EH2745 Computer Applications in Power Systems

Assignment 2– Machine Learning

Group:

Hadis Pourasghar Khomami

Arnau Sans Ibós

Rafael Pacheco Bubí

1. Kmeans and labeling the states of the system

A k-Means clustering algorithm is developed and used to identify the operational states in the database. This database contains 200 different points which should be classified in four different clusters. This four clusters or states are:

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

In order to assign an initial centroid for each cluster, the Forgy method is used. This method chooses K data points from the dataset randomly and uses them as the initial center. This avoids that some clusters may not find any point which would mean it would be empty and incorrect. If it chooses the same point twice, it chooses them all again.

Once the K-Means is designed, the result is a clustering of all the points in four groups as below:

Table 1 clustering the points

Cluster number	Points
1	49
2	47
3	53
4	51
Total	**200**

As it is seen in Table 1, The 200 points are classified in the 4 clusters. In order to check if the clustering was correct, the points should be plotted. For this, per each point the mean of voltages and angles is done. Therefore, each point has two variables and can be easily plotted in a 2D graph in Figure 1.

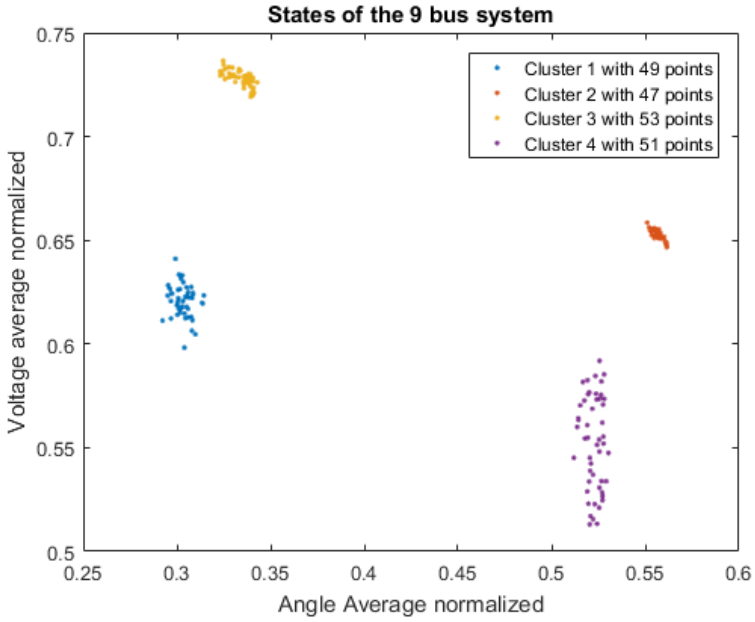

Figure 1Plot of the points in clusters

The points have been normalized in order to do the clustering so all the variables have the values between 0 and 1. As it is seen in Figure 1, the points are literally divided into 4 groups in the space. As a conclusion, the clustering was correct.

The next step is to define which cluster belongs to which state. For this, the variables of the final centroid of each cluster are back to the non-normalized values getting where it is seen the 4 centroids and their variables. The main purpose of this is to distinguish them in order to define the clusters correctly. Table 2 shows the centroid of each cluster with the value of its variables

Table 2 Centroids of the clusters

Bus Name	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1 Clark	$1 \angle 0°$	$1 \angle 0°$	$1 \angle 0°$	$1 \angle 0°$
2 Amherst	$1 \angle -20.76°$	$1 \angle 18.54°$	$1 \angle 1.83°$	$1 \angle 24.61°$
3 Winlock	$1 \angle -25.10°$	$1 \angle 19.68°$	$1 \angle -9.24°$	$1 \angle 19.36°$
4 Bowman	$0.90 \angle -15.22°$	$0.97 \angle -2.70°$	$0.99 \angle -5.29°$	$0.99 \angle 4.22°$
5 Troy	$0.85 \angle -27{,}52°$	$0.94 \angle -7.83°$	$0.98 \angle -9.83°$	$0.99 \angle 7.46°$
6 Maple	$0.95 \angle -28.10°$	$1.005 \angle 16.84°$	$1 \angle -9.24°$	$1.01 \angle 16.54°$
7 Grand	$0.90 \angle -32.91°$	$0.98 \angle 11.99°$	$0.98 \angle -8.62°$	$1.01 \angle 17.04°$
8 Wautaga	$0.94 \angle -26.99°$	$0.99 \angle 12.63°$	$1 \angle -4.04°$	$1.01 \angle 18.82°$
9 Cross	$0.81 \angle -29.86°$	$0.93 \angle -1.74°$	$0.96 \angle -8.97°$	$0.98 \angle 7.72°$

Then, each cluster must be related somehow with a state of the system. In a first look, it is seen that cluster 4 has only positive angles meanwhile the others has a large amount of negative angles.

In the electric power system, the lines have a certain impedance meaning that they will add some reactive power to the system and the angles will become negative if it is the case depending on the bus and the conditions. Therefore, if the values are positive, there may be a line or lines that are disconnected so this won't consume reactive power. As a result, **Cluster 4** can be labeled as a **Disconnection of a line for maintenance**.

Checking the remaining clusters without label, it is seen that Cluster 3 has, as a sum, higher voltages in the majority of the buses. This condition may happen when there is a low demand of energy while the generators are producing so the voltage in the system increases in the buses. Therefore, **Cluster 3** can be labeled as a **Low load rate during night.**

Considering the last two clusters, they should be labeled as Shot down of generator for maintenance or High load rate during peak hours. Respect the two labeled ones, they tend to reduce the magnitude of the voltage in all the buses except the ones where generators are. To label them, the power flow should be done.

The power flow in a line is expressed as:

$$P = \frac{U_k \cdot U_i}{X_{Line}} \cdot \sin(\theta_k - \theta_i)$$

Where:

- U_k is the voltage magnitude from bus k

-U_i is the voltage magnitude from bus i

-X_{Line} is the inductance of the line

-θ_k is the angle of the voltage from bus k

-θ_i is the angle of the voltage from bus i

In other words, the power flow of a line is calculated taking into account the voltages and angles of the two buses in which it is between. If there is a shutdown of a generator, the power flow between the bus where it belongs and the other buses which are related with a line will be affected. As a consequence, the result of the power flow will be lower than when there is a high load rate, where the power flow results will be higher. Therefore, **Cluster 1** can be labeled as **Shut down of generator for maintenance** and the remaining Cluster which is **Cluster 2** will be labeled as **High load rate during peak hours.**

This labeling is done automatically in the program and the displaying is found in Figure 2 .

```
The Disconnected line for maintenance are from cluster: 4
The Low load rate during night are from cluster: 3
The Shut down generator for maintenance are from cluster: 1
The High Load rate during peak hours are from cluster: 2
```

Figure 2 Print of the labeling of the clusters Java

2. KNN Classification

Once the K-means is working and all clusters are labeled, the next step is to classify some points in one of this clusters. The K nearest neighbors chose is 5. Normally, an odd number is selected in order to have less probabilities to have a draw between two clusters.

At the end, Figure 3 shows the clusters and all the points classified in this clusters.

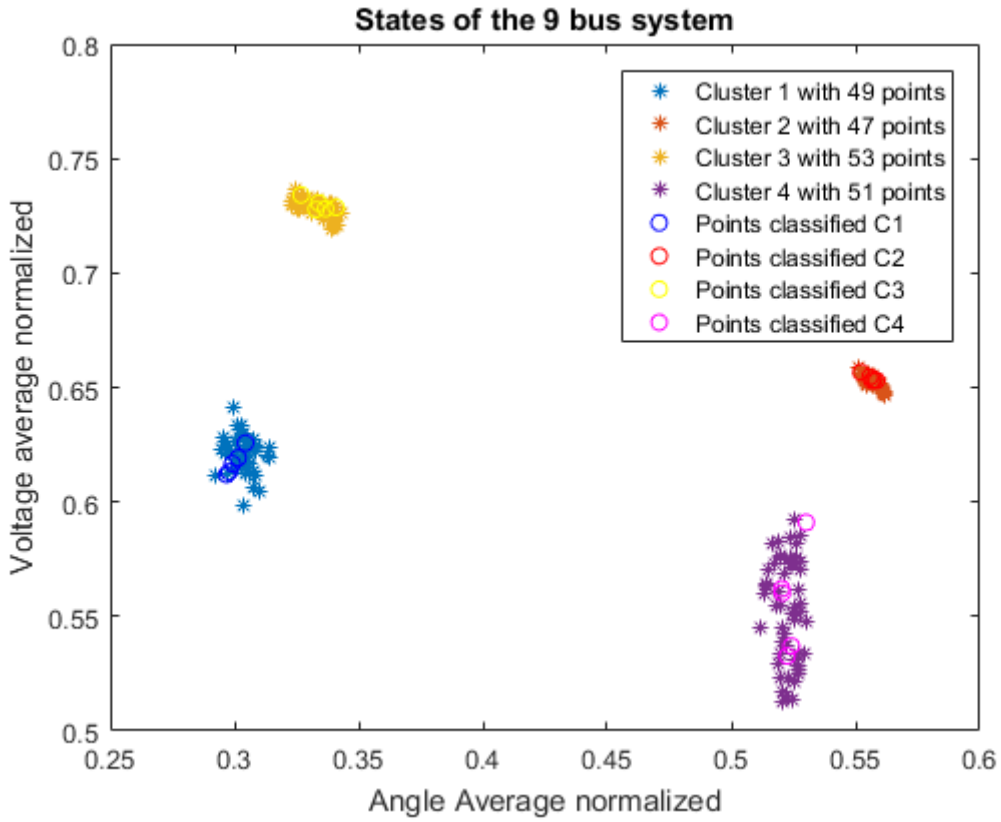

Figure 3 Clusters with the points and the points classified Kmeans - KNN

All the 20 points to classify are classified in the correct cluster which means KNN works.