各种Linux命令、spark的一些知识、Svm需要调的参数有哪些，过拟合了要怎么调

随机森林介绍一下

二面：

Python字典怎么实现的、爬虫怎么实现的？有什么困难么？直接用URL爬？

遇到反爬虫机制怎么办

分词原理、字典树的缺陷

讲word2vec，输入输出，怎么优化

CNN激励函数 优点

简单：斐波那契数列，返回数组；字符串转制，复杂度越小越好

分类准确度你满意么？为什么？业务中应该达到多少？能不能达到100%？

精确度和召回率

布隆过滤器作用

map查找的时间复杂度

你知道哪些语言里map的实现方式

讲一下你自己的优势

有什么问题

一.语言方面

1. 函数式编程和面向对象编程的区别？

只凭印象硬着头皮回答了出了在函数式编程里面函数是一等公民，然后支支吾吾半天，善解人意的面试官就让我停住了，说知道我这方面的深度了。

2. python水平怎么样？

我说会用但不熟，自己主要写java和c的。

3. java里面的Object里有哪些方法？

这个还我很熟呀，立马答道equals、hashcode、toString、wait、notify、notifyAll，面试官说行了，然后又打住了。盼望着多来几个java的题目，可面试官转移到我的项目了。

二.项目

1. 介绍下这个J2EE的项目？你在里面主要负责哪一块？

springMVC，spring-core,IOC,aop,orm,hibernate, blabla。。。。，说了一大堆，但感觉面试官对这个项目不是很感兴趣，也许只是简单想要看下自己能否表达清楚吧。

2. 介绍下你的这个建模大赛题目？用到了啥算法？解释下原理？

选的是一个场景特征提取问题，里面主要用到了PCA（主成分分析法），然后讲了一下实现过程，以及是怎样降维的。面试官听完后说我原理方面理解不是很头侧，然后给我讲了一下。(★ ω ★)，面试官还是很耐斯的。

3. 介绍下离散傅里叶变换呗？分析下它的时间复杂度，以及如何改进？

(￣▽￣)"，可算是遇到行家啦。dft基的正交性。O（n2)的复杂度，采用fft分组计算，蝶形算子可以降低时间复杂度。

三. 机器学习和大数据处理

1. spark的rdd，和hadoop的区别，在什么情况spark会优于hadoop？

一年前用过的spark，忘得差不多了。但还是说了一点，内存计算，减少io开销，transform和action等。

2. 知道overfitting吗？linear regression的l1、l2正则有啥区别？

overfitting过拟合还是知道的，本来想说cross-validation可以避免过拟合的，但当时一直没想起来，更加具体的也不知道啦( ▼-▼ )关于l1、l2正则，自己只知道其中一个可以得到稀疏的解。然后也就这样啦！

四. 算法与数据结构

1. 知道快速排序吗？讲下原理？时间复杂度？以及如何选取pivot？

主要是选取pivot上答得不是很好，只给出了一个随机选取pivot的方法。然后耐斯的面试又给我上了一课o(^▽^)o

整场面试主要靠回忆，其实难度倒是不大，考的面还是很广的。

<https://www.nowcoder.com/discuss/82161>