

K-Means Algorithm Pseudocode

1. Dataset loading:
Load data.csv into 2D numpy array "Data"
Load centers.csv into 2D numpy array "Centers"
Initialize 2 2D lists named "Clusters" and "Temp_Clusters", each containing 6 1D lists for the 6 centers.
2. Algorithm:
"iteration" = 0
While True:

 for each sample "S" in "Data":
 for each center "C" in "Centers":
 Determine distance between "S" and "C". Keep min distance "C"
index stored always
 You now have the closest center "C" (Suppose, "C" has index "i"
in "Centers")
 Append "S" in "i"th list of "temp_Clusters" (Just append the index
of "S" found in "Data")
 Now your "Temp_Clusters" 2D list contains 6 1D lists, each
containing samples belonging to corresponding centers

 for each 1D list "L" in "Temp_Clusters":
 Determine avg
 These 6 averages are the new centers. Assign these to "Centers".
 "Iteration"++
 if "Iteration" > 1:

 for each sample S in Data:
 See in which list it was in "Clusters" and which list it is now in
"Temp_Clusters"
 If the lists are different, then there has been a shift
 See how many such shifts has taken place and store the number in "Shift"
 if "Shift" < 10:
 Assign "Temp_Clusters" to "Clusters" and break the loop
(convergence)
 Assign "Temp_Clusters" to "Clusters"

 "Clusters" will contain your desired clusters and "Centers" will contain your
desired centers at the end of loop

Plotting:

Plot them with appropriate color.