



گزارش تمرین سری 4 یادگیری در مغز و ماشین

استاد درس : دکتر ابراهیم پور

دانشجو : سعید منصورلکوج - 99102304

توضیحات کلی :

این تمرین را به کمک نرم افزار Matlab انجام دادم و برای هر subject نیز یک فایل متلب جداگانه درست کردم که از روی نام گذاری ها کاملا مشخص می باشد.

همچنین 3 تا function مشترک میان تمام فایل ها وجود دارد.

`:getDataBeforeTime`

این تابع داده و زمان رخداد صرع را می گیرد و 9 دقیقه ی ابتدایی را از 10 دقیقه آخر برمیگرداند.

`:getMatrix`

این تابع دیتا با فرمت timetable را میگیرد و به صورت ماتریکس برمیگرداند.

دقت شود فرم اولیه ی دیتا مانند زیر می باشد:

3600x23 timetable

	Record Time	1 SignalLabel1_FP1_F7	2 SignalLabel2_F7_T7	3 SignalLabel3_T7_P7	4 SignalLabel4_P7_O1	5 SignalLabel5_FP1_F3	6 SignalLabel6_F3_C3
1	0 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
2	1 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
3	2 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
4	3 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
5	4 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
6	5 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
7	6 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
8	7 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
9	8 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
10	9 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
11	10 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
12	11 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
13	12 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
14	13 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double
15	14 sec	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double	256x1 double

`:getFeature`

این تابع یکی از مهم ترین توابع می باشد که دیتا را میگیرد و طبق توضیحات مقاله ویژگی ها را استخراج می کند:

یعنی 5 ویژگی را برای هر epoch استخراج میکند که این ویژگی ها شامل entropy آن epoch و همچنان میانگین، انحراف معیار، ماکس و مین می باشد.

`:T1TestSelection`

این تابع همانطور که از اسمش نیز مشخص می باشد مربوط به بخش *selection* مقاله می باشد که طبق مقاله و با استفاده از تابع آماده *ttest2* فرآیند انتخاب ویژگی ها را انجام میدهد.

در ادامه نیز همانطور که قبلا توضیح دادم برای هر *subject* یک فایل جداگانه ی متلب داریم که محاسبات مربوط به هر سوژه را در آن انجام می دهیم (محاسبات مشابه است) و در نهایت عمل *classification* را انجام می دهیم.

دو به یک:

در این بخش به تعداد نمونه های دارای صرع سمپل 9 دقیقه ای از تنها فایل بدون صرع نمونه بر میدارم. (گاهی اوقات برای اینکه زمان داده کم است ممکن است تداخل را هم در نظر بگیرم)

در این بخش مقدار p -value را 0.01 در نظر میگیرم و همچنین مقدار k برای KNN را برابر 3 در نظر می گیریم.

در روش دو به یک تقریباً تعداد نمونه های ترین دو برابر تعداد نمونه های تست می باشد:

Subject 1:

	Accuracy	FP rate	TP R
KNN	1	0	1
SVM	0.75	0	0.5

Subject 3:

	Accuracy	FP rate	TP R
KNN	1	0	1
SVM	1	0	1

Subject 8:

	Accuracy	FP rate	TP R
KNN	1	0	1
SVM	0	1	0

نتیجه ی بسیار بدی برای SVM بدست آوردیم با این مقدار p -value، 373 تا ویژگی انتخاب شده است که می خواهیم این مقدار را بیشتر کنم ببینم چه تفاوتی می کند پس مقدار p -value را برابر 0.1 قرار میدهم تا 1254 تا ویژگی انتخاب شود.

نتیجه :

	Accuracy	FP rate	TP R
KNN	1	0	1
SVM	1	0	1

مشاهده می شود که نتیجه خیلی تغییر کرد و می تواند نشان دهنده ی این باشد که SVM به صورت کلی آنقدر روش خوبی نیست یا تعداد ویژگی ها کم بود.

Subject 14:

در این بخش در داده های ورودی بعضی کانال های مقدارشان NaN بود که این مورد را handle کردم.

	Accuracy	FP rate	TP R
KNN	0.75	0	0.5
SVM	0.75	0	0.5

Subject 16:

	Accuracy	FP rate	TP R
KNN	1	0	1
SVM	1	0	1

نتیجه:

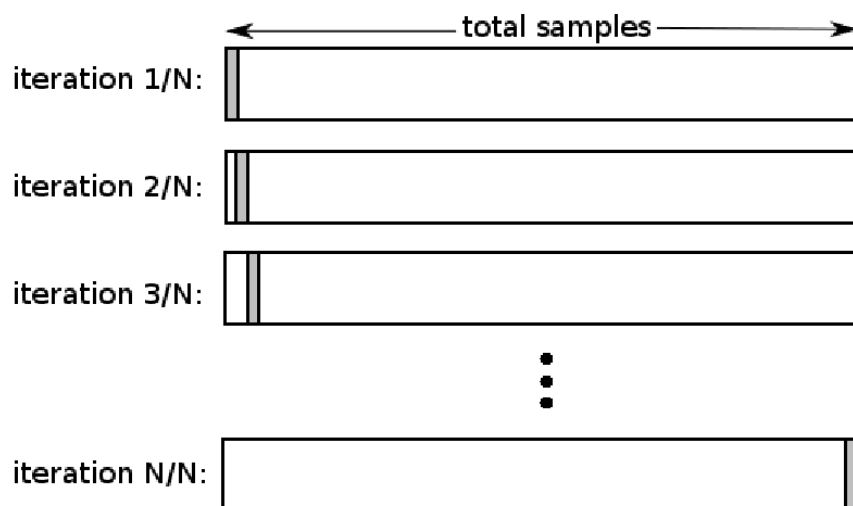
به صورت کلی یک دید خوبی از مسئله پیدا کردیم و از روی **subject8** به این نتیجه می رسیم که بهتر است **p-value** را کمی بیشتر از 0.01 در نظر بگیریم. (اما در ادامه هم مقدار را تغییر ندهیم تا تمامی نتایج با یک **p-value** باشد و اگر نتیجه ی نهایی خیلی بد بود آنرا با **p-value** های دیگر امتحان می کنیم)

اما این روش یک مشکل دارد.

وقتی تعداد داده های ما خیلی کم می باشد استفاده از این روش و **split** کردن داده ها به **test** و **train** و سپس تخمین از خوبی مدل به کمک تحلیل **test** روش مناسبی نمی باشد بلکه بهتر است از روش های دیگری استفاده شود.

:Leave one out

این روش در واقع همان **k-fold** می باشد که تعداد **fold** ها به اندازه ی تمام داده ها می باشد و در آخر تمامی نتایج بدست آمده با یکدیگر **concatenate** می شوند.



در این صورت نتایج خواسته شده به صورت زیر می شوند:

Subject 1:

	Accuracy	FP rate	TP R
KNN	0.92	0.17	1
SVM	0.75	0.17	0.67

Subject 2:

	Accuracy	FP rate	TP R
KNN	1	0	1
SVM	0.9	0.2	1

Subject 8:

	Accuracy	FP rate	TP R
KNN	0.92	0	0.83
SVM	0.92	0	0.83

Subject 14:

	Accuracy	FP rate	TP R
KNN	0.92	0	0.83
SVM	0.92	0	0.83

Subject 18:

	Accuracy	FP rate	TP R
KNN	1	0	1
SVM	1	0	1

نتیجه:

مشاهده می شود که دقت بسیار بالا می باشد (دقت شود که به صورت میانگین تعداد ویژگی های برای subject ها کمتر از 200 می باشد، نکته ی جالب دیگر این است که در همه ی موارد دقت KNN و SVM (SVC) مشابه هم می باشد که نشان دهنده ی این است که موردی که در بخش های قبلی داشتیم به خاطر روش split کردن و همچنین تعداد کم داده ها رخ داده بود که در اینجا این مورد اصلا مشاهده نشده است. (البته به صورت کلی KNN بهتر می باشد)

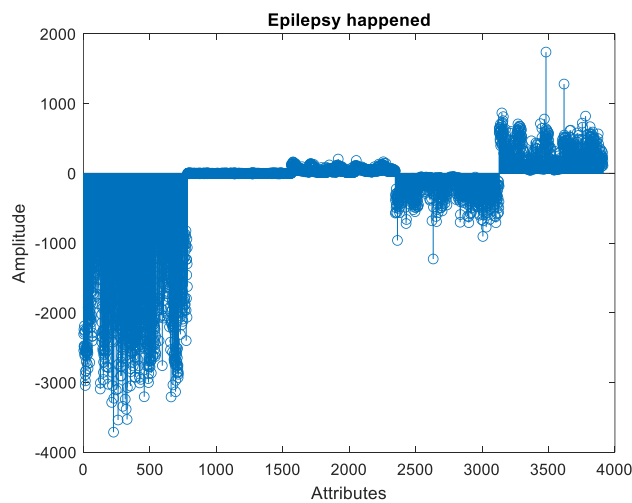
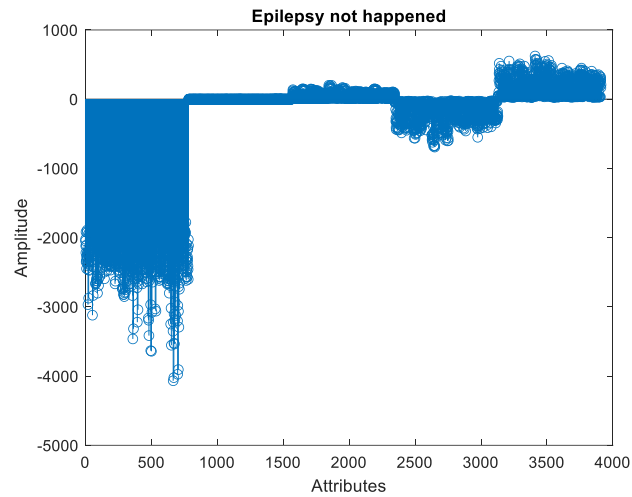
همچنین دقت شود که یکی از ایراد های جزئی این مدل می تواند این باشد که نرخ TP کمی از FP بهتر می باشد که معمولا در بیماری ها بهتر است که مدل طوری باشد که 1 TP باشد تا سبب مرگ بیماران یا مواردی مانند این نشویم. (البته باز اینجا مقادیر مقدار خوبی داشتند)

به صورت کلی میتوان گفت که نتایج عالی بوده است و مدل ما تقریبا مشکل خاصی نداشته است، نه از نظر دقت و نه از نظر محاسباتی و زمانی، میتوان گفت تنها مشکل مربوط به موارد بیولوژیک می باشد که حمله را تنها 1 دقیقه قبل از حمله می توان پیش بینی کرد که البته چون این افراد در بیمارستان حضور دارند چندان مشکل خاصی نمی باشند.

گزارش اصلی تمرین اینجا به پایان می رسد و در ادامه چندین نمودار مثال از روند حل می بینیم.

اطلاعات اضافی:

نمودار ویژگی‌های استخراج شده از subject 16 را مشاهده می‌کنید که بعد یکی از آنها صرع اتفاق افتاده است و برای یکی از آنها اتفاق نیفتاده است.



دقت شود که در این موارد هنوز فرآیند انتخاب ویژگی‌ها را انجام نداده ایم.

ویژگی‌های انتخاب شده را باید درست کنیم.

حالا در ادامه همین نمودار ها برای سابجکت آخر و نحوه ی training یک به دو می بینیم.

