

MLE 帖子总结

积分不够好多帖子都看不了，求加米，谢谢各位！（加米不会扣自己的大米’

纯经验帖子：

<https://www.1point3acres.com/bbs/thread-1068267-1-1.html>

<https://www.1point3acres.com/bbs/thread-1069790-1-1.html>

<https://www.1point3acres.com/bbs/thread-1068145-1-1.html>

<https://www.1point3acres.com/bbs/thread-1066375-1-1.html>

<https://www.1point3acres.com/bbs/thread-1065581-1-1.html>

<https://www.1point3acres.com/bbs/thread-1064289-1-1.html>

<https://www.1point3acres.com/bbs/thread-1063669-1-1.html>

<https://www.1point3acres.com/bbs/thread-1051733-1-1.html>

<https://www.1point3acres.com/bbs/thread-1050829-1-1.html>

<https://www.1point3acres.com/bbs/thread-713903-1-1.html>

<https://www.1point3acres.com/bbs/thread-1070630-1-1.html>

Research面: 大致上有两种, 一种考察你的基础知识, 会问有明确答案的问题, 比如 what's p-value, what are some ways to adapt LLMs to new tasks...

另一种比较天马行空, brainstorming类型。可能会问你基于你过去work的一些 research vision问题, 然后不停深挖问follow up问题, 我觉得research做得好的人这些都不太需要准备, 跟平时开会讨论研究差不多。

<https://www.1point3acres.com/bbs/thread-1070523-1-1.html>

1. xgboost和rf的区别

然后细问rf: how to choose parameters? what are the parameters? how to decide # of variables? feature importance plot中两个相关的var会怎么show up

2. ways to deal with collinearity

logistic: how to deal with too many vars?

PCA: how would the vars look like? dislike or alike? (cosine similarity会是什么呢?) how to decide which variables to choose?

3. optimization

convex SVD adam gradient?

learning rate啥的 我不是很会这部分

4. class imbalance 问题 (举例fraud) 怎么处理

我说dataset metric 和model 然后他细问了一些

5. variance vs bias

他提出了三个情况问variance和bias会怎么样 好像有reduce feature会怎么样 然后加data会怎么样 还有一个很obvious的我给忘了

<https://www.1point3acres.com/bbs/thread-1069392-1-1.html>

1. 深挖项目细节, 如为什么用这个模型, 穿插问了一些transformer相关

2. attention 机制, decoder 模型优点, 输出是什么

3. GRU vs LSTM

4. LLM fine tuning 方法，有无相关经验（楼主無）
广告

5. 不同LLM attention区别？ position embedding区别（这个我不太了解所以也可能挂在这）

6. medium linked list

8.反问

<https://www.1point3acres.com/bbs/thread-1068681-1-1.html>

Practical ML Coding: Implement pos tagging model in Colab: 包括feature engineering, train model, inference function and evaluation。最后问了一下如果不是面试有足够的时间会用什么模型。

<https://www.1point3acres.com/bbs/thread-1068245-1-1.html>

第一轮手撕batch norm, follow up merge conv+BN layer, 还简单聊了相机标定的几个问题

第二轮分析一段training code运行调用CPU/GPU的情况, follow up针对程序优化训练速度, 手撕conv 2d

第三轮是data structure, 自定义一个类, follow up另一个示例类, 对代码进行分析和debug

第四轮是manager面, 手撕residual block, 问了较多object detection的问题以及system design

<https://www.1point3acres.com/bbs/thread-1068145-1-1.html>

时间问了非常多的关于LLM的问题, transformer是啥alignment是啥解释整个training process让你帮公司设计使用AI你会怎么设计LLM limitation是啥你的project在industry怎么apply怎么combine multi-model等等

<https://www.1point3acres.com/bbs/thread-1068100-1-1.html>

面试官会把基本上所有的机器学习模型都给你撸一遍，从Linear Discriminative Analysis到Linear Regression，再到Tree-based models、Clustering Analysis、MLP、RNN、CNN、Transformer，再到各种Optimization和Reinforcement Learning。

<https://www.1point3acres.com/bbs/thread-1067452-1-1.html>

对GPU 的理解，谈谈对Tensorrt, Tensorrt-llm 使用的反馈并聊一下

1st engineering skill 和 c++: 聊了主要是模型如何做inference加速以及实际部署的场景, 比如dynamic batching, continus batching, flash attention, quantizaiton etc...。如何做 profiling 来发现bottleneck, 给了一个场景来分析bottleneck 在哪儿以及如何改进。C++ 基础知识, 虚函数, 异常处理, 如何用C实现c++多态, RAII, 内存管理
2nd problem solving: 给一个寄存器每次往里面加数, 多长时间overflow. 双蛋, 称球.
3rd AI knowledge: 简历deep dive, NLP/BERT, GEMM 如何加速, LLM 和transformer 的一些基础知识, Convolution 如何加速

4th GPU/CPU Pytohon: CPU/GPU 区别, GPU 内存结构, Python 一些高级特性, 比如装饰器, 传参, 内存管理, GPU里面的CUDA thread block 原理, 根据这个来不停拓展问题和以及给出一个场景让你判断问题出在哪儿。其中一个有意思的问题是pytorch 中 pin memory 原理是什么。。

<https://www.1point3acres.com/bbs/thread-1066874-1-1.html>

ML轮中规中矩, 面的组做时间序列, 就讨论了相关的project, 然后考了些简单的ML八股, 写了个cross validation

<https://www.1point3acres.com/bbs/thread-1066225-1-1.html>

就标准ml八股, 大概讲了model的选择, data collection, positive/negative label的选取, feature engineering, evaluation metrics, 中间我根据自己phd的方向讲了一下如

何处理exposure bias

<https://www.1point3acres.com/bbs/thread-1065999-1-1.html>

简单自我介绍之后马上问关于ML的问题:

1. LSTM vs Transformer
以下内容需要积分高于 188 您已经可以浏览
广告
2. Explain Transformer architecture
3. Input and output of Transformer
4. Compare BERT and Transformer
5. Compare FC and Transformer
6. Metrics for classification model
7. How to improve accuracy of search result
8. Code BCE, CE, Softmax, MSE
9. Data needed to train recommendation system

<https://www.1point3acres.com/bbs/thread-1065950-1-1.html>

3/ reading paper and deep diving. 栽到这一轮了。我准备的时候假设是像reading group, 讲给不懂的观众, 但是实际上面试官对paper很熟了, 直接跳过我准备的slides, 问了很多paper里面的缺点, 比如实验设置是否合理, 实验结果是否实用等等。确实没准备好。

1/ coding, 实现top_k llm token decoding.

<https://www.1point3acres.com/bbs/thread-1064928-1-1.html>

- 1, Precision/recall
2. ROC axis
3. Variance vs bias

1. reducing LLM latency at inference time
2. RNN technique to improve coherence and realism
3. The web service process is time-consuming, what system component of architecture
4. off-line to on-line LLM, what change
5. Path sum on 2D board from both top and bottom

<https://www.1point3acres.com/bbs/thread-1064309-1-1.html>

lstm和transformer区别优缺点, gpu utilization。问了之前项目然后深入讨论相关技术。

<https://www.1point3acres.com/bbs/thread-1064268-1-1.html>

DS questions: 给了一个场景然后是两个variable (cost & money), 问怎么进行拟合, 答 linear regression, 计算mse loss。怎么看模型性能以及怎么predict, 答用train, val, test 预测的时候用weight & bias。如果有 outlier 怎么办, 答需要先做data preprocessing 剔除 outliers, 再训练linear regression. 这里问怎么detect outliers, 答看distribution以及看根据数据intervals 有没有outliers (当时想到箱型图但是描述的不好), 后面追问怎么具体做, 对哪个variable 做, 突然有点懵逼, 就改口说要对y/x 这个计算一个值来发现 outliers, 因为x,y过大过小并不算outliers, 只要他们符合线性回归方程 (这个就扯到怎么定义outliers) 这里答的不好。后面问如果几百个features 怎么做, 那就高维线性回归, 不过要考虑overfitting 的问题, 得去做一些feature selection, 以及对feature做 normalize 以及scaling。又问如果不是linear 的关系, 怎么去拟合, 答加一个activation functions类似NN 去拟合非线性关系, 感觉这里答的不好, 面试官应该想问如何去拟合曲线。

<https://www.1point3acres.com/bbs/thread-1064120-1-1.html>

第一轮coding是ml coding, 具体的题目是:

给定一个stubhub用户数据集，其中包含一些缺失值，希望你通过不同的算法来补全缺失值（比如我被考到的利用平均值来补，利用另一列与其有线性关系的特征值补，利用knn模型补，利用cosine similarity来补），需要你对pandas，scikit-learn有比较熟练的使用

<https://www.1point3acres.com/bbs/thread-1063994-1-1.html>

Visa.csv

**交易单号，时间，货币，值，原因...
多行交易记录**

1. 遍历所有String, 转换字符串为符合要求的输出（分行符切分的，逗号分开的string）
2. 过滤掉不valid的记录，比如不是正整数，或者重复的
3. 过滤掉更多不valid的记录（这部分题意没理解明白，没做完，就不说了）
总之，test case很多看不了，所以过不了也不知道是什么原因，不清楚理解对题目意思没有，需要多考虑corner case，时间紧。

<https://www.1point3acres.com/bbs/thread-1063872-1-1.html>

很多transformer相关的问题，利口edit distance

<https://www.1point3acres.com/bbs/thread-1062320-1-1.html>

了ml八股，gradient exploding & vanishing，loss function avg across records和sum records的区别，min batch shuffle和不shuffle的区别，面试官指明不需要text book样的回答，但楼主不像textbook回答是什么意思 lol

<https://www.1point3acres.com/bbs/thread-1062021-1-1.html>

30 min tech interview

1. Talk about yourself, why you chose this field
2. I mentioned GPT. Talk more about GPT

3. Transformer architecture, encoder decoder, QKV, attention
4. Why we don't use encode-decoder model, why we choose decoder? (没回答好, ta解释了一大串有个简短交流, 回头复习一下qaq)
5. Model Fintuning, what is LoRA, what is lora used for
6. If we want to train a model with specific style output, what should we do
7. Continue training and LoRA training's difference, advantages and disadvantages
8. Talk about RAG

<https://www.1point3acres.com/bbs/thread-1061851-1-1.html>

ML fundamental, 问了基本的ML八股比如bias var tradeoff, over/under fitting, 以及广告相关的 (因为楼主面的广告组) feature processing比如embeddings, sequence features等。此轮也问了很多简历projects 的细节。

<https://www.1point3acres.com/bbs/thread-1061662-1-1.html>

一轮技术面: Transformer architecture, attention mechanism, embedding, NLP project walkthrough ; Leetcode - medium

二轮技术面: RNN, Gated RNN -LSTM.

<https://www.1point3acres.com/bbs/thread-1061624-1-1.html>

标准八股例如bias/var tradeoff, overfit/underfit之类。涉及一些deep learning但是考察不深入, 简单准备应该够用。

<https://www.1point3acres.com/bbs/thread-1061536-1-1.html>

1. ML : 介绍attention, self-attention, multihead attention

<https://www.1point3acres.com/bbs/thread-1061416-1-1.html>

vo1的美国小哥人真的很nice，题目也很新颖，就一道题，对手机里所有图片做 clustering，从embedding的options, evaluation, 到contrastive loss，到numpy实现 cos similarity, kmeans，一路主动接话题，给鼓励给hint，唉，可惜我自己坑坑巴巴，八股我是知道的，可是kmeans写不出vectorized的解，l2 norm还少写了 keepdims=true 我都被自己笑死了

<https://www.1point3acres.com/bbs/thread-1061333-1-1.html>

linear regression, xgboost, evaluation metrics, vanishing gradients, initialization

<https://www.1point3acres.com/bbs/thread-1061156-1-1.html>

第三轮是ML practioner，问了如何产生member embedding和pin embedding

<https://www.1point3acres.com/bbs/thread-1061154-1-1.html>

店面第二轮：不是LC题目，implement a tokenizer，需要识别alphanumeric, whitespace和punctuation。code中提供了判别alphanumeric和whitespace和 punctuation的API。

onsite第二轮：现场load一个dataset，用的pandas，需要对数据进行处理，比如说处理 label，解决imbalance的问题，解决missing feature的问题，建议提前熟悉下pandas语法，可以google。

<https://www.1point3acres.com/bbs/thread-1061066-1-1.html>

店面coding是手写K means，ML知识问了一些以前的project，multi-armed bandit, bayesian linear regression 以及a/b testing如何确定sample size。

- NLP depth：从给一堆document分类开始，聊到了context-agnostic embedding 和context-aware coding such as transformer，问了transfomer经典八股文，还追问了如果维度过高怎么办
- Recommendation depth: 这轮我实在记不得问题了，推荐youtube上看王树森的小红书推荐系统课程，能够cover所有了。
- ML breadth：问了如何e2e上线一个模型，问了bias variance trade off，问了决策树，问了bagging和boosting。

<https://www.1point3acres.com/bbs/thread-1060998-1-1.html>

(2) 描述推荐系统中Candidate generator 用的是什么模型?解释模型, 模型时间复杂度。

(3) 描述推荐系统中Ranker 使用什么模型, 解释模型和模型复杂度

释在chatbot项目中使用过的所有模型。并详细询问了NLP模型里模型的数学解析过程。

(3) 完整描述一个推荐系统系统设计和NLP结合的过程 (中间重点询问了feature部分如何获取和设计)

<https://www.1point3acres.com/bbs/thread-1060679-1-1.html>

他考了logestic regression是什么, 为什么要用logestic function, 什么是overfit, underfit, overfit怎么解决, 什么是l1,l2regularaiztion
这些都有答出来。但是他后来问我neural network东西

<https://www.1point3acres.com/bbs/thread-1060602-1-1.html>

问了如何用api来deploy一个llm model, dockerfile 应该有哪些内容, 之后是问了什么是vector embedding ;

<https://www.1point3acres.com/bbs/thread-1057801-1-1.html>

就text encoding, encoder/decoder structure, tasks (loss functions), output sampling, efficient training and inferencing.

就是一个llm的模型, 然后跑过能出结果就行。中间会有一些bug去解决。

<https://www.1point3acres.com/bbs/thread-1055929-1-1.html>

llm basic (opt/llama 中transformer的区别是什么, tokenizer 的几种类型。这两个问题我说我是做 llm 安全, 没有具体区分他们, 然后面试官华人就说 that's fine , 叫我介绍 transformer 是什么, 还有简述tokenizer机制) 和讨论了一些简单的 solution, 面试官 对 research 和 solution 的feedback 感觉还不错

<https://www.1point3acres.com/bbs/thread-1053031-1-1.html>

ML问答 - 被问了几个regression意义不明的问题，实在听不懂问题的考点是啥。
。 ML coding - 实现 k-fold split的function

<https://www.1point3acres.com/bbs/thread-1052607-1-1.html>

Backpropagation

- Describe math first - Implementation - Test case

<https://www.1point3acres.com/bbs/thread-1052250-1-1.html>

第二部分 ml快问快答，三四个问题，decision tree inference time, random restart, cross validation

<https://www.1point3acres.com/bbs/thread-1050390-1-1.html>

所以还考了一些ML 八股，包括transformer结构，T5和GPT区别等等。。

<https://www.1point3acres.com/bbs/thread-1049454-1-1.html>

写一个Decision Tree

<https://www.1point3acres.com/bbs/thread-1049437-1-1.html>

简历里有推荐算法经验，很详细问了推荐算法的经历，deep dive

问了Transformer，手写transformer

最近有什么关于Transformer position encoding的research？

问了archetecture 相关的问题，例如堆和栈是怎么实现的

ML Ops 在production要怎么知道模型需要re-train

<https://www.1point3acres.com/bbs/thread-1048712-1-1.html>

ml八股还记得的有：svm定义，描述一下过程；k-means过程，缺点是什么；知不知道VAE，VAE可以用来干什么

<https://www.1point3acres.com/bbs/thread-1048518-1-1.html>

往nlp各模型/方法的基础知识上靠，细到会问整个transformer结构与原理，以及各bert模型之间的区别。细到问我bert每一层多少个参数

<https://www.1point3acres.com/bbs/thread-1048175-1-1.html>

3.ML coding：给定一堆平面上的点，求到所有点l2距离之和最小的点的坐标。（本能反应是mean，但mean是最小化l2距离平方和而不是距离之和。虽然面试官提示后想了一下马上更正，自此面试官不悦。）

<https://www.1point3acres.com/bbs/thread-1047813-1-1.html>

- **ML coding session**: data pipeline on Colab. Language Modeling. 处理一些json数据很简单，然后找3个bug

- **LLM** Present a Paper. (slides. 15mins + 15mins QA)
- **ML** 八股 on LLM and Transformer. SFT + RLHF. 会问很细，感觉没有hands-on experimce的应该不太能crash这一轮。

<https://www.1point3acres.com/bbs/thread-1047771-1-1.html>

第一轮：在线筛选，手写了一个KMeans，问了几个followup question。目的是为了考验对KMeans的原理理解。

第五轮：ML轮，手写了一个Cross Validation。问了几个非常深入的问题，需要对模型的training有不错的理解。

<https://www.1point3acres.com/bbs/thread-1047359-1-1.html>

deep learning foundation, 和distributed training的快问快答, 整体压力是比较大的, 问的问题也很hard-core, 比如描述multi-GPU dnn training, distribute的是什么, 传回的是什么, NN size double的话traffic怎么变。

Onsite 第二轮: Deep learning foundation, 解释CNN, RNN, Transformer工作原理, convolution具体是啥, attention mechanism具体是啥, L1 L2 regularization, 之类的。

Onsite 第三轮: Algorithm coding, multinomial sampler, 本身不难, 在notebook上要跑通, 要自己写function以验证自己生成的sample distribution符合给的list of probabilities。重点在后面各种discuss在real world如何写这个程序跑的更快。Linear search换binary search这种算法上的优化反而不是他想要的答案, 把call random number generator放到loop外, 或者list换numpy array这种经验性的优化才是面试官想要的, 一边discuss一边在notebook上做各种runtime测试验证是否新的写法可以跑更快。

Onsite 第四轮: Pytorch coding, 给一个colab notebook, 用pytorch按照题目要求写一个简单的NN, 分析dimension变化, 分析一个实例下用什么loss function, 自己写出loss function (这部分不许直接调用pytorch的loss function), 给的部分代码里有小bug要找出, 最后要跑通, 其过程中可以google所需方程的signature, 也可以问面试官, 不要求把pytorch syntax背下来, 但是真没写过的肯定面不过去。这轮面试体验不好是感觉面试官很geek, 会揪着莫名奇妙的点不放, 自己很居高临下的那种感觉。

<https://www.1point3acres.com/bbs/thread-1047123-1-1.html>

然后是coding, 让用python写一个更新股票平均价格的class, 每加入一个新的股价, 都要重新计算平均值。

<https://www.1point3acres.com/bbs/thread-1046638-1-1.html>

round 5: ML fundamental. 从LR到deep learning, 随机问。evaluation metrics问了PR-AUC和ROC-AUC, 怎么处理imbalanced data, 如果选择upsampling minor class, 会对这两个图有什么影响。DL问了dropout, weight initialization, adam vs. RMS prop, etc.

<https://www.1point3acres.com/bbs/thread-1046494-1-1.html>

第二个组电面：implement KNN

<https://www.1point3acres.com/bbs/thread-1045815-1-1.html>

coding 面中奖了，上来让手写transformer attention. 后面又问了一系列细节问题，什么加mask, softmax

<https://www.1point3acres.com/bbs/thread-1042225-1-1.html>

Go through one of my project, and discuss with questions in between to test knowledge. Questions including: regularization tricks, negative transfer, hyperparameter optimization, vision transformer and cnn, low-data scenario optimization. Questions are in the forms of- If aaa happens what will you do? Do you know bbb, what is the insight of it?

Write a pytorch layer with forward and backward function using numpy

<https://www.1point3acres.com/bbs/thread-1070905-1-1.html>

bert的了解之类的，ml问了如何split data，解释llm