روش آنالیز جداکننده چندخطی با استفاده از ملاک تفاضلی

امین رستگار، منصور رزقی*؛ دانشگاه تربیت مدرس
دریافت ۹۴/۰۸/۱۷ پذیرش ۹۶/۰۳/۰۶

چکیده

روش‌های کاهش ابعاد نقش بسیار مهمی در بهبود کیفیت روش‌های یادگیری ماشین و تشخیص الگو ایفا می‌کنند. در این روش‌ها فرض می‌شود که داده‌ها به‌صورت یک بردار هستند. بنا براین داده‌های غیربرداری مانند تصاویر که به داده‌های با بعد بالا یا تانسور معروف هستند باید به بردار تبدیل شده و سپس فرایند کاهش ابعاد انجام گیرد. این عمل علاوه بر هزینهٔ محاسباتی ممکن است باعث از بین رفتن برخی روابط محلی بین دادها نیز شود. برای حل این مشکل در سال‌های اخیر روش‌های کاهش ابعاد چند خطی ارائه شده‌اند که سعی دارند فرایند کاهش ابعاد را بدون تغییر شکل داده‌ها انجام دهند. یکی از رویکردهای چندخطی مهم در این زمینه توسیع جداکننده خطی فیشر به‌روش‌های چندخطی است. مشکل مهم این روش‌ها نیاز آن‌ها به حل تعداد زیادی مسئله مقدار ویژه توسیع‌یافته در تکرارهای زیاد است. در این مقاله روشی برای حل این زیر مسئله‌ها ارائه کرده‌ایم. نتایج پیاده‌سازی کیفیت روش ارائه‌شده در جدا کردن کلاس‌ها را در مقایسه با روش اصلی جداکننده چندخطی به‌ویژه برای داده‌ها با نمونه کم نشان می‌دهند.

واژه‌های کلیدی: کاهش بعد، روش آنالیز مؤلفه‌های اصلی، آنالیز جداکننده خطی، روش چندخطی، آنالیز جدا کننده چندخطی

مقدمه

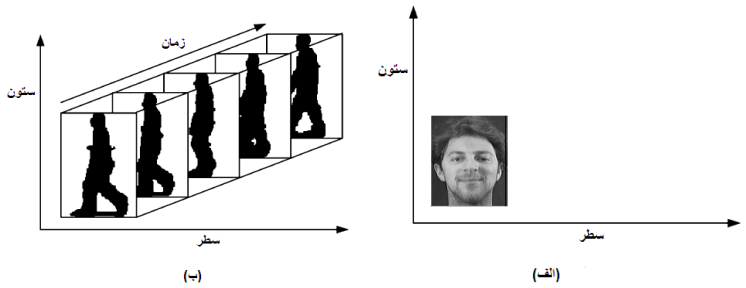
امروزه داده‌های زیادی تولید می‌شوند که با تجزیه و تحلیل آن‌ها اطلاعات مفیدی به‌دست می‌آید. اگر داده‌ها دارای حجم زیادی باشند، آنالیز و پردازش آن‌ها به‌سختی انجام می‌گیرد. به همین دلیل روش‌هایی وجود دارند که قبل از پردازش داده، حجم داده یا ویژگی‌های آن را کاهش می‌دهند و داده را به فضایی با ابعاد کوچک‌تر تصویر می‌کنند. به این روش‌ها که باعث سهولت تجزیه و تحلیل دادهٔ کاهش یافته و حذف اطلاعات تکراری یا غیرمرتبط می‌شود، اصطلاحاً روش‌های کاهش بعد می‌گویند [۱]. از جمله متداول‌ترین روش‌های کاهش بعد، روش‌های تحلیل مؤلفه‌های اصلی PCA[1]، روش جداکننده خطی LDA[2] و تجزیهٔ نامنفی ماتریسی را می‌توان نام برد [۲]، [۳]، [۴]. همگی این روش‌های هر داده را به‌صورت یک بردار فرض می‌کنند. ولی در عمل ممکن است داده‌ها دارای ساختاری غیربرداری باشند. داده‌هایی مانند تصاویر و ویدئو دارای ساختار دو یا چندوجهی هستند و به تانسور معروف هستند[۵].

برای کاهش بعد تانسورهای مرتبهٔ بالا با استفاده از روش‌های خطی متداول، باید ساختار داده تغییرکرده و به بردار تبدیل شود. در فرآیند برداری کردن، داده‌های تانسوری از فضای $\mathbb{R}^{m_1\times m_2\ldots\times m_N}$ تبدیل به بردارهایی در فضای $\mathbb{R}^{(m_1\times m_2\ldots\times m_N)}$ می‌شوند. سپس با استفاده از روش‌های کاهش بعد خطی، به فضای برداری $\mathbb{R}^n$ که $n<(m_1\times m_2\ldots\times m_N)$ تصویر می‌گردند. استفاده از روش‌های خطی برای کاهش بعد، داده‌های تانسوری دارای معایبی است که آن را در عمل غیرکاربردی می‌سازند. این معایب را بدین‌صورت می‌توان خلاصه کرد:

*نویسندهٔ مسئول rezghi@modares.ac.ir

1. Principle Component Analysis
2. Linear Discriminant Analysis

- یکی از مشکلات به کاربردن روش‌های خطی در داده‌های تانسوری، پیچیدگی محاسباتی زیاد آن است. برای مثال هر داده تانسور N بعدی عضو $\mathbb{R}^{m_1\times m_2\ldots\times m_N}$ به یک بردار $m_1 \times m_2 \ldots \times m_N$ تایی تبدیل می‌شود. این کار برای روش‌هایی مانند PCA و LDA که نیازمند حل مسئله مقدار ویژه یک ماتریس به ابعاد $(m_1 \times m_2 \ldots \times m_N) \times (m_1 \times m_2 \ldots \times m_N)$ برای این داده‌ها هستند بسیار از لحاظ محاسباتی پرهزینه است [۶].
- از طرف دیگر به‌دلیل زیاد بودن بعد (ویژگی‌های) هر داده با مسئلهٔ معضل ابعاد روبه‌رو هستیم. در مسئله معضل ابعاد با افزایش ویژگی‌های داده، تعداد نمونه‌های لازم برای داشتن یک روش یادگیری قابل اعتماد به‌صورت نمایی افزایش پیدا می‌کند [۶]، [۷]، [۸]، [۹]. در صورت کم بودن داده‌ها معضل ابعاد برای هر روش به‌صورت خاصی رخ می‌دهد. مثلاً برای روش LDA کم بودن داده‌های یادگیری منجر به بدحالتی ماتریس پراکندگی و بروز خطا در حل مسئله می‌شود.
- ایراد سوم استفاده از روش‌های خطی، تغییر ساختار داده‌ها است. زیرا با برداری کردن، جای‌گاه عناصر داده نسبت به‌هم تغییر می‌کند و پیوستگی و روابط محلی عناصر داده از بین می‌رود. مثلاً در داده‌های موجود در شکل ۱ برداری کردن تصویر الف باعث از بین بردن روابط مکانی پیکسل‌های تصویر می‌شود. هم‌چنین در داده ب که نشان‌دهندهٔ داده‌ای ویدویی است علاوه بر از بین بردن روابط همسایگی پیکسل‌ها که در تصویر مهم است، روابط زمانی آن‌ها را نیز به‌هم می‌ریزد.



شکل ۱. الف) تصویر چهره به‌صورت یک تانسور مرتبه ۲، ب) داده ویدویی به‌صورت یک تانسور مرتبهٔ ۳

در سال‌های اخیر برای پوشش ضعف‌های روش‌های خطی، روش‌هایی تحت نام روش‌های چندخطی که به‌روش‌های تانسوری نیز معروف هستند، برای یادگیری و کاهش ابعاد روی داده‌های تانسوری عرضه شده‌اند [۵]، [۶]، [۷]، [۹]،[۱۰]. در این روش‌ها، ساختار داده تغییر نمی‌کند و کاهش بعد به‌صورت مستقیم روی داده تانسوری انجام می‌گیرد. داده‌های تانسوری از فضای $\mathbb{R}^{m_1\times m_2\ldots\times m_N}$ به تانسورهایی در فضای $\mathbb{R}^{n_1\times n_2\ldots\times n_N}$ تصویر می‌شود که در آن برای هر $1\le i\le N$ رابطهٔ $n_i << m_i$ برقرار است. خلاصه مزیت‌های این روش‌ها را می‌توان بدین‌صورت خلاصه کرد:

- این روش‌ها روابط محلی بین داده‌ها را ازبین نمی‌برند. مثلاً در برداری کردن، دو پیکسل مجاور هم ممکن بود از هم دور شده و رابطهٔ محلی آن‌ها در نظر گرفته نشود. اما در روش‌های تانسوری ساختار داده دست‌نخورده باقی می‌ماند.
- در این روش‌ها معمولاً ویژگی‌های داده‌ها در بعدهای مختلف جدای از هم در نظر گرفته شده و تحلیل می‌شوند. بنا براین این امکان ایجاد می‌شود تا در یک بعد که قابلیت کاهش بیش‌تری وجود دارد کاهش بعد بیش‌تری

صورت گیرد. مثلاً در یک تصویر خاکستری که به‌صورت ماتریس است ممکن است مشاهده کنیم که سطرهای آن بیش‌تر از ستون‌ها قابلیت کاهش دارند.

- حل مدل‌های چندخطی معمولاً به حل چندین مسئله کوچک‌تر تبدیل می‌شود که این امر از حجم محاسبات و احتمال مواجه با مسئله معضل ابعاد می‌کاهد.

مهم‌ترین مشکل روش‌های چندخطی فقیر بودن ادبیات حل مسئله در این روی‌کرد است. به‌طوری‌که در سال‌های اخیر تنها روش‌های خیلی بدیهی خطی را توانسته‌اند به حالت چندخطی توسیع دهند. اگرچه این توسیع‌ها نیز موفق بوده‌اند.

با توجه به مزیت‌های بالا، روش‌های چندخطی[۳] در دهۀ اخیر مورد توجه قرار گرفته و کارهای متنوعی روی آن‌ها انجام گرفته است. به‌دلیل خاص بودن داده‌های دوبعدی که بیش‌تر تصویر هستند روش‌های دوخطی زیادی مانند روش تقریب تعمیم یافته رتبۀ پایین ، آنالیز مؤلفه‌های اصلی دوخطی و آنالیز جداکننده دوخطی به‌وجود آمدند که که بدون برداری کردن داده‌های ماتریسی مانند تصاویر دوبعدی، ابعاد سطری و ستونی آن‌ها را به فضایی با ابعاد کم‌تر کاهش می‌دهند [۱۱]، [۱۲]، [۱۳]، [۱۴]، [۱۵].

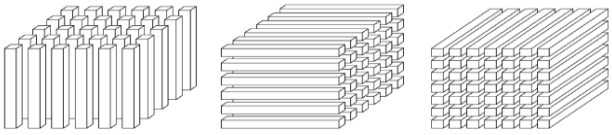
روش‌های چندخطی کاهش بعد مانند آنالیز مؤلفه‌های اصلی چندخطی و آنالیز جداکننده چندخطی نیز از روش‌های چندخطی مهم و پرکاربرد در سال‌های اخیر هستند [۵]، [۷]، [۹]،[۱۶]. این روش‌ها در کاربردهای بسیار زیادی استفاده شده‌اند [۱۸].

حل مدل‌های چندخطی مانند آنالیز جداکننده خطی و مؤلفه‌های اصلی چندخطی از یک فرایند تکراری تشکیل می‌شود که در هر مرحله از آن یک مسئله مقدار ویژه باید حل شود. بنا براین سرعت و دقت الگوریتم به‌کار گرفته شده در حل مسئله مقدار ویژه مهم است. در این مقاله ما از روشی تفاضلی برای حل این مسائل مقادیر ویژه به‌صورت فرایندی تکرای با تعداد تکرار کم استفاده می‌کنیم.

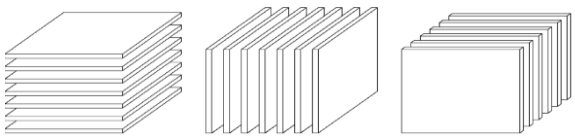
مفاهیم اولیه و مرور ادبیات

تانسور آرایه‌ای چندبعدی است که به تعداد بعدهای آن مرتبۀ تانسور گفته می‌شود. بردار تانسور از مرتبۀ یک و ماتریس آن مرتبۀ دوم است. به آرایه‌های با سه بعد یا بیش‌تر تانسورهای مرتبه بالا[۴] گفته می‌شود. برای نمایش یک بردار از حروف کوچک مانند a ، ماتریس‌ها از حروف بزرگ مانند A و تانسورهای مرتبه بالاتر از حروف اسکریپتی اویلر[۵] مانند $\mathcal{X}$ استفاده می‌شود. عنصر i ام برداری مانند a با a_i، عنصر (i,j) ماتریس A با $a_{i,j}$ و عنصر (i,j,k) تانسور مرتبۀ سه $\mathcal{X}$ با $x_{i,j,k}$ نشان داده می‌شود. مفهوم سطر و ستون را در تانسورها گسترش یافته و به فیبر معروف است. اگر یکی از اندیس‌ها متغیر و بقیه اندیس‌ها ثابت در نظر گرفته شود، یک فیبر[۶] به‌دست می‌آید. برای یک تانسور مرتبۀ سوم $\mathcal{X}$، می‌توان فیبرهای $x_{i,j,:}$ ، $x_{i,:,k}$ ، $x_{:,j,k}$ را تولید کرد. اسلایس[۷] برشی دوبعدی از تانسور است. با متغیر در نظر گرفتن دو اندیس و ثابت نگه‌داشتن بقیه اندیس‌ها، ماتریسی به‌دست می‌آید که به آن اسلایس گفته می‌شود. در شکل ۲ و شکل ۳ فیبرها و اسلایس‌های مختلف یک تانسور مرتبۀ سه نشان داده شده است.

3. Multilinear
4. higher-order tensor
5. Euler script
6. Fiber
7. Slice

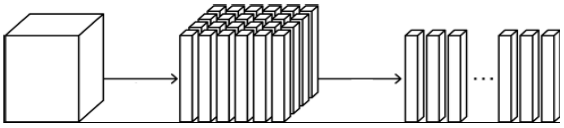


شکل ۲. فیبرهای مختلف یک تانسور مرتبهٔ سه



شکل ۳. اسلایس‌های مختلف یک تانسور مرتبهٔ سه

تانسورهای مرتبه بالا را می‌توان با عمل ماتریسی کردن به‌صورت یک ماتریس مرتب کرد. به‌صورت کلی ماتریسی کردن وجه-k تانسور $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N}$ از قرار دادن فیبرهای وجه-k آن در ستون‌های ماتریس به‌دست می‌آید و با $\mathcal{X}_{(k)}$ نمایش داده می‌شود. شکل ۴ نحوهٔ ماتریسی کردن یک تانسور از مرتبهٔ ۳ را در وجه اول آن نشان می‌دهد.



شکل ۴. فرایند ماتریسی کردن وجه- یک در تانسور مرتبهٔ سه

تانسورها را می‌توان از یک وجه در ماتریس ضرب کرد. در ضرب وجه-n تانسور $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N}$ در ماتریس $\mathrm{U} \in \mathbb{R}^{J \times I_n}$ که با $\mathcal{X} \times_n \mathrm{U}$ نمایش داده می‌شود، تانسوری با اندازهٔ $I_1 \times I_2 \ldots \times I_{n-1} \times J \times I_{n+1} \times I_N$ و هم‌مرتبه با تانسور $\mathcal{X}$ حاصل می‌گردد. عناصر تانسور حاصل‌ضرب از ضرب هر فیبر وجه-n تانسور $\mathcal{X}$ در ماتریس U با استفاده از رابطهٔ

$$(\mathcal{X} \times_n \mathrm{U})_{i_1, i_2 \ldots i_{n-1}, j, i_{n+1} \ldots, i_N} = \sum_{i_n=1}^{I_n} x_{i_1, i_2 \ldots, i_N}\, \mathrm{u}_{\mathrm{j i_n}}$$

محاسبه می‌شود. برای ضرب وجه-n تانسور در ماتریس روابط (۱) و (۲) برقرار است:

$$\mathcal{Y} = \mathcal{X} \times_n \mathrm{U} \iff \mathrm{Y}_{(n)} = \mathrm{U}\mathrm{X}_{(n)}, \qquad (۱)$$

$$\mathcal{Y} = \mathcal{X} \times_1 \mathrm{U}^{(1)} \times_2 \mathrm{U}^{(2)} \ldots \times_N \mathrm{U}^{(N)} \iff \mathrm{Y}_{(n)} = \mathrm{U}^{(n)}\, \mathrm{X}_{(n)} (\mathrm{U}^{(N)} \otimes \mathrm{U}^{(N-1)} \ldots \otimes \mathrm{U}^{(1)}),^{T} \qquad (۲)$$

که در این روابط $\otimes$ نشان‌دهندهٔ ضرب کرونکر ماتریسی است.

نرم تانسور $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N}$ مانند نرم فروبینیوس[8] در ماتریس از جذر مجموع توان‌دوم تمام عناصر تانسور $\mathcal{X}$ یعنی $\|\mathcal{X}\| = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \ldots \sum_{i_N=1}^{I_N} x_{i_1, i_2 \ldots, i_N}^2}$ به‌دست می‌آید.

۱. الگوریتم جداکننده چندخطی

روش جداکننده خطی LDA یکی از روش‌های معروف و پرکاربرد کاهش بعدخطی است که نسخهٔ چندخطی آن نیز ارائه شده است. در این بخش این روش و نسخه چندخطی آن را بررسی کرده و در بخش بعد روش خود برای بهبود نسخه چندخطی آن را ارائه می‌کنیم. روش LDA داده‌های برداری را به فضایی با ابعاد کم‌تر تصویر می‌کند به‌طوری‌که داده‌های تصویر شده دارای بیش‌ترین پراکندگی بین‌کلاسی و کم‌ترین پراکندگی درون‌کلاسی هستند. اگر

8. Frobenius

تعداد کلاس‌ها برابر K بوده و تعداد عناصر کلاس k ام را با N_k نشان دهیم ملاک جداشوندگی کلاس‌ها از هم‌دیگر و جمع‌شدگی درون‌کلاسی که به ملاک فیشر معروف است بدین‌صورت است:

$$J(W) = \frac{\mathrm{trace}(\mathrm{U^T S_B U})}{\mathrm{trace}(\mathrm{U^T S_W U})}. \quad (۳)$$

در اینجا S_B پراکندگی بین‌کلاسی و S_W پراکندگی درون‌کلاسی داده‌های اصلی بوده و بدین‌صورت تعریف می‌شوند:

$$S_w = \sum_{k=1}^{K} \sum_{x_n \in C_k} (x_n - m_k)\,(x_n - m_k)^T,$$

$$S_B = \sum_{k=1}^{K} N_k (m_k - m)(m_k - m)^T.$$

در اینجا m و m_k به‌ترتیب نشان‌دهندهٔ میانگین کل داده‌ها و میانگین کلاس k ام هستند. روش جداکننده خطی به‌دنبال یافتن ماتریس $U \in \mathbb{R}^{n\times l}$ است که تابع (۳) را ماکسیمم کند. در این‌صورت هر داده $x \in \mathbb{R}^n$ به‌صورت $y = U^T x$ به فضای با بعد l به‌صورت خطی منتقل می‌شود. نشان می‌دهیم که جواب ماکسیمم‌سازی (۳)، l تا بردار ویژه متناظر با بزرگ‌ترین مقادیر ویژه مسئلهٔ:

$$S_B u = \lambda S_w u,$$

است [۸]. این روش یکی از کاراترین روش‌های کاهش ابعاد خطی است که در بسیاری از کاربردها استفاده می‌شود [۱۸].

در سال‌های اخیر نسخهٔ چندخطی این روش نیز ارائه شده است [۵]. به‌کمک این ملاک، داده‌ها به زیرفضایی تانسوری تصویر می‌شوند که به‌صورت هم‌زمان پراکندگی بین‌کلاسی را ماکسیمم و پراکندگی درون‌کلاسی را مینیمم می‌کند. فرض کنید $\mathcal{X}_m \in \mathbb{R}^{I_1\times I_2 \ldots \times I_N}$ برای $m = 1,2 \ldots M$ تانسورهایی از مرتبهٔ N هستند که I_n اندازهٔ وجه n تانسور است. داده‌ها در K کلاس مختلف دسته‌بندی شده و N_k تعداد داده‌ها در کلاس k ام است. ملاک جداکنندگی تانسوری برای این داده‌ها بدین‌صورت تعریف می‌شود.

$$\tilde{\mathrm{U}}^{(n)}\big|_{n=1}^{N} = \arg_{\tilde{\mathrm{U}}^{(n)}|_{n=1}^{N}} max \frac{\sum_{k=1}^{K} N_k \|\bar{y}_k - \bar{y}\|_f^2}{\sum_{k=1}^{K}\sum_{i\in C_k} \|y_i - \bar{y}_k\|_f^2}, \quad (۴)$$

که در آن $y_i = \mathcal{X}_i \times_1 \tilde{\mathrm{U}}^{(1)^T} \times_2 \tilde{\mathrm{U}}^{(2)^T} \ldots \times_N \tilde{\mathrm{U}}^{(N)^T}$، داده تصویرشده i ام است. هم‌چنین

$$\bar{y} = \bar{\mathcal{X}} \times_1 \tilde{\mathrm{U}}^{(1)^T} \times_2 \tilde{\mathrm{U}}^{(2)^T} \ldots \times_N \tilde{\mathrm{U}}^{(N)^T},$$

و

$$\bar{y}_k = \bar{\mathcal{X}}_k \times_1 \tilde{\mathrm{U}}^{(1)^T} \times_2 \tilde{\mathrm{U}}^{(2)^T} \ldots \times_N \tilde{\mathrm{U}}^{(N)^T}.$$

به‌ترتیب تصویر میانگین کلاس k و میانگین کل داده‌های نمونه به زیر فضای تانسوری جدید هستند. این روش یکی از روش‌های معروف کاهش ابعاد چندخطی است که در کاربردهای مختلف در سال‌های اخیر به‌کار گرفته شده است[۵].

از آن‌جا که این یک مسئله بهینه‌سازی نامحدب است برای حل این مسئله از روش تناوبی استفاده می‌شود به این معنی که در هر مرحله تمامی ماتریس‌ها به‌غیر از یکی ثابت گرفته شده و مسئله بر حسب فقط یک مؤلفه حل می‌شود و این فرایند برای همهٔ مؤلفه‌ها تک به تک انجام می‌گیرد. این عمل تا یک تکرار مشخص یا برآورده شدن هم‌گرایی

ادامه پیدا می‌کند. از میزان تغییرات تابع هدف در تکرارها می‌توان به‌عنوان عامل توقف استفاده کرد. حال اگر عامل $\widetilde{\mathrm{U}}^{(\mathrm{n})}$ متغیر در نظر گرفته شده و بقیه ثابت باشند معادلهٔ (۴) با استفاده از روابط تانسوری (۱) و (۲) بر حسب $\widetilde{\mathrm{U}}^{(\mathrm{n})}$ به‌صورت (۵) نوشته می‌شود:

$$\widetilde{\mathrm{U}}^{(n)} = \mathrm{argmax}_{\widetilde{\mathrm{U}}^{(n)}} \frac{\sum_{k=1}^{K} n_k \|\bar{y}_k - \bar{y}\|_f^2}{\sum_{k=1}^{K}\sum_{i\in C_k}\|y_i - \bar{y}_k\|_f^2} = \mathrm{argmax}_{\widetilde{\mathrm{U}}^{(n)}} \frac{tr\left(\widetilde{\mathrm{U}}^{(n)^{\mathrm{T}}}.\mathrm{S_B}^{(n)}.\widetilde{\mathrm{U}}^{(n)}\right)}{tr\left(\widetilde{\mathrm{U}}^{(n)^{\mathrm{T}}}.\mathrm{S_W}^{(n)}.\widetilde{\mathrm{U}}^{(n)}\right)}, \quad (۵)$$

که در آن $\mathrm{S_B}^{(n)}$ و $\mathrm{S_W}^{(n)}$ بدین‌صورت محاسبه می‌شوند:

$$\mathrm{S_B}^{(n)} = \sum_{k=1}^{K} n_k \,.\,(\overline{\mathrm{X}}_{k(n)} - \overline{\mathrm{X}}_{(n)}).\,\widetilde{\mathrm{U}}_{\Phi^{(\mathrm{n})}}.\,\widetilde{\mathrm{U}}_{\Phi^{(\mathrm{n})}}^{\mathrm{T}}.\,(\mathrm{X}_{m(n)} - \overline{\mathrm{X}}_{(n)})^{\mathrm{T}},$$

$$\mathrm{S_W}^{(n)} = \sum_{k=1}^{K}\sum_{i\in C_k}(\mathrm{X}_{i(n)} - \overline{\mathrm{X}}_{k(n)}).\,\widetilde{\mathrm{U}}_{\Phi^{(\mathrm{n})}}.\,\widetilde{\mathrm{U}}_{\Phi^{(\mathrm{n})}}^{\mathrm{T}}.\,(\mathrm{X}_{i(n)} - \overline{\mathrm{X}}_{k(n)})^{\mathrm{T}}.$$

و $\widetilde{\mathrm{U}}_{\Phi^{(\mathrm{n})}}$ ضرب کرونکر زیر است

$$\widetilde{\mathrm{U}}_{\Phi^{(\mathrm{n})}} = \left(\widetilde{\mathrm{U}}^{(n+1)} \otimes \widetilde{\mathrm{U}}^{(n+2)} \otimes \ldots \otimes \widetilde{\mathrm{U}}^{(N)} \otimes \widetilde{\mathrm{U}}^{(1)} \otimes\ \widetilde{\mathrm{U}}^{(2)} \ldots \otimes \widetilde{\mathrm{U}}^{(n-1)}\right).$$

از روابط ماتریسی مشخص است که جواب مسئلهٔ ماکسیمم‌سازی (۵) یعنی $\tilde{U}_n \in \mathbb{R}^{I_n \times k_n}$ برابر با k_n تا بردار ویژه مناظر با k_n تا از بزرگ‌ترین مقادیر ویژه مسئله مقادیر ویژه تعمیمم یافتهٔ

$$\mathrm{S_W}^{(n)}\widetilde{\mathrm{U}}^{(n)^{\mathrm{T}}} = \mathrm{S_B}^{(n)}\widetilde{\mathrm{U}}^{(n)^{\mathrm{T}}}\Lambda,$$

است. از این رو، در هر تکرار از الگوریتم N تا مسئلهٔ مقدار ویژهٔ تعمیم یافته به بالا باید حل شوند. از این رو، روش حل این‌گونه مسائل سهم به‌سزایی در کیفیت و نیز سرعت روش‌ها دارد. در ادامه ما از روشی سریع و دقیق که مسئله مقدار ویژه تعمیم یافته را به‌صورت تفاضلی حل می‌کند استفاده می‌کنیم.

۲. الگوریتم جداکننده چندخطی با استفاده از ملاک تفاضلی

چنان‌که در بخش قبلی دیده شد در روش جداکننده چندخطی نیاز به حل مسئله ماکسیمم کردن تریس به‌صورت (۶) به دفعات زیاد است:

$$\max_{\mathrm{W}^T\mathrm{W}=\mathrm{I}} \frac{trace(\mathrm{W}^T\mathrm{AW})}{trace(\mathrm{W}^T\mathrm{BW})}, \quad (۶)$$

روش متداول برای ماکسیمم کردن نسبت تریس مذکور با استفاده از بردارهای ویژه حاصل از مسئلهٔ مقادیر ویژه تعمیم یافته است. استفاده از این روش در صورت بدحالت بودن ماتریس B کارامد نیست. به‌همین علت در این قسمت روشی جدید با استفاده از الگوریتم تکراری نیوتن شرح می‌دهیم که مشکل مذکور را نداشته و سریع‌تر به‌جواب می‌رسد [۱۷].

فرض کنید که $W_{opt} \in \mathbb{R}^{\mathrm{n}\times\mathrm{k}}$ جوابی بهینه برای مسئله ماکسیمم‌سازی نسبت تریس (۶) باشد. ρ^* به‌صورت (۷) تعریف می‌شود:

$$\rho^* = \frac{trace(\mathrm{W}_{opt}^T\mathrm{AW}_{opt})}{trace(\mathrm{W}_{opt}^T\mathrm{BW}_{opt})}, \quad (۷)$$

از آن‌جاکه ρ^* بیش‌ترین مقدار ممکن برای نسبت تریس است، بنا براین برای هر ماتریس متعامد $\mathrm{W} \in \mathbb{R}^{\mathrm{n}\times\mathrm{k}}$ رابطهٔ $\frac{trace(\mathrm{W}^{\mathrm{T}}\mathrm{AW})}{trace(\mathrm{W}^{\mathrm{T}}\mathrm{BW})} \leq \rho^*$ برقرار است و نتیجهٔ (۸) به‌دست می‌آید:

$$trace(\mathrm{W^T AW}) - \rho^{*}.trace(\mathrm{W^T BW}) \le 0, \quad (۸)$$

با توجه به رابطهٔ مذکور برای W_{opt} و ρ^{*} شرط (۹) وجود دارد:

$$\max_{\mathrm{W^T W}=I} trace(\mathrm{W^T}(\mathrm{A}-\rho^{*}.\mathrm{B})\mathrm{W}) = trace\left(\mathrm{W}_{opt}(\mathrm{A}-\rho^{*}.\mathrm{B})\mathrm{W}_{opt}\right) = 0. \quad (۹)$$

تابع $f(\rho)$ به‌صورت (۱۰) تعریف می‌شود:

$$f(\rho) = \max_{\mathrm{W}^T\mathrm{W}=I} trace(\mathrm{W}^T(\mathrm{A}-\rho.\mathrm{B})\mathrm{W}). \quad (۱۰)$$

با توجه به تابع f، ماتریس G به‌صورت (۱۱) تعریف می‌شود:

$$G(\rho) = \mathrm{A} - \rho\mathrm{B}. \quad (۱۱)$$

مقادیر ویژه G به‌ترتیب نزولی مرتب شده و به‌صورت (۱۲) برچسب می‌خورند:

$$\mu_1(\rho) \ge \mu_2(\rho) \ge \cdots \ge \mu_n(\rho). \quad (۱۲)$$

با توجه به ماتریس G و مقادیر ویژه آن، رابطهٔ (۱۳) برای تابع f برقرار است:

$$f(\rho) = \mu_1(\rho) + \mu_2(\rho) + \cdots + \mu_k(\rho). \quad (۱۳)$$

در ادامه با استفاده از تابع G، تابع f را می‌توان به‌صورت (۱۴) بازنویسی کرد:

$$f(\rho) = trace(\mathrm{W}(\rho)^T G(\rho)\mathrm{W}(\rho)). \quad (۱۴)$$

در این رابطه $W(\rho)$ از k بردار ویژه متناظر با k مقدار ویژه بزرگ‌تر $G(\rho)$ تشکیل شده است. در [۱۷] نشان داده‌اند که تابع f در (۱۴) دارای این خواص است:

الف) تابع f یک تابع غیرصعودی است.

ب) $f(\rho) = 0$ است اگر و فقط اگر $\rho = \rho^{*}$ باشد.

با توجه به این خواص واضح است که، ρ^{*} تنها ریشه تابع f است. بنا براین می‌توان با استفاده از روش تکراری نیوتن روی تابع f، ρ^{*} و W_{opt} را محاسبه کرد. برای استفاده از روش نیوتن نیاز به مشتق گرفتن از تابع f است مشتق تابع f بدین‌صورت به‌دست می‌آید:

$$f'(\rho) = -trace(\mathrm{W}(\rho)^{\mathrm{T}}\mathrm{BW}(\rho)).$$

با استفاده از مشتق به‌دست آمده، فرمول روش نیوتن بدین‌صورت می‌شود:

$$\rho_{new} = \rho - \frac{trace\left(\mathrm{W}(\rho)^T(\mathrm{A}-\rho.\mathrm{B})\mathrm{W}(\rho)\right)}{-trace\left(\mathrm{W}(\rho)^T\,\mathrm{B}\,\mathrm{W}(\rho)\right)} = \frac{trace\left(\mathrm{W}(\rho)^T\,\mathrm{A}\,\mathrm{W}(\rho)\right)}{trace\left(\mathrm{W}(\rho)^T\,\mathrm{B}\,\mathrm{W}(\rho)\right)}.$$

در الگوریتم (۱) الگوریتم تکراری نیوتن برای ماکسیمم کردن نسبت تریس با استفاده از ملاک تفاضلی آورده شده است. ملاک توقف می‌تواند بر اساس تفاوت مقدار هدف در دو تکرار متوالی باشد.

نشان داده شده است که این الگوریتم دقت و سرعت زیادی در حل مسئله مقدار ویژه تعمیم یافته دارد. حال از این روش در پیدا کردن بردارهای ویژه در هر تکرار از روش جداکننده چندخطی می‌توان استفاده کرد. با توجه به خواص این روش ، هم‌گرایی روش چندخطی جدید نیز نسبت به الگوریتم اصلی چندخطی سریعتر خواهد شد.

ورودی الگوریتم: ماتریس‌های A و B که ماتریس B نیمه‌معین مثبت است.

خروجی الگوریتم: ماتریس $W \in \mathbb{R}^{D\times M}$ که نسبت $\mathrm{Tr}(W^{T}AW)/\mathrm{Tr}(W^{T}BW)$ را ماکسیمم می‌کند.

ماتریس W با ماتریس واحد مقداردهی اولیه شود.

$$\rho = \mathrm{Tr}(W^{T}AW)/\mathrm{Tr}(W^{T}BW)$$

تا زمانی که هم‌گرایی (تفاوت مقدار هدف دو تکرار متوالی از یک حدی کم‌تر شود) اتفاق بیفتد کارهای زیر انجام گیرد.

ماتریس W برابر M بردار ویژه متناظر با M مقادیر ویژه بزرگتر ماتریس $A - \rho.B$ قرار داده شود.

$$\rho = \mathrm{Tr}(W^{T}AW)/\mathrm{Tr}(W^{T}BW)$$

انتهای حلقه

الگوریتم ۱. الگوریتم تکراری نیوتن برای ماکسیمم کردن نسبت $\frac{\mathbf{Tr(W^{T}AW)}}{\mathbf{Tr(W^{T}BW)}}$ با استفاده از ملاک تفاضلی

مراحل الگوریتم جدید روش جداکننده خطی با این ملاک تفاضلی در الگوریتم ۲ آورده شده است. نکته‌ای که باید به آن اشاره کرد این است که اگر چه در این مقاله روش پیشنهادی روی روش جداکننده چندخطی انجام گرفته است اما این روش روی هر روش چندخطی دیگری مانند روش تحلیل مؤلفه‌های اصلی چندخطی که نیاز به محاسبه مقادیر ویژه دارند، نیز به‌کار گرفته می‌شود.

ورودی الگوریتم: تانسورهای نمونه $\{\mathcal{X}_m \in \mathbb{R}^{I_1\times I_2\ldots\times I_N}\}_{m=1}^{M}$ که کلاس آن‌ها مشخص شده و $\{P_n\}_{n=1}^{N}$ ابعاد داده‌های تصویر

خروجی الگوریتم: ماتریس‌های تصویر $\left\{\widetilde{U}^{(n)} \in \mathbb{R}^{\boldsymbol{I_n}\times\boldsymbol{P_n}}\right\}_{n=1}^{N}$ که در داده‌های تصویر شده پراکندگی بین‌کلاسی ماکسیمم و پراکندگی درون کلاسی مینیمم شود.

ماتریس‌های $\left\{\widetilde{U}_0^{(n)} \in \mathbb{R}^{\boldsymbol{I_n}\times\boldsymbol{P_n}}, n = 1,2\ldots N\right\}$ باماتریس‌های واحد مقداردهی اولیه شود.

$\overline{\boldsymbol{X}}$ و $\overline{\boldsymbol{X}}_{\boldsymbol{k}}$ که میانگین کل و میانگین کلاس $\boldsymbol{k}$ است برای $k = 1,2\ldots K$ به‌دست آورده شود.

برای $t = 1,2\ldots T_{max}$

برای $n = 1,2\ldots N$

$$\widetilde{U}_{\boldsymbol{\Phi}^{(n)}} = \left(\widetilde{U}_{t-1}^{(n+1)} \otimes \widetilde{U}_{t-1}^{(n+2)} \otimes \ldots \otimes \widetilde{U}_{t-1}^{(N)} \otimes \widetilde{U}_{t}^{(1)} \otimes \widetilde{U}_{t}^{(2)} \otimes \ldots \widetilde{U}_{t}^{(n-1)}\right),$$

$$S_B{}^{(n)} = \sum_{k=1}^{K} n_k \,.\, (\overline{\mathbf{X}}_{\boldsymbol{k(n)}} - \overline{\mathbf{X}}_{\boldsymbol{(n)}}).\, \widetilde{U}_{\boldsymbol{\Phi}^{(n)}}.\, \widetilde{U}_{\boldsymbol{\Phi}^{(n)}}^{T}.\, (\mathbf{X}_{\boldsymbol{m(n)}} - \overline{\mathbf{X}}_{\boldsymbol{(n)}})^{T},$$

$$S_W{}^{(n)} = \sum_{k=1}^{K}\sum_{i\in C_k}(\mathbf{X}_{\boldsymbol{i(n)}} - \overline{\mathbf{X}}_{\boldsymbol{k(n)}}).\, \widetilde{U}_{\boldsymbol{\Phi}^{(n)}}.\, \widetilde{U}_{\boldsymbol{\Phi}^{(n)}}^{T}.\, (\mathbf{X}_{\boldsymbol{i(n)}} - \overline{\mathbf{X}}_{\boldsymbol{k(n)}})^{T}.$$

ماتریس $\widetilde{U}_t^{(n)}$ با ماتریس واحد مقداردهی اولیه شود.

$$\rho = \mathrm{Tr}\left(\widetilde{U}_t^{(n)^{T}}.\, S_B{}^{(n)}.\, \widetilde{U}_t^{(n)}\right)/\mathrm{Tr}\left(\widetilde{U}_t^{(n)^{T}}.\, S_W{}^{(n)}.\, \widetilde{U}_t^{(n)}\right)$$

تا زمانی که $\left\|\widetilde{U}_t^{(n)} - \widetilde{U}_{t-1}^{(n)}\right\| > \varepsilon$ کارهای زیر انجام گیرد.

ماتریس $\widetilde{U}_t^{(n)}$ برابر $\boldsymbol{P_n}$ بردار ویژه متناظر با $\boldsymbol{P_n}$ مقادیر ویژه بزرگتر ماتریس $S_B{}^{(n)} - \rho.S_W{}^{(n)}$ قرار داده شود.

$$\rho = \mathrm{Tr}\left(\widetilde{U}_t^{(n)^{T}}.\, S_B{}^{(n)}.\, \widetilde{U}_t^{(n)}\right)/\mathrm{Tr}\left(\widetilde{U}_t^{(n)^{T}}.\, S_W{}^{(n)}.\, \widetilde{U}_t^{(n)}\right)$$

اگر $\left\|\widetilde{U}_t^{(j)} - \widetilde{U}_{t-1}^{(j)}\right\| < \varepsilon$ برای $j = 1,2\ldots N$ از حلقه خارج شود.

الگوریتم ۲. الگوریتم جداکننده چندخطی با استفاده از ملاک تفاضلی

پیاده‌سازی

در این قسمت برای مقایسهٔ روش جداکننده چندخطی و روش پیشنهادی از دو مجموعهٔ داده بسیار معروف YALE و MINIST استفاده شده است. داده‌ها به دو دستهٔ داده‌های نمونه و داده‌های آزمایش تقسیم شده و با کمک الگوریتم‌های گفته شده کاهش بعد پیدا می‌کنند. داده‌های آزمایش پس از کاهش بعد با استفاده از روش کلاس‌بندی نزدیک‌ترین همسایه[۹]، کلاس‌بندی می‌شوند. در پایان نتایج به‌دست آمده از نظر درصد کلاس‌بندی درست داده‌های آزمایش ارزیابی می‌شوند.

در اینجا روش جداکننده خطی را با نماد LDA، روش جداکننده چندخطی را با نماد DATER[۱۰] نمایش می‌دهیم. قابل ذکر است که در مقالات دیگر از نمادهای دیگری برای این روش نیز استفاده شده است [۵]. هم‌چنین روش پیشنهادی نیز با نماد DATER(new) نمایش داده شده است. در اینجا برای پیش پردازش داده‌ها از PCA (برای روش خطی) و از روش HOSVD[۱۱] (برای روش چندخطی) استفاده شده است. برای شروع داده YALE را در نظر می‌گیریم. شکل ۵ تعدادی از داده‌های این مجموعه داده را نشان می‌دهد.

شکل ۵. نمونه ای از دادهای مجموعه داده YALE

مجموعهٔ داده YALE، شامل ۱۶۵ تصویر سیاه وسفید چهره ۱۵ فرد است که در جهات مختلف با شدت نور متفاوت گرفته شده است. تصاویر به ابعاد ۳۲×۳۲ برش داده شده و بعد از نرمال شدن استفاده می‌شود. به‌صورت تصادفی تعداد ۲، ۳، ۶، ۸ داده از هر فرد به‌صورت نمونه اختیار شده و از بقیه داده‌ها به‌منزلهٔ داده‌های آزمایش استفاده می‌شود. ابتدا روش چندخطی را با روش خطی مقایسه می‌کنیم. نتایج در جدول ۱ قابل مشاهده است. در این‌جا مشاهده می‌شود که برای تعداد نمونه کم، روش LDA نتایج خوبی به‌نسبت روش چندخطی ندارد. اما با افزایش تعداد نمونه‌ها چون از بدوضعی ماتریس پراکندگی کاسته می‌شود، شاهد بهبود کیفیت این روش هستیم. اما حساسیت روش چندخطی کم‌تر است. جدول ۱ نشان می‌دهد که با داده آموزشی کم، عدد حالت ماتریس پراکندگی روش خطی بسیار زیاد است اما با افزایش داده‌های آموزشی بهبود می‌یابد. در روش چندخطی همیشه عدد حالت ماتریس پراکندگی در حد معقولی باقی می‌ماند.

حال نتایج روش چندخطی را با روش پیشنهادی مقایسه می‌کنیم. در این‌جا از دو ملاک درصد پیش‌بینی درست کلاس‌ها و نیز نسبت فیشر استفاده می‌کنیم. در این‌جا نسبت فیشر میزان پراکندگی بین‌کلاسی به درون‌کلاسی را نشان می‌دهد و هر چه بزرگ‌تر باشد بهتر است. این ملاک برای ارزیابی بسیاری از روش‌های کاهش ابعاد استفاده می‌شود [۱۹]. این ملاک از روش درصد پیش‌بینی درست کلاس‌ها ملاک بهتری است چون به کیفیت روش کلاس‌بندی وابستگی ندارد. جدول ۲ نشان می‌دهد که با نمونه‌های کم روش پیشنهادی از نظر ملاک درصد پیش بینی نتایج بهتری نسبت به‌روش اصلی را دارد اما از نظر ملاک فیشر روش پیشنهادی همیشه از روش چندخطی اصلی بهتر عمل می‌کند. هم‌چنین از لحاظ زمان مصرفی نیز عموما بهتر عمل می‌کند.

9. Nearest neighbor
10. Discriminant Analysis with Tensor Representation
11. Higher Order singular Value Decomposition

جدول ۱. مقایسهٔ نتایج روش خطی LDA و چندخطی DATER برای مجموعه داده YALE

روش	۲ نمونه در هر کلاس		۳ نمونه در هر کلاس		۶ نمونه در هر کلاس		۸ نمونه در هر کلاس	
	درصد کلاس‌بندی	عدد حالت ماتریس پراکندگی	درصد کلاس‌بندی	عدد حالت ماتریس پراکندگی	درصد کلاس‌بندی	عدد حالت ماتریس پراکندگی	درصد کلاس‌بندی	عدد حالت ماتریس پراکندگی
LDA	۴۳/۲۵	$۰/۱۴\times۱۰^{۲۰}$	۴۵/۹۶	۸۳۳۳۳	۷۵/۲	۳۴۵	۸۱/۹	۱۵۹
DATER	۵۱/۷	۲۹/۸	۶۳/۷۵	۲۳/۶	۷۶/۸	۲۲/۹۸	۸۰/۵۵	۲۰

جدول ۲. مقایسه نتایج روش‌های DATER روش پیشنهادی (new) DATER روی مجموعه داده YALE

روش	۲ نمونه در هر کلاس			۳ نمونه در هر کلاس			۶ نمونه در هر کلاس			۸ نمونه در هر کلاس		
	درصد کلاس‌بندی	نسبت فیشر	زمان	درصد کلاس‌بندی	نسبت فیشر	زمان	درصد کلاس‌بندی	نسبت فیشر	زمان	درصد کلاس‌بندی	نسبت فیشر	زمان
DATER	۵۱/۷	۳/۱۲	۷/۲۰	۶۳/۷۵	۲/۳۵	۱۲/۰۸	۷۶/۸	۱/۵۴	/۲۳ ۳۸	۸۰/۵۵	۱/۳۸	۵۷/۳۰
DATER(new)	۵۶/۵۹	۵/۴۷	۶/۲۷	۶۴/۰۲	۳/۳۰	۱۱/۰۵	۷۴/۶۶	۲/۳۸	/۵۰ ۳۶	۷۸	۲/۰۳	۵۴/۲۴

به عنوان مثال دوم مجموعه MINIST را شامل ۶۰۰۰۰ از دست خط مختلف برای ارقام ۰ تا ۹ با ابعاد ۲۸×۲۸ است را در نظر می‌گیریم. داده‌های نمونه از هر کلاس به تعداد ۲، ۳، ۶، ۱۲ به‌صورت تصادفی انتخاب می‌شوند و تعداد ۸۰ داده به‌عنوان داده‌های آزمایش استفاده می‌شوند شکل ۶ تعدادی از داده‌های این مجموعه داده را نشان می‌دهد.

شکل ۶. نمونه‌ای از داده‌های مجموعه MINIST

جدول ۳: مقایسه نتایج روش خطی LDA و چند خطی DATER برای مجموعه داده MINIST

روش	۲ نمونه در هر کلاس		۳ نمونه در هر کلاس		۶ نمونه در هر کلاس		۸ نمونه در هر کلاس	
	درصد کلاس بندی	عدد حالت ماتریس پراکندگی	درصد کلاس بندی	عدد حالت ماتریس پراکندگی	درصد کلاس بندی	عدد حالت ماتریس پراکندگی	درصد کلاس بندی	عدد حالت ماتریس پراکندگی
LDA	۳۷/۵	0.384×10^{18}	۴۵/۸۳	۴۵۵	۵۰	۵۲	۶۶/۰۷	۲۸/۳
DATER	۴۹	۳۶	۵۰/۱	۲۱	۶۱/۸۷	۱۴/۱	۷۱/۵	۱۰/۴

در این مثال نیز رفتار روش‌ها شبیه به مثال اول است. هم‌چنین جدول ۳ نشان می‌دهد که از نظر درصد درست تشخیص کلاس، روش پیشنهادی تنها در داده‌های کم بهتر از روش اصلی عمل می‌کند. اما از لحاظ نسبت فیشر عملکرد بسیار بهتری از روش اصلی را در همهٔ حالت‌ها دارد.

جدول ۴: مقایسهٔ نتایج روش‌های DATER روش پیشنهادی (new) DATER روی مجموعه داده MINIST

روش	۲ نمونه در هر کلاس			۳ نمونه در هر کلاس			۶ نمونه در هر کلاس			۸ نمونه در هر کلاس		
	درصد کلاس بندی	نسبت فیشر	زمان	درصد کلاس بندی	نسبت فیشر	زمان	درصد کلاس بندی	نسبت فیشر	زمان	درصد کلاس بندی	نسبت فیشر	زمان
DATER	۴۹	۱/۶۴	۲/۸۵	۵۰/۱	۱/۱۲	۶/۳	۶۱/۸۷	۱/۰۱	۲۴/۰۱	۷۱/۵	۰/۶۷	۷۱/۰۷
DATER(new)	۵۱/۲	۲/۷۲	۲/۴۷	۵۴/۳۷	۱/۹۱	۴/۴	۶۱/۰۱	۱/۳۱	۱۸/۲۵	۶۸/۳۰	۰/۷۱	۷۱/۱۵

نتایج پیاده‌سازی سه الگوریتم LDA، DATER، DATER(new) نشان می‌دهد که در صورت کم بودن تعداد نمونه‌ها در کلاس، روی کرد تفاضلی استفاده شده در الگوریتم DATER(new) بهترین جواب را از لحاظ تشخیص

درست کلاس‌ها در بین این سه الگوریتم می‌دهد و با افزایش تعداد نمونه‌ها در هر کلاس به‌علت خوش‌حالت‌شدن ماتریس‌های پراکندگی درون‌کلاسی، نتایج سه الگوریتم به‌هم نزدیک می‌شود. از طرف دیگر چون کیفیت درصد درست تشخیص کلاس‌ها به کیفیت روش کلاس‌بندی نیز وابسته است، ما از یک روش ارزیابی دیگر که ربطی به کلاس‌بندی ندارد استفاده کردیم. این ملاک نسبت فیشر است. نتایج با این ملاک نشان می‌دهد که روش پیشنهادی در همه حالات بهتر از روش اصلی کار می‌کند. هم‌چنین زمان روش پیشنهادی نیز از روش اصلی کم‌تر است.

نتیجه‌گیری

در این مقاله یک روش چندخطی از روش‌های جدا کننده چندخطی داده شد. این روش بر مبنای بهبود حل مسئله ماکسیمم‌سازی تریس ساخته شده است. نتایج عددی برتری کیفیت این روش به‌ویژه برای داده‌ها با داده‌های آموزشی کم در مقایسه با دیگر روش‌ها را نشان می‌دهد. هم‌چنین این روش به‌دلیل جداسازی کلاس‌ها و نیز فشرده‌سازی داده‌های درون‌کلاسی که با معیار فیشر اندازه‌گیری می‌شود از روش اصلی بهتر عمل می‌کند. اگرچه ما این روش را روی جداکننده چندخطی اعمال کردیم اما این روش با هر روش کاهش ابعاد چندخطی مانند مؤلفه‌های اصلی چندخطی که به مسئلهٔ مقدار ویژه ختم می‌شوند به‌کار برده می‌شود.

منابع

1. Rezghi M., Abulkasim A., "Noise-free principal component analysis: An efficient dimension reduction technique for high dimensional molecular data, Expert systems with Applications, 41(2014) 7797-7804.
2. Murphy K. P., "Machine Learning: A Probabilistic Perspective", MIT Press (2012).
3. Duda R. O., Hart P. E., Stork D. H., "Pattern Classification (2nd ed.)", Wiley (2006).
4. Binesh N., Rezghi M., "Fuzzy clustering in community detection based on nonnegative matrix factorization with two novel evaluation criteria", Applied Soft Computing, Accepted (2017).
5. Li Q., Schonfeld D., "Multilinear Discriminant Analysis for Higher-order Tensor data Classification", IEEE Transactions on Pattern recognition and Machine Learning, 36 (2014) 2524-2537.
6. Wang J., Barreto A., Wang L., chen Y., Rishe N., Andrianm J., Adjouadi M., "Multilinear principal component analysis for face recognition with fewer features", Neurocomputing, 73, (2010) 1550-1555.
7. Kong S., Wang D., "A Report on Multilinear PCA Plus Multilinear LDA to Deal with Tensorial Data: Visual Classification as An Example", Arxiv (2012).
8. Bishop C., "Pattern Recognition and Machine Learning, Springer (2006).

9. Yan S., Xu D., Yang Q., Zhang L., Tang X., Zhang H. J., "Discriminant analysis with tensor representation, in Proc", IEEE Conference on Computer Vision and Pattern Recognition, vol. I (2005) 526-532.

10. Hao N., kilmer M., Braman K., Hoover R. C., "Facial Recognition Using Tensor-Tensor Decompositions", SIAM Imaging, 6 (2013) 437-463.

11. Yang J., Zhang D., Frangi A. F., Yang J., "Two-dimensional PCA: a new approach to appearance-based face representation and recognition", IEEE Trans- actions on Pattern Analysis and Machine Intelligence (2004) 131-137.

12. Ye J., "Generalized low rank approximations of matrices, Machine Learning,vol. 61, no. 1-3 (2005) 167-191.

13. Ye J., Janardan R., Li Q., "Two-dimensional linear discriminant analysis", in Advances in Neural Information Processing Systems (NIPS) (2004)1569-1576.

14 . Ren C. X., Dai D. Q., "Bilinear Lanczos components for fast dimensionality reduction and feature extraction", Pattern recognition, 43 (2010) 3742-3752.

15. Wu G., Xu W., Leng H., "Inexact and incremental bilinear Lanczos components algorithms for high dimensionality reduction and image reconstruction", Pattern Recognition, 48 (2015) 244-263.

16. Lu H., Plataniotis K. N., Venetsanopoulos A. N., "Multilinear principal component analysis of tensor objects for recognition", in Proc. Int. Conf. on Pattern Recognition (August 2006) 776-779.

17. Ngo T. T., Bellalij M., Saad Y., "The trace ratio optimization problem for dimensionality Reduction", SIAM J. Matrix Anal. and Appl (2010) 2950-2971.

18. Laohakiat S., Phimoltares S., Lurisinsap C., "A clustering algorithm for stream data with LDA-based unsupervised localized dimension reduction, Information science", 381 (2017) 104-123.

19. Raducanu B., Dornaika F., "A supervised non-linear dimensionality reduction approach for manifold learning", Pattern recognition, 45 (2012) 2432-2444.