Team number - 3

|  |  |
| --- | --- |
| **Name** | **Unity ID** |
| Rasika Pande | rhpande |
| Satish Gurav | sjgurav |
| Deepak Patil | dpatil |
| Uddhav Bhosle | ubhosle |
| Kumar Nielarshi | knielar |
| Atefeh Morsali | matefeh |

We are supposed to implement the following Gaussian-means (G-means) algorithm, taken from reference paper, in C language.

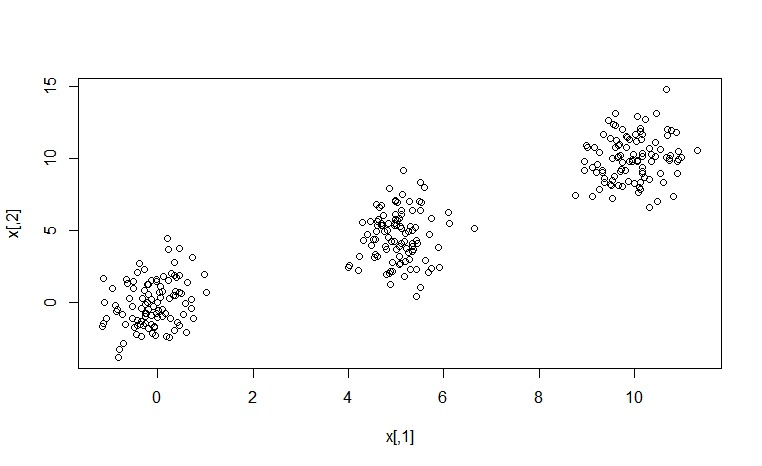
**Algorithm** G-means(X,

1. Let C be the initial set of centers (usually).
2. C kmeans(C, X).
3. Let { be the set of datapoints assigned to center .
4. Use a statistical test to detect if each {class(=j} follow a Gaussian distribution (at confidence level
5. If the data look Gaussian, keep . Otherwise replace with two centers
6. Repeat from step 2 until no more centers are added

We have implemented the above in C. Here is the synthetic dataset that we used 🡪



The above dataset is a two-dimensional dataset that can be visualized as below:



Our Experimental Results 🡪

1. First, we ran our algorithm for synthetic data that had two distinct clusters rather than three.
2. Our algorithm correctly predicted it as two distinct clusters.
3. Then, we ran algorithm on above provided data.
4. But, it predicted 4 distinct clusters instead of 3.   
   Following are the 4 centers that we got,

|  |  |  |
| --- | --- | --- |
| Cluster 0: | -0.102 | 0.913 |
| Cluster 1: | 5.020 | 4.180 |
| Cluster 2: | 5.303 | 7.414 |
| Cluster 3: | 9.961 | 9.990 |