Project Part 2

Unsupervised Learning (K-means)

Prepared by **Rohit Roongta** 26st October 2019

K-MEANS ALGORITHM

Introduction

K-means algorithm is an unsupervised learning algorithm used for clustering. It divides the data points into k-clusters of similar objects. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. Similarity is measured by calculating the distance (Euclidean Distance).

Objective

To implement the k-means algorithm on the given dataset, which contains a set of 2-D points, using to different strategies for choosing the initial cluster centres.

- Strategy 1: randomly pick the initial centres from the given samples.
- Strategy 2: pick the first centre randomly; for the ith centre (i>1), choose a sample (among all
 possible samples) such that the average distance of this chosen one to all previous (i-1) centres is
 maximal.

Formula:

- 1) Euclidean distance (D) = $\sqrt{(x_2 x_1)^2 + (y_2 y_1)^2}$
- 2) Strategy 2 = $argmax_i \frac{1}{n} \sum_j D(C_i, C_j)$ where 'i' represents the point index of the new center, $j \in n$ represents the point index of the existing centres
- 3) Objective function = $\sum_{i=1}^{k} \sum_{x \in D_i} ||x \mu_i||^2$

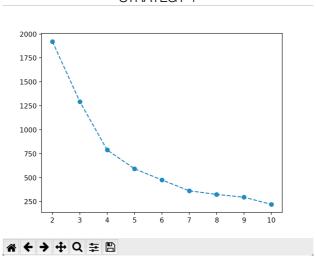
[D represents the data set and μ_i is the centroid]

Pseudo Code (Algorithm)

- Given: dataset and a number of k (ranging from 2 to 10)
- Initialize centroids (either strategy 1 or strategy 2)
 - o <u>Do</u> classify data points according to the nearest centroid (μ_i)
 - \circ Recompute μ_i
 - o <u>Until</u> they converge (no change)
- Return centroid list

Result

STRATEGY 1



* Objective Function Value: 1921.0334858562055

Initial Centroid List: [(2.641451406258703, 2.6220682177926107), (2.3643033531884674, 1.0520971253700746), (3.252246409117296, 2.4788534037773373), (7.529630088974836, 8.796171120440778)]
Final Centroid List: [(2.8337661044723124, 6.91895689893875), (3.2428534716041297, 2.551979054939448), (7.252626831256577, 2.4001582635520533), (6.579576431887845,

7.57333594820388)]
******** Objective Function Value: 788.9645806635216

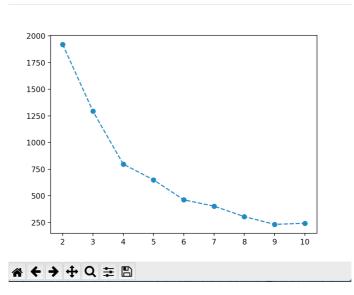
-K=5-

Initial Centroid List: [(6.856532248221895, 7.724688252918387), (4.807540932988571, 3.0346495391326473), (4.783632112114197, 7.106442879583119), (6.633523323424615, 0.9802070492997959), (3.009328298642023, 1.4506571693366839)]
Final Centroid List: [(7.299749694765005, 8.413318379491713), (5.339072124427115, 4.465511754739297), (3.1507276114491836, 7.121929064073764), (7.556167822397726, 2.235167959857534), (2.681986334188929, 2.094615867800809)]

Objective Function Value : 476.1187516763532

Final Celtrolia List: [6:2214497), (4:951859575817064, 4:117566936783995), (3:8427898922439203, 5:55466947744628), (5:7714422285783105, 9.040753944871518), (3:7261084438667154, 6:908979840290732), (4:951859575817064, 4:117566936783895), (3:8427898922439203, 5:55466947744628), (5:7714422285783105, 9.040753944871518), (3:7261084438667154, 5:2043243906921255), (3:5446126706973637, 0.9426188179429711), (2:5366923982985132, 2:2422267166268712)]

Final Centroid List: [(7:914309977818314, 8:519909807700076), (7:556167822397726, 2:235167959857534), (2:183214622031253, 7:703553409180906), (6:15468228405522, 5:701407205679413), (5:183123899476708, 3:2140910620351697), (3:060422754557576, 6:029718158624873), (5:021776597933489, 7:8240125818092165), (4:331939478230939, 4:336266052296369), (3:1690614508664035, 0.8143251472991676), (2:242047519125402, 3:251007486318421)]



Initial Centroid List: [(3.9872431110331834, 4.042547799500241), (2.389526061607529, 7.221955635454229), (4.783632112114197, 7.106442879583119), (3.120736964994129, 0.4897907901976055), (6.792518315497752, 2.5620809523410912)]
Final Centroid List: [(3.4955665791995627, 3.5661123157286907), (2.6756987971688004, 6.982927498982885), (6.489843945177897, 7.558852441314683), (3.145061482959145,

K = 6

Initial Centroid List: [(8.007064409166201, 2.7753199731436236), (8.076416518034618, 9.271620015802068), (6.470118291290549, 5.540355428924968), (5.481219654489373, 6.551717768804221), (2.9100822099635435, 7.519439837064963), (4.502364450590697, 2.928880404940474)]
Final Centroid List: [(7.5561678223937726, 2.235167958857534), (7.91430997818314, 8.519909807700076), (5.2305366674047375, 4.279342496049099), (5.2402829638043125, 7.531310293267821), (2.541652523310345, 7.002678323536489), (2.681986334188929, 2.094615867800809)]

***** Objective Function Value: 462.9263558248375

8.490601185961385), (6.215390304639243, 6.261392254422516), (7.444728022754422, 2.418493184045437), (2.499984886739594, 7.99174944862789517), (7.494728022754422, 2.418493184045437), (2.499984886739594, 7.99174944862789517), (7.49984886739594, 7.99174944862789517), (7.56483249146484, 8.556689279063415), (5.464277356727894, 6.837713536435891), (7.556167822397726, 2.235167959857534), (2.183214622031253, 7.703553409180906)]
*********Objective Function Value: 402.91835862050726

K = 8

Initial Centroid List: [(6.976905728972921, 7.965091986724786), (2.7845243023650736, 6.618471581818432), (4.807540932988571, 3.0346495391326473), (2.515552088741975, 6.41071773765037), (8.226274849635631, 2.260487012503675), (1.9256185314039267, 2.7385763244050594), (5.7714422285783105, 9.040753944871518), (3.047435882855882, 1.5580763520111716)]

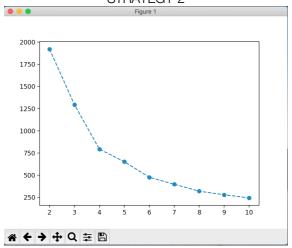
Final Centroid List: [(7.914309977818314, 8.519909807700076), (2.183214622031253, 7.703553409180906), (5.345603315617595, 4.203354778508424), (3.138347677015053, 5.933723217937643), (7.556167822397726, 2.235167959857534), (2.214068054222364, 3.286445548570339), (5.2402829638043125, 7.531310293267821), (3.1660397272577874,

0.8616886118323305)]
******* Objective Function Value: 305.3870732847836

K = 9
Initial Centroid List: [(8.226274849535631, 2.260487012503675), (4.91330965187404, 3.2577242509859516), (3.5446126706973637, 0.9426188179429711), (8.201293857837541, 9.212915410514553), (2.200114959631794, 1.5386322076504713), (8.094693454952871, 7.791998464329252), (4.758510501654859, 8.243174587119283), (2.078985690973731, 7.167393130690192), (4.729351540291086, 3.808390445364688)]
Final Centroid List: [(7.556167822397726, 2.235167959857534), (4.912514973376666, 3.5631409634126254), (3.1690614508664035, 0.8143251472991676), (7.914309977818314,

-- K = 10 -





Final Centroid List: [[6.497249620849709, 7.522972929895171), (5.477400388679034, 2.254981027984723), (2.5614644894663545, 6.088613382895479)]

Initial Centroid List: [(1.0521742722626426, 3.8894374123577706), (8.875780722296865, 8.960923612139595), (3.0410170219628694, -0.3613848735372547), (3.021056869801129,

Final Centrold List: [(3.2285300905383707, 2.5240486292057867), (6.856583330228901, 7.661434199806484), (7.252626831256577, 2.4001582635520533), (3.153427001856, 6.91292070093253)]

* Objective Function Value: 792.7110095863353

Initial Centroid List: [(3.0104761241001885, 6.542864553074541), (7.597313421987085, 1.165047426798992), (8.875780722296865, 8.960923612139595), (3.0410170219628694, -0.3613848735372547), (3.021056869801129, 9.26213795798542), (2.200114959631794, 1.5386322076504713)]

Final Centroid List: [(5.464277356727894, 6.837713536435891), (7.414192434680615, 2.3216911383868664), (7.756483249146484, 8.556689279063415), (3.145061482959145, 0.9077065486588153), (2.5633381461259046, 6.978224800606624), (3.4955665791995627, 3.5661123157286907)]

- K = 7 -

Initial Centroid List: [(2.4898969303766347, 8.400478632861432), (7.597313421987085, 1.1650474267989992), (8.875780722296865, 8.960923612139595), (3.0410170219628694, -0.3613848735372547), (3.021056869801129, 9.26213795798542), (2.200114959631794, 1.5386322076504713), (8.619479454228491, 2.98598319047101)]
Final Centroid List: [(2.5633381461559046, 6.97822480066624), (7.42803425946345, 2.28291866452028), (7.914309977818314, 8.51990807700076), (3.145061482959145, 0.9077065486588153), (5.021776597933489, 7.8240125818092165), (3.4955665791995627, 3.5661123157286907), (6.15468228405522, 5.701407205679413)]

***** Objective Function Value: 397.44812244413686

Tisal Centroid List: ([5.3321424077429915, 4.444481111506235), (7.914309977818314, 8.519909807700076), (3.1690614508664035, 0.8143251472991676), (4.923731090150018, 7.833731542772698), (7.056682932407747, 1.333196794745837), (2.5131744928356072, 6.883061799423911), (7.986758244802879, 3.012729309091755), (2.242047519125402,

3.251007486318421)] ******** Objective Function Value: 319.47442215977026

--- K = 9 --

 $Initial\ Centroid\ List: [(7.578050252952152, 3.8248701734522355), (3.021056869801129, 9.26213795798542), (3.0410170219628694, -0.3613848735372547), (8.875780722296865, 8.960923612139595), (2.200114959631794, 1.5386322076504713), (1.69565649157431, 7.680824581467082), (7.597313421987085, 1.1650474267989992), (7.5976350478966745, 7.9392400048313085), (1.3483715993964611, 3.963796377807688)]$

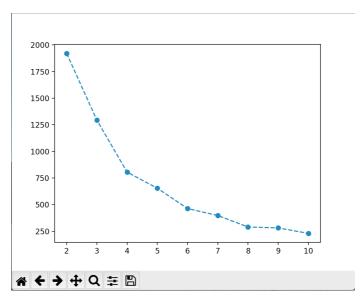
Final Centroid List: [(a.91251497337666, 3.5631409634126254), (4.859398750231154, 7.941638208906054), (3.392397796044764, 0.5065576443944665), (7.914309977818314, 8.519909807700076), (2.5755081335118746, 2.3149966532137327), (2.4631734791294, 6.989095694190998), (7.556167822397726, 2.235167959857534), (6.15468228405522, 5.701407205679413), (2.46204926319718, 4.294396379247468)]

- K = 10 -

K = 10

Ki =

** Objective Function Value: 243.45056737269005



Initial Centroid List: [(8.218975256689072, 8.951050496546863), (3.0410170219628694, -0.3613848735372547)]

8.960923612139595)

0.500525012.133353)]
Final Centroid List: [(6.783746087903545, 2.8501999866465444), (3.3426476922451784, 6.926028030100171), (2.852351493110535, 2.2818648297203246), (7.179286211861507,

- K = 5 -Initial Centroid List: [(2.100548909914467, 1.4414401856186492), (8.875780722296865, 8.960923612139595), (3.021056869801129, 9.26213795798542), (7.597313421987085, 1.1650474267989992), (3.0410170219628694, -0.3613848735372547)]
Final Centroid List: [(3.55458080959230115, 3.628130619771138), (6.8042102875897745, 7.656222203211083), (3.00668800078913, 7.034815026483998), (7.414192434680615, 2.3216911383868664), (3.145061482959145, 0.9077065486588153)]

* Objective Function Value : 653.9567908403943

Initial Centroid List: [(4.949560744079336, 3.22624796965257), (8.875780722296865, 8.960923612139595), (3.0410170219628694, -0.3613848735372547), (3.021056869801129, пица сепион съвт. 1 (н.эм-эконогичног укъж), 3.22024/ укочокост /, (к.о.готови/2222/оковъ, 8.96092/s0127.3995), (3.041017/0219628694, -0.3613848735372547), (3.021056869801129, 9.26213795798542), (7.597313421987085, 1.1650474267989992), (1.69656849157431, 7.680824581467082)]

Final Centroid List: ([6.2305366674047375, 4.279342496049099), (7.914309977818314, 8.519909807700076), (2.681986334188929, 2.094615867800809), (5.2402829638043125, 7.531310293267821), (7.556167822397726, 2.235167959857534), (2.541652523310345, 7.002678323536489)]

* Objective Function Value: 462.9263558248375

(3.145061482959145,

K = 8 Initial Centroid List: [(2.695113024232428, 5.939673524983461), (8.875780722296865, 8.960923612139595), (3.0410170219628694, -0.3613848735372547), (7.597313421987085, 1.1650474267989992), (3.021056869801129, 9.26213795798542), (2.200114959631794, 1.5386322076504713), (7.5976350478966745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.5386322076504713), (7.5976350478966745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.5386322076504713), (7.5976350478966745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.5386322076504713), (7.5976350478966745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.5386322076504713), (7.5976350478966745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.5386322076504713), (7.5976350478966745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.5386322076504713), (7.5976350478966745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.5386322076504713), (7.5976350478966745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.5386322076504713), (7.5976350478966745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.5386322076504713), (7.5976350478966745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.5386322076504713), (7.5976350478966745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.5386322076504713), (7.597635047896745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.53863220765047413), (7.597635047896745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.53863220765047413), (7.597635047896745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.53863220765047413), (7.597635047896745, 7.9392400048313085), (8.619479454228491, 2.020114959631794, 1.53863220765047413), (7.597635047896745, 7.9392400048313085), (8.61947945428491, 2.020114959631794, 1.020114959631794, 1.020114959631794, 1.020114959631794, 1.020114959631794, 1.020114959631794, 1.020114959631794, 1.020114959631794, 1.020114959631794, 1.020114959

1.1030474267999932), (3.02109069801129, 9.20213793798942), (2.200114999931794, 1.35903220769047103), (7.3976390476906743, 7.3932400404513063), (6.61947949422482298828319047101)]
Final Centroid List: ([2.596501078790139, 6.85941978482168), (7.914309977818314, 8.519909807700076), (3.1690614508664035, 0.8143251472991676), (4.818330576085663, 3.69502320071854), (4.859398750231154, 7.941638208906054), (2.242047519125402, 3.251007486318421), (6.15468228405522, 5.701407205679413), (7.556167822397726, 2.23167959857534)]

*** Objective Function Value: 289.93272604483815

Initial Centroid List: [(2.3605714528864787, 6.007966227729477), (8.875780722296865, 8.960923612139595), (3.0410170219628694, -0.3613848735372547), (7.597313421987085,

Final Centroid List: ((3.8949785598108813, 4.917516281920117), (7.914309977818314, 8.519909807700076), (3.1690614508664035, 0.8143251472991676), (6.182966504540978, 2.083050197023721), (4.986279019961613, 7.797378117091519), (7.941713964744382, 2.87966135379709), (2.297647316142252, 7.394026355610831), (2.242047519125402, 3.251007486318421), (6.15468228405522, 5.701407205679413)]

* Objective Function Value: 281.0199554193879

-- K = 10 --