

با اسمه تعالی



دانشگاه صنعتی شریف

دانشکده مهندسی برق

## درس علوم اعصاب محاسباتی

گزارش پروژه درس

سری چهارم

تشخیص ژانر موسیقی با استفاده از دیکود کردن دیتای FMRI

استاد درس:

دکتر کربلایی آقاجان

امیرحسین افشار راد

۹۵۱۰۱۰۷۷

طaha antchari

۹۵۱۰۱۱۱۷

۱۳۹۷ بهار

## قسمت اول) آشنایی با مقاله پژوهش اصلی

### Terminology

**EPI=echo-planar imaging**

**FLIRT= FMRIB's Linear Image Registration Tool**

**MCFLIRT= motion correction FLIRT**

**FNIRT= FMRIB's nonlinear image registration tool**

**GLM =generalized linear model**

### بخش ۱ - ۱

هدف اصلی این پژوهش را می‌توان پردهبرداری از شبکه‌های پردازش موسیقی و تجربه موسیقی دانست. فرضیه اصلی که در این پژوهش به آن پرداخته‌اند، به این شکل قابل بیان است: انکودینگ ویژگی‌های متمایز موسیقی، به صورت نمایشی از فضاهای ساخته‌شده توسط واکسل‌ها است که هر بعدی از این فضا، پاسخ‌های متمایز هر واکسل به تحریک موردنظر است.

البته در مقاله ۲ ادعای دیگر نیز بیان شده است، به شرح زیر:

۱. فضای چندبعدی پیشنهادشده تمامی اطلاعات موسیقی از جامع‌ترین ویژگی‌ها مثل ژانر و نوع موسیقی تا ریزترین اطلاعات آن مانند شدت صوت را کد می‌کند.

۲. ویژگی‌های متفاوت موسیقی به صورت جدا از هم در ناحیه‌هایی که اطلاعات موسیقی پردازش می‌شوند، کد می‌شوند.

در کارهای قبلی، تلاش برای بدست آوردن پاسخ تنظیمی گروه‌های واکسل‌ها به ویژگی‌های صدا در قسمت‌های اصلی پردازش صوت مغز بوده است اما در این پژوهش به قسمت‌های دیگر مغز که مسئولیت اصلی آن‌ها پردازش صوت نیست نیز پرداخته‌شده است. در این پژوهش با استفاده از شناخت‌های ۳ عرصه علم، شناخت موسیقی، بازیافت اطلاعات موسیقی و تحلیل الگوهای مالتی واکسل‌ها، سعی در شناخت کدینگ موجود برای موسیقی در مغز است.

### بخش ۱ - ۲

در این پژوهش از داده‌های آزمایش پیشین – Hanke (2014) , Hanke (2015) – استفاده شده است. شرح آزمایش مورد استفاده در این پژوهش به صورت زیر است:

داده‌گیری fMRI از ۲۰ نفر راستدست با شنایی سالم بوده است. تمامی سابجکت‌ها از آزمایش و ماهیت آن اطلاع داشتند. تحریک در این آزمایش، ۲۵ کلیپ صوتی ۶ ثانیه‌ای با فرکانس سمپلینگ ۴۴/۱ کیلوهرتز بوده است. ۲۵ کلیپ صوتی، ۲۵ کلیپ ثابت از ۵ ژانر، هر ژانر ۵ کلیپ بوده است. ۵ ژانر تحریک عبارتند از:

Ambient, Rock & Roll, Country, Symphonic, Heavy metal

که اطلاعات هر ۲۵ آهنگ در مقاله آمده است. هر کلیپ ۶ ثانیه از وسط آهنگ موردنظر گرفته شده است. همچنین انرژی کلیپ‌ها به گونه‌ای متناسب با یکدیگر تنظیم شده است تا RMS توان آن‌ها یکسان باشد.

به ۵۰ میلی ثانیه اول و آخر هر کلیپ موجی سینوسی میرا اعمال شده است تا گذار به آهنگ و پایان آن نرم باشد. بعد از کالیبره سازی اولیه، ۸ سری آزمایش اجرا شده است. برای حفظ توجه افراد در طی آزمایش ۵ بار از آنها سوال بله/خیر پرسیده می شود، یک بار به ازای هر ژانر. البته این پرسش ها تنها برای حفظ توجه فرد بوده اند و در تحلیل آزمایش استفاده نشده اند.

ترتیب آهنگ ها با استفاده از چرخه De Bruijn انتخاب شده است. بین آهنگ ها تاخیر های ۴، ۶ و یا ۸ ثانیه ای بوده که ترتیب این تاخیرها نیز برای هر فرد رندم می شود.

### بخش ۱ - ۳

بنابراین گفته مقاله، نحوه پیش پردازش و بهبود اثرهای حرکت و اعوجاج تصویر مشابه مقاله (Hanke 2014) است.

برای اثر حرکت سر، در ابتدای آزمایش و در طول آن با بررسی های زنده با رفرنس اثر این اعوجاج را کمینه می کنند.

تصاویر EPI با استفاده قالب های برنامه های FLIRT, MCFLIRT و FNIRT گروه بندی و تصحیح شده اند. برای تنظیم مغز سابجکت های مختلف، با قالب های مختص به گروه، تصاویر EPI بدست آمده اند. تصاویر کاربردی افراد، با استفاده از MCFLIRT، به رفرنس اولیه آزمایشان نگاشت می شود. از میانگین تمامی تصاویر حاصل شده برای هر سابجکت نیز برای بدست آوردن قالب مختص آن فرد استفاده شده است. با استفاده از برنامه FLIRT قالبی نهایی از تمامی افراد بدست آمده و از این قالب برای بیشینه کردن تقاطع مغز هر فرد برای درست کردن قالب گروهی EPI استفاده شده است. پس از این مراحل، هر تحریک داده های EPI با استفاده از مدل خطی کلی بر روی ویژگی های واکسل ها تصویر می شوند.

### بخش ۱ - ۴

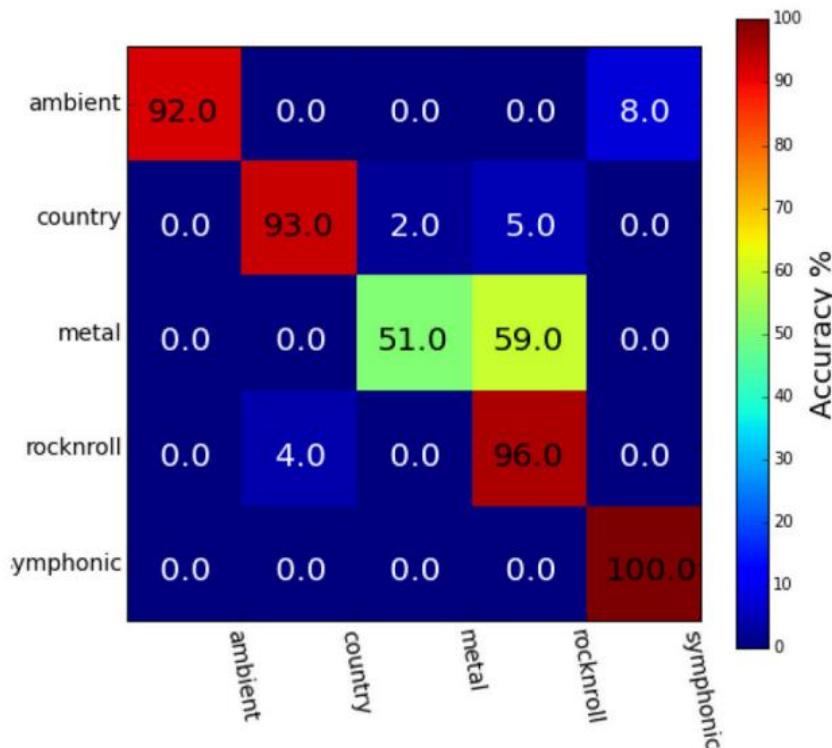
همانگونه که گفته شد از اهداف مقاله تشخیص موسیقی و تشخیص ژانر موسیقی است. در این راستا، برای تشخیص موسیقی، کل داده های یک آزمایش از ۸ آزمایش را برای کراس ولیدیشن نگه داشته و بقیه را برای یادگیری به SVM خطی تحويل میدهند. نتایج حاصل این کراس ولیدیشن به صورت میانگین برای تمامی ۲۰ سابجکت به صورت میانگین حدود ۲۱ درصد است که در این آزمایش تشخیص شانسی دارای احتمال ۴ درصد است.

برای تشخیص ژانر، در کراس ولیدیشن تمامی آزمایش های یک موسیقی از هر ژانر را برای تست نگه داشته و بقیه را به همان SVM خطی می دهیم. در این قسمت هم به صورت میانگین ۵۳ درصد جواب گرفته شد که برای این قسمت تشخیص شانسی دارای احتمال ۲۰ درصد است.

برای کاهش اثر واکسل های دارای نویز، برای SVM تنها از ۵۰۰۰ واکسل که دارای بزرگترین F-score بودند استفاده شده است.

**بخش ۱ - ۵**

برای ۳ ناحیه موردنظر از temporal cortex به صورت میانگین در خوشبندی آهنگ به درصد صحبت میانگین ۲۱ در کراس ولیدیشن رسیده‌اند و در خوشبندی ژانر آهنگ به درصد میانگین ۵۳ دست یافته‌اند. البته با خوشبندی رفتاری، که نظر خود شخص در مورد ژانر موسیقی پرسیده شده‌است، به درصد صحبت  $\frac{86}{4}$  درصد دست یافته‌اند که ماتریس خطای این طبقه‌بندی به شکل زیر است. همانطور که دیده می‌شود بیشترین خطای برای جداسازی ژانر metal از rock & roll و سپس ambient از symphonic بوده است که این را هم می‌توان به ویژگی موسیقی این ژانرها نسبت داد.

**بخش ۱ - ۶**

نتایج خوشبندی آهنگ در این پژوهش از پژوهش پیشین (Guntupalli 2013) بهتر است که یک علت آن را می‌توان میدان مغناطیسی ۷ تسلای نسبت به میدان مغناطیسی ۳ تسلای استفاده شده در آزمایش پیشین دانست چرا که با افزایش شدت میدان مغناطیسی، رزولوشن مکانی واکسل‌ها افزایش می‌یابد. اما در عین حال نتایج خوشبندی ژانر موسیقی از مقاله ذکر شده و همچنین مقاله (Casey 2012) کمتر است، با وجود این که دقت مکانی این آزمایش بیشتر بوده است. فرق عمده این پژوهش و دو پژوهش ذکر شده در انتخاب واکسل‌ها برای استفاده در عملیات خوشبندی است. در پژوهش اول ۱۰۰۰ واکسل از نقاط مختلف مغز و در پژوهش دوم حساسیتی بر روی واکسل‌ها نبوده است.

از این مقایسه‌ها نتیجه‌گیری شده است که خوشبندی با استفاده از داده‌های واکسل‌ها از کل مغز، اگرچه با دقت ۳ تسلای نتیجه بهتری نسبت به خوشبندی با استفاده از واکسل‌های temporal cortex، که وظیفه پردازش اصلی صوت را بر عهده دارند، اگرچه با دقت ۷ تسلای حاصل می‌دهد.

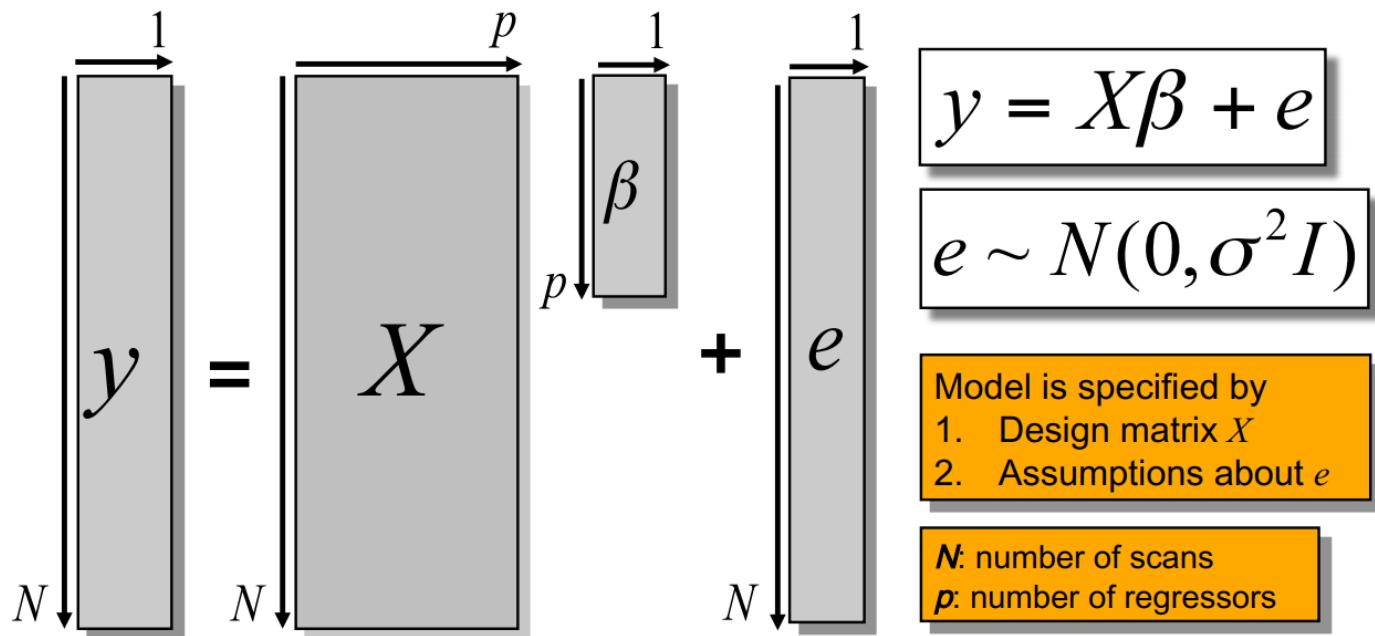
**Mentioned papers:**

1. Hanke, M., Baumgartner, F. J., Ibe, P., Kaule, F. R., Pollmann, S., Speck, O., et al. (2014). A high-resolution 7-tesla fMRI dataset from complex natural stimulation with an audio movie.
2. Hanke, M., Dinga, R., Häusler, C., Guntupalli, J. S., Casey, M., Kaule, F. R., et al. (2015). High-resolution 7-tesla fmri data on the perception of musical genres—an extension to the studyforrest dataset.
3. Guntupalli, J. S. (2013). Whole Brain Hyperalignment: Intersubject Hyperalignment of Local Representational Spaces. Dartmouth College, Ph.D. Thesis.
4. Casey, M., Thompson, J., Kang, O., Raizada, R., and Wheatley, T. (2012). “Population codes representing musical timbre for high-level fmri categorization of music genres,” in Machine Learning and Interpretation in Neuroimaging

### قسمت سوم) بررسی نقاط فعال مغز در هنگام شنیدن موسیقی

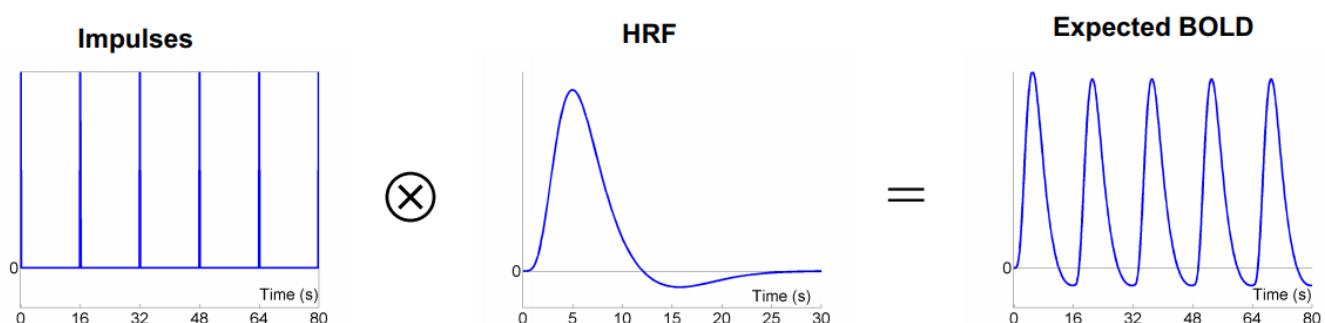
#### بخش ۱-۳

ابتدا یک توضیح کلی در مورد تئوری مورد استفاده ارائه می‌کنیم. برای هر واکسل، تعدادی رگرسور (regressor) (بر حسب نیاز و هدف مسئله) در نظر می‌گیریم، و یک مدل خطی تعمیم‌یافته (Generalized Linear Model – GLM) برآذش می‌کنیم و ضرایب  $\beta$  مربوط به هر رگرسور را مطابق رابطه  $y = \beta X + e$  تخمین می‌زنیم. شکل زیر، خلاصه‌ی این توضیحات است:



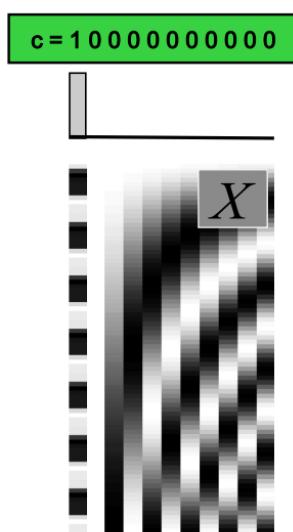
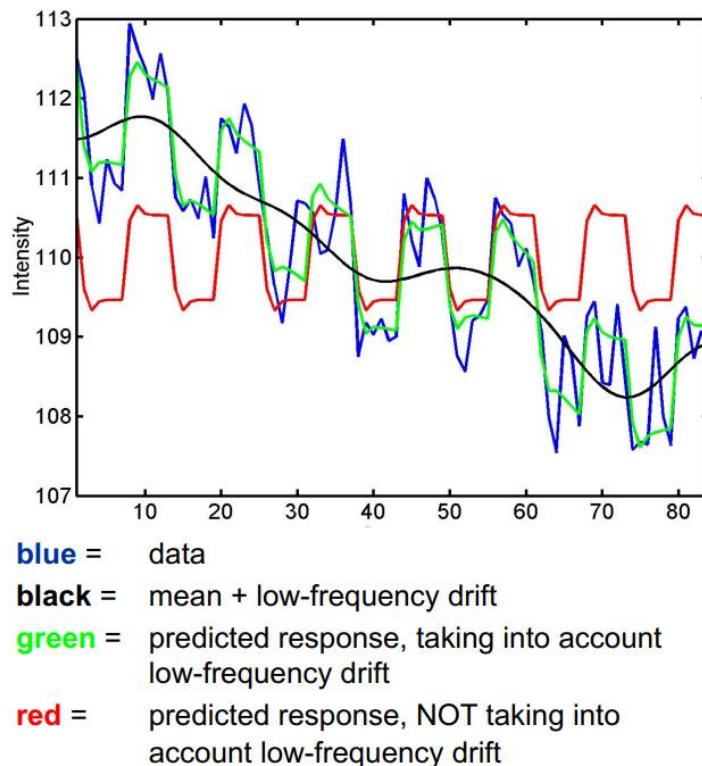
ماتریس  $X$  را ماتریس دیزاین (Design Matrix) می‌نامند که حاوی همه‌ی اطلاعاتی است که فکر می‌کنیم روی خروجی اثر دارند (که همان رگرسورها هستند).

برای ارائه‌ی تخمین مناسب، پاسخ کلی مغز را از یک مدل مبتنی بر کانولوشن (Convolutional Model) تولید می‌کنیم که در آن، پاسخ به هر تحريك را یک تابع مشخص، در نظر می‌گیریم (این تابع می‌تواند به دلخواه ما در نظر گرفته شود) که آن را HRF (Haemodynamic Response Function) می‌نامیم و پاسخ کلی مغز را از کانولوشن این تابع، با ضربه‌هایی در محل زمانی تحريك‌های مؤثر تولید می‌کنیم. شکل زیر، این فرایند را با HRF مورد استفاده در این پروژه (که مدل معروف و معتبری برای پاسخ مغز است) نشان می‌دهد:



expected BOLD response  
= input function  $\otimes$  impulse response function (HRF)

نهایتاً، برای آن که پاسخ مورد انتظار با پاسخ مشاهده شده منطبق باشد، اثر نویز را نیز لحاظ می‌کنیم. نمودارهای زیر، این فرایند را نشان می‌دهد که در آن، نمودار قمزرنگ نشان‌دهنده پاسخ ایده‌آل مغز و نمودار سبز رنگ، پاسخ تولیدشده با استفاده از اضافه کردن یک نویز فرکانس پایین (نمودار سیامرنگ) را نشان می‌دهد. این نمودار، تخمینی از پاسخ مشاهده شده، یعنی نمودار آبی رنگ است.



Q: activation during listening ?  
Null hypothesis:  $\beta_1 = 0$

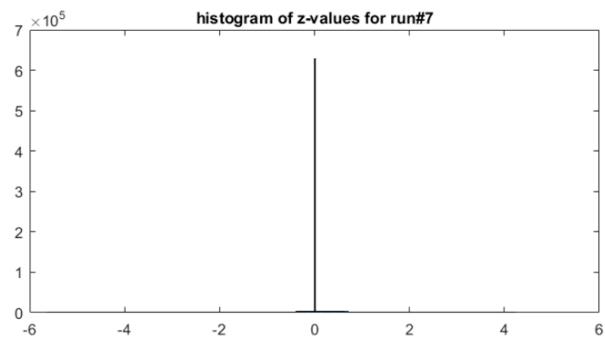
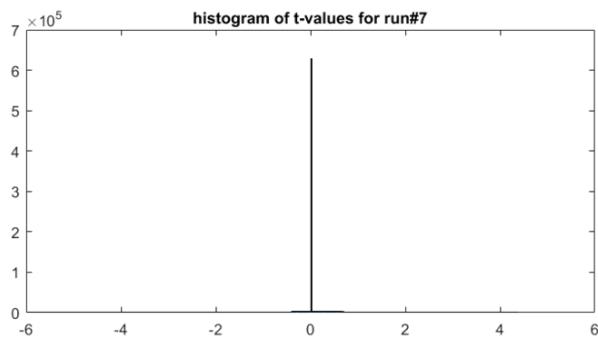
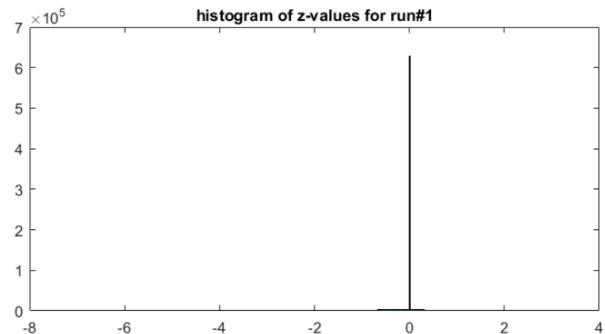
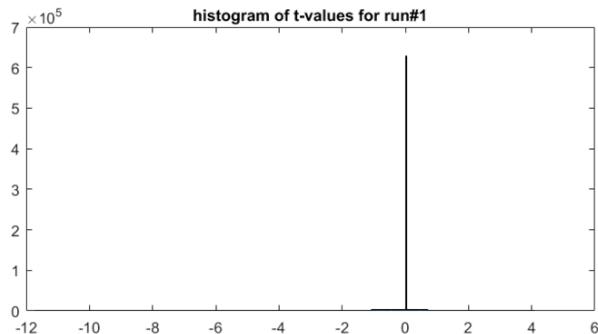
$$t = \frac{c^T \hat{\beta}}{Std(c^T \hat{\beta})}$$

آخرین نکته‌ی قابل ذکر آن است که کنتراست‌هایی که تعریف می‌کنیم، همان رگرسورها هستند که با توجه به هدف مورد نظر، می‌توانند متفاوت باشند. در نهایت، با تعریف کردن کنتراست مدل نظر و تشکیل ماتریس دیزاین،  $\beta$ ‌های مربوطه تخمین‌زده می‌شوند و با انجام یک آزمون فرضیه روی این  $\beta$ ‌ها، مقادیر t-value متناظر را به دست می‌آوریم. فرض صفر این آزمون فرضیه معادل صفر بودن  $\beta$ ، یعنی بی‌اثر بودن رگرسور مربوطه است. در نتیجه، برای هر واکسل و برای هر رگرسور، یک مقدار  $\beta$  و نیز یک مقدار t-value خواهیم داشت که در ادامه، برای مقایسه، می‌توانیم آن را تبدیل به z-value کنیم. شکل مقابل نیز نمایانگر رابطه کنتراست،  $\beta$ ، و آزمون t-test را ناشن می‌دهد.

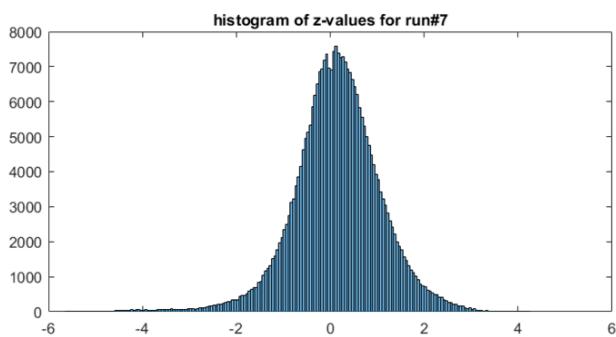
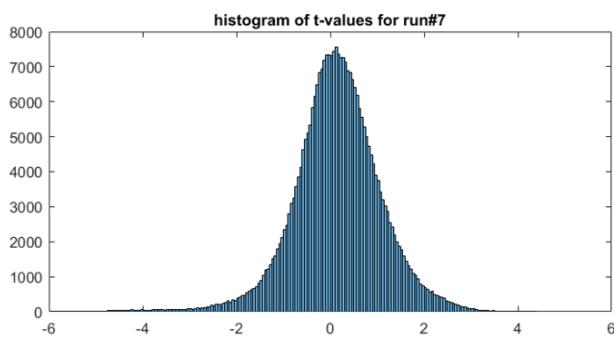
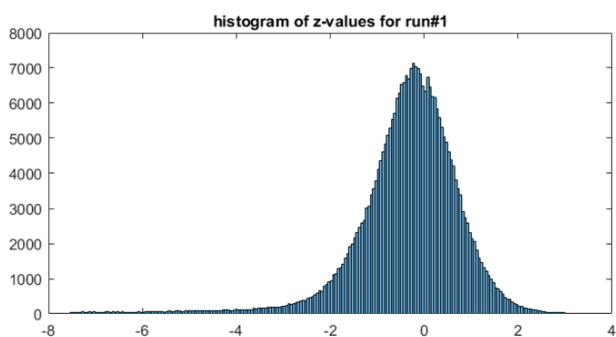
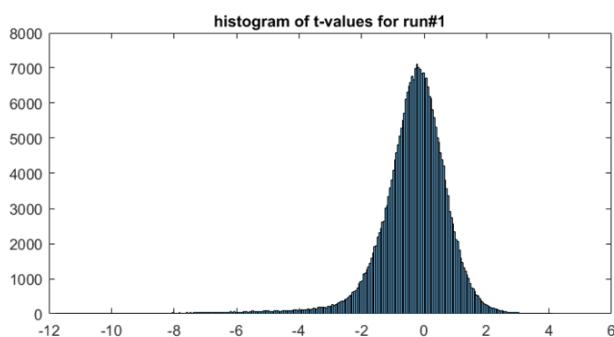
تا به این جا، توضیحات کلی مربوط به فرایند طی شده را ارائه کردیم که به همین میزان اکتفا می‌کنیم و در ادامه، به گزارش یافته‌ها می‌پردازیم. مرجع عکس‌های این توضیحات نیز از طریق لینک زیر قابل مشاهده است:

[www.fil.ion.ucl.ac.uk/spm/course/slides10-vancouver/02\\_General\\_Linear\\_Model.pdf](http://www.fil.ion.ucl.ac.uk/spm/course/slides10-vancouver/02_General_Linear_Model.pdf)

برای این قسمت، مطابق خواسته‌ی سؤال، مقادیر t-value را به (از طریق خروجی spm) و نیز مقادیر z-value را به کمکتابع spm\_t2z محاسبه کردیم. همه‌ی این‌ها، در استراکت data، که دو فیلد t و z دارد ذخیره شده است. با توجه به این که گزارش کردن مقادیر t‌ها و z‌ها در این متن، (به دلیل حجم بالای اعداد) هیچ ارزش عملیاتی ندارد، هیستوگرام مقادیر t و z را برای دو ران دلخواه (ران اول و ران آخر) رسم می‌کنیم.



مشاهده می‌شود که مقادیر z-value و t-value برابر به صفر، به قدری زیاد هستند که هیستوگرام‌های فوق را تقریباً بلااستفاده کرده است. بنابراین، مقادیر t-value و z-value برابر با صفر را در نظر نمی‌گیریم و هیستوگرام‌های فوق را مجدداً رسم می‌کنیم.



گزارش کار پژوهه درس علوم اعصاب محاسباتی سری چهارم  
دامنه و چگالی توزیع مقادیر  $t$ -value و  $z$ -value با انتظارات نسبی ما سازگار است و ضمناً در نظر داریم که مقادیر  $t$  و  $z$  برابر با صفر، درصد بسیار زیادی از دیتای موجود را تشکیل می‌دهد که می‌تواند طبیعی باشد، چراکه حجم قابل توجهی از عواملی که احتمال می‌دادیم در پاسخ معز مؤثر باشند، احتمالاً اثر چندانی ندارند. این مسئله را در بخش چهارم و در فرایند کاهش بُعد، به شکل دقیق‌تری مشاهده و بررسی خواهیم نمود.

**بخش ۲-۳**

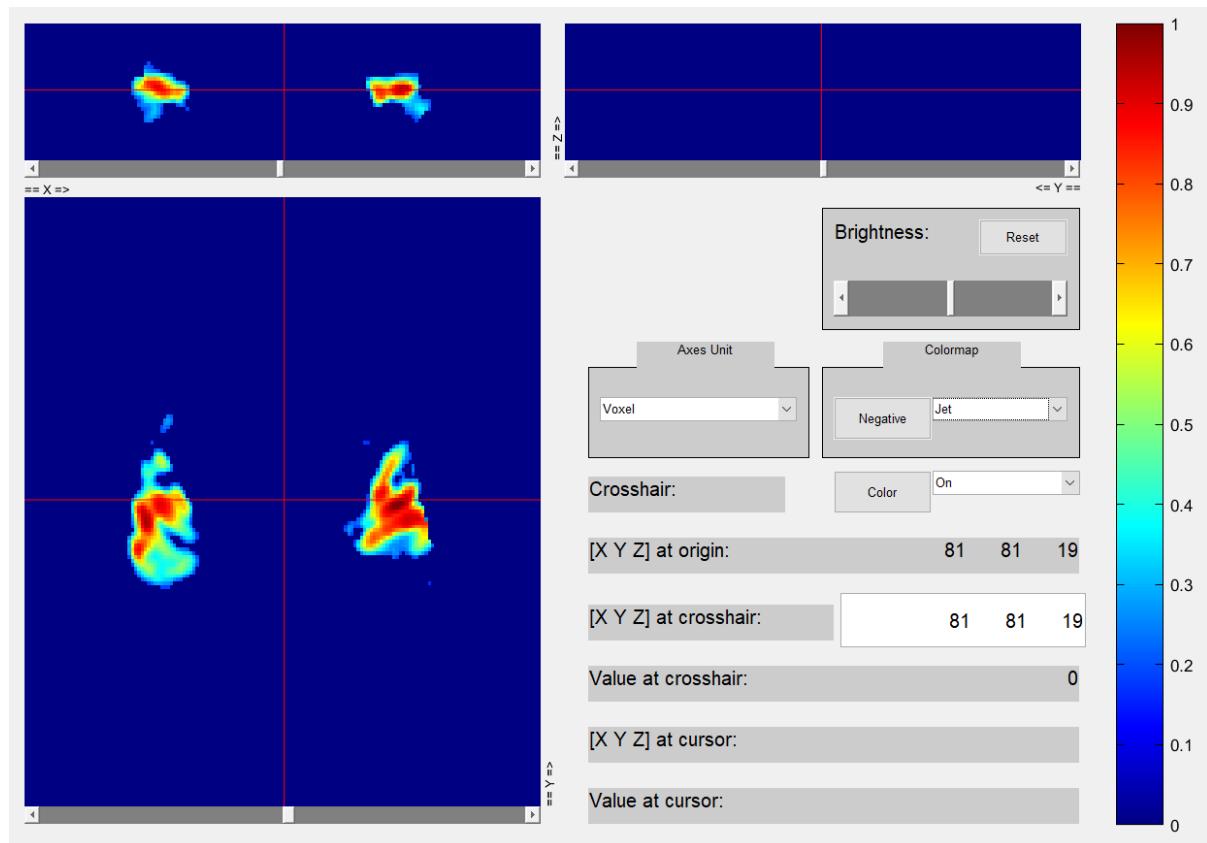
برای آن که بتوانیم تصویر مطلوبی رسم کنیم، تابعی به نام brainPlot که ماتریس z-value ورودی را با کمک توابع make\_nii و savi\_nii تبدیل به فایل nii می‌کند، سپس با استفاده از تابع convertnii2mat، مجدد آن را به ماتریس تبدیل می‌کند. این کار (که در عمل بیهوده به نظر می‌رسد) باعث می‌شود که در اثر استفاده از تابع convertnii2mat، نموداری از نقاط فعال (بر حسب مقادیر z-value) رسم شود.

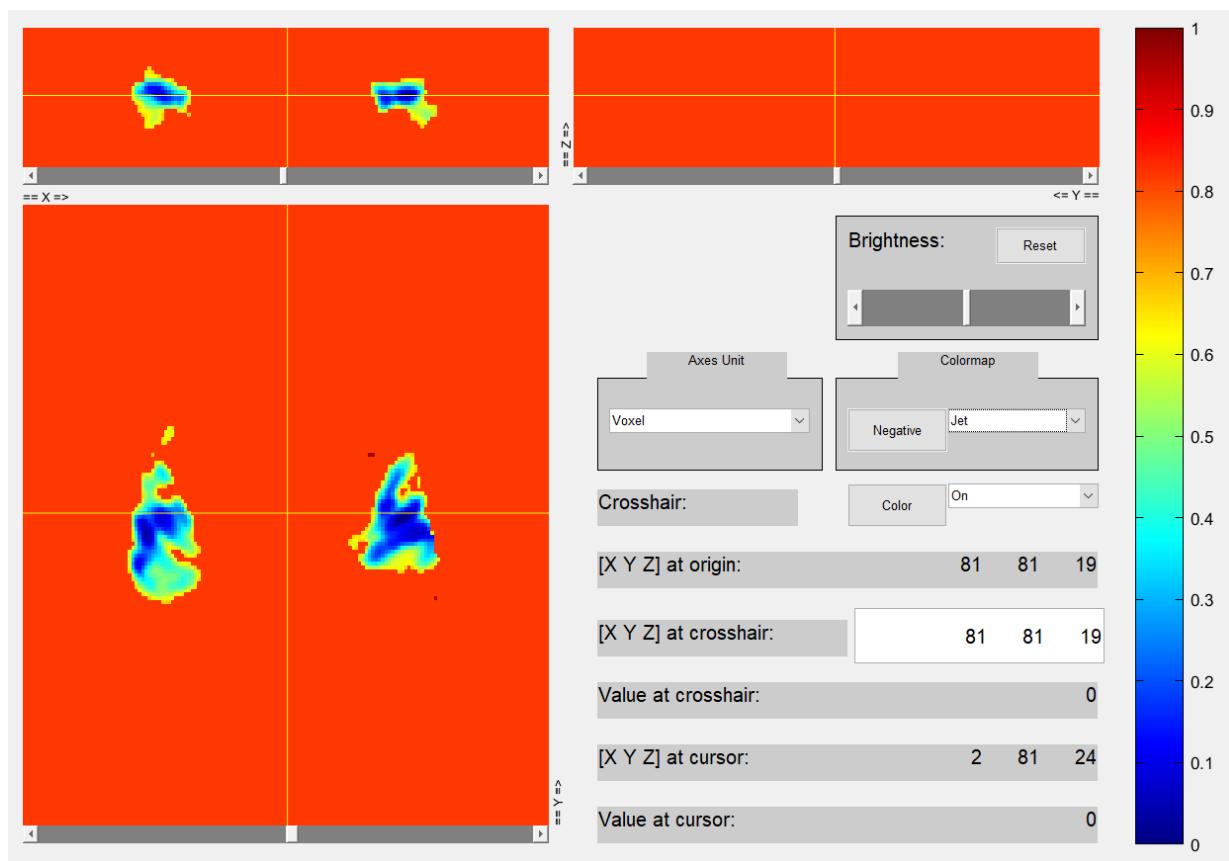
تعریف کتراست مناسب: در این بخش، با توجه به این که نقاط فعال در اثر هر ژانر را مورد بررسی قرار می‌دهیم، کتراست‌های مشابه با قسمت قبلی (یعنی در هر کتراست، برای یکی از ژانرهای مقدار ۱ و برای مابقی مقدار صفر) تعریف کردیم. همچنین، برای ارائه‌ی یک نتیجه‌ی واحد برای هر ژانر، روی مقادیر z-value مربوط به ران‌های مختلف میانگین گرفتیم.

**محاسبه حد آستانه (threshold):** برای آن که مقدار آستانه برای همه‌ی ژانرهای یکسان باشد، انحراف معیار تمامی z-value‌ها را محاسبه کردیم و حد آستانه را برابر این مقدار در نظر گرفتیم. (یعنی z-value‌هایی که فاصله‌ای کمتر از  $k\sigma$  از میانگین ژانر خود دارند را صفر کردیم) نتایج این بخش به ازای  $k = 3$  گزارش می‌شوند.

در رویکرد در ترسیم نقاط فعال، هم می‌توانیم مقادیر z-value را مستقیماً به عنوان ورودی تابع brainPlot (در ابتدای این بخش، در مورد این تابع توضیح داده شد) بدهیم و هم می‌توانیم قدر مطلق این مقادیر را به عنوان ورودی در نظر بگیریم. در واقع مقادیر بسیار مثبت و نیز بسیار منفی z-value‌ها، هر دو به عنوان نقاط اثرگذار و فعال قابل در نظر گرفتن هستند، چرا که اندازه‌ی  $\beta$  متناظر با آن‌ها بزرگ است. حال، چنان‌چه قدر مطلق مقادیر z-value را رسم کنیم، تمایزی بین نقاطی که بسیار منفی و بسیار مثبت دارند، قائل نمی‌شویم، در حالی که اگر مقادیر واقعی را رسم کنیم، این تمایز قابل مشاهده است. به عنوان نمونه، نقاط فعال ژانر rock را با هر دو روش در ادامه مشاهده می‌کنید:

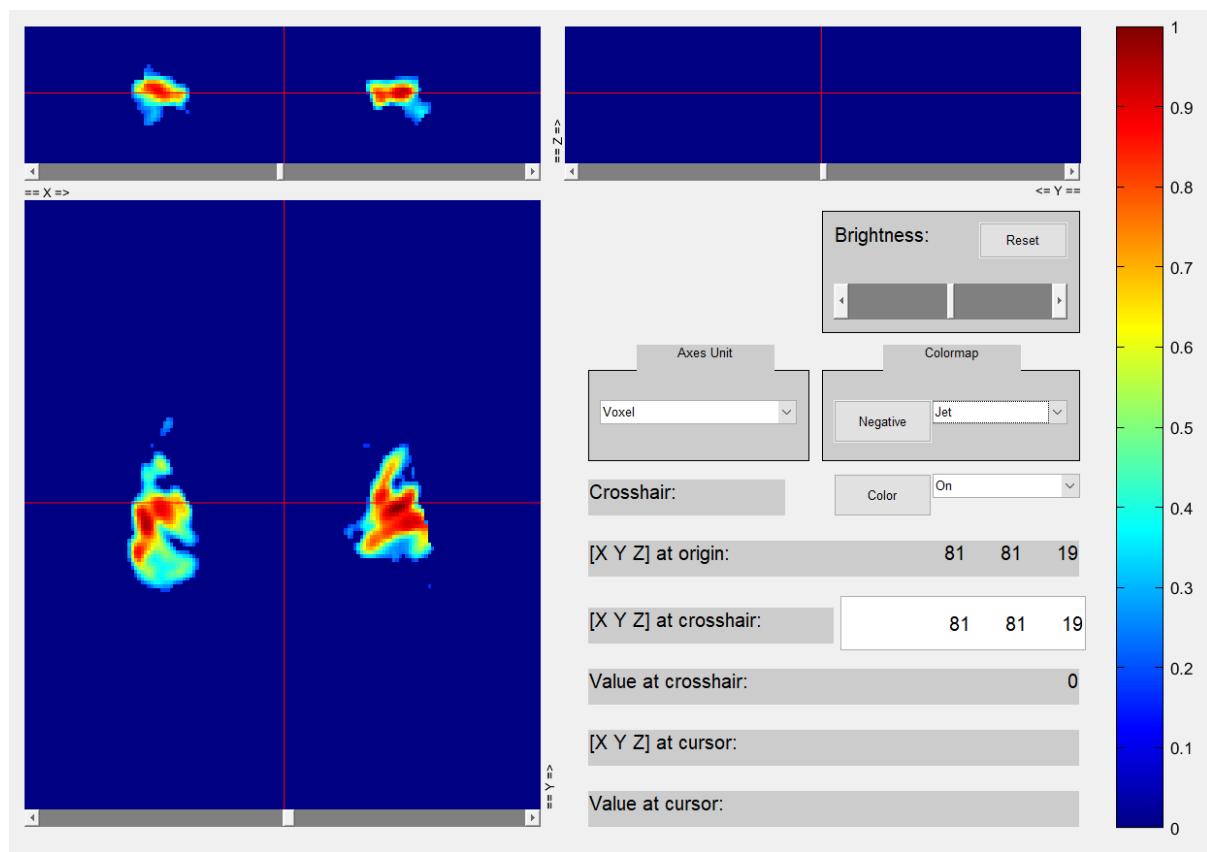
حالت اول: رسم قدر مطلق z-value‌ها



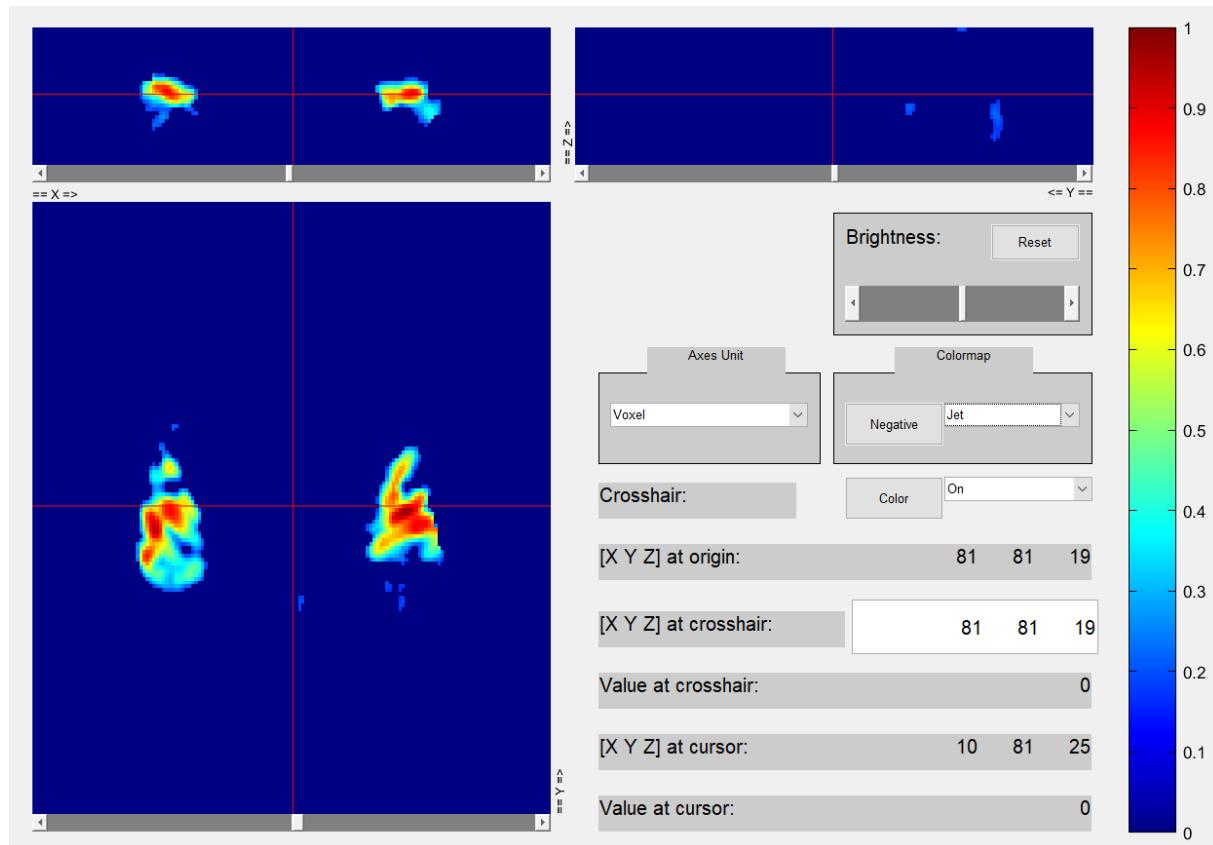


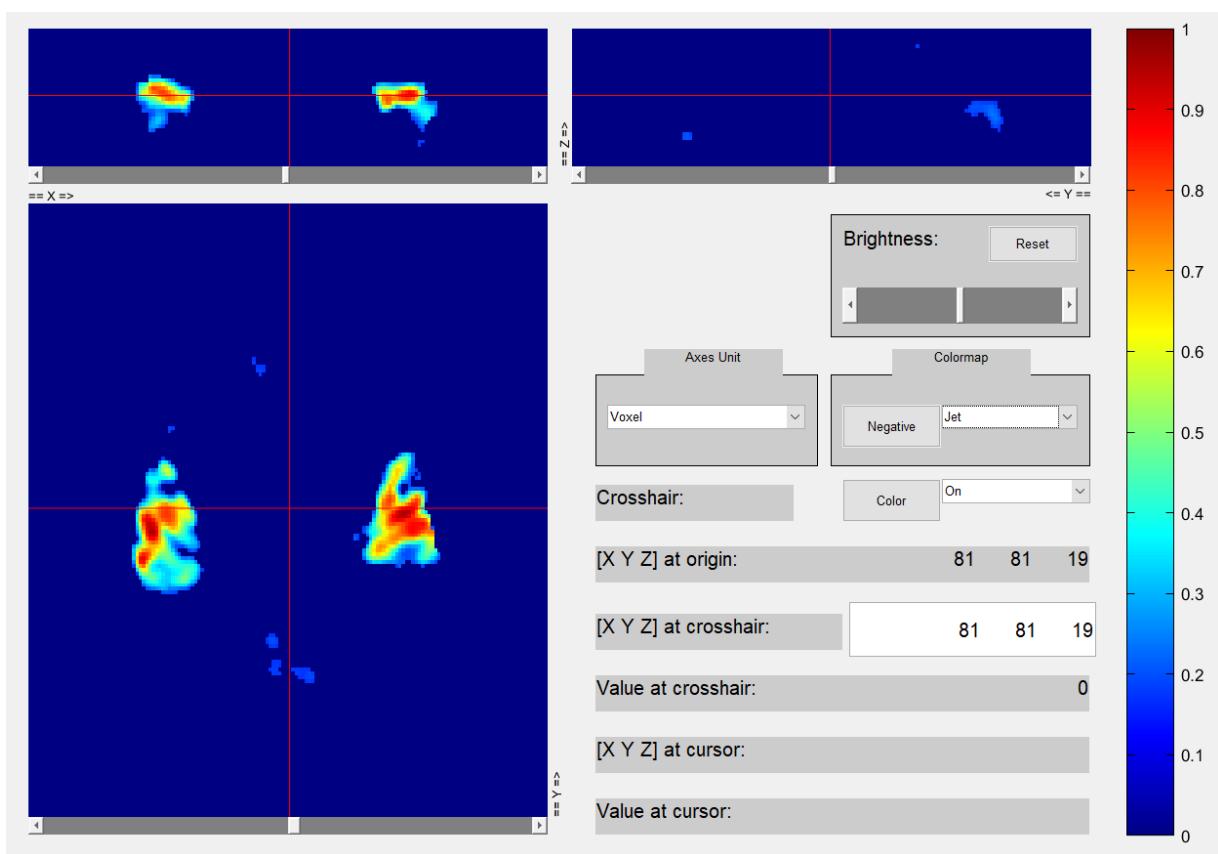
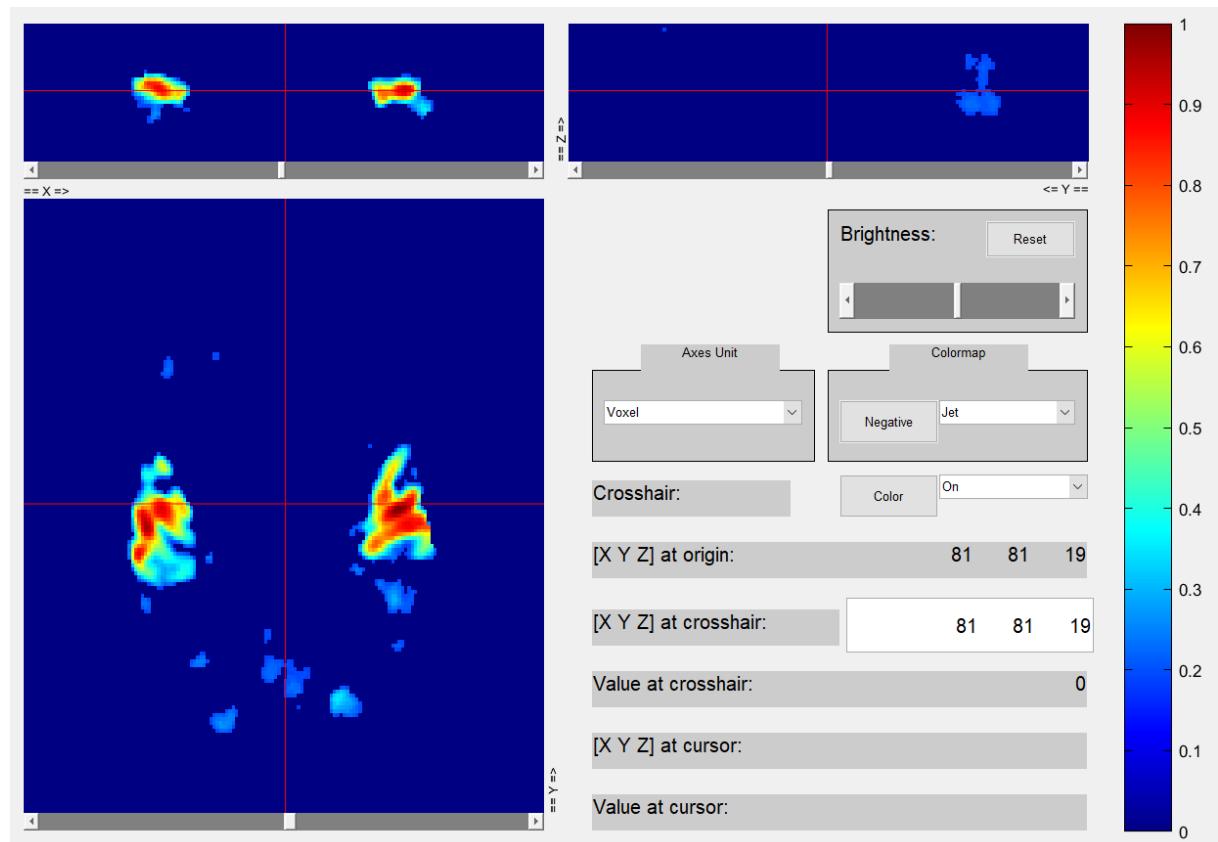
مشاهده می شود که مقادیر بسیار کوچک و بسیار بزرگ نمودار اول، همگی در نمودار اول به عنوان مقادیر بزرگ در نظر گرفته می شوند، و حد وسط نمودار دوم، در نمودار اول به متزله‌ی صفر است. همچنین شایان ذکر است که مقادیر z-value‌ها در رسم نمودارهای فوق، به بازه‌ی صفر تا یک مب شده‌اند.

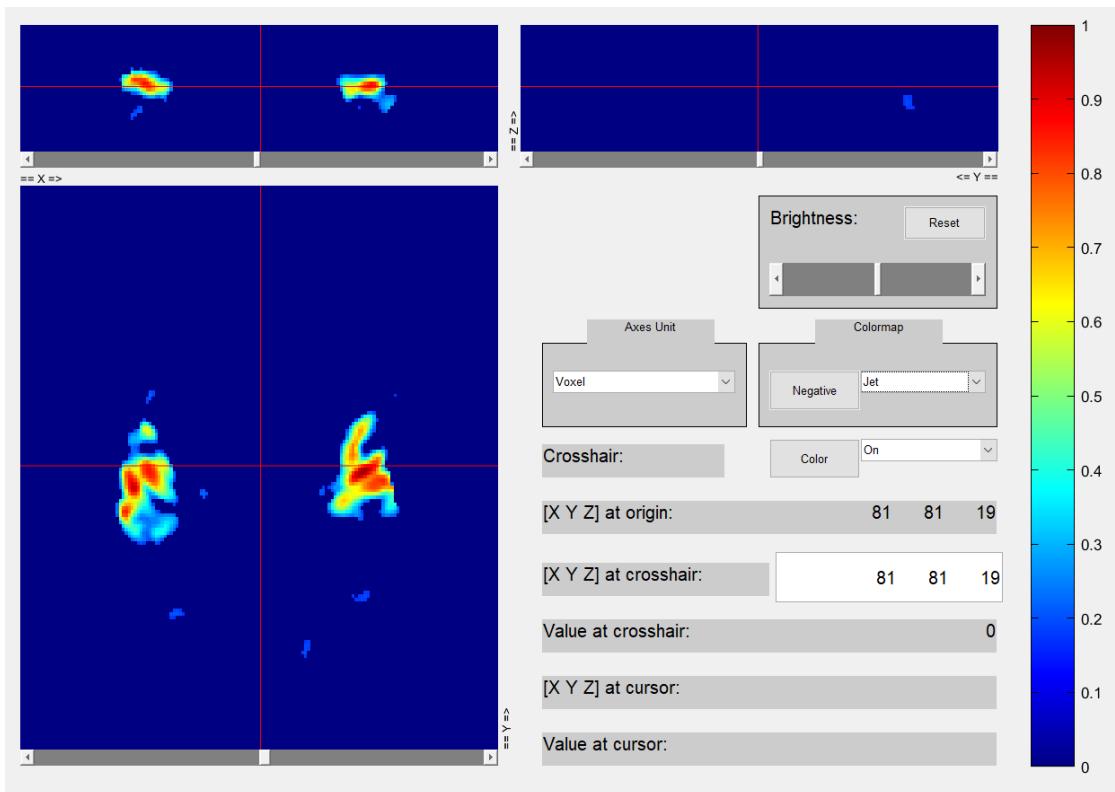
از این پس، نوع نمودار اول (رسم قدر مطلق‌ها) را انتخاب می کنیم، چراکه اگرچه بخشی از اطلاعات از دست می‌رود، اما هدف اصلی ما، یعنی تعیین بخش‌های فعال را به شکل بهتری دنبال می‌کند و تمامی نقاط فعال را، فارغ از نوع فعالیتشان، با رنگ‌های زرد و قرمز (مطابق طیف رنگی‌ای که در نمودار اول قابل مشاهده است) مشخص می‌کند. در ادامه (از صفحه بعد)، پنج نمودار مربوط به نقاط فعال مغز را (متناظر با پنج ژانر مورد بررسی) مشاهده می‌کنید.



نقاط فعال مربوط به ژانر: symphonic

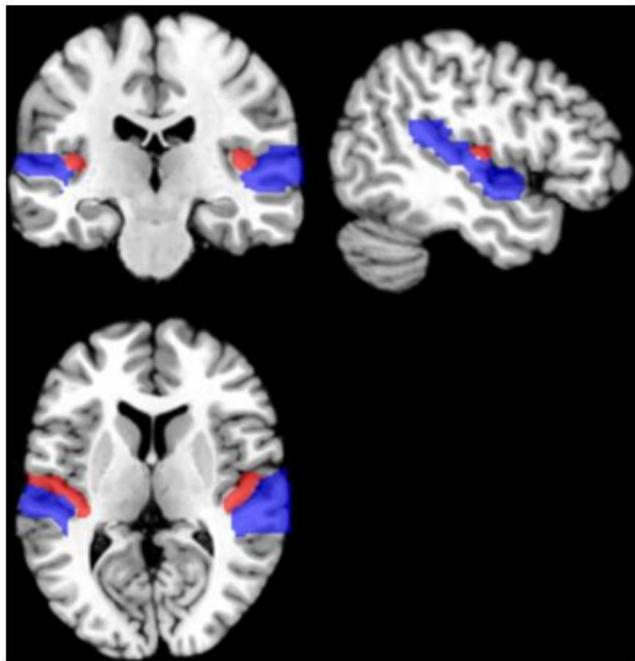


نقاط فعال مربوط به ژانر *country*:



آن چه که مشاهده می‌شود، آن است که برای تمامی ژانرهای تقریباً نقاط فعالیت ثابتی از مغز فعالیت بالایی دارند. این نقاط، در دو طرف مغز (چپ و راست) واقع شده‌اند. این مشاهده، این احتمال را در ذهن ما پیدید می‌آورد که ژانرهای تفاوت چندانی در فعالیت نقاط مختلف مغز ایجاد نمی‌کنند، یا به عبارت دیگر، فعالیت نقاط مختلف مغز مستقل از ژانر موسیقی پخش شده است.

یک نکته در اینجا توجه ما را جلب می‌کند، و آن این است که احتمالاً نقاطی که در اثر پخش این موسیقی‌ها فعال شده‌اند، به طور کلی مربوط به فعالیت‌های شنیداری مغز انسان باشند. بررسی امر کار آسانی است و می‌توان مطابق شکل زیر، این فرض را تأیید نمود:



این شکل که نقاط فعال ناحیه شنیداری (Auditory Cortex) را نشان می‌دهد، علت نتایج نه چندان مطلوب این بخش را روشن می‌کند. به دلیل آن که صرف شنیدن صدا، باعث فعالیت قسمت‌های شنیداری مغز می‌شود، تفاوت‌های دیگری که در اثر تفاوت ژانر موسیقی ممکن است وجود داشته باشد (هنوز از وجود چنین تفاوتی مطمئن نیستیم)، به وضوح دیده نخواهند شد، چرا که این ناحیه شنیداری مشاهده شوند. همین مسئله، ایده‌ی تولید کنترast‌های تفاضلی را مطرح می‌کند که در بخش بعد، مورد بررسی قرار خواهد گرفت.

مرجع عکس:

The role of the primary auditory cortex in the neural mechanism of auditory verbal hallucinations

April 2013 · Frontiers in Human Neuroscience 7:144

DOI: 10.3389/fnhum.2013.00144

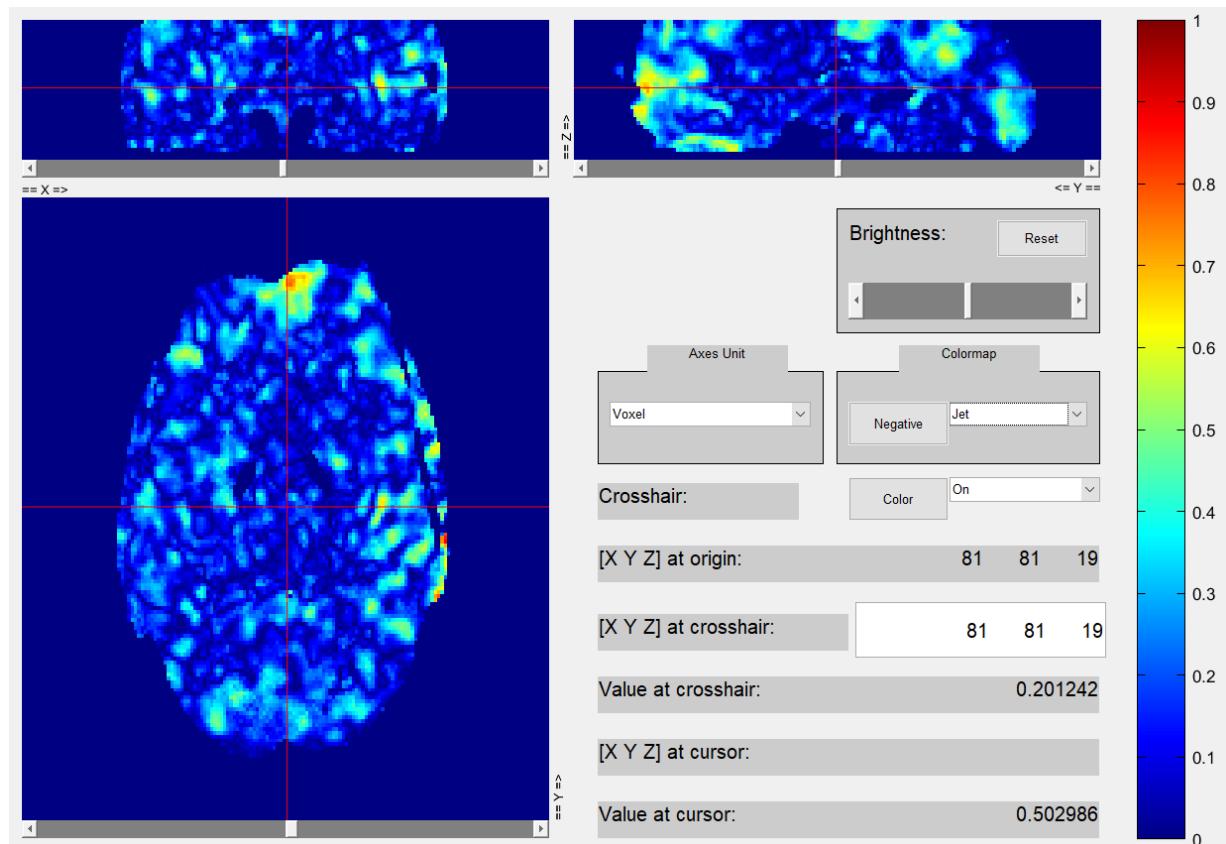
**بخش ۳-۳**

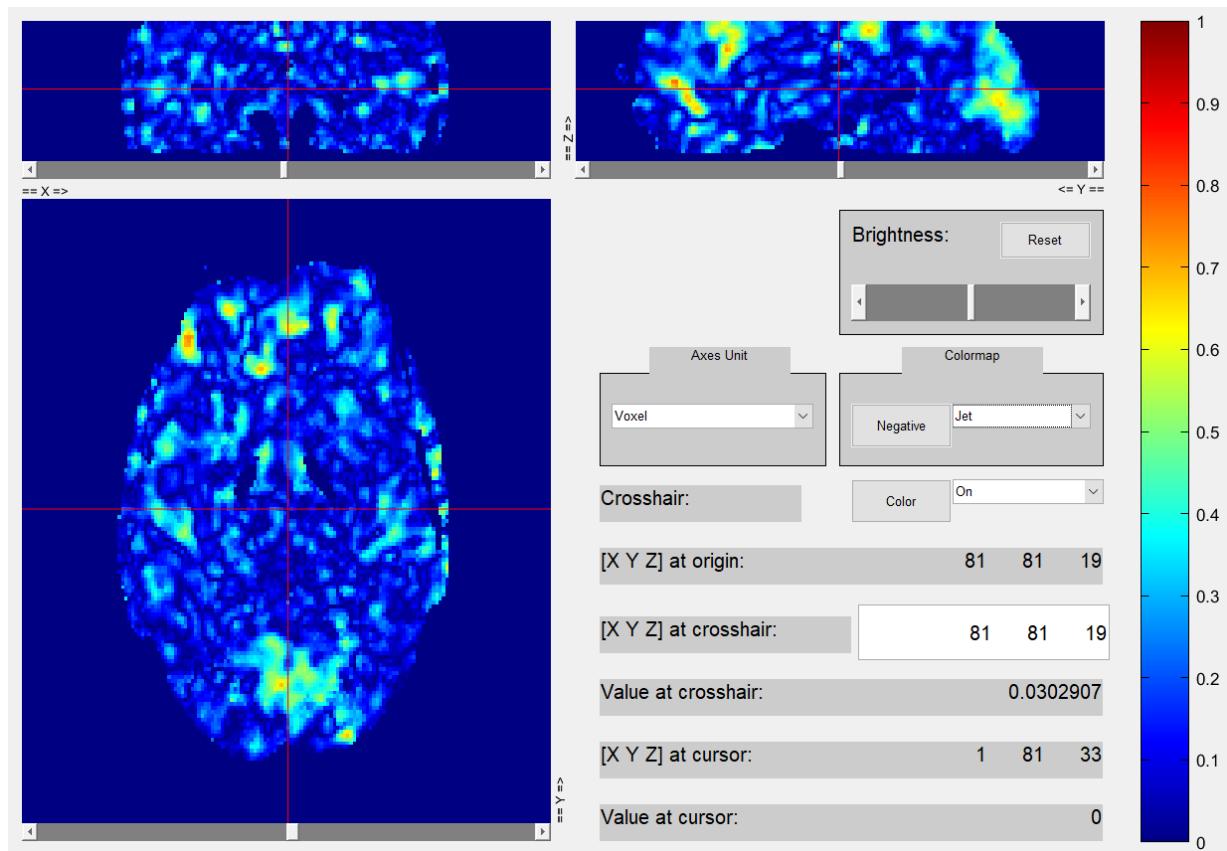
با توجه به مقدماتی که در بخش قبل بیان شد، برای آن که اثر ناحیه شنیداری که در تمامی ژانرها مشترک است را از بین ببریم، به سراغ تفاوت نقاط فعال در اثر پخش دو ژانر موسیقی مختلف می‌رویم. برای این کار، کنتراست‌ها را به صورت تفاضلی تعریف می‌کنیم. (به عنوان نمونه، یک بردار کنتراست تفاضلی مطلوب می‌تواند به شکل [1 0 0 0 -1] باشد)

در ادامه، نمودار نقاط فعال در تفاضل فعالیت دو ژانر را مشاهده می‌کنید.

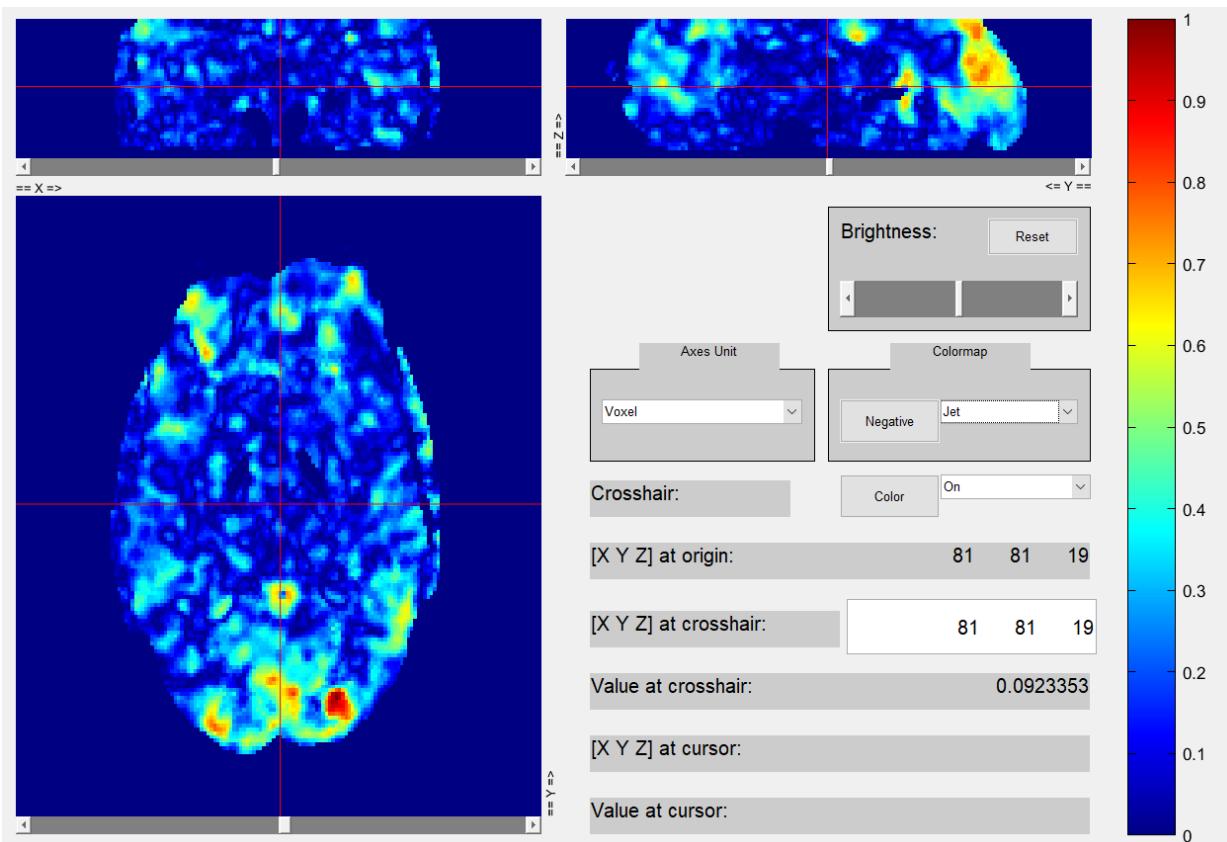
همچنین مجدداً مشابه بخش قبل، از قدر مطلق مقادیر z-value برای رسم نمودارهای نقاط فعال استفاده شده است. (توضیحات کامل در مورد روش‌های رسم نقاط فعال در بخش ۲-۳ داده شده است)

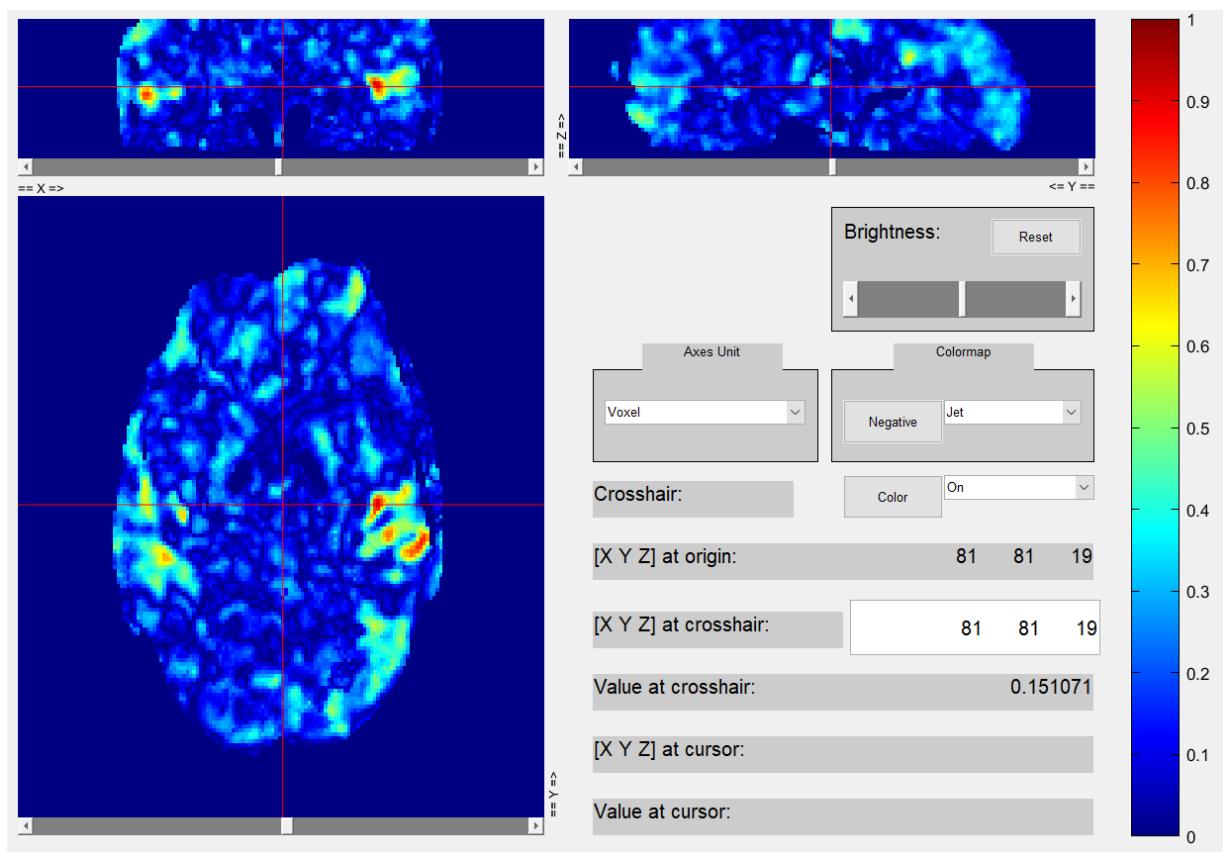
تفاوت نقاط فعال دو ژانر *symphonic* و *rock*



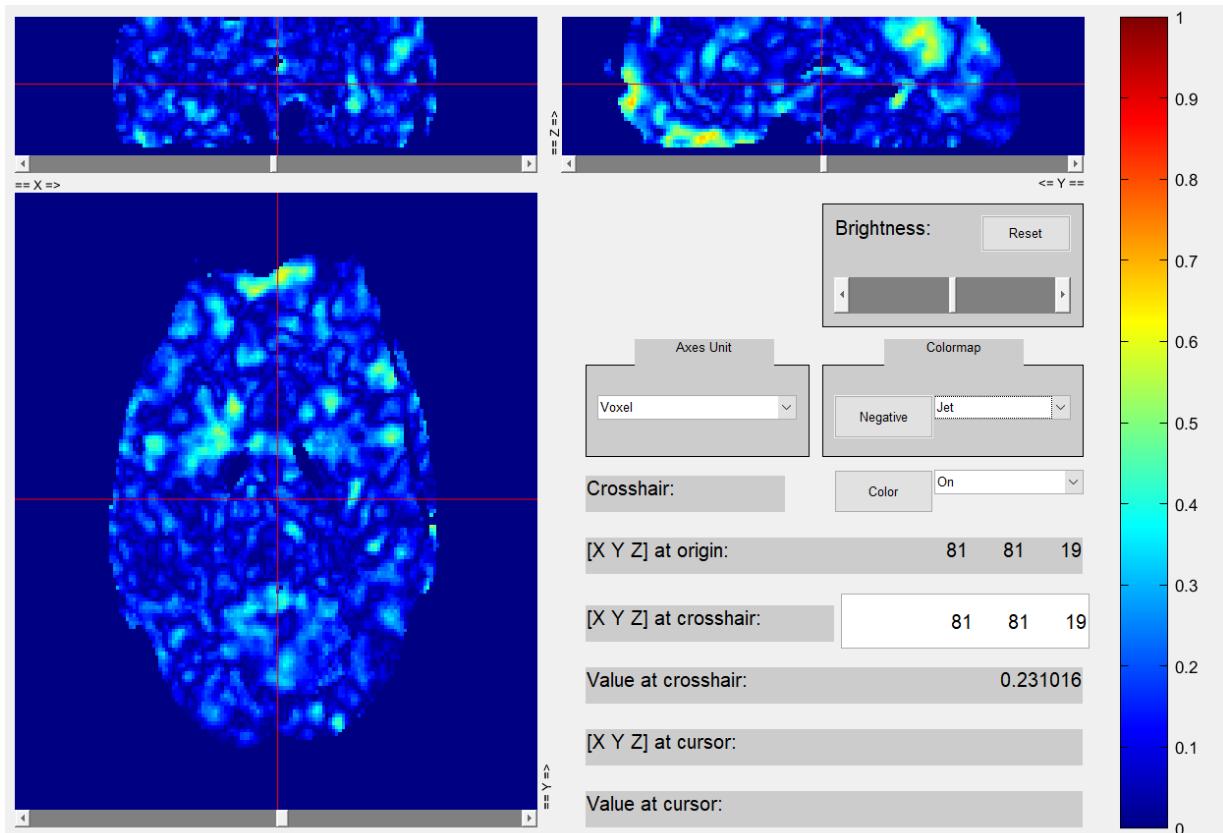


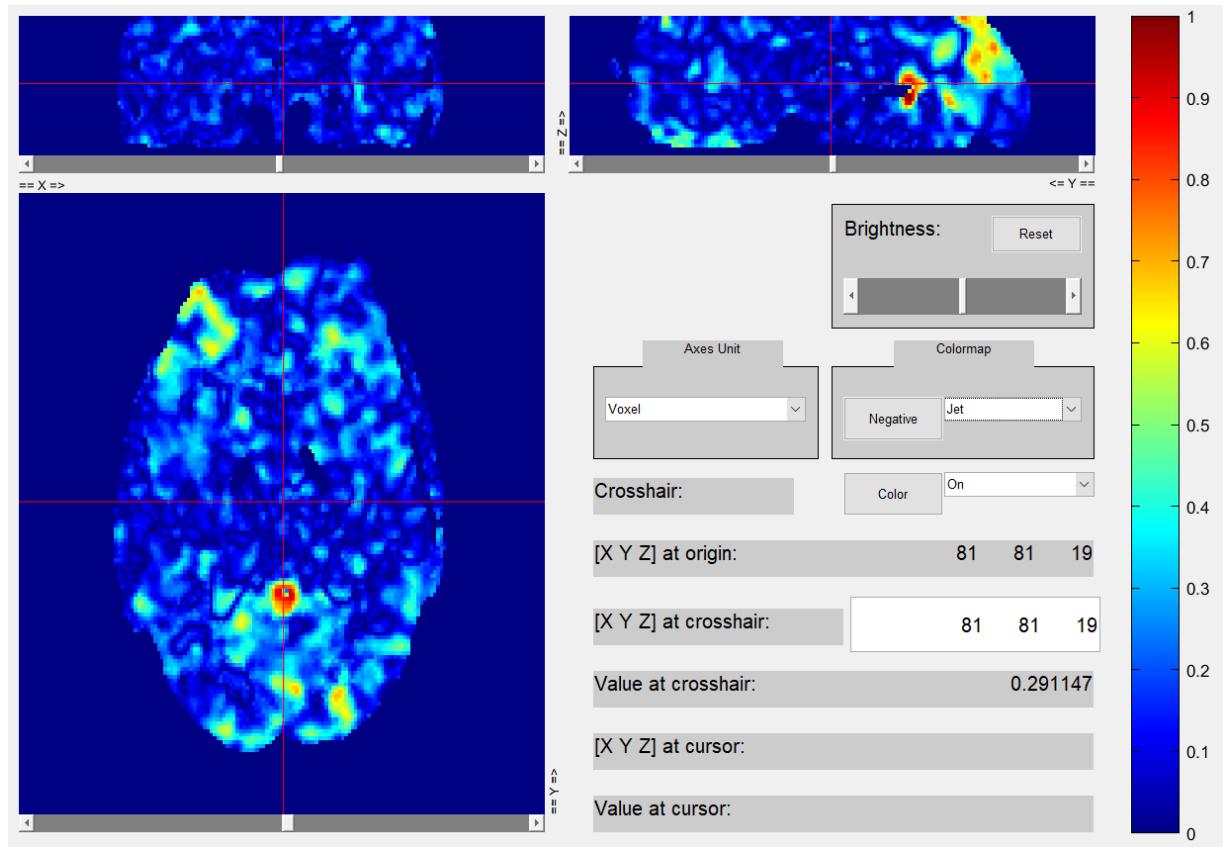
تفاوت نقاط فعال دو ژانر rock و country



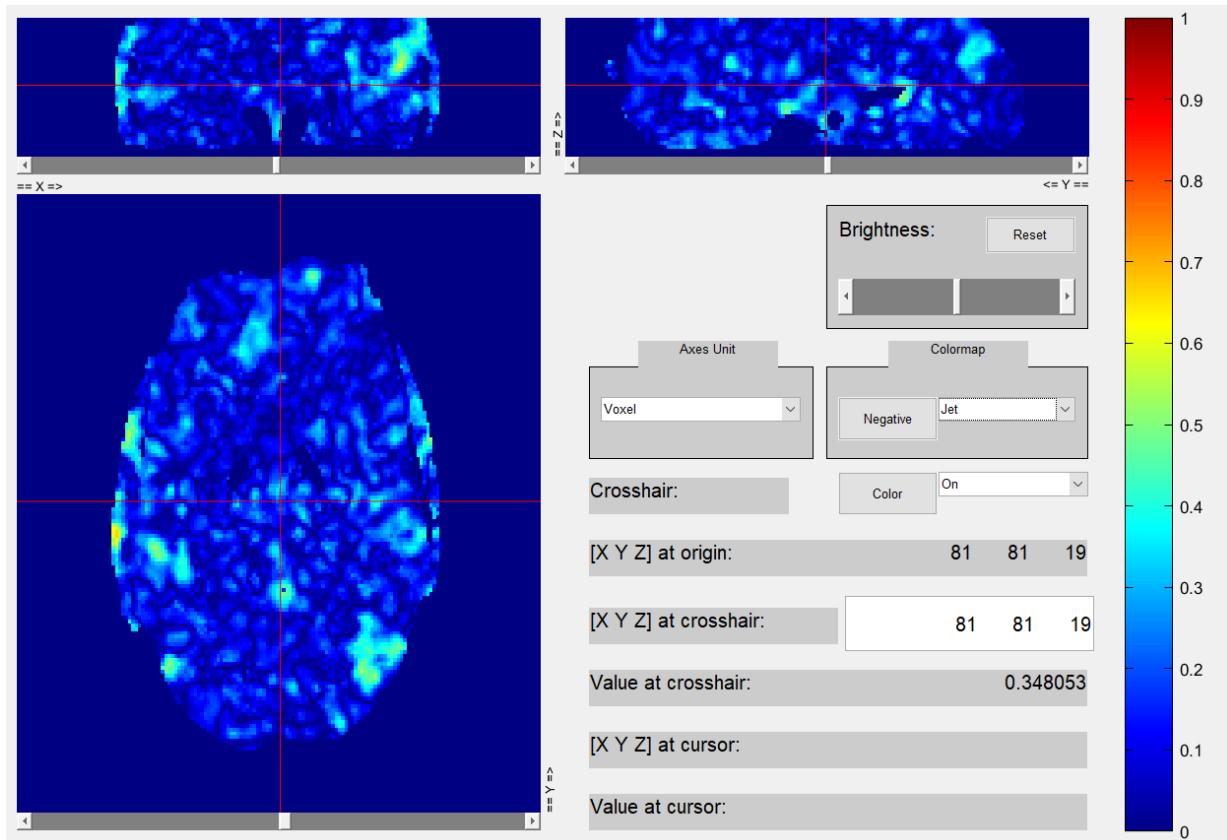


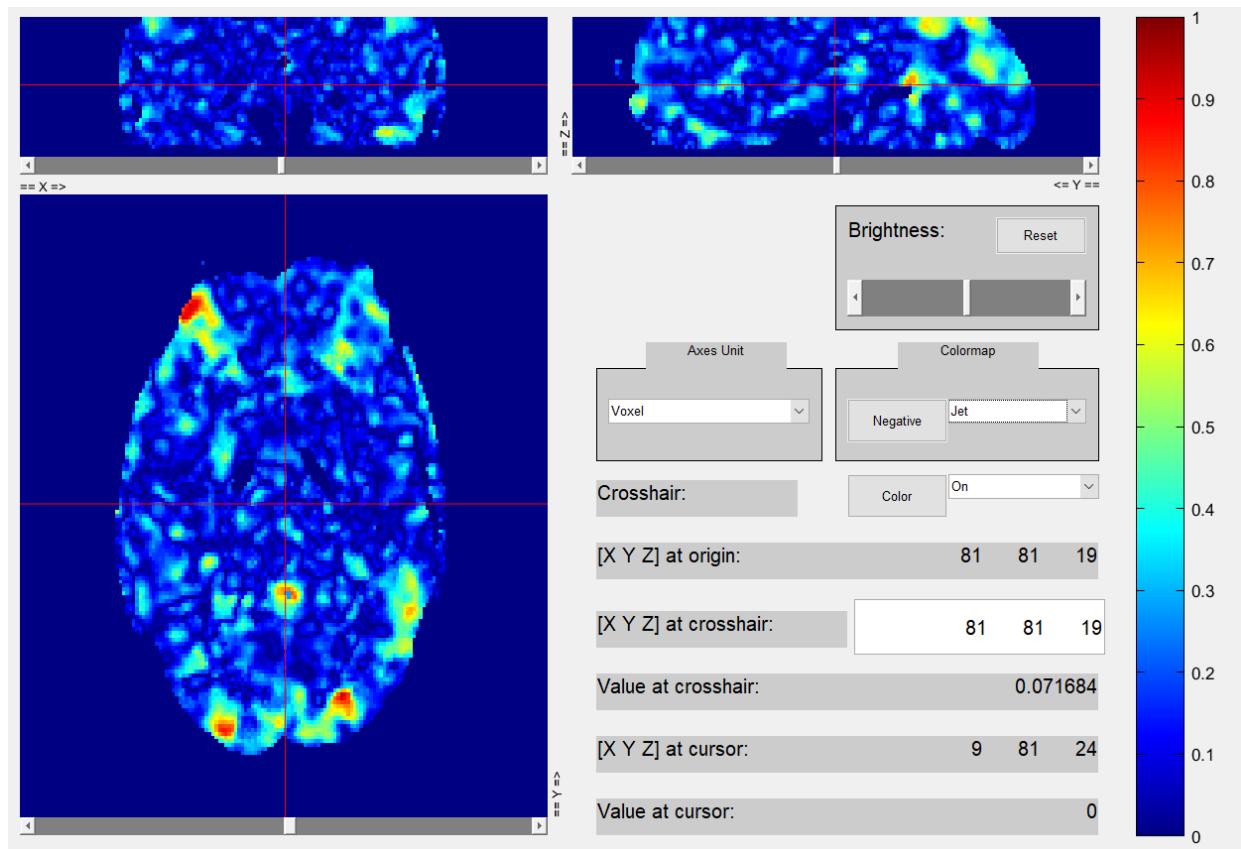
تفاوت نقاط فعال دو ژانر metal و symphonic



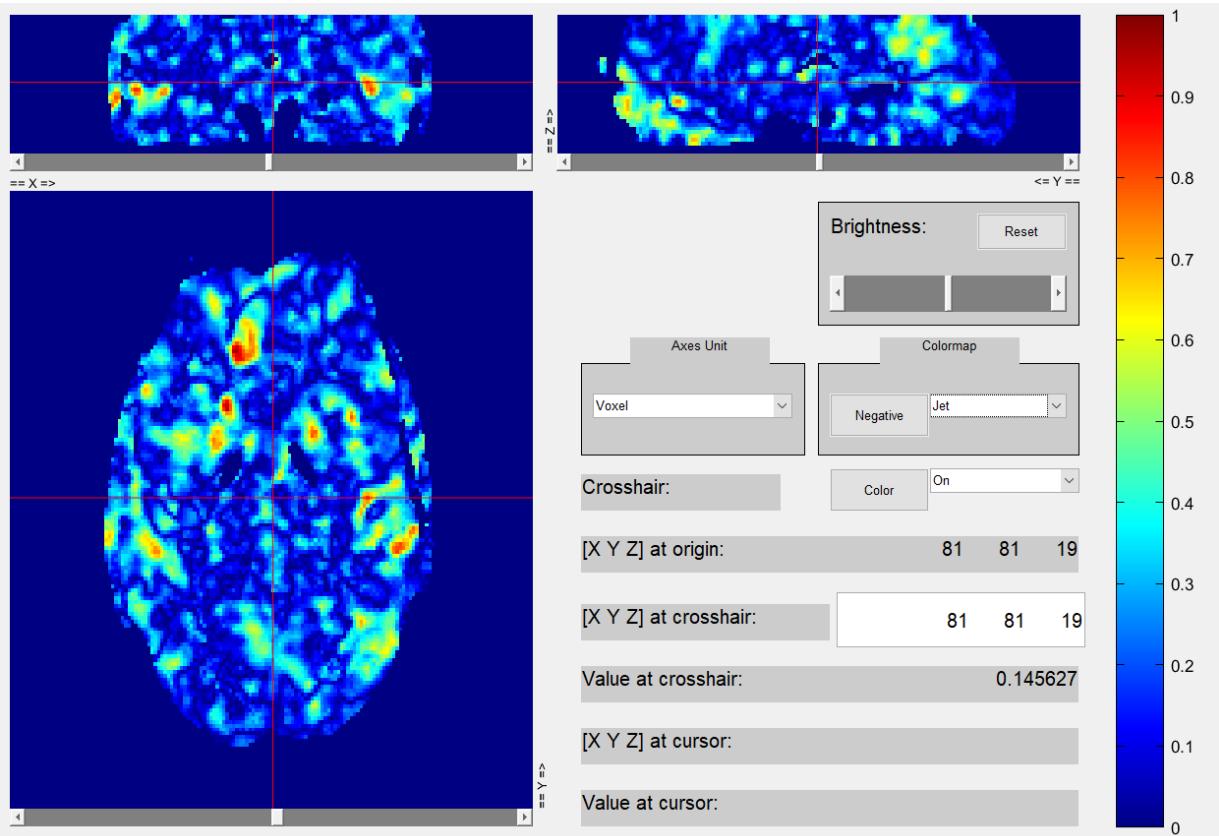


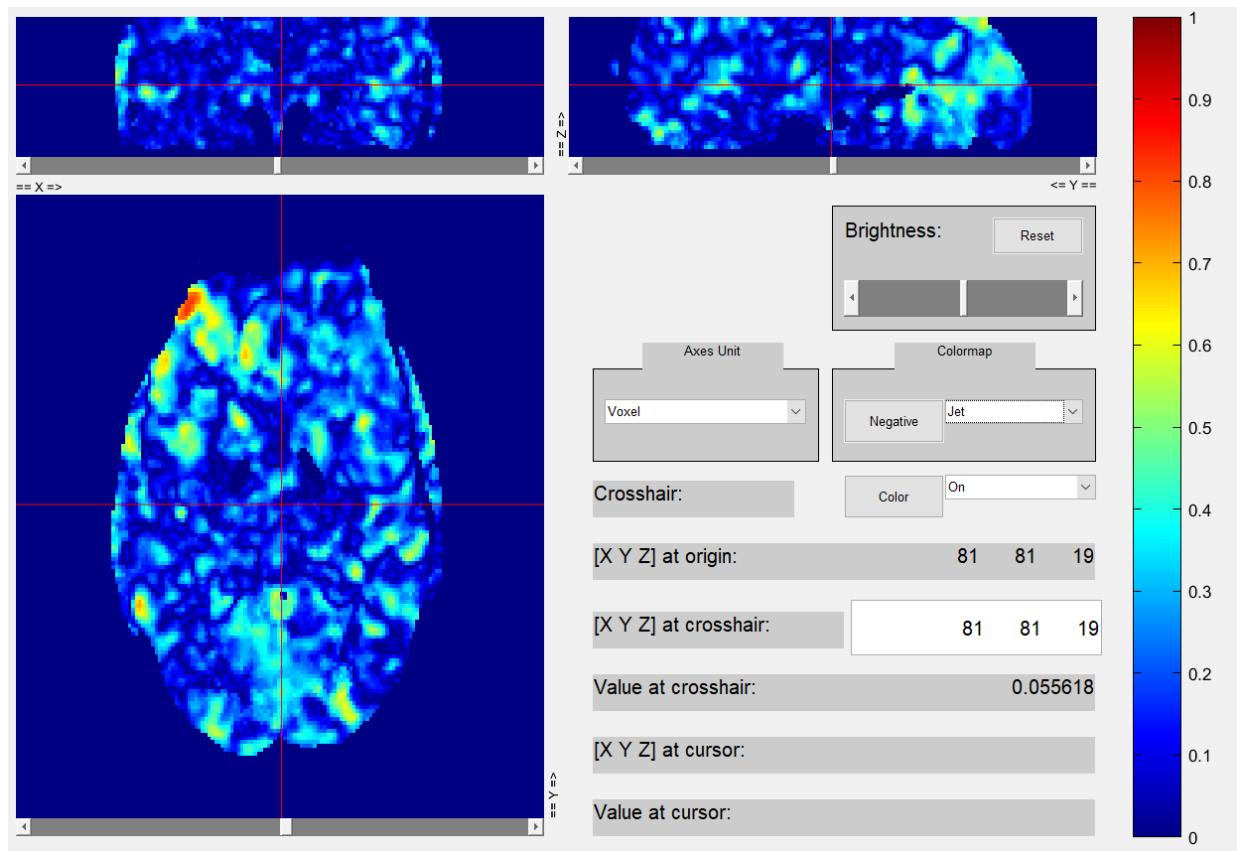
تفاوت نقاط فعال دو ژانر symphonic و ambient





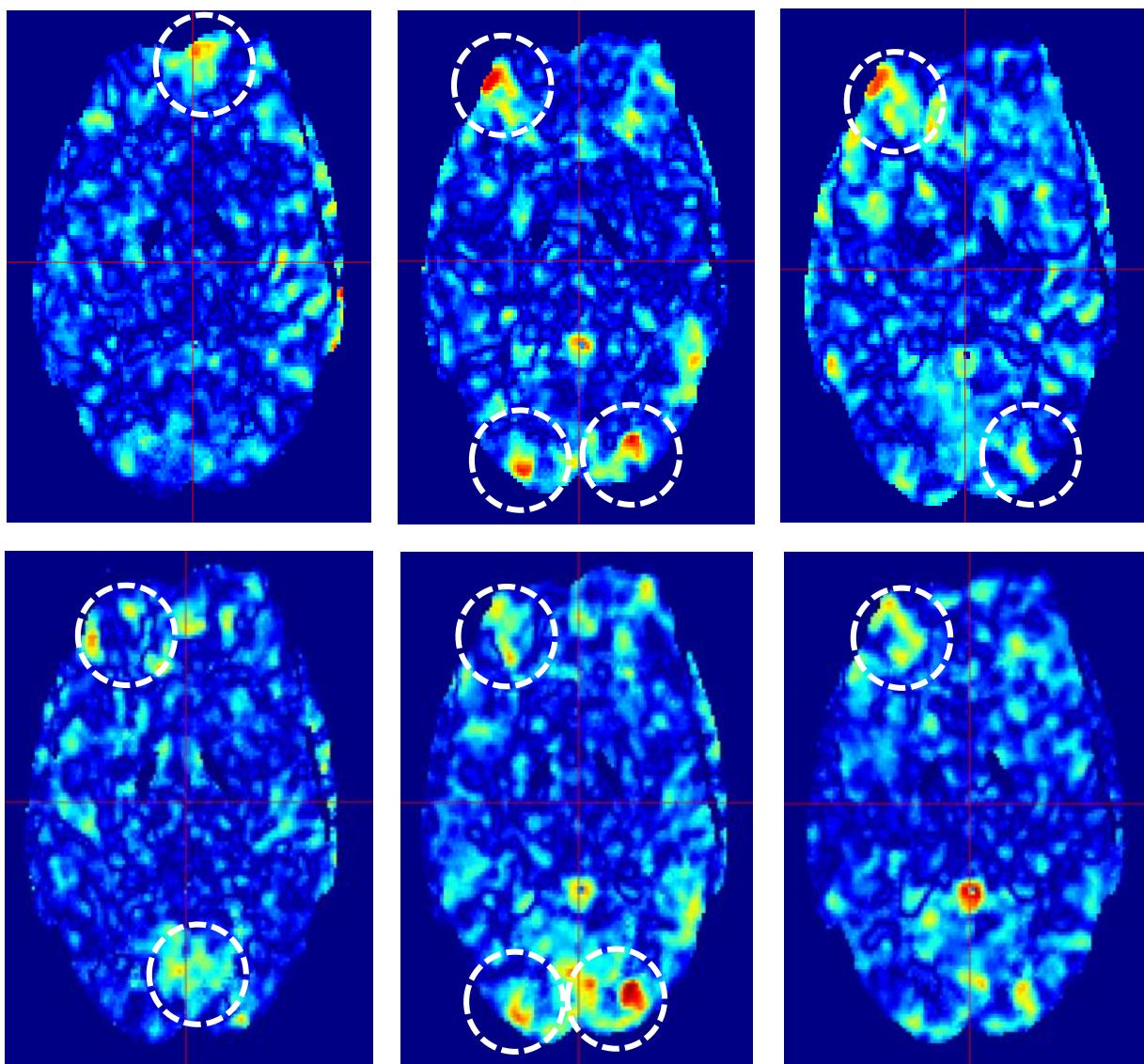
تفاوت نقاط فعال دو ژانر country و metal





در بررسی نتایج این بخش، اولًاً یک مشاهده مهم آن است که تقریباً در هیچ کدام از شکل‌ها، اثری از auditory cortex دیده نمی‌شود، و این بدان معناست که این ناحیه، صرفاً در اثر شنیدار تحریک می‌شود (یا حداقل، فعالیت آن ارتباطی با ژانر موسیقی ندارد) و در اثر پخش شدن ژانرهای مختلف فعالیت یکسان و مشابهی دارد. (تأیید نتایج بخش قبل)

اما در مورد قسمت‌های تمایزدهنده ژانرهای مختلف، به نظر می‌آید که دو قسمت جلو و عقب مغز (بر خلاف بخش قبل که دو قسمت راست و چپ مغز دارای فعالیت بودند) اثرگذاری بیشتری دارند، به خصوص نقاط معینی که در چند نمونه از شکل‌های زیر (که همان شکل‌های فوق هستند و به صورت گزینشی انتخاب شده‌اند) مشخص شده‌اند. (نقاط مدنظر با دایره‌های سفید مشخص شده‌اند)



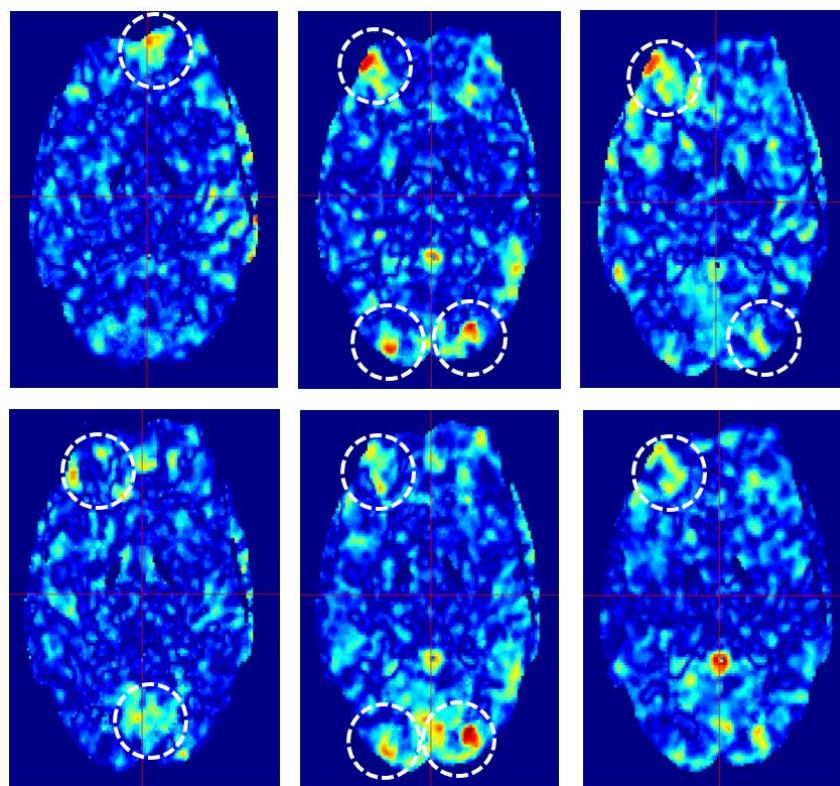
می‌بینیم که شاید بتوان تفاوت ژانرهای مورد نظر را به نقاطی در جلو و عقب مغز نسبت داد، اما نمی‌توان ادعایی قطعی و دقیق در این مورد ارائه داد، چراکه فعال شدن این بخش‌ها شرط نه لازم و نه کافی برای تمایز دو ژانر است (زیرا با بررسی ۱۰ نمونه از کنترast‌های تفاضلی، می‌توان نمونه‌هایی پیدا کرد که این نقاط فعال نیستند، در حالی که ژانر متفاوت است، و همچنان می‌توان نقاط فعال دیگری را در اثر پخش دو ژانر متفاوت مشاهده نمود) و صرفاً می‌توان از این نتیجه گیری به عنوان یک حدس معقول که طبقه‌بندی بر مبنای آن، احتمالاً نتیجه‌ای بهتر از حالت طبقه‌بندی تصادفی دارد، یاد کرد.

**بخش ۴-۳**

توضیحات این بخش، صرفاً یک خلاصه برداری و جمع‌بندی توضیحات دو بخش قبلی است.

آن چه که به طور روشن و آشکار مشاهده شد، فعالیت قابل توجه ناحیه auditory cortex واقع در دو طرف راست و چپ مغز (نرديک به گوش‌ها) در هنگام پخش شدن موسیقی بود. همان طور که احتمالاً انتظار می‌رفت، فعالیت اين نواحی، مستقل از ژانر موسیقی و صرفاً وابسته به تحريك صوتی و به وقوع پيوستان فرایند شنیدن است. اين باعث می‌شود که نقاط فعال مغز، به شکل قابل توجهی به اين موضوع اختصاص پيدا کنند، و چنان‌چه تفاوتی بين نقاط فعال مغز در اثر پخش ژانرهای مختلف موسیقی وجود دارد، به سادگی قابل تشخيص نباشد.

با در نظر گرفتن كنتراست‌های تفاضلی و رسم تفاوت نقاط فعال برای دوبعدی ژانرهای موسیقی، اولین مشاهده‌ی مهم، خاموش شدن بخش auditory cortex است که گفته‌های پيشين ما را مبنی بر بی‌اثر بودن ژانر موسیقی بر فعالیت اين ناحیه تأييد می‌کند. در ادامه، می‌بينيم که نقاطی در جلو و نقاطی در عقب مغز در عمدۀ شکل‌های تفاضلی رسم شده جلب توجه می‌کنند؛ به اين معنی که ميزان فعالیت نسبتاً بالايی دارند. در نهايّت، می‌توان ادعا کرد که اين نقاط، احتمالاً با مسئله‌ی ژانر موسیقی مرتبط هستند، اما نمی‌توانند به عنوان يك ملاک قطعی مطرح شوند، چرا که مواردی وجود دارد که در اثر تفاوت ژانر موسیقی، فعالیت اين نواحی تفاوتی ندارد و فعالیت تفاضلی آن‌ها نزدیک به صفر است؛ و نيز در بسياري موارد نقاط ديگري هستند که در اثر تفاوت ژانر موسیقی روشن می‌شوند. حداکثر چيزی که می‌توان به اين نقاط نسبت داد، آن است که در درصد قابل قبولی از نمونه‌ها (منظور از درصد قابل قبول، درصدی بيش از درصد طبقه‌بندی تصادفي است) می‌توانند نشان‌گر تفاوت ژانر موسیقی باشند. شکل زير (که در بخش قبل نيز مورد بررسی قرار گرفت)، محل‌های تقریبی اين نقاط را نشان می‌دهد.



يك نکته‌ی ديگر که اشاره به آن خالي از لطف نیست، آن است که همان طور که پيش‌تر اشاره شد، شکل‌های رسم شده برای نقاط فعال مغز، ميزان فعالیت (متناظر با مقادیر z-value) را بر بازه صفر تا يك مپ کرده‌اند. بنابراین، نقاطی که در نمودارهای تفاضلی شاخص هستند و بعضًا با رنگ قرمز مشخص شده‌اند، احتمالاً در مقایسه با نقاطی که در نمودارهای اولیه (که در آن‌ها، كنتراست‌ها صرفاً بر

مبناً یک زانر مشخص شده بود) با رنگ قرمز مشخص شده بودند، فعالیت بسیار کمتری دارند، چراکه نقاط فعال نمودارهای تفاضلی در نمودارهای اولیه جلب توجه نمی‌کردند و فعالیت آنها در مقایسه با نواحی auditory cortex به شکل معناداری کمتر است.

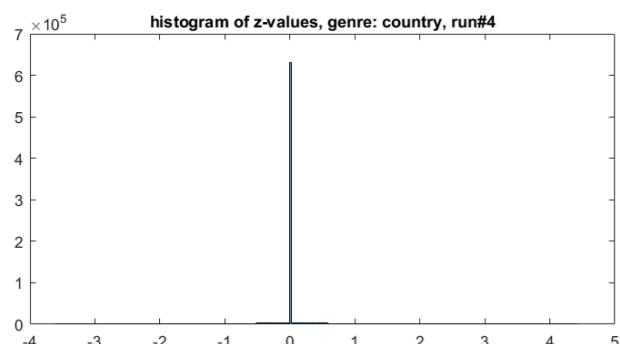
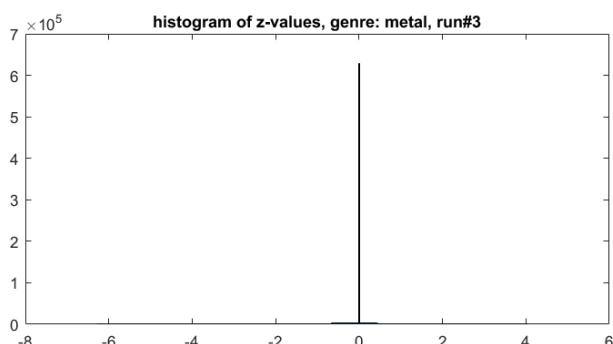
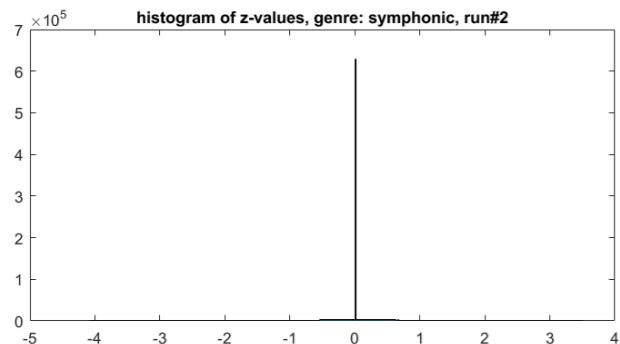
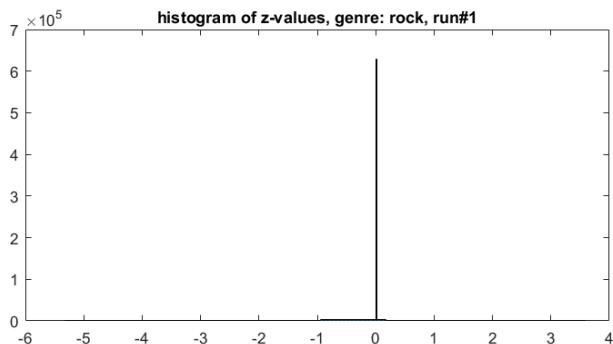
## قسمت چهارم) طبقه‌بندی ژانرهای موسیقی متفاوت

### بخش ۱-۴

در این بخش، همان طور که صورت سؤال بیان می‌کند، هر موسیقی به عنوان یک تحریک در فضای واکسل‌ها به حساب می‌آید، بنابراین ماتریس دیزاین در این بخش با بخش‌های قبلی تفاوت خواهد داشت و برای هر واکسل و در هر ران، ۲۵ رگرسور داریم که همان ۲۵ قطعه موسیقی‌ای هستند که در آن ران پخش شده‌اند. در نهایت پس از تبدیل  $t$ -value‌ها به  $z$ -value هم قرار دادن مقادیر  $z$ -value مربوط به هر قطعه موسیقی، به ماتریسی با ۱۷۵ سطر و  $n$  ستون می‌رسیم که ماتریس ویژگی‌ها (نسخه اولیه) است. یعنی ۱۷۵ نقطه در فضایی  $n$  بعدی داریم. در این مسأله، برای عدد  $n$  داریم:  $n = 160 \times 160 = 9216200$ .

در کد ضمیمه شده به همراه این گزارش، این ماتریس با نام `allFeature` شناخته می‌شود که تعداد ستون‌های آن، ۲ تا بیشتر از عدد  $n$  است که محاسبه شد، که این دو ستون به ترتیب حاوی لیبل (نوع موسیقی) و شماره ران مربوط به هر آهنگ می‌باشند.

مجدداً مشابه بخش ۳-۱، به دلیل آن که گزارش کردن مقادیر عددی به دست آمده مقدور نمی‌باشد، برای به دست آوردن شهودی از مقادیر عددی آن‌ها، به رسم چند هیستوگرام می‌پردازیم. در ادامه، هیستوگرام  $z$ -value‌های ۴ آهنگ که به صورت تصادفی از ران‌ها و ژانرهای مختلف انتخاب شده‌اند، مشاهده می‌کنید.

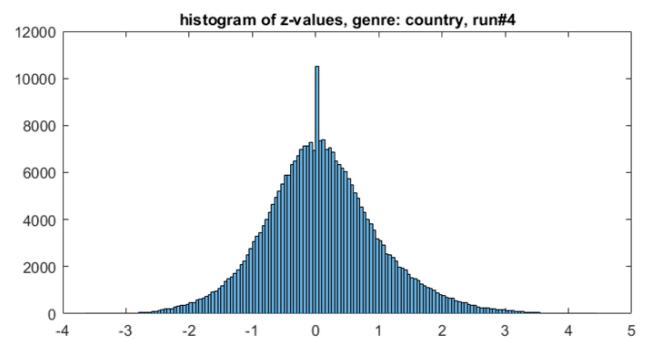
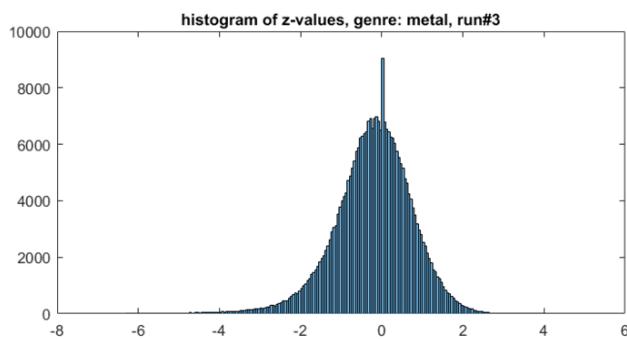
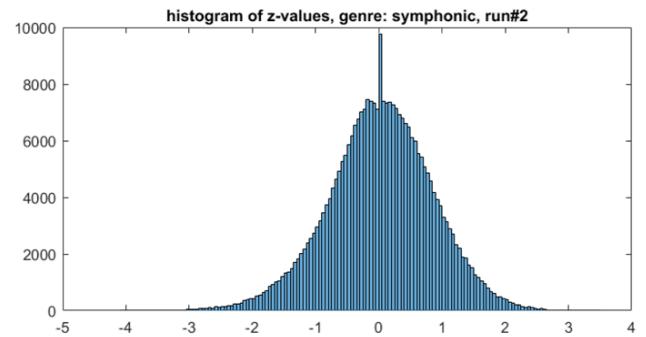
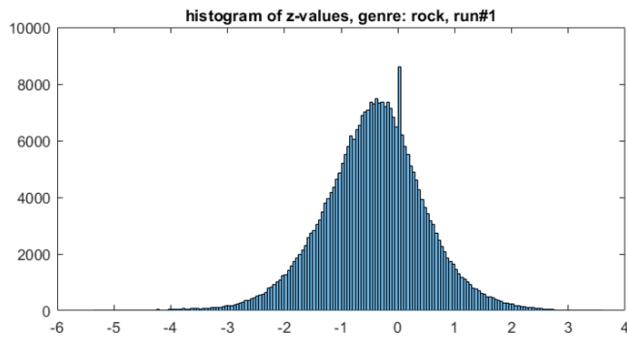


همان طور که در قسمت ۳-۱ نیز مشاهده شد، هیستوگرام‌های رسم شده چندان گویا نیستند، چرا که مقادیر  $z$ -value‌های برابر با صفر به قدری زیاد هستند که سایر مقادیر موجود در هیستوگرام‌ها اصلاً دیده نمی‌شوند.

در ادامه، می‌خواهیم مشابه بخش ۳-۱، مقادیر صفر را در نظر نگیریم و هیستوگرام‌ها را مجدداً رسم کنیم، اما این بار کمی متفاوت عمل می‌کنیم. در بخش ۳-۱، تمامی مقادیر صفر را نادیده گرفتیم. این جا، با توجه به این که در ادامه قرار است با استفاده از آزمون ANOVA، به کاهش بُعد پردازیم، به این مشکل برخواهیم خورد که واکسل‌هایی که مقادیر  $z$ -value آن‌ها در تمامی نقاط (منظور از نقاط، همان آهنگ‌ها هستند که نقاطی در فضای بُعد بالا محسوب می‌شوند) صفر هستند، خروجی‌های NaN خواهند داشت. (در مورد عملکرد ANOVA در ادامه بیشتر توضیح داده خواهد شد) بنابراین، واکسل‌هایی که مقادیر  $z$ -value آن‌ها برای تمام آهنگ‌ها صفر است را حذف می‌کنیم. این کار هم به رسم هیستوگرام‌های این قسمت کمک می‌کند، و هم در مسأله‌ی طبقه‌بندی

که در ادامه بررسی می‌شود، هیچ اثر بدی نخواهد داشت، چرا که واکسل‌هایی که به ازای تمامی ژانرها مقادیر یکسان (صفر) دارند، نمی‌توانند در طبقه‌بندی نقشی ایفا کنند.

با حذف این واکسل‌ها، دیتای جدید در ماتریس Feature ذخیره می‌شود و هیستوگرام‌های صفحه قبل به شکل زیر در می‌آیند.



مشاهده می‌شود که نتایج منطقی‌تر شدند، و بازه‌ی توزیع z-value‌ها و چگالی آن‌ها نیز منطقی است، و نیز مجدداً مقادیر صفر بیشتر هستند. (دقت داریم که برخلاف قسمت ۱-۳، همه‌ی مقادیر صفر را حذف نکردیم، بلکه واکسل‌هایی را حذف کردیم که برای همه‌ی آهنگ‌ها، z-value‌ای برابر با صفر دارند).

**بخش ۲-۴**

می خواهیم با استفاده از آزمون فرض ANOVA، تعدادی از واکسل های موجود در ماتریس Feature را انتخاب کنیم. در واقع در فرایند (ANOVA)، آزمون فرضی به نام F-test چندین بار انجام می شود و برای هر واکسل، کمیتی به نام F از رابطه زیر محاسبه می شود:

$$F = \frac{\text{between-group variability}}{\text{within-group variability}}$$

که در آن، تغییرات بین گروهی (صورت کسر F) از طریق زیر محاسبه می شود:

$$\sum_{i=1}^K n_i (\bar{Y}_{i\cdot} - \bar{Y})^2 / (K - 1)$$

( $n_i$  تعداد مشاهدات مربوط به گروه i و  $\bar{Y}_{i\cdot}$  میانگین نمونه گروه i و K تعداد گروه هاست)

همچنین تغییرات درون- گروهی (مخرج کسر F) نیز از رابطه زیر محاسبه می شود:

$$\sum_{i=1}^K \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2 / (N - K)$$

( $Y_{ij}$  زامین مشاهده از i امین گروه است، و N تعداد کلی نمونه هاست)

آن چه که توضیح داده شد، مربوط به One-Way Analysis of Variance بود. به شکل متناظر، می توان n-Way Variance نیز در نظر گرفت که در آن، نمونه ها متأثر از n متغیر مستقل هستند و در نتیجه، در خروجی نیز p-value خواهیم داشت.

• مرجع توضیحات: <https://en.wikipedia.org/wiki/F-test>

در این پژوهش، از One-Way ANOVA استفاده می کنیم که در مطلب با استفاده ازتابع anova1 پیاده سازی شده است. در ادامه باید p-value های کوچک (کوچک تر از یک مقدار آستانه) را نگه داریم، چرا که فرض صفر این آزمون آن است که (میانگین) توزیع دیتای موجود در گروه های مختلف یکسان است، بنابراین کوچک متناظر با واکسل هایی هستند که با احتمالی ناچیز، دارای توزیع یکسان برای گروه های مختلف هستند.

به عنوان شروع در این بخش، مقدار این حد آستانه را  $p = 0.01$  در نظر می گیریم.

بخش ۳-۴

با حد آستانه‌ای که در قسمت قبل در نظر گرفتیم ( $p = 0.01$ )، طبقه‌بندی کننده LDA را می‌سازیم و آن را روی خود دیتا تست می‌کنیم. نتیجه‌ای که مشاهده می‌شود، چندان عجیب نیست و طبقه‌بندی کننده، روی دیتای آموزش (train) دقیق صد درصدی دارد و خطای ندارد.

```
AccuracyOnTrainData =
```

```
100
```

**بخش ۴-۴ و ۵-۴**

دو بخش ۴-۴ و ۵-۴ را به موازات یک دیگر و کمی فراتر از آن چه که سؤال خواسته است انجام داده ایم، بنابراین نتایج را نیز به موازات گزارش می کنیم.

می خواهیم بهترین ترشهولد ممکن را برای طبقه بندی LDA بیابیم. برای این کار، دامنه‌ی وسیعی از ترشهولدها را در نظر می‌گیریم و با همه‌ی آن‌ها، فرایند کراس ولیدیشن را انجام می‌دهیم.

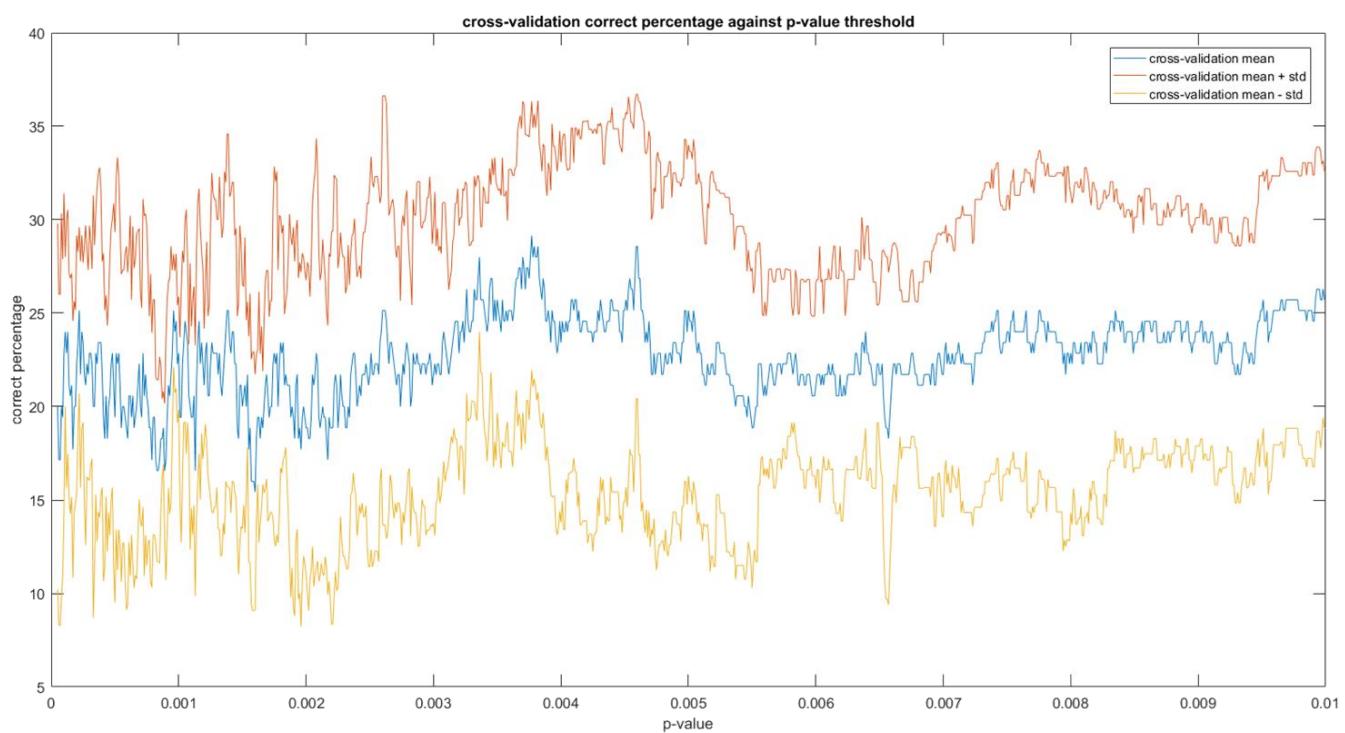
در این فرایند، روشی را پیش می‌گیریم که از وجود هر نوع بایاس در کراس ولیدیشن جلوگیری کند تا بتوانیم بهترین شاخص ممکن را برای عملکرد طبقه بندی کننده‌ی خود به دست آوریم. هر بار، یکی از ۷ ران آزمایش را از دیتای آموزش خارج می‌کنیم، و هر بار مستقلًا فرایند کاهش بُعد را تکرار می‌کنیم تا هیچ اثری از ران کنارگذاشته شده در تولید طبقه بندی کننده باقی نماند.

با انجام این کار، برای هر ترشهولد، به ۷ نتیجه می‌رسیم که هر کدام، درصد پاسخ درست طبقه بندی کننده در هر کدام از ۷ مرحله‌ی کراس ولیدیشن هستند. با محاسبه میانگین و انحراف معیار این ۷ عدد، به آماره‌هایی برای سنجش درصد پاسخ درست هر ترشهولد می‌رسیم.

در اینجا، ما طیف ترشهولدهای مورد بررسی را به صورت زیر تعریف کرده‌ایم:

```
threshold = 0.00005 : 0.00001 : 0.01;
```

در ادامه، نمودار درصد پاسخ درست طبقه بندی کننده در فرایند کراس ولیدیشن بر حسب ترشهولد را (به همراه دو کران بالا و پایین با فاصله‌ی یک انحراف معیار) مشاهده می‌کنید:



نمودار فوق می‌تواند ترشهولد بهینه را به ما معرفی کند. با محاسبه ماقادیر این نمودار، خواهیم داشت:

```
optimum_threshold =
```

```
0.0037700000000000
```

گزارش کار پروژه درس علوم اعصاب محاسباتی  
برای این ترشهولد، مقدار درصد پاسخ صحیح کراس ولیدیشن و مقدار انحراف معیار به صورت زیر است: (که البته به صورت کلی از روی نمودار نیز مشخص است)

MaximumAccuracy =

29.1429

StandardDeviation =

7.1979

همچنین تعداد واکسل هایی که در هر کدام از 7 مرحله‌ی کراس ولیدیشن برای این ترشهولد بهینه مورد استفاده قرار می‌گیرد، به صورت زیر است:

number\_of\_voxels =

892            585            805            445            910            729            1316

و نیز، هنگامی که با این ترشهولد، روی کل دیتای موجود طبقه‌بندی کننده را بسازیم، تعداد واکسل های انتخابی برابر با ۸۶۳ واکسل خواهد بود.

در این جا، مسئله‌ی پیدا کردن ترشهولد تمام است، اما اشاره به این نکته خالی از لطف نیست که با توجه به نمودار موجود در صفحه قبل، می‌بینیم که نوسانات این نمودار بسیار زیاد است، و میزان پاسخ‌گویی وابستگی شدیدی به مقدار ترشهولد دارد، به گونه‌ای که زمانی که ما برای به دست آوردن لیل‌های ران هشتم که خواسته‌ی اصلی سؤال است، به جای  $p = 0.00377$  (که ترشهولد بهینه است)، از  $p = 0.0038$  استفاده کردیم، از میان ۲۵ لیل، ۳ مورد متفاوت مشاهده شد، یعنی به ازای این تغییر جزئی، ۱۲ درصد در پاسخ تغییر ایجاد شد:

$$\frac{\Delta n/n}{\Delta p/p} = 15.2$$

مشاهده می‌کنیم که حساسیت این مسئله به قدری زیاد است که اصلاً بعد نیست ترشهولد بهینه‌ای که با کراس ولیدیشن به دست آورده، برای ارائه‌ی جواب نهایی و پیدا کردن لیل‌های مجھول بهترین انتخاب نباشد و بتوان با ترشهولد دیگری، جواب بهتری گرفت؛ اما با توجه به دیتای موجود، بیش از این کاری از دست ما بر نمی‌آید و به عنوان بهترین ترشهولد، همین عدد را اعلام می‌کنیم.

**Bonferroni Correction (۱)**

در آزمون فرض، خطای نوع اول خطای ناشی از آن است که فرض درست را به اشتباه رد کنیم. اصولاً در یک آزمون فرض، حدی به نام  $\alpha$  مشخص می‌شود که اگر فرض صفر با احتمالی کمتر از  $\alpha$  درست باشد، این فرض را رد می‌کنیم. حال، با زیاد شدن تعداد دفعات آزمون فرض، احتمال خطای نوع اول نیز افزایش می‌یابد، به طوری که اگر  $m$  بار این آزمون را انجام دهیم، به طور متوسط به تعداد  $m\alpha$  خطای نوع اول رخ می‌دهد که به ازای مقادیر بزرگ  $m$ ، تعداد قابل توجهی خواهد بود. تصحیح بنفرونی، مبتنی بر همین موضوع است و در این روش، به جای آن که سطح اطمینان را برابر با  $\alpha/m$  می‌گیرند. این باعث سخت‌گیرانه‌تر شدن آزمون می‌شود و مقدار خطای نوع اول را کاهش می‌دهد.

همچنین اگر  $H_i$ ‌ها خانواده‌ای از فرض‌ها با  $p$ -value‌های متناظر  $p_i$  باشند، احتمال آن است که حداقل یک بار مرتکب خطای نوع اول بشویم. همچنین اگر  $m_0$  تعداد کل فرض‌های صفر و  $m$  تعداد فرض‌های صفر درست باشد، مطابق با اثبات زیر نشان می‌دهیم که تصحیح بنفرونی، مقدار FWER را کوچک‌تر یا مساوی  $\alpha$  نگه می‌دارد:

$$\text{FWER} = P \left\{ \bigcup_{i=1}^{m_0} \left( p_i \leq \frac{\alpha}{m} \right) \right\} \leq \sum_{i=1}^{m_0} \left\{ P \left( p_i \leq \frac{\alpha}{m} \right) \right\} = m_0 \frac{\alpha}{m} \leq m \frac{\alpha}{m} = \alpha.$$

**Benjamini–Hochberg procedure (۲)**

در این روش، برخلاف روش قبل که مقدار FWER را در سطح  $\alpha$  نگه می‌داشت، کمیت دیگری در این سطح کنترل می‌شود. این کمیت FDR یا نرخ کشف اشتباه (False Discovery Rate) نامیده می‌شود، متوسط تعداد دفعات رخ دادن خطای نوع اول به تعداد دفعات رد شدن فرض صفر در انجام مکرر یک آزمون فرض است.

با نام گذاری موارد فوق، می‌توان FDR را به صورت زیر تعریف کرد:

	Null hypothesis is true ( $H_0$ )	Alternative hypothesis is true ( $H_A$ )	Total
Test is declared significant	$V$	$S$	$R$
Test is declared non-significant	$U$	$T$	$m - R$
Total	$m_0$	$m - m_0$	$m$

$V$  متغیر تصادفی تعداد خطاهای نوع اول

$S$  متغیر تصادفی تعداد دفعات درست رد کردن فرض صفر

$T$  متغیر تصادفی تعداد خطاهای نوع دوم

$U$  متغیر تصادفی تعداد دفعات درست پذیرفتن فرض صفر

$m$  تعداد دفعات انجام آزمون فرض

$m_0$  تعداد فرض‌های صفر درست

$$Q = V/R = V/(V + S).$$

تعریف میانگین یافته‌های اشتباه (FDR) :

$$\text{FDR} = Q_e = \text{E}[Q]$$

در این روش، الگوریتمی برای آزمون فرض ارائه می‌شود که FDR را کنترل می‌کند. این الگوریتم به شرح زیر است:  
 اگر  $H_i$  ها فرض‌های صفر با مقادیر p-value  $P_i$  متناظر باشند، ابتدا برای یک  $\alpha$  مشخص،  $P_i$  ها را به ترتیب صعودی مرتب کرده و آن‌ها را  $P_{(i)}$  می‌نامیم. سپس بزرگترین عدد  $k$  را پیدا می‌کنیم، به طوری که داشته باشیم:  $P_{(k)} \leq \frac{k}{m} \alpha$ . در ادامه فرض صفر را برای  $H_{(i)}$  هایی با  $i \leq k$  رد می‌کنیم.

نکته ۱) این فرایند در شرایط مستقل بودن  $m$  آزمون فرضیه (و در برخی شرایط خاص در حالت وابستگی) برقرار است.  
 نکته ۲) این فرایند FDR را در سطح  $\alpha$  نگه می‌دارد و در مورد FWER که در بخش قبل بررسی شد، الزاماً کنترلی ندارد. این امر باعث می‌شود که این روش سخت‌گیری کمتری داشته باشد که مسبب کاربرد وسیع‌تر آن شده است.

مراجع:

- [https://en.wikipedia.org/wiki/Bonferroni\\_correction](https://en.wikipedia.org/wiki/Bonferroni_correction)
- <http://statweb.stanford.edu/~ckirby/brad/LSI/chapter2.pdf>
- [https://en.wikipedia.org/wiki/False\\_discovery\\_rate](https://en.wikipedia.org/wiki/False_discovery_rate)
- <http://statweb.stanford.edu/~ckirby/brad/LSI/chapter4.pdf>

**Logistic Regression (۱)**

برای انجام طبقه‌بندی Logistic Regression از توابع mnrrfit و mnrvval در مطلب استفاده کردیم. خروجی تابع mnrrfit، ماتریس ضرایب تخمینی برای multinomial logistic regression است. با دادن این ماتریس به عنوان ورودی تابع mnrvval به همراه ماتریس ویژگی‌های دیتای تست، لیل‌های خروجی طبقه‌بندی کننده به دست می‌آید. ضمناً لیل‌های ورودی (هم در تابع mnrvval و هم در تابع mnrrfit با فرمت categorical داده شده‌اند)

با توجه به تعداد واکسل‌هایی که در هر مرحله از کراس‌ولیدیشن در نظر گرفته می‌شوند، ماتریس ضرایب multinomial logistic regression در هر مرحله  $n \times 150$  خواهد بود که  $n$  تعداد واکسل‌هاست. با توجه به این که تعداد واکسل‌ها به طور متوسط حدود ۸۰۰ تا هستند، یک بار انجام دادن کراس‌ولیدیشن (به ازای یک ترشهولد مشخص)، زمان زیادی (حدود ۱ ساعت) از مالگرفت، بنابراین دیگر رسم نمودار دقت کراس‌ولیدیشن بر حسب ترشهولد امکان‌پذیر نبود و مسئله‌ی logistic regression را صرفاً یک بار برای همان ترشهولد بهینه‌ای که در قسمت ۴-۴ و ۵-۴ بدست آورده بودیم ( $p = 0.00377$ )، حل کردیم.

نتیجه‌ی تست طبقه‌بندی کننده بر روی داده‌ی تست، (مشابه LDA) برابر با ۱۰۰ درصد بود.

نتایج کراس‌ولیدیشن به صورت زیر است:

`LogisticRegressionCrossValAccuracy =`

24      28      20      24      20      20      12

درصدهای فوق، برای هر کدام از مراحل هفت‌گانه‌ی کراس‌ولیدیشن هستند. میانگین درصد پاسخ‌گویی برابر با ۲۱.۱۵٪ و انحراف معیار برابر ۵.۰۱٪ می‌باشد.

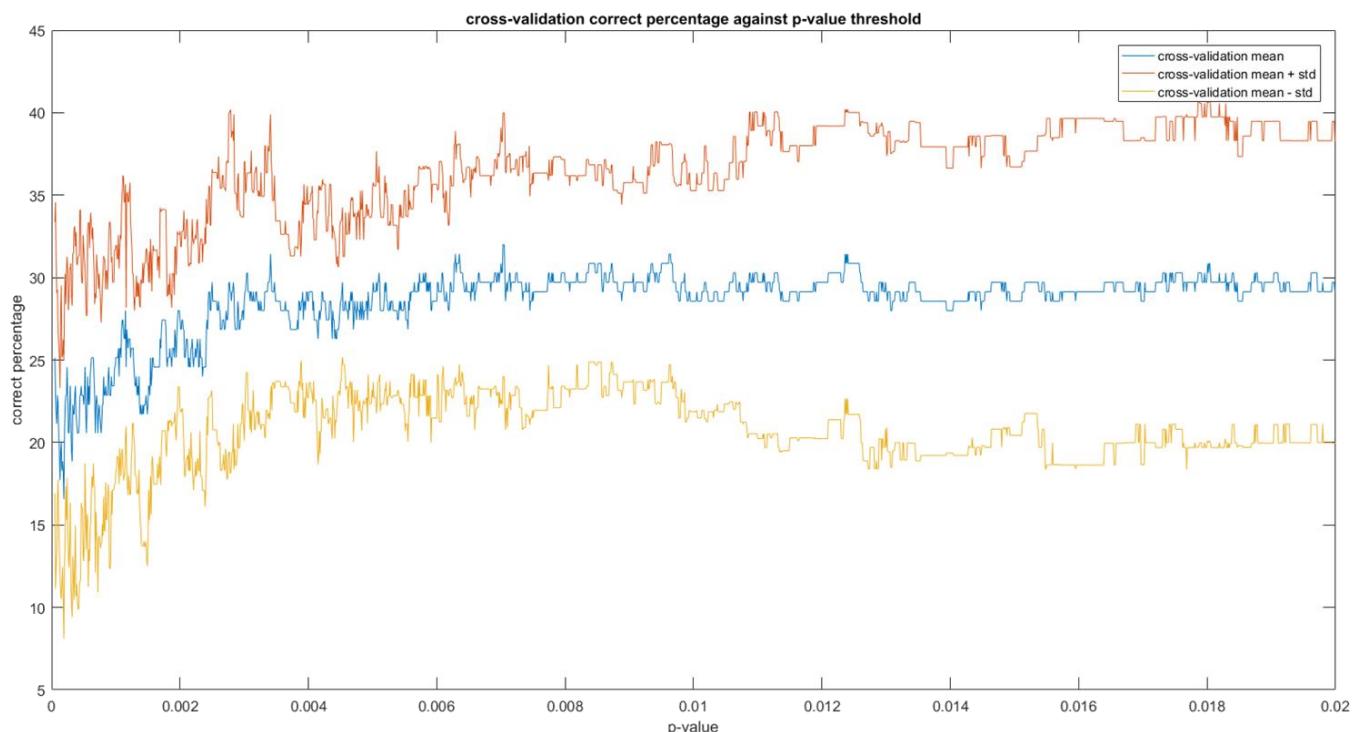
می‌بینیم که درصد پاسخ‌گویی عملاً بی‌فایده است، چرا که طبقه‌بندی تصادفی به شکل متوسط ۲۰٪ پاسخ می‌دهد، اما همان طور که ذکر شد، به دلیل محدودیت عملیاتی نتوانستیم این طبقه‌بندی را برای ترشهولدهای مختلف انجام دهیم و این احتمال وجود داشت که برای ترشهولدها دیگری، به درصدهای بهتری دست پیدا کنیم.

در این بخش، مشابه بخش قبل و بخش های ۴-۵، این بار با طبقه‌بندی کننده SVM به حل مسئله می‌پردازیم. این جا برخلاف Logistic Regression محدودیت عملیاتی نداریم و می‌توانیم مشابه بخش های ۴-۵، برای بازه‌ی وسیعی از ترشهولدها، کراس‌ولیدیشن را انجام داده و نتایج را مقایسه کنیم. محدوده ترشهولدها به صورت زیر تعریف شده است:

$$\text{threshold} = 0.00005 : 0.00001 : 0.02;$$

ابتدا ذکر می‌کنیم که درصد پاسخ‌گویی طبقه‌بندی کننده روی کل داده‌ی آزمایش (مشابه با LDA و Logistic Regression) برابر با ۱۰۰ درصد بود.

در ادامه، نمودار درصد پاسخ درست طبقه‌بندی کننده در فرایند کراس‌ولیدیشن بر حسب ترشهولد را (به همراه دو کران بالا و پایین با فاصله‌ی یک انحراف معیار) مشاهده می‌کنید:



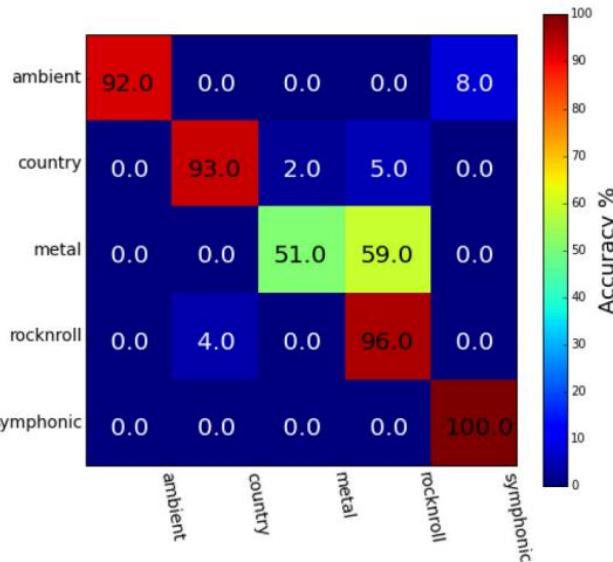
مشاهده می‌شود که در طبقه‌بندی SVM درصد ماقسیمم اندکی بیشتر از LDA می‌شود (افزایشی در حدود ۳ درصد، از ۲۹ به ۳۲) اما شاید نکته‌ی مهم‌تر آن است که SVM شدت نوسان کمتری در درصد پاسخ‌گویی دارد و این می‌تواند اعتبار این طبقه‌بندی را نسبت به LDA در این مسئله‌ی خاص بیشتر کند، چراکه با دلیل نوسانات شدید پاسخ LDA، جواب به شدت به مقدار دقیق ترشهولد حساس بوده و نمی‌توان با اطمینان از تکرار درصدهای بالای آن در آزمون‌های بعدی سخن گفت. (توضیحات بیشتر در بخش ۴-۵)

**بخش ۴-۴**

برای ترشهولد بهینه‌ی یافته شده در بخش ۴-۴ و ۴-۵، ماتریس کانفیوژن به صورت زیر است:

Confusion Matrix						
True Class	28.57143	31.42857	11.42857	20	17.14286	Rock
	25.71429	25.71429	20	14.28571	20	Symp.
	14.28571	8.571429	25.71429	8.571429	22.85714	Metal
	25.71429	25.71429	22.85714	42.85714	17.14286	Country
	5.714286	8.571429	20	14.28571	22.85714	Ambient
	Rock	Symp.	Metal	Country	Ambient	
Predicted Class						

آن چه که در بخش behavioral genre category-assignment در مقاله گزارش شده است، نتایجی به شکل زیر دارد:



مشاهده می‌شود که به صورت behavioral، تقریباً همه‌ی ژانرهای به درستی شناخته شده‌اند، و صرفاً افراد بعضاً به اشتباه ژانر metal را rock تشخیص داده‌اند.

در کل، شباهت‌های زیادی بین نتایج ما و نتایج behavioral genre category-assignment وجود ندارد، که طبیعی است، چراکه اولاً تشخیص به صورت رفتاری در مورد ژانر آهنگ ممکن است هیچ ارتباطی با محل کد شدن این آهنگ‌ها در مغز نداشته باشد؛ و ثانیاً نتایج گزارش شده از مقاله در ماتریس فوق به صورت میانگینی روی افراد مختلف است، در حالی که نتایج ما مربوط به یک سایجکت خاص است.

با این وجود، برخی شباهت‌های جزئی بین این دو ماتریس نیز وجود دارد، از قبیل کم بودن درصد پاسخ صحیح به ژانر metal یا درصد نسبتاً بالای تشخیص ambient به جای symphonic

در نهایت، دوباره ذکر می‌کنیم که از ابتدا هم انتظار نداشتم که نتایج این دو ماتریس کانفیوژن الزاماً به هم شبیه باشد، چراکه فرایند تشخیص ژانر موسیقی به صورت رفتاری و کد شدن ژانرهای موسیقی در نقاط مختلف مغز (در صورتی که اصلاً وجود داشته باشد) ممکن است کاملاً با هم بی ارتباط باشد، یا حداقل ارتباط آن با روشنی که ما در پیش گرفتیم به وضوح آشکار نشود.

#### بخش ۴-۹

فایل Predicted\_Label.mat خصیمه‌ی تمرین شده است. این فایل دارای فرمت cell است و ژانر ۲۵ آهنگ پخش شده در ران هشتم با طبقه‌بندی کننده LDA و با استفاده از ترشحولد پیدا شده در قسمت ۴-۴ و ۴-۵ پیش‌بینی شده است.

**بخش ۴ - ۱۰**

در این قسمت، هدف ارائه‌ی بهترین طبقه‌بندی کننده برای جداسازی ژانرهای موسیقی بود. با استفاده از 1-way ANOVA، به کاهش بُعد فضای مسئله، یعنی به حذف واکسل‌هایی که اطلاعات کمی می‌دهند، پرداختیم. برای پیدا کردن ترشهولد بهینه، فرایند کراس‌ولیدیشن را (با طبقه‌بندی کننده LDA) برای طیف وسیعی از ترشهولدها تکرار کردیم و نمودار درصد دقت بر حسب ترشهولد را رسم کردیم و ماکسیمم آن را پیدا کردیم. این ماکسیمم در  $p = 0.00377$  می‌افتد و مقدارش برابر با 29.1% بود. عملیات مشابه را برای طبقه‌بندی کننده SVM نیز انجام دادیم که درصد صحت اندکی بهبود یافت و تا 32% نیز رسید. همچنین به ازای ترشهولد بهینه‌ای که برای LDA یافته بودیم، طبقه‌بندی Logistic Regression را نیز امتحان کردیم که پاسخ مطلوبی نداشت و در حدود 21% پاسخ می‌داد. البته به دلیل محدودیت‌های عملیاتی نتوانستیم این طبقه‌بندی را به ازای مقادیر مختلف ترشهولدها تست کنیم. احتمالاً اگر چنین می‌کردیم، می‌توانستیم درصدهای بالاتری با این طبقه‌بندی کننده بگیریم. بنابراین در اینجا نمی‌توانیم Logistic Regression را با LDA و SVM مقایسه کنیم، چراکه اطلاعات کافی را در مورد آن در اختیار نداریم.

در مقادیر بهینه‌ای که یافتیم، به طور میانگین چیزی حدود ۸۰۰ واکسل برای طبقه‌بندی مورد استفاده قرار می‌گرفت و همان طور که بیان شد، درصد پاسخ‌گویی متوسط حدوداً برابر اب ۳۰ درصد بود، اما در مراحل مختلف کراس‌ولیدیشن، در بازه‌ای از حدود ۱۵ درصد تا ۴۰ درصد تغییر می‌کرد.

نکته‌ی دیگر آن که نمودارهای درصد دقت بر حسب ترشهولد، نشان می‌داد که به خصوص برای طبقه‌بندی کننده LDA، میزان حساسیت پاسخ به ترشهولد بسیار بالاست. (این مورد را در طبقه‌بندی پاسخ نهایی با دو مقدار  $p = 0.0038$  و  $p = 0.00377$  نیز مشاهده کردیم که تفاوتی ۱۲ درصدی در توزیع پاسخ‌ها مشاهده شد – توضیحات بیشتر در بخش ۴-۵ و ۴-۶) و این می‌تواند دلیلی باشد بر آن که نتایج پایداری و قابلیت اتکای زیادی ندارند، که البته در این مورد نتایج SVM نسبتاً پایدارتر به نظر می‌رسید.

نهایتاً با به دست آوردن ماتریس کانفیوژن مربوط به طبقه‌بندی LDA با ترشهولد بهینه دیدیم که بیشترین پاسخ صحیح مربوط به ژانر country است و در رتبه‌های بعدی، ژانر rock و بعد metal و symphonic در کنار هم و نهایتاً ambient قرار دارند. همه‌ی درصدهای بیشتر از ۲۰ هستند (بهتر از حالت تصادفی عمل می‌کنیم) و محدوده تغییرات از 22.8% تا 42.8% با میانگین 29.1% است. همچنین بیشترین تشخیص‌های اشتباه نیز در مثلث rock – symphonic – country مشاهده می‌شود.

## قسمت پنجم) یک سؤال دلخواه

در این بخش، از یک ملاک دیگر برای کاهش بُعد فضای مسئله استفاده می‌کنیم. ملاکی که معرفی می‌کنیم، شاخصی برای فاصله‌ی بین دو گروه بر مبنای یک فیچر است، به این معنی که مقداری عددی به عنوان فاصله‌ی دو کلاس ارائه می‌کنیم (در صورتی که بخواهیم این دو کلاس را به وسیله‌ی یک فیچر جدا کنیم). بنابراین به ازای هر فیچر و هر دو کلاس، یک مقدار قابل محاسبه است.

این ملاک که با نام  $J$ -value شناخته می‌شود، به صورت زیر محاسبه می‌گردد:

$$J = \frac{|\mu_0 - \mu_1|^2 + |\mu_0 - \mu_2|^2}{\sigma_1^2 + \sigma_2^2}$$

که در آن،  $\mu_0$  میانگین فیچر مربوطه روی هر دو کلاس،  $\mu_1$  میانگین آن روی کلاس اول،  $\mu_2$  میانگین آن روی کلاس دوم، و  $\sigma_1^2$  و  $\sigma_2^2$  به ترتیب واریانس آن روی کلاس اول و دوم می‌باشند.

می‌توان از طریق این فرمول مشاهده کرد که اگر میانگین دو کلاس برابر باشد، مقدار  $J$  صفر خواهد شد. از این جهت، کار ما تا حدودی شبیه به ANOVA است، اما دو تفاوت عمده وجود دارد، یکی مربوط به چگونگی محاسبه‌ی کمیت عددی‌ای که بر مبنای آن به جداسازی فیچرهای می‌پردازیم، و دوم نحوه جداسازی، که در طول پژوهه بر مبنای ترشهولد  $p$ -value بوده است، اما این جا متفاوت خواهد بود و در ادامه توضیح داده می‌شود.

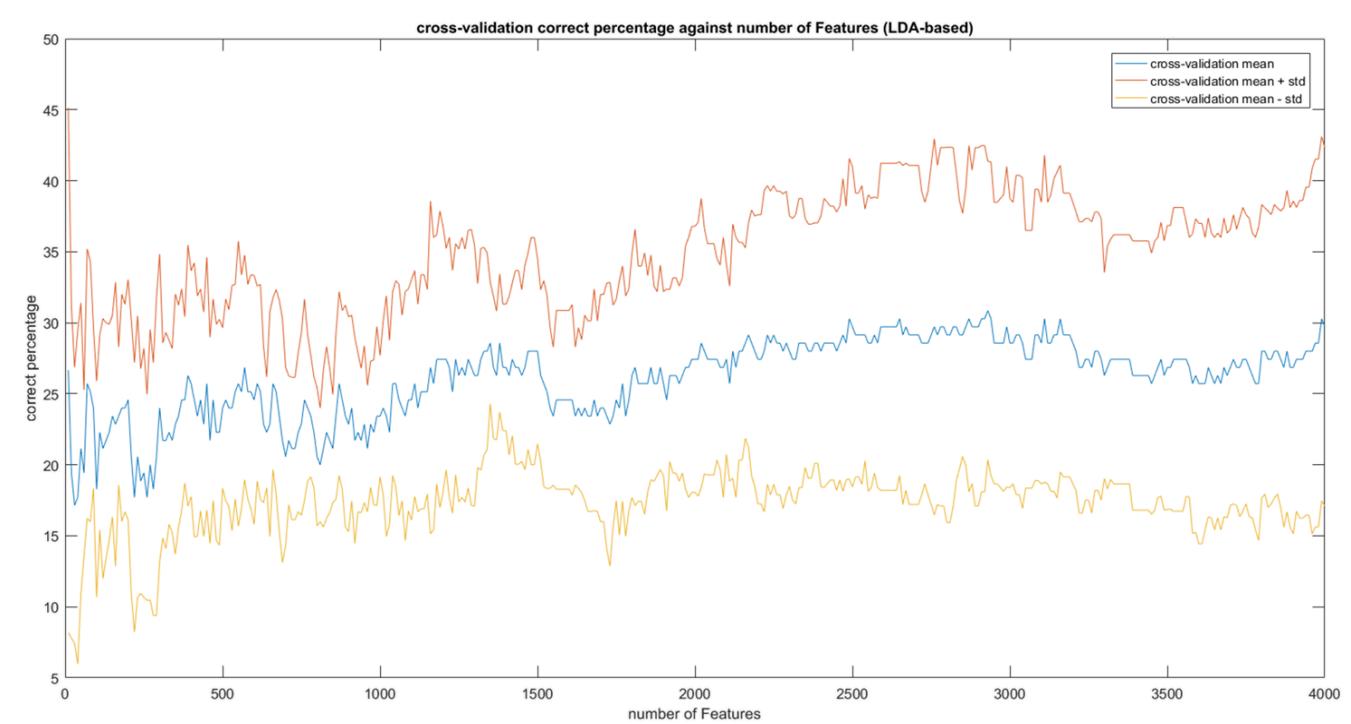
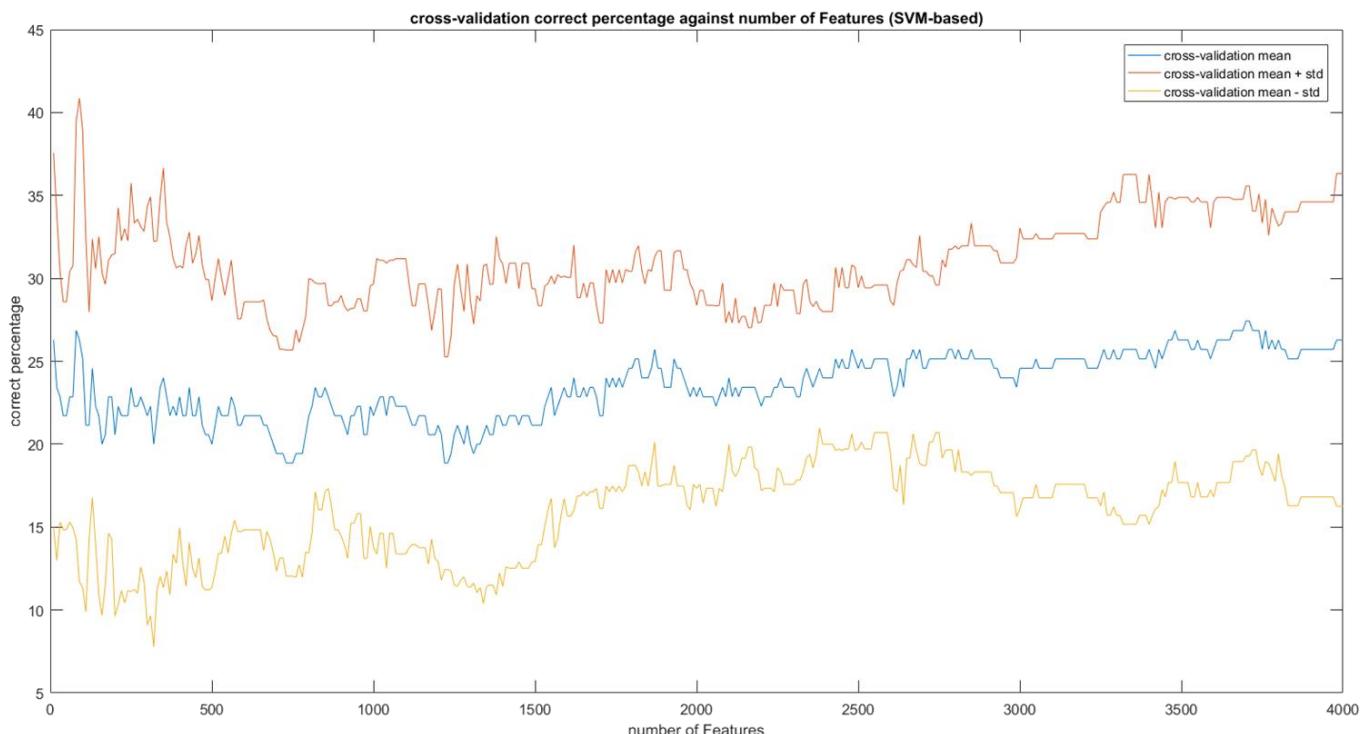
همان طور که گفته شد،  $J$ -value بین هر دو کلاس محاسبه می‌شود. بنابراین در این مسئله، با توجه به این که ۵ کلاس داریم، به ازای هر واکسل می‌توان  $10 = \binom{5}{2}$   $J$ -value محاسبه کرد. نهایتاً با محاسبه مقادیر  $J$  برای تمامی واکسل‌ها، ماتریسی  $n \times 10$  خواهیم داشت که  $n$  تعداد کل واکسل‌هاست.

برای کاهش بُعد، می‌توان روش‌های متفاوتی پیش‌گرفت. نکته‌ی اول آن است که کاهش بُعد را بر مبنای تعداد واکسل (ونه ترشهولدی روی مقادیر عددی محاسبه شده) انجام می‌دهیم. یعنی تصمیم می‌گیریم  $m$  واکسل را نگه داریم و از بقیه صرف نظر کنیم. در ادامه باید این  $m$  واکسل را انتخاب کنیم. می‌دانیم که هر واکسل  $10 = \binom{5}{2}$   $J$ -value دارد. یک روش می‌تواند این باشد که روی این  $10$  مقدار میانگین بگیریم و سپس،  $m$  واکسل با بزرگترین میانگین‌ها را انتخاب کنیم. ما از این روش استفاده نخواهیم کرد، چراکه ممکن است یک جفت از کلاس‌ها به شکل معناداری بهتر از بقیه از هم جدا شوند. در این صورت این کار منجر به آن می‌شود که واکسل‌هایی انتخاب شوند که همگی آن دو کلاس خاص را به خوبی از بقیه هم می‌کنند. برای جلوگیری از این مشکل، واکسل‌ها را در  $10 = \binom{5}{2}$  جدالگانه انتخاب می‌کنیم. به عبارتی، گروهی از واکسل‌ها به خوبی کلاس ۱ و ۲ را از هم جدا می‌کنند. گروه دیگر، به خوبی کلاس ۱ و ۳ را از هم جدا می‌کنند؛ و... کاری که ما انجام می‌دهیم، آن است که به تعداد  $10/m$  از هر کدام از این گروه‌ها را جدا می‌کنیم. یعنی ما برای جداسازی مناسب هر دو کلاس از هم، واکسل‌هایی را انتخاب کرده‌ایم. این کار با مرتب‌سازی ماتریس  $J$ -value در هر سطر، و انتخاب  $10/m$  واکسل برتر هر سطر صورت می‌پذیرد.

پی‌نوشت: ما پیش‌تر با این روش و کار با مقادیر  $J$  آشنا بودیم، اما پس از انجام این عملیات در این پژوهه و بررسی مجدد، متوجه شدیم که  $J$ -value‌ها، دقیقاً همان  $F$ ‌هایی هستند که در ANOVA به کار می‌روند و صرفاً در حد یک ضریب با آن‌ها تفاوت دارند. بنابراین در این نقطه متوجه می‌شویم که ارزش کاری که انجام داده‌ایم و نتایجش را در ادامه می‌بینیم، کمتر از چیزی بوده که فکر شر را می‌کرديم و ما عملاً روش جدیدی برای تعیین فاصله دو کلاس ارائه ندادیم و تنها بخش جدید این فعالیت، آن است که به بررسی دو به دوی کلاس‌ها پرداخته‌ایم که البته این خود، می‌تواند سؤال جالبی باشد که آیا بررسی دو به دوی کلاس‌ها و استخراج واکسل‌هایی که این وظیفه را به عهده دارند، تغییری در نتایج قبلی ما ایجاد می‌کند یا نه.

برای پیاده‌سازی این الگوریتم، تابعی با نام  $Jvalue$  نوشته‌ایم که وظیفه‌ی محاسبه‌ی مقادیر  $J$  را به عهده دارد. سپس به ازای مقادیر مختلف  $m$ ، عملیات کراس ولیدیشن را به نحوی که قبلاً نیز انجام داده بودیم (یعنی محاسبه  $J$ ها، در حالتی که یکی از ران‌ها از دیتای اصلی حذف شده است) را انجام دادیم و مقادیر درصد دقت کراس ولیدیشن  $SVM$  و هم برای طبقه‌بندی  $LDA$  به دست آوردیم.

نتایج دو طبقه‌بندی  $SVM$  و  $LDA$  به ترتیب در دو نمودار زیر (نمودارهای درصد دقت کراس ولیدیشن بر حسب تعداد واکسل انتخابی) قابل مشاهده هستند.



مشاهده می شود که درصد های به دست آمده با مقادیری که با آزمون فرض ANOVA به دست آمده بودند، رفتار نسبتاً مشابهی دارند که با توجه به پی نوشت موجود در صفحه ۳۷، این امر قابل انتظار بود. در اصل، نتیجه های که به دست می آید آن است که تقسیم بندی بر مبنای تعداد واکسل و بر مبنای جداسازی دوبه دوی کلاس ها، نتیجه های بهتری در مقایسه با رویکرد p-value ندارد.

ضمیناً یک مشاهده جالب دیگر را می توان در نمودارهای فوق، به خصوص نمودار مربوط به SVM انجام داد، و آن مشاهده این است که با افزایش تعداد واکسل ها از جایی به بعد، نتایج از نظر متوسط درصد پاسخ صحیح تقریباً به حالت اشباع می رسد؛ اما انحراف معیار زیاد می شود و بازه هی حول این مقدار اشباع شده بزرگ و بزرگ تر می شود. این می تواند نشان از ناپایدار شدن روش برای تعداد های زیاد واکسل های انتخابی باشد که احتمالاً می تواند اثر overfitting را به ما نشان بدهد. ضمیناً انتظار داریم که با افزایش بیش از حد تعداد واکسل ها، حتی متوسط درصد پاسخ نیز افت کند که اثر همان overfitting است.