

## KMean\_Armin

```
1 import random
2 import math
3 import sys
4 import matplotlib.pyplot as plt
5
6 def cluster(centroids, datapoints):
7
8     # for each point find the nearest Center
9     for p in datapoints:
10         minD = distance(p[0], centroids[0][0])
11         p[1] = centroids[0][0]
12         for c in centroids:
13             d = distance(p[0], c[0])
14             if(d < minD):
15                 minD = d
16                 p[1] = c[0]
17     return datapoints
18
19 # Update the centers
20 def calculateCentroid(centroids, datapoints):
21
22     # for each Center find it's points
23     for c in centroids:
24         x=0
25         y=0
26         count = 0
27         for p in points:
28             # if a Point's Label is center C
29             if p[1] == c[0]:
30                 # include it in (x1+x2,.../points + y1+y2,.../points)
31                 x += p[0][0]
32                 y += p[0][1]
33                 count += 1
34         c[0] = [x/count, y/count]
35
36     return centroids
37
38 def distance(p, c):
39     d = 0
40     for i in range(2):
41         d += ((p[i] - c[i]) ** 2)
42     return math.sqrt(d)
43
44 def unshared_copy(inList):
45     if isinstance(inList, list):
46         return list( map(unshared_copy, inList) )
```

## KMean\_Armin

```
47     return inList
48 ##### Input
49
50 # get the DataSet, K, M
51 with open(sys.argv[1]) as f:
52     k, m = [int(x) for x in f.readline().split()]
53
54     # each point = [ [coordinates], [Label] ]
55     points = [list([list([float(x) for x in line.split()]),[]]) for line in
56 f]
57
58 # get an Unshared Copy of Points ,
59 # so it doesn't have any reference to 'points'
60 # because we are picking up centers from points and then we change them
61 pointC = unshared_copy(points)
62
63 centers = []
64 ##### K-Mean
65 # Initialize First Centers Randomly
66 while len(centers) < m:
67     centers.append(random.choice(pointC))
68
69 # 5 Iterations for K-Mean
70 for i in range(5):
71     points = cluster(centers, points)
72     centers = calculateCentroid(centers, points)
73
74 ##### Plot
75 # Clusters
76 for p in points:
77     if p[1][0] == centers[0][0][0]:
78         plt.scatter(p[0][0], p[0][1], s=125, c='red')
79     else:
80         plt.scatter(p[0][0], p[0][1], s=125, c='blue')
81 # Centers
82 for c in centers:
83     plt.scatter(c[0][0], c[0][1], s=125, c='yellow')
84 plt.xlim(-1, 4)
85 plt.ylim(-1, 4)
86 plt.show()
87
88 # Final Centers : [[1.8, 2.867], [1.032, 1.212]]
89 #####
90
```