

باسمه تعالی



دانشگاه صنعتی شریف

دانشکده مهندسی برق

درس علوم اعصاب محاسباتی

یافتن معیاری برای عمق خواب مبتنی بر سیگنال‌های مغزی

بهراد منیری

معید جمال‌زاده

استاد

دکتر حمید کربلایی آقاجان

بهار ۱۳۹۷

بخش اول – آشنایی با پژوهش مقاله اصلی

سوال یک

این مقاله سعی دارد معیاری برای عمق خواب از طریق سیگنال‌های مغزی بیابد. در ابتدا به توضیح روش‌هایی برای تخمین عمق خواب از طریق توان Slow-Wave (سیگنال‌های مغز با فرکانس کم) یا Slow-Wave Power (SWP) می‌پردازد. این روش مشکلاتی دارد که یکی از آن‌ها تاثیر آناتومی بدن از جمله ضخامت جمجمه در این توان است. یک مشکل دیگر این است که SWP در بانوان حدود دو برابر آقایان است ولی پژوهش‌های قبلی نشان داده‌اند که در عمق خواب آقایان و بانوان تفاوت معناداری وجود ندارد. یعنی استفاده مستقیم از SWP برای عمق خواب چندان نتایج جالبی را در بر نخواهد داشت زیرا معیارهایی بی ارتباط با خواب در SWP تاثیر گذارند.

یک راه حل برای رفع بایاس‌های مربوط به آناتومی استفاده از $SW\%$ (Slow-Wave microcontinuity) است که به صورت کسری از Slow-Wave است که در حلقه فیدبک به مشاهده بعد جریان می‌یابد. در معرفی مدل فیدبک این تعریف واضح‌تر می‌شود. با این نورمالیزیشن مسأله تاثیر آناتومی حل می‌شود ولی همانند SWP این مشکل وجود دارد که رزولوشن زمانی این دو به دلیل اینکه پردازش آن‌ها در حوزه فرکانس و در بازه‌های زمانی بزرگ صورت می‌گیرد محدود است.

مشکل بزرگ‌تری که در این دو روش موجود است این است که مشخص نیست که ارتباط این سیگنال‌ها با عمق خواب دقیقا در چیست و تنها تخمینی از این متغیرها عمق خواب را توصیف می‌کند و به همین دلیل به ارتباط مستقیم تری برای این ارتباط باید وجود داشته باشد که وابستگی عمق خواب به SWP نیز از آن نشأت می‌گیرد.

هدف مقاله ارائه مدلی با جزئیات کامل برای یک estimator بر مبنای حلقه feedback و gain است و ادعا می‌کند که عمق خواب بهره این حلقه فیدبک وابسته است.

سوال دو

برای داده‌گیری از ۲۲ مرد و زن کاملاً سالم و با سن‌های متفاوت که از مشکل خواب رنج می‌بردند انجام شده است. در دو شب مجزا برای از هر شخص دیتا گرفتند و بهره مدل فیدبک و همچنین SWP را آنالیز کردند. در شب اول این افراد به صورت عادی به خواب رفتند و در شب دوم افراد تحت تاثیر داروی آرامبخش Temazepam بودند. در این مقاله تاثیر پارامترهایی مثل سن و جنسیت بر دو تخمین با مدل SWP و مدل فیدبک بررسی می‌شوند.

سپس به افراد پرسشنامه‌ای داده شده که در آن درباره کیفیت خواب سوال شده است. در ادامه در این باره بیشتر بحث می‌کنیم.

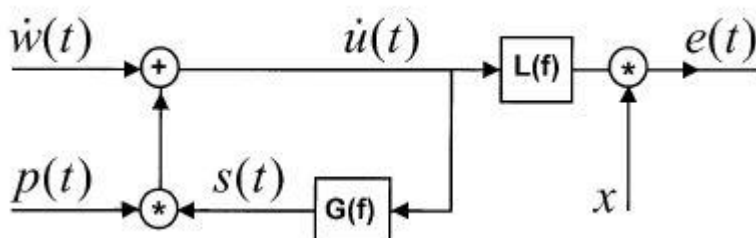
سوال سه

یک مشکل در پژوهش خواب در گذشته این بوده که کوریلشنی بین کیفیت خواب که توسط یک پرسشنامه از افراد پرسیده شده و طول کل زمان خواب Non REM که با روش SWP به دست آمده، که یک معیار برای سنجش عمق و کیفیت خواب در نظر گرفته می‌شده، وجود ندارد و مدت زمان خواب Non REM نمی‌توانسته بین یک خواب خوب و خواب بد تمایز دهد.

در این مقاله gain حلقه فیدبک به عنوان معیاری از عمق خواب شناخته می‌شود. پژوهشگران تفاوت این پارامترها را برای دو شب هر فرد محاسبه کرده و بررسی می‌کنند که این تفاوت کوریلشنی با تفاوتی که خود افراد در پرسشنامه ذکر کردند دارد یا نه. روش مقاله بر اساس ویژگی‌های فیزیولوژیکی است بنابراین انتظار می‌رود کوریلشنی که در گذشته دیده نمی‌شد را آشکار کند.

مقاله مدلی را پیشنهاد می‌دهد که سعی شده است تمام المان‌هایی که در خواب تاثیر دارند را در خود بگنجانید در این روش از حلقه فیدبک استفاده شده است و بهره ی $p(t)$ برای آن در نظر گرفته شده است که در شکل ۱ نشان داده شده است.

مقاله ادعا می‌کند که عمق خواب وابسته به بهره این feedback loop می‌باشد یعنی همان $p(t)$ و برای اثبات ادعای خود تلاش کرده است تا تمام الماهای دیگر را تا جای ممکن ثابت نگه داشته و تنها $p(t)$ را تغییر دهد تا بتواند ادعای خود را به اثبات برساند.



شکل ۱: مدل ارائه شده برای به دست آوردن عمق خواب

در این مدل \dot{w} نویز سفید استاندارد است. $p(t)$ بهره حلقه فیدبک است که ادعا می‌شود با عمق خواب ارتباط دارد. سیگنال حاصل از یک فیلتر پایین‌گذر، L ، عبور می‌کند و بعد از جمع شدن با x که معرف پارامترهای غیر خواب است، خروجی $e(t)$ به دست می‌آید. در این مدل G فیلتری است که Slow-Wave را جدا می‌کند یعنی $S(t)$ همان Slow Wave است.

افزایش عمق خواب در این مدل با ثابت نگه داشتن دیگر پارامترها با افزایش $p(t)$ شبیه‌سازی می‌شود. یک مدل که در این مقاله برای G بررسی شده، استفاده از تابع تبدیل زیر است.

$$G(f) = \frac{1}{1 + j \cdot Y(f)} \quad \text{with } Y(f) = \frac{f_0}{B} \cdot \left(\frac{f}{f_0} - \frac{f_0}{f} \right)$$

$$f_0 = 1 \text{ Hz} \quad B = 1.5 \text{ Hz}$$

نتایج این شبیه‌سازی به صورت بصری از سیگنال واقعی EEG قابل تمایز نیست.

علت استفاده پژوهشگران از این مدل این است که

۱. وجود حلقه فیدبک فرکانس پایین که در آن بهره فیدبک متناظر با عمق خواب است.
۲. فعالیت غیر قابل پیش‌بینی از منشاهای خارجی عامل کار کردن حلقه فیدبک است (نویز سفید w).
۳. افزایش بهره فیدبک باعث افزایش SWP در سیگنال می‌شود.

مدل‌های متفاوتی وجود دارد که این سه ویژگی را دارد بنابراین پژوهشگران مدل‌های متفاوتی را در مقاله بررسی کرده‌اند. تفاوت این مدل‌ها در نوع فیلترها و سیگنال تصادفی درایو کننده حلقه فیدبک و مکان gain فیدبک است. در آخر مدلی که حاصل آن شباهت بیشتری با سیگنال واقعی EEG داشته انتخاب شده است.

بخش دوم – آشنایی با دیتاست

سوال دو

تابع FeatureExtraction به عنوان ورودی آدرس فایل‌های edf و hypnogram را می‌گیرد و فرایند زیر را طی می‌کند.

۱. ابتدا به کمک تابع edfread فایل edf را خوانده و سیگنال‌های سابجکت مورد نظر را استخراج می‌کند. سپس با استفاده از

تابع AnnotExtract فایل Hypnogram مربوطه را خوانده و زمان‌های تغییر استیج خواب را استخراج می‌کند.

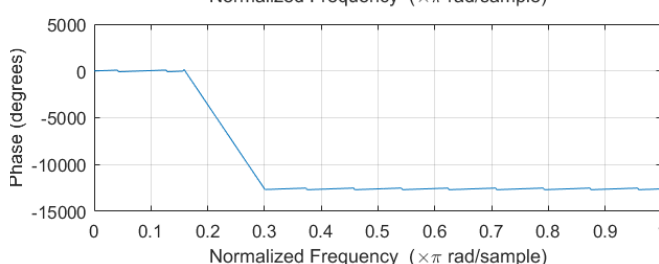
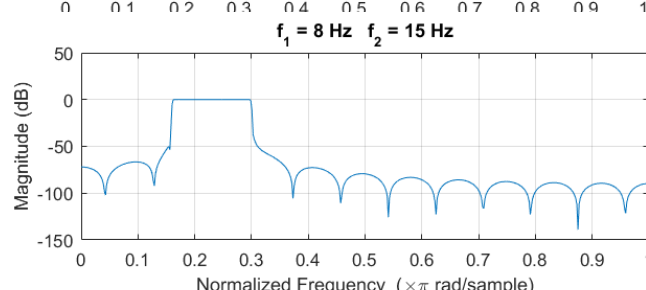
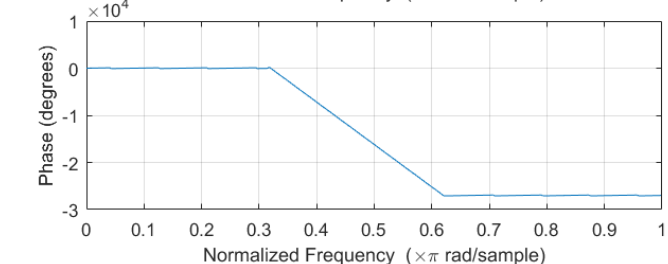
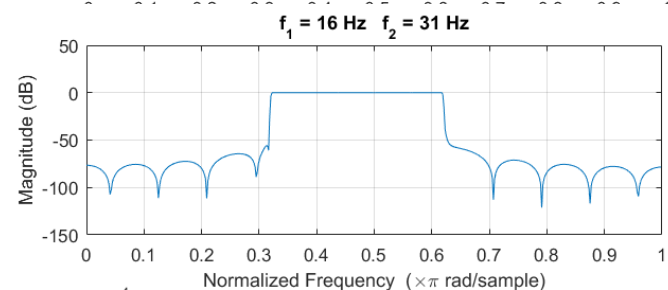
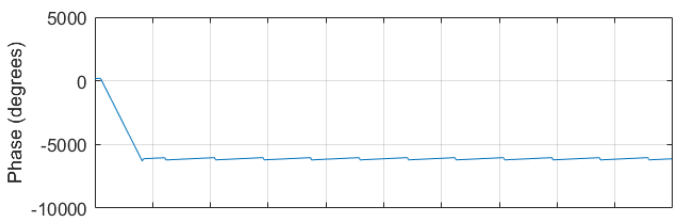
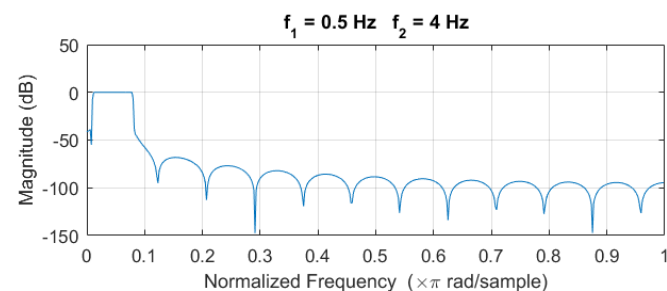
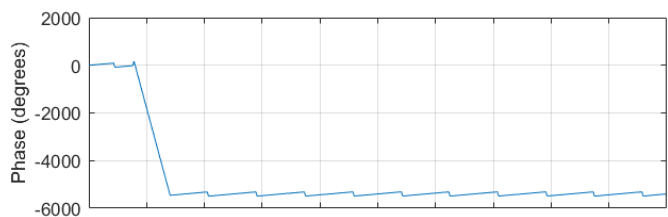
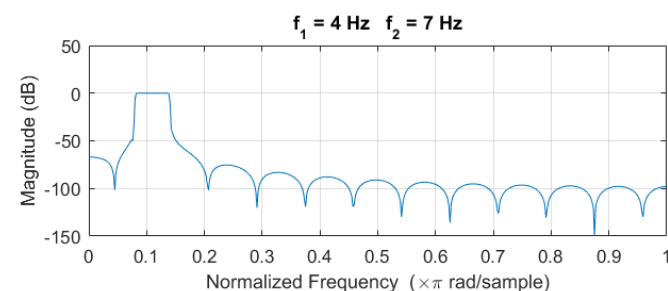
۲. با تابع BPF چهار فیلتر FIR میان‌گذر با Kaiser Window برای جداکردن باندهای فرکانسی آلفا، بتا، دلتا و تتای

سیگنال‌های EEG دو الکترود Oz و Fpz ساخته می‌شود.

می‌دانیم که عموماً نویز زیادی در فرکانس زیر 0.5 Hz روی سیگنال EEG وجود دارد و بنابراین بازه فرکانسی باند دلتا را از 0.5 هرتز تا 4 هرتز گرفتیم. در ویکیپدیا نیز همین بازه برای این باند ذکر شده است. فرکانس نمونه‌برداری داده‌ها ۱۰۰ هرتز است.

```
h_alpha = BPF (1001, 8 , 15, 100);
h_beta = BPF (1001, 16 , 31, 100);
h_theta = BPF (1001, 4 , 7, 100);
h_delta = BPF (1001, 0.5, 4, 100);
```

نمودارهای زیر پاسخ فرکانسی فیلترهای استفاده شده هستند.



داده‌ها را به بازه‌های ۱۰ ثانیه‌ای تقسیم می‌کنیم و به باندهای فرکانسی تقسیم می‌کنیم. توان هر باند فرکانسی در هر بازه را به صورت

$$P_{band} = \frac{\sum_{n=k}^{n=K+n-1} s^2[n]}{N}$$

تعریف می‌کنیم.

۳. در انتهای سیگنال مقدار زیادی داده ثابت داریم که کد ما نقطه‌ای که در آن سیگنال ثابت می‌شود را پیدا کرده و در خروجی بر میگرداند.

۴. در آخر نیز تابع ما بردار state که استیت خواب در هر بازه ۱۰ ثانیه‌ای است را بر می‌گرداند. زمان تمام نقاط شروع state جدید در فایل Hypnogram مضرب ده ثانیه است بنابراین انتخاب ده ثانیه به عنوان طول پنجره زمانی می‌تواند منطقی باشد.

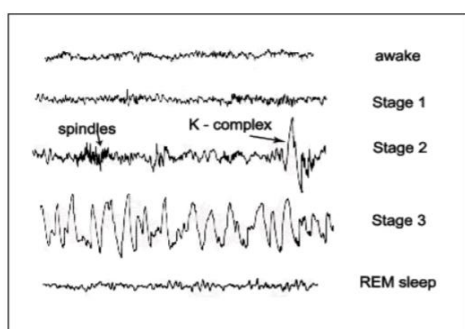
سوال سه

سیگنال EOG ولتاژ بین قرنیه و شبکیه چشم است. با این سیگنال می‌توانیم حرکت چشم را مشخص کنیم.

ادعا می‌شود ارتباط های زیر بین سیگنال‌های مغزی و stage های مختلف خواب وجود دارد. ما این اطلاعات را از [2] کسب کردیم.

توضیحات	استیت خواب
ولتاژ کم در EEG و وجود محتوا در فرکانس های مختلف. معمولاً پترن دندان‌اره‌ای در EEG دیده می‌شود. مقدار سیگنال EMG به کمترین حد خود رسیده و حرکات سریع چشم رخ می‌دهد که در EOG تاثیر می‌گذارند.	REM
همانند REM در این استیت هم در EEG دامنه کم است. بیشتر محتوا در فرکانس های ۲ تا ۷ هرتز است یعنی انرژی باند تتا و دلتا افزایش می‌یابد. در این استیت آرام چشم نیز وجود دارد.	Stage 1
در EEG پترن های Spindle و K-Complex دیده می‌شود که مثال هایی از آن در ادامه آمده است. Sleep Spindle در فرکانس های ۱۲ تا ۱۴ هرتز رخ می‌دهد.	Stage 2
بیشتر محتوا در فرکانس های ۲ هرتز و کمتر است و گاهی Sleep Spindle و K-Complex نیز به چشم می‌خورد.	Stage 3
این استیت شبیه استیت سه است با این تفاوت که امواج با فرکانس زیر ۲ هرتز در آن بیشتر می‌شود. یعنی افزایش انرژی باند دلتا.	Stage 4
سیگنال EEG در بازه فرکانسی بزرگی محتوا دارد و دامنه EMG نسبت به حالت های دیگر بیشتر است.	Wake

جدول ۱

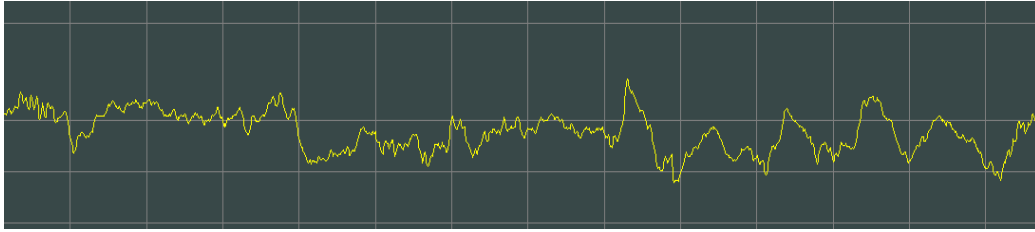


شکل روبرو نمونه‌ای از سیگنال EEG در استیت‌های مختلف خواب است که در مقاله [2] آورده شده است.

به طور کلی انرژی باند دلتا برای تشخیص استیت‌های NREM به کار می‌رود. این باند تقریباً همان Slow Wave ذکر شده در مقاله است.

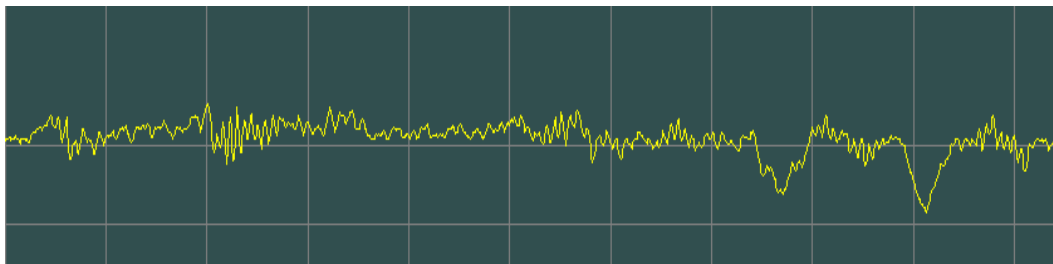
شکل ۲

۱. Stage 3 از خواب در دیتای EEG مربوط به الکتروود Fpz



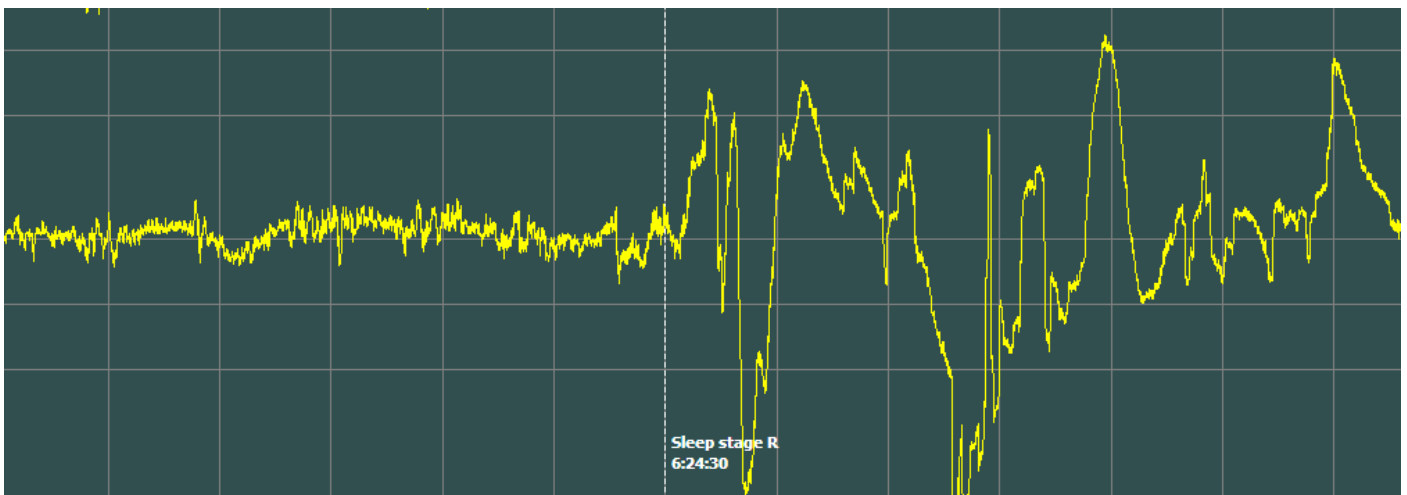
شکل ۳ Sleep Spindle , K-Complex در الکتروود Fpz استیت سوم خواب

۲. Sleep Spindle در Stage 2 خواب در دیتای الکتروود Fpz



شکل ۴ Sleep Spindle در الکتروود Fpz استیت دوم خواب

۱. در تصویر زیر سیگنال EOG در ناحیه گذر از Stage2 به REM دیده می‌شود. در REM اثرات حرکت چشم بر EOG مشهود است. شک



شکل ۵ شروع حرکات سریع چشم در REM

سیگنال الکتروود Fpz از جلوی سر و سیگنال الکتروود Oz از پشت سر است بنابراین انتظار کوریلیشن بالا بین سیگنال‌های این دو الکتروود نداریم. سیگنال EOG و سیگنال EEG باند Fpz روی هم تاثیر می‌گذارند زیرا از دو نقطه نزدیک به هم شده است و حرکت چشم که در سیگنال EOG نمود پیدا می‌کند باعث تغییر در الکتروودهای جلوی سر می‌شود، مثل الکتروود Fpz.

بخش سوم – یافتن معیاری برای خواب مبتنی بر PCA

سوال یک

PCA روشی است که با یک تبدیل متعامد، تعدادی متغیر احتمالا وابسته و با کوواریانس ناصفر را به مجموعه‌ای از متغیرهای نایسته و بدون کوریلیشن تبدیل می‌کند. اگر $n-1$ مشاهده از p متغیر تصادفی داشته باشیم، داده‌های ما در یک زیرفضای $\min(n-1, p)$ بعدی قرار می‌گیرند پس به همین تعداد جهت مستقل خواهیم داشت. نگاشت خطی ذکر شده در PCA به این نحو است که تصویر داده‌ها بر جهت اول بیشترین واریانس را خواهد داشت و به همین ترتیب جهت‌های بعدی واریانس‌های کمتری خواهند داشت و کمترین واریانس را در جهت آخر خواهیم داشت. به بیان بهتر، جهت k ام بر زیرفضای تولید شده توسط جهت‌های قبلی عمود است و بیشترین واریانس ممکن را با این شرط دارد.

فرض کنید x_i بردار دیتای ما باشد یعنی $x_i \in \mathbb{R}^n$ است. در ابتدا میانگین دیتا را صفر می‌کنیم یعنی

$$x_i := x_i - \frac{1}{n} \sum_{i=1}^n x_i$$

Sample Covariance را برای دیتا به دست می‌آوریم

$$C = \frac{1}{n-1} \sum x_i x_i^T$$

بر روی این ماتریس eigenvalue decomposition را اجرا می‌کنیم. فرض کنید V ماتریسی باشد که در ستون‌های آن بردارهای ویژه C قرار گرفته و Γ ماتریسی باشد که بر روی داریه‌های آن مقادیر ویژه ماتریس C قرار دارد و دیگر درایه‌های آن صفر هستند. یعنی

$$V = (v_1 | v_2 | \dots | v_n) \quad \text{و} \quad \Gamma = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix}$$

به طوری که $Cv_i = \lambda_i v_i$.

با از اجرای این decomposition داریم

$$C = V\Gamma V^T$$

فرض کنی در ماتریس‌های فوق $\lambda_1 > \lambda_2 > \dots > \lambda_n$ باشد. با تبدیل بالا، هر نقطه x_i به $(v_1^T x_i, v_2^T x_i, \dots, v_n^T x_i)^T$ نگاشته می‌شود.

حال یک خط در فضای \mathbb{R}^n با بردار مماس w در نظر بگیرید که از مبدا نیز می‌گذرد. تصویر بردار x_i بر این خط $\frac{w^T x_i}{\|w\|}$ است. از این به بعد فرض می‌کنیم $\|w\| = 1$ است یعنی w را نرمالیزه می‌کنیم.

واریانس تصویر داده‌ها بر روی این خط را به دست می‌آوریم.

$$\sigma_w^2 = \frac{1}{n-1} \sum_{i=1}^n (w^T x_i)^2 = w^T C w$$

که در آن C همان ماتریس کوواریانس داده‌هاست. هدف پیدا کردن w ای است که واریانس را بیشینه کند یعنی باید یک مساله بهینه‌سازی حل کنیم.

$$\begin{cases} ||w|| = 1 \\ w = \operatorname{argmax}_w (\sigma_w^2) \end{cases}$$

برای حل این مساله بهینه‌سازی با یک شرط از لاگرانژی استفاده می‌کنیم

$$L = w^T C w + \alpha (1 - w^T w)$$

$$\nabla L = 2Cw - 2\alpha w = 0 \rightarrow Cw = \alpha w$$

پس w یک بردار ویژه ماتریس C است و α مقدار ویژه این ماتریس است.

در صفحه قبل به دست آوردیم که واریانس تصویر داده‌ها برابر

$$\sigma_w^2 = w^T C w = w^T \alpha w = \alpha$$

است که در آن α مقدار ویژه ماتریس C مربوط به بردار ویژه w است پس جهت با بیشترین واریانس همان جهت بزرگترین بردار ویژه است.

به همین ترتیب می‌توان نشان داد که تصویر داده بر m بردار ویژه با بزرگترین مقدار ویژه بیشترین واریانس‌ها را خواهند داشت.

اطلاعات مفید در پراکندگی داده‌هاست پس برای کاهش بعد داده‌ها می‌توان روش PCA را بر آنها اعمال کنیم و به تعداد دلخواه بردار ویژه با مقدار ویژه بزرگ از آنها انتخاب کرده و داده‌ها را بر آنها تصویر کنیم. تصاویر بر روی این جهت‌ها بیشترین واریانس را خواهند داشت.

حال به بررسی خروجی‌های تابع `pca` متلب می‌پردازیم. این اطلاعات از Documentation متلب به دست آمده است. [3]

۱. خروجی `coeff`

این خروجی یک ماتریس مربعی است که در هر ستون آن یک بردار ویژه ماتریس کوواریانس داده ورودی است. به طوری که اگر هر ستون X یک متغیر و هر سطر X یک داده ما باشد $y = \text{coeff} * X$ بردار داده‌های ما در پایه بردارهای ویژه ماتریس کوواریانس است. طبق انتظاری که از PCA داریم این ماتریس یک ماتریس متعامد است.

۲. خروجی `score`

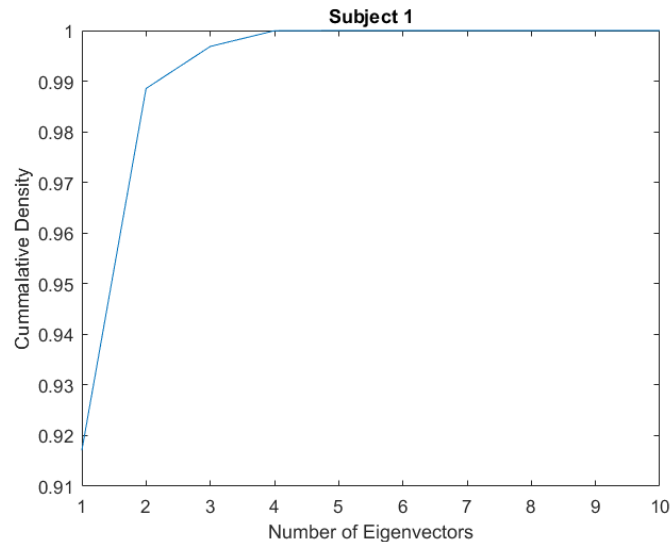
اگر X ماتریسی با میانگی صفر باشد و بر روی آن `pca` را اعمال کنیم، $\text{score} = X * \text{coeff}$ که در آن `coeff` ماتریسی است که در بالا تعریف شده است.

۳. خروجی `latent`

این خروجی یک بردار است که در آن مقادیر ویژه به ترتیب مرتب شده ماتریس کوواریانس قرار گرفته‌اند. این مقادیر ویژه برابر واریانس تصور داده‌ها بر روی بردار ویژه مربوطشان هستند.

در این بخش داده‌های بر روی داده‌های سابجکت اول PCA زده و نتایج را گزارش می‌کنیم. می‌دانیم که اعضای بردار latent مقداریر ویژه یا به عبارتی واریانس داده در جهت‌های مختلف PCA هستند. نمودار جمع تجمعی latent را رسم می‌کنیم. این نمودار نشان می‌دهد که با نگه داشتن N جهت اول PCA چند درصد از واریانس کل را حفظ کرده‌ایم.

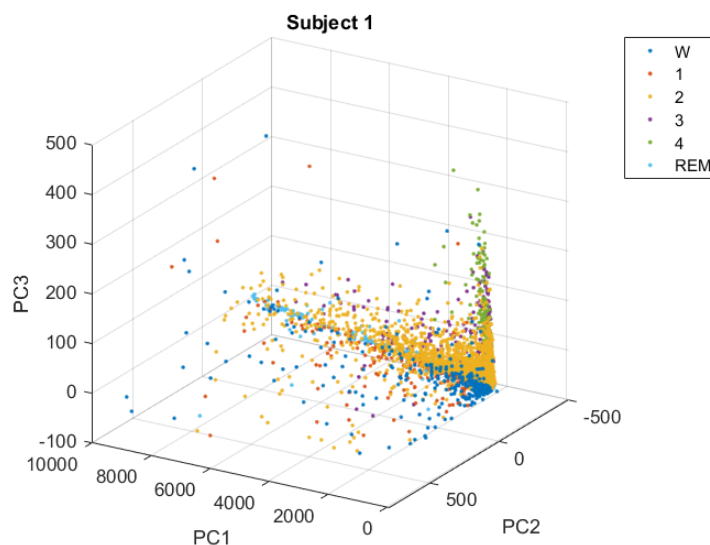
سه مولفه اول 99.6% از واریانس کل را دربر می‌گیرند.



شکل ۶ - داده‌های سابجکت اول

سوال سه

داده‌ها را در فضای حالت سه مولفه اول PCA با رنگ‌های مختلف برای هر استیت خواب رسم می‌کنیم. در بخش بعد این قسمت، این نمودار از جهت‌های مختلف مورد بررسی قرار می‌گیرد.



شکل ۷ - داده‌های سابجکت اول

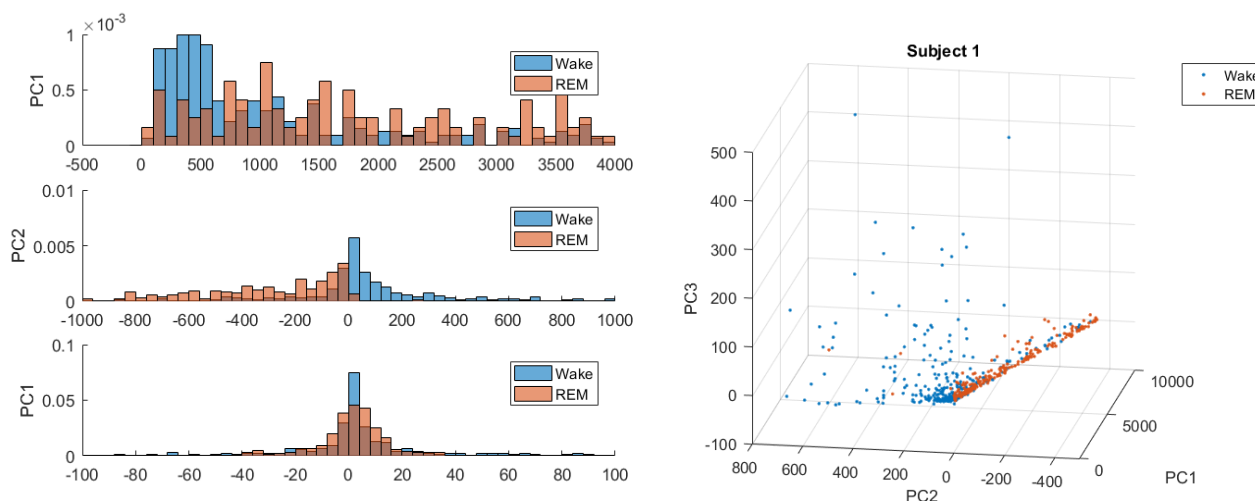
سوال چهار

در این بخش می‌خواهیم تاثیر جهت‌های مختلف PCA بر استیت خواب را بررسی کنیم. برای این کار از دو روش استفاده می‌کنیم. روش اول رسم داده‌ها در فضای سه بعدی و تشخیص از روی میزان پخش شدگی داده‌ها در این فضا است و روش دوم رسم هیستوگرام‌های تصویر داده‌های استیت‌های مختلف بر روی جهت‌های PCA است.

در بخش آخر تمرین با استفاده از SVM تلاش می‌کنیم استیت‌های مختلف خواب را طبقه‌بندی کنیم و با استفاده از Cross Validation طبقه‌بند خود را ارزیابی خواهیم کرد.

۱. جداسازی Wake و REM

با توجه به شکل‌های زیر به نظر می‌رسد PC2 به خوبی می‌تواند بیداری را از REM جدا کند. برای داده‌های بیداری هر سه مولفه حول مبدا و با پراکندگی کم پخش شده‌اند در حالی که داده‌های REM در جهت PC1 واریانس بسیار زیادی دارند. جهت PC3 هیچ اطلاعات مفیدی برای جداسازی REM و Wake در بر ندارد.



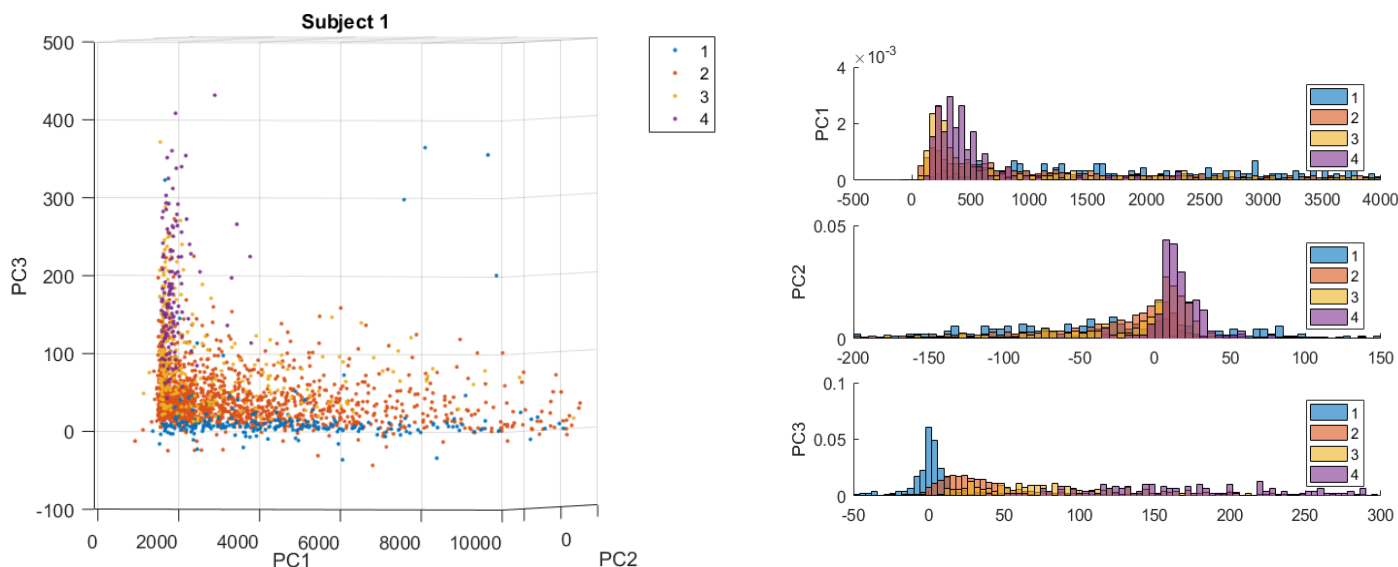
شکل ۸ - جداسازی REM و Wake

۲. جداسازی استیت‌های NREM

با توجه به نمودارهای زیر، با وجود پراکندگی زیاد داده‌های چهار استیت NREM در جهت PC1، این جهت نمی‌تواند برای جداسازی این استیت‌ها و تعیین عمق خواب به کار رود. جهت PC3 می‌تواند استیت‌های NREM را از هم جدا کند.

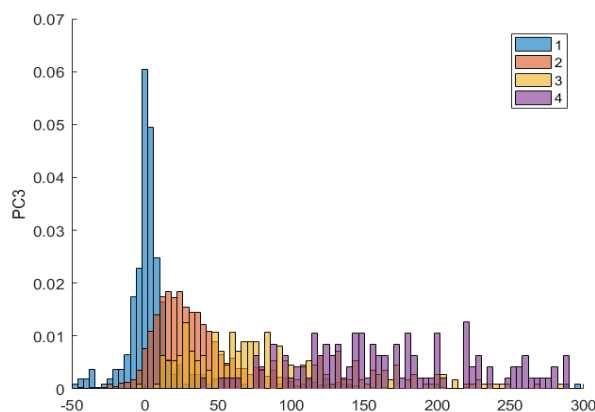
با توجه به شکل سمت راست مولفه PC1 تا حدی استیت ۲ و ۳ خواب را جدا می‌کند. این دو استیت به خوبی بقیه استیت‌ها جدا نمی‌شوند. بررسی دقیق‌تر این موارد در بخش آخر و با بررسی Confusion Matrix انجام خواهد شد.

داده‌های State1 در صفحه $PC1 = 0$ قرار گرفته‌اند در حالی که داده‌های State3 و State4 در جهت PC3 بسیار پراکنده‌اند.



شکل ۹ - جداسازی NREM

یک بار دیگر هیستوگرام تصویر داده‌ها بر جهت سوم را رسم می‌کنیم.



شکل ۱۰ - تصویر داده‌های NREM بر روی PC3

این جهت به‌تنهایی توانسته تا حد زیادی این چهار استیت را از هم جدا کند.

سوال پنج

همان‌طور که مطرح شد، coeff ماتریسی است متعامد که نقاط را از فضای حالت اصلی به فضای حالت PCA می‌برد. این ماتریس برای سابجکت اول به این صورت است. هر ستون متناظر با یک PC است.

	1	2	3	4	5	6	7	8	9	10
1	0.0140	0.0240	0.0409	0.9984	-0.0183	-0.0188	6.6210e-04	0.0015	7.2110e-04	-9.4170e-04
2	4.6627e-04	8.1659e-04	3.2183e-04	0.0228	0.2714	0.9353	0.0946	-0.1978	-0.0330	0.0438
3	2.4770e-04	0.0010	-1.2817e-04	0.0077	0.3531	0.0901	-0.6375	0.6137	-0.2462	-0.1534
4	7.7727e-04	0.0026	-3.9596e-04	0.0094	0.7988	-0.3163	-0.1274	-0.4519	0.1446	0.1429
5	0.0018	0.0412	0.9983	-0.0419	4.2164e-04	-2.6406e-05	-6.2114e-04	-0.0022	-0.0047	-3.7965e-04
6	-3.3286e-05	1.3026e-04	0.0050	8.8295e-04	0.0269	0.0930	-0.0553	0.2664	0.9506	-0.1133
7	9.3234e-05	3.6798e-04	0.0013	0.0014	0.0736	-0.0024	0.1764	0.3731	0.0121	0.9078
8	4.4347e-04	0.0013	6.3644e-04	0.0042	0.3963	-0.0899	0.7310	0.4122	-0.1162	-0.3422
9	0.9993	-0.0358	-8.6844e-04	-0.0131	-6.6384e-04	7.4921e-05	-1.3521e-04	6.4242e-05	4.0612e-05	-1.3112e-05
10	0.0354	0.9982	-0.0422	-0.0228	-0.0028	5.1205e-04	-7.5643e-05	4.6498e-05	1.0608e-04	-9.0200e-05

شکل ۱۱ - نگاشت از فضای حالت اصلی به فضای PCA

ارتباط فضای PC ها با فضای حالت اصلی بدین صورت است.

PC1	تقریباً برابر توان متوسط سیگنال EOG است و دیگر پارامترها تاثیر چندانی بر آن ندارند.
PC2	تقریباً برابر توان متوسط سیگنال EMG است و دیگر پارامترها تاثیر خاصی بر آن ندارند.
PC3	تقریباً برابر توان باند دلتای الکتروود Oz است و دیگر پارامترها تاثیر چندانی بر آن ندارد.
PC4	تقریباً برابر توان باند دلتای الکتروود Fpz است و دیگر پارامترها تاثیر چندانی بر آن ندارد.
PC5	تقریباً برابر $0.8 P_{\beta}(Fpz) + 0.3 (P_{\theta}(Fpz) + P_{\alpha}(Fpz) + P_{\beta}(Oz))$ است.
PC6	تقریباً برابر توان باند تتای الکتروود Fpz است و دیگر پارامترها تاثیر چندانی بر آن ندارد.
PC7	تقریباً برابر $0.7 P_{\beta}(Oz) - 0.6 P_{\alpha}(Fpz)$ است.

شکل ۱۲ - ارتباط فضای حالت اصلی و فضای PC ها

همان‌طور که انتظار داشتیم جهت سیگنال EOG که معیاری از حرکت چشم است، بیشترین واریانس را در داده‌هایمان دارد.

حال نتایج سوال قبل را به زبان جدید ترجمه می‌کنیم.

به طور مثال انرژی باند دلتای الکتروود Oz همان PC3 است که می‌توانست استیت‌های یک تا چهار را از هم جدا کند یعنی عمق خواب را تعیین می‌کند. این نتیجه همان نتیجه‌ای است که افرادی که بر روی سیگنال‌های مغزی کار کلینیکال می‌کنند نیز از آن استفاده می‌کنند. مقاله [4] درباره این موج در هنگام خواب بحث کرده است. هر چه انرژی باند دلتا افزایش یابد، خواب عمق تر می‌شود.

به طور مثال در ویکیپدیا صفحه Delta waves بیان شده است که

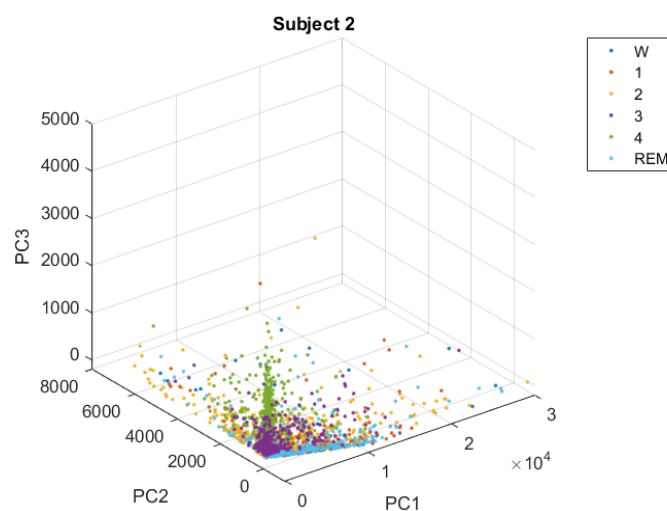
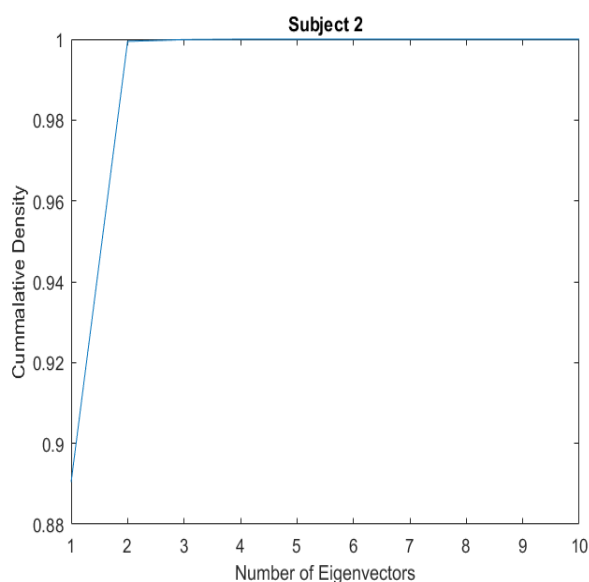
Delta waves, like other brain waves, are recorded with an electroencephalogram[1] (EEG) and are usually associated with the deep stage 3 of NREM sleep, also known as slow-wave sleep (SWS), and aid in characterizing the depth of sleep.

برای جدا کردن سیگنال‌های REM نیز از PC2 استفاده کردیم که توان متوسط EMG است.

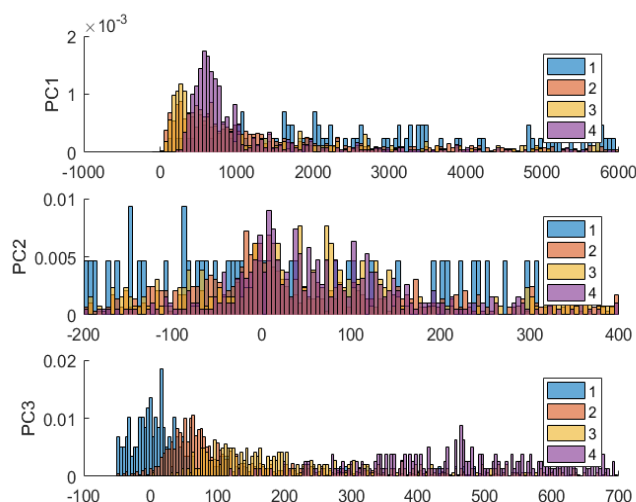
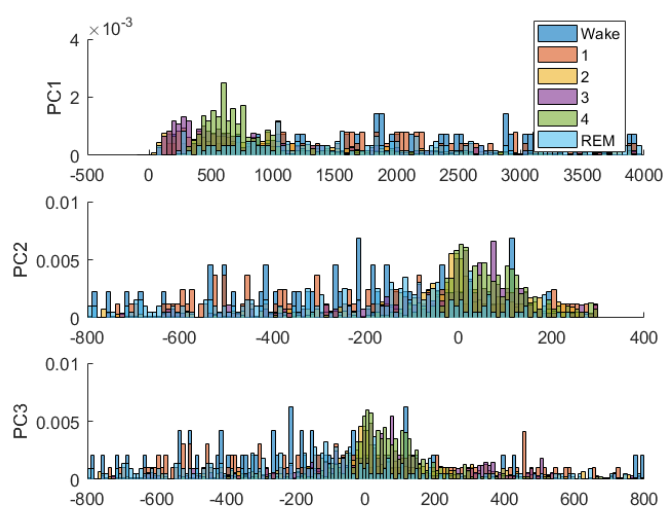
سوال شش

در این بخش به بررسی دیگر سابجکت‌ها می‌پردازیم.

• سابجکت دوم (ST7022J0-PSG)



شکل ۱۳ - نمایش داده‌ها



شکل ۱۴ - تصویر داده‌ها بر روی PCها

	1	2	3	4	5	6	7	8	9	10
1	-5.0876e-05	0.0020	0.1596	0.9866	-0.0173	0.0028	-0.0245	-0.0132	-0.0024	2.7336e-04
2	-1.3453e-05	3.6474e-05	0.0037	0.0278	0.1356	0.0704	0.6769	0.7152	-0.0677	0.0401
3	3.8022e-06	9.1475e-05	0.0032	0.0143	0.4766	-0.1474	0.5718	-0.6081	-0.0352	-0.2301
4	9.3903e-05	2.0811e-04	0.0026	0.0046	0.7464	0.5622	-0.3214	0.1057	0.0452	0.1011
5	0.0049	0.0080	0.9871	-0.1597	-0.0038	0.0030	0.0020	-5.8682e-04	-0.0015	9.5146e-04
6	-8.6441e-07	1.0255e-05	0.0050	0.0064	0.2412	-0.4651	-0.1182	0.1959	0.8089	-0.1374
7	1.9454e-05	6.0780e-05	0.0024	0.0032	0.2926	-0.5511	-0.0954	0.0204	-0.3016	0.7143
8	5.2204e-05	1.0108e-04	0.0026	-2.4291e-04	0.2308	-0.3706	-0.2963	0.2619	-0.4969	-0.6373
9	0.9952	-0.0980	-0.0040	8.9539e-04	-5.1809e-05	-2.9878e-05	3.7237e-05	-1.0560e-05	2.7166e-05	5.2646e-06
10	0.0980	0.9951	-0.0087	-6.2575e-04	-1.8804e-04	-6.2924e-05	6.4309e-05	8.2192e-06	7.6236e-05	1.3331e-05

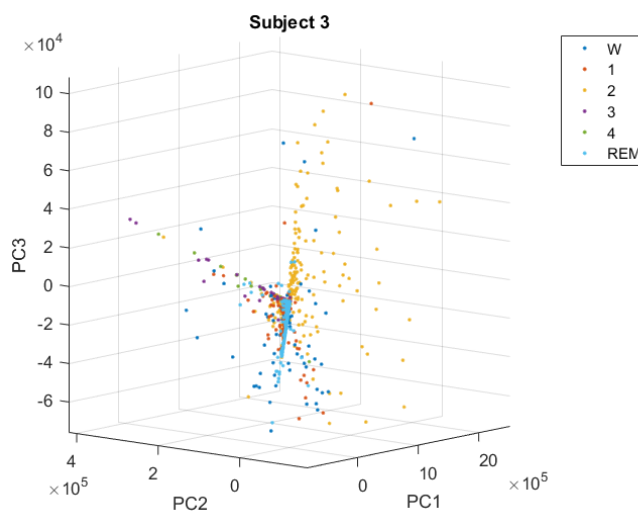
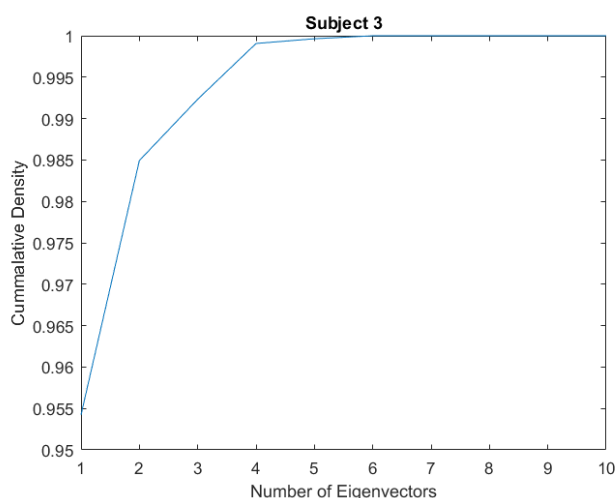
شکل ۱۵ - ماتریس Coeff

در این سابجکت، جهت سوم PCA توانسته عمق خواب را مشخص کند. با توجه به ماتریس Coeff در شکل ۱۵، این جهت متناظر با انرژی باند دلتای الکتروود Oz است که همان نتیجه سابجکت اول است.

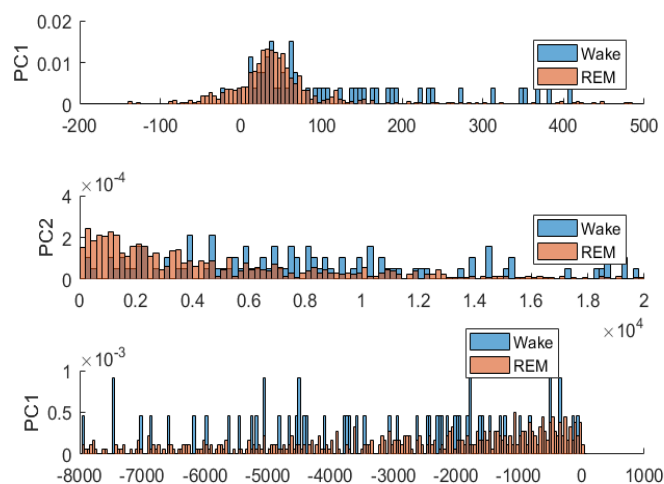
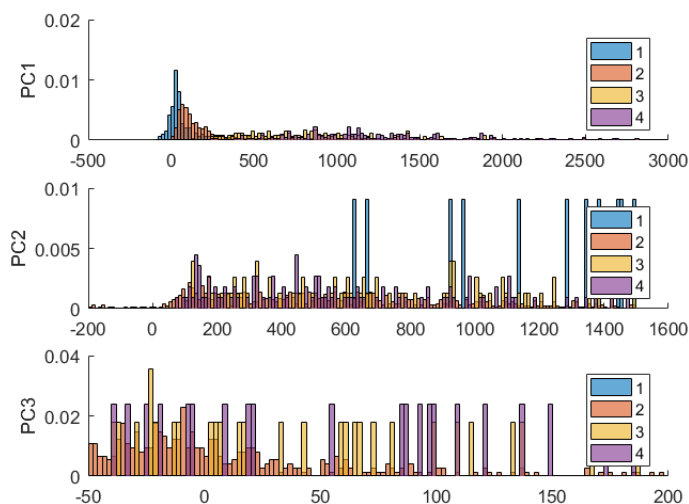
در این سابجکت علاوه بر جهت سوم PCA، جهت اول PCA نیز برای تشخیص عمق خواب موثر است. این جهت تقریباً متناظر با توان متوسط EOG است.

جهت دوم PCA یعنی توان سیگنال EMG در داده‌های REM را جدا می‌کند که باز هم با نتیجه حاصل از سابجکت اول هم‌خوانی دارد.

• سابجکت سوم (ST7041J0-PSG)



شکل ۱۶ - نمایش داده‌ها



شکل ۱۷ - تصویر داده‌ها بر PCها

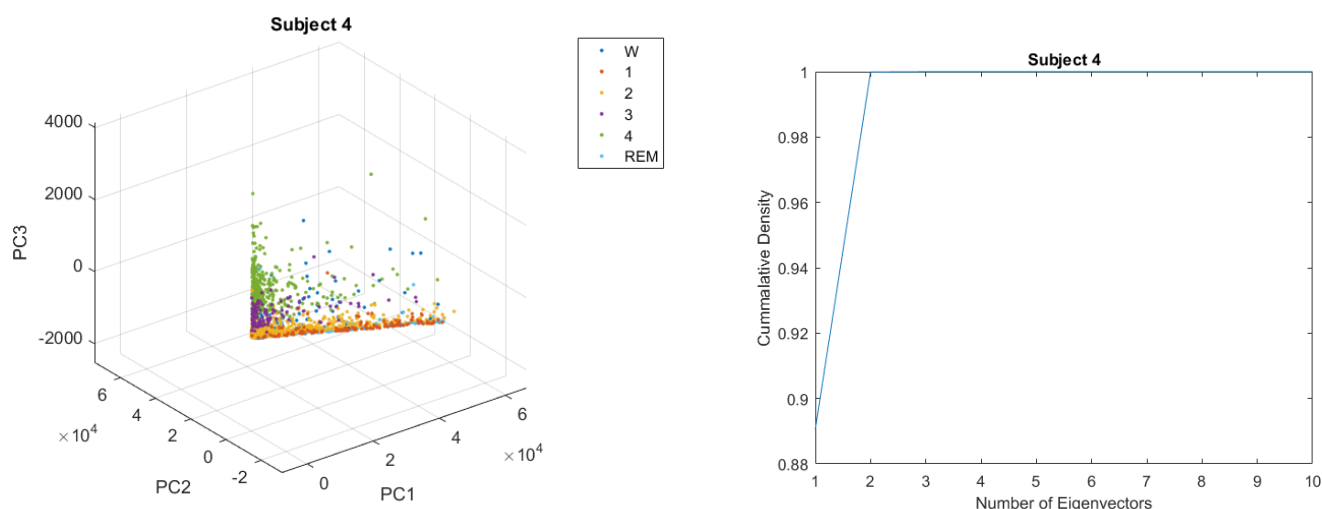
	1	2	3	4	5	6	7	8	9	10
1	3.4754e-04	0.0024	-0.0017	1.0017e-04	-0.0036	0.0089	0.9993	-0.0254	-0.0254	0.0023
2	3.1191e-05	9.8471e-05	2.0722e-05	1.8436e-04	-1.0809e-04	5.7378e-04	0.0314	0.3018	0.9074	-0.2907
3	2.5737e-05	2.7098e-05	4.7218e-05	1.6458e-05	9.6561e-05	1.4750e-04	0.0118	0.3865	0.1619	0.9079
4	2.3078e-05	4.3981e-05	-9.7780e-05	9.2268e-05	-4.6050e-05	5.0017e-04	0.0130	0.8711	-0.3870	-0.3020
5	0.9826	0.0011	-0.1780	-0.0520	-0.0086	0.0063	-7.3275e-04	-2.8162e-05	1.1006e-05	1.7913e-06
6	0.1360	-0.0159	0.7120	0.2206	-0.6371	-0.1404	1.5322e-04	7.4698e-05	-7.3154e-05	2.5405e-05
7	0.1055	-0.0156	0.4895	0.1439	0.7205	-0.4574	0.0075	1.4581e-04	-3.5236e-05	-1.8668e-04
8	0.0697	-0.0086	0.3701	0.1106	0.2735	0.8781	-0.0063	-4.1199e-04	-1.7197e-04	9.2934e-05
9	-0.0012	0.1866	-0.2810	0.9414	0.0069	-3.8669e-04	-9.9450e-04	-1.4011e-04	-9.6983e-05	8.1248e-05
10	0.0036	0.9821	0.0762	-0.1719	0.0022	-0.0018	-0.0021	7.4972e-06	-2.1985e-06	-5.0679e-06
11										

شکل ۱۷ - ماتریس $Coeff$

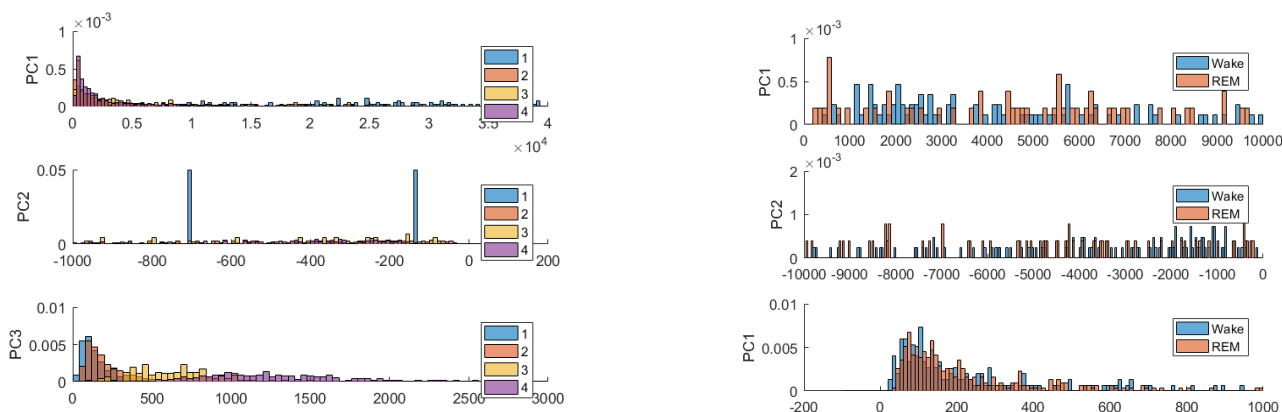
در این سابجکت PC2 داده‌های REM را جدا می‌کند. این جهت تقریباً انرژی سیگنال EMG است. این موضوع توسط افرادی که کار کلینیکال می‌کنند نیز تایید شده و در جدول (۱) در صفحه ۵ این گزارش و همچنین در [2] به آن اشاره شده است. در حالت REM سیگنال EMG به کمترین مقدار خود می‌رسد که با مشاهدات ما برای این سابجکت سازگار است.

جهت PC1 عمق خواب را نیز طبقه‌بندی کرده. این جهت تقریباً معادل انرژی باند دلتای الکتروود Oz است. که برای دیگر سابجکت‌ها نیز عمق خواب را تعیین کرده بود.

• سابجکت چهارم (ST7052J0-PSG)



شکل ۱۸ - نمایش داده‌ها



شکل ۱۹ - تصاویر دیتا برو روی PCها

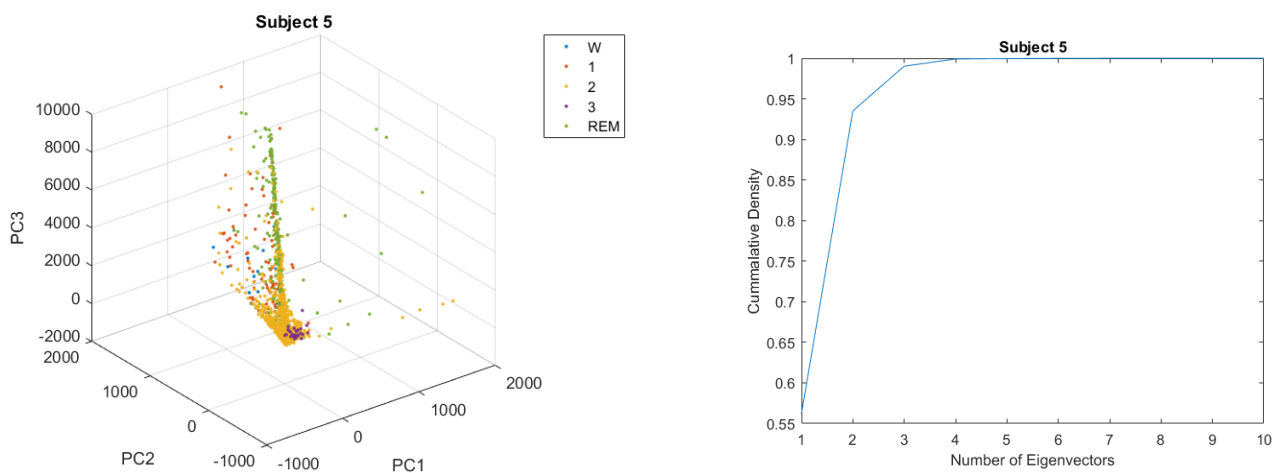
	1	2	3	4	5	6	7	8	9	10	11
1	0.0036	0.0065	0.7745	-0.6323	0.0123	-7.9282e-04	-0.0075	-0.0016	0.0107	0.0019	
2	2.3842e-04	3.2717e-04	0.0338	0.0257	-1.2821e-05	0.5331	0.7635	0.0659	-0.3526	0.0482	
3	1.2045e-04	2.5326e-04	0.0177	0.0261	0.0287	0.7842	-0.5620	-0.0932	-0.0764	-0.2299	
4	8.4333e-05	9.0634e-05	0.0105	0.0162	0.1306	0.0805	-0.0733	0.9732	0.1510	0.0281	
5	0.0039	0.0054	0.6296	0.7696	-0.0691	-0.0630	-0.0287	-7.4033e-04	-0.0416	-0.0034	
6	2.4404e-04	3.1786e-04	0.0383	0.0623	0.2406	0.1935	0.2866	-0.1555	0.8619	-0.2238	
7	1.3522e-04	1.8293e-04	0.0219	0.0419	0.5275	0.1293	-0.1118	-0.1219	0.0479	0.8204	
8	8.2526e-05	9.9982e-05	0.0141	0.0262	0.8006	-0.1901	0.0173	-0.0284	-0.3162	-0.4700	
9	0.7941	-0.6077	9.3495e-04	-5.3775e-04	-7.8638e-07	3.9283e-05	-3.8270e-05	-1.8482e-05	-4.2384e-06	-6.3073e-06	
10	0.6077	0.7941	-0.0099	-5.0196e-04	2.2855e-05	-1.0038e-04	7.8180e-06	-1.0359e-05	2.6183e-05	1.2549e-05	
11											
12											

شکل ۲۰ - ماتریس $coeff$

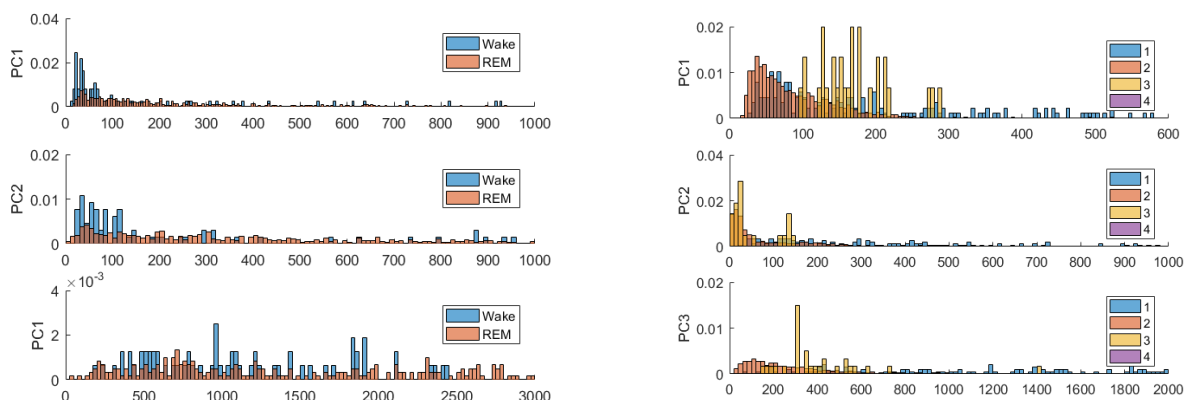
جهت PC3 عمق خواب را مشخص می‌کند یعنی دیتای NREM را جدا می‌کند. این جهت معادل $0.8 P_{\delta}(Fpz) + 0.6 P_{\delta}(Oz)$ است. که با مشاهدات قبلی و انتظارات هم‌خوانی کامل دارد.

جهت PC2 دیتای REM را کم و بیش خوب جدا می‌کند این جهت معادل $0.8 P_{EMG} - 0.6 P_{EOG}$ است.

• سابجکت پنجم (ST7061J0-PSG)



شکل ۲۱ - نمایش داده‌ها



شکل ۲۲ - تصویر دیتا بر روی PC ها

	1	2	3	4	5	6	7	8	9	10
1	0.0011	0.0016	0.0094	0.0038	-0.1916	0.6314	0.7486	0.0309	-0.0558	-0.0020
2	1.5148e-04	3.1554e-04	0.0011	0.0033	-0.0950	0.1701	-0.0937	-0.0663	0.9635	-0.1434
3	1.5095e-04	2.5709e-04	1.6998e-04	0.0029	-0.0642	0.1023	-0.0899	-0.0507	0.1094	0.9812
4	4.2740e-04	0.0012	0.0010	0.0119	-0.3734	0.5826	-0.5923	-0.3112	-0.2380	-0.1290
5	0.9610	-0.2636	-0.0083	-0.0817	-0.0113	-0.0017	-0.0023	0.0118	4.4809e-06	8.2766e-06
6	0.0673	-0.0164	0.0127	0.7404	0.6057	0.2550	-0.0597	-0.1065	-9.8623e-04	-3.4973e-05
7	0.0388	-0.0090	0.0211	0.4713	-0.4921	-0.3958	0.2289	-0.5698	0.0026	-0.0011
8	0.0240	-0.0049	0.0180	0.4712	-0.4488	-0.0268	-0.1259	0.7477	-0.0020	-6.3510e-04
9	0.0208	0.0531	0.9978	-0.0287	0.0130	-0.0012	-0.0082	1.2636e-04	-2.9219e-04	1.6679e-04
10	0.2635	0.9630	-0.0568	-0.0015	4.7703e-04	-0.0018	-1.4707e-04	2.7347e-04	6.3278e-05	-7.0354e-05

شکل ۲۳ - ماتریس Coeff

در این سابجکت PC1 جهت‌های NREM را از هم جدا کرده پس عمق خواب را مشخص می‌کند. بنابر ماتریس coeffs این جهت تقریباً معادل جهت توان باند دلتا در الکتروود Oz است. این نتیجه در تمام سابجکت‌ها عیناً تکرار می‌شد.

جهت PC2 یا به عبارتی سیگنال EMG خواب REM را جدا می‌کند. در این سابجکت جهت PC1 یعنی توان باند دلتا در الکتروود Oz نیز می‌تواند تا حدی REM را جدا کند.

سوال هفت

با توجه به مشاهدات ما بر اساس پنج سری دیتای استفاده شده در این تمرین، انرژی سیگنال دلتا معیار بسیار مناسبی برای طبقه‌بندی خواب به Stage های یک، دو، سه و چهار است بنابراین یک معیار عمق خواب است.

همچنین انرژی متوسط سیگنال EMG می‌تواند معیار خوبی برای جدا کردن دیتای مربوط به خواب REM باشد. توان EOG نیز می‌تواند روش موثری برای جدا کردن خواب REM باشد.

در تمام سابجکت‌های بررسی شده، بیشترین کشیدگی در یکی از جهت‌های زیر رخ داد

- انرژی باند دلتای الکترودهای Oz و یا Fpz
- توان متوسط سیگنال EMG
- توان متوسط سیگنال EOG

بخش چهارم: یافتن معیاری برای عمق خواب مبتنی بر Linear Regression

سوال ۱

جدول زیر نتایج حاصل از برازش خط روی دیتای یازده بعدی سابجکت اول است.

Linear regression model:

State ~ 1 + FpzDelta + FpzTheta + FpzAlpha + FpzBeta + OzDelta + OzTheta + OzAlpha + OzBeta + EOGPower + EMGPower

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	1.7963	0.043173	41.606	2.8496e-276
FpzDelta	0.00058774	3.3155e-05	17.727	1.2676e-65
FpzTheta	-0.00067884	0.00094374	-0.71931	0.47203
FpzAlpha	0.0024949	0.0013203	1.8896	0.058951
FpzBeta	-0.0012259	0.00062065	-1.9752	0.048379
OzDelta	0.00045411	4.0182e-05	11.301	8.6397e-29
OzTheta	0.028254	0.0019841	14.24	5.4496e-44
OzAlpha	-0.0020638	0.0035926	-0.57446	0.56572
OzBeta	-0.010751	0.0065699	-16.364	9.8854e-57
EOGPower	1.6672e-08	4.1874e-07	0.039815	0.96824
EMGPower	-2.4154e-06	1.2795e-06	-1.8878	0.059196

Number of observations: 2105, Error degrees of freedom: 2094

Root Mean Squared Error: 0.581

R-squared: 0.627, Adjusted R-Squared 0.625

F-statistic vs. constant model: 352, p-value = 0

شکل ۲۴ — نتایج Linear Regression بر روی داده‌های خام

Estimate مربوط به هر ستون به این معنی است که به ازای هر واحد افزایش پارامتر آن ستون، انتظار داریم در خروجی چه مقدار تغییر ببینیم. با توجه به این موضوع متوجه می‌شویم که متغیرهای **FpzTheta** و **FpzBeta** و **OzAlpha** و **EMGPower** بر روی عمق خواب تاثیر منفی و بقیه متغیرها تاثیر مثبت دارند. باید توجه داشته باشیم که Estimate ها واحدهای متفاوتی دارند و مقایسه آنها برای یافتن متغیرهای موثر بر عمق خواب بی معناست. در ادامه یک روش برای حل این موضوع ارائه خواهیم داد.

برای ضریب هر متغیر، یعنی Estimate، فرضیه صفر بودن را در نظر گرفته و آزمون فرضیه انجام می‌دهیم. ستون pValue برابر احتمال مشاهده این تحقق خاص یا یک تحقق extreme تر با فرض درست بودن فرضیه است یعنی هر چقدر PValue بیشتر باشد یعنی Estimate عددی بی معنی‌تر است. به طور مثال در جدول بالا مقدار PValue برای EOGPower برابر 0.98 است یعنی مقدار estimate این متغیر عددی بسیار بی معنی است. پس ادعا می‌شود که می‌توان از این بعد از دیتا برای تخمین استیت خواب صرف نظر کرد. این موضوع برای انرژی باند تتای الکتروود **Fpz** نیز برقرار است و به نظر می‌رسد می‌توان از تاثیر آن بر عمق خواب چشم پوشید. SE یک متغیر به این معناست که دیتا به طور متوسط چه میزان در جهت متغیر k از خط فاصله می‌گیرد. بُعد این کمیت همان بعد متغیر مربوطه است.

حال قصد داریم متغیرهای تاثیر گذار در این مدل را پیدا کنیم. به دلیل یکسان نبودن scale و واحد داده‌های جهت‌های گوناگون، نمی‌توانیم از مقایسه مستقیم مقادیر Estimate نظری در این خصوص بدهیم. در کل این کار، کار بسیار سختی است. یک روش برای این کار که در سایت Stat Stack Exchange مطالعه کردیم، حذف تک به تک متغیرها و انجام Cross Validation برای مدل خطی بدون استفاده از آن متغیر است. با این کار می‌توان بفهمیم حذف کدام متغیر اثر مخرب‌تری بر نتیجه Cross Validation دارد و این متغیر را به عنوان موثرترین متغیر معرفی کنیم. این روش به نظر بسیار معتبر است. [7]

۱۰ متغیر داریم پس ده بار عملیات زیر را انجام داده که در هر بار یکی از متغیرها را در آن حذف کرده‌ایم.

ابتدا یک پنجم داده‌ها را به صورت تصادفی به عنوان داده تست انتخاب می‌کنیم و کنار می‌گذاریم. عملیات زیر را ۱۰۰ بار با ۱۰۰ سری داده تست تصادفی انجام داده و حاصل را میانگین می‌گیریم.

با استفاده از داده‌های تست و ۹ متغیر باقی‌مانده، یک مدل خطی بر دیتا فیت می‌کنیم و با استفاده از این مدل خطی، برای داده‌های تست استیت خواب را پیش بینی می‌کنیم. خطای طبقه‌بندی را به این صورت "جمع توان دوم‌های اختلاف استیت پیش‌بینی شده از

$$E = \sum_{Test\ Data} |f - \hat{f}|^2 \text{ یعنی "تعریف می‌کنیم."}$$

متغیرهای مختلف را حذف می‌کنیم و E را در هر کدام گزارش می‌کنیم.

با در نظر گرفتن تمام ویژگی‌ها داریم $E_{initial} = 8.9$

متغیر حذف شده	درصد تغییر E نسبت به حالت اولیه
FpzDelta	8.86 e+5
FpzTheta	7.27 e+5
FpzAlpha	6.88 e+5
FpzBeta	7.50 e+5
OzDelta	4.68 e+5
OzTheta	9.54 e+5
OzAlpha	7.56 e+5
OzBeta	7700
EOG Power	7
EMG Power	6

حذف EOG و EMG تاثیر به مراتب کمتری بر روی نتیجه طبقه بندی می‌گذارند درحالی که حذف بقیه به طرز محسوسی نتیجه را بدتر می‌کند. انرژی باند دلتای الکتروود Fpz و انرژی باند تتای الکتروود Oz بیشترین تاثیر را بر روی این مدل داشته‌اند.

وجود باند دلتای سیگنال در کمیت‌های تاثیر گذار خبر خوبی است! اما می‌بینیم که EOG و EMG که در مدل قسمت قبل بسیار تعیین کننده بودند در اینجا بی اهمیتند. اگر از روش ساده اما غلط مقایسه Estimate ها نیز این کار را می‌کردیم به همین نتیجه می‌رسیدیم.

با توجه به PValue هم می‌شد گفت که قطعاً EOG Power بی اهمیت است چون PValue آن تقریباً 0.9 است!

سوال ۲

R_squared کمیتی آماری است که مقدار نزدیک بودن داده‌ها به مدل خطی را تخمین می‌زند که به صورت زیر تعریف می‌شود:

$$R_Squared = \frac{Explained\ Variations}{Total\ Variations}$$

اگر بخواهیم به صورت ریاضی تعریفی برای آن بکنیم فرض کنید y_i ها داده‌های ما باشند که دارای میانگین \bar{y} هستند و فرض کنید که داده‌ها فیت شده روی مدل خطی به صورت f_i باشند در این صورت تعریف می‌کنیم:

$$R^2 = 1 - \frac{\sum (y_i - f_i)^2}{\sum (y_i - \bar{y})^2}$$

در صورتی که تعریف کنیم $SS_{tot} = \sum (y_i - \bar{y})^2$ و $SS_{reg} = \sum (y_i - f_i)^2$ خواهیم داشت

$$R^2 = \frac{SS_{reg}}{SS_{tot}}$$

صورت همان عبارت Explained variation و مخرج هم Total variations را توصیف می‌کند

ما برای چهار state خواب دارای $R\text{-Squared}=0.6268$ هستیم. اصولاً از $R\text{-Squared}$ به عنوان یک معیار مقایسه می‌توان استفاده کرد و به سؤالاتی از این دست که اگر بر دو سری دیتا خطوطی برازش کنیم، کدام خط به طور نسبی تقریب بهتری از داده‌های مربوط به خود است.

بالا بودن مقدار $R\text{-Squared}$ تنها وقتی مهم است که می‌خواهیم با دقت بسیار بالایی پیش‌بینی کنیم که با توجه به ماهیت تصادفی مساله ما، این $R\text{-Squared}$ کاملاً پذیرفتنی است.

در مساله ما $R\text{-Squared}$ پایین (؟) و نیز مقدار PValue های کوچکی داریم این یعنی پارامترهایی در نظر گرفته شده پارامترهای خوبی هستند و شاید بتوان با افزایش تعداد متغیرها به $R\text{-Squared}$ بهتری رسید اما باید توجه داشت که این افزایش باعث می‌شود که نویز نیز وارد مدل شود و مشکل overfitting برای مدل‌مان پیش بیاید [5]. به هر حال کوچک بودن PValue ها بدین معناست که ارتباطی واقعی بین متغیرهای در نظر گرفته شده و استیت خواب وجود دارد.

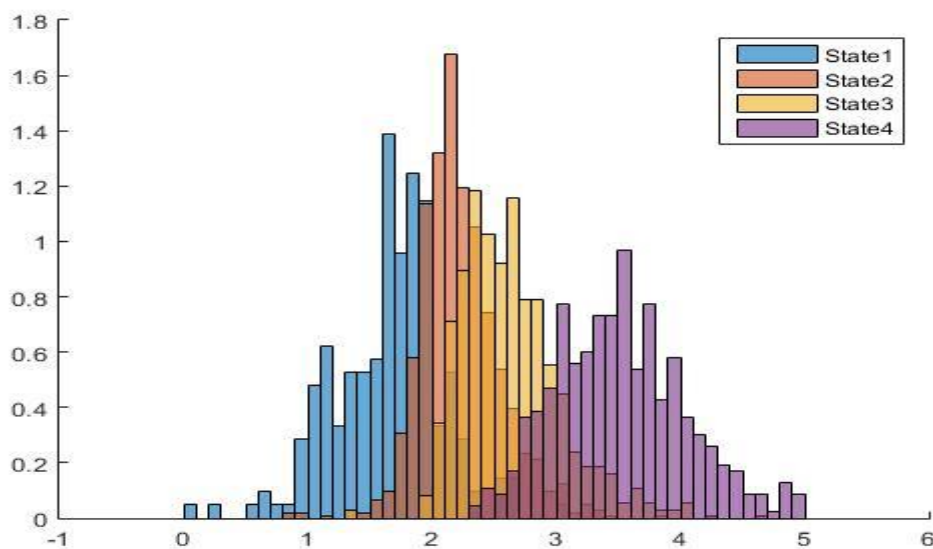
سوال ۳

در حال حاضر مدل ما به توصیف یک پدیده می‌پردازد و کمتر هدف پیش‌بینی دارد. در سوال‌های قبل تلاش کردیم بفهمیم که این با توجه به این مدل، کدام کدام کمیت در تاثیر استیت خواب موثر است و تعدادی متغیر به نظر بی تاثیر را پیدا کنیم. توانستیم به این سوال پاسخ دهیم که با عمیق شدن خواب کدام کمیت‌ها کم و کدام کمیت‌ها زیاد می‌شوند.

Cross Validation به کار گرفته شده در بخش قبل نیز قدرت پیش‌بینی مدل ما را نشان می‌دهد که برای بررسی بیشتر در بخش قبل به آن نیز پرداختیم و داده‌هایی که از هیچ اطلاعاتی از آنها را در طبقه‌بندی استفاده نکرده بودیم را طبقه بندی کردیم. در بخش بعد هم به بررسی توانایی پیش‌بینی مدل‌مان می‌پردازیم و می‌بینیم که با این مدل تا چه حد می‌توان یک طبقه‌بند شبیه به طبقه‌بند Bayes ساخت. در قسمت ۵ تمرین نیز با استفاده از SVM ها با کرنل‌های مختلف، داده‌هایمان را طبقه‌بندی کرده و ماشینی طراحی می‌کنیم که بتواند استیت‌های مختلف خواب را با دقت بسیار بالایی جدا کند.

سوال ۴:

در صورتی که هیستوگرام داده‌ها را رسم کنیم خواهیم داشت

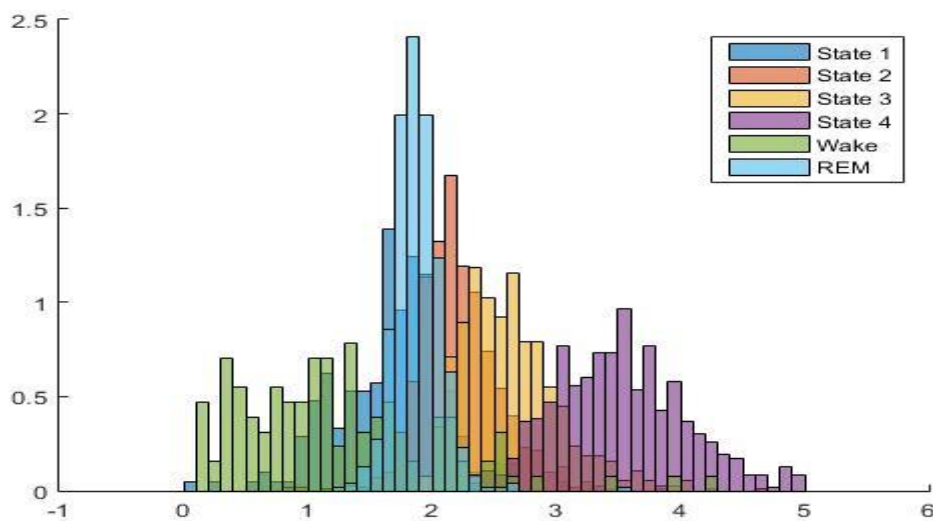


شکل ۲۵: پراکندگی چهار state خواب با روشن مدلاسیون خطی- سابیجت شماره ۲۲

در سوال ۶ هیستوگرام مربوط به دیگر سابیجت‌ها نیز آمده است.

سوال ۵:

در صورتی که دو دسته دیگر را نیز وارد این داده‌ها کنیم پراکندگی داده‌ها به صورت زیر خواهد بود:



شکل ۲۶: پراکندگی ۶ state خواب با روشن مدلاسیون خطی- سابیجت شماره ۲۲

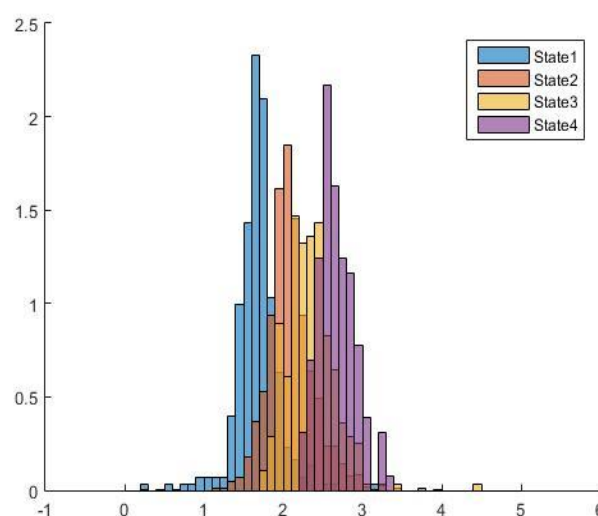
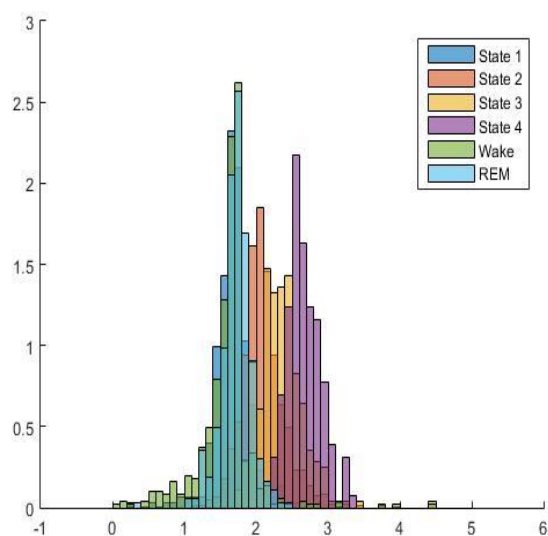
این مدل خطی توانسته داده‌های REM و Wake را از یکدیگر جدا کند همچنین داده‌های Wake از داده‌های خواب کاملاً جدا شده است. داده‌های خواب REM بین استیت ۱ و ۲ قرار می‌گیرند و این موضوع با انتظاراتمان هم‌خوان است زیرا می‌دانیم که خواب REM عمقی شبیه به استیت ۲ دارد.

سوال ۶

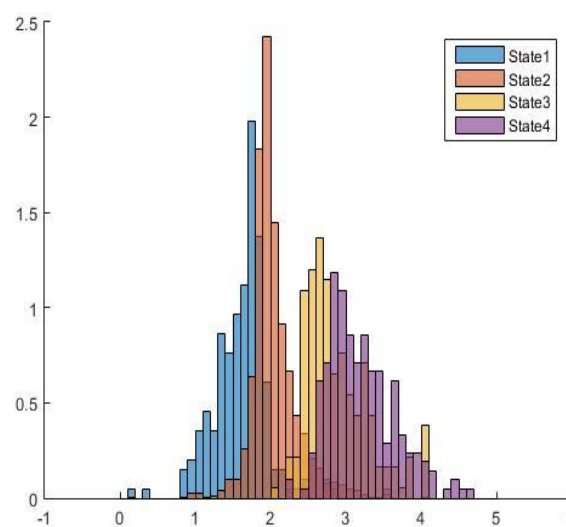
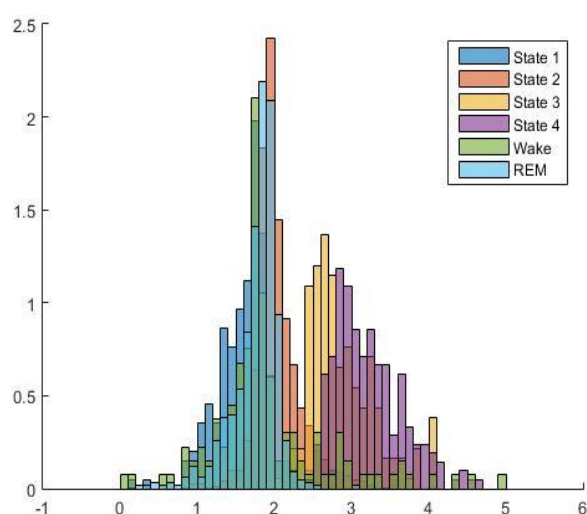
برای هر سابجکت اطلاعات آن مطابق جدول زیر است:

ویژگی‌های مدل خطی/سابجکت	Feature ها با تاثیر منفی	تاثیرگذارترین Feature	Feature های قابل صرف نظر	R_squared
ST7022	OzAlpha – OzBeta - Fpz Alpha – FpzBeta	FpzDelta - OzTheta	EOG and EMG	0.6268
ST7041	FpzAlpha - FpzBeta	FpzAlpha – FpzBeta - FpzDelta	همه به جز ردیف سمت چپ	0.59
ST7052	FpzAlpha – FpzBeta - OzDelta	FpzTheta – OzDelta – OzBeta	EOG and EMG and OzBeta	0.806
ST7061	Oz Delta – Oz Beta - EOG - EMG	FpzDelta - FpzTheta	EOG and EMG	0.347

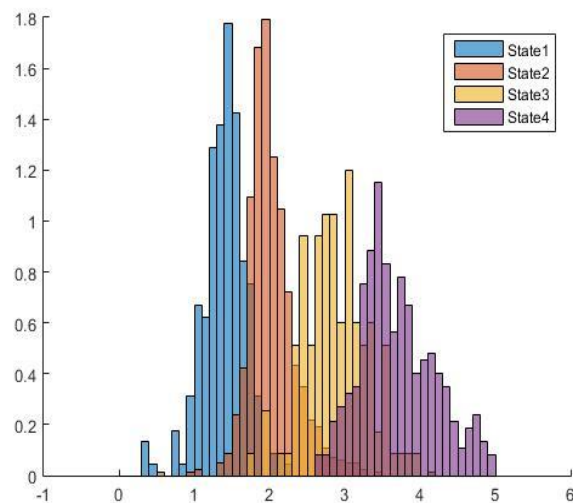
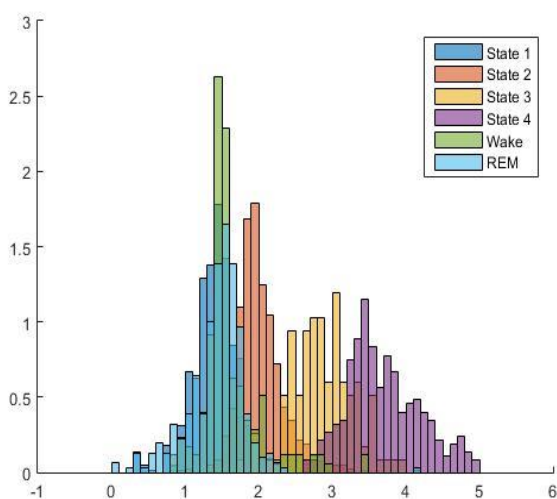
اطلاعات هر یک از سابجکت‌ها برای مدل خطی



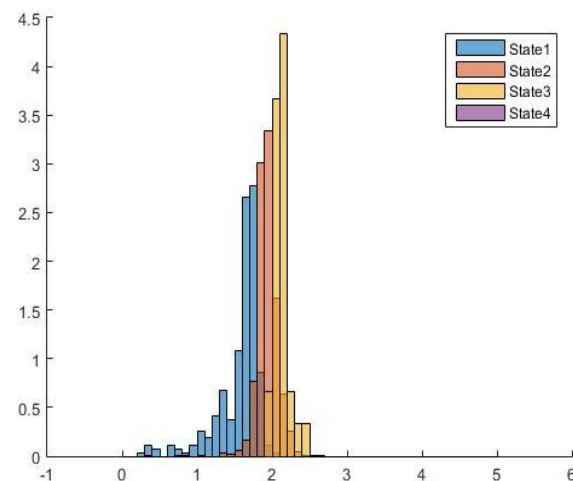
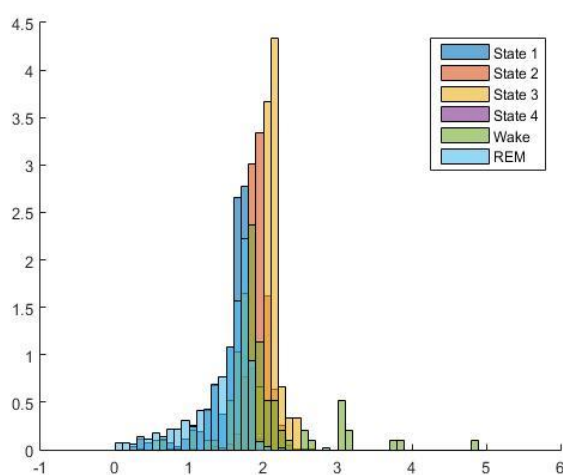
شکل ۲۷: پراکندگی‌های داده شماره ۱۱



شکل ۲۸: پراکندگی‌های داده شماره ۴۱



شکل ۲۹: پراکنندگی‌های داده شماره 52



شکل ۳۰: پراکنندگی‌های داده شماره 61

با توجه به نمودارها و R-squared ها مدل خطی در چهار state خواب NREM برای داده شماره ۵۲ بیشترین مقدار است و داده‌های این سابجکت از سایرین بیشتر به خط نزدیک هستند و برای داده شماره ۶۱ و ۱۱ نتیجه خوبی ندارد ولی برای هر ۵ داده مدل خطی نمی‌تواند حالت REM را از State 1 جدا کند و در بعضی حالت‌ها جداسازی حالت WAKE نیز از REM دشوار است که با مشاهدات ما در داده اول سازگار است.

سابجکت آخر نیز هیچ‌گاه وارد استیت چهارم خواب نشده است. به نظر می‌رسد که این مدل خطی برای سابجکت چهارم (۵۲) بهترین عملکرد را دارد.

سوال ۸

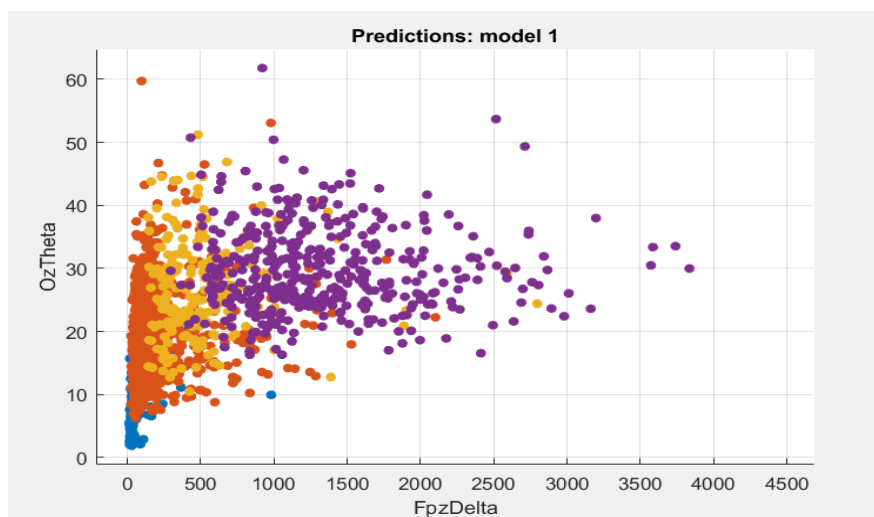
در تمام سابجکت‌ها EOG و EMG تاثیر بسیار کمی نسبت به سایرین داشته است و انرژی باند دلتا تاثیر زیادی داشته است. در روش قسمت قبل EOG، EMG و انرژی باند دلتا پارامترهای موثر بودند. این تفاوت قابل درک است زیرا در این روش Supervised Learning انجام می‌دهیم درحالی که در بخش قبل در حال انجام یک کار Unsupervised بودیم و لزومی به هم‌خوانی کامل این دو روش نیست. با توجه به مدل رگرسیون خطی و با توجه به داده‌های افراد مختلف، به نظر می‌رسد در استیت‌های عمق‌تر خواب انرژی باند آلفا و بتا کاهش می‌یابد.

بخش پنجم: طرح یک سوال

سوال ۱

در این بخش قصد داریم به کمک Support Vector Machine، داده‌های مربوط به عمق خواب را از هم جدا کنیم و سپس با 5-Fold Cross Validation، این طبقه‌بندی خود را ارزیابی کنیم. عملیات را بر داده‌های سابجکت دوم انجام داده و از Linear SVM استفاده خواهیم کرد.

نمودار داده‌ها در فضای OzTheta و FpzDelta به این صورت است. ادعای قبلی مان مبنی بر اینکه انرژی باند دلتا برای جدا کردن استیت‌های NREM مناسب است در این نمودار به وضوح دیده می‌شود.



شکل ۳۱

به کمک تابع fitcecoc متلب یک SVM خطی با دیتای داده‌شده می‌سازیم. تمام فرایند ساخت این ماشین به داخل تابع DepthTrainClassifier برده شده است.

با تابع crossval یک 5-Fold Cross Validation پیدا می‌کنیم. در این Cross Validation داده‌ها به صورت تصادفی به پنج دسته تقسیم می‌شوند و هر بار یک گروه برای تست کردن طبقه‌بند و چهار گروه دیگر برای آموزش طبقه‌بند استفاده می‌شوند. هر دسته دقیقاً یک بار برای آموزش استفاده می‌شود. انتخاب دسته را پنج بار تکرار می‌کنیم و میانگین درصد تشخیص درست Classifier مان را به عنوان دقت آن گزارش می‌کنیم.

در آخر نیز Confusion Matrix را رسم می‌کنیم که نشان می‌دهد که Classifier مان در تشخیص کدام دسته‌ها بیشتر اشتباه کرده است.

True class	1	2	3	4
	129	83	1	
	39	939	57	5
		168	184	29
4		6	48	417
	Predicted class			
	1	2	3	4

شکل ۳۲

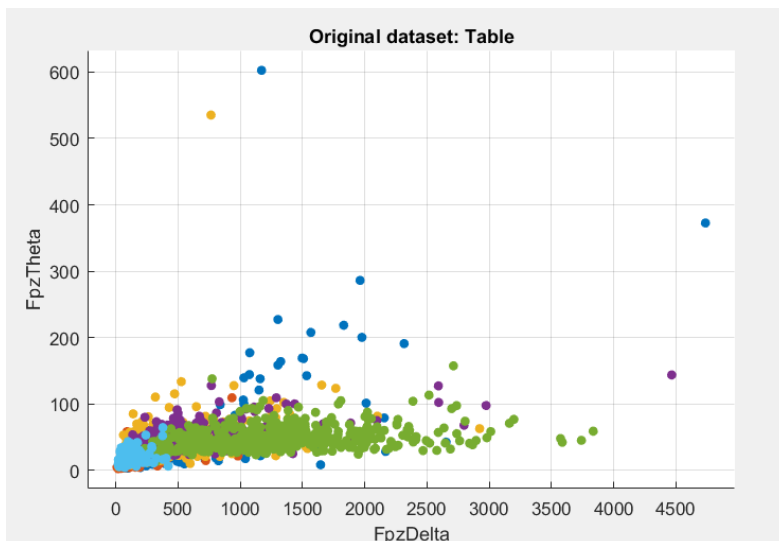
دقت این طبقه‌بندی در 5Fold Cross Validation برابر 79.8% است یعنی در 79.8% دیتای تست را به استیت درست مپ کرده است.

Confusion Matrix مربوطه نیز در شکل مقابل آمده است.

یافتن معیاری برای عمق خواب مبتنی بر سیگنال‌های مغزی ————— درس نوروساینس محاسباتی
 یعنی طبقه بندی ما بیشتر در جدا کردن استیت‌های مجاور خواب دچار مشکل شده است.

سوال ۲

در این بخش عملیات بالا را انجام می‌دهیم با این تفاوت که تلاش می‌کنیم تمام استیت‌های خواب را از هم جدا کنیم، فرایند کاملاً مشابه سوال ۱ است.



شکل ۳۳

دقت این طبقه‌بندی در **5Fold Cross Validation** برابر **73.8%** است یعنی در **73.8%** دیتای تست را به استیت درست مپ کرده است.

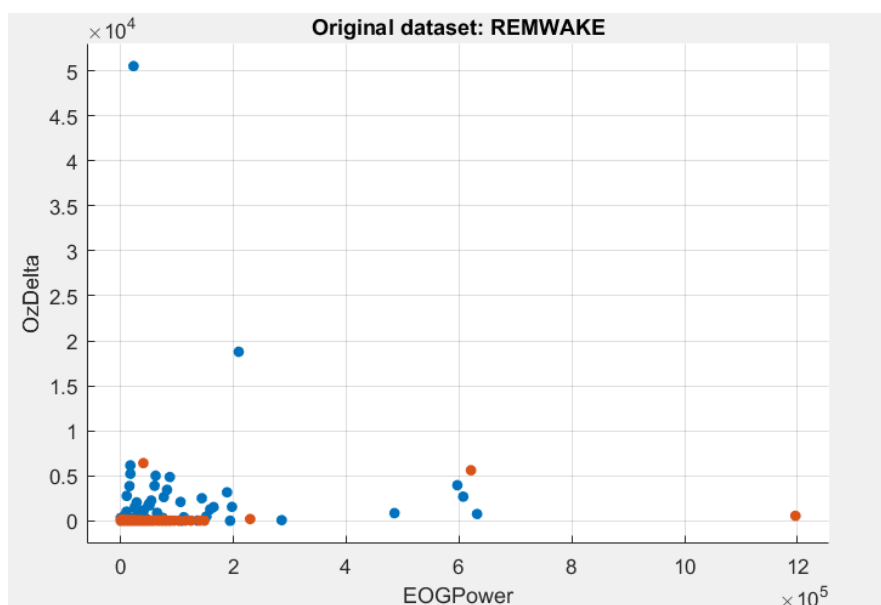
		Model 1					
True class	0	68	48	21	1	1	8
	1	16	65	31	1		100
	2	12	6	848	65	4	105
	3			170	179	31	1
	4			5	45	421	
	6	1	12	37			427
	Predicted class	0	1	2	3	4	6

شکل ۳۴

نتایج ماتریس Confusion بسیار جالب است زیرا دیده می‌شود که Classifier مان در تشخیص REM از State1 به شدت دچار مشکل شده است. این نتیجه معقول است زیرا می‌دانیم استیت یک و REM از نظر عمق خواب تقریباً یکسانند.

سوال ۳

در این بخش همانند دو بخش قبل عمل کرده و دیتای REM و Wake را طبقه‌بندی می‌کنیم.



شکل ۳۵

دقت این طبقه‌بندی در **5Fold Cross Validation** برابر **96.5%** است یعنی در **96.5%** دیتای تست را به استیت درست مپ کرده است.

ماتریس Confusion نیز بدین صورت است



شکل ۳۶

در آخر برای جمع‌بندی، دقت طبقه‌بندی کننده های داده‌های مختلف را در یک جدول کنار هم می‌آوریم.

دیتای مورد طبقه‌بندی	Accuracy
NREM	79.8%
NREM, REM and Wake	73.8%
REM and Wake	96.5%

References

- [1] Kemp, Bob, et al. *"Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG."* IEEE Transactions on Biomedical Engineering 47.9 (2000): 1185-1194.
- [2] K. Šušmáková. *"Human Sleep and Sleep EEG"*, MEASUREMENT SCIENCE REVIEW, Volume 4, Section 2, 2004.
- [3] MathWorks. *"PCA Function Documentation"*, Matlab R2016a.
- [4] Feinberg, I., T. Baker, R. Leder, and J. D. March. *"Response of Delta (0-3 Hz) EEG and Eye Movement Density to a Night with 100 Minutes of Sleep."* Sleep 11.5 (1988): 473-87.
- [5] S. Qureshi and S. Vanichayobon. *"Evaluate Different Machine Learning Techniques for Classifying Sleep Stages on Single-Channel EEG."*, 14th International Joint Conference on Computer Science and Software Engineering, 2017.
- [6] J. Frost, *"How to Identify the Most Important Predictor Variables in Regression Models"*, <http://blog.minitab.com/blog/adventures-in-statistics-2/how-to-identify-the-most-important-predictor-variables-in-regression-models>.
- [7] Stat StackExchange, *"For linear classifiers do larger coefficients imply more important features?"*, <https://stats.stackexchange.com/questions/202221/for-linear-classifiers-do-larger-coefficients-imply-more-important-features>