

## مدلی مبتنی بر الگوریتم ژنتیک برای انتخاب/استخراج ویژگی و بهینه‌سازی پارامترهای SVM

احسان عادل‌لی مسیب و محمود فتحی

دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر، آزمایشگاه پردازش سریع

E-mail: eadeli@comp.iust.ac.ir, mahfathy@iust.ac.ir

**چکیده** - ماشین‌های بردار حامی (SVM) نوعی الگوریتم یادگیری آماری هستند که در سالهای اخیر مورد توجه فراوانی قرار گرفته‌اند. SVM به طور کلی کاربردهای زیادی در بحث شناسایی الگو و برازش غیرخطی دارد. توابع هسته‌ای نقش مهمی در توانایی کلاس‌بندی آنها بازی می‌کنند. تنظیم مناسب پارامترهای هسته می‌تواند در دقت کلاس‌بندی آن نقش خیلی مهمی داشته باشد. در عین حال انتخاب مناسب ویژگی‌های ورودی نیز یکی از مسائل مهم در یک فرآیند کلاس‌بندی می‌باشد. در این مقاله مدلی مبتنی بر الگوریتم ژنتیک پیشنهاد شده است که از مجموعه بردار ویژگی ورودی زیر مجموعه‌ای را انتخاب یا استخراج (با استفاده از PCA) می‌کند و بطور همزمان پارامترهای کلاس‌بند SVM را بهینه می‌سازد. نتایج شبیه‌سازی روی برخی مجموعه داده‌های معروف کارایی این روش را نشان می‌دهد.

**کلید واژه** - الگوریتم ژنتیک، انتخاب ویژگی، شناسایی الگو، کلاس‌بندی، ماشین بردار حامی.

### 1- مقدمه

بهترین پارامترهای تابع هسته‌ای را تنظیم کنیم. این دو مسئله با هم کاملاً در ارتباطند. چراکه مجموعه‌ی ورودی روی انتخاب پارامترهای هسته تاثیرگذار است [8]. در مورد انتخاب مجموعه داده‌ی ورودی باید به این نکته اشاره کرد که در بسیاری موارد متعادل نبودن نسبت ویژگی‌های در دست به کل اطلاعات مورد نیاز در مورد هر کلاس (اثر Hughes [9])، باعث می‌شود کلاس‌بندهای معمولی در این مورد با شکست مواجه شوند. همچنین در بسیاری از مجموعه داده‌ها برخی از ویژگی‌ها در تصمیم‌گیری نقشی ندارند و به نوعی می‌توان آنها را اضافی تلقی نمود. پس انتخاب یک زیر مجموعه‌ی مناسب از ورودی‌ها می‌تواند هم در دقت کلاس‌بندی و هم در سرعت آن مثر ثمر باشد.

از روشهایی که در بحث کاهش ویژگی‌ها استفاده می‌شود، می‌توان به PCA اشاره کرد. این روش مقادیر ویژه ماتریس کوواریانس داده‌های ورودی را مورد تحلیل قرار می‌دهد [10]. محبوبیت PCA بدلیل سادگی استفاده و نتایج مناسب در برخی کاربردها باعث شده است که بسیاری فراموش کنند که PCA در بعضی کاربردها گزینه‌ی خوبی

ماشین بردار حامی (Support Vector Machine - SVM) در سال 1995 توسط Vapnik [1] پیشنهاد شد. اخیراً این مدل برای حل مسائل زیادی به کار رفته است. از این جمله می‌توان به کاربرد آن در بحث شناسایی الگو [2, 3]، برازش غیرخطی، دسته‌بندی متون [4] و کلاس‌بندی تصاویر ابر طیفی [5] اشاره کرد. یک SVM برای تعیین ابرصفحه‌ی جداکننده داده‌های آموزش از مکانیزمی استفاده می‌کند که در آن به کمک یک تابع هسته‌ای داده‌ها را در فضا به بعدی انتقال می‌دهد که بتواند این ابرصفحه را بر داده‌های ورودی برازش دهد [6, 7]. بنابراین انتخاب درست یک تابع هسته‌ای می‌تواند تاثیر مستقیم روی کارایی SVM داشته باشد.

در زمان استفاده‌ی SVM دو مسئله‌ی مهم وجود دارد که باید مد نظر قرار گرفته شوند. اول اینکه چگونه مجموعه داده‌ی ورودی را انتخاب کنیم، یا اینکه آیا انتقالی روی داده‌های ورودی قبل از ورود به SVM می‌توان ایجاد نمود که نتایج مناسب‌تری ایجاد کند یا خیر، و دوم اینکه چگونه

برای استخراج ویژگی نمی‌باشد. [11] نشان داده است که PCA در استخراج ویژگی‌های متمایز کننده در برخی توزیع‌های داده نمی‌تواند کارا باشد. همچنین نشان داده شده که حتی تمایز بهتری از همان بردار ویژگی اولیه ایجاد نمی‌کند. بنابراین، بسته به نوع توزیع داده‌های ورودی استفاده از PCA می‌تواند مناسب یا نامناسب باشد. این مسئله که آیا از انتقال PCA استفاده شود یا از بردار ویژگی اصلی، در کنار انتخاب مناسب پارامترهای SVM نوعی مسئله‌ی بهینه‌سازی است و الگوریتم ژنتیک این توانایی را دارد که به طور همزمان تصمیمی بهینه در مورد پارامترهای کلاس‌بند و بردار ویژگی‌ورودی آن اتخاذ کند. در این مقاله برآنیم تا با استفاده از الگوریتم ژنتیک مکانیزمی برای حل این مسئله طراحی کنیم.

در این زمینه کارهای فراوانی انجام نشده است. برای انتخاب ویژگی برای SVM توسط [12, 13, 14] روشهایی پیشنهاد شده است. برخی دیگر نیز مانند [15, 16] روشهایی مبتنی بر الگوریتم ژنتیک برای انتخاب ویژگی پیشنهاد کرده‌اند. اما نکته‌ی مهم آنست که هیچ‌یک از این کارها سعی نکرده‌اند پارامترهای SVM را نیز بهینه کنند. [15] با استفاده از الگوریتم ژنتیک نوع تابع هسته‌ای و پارامترهای آنرا بهینه کرده است ولی روی بردار ویژگی‌های ورودی تاثیری نگذاشته است. در عین حال Wang و Huang در [17] جزء معدود کسانی بودند که پارامترها و زیرمجموعه بردار ویژگی را در کنار هم بهینه کرده است. اما آنها هم هیچ‌گونه انتقالی در بردار ویژگی را در نظر نگرفته‌اند.

این مقاله به ترتیب زیر سازمان‌دهی خواهد شد: در بخش دوم مروری بر اصول SVM خواهیم داشت. بخش سوم الگوریتم ژنتیک را مختصراً شرح خواهد داد. پس از آن الگوریتم پیشنهادی مورد بررسی قرار خواهد گرفت. سپس بخش پنجم نتایج آزمایشات الگوریتم پیشنهادی را در بر خواهد داشت و در نهایت در بخشهای ششم و هفتم، نتیجه‌گیری و کارهای آینده را ارائه خواهیم کرد.

## 2- ماشین بردار حامی (SVM)

SVM ماشینی خطی با خصوصیتی جالب است، ابتدا الگوهای تفکیک پذیر را در نظر می‌گیریم. هدف این است که ابرصفحه‌ای به عنوان سطح تصمیم‌گیری بسازیم که

الگوها را جدا کند و حاشیه‌ی جداسازی بین نمونه‌های مثبت و منفی بیشینه شود. این اصل بر پایه‌ی دو مفهوم بنا شده است. اول میزان خطای آموزش یک ماشین آموزش، و دیگر شاخص VC (Vapnik-Chervonenkis). در مورد الگوهای تفکیک پذیر یک SVM مقدار صفر برای عبارت اول تولید می‌کند و عبارت دوم را کمینه می‌کند [6, 7].

## 2-1- الگوهای تفکیک پذیر خطی

نمونه‌های آموزش  $\{(x_i, d_i)\}_{i=1}^N$  را در نظر می‌گیریم، که  $x_i$  الگوی ورودی نمونه‌ی  $i$ ام و  $d_i \in \{-1, +1\}$  خروجی مطلوب است. فرض می‌کنیم الگوهای  $+1$  و الگوهای  $-1$  بصورت خطی تفکیک پذیر می‌باشند. معادله‌ی سطح تصمیم‌گیری (ابرفصفحه‌ی جداکننده) به شکل زیر است.

$$w^T x + b = 0 \quad (1)$$

که در آن  $x$  بردار ورودی،  $w$  بردار وزن و  $b$  بایاس آن است. برای بردار وزن  $w$  و بایاس  $b$ ، جداسازی توسط ابرصفحه‌ی فرمول 1 انجام می‌شود. فاصله‌ی نزدیکترین نقاط داده حاشیه‌ی جداسازی نامیده می‌شود ( $M$ )، هدف SVM آنست که ابرصفحه‌ای را به عنوان سطح تصمیم‌گیری پیدا کند که این حاشیه‌ی جداسازی را بیشینه می‌کند. اگر  $w_0$  و  $b_0$  مقادیر بهینه‌ی بردار وزن و بایاس را نشان دهند، ابرصفحه‌ی بهینه‌ی چند بعدی به شکل زیر نشان داده می‌شود.

$$w_0^T x + b_0 = 0 \quad (2)$$

جفت پارامتر  $w_0$  و  $b_0$  برای این هدف باید شرایط زیر را ارضا کنند.

$$\begin{aligned} w_0^T x_i + b_0 &\geq +1 \text{ for } d_i = +1 \\ w_0^T x_i + b_0 &\leq -1 \text{ for } d_i = -1 \end{aligned} \quad (3)$$

بنابراین صفحه‌ی مثبت را به شکل  $w_0^T x_i + b_0 = +1$  و صفحه‌ی منفی را به شکل  $w_0^T x_i + b_0 = -1$  نمایش می‌دهیم. در شکل 1 نقاط مربوط به بردار حامی مشخص شده‌اند. مقدار حاشیه‌ی جداسازی  $M$  را از رابطه‌ی 4 محاسبه می‌کنیم. این نکته واضح است که بیشینه‌سازی حاشیه‌ی جداسازی بین دو کلاس معادل است با کمینه‌سازی نرم اقلیدسی بردار وزن  $w$ .

### 3-2- ابرصفحه‌ی بهینه برای الگوهای غیر قابل تفکیک

اگر شرط 5 ارضا نشود خطا ممکن است در یکی از دو حالت زیر اتفاق بیافتد: 1- نقطه‌ی داده‌ی  $(x_i, d_i)$  داخل محدوده جداسازی بیافتد اما در طرف درست سطح تصمیم گیری. 2- نقطه‌ی داده‌ی  $(x_i, d_i)$  در طرف نادرست سطح تصمیم گیری قرار دارد.

در حالت اول یک کلاس‌بندی درست خواهیم داشت، در حالی که در حالت دوم کلاس‌بندی اشتباه صورت پذیرفته است. برای برخورد با این مسئله، یک سری متغیرهای غیر منفی  $\{\xi_i\}_{i=1}^N$  تعریف می‌کنیم و تعریف سطح تصمیم گیری را به شکل زیر تغییر می‌دهیم:

$$d_i(w^T x_i + b) \geq 1 - \xi_i \text{ for } i = 1, 2, \dots, N \quad (9)$$

متغیرهای  $\xi_i$  معیاری برای انحراف یک نقطه‌ی داده از شرایط ایده‌آل جداسازی الگو می‌باشند. برای  $0 \leq \xi_i \leq 1$ ، نقطه‌ی داده داخل محدوده‌ی جداسازی اما در طرف درست سطح جداسازی قرار می‌گیرد و برای  $\xi_i > 1$ ، نقطه در طرف نادرست ابرصفحه‌ی جداساز قرار گرفته است.

تابع هزینه که باید آنرا نیز کمینه می‌کردیم با توجه به  $w$  و متغیرهای کمبود به شکل زیر در خواهد آمد.

$$\Phi(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \quad (10)$$

پارامتر  $C$  پارامتری است برای کنترل مصالحه بین پیچیدگی ماشین و تعداد نقاط غیر قابل تفکیک.

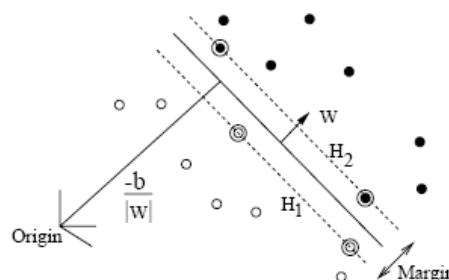
### 4-2- توابع هسته‌ای

اصولاً، یکی از نقاط قوت SVM، نگاهت غیرخطی بردار ورودی به یک فضای ویژگی با ابعاد بالا است، که از دید ورودی و خروجی پنهان می‌باشد. این کار توسط توابع هسته‌ای صورت می‌پذیرد. توابع هسته‌ای مهم عبارتند از:

ماشین یادگیرنده‌ی چندجمله‌ای:  $(x^T x_i + 1)^p$

RBF:  $\exp\left(-\frac{1}{2\gamma^2} \|x - x_i\|^2\right)$

$$M = \frac{2}{\sqrt{w^T w}} = \frac{2}{\|w\|} \quad (4)$$



شکل 1: ابرصفحه‌ی جداساز خطی.

### 2-2- یافتن ابرصفحه‌ی بهینه

«نمونه‌های آموزش  $\{(x_i, d_i)\}_{i=1}^N$  داده شده‌اند. مقادیر بهینه‌ی بردار وزن  $w$  و بایاس  $b$  را طوری پیدا کنیم که شرایط زیر را ارضا کنند،

$$d_i(w^T x_i + b) \geq 1 \text{ for } i = 1, 2, \dots, N \quad (5)$$

و بردار وزن تابع هزینه را کمینه کند:

$$\Phi(w) = \frac{1}{2} w^T w \quad (6)$$

باید فضای  $w$  و  $b$  را جستجو کرده و مناسب‌ترین  $w$  و  $b$  ها که بزرگترین حاشیه را می‌سازد انتخاب کنیم. برای این جستجو روشهای مختلفی پیشنهاد می‌شوند، اما یکی از بهترین و پر کاربردترین آنها بهینه سازی نمایی درجه‌ی دو است. در این روش به دنبال ماکزیمم عبارت 7 هستیم

$$\arg \max_u c + d^T u + \frac{u^T R u}{2} \quad (7)$$

ابتدا تابع لاگرانژ را می‌سازیم:

$$J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i [d_i (w^T x_i + b) - 1] \quad (8)$$

متغیرهای  $\alpha_i$  ضرائب لاگرانژ می‌باشند. حل این مسئله‌ی بهینه‌سازی معادل است با یافتن نقاط بحرانی این تابع است که باید نسبت به  $w$  و  $b$  کمینه و نسبت به ضرائب  $\alpha$  بیشینه شود.

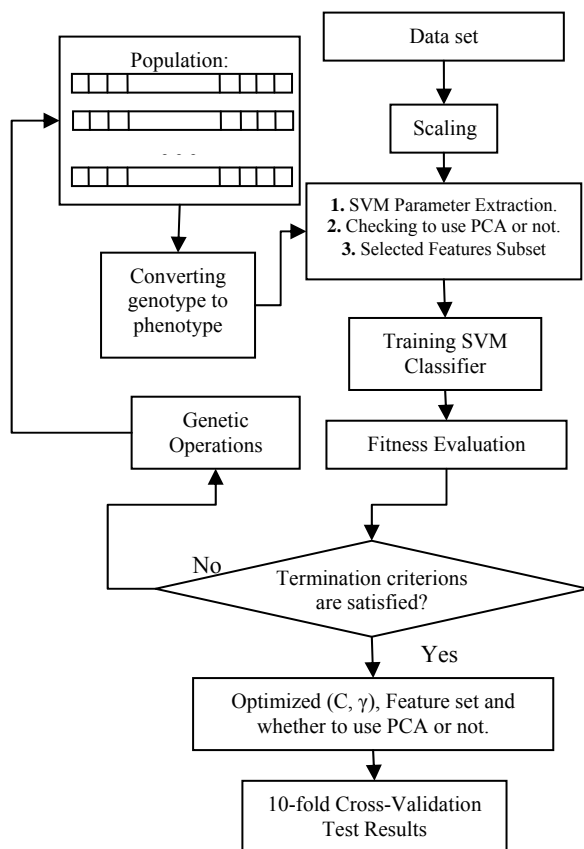
پرسپترون دو لایه:

$$\tanh(\beta_0 x^T x_i + \beta_1)$$

حد پایین آن است.

$$v' = \frac{v - \min_a}{\max_a - \min_a} \quad (11)$$

سپس الگوریتم ژنتیک وارد کار می‌شود. عملگرها و روش کار در ادامه بیان می‌شوند.



شکل 2: نمودار فرآیند کاری الگوریتم پیشنهادی.

#### 1-4- طراحی کوروموزم

شکل 3 نمونه‌ای کلی از کوروموزم را نشان می‌دهد. این کوروموزم رشته‌ای از بیت‌هاست و شامل 4 بخش است.

$g_c^i$  ( $0 < i \leq n_c$ )، بیت‌هایی برای پارامتر  $C$ ،  $g_\gamma^i$  ( $0 < i \leq n_\gamma$ ) بیت‌هایی برای پارامتر  $\gamma$  هسته‌ی RBF و  $g_f^i$  ( $0 < i \leq n_f$ ) بیت‌هایی برای انتخاب ویژگی‌ها،  $P$  نیز بیانگر اینکه آیا از PCA استفاده کنیم (مقدار 1) یا خیر (مقدار 0).  $n_f$  برابر تعداد ویژگی‌هاست،  $n_c$  و  $n_\gamma$  تعداد بیت‌هایی هستند که برای پارامترهای  $C$  و  $\gamma$  هسته‌ی RBF در نظر گرفته می‌شوند. مقادیر آنها بر اساس مقدار دقیقی که انتظار داریم تعیین می‌شوند.

#### 5-2- حالت چند کلاسه

برای حالت چند کلاسه چندین کلاس‌بند SVM دو کلاسه (بسته به تعداد کلاسهای مورد انتظار و بسته به نحوه‌ی ترکیب) آموزش داده می‌شوند. پس از آموزش برای رسیدن به نوع چند کلاسه نتایج آنها با هم ترکیب شده و خروجی مورد نظر بدست می‌آید. در این مقاله، از نوع ترکیب تصحیح خطای کدهای خروجی استفاده شده است.

#### 3- الگوریتم ژنتیک

الگوریتم ژنتیک توسط جان هالند در سال ۱۹۶۷ ابداع شده‌است. بعدها این روش با تلاشهای گلدبرگ ۱۹۸۹، مکان خویش را یافته و امروزه نیز بواسطه تواناییهای خویش، جای مناسبی در میان دیگر روشها دارد. روال بهینه یابی در الگوریتم ژنتیک براساس یک روند تصادفی- هدایت شده استوار می‌باشد. این روش، بر مبنای نظریه تکامل تدریجی و ایده‌های بنیادین داروین پایه گذاری شده‌است. در این روش، ابتدا برای تعدادی ثابت که جمعیت نامیده می‌شود مجموعه‌ای از پارامترهای هدف بصورت اتفاقی تولید می‌شود، پس از اجرای برنامه شبیه‌ساز عددی را که معرف انحراف معیار و یا برازش آن مجموعه از اطلاعات است را به آن عضو از جمعیت مذکور نسبت می‌دهیم (Fitness). این عمل را برای تک تک اعضای ایجاد شده تکرار می‌کنیم، سپس با فراخوانی عملگرهای الگوریتم ژنتیک از جمله وراثت (Crossover)، جهش (Mutation) و انتخاب نسل بعد (Selection) را شکل می‌دهیم و این روال تا ارضای معیار همگرایی ادامه داده خواهد شد.

#### 4- الگوریتم پیشنهادی

فرآیند کاری الگوریتم را می‌توان در نمودار موجود در شکل 2 دید. ابتدا داده‌ها را همگی به مقیاس واحدی می‌بریم. از فوائد مقیاس‌بندی، اجتناب از مشکلات ناشی از تفاوت بسیار زیاد مقادیر و ناشی از محاسبات و سر ریز شدن متغیرها را می‌توان بیان کرد. این کار باعث بالا بردن نتایج کلاس‌بندی نیز می‌شود. داده‌ها را در بازه‌ی  $[-1, +1]$  با فرمول 11 مقیاس‌بندی می‌کنیم. که در آن  $v$  مقدار اصلی،  $v'$  مقدار مقیاس داده شده،  $\max_a$  مقدار حد بالای آن ویژگی و  $\min_a$

$g_c^1$	$\dots g_c^i \dots$	$g_c^{mf}$	$g_v^1$	$\dots g_v^i \dots$	$g_v^{m'}$	P	$g_f^1$	$\dots g_f^i \dots$	$g_f^{mf}$
---------	---------------------	------------	---------	---------------------	------------	---	---------	---------------------	------------

شکل 3: نمونه‌ی کلی کوروموزم مورد استفاده.

می‌گیریم. ورودیها به این ترتیب برای SVM انتخاب می‌شوند و میزان دقت کلاس‌بندی محاسبه می‌شود. بنابراین می‌توان میزان Fitness را از رابطه‌ی زیر پیدا کرد.

$$Fitness = W_A \times SVM\_Accuracy + W_F \times \left(\sum_{i=1}^{n_f} g_f^i\right)^{-1} \quad (13)$$

#### 5-4- Selection

دیگر عملگرها انتخاب برای تولید نسل آینده است. در اینجا از Rolette Wheel استفاده می‌کنیم. به آنهایی که fitness بیشتری دارند احتمال بیشتری برای انتخاب می‌دهیم.

#### 6-4- شرط پایان

برای شرط پایان سه معیار زیر در نظر گرفته شده‌اند: 1- تعداد نسلها در الگوریتم از عددی بیشتر شود. 2- تعداد نسلهایی که بهترین عنصر آنها تغییر نکرده است به عدد خاصی برسد (جمعیت بهتر نشود). 3- دقت کلاس‌بندی عنصر با بهترین Fitness از حد آستانه‌ای بیشتر شود.

#### 5- نتایج آزمایشات

برای تست کارایی الگوریتم ارائه شده، این الگوریتم در محیط MATLAB 7.2.0.232 (R2006a) پیاده‌سازی شده و روی یک دستگاه کامپیوتر با CPU AMD Athlon™ 64 3200+ با 2GHz Clock و 512 MB Memory آزمایش و اجرا گردید. برای کلاس‌بند SVM از کتابخانه‌ی MATLAB [18] استفاده شده است. الگوریتم فوق روی تعدادی مجموعه داده اجرا شدند. اطلاعات مربوط به این مجموعه داده‌ها در جدول 1 آمده است. نتایج تست 10-fold cross validation مربوط به این داده‌ها و نتایج آن با یک SVM ساده که با پارامترهای  $C=1$  و  $\gamma=0.5$  و کل بردار ویژگی صورت پذیرفته مقایسه شده و در جدول 2 آمده است.

جدول 1: مجموعه داده‌های مورد استفاده

نام	تعداد کلاس‌ها	تعداد نمونه‌ها	تعداد ویژگی‌ها
Heart disease	2	270	13
Breast cancer	2	699	10
Iris	3	150	4
Ionosphere	2	351	34
Vowel	11	940	13
Wine	3	178	13

برای انجام این آزمایش هسته‌ی RBF در نظر گرفته شده پارامترهای C و  $\gamma$  آن بهینه می‌شود و تنظیمات به ترتیب

برای تبدیل مدل ژنی پارامترهای C و  $\gamma$  به مدل عددی آنها می‌توان از رابطه‌ی زیر استفاده کرد.

$$X = \min_x + \frac{\max_x - \min_x}{2^l - 1} \times d \quad (12)$$

که X مدل عددی رشته‌ی بیتی،  $\min_x$  مقدار مینیمم پارامتر،  $\max_x$  مقدار ماکزیمم پارامتر، d مقدار معادل دسیمال رشته‌ی بیت و l طول رشته بیت است.

#### 2-4- Crossover

برای عمل Crossover، عملگر Two-way Crossover کلاسیک در نظر گرفته شده است. به این ترتیب که دو عدد تصادفی کوچکتر یا مساوی با طول کوروموزم ایجاد می‌کنیم، و بیت‌های با اندیس‌های بین این دو عدد را در دو عنصر Crossover شونده عوض می‌کنیم.

#### 3-4- Mutation

برای عمل Mutation نیز از عملگر Mutation معمولی استفاده می‌شود. یک عدد تصادفی ایجاد می‌شود و مقدار مربوط به آن خانه عوض (Flip) می‌شود.

#### 4-4- تابع Fitness

دقت کلاس‌بندی و تعداد ابعاد انتخاب شده را می‌توان دو پارامتر بسیار مهم برای ارزیابی یک کوروموزم خاص دانست. عنصری با دقت کلاس‌بندی بالا و تعداد ویژگی یا ابعاد کمتر، میزان مطلوبیت بالایی دارد. برای این دو پارامتر دخیل ضرائب وزن  $W_A$  و  $W_F$  را تعریف می‌کنیم و بسته به میزان اهمیت هر یک از آنها را مقداردهی می‌کنیم.

برای محاسبه‌ی دقت SVM اینگونه عمل می‌کنیم که اگر  $P = 0$  باشد، بیت‌های  $g_f^i$  به عنوان ماسک بردار ویژگی در نظر گرفته می‌شوند. یعنی آن ویژگی که بیت معادلش مقدار 1 دارد انتخاب می‌شود و آنکه 0، انتخاب نمی‌شود. اگر  $P = 1$  باشد یعنی بر روی بردار ویژگی‌های ورودی باید انتقالی صورت گیرد. این انتقال را با PCA صورت می‌دهیم و به تعداد مجموع بیت‌های  $g_f^i$  که 1 هستند ابعاد PCA را در نظر

توانند جواب بهتری بدهند یا از لحاظ زمانی، زمان کمتری برای آموزش ایجاد کند.

## مراجع

- [1] Vapnik, V. N. "The nature of statistical learning theory". New York: Springer, 1995.
- [2] Pontil, M., & Verri, A. "Support vector machines for 3D object recognition". IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(6), 637-646, 1998.
- [3] C.J.C. Burges. "A tutorial on support vector machines for pattern recognition". Data Mining and Knowledge Discovery, 2(2):955-974, 1998.
- [4] Joachims, T. "Text categorization with support vector machines". In Proceedings of European conf. on machine learning (ECML) (pp. 137-142). Chemnitz, DE, 1998.
- [5] Y. Bazi, and F. Melgani. "Toward an Optimal SVM Classification System for Hyperspectral Remote Sensing Images" IEEE Transaction on Geoscience and Remote Sensing, Vol. 44, No. 11, November 2006.
- [6] Vladimir Vapnik, "Statistical Learning Theory", Wiley-Interscience; 1998.
- [7] Simon Haykin, "Neural Networks, A Comprehensive Foundation", Second Edition, Prentice Hall, 1999.
- [8] Frohlich, H., & Chapelle, O. "Feature selection for support vector machines by means of genetic algorithms". Proceedings of the 15th IEEE international conference on tools with artificial intelligence, Sacramento, CA, USA pp. 142-148, 2003.
- [9] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," IEEE Trans. Inf. Theory, vol. IT-14, no. 1, pp. 55-63, Jan. 1968.
- [10] R. Gonzalez and R. Woods, "Digital Image Processing", 2nd Ed., Addison-Wesley, 2001.
- [11] A. Cheryadat, L. Mann Bruce "Why Principal Component Analysis is not an Appropriate Feature Extraction Method for Hyperspectral Data", IEEE 2003.
- [12] Guyon, I., Weston, J., Barnhill, S., & Bapnik, V. Gene selection for cancer classification using support vector machines. Machine Learning, 46(1-3), 389-422, 2002.
- [13] Mao, K. Z. Feature subset selection for support vector machines through discriminative function pruning analysis. IEEE Transactions on Systems, Man, and Cybernetics, 34(1), 60-67, 2004.
- [14] H. LIU, Y. WANG, X. LU, "A METHOD TO CHOOSE KERNEL FUNCTION AND ITS PARAMETERS FOR SUPPORT VECTOR MACHINES", Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, p. 4277-4280, Guangzhou, August 2005.
- [15] Frohlich, H., & Chapelle, O. "Feature selection for support vector machines by means of genetic algorithms". Proceedings of the 15th IEEE international conference on tools with artificial intelligence, Sacramento, CA, USA pp. 142-148, 2003.
- [16] Raymer, M. L., Punch, W. F., Goodman, E. D., Kuhn, L. A., & Jain, A. K. "Dimensionality reduction using genetic algorithms". IEEE Transactions on Evolutionary Computation, 4(2), 164-171, 2000.
- [17] C. L. Huang and C. J. Wang, "A GA-based feature selection and parameters optimization for support vector machines". Expert Systems with Applications 31, pp. 231-240, 2006.
- [18] Support Vector Machine toolbox for Matlab Version 2.51, Anton Schwaighofer. January 2002.

زیر رعایت شده‌اند. اندازه‌ی جمعیت 100، نرخ Crossover 0/75، نرخ 0/02 Mutation و جایگذاری در نسل جدید با تبعیت از Elitism.  $n_f=13$  و  $n_c=13$  برابر با تعداد ویژگی‌های مجموعه داده‌ی مورد استفاده،  $W_A$  و  $W_F$  نیز بسته به میزان اهمیت دقت و تعداد ابعاد انتخاب شده مقداردهی شده‌اند. در اینجا  $W_A=0.8$  و  $W_F=0.2$  در نظر گرفته شده است. جستجوی پارامترهای C و  $\gamma$  به ترتیب در بازه‌های [1 200] و [0.00001 1.2] در نظر گرفته شده است. شرط پایان نیز به این ترتیب اتخاذ شده است که اگر در 50 نسل میزان Fitness بهترین عنصر جمعیت بهبود نبخشد یا تعداد کل نسل‌های تولید شده به 500 رسید یا دقت کلاس‌بند بهترین عنصر از 98٪ بیشتر شد، فرآیند تکرار به اتمام رسیده و بهترین عنصر جمعیت به عنوان خروجی در نظر گرفته شود.

جدول 2: نتایج کلاس‌بندی (تست 10-fold Cross-Validation)

مجموعه داده	Simple SVM (%)	Our Approach (%)
Heart disease	84.5	95.4
Breast cancer	91.24	94.12
Iris	95.5	97.9
Ionosphere	88	96.6
Vowel	92.55	96.1
Wine	97.3	98.5

## 6- نتیجه‌گیری

در این کار، به طور همزمان پارامترهای SVM و زیر مجموعه ویژگی‌های انتخابی را بهینه کردیم. در این بین تصمیم هم گرفته می‌شود که آیا نیازی به انتقال توسط PCA هست یا خیر. برای انجام این کار از الگوریتم ژنتیک استفاده شده است. قبل از این، هیچ کاری برای انتخاب ویژگی یا انتقال در ویژگی‌ها برای SVM انجام نشده است. نتایج آزمایشات خروجی‌های مناسبی را برای دقت کلاس‌بندی و موارد مرتبط نشان داده است.

## 7- کارهای آینده

در مورد بهینه‌سازی کلاس‌بند SVM کارهای فراوانی باقی مانده است. می‌توان دیگر انتقالات چون KLT را به مجموعه فرآیندهای قابل انجام در کورومومزهای الگوریتم ژنتیک اضافه نمود. یا اینکه نوع تابع هسته‌ای را نیز در هنگام بهینه‌سازی در نظر گرفت. در این آزمایش از نوع هسته‌ی RBF استفاده شد و پارامترهای این هسته بهینه شدند. نوع هسته‌های دیگری نیز موجود است که در موارد مختلف می-