

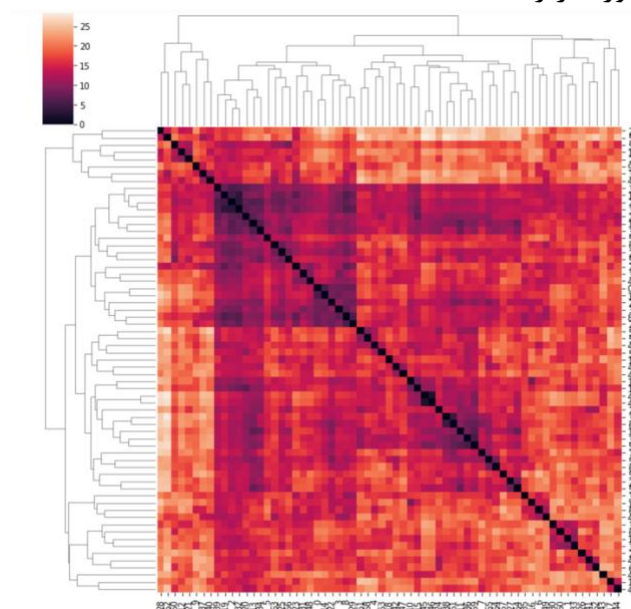
تمرین کامپیوتری دوم

(ثمین مهدی زاده ۵۲۶۰۱۰۱۸۱)

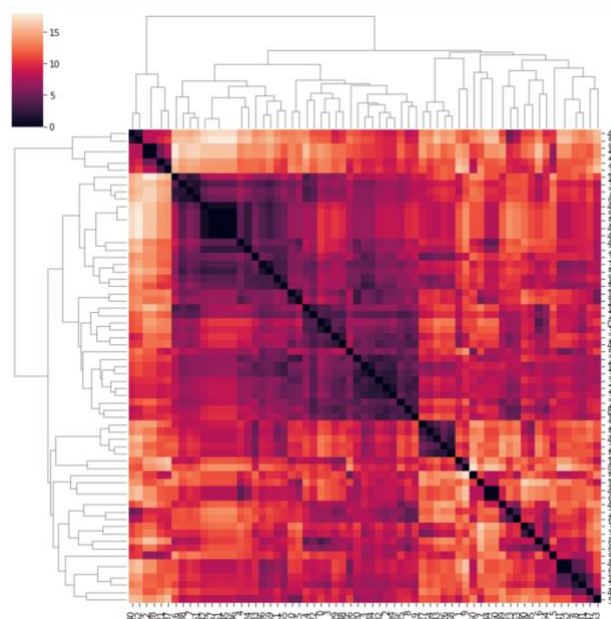
RSA.۷

در این قسمت در ابتدا به ازای مقادیر pse محاسبه شده برای vf,loc,ecc متناظر با هر فرد، یک ماتریس که بیانگر آن فرد است به وجود آورده و به کمک فاصله ی اقلیدسی تفاوت هر فرد با فرد دیگر محاسبه شد و کلاستر ها به دست آمد.

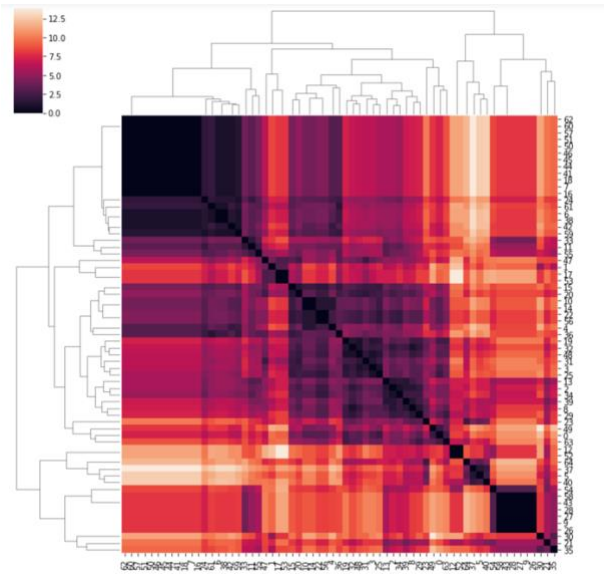
ماتریس های به وجود آمده به صورت زیر است:



شکل ۱. کلاسترهای به وجود آمده براساس loc



شکل ۲. کلاسترهای به وجود آمده براساس vf



شکل ۳. کلاستر های به وجود آمده بر اساس ecc

همان طور که مشاهده می شود کلاستر های به وجود آمده هنگامی که از ecc برای بازنمایی هر فرد استفاده می شود متمایز ترند به همین منظور برای ادامه کار از کلاستر های به وجود آمده توسط این پارامتر استفاده شده است. در صورتی که خوشه بندی را تا ۵ مرحله جلو ببریم خوشه های به وجود آمده به صورت زیر خواهند بود:

cluster1 = [62,60,57,51,50,46,45,44,41,18,7,16,24,61,6,38,42,59,33,11,55]

cluster2 = [47,1,17,53]

cluster3 = [15,20,10,14,22,56,4,36,19,32,48,31,325,13,2,34,39,8,29,23,49,0,63]

cluster4 = [12,52,64,37,5,40]

cluster5 = [54,58,43,28,27,9,26,30,21,35]

از آن جایی که تنها متغیر های متفاوت بین افراد دست برتر، چشم و جنسیت بودند برای هر فرد اطلاعات این سه مورد به همراه کلاستری که در آن قرار داشت ذخیره شدند و با انجام تست های آماری بر روی هر کدام از این متغیر های بین گروهی مشخص شد که هر سه می توانند در جدایی کلاستر ها نقش داشته باشند (یعنی حداقل دو کلاستر وجود دارند که تفاوت معناداری میان میانگین متغیر های آن وجود دارد).

نتایج تست های آماری بر روی هر یک از این سه متغیر به صورت زیر است که مشاهده می شود هر سه این متغیر ها از لحاظ آماری معنادار هستند. برای انجام تمام این تست ها از تست ANOVA استفاده شده است.

----- cluster ~ C(hndns) -----

OLS Regression Results

Dep. Variable:	cluster	R-squared:	0.025
Model:	OLS	Adj. R-squared:	0.010
Method:	Least Squares	F-statistic:	1.648
Date:	Fri, 08 Apr 2022	Prob (F-statistic):	0.204
Time:	00:20:48	Log-Likelihood:	-115.87
No. Observations:	65	AIC:	235.7
Df Residuals:	63	BIC:	240.1
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	3.3750	0.517	6.532	0.000	2.343	4.407
C(hndns) [T.r]	-0.7083	0.552	-1.284	0.204	-1.811	0.394

Omnibus:	14.518	Durbin-Watson:	1.786
Prob(Omnibus):	0.001	Jarque-Bera (JB):	3.753
Skew:	0.124	Prob(JB):	0.153
Kurtosis:	1.849	Cond. No.	5.53

شکل ۴. تست آماری *hndns* بر روی خوشه بندی

----- cluster ~ C(eye) -----

OLS Regression Results

Dep. Variable:	cluster	R-squared:	0.001
Model:	OLS	Adj. R-squared:	-0.015
Method:	Least Squares	F-statistic:	0.03951
Date:	Fri, 08 Apr 2022	Prob (F-statistic):	0.843
Time:	00:21:41	Log-Likelihood:	-116.69
No. Observations:	65	AIC:	237.4
Df Residuals:	63	BIC:	241.7
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	2.8000	0.296	9.460	0.000	2.209	3.391
C(eye) [T.r]	-0.0750	0.377	-0.199	0.843	-0.829	0.679

Omnibus:	19.852	Durbin-Watson:	1.842
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4.353
Skew:	0.163	Prob(JB):	0.113
Kurtosis:	1.775	Cond. No.	2.99

شکل ۵. تست آماری *eye* بر روی خوشه بندی

----- cluster ~ C(sex) -----

OLS Regression Results

Dep. Variable:	cluster	R-squared:	0.002
Model:	OLS	Adj. R-squared:	-0.013
Method:	Least Squares	F-statistic:	0.1514
Date:	Fri, 08 Apr 2022	Prob (F-statistic):	0.699
Time:	00:22:16	Log-Likelihood:	-116.64
No. Observations:	65	AIC:	237.3
Df Residuals:	63	BIC:	241.6
Df Model:	1		
Covariance Type:	nonrobust		

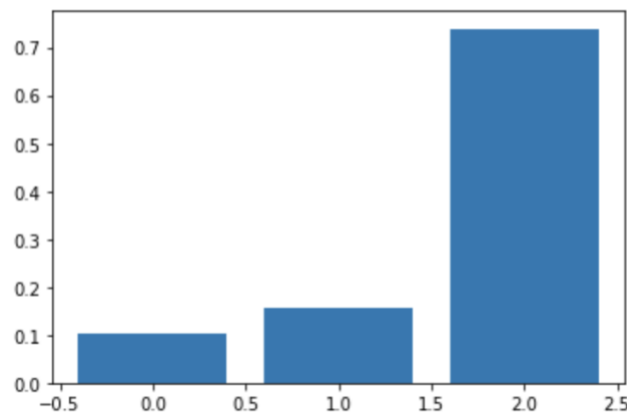
	coef	std err	t	P> t	[0.025	0.975]
Intercept	2.8571	0.323	8.855	0.000	2.212	3.502
C(sex) [T.m]	-0.1526	0.392	-0.389	0.699	-0.936	0.631

Omnibus:	18.967	Durbin-Watson:	1.864
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4.314
Skew:	0.173	Prob(JB):	0.116
Kurtosis:	1.786	Cond. No.	3.28

شکل ۶. تست آماری SEX بر روی خوشه بندی

برای مشخص شدن اهمیت هر یک از این متغیرها در خوشه بندی، بر روی داده (که شامل شناسه فرد، سه ویژگی گفته شده و شماره خوشه است) یک مدل را به کمک درخت تصمیم آموزش داده و برای آن feature_importances محاسبه شده است. نتایج به دست آمده به شرح زیر است که نشان می دهد مهم ترین ویژگی برای خوشه بندی افراد چشم برتر بوده است.

Feature 0: hndns, Score: 0.10345
 Feature 1: sex, Score: 0.15770
 Feature 2: eye, Score: 0.73885



شکل ۷. اهمیت ویژگی ها در خوشه بندی

- کدهای این قسمت در بخش آخر (representational similarity matrix) نوت بوک موجود است.