1. **Part 1 (EM for general GMMs with unknown variance):**

   In the bottom of the report, we explain why we select number of cluster, K=3.

   Initialization method:

   We initialize mixture's weight with equal weight. For example, if we are interested for 3 clusters, then for each component or cluster weight is 0.3333.

   For mixture mean, we select K (number of mixtures or cluster – we are interested) points from data randomly.

   For mixture variance, we calculate overall data variance and randomly scale from 0 to 9 for each mixture.

   Parameter for different initializations:

   Parameters for various initializations are included in text file (output1.txt). Here we only show, the one for which we receive highest log likelihood value (calculate according to `http://www.ics.uci.edu/~smyth/courses/cs274/notes/EMnotes.pdf`) and sveral others (randomly selected). However, if you run the program, the highest log likelihood may change and you will get different result.

   We set 3 for number of mixture components.

   Initial parameters' values and parameters' values in convergence are reported below:

   ==========Iteration : 1==========
   Parameter values after initialization:
   Mixture Weights:
   Weight for cluster 0 : 0.3333333333333333
   Weight for cluster 1 : 0.3333333333333333
   Weight for cluster 2 : 0.3333333333333333
   Mixture mean:
   Mean for cluster 0 : 24.5202360596
   Mean for cluster 1 : 4.36975217887
   Mean for cluster 2 : 13.7752299177
   Mixture variance:
   variance for cluster 0 : 607.6344447332143
   variance for cluster 1 : 67.51493830369049
   variance for cluster 2 : 67.51493830369049

   Parameter values after convergence:

Mixture Weights:
Weight for cluster 0 : 0.3333333189159592
Weight for cluster 1 : 0.33333334041226304
Weight for cluster 2 : 0.3333333406717546
Mixture mean:
Mean for cluster 0 : 15.481698203117045
Mean for cluster 1 : 15.481698203114524
Mean for cluster 2 : 15.481698203117096
Mixture variance:
variance for cluster 0 : 135029.8707670621
variance for cluster 1 : 135029.8794749699
variance for cluster 2 : 135029.87958010955

Current log likelihood: -40953.38523375209 (randomly selected)

==========Iteration : 5==========
Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.3333333333333333
Weight for cluster 1 : 0.3333333333333333
Weight for cluster 2 : 0.3333333333333333
Mixture mean:
Mean for cluster 0 : 14.0544545597
Mean for cluster 1 : 4.5405667087
Mean for cluster 2 : 7.01817043789
Mixture variance:
variance for cluster 0 : 607.6344447332143
variance for cluster 1 : 67.51493830369049
variance for cluster 2 : 67.51493830369049

Parameter values after convergence:

Mixture Weights:
Weight for cluster 0 : 0.33333331896711826
Weight for cluster 1 : 0.333333340464559
Weight for cluster 2 : 0.3333333405682995
Mixture mean:
Mean for cluster 0 : 15.481698203117126
Mean for cluster 1 : 15.481698203114574

Mean for cluster 2 : 15.481698203116999
Mixture variance:
variance for cluster 0 : 135029.87078778548
variance for cluster 1 : 135029.87949615446
variance for cluster 2 : 135029.87953820033

Current log-likelihood: -40953.38523375208 (got highest log likelihood for this iteration)

==========Iteration : 98==========
Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.3333333333333333
Weight for cluster 1 : 0.3333333333333333
Weight for cluster 2 : 0.3333333333333333
Mixture mean:
Mean for cluster 0 : 13.8956064056
Mean for cluster 1 : 6.12539333785
Mean for cluster 2 : 5.05758152249
Mixture variance:
variance for cluster 0 : 0.0
variance for cluster 1 : 270.05975321476194
variance for cluster 2 : 67.51493830369049

Parameter values after convergence:

Mixture Weights:
Weight for cluster 0 : 0.0
Weight for cluster 1 : 0.49999997977514077
Weight for cluster 2 : 0.5000000202248592
Mixture mean:
Mean for cluster 0 : 0.0
Mean for cluster 1 : 15.48169820311675
Mean for cluster 2 : 15.481698203114966
Mixture variance:
variance for cluster 0 : 0.0
variance for cluster 1 : 202544.80671820225
variance for cluster 2 : 202544.82310393875

Current log likelihood: -40953.38523375208 (randomly selected)

==========Iteration : 100==========
Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.3333333333333333
Weight for cluster 1 : 0.3333333333333333
Weight for cluster 2 : 0.3333333333333333
Mixture mean:
Mean for cluster 0 : 14.087839994
Mean for cluster 1 : 4.85650126778
Mean for cluster 2 : 15.3822433244
Mixture variance:
variance for cluster 0 : 0.0
variance for cluster 1 : 0.0
variance for cluster 2 : 67.51493830369049

Parameter values after convergence:

Mixture Weights:
Weight for cluster 0 : 0.0
Weight for cluster 1 : 0.0
Weight for cluster 2 : 1.0
Mixture mean:
Mean for cluster 0 : 0.0
Mean for cluster 1 : 0.0
Mean for cluster 2 : 15.481698203115881
Mixture variance:
variance for cluster 0 : 0.0
variance for cluster 1 : 0.0
variance for cluster 2 : 405089.6298221419

Current log likelihood: -40953.38523375208 (randomly selected)

Discussion:

The effect of parameter initialization is not much. For example, iteration 1 and iteration 5 converge almost in the same even though their initial means are different.

However, when initial variances are different specially zero (for one or two mixtures, see for iteration 98 and iteration 100), the variances increase for the non-zero variance or mean ones.

One mentionable issue is for all cases the highest log likelihood is almost same. And from our experiments, this seems like the data is well fitted for three clusters.

2. **Part 2 (EM for GMMs with known variance – 1.0):**

Initialization method:

This is same as EM for general GMM with unknown variance.

Parameter for different initializations:

Parameter for various initializations are included in text file (output2.txt) like for EM for general GMM.

We set 3 for number of mixture components.

Initial parameters' values and parameters' values in convergence are reported below:

==========Iteration : 1==========
Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.3333333333333333
Weight for cluster 1 : 0.3333333333333333
Weight for cluster 2 : 0.3333333333333333
Mixture mean:
Mean for cluster 0 : 24.7951879314
Mean for cluster 1 : 24.9580942842
Mean for cluster 2 : 25.1161751291
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0

Parameter values after convergence:

Mixture Weights:
Weight for cluster 0 : 0.33333333351466654
Weight for cluster 1 : 0.3333333331613732
Weight for cluster 2 : 0.33333333332396015
Mixture mean:

Mean for cluster 0 : 5.509279395865944
Mean for cluster 1 : 15.449160789778821
Mean for cluster 2 : 25.48665442939237
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0
Current log likelihood: -15100.775131483002 (got highest log likelihood for this iteration)

==========Iteration : 98==========
Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.3333333333333333
Weight for cluster 1 : 0.3333333333333333
Weight for cluster 2 : 0.3333333333333333
Mixture mean:
Mean for cluster 0 : 24.1334416138
Mean for cluster 1 : 12.6829688621
Mean for cluster 2 : 6.99190265353
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0

Parameter values after convergence:

Mixture Weights:
Weight for cluster 0 : 0.33333333332396015
Weight for cluster 1 : 0.3333333331613732
Weight for cluster 2 : 0.33333333351466654
Mixture mean:
Mean for cluster 0 : 25.48665442939237
Mean for cluster 1 : 15.449160789778821
Mean for cluster 2 : 5.509279395865944
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0
Current log likelihood: -15100.775131483002 (randomly selected)

==========Iteration : 100==========
Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.3333333333333333
Weight for cluster 1 : 0.3333333333333333
Weight for cluster 2 : 0.3333333333333333
Mixture mean:
Mean for cluster 0 : 26.0345168931
Mean for cluster 1 : 16.7539199179
Mean for cluster 2 : 13.3584080839
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0

Parameter values after convergence:

Mixture Weights:
Weight for cluster 0 : 0.33333333332396015
Weight for cluster 1 : 0.3333333331613732
Weight for cluster 2 : 0.33333333351466654
Mixture mean:
Mean for cluster 0 : 25.48665442939237
Mean for cluster 1 : 15.449160789778821
Mean for cluster 2 : 5.509279395865944
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0
Current log likelihood: -15100.775131483002 (randomly selected)

Discussion: Above we provide several iterations with parameters' values after initialization and convergence along with log likelihood. Here, only mixture means are different for various initializations remaining mixture weights and variance are same. And they converge almost on the same parameter values.

3. **Comparison with approach 1 and approach 2:**

   The log-likelihood is higher in case of EM for GMM with known variance. -15100.775131483002 for known variance and -40953.38523375208 for un-

known variance. This indicates approach 2 (with known variance) is better than approach 1. Moreover, known variance is best fitted for this given data. The mean are about 5, 15, and 25 for known variance while mean remains always about 15 with unknown variance.

4. **Cluster number selection**

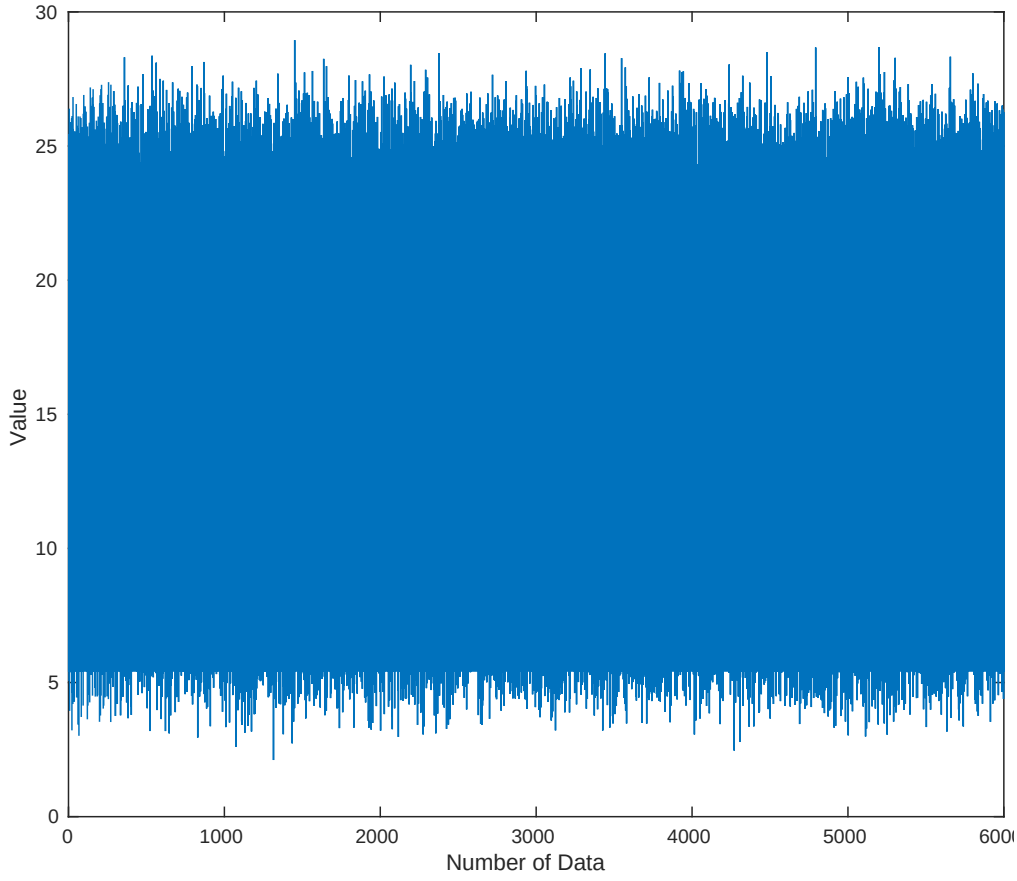

Figure 1: Given data

Frorm the plot (Figure 1), the value of the given 1-D array varies from 5 to 25. And the mean is about 15. To select proper cluster number, we did some experiments with both unknown and known variance. For unknown, it is tough to determine proper cluster numbers. For example, if we try with K=2, we get:

==========Iteration : 5==========

Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.5
Weight for cluster 1 : 0.5
Mixture mean:
Mean for cluster 0 : 15.4712290824
Mean for cluster 1 : 5.78718214968
Mixture variance:
variance for cluster 0 : 5468.71000259893
variance for cluster 1 : 1687.873457592262

Parameter values after convergence:
Mixture Weights:
Weight for cluster 0 : 0.4999999997262415
Weight for cluster 1 : 0.5000000002737585
Mixture mean:
Mean for cluster 0 : 15.48169820311548
Mean for cluster 1 : 15.481698203116238
Mixture variance:
variance for cluster 0 : 202544.81480016842
variance for cluster 1 : 202544.8150219723
Current log-likelihood: -42169.780551904994

After convergence, the weight for both clusters are same. And their mean and variance are also same.

For K=3, we get:

Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.3333333333333333
Weight for cluster 1 : 0.3333333333333333
Weight for cluster 2 : 0.3333333333333333
Mixture mean:
Mean for cluster 0 : 25.9906123068
Mean for cluster 1 : 15.9007610099
Mean for cluster 2 : 24.1955194766
Mixture variance:
variance for cluster 0 : 607.6344447332143
variance for cluster 1 : 607.6344447332143

variance for cluster 2 : 607.6344447332143
Parameter values after convergence:
Mixture Weights:
Weight for cluster 0 : 0.3333333332495957
Weight for cluster 1 : 0.3333333334410443
Weight for cluster 2 : 0.33333333330933673
Mixture mean:
Mean for cluster 0 : 15.481698203117107
Mean for cluster 1 : 15.481698203117052
Mean for cluster 2 : 15.481698203114536
Mixture variance:
variance for cluster 0 : 135029.8765734697
variance for cluster 1 : 135029.87665102314
variance for cluster 2 : 135029.87659764764
Current log-likelihood: -40953.38523375208

Though for K=3, the behavior of cluster weight, mean, and variance are same, the log-likelihood increases (-42169.780551904994 to -40953.38523375208). Notice able mean is same for two cases around 15.48.

For K=4, we get:

==========Iteration : 7==========
Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.25
Weight for cluster 1 : 0.25
Weight for cluster 2 : 0.25
Weight for cluster 3 : 0.25
Mixture mean:
Mean for cluster 0 : 13.6210160386
Mean for cluster 1 : 5.87143547534
Mean for cluster 2 : 25.8502412182
Mean for cluster 3 : 12.0993870535
Mixture variance:
variance for cluster 0 : 270.05975321476194
variance for cluster 1 : 270.05975321476194
variance for cluster 2 : 3308.2319768808334
variance for cluster 3 : 607.6344447332143
Parameter values after convergence:
Mixture Weights:

Weight for cluster 0 : 0.25000000308908393
Weight for cluster 1 : 0.25000000308455006
Weight for cluster 2 : 0.24999999375698614
Weight for cluster 3 : 0.25000000006941175
Mixture mean:
Mean for cluster 0 : 15.481698203114716
Mean for cluster 1 : 15.481698203115133
Mean for cluster 2 : 15.481698203114627
Mean for cluster 3 : 15.481698203116979
Mixture variance:
variance for cluster 0 : 101272.40870688374
variance for cluster 1 : 101272.40870504986
variance for cluster 2 : 101272.40492654704
variance for cluster 3 : 101272.40748366015
Current log-likelihood: -40090.33902503616

Thus as we increase K higher values like 5 and 6 (we did), the cluster weight is equal for all cluster. Mean remain same as earlier 15.48. However, overall, variance and log-likelihood decrease as we increase K.

For all results with unknown variance, please see file p2.txt, p3.txt, p4.txt, p5.txt, and p6.txt.

On the other hand, for known variance (for our case 1), variance remains same (1). However, mean can be divided in three types : about 5, 10, and 25. And only for K=3, we get these three means. For K>3, some means' values are repeated. Another noticeable matter,for K=2, log-likelihood becomes about -61703.70111161951 while for K>2, log-likelihood becomes fixed about -15100. And cluster weight becomes equal for only K=3. For other cases like K$\geq$ 4 for some clusters weight becomes negligible. Please see the following results.

This leads us to select K=3.

For K=2, we get:

==========Iteration : 42==========
Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.5
Weight for cluster 1 : 0.5
Mixture mean:
Mean for cluster 0 : 25.8608524395

Mean for cluster 1 : 5.22723134307
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
Parameter values after convergence:
Mixture Weights:
Weight for cluster 0 : 0.3345285098863927
Weight for cluster 1 : 0.6654714901136077
Mixture mean:
Mean for cluster 0 : 25.46063213794922
Mean for cluster 1 : 10.465348218077915
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
Current log-likelihood: -61703.70111161951

For K=3, we get:

==========Iteration : 1==========
Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.3333333333333333
Weight for cluster 1 : 0.3333333333333333
Weight for cluster 2 : 0.3333333333333333
Mixture mean:
Mean for cluster 0 : 25.489681819
Mean for cluster 1 : 25.4427005847
Mean for cluster 2 : 5.01078791723
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0
Parameter values after convergence:
Mixture Weights:
Weight for cluster 0 : 0.33333333332396015
Weight for cluster 1 : 0.3333333331613732
Weight for cluster 2 : 0.33333333351466654
Mixture mean:
Mean for cluster 0 : 25.48665442939237
Mean for cluster 1 : 15.449160789778821

Mean for cluster 2 : 5.509279395865944
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0
Current log-likelihood: -15100.775131483002

For K=4, we get:

==========Iteration : 32==========
Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.25
Weight for cluster 1 : 0.25
Weight for cluster 2 : 0.25
Weight for cluster 3 : 0.25
Mixture mean:
Mean for cluster 0 : 5.92000946619
Mean for cluster 1 : 16.4974121942
Mean for cluster 2 : 15.1200295577
Mean for cluster 3 : 6.14442244716
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0
variance for cluster 3 : 1.0
Parameter values after convergence:
Mixture Weights:
Weight for cluster 0 : 0.03583736694455412
Weight for cluster 1 : 0.33333333332396015
Weight for cluster 2 : 0.3333333330862974
Weight for cluster 3 : 0.2974959666451881
Mixture mean:
Mean for cluster 0 : 4.9261213682366085
Mean for cluster 1 : 25.48665442939237
Mean for cluster 2 : 15.449160790570458
Mean for cluster 3 : 5.5795285792107014
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0

variance for cluster 3 : 1.0
Current log-likelihood: -15099.991617722846

For K=5, we get:

==========Iteration : 14==========
Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.2
Weight for cluster 1 : 0.2
Weight for cluster 2 : 0.2
Weight for cluster 3 : 0.2
Weight for cluster 4 : 0.2
Mixture mean:
Mean for cluster 0 : 4.02885767126
Mean for cluster 1 : 6.86861226281
Mean for cluster 2 : 4.98745148385
Mean for cluster 3 : 16.063839289
Mean for cluster 4 : 17.7737100823
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0
variance for cluster 3 : 1.0
variance for cluster 4 : 1.0
Parameter values after convergence:
Mixture Weights:
Weight for cluster 0 : 0.035837852526796964
Weight for cluster 1 : 0.12836532854979993
Weight for cluster 2 : 0.16913015251314636
Weight for cluster 3 : 0.3333333330862968
Weight for cluster 4 : 0.33333333332396015
Mixture mean:
Mean for cluster 0 : 4.9261252043415045
Mean for cluster 1 : 5.579529183606829
Mean for cluster 2 : 5.579529183607208
Mean for cluster 3 : 15.449160790570463
Mean for cluster 4 : 25.48665442939237
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0

variance for cluster 2 : 1.0
variance for cluster 3 : 1.0
variance for cluster 4 : 1.0
Current log-likelihood: -15099.991617723968

For K=6, we get:

==========Iteration : 6==========
Parameter values after initialization:
Mixture Weights:
Weight for cluster 0 : 0.16666666666666666
Weight for cluster 1 : 0.16666666666666666
Weight for cluster 2 : 0.16666666666666666
Weight for cluster 3 : 0.16666666666666666
Weight for cluster 4 : 0.16666666666666666
Weight for cluster 5 : 0.16666666666666666
Mixture mean:
Mean for cluster 0 : 14.9305413166
Mean for cluster 1 : 4.8447642925
Mean for cluster 2 : 26.2825457652
Mean for cluster 3 : 14.9998522175
Mean for cluster 4 : 6.65203481473
Mean for cluster 5 : 25.195073364
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0
variance for cluster 3 : 1.0
variance for cluster 4 : 1.0
variance for cluster 5 : 1.0
Parameter values after convergence:
Mixture Weights:
Weight for cluster 0 : 0.1638714225394738
Weight for cluster 1 : 0.03584022338593533
Weight for cluster 2 : 0.14642663462880454
Weight for cluster 3 : 0.16946191054681992
Weight for cluster 4 : 0.2974931102038109
Weight for cluster 5 : 0.18690669869515558
Mixture mean:
Mean for cluster 0 : 15.44916079057038
Mean for cluster 1 : 4.926143933409409

Mean for cluster 2 : 25.486654429417378
Mean for cluster 3 : 15.449160790570584
Mean for cluster 4 : 5.579532134515156
Mean for cluster 5 : 25.486654429372884
Mixture variance:
variance for cluster 0 : 1.0
variance for cluster 1 : 1.0
variance for cluster 2 : 1.0
variance for cluster 3 : 1.0
variance for cluster 4 : 1.0
variance for cluster 5 : 1.0
Current log-likelihood: -15099.991617730111

All results for know variance (1) can be found in p21.txt (for K=2), p31.txt (for K=3), p41.txt (for K=4), p51.txt (for K=5), and p61.txt (for K=6).