فرهنگ و اندیشهٔ ریاضی
شماره ۵۹ (پاییز و زمستان ۱۳۹۵) صص. ۷۱ تا ۹۰

تجزیهٔ نامنفی ماتریسی: روشی برای تحلیل داده‌های نامنفی

مهسا یوسفی و منصور رزقی

چکیده

امروزه پیشرفت فناوریِ رایانه‌ای منجر به افزایش حجم داده‌ها و نیز به‌وجود آمدن پایگاه‌های بزرگ داده‌ها شده است. در نتیجه روش‌های مختلفی برای کشف دانش از آنها، معرفی شده است و یا در حال معرفی هستند. در این راستا، یکی از شاخه‌های نسبتاً جدید علمی موسوم به *داده‌کاوی* مورد توجه زیادی قرار گرفته است. ماتریس‌های داده در کاربردهای داده‌کاوی، غالباً نامنفی هستند که این ویژگی محدودیت‌هایی را در استفاده از روش‌های ماتریسی کلاسیک به همراه دارد. اگرچه به‌کارگیری این روش‌ها سبب کاهش بُعد داده‌های بزرگ می‌شود، اما تعبیری صحیح از داده‌های نامنفی از آنها به‌دست نمی‌آید. اخیراً روش جدیدی با نام تجزیهٔ نامنفی ماتریسی برای نمایش خطی داده‌های نامنفی پیشنهاد شده است که علاوه بر کاهشِ بُعد داده‌ها، محدودیت روش‌های کلاسیک را ندارد. در این روش، ماتریس بزرگِ متناظر با داده‌های نامنفی به دو ماتریس نامنفی کوچک تجزیه می‌شود. در این مقاله، ابتدا روش‌های کلاسیک را مرور می‌کنیم. سپس تجزیهٔ نامنفی ماتریسی با نسخه‌های مختلف آن معرفی و مسائل مهم داده‌کاوی مانند رده‌بندی و خوشه‌بندی برای این روش بررسی می‌شود.

۱. مقدمه

رشد روش‌های تولید داده‌ها بسیار سریع‌تر از توانایی ما در درک و استفاده از آنها است. برای درک این سرعت رشد، کافی است برای چند لحظه به تعداد افراد آنلاینی فکر کنیم که تجربیاتشان را با استفاده از متن، فیلم، عکس و ... با یکدیگر به اشتراک می‌گذارند. آن‌وقت می‌توان به سرعت تولید داده‌ها و افزایش

عبارات و کلمات کلیدی. تجزیهٔ نامنفی ماتریسی؛ کاهش بُعد؛ کمترین مربعات متناوب با قید نامنفی؛ خوشه‌بندی؛ رده‌بندی.

حجم آنها در هر ثانیه پی بُرد. اما داده‌ها فقط زمانی مفید هستند که مورد پردازش قرار گرفته باشند. در واقع، برای ما مهم این است که از حجم بسیار بالای داده‌های خام اولیه، به دانش پنهان شده و الگوی مفید در آنها دست یابیم. این الگو باید ساده، معتبر و قابل فهم برای توصیف ارتباط میان داده‌ها باشد. در این راستا، شاخهٔ نسبتاً جدید علمی موسوم به داده‌کاوی[۱] در حال گسترش است که می‌توان انگیزهٔ شکل‌گیری آن را انجام تحقیقات در زمینه‌هایی نظیر آمار، یادگیری ماشینی و مدیریت پایگاه داده‌ها دانست. با داده‌کاوی می‌توان اطلاعات باارزشی را از مجموعهٔ داده‌های بزرگ استخراج و بر اساس آنها تصمیمات مهمی اتخاذ کرد. در یک تقسیم‌بندی کلی می‌توان یادگیری ماشینی را به دو دستهٔ اصلی: یادگیری بدون نظارت[۲] مانند خوشه‌بندی[۳] و یادگیری با نظارت[۴] مانند رده‌بندی[۵] تقسیم کرد. در داده‌کاوی، از خوشه‌بندی و رده‌بندی به‌سان ابزارهایی جهت توصیف داده‌ها و مدل کردنِ پیش‌بینی آنها استفاده می‌شود. اساس مشترکِ رهیافت‌های مختلف در تحلیل داده‌ها، پیدا کردن یک مدل مناسب برای نمایش آنها است تا بینش و تصوری صحیح از آنها به‌وجود آید. هدف ما، توصیفِ داده‌کاوی به‌عنوان شاخهٔ جدید علمی نیست، بلکه بر آن هستیم که به اهمیت و کاربرد مدل(های) مناسب در آن بپردازیم. رهیافت بسیار معمولی که آن را مدل کاهشی[۶] می‌نامیم، مدلی است که سعی می‌کند از پیچیدگی داده‌های اولیه و با بُعد بالا بکاهد و ضمن حفظ ملزومات مسئله همراه با داده‌ها، ساختار پنهان آنها را آشکار کند. در واقع، این مدل ضمن کاهش داده‌ها، در سطحی نزدیک‌تر (واقعی‌تر) به سیستم اولیه قرار دارد. مدل‌های کاهشی می‌توانند خطی و یا غیرخطی باشند که در این مقاله، مدل‌های خطی را مطالعه می‌کنیم. در این مقاله، سه مدل کاهشی خطی به نام‌های تجزیهٔ مقدار تکین (SVD)[۷]، تحلیل مؤلفهٔ اصلی (PCA)[۸] و تجزیهٔ نامنفی ماتریسی (NMF)[۹] بیان می‌شوند و روی مدل اخیر بیشتر تمرکز خواهیم کرد. مدل NMF برای تقریب داده‌های نامنفی ذخیره‌شده در ماتریس نامنفی $A \in \mathbb{R}^{m\times n}$، به‌دنبال تولید دو ماتریس نامنفی دیگر مانند $W \in \mathbb{R}^{m\times r}$ و $H \in \mathbb{R}^{r\times n}$ با شرط $r \ll \min\{m,n\}$ است به‌طوری که معادلهٔ تقریبی $A \approx WH$ برقرار شود. در این معادلهٔ تقریبی، رتبهٔ ماتریس WH از رتبهٔ ماتریس A کمتر است. در بخش دوم، تجزیهٔ ماتریسی را توضیح خواهیم داد. روش‌های کلاسیک مشهور در تجزیهٔ ماتریسی را که در داده‌کاوی مورد استفاده قرار می‌گیرند، مرور می‌کنیم و محدودیت آنها را در توصیف داده‌ها نشان می‌دهیم. سپس روش جدید NMF را معرفی و نسخه‌های مختلف موجود از آن را به‌تفصیل بررسی می‌کنیم. بخش سوم، به دو کاربردِ منتخب از روش NMF برای مسائل خوشه‌بندی و رده‌بندی به همراه نتایج عددی اختصاص دارد. در بخش پایانی، به جمع‌بندی و نتیجه‌گیری مطالب این مقاله خواهیم پرداخت.

---

[۱] data mining [۲] unsupervised learning [۳] clustering [۴] supervised learning [۵] classification [۶] reduction model [۷] singular value decomposition [۸] principle component analysis [۹] nonnegative matrix factorization

۲. تجزیهٔ ماتریسی

به‌طورکلی، تجزیهٔ ماتریسی ابزاری برای تحلیل داده‌ها است. هر تجزیه تعبیرهای مختلفی را از ساختار ضمنی داده‌ها آشکار می‌سازد که البته این تعبیرها از نظر ریاضی هم‌ارز هستند. برخی از این تعبیرها در قالب مثال‌هایی در قسمت ۱.۳.۲ توضیح داده می‌شوند. یک پایگاه داده را می‌توان ماتریسی با m سطر و n ستون در نظر گرفت. برای مثال، در ماتریس متناظر با یک پایگاه دادهٔ تصویری، هر ستون می‌تواند معرف یک تصویر باشد (شکل ۱ را ببینید). تجزیهٔ ماتریسی بازنمایش ویژه و ساده‌ای از ماتریس دادهٔ اولیه از طریق تولید ماتریس‌های جدید است. لذا به‌کارگیری آن، به این دلیل که داده‌ها اغلب پیچیده هستند و منجر به ناکارآمدی روش‌های موجود در داده‌کاوی می‌شوند، معقول و مناسب خواهد بود. حتی می‌توان از آن برای پاکسازی داده‌ای[۱] در شرایطی که داده‌ها آلوده به نوفه باشند نیز استفاده کرد. با مطالعهٔ مرجع [۲۰] می‌توان جزئیات بیشتری در این باره به‌دست آورد. تجزیهٔ مقدار تکین، تحلیل مؤلفهٔ اصلی و تجزیهٔ نامنفی ماتریسی از روش‌های شناخته‌شده در مبحث تجزیهٔ ماتریسی هستند. در این مقاله، سعی بر این است که چگونگی محاسبه و آشکارسازیِ ساختار ضمنیِ ماتریس داده توسط NMF به‌تفصیل ارائه گردد. اما قبل از آن، دو تجزیهٔ دیگر را مرور می‌کنیم.

۱.۲. تجزیهٔ مقدار تکین. هر ماتریس $A \in \mathbb{R}^{m\times n}$ با $m \geq n$ را می‌توان به‌صورت

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T \tag{۱.۲}$$

تجزیه کرد که در آن، $\Sigma \in \mathbb{R}^{n\times n}$ ماتریس قطری با نمایشِ

$$\Sigma = \mathrm{diag}(\sigma_1, ..., \sigma_n), \qquad \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$$

است و ماتریس‌های $U \in \mathbb{R}^{m\times m}$ و $V \in \mathbb{R}^{n\times n}$ متعامد هستند. رابطهٔ (۱.۲) را تجزیهٔ مقدار تکین (SVD) ماتریس A می‌نامیم. این تجزیه را می‌توان به‌صورت بسط زیر نیز نمایش داد:

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T. \tag{۲.۲}$$

رتبهٔ ماتریس A با تعداد مقدارهای تکین ناصفرِ آن برابر است. در عمل، به علت وجود خطاهای مختلف، امکان دارد استقلال خطی ستون‌های ماتریس تغییر کند. به تعبیر ریاضی، این ناسازگاری به‌سببِ وجود یک یا چند مقدار تکین بسیار کوچک رخ می‌دهد. این رویداد منجر به افزایش تأثیر خطا می‌شود که البته می‌توان با نادیده گرفتن این مقادیر تکین کوچک، از آن جلوگیری کرد. در نتیجه تعداد مقدارهای تکین

[۱] data cleaning

بزرگ به رتبهٔ عددی[۱] ماتریس اشاره دارد که به‌روشنی از رتبهٔ اصلی ماتریس کوچکتر است ($r < n$). در نمایش زیر که به روش تجزیهٔ مقدار تکین برشی (TSVD)[۲] معروف است، ماتریس A_r جایگزین ماتریس A می‌شود:

$$A \approx \sum_{i=1}^{r} \sigma_i u_i v_i^T =: A_r. \qquad (۳.۲)$$

از این‌رو می‌توان گفت که هر یک از داده‌های اولیه به‌صورت

$$a_j \approx \sum_{i=1}^{r} (\sigma_i v_{ij}) u_i, \qquad j = 1, \ldots, n \qquad (۴.۲)$$

توسط u_iها بازسازی می‌شود به‌طوری که سهم (وزن) هر بردار u_i برابر با $\sigma_i v_{ij}$ است. به این ترتیب، ماتریس متعامد U ماتریس پایه‌ای برای A خواهد بود. مراجع [۶] و [۱۲] را برای مطالعهٔ جزئیات بیشتر در مورد SVD و کاربردهای آن ببینید.

۲.۲. تحلیل مؤلفهٔ اصلی. در تحلیل مؤلفهٔ اصلی [۲۳] نیز به‌دنبال یک تبدیل خطی متعامد هستیم تا داده‌ها توسط این تبدیل به دستگاه مختصات جدید انتقال یابند و بر اساس آن، آسان‌تر توصیف شوند. برای این منظور، محورهای مختصات جدید، پی در پی در جهتی قرار می‌گیرند که داده‌ها دارای بیشترین پراکندگی (واریانس) باشند. در این روش، پس از تعیین ماتریس دادهٔ نرمال‌شدهٔ A و ماتریس کوواریانس C متناظر با آن، بردارهای ویژهٔ ماتریس C محاسبه و بر اساس کاهش مقادیر ویژهٔ متناظرشان، مرتب می‌شوند. با توجه به متقارن بودنِ ماتریس C، بردارهای ویژه چون متعامدند، تشکیل پایه می‌دهند. برای کاهش بُعد داده‌ها و حفظ اطلاعات مهم، r تا از بردارهای ویژهٔ نرمال‌شدهٔ متناظر با مقادیر ویژهٔ بزرگتر انتخاب خواهند شد. این r بردار پایه‌ای، تبدیل خطی $Q = [q_1 q_2 \ldots q_r]$ (ماتریس پایه) را برای تبدیل داده‌ها از فضای چند بُعدی به فضایی با بُعد کمتر تشکیل می‌دهند. در روش تحلیل مؤلفهٔ اصلی، داده‌های اولیه روی زیرفضای تولیدشده توسط ماتریس پایهٔ C تصویر می‌شوند که به این ترتیب، ماتریس متناظر $H = [h_1 h_2 \ldots h_r] \in \mathbb{R}^{r \times n}$ (ماتریس اوزان) به‌دست می‌آید [۲۳]:

$$A \approx QH = \sum_{j=1}^{r} q_j h_j^T =: A_r. \qquad (۵.۲)$$

از این‌رو می‌توان گفت که هر یک از داده‌های اولیه، توسط r مؤلفهٔ اصلی به‌صورت زیر بازسازی می‌شوند:

$$a_j \approx Q h_j = \sum_{i=1}^{r} q_i h_{ij}, \qquad j = 1, \ldots, n. \qquad (۶.۲)$$

[۱] numerical rank [۲] truncated SVD

خیلی وقت‌ها به‌دلیل وجود شباهت‌هایی بین روش‌های SVD و PCA، این دو به‌جای هم در نظر گرفته می‌شوند. در این باره، مطالعهٔ [۲۴] پیشنهاد می‌شود.

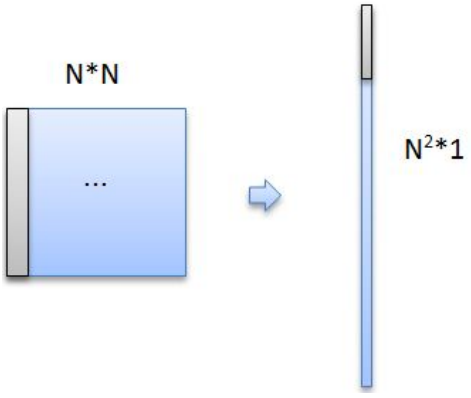
۳.۲. تجزیهٔ نامنفی ماتریسی.

۱.۳.۲. ***نظریه.*** بسیاری از داده‌های دنیای واقعی پیرامون ما از جمله شدت نور (پیکسل‌های مربوط به تصویر دیجیتال) و غلظت شیمیایی در بیوانفورماتیک (ژن‌ها)، داده‌هایی با مقادیر نامنفی هستند. روشن است که بردارهای پایه‌ای در SVD (u_iها) و PCA (q_iها) دارای خاصیت تعامد هستند. این خاصیت سبب ایجاد مؤلفه‌هایی با علامت‌های دلخواه در بردارهای پایه‌ای و اوزان می‌گردد. پس این روش‌ها نمی‌توانند شرط نامنفی بودنِ داده‌ها را تضمین کنند. این محدودیت، انگیزهٔ کافی برای تمرکز بر ابزاری بهتر و قوی‌تر مانند NMF را به‌وجود می‌آورد. این تجزیه، روشی سودمند برای نمایش خطی داده‌های نامنفی است به‌طوری که واقعیت فیزیکی (اطلاعات) آنها حفظ شود. NMF عوامل ماتریسی $W \in \mathbb{R}^{m\times r}$ و $H \in \mathbb{R}^{r\times n}$ را برای تقریب ماتریس نامنفی $A \in \mathbb{R}^{m\times n}$ با شرطِ $r \ll \min\{m,n\}$ چنان می‌یابد که رابطهٔ تقریبی $A \approx WH$ برقرار باشد. پیدا کردن چنین تقریبی، نیازمندِ تابع هزینه‌ای است که کیفیت تقریب را به‌خوبی نشان دهد. یکی از این تابع‌ها می‌تواند اندازهٔ فاصلهٔ دو ماتریس نامنفی از یکدیگر باشد. معمولاً این معادلهٔ تقریبی به‌صورت مسئلهٔ بهینه‌سازی

$$\min_{\substack{W\geq 0\\ H\geq 0}} f(W,H) \equiv \frac{1}{2}\|A - WH\|_F^2 \qquad (۷.۲)$$

صورت‌بندی می‌شود. در اینجا معنای نمادهای $H \geq 0$ و $W \geq 0$ این است که کلیهٔ درآیه‌های این ماتریس‌ها نامنفی هستند. در سراسر این مقاله، منظور از NMF رابطهٔ (۷.۲) خواهد بود. همان‌طور که قبلاً اشاره شد، تعبیرهای مختلفی از ساختار ضمنیِ داده‌ها توسط تجزیه‌های ماتریسی آشکار می‌شوند که از نظر ریاضی هم‌ارز هستند [۲۱].

مثال ۱.۲ (تعبیر عاملی). هر تصویر دیجیتال، یک آرایهٔ مستطیلی $N \times N$ از پیکسل‌ها است و هر پیکسل توسط شدت نورش نمایش داده می‌شود. چون شدت نور توسط مقادیری نامنفی معین می‌شود، می‌توان هر تصویر را یک ماتریس نامنفی در نظر گرفت. همانند شکل ۱، از پشت سر هم قرار دادنِ ستون‌های ماتریس $N \times N$، بردار N^2بُعدی متناظر با آن حاصل خواهد شد و مختصات این بردار، معرف پیکسل‌های تصویر هستند. به این ترتیب، ماتریس دادهٔ n ستونی متناظر با n دادهٔ تصویری تولید می‌شود. توجه به شکل برداریِ معادلهٔ تقریبی $A \approx WH$ با نمایش $a \approx Wh$، اهمیت ویژه‌ای دارد. در این نمایش، h و a معرف یک ستون از H و A هستند. این معادلهٔ تقریبی نشان می‌دهد که هر ستون از داده‌های ماتریس A را می‌توان به‌صورت ترکیبی خطی از r ستون ماتریس W نوشت (رابطه‌های (۴.۲)



شکل ۱. تبدیل داده از قالب ماتریس به بردار

و (۶.۲) را نیز ببینید). روشن است که ضرایب این ترکیب خطی مؤلفه‌های بردار h هستند. لذا به W ماتریس پایه و به ستون‌های آن، بردارهای پایه‌ای اطلاق می‌شود و همچنین H ماتریس اوزان خواهد بود. با توجه به نمایش $a \approx Wh$، می‌توان گفت که نمونهٔ a توسط عامل ماتریسی W بازسازی می‌شود.

مثال ۲.۲ (تعبیر هندسی). متنی متشکل از پنج جمله و دو موضوع متفاوت در اختیار است که جمله‌های آن را در قالب ۵ سند به ترتیبِ زیر مشاهده می‌کنید به‌طوری که موضوع سند آخر (football) مستقل از موضوع سایر اسناد (google) است [۱۲]:

The **google matrix** P is a model of the **Internet**.

P_{ij} is nonzero if there is a **link** from **web page** j to i.

The **google matrix** is used to **rank** all **web pages**.

The **ranking** is done by solving a **matrix eigenvalue** problem.

England dropped out of the top 10 in **FIFA ranking**.

ماتریس متن-سندِ متناظر با این متن، به‌صورت ماتریس دادهٔ

$$A = \begin{pmatrix} ۰ & ۰ & ۰ & ۱ & ۰ \\ ۰ & ۰ & ۰ & ۰ & ۱ \\ ۰ & ۰ & ۰ & ۰ & ۱ \\ ۱ & ۰ & ۰ & ۰ & ۱ \\ ۱ & ۰ & ۱ & ۰ & ۰ \\ ۰ & ۰ & ۰ & ۰ & ۰ \\ ۱ & ۰ & ۰ & ۰ & ۰ \\ ۰ & ۱ & ۱ & ۱ & ۰ \\ ۰ & ۰ & ۱ & ۱ & ۱ \\ ۰ & ۱ & ۱ & ۰ & ۰ \end{pmatrix}$$

تشکیل می‌گردد که در آن، اعداد ۱ و ۰ نشان‌دهندهٔ حضور و عدم حضور کلمات کلیدی در هر سند هستند. با این دانش، می‌توان میانگین بردارهای ستونی اول تا چهارم را نمایندهٔ خوشه‌ای با موضوع google و طبیعتاً بردار ستونی آخر را به‌عنوان نمایندهٔ خوشه‌ای با موضوع football دانست. هر نماینده را می‌توان یک بردار پایه‌ای برای ماتریس پایهٔ $W \in \mathbb{R}^{۱۰\times۲}$ در نظر گرفت و مختصات ستون‌های ماتریس A را برحسب این بردارهای پایه‌ای به‌صورت زیر محاسبه کرد:

$$\min_{H\geq ۰} \|A - WH\| . \tag{۸.۲}$$

ستون‌های ماتریس H همان مختصات جدید داده‌های اولیه هستند که با توجه به بیشینهٔ مقادیر هر یک از این ستون‌ها، می‌توان خوشهٔ متناظر با هر یک از داده‌ها را تعیین کرد (خوشه‌بندی). اگر در (۸.۲) قید مثبت بودن حذف شود، این احتمال وجود خواهد داشت که در H مؤلفه‌های منفی ظاهر شود. این اتفاق بیان می‌کند که ماتریس نامنفی A با ماتریس رتبهٔ پائین‌تر اما با درآیه‌های منفی، تقریب زده شده است که این در واقعیت، غیرطبیعی به‌نظر می‌رسد. چون در این شرایط، حضور یا عدم حضور کلمات در اسناد، با مقادیر منفی بیان خواهد شد که تعبیر صحیحی نیست. در تعبیر هندسی، وجود پایه برای تصویر کردن داده‌ها ضروری است، اما نکتهٔ قابل توجه این است که مفهوم پایه در NMF برای ستون‌های ماتریس W، برخلافِ دو تجزیهٔ دیگر، با تعریفِ پایه در جبرخطی، فرق دارد. به‌عبارت دیگر، ماتریس W تنها با نام ماتریس پایه، نامگذاری و شناخته می‌شود.

۲.۳.۲. ***روش‌های محاسبه.*** تابع هزینه در مسئلهٔ NMF (۷.۲) همزمان نسبت به دو عامل ماتریسی H و W نامحدب است. لذا نمی‌توان در پی یافتنِ کمینهٔ سراسری بود. از این‌رو همگرایی به نقطهٔ ایستا، هدفی است که در همهٔ الگوریتم‌های NMF دنبال می‌شود. بر اساس قضیه‌ای در [۳]، اگر (W, H) نقطهٔ کمینهٔ موضعی مسئلهٔ (۷.۲) باشد، آن‌گاه شرایط KKT[۱] به‌صورت

$$\begin{aligned}
&H \geq ۰,\ W \geq ۰\\
&\nabla_H f(W,H) = W^T WH - W^T A \geq ۰,\\
&\nabla_W f(W,H) = WHH^T - AH^T \geq ۰,\\
&H. * \nabla_H f(W,H) = ۰,\ W. * \nabla_W f(W,H) = ۰
\end{aligned}$$

برقرارند. نمادهای $*.$ و ∇ به‌ترتیب، نشان‌دهندهٔ ضرب مؤلفه‌ای عوامل و گرادیان تابع هستند. پس (W, H) یک نقطهٔ ایستا برای مسئلهٔ (۷.۲) است اگر و تنها اگر در شرایط KKT صدق کند.

[۱] Karush–Kuhn–Tucker conditions

در سال ۱۹۹۹ مطالعات جدیدی دربارهٔ NMF توسط لی و سانگ[۱] [۱۴] آغاز گردید. البته پیش از این، در مورد تجزیهٔ مثبت ماتریسی[۲] [۱۹] چندین اثر از پاترو[۳] در سال‌های ۱۹۹۴، ۱۹۹۷ و ۱۹۹۹ منتشر شده بود. با وجود این، اثر لی و سانگ به‌عنوان اساس و NMF استاندارد در نظر گرفته می‌شود. روش‌های تناوبی مختلفی به‌منظور سرعت بخشیدن به الگوریتم استاندارد NMF پیشنهاد شده‌اند. در این باره می‌توان به پژوهش‌های لین[۴] در سال ۲۰۰۵ [۱۶، ۱۷]، بری[۵] و همکاران وی در سال ۲۰۰۷ [۲] و سپس کیم و پارک[۶] در سال‌های ۲۰۰۷ و ۲۰۰۸ [۸، ۹] اشاره کرد. کلی‌ترین دسته‌بندی برای الگوریتم‌های NMF عبارت است از: قواعد بهنگام ضربی[۷]، روش کمترین مربعات متناوب[۸] و روش‌های گرادیان[۹].

۳.۳.۲. ***قواعد بهنگام ضربی.*** پرکاربردترین رهیافت برای مسئلهٔ (۷.۲) را می‌توان قواعد بهنگام ضربی دانست که توسط لی و سانگ در سال ۲۰۰۱ پیشنهاد گردید [۱۵]. گفتیم که تابع هزینهٔ مسئلهٔ NMF نامحدب است و لذا نمی‌توان نقطهٔ کمینهٔ سراسری را پیدا کرد. اما روش‌هایی در بهینه‌سازی عددی وجود دارند که برای پیدا کردن کمینهٔ موضعی به‌کار می‌روند. شاید روش‌های کاهش گرادیان [۳] از ساده‌ترین آنها باشند. به‌طور کلی، الگوریتم لی و سانگ را می‌توان یک روش کاهش گرادیان دانست که در قالب یک فرم متناوب برای بهنگام کردن مؤلفه‌های عوامل ماتریسی از قواعد

$$W_{ia} \leftarrow W_{ia} - \eta_{ia}(\nabla_W f(W,H))_{ia}, H_{bj} \leftarrow H_{bj} - \mu_{bj}(\nabla_H f(W,H))_{bj} \quad (۹.۲)$$

در خلاف جهت گرادیان استفاده می‌کند. در اینجا $\eta_{ia} = \frac{W_{ia}}{(WHH^T)_{ia}}$ و $\mu_{bj} = \frac{H_{bj}}{(W^TWH)_{bj}}$ به‌عنوان طول گام در نظر گرفته می‌شوند. توجه می‌کنیم که همگرایی روش‌های مبتنی بر گرادیان، به انتخاب طول گام حساس است. لی و سانگ برای کمینه کردن نرم فروبنیوس $\|A - WH\|_F^2$، قواعد بهنگام ضربی

$$W_{ia} \leftarrow W_{ia}\frac{(AH^T)}{(WHH^T)_{ia}}, \quad H_{bj} \leftarrow H_{bj}\frac{(W^TA)_{bj}}{(W^TWH)_{bj}} \quad (۱۰.۲)$$

را پیشنهاد کردند. روشن است که وقتی $A = WH$، هر یک از ضرایب این قواعد برابر با یک خواهند شد؛ به این معنی که $\nabla_H f(W,H) = ۰$ و $\nabla_W f(W,H) = ۰$. لذا این روش یک نوع روش نقطه ثابت است. تحت این شرایط، لی و سانگ ثابت کردند که نرم فروبنیوس $\|A - WH\|_F^2$ تحت قواعد (۱۰.۲) ناافزایشی است. البته نسخهٔ اصلی این قضیه در [۱۵] شامل قسمت دومی نیز هست که بعداً معلوم شد لزوماً درست نیست [۲، ۷، ۱۶]. لی و سانگ در قسمت دوم نسخهٔ اصلی قضیه ادعا کردند که این نرم، تحت قواعد (۱۰.۲)، در یک نقطهٔ ایستا تغییر نمی‌کند به این دلیل که الگوریتم لی و سانگ، یک روش نقطه ثابت است و لذا اگر با شروع از $H > ۰$ به نقطهٔ ایستای $H_* > ۰$ همگرا شود،

---

[۱] Daniel Lee and Sebastian Seung [۲] positive matrix factorization [۳] Paatero Pentti [۴] Chih-Jen Lin [۵] Michael Berry [۶] Hyunsoo Kim and Haesun Park [۷] multiplicative update rules [۸] alternating least squares method [۹] gradient methods

آن‌گاه $\nabla_H f(W_*, H_*) = ۰$. اما علی‌رغم مثبت بودنِ عامل ماتریسی در همهٔ تکرارها، ممکن است برخی از مؤلفه‌های آن به صفر میل کنند، زیرا ما با دستگاه‌هایی سروکار داریم که اعداد در آنها دارای دقت محدود هستند و در نتیجه اعداد بسیار کوچک، به صفر گرد می‌شوند. به عبارت دیگر، مؤلفه‌های صفر بهنگام نمی‌شوند. حال اگر در یکی از تکرارها، مؤلفه‌ای مانند H_{bj} صفر شود، آن‌گاه $H_{*bj} = ۰$. اما با توجه به شرط $\nabla_H f(W_*, H_*)_{bj} \geq ۰$، معلوم نیست قاعدهٔ بهنگام چگونه باید عمل کند، زیرا وقتی $\nabla_H f(W_*, H_*) \neq ۰$، روش باید نقطهٔ فعلی را بهنگام کند که در این صورت حتماً بهینه بودنِ جواب فعلی نقض خواهد شد. حتی گاهی ممکن است با صفر شدنِ مؤلفه‌ای، مقدار گرادیان منفی شود. کلیهٔ استدلال‌های بیان‌شده، برای قاعدهٔ دوم نیز برقرار است. با دلایل ذکر شده، الگوریتم لی و سانگ دارای همگرایی قوی نیست. معمولاً برای جلوگیری از به‌وجود آمدن مشکلات عددی، عدد کوچک مثبتی مانند $\epsilon = ۱۰^{-۹}$ به مخرج کسر هر یک از قواعد اضافه می‌کنند. الگوریتم لی و سانگ برای تابع هزینهٔ نرمدار، به‌صورت زیر معرفی می‌شود:

مقداردهیِ اولیه: $W^k \geq ۰$ و $H^k \geq ۰$ وقتی $k = ۱$

تکرار مراحل زیر تا برقراری شرط توقف:

(۱) اعمال قواعد (۱۰.۲) با افزودن مقدار ϵ به مخرج کسرها؛

(۲) $k = k + ۱$.

۴.۳.۲. ***روش کمترین مربعات متناوب.*** با توجه به نامحدب بودنِ تابع هزینه در مسئلهٔ NMF، تبدیل این مسئله به دو زیرمسئلهٔ محدب و در نتیجه استفاده از ویژگی‌های بهینه‌سازی محدب منطقی به‌نظر می‌رسد. در اینجا با ثابت نگه‌داشتنِ یکی از عوامل ماتریسی، می‌توان عامل ماتریسی دیگر را از یک مسئلهٔ کمترین مربعات به‌دست آورد. به این ترتیب در روش کمترین مربعات متناوب با قید نامنفی (ANLS)[۱]، مسئلهٔ (۷.۲) به دو مسئلهٔ کمترین مربعات با قید نامنفی به‌صورت

$$\min_{H \geq ۰} \frac{۱}{۲} \|A - WH\|_F^۲ \tag{۱۱.۲}$$

با عامل ثابت W و

$$\min_{W \geq ۰} \frac{۱}{۲} \|A^T - H^T W^T\|_F^۲ \tag{۱۲.۲}$$

با عامل ثابت H^T (H) بازنویسی می‌شود و تا برقراریِ شرط همگرایی، به‌طور تناوبی به‌دنبال هم می‌آیند. از آنجا که عامل سمت راست در زیرمسئله‌های (۱۱.۲) و (۱۲.۲) یک ماتریس است، در ادامه از نام NLS[۲] ماتریسی برای آنها استفاده می‌شود.

[۱] alternating non-negativity-constrained least squares [۲] non-negative least squares

در سال ۲۰۰۷، بری و همکاران وی [۲] الگوریتم پایهٔ ALS[۱] را برای NMF بیان کردند. در این الگوریتم، هر یک از زیرمسئله‌های (۱۱.۲) و (۱۲.۲) به‌صورت LS نامقید حل می‌شود و در صورت به‌وجود آمدنِ درآیه‌های منفی در عامل ماتریسی، مقدار صفر جایگزین آن درآیه‌ها خواهد شد. این عمل، ضمنِ برقراری شرط نامنفی بودنِ عوامل ماتریسی تولیدشده، به تنکی[۲] آنها نیز کمک می‌کند.

مقداردهیِ اولیه: $W^k \geq 0$ وقتی $k = 1$

تکرار مراحل زیر تا برقراری شرط توقف:

(۱) حل مسئلهٔ (۱۱.۲) به‌صورت نامقید برای عامل ماتریسی H؛

(۲) تعویض درآیه‌های منفی ماتریس H با مقدار صفر (عمل تصویر کردن)؛

(۳) حل مسئلهٔ (۱۲.۲) به‌صورت نامقید برای عامل ماتریسی H؛

(۴) تعویض درآیه‌های منفی ماتریس W با مقدار صفر (عمل تصویر کردن)؛

(۵) $k = k + 1$.

اگرچه این الگوریتم موجبِ تسریع محاسبات می‌شود، اما تحلیل همگرایی برای آن سخت خواهد بود، زیرا قیدهای نامنفی در دو زیرمسئلهٔ محدب دقیقاً برقرار نمی‌شوند. در حالت کلی، این الگوریتم به نقطهٔ ایستا همگرا نیست و از آن به‌طور مستقیم استفاده نمی‌شود. معمولاً این الگوریتم به‌عنوان یک راه‌انداز برای الگوریتم‌های دیگر به‌کار می‌رود. هر NLS ماتریسی به‌ترتیب با ثابت در نظر گرفتنِ W و H، با مسائل

$$\min_{H_{:j} \geq 0} \frac{1}{2} \sum_j \|A_{:j} - W H_{:j}\|_2^2, \qquad j = 1, 2, \ldots, n \tag{۱۳.۲}$$

و

$$\min_{W_{:j}^T \geq 0} \frac{1}{2} \sum_j \|A_{:j}^T - H^T W_{:j}^T\|_2^2, \qquad j = 1, 2, \ldots, m \tag{۱۴.۲}$$

معادل است. از این‌رو زیرمسئله‌های (۱۱.۲) و (۱۲.۲) به‌ترتیب به n و m مسئلهٔ NLS برداری تبدیل می‌شود. نماد $X_{:j}$ را برای نشان دادن ستون jام ماتریس X به‌کار می‌بریم. برای حل هر NLS برداری، می‌توان از الگوریتم NNLS [۱۳] با نام کاربردیِ مجموعهٔ *فعال*[۳] استفاده کرد. این الگوریتم در نرم‌افزار متلب با نام تابع lsqnonneg در دسترس است.

می‌دانیم که ناحیهٔ پذیرفتنی یک LS مقید با قید نامنفی، همان قید مفروض است. با این فرض، هر جواب مسئلهٔ NLS برداری باید در ناحیهٔ پذیرفتنی باشد و هر جواب پذیرفتنی دارای دو رده از متغیرها است که توسط آنها شناسایی می‌شود. ردهٔ اول، شامل متغیرهایی است که در درون ناحیهٔ پذیرفتنی و ردهٔ دوم، شامل متغیرهایی است که بر روی مرز ناحیهٔ پذیرفتنی قرار دارند. مجموعهٔ غیرفعال، توسط اندیس متناظر با متغیرهای ردهٔ اول و مجموعهٔ فعال توسط اندیس متناظر با متغیرهای ردهٔ دوم شناخته می‌شوند.

[۱] alternating least squares [۲] sparsity [۳] active set

سازوکار الگوریتم مجموعهٔ فعال در حل مسئلهٔ NLS برداری این گونه است که برای هر جواب پذیرفتنی آغازین، متغیرها به‌طور تکراری بین مجموعهٔ فعال و غیرفعال جابه‌جا می‌شوند. در چنین حالتی، تابع هزینه به‌طور پیوسته و یکنواخت کاهش پیدا می‌کند در حالی که پذیرفتنی بودنِ جواب، محفوظ می‌ماند. برای درک و دریافت جزئیات بیشتر این الگوریتم، مطالعهٔ مرجع [۲۰] پیشنهاد می‌شود. روشن است که استفاده از الگوریتم مجموعهٔ فعال برای همهٔ NLSهای برداری، حتی با انجام محاسبات خیلی سریع، منجر به اجرایی کُند و غیر قابل قبول می‌شود. به‌منظور ارائهٔ جایگزینی که برای حل زیرمسائل (۱۳.۲) و (۱۴.۲) مناسب باشد، می‌توان به اقدامات کیم و پارک [۸] در استفاده از روش مجموعهٔ فعال ترکیبیاتی (FC-NNLS)[۱] و نیز اقدامات لین [۱۷] در استفاده از روش تصویر گرادیان[۲] اشاره کرد.

الگوریتم FC-NNLS توسط بنتم و کینن[۳] برای حل NLS ماتریسی پیشنهاد گردید. در حقیقت، این الگوریتم استفادهٔ مستقیم از الگوریتم مجموعهٔ فعال است. اشاره کردیم که هر NLS ماتریسی را می‌توان به چندین NLS برداری تبدیل کرد. تعداد NLSهای برداری به تعداد ستون‌های ماتریس ضرایب زیرمسائل بستگی دارد. لذا می‌توان به‌جای اجرای الگوریتم تکراریِ مجموعهٔ فعال برای هر NLS برداری، فقط یک تکرار از این الگوریتم را برای یک NLS ماتریسی به‌کار برد. در این استراتژی ترکیبیاتی، ستون‌هایی از ماتریس ضرایب که دارای مجموعه‌های غیرفعال یکسان‌اند با هم در نظر گرفته می‌شوند و مسئله به‌ازای متغیرهای متناظر با مجموعهٔ غیرفعال، حل می‌شود. برای مطالعهٔ دقیق ساختار الگوریتم FC-NNLS همراه با شبه‌کد متلب آن به نام fcnnls، به [۱] رجوع کنید. اگرچه سرعت رهیافت پیشنهادی توسط کیم و پارک مطلوب است، نسخه‌هایی سریع‌تر از آن در مراجع [۱۰، ۱۱] ارائه شده است.

مقداردهیِ اولیه: $W \geq 0$

تکرار مراحل زیر تا برقراری شرط توقف:

(۱) حل مسئلهٔ (۱۱.۲) با استفاده از fcnnls؛

(۲) حل مسئلهٔ (۱۲.۲) با استفاده از fcnnls.

۵.۳.۲. ***روش‌های گرادیان***. برای تعیین عوامل ماتریسی در NMF، همانند قواعد بهنگام ضربی، می‌توان از قواعد بهنگام مبتنی بر روش گرادیان به‌صورت زیر استفاده کرد:

$$H = H - \alpha_H \nabla_H f(W,H), \qquad W = W - \alpha_W \nabla_W f(W,H). \qquad (۱۵.۲)$$

پارامترهای α_W و α_H به‌عنوان طول گام، وابسته به الگوریتم بوده و در خلاف جهت گرادیان انتخاب می‌شوند. در (۱۵.۲) برای تعیین هر یک از عوامل ماتریسی، عامل دیگر، ثابت در نظر گرفته می‌شود. اما نکتهٔ قابل توجه، چگونگیِ انتخاب طول گام‌ها است. می‌توان در ابتدا از مقدار یک برای آنها استفاده

[۱] fast combinatorial non-negative least squares [۲] projected gradient [۳] Mark H.Van Benthem and Michael R.Keenan

کرد و در تکرارهای بعدی این مقدار را نصف نمود. با وجود این، تضمینی در منفی نشدنِ عوامل وجود ندارد. از طرفی، اگر بخواهیم عمل تصویر کردن را انجام دهیم، به این معنی که هر مقدار منفی را با صفر جایگزین کنیم، تحلیل همگرایی سخت‌تر خواهد شد. قبلاً اشاره کردیم که می‌توان الگوریتم NMF/MUR را یک روش کاهش گرادیان در نظر گرفت. در آنجا (روابط ۹.۲) مشاهده شد که همگرایی حتی با انتخاب دقیق‌تر طول گام‌ها، آرام است. لذا در حالت کلی می‌توان نتیجه گرفت که رهیافت‌های مبتنی بر روش گرادیان، همگرایی مطلوبی ندارند. محاسبهٔ عوامل ماتریسی NMF را با روش‌های گرادیان می‌توان به سه مرحله: محاسبهٔ گرادیان، انتخاب طول گام و تصویر کردن عامل ماتریسی جدید در فضایی نامنفی، برای هر تکرار تقسیم کرد. گاهی می‌توان مراحل انتخاب طول گام و عمل تصویر کردن عوامل ماتریسی جدید را در یک تکرار و با هم در نظر گرفت تا کاهش کافی در مقدار تابع هزینه حاصل شود. به این ترتیب در هر تکرار، الگوریتم‌های مبتنی بر روش گرادیان از یک حلقهٔ داخلی به‌ازای هر قاعدهٔ بهنگام برخوردار خواهند شد. می‌توان به‌طور فرضی، کران‌های بالای بسیار بزرگی را برای قیود زیرمسئله‌های (۱۱.۲) و (۱۲.۲) در نظر گرفت. آن‌وقت هر یک از آنها یک مسئلهٔ مقید کراندار خواهد بود که لین برای حل این زیرمسئله‌ها، از رهیافتی مبتنی بر روش تصویر گرادیان [۳] استفاده کرد. راه‌های مختلفی برای انتخاب طول گام در این روش وجود دارد که در میان آنها، روش آرمیژو [۳] الگوریتمی ساده و مؤثر است. در این راستا، لین از یک سو کاهش هزینه را با توجه به درجه دو بودنِ تابع هدف و از سوی دیگر، انعطاف بیشتری را برای تعیین طول گام مورد توجه قرار داد. بر این اساس، طول گامی به‌عنوان حدس اولیه در نظر گرفته می‌شود و سپس در جریان الگوریتم افزایش یا کاهش می‌یابد. رهیافت لین این امکان را می‌دهد که طول گام‌های بزرگتر از یک نیز در جستجو، امتحان شوند. جزئیات بیشتر در مورد استفادهٔ لین از این رهیافت به‌همراهِ شبه‌کد متلب آن با نام alspgrad، در [۱۷] موجود است.

مقداردهیِ اولیه: $W \geq 0$ و $H \geq 0$
تکرار مراحل زیر تا برقراری شرط توقف [۱۷] :
(۱) حل مسئلهٔ (۱۱.۲) با استفاده از alpsgrad؛
(۲) حل مسئلهٔ (۱۲.۲) با استفاده از alpsgrad.

۶.۳.۲. ***مقداردهیِ اولیه و شرط توقف.*** برای NMF دو مبحث مقداردهیِ اولیه و شرط توقف، در دستیابی به عوامل ماتریسی بهین و همگرایی نقش به‌سزایی دارند. مشکلی که اغلبِ الگوریتم‌های NMF با آن روبه‌رو هستند، عدم تضمین همگرایی به نقطهٔ کمینهٔ سراسری است. مقداردهیِ اولیهٔ ضعیف (مانند روش تصادفی) اغلب همگرایی آرام و گاهی جواب‌های بی‌ربط و غلطی را نتیجه می‌دهد. اگر یک مقدار اولیه خوب نتواند به قدر کافی همگرایی به نقطهٔ ایستا را ضمانت کند، حتماً به اندازهٔ کافی از تعداد تکرارها خواهد کاست. کارایی بسیاری از الگوریتم‌های NMF تحت تأثیر انتخاب ماتریس‌های اولیه است و لذا

مهم است که روش‌های سازگار و کارا برای مقداردهیِ اولیه به عوامل ماتریسی در دست باشند. مقداردهیِ اولیه مبتنی بر روش NMF [۵] و مقداردهیِ مبتنی بر روش SVD [۴] از این گونه الگوریتم‌ها هستند. در این مقاله، از مقداردهیِ مبتنی بر NMF استفاده می‌شود که در آن، ابتدا c زوج ماتریس اولیه (W, H) به‌صورت تصادفی و یا خروجی از یک الگوریتم ساده و سریع مانند NMF/ALS تولید می‌شوند. معمولاً ۱۰ تا ۲۰ زوج ماتریس کافی است. سپس عواملی مانند $(W^{i_{\min}}, H^{i_{\min}})$ که در آن،

$$i_{\min} = \text{argmin}_{1 \le i \le c} \|A - W^i H^i\|_F^2$$

انتخاب می‌شوند. به این معنی که کمترین مقدار برای تابع هزینهٔ مسئلهٔ NMF به‌ازای این عوامل به‌دست می‌آید. در نتیجه از آنها می‌توان به‌عنوان عوامل ماتریسی اولیه استفاده کرد. برای هر الگوریتم تکراری، از جمله الگوریتم‌های تکراری NMF باید شرط خاتمه‌ای وجود داشته باشد. علاوه بر معیارهای توقفی که در مرجع [۵] برای NMF آمده‌اند، معیارهای قوی‌تر دیگری توسط لین برای رهیافت مبتنی بر روش تصویر گرادیان و توسط کیم و پارک برای رهیافت مبتنی بر روش مجموعهٔ فعال معرفی شدند. البته شرط توقفی که لین [۱۷] معرفی کرد، متناسب با همان رهیافت مربوطه است؛ در حالی که شرط توقف پیشنهادیِ کیم و پارک، برای همهٔ روش‌های محاسبه NMF از عمومیت لازم برخوردار است. ما نیز در این مقاله از معیار توقف کیم و پارک استفاده می‌کنیم. شرایط KKT را می‌توان به‌صورت فشردهٔ

$$\min(W, \nabla_W f(W, H)) = 0, \qquad \min(H, \nabla_H f(W, H)) = 0 \qquad (۱۶.۲)$$

بیان کرد. به عبارت‌های سمت چپ علامت تساوی، ماندهٔ KKT اطلاق می‌شود. عبارت

$$\Delta = \frac{\Delta_0}{\sigma_W + \sigma_H} \qquad (۱۷.۲)$$

ماندهٔ KKT نرمال‌شده است که در آن، σ_W و σ_H تعداد مؤلفه‌های W و H هستند که در حال حاضر، مقدار ماندهٔ KKT به‌ازای آنها صفر نشده است و

$$\begin{aligned} \Delta_0 &= \sum_{i=1}^{m} \sum_{a=1}^{r} |\min(W_{ia}, (\nabla_W f(W, H))_{ia})| \\ &+ \sum_{b=1}^{r} \sum_{j=1}^{n} |\min(H_{bj}, (\nabla_H f(W, H))_{bj})|. \end{aligned} \qquad (۱۸.۲)$$

بنابراین اگر Δ_1 مقدار Δ بعد از یک تکرار و ϵ یک تلورانس معین باشد، معیار همگرایی KKT که کیم و پارک ارائه کردند، به‌صورت زیر است:

$$\Delta \le \epsilon \Delta_1. \qquad (۱۹.۲)$$

۳. کاربردها و نتایج عددی

کاربردهای NMF از همان دههٔ ۹۰ میلادی با اقدامات پاترو با نام تجزیهٔ مثبت ماتریسی آغاز شدند و سپس با اقدامات لی و سانگ مورد توجه ویژه‌ای قرار گرفتند. در حال حاضر، طیف گسترده‌ای از این کاربردها در مقاله‌های گوناگون مورد بحث هستند. NMF یک روش کاهش بُعد است که علاوه بر داشتنِ این توانایی، می‌توان از آن به‌عنوان یک خوشه‌بند یا رده‌بند نیز استفاده کرد. خوشه‌بندی یا رده‌بندی داده‌ها از روی مختصات جدید آنها شدنی خواهد بود. به عبارت دیگر، پس از اِعمال NMF بر روی ماتریس داده‌ها و کاهش بُعد آن، انجام عمل خوشه‌بندی یا رده‌بندی در مختصات جدید ساده‌تر می‌شود. در این مقاله، به این دو کاربرد می‌پردازیم. در کاربردهای زیر، پارامتر r مقدار معینی دارد که توسط کاربر تعیین می‌شود. این مقدار می‌تواند با توجه به نوع کاربردها، تعداد خوشه‌ها و یا رده‌ها باشد.

۱.۳. *خوشه‌بندی.* مسئلهٔ خوشه‌بندی همان گروه‌بندی بدون نظارت داده‌ها در چندین گروه (خوشه) با ویژگی‌هایی مشابه است. اگر بحث تفسیر داده‌ها خصوصاً داده‌های نامنفی مطرح نباشد، k-means معمول‌ترین روش خوشه‌بندی است [۱۲]. پارامتر k در k-means معادل با پارامتر r در NMF است. در عمل خوشه‌بندی با NMF، حداکثر تعداد درآیه در ستون jام ماتریس H شناسایی شده و دادهٔ متناظر با این ستون به خوشهٔ متناظر با محل این عنصر، تخصیص می‌یابد. برای خوشه‌بندیِ داده‌ها با روش k-means، k نمونه به‌عنوان افراز اولیه تعیین می‌شوند به‌طوری که این نمونه‌ها، دارای کمترین فاصلهٔ اقلیدسی از میانگین کل باشند. در روش NMF نیز از همین افراز برای مقداردهیِ اولیهٔ عامل ماتریسیِ متناظر استفاده می‌گردد. اما برای شرط توقف روش k-means، وابستگی بهین بین داده‌ها کمتر از تلورانس δ در نظر گرفته [۱۲] و برای روش NMF از شرط توقف KKT استفاده می‌شود.

۱.۱.۳. ***خوشه‌بندی اسناد.*** یکی از مسائل مورد بحث متن‌کاوی[۱]، خوشه‌بندی داده‌های متنی با ویژگی‌های پنهان (مانند موضوع) می‌باشد. با فراخوانی ماتریس متن-سند در مثال ۲ و توضیحاتِ آن، روشن است که تعداد خوشه‌ها $k = r = ۲$، زیرا به وجودِ دو موضوع در اسناد آگاهی داریم. با این فرض، تعداد اسناد شناسایی‌شده برای موضوع google و موضوع football توسط هر دو روش k-means (با $\delta = ۱۰^{-۱۰}$) و NMF (با $\epsilon = ۱۰^{-۳}$) به‌ترتیب برابر با ۳ و ۱ به‌دست آمد. لذا هر دو روش در تشخیص موضوع تنها یک سند دچار ضعف بوده‌اند.

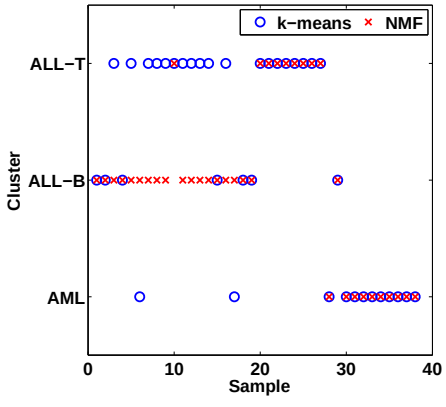
۲.۱.۳. ***تحلیل ژنی در ریزآرایه.*** داده‌های ژنی ALLAML[۲] متشکل از سرطان حاد لنفوم (ALL) با ۱۹ نمونه سلول از نوع B و ۸ نمونه سلول از نوع T و سرطان حاد لنفوم (AML) با ۱۱ نمونه سلول است. این داده‌ها شامل ۵۰۰۰ ژن هستند. برای ماتریس دادهٔ ریزآرایه $A \in \mathbb{R}^{۵۰۰۰\times ۳۸}$ نتایج خوشه‌بندی دو روش k-means با $\delta = ۱۰^{-۱۰}$ و NMF با $\epsilon = ۱۰^{-۶}$ و $k = r = ۳$ در جدول ۱ ارائه شده

[۱] text mining [۲] http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi

است. شکل ۲ نحوهٔ خوشه‌بندی نمونه‌ها را در ۳ خوشه توسط این دو روش نمایش می‌دهد. روشن است که روش NMF در خوشه‌بندی داده‌های ژنی، عملکرد بهتری نسبت به روش k-means داشته است. مشاهده می‌شود که این روش تنها در تشخیص خوشهٔ مربوط به نمونه‌های ۱۰ و ۲۹ ناتوان است.

جدول ۱. نتایج خوشه‌بندی داده‌های ژنی توسط k-means و NMF

	۱۹ نمونه ALL-B	۸ نمونه ALL-T	۱۱ نمونه AML
روش k-means	۶	۸	۱۰
روش NMF	۱۸	۸	۱۰



شکل ۲. نحوهٔ تفکیک داده‌های ژنی به سه خوشه توسط k-means و NMF

۲.۳. رده‌بندی. مسئلهٔ رده‌بندی داده‌ها عبارت است از تخصیص دادهٔ مفروض به یکی از چندین ردهٔ از پیش تعیین شده. رده‌بندی را می‌توان برحسب هر دو شیوهٔ یادگیری با نظارت و بدون نظارت توصیف کرد که در این مقاله، از NMF به‌صورت یادگیری با نظارت استفاده می‌شود. بخش مهمی از روند شناسایی الگو، مسئلهٔ رده‌بندی در حوزهٔ تصویر است که از آن می‌توان به دو مسئلهٔ استاندارد شناسایی چهره و شناسایی ارقام دست‌نویس اشاره کرد.

۱.۲.۳. ***شناسایی چهره.*** در شناسایی تصاویر چهره توسط NMF [۲۵]، همهٔ داده‌های تصویری آموزشی و آزمایشی روی فضای تولیدشده توسط بردارهای پایه‌ای (ستون‌های ماتریس W) تصویر شده و بردارهای ویژگی جدید (ستون‌های ماتریس H) تولید می‌شوند. سپس عمل مقایسه به‌منظورِ رده‌بندی

چهره‌ها توسط این بردارهای ویژگی انجام می‌گیرد. معیارهای مختلفی برای این مقایسه وجود دارد که معمولاً از ساده‌ترین شکل آن یعنی فاصلهٔ اقلیدسی استفاده می‌شود. پایگاه داده (ORL)[۱] شامل ۴۰۰

شکل ۳. سه رده از پایگاه داده ORL

تصویر سیاه و سفید از چهرهٔ ۴۰ فرد متفاوت (۴۰ رده و هر یک شامل ۱۰ تصویر) در اندازه ۹۲ × ۱۱۲ با نورپردازی و حالات چهرهٔ مختلف مفروض است. در این مقاله، ۲۰ رده از این پایگاه انتخاب و از هر رده، ۴ تصویر آموزشی و ۶ تصویر آزمایشی با تبدیل اندازه به ۱۶ × ۱۶ در نظر گرفته شده است. به این ترتیب، بر اساس مثال ۱، ماتریس دادهٔ $A \in \mathbb{R}^{۲۵۶\times ۸۰}$ تشکیل می‌گردد. ماتریس دادهٔ A از ۸۰ تصویر آموزشی تشکیل شده است که عوامل ماتریسی آن به‌صورت $A \approx W_۱H_۱$ تعیین می‌شوند. همچنین ماتریس دادهٔ A' متشکل از ۱۲۰ چهرهٔ آزمایشی بر $W_۱$ به‌صورت $A' \approx W_۱H_۲$ تصویر می‌شود. اشاره شد که همه داده‌های تصویری برای تولید بردارهای ویژگی جدید، باید روی فضای تولیدشده توسط بردارهای پایه‌ای تصویر شوند. لذا $H_۱$ و $H_۲$ به‌ترتیب شامل بردارهای ویژگی جدید در فضای تولیدشده توسط $W_۱$ برای بازسازی چهره‌های آموزشی و آزمایشی است. از آنجا که داده‌های آموزشی شامل ۴ تصویر از هر فرد است، متناظراً میانگین ۴ بردار ویژگی جدید (h_{m_i}) محاسبه شده و این عمل برای هر ۲۰ ردهٔ آموزشی تکرار می‌شود. سرانجام ردهٔ مربوط به چهرهٔ نامعلوم jام به‌صورت زیر تعیین می‌گردد:

$$\min_{m_i} \|(H_۲)_{:j} - h_{m_i}\|, \qquad i = ۱, \ldots, ۲۰, \quad j = ۱, \ldots, ۱۲۰. \qquad (۱.۳)$$

در جدول ۲ تعداد تشخیص درست چهره‌های آزمایشی ORL به روش NMF با رتبهٔ $r = ۲۰$ ارائه شده است. با توجه به این نتایج، میزان موفقیت روش NMF در رده‌بندی چهره‌ها ۸۴٫۱۶ درصد است. با استفاده از روش PCA، این نتیجه ۹۴٫۱۶ درصد برآورد شده است. به این ترتیب، روش NMF علاوه بر رفع مشکل تفسیر و نمایش مبتنی بر اجزاء، می‌تواند قابل رقابت با روش PCA نیز باشد. حتی این نتیجه را می‌توان برای NMF با معیارهای مقایسه‌ای بهتری در (۱.۳) بهبود داد [۲۵].

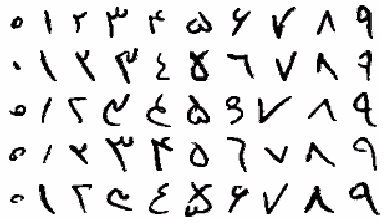
۲.۲.۳. ***شناسایی ارقام دست‌نویس.*** رده‌بندی تصاویر ارقام دست‌نویس [۱۲] با توجه به ماتریس پایهٔ (شاخص) هر رده انجام می‌پذیرد. برخلافِ مسئلهٔ قبل که روش NMF بر روی همهٔ داده‌ها اِعمال

[۱] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

جدول ۲. نتایج رده‌بندی چهره‌ها با روش NMF

شماره متناظر با ردهٔ چهره	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
تعداد تشخیص درست	۵	۶	۶	۵	۵	۶	۶	۶	۶	۵
شماره متناظر با ردهٔ چهره	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
تعداد تشخیص درست	۶	۶	۴	۶	۲	۴	۲	۴	۵	۶

شد، در اینجا برای داده‌های هر رده جداگانه به‌کار می‌رود. مجموعهٔ ارقام هُدی، اولین مجموعهٔ بزرگ ارقام دست‌نویس فارسی است. این مجموعه متشکل از ۱۰ رده با ارقام سیاه و سفید ۰–۹ است که تعداد ارقام آموزشی و آزمایشی موجود برای هر رده از آن به‌ترتیب، ۶۰۰۰ و ۲۰۰۰ نمونه است.



شکل ۴. نمونه ارقام دست‌نویس هُدی

برای بررسی عددی، مجموعهٔ ارقام آموزشی با ۱۰۰۰ نمونه و مجموعهٔ ارقام آزمایشی با ۲۰۰ نمونه (۲۰ نمونه از هر رده) و با تبدیل اندازه به ۱۶ × ۱۶ انتخاب شده‌اند. ماتریس A_i متناظر با ردهٔ ارقام هم‌نوع تشکیل و مسئلهٔ $\rho_i = \min_x \|A_i x - d\|_2^2$ برای شناسایی رقم آزمایشیِ نامعلومِ $d \in \mathbb{R}^{256}$ حل می‌شود. به این ترتیب، رقم d به ردهٔ iام تخصیص خواهد یافت. اما می‌توان از روش NMF برای تجزیهٔ ماتریس A_i استفاده کرد. لذا برای رده‌بندی رقم آزمایشی d کافی است مسئلهٔ کوچکتر زیر حل شود:

$$\rho_i = \min_y \|W_i y - d\|_2^2, \qquad y = H_i x, \quad i = 1, \ldots, 10. \tag{۲.۳}$$

نتایج حاصل از رده‌بندی ارقام با استفاده از روش NMF با رتبهٔ ماتریسی $r = 10$ در جدول ۳ ارائه شده که بر اساس آن، میزان موفقیت این روش ۹۳ درصد است. اما میزان تشخیص درست با استفاده از روش SVD در رده‌بندی ارقام [۱۸]، برابر با ۹۳٫۵ درصد است. با توجه به این نتیجه، شناسایی ارقام توسط روش NMF علاوه بر تفسیر صحیح، دارای دقت عملکرد قابل مقایسه‌ای با روش SVD است.

جدول ۳. نتایج رده‌بندی ارقام با روش NMF

ردهٔ ارقام	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
تعداد تشخیص درست	۱۹	۲۰	۱۹	۲۰	۱۸	۱۹	۱۷	۱۹	۲۰	۱۵

بحث و نتیجه‌گیری

ارائهٔ یک مدل کاهشیِ مناسب برای نمایش داده‌های نامنفی، برای درک و تفسیر آنها اهمیت دارد. NMF برخلافِ روش‌هایی چون SVD و PCA، بردارهای وزن و پایه‌ای را همراه با قیدهای نامنفی به‌دست می‌آورد. در تقریب ماتریسی، این رویداد برای سایر روش‌ها به‌دلیل متعامد بودنِ بردارهای پایه‌ای و در نتیجه احتمال حضور درآیه‌های منفی اتفاق نمی‌افتد. روش NMF با توجه به قیدهای نامنفی می‌تواند نمایشی مبتنی بر اجزاء را به‌خوبی معنا بخشد. لذا این روش نسبت به سایر روش‌های نامبرده، به تغییرات جزئی داده‌ها حساس است. هرچند دستگاه‌های رایانه‌ای، تفسیر داده‌های با بُعد بالا را تسهیل می‌کنند، اما به هیچ وجه، انعطاف‌پذیری و شهود انسان را ندارند. به این معنی که ویژگی(های) مجموعهٔ داده‌ها موجب می‌شوند که عملکرد روش‌ها برای آنها متفاوت باشد. مثلاً مشاهده شد که دقت روش NMF بیشتر از دقت روش‌های دیگر در کاربردهای مذکور نیست. از جمله بارزترین دلایل این رخداد، به‌کارگیریِ معیاری مانند فاصلهٔ اقلیدسی در این کاربردها است. هرچند نمی‌توان هیچ ملاک خاصی را در برتری معیارها نسبت به یکدیگر بیان کرد، اما می‌توان معیارهای جایگزین مناسب‌تری را نسبت به فاصلهٔ اقلیدسی در نظر گرفت که برای مطالعهٔ این جایگزین‌ها، مطالعهٔ مرجع [۲۵] پیشنهاد می‌شود. بارزترین مشکل NMF، عدم تضمین همگرایی به نقطهٔ کمینهٔ سراسری است. چون مسئلهٔ NMF محدب نیست، انتظار می‌رود چندین نقطهٔ کمینهٔ موضعی موجود باشد و لذا به‌دلیلِ عدم یکتایی جواب، نرمال کردن ستون‌های W در الگوریتم‌ها مطلوب به‌نظر می‌رسد. در الگوریتم‌های تکراری NMF چندین و چند تکرار لازم است تا همگرایی به نقطهٔ بهین اتفاق افتد که این امر موجب می‌شود تا NMF در مقایسه با PCA زمان بیشتری را صرف کند. اما چون $r \ll \min\{m, n\}$، این امیدواری وجود دارد که این روش نیازمند فضای ذخیره‌سازی کمتری باشد.

چگونگیِ انتخاب مقدار r در PCA بر اساس بزرگی مقادیر ویژه قابل تعیین است. اما در حالت کلی، هیچ معیار مشخصی برای تعیین مقدار بهین آن برای NMF در دسترس نیست. غالباً این مقدار در کاربردهایی نظیر کاربردهای بیان شده در بخش قبل به‌عنوان تعداد خوشه و یا تعداد رده، از قبل مشخص است. اما در حالت کلی، نحوهٔ انتخاب بهین این مقدار به‌صورت یک مسئلهٔ باز در حال مطالعه است. در مرجع [۲۵] به یکی از رهیافت‌های مورد بررسی در این زمینه پرداخته شده است.

قدردانی

نویسندگان بر خود لازم می‌دانند از داوران گرامی درارائهٔ دیدگاهشان، سردبیر محترم و سرکار خانم صمدیان برای پیگیری امور چاپ مقاله قدردانی نمایند.

مراجع

[1] Benthem, M. H. V., Keenan, M. R., Fast algorithm for the solution of large-scale nonnegativity-constrained least squares problems, *J. Chemometrics*, **18** (2004), 441-450.

[2] Berry, M. W., Browne, M., Langville, A. N., Pauca, V. P., Plemmons, R. J., Algorithms and applications for approximate nonnegative matrix factorization, *Computatinal Statistics and Data Analysis*, **1** (2007),155-173.

[3] Bertsekas, D. P., *Nonlinear Programming*, 2nd. edn., Athena Scientific, Belmont, Massachusetts, 1999.

[4] Boutsidis C., Gallopoulos E., SVD-based initialization: A head start for nonnegative matrix factorization, *Pattern Recognition*, **41** (2008), 1350-1362.

[5] Cichocki, A., Zdunek, R., Phan, A. H., Amari, S., *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, John Wiley & Sons, New York, 2009.

[6] Golub, G. H., Van Loan, C. F., *Matrix Computations*, 3rd. edn., Johns Hopkins University Press, Bartimore and London, 1996.

[7] Gonzales, E. F., Zhang, Y., *Accelerating the Lee-Seung Algorithm for Nonnegative Matrix Factorization*, Rice University, 2005.

[8] Kim, H., Park, H., Nonnegative matrix factorization based on alternating non-negativity-constrained least squares and the active set method, *SIAM J. Matrix Anal. Appl.*, **30(2)** (2008), 713-730.

[9] Kim, H., Park, H., Sparse non-negative matrix factorizations via aternating non-negativity-constrained least squares for microarray data analysis, *Bioinformatics*, **23** (2007), 1495-1502.

[10] Kim, J., Park H., Toward faster nonnegative matrix factorization: A new algorithm and comparisons, *Proc. 8th IEEE Int. Conf. Data Mining*, 2008, 353-362.

[11] Kim, J., Park H., Fast nonnegative matrix factorization: An active-set-like method and comparisons, *SIAM Journal on Scientific Computing (SISC)*, **33** (2011), 3261-3281.

[12] Elden, L., *Matrix Methods in Data Mining and Pattern Recognition*, Fundamentals of Algorithms, Society for Industrial and Applied Mathematics, 2007.

[13] Lawson, C. L., Hanson, R. J., *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[14] Lee, D. D., Seung, H. S., Learning the parts of objects by non-negative matrix factorization, *Nature*, **401** (1999), 788-791.

[15] Lee, D. D., Seung, H. S., Algorithms for non-negative matrix factorization, *Advances in Neural Information Processing*, **13**, 2001.

[16] Lin, C. J., On the convergence of multiplicative update algorithms for non-negative matrix factorization, *IEEE Transactions on Nueral Networks*, **18(6)** (2005), 1589-1596.

[17] Lin, C. J., Projected gradient methods for nonnegative matrix factorization, *Neural Computation*, **19(10)** (2005), 2756-2779.

[18] Mazack, M. J. M., *Non-negative Matrix Factorization with Applications to Handwritten Digit Recognition*, Working paper, University of Minnesota, 2009.

[19] Paatero P., Tapper U., Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values, *J. Environmetrics*, **5** (1994), 111-126.

[20] Rasmus, B., Sijmen, D. J., A fast non-negativity-constrained least squares algorithm, *Journal of Chemometrics*, **11(5)** (1997), 393-401.

[21] Skillicorn, D. B., *Understanding Complex Datasets: Data Mining with Matrix Decompositions*, Chapman & Hall/CRC, 2007.

[22] Tan, P. N., Steinbach, M., Kumar, V., *Introduction to Data Mining*, Addison-Wesley, Boston, 2005.

[23] Turk, M., Pentland, A., Eigenfaces for recognition, *Journal of Cognitive Neuroscience*, **3(1)** (1991), 71-86.

[24] Wall, M. E., Rechtsteiner, A., Rocha, L. M., Singular Value Decomposition and Principal Component Analysis, in: *A Practical Approach to Microarray Data Analysis*, Berrar, D. P., Dubitzky, W. and Granzow, M. (editors), 2003, 91-109.

[25] Yun, X., *Non-negative Matrix Factorization for Face Recognition*, Hong Kong Baptist University, Ph.D. Thesis, 2007.

مهسا یوسفی: دانشگاه صنعتی سهند تبریز، گروه ریاضی کاربردی
رایانامه: mahsa.yousefi1987@gmail.com , m_yousefi@sut.ac.ir

منصور رزقی: دانشگاه تربیت مدرس، گروه علوم کامپیوتر
رایانامه: rezghi@modares.ac.ir