Name: Harish A

Roll no: 205002038

Dept : IT A

Problem:

- Generate two sets of two dimensional normally distributed random numbers, with different means and variances (at least 40 to 50 numbers) in each set. Plot the scatter diagrams of both the sets. Based on the Shannon's entropy try to draw an optimized line which separates the two classes (i.e. sets)

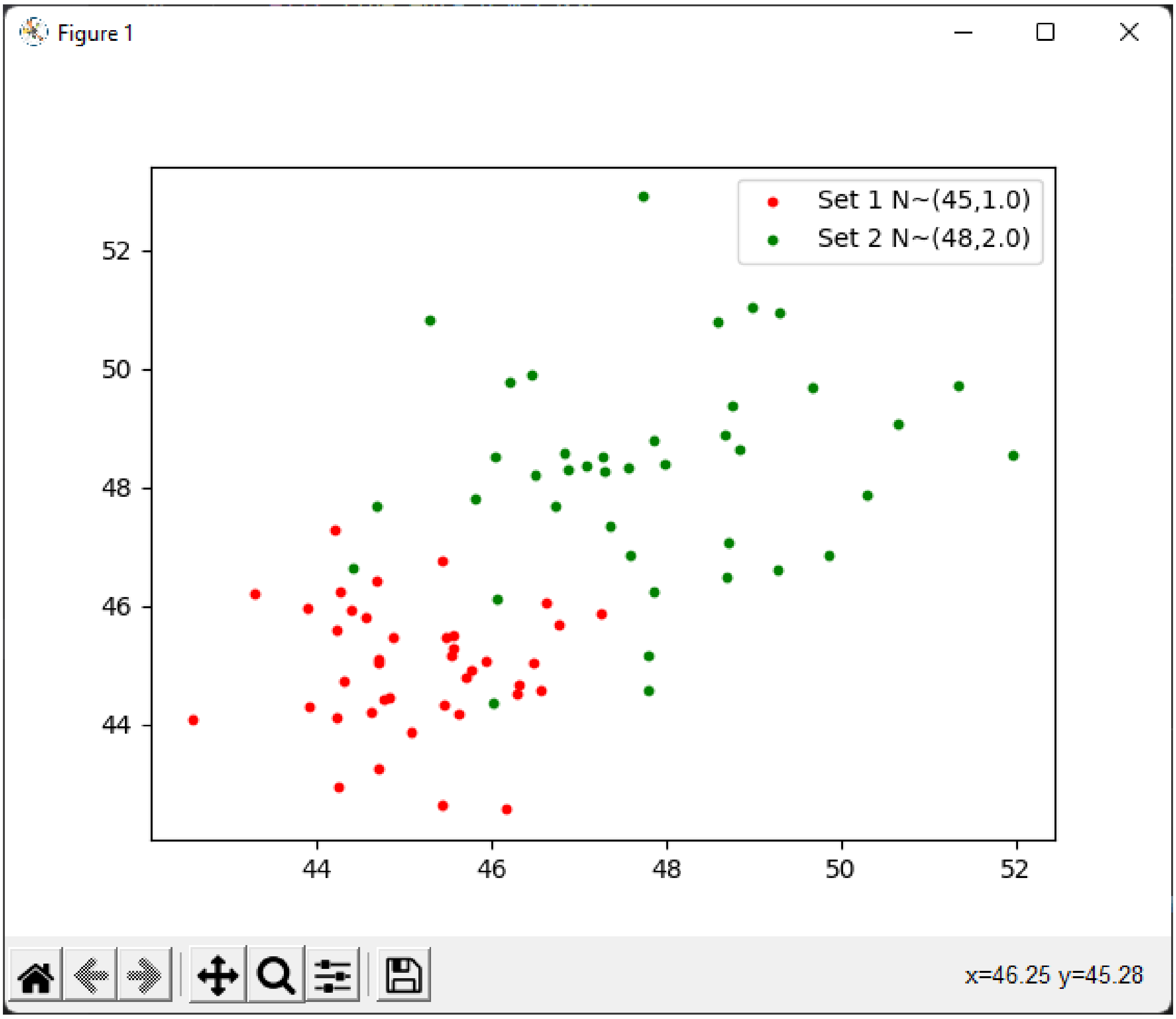
- If your register number is ending with 2 or 5 or 8 then the lower and upper limit to generate the random numbers 40 and 60

Datasets:

Set 1		**Set 2**	
Mean = 45	SD=1	Mean = 48	SD = 2
X	Y	X	Y
46.62607	46.05149	48.68348	46.47318
46.77099	45.68697	47.84502	46.24222
44.70248	43.24961	48.97622	51.02315
44.62976	44.20629	48.71258	47.05681
46.29156	44.50571	48.58619	50.78464
44.20317	47.29543	46.72356	47.67812
43.90035	45.97361	47.72151	52.90274
44.27353	46.23307	46.05658	46.10107
45.56406	45.48989	51.97304	48.55274
46.55249	44.5743	47.28887	48.26154
45.92464	45.05653	45.81625	47.80754
46.31784	44.65842	46.83808	48.58475
43.9195	44.30036	46.20055	49.78569
45.70762	44.79536	47.08493	48.35275
45.08374	43.86822	46.46056	49.90442
45.43526	42.63054	49.85189	46.86422
45.43636	46.75053	44.69198	47.69177
44.3847	45.93793	47.56287	48.33888
44.55243	45.7997	46.04131	48.50258
44.31065	44.72704	49.67909	49.69559
46.16449	42.56213	46.01139	44.35069
45.76427	44.91682	47.86301	48.80106
45.62573	44.16348	51.34588	49.72479
45.44955	44.3353	46.8699	48.31285
45.56172	45.28356	47.59027	46.86678

44.25247	42.95306	47.34314	47.33681
43.29322	46.21462	48.83918	48.6243
44.75628	44.41108	49.28499	50.93695
44.23102	44.11047	45.28393	50.81675
44.2324	45.59439	48.66306	48.87599
47.25375	45.86529	47.79578	44.58654
44.82994	44.4404	48.76061	49.38595
45.48242	45.47578	47.98618	48.39299
45.52865	45.15792	47.27948	48.52042
44.71176	45.03996	46.50529	48.21531
42.56648	44.08718	47.79523	45.16171
46.48003	45.04033	49.27814	46.61168
44.67973	46.41627	50.28434	47.86871
44.69974	45.11076	44.40954	46.62379
44.86153	45.45296	50.64589	49.07921

Plots:



Code for finding the optimum line:

```
import numpy as np

import matplotlib.pyplot as plt



x1=np.array([46.62607,46.77099,44.70248,44.62976,46.29156,44.2
0317,43.90035,44.27353,45.56406,46.55249,45.92464,46.31784,43.
9195,45.70762,45.08374,45.43526,45.43636,44.3847,44.55243,44.3
1065,46.16449,45.76427,45.62573,45.44955,45.56172,44.25247,43.
29322,44.75628,44.23102,44.2324,47.25375,44.82994,45.48242,45.
52865,44.71176,42.56648,46.48003,44.67973,44.69974,44.86153])

y1=np.array([46.05149,45.68697,43.24961,44.20629,44.50571,47.2
9543,45.97361,46.23307,45.48989,44.5743,45.05653,44.65842,44.3
0036,44.79536,43.86822,42.63054,46.75053,45.93793,45.7997,44.7
2704,42.56213,44.91682,44.16348,44.3353,45.28356,42.95306,46.2
1462,44.41108,44.11047,45.59439,45.86529,44.4404,45.47578,45.1
5792,45.03996,44.08718,45.04033,46.41627,45.11076,45.45296])



x2=np.array([48.68348,47.84502,48.97622,48.71258,48.58619,46.7
2356,47.72151,46.05658,51.97304,47.28887,45.81625,46.83808,46.
20055,47.08493,46.46056,49.85189,44.69198,47.56287,46.04131,49
.67909,46.01139,47.86301,51.34588,46.8699,47.59027,47.34314,48
.83918,49.28499,45.28393,48.66306,47.79578,48.76061,47.98618,4
7.27948,46.50529,47.79523,49.27814,50.28434,44.40954,50.64589]
)

y2=np.array([46.47318,46.24222,51.02315,47.05681,50.78464,47.6
7812,52.90274,46.10107,48.55274,48.26154,47.80754,48.58475,49.
78569,48.35275,49.90442,46.86422,47.69177,48.33888,48.50258,49
.69559,44.35069,48.80106,49.72479,48.31285,46.86678,47.33681,4
8.6243,50.93695,50.81675,48.87599,44.58654,49.38595,48.39299,4
8.52042,48.21531,45.16171,46.61168,47.86871,46.62379,49.07921]
)

plt.scatter(x1,y1,s=10,color="red",label="Set 1 N~(45,1.0)")

plt.scatter(x2,y2,s=10,color="green",label="Set 2 N~(48,2.0)")
```

```
m=-2

c=-50

def line_eq(x,y,m,c):

    return y-(m*x)-c

entropy=[]

min=10000000


while c<=50 and c>=-50:

    m=-2

    while m>=-2 and m<=2:


        belowlineSetA=[]

        abovelineSetA=[]

        belowlineSetB=[]

        abovelineSetB=[]


        for i,j in zip(x1,y1):

            if line_eq(i,j,m,c)>0:

                abovelineSetA.append((i,j))

            if line_eq(i,j,m,c)<0:

                belowlineSetA.append((i,j))


        for i,j in zip(x2,y2):

            if line_eq(i,j,m,c)>0:

                abovelineSetB.append((i,j))
```

```
            if line_eq(i,j,m,c)<0:

                belowlineSetB.append((i,j))



        if len(belowlineSetA)!=0 and len(abovelineSetA)!=0 and
(abs(len(belowlineSetA)-len(belowlineSetB))>=20 and
abs(len(abovelineSetA)-len(abovelineSetB))>=20):

            h=-
((max(len(belowlineSetA),len(abovelineSetA))/(len(belowlineSet
A)+len(abovelineSetA))*np.log2((max(len(belowlineSetA),len(abo
velineSetA))/(len(belowlineSetA)+len(abovelineSetA))))) +
(max(len(belowlineSetB),len(abovelineSetB))/(len(belowlineSetB
)+len(abovelineSetB)))*np.log2((max(len(belowlineSetB),len(abo
velineSetB))/(len(belowlineSetB)+len(abovelineSetB)))))

            if h<min and h>0:

                min=h

                entropy.append((h,m,c))

        m+=0.01

    c+=0.1



h,m,c=entropy[-1]

print("Minimum entropy of this system : ",h)





a=float(input("Enter the 1st dimension : "))

b=float(input("Enter the 2nd dimension : "))

if line_eq(a,b,m,c)>0:

    print('The given element lies in Set 2')

elif line_eq(a,b,m,c)<0:

    print('The given element lies in Set 1')
```

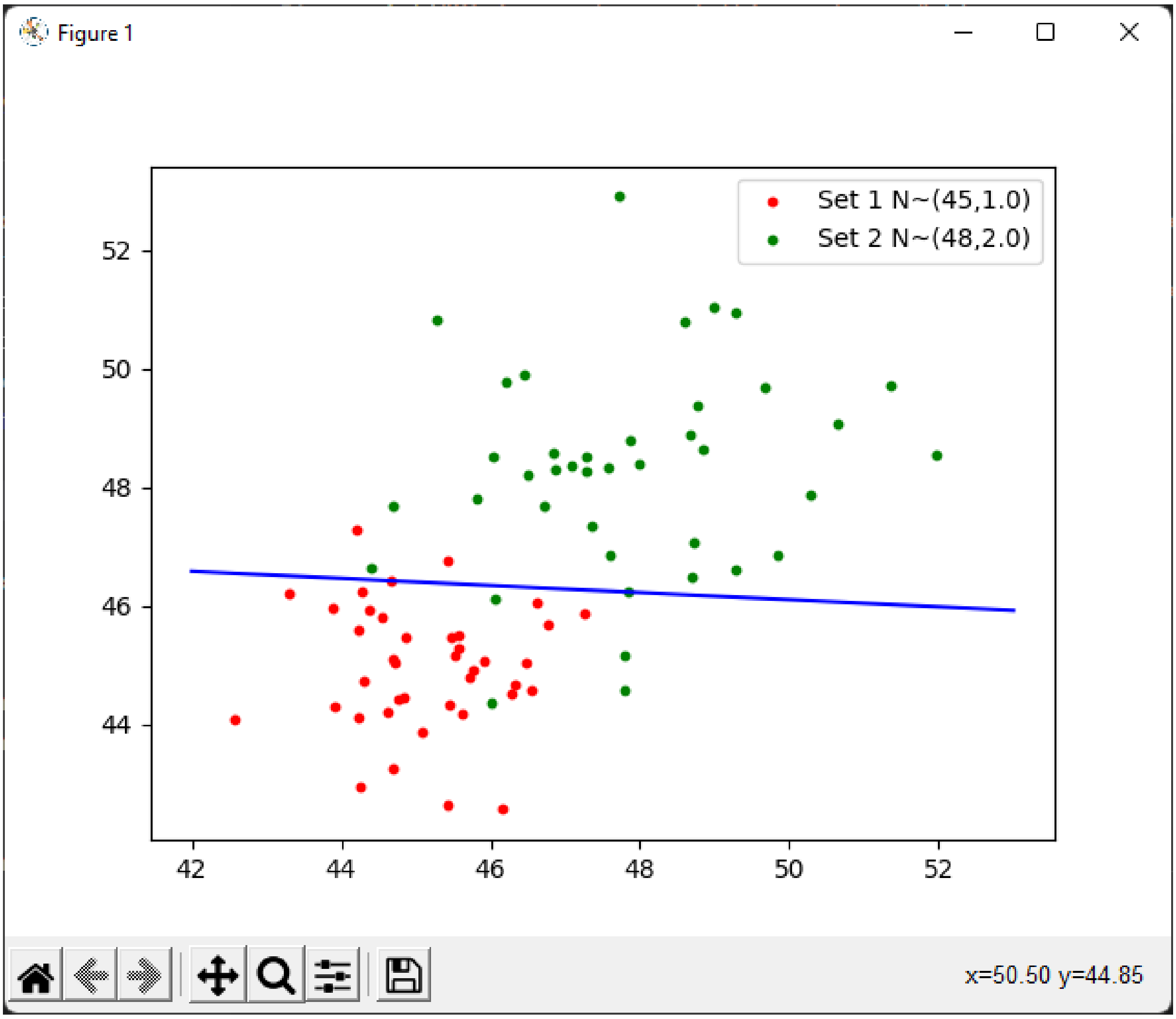
```
x=np.arange(42,54)

plt.plot(x,(m*x)+c ,color='blue')


plt.legend()

plt.show()
```

Plots with the optimum line:



Calculation for Minimum entropy:

$H= -\sum p(\log_2 p)$

n= Total no. of points in each Set= 40

n1= No. of points below the line belongs to Set 1 = 38

n2= No. of points above the line belongs to Set 2 = 36

p1=n1/n= 38/40= 0.95

p2=n2/n= 36/40= 0.9

$H=-(p1(\log_2 p1) + p2(\log_2 p2))$
$=-(0.95(-0.07400058144377693) + 0.9 (-0.15200309344504995)$
$= 0.20710333647213303$

Output:

Trail 1

Minimum entropy of this system : 0.20710333647213303

Enter the 1st dimension : 44

Enter the 2nd dimension : 46

The given element lies in Set 1

Trail 2

Minimum entropy of this system : 0.20710333647213303

Enter the 1st dimension : 49

Enter the 2nd dimension : 47

The given element lies in Set 2

Inference:

We have generated two dimensional two sets of data assuming,

Set1 with mean=45 and SD=1
Set2 with mean=46 and SD=1.5

By finding the entropy of the system by the separating line and came into conclusion that any point (mean, SD) can be classified into Set 1 or Set2.

If that point lies above the line, then the point belongs to Set2. If the point lies below the line, then the point belongs to Set 1.