# یک ماشین بردار پشتیبان فازی جدید با فازی‌سازی در دو مرحله

**جهانشاه کبودیان** [1]
kabudian@ce.aut.ac.ir

**محمدحسن مرادی** [2]
mhmoradi@aut.ac.ir

**۱**– آزمایشگاه پردازش هوشمند سیگنالهای صوتی و گفتاری، دانشکدهٔ مهندسی کامپیوتر و فناوری اطلاعات،
**۲**– دانشکدهٔ مهندسی پزشکی،
دانشگاه صنعتی امیرکبیر(پلی‌تکنیک تهران)، تهران.

## چکیده

یکی از روشهای نسبتاً جدید در شناسایی و دسته‌بندی الگوها، ماشین بردار پشتیبان یا SVM است. در این مقاله یک SVM فازی جدید پیشنهاد شده است که فازی‌سازی را در دو مرحله دخالت می‌دهد. در مرحلهٔ آموزشSVM ، اهمیت نسبی هر نمونه در میان نمونه‌های یک کلاس را با استفاده از یک تابع عضویت فازی که نسبت به داده‌های پرت(Outlier) غیر حساس است، دخالت داده، و در مرحلهٔ بازشناسی و تصمیم‌گیری نیز، بر اساس منطق فازی، یک تصمیم‌گیری نرم و ملایم را انجام می‌دهد. SVM فازیِ پیشنهاد شده، بر روی پایگاه دادهٔ MNIST از ارقام دستنویس آزمایش گردیده و به ازای کرنل‌های چندجمله‌ای و RBF دارای راندمانی بالاتر از SVM استاندارد برای شناسایی ارقام دستنویس می‌باشد.

**واژه‌های کلیدی:** ماشین بردار پشتیبان فازی، شناسایی الگو، منطق فازی، روشهای مبتنی بر کرنل(هسته)، شناسایی ارقام دستنویس.

## ۱- مقدمه

اولین الگوریتم برای طبقه‌بندی و دسته‌بندی الگوها در سال 1936 توسط Fisher ارائه شد و معیار آن برای بهینه بودن، کم کردن خطای طبقه‌بندی الگوهای آموزشی بوده است. بسیاری از الگوریتم‌ها و روشهایی نیز که تاکنون برای طراحی طبقه‌بندی‌کننده‌های الگو ارائه شده است، از همین استراتژی پیروی می‌کنند. در هیچیک از این روشها خاصیت تعمیم طبقه‌بندی‌کننده بطور مستقیم در تابع هزینهٔ روش دخالت داده نشده است و طبقه‌بندی‌کنندهٔ طراحی شده نیز دارای خاصیت تعمیم‌دهندگی کمی می‌باشد. اگر طراحی دسته‌بندی‌کنندهٔ الگو را بعنوان یک مسأله بهینه‌سازی در نظر بگیریم، بسیاری از این روشها با مشکل بهینه‌های محلی در تابع هزینه مواجهند و در دام بهینه‌های محلی گرفتار می‌آیند. مشکل دیگری نیز وجود دارد و آن، تعیین ساختار و توپولوژی دسته‌بندی‌کننده قبل از طراحی است، که بعنوان مثال تعیین تعداد بهینهٔ گره‌های لایه مخفی در شبکه‌های عصبی MLP، تعداد توابع گاوسی در شبکه‌های عصبی RBF و یا تعداد بهینهٔ حالتها و توابع گاوسی در مدل پنهان مارکف می‌باشد. همه این عوامل باعث می‌شوند که در عمل با روشهای پیشنهاد شدهٔ قبلی نتوانیم به یک دسته‌بندی‌کنندهٔ بهینه برسیم. محقق روسی بنام Vladimir Vapnik در سال 1965 گامی بسیار مهم در طراحی دسته‌بندی‌کننده‌ها برداشت [1] و نظریه آماری یادگیری را بصورت مستحکم‌تری بنا نهاد و ماشین‌های بردار پشتیبان(SVM) را بر این اساس ارائه داد. ماشینهای بردار پشتیبان دارای خواص زیر هستند:

۱- طراحی دسته‌بندی کننده با حداکثر تعمیم ۲- رسیدن به بهینهٔ سراسریِ تابع هزینه ۳- تعیین خودکار ساختار و توپولوژی بهینه برای طبقه‌بندی‌کننده ۴- مدل کردن توابع تمایز غیرخطی با استفاده از هسته‌های غیرخطی و مفهوم حاصلضرب داخلی در فضاهای هیلبرت.

در این مقاله، در بخش دوم، اصول و پایه‌های یک طبقه‌بندی‌کننده از نوع ماشین بردار پشتیبان را توضیح خواهیم داد. در بخش سوم، SVM در حالت جدایی ناپذیر را بررسی می‌کنیم. در بخش چهارم، SVM غیرخطی را معرفی خواهیم کرد و در بخش پنجم، چگونگی طراحی دسته‌بندی

کنندهٔ SVM در حالت چندکلاسه را شرح می‌دهیم. در بخش ششم، SVM فازی را توضیح خواهیم داد و در انتها به ارائه نتایج آزمایشهای انجام شده و نتیجه‌گیری خواهیم پرداخت.

## ۲- ماشین بردار پشتیبان(SVM)

فرض کنید تعدادی از بردارهای ویژگی یا الگوهای آموزشی بصورت $\{x_1, x_2, ..., x_i, ..., x_N\}$ داریم که هر کدام یک بردار ویژگی d بُعدی بوده و دارای برچسب $y_i$ است و $y_i \in \{-1,+1\}$. هدف حل یک مسأله دسته‌بندی دو کلاسه بصورت بهینه است. فرض کنید این دو کلاس را با تابع تمایز f(x) و با یک ابر صفحه H با معادله زیر بخواهیم از هم جدا کنیم:

$$H: w.x + b = 0 \qquad (۱)$$

$$f(x)=\text{sign}(w.x+b) \qquad (۲)$$

بردار وزن w، بردار عمود بر ابرصفحهٔ جداکننده و b مقدار Bias است و منظور از w.x حاصلضرب داخلی می‌باشد. Vapnik ثابت کرد که بُعد VC[۱] برای طبقه‌بندی‌کننده‌هایی از نوع ابر صفحات کانونی[۲]، دارای یک کران بالاست که این کران بالا با توان دوم نُرم بردار وزن یعنی $\|w\|^2$ نسبت مستقیم دارد[3]. در واقع اگر ما $\|w\|^2$ را محدود کرده و مینیمم کنیم، بُعد VC طبقه‌بندی‌کننده را می‌نیمم کرده‌ایم و تخمین ما از مقدار ریسک واقعی بصورت احتمالی دقیق‌تر بوده و خاصیت تعمیم دسته‌بندی‌کننده بیشتر خواهد شد.

$$\|w\| = \left(\sum_{i=1}^{d} w_i^2\right)^{1/2} \qquad (۳)$$

رابطه بین خاصیت تعمیم طبقه‌بندی‌کننده با نُرم بردار وزن $\|w\|$ را می‌توان به طریق دیگر نیز توجیه کرد: فرض کنید داده‌های دو کلاس جدایی‌پذیر باشند و بردارهای ویژگی مرزی کلاس اول روی ابرصفحه $H^+$ و بردارهای ویژگی مرزی کلاس دوم روی ابرصفحه $H^-$ قرار گیرند. ابر صفحات $H^+$ و $H^-$ به این صورت تعریف می‌شوند:

$$\begin{aligned} H^+ &: w.x + b = +1 \\ H^- &: w.x + b = -1 \end{aligned} \qquad (۴)$$

الگوهایی که بر روی ابر صفحات $H^+$ یا $H^-$ قرار می‌گیرند، بردار پشتیبان نامیده می‌شوند. ناحیه بین دو ابر صفحه $H^+$ و $H^-$ را حاشیه یا ناحیه مرزی[۳] گویند. فاصله بین دو ابر صفحه $H^+$ و $H^-$ برابر $\frac{2}{\|w\|}$ خواهد بود. طراحی ابر صفحه با بیشترین عرض ناحیه مرزی یا ناحیه مرزی بهینه[۴] بر این استوار است که با شرط درست طبقه‌بندی‌شدنِ الگوها، عرض ناحیه مرزی حداکثر شود، یعنی $\frac{2}{\|w\|}$ ماکزیمم شود و $\|w\|$ مینیمم گردد. هدف این است که اولاً الگوها درست طبقه‌بندی گردند و ثانیاً بر روی و یا خارج از ناحیه مرزی واقع شوند، یعنی:

$$y_i(w.x_i + b) \geq 1 \quad \text{for } i = 1,2,...,N \qquad (۵)$$

پس در واقع طراحی یک طبقه‌بندی‌کنندهٔ ابرصفحه‌ای با ناحیه مرزی بهینه بصورت زیر خواهد بود:

$$\begin{cases} \underset{w}{\text{Minimize}} & \frac{1}{2}\|w\|^2 \\ \text{Subject to} & y_i(w.x_i + b) \geq 1 \\ & \text{for } i = 1,2,...,N \end{cases} \qquad (۶)$$

واضح است که داریم :

$$\begin{aligned} w &= [w_1\, w_2 \ldots w_d]^T \\ \|w\|^2 &= w.w \end{aligned} \qquad (۷)$$

مسأله فوق یک مسأله بهینه‌سازی مقید از نوع محدب و درجه دوم است. برای حل این مسأله، تابع لاگرانژی زیر را تشکیل داده و ضرایب لاگرانژ $\alpha_i$ را بدست می‌آوریم:

$$L(w,b,\alpha) = \frac{1}{2} w.w - \sum_{i=1}^{N} \alpha_i \left( y_i (w.x_i + b) - 1 \right) \qquad (۸)$$

برای اینکه $(w,b,\alpha)$ جواب مسأله باشد، این جواب باید در شرایط KKT[۵] صدق کند و در نقطه جواب، مشتق L نسبت به w ، b و $\alpha$ برابر صفر باشد. با مساوی قرار دادنِ مشتق برابر با صفر به معادلات زیر خواهیم رسید:

$$\begin{aligned} w &= \sum_{i=1}^{N} \alpha_i y_i\, x_i \\ \sum_{i=1}^{N} \alpha_i\, y_i &= 0 \end{aligned} \qquad (۹)$$

با قرار دادن مقدار w از رابطه فوق در $L(w,b,\alpha)$ به مسأله

---

[1] VC Dimension

[2] Canonical Hyperplane

[3] Margin

[4] Optimal Margin Hyperplane (Maximal Margin Hyp.)

[5] Karush-Kuhn-Tucker Conditions

دوگان[6] برای بهینه‌سازی مقید خواهیم رسید:

$$\begin{cases} \underset{\alpha}{Maximize} & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\ x_i.x_j + \sum_{i=1}^{N}\alpha_i \\ Subject\ to & \alpha_i \ge 0 \quad for\ i=1,2,\dots,N \\ & \sum_{i=1}^{N}\alpha_i y_i = 0 \end{cases} \tag{۱۰}$$

پس از حل این مسألهٔ دوگان بهینه‌سازی، ضرایب لاگرانژ $\alpha_i \ge 0$ بدست می‌آیند. در واقع هر کدام از ضرایب لاگرانژ $\alpha_i$ متناظر با یکی از الگوهای $x_i$ می‌باشند، الگوهای $x_i$ را که متناظر با ضرایب $\alpha_i > 0$ (مثبت) هستند، بردارهای پشتیبان $sv_i$ می‌نامیم. مقدار بردار وزن و b از روابط زیر بدست می‌آیند:

$$\begin{aligned} w &= \sum_{i=1}^{N_{sv}}\alpha_i y_i\ sv_i \\ b_j &= y_j - \sum_{i=1}^{N_{sv}}\alpha_i y_i\ sv_i.sv_j \\ b &= \frac{1}{N_{sv}}\sum_{j=1}^{N_{sv}} b_j \end{aligned} \tag{۱۱}$$

تابع تمایز برای طبقه بندی یک الگوی ورودی $x$ بصورت زیر خواهد بود:

$$f(x) = sign\left(\sum_{i=1}^{N_{sv}}\alpha_i y_i\ x.sv_i + b\right) \tag{۱۲}$$

## ۳-ماشین بردار پشتیبان درحالت جدایی‌ناپذیر

در بخش قبلی، مسأله SVM را برای حالت جدایی‌پذیر حل کردیم، ولی در عمل تقریباً تمامی مسائل بصورت جدایی‌ناپذیر می‌باشند. برای حالت جدایی‌ناپذیر یک دسته از متغیرهای $\in_i$ بنام متغیرهای کمبود[7] تعریف می‌کنیم طوری که شرط زیر برقرار باشد:

$$\begin{aligned} & y_i(w.x_i + b) \ge 1 - \in_i \\ & \in_i \ge 0 \ ,\ \ for\ \ i = 1,2,\dots,N \end{aligned} \tag{۱۳}$$

واضح است که هر چقدر مجموع مقادیر متغیرهای $\in_i$ بیشتر شود، از حالت بهینه دورتر خواهیم شد و خطا بیشتر می‌گردد. پس مسأله بهینه‌سازی مقید را بصورت زیر تعریف می‌کنیم:

$$\begin{cases} \underset{w}{Minimize} & \frac{1}{2}w.w + C\sum_{i=1}^{N}\in_i \\ Subject\ to & y_i(w.x_i + b) \ge 1 - \in_i \\ & \in_i \ge 0\ ,\ for\quad i = 1,2,\dots,N \end{cases} \tag{۱۴}$$

برای این مسأله نیز شرایط KKT را در نقطه جواب تشکیل داده و به مسأله دوگان زیر خواهیم رسید:

$$\begin{cases} \underset{\alpha}{Maximize} & -\frac{1}{2}\sum_{j=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\ x_i.x_j + \sum_{i=1}^{N}\alpha_i \\ Subject\ to & 0 \le \alpha_i \le C \\ & \sum_{i=1}^{N}\alpha_i y_j = 0 \end{cases} \tag{۱۵}$$

همانطور که ملاحظه میشود حل مسأله SVM در حالت جدایی‌ناپذیر مشابه حل آن در حالت جدایی پذیر است با این تفاوت که محدوده تغییرات ضرایب لاگژانر $\alpha_i$ فرق می‌کند. پس از بدست آوردن ضرایب لاگرانژ $\alpha_i$ الگوهایی که ضرایب لاگرانژ آنها در رابطه زیر صدق می کنند، بردار پشتیبان می باشند:

$$0 < \alpha_i \le C \tag{۱۶}$$

مقدار w و شکل تابع تمایز هم مشابه حالت جدایی پذیر خواهد بود. ابر صفحه بدست آمده در حالت جدایی ناپذیر را ابر صفحه با ناحیه مرزی نرم[8] می‌نامند.

## ۴-ماشین بردار پشتیبان غیرخطی

ماشین‌های بردار پشتیبان ذکر شده در قسمت‌های قبل، برای دسته‌بندی الگوهای یک مسأله دو کلاسه، از مرزهای جداکننده خطی و از یک ابرصفحه استفاده می‌کند و در واقع حاصلضرب داخلی بردار ورودی با هر کدام از بردارهای پشتیبان در فضای d بُعدی ورودی محاسبه می‌گردد. Vapnik با استفاده از مفهوم حاصلضرب داخلی در فضاهای هیلبرت و قضیه هیلبرت–اشمیت[9] نشان داد که ابتدا می‌توان بردار ورودی x را با یک تبدیل غیرخطی به یک فضای با بُعد زیاد انتقال داد و در آن فضا حاصلضرب داخلی را انجام داد و ثابت کرد که اگر یک هسته متقارن، شرایط قضیهٔ Mercer را داشته باشد، اعمال این هسته در فضای ورودی با بُعد کم می‌تواند به عنوان حاصلضرب داخلی در یک فضای هیلبرت با بُعد زیاد تلقی شود و محاسبات را به شدت کاهش دهد[1]. بعنوان مثال تابع هسته می‌تواند به فرمهای زیر باشد:

---

[6] Dual Problem
[7] Slack Variables

[8] Soft Margin Hyperplane
[9] Hilbert-Schmidt Theory

$$k(x,y)=(x.y+1)^p \qquad p=2,3,...$$
$$k(x,y)=\exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \qquad (۱۷)$$
$$k(x,y)=\tanh(x.y+\theta)$$

این هسته‌ها به ترتیب هستهٔ چندجمله‌ای، هسته گاوسی و هستهٔ تانژانت هیپربولیک هستند. مسأله بهینه‌سازی دوگان در حالت جدایی‌ناپذیر و غیرخطی بصورت زیر خواهد بود:

$$\begin{cases} \underset{\alpha}{Maximize} & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\,y_i\,y_j\;k(x_i,x_j)+\sum_{i=1}^{N}\alpha_i \\ Subject\;\;to & 0\le\alpha_i\le C \\ & \sum_{i=1}^{N}\alpha_i y_j=0 \end{cases} \qquad (۱۸)$$

بردارهای پشتیبان الگوهایی هستند که ضرایب لاگرانژ متناظر آنها در رابطه $0<\alpha_i\le C$ صدق کند. تعدادی از بردارهای پشتیبان که ضرایب لاگرانژ متناظر آنها در رابطه $0<\alpha_i<C$ صدق کند و تعداد آنها $N_b$ است، برای محاسبه b استفاده می‌شود:

$$b_j=y_j-\sum_{i=1}^{N_{sv}}\alpha_i y_i\;k(sv_i,sv_j)$$
$$b=\frac{1}{N_b}\sum_{j=1}^{N_b}b_j \qquad (۱۹)$$

تابع تصمیم‌گیری بصورت زیر خواهد بود:

$$f(x)=Sign\left(\sum_{i=1}^{N_{sv}}\alpha_i y_i\;\;k(x,sv_i)+b\right) \qquad (۲۰)$$

## ۵- ماشین بردار پشتیبان در حالت چندکلاسه

تئوری طراحی ابر صفحه با ناحیه مرزی بهینه، برای یک مسأله دو کلاسه ارائه شده است ولی به راحتی می‌توان یک ماشین بردار پشتیبان ساخت که در حالت چندکلاسه نیز عمل دسته‌بندی را به خوبی انجام دهد. برای مسأله SVM در حالت چندکلاسه می‌توان چندین استراتژی در پیش گرفت[3]: ۱- یکی درمقابل بقیه[۱۰] ۲- دسته‌بندی کردن زوج– زوج[۱۱] ۳- تعریف توابع هدف چند کلاسه[۱۲] ۴- استفاده از کُدهای تصحیح خطا[۱۳] . در این مقاله ما از روش دوم برای یک مسأله چند کلاسه استفاده کرده‌ایم.

## ۶- ماشین بردار پشتیبان فازی (FSVM)

در دسته‌بندی کنندهٔ SVM استاندارد، اهمیت میزان خطا (مقدار متغیرهای $\in_i$) به ازای نمونه‌های آموزشی مختلف یکسان است، در حالیکه منطقاً نباید چنین باشد. با استفاده از منطق فازی، می‌توانیم میزان اهمیت هر نمونه را در فاز آموزش دخالت دهیم. همچنین می‌توانیم با استفاده از منطق فازی، در مرحله تصمیم‌گیری به جای یک تصمیم‌گیری خشن(Hard) (در SVM استاندارد)، یک تصمیم‌گیری نرم را انجام دهیم. در قسمت‌های بعدی به شرح چگونگی دو عمل فوق خواهیم پرداخت.

### ۶-۱- دخالت دادن میزان اهمیت نمونه‌ها

SVM استاندارد، نمونه‌های آموزشی را بصورت زوجهای $(x_i,y_i)$ در نظر می‌گیرد و $y_i\in\{-1,+1\}$ . حال برای دخالت دادن میزان اهمیت هر نمونه، نمونه‌ها را بصورت سه‌تایی $(x_i,y_i,s_i)$ در نظر می‌گیریم. $s_i$ در واقع میزان درجه عضویت نمونه $x_i$ به کلاس خودش است. برای دخالت دادن اهمیت هر نمونه، فرمولاسیون SVM را بصورت زیر تغییر می‌دهیم[6] :

$$\begin{cases} \underset{w}{Minimize} & \frac{1}{2}w.w+C\sum_{i=1}^{N}s_i\in_i \\ Subject\;to & y_i(w.x_i+b)\ge 1-\in_i \\ & \in_i\ge 0\;,\;\;for\quad i=1,2,...,N \end{cases} \qquad (۲۱)$$

با این تغییر، مسأله دوگان بهینه‌سازی در SVM بصورت زیر خواهد شد:

$$\begin{cases} \underset{\alpha}{Maximize} & -\frac{1}{2}\sum_{j=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\;x_i.x_j+\sum_{i=1}^{N}\alpha_i \\ Subject\;\;to & 0\le\alpha_i\;\le s_i C \\ & \sum_{i=1}^{N}\alpha_i y_j=0 \end{cases} \qquad (۲۲)$$

در واقع تفاوت SVM فازی با SVM استاندارد در این است که حد فوقانی ضریب لاگرانژ $\alpha_i$ برابر $s_iC$ است، در حالیکه در SVM استاندارد این حد برابر C می‌باشد. بردارهای پشتیبان الگوهایی خواهند بود که ضرایب لاگرانژ متناظر آنها در رابطه $0<\alpha_i\le s_iC$ صدق کند. تعدادی از بردارهای پشتیبان که ضرایب لاگرانژ متناظر آنها در رابطه $0<\alpha_i<s_iC$ صدق کند، برای محاسبه b استفاده می‌شود.

حال باید میزان عضویت $s_i$ را به ازای نمونه $x_i$ تعیین کنیم. در مرجع [6] و [7] ، $s_i$ براساس نسبت فاصله نمونه از مرکز کلاس به فاصله دورترین نمونه همان کلاس از مرکز کلاس

---

[10] One versus the Rest
[11] Pairwise Classification
[12] Multiclass Objective Functions
[13] Error-Correcting Codes

بدست آورده شده است. این روش به داده‌های پرت(Outlier) حساس است و مناسب نمی‌باشد. در اینجا ما $s_i$ را به طریق زیر محاسبه کرده‌ایم:

$$s_i = \exp\left(-\frac{1}{2}(x_i-\mu)^T \sum{}^{-1}(x_i-\mu)\right)+\varepsilon \qquad (۲۳)$$

در اینجا $\varepsilon$ یک عدد کوچک با مقدار 0.01 قرار داده شده است. $\mu$ و $\Sigma$ به ترتیب بردار میانگین و ماتریس کوواریانس کلاس مربوط به نمونه $x_i$ می‌باشد. برای سادگی و کاهش محاسبات، می‌توان ماتریس کوواریانس را بصورت قطری فرض کرد. با روش فوق و اعمال $s_i$ در فاز آموزش SVM ، در واقع میزان اهمیت هر نمونه ($s_i$) را در خطای آن نمونه ($\in_i$) ضرب کرده‌ایم و بدین ترتیب اثر داده‌های پرت را کاهش داده‌ایم.

### ۶-۲- تصمیم‌گیری نرم با استفاده از منطق فازی

در این مقاله، حل مسأله دسته‌بندی در حالت چندکلاسه و بصورت زوج- زوج(Pairwise) مورد نظر است. اگر بردار وزن و مقدار سطح آستانه برای جداسازی داده‌های کلاس i و j پس از آموزش به طریق گفته شده در بخش ۷-۱، به ترتیب $w_{ij}$ و $b_{ij}$ باشد، آنگاه تابع تصمیم‌گیری(تابع تمایز) برای جداسازی کلاس i و j چنین خواهد بود:

$$\begin{aligned} D_{ij}(x) &= w_{ij}^T x + b_{ij} \qquad , \quad i \neq j \\ D_{ji}(x) &= -D_{ij}(x) \end{aligned} \qquad (۲۴)$$

در SVM استاندارد، برای جداسازی کلاس i از سایر کلاسها از تابع تمایز زیر استفاده می‌شود:

$$D_i(x) = \sum_{j=1, j\neq i}^{C} sign(D_{ij}(x)) \qquad (۲۵)$$

C تعداد کلاس‌هاست. کلاس بردار x چنین تعیین می‌شود:

$$c(x) = \arg \underset{i}{Max}\ D_i(x) \qquad (۲۶)$$

مشکلی که در این توابع تمایز وجود دارد، این است که نواحی غیرقابل دسته‌بندی(Unclassifiable Regions) وجود دارند. یک راه برای حل این مشکل این است که تصمیم‌گیری را بصورت نرم و با استفاده از منطق فازی انجام دهیم. برای این کار، هنگام تصمیم‌گیری و تمایز بین دو کلاس i و j ، تابع عضویت $m_{ij}(x)$ را تعریف می‌کنیم. $m_{ij}(x)$ بدین معنی است که در هنگام دسته‌بندی بردار x ، به چه میزان کلاس i برنده است (و کلاس j بازنده) و میزان تعلق بردار x به کلاس i چقدر است:

$$m_{ij}(x) = \frac{1}{1+\exp(-D_{ij}(x))} \qquad (۲۷)$$

حال برای آنکه بردار x در تمام زوجهای کلاس (i,j) به کلاس i متعلق باشد و کلاس i در تمام این تصمیم‌گیری های دودویی برنده شود، از AND فازی و اپراتور Min استفاده می‌کنیم[8]:

$$m_i(x) = \underset{j, j\neq i}{Min}\quad m_{ij}(x) \qquad (۲۸)$$

پس از تعیین $m_i(x)$ ها، بردار x را به کلاس c(x) نسبت می‌دهیم:

$$c(x) = \arg \underset{i}{Max}\ m_i(x) \qquad (۲۹)$$

با استفاده از منطق فازی و انجام یک تصمیم‌گیری نرم و ملایم، اولاً مشکل نواحی غیرقابل دسته‌بندی برطرف خواهد شد و ثانیاً تصمیم‌گیری و دسته‌بندی بصورت نرم و فازی انجام می‌شود.

## ۷- آزمایش‌های انجام شده

در این مقاله برای مقایسهٔ SVM فازی ارائه شده و SVM استاندارد، مسأله شناسایی ارقام دستنویس را انتخاب کرده‌ایم. برای آنکه نتایج بدست آمده مطمئن و قابل تعمیم باشد، از یک پایگاه دادهٔ مشهور بنام MNIST که دادگان آن رایگان می‌باشد، استفاده کرده‌ایم. در دادگان MNIST که شامل ارقام دستنویس 0 تا 9 به زبان انگلیسی است، هر رقم(نمونهٔ آموزشی)، بصورت یک تصویر خاکستری 28×28 پیکسل ذخیره شده است. تعداد دادگان آموزشی ۶۰۰۰۰ رقم و تعداد دادگان آزمایشی ۱۰۰۰۰ رقم است. در این مقاله، برای آموزش مدلهای SVM ، از هر رقم ۵۰۰ نمونه برای آموزش استفاده شده و در دوره آزمایش و تست سیستم از ۵۰۰ نمونه دیگر استفاده کرده‌ایم. برای مسأله ارقام دستنویس که یک مسالهٔ ۱۰ کلاسه است، ۴۵ دسته‌بندی کنندهٔ فازی برای تمام زوجهای (i,j) آموزش می‌دهیم. ویژگی‌های استخراج شده از تصویر هر رقم، ویژگی‌های گشتاورهای هندسی مقیاس شده بودند[5]. ثابت می‌شود که گشتاورهای مقیاس شده نسبت به جابجایی و اندازه الگو غیر حساس می‌باشند. اگر $\mu'$ گشتاورهای هندسی مقیاس شده باشند، بردارهای ویژگی استفاده شده یک بردار ویژگی ۲۲ بُعدی x بصورت زیر است:

$$\begin{aligned} x_1 &= \{\mu'_{pq} \ : p,q = 0,1,2,3,4\} \\ x &= x_1 - \{\mu'_{00}, \mu'_{01}, \mu'_{10}\} \end{aligned} \qquad (۳۰)$$

پس از استخراج بردارهای ویژگی به روش فوق برای بالا بردن راندمان SVM، بردارهای ویژگی به طریق زیر نرمالیزه شدند:

$$M_x = \underset{i,j}{Max}\ \left|x_{ij}\right| \qquad (۳۱)$$

$$x'_{ij} = \frac{x_{ij}}{M_x} \qquad (۳۲)$$

منظور از $x_{ij}$ المان j ام از بردار آموزشی i ام است. در واقع ماکزیمم مقدار $|x_{ij}|$ به ازای تمام نمونه‌های تمام کلاسها محاسبه می‌گردد و سپس عناصر بردار ویژگی $x_i$ بر این مقدار($M_x$) تقسیم می‌شوند و با این کار تمام المانهای بردارهای ویژگی در فاصله [-1,+1] قرار می‌گیرند. این کار دو مزیت دارد : ۱- همگرایی در آموزش SVM را تسریع می‌کند(مخصوصاً در کرنل چندجمله‌ای) ۲- تعیین مقدار بهینهٔ $\sigma$ در کرنل RBF آسان‌تر خواهد بود. در این مقاله از دو نوع کرنل چندجمله‌ای و کرنل گاوسی(RBF) استفاده شده است. درجهٔ کرنل چندجمله‌ای برابر p=3 قرار داده شد و مقدار $\sigma$ در کرنل RBF برابر $\sigma$ =0.5 انتخاب گردید. مقدار این پارامترها در SVM استاندارد و SVM فازی یکسان بوده است. مقدار C برابر 100 قرار داده شد. جدول (۱) نتایج بازشناسی ارقام دستنویس بازای داده‌های آموزشی و آزمایشی و کرنل‌های چندجمله‌ای و گاوسی در SVM استاندارد را نشان می‌دهد. جدول (۲) ، همین نتایج را با استفاده از SVM فازی نشان می‌دهد. همانطورکه ملاحظه می‌گردد، روش SVM فازی پیشنهاد شده، همیشه از SVM استاندارد بهتر عمل کرده است، چون SVM فازی اولاً اهمیت نسبی نمونه‌های مختلف را در نظر گرفته و ثانیاً تصمیم‌گیری را بصورت نرم و ملایم با استفاده از منطق فازی انجام می‌دهد. همچنین مشاهده می‌شود که کارایی کرنل چندجمله‌ای بیش از کرنلRBF است. این می‌تواند بدان دلیل باشد که تنظیم پارامتر $\sigma$ در کرنل RBF بطور دقیق مشکل است.

**جدول ۱- نتایج بازشناسی ارقام دستنویس با SVM استاندارد**

| | Standard SVM | |
|---|---|---|
| | Polynomial Kernel | RBF Kernel |
| آموزشی | ٪۹۴/۴ | ٪۹۳/۹ |
| آزمایشی | ٪۹۳/۱ | ٪۹۱/۸ |

**جدول ۲- نتایج بازشناسی ارقام دستنویس با SVM فازی**

| | Fuzzy SVM | |
|---|---|---|
| | Polynomial Kernel | RBF Kernel |
| آموزشی | ٪۹۵/۱ | ٪۹۴/۱ |
| آزمایشی | ٪۹۴/۲ | ٪۹۲/۰ |

## ۸- نتیجه‌گیری و جمع‌بندی

در این مقاله یک SVM فازی جدید پیشنهاد شد که در مرحلهٔ آموزشSVM ، اهمیت نسبی نمونه‌های آموزشی را دخالت داده و همچنین در مرحله دسته‌بندی، تصمیم‌گیری را بصورت نرم و ملایم انجام می‌دهد و مشکل نواحی غیرقابل دسته‌بندی را ندارد. SVM فازی پیشنهاد شده بر روی پایگاه دادهٔ غنی MNIST آزمایش شد. بردارهای ویژگی استفاده شده، از نوع گشتاورهای هندسی مقیاس شده بوده و پس از استخراج ویژگی، نرمالیزه شده‌اند. در نهایت، SVM فازی با دو نوع کرنل RBF و چندجمله‌ای بکار گرفته شد و در هر دو حالت، SVM فازی از SVM استاندارد نظیر آن راندمان بالاتری داشته است. این نتیجه، تأثیر دخالت دادن اهمیت نسبی هر نمونه در مرحله آموزش و نیز انجام تصمیم‌گیری بصورت نرم را در مرحله دسته‌بندی نشان می‌دهد.

## مراجع


[1] C. Cortes, V. Vapnik, “Support-Vector Networks”, Machine Learning, Vol. 20, pp. 273-297, 1995.

[2] C.J.C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition”, Data Mining and Knowledge Discovery, Vol. 2, pp. 121-167, 1998.

[3] B. Schölkopf, A.J. Smola, Learning with Kernels, MIT Press, Cambridge, MA, 2002.

[4] B. Schölkopf et al., “Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers”, IEEE Trans. on Signal Processing, Vol. 45, No. 11, pp. 2758-2765, Nov. 1997.

[5] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Addison Wesley, Reading, MA, 1992.

[6] C.-F. Lin, S.-D. Wang, “Fuzzy Support Vector Machines”, IEEE Trans. on Neural Networks, Vol. 13, No. 2, pp. 464-471, March 2002.

[7] H.-P. Huang, Y.-H. Liu, “Fuzzy Support Vector Machines for Pattern Recognition and Data Mining”, International Journal of Fuzzy Systems, Vol. 4, No. 3, pp. 826-835, Sep. 2002.

[8] S. Abe, T. Inoue, “Fuzzy Support Vector Machines for Multiclass Problems”, European Symposium on Artificial Neural Networks (ESANN’2002), pp. 113-118, Bruges, Belgium, April 2002.