# EH2745 – Assignment II

# Computer Applications in Power Systems

## Group 10:

**Daniel Cillero Rancaño –** daniecr@kth.se

**Alejandro Vas Corrales –** alvc@kth.se

# EH2745 Assignment II – Identification and labelling of clusters

## 1. Introduction

The objective of the assignment is to analyse a database from a 9-bus power system identifying the different operational states from one learning set and once the different states are identified, assign the corresponding operational state for another set of data – or test set.

The different operational states to be identified are:

- High load rate during peak hours
- Shut down of generator for maintenance
- Low load rate during night
- Disconnection of a line for maintenance

The data is obtained from two different .csv files with all the values of the voltage and phase for every bus and time-step. In the learning set (measurements file) there are 200 time-steps, with voltage and phase data for every bus, therefore, 3600 values (200x9x2=3600). To identify the different operational states for every time-step the data should be modified to reduce the number of dimensions, instead of considering the 9 buses, we calculate the average voltage and phase value for every time-step and, moreover, we normalize the data to cluster it within a better range (removing negative values for example). After the clustering, we will be able to analyse the different operational states and label them accordingly.

Finally, using the test set (analog_values file) which has 20 time-steps we will use the KNN classification algorithm to assign to the test value the most common cluster around the neighbours previously classified using k-Means.

### 1.1. K-Means clustering and labelling

In this section we will specify how we have initialized the centroids in the k-Means algorithm and how we have labelled every cluster after analysing the results.

Once the learning set is modified – average calculation and normalization, we can cluster the data using the k-Means algorithm. In order to initialize the centroids we have used a variant of the k-Means++, firstly, we set the first centroid in the middle of the whole dataset and then, we calculate the distance from this centroid to every point, the second centroid will be set at the furthest point from the first centroid. The third centroid will be set as the point that is further from both the first and the second and finally, the fourth one will be set as the furthest point from the other 3.

It is not a random initialization, but it fits our needs in order to start the k-Means algorithm. After this first initialization the algorithm will recalculate the new centroids until the tolerance is accomplished, the obtained centroids (denormalized but still representing the average value of the power system) are:

Table 1. Clustering final centroids

| | |
|---|---|
| Centroid cluster 1 | 0.987, -5.909 |
| Centroid cluster 2 | 0.915, -22.687 |
| Centroid cluster 3 | 0.991, 12.907 |
| Centroid cluster 4 | 0.966, 7.688 |

In figures 2,3,4 and 5 from the Annex I, obtained directly from the Java application, a detail of the centroids position can be seen.

To label the different clusters the data should be analysed:

- **Cluster 1 – Shut down of generator for maintenance**.

  During a generator shut down, the power flow through the line where it is connected will be approximately 0. Knowing that the power flow could be approximated as:

$$P_{ij} = \frac{V_i \cdot V_j}{X_{ij}} \sin(\theta_i - \theta_j)$$

  It can be seen that the power flow through buses 3 and 6 in figure 1 will represent a generator shut down.

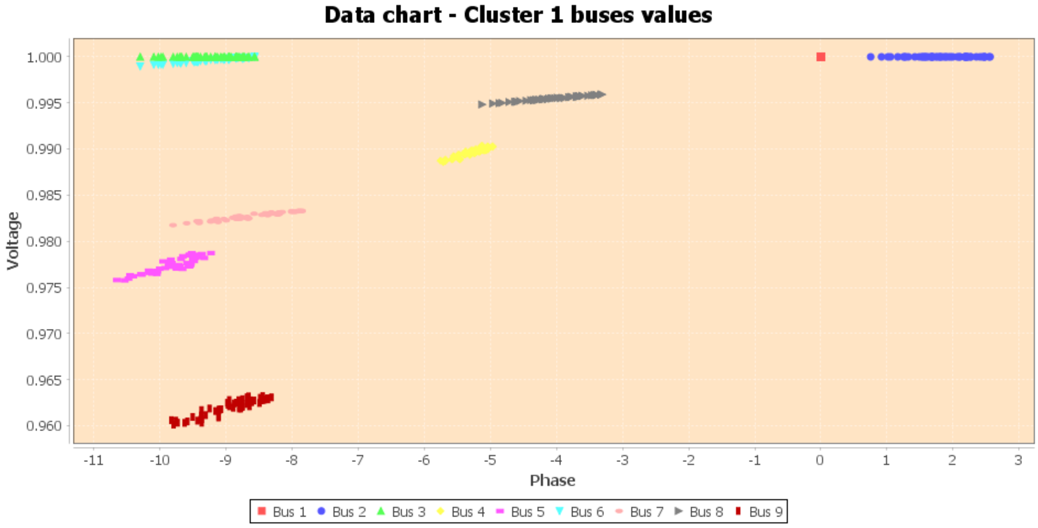


Figure 1. Scatter plot of the cluster 1 obtained from the learning set.

- **Cluster 2 – Disconnection of a line for maintenance**.

  During the disconnection of a line the flow of the power is shared with the remaining ones, increasing the current through them and so that the losses, the result will be a voltage drop. Figure 2 shows the buses with the lowest p.u. voltage level.

- **Cluster 3 – Low load rate during night**.

  During low load rate, buses with loads and their neighbours will increase their voltage as a consequence. In figure 3 it can be seen that buses 6, 7 and 8 have a p.u. voltage value above 1.

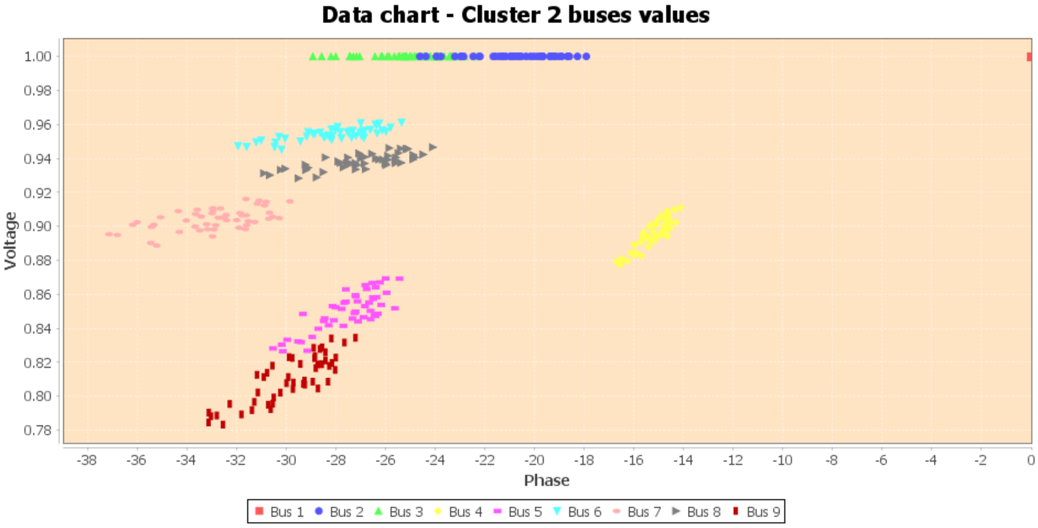


Figure 2. Scatter plot of the cluster 2 obtained from the learning set.

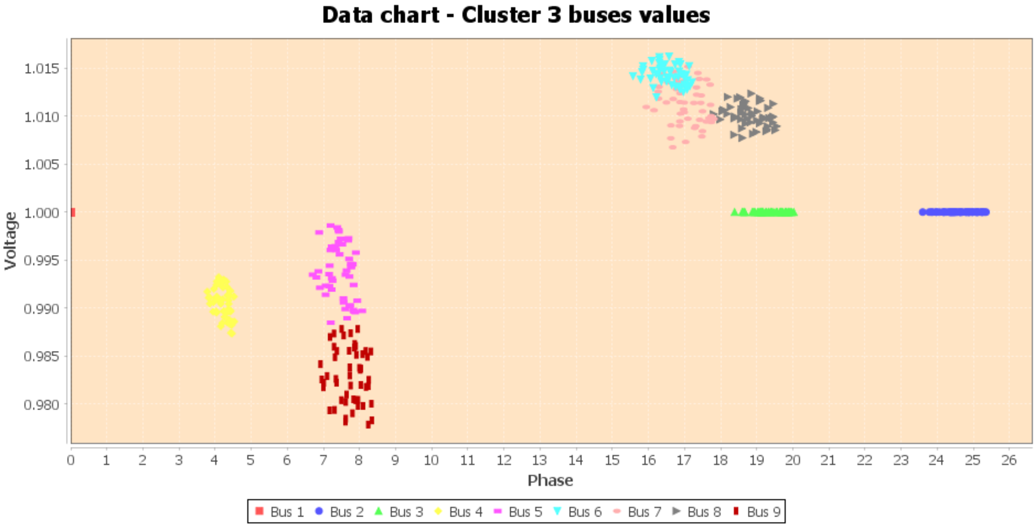


Figure 3. Scatter plot of the cluster 3 obtained from the learning set.

- **Cluster 4 – High load rate during peak hours**.

  During high load rate, the buses with loads – 5, 7 and 9 – will see a voltage drop, figure XX represents that situation.

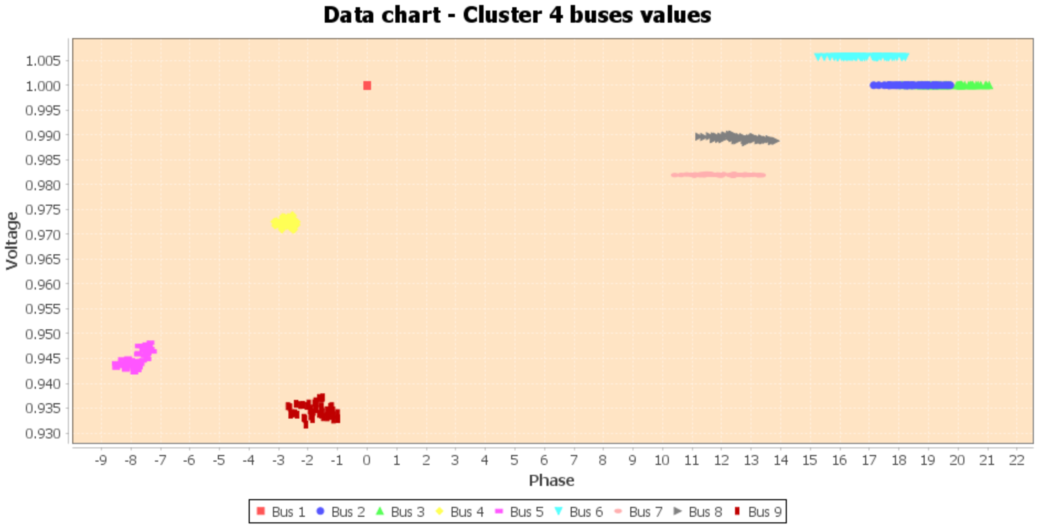


Figure 4. Scatter plot of the cluster 4 obtained from the learning set.

In figure 1 it can be seen the 200 average values from the learning set after the clustering and their centroids and also the test values with their assigned cluster after the KNN classification.

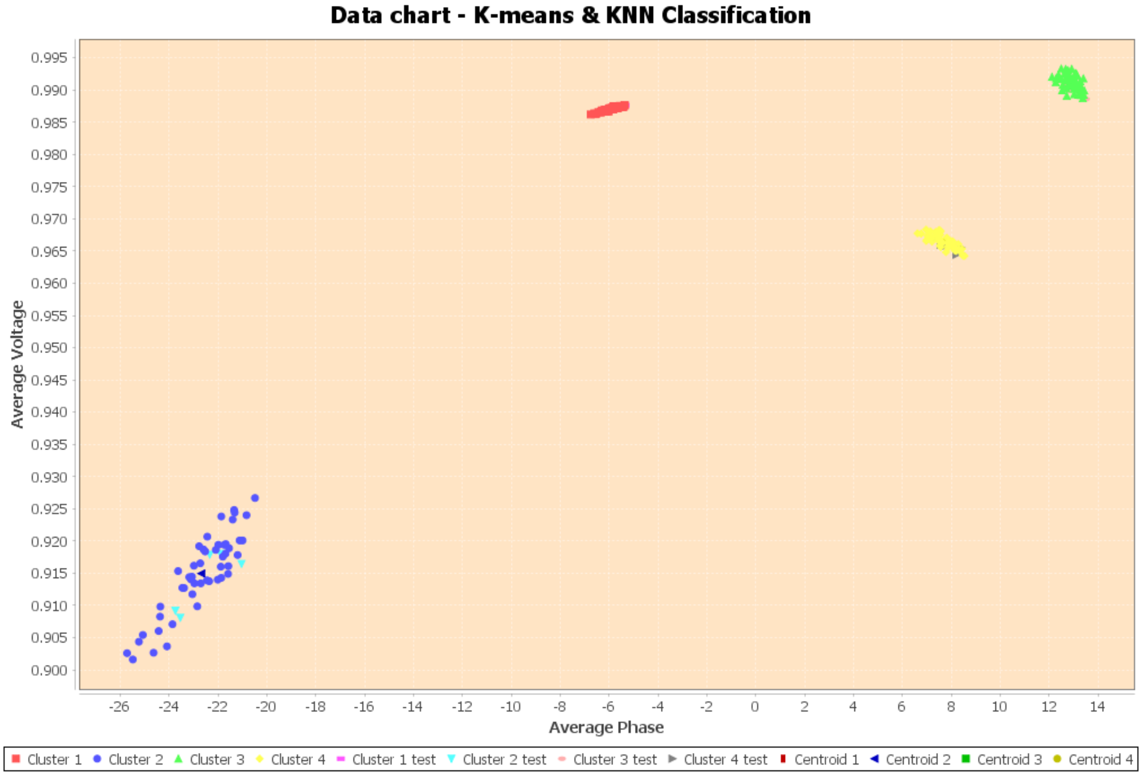


Figure 5. Scatter plot with learning set after clustering, their centroids and test values

## 1.2. KNN Classification

For the classification, we have decided to choose K=5 as the number of neighbours to decide the cluster for each test value. In the figures in the Annex I a detail of each cluster with the test values can be seen.

# Annex I

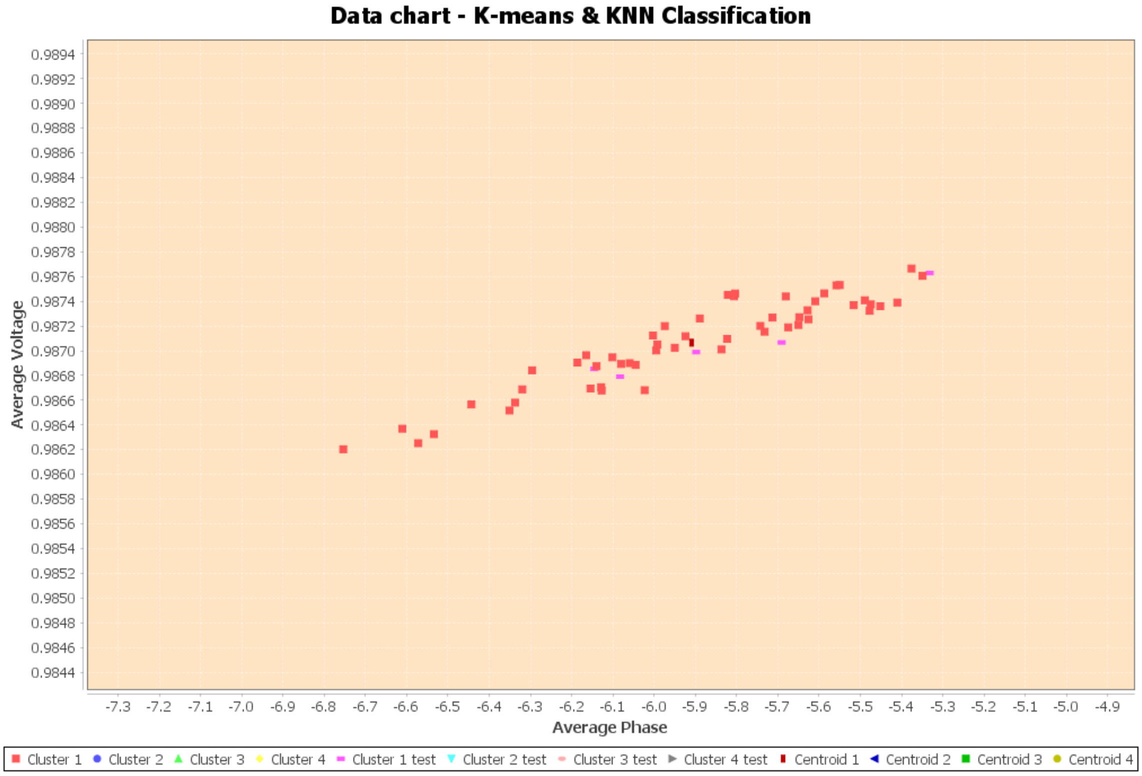


Figure 1. Detail of cluster 1, assigned values from test-data and centroid.

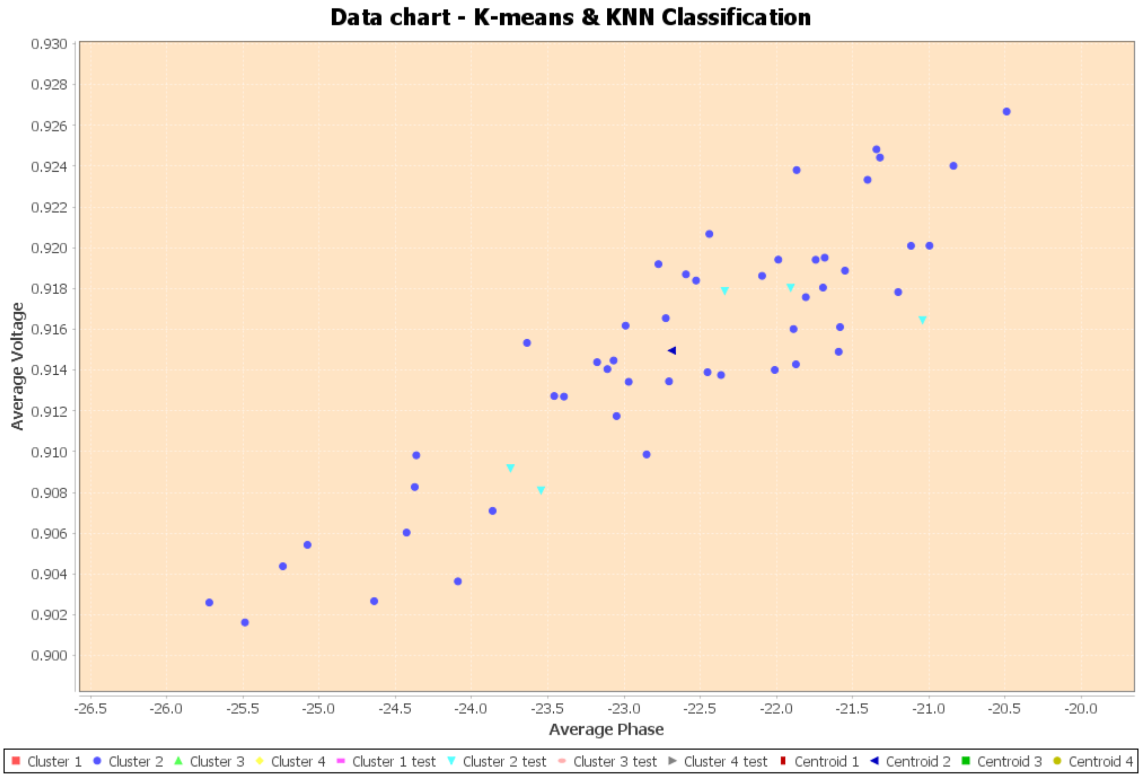


Figure 2. Detail of cluster 2, assigned values from test-data and centroid.

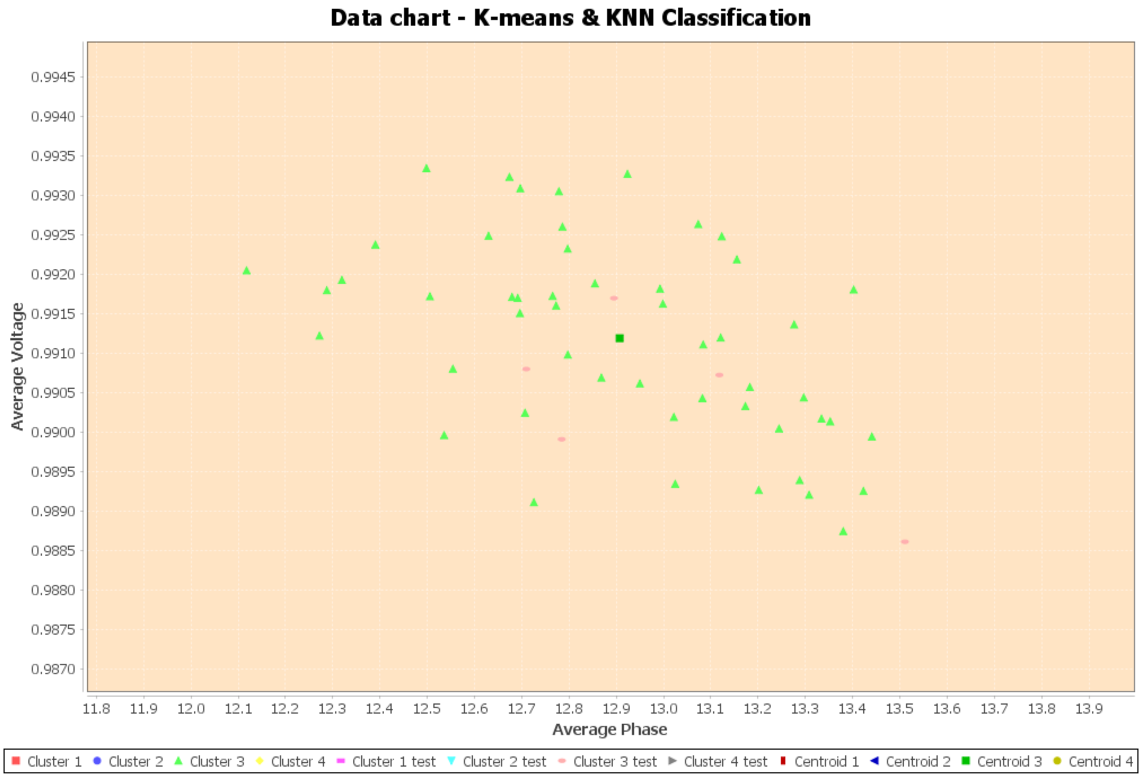


Figure 3. Detail of cluster 3, assigned values from test-data and centroid.

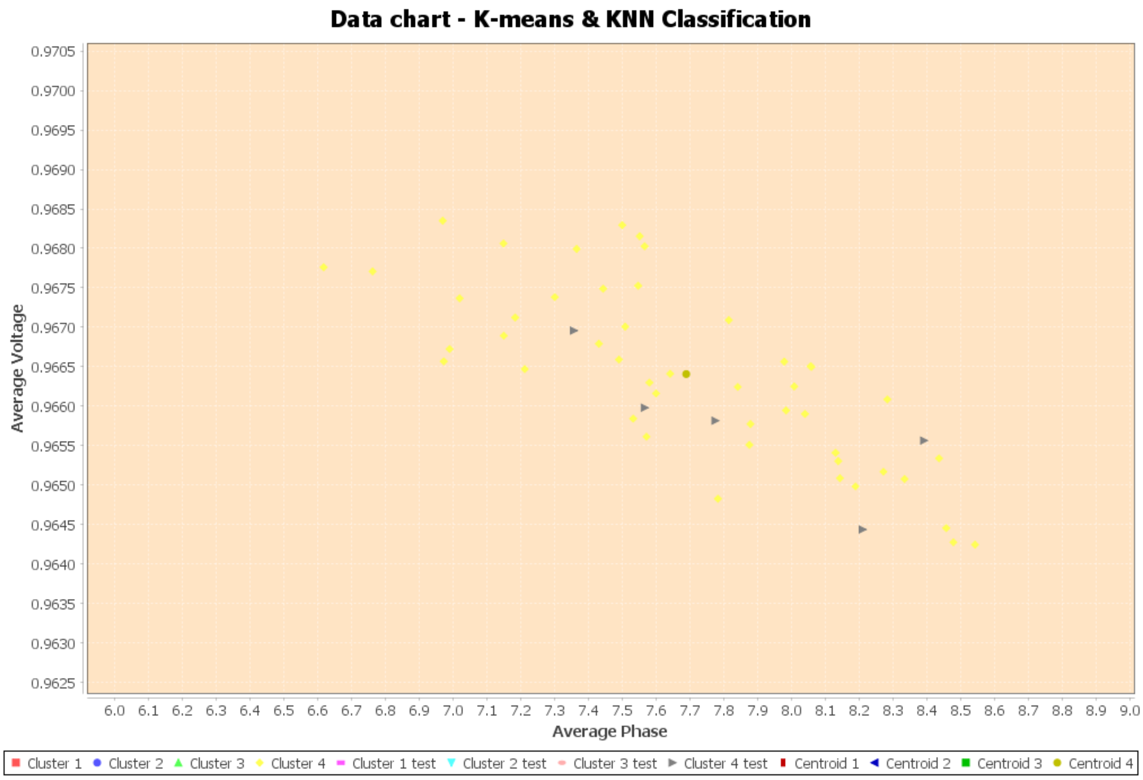


Figure 4. Detail of cluster 1, assigned values from test-data and centroid.