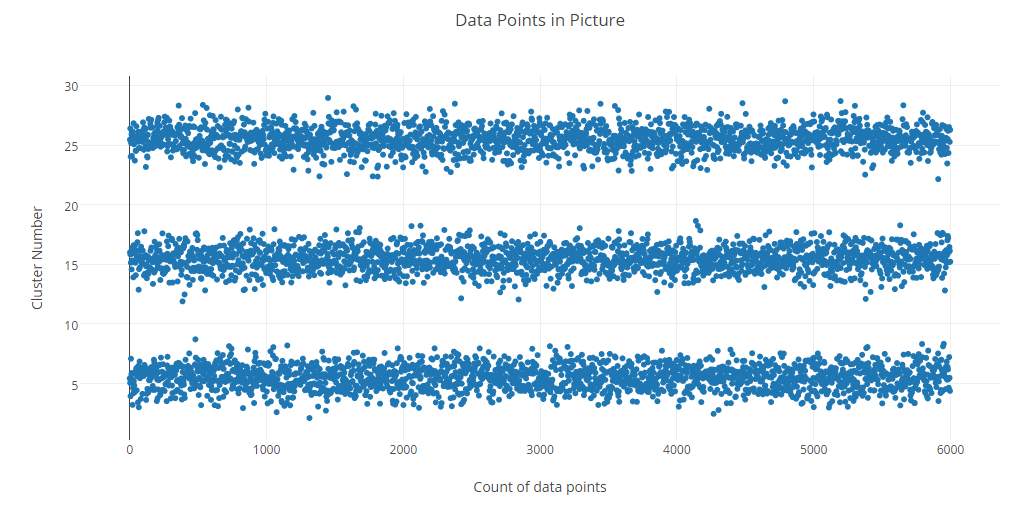
**Report on EM Algorithm**

**Analysis of EM algorithm:**

Following is the scattered graph of the given dataset. From the graph it is clear that the data points forms the three cluster’s 5, 15 and 25

* Visualization of clustering on given input data clearly provides the picture that the data is divisible into three different clusters.
* Since the input data passed is in such a way that it is majorly concentrated around three values (5, 15, and 25).
* I have tested for the three different values of cluster (K=2, K=3, K=4) for both Gaussian mixture model and Gaussian mixture model with known variance.



**Answers:**

1. What initialization strategy did you use? How sensitive was the performance to the initial settings of parameters?

* Random initialization strategy using java API ***Random*** function . That will initially select the ***“k”*** different means.
* Variance is calculated by calculating actual mean of the data and initializes the variance into multiple of the actual variance.
* If the initial setting of the mean is not proper or not nearer to actual mean value of the data points the algorithm will take time to converge that is number of iterations are more.

1. Which approach worked better, this one or the previous one? Why?

* First approach is better than with known variance as it always yields to the correct classification regardless of initialization values. Although approach with known variance, algorithm is converging fast with less iteration comparatively with unknown variance.

**Output:**

**K=2:**

**Gaussian Mixture model**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Gaussian Mixture Model** | | | | | | |
| Intial | | | Final | | | No of Iterations |
| Mean | Variance | Prior | Mean | Variance | Prior |
| 19.006141764366124 6.4739574111370395 | 101.27240745553573 101.27240745553573 | 0.5 0.5 | 20.320789630323002 5.486942558558723 | 27.943391046352538 0.9861220193851277 | 0.6737804155728314 0.32621958442716825 | 38 |
| 25.60956874412135 25.142986027609002 | 101.27240745553573 101.27240745553573 | 0.5 0.5 | 25.5167446455702 10.642822531041055 | 0.9256489488567436 27.651076347656076 | 0.32532614026182655 0.6746738597381732 | 2836 |
| 20.900960533415958 25.23039685222689 | 101.27240745553573 101.27240745553573 | 0.5 0.5 | 10.642822531041055 25.5167446455702 | 27.651076347656076 0.9256489488567436 | 0.6746738597381732 0.32532614026182655 | 76 |
| 25.32481212946004 19.82462414964363 | 101.27240745553573 101.27240745553573 | 0.5 0.5 | 25.5167446455702 10.642822531041055 25.5167446455702 10.642822531041055 | 0.9256489488567436 27.651076347656076 | 0.32532614026182655 0.6746738597381732 | 62 |
| 2.5294829001861787 22.791656466430908 | 101.27240745553573 101.27240745553573 | 0.5 0.5 | 5.486942558558723 20.320789630322984 | 0.9861220193851242 27.943391046352712 | 0.3262195844271675 0.6737804155728322 | 36 |

**Gaussian Mixture Model with Known Variance**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | **Gaussian Mixture Model with Known Variance** | | | | | | |
| Intial | | | Final | | | No of Iterations |
| Mean | Variance | Prior | Mean | Variance | Prior |
| 1 | 22.37105473042751 5.532111511202929 | 1.0 1.0 | 0.5 0.5 | 20.320789630322984 5.486942558558723 | 27.9433910463527 0.9861220193851242 | 0.6737804155728322 0.3262195844271675 | 29 |
| 2 | 25.822763510301186 18.94588331252536 | 1.0 1.0 | 0.5 0.5 | 25.5167446455702 10.642822531041055 | 0.9256489488567438 27.65107634765607 | 0.3253261402618266 0.6746738597381731 | 33 |
| 3 | 26.785013798387922 10.693778170635323 | 1.0 1.0 | 0.5 0.5 | 25.51674464557019 10.642822531041 25.51674464557019 10.642822531041 | 0.9256489488567616 27.65107634765542 | 0.3253261402618289 0.674673859738171 | 28 |
| 4 | 8.820192980274854 22.612370120648944 | 1.0 1.0 | 0.5 0.5 | 10.642822531041055 25.5167446455702 | 27.651076347656076 0.9256489488567436 | 0.6746738597381732 0.32532614026182655 | 42 |
| 5 | 17.982932892792356 5.667638440089535 | 1.0 1.0 | 0.5 0.5 | 20.320789630323 5.486942558558721 | 27.94339104635255 0.9861220193851272 | 0.6737804155728316 0.32621958442716825 | 26 |

**K=3:**

**Gaussian Mixture model**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | **Gaussian Mixture Model** | | | | | | |
| Intial | | | Final | | | No of Iterations |
| Mean | Variance | Prior | Mean | Variance | Prior |
| 1 | 19.853260019000206 10.881592009827711 15.30285366899137 | 101.27240745553573 101.27240745553573 101.27240745553573 | 0.3333 0.3333 0.3333 | 25.486654429329228 5.50927940017642 15.449160792031073 | 0.9980966181791335 1.0302576846296716 0.9671159353647738 | 0.33333333333014475 0.3333333337388757 0.3333333329309795 | 24 |
| 2 | 6.171792432579419 7.234890347800988 22.65352037834792 | 101.27240745553573 101.27240745553573 101.27240745553573 | 0.3333 0.3333 0.3333 | 5.50927940017642 15.449160792031073 25.486654429329228 | 1.0302576846296716 0.9671159353647738 0.9980966181791335 | 0.3333333337388757 0.3333333329309795 0.33333333333014475 | 61 |
| 3 | 2.482908892925725 4.166006657252687 5.489252262186944 | 101.27240745553573 101.27240745553573 101.27240745553573 | 0.3333 0.3333 0.3333 | 5.50927940017642 15.449160792031073 25.486654429329228 | 1.0302576846296716 0.9671159353647738 0.9980966181791335 | 0.3333333337388757 0.3333333329309795 0.33333333333014475 | 110 |
| 4 | 9.512588857723022 11.61654853564524 22.09807703595566 | 101.27240745553573 101.27240745553573 101.27240745553573 | 0.3333 0.3333 0.3333 | 5.50927940017642 15.449160792031073 25.486654429329228 | 1.0302576846296716 0.9671159353647738 0.9980966181791335 | 0.3333333337388757 0.3333333329309795 0.33333333333014475 | 37 |
| 5 | 4.597259952755077 8.35789737873739 17.297297986560466 | 101.27240745553573 101.27240745553573 101.27240745553573 | 0.3333 0.3333 0.3333 | 5.50927940017642 15.449160792031073 25.486654429329228 | 1.0302576846296716 0.9671159353647738 0.9980966181791335 | 0.3333333337388757 0.3333333329309795 0.33333333333014475 | 20 |

**Gaussian Mixture Model with Known Variance**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | **Gaussian Mixture Model with Known Variance** | | | | | | |
| Intial | | | Final | | | No of Iterations |
| Mean | Variance | Prior | Mean | Variance | Prior |
| 1 | 11.460680980051285 29.063486462301803 6.494102335094967 | 1.0 1.0 1.0 | 0.3333333333333333 0.3333333333333333 0.3333333333333333 | 15.449160792031073 25.486654429329228 5.50927940017642 | 0.9671159353647738 0.9980966181791335 1.0302576846296718 | 0.3333333329309795 0.33333333333014475 0.3333333337388757 | 5 |
| 2 | 21.11297388823476 5.51543033178098 14.875255133852846 | 1.0 1.0 1.0 | 0.3333333333333333 0.3333333333333333 0.3333333333333333 | 25.486654429329228 5.50927940017642 15.449160792031076 | 0.9980966181791346 1.030257684629684 0.9671159353647694 | 0.33333333333014475 0.33333333373887586 0.33333333293097933 | 3 |
| 3 | 1.5768673419851567 13.519912094310556 12.345842539375546 | 1.0 1.0 1.0 | 0.3333333333333333 0.3333333333333333 0.3333333333333333 | 5.50927940017642 25.486654429329228 15.449160792031073 | 1.0302576846296716 0.9980966181791335 0.9671159353647738 | 0.3333333337388757 0.33333333333014475 0.3333333329309795 | 14 |
| 4 | 22.80928300582625 23.13569846154568 23.27265593218723 | 1.0 1.0 1.0 | 0.3333333333333333 0.3333333333333333 0.3333333333333333 | 5.50927940017642 15.449160792031073 25.486654429329228 | 1.0302576846296716 0.9671159353647738 0.9980966181791335 | 0.3333333337388757 0.3333333329309795 0.33333333333014475 | 27 |
| 5 | 18.275557179751463 11.649914480503679 26.294940834079597 | 1.0 1.0 1.0 | 0.3333333333333333 0.3333333333333333 0.3333333333333333 | 15.449160792031073 5.50927940017642 25.486654429329228 | 0.9671159353647738 1.0302576846296716 0.9980966181791335 | 0.3333333329309795 0.3333333337388757 0.33333333333014475 | 8 |

**K=4:**

**Gaussian Mixture model**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | **Gaussian Mixture Model** | | | | | | |
| Intial | | | Final | | | No of Iterations |
| Mean | Variance | Prior | Mean | Variance | Prior |
| 1 | 9.785783130084253 19.532441714150426 24.806003578104647 25.838706996771045 | 101.27240745553573 101.27240745553573 101.27240745553573 101.27240745553573 | 0.25 0.25 0.25 0.25 | 5.509279394164894 15.45470122965074 12.048556662969482 25.486654429309432 | 1.030257646817398 0.94980592989638 0.008684053873664013 0.9980966182497611 | 0.3333333334282901 0.3327911326376828 5.422006020335687E-4 0.3333333333319935 | 95 |
| 2 | 4.638633650468238 11.404443906674587 8.369103392111464 14.682676100993486 | 101.27240745553573 101.27240745553573 101.27240745553573 101.27240745553573 | 0.25 0.25 0.25 0.25 | 5.509279394164894 12.048556662969482 15.45470122965074 25.486654429309432 | 1.030257646817398 0.008684053873664046 0.9498059298963801 0.9980966182497611 | 0.3333333334282901 5.422006020335691E-4 0.3327911326376828 0.3333333333319935 | 151 |
| 3 | 15.454701229650743 12.048556662969284 5.509279394164894 25.486654429309432 | 0.9498059298963677 0.008684053873680797 1.030257646817398 0.9980966182497611 | 0.33279113263768245 5.422006020338805E-4 0.3333333334282901 0.3333333333319935 | 15.454701229650743 12.048556662969284 5.509279394164894 25.486654429309432 | 0.9498059298963677 0.008684053873680797 1.030257646817398 0.9980966182497611 | 0.33279113263768245 5.422006020338805E-4 0.3333333334282901 0.3333333333319935 | 86 |
| 4 | 4.1675638858210124 12.247963756924761 10.71844474224073 3.0330995692969798 | 101.27240745553573 101.27240745553573 101.27240745553573 101.27240745553573 | 0.25 0.25 0.25 0.25 | 15.454701229650738 25.486654429309432 12.048556662969515 5.509279394164894 | 0.9498059298963819 0.9980966182497611 0.008684053873661427 1.030257646817398 | 0.3327911326376828 0.3333333333319935 5.422006020335212E-4 0.3333333334282901 | 99 |
| 5 | 2.712073011898524 2.2048328099272334 7.125226810744075 24.13687985654043 | 1.0819050400095365 0.8950760939955964 0.9980966183552422 1.0302576602762077 | 0.03565460528623365 0.2976787279254868 0.3333333333344932 0.33333333345378685 | 14.583966290534486 13.720032843331598 15.98193575115269 11.886734263638537 | 14.583966290534486 13.720032843331598 15.98193575115269 11.886734263638537 | 0.25 0.25 0.25 0.25 | 162493 |

**Gaussian Mixture Model with Known Variance**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | **Gaussian Mixture Model with Known Variance** | | | | | | |
| Intial | | | Final | | | No of Iterations |
| Mean | Variance | Prior | Mean | Variance | Prior |
| 1 | 4.719503595275976 2.3460454595480273 21.124312171964153 12.571902393228685 | 1.0 1.0 1.0 1.0 | 0.25 0.25 0.25 0.25 | 5.900707533160835 4.5456604331822374 25.486654429329228 15.44916078772276 2.3460454595480273 21.124312171964153 12.571902393228685 | 0.6953876182083206 0.5488925609369294 0.9980966181791335 0.9671159493979212 1.0 1.0 1.0 | 0.23704439477169154 0.09628893855844214 0.33333333333014475 0.33333333333972187 | 11775 |
| 2 | 5.4755293273129 9.161983188232421 3.2599952431354615 19.84782192707613 | 1.0 1.0 1.0 1.0 | 0.25 0.25 0.25 0.25 | 5.900707533161954 15.44916078772276 4.545660433184068 25.486654429329228 | 0.6953876182076967 0.9671159493979212 0.5488925609376951 0.9980966181791335 | 0.23704439477136724 0.33333333333972187 0.0962889385587664 0.33333333333014475 | 11616 |
| 3 | 27.625091528570454 11.85833431104021 12.522385693398448 20.8062468650538 | 1.0 1.0 1.0 1.0 | 0.25 0.25 0.25 0.25 | 25.486654429309432 5.509279394164894 15.454701229650741 12.048556662969283 | 0.9980966182497611 1.030257646817398 0.9498059298963673 0.008684053873681086 | 0.3333333333319935 0.3333333334282901 0.33279113263768245 5.422006020338855E-4 | 71 |
| 4 | 7.984784586766416 3.4949615142475956 14.181278337773051 14.365992253357376 | 1.0 1.0 1.0 1.0 | 0.25 0.25 0.25 0.25 | 5.900707533173684 4.545660433203238 15.44916078772276 25.486654429329228 | 0.6953876182011864 0.5488925609458446 0.9671159493979212 0.9980966181791335 | 0.23704439476795688 0.09628893856217612 0.33333333333972187 0.33333333333014475 | 12820 |
| 5 | 6.370951117254503 29.305319687245152 27.002633707196793 7.626824169663374 | 1.0 1.0 1.0 1.0 | 0.25 0.25 0.25 0.25 | 5.50927940017642 28.410993952705116 25.477933811867903 15.44916079204869 | 1.0302576846296716 0.02645335759114029 0.97541604146459 0.9671159354521583 | 0.3333333337388757 9.910715616895309E-4 0.33234226176714926 0.3333333329322855 | 388 |