# OBSERVATIONS:

**Generation of data from GMM model with some initial mean and variance**

The mean and variance thus generated from the Expectation maximisation will converge to the initial mean and variance provided **the number of data samples is high and error is very low**

Number of data samples per mode:
10000
Dimension of input(d):
1
Modes of gaussians(k):
1
Enter threshold:
.00000001
Initial mean [array([1.75805447])]
Initial covariance matrix [array([[1.61133617]])]
(10000, 1)
1e-05
error:861.9035060604438 Log_Likelihood:-16510.80434178555
error:0.0 Log_Likelihood:-16510.80434178555
mean
[[1.75331771]]

---

covar
[[[1.5908744]]]

---

w
[[1.]]

---

**Best fit for k depends on the data and cannot be generalized**

## Number of samples per mode

Increasing the number of samples of course increases the log likelihood but as we increase the number of samples, the GMM goes closer to the original Mixture model from which the data has been generated

Number of data samples per mode:

100

Dimension of input(d):

2

Modes of gaussians(k):

2

Enter threshold:

.00001

Actual mean

[array([1.75805447, 1.26938417]), array([1.4238598 , 1.97346674])]

Actual covariance matrix

[array([[5.20427875, 5.31665141],

[5.31665141, 5.62340179]]),

array([[3.63634167, 4.07888403],

[4.07888403, 4.88418427]])]

Estimated mean

[[2.70306638 2.13996621]

[1.51762317 1.59856692]]

---

Estimated covar

[[[0.81242182 0.64181127]

[0.64181127 0.51219033]]

[[4.66966278 5.11842019]

[5.11842019 6.17959049]]]

---

w

[[0.08052274]

[0.91947726]]

---

Not so close

################################################################################

Number of data samples per mode:
1000
Dimension of input(d):
2
Modes of gaussians(k):
2
Enter threshold:
.00001

Actual mean

[array([1.75805447, 1.26938417]),
array([1.38547855, 1.37876314])]

Actual covariance matrix

[array([[5.20427875, 5.31665141],
[5.31665141, 5.62340179]]),
array([[5.20751509, 4.63555808],
[4.63555808, 4.8558754 ]])]

#estimated
mean
[[1.74776215 1.26930338]
 [1.44852313 1.42812232]]

---

covar
[[[5.05368654 5.22516245]
  [5.22516245 5.61050961]]

 [[4.8244277  4.27773469]
  [4.27773469 4.48254526]]]

---

w
[[0.47078692]
 [0.52921308]]

---

(the w1,w2 values are around .5 meaning both are equiprobable because that is how the data is generated thus a close estimate)

**They are Close estimates**
**##########################################################################**

**Another interesting observation is that if the data is taken from a single multivariate gaussian**

**The a single value of k will persist( as the n increases and error goes to 0 ie. 1e-11) and remaining values will go to '0'**

**The mean and covariance estimated approaches the values of mean and covariance from which the data is generated**

Number of data samples per mode:
500
Dimension of input(d):
1
Modes of gaussians(k):
3
Enter threshold:
1e-5

We can clearly see the estimated w

w
[[0.8934973 ]
 [0.0923554 ]
 [0.01414731]]

---

W1 -**> 1**
W2 -> 0
W3 -> 0