Merlin Carson
CS446, Spring 2018

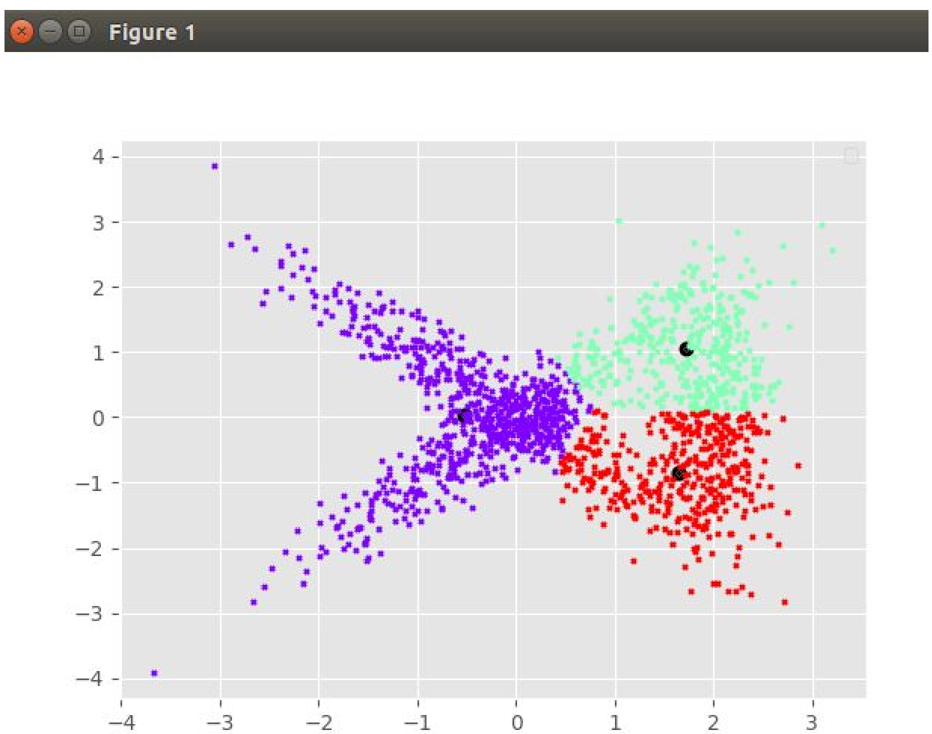
EM-Algorithms

1. K-Means -

Using Python2.7 and the class slides, I fairly easily implemented the k-means algorithm from scratch. The only libraries I used for this was np.linagl.norm for calculating euclidean distance between data and centroids and matplotlib for drawing the clusters. I ran the algorithm for K=3, K=5, and K=10 as the number of clusters. I ran the algorithm 10 times for each of those numbers of clusters and choose the run with the lowest MSE. Below I've including the results of the 3 experiments.

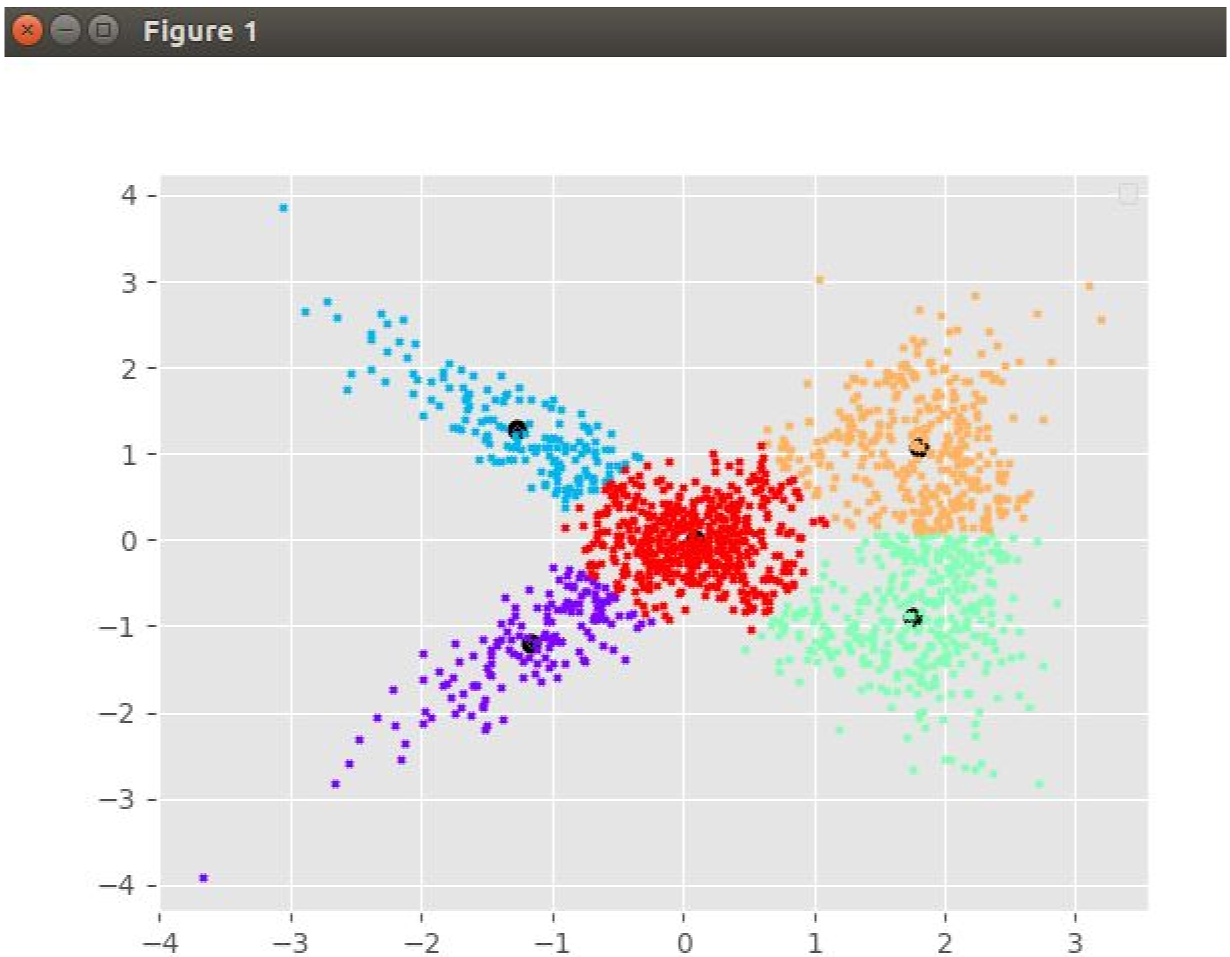
K = 3

```
MSE:  582398.0295824544
Converged after 10 iterations.
Plotting...
```



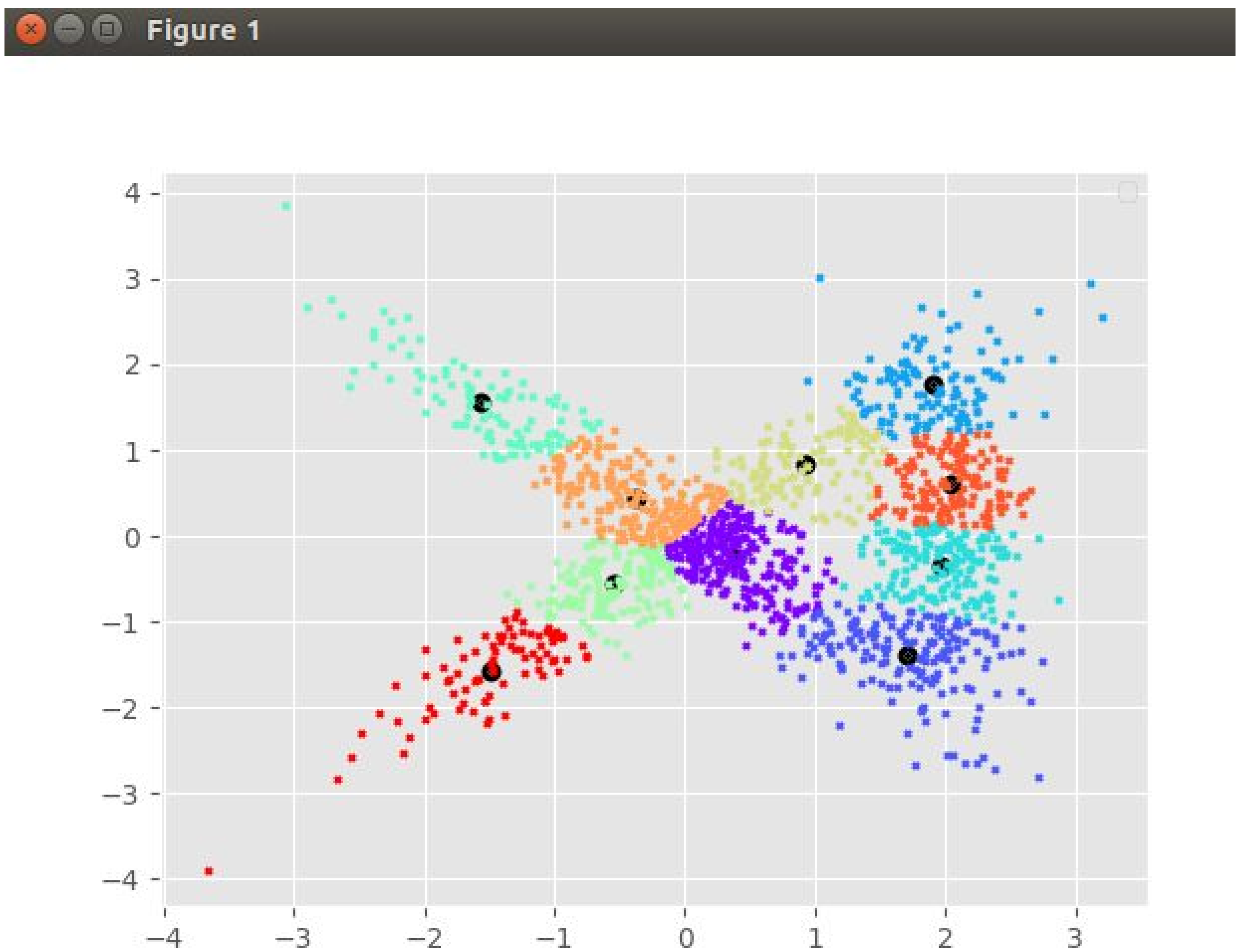
K=5

```
MSE:  376387.2643944216
Converged after 10 iterations.
Plotting...
```



K=10

```
How many clusters would you like: 10
How many iterations of k-means would you like: 5
MSE:  172843.36282012748
Converged after 20 iterations.
Plotting...
```



2. Gaussian Mixture Model

Programming the code for the Gaussian Mixture Model was a significantly harder task. After a little research and reviewing the class slides, I began my programming. The only library I needed for this part of the task was scipi.stats multivariate_normal for calculating the covariance matrices and the responsibilities. I also used Python's math library for the constant pi, the exponential function and the log function. My initial experiments classified all points inside of one Gaussian bubble. After some debugging, I determined that I was incorrectly calculating my covariance matrices. After fixing that my algorithm converged correctly. When K=3, the number of classes in the data set, the algorithm classified the points almost to perfection. I also ran the algorithm with K=2, K=4, and K=5. See results below.

K=2

```
('log likelihood: ', -4449.919599518197)
Means-

[-0.11598922 -0.00684397]

[1.94054629 0.17000736]

Covariance Matricies-

[ 0.96506927 -0.09561152]
[-0.09561152  0.87866293]

[ 0.12109816 -0.01151972]
[-0.01151972  1.40155003]

Priors-

0.6447907559490051

0.3552092440509949

Plotting...
```

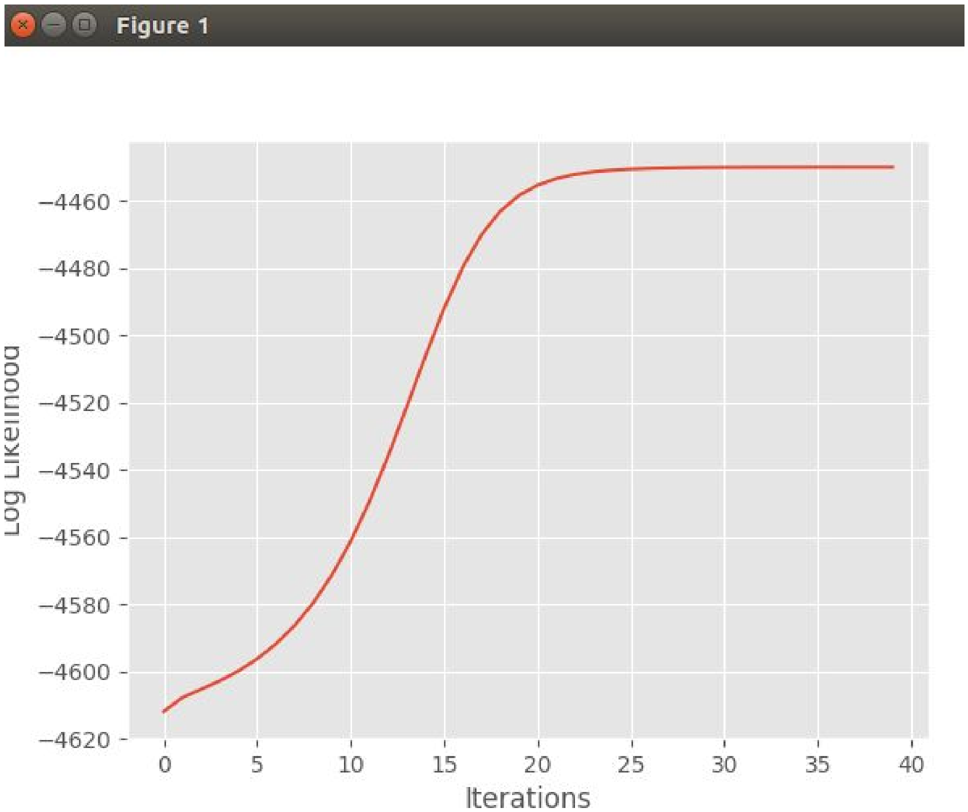


Figure 1
Log Likelihood
−4460
−4480
−4500
−4520
−4540
−4560
−4580
−4600
−4620
0
5
10
15
20
25
30
35
40
Iterations

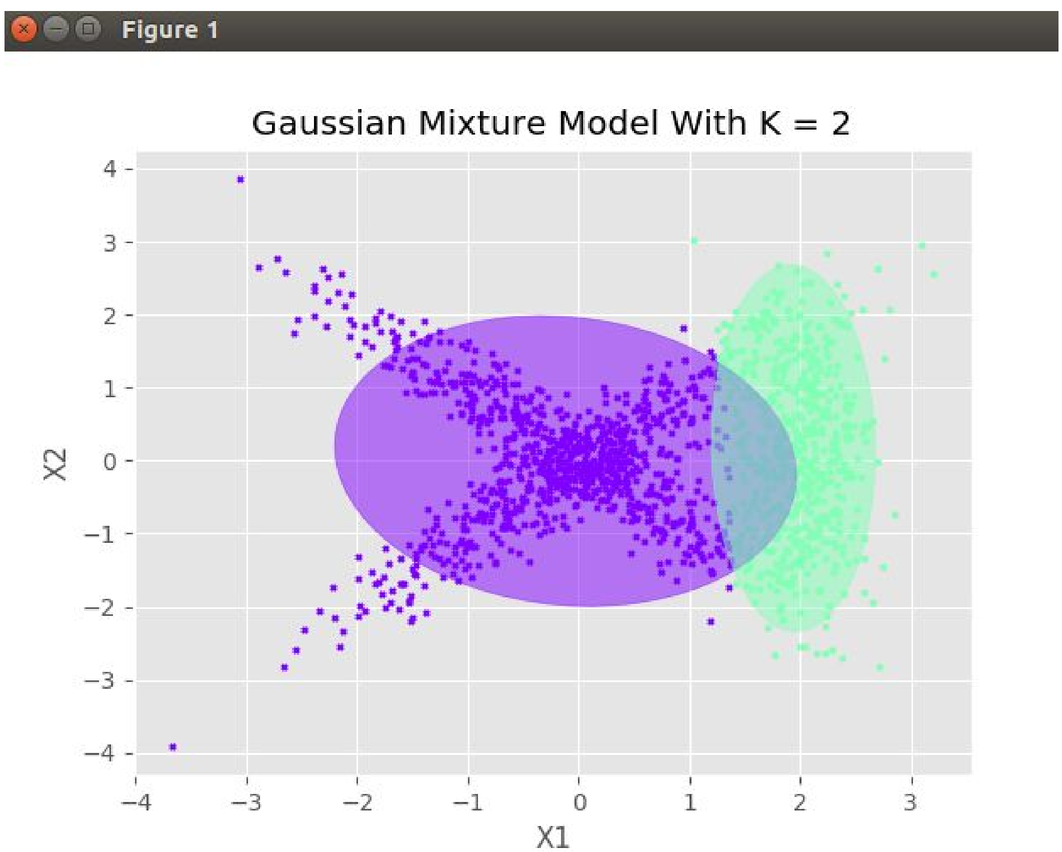


Figure 1
Gaussian Mixture Model With K = 2
X2
4
3
2
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1
2
3
X1

K=3

```
('log likelihood: ', -3927.3650764932295)
Means-

[ 0.00425201 -0.03969205]

[-0.09320913  0.11551842]

[1.96450773 0.09198051]

Covariance Matricies-

[1.05409804 0.99887186]
[0.99887186 1.06750198]

[ 1.05512966 -0.9869312 ]
[-0.9869312   1.03569525]

[ 0.10283618 -0.01077937]
[-0.01077937  1.0983756 ]

Priors-

0.33386836796317804

0.3380101325249987

0.3281214995118235

Plotting...
```

-

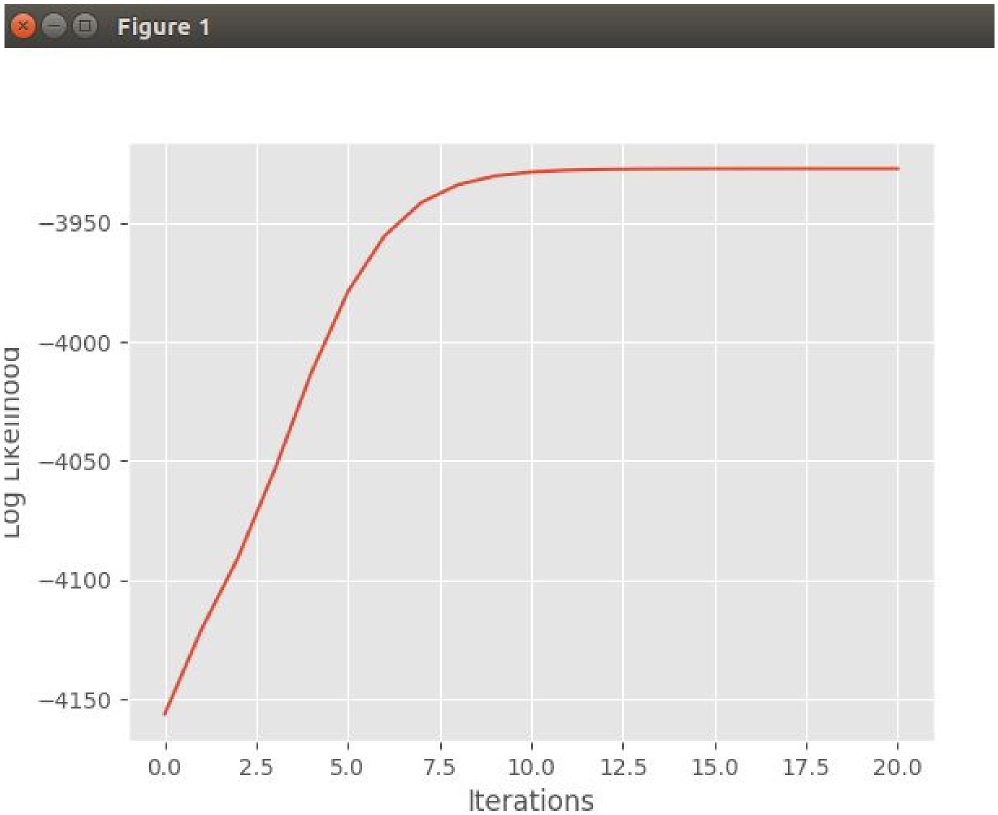


Figure 1
Log Likelihood
Iterations
−3950
−4000
−4050
−4100
−4150
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

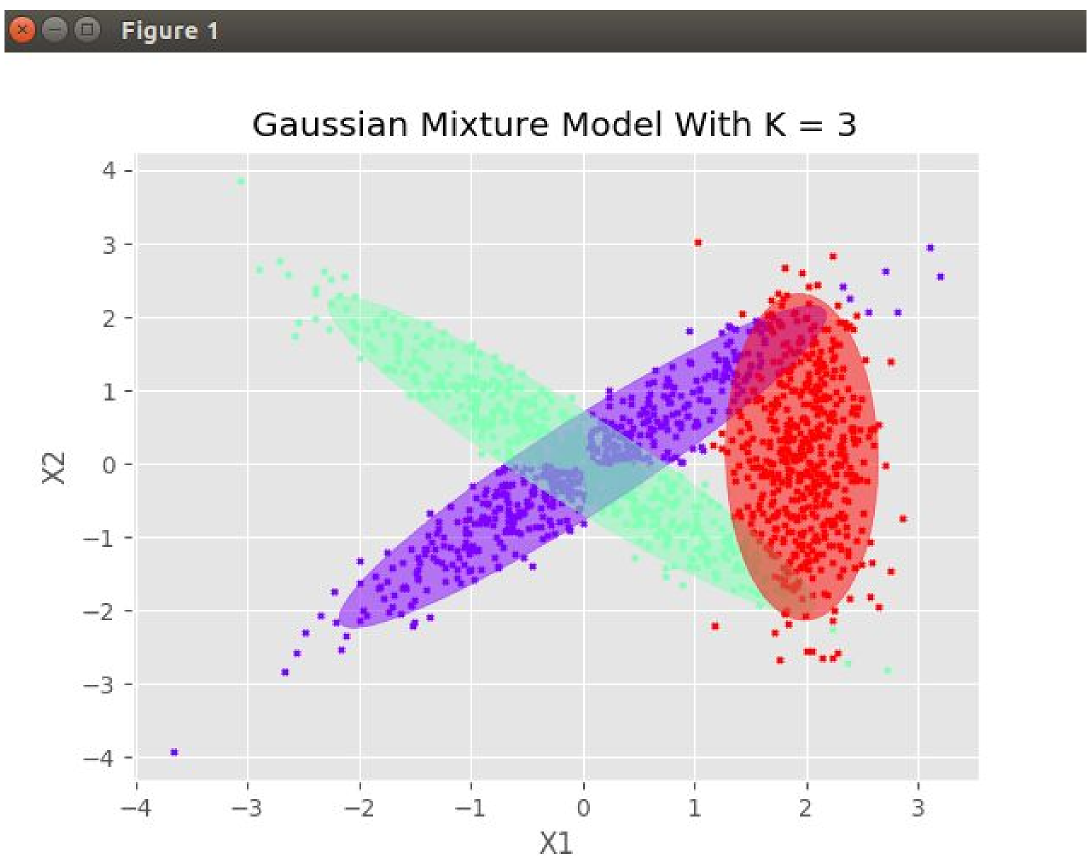


Figure 1
Gaussian Mixture Model With K = 3
X2
X1
4
3
2
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1
2
3

K=4

```
('log likelihood: ', -3923.333527010868)
Means-

[-0.10557344  0.12123675]

[2.04517193 0.96267517]

[ 1.91127569 -0.43871483]

[ 0.01368561 -0.03393161]

Covariance Matricies-

[ 1.04742846 -0.98805435]
[-0.98805435  1.04429006]

[ 0.08990954 -0.07909948]
[-0.07909948  0.56058257]

[ 0.10307781 -0.03901741]
[-0.03901741  0.67034965]

[1.0652867  1.00486085]
[1.00486085 1.0696691 ]

Priors-

0.33471622402929624

0.12213647846738546

0.20690714385985726

0.3362401536434612

Plotting...
```

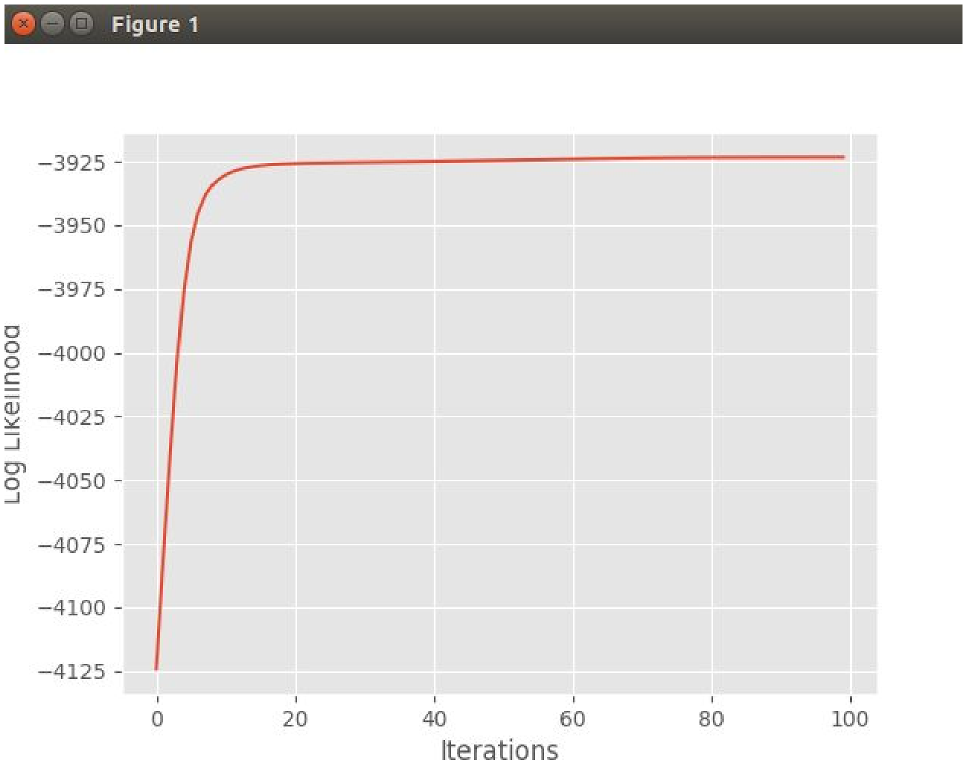


Figure 1
Log Likelihood
Iterations

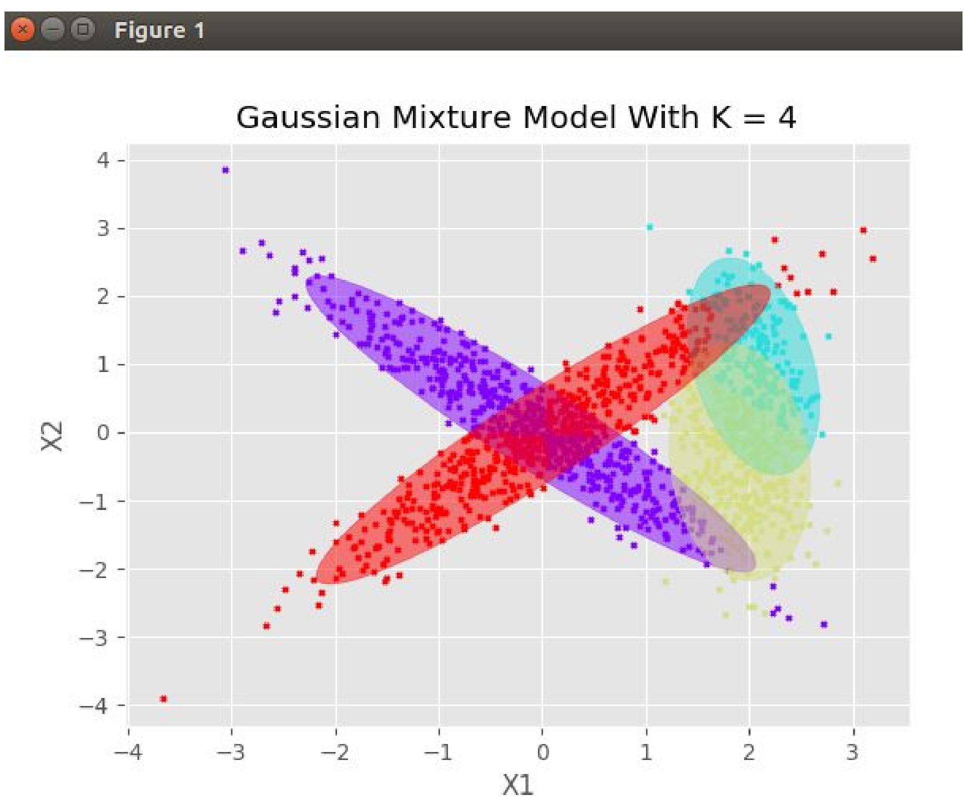


Figure 1
Gaussian Mixture Model With K = 4
X1
X2

K=5

```
('log likelihood: ', -3919.9792290408104)
(11, ' in a cluster, resetting!')
Means-

[-0.01012291 -0.04092371]

[ 0.17345901 -0.09615668]

[2.01739539 0.79534488]

[-0.12864707  0.13752146]

[ 1.9033811  -0.60894499]

Covariance Matricies-

[1.09885908 1.05856124]
[1.05856124 1.13631885]

[0.27520844 0.06604704]
[0.06604704 0.05479178]

[ 0.09559095 -0.05865564]
[-0.05865564  0.63122209]

[ 1.0655972 -1.0145897]
[-1.0145897  1.0764256]

[ 0.10386589 -0.03492633]
[-0.03492633  0.5837    ]

Priors-

0.3098522350273819

0.03666897953069773

0.16440167456964774

0.32050543291865446

0.16857167795361758
```

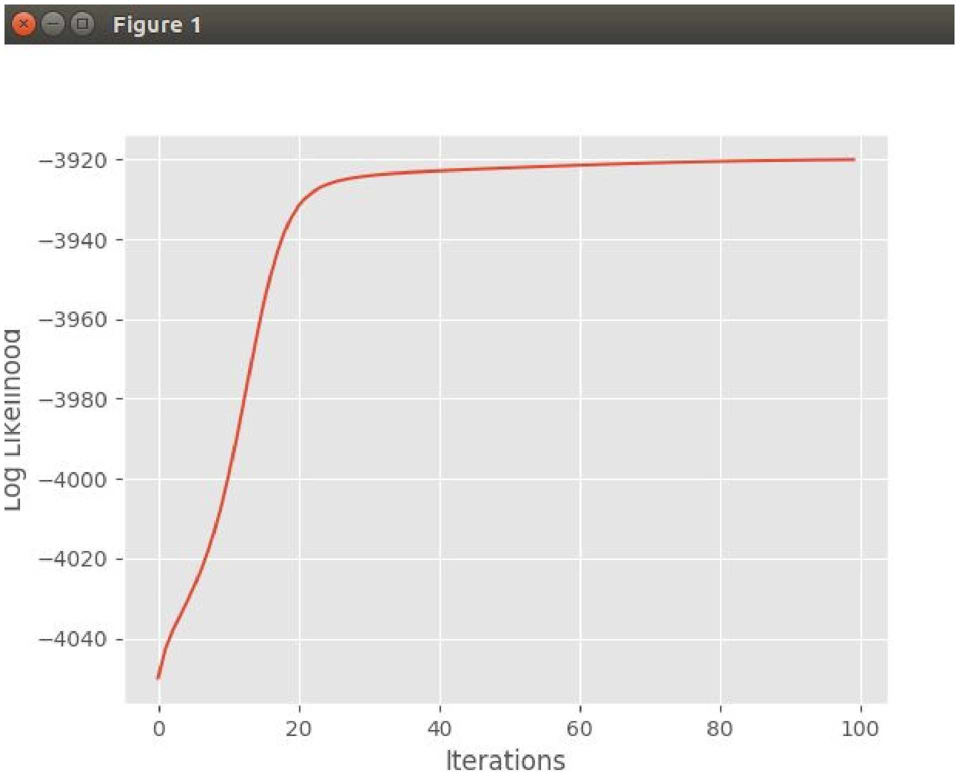


Figure 1
Log Likelihood
Iterations
−3920
−3940
−3960
−3980
−4000
−4020
−4040
0
20
40
60
80
100

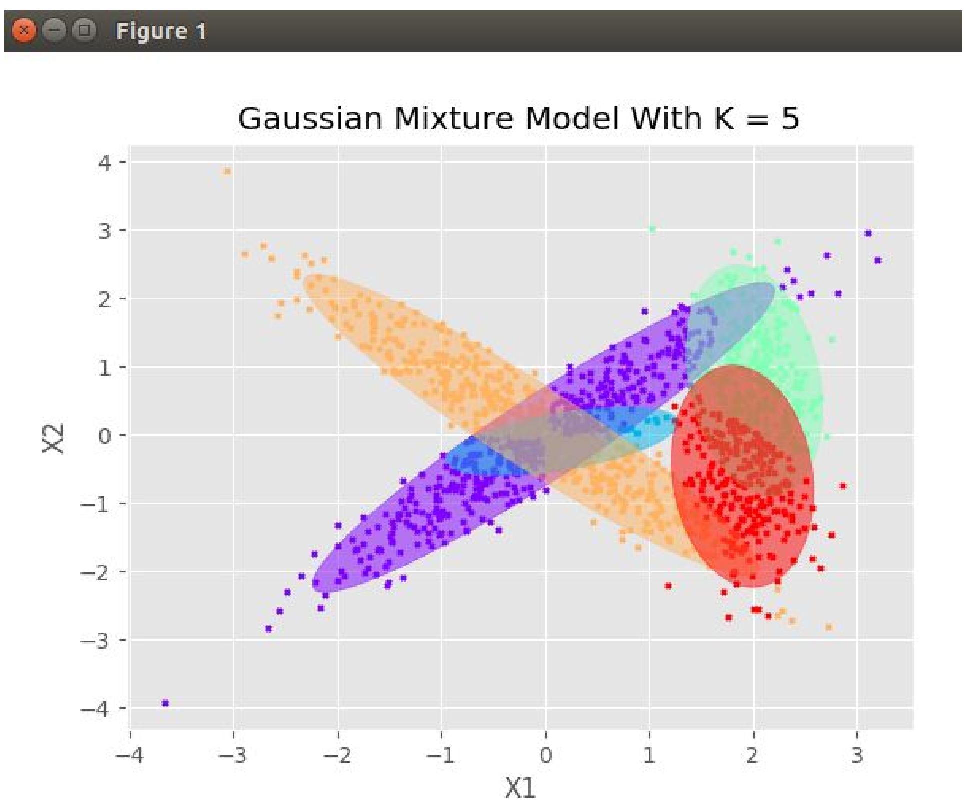


Figure 1
Gaussian Mixture Model With K = 5
X1
X2
4
3
2
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1
2
3