

دانشگاه تهران
پردیس دانشکده فنی دانشکده مهندسی برق و کامپیوتر

عنوان:

گزارش پروژه پایانی درس یادگیری ماشین

نگارش:

محمد اسدزاده

حامد خادمی خالدی

آفرین زمانیان دولتی

سید امین موسوی چاشمی

۱۴۰۱ زمستان

چکیده

گزارش پیش‌رو حاصل پیاده‌سازی و تحلیل نتایج پژوهش پایانی درس یادگیری ماشین است که هدف آن طبقه‌بندی یک مجموعه داده از موسیقی سنتی ایرانی (اجرا شده در ترکیبی از ۷ دستگاه و ۶ ساز) بر اساس دستگاه موسیقی آن می‌باشد.

بخش اول این گزارش شامل پیش‌پردازش داده‌ها و استخراج ویژگی بوده و سپس به طبقه‌بندی داده‌ها به کمک روش‌های مختلف پرداخته شده است. در مرحله بعد، سعی در بهبود کارایی طبقه‌بندها به کمک ensemble روش‌های کاهش بعد بوده و نتایج حاصل از آن ارائه شده‌اند. در ادامه به روش‌های یادگیری پرداخته شده و از شبکه‌های عصبی ترتیبی، که امروزه میان بهترین روش‌های دسته بندی موسیقی هستند، برای طبقه‌بندی داده‌ها استفاده می‌شود. در نهایت، عملکرد روش‌های مختلف خوشه‌بندی بر روی داده‌ها آزموده و تحلیل شده است.

فهرست مطالب

۹	۱	مقدمه
۹	۱-۱	مفاهیم اصلی
۱۲	۲-۱	ساختار دستگاه
۱۲	۱-۲-۱	ارتباط میان دستگاهها
۱۳	۳-۱	تشخیص دستگاه
۱۳	۱-۳-۱	چالش‌ها و مشکلات
۱۴	۲-۳-۱	کارهای پیشین
۱۵	۲	پیش‌پردازش
۱۶	۱-۲	پاکسازی داده‌ها
۱۶	۱-۱-۲	Resample کردن داده‌ها
۱۶	۲-۱-۲	نرمال‌سازی داده‌ها
۱۶	۲-۲	استخراج ویژگی
۱۷	۱-۲-۲	MFCC
۱۷	۲-۲-۲	Chroma
۱۸	۳-۲	تصورسازی ویژگی‌ها
۱۸	۴-۲	مقایسه ویژگی‌های MFCC و Chroma

۲۲	۳ طبقه‌بندی
۲۲	۱-۳ آماده‌سازی
۲۲	۱-۱-۳ توزیع کلاس‌ها
۲۲	۲-۱-۳ تعریف بردار ویژگی
۲۳	۳-۱-۳ نحوه ارزیابی طبقه‌بندها
۲۴	۴-۱-۳ ارزیابی ویژگی‌ها
۲۴	۵-۱-۳ کاهش ابعاد
۲۵	۲-۳ انتخاب مدل
۲۵	۱-۲-۳ SVM
۲۷	۲-۲-۳ KNN
۲۷	۳-۲-۳ MLP
۲۸	۴-۲-۳ XGBoost
۲۹	۵-۲-۳ ارزیابی طبقه‌بندها
۲۹	۳-۳ Ensemble Learning
۳۰	۴-۳ LSTM
۳۵	۴ خوشه‌بندی
۳۵	۱-۴ آزمایش‌های اولیه
۳۶	۱-۱-۴ معیارهای سنجش عملکرد
۳۷	۲-۱-۴ توابع و عملیات
۳۷	۳-۱-۴ فرکانس، Chroma ۱۲
۳۸	۴-۱-۴ فرکانس Chroma ۲۴
۳۹	۵-۱-۴ فرکانس Chroma ۳۶

۴۰	MFCC فرکانس	۶-۱-۴
۴۱	مقایسه عملکرد	۷-۱-۴
۴۳	آزمایش‌های بیشتر و تحلیل نتایج	۲-۴
۴۶	آزمایش بر روی پارامترهای Agglomerative	۱-۲-۴
۴۸	Tأثیر کاهش بعد با PCA	۳-۴
۵۰	نتیجه‌گیری	۴-۴

فهرست تصاویر

۱۱	۱-۱ نت‌های شاخص دستگاه‌های موسیقی ایرانی
۱۹	۱-۲ نمونه ویژگی‌های Chroma و MFCC برای دستگاه‌های مختلف
۲۰	۲-۲ نمودار ROC
۲۳	۱-۳ توزیع داده‌ها بر حسب دستگاه
۲۵	۲-۳ دقت طبقه‌بند KNN برای انواع ویژگی‌ها
۲۶	۳-۳ دقت طبقه‌بند KNN با روش‌های کاهش بعد
۲۶	۴-۳ دقت طبقه‌بند SVM
۲۷	۵-۳ دقت طبقه‌بند KNN
۲۸	۶-۳ دقت طبقه‌بند MLP
۲۸	۷-۳ دقت طبقه‌بند XGBoost
۲۹	۸-۳ دقت طبقه‌بند‌های مختلف
۳۰	۹-۳ دقت روش‌های Ensemble Learning
۳۱	۱۰-۳ معماری شبکه LSTM
۳۲	۱۱-۳ نتایج LSTM با تعداد مختلف ویژگی chroma
۳۲	۱۲-۳ نمودار آموزش شبکه عصبی
۳۳	۱۳-۳ دقت به تفکیک هر کلاس در داده train+validation

۳۴	۱۴-۳ دقت به تفکیک هر کلاس در داده test
۳۴	۱۵-۳ عملکرد k-fold بر روی LSTM
۴۲	۱-۴ معیارهای عملکرد خوشبندی بر روی دادهها
۴۳	۲-۴ دقت آموزش و تست خوشبندی بر روی دادهها
۴۴	۳-۴ ماتریس پریشانی برای خوشبندی با ward linkage
۴۶	۴-۴ ماتریس آشتفتگی برای حالت‌های مختلف linkage
۴۸	۵-۴ نمودار Scree برای PCA
۴۹	۶-۴ استفاده از PCA در رسم دادهها (n=2)

فهرست جداول

۱۲	۱-۱ گوشه اوج دستگاه‌های موسیقی سنتی ایرانی
۲۱	۱-۲ نتایج Cross Validation بر روی ویژگی‌های مختلف
۳۱	۱-۳ نتایج LSTM با تعداد مختلف ویژگی chroma
۳۸	۱-۴ خروجی بر روی داده‌های 12 Chroma، تمام ابعاد
۳۸	۲-۴ خروجی بر روی داده‌های 12 Chroma، میانگین ویژگی‌ها (۱۲ بعد)
۳۹	۳-۴ خروجی بر روی داده‌های 24 Chroma، تمام ابعاد
۳۹	۴-۴ خروجی بر روی داده‌های 24 Chroma، میانگین ویژگی‌ها
۳۹	۵-۴ خروجی بر روی داده‌های 36 Chroma، تمام ابعاد
۴۰	۶-۴ خروجی بر روی داده‌های 36 Chroma، میانگین ویژگی‌ها
۴۰	۷-۴ خروجی بر روی داده‌های 40 MFCC، تمام ابعاد
۴۱	۸-۴ خروجی بر روی داده‌های 40 MFCC، میانگین ویژگی‌ها
۴۵	۹-۴ نتایج و دقت خوشه‌بندی بر روی داده‌های تست با ward linkage
۴۹	۱۰-۴ نتایج و دقت خوشه‌بندی بر روی داده‌های تست با PCA
۵۰	۱۱-۴ نتایج و دقت خوشه‌بندی بر روی داده‌های تست با PCA(n_components=2)

فصل ۱

مقدمه

موسیقی ایرانی تأثیر زیادی بر موسیقی اطراف خود، بلکه حتی موسیقی اروپایی و شرق دور گذاشته است. از این رو، تحقیق بر روی آن می‌تواند به موسیقی مناطق و فرهنگ‌های دیگر نیز تعیین داده شود و از آن استفاده شود. ساختار تشکیل دهنده موسیقی ایرانی، برخلاف موسیقی کلاسیک غرب اروپا که بر اساس گام و فرم است، دستگاه نامیده می‌شود. تشخیص این دستگاه‌ها بر اساس فرم و الگوهای تکرار شونده می‌تواند به عنوان یک مسئله یادگیری ماشین و هوش مصنوعی تعریف شود. در ادامه به معرفی دستگاه‌ها و برخی از نکات و چالش‌های حل این مسئله پرداخته شده است^۱.

۱-۱ مفاهیم اصلی

تعاریفی که امروزه بیشتر شناخته شده هستند از موسیقی کلاسیک غرب آمده‌اند و سیستم‌های موجود نیز غالباً بر این اساس «گام»^۲ و «مُد»^۳ موسیقی را تشخیص می‌دهند. برخی از این مفاهیم در موسیقی ایرانی نیز وجود دارند و برخی دیگر به نحوه متفاوتی تعریف می‌شوند.

«فاصله»^۴ در موسیقی که میان دو نت در نظر گرفته می‌شود، به معنی تفاوت فرکانسی آن‌هاست. برای مثال، بسامد نت «لا» میانی (در پیانوهای امروزی) که یکی از مبنای‌های پرکاربرد موسیقی است

^۱ محتوای این فصل برگرفته از گزارش اولیه پروژه می‌باشد.

^۲ scale (French: gamme)

^۳ mode

^۴ interval

هرتز می‌باشد و نت «سی بمل» که بلا فاصله بعد از آن می‌آید، بسامدی تقریباً معادل ۱۶۴.۴۶۶ دارد. این فاصله یک «نیم پرده»^۵ نامیده می‌شود. از تفاوت‌های اصلی موسیقی ایرانی با غربی که باعث می‌شود بسیاری از مفاهیم استفاده شده در موسیقی کلاسیک غربی قابل تعمیم به آن نباشند، سازهای آن هستند که در آن نتها به جای نیم پرده و یک پرده معمول در سازهای غربی، یک‌چهارم پرده با هم فاصله دارند، این ساختار را ریزپرده^۶ می‌نامند.

در موسیقی سنتی ایرانی، از آنجا که موسیقیدانان هنوز با مفهوم بسامد آشنایی نداشتند، فاصله را بر اساس سیم‌های ساز عود محاسبه می‌کردند، چرا که سازهای ذهنی (بعد از حنجره انسان) کامل‌ترین نغمه‌ها را ایجاد می‌کردند. عبارتی مانند مطبوعیت (ملایمت)^۷ و نامطبوعیت (خشونت)^۸ برای توصیف ترکیب دو نت با هم بر اساس فاصله آنها به کار می‌روند. در موسیقی غربی دوستگی کاملی میان این دو وجود دارد، اما در موسیقی شرقی و ایرانی، درجه ملایمت در یک طیف تعریف می‌شود.

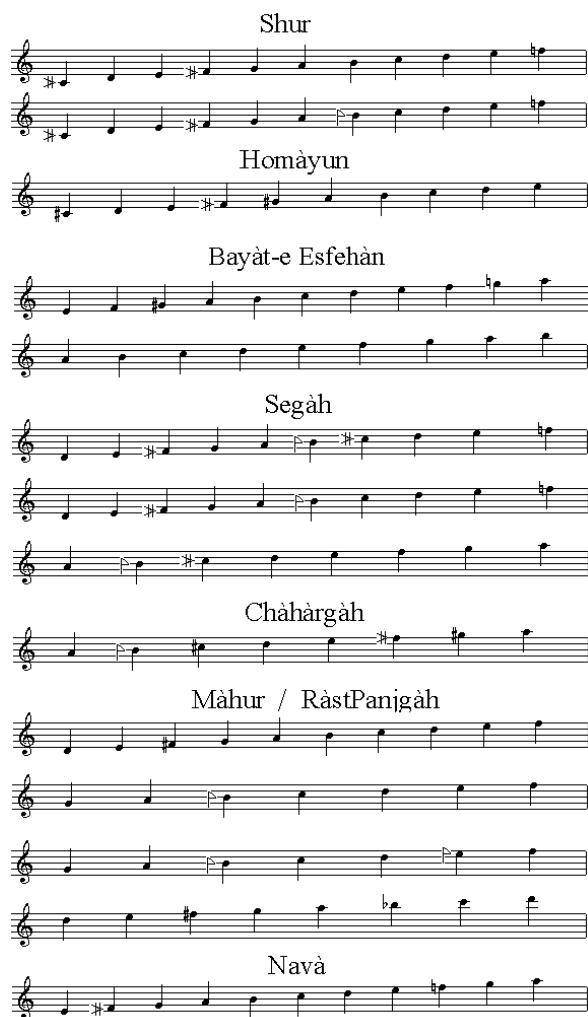
«اکتاو»^۹ فاصله میان دو نت همنام متواالی است. از لحاظ بسامد، بسامد یک نت دو برابر نتی است که یک اکتاو پایین‌تر از آن قرار دارد؛ برای مثال، بسامد نت لا میانی ۴۴۰ هرتز و نت لا اکتاو بعدی ۸۸۰ هرتز است. بر همین اساس، گام یک دنباله از نتها تعریف می‌شود که معمولاً در یک اکتاو قرار دارند. در موسیقی غربی، گام معمولاً بر اساس فواصل یک یا نیم پرده تعریف می‌شود، اما در موسیقی ایرانی این فواصل می‌توانند به ربع‌پرده نیز کاهش پیدا کنند.

مُد موسیقی یک مفهوم انتزاعی‌تر است که در رایج‌ترین تعریف، شکل خاصی از گام با مجموعه‌ای از حالت‌های مشخص ملودیک است. در واقع، مد مفهومی میان گام و ملودی است.

علاوه بر این موارد، مفاهیمی خاص موسیقی ایرانی نیز وجود دارند که به کمک آنها یک قطعه توصیف می‌شود. مهم‌ترین آنها «دستگاه» و «مقام» است.

دستگاه در موسیقی ایرانی، مجموعه‌ای از نغمه‌ها یا «گوشه»‌هاست که در گام و فواصل هم‌آهنگ هستند. در تقسیم‌بندی جدید، ۷ دستگاه موسیقی تعریف می‌شوند و در کنار آنها ۵ آواز نیز وجود دارد که در کنار آنها اجرا می‌شود. دستگاه‌ها را می‌توان بر اساس گام‌های به کار رفته در آن شناسایی کرد. شکل ۱-۱ نتهايی که برای کوک کردن سنتور در هر دستگاه استفاده می‌شوند را نشان می‌دهد.

semitone^۵
microtonal^۶
consonance^۷
dissonance^۸
octave^۹



شکل ۱-۱: نت‌های شاخص دستگاه‌های موسیقی ایرانی [۱]

در طبقه‌بندی‌های قدیمی‌تر موسیقی سنتی، به جای دستگاه از مفهوم «مقام» استفاده می‌شد که گسترده‌تر از دستگاه و مشترک با موسیقی سنتی عثمانی (ترکی) و عربی بود. پس از جدایی موسیقی ایرانی از عثمانی و تحولاتی که سعی در مدون کردن موسیقی سنتی ایرانی در دوران قاجاریه صورت گرفت، دستگاه به عنوان طبقه‌بند اصلی موسیقی ایرانی استفاده می‌شود.

در موسیقی ایرانی دوازده مقام اصلی بر شمرده شده‌اند و هر یک مجموعه‌ای از نغمه‌ها با فواصل معین و سازمان یافته هستند. اما از آنجا که مفهوم مقام بسیار قدیمی‌تر از مفاهیم دیگر ساختاری است، تعریف هر مقام نسبت به ردیف، گوشه، و دستگاه بیهمتر است. برای مثال، بعضی مقام‌ها معادل مفهوم مُد موسیقی و برخی ساختارمندتر هستند و فقط کمی پایین‌تر از مفهوم دستگاه قرار می‌گیرند. هر یک

جدول ۱-۱: گوشه اوج دستگاه‌های موسیقی سنتی ایرانی

دستگاه	شور	سه‌گاه	چهارگاه	راست‌پنجم‌گاه	همایون	نوا	ماهور
اوج	شهناز	مخالف سه‌گاه	مخالف چهارگاه	عشاق	بیداد	عشاق	شکسته

از مقام‌ها بار احساسی مخصوص به خود را دارد و حتی گاهی دوازده مقام را برگرفته از ماههای سال و دایره البروج و حتی ساعات شب‌انه‌روز دانسته‌اند. این موارد نشان می‌دهند که بیشتر از یک ساختار مشخص موسیقایی، مقام به درک فردی و هنری از اثر وابسته است و تشخیص آن به صورت الگوریتمیک سخت‌تر است.

۱-۲ ساختار دستگاه

«ردیف» به معنی مجموعه‌ای از گوشه‌های است که با ترتیب خاصی کنار هم قرار داده می‌شوند و در کنار هم دستگاه را تشکیل می‌دهند و پس از تحول مقام به دستگاه به عنوان اجزای شاخص دستگاه‌ها شناخته می‌شود. قطعات معمولاً با گوشه «درآمد» آغاز می‌شوند و در گوشه‌های بعدی نتها به تدریج اوج می‌گیرند یا به عبارتی زیرتر می‌شوند که بر همین اساس گوشه‌های «اوج» نام دارند که مهم‌ترین گوشه دستگاه تلقی می‌شوند. در ادامه گوشه‌های «فروود» می‌آیند که دوباره در آن نتها بهم می‌شوند. معمولاً قطعه در همان گوشه‌ای که آغاز شده پایان می‌یابد.

عباراتی نیز وجود دارند که به طور ریزتر به اجزای یک ردیف اشاره می‌کنند. در واقع هر ردیف معمولاً شکل دایروی پنجگانه‌ای شامل درآمد، چهارمضراب، آواز، تصنیف و رینگ دارد و دوباره به درآمد ختم می‌شود. گوشه درآمد، معرفی دستگاه و مدد آن است و در واقع فضای قطعه را معرفی می‌کند.

هر دستگاه، گوشه اوج شاخصی را دارد که نامی نیز به آن داده می‌شود. برای مثال، اوج دستگاه شور، شهناز نامیده می‌شود. در جدول ۱-۱، هفت دستگاه اصلی به همراه گوشه اوج آن‌ها آورده شده‌اند.

۱-۲-۱ ارتباط میان دستگاه‌ها

همانطور که در جدول ۱-۱ نیز مشاهده می‌شود، دستگاه‌ها نیز به طور کامل از هم مستقل نیستند و ممکن است گوشه‌های مشابهی در آن‌ها به کار رفته باشد. در واقع، برخی موسیقیدانان پنج دستگاه اصلی

را برای موسیقی ایرانی بر می‌شمارند و حتی محققانی از جمله داریوش طلایی معتقدند با شکستن هر دستگاه به دو «دانگ» می‌توان تمام ساختارها را با ترکیب چهار دانگ اصلی بازسازی کرد.

بر همین اساس، دستگاه شور، هرچند که قدمت کمتری از سایر دستگاهها دارد، به عنوان «مادر دستگاهها» و مهم‌ترین دستگاه موسیقی ایرانی شناخته می‌شود، چرا که بیشتر آوازها و بسیاری از گوشدها برگرفته از این دستگاه هستند و مشترکاتی با آن دارند که این امر می‌تواند تشخیص دستگاه را برای شنوندگان و سیستم‌های تشخیصی سخت کند.

از طرفی، در برخی قطعات سنتی بیش از یک دستگاه به کار می‌رود؛ قطعه با معرفی دستگاه اولیه در پیش‌درآمد شروع می‌شود و سپس به کمک گوشدهای خاصی به دستگاه دیگری می‌رود و دوباره در فرود به دستگاه اصلی باز می‌گردد. در این موارد، دستگاه قطعه همان دستگاه اولیه در نظر گرفته می‌شود.

۱-۳-۱ تشخیص دستگاه

۱-۳-۱-۱ چالش‌ها و مشکلات

دو نکته در بحث تشخیص دستگاه حائز اهمیت است، اول آنکه همانطور که در قسمت ۱-۲-۱ اشاره شد، برخی دستگاهها با هم شباهت بیشتری دارند و برخی قطعات نیز از بیش از یک دستگاه تشکیل شده‌اند. در وهله بعدی، برخی از قطعاتی که باید تشخیص داده شوند قطعات آوازی هستند و آواز تا حدود زیادی مستقل از خود دستگاه است و می‌تواند میان دستگاهها مشترک باشد.

از دیگر چالش‌های این مسئله، تعداد کم داده‌های آموزشی و در بسیاری از موارد کیفیت پایین ضبط قطعات است که از اجراهای زنده، خانگی و یا بسیار قدیمی گرفته شده‌اند. به همین دلیل، در تحقیقاتی که در این زمینه صورت گرفته، تقریباً در هیچ دو موردی دیتاست یکسانی استفاده نشده که این امر می‌تواند مقایسه عملکرد مدل‌ها را عملأ ناممکن کند.

همچنین، طول قطعات نیز می‌توانند از خیلی کوتاه (کمتر از یک دقیقه) تا قطعات بداهه بلند (بیش از ۲۰ دقیقه) متغیر باشند. قطعات کوتاه معمولاً بخشی از اجراهای کامل هستند که فقط یک یا چند گوشه از اجرا را پوشش می‌دهند. هر قطعه باید به سگمنت‌های کوچک‌تر تقسیم و سپس طبقه‌بندی شود. قابل ذکر است که داده‌ها باید به عنوان سری زمانی دیده شوند و ترتیب سگمنت‌های قطعه باید حفظ شود.

۲-۳-۱ کارهای پیشین

کارهای متعددی برای کلاس‌بندی موسیقی ایرانی بر اساس دستگاه انجام گرفته است که از میان آن‌ها می‌توان به [۲] اشاره کرد که با تقسیم هر اکتاو به ۶۰ فاصله، دستگاه‌ها و در ادامه مقام‌ها بر اساس یک آرایه از مدل‌های اسکلت ملودیک قطعه شناسایی می‌کند. عبدالی در ۲۰۱۱ از مجموعه‌های فازی به عنوان الگویی برای کلاس‌بندی دستگاه‌ها استفاده کرده است ([۳]). اما در مجموع کارهای کمتری به کمک یادگیری عمیق و یادگیری ماشین در این حوزه صورت گرفته که در جدیدترین آن‌ها [۴] در سال ۲۰۲۲، از یک شبکه عصبی تکرارشونده دو طرفه استفاده شده است.

فصل ۲

پیش‌پردازش

در این قسمت به پاک سازی داده های موسیقی و استخراج ویژگی ها آن ها می پردازیم. در این مرحله داده ها برای مرحله بعدی این پروژه یعنی دسته‌بندی و خوشه‌بندی آماده می شوند.

در این مرحله ما داده های موسیقی را را با طول ۲۰ ثانیه قطعه قطعه می کنیم و اگر موسیقی بر ۲۰ ثانیه بخش پذیر نبود (که غالباً این گونه است) باقی‌مانده موسیقی بر ۲۰ را از ابتدا و انتهای موسیقی کم می کنیم (چون ابتدا و انتهای موسیقی دارای اطلاعات مفیدی نیستند و عموماً با سکوت شروع و تمام می شوند). مثلاً اگر یک قطعه موسیقی با طول ۲۵ ثانیه داشته باشیم، برای برداشتن قطعه با طول ۲۰ ثانیه، ۲۰ ثانیه وسط انتخاب می شود و ۵.۰ ثانیه ابتدا و انتها از موسیقی نادیده گرفته می شود.

مجموعه داده ای که در اختیار گذاشته شده با اینکه در صورت پروژه بیان شده بود حداقل اندازه موسیقی ها باید ۲۰ ثانیه باشد. دو موسیقی وجود داشت که اندازه آن ها کمتر از ۲۰ ثانیه بود به همین دلیل آن ها نادیده گرفته شدند. برای خواندن موسیقی و پاک سازی و استخراج ویژگی ها از کتابخانه librosa استفاده شده است. چون اکثر داده ها از فرمت ۳mp است در ابتدا به فرمت wav تبدیل می کنیم به این دلیل که این فرمت بسیار lossless و نزدیک به نمایش دیجیتال صدا اصلی است ولی فرمت ۳mp یک فرمت فشرده برای ذخیره محسوب شود که آهنگ ها را با کیفیت کمتری توزیع کرد و هم چنین در کتابخانه librosa در خواندن موسیقی با فرمت wav بسیار سریع است.

۱-۲ پاکسازی داده‌ها

برای پاک سازی داده‌ها از دو روش resample کردن دادن داده‌ها و نرمال سازی داده‌ها استفاده کردیم که در ادامه شرح داده می‌شوند.

۱-۱-۲ کردن داده‌ها Resample

صدا‌ها به فرکانس ۰.۵۲۲ کیلوهرتز ریسمپل می‌شوند و این کار هنگامی که موسیقی را می‌خوانیم توسط کتابخانه librosa با متدهای load انجام می‌شود. همچنین در این کتابخانه از روش soxr_hq این کار انجام می‌شود.

مزیتی که این کار در پیش‌پردازش داده دارد اینست که داده‌های موسیقی چون داده‌های موسیقی دارای نرخ نمونه‌گیری متفاوتی هستند بنابراین یکسان‌سازی این نرخ باعث می‌شود مدل‌های یادگیری ماشین بتوانند تمایز بین داده‌ها موسیقی متأثر از تفاوت در نرخ در آن‌ها نباشد و به خود مقدار داده‌ها برگردد.

۲-۱-۲ نرمال‌سازی داده‌ها

با استفاده از کتابخانه librosa.util.normalize با متدهای normalize انجام می‌شود. در این متدها سیگنال‌های ساده به گونه‌ای نرمال می‌شوند که میانگین‌ها صفر شود و واریانس‌ها ۱ شود. این کار با این هدف انجام می‌شود که تمام سیگنال‌ها مانند آهنگ در یک scale قرار گیرند و در نتیجه ویژگی‌هایی که از سیگنال‌ها استخراج می‌شوند ویژگی‌های بهتری برای مدل‌های یادگیری ماشین باشند.

۲-۲ استخراج ویژگی

برای استخراج ویژگی از داده‌های صوتی به استخراج دو ویژگی MFCC و Chroma می‌پردازیم.

MFCC ۱-۲-۲

ویژگی (MFCC) Mel-frequency Cepstral Coefficient بیانگر پوشش طیفی^۱ یک سیگنال صوتی هست. این ویژگی حاصل لگاریتم گرفتن از طیف توانی^۲ در دامنه فرکانسی سیگنال صوتی و سپس تبدیل به مقیاس مل^۳ است. مقیاس مل بیانگر مقیاسی است که تقریباً در همان طیف فرکانسی شناوری انسان قرار می‌گیرد و فاصله میان باندهای فرکانس در آن برابر است. ویژگی‌های MFCC از ویژگی‌های متداول در کاربردهای یادگیری ماشین مانند تشخیص گفتار و طبقه‌بندی موسیقی است. از فواید این نوع ویژگی می‌توان به مقاوم بودن نسبت به نویز سیگنال صوتی و عدم حساسیت به طول آهنگ و سطح صدا اشاره کرد.

محاسبه ویژگی‌های MFCC از لحاظ زمانی بهینه است و ویژگی‌های تولید شده نسبت به نویز مقاوم هستند. این ویژگی‌ها به خوبی محتوای طیف صوتی را پوشش می‌دهند و قادر به بازنمایی ساختارهای تونالیته و ریتم است. در مقابل، اطلاعات محلی^۴ زمانی در ویژگی‌های استخراج شده بازنمایی نمی‌شوند و برخی از اطلاعات به دلیل ساختار کلی ویژگی‌ها در تبدیل از بین می‌روند. ویژگی‌های MFCC به گام‌های پیش‌پردازشی همچون پنجره‌بندی و قاب‌بندی نیز حساس است.

برای استخراج ویژگی MFCC از داده‌های موسیقی از متدهای librosa.feature.mfcc استفاده شده است.

Chroma ۲-۲-۲

ویژگی‌های Chroma از انواع ویژگی‌های قدرتمند برای داده‌های صوتی هستند که در استخراج آن هر داده صوتی به تعدادی دسته^۵ (مثلاً ۱۲ تا) نگاشت می‌شود. هر کدام از ویژگی‌های حاصل نشان‌دهنده یک نیم پرده (یا کرومای) متمایز از یک اکتاو موسیقی است. از آن جایی که نت‌هایی که دقیقاً یک اکتاو با هم فاصله دارند شبیه به یکدیگر هستند، دانستن توزیع کروما می‌تواند اطلاعات مفیدی درباره خود موسیقی به ما بدهد و حتی شباهت‌هایی را در صوت شنیداری بیابد که در طیف اصلی به آسانی قابل مشاهده نیستند. از این ویژگی نمایش کروما در طبقه‌بندی موسیقی نیز استفاده می‌شود.

spectral envelope^۱
power spectrum^۲
mel scale^۳
localized^۴
bin^۵

از دیگر برتری‌های ویژگی‌های Chroma، امکان حفظ و تقویت ویژگی‌های آهنگین و هارمونیک صوت در عین مقاومت در برابر تغییرات ناشی از ساز یا تن صدا است. این امر باعث می‌شود ویژگی‌های Chroma برای توصیف موسیقی و ویژگی‌های شنیداری آن بهتر عمل کنند. البته در استخراج این ویژگی‌ها بخشی از ویژگی‌های صوتی نیز از بین می‌روند. تعداد دسته‌ها و فرکانس کوک نیز در کیفیت ویژگی‌های خروجی تأثیر زیادی دارند.

برای استخراج ویژگی Chroma از داده‌های موسیقی ازتابع librosa.feature.chroma_stft استفاده شده است.

۳-۲ مصورسازی ویژگی‌ها

در شکل ۱-۲ برای هر کدام از دستگاه‌های موسیقی موجود در داده‌ها (شور - سه‌گاه - ماهور - همايون - راست پنجگاه - نوا - چهارگاه) یک قطعه به عنوان نمونه آورده و ویژگی‌های chroma و mfcc برای هر یک رسم شده‌اند.

۴-۲ مقایسه ویژگی‌های Chroma و MFCC

برای مقایسه بین این دو ویژگی، تعداد ویژگی‌ها را برای هر دو ویژگی chroma و mfcc با ۲۰ در نظر گرفتیم و با استفاده از طبقه‌بند MLP (با تعداد لایه‌های مخفی ۷ و ۸) نمودارهای ROC را رسم کرده (شکل ۲-۲) و مقدار AUC را برای هر ویژگی گزارش دادیم. هم‌چنین با استفاده از روش cross-validation با مقدار $CV = 5$ به مقایسه تأثیر دو ویژگی بر معیار دقت می‌پردازیم و میانگین دقت و انحراف معیار آن را روی داده‌های تست (بخش تست داده در هر fold اعلام می‌کنیم).

برای رسم نمودار ROC از روش one against all برای هر کلاس استفاده شده است زیرا برای طبقه‌بندی دو کلاسه است اما مسئله در دست از نوع چندکلاسه می‌باشد.

با توجه به نمودارهای ROC عملکرد ویژگی chroma بهتر بوده است. این را می‌توان را از روی مقدار AUC هر کلاس که مقدار بزرگتری نسبت به ویژگی mfcc در نمودار ROC دارد فهمید. همانطور که در شکل معلوم است در نمودار ROC هر ویژگی از نیمساز ربع اول فاصله گرفته است (که

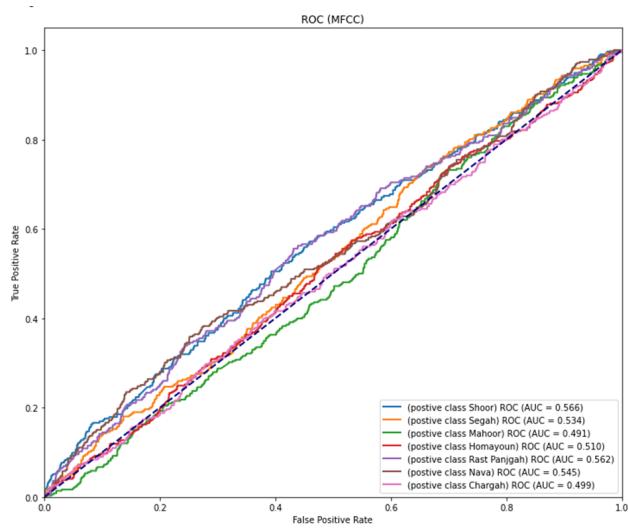


شکل ۲-۱: نمونه ویژگی‌های Chroma و MFCC برای دستگاه‌های مختلف

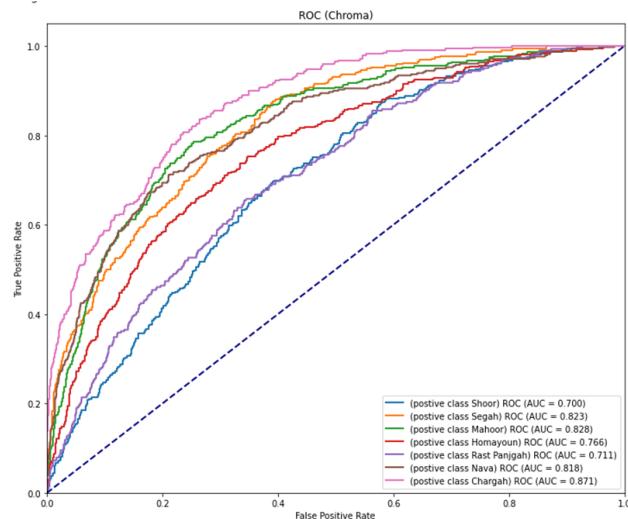
این معادل با بزرگ بودن مقدار AUC است).

با توجه به نمودارهای ROC عملکرد ویژگی chroma بهتر بوده است. این را می‌توان را از روی مقدار AUC هر کلاس که مقدار بزرگتری نسبت به ویژگی mfcc در نمودار ROC دارد فهمید. همانطور که در شکل معلوم است در نمودار ROC هر ویژگی از نیمساز ربع اول فاصله گرفته است (که این معادل با بزرگ بودن مقدار AUC است).

با توجه به نتایج بدست آمده از جدول ۱-۲، ویژگی chroma ویژگی بهتری برای طبقه‌بندی



(ا) نمودار ROC با ویژگی‌های MFCC



(ب) نمودار ROC با ویژگی‌های Chroma

شکل ۲-۲: نمودار ROC

است چون میانگین دقت روی داده‌های تست با این سری ویژگی‌ها بیشتر گزارش شده. این نتیجه‌گیری مطابق نتیجه قسمت قبل به کمک نمودار ROC است.

جدول ۱-۲: نتایج Cross Validation بر روی ویژگی‌های مختلف

	mean test score	std test score
MFCC	0.1509	0.0264
Chroma	0.2820	0.0400

فصل ۳

طبقه‌بندی

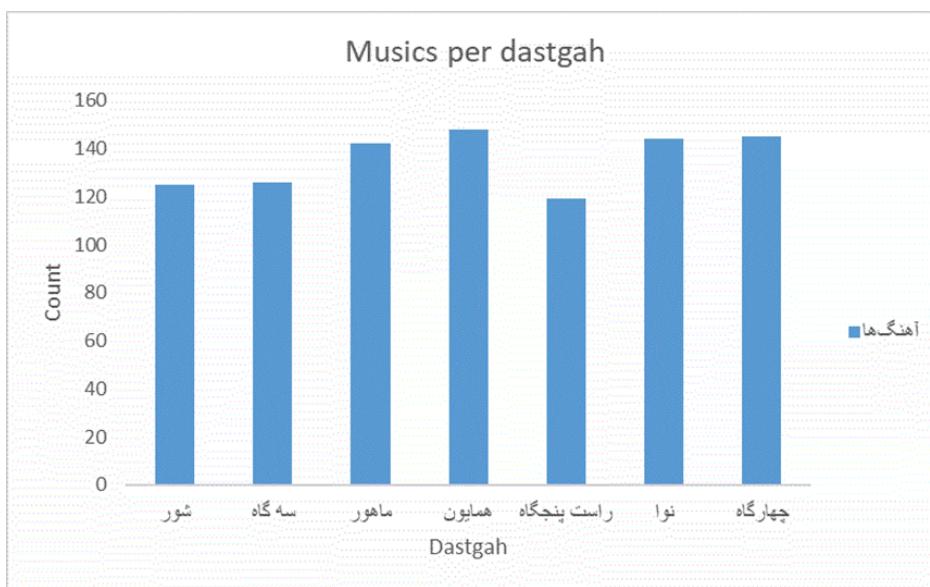
۱-۳ آماده‌سازی

۱-۱-۳ توزیع کلاس‌ها

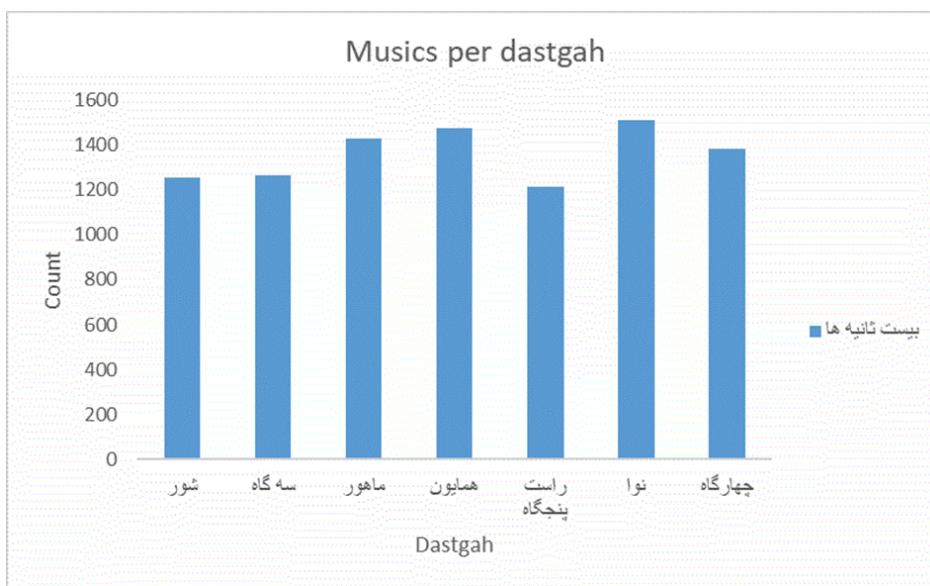
داده‌های موسیقی جمع آوری شده در ۷ کلاس (۷ دستگاه موسیقی) قرار دارند که توزیع آن‌ها به صورت شکل ۱-۳ است و نیز توزیع قطعات بیست ثانیه‌ای بر حسب دستگاه مشاهده می‌شود که تقریباً مشابه توزیع کل آهنگ‌هاست چرا که توزیع طول آهنگ‌ها در دستگاه‌های مختلف تقریباً یکسان است.

۲-۱-۳ تعریف بردار ویژگی

در بخش‌های قبلی ویژگی‌های MFCC و chroma معرفی شدند. می‌دانیم برای هر نمونه این ویژگی‌ها یک ماتریس دو بعدی است که یک بعد فریم‌های زمانی و بعد دیگر فرکانس می‌باشد. برای اینکه بتوانیم از روش‌های متداول یادگیری ماشین برای آموزش مدل استفاده کنیم، ابتدا باید ویژگی‌های هر نمونه را به صورت برداری نمایش دهیم. برای اینکار برای هر فرکانس، روی محور زمان میانگین و واریانس آن را محاسبه می‌کنیم. در نتیجه تعداد اجزای بردار ویژگی بدست آمده دو برابر تعداد فرکانس‌های اصلی می‌باشد.



(آ) توزیع کل داده‌های موسیقی



(ب) توزیع قطعات ۲۰ ثانیه‌ای از داده‌ها

شکل ۳-۱: توزیع داده‌ها بر حسب دستگاه

۳-۱-۳ نحوه ارزیابی طبقه‌بندها

در تمام مراحل این بخش وقتی می‌خواهیم مدلی را ارزیابی کنیم از معیار accuracy استفاده می‌کنیم. همچنین از روش k-fold cross validation برای بخش‌بندی‌ها داده‌های تست و آموزش استفاده می‌کنیم.

کنیم. از آنجایی که می‌خواهیم داده‌های تست ما ۲۵ درصد کل داده‌ها باشد، $k = 4$ قرار می‌دهیم. پس از چهار بار آموزش و تست مدل در نهایت میانگین دقت روی داده‌های آموزش، همچنین میانگین و انحراف معیار دقت روی داده‌های تست به عنوان شاخص‌های ارزیابی مدل بدست می‌آیند.

۴-۱-۳ ارزیابی ویژگی‌ها

می‌خواهیم از بین ویژگی‌های استخراج شده در مراحل قبل (MFCC و chroma) ویژگی را بیابیم که بهترین دقت طبقه‌بندی را نتیجه دهد. طبقه‌بندی را بر اساس تبدیل LDA پنج تا از ویژگی‌ها روی یک مدل KNN انجام می‌دهیم و نتایج آن را با یکدیگر مقایسه می‌کنیم. پنج ویژگی شامل موارد زیر است:

۱. MFCC-Cleaned با ۴۰ فرکانس

۲. MFCC با ۴۰ فرکانس

۳. MFCC-Cleaned با ۲۰ فرکانس

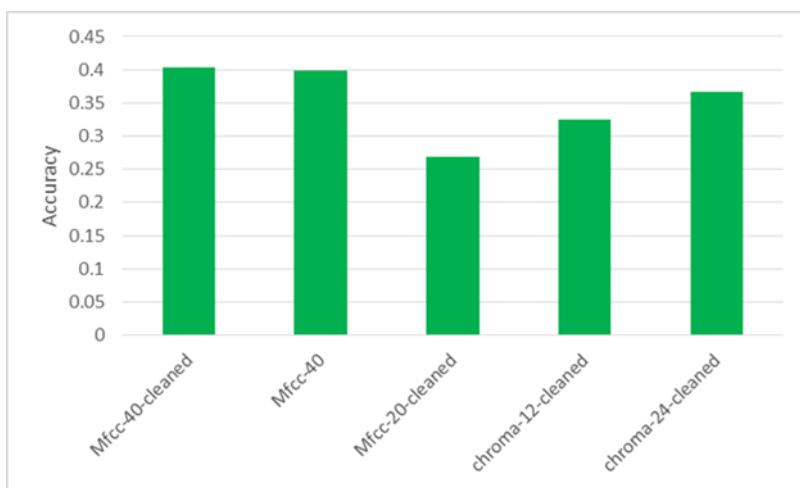
۴. Chroma-Cleaned با ۱۲ فرکانس

۵. Chroma-Cleaned با ۲۴ فرکانس

دقت بدست آمده برای هر یک از این پنج ویژگی در نمودار ۲-۳ قابل مشاهده است. با توجه به این نتایج طبقه‌بندی بر اساس ویژگی Mfcc-40-cleaned دارای بالاترین دقت می‌باشد. لذا از این پس این ویژگی را مبنای کار خود قرار می‌دهیم.

۵-۱-۳ کاهش ابعاد

روش‌های PCA، LDA و Forward Selection و Backward Selection از متداول ترین روش‌ها برای کاهش ابعاد ویژگی می‌باشند. LDA یک روش با ناظر است که با توجه به اینکه ۷ کلاس داریم ۶ بردار ویژگی خواهد داشت. تعداد ویژگی‌ها برای روش PCA را به گونه‌ای تعیین می‌کنیم که ۹۹ درصد واریانس داده‌های اولیه حفظ شود. در این صورت PCA سه ویژگی می‌دهد. Forward Selection و Backward Selection نتایج کاملاً مشابهی می‌دهند و ۲۰ ویژگی را انتخاب می‌کنند.



شکل ۳-۲: دقت طبقه‌بند KNN برای انواع ویژگی‌ها

دقت حاصل از هریک از این روش‌ها را بدست می‌آوریم و آن را با حالتی که کاهش ابعاد نداشته باشیم مقایسه می‌کنیم. در اینجا نیز از مدل KNN برای محاسبه دقت استفاده می‌کنیم.

همانطور که در نمودار ۳-۳ می‌بینیم کاهش بعد با LDA به طور قابل ملاحظه‌ای دقت مدل را افزایش داده. FC و BC نیز به بهبود دقت کمک کرده است، در حالی که روش PCA تاثیر مثبتی روی دقت نداشته است. در ادامه LDA را مبنای کار خود قرار می‌دهیم.

۲-۳ انتخاب مدل

در این پروژه از سه طبقه‌بند SVM ، KNN و MLP برای طبقه‌بندی دستگاه‌های موسیقی استفاده می‌کنیم. ابتدا با بررسی مقدار هایپر پارامترها برای هر طبقه‌بند بهترین مدل را برای هر یک انتخاب می‌کنیم. سپس به مقایسه این سه طبقه‌بند با یکدیگر خواهیم پرداخت.

SVM ۱-۲-۳

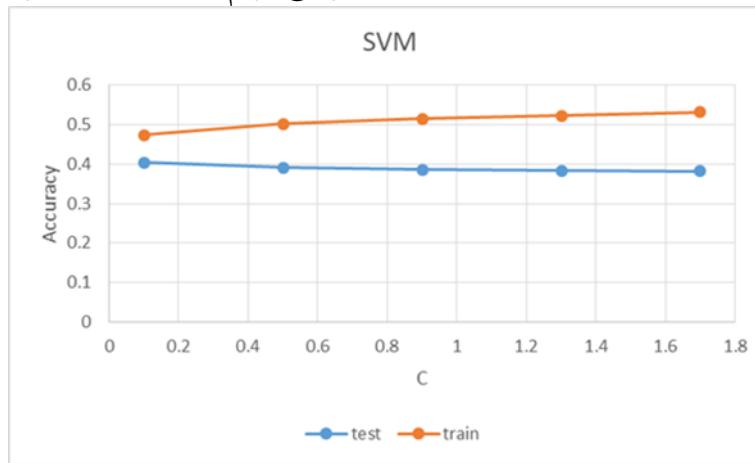
در این روش از یک hyperplane برای جداسازی کلاس‌ها استفاده می‌شود تا بیشترین مارجین را بین کلاس‌ها پیدا کند . همچنین با استفاده از کرنل‌ها می‌توان کلاس‌های که به صورت غیر خطی جدا شده اند نیز جداسازی نمود. و با توجه به این ویژگی‌ها از این روش برای پیاده‌سازی و کلاس‌بندی داده‌ها



شکل ۳-۳: دقت طبقه‌بند KNN با روش‌های کاهش بعد

استفاده شده است.

کرنل SVM را گوسی فرض کرده‌ایم. هایپر پارامتر ما در این طبقه‌بند C یعنی پارامتر مجازات خطای می‌باشد. پنج مقدار مختلف در بازه ۰.۱ تا ۱.۷ را در نظر می‌گیریم و دقت مدل را بررسی می‌کنیم.



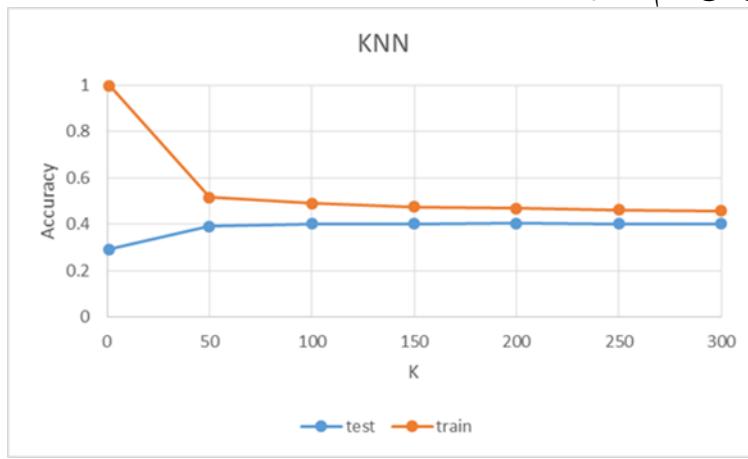
شکل ۴-۳: دقت طبقه‌بند SVM

با توجه به نمودار ۴-۳ با افزایش C دقت روی داده آموزش افزایش یافته اما generalization کاهش می‌یابد و خطای روی داده تست صعودی خواهد بود. همانطور که مشخص است مدل با $C = 0.1$ دارای بهترین دقت روی داده‌های تست می‌باشد و به عنوان مدل نهایی SVM انتخاب می‌شود.

KNN ۲-۲-۳

در این روش k تا از نزدیک ترین نقاط مربوط به یک نقطه انتخاب می‌شود و با توجه به آن نقاط ، کلاس نقطه مدنظر تعیین می‌شود. با توجه به این که این روش یک روش قدرتمند و در عین حال ساده است برای کلاس‌بندی داده‌ها از این روش استفاده شده است

برای هایپرپارامتر این طبقه‌بندی (K) هفت مقدار مختلف در بازه ۱ تا ۳۰۰ را در نظر می‌گیریم و دقت مدل را بررسی می‌کنیم (نمودار ۵-۳).



شکل ۳-۵: دقت طبقه‌بند KNN

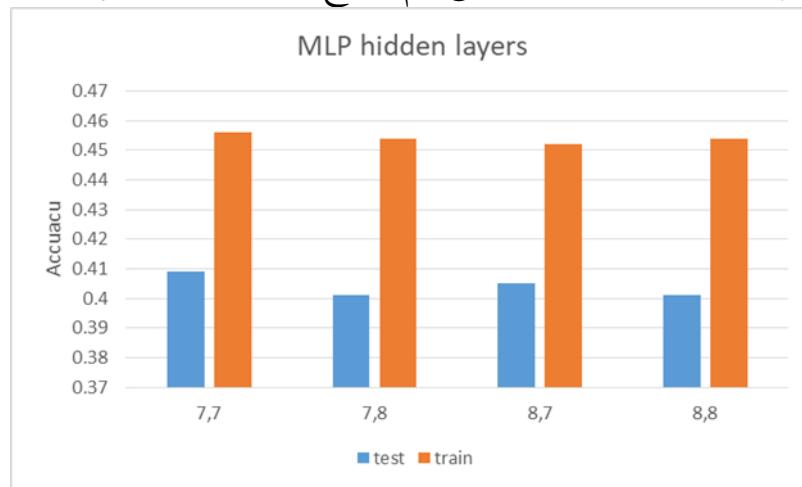
دقت در $K = 1$ روی داده‌های تست تقریباً صد درصد است. با توجه به نمودار بالا با افزایش K خطای روی داده آموزش کاهش یافته اما generalization نیز کاهش می‌یابد و خطای روی داده تست صعودی خواهد بود. مدل با $K = 200$ دارای بهترین دقت می‌باشد به عنوان مدل نهایی KNN انتخاب می‌شود

MLP ۳-۲-۳

روش Multi-Layer Perceptron یک روش بر پایه شبکه عصبی است. که ممکن است در این روش از یک یا چند لایه مخفی استفاده شود. و در نهایت با کمک روش‌هایی مانند backpropagation وزن‌های شبکه شروع به یادگیری می‌کنند. و با توجه به اینکه این روش برای داده‌های non-linear و پیچیده مناسب است از این روش نیز برای کلاس‌بندی استفاده شده است.

در اینجا مدل را پرسپترون با دو لایه مخفی فرض کرده ایم. هایپرپارامترهای ما تعداد نورون‌ها

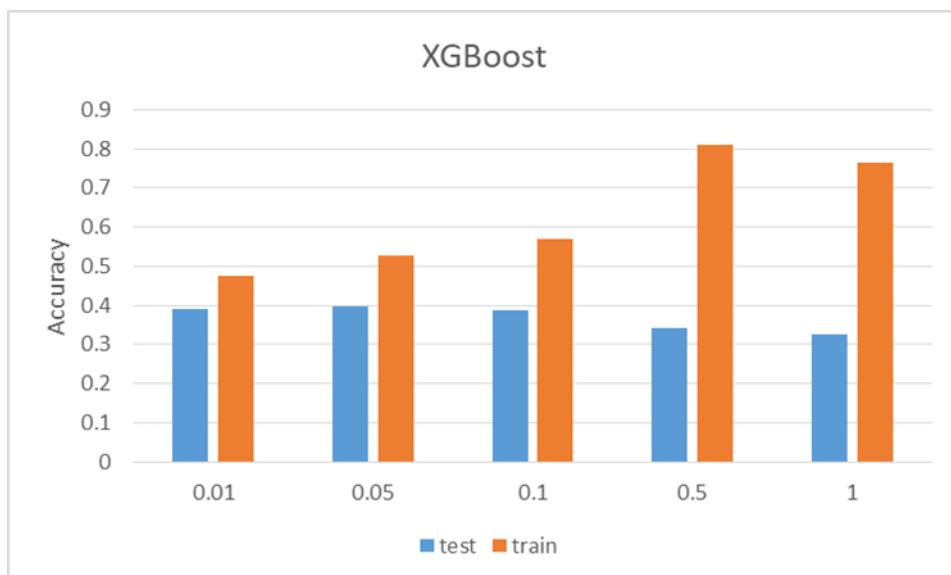
در هر لایه می‌باشد. تعداد نورون‌های لایه اول و دوم را مقادیر $(7, 7)$ ، $(8, 7)$ و $(8, 8)$ در نظر می‌گیریم و برای هریک دقت مدل را محاسبه می‌کنیم. نتایج بدست آمده در نمودار ۶-۳ آمده است.



شکل ۳-۶: دقت طبقه‌بند MLP

طبق مقادیر نمودار ۶-۳، ساختار لایه‌ها بصورت $(7, 7)$ بیشترین دقت و generalization را دارد و به عنوان مدل نهایی MLP انتخاب می‌شود.

XGBoost ۴-۲-۳



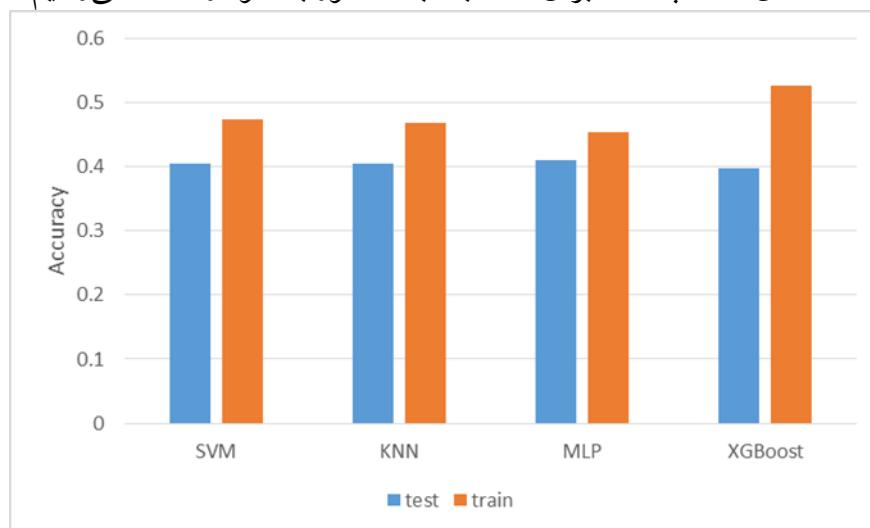
شکل ۷-۳: دقت طبقه‌بند XGBoost

در نهایت آخرین کلاسیفایر استفاده شده xgboost است. که در این روش از gradient boosting استفاده می‌شود. در حقیقت در این روش یک ensemble learning algorithms تشکیل شده است و با اضافه کردن درخت و استفاده از اطلاعات decision tree کاهش loss دارد. با توجه به اینکه این مدل نسبت به روش‌های قدیمی معمولاً دقیق‌تر و سرعتی بیشتری دارد از این روش استفاده شده است.

دقیق‌ترین مقادیر مختلف در نمودار ۷-۳ آورده شده است.

۵-۲-۳ ارزیابی طبقه‌بندها

با مقایسه دقیق‌ترین مدل‌های انتخاب شده برای سه طبقه‌بند مذکور به نمودار ۸-۳ می‌رسیم



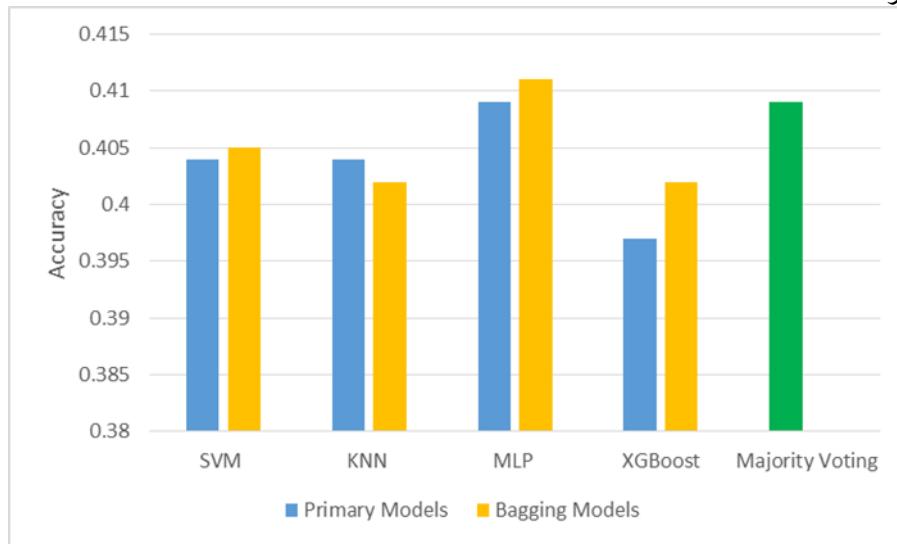
شکل ۸-۳: دقیق‌ترین طبقه‌بندها

دقیق‌ترین سه طبقه‌بند حدود ۰.۴۱ تا ۰.۴۲ است. اما روی داده‌های آموزش، SVM دقیق‌تری را ثبت کرده است. MLP بهترین و SVM کمترین generalization را دارند.

۳-۳ Ensemble Learning

در این بخش از روش ensemble learning به عنوان روش‌های Majority Voting و Bagging استفاده شده است. برای روش Bagging تعداد دیتاست‌ها ۱۰ در نظر گرفته شده است. دقیق‌ترین بدست آمده

مطابق نمودار ۹-۳ است.



شکل ۹-۳: دقت روش‌های Ensemble Learning

همانطور که انتظار می‌رفت، استفاده از Majority Voting باعث افزایش دقت شده است. اما استفاده از Bagging تنها در مدل MLP و XGB توانسته دقت را افزایش دهد و در دو طبقه بند دیگر حتی باعث کاهش دقت شده است.

LSTM ۴-۳

در این بخش از روش‌های شبکه عصبی برای کلاس‌بندی داده‌ها استفاده می‌شود. تمامی کدهای این بخش [در این لینک](#) آورده شده است.

برای پیاده‌سازی، از شبکه شکل ۱۰-۳ استفاده شده است که شامل دو LSTM متوالی و یک لایه fully-connected برای کلاس‌بندی می‌باشد. همچنین برای آموزش مدل از اپتیمایزر Adam استفاده شده است و در لایه‌های LSTM برای جلوگیری از بیش‌برازش^۱ از dropout، پنجاه درصد استفاده شده است. همچنین برای loss از categorical_crossentropy استفاده شده است. و در نهایت برای رسیدن به دقت بهتر از Early Stopping کمک گرفته شده است.

سپس برای پیاده سازی و آموزش از دو ویژگی مختلف MFCC و chroma استفاده شده است. با

^۱ overfitting

```

Build LSTM RNN model ...
Model: "sequential_1"

Layer (type)          Output Shape         Param #
lstm (LSTM)           (None, 862, 128)      90624
lstm_1 (LSTM)          (None, 32)            20608
dense (Dense)          (None, 7)             231

=====
Total params: 111,463
Trainable params: 111,463
Non-trainable params: 0
=====
```

شکل ۳-۱۰: معماری شبکه LSTM

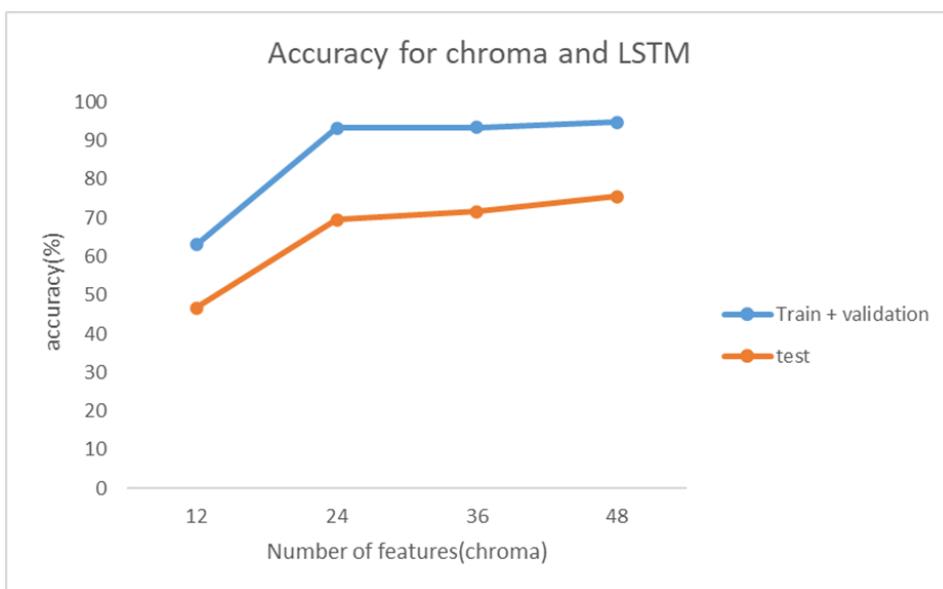
جدول ۳-۱: نتایج LSTM با تعداد مختلف ویژگی chroma

	feature count	12	24	36
accuracy	Train + validation	63.1	93.3	93.4
	test	46.7	69.5	71.6

استفاده از ویژگی‌های MFCC و تعداد ۲۰ ویژگی دقت (accuracy) برابر با ۵۰ درصد شد اما با توجه به اینکه دقت با ویژگی‌های chroma بهتر می‌شد در ادامه از این ویژگی‌ها برای آموزش استفاده شد.

برای آموزش تعداد ۹۵۳۸ آهنگ ۲۰ ثانیه‌ای وجود دارد که ۲۰ درصد آن مربوط به تست در نظر گرفته شده و همچنین ۲۰ درصد آن مربوط به داده validation. همچنین نسبت داده‌های train به test در هر کلاس به صورت مساوی در نظر گرفته شده است. در آخر دقت accuracy برای تعداد ویژگی‌های مختلف chroma در جدول ۳-۱ آورده شده است. که با توجه به نمودار ۳-۱۱ مشخصاً هر چه تعداد ویژگی‌ها بیشتر باشد دقت نیز بیشتر می‌شود.

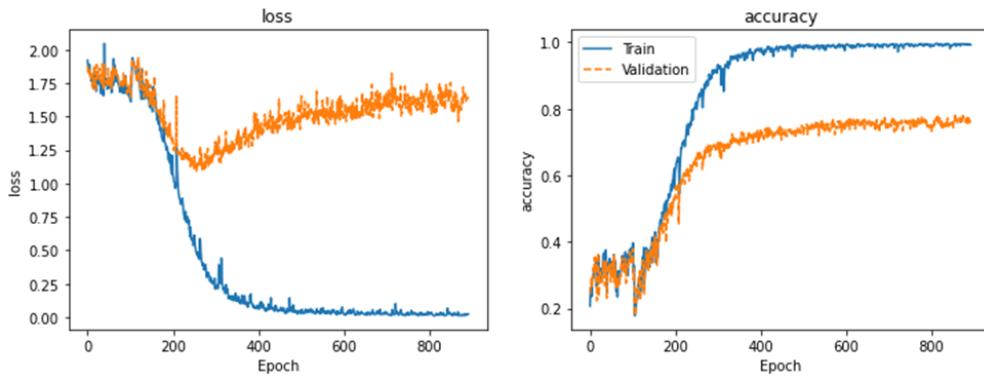
با تعداد ۴۸ ویژگی نمودار آموزش به صورت شکل ۳-۱۲ و در اپوک آخر نتایج به صورت زیر است:



شکل ۱۱-۳: نتایج LSTM با تعداد مختلف ویژگی chroma

train_loss: 0.0238 - train_accuracy: 0.9921

val_loss: 1.6476 - val_accuracy: 0.7661



شکل ۱۲-۳: نمودار آموزش شبکه عصبی

دقت بر حسب کلاس بر روی داده آموزش و ولیدیشن که شامل ۸۰ درصد داده هاست در شکل ۱۳-۳ و برای داده تست در شکل ۱۴-۳ آورده شده است.

همانطور که مشخص است شبکه عصبی عملکرد ضعیف تری بر روی دستگاه شور و راست پنجگاه دارد و در نهایت accuracy آن برابر با ۰.۷۵ درصد است.

در نهایت برای اعتبار بخشیدن به نتایج این بخش از روش k-fold استفاده می‌شود که این آموزش

	precision	recall	f1-score	support
0	0.932	0.949	0.940	1003
1	0.967	0.955	0.961	1011
2	0.948	0.962	0.955	1146
3	0.956	0.920	0.937	1181
4	0.926	0.933	0.930	972
5	0.951	0.955	0.953	1211
6	0.957	0.965	0.961	1106
accuracy			0.948	7630
macro avg	0.948	0.948	0.948	7630
weighted avg	0.949	0.948	0.948	7630

شکل ۱۳-۳: دقت به تفکیک هر کلاس در داده train+validation

روی 36 chroma صورت گرفته است که دقت مناسبی در مرحله قبل داشت (و با توجه به محدودیت استفاده شده است که نتایج دقت بر روی هر fold به صورت شکل ۱۵-۳). از $k = 4$ برای k-fold دقت به دست آمده را تایید می‌کند.

که به طور میانگین میزان accuracy برابر با 73.02 است که دقت به دست آمده را تایید می‌کند.

	precision	recall	f1-score	support
0	0.644	0.729	0.684	251
1	0.817	0.779	0.798	253
2	0.752	0.752	0.752	286
3	0.819	0.658	0.729	295
4	0.657	0.724	0.689	243
5	0.756	0.766	0.761	303
6	0.856	0.881	0.868	277
accuracy			0.755	1908
macro avg	0.757	0.755	0.754	1908
weighted avg	0.760	0.755	0.756	1908

شکل ۱۴-۳: دقت به تفکیک هر کلاس در داده test

```

Score for fold 1: loss of 1.64; accuracy of 73.29%
Score for fold 2: loss of 1.51; accuracy of 73.58%
Score for fold 3: loss of 1.56; accuracy of 72.56%
Score for fold 4: loss of 1.54; accuracy of 72.65%

```

شکل ۱۵-۳: عملکرد k-fold بر روی LSTM

فصل ۴

خوشبندی

۱-۴ آزمایش‌های اولیه

در این قسمت به خوشبندی داده‌ها با دو روش K-Means و Agglomerative پرداخته‌ایم. علاوه بر استفاده از پارامتر تعداد خوشبندی‌های مختلف، این روش‌ها هر کدام برای فرکانس‌های مختلف داده دو بار اجرا می‌شوند:

۱. تمام ویژگی‌ها به صورت concatinate شده

۲. با میانگین‌گیری از هر ستون (فرکانس)

در همه این حالات، داده‌های تست ۲۵٪ از کل داده‌ها را تشکیل می‌دهند و به صورت متعادل و با انتخاب شده‌اند. پیاده‌سازی توابع و مدل‌ها به کمک کتابخانه ScikitLearn انجام شده و کدها و نتایج کامل در دفترچه Clustering.ipynb موجود هستند. برای داده‌هایی که تعداد ابعاد بسیار بالا دارند، به دلیل دقت کمتر و پیچیدگی زمانی بالا، از اجرای خوشبندی agglomerative صرف نظر شده و تنها الگوریتم K-Means بر روی آن‌ها اجرا شده است. البته همراه در حالت میانگین ویژگی‌ها، هر دو الگوریتم اجرا و تست شده‌اند.

در نهایت، با بررسی حالات مختلف، بهترین حالت که مربوط به خوشبندی بر agglomerative روی داده‌های میانگین‌گرفته شده ۳۶ فرکانس Chroma هستند انتخاب شده و بر روی آن آزمایش‌ها و

تحلیل‌های کاملی صورت گرفته است.

۱-۱-۴ معیارهای سنجش عملکرد

Rand Score (Index) • تعداد تمام جفت نقاطی که در خوشهای متفاوت یا یکسان قرار می‌گیرند را با بروچسبهای واقعی داده‌ها مقایسه می‌کند.

$$RI = \frac{\text{number of agreeing pairs}}{\text{number of pairs}} \quad (1-4)$$

در ادامه از معیار Adjusted Rand Index (ARI) استفاده شده است.

Mutual Info • یک معیار متقارن شباهت میان دو بروچسب واقعی و پیش‌بینی شده برای یک داده است. برای دو خوشبندی ممکن U و V , $|U_i|$ تعداد نمونه‌های خوش U_i است و $|V_j|$ تعداد نمونه‌های خوش V_j :

$$MI(U, V) = \sum_{i=1}^{|U|} \sum_{j=1}^{|V|} \frac{|U_i \cap V_j|}{N} \log \frac{N|U_i \cap V_j|}{|U_i||V_j|} \quad (2-4)$$

در سنجش مدل‌ها از نرمالایز شده این معیار استفاده شده است Normalized Mutual Info (NMI).

Homogeneity Score • یک خوشبندی هنگامی یک‌دست است که هر یک از خوش‌ها شامل نقاط تنها یک کلاس باشند. اگر جای بروچسبهای حقیقی و پیش‌بینی شده در این معیار عوض شود، به معیار Completeness Score می‌رسیم.

Completeness Score • یک خوشبندی هنگامی به تمامیت می‌رسد که تمام نقاط هر کلاس باهم و در یک کلاس قرار بگیرند.

در جداول نتایج که در ادامه می‌آیند، برای سادگی عبارت score حذف شده و منظور از تمامیت و یک‌دستی، امتیاز این دو است.

۲-۱-۴ توابع و عملیات

برای هر یک از روش‌های خوشبندی از یک تابع wrapper استفاده می‌شود که پارامترهایی نظیر تعداد خوش و در خوشبندی agglomerative نوع لینک را دریافت می‌کند و در خروجی مدل مورد نظر را می‌دهد.

یک تابع برای برچسب‌زندن خوش‌ها بر اساس داده‌های اکثریت آموزش نیز استفاده شده که در رسم ماتریس پریشانی^۱ کاربرد دارد.

در نهایت، تابع evaluate() برای بازگرداندن معیارهای سنجش عملکرد مدل استفاده می‌شود که در قسمت پیشین به آن اشاره شد.

۳-۱-۴ Chroma فرانس

در این بخش، از ویژگی‌های استخراج شده Chroma بر روی قطعات ۲۰ ثانیه‌ای و با ۱۲ باند فرکانس در هر قطعه استفاده می‌شود. البته از این ابعاد با توجه به اینکه چندین برابر تعداد داده‌های آموزشی هستند انتظار دقت بالایی نخواهیم داشت.

ابتدا برای استفاده از تمام ابعاد هر یک از رکوردها نیاز داریم آن را «تخت» کنیم. برای این کار از کد زیر کمک می‌گیریم:

```
X12 = X.reshape(X.shape[0,] -1)
```

که در آن شکل کل داده‌های ورودی (پیش از تقسیم به آموزش و تست) از (12, 9538, 862) به (9538, 10344) تغییر می‌کند.

در ادامه هر یک از دو الگوریتم خوشبندی برای سه مقدار ۲، ۷ و ۲۰ خوش بر روی این داده‌ها اجرا می‌شوند. نتایج این بخش در جدول ۱-۴ مشاهده می‌شود. بهترین عملکرد برای خوشبندی K-Means با ۲۰ خوش حاصل شده که در آن دقت تست ۲۶٪ است.

در ادامه همین آزمایش را بر روی ابعاد کاهش یافته به کمک میانگین‌گیری انجام می‌دهیم. برای هر یک از داده‌ها، میانگین ۸۶۲ بخش برای هر یک از ۱۲ ستون گرفته می‌شود و در نهایت ۱۲ ویژگی

^۱Confusion Matrix

جدول ۴-۱: خروجی بر روی داده‌های Chroma 12، تمام ابعاد

method	k	dim	ARI	NMI	Homogeneity	Completeness	train acc	test acc
KMeans	2	10344	0.0002	0.0013	0.0009	0.0025	0.1707	0.1623
KMeans	7	10344	0.0076	0.016	0.0159	0.016	0.209	0.2109
KMeans	20	10344	0.0134	0.0408	0.0515	0.0337	0.2691	0.26
Agglomerative	2	10344	0.0021	0.005	0.0032	0.0119	0.1836	0.182
Agglomerative	7	10344	0.0091	0.0208	0.0203	0.0213	0.2132	0.2143
Agglomerative	20	10344	0.011	0.0366	0.0459	0.0304	0.2533	0.2474

خواهیم داشت که از لحاظ پیچیدگی زمانی بسیار بهینه‌تر است. برای میانگین‌گیری می‌توانیم از کتابخانه Numpy در قطعه کد ساده زیر استفاده کنیم:

```
X_mean = X.mean(axis=1)
```

شکل داده‌های ورودی در این حالت (9538, 12) می‌شود که بسیار کوچکتر از حالت قبل است.

نتایج آموزش این قسمت در جدول ۴-۲ مشاهده می‌شود. بهترین دقت این بخش نیز در مدل مشابه قسمت قبل حاصل شده است اما دقت آن در تست $27/3\%$ است که از حالت قبل اندکی بیشتر است. می‌توان از این موضوع نتیجه گرفت که میانگین‌گیری در این مورد کمک کرده است.

جدول ۴-۲: خروجی بر روی داده‌های Chroma 12، میانگین ویژگی‌ها (۱۲ بعد)

method	k	dim	ARI	NMI	Homogeneity	Completeness	train acc	test acc
KMeans	2	12	0.0003	0.0015	0.001	0.0028	0.1717	0.1639
KMeans	7	12	0.0077	0.0158	0.0158	0.0159	0.2093	0.2109
KMeans	20	12	0.0163	0.047	0.0595	0.0388	0.2728	0.273
Agglomerative	2	12	0.0012	0.0024	0.0015	0.0058	0.1777	0.1719
Agglomerative	7	12	0.0081	0.0206	0.0202	0.0209	0.2045	0.208
Agglomerative	20	12	0.0136	0.0424	0.053	0.0354	0.2673	0.2537

۴-۱-۴ فرکانس Chroma ۲۴

این بخش نیز بر روی ویژگی‌های Chroma اما این بار با ۲۴ فرکانس است. شکل اولیه داده‌ها در اینجا مشابه قبل است و تنها المان سوم از ۱۲ به ۲۴ تغییر یافته و در نتیجه (9538, 862, 24) برای کل دیتا است

جدول ۴-۳: خروجی بر روی داده‌های Chroma 24، تمام ابعاد

method	k	dim	ARI	NMI	Homogeneity	Completeness	train acc	test acc
KMeans	2	20688	0.0004	0.0019	0.0013	0.0037	0.171	0.1631
KMeans	7	20688	0.0127	0.0255	0.0253	0.0256	0.2273	0.2273
KMeans	20	20688	0.0185	0.0524	0.0661	0.0434	0.2638	0.2629

است. با تخت کردن ویژگی‌ها به ۲۰۶۸۸ بعد می‌رسیم. برای حالتی که از کل این ابعاد استفاده می‌کنیم، به دلیل دقت کم و زمان اجرای بالا، الگوریتم خوشبندی agglomerative را امتحان نخواهیم کرد.

نتایج این خوشبندی در جدول ۴-۳ دیده می‌شود. بهترین دقت تست در ۲۰ خوش بندی به دست آمده و معادل ۲۶٪ است که اندکی بیشتر از تمام ابعاد ۱۲ فرکانس اما کمتر از حالت میانگین آن است.

جدول ۴-۴: خروجی بر روی داده‌های Chroma 24، میانگین ویژگی‌ها

method	k	dim	ARI	NMI	Homogeneity	Completeness	train acc	test acc
KMeans	2	24	0.0004	0.0017	0.0012	0.0034	0.1728	0.1635
KMeans	7	24	0.0143	0.0271	0.027	0.0273	0.2297	0.2314
KMeans	20	24	0.0171	0.0536	0.0675	0.0445	0.2655	0.2625
Agglomerative	2	24	0.0047	0.0112	0.0075	0.0223	0.1711	0.174
Agglomerative	7	24	0.0163	0.0337	0.0334	0.0341	0.1975	0.226
Agglomerative	20	24	0.021	0.0614	0.0773	0.051	0.2648	0.2788

برای میانگین ۲۴ بعدی ویژگی‌ها از هر دو خوشبندی استفاده شده و بهترین نتیجه با خوشبندی agglomerative با دقت تست ۲۷٪ حاصل شده است که بیشتر از حالات قبل است (جدول ۴-۴).

۵-۱-۴ فرکانس، Chroma ۳۶

جدول ۵-۴: خروجی بر روی داده‌های Chroma 36، تمام ابعاد

method	k	dim	ARI	NMI	Homogeneity	Completeness	train acc	test acc
KMeans	2	31032	0.0008	0.0026	0.0017	0.0051	0.1706	0.1648
KMeans	7	31032	0.0104	0.0208	0.0208	0.0209	0.2108	0.2134
KMeans	20	31032	0.0133	0.0446	0.0562	0.0369	0.2591	0.2507

این بخش نیز بر روی ویژگی‌های Chroma و با ۳۶ فرکانس است. شکل اولیه داده‌ها در اینجا

(9538, 862, 36) برای کل دیتاست است. با تخت کردن ویژگی‌ها به ۳۱۰۳۲ بعد می‌رسیم. برای حالتی که از کل این ابعاد استفاده می‌کنیم، به دلیل دقیق کم و زمان اجرای بالا، الگوریتم خوشبندی agglomerative را امتحان نخواهیم کرد.

نتایج این خوشبندی در جدول ۵-۴ دیده می‌شود. بهترین دقت تست در ۲۰ خوشبندی دست آمده و معادل ۲۵٪ است که نشان می‌دهد تعداد ابعاد و در نتیجه پیچیدگی داده‌ها نسبت به مدل خوشبندی زیاد شده.

جدول ۶-۴: خروجی بر روی داده‌های 36 Chroma، میانگین ویژگی‌ها

method	k	dim	ARI	NMI	Homogeneity	Completeness	train acc	test acc
KMeans	2	36	0.0007	0.0024	0.0016	0.0048	0.1701	0.1652
KMeans	7	36	0.0121	0.0234	0.0233	0.0234	0.2076	0.2151
KMeans	20	36	0.0144	0.0456	0.0576	0.0377	0.2627	0.2537
Agglomerative	2	36	0.0077	0.014	0.0091	0.0297	0.1597	0.1941
Agglomerative	7	36	0.0169	0.029	0.0283	0.0298	0.2114	0.2382
Agglomerative	20	36	0.0212	0.0779	0.0952	0.066	0.2739	0.2876

برای میانگین ۳۶ بعدی ویژگی‌ها از هر دو خوشبندی استفاده شده و بهترین نتیجه باز هم با خوشبندی agglomerative با دقت تست ۲۸٪ حاصل شده است که بیشتر از حالات قبل است (جدول ۴-۶). این امر نشان می‌دهد که علاوه بر اینکه میانگین‌گیری بهتر عمل می‌کند، با افزایش ابعاد به ۳۶ همچنان عملکرد بهبود یافته و پیچیدگی بیش از حد نیست. نکته قابل توجه این است که معیارهای دیگر خوشبندی نیز در این حالت از تمام آزمایش‌های قبلی بهتر هستند (البته همچنان مقادیر بسیار از حد انتظار پایین‌تر هستند که به تعداد و نوع ویژگی‌ها در کنار پیچیدگی خود نمونه‌ها و تعداد کلاس‌ها برمی‌گردد).

۶-۱-۴ MFCC، ۴۰ فرکانس

جدول ۷-۴: خروجی بر روی داده‌های MFCC 40، تمام ابعاد

method	k	dim	ARI	NMI	Homogeneity	Completeness	train acc	test acc
KMeans	2	34480	0.0041	0.007	0.0047	0.0133	0.1694	0.1672
KMeans	7	34480	0.0057	0.0131	0.013	0.0133	0.1879	0.1907
KMeans	20	34480	0.0073	0.0262	0.0325	0.022	0.2092	0.2264

این بخش بر روی ویژگی‌های MFCC و با ۴۰ فرکانس است. شکل اولیه داده‌ها در اینجا (9538, 862, 40) برای کل دیتاست است. با تخت کردن ویژگی‌ها به ۳۴۴۸۰ بعد می‌رسیم. برای حالتی که از کل این ابعاد استفاده می‌کنیم، به دلیل دقت کم و زمان اجرای بالا، الگوریتم خوشبندی agglomerative را امتحان نخواهیم کرد.

جدول ۴-۸: خروجی بر روی داده‌های 40 MFCC، میانگین ویژگی‌ها

method	k	dim	ARI	NMI	Homogeneity	Completeness	train acc	test acc
KMeans	2	40	0.0042	0.0072	0.0049	0.0138	0.1693	0.1667
KMeans	7	40	0.0041	0.0104	0.0103	0.0104	0.1873	0.1865
KMeans	20	40	0.0084	0.0325	0.0407	0.027	0.2251	0.2297
Agglomerative	2	40	0.0016	0.0057	0.0038	0.0116	0.1698	0.168
Agglomerative	7	40	0.0064	0.0158	0.0153	0.0163	0.1924	0.197
Agglomerative	20	40	0.0108	0.0404	0.05	0.0339	0.2327	0.2335

نتایج این بخش (جدول ۷-۴ و ۸-۴) بسیار کمتر از حالت‌های قبل است که علاوه بر نوع ویژگی‌ها، به تعداد بیشتر آن‌ها نیز بر می‌گردد. برای مثال، دقت خوشبندی که در میانگین‌گیری قسمت‌های پیشین با افزایش باندهای فرکانسی بهبود می‌یابد اکنون به طرز قابل توجهی کاهش یافته است و می‌توانیم نتیجه بگیریم که ۴۰ ویژگی (در حالت میانگین گرفته شده) پیچیدگی مدل را بیش از حد زیاد کرده است.

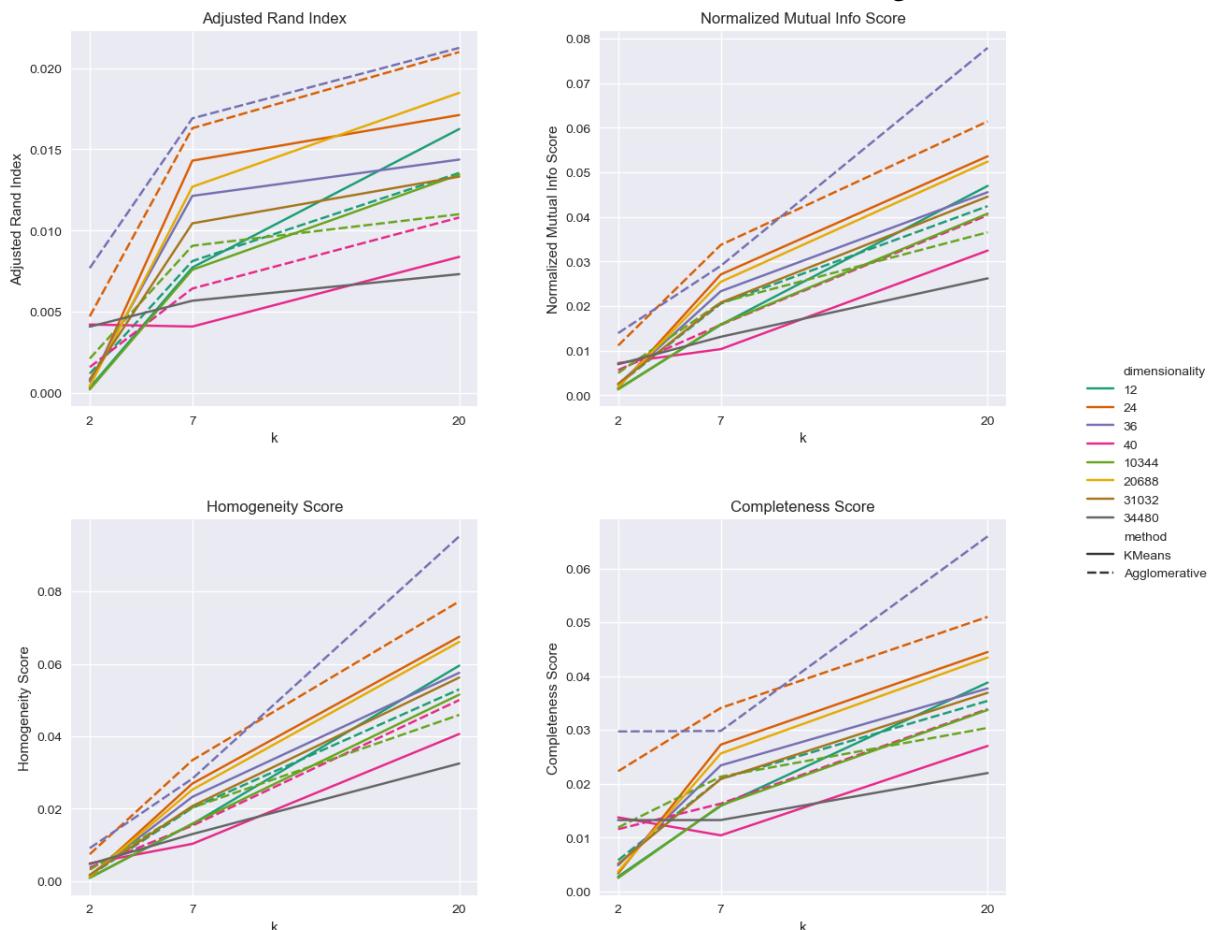
در ادامه، با توجه به نتایج آزمایش‌های انجام شده، بر روی بهترین گزینه که خوشبندی agglomerative بر روی داده‌های ۳۶ فرکانسی میانگین‌گیری شده و با ۲۰ خوشبندی است، آزمایش‌های بیشتری انجام می‌دهیم.

۷-۱-۴ مقایسه عملکرد

برای نمایش بهتر عملکرد هر یک از خوشبندی‌های بالا، می‌توانیم جدول‌های نتایج را در قالب چند نمودار نمایش دهیم. در نمودار ۱-۴ امتیاز‌های تمامیت و یک‌دستی و همچنین ARI و NMI برای خوشبندی‌های متفاوت مشاهده می‌شود. معیارهای خوشبندی در حالت کلی برای خوشبندی دوتایی بسیار ضعیفتر و در خوشبندی‌های ۲۰ تایی برای هر سری داده در بهترین حالت خود هستند. البته این اختلاف در معیار ARI کمتر مشهود است و در بقیه معیارها با شبیه بیشتری تغییر می‌کند.

توضیح علت عملکرد پایین مدل در خوشبندی دوتایی ساده است، زیرا تعداد خوشبندی‌ها از تعداد کلاس‌ها کمتر است و در نتیجه در صد خطای بسیار بالا می‌رود. همچنین معیار تمامیت در این خوشبندی

ممکن است بهتر از حالت هفت خوشهای باشد، چرا که احتمال بیشتری وجود دارد که هر خوشه شامل نقاط دورتر یک دستگاه نیز بشوند. از طرفی معیار یکدستی بسیار کم خواهد بود چون هر خوشه شامل داده‌های زیادی از کلاس‌های مختلف است.

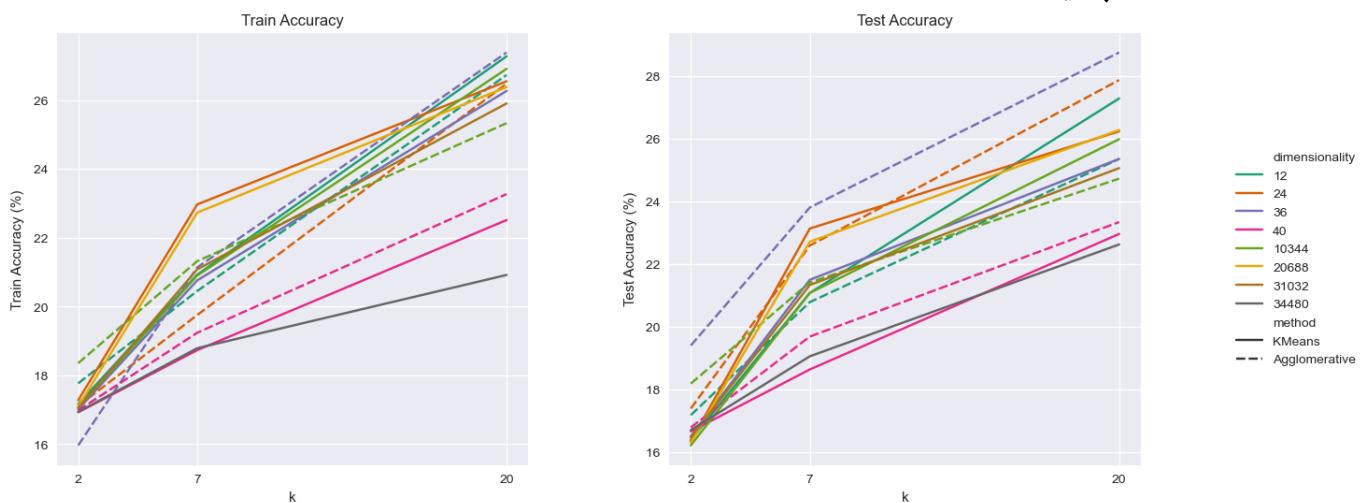


شکل ۴-۱: معیارهای عملکرد خوشبندی بر روی داده‌ها - (ساعتگرد از بالا چپ) ARI، NMI، امتیاز تمامیت و امتیاز یکدستی خوشبندی

در حالتی که خوشه‌ها برابر تعداد دستگاه‌ها هستند نیز الزامی وجود ندارد که به ازای هر دستگاه یک خوشه با آن برقسب وجود داشته باشد و ممکن است که داده‌های یک یا چند دستگاه به گونه‌ای میان داده‌های دیگر قرار گرفته باشند که به عنوان آن دستگاه برقسب نخورند و شباهت درون خوشه‌ای کمی به بقیه داده‌های آن خوشه داشته باشند. در اینجا مطابق انتظار و اکثراً در حالت‌هایی که ابعاد کمتری داریم، معیار تمامیت کاهش می‌یابد و یکدستی در تمام حالات افزایش یافته. در ادامه و با زیاد شدن خوشه‌ها، احتمال قرار گرفتن تمام داده‌های یک دستگاه در یک خوشه کاهش می‌یابد اما با توجه به اینکه خوشه‌ها منسجم‌تر و یکدست‌تر می‌شوند، نسبت به حالات قبل تمامیت نیز اندکی افزایش خواهد داشت.

در نمودار ۲-۴ دیده می‌شود که با وجود دقت نسبتاً نزدیک خوشبندی در آموزش، در تست عملکرد بسیار متفاوت است و به خصوص برای حالتی که ابعاد داده‌ها بالا هستند با پدیده‌ای مشابه بیش‌برازش روبرو هستیم، در صورتی که دقت تست داده‌های میانگین‌گرفته شده به طور کلی بیشتر است.

یک نکته حائز اهمیت دیگر اینست که در میان دو الگوریتم خوشبندی، و در تعداد خوشبندی مختلف، نمی‌توان مزیت خاصی در عملکرد هیچ یک از دو الگوریتم دید و نتایج آن‌ها در هم تنیده هستند. البته لازم به یادآوری است که الگوریتم agglomerative بر روی مجموعه‌های کمتری اجرا شده و این تحلیل کلی نیست، اما می‌توان فرض کرد به علت زمان بر بودن این الگوریتم برای ابعاد بسیار بالا مناسب نیست.



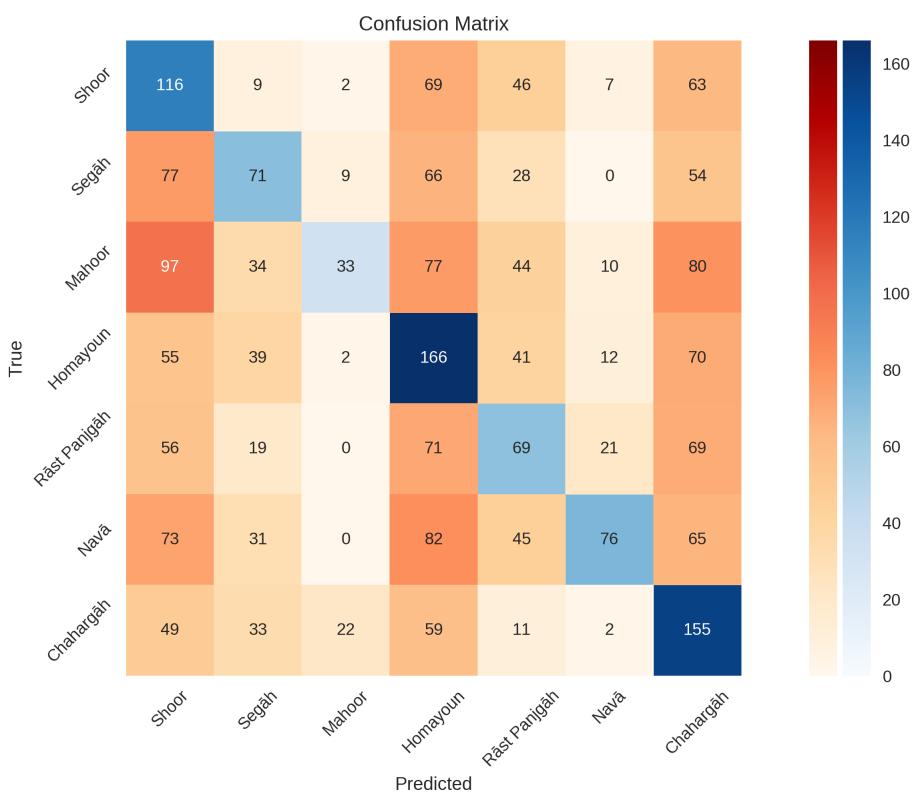
شکل ۲-۴: دقت آموزش (چپ) و تست (راست) خوشبندی بر روی داده‌ها

تحلیل کلاس به کلاس نتایج در قسمت بعد صورت می‌گیرد.

۲-۴ آزمایش‌های بیشتر و تحلیل نتایج

همانطور که نتایج جدول ۶-۴ نشان می‌دهند، بهترین دقت حاصل شده مربوط به حالت ۲۰ خوشبندی از الگوریتم agglomerative و تعداد ویژگی ۳۶ (میانگین‌گیری شده بر روی ۸۶۲ سطر از هر داده) است که چیزی حدود ۲۸٪ می‌باشد. در ادامه، نتایج این خوشبندی را بیشتر بررسی و تحلیل می‌کنیم و با تغییر پارامتر لینک در خوشبندی تلاش خواهیم کرد تا عملکرد آن را بهبود ببخشیم.

برای تحلیل نتایج و سنجش دقت به صورت باناظر، از برچسب‌های واقعی داده‌ها در تعیین برچسب



شکل ۴-۳: ماتریس پریشانی برای خوشبندی با ward linkage

هر خوشه بر اساس رأی اکثریت استفاده شده است. از این طریق می‌توانیم دقت پیش‌بینی در هر کلاس و همچنین ماتریس پریشانی را به دست آورده و تحلیل کنیم.

در تصویر ۴-۳ ماتریس پریشانی این خوشبندی آورده شده است. می‌بینیم که تعداد true positive برای دستگاه‌های همایون، چهارگاه و شور بیشتر از بقیه است و در مجموع به نظر می‌رسد با توجه به تعداد این سه برچسب، خوشه‌های بیشتری نیز به آن‌ها تعلق گرفته‌اند. در واقع، می‌توان گفت که داده‌های این سه دستگاه در فضای ویژگی‌ها پراکندگی بیشتری داشته‌اند و احتمالاً با دستگاه‌های دیگر درهم‌تنیده هستند. در جدول ۴-۹ معیارهای سنجش دقت به تفکیک کلاس بر روی داده‌های تست آورده شده‌اند.

این شباهت خوشه‌ها جالب توجه است، زیرا همانظور که در مقدمه (گزارش اولیه) گفته شد، دستگاه شور در واقع به عنوان «مادر دستگاه‌ها» شناخته می‌شود و دستگاه‌های دیگر مشترکات زیادی با آن دارند، این امر می‌تواند اشتباه‌های طبقه‌بندی را تا حدودی توضیح دهد. به خصوص در دستگاه‌های نوا و سه‌گاه که مستقیماً از شور منشعب می‌شود این شباهت در نتایج مشهود است.

جدول ۴-۹: نتایج و دقت خوشبندی بر روی داده‌های تست با ward linkage

	precision	recall	f1-score
Shoor	0.22	0.37	0.28
Segāh	0.3	0.23	0.26
Mahoor	0.49	0.09	0.15
Homayoun	0.28	0.43	0.34
Rāst Panjgāh	0.24	0.23	0.23
Navā	0.59	0.2	0.3
Chahargāh	0.28	0.47	0.35
accuracy			0.29
macro avg	0.34	0.29	0.27
weighted avg	0.35	0.29	0.27

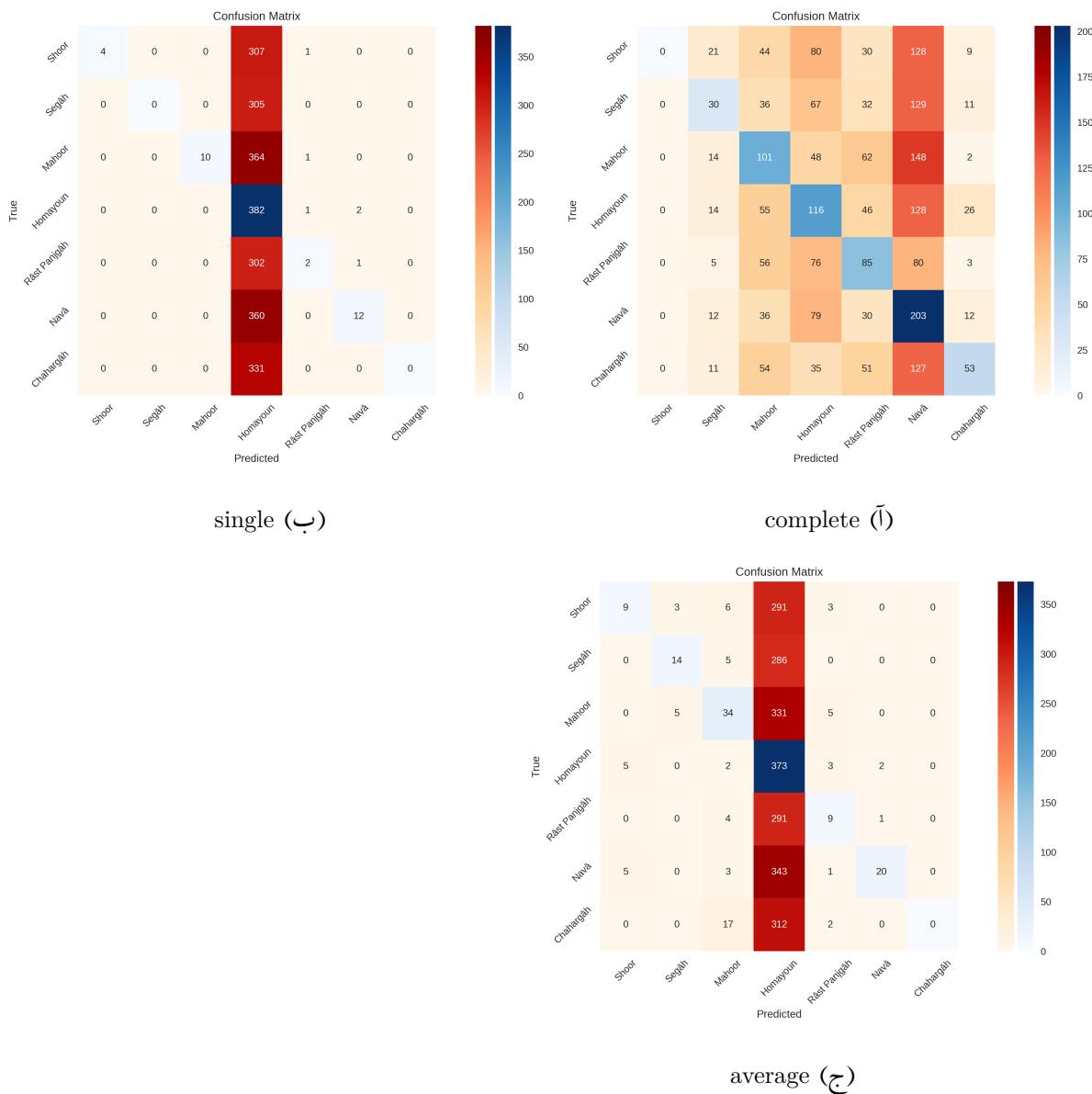
از طرفی، رفتار دستگاه راست پنجگاه و ماهور به هم شبیه است اما در ماتریس آشفتگی، این شباهت چندان مشهود نیست و نمونه‌های ماهور بیشتر به خوشبندی شور و چهارگاه نزدیک هستند و راست پنجگاه به همایون. البته لازم به ذکر است که ماهور یکی از گسترده‌ترین دستگاه‌های موسیقی سنتی و شامل تعداد زیادی گوشه‌های متنوع است که این امر می‌تواند دقت پایین در خوشبندی آن را توضیح دهد.

بیشترین امتیاز اف-۱ خوشبندی نیز طبق جدول ۴-۹ مربوط به کلاس چهارگاه و بعد از آن همایون است که به ترتیب $0/34$ و $0/35$ می‌باشد.

اما یک راه بهتر برای درک شباهت‌های کلاس‌ها، استفاده از پارامترهای لینک مختلف در خوشبندی agglomerative است. به طور پیشفرض، پیاده‌سازی کتابخانه ScikitLearn از لینک Ward استفاده می‌کند که واریانس خوشبندی که یکی می‌شوند را کمینه می‌کند. این معیار لینک با وجود عملکرد خوب، چندان شهودی نیست. در ادامه در سه حالت بیشتر شناخته شده «تکی»، «کامل» و «میانگین» این خوشبندی اجرا شده و نتایج بررسی می‌شوند. لازم به ذکر است که عملکرد خوشبندی در هیچ یک از این سه حالت بهتر از حالت «وارد» نیست و این کار بیشتر برای رسیدن به شهودی از نحوه کار

الگوریتم و فرارگیری داده‌ها در فضای ۳۶ بعدی انجام می‌شود.

آزمایش بر روی پارامترهای Agglomerative ۱-۲-۴



single (ب)

complete (ج)

average (ج)

شکل ۴-۴: ماتریس آشفتگی برای حالت‌های مختلف linkage

خوشبندی agglomerative علاوه بر پارامتر تعداد خوشبندی، دارای پارامتر دیگری به نام لینک^۴ است که نحوه سنجش فاصله برای تعیین نزدیکترین خوشبندی در مرحله یکی‌کردن^۵ را تعیین می‌کند. همانطور

linkage^۴
merging^۵

که پیشتر گفته شد، پارامتر پیشفرض خوشبندی‌های الگوریتم agglomerative در قسمت‌های پیشین بر اساس لینک‌های «وارد» بود که واریانس میان خوشبندی‌هایی که یکی می‌شوند را کمینه می‌کند. این روش معمولاً بهترین نتیجه را ارائه می‌دهد اما روش‌های شناخته‌شده‌تر دیگری نیز وجود دارند که به شرح زیر هستند:

- تکی (single): فاصله بین خوشبندی‌ها، فاصله نزدیک‌ترین نقاط آن دو خوشبندی در نظر گرفته می‌شود (حالت آسان‌گیرانه).
- کامل (complete): فاصله میان خوشبندی‌ها، فاصله دورترین نقاط آن دو خوشبندی در نظر گرفته می‌شود (حالت سخت‌گیرانه).
- میانگین (average): فاصله میان دو خوشبندی، میانگین فاصله تمام جفت نقاط دو خوشبندی از یکدیگر است. این روش با آنکه حد وسطی میان دو روش پیشین است، بار محاسباتی بیشتری دارد.

در شکل ۴-۴ ماتریس‌های آشفتگی برای سه پارامتر لینک مختلف در خوشبندی agglomerative بر روی داده‌های ۳۶ بعدی (میانگین‌گیری) شده مشاهده می‌شود. همانطور که می‌بینیم، در همه این خوشبندی‌ها، خوشبندی‌ها به سمت کلاس‌های مختلف بایاس دارند و به خصوص در پارامترهای لینک تکی و میانگین این موضوع مشهودتر است.

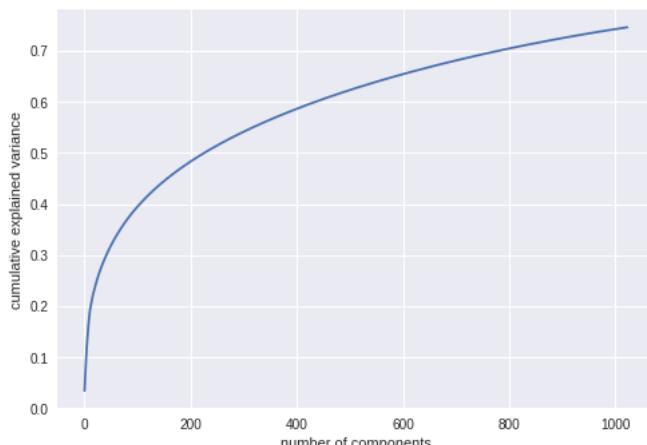
پارامتر لینک کامل که کمی سخت‌گیرانه‌تر است، دقیق‌تر نیز دارد (در حدود ۲۳ الی ۲۴ درصد در آموزش و تست) اما همچنان بایاس زیادی نسبت به کلاس نوا در آن مشاهده می‌شود. نکته جالب در این خوشبندی، عدم تشخیص تمام نمونه‌های دستگاه شور است که در حالت لینک «وارد» یکی از کلاس‌هایی بود که با دقیق‌تر از آن از کلاس نوا طبقه‌بندی شده‌اند بیشتر که تعداد تشخیص‌های صحیح آن از false negative کم است و می‌توان نتیجه گرفت که فاصله بیشتری از این دستگاه داشته است، البته این اختلاف همچنان بسیار کوچک است.

در هر دو حالت بعدی، بایاسی بسیار قوی نسبت به دستگاه همایون وجود دارد که ناشی از نحوه کارکرد لینک‌های تکی و میانگین است. برای مثال، در لینک ساده، در تئوری موسیقی، دستگاه شور به کلاس‌های دیگر نزدیک‌تر است؛ حال ممکن است یک خوشبندی از دستگاه سه‌گاه در نقطه‌ای نزدیک به یک خوشبندی از شور شود و این دو باهم ترکیب شوند. از طرف دیگر، در مرحله بعدی این خوشبندی

به یک خوشه با داده‌های دستگاه ماهور می‌شود و باز این سه خوشه ترکیب می‌شوند. دستگاه ماهور هم از نقطه‌ای دیگر به یک خوشه بزرگ‌تر از دستگاه همایون نزدیک می‌شود و این امر تا جایی ادامه پیدا می‌کند که تعداد زیادی نمونه از دستگاه‌های مختلف در یک خوشه قرار دارند. در نتایج شکل ۴-۴ مشاهده می‌کنیم که برخی دستگاه‌ها شاخص هیچ خوشه‌ای نیستند و میان داده‌های دستگاه‌های دیگر «گم شده‌اند». در خوشبندی لینک کامل و میانگین این اتفاق فقط برای یک دستگاه (به ترتیب شور و چهارگاه) رخ داده اما در حالت تکی که دقت کمتری دارد، هر دو دستگاه سه‌گاه و چهارگاه هیچ خوشه شاخصی ندارند. این در حالی است که دستگاه‌های دیگر هر یک چندین خوشه دارند.

۳-۴ تأثیر کاهش بعد با PCA

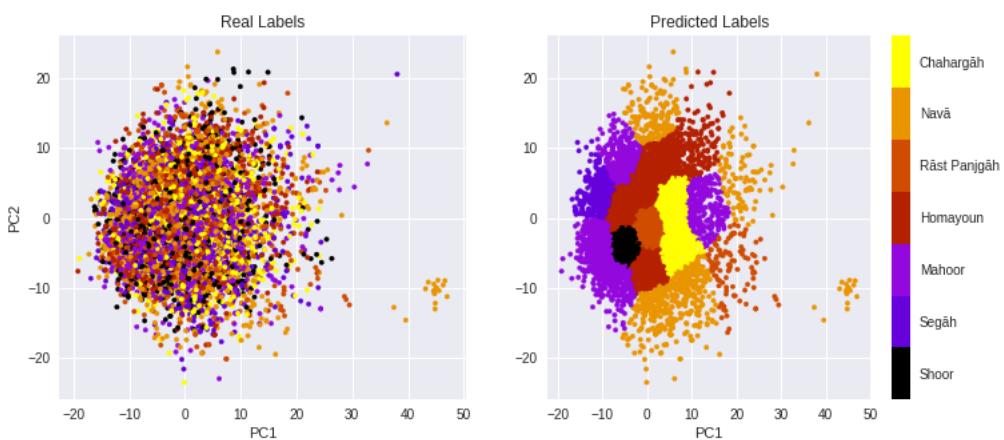
یکی از راه‌های بررسی شکل خوشه‌ها به صورت شهود، استفاده از روش کاهش بعد Principal Component Analysis (PCA) است. عملیات کاهش بعد را بر روی ویژگی‌های Chroma ۳۱۰۳۲ بعدی (برگرفته از تخت کردن همه ویژگی‌های ۳۶ تایی) انجام می‌دهیم.



شکل ۴-۵: نمودار Scree برای PCA

نمودار ۵-۴، در روش Elbow برای تعیین تعداد بهینه ابعاد استفاده می‌شود، اما از آن جایی که در داده‌ها پراکندگی در تمام جهت‌ها زیاد است و تعداد ابعاد نیز بالاست، تنها ۱۰۰۰ مولفه اول بررسی شده‌اند. نتایج تست بر روی سه مجموعه داده کاهش یافته به ۲، ۵۰۰ و ۱۰۰۰ بعد در جدول ۱۰-۴ مشاهده می‌شوند.

همانطور که مشاهده می‌شود، نتایج بر روی این داده‌ها از مدل اصلی کمتر است که البته قابل انتظار



شکل ۴-۶: استفاده از PCA در رسم داده‌ها: (چپ) برچسب‌های واقعی و (راست) برچسب‌های حاصل از خوشبندی ۲۰ تایی

جدول ۴-۱۰: نتایج و دقت خوشبندی بر روی داده‌های تست با PCA

n_components	method	ARI	NMI	homogeneity	completeness	test accuracy
2	kmeans	0.0076	0.0148	0.0139	0.0158	0.2151
500	kmeans	0.0283	0.0365	0.0354	0.0377	0.273
1000	kmeans	0.0213	0.0343	0.0319	0.0371	0.2495

می‌باشد، زیرا بخش از داده‌ها از دست رفته‌اند و روش PCA بر خلاف LDA جهت بیشترین پراکندگی داده‌ها را استخراج می‌کند و نه بیشترین جدایی‌پذیری.

با این وجود، استفاده از دو بعد، در ک شهودی خوبی از پراکندگی داده‌ها می‌دهد که در شکل ۴-۶ مشهود است. در این شکل داده‌های دستگاه‌های مختلف کاملاً در هم تنیده‌اند، و می‌توان گفت که خوشبندی تقریباً تصادفی قادر به تشخیص داده‌ها با ۲۱٪ دقت است. نکته حائز توجه، خوشبندی صحیح داده‌های پر تر مربوط به کلاس نوا (گوشه پایین راست تصویر) است که به دلیل دوری از بقیه داده‌ها به درستی طبقه‌بندی شده‌اند. در مجموع، خوشبندی دستگاه نوا بیشترین سهم و ۴ خوش از بیست خوش را تشکیل می‌دهند و تعداد نمونه‌های درست طبقه‌بندی این دستگاه به نسبت بیشتر از بقیه است و به طور مشابه، دستگاه شور تنها یک خوش کوچک و در نتیجه دقت کمتری (در حدود ۲۲٪) در مقابل ۳۸٪ نوا - جدول ۱۱-۴ دارد، این در حالی است که داده‌های دستگاه شور در تمام فضای ویژگی‌ها پراکنده‌اند.

جدول ۱۱-۴: نتایج و دقت خوشبندی بر روی داده‌های تست با PCA(n_components=2)، به تفکیک دستگاه Kmeans

	precision	recall	f1-score
Shoor	0.1809	0.2179	0.1977
Segāh	0.2195	0.118	0.1535
Mahoor	0.2058	0.3227	0.2513
Homayoun	0.2454	0.2753	0.2595
Rāst Panjgāh	0.2857	0.0328	0.0588
Navā	0.2154	0.3763	0.274
Chahargāh	0.2286	0.0967	0.1359
accuracy			0.21
macro avg	0.2259	0.2057	0.1901
weighted avg	0.2255	0.2151	0.196

۴-۴ نتیجه‌گیری

با بررسی تمام آزمایش‌های پیشین، می‌توان به این نتیجه رسید که با وجود آنکه روش خوشبندی agglomerative بر روی داده‌های میانگین گرفته شده قسمت ۵-۱-۴ بیشترین دقت را دارد، روش خوشبندی در کل برای این داده‌ها مناسب نیست و روش‌های طبقه‌بندی با ناظر پیشین در کنار استخراج ویژگی به کمک LDA نتایج بسیار بهتری در بر دارد.

در آزمایش‌های انجام شده، از دو روش شاخص برای خوشبندی مبتنی بر مرکز و فاصله (K-Means) و خوشبندی سلسله‌مراتبی (Agglomerative) استفاده شد. از میان سه نوع اصلی خوشبندی، تنها خوشبندی مبتنی بر چگالی آزمایش نشده که با توجه به فشردگی خوشبندیها و عدم تفاوت در چگالی (شکل ۶-۴) می‌توان به این نتیجه رسید که الگوریتم‌های چگالی محور مانند DBSCAN نیز نخواهند توانست داده‌ها را به خوبی خوشبندی کنند.

Bibliography

- [1] P. Heydarian and J. D. Reiss. A database for persian music. 2005.
- [2] N. Darabi, N. H. Azimi, and H. Nojumi. Recognition of dastgah and maqam for persian music with detecting skeletal melodic models. 2006.
- [3] S. Abdoli. Iranian traditional music dastgah classification. pages 275–280, 01 2011.
- [4] D. Ebrat, F. Didehvar, and M. Dadgar. Iranian modal music (dastgah) detection using deep neural networks, 2022.