Aditya Das 1001675762

Assignment 1
Problem 1

All the problems are done in Python.

Function execution :

The first file's name is DM1.py

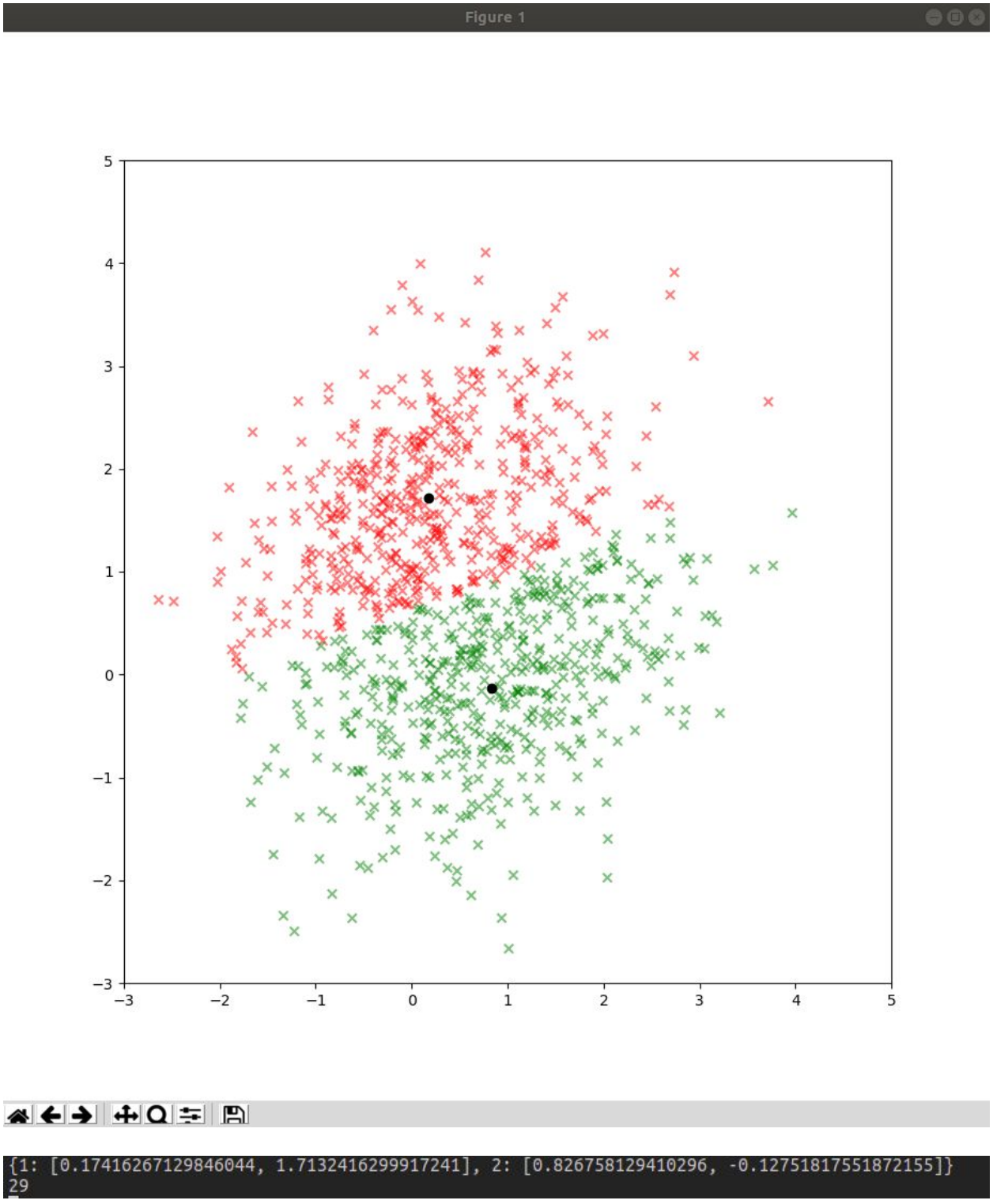
To run the file use the following command python3 DM1.py

1. The 2D Gaussian multivariate was generated using *np.random.multivariate_normal(mean1, Sigma1, 500).T.*
2. The centers and number of clusters for the sub problems are hard coded

```
30  Sigma1 = [[0.9,0.4],[0.4,0.9] ]
31  Sigma2 = [[0.9,0.4],[0.4,0.9] ]
32  #c=[[10,10],[-10,-10]]                          #centers for Part 2 of Question 1
33  c=[[10,10],[-10,-10],[10,-10],[-10,10]]         #centers for Part 3 of Question 1
34  #k=2                                            # cluster for Part 2 of Question 1
35  k=4                                             # cluster for Part 3 of Question 1
```

   They can be changed/swapped out by the values required. (Line 32:34)

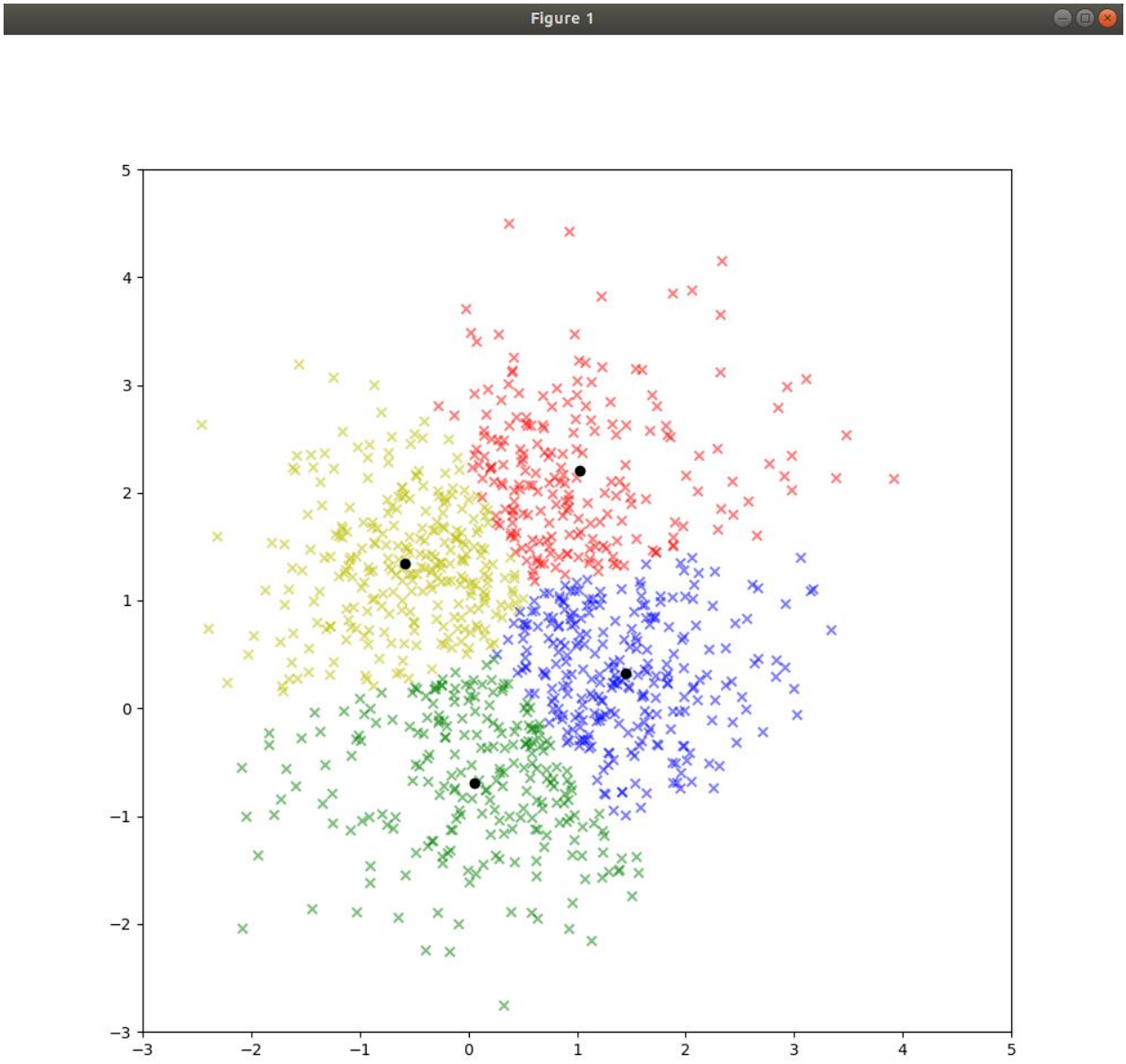
3. The numpy arrays concatenated and stored in pandas Dataframes.
4. The centers are then stored in a dictionary called centroids. These are dynamic and can change according to the number of centers given.(Max centers supported is 6)
5. The centers are then assigned to the generated dataset using center_assignment() function. The l2 norm is found and the least distance center is assigned to the point.
6. After assignment of centers, the old centers are copied using deepcopy and a loop runs until the difference between the old and new loops is less than 0.001 or the number of iterations reach 10000.
7. The loop updates the new centers found using the mean and assigns the new centers.
8. Finally the following graphs were generated



The above figure is 2 clustering 2 center graph.

The centers are shown in the snippet.

The number of iterations are 29.(30 including the first assignment)




```
{1: [1.0258480795731193, 2.201339832897128], 2: [0.05316939651517685, -0.6955382409110844], 3: [1.444085
7328372123, 0.3198895747501936], 4: [-0.5885478237211383, 1.3400724096043108]}
19
```

The above figure is 4 clustering 4 center graph.

The centers are shown in the snippet.

The number of iterations are 19.(20 including the initial assignment)

Problem 2

Function execution :

In this question sub question 1 till 3, plots are generated by running the file DM2.py

For sub question 4, the plots are generated using the file DM24.py

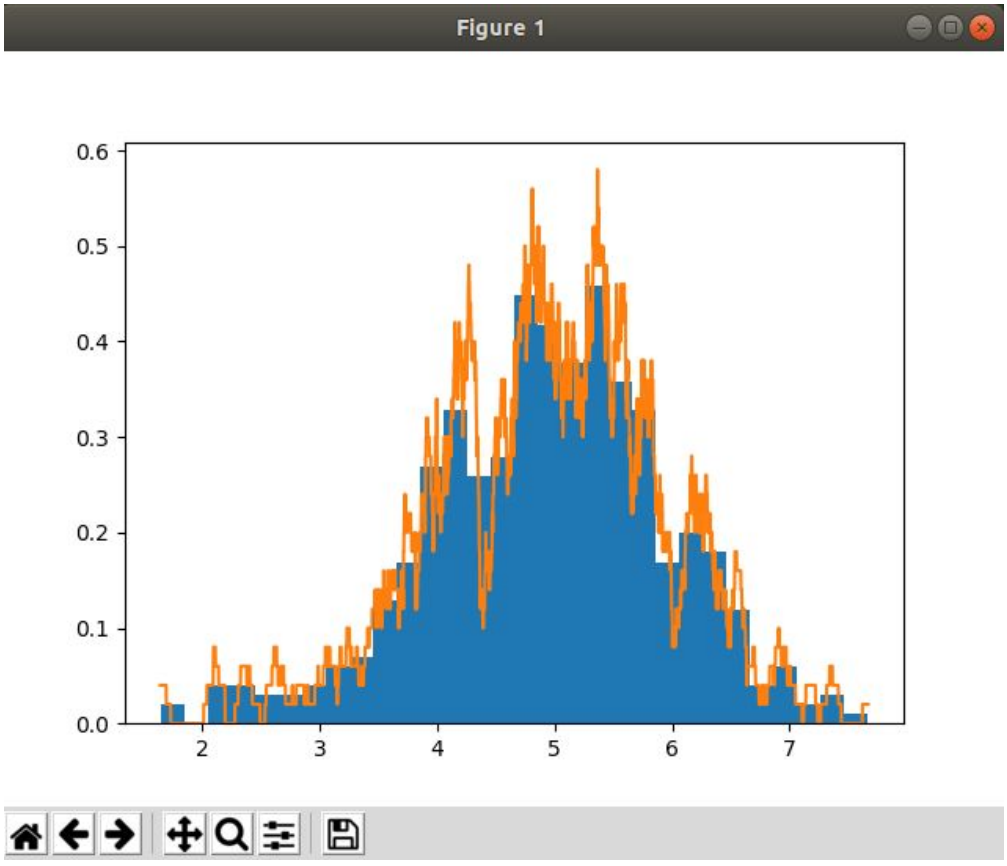
To run the files use the command: python3 DM2.py or python3 DM24.py

1.The 1D Random Gaussian data is generated using Sigma1 * np.random.randn(500) + mean1

2. The capital **X** is assumed as a discrete value from the min of x to the max of x with a step count of 0.001.

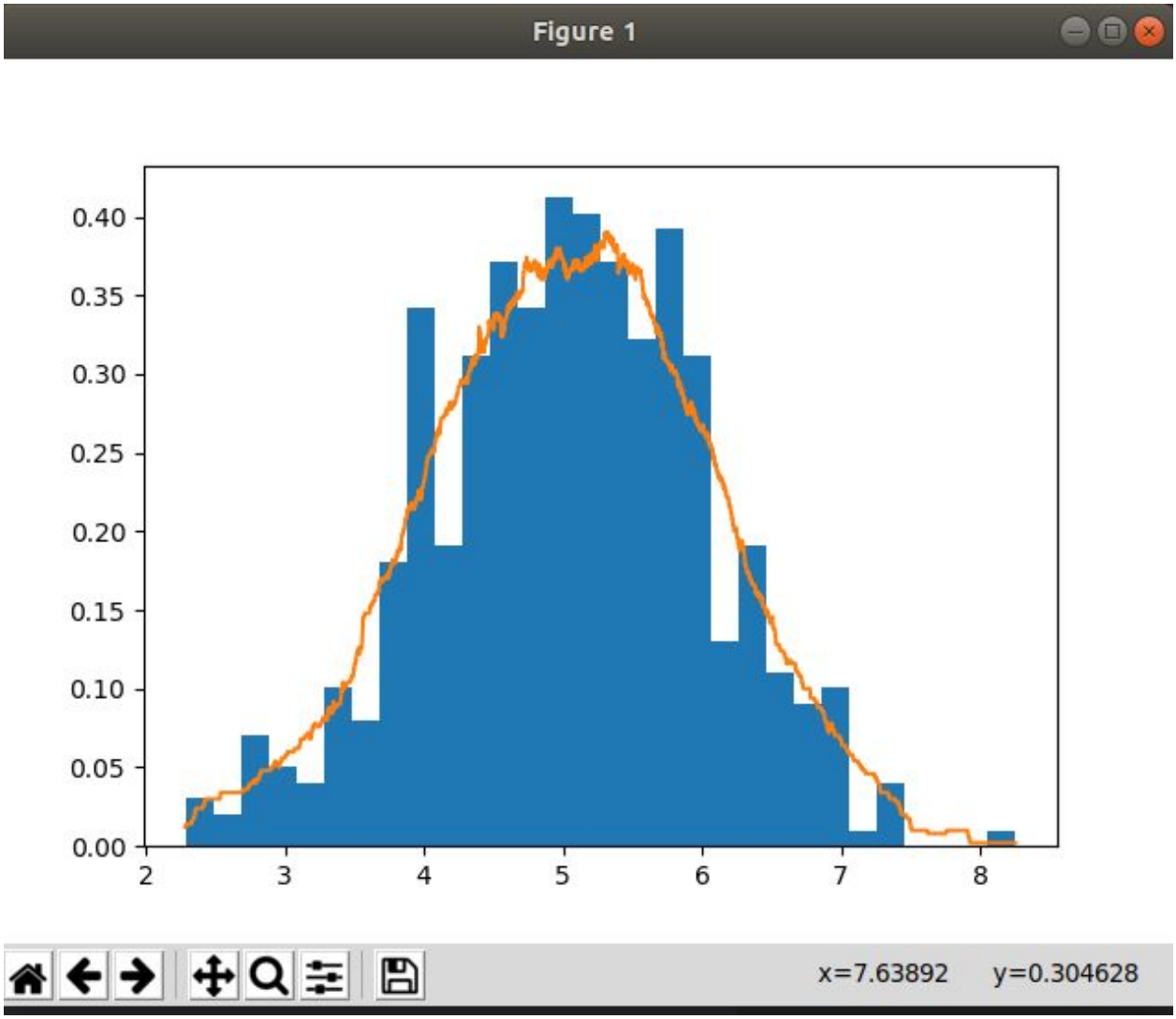
3. The h values are iterated over the entire file.

4. For part 2 of the question, the x value generated is used in the Kernel density estimation formula.

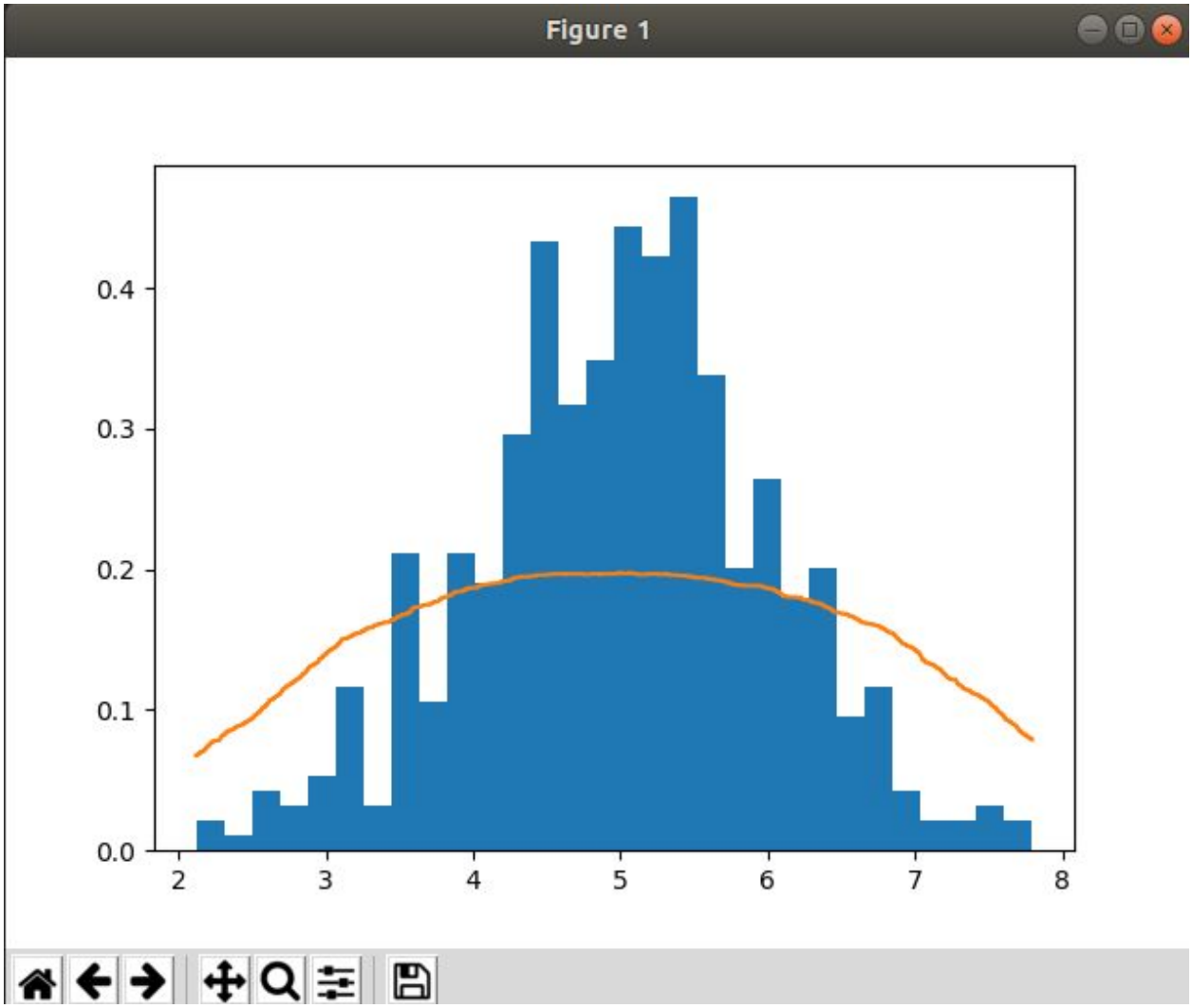


This plot is generated when mean = 5 sigma = 1 and h = 0.1

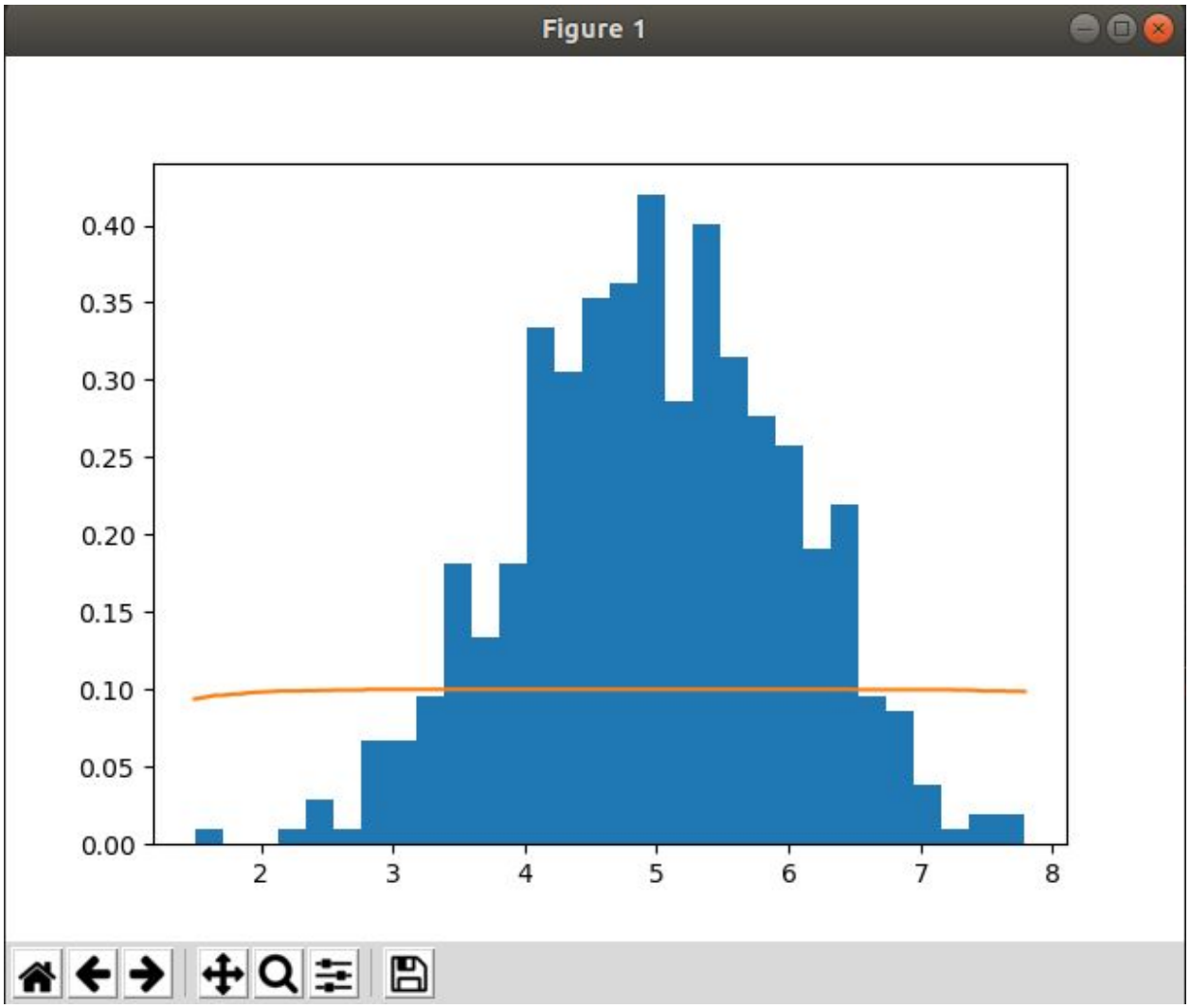
To get the next plots, we need to exit the generated plot.



This plot is generated when mean = 5 sigma = 1 and h = 1



This plot is generated when mean = 5 sigma = 1 and h = 5

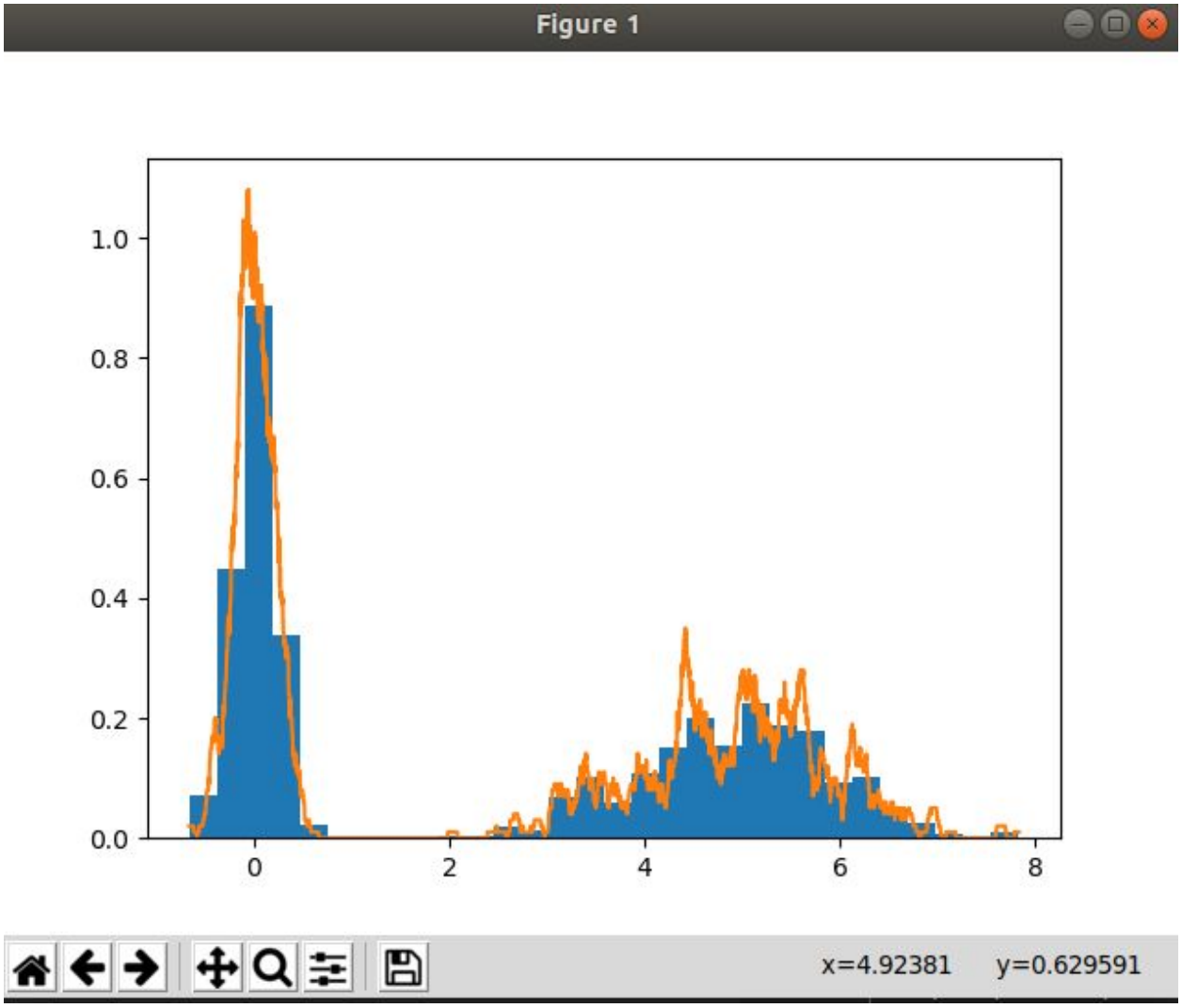


This plot is generated when mean = 5 sigma = 1 and h = 10

5. Subquestion 3 can be solved by uncommenting two lines in the file.(Line 15 and 16)

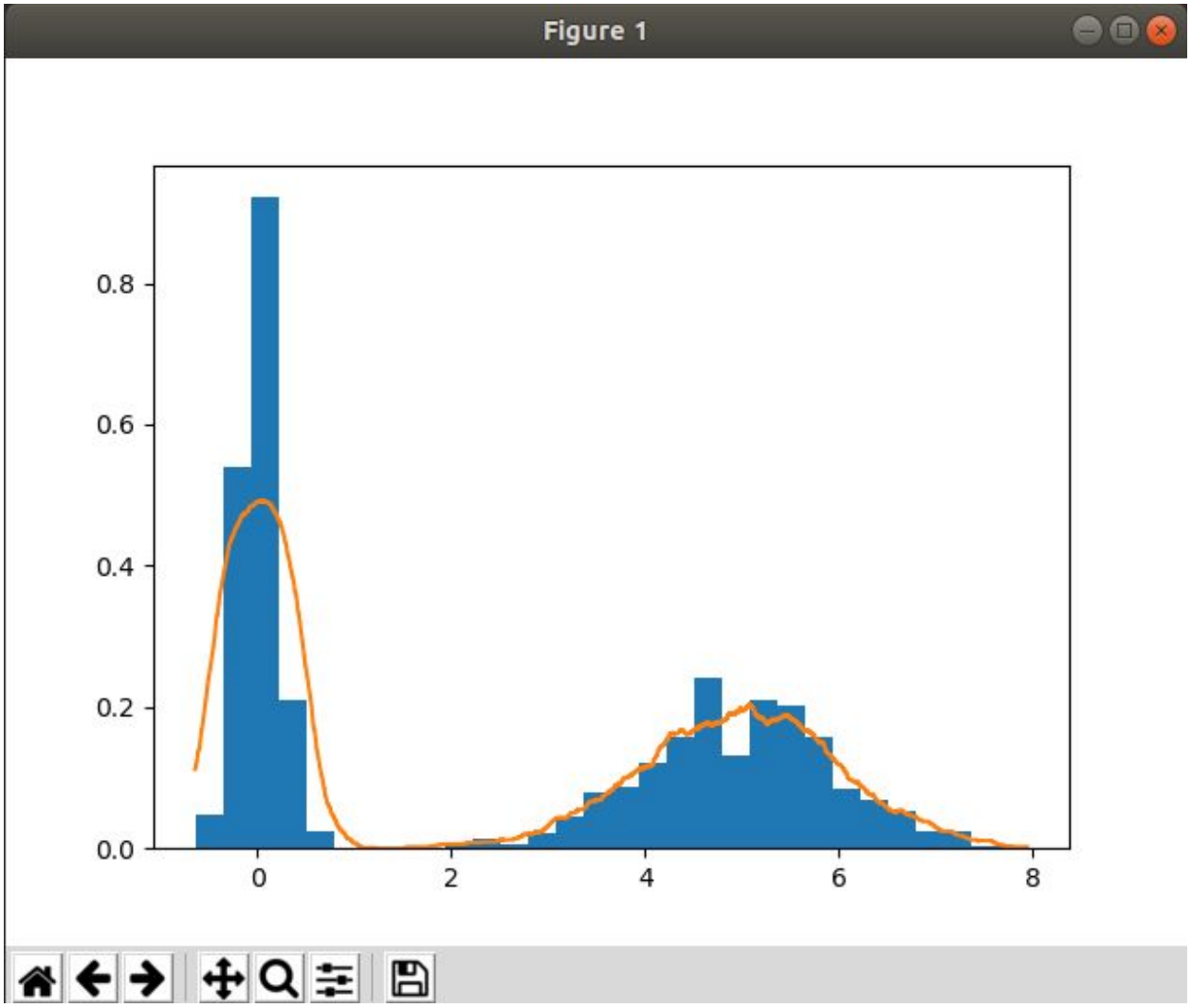
```
14      x = Sigma1 * np.random.randn(500) + mean1
15      #y = Sigma2 * np.random.randn(500) + mean2                    # 2nd set of Gaussian Data.
16      #x = np.concatenate((x, y))                                   # Concatenating the 2 sets of Gaussian Data.
17      X = np.arange(min(x),max(x) , 0.001)
```

6. The following graphs are generated after uncommenting the two lines.

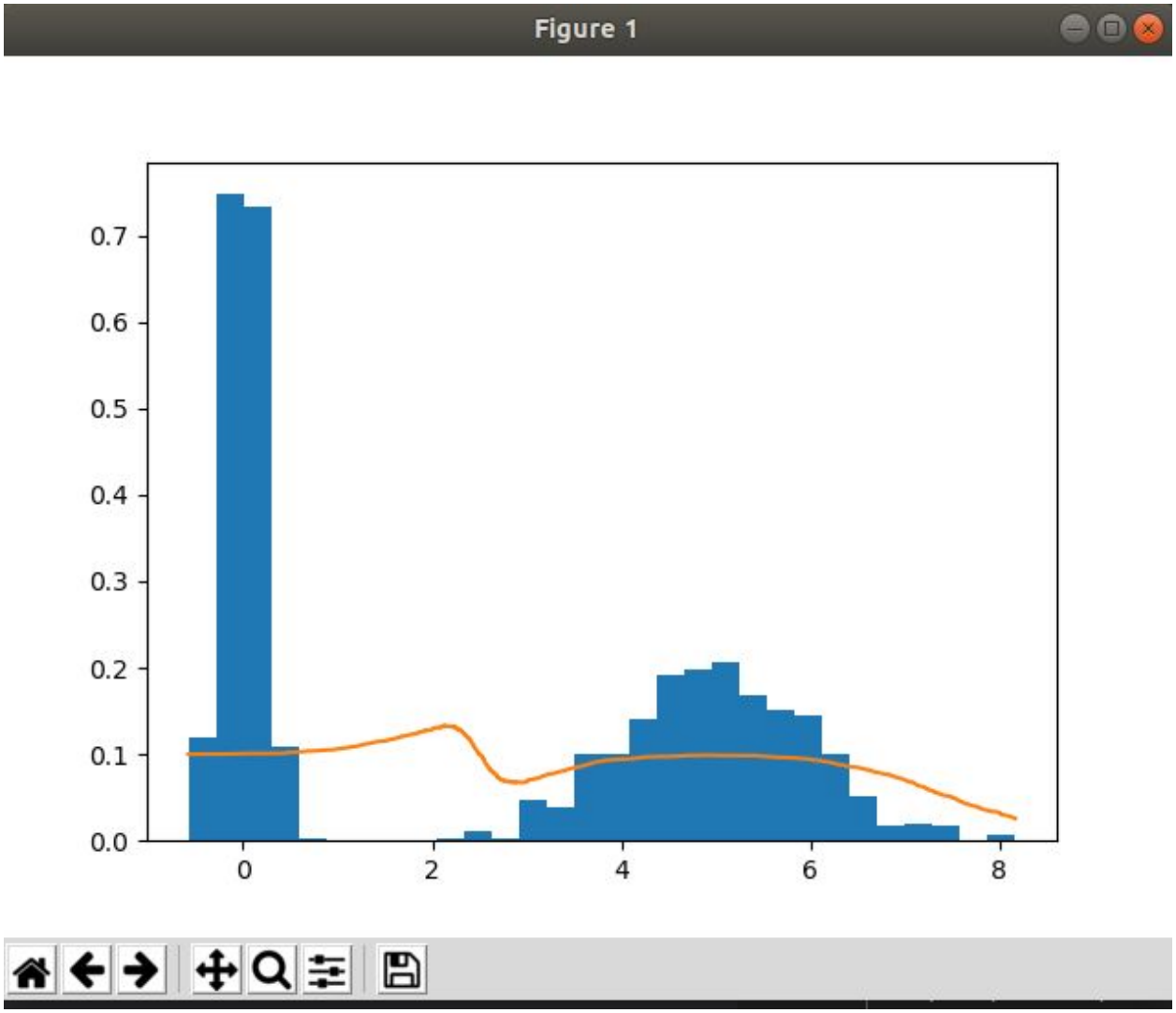


This plot is generated by combining two data sets, where the mean1 = 5, sigma1 = 1 and mean1 = 0, sigma1 = 2 and h =0.1

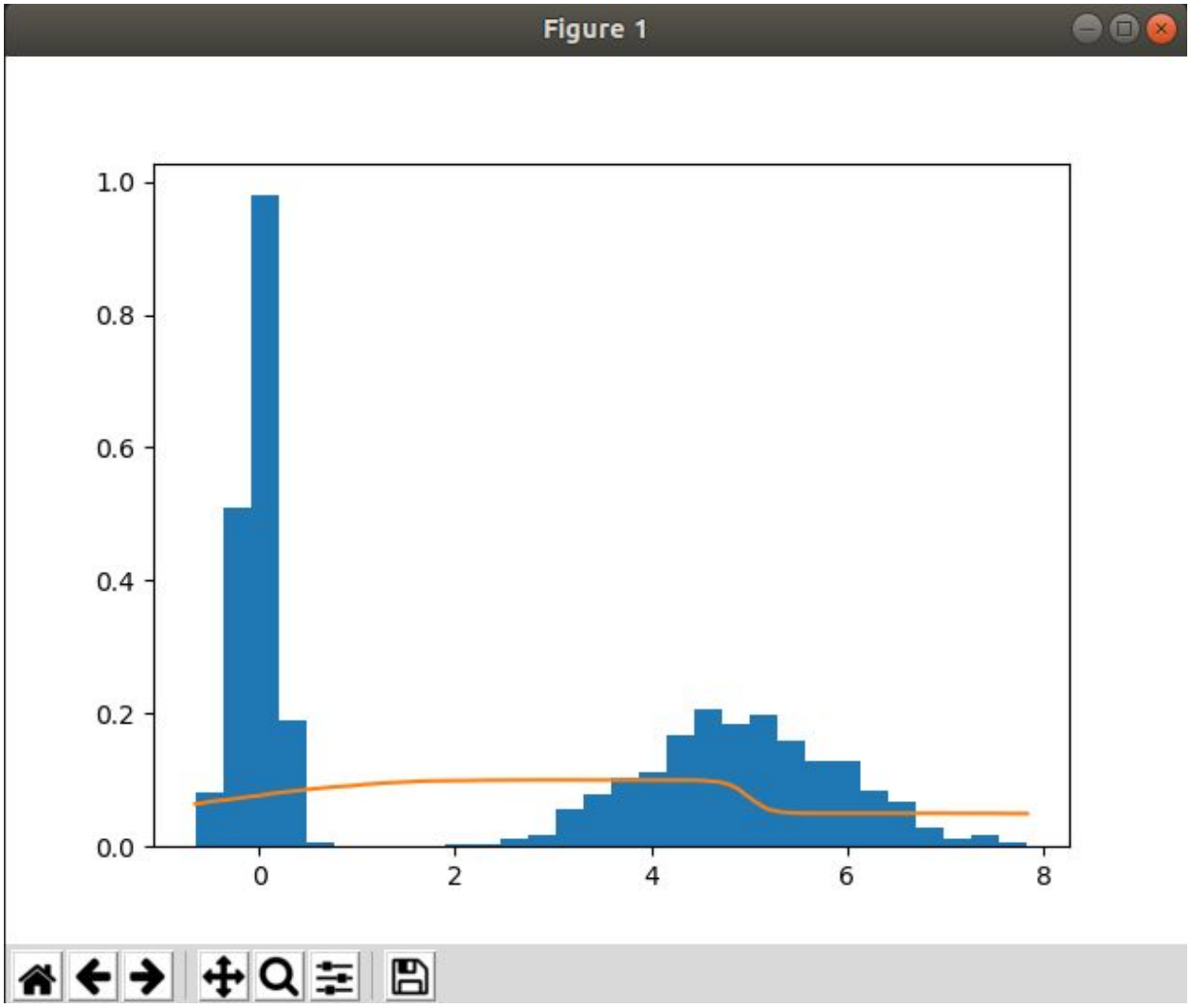
To get the next plots, we need to exit the generated plot.



This plot is generated by combining two data sets, where the mean1 = 5, sigma1 = 1 and mean1 = 0, sigma1 = 2 and h =1



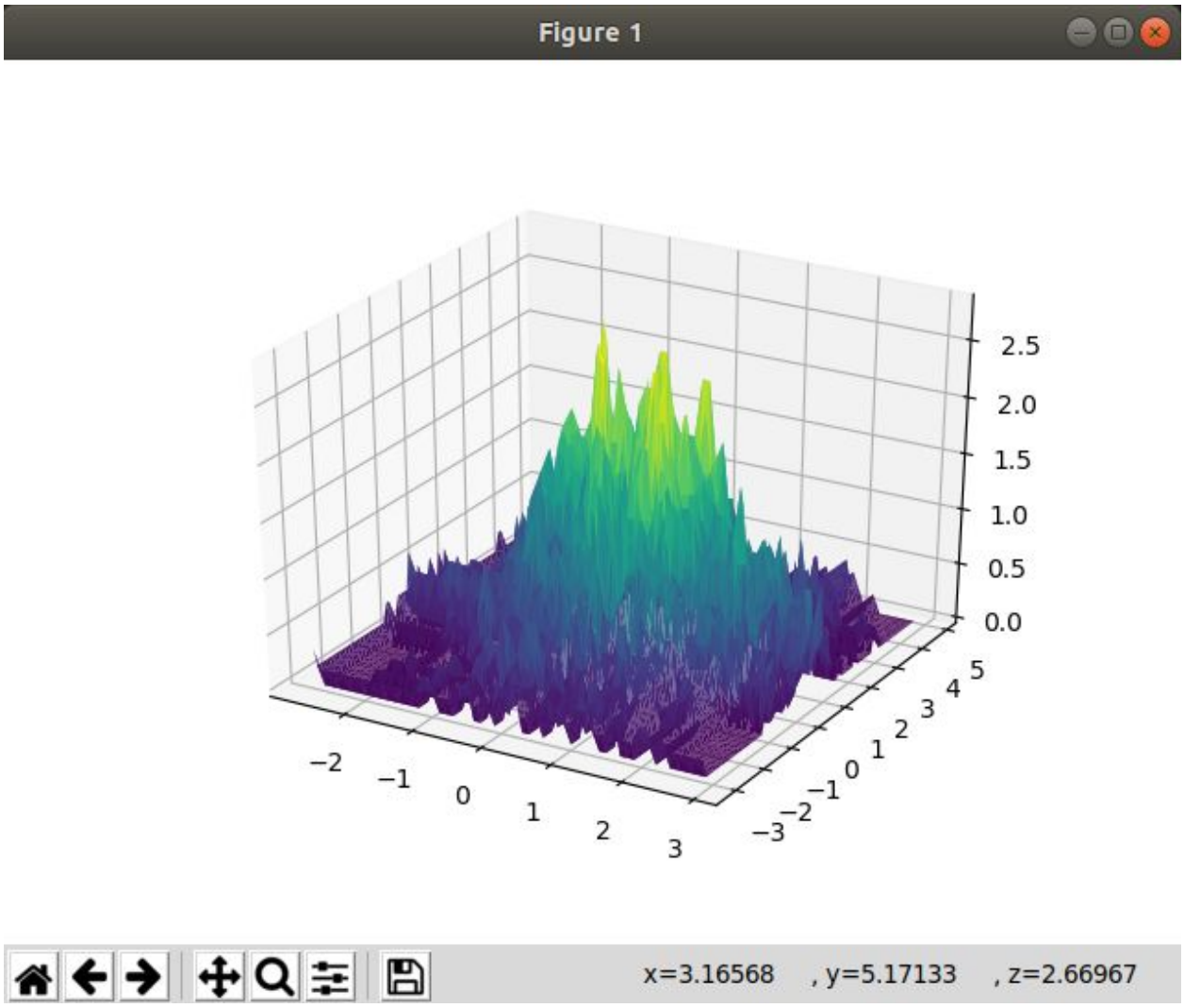
This plot is generated by combining two data sets, where the mean1 = 5, sigma1 = 1 and mean1 = 0, sigma1 = 2 and h =5



This plot is generated by combining two data sets, where the mean1 = 5, sigma1 = 1 and mean1 = 0, sigma1 = 2 and h =10

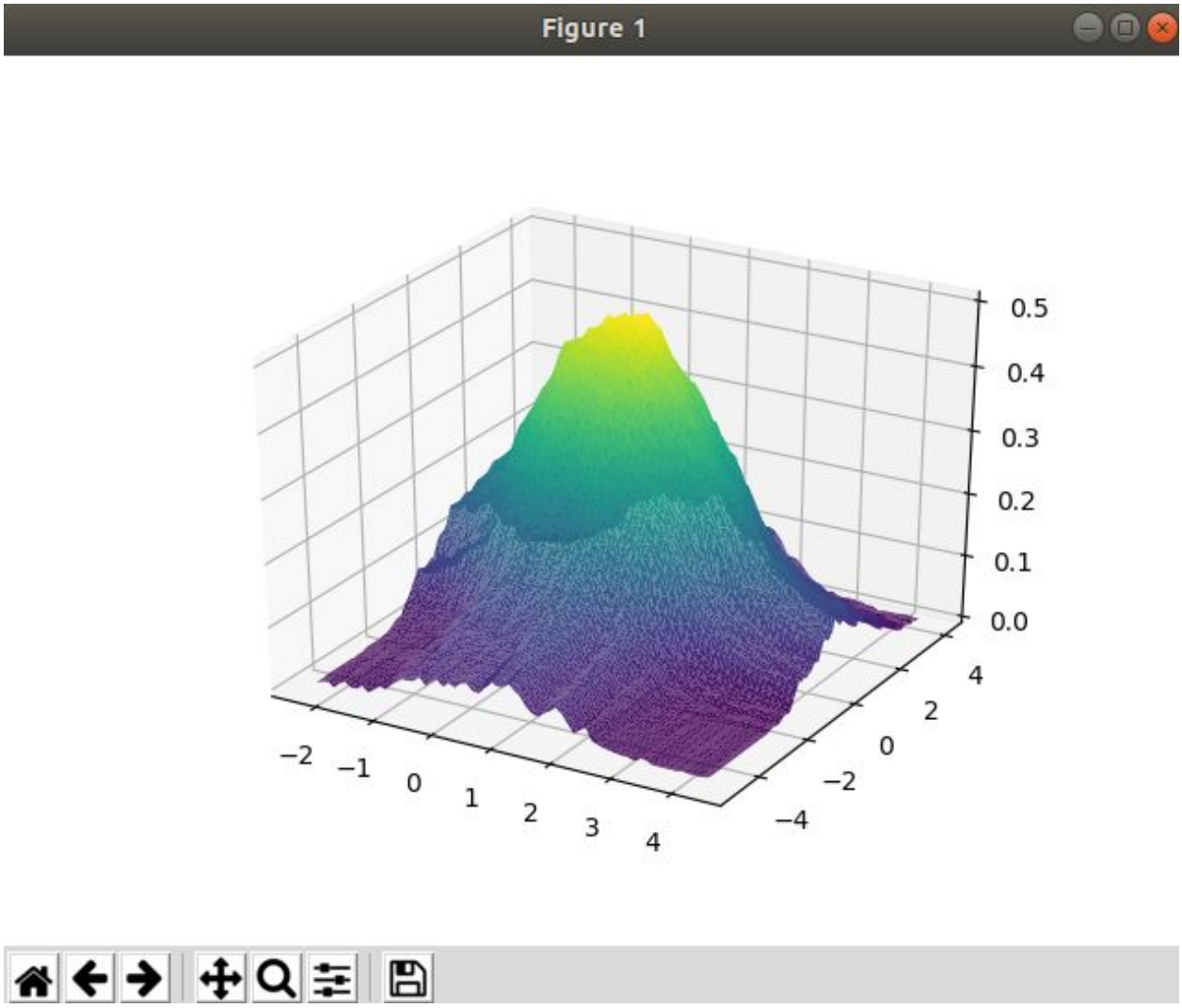
7. The subquestion 4 is solved by executing the file dm22.py

8.



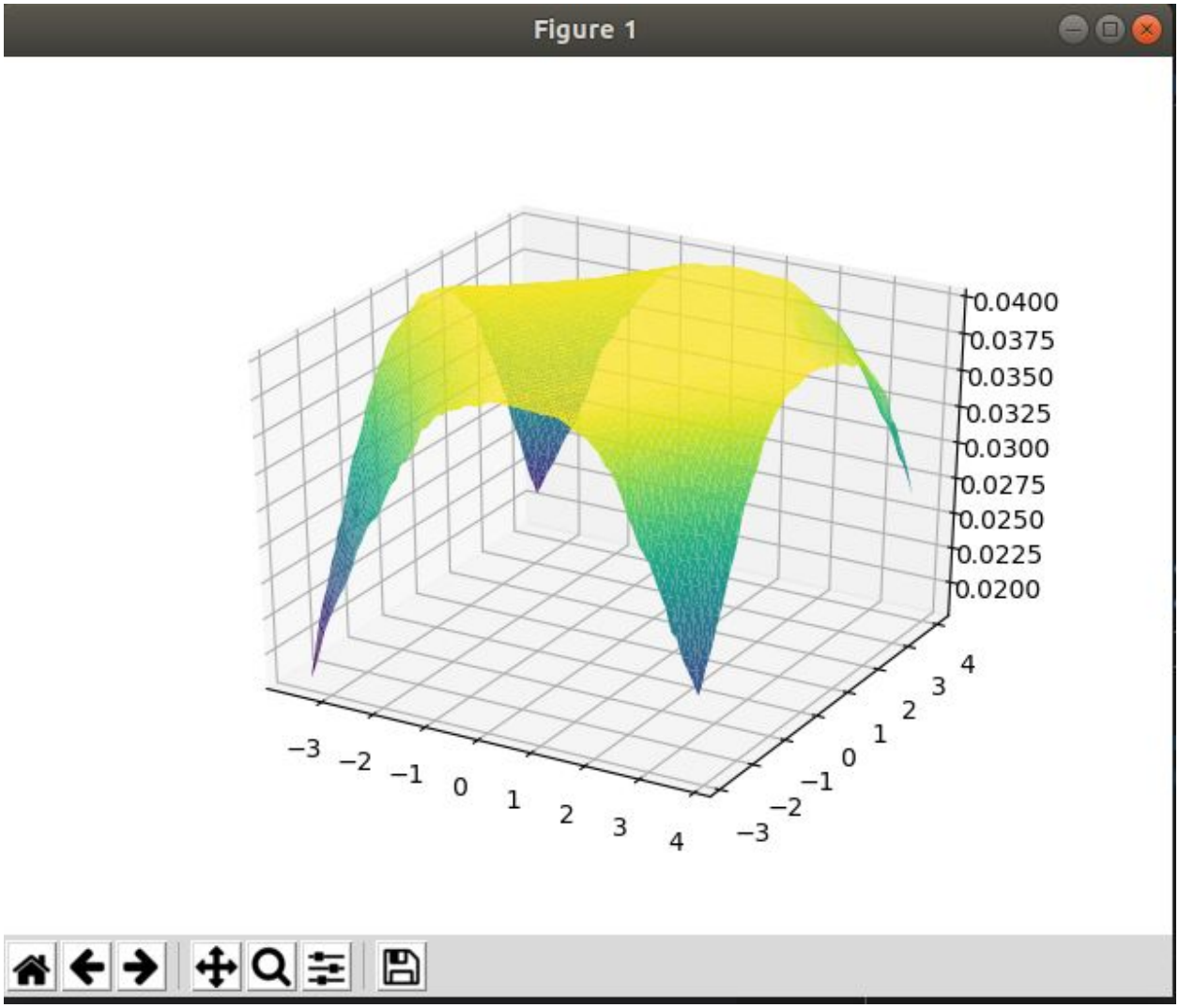
This plot is generated by combining two data sets, where the μ 1 = [1, 0], μ 2 = [0, 1.5],

Σ 1 =[0.9, 0.4; 0.4,0.9], Σ 2 =[0.9, 0.4; 0.4,0.9] and h = 0.1



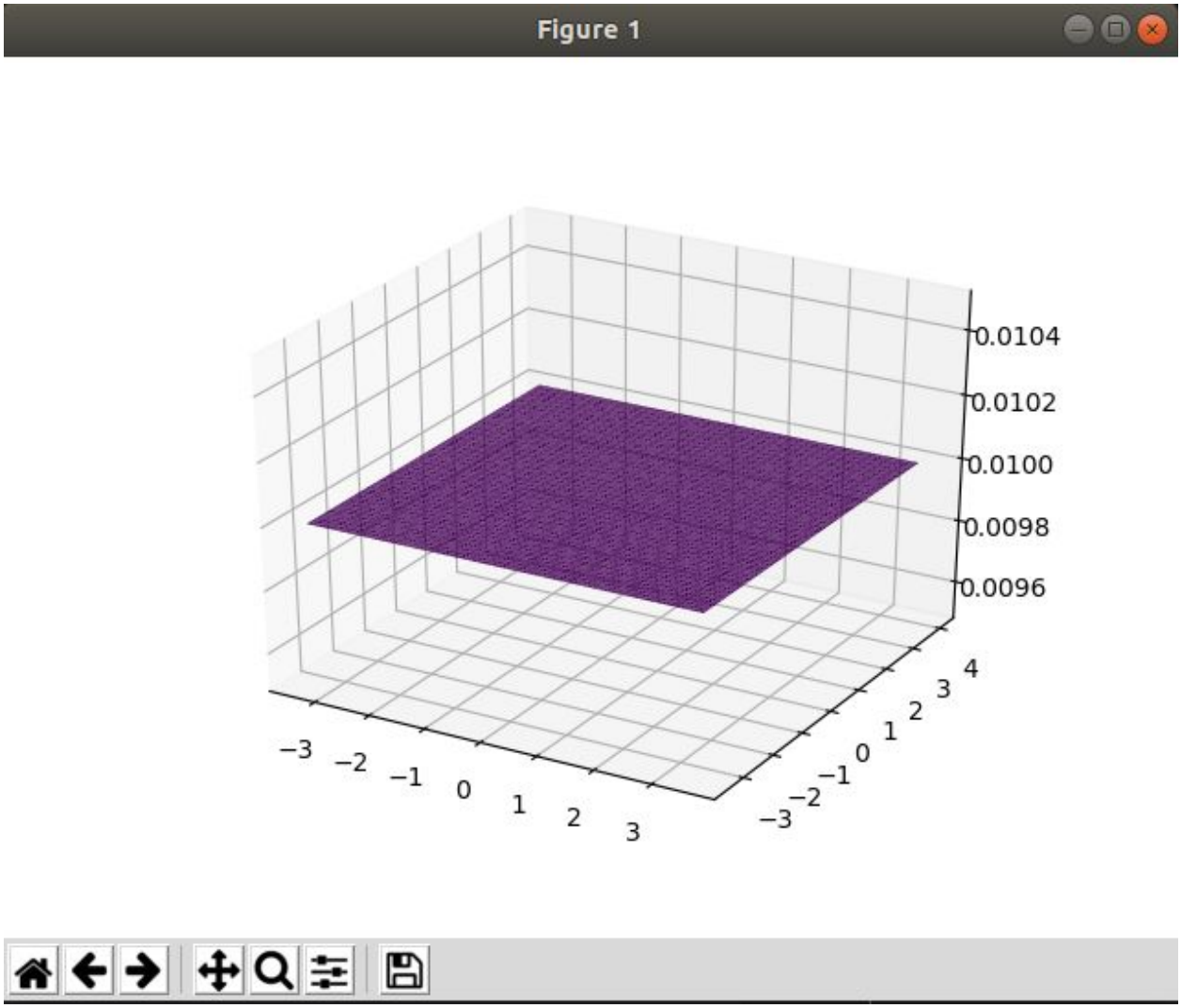
This plot is generated by combining two data sets, where the μ 1 = [1, 0], μ 2 = [0, 1.5],

Σ 1 =[0.9, 0.4; 0.4,0.9], Σ 2 =[0.9, 0.4; 0.4,0.9] and h = 1



This plot is generated by combining two data sets, where the μ 1 = [1, 0], μ 2 = [0, 1.5],

Σ 1 =[0.9, 0.4; 0.4,0.9], Σ 2 =[0.9, 0.4; 0.4,0.9] and h =5



This plot is generated by combining two data sets, where the μ 1 = [1, 0], μ 2 = [0, 1.5],

Σ 1 =[0.9, 0.4; 0.4,0.9], Σ 2 =[0.9, 0.4; 0.4,0.9] and h =10