一、 Equation Puzzle

下列乘法算式中：每个字母代表1个数字（0~9）。相同的字母代表相同的数字，不同的字母代表不同的数字。例如：

ACD \* BA = CDBE

试编程确定使得整个算式成立的数字组合，如有多种情况，请给出所有可能的答案。

例如：

你的程序输入只有一个参数，一个文本文件的路径，包含如下内容:

CCD \* JC = FJDHD

你的程序的输出是stdout标准打印输出：

220 \* 82 = 18040

注意: 保留原有空格

二、 Shuttle bus

诺基亚杭州研发中心将开通一辆承载能力无限的超级班车。如图，班车经过的站点构成M列N行的网格(1<= M,N <=100)，

起点(1,1) 站点(2,1) ... 站点(M,1)

站点(1,2) 站点(2,2) ... 站点(M,2)

... ...

站点(1,N) 站点(2,N) ... 公司(M,N)

现已知每个站点需要乘车的员工数（每个站点都不超过10人，终点CMN=0）：

C11, C21, C31, ... CM1,

C12, C22, C32... CM2,

... ...

C1N, C2N, C3N, ... CMN,

请问班车沿着网格从起点(1,1)出发到公司(M,N)，只能按图中站点向下或向右行驶, 一次最多能运送多少员工？

例如：

M=3， N=2

需要乘车的员工数C11 - CMN：

C11=5，C21=2，C31=4

C12=1，C22=6，C32=0

则行驶路线为(1,1)->(2,1)->(2,2)->(3,2), 一次最多运送5+2+6=13名员工

参数范围：

1 <= M, N <= 100

0 <= C11~CMN <= 10

你的程序输入只有一个参数，一个CSV文件的路径，包含如下内容:

M, N

C11, C21, C31, ... CM1,

C12, C22, C32... CM2,

... ...

C1N, C2N, C3N, ... CMN,

例如：

3, 2

5, 2, 4

1, 6, 0

你的程序的输出是stdout标准打印输出.

例如

13

三、 Test Lines

诺基亚杭州研发中心有多条云测试线，编号为1，2，3，...，testlineNum；有多个测试工程师，编号为1，2，3，...，userNum。

每名工程师需要在同一时间分别使用其中一条测试线(1 <= userNum <= testlineNum)。

每条测试线只能分配给一个用户。

每个工程师对测试线的需求是不同的，所以他们能申请的测试线不尽相同。

例如:

工程师1：可申请1, 3, 4, 6, 7；

工程师2：可申请2, 3, 4, 8, 9, 15。

要让每名工程师都能分配到一条合适的测试线，请计算有多少种分法？

数据范围：

userNum <= testlineNum

1 <= userNum <= 10

1 <= testlineNum <= 20

你的程序输入只有一个参数，一个CSV文件的路径，包含如下内容:

testline\_num, user\_num

user1\_testline\_list

user2\_testline\_list

...

例如

3, 2

1, 2

1, 2, 3

你的程序的输出是stdout标准打印输出.

例如

4

四、 SAFe Challenge

诺基亚将从现在开始逐步应用SAFe，每个ART（敏捷发布列车）上的所有功能和启用程序将在PI会议上一起计划。

我们假设：

- 所有的价值因素，如用户业务价值，时间紧迫性和风险降低/机会增加，已被统一为一个数字价值，表示利润或亏损。

- 客户会为他们需要的feature付款，所以feature将为公司带来利润。

- 客户不需要所有enabler，所以这些enabler只会消耗成本，即亏损。但是某些feature可能依赖于它们，因此有时必须完成它们。

- 所有feature和enabler可能相互依赖

作为RTE/PO，您必须选择一些功能和促成因素，以确保NOKIA获得最大利润。

输入：第1行是feature和enabler的总数（0 <= n <= 1000），然后每行代表一个feature或enabler，其中第一个数字是成本（负值）或利润（正值）值（-1,000,000 <= n <= 1,000,000），在此之后，数字是这个所依赖的所有feature和enabler的索引。它们由空格分隔。

输出：第一行是公司的最大利润，然后是所有选定feature/enabler的列表，每行都是一个feature/enabler的索引。如果有多种解决方案可以获得最大利润，那么选择项目数最少的解决方案。

你的程序输入只有一个参数，一个文本文件的路径，包含如下内容:

item\_num

item1\_dependence\_list

item2\_dependence\_list

...

例如：

6

-4

1

2 2

-1 1 2

-3 3

5 3 4

你的程序的输出是stdout标准打印输出，包含如下内容，index从1开始：

Maximum\_profits

selected\_item\_index1

selected\_item\_index2

...

例如:

3

2

3

五、 Machine Learing

1. 任务描述

根据（姓名，性别）数据建立机器学习/深度学习模型，以根据姓名预测性别。

数据：

训练集：包含(姓名，性别)的数据

测试集：只包含姓名

你需要做的：

1）从姓名中提取特征

2）从零开始建立并训练模型（不能使用第三方机器学习库）

3）根据姓名预测性别

4）提交你的代码

评分标准：

1）模型在测试集上的预测正确率

2）程序执行时间（包括训练和预测）必须小于10分钟，否则不计算正确率

训练集格式如下：

John,male

Lucy,female

...

测试集格式如下：

John

Lucy

...

你的程序输入只有一个参数，一个测试集文件的路径。

输出是标准打印输出，格式与训练集文件格式相同。

允许使用的python第三方库：

Numpy

注意： 如果你为了更高的性能而使用C/C++或Java，请告知组委会你需要的第三方库，组委会回酌情评估是否允许，如允许则在服务器上安装。

一般只有辅助矩阵运算的库才可能被允许，现成的机器学习库是被禁止的。

机器学习模型：

允许使用任何机器学习模型，如逻辑回归模型，神经网络等。

在线调试时，训练集文件路径是固定的，如下：

"/var/www/html/training\_dataset.txt"

训练集下载路径：

http://10.182.198.199/training\_dataset.txt

2.特征提取

我们可以用X表示特征 X = {x1, x2, ..., xn}， n为特征个数， 比如：

x1 = 1 如果字母a出现在名字中，否则为0

x2 = ‘a’ 在名字中出现的次数

对 'John','Lucy'等名字，可以提取如下特征:

names x1=是否包含字母‘c’ x2=字母‘h’出现次数 ...

John 0 1 ...

Lucy 1 0 ...

3.模型实现例子：

1) Neural Network from scratch http://www.wildml.com/2015/09/implementing-a-neural-network-from-scratch/

2) Logistic Regression from scratch: https://medium.com/@martinpella/logistic-regression-from-scratch-in-python-124c5636b8ac