

6|美团算法岗武功秘籍

1 美团面经汇总资料

第一节
美团面经
汇总资料
(整理: 江大白)
www.jiangdabai.com

- 1.1 面经汇总参考资料
- 1.2 面经涉及招聘岗位
- 1.3 面试流程时间安排
- 1.4 美团面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 美团面经-114 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

【机器学习工程师】、【数据挖掘工程师】、【广告平台的 NLP 工程师】、【AI 算法工程师】、
【美团点评到店事业群算法挖掘岗】、【无人驾驶算法工程师】、【美团地图部工程师】、
【nlp 和搜索方向工程师】、【智能推荐平台工程师】、【机器学习引擎框架开发】、【机器学习/数据挖掘算法工程师】、【推荐算法工程师】、【美团到店数据挖掘算法工程师】、【美团点评机器学习工程师】

1.3 面试流程时间安排

美团面试流程-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	项目细节问的很细， 对应用场景也会发散提问
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	在一面的基础上， 更多维度提问
第三面	技术Leader面	自我介绍+项目经验+公司发展	和项目相关， 主要关注解决问题的思路
第四面	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。

1.4 美团面经面试心得汇总

★ 写题比较简单，模型扣的细，细的程度不是理解网上博客写的就可以，而且要看原论文。一面的时候被憋的几乎想拿书包走人了，说的最多的话就是，这个记不清，这个不了解，建议大家深扣细节！

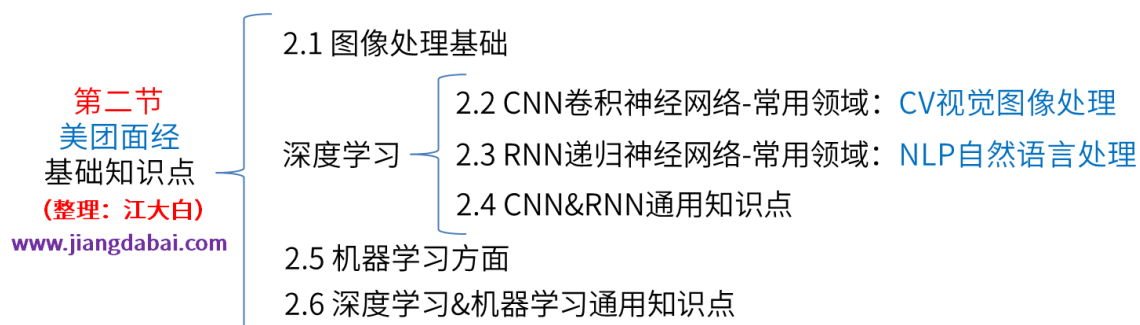
★ 美团的面试官水平真的高，技术面问的我都淌汗了，总监面也很专业。

★ 会经常问算法题目+机器学习相关的，根据项目问

★ 项目问的比较清楚，所有的东西都是由简历的内容进行拓展。一般都会有手写代码题，所以常见的一定得刷一刷。

★ 机器学习的偏多，底层的算法都会涉及到，但是简历上的项目建议要准备好，而且虽然是算法岗，但是也会问很多计算机基础相关的问题。

2 美团面经涉及基础知识点



2.1 图像处理基础

2.1.1 讲解相关原理

- 透视变换和仿射变换讲一下？
- 写出旋转变换矩阵（三维的那种）
- 写出带平移的旋转变换矩阵，写出带缩放的旋转变换矩阵，写出射影变换矩阵

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

- 是否了解图卷积？
- CNN 1×1 卷积核的作用？
- 卷积神经网络的权重是怎么更新的？
- dropout 的随机因子会对结果的损失有影响吗
- dropout 怎么反向传播？
- 卷积是空间不变性还是时间不变性？
- CNN 网络有哪些层？

- 感受野在 cv 中的作用，大小分别有什么影响 全局感受野和局部感受野的优缺点，哪些论文的方法是从这方面考虑并进行改进的，介绍一下他们的方法？

2.2.1.2 池化方面

- 平均池化和最大池化的反向传播是怎么运作的？
- 池化层如何反向传播？

2.2.1.3 网络结构方面

- Inception V3 和 ResNet50 网络模型用过吗？
- 简单讲一下 Inception 家族(各种优缺点以及改进)
- 讲一下 ResNet 的原理以及它解决了什么问题？怎么解决的？
- Resnet 为什么有效果？为什么能够保证很深的网络具备不错的效果？
- 对感受野的理解？例如 VGG 网络，最后一层卷积网络输出图片对于输入图片的感受野的大小？
- SEnet 的结构？SEnet 如何放到 Resnet 的 backbone 里？
- RoI Pooling 和 ROI Align 区别？

2.2.1.4 其他方面

- 神经网络中的偏置项 (b) 尺寸应该是什么样的？
- BN 知道嘛？讲一下 BN 的原理，作用？它有四个公式，每一个公式分别是什么，有什么各自的作用？
- BN 为啥可以缓解过拟合，详细讲一下？BN 有哪些需要学习的参数啊，BN 训练和测试是怎么做的？
- 除了 BN，你还知道那些其他的加速收敛的方法(楼主说到了归一化)，面试官说，和 BN 差不多的那些你了解吗？(GN, IN, FN)
- BN 一般用在网络的哪个部分啊？
- BN 底层如何计算，手撕 BN，BN 在训练、测试阶段的计算有什么区别
- 用公式推导小的 batchsize 会对模型训练有什么影响，我回答了 BN 方面的一些影

响，面试官说不行，从 BP 角度考虑。

- 如何解决梯度消失问题？
- 梯度消失，梯度爆炸讲一下？怎么解决？
- 分类问题的 loss 为什么选交叉熵，MSE 可以吗？基尼系数的公式为什么这么写？

2.2.2 公式推导

- 手推 BP 算法公式（就只有一层隐含层的那种）
- Softmax 的计算公式是什么？为什么使用指数函数？
- 用公式详解 BP 原理
- 通过公式解释链式法则以及 resnet？

2.3 深度学习：RNN 递归神经网络方面

- RNN 为什么梯度消失？ 答：tanh 激活函数 以及序列过长会导致梯度消失。还有一个原因是因为 RNN 是每一步都共享权重的。
- LSTM 跟 RNN 的区别，他和 RNN 相比有什么优势。
- LSTM 的信息传递机制是什么？
- LSTM 原理，与 GRU 区别？使用场景的不同点？
- BiLSTM 相比 LSTM 有哪些 case 上的提升。Attention 是如何加的取得了哪些效果的提升？
- LSTM 结构以及从数学层面谈为啥优于 RNN？

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

- 注意力机制的运行过程是什么样的？
- Local attention 和 global attention 的区别？
- attention 机制的作用以及选用的原因？

- 为什么设计神经网络解决问题，目前网络存在的问题是什么，后续可以怎么优化？
- 介绍 self-attention 计算，为什么用多头注意力？
- Dropout 的解释

2.4.2 模型评价

- 问了 Precision Recall?
- AUC 的作用，AUC 计算方法，AUC 的时间复杂度？

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

- 数据探索（拿到数据之后怎么分析的？）

2.5.1.2 特征工程

① 特征降维

- 降维了解吗？PCA 是什么，矩阵的特征值和特征向量的物理意义是什么？
- 为什么用 SVD？
- 说说矩阵分解？

② 特征选择

- 特征选择有哪些方法？（这个也是有很多，从过滤法、包装法、嵌入法来论述，条理依旧很重要）
- 特征筛选都有哪种方式（卡方检验的公式是啥）
- 了解 Embedding 吗？图嵌入的训练集是什么？

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）

- boosting 和 bagging 在不同情况下的选用？

- Xgboost 和随机森林的区别？

A.基于 bagging: 随机森林

- 随机森林怎么计算特征重要性？
- 讲下随机森林，随机森林的随机体现在哪里？

B.基于 boosting: Adaboost、GDBT、XGBoost

- GBDT 的原理，以及常用的调参的参数
- GBDT 怎么处理类别特征，例如 ID 特征
- XGBoost 中对 GBDT 有哪些优化？Xgboost 中的行抽样，可以起到哪些作用？
- XGBoost 原理，怎么防过拟合，gbdt 推导？
- XGBoost 和 lightgbm 的区别和适用场景？
- gbdt 推导和适用场景？
- LightGBM 和 XGBoost 是怎么处理缺失值的？
- RF 和 GBDT 的区别？
- Xgboost 跟 RF 的区别？你觉得在什么场景下，哪个性能会更出色？
- XGboost 里面预排序是怎么做的？
- 比较 GBDT、XGBOOST 区别
- xgboost 怎么处理 null 值？null 值分裂的时候在正常值的哪一边？
- 介绍 xgb 和 lgb，改进了什么？
- XGBoost 如果损失函数没有二阶导，该怎么办？

② 线性回归

- 线性回归跟逻辑回归的区别？

③ 逻辑回归 LR

- LR 的推导，损失函数？
- 讲一下 LR，LR 怎么优化的(说到了线性回归，然后面试官又追问了线性回归怎么优

化的？说了梯度下降，但面试官好像不满意，他是想让我说极大似然估计？

- 为什么说 LR 是广义线性模型？
- LR 的损失函数写一下，极大似然和最大后验的区别？
- 问了 LR 和 DecisionTree 的区别？
- GBDT+LR 的设计理念是啥，为啥要这样设计，为什么不用 RL，而是 GBDT？

④ SVM（支持向量机）

- 解释 SVM，问到了 SVM 原理，对偶过程等。
- SVM 介绍一下？为什么可以使用对偶来求解原始问题？
- 核函数了解吗？核函数解决什么问题？我说核函数解决了当前特征空间中样本不可分的问题，他说只要是样本不可分就用核函数吗？
- SVM 的松弛因子作用？
- SVM 的理论依据，如何推导？
- One-class svm 和传统 svm 的区别，你看过 one-class svm 的底层代码吗？
- 介绍一下 SVM，SVM 如何扩展维度？

⑤ 朴素贝叶斯（Naive Bayes）

- 解释贝叶斯深度网络，并说明其优缺点？
- 贝叶斯分类的前提假设？

⑥ 决策树（DT）

- 具体说明一下决策树如何划分，写出相应的公式？
- 决策树将一个特征全部乘以 2 会有什么影响？
- 决策树的分裂方式？（id3,gini,gdbt,xgboost）
- 决策树分类划分的选择有哪些？
- 如果你是算法的设计者，你会怎么设计决策树回归？
- 关于决策树的了解？回答了从 ID3，C4.5 到 CART 各自的优缺点

- C4.5 和 CART 如何处理连续变量以及样本量问题？

2.5.1.4 无监督学习-聚类方面

- 对机器学习的了解，项目中视觉词袋有用到 K-means，问 K-means 初始类中心的选择？

- GMM 介绍一下？

2.5.2 手推算法及代码

2.5.2.1 手推公式

- SVM 的推导，挨个步骤说清楚，对核函数的理解
- 推导 LR
- 手推了 GBDT 公式，问的很细，得对公式非常熟悉

2.5.2.2 手写代码

- 手写实现 kmeans 的伪代码？

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

- 写一下交叉熵损失函数？

2.6.2 激活函数方面

- 常见激活函数的异同？说一下，楼主说了 sigmoid, tanh, ReLU, leaky ReLU, PReLU, ELU, random ReLU 等。
- Softmax 和 Sigmoid 区别？
- Sigmoid 函数讲一下，ReLU 讲一下？它们俩的区别以及 ReLU 的优点
- ReLU 有哪些缺点啊，怎么解决的？其他的解决方法你知道吗？

2.6.3 网络优化梯度下降方面

- 梯度下降，何时收敛何时震荡？

2.6.4 正则化方面

- 正则化有什么用，为什么有用，L1 正则为什么能使参数稀疏，为什么能防止过拟合？
- L1、L2 如何选择？区别？
- 为什么 L1 更容易产生稀疏解，从数学公式回答？
- 分别介绍 L1 和 L2 正则化的方式和优缺点

2.6.5 过拟合&欠拟合方面

- 过拟合、欠拟合讲一下，怎么解决？
- 如何防止过拟合，你都采用了哪些方法，还有哪些你没有用到的方法？
- 过拟合和欠拟合哪种比较好处理？答：过拟合，因为欠拟合要么增加数据非线性来提高模型性能，要么增加数据集，做增广（图像方面）。过拟合可以加 dropout，BN，L1，L2 正则。如果实在不行，可能要重新查看数据集，看数据集是否有问题
- 神经网络和树模型的过拟合的一些参数以及解决方法？

2.6.6 其他方面

- 如果样本不平衡，怎么处理？

3 美团面经涉及项目知识点

第三节
美团面经
项目知识点
(整理: 江大白)
www.jiangdabai.com

- 3.1 深度学习：CNN卷积神经网络方面
- 3.2 深度学习：RNN递归神经网络方面
- 3.3 强化学习方面
- 3.4 机器学习方面

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

- 讲述一下 SSD 和 YOLO（我把 DSSD、DSOD、FSSD、RFBNet 等全讲了一遍，外带了一些语义分割的网络，简历上写了，全讲了）
- 简述 Cascade R-CNN 的提出为了解决什么问题？

3.1.2 目标追踪

- Deep sort 跟踪算法使用过吗？

3.1.3 图像分割

- 说一下你知道的经典图像分割网络(不限于医学图像)，我说了 FCN，SegNet，Deeplab 系列，顺便也说到了 U-Net。

3.2 深度学习：RNN 递归神经网络方面

3.2.1 自然语言处理 NLP

3.2.1.1 讲解原理

① Bert

- 简单问了 Bert 模型和 Xlnet 模型？
- Bert 里 Transformer 的多头 attention 是怎么做的？
- Bert 和普通的 Word2Vec 模型相比优势在哪里？他为什么会有这样的优势？
- Bert 为什么只用 Transformer 的 Encoder 而不用 Decoder？
- Bertde 的原理讲一下？

② Transformer

- 问 transformer 的 encoder 输出的 k, v 是不是相同的，如果相同，为什么需要两

个？

③ Word2vec

- Word2Vec 讲一下，想用 Word2Vec 构造的特征表达什么信息（项目）
- Word2vec训练过程的最后一步有什么办法可以优化softmax的计算？指数函数的计算会用查表来近似代替。
- Word2vec 与 glove 区别？

④ 其他

- Deepwalk 是手写还是工具包，有没有用 numpy？
- Seq2seq 中 scheduled sampling 如何做的，RL 部分训练过程中数据集如何构造？
- 问到了 RL+Seq2seq 的一些技术，比如 Seq2seq 怎样和 RL 结合，这里的 action 与 state 都是什么，如何设计 reward 等，为什么选取这样的 reward,具体训练流程是怎样的；
- 怎么样做实体抽取，怎样进行 aspect-level 情感分析，你们模型中增强学习的 reward 如何设计的？为什么这样设计？
- One-Hot 编码的优势？

3.2.1.2 损失函数类

- CTC 介绍一下？

3.3 强化学习

- GAN 里面经常会用到 KL 散度，来写一下 KL 散度的公式？
- 为什么要用强化学习，这个问题还能怎么解决，强化学习好处是什么？

3.4 机器学习方面

3.4.1 推荐系统

- 使用深度模型的话为啥使用 deepfm 而不使用 wide&deep 之类的？
- DNN 与 Wide&Deep 与 DeepFM 之间的区别？
- 传统的机器学习和深度学习在推荐上的异同点，深度学习的优势在哪里？
- 说说推荐系统算法大概可以分为哪些种类：(1) 基于内容；(2) 基于协同过滤：基于内存 (UB IB)；基于模型 (MF)
- 推荐系统了解多少，冷启动问题解决什么方案？
- 协同过滤 itemcf 怎么改进？
- 可不可以将深度学习关于图像方面的东西加入到推荐系统中去？
- 文本内容推荐中有哪些内容可以应用到商品团购推荐当中去？
- 因为项目用了 deepfm，所以还发散问了下 fm 的原理，以及 fm 怎么降低复杂度的？
- 对 fm，deepfm，wide&deep，DIN 的看法和了解？
- EE 有哪些别的方法，汤普森采样和 UCB 分别适用什么场景？
- 你还知道哪些 CTR 预估模型，你都用过哪些？
- FM，LR 的区别？FM 的计算复杂度？
- 介绍 Wide&Deep 和 deepFM，有啥区别，有啥优势？

4 数据结构与算法分析相关知识点

第四节
美团面经
数据结构与算法分析
(整理: 江大白)
www.jiangdabai.com

- 4.1 数据结构与算法分析：线性表、属、散列表、图等
- 4.2 算法思想实战及智力题
- 4.3 其他方面：数论、计算几何、矩阵运算等
- 4.4 Leetcode&剑指offer原题

4.1 数据结构与算法分析

4.1.1 线性表

4.1.1.1 数组

- 一个数组，找出第 k 大的数、这些方法的时间复杂度是多少？
- 给一个数组，用最小空间复杂度算出数目大于 1 的值
- 二维有序数组 找 target
- 先递增后递减的数组查找最大值（二分法）
- 从给定数组中找到三个和为定值 k 的数？
- 一个升序数组{2,6,7,8,8,9, 100}找出其中重复的数字，返回重复数字的起始位置和重复次数？
- 一维数组[1,n],n 可以认为正无穷，然后输入无数个区间[a,b],保证 $b > a$ ，求所有区间长度？
- 最长的可整合子数组的长度
- 旋转数组找目标值？
- 有序旋转数组查找？
- 求一个类似旋转数组的拐点位置？由于开始对题目描述理解有点问题，写了一个 $O(n)$ 的过程，被问到是否有更快的解决方法，在启发下写了类似下面的二分查找过程。
- 翻转数组，找一个值
- 两个有序数组，求中位数
- 给一个数组，里面只有一个数只出现一次，其余均出现两次，找出那个数（会挖坑，就是给你错误的输入，比如输入的数组大小是偶数）
- 如果只有两个数只出现一次，其余均出现两次，找出这两个数？
- 两个有序数组，找到它们第 K 大的数？
- 两个有序数组合并

- 打印 N 个数组整体最大的 Top K: 有 N 个长度不一的数组, 所有的数组都是有顺序的, 请从大到小打印这 N 个数组整体最大的前 K 个数?
- 顺时针打印二维数组: 关键考点是边界条件, 奇数偶数两种情况如何简化代码, 极端情况(例如 1×1 的矩阵)要确保能打印
- 数组的最大连续子数组和
- 一个数组, 分为两个子数组, 使得它们的和相等
- 对一个数组随机排序, 每个位置都随机, $O(n)$?
- 给定一个数组 $[3, 2, 5, 0, 2, 0, 0, 0, 7]$, 实现一个算法, 使 0 都放置在数组末尾, 其他元素保持顺序不变? 要求时间复杂度为 $O(n)$, 空间复杂度为常数。
- 给一个数组, 将其中的非 0 元素移动到做边, 0 移动到右边, 不改变非 0 元素的相对顺序?
- 删除一个排序数组的重复项
- 处理 IP 地址, 例如 192.18.23.57, 将四个数存入数组中 (数组范围在 0 到 255 之间, 为何是这个范围, 因为只有 8 位, 可能有字母或无效的情况出现), 用的队列, 面试官说想到这个很不错, 不过可以调函数实现 (可能在考察 C++ 的一些字符串处理库)
- 给定一个目标值 M 的数组, 返回数组是否存在和为 M 子集

4.1.1.2 链表

- 判断一个链表是否有环路, 并找到环路入口?
- 判断一个链表是否有环? 快慢指针可以解决
- 找相交链表的交点
- 合并两个排好序的链表
- 链表问了很多: 找中点, 是否有环, 环的入口, 是否有交点, 交点在哪里, N 个链表是否有交点, 复杂度分析?
- 链表倒数第 k 个值

4.1.1.3 字符串

- 翻转中间由各种符号隔开的字符串？
- 字符串转换为数字,比如'123'变成 123.【要考虑特殊情况 '0123' >123, '12 3'>123, '123' >123】
- 输入一个字符串,判断其是否是“(”和“)”的一一配对
- 两个字符串拼接起来,判断有没有回文?
- 输出最大无重复子串
- 给定一个字符串列表(长度为 n),给定一个滑动窗长度 L,求滑动窗里边最多有多少个不同字符?(要求时间复杂度 $O(n)$,空间复杂度 $O(1)$)
- 求两个字符串的最长公共子串
- 括号字符串是否能完全匹配
- 求一个字符串的顺序子串的个数
- 一个字符串,找到第一个只出现一次的字符, n 空间 n 时间,只能扫一次
- 字符串把多个连续空格合并成一个,输入是 char*,要求原地空间

4.1.2 树

- 问了满二叉树和完全二叉树,大概画了一下
- 从右边看被遮挡的二叉树,求露出的 node?
- 判断两个二叉树是否相同,递归可用
- 树的层序遍历,每一层单独输出,且从底向上输出
- 翻转一棵二叉树
- 输出二叉树最长的路径?
- 层序遍历二叉树
- 中序遍历二叉树
- 手撕反转二叉排序树递归和非递归 问了时间复杂度?

- 二叉搜索树如何序列化和反序列化，如果是普通的二叉树、多叉树呢？
- 给二叉搜索树插入节点，然后逐个返回父节点的值
- 文件查找树（递归、DFS）

4.1.3 排序

- 冒泡排序的复杂度说一下？
- 写一下归并排序
- 100 万数据中找前 10 大的数，随时找到 100 万数组中的某个数， $O(1)$ 复杂度
- topK，用了快排和堆两种思路，面试官让手推快排时间复杂度
- 海量数据，内存不够的情况下如何以最快速度进行排序？
- 排序算法，快排复杂度，最坏情况，怎么优化？
- 手写最小堆代码
- 给一个数组，找出最小的 K 个数，用了堆排序 $O(n\log k)$ ，有没有更快的方法？
- 一个很大的数组找出第 k 个数等（基于桶排序回答的）

4.1.4 搜索

- 平面一些点，距离近的算一类，输出可以有几类（我用的 dfs，复杂度高了点）希望复杂度降下来。
- 二叉树 DFS，BFS，如果层次遍历要从右到左

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 四边形四个顶点 S、A、B、C，每个顶点都与其它三个点相连，所以把它们平铺成平面，就是一个 2×2 的网格，两点之间都有连线（画出来就是一个框框里一个叉×）。把 S 坐标定为 $(0, 0)$ ，就是求：走 k 步回到 S 点的路线有多少个？
- 用加减乘除算根号 2

- 写了个 DFS：一个 $n \times n$ 的矩阵，一个阈值，找到大于这个阈值的点，若它的上下左右都满足条件，它们则聚为一类，输出每个类别所含点的坐标

- 8 升、5 升、3 升水桶各一个,如何分成两个 4 升,写出允许状态集合和允许决策集

4.2.2 智力题

- 抽蓝球红球，蓝结束红放回继续，平均结束游戏抽取次数？

- 五个外卖，先取后送，一共十个点，有多少种排列组合情况？

- 如何来判断一个骰子是否是均匀的？

- 找出 9×9 的数独的一个解？

- 两瓶墨水，一红一黑，用小勺从红墨水瓶里舀一勺放入黑瓶，搅拌均匀，然后从黑瓶里舀一勺放入红瓶，这时红瓶里的红墨水多还是黑瓶里的黑墨水多？如果不搅匀呢？

- 25 匹马，5 个赛道，最多几次可以知道前三名

4.3 其他方面

4.3.1 数论

- 假设 A 和 B 服从 $(0,1)$ 的均匀分布，并且 A 和 B 相互独立，求 $\max(A, B)$ 的数学期望？

- 给定一个数轴，在数轴上放一些点，给定一个长度的标尺，在标尺内最多能有多少个点(当时题意理解不对，说用滑窗，考官说思路对)

- 10 万个数字找最大的 k 个

- 输入两个整数 n 和 m，从数列 1, 2, 3.....n 中随意取几个数,使其和等于 m？要求将其中所有的可能组合列出来。

- 给定整数 n，求离根号 n 最近的整数？用了二分查找实现。

- 给定一段升序整数序列，求出积最小且和为定值的两个数，设置头尾两个指针即可

- 一个整数数组，找最长的先增后降的序列

4.3.2 概率分析

- 比如 10 只（5 双）鞋子抽 4 只，不能配对的概率？
- 支付宝集福活动，假设五张福的概率一样，那么平均多少张福可以集齐五福？
- 给定 n 个小球，有放回地采样。当 n 趋向于无穷的时候，某小球不被取到的概率是多少？
- 问个数学题，一副牌 54 张，去掉大小王，还有 52 张，从中抽取 2 张，问是一红一黑的概率多大？
- 一个商店，1 个小时卖出去 5 个包子，问下一个小时卖出 6 个的概率？
- 一个家庭有两个孩子，已知有一个是女孩子，问全是女孩子的概率是多少？
- 一根木棍随机砍两道，构成三角形的概率？
- 某村庄的习惯是一直生到男生为止，求村庄的男女比例？
- 一个小时平均闯红灯 5 次，问一个小时闯红灯 6 次的概率？（泊松分布）
- 一家人两个孩子，已知一个是女儿，问两个都是女儿的概率？（条件概率）
- 一个人打靶十次命中 7 次，命中率是 70%，这个概率是怎么估算出来的？
- N 枚真硬币是一面图案一面字， M 枚假硬币是两面图案，选了一枚抛 K 次都是图案，问是真硬币的概率——贝叶斯

4.3.3 并查集

- 两个集合求交集

4.3.4 其他

- 实现 `Math.sqrt()`
- 用最快的方法计算 2 的次幂? 2 的次幂溢出怎么办？
- 查找两个数的中位数(一步步深入, 先从空间复杂度 $O(n+m)$ 到时间复杂度 $O(n+m)$ 到 $O(\log(n+m))$) ?
- 通过概率未知的非均匀硬币生成 1 到 N 随机数？

- 求因子个数为 n 的最小数字(写了暴力的方案)
- 给一个数列，找到最长上升子序列并输出？（dp+回溯）
- [4,5,6, 1,2,3] 找到两段有序数列的分割点？先说了个顺序查找 $O(N)$ 然后说了个二分查找 $O(\log N)$ 并实现？
- LRU cache 设计
- DAG 遍历
- 基尼系数是什么？为什么用基尼系数不用熵？
- 判断点是否在矩形框内

4.4 Leetcode&剑指 offer 原题

- Leetcode 172
- Leetcode 300
- Leetcode 543
- Leetcode 695：岛屿的最大面积
- Leetcode 1482
- Leetcode 简单题：求 $n!$ 后面有几个 0，要求不把 $n!$ 求出来？
- Leetcode 简单题：给一个数组[1,2,3,0,0,1,0,2]，把 0 移到数组的后面，非 0 数字保持相对顺序。要求用双指针。
- 剑指 offer 原题：不用加号实现加法

5 编程高频问题：Python&C/C++方面

第五节
美团面经
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 网络框架方面

5.1.1.1 Pytorch 相关

- PyTorch 里增加张量维度和减少张量维度的函数？

5.1.1.2 Tensorflow 相关

- TensorFlow 参数服务器实现、为什么选 MXNet、参数服务器分布式训练过程、MXNet 中 KVStore 的实现细节？

5.1.1.3 其他

- 平时主要用什么深度学习框架？
- Mxnet 跟 TF 介绍一下

5.1.2 基础知识

5.1.2.1 线程相关

- Python 里面的多线程，多进程说一下？
- 线程之间怎么通信，进程之间怎么通信，python 的多线程有用吗？

5.1.2.2 内存相关

- 知道 python 的垃圾回收机制吗？

5.1.2.3 区别比较

- 数组和链表的区别？
- Python 里面的深拷贝和浅拷贝说一下？
- 创建对象时 new 和 init 有什么区别？
- python 的单下划线和双下划线有什么区别？

5.1.2.4 讲解原理

- Python 的 try except finally？

5.1.3 手写代码相关

- 有一个 n 个数字的序列，现在想把这个序列分成 k 段连续段，想知道分出来的 k 个连续段的段内数字和的最小值最大可以是多少？
- Python 里面的 lambda 表达式写一下，随便写一个？

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 区别比较

- C++ STL 容器以及命令有哪些，区别是什么？

5.2.1.2 讲解原理

- C++中的 export 讲一下；虚函数讲一下；构造函数能是私有函数吗？C++是面向对象的吗？
- C++学过吧，map 的底层实现是啥？
- 了解 Redis 吗？
- shared ptr 的特点是什么，可以引用传参吗？

5.2.2 手写代码相关

- 编写 C 语言 atoi 函数

6 操作系统高频问题：数据库&线程&常用命令等

第六节
美团面经
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

6.1.1.1 区别比较

- `count(1)`, `count(*)`, `count(列名)` 这三个有什么区别？

6.1.1.2 讲解原理

- 数据库事务，隔离级别，引擎，为什么用 B+树
- 数据库隔离级别、数据库最左匹配
- SQL 熟悉吗？讲一下 SQL 引擎

6.1.2 手写代码

- SQL 了解吗？写个 SQL，一个表字段有学生 id, 学科、成绩，选出所有学科成绩都大于 60 分的学生学号。

6.2 操作系统方面

6.2.1 线程和进程相关

- 进程的状态和转换，进程调度算法，进程间通信，消息队列有什么好处？
- 讲一下多线程和多进程的原理和区别？

6.2.2 常用命令

- Linux 查看进程命令

7 技术&产品&开放性问题

7.1 技术方面

- 树模型跟深度学习的区别？垃圾邮件分类哪个更适合？
- 如何判断一个字符串是手机号？

- 场景题，有很多正例，没有负例，然后有很多未标注数据，尽可能的从这些数据中寻找负例（我是从聚类、最近邻以及异常点检测的思路去想。one-class-svm 应该也是可以的）。场景题回答的好是很关键的，其余的都可以背下来，但是只有这部分才会考你对算法的理解和对问题的处理方式。
- 场景题：如何给商家选择头图，以及 topk 的实现？
- 场景题：怎么做美团 app 的猜你喜欢，只能用 LR 模型，（特征，那些特征，怎么获取，怎么处理）特征怎么离散化、怎么设计整个逻辑，在线怎么获取用户的特征？
- 有 100 万条诈骗电话黑名单，现在有个电话来了，快速判断这个电话是否在黑名单里，要求查询 1000 条和 100 万条所消耗的时间一样？
- 场景题：一个 query，一些结果商品，怎么做点击率模型，怎么处理商家恶意点击？
- 场景题，问电商推荐可以用的主要特征有哪些？
- 如何根据美团的商品评论，生成商品的描述。传统抽取方法，语料大后上深度模型。采用类似于 TF-IDF 的思想避免抽取的描述太大众化没有特点？
- 数据样本不平衡问题有哪些解决办法？
- 假设两个分布 A 和 B，我们一般怎么衡量两个分布之间的距离，一般用什么距离？
- 场景题：找 top10 个常出现的 ip？
- 场景题，如果输入某关键词进行搜索 对于返回的结果可以从哪些方面进行设计？
- 开放问题：如何去做门店的推荐匹配

7.2 产品方面

- 场景题：如何判断异常账号（从注册的时候看）
- 如何找到从交易记录中找到最近 10 天最大一笔交易的时间？
- 场景题：如何判断刷单？
- 场景题：判断是不是垃圾信息
- 如何对图片拍摄角度进行纠正？

7.3 开放性问题

- 你会如何去统计一下北京所有大学生在食堂吃饭的比例以及消费情况?