

Name: Naveenraj Palanisamy

NetId: NXP154130

Machine Learning Assignment (CS 6375.001)-(EM)

(Note: Runs in python 3.5)

(Note: File Path should be full path)

(Note: This is done as optional home work)

EM:

Folder Structure:

This PC > OS (C:) > 3 > Assignment3 > EM >				
<input type="checkbox"/> Name	Date modified	Type	Size	
__pycache__	4/16/2016 10:50 PM	File folder		
.project	4/7/2016 8:26 PM	PROJECT File	1 KB	
.pydevproject	4/7/2016 8:26 PM	PYDEVPROJECT File	1 KB	
audio_file.dat	4/8/2016 6:49 PM	DAT File	66 KB	
beep.wav	4/8/2016 5:07 PM	WAV File	23 KB	
EM_algorithm.py	4/9/2016 8:33 PM	Python File	6 KB	
em_data.txt	4/9/2016 8:31 PM	Text Document	82 KB	
em_data_temp.txt	4/7/2016 8:39 PM	Text Document	82 KB	
FileParsing.py	4/7/2016 9:03 PM	Python File	1 KB	
Main_class.py	4/16/2016 10:52 PM	Python File	1 KB	
outfile.dat.npy	4/8/2016 6:49 PM	NPY File	1 KB	
sample.mp3	9/24/2015 10:51 A...	MP3 File	3,746 KB	

Main_Class.py is the file need to run:

Running the program steps: ->

- 1) Go to 'command prompt'
- 2) Go to folder where python is installed.
- 3) Given command as (python.exe 'full path to-> Main_class.py' Number_of_Iterations varience_value K_value em_Dataalcation)

Sample Run:

C:\Users\NAVE\AppData\Local\Programs\Python\Python35-32\python.exe
C:\3\Assignment3\EM\Main_class.py 5 0.5 4 C:\3\Assignment3\EM\em_data.txt

```
Go to Command Prompt
Go to [25.705, 5.8859, 15.6422, 6.9153]
Given [0.6749, 0.9238, 0.6637, 7.8733]
mber -----
-----
[0.3264, 0.3188, 0.3175, 0.0372]
[25.5161, 5.5498, 15.5004, 12.3982]
[0.8838, 0.9049, 0.9087, 7.753]
-----
-----
[0.3285, 0.3244, 0.3217, 0.0254]
[25.4966, 5.5088, 15.4638, 13.5277]
[0.9381, 0.9502, 0.9581, 12.1225]
-----
-----
[0.3291, 0.3288, 0.3278, 0.0143]
[25.4894, 5.5093, 15.4554, 15.0722]
[0.9508, 0.9516, 0.9546, 21.8478]
-----
-----
[0.3313, 0.3316, 0.3313, 0.0058]
[25.4865, 5.51, 15.4527, 15.8182]
[0.9731, 0.9722, 0.9731, 55.6572]
-----
C:\Users\NAVE>
```

Algorithm is tested for different initialization values and algorithm works mainly based on the initialization parameter.

Different initialization values are passed as comment line argument.

Initially took values as below:

K: 2

Iterations: 4

Variance: 0.5

Take variance as 1 and run for 3 clusters. Means values are there in image.

```
mean [16.6813804426, 4.23545335138, 25.8852836479]
-----
[0.3327, 0.3332, 0.3333]
[15.456, 5.5077, 25.4867]
[0.9458, 0.9444, 0.9439]
-----
ially
-----
[0.3333, 0.3333, 0.3333]
[15.4492, 5.5093, 25.4867]
[0.9671, 0.9671, 0.9671]
-----
iance
-----
[0.3333, 0.3333, 0.3333]
[15.4492, 5.5093, 25.4867]
[0.9671, 0.9671, 0.9671]
```

According to me setting variance to 1 seems to be best, because it will reduce the change of cluster without any assignment in the initial iterations itself.