

## پروژه درس هوش محاسباتی

در این پروژه، هدف پیاده سازی یک سیستم هوشمند برای تشخیص تشنج در بیماران صرعی از سیگنال های EEG است. برای این مهم، یک دیتاست EEG در اختیار شما قرار گرفته است که شامل سیگنال های خام می باشد و شما باید مراحل مختلف پردازش یک داده را بر روی آن پیاده سازی کنید. ابتدا باید سیگنال های داده شده پیش پردازش شوند و سپس از آنها ویژگی هایی را استخراج کنید. پس از آن، باید ویژگی های استخراج شده را بهبود دهید تا به عملکرد دلخواه برسید و نهایتاً، با طبقه بندی سیگنال ها به کمک این ویژگی ها، به نتیجه نهایی (تشخیص تشنج) برسید. این مراحل در فاز نخستین پروژه بوده که در طی دو زیر پروژه انجام خواهد شد. سپس در فاز نهایی پروژه، به پیاده سازی یک سیستم با کمک آموخته های خود در این مرحله و روش های یادگیری عمیق<sup>1</sup> خواهید پرداخت.

### توضیحات دیتاست

دیتاستی که در اختیار شما قرار داده شده است شامل ۵۰۰ سگمنت سیگنالی به طول 23.6 ثانیه بوده که سیگنال ها از یک دستگاه با فرکانس سمپلینگ 173.61 هرتز جمع آوری شده اند. بنابراین شما به ازای هر داده، یک فایل ورودی به طول 4097 خواهید داشت. همچنین ۵۰۰ فایل داده وجود دارند که در پنج گروه مختلف طبقه بندی شده اند و توضیحات این گروه ها و دسترسی به داده ها را در لینک زیر می توانید مشاهده کنید.

[https://www.upf.edu/web/ntsa/downloads/-/asset\\_publisher/xvT6E4pczrBw/content/2001-indications-of-nonlinear-deterministic-and-finite-dimensional-structures-in-time-series-of-brain-electrical-activity-dependence-on-recording-regi?inheritRedirect=false&redirect=https%3A%2F%2Fwww.upf.edu%2Fweb%2Fntsa%2Fdownloads%3Fp\\_p\\_id%3D101\\_INSTANCE\\_xvT6E4pczrBw%26p\\_p\\_lifecycle%3DO%26p\\_p\\_state%3Dnormal%26p\\_p\\_mode%3Dview%26p\\_p\\_col\\_id%3Dcolumn-1%26p\\_p\\_col\\_count%3D1](https://www.upf.edu/web/ntsa/downloads/-/asset_publisher/xvT6E4pczrBw/content/2001-indications-of-nonlinear-deterministic-and-finite-dimensional-structures-in-time-series-of-brain-electrical-activity-dependence-on-recording-regi?inheritRedirect=false&redirect=https%3A%2F%2Fwww.upf.edu%2Fweb%2Fntsa%2Fdownloads%3Fp_p_id%3D101_INSTANCE_xvT6E4pczrBw%26p_p_lifecycle%3DO%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-1%26p_p_col_count%3D1)

---

<sup>1</sup> Deep Learning

## پیش پردازش و لود کردن دیتاست

قدم اول برای انجام هر پروژه دیتا محور، ایجاد یک ساختار مناسب برای دسترسی های متوالی به داده هاست. شما در طول روال پیاده سازی یک سیستم، مکررا نیاز به دسترسی به دیتاست خواهید داشت و در صورت عدم پیاده سازی یک راهکار مناسب برای لود کردن داده ها، زمان زیادی را به I/O هدر خواهید داد. برای این مهم، در کلاس حل تمرین توضیحات کامل داده خواهد شد و نیازی به پیاده سازی از سوی شما نیست.

بعد از لود کردن داده ها، باید داده ها پیش پردازش شوند. روال یک پیش پردازش عموما شامل بررسی داده ها، پیدا کردن محل اطلاعات اصلی در داده ها و محل نویز هاست. پیش پردازش داده ها به شکل عمومی مربوط به درس داده کاوی و به شکل خاص تر در این دیتاست مربوط به درس پردازش سیگنال است و لذا خواسته این پروژه از شما این مهم نخواهد بود و در کلاس حل تمرین کد مناسب پیش پردازش نیز در اختیار شما قرار خواهد گرفت.

## مهندسی ویژگی<sup>2</sup> - تکلیف اول

اولین قدمی که شما باید در انجام این پروژه پیش ببرید استخراج ویژگی های سیگنال هاست. ساخت یک بازنمایی مناسب از داده قدمی اجتناب ناپذیر در ساخت هر سیستم هوشمندی بوده و بدون آن، الگوریتم هایی که در درس یادگرفته اید کارایی خاصی نخواهند داشت. در تکلیف اول، شما باید بخش مهندسی ویژگی را جلو ببرید. در کلاس حل تمرین به شما توضیحات کافی از انواع دسته ویژگی های عمومی و اختصاصی داده خواهد شد و چندین گروه به شما معرفی می شود. انتظار می رود که حداقل ۱۵ ویژگی از ۳ دسته مختلف را پیاده سازی کنید. (همچنین علاوه بر کلاس، لیست این دسته ها به شرح زیر است)

- آماری
- ابتکاری
- آنتروپی ها
- LBP based features
- Time domain
- Frequency Domain

---

<sup>2</sup> Feature Engineering

## طبقه بندی سیگنال ها - تکلیف اول

پس از استخراج ویژگی ها، نوبت به طبقه بندی آنها میرسد. این دیتاست شامل تعداد زیادی مسئله است که توضیحات آنها نیز داخل کلاس حل تمرین داده خواهد شد. در خصوص نحوه پیاده سازی الگوریتم های مختلف نیز توضیحاتی داده می شود. شما در این بخش باید سه الگوریتم مختلف را پیاده سازی کنید، و سپس به ازای هر کدام، معیار های مختلف ارزیابی را بدست آورید. همچنین برای ارزیابی عملکرد روش ها، از k-fold cross validation با مقدار k برابر با 5 استفاده کنید.

الگوریتم هایی که باید استفاده کنید عبارتند از:

- Svm
- Random forest
- KNN

تاثیر تغییر طبقه بند، و تغییر پارامتر های مختلف هر طبقه بند را تست کرده و گزارش دهید. برای هر کدام، پارامتر های

- Accuracy
- Precision
- Recall

را گزارش کنید و سپس برای بهترین، نمودار recall و ماتریس گمراهی را نیز محاسبه کنید. همچنین تاثیر نرمال سازی داده های ورودی را برای هر الگوریتم تست کنید.

## نحوه تحویل

برای تحویل هر فاز شما باید یک گزارش کامل شامل تحلیل و ارزیابی بخش های مختلفی که انجام داده اید و نتایج ارائه دهید. ارائه حضوری نیز احتمالاً برای کل فاز های پروژه یکجا و در انتهای ترم خواهد بود، بنابراین سعی کنید کامنت گذاری مناسب در کد داشته باشید که در زمان ارائه برای توضیح بخش های مختلف دچار مشکل نشوید.

## رفع اشکالات

برای رفع اشکالات، در هفته هایی که کلاس حل تمرین نیست یک روز مشخص خواهیم کرد و من در اون روز در تایم مشخصی داخل دانشکده خواهم بود که برای رفع اشکال می توانید مراجعه کنید. همچنین داخل تلگرام نیز می توانید پیام بدهید.

