import pandas as pd

import numpy as np

import matplotlib.pyplot as graph

import math

from sklearn.cluster import KMeans

pathtofile = "k-means.csv"

data = pd.read\_csv(pathtofile, delimiter=",")

X = data.iloc[:,0]

K = 3

initialcenter = np.array([[3,3],[6,2],[8,5]])

def ClosestCentro(X, centro):

K = centro.shape[0]

id = np.zeros((X.shape[0],1), dtype=np.int8)

for i in range(X.shape[0]):

dist = np.linalg.norm(X[i] - centro, axis=1)

min\_dist = np.argmin(dist)

id[i] = min\_dist

return id

id = ClosestCentro(X, initialcenter)

print("The result for all examples:\n")

print(id[:300])

def ComputeCentro(X, id, K):

m, n = (np.random.randint(3,size=(2,1)))

K = np.array([3])

#centro = np.zeros((K,n))

centro = np.zeros((X.shape[0],1), dtype=np.int8)

for k in range(X.shape[0]):

#centro[k,:] = np.mean(X[id.ravel()==k,:], axis=0)

centro = centro

return centro

centro = ComputeCentro(X, id,K)

print(centro)

def plottingDataPoints(X, id, K):

colors = [graph.cm.tab20(float(i) / 10) for i in id]

#graph.scatter(X[:,0], X[:,1], c=colors, alpha=0.5, s=2)

def plottingProgressMeans(X, centro, prev, id, K, i):

plottingDataPoints(X, id, K)

#graph.scatter(centro[:,0], centro[:,1],marker='x', c='k')

"""

for j in range(centro.shape[0]):

graph.plot([centro[j, :][0], prev[j, :][0]],

[centro[j, :][1], prev[j, :][1]], c='k')

graph.title('Iteration number {:d}'.format(i+1))

"""

def runkMeans(X, initialcenter, maxiters, plotprog):

#m, n = X.shape

m, n = (np.random.randint(3,size=(2,1)))

K = initialcenter.shape[0]

centro = initialcenter

prevcentro = centro

#id = np.zeros((m,1))

id = np.zeros((X.shape[0],1), dtype=np.int8)

graph.ion()

for i in range(maxiters):

print("Clustering K means in iteration {}/{}...".format(i, maxiters))

id = ClosestCentro(X, centro)

if plotprog:

plottingProgressMeans(X, centro, prevcentro,id,K,i)

prevcentro = centro

centro = ComputeCentro(X, id, K)

return centro, id

K = 3

maxiters = 10

initialcenter = np.array([[3,3],[6,2],[8,5]])

centro, id = runkMeans(X, initialcenter, maxiters, plotprog=True)