学习

2.1 从人类如何学习说起

人类的学习大致有三种范式，机器学习的三大范式正是对它们的数学化与算法化抽象：

- 自己观察学习：婴儿无人教导，靠观察自动把相似的事物归到一起（分组）——对应无监督学习；
- 妈妈教我学习：妈妈给出输入（图片）与输出（标签），学习两者间的映射——对应监督学习；
- 不断试错学习：学走路靠摔倒（负反馈）与鼓励（正反馈）调整行为——对应强化学习。

一个自然的问题是：能否让机器也模拟这些学习方式，从数据中自我进化，而不再仅仅执行人写死的指令？

2.2 「Machine? Learning?」拆词

「机器学习」里的「机器」不是一台会学习的机器人，而是计算机，以及运行其上的学习算法/模型；「学习」指程序从数据中自己学，而非人把规则一条条写死。

命名由来：Arthur Samuel（1959）研究让计算机自己下跳棋，给出经典定义：

“The field of study that gives computers the ability to learn without being explicitly programmed.”

（即：让计算机无需被显式编程即可学习的研究领域。）

2.3 机器学习的定义

维基百科的定义（逐句解读）：

Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalise to unseen data, and thus perform tasks without explicit instructions.

机器学习是人工智能领域的重要分支，致力于研究能够从数据中学习并泛化到未知数据的统计算法，从而在没有明确指令的情况下自主完成任务。

形式化的定义（Mitchell, 1997）：

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .

给定任务 T 与性能度量 P ，利用经验 E 优化 θ ，使 h_θ 在 T 上的性能（以 P 衡量）随 E 提升。（对应关系：经验 E = 数据，任务 T = 学习要完成的事，性能度量 P = 衡量“学得好不好”。）

定义中的三个关键词：

- **数据 (Data)**：机器的燃料与经验，学习的起点；
- **学习 (Learning)**：从数据中自动总结规律，而非人工写死规则；
- **泛化 (Generalization)**：在未见数据上仍有效，是衡量模型是否“真学会”的黄金标准。

整个学期的内容，本质上就是在剖析这个定义：如何设计算法实现“学习”、如何用数学保证“泛化”、如何为“任务”准备“数据”。

2.4 数据、任务与模型

机器学习的三要素构成一个闭环：

- **数据 (Data)**：多模态存在——图像、视频、文本、点云、音频等；
- **任务 (Task)**：学习的目的，决定需要什么模型、学什么规律；
- **模型 (Model)**：从数据到任务的映射函数，连接两者的桥梁。

根据任务收集数据，选择/设计模型结构，再用优化算法让模型从数据中学习，最终得到能完成该任务的映射函数。

三、机器学习的三大范式

3.1 监督学习（分类与回归）

给定输入 x ，并告知其对应的正确输出 y ，学习映射 $x \rightarrow y$ 。按输出类型分为两类：

- **分类**： y 离散、有限（二分类 $Y = \{+1, -1\}$ ；多分类 $Y = \{0, \dots, K\}$ ）。例：手写数字识别、垃圾邮件过滤；
- **回归**： y 连续（ $y \in \mathbb{R}$ 或 $\mathbb{R}^k$ ）。例：房价预测、温度预测。

记号：训练数据 $D = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$ ； $x^{(i)}$ 为第 i 个样本（通常是 D 维向量）， $y^{(i)}$ 为其标签。目标：给定新输入 $x^{(n+1)}$ ，预测 $y^{(n+1)}$ 。

3.2 无监督学习（密度估计、聚类、降维）

不给「输入—输出」对，只给数据集，让系统自己发现数据内在的模式或结构。

- **密度估计**：给定样本 $x^{(1)}, \dots, x^{(n)} \in \mathbb{R}^D$ （IID 抽样自 $\Pr(X)$ ），估计 $\Pr(x^{(n+1)})$ 。例：异常检测、生成模型；
- **聚类**：把相似样本分到一组。例：客户分群、新闻分主题；
- **降维**：把 $\mathbb{R}^D$ 重新表示为 $\mathbb{R}^d$ （ $d < D$ ），保留关键信息。例：PCA、高维数据可视化。

备注：IID 即独立同分布；密度估计常作为监督学习的“子程序”。

3.3 强化学习（智能体与环境）

同样是学 $x \rightarrow y$ 的映射（状态 $\rightarrow$ 动作），但没有“哪个 y 最好”的直接监督信号，而是靠奖励试错学出策略。

智能体—环境循环：观察状态 $x^{(0)} \rightarrow$ 选择动作 $y^{(0)} \rightarrow$ 获得奖励 $r^{(0)} \rightarrow$ 环境按概率转移到新状态 $x^{(1)} \rightarrow \dots$

目标：学习策略 $\pi: x \rightarrow y$ （状态 $\rightarrow$ 动作），最大化长期累积奖励。

关键区别：动作既影响奖励，又影响后续能观察到什么，必须考虑长期后果。例：下围棋（AlphaGo）、自动驾驶、机器人控制。

四、数据表示：张量

4.1 张量的概念

张量是向量和矩阵的推广，是可超过两维的多维数组，是机器学习的“通用语言”；数据和模型参数都用张量表示。0 维张量 = 标量；1 维 = 向量；2 维 = 矩阵；3 维 = 三维数组（如 RGB 图像）。

4.2 常见数据的张量表示

- RGB 图像： $C \times H \times W$ （通道 $\times$ 高 $\times$ 宽），如 $3 \times 256 \times 256$ ；
- 视频： $T \times C \times H \times W$ （时间 $\times$ 通道 $\times$ 高 $\times$ 宽）；
- 点云： $N \times 3$ （点数量 $\times$ xyz 坐标）；
- 文本： N 个 token。

4.3 文本表示：从 One-Hot 到词嵌入

- 词表 V ：把所有词映射成数字 ID；
- One-Hot 编码：长度为 $|V|$ 的向量，仅对应位置为 1。缺点：无法表达词间相似性（“猫”“狗”点积为 0），数据稀疏；
- 词嵌入：低维、稠密、连续的向量，在训练中学习。语义相近的词距离接近，可捕捉“国王 — 男人 + 女人 $\approx$ 女王”的语义关系。

五、一个完整例子：手写数字识别

5.1 任务与数据

输入 28×28 黑白图片（手写数字 0-9），输出其类别。这是典型的监督学习多分类问题。

5.2 线性模型

- 重塑（Reshape）：把 28×28 矩阵拉平成 784 维向量 $\mathbf{x} \in \mathbb{R}^{784}$ ；
- 线性变换： $z = \mathbf{x}\mathbf{W}^\top + \mathbf{b}$ 。 $\mathbf{W}$ 是 10×784 权重（10 个“数字模板”）， $\mathbf{b}$ 是 10 维偏置， z 是 10 维 Logits（原始分数）；
- 分数越高，代表图片与对应数字模板越相似；

- 后续章节：Softmax 把 Logits 变成概率；损失函数 + 梯度下降更新 $\mathbf{W}, \mathbf{b}$ 。

5.3 PyTorch 实现（核心代码）

```
z = torch.matmul(x, self.weight.t()) + self.bias    # 对应  $z = \mathbf{x}\mathbf{W}^T + \mathbf{b}$ 
```

5.4 三大核心问题

- **Q1** 如何优化参数 ($\mathbf{W}, \mathbf{b}$)? ——训练集、损失函数、优化算法;
- **Q2** 超参数（学习率等）如何设置? ——用验证集调优;
- **Q3** 如何评估模型好坏? ——用测试集最终评估（只用一次）。

六、机器学习发展简史

- **推理期（1950s）**: 认为“智能 = 逻辑推理”，把知识编码成规则。局限：知识瓶颈、系统脆弱；
- **知识期（1970s–80s）**: 专家系统，把人类专家知识“教”给机器。局限：知识难总结、难交付、仍脆弱；
- **学习期（数据驱动）**: 让机器自己从数据中“学”，完成从“教”到“学”的关键飞跃；
- **大数据时代**: 机器学习与大数据共生，数据从“负担”变为“资产”。

七、应用案例

人脸识别、短视频推荐、语音识别、自然语言大模型、自动驾驶、火星车；以及文献筛选、画作鉴别、古文献修复、视频监控、皮肤癌早期识别（Stanford, Nature 2017）、无人配送等。

八、本章小结

8.1 核心概念

- 定义三要素：数据、学习、泛化；
- 三大范式：监督（分类/回归）、无监督（密度估计/聚类/降维）、强化（策略/奖励）；
- 数据表示：张量；文本从 One-Hot 到词嵌入；
- 模型与学习算法：模型是“产物”（训练后的 h_θ ），学习算法是“过程”（用数据调 θ ）；
- 数据集划分：训练集（调参）、验证集（选超参）、测试集（最终评估，只用一次）。

8.2 推荐阅读

Esteva A, Kuprel B, Novoa R A, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. Nature, 2017, 542(7639): 115–118.

九、课堂自测题

9.1 选择题

Q1 维基百科定义里，判断模型是“真学会”还是“死记硬背”的关键词是？

A. 数据 B. 学习 C. 泛化 D. 指令

Q2 下列哪个任务是“回归”？

A. 手写数字识别 B. 垃圾邮件过滤 C. 房价预测 D. 人脸识别

Q3 一段 10 秒、30fps、分辨率 256×256 的彩色视频，是几阶张量？

A. 1 阶 B. 2 阶 C. 3 阶 D. 4 阶

Q4 专家系统（把人类专家的知识“教”给机器）主要出现在哪个时期？

A. 推理期 B. 知识期 C. 学习期 D. 大数据时代

Q5 用一句话概括机器学习，最应该包含哪三个词？

A. 数据、算法、模型 B. 数据、学习、泛化 C. 训练、验证、测试 D. 监督、无监督、强化

9.2 参考答案与解析

Q1 C——泛化 = 在未见数据上仍有效。

Q2 C——回归输出连续值，其余是分类。

Q3 D—— $T \times C \times H \times W =$ 四阶。

Q4 B——知识期 = 人工总结并交付专家知识。

Q5 B——定义三要素：数据、学习、泛化。

附：中英文术语对照

机器学习	Machine Learning (ML)	人工智能	Artificial Intelligence (AI)
深度学习	Deep Learning (DL)	监督学习	Supervised Learning
无监督学习	Unsupervised Learning	强化学习	Reinforcement Learning
分类	Classification	回归	Regression
数据	Data	任务	Task
模型	Model	学习算法	Learning Algorithm
泛化	Generalization	假设函数	Hypothesis Function h_{θ}
参数	Parameter	超参数	Hyperparameter
经验	Experience	性能度量	Performance Measure
标签	Label	聚类	Clustering
降维	Dimensionality Reduction	密度估计	Density Estimation
智能体 / 环境	Agent / Environment	状态 / 动作 / 奖励	State / Action / Reward
策略	Policy	张量	Tensor
标量 / 向量 / 矩阵	Scalar / Vector / Matrix	点云	Point Cloud
特征	Feature	词表	Vocabulary
One-Hot 编码	One-Hot Encoding	词嵌入	Word Embedding
语料库	Corpus	数据集	Dataset
训练集	Training Set	验证集	Validation Set
测试集	Test Set	标注	Annotation / Labeling
数据泄露	Data Leakage	线性层	Linear Layer
权重 / 偏置	Weight / Bias	Logits	Logits (原始分数)
Softmax	Softmax	损失函数	Loss Function
优化算法	Optimization Algorithm	学习率	Learning Rate
准确率	Accuracy	专家系统	Expert System
大数据	Big Data		