

«به نام خدا»

تکلیف پنجم – سوال سوم – مرضیه علیدادی – 9631983
(کد های مربوط، در دو فرمت py و .ipynb. ضمیمه شده اند.)

3.

(a) دیتاست iris:

	0	1	2	3
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
...
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

150 rows × 4 columns

داده های آن را نرمال سازی کردم:

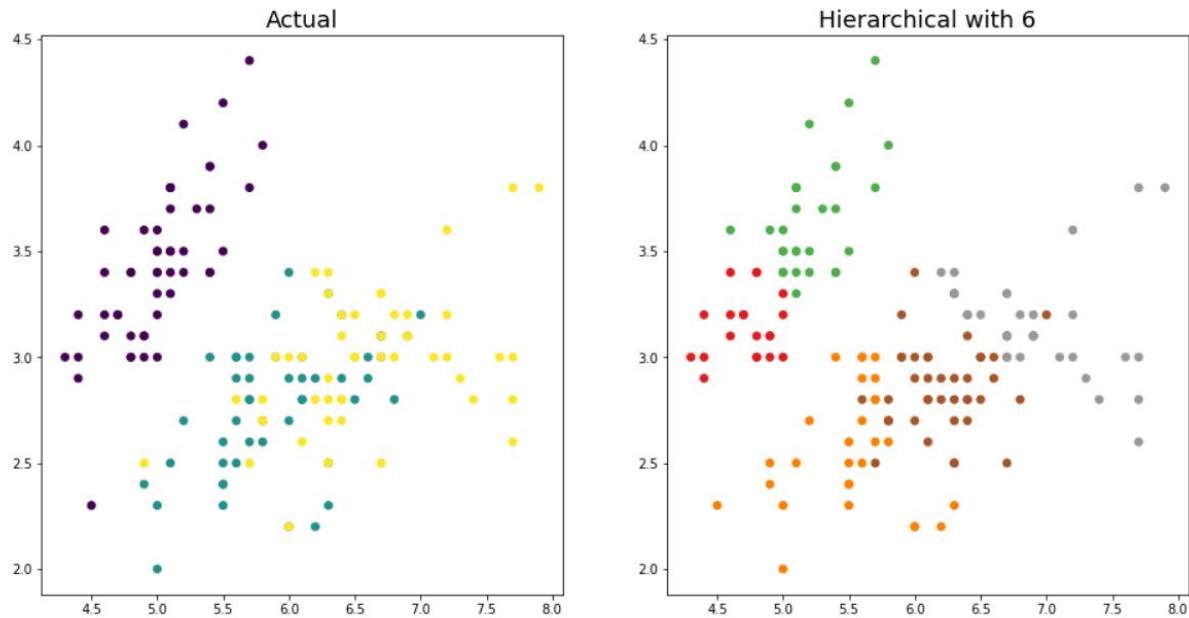
```
array([[ 6.17956139,  8.05688682,  0.79572398,  0.26326437],
       [ 5.93722565,  6.90590299,  0.79572398,  0.26326437],
       [ 5.69488991,  7.36629652,  0.73888655,  0.26326437],
       [ 5.57372204,  7.13609975,  0.8525614 ,  0.26326437],
       [ 6.05839352,  8.28708359,  0.79572398,  0.26326437],
       [ 6.543065 ,  8.97767388,  0.96623626,  0.52652874],
       [ 5.57372204,  7.82669005,  0.79572398,  0.39489655],
       [ 6.05839352,  7.82669005,  0.8525614 ,  0.26326437],
       [ 5.3313863 ,  6.67570622,  0.79572398,  0.26326437],
       [ 5.93722565,  7.13609975,  0.8525614 ,  0.13163218],
       [ 6.543065 ,  8.51728035,  0.8525614 ,  0.26326437],
       [ 5.81605778,  7.82669005,  0.90939883,  0.26326437],
       [ 5.81605778,  6.90590299,  0.79572398,  0.13163218],
       [ 5.21021843,  6.90590299,  0.6252117 ,  0.13163218],
       [ 7.02773649,  9.20787065,  0.68204912,  0.26326437],
       [ 6.90656862, 10.12865772,  0.8525614 ,  0.52652874],
       [ 6.543065 ,  8.97767388,  0.73888655,  0.52652874],
       [ 6.17956139,  8.05688682,  0.79572398,  0.39489655],
       [ 6.90656862,  8.74747712,  0.96623626,  0.39489655],
       [ 6.17956139,  8.74747712,  0.8525614 ,  0.39489655],
```


(c)

```

4      45
3      30
2      29
5      26
1      20
Name: cluster_labels_6, dtype: int64

```



(d) در قسمت قبل، تعداد برچسب ها برابر با 5 تا بود. در این قسمت، تعداد برچسب ها (با توجه به اینکه تعداد کلاستر ها افزایش پیدا کرده) افزایش یافته است و برابر با 11 تاست. همانطور که در نمودار ها مشخص است، با افزایش تعداد کلاستر ها و برچسب ها، از دسته بندی طبیعی داده ها دور شده ایم. در نتیجه، حالت قبلی، دسته بندی بهتری را برای ما حاصل می شد.

```

: 8      32
3   22
1   20
6   17
11  15
7   13
9    8
5    8
2    7
4    5
10   3
Name: cluster_labels_3, dtype: int64

```

