

B 大数据丛书
IG DATA SERIES

LEARNING DEEP
ARCHITECTURES FOR AI

人工智能中的 深度结构学习

【加】尤舒亚·本吉奥 (YOSHUA BENGIO) 著

俞凯 吴科 译



 Springer

 机械工业出版社
CHINA MACHINE PRESS

Architectures for AI

人工智能中的深度 结构学习

[加] 尤舒亚·本吉奥 (Yoshua Bengio) 著

俞凯 吴科 译



机械工业出版社

以人工智能为代表的新技术正在给人们的生产和生活方式带来革命性的变化。人工智能技术试图了解智能的本质,并产生一种新的与人类智能相似的方式做出反应的智能机器。让计算机理解现实世界中诸如图像、语音和语言等数据所蕴含的高层次抽象信息,并加以利用,是该领域最大的挑战之一。诸多理论和实践成果表明,以神经网络为代表的“深度结构”是解决该问题的最重要工具之一。

本书详细论述了采用深度结构的动机、原理和理论依据,讨论了训练深度神经网络的难点,继而详尽地介绍了自动编码器、受限玻尔兹曼机以及深度置信网络的概念和理论,并进行了理论分析。本书是深入理解深度学习的动机和原理的经典之作。

本书可作为高等院校相关专业本科生和研究生的教学辅助读物,对于人工智能相关人员,科学界和业界关注机器学习特别是希望深入理解深度学习理论基础的研究者和从业者,本书值得仔细阅读。

Learning Deep Architectures for AI/by Yoshua Bengio/ISBN: 9781601982940

© China Machine Press 2016. Authorized translation of the English edition

© Yoshua Bengio. This edition is published and sold by permission of Now Publishers Inc., the owner of all rights to publish and sell the same.

This title is published in China by China Machine Press with license from Now Publishers Inc. This edition is authorized for sale in China only, excluding Hong Kong SAR, Macao SAR and Taiwan. Unauthorized export of this edition is a violation of the Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书由 Now Publishers Inc. 授权机械工业出版社在中华人民共和国境内(不包括香港、澳门特别行政区及台湾地区)出版与发行。未经许可的出口,视为违反著作权法,将受法律制裁。

北京市版权局著作权合同登记 图字:01-2016-4515号。

图书在版编目(CIP)数据

人工智能中的深度结构学习/(加)尤舒亚·本吉奥(Yoshua Bengio)著;俞凯,吴科译.—北京:机械工业出版社,2017.6

书名原文:Learning Deep Architectures for AI

ISBN 978-7-111-56935-0

I. ①人… II. ①尤…②俞…③吴… III. ①人工智能 IV. ①TP18

中国版本图书馆CIP数据核字(2017)第115636号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

策划编辑:王康 责任编辑:王康 汤嘉

责任校对:樊钟英 封面设计:路恩中

责任印制:李昂

三河市国英印务有限公司印刷

2017年7月第1版第1次印刷

169mm×239mm·7.75印张·119千字

标准书号:ISBN 978-7-111-56935-0

定价:35.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

服务咨询热线:010-88379833

读者购书热线:010-88379649

网络服务

机工官网:www.cmpbook.com

机工官博:weibo.com/cmp1952

金书网:www.golden-book.com

教育服务网:www.cmpedu.com

封面无防伪标均为盗版

译者序

深度学习是近年在学术界和产业界都获得极大重视的机器学习技术。它在图像、语音等方面取得的巨大进展使得人们对于它的实际应用充满了兴趣。而这些实际应用算法大都是基于2006年受限玻尔兹曼机以及深度置信网络的理论突破而产生的。深入理解深度结构提出的动机和原理对于学习和发展深度学习算法具有重要的意义。目前出版的大多数深度学习书籍均以算法应用为主，本书则侧重于解释算法背后的动机，并详细分析深度结构的理论基础，是一本不可多得的深入浅出的理论小册子。

本书作者 Yoshua Bengio 是国际著名的深度学习领域开拓者，本书汇集了他对深度结构的比较系统的理论思考和对深度置信网络这一核心理论的详细分析。全书分为10章，以深度结构的引入动机、引入方法以及经典结构的理论分析为主线，详细介绍了受限玻尔兹曼机、自编码器、深度置信网络以及一系列算法变体的理论及其算法分析。本书可作为高等院校相关专业本科生和研究生的教学辅助读物，对于人工智能相关人员，科学界和业界关注机器学习特别是希望深入理解深度学习理论基础的研究者和从业者，本书值得仔细阅读。

本书由上海交通大学的俞凯和吴科共同翻译，翻译过程中还得到了上海交通大学智能语音实验室的周瑛、常烜恺、陈瑞年、丁文、林弘韬、石开宇、陈哲怀、张慧峰、杨闰哲、叶子豪、李慧琛、汤舒扬、白毅伟等同学的帮助，以及机械工业出版社的大力支持。翻译中难免有疏漏和错误，欢迎批评指正。

俞凯

目 录

译者序

1 引言	1
1.1 如何训练深度结构	4
1.2 中间层表示：在不同的任务中共享特征和抽象	5
1.3 学习人工智能的必经之路	7
1.4 本书大纲	8
2 深度结构的理论优势	9
2.1 计算复杂性	12
2.2 一些非正式的论证	13
3 局部与非局部泛化性	16
3.1 局部模板匹配的局限性	16
3.2 学习分布式表示	21
4 具有深度结构的神经网络	24
4.1 多层神经网络	24
4.2 训练深度神经网络的挑战	25
4.3 深度结构的无监督学习	31
4.4 深度生成结构	32
4.5 卷积神经网络	35
4.6 自动编码器	37

5 能量模型和玻尔兹曼机	40
5.1 能量模型和专家乘积系统	40
5.2 玻尔兹曼机	44
5.3 受限玻尔兹曼机	46
5.4 对比散度	50
6 深层结构的逐层贪心训练	58
6.1 深度置信网络的逐层训练	58
6.2 堆叠自动编码器训练	61
6.3 半监督与部分监督训练	63
7 受限玻尔兹曼机和自动编码器的变体	64
7.1 自动编码器和受限玻尔兹曼机的稀疏化表示	64
7.2 降噪自动编码器	69
7.3 层内连接	71
7.4 条件 RBM 和时序 RBM	72
7.5 分解式 RBM	74
7.6 受限玻尔兹曼机和对比散度的推广	75
8 DBN 各层联合优化中的随机变分边界	78
8.1 将 RBM 展开为无限有向置信网络	79
8.2 逐层贪心训练的变分证明	80
8.3 所有层的联合无监督训练	83
9 展望	87
9.1 全局优化策略	87
9.2 无监督学习的重要性	92

9.3 开放的问题	93
10 总结	96
致谢	98
参考文献	99

引 言

在超过半个世纪的时间里，使用计算机为我们的世界建模，展示我们所说的“智能”，一直是研究的重点。显然，为了实现这一点，大量的信息应该以某种方式存储在计算机中。这些信息的存储或以显式方式或以隐式方式进行。如果要完全人工地将所有信息处理为机器可以利用的形式，以便解决问题并推广到新的情境中，其工作量是无法想象的。因此，许多学者已转而使用学习算法来捕捉这些信息的大部分。虽然人们在理解和改进学习算法方面有了很大的进展，但是人工智能仍面临着挑战。我们拥有能让机器理解场景并用自然语言描述这个场景的算法吗？除了在极其受限的情况下，的确没有这样的算法。我们有能推导出足够的语义概念并且能用这些概念和大多数人进行交流的算法吗？答案是没有。以定义得最好的人工智能任务之一的“图像理解”为例，我们还没有找到一个学习算法能发现必要的视觉和语义概念，来解释网上的大规模图片。在其他人工智能领域也有类似的情况。

考虑一个例子：解释一个如图 1.1 所示的输入图片。当人们尝试解决特定的 AI 任务（比如机器视觉或者自然语言处理）时，通常会考虑直观地将问题拆解成多个子问题或是多个层级的表示，例如物体部件以及坐标模型^[138,179,197]，它们可以在不同的物体实例中被重用。目前最先进的计算机视觉模型就构建了多层模型，将像素点作为原始输入，最后用线性函数或是核函数分类^[134,145]。其中的中间模块混合了工程化的变换和学习，例如可以先提取那些对小的几何波动不变的低层级特征（如用 Gabor 滤波器做边缘检测），再逐渐对它们做转换（如使它们在参照物改变或反转时保持不变，有时候使用池化和子采样），然后检测出

最常出现的模式。如前所述，为了从图片中获取有用信息，目前最常用且合理的方案就是从原始的像素点出发，逐渐转换成更抽象的表征，例如从出现边缘的地方开始，到更复杂但仅出现在局部的形状，再到侦测与子物体和图像的部件相关的抽象类别，最后将这些信息整合，从而获取足够的信息来回答关于图像理解的问题。

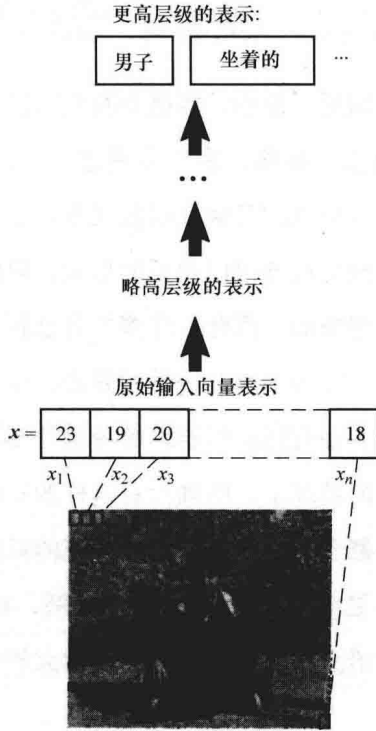


图 1.1 我们想要把原始的图像输入转换成更高级别的表示。这些表示是原始输入的函数，并且越来越抽象。例如边缘、局部形状、物体部件等。虽然语言概念可以帮助我们猜测这些更高级别的隐式的表示内容，实际上我们并不能提前知道所有层级的抽象概念所确切表示的东西

需要注意的是，假定能够做出复杂行为（或许可以被称作“智能”）的计算机需要学习一个高可变函数，即这个函数关于原始输入是高度非线性的，并且在不同目标领域里表现出非常大的波动和起伏。将学习系统的原始输入看作由多个可观测的变量组成的一种高维实体，这个实体的组成变量之间具有未知且错综复

杂的关联关系。举个例子来说，运用实物和光照的三维几何知识，可以把物理和几何上的微小改变（如位置、方向以及物体的光照）与图片上像素点的强度变化联系起来。将这些导致变化的因素叫作波动影响因子，因为它们是不同的解释视角，它们各自分别变化并且往往在统计上是独立的。在这种情况下，显式的物理因素的知识可以让我们获取一个整体的数学表达，可以用于描述因素之间的统计依赖，也可以让我们对与相同三维物体相关的图像（作为高维像素强度空间中的点）的形状有个粗略判断。如果机器能捕捉到影响数据统计波动变化的那些因素，以及它们产生观察数据的机制，那么就可以认为机器理解了真实世界中由这些波动影响因子覆盖的方面。不幸的是，一般情况下，对于隐含在自然图片里的大部分变化因素，我们并不能做解析性的理解。我们没有足够的先验知识来解释观测到的图像变化。正如图 1.1 所示，即使对于显而易见的类别也没有办法。一个像“人”这样的抽象类别其实对应着大量可能的图片，它们在像素点强度的欧式距离上可能截然不同。这类图片在像素空间中的存在非常复杂，甚至都不一定在互相连接的区域之中。在图片空间里，“人”这个类别可以看作一个高层级抽象概念。在这里所说的抽象概念可以是一个类别（如“人”）或是一个特征。特征指的是传感器输入信号的一个函数，它可以是离散的（例如“输入的句子是否是过去时态”），或者连续的（如“输入的视频展示了一个物体在以 2m/s 的速度运动”）。许多较低层级和中间层级的概念（也可以被称作抽象）对于构建一个检测“人”的系统是很有用的。较低层级的抽象和特定的感知有着更直接的联系，而更高层级的抽象则以中间层级的抽象为基础，它与实际感知的联系更微弱。

产生适当的中间层抽象是件困难的任务。此外，一个所谓“智能”的机器要掌握的图像和语义概念（例如“人”）也非常多。因此，深度结构就希望能以自动化的方式发现这些抽象，从最低层次的特征到最高层次的概念。理想情况下，我们希望人工干预尽可能少，不需要人为定义所有必要的抽象，或者是提供大量人工标注的数据。如果算法能自动处理网上存在的大量图片和文字，肯定有助于把人类的知识转换成机器可理解的形式。

1.1 如何训练深度结构

深度学习希望能学到特征的层次结构，其中较高层次特征由较低层次特征组合而来。自动地学习这些多层次的特征可以让机器学到从数据输入直接映射到输出的复杂函数，而不是完全依赖于人工特征。这对于处理高层次的抽象是很重要的，因为我们往往也不知道如何根据原始输入定义它们。随着机器学习方法数据量和应用范围的增长，自动学习强特征的能力将越来越重要。

结构的深度指的是机器学得的函数中，由非线性操作组成的层级数量。目前大多数训练算法所学到的都只是浅层结构（1~3层），但是哺乳动物的大脑用的是深度结构^[173]。原始感知的输入被多层次的抽象所表征，每一层次对应着大脑皮层的不同区域。人类也是用类似的方法处理的。大脑处理信息的方式似乎是经过了多层的转换和表示。这在灵长动物的视觉处理系统中得到了验证^[173]。其处理方式是有顺序结构的，从检测边缘的存在，到简单图形，然后是更复杂的视觉图案。

受大脑结构深度处理方式的启发，神经网络的研究者们在几十年来一直希望能训练多层神经网络^[19,191]，在2006年之前都没有成功[⊖]：他们在使用两到三层结构（也就是一到两个隐层）时有较好的结果，但是层次越深，结果越差。直到2006年才有突破性进展：多伦多大学的Hinton等人构建了深度置信网络（DBNs）^[73]，其学习算法每次只对一个层级用贪心的思想做训练，每一层采用受限玻尔兹曼机（RBM）^[51]，因此可以用无监督学习的方法训练。此后不久，基于自动编码器的相关算法也被提出^[17,153]，用的也是类似的思想——用无监督学习独立地训练中间层。最近，基于同一思想，也有一系列其他深度结构（受限玻尔兹曼机和自动编码器之外）的训练方法被提了出来^[131,202]（见第4章）。

⊖ 除了一种特殊的神经网络——卷积网络，我们在4.5节会提到。

2006 年以来, 深度网络不仅被成功地运用在分类任务上^[2,17,99,111,150,153,195], 在回归^[160]、降维^[74,158]、纹理建模^[141]、运动跟踪^[182,183]、物体分割^[114]、信息检索^[154,159,190]、机器人^[60]、自然语言处理^[37,130,202]以及协同滤波^[162]等方面都有成功的案例。虽然自动编码器、受限玻尔兹曼机以及深度置信网络使用的是无监督学习, 在上述的许多应用中, 它们已被成功地用于初始化深度有监督前馈神经网络的参数。

1.2 中间层表示: 在不同的任务中共享特征和抽象

由于深度结构可以认为由一系列层级组合而成, 随之而来的问题就是在每一层级里, 它的输出(也就是下一层的输入)都是如何表达原始数据的呢? 层级之间的连接是怎样的? 最近对于深度结构的研究重点之一就是中间层的表示: 深度结构的成功源于在中间层使用受限玻尔兹曼机^[73]、自动编码器^[17]、稀疏自动编码器^[150,153]或是降噪自动编码器^[195], 并采用无监督学习的方式学习。这些算法(会在 7.2 节具体介绍)可以看作是对“表示”(下一层级的输入)做转换, 将波动影响因子更好地拆解开。在第 4 章我们将会具体介绍, 无数的观测结果表明, 当每一层次有较好的表示后, 我们就可以用这些参数作为初始参数, 用监督学习中的梯度优化方法成功地训练一个深度神经网络。

在大脑中, 每个层次的抽象都是由一些“激活”(神经元激励)组成, 这些“激活”只占大量特征中的一小部分, 并且通常不是互斥的。由于这些特征不互斥, 它们组成了所谓的分布式表示^[68,156]——信息并不是局限在某一个神经元里, 而是分布在许多神经元之中。此外, 大脑对特征的存储似乎是稀疏的——只有大约 1% ~ 4% 的神经元在某个时刻是活跃的^[5,113]。3.2 节将会介绍稀疏分布式表示的概念, 在 7.1 节会进一步详细描述相关的机器学习方法。其中一些方法是受到大脑中稀疏表征的特点启发, 并用于搭建含有稀疏表示的深度结构。

稠密的分布式表示是这类表示的一个极端, 稀疏表示处于中间位置, 而完全的局部表示则是另一个极端。表示的局部性和所谓“局部泛化”的概念是紧密

相连的。许多现有的机器学习算法在输入空间里是局部的：为了在不同的数据空间中有不同的表现，这些算法需要有一套不同的参数（3.1节有详细介绍）。虽然当参数量很大的时候，统计效率未必很差，但是为了获得较强的泛化能力，往往需要加上一些先验知识才行（例如倾向于选择数值较小的参数）。如3.1节最后所讨论的，当这些先验知识不是针对特定任务的时候，它们可能会让模型变得很平滑。与这些基于局部泛化的模型相比，使用分布式表示所能区分的模式数目可能与表示的维数（即学习到的特征数目）呈指数关系。

在许多机器视觉的系统中，学习算法只限于这样一个处理链条的特定部分，其余部分仍需要大量人工参与。这会限制系统的规模。而且，智能机器的一个标志是能识别足够多的概念，而不只是识别“人”这个类别。因此，需要一个能处理很多不同任务和概念的计算。人工定义这么多任务显然不可能，所以自动学习在这种情境下变得非常重要。此外，任务之间和任务需求的概念之间的潜在共性非常重要，不利用这些条件是不明智的，而这一直是多任务学习^[7,8,32,88,186]的研究重点。多层次结构很自然地提供部件共享和复用：低层级的特征（如边缘检测）和中间层的特征（如局部目标）不仅对识别人是有用的，在很多其他的视觉任务里也起作用。

深度学习是基于学习跨任务可共享的中间表示的。因此，深度学习能利用无监督的数据和来自相似任务^[148]的数据解决大型任务中的数据匮乏问题。正如文献[37]显示的，它在几个自然语言任务中击败了最先进的算法。文献[2]也将相似的深度框架的多任务学习方法应用于视觉任务之中。考虑这样一个多任务情形，不同的任务有不同的输出，而这些输出从共享的高级特征池中获得。由于这些通过学习得来的特征可以在多个任务中共享，这就使得统计上的强度正比于任务的个数。而这些高级特征本身又能通过公共池中的低级别中间特征的组合来表达，统计强度再一次能用相似的方式获得，并且这个策略能在深度框架的每一个层级中使用。

此外，对于大量相关概念的学习有助于实现人类能做的“宽泛抽象”，而这目标无法通过为每个视觉类别独立的训练一个分类器去达到。如果每个高层的

类别都是由公共池中抽象特征的组合得到的，则通过这些特征的新组合就能很自然地推广到未见过的类别上。即使只有一部分这样特征的组合出现在训练样本中，由于它们表示了数据不同的侧面，新样本也会通过这些特征新的组合来有意义地表达。

1.3 学习人工智能的必经之路

在以上所提及的问题的基础上，我们把视野扩展到广义的人工智能中，对人工智能训练算法提出了一些能力上的要求。我们认为这些要求很重要，并且对研究有推动意义：

- 能学习复杂、高度变化的函数（其变化的数量远大于训练样本）。
- 能通过很少的人工输入，学习各个层级（低层级、中间层级、高层级）的抽象概念。这些抽象概念对于表示复杂的函数是有益处的。
- 能从大量样本中学习：关于样本数量的训练时间复杂度应该趋近于线性。
- 能从大部分无标注数据的数据集中进行学习（也就是半监督学习场景），其中有些数据没有完整或者正确地标注。
- 能表示大量任务之间的共性（即多任务学习）。这些共性之所以存在，是因为所有的人工智能任务都只是真实情况的不同表现方式。
- 能有很强的无监督学习能力（即能发现观测数据中的结构）。这对于突破目前很多任务的瓶颈是很有必要的，并且很多未来的任务也不能提前知道。

还有一些能力和本书没有直接的联系，但是也同样重要。例如能学习变长或变结构的上下文情境表达^[146]，从而让机器可以在上下文相关的情境下运行并针对观测数据流，做出一系列的行为决策；例如当决策会影响未来的观测和利益时，能有合理的考虑^[181]；为了收集更多关于真实世界的的数据，能采取行动做出探索（也就是主动学习的一种形式）^[34]等。

1.4 本书大纲

第2章回顾了一些理论成果（可以跳过，对于之后章节的理解没有影响），说明了结构的深度和任务需要相匹配，过于浅层的结构会使得计算元素急剧增加（对于输入规模可能是指数型的）。我们提出，过于浅层的结构是有害于学习的。如果用一个大型的浅层结构来表示任务（有大量计算元素），为了调整其中的每一个元素并学习一个高可变函数，我们会需要大量的样本。3.1节通过说明局部泛化和局部估计的缺陷，进一步说明了深度结构的动机。我们希望通过含有分布式表示的深度结构来避免这些缺陷（见3.2节）。

本书后面的部分阐述并分析了一些用于训练深度结构的算法。根据神经网络的相关文献，第4章介绍了一些关于训练深度结构的概念。我们首先指出了先前训练多层神经网络的困难之处，然后介绍了用于初始化深度神经网络的无监督学习方法。其中，很多算法（包括受限玻尔兹曼机）和自动编码器的训练算法相关：通过简单的无监督学习，在一个单层模型上得到输入的分布式表示^[25,79,156]。为了彻底理解受限玻尔兹曼机和相关的无监督学习算法，第5章介绍了一类基于能量的模型，可用于搭建含有隐变量的生成模型（如玻尔兹曼机）。第6章重点讲述了如何用逐层贪心训练算法训练深度置信网络（DBNs）^[73]以及堆叠自动编码器^[17,153,195]。第6章讨论了受限玻尔兹曼机和自动编码器的一些变体，它们被用于扩展和改进原本的模型。其中有些考虑了稀疏性和对时序依赖的建模。第8章讨论了如何通过变分方法联合训练深度置信网络的所有层级。最后，我们在第9章提出一些展望性的问题，比如在训练深度结构时的复杂优化问题。我们认为，目前训练深度结构的成功部分源于对于低层级特征的优化。我们讨论了一些延拓法的原理。它们通过最小化一个逐渐变得不平滑的代价函数，来实现深度结构的优化。

深度结构的理论优势

在本节中，通过理论上分析浅层结构的局限性，讨论研究基于深度结构的学习算法的动机。本书的这部分（包括本节和下一节）将讲述为什么会提出之后章节中描述的算法。跳过这部分并不会影响后面章节的阅读。

本节的主要观点是：过浅的结构（就可调参数的数目而言）不能有效地表达某些函数。这说明探究深度结构的学习算法是有价值的，深层结构可以表示一些被其他结构无法有效表示的函数。在简单和较浅的结构不能有效表示（甚至去学习）的目标任务上，我们可以寄希望于基于深度结构的学习算法。

当一个函数表达式具有较少的计算元素（Computational Elements）时，我们称这个函数表达式是紧凑的（Compact），即需要在学习过程中调整的自由度是小的。因此，对于有固定数量的训练样本，并且缺少其他输入到学习算法的知识来源时，我们希望目标函数[⊖]的紧凑表示将会带来更好的泛化性。

更准确地说，一个能够被深度为 k 的结构来紧凑地表示的函数，如果用一个深度为 $k-1$ 的结构来表示，可能需要指数级的计算元素。因为可承受的计算元素的数量取决于用来调整或选择它们的训练样本的数量，所以这带来的影响既是计算上的也是统计上的——使用深度不够的结构来表达某些函数时，可以预见泛化能力会比较弱。

考虑固定维数输入的情形，可以使用一个有向无环图表示机器执行的计算。图中每个节点都利用其输入执行一个函数的计算，每个输入都是图中另一个节点

⊖ 目标函数是我们想让机器发现的函数映射。

的输出或是来自图的外部的输入。整个图可以看作是一个电路，实现了对外部输入的函数计算。当计算节点允许的函数集合仅限于逻辑门时，诸如“与”“或”“非”（AND, OR, NOT），这就是一个布尔电路或逻辑电路。

为了形式化结构深度的概念，必须引入计算元素（Computational Elements）集合的概念。这种集合的一个例子是逻辑门，可以执行的运算的集合。再举一个例子就是由人工神经元（依赖于它的权重取值）执行的运算的集合。一个函数可以用给定计算元素集合中元素的组合来表示。用一个形式化这种组合的图来定义这个函数，其中每一个计算元素都用一个节点来表示。“结构深度”是指计算元素连接图的深度，即从一个输入节点到输出节点的最长路径。当计算元素集合是人工神经元可执行的运算的集合时，深度对应于神经网络中的层的数目。让我们用不同深度的结构的实例来探索深度的概念。考虑函数 $f(x) = x * \sin(a * x + b)$ 。如图 2.1 所示，它可以表示为简单运算的组合。这些简单运算如加法、减法、乘法以及 \sin 运算。在这个例子中，乘法 $a * x$ 和最后的关于 x 的乘法会用不同的节点来表示。图中的每一个节点都和一个输出值相关联。这些输出值都是使用一些函数在输入值上进行计算得到的。而这些函数的输入值又是图中其他节点的输出值。例如，在逻辑电路中，每个节点可以计算一个小的布尔函数集合中的某个布尔函数。该图作为一个整体具有输入节点和输出节点，并表达了一个从输入到输出的函数。一个结构的深度是从图中任意输入到任意输出路径的最大长度，如在图 2.1 中， $x * \sin(a * x + b)$ 的深度是 4。

- 如果在计算元素集合里包含仿射（Affine）操作及其与 S 型函数（Sigmoid）的可能组合，则线性回归和逻辑回归的深度为 1，即只有一层（Level）。

- 当我们把一个固定的核计算 $K(u, v)$ 与其他仿射（Affine）操作放在允许的运算集合里，带有固定核的核机器（Kernel Machines）^[166] 可以被认为具有两层深度。第一层对于每个 x_i （选定的代表性训练样例），都有一个计算元素计算核函数值 $K(x, x_i)$ ，把输入向量 x 与代表样本 x_i 匹配起来。第二层进行仿射组合（Affine Combination） $b + \sum_i \alpha_i K(x, x_i)$ 把匹配好的 x_i 和期望的响应关联起来。

- 当我们把人工神经元（仿射变换后接非线性变换）放进计算元素集合里，