
百度最新面经-算法工程师

一、

虚函数

霍夫曼树

malloc 和 new

bert 为什么设置那两种训练方式

过拟合解决方法

l_1 , l_2 正则为什么能降低过拟合

数组，可以分别从最左边最右边取个数字，求取得 k 个数的最大值， $O(1)$ 空间呢， k 的取值范围的条件（面试官所说的细节）

经理面

C++的基础 malloc 和 new 释放内存

vector 和 list

析构函数。const

稀疏词向量 用 skip-gram 还是 cbow 训练好

k 个的列表反转

两个 cpu，设计方法让处理时长最短，输入为数组

智力题 2 个蜡烛 1 个小时，如何记录 15 分钟

你平常怎么查 bug

二、自然语言处理

一面：

- 1.手写 softmax，手写 BN 公式。
- 2.word2vec，bert 详细介绍。
- 3.激活函数无限维的问题。
- 4.为什么使用交叉熵，不用平方差。
- 5.硬币兑换最小数。
- 6.100 层楼 2 个小球问题。
- 7.linux 命令,不会。
- 8.python 多线程，不会，进程内存互相访问，不会。。

二面：

-
- 1.手写 svm,
 - 2.手写 LR,
 - 3.手推前向传播。
 - 4.xgb 详细讲解。
 - 5.knn, k-mean。
 - 6.旋转数组：用额外内存和不用。
 - 7.判断是不是后续遍历中序二叉树。
 - 8.python 内存管理，内存池最大？
 - 9.python 可变不可变数据结构。
 - 10 python lambda 与 def 定义函数的区别

三面：

- 1.redis 和 mongodb 与 mysql。。真不会。
- 2.加快搜索速度方法。
- 3.计算相似度方法。
- 4.bert 微调。
- 5.研究生最大收获。
- 6.研究生与本科生区别。

三、凤巢广告算法

我是九月初投递的简历，大概九月中旬参加的面试（校招大规模，事先不知道部门），最后九月底收到了 offer。应该说我度的效率还是很高的。面试一共三轮，前两面偏技术，问得很广，很基础。三轮是经理面，有技术的问题，但相对更加开放，会更多谈到你对一些东西的看法和理解。

一面：

1. 上来先推了个 LR，问了 SVM 的基本原理，两个算法各自的使用场景以及它们之间的区别。
2. Boosting 和 Bagging 的概念和区别。
3. 常用的优化算法有哪些，举一个你熟悉的具体说一下，并分析它的优缺点。
4. 解释一下 AUC 的计算方法和它代表的意义。问了一个相关的问题，当时没有答的很好，就是一个数据如果分成两份，分别在两份数据上计算出 AUC 为 AUC_1 和 AUC_2，问整体数据的 AUC 是多少？面试的时候一直以为这个是能算出来的，所以一直在推导公式。最后的答案其实是无法得到，因为从局部的有序无法直接退出全局的有序，这个题目其实还是考查对于 AUC 这个指标的理解深度。
5. 介绍了一下实习的工作，没有细问

6. 代码题：用两个栈实现一个队列

二面：

1. GBDT 的原理，xgboost 与 GBDT 的区别；
2. 常用的模型评估指标有哪些，也问到了对 AUC 的理解；
3. 碰到数据不均衡的情况，如何解决；
4. 因为我简历里有 NLP 的内容，所以问到了很多 NLP 的相关知识：包括 Transformer 的结构，BERT 和 ELMo 的区别，word2vec 的原理；
5. 介绍实习经历；
6. 代码题： $O(n)$ 实现三色排序；

三面：

1. 聊实习经历，说说你在实习中最出色的工作，还有改进的余地吗？如何改进；
2. 说说你在研究生期间的工作，你碰到过最困难的事是什么，以及你是怎么解决的；
3. 是否了解 Hadoop？做一道大数据题目，topK 问题；
4. 说说你对凤巢的了解，如果你碰到了一个**场景，你会如何解决。
5. 职业规划相关...

整体来说，问得很细，有些地方可能由于时间久远有些疏漏。个人的一些经验是：

1. 基础知识复习复习好（包括机器学习、深度学习还有最重要的刷题），尽量提早做，说实话这些事我也是到春招末和秋招中才做得比较好，因此错失了很多机会；
2. 简历上的内容一定要很熟悉，这个不解释；
3. 准备一个漂亮的简历，所以春招找实习很重要！拿到一个大中厂的实习，你会发现事事都会顺利很多；
4. 找实习、找工作，宜早不宜迟，但有个问题是可能比较早的时候你还没有准备好，所以前一两个面试不要挑最想去的公司，打怪升级才是王道。

四、

一面

自我介绍

会什么机器学习算法，说了 lr、树模型

分类和回归有什么区别

讲讲 lr，公式，做多分类怎么办（说多个二分类或者 softmax）

讲讲 svm，推一下公式，没推出来，怎么处理非线性，说说核函数，具体怎么做，软间隔

异常检测项目，简单说一下，目前有什么方法？

机器学习和深度学习有什么区别？

正则化有什么，l1, l2, 分别有什么作用，怎么做到的，对应到概率是什么，拉普拉斯和高斯，具体说一下

文本分类，bert 了解吗？输入有什么改进

说一下 fasttext, 有什么好处

做个题目，爬楼梯（动态规划）

二面

简单介绍一下异常检测

图像数据 cnn 介绍、物理意义

生成方法为啥可以保留住信息

稀疏图像进行压缩，hash, 查找值多层 hash

怎么判断哪些位置是不是 0 问题，用二进制

多标签情况下，如何判断一个标签该不该留下，考虑绝对值和相对值

用假设检验方法

word2vec, 分层 softmax, 用的什么树，哈夫曼树，怎么做分层 softmax

五、

接到过百度三次面试邀请：

第一次面试前端（我从来没投递过前端的岗位）。当时是 HR 帮我内推的，可能觉得前端缺人，简历里面也涉及了移动端的项目，但是我真滴不懂前端就拒了。所以最好不要让 HR 帮你内推，或者你要明确告知投递的岗位。

第二次面试 C++ 研发工程师。因为简历里面涉及了 C++，有面试官捞起来问了，但是我估计我 C++ 达不到他们的要求所以也拒了（之前投递过其他公司的 C++ 开发岗，一面就挂了）。

第三次面试算法岗。我以为是 CV 岗，结果去了才知道是机器学习\NLP 岗。好吧，看来今年 CV 是真滴饱和，整个秋招的投递比大概是 CV < NLP < 机器学习。

如果之前知道面的是机器学习\NLP 岗，我可能就直接拒了，因为不懂 NLP。但是后来顺利拿到了 offer，期间也被问了 NLP 的知识，面试官说不懂也没关系，更看重思维。但是一起去的另一个朋友就没有我这么好的运气了，他也是做 CV 的，但是面试官问了一些 NLP、分布式的内容，没答上来就挂了。

所以，建议遇到机会大家多尝试，有时候看运气。

算法岗的面试一般有三轮，但是我面了四轮儿：

一面：持续一个半小时。讨论简历里面的项目，两道常见算法题，顺利搞定。

二面：一面过了的话会接着问二面，二面本来是 leader，但是当时 leader 有事儿就安排了其他面试官。讨论简历里面的项目，一道不常见算法题，一道数学题，顺利搞定。大概隔了一周。

三面：leader 亲自又面了一轮儿。不问算法，问的主要是机器学习，一些数学定义啥的。三面不太顺利。

四面：紧接着三面，是部门经理面试，主要是聊一些工作上的想法，对以后的期望，介绍部门情况之类的。

四面结束，现场给了口头 offer。