

## 面试不会的问题

一、一个文件有1000万行，每行一组两个数字，比如 (1, 4) (2, 3)，如果两组有一个数字相同，则整合成一组，比如 (1, 2) (1, 5)  $\Rightarrow$  (1, 2, 5)，设计算法，计算文件中共有多少个组，每个组的数据？

可以通过并查集去实现，关于并查集的具体说法[http://blog.csdn.net/dm\\_vincent/article/details/7655764](http://blog.csdn.net/dm_vincent/article/details/7655764)

二、有N个不同颜色的小球，有放回的抽样，抽取N次，M表示抽取N次之后颜色不重复的数目，求解M的数学期望

三、求解最大递增子序列

利用动态规划：

$L[i]$  以i结尾的为前面全部比他小的最大值+1

四、RBF核函数（高斯核函数）

五、判断[1, N]个数字是不是N的全排列

[http://blog.sina.com.cn/s/blog\\_6be26cc7010115hg.html](http://blog.sina.com.cn/s/blog_6be26cc7010115hg.html)