**Assignment Questions:**

Q1: From the following list = ["My","First","Probability","Program"], Randomly sample 1000 words. For every word, collect the neighbour word “My”, call it as a sample space for the word say “My”. If you do this for all words, you will get four different sample spaces. Now compute the probability of all words in the sample space.

**Code:**

N = int(input())

#Neighbourhood computation

mol = ["My","First","Probability","Program"]

itemall = []

source1 = []

source2 = []

source3 = []

source4 = []

import random

random.seed(N)

for j in range(1, 1000):

#random.seed(j)

i = random.randrange(0,len(mol),1)

#print(mol[i])

a=mol[i]

itemall.append(a)

for i in range(1,len(itemall)-1):

#a = itemall(i)

if itemall[i]== 'My':

source1.append(itemall[i+1])

elif itemall[i]=='First':

source2.append(itemall[i+1])

elif itemall[i]=='Probability':

source3.append(itemall[i+1])

elif itemall[i]=='Program':

source4.append(itemall[i+1])

#Neighbour of my

frequency1 = {}

for item in source1:

if item in frequency1:

frequency1[item]+=1

else:

frequency1[item]=1

#print(frequency1)

#{'Probability': 792, 'First': 744, 'My': 696, 'Program': 744}

#Neighbour of first

frequency2 = {}

for item in source2:

if item in frequency2:

frequency2[item]+=1

else:

frequency2[item]=1

#print(frequency2)

#{'First': 588, 'Probability': 660, 'Program': 804, 'My': 816}

#Neighbour of probability

frequency3 = {}

for item in source3:

if item in frequency3:

frequency3[item]+=1

else:

frequency3[item]=1

#print(frequency3)

#{'Program': 844, 'Probability': 528, 'My': 744, 'First': 720}

#Neighbour of Program

frequency4 = {}

for item in source4:

if item in frequency4:

frequency4[item]+=1

else:

frequency4[item]=1

#print(frequency4)

#{'First': 816, 'My': 720, 'Program': 900, 'Probability': 852}

print(frequency1['My']/len(source1),frequency1['First']/len(source1),frequency1['Probability']/len(source1),frequency1['Program']/len(source1))

print(frequency2['My']/len(source2),frequency2['First']/len(source2),frequency2['Probability']/len(source2),frequency2['Program']/len(source2))

print(frequency3['My']/len(source3),frequency3['First']/len(source3),frequency3['Probability']/len(source3),frequency3['Program']/len(source3))

print(frequency4['My']/len(source4),frequency4['First']/len(source4),frequency4['Probability']/len(source4),frequency4['Program']/len(source4))

Q2. For the following dataset, print a sentence probabilistically.

I love the world and the things in it.

I love the way cheetah runs.

I am a man of honor.

I will be a rich guy.

I am a teenager so I am a rebel.

I am an iconoclast and a fighter

I believe in education.

I love everything.

I watch a movie every day.

I hate pollution.

I love the work of god.

I love the beauty of this world.

I adore the way people try solve hard things.

I am nothing but a blade of grass.

I will unleash a lot of prophecies and will bring down hordes of legions unto this earth to destroy you.

I love doing things in a peculiar way.

I love and hate probability. It is so stupid and fun at the same time.

I love the way software programs work.

I love my room.

Test cases:

The system should take as input the number of words and the starting word. The system should print a sentence on its own.

For instance:

Input ={4,love}

Output={love the way software}

Code:

import re

def mooccur(N):

h=0

w=N[0]

for m in N:

count=N.count(m)

if(count>h):

h=count

w=m

return w

nmbr=int(input())

w=input()

out=w

dw=[]

f=open("Input.txt","r")

for k in f:

dw.append(re.split("\s|[.]", k))

for k in range(len(dw)):

while('' in dw[k]):

dw[k].remove('')

for p in range(nmbr-1):

nw=[]

for q in dw:

for r in range(len(q)-1):

if(w==q[r]):

nw.append(q[r+1])

word=mooccur(nw)

out+=' '+word

w=word

print(out)

f.close()

**Practice Questions:**

For the dataset available at the link: <https://www.kaggle.com/uciml/student-alcohol-consumption#student-mat.csv> (student-mat.csv), answer the following questions.

Q1. Calculate the probability a student gets an A(80%+), given that the student is:

1. Male

2. Female

Q2. Calculate the probability a student gets an A (80%+) in math, given that the student’s father’s job and mother job is:

1.Teacher

2. Engineer

Q3. Calculate the probability a student gets an A (80%+) in math, given that the student’s father’s education is: PG and mother education is: UG.

Q4. Calculate the probability a student gets an A (80%+) in math, given that the student’s family size is GT3.

You can use the following code if you want any help. You can modify the code to answer the question you want.

//solution for practice question 1.1

import pandas as pd

import numpy as np

df = pd.read\_csv('database.csv')

df.head(3)

//student gets 80+

df['grade\_A'] = np.where(df['G3']\*5 >= 80, 1, 0)

df['sex'] = np.where(df['sex'] == 'M', 1, 0)

df['count'] = 1

df = df[['grade\_A','sex','count']]

df.head()

pd.pivot\_table(

df,

values='count',

index=['grade\_A'],

columns=['sex'],

aggfunc=np.size,

fill\_value=0

)