Practical 7:
(for Lecture 7: Clustering)

Have a look at the simple k-means program you have been given. The program can work from a randomly generated set of points or a defined set. [Note, it can look for clusters of different ***k***s; but for plotting purposes it fixed at ***k***=3.]

1) Use the **`init_board`** method to randomly generate 15 data-points; store this output and set the **`data`** variable to it.

   Now run this dataset 10 times and note the clusters found by k-means.

   Carefully record how many of these clusters are the same/different (or roughly similar/different, if none are exactly identical). Discuss and comment on the variations you find across these runs.

2) Now, create your own dataset, by hand, with 10 data-points. You should construct this data-set to have three very clear clusters (a bit like the simple 6-point example shown in the program).

   Now run this set 10 times and note the clusters found by k-means.

   Report the results of these runs and the extent to which the same clusters are found. Modify some of the points in the dataset to ensure that most of the time the k-means finds the same 3 clusters. Report the changes you had to make to the dataset to achieve this consistent clustering.

3) Do some web searches (in Google and Google Scholar) on methods that have been developed to improve the clusters found by k-means (especially, with respect to the issue of it finding different clusters on different runs).

   Describe one improvement with reference to the literature you read on the topic (n.b., provide full citations to the relevant papers found).