



دانشگاه صنعتی شریف

دانشکده مهندسی برق

مسعود ناطقی ۹۶۱۰۲۵۶۷

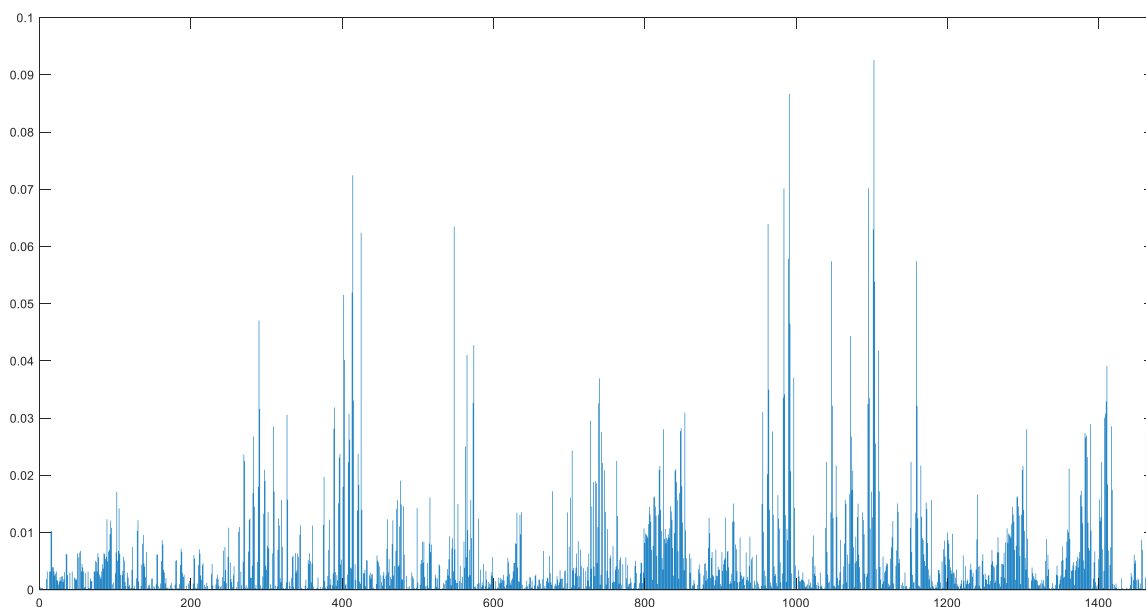
گزارش پروژه هوش محاسباتی

دکتر حاجی پور

## فاز ۱:

الف) برای استخراج ویژگی‌ها از <sup>1</sup>EEG Feature Extraction Toolbox استفاده کردیم. همچنین علاوه بر ویژگی‌های پیشنهادی در متن پروژه، از ویژگی‌هایی که معمولاً در task های تصویری از آن‌ها استفاده می‌شود مثل CSP و <sup>2</sup>obw نیز استفاده کردیم. البته باید دقت داشت بعد از استخراج ویژگی‌ها باید ماتریس ویژگی‌ها را تمیز کرد. چرا که بعضی سطرهای آن تمام صفر است، و در بعضی دیگر نیز مقادیر بی‌نهایت و NaN دیده می‌شود که می‌بایست آن‌ها را حذف کنیم.

ب) برای انتخاب ویژگی‌های بهتر از معیار فیشر تک‌بعدی استفاده کردیم و نمودار میله‌ای این معیار را برای همه ویژگی‌های به دست آمده رسم کردیم.



و در نهایت با قرار دادن یک آستانه، ویژگی‌هایی که مقدار **fisher score** بالاتر از آستانه دارند را انتخاب می‌کنیم. زیرا هر چه **fisher score** بیشتر باشد، احتمال جدایی پذیری داده‌ها بالاتر می‌رود.

ج) با ۴۸ ویژگی، شبکه را آموزش می‌دهیم. ابتدا برای تعیین ساختار شبکه و با استفاده از 5-Fold CV، دقت شبکه را تعیین می‌کنیم. این کار به ما در تعیین تعداد نورون‌های شبکه کمک زیادی می‌کند. روی ۵ دقت به دست آمده از هر یک از fold ها میانگین‌گیری می‌کنیم و به عنوان دقت شبکه بیان می‌کنیم. بهترین شبکه به ازای تعداد نورون ۶۱ عدد و با دقت ۷۳ درصد به دست آمد. اکنون ما ساختار شبکه را می‌دانیم و می‌توانیم برای آن که دقت آن را بالاتر ببریم، آن را با کل داده‌ها آموزش دهیم. این کار را چندین بار انجام می‌دهیم تا بالاترین دقت را به دست آوریم و نهایتاً به دقت ۸۰ درصد رسیدیم. خود شبکه در فایل MLP.mat ذخیره شده است.

<sup>1</sup> <https://www.mathworks.com/matlabcentral/fileexchange/84235-eeb-feature-extraction-toolbox>

<sup>2</sup> Occupied bandwidth

د) بهترین شبکه به ازای تعداد نورون ۲۵ عدد و شعاع ۹ و با دقت ۷۱ درصد به دست آمد. اکنون ما ساختار شبکه را می‌دانیم و می‌توانیم برای آن که دقت آن را بالاتر ببریم، آن را با کل داده‌ها آموزش دهیم. این کار را چندین بار انجام می‌دهیم تا بالاترین دقت را به دست آوریم و نهایتاً به دقت ۸۰.۷ درصد رسیدیم. خود شبکه در فایل RBF.mat ذخیره شده است.

ه) بهترین ویژگی‌های استفاده شده در این مساله عبارت بودند از:

۱. ویژگی همبستگی

۲. ویژگی CSP

۳. ویژگی log root sum of sequential variation

۴. mean teager energy

۵. mean curve length

۶. normalized second difference

۷. second difference

۸. first difference

از بین تمام ویژگی‌های بالا انتظار داشتیم که ویژگی CSP حتماً بین ویژگی‌های منتخب حضور داشته باشد.

البته از همه کانال‌ها در ویژگی‌های فوق‌الذکر استفاده نشد و تعداد بسیار کمی از کانال‌ها استفاده شدند.

و) خروجی لیبل‌های تست از شبکه MPL را در فایل testLabel\_MLP\_phase1.mat و خروجی لیبل‌های تست از شبکه RBF را در فایل testLabel\_RBF\_phase1.mat ذخیره کرده‌ایم.

## فاز ۲:

برای پیدا کردن ویژگی‌های مناسب از الگوریتم تکاملی استفاده کردیم. رمزگذاری آن را به صورت یک کروموزوم با ۳۰ ژن در نظر گرفتیم و هر ژن می‌توانست آلی از ۱ تا ۱۴۷۴ داشته باشد که ۱۴۷۴ تعداد کل ویژگی‌هاست. البته در ادامه و در کد پاسخ را از تابع **unique** رد کرده‌ایم تا ویژگی‌های تکراری را چندبار برنذاریم. برای تشکیل تابع هدف، ابتدا ۳۰ ویژگی که متناظر به کروموزوم هستند را انتخاب می‌کنیم بصورت 5-Fold CV و با طبقه‌بند SVM آموزش می‌دهیم. با توجه به اینکه در MATLAB می‌بایست تابع هزینه تعریف کنیم، تابع هزینه را بصورت منفی درصد دقت در نظر گرفتیم.

**Problem Setup and Results**

Solver: ga - Genetic Algorithm

Problem

Fitness function: @fitness\_SVM

Number of variables: 30

Constraints:

Linear inequalities: A:  b:

Linear equalities: Aeq:  beq:

Bounds: Lower: ones(1, 30) Upper: 1474\*ones(1, 30)

Nonlinear constraint function:

Integer variable indices: 1:30

Run solver and view results

☐ Use random states from previous run

Start Pause Stop

Current iteration: 72 Clear Results

Optimization running.  
Stop requested.  
Objective function value: -74.03273809523809  
Optimization terminated: Stop requested;

Final point:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1,408	17	437	454	1,347	1,320	852	933	810	764	1,058	255	966	1,097	1,301	764	962	137	963	846	506	765	249	967	1,189	917	141	706	1,282	380

**Options**

Population

Population type: Double vector

Population size: ☒ Use default: max(min(100, numel(x)), 10) ☐ Specify:

Creation function: Uniform

Initial population: ☒ Use default: [] ☐ Specify:

Initial scores: ☒ Use default: [] ☐ Specify:

Initial range: ☒ Use default: [] ☐ Specify:

Fitness scaling

Scaling function: Rank

Selection

Selection function: Stochastic uniform

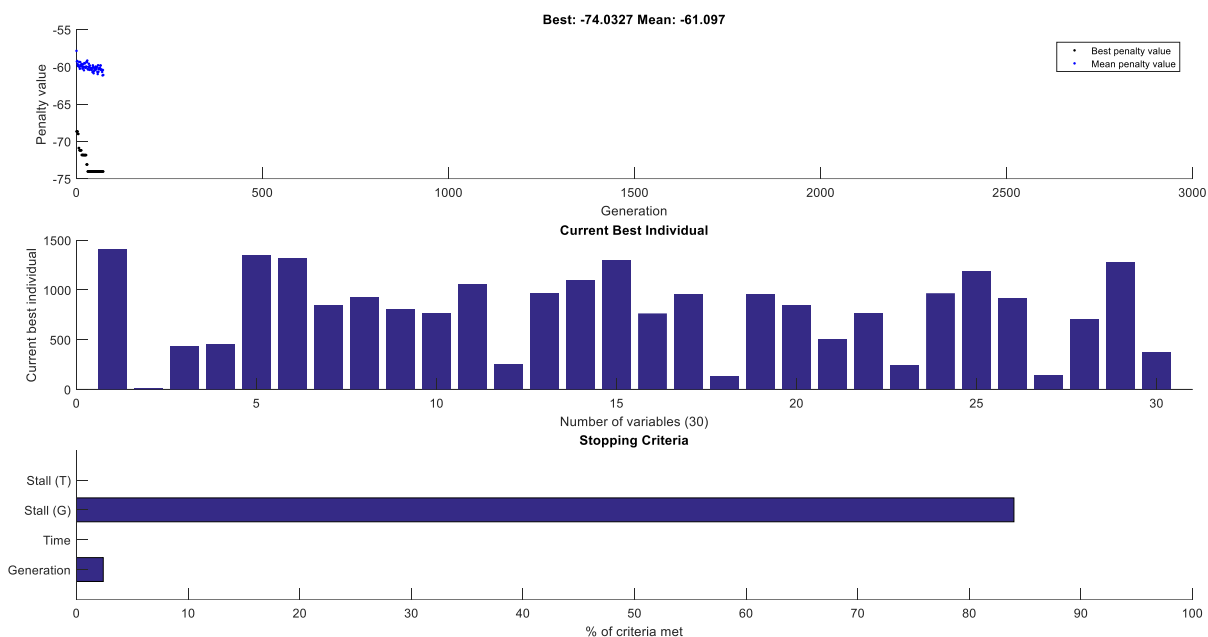
Reproduction

Elite count: ☒ Use default: 0.05\*max(nvars, 10) ☐ Specify:

Crossover fraction: ☒ Use default: 0.8 ☐ Specify:

Mutation

Mutation function: Constraint dependent



با توجه به اینکه ساعت ۷ صبح بود و چند شبی خواب کافی نداشتم، ادامه الگوریتم را لغو کردم. اما الگوریتم را برای ۱۵ و ۲۰ ویژگی نیز تکرار کردم و پاسخ خوبی به دست نیامد.

حال سعی می‌کنیم با این 29 ویژگی (دو تا از ۳۰ ویژگی یکسانند) ، شبکه ها را آموزش دهیم.

بهترین شبکه MLP با تعداد نوروں ۴۲ عدد و دقت ۶۶ درصد به دست آمد که با آموزش روی کل داده‌ها این دقت به ۷۸ درصد رسید. بهترین شبکه RBF با تعداد نوروں ۲۰ عدد و شعاع ۲۰ و دقت ۷۰ درصد به دست آمد که با آموزش روی کل داده‌ها این دقت به ۷۹ درصد رسید.

شبکه MLP در فایل MLP\_GA.mat و شبکه RBF در فایل RBF\_GA.mat ذخیره شده است. هم چنین لیبل‌ها نیز به ترتیب در فایل های testLabel\_MLP\_phase2.mat و testLabel\_RBF\_phase2.mat ذخیره شده است.

نتیجه کلی: الگوریتم ژنتیک مدل‌ها را ساده‌تر کرد اما درصد طبقه‌بندی‌ها کم شد. ممکن است به ازای درصد طبقه‌بندی کمتر تعمیم پذیری شبکه افزایش یافته باشد. بنده بسیار مشتاق مقایسه نتایج خود با لیبل‌های خواهم بود.