01 机器学习和统计学习

从现实能力到机器学习的基本过程

学习目标

- 认识人工智能、机器学习及相关领域，了解主要技术路线的变化与并存
- 用任务、经验、数据、表示、学习结果和评价描述一个机器学习问题
- 区分学习信号、任务类型和生成与判别的建模路径
- 解释有限数据怎样参与模型训练，以及训练、使用和评价如何连接
- 理解过拟合、正则化与泛化的基本关系，说明结论适用的条件

1 认识机器学习

建立共同问题与领域坐标，再形成基本概念地图。

现实中的预测、生成与行动

- **预测**根据已有信息判断类别、数值或未知结果
  - 人脸识别根据图像验证或识别身份
- **生成**产生新的文本、图像等内容
  - 语言助手根据上下文生成回答
- **行动**是对环境施加影响的行为，系统依据环境信息与目标选择行动
  - 机器人依据环境选择移动、抓取等行为
- 预测、生成与行动可以组合在同一系统中
- 系统怎样从数据或交互经验形成可重复使用的处理方式?

人工规则与从经验学习

- **人工规则**是由人明确规定条件与处理步骤的方式
  - 条件与步骤明确时，可以直接编写规则，如检查固定数据格式
- **机器学习**利用数据或交互经验，形成或改进完成任务的处理方式
  - 图像、语言或环境的变化很多，难以预先穷举全部情形，从经验学习提供另一条形成能力的路线
- 真实系统可以组合人工规则、搜索、规划与机器学习
  - 学习也需要人明确任务、选择可用经验和评价依据
- 分析学习系统，可以提出四个共同问题
  - **任务是什么**：系统接收什么，需要输出什么或完成什么行为？
  - **经验从哪里来**：系统使用哪些样本、记录或交互反馈？
  - **系统学到了什么**：形成了怎样的、可供后续使用的处理方式？
  - **怎样判断结果**：依据什么评价任务表现？

人工智能与机器学习

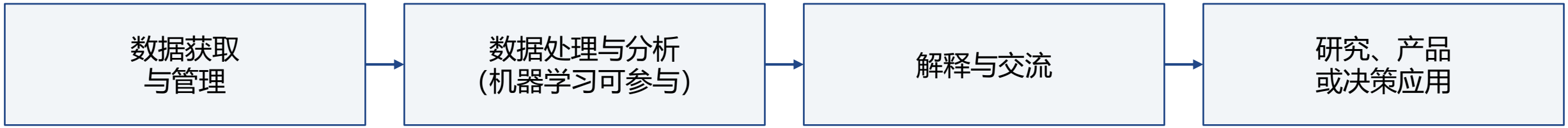
- **人工智能**研究和构造能够执行感知、推理、学习、规划、生成或行动等任务的系统
  - 关注系统需要完成的任务、所需能力及实现路线
- 机器学习利用数据或交互经验，形成或改进系统完成任务的能力
  - 关注如何利用数据或交互经验，形成或改进完成任务的处理方式
- 机器学习是实现人工智能能力的重要路线，也用于科学分析和预测
- 人工智能还使用规则、搜索、规划和知识表示；一个完整系统可以组合多条路线
- 人工智能与机器学习的研究及应用广泛交叉

统计学习的视角与深度学习方法

- **统计学习**依据有限观测，在统计假设下建立模型，用于预测、推断或结构分析，并关注不确定性和适用范围
- **深度学习**以多层神经网络为基础，依据数据形成复杂表示并完成任务
  - **表示**是系统用于描述和处理输入的信息形式
- 统计学习与深度学习采用不同的划分角度
  - 统计学习强调有限数据、统计假设与不确定性
  - 深度学习指以多层可训练结构为基础的方法体系
- 深度模型也可以接受统计分析；统计学习与机器学习的研究范围高度重叠

数据挖掘、数据科学与机器学习

- **数据挖掘**关注从实际数据中发现有用的模式、关系或知识
- **数据科学**综合数据管理、统计分析、机器学习、可视化和领域知识，完成从数据获取到解释与应用的实践
- 机器学习方法可以参与数据挖掘和数据科学实践
  - 分类、预测、聚类等方法支持数据分析
  - 知识发现还需要解释所发现的关系
  - 完整数据实践还包括数据质量、管理、交流和应用



机器学习可以参与“数据处理与分析”，也可支持相关环节；图示是一条简化的实践路径，实际工作可按任务需要调整顺序

相关领域的区别与交叉

- 同一项目可以从不同领域的角度分析。以“用动物图像构建猫狗识别系统”为例：

领域	在本例中关注什么	可能涉及的工作
人工智能	系统怎样完成图像识别任务	组织感知、识别及必要的规则处理
机器学习	怎样从已有图像形成预测方式	利用样本及标签建立分类处理方式
统计学习	有限图像能支持什么判断	分析采样、预测误差与不确定性
深度学习	是否用多层网络学习图像表示	采用网络形成表示并用于分类
数据挖掘	图像集合中有哪些有用模式	发现分组、异常或特征关系并解释其用途
数据科学	数据怎样服务完整项目	收集整理数据、分析结果并交流与应用

- 这些关注点可以交叉：用深度网络训练分类器，同时涉及机器学习方法；评价有限样本上的证据，则涉及统计视角

项目目标与领域选择

- 具体工作由项目目标决定
  - 猫狗分类可以选择深度网络或其他模型
  - 数据挖掘是否开展，取决于项目是否需要发现和解释数据中的模式
  - 各领域提供可以交叉使用的视角，项目据此组织工作

人工规则与早期机器学习

- 人工智能长期并用规则、搜索与从样本学习的方法；以下节点提供代表性坐标
  - **1955 达特茅斯提案**：提出关于人工智能的研究计划；会议在 1956 年举行
  - **1958 感知机**：展示从样本调整分类处理方式的早期路线
  - **1986 反向传播代表论文**：推动多层网络内部表示与学习方法研究
  - **1995 支持向量网络**：为分类提供另一条统计与几何建模路线
- 时间轴表示年代顺序；人工规则与学习方法持续并存

深度视觉与语言预训练

- 大规模数据、计算资源与深度网络共同推动表示学习；代表节点展示视觉与语言方向的进展
  - **2012 AlexNet**：在 ImageNet 大规模、多类别视觉识别中展示深度网络的进展
  - **2017 Transformer**：提出以注意力机制为基础的序列模型结构
  - **2019 BERT 论文**：以语言预训练连接下游任务；预训练是利用较通用数据先形成可复用能力
  - **2020 GPT-3**：展示大规模语言模型与少样本任务使用的进展
- ImageNet 是多类别视觉任务，本讲的猫狗识别是二分类任务
- 这些节点分别展示视觉与语言方向的代表进展

公众应用、多模态与机器人

- **2022 ChatGPT**使对话式语言模型成为公众可直接使用的应用之一
- **2022 Flamingo**是结合视觉与语言输入的代表模型；多模态指系统处理图像、文字等不同信息形式
- **2023 RT-2**探索把视觉和语言知识连接到机器人动作
- **2025 Gemini Robotics 1.5**是视觉—语言—动作方向的代表系统
- 这些节点分别涉及公众使用、跨信息形式和环境交互
  - 各系统的能力及适用条件以对应发布或实验材料为准

数学推理、世界模型与编码智能体

- **数学推理**：AlphaProof（2024）展示竞赛数学进展；2026 年 OpenAI 公布 Navier–Stokes 解答及 Lean 形式化证明
  - 公布解答、形式化验证与 Clay 正式评奖是不同确认环节；正式评奖仍须依据 Clay 的评价程序
- **世界模型**：李飞飞及团队在 2026 年提出一种功能分类：渲染器输出可见画面，模拟器描述可计算的结构与动态，规划器输出行动
  - 这是作者提出的工作分类；世界模型概念更早已有
  - 按这一分类，Genie 3 属于交互式渲染器；物理可靠性需要相应证据支持
- **编码智能体**：Claude Code 与 Codex 等系统可参与代码生成和修改；vibe coding 可作为快速原型方式，结果仍需测试与审查

猫狗分类的任务、经验与评价

- **任务** T规定输入与需要的输出：输入一幅动物图像，输出“猫”或“狗”，这是二分类
- **经验** E是形成或改进能力的可用观测：带有类别标签的猫狗图像
  - 标签是对应类别答案，训练时提供学习信号，评价时用于核对预测；预测的输入是一幅待识别图像
- **评价**在约定的数据和条件下，将模型预测的类别与真实标签比较，判断猫狗分类任务的完成情况
  - **性能指标** P把任务表现转化为可比较的量，如准确率
    - 准确率是预测正确的图像数占被评价图像总数的比例
  - 评价还要关注把猫判为狗、把狗判为猫等错误及其代价
- $T/E/P$ 分别描述任务、经验和性能指标，是分析学习问题的框架
  - P 来自 performance，表示性能指标；$P(\cdot)$ 表示概率
  - 当目标变为定位动物或识别品种时，输出及任务也随之改变

数据集的有限性与代表性

- **数据集**是为学习或分析组织起来的有限观测集合
  - 观测可以带标签，也可以没有任务标签
  - 一张照片记录一个动物在特定姿态、光照、背景和设备条件下的信息
- **代表性**须相对于任务、采样方式与预计使用范围判断
  - 数据量描述观测数量
  - 覆盖范围描述观测涉及哪些品种、姿态、设备或背景；猫狗数据可能遗漏其中一些情况
  - 训练时见到的条件，是否适当覆盖模型需要处理的条件？
- 有限数据提供的证据有适用范围，预测结论需要结合采样条件与预计使用环境判断

人工特征与学习表示

- **原始输入**是现实对象被记录后进入系统的数据形式，如动物图像的像素
- **特征**是描述输入、供模型处理的属性或数值，如图像像素值、颜色统计或纹理描述
- 表示是输入供后续处理使用的信息形式，可以由一组特征构成，也可以是网络内部的数值编码
  - 若原始图像记为 x，表示可记为 $z = \phi(x)$；ϕ 是转换过程，z 是所得表示
- 表示可以由人设计，也可以通过学习得到
  - 人工特征由人按任务知识设计，如轮廓、局部纹理、耳部形状或颜色统计
  - 学习表示由训练过程得到，如网络中间层的数值编码
- 人工特征与学习表示都受数据、任务及设计选择影响
  - 传统方法通常更依赖人工特征，深度学习更强调表示与任务处理共同学习
- 表示决定后续直接使用哪些信息
  - 表示可能保留背景等无关因素，也可能丢失有用信息

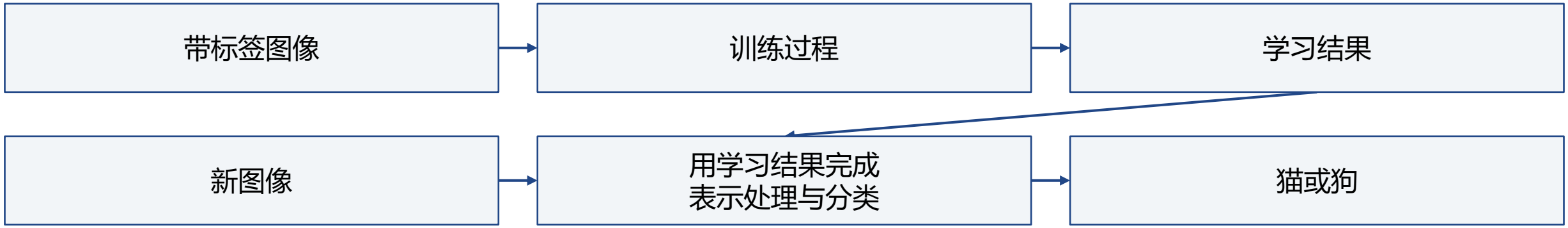
学习、模型与学习结果

- **学习**利用数据或交互经验，形成或改进可重复使用的处理方式
- **模型**用数学或计算结构表示输入输出关系、数据规律或决策规则，并据此处理对象
- **学习结果**是学习得到的可计算载体，如具体模型、表示或选择行动的规则
  - 可以表现为带系数的函数、树规则、概率关系、保存的样本或网络权重
  - 训练算法负责形成具体结果；训练后的具体模型可以是学习结果，学习表示也可以是学习结果

学习结果的表达、实现与使用

- **表达**：说明学习得到的具体模型、表示或规则
  - 可以采用带系数的函数、决策树、概率模型、保存的样本或训练后的网络等形式
- **实现**：程序结合所得参数、结构或样本，执行相应处理
  - 训练算法负责形成学习结果，预测算法规定如何使用结果处理新输入
  - 可执行性与能否概括成简洁可读规则是不同性质
- **使用**：将学习结果用于一幅新图像，得到“猫”或“狗”的预测
  - 学习结果支持任务能力；一次预测是使用结果产生的行为，表现需要评价

从有限经验到新图像的预测



连接：训练结果供预测使用，沿用同一表示处理与分类模型

- **训练与预测**：训练形成学习结果，预测用它处理新图像
  - 人工特征路线先确定表示处理，再学习分类模型；学习表示路线可共同学习表示与分类模型
- **推广条件**：学习尝试用经验中的任务相关关系处理新图像
  - 这些关系在预计使用条件下仍适用，有限经验才可能支持新预测
  - 新图像上的表现须用独立且与使用条件相关的数据评价
- **学习选择**：有限数据通常容许多种解释，学习还依赖候选形式、选择标准与求解方式

学习范式与信号来源

- **学习范式**按经验与训练信号区分学习方式

名称	经验或训练信号	例子
监督学习	输入与对应任务目标	图像与猫狗标签
无监督学习	无任务标签的观测及结构假设	按相似性组织图像
自监督学习	从数据自身关系构造目标	预测被遮住的内容
半监督学习	联合使用有标签与无标签数据	少量标注配合大量图像
强化学习	行动、奖励与持续交互经验	依据抓取奖励调整策略

- 范式与任务可组合；自监督构造监督式训练目标，与其他学习方式可以交叉

任务类型与表示学习目标

- 学习范式说明信号来源，任务类型说明系统要得到什么结果

任务或目标	需要得到什么	例子
分类	预定离散类别	判断猫或狗
回归	连续数值	预测价格
聚类	根据相似性或假设组织分组	无标签图像分组
生成	新数据样本或内容	生成文字或图像
决策	当前应采取的行动	选择机器人动作
表示学习	更有用的数据描述或内部编码	学习供分类使用的图像表示

- 表示学习常为最终任务提供中间表示；同一任务可以采用不同建模路径

同一分类任务的两种建模思路

- **共同任务**：用带标签的猫狗图像学习，判断新图像是猫还是狗
- **思路一：先描述各类数据，再判断类别（生成路径）**
  - 学习什么：猫、狗各自的特征分布，以及类别出现概率
  - 怎样预测：结合各类分布与类别出现概率，判断新图像属于哪一类
- **思路二：直接学习区分类别（判别路径）**
  - 学习什么：给定图像时的类别概率，或直接区分猫狗的决策规则
  - 怎样预测：把新图像输入学得的模型，得到类别判断
- 两种思路的输入与分类目标相同，学习对象和形成预测的方式不同

生成与判别的建模路径

- 在监督分类中，**生成方法**学习类别出现概率及各类数据的分布，再据此判断类别
  - 代表方法：朴素贝叶斯
- **判别方法**直接学习给定输入时的类别概率或决策函数
  - 代表方法：Logistic 回归、感知机、SVM
- **条件性用途**：生成路线可支持分类、采样或缺失信息处理；判别路线直接针对预测
  - 具体用途及表现取决于模型、假设与数据
  - 现代生成模型学习复杂数据及条件分布、生成新内容，范围比上述分类比较更广
- **分类维度**：“监督 + 分类 + 判别”分别说明信号、任务与路径；深度学习说明模型方法体系

从类别预测到行动选择

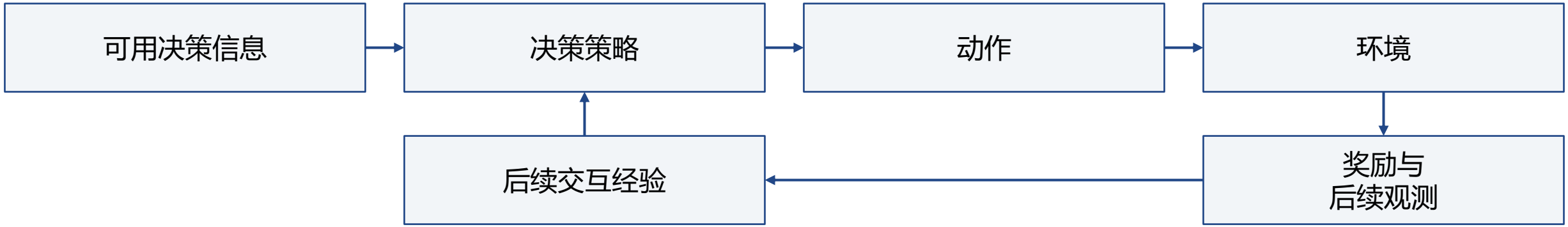
- 从猫狗识别转向机器人抓取，任务输出由类别变为行动
  - **智能体**获取环境信息并选择行动；**动作**是作用于环境的可选行为

比较项	猫狗分类	机器人抓取
需要的输出	图像对应的类别	当前采取的动作
经验形式	已有图像与类别标签	行动后的反馈与后续信息
输出的作用	给出类别判断	改变环境，影响后续经验

- **状态**描述决策相关的环境情况；**观测**是智能体获得的环境信息，可能只覆盖部分状态
- 在交互任务中，行动影响后续数据，学习需要考虑行动的连续效果

强化学习与具身智能

- **决策策略**依据可用决策信息选择动作，可用规则或条件分布表达
- **奖励**是界定决策目标的数值反馈；当前奖励与长期累积效果可能不同
- **强化学习**用交互经验改进策略，关注行动的累积效果



连接：交互经验可改进策略；动作会影响后续数据

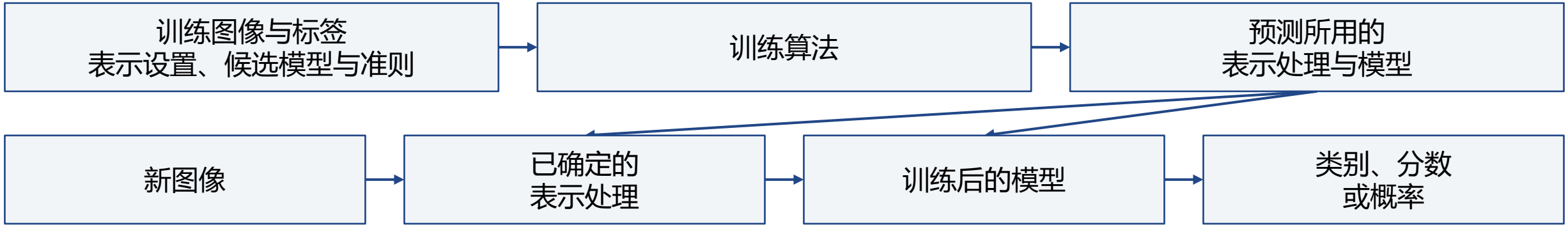
- **具身智能**涉及感知、身体与环境的持续耦合
  - 机器人抓取可结合示范学习、强化学习、规划与控制等方法
  - 示范学习从专家的行为示例学习行动方式

2 机器学习的基本过程

概念认识已成立；说明有限数据怎样形成模型以及如何评价。

监督学习中的训练与预测

- 猫狗分类可采用静态监督过程；无监督与持续交互学习可有不同组织
- **训练**：用图像与标签，按表示设置与候选范围形成学习结果
  - **候选模型**规定可选形式，**学习准则**规定比较与选择标准，**训练算法**求得具体结果



连接：训练结果中的表示处理与模型分别供预测路径沿用

预测输入与表示处理的数据边界

- **预测**：沿用已确定的表示处理与模型；输出为类别、分数或概率，由任务和模型决定
  - 图像是预测输入，模型给出类别判断；真实标签用于训练和评价核对
- **表示处理的数据边界**：需从数据确定的处理，用训练数据确定后供预测沿用
  - 表示设置可包含预先设计的人工特征处理，或与分类模型共同学习的表示

数据来源与目标使用条件

- **数据来源**：训练图像来自哪些动物、设备、场景与收集方式
- **目标使用条件**：预计识别哪些动物，在什么条件下采集与表示图像
  - 例如训练多为明亮室内照片，使用多为昏暗室外照片，需检查差异对预测的影响
- **使用依据**：训练与目标条件可以不同，二者的关联须支持预期使用
  - 明确哪些任务相关关系预期仍适用，并在目标条件下评价
- **概率描述**：概率关系描述特征与类别的共同出现，可用于分析训练与使用的关联
  - 概率关系描述未知规律；有限数据提供观察与估计依据，对未知概率的估计仍有不确定性

用特征与类别记录图像

- 判断训练经验能否支持目标使用，需要描述特征与类别的关系
- **观察单位**是一幅图像，记录其特征与类别
  - 本例将视觉信息简化为“耳部是否较尖”
  - X 记录特征：1 为耳部较尖，0 为其他；Y 记录类别：1 为猫，0 为狗
- **随机变量**将随机抽取的结果对应为数值；本例 X, Y 分别记录所抽图像的特征与类别
  - 从指定来源随机抽取图像，抽到的图像决定 X, Y 的取值
  - 一幅已观察图像给出一对确定取值；汇总记录得到特征与类别的计数表
- 该二元特征仅保留部分视觉信息，真实识别需更充分的信息

从交叉计数到列联表

- 先按“耳部较尖 / 其他”和“猫 / 狗”交叉分组，再数每种组合包含多少幅图像
- **列联表**按两个分类变量的取值交叉列出计数，行、列合计分别汇总各变量的情况
- 以下 100 幅图像的计数为教学构造，沿用 X, Y 的取值定义

教学特征	猫：$Y = 1$	狗：$Y = 0$	合计
$X = 1$：耳部较尖	30	10	40
$X = 0$：其他	20	40	60
合计	50	50	100

- 例如耳部较尖的猫有 30 幅；耳部较尖共 40 幅，猫共 50 幅
- 列联表把有限样本中的共同出现与各自合计放在一起，为计算经验比例提供依据

联合分布与边缘分布

- 将计数转为比例，可以描述特征与类别共同出现、以及各自出现的情况
- 本例中，**概率分布**列出离散随机变量各可能取值的概率；P 表示目标来源的概率，$\hat{P}$ 表示样本中的经验比例
- **联合分布**描述 X, Y 共同取值的概率；本例计算其中一项
  - 同时满足 $X = 1, Y = 1$ 的 30 幅，占全部 100 幅：$\hat{P}(X = 1, Y = 1) = 30/100 = 0.30$
- **边缘分布**描述单个变量的取值概率，由联合概率对另一个变量的所有取值求和得到
  - 耳部较尖的比例：$\hat{P}(X = 1) = (30 + 10)/100 = 0.40$
  - 猫的比例：$\hat{P}(Y = 1) = (30 + 20)/100 = 0.50$
- 经验比例描述有限样本；估计目标概率的适用性取决于采样与目标条件

已知条件如何改变概率

- **条件概率**描述已知条件成立时，目标情况出现的概率；此处离散条件须有正概率
  - $P(Y=1 \mid X=1)$ 表示已知耳部较尖时为猫的概率；| 读作“在……条件下”
  - 本例用经验比例计算：先筛出满足条件的图像，再用目标计数除以条件计数，条件计数须大于零
- **给定特征，判断类别**：耳部较尖的 40 幅图像中，30 幅是猫
  - $\hat{P}(Y=1 \mid X=1)=30/40=0.75$
- **给定类别，描述特征**：50 幅猫图像中，30 幅耳部较尖
  - $\hat{P}(X=1 \mid Y=1)=30/50=0.60$
- 条件方向不同，概率一般不同；本例共同计数均为 30，条件计数为 40、50，因此比例不同
- **条件分布**由固定条件下目标变量所有取值的概率构成，如给定耳部较尖时猫、狗的概率

从联合概率到类别判断

- 将输入与各类共同出现的概率转换为给定输入后的类别概率，可支持本例的类别判断
- 同一共同事件可按两个方向分解；x, y 是固定取值，条件事件须有正概率
  - $P(X = x, Y = y) = P(Y = y \mid X = x)P(X = x)$
  - $P(X = x, Y = y) = P(X = x \mid Y = y)P(Y = y)$
  - 本例经验比例满足 $0.30 = 0.75 \times 0.40 = 0.60 \times 0.50$
- **汇总各类贡献**：耳部较尖包括猫与狗两类
  - $\hat{P}(X = 1) = \hat{P}(X = 1, Y = 1) + \hat{P}(X = 1, Y = 0) = 0.30 + 0.10 = 0.40$
- **得到类别比例**：各类共同出现的比例除以耳部较尖的总比例
  - 猫：$0.30/0.40 = 0.75$；狗：$0.10/0.40 = 0.25$
  - 本例设两类误判代价相同，按较大类别比例选择：耳部较尖图像判为猫

生成与判别的学习与预测

- 同一分类任务可采用不同建模路径，其学习对象与预测过程可以不同
- $P(X,Y)$、$P(Y \mid X)$ 等表示分布；具体计算代入相应取值，沿用本例离散变量
- **判别路径**：直接学习估计类别条件分布 $P(Y \mid X)$ 的模型，或直接学习决策函数
  - 给定输入，模型直接给出类别概率估计或类别判断
- **生成分类路径**：学习估计各类特征分布 $P(X \mid Y)$ 与类别概率 $P(Y)$ 的模型
  - 将两项相乘，得到输入与各类共同出现的概率估计；比较各类贡献，或归一化得到类别概率估计
  - 归一化以各类贡献之和为分母，对应 $P(X)$；给定输入且该分母为正时，各类分母相同
- 概率分解是恒等关系；经验比例能否支持新图像，仍取决于采样及相关关系在使用条件下是否适用

从模型族到具体模型

- 概率关系为类别判断提供依据；模型参数由训练数据、候选范围、学习准则与算法共同确定
- **模型族或假设空间**是训练允许比较的一组候选模型，如一组线性函数、树或概率模型
- **参数**是模型内部由当前训练过程确定的量，如概率值或模型系数
  - 参数化候选写作 f_θ，θ 区分同一模型族中的具体候选
  - 训练得到 $\hat{\theta}$，固定后形成具体模型 $f_{\hat{\theta}}$
- **超参数**在当前参数训练之外设定，如树深限制或正则化强度，可以依据验证选择
- 本例采用二元特征概率分类模型，真实视觉识别需更充分的信息
  - 不同候选可能给出不同预测；比较预测表现需要代价尺度，确定参数还需选择准则

单次损失、经验风险与期望风险

- 衡量候选模型的预测表现，需要区分单次预测代价与平均代价
- **损失** $\ell(y, f(x))$ 衡量一次预测的代价
  - y 是目标答案，$f(x)$ 是预测；两类误判可设相同或不同代价
- **经验风险**是有限训练集上的平均损失
  - D 含 n 个样本，i 为样本编号；x_i 是输入或表示，y_i 是标签
  - $\hat{R}_D(f) = \frac{1}{n}\sum_{i=1}^{n} \ell(y_i, f(x_i))$
  - 将每次损失相加，再除以样本数，得到训练平均
- **期望风险**是目标概率规律下的平均损失：$R(f) = \mathbb{E}[\ell(Y, f(X))]$
  - $\mathbb{E}$ 表示按各情况发生概率加权平均；目标规律通常未知
- 经验平均能否代表目标平均，取决于数据条件及模型选择；训练目标、评价指标与实际代价也可能不同

类条件参数的极大似然估计

- 概率模型可按其赋予已观测数据的概率选择参数；分类器形成后，再用预测损失衡量类别判断的代价
- **Bernoulli（伯努利）模型**描述只有 1、0 两种取值的变量：取 1 的概率为 θ，取 0 的概率为 $1-\theta$，$0 \le \theta \le 1$
  - 猫类特征模型中，$\theta = P(X = 1 \mid Y = 1)$，是一个待估计的类条件参数
  - 从训练集 D 取出猫图像，特征观测记为 D_{cat}：50 幅中，30 幅特征为 1，20 幅为 0
- **似然**把模型赋予固定数据的概率或密度看作参数函数
  - 假定猫类特征观测相互独立、遵循同一个 Bernoulli 模型
  - 特征为 1 的每次观测贡献 θ，为 0 的每次观测贡献 $1-\theta$；按已观察序列的概率相乘
  - $L(\theta; D_{\text{cat}}) = \theta^{30}(1-\theta)^{20}$
- **极大似然估计**选择似然最大的参数；本例为 $\hat{\theta} = 30/50 = 0.60$
  - 这确定了猫类特征参数；完整分类器还需类别比例和狗类特征参数

概率分类器的参数与条件预测

- 沿用 100 幅构造数据；假定图像记录独立、来自同一联合模型，类别与类内特征均采用二元概率模型
- **完整模型的三参数估计**：固定以下极大似然估计与决策规则，形成分类器
  - 全部标签：$\hat{P}(Y=1)=50/100=0.50$，$\hat{P}(Y=0)=0.50$
  - 猫类特征：$\hat{P}(X=1 \mid Y=1)=30/50=0.60$
  - 狗类特征：$\hat{P}(X=1 \mid Y=0)=10/50=0.20$
- **条件预测**：类贡献 = 类条件概率×类别概率；猫概率 = 猫贡献÷两类贡献之和，均为模型估计
  - $x=1$：猫、狗贡献为 0.30、0.10；总和 0.40，猫概率 $0.30/0.40=0.75$
  - $x=0$：猫、狗贡献为 0.20、0.40；总和 0.60，猫概率 $0.20/0.60=1/3$

类别判断与训练经验风险

- **类别判断**：两类误判同代价时选概率较大类别；$x = 1$ 判猫，$x = 0$ 判狗
- **零一损失**：预测正确取 0，错误取 1
  - 狗被判猫 10 幅，猫被判狗 20 幅；训练经验风险 $(10 + 20)/100 = 0.30$
  - 独立测试还需未参与训练或选择的数据

模型、学习准则与求解算法

- **模型**给出允许的关系形式；**学习准则**给出选择标准；**算法**执行求解
  - 本例采用 Bernoulli 类条件模型，以极大似然为准则，通过计数计算比例获得参数
- **负对数似然**是似然取对数后再取负值；对数保持正数的大小顺序并将乘法转为加法，因此最大化正似然等价于最小化负对数似然
  - 猫类特征模型：$-\log L(\theta; D_{\text{cat}}) = -30\log\theta - 20\log(1-\theta)$，其中 $0 < \theta < 1$，log 表示对数
  - 独立观测下，各次观测的负对数概率相加；除以 50 得到平均负对数损失，最优参数不变
  - 负对数损失评价模型赋予观测特征的概率，零一损失评价类别判断；两种目标一般不等价

建模选择、归纳偏置与正则化

- 二元特征表示决定利用哪些信息，独立假设决定怎样组织观测的概率；这些都是数据之外的建模选择
- **归纳偏置**是学习系统在有限经验容许多个解释时采用的偏好或假设，来自表示、模型形式、准则或求解方式
  - **正则化**通过目标、约束或训练过程表达这类偏好，常用于控制过拟合风险；其效果取决于数据、模型及使用条件，不保证恢复真实规律或获得泛化
- 数据、表示、候选、准则与算法共同形成结果；使用依据还需独立数据与适当评价协议

训练集、验证集与测试集的职责

- 模型对未参与训练的相关对象是否有效，需要独立评价；数据按拟合、选择和最终测试分工
- **训练集**用于拟合模型参数及需拟合的表示处理
  - 训练表现受模型已见信息影响
- **验证集**用于选择模型和训练设置
  - 反复依据验证结果选择，方案可能逐渐适应该验证数据
- **测试集**在方案冻结后，用于估计相关新数据上的表现
  - 独立最终测试要求测试信息不参与训练或方案选择
- **数据泄漏**指本应独立的评价信息进入训练或方案选择，破坏信息边界
  - 同一动物的近重复照片跨集合、用全数据拟合预处理或提前使用测试标签，都可能引入这类泄漏
- 数据划分需考虑采样单位、样本相关性和目标条件，划分比例没有固定规则
  - 测试数据为其覆盖的条件提供证据

评价指标与错误分析

- **评价指标**将任务表现转为可比较的量
  - 猫狗分类的**准确率**是正确分类数占评价样本总数的比例
- 训练损失用于拟合，评价指标用于报告表现，实际使用代价反映错误后果；三者可能不同
- **错误分析**检查错误类型、不同数据子群的表现及失败条件，定位总体分数背后的差异
  - 按错误类型：分别检查猫被判为狗、狗被判为猫的情况
  - 按对象与条件：比较不同品种、姿态、背景、光照、设备或数据来源下的表现
- 总体指标概括表现，错误分析定位差异；解释原因还需检查数据与评价流程

过拟合与分布变化的诊断区别

- 背景变化时错误可能增多，解释差异需区分模型拟合、数据变化与评价流程等原因
- **过拟合**是学习结果过度适应有限训练经验，未能在相关未知数据上保持相应表现
  - 判断需结合训练与相关独立数据上的表现，训练误差低本身不证明过拟合
- **分布变化**是使用数据的条件范围或出现比例，相对训练或评价阶段发生系统变化
  - 如背景、设备或光照的变化
- **构造情境**：模型在训练照片上准确率很高，换背景后表现下降，并过度依赖背景纹理
  - 可能涉及偶然背景关联、数据覆盖不足、分布变化或评价流程问题
  - 仅凭换背景后表现下降，不能唯一诊断为过拟合
- **诊断线索**：比较同条件独立数据与背景变化数据上的表现，检查特征依赖和采样方式
  - 这些比较提供线索，确定原因仍需更多证据
- **改进需对应原因**：数据与流程问题检查覆盖与划分；模型表达或偶然细节拟合问题考察表达范围与学习偏好

模型容量、正则化与学习结果

- 针对模型表达或偶然细节拟合的问题，可从表达范围与学习偏好两方面考察
- **模型容量**表示一类模型能够表达或区分的数据关系的丰富程度
  - 容量过小可能无法表示任务关系，容量很大也可能拟合偶然细节
  - 具体结果取决于数据、噪声、表示和训练方式；参数数量只是部分模型的线索，不能普遍完整衡量容量
- **正则化**表达对候选结果或训练过程的偏好，如限制关系复杂度
  - 可能降低过拟合风险，也可能限制任务所需关系
- **选择与评价**：依据验证证据及任务约束选择模型与正则化强度，再对选定方案独立测试
  - 容量、样本量、归纳偏置与评价共同影响结果；模型更简单不必然更好，加入正则化不保证泛化
- 数据覆盖、分布变化和评价流程问题需分别处理，容量与正则化只回应其中相关的模型问题

泛化评价的条件与边界

- **泛化能力**是模型在明确任务和数据条件下，对未参与训练的相关对象保持有效表现的能力
- **评价证据**：独立测试提供目标条件下表现的有限估计
  - 解释测试结果需考虑样本覆盖、采样相关性和模型选择过程
  - 任务、设备、场景或对象范围变化时，需重新检查覆盖和表现；一次受控测试不能支持所有未来环境
- **形成模型与支持使用**
  - 形成模型：有限经验与表示、候选形式、学习准则和算法共同确定学习结果
  - 支持使用：训练结果还需独立评价，证据范围受任务与数据条件限制
- 相关新对象上的泛化、跨任务迁移和不受条件限制的通用能力，对应不同范围的要求

3 全讲总结

回到开场系统，综合四个共同问题与学习、评价的条件。

识别、生成与行动中的学习问题

- 猫狗分类中建立的任务、经验、学习结果与评价问题，也可用于分析识别、生成与行动系统
- **识别：人脸验证**需要判断两幅图像是否属于同一人
  - 经验来自相关人脸数据
  - 评价需考虑采集设备、人群和使用条件
- **生成：语言模型**依据文本及其他训练信号，形成表示与生成能力
  - 回答质量需结合事实、任务和使用语境评价
- **行动：机器人**可依据示范、交互反馈或其他经验，形成行为处理方式
  - 评价需明确具体任务与环境条件
- 三类系统共用任务、经验、表示、学习结果和评价等分析问题，具体数据结构、目标与训练方式各有不同

本讲总结

- **任务是什么**
  - 明确输入、输出及需要完成的预测、生成、结构发现或行动
- **经验从哪里来**
  - 说明有限数据或交互记录的来源、目标范围，以及数据怎样表示
- **系统学到了什么**
  - 说明学习结果由什么承载、怎样用于新对象，以及候选模型、学习准则、算法和偏好怎样参与形成结果
- **怎样判断结果**
  - 区分训练、选择与最终评价，检查指标、错误类型及任务和数据条件
- 学习需要选择假设，有限经验不能唯一决定模型；使用需要独立且相关的有限证据，训练表现不能独立证明泛化