

2023 年稳博投资量化研究员笔试题

笔试说明：

- 1、本次**笔试时长预计7小时**，请于笔试开始后48小时内交卷；
- 2、本 pdf 为笔试题目，答案按要求放在答卷文件夹中，已建立并命名为“姓名+笔试时间+交卷文件”，按要求答题后压缩该文件夹回传答案即可。

一、算法题-----深境螺旋

-----传说在遥远的古代，「王」与「祭司」会共同攀登指向天空的尖塔，寻求某种指引。

因为螺旋有两条路，所以你只需要完成两道题目中的任意一道即可，第一道题目较为容易，第二道题目会非常难。当然如果完成的话第二道题目的分数也会是第一道题目的两倍以上。

渊月祝福：如果对自己的代码功力不自信，也可以提交文字说明自己的思路，对每一档数据可以获得一半分数，要求思路说明必须详细，不可以过于模糊，不建议使用该种方式除非你觉得自己真的不善于写代码。

（一）左侧通道

做题时间建议：1H

运行时间：1S

可莉来到了左侧的台阶，发现面对的是一个特殊的地图。整个地图一共有N个地点，地点之间彼此相通，并且好心的旅行者还告诉她这张图是一棵树的结构。（什么是树呢？树就是有n个节点和n-1条边形成的无环连通无向图。——旅行手册介绍）

但是可莉面对复杂的地图还是迷路了。为了估算一下走出去的时间，她现在很想知道从任意编号的点出发，到距离该点最远的点需要多少步。

输入格式：

第一行一个整数N，表示一共有N个地点

第二到第N行，每行两个数，表示这两个地点之间有路

输出格式：

N个整数，表示从该地点走到距离该点最远的点需要的步数

样例：

输入：

4

1 4

2 4

3 4

输出：

3 3 3 2

样例解释：

地点1，2，3均与4号点相连，除此之外都没有路。

所以从地点1出发，最远点是地点2或者地点3，都需要通过4，算起点和终点长度为3

只有从地点4出发，到任意一个点都只要一步，所以算起点和终点长度为2

第八层：N ≤ 1000（10%的数据）

第九层：N ≤ 10000（30%的数据）

第十层：N ≤ 200000（60%的数据）

第十一层：N ≤ 10000000（100%的数据）

(二) 右侧通道

做题时间建议：1H

运行时间：1S

影来到了右边的阶梯，出现在眼前的是神樱树，树上的每个节点都寄宿着一位故友的灵魂。（什么是树呢？树就是有 n 个节点和 $n-1$ 条边形成的无环连通无向图。——旅行手册介绍）传说，只要在一定距离内，心有所向的两人便会互相吸引。所以现在影想请教你，这棵树上有多少对节点的距离在 K 以内。

输入格式：

输入会有多组数据，对于每组数据：

第一行有两个整数 N, K

之后 $N-1$ 行每行三个整数： U, V, L 表示节点 U 和节点 V 之间的边长度为 L

最后会以 $0\ 0$ 结尾代表输入结束

输出格式：

对于每组测试数据，输出一个整数，代表这样的节点对数

样例：

输入：

```
5 7
1 3 5
2 3 1
1 4 4
4 5 3
0 0
```

输出：

```
6
```

样例解释：一共有6对节点之间的距离 ≤ 7

分别为（1，3） 距离为5

（2，3） 距离为1

（1，4） 距离为4

（4，5） 距离为3

（1，5） 道路为1->4->5, 总距离为7

（1，2） 道路为1->3->2, 总距离为6

第九层： $\Sigma N \leq 1000$ （10%的数据）

第十层： $\Sigma N \leq 10000$ （30%的数据）

第十一层： $\Sigma N \leq 300000$ ， 数据大体随机（60%的数据）

第十二层： $\Sigma N \leq 400000$ ， 数据精心构造（100%的数据）

提示：本题最后两档有一定难度，如果觉得写不出路的话，有思路直接写思路，也可以根据思路给予得分。

二、金融数据计算-超额收益率计算

做题时间建议：30Min

现有一投资组合（附件 1），请利用所提供的行情数据（附件 2），计算 2021 年 11 月该组合相对于上证指数的每日超额收益率。

提示：T 日收益率 = T 日收盘价 / T-1 日收盘价 -1

T 日超额收益率 = T 日组合收益率 - T 日上证指数收益率

假设权重不变，vwap数据自己准备，也可以依赖现有数据

结果四舍五入保留 4 位小数，存入 result.csv 文件，格式如下：

1	date,excessReturn
2	20210507,-0.0156
3	20210510,-0.0008
4	20210511,0.0033
5	20210512,0.0049

三、机器学习题

做题时间建议：1H

现有一大小为 $m \times n$ 的股票数据集，其包含了多支股票多天的特征数据（一支股票每天可能有 1000-5000 行数据， m 大于 1000000），前两列为时间和股票代码，后 $n-2$ 列为特征，现在想要对其作 1D-CNN 的回归训练（假设 kernel 个数为 1，kernel size 为 3），请简要描述该如何进行数据处理及模型训练，并画出 CNN 流程图（每一层之间如何转化）。

四、数学题 做题时间建议：1.5H

（一）证明对任意实数列，...

$$\sum_{i=1}^n \sum_{j=1}^n \frac{a_i a_j}{i+j} \geq 0$$

恒成立。

（二）证明存在一个常数 $c > 0$ 对于任意整数 n 具有如下性质：

证明存在一个常数 $c > 0$ 对于任意整数 n 具有如下性质：对于任意 $n \times n$ 元素各不相同的矩阵 A ， A 的第 i 行为 A_i ，存在一个 $1, 2, 3, \dots, n$ 的排列 $M = \{a_1, a_2, a_3, \dots, a_n\}$ ，使得矩阵 $A[M] = [A_{a_1}^T, A_{a_2}^T, A_{a_3}^T, \dots, A_{a_n}^T]^T$ 没有一列有长度超过 $c\sqrt{n}$ 的递增子列。

（三） Z 是在半径为 a 的圆内随机取两点的距离，求 EZ^2

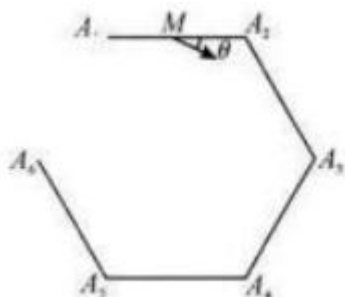
（四）在三角形 A 中随机取三点连成三角形 B ，证明： $E(B \text{ 的面积}) = A \text{ 的面积} / 12$ 。

五、物理题

（一）求光线出射角 θ 的范围

做题时间建议：30Min

如图所示，一正六边形 $A_1A_2A_3A_4A_5A_6$ ，除去 A_6A_1 每边均为一平面镜， A_6A_1 边为空气。现从 A_1A_2 中点 M 发出一条光线，此光线依次经 A_2A_3 ， A_3A_4 ， \dots ， A_5A_6 反射，最后从 A_6A_1 边射出，试求光线出射角 θ 的范围。



（二）电阻和随机行走

做题时间建议：30Min

1. 考虑 N 维空间 R^N ，如果在所有相邻整数格点（ ZN ）间都连接一个大小为 r 的电阻，那么两个相邻格点间的等效电阻是多少？

2. 考虑 N 维空间中的随机行走，每一步有相同概率到达相邻格点，那么从原点出发不返回原点而经过 $(0, 0, \dots, 0, 1)$ 逃出的概率是多少？
3. 是否可以简要说明解释上述 N 维网格上的随机行走常返的条件？（可以不严格证明，常返指回到原点概率为 1，也可以理解为无法从无穷远点逃出）