

الله الرحمن الرحيم



دانشگاه صنعتی اصفهان

دانشکده برق و کامپیوتر

خوشه بندی گراف های مغزی افراد ناسالم (اوتیسم)

گزارش پروژه کارشناسی مهندسی کامپیوتر

فاطمه نادی راوندی 9636753

اساتید راهنما

دکتر زینب مالکی ، دکتر فرزانه شایق

مهر ماه ۱۴۰۰

فهرست مطالب

1	فصل اول	7
1.1	مقدمه	7
1.2	تحقیقات مرتبط	7
2	فصل دوم	8
	مفاهیم و اطلاعات اولیه	8
2.1	مغز	8
2.2	بیماری اوتیسم	8
2.3	ماتریس اتصال مغزی – کانکتوم‌ها	9
2.4	مدل کردن مغز به عنوان یک شبکه پیچیده	10
2.5	ویژگی‌های گراف	12
2.5.1	ویژگی‌های اولیه	12
2.5.2	معیارهای تفکیک عملکردی	13
2.5.3	معیارهای تجمیع عملکردی	14
2.5.4	معیارهای مرکزیت	15
2.6	آشنایی با خوشه‌بندی	17
2.6.1	خوشه‌بندی سلسله مراتبی	17
3	فصل سوم	18
	پیش نیاز ها	18
3.1	تحلیل واریانس	18
3.2	آزمون‌های تعقیبی	20
3.3	نحوه‌ی ارزیابی خوشه‌بندی ها	20
3.4	نحوه‌ی انتخاب ویژگی ها	21
3.5	نحوه‌ی استفاده از آزمون <i>POST-HOC</i> و جداول آن	21
3.6	چگونگی مقایسه نمودار های جعبه ای	22
4	فصل چهارم	22
	نتایج و پیاده‌سازی	22
4.1	توضیحات کلی پیرامون روند انجام پروژه	22
4.2	بررسی ویژگی ها با استفاده از تحلیل واریانس	23
4.2.1	خوشه بندی کانکتوم‌ها با استفاده از تمام ویژگی ها – ۴۴ ویژگی	24

29.....	خوشه بندی کانکتوم‌ها با استفاده از ویژگی‌های منتخب - ۲۵ ویژگی	4.2.2
30.....	خوشه بندی کانکتوم‌ها با استفاده از ویژگی‌های برگزیده - ۱۵ ویژگی	4.2.3
31.....	مقایسه روند خوشه بندی‌ها در سه حالت مختلف.....	4.2.4
33.....	تحلیل روند خوشه بندی براساس ویژگی‌ها.....	4.2.5
34.....	بررسی ترشه‌لدهای مختلف.....	4.2.6

5 فصل پنجم..... 35

تحلیل و آنالیز..... 35

35.....	بررسی خوشه بندی‌های حاصل از دسته بندی داده‌ها بدون دانستن برچسب‌های واقعی.....	5.1
36.....	بررسی گروه‌هایی شاخص.....	5.1.1
38.....	تحلیل‌های کیفی.....	5.2
38.....	ماکسیمم ضریب خوشگی.....	5.2.1
39.....	میانۀ ضریب خوشگی.....	5.2.2
39.....	ماکسیمم مثلث‌ها.....	5.2.3
40.....	میانۀ مثلث‌ها.....	5.2.4
41.....	میانۀ مرکزیت بینابینی.....	5.2.5
42.....	ماکسیمم مرکزیت بر مبنای نزدیکی.....	5.2.6
43.....	میانۀ مرکزیت بر مبنای نزدیکی.....	5.2.7
44.....	ماکسیمم مرکزیت بر مبنای درجه.....	5.2.8
45.....	میانۀ مرکزیت بر مبنای درجه.....	5.2.9
45.....	نتیجۀ گیری.....	5.2.10
47.....	تحلیل‌های کیفی نهایی.....	5.2.11
47.....	مقایسه و تحلیل دو روش خوشه‌بندی متفاوت.....	5.3

6 فصل ششم..... 48

جمع‌بندی و نتیجه‌گیری..... 48

7 مراجع..... 48

8 ضمیمه‌ها..... 49

49.....	راهنمای استفاده از جداول.....	8.1
49.....	ترشه‌لد 0.4.....	8.2
51.....	تحلیل دسته بندی‌ها با تمام ویژگی‌ها - ۴۴ ویژگی.....	8.2.1
53.....	تحلیل دسته بندی‌ها با استفاده از ویژگی‌های منتخب - ۲۵ ویژگی.....	8.2.2
54.....	تحلیل دسته بندی‌ها با استفاده از ویژگی‌های برگزیده - ۱۵ ویژگی.....	8.2.3
55.....	ترشه‌لد 0.35.....	8.3
59.....	خوشه بندی کانکتوم‌ها با استفاده از ویژگی‌های منتخب - ۲۵ ویژگی.....	8.3.1
62.....	خوشه بندی کانکتوم‌ها با استفاده از ویژگی‌های برگزیده - ۱۹ ویژگی.....	8.3.2

65	0.3 ترشه‌لد	8.4
69.....	خوشه بندی کانکتوم‌ها با استفاده از ویژگی‌های منتخب - ۲۷ ویژگی	8.4.1
71.....	خوشه بندی کانکتوم‌ها با استفاده از ویژگی‌های برگزیده - ۲۲ ویژگی	8.4.2
74	تحلیل‌های کمی	8.5
74.....	بررسی ویژگی‌های مختلف براساس تعداد مثلث‌ها	8.5.1

فصل اول 1

1.1 مقدمه

اختلال طیف اوتیسم (ASD) یک ناتوانی رشدی¹ است که می‌تواند چالش‌های اجتماعی، ارتباطی و رفتاری مهمی ایجاد کند. اغلب، تفاوت ظاهری در افراد مبتلا به اوتیسم وجود ندارد که آن‌ها را از سایر افراد متمایز کند اما معمولاً افراد مبتلا به اوتیسم به شیوه‌ای متفاوت از سایر افراد ارتباط برقرار می‌کنند، تعامل دارند، رفتار می‌کنند و یاد می‌گیرند. با توجه به همین موضوع توجه ما به سمت ساختارها و شبکه‌های مغزی این افراد جلب شد که چگونه این اختلال در سطح شبکه مغز منعکس می‌شود.

پژوهش‌های قبلی صورت گرفته بر روی گراف‌های مغزی با استفاده از الگوریتم‌های کلاسیفیکیشن با هدف طبقه‌بندی آن‌ها بر روی افراد سالم و اوتیسم از دقت خوبی برخوردار نبوده؛ حال ما در این پژوهش به دنبال این قضیه هستیم که با انجام کلاسترینگ و بدون دانستن اطلاعاتی از قبیل سن و جنسیت افراد و صرفاً با بررسی بر روی خود مغز در نتیجه‌ی محدود کردن حیطه بررسی، ببینیم علت پایین بودن دقت چیست، آیا می‌توان با توجه به گسترده بودن طیف اختلالات این بیماری به دسته‌های متفاوتی از این بیماری دست یافت؟! ویژگی یا ویژگی‌های موثر در دسته‌بندی گراف‌های مغزی این افراد چیست؟

1.2 تحقیقات مرتبط

مقالات زیادی به بررسی مغز با استفاده از معیارهای شبکه‌های پیچیده سعی به بهبود در تشخیص اختلالات اوتیسم داشته‌اند. در چند سال اخیر سعی به بررسی این بیماری با استفاده از روش‌های یادگیری عمیق نموده‌اند، این روش‌ها ماتریس اتصال را به صورت یک تصویر در نظر می‌گیرند. ماهیت این ماتریس یک تصویر نیست بلکه گرافی است که دارای معیارهای شبکه‌های پیچیده است.

در پژوهش دیگری رویکرد به سمت استفاده از گراف مغزی با استخراج ویژگی‌های گرافی از این گراف که با استفاده از یادگیری بازنمایی سعی به بهینه‌سازی پردازش و کاهش پیچیدگی زمانی در شبکه یادگیری عمیق می‌نماید تا با افزایش دقت بتواند به تشخیص افراد مبتلا به اختلالات اوتیسم کمک کند. روش یادگیری بازنمایی سعی می‌کند تا ویژگی‌های گراف و گره‌ها را حفظ کند.

در پژوهشی دیگر برای ساختن یک شبکه مغزی از یک تصویربرداری رزونانس مغناطیسی عملکردی مغز² از طرح تقسیم‌بندی بر اساس ۲۶۴ ناحیه استفاده شده است. با استفاده از الگوریتم انتخاب ویژگی، ویژگی‌های متمایزی استخراج شده و در

¹ developmental disability

² fmri

نهایت مدل شبکه‌های کانولوشنی گراف³ برای تشخیص اوتیسم با ویژگی‌های به دست آمده در مجموعه داده آموزش داده‌شد است. همچنین از پژوهش‌هایی که روی گراف‌های مغزی صورت گرفته، از مدل‌های یادگیری ماشین برای اندازه‌گیری دقت کلاسیفیکیشن استفاده شده است. در یکی از این پژوهش‌ها باند‌های فرکانسی مختلف اتصال‌های عملکردی مغز را به عنوان ویژگی استفاده کرده و با استفاده از SVM ⁴ کلاس‌بندی کرده است. و همچنین در سال 2015 سه الگوریتم کلاس‌بندی SVM ، RF ⁵ و GBM ⁶ به این منظور استفاده کردند و در نهایت به دقت 67 درصد رسیدند.

فصل دوم 2

مفاهیم و اطلاعات اولیه

2.1 مغز

درک سیستم مغزی به دلیل پیچیده بودن آن مشکل است و به همین دلیل تحقیقات پیرامون این رشته در دنیا نوپا بوده و به چند دهه اخیر برمی‌گردد. به جرات می‌توان مغز را پیچیده ترین شبکه‌ی جهان دانست که ساختاری در هم تنیده از عصب‌ها آن را در احاطه خود در آورده اند و به لطفشان تک تک رفتارهای انسان و هر جان‌دار دیگری تحت کنترل در می‌آیند. چرا که شبکه‌ی مغز از میلیاردها نورون تشکیل شده که مطالعه‌ی ارتباطات میان آن‌ها و تشخیص عملکرد نورون‌ها در هر فعالیت برای شناخت مغز بسیار حائز اهمیت است. در واقع همین ارتباطات مختلف میان نورون‌ها است که می‌تواند منجر به بروز فعالیت خاصی در عملکرد انسان شود؛ این سیستم‌های پیچیده را می‌توان به عنوان مجموعه‌ای از "راس‌ها" (یعنی مناطق مغز) و "یال‌ها" (یعنی ارتباطات بین راس‌ها) توصیف و مدل کرد.

2.2 بیماری اوتیسم

به طور کلی اختلالات نافذ مربوط به رشد یا همان اختلالات فراگیر رشدی⁷ (PDD) به پنج گروه طبقه‌بندی می‌شوند: اوتیسم⁸، آسپرگر⁹، اختلالات از هم‌گسیختگی کودکی¹⁰، سندرم رت¹¹ و اختلال فراگیر رشد که به گونه‌ای دیگر مشخص نشده¹²

³ GCN

⁴ Support vector machine

⁵ Random Forest

⁶ Gradient Boosting Machine

⁷ Pervasive Developmental Disorders

⁸ Autism

⁹ Asperger

¹⁰ Childhood disintegrative

¹¹ Rett Syndrome

نام‌گذاری شده‌است در این بین سه گروه اوتیسم، آسپرگر و اختلال فراگیر رشد که به گونه‌ای دیگر مشخص نشده در حیطه اختلالات طیف اوتیسم^{۱۳} قرار می‌گیرند. شیوع این اختلال به شدت رو به افزایش است اما هنوز دلیل قطعی برای این اختلال بیان نشده است.

اوتیسم یا درخودماندگی یک اختلال رشد است و عموماً به اشتباه با کم‌توانی ذهنی یکسان پنداشته می‌شود که در سه سال اول زندگی ظاهر می‌گردد. این بیماری با تاثیر بر روی مغز کودک، رفتارهای اجتماعی و مهارت ارتباط برقرارکردن را مختل می‌کند. به کسانی که این اختلال را دارند "اوتیستیک" گفته می‌شود. کودک اوتیستیک حتی اگر کلام داشته باشد، نمی‌تواند درست ارتباط برقرار کند، آن‌ها ارتباط یک‌طرفه برقرار می‌کنند و به همین علت ارتباطشان موثر نیست؛ در این افراد حرکات تکراری دست زدن، پريدن پاسخ‌های غیر معمول به افراد، دل‌بستگی به اشیا و یا مقاومت در مقابل تغییر نیز دیده می‌شود. به نظر می‌رسد مساله ژنتیک مهم باشد اما عوامل محیطی و فاکتورهای ژنتیکی هر دو در بروز آن موثرند.

اختلال طیف اوتیسم بیان کننده یک طیف است که دارای سه ویژگی مشخص می‌باشد: نقص در ارتباط اجتماعی، انعطاف پذیر نبودن گفتار و رفتار، حرکات کلیشه‌ای؛ اگرچه اوتیسم به عنوان یک اختلال روانی مورد بحث قرار گرفته‌است اما همراه با دیگر اختلالاتی است که بیشتر فیزیکی یا سیستماتیک می‌باشد به واسطه گسترده بودن این طیف و همچنین تداخل آن با دیگر اختلالات بررسی و آنالیز آن بسیار مشکل خواهد بود.

2.3 ماتریس اتصال مغزی – کانکتوم‌ها

تشخیص افراد مبتلا به اختلالات اوتیسم یکی از مهم‌ترین اهداف در تحقیقات علوم شناختی است. موضوع مورد بحث در این زمینه بررسی افراد مبتلا به اختلالات اوتیسم و نواحی مغزی که باعث این اختلالات می‌شوند است. یکی از روش‌هایی که اخیراً مورد توجه قرار گرفته ماتریس اتصال است، این رویکرد باعث شد تا با استفاده از شبکه‌های پیچیده به تحلیل گراف‌های مغزی و مقایسه آن‌ها بپردازند. معمولاً ماتریس اتصال با استفاده از شبکه پیچیده تحلیل می‌شود. مقالات زیادی به بررسی مغز با استفاده از معیارهای شبکه‌های پیچیده به بهبود در تشخیص اختلال اوتیسم پرداخته‌اند. در این بخش به معرفی ماتریس اتصال مغزی یا کانکتوم می‌پردازیم که به ما کمک می‌کند درباره ی مغز به عنوان یک شبکه‌ی پیچیده، اطلاعات مورد نیاز را کسب کرده و عملیات مختلفی را بر روی این شبکه پیاده‌سازی کنیم. به یک نقشه‌ی جامع از اتصالات نورون‌ها در مغز **کانکتوم** می‌گویند که با نام ماتریس اتصال مغزی نیز شناخته می‌شود. به زبان جامع‌تر، یک کانکتوم تمام اتصالات نورونی درون مغز یک موجود زنده را شامل می‌شود. کانکتوم دارای دو بخش اصلی است: نورون‌ها و اتصالات نورونی. در این ماتریس $N \times N$ سطرها و ستون‌ها بیانگر N نورون و هر درایه‌ی a_{ij} نشان دهنده ی میزان ارتباط بین دو راس i و j به عنوان دو نورون در مغز است.

¹² Pervasive Developmental Disorders Not Otherwise Specified

¹³ Autism Spectrum disorders

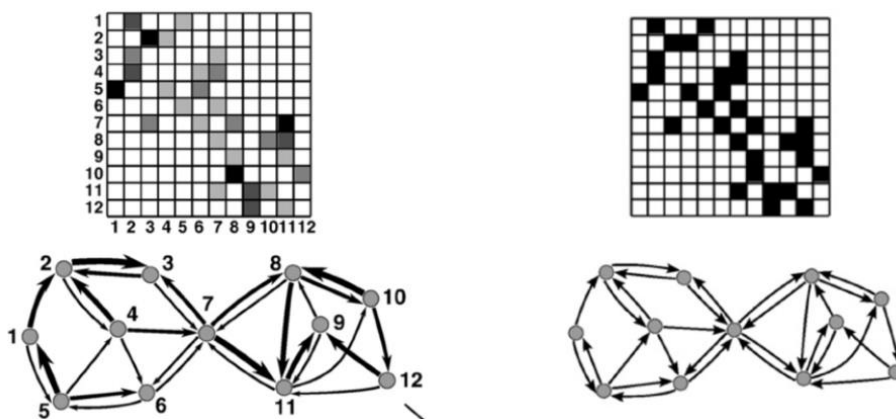
2.4 مدل کردن مغز به عنوان یک شبکه پیچیده

در ریاضیات، یک شبکه یا گراف مدل سازی ریاضی از مسئله‌های در دنیای واقعی است که به صورت مجموعه‌هایی از رئوس و یال‌ها تعریف می‌شود؛ حال برای بررسی مغز به عنوان یک شبکه‌ی پیچیده و بررسی معیارهای مختلف بر روی آن، در اولین قدم نیاز است بدانیم رئوس و یال‌های گراف را به چه صورت تعریف می‌کنیم. رئوس معمولاً نشان دهنده‌ی نواحی مختلف مغز است که تعریف این نواحی بستگی به آن دارد که در چه سطحی در حال مطالعه‌ی مغز هستیم. در تعریف رئوس باید به این نکته توجه کرد که یک قسمت از مغز در دو ناحیه جای نگیرد؛ چرا که با در نظر گرفتن یک راس برای هر ناحیه، ممکن است هویت بخشی از مغز در دو راس تعریف شده‌باشد و این بررسی گراف را دچار اشکال خواهدکرد. از طرفی رئوس باید به گونه‌ای تعریف شوند که تمامی نواحی مغز را پوشش دهند تا عملکرد تمامی نواحی در مطالعه‌ی ما دخیل باشند. یال‌ها نشان دهنده‌ی اتصالات آناتومیک یا عملکردی بین رئوس بسته به نوع مجموعه داده‌ای است؛ اتصالات آناتومیک معمولاً مربوط به رشته‌های ماده‌ی سفید بین نواحی مختلف مغز است.

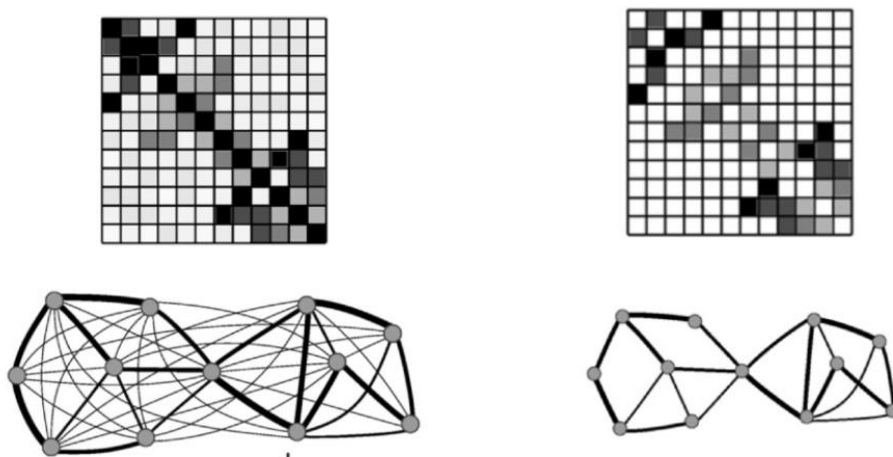
اتصالات عملکردی مغز مربوط به میزان ارتباط قسمت‌های مختلف مغز در یک زمان حین انجام یک فعالیت است که ممکن است این قسمت ها با هم ارتباط آناتومیک نداشته باشند. نشان دادن ارتباطات بین نواحی مختلف مغز در گراف به صورت یال‌های وزن دار یا غیروزن دار تعریف می‌شود. اگر گراف به صورت باینری (بدون یال‌های وزن دار) تعریف شود وجود یک یال نشان دهنده‌ی وجود ارتباط بین دو یا ناحیه‌ی مغز است. اگر گراف وزن دار باشد وزن هر یال نشان گر میزان قوی بودن ارتباط بین دو ناحیه یا چگالی ساینز رشته‌های بین دو ناحیه در مغز است.

در زیر تصویر ماتریس‌های مختلف از مغز را مشاهده می‌کنیم.

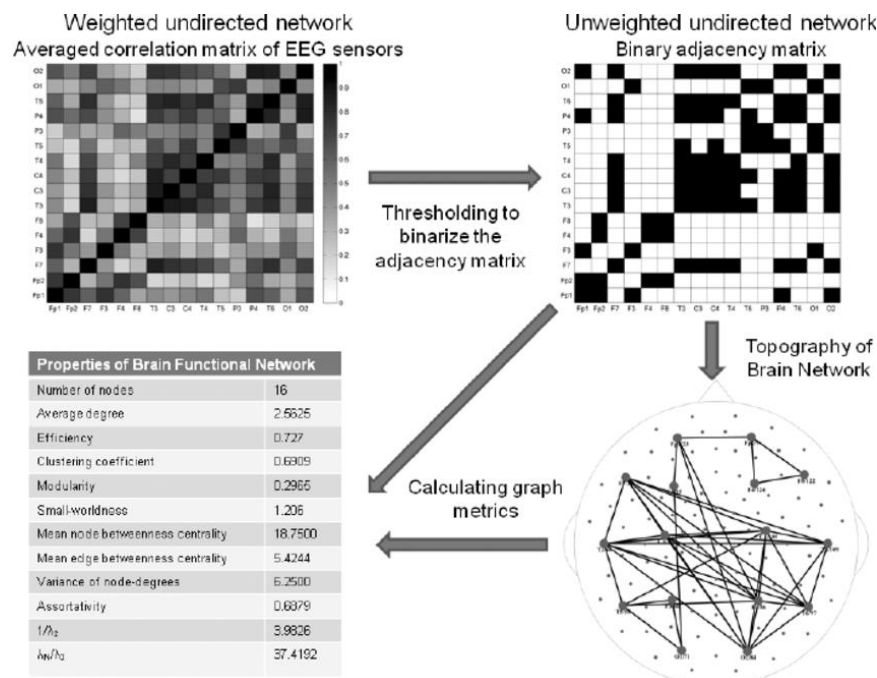
ماتریس اول که در عکس سمت چپ مشاهده می‌شود، مربوط به یک گراف وزن دار و جهت دار ساخته شده از 12 ناحیه در مغز است. مثلاً در تصویر سمت راست، از گراف وزن دار موجود در تصویر اول وزن‌هایش را حذف کرده‌ایم و صرفاً در صورت وجود اتصال بین دو راس i و j ، درایه‌ی a_{ij} را یک قرار داده ایم. همچنین در صورتی که بین دو راس i و j ، هیچ ارتباطی وجود نداشته باشد، درایه‌ی a_{ij} را صفر در نظر گرفته‌ایم.



در زیر ماتریس دیگری از همان نواحی در مغز را مشاهده می‌کنیم. این ماتریس یک ماتریس وزن‌دار و غیرجهت‌دار است. در اینجا هم سعی کرده‌ایم با در نظر گرفتن یک *threshold* برای وزن یال‌ها، یال‌های با وزن کم تر را حذف کرده و فقط یال‌هایی که در گراف حائز اهمیت هستند را نگه داریم.



از روش بالا، در قسمت‌های بعدی این پژوهش برای حذف یال‌های با اهمیت کمتر در بررسی معیارها استفاده خواهیم کرد که به طور خلاصه می‌توانید در زیر روند کلی آن را مشاهده کنید:



2.5 ویژگی‌های گراف

به طور کلی برای هر گراف ویژگی‌های مختلفی تعریف می‌شود که هر کدام می‌توانند در تعریف معیارهایی برای بررسی عملکرد نواحی مغز به ما کمک کند. برخی از این ویژگی‌ها که در پروژه نیز به کار رفته‌است به شرح ذیل می‌باشد.

2.5.1 ویژگی‌های اولیه

باید بدانیم که معیارهای مختلفی که برای بررسی گراف تعریف می‌شوند، شدیداً تحت تاثیر همین ویژگی‌های اولیه از گراف هستند. به همین علت نیاز است قبل از بررسی معیارها، این ویژگی‌های ساده را یک بار مرور کنیم.

2.5.1.1 درجه^{۱۴}

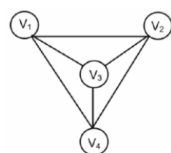
در هر گراف غیرجهت‌دار برای هر راس معیاری تحت عنوان درجه‌ی آن راس تعریف می‌شود که بیان‌گر تعداد یال‌های متصل به آن است. به بیان دیگر تعداد همسایه‌های آن راس یا همان تعداد رئوس متصل به این راس را مشخص می‌کند. درجه‌ی یک راس می‌تواند بیان‌گر میزان اهمیت آن راس در گراف باشد. چرا که تعداد ارتباطات آن راس با بقیه‌ی رئوس (در این جا ارتباط یک ناحیه از مغز با دیگر نواحی) را مشخص می‌کند.

2.5.1.2 قدرت راس^{۱۵}

مجموع وزن یال‌های متصل به یک راس را قدرت آن راس گوئیم. در گراف‌های جهت‌دار قدرت ورودی راس مجموع وزن یال‌های ورودی آن است و قدرت خروجی راس مجموع وزن یال‌های خروجی است.

2.5.1.3 تعداد مثلث‌های گراف

اگر در یک گراف سه راس داشته باشیم که دو به دو به یک دیگر متصل باشند گوئیم که این سه راس تشکیل یک مثلث داده‌اند. تعداد مثلث‌های موجود در یک گراف یکی دیگر از ویژگی‌هایی است که در بررسی گراف‌ها به ما کمک می‌کند. برای مثال در گراف زیر 4 مثلث داریم:



¹⁴ Degree

¹⁵ Node Strength

2.5.2 معیارهای تفکیک عملکردی

این دسته از معیارها در گراف مغزی بررسی می‌کنند که نواحی مختلف مغز تا چه حد این قابلیت را دارند که مستقل از دیگر بخش‌ها فعالیت مشخصی را انجام دهند. در واقع وجود کلاسترها یا ماژول‌هایی از مغز را بررسی می‌کنند که دارای این ویژگی باشند. این معیارها معمولاً بر اساس تعداد مثلث‌های گراف تعریف می‌شوند. هر چه تعداد این مثلث‌ها در گراف بیشتر باشد، تفکیک عملکردی بهتر صورت می‌گیرد.

در زیر به بررسی برخی از این معیارها می‌پردازیم:

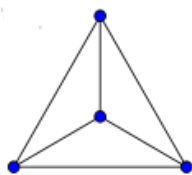
2.5.2.1 ضریب خوشگی^{۱۶}

معیاری است که تمایل راس‌ها به ایجاد یک خوشه در یک گراف را اندازه می‌گیرد. برای هر راس در گراف، تعداد مثلث‌هایی که اطراف آن راس وجود دارد را ضریب خوشگی آن گوییم. به بیان بهتر همسایه‌هایی از گراف که خودشان با یک‌دیگر همسایه هستند را بررسی می‌کنیم. برای هر راس v در گراف G ضریب خوشگی به صورت زیر تعریف می‌شود:

$$cc(v) = \frac{2N_v}{K_v(K_v - 1)}$$

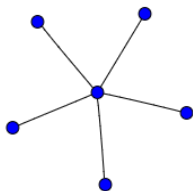
در رابطه‌ی فوق K_v درجه‌ی راس v و N_v تعداد یال‌هایی است که بین دو همسایه از راس v قرار دارد. اگر مقدار ضریب خوشگی برای یک راس برابر صفر شود به آن معناست که این گراف یک گراف ستاره‌ای است و اگر برابر یک شود به آن معناست که گراف خوشه است؛ یعنی هر راس از گراف به هر راس دیگر مسیری به طول یک دارد.

همان‌طور که مشاهده می‌کنید گراف زیر یک گراف خوشه است، با محاسبه‌ی ضریب خوشگی مطابق فرمول بالا، مقدار یک را خواهیم داشت.



$$cc(v) = \frac{2N_v}{K_v(K_v - 1)} = \frac{2 \times 3}{3(2)} = 1$$

همان‌طور که مشاهده می‌کنید گراف زیر یک گراف ستاره‌ای است، با محاسبه‌ی ضریب خوشگی مطابق فرمول بالا، مقدار صفر را خواهیم داشت.



$$cc(v) = \frac{2N_v}{K_v(K_v - 1)} = \frac{2 \times 0}{5(4)} = 0$$

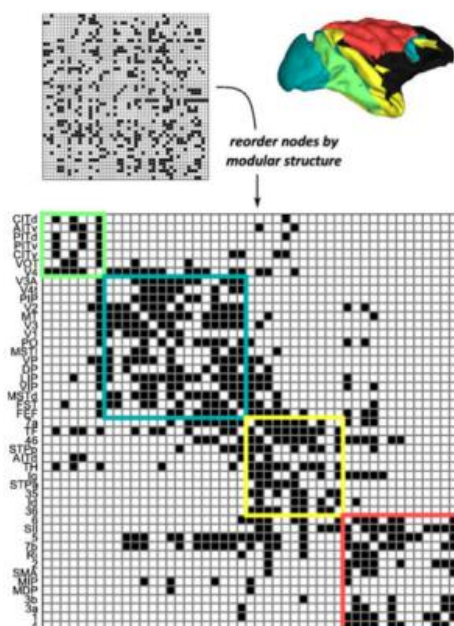
¹⁶ Clustering

2.5.2.2 ماژولاریتی

این معیار علاوه بر پیدا کردن ناحیه هایی که قابلیت تفکیک عملکرد دارند، سایز و ترکیب دقیق این نواحی را نیز تعیین می کند. با تقسیم رئوس به بخش های کوچک تر به گونه ای که داخل هر بخش تعداد ارتباطات بین آن ها ماکزیمم و بیرون از این بخش ها ارتباطات مینیمم باشد به این ترکیب ها دست پیدا می کنیم. به طور معمول معیارهای مورد بررسی ما به طور دقیق محاسبه شده و مقدار مشخصی را ارائه می دهند. اما برای پیدا کردن ماژولاریتی در گراف نیاز است از الگوریتم های بهینه سازی استفاده کنیم و جواب ما دقیق نخواهد بود. البته الگوریتم هایی که برای اینکار ارائه شده اند با دقت خوبی ماژول ها را تعیین می کنند.

در زیر نمونه ای از ماتریس مربوط به تصویر $fMRI$ به دست آمده از قشر نوعی از میمون را مشاهده می کنیم.

با به دست آوردن ماژول های قشر این میمون، ماتریس دوم را خواهیم داشت که داده ها برای هر ماژول از قشر قابل تفکیک است.



2.5.3 معیارهای تجمیع عملکردی

تجمیع عملکردی در مغز به آن معناست که بتوانیم اطلاعات از نقاط متفاوتی از مغز را با هم ترکیب کرده یا به هم ربط دهیم. بررسی این معیار معمولاً با استفاده از بررسی مسیرهای بین نواحی صورت می پذیرد. طول مسیرها به ما کمک می کند تا پتانسیل نقاط مختلف مغز برای تجمیع اطلاعاتشان را بررسی کنیم. هر چه مسیر کوتاه تر باشد احتمال آن که بتوانیم اطلاعات را تجمیع کنیم بیشتر است.

2.5.3.1 بازده سراسری^{۱۷}

میانگین معکوس طول کوتاه‌ترین مسیر را گوییم. بر عکس مشخصه‌ی طول مسیر، معیار بازده سراسری در گراف‌هایی که یال‌های کمتر بین رئوسشان وجود دارد محاسبه شده و معنا پیدا می‌کند. با توجه به این‌که در این گراف‌ها اگر بین دو راس یالی وجود نداشته باشد طول آن را بینهایت در نظر می‌گیریم، معکوس آن به صفر میل می‌کند و در نتیجه بازده برابر صفر را خواهد داشت با به طور کلی، مشخصه‌ی طول مسیر تحت اثر یال‌های با طول بیشتر در گراف و بازده سراسری تحت تاثیر یال‌هایی طول کمتر در گراف است.

2.5.3.2 خروج از مرکز^{۱۸}

خروج از مرکز یک راس، بیشترین فاصله‌ی راس مذکور از دیگر راس‌های موجود در شبکه می‌باشد. به بزرگترین خروج از مرکز راس‌های گراف قطر و به کوچکترین خروج از مرکز راس‌های گراف شعاع گراف می‌گویند.

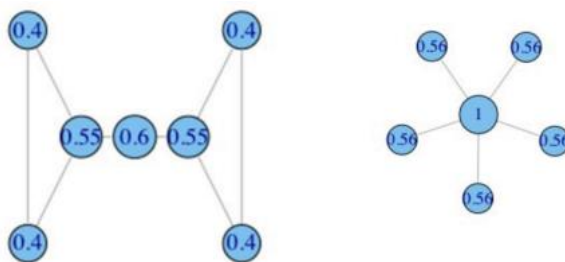
2.5.4 معیارهای مرکزیت

2.5.4.1 مرکزیت بر مبنای نزدیکی^{۱۹}

معکوس میانگین طول کوتاه‌ترین مسیرهایی که از راس i به دیگر رئوس موجود در گراف وجود دارد. این معیار طبق فرمول زیر محاسبه می‌شود:

$$Cc(v_i) = \left[\frac{1}{n-1} \sum_{j \neq i}^n g(v_i, v_j) \right]^{-1} = \frac{n-1}{\sum_{j \neq i}^n g(v_i, v_j)}$$

در زیر محاسبه‌ی این معیار را برای رئوس یک گراف داریم:



¹⁷ Modularity

¹⁸ Eccentricity

¹⁹ Closeness Centrality

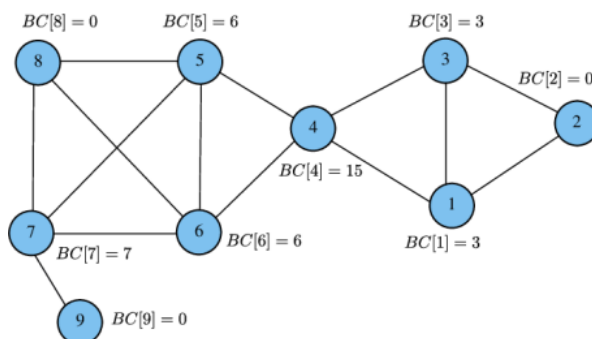
2.5.4.2 مرکزیت بینابینی^{۲۰}

مرکزیت بینابینی برای راس v ، نسبت تعداد کوتاه ترین مسیرهای بین هر دو راس s و t که از راس v می‌گذرد به تعداد کل کوتاه ترین مسیرهای بین دو راس s و t است. به بیان دیگر این معیار تعیین می‌کند که راس v در چه نسبتی از کوتاه ترین مسیرهای بین هر دو جفت راس در گراف ظاهر می‌شود. معمولاً راس‌هایی در گراف که به عنوان پُل بین بخش‌های مجزای گراف هستند دارای مرکزیت بینابینی بالاتری نسبت به دیگر رئوس هستند. برای محاسبه‌ی مرکزیت بینابینی هر راس v در گراف فرمول زیر را داریم:

$$c_B = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

که $\sigma(s,t)$ تعداد کل مسیرهای بین دو راس s و t می‌باشد و $\sigma(s,t|v)$ تعداد مسیرهای بین دو راس s و t می‌باشد به شرط آنکه v در آن مسیر حضور داشته باشد.

در تصویر زیر نمونه‌ای از محاسبه‌ی مرکزیت بینابینی را برای رئوس گراف مشاهده می‌کنیم:



2.5.4.3 مرکزیت بر مبنای درجه^{۲۱}

مرکزیت بر مبنای درجه برای یک راس v تعداد راس‌هایی است که به آن متصل است.

$$c_D(v) = \frac{\deg(v)}{n-1}$$

²⁰ Betweenness Centrality

²¹ Degree Centrality

2.6 آشنایی با خوشه‌بندی^{۲۲}

الگوریتم‌های متنوع و زیادی برای انجام عملیات خوشه‌بندی به کار می‌رود. تحلیل خوشه‌بندی^{۲۳} روشی بدون ناظر^{۲۴} است که به کمک آن می‌توان مجموعه‌ای از اشیاء را به گروه‌های مجزا افراز کرد. هر افراز یک خوشه نامیده می‌شود. اعضاء هر خوشه با توجه به ویژگی‌هایی که دارند به یکدیگر بسیار شبیه هستند و در عوض میزان شباهت بین خوشه‌ها کمترین مقدار است. در چنین حالتی هدف از خوشه‌بندی، نسبت دادن برچسب‌هایی به اشیاء است که نشان دهنده عضویت هر شیء به خوشه است. اگر چه بیشتر الگوریتم‌ها یا روش‌های خوشه‌بندی مبنای یکسانی دارند ولی تفاوت‌هایی در شیوه اندازه‌گیری شباهت یا فاصله بین آن هاست. الگوریتم‌های خوشه‌بندی انواع مختلفی دارد که در ادامه کار یکی از انواع این الگوریتم‌ها که متعلق به نوع سلسله مراتبی است معرفی خواهد شد.

2.6.1 خوشه‌بندی سلسله مراتبی^{۲۵}

خوشه‌بندی سلسله مراتبی در هر سطح از فاصله، نتیجه خوشه‌بندی را نشان می‌دهد. این سطوح به صورت سلسله مراتبی^{۲۶} هستند. برای نمایش نتایج خوشه‌بندی به صورت سلسله مراتبی از درختواره^{۲۷} استفاده می‌شود. این شیوه، روشی موثر برای نمایش نتایج خوشه‌بندی سلسله مراتبی است. ابتدا هر مقدار یک خوشه محسوب می‌شود. در طی مراحل خوشه‌بندی، نزدیکترین مقادارها (براساس تابع فاصله تعریف شده) با یکدیگر ادغام شده و خوشه جدیدی می‌سازند. توجه داشته باشید که مقادارهایی که روی محور افقی دیاگرام دیده می‌شوند، بیانگر موقعیت مقادارها در لیست داده‌ها است و نه خود مقادیر. همچنین محور عمودی نیز فاصله یا ارتفاع بین خوشه‌ها را نشان می‌دهد.

در الگوریتم خوشه‌بندی سلسله مراتبی معمولاً با کوچکترین خوشه‌ها شروع می‌کنیم. به این معنی که در ابتدا هر مقدار بیانگر یک خوشه است. سپس دو مقداری که بیشترین شباهت (کمترین فاصله) را دارند با یکدیگر ادغام شده و یک خوشه جدید می‌سازند. بعد از این مرحله، ممکن است دو مقدار یا یک مقدار با یک خوشه یا حتی دو خوشه با یکدیگر ادغام شده و خوشه جدیدی ایجاد کنند؛ این عملیات با رسیدن به بزرگترین خوشه که از همه مقادارها تشکیل یافته پایان می‌یابد. به این ترتیب اگر n مقدار داشته باشیم در مرحله اول تعداد خوشه‌ها برابر با n است و در آخرین مرحله یک خوشه باقی خواهد ماند. معمولاً به این شیوه خوشه‌بندی سلسله مراتبی، روش تجمیعی^{۲۸} می‌گویند. برای اندازه‌گیری فاصله بین دو خوشه یا یک مقدار با یک خوشه

²² Clustering

²³ Cluster Analysis

²⁴ Unsupervised

²⁵ Hierarchical Clustering

²⁶ Hierarchy

²⁷ Dendrogram

²⁸ Agglomerative

از روش‌های پیوند^{۲۹} استفاده می‌شود روش پیوند مورد استفاده در این پژوهش روشی به نام *ward* صورت است بدین صورت که سعی می‌کند واریانس خوشه‌های ادغام شده را به حداقل برساند.

فصل سوم 3

پیش‌نیازها

3.1 تحلیل واریانس^{۳۰}

تحلیل واریانس و روش‌های تجزیه واریانس، یکی دسته از مدل‌های آماری هستند که قادرند اختلاف بین گروه‌ها یا دسته‌ها را بررسی کنند. در ساده‌ترین شکل، تحلیل واریانس می‌تواند به عنوان یک روش برای آزمون فرض مقایسه میانگین در بین چند جامعه مستقل به کار رود. این کار به عنوان یک جایگزین برای آزمون فرض با استفاده از آماره آزمون T است.

تحلیل واریانس در حالت کلاسیک راه حلی است که سه عمل زیر را همزمان انجام می‌دهد:

1. تجزیه مجموع مربعات کل به مجموع مربعات اجزا حاصل از مدل خطی

2. مقایسه میانگین مربعات، به کمک آماره و آزمون F

3. آزمون پارامترهای مدل به منظور دستیابی به مدل آماری مناسب

فرض کنید قرار است در مورد یکسان بودن دو یا چند جامعه تحقیق کنید. یکی از شاخص‌های قابل استفاده برای بیان ویژگی‌های جامعه‌ها، می‌تواند میانگین باشد. با مقایسه میانگین و تشخیص برابری یا نابرابر بودن آن‌ها در بین جامعه‌ها، می‌توان رای به یکسان یا متفاوت بودن آن‌ها داد. بنابراین اگر یکی از میانگین‌ها با بقیه تفاوت داشته باشد، متوجه می‌شویم که جوامع مانند یکدیگر نیستند.

با توجه به این موضوع، می‌توان فرضیه‌های آزمون برای مقایسه میانگین k جامعه را به صورت زیر نوشت:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \dots = \mu_k \\ H_1 : \text{there are some } \mu\text{'s not equal with others} \end{cases}$$

در اینجا فرض مقابل یا H_1 بیان می‌کند که حداقل یکی از میانگین‌ها با بقیه تفاوت دارد. می‌دانیم که احتمال خطای نوع اول برای مسئله اصلی آزمون به صورت زیر نوشته می‌شود:

$$\alpha_t = P(\text{Error Type I}) = P(\text{Reject } H_0 \mid H_0 \text{ is true})$$

²⁹ Linkage

³⁰ Analysis of Variance (ANOVA)

این عبارت به معنی احتمال رد فرض صفر به شرط صحیح بودن آن است. صحیح بودن فرض صفر بیانگر برابر بودن میانگین‌ها با یکدیگر خواهد بود. برای انجام این بررسی ممکن است از آزمون مقایسه میانگین در بین دو جامعه استفاده کنیم و به صورت ترکیب‌های دوتایی مسئله را تحلیل کنیم.

تحلیل واریانس در مسائل مربوط به آزمون میانگین چند جامعه، بر اساس تجزیه «پراکندگی کل^{۳۱}» به «پراکندگی بین گروه‌ها^{۳۲}» و «پراکندگی درون گروه‌ها^{۳۳}» صورت می‌پذیرد. بنابراین اگر پراکندگی کل را با «مجموع مربعات کل^{۳۴}»، پراکندگی بین گروهی را با «مجموع مربعات بین گروه‌ها^{۳۵}» و پراکندگی درون گروهی را «با مجموع مربعات درون گروهی^{۳۶}» نشان دهیم، خواهیم داشت:

$$SST = SSB + SSW$$

بر این اساس، با توجه به حضور یک متغیر تاثیرگذار در جامعه‌ها، محاسبات مربوط به جدول تحلیل واریانس برای آزمون مقایسه میانگین چند جامعه طبق جدول «تحلیل واریانس یک طرفه^{۳۷}» انجام می‌شود. در زیر یک نمونه از چنین جدولی قابل رویت است:

منشاء تغییرات	درجه آزادی	مجموع مربعات	میانگین مربعات	آماره F
بین گروه‌ها	$k - 1$	SSB	$MSB = \frac{SSB}{k - 1}$	$F = \frac{MSB}{MSW}$
درون گروه‌ها	$n - k$	SSW	$MSW = \frac{SSW}{n - k}$	
کل	$n - 1$	SST		

در این جدول k تعداد گروه‌ها یا جامعه‌ها و n نیز تعداد مشاهدات است.

متغیری که باعث اختلاف در جامعه‌ها می‌شود را گاهی «عامل^{۳۸}» نیز می‌نامند. در نتیجه اگر فرض صفر در آزمون فرض میانگین چند جامعه رد شود، می‌توان گفت که متغیر عامل در تغییر میانگین جامعه موثر است.

در صورتی که مقدار p -value کمتر از 0.05 باشد بدین معناست که فرض صفر رد شده و پراکندگی درون گروه‌ها کمتر از پراکندگی بین گروه‌ها خواهد بود.

³¹ Total Variation

³² Variation between groups

³³ Variation within groups

³⁴ Total Sum of Squares- SST

³⁵ Between Sum of Squares- SSB

³⁶ Within sum of squares- SSW

³⁷ One Way Anova

³⁸ Factor

3.2 آزمون های تعقیبی^{۳۹}

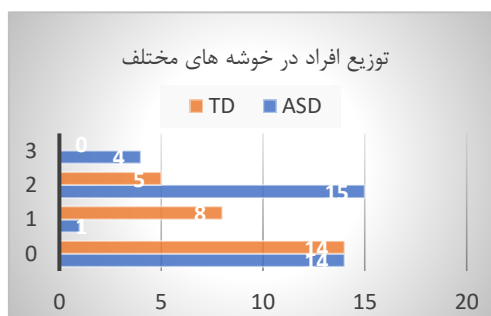
در آنالیز واریانس یک عامله در صورت رد فرض صفر (یعنی تفاوت معنادار است) ما می توانیم برای تشخیص تفاوت درون گروه ها از آزمونهای تعقیبی استفاده کنیم. در واقع ببینیم این تفاوت در بین کدام یک از گروه ها وجود دارد. تحلیل واریانس نشان می دهد که آیا نمونه ها متعلق به جامعه هستند یا خیر. در صورتی که فرض صفر، رد شود، معلوم نیست که کدام یک از نمونه ها در کدام جامعه قرار دارند. به عبارت دیگر، معنی دار شدن نسبت F به ما نمی گوید که اختلاف بین کدام جفت از میانگین ها معنی دار است. بلکه با آمار F ، تنها می توانیم پی ببریم که اختلاف بین میانگین گروه ها معنی دار است.

از این آزمون در تحلیل هایی که در ادامه می آید بدین منظور استفاده شده، در صورتی که انووا فرض صفر را رد کرد، با تحلیلی دقیق تر ببینیم در این ویژگی خاص خاص کدام دو جفت گروه تفاوت معنا داری با یک دیگر دارند.

3.3 نحوه ی ارزیابی خوشه بندی ها

برای آن که بتوان معیاری برای ارزیابی درست خوشه بندی های در هر تعداد خوشه داشت روشی که در ادامه توضیح داده شده است را اعمال خواهیم کرد؛ خوشه ای در دسته بندی های ما خوب ارزیابی می شود که اختلاف تعداد افراد سالم به ناسالم (و برعکس) در آن ها قابل ملاحظه باشد. اگر در یک خوشه تعداد افراد سالم و ناسالم مقدار برابری باشد آن گروه مناسبی به حساب نمی آید. در نتیجه اختلاف تعداد افراد سالم به ناسالم در حالت قدر مطلق آن نسبت به کل تعداد گروه اگر به یک نزدیک باشد گروه کاملاً سالم یا کاملاً ناسالم، گروه مثبتی است. به مثال زیر توجه کنید:

فرض کنید ما چهار خوشه داریم که نحوه ی توزیع افراد سالم به ناسالم در آن به شرح زیر است:



شکل (۱-۳) تعداد افراد سالم و اوتیسم در خوشه بندی های مختلف

- گروه با برچسب 0: تعداد افراد سالم و ناسالم در آن برابر است در نتیجه اختلاف آنها برابر با 0 می باشد.
- گروه با برچسب 1: دارای 8 فرد سالم و 1 اوتیسم اختلاف این دو عدد نسبت به کل اعضای مجموعه برابر است با $\frac{7}{9}$ می توان به بیانی این گروه را نمایانگر گروه سالم نامید.
- گروه با برچسب 2: دارای 15 فرد اوتیستیک و 5 فرد سالم که اختلاف آنها نسبت به کل اعضا $\frac{10}{20}$ می باشد.
- گروه با برچسب 3: هر 4 عضو این خوشه افراد اوتیسم هستند در نتیجه اختلاف آنها نسبت به کل اعضای مجموعه برابر با یک می باشد؛ این گروه نیز نمایانگر افراد مبتلا به اوتیسم خواهد بود.

حال برای ارزیابی باید جمع هر چهار گروه در بهترین حالت برابر با کل تعداد گروه ها باشد یعنی در اینجا برابر 4 باشد.

جمع اعداد : $0 + 0.78 + 0.5 + 1 = 2.28$ و به طور درصدی : $\frac{2.28}{4} = 0.57$ or 57%

بدین صورت معیاری برای ارزیابی عملکرد کلی خوشه ها با تعداد متغیر خواهیم داشت.

3.4 نحوه ی انتخاب ویژگی ها⁴⁰

در ابتدا باید دانست که ویژگی ای خوب محسوب می شود که در تست آنوا p -value کمتر از 0.05 داشته باشد.

در این پژوهش ویژگی ها در دو سطح متفاوت انتخاب شده است:

1- ویژگی هایی که مقدار p -value آنها چه در برچسب های اصلی و چه در خوشه بندی های صورت گرفته کمتر از 0.05 باشد به عنوان ویژگی های منتخب انتخاب می شوند.

2- ویژگی هایی که مقدار p -value آنها چه در برچسب های اصلی و چه در خوشه بندی های صورت گرفته کمتر از 0.02 باشد به عنوان ویژگی برگزیده انتخاب می شود.

علت استفاده از دو ترشهلد بدین منظور است که ویژگی هایی که پراکندگی درون گروهی کمتری نسبت به پراکندگی بین گروهی در آزمون آنوا که قبل تر توضیح داده شد به طور سختگیرانه تری انتخاب شوند و نتایج مورد بررسی قرار خواهد گرفت.

3.5 نحوه ی استفاده از آزمون $post-hoc$ و جداول آن

با توجه به آنچه در مورد تست $post-hoc$ در فصل گذشته توضیح داده شد هدف استفاده ما از این تست آن است که ببین کدام ویژگی ها بین کدام یک از دسته بندی ها تفکیک درستی را اعمال کرده اند و کدام ویژگی ها هستند که بین دو به دوی گروه پراکندگی درون گروهی کمتری را اعمال کرده اند به بیانی دیگر دو به دوی گروه ها در ویژگی مدنظر به خوبی از هم جدا شده اند یا خیر.

کدام گروه ها هستند که بیشترین تفکیک را از این منظر با دیگر گروه ها دارند؟

آیا نتایج این آزمون با بررسی برچسب های اصلی در داده های ما صدق میکند یا خیر؟

در فصل آینده این موضوع به طور مفصل بررسی خواهد شد.

جدول موجود در بخش 5.3 در سطر یکی به آخر نشان می دهد این برچسب در چند ویژگی با حداقل یکی از گروه ها درگیر اند و سطر آخر نیز نشان می دهد که این برچسب با چند گروه در ویژگی های مختلف درگیر است.

⁴⁰ Feature Selection

3.6 چگونگی مقایسه نمودارهای جعبه ای

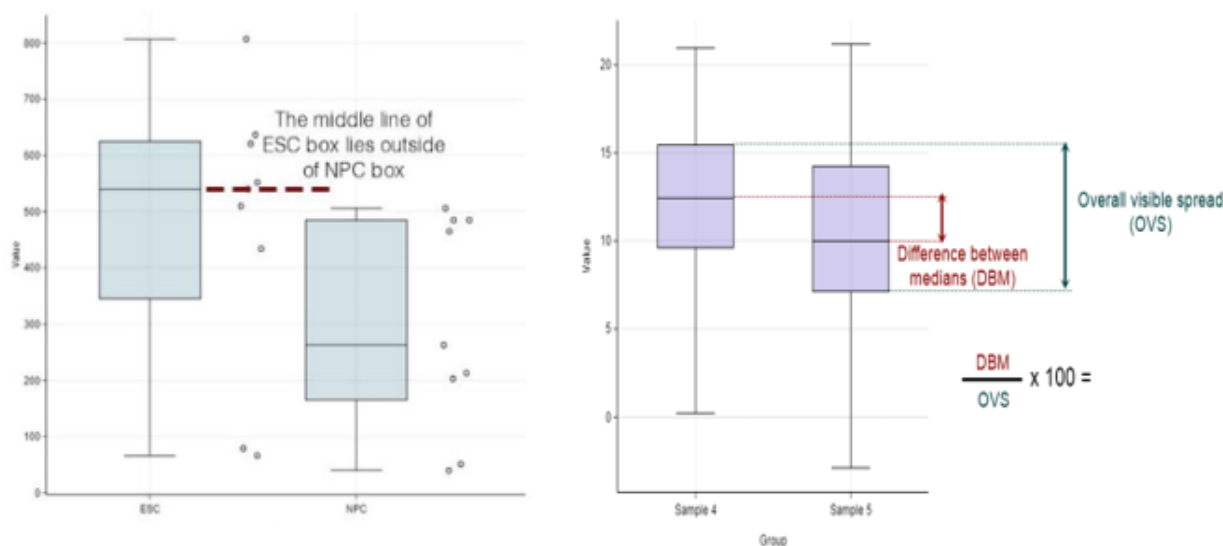
برای مقایسه ی دو نمودار جعبه ای و به چه میزان دو گروه متفاوت هستند از چند نماد استفاده شده است

$A+$: جعبه در دو نمودار به طور کامل از هم جدا شده اند.

$B+$: مقدار میانه ی هر دو جعبه خارج از ماکسیمم و مینیمم جعبه دیگر قرار دارد.

B : مقدار میانه یک جعبه خارج از ماکسیمم و مینیمم جعبه دیگر قرار دارد.

C : فاصله ی دو میانه نسبت به اندازه روی هم رفتن دو جعبه بیشتر از 0.33 باشد. (به اندازه کافی میانه ی دو گروه از هم جدا باشند)



شکل (۲-۳) موقعیت‌های مختلف قرار گیری جعبه‌ها

فصل چهارم 4

نتایج و پیاده‌سازی

4.1 توضیحات کلی پیرامون روند انجام پروژه

برای بررسی و شناخت اوتیسم مجموعه داده‌ی *Autism UCLA* فراهم شده است. این مجموعه داده شامل تصاویر *fMRI* و ماتریس اتصال مغزی افراد سالم و بیمار می باشد. تعداد بیماران مبتلا به اوتیسم در این مجموعه داده 42 و تعداد افراد سالم 37 نفر می باشد.

کانکتوم های مغزی در این مجموعه داده به صورت ماتریس مجاورت با ۲۶۴ نود می باشد که اعداد داخل این ماتریس ۲۶۴ در ۲۶۴ وزن های بین دو نود است برای رسیدن از این تعریف به تعریف متداول گراف باینری سه سطح ترشهلد در نظر گرفته شده که نهایت بهترین آن برای تحلیل های نهایی و پیدا کردن روند ها مورد استفاده قرار گرفت.

در ابتدای کار برای انتخاب ویژگی ها با علم به دانشتن برچسب ها فرایند پژوهش را جلو می بریم ولی در نهایت تحلیل ها را بدون برچسب انجام خواهیم داد؛ پس از رسیدن به یک گراف باینری ابتدا ویژگی های ذکر شده در فصل ۳ را استخراج کرده و مقادیر ماکسیمم، مینیمم، میانه، مقدار مربوط به چارک اول و سوم و میانگین هر ویژگی محاسبه می شود و در نهایت خوشه بندی گراف های مغزی را با استفاده از الگورینم مبتنی بر سلسله مراتب *Agglomerative Clustering* انجام می دهیم؛ خوشه بندی ها در دو، سه، چهار و پنج دسته صورت می گیرد تا علاوه بر پیگیری روند تقسیم و توزیع افراد در دسته های مختلف بتوانیم در مراحل بعدی ارزیابی دقیق تری روی هر خوشه به طور خاص داشته باشیم.

سپس با استفاده از تحلیل واریانس ویژگی های خوب را انتخاب کرده و بار دیگر خوشه بندی انجام می دهیم (یکبار خوشه بندی با استفاده از ویژگی های منتخب و بار دیگر با ویژگی های برگزیده انجام شده است). که در نهایت گراف های مغزی درون هر خوشه با استفاده از ویژگی های جدید تحلیل و ارزیابی خواهیم کرد.

4.2 بررسی ویژگی ها با استفاده از تحلیل واریانس

در ابتدا کانکتوم های به وسیله سه ترشهلد مختلف مورد ارزیابی قرار گرفته است که در ادامه نتایج حاصل از هریک از آنها آورده خواهد شد.

خوشه بندی گراف های مغزی با استفاده از ترشهلد 0.4

پس از رسیدن به گراف های باینری، ما اقدام به استخراج ویژگی ها از روی این گراف ها می کنیم؛ ویژگی های استخراج شده بر دو نوع اند: برخی ویژگی ها بر روی هر نود از گراف استخراج می شوند بدین صورت که گراف ما دارای ۲۶۴ نود است، ۲۶۴ ویژگی استخراج می شود؛ ویژگی هایی مثل تعداد مثلث های روی هر نود، مرکزیت بینابینی، ضریب خوشگی و...

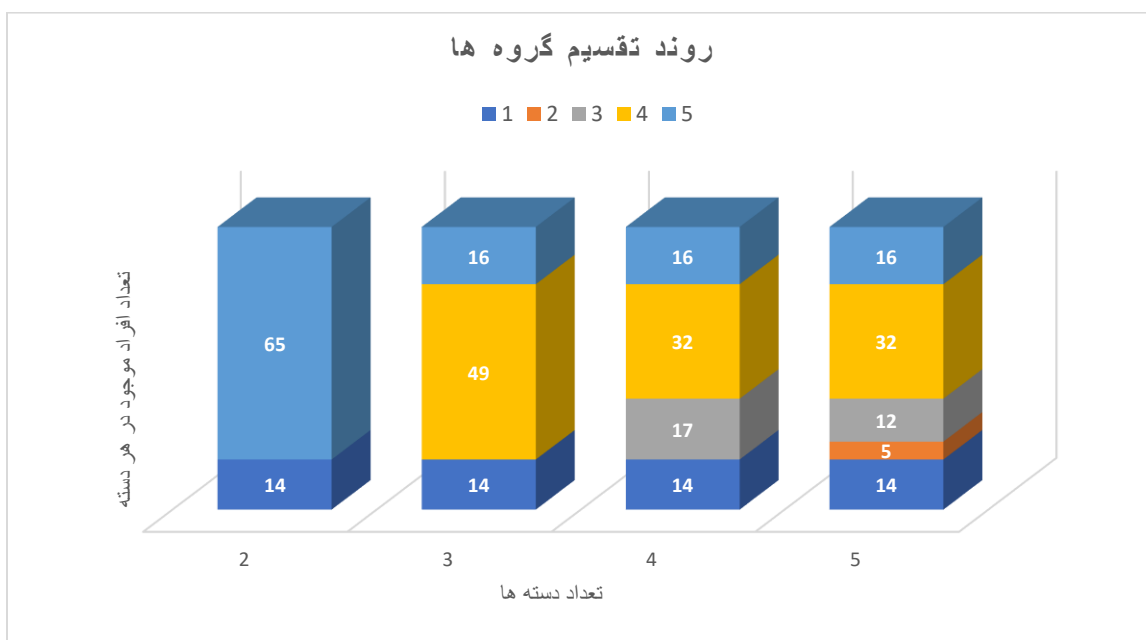
دسته دوم از ویژگی ها به طول کلی روی کل گراف تعریف می شوند مثل ماژولاریتی و بازده سراسری.

برای آن که تعداد ویژگی ها با تعداد داده ها همخوانی داشته باشند از ویژگی های دسته اول تنها مقادیر نود هایی مد نظر ما خواهند بود که مقادیر آنها ماکسیمم، مینیمم، میانه، مقدار مربوط به چارک اول و سوم و میانگین از کل نودهای هر گراف را دارا باشند.

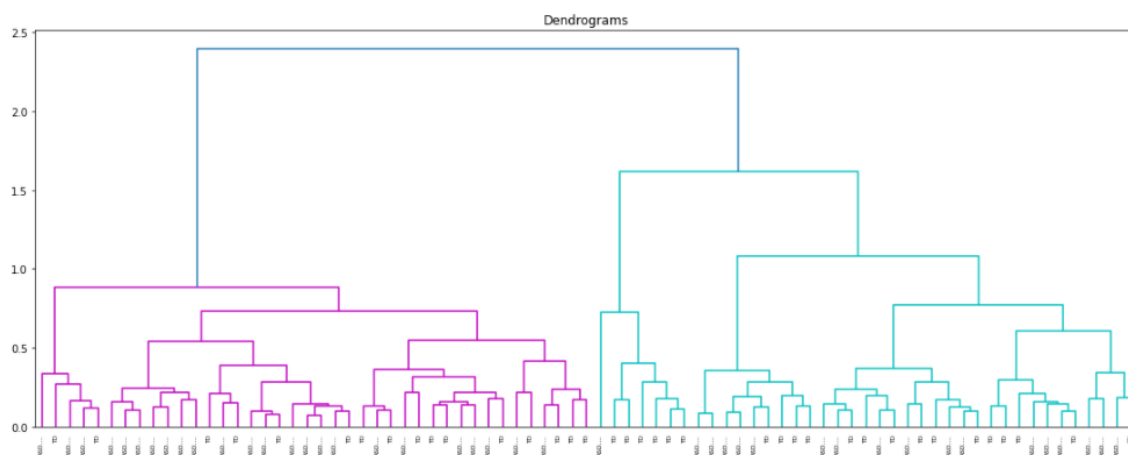
به طور مثال برای کانکتوم شماره یک برای ویژگی ضریب خوشگی نود های آن ۲۶۴ مقدار به دست می آید که از میان آن ها مقدار ماکسیمم، مینیمم، میانه، میانگین و دو مقدار مربوط به چارک اول و سوم در میان ویژگی ها استخراج خواهد شد.

در نتیجه ما در اینجا 7 ویژگی از دسته اول داریم که برای هر یک 6 ویژگی فوق استخراج خواهد شد به علاوه 2 ویژگی از دسته دوم لذا برای هر کانکتوم 44 ویژگی استخراج میشود. در نهایت این ویژگی ها با استفاده از متد $Z\text{-score}$ نرمال شده و برای خوشه بندی اولیه به الگوریتم خوشه بندی سلسله مراتبی ذکر شده آماده سازی می شوند. 79 کانکتوم مغزی در داده مجموعه ما با روش بدون نظارت *Agglomerative Clustering* به ترتیب دسته های دو، سه، چهار و پنج تایی خوشه بندی می شوند.

4.2.1 خوشه بندی کانکتومها با استفاده از تمام ویژگی ها - 44 ویژگی



شکل (۴-۱) - بررسی روند دسته بندی داده ها با کل ویژگی ها - 44 ویژگی



شکل (۴-۲) - دیاگرام مربوط به خوشه بندی با استفاده از 44 ویژگی

4.2.1.1 تحلیل واریانس

با استفاده از تحلیل واریانس که در فصل قبل به آن اشاره شد برای هر 44 ویژگی ذکر شده در قسمت قبل که به وسیله آن خوشه بندی های دو الی پنج تایی را انجام دادیم مناسب بودن ویژگی ها را بررسی می کنیم در صورتی که مقدار p در تحلیل واریانس بیشتر از 0.05 باشد بدین معناست که این ویژگی در این نوع از دسته بندی پراکندگی درون گروهی بیشتری نسبت پراکندگی بین گروهی دارد و ویژگی مثبتی ارزیابی نمی شود. برای نمونه به جدول زیر توجه نمایید:

Features		2 clusters	3 clusters	4 clusters	5 clusters	True Label
Betweenness centrality	Max	0.024289	0.074308	0.024402	0.047519	0.113084
	Chart3	7.373861e-10	4.192990e-10	2.884368e-12	1.515370e-11	0.031212
	Med	2.326690e-11	4.347093e-11	2.467396e-12	1.399424e-11	0.030819
	Mean	2.923158e-13	2.500668e-13	1.020958e-17	7.539236e-17	0.015452
	Chart1	9.254494e-07	5.899664e-06	1.693074e-09	8.309867e-09	0.049625
	Min	0.200060	0.332894	0.159761	0.123255	0.093287

مقادیر موجود در جدول بالا همان p -value حاصل از تحلیل واریانس می باشد.

ویژگی مورد بحث در اینجا مرکزیت بینابینی است؛ ستون آخر از این جدول مقادیر p را بین دو دسته اوتیسم و سالم با فرض دانستن برچسب ها به ما می دهد و ستون های 3 تا 6 نیز مقادیر p را به ترتیب برای خوشه های دوتایی، سه تایی، چهارتایی و پنج تایی نشان میدهد.

برای دانستن مفاهیم رنگ های موجود در جدول لطفا به راهنمای استفاده از جدول و همچنین برای مشاهده کامل جدول ویژگی ها در بخش ضمیمه ها مراجعه بفرمایید.

ماکسیمم مرکزیت بینابینی در میان خوشه بندی های صورت گرفته، ویژگی خوبی ارزیابی شده است ولی در بین برچسب های واقعی اوتیسم و سالم ویژگی مثبتی ارزیابی نشده است در نتیجه این ویژگی به رنگ نارنجی در آمده است.

بدین معناست که در میان خوشه بندی های صورت گرفته نود هایی که ماکسیمم مقدار مرکزیت بینابینی را دارند یعنی نود هایی که واسطه ارتباطات زیادی میان نود های دیگر هستند پراکندگی درون گروهی کمتری نسبت به پراکندگی بین گروهی دارند ولی در میان برچسب های واقعی ویژگی این نود ها سبب پراکندگی داده ها شده و لذا تاثیری مثبتی در روند خوشه بندی ما نگذاشته است.

چارک سوم، میانه و چارک اول ویژگی‌هایی است که مقدار p آن‌ها در برچسب‌های اصلی بین 0.05، 0.02 قرار دارد ولی در دسته بندی‌ها مقادیر کمتر از 0.05 دارند در نتیجه این سه ویژگی مثبت ارزیابی شده و در دسته ویژگی‌های منتخب قرار می‌گیرند. میانگین مرکزیت بینابینی چه در خوشه بندی‌ها و چه در برچسب‌های اصلی ویژگی خوبی ارزیابی شده است و با توجه به تعریف جز ویژگی‌های برگزیده قرار می‌گیرد.

4.2.1.2 بررسی ویژگی‌ها

برای مشاهده دقیق اعداد و نیز بررسی تحلیل واریانس به بخش ۹.۲ در فصل ضمیمه‌ها مراجعه بفرمایید.

1. قدرت راس

در کلیه دسته‌ها ماکسیمم، چارک سوم، مینیمم، چارک اول مقدار p -value کمتر از 0.05 دارد. در کلیه دسته‌ها میانگین و میانه مقدار p -value بیشتر از 0.05 دارد. لیبل‌های اصلی برای هیچ یک از این ویژگی‌ها مقدار p -value کمتر از 0.05 ندارد. نتیجه: در نهایت هیچ یک از این ویژگی‌ها جز ویژگی‌های منتخب انتخاب نخواهد شد.

2. خروج از مرکز

در کلیه دسته‌ها میانه، میانگین، چارک اول مقدار p -value کمتر از 0.05 دارد. در کلیه دسته‌ها ماکسیمم مقدار p -value بیشتر از 0.05 دارد. در کلیه دسته‌ها چارک سوم، مینیمم مقدار p -value نامعینی دارد که به این معناست که این ویژگی میان کانکتوم‌های مختلف مقادیر یکسانی دارند. لیبل‌های اصلی برای هیچ یک از این ویژگی‌ها مقدار p -value کمتر از 0.05 ندارد یا مقدار nan دارد. نتیجه: در نهایت هیچ یک از این ویژگی‌ها جز ویژگی‌های منتخب انتخاب نخواهد شد.

3. مثلث‌ها (تعداد مثلث‌هایی که هر نود درون آن حضور دارند)

در کلیه دسته‌ها ماکسیمم، چارک سوم، میانه، میانگین، چارک اول مقدار p -value کمتر از 0.02 دارد. در کلیه دسته‌ها مینیمم مقدار p -value بیشتر از 0.05 دارد. (به جز دسته دو)

لیبل های اصلی برای ماکسیمم، چارک سوم، میانه، میانگین مقداری بین 0.05 و 0.02 دارند.

لیبل های اصلی برای چارک اول مقداری کمتر از 0.02 دارند.

لیبل های اصلی برای مینیمم مقداری کمتر از 0.02 دارند.

نتیجه : در نهایت چارک اول به عنوان ویژگی برگزیده انتخاب می شود.

نتیجه : در نهایت ماکسیمم، چارک سو، میانه، میانگین، چارک اول به عنوان ویژگی منتخب اعلام می شوند.

4. ضریب خوشگی

در کلیه دسته ها میانه، میانگین، چارک اول مقدار $p-value$ کمتر از 0.02 دارد.

در کلیه دسته ها مینیمم مقدار $p-value$ بیشتر از 0.05 دارد. (به جز دسته دو)

لیبل های اصلی برای ماکسیمم، چارک سوم مقداری بین 0.05 و 0.02 دارند.

لیبل های اصلی برای چارک اول، میانه، میانگین مقداری کمتر از 0.02 دارند.

لیبل های اصلی برای مینیمم مقداری کمتر از 0.02 دارند.

نتیجه : در نهایت چارک اول، میانگین، میانه به عنوان ویژگی برگزیده انتخاب می شود.

نتیجه : در نهایت ماکسیمم، چارک سوم، میانه، میانگین، چارک اول به عنوان ویژگی منتخب اعلام می شوند.

5. مرکزیت بینابینی

در کلیه دسته ها میانه، میانگین، چارک سوم، چارک اول مقدار $p-value$ کمتر از 0.02 دارد.

در کلیه دسته ها مینیمم مقدار $p-value$ بیشتر از 0.05 دارد. (به جز دسته دو).

در کلیه دسته ها ماکسیمم مقدار $p-value$ کمتر از 0.05 دارد.

لیبل های اصلی برای چارک اول، میانه، چارک سوم مقداری بین 0.05 و 0.02 دارند.

لیبل های اصلی برای میانگین مقداری کمتر از 0.02 دارند.

لیبل های اصلی برای مینیمم، ماکسیمم مقداری بیشتر از 0.05 دارند.

نتیجه : در نهایت میانگین به عنوان ویژگی برگزیده انتخاب می شود.

نتیجه : در نهایت میانه، میانگین، چارک اول، چارک سوم به عنوان ویژگی منتخب اعلام می شوند.

6. مرکزیت بر مبنای نزدیکی

در کلیه دسته ها میانه، ماکسیمم، میانگین، چارک اول، چارک سوم مقدار $p\text{-value}$ کمتر از 0.02 دارد.

در کلیه دسته ها مینیمم مقدار $p\text{-value}$ بیشتر از 0.05 دارد. (به جز دسته دو)

لیبل های اصلی برای چارک اول مقداری بین 0.05 و 0.02 دارند.

لیبل های اصلی برای ماکسیمم، چارک سوم، میانه، میانگین مقداری کمتر از 0.02 دارند.

لیبل های اصلی برای مینیمم مقداری بیشتر از 0.05 دارند.

نتیجه : در نهایت میانه، ماکسیمم، میانگین، چارک سوم، به عنوان ویژگی برگزیده انتخاب می شود.

نتیجه : در نهایت میانه، ماکسیمم، میانگین، چارک اول، چارک سوم به عنوان ویژگی منتخب اعلام می شوند.

7. مرکزیت بر مبنای درجه

در کلیه دسته ها ماکسیمم، چارک سوم، میانه، میانگین، چارک اول مقدار $p\text{-value}$ کمتر از 0.02 دارد.

در کلیه دسته ها مینیمم مقدار $p\text{-value}$ بیشتر از 0.05 دارد.

لیبل های اصلی برای ماکسیمم، چارک سوم، میانه، میانگین، چارک اول مقداری کمتر از 0.02 دارند.

لیبل های اصلی برای مینیمم مقداری بیشتر از 0.05 دارند.

نتیجه : در نهایت ماکسیمم، چارک سوم، میانه، میانگین، چارک اول به عنوان ویژگی برگزیده انتخاب می شود.

نتیجه : در نهایت ماکسیمم، چارک سوم، میانه، میانگین، چارک اول به عنوان ویژگی منتخب اعلام می شوند.

8. ماژولاریتی

در هیچ یک از تعداد دسته بندی خوشه ها مقدار $p\text{-value}$ کمتر از 0.05 نبوده و به دین معناست که پراکندگی داخل دسته

ها بیشتر از پراکندگی بین دسته هاست.

این ویژگی حتی زمانی که دسته بندی بین گروه های واقعی اوتیسم و سالم قرار میگیرد هم مقدار $p\text{-value}$ زیادی دارد.

نتیجه : به نظر می رسد این ویژگی باعث میشود خوشه بندی درستی روی داده ها صورت نگیرد.

9. بازده سراسری

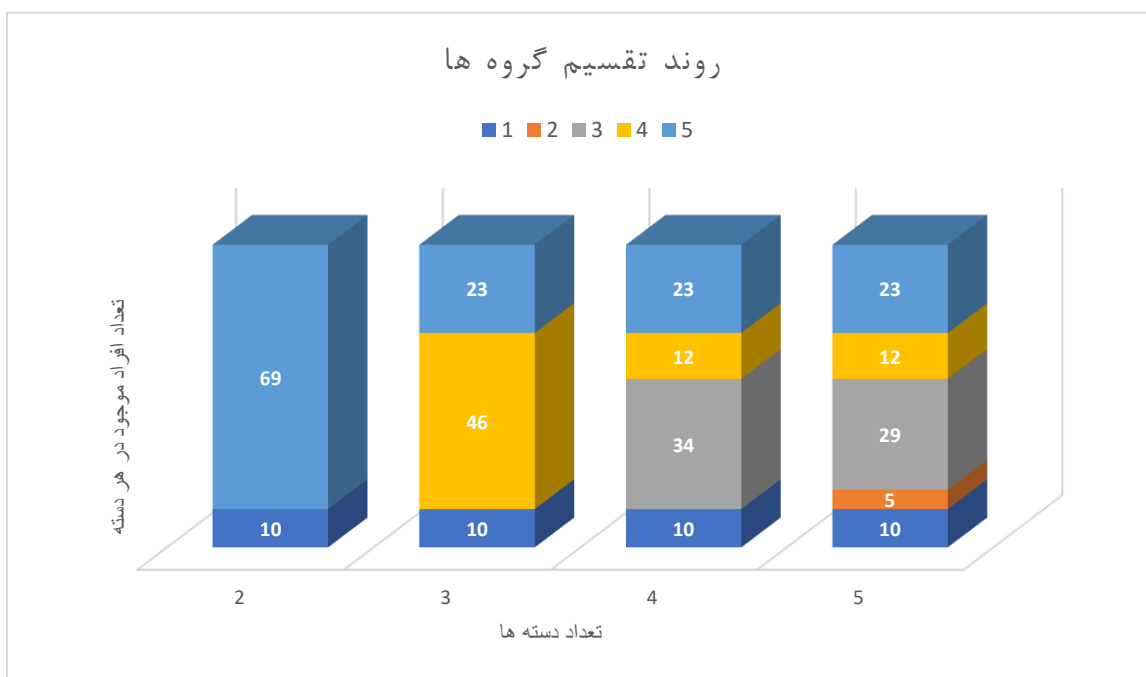
برای تمام دسته بندی ها مقدار p -value کمتر از 0.02 بوده و به دین معناست که پراکندگی داخل دسته ها کمتر از پراکندگی بین دسته هاست.

این ویژگی برای لیبل اصلی هم مقدار کمتر از 0.02 دارد.

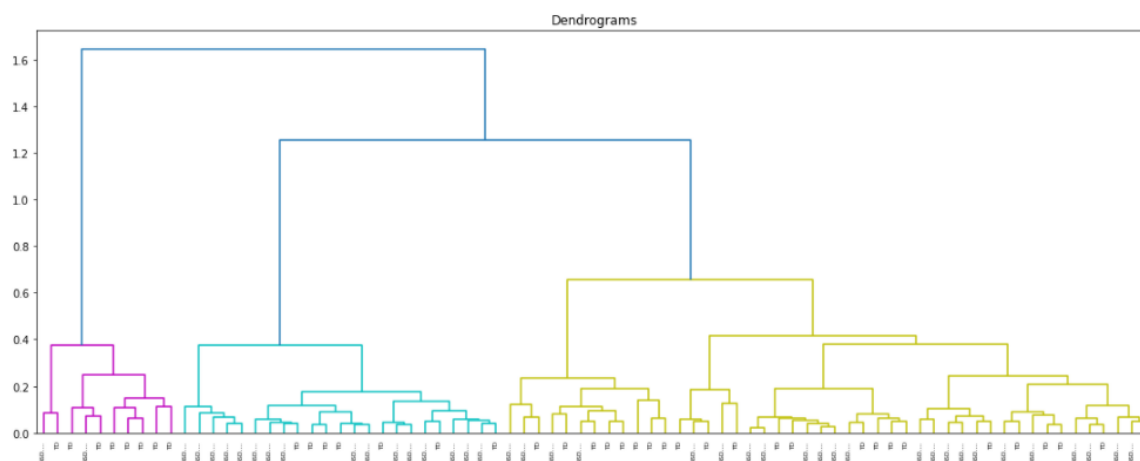
نتیجه : به نظر می رسد این ویژگی باعث میشود خوشه بندی درستی روی داده ها صورت گیرد و جز ویژگی های منتخب و برگزیده است.

4.2.2 خوشه بندی کانکتوم ها با استفاده از ویژگی های منتخب - ۲۵ ویژگی

با توجه به آنچه در قسمت قبل بررسی شد از میان ۴۴ ویژگی ذکر شده ۲۵ ویژگی وجود داشت که جز ویژگی های خوب ما طبقه بندی شده است حال میخواهیم بار دیگر ۷۹ کانکتوم مغزی موجود در مجموعه داده را اینبار با استفاده از این ویژگی ها همانند آنچه در قسمت قبل انجام شد خوشه بندی کرده و نتایج حاصل از آن را مورد ارزیابی قرار دهیم



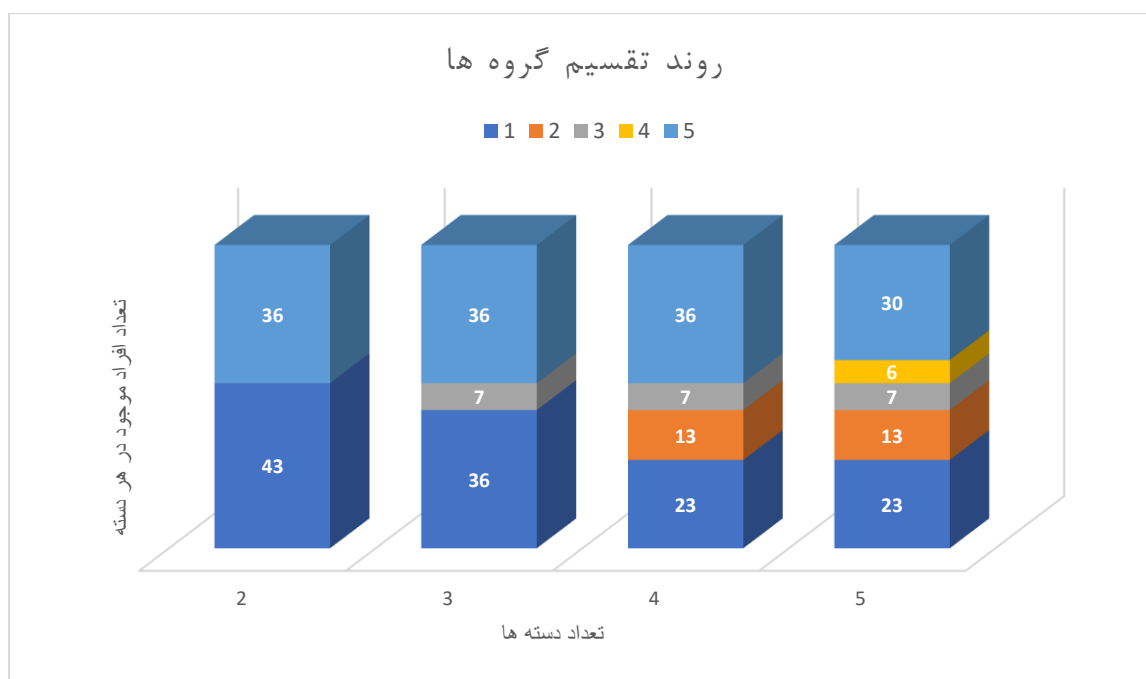
شکل (۳-۴) - بررسی روند دسته بندی داده ها با ویژگی های منتخب - ۲۵ ویژگی



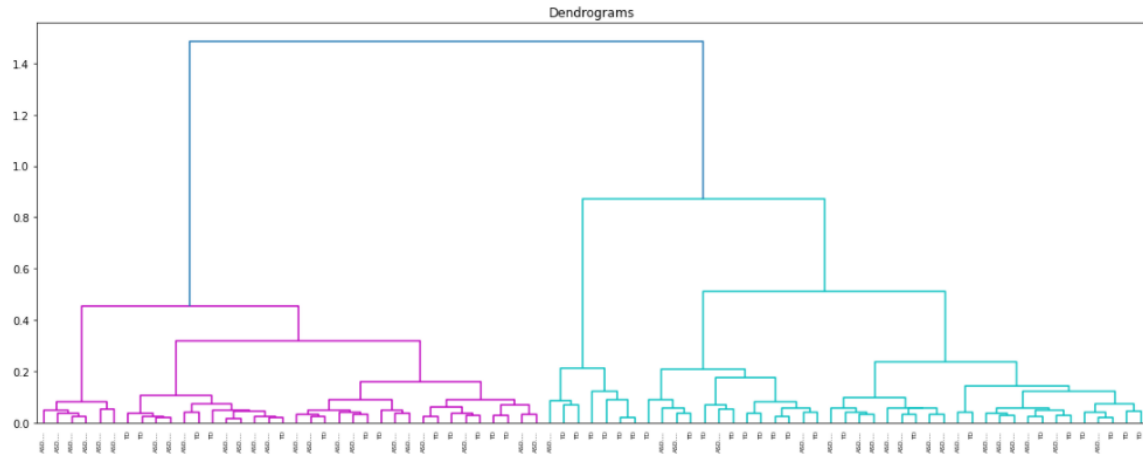
شکل (۴-۴) - دیاگرام مربوط به خوشه بندی با استفاده از ۲۵ ویژگی

4.2.3 خوشه بندی کانکتوم ها با استفاده از ویژگی های برگزیده - ۱۵ ویژگی

با توجه به آنچه در قسمت قبل بررسی شد از میان ۴۴ ویژگی ذکر شده ۱۵ ویژگی وجود داشت که جز ویژگی های خیلی خوب ما طبقه بندی شده است حال می خواهیم بار دیگر ۷۹ کانکتوم مغزی موجود در مجموعه داده را اینبار با استفاده از این ویژگی ها همانند آنچه در قسمت قبل انجام شد خوشه بندی کرده و نتایج حاصل از آن را مورد ارزیابی قرار دهیم می خواهیم بررسی کنیم آیا با انتخاب بهترین ویژگی ها می توان به خوشه بندی بهتری نسبت به آنچه تا به حال داشته ایم دست یافت.



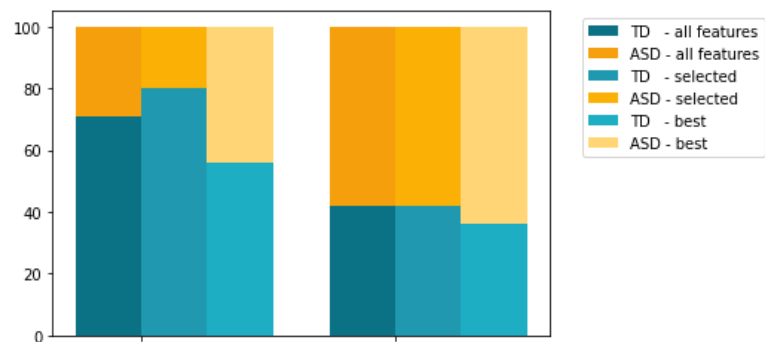
شکل (۴-۵) - بررسی روند دسته بندی داده ها با ویژگی های برگزیده - ۱۵ ویژگی



شکل (۴-۶) - دیاگرام مربوط به خوشه بندی با استفاده از ۱۵ ویژگی

4.2.4 مقایسه روند خوشه بندی‌ها در سه حالت مختلف

در این قسمت نحوه ی تغییر خوشه بندی‌ها را با کل ویژگی‌ها، ویژگی‌های منتخب و ویژگی های برگزیده نشان داده می شوند.

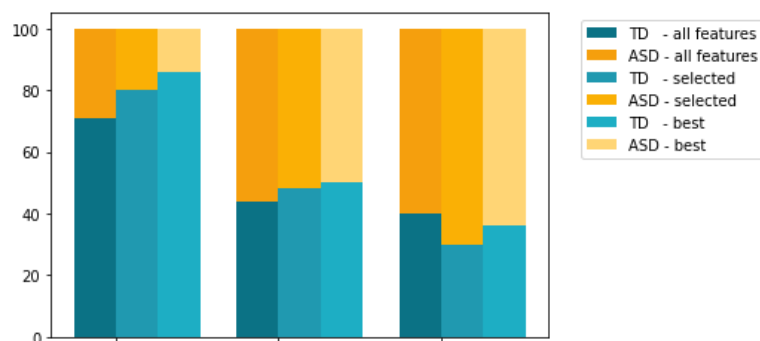


خوشه بندی بر اساس دو دسته

همانطور که در نمودار رو به رو نشان داده شده است در خوشه ی اول، بعد از خوشه بندی با ویژگی های منتخب، درصد افراد سالم افزایش میابد و مجدد با دسته بندی ویژگی های برگزیده کاهش را مشاهده میکنیم و در خوشه ی دوم این مقدار تقریباً ثابت میماند.

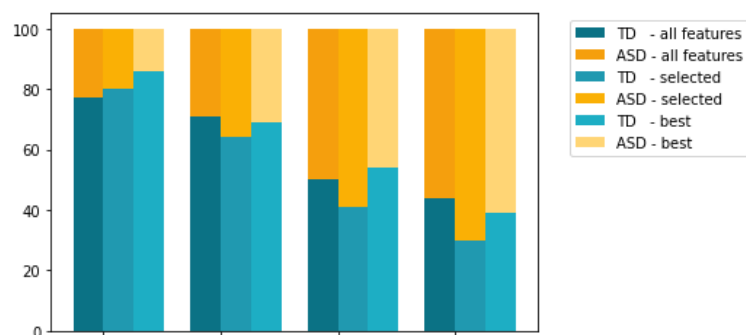
راهنمای نمودار: در هر خوشه میله ی سمت چپ نتیجه ی حاصل از تمام ویژگی ها، میله ی وسط نتیجه ی حاصل از ویژگی هایی است که با آستانه ی 0.05 روی p-value های حاصل از تحلیل واریانس انتخاب شده اند (ویژگی های منتخب) و ویژگی سمت راست نتیجه ی حاصل از ویژگی هایی است که با آستانه ی 0.02 انتخاب شده اند. (ویژگی های برگزیده)

خوشه بندی بر اساس سه دسته



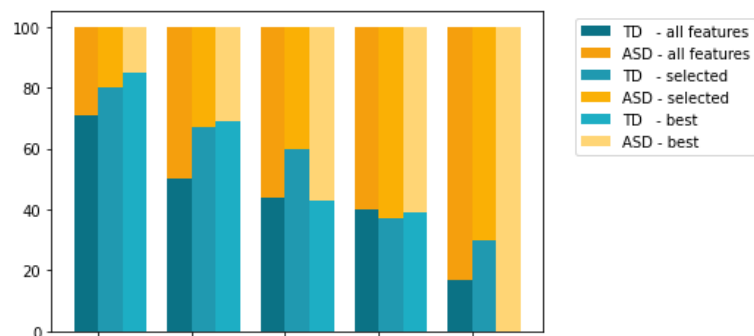
در نمودار رو به رو نشان داده شده است که در خوشه ی اول و دوم، بعد از خوشه بندی با ویژگی های منتخب و برگزیده، درصد افراد سالم افزایش میابد. و در خوشه ی دوم بعد از خوشه بندی با ویژگی های منتخب تعداد افراد اوتیسم افزایش میابد ولی با خوشه بندی با ویژگی های برگزیده تغییرات زیادی ایجاد نمی شود.

خوشه بندی بر اساس چهار دسته



در نمودار رو به رو نشان داده شده است که در خوشه ی اول، بعد از خوشه بندی با ویژگی های منتخب و برگزیده، درصد افراد سالم افزایش میابد. و در خوشه ی دوم تا چهارم بعد از خوشه بندی با ویژگی های منتخب تعداد افراد اوتیسم افزایش میابد ولی با ویژگی های منتخب تغییر محسوسی حس نمی شود.

خوشه بندی بر اساس پنج دسته

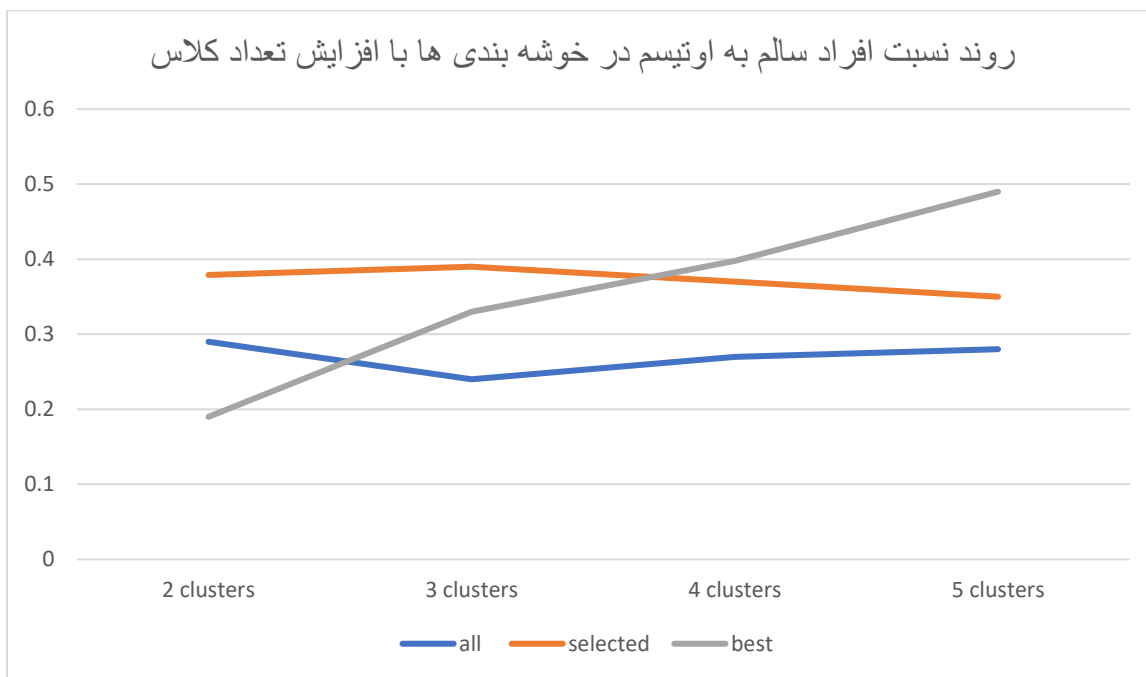


در نمودار رو به رو نشان داده شده است که در خوشه ی اول و دوم، بعد از خوشه بندی با ویژگی های منتخب و برگزیده، درصد افراد سالم افزایش میابد. و در خوشه ی سوم بعد از خوشه بندی با ویژگی های منتخب و برگزیده تغییر محسوسی حس نمی شود.

در خوشه سوم با خوشه بندی با ویژگی های منتخب تعداد افراد سالم رو به افزایش است.

در خوشه پنجم، بعد از خوشه بندی با ویژگی های منتخب تعداد افراد سالم در حال کاهش است اما با خوشه بندی با ویژگی های برگزیده خوشه به خوشه کاملاً اوتیسم تبدیل میشود.

4.2.5 تحلیل روند خوشه بندی براساس ویژگی‌ها



با توجه به نتایج به دست آمده مشاهده میشود استفاده از ویژگی های برگزیده⁴¹ دسته بندی های تفکیک شده بهتری را با افزایش تعداد خوشه ها در اختیار ما قرار می دهد.

همان طور که انتظار می رفت استفاده از تمام ویژگی ها بدون آگاهی از مناسب بودن یا نبودن آنها بدترین نتایج را داشته است و ویژگی های منتخب روند ثابتی را با افزایش تعداد کلاستر طی می کند.

خوشه بندی گراف های مغری با استفاده از ترشهلد 0.35

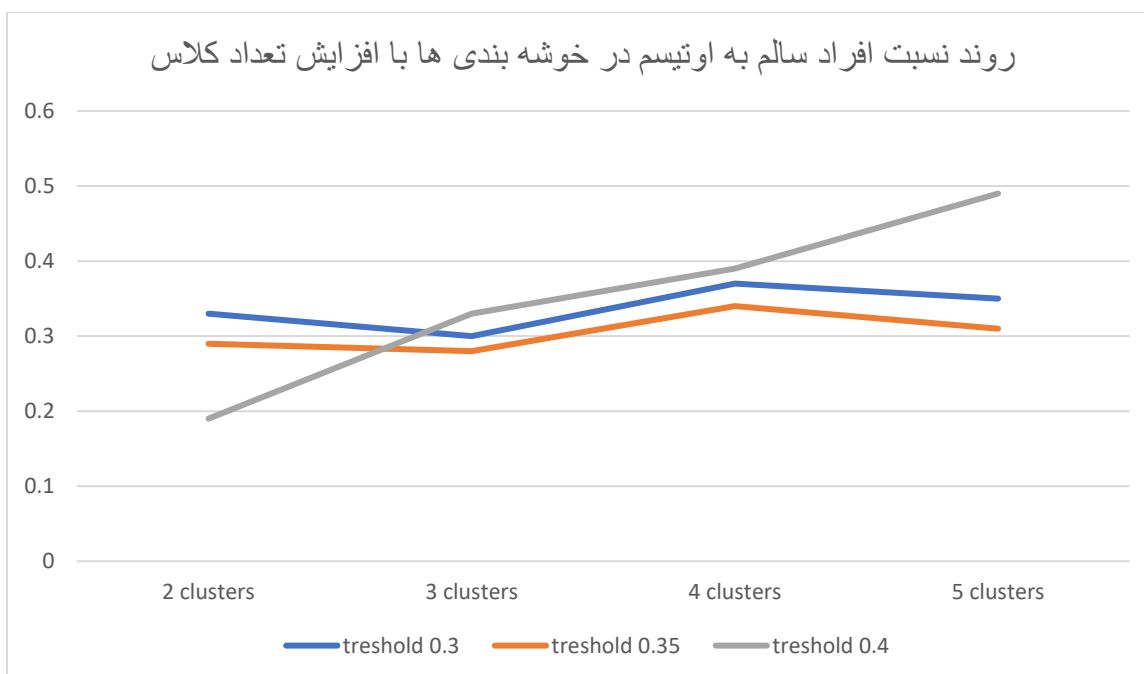
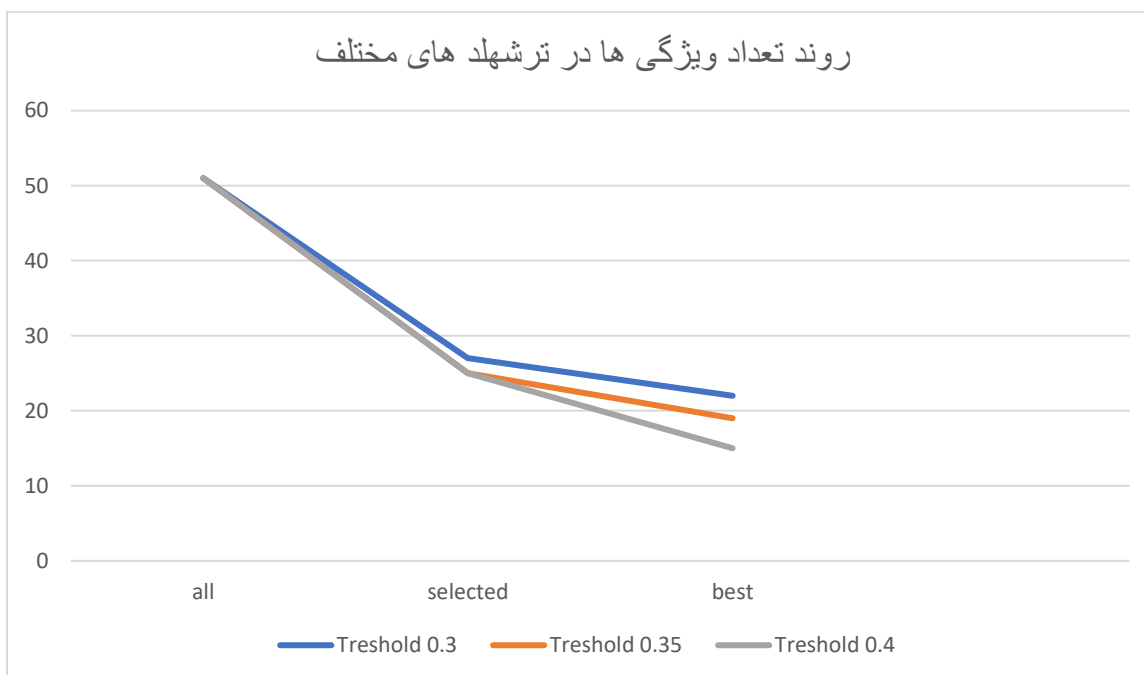
تمام آن چه در قسمت قبل بررسی شده است برای این ترشهلد نیز آورده شده است لطفا برای مشاهده فرایند و نتایج حاصل به بخش ۸.۳ ضمیمه ها مراجعه نمایید.

خوشه بندی گراف های مغری با استفاده از ترشهلد 0.3

تمام آن چه در قسمت قبل بررسی شده است برای این ترشهلد نیز آورده شده است لطفا برای مشاهده فرایند و نتایج حاصل به بخش ۸.۴ ضمیمه ها مراجعه نمایید.

⁴¹ best

4.2.6 بررسی ترشهلدهای مختلف



با افزایش میزان ترشهلد (از 0.3 به 0.4) یال هایی با مقدار وزن کمتر به مرور حذف می شود.

با کاهش میزان ترشه‌لد (از 0.4 به 0.3) یال‌هایی با مقدار وزن کمتر به گراف ما اضافه میشوند که در مجموع سبب افزایش تعداد یال‌های ما می‌شود.

می‌توان گفت در این حالت جزئیات بیشتری به گراف اضافه می‌شود در نتیجه ویژگی‌هایی که از اهمیت بالاتری برخوردارند تأثیر کمتری در خوشه‌بندی ما خواهند داشت. در نتیجه ما ترشه‌لدی را باید انتخاب کنیم که ویژگی‌های شاخص‌تر و به بیان دیگر مفیدتری در اختیار ما قرار دهد و ویژگی‌هایی که سبب برهم زدن یا خراب کردن خوشه‌بندی ما را دارد از ویژگی‌هایمان حذف شوند.

از اینجا به بعد تمام تحلیل‌ها بر روی ترشه‌لد 0.4 و با استفاده از ویژگی‌های منتخب (۲۵ ویژگی) صورت می‌گیرد.

فصل پنجم 5

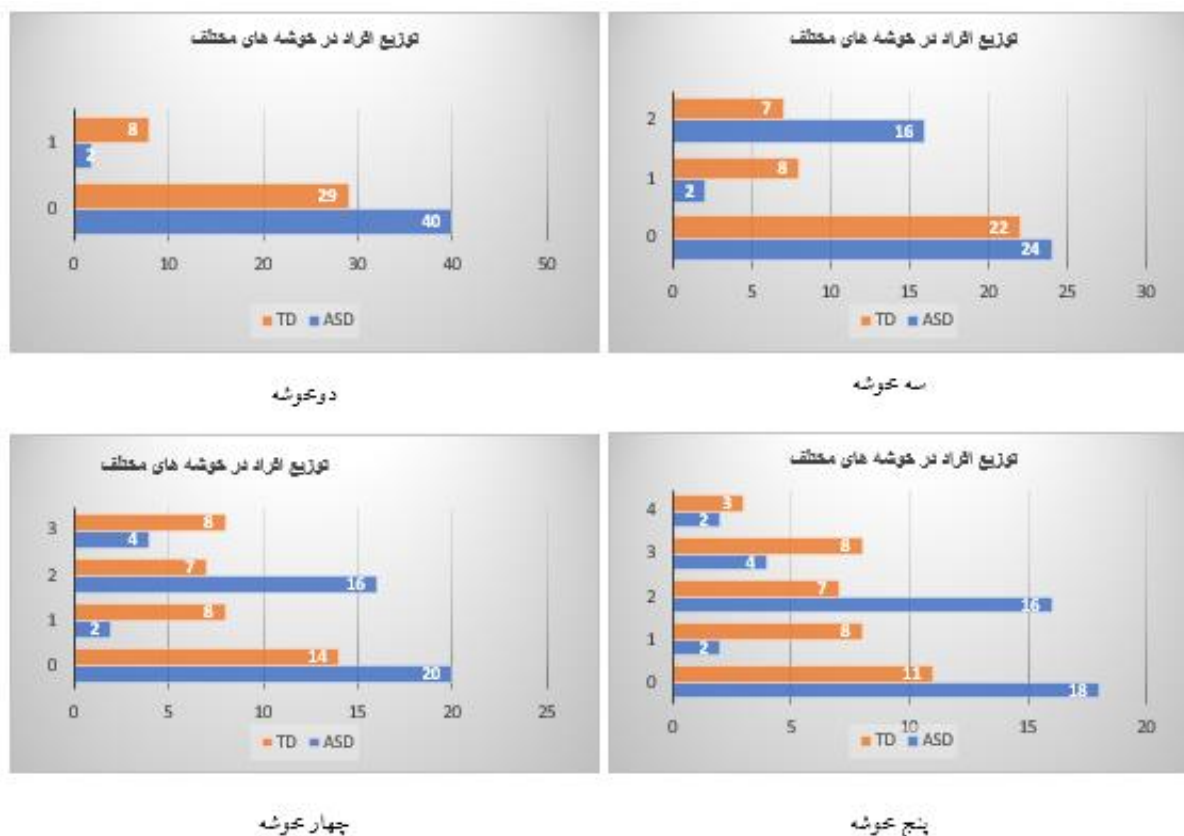
تحلیل و آنالیز

در این فصل تلاش بر این بوده که با توجه به نتایج به دست آمده از فصل قبل دسته‌بندی‌های مختلف را بررسی و نتایج را با یک دیگر مقایسه کنیم

5.1 بررسی خوشه‌بندی‌های حاصل از دسته‌بندی داده‌ها بدون دانستن برچسب‌های واقعی

تا به اینجا تحلیل و بررسی خوشه‌بندی‌ها با علم به دانستن برچسب‌های اصلی به جهت یافتن ویژگی‌های مناسب‌تر برای خوشه‌بندی بوده است. سوالی که پیش می‌آید آیا راه حلی وجود دارد که پس از دسته‌بندی کانکتوم‌های مختلف، بتوان تشخیص داد کدام یک از دسته‌های حاصل نسبت به بقیه دسته‌ها بهتر دسته‌بندی شده‌اند (نسبت افراد سالم به افراد ناسالم در آنها قابل توجه است و یا برعکس)

در ابتدا نگاهی به نحوه‌ی توزیع افراد سالم و اوتیسم در دسته‌بندی‌های دو تا پنج تایی می‌اندازیم:



شکل (۱-۵) - پراکندگی افراد سالم و اوتیسم در خوشه ها از نظر تعداد

5.1.1 بررسی گروه‌هایی شاخص

باید توجه داشت در زمان استفاده از انووا زمانی که ما به $p-value$ پایین (کمتر از 0.05) دست پیدا می‌کنیم بدین معناست که در میان دسته‌های موجود حداقل یک دسته وجود دارد که پراکندگی درون گروهی آن از پراکندگی بین گروهی آن‌ها کمتر است. در نتیجه ما در آزمون $post-hoc$ به دنبال خوشه‌هایی هستیم که پراکندگی درون گروهی آن‌ها نسبت به پراکندگی بین گروهی‌شان بیشتر بوده است. هر گروهی که در تعداد ویژگی بیشتری نسبت به مابقی گروه‌ها بهتر جدا شده باشد را ما به عنوان گروه بهتر معرفی خواهیم کرد. این آزمون بدین صورت عمل می‌کند که بر روی یک ویژگی خاص دو به دو گروه‌ها را بررسی کرده و نشان می‌دهد که آیا به واسطه این ویژگی این دو گروه تفکیک معناداری از هم داشته‌اند یا خیر.

در ادامه روند بررسی خوشه بندی سه تا پنج تایی را به اختصار بیان خواهیم کرد.

5.1.1.1 بررسی برای خوشه بندی سه تایی

	Label 0	Label 1	Label 2
جمع	4	4	0
تعداد گروه های درگیر	4	4	0

همان طور که مشاهده می شود خوشه با لیبل های دو بهترین تفکیک را نسبت به مابقی گروه ها داشته اند و گروه با لیبل صفر و یک اصلا به درستی تفکیک نشده است. اگر لیبل های اصلی این سه گروه را مشاهده کنیم گروه دو دارای شانزده

فرد ناسالم و هفت فرد سالم و گروه صفر دارای بیست و چهار فرد ناسالم و بیست و دو فرد سالم است، و گروه یک دارای دو فرد سالم و هشت فرد ناسالم می باشد.

5.1.1.2 بررسی برای خوشه بندی 4 تایی

	Label 0	Label 1	Label 2	Label 3
جمع	3	11	1	11
تعداد گروه های درگیر	4	12	1	13

همان طور که مشاهده می شود خوشه با لیبل های دو بهترین تفکیک را نسبت به مابقی گروه ها داشته اند و گروه با لیبل سه و یک اصلا به درستی تفکیک نشده است. اگر لیبل های اصلی این سه گروه را مشاهده

کنیم گروه دو دارای شانزده فرد ناسالم و هفت فرد سالم و گروه سه دارای چهار فرد ناسالم و هشت فرد سالم است، و گروه یک دارای دو فرد سالم و هشت فرد ناسالم می باشد.

5.1.1.3 بررسی برای خوشه بندی 5 تایی

	Label 0	Label 1	Label 2	Label 3	Label 4
0			x		x
1			x		x
2	x	x		x	x
3	x		x		x
4	x	x		x	x
5	x	x		x	x
6	x		x		x
7	x		x		x
8	x		x		x
9	x	x	x	x	x
10	x	x	x	x	x
11	x		x	x	x
12	x	x	x	x	x
13	x	x	x		x
14			x		x
15			x		x
16	x	x	x	x	x
17	x		x		x
18	x	x	x	x	x
19			x		x
20			x		x
21	x	x	x	x	x
22	x		x		x
23	x	x	x	x	x
24	x		x		x
جمع	19	11	22	11	25
تعداد گروه های درگیر	23	17	24	23	55

همان طور که مشاهده می شود خوشه با لیبل های یک و سه بهترین تفکیک را نسبت به مابقی گروه ها داشته اند. و گروه با لیبل چهار اصلا به درستی تفکیک نشده است. اگر لیبل های اصلی این سه گروه را مشاهده کنیم گروه یک دارای دو فرد ناسالم و هشت فرد سالم و گروه سه دارای چهار فرد ناسالم و هشت فرد سالم است، و گروه پنجم که بدترین تفکیک را داشته است دارای دو فرد سالم و سه فرد ناسالم می باشد.

5.1.1.4 بررسی برای خوشه بندی 6 تایی

	Label 0	Label 1	Label 2	Label3	Label4	Label5
جمع	11	25	24	16	25	25
تعداد گروه های درگیر	21	53	31	26	75	52

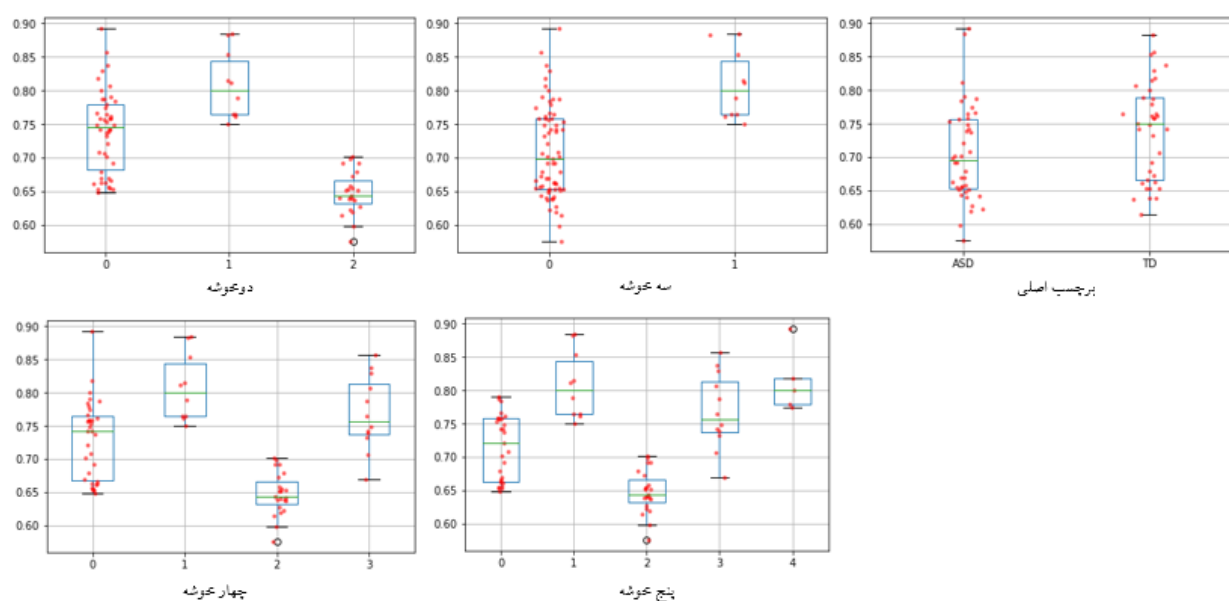
گروه 2 : 0 ناسالم ، 8 سالم گروه 5 : 1 ناسالم ، 12 سالم گروه 7 : 2 ناسالم ، 16 سالم

گروه 8 : 3 ناسالم ، 4 سالم گروه 3 : 4 ناسالم ، 2 سالم گروه 6 : 5 ناسالم ، 6 سالم

به نظر میرسد اتکا به نتایج این آزمون **تأحیدی** قابل قبول می باشد مخصوصاً زمانی که تعداد دسته ها رو به افزایش است.

5.2 تحلیل های کیفی

5.2.1 ماکسیمم ضریب خوشگی

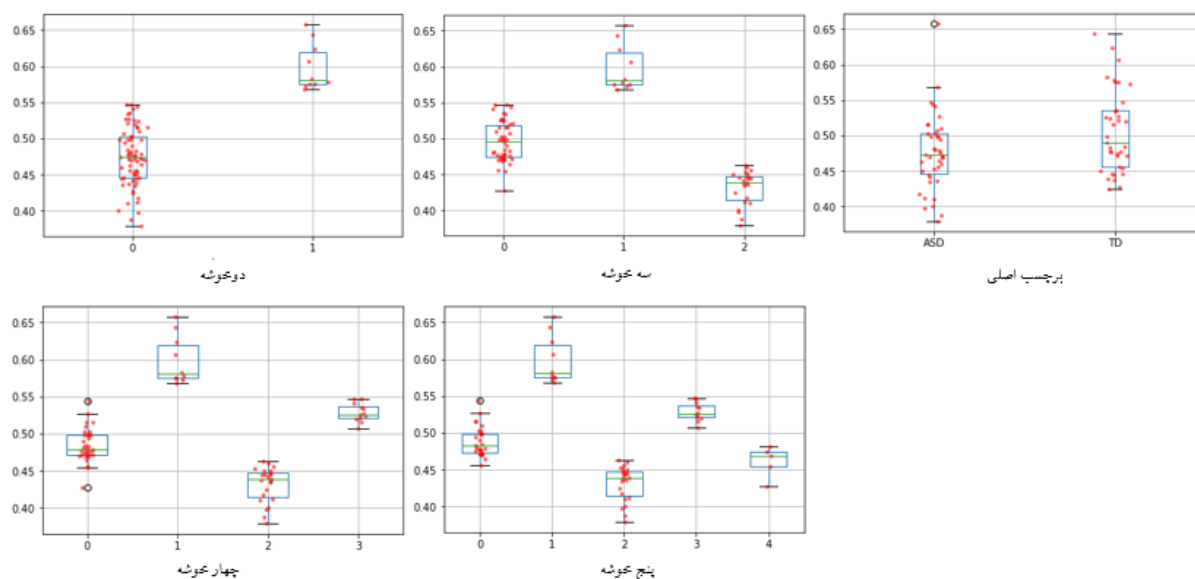


به طور تقریبی می توان گفت ماکزیمم ضریب خوشگی افراد سالم به نسبت افراد اوتیسم بیشتر است.

گروه هایی با برچسب دو که تا حدی نماینده اوتیسم هستند نسبت به گروه های یک و سه که نماینده افراد سالم در این دسته بندی ما هستند ماکزیمم ضریب خوشگی کمتری نسبت به دو گروه ذکر شده دارد (به خوشه بندی پنج تایی نگاه کنید).

گروه 4 که افراد سالم در آن بیشتر از ناسالم است بیشتر به سمت بالا و عکس این قضیه برای گروه صفر صادق است بیشتر به سمت پایین است.

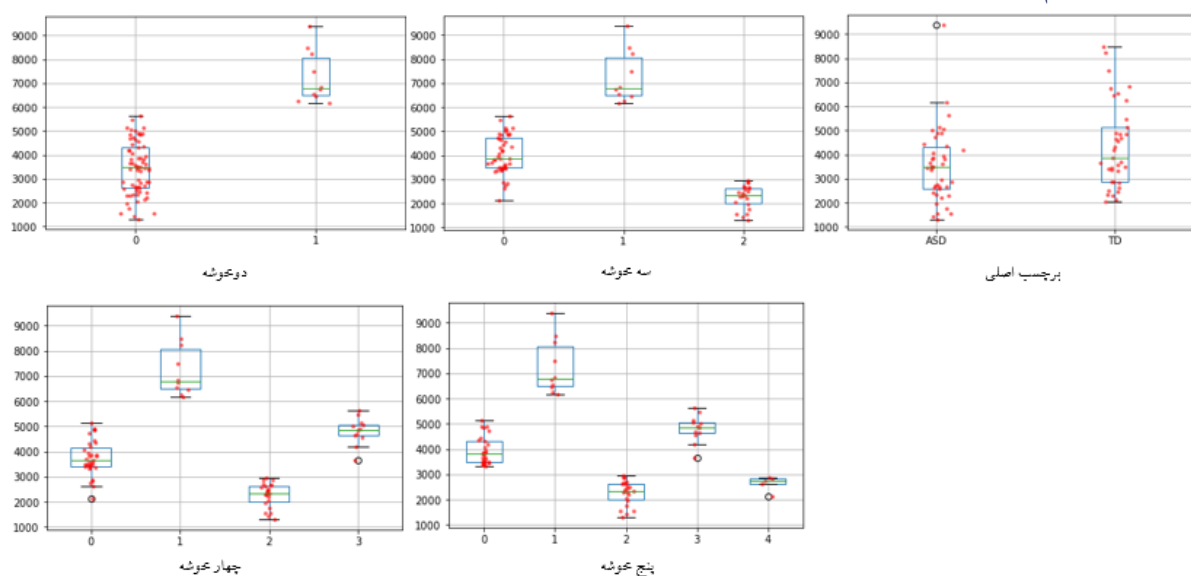
5.2.2 میانه ضریب خوشگی



به طور تقریبی می توان گفت میانه ضریب خوشگی افراد سالم به نسبت افراد اوتیسم کمی بیشتر است ولی پراکندگی بسیار مشابهی دارند در این حالت.

گروه هایی با لیبیل دو که تا حدی نماینده اوتیسم هستند نسبت به گروه های یک و سه که نماینده افراد سالم در این دسته بندی ما هستند میانه ضریب خوشگی کمتری نسبت به دو گروه ذکر شده دارد (به خوشه بندی پنج تایی نگاه کنید).

5.2.3 ماکسیمم مثلث ها



1

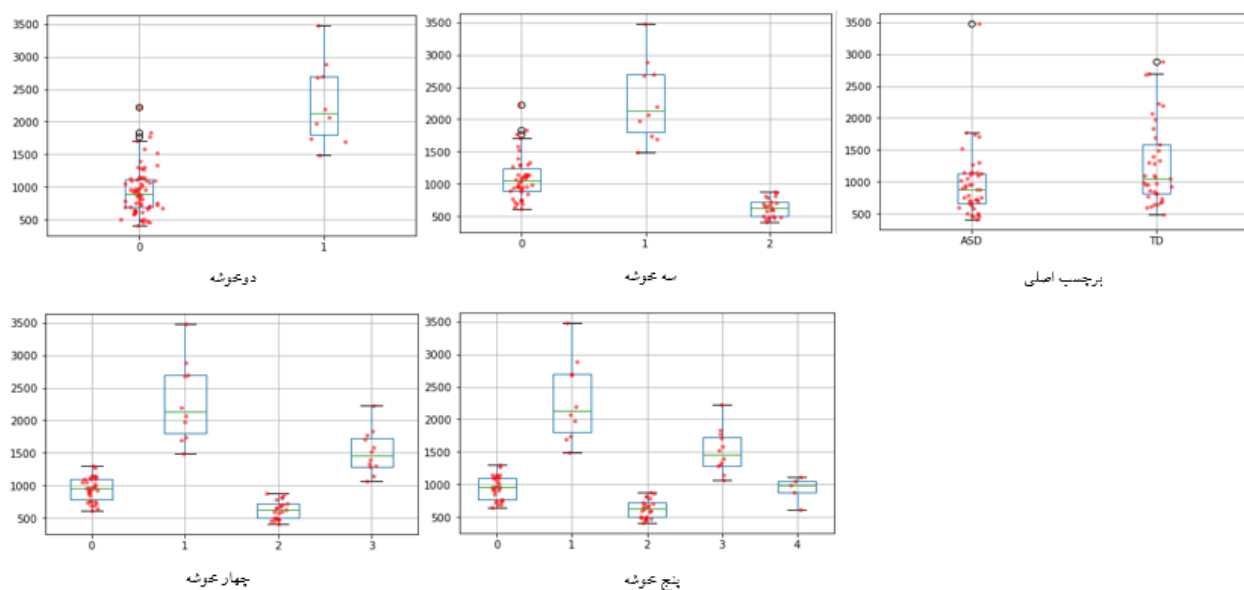
به طور تقریبی می توان گفت ماکزیمم تعداد مثلث ها روی هر نود افراد سالم به نسبت افراد اوتیسم کمی بیشتر است ولی پراکندگی

بسیار مشابهی دارند در این حالت.

گروه‌هایی با لیبل دو که تا حدی نماینده اوتیسم هستند نسبت به گروه‌های یک و سه که نماینده افراد سالم در این دسته بندی ما هستند ماکزیمم تعداد مثلث‌ها کمتری نسبت به دو گروه ذکر شده دارد (به خوشه بندی پنج تایی نگاه کنید).

گروه 4 و 0 در بین هر دو جعبه اوتیسم و سالم قرار دارند.

5.2.4 میانه مثلث‌ها



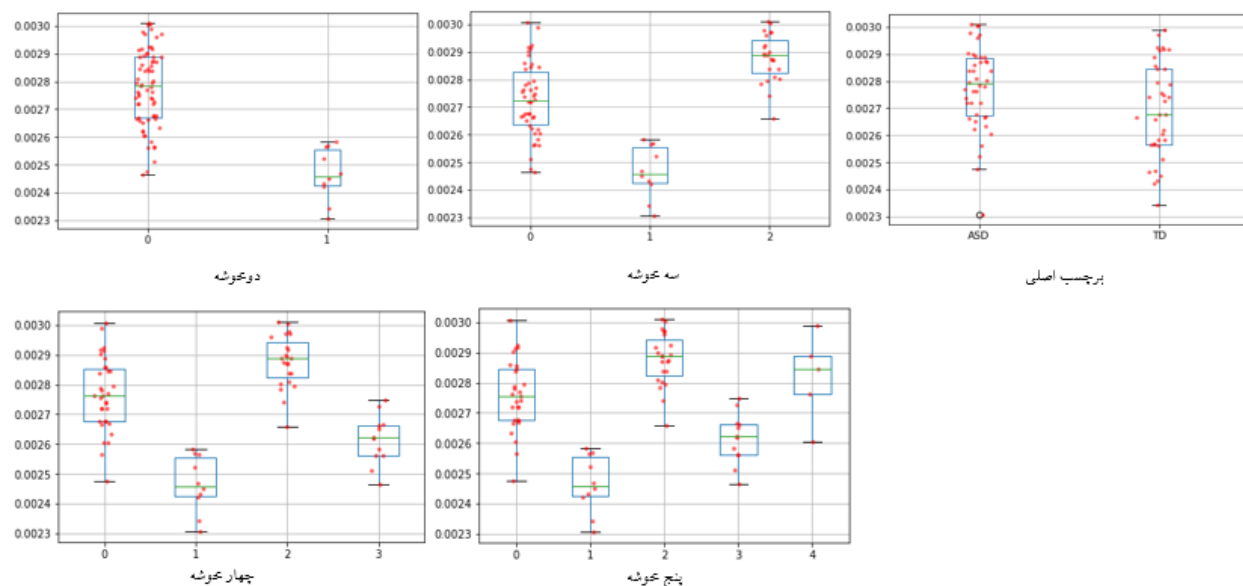
2

به طور تقریبی می‌توان گفت میانه تعداد مثلث‌ها روی هر نود افراد سالم به نسبت افراد اوتیسم کمی بیشتر است ولی طیف پراکندگی سالم بیشتر از اوتیسم می‌باشد ولی میانه‌هایی بسیار نزدیک به هم دارند

گروه‌هایی با لیبل دو که تا حدی نماینده اوتیسم هستند نسبت به گروه‌های یک و سه که نماینده افراد سالم در این دسته‌بندی ما هستند میانه تعداد مثلث‌ها کمتری نسبت به دو گروه ذکر شده دارد (به خوشه بندی پنج تایی نگاه کنید).

گروه 4 و 0 در بین هر دو جعبه اوتیسم و سالم قرار دارند.

5.2.5 میانه مرکزیت بینابینی



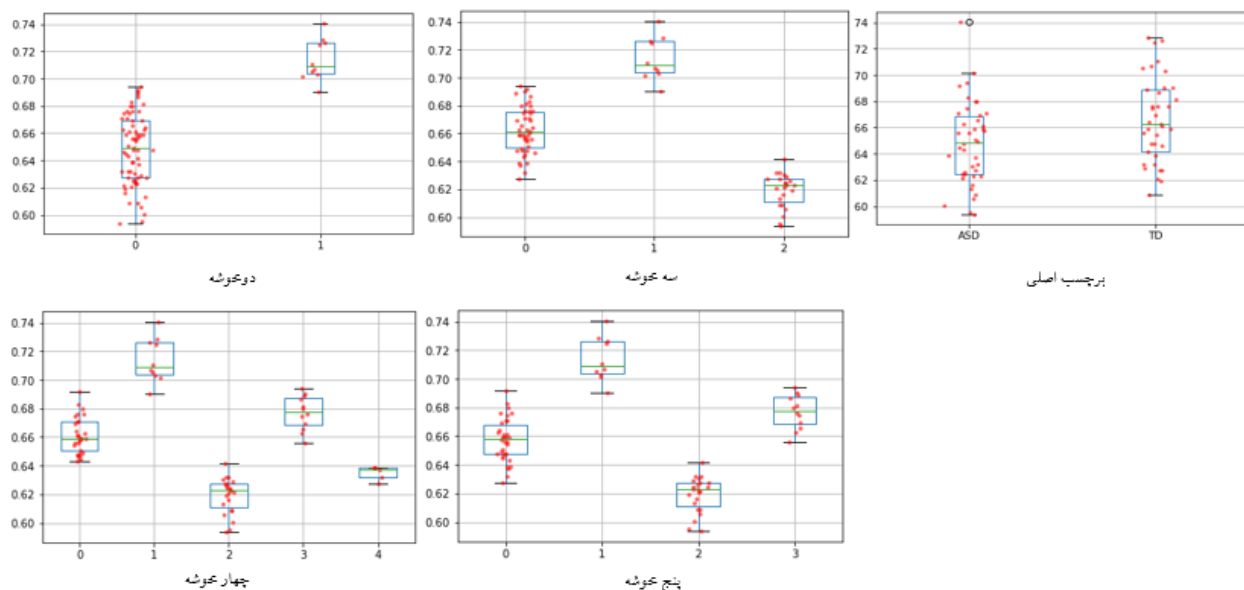
3

به طور تقریبی می‌توان گفت میانه مرکزیت بینابینی افراد سالم به نسبت افراد اوتیسم کمی کمتر است ولی طیف پراکندگی سالم بیشتر از اوتیسم می‌باشد.

گروه‌هایی با لیبیل 2 که تا حدی نماینده اوتیسم هستند نسبت به گروه‌های یک و سه که نماینده افراد سالم در این دسته بندی ما هستند میانه مرکزیت بینابینی بیشتری نسبت به دو گروه ذکر شده دارد (به خوشه بندی پنج تایی نگاه کنید).

گروه 0 و 4 در بین هر دو جعبه اوتیسم و سالم قرار دارند ولی بیشتر به اوتیسم نزدیک اند.

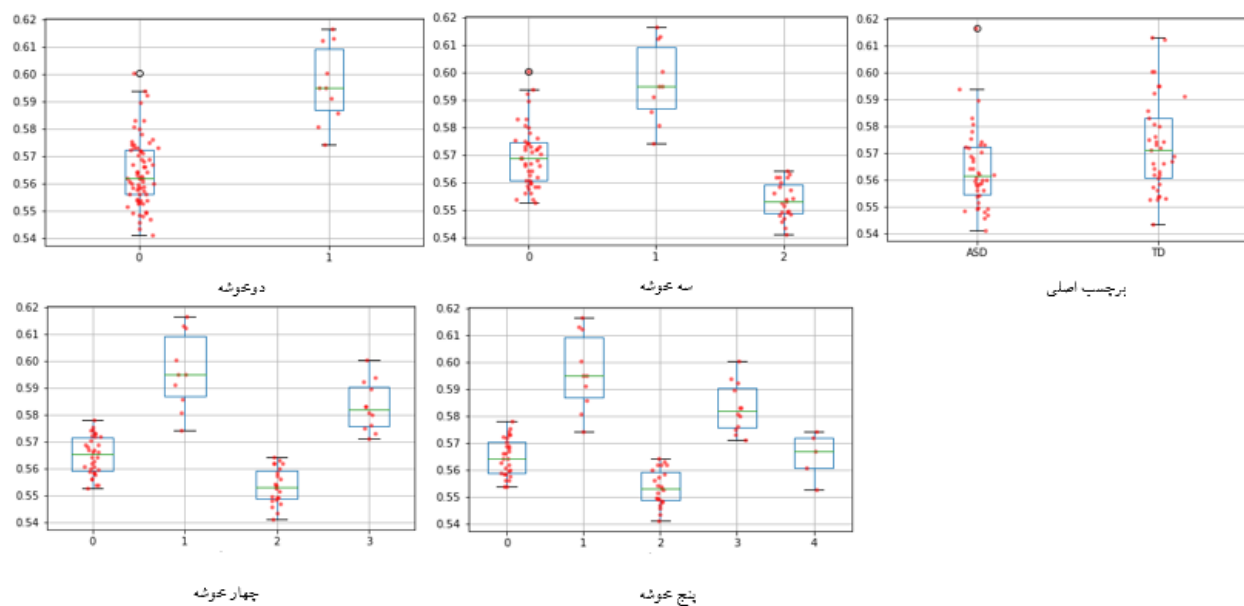
5.2.6 ماکسیمم مرکزیت بر مبنای نزدیکی



به طور تقریبی می‌توان گفت ماکسیمم مرکزیت بر مبنای نزدیکی افراد سالم به نسبت افراد اوتیسم کمی بیشتر است ولی میانه‌هایی بسیار نزدیک به هم دارند.

گروه‌هایی با لیبیل 2 که تا حدی نماینده اوتیسم هستند نسبت به گروه‌های 1 و 3 که نماینده افراد سالم در این دسته بندی ما هستند ماکسیمم مرکزیت بر مبنای نزدیکی کمتری نسبت به دو گروه ذکر شده دارد (به خوشه بندی پنج تایی نگاه کنید). گروه 4 و 0 نیز بین سه گروه قبل قرار دارند.

5.2.7 میانه مرکزیت بر مبنای نزدیکی

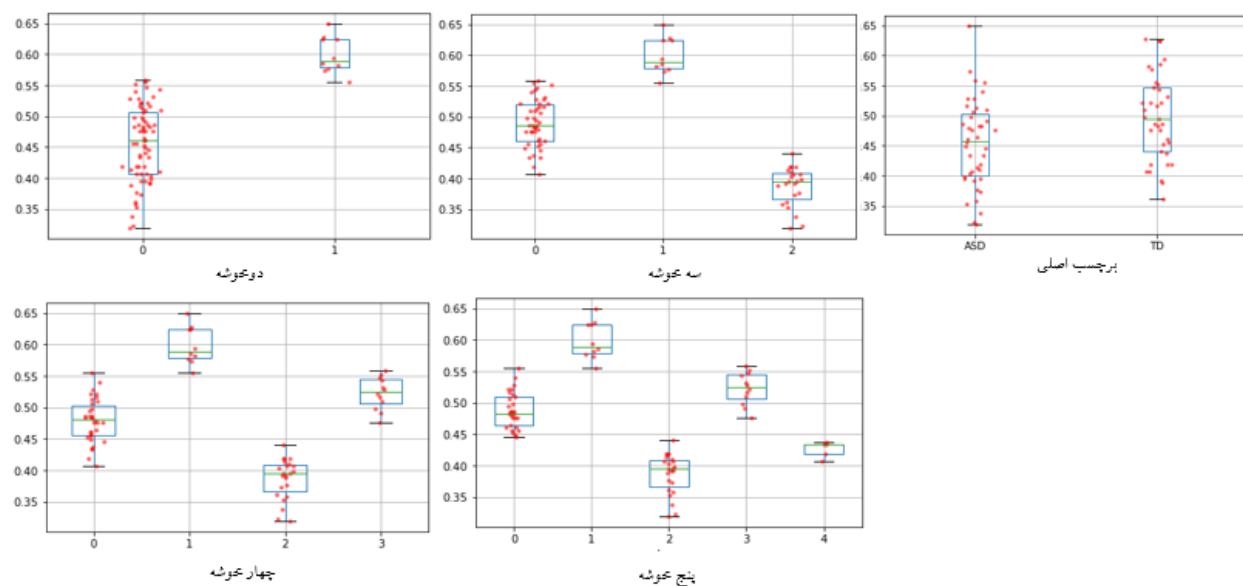


به طور تقریبی می توان گفت میانه مرکزیت بر مبنای نزدیکی افراد سالم به نسبت افراد اوتیسم بیشتر است.

گروه هایی با لیبل 2 که تا حدی نماینده اوتیسم هستند نسبت به گروه های 1 و 3 که نماینده افراد سالم در این دسته بندی ما هستند میانه مرکزیت بر مبنای نزدیکی کمتری نسبت به دو گروه ذکر شده دارد (به خوشه بندی پنج تایی نگاه کنید).

گروه 4 که نسبت سالم به اوتیسم بیشتری دارد بیشتر به سمت بالا و گروه 9 بیشتر به سمت پایین قرار دارد (به خوشه بندی پنج و میانه های گروه 4 و 0 نگاه کنید)

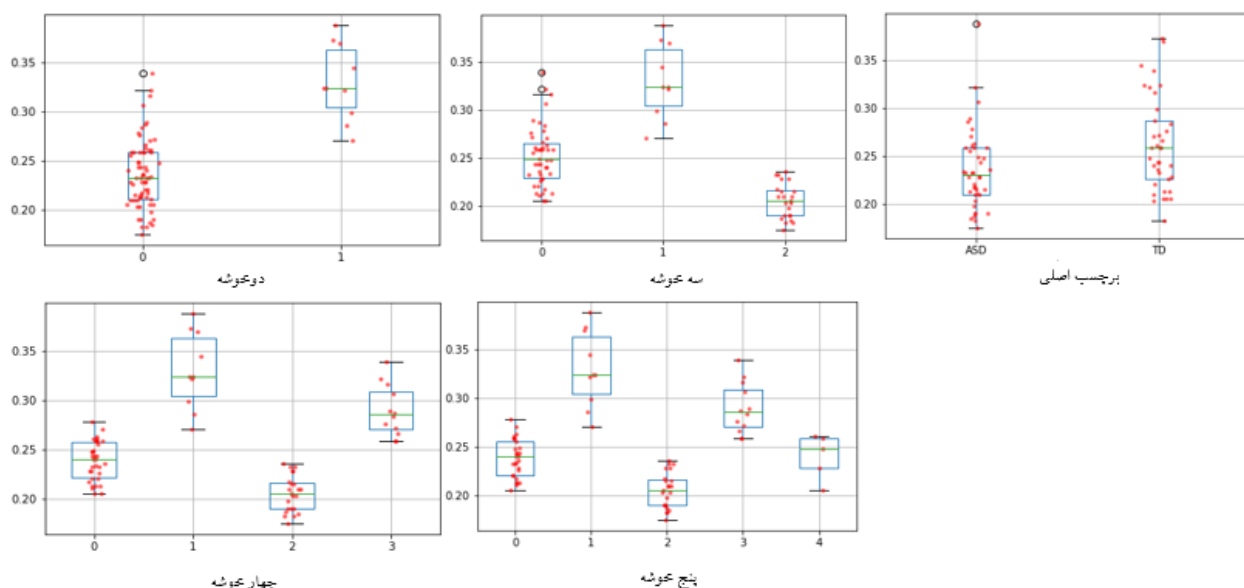
5.2.8 ماکسیمم مرکزیت بر مبنای درجه



4

به طور تقریبی می‌توان گفت ماکسیمم مرکزیت بر مبنای درجه افراد سالم به نسبت افراد اوتیسمم بیشتر است. گروه‌هایی با لیبیل 2 که تا حدی نماینده اوتیسمم هستند نسبت به گروه‌های 1 و 3 که نماینده افراد سالم در این دسته‌بندی ما هستند ماکسیمم مرکزیت بر مبنای درجه کمتری نسبت به دو گروه ذکر شده دارد (به خوشه بندی پنج تایی نگاه کنید).

5.2.9 میانه مرکزیت بر مبنای درجه



به طور تقریبی می توان گفت ماکزیمم مرکزیت بر مبنای درجه افراد سالم به نسبت افراد اوتیسم بیشتر است.

گروه هایی با لیبل 2 که تا حدی نماینده اوتیسم هستند نسبت به گروه های 1 و 3 که نماینده افراد سالم در این دسته بندی ما هستند ماکزیمم مرکزیت بر مبنای درجه کمتری نسبت به دو گروه ذکر شده دارد (به خوشه بندی پنج تایی نگاه کنید).

گروه 4 که نسبت سالم به اوتیسم بیشتری دارد بیشتر به سمت بالا و گروه 9 بیشتر به سمت پایین قرار دارد (به خوشه بندی پنج و میانه های گروه 4 و 0 نگاه کنید)

5.2.10 نتیجه گیری

برای برچسب های اصلی در تمام ویژگی های ذکر شده در قسمت قبل باید در نظر داشت که هیچ یک از جعبه ها به طور کامل از یک دیگر جدا نشده اند ولی می توان دید که به طور تقریبی مقدار مقادیر افراد سالم در آن ویژگی ها به نسبت اوتیسم بیشتر بوده (جعبه در پوزیشن بالاتری قرار می گرفت) به جز میانه ویژگی مرکزیت بینابینی که جعبه افرتد اوتیسم بالاتر از افراد سالم قرار داشت که بدان معنی است که مقادیر این ویژگی در افراد اوتیسم به نسبت سالم بیشتر است.

برای خوشه بندی های صورت گرفته این نتایج مشاهده می شود:

در صورتی که اختلاف یکی از افراد سالم به ناسالم و برعکس مقدار قابل توجهی باشد (مثلا 1 به 2 و بیشتر) در این صورت از روند لیبل های اصلی پیروی می کند مثلا اگر در لیبل های اصلی جعبه سالم بالاتر از اوتیسم قرار دارد در این حالت نیز خوشه ای که تعداد افراد سالم به اوتیسم آن بیشتر است نیز، بالاتر قرار می گیرد.

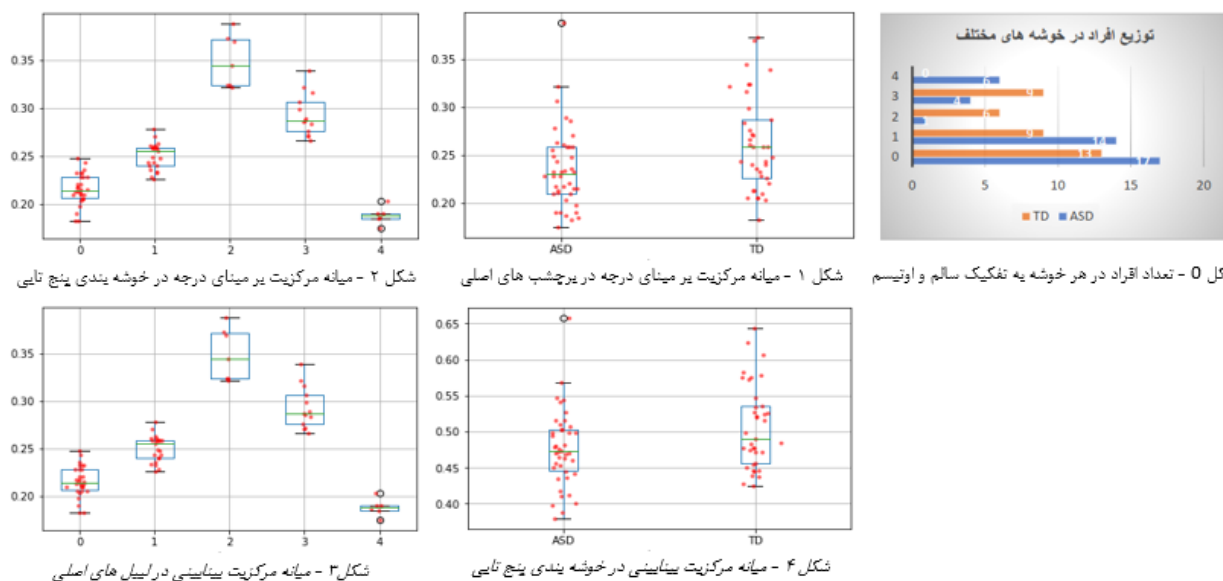
اما اگر اختلاف افراد سالم و ناسالم تا حدی نزدیک به یک دیگر باشد دو حالت پیش می آید:

حالت اول : اگر نمودار جعبه ای لیبل های اصلی نیز تا حدی از هم جدا شده باشند (اختلاف میانه ها بیشتر از 33 درصد از طول

جعبه باشد) گروهی که تعداد افراد سالم آن بیشتر از افراد ناسالم است از روند نمودار جعبه ای افراد سالم در لیبل های اصلی پیروی می کند (یعنی اگر افراد سالم مثلا پایین تر از افراد ناسالم قرار گرفته اند در این خوشه بندی نیز گروه مدنظر به سمت پایین متمایل تر است و جعبه گروه افراد ناسالم بیشتر به سمت بالا است).

حالت دوم: نمودار جعبه ای لیبل های اصلی نیز از یک دیگر به خوبی جدا نشده اند و میانه هایی نزدیک به یک دیگر دارند (اختلاف میانه ها کمتر از 33 درصد از طول جعبه باشد) در این حالت از وضعیت جعبه ها از قانون خاصی تبعیت نمی کنند.

برای اثبات آنچه گفته شد به نمودار های زیر توجه نمایید:



همان طور که در شکل ۱ مشاهده می کنید میانه جعبه افراد سالم از ماکزیمم جعبه افراد اوتیسم بیشتر است و این بدان معناست که تا حدی این دو جعبه از یک دیگر تفکیک شده اند. در شکل ۲ گروه با برچسب ۴ نماینده بیماران اوتیسم می باشد و گروه با برچسب ۲ نماینده افراد سالم می باشد (به شکل ۰ نگاه کنید). گروه ۲ بیشترین اختلاف را با گروه ۴ دارد. گروه ۳ که تعداد افراد سالم در آن بیشتر از افراد اوتیسم است به جعبه شماره ۲ نزدیک تر است و جعبه شماره ۰ و ۱ به جعبه شماره ۴ نزدیک تر است چون تعداد افراد اوتیسم در این دو گروه به نسبت افراد سالم بیشتر است. حال با مقایسه دو جعبه ۰ و ۱ می توان به این نتیجه رسید که چون تعداد افراد اوتیسم در گروه ۰ بیشتر از یک است در نتیجه در موقعیت پایین تری به نسبت جعبه شماره ۱ قرار دارد. همان طور که در شکل ۳ می بینید میانه های دو جعبه بسیار به هم نزدیک هستند در نتیجه نمی توان برای گروه هایی مثل گروه ۱ دز شکل ۴ اظهار نظر قطعی نمود.

- 1- به طور کلی میانه مثلث های روی هر نود در افراد سالم بیشتر از افراد مبتلا به اوتیسم می باشد.
- 2- به طور کلی ماکزیمم ضریب خوشگی در افراد سالم بیشتر از افراد مبتلا به اوتیسم می باشد.
- 3- به طور کلی میانه مرکزیت بر مبنای درجه در افراد سالم بیشتر از افراد مبتلا به اوتیسم می باشد.
- 4- به طور کلی میانه مرکزیت بینابینی در افراد سالم کمتر از افراد مبتلا به اوتیسم می باشد.
- 5- به طور کلی میانه مرکزیت بر مبنای نزدیکی در افراد سالم بیشتر از افراد مبتلا به اوتیسم می باشد.

5.3 مقایسه و تحلیل دو روش خوشه بندی متفاوت

در پژوهش دیگری در همین زمینه هم زمان با انجام این پروژه صورت گرفت از الگوریتم خوشه بندی متفاوتی استفاده شده است که مبتنی بر مدل ترکیبی گوسی است در ادامه ترکیب و مقایسه نتایج هر دو روش خوشه بندی آورده خواهد شد.

به منظور مقایسه دو روش در ابتدا باید دانست که هر دو روش روی مجموعه داده ای موجود تا حدی دقت برابری را نتیجه می دهند لذا نمی توان به طور دقیق گفت کدام روش ها بهتر یا بدتر عمل می کنند. اما با توجه به نتایج به دست آمده در یکی از خوشه بندی ها مدل ترکیبی گوسی توانسته به اختلاف 100 درصدی برای اوتیسم برسد که بدین معناست که تمام افراد حاضر در یکی از گروه های خوشه بندی ها تمام افراد حاضر اوتیستیک هستند. و بهترین درصد هم برای خوشه بندی افراد سالم با استفاده از این روش 80% می باشد. در روش خوشه بندی سلسله مراتبی اعداد بالا به ترتیب 70% و 80% می باشد که از این حیث تا حدی الگوریتم ترکیبی گوسی بهتر عمل کرده است.

حال از این دو روش چه استفاده هایی می توان کرد. یکی از راه کار ها هنگام تشخیص برچسب کانکتوم های مغزی افراد می باشد بدین صورت که داده جدید به همراه داده های قبلی مجدد خوشه بندی می شود با استفاده از هر دو روش، اگر داده جدید در هر دو روش به دسته ای تعلق گرفت که آن دسته به گونه ای نماینده ی افراد اوتیسم باشد (و برعکس) با احتمال خوبی می توان گفت که داده جدید نیز احتمالاً متعلق به فرد مبتلا به اوتیسم است (و برعکس)

فصل ششم 6

جمع‌بندی و نتیجه‌گیری

هدف اصلی این پروژه بررسی ویژگی‌های مختلف و میزان خوب عملکرد آن‌ها در خوشه بندی بوده است.

از نتایج ذکر شده در پروژه می‌توان پی‌برد که به دلیل گسترده بودن طیف بیماران اوتیسم و همچنین وجود دیگر اختلالات که معمولاً افراد مبتلا به اوتیسم با آن‌ها نیز درگیرند سبب شده ویژگی‌های استخراج شده از گراف مغزی انسان‌های سالم و ناسالم تا حد زیادی اشتراک داشته باشند و نتوان الگوی خاصی را در میان اوتیستیک‌ها شناسایی و به عنوان شاخصه مغزی افراد مبتلا به اوتیسم معرفی نمود.

با این حال قسمتی از طیف افراد سالم، دارای تفاوت محسوسی با طیف افراد اوتیسم است.

مراجع 7

- [1] Analysis of Variance (G. Larson, Jan 2008)
- [2] strategies for detection of autism syndrome disorder
- [3] Complex Brain Network Analysis and Its Applications to Brain Disorders
- [4] Machine Learning Methods for Diagnosing Autism Spectrum Disorder and Attention-Deficit/Hyperactivity Disorder Using Functional and Structural MRI
- [5] A Physarum Centrality Measure of the Human Brain Network
- [6] Diagnosis of autism spectrum disorder based on complex network features. Computer Methods and Programs in Biomedicine
- [7] Analysis of brain signals based on complex networks to improve the diagnosis of autism
- [8] Documentation
<https://scikitlearn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html>
- [9] Documentation <https://networkx.org/documentation/stable/reference/algorithms/>

ضمیمه‌ها 8

8.1 راهنمای استفاده از جداول

- فیلدهایی که با سبز مشخص شده نشان دهنده p -value کمتر از 0.02 می‌باشد.
- فیلدهایی که با آبی مشخص شده نشان دهنده p -value کمتر از 0.05 و بیشتر از 0.02 می‌باشد.
- فیلدهایی که با رنگ نارنجی نشان داده شده است نشان دهنده آن است که این ویژگی در دسته بندی ها خوب عمل کرده و مقدار p -value کمتر از 0.05 دارد ولی در دسته بندی واقعی مقدار p -value قابل قبولی ندارد.
- فیلدهایی که با سبز پررنگ نشان داده شده است برای ویژگی‌هایی است که چه در خوشه بندی و چه در لیبل‌های اصلی خوب عمل کرده اند.
- فیلدهایی که با آبی پررنگ نشان داده شده است برای ویژگی‌هایی است که در لیبل های اصلی مقداری بین 0.05 تا 0.02 دارند.
- فیلدهایی که با قرمز مشخص شده نشان دهنده آن است که آن ویژگی مقدار یکسانی بین کانکتوم‌ها داشته است.

8.2 ترشپلد 0.4

Features			2 clusters	3 clusters	4 clusters	5 clusters	True Label
Node strength	Max	0	0.002610	1.080526e-09	6.316354e-09	4.613776e-14	0.941256
	Chart3	1	0.008005	1.719046e-06	7.697699e-06	4.907477e-08	0.626687
	Med	2	0.441929	0.735088	0.884979	0.950935	0.872828
	Mean	3	0.433195	0.708611	0.088950	0.000560	0.615298
	Chart1	4	0.005297	3.715218e-08	1.909146e-07	2.431705e-10	0.780186
	Min	5	0.001640	8.438508e-09	4.767253e-08	4.689880e-12	0.932731
Eccentricity	Max	6	0.700085	0.588764	0.384646	0.553615	0.375530
	Chart3	7	nan	nan	nan	nan	nan
	Med	8	3.996119e-06	1.685546e-05	7.067146e-05	0.000232	0.063630
	Mean	9	0.000718	1.649070e-05	1.297364e-05	4.511391e-05	0.096772
	Chart1	10	0.008941	0.009822	0.005715	0.012895	0.646946
	Min	11	nan	nan	nan	nan	nan
Triangles	Max	12	2.596029e-16	2.068706e-15	1.567264e-21	1.394485e-20	0.028499
	Chart3	13	3.186189e-19	3.591038e-18	1.535198e-21	1.347852e-20	0.027942
	Med	14	2.576671e-15	2.107883e-14	2.773810e-16	1.953449e-15	0.027772
	Mean	15	3.969563e-18	3.477150e-17	6.003849e-21	5.123373e-20	0.026181

	Chart1	16	7.815514e-07	1.654203e-06	4.447694e-07	1.815101e-06	0.008797
	Min	17	0.390029	0.048910	0.111803	0.126611	0.825436
Clustering	Max	18	8.255417e-06	5.144355e-05	5.761122e-10	2.283350e-09	0.039066
	Chart3	19	7.778295e-15	4.896125e-14	2.157225e-21	1.774849e-20	0.036642
	Med	20	2.188789e-14	1.533783e-13	4.196251e-21	3.293272e-20	0.018362
	Mean	21	7.771187e-15	4.188102e-14	5.148993e-22	4.127667e-21	0.016783
	Chart1	22	5.316672e-12	3.140070e-11	5.909729e-17	4.191135e-16	0.008531
	Min	23	0.182536	0.010436	0.013847	0.025422	0.165042
Betweenness centrality	Max	24	0.024289	0.074308	0.024402	0.047519	0.113084
	Chart3	25	7.373861e-10	4.192990e-10	2.884368e-12	1.515370e-11	0.031212
	Med	26	2.326690e-11	4.347093e-11	2.467396e-12	1.399424e-11	0.030819
	Mean	27	2.923158e-13	2.500668e-13	1.020958e-17	7.539236e-17	0.015452
	Chart1	28	9.254494e-07	5.899664e-06	1.693074e-09	8.309867e-09	0.049625
	Min	29	0.200060	0.332894	0.159761	0.123255	0.093287
Closeness centrality	Max	30	1.473959e-12	7.758559e-12	6.838932e-21	6.050988e-20	0.012525
	Chart3	31	5.960411e-17	2.832831e-16	7.186934e-22	5.931704e-21	0.017322
	Med	32	1.709085e-12	5.506513e-12	8.660532e-15	5.723111e-14	0.015365
	Mean	33	1.271864e-14	1.904314e-14	5.712289e-19	4.466188e-18	0.015537
	Chart1	34	2.994479e-05	1.246569e-05	4.567713e-06	1.647356e-05	0.049554
	Min	35	0.649342	0.053985	0.120948	0.213999	0.792752
Degree centrality	Max	36	1.657617e-11	7.458952e-11	6.626859e-21	5.855422e-20	0.010146
	Chart3	37	2.967552e-16	1.475940e-15	2.855487e-22	2.362097e-21	0.014662
	Med	38	1.284304e-12	4.582800e-12	4.398781e-15	2.881507e-14	0.013343
	Mean	39	1.305024e-14	3.322150e-14	1.055381e-19	8.202995e-19	0.012537
	Chart1	40	6.444066e-06	4.070319e-06	1.168311e-06	4.589814e-06	0.015410
	Min	41	0.787436	0.056069	0.120459	0.195864	0.774425
Modularity		42	0.424567	0.705881	0.765085	0.881175	0.879755
Global efficiency		43	3.312281e-14	5.714312e-14	4.095319e-19	3.147598e-18	0.013273

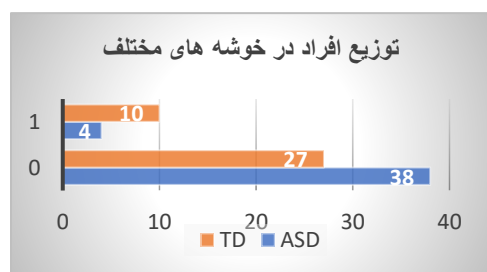
8.2.1 تحلیل دسته بندی‌ها با تمام ویژگی‌ها - ۴۴ ویژگی

برچسب‌های واقعی

از 79 فرد موجود در این بررسی، 42 فرد اول مبتلا به اوتیسم و 37 نفر بعدی افراد سالم در این مجموعه داده هستند.

خوشه‌بندی براساس دو دسته

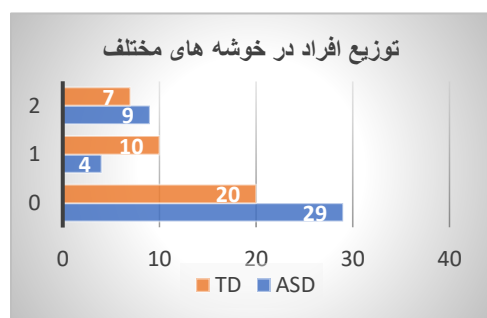
با توجه به این دسته بندی 65 نفر در گروه با برچسب صفر و 14 نفر هم در گروه دو با برچسب یک قرار گرفته اند.



برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.584	0.415	0.168	
1	0.285	0.714	0.429	
جمع			0.597	29%

خوشه بندی براساس سه دسته

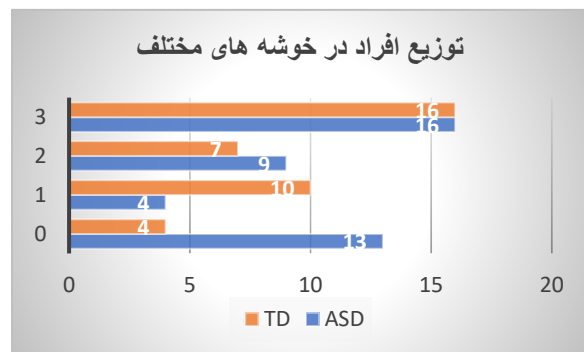
با توجه به این دسته بندی 49 نفر در گروه با برچسب صفر و 14 نفر هم در گروه یک و 14 نفر هم در گروه با برچسب دو قرار گرفته اند.



برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.591	0.408	0.182	
1	0.285	0.714	0.429	
2	0.562	0.437	0.124	
جمع			0.735	24%

خوشه بندی براساس چهار دسته

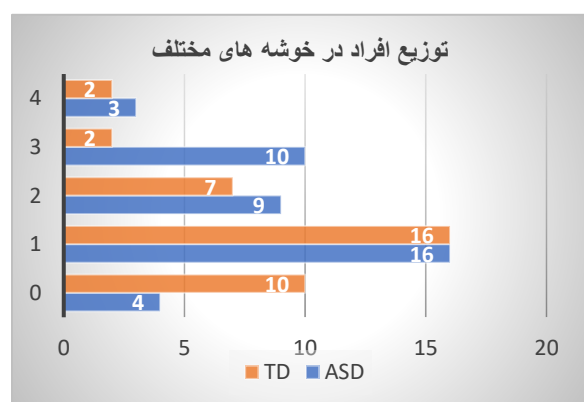
با توجه به این دسته بندی 17 نفر در گروه با برچسب صفر و 14 نفر هم در گروه با برچسب یک و 16 نفر هم در گروه با برچسب دو و 32 نفر هم در گروه با برچسب سه قرار گرفته اند.



برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.764	0.235	0.528	
1	0.285	0.714	0.429	
2	0.562	0.437	0.124	
3	0.5	0.5	0	
جمع			1.081	%27

خوشه بندی براساس پنج دسته

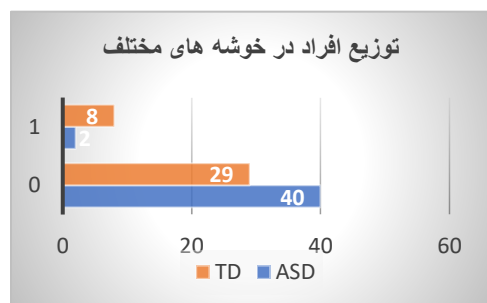
با توجه به این دسته بندی 14 نفر در گروه با برچسب صفر و 32 نفر هم در گروه با برچسب یک و 16 نفر هم در گروه با برچسب دو و 12 نفر هم در گروه با برچسب سه و 5 نفر هم در گروه با برچسب چهار قرار گرفته اند.



برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.285	0.714	0.429	
1	0.5	0.5	0.0	
2	0.562	0.437	0.124	
3	0.833	0.166	0.667	
4	0.6	0.4	0.2	
جمع			1.42	%28

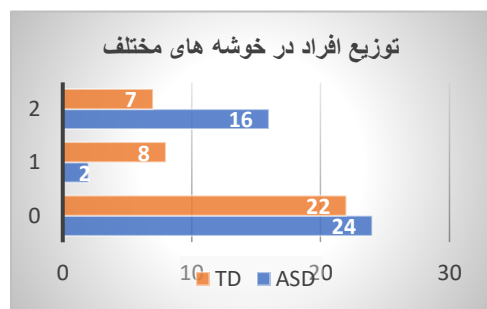
8.2.2 تحلیل دسته بندی ها با استفاده از ویژگی های منتخب – ۲۵ ویژگی

خوشه بندی براساس دو دسته



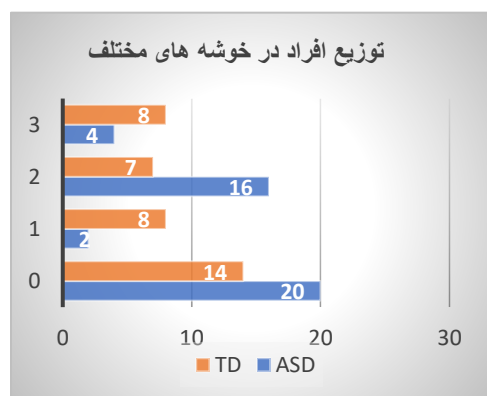
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.579	0.42	0.159	
1	0.2	0.8	0.6	
جمع			0.759	37.9%

خوشه بندی براساس سه دسته



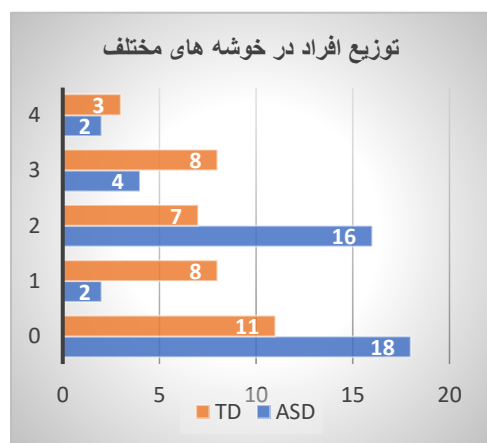
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.521	0.478	0.183	
1	0.2	0.8	0.6	
2	0.695	0.304	0.39	
جمع			1.173	39%

خوشه بندی براساس چهار دسته



برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.588	0.411	0.176	
1	0.2	0.8	0.6	
2	0.695	0.304	0.391	
3	0.333	0.666	0.333	
جمع			1.499	37%

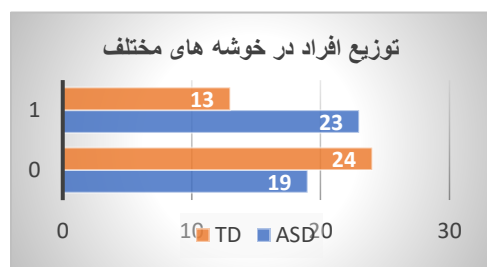
خوشه بندی براساس پنج دسته



برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.62	.0379	0.24	
1	0.2	0.8	0.6	
2	0.695	0.304	0.391	
3	0.333	0.666	0.333	
4	0.4	0.6	0.2	
جمع			1.764	35%

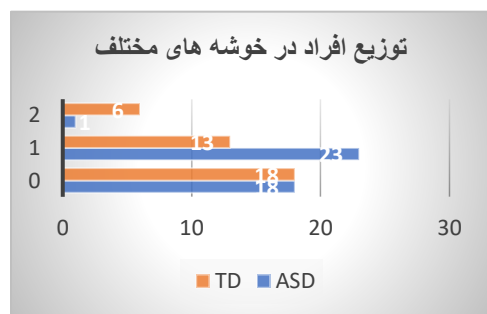
8.2.3 تحلیل دسته بندی ها بیا استفاده از ویژگی های برگزیده - ۱۵ ویژگی

خوشه بندی براساس دو دسته



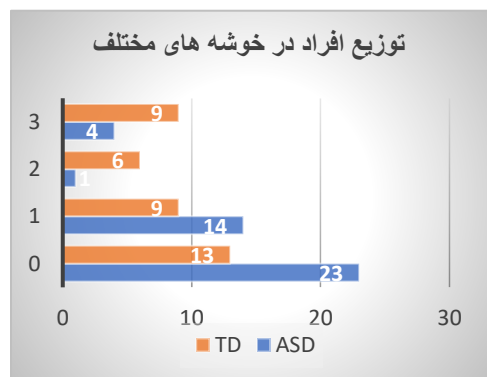
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.441	0.558	0.117	
1	0.638	0.361	0.276	
جمع			0.394	19%

خوشه بندی براساس سه دسته



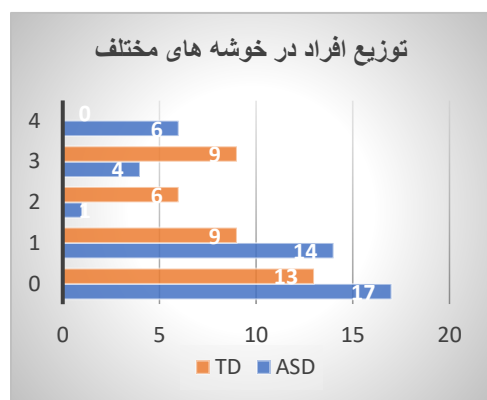
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.5	0.5	0.0	
1	0.638	0.361	0.276	
2	0.142	0.857	0.715	
جمع			0.991	33%

خوشه بندی براساس چهار دسته



برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.638	0.361	0.277	
1	0.608	0.391	0.216	
2	0.142	0.857	0.715	
3	0.307	0.692	0.385	
جمع			1.593	%39.8

خوشه بندی براساس پنج دسته



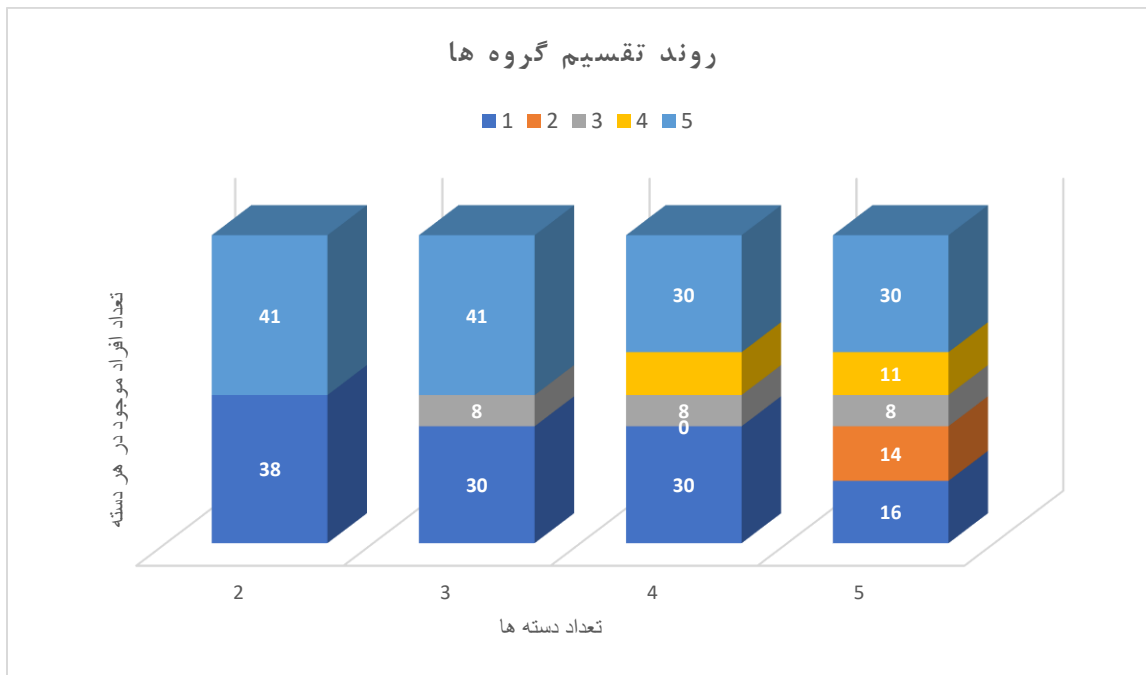
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.566	0.433	0.132	
1	0.608	0.391	0.216	
2	0.142	0.857	0.715	
3	0.301	0.692	0.391	
4	1	0	1	
جمع			2.454	%49

8.3 ترشهد 0.35

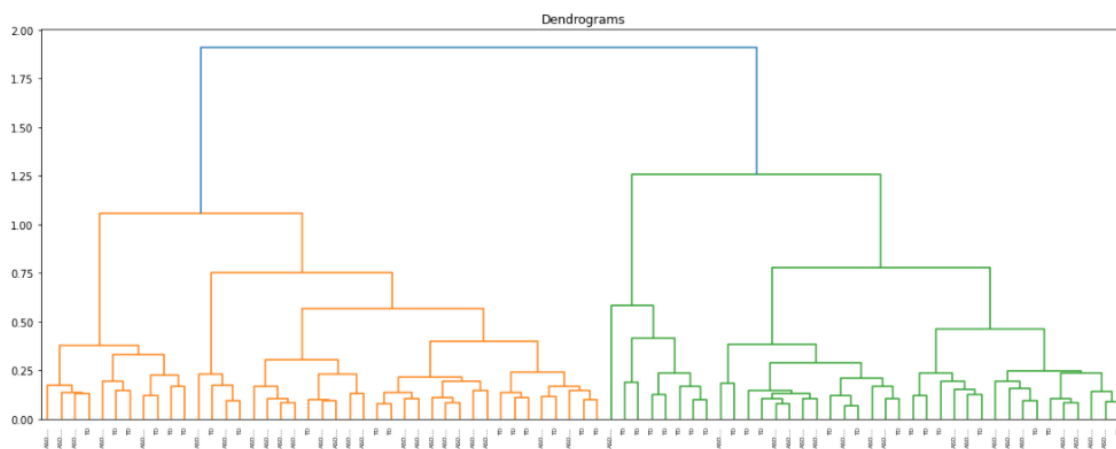
Features			2 clusters	3 clusters	4 clusters	5 clusters	True Label
Node strength	Max	0	7.632513e-11	3.328378e-10	2.341064e-11	4.159069e-11	0.416086
	Chart3	1	1.746786e-10	2.085830e-10	2.238848e-10	9.274886e-10	0.441810
	Med	2	0.923160	0.985921	0.803048	0.911917	0.981004
	Mean	3	0.246831	0.512084	0.702726	0.824521	0.209744
	Chart2	4	1.152859e-11	1.956314e-11	2.556146e-11	2.226415e-11	0.574589
	Min	5	3.507391e-11	5.844431e-12	8.909601e-13	5.245789e-12	0.575617
Eccentricity	Max	6	nan	nan	nan	nan	nan
	Chart3	7	0.000380	0.001695	0.003492	0.005392	0.633539
	Med	8	nan	nan	nan	nan	nan

	Mean	9	7.796868e-05	0.000400	0.000760	0.000848	0.526269
	Chart2	10	nan	nan	nan	nan	nan
	Min	11	nan	nan	nan	nan	nan
Triangles	Max	12	5.245584e-07	9.113200e-14	1.557271e-18	7.620327e-25	0.020745
	Chart3	13	6.825320e-07	6.371921e-17	1.375726e-20	3.011713e-26	0.020416
	Med	14	4.577919e-07	1.546253e-20	3.067408e-24	2.352328e-27	0.025444
	Mean	15	1.884543e-07	3.387791e-18	5.516699e-23	3.552922e-29	0.018680
	Chart2	16	1.931310e-06	2.777728e-11	6.165568e-13	3.148821e-13	0.008561
	Min	17	0.021646	0.000949	0.002950	0.006907	0.795203
Clustering	Max	18	4.635818e-05	1.080800e-05	5.058532e-09	5.788653e-11	0.089101
	Chart3	19	1.127378e-06	4.654289e-12	3.422955e-17	1.593064e-22	0.028553
	Med	20	2.239312e-07	1.023829e-13	1.103357e-19	5.732396e-25	0.015705
	Mean	21	1.340904e-07	3.928967e-14	2.064984e-20	6.887431e-26	0.013248
	Chart2	22	5.307198e-08	5.121016e-15	5.541636e-21	1.723486e-23	0.007516
	Min	23	0.019015	0.059006	0.076776	0.145630	0.149231
Betweenness centrality	Max	24	0.001144	0.000756	0.001430	0.002728	0.012981
	Chart3	25	3.337534e-08	7.305534e-14	2.303238e-16	2.094618e-19	0.091181
	Med	26	6.700558e-11	2.469694e-15	3.991323e-18	3.130764e-19	0.010830
	Mean	27	7.399057e-09	7.935308e-16	1.186190e-21	7.513244e-27	0.010313
	Chart2	28	4.210898e-06	1.305806e-07	2.888523e-09	1.236269e-10	0.043763
	Min	29	0.684431	0.015220	0.013056	0.015996	0.033128
Closeness centrality	Max	30	2.950119e-07	4.179676e-12	2.017718e-17	3.103271e-23	0.010378
	Chart3	31	2.114450e-07	6.926408e-16	8.061030e-21	7.480076e-27	0.016493
	Med	32	1.819381e-07	1.835298e-16	3.273998e-21	4.974075e-24	0.015955
	Mean	33	1.351908e-08	9.207189e-17	1.635215e-22	2.926521e-28	0.011258
	Chart2	34	1.326809e-06	2.898815e-09	6.850068e-11	2.921086e-11	0.016718
	Min	35	0.044723	0.035375	0.082528	0.150886	0.908276
Degree centrality	Max	36	2.732580e-07	3.016702e-11	7.424572e-17	3.132414e-22	0.009772
	Chart3	37	1.618835e-07	7.926977e-15	3.375185e-20	3.158363e-26	0.014518
	Med	38	1.369027e-07	9.528665e-16	6.086132e-21	9.475231e-24	0.015846
	Mean	39	1.113879e-08	7.681478e-16	8.275146e-22	4.586301e-27	0.010038
	Chart2	40	8.504134e-07	1.855039e-09	3.506439e-11	1.454460e-11	0.014470
	Min	41	0.107897	0.002247	0.006571	0.016093	0.742080
Modularity		42	0.742446	0.754466	0.731543	0.498209	0.146493

Global efficiency		43	9.684100e-09	7.704570e-16	9.202860e-22	5.284285e-27	0.010123
-------------------	--	----	--------------	--------------	--------------	--------------	----------

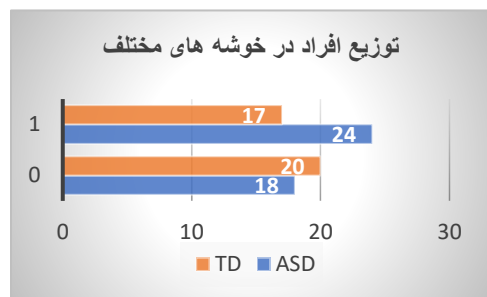


بررسی روند دسته بندی داده ها با کل ویژگی ها - 44 ویژگی



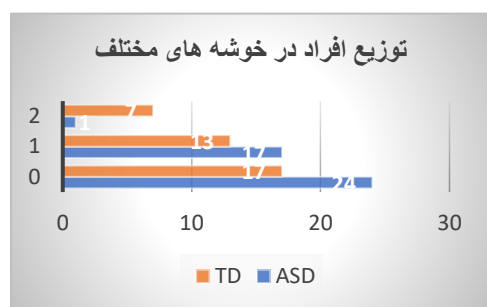
دیاگرام مربوط به خوشه بندی با استفاده از 44 ویژگی

خوشه بندی براساس دو دسته



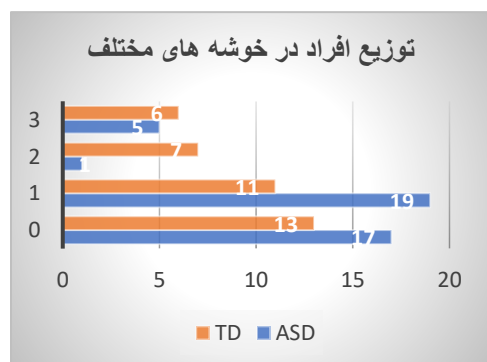
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.473	0.526	0.053	
1	0.585	0.414	0.171	
جمع			0.224	%11

خوشه بندی براساس سه دسته



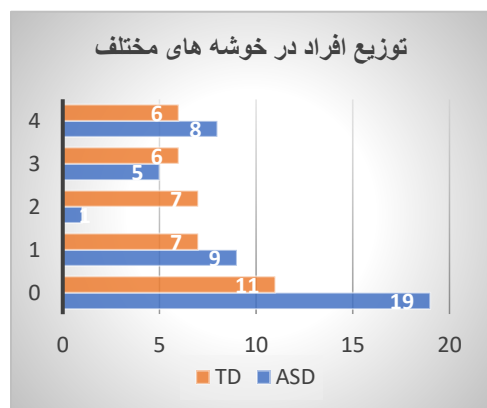
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.585	0.414	0.171	
1	0.566	0.433	0.133	
2	0.125	0.875	0.75	
جمع			1.054	%35

خوشه بندی براساس چهار دسته



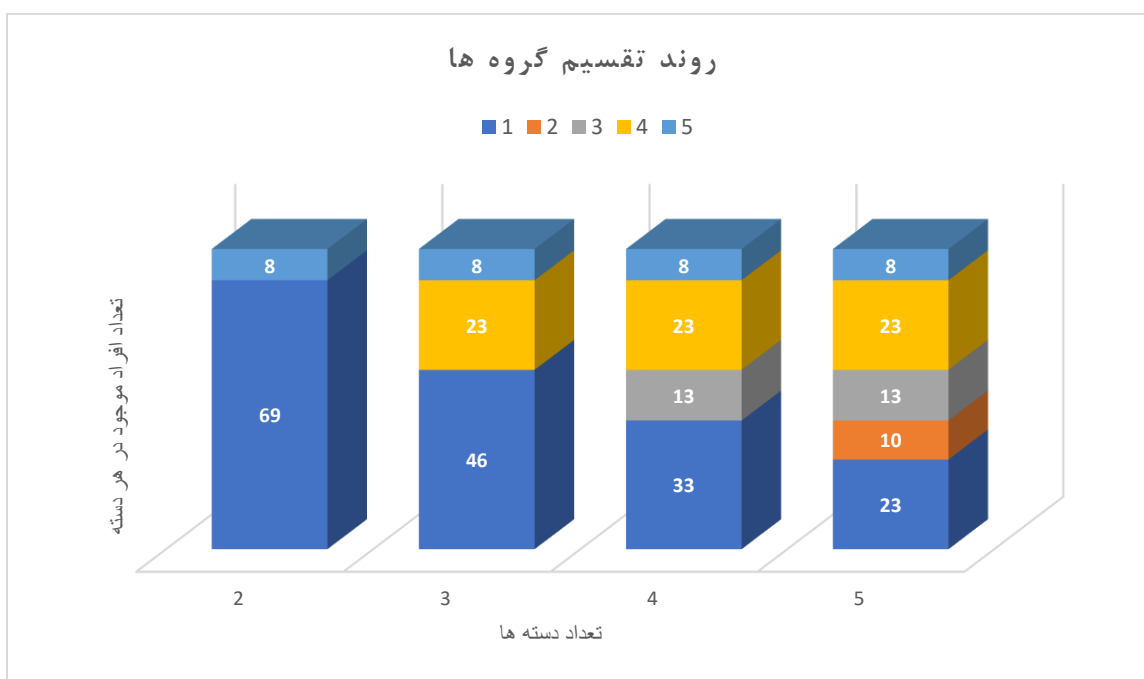
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.566	0.433	0.133	
1	0.633	0.366	0.267	
2	0.125	0.875	0.75	
3	0.454	0.545	0.091	
جمع			1.241	%31

خوشه بندی براساس پنج دسته

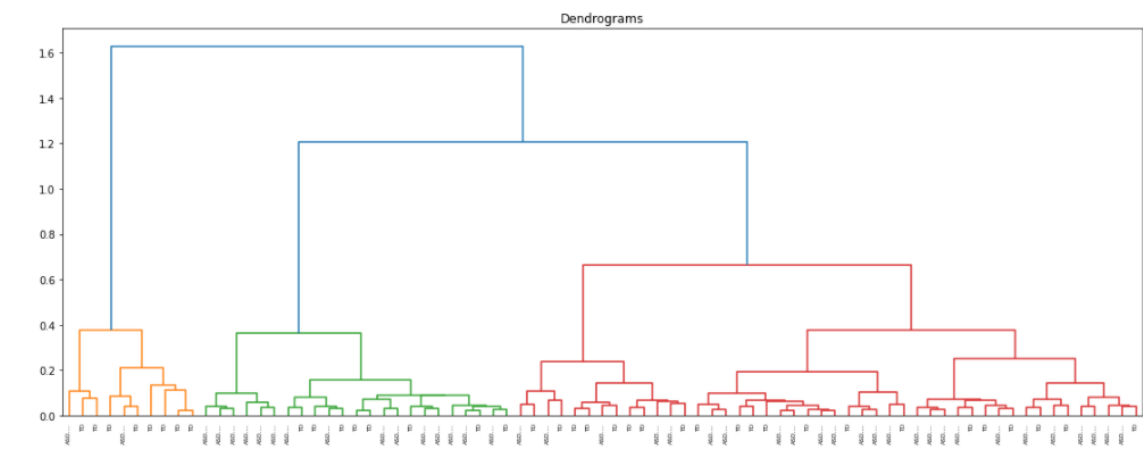


به درصد	اختلاف دو ستون قبلی	سالم کل گروه	اوتیسم کل گروه	برچسب
	0.267	0.366	0.633	0
	0.124	0.437	0.562	1
	0.75	0.875	0.125	2
	0.091	0.545	0.454	3
	0.142	0.428	0.571	4
27%	1.374			جمع

8.3.1 خوشه بندی کانکتوم ها با استفاده از ویژگی های منتخب - ۲۵ ویژگی

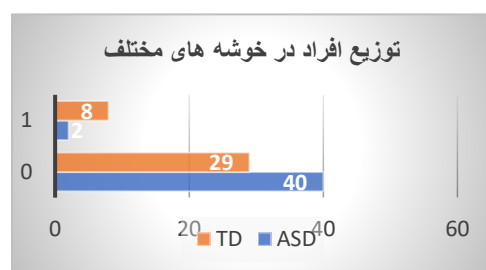


بررسی روند دسته بندی داده ها با ویژگی های منتخب - ۲۵ ویژگی



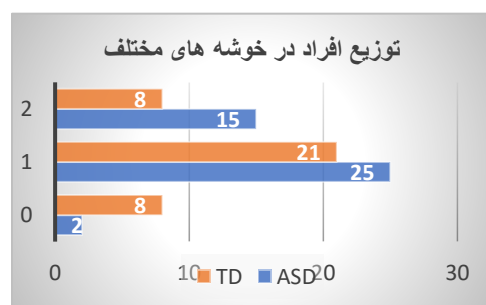
دیاگرام مربوط به خوشه بندی با استفاده از ۲۵ ویژگی

خوشه بندی براساس دو دسته



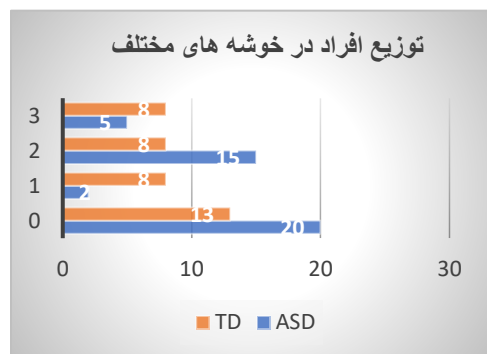
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.579	0.42	0.159	
1	0.2	0.8	0.6	
جمع			0.759	37.9%

خوشه بندی براساس سه دسته



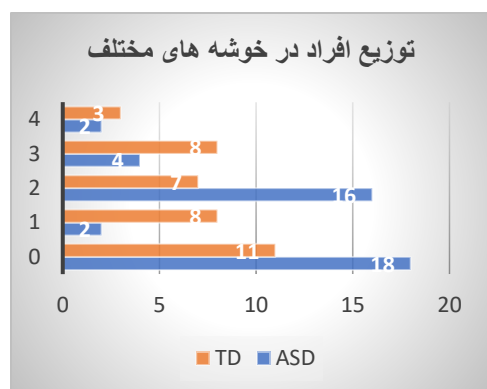
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.2	0.8	0.6	
1	0.543	0.456	0.086	
2	0.652	0.347	0.305	
جمع			0.991	33%

خوشه بندی براساس چهار دسته



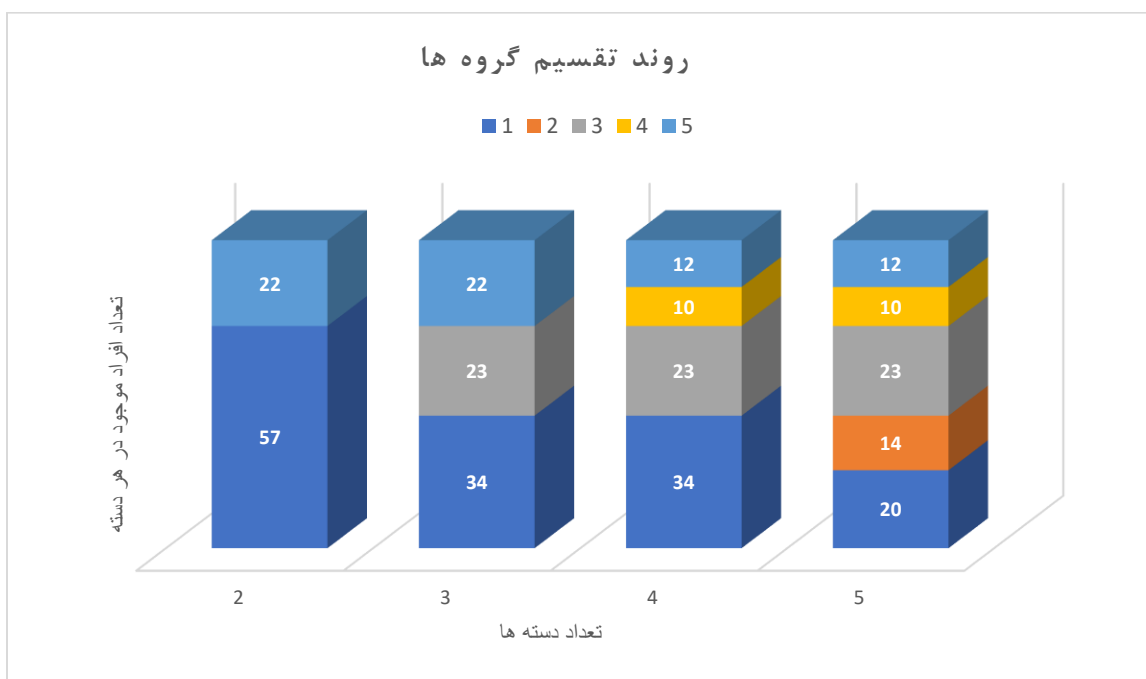
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.606	0.393	0.212	
1	0.2	0.8	0.6	
2	0.652	0.347	0.305	
3	0.384	0.615	0.231	
جمع			1.348	%33

خوشه بندی براساس پنج دسته

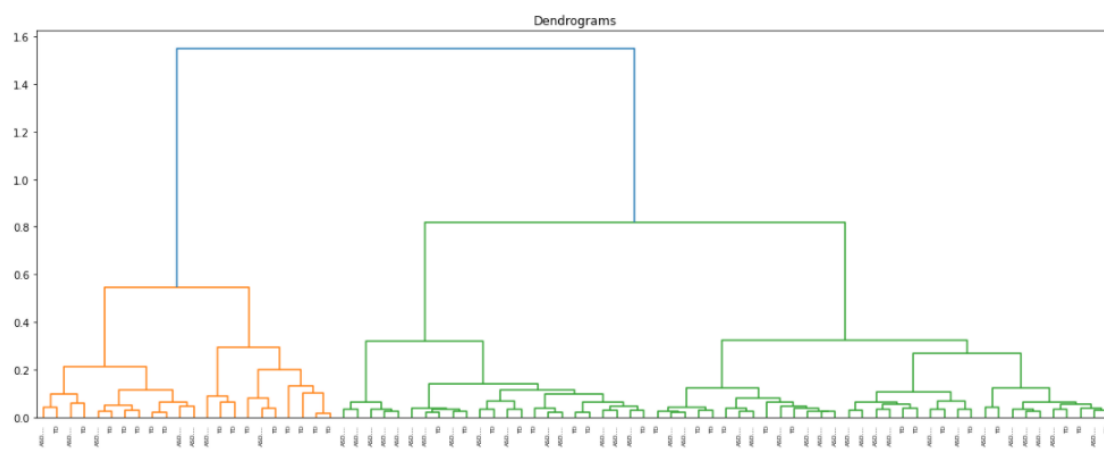


برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.2	0.8	0.6	
1	0.647	0.352	0.294	
2	0.652	0.347	0.305	
3	0.384	0.615	0.231	
4	0.562	0.437	0.124	
جمع			1.554	%31

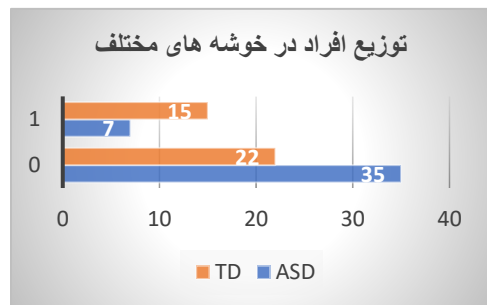
8.3.2 خوشه بندی کانکتوم‌ها با استفاده از ویژگی‌های برگزیده - ۱۹ ویژگی



بررسی روند دسته‌بندی داده‌ها با ویژگی‌های برگزیده - ۱۹ ویژگی

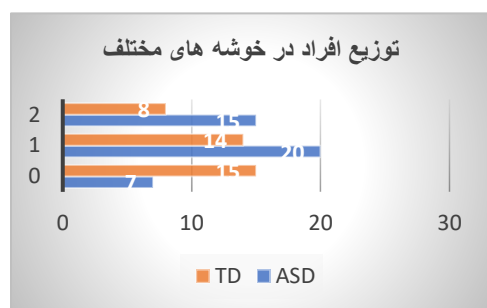


خوشه بندی براساس دو دسته



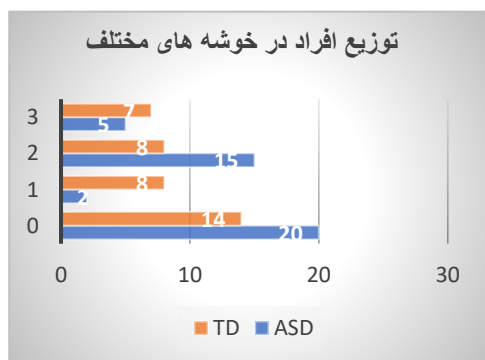
به درصد	اختلاف دو ستون قبلی	سالم کل گروه	اوتیسم کل گروه	برچسب
	0.228	0.385	0.614	0
	0.0363	0.0681	0.0318	1
29%	0.394			جمع

خوشه بندی براساس سه دسته



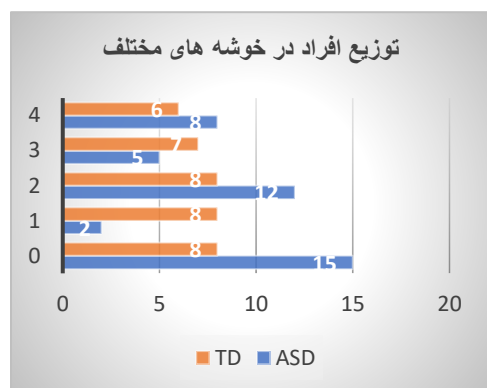
به درصد	اختلاف دو ستون قبلی	سالم کل گروه	اوتیسم کل گروه	برچسب
	0.363	0.681	0.318	0
	0.176	0.411	0.588	1
	0.304	0.347	0.652	2
28%	0.843			جمع

خوشه بندی براساس چهار دسته

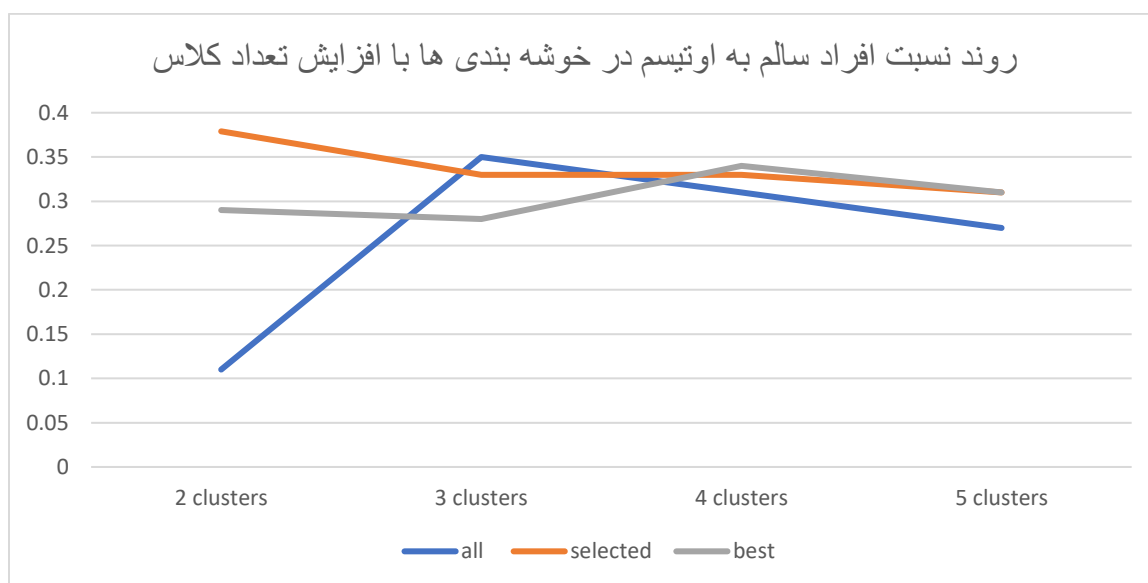


به درصد	اختلاف دو ستون قبلی	سالم کل گروه	اوتیسم کل گروه	برچسب
	0.176	0.411	0.588	0
	0.6	0.8	0.2	1
	0.304	0.347	0.652	2
	0.304	0.583	0.416	3
34%	1.384			جمع

خوشه بندی براساس پنج دسته



برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.652	0.347	0.304	
1	0.2	0.8	0.6	
2	0.6	0.4	0.2	
3	0.652	0.347	0.304	
4	0.571	0.428	0.142	
جمع			1.55	31%

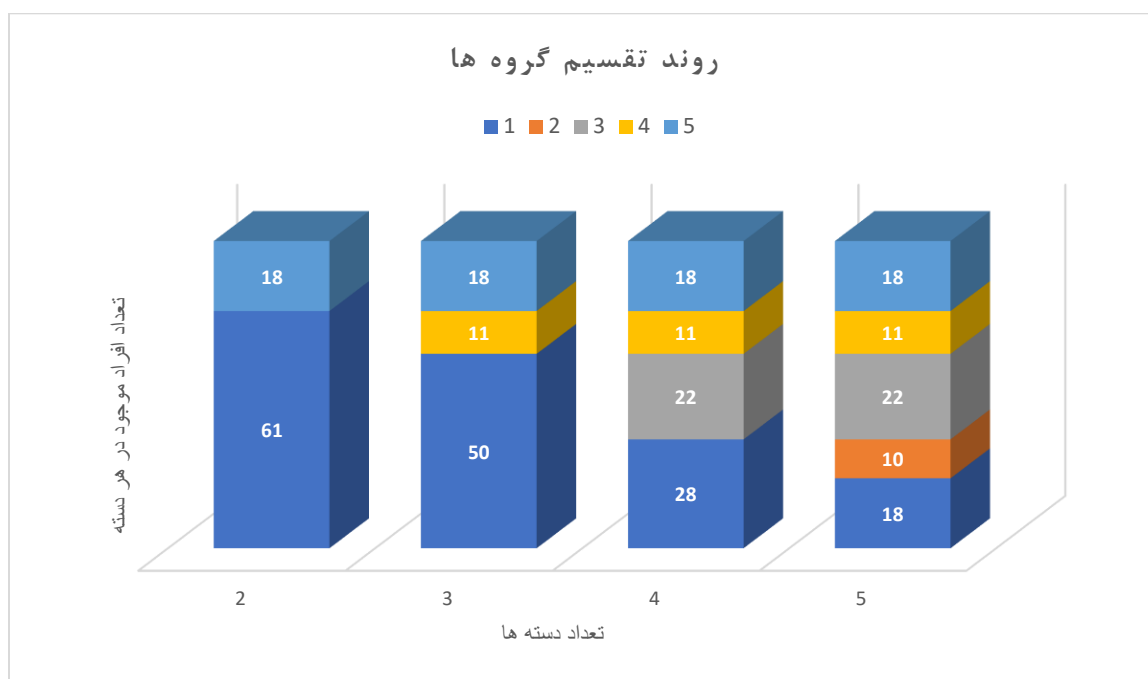


تحلیل روند خوشه بندی براساس ویژگی ها

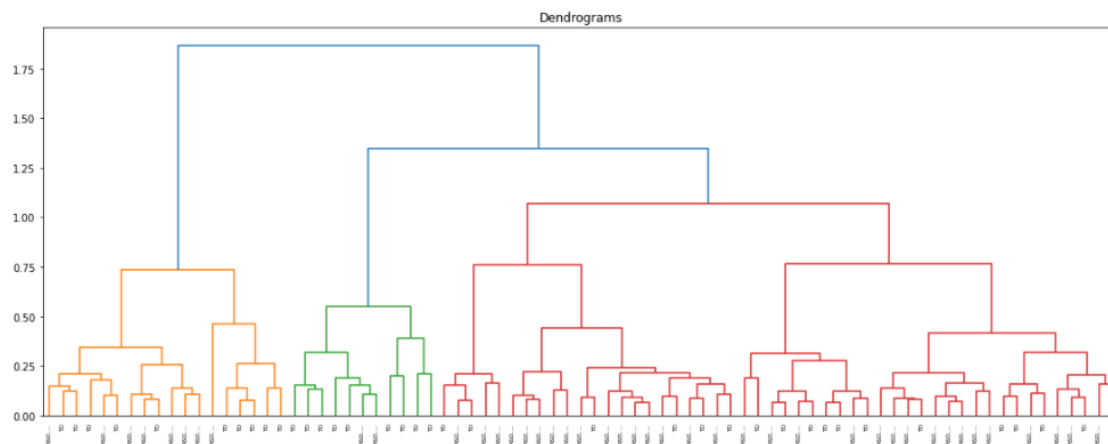
8.4 ترشهلد 0.3

Features			2 clusters	3 clusters	4 clusters	5 clusters	True Label
Node strength	Max	0	0.000267	5.236767e-11	1.814163e-13	1.690742e-16	0.167257
	Chart3	1	0.001869	0.000164	3.348059e-05	8.571481e-09	0.355187
	Med	2	0.474697	0.061553	0.081511	0.119003	0.833189
	Mean	3	0.803618	0.846632	0.858441	0.890640	0.630703
	Chart2	4	0.002558	7.188932e-06	3.672387e-07	1.742130e-10	0.513266
	Min	5	4.671405e-05	8.968978e-09	3.361809e-10	1.153263e-13	0.228745
Eccentricity	Max	6	0.672364	0.777295	0.109555	0.149939	0.950537
	Chart3	7	nan	nan	nan	nan	nan
	Med	8	nan	nan	nan	nan	nan
	Mean	9	0.410309	0.273006	0.026211	0.054354	0.306109
	Chart2	10	nan	nan	nan	nan	nan
	Min	11	nan	nan	nan	nan	nan
Triangles	Max	12	1.038795e-13	2.970796e-16	1.605154e-21	3.030373e-21	0.010409
	Chart3	13	1.111358e-13	2.038674e-16	3.680745e-19	1.008415e-18	0.021529
	Med	14	1.463803e-10	4.999856e-13	3.477605e-14	1.600068e-13	0.017418
	Mean	15	1.315730e-12	1.497629e-15	2.024748e-18	7.600440e-18	0.014278
	Chart2	16	4.867013e-05	2.845409e-07	5.579709e-08	2.360893e-07	0.005734
	Min	17	0.485089	0.064993	0.003981	0.009363	0.974898
Clustering	Max	18	1.274198e-08	8.728623e-11	1.826024e-12	6.994054e-13	0.046641
	Chart3	19	9.227775e-13	7.887419e-16	3.460003e-21	3.164259e-21	0.024925
	Med	20	2.314717e-12	1.047612e-15	1.331084e-20	3.305060e-20	0.014157
	Mean	21	1.270417e-11	3.645516e-15	2.001153e-20	4.723936e-20	0.011286
	Chart2	22	1.641794e-08	8.547038e-12	2.749524e-15	1.452508e-14	0.003888
	Min	23	0.013410	0.040815	0.005721	0.011289	0.011172
Betweenness centrality	Max	24	0.006492	0.000719	0.002406	0.003825	0.022369
	Chart3	25	1.990410e-08	1.259428e-09	3.935079e-13	6.212249e-13	0.080273
	Med	26	1.345303e-09	5.004576e-11	6.123681e-17	4.113718e-16	0.011395
	Mean	27	8.361857e-11	7.290729e-14	2.690205e-18	1.160535e-17	0.009107
	Chart2	28	4.427311e-09	5.949962e-10	1.238182e-13	6.219282e-13	0.024459
	Min	29	0.008694	8.448498e-05	0.000317	0.000951	0.008307
Closeness	Max	30	1.917601e-11	2.292892e-13	5.716463e-20	6.005180e-20	0.007763

centrality	Chart3	31	6.946560e-13	3.585926e-16	1.916329e-20	4.972094e-20	0.012146
	Med	32	2.138882e-09	3.210246e-12	5.812049e-14	2.758082e-13	0.011685
	Mean	33	1.997861e-11	1.616620e-14	1.671785e-18	7.098059e-18	0.009769
	Chart2	34	8.652946e-05	3.947797e-06	5.062653e-07	1.975294e-06	0.008870
	Min	35	0.305284	0.029057	0.000880	0.002186	0.390519
Degree centrality	Max	36	7.603550e-11	7.381998e-13	2.297829e-20	2.273291e-20	0.008037
	Chart3	37	1.577426e-12	4.677536e-16	2.323698e-21	5.040889e-21	0.010785
	Med	38	2.006818e-09	2.556482e-12	1.849185e-14	8.628510e-14	0.010847
	Mean	39	8.318004e-11	7.122834e-14	2.681424e-18	1.157015e-17	0.009043
	Chart2	40	6.962276e-05	3.713814e-06	3.842959e-07	1.511220e-06	0.008746
	Min	41	0.185012	0.010719	0.000893	0.002233	0.456062
Modularity		42	0.748177	0.698794	0.868185	0.137056	0.962672
Global efficiency		43	8.332484e-11	7.178236e-14	2.684261e-18	1.158150e-17	0.009064

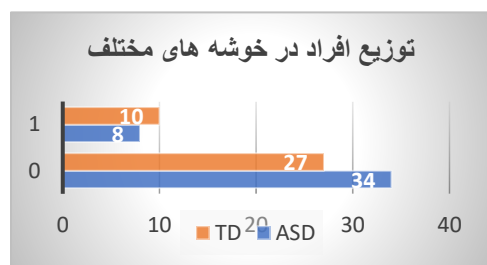


بررسی روند دسته بندی داده ها با کل ویژگی ها - 44 ویژگی



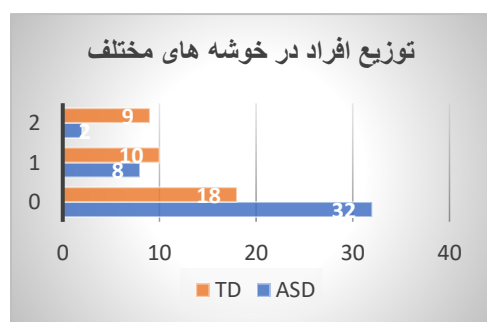
دیاگرام مربوط به خوشه بندی با استفاده از ۴۴ ویژگی

خوشه بندی براساس دو دسته



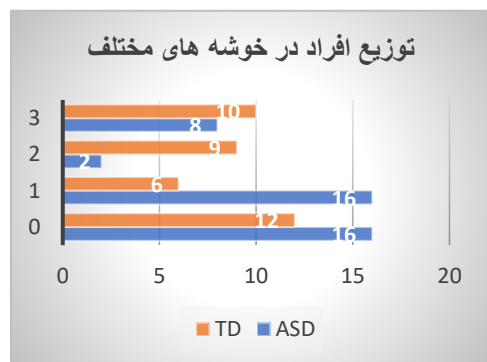
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.557	0.442	0.115	%11
1	0.444	0.555	0.111	
جمع			0.226	

خوشه بندی براساس سه دسته



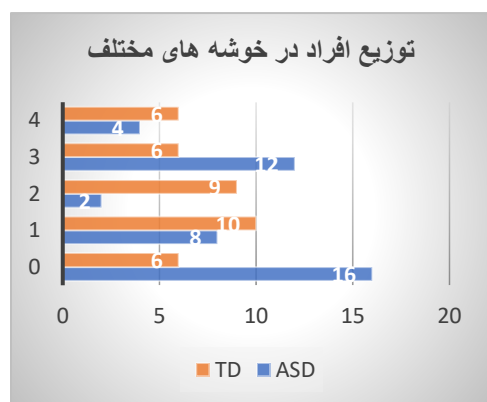
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.64	0.36	0.28	%34
1	0.444	0.555	0.111	
2	0.181	0.818	0.637	
جمع			1.028	

خوشه بندی براساس چهار دسته

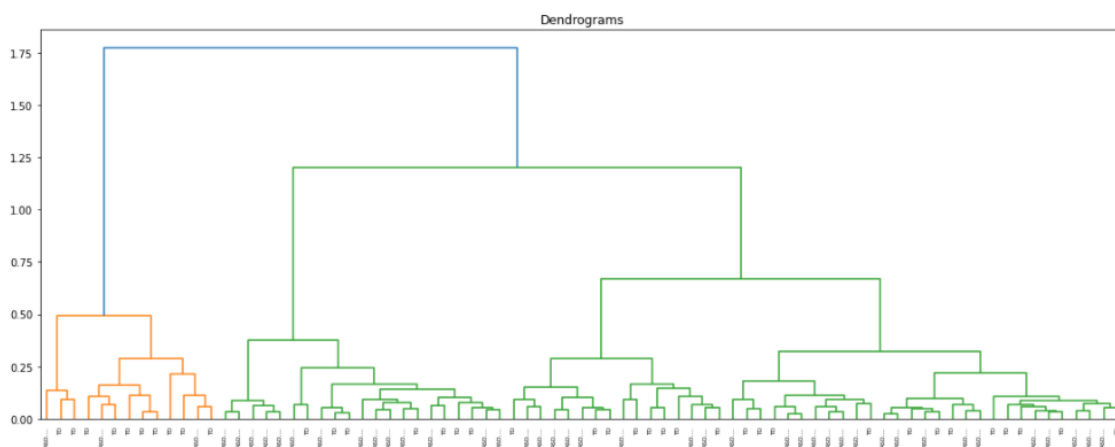


برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.571	0.428	0.142	
1	0.727	0.272	0.454	
2	0.181	0.818	0.637	
3	0.444	0.555	0.111	
جمع			1.344	
				%33

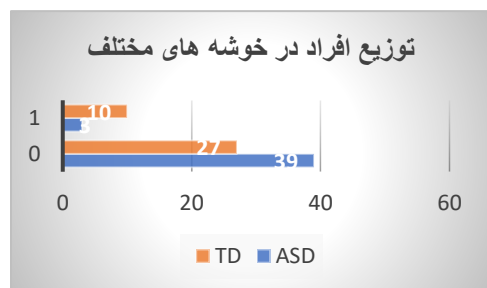
خوشه بندی براساس پنج دسته



برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.727	0.272	0.454	
1	0.444	0.555	0.111	
2	0.181	0.818	0.637	
3	0.666	0.333	0.332	
4	0.4	0.6	0.2	
جمع			1.734	%34

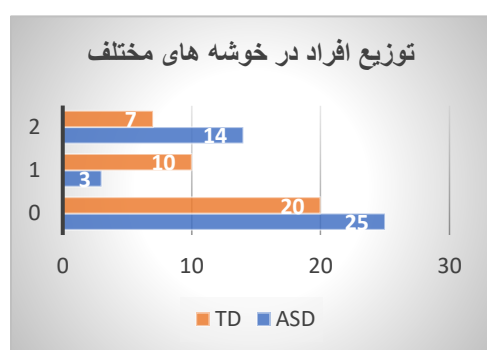


خوشه بندی براساس دو دسته



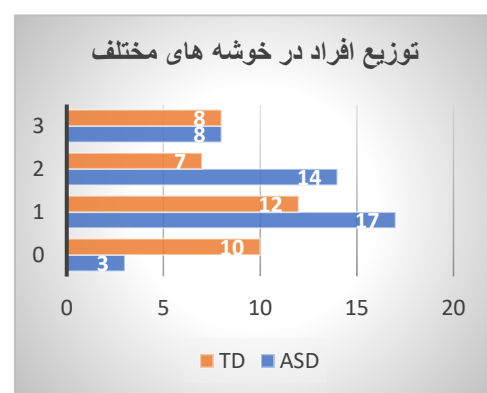
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبل	به درصد
0	0.59	0.409	0.181	35%
1	0.23	0.769	0.538	
جمع			0.719	

خوشه بندی براساس سه دسته



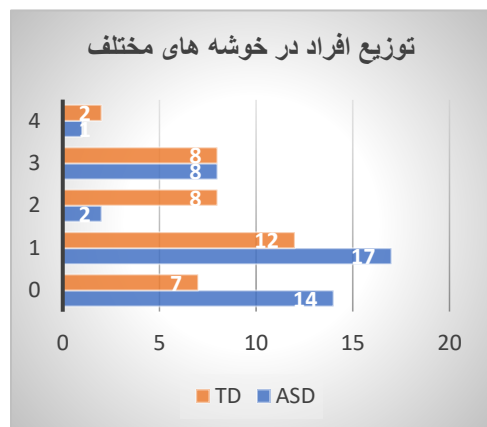
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبل	به درصد
0	0.555	0.444	0.111	%32
1	0.23	0.769	0.538	
2	0.666	0.333	0.333	
جمع			0.982	

خوشه بندی براساس چهار دسته



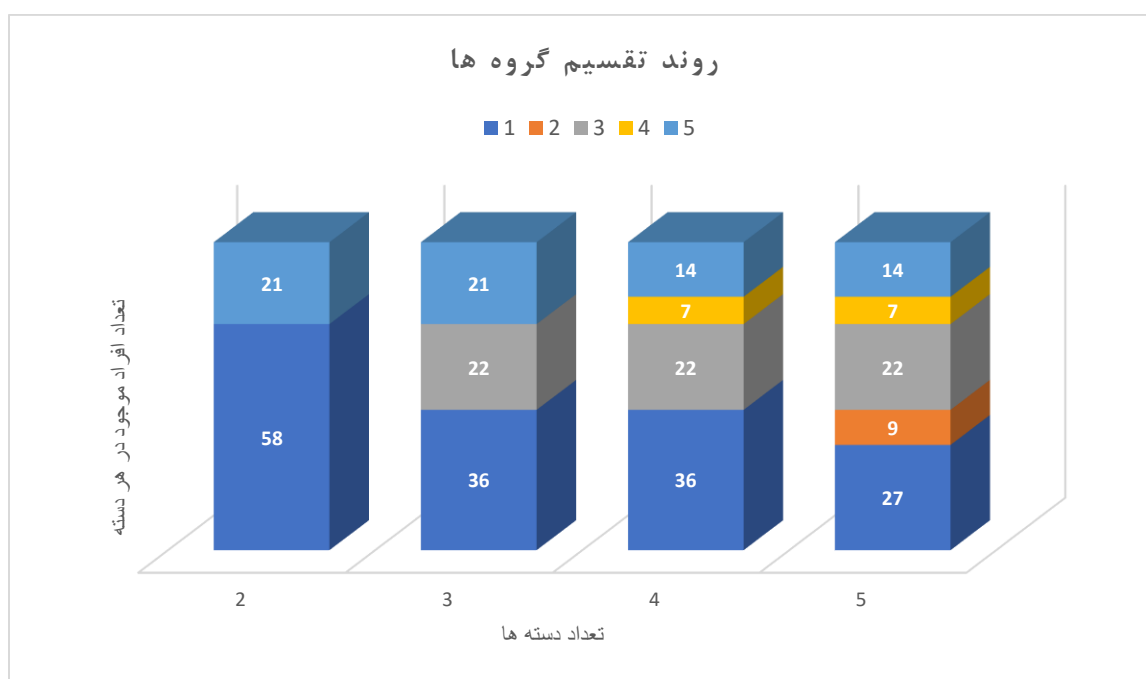
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبل	به درصد
0	0.23	0.769	0.538	%22
1	0.586	0.413	0.172	
2	0.666	0.333	0.333	
3	0.5	0.5	0.0	
جمع			0.882	

خوشه بندی براساس پنج دسته

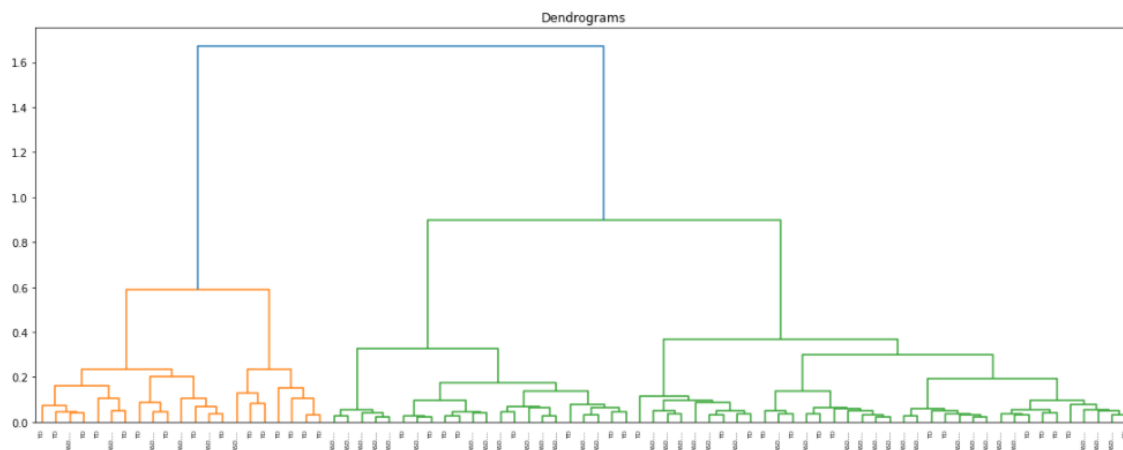


برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبلی	به درصد
0	0.666	0.333	0.333	
1	0.586	0.413	0.172	
2	0.2	0.8	0.6	
3	0.5	0.5	0.0	
4	0.333	0.666	0.333	
جمع			1.438	%28

8.4.2 خوشه بندی کانکتومها با استفاده از ویژگی های برگزیده - ۲۲ ویژگی

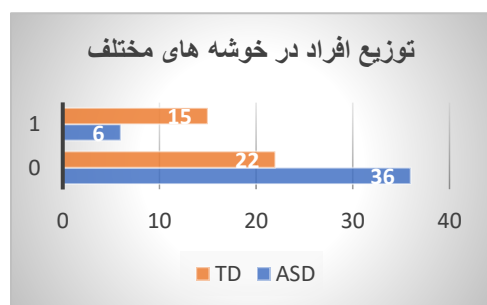


بررسی روند دسته بندی داده ها با کل برگزیده - ۲۲ ویژگی



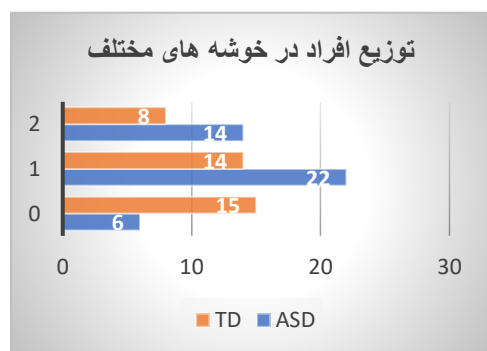
دیاگرام مربوط به خوشه بندی با استفاده از ۲۲ ویژگی

خوشه بندی براساس دو دسته



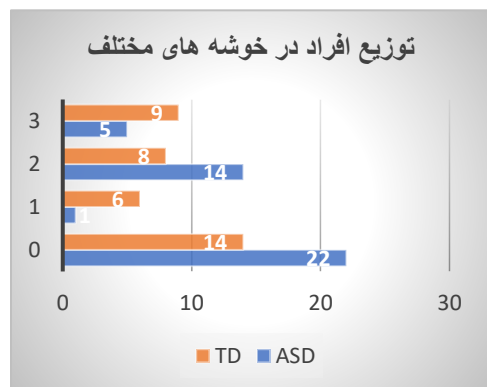
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبل	به درصد
0	0.62	0.38	0.241	33٪
1	0.285	0.714	0.428	
جمع			0.669	

خوشه بندی براساس سه دسته



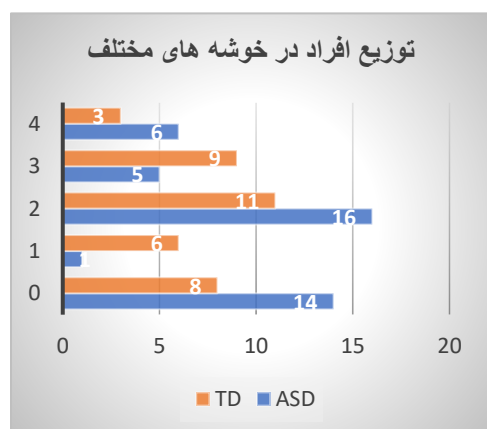
برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبل	به درصد
0	0.285	0.714	0.428	30٪
1	0.611	0.388	0.222	
2	0.636	0.363	0.272	
جمع			0.922	

خوشه بندی براساس چهار دسته

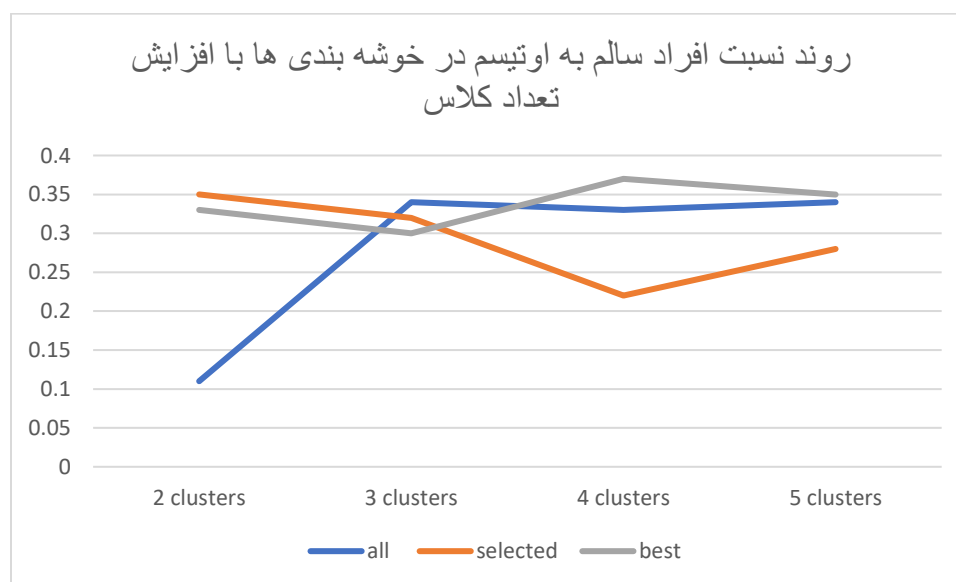


برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبل	به درصد
0	0.611	0.388	0.222	
1	0.142	0.857	0.714	
2	0.636	0.363	0.272	
3	0.357	0.642	0.285	
جمع			1.493	%37

خوشه بندی براساس پنج دسته



برچسب	اوتیسم کل گروه	سالم کل گروه	اختلاف دو ستون قبل	به درصد
0	0.636	0.363	0.272	
1	0.142	0.857	0.714	
2	0.592	0.407	0.184	
3	0.357	0.642	0.285	
4	0.666	0.333	0.333	
جمع			1.788	%35



تحلیل روند خوشه بندی براساس ویژگی ها

8.5 تحلیل های کمی

8.5.1 بررسی ویژگی های مختلف براساس تعداد مثلث ها

این بررسی بدین صورت انجام می گیرد که نشان می دهد تعداد مثلث های روی این نود تشکیل دهنده چند درصد از کل مثلث های موجود در این گراف می باشد. در این مثالی که در ادامه آورده شده است نشان داده می شود که چگونه می توان تحلیل دقیق تری از نحوه ی عملکرد تفکیک یک ویژگی بر روی گروه های مختلف داشت.

8.5.1.1 ماکسیمم ضریب خوشگی

همان طور که در تصاویر زیر مشاهده می کنید با توجه به نتایج حاصل از آزمون *post-hoc* امکان بررسی میان گروه های 4 و 2، 4 و 3، 3 و 2 به علت آنچه در این ویژگی پراکندگی بین گروهی کمتری نسبت به پراکندگی درون گروهی دارند بررسی نمی گردد پس به ادامه بررسی می پردازیم.

گروه 0 تا حدی از گروه 1 و به خوبی از گروه 4 جدا شده است اما تفکیک خوبی نسبت به دو گروه دیگر ندارد.

گروه 1 تا حدی از گروه 0 و 3 و به خوبی از گروه 4 جدا شده است.

گروه 2 از هیچ گروهی تفکیک نشده است.

گروه 3 تنها تا حدی از گروه 1 جدا شده است.

گروه 4 نیز از گروه های 0 و 1 به خوبی جدا شده است.

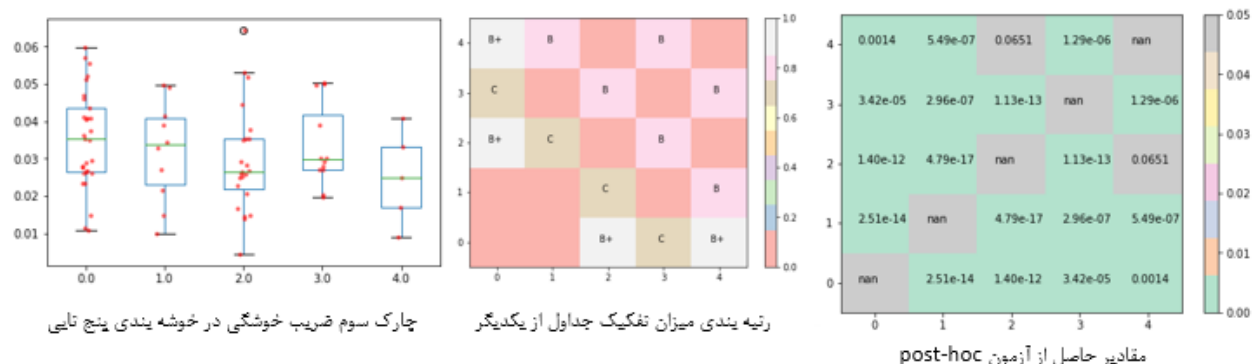
به نظر می رسد این ویژگی تفکیک معناداری بین گروه 4 و 0، 4 و 1 ایجاد کرده است.



8.5.1.2 چارک سوم ضریب خوشگی

گروه‌های 4 و 2 به خوبی از گروه 0 تفکیک شده اند، همچنین گروه 4 توانسته تا حدی از مابقی گروه‌ها جدا شود.

گروه 3 و 2 نیز تا حد کمی تفکیک شده اند اما بسیار بهم نزدیک اند.



6

8.5.1.3 میانه ضریب خوشگی

گروه 4 و 2 تا حد خوبی از گروه‌های 0، 1 و 3 جدا شده است.



7

8.5.1.4 چارک اول ضریب خوشگی

گروه 4 و 2 تا حد خوبی از گروه‌های 1 و 3 جدا شده است.

گروه 0 و 2 نیز تا حدی از هم تفکیک شده اند در این ویژگی.

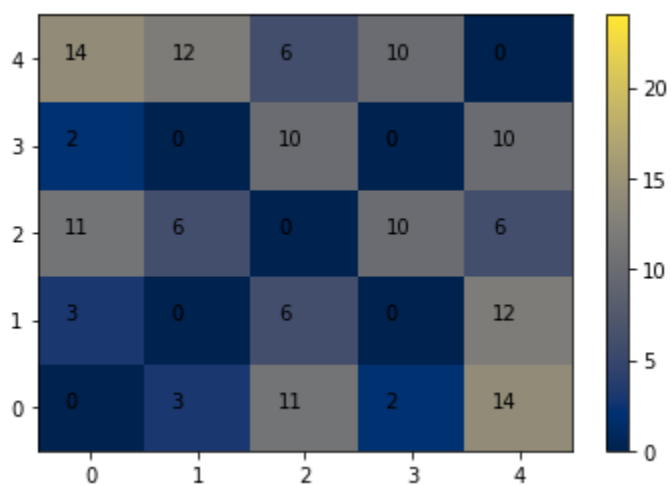


8

با مقایسه 4 جداول و نمودارهای بالا متوجه می‌شویم که گروه 4 از گروه‌های 0، 1 و 3 کاملاً متمایز می‌شوند و گروه‌های 2 و 4، 3 و 1 تا حد خوبی نسبت به یک دیگر تفکیک شده‌اند هم‌چنین گروه‌های 0 و 3 نیز از یک دیگر کاملاً متمایز شده‌اند. برای مقایسه‌ی بهتر نتایج بالا، برای هر یک از سطح‌های جدا شدن یک امتیاز در نظر می‌گیریم که به شرح زیر است

$$C = 1, B = 3, B+ = 4, A+ = 6$$

و برای هر جفت ویژگی نتایج حاصل از چهار جدول قبلی را در بالا باهم جمع می‌کنیم:



هر چه مقدار نتایج بالا بین دو جفت گروه بیشتر باشد به معنای تفکیک بهتر آن گروه‌ها در ویژگی مورد نظر می‌باشد.

پایان