机器学习算法面经---2018秋招

**职位：机器学习/数据挖掘算法工程师**

在牛客网得到很多宝贵的面试经验

现我也来开帖，分享一下到目前为止的面试经验~~~

**本人会一直写到秋招结束，记录下参加的每场面试**

**如果回答上有什么错误，请不吝赐教哈！！！ 谢谢~~~~**

**目前内推面了：阿里（一面跪）、京东（offer）、拼多多（录用意向书）、美丽联合（一面跪）、链家（offer）、美团点评（估计三面跪）**

**目前校招面了：360（一面跪）、百度（估计一面跪）、三星研究所**

**据说可以攒人品~~~~~**

更新线-2017-09-22 17:56---------------------------------------------------------------------------

**刚去导员办公室领了三方协议。感觉秋招快要结束了。**

唉，腾讯也没霸面上。。。

不过得提一下，三星研究所的效率真的挺高的。刚才打电话说了薪资，让签约。（也算是对周二面试有个交代吧）

效率真心高，而且笔试面试还有福利，送卡管午饭~~~

更新线-2017-09-22 10:14---------------------------------------------------------------------------

尴尬了，前几天犯懒了，没及时更新百度的一面，直到。。。现在还没有收到下一面通知，应该跪掉了。这周二又面了三星研究所。

昨天傍晚，当小伙伴们都焦急的等待华为结果时，我很淡定（**因为华为连面试机会都没有给我**）。结果收到了意外之喜。

拼多多居然发意向书了，（这公司再不发，都忘了有这一茬了）。毕竟是对自己面试的肯定，还是挺开心的。

秋招快要结束了，霸面腾讯不知道还能不能安排面试，下午领三方~~~

话不多说，百度一面，三星两面奉上。

更新线-2017-09-18 20:54---------------------------------------------------------------------------

**上周收到了京东的offer。虽然之前有面试官的口头offer，但是正式offer下来，才安心嘛。**

**接到offer还是很震惊的。出乎人意料啊。**

**然后，晚上就吃了麻辣香锅~~~**

**犒劳自己~**

**昨天面了百度，第一轮。这是迄今为止经历过时间最长的一轮面试。足足整了1个半小时。**

**把我都快榨干了，会的几乎全都被问了，不会的也被问了。**

**不过确实学到很多东西，之前很多东西没有实际做过，只看理论，只是知其然，经过面试官的讲解才明白其所以然。**

**先去洗澡，一会回来更新~**

更新线-2017-09-12 22:37---------------------------------------------------------------------------

**上周末收到了链家的offer。还是蛮开心的。毕竟是准备这几个月以来的第一个结果。**

**当天晚上，就给自己加了一个鸡腿~~~**

**\*\*\*继续努力，要让自己的能力尽可能接近用人单位所开出的工资的价值\*\*\***

**昨天早上搜狗一面，下午美团三面**

**昨晚百度做崩了。。。有道编程题脑袋秀逗了，搞错了**

**刚刚做了腾讯的，感觉真是画风新奇啊。跟其他公司那些浓妆艳抹的风格不一样啊。**

**没有编程，没有编程，没有编程**

**还有个求行列式的，略微尴尬啊**

更新线-2017-09-07 16:56---------------------------------------------------------------------------

**早上美团点评二面**

更新线-2017-09-05 20:24---------------------------------------------------------------------------

因为要面美团，所以在牛客上找面经。

**感觉应该把这几天面的写下来，不能等秋招结束再写了，那样找面经的同学（比如说我~）不就少了一篇面经可看嘛**

**因为时间紧迫，更新的问题没有详细解答。后面再来补**

**更新：链家二面、链家HR面、拼多多二面、拼多多HR面、京东HR面、360 一面、美团点评一面**

更新线-2017-08-29 14:06---------------------------------------------------------------------------

**记录一下链家的面试，本来准备把几面都写完。但发现写完一面，有种精疲力尽的感觉。。。**

这是我目前为止经历过时间最长的面试。约1个小时多一点点。

后面再写二面、三面吧。一面够长的了。

更新线-2017-08-27 20:10---------------------------------------------------------------------------

感谢各位大佬捧场，一会记录下今天的面试~~~

西安南雷村职业技术学院

---一个散养的没导师的硕。。。

以前玩大数据，现在玩机器学习，刚开始玩。

开八：

-----------------------------------------------------------------------------------------------------------

**2017-08-10-阿里菜鸟-机器学习-内推-1面-电话**

没有自我介绍。。。直接略过。。。

1、讲一下你觉得你突出的地方，有亮点的地方。

说了SVM和LR

2、LR为什么用sigmoid函数。这个函数有什么优点和缺点？为什么不用其他函数？

3、SVM原问题和对偶问题关系？

4、KKT条件用哪些，完整描述

5、说项目

6、有一堆已经分好的词，如何去发现新的词？

面试官给的提示：用这个词和左右词的关系。**互信息** 新词的左右比较丰富，有的老词的左右也比较丰富。还要区分出新词和老词。

7、L1正则为什么可以把系数压缩成0，坐标下降法的具体实现细节

8、spark原理

9、spark Executor memory 给16G  executor core 给2个。问每个core分配多少内存

面试官那边全程嘈杂，不知道在干啥，讨论问题？？

-----------------------------------------------------------------------------------------------------------

**2017-08-11-京东广告数据部-机器学习-内推1面-电话**

1、自我介绍

2、说一下进程和线程

说一大堆，再就说之间的区别

3、线程安全的理解

4、有哪些线程安全的函数

5、数据库中主键、索引和外键。以及作用

一个表可以没有主键，可以有索引

6、说项目

7、Spark原理

8、Spark是多线程模式，怎么退化为多进程模式。

在每个executor core设置为1，即每个executor是单线程的。

9、撸代码。实现一个java迭代器

数据：

int[][] data = new int[][] {

null,

new int[] {1,2,3,4},

new int[] {},

null,

new int[] {5,6,7},

new int[] {8},

null,

};

要求：遍历是跳过NULL。依次遍历每个元素：1,2,3,4,5,6,7,8

提示：

hasNext里面不应该改变迭代器内部状态，hashNext只判断

next返回值，并且指向下一个有效元素。

P.S. 面试官很忙，在我写代码的时候。还在跟另一个候选人约时间~~~

-----------------------------------------------------------------------------------------------------------

**2017-08-11-京东广告数据部-机器学习-内推2面-电话** **一面、二面连着玩~~~**

1、自我介绍

2、对于机器学习你都学了哪些？讲一个印象深的

说了SVM原理，拉格朗日法，对偶问题，以及好处。

3、SVM怎么防止过拟合

说了SVM里面的松弛变量。不知道对不对

4、我主动出击，有另一大类算法决策树，说不管是LR还是SVM都不能直观的感受到决策依据。而决策树易于理解，能够直观的感受到决策依据。

说了划分依据：信息增益（说了信息熵的来源，等概率时熵最大）、信息增益率、基尼系数。

说了划分方法（基于信息增益的）

说了C4.5比较ID3的优点。

5、决策树如何防止过拟合

剪枝，前剪枝和后剪枝。说了REP剪枝。C4.5是悲观剪枝

6、项目没问，说从上位面试官了解了。

7、撸代码

求连续子数组最大乘积，还让考虑边界问题（最后问了：连乘有可能导致溢出，存不下了）

-----------------------------------------------------------------------------------------------------------

**2017-08-15-拼多多-算法-内推1面-电话**

1、自我介绍

2、介绍项目

3、项目延展题：电商搜索框，每天有500W的搜索query。针对新来的一个query，给出和它最相似的100个query。

如果用RNN分类模型表征，那么向量不应该用最后一层的分类特征。应该用倒数第二层的更纯的特征。

现在假设500W的query已经是向量了。如何和这一个query比较。全部算距离不行，开销太大。

**应该怎么办？？？**

4、K-means聚类个数选择，做什么样的试验来确定K

5、两个4G的文件（每个文件可能有重复），里面全都是数字。现有内存1G，求这两个文件的交集。

2个4G的文件，分别hash成10个子文件，一个400M。

把一个子文件存储到hash表中，作为key。遍历另一个文件，看这个数字是否存在于刚才的hash表中。存在即可输出。

-----------------------------------------------------------------------------------------------------------

**2017-08-23-美丽联合-算法-内推1面-电话**

1、自我介绍

2、介绍项目

3、说了SVM

4、为什么要把原问题转换为对偶问题？

因为原问题是凸二次规划问题，转换为对偶问题更加高效。

5、为什么求解对偶问题更加高效？

我答了，因为只用求解alpha系数，而alpha系数只有支持向量才非0，其他全部为0.

6、alpha系数有多少个？

我答了：样本点的个数

7、避免过拟合的方法

答了：决策树剪枝、L2正则和L1正则

8、为什么L1正则可以实现参数稀疏，而L2正则不可以？

答了：L1正则因为是绝对值形式，很多系数被压缩为0,。而L2正则是很多系数被压迫到接近于0，而不是0

9、为什么L1很多系数可以被压缩为0，L2是被压缩至接近于0？

答了：图像上，L1正则是正方形，L2正则是圆形。

L1正则的往往取到正方形顶点，即有很多参数为0

L2正则往往去不到圆形和参数线的交点，即很多分量被压缩到接近于0

**哪位大佬知道哪里有L1、L2的实现代码？？？，求告知~~~~~~~**

10、问平时用啥语言比较多？

说了之前用java、scala多。现在用python较多。

11、问jvm 啥啥啥（没听清）。。。

答：不会

12、python...直接问你个开发中的实际问题吧，如果写的程序跑的非常慢，多方面分析这个问题？

答了： 1、检查程序是否有多层嵌套循环，优化

2、检查程序是否有很耗时的操作，看能否优化为多线程并行执行

3、检查数据量是否非常大，考虑是否可以用分布式计算模型。

**求大佬补充~~~~~~~~~~**

13、SQL中inner join 和outer join的区别？

14、试图给他说说SPARK，结果被严词拒绝（开玩笑的）。。。说时间紧迫，还是他来问吧。。。

15、Kmeans中，现在给你n个样本点不在欧式空间中，无法度量距离。现在给了一个函数F，可以衡量任意两个样本点的相似度。请问Kmeans如何操作？

答：想了一会，比如K=4的聚类。

1、首先，随机去4个点，作为初始类簇中心。

2、计算所有样本点与这4个点的F相似度。根据相似程度，把所有样本点分到4个类中。

3、在这4个类中，计算每一个样本点 i 到该类其他样本点的相似度和Si。取Si最大的那个点作为这个类的中心。

4、重复2、3步骤，直到类中心不再变化或者循环次数达到目标。

-----------------------------------------------------------------------------------------------------------

**2017-08-27-链家-算法-内推1面-现场**

**来了之后先做1个小时的题，5道算法题**

**因为比较长，所以采用 A：面试官   B：本人**

B 自我介绍

A 你自己学机器学习，怎么学的？

B 自己看书，周志华的西瓜书、机器学习实战。先找着撸代码，然后去深究里面的理论。

A西瓜书看到什么程度？

B刚开始看，看不太懂，然后就以机器学习实战为主，先照着撸代码，然后去西瓜书里深究里面的理论。

B我给您说说SVM吧，自学的时候留下很深的印象（试图抓住主动权~）

SVM是基于。。。说着手动起来写SVM的损失函数

A （打断）为什么样本点到决策面是 1/||w||

B 手推向量点到决策面的表达式（麻蛋，竟然一时紧张忘了。。。没推出来）

A 点到直线距离公式记得吧？

B 嗯嗯，又没写出来。只能说之前推过，现在一紧张忘了。。。

A 这个也无关紧要，继续

B 继续说SVM

A （打断）知道LR吧，知道LR和SVM有什么不同吗？

B 知道，首先这两个算法的分类思想不同，LR是基于概率推导的，SVM是基于最大化几何间隔的

A （打断）写一下，LR的损失函数

B 手写出来。其实这个sigmoid函数由那个什么族分布（真的忘了名字，其实是：指数族分布），加上二项分布导出来的。损失函数是由最大似然估计求出的。

A 怎么由最大似然估计导出的？推导一下

B 最大似然估计就是求让已知事件发生的概率最大的参数。

假设有5个样本，每一个的类别是yi，由LR计算出的概率是h(x)。那么每一个样本预测正确的概率为：

(H(x)^yi)\*((1-h(x))^(1-yi)) ----

（刚开始一紧张，把h(x)和yi写反了）面试官说是这样吗？你这样全为0，我感觉你在背公式。。。你再看看

我一看，卧槽这竟然写错了。赶紧改过来，然后表明是自己紧张了。

概率连乘后，然后取对数就是LR的损失函数了。

A 为什么损失函数有个负号？

B 这是因为要应用梯度下降法，引入的。不加负号也可以，梯度上升法。这都是一样的。

A OK，继续，LR和SVM有什么区别？

B SVM决策面只由少量的支持向量决定，而LR的话是所有样本都会参与决策面的更新。

A 对，所以说SVM怎么样？

B SVM对于异常点不敏感，而LR敏感。SVM更加健壮，决策面不受非支持向量影响。

A OK

A 知道过拟合吧？

B 知道，在训练集表现好，在测试集表现一塌糊涂。举个例子就是：学生平时考试成绩非常棒，但一到实际应用就很烂。

A 说说常见的过拟合的解决办法

B 数据，样本不够，如果现在的训练集只是所有样本空间的一个小小的部分，那么这个模型的泛化能力就非常差（边画图，边说）

A 嗯嗯，还有呢

B 可以加正则项，L1，L2正则。L1还可以用来选择特征

A 为什么L1可以用来选择特征

B 因为L1的话会把某些不重要的特征压缩为0

A 为什么L1可以把某些特征压缩为0

B 因为（画图）L1约束是正方形的，经验损失最有可能和L1的正方形的顶点相交，L1比较有棱角。所以可以把某些特征压缩为0

A 还有什么过拟合的解决方法

B 神经网络中，dropout方法。就是每层网络的训练，随机的让一半神经元不工作。达到防止过拟合的目的

A 还有吗？

B 决策树中可以用剪枝操作。

B 决策树过拟合，可以用随机森林。。。

A 什么？？？现在一个决策树已经过拟合了，还要再以它为基准训练随机森林？

B 。。。对，你说的对。我想错了。。。

B 我就知道这些方法了。。。

A OK，挑一个项目给我说说吧

B 说项目（不记得中间有没有再提问了。。。）

B 要不我给您说说spark框架吧，之前还用的挺多。

A 嗯（看简历和笔试题中。。。）

B 开始说。。。说到三分之一

A 好了！ 你不必说了。（大手一挥~）我看你5道笔试题都没写思路，现在把第二题代码写出来

注： 第二题就是检测括号是否匹配

B 我写了啊。。。（给他翻到其中一个的背面）

A 哦，（迅速扫过代码，），为什要把字符压栈呢？不压栈也可以的。

B 是吗？{abc()}这样的也是合法的吗？

A 当然啊（看了一眼题。）

B 好吧，我本来也准备看到字符就丢到，不入栈。但担心这种情况不合法，就给入栈了。

A 嗯，第三题呢？

B 没思路，没写

A 给我说说第四题

第四题：10分钟内，恶意IP访问检测（10分钟内访问次数超过1024即为恶意访问）

B 这是10分钟动态检测的，需要时间刻度精确到秒吗？

A 不需要

B 把10分钟内的<ip,次数>存入hashmap, 再把key,value互换存入treemap。因为treemap是基于key有序的，升序。然后直接拿出来最后一个和1024比较。

A 怎么实现动态的检测，当前检测0-10分钟，那么第11分钟怎么办？

B 把0-10分钟的摘出来，从10分钟内的hashmap中减去，再把10-11分钟内的加上。

我知道这样实现起来，效率应该不高，但这一会我只想到了这个。。。

A 嗯，其实可以这样，把每分钟的分开存储，动态的向后移动，取这10个的总的数据就行。

甚至可以每分钟只存储TOP200的，然后10个分钟的汇总，取TOP1

B 嗯，明白了。

A 说说循环依赖这个怎么解决的？

第五题：系统有很多相互依赖的包，怎么检测循环依赖

B 把它当做一个链表。记录当前的名字在hashset中。如果某一次遍历的依赖名字存在于这个hashet中。就认为有循环依赖。

A 学过数据结构吧？学过图吧？给你一个有向图，怎么检测有环？

B 维护一个访问的数组，记录哪些点被访问过，从一点开始遍历，如果遍历的点被访问过，就说明有环

A 从哪个店开始遍历？

B 从入度为0的点开始遍历

A 如果有多个入度为0的点呢？

B 嗯。。。都要以它为入口开始遍历。

A show me the code!!!

（我内心是崩溃的。。。）

B 纠结了一会，又给他说了一遍思路。

A 嗯，好吧，我没有什么想问的了。你呢？

B 请问您说的这个图的这个应该怎么。。。算了，我还是下去自己看吧。。。我还是想知道怎么解决。。。

A 你说的对啊，就把思路给我讲了一下，和我的差不多。

B 贵公司这里机器学习、深度学习有什么应用场景呢？

A 房屋估价啊什么的。

B 好的，谢谢。再次感谢，离开。

-----------------------------------------------------------------------------------------------------------

**2017-08-27-链家-算法-内推2面-现场**

1、自我介绍

2、之前写过spark？写过统计日志用户数？那手写一下统计用户数（scala手写）

3、项目中用到了聚类？手写一下Kmeans

4、一般工业界不这样用，用kd-tree加速

5、给你出道题写一下，一个文件每一行有3列（\t分隔），每个字符串是abcd，这种形式，中间有大写有小写。

现认为：abcDe 等于BcaDe （即：不区分大小写，无关顺序）

要求输出： 字符 空格 出现次数 空格 每一种字符（以|分隔）

实例输出： abcde 2 abcDe|BcaDe

6、记不得了。。。好像没了。。。

-----------------------------------------------------------------------------------------------------------

**2017-08-27-链家-算法-内推HR面-现场**

**这个后面再补吧，无关技术**

-----------------------------------------------------------------------------------------------------------

**2017-08-29-拼多多-算法-2面-内推-电话**

1、自我介绍

2、将项目

3、说SVM

4、好像还说了spark原理

**5、电话中断，面试官线上有BUG，去改BUG了。。。**

6、10分钟后电话来了

7、我主动说：我给您说一下决策树方面的吧

8、面试官说：不用了，来道题。。。

9、一个矩阵都是0,1 且每一行，0都在1前面。求1个数最多的那一行的序号

-----------------------------------------------------------------------------------------------------------

**2017-08-30-京东-机器学习-内推HR面-现场**

**这个后面再补吧，无关技术**

-----------------------------------------------------------------------------------------------------------

**2017-08-31-拼多多-算法-HR面-内推-电话**

**这个后面再补吧，无关技术**

-----------------------------------------------------------------------------------------------------------

**2017-09-04-360-大数据算法-1面-内推-视频**

1、自我介绍

2、说项目

3、说一下项目中用的Kmeans算法

4、知道哪几种聚类算法，说下原理

5、Kmeans有什么优缺点

6、项目用了RNN，说一下RNN原理

说了RNN原理，顺便说了LSTM/GRU的出现

7、为什么会出现长时依赖的问题

8、LSTM/GRU如何解决长时依赖的问题

9、写代码：

一个有序数组中查找某个数

一开始写了个遍历查找，面试官说，还能再快吗？

然后写了个二分查找

-----------------------------------------------------------------------------------------------------------

**2017-09-05-美团点评-机器学习-1面-内推-电话**

1、自我介绍

2、说项目

3、打断，问个扩展题：问答系统，有200W个FAQ，如何用分类模型做分类

思考ing，面试官提示：了解搜索引擎吗？

用倒排索引，把FAQ的问题分词，每个词对应多个FAQ。新来的query分词，每个词对应的FAQ拉出来。再在这个里面做分类。

4、继续说项目

5、说一下hadoop重要的2点

说shuffle，说map、reduce分别分配资源，可以细粒度控制资源占用情况，有利于超大任务平稳正常运行。

6、面试官说，其实是HDFS，正是由于有了分布式文件系统，才可以分布式计算

对，分布式文件系统。数据在哪里计算就在哪里，移动数据变成了移动计算。更高效

7、做题

给定二叉树前序、中序遍历结果。求后序遍历结果

8、一维空间中，2个线段，a1-b1 和a2-b2。判断是否两个线段有交集

他想要的答案是：一个线段里面的大坐标，小于等于另一个线段里面的小坐标。

-----------------------------------------------------------------------------------------------------------

**2017-09-07-美团点评-机器学习-2面-内推-电话**

1、自我介绍

2、说项目

3、用RNN了，说一下原理

说RNN，顺便说了长时依赖问题，介绍了LSTM，GRU

4、说情感分析的项目

5、每个句子都被打上标签正向或者负向情感，如果我想得出句子中的每个词的情感倾向，怎么做？

我不清楚该怎么做，就如下扯乎：

认为每个句子的情感倾向由每个词的情感倾向打分相加而得。

有的词正向：+1，+2，+3...

有的词负向：-1，-2，-3...

经过RNN，每一时刻的输出。。。扯完我现在想都想不通了。。。

**后来想了想可以用贝叶斯分类。不知道对不对，还请大佬指正啊~~~~~**

6、情感分析里用了SVM，说一下

说SVM，顺便跟LR对比一下

7、还知道其他分类算法吗

嗯嗯，知道，说了决策树，ID3，C4.5，再扯了扯bagging和boosting

8、做题

数轴上从左到右有n各点a[0], a[1], ……,a[n -1]，给定一根长度为L的绳子，求绳子最多能覆盖其中的几个点。要求时间复杂度O(n)，空间复杂度O(1)

-----------------------------------------------------------------------------------------------------------

**2017-09-11-美团点评-机器学习-3面-内推-电话**

1、自我介绍

2、说项目

3、场景题：一个景点有很多信息，位置、门票、类型等等。设计一个知识图谱。这个事情如果交给你来做，你会怎么推进

当时就一脸懵逼，只听过这个东西。没研究过。。。就硬着头皮瞎掰

4、我给介绍了SVM

5、你这机器学习这块，只学了这几个月。你认为你有什么优势能跟其他这个专业的人竞争？

麻蛋。。。确实没想过这个问题，继续瞎掰

6、又是场景题：有100亿网页，每个网页都有一个标签。有可能一个标签对应上百万标签，有的标签只对应几个标签。要做一个数据去重，每个标签只要1个网页。

7、工作中遇到了什么实际的难点问题，怎么解决的？

**面试官是一个和蔼的秃顶大叔，估计是总监级别。问的问题就是有深度，考察解决问题能力**

-----------------------------------------------------------------------------------------------------------

**2017-09-11-搜狗-机器学习-1面-校招-现场**

1、自我介绍

2、说项目

3、用RNN了，说一下原理

4、问RNN怎么训练的？

大概说了说，BPTT。这个不太懂

5、RNN的输入是什么呢？

有word2vec训练的词向量库，一个句子分词后，把词都换成对应的向量输入

6、继续说项目

7、项目用到聚类了？介绍一下

巴拉巴拉巴拉

8、说文本情感分类项目，文本向量用tf-idf这种有什么问题没有？

有，不能捕获到上下文之间的联系。以后尝试用doc2vec这种。

9、了解bagging和boosting吗？

巴拉巴拉

10、做题

1、全排列

2、数组第k大的数

3、数组左减右，求最大差

4、树的路径和

-----------------------------------------------------------------------------------------------------------

**2017-09-17-百度-机器学习-1面-校招-现场**

因为笔试做的比较烂，所以以为没有面试机会。于是去了现场企图霸面。结果小姐姐当场一查，居然我也在面试名单，是还没有通知。于是回去，愉快的等待第二天面试。

因为较长 A:面试官    B:本人

A你是上午最后一个，咱们可以多聊一会（内心是崩溃的。。。）

A（原本以为要套路的自我介绍。。。结果。。。）你用C++多吗？

B不多，用java、python较多

A那STL熟悉吗？

B不熟

A知道Trie树吗？

B不熟，一顿扯

A详细讲解Trie树。

A红黑树了解吗？（据说让搞红黑树就要挂，难道这里就是预兆？？？）

B说了说5个特性

A详细讲解红黑树、B树、2-3-4树。

B（一脸谦虚的认真听）

A来做道题

一副扑克牌，未拆封，是有序的排列。要给4个人发牌，要使发的每一张牌的概率相同。

即发第i次牌，发出10和发出2的概率要相等。

B想了想，问了问，说了说思路，谈论了一下。

A排序算法知道哪些

B巴拉巴拉

A快排了解吗？

B说了思想，说了如何划分集合。

A知道快排的非递归实现吗？

B不了解

A那写个mergeSort吧，规定要写代码的。

B一会就搞定了

A咱们问问机器学习吧

A随机森林了解吗？Bagging和boosting了解吗？

B介绍随机森林

A **RF的话，如果有一个特征和标签特别强相关。选择划分特征时，如果不随机的从所用特征中随机取一些特征的话，那么每一次那个强相关特征都会被选取。那么每个数都会是一样的。这就是随机森林随机选取一些特征的作用，让某些树，不选这个强相关特征。**

B搜嘎。。。

A知道为什么bagging吗？

B。。。

A bootstrap aggregating

B又介绍了boosting

A说说这个项目吧

B巴拉巴拉

A看你项目用了SVM，介绍一下

B巴拉巴拉（中间被打断）

A你们怎么过来说的都很像啊，你们都看什么书？

B。。。我看的周志华的西瓜书和李航的统计学习方法。。。

A继续

B。。。巴拉巴拉

A还用到了RNN，介绍一下

B巴拉巴拉

-----------------------------------------------------------------------------------------------------------

**2017-09-19-三星研究所-机器学习-1 2 面-校招-现场**

笔试：

早上笔试，一道题，3小时。。。

其实不是考编程，是考英语。。。

看题1小时，做题10分钟。。。

**结束后，发了一张公交卡（32元）~管了一顿泡菜料理。。。**

HR面试：

略。。。

技术面试：

1、介绍项目

2、介绍RNN

3、Python如何定义一个私有变量

4、Java多线程start和run方法的区别

5、Java hashmap和hashtable的区别

西安三星电子研究所说有关机器学习的有存储SSD方面的、物联网平台方面的