


<p>دانشکده مهندسی برق و کامپیوتر</p> <p>گروه کنترل</p> <p>زمان تحویل تمرین: 25/11/1400 ساعت: 23:55</p>	<p>به نام خدا</p> <p><b>پروژه پایانی</b></p> <p>شناسایی سیستم دکتر امین رضانی</p>	 <p>دانشگاه تربیت مدرس Tarbiat Modares University</p>
--	---	--

## مقدمه

در طول درس با روش ها و مدل های متفاوتی برای کاربرد های طبقه بندی و خوشه بندی آشنا شده اید. در این مرحله همواره چالش انتخاب مدل بهینه برای حل یک مسئله با روش های یادگیری ماشین وجود دارد. در این پروژه قصد داریم در ابتدا با یادگیری دانش اولیه نسبت به یک موضوع اطلاعات اختصاصی فیلد را بدست آورده و روش های مختلفی که در درس وجود داشته است را با یکدیگر مقایسه کنیم و مناسب ترین روش را انتخاب کنیم.

## دانش اولیه از موضوع

در حل یک مسئله، مطالعه تحقیقات انجام شده در آن موضوع اهمیت بسیاری دارد. مطالعه مقاله های مربوطه از چند جهت در پیشبرد تحقیق شما مفید خواهد بود. در درجه اول با اطلاع از تحقیقات پیشین مطمئن خواهید شد که روش/ایده/مدل شما قبلاً توسط محقق دیگری امتحان و بررسی نشده باشد. علاوه بر آن در انتهای مطالب علمی گاهی نویسندگان مسیری برای ادامه و پیشبرد تحقیقات ارائه می دهند که این پیشنهادات می تواند در شکل دهی ایده های شما و مسیری که برای تحقیقات خود انتخاب می کنید بسیار مفید باشد. در این قسمت از پروژه، از شما خواسته می شود تا مقاله ای که در اختیار شما قرار داده شده است را مطالعه کنید و خلاصه ای از آن تهیه کنید. در این خلاصه باید چند قسمت اصلی متداول را حتماً ذکر کنید. لذا خلاصه شما باید پاسخگو و شامل قسمت های زیر باشد.

۱. خلاصه ای از مقدمه
۲. در تحقیق از چه مجموعه داده ای استفاده شده است؟
۳. آیا داده های توسط خود نویسندگان جمع آوری شده است یا خیر؟
۴. از چه روش هایی برای پیش پردازش داده ها و انتخاب ویژگی ها استفاده شده است؟
۵. از چه مدل هایی برای طبقه بندی/خوشه بندی/درون یابی استفاده شده است؟
۶. عملکرد مدل ها با چه اطلاعاتی گزارش شده است؟ آیا این گزارش دقیق است یا خیر؟ در صورتی که پاسخ منفی است، شیوه بهتری برای گزارش عملکرد مدل پیشنهاد دهید.

۷. نتیجه گیری و دست آورد های پژوهش

## معرفی مجموعه داده

در سال‌های اخیر استفاده از سیستم‌های تشخیص و کنترل از راه دور برای تشخیص بیماری پارکینسون بر اساس اندازه‌گیری اختلالات در عملکرد قسمت موتور کورتکس مغز افزایش یافته است. در حدود ۹۰ درصد از بیماران، نشانه‌های اختلال در صحبت کردن در همان مراحل ابتدایی بیماری آشکار می‌شود. بنابراین تحلیل و پردازش صوت یک روش مناسب برای تشخیص زود هنگام بیماری می‌تواند باشد.

مجموعه داده این پروژه ویژگی‌های استخراج شده از صدای ۱۸۸ بیمار پارکینسونی (۱۰۷ مرد و ۸۱ زن) و ۶۴ فرد سالم (۲۳ مرد و ۴۱ زن) را در بر دارد. الگوریتم‌های استخراج ویژگی متفاوتی بر روی صوت‌های جمع‌آوری شده اعمال شده تا اطلاعات بالینی مناسبی برای تشخیص بیماری پارکینسون بدست آید.

---

## معرفی مسئله مورد بررسی

در ادامه با استفاده از مجموعه داده معرفی شده در قسمت قبل، مدل‌هایی برای تشخیص بیماری ارائه و یک مسئله طبقه‌بندی در دنیای واقعی را حل کنید. با توجه به این که کارایی مدل نهایی ارائه شده اهمیت به سزایی در تشخیص دقیق دارد، روش‌های مختلف پیش‌پردازش، انتخاب ویژگی و طبقه‌بندی را مقایسه کنید و در نهایت بهترین پروسه (روش کاهش بعد و طبقه‌بندی) پردازش اطلاعات برای مجموعه داده را پیشنهاد دهید.

دقت کنید که هدف اصلی آن است که میان گزینه‌های موجود روشی را بیابید که عملکرد بهتری را ارائه دهد. بنابراین باید با کوشش و خطا همراه با استدلال و شهودی که در مورد مدل‌ها دارید (این مدل را پیدا کنید. برای سنجش عملکرد مدل می‌توانید از معیاری‌های متفاوتی که در زیر به آنها اشاره شده است استفاده کنید.

accuracy

confusion matrix

ROC curve

AUC

F1 Score

precision

Recall

دقت شود که لزومی به استفاده از تمامی این متریک‌ها نیست! بلکه لازم است تعدادی را انتخاب کنید و مدل‌های مختلف را بر اساس این متریک‌ها با یکدیگر مقایسه کنید. برای اطلاعات بیشتر در مورد هر یک از متریک‌ها می‌توانید به [این منبع](#) مراجعه کنید.

---

## روش‌های پیش‌پردازش و انتخاب ویژگی

۱. مجموعه داده را یکپارچه سازی کنید.

۲. با توجه به این که در این مجموعه داده تعداد نمونه‌های دو کلاس نامتوازن است، چه روشی را برای مواجهه با این مشکل انتخاب میکنید؟

روش‌های متفاوتی مانند استفاده از متریک مناسب، ایجاد سَمپل جدید برای داده یادگیری و ... را می‌توان استفاده کرد و یا با توجه به جنس مساله از عدم توازن چشم‌پوشی کرد. (به دلخواه روشی را انتخاب کنید و توجه داشته باشید انتخاب شما، مقایسه عملکرد طبقه‌بندها را در قسمت بعد تحت تاثیر قرار می‌دهد.

۳. از ابعاد مجموعه داده مشخص است که با نحسی ابعاد مواجه هستیم. برای رفع این مشکل از روشهای زیر برای کاهش ابعاد استفاده کنید.

- LDA
- ICA
- PCA with Whitening
- PCA without Whitening
- Sequential Backward Feature Elimination

---

## روش های طبقه بندی

از بین روش های طبقه بندی که در درس یاد گرفتید و یا هر مدل دیگری که می شناسید دو مدل را انتخاب کنید و داده ها را طبقه بندی کنید. همانطور که پیش از این بیان شد، باید عملکرد این مدل ها را با استفاده از متریکی که در قسمت های پیش انتخاب کرده اید را بسنجید و در گزارش بیاورید. همچنین عملکرد مدل ها را مقایسه کنید و بیان کنید که از کدام برای طبقه بندی نهایی استفاده خواهید کرد.

**\*\* نکاتی در مورد عملکرد قابل قبول برای مدل نهایی:**

با توجه به انتخاب های متعددی که در قسمت های پیشین وجود دارد، ارائه یک دقت (متریک عملکرد) ثابت مناسب نخواهد بود. مهم ترین قسمت آن است که به درستی مدل های مختلف را آموزش دهید و با یکدیگر مقایسه کنید و در نهایت بهترین مدلی که بدست آورده اید را ارائه دهید. دقت این مدل باید بیشتر از ۸۰ درصد روی داده های آموزش باشد.

---

### چند نکته مهم :

- ۱- گزارش شما (pdf و word) باید شامل روش های بکار گرفته شده و روش انجام محاسبات و کد نتایج و جمع بندی باشد.
- ۲- برنامه شما باید با پایتون نوشته شود و سایر زبان ها مورد قبول نیست.
- ۴- حجم گزارش به هیچ عنوان معیار نمره دهی نیست. دقت کنید که گزارش شما تعیین کننده میزان یادگیری و تفاوت کد شما با منابع دیگر را می رساند لذا هرگونه مشابهت و تقلب با دیگر افراد نمره تمرین مربوطه و یا کلیه تمرین صفر خواهد شد.
- ۵- فایل گزارش و کدها حتما به صورت زیپ و به فرمت "**StudentNumber\_YourName\_PRJ\_SI**" نام گذاری شود.
- ۶- دقت کنید پروژه دارای تحویل آنلاین می باشد. لذا باید بر روی کد و گزارش مسلط باشید.
- ۷- در صورت وجود سوال و یا مشکل می توانید با آموزشیاران درس سوال خود را مطرح کنید.

شاد باشید و امیدوار