

ML&DEV[2] | 机器学习开发技能入门线路

机智的叉烧 北京科技大学 统计学硕士

关注他

10 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2篇文章，本系列下主要谈和机器学习算法工程有关的内容。上回讲到了有关数学知识方面的入门线路，给出了我的一些经验和学习建议，同时给了一些比较合适的材料，这次我来给大家讲一下开发方面的技能以及比较好的线路。

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

上期和大家讲过有关数学方面的线路，错过的同学可以点进去再看看~

机智的叉烧：ML&DEV[1] | 机器学习数学基础入门线路
12 赞同 · 2 评论 · 文章

开发技能学习的目标

机器学习领域作为非诚要学习机器学习甚至深度学习，开发绝对不可能绕开，因此还是要扎扎实实学习。在开始之前，我想分几个方向和大家谈谈要学到什么程度。

- 首先如果你是科研导向，或者你的工作是数据分析、产品数据，则其实只要明确一点——你只需要具备将你的想法翻译为代码能力即可，掌握一门编程语言足够，例如python，掌握的概念是根据需求实现一个算法，会使用常用的包，或者是能再有需要的时候找到你需要的资源套用即可，补充一点，大家可以学学sql，公司的数据分析、产品数据都能写的一手好SQL。
- 如果你的岗位是“算法工程师”，那你的需要的则更高，上面的要求你必须满足，还需要有过硬的计算机开发技能，实质上，“算法工程师”这个岗位和“后端工程师”、“Java工程师”类似，你就是一个开发人员，基本的开发技能都要懂。

将想法翻译为代码

这种能力我将它划分为3块。

首先是能够写一个简单的功能，例如排序，写一个函数，你能把它写出来，一样一行的，足够。

然后，对于比较复杂的功能，你能通过百度找到你需要的代码，拿过来可以套用。

完成上面两点，就算是“会写代码”，你可以拿着这个基础的代码能力完成一些任务了。

工程师的开发技能

作为一名工程师，那在技能上对你的要求就更加高了。从上到下，会的越多越好。

- 编程语言方面，python你应该是非常熟悉的。Java和c++至少会一个，要达到上面提到的程度。
- 数据库基础理论你要会，增删改查，SQL。
- linux操作系统会操作，会写shell脚本。
- hadoop大礼包，hdfs、mapreduce、spark要会一些。
- 能了解一些web后端技术，会写web服务，c++的话例如grpc，Java的spring框架等，python就是flask、tornado、Django等。

学习路线

python

网络资源很多很多，我就给一些我比较好的，又补充欢迎和我说，先强调一个点，只看3就行，2作为补充吧。

- 百度搜廖雪峰，他的官网下有教程，我觉得写的很好，学起来其实也很快。
- 慕课网有关python入门的视频，加一些面向对象的内容。

SQL

SQL是一个非常重要的技能，无论对开发岗，还是其他岗位的同学，只要和数据打交道，这个技能建议大家get起来，这里给大家一些建议吧：

- w3c上肯定是有。
- 我最近转载的一篇文章，是有关hive sql的，在很多数据规模比较大的公司，这个东西学起来可能更加实用：

(超长文)Hive Sql基本操作
mp.weixin.qq.com/s/kvPJGaaAZwMI1Oeo...

Java和c++

- w3c非常经典
- 同推荐慕课网视频

有条件的，弄个虚拟机开始自己操作，你要学会的不是图形操作，而是命令行操作，包括理解里面的一些操作。

- w3c的linux教程
- 《Linux Shell命令行及脚本编程实例详解》
- 慕课网和B站视频均可

模型部署技术

这块学习的要求主要是要学会把模型进行部署起来，让大家都能应用，一般地，学会写一个模型服务，首先是个服务，能有请求和相应，然后能装入模型，用户请求到你的服务，你的服务里有模型，这样就能够跑你的模型了。

具体的教程不提示太多，因为技术比较杂乱，选择也比较多，大家以百度教程为主，学会自己查阅自己喜欢的资料学习，以完成功能为目标。

hadoop大礼包

初级的算法工程不需要太复杂的技能，大家简单了解即可，但是随着逐渐成长，hadoop可能要自己写，然后大数据环境下需要spark，也是逐步开始学习。

学习建议

和大家简单谈一下我的经验。

- 学习线路的内容看起来写的非常简单粗略，不是因为内容本身就简单，而是因为网络上已经有大量教程助你完成学习，毕竟计算机方向的人很喜欢非常喜欢分享，说没有资料的，肯定是没有认真去查找和阅读。
- 站在岸上学不会有游泳，多动手，多看案例，自己动手尝试实践，这点和数学的推导是类似的。
- 解决技术上的问题，自己查>百度，无论是速度上还是效率上，都是如此。
- 学习阶段，自己实现>复制粘贴，实践阶段，复制粘贴>自己实现。
- 有时间多查查文档，每个工具的使用其实都会有非常完善的文档，阅读文档能让你对这个工具有很深的理解，甚至知道有一些隐藏功能，非常建议大家多去阅读。
- 有时间多阅读源码，是指一些工具的源码，虽然工具你可以信手拈来，但是阅读源码能让你对这个工具被回购的原理更理解，这有利于你了解工具的优缺点，在你选择解决方案时能有更多信息。

计算机能力的培训本就是一条很长的路，这是在我当前修下看到的了解到的技能，更多深入的技能有待我进一步修炼，另外各位前辈有什么好的建议，欢迎大家提出，我进行补充~

编辑于 2019-08-17 18:11

「谢谢各位支持！」

赞赏

还没有人赞赏，快来当第一个赞赏的人吧！

机器学习 计算机

赞同 10 3 条评论 分享 喜欢 收藏 申请转载 ...

写评论... | 你关注的 微调 等 4 个人关注了作者

- 3 条评论
- 默认 时间
- 往风

没有数据结构的学习么

2019-09-01 · 作者回复了

回复 赞
- 往风 · 机智的叉烧

期待更新

2019-09-01

回复 赞
- 机智的叉烧 (作者)

非常有道理！我想放在进阶篇的软技能里面~

2019-09-01

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！

推荐阅读

- ML&DEV[4] | 机器学习进阶线路

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
- 初学机器学习必备10大算法
- SUPER MARIO LAND Golden Coins 2 ©1992 Nintendo

用超级玛丽理解机器学习基础
- 走心整理——十个常用深度学

