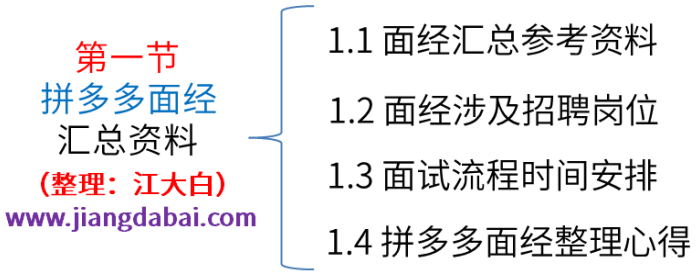
9|拼多多算法岗武功秘籍

1 拼多多面经汇总资料



1.1面经汇总参考资料

① 参考资料：

（1）牛客网：拼多多面经-87 篇，网页链接

（2）知乎面经：点击进入查看

（3）面试圈：点击进入查看

② 面经框架及参考答案：

（1）面经框架及参考答案：点击进入查看

（2）大厂目录及整理心得：点击进入查看

1.2 面经涉及招聘岗位

（1）全职岗位类

【NLP 工程师】、【图像视觉算法工程师】、【搜索广告算法工程师】、【数据分析工程师】、【核心搜索组搜索算法工程师】、【拼越计划算法工程师】

1.3 面试流程时间安排

拼多多面试流程-整理：江大白			
	面试类型	面试流程	备注（侧重点）
第一面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	根据项目深挖
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	根据项目深挖
第三面	HR面	基础人力问题	/

PS：以上流程为大白总结归纳所得，以供参考。

其他注意点：

● 要不是技术+技术+HR，要不就是技术+HR+技术

1.4 拼多多面试心得汇总

★ 两轮技术面试都是，先撕代码再聊项目，手写的相对较快，面试反馈比较正向，后面给了 SSP。

★ 面试的时候聊了很多实习期间的东西，基本上整体回顾了一下自己的经历，复杂网络、推荐系统等东西都聊了很多，比较重视深度和广度，细节也问的比较多。

★ 面试前的最佳复习资料还是自己的简历，简历上写的东西一定要会讲原理会推导。最后，不要在简历上写你不能熟练说出大部分细节点的知识！面试官基本还是看你简历提问多的。

★ 问项目的时候，一定要显示出自己对项目细节了如指掌，并且熟悉背后的原理，这样就算项目很水，如果能【显得自己基础牢靠】，面试官也会满意的。

★ 总体来说 ML 基础问题比较简单，算法题比较难（但不强求），主要问项目。

★ 拼多多感觉主要考察点就是从项目开始发散，所以项目里用到的东西一定要认真准备，当然没有项目可能连笔试机会也没，所以项目，论文，实习怎么也得有一个的。然

后编程题主要就是简单题为主，但是需要会简单题的优化算法。

★ 面试前就是几本常用的机器学习系列书籍，还有 svm，lr 这一类的推导，认真复习，自己想好对于自己 的项目可能会问到什么，提前准备好，还是很有用的。

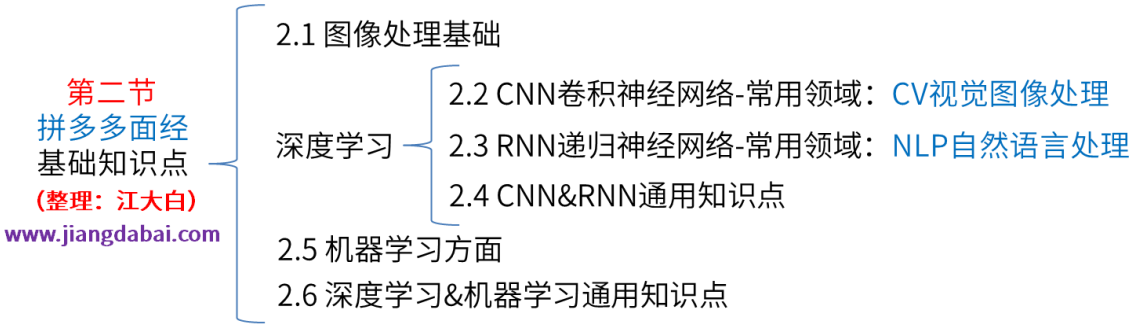
★ 总结：一面总体偏基础，于简历展开询问，要对简历内容熟稔于心，基础要牢固，同时也要有一定的深度广度。二面开始强调项目或者文章内容。

★ 感觉电商的场景还是比较单一的，主要是刀具等违禁品检测识别。

拼多多分工比较明确，这点应该是个优势，可以增加算法团队的专注度。

★ 整体而言题量不大，但是问的相当细致（尤其是 SVM 和 transformer 那块）。难度也不大，考察的知识点都集中在常规的模型和算法上面。

2 拼多多面经涉及基础知识点



2.1 图像处理基础

无

2.2 深度学习：CNN 卷积神经网络方面

2.2.1 讲解相关原理

2.2.1.1 卷积方面

- 卷积如何加速？
- 卷积和反卷积的原理说一下

- 讲一下 Dropout 层的原理 ？
- Dropout 的 train 和 test 的区别?
- Droupout 会改变数据分布，导致训练和测试样本分布不一致，怎么解决这个问题?

2.2.1.2 池化方面

- 池化和反池化的原理说一下?

2.2.1.3 网络结构方面

- Resnet 结构是怎么样的，有什么优点?
- Vgg 结构的优点是?
- inception 结构的优点?
- SeNet 的结构用过吗?
- Densenet 结构讲一下?

2.2.1.4 其他方面

- 数据增强的一些方法?
- 梯度消失和梯度爆炸是什么原因，怎么避免或解决?
- 讲一下 BN 层的原理?
- 什么是 roi pooling，怎么实现的?
- 什么是 roi align，怎么实现的?

2.2.2 数学计算

- 求普通卷积和 resnet block 的参数量、计算量，并对比两种结构?

2.2.3 公式推导

- 解释一下 CNN 原理，卷积后 feature map 尺寸大小怎么算，写一下公式?

2.2.4 手写算法代码

- 实现一个卷积操作，可以用 numpy?

2.3 深度学习：RNN 递归神经网络方面

2.3.1 讲解相关原理

- RNN 的改进有哪些？讲 LSTM 和 GRU，对于更长的序列怎么处理?
- 说说你理解的 LSTM（RNN 到 LSTM 讲 LSTM 的缺点，引出 transformer）?
- LSTM 如何调参?
- RNN 怎么解决，LSTM 为什么可以解决?

2.3.2 手绘网络原理

- 画出 LSTM 的结构图?

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

- 你认为 self attention 在提一些什么模式的特征呢?
- self-attention 和普通 seq2seq 的 attention 区别?

2.4.2 模型评价

- Recall 和 Precision 的原理讲一下?
- 用过哪些衡量分类性能的指标?
- 用了哪些评价指标，AUC 是什么?
- 写一个函数计算 AUC?

2.5 传统机器学习方面

2.5.1 讲解相关原理

2.5.1.1 数据准备

无

2.5.1.2 特征工程

① 特征降维

- 推一下 PCA 的公式?
- SVD 了解吗？主要的作用是什么？F 范数下的最优近似?

② 特征选择

- 特征如何选择，贡献性？共线性?
- 单特征选择有哪些方式？组合特征呢?

2.5.1.3 有监督学习-分类和回归方面

① 分类回归树（集成学习）

- GDBT 和随机森林的区别?
- 随机森林是 bagging 还是 boosting?

A.基于 bagging：随机森林

- 随机森林如何选择特征，如何评价特征重要性?
- RF 说一下，RF 树特别多会发生什么，RF 和 GBDT 哪个深一些，为什么?
- 随机森林的随机性在哪儿?

B.基于 boosting：Adaboost、GDBT、XGBoost

- XGBoost 和 GBDT 的区别，提升在哪里?
- 讲讲 XGBoost 原理？ XGBoost 优势？ XGboost 怎么解决过拟合？过拟合了，怎

么调参数?

- GBDT 关键参数有什么，树深和树的棵树都会导致过拟合，如果现在模型过拟合了，这两个参数调哪个?
- GBDT 是怎么选择特征的，答借助于 CART 树模型进行选择，类似于 ID3,C4,5 用信息增益和信息增益率? GBDT 还可以构建特征什么的?
- GBDT 树怎么生成的，残差从哪里来的，实际用的时候用的负梯度，绝对值损失函数的时候怎么求导?
- XGBoost 的损失函数是什么? 用过哪个损失函数?
- 项目中使用了 lightgbm，让讲一下 gbdt?

② K 近邻（KNN）

- KNN 是生成式还是判别式，时间复杂度，怎样优化?
- KNN 最后做回归，除了取平均还有什么其他方法?

③ 逻辑回归 LR

- 逻辑回归的介绍，如何训练，几种训练方法?
- LR 的定义，目标函数，优化方法?
- LR，随机森林，XGBoost 区别，细节?
- LR 和决策树的区别是什么?
- LR 损失函数：怎么用最大似然求损失函数?
- 说到过拟合，讲讲 LR 怎么应对过拟合?（L1, L2）为什么正则化可以?
- 逻辑回归如何分类出非线性超平面?

④ SVM（支持向量机）

- SVM 原理讲一下?
- 通俗介绍 SVM?
- 讲一下 SMO 算法?
- SVM 的损失函数（hinge）

- SVM 的推导-对偶--求解？核函数的选择？
- 为什么 SVM 这么有效？核函数能把特征映射到高维空间？
- 一般怎么选核函数？SVM 跟 AdaBoost 有什么联系？
- LR 和 SVM 介绍+区别，什么场景用 SVM 比较好？
- SVM 和神经网络的区别，两个分别适用于什么场景，SVM 有过拟合问题吗？
- 拉格朗日乘子，KKT 原理？

⑤ 朴素贝叶斯（Naive Bayes）

- 贝叶斯和传统统计学派区别吗?那从贝叶斯定理的角度说为啥 L2 可以解决过拟合?
- 知道共轭分布吗？

⑥ 决策树（DT）

- 决策树有几种（cart/id3/c4.5），他们分别有什么特点？
- 3 种决策树，区别，以及一个连续特征如何建树？
- 各种决策树模型的优劣（从最简单的 ID3 到最后的 LGB）？
- 树模型（决策树）复杂度？
- ID3 复杂度是多少，怎么优化呢？
- 决策树了解吗，GBDT 怎么做的，做分类问题用的什么树？

2.5.1.4 无监督学习-聚类方面

- 介绍一下 kmeans 的原理？
- 看简历上用过 GMM，讲解 GMM 原理？

2.5.2 手推算法及代码

- 推导 SVM？
- 推导 LR？
- GMM、EM 算法手推？
- 推导一下 GBDT 的公式？

- 手推 SVM+KKT 条件+SMO 算法简单描述?
- 推了一下 LR 的损失函数以及梯度传递

2.6 深度学习&机器学习面经通用知识点

2.6.1 损失函数方面

- 交叉熵的公式?
- MSE 的公式?
- 问逻辑回归算法的原理?
- 逻辑回归推导一下，损失函数+求解过程
- 为什么不用回归的 loss？均方误差?
- MSE 是不是凸函数，是 MSE 本身的问题吗?
- 逻辑回归表达式是不是凸函数?

2.6.2 激活函数方面

- 激活函数有哪些？Relu 和 Sigmoid 的区别?

2.6.3 网络优化梯度下降方面

- 项目中的超参是怎么调的，比如学习率或者优化方法，都用过哪些方法?
- adagrad 用过没？跟 SGD 有何区别?
- SGD 里面的“S”，什么意思?
- 训练里面用过多卡吧？里面对梯度有啥操作?
- 为什么要对多卡的梯度做平均?
- 为什么梯度下降可以求解最优化，数学理解?
- 深度学习里面优化器，ADAM 特点?
- 网络权重初始化方法、优化方法 Adam、多卡训练时超参数怎么变化（lr、迭代次数）

2.6.4 正则化方面

- 有哪些正则化、L1 和 L2、L1 梯度，在零点怎么办?
- L2, L1 原理(答：高斯先验， 拉普拉斯先验)
- L1 正则化损失函数如何求解?
- L2 正则化的特点，使用场景?
- L1 和 l2 正则化的区别，特点?

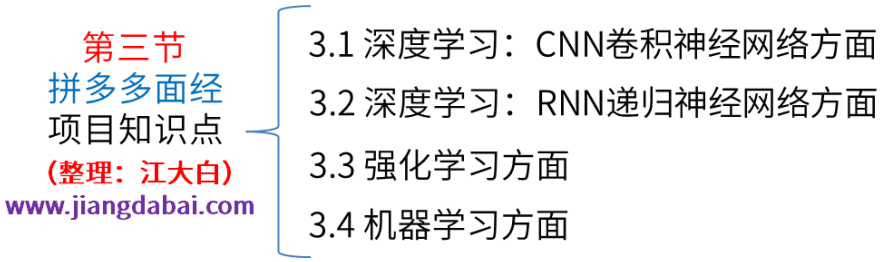
2.6.5 过拟合&欠拟合方面

- 深度学习中怎么应对过拟合?
- 如何解决过拟合？过拟合、欠拟合、如何调参?

2.6.6 其他方面

- 深度学习模型和机器学习模型的区别是什么 ？
- 数据不平衡怎么做?

3 拼多多面经涉及项目知识点



3.1 深度学习：CNN 卷积神经网络方面

3.1.1 目标检测方面

3.1.1.1 讲解原理

- 项目用到了 faster rcnn，然后就扯了 RCNN 系列，RCNN 和 CNN 的区别?

- rcnn 到 faster rcnn，介绍一下，然后具体每一步操作都问的很细，roi pooling 具体怎么做的、RPN 是怎么做的，输入输出是什么?
- Faster rcnn 的最后输出大小是多少、它的正负样本是怎么选择的?
- NMS 具体怎么做的，假设这是一个函数，那么这个函数输入输出是什么，中间操作又是怎么做的?
- 问我改个 6 点怎么从 Faster rcnn 改到 yolov3?

（1）backbone 要变

（2）ROI pooling 要删掉

（3）anchor 的设置要变

（4）正样本的设置方式要变

（5）regression loss 的实现要变

（6）yolov3 用了多尺度训练的方式

- 问目标检测目前的难点和未来的发展方向?
- Faster RCNN 中 RPN 的细节

3.1.1.2 手写代码

手写 SoftNMS 代码

手写 NMS 代码

3.2 深度学习：RNN 递归神经网络方面

3.2.3 自然语言处理 NLP

① Bert

- 说一下 bert 几个最关键的点?
- 问了下 bert 的输入以及和 gpt，elmo 的区别?

② Transformer

- 介绍 transformer 结构，手推计算讲解细节，画 Transform 结构图?
- 讲 transformer 如何并行化运算?
- 说一下 transformer 的优点?

③ CRF

- LSTM+CRF 层中 CRF 的作用?

④ HMM 隐马尔科夫模型

- HMM 和 CRF 的区别?

⑤ Word2vec

- Word2vec 原理?讲一下 word2vec 怎么实现?
- Word2vec 和 doc2vec 区别？和 fasttext 的区别?
- Word2vec 里有什么重要的方法，负采样和层次 softmax，分别是怎么实现的，为什么能提升速度和效率?
- 怎么用 word2vec 训练 embedding 的，怎么做的？召回分哪几路？具体候选集是多少?
- Word2vec 怎么训练的，为什么近义词能训练出距离比较小的向量?
- LDA 的词表示和 word2vec 的词表示有什么区别?

⑥ CNN 方面

- TEXTCNN 和 TEXTRNN 比较，fasttext 原理说明，为啥比其他效果差?

⑦ fasttext（词向量和文本分类）

- Fasttext 原理，为什么用 skipgram 不用 cbow，负采样怎么做到，公式是什么?

⑧ 其他

- EM 算法，为什么隐变量的问题要用 EM 算法?
- 怎么做文本分类，RNN 和 CNN 各自特点?

3.3 强化学习

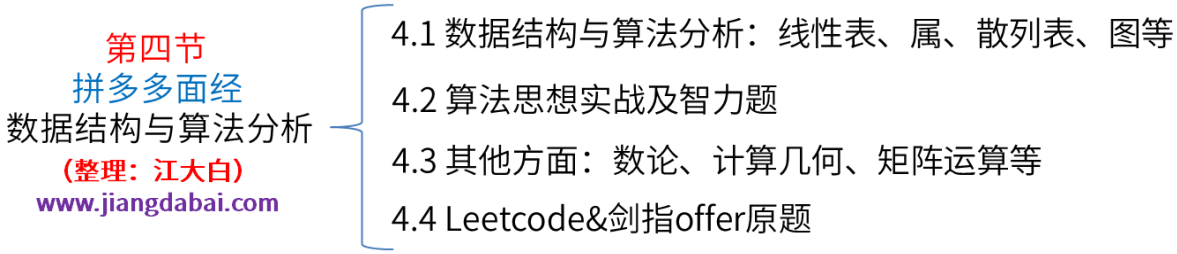
- 说一下 GAN 的原理，有什么应用，有什么变种?

3.4 机器学习方面

3.4.1 推荐系统

- 介绍 FM、FFM?
- DEEP FM 讲一下?
- deepfm 协同过滤算法
- 推荐和搜索有什么区别?（面试官是做搜索的）

4 数据结构与算法分析相关知识点



4.1 数据结构与算法分析

4.1.1 线性表

4.1.1.1 数组

- 求最长上升子序列?
- 一维数组，最长上升子序列?
- 【1,2,3,4，7,8,8,8,9】有序数组，给定 k，找到 k 的最后一次出现的索引?
- 找数组中第 k 大的数
- 数组里面有一个数出现的次数超过一半，这个数是哪个?

- 旋转有序数组的查找?
- 旋转数组查找 target 的开始和结束索引?
- 数组先升序再降序，找到最大的数返回？注意边界条件。
- 从[7,2,3,4,9,1,7,5,20,6]找两个数，使得后面一个数减前一个数的差最大?
- 合并两个有序数组，用了 python 的 pop 操作，让我解释原理?
- 给一个数组，输出子集
- 两个数组的公共元素
- 给出一个整数数组 a，给出一个整数数组 b（无重复元素），根据 b 的元素顺序，逐个判断 a 中的元素对应位置是否等于这个值，如果相等输出 1，不等对应元素输出 0。
- 无序数组构建平衡二叉树

4.1.1.2 链表

链表反转，我用了迭代写法

- 给个链表 1->2->3->4->5->6

把它变成 1->6->2->5->3->4

- 给定一个链表和一个数值，把小于数值的放到前面，大于的放到后面?
- 链表快排

4.1.1.3 字符串

- 反转字符串?
- 字符串编辑距离？（两个字符串 s,t，可以增删改，将 s 转化成 t 最小步数）
- 给定字符串 s，求与 s 编辑距离为 2 的字符串集合?
- 假设一个字符串中除了一个字符外每个字符都连续出现正好两次，找出那个字符？（例如 aabbcdd，accbbdd 等）先写了一遍 O(N)的解法，然后在面试官的引导下想到了 O(logN)的解法。
- 字符串 A 交换两个字符任意次数能得到字符串 B，则 AB 是同一类，给定字符串求

有多少类?

- Pattern 匹配：判断字符串 S 是否匹配 Pattern (如’abbc’)？刚开始说构建一个字典，同时遍历 S 和 P，但是会出现两个字符同时代表一个 S 中的字母，遂加了个 set 对 S 中出现过的字符串进行记录。
- 最长回文子串

4.1.2 树

4.1.2.1 二叉树

- 找二叉树中两个节点的公共父节点，我一开始说的递归，让我想个非递归的方法，我说了个先层次遍历然后用一维数组存放，然后再用完全二叉树的方法去找。
- 字典树
- 给一个二叉树和一个目标值，找到和等于这个值的所有路径?
- 给两个字符串 s1 和 s2，将其表示为二叉树，问 s2 是否能由 s1 进行任意多次交换任意节点的左右子节点得到?s1，s2 的二叉树分查不限。
- 又问了复杂度和平均递归深度?
- 最大二叉搜索树
- 给一个二叉树的根节点，一个节点 p， 一个节点 q，找出 p，q 最近的公共祖先
- 判断是否为平衡二叉树?
- 二叉树的层序遍历?
- 树的非递归遍历?
- 二叉树中序遍历，允许用递归写
- 二叉树的深度遍历?
- 二叉树的后序遍历
- 二叉树从左到右遍历叶子结点，输出从根节点到叶子节点的路径?
- 一个普通的二叉树，从里面找到一个节点数最多的二叉搜索树（子树），输出这个

树的节点个数?

给定 dic=['ab','abc','abcd','bcd','bcde','bde','efg']，str='abcdefg'，求最长匹配 dic 中的元素，比● 如本例就是输出['abcd','bcde','efg']。

实现：把 dic 的元素建成前缀树，然后搜索前缀树根节点，如果匹配就一直搜到根节点。

● 给了一棵多叉树的所有的边，让返回一个合理排序，要求子节点必须排在父节点前面?

● 数组实现二叉搜索树?

4.1.2.2 堆

● 建堆的复杂度（O（N））

4.1.3 排序

● 计算 Topk?

● 快排，时间复杂度，最坏情况复杂度，如何改进避免最坏复杂度?

● 写一个堆排序

4.1.4 搜索

● 一个二维矩阵上的 DFS 问题:给你一个二维矩阵，每个点是一个 0-1 之间的像素值，规定超过像素阈值的称之为高点击诱导点，由相连的高点击诱导点组成的区域面积大于面积阈值的称之为高点击诱导区域，统计这个二维矩阵中高点击诱导区域的个数?

4.2 算法思想实战及智力题

4.2.1 算法思想实战

● 给一幅地图上色，相邻区域颜色不能相同（我用 BFS 做的）?

● 有 10 万行的数据，每一行有三个数据，分别是起始 ip 地址，终止 ip 地址，和这个 ip 段对应的区域 比如说：

[11.12.10.01, 11.12.255.255, 北京] .....

然后写一个函数，输入一个 ip 地址，返回对应的区域?

- 一套扑克牌 52 张，随机发 N 张牌，判断 N 张牌张有没有同花顺?

4.2.2 智力题

- 一个圆被分成 M 个扇形，一共有 N 种颜色，相邻扇形不同色，一共有几种涂法?
- 给一个股票价格序列，只准买卖一次，求什么时候买入什么时候卖出能够获得最大利润?
- 有 n 堆砖头，第 i 堆砖头的重量为 Wi，合并第 i 堆砖头和第 j 堆砖头需要耗费能量为 Wi+Wj，问合并 n 堆砖头最少需要耗费多少能量，以及合并的具体过程是什么?
- 不同长度的绳子有不同的价值，一根绳子如何切分可以让总价值最大。动态规划求解即可?
- 一个棋盘，起始点在左上，终点右下，棋盘上有一些棋子，找到一条路径，使得经过的棋子最多，并且记录下这条路径?
- 现有一组会议起止时间，会议的时间信息[[s1,e1], [s2,e2],...]，总共同时需要多少会议室，例:[[0,30], [5,10], [15,20]], res=2。

4.3 其他方面

4.3.1 数论

- 面积期望
- 扑克牌 123456，抽到 6 停止，求期望?
- 抛硬币，正面继续抛，反面不抛。问抛的次数的期望?
- 100w 个桶，100w 个小球，小球随机入桶，空桶的期望 （二项分布）
- 袋子里面有 n 个不同的小球，每次随机从里面取一个出来。每个球被取到的概率为 probs[i]，取到观察后再放回去。每个球至少被取到一次，需要的最少次数的期望。
- 重复数字不连续排列?
- 输入正整数 N，从 1 开始打印不超过 N 位的所有数字，比如:

N=1, 打印 1，2，3,。。。，8，9

N = 2，打印 1，2， 3， 。。。，98， 99

N = 3, 打印 1，2，3， 。。。， 998， 999

一开始以为是快速幂，本质是用字符串来存储 N 位大数，从 0 开始用字符串加法逐次加 1，然后打印出和，等最高位进位的时候就停止打印

- 数学专业的题目：

（1）对数几率怎么理解?

（2）凸函数的定义还记得吗?

（3）x^2 是不是凸函数?

（4）凸函数对导数的要求?

4.3.2 概率分析

- 出了一个统计概率题，扔硬币，然后求极限那种?

概率 1: 11 个球，1 个特殊球，两个人无放回拿球，问第一个人取到特殊球的概率?

概率 2: 11 个球，1 个特殊球，两个人有放回拿球，问先拿到这个特殊球的概率?

- 求圆内接三角形过圆心的概率
- 从 n 个数据中抽取 m 个数据，保证每个数据被抽到的概率为 m/n?
- 求出 n 个色子之和为 s 的各种概率（可以用递归和动归做）
- 10W 个人有一个人生病，检测无病的 1%假阳性，有病一定检测出，现有一个人检测有病，则真实有病的概率?

4.3.3 矩阵运算

- 二维矩阵中，重复数字组成的最大面积 ？
- N*N 矩阵顺时针旋转 90°，要求时空复杂度尽可能低?

4.3.4 其他

- 为啥泰勒展开成二阶，作用?
- 已知 X 的 2 范式=1（nx1），a（nxn），求 aX 的 2 范式最小值（拉格朗日）

- 斐波那契数列 O(log(n)): x^2
- 最高效的方式实现幂函数?
- 数轴上的最长连续线段（要求 o(n)以内）
- 集合的所有子集
- 一亿个数字，找中位数?
- 给数字 n，依次打印 1～10^n-1 (全排列）？
- 手撕代码，实现 LRU 算法

4.4 Leetcode&剑指 offer 原题

- Leetcode 10
- Leetcode 45：跳跃游戏
- Leetcode 233：hard 题
- Leetcode 151 题
- Leetcode 原题：顺时针打印矩阵

5 编程高频问题：Python&C/C++方面

第五节
拼多多面经
编程高频问题
（整理：江大白）
www.jiangdabai.com

5.1 Python方面：网络框架、基础知识、手写代码相关

5.2 C/C++ 方面：基础知识、手写代码相关

5.1 python 方面

5.1.1 基础知识

- python 字典的底层数据结构?
- for 适用的数据类型?
- 深拷贝和浅拷贝，如何操作?

- 单引号，双引号，三引号的区别?
- python 中的多线程与多进程

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 内存相关

- 说一下 C++中如何防止内存泄漏?

5.2.1.2 区别比较

- 虚函数和纯虚函数有什么区别?

5.2.1.3 讲解原理

- 头文件、命名空间、传值和引用等

5.2.1.4 讲解应用

- Const function(const) const，这三个位置的 const 有什么区别?

6 操作系统高频问题：数据库&线程&常用命令等

第六节
拼多多面经
操作系统高频问题
（整理：江大白）
www.jiangdabai.com

6.1 数据库方面：基础知识、手写代码相关

6.2 操作系统方面：TCP、线程&进程、常用命令相关

6.1 数据库方面

6.1.1 基础问题

- SQL 的考察包括：取数，排序，日期的操作，等等。
- SQL 连接方式的区别?

- 主键和索引有什么关系?
- 索引的类型?
- mysql 底层索引，怎么实现的?

6.1.2 手写代码

- 表 1：userID,age,city

表 2：userID,pv,orderNum

Q1：城市为上海，北京，pv>1000,orderNum <10?

Q2：在表 2 增加一个 dt（时间，精确到天），还是上述问题

Q3：有什么优化方式?

6.2 操作系统方面

6.2.1 线程和进程相关

- 线程和进程的区别?
- 进程之间可以通过指针共享内存吗 ？
- 死锁的原因及解决办法?

7 技术&产品&开放性问题

7.1 技术方面

- 如何度量两个短视频的相似度?
- 情景题：计算两个短文本的相似度，你有什么方案？假设两个短文本里面的实体词无法准确抽取（没有大量的实体词库），你又要怎么修改方案?
- 视频抽帧做分类任务，有的图片比较模糊怎么办?
- 假设你有一批数据，一个新样本，怎样在这批数据中找到与它最相近的样本
- 如果让你设计一个预测商品点击率的怎么去做，常见的决策树 ID3,C4.5,CART 有什

么区别？用 cart 好还是 c4.5 好 ？

- 场景题：拼多多在线海量以物搜物的问题

如何标注 (active learning)

如何训练 (online learning)

如何使用？不断增加新的商品怎么办？(不重新训练，直接计算高维特征)

- 10w 的物料库，对应的用户信息，给用户推荐 top10 的物料，怎么设计算法?
- 排序用的什么模型？特征是什么？具体是什么？为什么用这个模型？和其他模型的比较?

给一句话，怎么提取出里面的地名，要先分词吗 ？

- 问了我机器学习、深度学习是通过什么渠道自学的？我回答的是看书（西瓜书、机器学习实战、python 深度学习、数据结构与算法等）、看网课（coursera 网站上吴恩达机器学习、深度学习的网课以及斯坦福大学网课）、参加比赛（kaggle）、刷题（leetcode、牛客）

7.2 产品方面

- 场景题：给两个商品，判断是不是一个款式?
- 场景题：怎么把 CV 一些模型用到拼多多手机找同款这个任务里面?
- 场景题：假设现在有一个商品的详情页面，希望你来根据商品的介绍，自动化生成一些评论，你怎么做？ 如何发现商家卖与店家不符的商品?

7.3 开放性问题

- 如何生成用户画像?