# Implementation of Ferraro et al.'s paper on soft clustering

*Biswajit Sahoo*

This post is **only** the implementation of Ferro et al's paper on soft clustering that appeared in the jornal 'WIREs Computational Statistics' in 2019. The paper can be found here. The paper contains enough details, along with software packages used, to reproduce its results. More details and relevant theory can be found in the paper itself.

```
clust_data = SoftClustering::DemoDataC2D2a
library(fclust) # for fuzzy clustering
library(ppclust) # for possibilistic clustering
library(SoftClustering) # for rough clustering
library(mclust) # for model based clustering
```

```
## Package 'mclust' version 5.4.3
## Type 'citation("mclust")' for citing this R package in publications.
```

```
fuzzy_clust = FKM(clust_data,k = 2)
possi_clust = pcm(clust_data,centers = 2)
rough_clust = RoughKMeans_LW(clust_data,meansMatrix = 2, nClusters = 2)
```

```
## [1] Iteration:
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
```

```
model_clust = Mclust(clust_data,G = 2)
```

## Results of Table 1

This table displays prototype scores.

```
round(fuzzy_clust$H,3)
```

```
##        Feature1 Feature2
## Clus 1    6.874    8.369
## Clus 2    4.072    3.213
```

```
round(possi_clust$v,3)
```

```
##           Feature1 Feature2
## Cluster 1    4.230    3.430
## Cluster 2    6.739    7.905
```

```
round(rough_clust$clusterMeans,3)
```

```
##      Feature1 Feature2
## [1,]    6.967    7.610
## [2,]    4.975    3.911
```

```
round(model_clust$parameters[[2]],3)
```

```
##           [,1]  [,2]
## Feature1 4.058 6.909
## Feature2 3.217 8.234
```

## Results of Table 2

This table shows membership degrees, typicality values, upper approximations and posterior probabilities for some uncertain assignments.

```
round(fuzzy_clust$U[fuzzy_clust$clus[,2]<0.7,],3)
```

```
##         Clus 1 Clus 2
## Obj 83   0.392  0.608
## Obj 97   0.369  0.631
## Obj 120  0.578  0.422
## Obj 124  0.680  0.320
## Obj 136  0.692  0.308
## Obj 143  0.651  0.349
## Obj 163  0.324  0.676
## Obj 181  0.687  0.313
## Obj 185  0.574  0.426
## Obj 187  0.634  0.366
## Obj 190  0.440  0.560
```

```
round(possi_clust$t[fuzzy_clust$clus[,2]<0.7,],3)
```

```
##     Cluster 1 Cluster 2
## 83      0.340     0.291
## 97      0.304     0.239
## 120     0.216     0.336
## 124     0.135     0.300
## 136     0.102     0.236
## 143     0.181     0.361
## 163     0.311     0.207
## 181     0.153     0.347
## 185     0.255     0.389
## 187     0.232     0.424
## 190     0.211     0.208
```

```
round(rough_clust$upperApprox[fuzzy_clust$clus[,2]<0.7,],3)
```

```
##       [,1] [,2]
##  [1,]    0    1
##  [2,]    0    1
##  [3,]    1    1
##  [4,]    1    1
##  [5,]    1    1
##  [6,]    1    1
##  [7,]    0    1
##  [8,]    1    1
##  [9,]    1    1
## [10,]    1    1
## [11,]    1    1
```

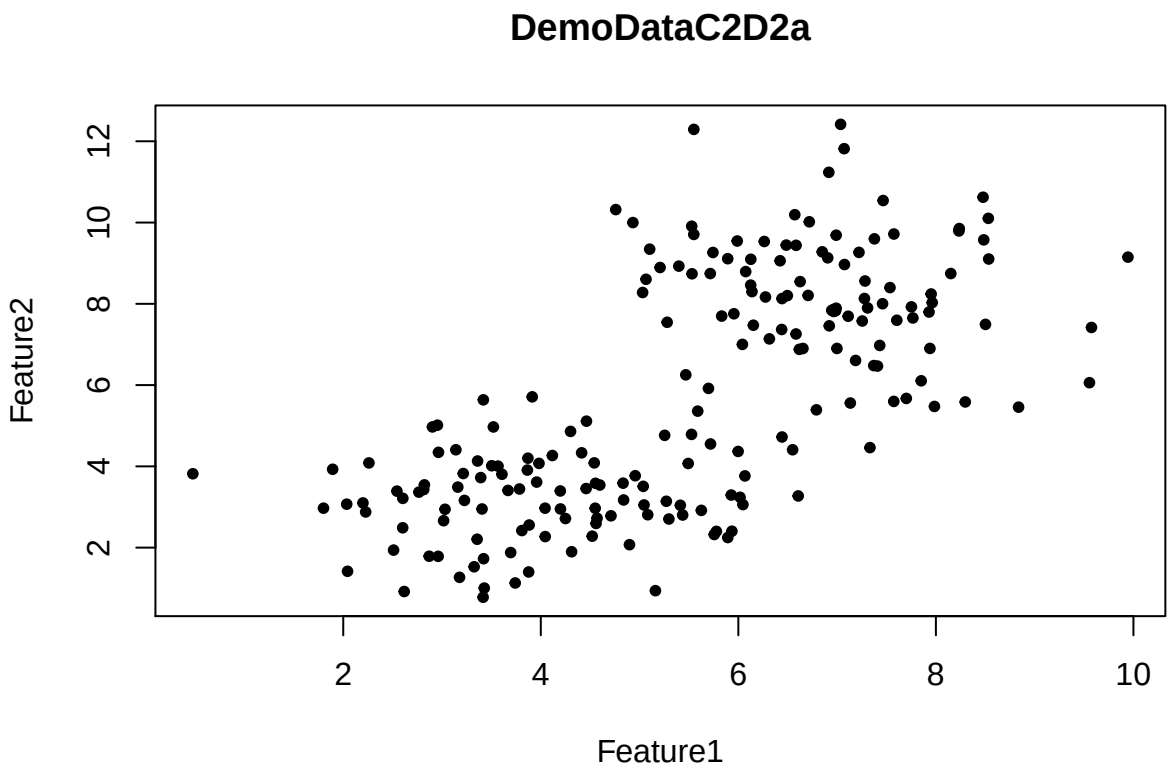
```
round(model_clust$z[fuzzy_clust$clus[,2]<0.7,],3)
```

```
##        [,1]  [,2]
##  [1,] 0.581 0.419
##  [2,] 0.596 0.404
##  [3,] 0.109 0.891
##  [4,] 0.016 0.984
##  [5,] 0.007 0.993
##  [6,] 0.041 0.959
##  [7,] 0.736 0.264
##  [8,] 0.019 0.981
##  [9,] 0.172 0.828
## [10,] 0.108 0.892
## [11,] 0.402 0.598
```

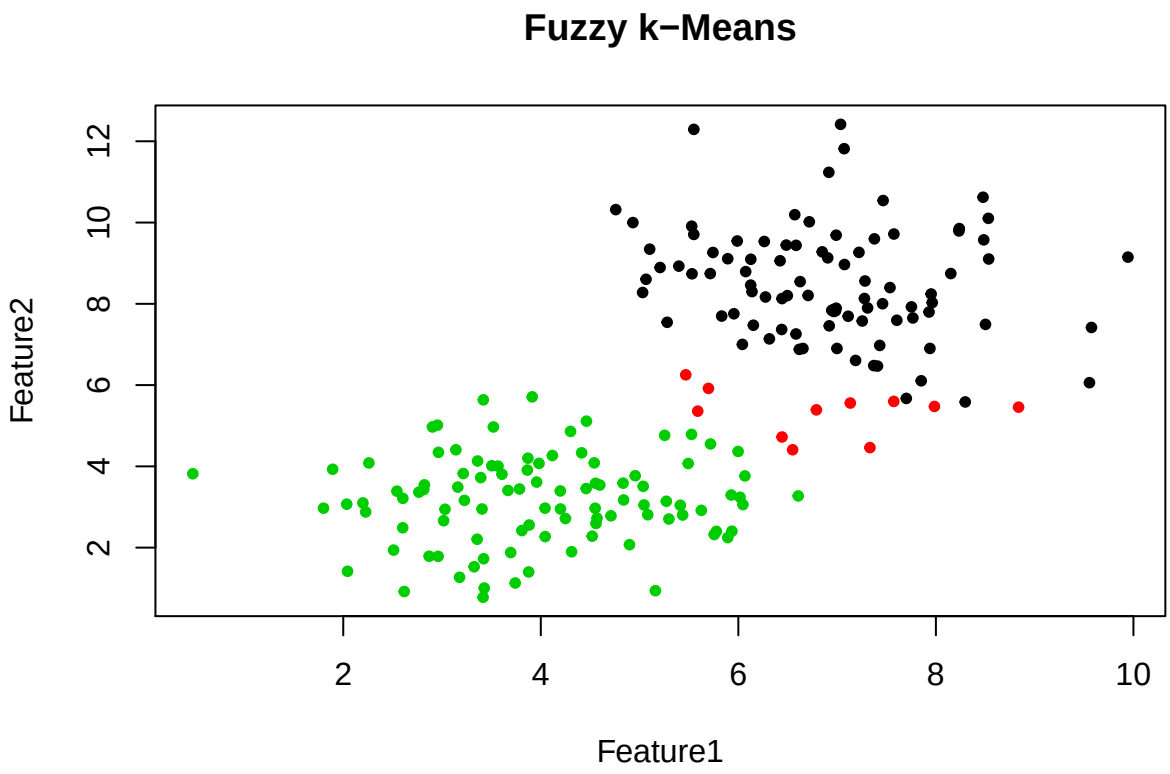
## Plot of Original data and Partitions obtained by various methods

We have changed the color of uncertain data points. In the original paper, those are shown in yellow, in this post those are red. We have also changed the markers from unfilled circles to filled circles.

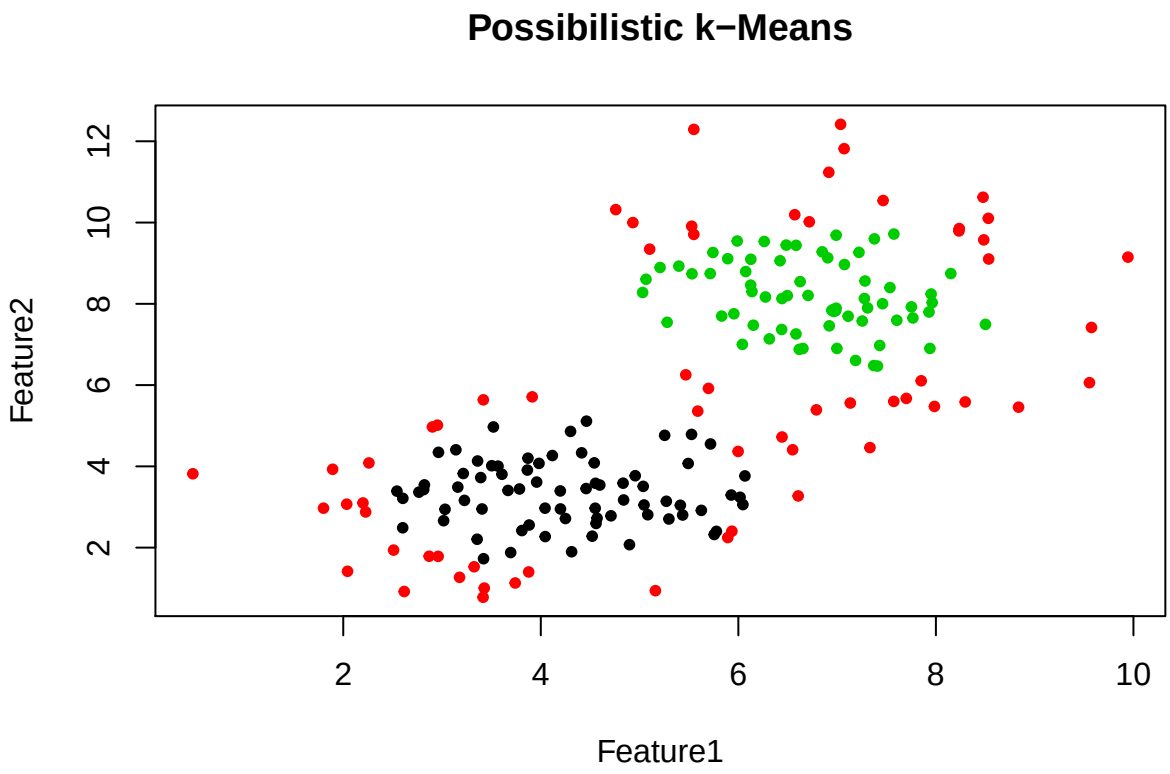
```
col_fuzzy = rep(NA,200)
col_fuzzy[fuzzy_clust$U[,1]>0.7] = 1
col_fuzzy[fuzzy_clust$U[,2]>0.7] = 3
col_fuzzy[fuzzy_clust$U[,1]<0.7 & fuzzy_clust$U[,2]<0.7] = 2

col_possi = rep(NA,200)
col_possi[possi_clust$t[,1]>0.43] = 1
col_possi[possi_clust$t[,2]>0.43] = 3
col_possi[possi_clust$t[,1]<0.43 & possi_clust$t[,2]<0.43] = 2

col_rough = rep(NA,200)
col_rough[rough_clust$upperApprox[,1]==1] = 1
col_rough[rough_clust$upperApprox[,2]==1] = 3
col_rough[rough_clust$upperApprox[,1]==1 & rough_clust$upperApprox[,2]==1] = 2

col_model = rep(NA,200)
col_model[model_clust$z[,1]>0.7] = 1
col_model[model_clust$z[,2]>0.7] = 3
col_model[model_clust$z[,1]<0.7 & model_clust$z[,2]<0.7] = 2

plot(clust_data, pch = 20) # Figure 1
title(main = "DemoDataC2D2a")
```



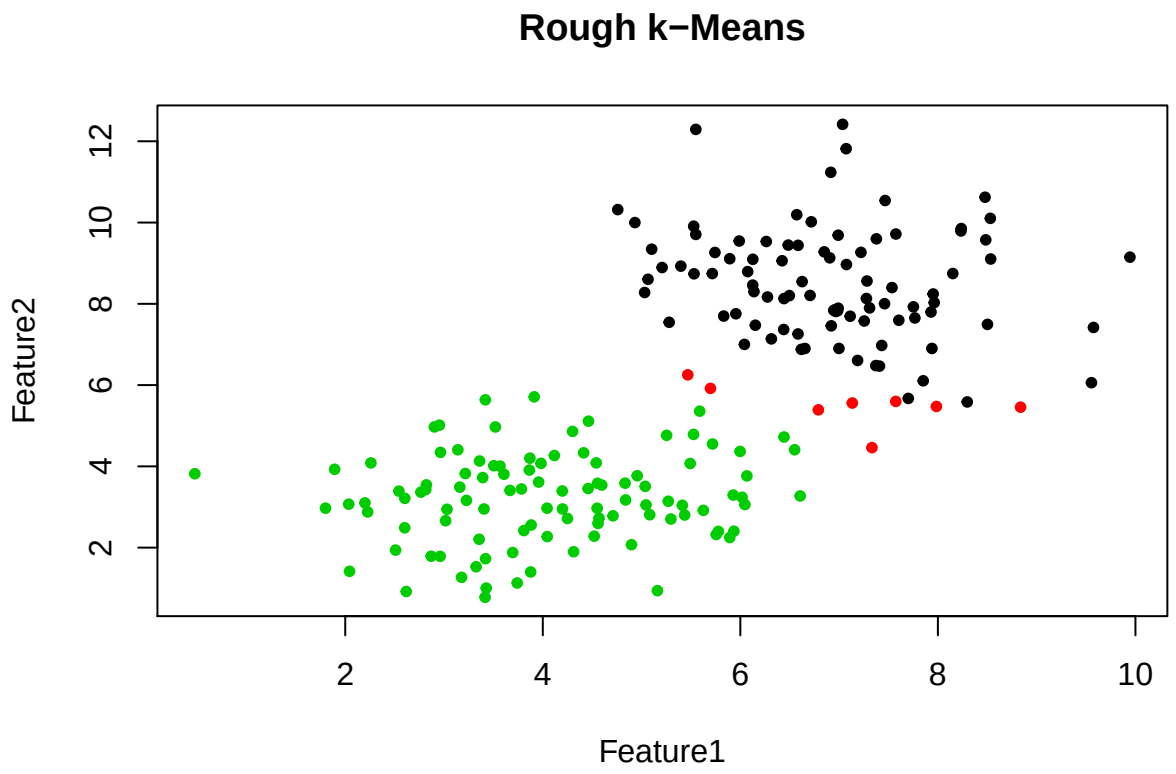

```
plot(clust_data, col = col_fuzzy, pch = 20);title(main = "Fuzzy k-Means")
```



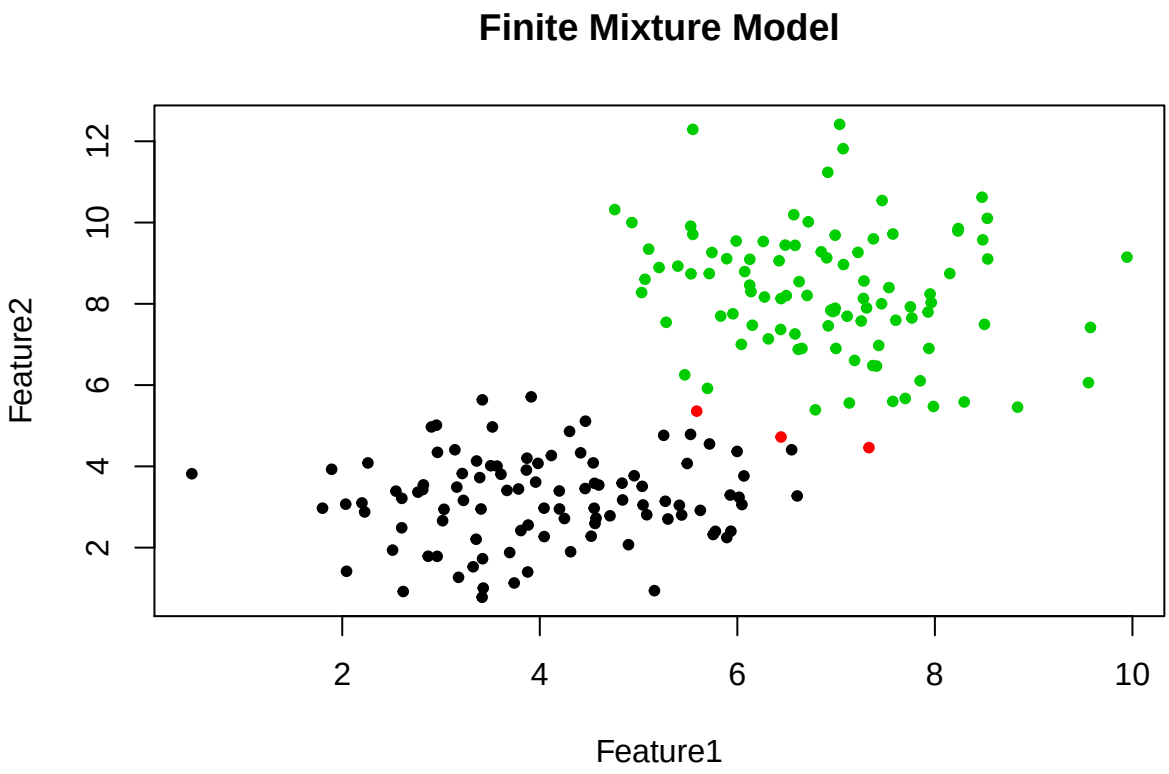

```
plot(clust_data, col = col_possi, pch = 20);title(main = "Possibilistic k-Means")
```




```
plot(clust_data, col = col_rough, pch = 20);title(main = "Rough k-Means")
```




```
plot(clust_data, col = col_model, pch = 20);title(main = "Finite Mixture Model")
```




Last modified: $19^{th}$ June, 2019.