PROBLEM 2 - MERGING CLUSTERS (10 points possible)

In this problem, you will finish implementing the `ClusterSet` class by writing code for the three missing functions: `mergeClusters`, `findClosest`, and `mergeOne`.

- `mergeClusters` will create a new cluster containing the union of the points in c1 and points in c2. This new cluster will be added to the cluster set, while c1 and c2 are removed from the cluster set. This funcion does not return anything.

- `findClosest` will use the "linkage" parameter to find the distance between two clusters. It will iterate over all pairs of clusters in the cluster set and return the tuple (c1,c2) of the clusters within the cluster set that are closest. Note that no matter what linkage criteria we are using, we will always return the cluster pairs that are closest to each other.

- `mergeOne` will make use of `findClosest` to determine which pairs of clusters to merge. Then, it will use `mergeClusters` to perform the merging on these two closest clusters. This function returns the tuple (c1,c2) representing the clusters that were merged.

To test how your code clusters the city data, you may use the `hCluster` function and uncomment the line `#test()` to run the hierarchical clustering algorithm. It may take up to a minute to cluster, so be patient. Notice that the last parameter of `hCluster` is a history flag. If toggled, it will print out more detail, in particular which clusters are merged at each step. During testing, you may also want to make up a new datafile that contains less datapoints, less features, and easier numbers to work with.

Hint: A simpler datafile and sample output

Below is a simpler datafile (test.txt). As with cityTemps.txt, the first line represents the name of a point. The lines after that, up until before #end represent how many features will correspond to each point and they are represented by numbers on the same line as the point name, comma delimited. The line #end represents the end of the column titles and the beginning of the datapoints. You must have an empty line at the end of the file.

```
#point_name
#feature_value1
#feature_value2
#end
a,3,1
b,6,2
c,6,5
d,6,2
e,5,5
f,1,4
g,5,8
```

Appropriate test lines with this data would be:

```
points = buildCityPoints('test.txt', False)
hCluster(points, Cluster.singleLinkageDist, 3, False)
hCluster(points, Cluster.maxLinkageDist, 3, False)
hCluster(points, Cluster.averageLinkageDist, 3, False)
```

And with a correctly implemented set of functions, one possible output is shown below. For such a small number of data points and features it is possible that your output would be slightly different, depending on which point gets chosen in case of a tie.

```
Final set of clusters:
  C0:a
  C1:b, c, d, e, g
  C2:f

Final set of clusters:
  C0:a, b, d
  C1:c, e, g
  C2:f

Final set of clusters:
  C0:a, f
  C1:b, c, d, e
  C2:g
```

Enter all code for the `ClusterSet` class below, including the functions in this class that were already defined for you. Do not paste the `Cluster` class code.

```
50            return len(self.members)
51    def toStr(self):
52            cNames = []
53            for c in self.members:
54                cNames.append(c.getNames())
55            cNames.sort()
56            result = ''
57            for i in range(len(cNames)):
58                names = ''
59                for n in cNames[i]:
60                    names += n + ', '
61                names = names[:-2]
62                result += '  C' + str(i) + ':' + names + '\n'
63            return result
```

Incorrect

Test results

	See full output
INCORRECT	
	See full output

You have used 1 of 30 submissions

Show Discussion

edX

EdX offers interactive online classes and MOOCs from the world's best universities. Online courses from MITx, HarvardX, BerkeleyX, UTx and many other universities. Topics include biology, business, chemistry, computer science, economics, finance, electronics, engineering, food and nutrition, history, humanities, law, literature, math, medicine, music, philosophy, physics, science, statistics and more. EdX is a non-profit online initiative created by founding partners Harvard and MIT.

Terms of Service and Honor Code

Privacy Policy (Revised 4/16/2014)

About edX

About

News

Contact

FAQ

edX Blog

Donate to edX

Jobs at edX

Follow Us

Twitter

Facebook

Meetup

LinkedIn

Google+