

۱ آشنایی با مقاله

۱.۱ هدف کلی مقاله

سوال اینکه اطلاعات صوتی چگونه در مغز انسان کد می‌شود. هدف کلی این مقاله و تمام مقاله های از این دست یافتن جوابی کامل تر برای این سوال هست. اما هدف این مقاله در مرحله های پایین تر این است که بتواند با دقت بیشتر از شانس (۲۰ درصد) ۵ ژانر مختلف موسیقی را با بهره گرفتن از الگوریتم های ماشین لرنینگ دسته بندی بکند. مقاله های قبلی بیشتر در قسمت هایی از مغز تمرکز داشته اند که وظیفه ی آن قسمت ها از قبل برای پردازش تحریک های شنیداری مشخص شده بود، و همچنین تحقیقات قبلی اندکی شرایط آزمایش محدودتر در نظر گرفته میشد، برای مثال در سال ۱۹۹۰ میلادی Krumhansl تلاش کرد که فواصل گامی که توسط هر سابیجکت درک می شود را متوجه شود، به این طریق که یه ملودی توسط یک کلارینت اجرا می شد و به صورت همزمان گام آن لحظه از موسیقی استخراج می شد و پاسخ واکسل ها به یک مدل رگرسیون داده می شد. همچنین برای مثال این را متوجه شده بودند که تمپو که در حقیقت همان سرعت آهنگ است در قسمت اولیه و ثانویه بخش شنوایی با دقت خوبی کد میشود و همچنین کارهای قبلی موفق شده بودند با استفاده از گام های بالارونده و پایین رونده پاسخ واکسل ها را با استفاده از مدل MVPA بدست آورند. حال این مقاله کوشیده است تا با استفاده و ترکیب روش های

۱. شناخت موسیقی

۲. بازیابی اطلاعات موسیقی

۳. Multivoxel pattern analysis

خلاهای موجود در پژوهش های قبلی را تا حد خوبی پوشش دهد. Multivoxel pattern analysis در حقیقت یک روشی است که در آن پاسخ هر واکسل به عنوان یک متغیر پیوسته و بعد از فضای ویژگی ها در نظر گرفته میشود و با استفاده از آن می توان پاسخ های مختلف واکسل های مختلف را به تحریک های مختلف بدست آورد. با بهره گیری از این روش و همانطور که در کمی قبل ذکر شد، پژوهش های قبلی موفق شده اند که اطلاعات بسیار مفیدی از نواحی اول و دوم شنوایی بدست آورند. در این پژوهش چندین فرض شده است که اصلی ترین آن این است که ویژگی های مختلف موسیقی به صورت نوروینی در فضای حالت واکسل ها کد می شوند. به این صورت که پاسخ هر واکسل می تواند یک مشخصه از موسیقی را نشان بدهد. و فرضیه ی دوم این است که واکسل ها اطلاعات موسیقی را به صورت سلسله مراتبی کد می کنند، به این صورت که ابتدا اطلاعاتی نظیر ژانر و سبک کد می شوند، سپس ویژگی های جزئی تر مانند گام و ... و آخرین فرض این که ویژگی های مختلف موسیقی در قسمت های مختلف کد می شوند.

۲.۱ آزمایش انجام شده

در این تحقیق از ۲۰ نفر آدم آلمانی زبان که میانگین سنی آنها ۲۶ سال است و همچنین بدون هیچگونه سابقه ی بیماری نوروساینتسی و همچنین با سطح شنوایی عادی استفاده شده است. در این آزمایش از ۲۵ موسیقی با سبک های ambient, country, metal, rocknroll, symphonic که مدت هر کدام ۶ ثانیه است و انرژی در تمام این آهنگ ها نرمال شده است و همچنین ابتدا و انتهای آهنگ در یک فیلتر برای جلوگیری از حالت های گذرا ضرب شده است. تفاوت این آهنگ ها در وجود یا عدم وجود صدای انسان و سازهای ضربی است. کل آزمایش ۸ ران اجرا شده است. و هر ران شامل ۵ قطعه از ۵ ژانر مختلف است. اسکن ها با فاصله های ۲ ثانیه ای انجام شده است، اما در بین آنها تاخیرهایی ۴ یا ۶ یا ۸ ثانیه ای که به صورت رندوم بوده است وجود داشته است. در گاهی اوقات در حین اجرای آزمایش از سابیجکت ها سوالاتی از قبیل "خواننده مرد است یا زن؟" یا "این آهنگ شاد است یا غمگین؟" پرسیده میشد برای اطمینان از اینکه سابیجکت به آزمایش توجه می کند.

۳.۱ پیش پردازش داده ها

تمام داده با استفاده از fmri با قدرت ۷ تسلا تصویر برداری شده است. در کل ۸ ران موجود است که هر ران ۱۵۳ تصویر با $TR=2s$ تصویر برداری شده است. اصلاحات مربوط به حرکت سر و همچنین اعوجاج دیتا در حین تصویر برداری آنلاین انجام شده است. همچنین عکس های بعدی همگی همراستا با عکس های اولیه شده اند. هر تصویر شامل ۳۶ برش که به صورت یک صفحه ی $1.4mm \times 1.4mm \times 1.4mm$ است در حقیقت تنظیم کردن این بخش ها به وسیله ی بهینه کردن دقت مکانی بوده است. در هر تحریک چند عکس مرجع موجود است که با میانگین گیری عکس های مرجع به یک عکس مرجع واحد می رسیم که روی این عکس یک تبدیل زده می شود. GLM بر روی دنباله زمانی عکس ها فیت شده است.

۴.۱ Genre and song classification

در کلاسیفای کردن ژانرها و صحت درصد داده‌ها ما از کراس ولیدیشن استفاده می‌کنیم در کلاسیفای کردن آهنگ‌ها از hold-out validation استفاده شده‌است. همچنین با SVM عمل دسته بندی بر روی دیتای اسکیل شده انجام شده‌است. برای بدست آوردن واکسل‌های مهم از روش Anova استفاده شده‌است.

۵.۱ نتایج پژوهش

در این قسمت شایان توجه است که دقت تشخیص آهنگ با شانس ۴ درصد و تشخیص ژانر ۲۰ درصد است و نتایج مقاله خیلی بهتر از نتایج با شانس است. و میانگین تشخیص ژانر در حدود ۵۳ درصد است. نکته ی جالبی که وجود دارد احتمال خطا در قسمت هایی خطای رفتاری رخ داده‌است بیشتر است از جاهایی که خطای رفتاری نداشته‌ایم. شکل ۳ مقاله نتایج رفتاری میانگین‌گیری شده از ۲۰ شخص است که دقت بسیار بالاتر از SVM دارد. و نتیجه‌جالب دیگر که در شکل ۲ مقاله موجود بود، یکسان بودن تقریبی درصد ها در قسمت‌های مختلف مغز است.

۶.۱ بحث در مورد نتایج

در مورد نتایج می توان اینگونه اظهار نظر کرد که در دسته‌بندی آهنگ دقت نسبت به تحقیقات پیشین افزایش داشته‌است که علت این را می‌توان در استفاده از دستگاه fmri با قدرت بالاتر جستجو کرد. اما دقت تشخیص ژانر نسبت به تحقیقات قبلی کمتر شده است. اما کراس ولیدیشن بهتری استفاده‌است. چرا که ما زمانی ۱ ران را به عنوان داده‌ی تست نگه می‌داریم در اصل کار بهتری انجام داده‌ایم نسبت به اینکه cross-validation را به شیوه‌ی مرسوم انجام دهیم. در تحقیقات قبلی استفاده از fmri با قدرت ۳ تسلا برای دسته‌بندی کردن ژانر نتیجه‌ی بهتری می‌داد و در تحقیقات قبلی از ۱۰۰۰ واکسل برای استخراج ویژگی استفاده میشد اما استفاده‌ی ۵۰۰۰ واکسلی شاید دلیل دیگر باشد. یک نتیجه‌بسیار جالب کورلیشن بالای بین پاسخ رفتاری و پاسخ کلاسیفایر است. ambient and symphonic, country and rocknroll و نتیجه‌ی بسیار جالب دیگر که مهر تاییدی بر ادعا و فرض مقاله مبنی بر سلسله مراتبی بود پاسخ بخش شنیداری است، این است که اگر در تشخیص یک آهنگ دچار اشتباه شده‌است کلاسیفایر، خیلی احتمال بیشتری دارد که با یک آهنگ از دسته‌ی خودش اشتباه شده باشد، تا نسبت به دسته‌ی دیگر.

۲ آشنایی با دیتاست

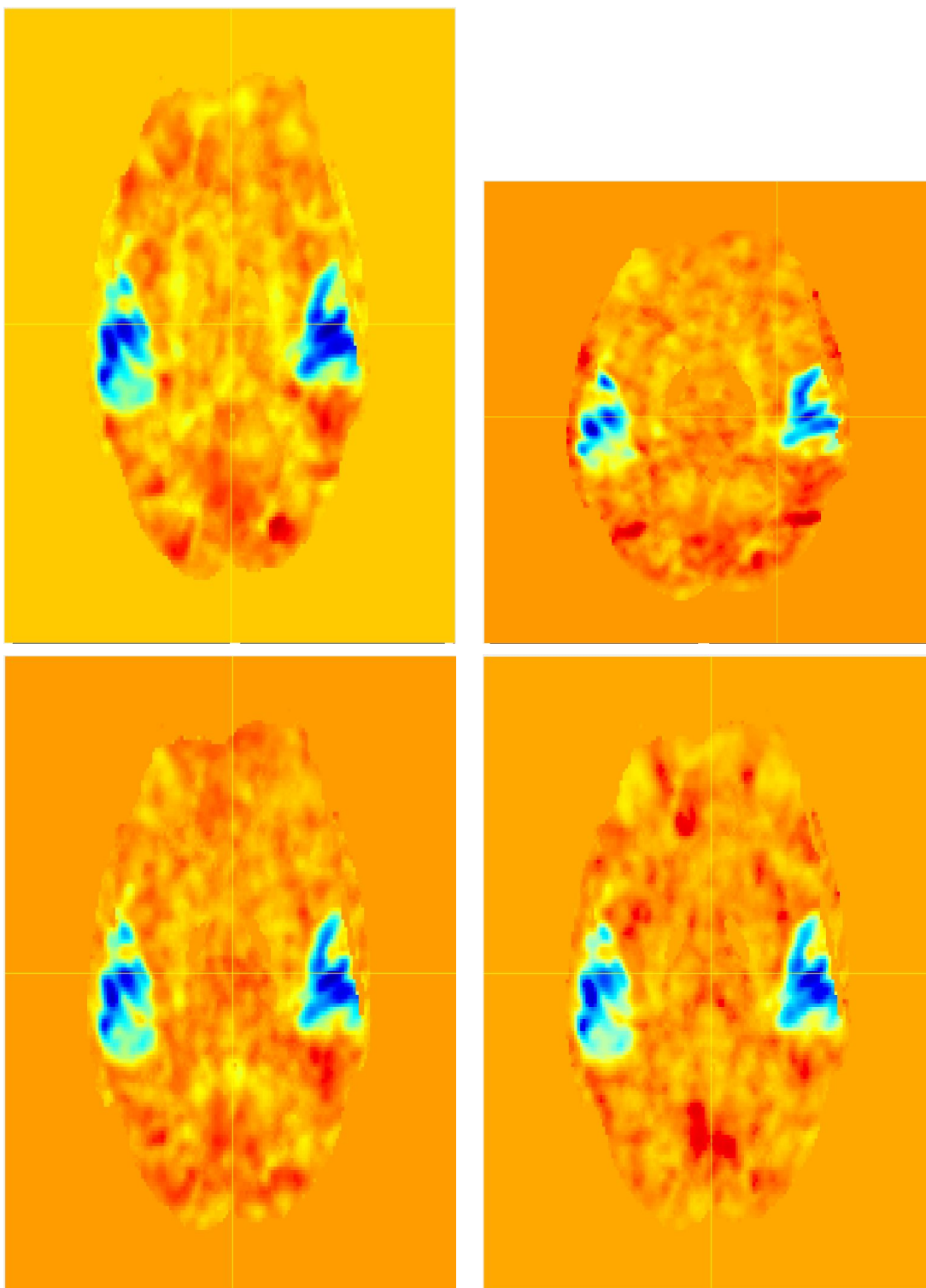
از آنجایی که این قسمت هیچ نمره‌ای ندارد ما نیز هیچ توضیحی برای آن نمی‌آوریم.

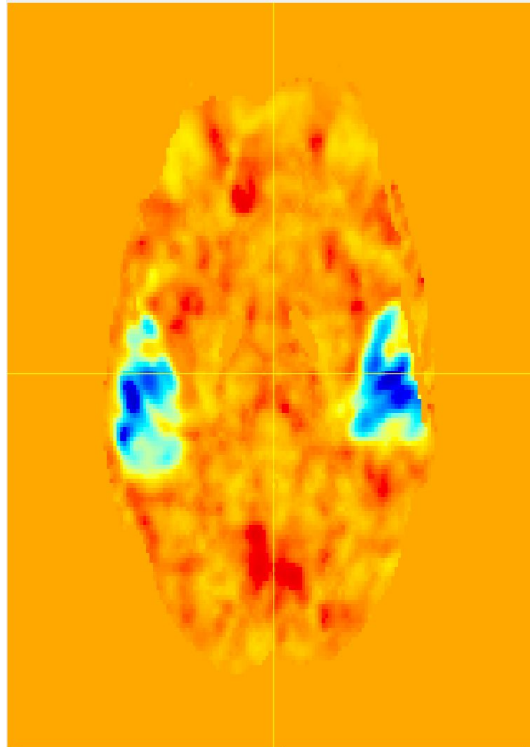
۳ بررسی نقاط فعال مغز در هنگام شنیدن موسیقی

۱.۳ بدست آوردن ماتریس Z-value

در این قسمت ما داده t-value را با استفاده از نرم‌افزار بدست می‌آوریم. ابتکاری که ما در اینجا بخرج دادیم نوشتن تابعی بود که بتاهای متناظر با هر کنتراست را بدون نیاز به رابط گرافیکی نرم‌افزار استخراج می‌کرد. باقی مراحل و جزئیات این روش درکد ضمیمه شده‌است.

۲.۳ فعالیت معنادار نقاط سر

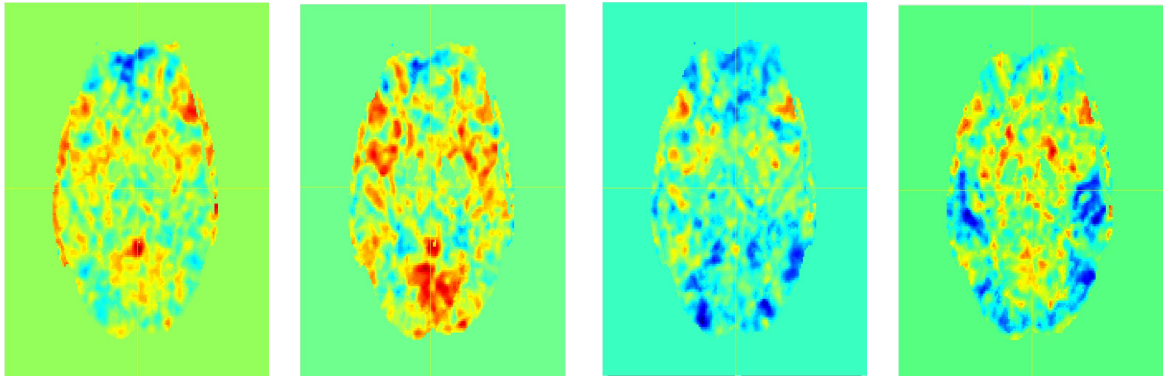




تنها اطلاعات مفیدی که از این عکس‌ها می‌توان برداشت کرد، کم بودن بتای متناظر با بخش شنیداری است، که به معنی آن است که این نقاط فعالیت قابل توجهی نسبت به نقاط دیگر مغز دارند. لازم به ذکر است که ما این عکس‌ها را به وسیله‌ی میانگین‌گیری بر روی ۷ ران موجود برای ۵ ژانر مختلف بدست آورده‌ایم. نکته‌ی فوق ارزشمندی که وجود دارد این است که به صورت دقیق و حتی شاید درست راجع به تفاوت این عکس نمی‌توان اظهار نظر کرد، اما با می‌توان گفت که بخش شنیداری مغز در کدام مورد بیشتر فعال است. تصویر راست از بالا و آخرین تصویر مربوط به سبک ambient, symphonic است که همانطور که می‌دانیم نسبت به سایر سبک‌ها ریتم آرام‌تری دارد و همانطور که در تصویر مشاهده می‌کنیم قسمت شنوایی مغز کمتر برای این سبک فعال است.

۳.۳ تفاوت معنادار

در این قسمت ما سبک متال را به عنوان سبک مرجع در نظر گرفتیم و باقی سبک‌ها را با آن سنجیدیم. لازم به ذکر است که این ۴ تصویر ضمیمه شده ناشی از میانگین‌گیری در چند عکس حساب شده است. نکته‌ای که در تایید نتیجه‌گیری ما در مورد سوال دوم این پارت صادق است این است که زمانی تفاوت متال با ambient را می‌سنجیم، می‌بینیم که بخش شنیداری در تصویر هنوز مقدار قابل توجهی آبی است. البته از این عکس می‌توان نتایج جالب دیگری نیز استخراج کرد که آن‌ها را به صورت خلاصه بیان می‌کنیم. در متال نسبت به ambient قسمت شنوایی و پشت سر فعالیت بیشتری دارد. در متال نسبت به کانتری پشت و جلوی سر فعالیت بیشتری دارند. در متال نسبت به rocknroll, symphonic جلوی سر فعالیت بیشتری دارد.



۴.۳ تعبیر ما

نقاطی که به شدت آبی یا همچنین به شدت قرمز بودند آنهایی بودند که آزمون فرض برای آنها رد شده است و در حقیقت اینها نقاطی هستند که اگر ما بخواهیم ژانر را دسته بندی بکنیم به شدت برای ما مفید خواهند بود. ما می توانستیم که در قسمت های قبل برای ماتریس Z-value یک ترشهولد قرار بدهیم به این نحو که بعضی نقاط کمتر مهم را حذف کنیم، اما مشاهده کردیم که دیگر با توجه به نرم افزاری که ما با آن کار کرده ایم نمی توانستیم کلیت مغز را نشان بدهیم فلذا از اینکار اجتناب کردیم. قسمت هایی از مغز که به صورت قطعی هنگام پردازش صوت فعالیت می کنند را ما در سوال ۲ این قسمت مشاهده کردیم، در مورد سوال ۳ این قسمت می توان نتیجه هایی گرفت که البته بنظر ما خیلی اصالت ندارد.

۴ قسمت چهارم

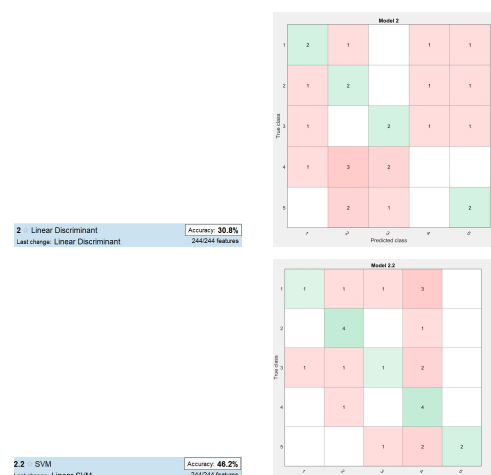
۱.۴ بدست آوردن ماتریس Train

پس از کلی عملیات طاقت فرسا که جزئیات آن ها در کد آمده است توانستیم بالاخره به این ماتریس برسیم. به علت طولانی بودن عملیات مجبور به این شدیم که ماتریس ها را ذخیره کنیم تا از تولید دوباره ی آنها اجتناب بشود.

۲.۴ Anova

در این مرحله نیز آزمون واریانس را بر روی بردار p-value را بدست میاوریم. حال قرار است با استفاده از این بردار تشخیص دهیم که کدام واکسل ها عملکرد خوبی دارند. هر ستون که مقدار کمتری داشت متناظر است با واکسلی که برای جداسازی مناسب است.

۳.۴ LDA classifier



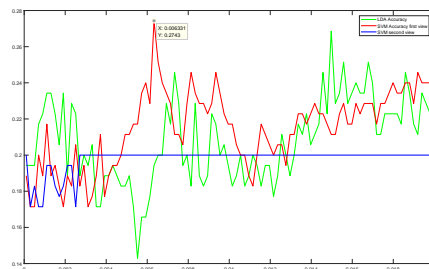
در این قسمت ۲ کار صورت گرفته است. با استفاده از روش

hold-out validation که در آن هر بار ۱۵ درصد معادل یک ران را به عنوان تست استفاده می کرد. یک بار این دقت برای

SVM و یکبار برای LDA بدست آوردیم. نکته ای که فوق العاده مهم است این است که اگر ما در SVM داده‌ها را نرمال کنیم به نتایج بسیار بهتری می‌رسیم در حالی که برای LDA لازم نیست دیتا را نرمال کنیم. منظور از نرمال کردن در راستای ستون‌هاست. که در شکل های بالا SVM که ماتریس ترین نرمال شده است و در شکل پایین LDA که ماتریس ترین نرمال نشده است. نکته ی بسیار جالب این است که این دقت‌ها با واکسل هایی بدست آمده اند که مقدار p-value کمتر از ۱ هزارم دارند که معادله ۲۴۴ واکسل است بدست آمده است.

۴.۴ سوال سوم و چهارم و پنجم

در این قسمت به علت پیوستگی بین پاسخ ها به تمام این سوال‌ها یکجا پاسخ داده می‌شود. ما در این قسمت به ۲ دیدگاه به SVM برای ترین کردن نگاه می‌کنیم. در linear SVM برای معمولاً خیلی بهتر است که دیتای خود را نرمال کنیم و سپس به کلاسیفایر می‌دهند که البته این نرمال کردن به ۲ طریق انجام میشود. اول اینکه ماتریس تست را با میانگین و واریانس ماتریس ترین نرمال کنیم و دوم اینکه این ماتریس را برخوردش نرمال کنیم. که نمودار ما نمایش می‌دهد که روش اول به مراتب بهتر از روش اول است. حال همانطور که در این نمودار مشاهده می‌شود به ازای $p=0.00633$ برای SVM این درصد حداکثر می‌شود



و برای LDA به ازای مقدار $p=0.01497$ این درصد حداکثر می‌شود. پس ما با این ۲ تا p-value خاص دقت را کلاسیفایر را گزارش می‌کنیم. برای مدل SVM انحراف معیار برابر 0.078 و میانگین برابر 27.43 و تنها از ۹۳ واکسل استفاده شده است و برای مدل LDA با همان ترشهولد مخصوص به خودش انحراف معیار برابر 0.091 و میانگین برابر 26.86 میشود و از ۵۰۰ واکسل استفاده میشود. که البته لازم به ذکر است که حداکثر درصد های هر کدام بسیار با میانگین متفاوت است.

۵.۴ سوال ۶ امتیازی

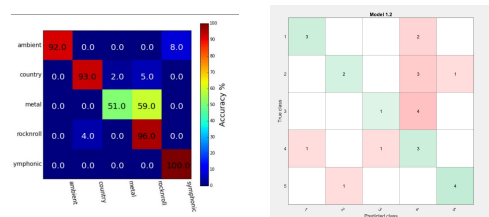
۶.۴ سوال ۷ امتیازی

تمام قسمت های قبلی نیز علاوه بر LDA از SVM نیز استفاده شده بود. حالا به بررسی روش Logistic Regression می‌پردازیم. در این روش که با استفاده از مینیم کردن خطاوری تصمیم‌گیری باینری بدست آمده است و با فیت کردن یک تابع sigmoid که احتمال را به ازای هر مشاهده به ما می‌دهد کار را انجام می‌دهیم. برای این روش از آماده‌ی متلب استفاده می‌کنیم. این تابع به هر مشاهده یک احتمال نسبت می‌دهد که ما به آن مشاهده عددی را نسبت می‌دهیم که بیشترین احتمال دارد. جالب است که روش رگرسیون بسیار وقت‌گیر بود و اینکه فقط توانستیم مقدار آن را برای $p=0.001$ بدست آوریم که دقتی میانگینی معادله 31.65 و انحراف معیار 0.087 بود.

۷.۴ سوال هشتم

در این سوال دیگر با کلاسیفایر linear SVM کار می‌کنیم همانطور که در خود مقاله کار شده بود. فقط نکته‌ای که قابل ذکر است این است که در confusion matrix ما عدد ۱ و ۲ و ۳ و ۴ و ۵ به ترتیب معادله ambient, country, metal, rocknroll, symphonic هستند.

مشاهده می کنیم که پاسخ ما تا حد بسیار عالی به پاسخ مقاله



نزدیک است و از نتیجه لذت می بریم.

۸.۴ سوال نهم

۲ بردار ضمیمه شده است که دومی مربوط به کار انجام شده در قسمت سوال دلخواه است.

۹.۴ سوال دهم

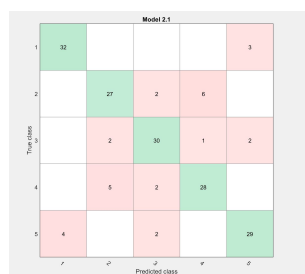
به صورت کلی ما درصدهایی نزدیک به مقاله دریافت کردیم اما نکته ای که مهم بود این بود که درصدهای ما زمانی که میانگین می گرفتیم به شدت پایین تر می آمدند. اما در حالی که ما با کراس ولیدشن به شیوهی مرسوم پیش می رفتیم حتی درصدهای بالای ۶۰ درصد نیز می گرفتیم. که البته این نکته خیلی هم دور از انتظار نبود، چراکه ما زمانی که از شیوهی معمولی کراس ولیدیشن استفاده می کنیم داریم و همچنین از آزمون آنوا استفاده می کنیم در حقیقت داریم مقداری اطلاعات اضافی وارد سیستم می کنیم.

۵ یک سوال دلخواه

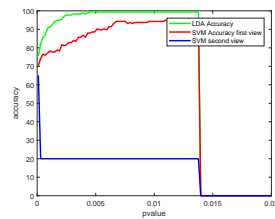
در این قسمت ما ۲ سوال اساسی بنظر بسیار مهم را مطرح خواهیم کرد.

۱۰.۵ سوال اول

سوال اول ما این مسأله را بررسی می کنیم که شاید دسته بندی تحریک ها براساس BOLD حتی از دسته بندی با Z-value بهتر باشد. برای این منظور ما اطلاعات موجود در فایل ها با فرمت niftiread را با استفاده از تابع convertnii2mat استخراج می کنیم. همانطور که در صورت مقاله توضیح داده شد، ما برای هر تحریک که شامل یک آهنگ ۶ ثانیه ای است، ۶ تصویر داریم که ۳ تصویر اول مربوط به زمان پخش آهنگ می باشند. ما میانگین ۳ تصویر ابتدایی را به عنوان پاسخ واکسل ها به تحریک حساب می کنیم. در نتیجه ما به ماتریس تحت عنوان BOLD-train ذخیره می کنیم و همچنین لیبل ها را با استفاده از فایل های tsv استخراج می کنیم. حال به انجام آزمون واریانس Anova بر روی ماتریس بدست آمده می پردازیم. نتیجه بسیار جالب هست که حتی زمانی که ترشهولد را تا مقدار یک ده هزارم پایین می آوریم ۳۰۴۱ واکسل مهم برای ما باقی می ماند. زمانی که کلسیفایر LDA را بر روی این ماتریس با استفاده از 5-fold cross validation پیاده سازی می کنیم نیجه بسیار جالب و هیجان انگیز است. درصد صحت ما 83.4 درصد است و درصد ها واریانس نسبتا کمی دارند.



شکل ۱: کانفیوژن ماتریس متناظر با کلسیفایر LDA



۲.۵ سوال دوم

حال به سراغ سوال دوم می‌رویم. موضوع اصلی این سوال بررسی وابسته به زمان بودن واکسل‌ها است. همانطور که در متن مقاله اشاره شد، ما ۲۵ قطعه موسیقی را در ۸ بار ران برای هر سابجکت پخش می‌کنیم، که البته ما لیبل‌های متناظر با ران هشتم را در اختیار نداریم فلذا از ۷ ران ابتدایی استفاده می‌کنیم. برای این قسمت پاسخ تمام واکسل‌ها را در نظر می‌گیریم و با هیچ آزمونی هیچ‌کدام را رد نمی‌کنیم. فرض بزرگی که ما در این قسمت انجام داده‌ایم این است که شرایط شنیدن یک موسیقی خاص در یک ران با ران دیگر هیچ تفاوتی نمی‌کند. ادعا این است که در پس از شنیدن تحریک یکسان پس از بارهای متوالی از پاسخ واکسل‌ها نسبت به بار اول کم می‌شود تا اینکه به یک حد خاصی اشباع شود. ما این را رسم نمودارهایی نمایش داده ایم و مشاهده می‌کنیم که نتایج تا مقدار خوبی با نتایج ما سازگار است. و شاید بتوان این ادعا را تایید کرد که واکسل‌ها برای تحریک یکسان در دفعات بعدی نسبت به بار اول مقدار کمتری اکسیژن مصرف می‌کنند.

