

به نام خدا



دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

عنوان: تکلیف چهارم

افروز شیخ الاسلامی

9729393

بهار 1401

مدرس: دکتر ناصر قدیری

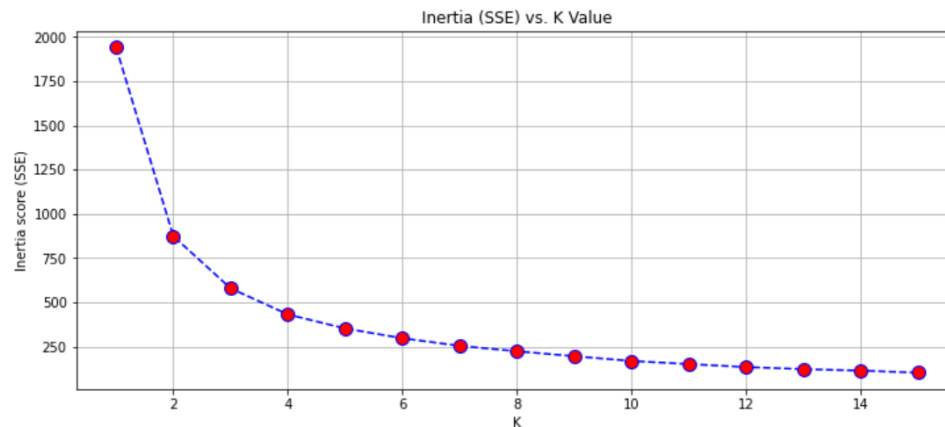
## فهرست مطالب

3	..... سوال اول
3	..... قسمت f
4	..... قسمت g
4	..... قسمت h
5	..... سوال دوم
5	..... قسمت a
5	..... قسمت b
6	..... قسمت c
6	..... قسمت d
7	..... قسمت e
7	..... سوال سوم

## سوال اول

### قسمت f

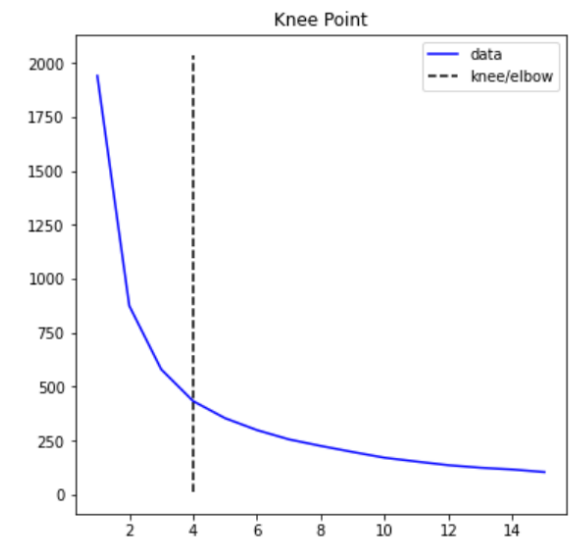
نمودار به دست آمده در قسمت e:



1 Figure

حدس من از  $k$  مناسب در این قسمت،  $k=7$  بود.

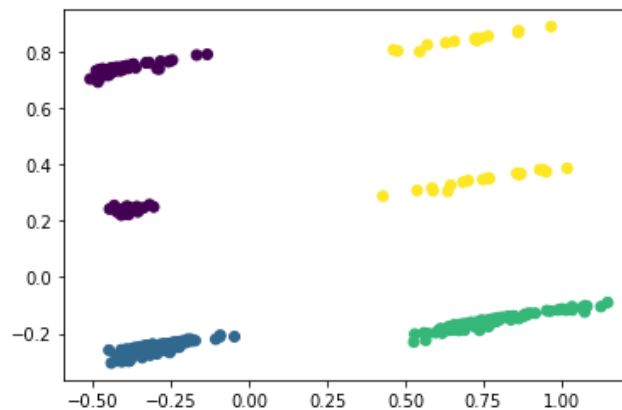
نمودار به دست آمده با استفاده از KneLocator :



2 Figure

واضح است که نقطه elbow که برابر  $k$  بهینه است، برابر 4 شده است. بنابراین برداشت من از نمودار قسمت e، اشتباه بوده است.

## قسمت g



شکل 3، نشان دهنده scatter plot نتایج به دست آمده از خوشه بندی ها است. 4 خوشه مورد نظر در این شکل مشخص است.

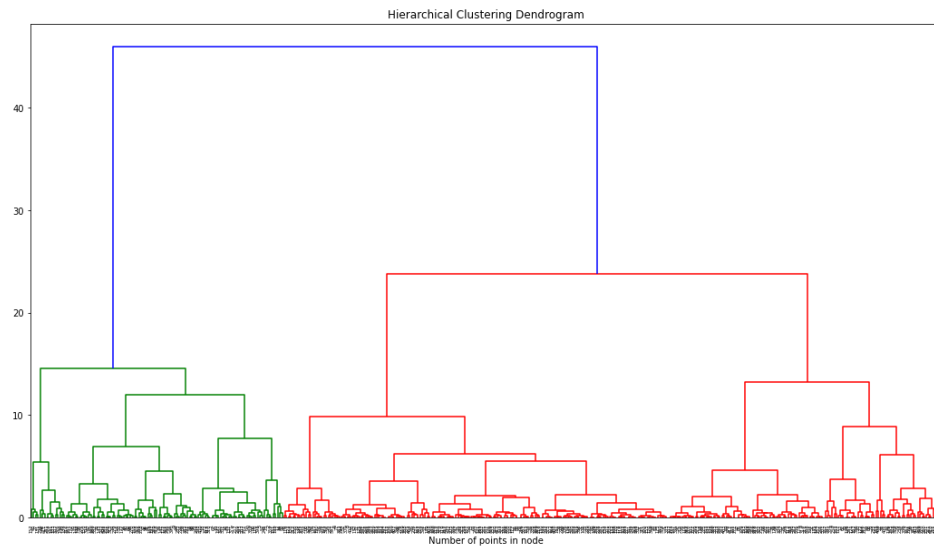
## قسمت h

	labels	0	1	2	3
pc1	count	83.000000	201.000000	97.000000	33.000000
	mean	-0.399504	-0.327134	0.771287	0.730236
	std	0.069608	0.063967	0.138220	0.156289
	min	-0.508485	-0.448453	0.527158	0.428090
	25%	-0.446993	-0.367193	0.672705	0.629182
	50%	-0.409372	-0.342961	0.742363	0.729482
	75%	-0.378911	-0.299626	0.840734	0.860878
	max	-0.133943	-0.046796	1.147341	1.017859
pc2	count	83.000000	201.000000	97.000000	33.000000
	mean	0.577831	-0.256017	-0.158487	0.571897
	std	0.236530	0.016350	0.027458	0.254025
	min	0.220682	-0.305004	-0.230151	0.287960
	25%	0.247857	-0.264949	-0.176947	0.347317
	50%	0.731972	-0.255558	-0.160252	0.382102
	75%	0.741903	-0.247273	-0.142837	0.841151
	max	0.791866	-0.206324	-0.090007	0.890406

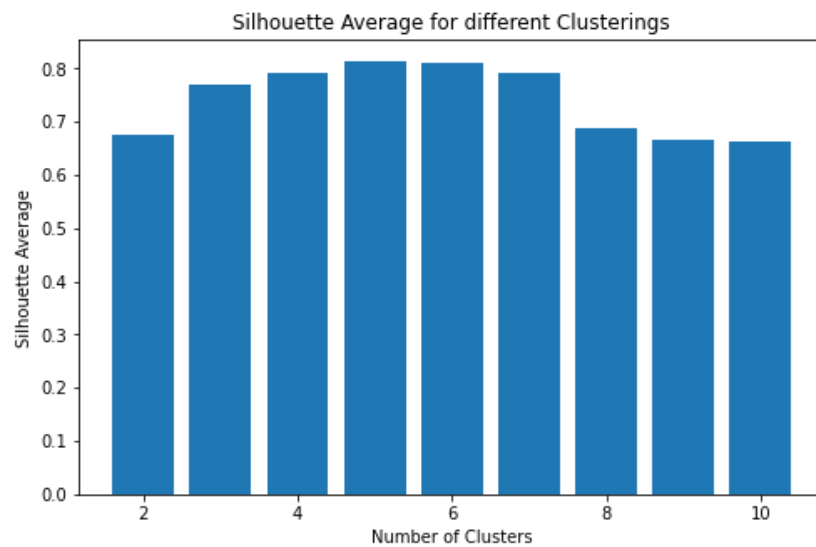
اطلاعات آماری هر یک از 4 خوشه، برحسب هر کدام از principal component ها در شکل بالا نمایش داده شده است. برای مثال از 414 داده، 83 داده در خوشه اول، 201 داده در خوشه دوم، 97 داده در خوشه سوم و 33 داده در خوشه چهارم قرار گرفته اند.

## سوال دوم

قسمت a

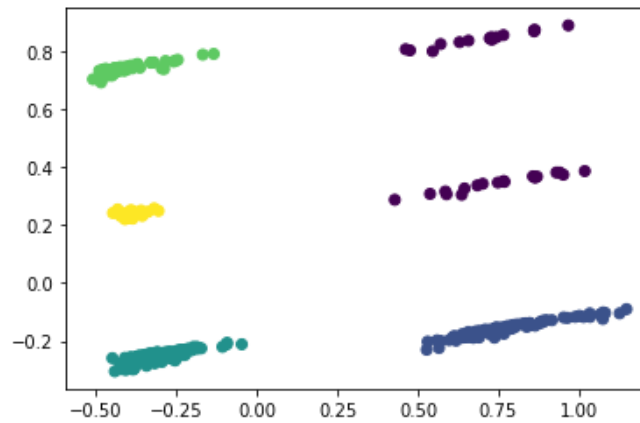


قسمت b



با توجه به نمودار بالا، بهترین  $k$  برابر 5 است، زیرا بیشترین مقدار silhouette را دارد.

## قسمت C



## قسمت d

	labels	0	1	2	3	4
pc1	count	33.000000	97.000000	201.000000	56.000000	27.000000
	mean	0.730236	0.771287	-0.327134	-0.404418	-0.389311
	std	0.156289	0.138220	0.063967	0.081552	0.032444
	min	0.428090	0.527158	-0.448453	-0.508485	-0.446596
	25%	0.629182	0.672705	-0.367193	-0.463910	-0.406107
	50%	0.729482	0.742363	-0.342961	-0.435958	-0.391605
	75%	0.860878	0.840734	-0.299626	-0.374994	-0.381524
	max	1.017859	1.147341	-0.046796	-0.133943	-0.305299
pc2	count	33.000000	97.000000	201.000000	56.000000	27.000000
	mean	0.571897	-0.158487	-0.256017	0.740705	0.240017
	std	0.254025	0.027458	0.016350	0.018241	0.009870
	min	0.287960	-0.230151	-0.305004	0.693553	0.220682
	25%	0.347317	-0.176947	-0.264949	0.731470	0.231973
	50%	0.382102	-0.160252	-0.255558	0.739614	0.241485
	75%	0.841151	-0.142837	-0.247273	0.747033	0.247278
	max	0.890406	-0.090007	-0.206324	0.791866	0.257945

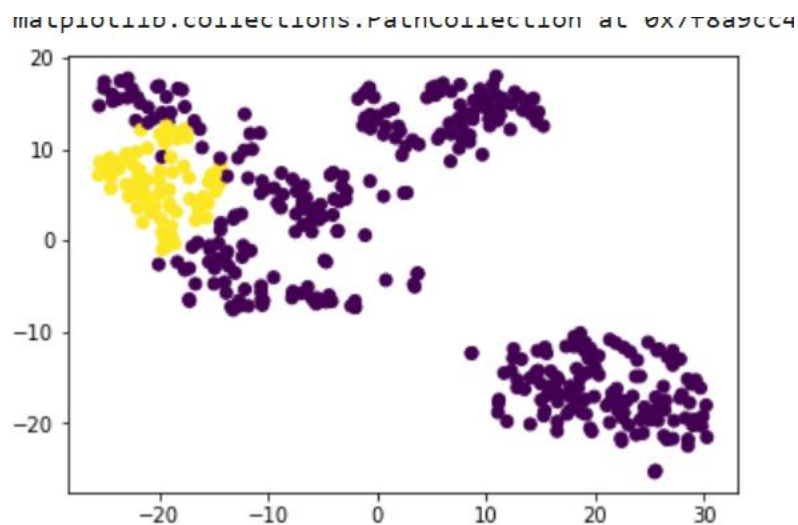
در نمودار بالا، اطلاعات آماری 5 خوشه تشکیل شده بر حسب هر کدام از principal component ها، قابل مشاهده است.

## قسمت e

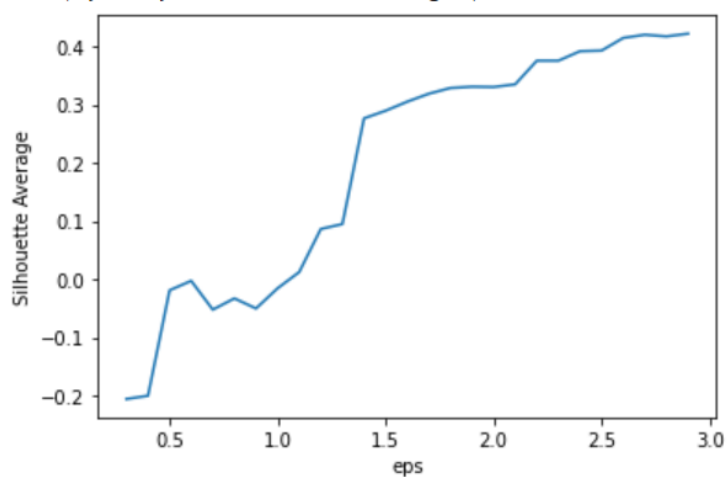
روش ward ، خوشه بندی به روش Agglomerative است که روش hierarchical بوده و داده ها را از پایین به بالا خوشه بندی می کند. Kmeans نیز روش clustering دیگری است که اصولا داده ها را به شکل کروی خوشه بندی می کند. ولی از آنجا که داده های ما در این سوال شکل کروی نداشتند، نمی توان به راحتی تفاوت روش ward و kmeans را متوجه شد.

## سوال سوم

نتیجه خوشه بندی اولیه:



تاثیر eps بر معیار silhouette :



بهترین مقدار  $\epsilon$  حدود 1.3 یا 1.4 در نظر گرفته می شود.

رسم نمودار scatter plot با  $\epsilon = 1.3$  :

