# «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,
                                                   0.263264371,
       [ 6.543065
                       8.97767388,
                                     0.96623626,
                                                   0.52652874],
       [ 5.57372204,
                       7.82669005,
                                     0.79572398,
                                                   0.394896551,
         6.05839352,
                       7.82669005,
                                     0.8525614 ,
                                                   0.26326437],
         5.3313863 ,
                       6.67570622,
                                     0.79572398,
                                                   0.26326437],
                                     0.8525614 ,
       [ 5.93722565,
                       7.13609975,
                                                   0.13163218],
       [ 6.543065
                       8.51728035,
                                     0.8525614 ,
                                                   0.263264371,
       [ 5.81605778,
                       7.82669005,
                                     0.90939883,
                                                   0.26326437],
         5.81605778,
                       6.90590299,
                                     0.79572398,
                                                   0.13163218],
         5.21021843,
                       6.90590299,
                                     0.6252117 ,
                                                   0.131632181,
         7.02773649,
                       9.20787065,
                                     0.68204912,
                                                   0.26326437],
       [ 6.90656862, 10.12865772,
                                     0.8525614 ,
                                                   0.52652874],
                                     0.73888655,
                                                   0.52652874],
       [ 6.543065 ,
                       8.97767388,
       [ 6.17956139,
                       8.05688682,
                                     0.79572398,
                                                   0.39489655],
       [ 6.90656862,
                       8.74747712,
                                     0.96623626,
                                                   0.39489655],
                                                   0.304996551
                                     0.9525614
```

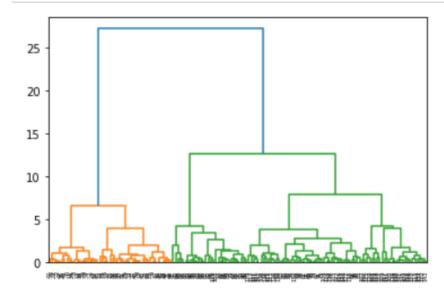
### جزئيات آن:

	0	1	2	3
count	150.000000	150.000000	150.000000	150.000000
mean	7.080243	7.037882	2.135951	1.578709
std	1.003350	1.003350	1.003350	1.003350
min	5.210218	4.603935	0.568374	0.131632
25%	6.179561	6.445509	0.909399	0.394897
50%	7.027736	6.905903	2.472428	1.711218
75%	7.754744	7.596493	2.898709	2.369379
max	9.572262	10.128658	3.921782	3.290805

### ماتریس مجاورت با استفاده از linkage:

```
array([[1.01000000e+02, 1.42000000e+02, 0.00000000e+00, 2.00000000e+00],
       [7.00000000e+00, 3.90000000e+01, 1.21167870e-01, 2.00000000e+00],
       [1.00000000e+01, 4.80000000e+01, 1.21167870e-01, 2.00000000e+00],
       [0.00000000e+00, 1.7000000e+01, 1.31632184e-01, 2.00000000e+00],
       [9.00000000e+00, 3.40000000e+01, 1.31632184e-01, 2.00000000e+00],
       [1.28000000e+02, 1.32000000e+02, 1.31632184e-01, 2.00000000e+00],
       [1.27000000e+02, 1.38000000e+02, 1.33836265e-01, 2.00000000e+00],
       [2.00000000e+00, 4.70000000e+01, 1.33836265e-01, 2.00000000e+00],
       [1.90000000e+01, 4.60000000e+01, 1.43378956e-01, 2.00000000e+00],
       [8.00000000e+01, 8.10000000e+01, 1.43378956e-01, 2.00000000e+00],
       [1.20000000e+02, 1.43000000e+02, 1.66143388e-01, 2.00000000e+00],
       [1.00000000e+00, 2.50000000e+01, 1.66143388e-01, 2.00000000e+00],
       [1.10000000e+01, 2.40000000e+01, 1.70512281e-01, 2.00000000e+00],
       [4.00000000e+01, 1.53000000e+02, 1.72216546e-01, 3.00000000e+00],
       [3.00000000e+01, 1.54000000e+02, 1.72216546e-01, 3.00000000e+00],
       [2.90000000e+01, 1.57000000e+02, 1.78366645e-01, 3.00000000e+00],
       [4.00000000e+00, 3.70000000e+01, 1.78909711e-01, 2.00000000e+00],
       [8.80000000e+01, 9.50000000e+01, 1.87721011e-01, 2.00000000e+00],
       [1.36000000e+02, 1.48000000e+02, 2.11968529e-01, 2.00000000e+00],
       [6 30000000a±01 7 90000000a±01 2 11969520a_01 2 00000000a±001
```

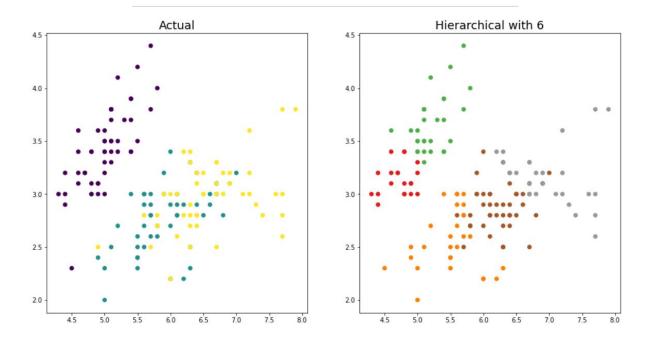
## dendrogram به این صورت خواهد شد:



از نمودار برداشت می شود که بهترین تعداد برای دسته بندی داده های این دیتاست، 3 تا است. و همچنین مشخص است که 2 تا از این 3 دسته، که سبز رنگ هستند، به هم نزدیک ترند.

```
4 45
3 30
2 29
5 26
1 20
```

Name: cluster\_labels\_6, dtype: int64

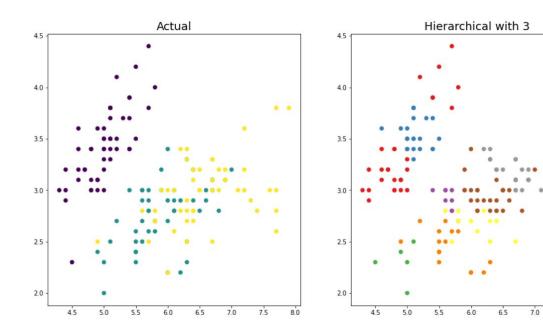


در قسمت قبل، تعداد برچسب ها برابر با 5 تا بود. در این قسمت، تعداد برچسب ها (با توجه به اینکه تعداد کلاستر ها افزایش پیدا کرده) افزایش یافته است و برابر با 11 تاست.

همانطور که در نمودار ها مشخص است، با افزایش تعداد کلاستر ها و برچسب ها، از دسته بندی طبیعی دادهها دور شده ایم.

در نتیجه، حالت قبلی، دسته بندی بهتری را برای ما حاصل می شد.

Name: cluster\_labels\_3, dtype: int64



7.5

8.0