Problemes APA
Problema 12: Clustering de les dades artificials de Cassini

Lluc Bové

Q1 2016-17

Volem analitzar un problema d'agrupament amb dades en 2D usant la rutina mlbench.cassini. Generem dades en 3 grups amb el codi:

```
library(mlbench)
N <- 2000
data.1 <- mlbench.cassini(N, relsize = c(1,1,0.25))
plot(data.1)
```

Veiem que les estructures externes tenen forma de plàtan i entre elles hi ha un cercle amb menys densitat de dades. El plot anterior mostr la veritat de les dades (els 3 grups generats). Si ara fem:

```
plot(x=data.1$x[,1], y=data.1$x[,2])
```

Veurem les dades en brut (el que rebrà el mètode de *clustering*). Es demana:

1. **Decidiu per endavant quin mètode de *clustering* hauria de treballar millor i amb quins paràmetres**
2. **Apliqueu k-means un cert nombre de vegades amb $k = 3$ i observeu els reultats**
3. **Apliqueu k-means amb una selecció de valors de k al vostre criteri (20 cops cadascun) i monitoritzeu l'índex de Calinski-Harabasz mitjà; quin k es veu millor?**
4. **Apliqueu l'algorisme E-M amb una selecció de valors de k al vostre criteri (10 cops cadascun) i observeu els resultats. Comproveu els resultats contra les vostres expectatives (apartat 1).**