

بهمن ۱۴۰۰

یازدهمین سمینار بین‌المللی

جبر خطی و کاربردهای آن



کاربرد تقریب کم-رتبه نیستروم در محاسبه هسته متعامد لژاندر جهت افزایش دقت تشخیص زودهنگام کووید-۱۹

علیرضا افضل آقائی نائینی^۱ * و کورش پرند^۱

^۱گروه علوم داده‌ها و کامپیوتر، دانشکده ریاضی، دانشگاه شهید بهشتی، تهران، ایران

چکیده

همه‌گیری کووید-۱۹ به یکی از بزرگ‌ترین چالش‌های قرن بیست و یکم تبدیل شده است. علائم رایج این بیماری، شامل تب، سرفه و تنگی نفس می‌باشد. شباهت این علائم به آنفولانزای فصلی، امکان تشخیص سریع و کم‌هزینه کووید-۱۹ را از متخصصان سلب کرده است. در این پژوهش یک الگوریتم هوشمند جهت دسته‌بندی افراد مبتلا به کووید-۱۹ از میان افراد مشکوک ارائه می‌شود. برای این منظور، ابتدا کاربرد چندجمله‌ای‌های متعامد به عنوان توابع هسته بررسی شده و سپس جهت غلبه بر پیچیدگی‌های محاسباتی، از تقریب کم-رتبه نیستروم و تجزیه مقادیر منفرد استفاده می‌کنیم. در پایان به کمک دسته‌بند گرادیان کاهشی، یک مدل بلادرنگ جهت تشخیص بیماری ارائه خواهیم کرد. نتایج حاصله بیانگر دقت ۹۸ درصدی الگوریتم ارائه شده می‌باشد.

واژه‌های کلیدی: کووید-۱۹، تجزیه مقادیر منفرد، چندجمله‌ای‌های متعامد، یادگیری ماشین، هوش مصنوعی
 کد موضوع بندی ریاضی [۲۰۱۰]: 68T10, 15A23

۱ مقدمه

کووید-۱۹ (COVID-19) یک بیماری عفونی ناشی از کروناویروس سندرم حاد تنفسی ۲ (SARS-CoV-2) است که در دسامبر سال ۲۰۱۹ در ووهان چین پدیدار شد. این بیماری واگیردار با تحت تاثیر قرار دادن سیستم تنفسی افراد، منجر به بروز علائمی همچون مشکلات تنفسی، تب و سرفه خشک می‌گردد. معمولاً افرادی که سیستم ایمنی قوی دارند، پس از ابتلا به این بیماری، بدون انجام درمان خاصی در نهایت بهبود می‌یابند. با این حال، افراد مسن و کسانی که دارای بیماری‌های زمینه‌ای همچون بیماری‌های قلبی-عروقی، دیابت، بیماری‌های تنفسی و سرطان هستند، آسیب پذیری بیشتری دارند. بر اساس یافته‌های سازمان بهداشت جهانی علائم کمتر شایع این ویروس، شامل سردرد، گلودرد، اسهال، از دست دادن بویایی و از دست دادن گفتار می‌باشد.

علی‌رغم تلاش‌های جامعه برای مهار ویروس، کووید-۱۹ با جهش‌های متوالی همچنان در حال گسترش است، شیوع بیماری می‌تواند منجر به افزایش تقاضا در منابع بیمارستانی و کمبود تجهیزات پزشکی، کارکنان مراقبت‌های بهداشتی و البته کیت‌های تست کووید-۱۹ شود. دسترسی محدود به این کیت‌ها می‌تواند تشخیص زودهنگام بیماری را مختل کرده و ارائه بهترین مراقبت ممکن برای بیماران مشکوک را سخت کند. در نتیجه، وجود یک سیستم پیش‌بینی خودکار که هدف آن تعیین وجود ویروس کووید-۱۹ در یک فرد مشکوک است ضروریست.

الگوریتم‌های یادگیری ماشین، از ابزارهای مفید و کاربردی در تشخیص زودهنگام انواع بیماری‌های مختلف همچون دیابت، سرطان و... هستند. این الگوریتم‌ها با دریافت علائم بیماری، با دقت مناسب به پیش‌بینی وجود یا عدم وجود بیماری در بدن فرد می‌پردازند. دانشمندان این مدل‌ها را جهت تشخیص بیماری کووید-۱۹ نیز توسعه داده‌اند [۳، ۵]. در این رویکردها، الگوریتم به‌طور صریح تعریف نشده و به مدل اجازه داده می‌شود برخی از بخش‌های خود را با کسب تجربه به‌روز کرده و نتیجه نهایی را بهبود دهد. در اینجا منظور از تجربه یک مجموعه داده شامل چند نمونه مختلف بوده که هر کدام دارای چندین ویژگی مستقل و یک برچسب صحیح می‌باشند.

* سخنران. آدرس ایمیل: alirezaafzalaghaei@gmail.com

در میان رویکردهای حل مسائل یادگیری ماشین، روش‌های مبتنی بر هسته، دسته‌ای از الگوریتم‌ها هستند که مدل‌سازی خود را بر اساس شباهت میان داده‌ها انجام می‌دهند. در این روش‌ها، تابع هسته داده‌ها را به طور ضمنی به یک فضای با ابعاد بالا نگاشت کرده و عملیات یادگیری در آن فضا انجام می‌شود. از معروف‌ترین توابع هسته می‌توان به هسته خطی، چندجمله‌ای و تابع پایه شعاعی گوسی اشاره کرد. اخیراً نشان داده شده است که چندجمله‌ای‌های متعامد نیز می‌توانند به عنوان تابع هسته به کار گرفته شده و دقت مدل یادگیری ماشین را افزایش دهند [۴].

استفاده از توابع متعامد به عنوان هسته در مدل‌های یادگیری ماشین، منجر به افزایش پیچیدگی محاسباتی و زمان اجرا شده است. در این پژوهش به کمک روش‌های تقریب کم-رتبه، یک تقریب از ماتریس هسته را بدست آورده و به تعیین وجود یا عدم وجود ویروس کووید-۱۹ در افراد مشکوک می‌پردازیم. به این منظور در بخش بعد به معرفی پیش‌نیازهای لازم پرداخته و در بخش سوم مجموعه داده مورد استفاده تحقیق را شرح می‌دهیم. در بخش چهارم یک الگوریتم سریع برای تعیین وجود ویروس کووید-۱۹ در بدن افراد ارائه خواهیم کرد. در پایان نیز با ارائه نتایج و جدول یافته‌های تحقیق به نتیجه‌گیری می‌پردازیم.

۲ پیش‌نیازها

در این بخش به معرفی توابع متعامد و چندجمله‌ای‌های لژاندر، روش تقریب کم-رتبه نیستروم برای ماتریس‌های نیمه معین مثبت متقارن و همچنین دسته‌بند گردایان کاهش می‌پردازیم.

۱.۲ چندجمله‌ای‌های متعامد

چندجمله‌ای‌های متعامد دنباله‌ای نامتناهی از چندجمله‌ای‌هایی است که دو به دو بر یکدیگر عمودند. خاصیت عمود بودن توابع به کمک تعریف ضرب داخلی

$$\langle f, g \rangle_w = \int_a^b f(x)g(x)w(x)dx \quad (1)$$

بیان می‌شود. چندجمله‌ای‌های لژاندر، چبیشف، ژاکوبی و گگنباور از معروف‌ترین توابع متعامد بوده که در بازه‌ی $[-1, 1]$ براساس تابع وزن مربوطه خاصیت تعامد دارند. چندجمله‌ای‌های لژاندر به دلیل داشتن تابع وزن واحد $w(x) \equiv 1$ در بین پژوهشگران از محبوبیت خاصی برخوردار است. این چندجمله‌ای‌ها با فرمول بازگشتی زیر تعریف می‌گردند.

$$\begin{aligned} P_0(x) &= 1, \quad P_1(x) = x, \\ (n+1)P_{n+1}(x) &= (2n+1)xP_n(x) - nP_{n-1}(x), \quad n \geq 1. \end{aligned} \quad (2)$$

جهت تعریف این رده از توابع در بازه‌ی متناهی دلخواه $[a, b]$ می‌توان از نگاشت $\varphi(x) = (2x-a-b)/(b-a)$ استفاده کرد و چندجمله‌ای‌های لژاندر انتقال یافته را به صورت $\tilde{P}(x) = P(\varphi(x))$ محاسبه نمود. جهت مطالعه بیشتر در مورد این توابع خواننده را به مرجع [۴، ۱] ارجاع می‌دهیم. لازم به ذکر است این توابع در افزایش سرعت حل دستگاه‌های معادلات خطی نیز به کار گرفته شده‌اند [۲].

۲.۲ روش تقریب هسته نیستروم

در روش‌های مبتنی بر هسته برای یک مجموعه داده‌ی $D = \{(x_i, y_i)\}_{i=1}^N$ ماتریس نیمه معین مثبت متقارن $\Omega_{N \times N}$ به صورت

$$\Omega_{i,j} = K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle, \quad i, j = 1, 2, \dots, N. \quad (3)$$

تعریف خواهد شد. در اینجا تابع $K(x, t)$ هسته‌ی مدل یادگیری ماشین است. این تابع، ضرب داخلی تابع گسترش ویژگی $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}^k$ ، که در آن $k > d$ است را به صورت ضمنی محاسبه می‌کند. در مسائلی که تعداد داده‌های مسئله بزرگ باشد، بدست آوردن ماتریس Ω از نظر محاسباتی کارا نیست. از این رو دانشمندان رویکردهایی جهت تقریب این ماتریس، با تعداد محدودی از داده‌ها ارائه کرده‌اند. یکی از معروف‌ترین این روش‌ها، تقریب کم-رتبه نیستروم می‌باشد. در قضیه‌ی ۱.۲ به معرفی این روش می‌پردازیم.

قضیه ۱.۲. ماتریس نیمه معین مثبت متقارن $A \in \mathbb{R}^{n \times n}$ را در نظر بگیرید. فرض کنید $C_{n \times m}$ یک ماتریس شامل $m \ll n$ ستون از ماتریس A باشد که به صورت تصادفی انتخاب شده است. در این صورت می‌توان ماتریس A و C را به صورت زیر بازنویسی کرد.

$$C = \begin{bmatrix} W \\ S \end{bmatrix} \text{ و } A = \begin{bmatrix} W & S^T \\ S & B \end{bmatrix}. \quad (۴)$$

در اینجا $W \in \mathbb{R}^{m \times m}$, $S \in \mathbb{R}^{(n-m) \times m}$ و $B \in \mathbb{R}^{(n-m) \times (n-m)}$ می‌باشد. در این صورت ماتریس W نیز نیمه معین مثبت خواهد بود. حال برای هر $k \leq m$ تقریب نیستروم مرتبه k ماتریس A به صورت

$$\hat{A}_k = CW_k^\dagger C^T \quad (۵)$$

تعریف می‌شود. در اینجا W_k بهترین تقریب با رتبه k ماتریس W و $W_k^\dagger$ شبه معکوس مور-پنروز این ماتریس است.

اثبات. برای اثبات به مرجع [۶] مراجعه کنید. $\square$

با استفاده از این قضیه می‌توان یک تقریب کم-رتبه از ماتریس هسته Ω ارائه کرد. برای این منظور ابتدا یک مجموعه اندیس $I \subseteq \{1, 2, \dots, N\}$ انتخاب کرده و ماتریس

$$W_{i,j} = K(x_i, x_j), \quad i, j \in I \quad (۶)$$

را تشکیل می‌دهیم. جهت محاسبه بهترین تقریب مرتبه k ماتریس W از تجزیه مقادیر منفرد استفاده می‌شود.

$$W = U \Sigma V^T \quad (۷)$$

در اینجا $U, V \in \mathbb{R}^{m \times m}$ ماتریس‌های متعامد و Σ نیز یک ماتریس قطری است که عناصر آن شامل مقادیر منفرد ماتریس W بوده که به صورت نزولی مرتب شده‌اند. ماتریس کم-رتبه W_k با استفاده از این تجزیه تقریب زده می‌شود:

$$W_k = U_k \Sigma_k V_k^T. \quad (۸)$$

شبه معکوس ماتریس W نیز به کمک تجزیه مقادیر منفرد محاسبه می‌گردد:

$$W_k^\dagger = \sum_{i=1}^k \sigma_i^{-1} U_{:,i} U_{:,i}^T. \quad (۹)$$

در اینجا σ_i بیانگر i -امین بزرگترین مقدار منفرد ماتریس W و $U_{:,i}$ بیانگر i -امین ستون ماتریس متعامد U است. برای تقریب ماتریس هسته Ω لازم است ماتریس C نیز تعریف شود، برای این منظور

$$C_{i,j} = K(x_i, x_j), \quad i \in \{1, 2, \dots, N, j \in I\} \quad (۱۰)$$

را توسط مجموعه اندیس‌های I به دست می‌آوریم. حال به کمک معادله (۵) می‌توان تقریب کم-رتبه ماتریس هسته را بدست آورد.

۳.۲ دسته‌بند خطی گرادیان کاهشی

دسته‌بند گرادیان کاهشی یک مدل خطی در یادگیری ماشین است که به جهت مقیاس پذیری و کارایی محاسباتی مورد توجه دانشمندان قرار گرفته است. در این مدل برای مجموعه داده‌ی $D = \{(x_i, y_i)\}_{i=1}^N$ ، که در آن $x_i \in \mathbb{R}^m$, $y_i \in \{-1, 1\}$ است، مدل خطی

$$y(x) = \text{sign}(w^T x + b) \quad (۱۱)$$

با پارامترهای مجهول w, b در نظر گرفته می‌شود. با مفروض بودن تابع خطای محدب $L: \mathbb{R}^2 \rightarrow \mathbb{R}$ که با دریافت پاسخ صحیح و پیش‌بینی مدل، خطای پیش‌بینی را محاسبه می‌کند، الگوریتم گرادیان کاهشی به صورت

$$w_i^{\tau+1} = w_i^\tau - \eta * \frac{\partial L}{\partial w_i} \quad (۱۲)$$

نوشته می‌شود. با اجرای این الگوریتم به تعداد تکرار مشخصی، پارامترهای بهینه برای پیش‌بینی دسته‌بند به دست می‌آید. انتخاب توابع خطای مختلف، مدل‌های خطی یادگیری ماشین همچون رگرسیون لاجستیک و یا ماشین بردار پشتیبان را شبیه سازی می‌کند.

۳ تشخیص کووید-۱۹

در این بخش پس از معرفی مجموعه داده‌ی مورد استفاده، به ارائه روش پیشنهادی بر اساس توابع متعامد لژاندر و الگوریتم دسته‌بند گرادیان کاهشی می‌پردازیم.

۱.۳ مجموعه داده

در این پژوهش از مجموعه داده متن باز Symptoms and COVID Presence^۱ استفاده می‌کنیم. این مجموعه شامل ۵۴۳۴ نمونه‌ی مختلف است که هرکدام شامل ۲۰ ویژگی مستقل و یک ویژگی وابسته است. از این تعداد ۴۳۸۳ نفر مبتلا به کووید-۱۹ بوده در حالی که ۱۰۵۱ نفر علائم این بیماری را داشته اما به کووید-۱۹ مبتلا نیستند. در جدول ۱ ویژگی‌های این مجموعه داده، شرح داده شده است.

نام مشخصه	درصد مبتلایان به کووید-۱۹	نام مشخصه	درصد مبتلایان به کووید-۱۹
مشکل تنفسی	۷۶٫۹٪	احساس خستگی	۵۰٫۸٪
تب	۸۵٫۷٪	مشکل گوارشی	۴۶٫۹٪
سرفه خشک	۸۸٫۵٪	سفر خارج از کشور	۵۵٫۹٪
گلو درد	۸۳٫۷٪	تماس با بیمار کووید-۱۹	۵۸٫۹٪
آبریزش بینی	۵۴٫۲٪	شرکت در گردهمایی بزرگ	۵۵٫۷٪
آسم	۴۸٫۵٪	بازدید از مکان‌های عمومی	۵۴٫۸٪
بیماری مزمن ریه	۴۵٫۸٪	اشتغال اعضای خانواده در مکان‌های عمومی	۴۵٫۵٪
سردرد	۴۹٫۷٪	استفاده از ماسک	-
بیماری قلبی	۴۷٫۱٪	استفاده از محصولات ضد عفونی	-
دیابت	۴۸٫۶٪	مبتلا به کووید-۱۹	۱۰۰٪
فشار خون بالا	۵۱٫۵٪		

جدول ۱: ویژگی‌های مجموعه داده و درصد افرادی که دارای هر یک از علائم فوق بوده و مبتلا به کووید-۱۹ می‌باشند.

پیش پردازش داده‌های ورودی جهت آموزش یک مدل یادگیری ماشین، تاثیر زیادی در بدست آوردن دقت مناسب مدل دارد. پس از انجام عملیات پیش پردازش این داده‌ها، دریافته‌ایم که دو ویژگی استفاده از ماسک و استفاده از محصولات ضد عفونی، برای تمامی نمونه‌ها دارای مقدار خیر بوده است. از این رو این دو ویژگی از مجموعه داده حذف می‌گردد.

۲.۳ چندجمله‌ای‌های لژاندر به عنوان هسته

انتخاب درست توابع هسته در مدل‌های یادگیری ماشین می‌تواند تاثیر زیادی بر دقت مدل داشته باشد [۵]. یک شرط لازم و کافی برای قابل استفاده بودن تابع دلخواه $K(x, t)$ به عنوان تابع هسته مدل یادگیری ماشین، توسط مرسر^۲ ارائه شده است [۴]. توابع هسته رایج در زمینه یادگیری ماشین که در شرط مرسر نیز صدق می‌کنند عبارت‌اند از:

- هسته خطی: $K(x, t) = x^T t$

- هسته چندجمله‌ای: $K(x, t; d) = (1 + x^T t)^d$

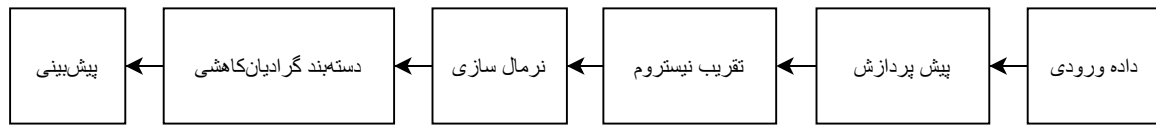
- هسته پایه شعاعی: $K(x, t; \gamma) = \exp(\gamma \|x - t\|)$

در این پژوهش چندجمله‌ای‌های متعامد لژاندر انتقال یافته به بازه $[0, 1]$ از درجه $q - 1$ به عنوان تابع هسته مدل‌های یادگیری ماشین به کار گرفته می‌شوند. این توابع طبق تعریف هسته، برای ورودی با تک ویژگی به صورت

$$K(x, t; q) = \langle \mathbf{P}_q(x), \mathbf{P}_q(t) \rangle = \sum_{i=0}^q P_i(x) P_i(t), \quad (13)$$

¹<https://www.kaggle.com/hemanthhari/symptoms-and-covid-presence>

²Mercer's theorem



شکل ۱: الگوریتم تشخیص ابتلا به کووید-۱۹

تعریف می‌شوند. همچنین برای مجموعه داده با چند ویژگی $x, t \in \mathbb{R}^d$ با استفاده از خواص تابع هسته، هسته‌ی جدید به صورت

$$K(x, t; q) = \prod_{i=1}^d K(x_i, t_i) = \prod_{i=1}^d \sum_{j=0}^q P_j(x_i) P_j(t_i),$$

فرموله بندی می‌شود. لازم به ذکر است ورودی‌های مسئله نیز باید در بازه $[0, 1]$ مقیاس بندی شده باشد [۴].

۴ مدل پیشنهادی

در این پژوهش جهت تشخیص ابتلا به بیماری کووید-۱۹ الگوریتم ۱ را پیشنهاد می‌کنیم. این الگوریتم در فاز اولیه به کمک پارامترهای ذکر شده در جدول ۲ آموزش می‌بیند. برای این منظور داده‌های ورودی پس از انجام پیش‌پردازش و افزایش ویژگی نرمال شده و سپس به الگوریتم دسته‌بند گرادیان کاهشی داده می‌شوند. نرمال سازی ویژگی‌ها با حذف میانگین و تغییر مقیاس واریانس به یک انجام می‌شود. استفاده از تابع خطای hinge به همراه روش منظم سازی تیخونوف یا L2 در مدل دسته‌بند گرادیان کاهشی باعث می‌شود الگوریتم ارائه شده همانند فرم اولیه مسئله ماشین بردار پشتیبان عمل کند [۴]. تابع خطای hinge به صورت

$$L(y_i, f(x_i)) = \max(0, 1 - y_i f(x_i)), \quad (14)$$

تعریف می‌گردد. در اینجا $y_i = \pm 1$ برچسب صحیح نمونه x_i و $f(x_i)$ پیش‌بینی دسته‌بند گرادیان کاهشی برای این نمونه می‌باشد. در فاز آزمایش مدل، از پارامترهای بدست آمده در فاز آموزشی استفاده می‌کنیم. برای این منظور هر داده‌ی جدید پس از انجام پیش‌پردازش لازم و افزایش ویژگی به کمک تقریب نیستروم، به کمک پارامترهای میانگین و واریانس داده‌های آموزشی نرمال می‌شود. سپس با استفاده از وزن‌های بدست آمده توسط فاز اولیه الگوریتم گرادیان کاهشی، دسته‌بند پیش‌بینی خود را ارائه می‌کند. در بخش بعد با ذکر روش تنظیم ابر-پارامترهای مسئله و معیارهای بررسی دقت مدل، به مقایسه روش ارائه شده با دیگر روش‌های موجود می‌پردازیم.

۱.۴ نتایج

یافتن ابر-پارامترهای بهینه برای مدل به وسیله‌ی اعتبارسنجی متقابل 5-Fold انجام شده است. پس از انجام آزمایشات مختلف، پارامترهای مدل پیشنهادی مطابق با جدول ۲ بدست آمد. جدول ۴ زمان لازم برای محاسبه ماتریس هسته به روش ساده و همچنین روش تقریب هسته نیستروم به ازای مقادیر متفاوتی از تعداد کامپوننت‌ها (نمونه‌های مورد استفاده جهت تقریب) آورده شده است. مشاهده می‌شود با انتخاب زیرمجموعه کوچکی از داده‌های آموزشی، می‌توان ماتریس هسته را با دقت مناسب در زمان کوتاه تر تقریب زد. در جدول ۵ با انتخاب ۲۰٪ از داده‌ها به عنوان مجموعه داده‌های آزمایشی، معیارهای دقت، صحت، بازخوانی و F1 گزارش شده است. با در نظر گرفتن مثبت صادق (TP)، منفی صادق (TN)، مثبت کاذب (FP) و منفی کاذب (FN) این معیارها به صورت زیر تعریف می‌شوند:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}, \quad F1 = \frac{2 * (Recall * Precision)}{Recall + Precision} \quad (15)$$

در جدول ۳ دقت مدل ارائه شده با رویکردهای مختلف مقایسه شده است. مشاهده می‌شود مدل ارائه شده دقت بالایی در تشخیص افراد مبتلا به کووید-۱۹ در میان افراد مشکوک دارد. لازم به ذکر است تعیین وضعیت بیماری افراد در فاز تست بلادرنگ بوده و به کمتر از یک صد ثانیه زمان نیاز دارد.

مدل	نام پارامتر	مقدار پارامتر	هسته	پارامترها	دقت
نیستروم	تعداد کامپوننت‌ها	۵۰۰	خطی [۳]	-	۹۶/۸۰٪
دسته‌بند	منظم‌سازی	L2	چندجمله‌ای [۵]	-	۹۵/۴۸٪
دسته‌بند	ضریب منظم‌سازی	۱۰-۶	چندجمله‌ای نرمال شده [۵]	-	۹۵/۳۹٪
دسته‌بند	تابع خطا	hinge	تابع پایه شعاعی	$\gamma = ۰/۰۱$	۹۶/۴۰٪
الگوریتم گرادیان کاهشی	طول گام	تطبیقی	تابع پایه شعاعی	$\gamma = ۰/۱$	۹۷/۸۲٪
الگوریتم گرادیان کاهشی	حداکثر تکرار	۱۰۰۰	چندجمله‌ای لژاندر	$q = ۳$	۹۷/۷۰٪

جدول ۳: تاثیر پارامتر هسته در دقت مدل ماشین بردار پشتیبان

روش	تعداد کامپوننت‌ها	زمان (ثانیه)
محاسبه کل هسته	-	۲۱/۱
نیستروم	۲۰۰	۱/۶۵
نیستروم	۵۰۰	۴/۳۸

جدول ۲: پارامترهای مورد استفاده برای آموزش مدل

صحت	بازخوانی	F1	تعداد داده‌ها
غیرکووید-۱۹	۰/۹۳	۰/۹۷	۲۱۰
کووید-۱۹	۰/۹۹	۰/۹۸	۸۷۷
دقت	۰/۹۸	۰/۹۸	۱۰۸۷

جدول ۵: نتایج دسته‌بند پیشنهادی

جدول ۴: مقایسه سرعت اجرای محاسبه هسته و تقریب نیستروم

۵ نتیجه‌گیری

در این تحقیق، با ترکیب روش‌های جبرخطی و توابع متعامد، یک روش کارآمد جهت حل مسائل یادگیری ماشین ارائه شد. در این رویکرد به کمک تجزیه مقادیر منفرد یک تقریب از ماتریس هسته بدست آمد. سپس به کمک روش دسته‌بند گرادیان کاهشی یک الگوریتم جهت تشخیص وجود یا عدم وجود ویروس کووید-۱۹ در افراد مشکوک ارائه شد. بررسی دقت مدل بر روی یک مجموعه داده شناخته شده انجام شد و نتایج بدست آمده بیانگر دقت بالای مدل در تشخیص این ویروس بود.

مراجع

- [۱] ک. پرند، م. رزاقی، ج. امانی راد، م. دلخوش، م. معیری. رویکردهای نوین از روش‌های طیفی در محاسبات علمی: نظریه و کاربردها. دانشگاه شهید بهشتی؛ ۱۳۹۸.
- [۲] م. حجاریان. نخستین درس در جبر خطی عددی، انتشارات شهید بهشتی، تهران، ۱۳۹۷.
- [3] Aliyu S, Zakari AS, Adeyanju I, Ajoge NS. A Bayesian Network Model for the Prognosis of the Novel Coronavirus (COVID-19). In International Conference on Computational Science and Its Applications 2021 Sep 13; (pp. 127-140).
- [4] Parand K, Aghaei AA, Jani M, Ghodsi A. A new approach to the numerical solution of Fredholm integral equations using least squares-support vector regression. In Mathematics and Computers in Simulation 2021 Feb 1;180(pp. 114-28).
- [5] Villavicencio CN, Macrohon JJ, Inbaraj XA, Jeng JH, Hsieh JG. COVID-19 Prediction applying supervised machine learning algorithms with comparative analysis using WEKA. In Algorithms. 2021 Jul;14(7) No. 201.
- [6] Williams C, Seeger M. Using the Nystrom method to speed up kernel machines. In Proceedings of the 14th annual conference on neural information processing systems 2001 (pp. 682-688).