

ML&DEV[4] | 机器学习进阶线路

机智的叉烧 北京科技大学 统计学硕士

关注他

8 人赞同了该文章

导读：大家好，我是机智的叉烧，这是我ML&DEV系列下的第4篇文章，本系列下主要谈和机器学习算法工程有关的内容。在之前的章节中我谈到了机器学习的入门，本期就谈谈如何在入门之后进阶，提升自己的段位，完成新手小白到一名高手的蜕变。另外，这也是机器学习线路的最后一篇，后续ML&DEV将会有其他精彩内容，敬请期待！
更多文章欢迎关注：
我的专栏：数学 数据 计算机
我的公众号：CS的陋室

之前讲到了机器学习入门，主要通过计算机基础的掌握、数学基础的掌握，再到机器学习的基础入门，那么在此基础上，包括我在内其实都会遇到一个瓶颈——学完这些，我需要什么继续做什么，我能够做什么，这其实是一个不小的问题，我为此思考了很久，现在来总结一下，希望对大家有所帮助。

当然的，这篇文章写起来其实就很虚，能感觉到有部分同学会觉得没有大家所谓的“干货”，其实吧很多思维上的东西，就是比较虚，你学起来可能也比较未知不觉，但是一些内化到体内的东西，有时候输出出来确实就是不一样。

懒人目录：

- 数学能力进阶
- 开发能力进阶
- 应用能力进阶
- 分支领域

数学能力进阶

基本的数学知识，其实已经在入门阶段完成，而在此阶段，其实并不需要再特别学新的数学知识了，而是锻炼“数学思维”，说起来可能有点玄学，但是实际上有这个思维在机器学习中是能起到重要作用的，虽然你很难感知。在我的理解的比较常用的数学思维主要是下面的内容，求同存异哈：

- 能够自主地进行公式推导。（绝对基础能力）
- 更容易接受新的数学知识。在你学习的时候能够体现。
- 遇到一个事情不是想当然而是去考虑是否合理。
- 能够提出比较大胆模型改进。
- 抽象能力与抽象思考能力，抽象是指不能把你眼前这个直观的问题抽象成一个模型，抽象思考我的理解是能够把抽象的模型进行推导和分析。
- 数据敏感性，对数据有自己的独特理解，能识别或者利用数据陷阱。

可能专业的说法会将数学思维分成类似逻辑思维、空间思维等，那个说的太玄学，我只把一些自己能想到、会用到的东西说给大家听就行了。

说说怎么培养，我说几个比较有用的提升方法。

- 多看论文，里面理论部分不能跳过（例如问问自己，CNN的反向传播是怎么传的）
- 看论文的时候，多想想他的创新点是怎么来的，优点和缺点是怎么体现出来的。
- 看点这方面的书，推荐一本——《统计陷阱》，给很多人推荐过，里面的套路可以知道怎么识别，也可以学习怎么使用。
- 东西别只管拿来主义，多想想是否适合你这个场景，这里其实是两个问题，工具的具体性质如何以及你的场景又是如何，两个理解了才能知道适不适合。

这是一个修炼过程，你在提升的过程中很难发觉，但是确实是一些很有利于你解决问题的思维方式，慢慢训练，让你的脑子在遇到问题时快速往这里想，形成模式，很有用。

开发能力进阶

说白了，开发能力同样是一个很抽象的能力，在我一个段位不是很高的开发玩家角度，你可能需要提升这些方面的能力：

- 你的代码要足够高效，速度上、空间占用上，或者在有限资源下能够给出尽可能可行的方案。说白了就是数据结构与算法，这块你是否知道。
- 代码的可读性，可拓展性等。不要写完的代码过几天你自己都看不懂。
- 掌握一项新技术的能力。能形成自己的学习模式，知道要掌握什么，或者精准识别什么东西是自己马上或者未来要用的。
- 需求分析与实现能力。现实问题大都抽象，怎么落实到代码层面，其实很关键。

说的都是能力，那么怎么提升呢：

- 有关数据结构与算法，还是很建议大家一方面多看书，简单的复杂的可以循序渐进的看，另一方面多动手，LeetCode之类的不见得是你需要笔试你再来做，偶尔做做怡情也无法。
- 上github上多看看大项目，高分项目，你就知道该怎么整了。
- 有关学习能力，慢慢摸索吧，自己也没有总结出来，能告诉大家的是，尤其是准备毕业的同学，一定要意识到再也没有老师，也没有最匹配你需求的课程了，课程只是个学习方法，具体要什么你要学会通过网络资源，零散性或者系统性获取。
- 有机会就做项目，没机会做做kaggle之类的，多思考问题要怎么整理抽象，怎么选择最佳方案等。

分支领域

很多人到了这个阶段都会选择到分支领域，在某个方向上进行深入研究，当然是没有问题的，这也是一个突破瓶颈最简单快捷的方法（这里指的简单是指选择上的简单，问题上的难度丝毫没有下降，甚至还可能上升）。目前的比较火的选择是推荐系统、自然语言处理和机器视觉，当然还会有很多比较现实的问题。

推荐系统，我的理解更加抽象，是一个连接服务提供方和服务消费方两者的模型，选择最合适的进行匹配，做类似的问题的，都叫做推荐系统，于是哪怕是高德线路规划、滴滴的派单，其实都可以抽象的理解为推荐系统。由于场景的各异，会涌现很多不同的方案，视频、图片、文字、声音、线路、司机这些都是待推荐的产品，这些在机器学习的视角下，就是改善特征、设置合理的模型结构，例如wide&deep等，从而使推荐效果更好。

自然语言处理，这是一类问题，里面包含了很多任务，分类、匹配、标注、生成等，甚至包括语音识别，问题非常明朗，目标很明确，但是对于语义的理解和转化，换言之，特征的构建和转化其实并不简单，我们人自己都无法对一个问题说出文本上对解决任务有用的统一特征，可能更多是规则吧，因此在进行处理的时候，借助相对黑箱的深度学习可能会有更好的效果，至于深度学习上有什么好用的结构，简单的CNN之类的，而难度提升，要求对信息有更高的识别能力，transformer之类的也应运而生，在我的视角下，很多创新型，让人有眼前一亮感觉的深度学习结构，都是从NLP领域产生，毕竟NLP似乎真的是比较难的一个领域。

有关CV，我本身不太懂，不乱说啦~

有关这块的路线，我会根据需求写我的攻略，之前多少写过一些，不知道大家看的是否满意：

NLP&TMM | 我的NLP学习之路
mp.weixin.qq.com/s/hUNMuBLxpS8PwHt...

独当一面

R&S | 手把手搞推荐[0]：我的推荐入门小结
mp.weixin.qq.com/s/wXFNuGi26g7RhBLB...

手把手搞推荐系统

小结

总的来说，要进阶，其实和初步学习的方法差不多，难在思路上，不知道自己该看什么做什么，该怎么提升，那么我来总结一下吧。

- 多看看别人怎么做的，思路、推导、代码。至于如何定义这个“看”，其实就是能不能够从中抽取一些你能用在未来的一些经验。
- 多动手，看一些有深度的材料，博客可能满足不了大家的进阶水平了，看论文，看github，看一些比较有深度的书，并多动手尝试。
- 分支领域非常推荐，也更有利于找工作。

前两条又总结为一个成语——“手不释卷”。

最后补充一句，直接看到最后估计会觉得我说的是废话，所以建议还是全文直接看完哈。

发布于 2019-09-09 18:30

「真诚赞赏，手留余香」

赞赏

还没有人赞赏，快来当第一个赞赏的人吧！

机器学习 深度学习 (Deep Learning)



写评论... | 你关注的 夕小瑶 等 4 个人关注了作者

2 条评论

默认 时间



青羽

看完了叉烧大佬的建议，很感谢啊，准研一，在nlp与cv方向纠结，小白一个，决定入坑nlp。
2019-09-09 · 作者回复了

回复 赞



机智的叉烧 作者

加油加油！！

2019-09-15

回复 赞

文章被以下专栏收录



CS的陋室

微信公众号同名专栏！不定期更新技术文章！

机器学习进阶线路

赞同 8 2 条评论 分享 喜欢 收藏 申请转载



知乎

首发于CS的随室

摘要：本文介绍了机器学习的基础知识以及在企业中的一些应用。通过本文的学习可以快速了解机器学习的基本概念、监督学习以及PAC学习理论，并了解机器学习在实际应用中现状。 演讲嘉宾介绍：...

阿里云栖... 发表于程序员进修...

Python机器学习

机器学习过程

郭志敬

机器学习应该如何入门?

Giant 发表于NLP情报...

一个2010年入坑的后台如何转向做算法和 机器学习的历程和感悟，附录一个书单写在前面：本文主要介绍的是自己的一些经历，以及读过的一些书，实际应用...

贺嘉 发表于腾讯云的1...

写文章

...