

دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده‌ی برق و کامپیوتر



پروژه نهایی یادگیری ماشین

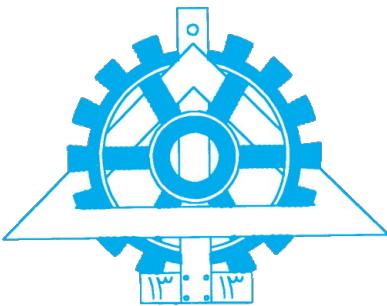
محسن فیاض، محمد رضا عظیمی، ثمین مهدی زاده، مهدی صابری

استاد راهنما

دکتر ابوالقاسمی

۱۴۰۰ زمستان

سُبْحَانَ رَبِّ الْجَمَلِ



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده‌ی برق و کامپیوتر



پروژه نهایی یادگیری ماشین

محسن فیاض، محمدرضا عظیمی، ثمین مهدی زاده، مهدی صابری

استاد راهنما

دکتر ابوالقاسمی

۱۴۰۰ زمستان

فهرست مطالب

۱	فصل ۱: فاز اول	
۲	سازها و آهنگها	۱.۱
۲	نی	۱.۱.۱
۳	پیانو	۲.۱.۱
۴	سنور	۳.۱.۱
۵	سه تار	۴.۱.۱
۶	تار	۵.۱.۱
۷	ویولن	۶.۱.۱
۸	تفاوت‌ها و چالش‌ها و سختی‌ها	۲.۱
۱۱	فصل ۲: تمیزسازی داده‌ها	
۱۱	استخراج ویژگی‌ها	۱.۲
۱۴	PCA	۲.۲
۱۶	LDA	۳.۲
۱۹	فصل ۳: طبقه‌بندی	
۲۰	Gaussian Naive Bayes	۱.۳
۲۲	Linear SVM	۲.۳
۲۴	RBF SVM	۳.۳
۲۶	MLP	۴.۳

۲۹	۵.۳ جمع‌بندی و مقایسه
۳۱	فصل ۴: خوش‌بندی
۳۲	K-Means ۱.۴
۳۹	Agglomerative ۲.۴
اول	مراجع

فهرست شکل‌ها

۲	سازنی	۱.۱
۳	ساز پیانو	۲.۱
۴	ساز سنتور	۳.۱
۵	ساز سه‌تار	۴.۱
۶	ساز تار	۵.۱
۷	ساز ویولن	۶.۱
۸	گستره صوتی سازهای مورد بررسی	۷.۱
۲۱	Gaussian Naive Bayes	۱.۲
۲۳	Linear SVM	۲.۲
۲۵	RBF SVM	۳.۲
۲۷	MLP	۴.۲

فهرست جداول

۱.۱ تفاوت‌های سازه‌ای مورد بررسی ۹

فصل ۱

فاز اول

در این فصل راجع به انواع این سازها و آهنگ‌های ساخته شده با آن‌ها اطلاعاتی جمع آوری کرده و توضیح می‌دهیم. در این فصل نکاتی را در ارتباط با تفاوت‌های این آهنگ‌ها به همراه چالش‌ها و سختی‌های مسئله ذکر می‌کنیم.

۱.۱ سازها و آهنگ‌ها

۱.۱.۱ نی



شکل ۱.۱: ساز نی

نی از جمله قدیمی ترین سازهای اصیل ایرانی است و در دسته سازهای بادی قرار می‌گیرد. این ساز از یک لوله استوانه‌ای از جنس نی ساخته می‌شود و در سرتاسر آن سوراخهایی وجود دارد. نی از جمله سازهای بی‌زبانه است و صدای آن از طریق دمیدن نفس نوازنده به درون آن ایجاد می‌شود و سپس نوازنده نی، با باز و بسته کردن سوراخها توسط انگشتان هر دو دستش، طول موج ارتعاش را کم و زیاد می‌کند و به این طریق صدای خارج شده زیر و بم می‌شود. ساز نی قابل کوک کردن نیست و نمی‌توان آن را با سایر سازها هماهنگ کرد، به همین خاطر از نی بیشتر در تکنووزی‌ها استفاده می‌شود. نی به عنوان تنها ساز رسمی بادی موسیقی ما، سازیست که قابلیت تولید دو نوع صدای کاملاً متفاوت از هم را دارد. در سازهای بادی موسیقی کلاسیک معمولاً برای تغییر رنگ صدایی از وسیله‌ای به نام سوردين استفاده می‌کنند که بنا به نظر آهنگساز از این وسیله برای بیان بهتر حالت اجرای یک قطعه استفاده می‌شود. سوردين وسیله‌ای است که در ساز قرار می‌دهند و باعث می‌شود صدای ساز به نحوی محسوس تغییر کند، به طوریکه اگر شنونده ای با این وسیله و کاربرد آن آشنا نباشد، تصور می‌کند این صدا از ساز جدیدی تولید می‌شود. در سازهای ایرانی این وسیله تعریف نشده و این سازها با رنگ طبیعی خود اجرا می‌شوند که صرفاً یک نوع صداست. [۲]

۲.۱.۱ پیانو



شکل ۲.۱: ساز پیانو

پیانو یک ابزار موسیقی صوتی (آکوستیک) است که در آن سیم ها به وسیله چکش مورد اصابت قرار می گیرند. پیانو یک ساز کوبه ایست. با این که از سیم های زیادی تشکیل شده است، اما صدا با ضربه زدن چکش ها به سیم ها ایجاد می شود. بنابراین پیانو در بخش سازهای کوبه ای یک ارکستر سمفونی جای دارد. این ساز توسط یک کیبورد نواخته می شود. کلمه‌ی پیانو مخفف کلمه‌ی ایتالیایی فورته پیانو هستش، که مدل های اولیه پیانو در سال ۱۷۰۰ با این نام شناخته شده بودند. نواختن این ساز به این صورت که نوازنده با انگشتان یا انگشت شست هر دو دست کلید ها را می فشارد تا چکش ها به سیم ها ضربه بزنند و ملوಡی نواخته شود. کلمه‌ی پیانو به معنای نرم و کلمه‌ی فورته به معنای بلند است. میزان صدای تولید شده از این ساز بستگی به میزان فشار انگشت پیانیست بر کلید ، دارد: هرچه سرعت فشار بر کلید بیشتر باشد، فشار چکش نیز بیشتر می شود و صدای آهنگ تولید شده بالاتر خواهد بود. [۱]

۳.۱.۱ سنتور



شکل ۳.۱: ساز سنتور

یکی از سازهای سنتی موسیقی ایرانی ساز سنتور می باشد که با صدای دلنشیش افراد زیادی را جذب خود می کند. سنتور از جمله سازهای زهی- ضربه ای است که شکل کاملاً منظم و هندسی دارد. این ساز ساخته شده از یک جعبه تشدید صدا به شکل ذوزنقه است که ۸ تا ۱۰ سانتی متر ارتفاع دارد و تمام سطوح آن از چوب گرد و سر ساخته شده است. همچنین دو حفره به شکل گل روی سطح سنتور وجود دارد که به شفاف شدن صدای سنتور کمک می کند. دور دیف پایه چوبی به نام خرک بر روی سطح فوقانی سنتور وجود دارد که از روی آن سیم هایی عبور کرده است. هر کدام از این سیم ها وظیفه تولید دامنه معینی از صدایها با بسامد معین را دارد.

سیم های سنتور دو دسته اند: سیم های زرد که سبب تولید صدای بم می شوند و بر روی خرک های ردیف راست به تناوب قرار گرفته اند، سیم های سفید که برای تولید صدای زیر مورد استفاده قرار می گیرند و بر روی خرک های سمت چپ به تناوب قرار گرفته اند. نوازنده سنتور به وسیله دو چوب نازک به نام های «مضراب» یا «زخمه» صورت می گیرد. این چوبها یک واسط میان دست و سیم های سنتور هستند و نوازنده به وسیله آنها سنتور می نوازد.

۴.۱.۱ سه تار



شکل ۴.۱: ساز سه تار

یکی از اصیل‌ترین ساز‌های ایرانی ساز سه تار است. این ساز از جمله ساز‌های ذهنی و مضرابی موسیقی ایرانی است که با ناخن انگشت اشاره دست راست نواخته می‌شود. در گذشته این ساز را در خانواده تنبور قرار می‌دادند اما امروزه ساز سه تار را به تار نزدیک تر می‌دانند و معمولاً نوازنده‌گان تار با ساز سه تار نیز آشنایی دارند. ساز سه تار در گذشته دارای سه سیم بوده و اکنون به چهار سیم افزایش پیدا کرده است. البته سیم سوم و چهارم بسیار به هم نزدیک هستند و در هنگام اجرا به طور همزمان نواخته می‌شوند. سه تار دارای صدایی مخللی و ظرفیت بوده و از آنجایی که با کنار ناخن انگشت سبابه دست راست نواخته می‌شود، ارتباط مستقیمی با اعصاب و روان نوازنده پیدا می‌کند و از این رو سه تار را اغلب همدم اوقات تنهایی خوانده‌اند. اغلب شنونده‌گان، ساز سه تار را دارای لحن و نوای غمگینی احساس می‌کنند، اما نوازنده‌گان معاصر موسیقی ایرانی در تلاش برای توسعه موسیقی مدرن و نوی ایران، آثاری آفریده‌اند که با حال و هوایی که تا دو دهه پیش از این ساز تصور می‌شد کاملاً متفاوت است.

۵.۱.۱ تار



شکل ۵.۱: ساز تار

تار ۶ سیم دارد که به صورت جفتی کنار هم قرار می‌گیرند. جفت سیم اول از پایین که رنگ آنها سفید است (دو) کوک می‌شوند. جفت سیم دوم که در وسط قرار دارند و زرد رنگ هستند (سل) کوک می‌شوند و جفت سیم آخر دقیقاً مثل سیمهای سه تار هستند و کوک اصلی و رایج آنها (د) می‌باشد. اما نسبت به دستگاهی که قرار است در آن بنوازیم کوک آنها را تغییر میدهیم. این ساز از دو قسمت کاسه و دسته تشکیل می‌شود. کاسه که مهمترین بخش ساز است معمولاً دو تکه ساخته می‌شود. چوب مورد استفاده بیشتر از درخت توت و گاهی از درخت گردوست و البته بهترین کیفیت با چوب توت بدست می‌آید. برای ساخت آن، ابتدا دو طرف کاسه تار به صورت جداگانه تراشیده می‌شود و سپس داخل هر یک خالی شده و بعد آنها را به هم می‌چسبانند. کاسه از یک قسمت بزرگ به نام شکم و یک قسمت کوچک به نام نقاره تشکیل می‌شود. تار گستره سه اکتاو را پوشش می‌دهد و این گستره در سیم‌های سل و دو قرار دارد.

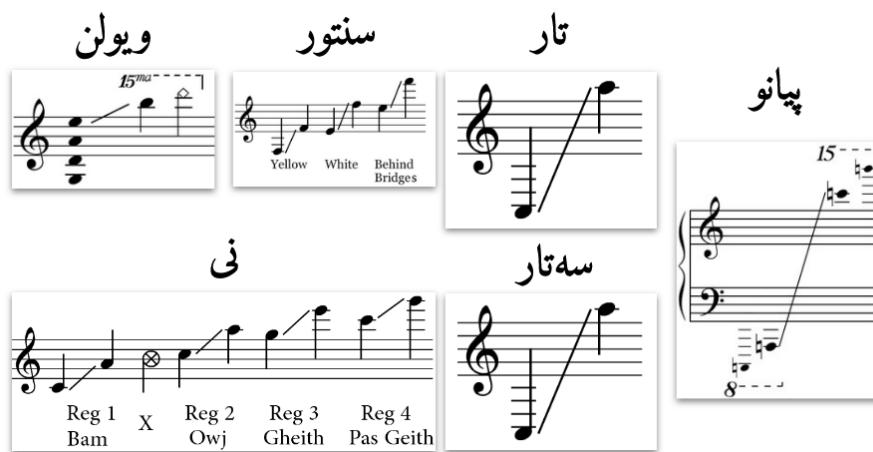
۶.۱.۱ ویولن



شکل ۱: ساز ویولن

تفاوت‌های صدای ویولن با سایر سازهای زهی در حالت کلی نتیجه ترکیب عوامل زیر است: ضخامت سیم (قطر یا گیج)، انتخاب سیم، نقطه تماس آرشه، سرعت آرشه‌کشی و فشار آرشه. سازهای زهی همگن ترین سازها در بین تمامی گروه‌های سازها در ارکستر سمفونی هستند. بدنه‌ی ویولن طوری طراحی شده است که به عنوان واسطه‌ای میان سیم و هوا عمل کند. سیم مرتعش نیرویی را که از لحظه زمانی متغیر است بر روی خرک وارد می‌کند که سبب ارتعاش تمام بدنه‌ی ساز به همراه سیم می‌شود. صدایی که ما می‌شنویم ناشی از ارتعاشات کوچکی است که در قسمت چوبی ساز ایجاد می‌شود. صدا از طریق بازآوایی و تشدیدی که در بدنه‌ی ویولن صورت می‌پذیرد، تقویت می‌شود و صفحات بزرگ و سبک (صفحات جلویی و پشتی) ساز را قادر می‌سازند تا امواج صوتی با برد قابل توجهی ایجاد کند که از این لحظه بی شباهت به بخش مخروطی شکل بلندگوها نیستند.

۲.۱ تفاوت‌ها و چالش‌ها و سختی‌ها



شکل ۷.۱: گستره صوتی سازهای مورد بررسی

گستره صوتی سازها از مهم‌ترین تفاوت‌های سازها است که برای هر طبقه‌بندی می‌تواند ویژگی مهمی باشد. در ۷.۱ گستره صوتی سازها مشخص شده است. یکی از نکات مهم این است که پیانو گستره‌ای خیلی بزرگ‌تر نسبت به سازهای دیگر دارد و با ۸۸ کلید خود می‌تواند فرکانس‌های گسترده‌ای را تولید کند. یکی از تفاوت‌های این سازها این است که برخی از آن‌ها مانند پیانو و سنترور فقط فرکانس‌های گستته مشخصی را می‌توانند اجرا کنند، در حالیکه سازهایی مانند ویولن و تار می‌توانند در گستره صوتی که دارند به صورت پیوسته هم فرکانس‌ها را تولید کنند. مانند اجرای ویراتو در سازهای زهی به این صورت است که دست چپ نوازنده، کمی بر روی سیم‌ها به جلو و عقب می‌لغزد و در نتیجه صدای حاصل از آرشه یا مضراب، به صورت متنابه، اندکی زیر و بم می‌شود. همچنین باید به جنس صدای هر کدام از این سازها اشاره داشت. در ۱.۱ طبقه‌بندی سازها مشخص شده است. مشخص است که ساز نی که یک ساز بادی است نسبت به سازهای دیگر جنس صدای متفاوتی داشته باشد چون از دمیدن هوا صدا ایجاد می‌شود. اما در تمام سازهای دیگر زهی وجود دارد که به لرزش در می‌آید و صدا تولید می‌کند. میان این سازها نیز تفاوت‌هایی در شکل به لرزش درآوردن زه است. به عنوان مثال در ویولن با استفاده از کشیدن یک آرشه صدا تولید می‌شود، در سنترور با استفاده از ضربه زدن با مضراب به روی سیم، در تار و سه‌تار با استفاده از مضراب یا کنار ناخن انگشت سبابه، و در پیانو چکش‌هایی که با مکانیزمی از طریق کلاوه‌ها به تارها ضربه می‌زنند.

از چالش‌های تفکیک این سازها قطعاً شباهت سازهای زهی به یک دگیر و مخصوصاً تفکیک تار و سه‌تار

ساز	طبقه‌بندی
نی	سازهای بادی چوبی
پیانو	ساز شستی دار که توسط چکش به صدا در می‌آید
سنتور	سازهای زهی - سازهای مضرابی
سه‌تار	سازهای زهی - سازهای مضرابی
تار	سازهای زهی - سازهای مضرابی
ویولن	سازهای زهی (آرشه‌ای)

جدول ۱.۱: تفاوت‌های سازهای مورد بررسی

خوهد بود، زیرا بسیار صدای شبیه به یک دیگر دارند، به شکلی مشابه نواخته می‌شوند و دامنه فرکانسی نسبتاً مشابهی را نیز پوشش می‌دهند. همچنین اگر با استفاده از یک ساز که گستره فرکانسی بزرگی دارد مانند پیانو، فقط در یک باز خاص و در یک کلید خاص که مشابه گستره و کلید ساز دیگری باشد نواخته شود می‌تواند گمراه کننده باشد. البته باید توجه داشت که سازهایی مثل پیانو و ویولن که غربی محسوب می‌شوند کلیدهایی با فواصل مشخص کلاسیک را دارند، در حالیکه سازهای ایرانی می‌توانند کلیدهای کاملاً متفاوتی و فواصلی مانند ربع پرده را داشته باشند که در فواصل کلاسیک دیده نمی‌شود. بنابراین فواصل خاص فرکانسی که مخصوص بعضی سازها است نیز می‌تواند گزینه خوبی برای تفکیک سازها باشد.

فصل ۲

تمیزسازی داده‌ها

۱۰.۲ استخراج ویژگی‌ها

برای استخراج ویژگی‌ها از داده صوتی نیاز است تا از یک کتابخانه مخصوص این کار استفاده شود زیرا ساختار و فشرده‌سازی این فایل‌ها پیچیدگی‌های خاص خود را دارد که نیاز نیست درگیر آن شویم. ما برای این کار از کتابخانه librosa استفاده می‌کیم که مخصوص پایتون است و ویژگی‌های مناسبی را در اختیار ما می‌گذارد. ویژگی‌های متنوعی در <https://librosa.org/doc/main/feature.html> وجود دارد که ما به طور خاص از موارد زیر استفاده می‌کنیم.

”chroma_stft“ •

ویژگی‌های مبتنی بر کرومومتری برای تجزیه و تحلیل موسیقی هستند که می‌توان صوت‌های زیر و بم را به کمک آن‌ها طبقه‌بندی کرد. این ویژگی‌ها در واقع هارمونیک و ملودیک موسیقی را به تصویر می‌کشند.

”spectral_centroid“ •

میانگین وزن دار فرکانس‌های موجود در صدا به کمک این ویژگی محاسبه می‌شود به گونه‌ای که اگر فرکانس‌های صدا در سراسر صوت یکسان باشد طیف فرکانس‌ها حول یک مرکز خواهد بود و در صورتی مه فرکانس‌های بالا در انتهای صوت وجود داشته باشند این مرکز به سمت انتهای کشیده خواهد شد.

”spectral_bandwidth” •

این ویژگی واریانس از مرکز طیف را نشان می‌دهد و متناسب با انرژی پخش شده در باندهای فرکانسی است.

”spectral_rolloff” •

خروجی این ویژگی فرکانسی است که درصد مشخصی از انرژی کل موج زیر آن قرار می‌گیرد. برای مثال فرکانسی که حدود ۸۰٪ انرژی کل طیف زیر آن قرار دارد.

”rms” •

این ویژگی مقدار rms به ازای هر فریم را باز می‌گرداند. این مقدار با توان موثر سیگنال مرتبط است.

”zero_crossing_rate” •

این ویژگی بیان می‌کند که در طول یک سیگنال نرخ تغییر علامت به چه اندازه است. یعنی نرخی که در آن سیگنال از مثبت به منفی و یا برعکس تغییر می‌کند.

”mfcc” •

یکی از مهم ترین ویژگی‌ها برای شناسایی صوت این ویژگی است. mfcc مجموعه‌ای از ویژگی‌های است که شکل کلی یک طیف را نشان می‌دهد. پس از کار با این ویژگی‌های ذکر شده دقت مناسبی حاصل نشد، که پس از بررسی بیشتر، این ویژگی را اضافه کردیم. این ویژگی تقریباً مشابه شناوی انسان عمل می‌کند و می‌تواند بازنمایی خوبی از صدا به ما بدهد. به همین دلیل این ویژگی برای طبقه‌بندی، تشخیص گفتار، طبقه‌بندی ژانر و غیره کاربرد دارد.

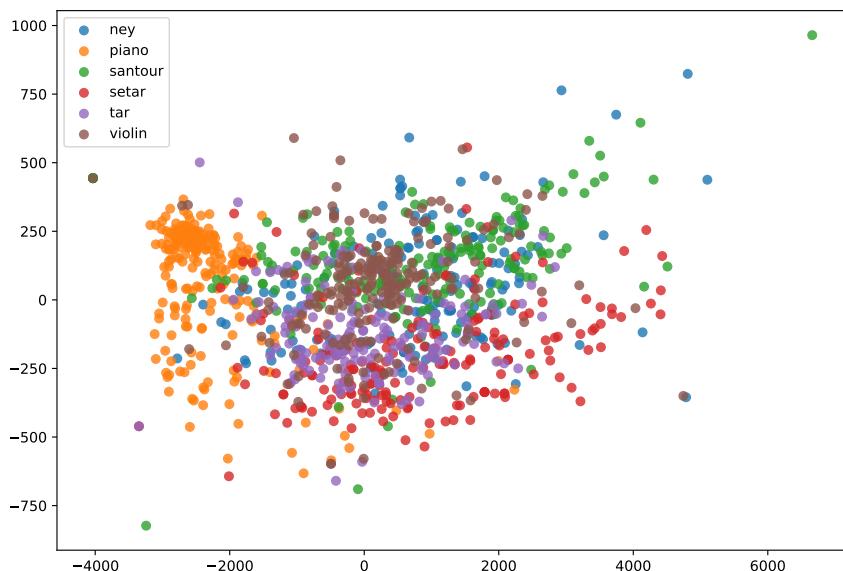
با توجه به مشکلاتی که در جمع‌آوری داده‌ها بوده از هر آهنگ، از بعد از ۶۰ ثانیه آن صدا برداشته می‌شود تا اگر هنوز آهنگ شروع نشده بود مشکلی ایجاد نشود. همچنین برای کم کردن هزینه محاسباتی، به طول ۳۰ ثانیه از هر آهنگ برداشته و پردازش می‌شود. در این مسیر بعضی از داده‌ها قابلیت پردازش نداشتند که به صورت مناسبی کنار گذاشته شدند و به این ترتیب داده تمیزسازی شد.

پس از این مراحل، به شکلی مناسب در یک فایل pickle ذخیره‌سازی شد که می‌توان به راحتی آن را خواند و با آن کار کرد. در این فایل علاوه بر ویژگی‌های استخراج شده، نام هر کدام از آن‌ها، نام فایل اصلی آن، کلاس آن ذخیره شده است.

		data	target	data_paths
0	[0.16597863, 2516.9022504985023, 2215.20888780...	0	/content/ML_data_G2/ney/132.mp3	
1	[0.16874595, 1968.798443479041, 2197.522508426...	0	/content/ML_data_G2/ney/99.mp3	
2	[0.17199844, 2048.163235103862, 1937.130002929...	0	/content/ML_data_G2/ney/8.mp3	
3	[0.23767014, 1717.2538157279628, 1829.60675558...	0	/content/ML_data_G2/ney/50.mp3	
4	[0.23016322, 1429.2651745338544, 1829.95074745...	0	/content/ML_data_G2/ney/210.mp3	
...	
1420	[0.25593293, 1271.0261494871563, 1490.26726038...	5	/content/ML_data_G2/violin/45.mp3	
1421	[0.17380463, 2120.149764892853, 1917.669420651...	5	/content/ML_data_G2/violin/261.mp3	
1422	[0.26513362, 1126.9594266773263, 1520.30768744...	5	/content/ML_data_G2/violin/115.mp3	
1423	[0.23881607, 2181.158352387854, 1826.625519289...	5	/content/ML_data_G2/violin/180.mp3	
1424	[0.2077015, 1969.5485336204406, 1667.757390786...	5	/content/ML_data_G2/violin/191.mp3	

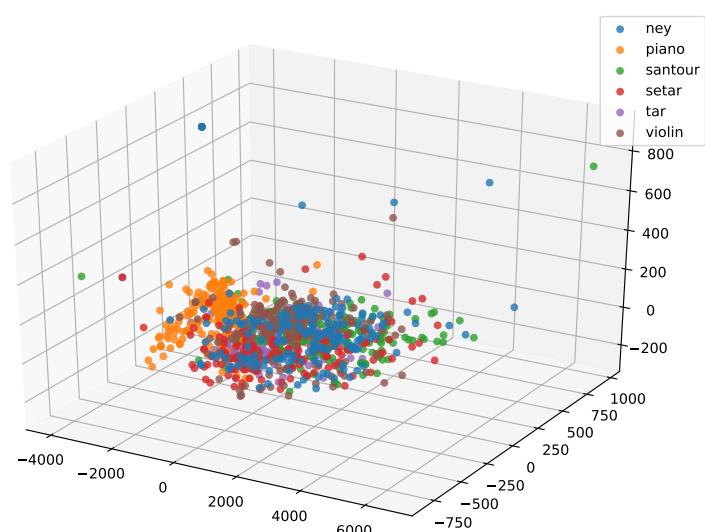
PCA ۲.۲

روش PCA با ترکیب فیچر ها سر و کار دارد و این قدرت را دارد که این ترکیب را انجام دهد، در صورتی که روش‌های دیگری در درس هم گفته شد مانند forward/backward selection که تنها می‌توانستند استفاده از یک ویژگی را تصمیم‌گیری کنند نه ترکیب چند ویژگی در یک بعد جدید. با روش PCA و با تعداد Component برابر ۲ و ۳ ویژگی‌ها را کاهش دادیم که در زیر آمده است.



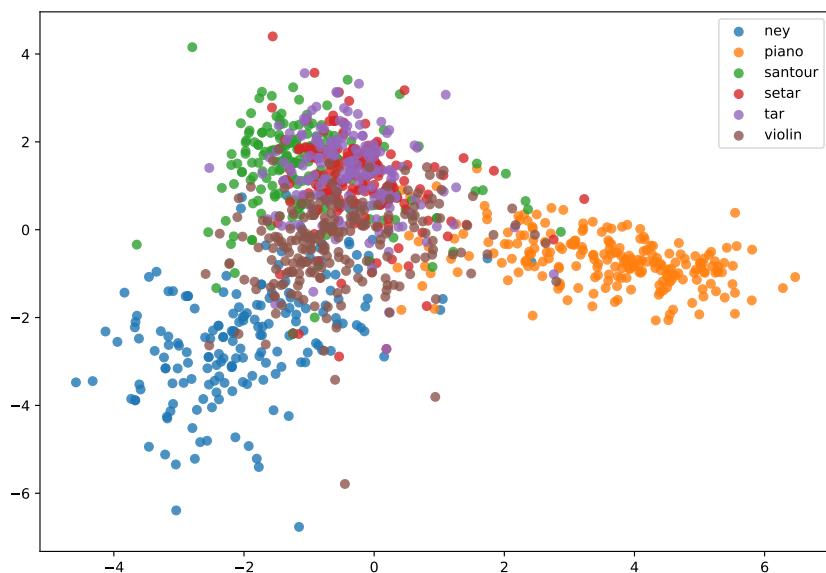
همانطور که دیده می‌شود، با استفاده از روش PCA کلاس‌ها چندان از هم جدا نشدند. علت این اتفاق این است که PCA سعی می‌کند تا واریانس تصویر داده‌ها بر روی ابعاد جدید بیشینه شود که معادل بیشینه کردن امکان بازگردانی داده‌ها به فضای اصلی است. این هدف لزوماً کلاس‌ها را جدا نمی‌کند، و به همین دلیل دیده می‌شود که اگرچه سعی شده داده‌ها بیشترین جدایی را داشته باشند، اما باز کلاس‌ها نزدیک هم افتاده‌اند. شکل ۳ بعدی PCA نیز در زیر آمده است.

و باز هم دیده می‌شود که واریانس سعی شده بیشینه شود اما کلاس‌ها خیلی خوب جدا نشده‌اند که طبیعی است. برای اینکه کلاس نیز در کاهش بعد در نظر گرفته شود باید از روش دیگری استفاده کنیم که در بخش بعدی توضیح داده می‌شود.



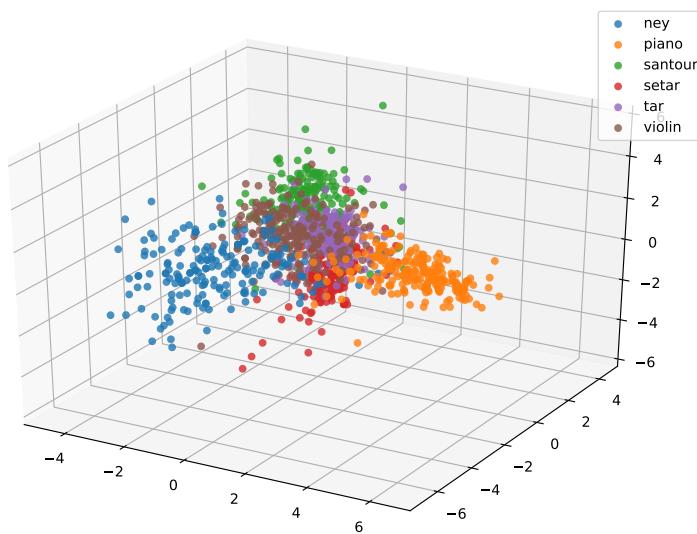
LDA ۳.۲

در کنار روش PCA و که می‌توان دو فیچر را مثلاً با هم ترکیب کرد. این قابلیت می‌تواند نتایج بهتری را برای ما فراهم کند روش LDA نیز این قابلیت و برتری را دارد. اما تفاوت با PCA این است که در کاهش بعد LDA کلاس‌ها نیز تاثیرگذار هستند و سعی می‌شود تا کلاس‌ها بیشترین جدایی را داشته باشند. حالت ۲ بعدی LDA در زیر آمده است.



همانطور که دیده می‌شود، نتیجه این روش نسبت به PCA جدایی خیلی بهتری را بین کلاس‌ها برای ما فراهم می‌کند. بنابراین استفاده از این روش برای ما مناسب‌تر است.
حالت ۳ بعدی LDA در زیر آمده است.

این حالت هم تمایز بیشتر کلاس‌ها نسبت به روش PCA را نشان می‌دهد که طبق تعاریف توضیح داده شد. نکته جالب توجه این است که در این بازنمایی می‌توانیم بینیم که اولاً پیانو توانسته به راحتی از بقیه کلاس‌ها جداسازی شود. این مورد را می‌توان به علت تفاوت شاخص گستره صوتی و نوع صدای این ساز دانست. همچنین ساز نی نیز به خوبی جدا شده که می‌توان علت آن را اینکه تنها ساز بادی بین سازهای مورد بررسی است دانست. از طرفی دیده می‌شود که تار و سه‌تار و سنتور خیلی مشابه هم شده‌اند. این مورد هم طبق فصل اول این نوشتار



قابل انتظار بود چون ساختار و صدای این سازها خیلی مشابه هم هستند و طبیعتاً ویژگی‌های نزدیکتری به هم دارند. در مجموع طبق مشاهده روش ۳ بعدی می‌تواند جداسازی بهتری نسبت به روش دو بعدی بدهد که نشان می‌دهد تعداد ویژگی‌ها اهمیت زیادی دارد.

فصل ۳

طبقه‌بندی

در این بخش داده‌ها را با نسبت ۸۰ به ۲۰ به دو دسته آموزش و تست تقسیم‌بندی کردیم. این کار با استفاده از train_test_split انجام شد. در ادامه با طبقه‌بندهای متفاوت این داده را طبقه‌بندی کرده و دقت و نتایج روی بخش تست را بررسی و تحلیل می‌کنیم.

Gaussian Naive Bayes ۱.۳

در ابتدا با استفاده از Naive Bayes و فرض نرمال بودن توزیع احتمال داده‌ها به شرط کلاس این دیتاست را طبقه‌بندی می‌کنیم.

$$P(\text{class} | \text{data}) = \frac{P(\text{data} | \text{class}) \times p(\text{class})}{p(\text{data})}$$

"Gaussian" because this is a normal distribution

This is our prior belief

We don't calculate this in naive bayes classifiers

ChrisAlbon

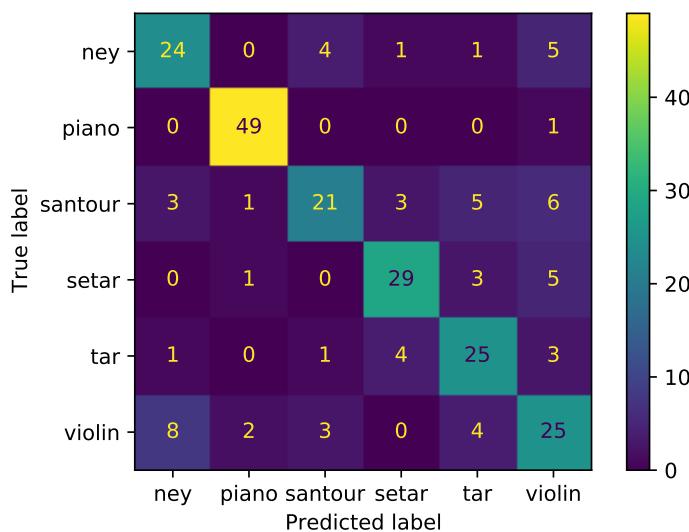
نتایج این دسته‌بندی به شکل زیر شد.

Class	Precision	Recall	F-score	Support
ney	0.67	0.69	0.68	35
piano	0.92	0.98	0.95	50
santour	0.72	0.54	0.62	39
setar	0.78	0.76	0.77	38
tar	0.66	0.74	0.69	34
violin	0.56	0.6	0.57	42
accuracy	-	-	0.73	238
macro avg	0.72	0.72	0.71	238
weighted avg	0.73	0.73	0.72	238

در جدول مشخص است که دقت حدود ۷۳ درصد قرار دارد و همچنین می‌توان دقت هر دسته را به صورت

می‌جزا مشاهده کرد. به نظر می‌رسد که تشخیص پیانو برای مدل از همه سازهای دیگر ساده‌تر بوده و بیشترین دقت را کسب کرده است که با توجه به گستره صوتی و تفاوت قابل توجه این ساز با دیگر سازها دور انتظار نبود. اما این فرض نرمال گوسی بودن توزیع به نظر لزوماً صحیح نیست و باعث شده است دقت به اندازه مطلوب بالا نباشد که در بخش‌های بعدی پیشرفت را خواهیم دید.

و همچنین confusion matrix در زیر نمایش داده شده است. به ازای هر ساز می‌توان دید که چه سازی



شکل ۱.۳ Gaussian Naive Bayes :

بیشتر با آن ساز اشتباه گرفته شده است. به عنوان مثال نی با ویالون بیشتر از بقیه اشتباه گرفته شده است که می‌توان از روی صدای این دو که ثبوت دارند و متفاوت از ضربه ای بودن باقی سازها است فهمید. پیانو و ویالون با هم اشتباه گرفته شده‌اند که هم می‌توانند بخاطر اشتباه داده باشد که شاید یکی از داده‌ها فقطعه‌ای بوده باشد که هم پیانو داشته هم ویالون، و خطای جمع‌آوری داده باشد، اما اگر این نباشد هم منطقی است چون پیانو و ویالون تنها سازهای کلاسیکی هستند که در این مجموعه داده داریم و بقیه سازهای ایرانی هستند که گام‌ها و فواصل متفاوتی نسبت به سازهای کلاسیک دارند. از طرفی تار و سه تار و سنتور با هم بیشتر اشتباه گرفته شده اند که این را هم از روی شباهت صدا می‌توانستیم حدس بزنیم.

در مجموع، نیاز به طبقه‌بند قوی تری داریم که در بخش‌های بعدی بررسی می‌کنیم.

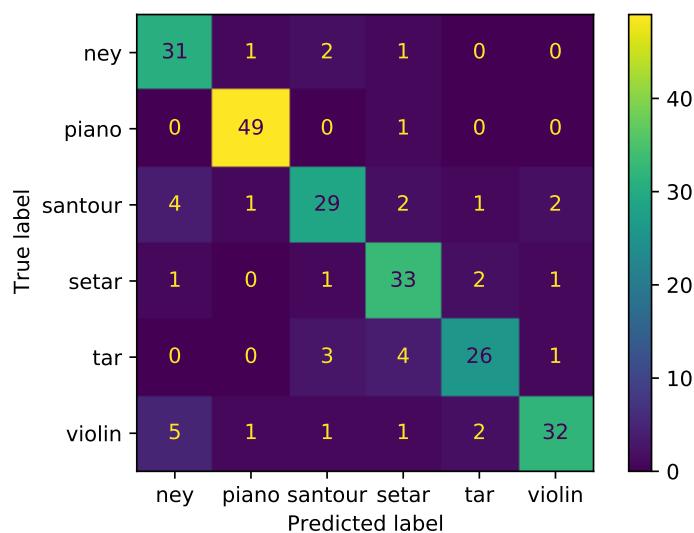
Linear SVM ۲.۳

در این بخش از SVM [۳] که در درس خواندیم استفاده می‌کنیم. این روش سعی می‌کند تا margin را بیشینه کند. در ابتدا از حالت خطی این طبقه‌بند استفاده می‌کنیم.

$$k(x, x') = x^T x'$$

Class	Precision	Recall	F-score	Support
ney	0.76	0.89	0.82	35
piano	0.94	0.98	0.96	50
santour	0.81	0.74	0.77	39
setar	0.79	0.87	0.82	38
tar	0.84	0.76	0.8	34
violin	0.89	0.76	0.82	42
accuracy	-	-	0.84	238
macro avg	0.84	0.83	0.83	238
weighted avg	0.84	0.84	0.84	238

طبق نتایج می‌بینیم که توانستیم به دقت ۸۴ درصد برسیم. به جز پیانو که در بخش قبل هم دقت خوبی داشت، در این طبقه‌بند توانستیم دقت باقی سازها را نیز افزایش دهیم که در نهایت باعث بهبود نتیجه کل شده است. همچنین در ماتریس نیز بهبود دیده می‌شود. نتایج بسیار مشابه قبل است با کمی بهبود و باز هم دیده می‌شود که نی با ویولن بیشتر اشتباه گرفته می‌شود و باز هم بین تار و سه تار و سنتور اشتباهات بیشتری داریم. در مجموع با اینکه بهبود در نتایج دیدیم ولی فرض جداسازی خطی محدود کننده است و می‌توانیم با فرض‌های بهتر دقت را بهتر کنیم که در ادامه بررسی می‌شود.



Linear SVM : ۲.۳

RBF SVM ۳.۳

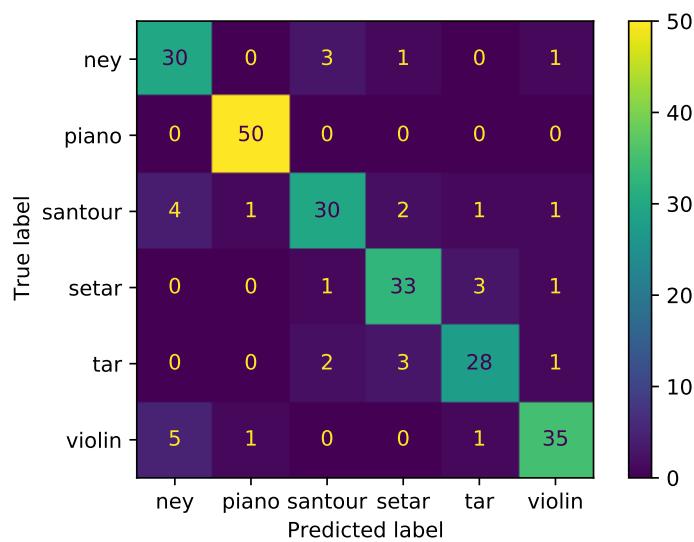
برای اینکه فرض خطی را کنار بگذاریم، این بار از SVM با کرنل به شکل RBF یا Radial Basis Function استفاده می‌کنیم که امکان فرض شعاعی را به جای خطی می‌دهد و دست طبقه‌بند را بازتر می‌کند.

$$k(x, x') = \exp\left(\frac{-(x - x')^2}{2\gamma^2}\right)$$

Class	Precision	Recall	F-score	Support
ney	0.77	0.86	0.81	35
piano	0.96	1.0	0.98	50
santour	0.83	0.77	0.8	39
setar	0.85	0.87	0.86	38
tar	0.85	0.82	0.84	34
violin	0.9	0.83	0.86	42
accuracy	-	-	0.87	238
macro avg	0.86	0.86	0.86	238
weighted avg	0.87	0.87	0.87	238

در نتایج می‌بینیم که دقت به ۸۶ افزایش پیدا کرد که نشان می‌دهد فرض خطی واقعاً برای این دیتا محدود کننده بوده و یک فرض شعاعی می‌تواند مرزهای تصمیم بهتری ایجاد کند.

در ماتریس هم بهبود دیده می‌شود و همچنین تشابه‌های سازها بر این اساس که قبل از توضیح داده شد اینجا هم هست. نکته جدیدی که می‌توان گفت توجه به بخش‌هایی مثل اشتباه گرفته شدن سنتور با نی است. در این مورد احتمال دارد مشکل از جمع آوری داده باشد و همچنین سازی در این موارد بوده باشد، یا حتی صدایی مشابه ولی جدای از سنتور بوده باشد که باعث شده نی تشخیص داده شود. در کل به علت تفاوت مشخص این دو صدا، احتمال اشتباه در داده بیشتر است، ولی باز هم امکان دارد مدل هنوز به اندازه کافی قوی نباشد.

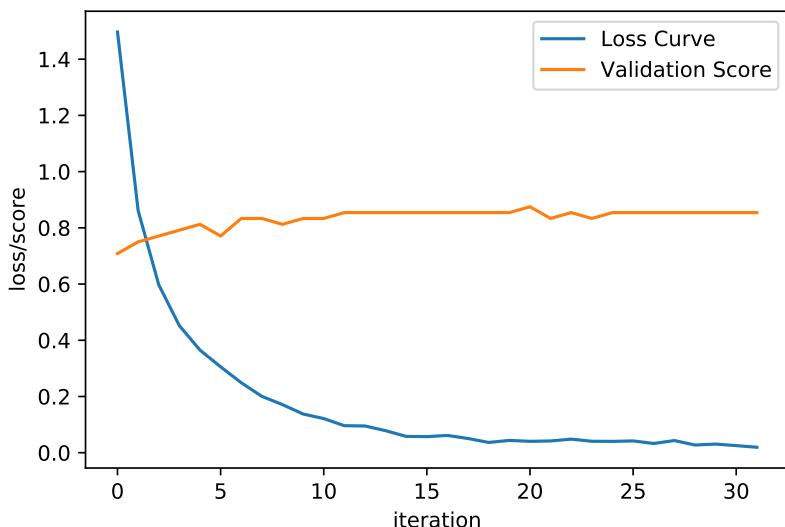


شكل ۳.۳ RBF SVM :

در مجموع مدل خوبی ساخته شد، اما هنوز هم جای بهبود داریم که در بخش بعدی با یک شبکه عصبی این کار را انجام می‌دهیم.

MLP ۴.۳

در این بخش وارد شبکه های عصبی [۴] می‌شویم که می‌تواند به صورت کلی و با پوشش بیشتری مرزها را بسازند و برای هر مسئله‌ای می‌تواند گزینه مناسبی باشند. به صورت خاص از Multi-layer Perceptron classifier که تابع log-loss را با استفاده از LBFGS یا نزول گرادیان تصادفی بهینه می‌کند استفاده می‌کنیم. از نظر ساختار لایه ها را با $256, 512, 512, 256$ نورون به ترتیب می‌سازیم که ساختار مورد استفاده‌ای است و به صورت کلی ابتدا از ابعاد کوچک به ابعاد بزرگتر رفته و سپس دوباره کاهش بعد می‌یابد. برای این طبقه‌بند، همچنین از early_stopping=True استفاده کردیم که باعث می‌شود داده آموزش به دو بخش آموزش و validation تقسیم شود و در هر iteration روی آموزش یادگیری انجام شده و با استفاده از بخش validation یک امتیاز داده شود و پس از مدتی اگر این امتیاز پیشرفت نکند و یا حتی بدتر شود ولی هنوز loss آموزش کاهش یابد یعنی به overfit رسیده‌ایم که مدل در اینجا اصطلاحاً early_stopping انجام می‌دهد و مدل بهینه را قبل از این اتفاق دهد. نمودار loss آموزش و امتیاز بخش validation در زیر آمده است.

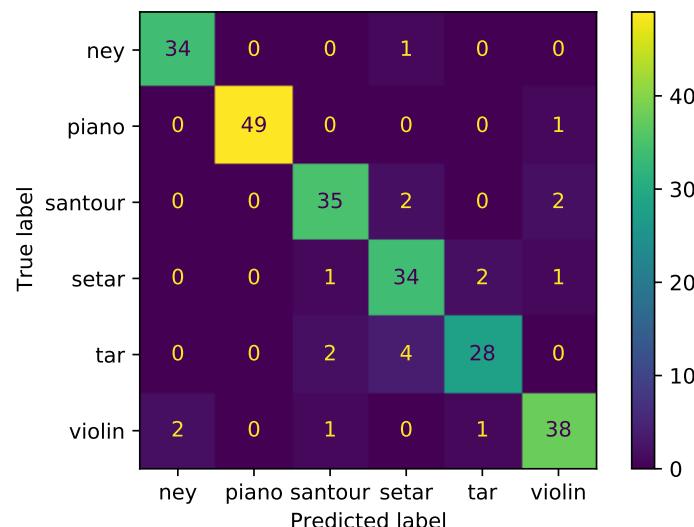


این بررسی بر روی داده validation انجام می‌شود چون تعداد iteration آموزش مدل رانیز می‌توان هایپرپارامتر در نظر گرفت و نباید برای تنظیم آن از داده تست که اصلاً نباید دیده شود استفاده کنیم. بنابراین به همین شکل یک بخش validation در نظر گرفتیم صحیح ترین کار است و سپس مدل نهایی را روی بخش تست هم آزمودیم که در ادامه توضیح داده می‌شود.

نتایج نهایی در زیر آمده است.

Class	Precision	Recall	F-score	Support
ney	0.94	0.97	0.96	35
piano	1.0	0.98	0.99	50
santour	0.9	0.9	0.9	39
setar	0.83	0.89	0.86	38
tar	0.9	0.82	0.86	34
violin	0.9	0.9	0.9	42
accuracy	-	-	0.92	238
macro avg	0.91	0.91	0.91	238
weighted avg	0.92	0.92	0.92	238

در نتایج می‌بینیم که دقت به ۹۲ درصد رسیده است که بهترین دققی است که بین تمام طبقه‌بندها توانستیم بدست بیاوریم و قدرت شبکه‌های عصبی را نشان می‌دهد.



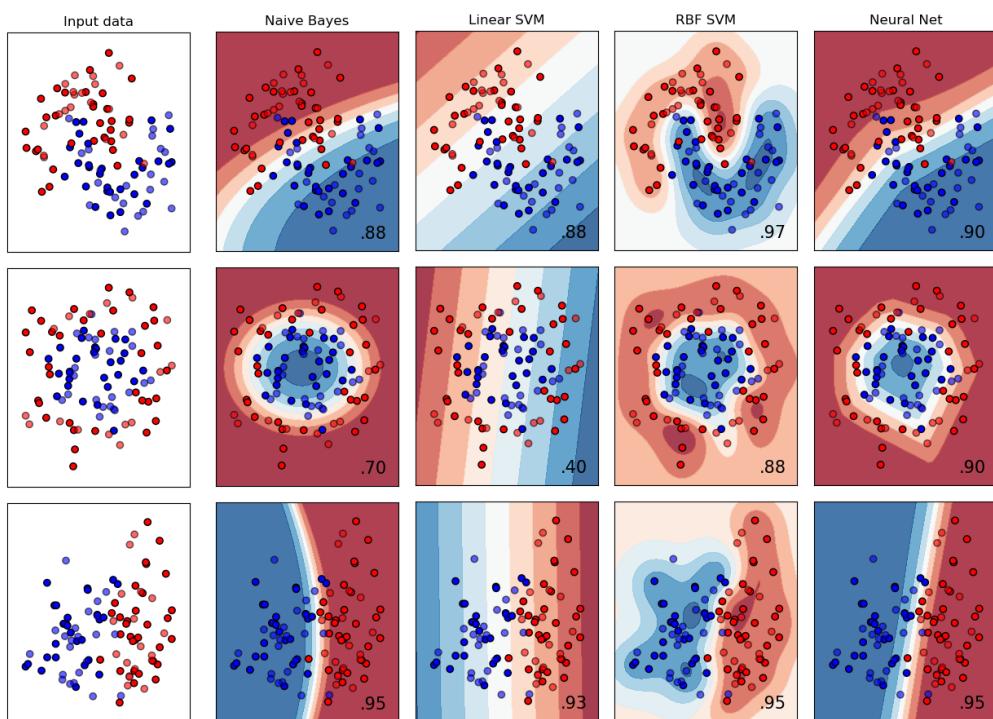
شکل ۴.۳ MLP :

در ماتریس هم بهبود دیده می‌شود. تنها اشتباہات باقی مانده مربوط به تشابه نی و ویولن، همچنین تشابه سنتور و سه تار و تار است که مشخصا در یک مربع ۳ در ۳ دیده می‌شود که خطأ متراکم‌تر در این بخش است و این تشابه از ابتدا هم اشاره شده بود که می‌تواند طبقه‌بند را به دردرس بیاندازد.

در مجموع آخرین مدلی که با MLP ساختیم توانست بهترین نتایج را با دقت ۹۲ درصد به دست آورد که دقت بسیار خوبی است و نشان می‌دهد مسیری که طی کردیم درست بوده و در هر مرحله با استفاده از طبقه‌بند قدرتمندتر توانستیم بهبود ایجاد کنیم و در انتها با شبکه عصبی که در کلاس هم گفته شد قدرت مثال زدنی دارد و تحولی ایجاد کرده است توانستیم بهترین نتایج را بگیریم.

۵.۳ جمع‌بندی و مقایسه

در نهایت در شکل زیر انواع نتیجه طبقه‌بندهایی که استفاده کردیم دیده می‌شود.



در این تصویر مشخص است که روش گاوسی بیز که استفاده کردیم سعی می‌کند برای توزیع هر کلاس یک توزیع نرمال در نظر بگیرد که دیدیم در دیتا ما مناسب نبود. از طرفی دیده می‌شود که روش خطی SVM مرزهای تصمیم خطی را ایجاد می‌کند که باز هم در دیتا ما دیدیم که این فرض خطی جدا شدن کلاس‌ها مناسب نبود. سپس از روش SVM با کرنل RBF شعاعی استفاده کردیم که هم در شکل دیده می‌شود که انعطاف بیشتری دارد و هم در نتایج دیدیم که نسبت به فرض خطی بهتر عمل کرد. و در نهایت با استفاده از MLP چندلایه توانستیم مرزهای پیچیده‌تری را همچنان که در شکل هم مشخص است ایجاد کنیم و توانستیم نتایجی بهتر از باقی طبقه‌بندها بیینیم. بنابراین در مجموع می‌بینیم که هر چه طبقه‌بند ما انعطاف بیشتری در مرزبندی به ما بدهد، می‌توانیم نتایج بهتری را بگیریم. البته باید واریانس و بایاس را در نظر داشت و از مدل بیش از حد پیچیده هم هر جایی استفاده نکرد و با دقت مدل را انتخاب کرد.

فصل ۴

خوشه بندی

در این بخش داده های به دست آمده در قسمت های قبلی به کمک روش خوشه بندی جدا می شوند. برای خوشه بندی نیز از دو روش K-Means و Agglomerative استفاده شده است.

خوشه بندی یکی از روش های unsupervised است و معمولاً از آن برای پیدا کردن یک پترن خاص درون داده ها استفاده می شود. در ادامه دور روش خوشه بندی جهت دسته بندی صوت های موسیقی به کار برده شده است و به مقایسه ای این دو روش پرداخته می شود.

K-Means ۱.۴

الگوریتم k-means تلاش می کند تا مرکز دسته ها را به گونه ای بیابد که مجموع فاصله ای نقاط هر کلاستر از مرکز دسته ای آن کمینه شود:

$$\sum_{i=0}^n \min_{\mu_j \in C} (\|x_i - \mu_j\|^2)$$

ابن الگوریتم در دو مرحله انجام می شود:

۱. در مرحله ای اول مراکز دسته ها انتخاب می شوند. ساده ترین و معمول ترین روش این است که k نمونه

از این داده ها به عنوان مراکز دسته انتخاب شوند.

۲. پس از انتخاب مراکز دسته k-means دو مرحله زیر را تا جایی ادامه می دهد که دیگر مقادیر ایجاد شده

تغییر نکنند:

۱.۲ اختصاص هر نمونه به نزدیک ترین مرکز دسته.

۲.۲ به روز رسانی مرکز ها و ایجاد مرکز جدید با میانگین گیری بر روی نمونه های موجود درون هر

کلاستر.

یکی از مشکلاتی که این روش با آن مواجه است این است که نتایج شدیداً به انتخاب اولیه مرکز دسته ها وابسته است. علاوه بر این الگوریتم گفته شده با این فرض جلو می رود که دسته ها به صورت convex و isotropic هست که همواره فرض درستی نیست. هم چنین ضعف این روش در ابعاد بالا نیز دیده می شود به همین جهت بهتر است پیش از اجرای این الگوریتم از روش های کاهش بعد مانند PCA و یا LDA استفاده شود.

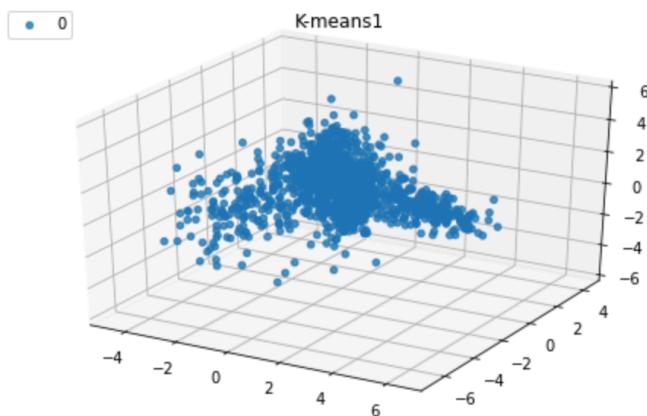
همان طور که در فصل دوم گفته شد پس از امتحان روش های متفاوت بالانجام LDA در سه بعد نتایج بهتری به دست آمد. در ادامه نتایج حاصل از خوشه بندی با تعداد خوشه های ۱ تا ۶ برای داده های به دست آمده آورده شده است.

در هر کدام از تصاویر ابتدا contingency matrix و سپس نمودار به دست آمده از خوشه بندی به کمک مولفه های LDA نشان داده شده است. هر سطر در contingency matrix نشان دهنده لیبل های واقعی و ستون ها برابر کلاستر های اختصاص داده شده است. برای مثال داده های موجود در سطر اول بیان می کند که توزیع

داده ها با لیبل ۰ درون کلاستر های اختصاص داده شده به چه شکل است. هر کدام از سطر ها به ترتیب از بالا به پایین بیانگر ساز های نی، پیانو، سنتور، سه تار، تار و ویولن می باشد.

$$k=1 . 1$$

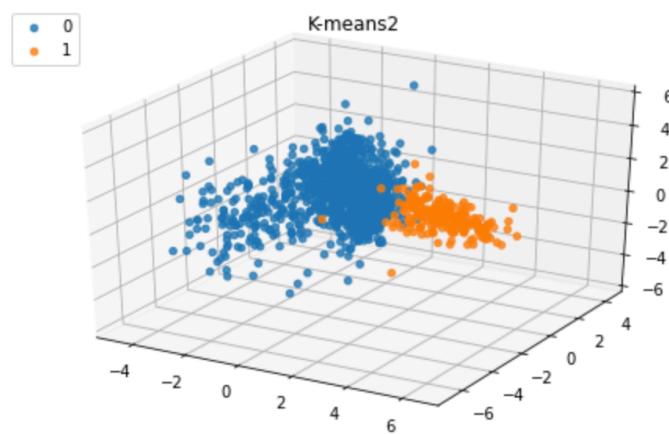
```
K-means1
[[183]
[220]
[200]
[194]
[175]
[214]]
```



هنگامی که k برابر یک است همه ی داده ها تنها در یک دسته قرار میگیرند و ماتریس به دست آمده تنها بیان میکند که در هر دسته چه تعداد داده وجود داشته است. برای مثال می توان گفت تعداد داده ها با لیبل نی برابر ۱۸۳ و با لیبل پیانو برابر ۲۲۰ است.

$k=2$. ۲

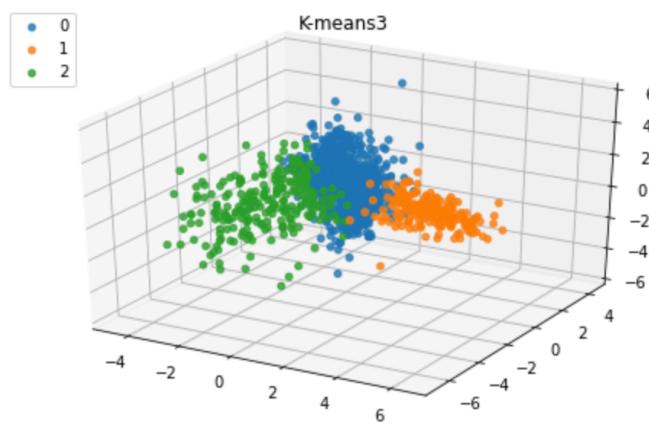
```
K-means2
[[181    2]
 [ 19  201]
 [195    5]
 [191    3]
 [173    2]
 [205    9]]
```



مشاهده می شود با قرار دادن داده ها تنها در دو دسته ساز پیانو به طور قابل توجهی در دسته های متفاوت از ساز های دیگر قرار گرفته است. می توان گفت این ساز به نسبت ساز های دیگر تفاوت بیشتری داشته است. این تفاوت احتمالاً به این علت بوده که ساز پیانو نسبت به سازهای دیگر گستره خیلی بزرگتری دارد که به طور قابل توجهی آن را از سایر ساز ها جدا کرده است.

k=۳ . ۳

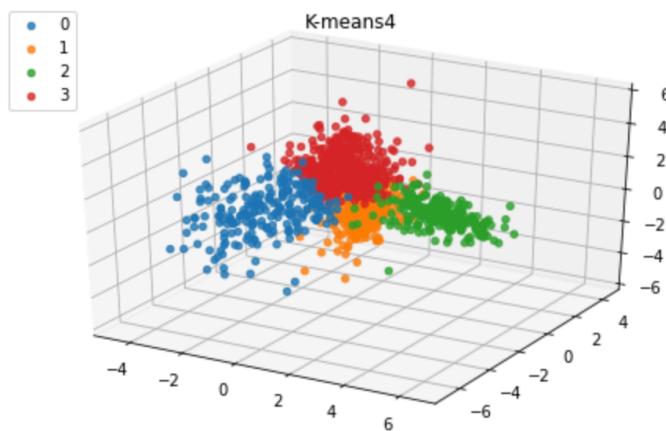
```
K-means3
[[164 16 3]
 [ 3 16 201]
 [ 12 184 4]
 [ 8 184 2]
 [ 4 169 2]
 [ 70 135 9]]
```



مشاهده میشود با افزایش تعداد دسته ها به ۳ ، ساز نی نیز از سایر ساز ها جداسده و درواقع چهار ساز سنتور، تار، سه تار و ویولن در یک دسته قرار میگیرند. ممکن است جدا سازی ساز نی در کلاسترینگ به این علت باشد که این ساز تنها ساز بادی موجود میان سایر ساز هاست و مدل به درستی توانسته است این تفاوت را تشخیص دهد و الگوی جداگانه ای برای آن بیابد.

$k=4$. ۴

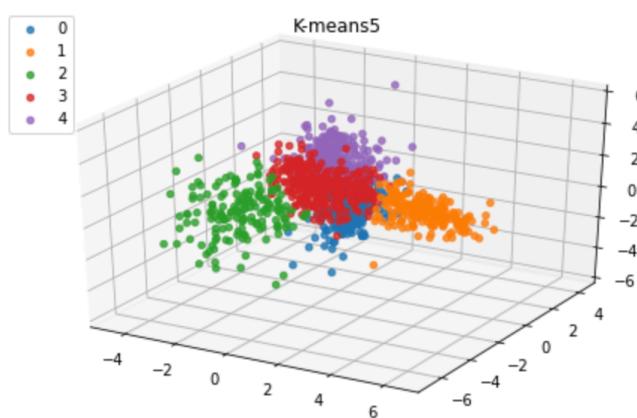
```
K-means4
[[ 4   12    5  162]
 [202   15    2    1]
 [ 4   10  178    8]
 [ 3  170   16    5]
 [ 2   81   88    4]
 [11   23  138   42]]
```



ماتریس به دست آمده نشان می دهد دو ساز ویولون و سنتور در یک دسته قرار گرفته اند. همچنین اکثر داده های مربوط به تار در دسته های مربوط به سه تار و سنتور قرار گرفته اند. همان طور که در بخش های پیشین نیز گفته شد داده های تار، سه تار و سنتور شباهت های زیادی به هم دارند به طوری که در طبقه بندی زده شده در قسمت قبل نیز این شباهت دیده می شود و سه ساز گفته شده با هم اشتباه گرفته می شدند. علت این موضوع نیز می تواند مربوط به قرار گرفتن این ساز ها در دسته های زهی-مضربایی باشد. اما قرار گرفتن اکثر داده های ویولون و سنتور در یک دسته می تواند یا ناشی از خطای داده ها و وجود آهنگ های همنوازی باشد که شامل هر دو این ساز ها بوده اند و یا ممکن است به علت وجود آهنگ های سنتی میان داده های سنتور و ویالون، این دو ساز با هم در یک کلاستر قرار گرفته باشند. اکثر داده های مربوط به پیانو و نی نیز به درستی جدا شده اند و هر کدام در دسته های مجزا قرار گرفته اند.

$k=5$.

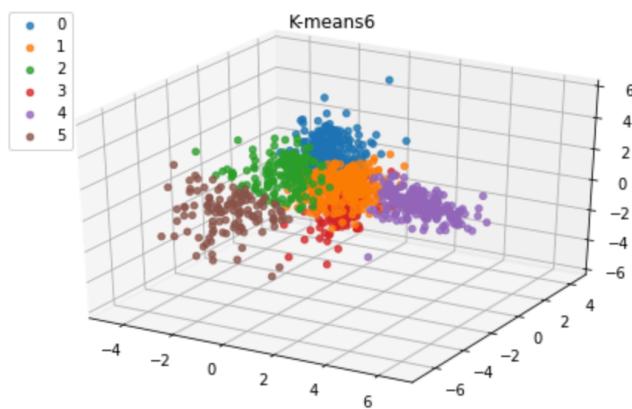
```
K-means5
[[ 32 142   2   2   5]
 [ 18    0   0 197   5]
 [ 34    7 149   4   6]
 [ 15    3  11   2 163]
 [ 35    2  78   2  58]
 [163   15  23   7   6]]
```



هنگامی که دسته بندی را با تعداد کلاستر ۵ انجام می دهیم اکثر داده های مربوط به ساز های یکسان در یک کلاستر قرار میگیرند و مشاهده میشود همهی ساز ها یه جز تار با اختلاف زیادی در یک کلاستر تجمع پیدا کرده اند(هر کدام از آن ها در یک کلاستر مجزا). بیشتر تجمع داده های تار نیز که مربوط به اعداد ۷۸ و ۵۸ در سطر چهارم است درون همان کلاستر های مربوط به ستور و سه تار است که می تواند به علت شباهت تار به این ساز ها وزهی -مضارایی بودن آنان باشد.

k=6

```
K-means6
[[ 20   2   1   4   46 110]
 [ 23   0 196   0   1   0]
 [ 28 146   3   4  18   1]
 [ 37   7   2 144   2   2]
 [ 79   57   2  32   5   0]
 [ 95  23   5   1  88   2]]
```



اعداد مشاهده شده در سطر ها نشان می دهد که تا سطر چهارم که در برگیرنده تمام ساز ها به جز تار و ویالون است دسته بندی با دقت خوبی انجام شده است و مدل توانسته است الگوهای میان این ساز ها را به درستی استخراج کرده و هر کدام را در یک دسته مجزا قرار دهد. برای ساز تار دو عدد ۷۹ و ۵۷ نزدیک به هم هستند. عدد ۷۹ در زیر آن که مربوط به ساز ویالون است نشان می دهد که اکثر داده های ویالون و تار در یک کلاستر مشترک قرار گرفته اند. قرار گیری اکثر داده های این دو ساز در یک کلاستر مشترک می تواند ناشی از تم سنتی موجود در این سازها باشد اما وجود خطأ در داده ها و یا از دست دادن بخشی از اطلاعات به دلیل کاهش بعد نیز می تواند علت قرار گیری این دو ساز در یک کلاستر مشترک باشد.

به طور کلی می توان گفت با توجه به ضعف های موجود در k-means و فرض های گرفته شده توسط آن بر روی شکل کلاستر ها، شاید روش فوق حتی با وجود استفاده از LDA نتایج خیلی خوبی بر روی داده ها ندهد اما با این حال همچنان مشاهده می شود که در تعداد دسته های ۵ و یا ۶ مدل با دقت خوبی موفق به پیدا کردن پترن ها و الگوهای موجود در صوت های ناشی از سازهای متفاوت و ایجاد تمایز میان آن ها شده است.

Agglomerative ۲.۴

یک روش خوش بندی سلسله مراتبی است که از رویکرد پایین به بالا یا bottom up استفاده می کند. در ابتدا هر نمونه به عنوان یک خوش در نظر گرفته می شود و سپس خوش ها به طور متوالی با یکدیگر ادغام می شوند. جهت ادغام خوش ها از رویکردهای متفاوتی استفاده می شود که در ادامه به آن ها اشاره شده است:

Linkage Single •

نزدیک فاصله میان دو نمونه از دو کلاستر متفاوت به عنوان فاصله در نظر گرفته می شود.

Linkage Complete •

دورترین فاصله میان دو نمونه از دو کلاستر متفاوت در نظر گرفته می شود.

Centroid •

فاصله میان مرکز های دو کلاستر به عنوان فاصله جدید تعریف می شود.

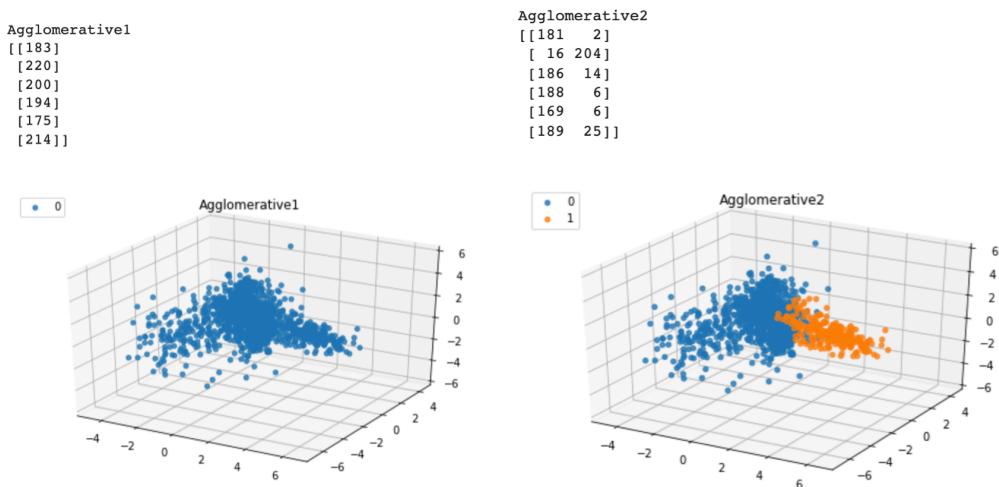
Linkage Average •

میانگین فاصله میان هر دو نمونه از دو کلاستر متفاوت برای فاصله در نظر گرفته می شود.

dendrogram برای فهم بیشتر داده و امکان خوش بندی در سطوح مختلف اشاره کرد. اما برای استفاده از این روش لازم است تا ماتریس فاصله و تعریف فاصله ای میان دو کلاستر به درستی انتخاب شود. برای مثال در صورتی که ویژگی ها از دو نوع مختلف عددی و کمی باشند متریک فاصله برای آن ها چگونه باید تعریف شود؟ علاوه بر این، این روش در صورت گم شدن تعدادی از اطلاعات نمونه ها به خوبی عمل نمی کند و هزینه محاسباتی بالایی دارد.

در ادامه نتایج حاصل از خوش بندی با تعداد خوش های ۱ تا ۶ و با استفاده از این روش آمده است که راجع به هریک توضیح داده می شود .

k=1، 2 . 1

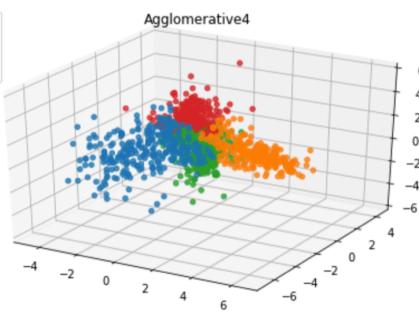
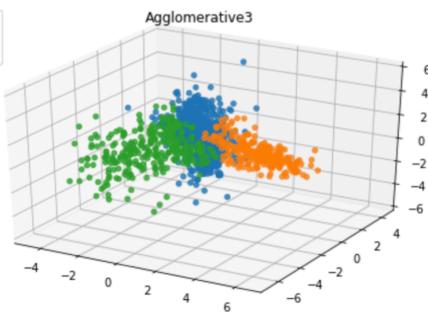


مشابه قسمت K-Means در اینجا نیز خوشه بندی با تعداد ۱ خوشه معنای خاصی ندارد و ماتریس به دست آمده تنها نشان می‌دهد که داده‌های مربوط به هر کلاس چه تعداد هستند. در صورتی که تعداد خوشه‌ها برابر ۲ باشد مجددا مشاهده می‌شود که داده‌های پیانو اکثرا در یک دسته و داده‌های مربوط به سایر سازها در دسته‌ی دیگر قرار دارد اما نکته‌ای که باید در این جا به آن توجه کرد این است که در حالت تعداد کمتری از داده‌ها ویولن و سنتور در دسته‌ی پیانو قرار گرفته بودند که یعنی روش قبلی K-Means تمایز بهتری میان این دوساز و پیانو پیدا کرده بوده است البته این تفاوت خیلی چشم‌گیر نیست و می‌توان گفت هر دو روش تقریباً عملکرد یکسانی داشته‌اند.

k=۳، ۴ . ۲

```
Agglomerative3
[[ 12   2  169]
 [  9 204   7]
 [176  14  10]
 [179   6   9]
 [163   6   6]
 [ 98  25  91]]
```

```
Agglomerative4
[[169   2   10   2]
 [  7 204   9   0]
 [ 10  14  34 142]
 [  9   6 175   4]
 [  6   6 142  21]
 [ 91  25  65  33]]]
```



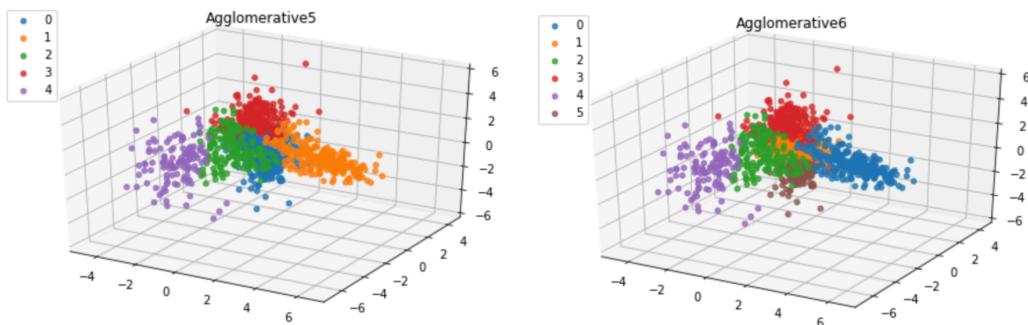
دسته بندی داده ها در ۳ خوش موجب شده ساز پیانو به دلیل پوشش گستره‌ی فرکانسی بالا و سازنی به علت تنها ساز بادی موجود میان سایر ساز ها تقریباً به خوبی تفکیک شوند. همچنین سه ساز تار، سه تار و سنتور که جزء سازهای ذهنی-مضرایی هستند در یک دسته قرار گرفته اند. داده های ساز ویولن نیز اکثرا در خوش های مربوط به نی و ساز های ذهنی-مضرایی قرار گرفته است. شباهت ویولن به نی می تواند به علت وجود صوت های حزن انگیز و شباهت به ساز های ذهنی-مضرایی به دلیل وجود قطعات سنتی باشد. مشاهده می شود که نتایج به دست آمده از این روش نیز شباهت زیادی به روش K-Means دارد و خوش بندی ساز ها در قسمت قبل نیز به همین صورت بود.

در حالت ۴ تایی داده های نی، پیانو و سنتور تقریباً به درستی دسته بندی شده اند و داده های مربوط به سازهای یکسان در یک خوش قرار گرفته اند. اکثر داده های مربوط به تار و سه تار در یک کلاستر و اکثر داده های مربوط به ویولن در خوش مربوط به نی و پس از آن در دسته مربوط به تار و سه تار وجود دارد. می توان گفت برای حالت ۴ تایی خوش بندی به دست آمده مناسب است چرا که میدانیم صوت مربوط به تار و سه تار تا حد زیادی شبیه هم است. ساز های نی، پیانو و سنتور نیز که از آنها فاصله دارند در دسته های مجزا قرار گرفته اند. می توان گفت در اینجا احتمالاً روش agglomerative از K-Means به دسته بندی بهتری رسیده است چرا که در حالت K-Means داده های تار میان سنتور و سه تار تقریباً به طور یکسان پخش شده بود اما در اینجا به طور مشخص به سه تار نزدیک تر است که از لحاظ شهودی نیز می دانیم که صوت تار به سه تار به نسبت سنتور نزدیک تر است.

k=5, 6

```
Agglomerative5
[[ 10  2  56   2 113]
 [ 9 204  7  0  0]
 [ 34 14 10 142  0]
 [175  6  6  4  3]
 [142  6  6 21  0]
 [ 65 25 87 33  4]]
```

```
Agglomerative6
[[ 2  7  56   2 113  3]
 [204  5  7  0  0  4]
 [ 14 32 10 142  0  2]
 [ 6 56  6  4  3 119]
 [ 6 133 6 21  0  9]
 [ 25 63 87 33  4  2]]
```



در دسته بندی ۵ تایی مشاهده می شود داده های نی، پیانو و سنتور تقریبا در کلاستر های مجزا از هم قرار گرفته اند. همچنین داده های تار و سه تار که پیشتر نیز به شباهت آنها اشاره شد در یک دسته و اکثر داده های ساز ویولن نیز به یک دسته مجزا قرار گرفته اند که نشان می دهد مدل تقریبا الگوهای موجود را به درستی یافته است. در سطر ۵ که مربوط به داده های ساز ویولن است مشاهده می شود که پس از ۸۷ که بیشترین تجمع داده درون کلاستر را نشان می دهد عدد ۶۵ قرار دارد که مربوط به کلاستر تار و سه تار است و از آن جایی که این اختلاف خیلی زیاد نیست می توان گفت مدل کمی در تشخیص ساز ویولن دچار مشکل است و آن را در موقعی مانند تار و سه تار در نظر می گیرند. در مدل K-Means که دسته بندی ساز ویولن به خوبی انجام شده است اما مشکل در تشخیص ساز تار وجود داشت که تعداد زیادی از داده های آن میان دو خوشه مربوط به سنتور و سه تار پخش شده بود.

هنگامی که خوشه بندی با ۶ دسته انجام می شود می بینیم که تقریبا در هر کلاستر به طور خاصی تنها یک ساز تعداد نمونه های بیشتری دارد و سازها تقریبا از هم جدا شده اند. اما همچنان تفکیک ساز ویولن کمی سخت بوده و تعدادی از داده های آن میان چند کلاستر پخش شده اند. با مقایسه ای این نتایج با حالت K-Means با ۶ خوشه می بینیم که دسته بندی در این حالت مجددا بهتر بوده چرا که در اینجا تنها ساز ویولن است که مقداری در کلاستر های دیگر نیز پخش شده است اما در حالت قبلی این مشکل برای دوساز ویولن و تار وجود داشت.

مراجع

- [۱] آنچه که باید درباره ساز پیانو بدانیم. <https://saednews.com/fa/post/anche-ke-baid-darbare-saz-piano-.bdanim>
- [۲] ملکی، علی نجفی. نوازنده‌گی نی. <http://www.harmonytalk.com/id/1030/>
- [3] Boser, Bernhard E, Guyon, Isabelle M, and Vapnik, Vladimir N. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152, 1992.
- [4] McCulloch, Warren S and Pitts, Walter. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.