算法（不含深度学习）

一、SVM

SVM的定义描述、SVM解决多分类问题、SVM公式的推导、SVM的对偶（利用拉格朗日对偶）形式的优点、SVM的核函数、SVM的损失函数（合页损失函数）

资源：

详细介绍SVM的文章（July大神的文章，对SVM的起源、推导、证明以及逻辑回归都有介绍）

http://blog.csdn.net/v_july_v/article/details/7624837

二、逻辑回归（LR）

逻辑回归的损失函数，要会推导逻辑回归的求解表达式

http://blog.csdn.net/pakko/article/details/37878837

三、牛顿法和梯度下降

牛顿法的基本介绍 http://blog.csdn.net/luoleicn/article/details/6527049

牛顿法与梯度下降法对比 http://www.myexception.cn/cloud/1987100.html

梯度下降直接求解函数的最值，而牛顿法求解的是函数的一阶导数等于0的值，通常情况下牛顿法比随机梯度法收敛更快，但是牛顿法计算比较复杂

二、布隆过滤器

解决面试时的大量数据判断是否重复问题---百度搜索每天会搜索很多内容，如何判断某个内容是否被搜索过----此时就可以使用布隆过滤器。

BloomFilter介绍：http://blog.csdn.net/dadoneo/article/details/6847481

三、推荐系统

推荐系统分类：协同过滤、基于内容、基于知识、混合推荐

协同过滤 http://www.cnblogs.com/luchen927/archive/2012/02/01/2325360.html

基于内容 http://www.cnblogs.com/breezedeus/archive/2012/04/10/2440488.html

基于知识 http://blog.csdn.net/puma_dong/article/details/42001683

三、查找

1、二分查找

二分查找主要有两种，一种是数组，即排好序的数组，这个查找速度永远是logN的，但是在插入或者删除的过程中，代价相当大（具体参照插入排序算法）

第二种是二叉查找树，插入代价和删除代价特别小，但是容易造成树畸形，即不平衡，造成查找速度远远达不到logN。改进二叉查找树是红黑树，此树是平衡的二叉树

四、排序

1、堆排序

堆排序算法是通过构建最大最小堆，并且不断删除最大最小数据来完成目的的。

堆排序中最重要的两个步骤一个是下滑（就是假设子树都是堆，把父节点下滑到合适位置，在删除节点和构建堆的过程中会调用该方法），另一个是上进（就是整个树已经是堆，我把一个新节点添加到堆的顶部，通过不断的上进完成要求）。

堆中的二叉树是完全二叉树，因此，在保证了排序速度跟归并、快排相同情况（N*logN）下，并不占用额外内存。
该文章简单介绍了堆：http://blog.csdn.net/morewindows/article/details/6709644/

2、归并排序
归并排序的思想是典型的分而治之，每次对两个排好序的数组执行归并，归并排序需要一个额外的数组
归并排序的具体介绍：http://blog.csdn.net/morewindows/article/details/6678165/

3、快排
最经典的排序，每次选择一个中间值，把大于他的放到左边，把小于他的放到右边，当数据是排好序的时候，快排自身的效果最差。因为，我每次排序的都是一个数组，而非两个数组
快排具体介绍：http://blog.csdn.net/morewindows/article/details/6684558

五、字符串相关
1、回文
判断字符串是否是回文（两种方法，从中间向两边遍历或者从两边向中间遍历）
查找最大回文子串（暴力破解+动态规划）
http://blog.csdn.net/kangroger/article/details/37742639

六、动态规划
1、求解最大子序列和
http://www.tuicool.com/articles/NfmyIf

2、查找一个数组中和为sum的三个数或者四个数
http://blog.csdn.net/qq_26437925/article/details/52787136

3、背包问题
http://blog.csdn.net/mu399/article/details/7722810

七、树与图
1、树的遍历
树的前序、中序、后序遍历，非递归实现 http://www.2cto.com/kf/201407/314705.html

八、其他
1、LDA是主题文档模型（判决两个文档的相似度。比tf-idf要高端）
关于LDA的简单介绍（是什么）：http://blog.csdn.net/huagong_adu/article/details/7937616
关于LDA的具体推导（July大神）：http://blog.csdn.net/v_july_v/article/details/41209515?utm_source=tuicool&utm_medium=referral

九、如何解决样本不平衡的问题
（1）进行上下采样
（2）正负样本均衡程度实在很差的情况下可以考虑一分类问题（one-class SVM）

十、线性回归
（1）线性回归为什么要用平方损失，其实是假设产生的误差符合高斯分布，目标是让误差的极大似然估计值，均值和方

差均为0，结果就是平方损失。

http://blog.csdn.net/saltriver/article/details/57544704