

[机器学习](#) [深度学习 \(Deep Learning\)](#)

深度学习如何入门？

本题已加入知乎圆桌 » [「机器学习·学以致用」](#)，更多「机器学习」相关话题讨论欢迎关注。...

显示全部 ▾

已关注

写回答

邀请回答

👍 好问题 137

💬 14 条评论

🔗 分享 ...

查看全部 290 个回答



Jacky Yang

深度学习，机器学习，人工智能

🌟 编辑推荐

共 3 项收录

JunMa 等 13,981 人赞同了该回答

关于深度学习，网上的资料很多，不过貌似大部分都不太适合初学者。这里有几个原因：1.深度学习确实需要一定的数学基础。如果不用深入浅出地方法讲，有些读者就会有畏难的情绪，因而容易过早地放弃。2.中国人或美国人写的书籍或文章，普遍比较难一些。我不太清楚为什么，不过确实是这样的。

深度学习，确实需要一定的数学基础，但真的那么难么？这个，还真没有。不信？听我来给你侃侃。看完，你也会觉得没那么难了。

本文是针对初学者，高手可以无视，有不对的地方，还请多多批评指正。

这里，先推荐一篇非常不错的文章：《1天搞懂深度学习》，300多页的ppt，台湾李宏毅教授写的，非常棒。不夸张地说，是我看过最系统，也最通俗易懂的，关于深度学习的文章。

这是slideshare的链接：http://www.slideshare.net/tw_dsconf/ss-62245351?qid=108adce3-2c3d-4758-a830-95d0a57e46bc&v=&b=&from_search=3

没梯子的同学，可以从我的网盘下：链接：pan.baidu.com/s/1nv54p9... 密码：3mtY

要说先准备什么，私以为，其实只需要知道导数和相关的函数概念就可以了。高等数学也没学过？很好，我就是想让文科生也能看懂，您只需要学过初中数学就可以了。

其实不必有畏难的情绪，个人很推崇李书福的精神，在一次电视采访中，李书福说：谁说中国人不能造汽车？造汽车有啥难的，不就是四个轮子加两排沙发嘛。当然，他这个结论有失偏颇，不过精神可嘉。

▲ 赞同 1.3 万 ▼

💬 515 条评论

猪的增长率是10头/年，也就是说，导数是10。函数 $y=f(x)=10x+30$ ，这里我们假设王小二第一年卖了30头，以后每年增长10头， x 代表时间（年）， y 代表猪的头数。当然，这是增长率固定的情形，现实生活中，很多时候，变化量也不是固定的，也就是说增长率也不是恒定的。比如，函数可能是这样： $y=f(x)=5x^2+30$ ，这里 x 和 y 依然代表的是时间和头数，不过增长率变了，怎么算这个增长率，我们回头再讲。或者你干脆记住几个求导的公式也可以。

深度学习还有一个重要的数学概念：偏导数，偏导数的偏怎么理解？偏头疼的偏，还是我不让你导，你偏要导？都不是，我们还以王小二卖猪为例，刚才我们讲到， x 变量是时间（年），可是卖出去的猪，不光跟时间有关啊，随着业务的增长，王小二不仅扩大了养猪场，还雇了很多员工一起养猪。所以方程式又变了： $y=f(x)=5x_1^2+8x_2+35x_3+30$ 这里 x_2 代表面积， x_3 代表员工数，当然 x_1 还是时间。上面我们讲了，导数其实就是变化率，那么偏导数是什么？偏导数无非就是多个变量的时候，针对某个变量的变化率呗。在上面的公式里，如果针对 x_3 求偏导数，也就是说，员工对于猪的增长率贡献有多大，或者说，随着（每个）员工的增长，猪增加了多少，这里等于35---每增加一个员工，就多卖出去35头猪。计算偏导数的时候，其他变量都可以看成常量，这点很重要，常量的变化率为0，所以导数为0，所以就剩对 $35x_3$ 求导数，等于35。对于 x_2 求偏导，也是类似的。求偏导我们用个符号表示：比如 y/x_3 就表示 y 对 x_3 求偏导。

废话半天，这些跟深度学习到底有啥关系？有关系，我们知道，深度学习是采用神经网络，用于解决线性不可分的问题。关于这一点，我们回头再讨论，大家也可以网上搜一下相关的文章。我这里主要讲讲数学与深度学习的关系。先给大家看几张图：

network

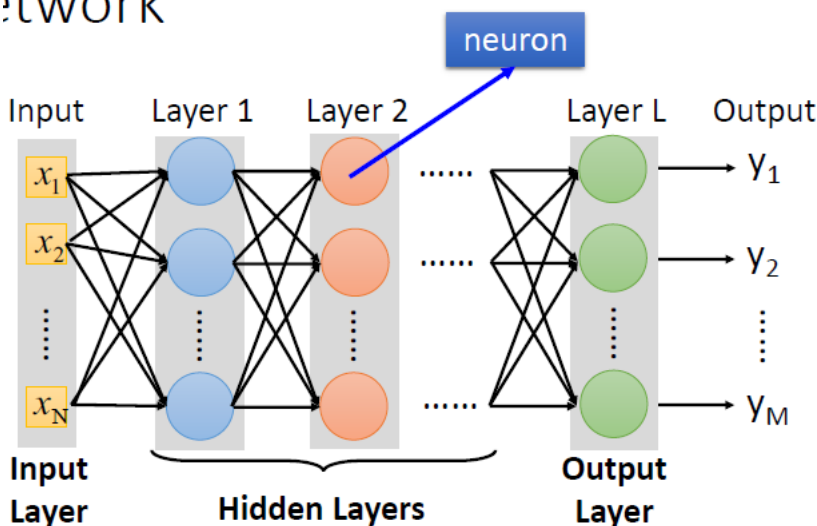


图1. 所谓深度学习，就是具有很多个隐层的神经网络。

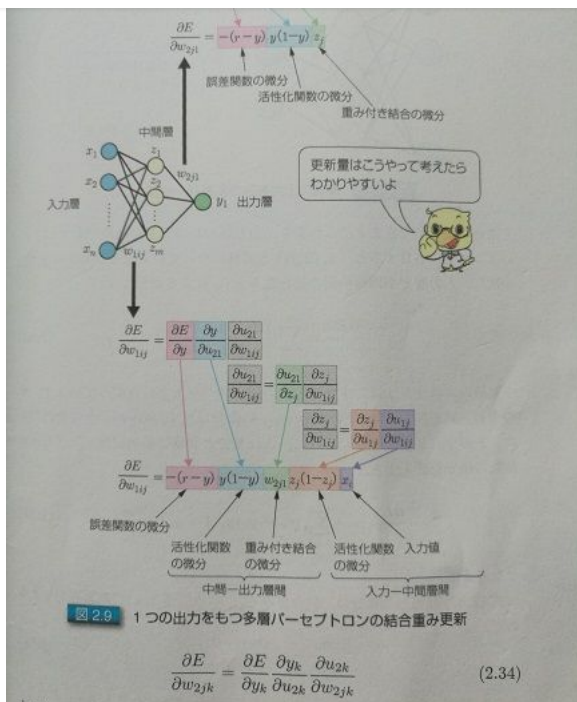


图2.单输出的时候，怎么求偏导数

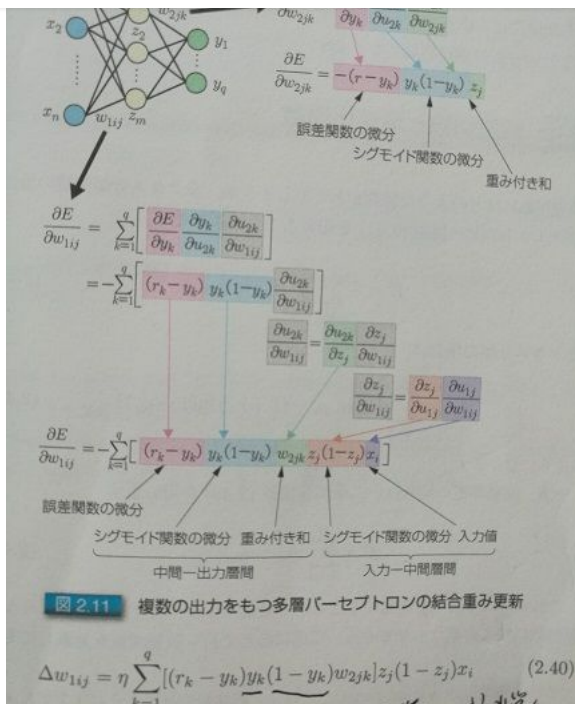


图3.多输出的时候，怎么求偏导数。后面两张图是日语的，这是日本人写的关于深度学习的书。感觉写的不错，把图盗来用一下。所谓入力层，出力层，中间层，分别对应于中文的：输入层，输出层，和隐层。

大家不要被这几张图吓着，其实很简单的。干脆再举一个例子，就以撩妹为例。男女恋爱我们大致可以分为三个阶段：1.初恋期。相当于深度学习的输入层。别人吸引你，肯定是有许多因素，比如：身高，身材，脸蛋，学历，性格等等，这些都是输入层的参数，对每个人来说权重可能都不一样。2.热恋期。我们就让它对应于隐层吧。这个期间，双方各种磨合，柴米油盐酱醋茶。3.稳定期。对应于输出层，是否合适，就看磨合得咋样了。

大家都知道，磨合很重要，怎么磨合呢？就是不断学习训练和修正的过程嘛！比如女朋友喜欢草莓蛋糕，你买了蓝莓的，她的反馈是negative，你下次就别买了蓝莓，改草莓了。-----

----- 看完这个，有些小伙可能要开始对自己女友调参了。有点不放心，所以补充一下。撩妹和深度学习一样，既要防止欠拟合，也要防止过拟合。所谓欠拟合，对深度学习而言，就是训练得不够，数据不足，就好比，你撩妹经验不足，需要多学着点，送花当然是最基本的了，还需要提高其他方面，比如，提高自身说话的幽默感等，因为本文重点并不是撩妹，所以就不展开讲了。这里需要提一点，欠拟合固然不好，但过拟合就更不合适了。过拟合跟欠拟合相反，一方面，如果过拟合，她会觉得你有陈冠希老师的潜质，更重要的是，每个人情况不一样，就像深度学习一样，训练撩妹而言，她会觉得你受前任(训练集)影响很大，这是大忌！如

深度学习也是一个不断磨合的过程，刚开始定义一个标准参数（这些是经验值。就好比情人节和生日必须送花一样），然后不断地修正，得出图1每个节点间的权重。为什么要这样磨合？试想一下，我们假设深度学习是一个小孩，我们怎么教他看图识字？肯定得先把图片给他看，并且告诉他正确的答案，需要很多图片，不断地教他，训练他，这个训练的过程，其实就类似于求解神经网络权重的过程。以后测试的时候，你只要给他图片，他就知道图里面有什么了。

所以训练集，其实就是给小孩看的，带有正确答案的图片，对于深度学习而言，训练集就是用来求解神经网络的权重的，最后形成模型；而测试集，就是用来验证模型的准确度的。

对于已经训练好的模型，如下图所示，权重（ w_1, w_2, \dots ）都已知。

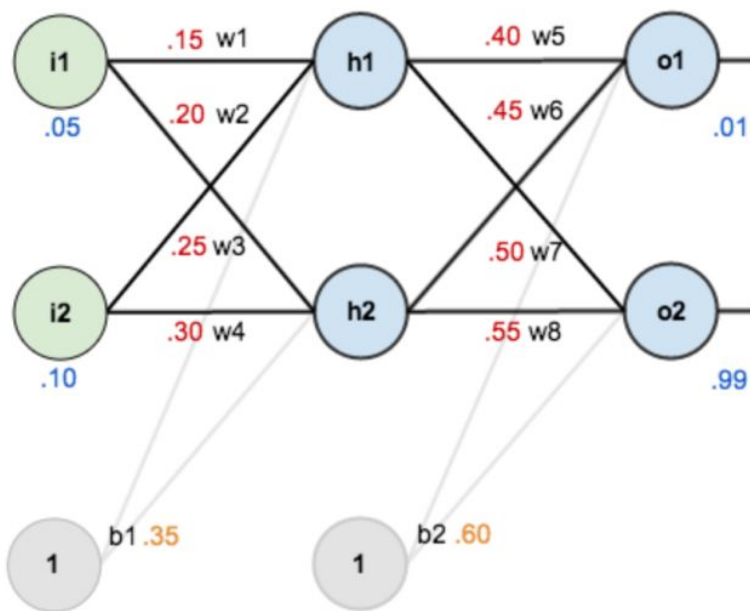


图4

$$net_{h1} = w_1 * i_1 + w_2 * i_2 + b_1 * 1$$

$$net_{h1} = 0.15 * 0.05 + 0.2 * 0.1 + 0.35 * 1 = 0.3775$$

我们知道，像上面这样，从左至右容易算出来。但反过来呢，我们上面讲到，测试集有图片，也有预期的正确答案，要反过来求 w_1, w_2, \dots ，怎么办？

绕了半天，终于该求偏导出场了。目前的情况是：

1.我们假定一个神经网络已经定义好，比如有多少层，都什么类型，每层有多少个节点，激活函数（后面讲）用什么等。这个没办法，刚开始得有一个初始设置（大部分框架都需要define-and-run，也有部分是define-by-run）。你喜欢一个美女，她也不是刚从娘胎里出来的，也是带有各种默认设置的。至于怎么调教，那就得求偏导。

2.我们已知正确答案，比如图2和3里的 r ，训练的时候，是从左至右计算，得出的结果为 y ， r 与 y 一般来说是不一样的。那么他们之间的差距，就是图2和3里的 E 。这个差距怎么算？当然，直接相减是一个办法，尤其是对于只有一个输出的情况，比如图2；但很多时候，其实像图3里的那样，那么这个差距，一般可以这样算，当然，还可以有其他的评估办法，只是函数不同而已，作用是类似的：

$$E_{total} = \sum \frac{1}{2} (target - output)^2$$

不得不说，理想跟现实还是有差距的，我们当然是希望差距越小越好，怎么才能让差距越来越小呢？得调整参数呗，因为输入（图像）确定的情况下，只有调整参数才能改变输出的值。怎么调整，怎么磨合？刚才我们讲到，每个参数都有一个默认值，我们就对每个参数加上一定的数值 Δ ，然后看看结果如何？如果参数调大，差距也变大，你懂的，那就得减小 Δ ，因为我们的目标是要让差距变小；反之亦然。所以为了把参数调整到最佳，我们需要了解误差对每个参数的变化率，这不就是求误差对于该参数的偏导数嘛。

关键是怎么求偏导。图2和图3分别给了推导的方法，其实很简单，从右至左挨个求偏导就可以。相邻层的求偏导其实很简单，因为是线性的，所以偏导数其实就是参数本身嘛，就跟求解 x_3 的偏导类似。然后把各个偏导相乘就可以了。

这里有两个点：

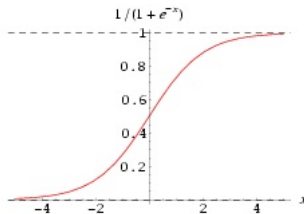
这里有两个点：一个是激活函数，这主要是为了让整个网络具有非线性特征，因为我们前面也提到了，很多情况下，线性函数没办法对输入进行适当的分类（很多情况下识别主要是做分类），那么就要让网络学出来一个非线性函数，这里就需要激活函数，因为它本身就是非线性的，所以让整个网络也具有非线性特征。另外，激活函数也让每个节点的输出值在一个可控的范围内，这样计算也方便。

貌似这样解释还是很通俗，其实还可以用撩妹来打比方；女生都不喜欢白开水一样的日子，因为这是线性的，生活中当然需要一些浪漫情怀了，这个激活函数嘛——我成年们生活中的小浪漫小惊喜，是不是？相处的每个阶段，需要时不时激活一下，制造

几个回合，她就离不开你了。因为你的非线性特征太强了。

当然，过犹不及，小惊喜也不是越多越好，但完全没有就成白开水了。就好比每个layer都可以加激活函数，当然，不见得每层都要加激活函数，但完全没有，那是不行的。

由于激活函数的存在，所以在求偏导的时候，也要把它算进去，激活函数，一般用sigmoid，也可以用Relu等。激活函数的求导其实也非常简单：



求导： $f'(x) = f(x) * [1 - f(x)]$ 这个方面，有时间可以翻看一下高数，没时间，直接记住就行了。至于Relu，那就更简单了，就是 $f(x)$ 当 $x < 0$ 的时候 y 等于0，其他时候， y 等于 x 。当然，你也可以定义你自己的Relu函数，比如 x 大于等于0的时候， y 等于 $0.01x$ ，也可以。

另一个是学习系数，为什么叫学习系数？刚才我们上面讲到 Δ 增量，到底每次增加多少合适？是不是等同于偏导数（变化率）？经验告诉我们，需要乘以一个百分比，这个就是学习系数，而且，随着训练的深入，这个系数是可以变的。

当然，还有一些很重要的基本知识，比如SGD（随机梯度下降），mini batch 和 epoch（用于训练集的选择），限于篇幅，以后再侃吧。其实参考李宏毅的那篇文章就可以了。

这篇拙文，算是对我另一个回答的补充吧：[深度学习入门必看的书和论文？有哪些必备的技能需学习？ - jacky yang 的回答](#)

其实上面描述的，主要是关于怎么调整参数，属于初级阶段。上面其实也提到，在调参之前，都有默认的网络模型和参数，如何定义最初始的模型和参数？就需要进一步深入了解。不过对于一般做工程而言，只需要在默认的网络上调参就可以了，相当于用算法；对于学者和科学家而言，他们会发明算法，难度还是不小的。向他们致敬！

写得很辛苦，觉得好就给我点个赞吧:)

关于求偏导的推导过程，我尽快抽时间，把数学公式用通俗易懂的语言详细描述一下，前一段时间比较忙，抱歉:)

更多回答



张宏伦

上海交通大学 工学博士

527 人赞同了该回答

实战是学习的最好路径。

强烈推荐下自己最近开源的项目，《深度有趣-人工智能实战项目合集》，呕心沥血大半年才完成的作品。

30多个精选实战项目，包括图片分类、人脸识别、物体检测等计算机视觉应用，中文分词、词向量训练、古诗自动生成等自然语言处理项目，DCGAN、CGAN、ACGAN、CycleGAN等经典生成式对抗网络，以及图像标题生成、机器翻译、语音识别和分类、图像风格迁移等有趣的人工智能案例。



项目使用Python3、TensorFlow1.9和Keras2.2，所涉及的文档、代码、数据、模型 **全部免费开源**，是的，**全部免费开源**。

182页的项目说明文档使用Latex编写，主要是为了达到更好的排版效果，代码也经过精心整理、反复校对、确认无误。

- 开源项目：<https://github.com/Honlan/DeepInterests>，欢迎star，这里可以下载到182页的项目说明文档；
- 项目网盘：pan.baidu.com/s/1zQRT5...，这里可以下载到所有项目所涉及的代码、数据、模型等资源，分为完整版和精简版两个版本，前者包括项目所涉及的完整资源，后者只包括最后执行所需的必要文件；
- 知乎专栏：[深度有趣](#)，里面包含了全部的线上版文档和代码。

希望这个项目能被更多人了解、从中有收获并找到乐趣。

▲ 赞同 1.3 万 ▼

● 515 条评论

▲ 赞同 527 ▼

● 25 条评论

➦ 分享

★ 收藏

♥ 喜欢

...

收起 ^



盐选成长计划

知乎 官方帐号

本回答节选自盐选专栏，有助于解答该问题。

盐选专栏名：《超级搜索力：快速链接信息、人脉、财富、资源》

作者：刘 sir 「个人发展学会」创始人

努力学习不一定就能够收获等同的回报，还需要你拥有技巧。找到知识之间的底层逻辑并建立知识链、找到考试的目的并为此建立一个知识体系、活学活用并随时调取知识，这是能让你变得高效、主动、卓有成效的秘籍。

学新知识时，人们常会用关键词或关键信息与旧知识相联系，例如记 dogma 这个词，dog 是狗的意思，ma 是表示妈妈的拼音，dogma 是「教条」的意思，这个词就可以记成「狗妈妈就是小狗的教条」。而考试会加强新知识和旧知识之间的联结，帮助人更好地学习新知识。

温故而知新的一种办法，就是向内检索。想要进入一个新的知识领域，获得全新的信息，就要向外检索了。这种检索不是简单地在搜索引擎中敲几个关键词，而是要首先有目的地建立思维模型，其次要把自己搜寻到的知识分门别类地放到这个思维模型中去，最后还要找到这些知识的应用场景。

有时候你会在大学生中间发现这种让人痛心的现象：他们空守好几版教材、历年真题试卷和一个图书馆，却不知道怎么查、怎么用，最后把力气花在了那些最不值得的事情上。

你说是他们不会使用搜索引擎、不会去图书馆检索书目吗？不是的。他们的困难不在这里。据我观察，很多人的真正问题是，他们抓不到考试的重点，沉沦在题库的汪洋大海中，找不到考试的意义，甚至憎恨考试。即使通过了考试，也没有提高自己在相应领域的能力。

针对这样的问题，我和你分享三个解决方法。

第一、找到知识之间的底层逻辑，建立知识链。

找到一些零散知识之间的底层逻辑，就是在锻炼你建立完整的知识链的思维。

那么如何找到零散知识之间的底层逻辑？具体来说有 4 个技巧。

(1) 学会找相同。

整体上找相同和把它拆解到局部找相同。中间有很多的维度，你要去定义一个类型或者角度，然后去找到事物之间的联系。比如说比喻，其实就是建立联系的一个基本方式。善用比喻，其实就善于去找到事物之间的底层联系。

▲ 赞同 1.3 万 ▼

● 515 条评论

在互联网的一个分类方式里面，就是用标签去分类。比如益智类，教育类等。学会打标签，用一个词去定义一件事物，去归纳一样知识的底层逻辑就是标签法，也是我们说归纳法。

知识其实也是可以通过打标签的方式，去找到思路之间的这种底层逻辑。比如说，这个知识的共性是什么，去定义一个标签，然后进行归纳。

(3) 学会演绎法。

我们通常不善于去记住单点的知识，更善于记住逻辑。所以需要把寻找到的知识线索，按着一定逻辑顺序进行推演。比如按着是什么，为什么，怎么办；或者按着时间、相关性的强弱、重要程度，这些逻辑顺序，来进行推演。

如果把它放到一个老练的知识里面去讲的话，你就应该会在你的知识上之间去找到一个线索去串

最低 0.3 元/天开通会员，查看完整内容 >

购买该盐选专栏查看完整 25 篇内容



超级搜索力：快速链接信息、人脉、财富、资源

刘 sir 「个人发展学会」创始人

盐选专栏

¥99.00 会员免费

发布于 08-05

▲ 赞同 ▼

● 添加评论

➤ 分享

★ 收藏

♥ 喜欢

...

收起 ^

查看全部 290 个回答