Assignment 4

Due: Tue, Dec 8, 23h59, by email to `jaime.cardoso@fe.up.pt`

1. **Batman and Joker** Even after all the attempts by the Gotham police to capture him, Joker still continues to destroy the property of the city. This time he has planted a deadly computer virus at different ($\hat{K}$) locations in the city. The virus has this strange property that it spreads to *nearby* computers very fast once triggered, but does not go too far from the source. Joker has triggered this virus 2 times already, but with slightly different properties. It is impossible to know which computer exactly is the source of the virus.

   One the other hand, Batman is now old, and does not have any skills in Machine Learning. He has to choose a candidate to be the next Batman (and win the Bat-mobile) based on how well someone is able to identify the spread and source of the virus across the city from the data about the locations of the infected computers.

   (a) As your first task, analyze the geographical spread of the infected computers by plotting the two data-sets. (You may find Joker's hidden signature in the impact of the virus).

   (b) Implement K-means and Gaussian Mixture Model techniques to cluster the input data. Choose two different values of $K = \{4, 6\}$.

   (c) For both the data-sets, and for both the clustering approaches, start with 5 different initial means, and show the obtained classification.

2. Although not discussed in the classes, the evaluation of the quality of a clustering and the comparison of different clustering methods is an important topic (similarly to the evaluation of supervised methods). Study and explain in a few sentences how, given a dataset, you could compare the quality of the clustering generated by two different clustering algorithms (say k-means and Mixture of Gaussian model).