هوالحبیب

دانشکده مهندسی برق دانشگاه صنعتی شریف

هوش مصنوعی و محاسبات زیستی

ترم دوم سال تحصیلی 00-01

گزارش پروژه

نام و نام‌خانوادگی: ارشاک رضوانی

شماره دانشجویی: 98106531

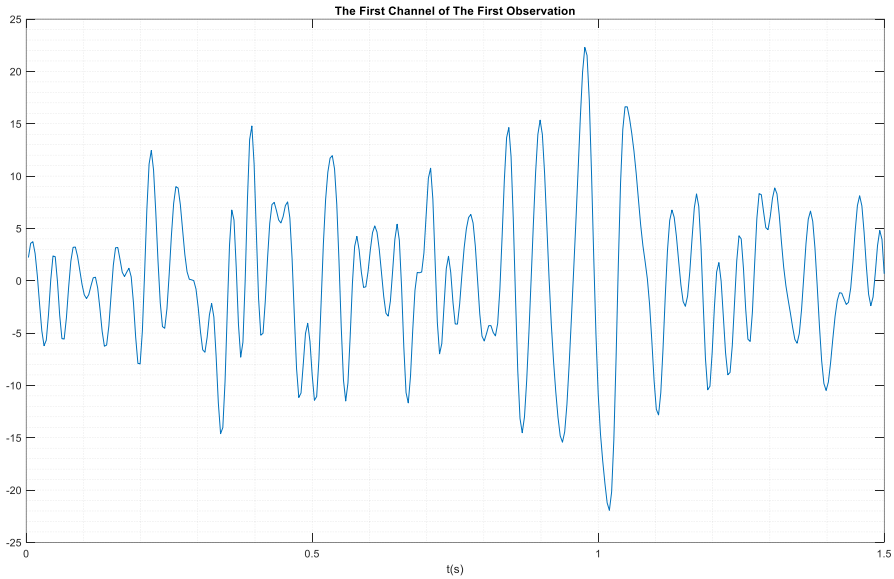
بهار 1401

این پروژه از سه بخش اصلی تشکیل شده است:

1. استخراج ویژگی(Feature Extraction)
   a. ویژگی‌های آماری
   b. ویژگی‌های فرکانسی
2. انتخاب ویژگی(Feature Selection)
   a. معیار فیشر(1D Fisher Score)
   b. استفاده از Chi-Square Test
   c. استفاده از الگوریتم MRMR (Minimum Redundancy Maximum Relevance)
   d. استفاده از PCA و الگوریتم ژنتیک
3. اموزش شبکه
   a. MLP
   b. RBF

استخراج ویژگی:

ابتدا برای درک بهتر سیگنال، کانال اول اولین نمونه را رسم کردم:



1245 ویژگی آماری و 300 ویژگی فرکانسی استخراج کردم.

ویژگی‌های آماری:

1- واریانس 30 کانال
2- کورلیشن میان 30 کانال که انتخاب 2 از 30 کانال است (435 ویژگی)
3- کورتسیس 30 کانال (Kurtosis)
4- اسکیونس 30 کانال (Skewness)
5- ضرایب مدل AR مرتبه 10 هر کانال(300 ویژگی)
6- چگالی دامنه در بازه‌های مختلف(هیستوگرام 30 کانال در 14 بازه، در کل 420 ویژگی)

ویژگی‌های فرکانسی:

1- فرکانس ماکسیمم 30 کانال
2- فرکانس میانگین 30 کانال
3- فرکانس میانه 30 کانال
4- انرژی نسبی باندهای باند توان(7 باند برای 30 کانال، در کل 210 ویژگی)

انتخاب ویژگی:

دو ماتریس یکی برای ویژگی‌های آماری(ابعاد 120x1245) و یکی برای ویژگی‌ها اماری(ابعاد 120x300) تشکیل داده شد، دو ماتریس را در کنار هم قرار دادم و ماتریس کل ویژگی‌ها را تشکیل دادم(ابعاد 120x1545، 1545 ویژگی از 120 مشاهده که هر ستون مربوط به یک ویژگی و هر سطر مربوط به یک مشاهده است) سپس ستون‌های ماتریس را نرمالایز کردم یعنی میانگین هر ستون را از آن ستون کم کردم و تقسیم بر واریانس کردم(میانگین داده‌ها را 0 و واریانس آن‌ها را 1 کردم)

به چند روش شروع به انتخاب ویژگی‌ها کردم(در این بخش هر ویژگی ستون‌های ماتریس است)، در کل به علت کم بودن داده‌های ورودی(120 داده)، 10 ویژگی بیشتر برای آموزش شبکه استفاده نشد.

1- استفاده از معیار فیشر(Fisher Score)

معیار فیشر را برای هر کدام از داده‌ها به صورت تکی‌حساب کردم و سپس به ترتیب در یک بردار چیدم، 10 ویژگی برتر انتخاب شده توسط این معیار:

1- هیستوگرام کانال 9 ام در بازه 12 تا 20
2- انرژی نسبی gamma گاما در کانال 17 ام
3- هیستوگرام کانال 7 ام در بازه 12 تا 20
4- انرژی نسبی باند high-beta در کانال 6 ام
5- ضریب 9 ام مدل AR مرتبه 10 کانال 6 ام
6- ضریب 4 ام مدل AR مرتبه 10 کانال 20 ام
7- ضریب 4 ام مدل AR مرتبه 10 کانال 19 ام
8- کورلیشن کانال 10 ام و 22 ام
9- فرکانس میانگین کانال 8 ام
10- ضریب 9 ام مدل AR مرتبه 10 کانال 16 ام

2- استفاده از Chi-Square Test

فیچرها را با استفاده از این معیار مرتب کردم و 10 ویژگی برتر انتخاب شده به ترتیب زیر هستند.

1- کورلیشن کانال 13 ام و 14 ام
2- کورلیشن کانال 12 ام و 25 ام
3- ضریب 3 ام مدل AR مرتبه 10 کانال 8 ام
4- کورلیشن کانال 15 ام و 19 ام
5- ضریب 9 ام مدل AR مرتبه 10 کانال 20 ام
6- کورلیشن کانال 10 ام و 18 ام
7- ضریب 4 ام مدل AR مرتبه 10 کانال 7 ام
8- کورلیشن کانال 9 ام و 12 ام

9- ضریب 5 ام مدل AR مرتبه 10 کانال 8 ام
10- ضریب 1 ام مدل AR مرتبه 10 کانال 12 ام

3- استفاده از الگوریتم MRMR (Minimum Redundancy Maximum Relevance) (حداقل افزونگی-حداکثر ارتباط)

این معیار سعی می‌کند ویژگی‌هایی که کمترین همپوشانی نسبت به هم و حداکثر ارتباط را با نتیجه خروجی دارند را پیدا کند.

10 ویژگی برتر انتخاب شده با این روش:

1- کورلیشن کانال 10 ام و 20 ام
2- هیستوگرام کانال 19 ام در بازه 0 تا 0.5
3- فرکانس ماکسیمم کانال 6 ام
4- هیستوگرام کانال 20 ام در بازه 4- تا 2-
5- هیستوگرام کانال 15 ام در بازه 12- تا 8-
6- هیستوگرام کانال 4 ام در بازه 2- تا 1-
7- هیستوگرام کانال 9 ام در بازه 4- تا 2-
8- هیستوگرام کانال 13 ام در بازه 12- تا 8-
9- هیستوگرام کانال 17 ام در بازه 2 تا 4
10- کورلیشن کانال 13 ام و 15 ام

4- استفاده از PCA و الگوریتم ژنتیک

اجرا کردن یک جستجو بر روی 1545 ویژگی برای انتخاب 10 تا فضای جستجوی بسیار بزرگی دارد و حتی نمی‌شود به جواب بهنیه نزدیک شد پس با استفاده از PCA(Principal Component Analysis) فضای جستجو را به 120 ویژگی(با توجه به ابعاد ماتریس ویژگی که 120x1545 است) محدود کردم و سپس 20 ویژگی‌ای که بیشترین میزان واریانس داده‌ها را در خود گنجانده بودند را انتخاب کردم(یک ماتریس جدید تشکیل دادم به ابعاد 120x20 که هر ستون یک ویژگی را نشان می‌دهد) و یک الگوریتم ژنتیک برای انتخاب 10 ویژگی از این 20 ویژگی اجرا کردم که خود باز فضای جستوجوی بسیار بزرگی را دارد(ولی قابل اجرا است).

کدگذاری به صورت یک بردار 20 تایی از صفر و یک است که دقیقا 10 تا از مولفه‌های آن یک است که یک بودن یک مولفه به معنای آن است که آن ویژگی جزو 10 ویژگی انتخاب شده است، برای تابع سازگاری از معیار فیشر چندبعدی استفاده کردم که مطابق فرمول با توجه به Genotype داده شده از روی ماتریس 20 ویژگی برتر PCA، 10 ویژگی را جدا کرده(Phenotype) و با استفاده فرمول فیشر چند بعدی یک میزان Fitness به هر پاسخ نامزد نسبت می‌دهم، عملگرهای ژنتیکی را توابعی انتخاب کردم که قیدهای مسئله را نقض نکنند(پیشنهادی خود متلب برای بهینه‌سازی مقید با استفاده از الگوریتم ژنتیک).

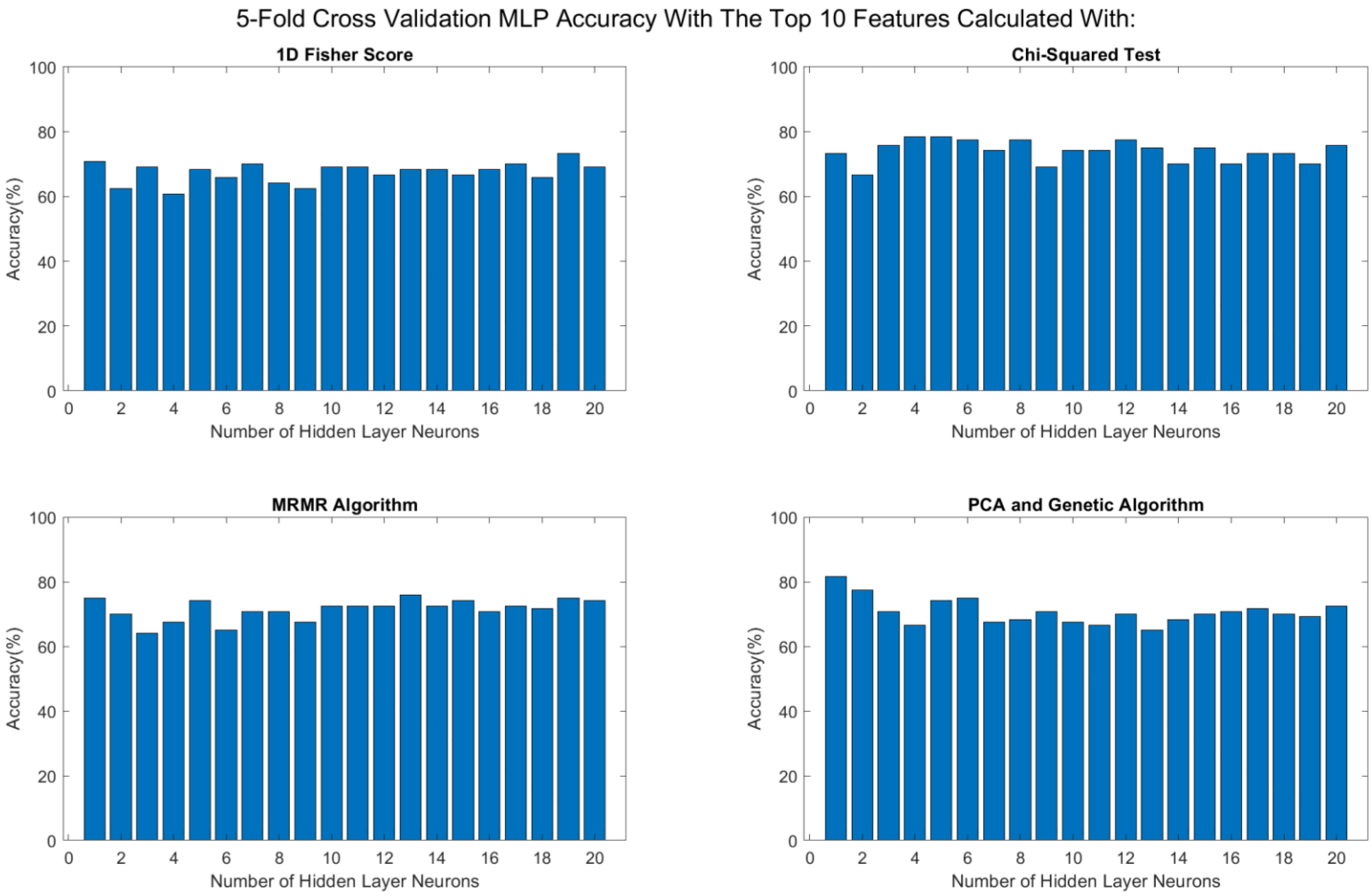
آموزش شبکه:

با استفاده از 5-Fold Cross Validation به 120 داده به صورت رندوم اندیس‌های یک تا پنج دادم سپس شروع به آموزش دو نوع شبکه MLP و RBF کردم. سپس خروجی‌های صحیح(Training Target) را One-Hot کردم، یعنی یک ماتریس 120x2 که هر ردیف یک داده را نشان می‌دهد که بسته به کلاس آن داده ستون مربوط به آن یک است که یعنی شبکه‌های ترین شده دارای دو نورون در لایه خروجی هستند که هر کدام بیشتر باشد یعنی شبکه آن داده را به آن کلاس نسبت داده‌است.

- MLP:

یک حلقه تودرتو بر روی چهار ست 10 تایی از ویژگی‌ها و بر روی تعداد نورون‌های لایه‌های میانی(از 1 تا 20 نورون) اجرا کردم تا با استفاده از 5-Fold Cross Validation بهترین خروجی(صحت) را پیدا کنم، توجه شود که به علت کم بودن تعداد داده‌های آموزشی و برای جلوگیری از اورفیت شدن شبکه فقط یک لایه پنهان برای شبکه در نظر گرفتم.

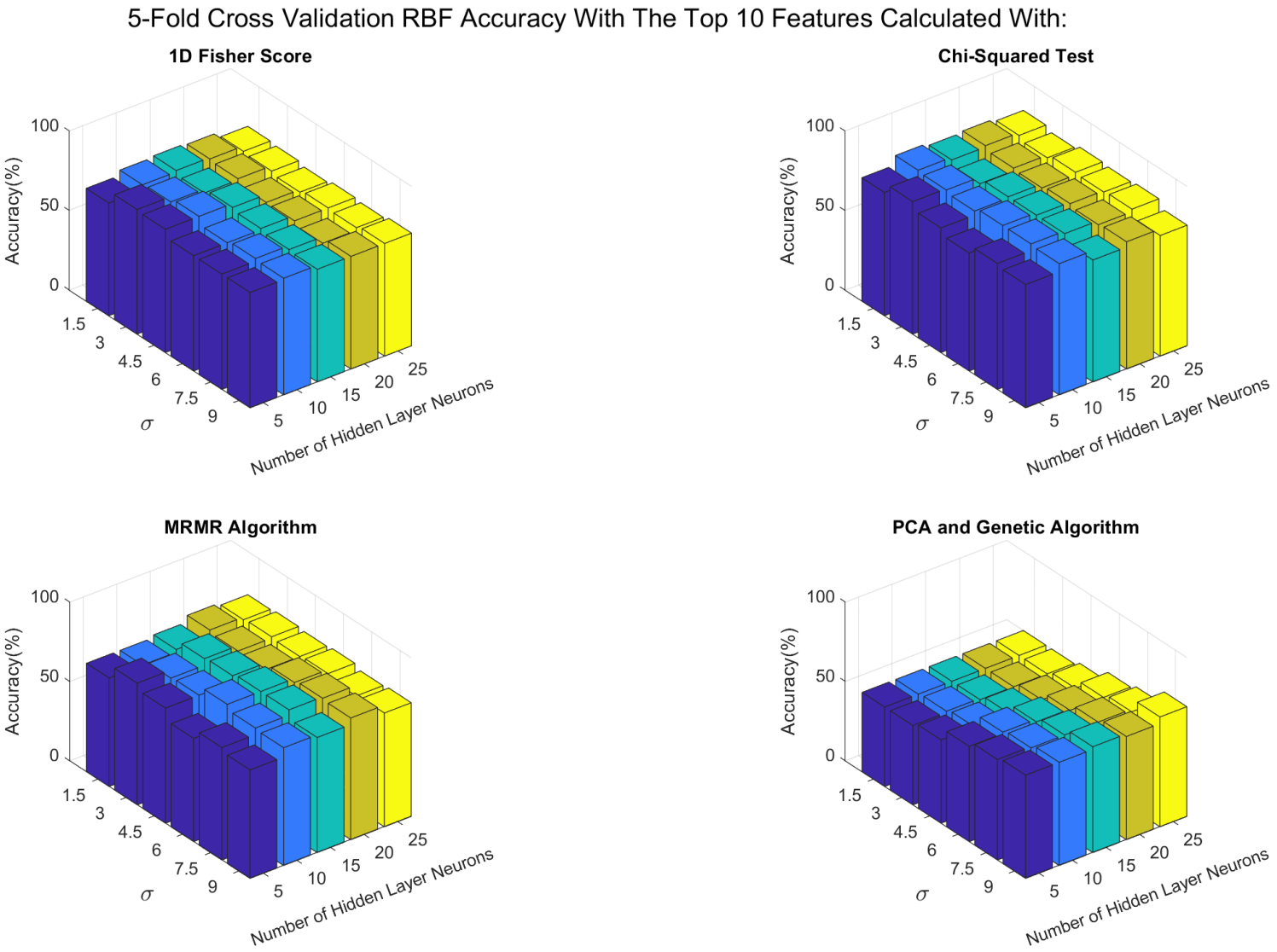
نتایج شبیه‌سازی:



مشاهده می‌شود که بیشترین دقت MLP بدون الگوریتم ژنتیک (78.33٪) با 4 یا 5 نورون لایه پنهان و با استفاده از ویژگی‌های جدا شده توسط الگوریتم Chi-Squared Test حاصل می‌شود. بیشترین دقت با استفاده از PCA و الگوریتم ژنتیک 81.67٪ است با 1 نورون لایه پنهان.

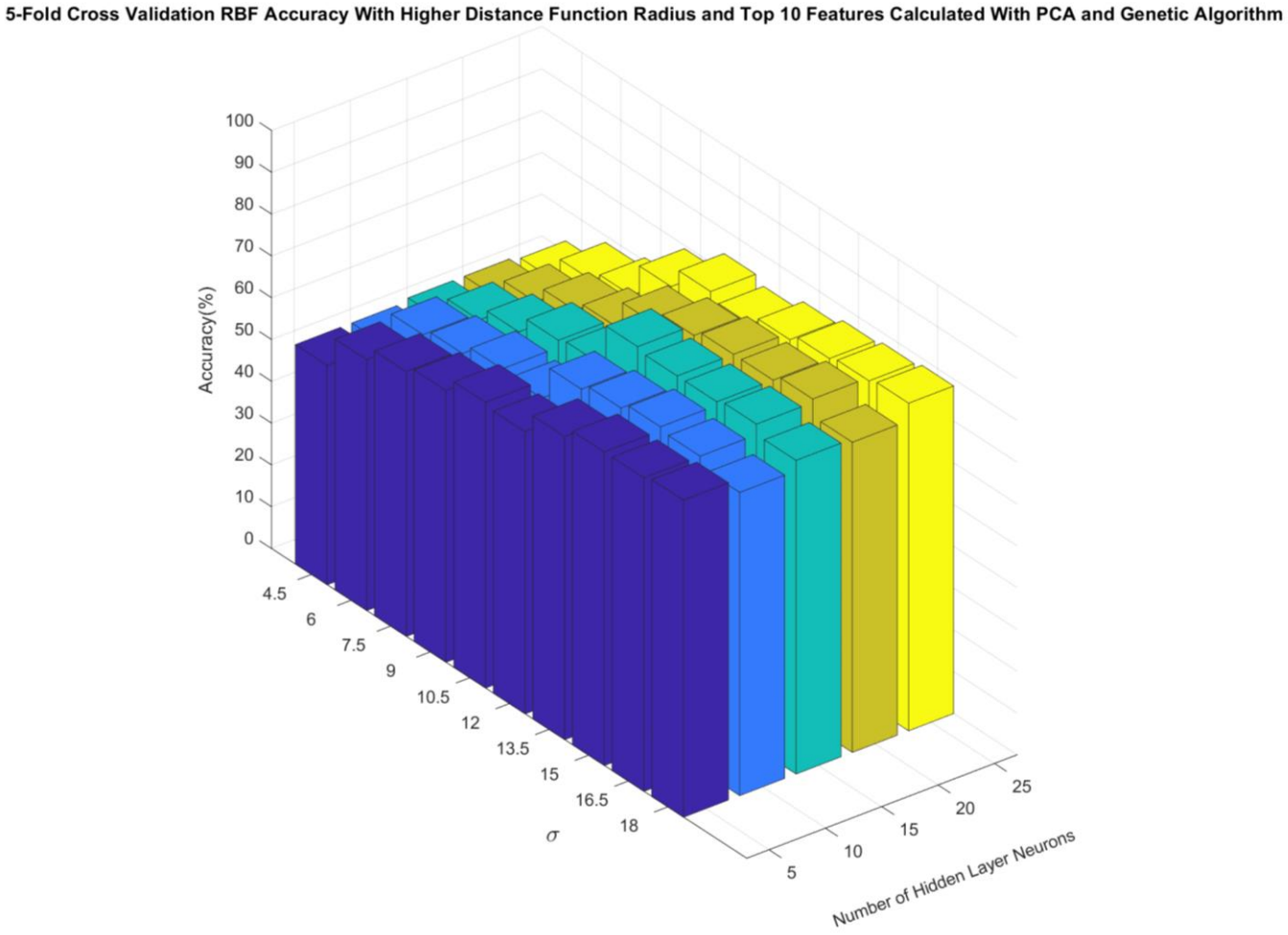
- RBF:

یک حلقه تودرتو بر روی چهار ست 10 تایی از ویژگی‌ها و بر روی شعاع توابع فاصله شبکه و بر روی تعداد نورون‌های لایه آخر شبکه اجرا کردم تا با استفاده از 5-Fold Cross Validation بهترین خروجی(صحت) را پیدا کنم:



مشاهده می‌شود که بیشترین دقت به دست آمده(83.33٪، از MLP بیشتر) با استفاده از ویژگی‌های جدا شده توسط Chi-Squared Test و شعاع تابع فاصله(σ) مساوی 3 و تعداد نورون‌های لایه پنهان 5 به دست آمده است. چون مشاهده می‌شود که با بیشتر شدن شعاع تابع فاصله(σ) میزان دقت شبکه با استفاده از ویژگی‌های جدا شده توسط الگوریتم ژنتیک بیشتر می‌شود پس یک بار فقط بر روی این ویژگی‌ها با با مقدار شعاع‌های بزرگ‌تر نیز شبکه را آموزش دادم تا دقت بیشتر بگیرم، خروجی در صفحه بعد موجود است.

دقت RBF با شعاع‌های بزرگتر و ویژگی‌های جدا شده با PCA و الگوریتم ژنتیک:



مشاهده می‌شود در این حالت بیشترین دقت به دست آمده 78.33% است که با 20 نورون لایه پنهان و شعاع تابع فاصله(σ) مساوی 16.5 حاصل می‌شود.