

 k-means-from-scratch.py ×

```

1 |
2 #Get path of working directory
3 path = %pwd
4
5 #Path of the input file
6 input_file_path = path + '\\' + 'prog2-input-data.txt'
7
8 #Read input file's content and save numbers to file_content as floats
9 file_content = [float(i.rstrip()) for i in open(input_file_path)]
10 # Find each number's closest centroid:
11 def find_distance(centroids, numbers):
12     """
13     DESCRIPTION:
14         For a given set of numbers and centroids, this function
15         assigns each number to the closest centroid.
16     -----
17     Parameters
18     -----
19     centroids : list
20         List of centroids.
21     numbers : list
22         List of numbers.
23
24     Returns
25     -----
26     centroids_numbers : dictionary
27         A dictionary of clusters and numbers closest to those clusters.
28     """
29
30     #Create dictionary that will hold cluster numbers and numbers
31     centroids_numbers = {}
32     #Initialize the dictionary with cluster numbers (k) as keys
33     for centroid in range(len(centroids)):
34         centroids_numbers[centroid] = []
35     #Go through numbers list, for each number, calculate it's distance to
36     #each centroid, find the centroid it is closest to, update centroids_nu
37     #dictionary with this centroid and number.
38     for number in numbers:
39         dist_to_centroid = []
40
41         for centroid in centroids:
42             dist_to_centroid.append(abs(number-centroid))
43

```

Name	Type	Size	Value
f	TextIOWrapper	1	TextIOWrapper object of _io module
file_content	list	10	[1, 8, 4, 5, 1, 1, 2, 1, 9, 8, 7, 6, 11, 32, 3, 2, 0, 5, 6, 5]

Console 1/A x

```
Enter the number of clusters: 5
```

Iteration 1

```
0 [1.8]
1 [4.5, 6.5]
2 [1.1, 0.5]
3 [2.1, 3.2]
4 [9.8, 7.6, 11.32]
```

Iteration 2

```
0 [1.8, 2.1]
1 [4.5, 6.5]
2 [1.1, 0.5]
3 [3.2]
4 [9.8, 7.6, 11.32]
```

Iteration 3

```
0 [1.8, 2.1]
1 [4.5, 6.5]
2 [1.1, 0.5]
3 [3.2]
4 [9.8, 7.6, 11.32]
```

```
Point 1.8 in cluster 0
Point 4.5 in cluster 1
Point 1.1 in cluster 2
Point 2.1 in cluster 0
Point 9.8 in cluster 4
Point 7.6 in cluster 4
Point 11.32 in cluster 4
Point 3.2 in cluster 3
Point 0.5 in cluster 2
Point 6.5 in cluster 1
```

In [2]: