

北京大学信息科学技术学院考试试卷

考试科目：算法设计与分析 考试时间：2013年6月17日

姓名：_____学号：_____

任课教师：肖臻

题号	一	二	三	四	五	六	七	八	九	十	总分
分数											
阅卷人											

北京大学考场纪律

1、考生进入考场后，按照监考老师安排隔位就座，将学生证放在桌面上。无学生证者不能参加考试；迟到超过 15 分钟不得入场。在考试开始 30 分钟后方可交卷出场。

2、除必要的文具和主考教师允许的工具书、参考书、计算器以外，其它所有物品（包括空白纸张、手机、或有存储、编程、查询功能的电子用品等）不得带入座位，已经带入考场的必须放在监考人员指定的位置。

3、考试使用的试题、答卷、草稿纸由监考人员统一发放，考试结束时收回，一律不准带出考场。若有试题印制问题请向监考教师提出，不得向其他考生询问。提前答完试卷，应举手示意请监考人员收卷后方可离开；交卷后不得在考场内逗留或在附近高声交谈。未交卷擅自离开考场，不得重新进入考场答卷。考试结束时间到，考生立即停止答卷，在座位上等待监考人员收卷清点后，方可离场。

4、考生要严格遵守考场规则，在规定时间内独立完成答卷。不准交头接耳，不准偷看、夹带、抄袭或者有意让他人抄袭答题内容，不准接传答案或者试卷等。凡有违纪作弊者，一经发现，当场取消其考试资格，并根据《北京大学本科考试工作与学术规范条例》及相关规定严肃处理。

5、考生须确认自己填写的个人信息真实、准确，并承担信息填写错误带来的一切责任与后果。

学校倡议所有考生以北京大学学生的荣誉与诚信答卷，共同维护北京大学的学术声誉。

装订线内不要答题

得分

一、填空题（每小题 2 分，共 20 分）

1. 已知归并排序的递推公式及复杂度为 $T(n) = 2T(n/2) + O(n) = O(n \log n)$ ，给出下面算法的递推公式及复杂度：

在未排序的数组中按照顺序统计学方法查找第 i 小的元素：_____

按 pre-order 遍历二叉树：_____

在排好序的数组中用二分法查找给定的数值 x ：_____

2. 给定数组 $A[1 \cdots n]$ ，用分治策略同时求出 $A[1 \cdots n]$ 元素的最大值和次大值，最坏情形下，算法比较次数约为 _____

3. 假设采用低位优先法的基数排序方法对十进制序列 $A = \{57, 43, 88, 49, 26, 71, 68, 51, 12\}$ 进行不减排序，对第 1 位（个位）排序后的中间结果为

4. 下面是求最长公共子序列的递推算法，请填补程序中的空白

```
LCS(x, y, i, j) {
    if (i == -1 or j == -1)
        then 0
    else if (x[i] == y[j])
        then c[i, j] = LCS(x, y, _____, _____) + 1
    else c[i, j] = max{LCS(x, y, _____, _____), LCS(x, y, _____, _____)}
}
```

5. 假设某问题用动态规划算法的状态转移方程为： $s[i, j] = \max(s[i-1, j], s[i, j-1])$ ，其中 $0 \leq i \leq n, 0 \leq j \leq m$ ，那么该算法的时间复杂度为_____

6. 给定一个有 n 个元素的最大二叉堆，现在需要查找其中 key 值为某个特定的数值的元素，该查找操作的时间复杂度为_____

7. 课上讲过的 d 叉堆与二叉堆很类似，但其中的每个非叶节点有 d 个子女，而不是两个。在高度为 h 的 d 叉堆中，最多的元素个数是_____，最少的元素个数是_____
8. 在 n 皇后问题中，设 $P(x_1, x_2, \dots, x_i)$ 为真表示向量 $\langle x_1, x_2, \dots, x_i \rangle$ 中 i 个皇后放置在彼此不能攻击的位置， $0 < k < n$ ，则多米诺性质是指：_____
9. 设组合优化问题的求解时间为 $f(n) = c2^n$ ，如果分解为 k 个子问题，每个子问题大小为 n/k ，组合时间 $T = O(f(n))$ ，则分解为子问题后的求解时间为_____
10. 假设原问题 L 的定义是“判断一个图存在哈密顿回路”。那么该问题的补问题 \bar{L} 的定义是_____

得分	二、判断题（正确打✓，错误打✗。）（每小题 1 分，共 10 分）

1. 设 $a \geq 1, b \geq 1$ 且 a, b 为常数，设 $f(n)$ 是一个函数， $T(n) = aT(n/b) + f(n)$ ，
若对于某常数 $\epsilon > 0$ ，有 $f(n) = O(n^{\log_b a + \epsilon})$ ，则 $T(n) = \theta(n^{\log_b a})$ []
2. 快速排序是一种稳定的排序方法 []
3. 随机快速排序用一个随机的参照元素（pivot）来划分数组，其运行
时间与输入数组中元素的排列顺序无关 []
4. 堆排序和快速排序具有空间原址性：即在排序过程中任何时候都只
需常数个额外空间存储临时数据 []
5. 一个所有元素各不相同的最大值堆中，其最小元素位于叶节点 []
6. 能够用回溯法和分支限界法解决的问题属于易解问题 []
7. 给定流网络 $\langle G, s, t \rangle$ ， $\langle S, T \rangle$ 是它的任意一个最小割， f 是它的任意
一个最大流。在对应的剩余网络 G_f 中，不存在从 S 中任何结点出发达到
 T 中任何结点的路径。 []
8. 最大流 f 的剩余网络中 G_f 中不存在从源点 s 到汇点 t 的路径，也不存
在从汇点 t 到源点 s 的路径。 []
9. 假设某个问题是 NP 问题。如果 $P \neq NP$ ，那么该问题不可能有多项
式时间的解法。 []
10. 如果一个 NPC 问题能归约到问题 A，则问题 A 也是 NPC 问题 []

得分

三、单选题。（每小题 2 分，共 20 分）

- 求解递归式 $T(n) = 2T(n/2) + n \log n$ []
 A) $T(n) = \theta(n)$
 B) $T(n) = \theta(n \log n)$
 C) $T(n) = \theta(n \log^2 n)$
 D) $T(n) = \theta(n \log n)$
- 用分治法求第 n 个 Fibonacci number 的复杂度是_____，求 a^n 的复杂度是_____ []
 A) $\log(n)$, $\log(n)$
 B) n , $\log(n)$
 C) $\log(n)$, n
 D) $\omega(\varphi^n)$, $\log(n)$
- 如果要把两个分别有 n 个元素的二叉堆合并为一个堆，那么该操作的复杂度是 []
 A) $\theta(1)$
 B) $\theta(\log n)$
 C) $\theta(n)$
 D) $\theta(n \log n)$
- 使用好的优先队列对下列哪个算法能改善算法的运行时间？ []
 A) 最小生成树 Kruskal 算法
 B) 单源最短路径 Dijkstra 算法
 C) 最大流 Bellman-Ford 算法
 D) 矩阵乘法的 Strassen 算法
- 赫夫曼（Huffman）编码采用的是 []
 A) 穷举法
 B) 动态规划
 C) 贪心法
 D) 分治法

6. 如果图 $G=(V, E)$ 中的每条边的长度均为 1，则求给定起点的单源最短路径问题的时间复杂度为（选取最紧的渐进界）： []
- A) $\theta(V^2)$
 - B) $\theta(E + V\log V)$
 - C) $\theta(E\log V)$
 - D) $\theta(V + E)$
7. 给定连通图 $G=(V, E)$ ，假设用普通二叉堆实现，那么 Prim 和 Dijkstra 算法的复杂度分别是 []
- A) $O(E\log V)$, $O(E\log V)$
 - B) $O(E\log V)$, $O(E+V\log V)$
 - C) $O(E+V\log V)$, $O(E\log V)$
 - D) 以上答案都不正确
8. 下列哪个问题不是 NP 完全问题？ []
- A) 团集问题
 - B) 顶点覆盖问题
 - C) 欧拉回路问题
 - D) 4-CNF-SAT 问题
9. 对于 MAX-2-CNF 问题，我们希望知道最多有多少个从句可以被满足。采用给每个变量随机赋值的方法，可以得到该问题的一个近似解。该算法的随机近似比是 []
- A) $7/8$
 - B) $8/7$
 - C) $3/4$
 - D) $4/3$
10. 以下关于均摊分析的说法**错误**的是 []
- A) 势能法中的“势”是与整个数据结构而不是其中的个别对象发生联系的。
 - B) 使用记账法时数据结构中存储的“总存款”不能是负的。
 - C) 聚集法假设操作序列符合某种概率分布下通过总运行时间求平均得到平摊时间。
 - D) 平摊分析保证在最坏情况下，每个操作具有的平均性能。

得分

四、Dijkstra 算法（5 分）

图论中求最短路径的 Dijkstra 算法要求图中没有权重为负数的边。有人提出下面的方法把包含负边的图转换成没有负边的图：找到图中权重最小的负边，设其权重为 $-W$ 。把图中所有边的权重都加上 W ，然后再用 Dijkstra 算法求最短路径。请问这种求解最短路径的方法是否正确？证明你的结论或举出反例。

得分	五、任务分配问题（5 分）

该问题定义如下：设有 n 项任务， k 台机器。任务 i 可以在任意一台机器上完成，所需要的时间均是 t_i 。求一个最佳的任务分配方案，使得完成时间（从时刻 0 计时，到最后一台机器的停止时间）达到最短。

有人提出用“贪心算法”来解这个问题：把 n 项任务按照所需要的时间从大到小排序。每当机器有空闲时，在当前尚未分配的任务中抽取所需时间最长的任务分配到任意一台空闲机器上。请问这种贪心算法是否可以得到最优解（即题中所定义的完成时间最短）？证明你的结论或举出反例。

得分

六、0-1 背包问题（5 分）

该问题定义如下：有 n 件物品和一个最大承重量为 W 的背包。第 i 件物品的重量为 w_i ，价值为 v_i 。每件物品只能作为一个整体来选择放与不放。求应该将那些物品放入背包中使其价值总和最大？（注意：这段话只是背景介绍，不用回答。只要回答下面所提出的问题就行了。）

这个问题可以用回溯和分支限界法求解。现在如果允许背包问题中的某些物品的重量为负数（价值仍然是非负数），那么该如何求解？

得分

七、开始时间可变的活动安排问题（9 分）

活动安排问题的定义如下：假设有 n 个活动 $a_1, a_2 \cdots a_n$ ，每个活动有一个开始时间 $s_1, s_2 \cdots s_n$ 和一个结束时间 $t_1, t_2 \cdots t_n$ 。这 n 个活动都要使用一个公共的教室，这个教室在任何时间只能供一个活动使用。每个活动一旦开始就要一直持续到结束，中间不能打断。如何安排这些活动，使得能成功进行的活动数量最多？（注意：这段话只是背景介绍，不用回答。只要回答下面所提出的问题就行了。）

现在考虑该问题的一个变种 —— **开始时间可变的活动安排问题**：每个活动有一个截止时间（即 **deadline**） $t_1, t_2 \cdots t_n$ 和一个时间长度（即举行该活动所需的时间） l_1, l_2, \dots, l_n 。也就是说，该活动的最晚开始时间是 $t_i - l_i$ ，在此之前的任何时间只要教室有空都可以开始。这 n 个活动仍然要使用一个公共的教室，这个教室在任何时间只能供一个活动使用。每个活动一旦开始就要一直持续到结束，中间不能打断。如何安排这些活动，使得能成功进行的活动数量最多？写清楚算法设计（无需给出伪代码）和时间复杂度分析，并简要证明该算法的正确性。

得分

八、包含多个时间段的活动安排问题（9 分）

现在考虑活动安排问题的另一个变种（**注意：这道题与前面那道题是相互独立的**）—— **包含多个时间段的活动安排问题**：每个活动包含多个时间段，每个时间段都有一个固定的开始时间和一个固定的结束时间。这 n 个活动都要使用一个公共的教室，这个教室在任何时间只能供一个活动使用。只有当某个活动的所有时间段都可以被安排时，才能安排该活动。每个时间段一旦开始就要一直持续到结束，中间不能打断。如何判断是否能够安排其中至少 k 个活动？

- （1） 证明该问题是 **NPC** 的。可以用任何一个已知是 **NPC** 的问题进行归约。（6 分）
- （2） 解释一下为什么不能用同样的方法证明原始的活动安排问题（即每个活动只能包含一个连续的时间区间）是 **NPC** 的？（3 分）

3) 在 2)的基础上, 沿着 $3 \rightarrow 1$ 这条边把解划分为两组。分别给出包含 $3 \rightarrow 1$ 这条边和不包含 $3 \rightarrow 1$ 这条边的解的费用矩阵和费用下限。(4 分)

得分

十、网络流应用（9 分）

地震救灾中，伤员救治是一个重要的工作，只有尽快把伤员送到医院获得及时医治才能最大限度的减少伤亡。假设灾区有 k 所医院，每个医院 i 有床位 b_i 个。现共发现 n 个伤员，每个伤员 j 距离每个医院 i 的距离以运送时间计算记为 d_{ij} 。我们希望伤员既能在黄金半小时内送到医院，又不能把一个医院的床位用完导致无法医治后续发现的伤员，因此需要尽可能地保持医院床位占用率平均。

- 1) 请设计一个最大流算法，判断是否可以将 n 个伤员在黄金半小时内送往 k 所医院，并保证每个医院 i 的床位占用数不超过 $\left\lceil \frac{nb_i}{\sum b_i} \right\rceil$ ？画出流网络图，说明源点、汇点、中间节点、边、容量、流量分别代表的含义。（5 分）

2) 如果无法保证医院床位占用率绝对平均，如何设计最大流算法，判断能否保证伤员在黄金半小时内得到救治？如果能，又如何尽可能的使床位占用率的最大值最小化（也就是使各医院的床位占用率尽可能地平均）？（4 分）