

## 网易 2020 校招笔试- 算法工程师（提前批）

1. 有一类二叉树用三叉链表来存储的时候除了带有指向左右孩子节点的两个指针，还有指向父节点的指针，那么这样一棵二叉树有 2 个节点，那么有多少指针指向 NULL（注：根节点的父指针指向 NULL，对于不存在的节点表示为 NULL）？  
A.1      B.2      C.3      D.4      E.5
2. 图 N 是有 7 个顶点的强连通图，那么 N 是有向图，那么 N 最少有（）条边？如 N 为无向图，N 最少有（）条边？  
A.21,21      B.21,20      C.7,7      D.7,6      E.8,7
3. 下列最短路径算法的叙述中正确的是（）  
A. Dijkstra 算法通常用于求每一对顶点间的最短路径；  
B. Dijkstra 算法不允许图中带有负权值的边，而 Floyd 算法则可以适用；  
C. Floyd 算法通常用于求某一顶点到其他各顶点的最短路径；  
D. Floyd 算法允许有包含负权值的边组成的回路，而 Dijkstra 算法不允许；
4. 以下是泊松分布概率密度函数的是

$$\binom{n}{k} p^k (1-p)^{n-k}, \quad k=0, 1, 2$$

A .

$$\frac{\lambda^k e^{-\lambda}}{k!}, \quad k=0, 1, 2, \dots$$

B.

$$\frac{1}{\theta} e^{-x/\theta} \quad x > 0$$

C.

$$D. \quad 1/(b-a), \quad x \in [a, b]$$

5. 判断一个数组或序列是正序,倒序还是乱序,需要我们将这个数组完整的遍历一遍通过构建有序序列，对于未排序数据，在已排序序列中从后向前扫描，找到相应的位置并插入

的排序算法是 ( )

- A.选择排序      B.希尔排序      C.插入排序      D. 归并排序

在总体  $N(52, 6.3^2)$  中随机抽取一容量为36的样本, 求样本的均值  $\bar{X}$  落在50.8到53.8之间的概率。(注:  $\Phi(0.16) = 0.5636$ ,  $\Phi(0.19) = 0.5753$ ,  $\Phi(1.71) = 0.9564$ ,  $\Phi(1.14) = 0.8729$ )

6.

- A.0.5317      B.0.1389      C.0.4365      D.0.8293

7. 若将 KNN 看成是 kernel method 的一种实现, 则 kernel 函数为

$$K_{\lambda}(x_0, x) = \frac{1}{\lambda} \exp\left[-\frac{\|x - x_0\|^2}{2\lambda}\right], \text{ 其中 } \lambda \text{ 为超参}$$

A.

$$K_k(x_0, x) = I(\|x - x_0\| \leq \|x_{(k)} - x_0\|) \text{ 其中 } k \text{ 为超参, } x_{(k)} \text{ 为与 } x_0 \text{ 第 } k \text{ 近邻的样本}$$

B.

$$K(x, z) = x \cdot z$$

C.

$$K(x, z) = (x \cdot z + 1)^p$$

D.

8. 当  $x$  在  $[x_1, x_2]$  非 0, 其余情况为 0, 则以下是  $x$  的概率密度  $p(x)$  的最大熵估计是

- A.均匀分布      B.高斯分布      C.指数分布      D.泊松分布

9. 下面对机器学习算法的阐述不正确的有:

- A. KNN 算法不仅可以用于分类, 还可以用于回归;  
B. 分类树的核心思想是递归划分自变量的空间;  
C. n-gram 对长文本分类表现良好;  
D. LDA 实质是多层贝叶斯的结合;

10. 下面对神经网络结构中不同层的说法不正确的是:

- A. Fully connected 通过非线性变化, 提取特征之间的关联;  
B. Pooling 主要对卷积后的特征层进行操作, 进行特征的压缩;  
C. Activation Function 可以为模型引入线性因素;  
D. Dropout 有效防止模型过拟合;

11. 请选择一个自己最熟悉的深度学习模型 (比如 lstm 等) 或者统计机器学习模型 (比如树模型、LR、FM 等), 详细介绍 loss function 推导过程、网络结构、优化方法的求解过程、模型优缺点以及最佳应用场景, 越细越好

12. 社区内容推荐方案:

(1) 有百万级别用户群 A

(2) 我们有用户群 A 听歌记录表 a, 表 a 中的数据为: 用户 id, 用户喜欢的歌曲 id, 听

歌的时间 (年月日小时分钟秒, 比如 20190816112002), 表 a 的记录为亿级别

(3) 我们有用户群 A 收藏歌曲记录表 b, 表 b 中的数据为: 用户 id, 用户收藏的歌曲 id, 收藏歌的时间 (年月日小时分钟秒, 比如 20190816112005), 表 b 的记录为千万级别

(4) 我们有用户群 A 点击查看过的社区内容记录表 c, 表 c 中的数据为: 用户 id, 用户点击查看过的社区内容 id, 点击发生的时间 (年月日小时分钟秒, 比如 20190816112005), 表 c 的记录为千万级别

(5) 我们有用户群 A 评论过的社区内容记录表 d, 表 d 中的数据为: 用户 id, 用户评论过的社区内容 id, 评论 id, 评论发生的时间 (年月日小时分钟秒, 比如 20190816112005), 表 d 的记录为千万级别

(6) 表 e: 歌曲 id, 歌曲名, 演唱者, 歌词文本

(7) 表 f: 社区内容 id, 发布者用户 id, 文本内容

(8) 表 g: 评论 id, 评论发布者用户 id, 文本内容

请根据上面提供的数据, 设计一套社区内容推荐方案, 能尽量利用全部的信息, 给用户精准的推荐社区内容

### 13.[编程题]序列交换

时间限制: C/C++ 1 秒, 其他语言 2 秒

空间限制: C/C++ 256M, 其他语言 512M

小易给你一个包含  $n$  个数字的数组  $a_1, a_2, \dots, a_n$ 。你可以对这个数组执行任意次以下交换操作:

对于数组中的两个下标  $i, j (1 \leq i, j \leq n)$ , 如果  $a_i + a_j$  为奇数, 就可以交换  $a_i$  和  $a_j$ 。

现在允许你使用操作次数不限, 小易希望你能求出在所有能通过若干次操作可以得到的数组中, 字典序最小的一个是什么。

输入描述:

第一行一个整数  $n$ ;

第二行  $n$  个整数  $a_1, a_2, \dots, a_n$ , 表示数组, 每两个数字之间用一个空格分隔。

输入保证  $1 \leq n \leq 10^5; 1 \leq a_i \leq 10^9$ 。

输出描述:

$n$  个整数, 每两个整数之间用一个空格分隔, 表示得到的字典序最小的数组。

输入例子 1:

4

7 3 5 1

输出例子 1:

7 3 5 1

输入例子 2:

10

53941 38641 31525 75864 29026 12199 83522 58200 64784 80987

---

输出例子 2:

12199 29026 31525 38641 53941 58200 64784 75864 80987 83522

#### 14. [编程题]数字圆环

时间限制: C/C++ 1 秒, 其他语言 2 秒

空间限制: C/C++ 256M, 其他语言 512M

小易有一个长度为 $n$ 的数字数组 $a_1, a_2, \dots, a_n$ 。

问你是否能用这 $n$ 个数字构成一个环(首尾连接), 使得环中的每一个数字都小于它相邻的两个数字的和(每个数字都必须使用并且每个数字只能使用一次)。

输入描述:

第一行包含一个整数 $t$  ( $1 \leq t \leq 10$ ), 表示测试用例的组数。

每个测试用例输入如下:

第一行一个整数 $n$ , 表示数字的个数;

第二行 $n$ 个整数 $a_1, a_2, \dots, a_n$ , 每两个整数之间用一个空格分隔。

输入数据保证

$$3 \leq n \leq 10^5$$

$$1 \leq a_i \leq 10^9$$

输出描述:

输出应该包含 $t$ 行, 对于每组用例, 若能输出"YES", 否则输出"NO"。

输入例子 1:

1  
5  
17 6 17 11 17

输出例子 1:

YES

输入例子 2:

1  
3  
1 2 4

输出例子 2:

NO

#### 15. [编程题]优秀的 01 序列

时间限制: C/C++ 1 秒, 其他语言 2 秒

---

空间限制：C/C++ 256M，其他语言 512M

给定 01 序列 S, 序列 S 是优秀的 01 序列, 优秀的 01 序列定义如下:

- 1、如果序列 S,T 是优秀的, 则序列 S+T 是优秀的, +被定义为按顺序连接两个序列, 即 "010"+"110"="010110"。
- 2、如果序列 S 是优秀的, 则序列 rev(S)也是优秀的。rev(S)被定义为按位翻转(0 变 1,1 变 0) 序列 S, 并删去前导零。例如 rev("1100101")="11010"。

现在请你判断序列 T 是不是优秀的

输入描述:

第一行数据组数T, 表示有T组数据。  
每组数据的第一行是一个01序列, 表示序列s。第二行是另一个01序列, 表示序列T。  
 $1 \leq |S|, |T| \leq 1000$ , s, T不含前导零。

输出描述:

对于每组数据, 一行输出"YES"或者"NO", 表示序列 T 是不是优秀的。(大小写敏感)

输入例子 1:

1  
1100  
110011

输出例子 1:

YES

输入例子 2:

1  
1000  
100001111

输出例子 2:

NO

## 16. [编程题]倒数排列

时间限制：C/C++ 1 秒，其他语言 2 秒

空间限制：C/C++ 256M，其他语言 512M

有一天, 小易把 1 到 n 的所有排列按字典序排成一排。小易从中选出了一个排列, 假设它是正数第 Q 个排列, 小易希望你能回答他倒数第 Q 个排列是什么。

例如 1 到 3 的所有排列是:

1 2 3

---

1 3 2

2 1 3

2 3 1

3 1 2

3 2 1

若小易选出的排列是 1 2 3，则  $Q = 1$ ，而你应该输出排列 3 2 1

输入描述:

第一行数字 $n$ , 表示排列长度  
接下来一行 $n$ 个数字, 表示选出的排列  
 $1 \leq n \leq 300000$

输出描述:

一行  $n$  个数字, 表示所求的排列。

输入例子 1:

3

1 2 3

输出例子 1:

3 2 1

输入例子 2:

5

3 1 5 2 4

输出例子 2:

3 5 1 4 2

选择题答案

DDBBCDBACC