

باسمه تعالی



دانشگاه صنعتی شریف

دانشکده مهندسی برق

## درس علوم اعصاب محاسباتی

گزارش پروژه درس

سری سوم

P300 Speller

استاد درس:

دکتر کربلایی آقاجان

امیرحسین افشارراد

۹۵۱۰۱۰۷۷

طاها انتصاری

۹۵۱۰۱۱۱۷

بهار ۱۳۹۷

## قسمت اول) آشنایی با مقاله پژوهش اصلی

### Terminology

BCI := brain computer interface = رابط مغز و کامپیوتر

ERP := event related potential =

سیگنال اندازه گیری شده از مغز بنا به تاثیر مستقیم یک پیش آمد حرکتی، سنسوری و یا شناختی و خودآگاه

P300 :=

که در پروسه ی تصمیم گیری نمایان می شود ERP جزئی از سیگنال

### سؤال (۱)

این پژوهش در صدد این می باشد که با استفاده از BCI مبتنی بر P300 سیستمی برای گفتار کلمات بسازد. هدف اصلی این پژوهش رسیدن به نتیجه بهینه در مدت زمان کمینه می باشد. برخلاف پژوهش های پیشین که از سابجکت مورد نظر در طی مدت زمان طولانی تری داده می گرفتند، این آزمایش تنها در ۵ دقیقه داده گیری، سیستم طبقه بندی کننده را آموزش می دهد.

در این پژوهش، با استفاده از روش بالا، و به دو روش که در ادامه خواهند آمد، توانسته اند به دقت ۱۰۰ درصد در 72.8% افراد در یک حالت و 55.3% افراد در حالت دیگر برسند.

در پژوهش های پیشین که مثلاً از سیستم motor imagery BCI استفاده شده بود کمتر از 20% افراد به دقت بیش از ۸۰ برسند.

### سؤال (۲)

داده ها از ۱۰۰ نفر سالم، بدون بیماری ALS، که پیش زمینه ای در آموزش EEG نداشته اند گرفته شده اند.

برای آموزش طبقه بندی کننده از ۲ روش برای مشخص کردن کلمات استفاده شده است. افراد برای انتخاب روش مختار بودند و ۱۹ نفر هر ۲ روش را انتخاب کردند.

کاراکترها بر روی صفحه ای تاریک ماتریسی ۶\*۶ چیده شده اند

در روش اول بر روی صفحه هر بار تنها یک کلمه روشن و خاموش می شود و داده مربوطه ضبط می شود.

در روش دوم هر حرف از تقاطع یک سطر و ستون مشخص می شود. به این صورت که ابتدا داده مربوط به یک سطر و سپس داده یک ستون ضبط می شود و نتیجه را برای حرفی که از تقاطع دو سطر و ستون بدست می آیند به حساب می آورند.

روشن و خاموش شدن کلمات و سطر و ستون ها البته بر حسب زمان خاص می باشد و رندم نیست.

در روش اول هر حرف به مدت ۶۰ ثانیه روشن و ۴۰ ثانیه خاموش می باشد. در روش دوم هر سطر یا ستون به مدت ۱۰۰ ثانیه روشن و ۶۰ ثانیه خاموش می شود.

لازم به ذکر است که روش دقیق بدین گونه می باشد که مثلاً در روش دوم قبل از اینکه سطر و یا ستون مورد نظر اندازه گیری شود، بقیه سطر ها و ستون ها همگی یک به یک روشن و خاموش می شوند.

فرکانس داده گیری ۲۵۶ هرتز می باشد.

در ابتدا به نظر می‌رسد که در روش اول که هر بار تنها یک حرف روشن می‌شود، نتیجه بهتر باشد و در خروجی بایستی سیگنال P300 با دامنه بزرگتری را مشاهده کنیم اما نتیجه مقاله بیانگر این می‌باشد که روش دوم نتیجه‌های بهتری دارد

در هنگام آموزش بنا به گفته مقاله کلمه WATER آموزش داده می‌شود و در هنگام تست کلمه LUCAS خواسته شده است. البته این گفته‌ها با داده‌های موجود همخوانی ندارند

در هنگام آموزش توسط کلمه WATER به سابجکت هیچ فیدبکی از عملکردش داده نمی‌شود اما در هنگام تست فیدبک در همان زمان آزمایش ارائه می‌شود

از ۳۵ نفر نیز نظر سنجی برای بررسی اثر پارامترهای مختلف گرفته شده است که در این نظر سنجی تنها اثر خواب کمتر از ۸ ساعت در داده‌ها نمایان می‌باشد و دیگر پارامترها از جمله جنسیت از لحاظ آماری تاثیری نداشته‌اند.

### سؤال (۳)

در این پژوهش از این اصل استفاده شده است که هرچه عملی نادرتر باشد در سیگنال P300 حاصل دامنه بزرگتری را مشاهده خواهیم کرد

از آنجایی که در صفحه‌ی روبه‌روی شخص یک صفحه شامل ۳۶ کاراکتر می‌باشد پس وقتی که فرد تنها به دنبال یک حرف خاص است، وقتی آن حرف نمایان می‌شود در سیگنال P300 حاصله دامنه بزرگی به‌وجود می‌آید. البته این دامنه با توجه به آزمایش‌های شده و مقالات پیشین در روش روشن شدن سطر و ستون بزرگتر می‌باشد و جداکننده بهتری می‌باشند، همانطور که نتیجه مقاله نیز بیانگر این می‌باشد.

در هر بار روشن و خاموش شدن، تارگت، تنها حرفی می‌باشد که انتظار آن را در آن لحظه می‌کشیم. بدین گونه که اگر چه ما مثلاً می‌خواهیم کلمه WATER را بگوییم اما وقتی که نوبت حرف W است اگر حرف A در جایی روشن شود جزو تارگت ما نیست.

بنا بر گفته مقاله، تمامی ویژگی‌های زمانی و مکانی داده‌ها به طبقه‌بند LDA داده می‌شود و این تابع به هر پارامتر با توجه به اهمیت آن وزن می‌دهد. جمع این پارامترهای وزن دار برای یک حالت، حاصل طبقه‌بندی است که مشخص می‌کند این حالت موردنظر و کاراکتر خاص تارگت می‌باشد یا نه. پس در واقع معیار ما برای جداسازی داده‌ها جمع پارامترهای خروجی طبقه‌بندی کننده LDA می‌باشد

### سؤال (۴)

هدف اصلی مقاله بررسی درصد پاسخ صحیح در جامعه آماری بزرگ می‌باشد. از بین ۳۸ نفر که در روش اول مشارکت داشتند 55.3% و از بین ۸۱ نفری که روش دوم را انتخاب کرده بودند، 72.8% به جواب کاملاً صحیح رسیدند. (۱۹ نفر در هر دو روش شرکت کردند)

همانطور که در سوال ۲ نیز گفته شد، پژوهش اثر آماری پارامترهای مختلف را نیز بررسی کرده است. این پارامترها شامل موارد زیر است:

جنسیت، تحصیلات، طول ساعت کار، سیگار، قهوه و خواب در بین پارامترهای بالا تنها اثر خواب مشهود بود و هیچ یک از پارامترهای دیگر تاثیر آماری چشمگیری نداشتند

مثلا در مورد جنسیت، مردها به میانگین 90.1% و زنان به میانگین 81.9% رسیدند که همانطور مشاهده می شود، اختلاف چشمگیر نیست.

برای مقایسه سرعت عمل، پژوهش دیگری که تصمیم در آن باینری بوده ولی از BCI مبتنی بر motor imagery استفاده کرده را در نظر میگیریم. در این پژوهش در هر دقیقه ۶ تصمیم باینری، راست و یا چپ گرفته می شود. اما در مقاله فعلی برای روش دوم که درصد صحت بهتری نیز داشت، هر ۲۸/۸ ثانیه تصمیم بین ۳۶ حالت مختلف

گرفته می شود که حاکی از سرعت عمل بالای این روش دارد

قسمت دوم) آشنایی با دیتاست

فایل هر سابجکت شامل ۱۱ سطر می باشد. سطر اول نمایانگر زمان می باشد که با فرکانس سمپلینگ ۲۵۶ هرتز اندازه گیری شده است. سطر دوم تا نهم هر یک، یکی از ۸ الکترودی می باشند که از آنها داده گرفته ایم. این الکترودها موارد زیر می باشند:

Fz, Cz, P4, Pz, P3, Po8, Oz, Po7

سطر دهم نمایانگر شماره سطر و یا ستونی میباشد که روشن می باشد و سطر ۱۱ بیانگر این است که آیا در سطر و یا ستونی که اکنون روشن می باشد حرفی از کلمه ی مورد نظر ما وجود دارد یا نه (منظور از کلمه مورد نظر، کلمه ی LUCAS می باشد و نه کلمه ی آموزش WATER). در صورت مثبت بودن جواب، عدد یک و در غیر این صورت عدد صفر در آن ستون مذکور از این سطر جای میگیرد.

در مورد سطر ۱۰ ام داده ها توضیحات زیر لازم می باشد:

شماره گذاری سطر ها و ستون ها از عدد ۱ تا ۱۲ می باشد. اعداد ۱ تا ۶ بیانگر شماره ستون با شروع از سمت چپ و اعداد ۷ تا ۱۲ بیانگر شماره سطر ها با شروع از بالا به پایین می باشد.

اگر از روش سطر یا ستون، RC استفاده شده باشد مشکلی در نمایش سطر و یا ستون مربوطه نیست اما اگر از روش کاراکتر تنها، SC استفاده شده باشد، عدد موجود از آنجایی که باید تنها یک کاراکتر را مشخص نماید، اگر بخواهیم از همین روش شماره سطر و ستون استفاده کنیم، بایستی هم شماره سطر و هم شماره ستون مربوطه آن را بدهیم و یا درایه ها را به ترتیبی بشماریم.

با بررسی های انجام شده معلوم شد که اولاً تنها ۲ مورد از ۱۰ سابجکت از روش SC استفاده کرده اند (سابجکت ۱ و ۲) و ثانياً روش شماره گذاری به شماره گذاری درایه ها بوده است و در واقع به ترتیب حروف انگلیسی شماره گذاری شده اند (شماره گذاری سطری). شکل زیر شماره گذاری را در هر ۲ حالت نشان میدهد:

1	2	3	4	5	6
7 A 1	B 2	C 3	D 4	E 5	F 6
8 G 7	H 8	I 9	J 10	K 11	L 12
9 M 13	N 14	O 15	P 16	Q 17	R 18
10 S 19	T 20	U 21	V 22	W 23	X 24
11 Y 25	Z 26	0 27	1 28	2 29	3 30
12 4 31	5 32	6 33	7 34	8 35	9 36

## قسمت سوم) بررسی ERP به روش سنتی

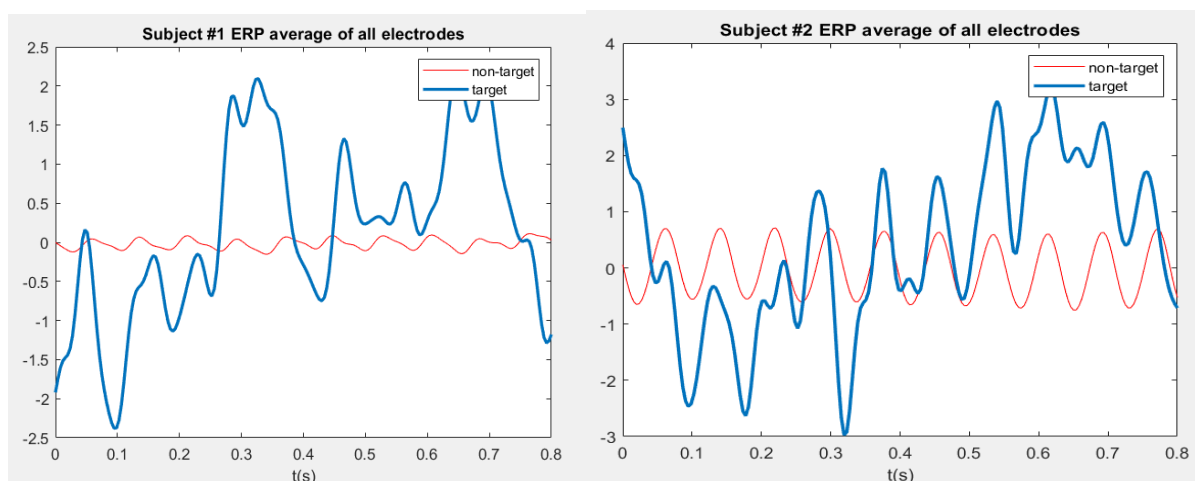
## سوال ۳، ۴ و ۵

برای ۲ فرد اول که آنها از روش SC و یا همان روشن شدن تک حرف استفاده کرده اند، خروجی non-target کاملاً نوسانی و حتی یک سیگنال پریودیک با میانگین صفر می‌باشد و دامنه‌ی این سیگنال‌ها نیز در مقایسه با target ها کمتر می‌باشد. البته در این سیگنال ها اگرچه سیگنال Non-target یک سیگنال نوسانی با دامنه کمتر می‌باشد ولی با توجه به شکل های موجود، به نظر می‌رسد که تمایز بین target و non-target به این آسانی میسر نباشد چرا که شکل ها در هم تنیده می‌باشند و نمیتوان مرزی را مشخص کرد که با آن بتوان داده ها را به خوبی جدا کرد. که این می‌تواند علت کم بودن درصد تشخیص درست داده ها در روش اول-SC- باشد.

Target ها در دو نفر اول اگرچه پریودیک نمی‌باشند اما به طور پیوسته در حال نوسان اند

شکل های زیر میانگین کل ۸ تصویر برای هر یک از افراد می‌باشد. میتوان با استفاده از این شکل ها معیاری نسبی از زمان مناسب برای تمایز بین تارگت و غیرتارگت بدست آوریم

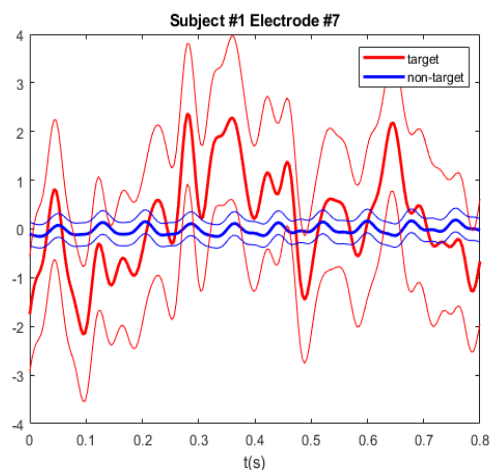
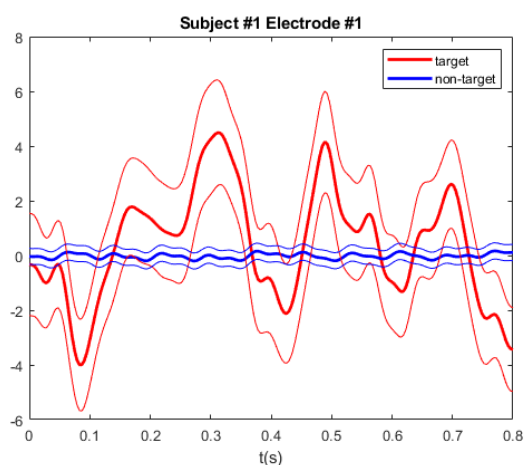
ابتدا ۲ نفر اول که از روش SC استفاده کرده‌اند را بررسی کنیم.



برای شخص اول، به نظر می‌رسد که ۳ زمان ۰/۱، ۰/۳۵ و ۰/۷ ثانیه نمایانگر بزرگترین تفاوت بین داده های تارگت و غیرتارگت می‌باشد و برای جداسازی تارگت می‌توان از این ۳ زمان بهره برد. البته با توجه به ماهیت سیگنال P300، زمان ۰/۷ دور از انتظار می‌باشد و در زمان ۰/۱ شاید اثر تصمیم هنوز پخش نشده باشد (propagate نشده باشد) به همین دلیل، بهتر است زمان ۰/۳۵ را انتخاب کنیم.

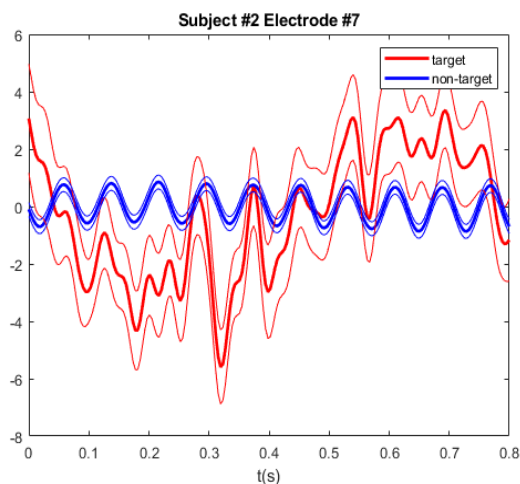
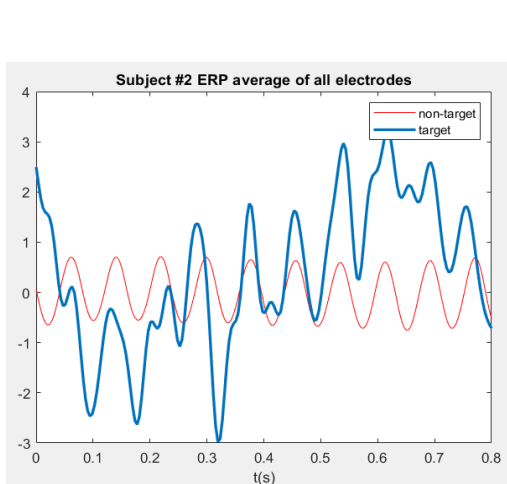
برای انتخاب الکتروود نیز با توجه شکل های موجود، الکتروود شماره ۱ انتخاب بهینه می‌باشد. برای این انتخاب از معیار بیشینه بودن جمع اختلاف توان ۲ های داده ها نیز استفاده شده است یعنی  $\max(\sum (y-x)^2)$

در زیر نمودار الکتروود اول و الکتروود هفتم آمده است. همانطور که مشاهده می‌شود با توجه به در هم تنیدگی داده ها در کانال ۱۷ جداسازی داده ها در این کانال سخت تر می‌باشد.

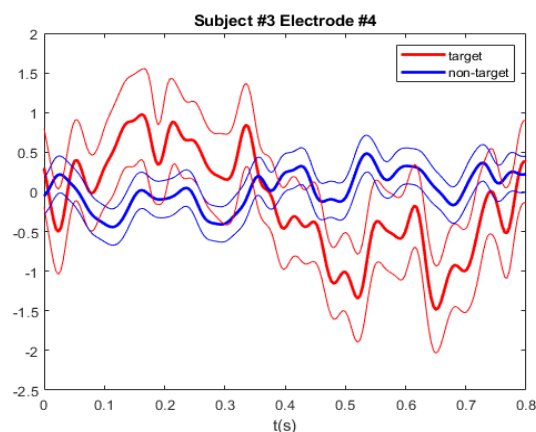
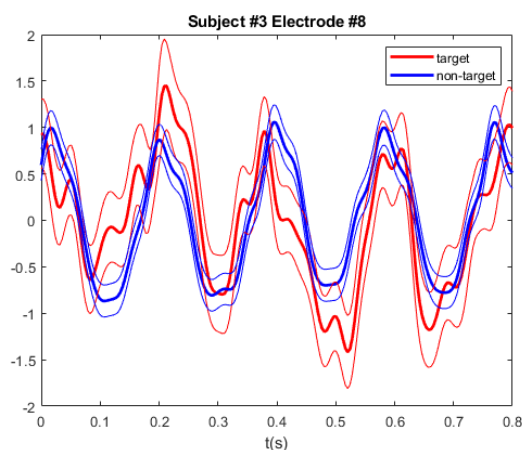
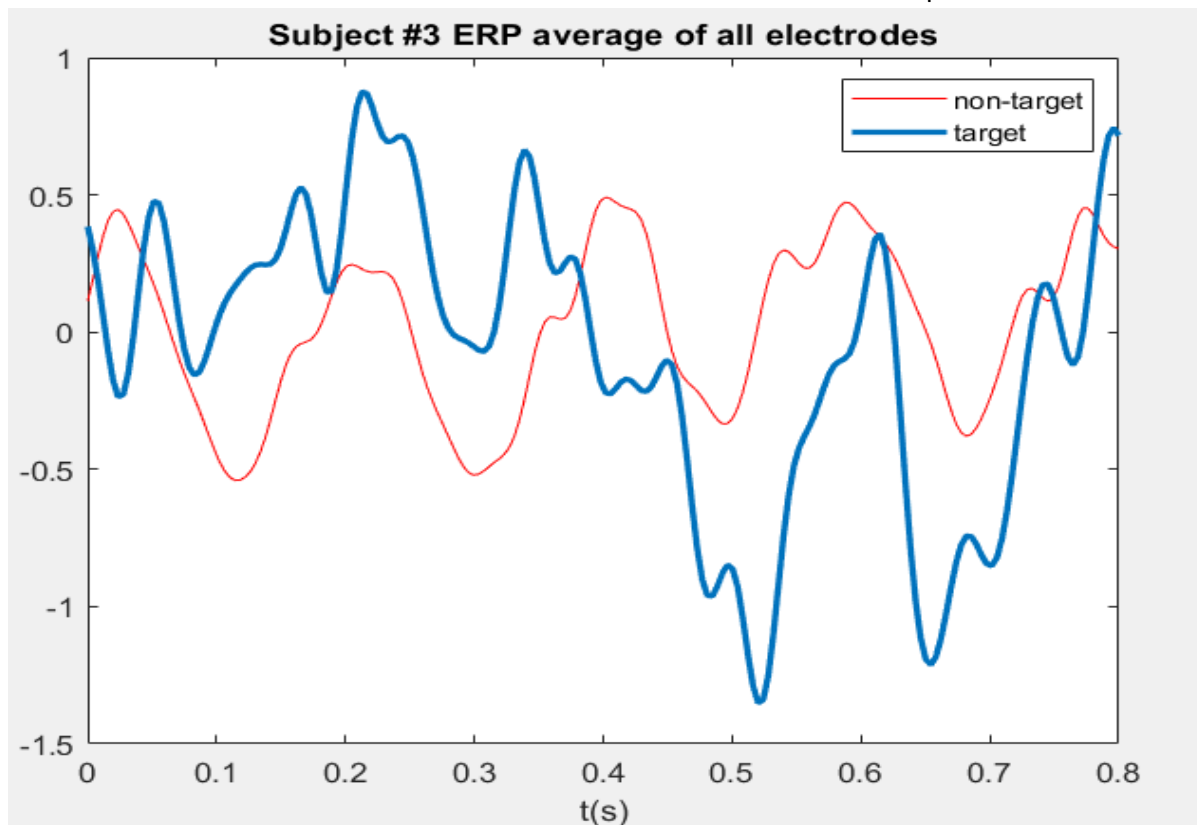


برای شخص دوم نوسانات میانگین کل داده ها بیشتر از شخص اول می باشد. زمان مناسب جهت بررسی، با توجه به ماهیت P300، ۰/۳ ثانیه انتخاب می شود.

برای این شخص نیز با همان معیارهایی که برای شخص اول معرفی شد، میتوان کانال ۷ام را انتخاب کرد.

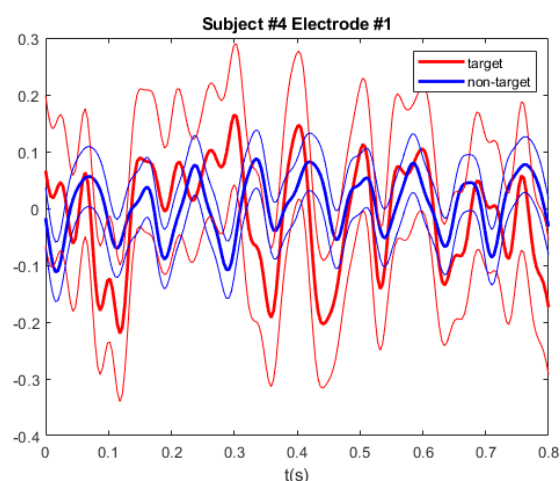
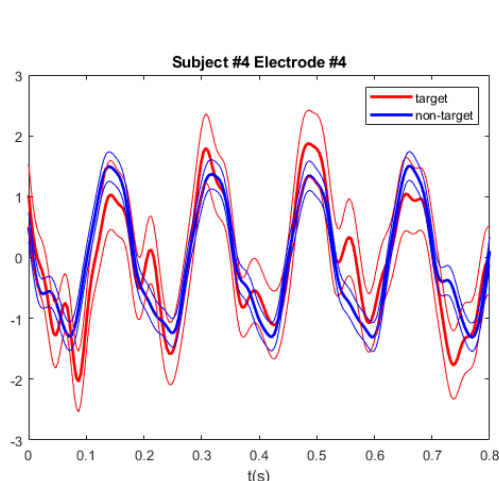
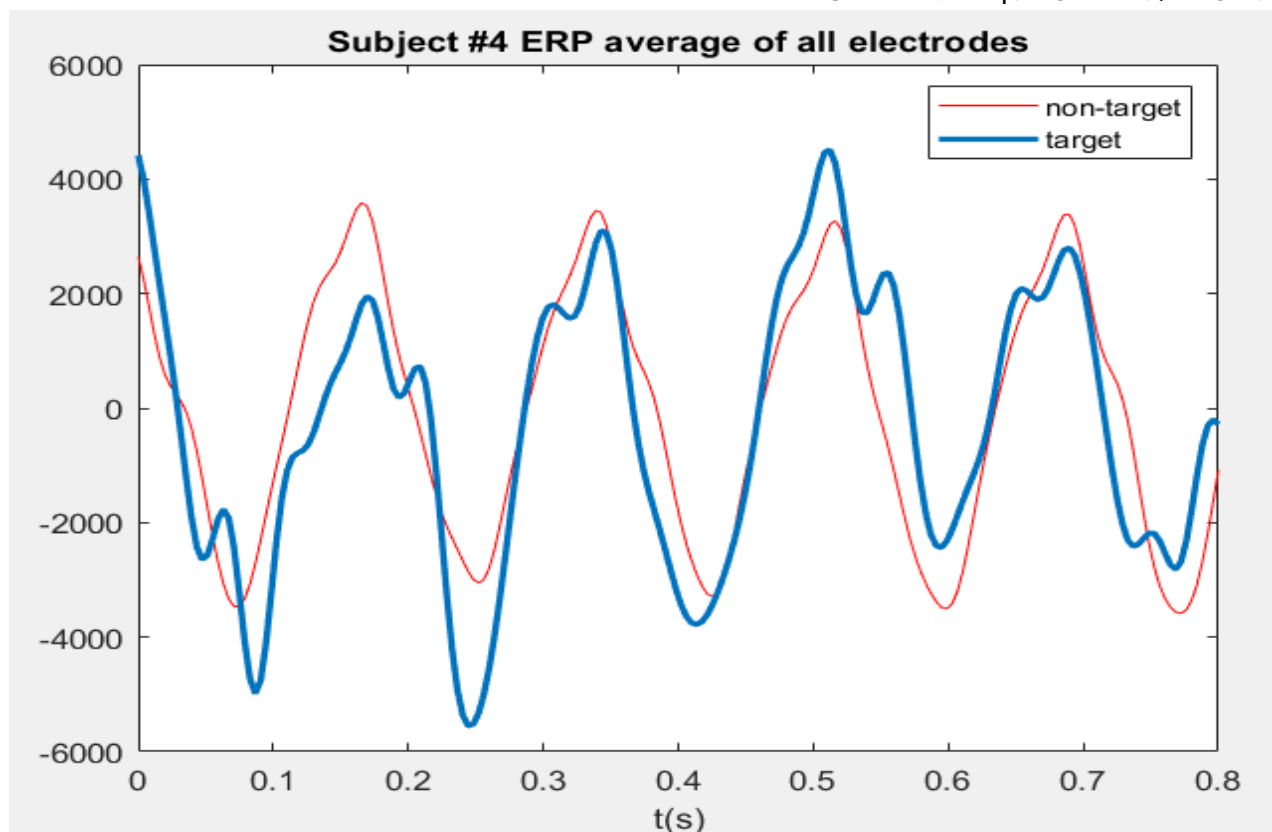


برای شخص سوم زمان ۰/۲ بهینه می باشد. برای این شخص معیار جمع مجذور اختلافات برای کانال های ۱ و ۴ نزدیک می باشد. با توجه به معیارهای شکلی کانال ۴ بهتر می باشد و کانال ۸ بدترین. داده های غیرتارگت رفتار نوسانی کمی از خود نشان می دهند



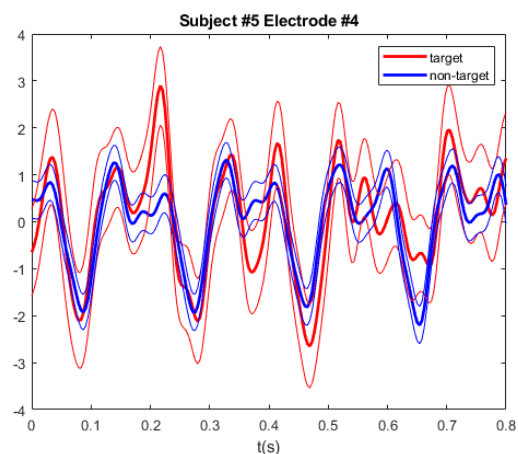
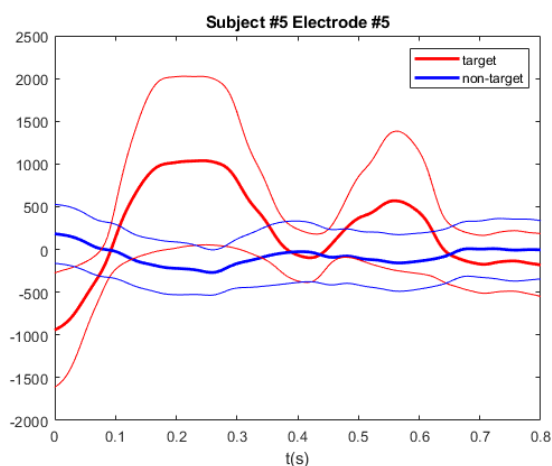
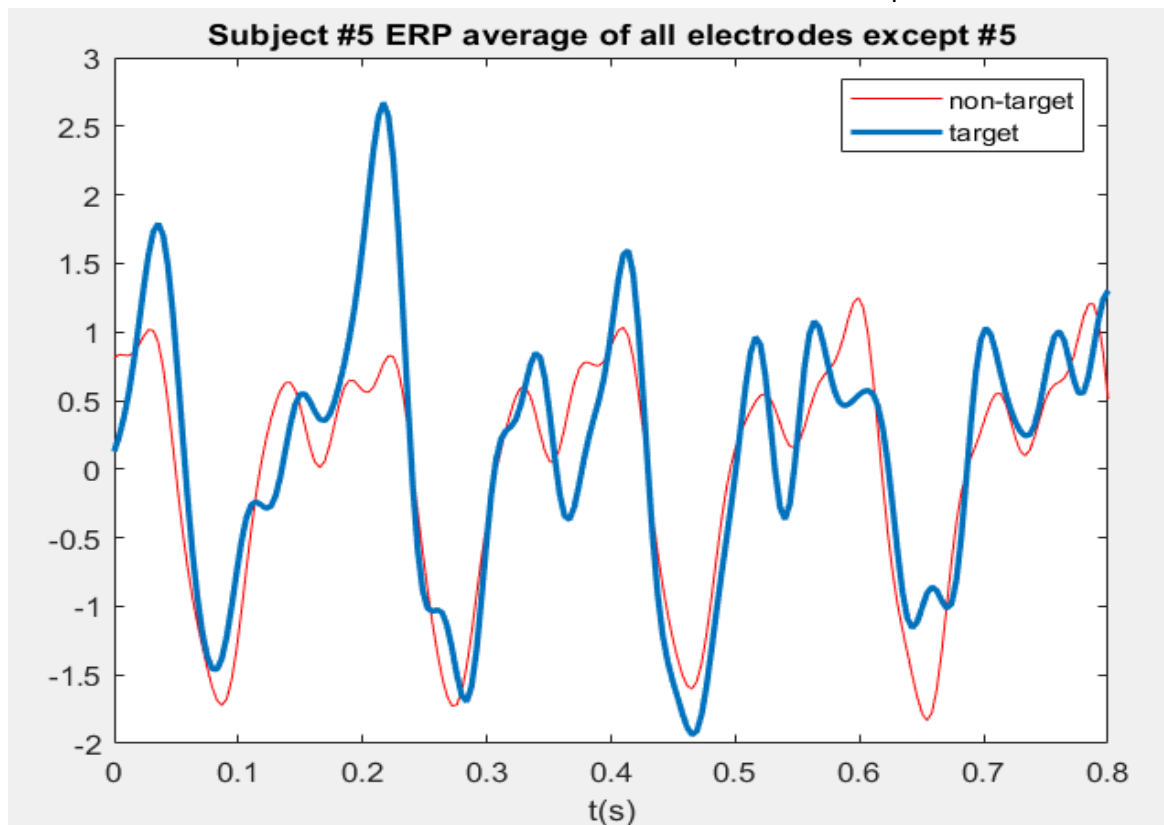
برای شخص چهارم ، برخلاف بقیه، داده ها بسیار در هم تنیده شده اند و تمیز بین داده ها بسیار سخت می باشد. همچنین داده های کانال های ۷ و ۸ بسیار بزرگ می باشند که به نظر اشتباه اندازه گیری می باشد. به همین دلیل ارائه زمانی یکتا برای جداسازی تارگت از غیر آن برای این شخص میسر نمی باشد. برای مثال کانال های ۱ و ۴ به شکل زیر می باشند برای این سبجکت، رفتار نوسانی هم در داده های تارگت و هم در داده های غیرتارگت قابل مشاهده است و این ۲ سری داده با یکدیگر نزدیکی زیادی دارند.





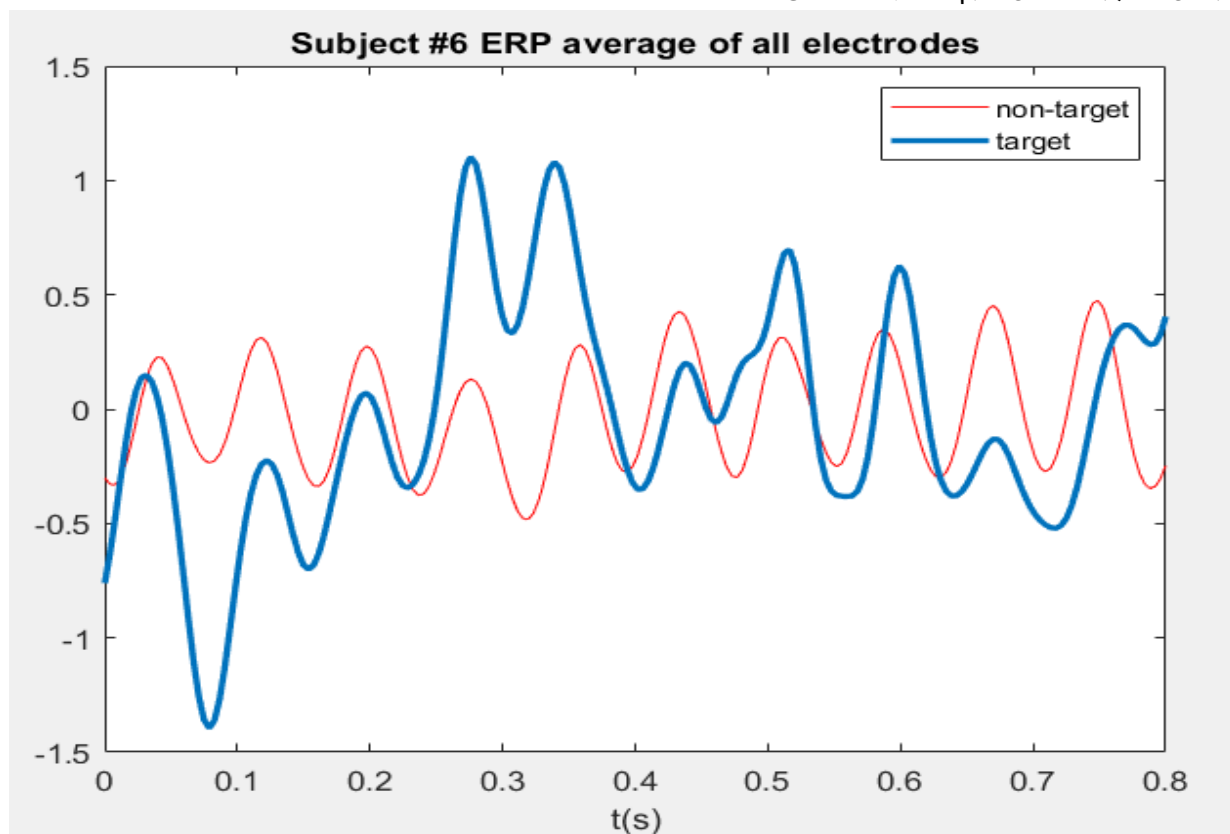
کانال ۵ ام سیگنال شخص 5 ام برخلاف بقیه سیگنال ها، حالت نوسانی ندارد و اندازه بسیار بزرگی دارد که شاید نوعی artifact باشد. البته این که اگر این سیگنال دارای artifact می باشد چرا در بقیه کانال ها چیزی مشاهده نمی شود جای سوال است. در هر حال، این کانال از مقایسه خارج شده است. با این وجود نیز هنوز نمی توان برای جداسازی زمان واحدی را ارائه کرد. البته در زمان ۰/۲ اندکی مقدار سیگنال تارگت بزرگتر می باشد ولی با توجه به بازه اطمینان، به درد نمی خورد. رفتار نوسانی غیرتارگت ها اگرچه موجود است ولی کم می باشد.

در زیر نمودار کانال ۵، که به نظر می رسد artifact داشته باشد، و نمودار کانال ۴ آمده اند

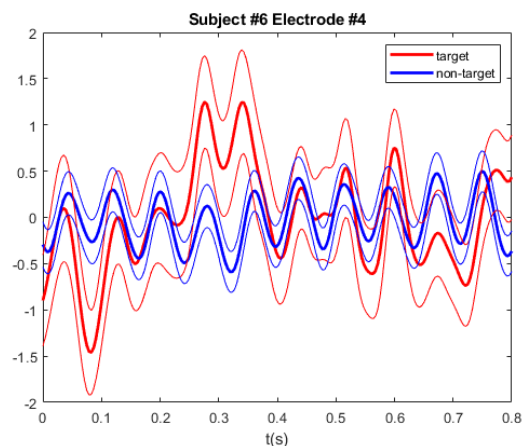
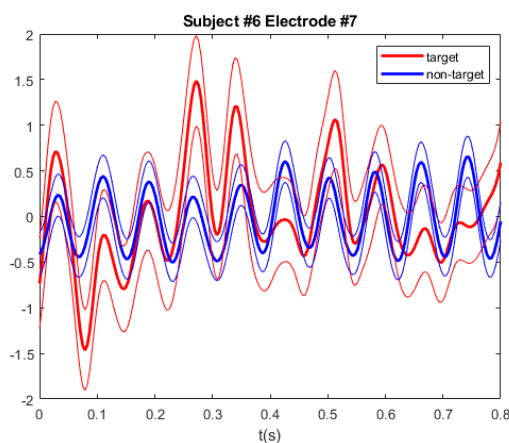


البته اگر کانال ۵، artifact نباشد برای جداسازی داده ها بهترین جواب می باشد

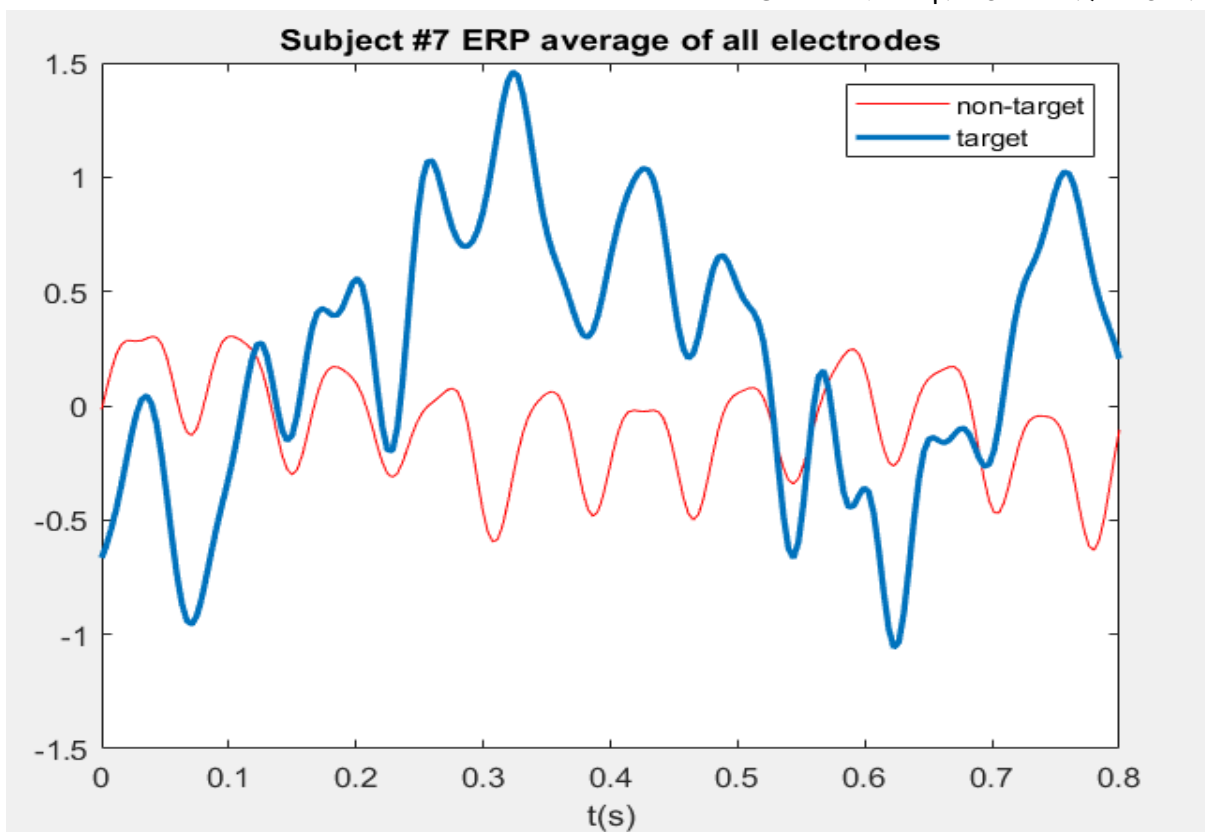
رفتار نوسانی برای سیگنال های غیرتارگت شخص ۶ کاملاً واضح می باشد. برای جداسازی نیز بازه زمانی حول ۰/۳ مناسب می باشد



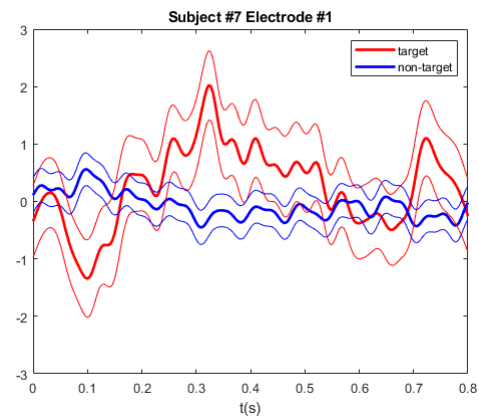
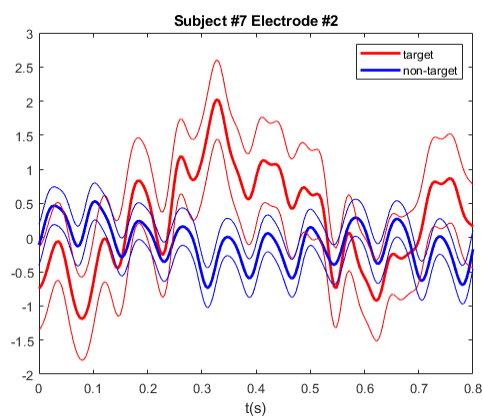
برای این شخص، تمامی الکترودها در بازه زمانی کوچک حول  $0/3$  ثانیه رفتار به نسبت مشابهی دارند به گونه ای که برای همه ی آنها بازه اطمینان داده های تارگت هیچ تداخلی با غیر تارگت ها ندارند و نمی توان الکترودی یکتا فقط بر حسب ویژگی هایی که تا به حال داریم انتخاب کرد. برای مثال در زیر 2 الکترودها 4 و 7 آمده اند که به نوعی بیشترین تفاوت را داشته اند اما با این وجود بسیار شبیه به هم می باشند.



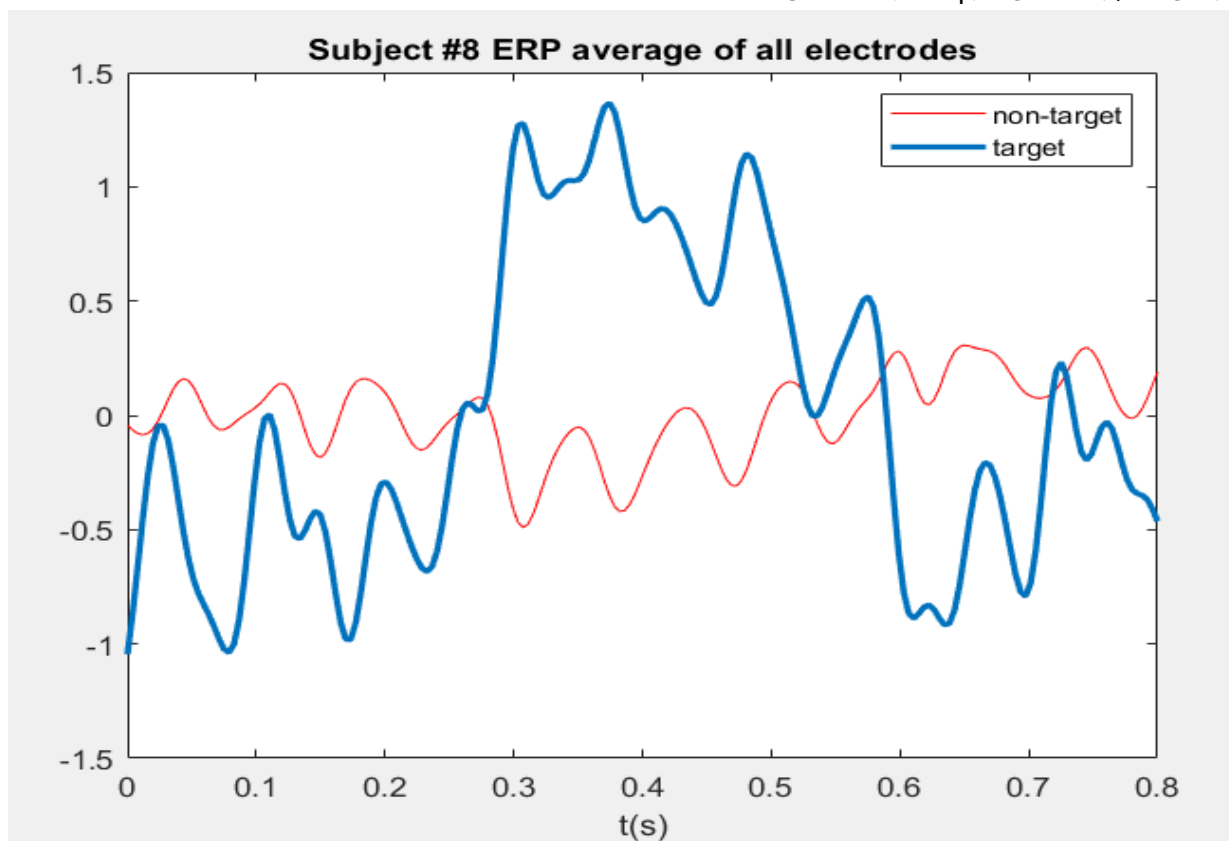
میانگین کل داده ها برای فرد 7 نشانگر این است که می توان از بازه  $0/25$  ثانیه تا  $0/5$  ثانیه برای جداسازی تارگت و غیرتارگت استفاده کرد



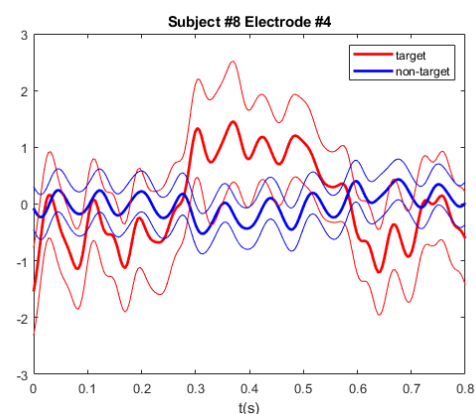
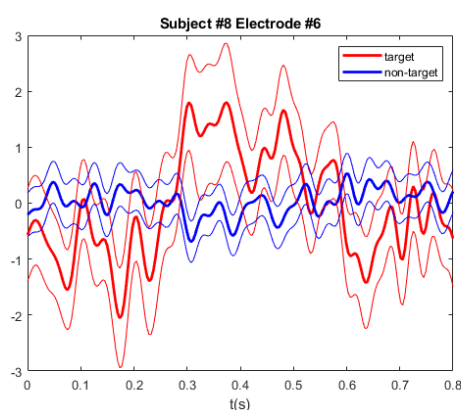
داده های غیرتارگت در این فرد نیز به صورت خوبی نوسانی می باشند. برای جداسازی میتوان از هر ۲ کانال ۱ و ۲ که به صورت خوبی تمایز بین داده های تارگت و غیرتارگت در آنها مشخص استفاده کرد که در زیر آمده اند.



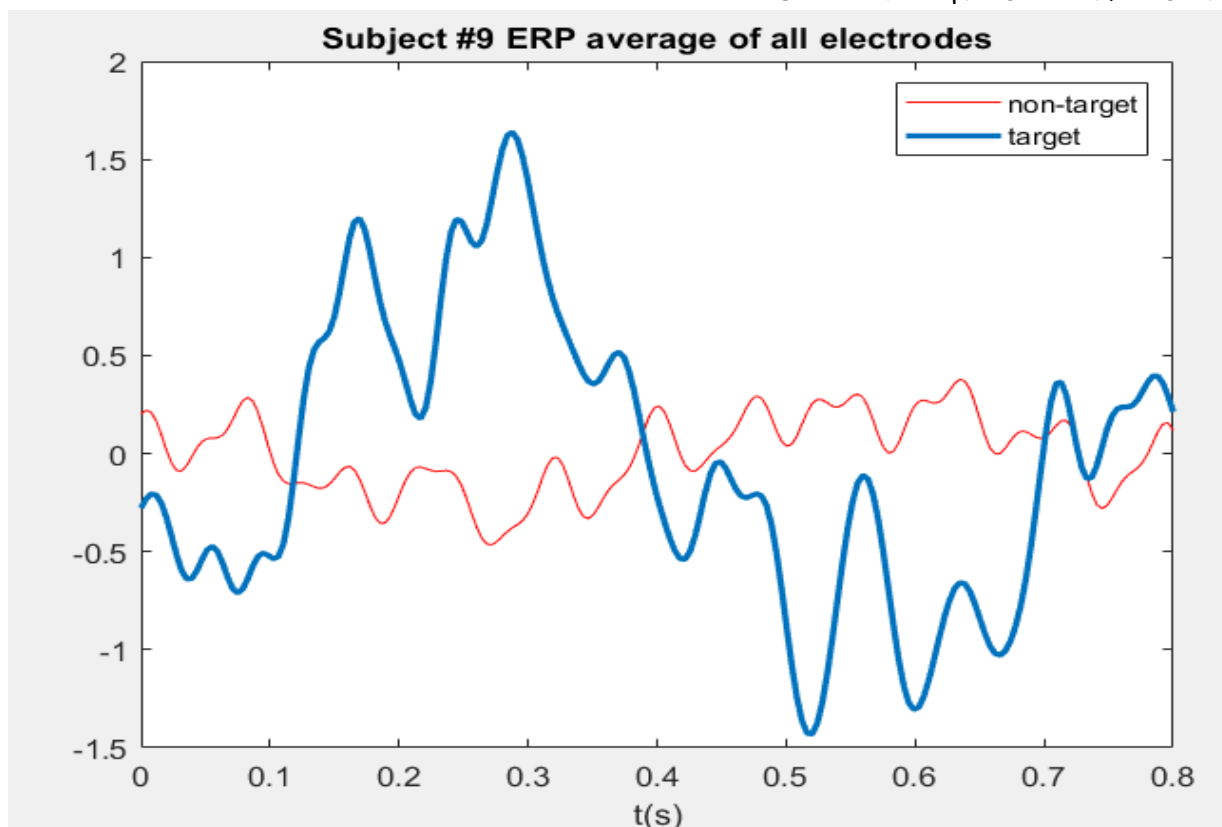
برای شخص ۸م نیز بازه زمانی ۰/۳ تا ۰/۵ ثانیه برای جداسازی داده ها مناسب می باشد. همچنین نوسانی بودن داده های غیرتارگت معلوم می باشد



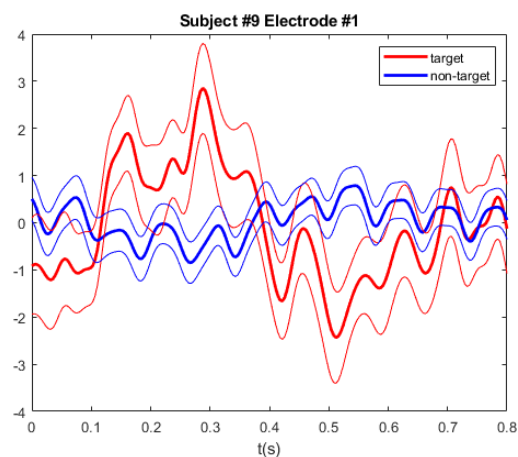
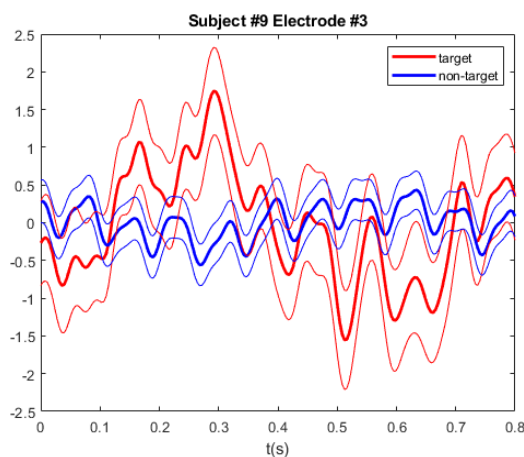
برای جداسازی داده ها استفاده از کانال های ۴ و ۶ می تواند مناسب باشد. در کانال ۶ اختلاف بزرگتری را شاهد هستیم و در کانال ۴ بازه ی اطمینان به صورت بهتر از داده های غیر تارگت جدا شده است.



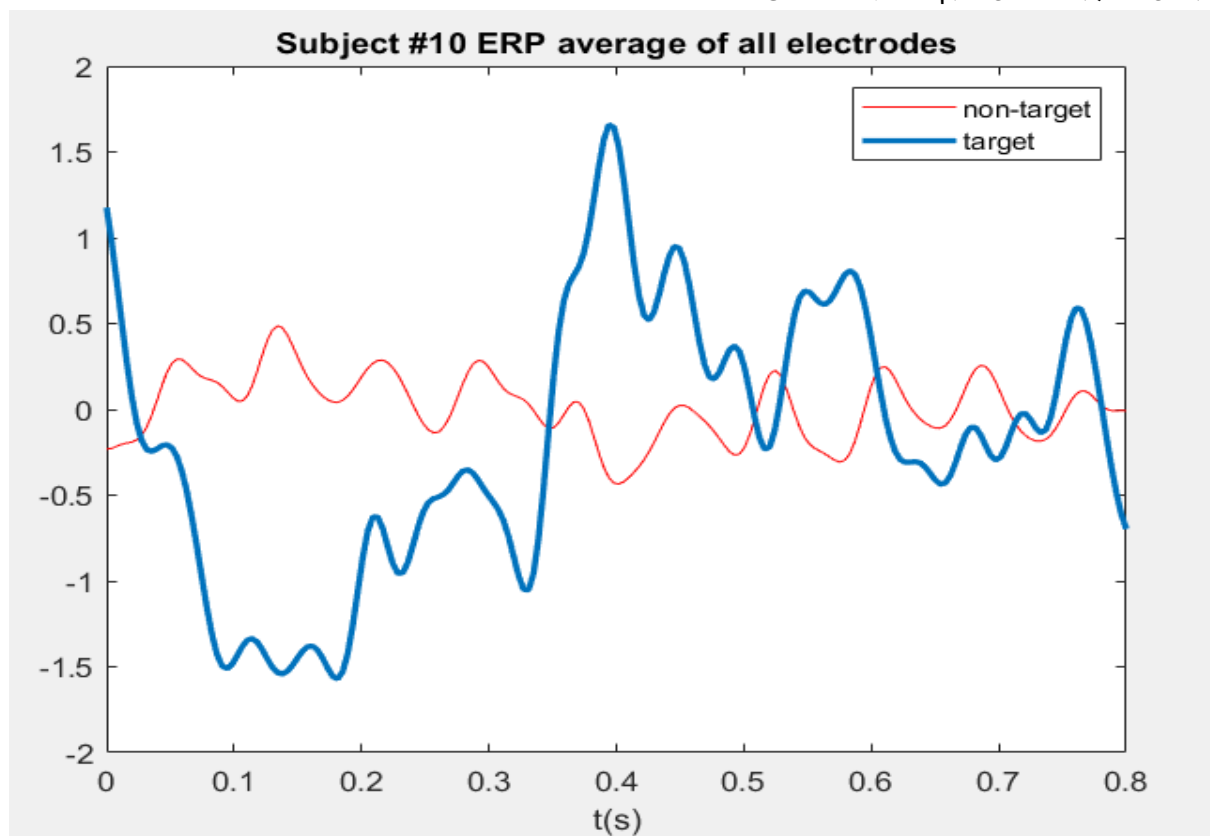
برای فرد شماره ۹ زمان ۰/۳ بهتر می باشد. ( البته بازه ۰/۱ تا ۰/۳ به نوعی بهینه می باشند اما با توجه به تاخیر در انتشار امواج P300، زمان ۰/۳ انتخاب می شود.) نوسانی بودن داده های غیرتارگت نسبت به ۲ سابجکت قبل کم تر است.



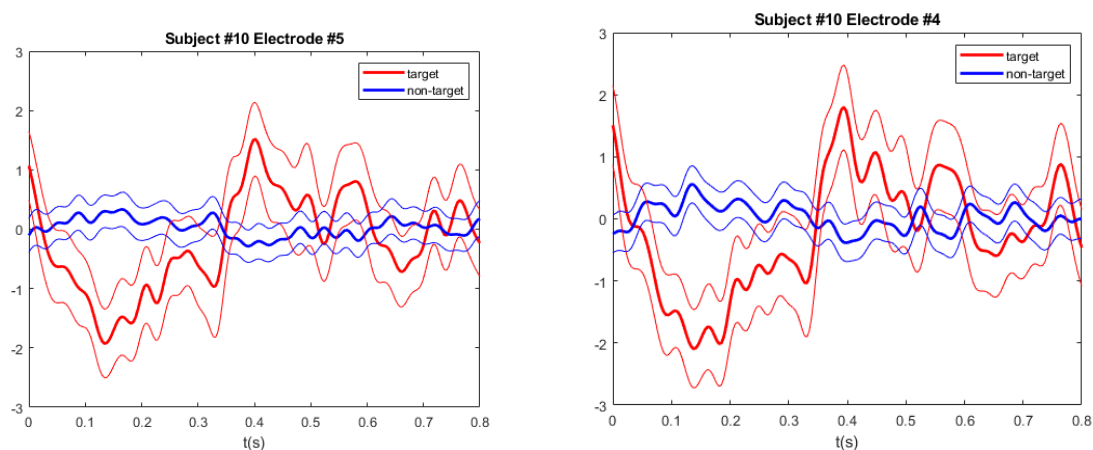
برای انتخاب الکترود مناسب، کانال ۱ و ۳ بهترین انتخاب ها هستند.



رفتار نوسانی غیرتارگت ها برای فرد ۱۰ نیز کم می باشد. برای این فرد میتوان از ۲ نقطه زمانی برای تمیز بین داده ها استفاده کرد. در بازه نزدیک قبل ۰/۳ ثانیه داده های تارگت از لحاظ اندازه منفی و قابل تمایز با غیرتارگت هستند. در بازه نزدیک ۰/۴ نیز داده های تارگت با اندازه های بزرگتری نسبت غیرتارگت ها قابل جداسازی هستند.



کانال های مناسب برای جداسازی داده های نیز الکترودهای ۴ و ۵ می باشند



با توجه به توضیحاتی که در بالا آمد و برای هر سابجکت زمان نسبی مناسب و الکترودهای خوب را معرفی کردیم، معلوم می شود که زمان مناسب برای افراد مختلف است البته با توجه به سیگنال P300 همه این زمان ها در بازه ۲۰۰ تا ۵۰۰ میلی ثانیه می باشند الکترودهای مناسب نیز برای افراد متفاوت بودند البته الکترودهای ۱، ۴ و ۷ بیشترین استفاده را در بین الکترودها داشتند و چندین بار تکرار شده اند

نوسانی بودن داده های غیرتارگت برای روش SC ، سابیجت اول و دوم، کاملاً مشخص و حتی سیگنالی پریودیک بود. در بقیه سابیجت ها رفتار نوسانی داده های غیرتارگت اگرچه موجود بود اما عموماً کمتر بود. البته برای سابیجت شماره ۴، علاوه بر سیگنال غیرتارگت، سیگنال تارگت نیز به صورت کاملاً نوسانی بود و داده های این سابیجت را نمی توانستیم به این راحتی جدا کنیم.

سابیجت شماره ۵ نیز رفتاری نوسانی دارد، هم سیگنال تارگت هم سیگنال غیرتارگت. و الکتروود شماره ۵ نیز به نظر می رسد که artifact داشته باشد.

از آنجایی که از پارادایم SC ما تنها ۲ سابیجت در دست داریم تصمیم گیری سخت می باشد و نیاز به داده ی بیشتری هست. جداسازی داده های سابیجت ۲ سخت تر بود ولی برای سابیجت شماره ۱ مشابه خیلی از حالت های دیگر سابیجت ها شاهد بودیم که سیگنال تارگت در بازه ای از زمان با بازه ی اطمینان مناسب از غیرتارگت ها قابل جداسازی است. یکی از خوبی های روش SC این است که سیگنال های غیرتارگت در این روش به صورت پریودیک و با دامنه به نسبت کم ظاهر می شوند. بنابراین به نظر می رسد که ادعای مقاله چندان قابل اطمینان نیست

میتوان انتظار داشت که سابیجت های ۱، ۴ و ۵ نتیجه ی چندان خوبی نگیرند. در عوض سابیجت های ۶، ۷، ۸، ۹ و ۱۰ نتیجه خوبی بگیرند.

## سوال ۶

ادعای مقاله مبنی بر بهتر بودن پارادایم SC لزوماً درست نیست. همچنین آن چه که از نظر زمان مناسب برای جداسازی از قبل در نظر داشتیم، تقریباً در این جا نیز تأیید می شود و مفهوم P300 خود را نشان می دهد. در مورد شماره الکتروودهای جداکننده نیز دیدیم که در رأس، شماره ۱ و ۴ و ۷ حضور داشتند. البته یک نکته مهم که باید به آن توجه کرد، تعریف جداکنندگی در این شرایط است. در واقع جداکنندگی با تعبیر دقیق تری در قسمت بعد بررسی می شود، اما می توان ایده های مختلفی برای فاصله دو تابع زمانی در نظر گرفت که می تواند شامل correlation و نیز مجموع (انتگرال) مجذور اختلاف آنها باشد.



**قسمت چهارم) پیاده‌سازی الگوریتم P300-Speller**

به دلیل آن که بخش‌های ۱ تا ۳ این سؤال را به صورت موازی برای همه‌ی سابجکت‌ها انجام دادیم، آن‌ها را در کنار هم گزارش می‌کنیم.

**سؤال ۲)**

Fields	Train_Percentage	Test_Percentage	CrossVal
1	98.1852	96.0741	95.6296
2	97.9259	96.2963	95.2222
3	94.5556	71.3333	68.0000
4	92.5556	69.4444	68.8889
5	94.2222	66.6667	68.8889
6	94.1111	68.1111	67.8889
7	92.1111	68	66.4444
8	93	69	66.6667
9	94.3333	69.4444	69.2222
10	93.6667	69.3333	70.0000

ابتدائاً مشاهده می‌شود که درصد پاسخگویی طبقه‌بندی‌کننده روی داده‌های Train همواره بالای ۹۰٪ و در برخی موارد بسیار نزدیک به ۱۰۰٪ است، که این امری طبیعی است؛ چرا که از همین دیتا برای ایجاد طبقه‌بندی‌کننده استفاده شده است. در واقع، چیزی که می‌تواند عجیب باشد آن است که چرا این مقادیر همگی برابر ۱۰۰٪ نیستند، که در پاسخ می‌توان گفت که طبقه‌بندی‌کننده‌ها، به خصوص طبقه‌بندی‌کننده‌های خطی، در بسیاری از موارد نمی‌توانند داده‌های کلاس‌های مختلف را به تمامی از هم جدا کنند، و با توجه به ماهیت توزیع داده‌ها، ممکن است جدا کردن آن‌ها (مثلاً عبور دادن یک خط از میان آن‌ها) اصولاً ممکن نباشد. در این حالت، طبقه‌بندی‌کننده بهترین انتخاب ممکن را ارائه می‌دهد.

در مورد پاسخ به داده‌های تست نیز درصد‌های قابل قبولی را مشاهده می‌کنیم، که همگی نزدیک به ۷۰٪ و یا بیشتر هستند، اما چیزی که باید مورد توجه قرار بگیرد، آن است که این درصد‌ها الزاماً خبر از موفقیت ما در ساختن speller نمی‌دهند، چرا که این درصد‌ها نشان‌دهنده برآیند داده‌هایی هستند که درست تشخیص داده شده‌اند، در حالی که همه‌ی داده‌های درست تشخیص داده شده به یک اندازه اهمیت ندارند. در واقع چیزی که مهم است، داده‌هایی است که به عنوان target تشخیص داده می‌شوند. مهم‌ترین چیز، کمبود خطا در تشخیص این داده‌ها است. یعنی اگر یک داده‌ی target به اشتباه non-target تشخیص داده شود، ضرر کمتری به عملکرد speller می‌رسد تا زمانی که داده‌ی non-target به عنوان target دسته‌بندی شود. (علت این امر نیز روشن است، چرا که حضور دیتای اشتباه در دسته target به تخمین ما بایاس می‌دهد)

همین مسأله به وضوح توجیه می‌کند که چرا درصد مربوط به داده‌های پارادایم SC تفاوت معنی‌داری با RC دارد، چرا که در داده‌های SC تعداد قابل توجهی از داده‌ها non-target هستند، و این باعث می‌شود که تشخیص فراوان آن‌ها درصد صحت پاسخگویی کل را بالا ببرد.

بنابراین، شاید مهم‌تر از تمامی درصد‌های فوق، درصد داده‌های target درست تشخیص داده شده، در میان کل داده‌هایی باشد که به عنوان target شناسایی شده‌اند. در ادامه این درصد‌ها را نیز مشاهده می‌کنید:

Test\_Percentage =

10.8108    12.0000    20.9677    16.9312    15.2778    17.8404    15.5000    15.1351    16.9312    20.2830

این درصدها از چپ به راست مربوط به سابجکتهای ۱ تا ۱۰ هستند. همان طور که مشاهده می‌شود، درصدهای فوق چندان امیدوارکننده نیستند، چرا که نشان می‌دهند چیزی در حدود بیش از ۸۰٪ داده‌هایی که به عنوان target شناسایی می‌شوند، در واقع non-target هستند و این یعنی احتمالاً speller ما نمی‌تواند عملکرد خیره‌کننده‌ای داشته باشد.

همچنین نکته‌ای که در مورد درصد بالای داده‌های پارادایم SC در جدول صفحه قبل بیان شد، مجدداً در این جدول نیز خود را نشان می‌دهد؛ و می‌بینیم که درصد این دو سابجکت نسبت به سایرین به شکل معنی‌داری پایین‌تر است. این مسأله، مؤید صحبت قبلی ما در مورد اثر زیاد بودن داده‌های non-target در طبقه‌بندی کننده، و نیز این نتیجه‌ی احتمالی که "عملکرد RC بهتر از SC خواهد بود" می‌باشد.

نهایتاً در مورد درصدهای حاصل از 5-fold Cross Validation نیز، قضاوت شبیه به چیزی است که در مورد Test\_Percentage گفته شد. در واقع عملکرد Cross Validation شبیه‌سازی حضور دیتای تست است که می‌بینیم درصدهای ناشی از این دو نیز بسیار به هم نزدیک هستند و همان توجیهاتی که در مورد درصدهای پاسخ‌گویی دیتای تست، و نیز عدم اعتبار آن‌ها برای مسأله‌ای که با آن مواجه هستیم بیان شد؛ دقیقاً برای درصدهای Cross Validation نیز صادق است.

## سؤال (۳)

در این بخش، با وجود تلاش بسیار زیادی که کردیم، متأسفانه به نتیجه‌ی جالبی نرسیدیم. در شکل زیر، دو نمونه از خروجی‌هایی که طی روش‌های متعدد به آن دست یافتیم گزارش می‌شود:

abc trainWord	abc testWord	abc predictedWord2	abc predictedWord
'LUKAS'	'LUKAS'	'FEOPM'	'FEOPM'
'LUKAS'	'LUKAS'	'CTHJC'	'CTHJC'
'LUKAS'	'LUKAS'	'CO1ML'	'CORMK'
'LUKAS'	'LUKAS'	'5SP8C'	'5SP8C'
'LUKAS'	'WATER'	'AXQEB'	'4XQAA'
'LUKAS'	'WATER'	'XYWLH'	'W1WLN'
'LUKAS'	'WATER'	'9LTHS'	'LL5GS'
'LUKAS'	'WATER'	'M9KKT'	'M9KKV'
'LUKAS'	'WATER'	'BFDQY'	'BFDQY'
'LUKAS'	'WATER'	'SNWM4'	'4NWM4'

با توجه به این که خروجی‌های مطلوبی به دست نیاوردیم، در ادامه هر آن چه که انجام دادیم و ایده‌هایی که به کاربردی را ذکر می‌کنیم (تا معلوم شود که سطحی از روی این مسأله نگذاشته‌ایم!):

ابتدا باید مدل LDA را درست می‌کردیم. برای این کار، تابعی به عنوان FeatureExtraction نوشتیم که درون آن، بسیار شبیه به TrialExtraction است، با این تفاوت که داده‌های target و non-target را از هم جدا نمی‌کند، بلکه آن‌ها را با همان الگوی تابع TrialExtraction اما با ترتیب زمانی درست به عنوان خروجی می‌دهد. علت این امر آن است که در ادامه و برای شناسایی سطر/ستون/کاراکتر نشان داده شده، به اندیس زمانی تحریک نیاز خواهیم داشت.

پس از ساختن مدل LDA، نوبت به شناسایی کلمه spell شده با استفاده از طبقه‌بندی داده‌های train می‌رسد. برای این کار، تابعی به عنوان WordRecognizer نوشتیم. البته به دلیل تعدد روش‌هایی که آزمایش کردیم، دو تابع WordRecognizer و WordRecognizer2 در کنار این فایل ضمیمه شده است که محتوای هر دو در ادامه توضیح داده می‌شود.

تابع WordRecognizer2 خروجی‌های حاصل از پیش‌بینی انجام شده توسط مدل LDA و ماتریس داده‌های تست را در نظر می‌گیرد. این خروجی‌ها شامل صفر و یک‌هایی هستند که به معنی non-target و target هستند. ابتدا با استفاده از اندیس این صفر و یک‌ها، زمان متناظر وقوع هر یک را پیدا می‌کند. این تابع، برای شناسایی حرف بیان شده، روی تمامی زمان مربوط به نمایش آن حرف، مُد می‌گیرد. این بدان معنی است که برای دیتای SC، در بازه‌های ۵۴۰ درایه‌ای و روی حروف نشان داده شده (که با اعداد ۱ تا ۳۶ شناخته می‌شوند) مُد می‌گیرد و حرف نهایی را به عنوان جواب اعلام می‌کند. این فرایند برای دیتای RC روی بازه‌های ۱۸۰ تایی صورت می‌گیرد، به این صورت که یک بار روی اعداد ۱ تا ۶ موجود در بازه، و بار دیگر بر روی اعداد ۷ تا ۱۲ موجود در بازه مُد گرفته می‌شود، سپس از تلاقی سطر و ستون مُد، حرف اسپل شده شناسایی می‌شود.

تابع WordRecognizer عملکردی مشابه تابع قبلی دارد، با این تفاوت که بیشتر وارد جزئیات می‌شود. به عبارت دیگر، بازه‌های ۵۴۰ تایی مربوط به دیتای SC، متشکل از ۱۵ بازه‌ی ۳۶ تایی است که به معنی آن است که هر حرف، ۳۶ بار روشن می‌شود. این تابع، هر بازه ۵۴۰ تایی را به ۱۵ قسمت ۳۶ تایی تقسیم کرده، در هر قسمت به دنبال حرف مورد نظر می‌گردد، و نهایتاً روی خروجی مُد می‌گیرد. برای دیتای RC، مسأله کمی پیچیده‌تر است، چرا که این بار باید هر بازه ۱۸۰ تایی به ۱۵ بازه ۱۲ تایی تقسیم شود، که ممکن است در یک بازه، به عنوان مثال صرفاً یک سطر تشخیص داده شده باشد و اثری از ستون نباشد. در این قسمت، تابع یک ماتریس

۶\*۶ در نظر می‌گیرد، و در هر بازه ۱۲ تایی، به درایه‌های متناظر با سطر و ستون مُد، یک واحد اضافه می‌کند. پس از طی کردن کل بازه ۱۸۰ تایی، درایه دارای ماکزیمم عدد موجود در این ماتریس، متناظر با محل حرفی خواهد بود که اسپل شده است.

علاوه بر نوشتن این دو تابع، موارد دیگری نیز در طی فرایند حل این مسأله آزمایش شدند که به شرح زیر می‌باشند:

۱- قرار دادن  $\text{cost}$  بیشتر برای تشخیص اشتباه یک نمونه  $\text{target}$  (که در واقع  $\text{non-target}$  بوده است). این کار توسط ورودی ' $\text{cost}$ ' تابع  $\text{fitediscr}$  و یک ماتریس  $2 \times 2$  انجام می‌شود.

۲- تصمیم‌گیری بر مبنای مقادیر  $\text{score}$  خروجی تابع  $\text{predict}$ . این مقادیر، به نوعی ضریب اطمینان حاصل از پیش‌بینی طبقه‌بندی کننده هستند که در حالت دو کلاسه، ماتریسی  $2 \times n$  است که  $n$  تعداد نمونه‌هایی است که می‌خواهیم دسته‌بندی کنیم. در این حالت، هر دو  $\text{score}$  متناظر مجموعی برابر با ۱ دارند که در واقع، جمع ضرایب اطمینان (احتمال) حضور در دو دسته آن‌ها طبعاً برابر با ۱ است. در حالت عادی دسته بندی به این صورت انجام می‌شود که احتمال حضور در هر دسته که بیشتر باشد، نمونه در آن دسته قرار می‌گیرد. کاری که در این قسمت انجام دادیم، بالابردن سطح احتمال لازم برای قرار گرفتن در دسته  $\text{target}$ ‌ها بود، (مثلاً تا ۸۰ یا ۹۰ درصد). با قرار دادن این حد سخت گیرانه، از حضور  $\text{non-target}$ ‌ها در دسته  $\text{target}$ ‌ها جلوگیری بیشتری انجام می‌شود. (همان طور که گفته شد و در بخش پنجم نیز دوباره بحث می‌شود، در این مسأله، مهم‌تر آن است که  $\text{target}$ ‌های شناسایی شده درست باشند، اگرچه تعدادشان کم بشود)

۳- تعویض میزان طول زمانی سیگنال ERP (که پیش فرض آن ۸۰۰ میلی ثانیه بیان شده است)

۴- بررسی اثر کم کردن میانگین از دیتای اصلی یا عدم انجام این کار (که تفاوت چندانی نداشت، به خصوص به دلیل وجود فیلتر فرکانسی) (البته به دلیل وجود فیلتر فرکانسی، در واقع در حالت ایده‌آل کم کردن میانگین هیچ اثری ندارد، چرا که فیلتر فرکانس صفر را از بین می‌برد؛ اما آن چه در واقعیت اتفاق می‌افتد آن است که به دلیل مناسب نبودن الگوریتم‌های فیلتر کردن، چنانچه سطح DC سیگنال نسبت به فرکانس‌های بالاتر، خیلی زیاد باشد، فیلتر به درستی انجام نمی‌شود)

**سؤال (۴)**

جدول زیر، اندیس ۱۵ ضریب با بزرگترین اندازه‌ها را در بردار Coeffs نشان می‌دهد: (برای سابیجکت ۳)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	155	103	356	269	306	208	357	408	406	350	1	3	207	150

با به دست آوردن باقیمانده بردار فوق بر ۵۱ خواهیم داشت:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	2	1	50	14	0	4	0	0	49	44	1	3	3	48

و نیز خارج قسمت به صورت زیر خواهد بود:

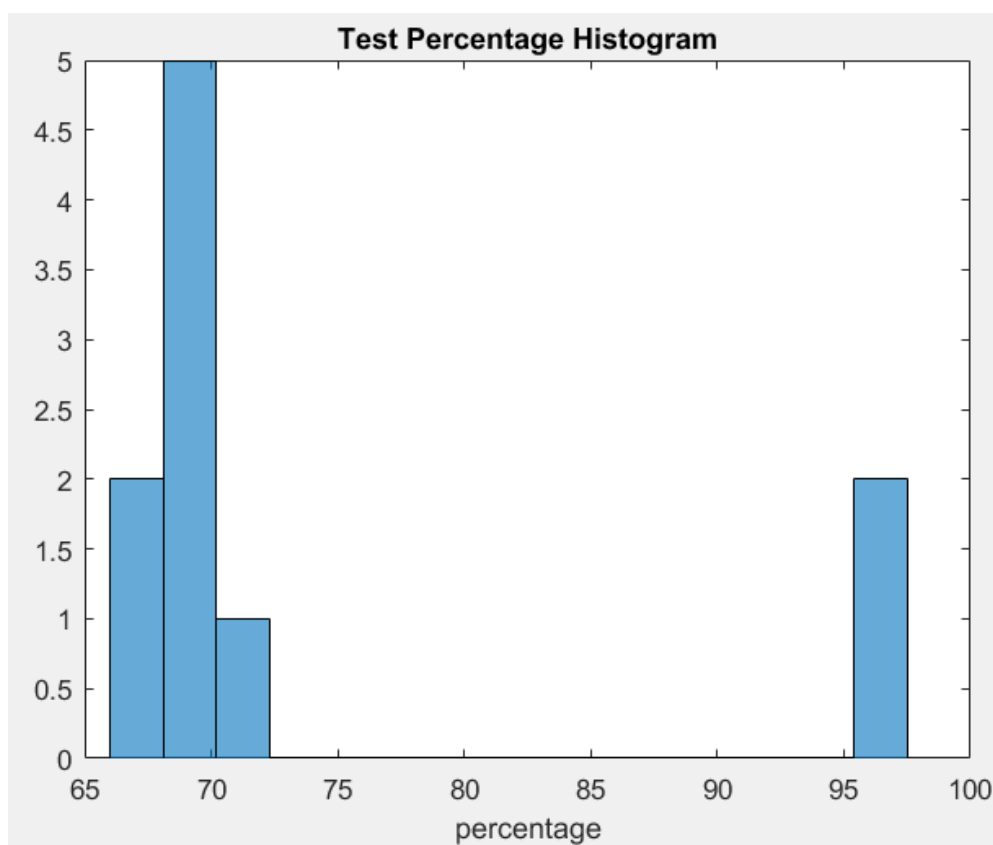
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	4	3	7	6	7	5	8	9	8	7	1	1	5	3

باقیمانده‌ها نقاط زمانی، و خارج قسمت‌ها شماره الکتروده هستند.

چیزی که مشاهده می‌شود، دور از انتظار (و شاید ناشی از ایراداتی است که در دیتاست داشته‌ایم، که باعث شد در بخش قبل جواب نگیریم) و آن این است که زمان‌های بسیار ابتدایی و نیز زمان‌ها انتهایی اثرگذار هستند، و در میان الکترودها نیز الکتروده ۷ و ۶ (الکتروده ۶ در ادامه نقش مهمی دارد و در ۱۵ تای اول زیاد مشاهده نمی‌شود) اثرگذار هستند.

## سؤال (۵)

(a)



(b) سیستم برای هیچیک از افراد موفق نبود.

(c) حدود زمانی برای تمامی افراد مشترک و مطابق با مطلب بیان شده در سؤال قبل است. این تطابق با دقت کمتری در مورد الکترودها نیز برقرار است، و می‌توان به طور کلی، مؤثرترین الکترودها را شماره ۷، و در ادامه شماره‌های ۸، ۵ و ۶ دانست.

(d) متأسفانه همان طور که گفته شد، در فرایند تشخیص لغات کاملاً ناموفق بودیم و نمی‌توانیم از این طریق انتخابی داشته باشیم. اما یک نکته‌ی دیگر که مطرح شد، جدولی بود که در سؤال ۲ بخش ۴ رسم کردیم که درصد target‌های درست تشخیص داده شده در میان کل target‌های تشخیص داده شده بود، که با توجه به توضیحاتی که در همان بخش دادیم (و ارزش شناسایی درست target‌ها، که بسیار بیشتر از non-target‌ها مهم است) روش RC می‌تواند مناسب‌تر باشد. (در صورتی که به خاطر ندارید، به توضیحات قسمت ۲-۴ مراجعه کنید)

(e) از نظر زمانی، انطباقی نداریم که احتمالاً به دلیل اشتباهی است که در این قسمت مرتکب شده‌ایم. از نظر شماره الکترودها، حضور شماره ۷ به عنوان الکترودها شاخص در هر دو سؤال مشاهده می‌شود. همچنین در مجموع با توجه به توضیحات (و نه مشاهدات) این سؤال، می‌توان در هر دو سؤال ادعا کرد که RC عملکرد بهتری دارد.

سؤال ۶)

همان طور که بیان شد، در این کاربرد خاص از مسأله‌ی طبقه‌بندی، یعنی طراحی speller، مهم‌ترین چیز آن است که در میان target‌هایی که تشخیص می‌دهیم، اشتباه کمتری وجود داشته باشد؛ اگر چه این مسأله موجب کم شدن تعداد آن‌ها بشود. با استناد به همین موضوع، دیدیم که احتمالاً درصدهای حاصل از Cross Validation و نیز تشخیص ماتریس test و train قابل اتکا نیستند، چرا که حجم زیادی از آن‌ها مربوط به دیتای non-target است که اگر چه موجب افزایش درصد می‌شود، اما در عملکرد دستگاه بی‌تأثیر خواهد بود. همچنین مشاهده شد که الکترودهای ۷، ۸، (و ۶ و ۵) و نیز زمان‌های ابتدایی و انتهایی در جداسازی مفیدتر هستند. (این نتایج – به خصوص نتایج زمانی – کمی عجیب است، که البته در این جا می‌توانیم یک نتیجه جالب بگیریم و آن نتیجه این است که با توجه به این که به دلیل نامعلومی ما در تشخیص کلمه موفق نبودیم، قاعدتاً این نتایج نیز معتبر نیستند، که با شهود ما از مسأله تا حدی منطبق است!)

**قسمت پنجم) یک سؤال دلخواه!**

در این قسمت می‌خواهیم برای به دست آوردن ویژگی و دسته‌بندی، از قدم فراتر از حوزه زمان بگذاریم. انتخاب اول، حوزه فرکانس است، اما می‌خواهیم برای رفع نسبی نقص‌های هر دو حوزه، حوزه زمان-فرکانس را امتحان کنیم. این کار با گرفتن تبدیل فوریه زمان کوتاه (STFT) امکان‌پذیر است. برای این کار می‌توان از تابع spectrogram متلب استفاده کرد، اما ما در این بخش چنین نمی‌کنیم و خود، به پیاده‌سازی این روش، به نحو متناسب با خواسته‌ی خود و شرایط مسأله می‌پردازیم.

با توجه به شرایط مسأله (این مسأله ارتباطی به خواب ندارد و صرفاً یک مسأله‌ی تمرکز و تشخیص است)، به سراغ دو باند فرکانسی آلفا و بتا می‌رویم. برای تعریف ویژگی، از توان متوسط این دو باند فرکانسی در بازه‌های زمانی کوتاه استفاده می‌کنیم (که به تعبیری، همان STFT است).

به این منظور، با استفاده از ماتریس‌های مکعبی تولید شده توسط تابع TrialExtraction، در بازه‌های زمانی به طول پنج نمونه (sample)، اقدام به محاسبه متوسط انرژی باند آلفا و باند بتا برای هر یک از کانال‌ها می‌کنیم و میزان این دو انرژی را در یک سطر به دنبال هم قرار می‌دهیم. بنابراین، آنچه به دست می‌آید، یک ماتریس سه بعدی است که یک بُعد آن مربوط به trialها (لحظات نشان داده شدن یک سطر/ستون/کاراکتر)، یک بُعد مربوط به الکتروود، و بُعد آخر انرژی متوسط باندهای فرکانسی در طول بازه‌های زمانی است. در ادامه برای ساخت ماتریس Feature، این ماتریس سه بعدی را به یک ماتریس دو بعدی تبدیل کرده، یعنی اطلاعات مربوط به الکتروودهای مختلف را به دنبال هم می‌چینیم.

در نهایت، برای سنجش کیفیت این ویژگی‌ها، به طبقه‌بندی توسط LDA و SVM و محاسبه Confusion Matrix برای هر کدام می‌پردازیم. برای جلوگیری از طولانی شدن محتوا، ۴ نمونه از سابجکت‌ها را مورد بررسی قرار می‌دهیم:

(ابتدا نتایج حاصل از بررسی ۴ سابجکت را گزارش می‌کنیم، در نهایت به تحلیل و جمع‌بندی داده‌ها می‌پردازیم)

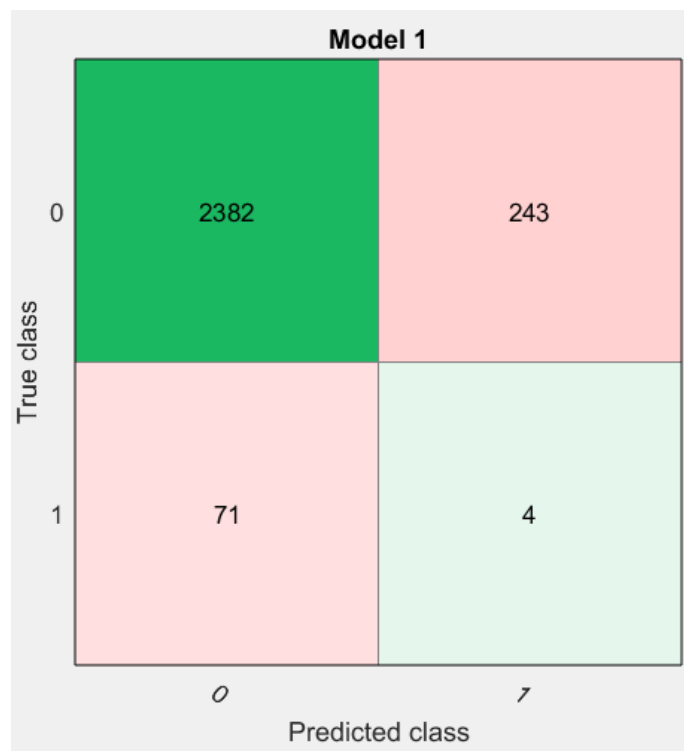


## سابجکت شماره ۱:

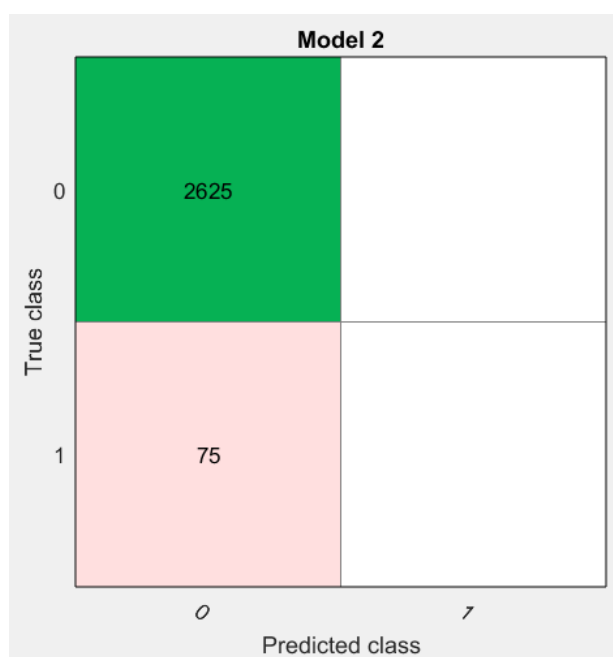
الف) درصد پاسخ گویی صحیح حاصل از 5-fold Cross Validation

<b>1</b> ☆ Linear Discriminant	Accuracy: 88.4%
Last change: Linear Discriminant	160/160 features
<b>2</b> ☆ SVM	Accuracy: <b>97.2%</b>
Last change: Linear SVM	160/160 features

ب) Confusion Matrix مربوط به LDA



ج) Confusion Matrix مربوط به SVM

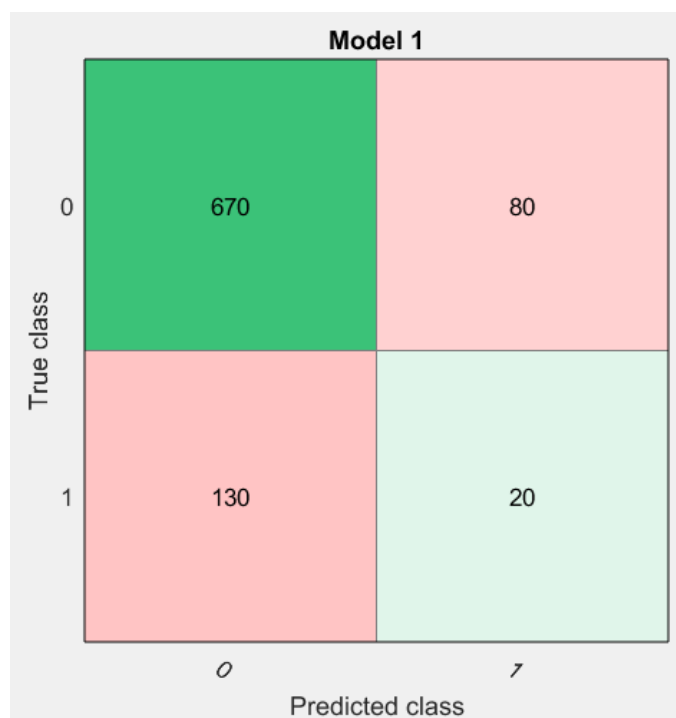


### سابجکت شماره ۳:

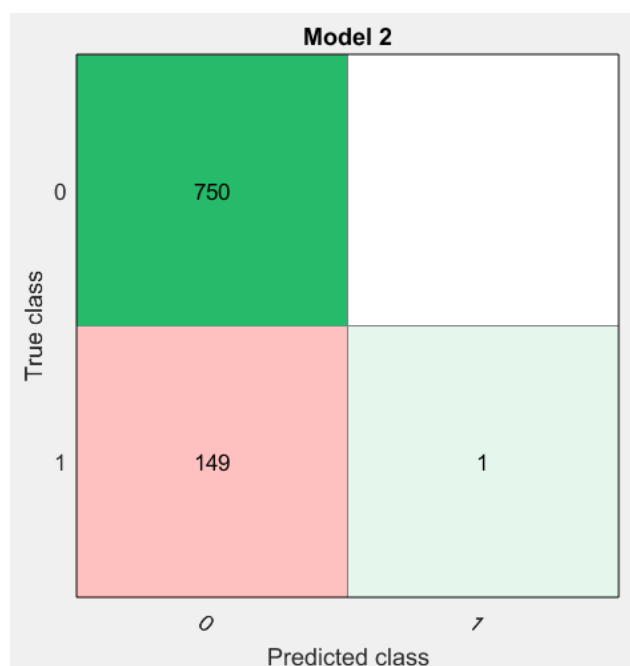
الف) درصد پاسخ گویی صحیح حاصل از 5-fold Cross Validation

<b>1</b> ☆ Linear Discriminant	Accuracy: 76.7%
Last change: Linear Discriminant	160/160 features
<b>2</b> ☆ SVM	Accuracy: <b>83.4%</b>
Last change: Linear SVM	160/160 features

ب) Confusion Matrix مربوط به LDA



ج) Confusion Matrix مربوط به SVM

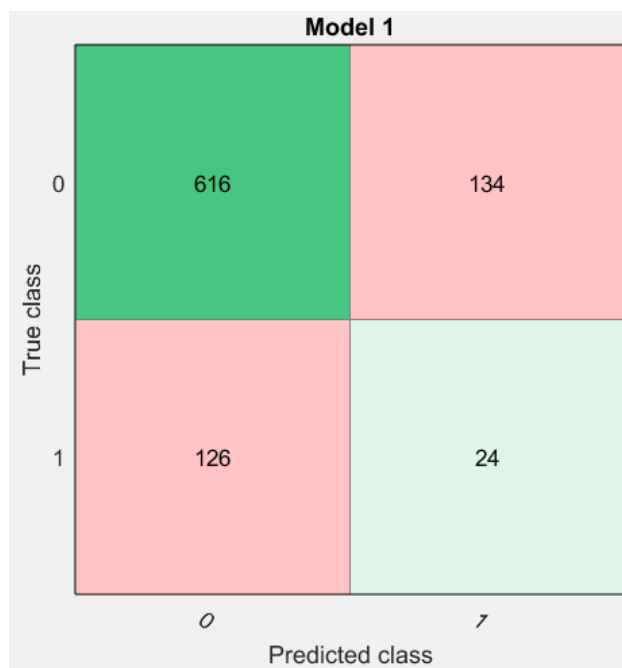


## سابجکت شماره ۵:

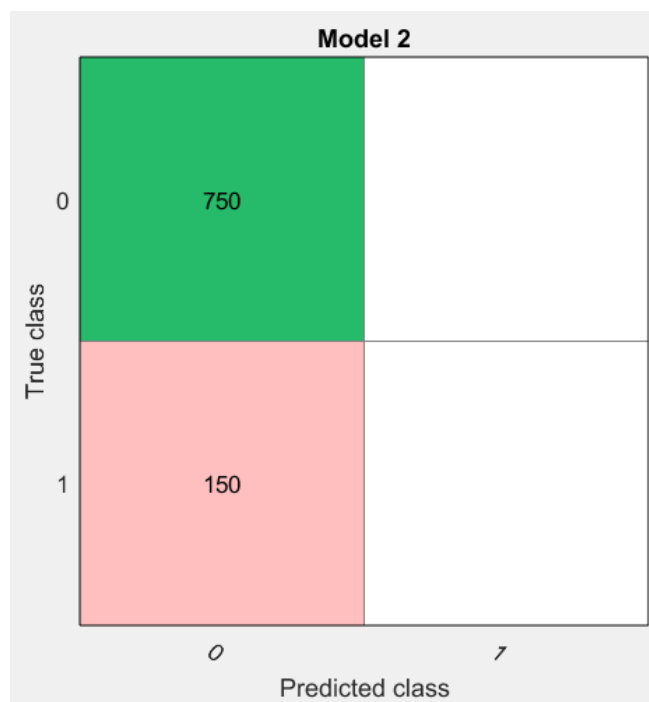
الف) درصد پاسخ گویی صحیح حاصل از 5-fold Cross Validation

<b>1</b> ☆ Linear Discriminant	Accuracy: 71.1%
Last change: Linear Discriminant	160/160 features
<b>2</b> ☆ SVM	Accuracy: <b>83.3%</b>
Last change: Linear SVM	160/160 features

ب) Confusion Matrix مربوط به LDA



ج) Confusion Matrix مربوط به SVM

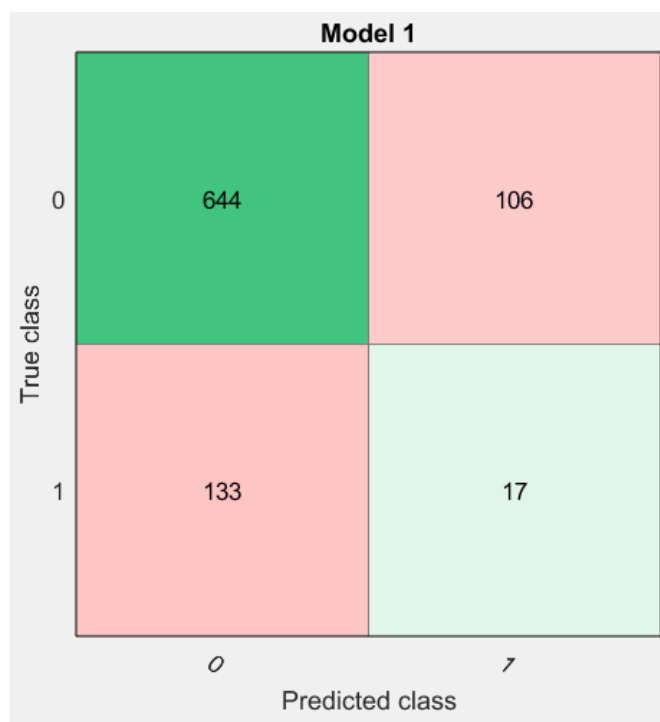


## سابجکت شماره ۷:

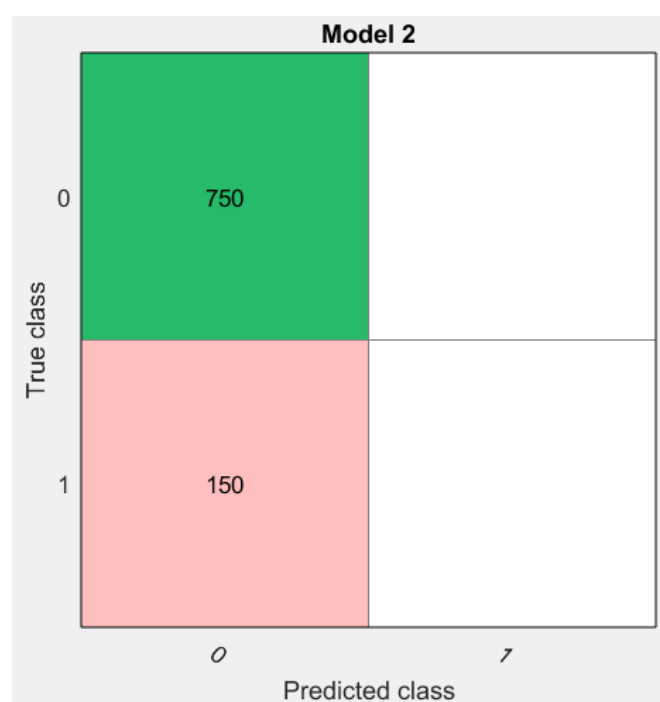
الف) درصد پاسخ گویی صحیح حاصل از 5-fold Cross Validation

1 ☆ Linear Discriminant	Accuracy: 73.4%
Last change: Linear Discriminant	160/160 features
2 ☆ SVM	Accuracy: <b>83.3%</b>
Last change: Linear SVM	160/160 features

ب) Confusion Matrix مربوط به LDA



ج) Confusion Matrix مربوط به SVM



**- تحلیل نتایج و جمع‌بندی:**

تحلیل نتایج فوق را به شکل عناوین زیر بیان می‌کنیم:

۱- درصد پاسخ‌گویی SVM همواره بیشتر از LDA است.

۲- برای سابجکت‌هایی که از روش RC استفاده کرده‌اند، درصد LDA تقریباً حدود ۷۳٪ است (به طور متوسط) که این مقدار برای سابجکت SC به ۸۸٪ افزایش داشته است. همچنین برای SVM نیز برای سابجکت‌های RC درصد پاسخ‌گویی برابر ۸۳٪ و برای سابجکت SC ۹۷٪ است.

۳- میزان درصد پاسخ‌گویی در تمامی مواردی که از یک متد استفاده کرده‌اند، با تقریب خوبی ثابت است.

۴- اگر میزان درصد پاسخ‌گویی SVM بیشتر است، اما با نگاه کردن به Confusion Matrix‌ها می‌بینیم که در SVM، عملاً تمامی داده‌ها در طبقه‌ی non-target قرار می‌گیرند، و به دلیل آن که تعداد نمونه‌های non-target بسیار بیشتر از نمونه‌های target است، درصد پاسخ‌گویی این روش بالا رفته است. علت افزایش درصد در سابجکت SC نیز، همین مسأله است، چرا که نسبت تعداد target‌ها به non-target‌ها برای این سابجکت، به مراتب بیشتر از سابجکت‌های RC است.

۵- با توجه به توضیحات فوق، احتمالاً استفاده از روش LDA می‌تواند اثربخشی بیشتر داشته باشد. با این وجود، با بررسی Confusion Matrix‌ها، می‌بینیم که تعداد نمونه‌هایی که non-target بوده‌اند، اما target طبقه‌بندی شده‌اند، در مقایسه با آن دسته از نمونه‌هایی که به درستی target شناخته شده‌اند، زیاد است؛ و این باعث می‌شود که speller عملکرد خوبی نداشته باشد. در واقع چیزی که باید در طبقه‌بندی بیشتر مورد توجه قرار بگیرد، آن است که سخت‌گیری بیشتری در تشخیص نمونه‌های target انجام شود، تا اگرچه تعداد آن‌ها کمتر می‌شود، اما بتوانیم با اطمینان بیشتری از آن‌چه که به دست می‌آید، استفاده کنیم. سخت‌گیری در طبقه‌بندی می‌تواند با تعریف cost نامتقارن برای دو طبقه (cost بیشتر برای انتخاب اشتباه یک نمونه در گروه target، در حالی که واقعا non-target است) و نیز با در نظر گرفتن مقادیر score حاصل از هر طبقه‌بندی، و قرار دادن آستانه‌ی انتخاب یک نمونه در دسته‌ی target، در جایی بالاتر از ۵۰٪ (یعنی مثلاً هر نمونه باید score برابر با حداقل ۷۰٪ برای قرارگیری در دسته target داشته باشد تا در آن دسته پذیرفته شود) صورت پذیرد.

۶- برخی تحلیل‌های دیگر مانند Quadratic SVM و Quadratic Discriminant Analysis نیز بر روی داده‌ها انجام شد که نتایج به مراتب بدتری (در مقایسه با هر آن‌چه که تا کنون بررسی کرده‌ایم) به دست آمد، که به همین دلیل از بیان جزئیات نتایج آن‌ها صرف نظر می‌کنیم.

۷- در نهایت، با مقایسه جواب‌های این بخش با آن‌چه که در بخش چهارم از این پروژه صورت پذیرفت، می‌بینیم که ویژگی‌هایی که در این بخش انتخاب کردیم، افزایشی در حدود ۱۰٪ در درصد پاسخ‌گویی طبقه‌بندی LDA ایجاد کردند، که این با توجه به ویژگی‌های مثبت حوزه زمان-فرکانس قابل انتظار بود. اما می‌بینیم که با توجه به Confusion Matrix‌ها و ماهیت خاص این مسأله، که در آن، هدف آن است که هر آن‌چه به عنوان target انتخاب می‌شود درست باشد (و قرار دادن target در دسته non-target خسارت چندانی وارد نمی‌کند)، به نظر می‌رسد که استفاده از روش حوزه زمان (P300) می‌تواند تقریباً به همان اندازه پاسخگو باشد، و با توجه به سادگی آن روش، شاید ارزش نداشته باشد که حوزه زمان را کنار بگذاریم.