**Problem statement:** Analyzing changes in evolving data fromadvanced habitation systems.

The goal of this task is to create a framework that monitors and provides change detection in a multimodal system. **Our objective** is two-fold (a) to provide the mining of patterns from older data as changes in data could reflect long term/ previous trends. (b) Mining for patterns over near data – as recent changes in data could indicate the recurrences of the previously known pattern or an upcoming event.

**Reported outcomes:** Implementation of a data streaming platform and using known techniques to capture concept drift (clustering) in evolving data.

**Specific Aims:**

1. Using the R statistical packages stream for the definition of micro and macro clusters over simulated streaming data.
2. Simulated data streams with static structure, i.e. the statistical properties of the data stream does not change.
3. Simulated data streams with concept drift, i.e. the statistical properties of the target variable, the model is trying to predict, changes with time in unforeseen ways.

**Key Accomplishments:**

1. Investigated the known functions in R to build data stream framework:

* Data Stream Data (DSD), simulates or connects to a data stream
* Data Stream Task (DST), performs a stream mining task.

**Red Flags:**

1. We are not handling real world multimodal data.
2. We are yet to integrate change detection with the data stream framework.

**Future Work:** To integrate the bearing dataset to the MOA-based techniques.

**Timeline (tentative timeline for the upcoming week)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Future Specific aims** | **09/25** | **09/26** | **09/27** | **09/30** | **10/01** |
| Task 2.1: Investigate MOA-based techniques as extension to R techniques |  |  |  |  |  |
| Task 1: Investigate NASA bearings datasets |  |  |  |  |  |
| Test 2.2: Integrating NASA dataset to the MOA-based techniques |  |  |  |  |  |

**References:**

[1] Hahsler, M., Bolanos, M., & Forrest, J. (2017). Introduction to stream: An Extensible Framework for Data Stream Clustering Research with R. *Journal of Statistical Software*, *76*(14), 1-50.

**Appendix A**

**Results**

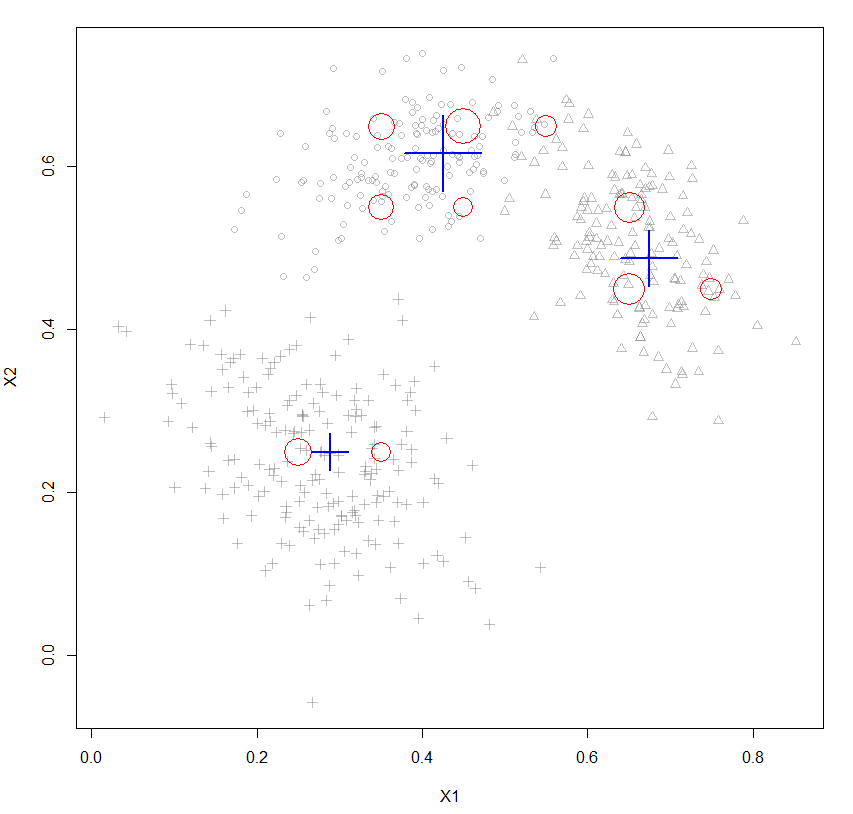


Figure 1: Data stream clustering result of simple simulated data set with three random Gaussians.

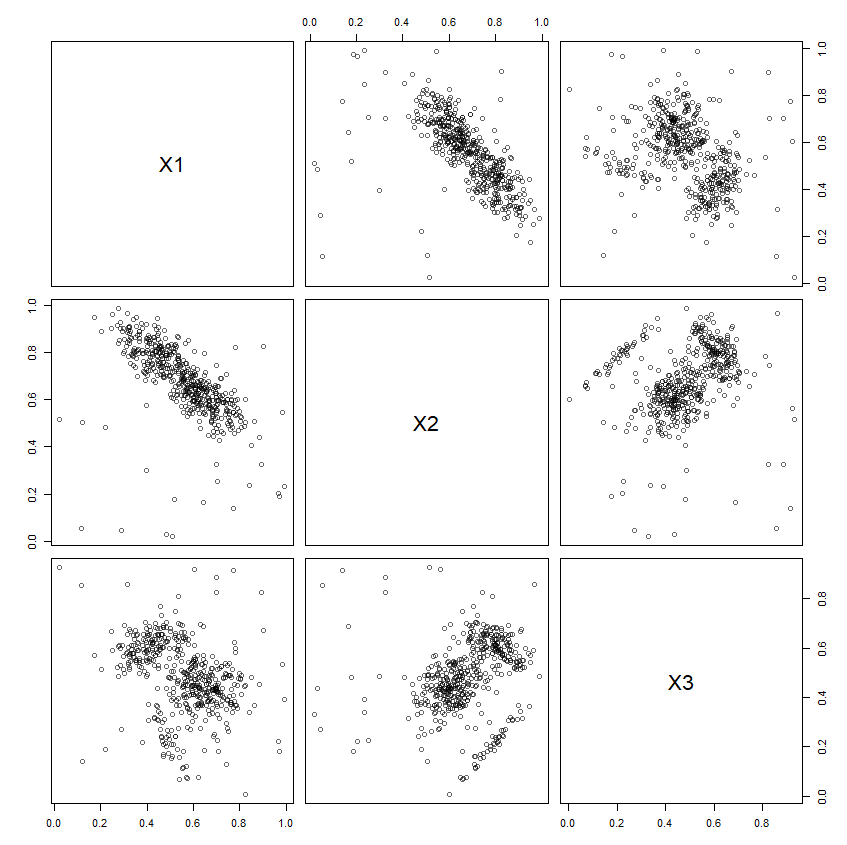


Figure 2: Plotting 500 data points from the data stream.

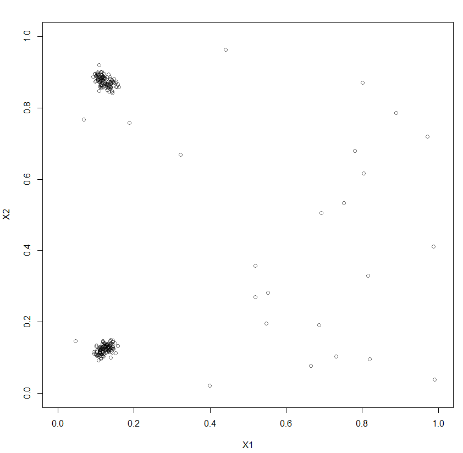
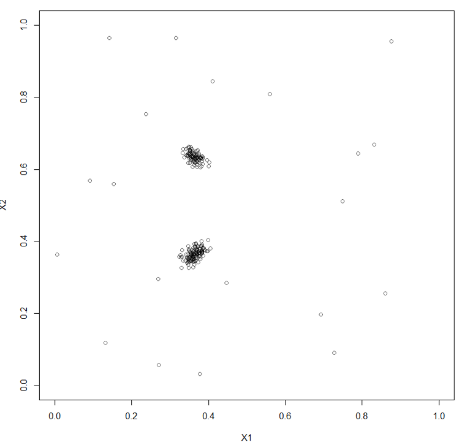
 

Figure 3(a): Position 1 Figure 3(b): Position 2

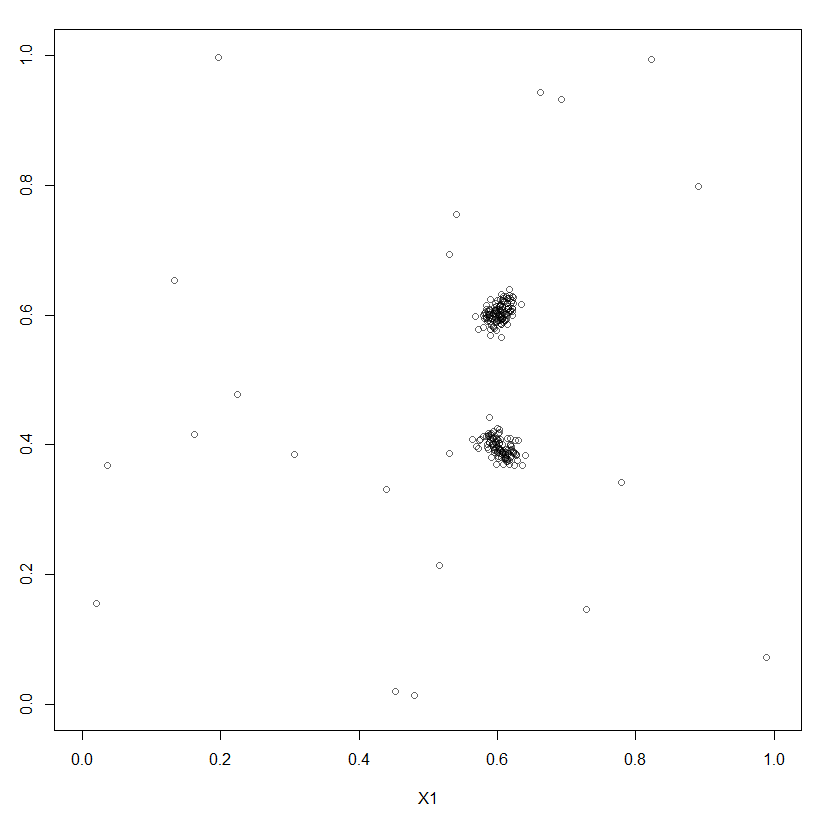
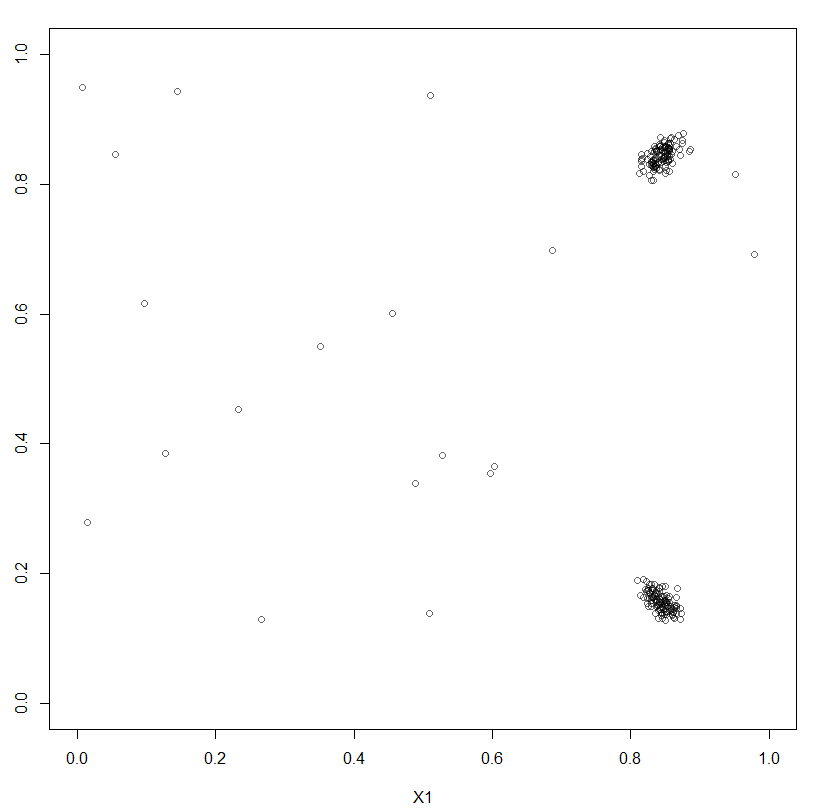
 

Figure 3(c): Position 3 Figure 3(d): Position 4