

名企 AI 面经 100 篇

--揭开三个月薪资翻倍的秘诀

前言

“这 100 篇面经，含金量很大”，这是我看到这份面经合集之后，最想说的第一句话。其次便是触动很大，我个人从 2010 年开始在 CSDN 写博客，坚持了十年，创业则已近五年，经历且看过很多的人和事，但看到这一篇篇透露着面经作者本人的那股努力、那股不服输的劲的面经的时候，则让我倍感励志。

比如“双非渣本三年 100 次面试经历精选：从最初 iOS 前端到转型面机器学习”这篇面经，便让我印象非常深刻。在佩服主人公毅力和意志的同时，也对他愿意分享对众多人有着非常重要参考价值和借鉴意义的成功经验倍感欣慰。

当然，本面经合集的干货远远不止于此，这 100 篇面经分为机器学习、深度学习、CV、NLP、推荐系统、金融风控、计算广告、数据挖掘/数据分析八大方向。分享面经的作者各种背景都有，比如科班或非科班；985、211 或双非院校；研究生或本科甚至大专；学生或在职。至于传统 IT 转型 AI 的就更多了，有从 Java、PHP、C、C++ 等偏后端服务转型的，也有从 Android、iOS、前端等偏客户端开发转型的，当然也有数据分析、大数据方向等转型的，但令人振奋的是，他们都转型成功了，而且他们中的很多人都通过集训营、就业班三个月到半年的学习，成功实现薪资翻倍——这些成功的经验就更值得借鉴了。就业部同事史老师、邢老师特地将这些宝贵的经验整理出来，希望可以帮到更多人。

截止目前，已有超过 1000 人通过集训营、就业班成功拿到心仪的 offer，但路还很长，希望他们能和我司全体同事、全体讲师共同促进技术的传播、人才的培养、产业的发展，共同让全球十亿人受益。

七月在线创始人兼 CEO July

二零一九年

目录

第一部分 机器学习	9
校招面试分享：我是如何连拿华为、苏宁、携程、招商 4 个 offer.....	9
文科生转行面经：我是如何从数据分析到算法并薪资翻倍.....	13
应届海龟小硕跳槽转岗求职经历.....	19
双非研究生学习历程及面试总结.....	21
AI offer 面经薪资近乎翻倍，题库里的 xgboost 笔记看了不下十遍.....	24
机器学习工程师面试心得：知识得系统性学习.....	26
毕业一年后端转型机器学习的面经：四大金刚值得好好用.....	28
一个半桶子的自述：19 应届且双非院校的转型之路.....	30
一个大龄码农的奋斗历程：在焦虑中转型拿到年薪 40 万.....	33
集 4 学员面试经验：经过五轮面试最终拿到年薪 40 万.....	35
普通嵌入式研发转行 AI.....	38
双非渣本三年 100 次面试经历精选：从最初 iOS 前端到转型面机器学习.....	40
双非渣硕，机器学习面经.....	45
机器学习面试心路历程：排序算法和 xgboost 被问的概率很大.....	48
我和七月在线的 5 年时光：从经济学专业到“高级”程序员.....	50
我的面试及就业分享：从后端开发转岗到机器学习算法的历程和感受.....	54
算法面经：应届生如何通过四大金刚拿到年薪 30 万.....	57
集三过后月薪 double.....	60
超励志！本科化学硕士力学，从干过销售只会 VB 到成功转行量化交易.....	61
面试经验，面试会经常问到的问题.....	64

【面经】一个机械转行算法的菜逼应届生如何进华为.....	66
机器学习之面经.....	68
【算法/机器学习】【春招+秋招】个人多厂面经分享.....	71
本科校招算法岗面经：拿到滴滴、头条 offer.....	77
Google 面经，已拿到 offer 哦！	80
18 应届毕业拿到机器学习 offer 的面经：嘉和美康、贷鱼科技.....	89
985 本科生程序员，校招拿到 40 万的 Offer，老员工被倒挂：难受！	94
2019 校招美团京东小米去哪儿华为面经机器学习面经.....	97
集 4 本科应届拿到近 20 万的 AI 面经：曾一度因是双非院校而没面试机会.....	101
集 4 AI 面经：从后端转行 AI，薪资翻了一倍多.....	103
工作 4 年的程序员面试经验.....	105
20 届找实习之阿里，头条，网易，小米，搜狐，商汤，滴滴,百度面经合集.....	108
连拿 4 个 offer 的校招面经-应用开发到算法工程师半年的转型.....	118
第二部分 深度学习	122
C 语言程序员的漫漫二转之路：曾报两个机构的培训班.....	122
算法菜鸟的秋招总结.....	124
深度学习的学习和面试总结：我是如何从客户端转行 AI 的.....	126
深度学习社招面试心得：算法+CV 基础知识+相关项目经历.....	128
学渣抓住时机进入大厂的心路历程.....	132
我是如何通过集训营在公司内部转岗拿到年薪近 100 万的.....	135
非科班生拿 BAT 算法 SP offer 面经（阿里巴巴，腾讯.....）	137
我的算法岗校招面经：微软、谷歌、阿里、头条、地平线、网易游戏等.....	144

我是如何一步步拿下 Google Offer 的?	156
多年 C 语言开发转 CV 算法拿到年薪 40 万的面经.....	164
海康、YY、抖音、CVTE、招银科技、比特大陆（计算机视觉算法或深度学习算法岗）面经.....	173
第三部分 CV 计算机视觉	177
双非应届小硕关于秋招的血与泪：投过很多简历拿到 2 个 CV offer.....	177
双非小白硕士的计算机视觉工程师面试经历.....	184
算法工程师面经：我是如何从产品转到 CV 的.....	186
社招转行 CV 面试经验分享：面试当场秀代码.....	188
社招转行 CV 算法的面经：项目的问题一般这么问.....	191
零基础的“大叔”：如何从仅会点 C 到转行机器学习并拿到 56 万年薪.....	195
从测试到 CV 算法工程师的转型之路：最终连拿 4 个 offer.....	204
大龄社招转行 CV 算法的一些心得：曾整晚整晚训练模型、调参、再训练.....	209
双非二本面经，计算机视觉方向可以来看看.....	211
深度学习算法应届硕士秋招记录,最高拿到 52 万年薪 offer:内含鹅厂头条面试真题及 offer 截图.....	213
第四部分 NLP 自然语言处理	218
我的 NLP 秋招之旅：如何一步步准备并拿到大厂 NLP offer.....	218
双非渣硕应届算法面经：这二三十家公司都面了我这些问题.....	222
从迷茫不会写代码到拿到 34 万的 NLP offer：一定要多写代码.....	226
Android 三年转 AI 狗血经历：曾报一培训机构但没找到工作.....	232
坎坷的学习和面试经历：5 年 Android 成功转行 NLP.....	237
双非本科应届只学过 C 的 NLP 找工作经历.....	241

NLP 面试分享.....	244
NLP 面试总结.....	246
外行工程师转行路：传统 IT 转型 NLP 之如何抉择方向 项目 简历.....	248
NLP 水硕校招面经：SVM、注意力机制、bert 基本必问.....	250
双非渣本的转行之路：从机械设计到 NLP 并薪资翻倍.....	253
我的 NLP 面试经历及总结：要把 tfidf、word2vec、注意力机制、transformer 都熟悉掌握.....	255
我是如何通过 3 个月的学习，从运维开发转 NLP 的.....	258
NLP 常见面试问题总结：Word2Vec、LSTM、BERT 等.....	261
985 小本 3 年后台转算法.....	262
渣硕产品转 NLP 实录：从只会 matlab 到 NLP 年薪 30 万.....	265
我如何从传统 IT 转型 NLP 并薪资翻倍的：NLP(知识图谱)面试总结.....	268
跨专业转行 AI 的艰辛历程与 NLP 面经.....	272
一个机器学习集 4 学员的工作感悟.....	276
跨专业转 NLP 的扎心历程：试用期被辞再找拿到更高 offer.....	278
传统 IT 企业转岗 NLP.....	281
NLP 开发的体会.....	284
来自菜鸟程序员血淋淋的教训：从最初各种被拒到最新年薪近 60 万.....	286
双非小白找工作教训.....	289
留学生数据分析转 NLP 的面经：seq2seq、transformer、bert 基本必考.....	291
今年转行从机械到算法.....	295
励志！充满干货的 AI 面经：纯电力员工如何成功转行 NLP 并薪资翻倍.....	298

NLP 岗位面试经历：很多面试题都能在题库里找到.....	305
AI 面经：我是如何从机械行业本科社招成功转行 NLP 并薪资翻倍的.....	307
刚毕业双非本科如何拿到 NLP offer 并薪资翻三倍的.....	310
集 4 薪资翻倍的 AI 面经：曾担心学完集训营依然找不到工作.....	312
ML 集训营 3 学员专访 2：我是如何从 iOS 成功转岗 NLP 拿到 45 万的.....	314
第五部分 推荐系统/计算广告	316
推荐算法工程师面试心得分享：课程进行过半拿到 offer 并薪资翻 1.5 倍.....	316
推荐算法面经 最重要的是必胜的信念：普通大学没经验但有开发能力就能转行成功.....	318
水硕跨专业：拿到自己满意的推荐系统方向 offer.....	320
大数据转行算法工程师面试经验.....	322
双非渣本跨专业转行面经：从 Linux 运维到推荐算法 40 万.....	323
集训营四期面经.....	327
推荐算法面试分享：如何从大数据起步拿到年薪近 40 万.....	331
算法工程师（推荐方向）面试经历：简历、数据结构、机器学习、推荐系统.....	333
第六部分 金融风控	335
经济学编程小白转数据分析.....	335
集四 45 万 offer 的 AI 面经：SVM 和 XGBoost 这类是面试常考算法.....	338
金融风控 100 道面试题：传统银行开发转行互金 top3 公司并年薪 40 多万.....	341
普通 IT3 年转行经验谈：从曾质疑培训机构到培训后年薪 35 万.....	348
某贷网、某蛋科技、百度金融、蚂蚁金服等互联网金融面试经验.....	350
一位从传统 IT 跳槽到互联网做数据分析师的面经.....	363
第七部分 数据挖掘	365

非科班从传统行业到 JAVA、爬虫到转型数据挖掘.....	365
普普通通的转行之路.....	367
面试分享，从开发转算法.....	369
双非硕士面试之旅：数据分析、特征工程、模型构建、模型优化等流程要熟练.....	371

第一部分 机器学习

校招面试分享：我是如何连拿华为、苏宁、携程、招商 4 个 offer

我是机器学习集训营第六期的学员，参加集训营的时候是一名正在参加秋招的研三准毕业生，机器学习方面的项目经验很少，秋招一开始阶段并不顺利。

因为我起初接触机器学习的时候就看过七月的一些 nlp 方面的小课，讲的很简单易懂。为了找个好工作所以就咬咬牙保留集六，如果当时有 nlp 的集训营 我就直接报这个了。学习的时间有四个多月，我一边参加秋招一边学习巩固知识，优化简历。

整个找工作结束，拿到了好几个 offer，其实四个月我学的还是不够的，我现在也在不断的看一些以前上课的视频，继续充实自己。下面介绍一下我面试时的情况，也给大家一些经验。因为我是校招，所以下面这些面试经验校招生可以好好看看。希望能对你们找工作有好的用处。

首先，集训营老师给的意见是自己确定好一两个方面，最好定两个简历模板，然后不断优化。譬如我当时就一心想去搞 nlp，所以就只做了一份简历。大家不要和我学，哈哈。首先不管我们是要走 cv、nlp、推荐什么方向的，面试的时候都是基于简历上我们自己写的。不过既然都选择了 NLP 或 CV 等，我们都必须熟悉且能手推出 LR、adaboost、xgboost、svm 这些经典的机器学习算法，这是最基础的了。下面我就介绍下我面试公司的一些情况

华为：笔试完 总共两面 获得 offer

一面较轻松，因为面试的是一个脾气很好的小哥。上来就问我平时都使用什么语言，然后用最熟悉的语言先练练手，写一个快排，面了很多家感觉手撕代码大部分都会先来个快排。然后看着我的简历开始一个个问：

最熟悉那个机器学习模型，xgboost 多分类 参数怎么设置 数据不平衡的时候 xgboost 可以直接训练吗？会不会有影响。讲一下 xgboost 的推导，大概讲讲就好。

然后开始看项目，忘了介绍，我简历一共三个项目，一个是集六中的 ai 法官项目，一个是京东的潜在用户预测，一个是我自己的毕设论文项目。

首先面试官对 ai 法官比较感兴趣，让我说一下模型的整体框架，对项目细节问的比较多，比如为什么要选择用 cnn 来作文本，为什么不选用 RNN 一类的来做。然后询问 word2vec 和 glove 的区别，模型输入是只有 word2vec 吗？用别人训练好的 word2vec 来训练你的模型有没有什么影响？有没有考虑用 RNN 的模型来做，怎么做？

二面的时候应该是部门的老大，整体问的问题都比较前沿，问了了解过 transformer 吗？

你觉得 bert 的优点是什么？说一说 transformer、rnn、cnn 这些模型的优缺点，比较一下这三种特征抽取器。

整体下来感觉基础知识要牢靠，面试之中的问题都是从我的项目出发，考察我的解决问题的一些思路 and 想法。

苏宁 总共一面 获得 offer

不知道为什么，今年校招只有一面。俩面试官同时面试我。坐下来时就很慌。

一位一直询问机器学习的问题：

简要介绍一下你做一个机器学习问题的完整流程。

如何判断过拟合，模型过拟合了怎么解决。

LR 和 SVM 的区别

介绍一下随机森林

为什么随机森林不需要做特征选择

svm 的核函数都用过哪些

rnn 的优缺点

cnn 的优缺点

cnn 怎么做文本分类，优势是什么，缺点是什么

LSTM 改进了 rnn 的什么问题

也记不清了 反正问的很多，感觉他有个面试题库 挨个问我。

另外一位是做 nlp，开了我的项目就问了一个问题，命名实体识别中，你如果一开始分词分错了怎么办，这样后面的实体识别不就蹦了吗？ 你是怎么考虑的？

果然一对二不占优势。不过问的问题除了最后一个其它都是比较基础的，比较好回答。

搜狗 加上 hr 面 一共两面

很神奇 苏宁和搜狗都是只有一面技术面。

直接说问题吧

因为面试官是做搜索方面的，所以上来就让写个代码。写一个最大字符串匹配的问题。

然后开始问：

word2vec 是怎么训练的，negative sampling 解释一下

tfidf 的原理

lstm 解决了 rnn 什么问题？

cnn 在文本中的用法，pooling 的作用

深层网络容易过拟合还是浅层网络容易过拟合？

注意力机制的理解，从公式出发

relu 和 sigmoid 有什么区别

xgboost 里面的 lambdarank 的损失函数是什么

xgboost 在什么地方做剪枝

lightbm 和 xgboost 有什么区别? loss 方面

问了很多问题, 记得的也就这么多了, 隔了好久了。

整个秋招到最后 又拿到了携程、招商银行网络科技的 offer

总结:

首先传统机器学习算法: svm、lr、softmax、决策树、随机森林、GBDT、xgboost、adaboost、bp 神经网络、朴素贝叶斯等等这些都必须自己手推一次或者多次。深度学习方面: CNN、RNN、LSTM、常用激活函数 (tanh、relu 等)、adam 优化函数、梯度消失这些原理或者是结构都应该能手画出来。

nlp 方面: 强烈建议 tfidf、textrank、word2vec、注意力机制、transformer 都应该非常熟悉。源码看不看随意, 但是必须自己去运行实践一边, 因为不这样去搞, 好多处理我们都不知道是为什么, 而面试最爱问的就是为什么这样做。最后祝大家找到心仪的工作, 前程似锦。

文科生转行面经：我是如何从数据分析到算法并薪资翻倍

一、个人背景

本人文科生一枚，上学时就对数据分析比较感兴趣，所以毕业论文导师选的也是偏方法偏统计方向的，并且自学了 SAS。毕业之后，在一家小公司里面做数据分析工程师，自学了 Python。公司主营广告业务，所以工作期间主要接触的都是互联网广告相关的业务数据，一半的时间是在协助运营取数。一半的时间在做建模相关、用户画像相关的工作。2019 年秋招找工作时，工作刚满 3 年。

二、为什么选择七月的集训营

很早就知道七月开始有集训营的课程，但是犹豫了比较久。在这半年期间也尝试过其他的在线的学习平台，比如说优达学城上的自动驾驶入门课程和 Coursera 上的吴恩达的 ML 的课程。我个人还是不太喜欢这种纯在线学习的方式，因为在线课程如果没有按照班期的课程进度跟上学习，很容易就落下了。对我个人来讲，我觉得线下集训营的方式比自己通过在线教育课程来学效率高很多。

第七期课上的老师都是有着非常多的行业工作经验和面试经验的，对于实际业务和面试上知识的重点、改简历的技巧和面试等的情况都非常了解，对于在职转行面试找工作还是非常有帮助的。这些肯定是靠自己学习，不可能得到的信息。

另外就是集训营每个知识模块都配套有相应的练习题，如果是自己，可能就看过一遍就完了，掌握的并不扎实。但是在集训营里大家一起学，按时把每周的作业题做完，对于知识的理解和巩固是非常有帮助的。而且，作业中涉及的内容与实际工作中需要的技能都非常相关，把所有作业都自己实现一遍，并且做好相应的笔记，这样进步会比较快。实际上，笔记中总结的内容，可能就是未来面试时候会遇到的面试题，如果自己实现一遍并且整理好笔记，对于面试也是非常有帮助的。

三、报名之前的准备

在最开始知道七月开始有集训营之后，就看过官网上面宣传的各种资料。比如，零基础也可以报名参加之类的说法。个人认为，在学习之前还是应该先熟悉一下相关的知识，这样学起来不会那么痛苦，毕竟是在工作的人，下班后的时间比较有限。在开始学习之前，自己能够先对这机器学习有个了解，再加上集训营的学习，效率更高，所以我在真正报名之前，我是集 7 的学员，自己会先把先关的课程提前听一部分，加上研究生期间的统计学基础和自学的线性代数的基础，工作上也会使用逻辑回归进行点击率预估的工作，其实还好。有一些基本的认识和编程基础就好，不需要准备太久。还是需要考虑招聘市场的情况，现在越来越多人转行算法，还是要抓紧时间学习，积累项目经验，尽早入坑。

四、金九银十面试

1. 面试时间的选择

我是集训营第七期的学员，正好是在 2019 年的 4 月初完成了整个集训营的学习，刚好赶上了 2018 年底开始的互联网寒冬。当时的就业市场情况比较一般，我就没有选择在这个时间段找工作，而是想等到秋招再找工作。运气比较好，2019 年 9-10 月的就业市场明显比春招的时候好了很多。

让我比较后悔的一点是，即使没有计划在 3-4 月就找工作，但是也没有投简历去练习一下算法岗的面试。如果春招能去面试一下，秋招的时候会更加自信和从容一些。所以还是建议大家，不要等到真正计划要去找工作的时候才开始去投简历面试，最好还是提早准备好简历，开始投简历去面试，因为对于没有面试过算法岗位的同学，还是要去适应一下这个过程，并且去组织语言与面试官去沟通的。

2. 准备简历

在简历的准备上，我写了两个版本的简历，一个是偏向数据分析的，就是偏向业务的，一个是偏向算法的，因为我两个方向都想了解一下，最后看薪资和岗位的成长价值来决定的。对于社招的同学，不管是分析师还是算法工程师岗位，最重要的还是实际工作中的项目经历。在集训营的学习过程中，大家需要去完成一些项目，可以根据岗位需要进行修改，写到自己的简历里面去。当然，最好是和现在自己工作更相关一些的项目。集训营也会提供简历修改的服务，针对自己的就业方向，可以自行找相应方向的老师讨论并修改简历。课堂上，老师也会讲一些关于简历修改的内容，也是很有帮助的。

3. 简历投递

因为我没有计划海投简历，所以有选择性的集中投递简历，将得到的面试机会的面试时间都集中安排在了一周内进行，如果有各方面都比较合适的就准备入职了。

我在今年的 8 月底 9 月初开始集中投简历，很快就有面试的机会，分析师岗和算法岗两个方向上都拿到了面试的机会，最终还是选择了算法岗。

我在 boss 直聘、拉勾和脉脉上都投递了简历。我个人的经验是：

- 脉脉的回复速度是最快的，一般 2 天之内就都有回复了，其次是 boss 直聘，拉勾的投递效率和回复速度都非常慢。
- 主动邀请你投递简历获得面试机会的可能性更高，宜信、滴滴和现在这家公司都是 hr 或者部门负责人主动邀请我投递的。

4. 面试情况

很幸运，第一个得到面试机会的公司就拿到了算法岗的 offer，各方面也都比较合适，在确定薪资等问题之后的 2 周就入职了。总体来说面试经历比较顺利，但是等待面试结果的过程还是比较痛苦的。

1) 分析师岗

偏向业务的拿到了两个面试机会,一个是宜信的商业数据分析师,一个是滴滴的数据分析师。

宜信

宜信的面试非常顺利就过了,跟部门的两个面试官都聊得非常好,拿到了 offer。其实,宜信的企业文化非常好,面试的部门也不错,是核心业务部门,因为没有什么社招的经验,在谈薪资的策略上没有把握好,对于薪资不太满意,就没有去。

滴滴

滴滴的面试准备,主要就是面试前一天,在网上搜了一些滴滴数据科学部门相关的面试题,重点准备了一下 SQL 就去了。

滴滴的数据科学部门的面试挂了,但是面试还是很有收获。让我了解到了数据科学方向上的工作的特点,还有就是滴滴的面试题,非常偏向业务,4 道完全基于滴滴业务场景的题目,即在不同场景下的司机侧和乘客侧的运营策略的问题。如果以前对于业务没有了解(不管是互联网中的哪个细分行业的,因为业务知识有些事可以迁移的),或者没有迁移的业务分析能力,基本不太可能面过,因为面试官就坐在你对面看着你答题,答题时间还不能太长。面试数据科学相关大部门下面的岗位,对于统计学和概率相关的知识要求比较高,在面试题里面有 4 道是关于这个方向的,剩下的一道就是 SQL 题了。这些只是笔试题,面试交流过程中,还针对实际的业务场景,要求设计一个 A/B test,去验证问题。我在这个问题上答的不太好。

2) 算法岗

算法岗,我拿到了现在入职的这家公司和腾讯微视部门的面试机会,腾讯约了面试的时候,但是部门的人并没有电话沟通,因为已经确定入职现在这家公司了,我就没有去和当时来电话的 hr 在沟通。至少这是一种认可,帮助我增加了很多的信心。

面试准备

1. 我在网上找了一些面经和机器学习算法岗面试题的总结，在集训营上边学习边刻意的根据这些面试题去总结知识点，以自己的方式进行表达。
2. 集训营算法知识类的课程讲完之后，我开始每天利用上下班地铁通勤各一小时的时间，开始刷七月在线 APP 上面试大题中的机器学习的题目，到 4 月中下旬把 140 多道题目从头到尾都认真的过了一遍。先过考的比较多的题目，再过比较细节的题目。
3. 面试之前的几天，集中把《百面机器学习》这本书和面试中常考的几个算法和一些知识点，全部刷了一遍，做到能够手推公式（实际上，在集训营上课期间，我已经将书中面试相关的重点内容，都已经快速过完一遍了）。逻辑回归是面试必备的，树模型和集成学习也要熟悉。

面试过程

面试现在这家公司岗位的过程中的问题并不太难，只是我没有面过算法岗，自己最开始比较紧张，但是后面聊起来了就放松多了。面试的问题主要是围绕着简历写的项目进行的（面试题目只写了现在还记得的）。

- 简单介绍一下自己做过的项目
- 广告点击率预估：特征挖掘思路，用了有哪些维度上的特征，预测的准确率如何，比上一版本的模型预测准确率提高了多少，模型优化的思路，手推逻辑回归
- 文本多分类项目：Xgboost 与 GBDT 的差异，为什么 Xgboost 训练速度更快，集成学习有哪几种方式，textCNN 模型预测的效果如何，word2vec 词向量的长度设置为多少维，word2vec 为什么选择这个数量的维度进行数据处理
- 使用 spark 进行数据处理，数据量特别大的时候，spark 会比较慢，这种情况你是怎么处理的？
- 手写二分查找

- 写 SQL

- 知识图谱项目：说下项目的基本情况，如何对数据进行处理的，结果如何

五、关于文科生转行的建议

1. 如果有转行算法岗的想法，首先应该花一些时间对算法岗在实际工作中的工作内容和日常进行一些了解，确认自己是否适合，是否符合自己的预期，是否认可这个岗位的价值，或者你只关注薪水，那就立刻开始行动吧！不要把大量时间总是花在思考“我能不能转行”这种事情上。

2. 确定要转行后，就要抓紧开始行动。开始了解机器学习，对机器学习有一个宏观的认识。开始看七月在线上的机器学习相关的视频课程，如果曾经没有接触过编程，每天下班后都要花时间，开始学习 Python 入门的课程，动手开始写代码了。

3. 最重要的一点是要相信自己，对自己有信心。

4. 面试找工作阶段，被面试官怼是比较正常的事情，不要气馁，多积累面试经验，并且把相应方向上的知识基础不断夯实。

六、总结

1. SQL 基本为必考项目，我的 3 个面试都考了。如果对于 SQL 掌握的不好，面试前需要专门进行准备，一般会考到开窗函数。

2. 手推公式环节上，逻辑回归是必备，其他重点算法都要掌握。

3. 如果不是面试 AI 相关的岗位（NLP 或者 CV 之类的），机器学习中的 Xgboost 已经是面试必备的，基础知识要扎实。

4. 谈薪资要有点策略。

P.S. 这个面经，如果不是老师一直催，我想我应该是写不出来的，拖了一周多才写完。希望对其他学员有帮助。

应届海龟小硕跳槽转岗求职经历

个人背景

本科毕业于 985 高校，电子信息工程专业，绩点不高，也没有什么特别的经历。硕士毕业于法国一所工程师学校，方向是 robotics。念了两年硕士毕业之后，就回国开启了找工作的慢慢长征。

毕业回国后的秋招连连碰壁，再加上当时回来的时候已经比较晚了，稍微好一点的公司都已经结束了校招的秋招工作。后来一直到几个月后的春招，才在上海找到一份软件工程师的工作，公司主要是做医疗试剂和医疗器械的，规模比较小，总人数大概一百多人。我的主要岗位职责是给全自动的医疗器械写底层的软件控制程序。在这家公司工作了有将近一年，当时还是觉得公司所提供的平台较小，个人发展有些受限，跳槽的想法一直萦绕在脑海。也是偶然的机会在微博上看到七月在线集训营的报名，加之之前学 robotics 的时候也自己学过一些机器学习相关的课程，就报名了课程想再多学一些，为未来的换工作做准备。后来还挺幸运的吧，大概在课程进行到一半左右的时候，就收到了一直心仪的公司的 offer，一家 500 强外企，岗位还是软件工程师，所在组里面主要是给芯片做人脸识别还有音频等一系列功能的集成。最终确定的 offer 年薪也由之前的 15w，上升到了 20 多 w。

面试经历和个人总结

其实最终拿到 offer 的这家公司，在之前的校招中就有投递过两次简历。第一次是石沉大海，第二次收到了笔试通知，但是笔试完就没有消息了。一直到后来第三次投递的时候，才被我现在的 manager 从简历池里打捞起来，给了后续的面试机会。当时进行了两轮面试，第一轮是全英文的沟通，主要包括自我介绍，想要换工作的原因，以及对未来发展的打算，是 manager 直面的，也询问了几个和 Linux 相关的问题，总体涉及技术的地方不多。

第二轮面试是中文的技术面，面试官是另一个比较资深的工程师，主要是针对我简历中提到的一些项目和所使用的算法进行了提问。例如，我的毕业设计中，采用了机器学习中的 k-means 算法对从 sensor 所获得的 data 进行了分类，所以面试官详细询问了这个算法的执行过程，以及和其他几类主要聚类算法的异同，为什么要使用这种方法。包括之后，对于一些误差偏差数据的处理等。另外，由于我之前的一份工作是做 MCU 的芯片编程，也询问了一些关于 SPI、I2C 通信协议的机制。给我的感觉是，只要是简历上出现的项目或者算法，一定要吃透，切勿不懂装懂。

虽然之后的工作还是会和芯片打交道，但机器学习或者说 AI 的知识，对我来说，真的是非常好的加分项。因为很多做半导体芯片的企业，也都在打造一些 AI 方面的突破，进行整合。所以感觉自己还是很幸运吧，不过也很感激自己一直坚持投简历，并且在为这个目标做着准备。我想，在求职的道路上，还是不能放弃，一直努力，坚持下去，可能机会就在下一路口，等着你。

后续

最后还是非常感谢七月在线，在线下的学习中也认识了一些可以一起交流的小伙伴，上海这边线下辅导的老师人也很 nice，有问必答，很多线上的课程，之后需要的时候，也还会再拿起来看一看，是个很不错的持续学习平台。

双非研究生学习历程及面试总结

个人学习经历

本人本科是通信工程专业，本科毕设做的是基于支持向量机的时间序列预测，研究生是信息与通信工程专业，在 2017 年 9 月选导师的时候了解到导师新开了机器学习方向，本人也想往这方面发展，于是选了本人的导师。

刚开始的时候就在网上搜机器学习的资源，比较有名的书籍有李航的统计学习方法（当时还没有出第二版）、周志华的机器学习、机器学习实战、花书等，视频有吴恩达的机器学习、李宏毅机器学习，进阶的还有斯坦福 CS231、CS224 等课程。于是研一期间一边上课一边自学，主要是理论方面看统计学习方法和西瓜书（个人以统计学习方法为主，遇到看不懂的就再去看西瓜书，个人建议是两本书结合着看，相互补充），等到 11 月大概过完一遍了，就开始看吴恩达的视频课（本人看吴恩达的主要是因为吴恩达的课太火了，后面也看了一些李宏毅的课程，也很不错）。吴恩达的课程会涉及深度学习，所以在这个过程中又穿插着看花书，主要还是以视频为主。这个过程持续的比较久，到 18 年 5 月份的时候参加了天池的一个比赛，虽然没有取得很好的成绩，但是开拓了视野，明白了比赛和项目的重要性。参加完比赛就在做实验室的项目了，也是使用 Python 做数据分析，用了 LightGBM 做预测，一直持续到九月。

这时候自己感觉对于基本的概念知识点都了解，但是缺乏系统性，也缺乏项目和实战经验，一个人学习的话也很难抓住重点，就考虑系统的学习一下。此时经过比较了好几家类似的机构，最终选择了七月在线这个平台，加入了集六的大家庭。

集训营开课之前会有一段时间来学习预习课程，主要有数据结构、数学、Python 和机器学习，我由于时间有限只看了 Python 和机器学习的部分内容（时间充足的话一定要提前看完，好好掌握，不然开课之后就会出现跟不上的情况）。

开课之后三个月的时间是每天都有学习计划，我周内要做实验室的项目，只能晚上和周末进行学习，导致有些地方学得不扎实（建议一定要多看几遍视频，好好自己整理复习）。课程中给的资料都很棒，省了自己很多时间，而且还有一些很棒的案例，自己消化一下就可以拿来用（小抄全集是真的好用）。学习结束之后会要求组队做比赛或者项目，这是一个很好的锻炼自己的机会。学习完到 2019 年了，自己感觉还有好多内容没有消化，需要自己整理沉淀。6 月份之前在忙着写论文，直到 6 月份才有时间复习整理，投了一些暑期实习的岗位，从七月份开始到九月份也实习了三个月时间。期间也投递了一些秋招提前批和正式批，参加了一些笔试和面试。

面试总结

某中小公司（自然语言处理算法工程师）：

一面：用过哪些中文分词工具（jieba），原理是什么（HMM），为什么使用它而不是别的（简单好用开源），遇到的问题（准确率一般）。

TF-IDF 算法（计算公式，为什么加 IDF 项，为什么取对数，为什么加 1），好处（计算速度快），存在的缺陷（静态，依赖于语料库）。

了解 PageRank 算法吗（了解，但是具体公式忘了，说了一下思想），如何将 PageRank 算法用于文本（其实就是 textrank，说了一下不一样的地方）。

场景题：目前有一批文本内容，每一条包括题目、关键词和正文三部分，如何提取关键词？

（我的回答是：从题目中提取名词，从正文中使用 tf-idf 提取 top-k 的关键词，和原本关键词一起去重后作为召回的关键词，可以通过给标题、关键词和正文赋予不同的权重来对关键词做重要性排序）

二面：聊实验室做的项目，数据量有多少，数据来源，数据缺失值处理方式，异常值处理方式，使用的评价指标有哪些。

为什么使用 lightgbm, 效果怎么样, 调节了哪些超参数, 怎么调节的, 如何判断有没有过拟合。有没有和其他模型 (如 SVM) 进行对比, 对比结果如何, 为什么会出现这种结果。有没有用过 xgboost, lightgbm 与 xgboost 的异同点, , GBDT 的原理, 和 RF 的异同点, 三种决策树算法的异同 (分裂规则, 适用问题)

大型互联网公司

一面: 聊做的实验室项目, 先阐述项目要解决的问题, 再说明设计的解决方案, 方案的设计思路, 创新点等。

算法题 1: 类似于两数之和, 不过给定数组中的元素允许重复使用, 而且不限定使用的元素个数。比如给定数组为[2,3,5], 目标值为 8, 要求输出为[[2,2,2,2],[2,3,3],[3,5]]。可以使用递归来解决, 我并没有写出来, 讲了下思路。

算法题 2: 给定一棵二叉树, 返回从二叉树右边看到的结点。

二面: 牛顿迭代法求一个整数的平方根, 爬楼梯问题, 求斐波那契数列的第 n 项 (要求时间和空间复杂度为 $O(1)$)。共轭分布的定义、二项分布和多项分布的关系。

总结

数据结构与算法: leetcode 剑指 offer

机器学习理论知识: 统计学习方法/西瓜书 百面机器学习

机器学习实战: 机器学习实战: 基于 sklearn 和 tensorflow

深度学习理论: 花书

AI offer 面经薪资近乎翻倍，题库里的 xgboost 笔记看了不下十遍

本人是机器学习集训营五期的学员，vincent，本科二本，硕士双非，之前做的数据分析都没怎么用上模型。17 年毕业，一开始做的工作不是机器学习，后来 18 年年中换了一份工作，也开始报七月在线课程。

七月在线的各个老师的课程让算法显得通俗易懂。下半年边学习课程边在一家外包公司工作，帮民生银行做数据挖掘项目，期间用到 pandas, numpy, rf, xgboost 去处理项目数据建模。用的是包，而算法过程是从十月份开始看。从推导 lr 开始，再到 svm, adaboost, xgboost，集成算法一开始看都是一脸懵逼，网上看很多资料，看了火大，抄来抄去，看不懂。

首选校长的博客，看 xgboost 那篇博客我看了不下十遍，adaboost 的过程看了也有好几遍，直到确定自己懂了并且能写出来，推导损失函数。另外，对 cnn 的文章，校长的博客也是首选。

(小编注：校长 July 的 SVM、adaboost、xgboost、CNN 等笔记都能在七月在线的题库里搜索到，题库也是校长一直提倡的四大金刚之一：课程 题库 OJ 竞赛)

到了面试公司了，寒老师给了我简历上的建议，也认真思考了一下。今年面了几家公司，前后花了一个多月找的工作。算法内容从自己懂再到讲述是不同的过程，第一次面试我连 xgboost 的损失函数都没写出来，直接就被拜拜了。后面就一直看校长的博客，反复看，看集成算法 bagging 和 boosting，再看 cnn。

后续参加几家公司的面试。主要问的是二分类的评估指标，svm 核函数的作用，集成算法的过程，介绍决策树再到随机森林，包括决策树的信息熵信息增益和增益率、基尼系数，

各个指标选取最大还是最小；xgboost 的整个过程；关于建模特征选取的 rfm 模型，faster rcnn 与 rcnn 的区别，还有数据结构的排序。

最后我选择一家做医药机器学习的公司，这家公司也有做 cv 的产品，工资是近乎翻倍，挺满意的。后续沿着机器学习加深度学习两方面的应用发展。再次感谢七月在线，感谢寒老师，特别感谢校长的博客，祝福七月在线的朋友都能找到好工作

机器学习工程师面试心得：知识得系统性学习

我是去年自学过几个月，感觉学得懵懵懂懂的。面过几个公司，效果都不好，所以萌生了报班的想法。正好遇到七月招生，就报了。我是集 7 的学员。经过系统的学习，之前自学时候不懂的东西，很快就明白了。面试的时候，基本都能回答上来。几个公司的技术面试都通过了。最后接的 Offer 是一个智能制造的机器学习工程师。

总的来讲，基础知识要理解清楚，热门和前沿的技术能掌握更好，不能掌握起码也要了解，面试的时候提到不要蒙。常见的问题都差不多，只要基础知识掌握好就能对付 80%。

基础知识：

常见算法的推导和特点、常见的特征提取方法、处理过拟合的方法、模型评估、模型集成、常见的网络结构、优化方法、梯度消失和爆炸都必须熟悉。这些都是课上讲到过的，只要多温习几遍也就熟悉了。如果有难点确实没理解的，可以拿时间专门攻克，这样可以增强信心。也可以直接请教别人。

项目方面：

我写的是一个 NLP 项目和一个 CTR 预估的项目。

CTR 预估是用 deep-cross 网络。

NLP 项目现在的标配是 Bert 的 fine-tune，但也不能只会 bert 微调。基本的 CRF、

LSTM、CNN 都要掌握。NLP 主要就是表示、标注和分类。

项目经验少没关系，只要有一两个就可以。因为 AI 总共才火几年，你又不是专门研究的博士，也不可能多。重要的是理解算法原理和思想，能够将现实问题转化为机器学习的问题，不是编程经验。

注意知识的系统性和自己的优势

面试的时候，抓住自己熟悉的问题一定要说透，把相关的都说。这样也能把面试官的注意力吸引过来。同时占用了时间，被问到不熟悉领域的机会就少了。

比如问到 xgboost，就从决策树（熵）到 xgboost 都说，同时还要说 boost 集成的方法，还可以扩展到其他的集成方法（bagging, stacking），还可以对比各种集成的特点，甚至可以扩展到神经网络，因为神经网络也可以看作是一种集成。

比如问到词向量或嵌入，就从 w2v, glove, elmo, 到 bert 都说一遍。如果时间充裕，可以研究各种表示方法，很有意思的。

比如问到关键词提取，就把 tf-idf, textrank, lda, 都说一遍。如果还知道其他的关键词提取方法就更好。

总之，要让面试官看到你是系统的理解和掌握了，不是零碎的知识。

毕业一年后端转型机器学习的面经：四大金刚值得好好用

本人背景介绍，我是报的机器学习集训营第七期，18年硕士毕业，毕业校招签约了国企后端开发。之前没做过机器学习算法项目，在读研期间也是从事后端开发，所以，你们想转型算法的，只要有决心跟着七月老师的课程走就没问题。当然，大家最关心的薪资问题嘛，也是翻了不少，不过也有工作城市原因。

在学习过程中，前期是下班后回家学习，在学习过程中感觉力不从心，后期就辞职了，破釜沉舟。这得根据每个人的情况而定，每个人有每个人的学习方式。课程安排是递进形式，前两周学习 python 和一些 numpy、pandas 常用的库。然后学习将数据图形化的一些方法，最后学习机器学习算法原理，结课是自己选好方向做项目，包括 cv、nlp 和推荐。这一套学完后你会对整个机器学习算法和项目有很深的了解，这也是很重要的，因为面试时基本是从项目中穿插问一些算法，所以这两部分得好好学习。

面试：前前后后也拿到好几个 offer，有电商互联网、在线教育（C 轮）和在线租车等上市公司，国内最大的安全公司最后领导面了，结果告诉我 hc 被锁了，哈哈。

先说说电商互联网，因为是异地，所以一面是电话面，首先是自我介绍，基本都这样，我现在不过脑子都能说出来了。接着是问你做的项目，我因为是做的推荐，问了数据有什么，多大，推荐系统怎么召回怎么排序，这个项目的目的是干嘛，word2vec 原理，因为我用到了 DNN。又问我在项目中怎么用的，以及原理，最后还问我熟悉其他算法吗，给他介绍了一两个。

二面得去现场面试，先做了份试卷，在这里也说明一下，一般公司有笔试的不要怕，都挺简单的，我面过一个公安部，把 svm 给推导出来了，面试官惊讶问我是不是用手机了。回归主题，二面面试官也还是从项目中着手问，在过程中手画了 LSTM 结构图（七月在线

的题库中有)，把项目说完后，问我 hadoop 和 spark 熟不熟，能不能从日志中找出想要的信息，接着问我怎么刻画用户画像，最后问为什么想换地方来这工作等一些个人问题。

对于一些常见的算法 LR、SVM 基本上必问，XGB 和 GBDT 区别等等。简历中写的东西都得会，项目得很熟悉。最后，面试大厂时数据结构也很重要，我面 bat 中某一家时，前几面一上来就是写代码，然后再聊其他的（小编注：校长 July 经常强调四大金刚是有道理的：课程 题库 OJ 竞赛）。

面试的一些体会：一定要多面，不要害怕，从面试中发现自己的不足。学习算法或者转到这一行，可能前期很辛苦，但是如果坚持下去肯定会不一样，算法的天花板我觉得也高一些。自己认定的路就坚持走下去，这也是对我自己说的。最后，希望大家都在七月好好学习，找到满意的工作。谢谢。

一个半桶子的自述：19 应届且双非院校的转型之路

我是七月在线机器学习训练营第 6 期的学员，也是 2019 届应届毕业，一个双非院校普通通的研究生，我大学学的是通信专业，那个时候我还从没听说过机器学习这个东西。研究生进的是软件学院，学的是物联网技术这个专业，这个好像和机器学习也没有很大的关联。

我第一次接触机器学习是在研二，因为当时有一些工作涉及到了一些算法模型且毕业论文也需要用到这方面的东西，所以开始了机器学习的学习之旅。随着学习的深入渐渐感觉到很吃力，自学机器学习果然还是很难得，那些理论公式推导看的人头晕。当时一个人在那里慢慢啃，也没人教，这样过了一段时间，跑去向老板诉苦，然后老板大手一挥，753G 的学习视频丢给我了，当时就懵了，这要我怎么学。

最后把视频大致看了一下，发现基本上是七月在线的这几年的机器学习课视频（也不知道他在哪弄的），这就是我第一次和七月的邂逅，看了一些视频讲解，果然有人指导就是不一样，一些以前看的迷迷糊糊的东西老师一讲解就豁然开朗了。经过了一段时间，我发现我对这个还蛮感兴趣的，于是就打定主意，以后就做人工智能了，物联网就被我果断放弃了。既然打定主意了就要行动起来，由于周围没人做这方面的，而且能系统的学习总好过一个人在那里毫无方向的看视频，最后我就报了机器学习集训营。

当时那段时间确实很忙，一方面要完成老板布置的任务，几篇小论文要写，而且大论文也要做各种实验，还要采集数据等相应的准备工作，还要看论文，所以一般只能到晚上跟着直播学习，课下也不能花很多时间巩固，最主要的由于当时不在北上广等大城市，无法参加七月的线下课程（这个是最亏的），所以好多实训项目都没能到现场去实现，每次只能在群里下载代码自己慢慢看，由于我的专业原因，以前没怎么接触过编程，所以代码方面不太行，所以课程的案例代码我都会对着敲一遍，当然也总会遇到很多问题，对一段代码我会反反复复的看，每一条理解清楚，就像背书一样将它们背下来，现在想想其实也挺好，既然没有很

多时间来慢慢来，那就先“填鸭”吧。当初为什么会选择学习机器学习呢，一方面是论文需要，另一方面是我一直对数学很感兴趣，虽然有时候那些数学公式让我头疼，但是乐此不疲，可能这也是我坚持学习的一个动力吧。

学校的生活总是平淡无奇的，就这样在写论文、看视频、写代码的生活中度过了几个月，迎来了春招，很多公司来学校招人，投了很多家，基本都会有回应（对于校招，大多数公司都会给你机会的），首先去做个笔试，笔试过了，然后再面试。我就挑两个我印象比较深刻的面试经历和大家分享吧

当然，面试的时候一般会让你先介绍你简历上的项目实践，然后根据你简历上的东西会问一些东西，下面是我还记得的：

某互联网公司（计算机视觉岗位）

对目标检测的算法了解多少，请简要说明

简要说说霍夫圆和霍夫线

说说什么是梯度弥散和梯度爆炸，为什么会发生梯度弥散和梯度爆炸，怎么防止

说说 kmeans

简历中有小目标检测，那你这个小目标是怎么定义的

小目标检测用的什么网络结构，在网络结构上有什么创新？（用的 ResNet101）

为什么用 ResNet,

ResNet 有什么特点

另一家（计算机视觉岗位）

谈谈 CNN 的各层

说说 RCNN、fate-RCNN 和 faster-RCNN 的区别

faster-RCNN 的 roi pooling 是怎么工作的

CNN 中为什么用 relu, 还可以用别的吗

什么是感受野

深度学习中有哪一些常用的损失函数

你的数据集做过数据增强的处理吗? (做过)

那都做过那些数据增强技术?

目前, 我也找到了我喜欢的工作, 刚开始已经在另一家入职了, 但那个工作环境确实不太喜欢, 果断离职了, 在七月中旬入职这一家, 算算快两个月了, 还是很开兴的。找工作其实就是个体力活, 就像“六项精进”中说的: 不要有感性的烦恼。加油!, 被抛弃不放弃。

说了这么多, 我其实已知觉得自己就是个半桶子, 没有花很多时间来学习, 虽然学过几个月, 但真正学习机器学习的时间有限, 所以我这个半桶子都能找到喜欢的工作, 你们又有什么好担心的呢? 希望大家都能找到自己心仪的工作。

再次感谢七月在线, 希望你能越办越好。

一个大龄码农的奋斗历程：在焦虑中转型拿到年薪 40 万

1、先介绍一下背景：我是一个普通毕业 7 年以上的大龄码农。毕业于一所普通院校，学历是本科，毕业后一直从事后台开发，也就是纯写业务逻辑的那种，没啥技术含量，没有大公司的工作背景，通过 4 个月机器学习集训营的培训，顺利找到机器学习算法的工作，年薪达到了 40W，自己感觉很满意。

2、我为什么来 7 月学习算法 工作时间长了后会产生焦虑，焦虑的原因是随着年龄的增大，发现跳槽越来越难了。说一说我 2 年前心惊胆战的找工作经历：因为前公司经营业绩不佳，我们项目组被裁，裁了以后就去各大招聘网站海投简历，本想着跟以前一样会受到好多面试邀请，然而现实却是非常残酷，收到面试邀请寥寥无几，那段时间特别焦虑，担心自己一直找不到工作，最后用了好几个月才有一家创业公司给 offer，回顾以前几次跳槽都会很快找到满意的工作，但是这次一次跳槽找工作真的太费劲了，让我印象深刻。后来静下心来反思自己的问题，总结出来的原因：自己的能力与工作年限严重不相匹配，没有核心竞争力。再加上看到网上经常说的大龄程序员面临失业的新闻比较多，感觉自己的程序员生涯快结束了，特别焦虑。想着去学习一些新技术来提升自己，在参加集训营之前，也看过机器学习的视频课程，苦于一个人单人作战加上自律能力比较差，感觉自己一直没有入门。带着各种焦虑和看视频效率低下的各种困惑来参加七月的机器学习集训营，希望通过学习机器学习能延长自己的码农生涯。

3、我是在职学习，并没有大部门同学脱产学习的勇气，因为还有房贷，老婆孩子要养。在职学习的弊端还是很多的，没有大片的时间集中学习，每天上班就很累，还是每天逼着自己看在线视频，做练习题，做阶段考试题目。在职学习太辛苦，效果也不如脱产的同学技术扎实，如果大家经济条件允许的话还是推荐大家脱产学习。脱产学习有大片的时间可以集中学习，遇到问题可以跟同学之间当面交流沟通。好在七月集训营的课程安排合理，每次视频

课都有课后作业，每个小单元完了，还有单元测试，每周会有线下交流课程，能最大限度的保证大家的学习质量。

4、面试的一些教训：

A 刚面试的话先拿几家小公司练练手，感觉差不多了再去大公司投递简历。如果是大神的话可以无视。

B 如果面试机器学习的岗位的话：LR,树模型, Random Forest, xgboost, svm, knn,kmeans 这些算法的理论一定要很熟，各种机器学习评价指标，这里我就不一一列举知识点了，老师讲课会把重点划的很明白。

C 面试中问项目中最多的的是数据划分，数据清洗，特征工程以及线上测试，模型保存等。要多多重视。

D 一定要能熟练手动推到：LR, xgboost, svm。

E 没事刷一刷 leetcode,大公司面试必问传统算法的相关题目。

5、面试对我来说不是一个很轻松的事情，因为基础比较差，从刚开始面试中各种被鄙视，各种面试一轮游之后。慢慢的打基础，回看集训营的视频，跟同学们讨论面试问题，打磨自己的技术，最终才找到合适的工作。通过学习机器学习集训营，我发现好多大公司都约我面试了，京东，滴滴，新浪，高德等,虽然没有遗憾没有通过这些面试，但是还是很欣慰。大家面试中也不要因为几次的失败而否定自己，祝各位学弟学妹好运

集 4 学员面试经验：经过五轮面试最终拿到年薪 40 万

某大型互联网公司的笔试题

- 1、Adj-Rsq 是否会小于 0;
- 2、一个排列题，总共 13 张扑克牌，翻开第一张，然后第二张放到最后，再如此重复下去，加入 13 张牌是黑桃 A 到黑桃 K，翻盘顺序是 A-K，那么刚开始的时候黑桃 K 应该排在第几位；
- 3、抽奖的题目，总共有五张卡片，其中有一张是有奖的，其他都没将，A 抽了一张没有看有没有中奖，B 抽了一张立刻查看并发现没有中奖，C 从剩下的三张中抽了一张，问 C 中奖的概率；
- 4、是置信区间与样本量和置信水平的关系；第五题是显著性水平 α 与第一类错误和第二类错误的关系；
- 6、 $\{1,2,3\}$ 的值域 (range) 是什么，平均数，标准差和方差等的关系；
- 7、已知 x 和 Y 的联合密度函数为 $P(x,y)$ 则熵值 $H(x,y) > H(x)H(y)$;
- 8、考得线性最小二乘估计中 $L2$ 正则什么的，不是很了解；
- 10、python，`x=50; def func(x);x=5 func(x);print("x is",func(x))`，问会输出什么结果；
- 11、考得 SQL 中的 Union 聚合，程序没看懂，因此也记不下来了；
- 12、后面还考了 EM 算法，隐马尔科夫连 (HMM)，SVM (问 SVM 分类时大部分向量都不是支持向量，那么分类的时候删除这些变量会不会影响分类结果)，决策树 (训练集和预测集，不太记得了)
- 13、考了节点 A-B-C-D-E-F-G-H-I 中如果出现错落怎么监测怎么发现，怎么知道某处出现了拥堵~~

14、还有一个我认为比较人性化的题目，就是你奶奶没读过书，但对所有认知都很全，你要如何向她介绍你的工作和工作内容

15、怎么分别网红店铺和普通店铺，用什么方法~~

16、C 中 int 型与 float 型的和是什么类型。

某互联网面试过程，经过五轮面试拿到 offer

一面是电话面试，因为我是统计学专业问的问题既有涉及统计方面的知识，大部分还是问算法里的知识点，我面的岗位是 NLP 方向的，问的问题有：

bagging boosting 的区别

异常值的识别和处理

拉普拉斯平滑

缺失值处理

arima 模型 p d q 识别

什么情况会导致过拟合

主成分分析

因子分析

决策树怎么剪枝

L1 范式 L2 范式的区别

各种距离算法的区别与联系

CNN 算法的原理，以及 word2vec 的特点

tfidf 算法的原理，具体场景怎么计算

还问了自己以前工作中的一些项目经验。

二面是现场面试，主要介绍自己在工作中参与的项目，面试官主要根据我的描述，提一些相关问题，所以要去面试的同学们最好对自己做过的项目要很熟悉，每个流程都很了解，还有最重要的是一面中没回答上来的问题或者回答的不透的问题回去一定要继续研究因为后续面试可能还会接着问到。

最后几轮面试基本都差不多，面试过程中就是要自信一点，面试之前要充分复习，了解算法的原理，多积累

普通嵌入式研发转行 AI

我是机器学习集训营第八期的学员，德国克劳斯塔尔工业大学硕士。在校时实验室研究信号处理，主要是面对生物信号的去噪处理。毕业后入职了一家国企事业编制，主要是一家海军武器研发与试验单位，主要从事嵌入式软硬件开发及分布式控制系统的开发。

就在入职后第二年有了一次和 AI 接触的机会，具体项目因为涉密就不展开叙述了，从此在兴趣上喜欢上了 AI，但是之前没有接触过相关的学习，这时单位要组建 AI 研发小组，我是非常感兴趣的，由于是内部竞聘，也许我的面试经验只能给与少数的同学带来帮助，但是既然要写学习经验就还是详细叙述一下。

1. 开始自己看书学习，从一本叫《深度学习》（Ian Goodfellow 等）的书开始，书确实不错，但是由于同时还有嵌入式的工作任务，学习进度缓慢
2. 办公室的同事向我介绍了 7 月在线，我详细的看了课程设置，感觉不错，于是报名
3. 学习进度设置是合理的但是作为产品研发工程师的我，工作任务及压力确实挺大，接连的出差，彻底的打乱了我的学习计划
4. 就在这个时候 7 月的老师联系了我，我才知道 7 月的老师是那么的负责，于是和老师协商停课，再下一期继续学习
5. 第二期开始了先从 Python 入手开始学习随着学习的深入自己也开始研究图像分类，先从 Tensorflow 开始做猫狗分类，然后接触 Vgg16，确实收获很多
6. 过程中发现了很多问题，由于工作时间，只能慢慢跟着 7 月的视频进行学习，7 月的老师非常的负责，每问必答
7. 就这样内部招聘开始了，资格审查通过后就是面试

内部面试的侧重点我认为在学习能力的认可上，在更加 7 月学习的构架上我一一的介绍了机器学习的每个知识点，面试的老师深入的询问了一些算法应用的问题，面试顺利搞定，而 7 月给予了我很大的帮助。

还有 10 天就去新部门报到了，准备跟随 7 月进一步的深入学习，这就是一个一边是嵌入式开发任务进行中，一边学习 AI 的我的经历。

希望能给像我这样类似经历的你一点点帮助。

最后我必须要说感谢 7 月对我带来的帮助。

双非渣本三年 100 次面试经历精选：从最初 iOS 前端到转型机器学习

虽然说起来可能会让大家感觉有些诧异，但到目前为止，我参加面试的公司数量应该是超过 100 家的，当然，岗位不仅限于机器学习，算是一段很不成功也很坎坷的经历，在此跟大家分享一下，本篇文章重点不是介绍机器学习在面试时，而是我本人关于面试本身的一点心得体会，希望对七月的学弟学妹们有所启发。

我求职过的岗位有 iOS 开发，web 前端开发，机器学习算法，都是技术岗，其中 iOS 求职过的公司应该算最多，差不多有六七十家左右，web 前端应该三十家，而机器学习差不多二十家。在此我选几次印象比较深刻的面试经历重点介绍一下。

1、面试地点：北京 岗位：iOS 开发 时间：2016 年 4 月 期望薪资：9K（面试失败）

首先，我强烈建议求职的同行们应该尊重并且在面试之前就去了解每一个给你发面试邀请的公司，因为这不单单是为了求职的成功率，如果公司不适合你以后发展，或者说岗位不合适，那么即使是发了 offer，也请拒绝。

有人可能会想难道提前了解自己求职的公司不是常识吗，在这里我想强调一点，倘若你去面试的是一家名气还算不错的公司，你早就已经心仪已久，这种情况下就算不说你也会去找这家公司的资料看，但如果你面试的是一家根本没听过的小创业公司（这是绝大多数人会面临的问题），而且这种面试一周可能会接到十个，你还会为每一次面试都去做这种准备吗？

在去这家公司之前，我翻了一下当时拉钩上前面几位求职者的评价，虽然不是很多，但也可以看出公司的一些东西。比如创业公司，公司技术这边本身有 iOS，甚至有一个求职者竟然直接在面试评价中直接写了公司想 8k 招一个 iOS 开发。

当时记得是和另一个面试者一起面的，至于原因呢，竟然是他提前了一个小时来参加面试。面试基本上是我先回答的，如果另一位面试者有其他想法就补充，虽然具体问的问题，记不太清了，感觉当时自己答的还不错，隐约记得应该是问的单例，代理和 block 的几个问题。面试官有两个，一个是做技术的，另一个应该是公司的 CEO，至于公司本身是做体育赛事记录的，看得出 CEO 本身也是个很有体育运动精神的人。聊到薪资，因为有了准备，所以我要了 9k（只能说我是贪心的，我想多要点）。

我记得最清楚的是面试的最后一个问题，“你有梦想吗？”。我不假思索自作聪明的回答，有啊，我的梦想就是成为 iOS 编程大牛。现在想想，似乎这个答案简直比回答没有更平庸。

在此我希望读者朋友们也记住这个问题，只要你还有一丝不甘平庸的心态那么它或许就会在你选择惰性生活的时候给你当头一棒。

有些时候我们为了更好地达成自己的一些目的，难免会在话术上使用一些小套路，但此时，请不要丢失掉真诚。但凡是经历过多年职场洗礼的人，都能分得清哪些是套路，哪些是真诚，与其耍一些小聪明，倒不如把最真实质朴的一面展现出来。

2、面试地点：北京 岗位：前端开发 时间：2016 年 11 月 期望薪资：16K（面试失败）

公司对于每一个岗位都是有预算的，希望用尽可能低的价格招到合格的人，退一步讲，哪怕是技术非常过硬，但公司的预算摆在那里，肯定也是没办法用超过预期的薪水录用求职者的。这一点在我之后的这次前端面试时，体会尤其明显。那家公司的技术总监是做后端的也会写一些前端，也会一些 JavaScript，那次面试我表现的确实不错不仅仅只是感觉，我和那家公司技术总监首先在技术上聊的很愉快，他问的几个技术问题我回答的都还算满意，但聊到薪资时候，我说我要 18k，然后他不好意思的说我们的岗位预算是 10k，如果你愿意来

的话，我感觉最多能跟老板争取到 13k，当然你也可以坚持你之前的要的期望薪资，后来我说我最少能接受 16k，晚上那边就邮件告诉我，面试没有通过。

3、面试地点：北京 岗位：前端开发 时间：2016 年 12 月 期望薪资：18K（面试成功）

这是第一家，我成功混进去的公司，但是因为工作压力巨大，并且经验的欠缺，再加上部分同事的白眼和冷嘲热讽，仅仅入职 10 天，我就辞职了。

当初我看到这家公司的招聘信息是在拉钩上，在拉钩上投递简历的同时，这次我走的路子比较野。在网站上搜索这个公司的名字看是否有官网（竟然真的有）。官网留下了公司的联系电话，我直接就打过去了，接电话的正是这家小公司的 CEO，我说想应聘公司的前端开发，对方很痛快的就和我约定了面试时间，就这样我成功的争取到了面试机会。

第二天去面试，也是一个后端技术负责人面试我，基本上问的问题都是简历上我罗列的那些知识点。这些我早就烂熟于胸了，在此需要说明一点就是，如果有些知识点你感觉掌握的不是很牢固，千万不要往简历上写，再或者掌握某些框架不是很牢固，尽量把这些东西写在后面，要不然上来就说不会，这次面试就凉凉了。但是你再简历上写的知识点又不能太少，否则的话对方会感觉没东西可问，也会很尴尬。

总之，这次面试很成功，我本以为我可以靠加班学习来弥补工作经验的不足，但很可惜，这家公司并没有给我这样的机会，工作时间是早上 9 点半到晚上 10 点，周六全天，但是周日即使放了一天假期也会让你在家里改 bug。总之，如果是刚培训出来的同学，请不要尝试去这种公司，因为这里没有时间学习，只能不停的产出，十分不建议。

4、面试地点：北京 岗位：前端开发 时间：2017 年 1 月 期望薪资：13K（成功）

这次面过之后，我在这家公司待了一整年，在这一年的时间里，我接触过的 web 前端项目包括 jq, angular, vue 以及 hbuilder 打包的混合式开发等技术，对于一个刚培训出来的人来讲，这家公司真的是太完美了。第一轮面试是我的直系领导面的，其实我当时因

为没有工作经验，技术水平菜的很，但因为我们领导面试时有一个巨大的漏洞，我还是侥幸蒙混过关了。

漏洞就是他问的问题都很泛，也就是每当问到一个知识点的时候，我只需要表示我了解过，并简单做几分钟的阐述，甚至可以说就是背一遍面试题，他感觉可以了，就不会深入刨下去的问了。这其实真的是很不好的面试习惯，包括后来我入职之后很久了再和他聊天的时候说到这个问题，他也说确实当时面试没总结好到底该问什么。

虽然现在培训机构巨多，每天 hr 送过来的人也很有可能是没有工作经验刚从培训机构毕业的，但想通过几个问题守好公司入职的大门还是可以做到，那就是通过一个知识往下深刨，连接更深入的知识点，然后通过求职者的表现，就能很容易的判断出求职者的水平。

如果面试官真的表现出了专业的这一面也不是无法可解的，我的解法就是想办法把面试官带进你自己的节奏里，引导面试官问你熟悉的问题，并把你熟悉的问题彻底展开，把所有你能串起来并讲的差不多的点都聊一下，掌握主动性，或许你会感觉自己本来理解就不深，岂不是说的多错的多。非也，很多台词不是你现场现编的，是需要你在面试之前就把一些东西背下来的。

如果能更好的贴合公司的需求那么你被录用的可能性就会大很多，这是我的领导后来聊天的时候告诉我，其实我来面试之前已经决定了要录用一个人了，但公司这边有混合式开发的岗位要求，所以我会 iOS 的优势成功使我成功截胡了。

5、面试地点：北京 岗位：算法工程师 时间：2019 年 3 月 期望薪资：21K（面试成功）

这也是我现在工作的公司，目前还处于创业期，公司 30 人，并没有专门的算法工程师，所以很大程度上讲是没办法判断求职者是否合格的，没有技术守门员，那么就只能通过交流的细节来判断求职者是否尊重这次面试以及表达能力是否达标，还有就是薪资要求是否合适。

首先刚来公司是做一道 python 的排序算法逻辑题。第二题是正则表达式，这个就没办法了，写也不知道写什么，只能空着。第三道是一道关于算法推荐的理论题，说实话，我之前专门准备过这方面的内容，毕竟我是想求职推荐算法工程师嘛，洋洋洒洒的写了两页纸。作答第三题的字数，我认为是公司判断我合格的一个关键原因，所以面试的时候哪怕是不会的东西，请你也尽可能的做到尊重这次面试吧，因为谁也没法告诉你究竟结果是什么样子。面试一共三轮，第一轮是产品总监，第二轮是后端开发，第三轮是人力总监。

之后不久我就接到 offer 来到了这家公司，说实话一点不忙，感觉很不错，有充分时间学习。

总结一下我对目前已经涉及过领域的理解，目前的时代，机器学习算是一个年轻的行业，目前的市场还是供远远小于求的。但从编程的角度讲，机器学习的入门课程门槛之高，确实是其他语言远远不能比拟的，主要原因还是因为数学公式原理的推导太过难以理解，而且很多公司的要求是硕士学历起步，像我这样的渣本二流院校的毕业生虽然找工作不算太难，但感觉也有点浑水摸鱼的意思。

但有意愿转型机器学习的朋友们也不要被难度所吓倒，个人认为我的人生中真正考验智商的只有高考，这种是在规定时间内出成绩的游戏，而且是好几门一起考，想门门精通确实不易。

至于转行转岗，只要时间坚持投入足够持久，必定会取得收获，所以肯定是没有那种学不会的说的，只能说投入的时间不够而已，即使转型失败了也绝对不是因为智商上的差距，因为努力的付出量级还远远没到拼智商的程度。

还有一点就是我说的学习是每天全天投入去学习，而不是每天高兴就学两小时，不高兴就玩两小时，如果是这样，可能三年都没法学成。最后，也是最关键的一点，那就是哪怕是你面试的时候被虐的面红耳赤，也请你保持绝对的自信。因为，哪怕是你牺牲尊严，也不会获得尊重，没有任何一个面试官会愿意招一个自己都感觉自己很 low 的人。

双非渣硕，机器学习面经

自我介绍

先自我介绍一下，我是七月在线机器学习集训营第三期的学员。应该也算是老学员了吧哈哈。在七月在线学习时我还是研二学生，就是通过集训营，我才找到了对口的实习工作，然后顺利转正，今年七月开始正式上班。工作在一个二线城市的外企，应届毕业生刚进来工资不高只有 13k，但在这个城市可以够生活和还房贷的，而且朝九晚五不加班，假期和公务员一样放，我也挺满足的了。下面开始说说我当初应聘这个公司实习生的面试经验吧。

面试经验

记得那是 18 年刚过完春节，我在七月在线的课程学习也进行了一半了。在学校已经进入研二下学期了，由于老板根本不管我（不知道是不是由于学校不好的原因），所以我决定自救（不能硕士毕业连个工作都找不到吧）。我开始制作简历找实习工作，我把在七月在线学到的东西都写进了简历中，那时学了 Python、爬虫、sklearn、xgboost 相关知识。简历制作好之后就开始了投简历，只要是 python 相关的实习工作我都投（在实习僧上投的），等了一段时间后就有公司发来面试通知了，第一个公司的实习工作是在一个猎头公司，需要写爬虫去爬取招聘网站上的职位信息。想着在七月在线也学习了爬虫的相关知识，于是就去面试了，面试问题如下：

1. python 常用的爬取数据的框架或者方法有那些？
2. scrapy 的工作流程是什么？
3. scrapy 怎么做分布式爬虫？
4. 有哪些比较常见的反爬机制？
5. 反爬机制对应的破解方法是什么？
6. 如何开启 scrapy 的去重？

其实问的都是比较基础的问题，比较简单，基本上都回答出来了，然后面试官给我指定了一个网站，让我回家写个爬虫去爬取该网站两天后发给他，我回家一会就写好了（七月在线的课后作业中有的，只需要把网站改一下就好了），然后发给面试官，下午就给我发了 offer，嘻嘻。

在猎头公司实习了有两个月，此时已经来到了 18 年的 5 月份，七月在线的课程也都全部结束了，有天晚上我躺在床上看实习僧（想找机器学习的工作），看见了离实习公司不远的的一个外企在找机器学习实习生，我抓紧投了简历，第二天就有回应了，公司的 HR 打电话来和我约面试时间，于是约到下个周一（几天时间突击下，准备准备）。周一请假去面试（说身体不舒服，不敢说去别的公司面试，哈哈），一进到公司，就感觉到外企就是不一样，办公环境就是好，就是有钱。面试官有三个，一起面试我，一人一台 macbook pro，当时真的好羡慕他们，心想我也有一台就好了，扯远了，说说这家外企的面试经验吧：

1. 自我介绍
2. 介绍自己做过的机器学习项目
3. 熟悉的机器学习算法
4. SVM 的原理
5. SVM 为什么采用间隔最大化？
6. SVM 怎么处理过拟合？
7. SVM 有哪些可以调节的参数？
8. 常用的 SVM 核函数？
9. 如何使用 SVM 处理多分类问题？
10. GBDT 与 XGBoost 的区别
11. XGBoost 如何解决缺失值问题？

12. XGBoost 中的树剪枝

13. RNN 原理?

14. 神经网络处理过拟合的方法?

可能还有些问题，我记的不是很清楚了，但基本都是七月在线课程中说过的知识点。面试结束后由一个面试官送的出门，并和我说希望我过来工作，当时就感觉稳了，美滋滋。

最后辞掉了爬虫的工作，来到现在的公司做机器学习了，在公司实习了半年，通过了实习生的转正考核，顺利转正。先在我也在用 macbook pro 写面经了，哈哈哈。

总结：

总之，现在面试中遇到的机器学习问题，七月在线的课程中都会有讲解到。本人特别喜欢寒老师的课，讲课很通俗易懂。可以说没有七月在线，就成就不了现在的我，感谢七月在线，感谢 July 校长，希望七月在线越办越好，永远支持七月在线

机器学习面试心路历程：排序算法和 xgboost 被问的概率很大

我是机器学习训练营四期的学员。在 2018 年那个烈日灼灼的夏日，每个周末早上早起，坐将近两个小时的地铁去上机器学习的线下课，但线下课的质量没有辜负我的辛苦通勤。临近结课，我开始了我的面试之旅。我的处女面献给了什么值得买，该公司在丰台。

我以前没听过这家公司。处女面难免会准备得不充分，面试官针对我简历上的项目问了几个问题，刚开始我回答得不错，有点沾沾自喜，结果在一道问题上卡壳了。除了针对简历上的问题，常规的数据结构与算法题没有考察，而让我做了一道推测北京出租车数量的问题。这种问题没有固定的答案，只要言之有理，有一定的逻辑，能自圆其说就行。最后面试官主动加了我微信，让我回去等消息。后来从面试官的朋友圈中看到，面试官去深圳敲钟去了，这个公司今年在 A 股上市了，看来业务做得不错。后悔没有加入这家公司。

中途参与了海淀那边几家互联网金融公司面试，没有考察数据结构与算法，只是针对简历上的项目经验问了一些问题。虽然我简历上的项目与互联网金融有些关系，但是我对互联网金融行业有些知识不了解，最后面试失败。

连续几家公司面试都失败，我很受打击。这个时候要学会鼓励自己。当上帝关上一扇门的同时，也会打开一扇窗。我想起《老人与海》中的名言：一个人并不是生来就要被打败的，人尽可以被毁灭，但却不能被打败。就像歌词里唱的那样：爱要越挫越勇，爱要肯定执着。于是舔着伤口，总结面试经验，继续投递简历，继续面试。

后来去 58 到家面试，面试官是我未曾谋面的校友，我俩一见如故，面试过程中我很放松，他针对我简历上的项目问了一些问题，我应答如流。然后他找来技术总监来面我，还是针对简历上的项目进行提问。最后技术总监找来一张白纸，让我写冒泡排序。在算法与数据结构这块我一直没有准备。没想到一道冒泡算法排序葬送了这次面试。后来面了一家创业公司，让我写 xgboost 推导过程，这一块我做了准备，没有丢分。最后入职这家创业公司。

最后做几点总结：

- 1, 简历投递路径：boss 直聘、拉勾网、智联招聘、前程无忧。如果身边有朋友能够内推，当然会更好。
- 2, 基础算法原理、推导要熟记，例如逻辑回归、SVM、XGboost。
- 3, 针对简历上项目可能要被问到的问题，要做好预案，罗列出可能被问到的问题，针对每个问题，要做好应答准备，最好落实到文字上，这样印象会更深。如果你是面试官，看到这样一份简历，你会问哪些问题。可以找身边做开发工作的朋友做面试官，给你模拟一下面试。
- 4, 算法与数据结构有时间就准备，至少把最基本的一些算法准备一下，例如十大排序算法。如果实在记不住，就把原理记住。如果面试的时候遇到这些问题，写不出代码，就把原理说一下。这一块不是所有公司都会考查，但越是一些名企，考查的概率越大。
- 5, 多去参加面试，每次面试都要总结经验，小公司也要去，如果不想去小公司，最好也要去面一下，就当练手。有些问题只有在面试过程中才会暴露。例如我在一次面试过程中居然笑场，犯了这么低级的错误。失败的经验尤其宝贵。
- 6, 心理建设：刚开始可能会失败，但别灰心。要学会鼓励自己，这一次离成功更近一次。面试有点像相亲，也是讲究缘分的。有些公司没面上，不代表你能力不行，只能说是不合适，毕竟面试是双向选择。每次面试的时候要自信。当你自信的时候，你才发挥得更好。

我和七月在线的 5 年时光：从经济学专业到“高级”程序员

很早之前就像写点东西，来记录我如何从一个非专业的学经济学学生，一步步的变成一个“高级”程序员的过程，一步步走来路上有过犹豫，有过茫然，有过自我怀疑，有过胜利的喜悦，借着国庆与祖国母亲普天同庆之际，抽空来记录下。

我是一个三本毕业,来自一个大家没有听过的学校毕业，学的是经济管理，毕业之后，按照家人的意思，来到了家人朋友的公司开始了实习（公司在一个 4 线城市）。初入社会对这个社会感到了各种好奇，都说学生时期是人生最幸福，最简单的时光，现在看来一点不假，甚至有点怀念，刚进入实习单位老板安排我到了销售部跟着经理学习“深入了解这个社会”，主要是一些人际交往，人情礼节，做人做事的方式，加上我比较外向，对此感到甚是快乐，就这样做了半年的时间，跟着领导见各种客户，参加各种饭局。

慢慢的我感到了社会的深浅，慢慢厌倦了这样的没有实现自我价值的生活。我觉得自己要做点什么，要为社会做点什么（那时候我也是一个愤青了，哈哈），然后和家人商量自己想去大城市闯荡一番，正好高中挚友在北京是搞 IT 的，平时也经常聊在大城市各种新奇的事物，以及挚友最近又做了一个什么项目，从交谈的过程中，我感到了他内心的成就感，发现我的生活中就是缺少了这样的东西，家里人虽然也不舍得我一个刚毕业一年不到的孩子独自去大城市闯荡，但是最终还是答应了。我也在实习单位交接了下，到学习毕业答辩，准备好了一切，我背着我的被子，行李买了张火车票，来到了北京。

在同学，朋友的介绍下来了一家外包公司干了稍微占计算机行业边的工作，现在看起来就是 IT 运维了，还包含了一些修网线的工作，对就这样挣着钱慢慢生活了 3 个月，感觉对未来生活的茫然，因为自己也没有技术，也没有深入的学习过专业的东西，和老友交流了内心的想法，老友给我推荐了 cplus，以及 linux，apue 书，并且告知我，这是一些计算机人必知的东西，然后我就开始了计算机专业的学习，就这样没有基础的情况下，用了半年的时

候，我把这些书勉强看了一遍（对，是看了一遍，看完了之后，只是对一些简单的东西有点印象），心里产生了很大的自我否定，怀疑，也算是自己运气比较好把，过程中朋友一直鼓励我，指点我，我也自己通过互联网学习各种计算机的东西。

就这样第一份工作干了一年半，有了想去大公司闯荡的想法，网上开始整理简历，找朋友推荐，结果还是不尽人意，原因也比较简单，自己理论，实践技能不足，大概得面试了 10 多家公司之后，我整理了各个公司的面试问题，继续面试，在面到第 13 家的时候，一个外包公司愿意和我谈了合同，虽然从一个外包公司到另外一个外包公司，但是这次干的东西是我心里一直想干的事情了，就这样来到这个外包公司外派的公司，是一个大国企银行，这是我真正开始搞计算机相关工作的开始了，开始工作干的东西都比较简单（现在看来），遇到各种不会的东西我都会去请教身边的同事，以及我的老友。

就这样慢慢的干了 1 年时间，从开始的满脸懵逼，到最后的得心应手（其实就是复制，粘贴了），慢慢感觉到了技术不好，只能干这些大公司人不愿意干的东西。毕竟内心还是想到一个技术比较强的公司去深造下，扩展下自己的视野，也增加自己的技术能力，但是大公司怎么是想进就进呢，没有一技之长的话也是做白日梦了，询问了很多 IT 朋友，以及老友，最后得到了一个结论：当你技术不强时候，数据结构，算法就是你的敲门砖。

这个时候我开始了算法，数据结构的学习之路，买了本算法导论开始了学习，当我拿到书的第一瞬间，我整个人都愣住了，这是什么东西啊，完全看不懂啊，完全不知道在讲什么，内心的茫然，瞬间无比的失落，对自己想去大型互联网公司完全失去了信心。这个时候老友出现了说：没事，数据结构，算法本来就是很难东西，慢慢来。

于是我开始了网上各种搜索信息，在 CSDN 上看到了一个大神写的博客，对这个人就是 July，七月在线的老板，我应该是看了 July 所有的博客，感觉很有深度，心里的膜拜之心油然而生。后来听说，July 要做一个算法的培训班，心想多好的机会，终于可以跟着

各种技术大牛学习了，那时候是 2015 年，对于七月在线应该是第一次举办算法培训，我很开心的参与了，跟着网络班开始了算法学习，但是算法，数据结构这种东西真的是太难了，不过是和一群兴趣相同的人一起，大家一起上课，做习题，练习，每周六周末学习，就这样学了 2 个月左右，加上自己的练习，满怀信心的准备其他技术知识，准备尝试大型互联网公司。

真的是希望越大，摔的越狠，当然不是算法问题导致的了，大公司要求的还是比较多的，没有办法只能硬着头皮开始学了，计算机原理，apue，linux，巩固算法，就这样又过了半年，因为个人原因，女朋友在上海，我离开了北京，来到了上海，凭借学到的东西，开始面试，自己也比较幸运了，面试过程中遇到了很多 nice 的技术大牛，他们对我一个非专业毕业的面试者表示了肯定，我也顺利的拿到了 offer，虽然不是一个大型互联网公司，但也是一个互联网金融公司了，我开始了自己的新的旅程。

期间我也没有停止学习的脚步，因为公司是做互联网金融的，我在七月在线参加了量化交易与机器学习的课程，对，工作中一边使用，一边运用学习的东西，成长速度是很快的，我把学习到的想法经常和同事，领导一起讨论，领导对我也是很满意，让我去做了一些学以致用用的东西，比如股票的相似 k 线，行情预测功能，慢慢的我也成了组里的骨干成员。

这个时候我开始正式接触到了机器学习的东西，深刻的理解到这个技术的未来场景，机缘巧合下，七月在线开始了机器学习的培训班（线上线下结合的机器学习集训营），我心想天赐良机啊，于是毫不犹豫的开始了人工智能的学习。当然学习的过程还是比较辛苦的，上课的地方在浦东，我住在嘉定，每周末地铁来回 4 小时，但是学到的东西还是很开心。从 python 开始入手，学习各种分析工具，pandas，seaborn，也是很又意思，也有生产中的大数据处理数据分析，建模，以及最后的深度学习，整个过程也是很 nice。

虽然深度学习真的有点难，不过老师的讲解比较清晰（老师是阿里高 p 技术专家），我也跟着学到了很多的生产中运用的技巧，在工作中也是使用了同样的技术，用人工智能去做了一个公司的股票行情的分析项目，领导对此也是很满意，当然最高兴的是我了，看到学到的东西在公司被认可，在生产中使用，这应该是每个程序员最开心的事情，当然年底领导也没有亏待我，给了我 A 的评价，当然年终奖也少不了啊，哈哈，时间关系，整个过程的技术写的比较少，后面再慢慢补充，未完待续。

我的面试及就业分享：从后端开发转岗到机器学习算法的历程和感受

大家好，我是集六的学员，现在在一家互联网公司做商品排序算法，已经工作了几个月了，前几天七月的就业老师邀请我做一个面试和就业分享，我很乐意去做这样的分享，一方面或许会对学弟学妹们有帮助（这里我就大言不惭的称自己为学长了，哈哈），另一方面我想借这次机会对我这一年来我的学习、面试和就业经历做一个总结。

首先声明一下我这里不会罗列一些笔试或面试点，因为我觉得之前的学长学姐们已经总结的很全面和到位了，我主要说一下学习和面试中我自认为需要注意的一些点以及我就业以来的一些感受。

先说一下情怀，情怀虽然不值钱，但是我觉得是最重要的东西，我关注机器学习和人工智能已经很长时间了，记不清从何时开始的，因为我一直对科技的前沿技术感兴趣，我微信里的公众号有三分之二都是技术相关的，我本科是物理专业，大学毕业后做了后端开发，后来也是靠着对前沿技术的热情和好奇心促使我转岗到了且称之为人工智能研发岗位，尽管目前技术水平还比较菜。

首先说明一下这篇分享适合像我当初一样基础比较薄弱或者学习和面试中比较迷茫的同学，理论基本功很扎实的未来潜力大牛可以当做一篇小说来读。

先说一下学习的经历：

我是从去年年初开始系统的学习，在此之前都是凭兴趣断断续续、哪里好奇搜哪里，开始系统学习之后，学习时间也都是从日常工作间隙中挤出来的，一天能挤出两个小时来已经很不错了，先是从最基础的数据结构和算法开始的，找了网上的一些教程看，后来又找了一套机器学习算法的视频从头到尾开始看，买书也是我毫不吝啬的，现在我买的书已经放了满满一落地书架，当然不止机器学习的书，无论是看视频教程还是看书，只看一遍是远远不够

的，看了和真正吸收成为自己的知识完全是两码事，但是也不必从头到尾照本宣科的去，还是多跟着自己的好奇心把知识点一点一点攻破，最后你会发现能把这些点连成线。

自学了大半年之后去年十一左右我从公司离职，报了集六开始全职学习，当时也没有太多压力，就想着一边学，一边用学习剩下的时间做其他自己感兴趣的事，打球啊看书啊什么的，这样一直到去年年底，今年年初开始面试。

接下来说下面试的经历：

我本来是想找 NLP 的职位，但进公司之后领导把我安排到了做排序算法。我从开始找工作到收到 offer 经历了大约三个月时间，收到了一个 offer，时间也可能有点长，但是每个人的情况不一样，我也不是什么大牛，我认为自己是一个有好奇心的喜欢挑战的不安于现状的技术小兵。可能有些人找了几家公司就收到 offer 就入职了听起来很牛逼但我觉得并不大好，因为毕竟没有工作经验，多面几家，了解下不同的公司岗位的情况，把面试当做一种学习的经历会对未来长久的工作生涯更有好处，所以找工作不要急躁，慢慢来，收到 offer 固然很重要，面试的学习过程也很重要。

下边说一下我认为面试中需要注意的几个点：

- 1、面试经常被问到一个问题就是模型的 bad case 是怎么分析和解决的，我们上课或学习时一般都是跑完模型看准确率还不错就完事了，其实现实工作中跑完模型仅仅只是一部分，甚至是最省事的一部分，跑完模型之后的 bad case 分析和优化可能占据了大部分时间；
- 2、一些基础算法甚至比近期刚发表的高大上的模型更重要：集训营刚毕业的同学能去到 BAT 或 TMD 这样的大厂的毕竟是少数，多数人面到的更多的是一些中小型的互联网公司，这些公司那些高大上的模型可能都没有提上工作日程，像 NLP 中的 BERT，目前也就大厂和一些很有潜力的创业公司在用吧，tfidf、n-gram、LR、GBDT 这些基础的东西更容易被问到；

3、LeetCode 手写算法：这种题多多少少会有的，我面过的公司一半以上会问到，我当时也是准备了一些典型的题，面试中写不上来是经常的事，但是我觉得即时写不上来思路能说明白也很不错，不一定非得完完整整写出来，像 NLP 岗位中的求最长公共子序列就被问到了三次；

4、工资要合理一些，不要被市面上 AI 岗动不动三四十万甚至上百万年薪冲昏了头脑，结合自己的实际水平，可以请教七月的老师定一个适合自己的薪资；

5、再就是心态问题，面试受打击是很正常的，不要灰心不要气馁，时间能够解决所有问题，不要纠结于一城一池的得失，不断去面试，遇到问题不断的调整，不管是简历的调整还是心态的调整。

最后说一下工作后的感受：

毕竟是转岗，跳出了之前的舒服圈，来到了一个陌生的岗位和工作环境，过程中遇到一些问题也是很正常，也是需要不断的调整自己，面试是一个挑战，入职之后适应也是一个很大的挑战，我刚入职不久，现在我就是在不断的学习、不断的调整、不断的适应，一切都会达到自己想要的样子，一切也都是一个过程。

最后做一个总结：

这个总结不仅仅是对学习，不仅仅是对面试，不仅仅是对工作，更是一个对人生的看法，我觉得我们无论什么时候都是一个学习者，未来充满着未知，充满着挑战，未来仿佛一切都不很确定，但有一点是确定的，那就是我们每个人都应该说自己想说的话，做自己想做的事，追随自己的直觉走脚下每一步路。

最后用乔布斯在斯坦福大学做演讲时的一句话做结尾：

Most important, have the courage to follow your heart and intuition. They somehow already know what you truly want to become. Everything else is secondary.

算法面经：应届生如何通过四大金刚拿到年薪 30 万

我是机器学习集训营的学员，硕士某 211。说来自己进入机器学习这个领域也是机缘巧合，本科是电气行业，但在硕士期间跟随实验室做一些数据挖掘的项目，算是踏入了这个领域。在校学习期间，自己也从知乎等其他平台找到了一些视频、书籍，但终归不够系统，而且实验室项目数量有限，始终觉得自己在机器学习的入门槛徘徊，后来偶然接触到了七月在线的视频，觉得老师讲课很生动，于是就在七月平台上购买了其他课程进行学习，但人毕竟有惰性，视频看着看着也就搁置了。

在研二接近结束的时候，了解到了七月的机器学习训练营，线上线下的模式让我很心动，再加上秋招的压力，自己还是狠了心报了班来参加，尽管后来因为种种原因没有完成最终的课程，但也收获了不少东西，简单总结一下就是一下几点：

首先就是校长经常强调的四大金刚（小编注，四大金刚指的就是：“课程 题库 OJ 竞赛”哦），个人觉得这个是一个算法工程师需要一直坚持的事情；其次，在机器学习的过程中要注重理论的推导和实践的结合，公式要反复推，项目要亲自实践，实践中会发现很多新的问题。

当然希望以后有机会可以去七月回炉重造，哈哈~

秋招当时由于家里有些事，所以面试经历不多，现在就简要回顾一下实习和秋招的经历，时间太久，有些问题已经记不清楚了。

实习：

A. 某互联网找房平台

算法题（关于二分查找的），对简历上的项目进行介绍，接着面试官根据项目用到的算法进行提问，当时项目经历主要跟 xgboost 相关，因此面试官从决策树的相关问题开始提问，面试问题如下：

1. 决策树的主要类别
2. 决策树的减枝
3. GBDT 和 xgboost 的区别
4. GBDT 的基本原理
5. L1、L2 正则化
6. 概率题（主持人问选手选了是否改选的问题，记不清了，七月概率课上有讲）

B. 某出行公司

面试主要两部分，项目经历以及算法题，算法题一道剑指 offer 上的原理和一道动态规划的题目。

1. 偏差与方差的区别
2. 什么是过拟合和欠拟合，如何判断模型欠拟合
3. K 均值算法的优缺点以及如何调优
4. Xgboost 的算法原理

秋招：

C. 某互联网银行

一共有三面，一面是技术面，后两面是群面：

一面：面试官一对一面谈，介绍简历项目

问题：

1. 常见的 CNN 的结构
2. LSTM 梯度消失爆炸，LSTM 的结构
3. Xgboost 的调参细节
4. Xgboost 相比于 GBDT 的优势

D. 某大厂

简历项目介绍，关注项目中的细节问题

问题：

1. 机器学习中的采样方法，现场实现
2. 给定业务场景，写一段代码

总结：对于在校生的建议，大家有精力的一定要先去实习，实习经历对秋招有很大的加分，而且实习转正的机会要比秋招容易得多；对于秋招，面试官考察的是大家的基础，所以要多刷题，对于数据结构中的图、二叉树以及各种排序都要能熟练的写出来，算法题是否能够熟练的写出来往往决定最后面试的成败，毕竟大家进去公司还是要做项目，要有工程实现的能力。另外，学习中对于一些算法建议手动实现，一是加深理解，二是锻炼 coding 能力。对于一个项目要了解透彻，了解为什么要用这个算法。机器学习的过程是一个长期积累的过程，github 上也有很多项目代码，大家多多学习，不断积累，有时间也多阅读一些顶会的文章，了解机器学习的最新动态。最后预祝大家都能找到一份满意的工作！

集三过后月薪 double

我是集三学员，在学习集三前在一家公司做一些基于规则的算法，期间也在网上自学了一些公开课。自学了一段时间，发现虽然理论知识了解了一些，但是仍然不知道这些知识怎么在实际中落地。偶然直接了解到七月在线，便看了看当时集三的课程配置，觉着内容挺全的，也有一些是网上的公开课没有的内容，便报名学习了。

在为期三个月的学习过程中，老师从 python 基础到机器学习，基本把算法相关的方方面面给大家过了一遍，我觉得最宝贵的就是能从老师那里学到工业界这些算法是怎么用的，这个是和公开课不一样的地方。最后还组织了一堂模拟面试课，这个对大家也是非常有帮助的，让大家了解到面试中容易被考察的点，以及相关的一些算法方向的行情。

我给大家的建议是在学习过程中要认真完成布置的作业，另外最好确定一个自己将来想发展的方向，比如是 CV，还是 NLP 或者推荐等，确定方向后，在把基础知识掌握好的同时，着重了解一下自己的那个方向的专业知识，各个方向都有不同的模型和算法，着重了解一个方向后对找工作会比较有帮助，七月也有一些专业方向的课，可以去看一下。

谈到就业，其实我学完集三后面试和换工作并不多，不过学完后跳槽薪水涨的还比较可观，算是 double 了一下，自己比较满意。面试中我被问到 leetcode 的题并不多，更多的是关于机器学习算法的，比如 LR 的 loss 是怎么来的，深度学习中的 dropout 为什么有效等。这些基本课上也有时有所覆盖的。

在日常工作中我觉得除了对基础算法有基本的了解外，工程能力也应该有一些，比如 tensorflow 等。这能使我们快速上手实现模型，快速迭代进行实验。

超励志！本科化学硕士力学，从干过销售只会 VB 到成功转行量化交易

把时间切成一天一天过，没觉着这个时间序列有什么特别，但今天安静回顾过往，发现这条路蛮曲折。介绍下背景，我本科学的化学，研究生学的力学，都是劝退专业。2015 年毕业，在一个小公司干了两年机械，又干了一年销售，2018 年报了七月在线集训营第五期，今年三月份离职，恶补了下数学知识，七月份入职当前公司。先定义下我所说的“编程零基础”，大学学过 VB，之后再没碰过编程，报名集训营上课的时候，不清楚 `import pandas as pd` 是什么意思，不清楚 `import` 是干啥，也不知道 `as` 是干啥，更别说看了很多 `import` 后突然发现有的地方写的是 `from sth import sth` 那种痛苦了。

关于化学、力学、机械、销售这几个概念背后的种种，我就不多谈了，相信大部分人都了解那一桩桩无奈。只提一点：局部最小值。Vapnik 的《统计学习理论的本质》里面提到了在数学中有一种情况，变量初始化后无法直接求解最大值，只能求解初始化值附近的极大值；此时就只能通过多次随机初始化，依据不同初始化位置求出不同极大值，再对比不同极大值的大小来求得一个可得的最大值。这个理论所讲的具体数学理论我不清楚现在是一个什么样子，但是当我看到这段描述的时候，醍醐灌顶一般：这不就是我们这种劝退专业出身的人所面对的问题么！读大学的时候，不清楚社会发生了什么，随便报个专业有学上就不错了，工作后只能在自己的专业范围内求一个极大值，而这个极大值跟其他人（比如阿里的程序员）相比实在太小了。

对我而言，化学是一次初始化、力学又是一次初始化、机械、销售都是一次次的随机初始化，包括一开始报名七月在线的集训营，本质上也是一次初始化，不同初始化点附近可能存在的极大值是不一样的。

如烟过往如上，新篇开启如下。

2018 年 8 月报名集训营后，我还在上班，每周一从杭州出发去山东出差，基本每天换一个城市，周三晚上找个咖啡馆上视频课程，周五晚上赶回杭州，周末去上线下课程，下周照旧。基本每天就 7 点到 12 点左右可以学习，白天还要拜访客户，当然偶尔也会偷懒，不见客户只学习。大概这样努力了三个月，到集训营结束。

十月份，跟线下课老师聊了下，从应用场景上来说，我觉得自己想做的还是量化交易，不是机器学习。这两个其实还是有区别的。学了算个月，这算是收获吧：知道量化交易与机器学习不是一回事。清晰记得是双十一，正式开始向量化交易进发，刚开始看了七月在线的《机器学习与量化交易》这门神课，那会这个课对我来说实在是太过神奥，完全看不懂。于是转向其他新手友好课程。继续周一从杭州到山东，每天一个城市，周五回杭州，每天晚 7 点到 12 点学习的生活，因为遇到的问题实在太多了，于是在 CSDN 把遇到的问题都记录下来，一天天就这么过，刚才看了下，目前写了 509 篇了。

一直到今年三月份，工作本身无望，我自己也越来越强烈感受到量化交易对我的“召唤”，一狠心一咬牙，辞职了。

三月中旬，正式离职，来到上海，每天去浦东图书馆三楼最边上固定一个位置。对于代码，我都是一行一行手写在纸上，弄懂每一个单词，画出全部逻辑关系及彼此引用；一遍不行，就两遍；

大概四月多，再次看七月在线冯老师的《机器学习与量化交易》，有了前面的基础（需要感谢另一位老师），此时慢慢能看懂了，循着老师的课程，开始理解机器学习，发现机器学习的基础是统计学，统计学的基础是概率论，研究概率论发现有贝叶斯学派与频数学派，继而是大数定律...一层层深入，在浦东图书馆一本本看书，基本上把图书馆概率、统计、贝叶斯、量化相关的书都看了一遍。就这样，一点点的啃，个中波折就不谈了。

一直到七月，开始面试，我没有撒谎优化简历，实实在在写的过往经历，不过重点突出了我已有的工程实现能力：有回测框架、有实际可用策略，用 NLP 分析文本预测大盘...总之，是表明我能直接上手做项目。然后，大部分公司都没理我（这才是残酷的现实），后来面了几家，现在这家彼此相合，我就入职了。关于面试过程，问了些实际写策略会遇到的问题，我自己都遇到过，不算难。机器学习的知识是加分项。当然，不是顶尖量化公司，百十号人，中规中矩的公司吧。

七月份找工作，七月份入职，大概，七月在线是福将吧。七月...要说七月在线对我的帮助，我仔细想了，在于给我开了一条继续学习的路。工作后，以为“见多识广”以为学习没有用，现在回过头看，工作后的学习才是真正让自己成长的。实事求是，条件够了，事儿自然就成了，事儿还没成，总是有个条件不够。

面试经验，面试会经常问到的问题

理论基础篇

理论基础比较重要，一般面试第一轮会考。

一、数据结构。有些公司比较喜欢考数据结构，链表结构等等，虽然工作中都有现成的工具，但是这些还是要掌握一些，需要会写一些常用的排序、搜索算法。

二、机器学习算法模型。需要会逻辑回归损失函数推导，决策树、GBDT、XGB、随机森林等模型原理。

我这准备了一些题目供大家参考：

- 1.回归模型和分类模型的评价指标都有什么？
- 2.ROC 曲线如何绘制？
- 3.什么是 word2vec？ Attention 机制是什么， bert 是什么模型
- 4.gbdn 和 xgb 的区别？
- Loss 有正则，是 GBDT 拓展，缺失值自动学习
- 5.决策树如何决定在哪个特征处分割？
- 6.bagging 和 boosting 的区别？
- 7,cnn 是如何工作的，解释一下反向传播
- 8.L1 和 L2 正则区别？
- 9.强化学习算法，比如 Q-learning 是如何计算的
- 10.时间序列算法
- 11.lgb,xgb 有什么区别
- 12.pyhon 中的数据结构有几种，分别是什么
- 13.python 中 zip 的用法

项目经验篇

面试过程中最好是自己能掌握主动权，主动抛出做项目时遇到的难点，如何去解决的。

尽量自己多说，自问自答。比如数据量太少了，我怎么怎么弄然后搞定了，特征太多了，又如何如何筛选特征、如何构造特征等等。

有个比较常问的问题是如何衡量你现在做的项目做的好坏？这一般用比较偏业务的指标来衡量，可以在 AB 测试用使用成交率、利润率、点击率等业务指标来衡量好坏，而不是用评判模型的准确率、ROC 等等测试指标。

项目中用到的模型一定要特别熟练，比如用到了 LSTM，基本的 t 三个门如何计算的都要能讲出来。数据分析能力也是很多领导比较看重的，因为有很多工作其实可能不是在做模型，而是在做策略，那么此时数据分析能力就很重要了，而且总不能一直在写代码吧，写好 PPT 就需要贴上各种分析报告。

祝各位面试成功

【面经】一个机械转行算法的菜逼应届生如何进华为

如果需要帮助，请点击关注或者回复评论，我会尽量回复大家的问题。

一、本人基本情况

普通 211 院校机械硕士，本科二本机械专业。

二、学习历程

研一研究项目课题、研一下自学 Udacity《机器人开发》课程，确定了自己的研究方向：计算机视觉。学习相关算法到校招不足 4 个月。参加两个天池和 kaggle 的比赛都是简单的图像分类。

三、校招情况

跑到上海拿到一个创业公司的计算机视觉工程师的实习 offer（不一定转正），被各各大小厂全部拒绝，很多笔试都没进，面试一家都没有（除了华为）。正式 offer 只有 10 月底拿到华为一家。可以看出是一个十足的菜逼了。

四、华为面试情况

先是投了华为的优招，做了华为的编程题和性格测试。编程题只做出了一题半，其实只要做出一题很简单的就能进面试了。性格测没过。。。所以优招没进，自动转校招了。校招他们直接联系的我进面试。面试一面是一个男性面试官问了一些简历上的项目，这个项目主要你要很熟悉简历上项目的内容。面试官很重视这个比赛参数人数与你的排名，毕竟有的菜逼笔试参数人数不足 20 队，你进前三也没用。问完项目会问一道简单的排序算法，手写，还要记得他们的复杂度，这个我之前从没学过，是暑假看别人面经时候才知道有这东西，后面突击学的。然后面试官会问你的学习历程，自己要把这个流程很熟悉，重点要突出自己的学习能力，特别是自学能力。一面结束（20 分钟）然后就进了二面了，这二面面试官就问我：“你没有实习经验啊”，然后我拿出了我的实习 offer（还没去，只是一个 offer）（所

以最好要有一个实习经验)。有说我本科不好,我说高考没考好,所以后来努力考上了福大(突出学习能力)。又问我的一个兴趣爱好,我说健身,告诉他我减肥 30 斤(突出自己的意志力)。二面结束(全程 9 分钟)。

最后拿到华为软件岗的 offer(投的算法,不过 hr 说有可能做算法,也有可能做软件,只能先给软件岗,我实在没 offer 了,就高兴地签了,16k,14-16 个月)

五、总结

华为面试,面试官对于你的能力更注重相关的项目,最好有个比赛经验,这样面试有的东西聊,然后二面会看你的实习经验,哪怕你没有实习经验也要去搞一个实习 offer 表示我有人要,会有实习经验。华为对能力要求不是那么高,有点基础就好。面试官和我强调我做的这些东西进了华为都没什么用,都要重新学。我表示我从研一开始都是自学的,这个没问题。所以除了一点的基础能力再在面试中突出自己的一些个人品格:自学能力和个人的意志力这类的就可以简单进华为了。

这就是一个被各个公司拒了的菜逼转行应届生进华为的校招面经。

机器学习之面经

今天来写一份面经，与大家分享：（这个只是一个我回答的答案，只是提供一个参考，希望对大家有所帮助，如果有改进的地方可以在下方给出评论，大家一起学（涨）习（薪））

1.这个预测怎么做的特征，为什么做这个特征，多少特征，xgb 的深度 和 树的颗数大致是多少

（1）首先要介绍下这个项目原始数据中有多少特征，分别什么含义，例如有入住日期、订单日期、订单量等等

（2）根据入住日期为 x 轴，订单量为 y 轴，绘制了一个折线统计图，发现具有周期性质，所以在入住日期上提取出月份、星期几等特征

（3）依据（2）中计算的月份、星期几等做的统计特征，包括均值、方差、中位数、最大、最小等

这里中位数比较中，因为用的 MAE 这种指标，所以中位数比较重要。

一共大概有 30~40 特征

xgb 树深度和树的颗数要结合你的数据量、特征维度等进行设置。

2.你参加这个比赛，能打到这个名次与其他选手有什么区别，或者比他们好在什么地方，差在什么地方？

这个我回答:没看过别人的 solution，以后可能会看看，至于有什么好的地方，可以介绍下特征，如何思考的，还有哪些需要改进的。

比如说利用了 NLP 的 word2vec 对 item 进行了处理。

3.为什么离开上一家公司

这个你懂的，当然是想深入发展了

4.linux 启动了好多 python 脚本，用一行命令去关掉所有 python 脚本

```
ps -ef | grep -i "python" | cut -d " " -f2 | xargs -l{} kill -9 {}
```

然后解释下思路,主要考察一下命令掌握情况,上面这么写是会有问题的,主要是 cut 命令的用法。思路! 思路! 思路!

5.能介绍下 xgb 么? 和 lgb 有什么区别?

可以结合论文,然后介绍下 xgb 是用泰勒展开,取到二阶导数

lgb 速度要比 xgb 快,使用了直方图方法、并且可以处理 category 类型的特征。

6.svm 能介绍下么?

就是推导,这个要自己看下公式,我写到一半就不让写了,可能觉得无聊了

7.LR (逻辑回归) 损失函数? 怎么来的? 为什么这么定义? 里面取 log 是为了什么?

这部分也是考察对 LR 的推导。一步一步写出来即可,前提假设为二项分布,取 log 可能是在数值计算方面因为都是小数,所有最后连乘可能就接近 0 了,还有就是通过对似然函数求对数,然后通过梯度下降进行优化。

8.如何选择的训练集和验证集?

如果里面包含时间,当然后最后几个月、几个星期或几个小时做验证集合,这个根据业务要求;

如果是不平衡分类的话,可以用分层采样;

也可以使用 k 折交叉验证。

9.如何防止过拟合的?

这个可以说一说正则化,如果深度学习可以用 dropout

还有就是 BN,怎么做的归一化

卷积操作如何做的,每一个像素和卷积核对应点相乘在求和

早停止,切出来一部分数据进行验证,当评价指标在 N 轮迭代中不在上升了,就停止(可

能早停止不会防止过拟合，我是在这里简单和面试官说了一下）。

10.训练和验证集的学习曲线 gap 什么情况？

gap 会有，但是不要太大，比如说可能会有 0.01 这种 GAP

11.针对某个项目问评价指标是什么？然后接着会问怎么在具体业务中来衡量这个到底是否好呢？

AUC, auc 评价模型的好坏，但是在有些问题上还是要看准确率和召回率，也就是在具体业务中，还是要关心准确性，例如：如果判断是否会产生购买行为，AUC 是很高，但是实际中还是关心模型判断为具有购买行为的用户到底有多少。

12.面试过程中可能无意间问到别的问题。

例如，为什么这么切分训练集和测试集合，会不会有问题？缺失值怎么处理？有没有异常值，有的话怎么处理的？

这些问题，都需要在你的业务中仔细了解一下，或者准备一下常用的方法，例如去异常的方法，均值 ± 3 倍的方差，利用分位数来处理异常值；缺失值怎么填充，可以利用历史同期的等等。

【算法/机器学习】【春招+秋招】个人多厂面经分享

本人（男，学历双 985）从三月份从百度实习回来开始，一直准备机器学习算法面试，前前后后面了不少公司，以下面经一并送上。

【春招阶段】——按时间顺序

【阿里巴巴蚂蚁金服算法实习生】

19min 简历面百度的项目介绍

CNN、LSTM、RNN 介绍下 LR

结果：挂完电话没出 20min 显示 reject

【京东北京算法实习生，销量预测相关】

33min 电面组里有十台 GPU，每个上面四块卡多分类项目：数据怎么预处理的，模型的调优以及改进，其中有没有数据不平衡的问题。百度的 DBSCAN 项目，为什么不打标签，做监督学习，DBSCAN 中的参数。最后问了 ResNet 后更新一点的网络，这个没回答上，我说了 GAN 增强对抗网络，但是好像不是之后出的。对 CNN 也有一些了解。问我对其他算法的理解程度，回答的不太好。应该说对里面的公式都推导过，每个算法怎么来的怎么工作的都明白。word2vector 怎么做的，讲一下其中的原理。

结果：本以为答的还可以，结果也是一面挂，面试官短信说不具有不可替代性。

【58 同城算法实习生】

现场面试 30min 在南大宣讲会上如何进行分词，算法原理，分词的几种方法，正向最大匹配法，逆向最大匹配法

word2vec 算法原理，输入是什么，输出是什么，这个要详细了解

LR 公式推导 损失函数 如何多分类 什么情况下用 LR 在一个有序数组中查找一个数，怎么最快 二分法 $O(\log n)$ 在一个无序数组中查找第 k 大的数字，怎么最快 堆排序 $O(n+k \cdot \log n)$ 建立堆要 $O(n)$ ，每次找出最大的要 $O(\log n)$

结果：很多基础问题没答对，甚至二分写错了，LR 损失也没说上，当时还是 too young too naïve，很明显没有通过面试。

【招商银行信用卡中心数据与 AI 岗】

酒店面试 20min 左右在项目中遇到过哪些问题，怎么解决的？对文本进行向量化，数字化，然后开始模型训练。怎么提升的准确率，你做了哪些？从 85% 到 90% 多你花了多久？我说 20 天，平均每天半天的时间。你认为你是努力大于聪明还是聪明大于努力？你的缺点是什

么？你有没有女朋友？你以为打算在那边发展？为什么选择金融？对互联网有没有考虑？
互联网金融

结果：通过面试，拿到实习资格，但选择了直通终面，因为去了其他公司实习，后续有讲。

【网易云音乐推荐算法实习生】

55min 杭州网易现场面如何去掉电视剧中的广告？详细介绍下 word2vector 介绍百度的项目，用到的聚类，为什么不用 Kmeans，和 DBSCAN 有什么区别，Kmeans 的复杂度是多少

LR 推导，推到梯度那里聊聊树，ID3 为什么相比 C4.5 能会过拟合，信息熵，条件熵，信息增益，信息增益率

CNN 的项目有没有，CNN 作文本也可以，回去跑一个 tf 的 mnist 模型说一下 CNN 的过程 Xgboost 相比 GBDT 的好处一个求概率的题，5000 个黑球 5000 个白球，每次摸两个，如果同色，放一个黑球进去，如果不同色，放一个白球进去，求最后剩一个黑球的概率。

L1 L2 正则化，怎么挑选特征的，L2 对于那些特征作用小的系数就很小么？

L1 为何能选取稀疏矩阵？写个算法吧，快排，说了下递归，结果面试官说让我用非递归实现

结果：非递归快排没写出，一面挂。

【携程算法实习生】

一面：酒店预订的五点，6:40 才开始面，现场 13min 介绍下 BSpline 算法（介绍了下业务场景，阅读的 KDD 论文，先用 MATLAB 实现，然后用 JAVA 实现的等等）介绍下 DBSCAN 算法的业务场景，介绍下 DBSCAN 的原理你的新闻多分类怎么做的，one vs rest 还是 one vs one，sklearn 里是怎么实现的

GBDT 也可以多分类，为什么你选用了 LR 新闻多分类最后只使用了准确率作为参考么，有没有层次分类，或者分层查看哪一层更准确，哪一层错误率高

结果：面试官心不在焉，明显不想招人了，未通过一面。

【腾讯应用研究机器学习——被调岗系统测试】

一面：55min 百度的 BSpline 算法，怎么做的，原理，有没有投入到线上使用，最后效果怎么样手写快排看我的 github 跟我介绍一下你的 LR 在百度做的数据分析，分析的哪些数据介绍在百度的业务如何查找一个文件夹下所有文件的大小和，用什么数据结构，我说用树，然后遍历，遍历用后序比较节约时间，最后访问根节点矩阵迷宫问题（我说了动态规划）
二面：40min 从浏览器输入一个网址，到返回的整个过程。是怎么寻找到服务器的

QT4A 的底层有看过么，是怎么实现的。我说看过，用的是源码，adb shell 的命令行，查找的控件，UISpy，有没有对控件做一些封装一万个数去排序，用什么方法比较好，随机顺序。当 N 很小时，快速排序慢，归并排序快，N 很大时，无序，堆排序快，有序，快速排序快

结果：面试中表现出不希望被转岗，未通过。

【微软苏州算法实习生】

一面 50min 稍微问了下项目和多分类系统开始写代码，写一个 buffer，带有读和写操作，写是从每次读的后一个位置开始写完后开始询问如何保证线程安全，写一下如何实现写读安全，两个读的时候不加锁（没有答上）二面 50min 自我介绍询问 java 基础，构造函数私有的情况，单例模式 protected 和 private 的区别 final 和 finally 的区别图的最短路径问题写一个实际场景的算法，trigger 和 alarm 的时间关系三面 75min 自我介绍多分类的项目，数据怎么获取的，如何进行训练开始写算法，四道
topk 要求空间复杂度 $O(1)$ ，时间复杂度没要求，有重复数字，不能改变原有数组矩阵找数 剑指 offer 原题链表找公共节点链表找环

结果：一共手撕 6 道题目，通过面试，拿到了 cong 邮件实习资格，但最终没去实习。

【今日头条算法实习生】

50min 一个女面试官 很和蔼 写了 20 多分钟的程序一个新闻多分类系统，怎么做的，数据预处理怎么做到怎么获取的数据，去掉停用词，特征维数的选取
xgboost 和 gbd 的区别，shinkage，列抽样，并行化，优化方法，XGBoost 还支持线性分类器项目都很基础，所以没有什么可以问的
xgboost 用的不都是 CART 树
DBSCAN 的原理分词的处理有没有用过跳词？？？？？分词的语料库用的哪里的其他的分词工具了解不 jieba 树模型不适合 one hot 编码
auc 是多少树模型要不要做归一化测试集怎么拿到的。。。写一个算法：找到二叉树每一层的最大值

结果：面试官说项目经验少，建议转开发，我拒绝了，当天面试结束。

【拼多多算法实习生】

一面，40min 左右自我介绍说一下新闻多分类这个项目怎么做的讲一下 word2vec
word2vec 里面有 CBOW 和 skip-gram 两种方式，他们的区别（这个没答好）说一些特征选择的方法，PCA 的原理讲一下了解 SVD 么，说一下 SVD 和 PCA 的区别（这个没答好）讲一下 EM 算法，E-step 和 M-step 都是怎么做的（这个没答好）介绍一下 LR 了解梯度下降么，牛顿法和梯度下降的区别
SGD 和 BGD 的区别
RNN 和 LSTM 讲一下深度学习了解么，讲一下 CNN 的结构介绍下 CNN 的优化方法，adam 和动量（这个没答好）
GBDT 和 XGBoost 的区别了解 LightGBM 么，说一下 LightGBM 和 XGBoost 的区别讲一下 DBSCAN 的原理如果 Kmeans 没有设置 k，需要怎么计算背包问题：了解动态规划么，一个书包体积是 V，有若干个 m V 的物体，最后让书包最重，把方法和递推公式说一下讲一下快速排序的思想，最坏时间复杂度，什么情况造成的，如何避免（三数取中）（对方是拼多多广告团队，主要做点击率预估、广告分发等，算法团队还有图像、搜索等部门）

结果：一面通过，邮件邀请二面，因为当时已经确定实习地点，所以拒绝了二面。

【华泰证券证投部大数据算法研究员】

14min 问项目，问图像的项目，如何增加鲁棒性，然后问 java 的 Integer 和 int 的区别，初始值都是啥，然后问 python 如何得到一个文件夹下的目录

sql 中的游标知道么问我的数模做了什么题目，用的什么模型。你为什么选择金融行业？你觉得自己在这个行业有哪些优势？

结果：通过，5-8 月去了华泰证券实习。

以上是我的春招经历，最终拿到了招行信用卡、微软和华泰证券三个实习 offer，我选择了华泰证券四个月的实习。以下是秋招的面试经历。

【秋招阶段】——按时间排序

【华为 fx 项目深度学习工程师】

电面 45min 自我介绍问项目：数字图像水印，实现的原理，对图像压缩之后的效果逻辑回归的新闻多分类：数据处理方面，如何调优验证码识别：怎么实现的多任务深度学习方面：验证码识别中网络的结构，自己做了哪些修改，做过哪些优化（回答了添加全连接层，batch_size, epoch 个数等）

vc 维两个矩阵同构和相似的定义

BP 环节什么最重要，回答了激活函数优化中如果遇到悬崖边怎么处理微积分中梯度的物理意义（速度下降最快的方向和下降的速度多快，一个量化值）在学校学习过哪些这方面的课程（数据挖掘，线性代数）应该再说一下 cs231 的两个正太分布的相加还是正太分布么，相乘还是正太分布么？平时看过哪些书，作者是谁？简历要对口，投的深度学习，就要主要写深度学习的项目，机器学习的岗位就主要写机器学习。查准率查全率好像说反了。。。

结果：面试官说我岗位不合适，建议优招换下岗位。

【vivo 杭州人工智能事业部推荐算法工程师】

一面 男面试官 27min 自我介绍，问了是否是保研，是否双 985 介绍项目介绍几个机器学习模型中需要注意的地方还是介绍项目二面 hr 20min 男 hr 自我介绍问学校项目和实验室做项目的区别问项目中最棘手的地方是哪里问自己在 nlp 方面的优势偏向哪个城市问我的有没有什么问题

结果：通过面试，很快拿到了录用书，然后电话讲薪资，直接给了 SSP。

【招商银行信用卡中心数据与 AI 终面】

20min 春招直通自我介绍通过谁知道的招行卡中心，学长为什么想来这里对以后的打算等

结果：通过面试，拿到 offer

【阿里妈妈联盟技术算法工程师】

电话 80min 自我介绍 Java 容器的了解，map 的存储方式，如何构建 map，有序的 map，treemap，如何扩容，n 个数插入进 hashmap，时间复杂度，每次扩容时要移动所有元素 Python range xrange 哪个是使用迭代器循环，哪个是先生成 list 再循环数据结构，介绍下链表，了解递归么，然后电话描述如何递归翻转链表机器学习，介绍 lr，数据成什么分布，loss 函数，似然函数，求对数，怎么求最优，什么方法优化，使用时的注意事项，如何防止过拟合，l1 l2 的原理，什么时候用 l1 什么时候用 l2 深度学习，了解哪些，lstm dnn 深度学习有哪些过拟合方法，了解哪些激活函数，激活函数怎么选用，各有什么优缺点详细介绍下 lstm 三个门 每个门的输入和输出 输入是什么介绍一个你所做的项目，介绍了华泰的特征选择模块你有什么要问的问题

结果：面试已过去一个多月，至今未二面，状态仍是待安排面试。

【华泰证券证投部算法研究员】

hr 面 7min 自我介绍问我对金融的看法为什么选择金融行业未来 3-5 年内的职业规划谈一谈自己的缺点

结果：通过

【拼多多算法工程师】

一面：50min 百度的项目华泰证券的项目算法题，求最小编辑距离二面：50min 华泰证券项目 Adam 和 SGD 的区别

LR 的推导交叉熵的物理含义，softmax 的物理含义

xgboost 的工程上的改进求数字数组中最长的斐波那契数列长度三面：hr 面对拼多多的看法自己在专业课这么多的情况下为何还参加了那么多的竞赛和课外活动社团活动锻炼了自己哪方面的能力，这些能力在现在对你有什么影响你期望的 offer，我提了 40w 你为什么选择来拼多多（面向工资编程）

结果：通过，已拿到 offer，但给的稍低

【百度机器学习工程师】

65min 男面试官 玄武假日酒店 1、自我介绍 2、介绍新闻多分类的项目，数据获取，特征获得，特征处理 3、百度地图实习项目，DBSCAN4、算法题一：两个语料库，一个千万级别，一个万级别，返回万级别中的词条有没有出现在千万级别的语料库中，考虑时间复杂度空间复杂度等，词典树（多叉树）5、算法题二：循环移动，一个数组，给定移动第几位，以及右移多少位，若移动到数组尾部则从头来 6、介绍下华泰证券实习中的项目，详细介绍下项目如何优化的 7、写一下 LR 的损失函数 8、讲一下 L1 和 L2 的区别 9、说一下 BP10、Java 中的 ArrayList 是链表还是数组，为什么插入和删除可以 O(1)二面，55min，很和蔼的一个面试官 1、自我介绍 2、涨停股的预测错误率，在华泰做的啥 3、百度地图做的啥，留用了为啥不留 4、数组查找题，升序数组，旋转一下，在旋转数组中查找有没有目标数字，

二分法当时又写错了一点 5、概率题，三个硬币，一个正正，一个反反，一个正反，随机取了一枚，抛出是正面，求背面也是正面的概率 6、论文作者归类，如何查找作者三面 55min1、自我介绍 2、算法题，给定一个字符串，给定一个很多子串的集合，求满足能完整无重叠的拼成字符串的所有集合 3、上题修改为求用最小子集的拼成字符串的集合 4、递归，动规，记忆搜索 5、LR 的损失函数，梯度下降的矩阵形式 6、对加班的看法 7、有哪些 offer，身边同学的 offer 情况 8、想来北京还是上海 9、提问题，想不想来度秘做 nlp，我拒绝了说想做广告或推荐 10、逻辑回归损失函数中为什么要除 m 样本数

结果：本以为面的不错，结果没有拿到 offer。

【腾讯机器学习应用研究方向】

一面 qq 语音 63min1、自我介绍 2、java 问 `String`、`StringBuffer`、`StringBuilder` 的区别，问 `hashmap` 是不是线程安全，`currentHashMap` 线程安全的原理 3、华泰的项目（特征是什么，用了这么多模型最后选了啥，`bagging`，有哪些特征，开高低收量，9:15-9:20 集合竞价大单撤单比例，对模型的可解释性如何操作，有没有做过特征组合）4、问我的博客啥时候写的 5、百度的项目（路人轨迹拟合，DBSCAN 算法，为什么选用这个算法）6、新闻多分类（多少个类，怎么做的特征，如何拼接在一起，面试官持怀疑，这样拼接词向量能到 97% 准确度？？）7、算法题：一棵普通的二叉树（不是二叉搜索树），已知两个节点，求他们最近的公共父节点）用了递归 8、智力题：一个猴子，离家 25 米，有 50 根香蕉，猴子手里最多能拿 25 根香蕉，每走一米吃一根，问最后回到家最多能剩多少根香蕉？二面 qq 视频+qq 语音 30min1、自我介绍 2、这么多模型是在哪里学的，有没有上课学（研一数据挖掘课程+自己找项目做）3、RF、GBDT、XGBoost 的区别 4、遇到的最有挑战性的项目（百度 BSpline）5、新闻多分类数据集怎么获得的，特征怎么做的 6、了解最新的一些 nlp 常用的模型么 7、算法题：一个有序整数数组， $(a, a+1, a+2, a+3, \dots, a+m, a+m, a+m+1, \dots)$ 大小为 N ，其中有一个重复的数字，编写函数实现返回重复数字。8、（这题在逗我？？最后做错了，用了 $O(n)$ 的查找复杂度，其实可以二分的，因为是连续数组）9、现在拿到的 offer 有哪些，是不是主要想在江浙沪发展。10、你有什么问题么

结果：二面估计是因为编程题想太简单了，未通过。一个多月后又被安全部门挖到简历邀请面试，不过因种种原因拒绝了面试。

华为 FX 科学家计划没通过以后就没有面试过，邀请过现场面没去拒绝掉了；网易考拉

推荐算法秋招给了两次杭州面试的机会，最后也都没去

以上就是个人所有的面经，最后只剩华泰和拼多多两家 offer，其他都已经拒掉了。希

望我的面经对大家有所帮助，同时大家可以给个建议，选择哪里比较好。

本科校招算法岗面经：拿到滴滴、头条 offer

我是七月在线集 4 学员，本科，以下是我的校招面经

今日头条

- 1、FM
- 2、EM
- 3、描述一下最大熵。
- 4、说下了解的概率图模型。
- 5、LR 损失函数及其求导。
- 6、交叉熵与 softmax。
- 7、GDBT 与 RF 的区别。
- 8、GDBT 和 XGBoost 如何处理分类问题。
- 9、误差分析 (bagging 为了降低偏差，所以 RF 中决策树用较深的，boosting 为了降低方差，所以 GDBT 中决策树用较浅的)。
- 10、加法模型，GDBT，XGBoost。
- 11、泰勒公式 XGBoost 中为什么没有三阶导数（因为泰勒展开最高二阶，高阶都是零
- 12、SVM 损失函数（合页损失）。
- 13、LR 与 SVM 的区别。
- 14、合页损失与 logloss 的区别。
- 15、了解 bert 么？
- 16、DSSM
- 17、word2vec 的优化方法。
- 18、fastText

19、RNN/LSTM/GRU

20、4~5 道 leetcode 题，记不太清了，重点考察算法复杂度优化。

滴滴

1、Seq2Seq

2、TextCNN

3、RNN/LSTM

4、XGBoost

5、RF

6、LDA topic model

7、GDBT

百度

1、编码实现 TF-IDF。

2、SVM 与 LR 的区别。

3、wordcount

小厂 1

1、深度学习是否需要交叉验证。

2、既然已经有准确率，召回率这样的更直观的评价指标，为何还需要 ROC、AUC？这样的指标意义何在？及其计算方式。

3、词向量的方法。

4、不平衡类别的处理方式。

5、MAP 是什么?如何计算。

6、数据扩充方法。

7、深度学习调参经验。

小厂 2

1、给定一个集合，输出这个集合的所有子集。

2、elastic search 的存储机制。

3、描述一下 EM 算法，并举一个例子。

4、Python 语法。

Google 面经, 已拿到 offer 哦!

原文标题: 10月8日 Google Onsite 面经+附上求职总结回报各位 已签 offer 真心感谢

我面的职位是 Software Engineer, Tools and Infrastructure, 所以开发和测试的问题都会问到

Phone interview 1: 白人小哥.给一个 Interval 的 class, 就是一个区间,左闭右开,比如 [1, 3) 意思是从 1 到 3 除了 3 的所有 integer. 让我在这个 class 里 implement 一个 method, 判断与另一个 Interval 是否有 overlapping. 第二问是写一个 method, 返回在 Interval 1 而不在 Interval

2 的区域. 一道如此简单的题...因为我前面实在太紧张了全身是汗面的吭吭哧哧...面完以后我都准备 move on 了, 看来小哥最后还是放了我一码让我有了二面.

Phone interview 2: 白人小哥.更简单了... Leetcode 原题 Plus One. 如果现在 Google 要 release 全新版本的 Chrome, 我要怎么保证这个新的 Chrome 全方位的 work? 意思就是测试些什么,怎么测试这个新版本的 Chrome,才能放心的 release 出去. 最后他还开心的问了我的名字到底怎么念,我就知道大概有个底儿了~

Onsite:

第一轮: 印度哥哥. 大早上的堵了一个多小时开过去腿都麻了,上来就开始 coding,直接蒙圈儿. 第一题, 给一个 array 比如[4,2,1,3,5],根据这个 array 现在我们能有了一个新的 array => 每个数是在原 array 里, 在它左边的所有比它大的 number 的个数,就是[0,1,2,1,0]. 题目是现在给了这个[0,1,2,1,0] 要求原 array,

原来 array 的 range 是 1~n. visit 1point3acres.com for more. 第二题, 知不知道 binary search? 但是现在 array 是 unsorted 的可是依然看做 sorted array 来做 binary search, 返回在 array 里面所有可以在这种情况下 binary search 出来的数.

第二轮: 韩国哥哥. 经典的地里出现过的 String 压缩编码解码类似题, 后悔当时看到没有好好写过一遍. 给一个 String 比如 "abcdffffffxyz", 写两个 methods, encode 和 decode. encode 就是比如 "ffffff" 变成 "7xf", decode 就是要变为原字符串. 我说 "ff" 怎么办, 他说变成 "2xf" 你不觉得更长了吗?

我才明白了, 应该是 encoded 后的 String 要比原来的短, 不然为啥要 encode, 的亏我问了这个问题... 然后又问他, 如果原 String 本来就是 "5xt" 这种结构, decode 不就无法辨认了吗? 他说很高兴你提出了这个问题, 但是不用管它, 一会再讨论, 先写吧.

写完以后他就问我如果原 String 本来就是 "5xt" 这种结构, 我 encode 应该怎么处理? 我就傻了...

因为一直觉得 encode 后的字符串长度一定要比原来的短, 所以根本想不出来他要的解法.

说了四五种方法他都不满意,

最后给我 hint 说, 要是有个 "1xt" 这样的你怎么处理? 当时脑洞大开想出来了... 其实是要变成三个 "1xt" 这种结构,

比如原 String 就是 "5xq", 就 encode 成为 "1x51xx1xq" 就好了.

但是这种方法违背了 encode 后要变短的 rule, 所以我是真没想出来..... 还讨论了好多种情况,

最后一种是 "1aaaaa" 这种情况怎么变, 我说 "1x15xa". 他说这是 6 个字符, 能不能只用 5 个?

实在想不出来, 这时候第三个小哥进来了, 韩国哥哥就过来告诉我说, 其实看做 1a 和 aaaa 两

部分 encode 就好.... 面完我就觉得跪了....

第三轮: 中国小哥. 第一个问题是测试的,比较简单. 测试 Calculator,input 就是比如俩数一个 operator, 都有什么 case, 怎么测,应该有什么预期结果或错误. 第二题, 一个 array, rearrange 成为另一个 array, 现在给了这两个 array, 求是怎么变化成第二个 array 的. 挺简单的就用了 Hashmap 秒了.... 然后问我,那现在给你原 array,也知道了是怎么变化的了,所以我们现在可以用原 array 求出变化后的 array 对吗?

但是我要 run 这个 method 好多次比如 k 次, 怎么最快能求出 array 被 rearrange 了 k 次以后的结果? 最后我就推倒出求 LCM.

面完他亲切的用中文跟我说,我是他见过面的最好的,时间复杂度最低 trade off 也说的好的.

谢谢小哥给了我信心~么么哒~

第四轮: 印度姐姐. 假装没有准备的样子现场想题目... 谢谢姐姐没有对我下死手 T T 海上有一片岛, 每个岛就是一个 node, 岛和岛之间有的连着有的没连着. 所有连着的岛是一个 Group. 求在这片海上, 包含岛屿个数最小的 group 的岛的个数,和最大的 group 的岛的个数. 就是返回两个数值, 肯定就是 int[2]嘛.

先讨论了用什么数据结构存储, 跟她说了 trade off. 然后开始写. 全程想给我挑错, 不断质疑我的代码...

还好我这一轮在高压下还是写的极其顺畅, 一个 bug 没有出现, 对她也是笑脸相迎, 躲过一劫...

第五轮: 中国大哥. 竟然中文给我面试, 也是感动哭... 第一题, 一个二维数组代表了一个岛. 周围都是海, 岛的左侧和上侧通向 Pacific, 右侧和下侧通向 Atlantic. 每个数字都代表了那

个位置的海拔高度. 现在下雨了,
雨只有从海拔高的地儿能流向海拔低或者一样的地儿. 返回岛上的分水岭的点, 就是在某个
/某些点上, 雨水既能流进 Pacific,
又能流向 Atlantic. 大哥可能也知道白板写不下,让我写纸上. 足足写了 4 页 A4 纸,当然字也
写的大...手都写疼了... 第二题,
给个 Google map, 你就测吧.... 我的 offer 效率很高我完全没想到, 5 个工作日从 onsite 到
签 offer,
真心感谢 hr 姐姐. 因为我有个 WalmartLabs 的 competing offer 正好是那天截止,
hr 的意思也是我的 feedback 很好,所以 HC 没有犹豫,也马上有组想要我, 所以 hr 加班加点
在进 HC 当天就跟 offer
team 合作把 offer 弄出来了. 这里再次感谢各位面试官对我高抬贵手, 以及
WalmartLabs.....

下面是求职总结 自我介绍一下. 毕业快两年了. Master 读的是普通学校的水专业
Information Science. 我是极其水,几乎零基础,项目都抱别人大腿, 当时没有好好学习知识,
现在别提多后悔了... 去年开始找工作的时候,
连 HashMap 是什么都不知道... Leetcode 就刷了七十道, 总拿女生不要太累了要不就这样
吧这种谎言安慰自己...
找到了现在的 SDET 工作,公司和待遇自然也是好不了. 今年三月开始奋发图强, 八月初开始
投简历,
十月十五日当天拿到 Google 和 WalmartLabs 的两个 offer
(对没错,WalmartLabs 真的不给考虑时间,当天 deadline). 总之是零基础, 靠努力学习和坚
持不懈活了下来.

如果你跟我的情况类似,希望我的经验可以帮到你~

CC150. 看了一遍 version 5. 据说新版本 v6 页数多了一倍... 还是希望大家坚持把每一章都看了,对新手帮助很大.

Leetcode. (没有做过 Lintcode 等等, 我觉得选一个刷题网站全弄会了其实就可以...)

三月到十月共刷 5 遍,订了 subscription,每遍仔细做了每一道题. 过程确实相当痛苦,每天还要上班,一有空就得刷题,周末也不得停歇.

尤其前两遍,基本都是看完答案背着写一遍的,可能也是因为我比较笨... 但是幸亏坚持下来了~ 当我做到第四遍的时候, 有一种通了的感觉,

即使拿到一道新题, 也能比较快的有思路. 这里还是希望我们找工作的朋友们好好刷

Leetcode, 弄通弄懂算法精髓, 举一反三.

不要有侥幸思想, 非牛人刷了一遍就想找到非常满意工作的真的少之又少.

Geeksforgeeks. 讲解非常清楚明白易懂, 尤其在学习数据结构和算法上面帮助很大.

它对于很多算法都会有一系列的问题的讲解, 看过之后基本对于某一部分的题目都没什么问题了. 比如 Trie,BST 的讲解, longest

common subsequence 一个系列,KMP 算法等等,我都是得益于它.

一个算法想不明白,可以先去搜搜 geeksforgeeks 有没有讲解

Data structure & Algorithms. 最基础也是最重要的部分, 千万别小瞧基础.

为什么一再强调数据结构与算法基础? 就算刷了十遍 leetcode/lintcde 等等刷题网站,面试还是会有没见过的题的.

遇到完全没见过的题该怎么办,没有强大的基础知识储备,怎么看穿这道题的本质,怎么很快有思路? 一切都要靠基础, 甚至比刷题本身更重要.

每个数据结构一定要做到彻底明白概念, 结构, 功能, 怎么用. 一定要亲自在 IDE 里面至少

implement 一遍!! 推荐书目:

我精读了 Data Structures and Algorithms in JAVA. 非常适合初学者,

每个数据结构的实现和用法都写的极其详细. 精读了每一章, 并且 implement 了两遍里面所有数据结构.

我也听有的人说精读 Introductions to Algorithms, 但是这本书感觉不是很适合我这种初学者...

如果你有基础或者是科班出身,面试之前详读 Intro to Algo 我觉得会很受用的. 刷题的过程中,会遇到很多没见过的算法和数据结构的用法.

比如说 graph, 在 leetcode里面会有用到 dfs, bfs, topology, dijkstra 等等算法, 每当遇到一种没见过的,

就脱离这道题, 去网上搜索这究竟是什么, 怎么实现的, 怎么用, 在 IDE 里面自己亲自实现算法本身, 很生疏的算法要多实现几遍.

这个过程是让我提高最大最快的一步.

Java Conception. 内推我 Google 的大牛朋友让我看 Thinking in Java 和 Effective

Java 这两本书. 虽然没有看完,但是确实是非常棒的两本书, 尤其是 Thinking in Java.

据说每看一遍都会对 Java 有全新的认识, 非常值得一看. 如果实在没有时间,只是为了面试紧急补课,

至少看会这个网站的 Java 面试题 <http://www.programmerinterview.com/>

Big Data.

今年以来,我发现几乎所有公司的面试都不约而同的添加了大数据相关的问题,就连

Walmartlabs 的 SDET 职位的面试中都遇到了,不得不说大数据真是现在一个很猛的 trend...

在面 Bloomberg 的时候就是因为大数据的问题不会而吃了亏挂了,回家以后恶补了很久...

这里推荐这个 blog,很多朋友都应该看过:

[http://blog.csdn.net/v_july_v/ ... 82693](http://blog.csdn.net/v_july_v/...82693)

我很想知道写这个 blog 的是个怎样的人,真心膜拜... 他的总结几乎囊括了所有大数据方面的知识背景,实在赞叹.

对于这个帖子里面提到的知识点,他都有专门介绍的链接,全面又方便. 如果想面试无敌的话,每个知识点都要自己多查资料弄懂,每道题都自己过一遍.

对于里面提到的不同方法要多比较, 每种方法什么时候适用, trade off 是什么都要清楚. 重中之重是 Map

Reduce 和 External sort.

Thread & Locks. 考得不多但是面 ebay 碰到了. 主要知识点: thread 和 process 区别, multithread, lock, semaphore, 对 resource 分配, deadlock, 怎么解决/预防 deadlock. 还有 BlockingQueue 和 Producer-Consumer 经典题要会 implement. 这里有几个经典问题:

[http://www.careercup.com/quest ... 87648](http://www.careercup.com/quest...87648)

[http://www.careercup.com/quest ... 96992](http://www.careercup.com/quest...96992)

OOD. 老老实实实现了两遍 Singleton, Factory, 还有 MVC pattern.

设计一个 class 应该也算在 OOD 范围里: 写过无数遍 LRU, Trie, Iterator, BST 以及变种, BlockingQueue 等等, 生怕被问到...

System Design. 这个对不住大家,我最后没面到过系统设计,所以不太知道自己这点准备到底充不充分...

如果你要面 Facebook 几乎肯定是要考系统设计的,还是得好好准备. 一定要看 FB 的 engineering blog, 看的越多越好.

基础的概念至少要会: load balancer, cache, memcache, consistent hashing, round robin, master slave, sharding, pre-computed, map reduce, difference with SQL/NoSQL.... 有很多牛人总结的系统设计帖,我就不多置喙了,这里推荐几个帖子.

[http://massivetechinterview.blhtml](http://massivetechinterview.blogspot.com/2011/03/system-design.html)

[http://www.mitbbs.com/article_html](http://www.mitbbs.com/article_1/201103020201.shtml)

[http://blog.csdn.net/sigh1988/ ... 90337](http://blog.csdn.net/sigh1988/article/details/90337)

还有这个公开课,太棒了,新手入门必备,谢谢成哥推荐~

[https://www.udacity.com/course ... 37165](https://www.udacity.com/course/cs371)

Resume. 就一点,要把自己简历上每个项目都弄熟,写下项目介绍背下来,这样被问到的时候可以张口就来.

也要把你面试的单位的简介自己总结一遍背下来,还有你为什么想来我们单位,如果你有工作你为什么想跳槽,你觉得为什么适合这个职位等等.

其实这些都是标答,只要好好准备过一次就能适用于各个公司...

这里有一个我总结的软加分项. 尤其对妹子,说实话妹子是可以很占优势的,特别如果你是个漂亮妹子~

你的性别,说话的态度,眼神,都可以成为你的加分项,一定要利用这一点.

为什么我突然说这个,不是说这只是个锦上添花的事情,而是因为这点非常重要,其实男生也一样.

一个面试官想找的不仅仅是能够做出题的人,更需要的是找到一个合适的 teammate.

你是不是好说话,是不是能聆听而不是一味反驳别人坚持自己,是不是能马上接纳别人,接受别人的 idea 并且有接受新知识的能力,从某些方面来说,比仅仅能做出这道题重要得多.

所以面试的时候,那天早上就告诉自己今天是去跪舔的,别耍态度,如果你是大神可以除外...

最好全程微笑,遇到不会的题的时候更要微笑.

把想题的过程全部说出来, 不能成为心理活动, 让对方知道你在非常努力的思考, 而且态度

很好, 所以就算你没有完全想出来,

他是非常愿意给你 hint 的. 态度决定很多事,甚至人生.

好啦,我啰啰嗦嗦了那么多真是不好意思 T T....

但愿这篇总结能给任何人一点点的帮助我就没白写~多努力就有多幸运.希望大家都能坚持

到底,不倒在黎明前,最终拿到很多大 offer,进入自己梦想中的公司,开启人生新的篇章!完。

18 应届毕业拿到机器学习 offer 的面经：嘉和美康、贷鱼科技

我是今年应届毕业生，从报名参加了七月在线的机器学习集训营，完成了 3 个月的集训营课程后，便开始求职，并最终拿到机器学习 offer。在求职过程中遇到了一些挑战，在这里与大家一起分享，希望能给大家一些帮助。

首先简历要有一定的含金量，也可以自己找项目来进行充实。

(1) 找准工作方向，然后写对应的工作经历，可以两个项目，一个 kaggle/天池的比赛；

(2) 项目的来龙去脉要搞清楚；

(3) 基础技能部分写自己比较清楚的，这部分是面试官必问的，而且会往深处问，有实践的如：SQL,编程这是还要手写。

接下来我按面试的流程给大家分享一下我面试的流程与面试过程中的遇到的一些问题以及面试之后的反思与总结。

1.公司：嘉和美康 岗位：机器学习开发工程师 行业方向：医疗

笔试加技术

(1) 开始面试前会收到一份笔试题，题目包含数据结构算法题，大数据分析常使用的工具 SPARK,以及 NLP 相关的问题。笔试数据结构这部分推荐大家刷一刷剑指 offer，其他的问题集训营的课程一般会有介绍，大家平时多练一练，课程认真听一般不会有什问题。

(2) 讲一讲项目。自己做个简短的介绍，这个项目的背景，做这个项目有多少人员，在其中担当什么角色，做了哪些工作，最后达到了什么效果。在介绍的时候面试官会看看简历，然后结合我们的介绍进行提问。

(3) 面试官开始根据项目的流程进行提问！这个和简历布局有关，我在项目标题下面提出了使用了哪些模型。面试官一般就从使用的模型开始问起！为什么选择这几个模型，这就要求大家在学玩模型之后对各个模型使用的场景，计算效率，准确率有一个了解。一般分类树

模型较有优势，分类能力强，而且可以查出分类出错的原因。讲了应用方面的知识后就是模型的原理，大公司尤其注重这一块，这一块也决定我们日后能不能自己进行模型的优化和调优。基础的要了解模型的思想，然后对公式进行推理，最后的难关就是自己有能力对模型进行建模。初学者进小公司只需会用，模型的思想清楚即可，大公司对模型的自己建模会在乎一些。

(4) 特征工程：特征工程这一块特征怎么找出来的？特征怎么筛选？特征一般需要对业务有一定的了解，将一些认识的到的重要特征列出来，对于遗漏或未知的特征用特征组合或特征筛选挑出比较重要的特征。特征筛选有一些方法：包裹型，过滤性，嵌入型这些要讲一讲。

(5) 数据：数据来源，数据量，数据预处理。数据来源我的是信息系统里面有的，直接使用，当然大家的数据来源不一定一致，大家结合自己的项目数据来源来说即可。数据量是企业比较关注的一个点，个人觉得得千万级别，大公司可能要上到百亿级别。这个数据量要咨询一下寒老师。数据预处理包括分桶，离散化，等相关处理，这个过程要讲详细一点。

2.公司：贷鱼 岗位：机器学习工程师 方向：金融

一面技术

(1) 看简历基础技能部分，然后挑一些实践较重要的技能进行口述。这一部分主要问了两个问题，挑一个较熟悉的模型讲清楚这个模型的原理？spark mllib/sql 针对这两个实用的技能进行询问，出了一道题来考察：spark 可以建立视图吗？使用视图有什么优缺点？spark 肯定可以建立视图，使用视图的缺点是计算效率慢。大家平时复习的时候与效率有关的问题要了解一下。

(2) 结合项目问特征工程，模型。数据量多大，数据来源，数据处理过程，会挑一个模型让我讲解。这部分与第一家公司大致相同，简历上的模型大家要清楚。

二面技术

这次面试的是老总，其实技术还不如一面，只不过老总阅历较丰厚，会通过人的表现来判别一个人的技术水平。问题还是哪些老问题，就是要表现的自信一点。老总更多关心能不能完成工作，给一定的压力来试探一下。能不能服从加班？能不能按时完成工作？能不能快速学习新技术？有没有对市场对业务的观点和看法？着一些考察个人综合素质。

三面人事

(1) 接受最低薪资是多少？然后讲一堆人情话，大家刚出来都是从几千元开始的，你过来又可以学习，能不能降低。有没有企业给 offer?找工作多久了？现在住哪里？以前的工作为什么辞掉？

有没有企业给 offer，这个问题是企业之间照对比。找工作多久，可以看面试者的水平，太久肯定不行，太短企业也不接受。住宿，北京住宿较贵，上班通勤较麻烦，离公司比较远一般会放弃。

最后面试的总结与心得：

1.面试看简历：

简历的项目真实，面试官一般问算法的较少，有真实项目算法只要能答出基本思想即可，有真实项目一般问比赛的比较少。我的简历两个真实项目一个比赛，面试官直接问我真实项目，最后附带问问主要技能以及工作经历。我的简历放在拉钩上面，项目写的比较详细，自己投的很少（想转 NLP）就有企业直接联系我，这个也感谢寒老师为我的简历把关。

2. 大中型公司一般会有笔试：

笔试一般包含一两道数据结构的算法题，希望进大中厂的同学要准备一下，我就是被卡在数据结构上。如果数据结构实在没时间准备可以找小伙伴大家相互帮助下（在企业笔试没有面试官监督，可以拍照给小伙伴看看，大家一起做，一起找答案），一般过了初面，这

些也就不太重要了（个人理解）。也可以自己将常考的题目自己多练习，有些公司会出 leetcode 或剑指 offer 上面题目的变种题，就是会在原题上面在加深，这类的题目要求大家对算法题了解的情况下继续加深，打算法的本质搞懂搞透（真心比较费时间）！

3. 真实项目里面的核心类容要清楚：

如样本数量，模型融合，模型选择，评估方法，最后的实施部署。项目写的好吸引力就强，样本量，服务器配置，模型搭建过程遇到哪些问题，如何解决的，找一两个亮点，这样面试官也会觉得是我们自己做过的。这个问题大家可以参加 kaggle 比赛进行迁移学习，将比赛中遇到的一些难题进行迁移，其他相关问题可以问问资深开发大佬或者寒老师，有这些背景，面试官再问也就不会露馅，也能让面试官相信我们是真实做过的。建议：大家最好自己一个人完成比赛，这样能够将项目中遇到的问题自己过一遍，知识进行融合、实践，对自己能力提升很快。

4. 项目相关信息：

主要考察面试者是否真实参与该项目。面试官一般会变着发着尝试从我们的回答中找到我们是突击学习的。项目的背景，开发团队人数，主要责任。

5. 人事：

当人事来找面试者的时候基本就是商量工资的时候，前面的面试基本都通过了，这个时候不要以为自己刚毕业或者技术不娴熟就少要，斩钉截铁直接告诉人事自己的要求。一般人事也会砍点价，面试的时候期望薪资适当提高一点。有些企业会问很多个人相关信息，学校里实习在干什么，学机器学习多久了，用 python 多久了，以前的实习工资，来北京多久，找工作多久，手头有几个 offer。这些在准备面试前，自己心里要有准备。

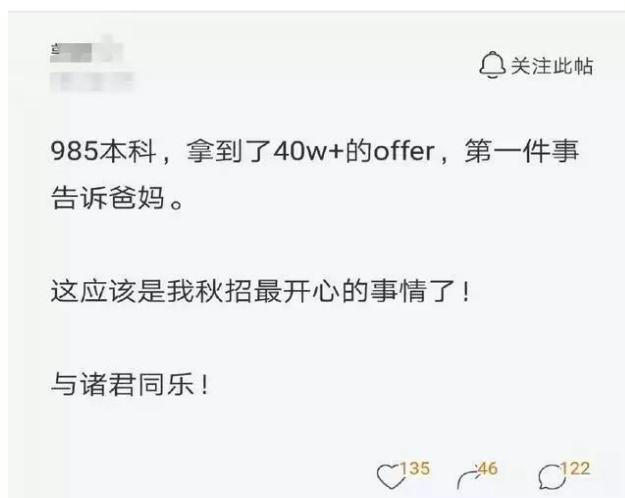
6.感谢：

来七月学习还是一个很不错的选择！有老师带，有学员可以交流，有宁哥指导就业，经

常发红包给我们鼓励，这对毅力不足的我来说是个非常好的条件。七月也提供了足够的设备（虽然 CV 不够用），但我做文本分析的还是足够了，感谢七月的各位老师提供的帮助。最后分享点学习经验：学完机器学习在进入 NLP 和 CV 要简单很多，我就是过渡时间太长，浪费了一段时间，祝大家能够找到一个好工作！

985 本科生程序员，校招拿到 40 万的 Offer，老员工被倒挂：难受！

青出于蓝胜于蓝，一代更比一代强。一名 985 高校的本科毕业生拿到了年薪 40 万的岗位，忍不住在互联网论坛炫耀起来，其拿到的是某互联网公司的算法岗位，也就是我们熟知的程序员一种，并称自己的年薪倒挂了不少老员工：**985 本科，拿到了 40 多万的 offer，第一件事就是告诉爸妈，这应该是我秋招最开心的事情了，与诸君同乐！**



此帖一出，立即引起了同行们的热议，大家纷纷投来羡慕的目光：校招 40 万倒挂不少老员工，老员工这几年的工作意义在哪？甚至还有网友认为互联网的泡沫太大，竟然给应届生这么高的薪水，其他行业只有羡慕的份。

如下是另外一位网友的观点：看到很多人说是泡沫，但实际上现在应届生的年薪是名至实归，因为基础、综合素质对得起这个年薪，几年前的毕业生根本就没有现在的学生这么刻苦，更早的 80 后更应该扪心自问，你们能保证不会经常网吧包夜、沉迷游戏？2008 年的大学校园混子居多，但这两年完全不同，毕业生质量自然和 10 年前不可同日而语。



有其它网友随之附和：因此老员工就不要不平衡了，今年这种倒挂局面还不是你们自己一手造成的？很多人离开大学恐怕就再也没有充过电。资深 HRBP 也反应，近十年的应届生质量一直在直线增长，常常超出预期。

无论技术基础、还是人生规划和视野、对劳动法律法规的等社会经验的熟悉程度、待人接物、思维能力和口才、衣品、为人处事和心智，碾压几年前的毕业生不是一点半点；不好意思，90 后程序员早已不是过去传统程序员的刻板印象，90 后平均智商远高于 80 后，技术水准碾压 80 后只是顺带的事，自然有大量空闲时间搞其他事情，比如社交，这就是 90 后码农多才多艺的原因。



不过对于这样的论调，对于这名 985 本科生的炫耀，有网友表示不服：楼上好几个吹九零后的，我也九零后，说句公道话，八零后不是智商差，只不过时代不同享受的教育资源

和社会资源不同。年纪轻轻，别这么狂，站这么高小心都不知道咋摔死的！

看到上面有人说新的应届生比老员工学得更多能力更强，我默默笑了，感情本科生就这个认识水准？人家毕竟比你多吃几年饭经验比你多人脉比你广，而且我真心不知道程序员暂时拿这个溢价的钱高兴什么，40岁以后总要还回来的。



2019 校招美团京东小米去哪儿华为面经机器学习面经

暑期参加了，七月在线的机器学习集训营，认识了很多优秀的同学，大家一起学习和交流，比自己一个月学习效率高很多，寒老师，David 课都讲的很好，宁哥的服务真的是一级棒。

总结：校招问的问题还是挺基础的，主要是根据你的简历去问的，所以把简历好好写写，不会的就别作死写上去。主要问的还是机器学习比较多，深度学习问的比较少也很基础。时间隔得有点久了，所以有的忘记了。总的来说把李航那本刷两遍问题不大。

美团

一面：

1.项目

2. xgb 相关，推导，参数与 GBDT，lightgbm 区别

3. 机器学习评价指标之类的基础问题，问的比较细，给了场景让选择指标和理由

3. 最大字数和

二面：

1.项目

2. word2vec, tfidf, fasttext

3. svm 推导，相关问题

4. CNN, RNN, 相关的问题

5. 最大合理 “ () ”

三面：

1.实习

2. 从词袋模型到 word2vec，以及 Word2vec 参数更新，两种加速方式，为什么

3. 树模型都被问了一遍

4. 一道 medium 的算法题

HR:

1.吹逼

2. 北京上海选哪, 几个部门选一个

HR 面的比较挫, 影响了后面的部门选择, 还有投递的时候意向部门不要乱填, 基本就是你的未来部门, 所以好好选, hr 别怂, 我有点怂, 所以 hr 面评价比较低, 导致被捞的比较晚。。。。

京东

一面:

1.项目

2. 实习

3. 写了个比较简单的题, 忘了

二面:

1.项目

2. 场景题

3. 好像没写题

京东面的比较随意, 不过工资给的挺高的, 比较迷。京东笔试就是摆设, 想去的话, 一定招人内推。。。。

小米

一面:

1.项目

2. xgb 相关, 推导, 调参, 不同模型的比较

3. 树模型相关，全被问了

3. 二叉树遍历非递归

二面：

1. 实习

2. GBDT 推导，应用场景之类的

3. word2vec, fasttext, 以及两者对比

3. LDA 相关

4. svm 与 lr 对比，问的比较细（整个秋招只被问过一次 LR 就是这。。。还是附带的问的）

5. 场景题

6. 翻转链表的变种

HR：

小米没有 hr 面，直接就是 hr 谈薪资

小米比较抠，低于市场价，谈薪资的时候别怂，要个市场价问题不大

去哪儿

一面：

1. 项目

2. 实习

3. 评估指标

4. xgb 相关

5. 朴素贝叶斯

6. 深度学习相关问题

7. spark, hadoop, 大数据的基本问题

8. 问了操作系统和网络的基本知识

9. 写了个统计 ip 地址并排序的题

HR 面：

一面是算法组 leader，直接就 HR 面，日常吹逼，谈薪资，现场给 offer

华为

一面：

1.c++相关问题，表示不会

2. 大数据相关

3. 聊天

二面：

聊天，没记错 5 分钟结束了，可能还没有

还有些其他的公司就不写了，问的东西都差不过，只是难度深浅而已。总的来说简历如果不改，加上适当引导，问的问题都差不多的，问你一个问题最后扩展说说，以此来表现你比别人强。

秋招今年开始的贼早，7 月份就开始，我是将近 10 月才开始的，所以很多公司都没面，所以要早做准备，提前批不要放过了，多以此机会。

应届生好好刷题很关键，有一个很有意思的事情，如果面试官想要你一般出的算法题都会比较简单，不想要你就是 3 维 DP 劝退，当然如果你写出来也会让你过的。

集 4 本科应届拿到近 20 万的 AI 面经：曾一度因是双非院校而没面试机会

1 首先再次恭喜本科应届而且是双非院校，通过三个月集训营的学习，拿到近 20 万的年薪，请简要介绍下你自己

汪同学：我是一个普通二本学校的应届毕业生，专业也是计算机大类的，之前在学校时候因为兴趣爱好自学过 python 语言，爬虫。后来对数据和算法感兴趣，在网上看到七月在线，就报了集训营。

2 面试时遇到了哪些困难和挑战？有何面试经验与大家分享，遇到了哪些典型面试题

汪同学：最大的困难就是学校不好，简历通过不了几个，面试机会太少了让人非常消极。

面试时主要能把自己的项目能说的清楚尽量说的细节一点，比如，怎么做预处理的，怎么做特征的，怎么调优这样。

再然后别人就会问你一些数据结构和算法题，这个只能刷题了（所谓四大金刚：课程 题库 leetcode kaggle/天池，缺一不可）。还有机器学习算法至少要非常熟悉几个你常用模型，每个细节都有可能是面试官要考的点。

3 你觉得作为本科应届，成功找到一份算法的工作，关键在哪几点？

汪同学：我觉得最关键的就是坚持吧，然后还有一点运气，一直没有什么面试，给人打击很大，别人岗位上都要有工作经验，没工作经验的最少硕士学历，对本科应届还是双非院校机会是少之又少，一定要准备充分不要浪费每一次面试机会。

4 在集训营上课的最大心得是什么

汪同学：集训营上课我觉得最大的好处就是可以和一群好朋友一起学习，氛围挺好，一起做项目很有帮助，在群里互相解答问题提问题。课程之外，需要自己找准方向深究下去。

5 怎么看待今年算法校招火热的情况

汪同学：就说明算法岗需求量很大，算法工程师的发展空间也还很大。

然后我自己的打算就是先进入这个行业再说，工作中不断学习加强知识储备，再有机会读个硕士，毕竟我觉得算法这东西是有未来的。

集 4 AI 面经：从后端转行 AI，薪资翻了一倍多

1 恭喜本科从后端成功转行推荐算法，请简要介绍下你自己（教育经历、工作经历）

我是一个非计算机专业的普通本科生,大学毕业后在老家做过一段时间后端开发,后来来北京找工作在一家电商工作了一年半,一共的工作经历不到两年。

从去年开始对人工智能方面特别感兴趣就自学了一段时间,其实也有一部分原因是觉得做开发受制于产品,感觉自己成了写代码的机器,这背离了我一开始想要做计算机这个行业的初衷,所以就一直在自学但是进度很慢.而且始终不知道该干什么,后来突然想到可以进行培训,就来到了七月。

2 你觉得传统 IT 和 AI 两类工作之间，最大的区别是什么

说实话因为我现在也刚入职不久,挣扎于大数据和以前没用过的编程语言,但就我的工作情况来看与传统 it 的大部分架构,框架,需求都被确定的情况下,算法确是一个更偏向探索和挖掘的感觉的东西,你可以通过各种办法来提升模型的准确度,也能让自己的模型更加接近“智能”,这种兴奋的感觉如果没有工期的要求有时会觉得不给钱我也愿意干，哈哈

3 你觉得从传统 IT 转行 AI 好转么，需要注意哪些，以及遇到了哪些障碍，后来咋克服的
编程语言如果没有基础会学的很困难(更不要提算法)。我的话碰到的障碍大部分来自于理解算法,其实也没有更好的办法,只能多理解几遍,手推公式什么的实在做不到也应该大致说出来算法的原理,不然面试会被当做外行

4 你觉得 AI 好学么，上集训营最大的心得是什么

说实话我觉得不好学,和传统 it 大部分实践相比,算法都是抽象的,特别对于离校时间长的人来说相对更困难,但是什么事都难不倒一颗想要上进的心只要你想做,对自己有信心,不用管别人说什么,你一定能做到,要一直给自己打气,相信自己,更加努力,一定可以。

5 有些人担心学不会 AI，对你 你有什么建议

来之前记得把数学好好看看,找工作之前把大数据相关的都学习一下,有些公司要求会 java,scala 我的建议是有富裕的时间最好学习一下,1 确实是技能,2 能给你加分.

至于担心学不会,我还是觉得没有什么学不会的只有你的信念够不够强,不要被困难击倒.

最后一定要记住找工作项目很重要,一定要找寒老师谈一下,最后如果想要做推荐算法推荐一本书美团机器学习实践,项亮那本书不是很落地。

工作 4 年的程序员面试经验

我是 14 年毕业的大专生，学历比较低。不过也还好，程序员这个行业还是可以的。下面说一些面试经验：

首先，感谢之前面试我的所有面试官，因为他们都是好人!!!加叹号表示重要，我大学毕业的第一次面试，是在华为面试的是一个测试的职位，当时碰到了一个大王，当时是应届生，本身会的东西不是很多，所以自己去面试还是有一些忐忑不安的。当时是上午 9 点我去的生命科学园附近，具体的地点记不清了，然后约的面试时间是 11 点。提前到了，等了很久。后面见到我的面试官王大哥，人很帅，我们聊的具体内容是什么呢？不是问问题，是聊天！我本身不是很健谈，不过还是描述一些事情的，和王大哥先简单聊聊哪里人，其实我早已记不清了，后来就聊起了学习经历，讲述了从大学入学进入实验室做的一些事情，说了一些自己做了什么，做错了什么，又喜欢着什么。将自己的经历说了出来，王大哥并没有问哪个具体问题，他问的都是我经历中的一些事情，比如在做车辆自动控制时，我做的内容，遇到的内容和解决的问题。很感谢不官方的大哥。聊完之后，王大哥简单说了说这个职位的要求，问我自己感觉如何，我说我只是一个应届生，我愿意去做，但是感觉自己还不够好，还是需要学习更多的。王大哥最后说，2 周之后来办入职，但是工资只有 4000。我说可以，其实在当时我的心里有饭吃就好。

在面试过华为之后，我很快又遇到了第二个面试，是在一个初创公司，本身公司不是很大。刚去到公司面试的时候，是徐经理面试的我，他是一个很热情的人，在和他面试的时候，感觉最重要的还是把自己能做的事情描述的很清楚，能够说出自己的好与不好，他很关照我，在面试的过程当中，他问的问题也不是很多，我记得比较清楚的问题是：

- 1、你在我们公司能做什么？
- 2、你感觉你能一个月给公司挣多少钱？

3、你觉得需要我为你提供什么？

其他的问题，我记得就不是太多了，同样我也拿到了这家公司的 offer。工资同样不高只有 3000，但是如果表现优秀可以加薪，后来我来到了这家公司，经过不到一年后工资变为了 8000。

在后面的面试也有很多，但是还是重点描述这两家的面试。原因是:这是我最早的两份面试，对于我后续的发展有着重要的作用，我本身是一个专科生，面试的公司其实也不少，但是我最喜欢的是什么呢？不是面试问题，而是面试的故事，我喜欢和每个面试官聊天，因为通过面试官可以知道他的想法和公司一部分的经营理念。

关于经营理念，就想到了曾经在乐视的面试，告诉我，公司的梦想是喝着乐视红酒，看着乐视电视。。。省略了，记不清了，每个公司都有自己的经营理念，都是不同的，我没有进过多少大厂，我进过的公司也就是普通上市公司，或者私营企业并没有了解过什么世界 500 强。可能是我学历太低了。

上面的啰嗦完了，现在简单的说一下，我的面试总结：

- 1、面试前，先了解公司，想要先去公司官网或者其他渠道去了解这家公司的情况
- 2、岗位了解，面试岗位职责对于公司的作用，哪些地方是自己可以做到的，可以去做的更好的
- 3、面试不是考试，我喜欢享受和面试官聊天，如果只是为了考试而考试，或者为了考核去面试，我不会去面的，可能会道歉离开
- 4、面试是一个相互的过程，希望能更多的了解公司
- 5、面试准备，公司相关问题，职位相关问题，项目经历。准备知识点问题，应该从项目的设计，项目的优化，项目的运营多方面考虑，然后准备细致化问题，设计方案筛选，实现难度，实现细节，基础知识掌握度，成体系，有想法的做项目，无论方案和结果是不是最好，

但都是自己可以独立去支撑项目的方式

6、无论面试如何，自己都是需要学习进步的，我从 14 年学 java，学 android，15 年学 hadoop，16 年学前端，17 年学习机器学习。我学的非常不好，但是我感觉我在进步我就很开心。

希望我的经历能够对大家有所帮助，我会的东西很少，我是个学渣，希望大家不会讨厌我的灌水帖！

20 届找实习之阿里，头条，网易，小米，搜狐，商汤，滴滴，百度面经合集

总结：

个人感觉回答面试官问题不必太官方和书面化，腾讯 sng 招实习被发现照着百度百科念。

结果可想而知。用自己的话和理解去回答就好。可能应届生会抱怨设计题，个人感觉尽量去

做，把自己平时积累的知识点，思想和能力尽量去展示就好，可能面试官对你的回答期望不

是太高。每面完，我就问面试官对我们在校学习有什么建议，无外乎，努力，多写代码多动

手，保持持续学习的心态，关注一些新技术。加油你们都是最棒的！

头条一二面（挂）

头条 1 面：

自我介绍

项目介绍

平时使用开的环境

redis 获取一个 hash 的长度

用 redis 做过控频吗

能接受转语言吗

http 状态码 301,302

手写快排

头条 2 面：

网络部分：

DHCP 如何实现分配 ip 的

发现阶段（DHCP 客户端在网络中广播发送 DHCP DISCOVER 请求报文，发现 DHCP 服务

器，请求 IP 地址租约）、提供阶段（DHCP 服务器通过 DHCP OFFER 报文向 DHCP 客户端提供 IP 地址预分配）、选择阶段（DHCP 客户端通过 DHCP REQUEST 报文确认选择第一个 DHCP 服务器为它提供 IP 地址自动分配服务）和确认阶段（被选择的 DHCP 服务器通过 DHCP ACK 报文把在 DHCP OFFER 报文中准备的 IP 地址租约给对应 DHCP 客户端）。

OSI 七层模型，每层都说下自己的理解和知道的，说的越多越好

TCP 和 UDP 的区别，TCP 如何保证可靠传输的

说一下常使用的协议及其对应的端口

Linux 部分：

I/O 多路复用

进程间如何通信

动态链接和静态链接的区别

进程的内存分布

如何查找一个进程打开所有的文件

设计模式：

工厂设计模式

单例设计模式

讲一下，说一下应用的场景。

网易两面（过,周一 hr 面）

自我介绍

hashmap

怎么优化这段代码

优化代码

两道简单算法

1、求连续子数组的最大和问题

2、走格子，从一个出发点到终点，只能向上和向右有多少种走法，连障碍点都没有

Java 如何进行垃圾回收的

你用过哪些设计模式？回答：单例，观察者，工厂，责任链，适配器

JDK 中哪些实现了单例模式？`java.lang.reflect.Proxy` 类，`java.lang.Runtime` 类，线程池。

两道设计题

第一道设计题：

比如我有个电商平台，做每日订单的异常检测，服务端代码应该写。、

第二道设计题：

比如窝网易邮箱春节要做一个活动，我网易邮箱用户量有 3.4 亿吧！但是我只允许 8 千万人参加，最后在 8 千万人里面产生一二三等奖，有一台物理机和一台数据库服务器。你怎么来设计。

二面

为什么会有内核态，保护模式你知道吗？

文件是怎么在磁盘上存储的？

有了进程为何还要线程呢，不同进程和线程他们之间有什么不同。

进程是资源管理的最小单位，线程是程序执行的最小单位。在操作系统设计上，从进程演化出线程，最主要的目的就是更好的支持 SMP 以及减小（进程/线程）上下文切换开销。

dhcp 数据包格式及其原理

innodb 聚集索引 b+树叶子节点和磁盘什么顺序相同。

阿里一面(过)

1、自我介绍

2、全程怼项目，问的及其细，包括你为什么选择这种方案来实现，具体怎么存储怎么做的，相当细，遇到这类面试官，项目不够熟，就会翻车了。

3、redis 原子操作怎么用比较好

4、消息队列会吗？

阿里二面

操作系统模块：

1、说一下并行和并发的区别

2、你知道进程吗？有进程为何还有线程？

3、操作系统是如何调度进程呢的

4、比如 linux 下我打开一个文件调用 read 函数，整个过程说的越详细越好。

5、同步异步阻塞非阻塞 IO

其他记不起来了。其实这些扩展开来就能说很多，就包含文件系统，进程管理和调度，就差内存管理机制了，我自己说了，虚地址保护模式。

网络模块：

TCP 报文格式

TCP 和 UDP

拥塞控制

超时重传

三次握手

滑动窗口

常见协议的端口号

数据库模块：

索引数据结构

索引优化

存储引擎

事务（spring 框架如何实现事务的呢）

分库分表

Java 模块：

hashmap

Arraylist 和 linkedlist

volatile

线程池,如何根据 CPU 的核数来设计线程大小，如果是计算机密集型的呢，如果是 IO 密集型的呢？

countdownlatch 和 CyclicBarrier 底层实现原理

线程状态以及 API 怎么操作会发生这种转换

JVM：

他说只问三个问题，

1、垃圾什么时候被回收

2、是怎么回收的呢

3、忘了。。。。

分布式：

1、redis

2、分布式事务解决

3、消息队列

4、分库分表

5、CAP

暂时只能记起这么多

明天醒了再补，时长 1 小时 17 分钟。你打住我知道你的掌握程度了期间这句话说了很遍。

我知道了，你不用说的这么详细。我们不在这个上面纠缠了，下一个。

阿里三面：

自我介绍

hashmap

红黑树

微信附近的人这个功能怎么实现的

项目问的很多

就聊了 20 分钟

小米一面(过，下午二面)

小米一面：

自我介绍

项目介绍

类加载机制

垃圾回收算法

G1 垃圾回收器

hashmap, hashtable, concurrenthashmap(JDK1.8)

BIO 和 NIO

三次握手

流式计算

悲观锁，乐观锁

spring bean 生命周期

进程和线程

linux 下排查某个死循环的线程

Arraylist 和 linkedlist

解决 hash 冲突的方法

跳跃表

小米二面：

自我介绍

项目

如果你的项目出现了内存泄露，怎么监控这个问题呢

一致性哈希算法

两个单链表相交，找到交点共享文档手写代码

nio 和 bio，bio 底层怎么实现的呢

select poll epoll 区别

redis 为何效率高

熟悉开源消息队列吗？讲一下

nginx+redis 集群+session 共享或者同步

问下你有面试其他公司吗？都有哪些 offer 呢

上午被他们看完博客，说一面太简单了。

明明面了一个多小时，记不起来问题了（记起来了再补）

搜狐(offer)

项目问得很细。

自旋锁

标记清除和标记整理的区别和优缺点，为何标记整理会发生 stoptheworld

mysql 索引，最左匹配原则，!= 为何会索引失效

写 sql 举例最左匹配原则

解释下乐观锁和悲观锁

信纸上写反转单链表

你是找实习的，怎么不早说呢，你过了。

其他记不起了。

商汤

让你设计一个 cache 如何设计

redis 过期策略是怎么实现的呢

mysql 存储引擎，更适合做哪种应用，

哪些情况下索引会失效

建立索引需要注意哪些情况

垃圾回收

CAS

新生代对象如何迁移到老年代对象

G1

ReentrantLock 实现原理

spring aop

hashmap

concurrenthashmap

面试官我要招一个尽快入职，主要你时间不匹配，说这句话我就知道 gg 了，不过我会把你简历留着，再招的话可以通知你。

滴滴 (offer)

滴滴一面：

很奇怪，这面试官不说话。让你自己讲，中间偶尔问下问题，这是故作高深吗？把毕生功力都展示出来了。

比如我虚地址保护模式是 80286 开始支持的，描述子里面放读写权限，地址空间，段基址，加上偏移量完成从虚地址到物理地址的转换。

CAP 理论

mysql 能讲的都讲了

Java 差不多也是的，多线程，thread，runnable，从 cyclebarrier，countdownLatch，线程池到自定义线程池（缓存，拒绝策略，核心数）。

最后说一到两周给通知。

百度

百度一面：

项目介绍

泛型

volatile

synchronized

项目异常你怎么处理的

concurrenthashmap

百度二面（过）：

项目介绍

进程和线程区别

linux 系统启动过程（越详细越好）

守护进程和僵死进程

mybatis 源码

mysql 能说的都说了

面试官说看了你博客，我对你有个大概了解，博客上面重复问题都没有问，可能面试官对我印象比较好吧！技术就怎么问，我们两都去聊其他了的，他问我怎么学习新技术，我就问他在校学习技术建议，然后就给我讲他们面 19 实习的时候，应届生还是在乎基础，设计题就是考你积累这么多基础知识，如何把有些思想和点用到项目中解决实际问题。应届生主要是基础的打造和可塑性。

连拿 4 个 offer 的校招面经-应用开发到算法工程师半年的转型

金九银十的校招已经过去，关于面试中的问题，大家都说了很多，也说得很好。所以我的重点就不放在面试问题上了，我就主要说说从春招实习到秋招面试过程中，一些大家需要注意的事项和怎么去准备。

先介绍一下自己的情况：之前基本自己做的项目主要是开发，后来临近秋招，经历了多个日夜的思前想后，还是决定走算法路线，同时，开发的基础也不能丢。所以，自己给自己招牌定下来的目标就是机器学习/数据挖掘方向。折腾了几个月，最后拿到的 offer 有京东、滴滴、链家、招行卡中心、网易等。

一般实习春招在三四月份就开始了，这个时间点一定要注意。不管你老板是否允许你去实习，你都要去实习面试！。尤其是针对在校生，一定要去看看，去体验。主要了解面试流程、面试考点以及面试中可能存在的坑，比如 hr 面等。这个过程要注意了，建议先尝试一下自己不太可能去的公司。比如，你的目标是 bat、flag，那么就可以先去面一些二线互联网公司，比如携程、58 这种。拿这些公司先练手。这些是春招面试的战略层面，战术层面涉及到面试中的具体环节也有很多要注意的。

面试过程中，一定要注意这些点：

1、会的，熟悉的问题，一定要主动往深了去说。比如问你知道哪些排序算法。说完之后顺带一句，不过大多数主流（STL 等）都用快排。因为虽然堆排序更稳定，但是快排有自己的优势。。。。又比如，问你 svm 的原理，你不要只介绍 svm 的思想。可以把 LR、SVM 对比起来说，这两个的异同、适用的场景等。然后从 svm 的对偶问题、求解方法、拉格朗日、kkt 都说一遍。然后继续延申到 svm 的线性可分和线性 svm 问题，然后再到核函数问题。核函数又可以继续拓展下去，几种核函数，各自特点等等。总之，自己会的，要主动深入下去，不要停留在表明，主动加大难度。

2、不会的问题，要表达自己的思路，表达自己思维的过程。比如问你怎么求一个无序数组的中位数，你如果不能立马想到最优解，那么可以先说出最蠢的方法，然后说当然，这个复杂度很高，我在想想有没有更优方法。在叙述的过程中，其实就相当于在帮你理思路，因为中位数就是保证大于小于某个数的个数相同，那么在叙述过程中你可能就和快排对应上了，然后接着阐述。如果实在想不到，你也在这个过程中表达了自己的思路，不至于 0 分。

3、面试过程，建议录音。因为作为学生，无论是开发还是算法，都不太可能有很深的项目背景，所以基础就显得很重要。而基础这块，基本上就没什么花样了。面试过程录音，面完之后复盘，反复听，看自己哪个地方知识漏洞。是动态规划不会，还是公式推导不记得。下来之后针对性练习，面多了之后你会发现，基本上问的都是这些问题。

4、算法的同学加一条，加强编程能力。很多做算法的同学，其实工程能力是很弱的，大多都是 python，甚至还有 matlab 的。但是工业界，数据量的规模就很大了，跟你平时玩的完全不是一个数量级。另外，这几年算法工资比开发高很多，所以很多开发转算法的，比如我。所以，光会推公式会 python 太局限了。玩玩 linux、各种数据库、玩玩 hadoop，这些都是你的加分项。

5、重视 hr 面。技术人通常都不太重视 hr 面，但是很多 hr 都是由一票否决权的。里面也有很多坑，这些去找各种面经就行了。我这边不做具体阐述，只画重点。

上面说的是面试过程中的注意事项。其实简历关也很重要。看了下时间，已经晚上一点多了，长话短说。

重点项目放前面；不相关的小项目不要写；自己了解不深的不要写；学生会工作就别写了（如果我是面试官，这是扣分项）；高质量论文加粗；gpa 如果很高 top5，可以写；各种算法比赛，名次很高可以写（比如天池、kaggle top 级别）；小比赛如果不是 top1、2，就别写了，意义不大；兴趣爱好啥的，比如你喜欢篮球还是足球，就别写了，先过技术

关，这些留着跟 hr 谈人生谈理想去；简历要有点审美，别用网上免费的那种模板，成千上万人都用一个模板，想想就可怕。

经过了实习面试的洗礼，如果能拿到你目标公司的实习机会，建议去实习，多花点时间，争取转正。如果不是很喜欢的实习，比如你一定要去 bat，结果拿了京东的实习。如果没有更好的实习机会，可以去实习，但是要注意，实习期间也要提升自己，准备秋招。

实习期间要注意，很多秋招提前批，都在暑假的末尾，也就是你实习快结束，要转正答辩的时候。这时候分清楚重要性，如果想留下来，那么专心转正吧，如果不想留下来，那么抽点时间复习知识点，准备秋招提前批。

如果不想转正，或者没有实习。记住，提前批一定要参加！！因为大厂，尤其是算法这块，提前批很重要。几乎绝大部分的 hc 都是这时候发完了，也就是秋招正式批的时候，任你再优秀，也没名额了。我就是吃了这个亏，错过了提前批，前车之鉴。而且，大部分的 sp、ssp 也都是在这时候产生。

这时候，经过了实习转正、提前批，你手里应该有些 offer 了。如果没有 offer，也不用急。在别人实习、提前批的时候，大量刷面经，渠道有牛客的剑指 offer 刷题、leetcode 中等难度等(小编注:七月在线也上线 OJ 了哦:<https://oj.julyedu.com/problemset.php>)。算法这块，夸一下七月在线这个平台，不是因为在这个平台才说他的好。算法面试大题覆盖面很广，值得刷一遍。但给平台提个意见，很多题目答案不完整，也没有特别整理过。如果把每个算法大题答案整理好，付费访问都值。

这时候，正式秋招，开始了。面试过程的注意事项和前面一样，不再赘述。如果你在正式秋招之前已经拿了 offer，且 hr 已经沟通了工资。那么你秋招就可以放开了，拿到一个 offer，可以利用你手头有的 offer，和公司去谈公司。比如 a 公司给你 23k，b 公司只给你 20k 的算法侮辱价（这是大佬口中的侮辱价，怎么就没多少公司来侮辱我呢），那么你有点

喜欢 b 的氛围，又嫌钱太少，那么你就可以跟 hr 商量，谈下价钱。请注意，如果你非常想去 b，即使 a 多了 3k，你只是略微有点不舒服，那么忍着吧。很多公司很忌讳这些，如果你去要求涨薪，会直接取消你的 offer。谈工资的前提是，你有很多个备胎！

基本上就这么多，还有些很碎的问题，这次就不再说了，已经快两点了。码农要身体健康，才有持续战斗力。祝大家求职顺利。

第二部分 深度学习

C 语言程序员的漫漫二转之路：曾报两个机构的培训班

我是个非 985/211 的普通一本本科生，非计算机专业，在毕业前一年就计划转 IT，经过一年的努力转型还算顺利；在经过了近两年的应用开发工作后，时常关注和思考行业动向和个人发展规划，权衡再三终于在架构师之路和 AI 算法之路中选择了前沿的 AI，此为“二转”。

于是在工作的闲时抽时间学习（报了一个非七月的 AI 系统学习班），在学习班将近开始深度学习项目阶段之际，正值项目完成，对投身 AI 行业有了信心，18 年末毅然辞职专心学习，计划赶上来年 3~4 月的求职黄金期；然而现实不随人意，3~4 月的尝试均以失败告终，事后分析原因：AI 系统学习班所教知识只有广度深度不够，与就业市场需求脱节，深度学习没有实际的项目（授课的讲师也只会机器学习），却给人制造一种基本会这些就能轻松入手一份月薪 15k 工作的感觉（有个拿到 10k 的 cv 岗的小伙伴因放弃 offer 而后悔多月），个人也没有明确的细分方向导致无一精通。

失败后继续回家自学，但效率低下，经常遇到难以解决的问题，后来在转行小伙伴的介绍下，谨慎考虑报了七月在线的 CV 就业 2，在讲师和助教的帮助下，我确实在深度学习方面有了很大的提升，从原理概念到推导，到训练方法和各种实用的思想和技巧；其中最深刻的感受是玩转了云计算的 GPU，自己搭过、跑过、调过模型才算是真正的入行，相关的论文要看，一手的资料内容丰富程度和博客不在一个等级。

说到面试，到十一假期之后我才开始投递简历，分小批次投，前期基本每天有改动，还会给就业老师，授课讲师及同学审批，然后修改；我只经历了两次电话面试，一个成功，一个没有然后了；所问的内容，一定是按照个人简历上的东西来提问的，按面试课程建议的顺

序写好，面试官也基本会按照这个顺序来提问；因为是电话面试，可以提前准备好话术，笔记记下关键词，或者一些可能项目细节的描述，问到就能非常流畅的答出，整个面试过程就会融洽，面试官给的好感度就会大幅提升。

事后分析拿到 offer 的原因，很大一部分归结于运气，另外在于理性的选择，如上所述个人经历的背景信息，结合面试课程老师的建议，我对就业的方向就有了明确的目标，中小型企业-深度学习 cv 岗，根据 3~4 月面试的经历我做了针对性的准备。

中小型公司喜欢问项目，问一些问题的解决方法，业务细节的描述，基本不会让人当面去写数据结构算法题（我在 LeetCode 刷到 50 之后基本停止了），所以我在学习后期做项目从头自己实现（或搬运）了代码，过程死磕了很多细节问题（从数据扩增到模型选择训练到模型压缩剪枝到部署，相关的方法都要了解用什么能够实现的程度）。

简历写项目时也尽量写上有感悟的地方（要准备充分别坑了自己），尽量给面试官营造一种我来就能干活的感觉，面过之后总结再回看，面过一两次后基本就能掌控面试的节奏了，然后只要有足够多的尝试机会（面试邀请）就完全不慌~

祝诸君好运！

算法菜鸟的秋招总结

首先介绍一下，我是上海末流 985 的硕士，非科班，只是专业跟 cs 或多或少有一些关系，本科学过数据结构，计算机网络什么的，Java 也学了，但全忘了，现在完全不记得了，研一零零散散有看一些算法的东西，但自己都是瞎看，今天看看这个人的博客，明天看看另一个人的博客。

直到去年 10 月份猛然看到了七月的集训营，因为之前有买过七月的小课，所以一看到这个集训营的课程我就跟我父母商量了想上这个课（哈哈，因为毕竟自己没钱得问父母要钱），父母听了后第二天就把钱给了我让我报名了，哈哈，在这非常感谢父母对我的无条件支持哈。

进入正题，今年秋招我开始的比较晚，最后考虑去的是四大独角兽之一的算法岗和 a 家的算法岗。在整个秋招过程中，我感受最深的是简历项目和基础知识真的很重要，虽然大家都说算法岗一定要顶会 paper，但我的经历是只要项目够清楚基础够扎实，是完全可以 pk 掉有顶会 paper 的人的，而对于非科班或者转行的人来说，项目最头疼了，但是七月无论是集训营还是单次课都是有项目的，只要自己理解深刻动手实现一下再自己加个项目背景完全就够了。关于基础知识，七月的老师总结的特别好，面试中的常考点全都讲到了，疯狂为七月的各位老师打 call。

对于校招来说的话，面试主要由三个部分组成，首先面试官会让你介绍项目，项目就得自己滚瓜烂熟了，在介绍项目的过程中面试官可能会问一些基础知识，大厂的面试官特别关注你为什么要这样做，这样做有什么好处，不这样做是什么情况，会问很多这种问题，所以大家准备的时候一定要熟悉项目，这也是对一些基础知识的变相考察。

如果你的项目跟面试官的业务特别匹配的话，面试官可能就是一直问项目，如果项目不是很匹配的话，那也别担心，这个时候就进入第二个环节，基础知识了。基础知识这一部分，其

实一般挺常规的，我投的主要是 Nlp 和机器学习的，所以一般都会问 LSTM 怎么缓解梯度消失，会问 attention，会让介绍 transformer 和 bert，还碰到过分析一下 cnn/lstm/transformer 的优缺点。

关于深度学习，常被问到的是 BN 和 Dropout，在 ML 上，被问频率最高的是介绍 gbd，对比 bagging 和 boosting 模型，对比 lightgbm 和 xgboost，xgb 相对于 gbd 有什么优点，主要我碰到的基础知识就是这些。

这些基础知识理解最重要，不要死记硬背，死记硬背面试官一眼就可以看出来，有自己的理解很重要，这一点很加分。

一般面试完前两个部分后，面试官就会让手写代码了，如果你前面表现的比较好的话，手撕代码面试官一般就不会为难你了，一般就是给个简单的意思一下，如果前两面面试官觉得一般的话，代码可能会问的难一些了，至于手撕代码的题七月的官网是可以练习的，我当时就是在这上面练习的，这一部分准备的话也不太好准备，就掌握一些常见题型就可以了，那些难的 dp 啥的没时间的话就可以放一放了。

最后，非常感谢七月的各位老师，老师们真的非常优秀也非常有耐心，正是因为有了七月各位老师的帮助，我才能成功转行并成功拿到大厂的算法 offer，也希望我的这些经验可以帮助到后面的学员~大家要有信心，只要自己知识掌握好，科班非科班其实差别不大！

深度学习的学习和面试总结：我是如何从客户端转行 AI 的

最近工作毕竟忙，七月在线的老师催着我写总结，于是终于抽出来几天的时间来写。首先感谢七月在线这个平台，我学习到了很多知识。

我是深度学习集训营第一期的学员，985 学校硕士，计算机专业。大学的时候学过机器学习和深度学习的课程，那个时候感觉学到的比较抽象，也没有去真正重视起来，没有认真去做技术实现，毕业后从事了当时比较火的客户端开发。

最近人工智能突然火爆起来，感觉客户端门槛和天花板太低，也挺后悔当时没有选择人工智能类的专业，想在公司内部转岗，但是难度太大，只好想着出去跳槽找工作。但是介于自己没有这方面的项目经验，所以一直不知所措。于是有一天我看到了七月在线的深度学习集群营，看到了里面一些项目，眼前一亮，都是实战类很贴近企业级项目，于是马上报名。

首先七月在线送了机器学习和深度学习的基础课程，我闭门在家，每天晚上下班看视频复习基础，周六日的时候去听集训营的课程。课程的侧重、编排与内容的设置，还有 seven 大叔讲的项目让我受益匪浅，我一边学习一边课下敲着代码。

关于项目和知识的讲解与总结，七月给了很多特别好的资料与案例，并且逐行讲解清楚，不明白的还可以再单独问，直到自己明白为止。

七月每期的学员，都是求变求突破的人。七月把大家集合到一起，共同讨论共同进步，这种氛围真的非常棒。如果自己一个人闭门造车，承受的压力要比这样大很多。

学习大概了 2, 3 个月后，我开始找这方面的工作，我先后面试了互联网公司，金融科技公司等。现在我把面试经验写出来供大家参考。

编程题： 给定一个数组，它的第 i 个元素是一支给定股票第 i 天的价格。

设计一个算法来计算你所能获取的最大利润。你可以尽可能地完成更多的交易（多次买卖一支股票）。

注意：你不能同时参与多笔交易（你必须在再次购买前出售掉之前的股票）。

简历上项目如何设计的；

编程题：快速排序

编程题：类似好像是求两个有序数组的中位数

SVM 原理

反向传播推导

常用的激活函数

pooling 池化的作用

为什么 L1 和 L2 正则化可以防止过拟合？

Relu 和 sigmoid 相比的优势

你写的项目里会比较很细致的问，比如为什么用这个模型，数据量多少，怎么做验证的等等。

很多基础都是七月在线上的那些面试题。

我大致就记得那么多，希望能够帮助大家，祝福大家都能找到好工作。

深度学习社招面试心得：算法+CV 基础知识+相关项目经历

我是 CV 就业班第一期的学员，报名的时候是一名已经从澳洲毕业的硕士毕业生。由于在研究生阶段学得比较杂，对自己的职业规划也不是很清晰，所以在秋招的时候不是很顺利。在研究生阶段上过机器学习和深度学习的课程，毕业论文也是做图像分类的项目。所以项目经历都是一些课堂作业和一些比较简单的开源项目。只了解一些机器学习和神经网络的知识，对目标检测，图像识别，分割这些算法完全不了解。自己试着在网上找些 Udacity 上的深度学习课程学习，但是都是些比较基础的知识，没有比较前沿和工业级的项目。

于是就下定决心上七月在线报班跟老师学，4 月份报名，学习了两个月，主要是看视频直播，跟着老师做 kaggle 项目，修改简历，最后模拟面试。到了 6 月中旬有两个项目经历就开始在 BOSS 直聘上投简历。在课堂上跟着老师做了一个肺炎检测的项目。除此之外，由于本人对自动驾驶很感兴趣，所以做的项目都与自动驾驶相关。例如，道路交通标志检测，道路语义分割和道路车辆检测等。

下面就介绍面试公司的一些情况：

虎牙直播：

电话面试：

- 1.自我介绍
- 2.介绍你的项目
- 3、focal loss 为什么可以解决正负样本不平衡的问题？
- 4、你是怎么对两个模型进行集成的？
- 5、样本不平衡有哪些解决办法？

间隔了好久没有面试，所以有些紧张。所以回答得不好。挂完电话，邮箱就收到了拒信。

Momenta 自动驾驶公司 笔试挂了

主要考 c++ 和 python 基础

例题：

1. Python 解释器 [3]*3
2. 比较两对象的内置函数、
3. STL 容器
4. 5 mod 3
5. 虚函数
6. 函数没有 return

编程题：

1. 求一个数的质数之和等于这个数，如 10，可以分成 (5, 5)，(7, 3)，有 2 对质数，所以程序返回 2。

2, 缺失的括号

一家自动驾驶公司：

深度学习笔试

1. 模型的评价指标
2. 解释一下反向传播算法, Pooling 层的作用
3. 解释 ResNet 和 Inception 的结构
4. 解释一个你熟悉的机器学习算法
5. 模型训练的优化方法

C++ 笔试题

1. 定义一年有多少秒
2. 定义一个指针函数

3. 编程题：找出数组中第二大的元素，时间复杂度为 $O(n)$ 。

一面：

介绍项目经历，Faster R-CNN 的原理，YOLOV3 的原理。由于编程题是用 Java 写的所以问是不是不会 C++。我就回答，不会。

二面：

觉得题目做得怎么样？我说：一般。介绍项目经历。然后给我看他们做的道路车辆检测的视频，问我能不能做到这个水平还能不能优化。

总结：去自动驾驶公司一定要准备 c++，必考。

一家无人机公司：

笔试题

1. 决策树的剪枝
2. 逻辑回归的损失函数
- 3、 数据降维有哪些方法
- 4、 KNN 与 K-means 的区别
- 5、 数据集中，特征怎么选择？
- 6、 如何解决数据不平衡问题？
- 7、 模型的评价指标有哪些？
- 8、 什么是梯度爆炸和梯度消失？
- 9、 深度学习的损失函数有哪些？

一面：

介绍项目，RetinaNet 怎么处理正负样本分布不平衡的问题？

YOLOv3 的原理？

segNet,FCN,U-Net 有什么区别？转置卷积的原理？

空洞卷积？

训练时梯度爆炸，怎么解决？

小物体检测效果不好，怎么优化？

Batch size 的大小对模型训练有什么影响？

如何改善训练模型的效果？

为什么要做数据归一化？

二面：

介绍项目

怎么检测倾斜的车牌，怎么检测直线

手写快速排序

三面（人事）：

这段时间在干什么？能接受有时加班到 8 点吗？期望工资多少？

面试完第二天收到了 offer

经过 4 个月终于找到了工作，可以开始进入下一阶段了。

总结：

- 1、了解自己。知道自己想什么，自己能做什么，提前做好规划
- 2、为下家准备项目。例如我想进入自动驾驶公司，所以我做的项目都是与自动驾驶相关的。
如果想进入医疗公司，就做些医疗检测的项目。
- 3、展示项目。用 PPT 介绍项目或者视频展示，这样比只靠嘴巴说，更生动也更有说服力

学渣抓住时机进入大厂的心路历程

哈哈，先要说史老师的微信头像真好看。

每个人的天赋和学习方法都是不同的，所以我这里献丑了，将我自己的学习方法和经历写出来，也算给自己的过去一个交代。

1.人的天赋

以下是从原则一书中学习到的一些关于人的天赋概括，个人认为人的能力要顺应个人天赋所在来进行有意识的自我培养，大体上分为 7 个天赋，我认为在计算机技术领域，初期的技术提升是需要一些数学天赋的，如果想在技术领域有所建树则非常需要数学天赋，在职业路径中认识自己非常重要。

语言天赋 / 数学天赋 / 空间感天赋 / 运动天赋 / 心智天赋 / 交际天赋 / 环境天赋

在思维方式上，大概分为两种。

线性思考者，左脑发达，“聪明”的人具有高智商，是非常善于分析的思考者，能够解决复杂的思维问题。

横向思考者，右脑发达，“机智”的人进行整体推理，把握主题并擅长综合分析。

在对于单一事件的处理上有以下 7 个方面

细节-全局 / 记忆-思考 / 工作-任务 / 规划-观察 / 情感-理智 / 风险承担-风险规避

内向-外向 （对于真相的探寻意愿）

2.经历

一、安卓逆向工程师 -2015-2017

二、计算机视觉算法工程师 -2017-2018

三、创业 -2019

3.选择比努力更重要

在第一部分中写了个人天赋，在求职过程中，也需要有和个人天赋相配套的策略。

我善于分析行业情况，进行全局思考。

在 2015 年的时候，在辽宁上学，还没毕业，需要进行实习，学校给了我们两个选择，一个是去大连中软进行进一步的学习，另一个是去北京直接在公司里进行实习，我在学校的时候每次考试都考倒数几名，连专业课都经常挂科，我就想都没想就报名去北京直接在公司里面进行实习，这么果断的原因就是因为我知道自己不是那种在课堂上学习的人，即便去了大连又如何？我也并不擅长考试，对于记忆类的东西很烦。

在选择公司的时候我知道对于我这种几乎一点编程都不会，连面向对象的三个特性都不知道的人是不可能面试成功的，所以我去面试了一家在当时移动安全领域很厉害的公司的测试岗，我当时的想法是，先进入这家公司，从事技术相关的职位，以后再通过努力获得成长。而我也真的如愿以偿在这家公司从测试做起，后来转岗到逆向月薪是 5k 了。

在 2016 年，我跳槽去杭州工作了半年，跳槽后的工资是 10k，我跳槽的时候也是选择去了阿里的外包团队，因为我知道以我这种三流学校毕业的人，在技术上又没有什么优势，只能先进入外包项目先做着，然后再寻求机会，可惜当时人工智能出来了，我知道安全行业在不远的将来，不会再需要重复性的去分析病毒，我知道 AI 算法才是未来，所以我辞职回到了北京。

在 2017 年的时候，我自学了一些简单的 Python 和数学知识，我仍就知道自己的深度学习的算法方面的技术很薄弱，我就去面试了英特尔的外包团队，想要像以前一样，先从技术相关的职位做起，然后慢慢的成长。在英特尔看到算法的神奇，也看到研发算法的艰难，当时一起共事的都是硕士、博士，我对自己的积累深深地感到羞愧，非常感谢在英特尔工作的时光，让我学到很多工作的方法，比如做事情一定要形成闭环，必须要分清哪件事情最重

要，我觉得这些是我毕业三年学到的最重要的东西，当时的同事真的都很优秀，真想和这样的同事一起工作一辈子。请允许我不透露当时的薪资。

可惜时不我待，奔三的我，面临着买房、结婚、生娃、我不能继续我的技术梦了，不然我真的很想就那么上班，我害怕自己 35 的时候就被辞退了，之后我选了创业。

可能是我这个人运气好一点，每一次都能进入在某个领域独当一面的公司，其实是和时机相关的，我在进入这些领域的时候都是行业刚刚兴起，或者特别需要人的时候，在选公司的时候，一定要尽量选择大公司，因为大公司的做事方法和对于人员的培养不是小公司可比的。

4. 个人财务的发展路径

借用货殖列传中的一句话，无财作力，少有鬻智，既饶争时，此其大经也。

天下熙熙，皆为利来，天下攘攘，皆为利往，个人认为追求财富没有什么可以害羞的，司马迁说在没有钱的时候出卖体力赚钱，稍微有一些财富的时候就用智慧去赚取钱财，如果有大笔的流动资金就可以用时间和时机去赚取财富。

在没有钱财的时候，也需要用智慧去判断时机，来让自己获得更好的成长和更快的晋升。

我是如何通过集训营在公司内部转岗拿到年薪近 100 万的

我是集训营的学员，通过集训营在公司内部转岗涨薪拿到年薪近 100 万。

在中国一线城市做互联网相关工作，年薪百万其实不难，人人都可以做到，这里圈点几个重要步骤，让大家少走点弯路。

第一步，先进一家大型知名互联网公司（BATMD 等都行，市值有几百亿美金+，名字大家都知道，职级在业界有认可度，就行了）。学校不太差的话，校招刷刷题都可以进。学校差则可以先选择一家小公司积累点经验，但目的性一定要强，别去了就傻傻的做需求，你要有针对性的积攒经验。

不管是需要刷题的校招同学，还是需要刷经验社招同学，选择一家靠谱的培训机构（比如七月在线，就业针对性非常强，老师也都非常专业），都会使你事半功倍。不要怕花点学费，你的时间远比那点学费宝贵。

第二步，进入公司之后先大致了解一下公司的职级体系、晋升制度，脉脉上有很多相关的内容，没事可以多刷一刷（脉脉负能量也挺重的，大家要提高心理素质，不要被带节奏，几乎在所有人眼里，自己家公司都是问题多多、槽点满满）。

对于基层程序员晋升之路，其实各家公司都差不多。大家进去了一边做需求，一边学习积累经验（千万别只做需求，需求有人推着你做，但学习积累是自己的事，自己要规划好）。在这个过程中，要注意自己在老大和团队的 visibility。怎么理解这个 visibility 呢？有很多同学进入一家公司后很长时间，所做的事老大都不太清楚，同事们也不太清楚，大家感觉这个人存在感有点弱，这就是 visibility 不强。当然你 visibility 也别太强，这样也会引得同事们不爽，这个程度要自己把握。我个人经验是，一是要找机会参与团队内重点项目。这种项目往往是高 T、高经直接带，并且经常要向上汇报。你平时自己做 N 个小需求不一定有人知道，这种重点项目，会大大放大你的产出（当然你的缺陷也会被放大，一个致命 bug

可能就要去准备简历了)。如果没有这样的机会,也没事,可以自己留心创造机会。比如在你平时做需求过程中,一定会有一些需求,你感觉可以做出亮点、做出超预期的东西的。你就挖掘这种机会,一旦有,就围绕这个需求,设计一套解决方案或架构方案,然后先找团队内自己关系比较好的高工 review 一下,看看是不是靠谱,如果靠谱就一起再往大范围推。一般大公司都有技术评审制度,评审时多是高 T、高经甚至总监一起参与,你能去讲讲方案,visibility 一下就上去了,如果你的方案很棒,能立项并漂亮的完成,那晋升就很轻松了。以上是一些个人经验吧。一般我们去到一家知名互联网公司,不要太在意起点职级和起薪(牛博、ACM 大牛请绕行,你们要争取 P7、P8、P9 的起点,我们和你们不是一个星球的)。一般人从普通应届生为起点,做的比较顺利的话,到年薪百万左右的级别,5 年差不多就能达到了---基本就是阿里 P7、腾讯 T3.1, 百度 T6 这个范围吧。有些大公司压薪比较严重,也见过到了这个职级也只有五六十万的,没关系,一跳槽也就百万了。这个一般人都可以做到,再往后的晋升之路则会有些难度。当然,到达这个“大众职级”后选择跳槽也是不错的时机,不怕苦不怕累的话,跳到头条、PDD 之类的公司,面试面的好能拿到 150 左右。

大概讲这些,在中国互联网圈子,自己多用心,百万年薪不是很难达到,人人皆可。当然有些同学说,“我完全没 follow 你说的这些,我就一毕业加入了一家做黄段子社区的公司,然后这个公司稀里哗啦的就估值 750 亿了,我一掰手指头自己也有上亿身家了,哦我说的还是美金。”那请您参考标题上的“创业的请绕行”,技术好可以学,思路好可以学,命好咱学不会啊!咱们讲的是大部分人可学、可 follow、可复制的方法。另外再强调一下,在一些关键节点,找一家靠谱的培训机构帮助你。当然所有的事情你都可以自己做,也都可以做到,但一家靠谱的培训机构能让你快速抓到关键点,少走弯路,节省你最最 valuable 的时间。

非科班生拿 BAT 算法 SP offer 面经（阿里巴巴，腾讯.....）

> 拿到阿里巴巴算法 offer 评级 A, 腾讯算法大 SP offer, 华为 offer, 还有几个小公司的, 百度说转开发, 就放弃了。现在很多非科班的学生想往 IT 或者算法这个方向转, 但又不知道非科班的身份是否拼的过计算机专业同学, 那么我现在告诉你, 完全没问题, 只要你愿意花时间学习, 照样拿高薪 offer。

0.本人履历

已经说了我是非科班生, 也不是通信、电子那种擦边专业, 楼主本硕学的是能源与动力工程, 本科 211 硕士 985。本科没有任何代码或者算法基础（本科学的 C 语言不算）, 从研一下学期开始学习算法, 学习了整整一年时间, 拿到自己满意的 offer。

1.阿里巴巴（实习和正式 offered）

春招的时候面试的实习生, 面试通过了, 最后也顺利转正（评级 A）, 下面我说下实习面试的一些问题。

1. 一面, 三月初投的简历, 大概一周后接到面试电话
 - 介绍一下你自己（介绍自己学历、项目、自己做的小 demo）
 - 简历上面写的 blog 是你自己开发完成上线的？（是的.....）
 - 看到你参加了阿里天池的算法比赛, 简单说下比赛背景和实现方案
 - 你觉得和其他队伍相比有哪些差距？
 - 现在让你继续做你会从什么地方改进？（特征挖掘方向说的）
 - 机器学习常用的一些方法, 以及原理
2. 二面
 - 介绍自己

- 做过什么项目，用的什么算法，请简要介绍
- 算法是否看过开源的源代码，或是否自己手动实现过？
- 有没有工程性质的项目，简单介绍
- 你觉得机器学习算法在阿里巴巴哪些场景中可以落地？

3. 三面（代码面）

- 介绍一下你自己
- 代码面试，第一个为排序题目，我用的插入排序，问是否可以改进；第二个将字符型数字转为 int 类，不用工具类。
- 脑力题：国王有一万桶酒，其中一桶为毒酒，喝下后 24 小时死亡，可以无限使用死刑犯试酒，请问 24 小时最少需要多少犯人，怎么喝？

4. 四面

- 介绍自己
- 做过的项目
- 对风控课精准营销是否了解？一般用什么方法
- 说说你自己本身的特质，觉得有什么优点
- 你不是计算机专业的，算法的知识是怎么学习的，怎么权衡自己专业和算法学习，是否有影响本专业的学习？
- 你有什么想问的？

5. HR 面

- 介绍自己
- 你觉得你和周围的人项目有哪些优点
- 和计算机专业的相比优势和劣势

- 对风控业务是否了解?
- 是否知道阿里巴巴最近相关产品动态
- 你有什么想问的?

阿里巴巴面试周期非常的慢，等的非常着急。

2. 腾讯 (实习 offered)

拿到了腾讯的实习 offer，由于去了阿里巴巴，所以放弃了，自己觉得有点可惜。

1. 一面

- 从学业、项目等方面介绍自己
- 会什么语言，什么语言用的比较多
- 之前实习的经历，主要做什么工作
- 详细介绍算法比赛的经历，介绍打比赛时候的特征处理、算法建模等，以及为什么这样处理特征，有什么好处？为什么选择这个算法，优点是什么？你觉得和其他队伍的差距在哪里，有什么地方是需要改进的
- 实习期间的目标，希望做一什么工作
- 反作弊和用户画像你比较喜欢那个方向？

2. 二面

- 详细介绍自己
- 介绍机器学习以及深度学习方面学的内容，机器学习和深度学习有什么区别，各自的重点在哪里
- 项目介绍
- 介绍图片分类 demo，怎么对数据集进行的处理？
- 一个物品正方和反方，分类器是如何判断属于同一类物品？

- 卷积相比于神经网络的好处
- 计算机视觉有哪些解决梯度消失和梯度爆炸的方法
- BN 原理
- 数据不平衡处理方法
- 能实习多长时间

3. 三面 (HR 面)

- HR 面比较随意，问一些性格的内容，能实习多长时间.....
- 能实习多长时间

3. 腾讯 (秋招 offered)

腾讯给的大 sp，有的人交 SSP

1. 一面

- 介绍自己
- 有哪些常用的机器学习方法？
- 在比赛中主要用了哪些方法？
- 说说 GBDT 和 XGBoost 的联系和优缺点？（这个一定要回，高频问题）

2. 二面

- 介绍自己
- 项目中用大的机器学习算法
- 会哪些语言，常用哪些语言，会 C++ 吗（感觉面试腾讯的话 C++ 得会）
- 场景题：王者荣耀中怎么给用户匹配对手，让玩家感受到挑战，同时差距也不是特别大。
- 如何运用算法检测作弊用户？如何检测恶意言语？如何检测恶意评论（同音不同字怎么处理）？

- GBDT 和 XGBoost 的联系和优缺点?
- 是否有其他 offer, 是哪些公司的, 怎么考虑手上的 offer?

3. 三面

- 简单介绍一下自己
- 前面的面试官已经跟你聊了不少, 你对我们这边的业务有什么了解, 有什么想法?
- 你手上也有其他 offer, 如果拿到腾讯的你怎么考虑?
- 你有什么想问我的?
- 聊工作内容

4. HR 面

- HR 面比较轻松, 介绍自己
- 说说你身上的几个特质
- 你不是计算机相关专业的, 你是怎么学习的呢?
- 年底能提前来实习吗?
- 期望薪资多少?
- 你有什么想问的?

4.华为 (offered)

华为是参加的秋招, 没有走提前批。

1. 笔试

- 华为的笔试有点忘记了, 有性格测试, 还有行测那种题

2. 现场一面

- 一段 C++ 代码, 查看有什么问题, 比较简单, 虽然没学过 C++, 但是能看出来, 在 C++

中 char 类型为一个字节，这种会无限循环。

```
c++
char a = '1'
for(int i=0;i++;i<sizeof(a))
{

func()//一个看书
}
```

- 会不会 C++ (不会)
- 用什么语言比较多 (python)
- 生产环境中要将机器学习算法上线用 Python 是不行的，速度跟不上，需要用 C++ (我表示会学习，并且学的很快)

3. 现场二面 (二面至少等了 2 个小时，人太多了)

- 介绍自己
- 介绍自己学习的计算机视觉的内容
- 做图像定位项目中运用的工具和方法 (用了 opencv 和 tensorflow)
- 那你是用了开源工具，创新点在哪里？
- 那你还是用的别人的东西，只是做了组合，没有创新点 (/手动狗头)
- 计算机视觉中 object detection 是怎么实现的，有哪几种常用的方法？
- 详细解释 Bounding Box Regression
- 说说自己的优缺点

还拿到了一起其他公司的 offer，薪资都不错，也都是算法岗位，觉得都挺好的，把论文搞完了细写

学习路径

1. 研一下学期，大概 2017.4 开始学习机器学习，学习了很多七月在线的课程，讲的不错，自己的理论知识基本是从这个课程中学到的，不明白的就查网上的技术博客，这个视频看了好几遍，自己总是忘。看完后会对给的代码进行复现，自己敲一遍(不止一遍)。

1.1 传统机器学习学完后开始接触深度学习，主要是计算机视觉方面的东西，吴恩达老师深度学习的课讲的非常好，每节课步长，但是总结的很到位很容易懂。把这些学完研二上学期也基本结束了。

1.2 研二下学期导师项目上的事情比较多，非常忙，只能抽空学习自己的东西。主要方向是加强代码能力，学习 TensorFlow 框架的使用，毕竟代码才是生产力，。2017 年底觉得做机器学习算法都是在海量数据的环境下，所以觉得学习一些大数据的知识还是很有必要的，开始自己根据教程搭建 Hadoop 平台，学习用 scala 编写 spark 机器学习代码，实现一些简单的数据处理和个性化推荐系统等。

1.3 2018 年初的寒假开始准备实习面试，毕竟这行有大厂实习经验对秋招是及有好处的。根据《剑指 offer》学习数据结构的面试算法题，梳理机器学习的基础知识。2018.3 开始投递各大厂的简历，拿到了阿里巴巴、腾讯、等公司的实习 offer，百度让我转开发岗，不想搞开发，然后就放弃了。

面经写的比较匆忙，面试的回答基本没怎么写，希望后面有时间补充。

我的算法岗校招面经：微软、谷歌、阿里、头条、地平线、网易游戏等

基本情况

收割情况：一共投递 8 家，收获微软、阿里、网易游戏、地平线 4 家 special offer，今日头条转岗 offer，腾讯没给面试机会，摩根士丹利面试冲突放弃，谷歌等待 offer 中。

个人情况：某 985 硕士，常用语言 C/C++、Python，研究方向为深度学习、计算机视觉，有论文，ACM 等算法竞赛和数据挖掘类竞赛获奖经历，项目众多。有算法岗的都是投了算法岗，有深度学习的就投深度学习算法岗，有图像的就投图像深度学习算法岗。

阿里巴巴

实习面试过了，校招直通终面，8 月 16 日通知我面试安排在 8 月 17 日，且面试日期不可修改，没能去参加摩根士丹利的直通车，不知道有没有上黑名单。

终面技术面

8 月 17 日，45min

- 聊实习，聊项目。
- 算法题：2 乘 n 的格子用 1 乘 1 和 2 乘 1 的砖块盖有多少种方法？典型 DP，写了转移方程，面试官就说不用写代码了。
- 现在的搜索技术很少上深度学习或者说很深的网络，你觉得是为什么？如果要用深度学习，你觉得应该往哪些方向思考？deep learning 比较吃资源，不太适合业务规模比较大的系统，比如：双十一的压力，如果一定要用，可以考虑深度模型压缩，量化，矮胖网络的并行计算等方向。
- 聊聊人生理想。

HR 面

8 月 17 日, 45min

- HR 先介绍了淘宝搜索这边的基本情况。
- 问了 ACM 和论文经历, 平时都怎么训练, 怎么做科研。
- 为什么拒了阿里的实习 offer?
- 校招拿了几个 offer? 选择 offer 的时候你更关注什么? 成长性, 相关性, 还有...当然是薪水啦!
- 阿里比微软更能让人快速成长, 安利一波。

结果

8 月 30 日, 拿到 offer 意向书。

10 月 17 日, 拿到 special offer, 薪资很高, 但考虑其他家给出的工资, 加班情况, 放弃了。在得知我的动摇之后, 阿里派了一位很有资历的前辈和我聊了很久个人的发展, 他并没有一味的站在公司的立场上劝我加入, 只是和我分享了一些自己的个人经历, 并客观对比了我所纠结几家的优劣, 我觉得分析的还是很有道理的, 没有一味吹嘘阿里, 也说了一些缺点, 让人感觉很真诚。

今日头条

头条的面试时间安排灵活且集中!!!

笔试

8 月 12 日, 5 道简单、中等难度的算法题, 做了 4.5 题。

一面

8月18日, 45min

- 聊实习, 聊项目。
- 算法题: 上 n 个台阶, 每次一步或两步, 求走法数量, 简单 DP。
- 算法题: 上一题变形, 最多只能连续走 m 次两步。依然 DP, $dp[i][j]$ 表示到第 i 级台阶, 前面连续走了 j 次两步的走法数量。写出转移方程, 面试官没有让写代码。
- 写一个快速排序。一紧张, 左侧大元素和右侧小元素交换没写, 可能面试官也看出来我没有背过, 告诉我不要紧张...
- sigmoid 和 softmax 的区别和联系。sigmoid 二分类, softmax 多分类, softmax 在二分类下退化成 sigmoid。
- 分类一般用什么损失函数? 写交叉熵的公式。

二面

8月18日, 45min

- 聊实习, 聊项目。推导了我在项目里用的做过改进的 LSTM 的公式。
- 针对实习提问题: 为什么要加入 reading-gate? 为什么不直接把控制向量加在输入中?
和具体实习做的东西有关, 在面经里就不多说了。
- 深度学习里的正则方法有哪些? 数据加噪, 增强、范数正则、dropout、集成学习、提前终止训练、多任务联合。
- weight decay 和范数正则有什么关系? 这个当时我没想到, 后来下来一查恍然大悟, weight decay 的系数其实就是 L2 正则前面的 λ 系数, 求个导, 梯度下降的时候和 w 合并一下就看出来了。
- Mask RCNN 做了哪些改进? 分割和检测任务联合; 多个二分类 mask, 避免类间竞争;

ROI 对齐，但是我记得不是太清楚了，说了用求周围均值的方法填充边界，后来面试官提示了才想起来是插值。

- 算法题：最大子串和问题。DP，写完代码，面试官说，用的 c++11 写法太多了，他不太熟悉 c++11，正好时间又不够了，是 $O(n)$ 就行吧。

三面

8 月 18 日，45min

- 聊实习，聊项目。问了和二面第二个一样的问题。这个问题确实是精髓所在。
- Mask RCNN 和二面第 5 个问题一样。
- NLP 和 CV 结合在头条的产品里可能有哪些应用？已经实现和未实现的都可以。必须 image caption 啊，虽然很难，但是在头条的新闻和抖音都有很大的用途的。然后聊了一些前沿的方向。
- 更喜欢做 CV 还是 NLP？更喜欢做产品还是研究。

面试官说后面有可能会让 CV 背景的同事再面一轮。HR 打电话，反馈结果不错，等通知就行了。

结果

9 月 3 日，问我是否考虑转安卓，IOS 开发？如果同意的话，不用其它面试直接发 offer 了。
我??? 不转...

9 月 7 日，又问我是否想转，承诺薪资不会比算法低。可是我实在不懂安卓和 IOS 开发啊，怎么转？没有下文...

网易游戏

猪场食堂还不错。

笔试

8月12日，线下笔试，综合性较高，考察了传统机器学习，深度学习，强化学习，分布式计算，编程基础，数学基础。做完一脸懵逼，听说均分只有30多分？

一面

8月22日，45min

- 聊实习，聊项目。
- 讲讲 LSTM 和实习项目中做的改进？
- 有一个多行的超大文件，内存装不下，如何中找出相同的行？我想了一会儿，用哈希，面试官点头，应该是对的吧。
- python 中，线程，进程，协程区别。线程，进程好说，协程我没怎么用过啊，python 写实验代码差不多了，老实回答：不太清楚协程。
- 我看你本科是学的信息安全，你能跟我讲讲 SSL 吗？差不多都忘了，讲了一些基本原理和中间人攻击的东西。
- 计算机网络怎么样？学过，忘了很多。然后就没有问了。

这充分说明不会就是不会，不要硬扯。

二面

8月22日，45min，面试官是个研究员

- 聊实习，聊项目。
- 推一遍 LSTM，用公式写出实习项目中做的改进并说说怎么用 tensorflow 实现的。推了一遍公式，实现是继承了 RNNCell 魔改，大致写了一下类的设计。

- 聊我的论文。
- 说说 DC-GAN。
- 说说 Faster-RCNN。
- 说说 BN 和 LN。我把 LN 听成 LRN 了，后来想再说的时候，面试官说不用了。
- skip connection 作用。防止梯度消失。
- 说说决策树。信息增益那几个公式我混淆了，就说了一个决策树的大致思想。
- tensorflow 有哪些缺点？速度慢，频繁更新，修改 API，结构混乱，同样的功能可以在多个地方找到不同的实现。

三面

8 月 22 日, 20min

- 聊我的论文。
- 说说 Faster-RCNN, YOLO, SSD, FPN。
- 算法题，简单 DP，最长上升子序列，对话很有趣：

Q: “写个堆排序。”

A: “我忘记堆怎么维护的了”

Q: “那如果告诉你怎么维护的你可以写么？”

A: “可以。”

Q: “那我们来写个最长上升子串吧！”

A: “是子串，不是子序列，是吧？”

Q: “你就写子串吧，子序列更难一些”

A: “子串太简单了，我还是写子序列吧！”

2 分钟后，写好了

Q: “lower_bound(), 我还以为你这么快写完了呢?”

A: “那我再手写一个二分?”

Q: “不用了，可以了，可以了。”

开始安利网易游戏。

HR 面

8 月 22 日，30min

- 这些项目是自己接的私活儿还是老板给的？老板。
- 你最有成就感的项目是哪个，你担任什么样的角色？XXX 项目，核心开发者。
- 这个项目遇到的最大问题是什么？你是怎么解决的，用了多久？原来的算法效果完全不能用啊，我干了 XXX，马上提升了 XX 指标 XX 点。
- 近期比较大的挫折是什么？很久以前，本科 ACM 打过铁。
- 如何看待微软的工作氛围？没有网易这么充实（轻松惬意）。
- 有哪些 offer，如果我们开出相同的价格，你会怎么选择？有 balabala。优先考虑，a.成长性，我是希望加班的（别不给钱就行） b.薪水。在工作比较对口，能让人快速成长，不是天天写业务的情况下，优先考虑巨硬啊，毕竟生活成本比较低，而且第一年惯例有 package，薪水不会差太多。*当然，如果有公司的 base 能多 10w，可以考虑。*
- 家在哪里，家人反对来杭州吗？如果过来会干几年，会不会跳槽？
- 你问了之前面试官哪些问题。
- 玩过哪些游戏？可多了，一一列举。
- 研发比较辛苦，安利网易游戏算法岗还是挺轻松的。

结果

9月7日收到 offer 意向书。

10月18日收到 special offer，薪资比其他几家差一点，考虑到加班和游戏行业的局限性，放弃了。

地平线机器人

没有内推，但是简历被捞起来参加提前批了。

一面

9月3日，45min，全程聊论文和项目。

- 重点探讨了小目标检测的 trick 和类别不均衡问题。应该是把自己踩过的坑都问了我一遍，好在我踩得坑也不少！
- 讲讲 RCNN 系列的演变过程。
- 讲讲 YOLO，SSD。
- 讲讲 RFCN，重点阐述 position sensitive score map。
- 讲讲 FPN。

二面

9月5日，45min

- 聊实习，聊项目
- 有一条历史时间线，上面有很多人，每个人有出生和死亡日期，设计一个结构，能快速查询和某一个人同时出现的所有人，比如：查询曹操，那么结果应该是夏侯惇，曹仁等...。先说了线段树，分析了时间和空间复杂度，空间复杂度让我推了一个级数和。

- 看起来空间复杂度是可以接受的, 那么如果共同出现不是精确到天要精确到秒呢? 可以在叶子节点上排序。
- 有没有更优的策略? 我没想出来, 面试官提示了哈希, 我就顺着他的想法说了, 以日期为 key, value 是这天活着的人的有序集合。ok, 面试官很满意, 然而我觉得这样的做法并没有比我的线段树更优, 因为对于查询的人, 他活着的每一天我都要查 hash 表, 这样复杂度其实是 $O(n)$, n 是这个人活着的天数, 大约是 $1e5$ 这个数量级的, 而线段树支持区间查询, 只需要 $O(\log m)$, m 是有文字可考的历史线上天数, 按中华上下五千年算, m 大约是 5000×356 的级别, 加上 \log , 远远小于 $1e5$ 。嗯嗯, 所以我的答案应该是比他的标准答案还要好。
- 打开百度图片搜索, 可以看到虽然同一行所有图片的高度是一样的, 但他们的原始高度肯定是不一样的, 这个后面可能是怎么实现的? 以每一行的第一张图片高度为基准, random 一个偏差, 以基准加偏差作为这一行的高度, 所有图片 resize 到这个高度, 然后往固定的行宽里填充, 如果变形太多, 就取一个放到下一行。
- 手撕代码, 自定义结构, 实现这个自动调整图片排列的功能。时间不多, 撸了个大概。

三面

9月19日, 60min

- 聊项目, 聊实习, 推公式。
- 有一个图像分类的问题, 类别有 2000 万类, 不考虑计算资源和网络结构合理性的问题, 暴力地在最后做 2000 万的全连接。只看对前 900 万类的分类, 效果很好, 如果加上后面的 1100 万类, 网络无法收敛, 可能是什么原因? 既然不考虑网络结构和计算资源, 那么只可能是数据的问题, 有可能是类别不平衡导致的, 可以用重采样或者部分类别做数据增广的

方式解决。

- 如果数据确保没有问题，依然出现这样的情况呢？没想出来，好在他提示了。
- 面试官提示，我做了一个实验如果把后面 1100 万的类标从 0 开始标记，问题就消失了，即 label = 11000000 的变成 0，11000001 的变成 1，以此类推。这个问题确实不好想，但是我踩过坑啊，类标是转成浮点数计算的，所以位数上去以后会损失精度。
- 说一下浮点数在计算机里的表示，为什么会损失精度？具体表示是本科学的记不清了，大概是由符号位，有效数字位和幂次位组成，所以超过有效数字位的精度会被忽略掉。
- 怎么解决这个精度问题？

手写高精度的 loss;级联分类器，前几个分抽象类，最后一个分到具体类。

- Re-id 的东西问了一点。

结果

9 月 19 日收到 offer 意向书，当天发，真赞！

10 月 22 日收到 special offer，在南京的企业里真的算非常高了。

微软

实习生转正面

印度大佬英文面。

- 聊了实习做的东西和 team 里其它人做的一些东西。
- 算法题：丑陋数变体。手撕代码，推出了 int 型的丑陋数上限并分析复杂度。
- 愉快地聊人生理想。

结果

10月9日收到 offer 意向书。

10月23日收到 special offer。

谷歌

基本情况

七月底和八月底有两场，校招必须通过至少一场。

面试总共五轮，一轮视频面，四轮 onsite 面试，其中包括两轮中文技术面，一轮英文技术面，一轮英文非技术面。

因为邮件里特意强调了不要泄露面试题，所以尊重一下，还是不说具体题目了。就说说面试感受吧。

面试安排真的是所有公司里最奢华的了，陆家嘴五星级酒店江景房，报销路费伙食费。定时有邮件通知面试进度，不会像有的公司十天半个月没有消息，所以虽然时间拖得很长，但并没有很焦虑的感觉。

结果

战线拖得很长，还在焦急地等待结果中。

摩根士丹利

大摩直通车

就等参加呢，和阿里面试冲突，两边时间都不可调整，遂放弃。

腾讯

简历未被捞起来。

感受

- 实习和项目，对于找工作太重要了，如果实习单位足够好，几乎所有面试官都会问，基本上一聊实习和项目 20-30min 就过去了。
- 算法题不用刷太难的吧，面下来感觉大部分都是 easy 级别的，重点在于 bug free 和手撕代码的能力。
- 以我面的情况看，算法岗基本是一轮工程面，考察 coding 和思维，一轮算法面，考察对领域的认知情况，一轮 boss 面，把把关。
- 会就是会，不会就是不会，千万别说会一点。
- 面试的过程不仅是公司考察面试者的过程，也是面试者考察公司的过程，千万不要妄自菲薄。
- 七八月份还雄心壮志想要收割十几二十个 offer，后来觉得还是太累了，能把自己真正想去的公司 offer 拿到手就可以了。

我是如何一步步拿下 Google Offer 的？

历时两个月，我终于踏进了梦想的公司.....在内心深处我深知必须努力尝试，因为进入 Google 并非不可能。只有尝试过我才会知道结果，有点像薛定谔的猫，这种想法让我感到不安。

与其他众多的技术人员一样，从一开始 Google 就是我梦寐以求的公司。此次在我通过 Google 的面试后，很多朋友都来找我，希望我能给他们接下来的面试提供一些技巧和指导。在与他们交流的过程中，我总结出了一些心得，而且觉得既然我明白面试的困难，就更应当把这些心得记录下来。整个过程非常艰难，尤其是如果你正在从事一份全职工作的话会更加费劲。在我准备的过程中，网上的很多文章给予了我很大的帮助，现在轮到我回馈开发者社区了。希望对大家有所帮助！

有趣的是，我在 2014 年参加过一次 Google 的面试，但失败了。那次面试非常糟糕，所以我完全记不清当天的情况了，那么只好谈一谈我这次的面试经历了。另外，在本文的最后，我会提到在准备面试期间我犯的一个让我十分后悔的错误。

动机

在我从事软件工程师三年半的时间里，我曾在一家创业公司与一些金融科技公司共事。每段经历都教会了我很多东西，每一步获得的知识都让我距离进入 Google 的目标更近了一步。这一次，当我决定寻找下一个机会的时候，我做了一个明确的决定：我想花点时间来为进入科技大公司而认真做准备，这一次我不会再屈就妥协。尽管我也面试了几家大型的科技公司，但老实说，我能够进入 Google 的机会连 1% 都没有，因为 Google 软件工程师面试的难度是业界数一数二的。但在内心深处我知道我必须努力尝试，因为进入 Google 并非不可能。只有尝试过我才会知道结果，有点像薛定谔的猫，这种想法让我感到不安。

确定面试的要素

根据我的经验，一般面试准备都涉及掌握五大重要内容：写代码、设计、语言特定的概念、以前的项目经历、简历，以及最后一项（但并非最不重要）是具体了解该公司。

写代码

在我的职业生涯中，我参加过面试的所有公司都至少有两轮纯算法的问题，他们会给你一个问题的，而你需要提出一个解决方案并写出代码。

例如，编写代码在控制台上输出数字 1 到 10（当然这只是一个例子）。

我会在本文的其余部分详细阐述本节的具体细节。

设计

根据我参加的面试的级别，有些公司会给出一轮设计问题，他们会给你一个相当抽象的系统，你需要为该系统做一份设计。

例如，如果你需要从零开始构建一个社交网络，那么你将如何设计这个系统？

对于接下来的 3 个方面，有些公司会专门再给出一轮，而有些公司则可能将它们融合到写代码或设计的环节。

熟悉首选语言

在大多数面试中，你都需要提供一门你最熟悉的语言。除了编程能力之外，在大多数情况下，你应该了解这门语言的基础知识。有些公司可能会问你一些基础知识，而有些公司则可能会深入问你某个领域。

例如，Java 中的继承是什么？

以前的项目/简历

这是有时我会完全忘记准备的环节之一，因为我会全神贯注到写代码和设计等方面，但面试不只是写代码！面试是为了证明你是一个潜在的队友。这就需要向他们解释在过去的项

目中你如何与其他人合作。这个环节的问题因人而异，比如他们会根据简历上的技术或项目提问等。

例如，谈谈最让你感到自豪的项目吧。

了解公司/好奇心

认真了解公司并专门做准备通常是最容易被大多数候选人所忽视的环节，而我只能说，这才是真正能够让你脱颖而出的环节。他们可能不会直接问及有关公司的问题，但是做一些了解可以帮助你更好地为上述环节做准备。你阅读我的这篇文章正是因为你想在即将到来的某次面试中取得好结果，那么说明你已经在朝着正确的方向努力了。

例如，为什么你选择我们公司作为你的下一个目标呢？

面试准备

我利用了各方面不同的资源来为面试做准备，下面我列出了在某个具体领域内曾对我有所帮助的资源：

课程学习

我在做准备的时候，购买了以下 3 门课程。这些课程帮助我明确了基本知识。注意：我未能完全学完这些课程，但完成了大部分内容。此外，我选择参加这些课程是因为与阅读文本相比，我个人通过视频的学习可以更快更好地掌握内容。再加上让别人讲给你听，比你自学学得好。我已经很久没有参加过面试了，所以这些课程还帮助我进行了编程练习。

- 1、数据结构与算法分析：了解如何从头开始实现数据结构。
- 2、面向软件开发人员的 Java 编程精讲：Java 编程的基础知识。
- 3、全面解析软件工程面试：通过一些实时的练习，了解 Google 面试的整个流程。

Leetcode 高级会员

对于我这个级别的人员来说，大多数科技巨头的面试都会围绕编程与算法，因为我没有太多行业方面的经验。所以，Leetcode premium 的高级会员非常适合我。我需要缴纳 35 美元的月费，但是我花的 70 块钱的每一分都物有所值。Leetcode 是我练习编程的唯一资源，我强烈向那些想要进入科技大公司的人推荐。编程的准备工作范围很广且内容很深，所以我在后面的小节“一个月集中精力做技术准备”中详细介绍了这个环节，解释了我充分利用该资源的方法，以及在此过程中我犯的一些错误。

YouTube 视频

在准备期间 YouTube 帮了我很多忙。不仅仅是为了学习，也是为了放松。再怎么疯狂我也从未想过短短两个月后我会参与 YouTube 的工作！奇迹真的出现了！

在你发现有些概念理解起来有困难或者只能理解一部分的时候，你可以看视频。下面我列出了一些给了我很大帮助和改变了我的命运的一些视频。然而，在浏览获取信息的时候，我曾观看过许许多多其他的视频。你永远无法预知哪条小消息可能会帮到你。

<https://youtu.be/ko-KkSmp-Lk>

https://youtu.be/XKu_SEDAykw

<https://youtu.be/oWbUtlUhwa8>

https://youtu.be/78t_yHuGg-0

<https://youtu.be/nYFd7VHKyWQ>

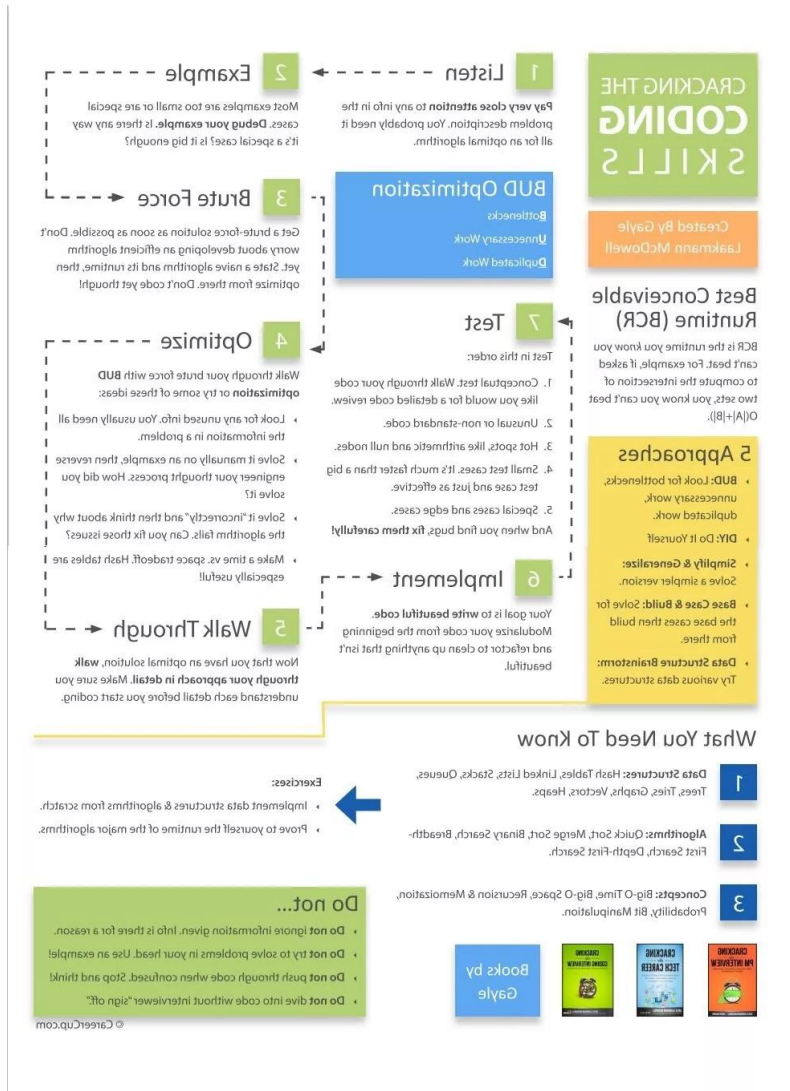
<https://youtu.be/CZ3wluvmHeM>

其他资源

Gayle Laakmann McDowell 的图书可以帮助你做好编程面试的准备：

《Cracking the Coding Interview》（攻克编程面试）

一步一步攻克面试（超级有用的图）：



准备设计面试：

Hired in tech 的课程：

<https://www.hiredintech.com/cl...on/52>

来自 Ramon Lopez 的有关系统设计面试方面的播放列表：

<https://youtu.be/KmAyPUv9gOY>

一个月集中精力做技术准备

玩转 Leetcode

Leetcode 是一个网站 (<https://leetcode.com/>)，你可以在上面编写和运行与面试

问题类似的程序。注意：在 Google 的面试中，他们问我的问题网上并没有，但基本概念

仍然相同。因此，归根结底你还是要掌握解决问题的艺术、识别规律并进行练习。

Leetcode premium 有 4 个级别：

简单 中等 困难 公司特有的问题

每个级别又有多个子类别：

数组 字符串 链接列表 回溯 搜索/排序 树 图表 数学（概率等）

计划

现场面试前的一个月我开始了编程集中训练，最后总共解决了大约 60-70 个问题，这是攻克面试所必须付出的努力，关键在于计算机科学的基础知识。

每天我都会解决两个以上的问题。简单的问题花费的时间可能有所不同，大概需要 30 分钟到 1 小时。中等难度的问题实际上对我来说很难（有时反倒是连简单的问题对我来说很难）。中等难度的问题限时为 1 个小时，开始 15 分钟后，如果你不知道解决方案，那么就需要查找并尝试实现。1 个小时后，如果你的解决方案行不通，或无法通过某些测试用例，那么就只能停止并查看解决方案。下面是一个粗略的计划，但是你可以随意改编，以免准备过程过于枯燥。

- 1、完成数组和字符串的简单问题（前 5 天），赢取一些自信；
- 2、每天做一个中等难度的数组或字符串问题，以及一个简单的链表或回溯；
- 3、树和图表的简单问题；
- 4、排序和搜索；
- 5、其他剩余的问题

我的准备工作主要集中在掌握大多数简单问题和中等难度问题上，困难的问题我解决了不到 10 个。如果我无法理解某个特定的解决方案，那么我会搜索 YouTube，找个人来给

我解释这个解决方案。回溯是我最难理解的一个方面。

一个让我很后悔的错误

Leetcode 高级会员中有一个环节是某个公司特有的问题，也就是 Google、Facebook 等公司特有的问题。我犯的最大的错误就是一开始就准备了 Google 的问题。那些问题超级难，结果让我丧失了动力，最后只好不了了之。所以，一定要从简单的问题开始，然后逐步尝试中等难度和困难的问题。

我最喜欢的 Leetcode 高级会员功能

下面是我最喜欢的三个功能：

- 1、针对你的解决方案运行各种测试用例；
- 2、生成一个图表，让你可以与其他高级用户比较运行时；
- 3、查看其他 Leetcode 用户提交的最佳运行时的示例解决方案。

小贴士

不要给自己压力。在准备面试的过程中，我一如既往地生活，并 100% 地完成当下公司的工作。

每天花费的时间不需要太多，但是要坚持一两个月，有效利用每天的准备时间。周末的时间十分宝贵，要充分利用！我把周末称作“周末加油站”，因为如果我能在周末完成所有 Leetcode 上的简单问题和一些中等难度的问题，那么就会感觉元气满满。

招聘人员也是你的朋友！在参加面试之前，可以随时向招聘人员询问招聘的形式、期望、准备材料以及提示等等。这可以帮助你注意力集中在重要的事情上，因为计算机科学这个领域太大了。

首先从常规的准备开始，比如参加 Udemy 课程、观看 YouTube 视频、阅读文章等等。把 Leetcode 放到最后，因为如果最近你编写了类似的问题（在 1 个月以内），对

你的面试帮助才最大。我建议你不要一整周都专门用于编程，应该结合其他方面的学习。

时不时地做一些设计，在中途观看一些视频，学会享受学习的过程！

特别对于 Google 来说，有一件事情对我的帮助很大，那就是我要求招聘人员给我提供了一台 Chromebook 来写代码，因为通常白板会给我制造麻烦。我的做法是：利用白板整理出解决方案，方便面试官查看，然后最后用 Chromebook 写出代码。现如今很多公司都鼓励这种做法，但是你必须提出这样的要求！

有时只是在网上随便闲逛也会有所收获，我曾阅读过其他人攻克面试的文章，在很多方面给予了我帮助。

做最坏的打算，曾经我问自己：“如果进不了 Google，该怎么办？”我的答案是：“没关系，我有一个备用计划。我会尝试其他顶尖的公司，并继续努力，直到我拿到我想要的结果。”为此我保持冷静，并取得了更好的表现。

多年 C 语言开发转 CV 算法拿到年薪 40 万的面经

本人背景：985 本，专业软件工程，到目前为止工作 6 年，做过 C 语言开发，python 开发等工作；机器学习方面在 2017 年年底开始接触，自己主要学习了七月在线的深度学习集训营和计算机视觉第二期、吴恩达的深度学习课程、李飞飞的 cs231n、udacity 的机器学习纳米学位等课程，2018 年开始在自己原有单位实用深度学习方法做网络流量方面的一些分析工作，主要基于类似 TextCNN 和 img2txt 的一些方法。

因为对 CV 比较感兴趣，想要转往该方向，今年 10 月中旬后开始更新简历找 CV 方向的工作。这此跳槽中感觉颇为艰辛——很多 CV 相关工作都要求硕士学历，年底又碰上互联网寒冬缩招裁员的，我这样之前做了多年开发，并没有 CV 直接相关工作经验的，猎头来推的都是开发工作，招聘网站上投简历也基本没什么回应，我自己也因为时间少以及公司方面有所筛选，因此 CV 方面的算法工作实际去 onsite 参加面试的机会都很少。虽然可能感觉并不是比较成功的一次求职经历，很多地方都没做到本来可以达到的更好的状态，依然和大家分享一下，也可以作为经验和教训。话说因为原单位的项目没能完结掉，现在拖了两个月才刚入职，最近听闻各种互联网寒冬裁员的报道，真是让人非常担忧，也希望分析能够攒点人品，保我别在试用期里被裁员吧~

下面的面经只记录了 CV 算法相关的一些问题，数据结构算法的考题之类没特意留意记录下来，HR 面也都不涉及。就数据结构算法这方面我自己也就是准备面试期间看看剑指 offer 和 leetcode 的简单题的水平，现在已经不怎么刷题了，年轻的童鞋们还请自己加油...此外面试题目下都是我个人当时的回答，存在各种问题，仅供参考，欢迎大家斧正指点。

一、某上市互金企业 计算机视觉算法工程师

先是介绍了一下自己的项目，由此引申到了下面的问题：

(1) inception 系列的结构特点、resnet 的结构特点。

我画了一下相关的结构，inception 主要提到了每个 block 里多个不同的 filter size 的分支最后合并起来，这是让网络能够自主地学到用哪种 filter size 更好而非人工去设置，还提到了 v2 和 v3 在结构上的变化如用两个 3×3 代替 5×5 ，用 1×3 和 3×1 代替 3×3 等，这些都降低了参数量，增加了非线性；resnet 主要提到了 short cut，因为这个的原有所以能够解决梯度消失的问题，因为有 short cut 把经过几层卷积前的结果直接和卷积后的结果相加，即使最差情况经过卷积后结果的权重都是 0 了，加上卷积前的结果也还是和卷积前一致的。

(2) 面试官表示上述的描述，都是在结构上能够看出来的，有没有什么个人更深层次的理解没能想到还有什么深层次的理解...

(3) 有没有一些自己针对一些场景设计的结构、loss 函数等

提到了自己项目中设计的一些卷积网络的结构，一些竞赛里涉及实用多任务学习的方式来获得更好的效果（比如同时做分类和 detection 的任务）；自己没有设计过什么特殊的 loss 函数。

(4) 有没有除了翻转拉伸加噪点等等之外的，自己针对场景设计的独特的数据增强方式？

提到了 u-net 里那种加网格然后加以扭曲的方式（根据 grid 做 elastic transform），不过这也不算是独创的...还提到了在 kaggle 走神司机检测竞赛里看到的，把图片切割成左右两半再和别的切出来的拼起来的方式（原理是根据模型对原图生成的热力图分析，模型实际关注的就是图片中司机头部和手部两个位置，分别在图片左右两半）。另外最近我还学到了针对 CV 的一种新数据增强方式（海康威视在 2016 年 imagenet 的场景分类任务中提出的），在这里做个补充：在场景分类（比较高层的语义图像分类项目）中如果随机在原图中扣取一定的图片来作为数据扩增，很可能并不有效（比如海滩的场景中扣出来树和天空的图片，但是树和天空完全可能出现在各种其他场景中，这样的数据增强可能反而会造成混乱，反而降低效果），新方法是使用原本数据训练出来的模型对训练图片生成热力图，然后抠出来相应较

强部分的图片作为数据增强——这是一种监督式的数据增强方法。

(5) 怎么做人脸年龄的检测？

对这个问题我完全没有思考过，后面都是自己当时随便说说的，实际上至今也没想到很好的方法——先尝试对每个年龄分类，比如 100 个类。或者尝试能否把比如 100 个年龄分成几个稍微少一点的类，比如每 10 个年龄一分组来进行粗略的分类，随后再对每个年龄分组的里面训练细分类的模型...又思考能否使用细粒度的方式，使得不同年龄训练得比较分开然后通过检索的方式...

(6) 怎么做有没有 PS 痕迹的检测，具体来说比如一个证件照和人脸的合影，如何确认人脸部分不是 P 上去的？

首先想到的是能否获得足够多的 PS 图片来做分类，简单得来分类是否有 PS；随后又想到先 detect 出来身份证上的人脸部分，和照片里真人脸的部分比对，是否能发现非常大程度上一致，可能是把身份证上的头 P 到了真人头部分...（和上个问题一样，应该是拍拍贷业务常常涉及到的对接待人上传的照片审核的业务场景，对这块都没啥思考）。

(7) 如果用添加 PS 处理的图片来做数据增强，万一美颜摄像的也认为是有 PS 的怎么办？我相信 PS 的模式回合美颜的存在一些区别，能否也收集到美颜摄像的图片，多加一个单独的美颜的类别识别？

(8) 遇到一个新的问题，会是如何着手去做？

先去网上检索有没有相关领域比较权威的做法，比如 paper 啦，开源实现啦，相似的竞赛啦等等，有的话进行参考，然后根据自己的问题进行相应的修改尝试，能否取得好的效果。找不到的话，会看看有没有一些相似的，或者其他接近领域的想法，能否能够借鉴的，随后引申到自己项目借鉴的 TextCNN 和 img2txt 上去了。

二、某工业 AI 创业公司 计算机视觉算法工程师（这场面试时候比较认真都录音了，所以

题目都记得很清楚...其他面试时候好像都懒了没搞...)

同样在介绍了自己的项目后，问到了下面这些问题：

(1) LSTM 和 RNN 的区别：

LSTM 比 RNN 多了很多控制的门，通过这些门它能够学到哪些记忆需要被保留，哪些要遗忘；所以它能够一定程度防止梯度消失，记住比较长的之前的内容。（之前还准备记了 LSTM 各种门的结构，不过也没细说）。

(2) LSTM 相对于 RNN 的优点：

普通 RNN 因为梯度消失的关系记不住很远的内容。

(3) 为什么要做卷积，为什么卷积在图片处理上效果非常好？

卷积像是一个滤波器，对于图片每个位置都去获取一个它所关注的模式，具有平移不变性，对每个位置都去获取其所关注的部分。

(4) 为什么不直接在图像上做全连接而是用卷积？

全连接参数非常多，容易过拟合，还会把位置信息给去除掉。

(5) 还有最重要的还有的是什么？主要问题还有参数太多的问题

卷积还有具有稀疏性，参数共享的特点。

(6) 说一些经典的 CNN 的网络如 alexnet, resnet 和 inception 等

(基本上说的和上面讲的差不多那些东西，不展开了，大家自己查相关资料)

(7) global pooling 和全连接比的好处在哪儿？

全连接的参数量很多，在整个网络中占有的参数比例太大，可能会容易导致过拟合。

(8) global pooling 还有什么好处？

global pooling 还可以使得模型不必像全连接那样必须固定输入图片的尺寸了（想不到其他的了，不知道这个是不是也算是好处）。

(9) 1*1 卷积有啥作用？

能够升维降维，能够对不同 channel 信息进行混合。

(10) 谈谈 dropout？

现在用的不多，主要是全连接的时候会接 dropout。

(11) 只有全连接时候用 dropout 是吗？

额，我是这样的，感觉面试官不太认同...（后来查阅了相关的资料，还是觉得在卷积层加 dropout 效果并不好，主要是因为图像每个像素位置和周围可能在常常可能比较相关有不少冗余的部分，经过 CNN 和池化以后很可能就对某些个点 dropout 不敏感了；现在有些新的 dropout 方法，如 15 年的 spatialdropout 方法，随机选择某些 channel 来 drop；又如 18 年的 dropblock 方法，在卷积后的结果上选择一块块连起来的部分 drop）。

(12) dropout 主要作用是什么？

是防止过拟合。

(13) 为什么能防止过拟合？

dropout 在每个 batch 的训练中会随机失活一些连接，假设在没有失活的情况下某个连接权重很大，这种情况就容易过拟合，当把这个连接失活以后，就要依赖别的连接去重新训练能够适应的情况；从另外一个角度来说，这也有一种集成学习的效果，因为每次随机失活一部分训练，都像是训练了不同的网络。

(14) 介绍一下 BN：

在 batch 中，在一个层的输出结果后多做了一个标准化的操作，也就是减去均值除以标准差的操作，相当于就是把分布拉到了 0 均值单位正太分布的地方，这样的分布就使得数据往激活函数的线性部分更加靠拢，缓解了梯度消失的情况，可以加速训练收敛；但是这样也就改变了学习到特征，为了防止这个问题也为了防止模型的非线性变弱，BN 还有两个独立

的拉伸 w 和偏移 b 的可训练参数，用来重新对数据继续拉伸和偏移，让数据再被拉到 0 均值单位正太分布后再重新往非线性上恢复，这样也就使得网络的非线性能力得恢复，同时也使得每层数据分布与之前层的结果解耦，在线性防止梯度消失和非线性上取得均衡。

(15) 平时会用到 BN 吗？

A: 因为比较多的情况是用迁移学习，改动模型不多的时候不会用；在自己的项目设计网络时候会有用到。

(16) 说到迁移学习，你是怎么做迁移学习的呢？就是预训练的网络，最后几层再那个？

对，就是先把最后一层分类去掉改成需要分类的几个输出；然后逐渐往上开锁定的层。

(17) 怎么监控模型的好坏，有啥评价指标？

在训练的时候先看 loss，有没有在正确地下降；等到能够 overfit，loss 下降收敛以后再去其他的其他的评价指标，比如准确率，f1 score 等；我自己项目中主要也就看看准确率。

(18) 如果有个非常不平衡的数据集，可能正样本占 1%，负样本 占 99%，这个时候用什么评价指标？

会看是不是一个类似异常检测的问题，那一个占比很小的类是否很重要

(19) 不一定

我表示没有要留意是不是很小的那个类别的话那就不用 f1 score 吧...

(20) ROC 和 AUC 记得怎么画吗？

这俩我当时不怎么记得了；就记得 PR 曲线...（提起了这俩，就这个就知道上面讲错了吧，不平衡的数据可以看 AUC 的，对这种情况不敏感...）

(21) 其他一些机器学习如线性回归，逻辑回归，树等方法有所了解吗？讲一下 xgboost？

没用过不大了解...（这个没用到过一直没看，果然大家都喜欢问...）

(22) 说一下二分类的 log loss 函数:

就写了一下公式...

(23) 发现一个网络训练地过拟合了的话, 能想到的要做的步骤会是什么?

训练到过拟合以后, 先加一些限制过拟合的比如 dropout, l2 正则化等; 如果还是比较厉害, 我会考虑是否模型过于复杂了, 可能会选择换轻量级的模型, 或者把模型调整简单比如把层数, channel 数消减; 然后就是做数据增强, 如果有可能的话最好能收集更多的数据。

(24) 做过 NLP 的东西吗? 对有什么了解过的算法?

做数据迁移可能会用别人训练好的 word-embedding; 训练 embedding 可以用 skip gram...NLP 方面也了解地不多。

三、平安金融壹帐通 算法落地工程师

一样地先介绍了自己的项目, 然后问了下面这些:

(1) 卷积怎么算的

这个很基础, 不详述了...我就是回答时候稍微提到了点 depth-wise 之类特殊点的卷积。

(2) RNN 是什么样的, LSTM 是什么样的, 什么用; LSTM 的起始 cell state 是多少?

基础的概述也就不展开了; 关于 LSTM 的其实 cell state 当时没有答上来, 平时我好像都是用的框架默认的, 没试过可以调这个...后来差了点相关资料

([https://r2rt.com/non-zero-inithtml](https://r2rt.com/non-zero-init-...html)), 有兴趣大家自己看, 篇幅有限也不展开了。

(3) 除了 CNN, RNN, 还用过其他什么深度学习的模型的吗? 比如 GAN 是什么, 讲一下。

关于 GAN 的描述也直接网上查就可以了; 当时自己也说得不太好, 学了以后没用过也没好好关注过...另外其他的还可以提提 deep and wide 模型之类的。

(4) 就我做过的一个根据 triplet loss 和多任务学习实现的车辆细粒度型号分类项目, 问

了具体的实现。

triplet loss 是做人脸识别，各种 reid 的细粒度分类的常用 loss，（另外一种孪生网络），通过 triplet loss 来学习 anchor、positive 和 negative 三个样本之间差异，让 anchor 和 positive（同一类的两个训练数据）的距离比 anchor 和 negative（和 anchor 不同类的训练数据，应通过 hard-mining 精良选择和 anchor 相似的数据）的距离远。具体定义可以直接查询网上定义。

(5) tensorflow 的模型怎么存的，怎么上线的？

因为好久没有生成 pb 文件了，我都忘了 pb 文件的后缀了，直接就和面试官说我是直接存的（其实我指的就是生成 pb，不是训练时候的 save...）；上线这方面我还没啥经验，我之前的项目实际上都还没有实际上线的，都还没很好落地呢...

(6) 因为提到了可以用知识蒸馏的方式来压缩模型，让我介绍知识蒸馏怎么做的；为什么蒸馏出来的小模型有用？

参考 <https://zhuanlan.zhihu.com/p/39945855>

(7) 怎么设计的网络，filter 尺寸，个数等等？

这个部分我建议参考“解析卷积神经网络——深度学习实践手册”这本小书里“网络超参”设定里提到的一些内容。还提到了根据数据大小规模会相应修改模型复杂程度，比如数据少而且过拟合严重，我可能会尝试把原有模型的 channel 先都减半之类的操作。另外还可以参考最小的各种研究结果的实现，参考 res-block, se-block, inception-block 等经典结构作为一些组件插入自己的设计中。

(8) tensorflow 训练中如何优化训练速度

tensorflow 有 profile 工具可以观察具体各个部分的耗时，可以针对性地修改模型进行调整。

还有比如可以把 input-pipeline 放到 cpu 上做 prefetch，然后同时 gpu 上在进行训练，

两者互相之间是一定并行的关系；在 pipeline 上加 cache 会很显著地提高数据阐述效率，提高 GPU 利用率。

海康、YY、抖音、CVTE、招银科技、比特大陆（计算机视觉算法或深度学习算法岗）面经

本人 211 小硕，本科非 211（说实话这个挺吃亏），研究方向：计算机视觉、模式识别

一、第一个面试是海康，电话面：

首先说项目，然后挑项目里的算法模型进行提问，比如你用的什么模型提取人脸关键点、用什么模型进行人脸检测等，要求把每个模型的原理能表述清楚，最后问了两个 python 的问题：怎么遍历文件夹和 python 的数字精度是多少（汗_-||，第二个完全没注意过啊）。这次面试基本没有问机器学习的内容，可能会留在二面问吧，毕竟第一次面试太紧张了，说话都哆嗦，一面就挂了。所以面试一定要调整好心态，把面试官当做是学术交流的对象，不断地给他输出你会的知识点，不要把他当成上级，显得畏手畏脚就不好了。

二、YY，笔试过了，武汉现场面：

先说项目，与海康不同的是，面试官挑出项目中的一个模型，然后不断深入进行提问。比如项目中用了 CNN，那么请介绍 CNN 的特点，为什么 CNN 有效，哪些方法可以防止过拟合，Dropout 的原理是什么，梯度消失和梯度爆炸原因、面对类别不均衡的数据集怎么办（后悔那晚没听计算机视觉第二期的直播课啊，seven 老师讲了两种方法，我当时只说了一种，但是面试官可能还是需要另一种的答案），面对数据集小且不均衡的数据集怎么办等等。。基本是深度学习的问题，大体我都答上来了，本以为过一面没问题，但是莫名地把我给挂了。我觉得这可能是一面刷了一波学历吧。。

三、抖音，过笔试，视频面：

项目说完，中间稍微提了些项目的问题，接着开始问机器学习的内容：SVM 的原理、SVM 的软间隔和硬间隔的区别、惩罚系数对 SVM 的影响等，决策树相关的知识（具体的问题有

些忘了，反正也是介绍原理之类的)。最后问我会什么数据结构和算法吗，我说还行吧，结果丢了我一道编程题：给定一个正整数数组，求一个连续子数组满足子数组的最小值跟子数组和的乘积最大，返回这个积和子数组区间。一脸懵逼，中途面试官提供了一个思路，但我觉得时间复杂度仍很高，最后放弃了。。后面想想，也许就按他思路写一遍，能出来个结果就行，不一定非得最优，最重要的是考察你的写代码能力。

四、CVTE（医学图像处理岗），电话面：

由于 CVTE 的视觉岗满人了，我就投了这个岗位。。首先问我了解过医学图像吗？知道医学图像分割吗？我说不会，但是接触过普通图像的分割，也不是很了解，只是扯了下用 FCN 的图像分割。接下来就是机器学习问题轰炸：SVM 原理、核函数怎么用的、正则化方法、决策树算法有哪些、xgboost 和 GBDT 的区别、boost 和 bagged 分别是降低偏差还是方差，为什么？等，考察的大多数是决策树集成学习方面的问题。最终以岗位不合适把我挂了。

五、招银科技，笔试加电话面加现场面：

电话面基本是机器学习的问题：了解哪些机器学习模型、SVM 的支持向量是什么、什么是过拟合、过拟合了怎么办、梯度消失怎么办、boost 和 bagged 方法区别，还问了求大数据的 top-k 问题。现场面 3 轮：一面是应该是技术人员，他问我用什么语言，我说 python，他还感觉有点诧异，也许银行大部分招 java 岗吧，然后给了我一张纸，要写各种排序算法的时空复杂度和是否原址排序，又写了一个爬楼梯的代码、手推了下梯度下降算法；二面应该是技术总监了，详细聊了下项目，期间我问了招银有没有做视觉的，他说很少，我心里就想凉了一半，对金融推荐系统这一块我不熟悉，而且这边的总监对我做的项目也不了解；三面 HR，主要聊本科的大学生活。面完回去等通知，等到现在还没消息，估计是凉了，岗位实在不合适。

六、比特大陆，笔试加视频面：

时间安排的比较紧，一人 30min，不能超过时间，一面介绍项目，问 SVM 原理，欠拟合和过拟合怎么办，用什么激活函数，项目中的模型用的学习率等，最后给了编程题：把一个随机顺序的数组按 $\text{mod}3=0$ 放左边、 $\text{mod}3=2$ 放中间、 $\text{mod}3=1$ 放右边的顺序进行排序，要求空间上原地排序，时间上单次遍历的时间。由于前面用的时间太多，编程题时间比较紧，我编到一半就让我说说思路就好了，然后我说了我的思路，他说恩不错，可以。。。就过了一面，第二天二面，换了个人，也是开始介绍项目，又问 SVM 原理（--，我也是醉了，不跟一面的人沟通下么），balabala 说了一堆之后，就直接给编程题：给定一个整数数组，求任一元素减去该元素右边任一元素的差值的最大值。先写了一个，说时间复杂度太大，我就一直卡在用动态规划上去了，当天状态也不好，最后没优化出来，给挂了。其实面完的时候，瞬间想出来了，根本不用去踩动态规划的坑 T T。

以上就是我的部分面试经验了，下面我总结一下：

- 1、面试千万不要紧张，千万不要紧张，千万不要紧张（重要的事说三遍），把面试当做同学或同事间的正常交流，而且大多面试官很友好，不会说因为你不会哪个知识点就鄙视你。面试也是你学习知识的一个途径嘛。
- 2、学历真的还挺重要的，好多大公司一面在面试结果差不多的情况下优先 985,211，所以学历不是很好的童鞋要尽量把自己会的都展示出来，秀出自己技能！（我是么得秀）
- 3、面试的时候千万不要自己提起你听过但又不熟悉的知识点，不然面试官一问你，你就尴尬了。
- 4、基础的算法还是要了解下的，像我这种本科不是计算机专业的，只能靠刷 leetcode 或牛客网、七月在线上的题来积累下经验了。
- 5、机器学习，深度学习的理论知识要掌握清楚，这是必备技能。

说了这么多，希望我的惨痛的面试经验能对大家有用吧 T T。在全民皆算法岗的现在，打铁

还需自身硬呐，自己修为还是不够，还是得继续搬砖修炼，春招再战！

觉得有用的童鞋请点下右上角的关注，谢谢啦~~~~

第三部分 CV 计算机视觉

双非应届小硕关于秋招的血与泪：投过很多简历拿到 2 个 CV offer

从 8 月份开始求职开始到 10 月出拿到第一个 offer，然后继续第二个 offer，有太多的艰辛和痛楚，自己也做了很多的记录。希望和大家一起分享自己的点滴求职经历。

说心里话，就自己而言，我是一个双非院校的硕士，做的也算是和算法相关吧，但是选择和大家一样在 8 月份与哪些名校进行竞争，真的是神仙打架的月份，投出去的简历不计其数，几乎没有任何回信，算法岗秋招对于学校说没有关系真心胡扯，但是要是说全部都看学历和学校也是胡扯，自己也接到了旷视科技和意图科技的面试，虽然没能通过，但是给了不少的启发，也拓宽了自己的视野。

对于求职个人给出以下几点建议：

1. CV 就业 2 课程里的内容好好的学习，别在面试的时候自己把学的东西写进去，但是又答不上来，一定一定要做笔记，我个人是以写 CSDN 博客的方式来进行记录自己的学习路径，也可以写知乎或者博客园，这样在面试的时候，面试官也会看你的 blog；
2. 对于课程安排的实践项目中，一定要动手实践，对于结果也要做仔细的分析，不要抛出结果结算了，代码算法的实现也要理清楚，不然写进简历，面试官极有可能会问你，某一层的输出维度是多少？还有就是其中某个部分的原理是什么？（后面附具体的问题）；
3. 虽然个人找的是深度学习和计算机视觉算法方面的东西，但是机器学习的经典算法也要理解到位，尤其 SVM 和 K-Means、KNN（几乎我每场都被问了），还有就是对于自然语言处理的一些东西也要了解，视野不可以太窄，具体内容，课程上的课程就够了；

4. 还有就是对于算法的应用场景也要有所思考，别只了解一些理论的东西，课程中老师闲扯的东西，尽量自己在做深入的思考（比如，车辆 ReID 的内容，可不可以延伸到行人 ReID（CV 就业班的内容），然后到姿态估计（机器学习集训营第 9 期有这个好像），然后两者结合增加模型的精确度，再进一步能不能对于视频的内容进行动态识别，等等）；

5. 多面试，然后记住当时不会的问题，回去之后多和老师讨论，然后自己进一步学习和复盘，还有就是一定一定不能怂，不会就是不会，会的就是会，真实很重要，这点特别感谢这个秋招一来就业邢老师的陪伴，期间心态都快崩了，邢老师不断地鼓励，感激！！

其他的也不多逼逼了，经历了就会知道其中的艰辛，别退缩，别自卑，好好的跟着班级走，别去觉得课程项目简单或者怎么样，面试不会问你多么多么高深的问题，不是每家公司都是研究所，人家需要实现后的落地，然后盈利，不可能一直砸钱搞研究，具体的公司研究倾向性，这个得自己网上搜索或者和就业处的史老师研究一下，这个公司到底是干嘛的？别怕麻烦老师，人很好的都。

下面给出面试时被问得问题，写出一部分：

旷视科技面试：

- 1、一开始就是手撕代码（基础的弄出来了，优化的部分没弄好）
- 2、SENet 的原理说一下，和 Attention 机制的区别？
- 3、EfficientNet 的原理？基于什么模型的搜索？
- 4、剪枝和量化有了解吗？具体在实现的时候是基于什么剪枝的，是 channel 还是 weight？
- 5、int8 量化的原理？
- 6、模型的融合模型准确率有点大，为什么还要选择融合？（这个是简历中的问题，被揪出来了！！！没考虑到的）

接下来也没有多问了，可能觉得我太菜了，因为是在浙大那边，一面定胜负，结果跪了。

中电海康面试：

- 1、手撕代吗，还好比较基础
- 2、对于 pytorch 框架和 tensorflow 他们的区别是什么？
- 3、对于 SENet 的效果提升，有什么看法？
- 4、讲一下 K-Means 的原理以及步骤？
- 5、讲一下决策树的三要素？
- 6、怎么解决过拟合的问题？
- 7、对于图像拼接，基于什么特征？为什么选择这个特征？

最后打开电脑看了一下我的 CSDN，说我学习挺认真，还是很开心的，毕竟写博客那么久，根本没人看。。。。结束就让回去了，说等消息。

魔点科技面试：

- 1、对于 EfficientNet，说明一下原理？
- 2、SENet 的原理，以及画出 block 的图？
- 3、ReID 项目中，Triplet loss 的原理？
- 4、说明一下怎么解决的困难样本的问题？
- 5、对于 ReID 模型，模型的 FLOPS 是多少？可以到达多少帧？

韵达面试：

- 1、简单介绍自己以及项目；
- 2、大概说明细粒度分类项目中的数据，多少个类，每个类多少张图片？
- 3、使用说明设备进行的训练？
- 4、在 ReID 项目中，解释一下 AlignedReID 方法？
- 5、为什么要进行切片？

6、为什么切片后要进行动态对齐？

7、ReID 数据集你是怎么标注的？

8、什么是表征学习以及度量学习？

然后就是总监面

1、主成分分析与因子分解的区别？

2、现在用的说明学习框架，简单说明一下原理？

3、有没有接触过金融类软件？

最后是老总面

大概聊了一下对于机器学习以及深度学习的看法，然后就是实际的应用场景，聊的比较开心，第二天就给了 offer。

小公司的就不具体说了，大概和上面的一样，总结完也可以看出来，面试的好坏更多在于你的准备，简历上的东西一定要非常非常熟悉，目前个人才接到 2 个 offer，但是更倾向于其中的一个，公司目标明确，非常清楚自己在做什么，分工也非常明确，也与重点实验室有合作。个人觉得刚刚毕业不能完全看待遇，也要看团队的情况，能不能学到知识，实现自己的价值。而不是将就地说曲线救国。

机械转算法岗位：我是如何一个月连拿 4 个 offer 的

我是机器学习集训营第八期的学员，某中上 985 机械硕士在读，目前刚刚参加完秋招。在校时主要做机器人相关，但研二的时候发现自己各方面的好像都会一点，但是都不精，就很担心找工作的事。因为也一直在关注人工智能方面的情况，自己也自学了一段时间，就是看看书，网上找点视频，但是很散，不系统。

恰好在网上看到七月在线的宣传，加上有朋友推荐，我就报了集训营的课程。在此之前自己虽然也对 c++ 比较熟悉，但是在学完集训营的课程之后去面试的过程中才发现，自己的算法能力有待提高，在这里跌倒了好几次，后面会重点介绍，大家要找算法的话！代码能力一定一定要提升！多刷刷 leetcode 和剑指 offer。接下来就详细说一下学习过程吧！我平时都在实验室做东西或者写文章啥的，所以都是借助周末或者晚上看集训营的课程，真的是从最基础的 python，到数学，再到算法推导以及项目实践，仔仔细细过了一遍，很多难的地方我至少看过了两遍，并且手推了 svm 和 Xgboost 等。不得不说，七月的课程设置还是蛮细致的。这不是打广告，亲身实践吧！

因为我是利用课余时间学习的，所以战线拉得比较长，我总共花费了将近半年的时间，才学完所有的基础课和集训营课程，这时大概已经八月初了，然后花了大概半个月对学的东西做了一个总结和归纳。然后就把简历做好，然后还让七月的老师帮我做了修改和提了建议，不然感觉好多家会直接挂简历。然后大概八月底我就开始准备去面试啦。

秋招面试持续了一个月多一点，拿了三个 offer，还有一些是最后没有去面试的，也有一些公司挂掉了。下面是我把几次重要面试的情况总结如下，大家有兴趣可以参考一二。

A，云从科技。cv 岗位

一面

简历上项目，会问道一些具体的细节，所以对项目要很熟悉；

对常见的机器学习算法有哪些比较熟，我当时说了 svm 和聚类，然后就问了我聚类的原理，

以及如何判断是否聚类完成

降低过拟合的方式；

神经网络降低过拟合的方式（dropout，早停）；

BN 层作用；

BN 层加在激活函数前与后效果有何不同；

机器学习中常用 loss；

精确率召回率；

重点是项目，问的很细

还问了我对那些排序算法比较了解，他们的时间复杂度和空间复杂度都是多少，

最后出了一个堆排序的题让我做了以下（见过类似的）

二面

二面比较尴尬，感觉是主管那种。

一上来就让我写代码，然后我很尴尬的没有写出来，虽然最后介绍项目还不错，但是主管对

我好像没有啥兴趣了，就问了两个简单的问题，

1.你用了 resnet，为啥不用其他的呢？有没有试一试

2. 你在什么上面跑模型啊？服务器还是电脑啊？为啥不用服务器呢？

然后就挂了，我也很迷，所以代码很重要很重要很重要！！

中兴通讯

一面

首先是介绍项目，中兴问的不是特别详细。

然后问了我关于项目上的问题，对机器学习有啥理解；

然后问了过拟合和欠拟合的区别；

还问了 svm 都能用于什么样的分类，核函数都有那些，有啥区别；

然后问了我对常见的数据结构，包括链表，堆栈等的情况，最后写了一个链表反转的题目

二面

二面基本就是聊天，问优缺点，以及学习能力，还聊了聊英文，好在我英语还不错。

还有其他一些公司，比如中国移动子公司、oppo、迈瑞等，基本都是先问项目，然后问基础知识，最后手撕一道代码题！所以代码能力很重要，最后综合考虑，我签约了中兴，因为薪资（16k*14-16）和地点的原因

上完七月的课程，只要自己稍微总结以下，就基本满足面试要求了，但代码要自己好好练习！

总结起来，失败的地方有这么几点。

第一，手写代码的功夫有点弱，各位同学务必花时间刷一下算法和编程；

第二，请务必耐住性子，心态一定要稳，挂了也不要沮丧，拿到 offer 也别沾沾自喜，做好基本的，掌握基本知识，做好实战课题，多复习，多请教。多试一试！

感激七月这个平台，真的对我帮助非常大，感谢各位老师对我的帮助！真的要静下心来好好的一步一步做，这个浮躁的时代，静下心来，耐住性子，很难，但是做到了很牛，所以我还会继续努力，希望大家和我一起加油！

双非小白硕士的计算机视觉工程师面试经历

前景介绍：本人目前在一家互联网公司担任图像识别工程师，本科通信，硕士还是通信，计算机基础差成鬼

由于本硕都是通信专业，基础差的一批，但是硕士导师主攻模式识别，于是研究方向在研一快结束的时候，稀里糊涂的决定了是计算机视觉。由于自己是导师的大弟子，连个可以请教的师兄都不存在。最开始疯狂的买书，买了各种各样深度学习的书，看了一两本，精度和理解的能力太差，学的很慢；后来在导师的推荐下看吴恩达的视频从机器学习开始学习，觉得理论可以了，但是一写代码还是无从下手；于是开始找别的视频，找到七月在线，开始跟着寒老师学习，从机器学习到深度学习，边学习边复制一些寒老师的代码(就是看代码，然后自己再抄一遍)；最后在研三上学期面临找工作很慌张，会的还是太少，于是参加了七月上海线下的深度学习课程，虽然当时因为要找工作发论文等事情学习的没有那么认真，但是课上学习的东西基本都掌握了，为后续找工作做了铺垫。

找工作经历：

大公司：腾讯：一面被刷了，不过那是大三刚开始的时候，学的不扎实，被刷掉是必然；平安产险：现场笔试，然后面试，连续两天；印象比较深的是笔试的内容很多挺寒老师讲过，所以答得不错，在人山人海中通过了笔试，第二天去面试的时候发现身边都是些 985 的男生，作为一个双非的妹纸瑟瑟发抖，在项目经理面前手撕代码的时候瑟瑟发抖，没有通过，不过面试官人不错，看得出我紧张，一直在鼓励我；还有一些银行的开发岗，也过了，但是都是 C++ 岗位，本身不太会，虽然面试官说语言是相通的，会 C 的话 C++ 很容易上手，但是最后都拒了。

小公司：收到很多家，有一家偏研究的图像识别小公司，岗位偏研究，算法用 matlab 和 C++，最后感觉跟我硕士期间学的东西基本无关，放弃；还有一些都不是跟计算机视觉相

关的岗位都拒绝了，最终在 18 年 12 月份签了一家互联网公司，担任图像识别岗位，现在主要做一些目标检测相关的项目。

总之找工作学习很重要，基础很重要，目标很重要，心态也很重要

算法工程师面经：我是如何从产品转到 CV 的

在研究生阶段是物体检测，用的是传统方法，一方面完成项目，另一方面也在自行的学习（对七月在线的关注就从自学开始。从最初的机器学习课程，到深度学习第三期，到 18 年的 vip，都是不错的课程。这些课程能够系统的整理知识点，一线的老师用一些实战的项目帮助大家学习、了解领域内前沿的进展）。毕业后我来到了一家国企，从事一些产品工作。

今年四月初，正处于换工作阶段。由于有一阶段没有对前沿科技、技术的摄入，需要一段时间进行补充。正巧看到了七月 cv 就业班的第一期正在招生，相比于个人的集中复习，报班学习能够提供基础知识巩固、技术大牛讲解、实验平台的使用，自己只需要闷头学就好了。也是为了让自己的学习有个“抓手”，就报了 CV 就业班。

由于我对自动驾驶很感兴趣，因此就比较关注自动驾驶的相关项目，主要准备的项目有道路交通标志的检测、车辆的重识别技术。除了深度学习，我将经历过其他的项目，如三维重建、点云匹配、传统的目标检测方法等。相比于专门在做一个岗位，我更希望有更多的可能性，因此面试的公司主要针对一些初创的小厂，岗位主要考虑传统方法与深度学习相结合的就可以。

下面简述一下我的面试经历：

1、AR 公司

三维重建的基本过程 / 相机模型、坐标转换 / yolov3 的原理

2、机器人公司

三维重建的基本过程 / 位姿定义 / 对极约束 / ssd 的原理

3、自动驾驶公司

笔试语言：c++ / 反转链表 / 密度聚类

一面介绍项目：

- 1、传统物体识别方法；
- 2、车辆重识别技术
- 3、项目中的细节处理思路和逻辑
- 4、常见的局部特征算子（主要是 sift）

三面（人事）

离职原因

最近在做什么

职业发展规划

能否适应加班节奏

最后拿到了 AR 公司和无人驾驶公司的 offer，选择去了无人驾驶公司。虽然工作强度会大一些，但对于我来说，有机会挑战新的方向，为啥不试一下呢？

感受：七月在线的 CV 就业班有一线大牛讲解前沿的项目、协助调整简历、准备项目、整理面试策略，让我避开了很多面试的坑，按部就班的学习也提高了这段复习时间的利用率。当然，除了七月在线的课程之外，语言基础，leetcode 啥的，还是不能落下的。

总结：

一、准备阶段要明确自己的目标行业和岗位。有了明确的目标，针对性的准备自己的项目是性价比较高的操作。各个行业的视觉技术的侧重点都不一样，如果有一个对口的项目会拉进个人与面试官的关系。

二、对于一些项目，最好有成品或者效果，会更方面面试官理解。

三、自动驾驶行业需要有一定的 c++ 的基础，leetcode 还是不要放松的。

社招转行 CV 面试经验分享：面试当场秀代码

本人普通本科毕业三年，电子相关行业，报的机器学习集训营第七期。目前只能说是跨进 CV 的大门，离真正的 CV 算法工程师还有很长的路要走。社招面试更看重的是项目经验，问的比较多的也是项目中的一些问题，对于一些基础知识的考察比较少，应该是默认都会的了，由于只面试过一些中小厂，所以一些经验可能不适用于大厂面试和校招。

中小厂面试一般 1-2 轮，公司更看重业务能力，所以面试更多的是根据简历中的项目和工作经历来提问，简历中有几个有含金量的相关项目经历即可，不需要写很多其他的内容，写到简历上的项目务必确保真的理解和熟悉，因为面试官一般会顺着你项目中用的技术来延伸提问。公司的产品最终都是要落地的，而工程落地一般是 C++ 实现的，所以面 CV 岗有 C++ 经验绝对是个加分项，自己之前一个面试就是由于 C++ 不过关而直接挂掉的。

现在的 CV 算法更多的是用深度学习来做的，但一般公司对图像的前期处理还是会用 OpenCV 来做，所以一般也会问到传统的图像问题，OpenCV 掌握一些常用的 API 即可，有时间的可以深入学习。

面试一般会问一些基础的深度学习的知识，自己面试中遇到比较多的问题是：

- 1、什么是过拟合，如何解决
- 2、1*1 卷积的作用是什么
- 3、数据不平衡的解决方法
- 4、什么是感受野。

除了这些，还被问到过传统的图像处理的知识，比如 SIFT，图像的边缘处理等，传统图像处理的知识有时间的话可以多看下，时间不够的话了解一下几个最常用的就可以了。

面试的大部分时间都是根据项目和相关工作经历来提问，比如用到的 backbone 网络的知识 (如网络的结构和 ResNet 能做到很深的原因)、训练中的调参问题 (learning rate 、batch size 等)、使用的优化器、损失函数和评估指标以及训练过程的可视化，对于数据的预处理或者是否做了图像增强以及为什么要这样做，样本中有没有噪声以及怎么剔除，分类问题中各个类别是否平衡等。

除了这些基础知识，面试官更关注的是运用这些知识来解决问题的能力，比如会问有没有自己设计过一些网络，最初的评估指标是多少，是如何优化提高的，中间遇到了哪些问题，这些问题的原因是什么，最终是怎么解决的，有没有更好的或者可替代的解决方案。

最后，面试官可能会假设一个业务场景，给出一个实际遇到的问题，让你给出一个方案，有时找不到最好的解决方法没关系，可以先给出一个可行的方案，然后面试官一般会慢慢引导你不断的优化自己的方案。

自己有做过一个 Kaggle 的比赛，排名还算可以，最后把代码放在了 Github 上，其中一个面试官想要看一个可量化的项目结果，于是自己把代码给他看，好感度立马上升了不少，最后顺利拿到 offer，如果面试官能认可你的项目代码，面试基本就没啥大问题。

最后的一些个人建议：

多看论文，了解最新的行业进展

多读论文，不仅可以提高自己的专业水平，还能给面试官很大的好感。除了一些经典的论文，其他的论文不一定要全部理解吃透，只需知道这篇论文主要用什么方法解决了什么问题，相比于之前的方法有了哪些提升，提升有多大，有哪些方法可以应用到实际的业务场景中

认真的完成一个项目

可以打一下 Kaggle 或者其他的竞赛，其他的项目经历也可，对于没有 GPU 资源的一些同学，七月的部分课程是有免费的 GPU 可以使用的，要充分利用好七月给到的资源，面试中问到的一些知识都是比较深的，如果自己只是简单的完成一个项目是完全不够的，很容易被面试官问跪，这也是为什么简历上的项目自己一定要真的动手做过，并且理解的很透彻提高编程水平，没事刷刷算法题

大家的工作内容主要就是写代码，所以编程水平的重要性不言而喻，中大厂的面试一般都有算法题，主要就是排序问题，比较难的就是动态规划的题目。没事可以多刷刷 OJ，尤其是对于非 IT 出身的转行 CV 的，多刷算法题提高自己的编程水平很有必要，OJ 的话选择 LeetCode 或者七月的 OJ 都是不错的

以上就是我的一些 CV 面试分享，希望对大家有用，同时祝大家找到满意工作，一起加油~

社招转行 CV 算法的面经：项目的问题一般这么问

我是 CV 就业班 1 期学员。硕士毕业在传统行业工作近四年，期间也有接触和学习机器学习/深度学习的相关知识。之所以决心转行 CV 方面的工作，主要是实在不想做之前的工作，觉得应该追随自己的心灵做自己想做的。

当然，目前 CV/人工智能这方面已经是高端人才稀缺、普通想上车的人一抓一大把的地步，自己这种状况很有劣势，但是我觉得自己没有退路，不干这个不赌一把还能干什么去呢？现实是困难重重，但是自己心底的念头也是不容易被压制的，不知道大家是否有这种同感。

大概两个月学习了基础知识和做项目，开始准备简历，开始找工作。由于我求职的地点锁定在某非一线城市，再加上我也不是计算机科班出身等客观不利因素，面临的机会相对较少。老师对项目的指导和简历的优化还是很到位的，我最终面了四家公司，拿了两个 offer，算是成功转行到 CV 算法岗。

某二线大厂 国企 内推 Fail

先做了笔试题，对深度学习部分，觉得做得较好；但是编程题部分几乎没有做出来，题目包括实现线程安全的 string 类、c template 实现 list 类、opencv Mat 操作编程；笔试还涉及到基本的图像处理的题目。不知道笔试的重要性有多少，但是面试官在面试的时候也看了几下笔试卷子。

面试官问了些项目相关的问题，比如为什么要用 triplet。这些方面我就回答偏了，我回答成了实现细节了。也是事后在老师帮我面试复盘的过程中，我才醒悟到自己回答问题的各种思维和思路的缺陷。

面试官还问了逻辑回归与线性回归的区别、用 tensorflow 实现逻辑回归（手写）、1*1 卷积的理解、python 装饰器这些基础知识，我自己觉得这些都答的挺好的，觉得自己基础知识方面学习的还算扎实，然而最终还是没通过，因为想想基础知识问答这种可能实在不是

社招的画风,社招更关注项目和经验,自己项目很多问题没有答出较深入系统的理解和思考,不通过也能理解。

某二线大厂 互联网 内推 Fail

由于是内推的,没有笔试。就是跟面试官随便聊聊的方式,很轻松和随意,面试官人很 nice,但是最终还是 Fail 了。可见,遇到很 nice 的面试官,似乎在谈笑风生,也千万别放松大意,一言一行都是在被评判的。面试官完全没有问基础知识,就是在聊经历经验和简历上的项目。一开始还问到了我传统的工作经历上的一些项目和经历,完全没有准备这方面,有些东西没有跟他解释清楚,囧。

面试官问了很多项目某个过程遇到什么困难、难点在哪里、为什么要用这种方法你是如何思考和选择的等问题,这些问题很常见,但是也并不好回答,我感觉自己并没有吸取到第一次面试失败的教训,回答的还是不好。还有就是并行、并发也被问到了,建议大家最好将这些并发、内存管理等常见的编程技术结合进项目和经验里面,不仅仅是要刷面试题。我当时回答自己没有深度学习中用过并发,事后再想想似乎不对,Tensorflow 的数据并行也可以算是并发吧。觉得面试中的表达与表现也是需要注意和加强的。

面试过后就没有消息了,想想自己面试过程确实有些问题没有回答好,而且可能这个岗位也不是很缺人,因为我之前在 boss 直聘上联系对方 hr 都没有人回复,都不收简历的。如果我主动点面试后联系一下说不定也是还有机会的,但也没有去管。

某上市公司 传统行业 投简历 offered

试过程也是聊了聊我的经历和项目,然后重点都在聊他们的 case 我会怎么做,数据将怎么处理、网络不收敛怎么办、类别不平衡怎么办、数据和模型都跟别人用的一样的,怎么做出比别人更好的结果呢?等等这些具体的问题。技术面之后等了会儿领导来面了下,也就各方面再了解了下。

某创业公司 内推 offered

面试前还跟之前一样，在紧张得准备各方面的基础知识，但是面试过程也还是没有问基础知识，主要也是在聊经历和项目。然后 HR 面了下，问了些 HR 会问的常规问题，考察性格、团队精神等方面吧。

最后总结一下自己的经验，觉得还是有一些感受和思考，希望能帮助到大家。

1、很羡慕那些还在校的宝宝们，社招转行还是比较痛苦的。早两年我就有这种转行的想法，但是全因为自己的怠惰吧。所以，在校的同学们要加油了，想想还是校招容易。社招转行要趁早，综合衡量下自己的背景、能力、经济条件，当前这风口挤满了人，比较艰难，当然也不是没有机会的，如果不像我一样固守一地的话机会更加多。我觉得做职业选择还是看自己的微观条件，风口和趋势像股市涨跌一样难以预测和把握，重要的是我们个人到底想干什么；同时保持终身持续的学习习惯。对于决心转行的同学，我觉得背景好能力强的可以直接去最好的企业；稍弱一些的可以先上车

2、项目最重要！简历上项目写不好没有面试机会，面试项目问题不好好准备也很难通过。项目涉及的各种技术及其关联扩展的详解、为什么要用这种方法思考过程是怎样的、有哪些难点、遇到哪些困难、后期做了哪些优化等问题，都要准备好再去面试。我在这方面也有一定的怠惰和心存侥幸。举个我回答失败的例子，面试官问我为什么用 triplet，我在详细说 triplet 是如何做的，甚至说出 triplet 是一种常见的方法这样 naive 的话。正确的方法应该是从面临的问题出发、从高屋建瓴的技术宏观视角出发，要解决什么问题，有哪些可行方法，互相对比一下，再看这种方法有什么优势。

3、基础知识（机器学习、深度学习、框架、C、Python 这些）在社招中应该已经不是加分项了，应该是必会题。事实上，一次面试就是一种画风，我遇到的问基础知识的很少，有些要问也是融入到经验中去问。社招也许大厂会问很多基础知识，社招面试的画风应该是聊

经历、项目、聊技术而不是问各种能找到标准答案的问题。当然，基础知识还是必备的，从长远看，吃这碗饭这些都是本应该就融入身体的东西。

最后希望自己新的工作能顺利，也祝愿大家都找到满意的工作。

零基础的“大叔”：如何从仅会点 C 到转行机器学习并拿到 56 万年薪

各位同学好！

首先感谢七月在线平台，让我成功转行并且拿到还算满意的年薪（将近 60）。先简单介绍我的情况：大龄（30++），非互联网行业，硕士，仅会一点 C++，还遗弃多年。

经同学点拨，决定转行机器学习（不站在风口上，怎么也飞不起来）。于是辞职，gap year 出去浪一年（完美错过了 AI 风口，并且赶上互联网裁员）。回来也是在同学建议下，报名七月。整个学习过程可以用惨不忍睹形容，编程基础薄弱，效率低，顶着巨大压力，咬牙学习三个月，后改方向（深度学习），又学习两个月。

找工作过程，相对比较幸运，因为有多年的研发经验，并且在项目准备上较为充分，所以还算是挺顺利找到满意薪资的企业。年纪大了，也不太想到互联网行业拼命。

回顾了一下，集训营对我的帮助还是巨大的：

- 1、 适合零基础或者基础薄弱的，譬如我。和我情况相似的人群，强推！！
- 2、 课程安排合理，集训营课程虽然不深，但是从最实用的角度，把在机器学习相关知识贯穿，形成一个很好的闭环。能够让你了解无论大厂还是小企业所需掌握的基础技能。
- 3、 实战应用还是非常好的，后期在实际工作中也可以借鉴其中代码和思想，对面试也有帮助。

学习过程中的一些建议：

- 1、 机器学习集训营，内容还是较为基础的，覆盖面比较全面，所以建议学习初期能够对机器学习、深度学习做一个全面的了解，结合自己当前工作尽早的确定方向。

我在这方面比较吃亏，按部就班的学习，后期确定深度学习方面发展，结果很大一部分时间算是浪费了。又非常痛苦的用两个月时间主攻深度学习。

- 2、 学习过程中，可以结合自己场景，搞一些相近的数据或者项目，尽早准备。
- 3、 集训营赠的课程可以先不选，等方向定了之后，再针对性的选择辅助课程。七月的老师还是非常 nice 的。
- 4、 深度学习需要 GPU，量力而行。
- 5、 时间有限的，确定方向后，完全可以选择性学习，更有针对性。

面试建议（适合刚转行过来的）：

- 1、 准备项目，越早准备越好，结合自己的实际工作或者找到合适的项目。一定要自己跑出结果，熟悉项目的每个环节。我是结合自己实际工作的项目（我也算是搞算法的），转化为适合机器学习&深度学习方法的项目。所以我在项目上准备还是非常充分的，目标、重难点等都非常清晰明确。
- 2、 工作后回顾，课程内容还算是入门级的，实际项目会遇到非常多的问题，经常要结合不同的技术，还有框架、效率等方面的问题。所以准备项目过程中，可以多发散一些思维。
- 3、 互联网公司对代码能力考核还是比较严格的（我代码比较薄弱，并且进入的是非互联网企业），建议有富裕时间的，可以刷一刷，当然在职的，建议以项目为主进行准备。
- 4、 时间有限的，一定要抓重点学习和准备，不建议什么都要兼顾，先入门，再跳槽嘛（目标是大厂的请忽略）。找准方向，抓项目、抓项目、抓项目。什么公式推导、代码熟练度我几乎放弃了，入职我推迟了几周，专门补的代码能力。
- 5、 智联上投了几个，可能是看我定位较高，没人理我。猎聘上更新简历，就有不少猎头联系。
- 6、 关于面试问题

从项目角度出发，说一说吧，简历上项目描述的每个细节要准备充分，均可能被问到。具体问题我不举例，可能不太方便。

- (1) 项目：描述、目标、项目的实际意义等
 - (2) 技能：使用的框架、语言以及项目中你担任的角色和任务
 - (3) 数据：数据来源、数据预处理、数据量等
 - (4) 方法：使用模型、技巧、技术等，这里面会针对性问一些技术层面的问题，需要做好相关准备。
 - (5) 结果：最终结果、遇到的困难&解决方案等，这方面建议发散思维，并一定要实际操作发现一些问题并且解决问题，效果可以略微夸大。困难和解决方案能很好的体现你的价值。
 - (6) 一些技巧：老师课程中闲扯的东西，最好都记录下来，能够让你面试的时候显得经验更为丰富。例如一般来说模型方面能够带来的收益是千分量级，而数据处理可以带来百分量级的收益等等。
 - (7) 我面试的前两家互联网相关的公司，技术问题问的我比较懵，经常问对一些技术如何理解的问题，get 不到面试官的点。后来的面试准备较为充分，也能够拿到更多的主动权。
 - (8) 传统行业的面试相对轻松，他们重点关注项目，因为他们的技术水平一般也不太高。
 - (9) 企业选择也有技巧，和你原行业契合度较高而薪资也 OK 的可以积攒一些经验后再投。
- 我比较幸运吧，没有面试几家企业，拿到了满意的 offer，这个有我其他方面的工作经验比较契合的原因。
- 7、关于书籍，我看的书并不多，西瓜书、百面等没看多少，因为后来转了深度学习，七月在线深度学习方面的所有视频刷了几遍。
- 8、小结：整个学习过程还是非常痛苦但是很充实的，最终的结果也很幸运。集训营的课程对我这种零基础、自制力差的人非常适合，他能够在最短的时间里让你建立起完善的知识架构。入门了，什么都好说。

励志！普通 IT 研发转行 AI 的学路、心路，和充满干货的面经

不是七月的就业老师催着写总结，我可能还在拖着不去回顾。人都有惰性，七月的老师催促自己总结这段经历，一如刚报名时催着大家预习、考试、提交项目。七月好像自身蕴含这种属性，催促你前进。

我是机器学习集训营第六期的学员，双非院校硕士。在校时实验室研究图像识别，但自己当时做技术实现，理论方面研究得比较浅。毕业后做了两年软件研发，发现实际工作跟上学时想的差别挺大，而且编程方面自己优势不大，于是觉得还是把以前接触过的图像识别加深下，顺带学下机器学习，转岗吧。当时实在是受够了长时间加班，而且对所做的技术方向有了较全面的认识，觉得发展很有限。

在新公司干了九天我就辞职了，辞职后真得觉得充满干劲，因为前面有片浩瀚的大海等着自己去冲浪。把不如意的现状打破，解下枷锁，全力投入新战场，即使暂时没有收入，即使积蓄仅够两个月生活。但是顾不得了，要干，就马上干。

我先自己闭门在家，看完了冈萨雷斯的《数字图像处理》，是真的看懂，不是随便翻翻。看完花了一个多月。后来报了一门课（还不是七月的）：深度学习在计算机视觉的应用。学了两课就震惊了：原来，传统图像处理已基本被深度学习碾压！而且当下 CV 应 FasterCNN，Yolo 等完全没听过。震惊之余充分认识到，闭门造车是绝对不行的，自学容易学偏，不得要领。正巧当时看了几部网上流出的七月在线的视频，授课时间很早了，视频还没有后来的 ID 水印。看后由衷地觉得讲得确实好，清晰明了。于是就搜七月在线官网，挑选一番，报上了机器学习集训营。也是咬牙决定的，要干，就马上干。

报名后又是一次次的意想不到，课程的侧重、编排与内容的设置，真的与之前闭门看图像处理时认为的天壤之别。这边绝对是按照实战需要进行的规划。而且，有几集课程和几篇

课件对理论概念的剖析，真的是神来之笔。这要比自己闷头看经典书省了大力气，比力气更宝贵的是时间。

再就是七月这边安排的节奏其实挺紧张的，我辞职全天候地学，都觉得吃力。当然用另一个词形容就是充实。课前送配套的预习课程，会反复提醒你看完。线上课讲解的信息量巨大，一般是连续 2 小时，我都是第二天再听才能勉强消化。真消化完，东西就基本掌握了。课后有作业，周末有考试。一遍遍催促你对知识咀嚼消化吸收。

关于项目和知识的讲解与总结，七月给了很多特别好的资料与案例，并且逐行讲解清楚，不明白的还可以再单独问，直到自己明白为止。

七月每期的学员，都是求变求突破的人。七月把大家集合到一起，共同讨论共同进步，这种氛围真的非常棒。如果自己一个人闭门造车，承受的压力要比这样大很多。

学完后已经 1 月份了，马上找工作心里没底，于是把简历让七月的老师帮改了一下，又温习了下课程，就过年了。打算年后赶春招。

春招真的不容易。我之前只简单做过一点图像处理，但是求职方向按照 NLP 准备的，数据挖掘类的也一起投，这就有点难办了。由于个人规划原因换过两次工作地点，加之简历润色得不太够，导致面试接连失利。春招面试持续了一个半月左右，拿了两个 offer，一个是某创业公司的 NLP 算法岗，25k/月但没过去，再就是我现在入职的这边，某高校下属科研院所的研究岗，兼带一部分开发的任务。下面是我把几次重要面试的情况总结如下，大家有兴趣可以参考一二。

A，某中型互联网金融公司。

简历上项目，最初提了多少套方案，为什么最终选这套；

机器学习训练误差由哪些构成；

偏差与方差；

降低过拟合的方式；

神经网络降低过拟合的方式（dropout，早停）；

BN 层作用；

BN 层加在激活函数前与后效果有何不同；

如何判断异常点；

逻辑回归的分布函数；

逻辑回归的参数求解在优化方面属于什么类型；

机器学习中常用 loss；

精确率召回率；

PR 曲线面积代表什么；

常见的语言模型；

文本表达方法；

新词如何发现；

句子中关键词如何提取（tfidf，textrank）；

如何计算两个句子相似度；

项目中遇到了哪些困难，如何解决的；

简历项目现在进行改进的话，可以从哪些方面；

seq-seq 有哪些结构形式；

soft-attention 与 hard-attention 的区别；

数据分类不均匀的话，要做哪些处理；

手写逻辑回归极大似然函数的数学推导；

Glove 与 word2vec 的比较；

讲一下 Attention 模型;

讲一下 Bert; 讲一下 fastText;

文本处理常用步骤;

分词分得不准确的话, 该如何处理。

B, 平安第一次面 (NLP) 。

DenseNet 的网络结构?

与全连接有何不同?

简历项目中网络用了多少层 FeatureMap, 每层面积?

项目用了多大的数据量?

达到了怎样的分类效果?

FeatureMap 提取了哪些特征?

画一下项目中采用的网络结构。

聊项目, 解释项目中为什么这么做。

词向量怎么训练的?

word2vec 两种常用模式?

词向量后面的 Softmax 如何优化?

CBOW 和 Skipgram 哪个更适合采用? (大规模训练的话 Skipgram 要更好)

Glove 的原理? (简历上有体现)

编程题: 连续数组中值的去重; 整数 $1 \sim n$, 每个数字的二进制表示中 “1” 的个数, 把此个数保存到数组。

C, 平安第二次面 (NLP, 不同事业部) 。

讲一下 fastText (简历上有体现), 说一下与 word2vec 的联系;

Huffman 树;

简历上项目如何设计的;

如果用 seq-seq 进行 embedding, 做相似度计算, 会如何 (替代 Glove 的话);

Glove 如何训练的, 用的多少维;

马尔可夫过程是怎样的;

怎么调试 tensorflow;

解释下 TextRank (简历有体现);

PageRank 是怎么回事;

为什么数据量越大 SVM 训练越慢;

SVM 原理;

分词是如何处理的;

语料中遇到新词如何处理;

决策树和 SVM 在数据预处理上有何不同 (缺失值); SVM 对文本要先做什么处理;

fastText 最初先对文本如何处理 (固定格式);

词性标注比较好的方法是哪些, 有没有最新的模型。

下面是上机编程, 记得有道题要求把不同的字符串进行分类。期间涉及到了对 python 的调试, 表现不够好。

D, 平安第三次面 (数据挖掘岗)。

先做了一份笔试题, 上面都是常规的概念问答。例如数据预处理、清洗的方法等等, 记得七月给的总结资料中几乎都有。

然后对着简历问。印象比较深刻的问题有: tensorflow 出错误了, 怎么 debug; 手推 SVM;

LSTM 的结构, 优势在哪里。其余的大多是常规问题。再就是根据简历问项目。

总结起来，失败的地方有这么几点。

第一，手写代码的功夫确实有退化，社招的同学务必花时间刷一下算法和编程题库；

第二，准备的时间还是有些短，三四个月的时间拿下理论和实战，会比较仓促，不够扎实，如果可以，延长到 7-8 个月，应该就不是当时那个心态了；

第三，最好还是不要离职脱产学习，如果脱产，请保证积蓄充足，不然拖到后面财务压力影响心情；

第四，工作头几年，请务必耐住性子，不要把简历跳花了，但如果确实有重要的抉择，优先级高于工作的问题，也不必在乎。

感觉比较感激和庆幸的几点：

一是遇到了七月这个平台，它不一定适合所有人，但是于我，帮助很大；

二是感谢线上的老师和深圳这边线下的老师，这个浮躁的时代，能有人耐下心来花大段个人时间，不厌其烦地向你传授宝贵经验，似乎已经超出了合同和契约，背后是有些情怀的。

三是认识了更多在这个领域奋进中的朋友。

要看的书有这么几本，百面机器学习，西瓜书，这两本要看熟（但很多章节可略过）；

统计学习方法，最好翻两三遍；NLP 做客服机器人的话，推荐《自然语言处理实践 聊天机器人技术原理与应用》，这本比七月对聊天机器人的讲解更系统全面

从测试到 CV 算法工程师的转型之路：最终连拿 4 个 offer

前前后后经过六个多月的努力，最后终于算是转行了.....

本人普通 985 硕士，研究生期间模式识别与智能系统专业，研究方向目标检测。刚毕业家里希望留在离家近的二线城市，但是 17 年秋招的时候，相关算法岗位非常少，于是先就业选择做测试工程师。

一年多的测试工作以黑盒测试为主，感觉自我发展空间确实有限，恰好一个师兄购买过七月很多课程，并最终转行成功，第一次了解到七月在线，便是通过师兄推荐知乎上七月在线的机器学习面试题了，由此便想着转行，希望通过在七月在线上的学习，使转行成为可能。一个人的学习过程中，知识体系通常很零散，虽然网上的教程很多，但是搜集教程并挑选出适合目前自己学习程度的课程也会花费比较大的精力，并且，网上的随处可见的教程、几个烂大街的项目对于简历过筛和面试还远远不够。3 月份看到七月在线推出了 [CV 就业第一期](#) 课程，于是立马就报名了。

第一期的课程包含的项目还是比较有含金量的，比如网络训练的 trick 等对于面试也非常有用。在 boss 直聘上自己也投了很多家公司，最后面试的大概是 6 家，拿到了 4 家公司的 offer。

1、深圳公司，方向医疗图像（无 offer）

一面，在深圳面试

全程主要问项目的情况，主要用了什么算法，

为什么用了这个算法效果就要好？

滤波算法有哪些？

研究生学习了哪些课程？

为什么离职？

对地域有什么要求?

为什么想转行?

二面, 视频面试

又把项目大致介绍了一遍, 后面主要是 HR 面

觉得自己有哪些优缺点?

觉得自己最遗憾的事情是什么?

为什么想转行?

自己的职业规划?

由于工作内容偏向于图像预处理及数字信号处理, 因此放弃了该 offer

2、广州新零售创业公司 (offer)

一面,

因为异地面试, 因此面试官要求做一份技术 ppt 用于展示

项目陈述, 主要用了什么算法, 解决了什么问题, 算法细节实现?

项目数据集是多少?

为何使用这个算法比另一个算法要好? 最后训练结果是多少?

二面

CTO 面试, 简单介绍了自己所做的项目, 选择自己最熟悉的模型介绍一下, 所以就介绍了

一下 FasterRCNN 网络, 并比较了 FasterRCNN 在 RCNN 系列中的改进点,

说明 RPN 网络的作用, 实现细节?

anchor 初始化的实现细节?

Faster RCNN 如何训练?

NMS 实现细节?

自己的项目使用了什么评价方法？含义是什么？

为什么想转行？怎么选择的方向？

未来 5 年的职业规划是什么？

三面

CEO 面试，同样是简单介绍了自己的项目经历，选择该公司的理由，

离职原因？为什么想转行？

觉得自己应聘这个岗位有什么优势？

CEO 面试中对技能，知识的考查比较少，主要看是否符合公司的价值观。

三面过后顺利拿到了 offer。

3、北京机器人初创公司（岗位不匹配）

一面，

技术面，介绍项目经历，

学习过哪些课程？

如何找出一幅图像中一个像素到另一个像素的最短路径？

一面结束后收到回复不匹配该岗位

4、上海独角兽公司（offer）

一面

介绍项目经历，根据项目经历询问，

了解哪些 anchor free 的网络，

了解哪些最新的模型？

是否会 CUDA 编程？

one-stage 和 two-stage 目标检测方法的区别和优缺点？

SLIC 算法介绍?

问的并不深，主要考查会的范围

二面

介绍了项目经历，然后二面面试官和我随意聊了会天

三面

HR 面，谈谈深度学习和传统目标检测的区别，聊了聊期望薪资，选择公司的要求

5、北京新零售创业公司 (offer)

一面

介绍项目经历，根据项目经历提问，

为什么离职？选择公司主要看中哪些方面？

未来 3 到 5 年规划？

一面为部门经理面试，考查综合性较强

二面

介绍项目经历，算法最终实现的效果？

为什么离职、转行并选择跨城市工作？

是否了解图像拼接，使用了什么特征描述子？

三面

HR 面，和以上 HR 面问的问题都比较类似，就不再列举了

最终也是拿到了 offer

6、西安创业公司 (无 offer)

一面，

手写快速排序算法，时间复杂度？

熟悉哪些分类算法？

使用决策树分类算法时，如何处理过拟合问题？

整个面试过程非常短，并没有涉及过多计算机视觉方面的知识，也没有问及所做的项目，在面试过程中已经感觉到和自己的方向并不相符，于是在面试过程中也没有过多的说明和展示，很快结束了面试。面试官说回去等消息。

由于自己是从家这边二线城市的软件测试岗位转到 cv 算法岗位，所以面试官都会问到为什么离职及转行的问题，技术面中 90%是根据自己的项目内容进行提问，因此项目准备很重要，对项目中的每一个细节要清楚，对算法的实现也要熟悉。因为简历中涉及传统的图像识别的内容，也涉及基于深度学习方法的目标检测内容，因此要对两者的区别有所理解。面试介绍自己的项目时，应该说明项目的难点，使用了哪些算法进行对比，或者选择该方案的原因，并适当带出自己了解到的最新论文，给面试官留下好印象。

推荐大家去刷一刷七月在线的[面试题](#)，同时多了解了解最近岗位的面经，这方面七月的一些关于面试的直播视频也很值得一看，最后祝大家求职顺利，拿到自己满意的 offer ~

大龄社招转行 CV 算法的一些心得：曾整晚整晚训练模型、调参、再训练

本来打算花三个月时间转岗的我花了整整六个月！

大概是今年 3 月份的时候,我实在是对当前的工作失去兴趣,每天胡乱的忙,没有目的没有目标,觉得每过一天都在浪费时间,浪费生命,于是乎起了转岗的念头.由于本科,硕士期间我有时候会看 July 写的一些算法博客,所以了解到 7 月在线,其实在 3 月份之前,大概去年开始我就在 7 月购买了一些机器学习和 CV 图像的课程晚上回去学习,发现自己对 CV 还是比较感兴趣,到了 3 月我做决定之后,参加了 CV 就业班的第一期.

本人研究生毕业后国内国外参加工作加起来已经 8 年,目前工作在一家小公司当产品主管,工作内容包括架构设计,研发管理,项目管理,团队管理,可由于公司后期研发投入太弱,大部分都是项目和开发任务,我的工作逐渐从软件强相关慢慢快变成为了一个 PM,并不是我个人对 PM 这个职位有什么排斥,而是我逐渐发现自己正在慢慢失去求知欲,疲于应付各种杂事,认同感缺失.

我的工作经历有些许机器学习的内容,比如用 SVM 做图像分类,比如用 Spark 做分布式的机器学习,但所涉及范围和深度都远远达不到立即转岗跳槽做图像的程度. 特别是深度学习,研究生阶段还以 SVM 和随机森林等算法为主,深度学习当时并未流行起来.所以其实对于深度学习的知识全靠自己自学.

报班前 3 个月每天白天工作,晚上上课学习,开始很有激情,通过三个月时间把神经网络和 CV 的基础知识网络抓了一遍,还参加了一些 Kaggle 的图像分类比赛,整晚整晚的电脑训练模型,调整参数,再训练,下班了晚上回来看结果,调整,在训练...如此反复,有了对于使用 CNN 进行图像分类的基本概念.但三个月过后,开始对自己有点怀疑了,因为 CV 涉及到的内容和范围

确实有点大,而且从分类,到检测,到追踪,后面面试中也会发现有时候需要运气.所以后三个月找工作的过程中有点心浮气躁,并未能在知识和技能上有前三个月那样的提高.

A 公司: 传统企业转型,做机器人及图像识别方面的工作

面试通过,并未有太多深入的技术面试,因为团队还在组建,顺利拿到 offer

B 公司: 大型国企

1 面:电话面 1 个半小时,从项目到对于神经网络的理解,对于结合项目中的使用,想法仔仔细细问了一遍,还好涉及的内容都是我之前有所准备的,或者是遇到过的,结果感觉还可以.

2 面:On Site, 面试官主要针对项目和本人的具体工作做了了解

Offered

C 公司: 新兴互联网 AI 公司,B 轮融资

1 面:了解我适合哪个部门

2 面:On site,相关部门的技术 leader 给我面试,上来就拿白纸让我写 NMS 的实现算法,我写完,又聊了其他的关于检测和跟踪的技术细节,还有过往项目中关于 CV 算法的实现细节.这部分需要好好准备,简历上所有写的东西都要有十足的把握说服面试官.

顺利拿下 Offered

总结一下,项目和行业经验对于社招很重要,深度学习结合自己对行业的理解,说服别人,不仅仅是技术的实现,否则对于大龄社招我又和应届毕业生有什么区别和优势?

双非二本面经，计算机视觉方向可以来看看

自我介绍

先简单介绍一下，本人二本毕业三年，之前在海康做 C++ 开发。机器学习集训营 4 期学员，现在在一家国企做图像识别，薪资 20k，低到爆炸，但五点半下班呀，哈哈哈哈哈哈
哈哈~开心 我有过两阶段面试经验，我慢慢道来

第一阶段

刚培训完机器学习集训营 4 期，说实话双非二本夸专业找算法真难啊！！我一共面试过 10 家公司，绝大多数都表达了不找转行的。所以大家面试的时候有过硬的项目当做之前的工作经历绝对很有必要，我是因为海康名气在杭州有点大，机器学习的项目根本放不到海康这种做 CV 的公司经历里，所以很难搞。

面试的问题并不是很难：

- 1.CART 树分裂过程？
- 2.信息熵怎么得到。
- 3.为什么逻辑回归叫回归。
- 4.为什么逻辑回归不用均方差。
- 5.loss 不下降的原因。
- 6.坐标轴中 n 个点，取出所有 $x_1+x_2=y_1+y_2$ 的点。

这些都是我当时没答上来的，都很基础哈哈。

第二阶段

由于对推荐系统之类的没啥兴趣，所以我放弃了第一阶段的一个 offer，时间推移到大半年后，也就是 19 年初。我在七月在线看完了计算机视觉的课程，专攻 CV，然后面试，说实话杭州的 cv 公司真心少，这次面试依旧避免

不了学历低和项目 easy 的问题，但只要问题拎得清，还是有机会滴。值得提的一点是：七月在线没有详细讲过 opencv, 搞视觉的同学一定要自己实践过 opencv 啊! 不会 opencv, 就别去丢人啦 (我已经丢过人了-_-|||)

面试问题：

1. 数据不均衡怎么办
2. 介绍最熟悉的网络架构：比如 faster-rcnn, yolo
3. batchsize 怎么设定，会有什么效果
4. faster-rcnn 里的 rpn 详细说一下
5. ssd 和 yolo 和 rcnn 的区别
6. 有哪些语义分割的网络，简述 U-Net 特点
7. 为什么要做 Batch Normalization
8. opencv 中可以用什么算法进行轮廓检测

总结

七月在线的课程给我打下了基础，大多数面试问题在七月的习题里都有，如果你不是硕士，不是 985、211, 还是转行的话，那你不仅要有七月学到的基础，还要扩充自己的知识，给个很好的途径，阿里天池某些比赛的决赛答辩视频，答辩者会提供很不错的 idea，把这些讲给面试官，绝对棒棒哒！

深度学习算法应届硕士秋招记录，最高拿到 52 万年薪 offer： 内含鹅厂头条面试真题及 offer 截图

本人背景是:211 985 软件工程本硕 本科均分 89 硕士均分 93 深度学习的项目有 3, 4 个，主要就是图像视觉方向。从 9 月来拿到的 Offer 一共有
腾讯，字节跳动，龙湖科技部，威马汽车，易车网，51Vr，周同科技，宽凳科技，美菜网，茄子快传，zenjoy

鹅厂是 9 月的事了，岗位是 C++ 开发，不是就不多说了，春招准备晚，一开始没敢投算法，怕机会太早用掉，分享一下相关面试题：

1 个人相关：你觉得自己的缺点是什么，硕士期间最大的成就是什么。

其实就是优缺点的回答，第一个千万别说什么太爱专研，太专心之类的东西。。我说的是一直在学校实验室工作，虽然项目多，但是深度都还不够，有可能自己的发散性思维没有得到太大锻炼，所以想在腾讯这个大平台好好历练一下自己。第二个就是别说什么主观的东西例如什么意志得到了磨练，品格什么，有了什么什么的精神，从你自己的专业水平来答，项目的成就啊论文啊研究啊

2 专业知识相关：多态，虚函数，const 与 define 的比较等等有些忘了，这个自己去搜吧。

手写编程题是：二叉树节点最近公共祖先 跟 数组中最长连续序列

offer 我 10 月份找算法的就拒了，30 出头吧：



然后就是头条了：

9 月出头来我校宣讲的时候，就说了今年算法爆炸多，竞争特别大，不管是想投算法还是投了算法还是面试中的算法都可以随时转岗开发，开发今年给 40。

头条简历过了后会一直给你发编程笔试，他们好像觉得硕士的项目都是垃圾，还不如会写算法题，是我面过写的最多的，笔试就是 5 道题，最好全 A，面试也一直写。

笔试的题目我忘了，网上应该都有，自己搜一下，有链表递归贪心最优。

面试面了三次，考的非常非常多。。。基础又细又多，每次都手写编程题了，因为我 10 月还串深度学习的知识点，有些没答好，所以还手写了好几个编程题。

问到的题目大概如下（记性真的捉急了，学算法的一定要早睡。。。：

1 LSTM 跟 RNN 的原理跟公式，图能画一下吗

2 BPTT 推导

3 Inception 机制是什么 卷积为什么要拆开

4 全连接层跟全卷积是什么，分别应用场景

5 机器学习了解吗，SVM 推导一下

6 python 比较熟是吧 讲一下 python 的垃圾回收机制还有迭代器

7 最大似然了解吗 交叉熵信息熵 kl 散度了解吗

8 做过 GAN 是吗，kl 很 js 写一下

9 无向无环图中，最短路径的最大值；分糖果（大家搜一下我不太记得了；还有一个 leetcode 有，电话棋盘，每次只能移到四个角的位置，最多可以打多少不同的电话，DP；还有一些零零散散的链表，比较挤出了。

10 深度学习问很深，当时很多答的马马虎虎，然后又填了接受转岗，最后拿到 offer 了，不过也是去开发了。



拿到了龙湖集团数字科技部门深度学习算法的 offer 了

龙湖还是一个很大的房地产公司，也上市很久了，最近在寻求转型，今年第一年招收人工智能，11.11 开始到 11.16，目标群体主要 985.211 大学，一天宣讲两次。这次只在这些大学招聘了



我也是直接接到 hr 电话，简历直接给 hr，得知我有腾讯跟字节跳动 offer 后直接告诉我，薪水可以在我原有 offer 基础加 30%。



这个项目士官生好像只校招，我是算法 深度学习那个。

面试过程也非常简单，12 点宣讲后一些人直接立刻终面，我是第二个，其实就问了些项目 + 有的 offer（不过提前一天 hr 让我打印或者截图出来 这种事最好不要撒谎 校招诚实最重要。

问到的方面:

自己想做的研究方向，导师跟同学怎么看你，讲一个最重要的项目

大概 15 分钟。然后就出来，领我去 hr 房间，直接给 offer 了。

一切发生的有点快，一开始很开心，50 万真的很高，回来冷静了一下，的确是新成立，然后正在转型，不如互联网大厂发展成熟，回来问了一下别的过了的，大概 35-40 吧，我的应该也是中等偏高了。

因为总裁直接终面没有笔试，我就只分享面试经验：

简历：简历很重要，废话一定不要写多，保留一些基本信息，算法岗，重点放在项目上。

3.4 个最好，国家级别的奖项跟论文专利可以写，别的就不用了

自我介绍：好神奇，没有让我介绍，直接问项目，但是我的意见是，准备3份：一份10几

秒 一份 30 秒 一份详细的可以讲久点，讲什么就人各有异，重点还是不讲废话，讲重点讲

优点

问答：不要撒谎 不要撒谎 不要撒谎 真诚看着面试官 有 1 答 1 就算不会也不要编 简历上的都要会 不会不要写 问到不会的了 可以向面试官寻求一些帮助都没事 重点是要思考 跟组织好语言 输出完你能想到的观点

其他：因为他们招的晚，好多地方没顾上，你拿到的 offer 就成为了他们考量的一个重要标准，像我的鹅跟头条，就给我加了很多分，所以建议大家大厂一定要投，一定要用心准备。

第四部分 NLP 自然语言处理

我的 NLP 秋招之旅：如何一步步准备并拿到大厂 NLP offer

Pain is inevitable. Suffering is optional.

首先介绍一下本人的简单情况。

本人是在校硕士，是机器学习集训营第六期的学员。本科学校 211，硕士学校 985。EE 专业。在校的实验室主要做的比较传统方法，有一些项目比较偏图像。看师兄师姐们的建议，已经当下的形势，决定还是学习机器学习方面的岗位。由于我个人比较认同兴趣驱动，结合个人的兴趣，最终选择了 NLP 方向。在学习的过程中发现手头资源不多，且难以调动，最终多方打听，选择了七月在线学习平台。主要是想通过集中系统的学习，来提高自己的实力，为秋招做准备。下定决心后，就开始了漫长的学习过程。以下是我的学习路线，请大家参考。

1. 西瓜书，蓝皮书。这两本是机器学习的经典之作，相信大家都不会陌生。个人的建议是，第一遍读的时候，太细节的地方可以放掉。先有一个大体的感觉。因为这两本书本来读一遍是不够的。我在看这两本书的时候，在书上做了很多笔记，都是自己的理解。再看的时候，说不定理解会变化，且更加深刻。
2. 七月在线课程。七月在线的机器学习集训营，有比较详细的课程计划表。说实话，时间还是挺密集的，我个人由于先看过书，但是有一些地方还是觉得内容较多，需要及时的复盘，补充。课程持续时间大概两三个月，也是一个集中学习的时间。当时我尽量把学校的事情完成之后，再来完成相关机器学习的学习的。现在想起来，那段时间紧张且充实。
3. 七月在线项目。老师给了一些项目让我们学习，这都是非常好的学习资源。不懂的地方可以问老师，问同学，七月在线给我们提供了一个良好的学习氛围。老师给的项目，有一些

简单的，有一些比较难的。我下来之后，基本都会自己跑一遍，对自己感兴趣的项目，还会自己重新敲一遍，特别是一些实现的细节，或者一些 tricks，我都会做一些记录。这些都是真实的项目技巧，也为我后来的项目做了准备。

4. 数据科学比赛。在上完七月在线的课程之后，一定要通过实际操作。一来是检验自己的学习效果，二来也是对自己能力的提高。多和比赛群里面的小伙伴交流，即使没进决赛也没关系，前期慢慢累积经验，逐渐会看到效果的。最终，我在天池的两个比赛中都拿到了前 1%，另外还有一个比赛是获得了最终的第五名的。这些不错的比赛成绩，以及在比赛过程中的一些困难，解决方式等细节，都是可以写进简历当中去的。

5. 找实习。对于校招生来说，没有相关的工作经验，实习其实对于每个人来说，都是比较宝贵的经历。很多秋招企业，也比较喜欢实习过的学生。我的情况是，找到了几个实习，但是导师这边不太肯放，所以我最终没有去实习。但是，即使没有去实习，我建议还是要去找实习，这样会给自己积累很多经验。最终拿到了三家公司的实习 offer，以导师不放为理由，遗憾拒绝。我虽然没有去参加实习，但是我利用这几个月好好的对之前的知识做了总结和复盘。并且还对于当前的各种 NLP 的新模型进行了学习。

6. 刷题。重要的一环，基本面试环节都会要做题。基本上面试的时候如果一道编程题都做不出来的话跟挂了没什么区别。建议在研一或者研二开始就坚持刷题，主要是剑指 Offer 和 LeetCode，我总共刷了 200 道左右。剑指 Offer 是赶时间用的，这本书一定要多看几遍，很多面试过程中遇到的题目就是原题。很多时候代码不一定要写的很精确，大概一个思想正确就好，面试官一般不会抠细节。所以，我的建议是思路至上，细节尽量。

7. 秋招。今年秋招，算法岗其实挺紧张的。正开始秋招之前，深刻感觉到今年秋招的难度比上一届更大。特别是 NLP，岗位的需求量比 CV 少很多。身边一些目标找算法的同学，最后没有坚持住，转了开发，甚至转了测试开发的都有。但是我凭借坚定的信念和意志力，最

终还是如愿拿到了 NLP 算法岗的 offer。也算是大厂的比较核心的团队。想来是自己的第一份工作，把方向定准，看好前景就是我最大的目标。

接下来，我总结了一些在面试中的一些问题。供后来人学习。

值得一提的是，面试官往往会根据你项目中用到的东西提问。所以不太清楚的东西，不要往简历上写，写上去的东西必须是清楚的。

1. GBDT , XGBOOST, lightGBM 之间的异同。
2. 随机森林与 gbdt 的异同
3. bagging、boosting、stacking 的异同。
4. embedding 的作用是什么？
5. 神经网络中的梯度消失和梯度膨胀是什么，怎么解决？
6. 激活函数的作用
7. 如何检验过拟合，数据量很小怎么办？
8. CRF 的算法的解释。
9. 介绍 momentum, RMSprop 和 Adam 等优化方式
10. word2vec 算法推导
11. word2vec 的优化
12. attention 的原理，计算公式。
13. seq2seq 的缺点，如何改进？
14. 循环神经网络 RNN 怎么解决长期依赖问题？LSTM 的结构是怎样的？
15. 怎么理解“长短时记忆单元”？RNN 中的隐状态 h_t 与 LSTM 中的记忆状态 C_t 有什么区别？
16. LSTM 与 GRU 关系是怎样的？

- 17.LSTM 和 GRU 的区别?
- 18.word2vec 和 fastText 对比有什么区别?
- 19.glove 和 word2vec 对比有什么区别?
20. elmo、GPT、bert 三者之间有什么区别?
21. elmo 为什么用的是两层单向的 LSTM 而不是 bi-LSTM?
22. Transformer 原理。手撕
23. 在做 NER 任务时, lstm 后面可以不用加 CRF 吗?
- 24.TextRank 原理?
25. BERT 原理。
26. BERT 适合哪些场景, 不适合哪些场景。
27. BERT 训练过程中用到涉及到的哪些任务, 这些任务如何去实现完成的。
28. 如何使用 bert? Bert 调参数的心得。
29. 聚类方式。K-means, 层次聚类, DBSCAN...
30. 样本不平衡的解决方法?
31. 介绍一下, 一些基于 BERT 的改进的模型。

最后, 在这里感谢七月的各位老师, 给予了很多帮助。

在以后的学习中, 我依然会选择七月在线, 认可七月在线, 相信七月在线。

最后的最后, 祝七月在线越办越好!

双非渣硕应届算法面经：这二三十家公司都面了我这些问题

本来不想写的，很惨，不过还是写写吧。

应届硕士，找工作从今年 3 月份春招找实习，一直到 10 月份秋招找工作，战线拉得有点长，但总算上岸了，拿了几个 offer，反正是一路坎坷。崩溃过，但都坚持下来了。

学校是非 985/211，学历硕士，研二下学期报的七月在线集训营第七期，学完以后就准备春招找实习了，本来是没意识到要找实习的，但看到周围的同学都在找实习，就写了份简历，跟风找实习。

试着投了一些大厂，有的简历没过，有的笔试没过，看到实验室大佬找到实习，很受打击。意识到可能简历项目没有写好，自己编程太菜，觉得编程这么菜，实习也没啥指望，即使实习了，秋招正式找工作，编程也还是不过关，就放弃了春招，专心练编程。反正你们别这样干，我实验室俩大佬都是实习转正了，秋招稳的一匹。

本人因为编程只会 python，所以刷题就用 python，网上找了个剑指 offer 的 python 版本，跟着刷题。别人都说剑指 offer 的题不难，可能我他娘的是个人才，坑坑巴巴的一天也就刷个两三道，慢的时候一天也就一道。

反正别管别人怎么说，按照你的节奏来。差不多从四月份到五月中旬基本刷了一遍。然后继续再刷一遍。差不多到了六月份，剑指 offer 上面的题基本做到看到原题，就能敲出来。后来就在 leetcode 上刷一些中等难度的题，因为自己三天打鱼，两天晒网，差不多也就刷了 50 多道。哈哈哈哈 刷题期间，为了保持自己对算法的学习状态，就跟七月带我报名的负责人商量下，又跟了第八期集训营。

集训营里面的课程，我是觉得最好能够跟上直播，有问题及时问老师解决，不要怕丢脸，不要怕问老师，反正有问题就问，即使老师没时间回答你，但是集训营的群里有些学员确实

很牛，在群里其他学员也会回答你。还有就是布置的课后作业最好都做了，很有帮助。不过，我比较懒没有跟着做，我都是等着都做完，看他们答案的。

8月中旬，我第二次跟的集训营课程基本结束。找负责就业的老师改了下简历，又找集训营带课的老师改了下项目，和技术栈。简历上写了一个比赛项目，一个智能问答的项目。然后，开始疯狂投简历。

招商科技，小红书，腾讯阅文，远景，三七互娱，欢聚时代，龙湖，字节跳动，依图，美团点评，经纬恒润，携程，奇安信，瓜子二手车，度小满，深信服。太多，一时间记不清了，反正差不多有好多家。校招投简历，一定要在牛客上找内推码，加入内推群。这样基本上简历都能过。

简历没过的：

深信服 nlp 岗，欢聚时代，龙湖（龙湖这个很早投了，据说工资开到一年 40，50w，但是简历投的数字科技里面的算法，到现在还没消息），OPPO，vivo，蘑菇街，猿辅导

笔试没过的

远景：投简历后让做一个英语测评，我六级都没过，做个毛线的英语，笔试没做，招商科技：编程题有一道没做，选择题好多不会，字节跳动：太难，不会，顺丰算法工程师，大疆，阿里（真难），奇安信，度小满

面试没过的

瓜子一面挂：介绍项目，链表，二叉树，梯度下降，通过硬币生成随机数，逻辑回归。

作业帮一面挂：自我介绍，介绍项目，然后直接编程。凉

小红书二面挂：一面，自我介绍，介绍项目，直接编程（leetcode 上的原题，秒解），因为简历上有个比赛项目，问了 XGBoost，问我做了多少特征，二面：还是聊项目，逻辑回

归，因为简历上比赛项目，问我优化方案，AUC，比赛多少人，自己排名多少（简历上我没写）

小米一面凉：问项目，word2vec,transformer,问我 bert,我没看，我说以前看过，现在忘了。。。。又问了 tfidf，让我写代码，不会，凉

字节跳动：一面面完，说算法岗满了，问我转不转客户端，（客户端我又不会，就说不转）就没下文了

后面：腾讯阅文一面凉，经纬恒润没有笔试，直接让面试的，做智能驾驶这块的，我投的nlp 算法,问我对智能驾驶了解吗，我会个毛线。凉，美团笔试也没做，后来让去成都面试了，当时心灰意冷，没去，携程也是没去现场面，
到这时候我的心态就崩了，所以心态很重要。

面完这些，发现自己跟大厂无缘，还有几家网申陆陆续续的笔试面试，都拒了，实在是糟心。感觉好难，都绝望了。

然后，打算跑学校宣讲会，我觉得当时这条路很正确，宣讲会流程快，笔试简单，基本上一两面，面完一周之内就有结果。不像大厂那么煎熬。

宣讲会：

广联达，宣讲会现场笔试，过，面试算法岗，结果面我的是搞开发的，问了项目，问我实验室方向，然后又问了我开发的问题，我一脸蒙逼，面试官比较好，又问我一些操作系统，数据库（因为我简历上写熟悉计算机基础），感觉答的还行，结果还是挂了

浪潮：这个是群面，三个面试官，面我们四个，看我是搞算法的，自我介绍完，问我会不会java,我说本科用java 做过毕设，然后让我转软件开发，当时没有一个 offer,我就说可以，过了几天，就去签了两方

成都的托尔思：宣讲会现场笔试，过几天给了面试通知，问项目，没有现场编程，问一些机器学习基础和自然语言的一些翻译模型，给过了，给了 offer,给的月薪 13.5k

四方伟业：宣讲会现场笔试，第二天面试，让做算法开发，因为面试官说它们做传统的机器学习，问我会不会用大数据做算法，sklearn 里面多少算法，我说不清楚，他说 200 多个，问我数据挖掘和算法开发的区别，问我自己 python 写过机器学习算法没有，也给了 offer

中兴：因为跟学校合作关系，我又是计算机的，基本上没咋问，直接让我转开发，也给了 offer

还投了三大运营商，不过这几个都不是算法岗，也给了 offer

秋招到这里基本结束，反正校招的话，心态很重要，工作肯定会有，只是自己是否愿不愿意去，笔试面试多总结，把剑指 offer，leetcode 多刷题，多看基础，语言不要只会 python，最好再会 java 或者 C，另外 sql 数据库最好会一些，有时候面试会问，还有你会 spark 肯定加分，我做 nlp 的，但是我也投机器学习岗，数据挖掘岗，不要只投自己那个方向的岗位，给自己多些选择，基本面试问的比较基础，自己的简历上的东西，自己一定要搞透彻，别乱编写简历，给自己挖坑。

多跑宣讲会，基本上宣讲会比网申靠谱，流程快，差不多到这里了，春招实习最好好好准备一下，我实验室的就有俩大佬华为直接转正了。

现在手里的 offer 都是来学校宣讲会得来的，大厂一个没有，拿到的 offer 有：拓尔思 nlp 算法岗，四方伟业算法开发，浪潮软件开发，中兴西安软件开发，联通省公司数据挖掘，电信省公司数据挖掘。打算先签个三方，然后看看明年春招能不能大厂捡个漏。哈哈哈

从迷茫不会写代码到拿到 34 万的 NLP offer：一定要多写代码

就业的史老师之前问我就业情况，也看到群里老师给了很多信息的帮助。现在回顾一下也挺好的。这篇文章，我会大致介绍自己的基本情况、学习的一个历程、面试的一些经验和寄语。

基本情况：本科双非硕士，目前已有的好点 offer 有某知名互联网公司的核心部门的算法岗，加补贴大概 34 上下，于我是很幸运了。硕士之前，没怎么写过代码。

偷懒，该还的迟早的还！我最开始是插班报的集六的课程。当时自以为研究生是计算机专业了，在报班之前，自以为是计算机专业，大家之前上过的一个多月课程，基本就没认真听，后面讲机器学习、深度学习的课程时，也是因为各种事（主要是自己懒）没有真正跟上，七月的老师一再的催作业，但是夜郎自大觉得听起来好像很简单，作业基本就没怎么做，到整个集六结束。

虽然那时候（寒假）老师建议我刷下剑指 offer。今年的春天实习招聘基本结束时，才发现我连个像样的实习都没找到，才发现老师虽然讲的明白，但是自己根本没认真学，算法功底不扎实，数据结构题也不行。出了问题了，没办法了，后面自己又加钱报了下一期的。后面的这段时间里，自己反思了自己的行为，虽然一边在实习，但是尽量抽时间重新把新课的清楚，你还得自己会真的去推内容重新学了一遍，主要是算法推导上，有些东西你听的懂不是你真的懂了，是老师讲的比较导，去写才行，千万不能偷懒。另外还有老师给的部分代码一定要自己去运行，去写，最好能自己从头写！我当时不扎实的工作，导致我后面秋招一再碰壁，一度自闭。

我秋招找工作的一个大致情况，大概笔试面试有好多家公司（emm 包括校招提前批将近三位数），有很多是在笔试的时候就挂了，其实自己应该早点刷题，现在想想，其实校长在一开始上课的时候就有提到四大金刚，其中数据结构题（比如剑指 offer、LeetCode）

就是个敲门砖，而且确实刷题也真的能提高编程能力（个人意见），保证至少基本的剑指 offer 刷完吧，66 道题，其中有个别难偏的题确实可以跳过。一遍不可能一次性刷完的，多刷几遍，熟练度很重要，在线笔试与面试，都是需要自己手敲（手写）。

秋招整个战线拉得很长，面试中遇到了不会的问题，记录下来，再回过头来看七月的视频。比如手推 xgboost，AI LAB 陈博士的视频就讲的挺好的。这样反复的迭代，另一方面自己在实习的项目，也反复的去抠细节，把代码与原理弄得更加透彻。

一次失败不要紧，我有个 P 大的朋友（本科非计算机，研究生也过的较水），前后笔试面试有 118 家！他大概到了 100 家才收到第一个 offer，所以大家都一样难。我相对幸运，不会一定要利用好咱报班的资源，反复的去看咱们的课程与搜索资料，还有去群里找老师提问。我找找实习时用的项目，就在群里提问了很多次。要相信自己，随着自己能力的不断提升，一定是会有好的 offer 的，我和他最终去向都还不错，从 0 到 1 突破了，后面就很快了，希望大家一定要坚持！坚持去反思与总结，并坚持学习。算法肯定是未来的趋势，而且待遇也是真的很好，搞开发的要 sp 才能拿到 30（除了头条），但是算法普遍白菜就有 28 了。很庆幸，自己坚持过来了，其实收到第一个 offer 通知时，我眼睛是真的湿润了。。。要相信天道酬勤！咱们能花钱报班，有想转变想改变的志气，就一定要用行动去提升自己！而且，算法虽然竞争比大了，但是更说明算法这个岗位确实是很棒，不然为什么那么人都争着往这个方向转！

一些杂七杂八的话：我报班其实是借款报的，但是现在想想，正是因为当时是借款报的，后面啪啪打脸的时候，我就想到一定要对的住自己花的钱，一定要努力，更何况只要自己愿意学，七月的老师还是很愿意教的，这一定程度上促进了我后期的更加努力。

一定要写代码，一定要写代码，一定要写代码！你不要抱着报班就能 100% 找到好工作的想法，如果你花个一两万块钱，就能让你 100% 实现从年收入 15 万到 30 万的转变，那

这肯定谁都愿意干啊!!! 还轮的到我们这样的普通人吗?! 一定要好好利用教育资源与自己买的教育服务, 好的东西一定是值钱的, 自己的能力是值钱的, 教育服务也是。对自己投资, 尤其是教育的投资, 确实回报最大的 (算是给他们打个广告吧, 但是这确实是我的观点) 最后, 其实还是需要好好感谢七月的老师, 尤其是褚老师的课程, 以及助教余远根, 他是我见过超级认真且耐心的助教老师 (校长, 这样的助教老师应该多招要加鸡腿!), 当时那个 QA 的 demo 完全是靠助教他写的注释才搞懂是怎样一回事。

对了, 还有当时服务器平台的运营老师, 网名应该是叫 Sniper (小编注: 现在是范浩管 GPU 了), 人也很有耐心。谢谢北京杨老师课程的精彩讲解, 他在讲授满是经验的时候, “大家要提升的是能力, 能力上去一定会有好的待遇” 非常认同了, 也成为了我后面与现在的信念! 还有谢谢就业老师们的信息提供与讲解, 谢谢曾经帮过我的所有老师、助教与同学! 面试主要都是围绕简历来问, 我感觉不同人的简历不一样, 参考性不大。

附近期面过的几家公司吧:

头条面试

整体面试小哥考核基础考的比较深入, 注重算法本质的与原理的深入考察

1) 写下 bilstm+crf 的损失函数? 损失函数中的 max 有什么作用, 为什么要减去这个 max?

-->具体的 emission score 与 transition score 是如何得到的?

2) RNN 与 lstm 有什么区别? lstm 解决了 RNN 的什么问题? 如何解决的, 写下解决的原理 (其想要的是本质答案, 关于激活函数的多次幂-- (好像是反向更新参数时 RNN 的激活函数 f 为指数次, lstm 变成了求和的形式-))

lstm 解决的是梯度消失的本质原因

transformer 整体架构 transformer 的提出 Q、K、V

Multi-Head attention、多头的理解

Position Embedding 位置编码

Mask 操作

粉笔蓝天科技算法岗：

除了问实习经历中的项目以外，两轮面试中还问到了：

推导 LR 的损失函数与梯度下降更新参数（必须要手推一下，当时就没准备好）

如何理解过拟合，怎样解决过拟合现象（当时没说全，过拟合现象的原因没有讲清楚）

如何理解 dropout 机制，dropout 机制是在解决一个什么问题，其中的原理（需要再整理）

L1 正则与 L2 正则的作用

LSTM+attention 机制是如何做句子分类的（这个当时没有说清楚）

实体抽取的项目中，bilstm 是如何和 crf 链接的？bilstm 起的作用是什么？能否用 cnn 代替？

（这个当时应该没答对，重新再查一下）

在 xx 的实习中，做的模型调优工作有哪些（当时答的不全）

Transformer Encoder 与 Decoder 有哪些不同？

Encoder-Decoder attention 与 self-attention

mechanism 有哪些不同？

multi-head self-attention mechanism 具体的计算过程是怎样的？

Transformer 在 GPT 和 Bert 等词向量预训练模型中具体是怎么应用的？有什么变化？

NLP 特征抽取机制有哪些？（基于 DN），各有哪些优缺点？

Multi-Head Attention 和 Scaled Dot-Product Attention

为什么是缩放点积，而不是点积模型？

为什么是双线性点积模型（经过线性变换 $Q \neq K$ ）？

相较于加性模型，点积模型具备哪些优点？

多头机制为什么有效?

算法题:

二维数组中找值, 一个很大的二维数组, 满足, 同一行中从左往右, 元素值递增。同一列中, 从上往下递增, 写代码。

(从右上角开始比较, 注意一行内用二分查找算法) (当时二分查找的边界好像没写对!)

给定一个整数型递增数组, 元素值不重复, 求其能组成的所有元素个数 (先转换成字符串, 然后变成字符串的全排列问题—当时全排列的划分问题好像没写对)

小黑盒 max 算法:

整体问了很多计算机基础的内容 (面试官是清华 cs 专业的)

提到相似度项目的 F1score 很高, 能否解释下原因? 解释下 F1score 的公式是怎样的? (当时公式都没有答清楚, 这个必须要会手写)

解释 acc 与 precision

Python 的数据库连接操作用什么实现, 服务器开发时连接数据库的原理是什么

线程与进程区别和联系, Python 线程与进程的特点, 多线程如何处理

面向对象设计语言的特点

怎样理解 http, 进一步提问计算机网络的七层协议

算法题:

手写快速排序的变形题, 问了有比快速排序时间复杂度更低的算法不?

9 月下旬面试

1. 双向 mask 机制的理解, --这个可以参考下 paperweekly 公众号 (知乎或者微信都有),

从语言模型到 Seq2Seq:transformer 如戏, 全靠 mask 的文章

2. attention 为什么比 RNN, CNN 好。

--attention 序列的对齐机制

3. transformer 为什么比 lstm 好。

4. 预训练向量模型进化的理解, 从 word2vec 到 bert ,到 ERNIE 与 XLNET , Bert 与 ERNIE 和 Xlnet 需要有所了解

Android 三年转 AI 狗血经历：曾报一培训机构但没找到工作

本人情况：数学专业，二本，Android 开发三年

2018.5.1 开始学 ai，当时一边工作，一遍学 ai，偶尔还要根据工作需要学些 Android 知识。所以天天熬夜学，找了一家培训机构，一直学到 2019.2 月份，奈何市场变化太快，简历投出去直接没动静，那家培训机构讲的也不系统，所以找工作非常被动，后来自己把图像的东西补了下，简历投出去依然没动静。

6 月份的时候，一起培训的一个朋友给我一个链接，就是七月的网站，我仔细看了几天七月的课程，觉得不错，就准备报，一起的另外两个不准备报，因为觉得学费贵(19000 大洋，的确不便宜)，因为我知道自己想要的是什么，所以直接就报了课程，他们两个纠结了一段时间我给他们陈明利害后，他们也报了，我和其中一个报的 cv 就业班 2，另一个金融就业。报了七月的课程，还没有到开课时间，不过助教已经给了我们说了要学哪些东西，而且很负责，会问学习进度，即使是周末有问题问了，助教也会回答（助教周末也需要休息，但是问他们问题了，他们只要看到了就是在外边只有手机也会回答），开课的时候自己就按照老师的课程节奏来学东西。

从 5 月到 10 月，自己的安排。开课前：重点学了上一期的 cv 就业班课程，以及一部分深度学习的课程，leetcode 刷了 50+ 题，动态规划，python 基础等。开课后：主要是跟着老师的课程走，虽然课程的内容我报七月之前都学过，但是通过老师讲解之后理解才更深刻，以前没弄懂的，现在都弄懂了。然后就是做项目。

在七月的学习过程中，自己有做的好的地方，也有不好的地方，现在列出来供大家参考。

不好的地方：

1. 和老师、同学沟通太少，七月的学员卧虎藏龙，跟学员老师多交流，既能防止自己一个人学习闭门造车，又能形成良性的竞争，鞭策自己前进，否则一个人学的话，极易自我满足，自我懈怠。我前期交流的比较少，所以才开始有些懈怠，因而也落后好多。
 2. 上课笔记做的太少：老师上课的每一句话都可能是一个面试题，所以老师的课程需要多听几遍，一定要有详细的笔记，不然马上就忘了，而且更多的情况是书本一打开什么都会，书本一关上感觉自己什么都懂但就是一句都讲不出来。笔记/博客是帮助自己整理思路的过程，自己不动手整理，思维的触角根本深入不到那些细节，而这些细节才是理解某个知识点的关键。自己开始为了图快，没有整理笔记，现在只能从头开始。
 3. leetcode 刷题太少：100 道左右是必须的，少了数量不够（当然越多越好，但是时间有限啊），才开始算法题 1-2 小时做一道很正常，面试会现场考算法，如果自己练的不够，或者自己没有深入思考过，给你一个题，基本一句话也说不出。我就遇到这个问题，其中一个面试就是这么挂的。
 4. 多动手做项目：在正式的大项目开始前，自己最好能动手做 1-2 个项目，哪怕 MNIST 自己动手从头开始写，自己调几次参，CIFAR-10，CIFAR-100 调参，kaggle 上调一个项目，能动手做一下。否则等真正的大项目开始后，自己根本不知道如何下手，更别说跑出来一个较理想的结果。而且也不要为了验证 trick 而去调参，即不要为了调参而调参，因为我在和一些学员交流的时候发现了这个问题。解决问题的思路很重要，为了调参而调参根本没办法训练思维。
 5. 不要辞职学习，因为我报七月之前已经离职了，所以这是自己一个非常大的失误。离职学习除了经济上的压力外，心理上的压力会更大，对后面找工作非常不利。
- 个人觉得可以推广的地方

1. 关注一些自己从事方向的公号，论坛等（ai 所有的方向也要有所了解），每天去看看发生了哪些进展，工业界学术界有什么新的成果，有什么新书出版，让自己对整个行业的发展有个概念，ai 更新换代这么快，不关注最前沿的东西大概率面临被淘汰的命运。
2. 在准备面试的时候，可以把写好的简历分解成一份文科期末考试试卷，即名词解释题，简答题，论述题，计算题等。当作一场考试(是真的用笔/电脑去写)，看自己不看书，这些问题自己可以回答到什么水平，这个水平基本就是自己对算法理解的真实水平，帮自己找简历中的漏洞。平时也可以用这种方法帮助自己理解一些算法。
3. 每次调完参数，产生一些新的感悟，立刻记录下来，日积月累自己对算法的理解才能越来越深刻，而且这些东西到时候可以直接在面试中讲出来，不用搜肠刮肚的去想怎么讲项目。
4. 每次面试的时候录音，回来检查自己的知识漏洞以及表述问题。

面试问的一些问题：

一 机器学习

1. 讲讲 svm 的核函数
2. svm 经过核函数变换后，最重要的是做什么
3. 讲讲 RandomForest，决策树，ID3,C4.5 ,CART
4. numpy 中的数组和 list 有什么区别
5. k-means
6. PCA，KPCA
7. GDBT，XGBoost，LightGBM 以及与 RandomForest 的区别（这个自己完全没掌握）

二 深度学习

1. 讲讲 resNet
2. 说说 bn 的作用

3. 说一下 word2vec
4. RNN, LSTM
5. 目标检测中, boundingbox 是如何调整的
6. 目标检测算法中有哪些分类, 分别讲讲
7. 现在让你提高模型的精度, 你有什么办法, 提高模型的召回率, 你有什么办法
8. 有哪些优化方法
9. 你是做图像的, 现在让你去做一个 nlp 或者语音的项目, 你准备怎么做
10. 给定一个网络结构和 loss, 现场推导反向传播算法。
11. 1x1 的卷积核有什么作用

三 编程能力

1. 这个一般会考算法题, 数据结构这块一定要清晰, 这个时候 leetcode 刷题就很关键了, 七月题库里面也有许多算法题, 要多做做。
2. 给定一个数组, 里面有各种数据类型的元素, 现在让你找出一个连续的子数组, 使得其所有元素的和最大(动态规划问题)

3. 给你个 csv 文件, 有时间和温度两列, 根据时间求温度的平均值, 这个编程题你怎么做

四 讲项目

这里如果自己对项目比较熟悉的话, 可以说是面试加分项目, 但是还是会有自己没考虑到的问题, 问的问题大概如下:

1. 你网络输入的 tensor 是什么样的, 最终输出是多少
2. 为什么要选用这个网络, 网络结构是什么样的, 网络中 64,48 是什么意思
3. 有多少照片, 最多的类别是多少张, 最少的类别是是多少张
4. 如何处理数据不平衡问题, 对应的 api 是什么, 在什么地方加入数据不平衡处理

5. 最好的结果是多少，你的结果排名多少

6. 学习率多少，采用什么的衰减策略，为什么用这个衰减策略

五 其他问题

1. 自我介绍

2. 你是做 Android 的，为什么放弃自己的优势跑来做 ai

3. 你是数学专业的对吧，谈谈你对测度论的理解(这个是选答题，我也回答的很不好，测度论，懂的人都知道)

4. 讲一下 java 的内存模型(自己提到了不懂 python 的内存模型，只知道 Java 的，算是给自己挖了个坑，不过自己回答过关了)

六 总结

1. 回答问题的时候，千万别问什么就回答什么，比如问 resnet，最好从 vgg 开始讲，然后是 resnet，inception；densenet 可以稍微提下，算法产生的背景，解决策略，最近新的成果，以及同类的其他解决办法，异同点。从横向和纵向两个角度来讲效果会比较好，这也是自己学习的时候路径，当然还有个程度维度，因为自己不可能一次性学到位。

2. 可以适当提下最新的研究成果，不用答得很细，比如问优化方法，可以提下 RAdam，所以适当阅读一些论文也是有必要的，虽然不是重点，但无论对自己知识的更新还是对面试都有帮助。

3. 有问题多找老师，助教，就业辅导，他们人都很 nice。多和学员交流。

非常感谢七月这个平台，不仅帮助自己成功入行 ai，也认识了很多有追求的同龄人，一起成长。虽然拿到 offer 了，但真正的挑战才刚刚开始，还有许多不会的东西需要学，所以接下来还要拼命学。

祝七月越办越好！

坎坷的学习和面试经历：5 年 Android 成功转行 NLP

我是传统 IT 行业从事 Android 开发，工作 5 年了，目前感觉上升空间较小，随着年龄的增加对于自己目前的现状有些慌，一直在寻找一些新的机会进行尝试。偶然的机会从朋友那里听到他参加了七月在线的机器学习培训，但是最初没有做好准备，虽然一直对于 AI 有接触，（公司产品有用到语音识别相关的功能）但是也只限于皮毛，没有深入了解，最初没有立即报名学习。

经过一段时间的反复考虑，在妻子的鼓励下报名参加了机器学习的训练班。最初的期望还是很高的，但是说实话在职学习的压力还是很大的，每天除了完成工作任务之外，每晚的学习时间非常紧张，基本只有 2-3 个小时的时间，最初的阶段应该和大部分同学一样，感觉压力并不大，但是到了深度学习阶段难度陡然上升，不巧这段时间家人和自己的身体健康都出现了问题，频繁的往返医院，学习时间越来越少，后来自己由于身体原因做了一次手术，这段时间的学习基本上不能持续了，只能安心修养身体，手术恢复后回到公司优先处理了患病期间的一些遗留的工作，准备重新开始学习。

在此很感谢七月在线的老师，向老师说明了我的情况之后，老师给我申请了重新个随下一期学习班的机会，这次机会对我来说很宝贵。就这样我开始了新的一期学习。在这一期的学习过程中，算法部分对我来说还是难度比较高的，可能工作比较久了数学基础还是比较差。公式推导对我来说还是有些困难，但是随着时间的推移，课程反复听和思考之后，勉强能够跟上老师的思路了，一些基本的公式和概念能够理解了。但是自己推到的话还是有些难度，而且公司的工作强度也越来越大，课程难度和安排也越来越紧张，只能挤时间出来学习。学习过程中的实践项目讲解还是比较合理的，而且关于项目和知识的讲解与总结，七月给了很多特别好的资料与案例，并逐行讲解清楚，不明白的还可以再单独问，直到自己明白为止。

我自己求职选择的方向是 NLP 方向, CV 等方向对于自身的背景要求还是比较高的, 因为自己是转行的, 所以还是没有很大的信心, 简历让七月的老师也帮忙进行了查看和修改, 目前也参加了几次面试, 对于自己的一些经历还是可以一起分享一下

第一家、某小型互联网创业公司。

大概问了如下一些问题:

1. 损失函数有哪些
2. 欠拟合, 过拟合以及降低过拟合的方式;
3. 神经网络降低过拟合的方式 (dropout, 早停);
4. BN 层作用;
5. BN 层加在激活函数前与后效果有何不同;
6. 精确率召回率;
7. ROC 曲线;
8. 常见的语言模型;
9. 文本表达方法;
10. 新词如何发现;
11. 常见分词的算法 (tfidf, textrank);
12. 如何计算两个句子相似度;
13. 项目中遇到了哪些困难, 如何解决的
14. Glove 与 word2vec 的比较;

第二家、平安 NLP

讲一下 fastText (简历上有体现), 说一下与 word2vec 的联系;

简历上项目如何设计的, 遇到过哪些印象深刻的问题;

如果用 seq-seq 进行 embedding，做相似度计算，会如何（替代 Glove 的话）；

Glove 如何训练的，用的多少维；

马尔可夫过程是怎样的；

怎么调试 tensorflow；

解释下 TextRank（简历有体现）；

PageRank 是怎么回事；

手推 SVM，介绍 SVM 原理；

LSTM 的结构，优势在哪里；

介绍一下 RNN；

分词是如何处理的；

语料中遇到新词如何处理；

决策树和 SVM 在数据预处理上有何不同（缺失值）；SVM 对文本要先做什么处理；

fastText 最初先对文本如何处理（固定格式）；

词性标注比较好的方法是哪些，有没有最新的模型。

第三家、猎户星空 自然语言处理

先做了一份笔试题，上面都是常规的概念问答。例如数据分析，特征提取的方法等等，记得七月给的总结资料中几乎都有。

然后基本对着简历问，包括简历上写的 SVM，LSTM，RNN，决策树等相关知识，另外问了一下实际项目中用到的算法，当初为何这样选择，常用的一些调试手段，项目过程中遇到的问题，如何解决的，当有一个新的棘手的问题出现的时候，你的思路是怎样的等等

经历过一部分面试后，总体感觉手推算法的能力还是很欠缺，这个可能还是和数学功底有关系，另外知识点的细节掌握不够，对于知识的整体性和贯穿性理解不够，大部分内容，让自

己讲的时候容易出现不知如何组织语言，不知如何引着面试官的思路跟着自己走，其实还是知识的理解不够透彻，不够系统。

另外在职学习学习的时间还是太短了，无论是理论还是实践感觉都不够深入，建议有一定经济基础的还是能够全身心的投入学习，不然每天工作回家后脑子吸收东西的能力真的很差（我个人是这样的），课程结束后也需要很多的时间去消化吸收，而且身体很重要，学习的同时不要忘记锻炼身体。

推荐的书籍，西瓜书，统计学习方法，最好翻两三遍；Python 机器学习等，可以根据自己的需要去选择。

双非本科应届只学过 C 的 NLP 找工作经历

看了很多的同学的面经，也简单的总结一下面试的经验，说是经验也不算吧，毕竟也还没找到算法相关的 offer，在七月的官网上已经有很多同学分享的面试经验，他们有很多很多的干货，是很不错的资源，但没能让你看清当今行业的形势，所以希望你在看到这边文章时，能对当前的业界的一个比较真实的认知，从而可以比较一下自己目前的知识掌握是否已经足够，自己的付出是否已经足够。

本人因面试算法的经历不多，所以本文主要是介绍自己为主，希望能给你一些帮助。

本人背景，双非本科，19 年应届(本科)，通信工程专业，在没有加入集训营之前，有学过 c 语言，数据结构和算法算是入门吧。

在目前为止，自个儿觉得已经掌握的内容吧，传统机器学习方面：LR、SVM、DT、RF、AdaBoost、GBDT、Xgboost 这些算法的公式推理，当然大部分没有在实际的项目用过。深度学习呢，像 cnn、rnn、lstm、gru 这些的基础原理，长啥样什么也算是掌握吧。在 nlp 方向上，word2vec、textcnn、attention 机制、transformer、bert 这些内容问原理和画图，也算是没问题的。

项目是很重要的，我的简历上有写两个项目一个是让 AI 当法官，一个是智能客服系统，都是集训营提供的项目，自己也有打一次比赛，2019 ccf 金融负面及主体判定，主要是 bert 及其相关扩展模型的使用。

到此，也算你已经对我有一定的了解的，下面就来讲讲自己的求职经历：

在拉勾上投过 20 份左右吧，全部石沉大海，在智联招聘上也投过几家，有一家给了面试机会，但是因为有些事情没能过去，所以凉，之后转战 boos 直聘，也投过不少家公司，最终一家给了面试，电话面试，给了口头 offer，但是后面通知说公司并购了一个公司，那边的团队也要过来，有几个和我一样的应届，所以凉。

此时已经是国庆了，所以只能国庆后又投，国庆之后也是只有一家给了面试机会，目前主要进行了笔试和简单初面了解。至此自己的关于算法的求职面试也就这样了。整体下来，算是面了两家。

第一家(小公司)电话面试了，只有一面，主要问了下面的一些问题：

- 1、自我介绍
- 2、让我选一个项目跟他讲讲，说了比赛那个，只要是用了 bert。所以聊了 bert 和一些自己改动。
用这些方法或这些改动有什么学术上的思考吗？
- 3、用到了什么激活函数，怎么选择激活函数的使用。
- 4、讲一下高方差和高偏差。
- 5、讲一下 seq2seq。
- 6、讲一下 cnn，为什么 cnn 中要 padding，池化呢。
- 7、rnn、lstm 和 gru 的区别？
- 8、在自然语言中如何选择使用 cnn 和 rnn。
- 9、因为项目中使用到了 macro-f1，所以问了 marco-f1，为什么不用 Precision、recall？
- 10、bert 算是去年的工作了，有了解今年有哪些新的技术吗？简单讲讲。
- 11、简单介绍 SVM。
- 12、了解 python 的深浅拷贝吗？装饰器呢？

之后就是聊了一下学习方法。

第二家：目前只是笔试和简单初面，基本没有技术细节。

笔试：因为是小公司，所以没有考算法，只要你参加了集训营是没有问题的。

面试：只是简单聊了项目和自己的一些背景，问了 word2vec 的原理。

到此，算法的面试经历也算是讲完了，下面主要是总结和建议：

总结：自己背景不好，双非本科，无工作经历，校长说的四大金刚没能利用好，自己没能好好利用七月提供的资源，比如说算法 oj，

自己没有能去刷题，所以不敢投一些比较好的公司，导致少了很多很多的机会。

建议：

- 1、如果你和我一样的背景，那么建议你一定要花更多的时间去学习，包括 leetcode 的刷题，这一部分很重要，不仅可以提高自己的编程能力，还能提高你自己的自信。
- 2、在项目方面，不仅要去做课程提供的项目，还要自己去参加比赛，因为参加比赛真的可以收获很多，且可以给你的简历提高通过率。
- 3、课程一定要多看，多刷，一定要搞懂。
- 4、在简历上面的东西一定要搞得明明白白，不要似懂非懂，可以通过写文章来加深自己的理解。
- 5、如果你和我一样，那么我希望你在看到我这边文章的时候，付出跟多更多的精力、时间去学习，不要认为自己的知识已经够了，真的还不够。
- 6、最后，七月在线是一个很好的平台，一定要好好利用它的资源，包括课程、老师、助教这些都是你快速成长的好助手。

最后，祝愿所有同学都能有一个自己满意的工作。

NLP 面试分享

题主背景：

985，211 高校硕士，NLP 专业方向，

但是毕业后校招去了某国企，从事底层不用脑工作，不想虚度光阴，奋而干回老本行。

面试能力：

项目做过简单的，学校的项目是 java 做的，工作用的 python 比较生疏，大数据啥的也只是听过，有算法理论推导能力和简单的数据结构理解。

面试过程：

第一周：（全败）

理由：准备不够充分，经常是不知道面试官提问的点，把简单问题想复杂了，或者是话到嘴边说不出来，面试程序也经常会出问题，太过于紧张，没有找到面试感觉，项目漏洞没有补全。

应对方法：

记下自己不会的点，以及项目的盲点和程序的 bug，尽量拖着面试官多聊天，暴露自己的知识盲区，反正不想着过，不如多暴露，回去多修补。回家仔细修补知识盲区，面试到第四家的时候已经略有小成，但还是没过。

第二周：（全过）

理由：心态放轻松，多次面试发现面试官提问的点只有那些，面试盲点慢慢变少了，而且通过面试官的反应和提问水平，引出一些比较高端的问题，比如问我决策树，我从决策树讲到 xgboost，面试官就去找下一个了，但是代码方面还需要加强。

应对方法：

总结来看，第二周面的大多是国企和小公司，所以通过也在情理之中，最终去了小公司

面试总结：

心态调整好，最好找公司练练手，找感觉，然后多推导，多写，多练习

工作后回顾：

小公司就是需要多面手，什么都要弄一点，回看过去，不论简历还是项目，自己都是个 lowB，工作后发现工作的东西的确有的时候和面试没什么关系，但是可能会用到一些基础，更多的是个人习惯和上进心，也发现理论一定一定要结合实践，实践过后再反来看理论，真的通俗易懂。

最后，希望帖子对大家有帮助，愿大家都有满意的工作

NLP 面试总结

面经～面经～，说真的写这类东西都是考验技术含量的，这里感谢在老师孜孜不倦的教导下终于排出了这篇传说中的面试经验，也算是在七月的总结篇！

我是集七的学员，选的 NLP，之前是大数据开发的，双非辣鸡一枚，接触七月是在我一个老哥在七月学习之后的薪资让我有了怦然心动的感觉，之后一路顺利的加入了这个大家庭，学了一段时间后辞职做的《毕业设计》，后来找到了一家不错的电商公司，相比之前更满意现在的生活～废话不说，一下是我面试了一些比较靠谱的公司后后总结一些经验，也是我觉得比较重要的一些，也是常问到的地方。

数据来源？

数据预处理（处理方法有哪些），存储？

模型选定，为什么选这个模型，是否还有备用模型。

模型的度量指标。

模型训练和优化方法，具体看你项目里的来说明。

细节：

请简要介绍下 SVM

请简要介绍下 tensorflow 的计算图

欧氏距离、曼哈顿距离

overfitting

LR 和 SVM

BAGGING Boosting

Xgboost 优化点

谈谈判别式模型和生成式模型？

各种聚类算法的优缺点 划分：K-means 层次：AGNES 密度：DBSCAN 模型：EM

Xgboost 优化点、LightGBM 与 XGboost 的联系及区别，是都试过了么

tf-idf 怎么做的

为什么 tf-idf char 会对结果有提高？

RNN 怎么做的？

RNN 为何引出 LSTM，其优缺点，解决的问题时？

glove 和 fasttext 怎么训练的？

词向量如何训练的？（word2vector）

word2vector 为啥语义相近的词词向量相近？怎么达到效果的？

RNN 网络结构，各结构有什么作用？

LSTM 输出是所有时刻的？还是最后时刻的？

简单说下 seq2seq

lgb 怎样的？缺失值处理过程？特征如何做的？

之后都是说自己项目遇到的难题以及解决方案，加上 linux 系统的一些命令，以及各种编程语言的使用方法（自己简历上写的），自己项目不熟悉的地方千万不要写上去，不管是语言及各种编程工具的使用等，面试时一定要礼貌耐心的回答问题不要急，急了会忘记东西，回答问题会就答，不会的话不要乱猜，猜错了印象会变的很不好，减分操作～

手写代码是个难题，有时间多练练

面试之前了解这个公司的体制和需要的技术

最好还是一边在职一边学习，多努力点，希望大家都能找到理想中的工作～

外行工程师转行路：传统 IT 转型 NLP 之如何抉择方向 项目 简历

一名传统行业搬砖快 3 年的工程师，你懂的，生活所迫，决心改头换面。在一位同公司的朋友推荐下，开始接触七月在线的课程，他通过七月的学习拿到了阿里的 offer。去年 6 月，开始在七月学习机器学习第 9 期的理论课程。从机器学习的各种理论到神经网络的初步了解，算是打了个基础。

后来又报了 10 月份的集 6 训练营（小编注：即[机器学习集训营第六期](#)），训练营给人的感觉，从 python 基础到数据处理，到可视化，到模型理论，训练，优化和项目，一整套的较为全面的课程。当然理论课进度是比较快的，课后还需要花時間进行自学和复习。

集训营 3 个多月全程下来，就开始准备面试了：

首先要定下自己要做的方向，CV，NLP，推荐，风控等等；其次就要动手做几个跟自己方向契合的项目，最好从头到尾做下来，面试的时候，面试官真的会从各个角度来考你聊不上来会减分；接着是简历准备，根据实际情况，写写自己熟练的知识和技能，写在简历上的技能一定要熟练，不然会显得虚假。

要转行的朋友（尤其像我这种来自传统行业的），要想办法写点原来工作中与机器学习或者 AI 有联系的项目，背景显得不会那么空白。方向的确定，项目和简历准备，可以咨询集训营的老师，老师们真的是非常热心的，不要犹豫，一定要多和大佬聊天，问问题。

下面重点聊聊面试过程吧。

简历最初挂出来的几天，几乎无人问津。随着在 boss 直聘上的活跃度上升，开始有各种邀请和猎头来联系。面了不下 20 家吧，感觉就是简直就是刷怪升级。经常有这种情况，在看书和看视频时感觉自己理解了和会的東西，被面试官问起来居然答不上来，或者发现原

来自己理解错了。因此学习的时候，一定要扎实一点，该推导的，就一步一步推导清楚，不要像我一样浮躁哈。

也许我的路子真的很笨，不过也确实有效果，经过几轮面试的查漏补缺，发现了自己在很多方面的不足，通过一边面试一边加强学习，知识得到不断巩固。

面试过程来看，基础的数据结构和算法是会考的，比如百度考了哈夫曼编码，二叉树非递归遍历等等。常见的机器学习理论和推导也是必考的，xgb 的原理，SVM，还有优化调参都是常见的问题。深度学习这块，卷积如何实现的，各种网络的结构和意义也是经常聊的。然后就是聊项目了，各种技术细节，是一定要清楚的。

一直找的是 nlp 方面的工作。词袋，tfidf、word2vec、LSTM、注意力机制、bert，是经常被问到的，需要根据自己情况做一些准备，被问到时最好往自己熟悉的方向展开讨论。

总的来说，机器学习知识是基础，但大部分公司都很在乎你在深度学习上的知识和经验，同学们在项目上一定要体现在深度学习方面的深度和广度来，切记。

最后战果跟论坛大牛们比起来，简直是献丑，最初拿到了 360 的 offer，但钱少离家远，拒了。后来陆续拿到 3 家 offer，有互金、教育等，大概 25K 吧.....，一入江湖岁月催，就着急着先转行上班了，曲线救国，来年再争取高薪。

NLP 水硕校招面经：SVM、注意力机制、bert 基本必问

先介绍一下个人背景，我是 NLP 一对一的，今年秋天毕业海外硕士，留学地区在留学鄙视链 3 或者 4 的位置，学校很一般。本科末流 985，专业都是 cs 方向的。读研期间参加过这边一个小的学术性质的搜索引擎的比赛，主要是教授要求，subtask 的 state-of-the-art，发过几篇论文。社招要求可能要求不一样我肯定没有发言权，所以只针对校招谈一点点。其实感觉自己手里的牌还行，就是自己一天心浮气躁的，又不太聪明的亚子，就说很菜。本文跪经居多。

从三月开始找工作到后面确定这段时间最大的感受就是：

- 1、没有正经的实习经历真的是硬伤（对于想找好工作的在校生来说）
- 2、熟悉知识，会用，能等到结果的程度根本不足以应付面试。
- 3、春招地狱难度，百度好像根本就没放岗位，内推都推不来面试，当然，也有可能是我简历挂了。

由于这边学制时间的问题，我只能先参加今年春招然后秋招。春招最后算是找到了去处。总之在校生请珍惜实习机会和秋招机会。

结合面试经历说些具体的：

达摩院，2 面跪

一面三道代码题，远程敲（感觉硕士以下不刷题能找到算法岗的概率为 0）。双指针，k-d tree，正则表达式。二面是总监，讲了些项目经历，transformer，seq2seq，svm，lda，learning to rank，基本简历里写的都会问到，棘手问题：项目里你的作用的是什么，有哪些是你自己想出来的？

香依，4 面跪

流程基本是每次两道代码题，加结合项目问一些问题。4 面的时候问了很多搜索的东西，因为自己主要准备常见 nlp 的内容，搜索相关的基本只是某开源搜索引擎的一些使用而已，被各种问住。例：为什么用了 svmrank，它属于哪种 ltr 方法，有什么好处和不足，怎么优化呢？

这个时候意识到了自己很多都只是会用现成的东西而已，实际上有很多不足，就报了七月的 NLP 一对一就业班。因为本身就是这个方向的，所以就和褚博士商量的以我这边主动推进为主，褚博士提供一些我需要的资料并及时对我进行指导，让零零散散的脑子里的东西，变成可以作为答案的知识点。

华为，面试非常愉快，但是进了 offer 池后等了一个多月跪

华为笔试很简单，3 道题，300 分 600 分 900 分，300 分的白给，600 分的需要写一写，但是像是体力活，900 分的有些难。

技术面试，就问问项目经历，svm 推导。综合面试聊天，也问了一些项目经历，但是就是聊天的程度。最后说对我的评价很好。我当时就以为稳了，后来赶上川川制裁华为，就一直拖进度。最后反正是被刷了。

目前的准备去的一家创业公司：

nlp 的算法岗，感觉 boss 很好，很尊敬。总共 3 面，一面 4 道代码题和一道思路题（就是谷歌爱考的从楼上扔鸡蛋什么的那种）。代码题基本都是 leetcode 原题，7 月这边也有，刷过了就还行。二面两道代码题，其中一个是在 linux 上开发的一些基本操作。然后问项目和基础，比如：lstm 的计算过程，举例说明，每一步说明各个参数的维度和变化。lr 和 svm 的原理，二分类时这两个选哪种，为什么。cnn 的算法复杂度。L2L1 正则化的原理。还有 adam 那些优化的计算过程。还有些 bert 和注意力的问题。三面就是谈待遇了，一个月两万多，据说不加班。

总结：春招时看，nlp 岗位真的没有 cv 岗位数量多。svm 基本必问，注意力必问，bert 基本必问。简历上项目写详细，防止了面试官瞎问，比如一些 boost，我根本没实际用过就试着跑过模型也没啥细节感受，虚得很，但是限于简历上的够问了，基本没被问到。“有问题，用模型，懂源码，会推导”不然就会被尴尬地问住（自己的血泪教训）最后祝大家都能找到心仪的工作。

双非渣本的转行之路：从机械设计到 NLP 并薪资翻倍

如果不出意外，我的条件应该是集训营最差的那个了吧，没有硕士光环，本科还是没听过的，三线小城市渣渣学校，英语水平属于那种四级飘过的那种，数学水平嘛，毕业两年都还给老师了，最主要的是我本科专业是机械设计及其自动化。现在回想起来，实在是找不到比我更渣的了，哈哈。那你可能会问，你这么渣怎么来学人工智能了？答案：其实我并不知这个很难，就一头扎进来了。还有一个原因是我高中数学还算不错，对自己有些盲目自信。这篇分享，我并不想写成屌丝逆袭的鸡汤文，事实上，距离逆袭还差很远，现在只是半只脚入门而已。主要是想分享一下我学习、找工作的历程，如果你还在迷茫，希望能给你点帮助

课程结束后，一个月的时间找到工作 NLP，目前工作的内容是对话机器人和知识图谱。

在报名集训营之前，我学习了 python，并且顺利从机械转行，做过几个月的爬虫工作，编程有一点的基础。但由于工作很忙，没时间学习，我又果断辞职，全心全意学习机器学习。脱产无疑是很痛苦的，没有经济来源，对后面学习完能不能找到工作的疑虑就很大，压力不言而喻。如果你正有辞职学习的打算，要慎重考虑；如果你还在学校的话，那就恭喜恭喜啦。

开始学习的时候，没什么方向，走了很多弯路，但也是跌跌撞撞把数学基础和部分机器学习的内容过了一遍。集训营课程对我帮助很大，这里有很多经验丰富的老师和志同道合的同学，学起来不会那么枯燥无味，少走了很多弯路。当然，最喜欢的还是集训营的线下课，每周跟线下的同学老师交流，会增加很多信心。说实话，如果不是七月的集训营，我现在不能找到工作，要打一个大大的问号。

集训营学习之前，我把赠送的机器学习预习课程看了两遍（不只是单纯的看，老师讲的每个知识点都记下来、理解，一般两个小时的视频，我要花一天的时间），深度学习预习课程看了一遍，有些算法，如果听不懂就听第二遍，还不懂就第三遍，我的方法就是死磕，毕竟咱水平只有这样，笨人总有笨人的方法，哈哈。

集训营的课程，内容很多，很紧凑，每课一定要跟着练习，完成作业和每周的考试，七月的老师和助教也会很负责的批改。集训营最大的一个看点就是项目部分了，NLP 项目的对话系统，我是反复看了很多遍，最后写在简历上的。

最后就是面试部分了，集训营的课程里有相关部分内容，老师会敲黑板，划重点，教你简历怎么写，项目怎么写，老师们经验丰富，照着做就对了，遇到不会一定要请教老师，特别的请老师帮你看看简历。具体到面试的时候，像我这样学历的，投出去的简历，一点回音都没有，那时也只能海投，起码可以增加面试机会，我想大多数人不会有我这种困境。面试的公司里都是中型、中小型和小型的创业公司。

小型的创业公司，基本不会遇到懂算法的面试官，大多会根据他们实际业务出发，考考你如何用算法解实际的业务的问题，比如我遇到一个电商公司，他们主要面向国外市场，希望做一个邮件自动回复的系统（国外好像都是用邮件）

中小型的公司，可能有些懂算法面试官了，这里就需要你对各算法的理论掌握熟练。中型的公司，会有比较厉害的面试官，各种算法的推导，还有实战经验都会考到（我实战经验不多，kaggle 都没打过，场面很尴尬）。

总体来说，NLP 的面试中关于机器学习的问题不太多（我没面试过大厂），大多是深度学习的问题。

面试手推过的算法：SVM，LSTM（很多次）TFIDF（手写代码实现），bert（面试必考）---我准备面试的时候每个重要算法都手推过不下十次

项目：面试介绍项目是一定有的环节，不懂的技术不要写在上面，不然问到答不出来很尴尬（用过的技术都不会，扣很多分）

以上是我的 NLP 转行之路，希望有能帮助到你的地方。

我的 NLP 面试经历及总结：要把 tfidf、word2vec、注意力机制、transformer 都熟悉掌握

我是机器学习集训营第七期的学员。

参加集训营的时候是一名打算找工作的国企换岗人员，之前在国企待得比较不自在，因为本身学的通信，后来在航空系统里面做机载设备，所以和本身专业也不是特别对口，加上体制的原因吧，就出来了。在机器学习方面之前都是自己看看书和视频，但是没有实际操作过。

我开始接触机器学习的时候就看过 July 的一些文章，讲得比较通俗易懂。以前听别的同学说过，可以报培训班来进行再充电和换岗，所以为了找个好工作就报名了。因为报名晚了点，所以学习的时间有四个月，从二月开始到五月底结束。因为我是社招，所以虽然过了春招，倒是不影响应聘的时间，一般来说，各个企业的社招都是全年在找人。

基本上五月份课程结束后，加上办理离职手续的事情，六月份调整了一个月，七月一号正式投简历开始找工作，拿到了几个 offer，因为我之前在国企，所以虽然对体制不太满意，不过还是对大公司以及国企比较有感觉吧，所以面的基本也都是大公司和国企，所以面的公司也不多，大概有五家吧，因为总共只有两周的面试时间，7 月的十九号我就已经入职了，虽然也是在国企，但是岗位是 NLP 的岗位，而且入职就开始干活了，现在其实挺汗颜的，因为还要经常补一补 CRF 和 viterbi 算法相关的知识以及 python 网络编程的一些知识。前两天一个简单的测试要内部上线，封装接口和 Docker 打包还花了一个周六加班研究了一整天。而且真心感觉培训班学的还是不够的，因为实际项目里发现 NLP 的机器学习一些算法是需要自己实现的，这个文章我是周天的下午写的，上午的时候还在看 Bi-LSTM+CRF 实现 NER 的 CRF 部分，感觉之前学的还是不扎实，不过现在我们部门气氛挺好的，因为也是国企新成立的做人工智能的部门，而且有时候发现没有遇见的问题在集训营的群里问老

师，一般也会很快有答复，所以虽然感觉不断的学习也挺有压力的，但是能做自己喜欢的事情还是挺开心的。

首先，集训营老师给的意见是自己确定好一两个方面，最好定两个简历模板，然后不断优化。譬如我当时就一心想去搞 nlp，所以就只做了一份简历。大家不要和我学，哈哈哈。首先不管我们是要走 cv、nlp、推荐什么方向的，面试的时候都是基于简历上我们自己写的。不过既然都选择了 NLP 或者 CV 等等，我们都必须熟悉且能手推出 LR、adaboost、xgboost、svm 这些经典的机器学习算法，这是最基础的了。下面我就介绍一下我面试的公司的一些情况

中兴:

主要谈了谈集训营的 NLP 项目，问的都是 NLP 基础问题，词向量、RNN 和 NER，比较简单吧，可能也是因为简历上写的就是这些。

某附属医院的人工智能：

首先说下结果，这个面的电话面就挂了。然后说下总结吧，当时对这个期望特别高，因为这个是大学附属医院的人工智能岗位，带头人是大学的一位人工智能方向的海归博士，所以特别期待能够拿到这个岗的 offer。但是也和我准备有关系，电话面的时间是晚上的八点半，我等到八点五十的时候见还没有电话以为要等第二天白天了，所以晚上九点钟电话打来的时候我正在看手机，感觉不是非常在状态吧，然后因为负责电话面就是那个海归博士，所以面的好细，先是问了奇异值分解，然后有问了 numpy 里面的矩阵相乘使用*和使用 dot 的区别以及 numpy 里面的矩阵初始化有哪些方法，接着问了如果做 NLP 的 translation 时如果训练集特别少怎么办，就是翻译的语句对只有 10000 条时有哪些方法可以提高神经网络模型翻译的准确率，最后还问了有没有自己设计过深度学习的模型。上面的几个问题有的

基础的我答出来了，后面两个当时真的没有思路，后来面试完了我看了一下才发现《百面机器学习》对迁移学习也是有介绍的，只怪自己没有好好看书呀。

国企一：

一面的时候问了精确率、召回率和 F1 值，然后简单聊了聊集训营的 NLP 项目。

二面的时候问了 RNN 网络的结构和初始化以及 CNN 网络的卷积体现在哪些方面。

最后就简单聊了聊待遇和加班情况之类的。

国企二：

上来就问我 HMM 模型的三类问题都是哪三类，LR 和 SVM 的区别，然后就是问了我 NLP 的项目相关内容，接着又问了如果想要提高分词的准确率有哪些方法，以及不了解知识图谱（当时是真不了解，可也只能硬着头皮聊了两句），最后问了些 python 的相关问题。

总结：

传统机器学习一定要掌握：svm、lr、决策树、随机森林、GBDT、xgboost 和朴素贝叶斯这些基础知识，最好能手推。

深度学习基本都是：CNN 以及卷积的意义、RNN 以及 RNN 的初始化、LSTM、常用激活函数（tanh、relu 等）这些原理。

自然语言处理方面：一定要把 tfidf、word2vec、注意力机制、transformer 都熟悉掌握。

最好自己去运行几次，最后就是祝大家找到称心的工作

我是如何通过 3 个月的学习，从运维开发转 NLP 的

大家好，我是机器学习集训营第六期的学员热苏斯，本科毕业于国字号双非大学，CS 专业，现坐标 cd，之前做过 linux 运维开发和 python 后端开发。转行 nlp 的主要原因是觉得这玩意挺高级，没套路，能折腾。

我记得集六是 2018 年 10 月份开的班，我同时还在学习其他平台的 ml 和 nlp 课程，所以感觉时间紧张紧张就到年底了。过完年我复习了 1 个月就开始投简历面试了。投简历只用了 boss 直聘和拉勾，可能也是赶上互联网寒冬吧，再加上本科双非毕业，虽然不少公司企业收了 my 简历，我也积极沟通过，但是最后只收到了四家公司的面试邀请，三家是创业公司一家是骗子公司，真的惨。面了一个周吧，都还答得可以，但最后只拿到了一个 offer，就赶紧入职了，先上车再说哈哈。

面试过程如下：

公司 A：

三道算法题，也都是 leetcode 上面的，一个是混淆字符串，一个是解一元一次方程组我用的二分法，还有一个记不清楚了。

面试就是正常聊项目、特征工程、模型。lr，svm，决策树这些机器学习的基础模型都有聊到。问的不深，基本上都是问下用过哪些模型，主要参数是啥，怎么调参调优，会用哪些框架。哦对还问到了爬虫，爬虫比较好回答，之前工作中也写过。

然后就是让我回家等消息了。

公司 B：

笔试题全是跟 pandas 相关的，复杂点的用 apply、groupby、agg 基本上就都解决了，感觉比较简单。

女面试官，先是从项目开始，项目聊了二十分钟吧，挨个介绍了项目的背景、流程、价值，用了哪些模型，怎么调优的，还有没有改进点等等。之后问了过拟合怎么解决，lr、svm、决策树等常用的机器学习模型的常用参数和含义，特征工程一般怎么做，聚类一般用哪些算法或者模型，knn 和 k-means 什么区别，讲一讲 lstm，有没有用过分布式训练，记得起来的的就是这些了。

然后就是让我回家等消息了。

公司 C:

全程都是 hr 跟我在聊，薪资开的是相当之高，最后说办公地点在菲律宾，至少要待半年才能回国，让我好好考虑下能不能接受出国，能接受的话马上就跟菲律宾研发中心联系进行远程技术面试，我去厕所抽了根烟，然后回家了。我怕是没这个命把钱带回来哦。同学们面试也要注意，这种公司前段时间被曝光过，这些公司一般都是在东南亚做博彩的，等你签完合同到东南亚办公地点之后，公司就会扣押你的护照和相关证件，跟传销组织一样，让你超负荷工作，基本的自由都被限制了，别说回家了。写了 bug 还要被保安拿棍子打，千万不要上当受骗了，天上不会掉馅饼的。

公司 D:

面试官跟我聊了大概十分钟我之前的工作内容，然后就开始讲项目：

- 1、讲项目的过程中问过特征工程怎么做的，用了哪些模型，为什么要用这些模型，有没有考虑其他模型，项目的提升点在哪里；
- 2、文本相似度计算常用算法有哪些，tfidf 和 textrank 的不同和适用场景，解释一下什么是语言模型，常用的分词工具有哪些，自己认为那款分词工具最好用，为什么，新词（未录入词）如何发现；

3、情感分析方面做过哪些工作，接触过舆情没有，讲一讲对 LDA 的理解，没有做过 NER

怎么做的，讲一讲你对隐马和条件随机场的理解，怎么理解语言模型；

4、讲一讲 word2vec，讲一讲 lstm 和 fasttext，讲一讲 attention，讲一讲 bert 和 transformer；

5、会不会 scala 和 spark，有没有用过 cdh，熟悉哪些大数据的组件。

能记起来的就这些了，聊了一个小时吧，第二天发了 offer，一周后我入职的，现在在做企业画像。

给大家的建议：

1、求职方向最晚在集训营开班一个月的时候就要定下来，搞 CV 还是搞 NLP，方向定了之后你会节省很多很多时间。我就是定的有点晚了，后来发现 nlp 相关的东西还有很多，但是明显感觉时间不够用了；

2、简历一定要专，之前与 NLP 或者 CV 无关的工作经验尽量少写，项目这块一定要写详细，面试应该是有 3/5 的时间是在聊项目，很多话题也是从项目上延伸出来的，不求多但求精；

3、实在不会讲项目怎么办，没事，就踏踏实实老老实实讲你做过哪些项目，之前干过啥，现在学过啥，当前的技能水平能干点啥，转行就说转行，不要虚，实话实说，除了工资少了点（可能是少了很多哈哈）其他也没啥损失，先上车再说。

NLP 常见面试问题总结：Word2Vec、LSTM、BERT 等

我是集训营的学员，我大致分享一下遇到的面试题

1. 基础的数据结构：插入排序、选择排序（记下时间复杂度），链表新增、删除，二叉树的遍历，其他场景算法题大多出自 leetcode
2. 逻辑回归损失是什么，手动推导一遍
3. 对集成学习，SVM 的理解，以公式的形式写出来最好
4. 对 HMM，CRF 的理解，CRF 的损失函数什么，维特比算法的过程
5. 手写一个 tfidf
6. word2vec 的 CBOW 与 SkipGram 模型及两种训练方式(负采样\层级 softmax)，两种训练方式的区别和应用场景，
7. word2vec 和 fasttext 的区别，训练 word2vec 的有哪些重要参数
8. LSTM 的单元结构图和 6 个公式要记住
9. 有几种 Attention, Attention 和 self-Attention 是具体怎么实现的，对应什么场景
10. BERT 的模型架构，多少层，什么任务适合 bert，什么任务不适合，应用在你写的项目改怎么做
11. tensorflow 手写一个卷积代码，BiLSTM + CRF 模型的原理，记住常用基础 api(比如 jieba 添加默认词典 api，分词 api)
12. 问项目阶段，会问数据集怎么得到、模型的训练、怎么部署、项目人员周期，开发中出现问题怎么解决等

985 小本 3 年后台转算法

我毕业于某 985 高校计算机专业，小本。毕业后做了一年多销售，然后转行做了两三年的后台研发。去年转算法（NLP），最终确定的 offer 年薪 40 万。

面试题

直奔主题 —— 面试题 ~ ~ （以下公司全部使用 “*” 补齐 4 个字，无论是否真的是 4 个字）

- 京 ** 团 （推荐算法）

1. 项目相关，为什么使用这个模型？
2. 数据清洗流程，概述以下
3. 为什么集成这些模型？什么模型不适合这种方法集成？（《百面》中有相关论述）
4. 常见推荐算法

- 美 ** 评 （NLP）

1. 一道动态规划的题目，难度大概就是爬楼梯那种级别的
2. 推导下 SVM
3. LSTM 梯度消失的原因
4. 数学智力题
5. 快速排序的原理，怎么消除快排的缺点（极客时间的算法课中有相关论述）

- 金 ** 团 （NLP）

1. LSTM 的图，GRU 的图
2. 梯度消失及其缓解策略
3. BERT 和 GPT 的区别
4. 手推 SVM、LR

5. 项目细节，比如清洗流程、Embedding 的维度之类

- 望 ** 慧 (机器学习)

1. 为什么使用这么复杂的模型？
2. 对于样本中缺失的特征怎么处理？（医疗领域很关注）
3. 写一个函数，均匀采样随机数生成正态分布随机数
4. 手写快

- 五道口某金融公司 (NLP)

1. 忘记这个公司的名字了，CTO 是以前网易有道的大神
2. 推导 SVM
3. 推导 LR
4. 会 Spark 、 Hive 之类吗
5. 链表反转

- 百 * (NLP)**

1. 推导 LR
2. Doc2Vec 两种方法孰优孰略？
3. 样本量多少？为什么使用深度神经网络？
4. 读取文件，计算 TF-IDF
5. ID3 C4.5 CART 的分裂标准，写出公式

- 总结：

1. 面试基本组成就是：算法题、机器学习/深度学习/NLP 之类子领域基础知识、项目、其他基础知识

2. 算法：CS 出身的同学们基本都会被考算法题，非 CS 的据说有的不考。算法题基本上刷一刷 LeetCode 的常见题目就可以了，无非就是字符串操作、广/深度优先搜索、链表、动态规划、树。面试算法其实不难，我刷了大概 100 多道的时候面试算法题就都没问题了。刷算法题我建议找一个 LeetCode 题目的分类帖子，然后从容易的开始刷，换着类型刷。上来直接写答案提交，如果觉得实在太难，就搜答案或者看讨论区的答案，理解后自己实现。（这种方法虽然不是正经的学习方法，但是现在是面向 offer 编程）一般切两三次就烂熟于心了。另外，对于一些很难的题目，我不建议死磕，因为面试中遇不到的。动态规划的难度也不会出到很高的那种难度，如果有，那就告辞 ~ ~ ~

3. ML/DL：机器学习反正就是看看 百面，DL 方面至少要对 CNN/RNN/LSTM/GRU 这种了解透。当然，也分情况，比如 NLP 是 LSTM 向的，CNN 没人问过我。

4. 项目：项目最好自己去做一下，至少要跑通。数据规模和来源、预处理流程、评价标准以及结果、哪些模型影响比较大、模型设计细节之类，这些都要记住。

5. paper 和 数学：如果时间不多（大部分人都不多），我认为可以先不管（这东西很耗时间），但是就职后最好补上。

6. 其他基础知识：如果你和我一样是 CS 科班出身，有可能会被问 Linux 命令、网络工程之类的东西。这个就看积累了，但是至少基本的 Linux 命令总要会。

7. 大数据相关：小公司一般会很关注 Spark、Hive 之类，

碎碎念

1. 不要裸辞。
2. 希望你是真的喜欢算法这个行业，有意愿持续学习。（想想每天都要看看论文——当然，你也可以不看，就是可能会在未来被淘汰）

渣硕产品转 NLP 实录：从只会 matlab 到 NLP 年薪 30 万

首先做个自我介绍：本人是 2018 年 3 月毕业的渣硕，由于从一开始就想做产品所以从本科到研究生我就没想过学习编程和相关的算法，所以毕业之前只会用 matlab，更别说机器学习或者深度学习什么的了。

毕业后在 Megvii 做产品经理，工作中发现想要做好产品，懂技术是必不可少的，所以在去年 10 月份的时候有了转行的打算，通过同学的介绍就到了七月的集六训练营。

当时一开始学的时候，比较吃力，因为在我潜意识里就不喜欢编程。但是随着学习的深入感觉还是挺有意思的，所以我就在 11 月份脱产学习。说实话在 Megvii 内部学习的机会也很多，但是工作的强度太大，对于学习一个新的领域来说我还是选择脱产。

在 8 月出开始找工作，一开始准备投一些小公司刷经验，但是幸运的是面试了两家，也得到了这两家的 offer。考虑到本人没有实际经验，也不是应届生，所以就选择其中一家工作了，现在我就聊一些干货吧。

首先，我要说七月的课程编排的已经很合理只要跟着学习，然后课后练习及课程相关的项目一定要自己熟练的做几遍，有些函数的参数一定要理解里面的意义。但是最重要的还是要自己完成练习及项目 ps 复制粘贴模板有时确实可以解决当前的问题，但是我个人认为只有在过程中出现问题，我才能记忆深刻否则同样的问题会发生很多。

我记得当时照着答案输入，有时候都会出现问题，更何况以后独立的编写呢。ps 这段时间我主要做练习是发现我的不足而不是能够完成具体的那个功能。

其次，一定要选择自己做的方向，不要盲目什么都去学，因为对我而言多学一部分知识就会花费一部分时间，多花费一部分时间我就会投入更大成本，毕竟是脱产。在选定好自己的方向后，一方面要学习相关领域的知识，一方面要提升在编程方面的能力。

在提升相关领域的知识方面我会在后面提供一些我收集的资料，在提升编程能力方面很多人建议刷 LeetCode，对于那些有开发经验的人来说我也建议，但是对于像我这样的小白我建议先刷剑指 offer 然后在刷 LeetCode。

我的理由是：

- 1、leetcode 题量相对较大
- 2、解题思路偏用 (c++或java) 解题，这个对于小白可以说是噩梦（数据结构这个真的对于我这样的小白很难，循序渐进的学习还算可以，硬钢我是做不到）
- 3、寻找解题的解析太浪费时间

此外，在七月的面试题库中有很多面试题，个人感觉已经很全面了，多刷几遍后在看看相关领域其他人的面经，在面试理论部分已经相当够用了。自己也要找一些课外的资料来补充，就拿我找的资料为例：

- 1、[https://www.cnblogs.com/jiangxhtml](https://www.cnblogs.com/jiangx...html) 这个是文本分类的实战基本涵盖了主流模型。
- 2、腾讯视频搜索 ‘60 分钟带你掌握最强 NLP 模型 BERT’ 这个对于理解 transformer, ELMo,GPT,Bert 很有帮助
- 3、[//github.com/Dikea/Dialog-System-with-Task-Retrieval-and-Seq2seq](https://github.com/Dikea/Dialog-System-with-Task-Retrieval-and-Seq2seq) 智能对话方面的项目，工程文件很好看懂，框架架构也比较适合初学者
- 4、知乎的平安寿险 AI 团队详细的介绍了智能交互的框架，以及每个部分需要的技术。
- 5、[https://blog.csdn.net/weixin_3 ... 35607](https://blog.csdn.net/weixin_3...35607) 关于 XLnet 的解读

最后，我讲一下我的面试经历因为是小公司具体叫什么我就不说了 说一下公司的业务方向：

第一家的业务方向是安防，

第一面主要问了简历项目，Istm 的原理，Istm 为何能长时记忆，以及关于文本的随机采样。

二面有两位面试官同时面主要问的有：编程题：从上往下打印二叉树、最小的 k 个数、两个列表的公共点；深度学习基础知识：w2v 原理，transformer 原理，对 AE,AR 语言模型的理解。

最后问了一个业务场景在如何设计智能对话系统。

第二家的业务方向是金融咨询，

第一面是主要问了 svm lr 等原理，简历上的项目。

第二面写了一道算法题 ‘二叉树中和为某一值的路径’，然后问我做文本分类一开始你一般会统计那些特征来确定模型的选择。

最后问我是否会用 spark，我说不会，然后就给我介绍了他们所做的项目。

我如何从传统 IT 转型 NLP 并薪资翻倍的：NLP(知识图谱)面试总结

我是机器学习集训营第五期的学员，先学的机器学习后来转了 NLP,主要学的方向是知识图谱，学习的时间大概是四个月，知识点比较多，学起来也很费神，主要难攻克的点有：(1)HMM;(2)CRF;(3)一些主流模型：bert, transformer,等。这些知识点一遍很难看懂，多看几遍，慢慢的有全局概念后就慢慢明白了。

学习 NLP 的主要难点，我觉得是：(1)知识点难懂，知识点多(2)模型复杂，实际应用比较占资源。但是为了有个好的基础，大家还是要扎实。因为前期自己掌握的资源有限，面试了一家比较厉害的企业将相关知识点补充后，后面再去面试基本没什么大的问题。下面讲述一下自己的面试过程，希望能够对各位同学有所帮助。

深度智药

这家公司实力比较强悍，很多是 bat 出来的，问的问题很细。简历上的知识点问的很细，word2vec 有几种模式(CBOW, Skip-gram),对应的流程图，提及 word2vec 极有可能顺带提及 fastText。提及词向量也会顺带牵出其他词向量的方法，词向量发展的各阶段模型，以及各个模型的优缺点。最后是词向量的工程实现，比如 word2vec 是哪个包下面的模型，自己怎么训练的，有哪些参数可以设置，参数怎么设置的，为什么这么设置。

seq2seq 的原理图，attention 是怎么回事，为什么会有效果，这里容易引进 transformer，然后这个话题就有点复杂了，看水平的时候，transformer 模型有点复杂，自己先看懂，然后画模型流程图，找出模型各个部分的区别，对比分析，说说自己的理解。最后一般会顺带问，在我的项目或者实际中有没有应用，这个自己在学的过程中最好自己应用一下，会加分不少。

基础技能：这部分会将知识点过一遍，写在简历上的知识点，要能够有很好的理解，有原理图的画的很熟，面试官有时候会牵出一些相关的知识点，这个时候自己看过就简单描述一下，没看过，回去就补一补。

项目：自己项目的用途，项目背景，项目相关信息(知识图谱为例：这是为了干什么，用户是哪些群体，解决了什么问题)，数据从哪里来的，数据量有多大，数据怎么处理的，模型的选择，模型的训练（训练时长，训练的设备（几张 1080ti 或者其他设备）），模型后期怎么维护的。

这次面试官比较厉害，也给我了一个总结！工资按要求写，不要怕写高，只要水平达到，工资不是问题，知识储备要加强。回宿舍后自己总结了一下，理论不够扎实，实际操作太少，很多基础操作都没有搞清楚，这个就掉价了。

某研修学院

面试第一家，备受打击，去面试时提心吊胆。学校找了计算机学院的主任和院长来面试，下的够哆嗦，后面发现问题都是很肤浅的。机器学习的本质是什么，我觉得就是：统计。从大量数据中统计分析数据本质规律！他们看简历也不懂，然后提了一个实际问题，怎么从文本判断一个老师有没有去上课！我的回答是：这是一个二分类问题，然后先切词，然后词向量，最后根据数据来训练一个模型做判断！本质就是，统计哪些词与上课高度相关，哪些词与没上课高度无关。然后校长一听就明白了！我觉得这个地方就是将复杂问题，简单形象的表达出来即可。

后面就开始和我谈薪水之类的！最后给了 offer，但个人觉得对技术积累没有帮助，放弃了！

渊亭科技有限公司

面试了几家,有些底气了!这一家刚开始就是笔试,笔试题目也就是我们常看的知识点,有一些有些超纲,但是个人觉得,自己什么水平就要什么价格。笔试题一般有一些难度,不要因为这个放低自己的要价。

面试也和深度智药差不多,问问基础知识点,问问项目,但是一般水平的公司不会问的特别细,这也是评估一家公司水平的考核方法。由于前期的弥补,知识点都搭上来了,项目也讲清楚了,后面也很快就进入谈薪资,发 offer 的阶段了。

后面几家公司面试流程也差不多,我就不再多说面试的常见的问题了,讲讲自己每个公司面试的一些体验!

苏州闻道

这家公司首先给我电话,第一个问题就是:机器学习的本质是什么?这个问题前期思考过,我觉得就是统计。然后下午就让我去面试,然后找了几个面试过来面试我,结果面试官都不懂,基本被我牵着鼻子走,我也讲的很顺,估计他们是因为我的熟练程度,感觉就很满意。面试结束,人事很快就给我电话了,谈期权,谈薪资,也就很顺的拿到了 offer

上报界面

刚接到上报的面试通知的时候对上报界面没什么概念,后来同学们说这家公司很牛!我也是被自己的无知给吓到了,开始踏踏实实复习知识点,复习各个知识点的原理,流程图等。早早的去到公司,果然是不到时间公司不开门,气势很足。默默的在外面等了一个小时。后来就是走了很多流程,有事漫长的等待,都有点着急了,来了一个小伙子,叫我面试。知识点这块面试官也不是很懂,基础知识,项目都讲了一下,问了几个简单的问题(LSTM 原理,GRU),然后进入二面,二面也就聊了聊家常(在哪里工作,干了什么,预期薪资),然后就是等人事通知。果然周五收到了人事通知,询问到岗时间,薪资,入职要求等!

苏宁

面试的差不多了,也想看看大公司的水平,就投了各个大公司。本以为大公司不会理我,自己也没有把握,只是觉得有那么多 offer 了,然后全看看大公司的水平,给后期留下一个奋斗的目标。结果百度给我电话,遗憾的是方向不一样。

后来苏宁给我电话了,积极准备,忐忑的心情去当炮灰。进去就是笔试,题目不多但很有代表性,考察基础。写完了,自己也感觉不是很好,面试官过来问了下项目,聊了聊知识点,但是面试官问了一个很怪的问题,就是:能不能加班!我觉得适当加班也可以,就很爽快的说,加班没问题!然后聊了一些苏宁也是大平台,在这边工作算是一个跳板,可以进大公司。最后就是等通知。两周过后通知我去办手续,准备去上班!很惊讶,自己居然过了!后来想了想,苏宁加班厉害,愿意加班,就有很大的优势!当然苏宁的工作强大也比较大!

面试的流程大致就是这样,自己的总结:在学知识点的时候面试信息适当看看就好了,很多负面信息会影响自己学习,理论学的差不多了,自己一定要多动手,自己调模型,多总结一些自己的理解,模型的参数要记一记。面试也是有个过程的,大概有个三四家公司练手,后面的就熟了,自己也可以判断出公司的水平,找一个自己满意的公司,然后就顺利入职了!哈哈,也祝各位同学好好学习,有个理想的工作!

跨专业转行 AI 的艰辛历程与 NLP 面经

回想着自己参加七月的机器学习集训营已经是一年前了，我记得机器学习集训营 5 期就是 8 月 1 号开课的，那时的我对机器学习停留的认知依旧只是西瓜书上的那一系列的理论知识，虽然我对所有的理论知识都有所了解，但是却很难将理论知识与实际项目结合在一起。没有老师的指导，一个人不停的在河边摸着石头过河是一件非常折磨人的事情，没有明确的目的与方向会让自己觉得一点都不踏实。

虽然在参加集训营五期之前我已经找到了 NLP 的工作，但是正是因为自己是从别的专业转行过来，项目经验的缺乏就像悬在头上的达摩克利斯之剑，害怕稍有不慎，就会掉下来。不过，也正是由于这份担心，我才选择了机器学习集训营，想夯实一下机器学习的理论知识的同时，逐项积累一些比较好的项目经验。

在报名之后，我才发现以前自己闭门造车的学习，学习进度滞后的有点过分，不停的追求数学公式的推导，使得自身对于项目的实践与理解产生了严重的偏差。反观机器学习集训营，其学习课程由浅及深，梯度式学习让自身的压力也降低了很多。机器学习集训营的内容真的很丰富，虽然线上讲解的时间只有 2 小时，但是里面所包含的信息确实很巨大，这些知识要将其吸收并转化为自己的知识，可不是一朝一夕可以完成的。但一旦消化完成，基本上也就掌握了这些知识，当然好记性不如烂笔头，我在集训营的学习过程中使用 OneNote 做了接近五万字的学习笔记，便于后期能在需要的时候将知识捡起来。

说实话，集训营的课程安排还是比较合理的，每周几节线上课程，用于讲解理论知识的同时还会具体剖析一些代码案例，这一点是在任何书本上无法学习到的，讲师还会为那些基础较差的学员进行讲解（比如我），直到听懂为止。讲师一般会留下一些课后作业，让学员能在课后巩固学习到的知识并为之付出实践，周末还会有线下课的讲师会和我们讲解一些比较经典的机器学习方面的项目案例。

在完成了机器学习集训营的课程后，自己在实际工作中面对一些复杂的需求也确实能得心应手的去面对，尤其是在一些数据挖掘的工作任务中，我成功应用了在课程上学习到 seaborn, pyecharts 和机器学习的知识做出了一系列精美的数据分析报告，得到了上级领导的表扬，直到这一刻我也才明白机器学习集训营的课程还是真的很接地气的。虽然我并没有因为有机机器学习集训营的学习经验而去换工作，但是我却因为它而成长了许多，也能在实际工作中独当一面了。

这里我也分享我闭门造车的过程的一些面试经历，面试的公司中小型的互联网公司，因为地标在广州，寻找工作的时间又是在招聘淡季，薪资不是很高。这里我将几次重要的面试情况进行了总结，大家有兴趣可以参考一二。

Python

1、浅复制与深复制的区别？

SQL

- 1、数据库查询有哪些常用的优化方法？
- 2、主要是通过什么方法来判断某些 SQL 语句需要进行优化的？
- 3、除了使用索引以外，还有什么方法可以加快查询速度？

ML

- 1、你在构建线性模型的时候，有没有遇到多重共线性的问题？是怎样解决的？
- 2、说一说逻辑回归和 SVM 的区别和联系？
- 3、简单介绍一下逻辑回归？
- 4、决策树模型中的预剪枝和后剪枝是什么意思？
- 5、你所知道的决策树做最优属性划分使用的指标有哪一些？
- 6、请你简单介绍一下 SVM？

- 7、SVM 中有用到对偶方法，请问对偶方法是什么，在满足什么条件下，对偶问题的解才能等于原问题的解？
- 8、工程上，经常遇到一些约束问题，一般要通过什么方法来进行求解？（拉格朗日乘子法）
- 9、SMO 算法在求解 SVM 的超平面方程中，扮演了一个什么样的角色？
- 10、请说一说朴素贝叶斯模型？它的假设条件是什么？
- 11、你了解的聚类算法有哪些？（我只答了 k-means, DBSCAN）
- 12、请说一说 K-means 的思想？
- 13、你知道那些降维算法？知道 t-sne 吗？
- 14、请说一说 PCA？
- 15、知道那些矩阵分解的方法？
- 16、请说一下你是通过什么方法来判断模型欠拟合和过拟合的？
- 17、你主要绘制的可视化的图表有那些？（面的一家数据公司）
- 18、请说一下 GBT 和 XGboost 的优缺点和区别？
- 19、简单说一下 XGboost？
- 20、请简单介绍一下随机森林？随机森林里面的随机代表了什么？

NLP

- 1、你知道哪些文本分类的方法？（我从机器学习和深度学习两个方面回答的）
- 2、在你从事的专业领域，遇到未登录词怎么办？你是使用什么方法进行分词的？（我说从搜狗输入法的词库下载的，然后结合 jieba 的自定义词库，完成的分词）
- 3、你是怎么去提取文本的关键词的？
- 4、如何去判定两个文本的相似性的？相似性的度量标准有哪些？你可以说一说余弦相似度和皮尔逊系数间的差异 吗？

- 5、可以简单介绍一些 word2vec 吗？
- 6、请简述 CBOW 和 FastText 之间的区别？
- 6、TextCNN 的结构可以说一下吗？卷积核的层数应该怎么取？
- 7、说一下 SVD 的时间复杂度？
- 8、PLSA 的核心思想是什么？
- 9、可以简单介绍一下 LDA 吗？求解主题使用的优化方法是什么（变分推断法和 GIBBS 采样）
- 10、如果给一段用户的对话，如何提取对话里的关键词将商城里面的商品推荐给他，这些关键词应该是什么？（我说要收集大量用户的历史行为记录，构建一个模型，当用户日常对话留言中有这些特征时，就可以把商品推荐给他）
- 11、用户与智能设备之间的语音交互中，如果将用户的对话对应到设备的一系列操作命令上，如果出现了一对多 或者多对一的关系，该怎么处理？
- 12、请问 A/B 测试是什么？
- 13、你知道注意力机制吗？

一个机器学习集 4 学员的工作感悟

大家好，我是集 4 的一名学员，看到七月在线现在集训营已经办了这么多期了，我也算是一名老学员了哈。

因为我是公司内部转岗，所以相对面试简单容易很多，所以谈不上什么面试经验吧，这里简单分享一下我参加七月在线整个的学习经历和工作后一点感想。

我是同学在网上看到七月在线的网站，然后推荐给我的。我之前是做安卓底层相关开发的，当时也纠结了好久要不要转，看到七月的课程后决定试一试。我在学员当中算大龄了，因为当时已经工作了 6 年多，报名的不算早，因此当时赠送的课程并没有看完，这也导致后期上课的时候有些吃力，当时有的同学已经把赠送的课程看了两遍了，所以上课时候很顺利轻松。所以新学员有时间的话，尤其刚报名的，提前把赠送的课程好好看下，这对以后正式上课很有帮助的。

还有就是七月在线学习的中后期，当时老师会每个机器学习方向（推荐，NLP，CV）都提供针对性的项目和练习，我当时想做 ASR，因为之前自己也看过一些相关的东西，甚至买了一下资料看。但是后来有次跟老师沟通我的情况后，老师觉得我还是做 NLP 更好一些，我想了想后也决定换做 NLP 方向了，现在从事的也是 NLP 相关的。

所以这里想说的一点感悟就是尽早确定自己以后想从事的方向吧。自己想好了，也最好跟七月的老师沟通一下，根据自己的情况确定自己的方向，做到有的放矢。集训营的课程上完后，当时公司正好有内部转岗的机会，就去试了。

因为是内部转岗，所以面试相对简单，也算顺利。好多学员已经分享了成功的面试经验，我就不提了哈。

这里简单说下工作之后的一点感悟吧。我工作中还是偏工程的，平时模型调参工作不是很多，主要是偏重技术落地，工程实现上。语言主要是 c/c++，Python。我觉得工作后，数据结构和算法功底还是很重要的，我也是随着工作在补这方面的东西。

另外一个分析源码，这是我觉得另外一个重要的东西，有时候学习算法，光看书本，可能印象不深刻，但是找来相关开源实现，读下源码，会理解的深刻很多，对编程的一些技巧也能学到很多。

这就是自己工作后的一点感触，一起加油！

跨专业转 NLP 的扎心历程：试用期被辞再找拿到更高 offer

转眼已经到了 9 月，又是一年金九银十的招聘季节。

大家好，我本科毕业，之前在一家国企工作。。。这一套自我介绍的话语，感觉自己说了 30 遍都不止。我是七月集五的一名学员，参加七月的机器学习集训营已经是一年前了，当时我还是在一个五线小城市做着国企中的一个螺丝钉，一个双非本科毕业四年每天过的浑浑噩噩的人，每天过着一眼望得到头的日子。

偶然的空闲我看到了七月在线的机器学习入门的课程，虽然自己只在大学学过一点 vb 编程，但是自己似乎可以趁三十岁之前在次拼一把，重新选择自己的生活。于是在考虑了一周后下定决心，报名参加了七月的第五期集训营，然后在跟着线上学了一个多月以后，由于自己所在的小城市没有线下课，裸辞，转战线下，继续每天听课，线下和老师交流，学习。

当时感觉自己又回到了高中的日子，每天很充实，一直在学习，一直到十月底课程结束。当时学完课程，发现机器学习集训营的内容很多自己都没完全吸收，虽然线上课讲解的时间只有 2 小时，但是里面所包含的信息确实很巨大，这些知识要将其吸收并转化为自己的知识，可不是一朝一夕可以完成的。于是给自己定了个目标在金三银四之前在准备一段时间，于是我将在集训营的学习过程中使用印象笔记做的学习笔记，重新进行梳理，结合代码，一篇一篇的看过去，建立自己的学习体系。

三月，我开始正式投简历，第一波投递仿佛石沉大海一般，没有一点消息，于是找到了寒老师和线下老师，在老师的不厌其烦的帮助下改了接近三版的简历，终于陆陆续续收到了面试邀请，在面了第三家的时候就收到了 offer，虽然薪资待遇和动辄 30 万+的计算机大佬们不能比，但是也算是能踏入算法的路，而且也比自己当初在国企工资要高很多，于是顺利入职。

写到这可能有人会觉得就完了，那就不能算是扎心历程了，因为扎心的才刚刚开始。当时快到自己三个月试用期转正的日子，领导找到我告诉我不能给我转正，然后说了一些客观原因，自己当时也感觉自己还有学要学习的，于是又一次离职，重新开始。

八月中，在自己又自己沉淀了两个多月后，再次投简历、面试，这次没有之前那么顺利，但抱着一股不信邪的倔强，在面试了 20 多家公司后，终于收获了一个 nlp 算法 offer，而且薪资比之前的还要高。虽然不知道自己后面的路是怎样，我只知道相信自己，尽力去做就好。

以上就是自己复盘过去一年的心路历程，虽然历尽挫折，但是一直没有放弃，最终还算收获了不错的结果。最后把这句话送给那些在犹豫的同学，面对未来，我们每个人都是既害怕又向往，如果你犹豫那么你永远不会迈出那一步，只有沉下心去做才会有所收获，just do it!!!

同时在这里感谢七月的各位老师，尤其是线下的老师，对我给予了很多帮助，最后祝七月越办越好~

附一些常问 NLP 面试题：

1. xgboost 是用同一个模型吗？
2. 分词模型构建，字典，实际应用？
3. jieba 分词步骤？
4. 一些基础模型特点？
5. softmax 和 sigmoid 区别？
6. svm 不同核方法区别？
7. 不均衡采样如何进行预测？
8. dropout 和 BN 的区别？

9. 简单介绍下聚类?
10. 无监督学习评价指标?
- 11.attention 注意机制? 还有简单介绍下 self attention?
- 12.如何使用 bert?
- 13.Bert 训练的一些参数?
- 14.Gan 的原理?
15. LSTM 与 GRU 区别?
16. rnn 和 cnn 区别
17. embedding 物理意义?
- 18.word2vec 伪代码推倒?
- 19.fine-tune 的原理?
- 20.pooling 层作用?
- 21.lr 为什么损失函数为什么用极大似然
- 22.GBDT 的原理 (理论基础)
23. 决策树节点分裂时如何选择特征, 写出 Gini index 和 Information Gain 的公式并举例说明 (理论基础)
24. 分类树和回归树的区别是什么? (理论基础)
25. 与 Random Forest 作比较, 并以此介绍什么是模型的 Bias 和 Variance (理论基础)
26. XGBoost 的参数调优有哪些经验 (工程能力)
- 27 XGBoost 的正则化是如何实现的 (工程能力)
28. XGBoost 的并行化部分是如何实现的 (工程能力)
29. 如果选用一种其他的模型替代 XGBoost, 你会选用什么? (业务理解和知识面)

传统 IT 企业转岗 NLP

我是机器学习集训营第五期的学员，计算机专业本科 211，硕士 985。硕士阶段主要涉及分布式计算、存储等领域，所以本身对机器学习的了解并不多。毕业后一直在一家美资传统 IT 企业，一晃五年多过去了，随着时间的推移，渐渐感受到对有限的上升空间的焦虑，以及经济方面的压力。在纠结、焦虑的过程中，也曾尝试着积极行动起来，努力地为自己寻找合适的方向，自学并最终实现转岗。但实际操作起来，就像无头的苍蝇一样没有头绪，或是选不准主攻方向，或是学习过程中把握不住重点，总之收效甚微。

一个偶然的的机会，朋友跟我聊到 AI 的广阔前景，加上自己还不错的数学功底，所以决定转岗机器学习。考虑到之前自学过程中遇到的一些问题，思忖再三之后决定参加七月的机器学习集训营。报名之后，距离开课还有一个半月，七月的工作人员先提供了一些基础课程以便我们快速掌握机器学习相关的基础知识，为后续的集训营学习打下扎实的基础。当然，除了七月的预习课程外，还看了吴恩达的斯坦福公开课，对经典的机器学习算法有了大概的了解。

正式开课后，就是持续三四个月的高强度学习。由于我是在职参加培训，另外家里小朋友也会分散部分精力，所以整个培训以及后续自我完善的阶段确实非常辛苦，几乎每天都熬到一点以后才休息，放弃了所有周末和节假日的休息。好在，七月合理编排的课程和老师们深入浅出的解析帮我们快速理解和消化，并且提炼出需要重点关注的内容，达到事半功倍的效果。平时自己遇到的疑惑，老师们也会在群里及时耐心地解答。

培训结束后又复习了一段时间，一眨眼又到了来年的四月份了。金三银四的求职季，怎能错过。赶紧请老师再帮忙修改了下简历，便开始了求职之路。从开始投简历开始，整个过程大概持续了一个半月，断断续续地面试了几家公司，拿到了几家公司的 NLP 岗位的 Offer，基本上涵盖了创业公司、互联网公司、券商这几种类型。下面是我几次重要面试的情况总结

如下,整体给我的感觉是基本都只关心项目经历,会根据项目描述穿插着问一些远离的细节。

大家有兴趣可以参考一二:

A. *火通信

自我介绍

介绍一下简历上的项目,问了一些项目方面的细节。

项目中用到的 fasttext 是自己实现的还是调包

项目中数据的来源以及正负样本分布情况,负样本如何采样的。

BERT 中为什么要做 positional embedding

常见的 attention 机制有哪几种类别

句法分析

读过哪些论文

QA 项目中具体如何检索出问题的答案的

介绍了他们目前正在做的一项任务,询问我的思路。当时没太理解他的问题描述,后来想想

应该是类似于事件抽取的问题。

会不会用 tensorflow,给了一道题,让用 tensorflow 实现。

B.某云商

自我介绍

逐个介绍简历上的项目

情感分析中如何处理否定词

项目中模型用的是什么损失函数

挑一个最了解的算法进行讲解

QA 项目中具体如何进行问题的匹配和检索的

这家公司的面试比较简单，面试官也比较耐斯。

C.某背单词 app

自我介绍

面试官似乎比较关心之前工作的具体内容,包括未涉及机器学习的部分工作内容也仔细询问了一些细节

项目的数据量多大，是否存在正负样本不均衡的问题，如何处理。

项目中使用的模型的具体结构以及为什么要用这种结构

项目中为什么要使用这么复杂的模型，训练时间需要多久，是否尝试过更简单的模型

项目持续的时间、是否已经上线，是否通过手动调参

项目由几个人完成、具体分工，自己完成哪些部分

BERT 训练过程中用到涉及到的哪两个任务，这两个任务分别有什么作用

在目前的公司使用什么版本控制工具

是否熟悉 docker，模型如何上线的

最后给了两道算法题，让描述思路。

- (1) 复杂度为 $O(n)$ 的算法，计算一个未排序数组中排序后相邻元素的最大差值。
- (2) 用 10 只小白鼠找出 1000 瓶液体中有毒的一瓶

NLP 开发的体会

作为一个 python 开发老司机，做 NLP 开发也有很长时间了，有一些心得体会，想站在开发者的角度，说说做算法相关的工程开发需要的技能，这也是我面试中着重考察面试者的技能点，希望能抛砖引玉。

做 NLP 开发，或者更进一步说，做开发，我觉得比较重要的是对原理的把握。看到很多论坛里和群上说，他们做的是代码的搬运工，直接百度抄的。对这种代码，其实要保持一份警惕，因为我自己也有自己的网站，有时候思考或总结了一些东西发上去，过一阵子常有发现发上去的东西有错误或不严谨的地方，但经常懒得去改，那反过来想，怎么能保证别人写的东西没问题呢？无论是什么代码，本着对项目对自己负责的精神一定要过一遍，理解原理。当真正去理解之后，有时候你会发现，这些抄过来的代码，逻辑可能还不够严谨，代码可能还不够规范，还不如自己实现一遍。当真正做到理解原理这一步，什么新的包就可以很容易上手了。而且，你长久跟一个项目就会发现，项目的需求是一直在增加，一直在变化的，掌握那些包的简单用法远远不够，还要深入去阅读源码。比如 jieba 分词，简单的用法仅仅是用 jieba 加载自定义词典，用 jieba.cut 分词，但后来随着项目的发展，我们需要对不同的用户建一套自己的分词词典，这样原来简单的 jieba.cut 就远远不够了，阅读源码发现 jieba.cut 其实调的是内建的 Tokenizer 对象的 cut 方法，这样要为每个用户建自己的分词器，只要建自己的 Tokenizer 对象就可以了。

第二个，比较重要的是抽象思维吧。要深入去理解面向对象编程、装饰器、元类等，去思考代码架构，这些对你写代码很有帮助。我接手的代码原先是算法工程师写的，他们之前既做算法又做工程，整个项目到处是重复的逻辑，通篇的函数，一个函数可以写 2000 多行，看得我简直内分泌失调。第一，一个问答接口函数写 2000 多行说明其实之前写这个功能的人对功能内部其实对问答的处理流程没有思考清楚，另外这样对 python 这种动态语言很容

易导致变量名覆盖和相互影响，更重要的是，这种写法完全没有抵御需求变化的能力，因为牵一发而动全身。稍微简单整理就会发现，问答的过程至少可以简单地分为问题预处理、匹配、组织协议数据等过程，每个过程用单独的方法去处理都不会写得这么乱，所以写代码一定要对让各功能独立成为一个模块，这样的自由组合当需求改变时，变化会非常小。第二，因为问答的过程涉及很多参数，我们用了 redis 缓存数据，所以函数式的编程导致到处要传大量参数，到处在取 redis 数据，大量重复的 IO 操作影响了性能。所以在这里，面向过程的编程明显是不合适的，面向对象，将大量公共参数作为实例变量，redis 数据做本地缓存，就可以解决这些问题。

第三个，熟悉 numpy、DataFrame 的用法，熟悉 NLP 各种算法。虽然做开发不需要研究算法，但需要对算法没有理解透的话，当算法工程师把他们的代码扔给你的时候，你无从码起。更重要的是，他们拿过来的代码并不是可以拿来即用的，三四秒的接口响应时间谁受得了？他们更注重的是效果，而作为开发除了效果还要看性能。没有对 numpy、DataFrame 的计算原理有一定的理解，没有对 python 的 GIL 锁、多线程、多进程有一定的理解，性能的优化是无从谈起的。

基于上面这些，我们公司招工程化人员的时候，更看重的是对原理的把握和挖掘能力，特别是对算法原理的把握，面试前起码把各种算法都实现一遍吧。

来自菜鸟程序员血淋淋的教训：从最初各种被拒到最新年薪近 60 万

本人只是一枚老码农，不生产代码，只做代码的搬运工。报过七月的集训营和 NLP 就业班，刚接到 offer(月薪 4 万多 \times 14，接近 60 万)，不像大神一样可以拿下很多大厂的 offer，我就总结一下我的教训吧，希望大家引以为戒。

首先拿到进入面试的门票，就是你的简历。如果你是双非本科，尽可能突出项目和经历，通过工作经验获得面试的机会，突出熟悉的算法，机器学习和深度学习都要有所准备。工作经验中突出你的算法项目，展现出你的突出贡献，突出你熟悉的算法，为面试问题做准备，先获取面试机会。

其次，准备 code 题，刷剑指 Offer，刷 LeetCode，由易到难，步步入坑，不要一下跳进去，不然很容易进入懵逼的状态，很难缓解尴尬的局面。由简单的数据结构原理入手，一般也就快排、堆排序、第 k 大的数等，也比较常见。放低姿态，不要以为学了点皮毛，就天下第一 老子无敌了。必须把基础打牢。

进阿里面试，在项目上和面试官叨叨了近一个小时，感觉吹的还行，结果让做道代码题（取数据流中第 k 大的数），当时给出了排序的代码，在循环遍历，面试官鄙视了我一眼后，还很有耐心的引导，半天没憋出个屁来。最终 game over。

天眼查面试，让写一个循环遍历，被我鄙视了，老子写这么简单的。然后没结果了。

京东的面试，还有几家公司都是有代码题，没给你出代码题的，自己多加小心，一般这家公司啥也不懂，希望你是万能的。

面试了某某在线，没代码题，就是干聊，一点稀的没有，薪资不合适，也就告辞了。

也有一些传统的企业招算法，比如某某幸福，也是尬聊，这次有个真的 nlp 工程师，我可能吃药多了，怼了他一顿，然后就没了然后了。所以说心态很重要，越是大佬，越是那种低

调，待人有礼，保持良好心态，面试官很多时候也会引导你的。但是 code 题一定要刷，不会 code 那就莫装逼。不然像我一样丢了很多机会。

再次，回答问题要有逻辑感觉，问题现象-----问题原因-----解决问题。

比如问怎么解决过拟合？这是送分题，加正则化、什么 droupout，最好多讲一些比如你是怎么判断的过拟合，过拟合有什么好处，什么坏处，为什么产生这种问题，从哪几个方向可以解决过拟合（训练集、测试集的不同表现-----模型学习到了内容，但泛化能力弱-----数据少了、模型太复杂了-----增加数据、降低模型复杂度（L1、L2）、bagging 思想）

如果你只是回答正则化，那面试官肯定问你，什么是正则化，L1、L2 的区别？怎么选？为什么这么选？等等都是挖坑的线路，如果你能都准备好那么还 ok。

你如果准备 xgboost，那么只是回答推导公式，没用的兄弟。

本人面试某车网，我回答了 xgboost 的推导，面试官直接问 xgboost 为什么快？

怎么分裂的？特征重要度怎么计算的？分位点怎么确定的？怎么自定义损失函数？

等等所以你准备的算法基础，要全面，在网上查看知乎相关的讨论，一定深入算法，通过网络查找关于该算法的面试题，从面试各个角度分析算法。

不要只停留在了了解算法原理的基础上。注意细节的深入。

面试某氧 App，随机梯度下降的抽样是有放回，还是无放回？为什么？随机梯度下降，有噪声怎么办？注意心态，不要慌，稳赢。人一懵逼，绝对死翘翘了。

同样在这家面试，因为前一个回答的不好，回答 auc 的时候就慌了，面试官问什么是 auc？

图中四个点分别代表什么？这稍微推理一下就可以了，当时被虐了，不敢嚣张跋扈了。

杜绝问答式的面试，提倡聊天式的回答。

最后就是 hr 了，记住要低调，低调做人，高调赚钱，态度要端正，嘴巴要紧，不要轻易松口，也要尊重 hr 和公司文化。千人千面，每个人遇到的问题不一样，但只要想好想细自己的场景，按照问题现象、问题原因、问题解决方案的路子，应该都可以找到工作。如果你已入坑，那赶快去准备。如果还未入坑，入坑有风险，请谨慎而行。

最终还是实力说话，我命由我不由天，是魔是仙我说了算，只要你肯努力，就一定可以。说你不行，灭了他个鸟命。

我面试中遇到的问题：

L1、L2 的区别？不要只回答截断和缩放，要展开一些，最好在二维图中画出，并解释。

随机梯度下降是什么抽样？为什么？

什么是 auc，图中四个点代表什么？auc 怎么计算？面积 auc 相同的时候怎么判断？

bias 与 variance，过拟合中谁高谁低，欠拟合的时候呢？

xgboost 的分位点怎么计算？怎么自定义损失函数？

解释 k-means？LR 的损失函数？svm 的核函数怎么选择？

不同时间浏览商品的权重标签怎么设计？

池化层的作用？

了解哪些采样方法？

你了解哪些主题模型？

今日头条的热度词怎么实现的？

spark 产生很多小文件怎么解决？

什么是凸函数？（这个是我自己挖的坑，大家千万别给自己挖坑）

文本分类的时候，我的真实样本标签可能会经常改变，你有什么思路解决？

双非小白找工作教训

本人不是大佬，所以叫讲一些教训，讲一些琐碎的知识点

一致性

注意分区 （项目的罗列方式：参与的什么项目 使用的三元组解决方案，使得上线业务点，、、结果 抓到 在什么。。有相当的经验）

标准的关键字

技能点 放在上面 能让人家抓到主题

个人主页 有的话，GitHub ???

多用数字说话，罗列事实 这个项目，现在到哪一步了，有什么样的进展、成果

要专业，表达出务实能干的字眼，开头简要介绍自己为何适应这个岗位 之前在哪里工作 有什么样的开发经验，（怎么样适合这个岗位）

项目不要冗长

时髦的技术不要罗列，命中率，准确率，简明的一种：做了什么 达到什么效果

相关的专利

HMM 必须掌握，NLP 的技能点！！！！，具体的 model，实现，引导

公式要不要推

我在。。。。作为 baseline，然后利用。。。。。（想让面试官问的点）

个人简历，吐字要清楚，思路清晰，分析遇到什么问题，怎么解决。

异常点，怎么定义异常数据，行业统一的规范，

数据分析遇到的技术，一定要脱稿

如果编，出现错误怎么办？

过渡，故事讲圆，让他觉得很懂，听上去合理的答案

贴合业务场景

Cnn 为什么，综合

为什么选择这个。。。。inception-v3?

采用那个什么学习框架，

做过什么，擅长什么？只要掌握就行，

准备的怎么样，自己能够脱稿讲

知其然知其所以然

过拟合的解决方法

resnet 深度残差

python 的不可变类型 可变类型

元组哪种

元组能否做字典的键

jieba 分词原理

svm 怎么算的

核函数原理，有几种

剃度下降，牛顿法

激活函数有几种，sigmoid 优缺点，剃度消失原因

100 万个词向量，找到和已知的一个最相似的，怎么做比较快

3 种特征选择方法

lasso 回归 岭回归

大概就是以上这些了

留学生数据分析转 NLP 的面经: seq2seq、transformer、bert 基本必考

作为留学生,在寻找国内工作的时候相比于国内同学来说,在信息和精力方面来讲相对处于劣势,对于国内学生来说毕业的前一年被称为应届生,可以参加各项校招,对于留学生来说,提前一年来找会有一些时间、课业、距离上的不便,就拿互联网公司的笔试来讲,可能需要半夜 3 点或者凌晨 5 点爬起来做 1.5 到 2 个小时的编程题。

不过现在比较好的是大多数互联网大厂都接受远程面试,从一面到终面最后签约都可以远程完成,但如果想要参加国企央企类的校园招聘,就必须得进行面对面的交流了,对于留学生毕业的当年还是否属于应届生这个问题,对于不同的毕业时间、不同的公司可能会有些不同,如果是在 9 月前毕业的,就只能参加很多互联网大厂的校招了。

基于对于在国内招聘信息以及求职规划不明确的情况下,我报名了七月在线的一对一就业班,方向是 NLP,之前专业比较偏向于大数据挖掘分析,专业不是针对于某一个小方向,而是属于大且杂,不过之前对于深度学习、NLP 也有一定的了解,也动手推导以及代码写过神经网络框架以及 back propagation 的过程,但是做过的项目并没有很多,并且当时 leetcode 也是刚刚开始刷。

不得不说对于找算法类的工作,leetcode 绝对是基础,在面试时,基本都需要手撕代码,所以提前准备好这个方面真的十分重要,一对一辅导班中导师有推荐一些可以学习以及练手的项目,并在学习之后也推荐了一些可以写在简历上并可以继续研究的项目,对于一些问题也有很详细的解答,接下来我把面试中一部分的经验总结一下,希望可以帮助到大家。

面经中因为有一些时间有些久了,记得不太清楚,只把一些我记得的写了下来,并且没有分一二三面,因为实在记不清了,所以混在了一起

1 作业帮

算法题

不同二叉搜索树个数

最长回文字符

xgboost 和 RF 区别与联系

seq2seq 框架

seq2seq 和 transformer 区别

HMM 与 CRF 区别

还有一些项目的细节，以及一些其他的记不清楚了

概率题：如何判断一个城市的平均收入，如何判断哪些收入属于高收入，哪些属于低收入，

如何对平均值进行估算

2 美团

聊的项目比较多，有问很多项目的细节

数据从哪里来，数据量有所少，为什么使用这个模型网络中用了多少层

文本如何表示如何做预训练，预训练区别和优缺点等

seq2seq, transformer, lstm, rnn 等框架的基础推导

目前所使用的这些框架有什么优缺点

GBDT 的原理

XGBOOST 和 GBDT 相比的优点

SVM 推导

fasttext 和 word2vec 区别

3 某医疗公司

解释 attention

multi-head 好在哪，每个 head 是使用全部还是局部

softmax 是否可以作为激活函数，为什么

LSTM 中为什么使用 sigmoid 和 tanh 函数

CRF 详解

BLEU 详解

seq2seq 详解

处理 UNKOWN 数据时应该如何处理最好

word2vec 是否可以用作预训练

4 头条

算法：

括号配对问题

股票买卖最大利润（只够买卖各一次已经无限制次数）

还有一道 SQL 题，但忘记是什么了好像是计算工资类的

数据预处理方法

如何处理缺失值

Bagging 和 Boosting 区别

seq2seq, transformer, Bert 详解

attention 有哪些、区别

为什么使用 attention 优点

word2vec 和 fasttext

生成式学习和判别式区别

tf-idf

l1 l2 正则

sigmoid 梯度消失问题

概率：一桌人戴帽子，看到别人但看不到自己的，可以举报，第三轮举报，问至少有几个人

戴帽子（其实和之前的打狗的概率题是一个意思）

还有一些项目相关的以及一些其他问题（比较难写出来）

今年转行从机械到算法

教育背景：

本科毕业某渣本机械设计制造及其自动化，硕士就读苏州某 211 机械工程，课题方向激光 3D 打印。。。

个人的竞争优势：

- (1) 飞利浦 AI 图像算法的实习，完完整整做了 2 个项目，并有一些成果；
- (2) 研究生数学建模国 3；
- (3) 校数学竞赛一等奖，二等奖；
- (4) 一作 sci 论文一篇（在飞利浦的工作），一作 EI 论文一篇；
- (5) 本科成绩 top1，研究生入学第一；
- (6) 深度学习文献读的多，相关知识较为扎实。

面试经验：

因为想在苏州发展，只参加了苏州的招聘会，结果签了南京的公司。。。早知道去杭州找工作了。。。

公司 1：安硕信息(做 nlp，14k，苏州)

笔试+两轮技术面

笔试内容：（1）给了一段 AI 方面的论文让你翻译；

- （2）手撕代码，寻找链表中的环；
- （3）常见的树模型有什么区别；
- （4）K-means 算法流程。

技术面：

主要针对简历还问，主要问了我的实习项目还有数学建模国奖的解决方案。

公司 2： 万得资讯（做 cv， 17k， 南京）

笔试 + 2 轮面试

笔试内容：（1）介绍常用的分类网络，及其优缺点；

（2）给了一个网络结构让计算感受野；

（3）给出一个人脸防伪的方案，及介绍有什么缺陷和改进方向。

技术面

手撕代码用 tensorflow 实现个简单的猫狗分类。如何防止过拟合和欠拟合（为什么那个方法能防止过拟合或欠拟合）。手撕代码用 numpy 卷积。手撕数学如何坐标系变换。。。

公司 3： 美的图算算法岗（挂）

在线笔试+群面+2 轮电话面

笔试内容：

都是选择填空题，常见的一些概念，还有梯度计算，感受野计算等

群面：

10 个人一组，共同讨论一个项目并给出设计方案。这属于团体项目，所以一定要找到自己的定位，要踊跃发言自己的看法。

电话面：

这家公司苏州的业务是做扫地机器人的，偏 SLAM 方向之前没有相关的经验，把我刷了。。。

整个秋招拿到了 8 个 offer，很多挂掉的都是主管面挂掉的，每次主管面都会大大地质疑我的专业。。。转行人的心塞。。当然也有些骨骼新奇的主管，觉得你一个大机械居然来找算法工作还通过技术面，那是真的热爱啊~哈哈。

一些建议：

（1）如果做深度学习的话一定要多看文献，不然很容易浮于表面；

(2) 如果实验室不是做这个方向最好找一个实习;

(3) 一定要刷算法题, 像《剑指 offer》和 leetcode, 还有七月在线的题库可以多刷刷, 整理的挺全的。

(4) 找工作心态一定要好, 屡败屡战总会上岸~

希望大家都能找到自己心仪的工作~~~

励志！充满干货的 AI 面经：纯电力员工如何成功转行 NLP 并薪资翻倍

大家好，我是集五的学员，经过长时间的学习，我成功从电力行业转行 AI，薪酬翻倍。

但需要说明的是，作为电力行业的基层员工，工资本来就低，所以工资翻倍对于各位本就在 IT 行业内的朋友来说，没有什么参考价值，纯粹贻笑大方，只能我自己偷偷窃喜一番。

应 July 校长的邀请，我腆着脸来班门弄斧，和大家分享一下面经和经历。单刀直入，先分享我的面经，至于转行的部分和我的经历，我放在最后面，这样有兴趣的朋友，可以参考。

我选的方向是 NLP。

通识部分 (小编注：绝大部分面试题都可以在七月在线 APP 中通过搜索相关关键词搜到哦)：

1.LR

介绍一下？公式推导？损失函数及其图像？与线性回归的关系？你在什么地方用过？怎么用的？

2.正则化

解释下正则化？

3.DT

ID3？C4.5？CART？区别，优点、缺点，计算公式？

为什么 DT 不适合稀疏数据？

预剪枝？后减枝？

3.1 Adaboost

介绍一下 Adaboost？手推 Adaboost

3.2 GBDT

GBDT 腾讯面试必问，因为我去的鹅厂现场面试的，被问到了 GBDT 的相关内容，而旁边

一个面试官在电话面试，也是在问 GBDT 的内容。

介绍下 GBDT？GBDT 的基模型是什么，是分类树还是回归树？Gradient 的体现？为什么要使用多棵树来不断缩小残差？而不是使用一棵拟合得更好的数？GBDT 如何构建特征？

GBDT 如何用于分类，写出公式？

3.3 XGBoost

手推 XGBoost 必备的。

你在什么地方用到了 XGBoost？它和 GBDT 的区别在哪里？你为什么不用 LightGBM？

XGBoost 不手推一下，说的内容感觉会很苍白。

3.4 随机森林

解释下随机森林？

你用它做过什么？为什么可以用它？

解释下 bagging？

4.SVM

手推 SVM，必备的。

SVM 的损失函数是什么，为什么选择它作为损失函数？

为什么用对偶？

核函数是什么，有什么作用？你用过哪几种核函数？这里最好写出来。

5.深度学习部分

解释一下交叉熵、相对熵，它们的关系？互信息呢？

Cross Entropy 是什么，它和 log loss 的区别？

Dropout 防止过拟合的原理？

CNN 如何用于 NLP？CNN 和 RNN 在 NLP 中能到达什么效果，怎么选择用哪一个？

梯度消失？梯度爆炸？

解释下 LSTM，画出结构？GRU？为什么用 LSTM 不用 RNN？

6. 特征工程

如何处理缺失值？异常值？为什么这么处理？

怎么降维？遇到的数据最大多少维？你怎么做？为什么这么做？

类别型数据的处理？连续性数据的处理？你怎么分桶的？

数据不平衡怎么做？要注意什么？

7. 聚类算法

你用过什么聚类算法？解释下原理？

K-means 有什么问题？如何解决这些问题？

说一说“中餐馆问题”——抱歉，这个我现在还没搞明白。

8. 降维

解释下 PCA？LDA？SVD？区别？手推。

解释下协同过滤？

9. 模型评估

Accuracy、Precision、Recall、F1-score、ROC、AUC

以上，概念、公式？

数据不平衡用什么评估指标？

P-R 曲线和 ROC 的区别？

NLP 方向的问题：

1. TF-IDF

原理、公式？为什么用 log？为什么“+1”？缺点？

2.NNLM、Word2vec、Doc2vec、FastText

介绍一下，解释一下模型？

怎么选择使用 Word2vec、Doc2vec？

解释下 Huffman Tree 或 Negative Sampling？详细说下 Huffman Tree 或 Negative Sampling 的过程？

Word2vec 有什么优点、缺点？

Word2vec 的模型参数怎么设置的？

余弦相似度？为什么使用余弦相似度？

3.Seq2seq、attention 机制、self-attention 机制、Transformer、bert

介绍下模型？

attention 机制、self-attention 机制的计算过程？为什么使用 attention？它们的优点？

bert 的结构和应用？bert 的为什么这么厉害？

CNN→RNN→LSTM/GRU→seq2seq→seq2seq-Attention→Transformer-self

Attention→bert

这是一条线，我在说 RNN 的部分时，就把这条线顺着说完。

NNLM→Word2vec→Doc2vec→FastText→ELMO→GPT→Bert

这是另一条线，被问到 Word2vec 时我先退回去说 NNLM，然后在 Word2vec，要是面试官不反感，又继续沿着线走。

其他：

1.根据简历上写的项目来进行具体的提问？

数据量？数据来源？为什么想到真么做？这么做优点、缺点？

Python 迭代器和生成器？

2. 算法题

不说了，我面了那么多，一道题都没答上来。校长说的四大金刚之一：leetcode（小编注：校长的口头禅四大金刚指的就是：课程 题库 OJ 竞赛），因为自身原因，我战略性放弃了，之后一定要补起来。

去某手机厂面试的时候，上来直接来 3 道题，把我干懵了；

对于这方面能力和我不相伯仲的朋友，我建议电话面试优先！一般电话面试不会去问算法题。

3. 非技术面

为什么转行？

这是转行的朋友肯定会被问到的问题，一定想好说辞，好好准备。

未来规划？

薪资期望？

你遇到的最困难的事？你如何解决的？

评价下你自己？

你有什么问题要问我们？

我的笔记全在印象笔记上，自己又根据七月在线的题库对问题和答案进行了整理后，制作了 Anki 卡片，用来记忆各种知识点，如果有朋友需要，请看本文的置顶评论。（另外关于 Anki 和它的使用，自行百度即可，记忆知识点的神器。）

转行的经历

学历本科，专业是动力，完全与 IT 无关，毕业后一直在电力系统做设计工作。后来工资实在太低了，活不下去了，得知我单位的一个好兄弟，转行进了华为，年薪 50W，目瞪口呆之下询问了一波他的经历，知道了七月在线和机器学习。再三权衡之后，也下定决心准备转行，从事 AI 相关的工作。当然，他是清华核物理硕士，数学本来就很骚，大学期间学

过 Java，我和他没有可比性。

我没什么编程基础，大学里学的 C++ 早被门外的旺财吃了；因为是工科，对高数、概率论、线代还有点记忆；机器学习、深度学习相关算法，在此之前从没接触过。

报名了集训营第五期，是我的开端。不会就学呗，不懂就问、就查呗，CSDN、Github、stackoverflow、群里提问，线下课有问题就问，没什么方法。看一遍不懂就看两遍，不行就三遍以上，搞懂，能推导。没什么捷径，真的。我是用了七八个月的时间，才达到了目前这个嘴炮菜鸡水平，未来的路还很长，要学习和弥补的东西还很多，不过转行了，做中学吧。我另一个兄弟，也是我的同事，集六的学员，在我之前拿到的 offer。

推导公式，很重要，非常重要，就算面试官不让你写，你也抢过纸和笔，写出来，不然我感觉真的太苍白无力了。而且现在水涨船高，SVM 和 XGBoost 的推导已经是常规操作了，不是特殊技能了。

Tensorflow 必须掌握，别说自己会 keras 了，丢人，真的，不开玩笑，至少我遇到公司都是这样。简历上最好把深度学习的知识具体的体现出来，比如你在某个项目上搭的模型是什么样子，这样更容易获得面试机会。

最好是会使用 Hadoop 和 Spark，这也慢慢成了必备技能了，我目前还不会，虽然集训营讲了，得赶紧在入职前学习一波；最好会 SQL，这个我就完全不知道了，后续又学习。我觉得最重要的课程是七月在线的机器学习第八期或第九期；西瓜书、统计学习方法、花书，要是急你不看也行，但是我上面提到的问题，你必须都知道。

关于找工作，我是海投了一波，广撒网，双手合十，一切看天。真的得看运气，说不好。拿到 Offer 之前，少吹牛。我去年年终奖比别人少了一倍，怒火中烧，年会吃饭直接避开领导，当面回绝敬酒，只和身边二三关系要好的同事酒到杯倾，一饮而尽，并笃定说年后自己要走。但是几次面试折戟之后，心里开始慌了，想到前路茫茫和自己之前恬不知耻的吹

牛夸大，实在是羞愧难当。但是家境又确实不允许我辞掉工作去完全学习和等待面试，得断粮饿死，所以又准备腆着脸继续忍辱负重，希望守得云开见月明。没必要的牛皮给自己又添上了沉重的心里压力，真的没必要。“闷声发大财”，自己实力过硬是关键。

我现在还是孤陋寡闻的小菜鸡，要学习的东西还非常多，不敢说分享成功经验，只能说分享下自己的经历。现在成功转行了，我还是很开心的。当然，我能够成功转行，要感谢很多人：助教 Yang，助教马总，寒老师、July 校长等，集五的波哥、成哥、飞哥、Miss Jiang、锋哥、长林哥等等学员，还有甚者线下的 Paul、安哥、瑞哥等，集四的九州、思佳，集六茂茂。你们对无知的我的给予的帮助，给了我非常大的助力，一直铭记于心，十分感谢。

NLP 岗位面试经历：很多面试题都能在题库里找到

本人是集七学员，工作之余也看了不少经典书《统计学习方法》、《机器学习》（周志华）等，看了不少技术博客，但是总觉得没有系统地学习机器学习，不能够对机器学习有教全面的认识。

后来就报了七月在线平台的培训课程，由浅入深，稳扎稳打，收益匪浅，四月份结束培训，便开始找寻相关工作，之前课程也有指导找工作的方向，算法岗位细分也不少，CV、NLP、风控、推荐方向等，课程老师也会详细介绍每个方向的重难点，并给相应方向的许多项目学习，项目驱动学习，才能更好更快地学习，也是快速进步的必要条件。

在选择算法方向上，考虑到本人工作经历有过 NLP 项目经历，于是便选择了 NLP 方向，老师也帮忙认真指导了之前的简历，包括简历的格式、项目描述的完善以及其他针对性的修改意见。

现介绍我在面试过程中，一些经常被问到的技术与项目上的问题：

机器学习相关：

- 1 □ L1 与 L2 的作用；
- 2 □ SVM 的原理；
- 3 □ bagging 与 boosting 方法的关系与区别；
- 4 □ 控制过拟合的方式；
- 5 □ Xgboost 相较 boosting 方式的优势，以及 lightGBM 相较 Xgboost 的提升；
- 6 □ word2vec 的原理与改进方式；

深度学习相关：

- 1 □ 卷积神经网络相较于传统神经网络的优势；
- 2 □ 常用激活函数以及其变种；

- 3 ☐ LSTM、GRU 是如何解决梯度消失问题的；
- 4 ☐ 神经网络是如何解决过拟合问题的；
- 5 ☐ 介绍下 attention model 的基本原理与应用场景；
- 6 ☐ 讲讲你对 bert 模型的理解；

项目相关：

- 1 ☐ 介绍下自己做过的 NLP 相关的项目；（本人主要讲的是智能机器人）
- 2 ☐ 模型的算法用的是啥？准确率？如何评价模型的效果？
- 3 ☐ 针对算法的优化？
- 4 ☐ 文本特征提取的方式有哪些？
- 5 ☐ 文本分类的模型有哪些？

其实以上这些面试题好多都是可以在七月在线 APP 上的题库可以找到，在工作之余好好看看，结合项目加深理解，找到 NLP 的工作还是相对比较容易的，希望大家能沉下心，切勿急躁，好好学，肯定能找到自己满意的工作。

AI 面经:我是如何从机械行业本科社招成功转行 NLP 并薪资翻倍的

抛砖引玉，大家有什么问题，可以回复讨论，我尽量答复。

1、本人情况

一本机械专业毕业，从事传统制造业多年，处于“鄙视链”的底层。

2、学习经历

参加了七月在线机器学习集训营 第六期。本人除了数学，其他基本零基础，毕业又多年，还是在职学习，学习还是比较吃力的。我自己的学习宝典：

➢ 七月在线的视频：注意老师说的每一句话，都可能在面试中被问到。

➢ 书：用于补充理论。

➢ 网络视频：很多公开课和一些大神的学习视频。

➢ 博客：网络视频还是搞不懂的情况下。

➢ 论文：最新的研究成果，google 大法好！

➢ 老师和同学：七月在线真是卧虎藏龙的地方啊！

我是在具备了以下这些技能下才开始投递简历的，机器学习模型知道原理，部分会推导，深度学习模型能画出模型结构，以及课上老师提到的常见问题基本都掌握：

➢ 机器学习：SVM、LR、DT、RF、Boost、k-means、DBSCAN、NB、HMM、CRF。（其中图模型基本没被问到）

➢ 深度学习：DNN、CNN、RNN\LSTM\GRU、Word2vec\Doc2vec、FastText。

➢ 数据结构和算法：链表、二叉树、排序、动态规划（被考到一次手写）。

面试过程也是学习的过程，面试过程中所获得的技能有：FAISS、Transformer（恶补）、

Self-Attention、Bert、Linux 常见命令、知识图谱（了解）。

3、面试经历

抓住金三银四的契机投了不少简历，大撒网。很多公司的 JD 上写的要求硕士甚至博士学位，但这个不是硬规定，只是加分项，可能会对“身价”有轻微影响，厚着脸皮投吧。要求的相关专业同样如此。这样，年后我大概投了 50+ 公司。

年后的第一周就面试了 7 家公司，基本都是小公司。面试官不太偏重理论，基本流程是：第一阶段问一下项目大概情况、以及所用模型的原理，这个阶段我没有什么问题；第二阶段，他们以工作中遇到的场景，问我有什么具体的解决方案。

第二阶段的场景把我难住了，这里列出我遇到的两个场景：

- 1) 一篇文档，如何自动获得词与词之间的关联度。这里我用的是 textrank 的思路回答的，大家有什么更好的方案，请告诉我，谢谢！
- 2) 用户注册 APP，填写性别，但用户自己填写的性别会有大约 10% 的错误率，这里并不知道哪些是错误的标签，请问如何通过用户行为特征，对这些性别标签进行纠错呢？现场我给出了传统半监督的方法，但因为并不知道错误/正确标签的数据是哪些，所以这个方案是行不通的。

面试后，群里大佬们给出的解答：先聚类成两类，然后离中心点近的样本权重高，离中心点远的样本权重低，用带权重的样本去学习模型，借鉴的是 attention 机制。大家如果有更好的方案请务必告诉我，谢谢！

第二周，大厂们的 HR 开始行动了，大厂一般会有好几轮面试，这里也只说说技术面。一般流程：首先看你简历上写的专业技能，抽出几个典型的模型问原理，迭代过程，当然会有推导；接下来是项目中各个环节的处理方式，特征工程为什么这样处理，模型的表现，以及你认为优化点有哪些，有没有进行尝试。

大厂们的工程师技术确实过硬，提到的问题基本都是项目中的关键点。当然，也会给我

场景，问有没有解决方案，但我遇到大厂的场景基本不那么复杂，抽象为分类、回归问题就能解决。

4、总结

因为个人原因，学习阶段确实花了很多时间。比如 xgboost 的学习，从刚开始的懵逼，到最后会推导，了解生成方式，了解模型中每个参数，甚至调参策略等，前后我至少花了 20 个学时。学习下来，个人认为，AI 领域的知识其实并不难，只要肯花时间，肯定能够理解的，不像我之前从事的行业，网上公开的资料基本没有，那才叫煎熬啊。除了理论，实践当然也很关键，参加比赛真的对我提升不少，特别是代码能力。

前后面了近 10 家公司，总体感觉今年的要求普遍比去年高，像 NLP 最新的模型

Transformer\Bert，好像成为 NLP 算法工程师的标配了，40%的几率会被问到。面试中肯定有自己不会的问题，这个时候我会说：我在项目中没有遇到类似的场景，但接触后肯定能够很快上手，再厚着脸皮请教面试官。由于是初涉及这个行业的新手，也遇到过不少坑，面试是双向选择的过程，最好在面试中就能识别出哪些公司适合自己。

Anyway, 学习阶段最重要的是：自律，面试阶段最重要的是：自信。再次感谢七月在线平台，感谢校长及各位老师和各位大神同学，以及小杜。最后，无论大家成功与否，都祝大家前程似锦！

谢谢！

刚毕业双非本科如何拿到 NLP offer 并薪资翻三倍的

在去年毕业没几个月的时候，因为不满意自己工作，于是来到了七月在线机器学习集训营学习，这也是我第一次真正接触七月在线这个平台，学习了一段时间收获了很多，不仅仅是机器学习、深度学习知识，还学到一些实用的 python 数据分析、spark 入门知识以及 NLP 的知识。于是赶着金三银四好时光投了一些简历，拿到了薪资还比较满意的 NLP offer（接近之前三倍，之前太低了）

下面我来分享一些我最近面试经历：

1.个人情况介绍：

去年双非本科毕业、计算机专业、写了一些博客(<https://tkhajwjy.github.io/>)（顺便复习知识）

2.简历问题：

简历上技能一定要写自己熟悉的知识，不熟悉的不要写；项目要贴合自己之前工作业务，要不然到时候解释不通，如果实在没有，可以写比赛。

机器学习问题：

逻辑回归公式？交叉熵公式？用 XGBoost 做过什么项目？GBDT 和 XGBoost 区别与联系？集成算法有哪些，介绍一下？

深度学习问题：

梯度消失、梯度爆炸原因，怎么解决？textCNN 原理？pytorch 会用吗？

NLP 问题：

TF-IDF 介绍一下、缺点是什么？word2vec 是什么，怎么优化的？BERT 了解么，用过吗？

项目问题：

数据从哪来的、数据量多少？项目最终成果是什么样的？

你主要负责项目哪一部分？

数据怎么预处理？会不会 spark？

模型如何部署上线？

如何评估模型，解释下精确率、召回率、F1？

有的时候线下模型准确率很高，但在线上效果很差是什么原因？

3.其他问题：

为什么离职？为什么不考研？职业规划是什么？你期望薪资最低是多少(不要说自己真实想法，要比心理预期高一点)？项目紧的时候能不能接受周六上班？有什么要问我的？

小结：

非常感谢七月在线为我提供了好的学习平台，在这里不仅遇到了优秀的老师，还遇到了很多优秀的同学，在群里提问都会有人解答，而且线下还有安静的教室。虽然自己暂时找到了停靠点，但是 NLP 还有很多知识去学习，接下来我还会依旧选择在七月在线学习，希望七月在线越办越好，造福更多学子！

集 4 薪资翻倍的 AI 面经：曾担心学完集训营依然找不到工作

1 恭喜从移动端成功转行 NLP，较培训之前的上一份工资年薪涨了 15 万，请简单介绍下你自己（教育经历、工作经历）

赵同学：hello 各位好，我是 16 年毕业于一所普通的一本学校，专业是软件工程，毕业后在一家游戏公司做移动客户端开发，去年十月开始接触机器学习，但由于没有工作经验面试机会都没有，参加完第四期培训后顺利找到了月薪 double 的工作，任职 nlp 工程师。

2 是什么原因促使你从传统 IT 转行 AI

赵同学：目前 AI 行业大热，也是未来互联网的发展趋势，相比于其它岗位，薪酬也较高。其次，对于移动端开发来说，大部分都是业务代码与重复率较高的功能代码。而对于机器学习来说，无论是算法的运用还是特征的构建，都充满了变数，更有挑战性。

3 当初报名集训营有哪些顾虑，后来如何消除的，请详细说说？

赵同学：最担心的还是怕因为没有工作经验集训完依然找不到工作。集训营期间对机器学习有了一个更加具体，全面的学习，加上后期的实战项目，有了更多能写到简历上的点，深圳的线下 nlp 项目也成了我找工作的敲门砖，非常对口，老师也针对一些面试中会问到的点给我们做了详细的说明解答。

在简历方面，七月的老师也提供了许多优化方案，集训完很快就收到了面试机会，面试问题也都是老师们经常提到的点。

4 来 ML 集训营 4 上收获到的最大技能或心得是什么

赵同学：首先是对机器学习有了一个更全面的认识，从基础的 python、pandas 到后期的算法，所有知识点都联系的很紧密，不像自学期间那么盲目，对机器学习的算法原理也了解的更加透彻，对于一些复杂的算法 svm 、xgboost 能做到熟练的推导出来。面对面试中遇到的问题也是游刃有余。

其次，在线下课也认识了许多优秀的老师与同学，学习氛围很好，我们也有学习的动力。

5 未来三年的职业规划是什么？

赵同学：首先肯定是先对目前的工作要有一个更深入的了解，包括项目的架构，算法原理，之后会继续学习 nlp 方面的知识，把一些难点学精学深，在多层会话、知识图谱方面能有更深入的了解。其次，在其他领域也要去不停的学习，紧跟当前的热门，希望在未来能有更好的机遇。

ML 集训营 3 学员专访 2：我是如何从 iOS 成功转岗 NLP 拿到 45 万的

1 首先再次恭喜 iOS 成功转岗 NLP 拿到年薪 45 万 offer，请简要介绍下你自己（教育、工作经历）

石同学：本科毕业以后，互联网教育、电商和企业服务干过，主要做移动端开发，早年也做过后端业务系统。业务系统做过.NET，移动端做的比较久，数据方面，SQL 用得还行，其实工作中还是偏业务。

2 为了这次转岗，过去半年，你做了哪些工作

石同学：最开始是自学，效率确实低而且不知道从哪下手，后来报了课程边工作边晚上周末学习，但是后来发现转型学习需要的时间太不够，所以辞职做转型，专心学了一个半月，基本是不断学习-实践、学习-实践，然后学完开始边复习边找工作，之后准备边工作边学习。基础还有不少不扎实的地方，需要继续学习。

3 从集训营 3 收获到的最大技能是什么

石同学：比较完整的知识体系吧，机器学习和深度学习课程能学到很多，从 python 语言，爬虫、pandas、numpy 数据处理基础库和数据处理技能，还有可视化技能，在做比赛和业务分析的时候感觉特别有用，大数据处理，寒老师和 David 讲的深入浅出，收获很大，机器学习、到深度学习，再到最后的项目课都获益颇多。周末的线下部分可以当面答疑，解决很多疑惑。

4 对于传统行业转岗 AI 的同学 有何建议

石同学：最好是从基础课先学，比如机器学习第八期就有很多干货，但是很多看完就忘了，需要再看。行业发展太快，技术趋势也不断地在变化，其实还是不能浮躁，踏实学习、工作，这个也想学那个也想学，最后都学不好。项目很重要，每个项目都练过，比赛真的自己做过

基本就不会忘，但是真的不能着急，业务背景公司也比较看重，谈薪水时候能加分。

5 未来三年有何职业规划

石同学：先入行，工作中边做边做，通过实际业务场景锻炼技能，慢慢深入这个行业，继续升级技能树。毕竟还是很有前景的方向，有很多可以做的事。深度学习比较火，想在方面继续深入，需要学的东西也太多了，但是不能心急，扎实基础，稳步前行。

第五部分 推荐系统/计算广告

推荐算法工程师面试心得分享：课程进行过半拿到 offer 并薪资翻 1.5 倍

本人背景介绍，本人 985 硕士一枚，已毕业一年半，目前从事推荐算法工作，之前工作做过相关算法工作，但是感觉学习的不是很系统，所以希望通过就业班来提升自己（小编注：这位同学学的是还在上课的推荐系统就业 3，课程刚进行一个多月就找到了新工作，薪资较之前翻了 1.5 倍，厉害），闲话不多说了，直接说经验哈！

我第一家面试是京东到家。

第一面，面试官是女的，问的问题也很专业，我首先介绍了一下自己的基本情况，然后就直接介绍我的项目，和我做过的一些工作，面试官根据项目问了 word2vec 的一些原理（因为项目涉及到了），还有其中损失函数的公式，然后又问了其中，什么是层 softmax，如何构建的哈夫曼树，负采样是如何采的，这些是根据项目问的一些问题；

然后问了一些使用的模型，问了 xgboost 和 lr 是如何组合使用的，特征是如何选择的，gbdt 的梯度是如何体现的等；

最后出了一个动态规划的算法题，一面结束，总体感觉第一面的面试官问的东西还是比较细的，而且都是根据你简历里的东西去问的一些东西；

第二面，面试官应该是一个 leader 级的人，他问了一些基础的召回问题，然后又问了一些我换工作的原因，二面的技术性不是很强（可能是一面问的太多了吧）；

hr 面，主要问了一下，换工作的原因，期望的薪资，然后我又问了一些公司的基本情况，hr 面因为这个是我面的第一家，所以价格可能要的有点高，然后就说可能给不到这个价位，问我是最低价吗，我当时觉得我提的挺合理的就回答了个是，然后就没有然后了。

第二家是已经 D 轮的互联网创业公司，这家公司是根据我的项目问了一些比较基础的召回问题，因为我用了 als，他会问我一些物品的打分是怎么打的，然后会问一些如果是异常用户行为，比如用户异常刷贴，那 als 又将如何打分，然后又问了一些使用的模型，比如，xgboost+lr，weed&deep，FM，DeepFM 等，之后又问了一下，特征工程是如何做的，最后出了一道树的算法题（比较简单）。

我整个的面试，其实主要是猎头和 hr 给我推的一些职位（一共 5 家，还有给我推数据挖掘岗的，直接忽略了），其中合适的就 3 家，我还没有投一些大厂，大厂一般对于社招我感觉希望招聘 3 年以上工作经验的，我的工作经验还是少，而且又因为面试不方便，所以拿到 offer 后，感觉跟 leader 还聊的来，就直接答应去了，因为不太想在原来的岗位干

面试心得总结：我觉得作为一个合格的算法工程师，首先要有扎实的基础功底，然后需要在一个方面有自己深度的理解，比如 word2vec 算法，知道怎么用，但也要知道他的内部原理和公式的一些推导，和在什么样的场景下使用，总体来说算法很多，我们学习大量的算法是好事，但是对于算法的深度必须要有，注重广度同时深度很重要，而且对于面试来说，我们也同时需要具备数据结构算法的一些知识，比如：树，动态规划等。

说了这么多其实，自己的决心跟努力是最重要的，不管你现在是后端岗还是数据岗，或者是在校学生，如果认定了做算法，就坚持下去，借用现在最火中国动漫的一句话，“是魔（坏）是仙（好），只有我自己说了才算”，这是我自己的一些拙见，希望能帮到各位。

推荐算法面经 最重要的是必胜的信念：普通大学没经验但有开发能力就能转行成功

现在回想起来,距离集 4 结束已经一年了,我在去年参加完集训营后找到推荐系统的工作到现在也已经一年了,回想起自己当初的面试和以后也要继续的面试我想为大家分享一些自己相关的经验。

首先我自己本身并不是计算机专业的学生,大学也不是什么好大学,但是我之前确实是有开发经验的,从 2016 年到 2018 年大概不到两年在北京一家电商工作,做网站开发。

是的 这份工作也是通过我当时毕业后找的培训班进行学习后来北京找到的,我想这和我们大部分朋友的经历不太一样,毕竟 ai 这个行业看起来逼格这么高怎么也得是 985/211 的硕士才能找到,也确实,我现在的身边的朋友也基本都是这样的,但是不是说像我这样转行的,一般大学的,没有经验的就不能从事这个行业了呢?

不是的.其实你和进入这个行业真正的距离,是你的想象力和信心,只要你有这两者那中间的过程就只是一些失败和努力而已,而找一个补习班,真的能从各种方向提升这个进度和速度,我想也许这就是大家常说的教育投资吧哈哈。

如果你是有想进入这个行业的愿望但又担心自己不能做这份工作的,我建议你考虑清楚,自己适不适合做程序员?自己有没有一些开发能力?如果有,那就放心大胆的做吧,这个行业现在是风口行业,只要你有这些能力,找一份还挺棒的工作还是一件不是很困难的事的说了这么多也没说到点子上,这篇文章是关于帮助大家找到工作的?哈哈,好我接下来就说一下在面试中需要注意的几件事

1、数据结构 对于想找大公司工作的大家,我想这部分是躲不过去了,如果不是想找稍微小一点的公司,这部分考察力度可能相对小一些,建议把剑指 offer 这本书看看,每天刷几道题

2、算法相关 对于这部分我不推荐大家没事去手推 svm 啥的,但一定要把自己项目中涉及到的算法都弄清楚,而且需要对大多算法有个基础的认识

3、工具相关 大公司一般会使用一些分布式框架和存储 如果时间足够的话 对 spark,hbase 等一些常用的分布式框架要多了解一下,注意如果时间不太够也建议了解一下,不然面试聊到这里会很尴尬

4、语言相关 对于除了 python 之外的开发能力的要求各个公司也是不相同的,如果你恰巧还会 java/c 等语言那肯定是更好,但是不会的话也不会太减分,但是注意自己的代码水平,面试前建议多刷 leetcode

5、如果你过了前面这些关,那恭喜你面对最后一个 boss 了,那就是项目,大家都说,那我没做过相关项目怎么写项目呢?这里需要和集训营的老师 and 同学多沟通,自己把自己的项目想通顺,不要有死角,不然让人看起来太像培训出来的,这部分是最重要的,可以在网上查阅一些你要写的项目的资料,再让老师进行修改,都让老师帮忙显然是不可能的

6、如果以上你都准备好了,那你只需最后一件事,清楚表达能力,建议大家在面试前多找小公司试一下,之后再投自己心仪的公司

以上就是我对面试需要的技能的总结,记住一切的一切最重要的是自信,不要因为几次面试失败怀疑自己,坚持下去,一定会成功的。

水硕跨专业：拿到自己满意的推荐系统方向 offer

首先，我是集训营第四期学员，双非的硕士，专业是轨道振动与噪声，一个完全和 AI 不搭边的方向。集训营 18 年 6 月份结束课程之后准备简历，但看到现在轨道方向和 AI 交叉的产品不是很多，所以决定考博士（有结合 AI 的方向），今年 5 月份复试失败，重新开始了学习找工作的过程。

其次，说说在集训营中的情况，感谢七月在线带入门，那时候依稀记得在集训营群里大家讨论问题的场景，老师也及时的回复，解答同学们问题，在此表示感谢七月在线的老师的指导，不仅是学习的指导还有人生规划的指导。

在接下来的学习过程，由于那时候给我们上课的是寒老师，受他的影响，决定推荐算法的方向。由于自己底子还可以，《统计学习方法》（第一版）里面的内容基本上从第一章可以讲到最后一章。看到七月在线官网上的《推荐系统实战》课程，当时课程已经结束，就购买跟着课程走，基本上懂得了 CTR 模型方面知识；在下载了很多课程中涉及到的论文结合强大 github，一定要在 github 上看论文相关的项目，看论文非常的空洞。

项目部分，在网上和 github 上找到了几个比赛和简单项目，分别是《京东-用户对品类下店铺的购买预测》（今年 6 月份结束）、《字节跳动-短视频内容理解与推荐》（5 月份结束），还有几个小的项目音乐推荐，电影推荐。自己拿着代码一边看一边跑，了解比赛和项目的构造，数据分析，特征工程，模型的构建等等。

简历部分，简历是公司通行证第一关，一份好的简历即使你是编的最起码有个面试的机会。自己有 HR 和猎头朋友，在他们的指导下

总结归于 2 点：

一点是简历刚开始都是通过 HR 进行筛选，而 HR 并不懂算法和大数据的东西，都是技术总监或者项目经理给 HR 招聘职位的 JD，有点像主题模型一样筛选简历，简历上的项目

匹配的程度有多大邀请面试机会就有大，所以简历多写几个项目是好的，因为匹配程度概率上来说更大。

二是简历的构造部分可以参考猎聘网上的简历制作格式，项目部分是 HR 最看重的；项目描述，项目职责及项目业绩写清楚，特别是项目职责像 python 列表一样写清楚项目的过程，不要信简历一页纸最好。

面试部分，因为知道自己什么水平，Letcode 和剑指 offer 基本没刷，所以大公司肯定不会去投的，在招聘网站上投了一些中小企业，刚投一天，第二天就有 3 个面试，第一个面试过程很好，一上来就问我简历上的项目是怎么做的，那就好办了，从背景、数据分布、数据处理，特征工程，然后到模型融合，评估指标等等，一讲就是 20 多分钟过去了，在问了我细节的问题，自己有跑过一遍，问题都不大；在问了一个 CTR 模型有哪些，我从 FM 模型开始，一直讲到 2018 的 XdeepFM 模型（主要讲模型的创新点及构造）讲了 30 多分钟，面试完直接 HR 面，拿到自己满意的 offer。

亲爱的七月在线学弟学妹们，希望你们在七月在线能学业有成，找到满意的 offer。写到最后的阶段，诚惶诚恐，我的经历相比较于七月在线社区面试板块来说是最差的，基础也是最不扎实的一个，面试经历也是最少的，但我的经历希望能给到你们帮助，谢谢！

大数据转行算法工程师面试经验

基本情况：大数据开发经验 2 年，有算法方面的项目经验和 kaggle 奖牌,然后因为自信和项目经验不足等因素，报了推荐算法就业班二期。这个班的内容主要由几个课程组成：数学（录播），机器学习第九期（录播）深度学习（录播）推荐实战（录播）和大概 7 节的直播课程和项目讲解直播课程。具体效果看个人吧，但是如果是零经验立志要转行的话，报一个班实际上收益还是不错的。

总共面了 5 家左右吧，只有一家给了 offer，title 是算法工程师。

说一说面试主要问的问题：

1、普通的算法（leetcode 题库看一看，刷一刷）频率不高，5 家的话记得是由 1 家有问到。

比重不多，5%左右。

2、项目（项目背景，技术选型，如何实现，模型选取，主要负责部分，对应算法原理，效果评估）这个比重会高很多，大概 80%都是从这里进行问题的衍生，比如写了 gdbt，就会问原理，优缺点，为什么。AUC 这个指标实际上是做什么用的。比重 80%

3、常见算法及原理描述，优缺点，还有推导。笔试的话遇到过 2 次逻辑回归的推导（svm 没有遇到）。比重 15%。

其实上面的这些问题，在课程里面都有，只是需要多刷几遍，因为只看一遍的话很多东西会没接收到，也记忆不深刻。

最深感受就是：一定要勇于去尝试。我之前一直不敢去外面面试算法，然后去了大约 3 次以后，该遇到的问题也遇到过了，回来总结和补充之后，后面的面试也有底气了，整个人给面试官的感受是会不一样的，成功几率就会大很多。最后小 tips：先面试小的公司，存存经验，觉得自信上来差不多了，再去面试心仪的公司。

最后的最后，祝大家能找到心仪的工作！

双非渣本跨专业转行面经：从 Linux 运维到推荐算法 40 万

我是集训营 3 期的学员，2016 年本科毕业于一所排名 500 开外的野鸡二本学校，学的电气类专业，毕业之后来帝都做了一年半的 Linux 运维，感觉做运维没有啥前途，只能勉强混口饭吃，所以想转到一个有发展前景的方向，然后在网上找各种 IT 培训机构，发现七月在线还不错，于是 2017 年末开始脱产来七月在线的集训营学习，这个课程是晚上 8 点到 10 点线上直播上课，课后会布置作业并且有老师和助教批改，我由于之前的计算机基础很差，所以需要白天将录播视频多看几遍才能勉强理解其中的概念和原理，然后做练习题进行巩固和查漏补缺。

脱产学习了 3 个多月课程结束后。我开始写简历投了几家小公司，去面了两家之后发现自己太菜了，因为我基础概念不扎实，缺乏项目经验，没有学过数据结构与算法，也没有刷过 LeetCode，感觉这个状态是不可能找到工作的，所以又自己看视频、刷题学了 2 个月，然后才又开始投简历找工作的，刚开始面试的时候很多问题回答的不太好，经过不断地面试和总结，最后在 2018 年 5 月份，终于收到了两家 AI 创业公司的 offer，做的都是推荐系统，因为连续找了一个多月工作很心累，想着先入坑再说，后来没有继续面其他公司了，最后去了待遇是 18k * 14 的那家。由于 18 年年底经济环境不好，公司融资出现问题，所以我只工作了半年多就跳槽到另一家有现金流的稍大型的公司，因为在上家公司积累了一些比较硬核的技术，所以此次跳槽薪资也涨到了年薪 40w。

感谢七月在线让我成功从前途渺茫的运维转型到了新兴的 AI 领域，让我实现了一次人生的蜕变，给我的事业发展带来了更加光明的未来。说到面试，简历自己一定要好好写，把项目内容写得详实一些，简历中有的，一定要保证自己都会，没有的不要写上去，尽量不要鼓吹自己会什么。此外，还有 leetcode 刷题的课程，熟练的完成编程题，在面试中的印

象分会很高，关于项目，首先要先了解一下你将要面试的公司，业务是否贴近，这样找工作的岗位，更有针对性，提高匹配度也能提高面试成功率。

以下是我面试中遇到的一些问题和自己的总结，希望对同学们以后的面试有所帮助：

数据结构算法

- 1.用最熟悉的语言实现二分查找、冒泡排序、快速排序
- 2.找出一定范围内的所有质数
- 3.找出数组中的第 K 大元素
- 4.求两个数组的交集
- 5.求最长公共子串
- 6.如何计算一个给定二叉树的叶子节点数目

机器学习基本概念与算法

- 1.请简单介绍下逻辑回归（LR）
- 2.写一下 sigmoid 激活函数的公式
- 3.逻辑回归与线性回归有什么区别
- 2.逻辑回归为什么要对特征进行离散化？为什么要对特征进行标准化？
- 3.哪些算法不需要对特征进行标准化处理？
- 4.请你说说支持向量机（SVM）？SVM 与 LR 之间有什么样的区别和联系？
- 5.说说你知道的 SVM 核函数有哪些
- 6.有遇到过拟合的情况吗？过拟合是怎样解决的？
- 7.说一说 bagging 和 boosting
- 8.L1 正则和 L2 正则的区别
- 9.特征选择有哪些方法

10.随机森林与 GBDT 有什么区别和联系

11.KNN 和 Kmeans 有什么区别

12.说说 xgboost 与 lightgbm, 他们有什么优缺点

深度学习

1. 什么是梯度消失? 什么是梯度爆炸?
2. 如何避免梯度消失与梯度爆炸
3. 详细推导一下 CNN 模型的工作原理
4. 什么是正向传播? 什么是反向传播
5. 你有哪些调参经验? 调过哪些超参数

NLP

- 1.什么是 TF-IDF? 写一下他的公式
- 2.什么是 word embedding
3. 推导一下 word2vec
4. 什么是 fine-tuning?
5. 了解 BERT 吗?

推荐系统

1. 说一说基于用户和基于物品的协同过滤?
2. 推导一下 FM 模型和 DeepFM 模型?
3. 推荐系统里有哪些常用的评估指标?
4. 用过 Hadoop 和 spark 吗? 用来做什么处理

可以发现, 面试题基本都中规中矩, 在课程中基本都可以学到。主要是面试时, 是否能够回答到点子上, 理解上有没有深度, 自己的理解、思考和体会。课后也要多做项目, 先照

着源码敲，多看几遍源码，然后尝试不看源码自己写出代码，一定要多练多总结，才能有提升。

推荐的书籍，我觉得书不在多而在精，建议把《统计学习方法》，《百面机器学习》和《推荐系统实践》这几本书多读几遍，读透彻了就行。

在七月认识了很多老师和朋友，学习过程中一起交流，找工作过程中相互鼓励，找到工作后依然经常联系，相互交流技术和人生，很感谢这段经历，感谢七月在线这个平台。

集训营四期面经

我是集训营四期的学员，2018 年开始在七月这边学习，之前自己陆陆续续写了一年，看了网课，感觉网课中的理论的有些难理解，而且实际做项目中，也不知道高深的理论该怎么食用，集训营的老师讲的十分通俗，学起来也比之前轻松了许多。

平时跟课程，需要克服自己惰性，不过最后还是检查下来了，也十分感谢努力的自己，除了七月的集训营，也重点看了算法的小课，老师手撕 DP 的场景，还历历在目。

说道面试，简历自己一定要好好写，把项目内容写得详实一些，简历中有的，已经要证明自己都会，没有的不要写上去，尽量不要鼓吹自己会什么，因为你基本上不会比面试官厉害，集训营的课程在内容上也有难度，授课老师不错。此外，还有 leetcode 刷题的课程，熟练的完成编程题，在面试中的印象分会很高，前前后后也算面了一两个月，面的公司大小都有，关于项目，其实首先要先了解一下你将要面试的公司，业务是否贴近，这样找工作的岗位，更有针对性，提高匹配度也能提高面试成功率。每个人的项目差距很大，这里主要总结性的列举一下面试题目。

初创公司

- 1、auc 的计算？有什么物理含义？
- 2、手写算法题：DFS，BFS
- 3、LR 公式推导，求梯度
- 4、0-1 背包问题
- 5、图像类数据，有什么扩充的方法？
- 6、说一说 xgboost 和 gbdn 的差别？
- 7、CNN，RNN，LSTM
- 8、准确率，召回率，F1-score 的含义。

- 9、深度学习调参经验
- 10、L1, L2, Dropout
- 11、SGD 和牛顿法
- 12、bagging 和 boosting 的区别是什么?

中型互联网公司

- 1、单源求最短路径。
- 2、动态规划, DFS 搜索
- 3、TF-IDF 的计算方式, 用 MapReduce 实现
- 4、朴素贝叶斯, 怎么平滑?
- 5、讲一下 gbd+lr ?
- 6、如果解决过拟合现象?
- 7、讲一下 SVM, 说说核函数的功能和设计

大型互联网公司

- 1、gbd+ 公式推导
- 2、xgboost 公式推导
- 3、FM 公式推导
- 4、算法题: 全排列
- 5、word2vec, 层次 softmax 和 skip-gram
- 6、MapReduce 手写 word count
- 7、kmeans 算法为什么一定会收敛, 讲一下 EM 算法, kmeans 算法有什么缺点?
- 8、gbd+ 和随机森林有什么区别?
- 9、算法题: 编辑距离

互金公司

- 1、L1 和 L2
- 2、使用蒙特卡罗方法计算圆周率
- 3、编程题 topk 问题
- 4、霍夫曼编码 (word2vec 中的层次 softmax)
- 5、编程题：八皇后问题
- 6、spark 用过么？和 hadoop 有什么区别？
- 7、讲一讲 MapReduce 的过程
- 8、xgboost 原理，为什么使用二阶？
- 9、SVM 公式推导
- 10、讲一讲交叉验证

大型互联网公司

- 1、了解哪些激活函数，怎么用？
- 2、了解 GAN 么？
- 3、特征方程的几何意义？
- 4、将一下 pca
- 5、讲一下 MapReduce 的过程，用 word count 来举例
- 6、编程题：括号匹配
- 7、主题模型了解么？
- 8、什么是极大似然估计
- 9、推导一下 LR
- 10、word2vec 有哪些优化方法？

11、GMM 和 KMeans 有什么异同，为什么一定会收敛？

12、推导一下 SVM 的损失函数，解释一下核函数的设计思想

可以发现，面试题基本都中规中矩，在课程中基本都可以学到。主要是面试时，是否能够回答到点子上，理解上有没有深度，自己的理解、思考和体会。课后也要多写项目，多练多总结，才能有提升。

推荐的书籍，我觉得有很多，比较有代表性的就是《统计学习方法》，《百面机器学习》，《数学之美》和《推荐系统实践》这些书。

在七月认识了很多老师和朋友，这一路走过来少不了相互的鼓励，最后同学们通过不断地努力，也有了不错的工作，很感谢这段经历。

推荐算法面试分享：如何从大数据起步拿到年薪近 40 万

本人是集训营第七期的学员，之前从事大数据相关，目前拿到了一家 D 轮推荐算法岗位，面试过程中有些心得可以和大家分享

1, 先从头说起吧，在开营之前已经把数据结构和送的 python 基础课刷了一遍。数据结构一定要循序渐进，对于常考的数据结构一定要自己实现过，如果面试官对做的项目认可，那么数据结构题也做得出来，offer 成功的几率会变大。

2, 开课期间，看了老师的课程，老师讲的很深入，上课的知识一定要多看几遍，可以结合帖子和课程一起看，作业一定要做，基础知识一定要了解透彻，面试的时候也会要求写公式，但是不会太难。

3.项目，我选的是咱们七月的电影推荐，推荐的算法用的不是很难，召回部分有的是隐语义模型（ALS），还有基于电影内容的词袋模型，面试中经常会问到怎么优化相似度矩阵，怎么怎么消除长尾，因为之前从事大数据，如果大家要做电影推荐，要把项目的数据流说清楚。除此之外，我还自学了一些其他召回方法。比如 personalrank, contentbased, wordtovect, 排序部分，常用模型，lr, gbdt, gbdt+lr, wideanddeep, 自己一定都要实现一遍，对原理一定要非常清楚。

4, 面试，一共面了 5 家公司，拿到 offer 就去了，因为每个公司的要求不同，对岗位面试难度也有所不同，对于推荐我认为一定要具备两方面的能力，一个是算法能力，一个是工程能力，算法能力就是常用的召回排序，工程能力就是大数据，面试官也会根据需求问一些场景题，不一定答的很准确但是一定要有自己的思路，手写数据结构必问。

面试过程中尽量把自己擅长的东西说出来，就是尽量占据主动，如果一直被面试官问，很容易答不上来，项目一定要说的详细，不能有一点说的不明白的地方，这样很容易被面试

官怀疑，你是不是自己亲手做过，还有就是对于项目成果尽量用生产指标去支撑，这样更有说服力。

最后数据结构一定要天天刷不能断，对于自己的硬实力是很有帮助的，强悍的编码能力会让你在面试中如鱼得水。

算法工程师（推荐方向）面试经历：简历、数据结构、机器学习、推荐系统

简历

我是推荐就业班的，找工作第一步就是投简历了，好的简历能提高不少面试机会。之前就因为没有一份好的简历，投递之后迟迟没有面试通知。

简历中最重要的就是和招聘 JD 中相关的项目经历。我在简历中主要提了在项目中使用了哪些算法，线下线上的指标影响。对于写到简历中提到的算法要能够很好掌握，面试官一般会根据简历项目中的算法进行提问。

数据结构和算法

作为一个算法工程师，需要具备良好的编程能力和逻辑思维能力。面试过程考察一个候选人的编程能力和逻辑思维一般通过数据结构和算法题来考察。我面试的公司中技术面基本都会问到此类问题，或是让手写代码，或是线上编程，又或是口头说一下解题思路。

我在面试中遇到的问题一般有：链表反转，有序列表合并，基本排序算法（快排，归并排序，堆排），二分查找等。这类问题多刷刷剑指 offer 和 leetcode 就好。

机器学习相关

虽然现在已经有公司将深度学习应用于推荐领域了，但是我在面试过程中发现大部分公司问的更多是基础的机器学习的问题（可能和我项目里没用到深度学习有关）。因为是做推荐系统的，所以协同过滤基本都是必问的，item-base、user-base、矩阵分解的原理，推导，优缺点都会被问到。经典的机器学习模型 LR、SVM 和 GBDT 的推导也有被问到，一般从损失函数开始一步步展开。正则化 L1、L2 的作用和特点，正负样本如何划分，模型调参一般调哪些参数，用的哪个包实现的还是自己实现的。模型输入的特征维度和数据规模有的面试官会问到。由于项目中处理过文本内容，所以面试官也问了 word2vec 和文本词袋模

型相关的问题。

推荐系统相关

推荐项目一般会问有哪些评价指标。推荐系统一般有模型线下评价指标如召回率，准确率，覆盖率，AUC 等等，线上看系统 AB 测试的点击率表现等。推荐系统的整个数据流程是如何处理的。协同过滤如何计算得到 N 个最近邻问题。数据比较稀疏的样本是如何处理的。为什么要有召回和排序的流程。如何用用户画像进行召回，数据的存储及性能如何解决。

总结

作为算法工程师基础编程能力要有，总不能连代码都不会写。虽然很多中小企业用的推荐系统还是经典的协同过滤那一套，但是也需要表达出你对推荐系统更进一步的理解，对现在前沿的推荐技术的了解。推荐系统是比较倾向业务的系统，需要知道从客户端获取数据到最终给用户推荐内容的整个流程是如何进行的。

第六部分 金融风控

经济学编程小白转数据分析

个人经历：

本人是本科应用数学，研究生经济学，没有太多的编程基础。一次偶然的机会在朋友圈里看到有人转发了七月的公开课，才与七月在线结缘。当时刚刚毕业非常迷茫，不知道自己应该做什么，在学校学习了太多理论知识，却不知道怎么应用起来。求职面试相当不顺利，当时 python 大火，于是就报名了七月 python 的课程，听课学习。对于没有基础的小白，有些地方还是有点吃力的，不过贵在坚持，有了 python 基础后面求职也算顺利一些。凭借会一些 SQL，再加上 python 基础，在一家中小型互联网公司入了行。

工作半年多的时间，对互联网行业也慢慢有了了解，工作上面也渐渐得心应手，18 年 11 月的时候正好看到集 7 开始报名了，纠结了很久我决定挑战自己一把。集训营的课程由浅入深，每周都有作业，答案解析。不得不说，纸上得来终觉浅绝知此事要躬行，因为作业的关系，很多以前模棱两可的概念，不熟悉的地方都得到了巩固，新的知识也快速地上手了。学完集七之后，看看同学中的大神，深知自己水平还不够，于是又潜心看书学习了一段时间，19 年的 8 月开始，陆续投了一些简历，一半是是关于数据挖掘，模型算法，另一半是数据分析师。最终选择一家互金公司风控组数据分析岗位，薪资提升 1.5 倍

下面分享一些常被问到的问题：

模型类：

- 1、 逻辑回归推导
- 2、 逻辑回归优缺点
- 3、 xgboost 的节点是如何分裂的

4、 xgboost 和 GBDT 有和不同

5、熟悉哪些聚类算法，解释下 K-means 算法

Python 基础类：

1、 pandas 如何对进行时间转化

2、 对 pandas 里面 groupby 进行考察

SQL 类：

1、 如何不用窗口函数输出每个人的订单的排序，比如 user1 的第一笔订单输出 1，第二笔输出 2

2、 用 Sql 怎么解析 json

3、 用 sql 求出每个用户平均下单间隔

比赛经历：

因为今年九月参加了一个 kaggle 比赛，在面试中经常会被问到一些这方面的问题

1、 如何进行数据清洗

2、 特征工程师如何做的

3、 为什么使用了 lighgbm，有何优势

4、 尝试用过其他线性模型么

分析思路类：

之前参加头条的面试被刷下来了，头条的面试很灵活，完全靠平时的储备

1、 你手机里现在最常用的 app 是什么（我答的知乎）

2、 在知乎的场景下，如果你发现用户浏览的平均时长减少了，你会怎么分析原因

3、 如果今天知乎想要上一个新的业务，你会从哪个方面分析该不该上这个业务

逻辑思维题：

有些题还是有些难度的，但我觉得当时做不出来也不是特别大的问题，面试官主要看有没有思路吧，平时也可以在网上搜一搜，很多题还是很有意思的。

1、现在有 5 个赛道，25 匹马，每个赛道一轮比赛只能跑一批马，请问最少需要几轮能够角逐出前三名。

2、在一个魔法岛上有一只羊和 100 只老虎，无限的草。羊和老虎都可以吃草生活，，羊只能吃草，但是如果老虎吃了羊，老虎就会变成羊。假设每只老虎都是理性的（我的理解这句话得意思是即老虎都想吃羊，但是也都不想被吃，这是一道博弈论的题）

集四 45 万 offer 的 AI 面经：SVM 和 XGBoost 这类是面试常考算法

1、再次恭喜集 4 毕业不到半个月，成功拿到 45 万年薪 offer（薪资相较培训之前的上一份工作翻了一倍多），请简要介绍下你的教育经历、工作经历

邱同学：我本科毕业于一个普通的一本学校，随后考上某 985 学校的研究生并在 2016 年毕业，专业方向是通信工程。毕业后的两年时间里一直在武汉的某一家做大数据的公司内任职，涉及到部分的 java 开发和数据分析工作，在这个过程中开始接触到 python 和机器学习的内容，也为后面参加集训营打下一定的基础。

2 看你在集 4 每周一次的在线考试中，成绩基本都是 100 分，与大家分享下你的学习经验？

邱同学：集训过程中的成绩这部分，第一是来开课之前提前把咱们集训营送的 python 入门视频、机器学习第 8 期以及深度学习第 3 期的视频刷了两遍，打下一个还不错的基础；第二是开课后认真消化和吸收老师上课教授的内容，及时整理和及时复习；第三是认真对待老师安排的课后练习，做作业的过程也是一个稳步提升和复习的过程；当然最重要的一点是要足够努力，这三个月的时间我几乎没有休息过几天，想拿到好的 offer 意味着需要付出比别人更多的时间和精力。

3 三场面试场场拿下 offer，被称为 offer 收割机的你，有何经验/面经跟大家分享(特别是面试最常考哪些题)，以及如何更好的准备 AI 面试？

在面试的过程中主要是针对简历上的项目来提问的，所以简历上写的项目一定要把每个细节和知识点搞清楚；另外机器学习一些常用的算法和原理要做到烂熟于心，面试官其实主要都是针对基础的知识点进行考察，如果连 LR、过拟合和欠拟合这样的基本概念都搞不清楚的话会很容易被面试官 challenge，此外，对于 SVM 和 XGBoost 这种常考的算法我几

乎前前后后手推了超过 10 次，到后面几乎都能闭着眼睛在脑海里推导出来，而在面试的过程中也确实被问到了相关的问题，考察对于算法的理解。

小编注：除了集训营课程，下面这两篇笔记写得通俗详尽，大家也可以看下

支持向量机通俗导论（理解 SVM 的三层境界），[博客版](#)、[公号版](#)

通俗理解 kaggle 比赛大杀器 xgboost，[博客版](#)，[公号版](#)

当然，线上直播的过程中有时确实不能把每个知识点都讲得非常仔细，但是老师会把所有的框架和知识点都串讲一遍，所以更多地需要自己在课下进行巩固（用好校长 July 所说的四大金刚：课程 题库 LeetCode kaggle/天池），这也是机器学习集训营和其他 Java 之类的培训之间的主要区别，很多机器学习的算法需要自己反复去理解和体会，算法思想和原理远远不是把一些常识性的东西复制性地背下来就可以了，而是要有自己的思考和理解，只有基础比较牢靠，才能在后面面试的时候游刃有余。

最后，一两次面试没有成功其实说明不了什么，反而自己应该在面试的过程中积累一些经验，为了后面的面试打下基础。

4 你觉得在集训营得到最大的收获和心得 分别是什么

邱同学：因为我是离职后来线下参加集训营，最大的收获是结识了优秀的寒老师和 David 老师，以及一起学习和努力的小伙伴，在这三个多月的时间大家都是互相帮助，有问题大家一起讨论，学习氛围非常好。学习上感谢两位老师的付出和指点，生活上也感谢宁哥的照顾和七月在线提供的学习资源，正是这一切才使得我能顺利从集训营毕业并拿到不错的 offer。

5 对于从传统 IT 转行 AI，你有什么经验跟大家分享，比如半年的时间应该怎么准备转行

邱同学：在我面试的过程中，可以感觉到目前市场上对于机器学习和深度学习工程师的需求还是很大的，所以对于想转行进入机器学习这个领域的同学来说，这是一个契机但同时

也伴随着挑战，需要自己对于市场和自己的职业提前做好规划。如果有半年时间准备的话，我的建议一是需要明确自己想从事的方向和领域，二是提前做好准备并做好预习，三是跟着老师方向并付出足够的努力，按照这样的学习路线我相信半年时间足够能有一个稳步提升的表现。

6 未来三年有何职业规划

邱同学：对于我个人而言，首先是需要花一段时间去了解自己所在岗位的真实业务场景，熟悉项目中数据和模型运行的整个流程，期望在 1-2 年能对所在部门和方向有一个较为全面的认知；在这个基础上再去尝试一些新的方向和领域，扩展自己的人脉和认识面，以获得更多的机会和发展空间。

金融风控 100 道面试题：传统银行开发转行互金 top3 公司并年薪 40 多万

我是机器学习集训营第四期的学员，我 2016 年从北京某 211 学校计算机科学与技术/生物技术双学位本科毕业，在国企中做了两年传统的银行系统开发，主要是 c 语言开发贷款系统，在做本科毕业设计的时候就用过计算机神经网络，对这方面一直都在关注着，一直想进入人工智能领域，苦于一直没有找到合适的学习机构和学习方法。

后来我同事给我分享了七月在线的人工智能课程，我看了一下课程表，里面是我想学的内容，经过考虑，我和同事一起报名了集训营 4 期的课程，这个课程是晚上 8 点到 10 点授课，有批改作业等环节，刚好适合我们这种平时白天上班晚上回来学习的上班族。

边上班边学习了两个多月课程结束后，我开始投简历，先后收到了百度，阿里，搜狐，滴滴，京东等大厂和很多小公司的面试邀请，由于是刚开始面试，准备的不是很充分，也是想试一试水，因为缺乏项目经验，且数据结构方面的算法知识我没有复习，刚开始面试的时候很多问题回答的不是很好，经过不断地面试总结和再面试，后面回答得就比较顺畅了。

最后 2018.9 月底我拿到了我比较满意 A 公司的 offer，做的是金融风控，待遇是 22k*15+ 期权，还有京东的 offer，京东的 offer 是做推荐的，因为我个人比较喜欢金融，所以选择去了 A 公司，在 A 公司工作了快 1 年，在 A 公司期间学到了很多，做了不少的项目，金融风控方面的项目经验丰富了很多，积累了很多技能和业务知识，我现在跳槽到了 B 公司，B 公司是比 A 公司做得更早和更大的 Top3 互联网金融公司，平台更好，工资也涨到了 28*15。

感谢七月在线让我成功从传统的国企转到了现在头部的互联网金融公司，成功转到了心仪的 ai 金融领域，给我的事业发展带来了更广阔的天空，实现了自我突破。

以下是我面试中被问到的和自己总结的，我觉得面试时比较重要的知识点，希望对新学员有所帮助（小编注：绝大部分面试题的答案都在题库里哦：

<https://www.julyedu.com/question>)：

- 1、用 python 写出 AUC 的计算过程
- 1、常见的激活函数有哪些，各自的区别和使用场景
- 2、梯度弥散和梯度爆炸
- 3、Relu 在 0 值处是不可导的，谈谈对这个问题的看法
- 4、在你自己的场景里如何解决样本不平衡问题以及为什么要这样做
- 5、你所参与的项目里实时性如何保证
- 6、项目的背景 数据量 数据如何切分 建模的流程以及这么做的考量点是什么
- 7、金融的话，主要是 LR+树模型，当然模型问的东西一般都会根据简历上的项目来问
- 8、数据结构面试问题，我知道的常问题；10 大排序算法，中后序遍历 有序数组合并 打印倒三角
- 9、Xgboost 的推导，这个是在说项目的时候涉及到了这些，所以就让我推了一些，再说了一些细节的东西，比如正则化的作用，和 GBDT 的区别。
- 10、排列组合算法
- 11、自己对接的业务现在用什么方法去做
- 12、如何把一堆正数和负数分开，回答完后会再问如果有 0 怎么办
- 13、了解什么机器学习算法，说一下 LR 的损失函数，聚类的方法，K 值怎么选取
- 14、怎么计算 TF-IDF
- 15、怎么进行类别划分（根据什么，怎么训练）
- 16、分类与聚类的区别

- 17、特征怎么选择（特征工程还需要学习）怎么知道选出来的是最好的
- 18、SVM 核函数的选择、核函数的主要作用是什么，SVM 与 LR 的异同点（从样本集、目标函数、优化方向）
- 19、常用的性能评估指标与计算方式，它们之间的比较
ROC、AUC、精确率，错误率，召回率
- 20、Bagging 与 boosting 的区别，（RF 与 xgboost 的区别） bagging 与 boosting 的偏差与方差
- 21、怎么根据数据集选择模型，不同模型之间的差异
- 22、简历中写的东西要了解的面面俱到，尤其细节
- 23、SVM 为什么使用拉格朗日乘子法？软间隔怎么做？
- 24、关于 word2vec 上下文词怎么表示中心词，输入是什么，上下文词权重，中心词怎么得到、怎么把词向量放到模型中训练，（LR 的输入）分词之后拿到重点词之后怎么处理
- 25、项目中的特征组合是怎么组合的，为什么，有何指导意义，组合形成了什么，对脱敏数据怎么进行组合
- 26、项目中把项目名称“某”词去掉，省得面试官觉得是假项目
- 27、什么是负采样，原理是什么？正负样本不平衡时该怎么采样，采样出的结果是什么（比例）
- 28、模型细节，怎么调参
- 29、连续值特征离散化有什么好处？
- 30、朴素贝叶斯的输入是什么，怎么进行分类，如何进行计算的
- 31、有一个几百 GB 级数据文件，数据中每行都是一个词，如何统计输出 Top10（集群与不用集群）

32、Xgboost/RF1 GBDT 是怎么选择出特征重要度的（内部怎么计算的）

33、L1 和 L2 的区别，各自的优点

34、评价指标：AUC、ROC、精确率、召回率

35、混淆矩阵、降维

36、分类模型：

38、LR 逻辑回归

39、LR 的损失函数是什么，为什么不用 L2loss

40、逻辑回归和最大似然函数的关系

41、LR 样本均衡怎么处理

42、 如果对 LR 样本采样，怎么在预估时还原真实的概率 P

43、L1 和 L2 的区别

44、DT 决策树

1) 决策树建模过程（停止生长和生长方式）

2) ID3、C4.5、CART 区别/联系

3) DT 缺失值，处理方式

4) 回归树构建方式

5) DT 怎么控制过拟合

45、NB 朴素贝叶斯

1) 贝叶斯公式

2) 有什么假设条件

3) 遇到计数为 0 的项怎么计算概率 p

46、SVM

- 1) SVM 原理
- 2) 求解 SVM 时为什么把原问题转换为对偶问题
- 3) 为什么引入核函数
- 4) 有什么 kernel,怎么选
- 5) 什么函数可以作为 kernel function
- 6) SVM 对缺省数据敏感吗
- 7) rbf、kernel 为什么能把原始维度映射到无穷多维

47、集成模型

- 1) RF/GBDT/Adabost/Xgboost
- 2) xgboost 和 GBDT 的区别
- 3) xgboost 为了控制过拟合做了什么优化
- 4) xgboost 的并行化是怎么做的
- 5) xgboost 树生长的精确分裂与近似分裂分别怎么做的

48、为什么要进行特征离散化

49、KS, AUC 详解, 会画图解释推导

50、时间复杂度, 空间复杂度、边界考虑、代码风格

51、算法要从头到尾讲清楚

52、从面试中积累问题、反思、回顾、积累答不上来的问题

53、hadoop/Spark 加分项, 不会大量花时间在这上面; 数据流程会问, 数据处理能力

54、卷积神经网络的卷积层为什么加 padding

55、gru 和 lstm 的区别

56、lstm 的结构和原始 rnn 区别为啥能缓解梯度弥散

- 57、数据不均衡怎么办
- 58、损失函数、优缺点
- 59、为什么 dropout 能缓解过拟合
- 60、神经网络需要调哪些超参数、参数有啥作用，对结果有啥影响
- 61、为什么用神经网络做文本分类
- 62、有哪些激活函数，求导、特点
- 63、什么时候会出现梯度弥散、梯度爆炸、之前的网络怎么处理这些问题的
- 64、卷积神经网络的结构是什么样的网络
- 65、典型的神经网络有哪些、每一个网络的出现有啥优化。
- 66、介绍一下在做项目的时候怎么判断网络过拟合了
- 67、手推 LR、GBDT、xgboost 和自己项目里用到的算法
- 68、随机森林和 xgboost 的区别
- 69、把算法里面的思想和原理讲清楚，项目里面用到的算法要全方位弄清楚
- 70、Xgboost 为什么要用二阶泰勒展开
- 71、Xgboost 什么时候是残差
- 72、为什么 xgboost 不用 lr
- 73、xgboost 怎么处理缺失值？
- 74、xgboost 中哪些参数可以控制过拟合？
- Xgboost 调参优化的一般方法
- GBDT vs Xgboost vs LightGBM
- 75、最大似然估计原理
- 76、逻辑回归与线性回归的联系、异同？

77、为什么逻辑回归比线性回归好

78、LR 和 SVM 的关系

79、你们公司线上模型的 ks,逾期率，线上有多少模型

80、工作中用的数据源及其底层的逻辑

91、你觉得做模型最重要的是什么？即一个好的模型算法工程师所必须的技能有哪些？

普通 IT3 年转行经验谈：从曾质疑培训机构到培训后年薪 35 万

2016 年从英国某大学硕士毕业之后，面对茫茫的就业大军，拿到了两家航天系国企的 java 开发岗 offer，其实在决定参加集训营之前，已经关注七月有大概一年的时间了，那时候人工智能机器学习的相关内容已经挺火的了，但是知乎上也有一些质疑和否定培训机构的声音，导致自己迟迟没有下定决心去系统的学习。再加上当时有看到同事桌上的西瓜书，随手翻了翻感觉好难，那么多公式那么厚，心里上的为难情绪也给自己打了退堂鼓。就这样过了一年，一个同事也和我谈论起个人发展和七月在线，我们都觉得在职业上和技术上已经很难取得进步，再加上枯燥的工作和大量的个人时间（5 点下班），让我们有了开始系统性学习人工智能的决心。

就这样在相互的鼓励下，我们都报了七月的集训营，大家都知道，老师可能是一个机构最核心的资源，传道受业解惑尤为重要。令我惊讶的是，集训营从开始的入门，pandas 基础，可视化，数据挖掘，到机器学习相关内容的讲解，再到后面的项目设计，都是经过精心设计的。尤其是机器学习的内容，该重点看哪一块，核心知识点是什么，哪些内容容易混淆，跟着老师过一遍，你就会发现，其实西瓜书看起来也很快。

这里其实我想说，每个人的基础和学习能力是不一样的，面对一个新的知识，新的领域，可能仅仅是有兴趣是不够的，自学成才自然是好的，但是还是会有不少同学可能像我一样，翻着书看着那些复杂的公式就会有为难情绪，很容易陷入闭门造车的弯路，从而也磨灭了自己难能可贵的兴趣。第三就是有了心中的那团火，有了老师，自己的付出是必不可少的，上面我也说了，由于国企的工作性质，我和我的同事有着大量的时间来学习，有时候我下班不走，留在办公室等着视频课，课上的内容量是巨大的，2 小时满满的干货，需要你在当天和最晚第二天消化掉，因为周五还有考试，周末还有量更大的线下课，就是这种反复的强化催促和压力，才能让你不断的进步，这也往往是培训比自学好的地方。

在老师的点拨下，如何梳理成自己的知识体系，就显得尤为的重要了，幸好有一群志同道合的朋友，大家相互分享，气氛很棒，我自己会把学过的内容用思维导图整理出来，每一节每一章都梳理出来，老师也会在整个学习期间分阶段地给大家整理很多资料和面试题，有了知识点，知识结构和面试题，后面就真的到了实战的环节。面试我觉得就是熟能生巧，戒骄戒躁，每次面试后都要去总结，哪些答得不好，再回到知识点，思维导图去找，查漏补缺。同时七月也会找来历届已经工作的优秀学员来分享自己的面试经验，这些经历过一线企业真实面试的资源，才是最宝贵的

找工作那会赶上年底，我自己大概面了 6-7 家左右，最终拿到 3 家公司的 offer，一个是知名外企，23k*14，一个是某大型互金公司，25*14，最后一个是股份制银行，18k。

某贷网、某蛋科技、百度金融、蚂蚁金服等互联网金融面试经验

我报名参加了七月在线的机器学习集训营，完成了 3 个多月的课程后，便向这个神秘而又充满诱惑力的职业发起挑战，下面就介绍一下面试过程中积累下来的经验，与君共享。

首先，简历是你能被邀请去一个公司进行面试的敲门砖，一般技术岗的简历要突出特色与重点，忌讳繁杂与花哨，一般可以将简历分为基本信息、个人技能、工作经历、项目经验四个大的方面，这四项也是面试官最感兴趣的，然后再对大项进行详细的划分，例如：个人技能和项目经验这两个肯定就能再分出来很多小项，小项在写的时候也要有完整的结构和严密的逻辑。一方面是让面试官在你的技能与经验迅速找到兴趣点，另一方面是让自己对自己有一个清晰的认识，这样面试的时候讲起来才能更有条理。

其次，就说一说我面试的部分公司和面试中经常遇到的问题。问到的问题我会清晰的列出来，问题的回答方式和我的答案也会写在下面，括号内的可能只说出相关知识点，部分详细的答案还需要各位看官自行去百度。

1.公司：某蛋科技，岗位：数据挖掘工程师，行业方向：互联网金融

一面技术：

(1) 介绍一下自己的逾期还款率项目

主要讲这个项目在什么场景下（P2P 借贷平台）应用什么技术手段（数据处理，特征工程，LR、RF、GBDT 训练，做 Ensemble,调参）解决什么问题（对客户是否会逾期还款的预测，最后 AUC 达到了一个点：一般说 70%-80%，太低没多大用，太高有点假或者有过拟合的可能性，有明显的好坏用户的分辨能力）

(2) 主要做了什么工作

这块肯定要挑自己擅长的说，因为接下来面试官会根据你说的擅长的部分进一步进行提问与了解，因为面试的是数据挖掘的岗位，所以我主要讲的是特征工程（数据处理，特征选择：

Filter、Wrapper、Embedded、特征衍生)，捎带着说了一点模型训练的事情

(3) 常用数据从哪里来，大数据处理的方法和技术用过哪些

其实对于一个互金公司的数据来源无非是两个方向，一个是自己业务中积累下来的，还有就是从外界买来的，但是这个问题如果只回答是自己公司的，有些面试官又会觉得不可思议，因为自己公司的数据不可能是全面的，而互金行业对数据量要求又大，所以不要入了面试官的坑。大数据处理方法就回答 spark 和 map-reduce，着重还是说 map-reduce。

(4) 有一个 1T 大小的每行只有一个词汇的文本，怎么统计词的个数

这个问题其实就是考察大数据处理中 map-reduce 的原理了，原理很简单就是‘分——合’的思想，就是对 HDFS 上的资源进行分片，map 阶段将数据分成 key-value 对，然后在 reduce 阶段再对 key 对应的 value 进行计数。这样就统计出了词的个数。

结果：技术面顺利通过，二面 CRO，问我对互金的看法和风险化解等等业务上的问题，由于是第一场面试，对互金方面不算太了解，所以直接折在了二面。

2. 公司：灵犀联云，岗位：算法工程师，行业方向：数据服务

一面算法：

(1) 编写二分查找算法

思想就是从中间查找，比较大小后再找左或右半边，这个百度一下，各个语言版本都有，这个问题需要能够手写出来的，回去背吧

(2) 二分查找的时间复杂度 ($O(\log n)$ ，最坏 $O(n)$)，时间复杂度，空间复杂度怎么计算的
其实这个就是看你代码执行的时间和占用空间，简单方法就是把常见的算法的时间复杂度和空间复杂度记下来，如果实在记不住就百度一下计算的过程

(3) 简单问了一下工作经历和所作的工作

(这跟第一个面试问的一样，然后就从头到尾的介绍一遍)

(4) 在工作中用到了哪些算法，简单介绍

这个肯定挑自己掌握度高的和简历中写的介绍，例如 LR, RF,SVM, 然后说原理和推导过程。

二面技术：

(1) 分类和聚类的区别

从样本数据、模型、方法：分类(LR, DT...), 聚类(K-Means,K-Means++), 、产生的结果等方面阐述区别

(2) 特征是怎么进行选择的或者说怎么知道选出来的就是最好的

这个问题上面已经提过三种方法：Filter、Wrapper、Embedded，主要从划分的属性上会像决策树生成过程中计算信息增益、增益率、Gini 一样评价一个属性的评分

(3) L1 正则化和 L2 正则化的区别

这个问题主要从公式，正则化形成过程，功能等方面进行回答，附一个我觉得讲的比较好的正则化的博客

https://blog.csdn.net/yuyang_1 ... 85483

(4) 手推 SVM

常用的机器学习算法大家最好还是可以手推出来的，因为有的面试官真的让你手推啊，手推之后回答了一下软间隔与硬间隔，核函数有哪些，区别，拉格朗日乘子法（西瓜书 SVM 那一章安排的明明白白）

三面人事：

这个人事就很牛的样子，完全不是公平性谈判，打压你的信心，说什么他们那都是清华北大的毕业的，我不是名校出身，处于技术初级算法工程师阶段，希不希望有一个好的平台什么

的，反正就是各种忽悠，然后压榨工资。这种情况也慢慢跟他聊，拿到 offer 是关键，只不过把它当备胎。

结果：最后谈到了 16k/月*14 个月，第二天发 offer。

3. 公司：某贷网，岗位：算法工程师，行业方向：互联网金融

一面技术：

(1)介绍自己的项目和主要工作

基本上是面试时的第一项，后面这一项就不写了

(2)bagging 和 boosting 的区别

这个具体化其实就是 RF 和 GBDT 的区别：a.样本选择方式；b.样例权重；c.预测函数；d.并行计算

(3)语料库怎么构建词向量

这个问题就是骗自然语言处理方向的问题了，因为这是我在第一次写简历的一个败笔，写了一个自然语言处理的项目。但是也是硬着头皮说用到了 word2vect 进行构建词向量的...

(4)TF-IDF 的计算方法，含义，应用场景

含义：词频-逆向文件频率，是一种统计方法，用以评估一字词对于一个文件集或一个语料库中的其中一份文件的重要程度

计算方式：TF 词在文件中出现频率；IDF：总文件数目除以包含该词语之文件的数目，

应用场景：其实就是自然语言处理中看一个词的重要性，然后对于文本分类这样的项目能够起到作用。

(5)word2vect 的原理，两种模型形式，霍夫曼树？

word2vect 其实就是一个输入层，隐藏层和输出层这样的三层神经网络结构，一般分为

CBOW 和 Skip-Gram 两种形式，霍夫曼树是在训练 CBOW 和 Skip-Gram 优化的数据结构。上面这三个问题具体还包含很多小点，由于我不是主要面深度学习的，所以就不加以赘述了，大家可以上网查一下，附一个博客链接，一目了然：

https://blog.csdn.net/yuyang_1 ... 85421

二面直接上老大，问我有没有自然语言处理的实际落地项目，他们现在急招的是自然语言处理的岗位，算法工程师不急。听到这很忧伤，自然语言处理我没做重点复习啊，只能硬着说做过一个短文本分类小项目，然后做下介绍，闲扯了 10 分钟，感觉已经没戏了，所以有想面自然语言处理方向的同学一定要好好复习。

4. 公司：永辉云创，岗位：机器学习算法工程师，行业方向：机器学习

一面技术：

(1)写一个快排，说一下快排主要快在哪？

这个可以百度一下，原理，计算方法，代码应用尽有，主要快在它的思想是分而治之的思想，其实东西都是这个思想，例如分布式、hadoop 等

推荐一个关于快排的博客：

<https://blog.csdn.net/pythonda ... 84385>

(2)手推 SVM，什么是核函数，为什么要用核函数，SVM 核函数怎么选择？

手推公式我这里就不写了，认准切入点是我们要找到一条直线正好把正负样本分割在两端，而且让它们离直线越远越好，之后进行了一系列的计算。

核函数是为了解决样本不是线性可分的问题提出来的，主要的核函数有线性核，高斯核，Sigmoid 核，核函数选择主要从训练集样本量，特征量上进行考虑。总之 SVM 的东西还是很多的，推导过程中还有拉格朗日乘子法、SMO 算法、软间隔、硬间隔等问题。

(3) SVM 和 LR 的异同点?

这个我整理了一篇博客, 直接上链接:

[https://blog.csdn.net/yuyang_1 ... 86204](https://blog.csdn.net/yuyang_1...86204)

(4)连续值特征做离散化有什么好处?

同样这个问题博客见:

[https://blog.csdn.net/yuyang_1 ... 86284](https://blog.csdn.net/yuyang_1...86284)

(5)特征工程中对时序型特征怎么处理

时序特征其实是在预测上常用的一个特征, 简单的来说我们可以根据特征中的时间特性来划分不同的时间区段(比如, 现在马上到双十一了, 在时间特征上我们可以剥离出双 11, 双 12, 周末, 工作日等), 这样会让特征有时序型(其实就是特征工程中对时间类别特征的处理), 另外还有成熟的时序特诊处理方法(这个需要自行去搜一下 paper), paper 较长, 这里就不赘述了。

(6)推荐系统的算法了解多少? 简单介绍一下协同过滤, 怎么计算商品之间的相似度?

其实我了解的真不多, 所以我就简单说知道协同过滤(基于物品和基于用户), 其实就是一个人以类聚, 物以群分的过程, 然后说一下距离计算的方法: 欧氏距离、Jaccard 相似度、Pearson 相似度、余弦相似度, 最终形成一个打分的矩阵, 然后根据得分进行推荐。

(7)讲一下隐语义模型

思想是通过隐含特联系用户兴趣和物品, 找出潜在的主题和分类, 基于用户的行为对 item 进行自动聚类划分到不同类别/主题。知道这不是自己的强项, 就跟面试官说只是了解, 没有实际用过, 但是我会回去好好研究研究, 衣服信誓旦旦的样子。

(8)说说 hadoop 的原理

这又是大数据了, 上面已经提到过了, 大数据这块的东西大家还是好好看一下, 因为现在搞

算法，搞机器学习、深度学习都是在大数据的背景下的，所以这以后是一个基本的技能，面试官问的也比较多。

(9)简历中的销售量预测的项目的评价指标召回率和 ACC 怎么协调

这个问题就看实际业务场景是更偏重于哪一方面了，我的回答(不一定对)我们是先以准确率为准进行模型调优，达到我们可接受的数值时，然后往上调召回率，最终使两个指标都稳定。

二面人事：人事没问什么特殊的问题，基本都是一个讨论，然后问一下工资，告知准备入职手续和材料。

结果：永辉云创是永辉超市注资成立的，初创阶段，业务方向也是依托于用户超市数据搞大数据处理和探索人工智能其它方面。

顺利拿到 offer，20K/月

5. 公司：百度金融，岗位：算法工程师，行业方向：互联网金融

(1)简单介绍自己，主要负责哪部分工作？

这个就不说了，大家对这自己的简历都练几遍。尽量多说一点，这样面试官就少问一点。

(2)项目中用到的数据的量有多少？特征有多少维，都用作训练吗？

这个问题是很多面试官都喜欢问的，要根据自己的项目实际情况进行阐述，数据量较大的情况一般是不能全部用作训练的，要不然那模型得跑多长时间啊，还有就是特征维度，一般几十到几千维都有，用作训练的其实也就几十维到百维这样，特征工程那块肯定要看一下哪些特征是有用的，进行特征选择，如果全用会给训练带来大量的工作反而起到副作用。

(3)对大数据了解多少？hadoop 的原理？

大数据处理其实上面已经提到过，用到的工具有 spark、hadoop，然后 hadoop 的原理就

是分而治之，map-reduce 都干什么，建议写两个这样的程序你就能明白它是怎么工作的，这个很主要，因为工作中就是这么用的。

(4)特征怎么进行组合的，对模型训练有什么指导意义？

特征组合一般都是拿差异性、关联性较大的特征进行组合，这样不会损失特征的作用，还可以拓展出未知的特征，例如将 longitude 与 latitude 组合，产生的组合特征则代表一个明确的城市街区。这样能产生比单独考虑单个特征更强烈的位置信息的信号，所以特征组合就是在衍生我们不能直接接触的特征，然后对模型训练有积极影响。

(5)在模型训练过程中都用了什么算法？

LR,RF,GBDT..., 然后——解释每个算法的特色、原理、中间再加点比较（因为都是分类算法）

(6)混淆矩阵，ROC、AUC、召回率、KS、Gini、IV

混淆矩阵：

		预测		
		1	0	合计
实际	1	True Positive (TP)	False Negative (FN)	Actual Positive(TP+FN)
	0	False Positive (FP)	True Negative(TN)	Actual Negative(FP+TN)
合计		Predicted Positive(TP+FP)	Predicted Negative(FN+TN)	TP+FP+FN+TN

这个是后面计算 ROC,AUC,召回率，准确率的基础，所以还是要把这个表格记住：

TP：正确肯定——实际是正例，识别为正例

FN：错误否定（漏报）——实际是正例，却识别成了负例

FP：错误肯定（误报）——实际是负例，却识别成了正例

TN：正确否定——实际是负例，识别为负例

AccuracyRate(准确率): $(TP+TN)/(TP+TN+FN+FP)$

ErrorRate(误分率): $(FN+FP)/(TP+TN+FN+FP)$

Recall(召回率, 查全率, 击中概率): $TP/(TP+FN)$, 在所有 GroundTruth 为正样本中有多少被识别为正样本了;

Precision(查准率): $TP/(TP+FP)$, 在所有识别成正样本中有多少是真正的正样本;

TPR(TruePositive Rate): $TP/(TP+FN)$, 实际就是 Recall

FAR(FalseAcceptance Rate)或 FPR(False Positive Rate): $FP/(FP+TN)$, 错误接收率, 误报率, 在所有负样本中有多少被识别为正样本了;

FRR(FalseRejection Rate): $FN/(TP+FN)$, 错误拒绝率, 拒真率, 在所有 GroundTruth 为正样本中有多少被识别为负样本了, 它等于 $1-\text{Recall}$

ROC 曲线就是以假阳性概率 (False positive rate) 为横轴, 真阳性 (True positive rate) 为纵轴所组成的坐标图

上面大家可能学习过程中都能接触到, 但是 KS、Gini、IV 是互金行业风控模型中常用的指标, 一被问到我就懵逼了, 完全不知道是啥啊, 只能支支吾吾说没用过这些指标。

这几个指标内容也比较多, 在这附一个讲的比较好的博客地址:

[https://blog.csdn.net/htbeker/ ... 97557](https://blog.csdn.net/htbeker/... 97557)

结果: 可能大公司要求就是严格, 亦或是一个问题就能看出来你在这个行业待了多长时间, 所以: 一面, 卒。

6. 公司: 某金融集团, 岗位: 算法工程师, 行业方向: 互联网金融

一面技术:

(1) 简单介绍一下自己

千篇一律, 当练口才了。

(2)手写快速排序算法，讲原理，最好、最坏时间复杂度，空间复杂度

这个上面也提到了，主要看对算法的理解程度，所以面试中可以不去看那些数据结构，但是常用的算法一定要能默写出来，而且知道为什么快。

(3)简历中写的申请评分卡是怎么做的？

基于 Logistic 回归的申请评分卡模型开发（主要讲了流程）：

①数据准备：收集并整合在库客户的数据，定义目标变量，排除特定样本。

②探索性数据分析：评估每个变量的值分布情况，处理异常值和缺失值。

③数据预处理：变量筛选，变量分箱，WOE 转换、样本抽样。

④模型开发：逻辑回归拟合模型。

⑤模型评估：常见几种评估方法，ROC、KS 等。

⑥生成评分卡

(4)做一个风控模型一般流程是什么样的

其实这个问题就是把上面的问题又说了一遍，但是我回来在网上查了一下大概是这个样子的：业务定义(确定义务才能确定其他的)->风险定义(其实就是定义好用户和坏用户，或者说是正负样本)->风险化解(这块就是用算法训练各种评分卡)->风险策略(针对作弊行为或坏账进行的后期行为)

(5)手推 LR，并对每一步进行解释：引入 sigmoid，逻辑回归的公式，极大似然，求解参数时的梯度下降算法

这个没什么可讲的，就是去手推吧，推几遍这些都记住了，包括哪一步该干什么。

(6)LR 和 SVM 有什么异同点，怎么理解 SVM 的核函数

异同点和核函数都在上面说过了，这块回顾一下吧(其实整个这个面经看下来可以看出来，面试官们常问的问题也就这些，只是会从它们的业务中带出问题来)。

(7)项目中样本量是多少？正负样本比例如何？怎么解决正负样本不均衡的问题？

样本量自己说了几十万，正负样本 10:1(肯定有人问，为什么说 10:1，因为在互金行业的正负样本比是十分不平衡的，每有一个负样本就说明至少出现一笔坏账，在几百万数据的情况下，有 1W 坏账这个公司就别活了，当然我说这个比例也不是很准确的，但是我解释取样本的时候做了采样)。样本不均衡解决方法：负例采样，smote 算法。然后讲了 smote 算法怎么做。

(8)特征工程做了哪些工作?为什么？

这个当时就把自己实际用的特征工程方法说了：对 null 值、异常值文本型、时间型、类别型，组合，统计缺失率等通通说了一遍，然后进行特征选择。这么做当然是为了模型能快速训练出效果，使数据收敛，提升评估指标等等啊，这么做肯定说产生好的影响。

(9) 项目中使用模型训练的评估指标是什么？AUC 多少？KS 值多少，KS 值是怎么计算的吃了百度金融的亏，这回我可答上来了，上面那个连接里面的每个概念、计算方法我都扣了好几遍。AUC：0.73(不能太高也不能太低，有零有整，很真实)，KS：0.28（一般>0.2 便说明模型有很好的区分能力）

这个问题其实也是一个不小的坑，因为你说的太假一下子就能看出来，面试官都是在实际工作中摸爬滚打出来的，你说 AUC0.9 那他得仰视你。

二面技术+业务：

(1)在工作中都会用哪些特征来进行模型的训练，哪些数据比较有效

这个主要是跟用户个人和信用方面息息相关的：例如个人基本信息(年龄、地址、学历、工作...)、设备属性(手机品牌啊，运营商...)、业务行为(借贷行为，还款，消费，旅游....)，有效的肯定是跟业务关联多的。因为从这些特征中可以直观的判断这个人的借款需求和还款能力。

(2)对风控模型的理解，完成一个风控系统都需要做哪些工作

①尽可能全面搜集客户信息

②根据业务进行模型训练，然后对用户进行评分，完成授信

③后期监控，防止产生坏账和作弊行为

(3)给一个场景:有 50W 贷中数据和 50W 贷后数据，特征维度有 300 维，其中还包含了空值，怎么用这些数据做一个申请评分卡？

讲思想：用贷后数据进行数据分析，做数据处理（空值、异常值）、特征工程，然后计算特征的 IV 值和特征选择方法筛选特征，再把处理完的特征放到 LR 中进行训练(训练集，验证集按照 7:3 划分)，对模型调优，达到一定指标。贷中数据也别闲着，放到模型里进行预测，放到库中，待时间线临近来验证模型实际应用中的效果。吧啦吧啦，说的口干舌燥，

(4)之前做的项目风险是怎么定义的？

这个不同公司有不同的定义，银行可能定义 M1 期以后是坏用户，而 P2P 比较脆弱，承担风险能力弱，所以 M0 期可能就定义为坏用户。

(5)有哪些手段防止用户的反欺诈行为？

这个问题回答的时候尽量从业务的场景去分析，第一，尽可能全面的收集用户的资料，用户资料越全后期用户的行为越可控；第二，授信前要有严格的授信策略(例如申请评分、贷中行为评分、贷后行为评分)；第三，后对贷后用户进行监控和跟踪，采取电话回访或者其他方式确定用户真实存在。第四，就是不断完善自己的风控系统。

(6)给我找了一批数据，然后发给我，让我回去用这些数据训练一个预测好坏用户的模型。

这个就是实际动手操作了，并不难，回来对数据进行简单的处理，做点特征工程，然后用 LR 和 RF 分别训练了一个模型，然后做一下比较，用 AUC 和 ACC 评估指标说一下效果等等。

三面人事：

人事这关就是问你为什么辞职啊，上家公司待遇啊，对自己的定位啊，对整个行业的了解啊什么的，跟人事聊天就是让她感觉你很积极向上，离职原因不要贬低自己原来的公司，可以说自己想拓展一些技能啊，进行更多的挑战啊什么的。人事问还有什么问题的时候也尽量问一些关于公司的文化啊，人才培养啊，进阶过程什么的，不要上来就问工资，这个问题等人事主动问你比较好。

结果：拿到 offer，顺利入职

总结：BB 这么多，都是我一家公司一家公司面过来的，还有几家公司跟我想从事的方向不一致，问题也都跟这些差不多就没多赘述，问题都是真实记录的，所以有要找互联网金融方面工作的朋友可以参考参考，有些问题我说的不对的也希望指正(改不改看我心情)，我也学习学习。

一位从传统 IT 跳槽到互联网做数据分析师的面经

本人因工作中建模需要参加七月在线的集 4，同时购买了两个 NLP 课程，经过前后四五个月收获很多。我之前是在一家传统快消公司负责电商相关数据分析工作，在经过集四培训后正式提离职。经过一个个多月面试，最后入职一家互联网公司任数据分析师岗。下面分享这一段时间面试经历。

中间面了二十多家公司，由于之前工作偏运营分析，因此此次投递的是数据分析，面试内容包括运营分析+基础算法内容，以下整合百姓网、哈罗单车、京东商城、饿了么、拍拍贷、GIO、蚂蚁金服、今日头条等公司的面试遇到的相关问题：

- 1、 之前的项目经历（有 1-2 个能逻辑清晰、有结论的项目描述）
- 2、 比如网站内容频道留存率下降，可以从哪些维度分析？（整理一套条理清晰、完整成熟的分析架构）
- 3、 常规 SQL 代码：行列转置、时间函数、聚合函数、左右拼接、子查询、数据更新、笔试代码（思考如何设置筛选条件及拼接条件）等等

备注：我常用的是 SQL Server 在一家一直被质疑代码在 hiveSQL 是否可以运行，以及运行效率问题？如果对 HIVESQL 熟悉的可以大致比较，不熟的话也不要紧，多家公司不要求。

- 4、 解释下 Logistic 回归，和线性回归的差异？sigmoid 函数在概率为 0.5 的时候自变量正好为 0，那如果业务场景需要想要改成概率为 0.6 的时候自变量为 0，应该怎么做？
- 5、 对哪些算法比较熟，解释下随机森林，他的随机性主要体现在哪些方面？对 Xgboost 或 XBDT 是否了解，他们之间的差异是什么？
- 6、 熟悉哪些聚类算法，解释下 K-means 算法？什么时候或条件下停止迭代？
- 7、 什么叫特征工程？有哪些评估模型好坏的指标？
- 8、 除了上面问题，经常会被说，公司希望建模的同学会用 java，这话估计是对我不满意

才说的吧。

9、 常会被问到的还有：目前薪资、期望薪资？以及有什么 问题想问他的？

备注：对于期望薪资，开始面试时候拿不喜欢的公司试水的时候，可以大概评估试试，如果到了喜欢的公司认真面试的时候，薪资建议面议，一般到 HR 那都能聊到还可以的。

面了很多公司，基本问题都是那些，而且除了前工作项目外，很多培训都有讲过。所以主要认真准备项目内容和基础算法：①内容熟悉、②描述逻辑清晰、③至少要准备 1-2 块可深挖的板块（前期先准备，面了一两家再优化），还得谢谢张老师中间给了很多帮助，祝大家都能找到有发展又多金的工作！

第七部分 数据挖掘

非科班从传统行业到 JAVA、爬虫到转型数据挖掘

我可能是一个不太安分的人吧，但到目前为止，换过很多的岗位，从刚毕业在四线城市呆在传统行业两年，到自学 4 个月 JAVA 到一线城市找了个 JAVA 开发的工作，内部转岗做 JAVA 爬虫，再自学 python 跳槽做 python 爬虫，报七月的集训营后内部转岗做了数据挖掘，到如今跳槽到一个更大的平台做数据挖掘工作。

我求职过的岗位有 JAVA 工程师，爬虫工程师，数据挖掘工程师，推荐系统工程师，都是技术岗，没事换工作面试也比较少，因为之前换工作都是在自己准备比较充分的情况下才去换工作，每次找工作面 4-5 家就差不多了。

1、岗位：JAVA 工程师 最终薪资：9.5K

毕业的时候在四线城市做了两年传统行业的工作，日复一日机械式地画图，感觉生活枯燥无味，于是就有了加入码农一族的想法。

由于当时工作比较清闲，每天下班后花 4 个小时左右的时间自学了 JAVA，自学几个月后就裸辞跑到一线城市找工作，面了 4 家拿了 3offer (两个外包)，然后入职了其中一家，9.5k。由于老员工离职，入职没多久就转岗做爬虫了。

在这个阶段，对自己的定位就是入行，所以当时也没有太考虑到公司的性质，待遇相关的，但是奉劝大家尽量不要去外包，因为待遇的确很差，很久才会有一次调整，而且对未来就业也不好，很多公司对外包出来的有歧视。

2、岗位：python 爬虫工程师 最终薪资：16K

这次入职的一个公司，是一个互联网公司，在对应的行业上处于一个独角兽的地位，当然这独角兽是 HR 自己说的，但入职了你会发现其实不过而而，好在在里面遇到一个比较好的领导，由于个人也比较喜欢钻研，最终转岗做了数据挖掘。

说真的，在这家公司对公司很失望，但对自己所在的团队还是很心存感谢的，给了很多学习的机会，也学到了很多。不过人员配置比较少，当时只有我一个人做数据挖掘的工作，在上线了两个机器学习相关的项目之后，业务这方面的需求也越来越多，要求越来越高，感觉力不从心后就才有了跳槽的想法。

3、岗位：数据挖掘工程师 最终薪资：30w+（年）

之前有了跳槽的想法，本来的打算就再往后几个月的，但没多久就在网上认识了一个猎头，是猎头帮忙推送了一个大公司，打着破罐子破摔的想法就有了这一次面试，可能自己也比较幸运，面试官主要看的是项目落地能力，而自己之前都是项目包办的，从立项到上线部署一条龙服务，于是比较顺利地拿到了 offer。

总之，这次面试很成功，我面试前本就不抱太大的期望，所以除了项目用到的技术，别的也没有写太多（个人技能部分没写），因为不是很了解的东西写太多会暴露更多的问题。很多同学可能面试了很多，可能不是你不够好，而是没遇到合适你的团队，有些团队看重项目落地能力，而有些看重基本功，多面一些，总会遇到合适你的那一个。

总结一下我对目前已经涉及过领域的理解，目前的时代，相对于传统 IT 来看，机器学习算是非常年轻的行业，目前的市场还是供远远小于求的，当然，这两年对从业者的要求也提高了很多，因此需要不仅要打好基础，同时也更要注重项目能力，面试官都是从你的项目来深入提问你所使用的技术，只有真正参与了项目，你才能注意到一些需要关注的技术要点。

最后说一句，对于刚毕业的同学来说，尽可能找一些大厂，有大企业背景对你未来跳槽很有用，大厂之间跳槽比小厂到大厂容易得多，这是我面试中从 HR 和猎头处得到的看法。

普普通通的转行之路

自 2018 年 6 月底毕业之后，依据专业和所学，校招进入了一家国企制造业单位，做了半年的工程师。宿舍、公司、食堂三点一线构成了我的日常，其实工作不外乎如此，或者说生活本质就是这样子的，工作，睡觉，吃饭。其实车间、粉尘、化学用品，防尘服什么的自身并不排斥。这份工作待遇也还行(仅相对于制造业而言)，工作也很稳定，自己不作的话（不过我还是作出来了，哈哈，愧对前公司的培养）。

然而，当了解了整体现状和发展趋势后，内心总是有一个声音对自己说，现在走的路我已经能够看到二三十年后的自己是什么样子了，如果这样，有必要继续下去吗？稳定是最大的谎言，这是外部环境给予的，而不是自身能力得到的。结合自身喜好与当前大环境，为自己的未来选定了机器学习方向。

自 2018 年底确定转行目标之后，开始在网络上收集相关资料信息，那时的我一些简单的专有名词都不清楚，hello word 也仅在 2011 年上学期 c 语言程序设计写过。这样经历一个多月后，到了 2019 年 1 月，发现自己这样何时才能实现自身目标，闭门造车的前提是你得知道车子大致结构吧，而且专业人士的指导，可以节省自己很多的精力和时间。

一月份裸辞后就考虑报班了，收集了一些市面上的培训班信息，和周围同学朋友，培训班的招生老师聊过之后，果断选择了七月在线，成为了机器学习集训营第七期的一名学员。选择七月的原因在于，

- 一、是它是唯一一个建议在职学习的培训班（当然自己是破釜沉舟，裸辞了）；
- 二、是其他培训班的课程包含所有（前端、后端），并不如七月的专精，七月的课程紧紧围绕着算法展开；
- 三、是一月份开班的就它一个了，其他的差不多要开年后三月份了。

七月的课程做的真的很良心，身边很多同期学员都是从其他互联网技术岗或者在校生，校招是一次很好的机会。然而作为一个基础薄弱的人来说，学习起来吃力是一方面，经验背景也是小白一只，再加上 2019 年的就业环境。

坦白说，我这不算是啥面经啦，只是简单和大家聊一下这段时间的历程，因为自己现在所转的岗位只是简单的互联网数据分析师岗，和向往的算法岗还是有很大差距的。其实，一个大的目标很多时候并不能一步到位，但是拆分为一个个小目标，按自己的规划前行就好。

2019 年的目标在于转换行业背景（目前任职公司），积累理论知识和项目经验（七月在线课程）和编程技能（多敲多学），目前一切都在规划中稳步进行。最后，祝愿大家都能心想事成、达成所愿。

面试分享，从开发转算法

我是机器学习集训营第四期的学员，双非院校硕士。

刚毕业时做的是 php 后端开发，面对是拿来就用的框架和繁琐的业务

转行的动机：

- 1.业务代码繁杂重复
2. 不想发代码（还是要打的）
3. 对未来人工智能发展吸引
4. 自身也有点兴趣
5. 目前公司学习机会减少，薪资提升难

所以想想不如去接触下这块，大学也有一定计算机（软件工程专业）和数学基础，所以跟朋友一起投身了七月，互相学习进步，刷课程视频，挤时间看直播，在过程中也认识很多同僚，尤其是寒老师，非常喜欢他的教学

虽然我学完后没有直接找工作，因为方向不太明确，当时整个教学内容偏大众一点，学习基础，实践项目也偏简单，很多内容需要自己涉猎和跟老师请教，包括 nlp 这块学习的内容并不多，当时记得讲过比较重要的有 lstm，word2vec 等，所以借助业余的时间再自己继续学习一下，想着往 nlp 方向发展，到今年才开始找工作，但是过程并不是非常顺利。

前期面试准备：

- 1.自我介绍
- 2.基础知识
- 3.面试项目
- 4.刷面经
- 5.刷 leetcode，剑指

自我介绍，就提前准备好，介绍下自己的学历背景专业，公司，技能，项目

基础知识，因为我投的是机器学习岗和 nlp，基础的 lr svm xgboost, nlp 方面 word2vec, tfidf, rnn lstm gru, textcnn, Bert, 基本概念要有, bert 看完基本 transformer, attention 就懂了，最好可以跑一下

面试项目，我是写了两个项目和两个竞赛，主要搞清楚整个项目建模流程，数据量多少，最后结果是什么，对比来说，自己提前练习下怎么说，尤其是数据量这块，看怎么吹真实点
起初开始投简历，效果并不好，简历内容比较少，简历的项目偏简单没什么含金量，面试官没什么聊的，后来面多了，也有点经验了，后来优化了下简历，丰富了项目内容

在深圳这边，投的机器学习岗机会比较多，会笔试让你撸算法题，leetcode 和剑指就够，让你写个快排。给你很大数列，要求在有限内存里排序。让你简单实现 bp 算法或者梯度下降，写个 tfidf，或者电脑撸码看你的代码风格等，也有一面聊聊项目就给 offer 的，有手推 lr, xgboost 边讲边写，svm 没遇过推的。

nlp 这块，面试遇到过，解释 textcnn，还有问到 bert，有笔试考你 pandas 操作的，词向量一定要会，word2vec 等，现在 bert 大行其道，一定要会，rnn 原理

最后选择了一个金融的数据挖掘岗，收入是有翻番，日常工作就是数据清洗，根据业务去建模提升业务产出，也接触一些客服机器人和机器学习平台，还在不断学习，祝大家成功

双非硕士面试之旅：数据分析、特征工程、模型构建、模型优化等流程要熟练

我是七月在线机器学习系训营第 4 期的学员，是一名 19 届双非院线的普通应届生，研究生是控制工程专业。作为小白的我，那时候根本不知道机器学习是什么。第一次接触机器学习是在研一的暑假，导师给的一个课题思路，让用机器学习的方法去解决，也因此让我接触到了机器学习。刚开始的学习是很痛苦的，一些算法模型的，看都看不懂，当时通过师兄的介绍花了一个暑假去学完了吴恩达老师的机器学习课程，只让我对机器学习有了简单的认识，也让我感受到了它的乐趣。

真正接触到七月在线是在研二的上学期，当时感受到师兄们找工作的压力，决心要好好学习机器学习，找一份相关的工作，师兄给我推荐了七月在线的课程，起初是在网盘上下载了七月在线视频，有理论有实战，被老师丰富的知识经验所吸引，所以就报名了机器学习集训营第四期，感谢七月在线带入门，那时候依稀记得在集训营群里大家讨论问题的场景，老师也及时的回复，解答同学们问题，在此表示感谢七月在线的老师的指导，不仅是学习的指导还有人生规划的指导。老师们的简历指导和面试经验也对我的面试求职起了巨大的帮助。

说道面试，简历按照老师们的讲解一定要好好写，把项目内容写得详实一些，简历中有的，已经要保证自己都会，没有的不要写上去。项目部分，就写了上课所讲的金融实战项目和 kaggle 比赛项目，一定要拿着代码一边看一边跑，了解比赛和项目的构造，数据分析，特征工程，模型的构建和模型优化等等。算法方面，要学会手推公式，像经典的 LR，SVM 和 XGBOOST 等算法一定要多多在纸上练习推导。

此外，还有 leetcode 刷题的课程，如果你想要去大公司，一定要熟练的完成 leetcode 刷题。

下面给大家分享一下我的一些面试过程。

CVTE 研究院数据挖掘岗

1. 自我介绍（研究方向）
2. 研究方向上目前的研究成果有哪些（介绍了自己的两篇论文和专利）
3. 参加的竞赛
4. 围绕 kaggle 分享预测竞赛开始发问：
 - 1) 自己在竞赛中的角色
 - 2) 比赛的难点
 - 3) 如何解决该难题
 - 4) 竞赛中上分最高的几个手段是哪些
 - 5) 特征工程如何做的，有哪些特征，为何使用这些特征
 - 6) 单模型最高与模型融合的结果差距多少？如何进行的融合
5. 阐述 SVM 原理，为何使用对偶性以及其对偶性原理，SVM 如何解决多分类，SVM 与 LR 对比。
6. 阐述 GBDT、xgboost、lightGBM 的异同，xgb 的优势，lgb 的优势，二者的并行如何实现。
7. 阐述 word2vec 的原理，为何使用哈弗曼树，如何构建哈夫曼，负采样的原理。
8. 对于一个取值较多的类别变量在不能进行 onehot 的情况下如何使用？
9. 阐述 Hadoop 与 Spark 的区别？
10. Spark 的工作原理是什么？

平安科技算法岗

1. 先是自我介绍，然后提了一些项目的问题，抠得比较细，抓到一个点就一直问下去。
2. 写公式：交叉熵、softmax、He 初始化

3. 写公式并说一下 KNN、 kmeans、 朴素贝叶斯的原理
4. 池化层如何反向传播
5. 上采样有哪些方式
6. 在分类任务中会用到哪些 loss function
7. 一个智力题
8. lc 的一道中等难度的算法题，讲算法，分析复杂度，最后代码实现

招银网络

1. 项目相关
2. 监督无监督区别，无监督实际场景？
3. 给定矩形内有若干点，求距离最小的两个点的距离，不能用穷举法？
4. 常见数据结构，栈队列区别和常见应用，完全二叉树和满二叉树。
5. 过拟合问题以及解决方法
6. GBDT 和 Random Forest 介绍
7. 激活函数，损失函数
8. 梯度消失、梯度爆炸原因和解决方法

这里只介绍了其中几个秋招面试的公司，可以发现，面试题都是比较基础的，课程中基本都有学到过。主要在于你面试时是否可以答到点子上，是否有自己的理解和思考，还有就是手推公式和 leetcode 刷题。推荐一些我用过的书籍，《剑指 offer》，西瓜书，《百面机器学习》等。

最后，希望我的一些经验对你们有所帮助，真心祝愿七月在线的学弟学妹们，希望你们在这里七学有所成，得到满意的 offer。谢谢！