

Tencent

2018年3月28日 13:03

2018.3.28 一面 1小时 数据平台部

1. 自我介绍、项目介绍
2. 问项目相关的
 - 实验结果：F1值，R和P，实验的时候没有太记得P和R，面试官都已经准备好纸和笔打算记下来P和R了，我说我不记得了.....
 - 模型
3. tensorflow怎么读入超大量的数据（面试官问题没表述太清楚，最后才大概了解到对方应该是想问当数据量十分大的时候不能一次性load进来所有的数据要怎么办。tf好像有个readdata的模块，他问我了解具体内部实现吗，我说不知道。。。)
4. 算法：
 - 自己定义一个数据结构（我用C++定义了一个单链表，现在想起来都是血和泪，不过作为一个不精通任何一门语言的人，怎么写都是坑。。）
 - 单链表检查是否成环
 - 一个长字符串，每个字符用ASCII码表示，统计每个码出现的次数
5. 梯度更新算法有哪些：
 - SGD, Adam, RMSPROP（忘记咋写了）
 - 计算公式？（不会）
 - Adam还有好几个，比如adadelta，了解吗？（不了解）
6. 说说正则化
 - 作用
 - 有哪些（L1, L2）
 - 这些正则是对什么的约束？输入还是参数还是输出（参数）
 - 对输入有哪些正则？比如有的输入特别大的特别小？（我说归一化）怎么归一？（max-min归一到一个区间orz...）
7. 中途换了一个面试官（一直保持有俩人一起面我）
8. 又问一遍项目，当中一些细节
9. 又问了一遍正则化
 - L1和L2区别
 - 我说L1可以使得某些参数为0，然后他问，那无法求导怎么办？我当时画了那两个L1 L2切等高线的图，然后他就问L1在坐标轴上的点无法求导，怎么处理？（完全不知道，我说参数都为0了还要求导吗。。。反正就不了了之了）
 - 贝叶斯角度解释L1 L2
10. BP求导，但是他简化了很多，只要求某一层对该层W的某一个值的导数，给出了残差（问了半天残差是啥，我不知道这个定义的意思，他直接写给我了就是误差对当前层的输入的导数。。反正挺简化的，链式法则用一次就可以了）
11. 画LSTM内部图

12. 说一下SVM

- 我真的很讨厌这种直接问“说一下XXX”，不知道怎么组织逻辑介绍orz
- 写出函数间隔和几何间隔的公式
- 为什么要引入拉格朗日？（一脸懵逼，还能为什么，不就是要求解这个最优化吗。反正说得有点乱，大概就说了一下后面有对偶问题之类的，后来甜甜说一个是对偶问题可以简化求解，一个是对偶问题可以用核函数，因为把 w 表示成了 x 的内积。我觉得很有道理，也许面试官想让我说核函数来着...
- 拉格朗日最优解的KKT条件

13. tensorflow是不是都是直接调用的API，我说模型还是自己搭的，他说没有自己实现比如BP之类的吗是不是函数都是直接调用的，我说嗯。。。

问了一下对方的情况：

- 我面的是TEG数据平台部
- 部门主要做广告，好像还有推荐？？就是广告点击率预测，所以如果是文本广告的话需要NLP理解文本意思
- 和TEG其他做NLP的区别：他们是按业务划分的，别的部门也有新闻推荐blabla的具体忘记了

总之这次面试十分速战速决，大家加油(ง •_•)ง