

7|京东算法岗武功秘籍

1 京东面经汇总资料

第一节
京东面经
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 京东面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 京东面经-105 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) 实习岗位类

【算法工程师实习】、【京东搜索部门算法实习】

(2) 全职岗位类

【京东保险算法岗】、【数据分析工程师】、【搜索与推荐平台算法工程师】、【京东数科算法工程师】、【京东云下的应用研发部算法推荐工程师】、【语音识别算法工程师】、【机器学习算法工程师】、【推荐系统算法岗】、【图像算法工程师】、【京东零售部算法工程师】、【达达京东到家算法工程师】、【京东广告算法工程师】、【京东 NLP 算法工程师】、【广

告质量部算法工程师】、【京东物流算法工程师】

1.3 面试流程时间安排

京东面试流程-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	主要问项目和比赛经历
第二面	技术Leader面	自我介绍+项目经验+公司发展	偏实际和业务场景的问题 以及合作能力
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：

- 有的 HR 问的问题常规，项目中最有成就感的一次 最有挫折感的一次?为什么想要加入 jd?
- 有的 HR 问的很独特，比如：
 - (1) 你认为成为好朋友是契机重要还是相处过程重要
 - (2) 和人相处过程中是否有遇到突破你底线的事情
 - (3) 关注时事吗？说一个时事，以及你的感受
 - (4) 你觉得自己是哪种动物？

1.4 京东面经面试心得汇总

★ 问得比较杂，比较综合，比如操作系统、计算机网络、linux 系统、数据库、机器学习、大叔局，编程。

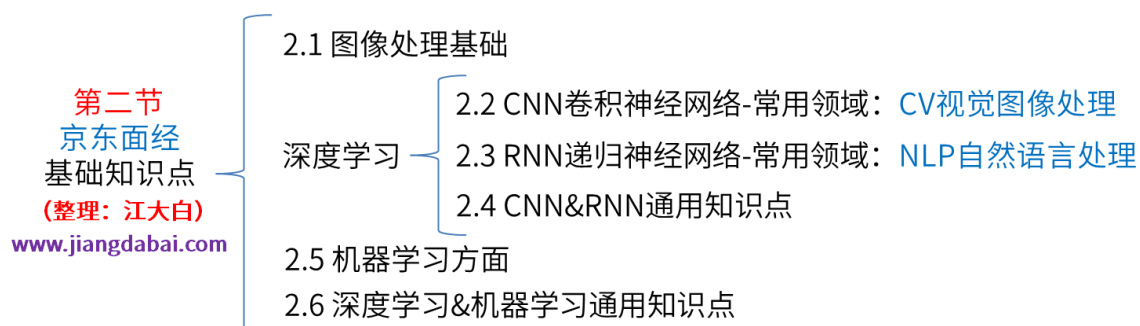
★ 京东的面试专注于考基础知识，基本不涉及特别深入的理解，我的面试时间相对来说比较长，一般都是 20 分钟左右的面试流程，面试的是最后是否会录取你的部门，京

东面试的流程比较快,但是发 offer 可能会比较慢,楼主在二面当天晚上就加到了 leader 的微信,期间一直询问我有没有收到 offer,但是最终时隔 20 多天才收到 offer,可能是校招组和内部沟通还是有时间的延迟吧,不过总体来说京东的面试体验挺好的,守时而且面试难度一般,面试官态度很好。

★ 各个方面都会问一些,针对会的问题会延伸问

★ 对于整个行业的动态,了解的比较多

2 京东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

- 膨胀腐蚀的原理讲一下?
- 传统去噪算法了解哪些, BM3D, NLM, 介绍下?
- 聊了傅立叶变换, 小波变换, 离散余弦变换?
- 图像的有椒盐噪声用什么滤波器?
- 中值滤波与高斯滤波的原理与运用场合?
- canny 算子和 sobel 算子的原理与运用场合?
- 霍夫变换检测圆的原理?

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

- 感受野受哪几个参数的影响，给了一个距离例子，计算感受野的大小？
- 上采样方式 subpixel，反卷积，resize
- 卷积核大小为什么是奇数？
- CNN 为什么参数共享？
- 简要介绍一下 dropout（训练测试时的步骤，为什么可以防止过拟合）
- Dropout 前向和反向的处理？
- dropout 原理，在测试时需要怎么补偿？
- Dropout 什么原理？
- CNN 为什么比 DNN 好呢？
- CNN 的权重共享平移不变的作用和意义怎么体现的？

2.2.1.2 网络结构方面

- 把 CNN 的发展历史从 2010 年开始按时间轴顺序说一下，说 Dense Net, Xception, 胶囊网络这些，然后问为什么 pooling 层不好，哪里不好，要用胶囊网络？
- 画一画 ResNet 的一个 BottleNeck？
- Resnet 说一下 shortcut, 两个 mapping、为啥可以无损传播梯度，为啥可以缓解网络退化
- Inception 网络多层卷积之后是 concat 还是逐像素相加？
- Xception 网络含义？
- ResNet、DenseNet 含义，处理方式，有什么好处，具体 concat 还是逐像素相加？
- 了解哪些模型，讲下它们的原理（VGG，Inception V1-V4，Resnet）
- Vgg 网络名字的由来？

2.2.1.3 其他方面

- 神经网络如何加速？
- CNN 和传统的全连接神经网络有什么区别？
- BN 怎么实现的？inference 时候具体怎么做的？
- BN 的参数，原理说一下？怎么做的标准化，作用是什么？为什么减少过拟合？
- Batch normalization 原理，先归一化然后恢复有何意义？
- BN 和 LN，问的很细，包括二者区别，为什么 BN 不在 RNN 中使用？
- 描述下前向传播、后向传播？
- 样本分布不平衡时，模型效果为什么不好？说明理由？
- DNN 和 CNN 区别？
- 梯度消失和梯度爆炸的原因，怎么解决？
- 解决梯度爆炸的方式（算法层面）？
- 梯度爆炸梯度消失（要求举具体的例子做为说明）
- 你的参数是怎么初始化的。全部为 0？随机初始化？高斯分布中随机取点？

2.2.2 数学计算

- 卷积参数量计算，尺寸计算？

2.2.3 公式推导

- 神经网络分类的 softmax 数学公式，如何计算

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 为什么会出现梯度消失的问题？
- 介绍一下 LSTM？GRU 跟 LSTM 有什么区别？
- LSTM 怎么解决 RNN 的问题？

- GRU 改进了，为什么不用？RNN 上面最近两年有什么新网络改进？
- LSTM 里面怎么处理输入变长的序列？追问那一个 batch 里面长度不一样怎么办，训练会出现什么问题？继续追问选择一批训练的数据满足哪些约束，随机选么？说一下对 LSTM 里面的 Attention 机制的了解？
- RNN，LSTM 原理？区别？为什么 lstm 门用 tanh？
- Transformer 和 lstm 的优缺点？
- LSTM 有几个门？各个门的作用是什么？公式是什么？LSTM 解决了什么问题？

2.3.2 手绘网络原理

- 写一下 LSTM 的结构和前向的传播公式

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

- 如果数据量很大，内存不够怎么办？
- transformer：位置编码，为什么用位置编码，self-attention
- self-attention 的作用和功能？
- 加不加 self-attention 在计算效率上有什么不同？

2.4.2 模型评价

- 知道哪些评价指标？
- 介绍下 AUC 和 F1-score？F1 值的计算公式说一下？怎么理解 AUC？
- 手写 AUC 的计算（小矩形积分得到总面积即可）
- 样本不均衡对 roc 曲线有影响吗？对 pr 曲线呢？为什么？
- 画一下混淆矩阵，写一下精确率和召回率的公式？
- 验证集是做什么的，测试集效果怎么评估？
- 介绍回归、分类用到的评价函数？

- 分类器评估标准（准确率，召回率，F1 值，ROC，AUC）
- 手写 recall，precision，f1score 公式以及公式中各个指标代表的含义？

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

- 有没有用过机器学习的降维方法？
- 讲一讲 LDA 算法（线性判别分析）？
- 介绍 PCA
- 特征工程预处理的流程？

② 特征选择

- 特征工程对于连续特征，我们通常有两种处理方式：1. 连续特征离散化；2. 特征缩放，这两种分别在什么情况下做？
- 讲了一些特征工程的技巧？
- 如何对连续特征进行离散化处理，为什么要这样做？
- 做数据分析选特征的时候有哪些评判指标？
- 所有模型都要求对数据进行标准化么？

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）

- 讲一讲树模型（RF, GBDT, XGBOOST）
- GBDT 和随机森林的区别来说一下？

- GBDT 和随机森林的树的深度哪一个比较深？为什么？
- 知道哪些集成方法？
- 如何用回归树实现分类算法？
- 常用的线性分类算法有哪些？
- 常用的非线性分类算法有哪些？
- xgb 和 gbdtd 的区别?gbdtd+lr 实现细节？
- 讲一下 XGB 的原理，优缺点，推一下公式？
- LGB、XGB 的区别和联系，并行是如何并行的？
- 除了树模型，Bagging 能不能接其他的基模型？
- lightgbm 讲一下，具体是怎么做的？和 xgboost 的区别讲一下？
- lightgbm 的直方图加速讲一下？具体是怎么来做的？叶子节点是怎么分裂的？说一下？
- Xgboost 和 LGB 原理？
- 说一下 GBDT 的原理，boosting 和 bagging 是怎么减少偏差的？
- gbdtd 各基学习器之间是如何产生联系的？
- 手写 xgboost 的目标函数，xgboost 构建树时候节点分裂的公式是什么？
- xgboost 如何调参,xgboost 可以自定义损失函数吗？
- 给定一个场景如何自定义损失函数？
- 如果样本的权重不一样如何自定义损失函数？
- sklearn 的 xgboost 支持哪些损失函数？
- 分类和回归算法都有哪些损失函数？
- 模型融合如何做的？
- bagging, boosting 和 stacking 的原理以及他们的区别是是什么？
- XGB+LR, XGB 充当什么角色？

- 为什么 XGB+LR 可以提高模型效果？
- 如何在 XGB 模型选择树的棵树时早停？
- XGB 的损失函数进行了二阶泰勒展开，为什么可以用泰勒展开？为什么用二阶而不是三阶四阶？
- 说一下自己理解的 XGBoost？
- XGBoost 和 GBDT 的区别是什么？

② 逻辑回归 LR

- 说了逻辑回归，在什么情况下你会选择用逻辑回归？
- 用 L-BFGS 来推导一下 logistic regression 的迭代公式？
- LR 损失函数介绍一下，如何优化？
- 讲一下逻辑回归？当数据量特别大的时候，逻辑回归(LR)怎么做并行化处理？
- LR 为什么不用 mse，svm 为什么用 hinge 不用 logloss，我不会，面试官耐心画图给我讲原理。问 svm 为什么要用核函数。
- 如果自己写一个 LR 的话，要包含哪些模块（我只说到了训练部分，面试官补充了还有分类预测的模块）
- LR 不做标准化有影响吗，神经网络呢？
- 简单介绍下 LR，写一下极大似然的函数？
- 送入 LR 前，如何处理数据（特征工程）
- 知道最大似然估计和最大后验概率估计么？
- 讲一下最大似然估计的原理？然后给出一个二项分布，让用最大似然估计手推出该分布的参数？
- 逻辑回归背后的数学原理是什么，如何推导的？

③ SVM（支持向量机）

- 讲一下 SVM (建模思想，误差函数推导，核，优化) SVM 的核函数有哪些？你都用过哪些？

- SVM 的推导对偶除了方便计算以外还有什么好处？
- 解释 SVM 的核函数，核函数的含义以及为什么能起作用？
- 为什么不用 SVM 做分类？从原理上讲一下 SVM，SVM 怎么解决多分类问题？
- LR, SVM 的原理，LR 和 SVM 区别，SVM 损失函数
- SVM 中有哪些调参经验？
- SVM 和 LR 的区别？
- SVM 数学上的实现？
- SVM：拉格朗日乘子，KKT 条件，对偶问题，核方法是什么，用过哪些核函数？

④ 朴素贝叶斯 (Naive Bayes)

- 贝叶斯思想了解吗？写一下公式并解释一下？
- 说一下朴素贝叶斯，为什么叫朴素贝叶斯？
- 朴素贝叶斯的好处？为什么那么多人用？

⑤ 决策树 (DT)

- 决策树原理，CART 树？
- 决策树 ID3 算法的特征选择指标，口述一下数学公式(信息增益)
- 决策树的启发式算法有哪些，不同算法分别用了什么准则来选择特征？
- 介绍一下 ID3、CART，逻辑树
- 说一下 C4.5 的过程，围绕决策树

2.5.1.4 无监督学习-聚类方面

- 聚类了解吗，说一下 K-means 聚类的原理以及过程？K-means 聚类怎么衡量相似度的？(我说欧式距离)K 的优化方法？
- 说一下 kmeans 聚类算法的原理，对于 k 和中心点怎么确定
- 如何评估我们的聚类结果，以及如何提升？
- 介绍 K-means 聚类，以及每次聚类结果是否一致，为什么？

- 介绍下 kmeans? 有什么改进方法么? 遇到很多维数据时会发生什么?
- 讲一下混合高斯模型, EM 的核心思想是什么?

2.5.2 手推算法及代码

2.5.2.1 手推公式

- 写一下 Gini 系数、信息增益、信息增益率的公式?

2.5.2.2 手写代码

- 手推 SVM: 空间上一点到超平面距离, SVM 整体代价函数, 如果进行对偶形。
- 详解 GBDT, 用伪代码实现树的生成和 boosting 迭代过程?

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

- 用过哪些损失函数, 有什么区别?
- 分类和回归都用什么损失函数, 分类为什么不用平方损失?
- 描述一下交叉熵, 求导?
- 交叉熵损失函数定义, 举例计算过程?
- 写一下 logistic regression 的损失函数
- 通过交叉熵做 loss, 怎样体现, 输入 x_1, x_2 , 输出 y_1, y_2 交叉熵如何计算, 比如 y_1 和 y_1' 越接近 loss 为何越小?

2.6.2 激活函数方面

- 了解哪些激活函数, 为什么要激活函数?
- sigmoid 和 relu 对比?
- 激活函数为什么要零均值输出?
- 为什么要用非线性激活函数, relu 右侧导数是 1, 为什么能作为激活函数?

2.6.3 网络优化梯度下降方面

- 参数优化方法说一下(梯度下降的三种方式的优缺点)
- 什么是梯度下降，有哪些优化算法，区别是什么，它们（SGD,BGD,mini-BGD）的区别？
- SGD 和 ADAM 的区别和联系？ADAM 算法比 SGD 优化好在哪儿？
- 深度学习里面的优化方法 momentum 和 Adam 来分别讲一下原理和公式？
- SGD 和 Batch 梯度下降区别？
- 怎么用的动态学习率，人工干预还是自动的？
- 手写 adam 更新公式？
- 平时怎么选择优化器？讲一下 adam 的优点？
- 深度学习常见优化方法有哪些？
- Momentumt 的公式，RMSProp,adam 的公式以及公式中参数代表的含义，以及他们分别解决了什么问题？

2.6.4 正则化方面

- 正则化有哪些方法？
- 过拟合问题。我答了几个方法。然后着重问了一下正则化的内容，l1、l2 正则化
- 正则化是怎么防止过拟合的？
- 介绍一下正则，L1 L2 的比较，为啥 L1 更稀疏？
- L1 和 L2 的数学解释，L1、L2 有什么区别，适用于什么场景？L1 为什么不用于卷积神经网络？
- 从多个角度分析 l1 和 L2 正则化为什么能防止过拟？

2.6.5 过拟合&欠拟合方面

- 过拟合是什么，如何解决，应对措施？
- 讲了讲深度学习训练中过拟合/Loss 不降等常见问题的处理方法？

- 模型效果不好的前提下，如何区分是过拟合还是模型复杂度不够？
- 从模型结构上如何解决过拟合？
- 讲一下偏差和方差(楼主从欠拟合和过拟合来讲的)
- 如何判断一个模型是处于高方差还是高偏差？
- 高方差如何调节模型，高偏差如何调节模型？

2.6.6 其他方面

- 写一下欧式距离的公式？
- 传统机器学习都是一次把全部的数据送进模型，现在深度学习为什么一次就一批？
- 数据类别不平衡怎么处理？

3 京东面经涉及项目知识点

第三节
京东面经
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

- faster rcnn 流程以及 RPN 的具体过程？
- 问项目相关，提升小目标检测效果的方法，kcf 的原理等等
- NMS 和 IOU 的计算
- Two-Stage 和 One-stage 结构的不同？
- 介绍了下 Faster RCNN, 问了 ROI Pooling。和卷积中普通的 Pooling 有什么区别？

你们做目标检测一般用什么损失函数？写一写 KL 散度和交叉熵函数？

3.1.1.2 损失函数

- 用过 Focal loss 吗?
- Yolo 的损失函数, v1 和 v3 损失函数的区别?

3.1.1.3 手写代码

- 手写 nms

3.1.2 图像分割

- 讲一下 unet 和 deeplabv2 的流程, 顺便问了下 deeplabv3?

3.1.3 OCR

- 讲一下文本分类模型
- 给一个新的文本分类任务, 会怎么选模型?

3.1.4 图像分类

- 常见的分类算法以及评估指标?

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

① Bert

- Bert 原理?
- Bert 内部结构(矩阵参数信息等)
- 问 Bert 的 attention 和普通的 attention 的区别, 具体怎么做的, 多头为什么要多头?
- bert 怎么分词? bert 的输入具体是啥?

② Transformer

- 画一下 transformer, 介绍一下结构, 说一下维度

- transformer 介绍

③ CRF

- 说一下 CRF 模型？

④ HMM 隐马尔科夫模型

- 讲一下 HMM 模型？

⑤ Word2vec

- 简单介绍下 word2vec 原理，对比下 CBOW 和 skip-gram 的区别？
- Word2vec，看过源码吗？源码里面是如何负采样的，为什么要层次化 softmax，sigmoid 在源码里面的计算方法是什么？
- Word2vec 三层结构很简单为什么效果这么好，word2vec 激活函数？
- Word2vec 和 bert 区别？
- 为什么 w2v 向量在语义空间内有很好的数学性质，比如相加减？

⑥ Deepwalk&Node2vec

- deepwalk 介绍一下，优缺点？（本来问的 word2vec，我说没有做过 NLP 的东西，介绍 deepwalk 可以吗，面试官说可以），deepwalk 的损失函数？

⑦ 其他

- 如果 onehot 等操作之后维度过高你会怎么做？
- 分词与实体识别的区别关系？
- fasttext 原理，同样要求画框架？
- 在无上下文的情况下如何看两个词是否是同义词？

3.3 强化学习

3.3.1 讲解原理

- 生成式模型和判别式模型的区别，都有哪些？
- 生成式模型和判别式模型具体讲下？

- G 和 D 具体结构？
- GAN 算法的二进制交叉熵函数怎么实现的（极大似然估计）？
- 介绍一下 GAN 算法？

3.3.2 损失函数

- G 网络的三种 loss 是怎么计算的，即 L1 loss L2 loss gan loss？
- 判别器 loss 如何度量？

3.4 机器学习方面

3.4.1 推荐系统

- 介绍一下 wide&deep 算法的原理？
- FM 模型与 LR 区别？怎么训练？FM 模型的具体公式，FFM 在此基础上有什么改进，如何确定每一个特征所属的 field？
- 搜索引擎的拼写纠正怎么做的？(楼主说了朴素贝叶斯和词袋模型)那如果第一个字母就输错了怎么办？词向量这一块有了解过吗？
- 讲一下 deepfm 的原理？
- 推荐里面的低秩矩阵分解具体是怎么做的？
- 个性化推荐是怎样过程，组员如何分配任务？如何进行优化改进？如果要考虑输入集的权重，在 fp-growth 中如何实现？如何评判他的一个推荐标准？（用户采纳度、收藏或者点进去看）fp-growth 在这个项目中的优缺点？

4 数据结构与算法分析相关知识点

第四节
京东面经
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

- 求数组里面连续子段和最大值？
- 旋转数组查找 k
- 旋转数组中查找某给定数（二分查找）
- 一个数组中出现次数最多的 K 个数？
- 2^n 个数组，每个数组长度都是 m，每个都是有序，合并成一个有序的数组？
- 数组中最大子数组的和，矩阵中最大子矩阵块的和？
- 一个数组里面，每 K 个数是一个递增的有序数组，将整个数组排序？
- 给定一个有序数组，统计目标值的个数。（二分查找，找到目标值的下界和上界。）

4.1.1.2 链表

- 反转链表
- K 个一组，反转链表
- 单链表的分组翻转（即 k 个一组翻转链表）
- 如何判断一个链表上是否有环？
- 快慢指针如果快指针走 3 步的话呢 还能奏效吗？如果快慢指针的起点不一样呢，还能奏效吗？

- 链表转化，1->2->3->4->5->6->7 转化为 1->7->2->6->3->5->4?
- 合并两个有序链表，合并 k 个有序链表（不让用递归），最小的 k 个数（指定用 partition，为什么时间复杂度是 $O(n)$ ）

4.1.1.3 字符串

- 字符串转 int?
- 字符串拼接成最大的数字的排序?
- 括号匹配：给定 n 对括号，求问总共有多少种排列方式？要求必须满足左右括号的顺序?
- 实现字符串的 lr_trim 算法，去掉字符串首尾的多个空格，函数名是 void lr_trim(char * s)

求字符串的最大回文子串(动态规划)

- 给一个小字符串数组，一个目标字符串，目的是找出数组中是目标字符串的子串的最大长度，我的想法就是维护字典树，加 KMP 模式匹配优化。
- 多个字符串，给定前缀和长度比例阈值，返回符合条件的字符串个数?
- 给出一个字符串，写出该字符串的全部排列组合?
- 反转字符串
- 给定 2 个字符串求最长公共子串的长度

4.1.2 树

4.1.2.1 二叉树

- 什么是二叉树、用伪代码说一下求二叉树的深度
- 二叉树镜像
- 给一个二叉树，和一个节点，找出该节点二叉树中序遍历下的下一个节点，如果树有父节点则个属性，在空间复杂度 $O(1)$ 的情况下找出来?
- 判断两棵树是否相同?

- 求二叉树每一层的最大值？
- 用伪代码说一下求二叉树的深度，如果用递归，具体实现，代码？
- 二叉树前序遍历？
- 二叉树的层次遍历

4.1.2.2 堆

- 寻找无序数组中的第 K 大的数，这个只需要说思路和复杂度？用最小堆 $O(N\log K)$
- 无序数组中找第 K 大的数，时间复杂度是多少？为什么是 $O(n)$ ，而不是 $O(n\log k)$ ，来推导并且证明一下你的解法的时间复杂度(级数求和)、 $O(n\log k)$ 的解法是怎么做的，说一下(堆排序)？

4.1.3 图

- 问最短路径算法有哪些？介绍一下 a^* 算法。回答：有起点到中间点的距离加上中间点到终点的一个估计距离。面试官又问，如果去掉第一项这个问题会变成什么？

4.1.4 排序

- 各种排序算法说下，写个插入排序
- 冒泡和快排是否稳定？还有哪些稳定的算法？
- 无序数组找第 k 大？快排&堆。写了快排。分别讲了一下两种思路，分析两个的时间复杂度。
- 如果几亿的数分别在不同机器上，怎么找第 k 大？
- 海量数据如何找到中位数和第 200W 个数
- 有 4 亿个数据，内存只能存 1 亿个数，找出第 8000 万大的数？你用堆排序做是吧？堆排序这种方法有什么缺点？如果我要找第 1.3 亿大的数呢？
- 手写快排，推复杂度
- 堆排序 (C++)
- 1000w 个数，数范围 $[-1000, 1000]$ ，写个排序？

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 上台阶，一次能上一个或者两个，问上 n 个台阶有多少种方法？（思路+撕代码）
- 纸牌博弈
- 最短路径问题
- 优先队列，列表中出现次数最多的数字，复杂度小于 $n\log 2n$ 。
- 有一个 1G 大小的一个文件，里面每一行是一个词，词的大小不超过 16 字节，内存限制大小是 1M，返回频数最高的 100 个词。

4.2.2 智力题

- 一维消消乐，红红黄黄绿绿绿绿黄蓝 \rightarrow 消一次变成 红红蓝 （三个以上的干掉）？

4.3 其他方面

4.3.1 数论

- 两个独立变量满足 0 到 1 均匀分布，求两个变量最大值的期望。

$$E(\max\{X_1, X_2\}) = \frac{X_1 + X_2 + |X_1 - X_2|}{2} = \frac{1}{2} + \frac{1}{2} \times E|X_1 - X_2|$$

4.3.2 计算几何

- 凸函数有什么优点？如何证明一个 n 元函数是凸函数？
- 拉格朗日乘子法能否求解非凸的目标函数问题？
- 凸优化的相关知识。
- 马尔科夫链的相关知识
- 线性代数里面的矩阵分解你知道吗？具体是怎么做的？

4.3.3 概率分析

- 最大似然估计解释下？
- 极大似然估计和最大后验估计的区别是什么？
- 计算抛骰子，抛 1 或者 6 庄家赢，2, 3, 4, 5 你赢，
庄家连续赢了三次，这个概率是多大，这样能说明骰子有问题吗？
如果抛了 100 次，庄家赢了 40 次，能说明有问题吗？
那怎样才能证明这个骰子有没有问题？

- 54 张扑克牌，大小王在同一堆的概率？
- a,b 丢硬币吃苹果，问吃到的概率？
- M 个样本有放回采样 N 次，问某条样本一次没被采中的概率？
- 某人有两个孩子，其中一个是女孩，两个孩子都是女孩的概率是多少？

4.3.4 矩阵运算

- 二维矩阵，从左到右从上到下递增，找 target？
- 给出一个二维矩阵，顺时针由外层到内层打印该二维矩阵？

4.3.5 其他

- 求 $\sin x$ ？
- 实现 `int sqrt(int x)` 函数
- 正则表达式匹配
- 已知有个 `rand7()` 的函数，返回 1 到 7 随机自然数，怎样利用这个 `rand7()` 构造 `rand10()`，随机 1 ~ 10？
- `rand1()` 生成 `rand5()`？
- 给你均值方差，让你利用正态分布随机生成 1000 个点？
- 乱序数据找第 K 大的数

- 字符流采 10 个字符，保证每个字符的采样概率一样？
- 动态规划，左右，求最大的路径和？

4.4 Leetcode&剑指 offer 原题

- Leetcode 4: 两个排序数组的中位数
- Leet code 47: 全排列 II
- Leetcode 143: 重排链表
- Leetcode 152: 求数组的最大连续子序和。时间空间复杂度？空间复杂度优化？空间复杂度优化？如果是连续乘积呢？
- Leetcode 215: 求数组中第 K 大的数
- Leetcode 206: 反转链表
- Leetcode 279: 完全平方数
- 剑指 offer 11: 旋转数组的最小数字
- 剑指 offer 51: 数组中逆序对

5 编程高频问题：Python&C/C++方面

第五节
京东面经
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

- 项目用的什么平台，keras、tensorflow、pytorch 都是哪家公司的，为什么喜欢用 pytorch？

- tensorflow 与 pytorch 区别?

5.1.1.2 Tensorflow 相关

- tensorflow 或者 caffe 的底层代码看过吗? 卷积是怎么实现的? GPU 进行并行计算时如何计算矩阵卷积的?
- tensorflow 中两个矩阵乘法的区别?

5.1.1.3 其他

- 常用的深度学习框架都有哪些?
- keras sequential 与自定义模型构建区别?

5.1.2 基础知识

5.1.2.1 线程相关

- 讲一下 Python 的多线程

5.1.2.2 内存相关

- Python 需要和 C++ 一样释放内存吗?
- Python 垃圾回收
- 讲一讲 python 内存

5.1.2.3 区别比较

- python 参数 * 和 ** 区别?
- list 和 tuple 区别?
- xrange 和 range 的区别?

5.1.2.4 讲解原理

- python 的 with 什么意思?
- python 的 dict 实现, 哈希表查找的时间复杂度一定是 $O(1)$ 么? 为什么? 怎么解决?
- 如果哈希表发生大量冲突, 怎么解决 (想到了二叉搜索树, 面试官问了解红黑树么)

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 内存相关

- C 的内存对齐，给了几个 struct 计算占用内存？

5.2.1.2 区别比较

- C 和 C++中的 struct 和 class 的区别？
- 介绍面向对象和面向过程的区别？

5.2.1.3 讲解原理

- C++继承、重载、虚函数的相关知识
- C 纯虚函数、虚函数表说一下
- 全局变量，静态全局变量存储位置

5.2.2 手写代码相关

- 给定一个数组，相邻元素之差的绝对值 ≤ 1 ，如[1,2,3,2,2,1,2]，如何快速查找某个数？
- $a=1, 2, b=(1, 2)$ ，问 ab 输出？不用中间变量交换 ab，用异或？

6 操作系统高频问题：数据库&线程&常用命令等

第六节
京东面经
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

- 口述一道 SQL 题，差集？
- 数据库相关，主键和外键的区别？
- 数据库的索引介绍一下（B+树）
- sql 常用关键字的执行顺序

6.2 操作系统方面

6.2.1 TCP 协议相关

- TCP/IP 四层模型
- 计算机网络：TCP 和 UDP 的区别，OSI 七层模型、全双工和单工？

6.2.2 线程和进程相关

- 多线程多进程问题，cpu，磁盘 io 哪种多线程带来效果好？
- 多进程适合处理什么类型的问题？（应该是想让回答 I/O 型？？）适合解决计算复杂的问题么？

6.2.3 常用命令

- Linux 基本命令，说下对管道的理解？
- 强制杀死某个进程的命令、其他的命令？
- Linux 更改文件权限的，那改可执行的是多少？777？详细解释下？
- 查看文件前十行？

7 技术&产品&开放性问题

7.1 技术方面

- 京东有 20 万人，要做通讯录，包括名字和电话，名字可以重复，电话不可以，要

实现快速的增删改查的话，用什么数据结构比较好？

- 假设京东有 1000 台服务器，每台上面有 100g 的日志文件，然后现在要在自己的服务器上统计出现次数最多的 ip 地址（服务器上不能进行统计）？
- 场景题：有用户 feed 流和点击信息，如何做推荐？
- 场景题：每天的用户，商品，销量订单记录，求销量前 10 的商品，用 SQL 或者什么编程语言写一下？
- 场景题：商品销量的时间序列数据如何分解？
- 场景题：解一个在北京地区的不同库房分配某一个商品的混合整数规划问题（要求说出目标函数，限制条件和求解法，也算磕磕碰碰答上来了）
- 场景题：打开京东 APP，点击一个商品，详情页会显示 XXX 也买过，这个具体是怎么做的？(扯到了协同过滤和冷启动)
- 场景题：现在我们有俩个排序模型，分别是 A 和 B，他们分别预测出了对应的排序的结果在我们的 APP 上，我们有真实地用户数据，那么怎么来评估这两个模型的好坏呢？用说出数学公式。
- 场景题：在京东有许多的不同的消费群体，我们如何找到学生群体并把他介绍给我们的用户。（大概就是聚类的详细建模的过程）
- 业务代码：
If ($a < 3$):
If ($b \cdots \cdots$):
If ($c \cdots \cdots$):
If ($d \cdots \cdots$):
If ($e \cdots \cdots$)
If ($a \geq 3$):
If ($b \cdots \cdots$):
If ($c \cdots \cdots$):

If(d…….):

If(e……)

这种 if 语句嵌套太多，条件也太多了，怎么来优化它？（面试官一步一步提示：决策树相关）

● 场景题：京东搜索里，输入一个关键词搜索某件商品，但是现有数据库里没有该关键词，该商品对应的是另一个关键词，这种情况下怎么解决？

7.2 产品方面

● 基于京东的数据（地理位置、活跃度等）来判断黄牛卖家？

7.3 开放性问题

● 如何给一个完全没有接触过机器学习的人介绍机器学习，机器学习是做什么的？

● 推荐岗位相关：你想找推荐，我们是机器学习+组合优化，偏向运筹学，考虑么？