第1章
认识人工智能

学习人工智能首先要对它在总体上有一个正确的认识，这便是本章的任务。我们将从名词溯源、发展简史、思想根基和理论脉络几个方面出发认识人工智能的本质和基本属性，提出并阐释人工智能的能力属性、工具属性和实用属性，分析和描述人工智能基本能力的内在逻辑关系，用圈层结构概括人工智能的知识体系。然后，本章将介绍人工智能的数学基础，阐述人工智能的数学原理以及机器学习在其中的核心作用。

1.1 何谓人工智能？

“智者灵也，能者动也。人工智能则乃人类所创造之灵动者也。其魅自神奇，其功在造化。其诱人也深焉，其助人也劲焉！”这是作者在多年前写下的一段话，尽管人工智能领域经历了深度学习和大模型所带来的重大技术变革，但这段话并没有过时，因为它道出了人工智能与人类的正确关系，那就是：人工智能是人类神奇的助手和工具，它能为人类创造强大的生产力。

学术上，人工智能这一名词诞生于1956年，其英文是Artificial Intelligence，简称AI。当时，约翰·麦卡锡(John McCarthy)、赫伯特·西蒙(Herbert Simon)、艾伦·纽厄尔(Allen Newell)、马文·明斯基(Marvin Minsky)、克劳德·香农(Claude Shannon)等聚集在美国的达特茅斯小镇，探讨如何让机器模拟人类的思维能力。他们探讨了许多相关问题，而会议达成的最重要共识就是将他们所谈论的内容用Artificial Intelligence命名。

历史证明，这的确是一项划时代的成果，因为这个名词概括出了机器模拟人类的思维能力这一问题的实质，那便是Artificial Intelligence，也就是人工智能！这一名词的提出，也标志着人工智能研究正式启航。

那么，AI究竟是什么意思呢？要回答这一问题，最好的方法是解读Artificial Intelligence这两个英文单词的原意。Artificial很简单，就是人造的或人工的意思，关键是如何理解Intelligence。根据词典的定义，Intelligence是指学习、判断、理解等(人类)大脑的能力。这里的重点是能力(Ability)，也就是说，Artificial Intelligence就是人造的大脑能力。需要特别强调的是，能力是客观存在的，是可观测和可度量的，是与精神或意识完全不同的概念。也就是说，人工智能不涉及精神和意识。这一点十分重要，是人工智能沿着正确方向发展的基本保证。

汉语的“人工智能”便是 Artificial Intelligence 的直译，这个翻译本身虽然没有错误，但其含义却变得模糊了。因为“智能”的意义不像“Intelligence”那样单纯，“智能”是复合词，智是智，能是能，合在一起有智慧和能力双重意思。因此，汉语的“人工智能”的意义并不像 Artificial Intelligence 那样易于从字面上进行解读。还有一个重要问题就是：汉语的“智能”不仅有名词的意思，还有形容词的意思。这便难免会使人将关注的重点放在人工智能的程度强不强、高不高上，从而对人工智能产生错误的解读。例如，在大模型技术出现后，一些人认为人脸识别等早期技术不再是人工智能。

关于人工智能的严格定义，学术界一直在探讨和争论。但时至今日，并没有达成一个统一的意见。各派学者从不同的角度给出了形形色色的定义，表面上各不相同，本质上却大同小异。例如，理论学派将人工智能定义为如何使计算机系统能够履行那些只有依靠人类智慧才能完成的任务的理论研究，技术学派将人工智能定义为使计算机去做过去只有人才能做的智能工作的技术研究，知识学派将人工智能定义为怎样表示知识以及怎样获得知识并使用知识的科学。而跨越各个学派的一种简单观点将实现机器问题求解作为定义人工智能的核心要义。事实上，这些争议并未对人工智能的发展产生实质性的影响，这说明这些定义以及它们之间的差异并不重要。本书认为，与其纠结于人工智能的定义，不如从 AI 这一名词的本义出发，紧紧把握它的能力属性、工具属性和实用属性，这更有利于认识人工智能的实质。

1.2 人工智能发展简史

人工智能的发展历史是人类在创造人工智能的同时不断认识其特征和本质的历史。从 1956 年人工智能元年以来，人工智能的发展经历了两次大的曲折，史称人工智能的两次寒冬。此后在深度学习和大模型技术的强力推动下，人工智能的整体水平和能力迅猛提升，达到了重构人类经济社会格局的高度，被称为新质生产力的核心要素。简要回顾这段历史，了解其中主流技术与理论思想的变迁，揭示发展人工智能的正确路线和正确方法，是认识和学习人工智能的第一步。

人工智能研究启航以后，早期以机器逻辑推理和数学定理证明为主要研究内容。在取得若干成功之后，全球范围的研究热潮迅速形成。乘着这一热潮，一些偏离方向不切实际的研究项目也纷纷上马。结果，到了 20 世纪 70 年代，大批项目的研究目标落空，令支持这些研究的各国政府陷入尴尬，不得不宣布终止对这些项目的资助，人工智能的发展进入了第一个寒冬。

20 世纪 80 年代，以知识工程为核心的专家系统技术异军突起，在很多重要领域得到应用。同期，日本又提出了雄心勃勃的第五代计算机研究计划，其核心便是制造具有人类感知能力和思维能力的计算机，也就是智能计算机。在这两大动力的共同作用下，人工智能的研究再次进入高潮。各国纷纷投入大量的研究力量和资金对相关项目进行开发和资助。可惜好景不长，到了 20 世纪 90 年代，由于开发的技术达不到市场预期，大量的投入得不到回报，故人们再次对人工智能技术丧失信心和兴趣，人工智能的发展进入了第二个寒冬。

虽然经过了两次寒冬的打击，但是研究人员并没有放弃。2000 年以后，以杰弗里·辛顿（Geoffrey Hinton）为代表的一批科学家历时 10 多年研究的深度学习理论和方法，通过与应用紧密结合逐渐崭露头角，在一系列重要技术竞赛中取得了令人刮目相看的性能。2012 年，在 ImageNet 图像识别大赛中，深度卷积神经网络 AlexNet 的性能达到甚至超越了人眼的识别精

度，以显著优势拔得头筹。以深度神经网络为内核的 AlphaGo 在 2016 年与 2017 年分别战胜了人类围棋顶级选手李世石和柯洁，震惊了全世界。

自此之后，以深度学习为特征的新一代人工智能在图像识别、语音识别、自然语言处理、机器翻译等重要应用中不断刷新性能，人工智能技术整体上发展势头之猛、涉及领域之广、颠覆作用之大，令人猝不及防。为抢占发展先机，占领科技制高点，各国政府纷纷制定发展人工智能技术的国家战略，教育界、科技界、产业界积极响应，形成了人工智能发展的新时代洪流。

2022 年 11 月，OpenAI 发布了大语言模型 ChatGPT，标志着人工智能进入了大模型时代。大模型以人机对话的方式在多个领域提供智能服务，在与人的对话过程中，机器的“善解人意”和“无事不通”的人性化表现令人赞叹。生成的语言、程序代码、动画、图像、视频等内容流畅自然，问题回答、信息咨询、心理咨询、文案辅助、计划制定、娱乐消遣等功能灵巧实用。大模型涌现出的上下文理解和思维链推理等能力出乎意料，其所具备的完成多样化任务的功能令人感到通用人工智能已经不再遥远。

至此，人类社会对人工智能的关注已经超越了所有的技术领域，人工智能成为领跑公共舆论的最热门话题。产业界对大模型研发的投入几乎瞬间爆发，各大头部企业纷纷在不到一年的时间内发布了自己的大模型。在我国，以百度的文心一言、华为的盘古、阿里云的通义千问等为代表的一大批企业大模型展现了不俗的水平和能力。

2024 年 2 月，OpenAI 发布了视频生成大模型 Sora，使大模型具备了对三维物理世界进行精确建模的能力，从而引发了人们对利用大模型构建“世界模型”的思考，并为之努力。而这个方向的研究给人工智能带来的改变目前是难以估量的。

上述发展简史是如何体现发展人工智能的正确路线和正确方法的呢？这个问题需要从它的技术变迁中寻找答案。总体上，我们可以用两句话概括人工智能近 70 年的技术变迁：一是核心内容由逻辑演算和规则决策转变为与实用紧密结合的机器学习；二是基本方法由逻辑精准推理转变为相关性概率推断。这说明机器学习是发展人工智能的“牛鼻子”，抓住它就能“牵引”人工智能的整体发展，就能为实现判别、推理、决策等能力奠定基础。同时，基于深度神经网络的概率推断方法是实现人工智能最有效的方法。

人工智能的发展历史不仅是技术发展史，也是理论思想发展史。要回顾和总结它的理论思想发展史，需要结合代表人物的贡献加以梳理。

人工智能的首要奠基人当属英国数学家和逻辑学家艾伦·麦席森·图灵（Alan Mathison Turing，1912—1954）。他提出的图灵机模型直接给出了计算机以存储器为核心的体系结构，为计算机的发明做出了决定性的贡献。图灵机模型阐述的思想是，记忆是计算的基础，只要有足够的记忆能力将所有中间结果和操作都记忆下来，任何计算都可由机器实现。计算机的发明完全证明了这一点。

图灵对人工智能更直接的贡献是：他首次提出了机器能否思维，以及是否可以实现机器智能的问题。为了判断机器智能，他给出了具体的测试方法，即图灵测试。这是一个十分重要的准则，它明确地定义了人工智能的能力属性。

简而言之，所谓图灵测试就是测试机器在与人的交互过程中，回答问题的能力是否达到了人类水平。如果测试者分辨不出问题的答案是人还是某个机器给出的，那么便判断该机器具有智能。图灵测试不涉及机器的结构和工作原理是否与人脑类似，机器是否具有主观意识等问题，只要它解答问题的能力达到了人类的水平，便判断它具有智能。这显然是将人工智能的能力属性作为第一原则的理论思想。而恰恰是这一思想为人工智能沿着正确方向不断发展提

供了不竭动力。

美国麻省理工学院教授艾弗拉姆·诺姆·乔姆斯基(Avram Noam Chomsky,1928—)提出了以脑功能和语言为基础研究认知的思想,倡导从语言和认知出发,将人工智能的研究与人类智能的具体载体联系起来。他所提出的转换语法理论将自然语言处理转化为计算问题,为计算机处理自然语言奠定了基础。语言是人类思维的工具,也是人类智能最为重要的表现形式之一。因此,机器能否很好地处理自然语言是其是否具有智能的十分重要的标志。乔姆斯基对自然语言处理的研究为人工智能开辟了一个不可或缺的发展方向。大语言模型所展示出的多方面的通用能力充分说明从语言出发研究人工智能是一条十分正确的路线。不仅如此,通过将语言的概念泛化为各种有意义的符号(Token)序列,现今的大模型的理解和生成对象已经远远超越了狭义语言的界限,而将程序代码、图像、视频、基因序列、化学分子式等事物通通作为"语言"进行处理。大模型的成功深刻揭示了研究人工智能与探索广义语言规律之间紧密的内在联系。

Artificial Intelligence 这一名词的提出者之一,同为美国麻省理工学院教授的马文·明斯基(Marvin Minsky,1927—2016)是最早研究基于神经网络实现人工智能的学者之一。1951年,明斯基构建了世界上第一个神经网络模拟器,并将其称为学习机。1954 年,他以"神经网络和脑模型问题"(Neural Nets and the Brain Model Problem)为题完成博士论文,获得博士学位。1969 年,他发表著作《Perceptrons》,为神经网络的理论分析奠定了基础。同时,这部著作也指出了感知器网络在建模能力方面的不足,对后期人工智能的发展走向产生了重大影响。明斯基坚信人的思维过程可以用机器去模拟,而人工神经网络是其首先尝试的手段。1969年,明斯基获得图灵奖,是第一位获此殊荣的人工智能学者。如今,人工神经网络已经成为实现各类人工智能系统不可或缺的数学模型,而如何将更多的神经元组织起来以形成更强大的智能函数是现今技术创新的本质问题。

我国著名数学家吴文俊(1919—2017)是中国人工智能领域卓越的开拓者和贡献者。他在20 世纪 70 年代提出了数学机械化的命题,他认为中国传统数学的机械化思想与现代计算机科学是相通的,于是他开始利用现代计算机技术进行几何定理证明的研究,并取得重要成果。他的方法在国际上称为"吴方法",不仅被用于几何定理证明,还被用于开普勒定律推导牛顿定律、化学平衡问题与机器人问题的自动证明等。吴文俊的工作在人工智能早期的数学定理证明和逻辑推理等研究中独树一帜、成果卓著,这使他成为对人工智能发展做出重要贡献的中国人的杰出代表。吴文俊将人工智能与数学紧密结合的思想阐释了人工智能的数学本质,而正是人工智能的数学本质才使其能够不受伪科学和玄学的干扰而牢牢植根于科学范畴。

当代最负盛名的人工智能学者非加拿大多伦多大学教授杰弗里·辛顿(Geoffrey Hinton,1947—)莫属。他长期致力于机器学习的研究,将机器学习作为发展人工智能的主要动力。为构建高效的机器学习数学模型,他和他的团队将目光投向了深度神经网络模型。辛顿认为,相较于拥有相同数量神经元的浅层神经网络,深度神经网络具有更强的函数实现潜力。并且,拥有更多层次隐变量(神经元)及其参数的深度神经网络具有更大的潜力。而使潜力变成 AI 能力的关键是找到高效的参数学习算法。这一理论思想成为其开创深度学习研究并取得成功的基本信条,同时也决定了深度学习一直以来的技术路线。预训练大模型出现之后,人们发现了所谓的"伸缩率(Scaling Law)",即随着模型参数的增长,模型的性能也会以某个幂律关系增长。这一关系也被人们简单地概括为"大力出奇迹",因为模型参数越多,训练模型所需的算力越大。伸缩率的发现进一步验证了辛顿深度学习理论思想的正确性。

辛顿的另一个重要贡献是通过将概率论的方法与深度神经网络紧密结合，实现各类人工智能模型，从而将人工智能所依赖的数学手段由早期的基于符号系统的逻辑推理转变为基于随机变量的概率推断。概率论的方法不仅使大数据成为人工智能的关键要素之一，而且也使系统的性能直接伸缩于大数据的尺度。这是由于概率模型的推断精度总是取决于样本数量。基于概率论的大模型所展现的非凡能力越来越令人相信各种物理规律的概率性质。这种认识正在产生多方面的深刻影响，包括刷新人们对量子力学的认知。

大模型的突破性进展在给人类带来了惊喜的同时，也带来了担忧。许多人怀疑人工智能的发展速度是否太快，是否会脱离人类的掌控，甚至连辛顿也对此深表忧虑。于是，人工智能安全随即成为热点话题。包括中国和美国在内的许多国家纷纷在政府层面组织人工智能安全方面的研讨，联合国也在紧急制定相应的政策和条约。目前关注的焦点主要集中在人工智能的滥用方面，例如，如何避免其被用于恐怖组织和大规模杀伤，如何避免其被用于挑唆政治动荡，如何避免人工智能产生职业、地域及种族歧视等。

大模型的突破性进展还引燃了关于通用人工智能的热议。所谓通用人工智能也同样没有统一的定义。简单地说就是适用于所有任务、所有场合的人工智能。说到底，这只是一个含义和本质还没搞清的名词，而且是否应该或需要开发通用人工智能一直存在巨大的争议。大模型功能的多样性和适应性似乎让人们看到了通用人工智能的端倪，甚至有人将其直接看作通用人工智能。但大模型并不是通用人工智能，它仅仅是将人工智能向通用方向推进了一步，虽然幅度可观，但距离想象中的通用人工智能还十分遥远。

更具实质意义的是，大模型以能力为核心的发展路线进一步体现了人工智能的能力属性。大模型的成功也让人们看到，将人工智能作为辅助人类的工具更好地解决实际问题是发展人工智能的唯一正确道路。这也进一步彰显了人工智能的工具属性和实用属性。

认识人工智能的能力属性、工具属性和实用属性是正确认识人工智能的基础，有了这个基础，人们就能遵循人工智能的主流思想和技术，逐步深入、系统地学习它的知识，就能不受“伪人工智能”的影响，避免坠入“人工意识”“人工精神”“人机敌对”等幻想陷阱。

1.3 人工智能的内在逻辑及知识体系

人工智能是大脑能力的机器实现。在技术发展过程中，人们将记忆、计算、学习、判别、决策等作为基本能力加以研究和实现，并按照它们之间的内在逻辑加以贯穿，人工智能才逐步达到了今天的整体水平(见图1.1)。

记忆是实现所有智能的基础，这也是人类非常直观的经验。一个人如果丧失了记忆，便会成为痴呆症患者。如1.2节所述，图灵机模型的核心是记忆，有了充足的记忆空间，图灵机可以完成任何计算。现今的大模型也可以被看作是对超大规模训练数据的高效有机记忆体，即通过编码压缩、层次化关联、自组织映射等方法实现的抽象浓缩记忆结构。无论是大模型的预训练过程还是推理过程，内存占用和访问开销都是瓶颈问题。这也旁证了记忆在人工智能系统中的根基作用。记忆包含两个过程，即“记”的过程和“忆”的过程。前者是将外部知识保存到大脑的过程，后者是将保存在大脑中的知识提取出来的过程。记忆既有准确性问题，又有速度问题。人们希望在保证准确性的前提下，“记”和“忆”的速度越快越好。记忆又分为短期记忆和长期记忆，长期记忆也会被淡忘。记忆单元是有组织的，而不是相互孤立的。例如，联想

记忆可以显著提高记忆效率。这些问题恰好也是人工智能模型要处理的基本问题，即如何有效模拟记忆的上述生物机制决定着人工智能的整体水平。

计算是一种高度抽象的智能，也是其他动物所不具备的能力。计算机是人工智能的开山之作，它不仅完美实现了机器对人类计算能力的模拟和延伸，同时也为机器模拟其他智能提供了必要和有利的条件。因为无论是学习还是判别决策，都需要进行大量和高效的计算。计算也同样有准确性问题和速度问题。通常是在保证准确性的前提下，计算速度越快越好，但也有允许牺牲一定精度而追求速度的场合。除了确定性计算之外，还有概率性计算。现阶段的概率性计算往往是基于大数据的计算，是对大数据中所蕴含的不确定性的概率建模。

历史上，在实现了计算能力之后，人们的下一个着眼点主要落到了机器推理。但是，机器推理的研究过程并不顺利，期间经历了两次人工智能的寒冬。而机器学习，特别是深度学习的突破性进展，使得学习最终成为计算之后机器所实现的关键能力。人类的学习是通过观察、阅读、听闻、体验等经验获取知识的过程；而机器学习则是将这些经验变成数据或大数据，再对人工智能模型进行训练，使模型从数据中提取知识并予以保存，以便日后利用。学习既有效果问题，又有效率问题。效果问题关注获取知识的质量及结构，效率问题关注学习的速度和进度。而学习的效果和效率正是机器学习所追求的主要目标。从与其他能力的关系上讲，学习与记忆、计算，以及判别、决策都直接相关。概括来讲，计算是学习的手段，记忆是学习的结果，而学习是判别、决策的基础。

判别能力是进行事物分类、模式识别所需要的能力，是智能的典型外在表现。判别过程往往包含推理过程和预测过程，而推理和预测依赖于学习所积累的系统性知识。在机器学习的基础上，机器在不同任务中的判别能力也已经得到高度的实现。如人脸识别、语音识别、异常检测、样本分类等。

决策是以判别为前提加以实现的能力。决策时往往需要在各种判别条件下进行代价计算，以代价最小为目标进行策略选择。在人工智能系统中，决策能力一般作为末端能力加以输出，而记忆、计算、学习、判别等能力作为系统的基础能力在前端加以整合集成。随着这些基础能力的提高，机器的自动决策能力和水平也在不断取得令人惊叹的突破，如博弈程序、机器人运动、蛋白质折叠结构预测、气象预测等。

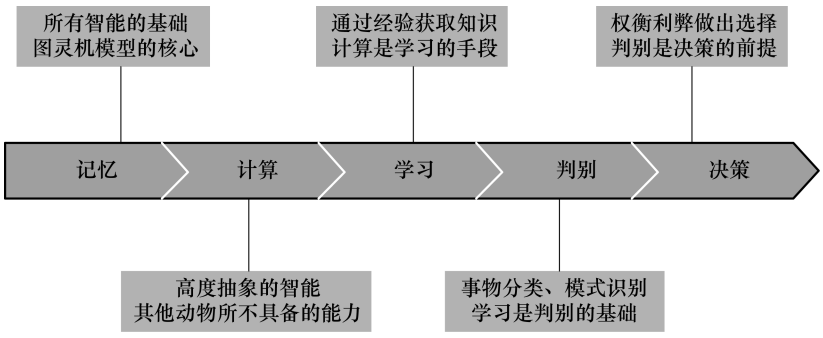

图 1.1　人工智能的内在逻辑关系

事实上，人工智能的理论发展路径也恰好遵循了上述逻辑。计算机的发明实现了记忆和计算的机器模拟，而其后便有待于学习、判别和决策方面的突破了。机器学习特别是深度学习的突破性进展带来了人工智能系统能力的显著跃升，在深度学习模型中，记忆、计算、学习、判别、决策等能力有机融合，使其整体智能得以爆发式增长。

从1956年至今，人工智能研究的里程碑成果正是循着这一逻辑不断产生，形成了次第深入、一脉相承的理论体系（见图1.2）。

1957年，被称为感知器的神经元模型问世，标志着机器学习有了基元模型。神经元模型不仅能判断输入向量与其权重向量的匹配度，权重向量还可通过误差反馈学习进行调整，这便提供了机器学习的一种基本机制。

1966年，隐马尔可夫模型（HMM）问世，将包含隐变量的概率模型用于数据分布的建模。隐变量的引入赋予了模型捕捉观测数据之间相关性的潜在因素的能力，指引了机器学习模型的发展方向。

1974年，反向传播学习算法（BP算法）问世，解决了带有隐层的神经网络学习问题，建立了机器学习的一个基本数学原理，成为一直以来训练神经网络的最主要方法。

1977年，期望最大算法（EM算法）问世，实现了对具有隐变量概率模型的估计，成为一种与BP算法并立的机器学习基本方法。

1985年，贝叶斯网络问世，给出了事物相关性的概率推断模型，进一步强化了概率论方法在智能模型中的作用。

1993年，支持向量机（SVM）问世，建立了从训练样本中选择少量样本构建最优分类器的理论，将机器学习引向了更深入的发展阶段。

1998年，卷积神经网络（CNN）问世，通过引入卷积层和池化层，实现了一个7层的深度神经网络，标志着深度学习取得突破。

2006年，用于高效数据建模的深度自动编码器在 *Science* 上发表，深度学习在理论上取得了重要突破。

2010年，深度强化学习方法问世，随后在多种博弈决策应用中获得显著成功。

2012年，综合运用多种深度学习理论和技术的大型深度神经网络AlexNet问世，显示出惊人的图像识别能力，成为被后人所仿效的经典实用化深度模型。

2014年，生成式对抗网络（GAN）模型问世，开创了两个学习体相互对抗，在对抗中相互强化的新的机器学习范式。

2017年，专注于自然语言上下文语义编码的转换器模型（Transformer）问世，在自动编码器结构中，用注意力机制完全替代了卷积神经网络和循环神经网络，人工智能具备了更高效率的机器翻译等语义序列转写能力。后续的发展表明，Transformer的真正优势在于它优异的可伸缩性，通过将更多的Transformer并联和层叠，可以处理更长的上下文，从而为构建大模型打下了基础。

2018年，基于Transformer的生成式预训练模型GPT问世，GPT将众多Transformer的基本单元加以层次化集成，构建大型语言模型，通过大规模文本数据的预训练，GPT学习了丰富的语言知识。此后GPT快速发展，版本不断更新，人工智能进入了大语言模型时代。

2021年，去噪扩散概率模型（Denoising Diffusion Probabilistic Models，DDPM）被提出，很快成为人工智能生成内容（AIGC）的有效工具。DDPM进一步强化了概率论方法在人工智能大模型时代的关键作用。

上述里程碑成果分属三个阶段，1957年至1993年的成果属于基本模型阶段，1998年至2017年的成果属于深度模型阶段，2018年之后的成果属于大模型阶段。三个阶段的模型存在逐级包含的关系，即深度模型包含基本模型，大模型包含深度模型。

这些成果是当代人工智能理论的核心内容，本质上是一系列机器学习的数学模型，它们的

特点是从神经元模型到神经网络模型，再到深度神经网络模型，直至如今的大模型，神经计算和概率是各种模型的精髓。

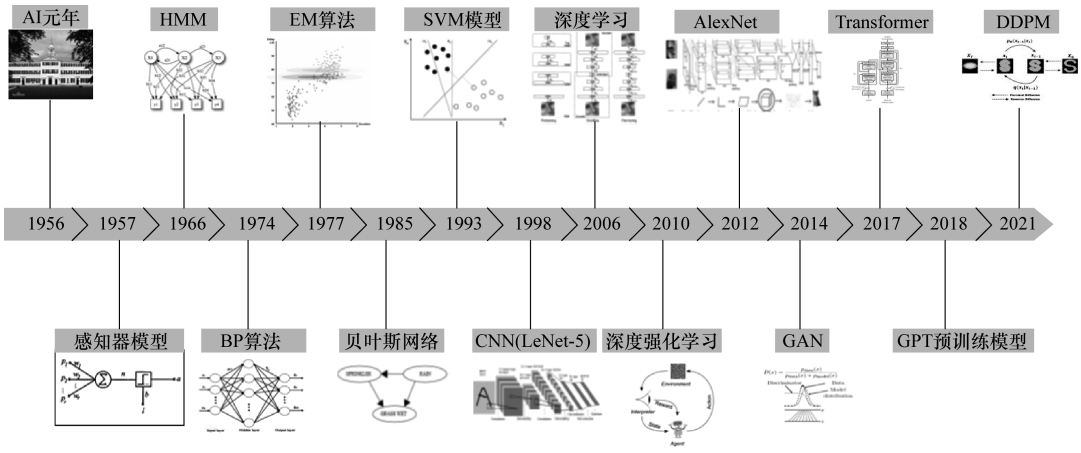

图 1.2　人工智能理论研究里程碑

伴随着理论发展，人工智能的技术创新也持续地在博弈、问答识别、运动等方向上推进。如图 1.3 所示，在博弈方向上，从黑白棋、西洋跳棋、国际象棋、围棋，到目前的电竞，机器的博弈水平正在不断地超越人类。在问答识别方向上，基于语音识别、图像识别、自然语言处理技术的人机对话系统的能力逐步提高。在运动方向上，人体运动捕捉、运动机器人、自动驾驶等技术正在走向成熟。AlphaFold 技术被用于蛋白质三维结构的预测，并取得了重大突破。2022 年 ChatGPT 的问世，使人工智能的技术水平再次大幅跃升，ChatGPT 在信息咨询、数学求解、代码生成、知识推理、科学实验等不同方面均展现出了非凡的能力。

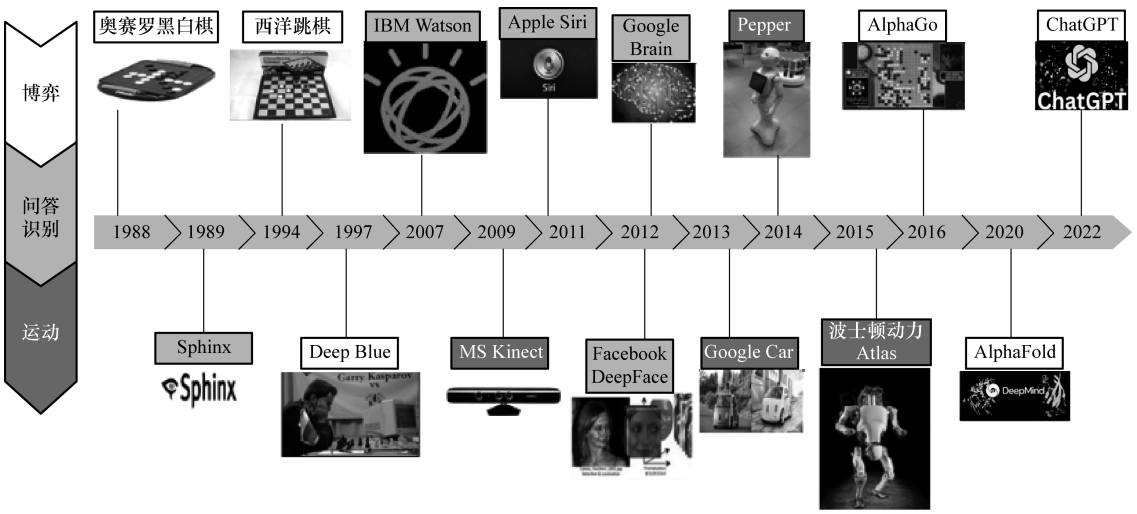

图 1.3　人工智能技术创新的主要方向

一方面，伴随技术水平的迅速提高，人工智能正在深入渗透人类社会的方方面面。在赋能产业方面，智能制造、智能物流、智能交通、智能电力等产业的变革开展得如火如荼。在造福民生方面，智慧教育、智慧医疗、智慧理财、AI 娱乐、AI 艺术等新生事物强势登场。在人工智能技术的推动下，人类社会整体上正在发生深刻的变化。

另一方面，人工智能巨大的变革作用也引起了人们合理的忧虑和关切。首先，人工智能存在现实的安全隐患。例如，包括大模型在内的人工智能系统易于受到对抗性攻击，受到攻击后，系统不仅会丧失正常功能，还可能产生有害功能，如生成虚假信息、协助欺诈、发布错误命令等。其次，人工智能的伦理问题十分复杂深刻，涉及个人隐私、数据保护、公平公正、责任义务、工作就业、军事战争等方方面面。最后，人工智能的社会角色问题也十分引人关注。人工智能可以担当辅助工具、决策支持、自主决策、伙伴协作、参与服务等不同社会角色，在界定人工智能的社会角色时，需要平衡技术的潜力和风险，以确保人工智能的发展和应用符合社会的期望和价值观。同时，也需要考虑伦理、法律和社会文化等方面的因素，制定相应的政策和准则来引导人工智能的发展和应用。

人工智能的理论和技术成果已经形成了一个完整的知识体系。整体上，这个知识体系呈现如图 1.4 所示的 4 个层次的圈层结构。从外向内，第一层是应用技术层，包含实用系统开发技术和领域知识，如问答系统、计算机视觉(CV)、自然语言处理(NLP)、信息搜索、智能机器人、智能制造、智慧医疗等；第二层是软硬件平台层，即通用开发工具层，包括开源架构和硬件芯片，如 Tensorflow、Caffe、Pytorch、Keras、Torch、Theano、MXNet、PaddlePaddle、GPU、ASIC 等；第三层是算法与模型层，包含基本计算流程和结构，如 BP 算法、卷积神经网络(CNN)、循环神经网络(RNN)等。人工智能研究的里程碑成果均处于这三个层次。核心层是数学原理层，数学原理提供了人工智能系统的根本机理，即人工智能本质上是参数可学习的复合函数。

之所以把这个知识体系描述为圈层结构，一方面是因为相邻层次之间存在基础与上层的关系，另一方面是因为各层所包含内容的多寡和精杂不同，即越向外范围越广泛、内容越丰富，越向内技术越基础、原理越单纯。

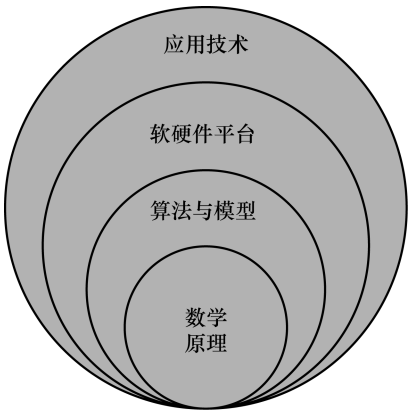

图 1.4　人工智能的圈层知识结构

1.4　人工智能的数学基础

人工智能的基本属性是能力，这种能力的特征是可用数学模型加以实现。因此，人工智能的数学原理在其整个知识体系中处于核心地位。在数学上，人工智能等价于函数。因其与一般数学函数相比具有自我优化的特点，故本书将其称为智能函数。智能函数是由基元函数组

合而成的复合函数。而所谓基元函数是指那些运算简单、功能单一的函数。尽管如此，将众多的基元函数“聚沙成塔”的智能函数却能产生异乎寻常的能力。这便是人工智能的基本数学原理。

本节首先阐述智能函数的形式和特点，然后介绍现有人工智能模型中常用的基元函数的计算原理和功能，最后介绍智能函数中参数学习的方法、方式和结构，并对深度学习的特点进行讨论。

1.4.1 人工智能的数学本质

在数学上，人工智能系统可以用智能函数(Intelligent Function)表示，其形式为

$$p=f_\theta(x) \tag{1.1}$$

其中：x 为输入变量，对应需解决的问题，如棋局、人脸、文本、声音等的观测值，通常用向量表示；p 是获得的解决问题的策略，如下棋策略、分类策略、编码方案、问题答案等，通常也用向量表示；f_θ 是智能函数，即人工智能系统，其参数 θ 通过机器学习调整，而函数的性能随着参数的优化逐渐逼近理想目标，f_θ 通常是包含众多基元函数的复合函数(Composite Function)。

经过几十年的发展迭代，智能函数 f_θ 的模型已经非常丰富，代表性的主流模型包括SVM、GMM、HMM、CRF、MCTS、MLP、CNN、RNN、Transformer、GPT 等。而常用的基元函数有线性变换、非线性激活函数(Sigmod 函数，ReLU 函数，Softmax 函数等)、离散卷积变换等。

1.4.2 常用的基元函数及其功能

构成智能函数的基元函数多种多样，但在现有的系统模型中，发挥关键作用的基元函数种类并不多，主要包括以下几类。对这些基元函数的学习和理解，是学习人工智能理论知识的起点。

1. 神经元函数

神经元函数模仿大脑神经元的工作机制，通过阈值触发机制来判断未知向量与已知向量的相关性，如果超过阈值则信息向下一级传递。如图 1.5 所示，神经元进行输入向量 $\boldsymbol{x}$ 与权重向量 $\boldsymbol{w}$ 的内积，来获取二者的相关性。而神经元的激活函数 f 对这一相关性进行非线性处理，简单逻辑函数的方法是：超过阈值的输出 1，否则输出 0，输出 1 时称神经元被激活。

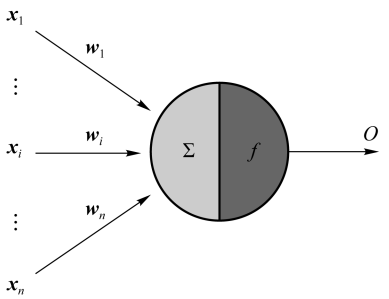

图 1.5 神经元函数

神经元的智能体现在两个方面，一是判断两个向量的相关性，二是作为判断基准的权重向量可以通过学习按需调整，参数的调整是学习过程中的重要环节。神经网络由神经元层层排布构成，一般情况下，神经元越多，神经网络功能越强。

2. 线性变换

一般的线性变换过程可以用 $\boldsymbol{y}=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}$ 来表示，用于求解未知向量 $\boldsymbol{x}$ 与多个已知向量 $\boldsymbol{w}_i$ 之间的相关性。如图 1.6 所示，线性变换完成神经网络中的基本计算过程，即将前一层神经元的输出作为本层各个神经元的激活输入。$\boldsymbol{W}$ 的各列向量 $\boldsymbol{w}_i$ 对应本层各神经元的连接权重向量，$\boldsymbol{x}$ 为输入向量，$\boldsymbol{y}$ 为各神经元的权重向量 $\boldsymbol{w}_1$，$\boldsymbol{w}_2$，$\boldsymbol{w}_3$，$\boldsymbol{w}_4$ 与 $\boldsymbol{x}$ 向量内积后形成的向量，$\boldsymbol{y}$ 的各个元素对应 $\boldsymbol{y}$ 列各神经元的激活输入。

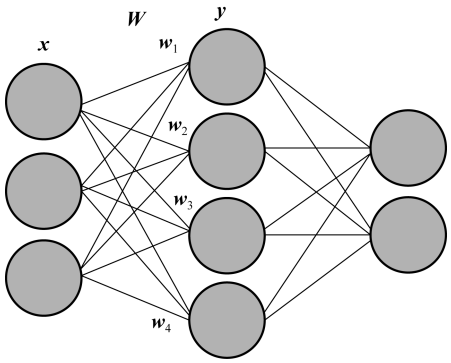

图 1.6　神经网络中的线性变换

3. 离散卷积

离散卷积是两个离散序列之间按照一定的规则将它们的有关序列值分别两两相乘再相加的一种特殊的运算。离散卷积运算广泛应用于各种工程应用场景，其具体表达式为 $\boldsymbol{y}=\boldsymbol{x}*\boldsymbol{w}$。

卷积运算可以计算未知向量 $\boldsymbol{x}$ 的多个局部向量与已知向量 $\boldsymbol{w}$ 的相关性。

现以图像处理中的二维卷积为例，对卷积操作进行解释，如图 1.7 所示。其核心操作是在特征图二维向量 $\boldsymbol{x}$ 上滑动二维卷积核向量 $\boldsymbol{w}$，并分别计算 $\boldsymbol{x}$ 的各个感受野（被覆盖的区域）与 $\boldsymbol{w}$ 的内积，$\boldsymbol{y}$ 是由各内积值构成的向量，故卷积的本质是移位内积。

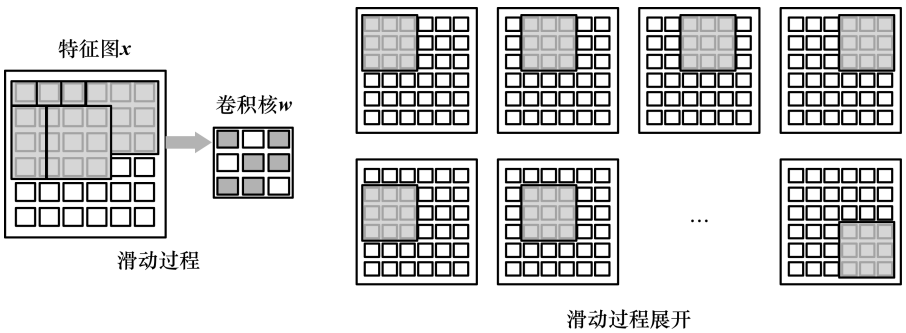

图 1.7　图像处理中的二维卷积

4. 池化（Pooling）

池化（Pooling），也称欠采样或下采样。主要用于特征降维，即压缩数据和参数的数量，降低卷积层输出的特征向量维度。

最常见的池化操作为平均池化(Average Pooling)和最大池化(Max Pooling)。平均池化是计算被池化集合中所有元素的平均值,而最大池化是计算被池化集合中所有元素的最大值。

图 1.8 显示的是对一个 4×4 的特征图用 2×2 池化器进行步长(Stride)为 2 的最大池化操作,从而将 16 维特征降低至 4 维。

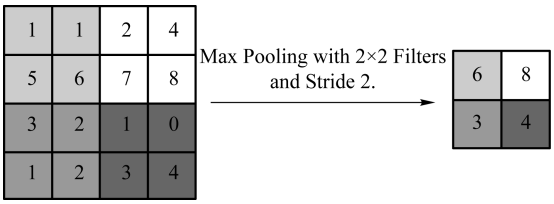

图 1.8　最大池化示意图

5. Sigmoid 函数

Sigmoid 函数,也被称为 Logistic 函数。它的公式如下:

$$S(x)=1/(1+e^{-x}) \tag{1.2}$$

Sigmoid 函数的函数图像是一条 S 形曲线,如图 1.9 所示。

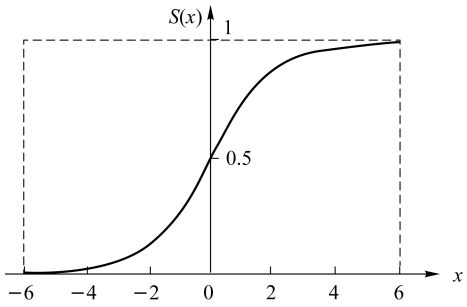

图 1.9　Sigmoid 函数曲线

从图 1.9 中我们可以观察到,Sigmoid 函数是一个值域为[0,1]、S 形的单调连续可导逻辑函数。Sigmoid 函数的功能是把输入的正负实数值压缩至 0 到 1 之间,当输入 $x=0$ 时,输出为 0.5;当输入 x 从 0 向负方向偏离时,输出迅速趋近于 0;当输入 x 从 0 向正方向偏离时,输出迅速趋近于 1。即输入 x 在偏离零点后,输出呈指数上升或下降,从而迅速饱和。故此函数也被称为逻辑(Logistic)函数。

Sigmoid 函数常用作传统神经网络的神经元激活函数。输出范围在 0 和 1 之间是它的一大优点,用处是可以把激活函数看作一种“分类的概率”,如激活函数的输出为 0.9,便可以解释为样本有 90%的概率为正样本,这种优化稳定的性质使 Sigmoid 函数可以作为输出层。函数连续便于求导且处处可导则是它的另一大优点。

然而 Sigmoid 函数也具有其自身的缺陷。Sigmoid 的饱和性容易产生梯度消失。从几何形状不难看出,原点两侧的导数逐渐趋近于 0,在反向传播的过程中,Sigmoid 的梯度会包含一个关于输入的导数因子,一旦输入落入两端的饱和区,导数就会变得接近于 0,这将导致反向传播的梯度变得非常小,此时网络参数难以得到更新,难以有效训练,这种现象称为梯度消失。一般来说,Sigmoid 网络在 5 层之内就会产生梯度消失现象。

6. ReLU 函数

ReLU 函数是常见的深度网络神经元激活函数中的一种,公式如下:

$$y=\max\{0,x\} \tag{1.3}$$

ReLU 函数是一个分段线性函数，其曲线形状如图 1.10 所示。

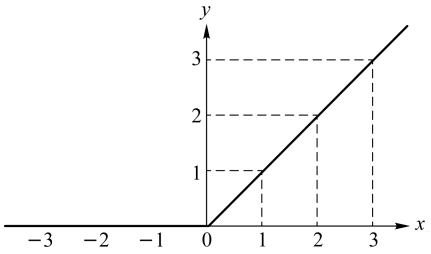

图 1.10　ReLU 函数曲线

从表达式和曲线可以看出，ReLU 是在输入 x 和 0 之间取最大值的函数。通过对非正输入零输出，对正值输入等值输出，将双极性输入变为单极性输出，故该函数也被称为整流线性单元。ReLU 函数使得同一时间只有部分神经元被激活，从而使得神经网络中的神经元有了稀疏激活性，而往往训练深度分类模型的时候，和目标相关的特征只占少数，也就是说通过 ReLU 函数实现稀疏后的模型能够更好地挖掘相关特征，拟合训练数据。

ReLU 函数的优势在于：

① 没有饱和区，在 $x>0$ 区域上，不会出现梯度饱和、梯度消失的问题；

② 没有复杂的指数运算，只要一个阈值即可得到激活值，计算简单、效率提高；

③ 实际收敛速度较快，比 Sigmoid 或 Tanh 函数快很多。

7. Softmax 函数

Softmax 函数，又称归一化指数函数。它是二分类函数 Sigmoid 在多分类上的推广，目的是将多分类的结果以概率的形式展现出来。Softmax 函数在神经网络中应用时多取如下形式：

$$P(y=i \mid x)=y_i=\frac{\mathrm{e}^{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}_i}}{\sum_{k=1}^{K}\mathrm{e}^{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}_k}} \tag{1.4}$$

式(1.4)的意义为将向量 $\boldsymbol{x}$ 输入到 K 个代表不同类别的神经元后，Softmax 函数判断向量 $\boldsymbol{x}$ 属于第 i（$i=1,\cdots,K$）个类别的概率。其中，$\boldsymbol{w}_k$ 为第 k 个神经元的连接权重向量。

为了实现将分布在负无穷到正无穷上的预测结果转换为概率的目的，Softmax 函数利用了概率的两个基本性质：①预测的概率为非负数；②各种预测结果的概率之和等于 1。

① 将预测结果转化为非负数

图 1.11 为指数函数的曲线，可以看出指数函数的值域为零到正无穷。Softmax 函数的第一步就是将模型的预测结果转化到指数函数上，这样保证了概率的非负性。

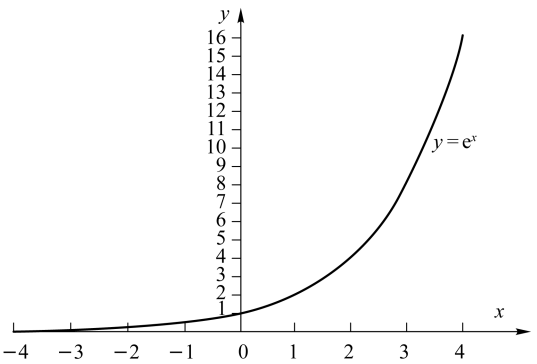

图 1.11　指数函数曲线

② 确保各种预测结果的概率之和等于 1

为了确保各种预测结果的概率之和等于 1，只需要将转换后的结果进行归一化处理。方法就是将转化后的结果除以所有转化后结果之和，可以理解为转化后结果占总数的百分比，这样就得到了近似的概率。

具体来说，Softmax 函数可以实现将未知向量 $\boldsymbol{x}$ 与各已知类别向量 $\boldsymbol{w}$ 的相关性转换为归属概率，所以其常用作神经网络顶层基元函数，给出系统的最终输出。图 1.12 的例子展示了 Softmax 函数将未知向量 $\boldsymbol{x}$ 与三个类别（$\boldsymbol{w}_1$，$\boldsymbol{w}_2$，$\boldsymbol{w}_3$）的相关性转换为归属概率的过程。

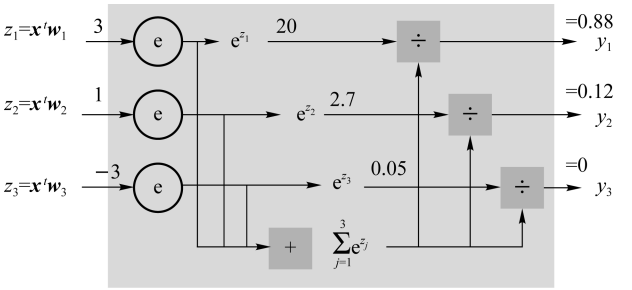

图 1.12　Softmax 函数计算过程

通过以上计算，Softmax 函数可将任意一组数值转化为一组概率值，并且数值间的差异被指数放大后，最大值所对应的概率被显著突出。图 1.12 的例子是将{3，1，−3}这组数值转换为{0.88，0.12，0}这组概率值。可见数值 3 对应的概率被显著突出。Softmax 函数的这一性质在包括大模型在内的许多人工智能模型中发挥了关键乃至神奇的作用。

1.4.3　参数学习方式

1. 监督学习

监督学习是指利用有人工标注的训练数据实现智能函数中的参数学习。例如，在分类任务中，我们需要利用训练数据建立一个分类函数，这时每个训练数据都会有一个类别标签。图 1.13 显示的是一个二分类问题，两类样本数据分别被标注了不同的颜色，利用这些数据，我们很容易找到一组参数来确定一条直线函数，以将这两类数据分开。这便是监督学习。

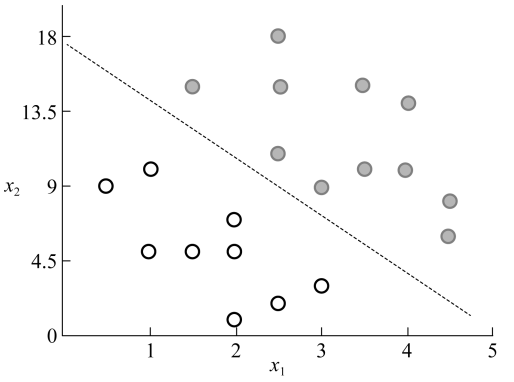

图 1.13　监督学习中的分类问题

2. 无监督学习

有别于监督学习，无监督学习所利用的数据没有人工标注信息。如图 1.14 所示，训练数据只是二维空间中的一些数据点，没有每个数据属于哪个类别的信息。无监督学习的目标是发现这些数据潜在的内部结构。例如，图 1.14 显示的任务是发现虚线所示的两个簇结构。无监督学习通常的方法是先随机设置参数的初始值，然后反复迭代“样本归属”和“更新参数”两个步骤，直至收敛。

无监督学习常用于聚类任务。聚类的目的在于把相似的东西聚在一起，而并不关心这一类是什么。因此，一个聚类算法通常只需要知道如何计算相似度就可以开始工作。

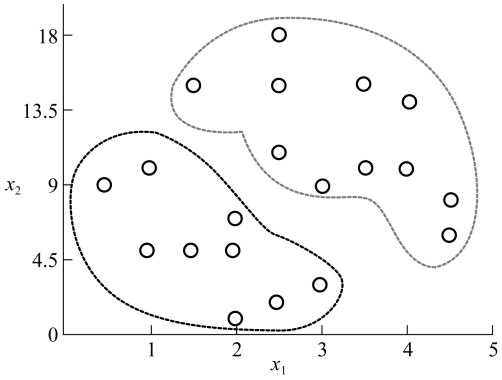

图 1.14　无监督学习中的聚类问题

3. 半监督学习

半监督学习是利用少量标注样本和大量未标注样本进行的学习，是介于监督学习和无监督学习之间的学习。目的是同时获得监督学习模型精度高和无监督学习人工标注成本低的优点。

如图 1.15 所示，两类训练数据中均只有一个样本做了类别标注，其余的都是未标注样本。学习的目标是找到将两类数据分开的那条曲线。半监督学习通常根据样本之间的聚类距离和流形结构传播标签，完成标签传播后，再进行有监督学习。

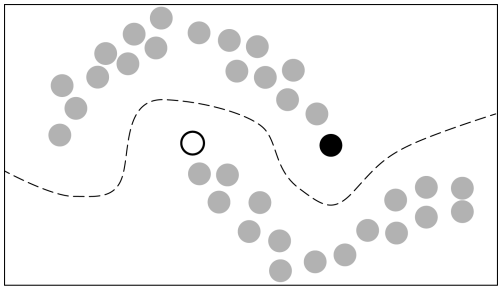

图 1.15　半监督学习中的分类问题

4. 自监督学习

自监督学习不依赖于外部的标注数据，而是利用数据本身的结构和属性来生成训练信号。自监督学习是完全数据驱动的学习方式，通常将数据中指定的部分作为预测对象来提供训练信号，例如，预测文本中的下一个单词、图像中被有意遮盖的部分或视频的下一帧。自监督学