بسم الله الرحمن الرحیم

دانشکده ریاضی و علوم کامپیوتر

پایان‌نامه برای دریافت درجه کارشناسی ارشد در رشته علوم تصمیم و مهندسی دانش

تجزیه نامنفی ماتریس در خوشه بندی داده‌ها

استاد راهنما

دکتر محمود امین طوسی

استاد مشاور

دکترامین رفیعی

پژوهشگر:

الهام سنچولی

بهمن ۱۳۹۶

باسمه تعالی

فرم ارزشیابی و صورتجلسه دفاع از پایان‌نامه کارشناسی ارشد

فرم ۱۱۳-ت

جلسه دفاع از پایان‌نامه آقای /خانم **الهام سنچولی** دانشجوی رشته علوم تصمیم و مهندسی دانش به شماره دانشجویی ۹۴۱۳۱۳۷۰۵۱ با عنوان:

تجزیه نامنفی ماتریس در خوشه بندی داده‌ها

در مورخه ..... در دانشکده ریاضی و علوم کامپیوتر تشکیل و توسط هیات داوران مورد ارزشیابی قرار گرفت و نمره ..... برابر درجه ..... برای آن تعیین گردید .

به‌این ترتیب از این تاریخ آقای/ خانم الهام سنچولی به عنوان کارشناس ارشد در رشته مذکور شناخته می‌شود .

موارد ارزشیابی	موارد	حداکثر نمره	نمره کسب شده
۱- کیفیت نگارش	رعایت اصول نگارش انسجام در تنظیم بخشهای مختلف ،کیفیت تصاویر ، جداول و اشکال ،تنظیم فهرست ها ، منابع و ماخذ	٤	
۲- کیفیت علمی	بررسی تاریخچه و سابقه تجربی و نظری موضوع انسجام منطقی دربخش های مختلف پایان نامه ،ابتکار و نوآوری ،اهمیت و ارزش علمی پایان نامه ،استفاده از منابع معتبر و جدید ،کیفیت تجزیه و تحلیل یافته ها و نتیجه گیری ،روشن بودن روش کار ،هدف ها و فرضیه های تحقیق ،جدید بودن روش تحقیق	۱۰	
۳-کیفیت ارایه در جلسه دفاع	تسلط بر موضوع و بیان واضح وتفهیم آن ،توانایی در پاسخگویی به سوالات مطرح شده در جلسه ، رعایت زمان ارایه , روش ارائه	٤	
٤-ارزشیابی گزارشات	گزارش های دوره ای پیشرفت کار(حداقل ٤مورد)	۱	
٥- خروجی پایان نامه	مقاله مستخرج از پایان نامه: این نمره به صورت زیر اختصاص می یابد ۱) چکیده کنفرانسی هر مورد ۰/۲۵ نمره تا سقف۰/۵ نمره ۲) مقاله کامل در مجموع مقالات همایشهای معتبر یا مقاله در مجلات علمی-ترویجی معتبر پذیرفته شده یا چاپ شده هر مورد ۰/۵ نمره تا سقف ۱نمره ۳)مقاله پذیرفته شده یا چاپ شده در مجلات علمی پژوهشی معتبر ۱ نمره ٤)مقاله ارسال شدهبه مجلات علمی پژوهشی معتبرهر مورد ۲۵/. نمره تا سقف ۰/۵ نمره ٥) دستگاه ساخته شده دارای گواهی ثبت اختراع یا به سفارش سازمان ها تا سقف ۱ نمره ٦) دستگاه ساخته شده کاربردی که به تایید رئیس دانشکده رسیده باشد تا سقف ۰/۵ نمره	۱	
جمع			

درجه معادل کسب شده: (از ۱۹تا ۲۰ عالی☐ از۱۸تا ۱۸/۹۹بسیارخوب ☐ از ۱٦تا ۱۷/۹۹خوب ☐ از ۱٤ تا ۱۵/۹۹قابل قبول☐ کمتر از ۱٤ غیر قابل قبول☐)

مشخصات هیات دوران

ردیف	نام و نام خانوادگی	سمت	مرتبه علمی	محل کار	امضـــــــــــا
۱	دکتر محمود امین طوسی	استاد راهنما	استادیار	دانشگاه حکیم سبزواری	
۲	دکترامین رفیعی	استاد مشاور	استادیار	دانشگاه حکیم سبزواری	
۳	...	استاد داور	استادیار	دانشگاه حکیم سبزواری	
۴	دکتر	نماینده تحصیلات تکمیلی	استادیار	دانشگاه حکیم سبزواری	

امضا

مدیر گروه

امضا

رئیس دانشکده

سوگند نامه دانش آموختگان دانشگاه حکیم سبزواری

به نام خداوند جان و خرد کزین برتر اندیشه بر نگذرد

اینک که به خواست آفریدگار پاک، کوشش خویش و بهره گیری از دانش استادان و سرمایه‌های مادی و معنوی این مرز و بوم، توشه‌ای از دانش و خرد گردآورده‌ام، در پیشگاه خداوند بزرگ سوگند یاد می‌کنم که در به کارگیری دانش خویش، همواره بر راه راست ودرست گام بردارم. خداوند بزرگ، شما شاهدان، دانشجویان و دیگر حاضران را به عنوان داورانی امین گواه می گیرم که ازهمه دانش و توان خود برای گسترش مرزهای دانش بهره گیرم و از هیچ کوششی برای تبدیل جهان به جایی بهتر برای زیستن، دریغ نورزم. پیمان می‌بندم که همواره کرامت انسانی را در نظر داشته باشم و همنوعان خود را در هر زمان و مکان تا سر حد امکان یاری دهم. سوگند می خورم که در به کارگیری دانش خویش به کاری که با راه و رسم انسانی، آیین پرهیزگاری، شرافت و اصول اخلاقی برخاسته از ادیان بزرگ الهی، به ویژه دین مبین اسلام، مباینت دارد دست نیازم. همچنین در سایه اصول جهان شمول انسانی و اسلامی، پیمان می‌بندم از هیچ کوششی برای آبادانی و سرافرازی میهن و هم میهنانم فروگذاری نکنم و خداوند بزرگ را به یاری طلبم تا همواره در پیشگاه او و در برابر وجدان بیدار خویش و ملت سرافراز، بر این پیمان تا ابد استوار بمانم.

نام و نام خانوادگی: الهام سنچولی

تاریخ و امضا:

تأییدیه‌ی صحت و اصالت نتایج

باسمه تعالی

اینجانب الهام سنچولی به شماره دانشجویی ۹۴۱۳۱۳۷۰۵۱ دانشجوی رشته علوم تصمیم و مهندسی دانش مقطع تحصیلی کارشناسی ارشد تأیید می‌نمایم که کلیه‌ی نتایج این پایان‌نامه حاصل کار اینجانب و بدون هرگونه دخل و تصرف است و موارد نسخه برداری شده از آثار دیگران را با ذکر کامل مشخصات منبع ذکر کرده‌ام. درصورت اثبات خلاف مندرجات فوق، به تشخیص دانشگاه مطابق با ضوابط و مقررات حاکم (قانون حمایت از حقوق مؤلفان و مصنفان و قانون ترجمه و تکثیر کتب و نشریات و آثار صوتی، ضوابط و مقررات آموزشی، پژوهشی و انضباطی ...) با اینجانب رفتار خواهد شد و حق هرگونه اعتراض درخصوص احقاق حقوق مکتسب و تشخیص و تعیین تخلف و مجازات را از خویش سلب می‌نمایم. در ضمن، مسؤولیت هرگونه پاسخگویی به اشخاص اعم از حقیقی و حقوقی و مراجع ذی صلاح (اعم از اداری و قضایی) به عهده ی اینجانب خواهد بود و دانشگاه هیچ گونه مسؤولیتی در این خصوص نخواهد داشت.

نام و نام خانوادگی: الهام سنچولی

تاریخ و امضا:

مجوز بهره برداری از پایان‌نامه

بهره برداری از این پایان‌نامه در چهارچوب مقررات کتابخانه و با توجه به محدودیتی که توسط استاد راهنما به شرح زیر تعیین می‌شود، بلامانع است:

☐ بهره برداری از این پایان‌نامه برای همگان بلامانع است.

☐ بهره برداری از این پایان‌نامه با اخذ مجوز از استاد راهنما، بلامانع است.

☐ بهره برداری از این پایان‌نامه تا تاریخ ..... ممنوع است.

استاد راهنما: دکتر محمود امین طوسی

تاریخ و امضا:

تقدیم :

با بوسه بر دستان
مادر شفیق و پدر مهربانم

همیشه دستان ربنایم را به سوی کسی می گیرم که مهربانی نگاهش را باور کرده ام، و امروز که اینجایم و از بی کران دانش خدا، به قدر بضاعتم آموخته ام، تمام قامت می ایستم و بلندترین شکرانه ام را قنوت می شوم.

و قدم های مادرم که یادم دادند، نماندن را، دویدن را و رسیدن را، می بوسم.

و از پدرم ممنونم که روی پا ایستادنم را قانون کرد.

از استاد راهنمای عالمم جناب آقای دکتر محمود امین طوسی، که صبورانه زکات داناییشان را برایم کنار گذاشتند تا جا نمانم از دانستن، بی حساب ممنونم.

همچنین از استاد مشاور ارجمندم، جناب آقای دکتر امین رفیعی، که قبول زحمت فرمودند، سپاسگزارم.

مراتب قدر دانی خود را از داور محترم این پایان نامه، جناب آقای دکتر..... ، اعلام می کنم.

هر چقدر از سپاس می شناسم را هدیه می کنم به خواهران بی دریغ مهربانم و برادران همیشه خوبم که بی نظیرند در همراهی.

در آخر از خدای یگانه ام، برای تمامی چشم های آشنا و ناآشنایی که کوچه های رفتنم، به حضور روشنشان نور گرفته است، یک آسمان بهار طلب می کنم.

الهام سنچولی

بهمن ۱۳۹۶

فهرست مطالب

فهرست جداول د

فهرست تصاویر ه

چکیده ۱

پیش‌گفتار ۲

فصل ۱: مفاهیم و تعاریف اولیه ۳

۱-۱ مقدمه . . . ۳

۱-۲ مفاهیم مورد نیاز ماتریس . . . ۳

۱-۲-۱ اَعمال ماتریسی . . . ۴

۱-۲-۲ زیرفضاهای برداری . . . ۶

۱-۳ تعامد بردارها و ماتریس ها . . . ۸

۱-۴ نُرم . . . ۱۰

۱-۴-۱ انواع نرم . . . ۱۰

۱-۵ نرم‌های ماتریسی . . . ۱۱

۱-۶ مساله حداقل مربعات . . . ۱۳

۱-۶-۱ مقدارهای ویژه و بردارهای ویژه . . . ۱۴

۱-۷ تجزیه ماتریسی . . . ۱۵

۱-۷-۱ کاهش ابعاد . . . ۱۶

۱-۸ روش تجزیه PCA . . . ۱۸

۱-۹ تحلیل تفکیک خطی LDA . . . ۱۹

۱-۹-۱ بررسی مدل LDA . . . ۲۰

۱-۱۰ تجزیه مقدار منفرد SVD . . . ۲۲

۱-۱۱ تقریب مرتبه پایین ۲۴

فصل ۲: تجزیه نامنفی ماتریس **۲۶**

۲-۱ مقدمه ۲۶

۲-۱-۱ پیشینه تحقیقات انجام شده ۲۶

۲-۲ بیان مسأله ۲۷

۲-۳ کاربردها ۲۹

۲-۳-۱ کاهش بعد ۳۰

۲-۳-۲ کاربرد در پردازش تصویر ۳۰

۲-۳-۳ کاربرد در پردازش متن ۳۱

۲-۳-۴ کاربرد در پردازش تصاویر فراطیفی ۳۲

۲-۴ الگوریتم هایی برای تجزیه NMF ۳۴

۲-۴-۱ الگوریتم قوانین به روز رسانی ضربی و کمترین مربعات غیرمنفی ۳۵

۲-۵ تجزیه های مربوط به NMF ۳۷

۲-۶ NMF برای خوشه بندی ۴۱

۲-۶-۱ خوشه بندی NMF و K-means ۴۱

۲-۷ خوشه بندی K-means کرنلی و NMF متقارن ۴۲

۲-۷-۱ تجزیه نامنفی K-means کرنلی ۴۴

۲-۸ خوشه بندی طیفی و NMF ۴۶

۲-۸-۱ قضیه محدودیت NMF ۴۸

۲-۹ قابلیت های خوشه بندی NMF ۴۹

۲-۹-۱ مثالها ۴۹

۲-۹-۲ تجزیه و تحلیل ۵۰

فصل ۳: برخی الگوریتم های مربوط به NMF **۵۲**

۳-۱ مقدمه ۵۲

۳-۲ مسائل کاربردی در الگوریتم NMF ۵۲

۳-۲-۱ مقدار دهی اولیه ۵۲

۳-۲-۲ معیارهای توقف ۵۳

۳-۲-۳ روشهای تعیین طول گام ۵۴

۳-۲-۳-۱ روش کاهش گرادیان ۵۴

۳-۲-۳-۲ روش نیوتن ۵۵
۳-۳ الگوریتم های مربوط به NMF ۵۶
۳-۳-۱ قواعد به هنگام ضربی ۵۶
۳-۳-۲ روش کمترین مربعات متناوب ۵۸
۳-۳-۳ الگوریتم FC-NNLS ۶۱

فصل ۴: آزمایشات ونتایج **۶۴**
۴-۱ روش NMF پیشنهادی ۶۴
۴-۱-۱ تعیین طول گام ۶۵
۴-۲ انتخاب مقدار r ۶۵
۴-۳ نمونه برداری ۶۶
۴-۴ مجموعه داده های مورد استفاده ۶۷
۴-۴-۱ بانک داده ORL ۶۷
۴-۴-۲ مجموعه داده iris ۶۸
۴-۵ آزمایشات و نتایج ۶۸
۴-۵-۱ مقایسه روش SVM و روش NMF روی پایگاه داده ORL ۶۸
۴-۵-۲ مقایسه روش SVM و روش NMF روی پایگاه داده iris ۶۹
۴-۵-۳ روش NMF پیشنهادی ۷۰
۴-۵-۴ روشی دیگر برای تجزیه NMF ۷۰

فهرست منابع **۷۲**

پیوست آ: برنامه های MATLAB روش های مطرح شده در پایان نامه **۷۵**

واژه‌نامه فارسی به انگلیسی **۸۴**

واژه‌نامه انگلیسی به فارسی **۸۵**

فهرست جداول

۲-۱ ماتریس ورودی V ۳۳
۲-۲ ماتریس W حاصل از تجزیه ماتریس V ۳۳
۲-۳ ماتریس H حاصل از تجزیه ماتریس V ۳۳

۴-۱ مشخصات بانک داده ORL ۶۷
۴-۲ روش SVM روی پایگاه داده ORL ۶۸
۴-۳ روش NMF اولیه روی پایگاه داده ORL ۶۹
۴-۴ روش SVM روی پایگاه داده iris ۶۹
۴-۵ روش NMF اولیه روی پایگاه داده iris ۶۹
۴-۶ روش NMF پیشنهادی روی پایگاه داده ORL ۷۰
۴-۷ روش NMF جدید روی پایگاه داده ORL ۷۱

فهرست تصاویر

۱-۱ تبدیل داده از از قالب ماتریس به بردار ۱۶
۱-۲ تبدیل PCA ۱۸
۱-۳ تفکیک کلاس ها ۲۱
۱-۴ افکنش کلاس ها ۲۱
۱-۵ تجزیه SVD ۲۳
۱-۶ تجزیه SVD کوچک شده ۲۴

۲-۱ ضرب ماتریس وزن با ماتریس ویژگی ۲۸
۲-۲ عدم انطباق NMF ۲۹
۲-۳ نمونه تصاویر بانک ORL ۳۶
۲-۴ اعمال NMF به پایگاه داده ORL ۳۶
۲-۵ تجزیه NMF برای دواجرا با شروع اولیه ی تصادفی روی داده های تصویری ۴۹
۲-۶ تجزیه NMF برای دو اجرا با شروع اولیه تصادفی برای داده های عددی ۵۰
۲-۷ مجموعه داده دو بعدی از ۳۸ نقطه داده و خوشه بندی انحصاری و هم پوشی روی آنها ۵۱

۴-۱ نمونه تصاویر بانک ORL ۶۷

فرم چکیده ی پایان‌نامه ی دوره ی تحصیلات تکمیلی

مدیریت تحصیلات تکمیلی

دانشگاه حکیم سبزواری

نام خانوادگی دانشجو:سنچولی	نام:الهام	ش. دانشجویی:۹۴۱۳۱۳۷۰۵۱
استاد راهنما: دکتر محمود امین طوسی		
استاد مشاور: دکترامین رفیعی		
دانشکده ریاضی و علوم کامپیوتر	رشته: علوم تصمیم و مهندسی دانش	
مقطع: کارشناسی ارشد	تاریخ دفاع:بهمن ۱۳۹۶	تعداد صفحات: ۸۶
عنوان پایان‌نامه: تجزیه نامنفی ماتریس در خوشه بندی داده‌ها		
کلید واژه‌ها:تجزیه نامنفی ماتریس، خوشه بندی ،کاهش بعد، استخراج ویژگی		

چکیده: یکی از مهمترین مباحث پایه ای در زمینه های آمار، یادگیری ماشین، شناسایی الگو و داده کاوی [۱] طبقه بندی می‌باشد. در الگوریتم های طبقه بندی مجموعه داده اولیه به دو مجموعه داده با عنوان مجموعه داده های آموزشی [۲] و مجموعه داده های آزمایشی [۳] تقسیم می‌شود، با استفاده از مجموعه داده های آموزشی مدل ساخته می‌شود و از مجموعه داده آزمایشی برای اعتبار سنجی و محاسبه دقت مدل ساخته شده، استفاده می‌شود. در الگوریتم های طبقه بندی چون ویژگی طبقه مربوط به هر رکورد مشخص است بنابراین جزء الگوریتم های با نظارت محسوب می‌شوند. الگوریتم های با نظارت شامل دو مرحله با عنوان آموزش (یادگیری) و مرحله ارزیابی هستند. یکی از روش‌های مورد استفاده برای طبقه‌بندی روش تجزیه نامنفی ماتریس است. تجزیه نامنفی ماتریس(NMF) از موضوعات به روز در زمینه جبر خطی است که برای تشخیص الگو، تحلیل داده ها و کاهش بعد به کار می رود. در این حوزه هدف، تجزیه ماتریس A شامل داده های نامنفی به حاصل ضرب ماتریس پایه U و ماتریس ضرایب V^T به صورت $A = UV^T$ با درایه های نامنفی است. در حالت کلی نیاز به تحلیل و استخراج ویژگی ما را خود به خود به سمت کاهش بعد و فشرده سازی سوق داده است. این عمل برای کاهش هزینه ذخیره سازی اطلاعات مطلوب است. در این پایان نامه روشهای مختلف موجود برای NMF مورد بررسی قرار گرفته است. استفاده از آن، بسیاری از ویژگی های داده های اصلی را حفظ می کند و تضمین می نماید که هم ماتریس پایه و هم ماتریس ضرایب نامنفی بمانند. به ویژه در مورد تصاویر افراد، تجزیه NMF برای نشان دادن ویژگی های اصلی صورت مانند چشم ها، گونه ها و لب ها به کار رفته است.

---

[۱] Data Mining [۲] Train Dataset [۳] Test Dataset

Data Mining [۱] Train Dataset [۲] Test Dataset [۳]

پیش‌گفتار

اساس مشترک رهیافت های مختلف در تحلیل داده ها، پیدا کردن یک مدل مناسب برای نمایش آنها است تا بینش و تصوری صحیح از آنها به وجود آید. رهیافت بسیار معمولی که آن را مدل کاهشی [۱] می نامیم، مدلی است که سعی می کند از پیچیدگی داده های اولیه و با ابعاد بالا بکاهد و ضمن حفظ ملزومات مسئله همراه با داده ها، ساختار پنهان آنها را آشکار کند. در واقع، این مدل ضمن کاهش داده ها، در سطحی نزدیک تر (واقعی تر) به سیستم اولیه قرار دارد. مدل های کاهشی می توانند خطی و یا غیرخطی باشند که در این پایان نامه، مدل های خطی مورد مطالعه قرار می گیرند. در این پایان نامه، سه مدل کاهشی خطی به نام های تجزیه مقدار تکین (SVD) [۲] تحلیل مؤلفهٔ اصلی (PCA) [۳]، و تجزیهٔ نامنفی ماتریسی (NMF) [۴] بیان می شوند. در فصل اول ابتدا برخی مفاهیم اولیه مورد نیاز ارائه شده و سپس تجزیه های تحلیل مولفه اصلی، (SVD) و (LDA) به اختصار شرح داده خواهد شد. در فصل دوم تجزیه نامنفی ماتریس را تعریف کرده برخی از انواع مختلف آن را به اختصار شرح داده و هم ارزی این تجزیه را با برخی الگوریتم های خوشه بندی نشان خواهیم داد. در فصل سوم برخی از پرکاربرد ترین و مهم ترین الگوریتم های مربوط به تجزیه نامنفی ماتریس را معرفی خواهیم کرد. فصل چهارم شامل آزمایشات انجام شده و نتایج این پایان نامه خواهد بود.

مقالات اصلی مورد استفاده در این پایان نامه عبارتند از:

title : On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering

author : Ding, C.,and He, X., and D.simon, H.

journal : SIAM international conference on Data Mining, SDM

year: 2005,

pages: 606-610

title : Two algorithms for orthogonal nonnegative matrix factorization

author : Pompili, F.,and Gillis, N.,and Absil, P.-A.,and Glineur, F

journal : Neurocomputing 141

year : 2014

pages : 15-25

---

[۱] reduction model [۲] singular value decomposition [۳] principle component analysis [۴] nonnegative matrix factorization

فصل ۱

مفاهیم و تعاریف اولیه

۱-۱ مقدمه

در این فصل ابتدا به مفاهیم و تعاریف مورد نیاز این پایان نامه از قبیل زیرفضاهای برداری، اعمال ماتریسی، نرم و متعامد سازی پرداخته، سپس مساله حداقل مربعات مطرح می شود. در انتها تعریفی از تجزیه ی ماتریس و برخی از انواع آن بیان خواهد شد.

۱-۲ مفاهیم مورد نیاز ماتریس

ماتریس حقیقی $m \times n$ یک جدول m سطری و n ستونی از اعداد حقیقی است. در صورت تساوی تعداد سطرها و ستون های یک ماتریس به آن ماتریس مربعی اطلاق می شود. مجموعه تمام ماتریس های حقیقی $m \times n$ با $\mathbb{R}^{m\times n}$ نشان داده می شود.در این پایان نامه از حروف بزرگ برای نمایش ماتریسها استفاده می شود. سطر iام ماتریس A با A_{ij} یا $[A]_{ij}$ نشان داده می شود.

یک بردار ستونی، ماتریسی شامل تنها یک ستون است. به طور مشابه هر بردار سطری، ماتریسی شامل تنها یک سطر است. به طور پیش فرض، هر بردار یک ماتریس ستونی است. مجموعه تمام بردارهای با بعد n، با $\mathbb{R}^n$ نشان داده می شود. بردارها اغلب با حروف کوچک مشخص می شوند مگر این که بردار سطر یا ستونی از یک ماتریس باشد.

ماتریس مربعی D ماتریس قطری است، هرگاه تنها قطر اصلی آن شامل عنصر یا عناصر غیر صفر باشد، یعنی برای $i \neq j$ داشته باشیم: $D_{ij} = 0$. همچنین D_x ماتریسی قطری است که بردار x را روی قطر اصلی خود دارد و سایر درایه های آن صفر است.

برخی ماتریسهای خاص دیگر عبارتند از:

- ماتریسی که کلیه عناصر آن برابر ۱ است:

$$۱_{m\times n} = ۱_{m\times ۱} \times ۱_{۱\times n}$$

که در آن $۱_{۱\times n} = (۱, ۱, \ldots, ۱)$ و $۱_{m\times ۱} = (۱, ۱, \ldots, ۱)^T$

- بردارهای یکه:

$$e_i = (۰, ۰, \ldots, \overbrace{۱}^{\text{i امین مکان}}, \ldots, ۰)^T$$

- ماتریس واحد I_n:ماتریس قطری که عناصر قطری آن برابر ۱ است. این ماتریس با

$$I_{ij} = \delta_{ij} = \begin{cases} ۱ & \text{هر گاه } i = j \\ ۰ & \text{هرگاه } i \neq j \end{cases}$$

تعریف می شود که به آن دلتای کرونر اطلاق می شود.

- ماتریس جایگشت: ماتریس مربعی که در هر سطر و ستون آن تنها یک درایه غیر صفر برابر ۱ وجود دارد.

- ماتریس گزینشی: هر زیرماتریسی از ماتریس های جایگشت [۱].

۱-۲-۱ اَعمال ماتریسی

در این قسمت برخی عملگرهای پایه ای ماتریس بیان می شود.

- ترانهاده ماتریس A^T: $[A^T]_{ij} := A_{ij}$

فرض کنید A یک ماتریس $m \times n$ باشد. ترانهاده [۱] ماتریس A با نماد A^T نمایش داده می شود،که یک ماتریس $n \times m$ است که به صورت زیر تعریف می شود [۲].

$$A = (a_{ij})_{۱\leq i\leq m, ۱\leq j\leq n} \iff A^T = (a_{ji})_{۱\leq j\leq n, ۱\leq i\leq m}$$

- جمع ماتریس $C = A + B$: $C_{ij} = A_{ij} + B_{ij}$

- ضرب ماتریس $C_{ij} = \sum_k A_{ik}B_{ik}\ C = A.B = AB$

[۱] Transpose

- نمایش برداری ماتریس $A \in \mathbb{R}^{m\times n}$:

$$\text{vec}(A) = \begin{pmatrix} A_{:1} \\ \vdots \\ A_{:n} \end{pmatrix} \in \mathbb{R}^{mn\times 1}$$

ضرب کرونر ماتریس $A \in \mathbb{R}^{m\times n}$ و ماتریس $B \in \mathbb{R}^{p\times q}$:

$$A \otimes B = \begin{pmatrix} A_{11}B & \dots & A_{1n}B \\ \vdots & \ddots & \vdots \\ A_{m1}B & \dots & A_{mn}B \end{pmatrix} \in \mathbb{R}^{pm\times qn}$$

$A \leq B$ به این معناست که به ازای هر i و j داریم: $A_{ij} \leq B_{ij}$ و $A \leq \alpha$ به ازای $\alpha \in \mathbb{R}$ به معنای $A \leq \alpha 1_{m\times n}$ است. قدر مطلق ماتریس $|A|$ به صورت $[|A|]_{ij} = |A_{ij}|$ برای هر i و j تعریف می شود.

ماتریس مربعی ،A معکوس پذیر نامیده می شود هر گاه ماتریس B چنان موجود باشد که:

$$AB = BA = I$$

ماتریس B معکوس ماتریس A نامیده می شود و با $B = A^{-1}$ نشان می دهیم. هرچند تمام ماتریس ها معکوس پذیر نیستند.

تعریف ۱-۲-۱. ضرب هادامارد [1] : در این روش دو ماتریس باید ابعاد یکسانی داشته باشند. ضرب هادامارد دو ماتریس A و B را به صورت $A \odot B$ نمایش می دهند ماتریس حاصل ماتریسی به همان ابعاد است که در آن هر درایه از حاصلضرب دو درایه متناظر در A و B بدست می آید.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \odot \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} & a_{13}b_{13} \\ a_{21}b_{21} & a_{22}b_{22} & a_{23}b_{23} \\ a_{31}b_{31} & a_{32}b_{32} & a_{33}b_{33} \end{pmatrix}$$

[1] Hadamard

ضرب کرونر ماتریس $A \in \mathbb{R}^{m\times n}$ و ماتریس $B \in \mathbb{R}^{p\times q}$:

$$A \otimes B = \begin{pmatrix} A_{11}B & \dots & A_{1n}B \\ \vdots & \ddots & \vdots \\ A_{m1}B & \dots & A_{mn}B \end{pmatrix}$$

تعریف ۱-۲-۲. تانسور: تانسور عنصری هندسی است که در ریاضی و فیزیک به منظور گسترش مفاهیم اسکالرها، بردارها و ماتریس ها به ابعاد بالاتر معرفی می شوند. تانسور آرایه ای است از اعداد که در یک جدول چیده شده اند. این جدول در حالت کلی می تواند به صورت $N \times M \times O \times P \times \dots$ باشد که حروف بزرگ هر کدام می توانند نمایندهٔ یک عدد طبیعی باشند و $\times$ نشان دهنده عمل ضرب بین آنهاست. تانسور در ساده ترین حالت می تواند یک عضو داشته باشد که به آن تانسور، اسکالر گوییم. در حالت کمی پیشرفته تر تانسور می تواند به صورت بردار باشد. یعنی وقتی شما بردار A را به صورت (x, y, z) نشان می دهید در حقیقت یک تانسور دارید. در حالتی باز هم پیشرفته تر تانسور می تواند دو بعدی باشد (به صورت ماتریسی). یعنی مثلاً جدول ما ۲ × ۲ باشد یعنی دو سطر و دو ستون داشته باشد. چنین تانسوری دارای ۴ عضو است. به طور کلی تانسورهای دو بعدی و بالاتر از دو بعد را با نام ماتریس هم می شناسند. ماتریس ها از آن جهت مورد استفاده قرار می گیرند که باعث ایجاد نظم بین داده های یک مسئله و دسته بندی اطلاعات آن می شوند.

۱-۲-۲ زیرفضاهای برداری

یک زیرمجموعه غیر تهی Eاز فضای برداری $\mathbb{R}^n$ زیرفضایی از $\mathbb{R}^n$ است اگر و تنها اگر به ازای هر a و b در E و اسکالر c بردارهای $ca + b$ درE باشد. هر گاه مجموعهE یک زیر فضای خطی از $\mathbb{R}^n$ باشد، به وسیله مجموعه تمام ترکیبات خطی بردارهای $\{v_1, v_2, \dots, v_k\}$ از $\mathbb{R}^n$ به صورت زیر نوشته می شود:

$$E = \{\sum_{i=1}^{k} \alpha_i v_i | \alpha_i \in \mathbb{R}\}$$

E گسترش یافته V و V مجموعه مولد E نامیده می شود. با فرض یک زیر فضای E، مجموعه های متعددی برای آن وجود دارد. مجموعه ای که هیچ برداری از آن نمی تواند بدون تغییر E حذف شود، مجموعه مستقل خطی بوده و پایه E نامیده می شود. تعداد پایه های E ثابت بوده و بعد E نامیده می شود.

رتبه ماتریس A از مرتبه $m \times n$ بعد زیر فضای تولید شده توسط ستون های A تعریف می شود:

$$\text{rank}(A) = \dim(\text{span}(A_{:1}, A_{:2}, \dots, A_{:n})) \leq \min(m, n).$$

تعریف ۱-۲-۳. رتبه. رتبه ماتریس $A_{n\times n}$ برابر با ماکزیمم تعداد ستونهای (سطرهای) مستقل خطی در آن ماتریس است،

که با نماد $\text{Rank}(A)$ نشان داده می شود [۳].

یک زیر فضای خطی تحت جمع و ضرب اسکالر بسته است [۱] یعنی:

$$\forall u, v: \quad u, v \in E \Rightarrow u + v \in E,$$

$$\forall u: \quad u \in E, \forall \alpha \in \mathbb{R} \Rightarrow \alpha u \in E.$$

تعریف ۱-۲-۴. (تابع کرنل [۱]). κ یک تابع کرنل است که برای همه $x, z \in X$ رابطه زیر برقرار باشد:

$$\kappa(x, z) =< \phi(x), \phi(z) >$$

مثال ۱. فضای دو بعدی ورودی $x \subseteq \mathbb{R}^2$ را با نگاشت ویژگی زیر در نظر می گیریم:

$$\phi : x \longmapsto \phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2) \in F = \mathbf{R}^3$$

فضای فرضی تابع خطی F به صورت زیر خواهد بود:

$$g(x) = \omega_{11}x_1^2 + \omega_{22}x_2^2 + \omega_{12}\sqrt{2}x_1x_2$$

نگاشت ویژگی، داده ها را از فضای دو بعدی به فضای سه بعدی به گونه ای نگاشت می کند که روابط خطی در فضای ویژگی متناظر با روابط درجه دو در فضای ورودی هستند. ترکیب نگاشت طرح با حاصل ضرب داخلی در فضای طرح می تواند به صورت زیر محاسبه شود:

$$\begin{aligned} < \phi(x), \phi(z) > &=< (x_1^2, x_2^2, \sqrt{2}x_1x_2), (z_1^2, z_2^2, \sqrt{2}z_1z_2) > \\ &= x_1^2z_1^2 + x_2^2z_2^2 + 2x_1x_2z_1z_2 \\ &= (x_1z_1 + x_2z_2)^2 =< x, z >^2 \end{aligned}$$

بنابراین تابع

$$\kappa(x, z) =< x, z >^2$$

یک تابع کرنل با فضای ویژگی F است. این به این معنا است که می توان حاصل ضرب داخلی بین تصویر دو نقطه در فضای ویژگی را بدون محاسبه صریح مختصات آن ها، محاسبه کرد.

[1] Kernel function

تعریف ۱-۲-۵. ترکیب خطی. اگر $\alpha_1, \alpha_2, \ldots, \alpha_n \in F$ و $v_1, v_2, \ldots, v_n \in V$ باشد، آن گاه

$$u = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

را ترکیب خطی $v_1, v_2, \ldots, v_n$ می نامند [۳].

تعریف ۱-۲-۶. پایه. مجموعه بردارهای $v_1, v_2, \ldots, v_n$ در فضای برداری V، برای آن فضا تشکیل یک پایه می دهند اگر:

$$V = \text{sp}\{v_1, v_2, \ldots, v_n\}$$

و $v_1, v_2, \ldots, v_n$ مستقل خطی باشند [۳].

تعریف ۱-۲-۷. بعد [۱]. تعداد بردارهای پایه در یک فضای برداری مانند V را بعد آن فضا می نامند و با نماد $\dim(V)$ نشان می دهند. به عبارتی بعد یک فضا برابر با حداکثر تعداد بردارهای مستقل خطی در آن فضا است. بنابراین در یک فضای n بعدی حداکثر بردارهای مستقل خطی n عدد می باشد [۳].

۱-۳ تعامد بردارها و ماتریس ها

تعریف ۱-۳-۱. دو بردار

$$x = [x_1, x_2, \ldots, x_n]^T \quad , y = [y_1, y_2, \ldots, y_n]^T$$

در $\mathbb{R}^n$ را در نظر بگیرید. ضرب داخلی x, y به صورت زیر تعریف می شود:

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$$

ضرب داخلی روی یک ماتریس به صورت زیر بیان می شود:

$$\langle x, y \rangle = y^T x = x^T y$$

اگر $x, x_1, x_2, y, y_1, y_2, v_2 \in \mathbb{R}^n$ و $a_1, a_2 \in \mathbb{R}$ باشد، آنگاه ویژگی های زیر برای ضرب داخلی برقرار است:

- $\langle x, y \rangle = \langle y, x \rangle$

[۱] dimension

- $\langle a_1 x_1 + a_2 x_2, y\rangle = a_1\langle x_1, y\rangle + a_2\langle x_2, y\rangle$
- $\langle x, a_1 y_1 + a_2 y_2 y\rangle = a_1\langle x, y_1\rangle + a_2\langle x, y_2\rangle$
- $\langle x, x\rangle = 0$ تساوی زمانی برقرار است که $x = 0$
- $\| x \|_2 = \sqrt{\langle x, x\rangle}$

اگر x و y دو بردار غیر صفر باشند می توان فاصله بین دو بردار را از فرمول زیر بدست آورد:

$$\cos(\theta) = \frac{\langle x, y\rangle}{\| x \|_2 \| y \|_2}$$

در شرایطی که $x = 0$ یا $y = 0$ باشد، آنگاه $\theta = \frac{\pi}{2}$ می شود که در این صورت دو بردار x و y را متعامد گویند. واضح است که x و y متعامد هستند اگر و تنها اگر $\langle x, y\rangle = 0$ [۴].

تعریف ۱-۳-۲. مجموعه بردارهای حقیقی $i = 1, 2, \ldots, n, a_i$ را متعامد یکه گوییم هر گاه داشته باشیم:

$$\forall i, j, \qquad a_i^T . a_j = \sigma_{i,j,}$$

فرض کنید $j = 1, 2, \ldots, n, a_j$ متعامد باشند یعنی $a_i^T a_j = 0$ ، $i \neq j$ ، آنگاه a_j ها مستقل خطی هستند [۵].

تعریف ۱-۳-۳. ماتریس $n \times n$، A را متعامد گوییم، هرگاه

$$A^T A = A A^T = I_n \qquad A^{-1} = A^T$$

به عبارت دیگر یک ماتریس مربعی $A = (A_1, A_2, \ldots, A_m) \in \mathbb{R}^{m \times m}$ که ستونهایش متعامد یکه هستند را ماتریس متعامد می گویند.

تعامد ستون ها، نتیجه می دهد که یک ماتریس متعامد، رتبه کامل است و پیدا کردن معکوس آن ساده است. چون A^T نیز متعامد است، بنابراین سطرهای یک ماتریس متعامد نیز یک مجموعه متعامد تشکیل می دهد [۳، ۵].

تعریف ۱-۳-۴. اگر A یک ماتریس متعامد باشد، آنگاه $\det A = 1$ [۳].

تعریف ۱-۳-۵. متعامد. بردار u را بر زیرفضای V_1 متعامد گویند، اگر بردار u بر هر بردار در زیرفضای V_1 متعامد باشد. زیرفضای V_1 و V_2 را متعامد گویند، اگر هر بردار در زیرفضای V_1 با هر بردار در زیرفضای V_2 متعامد باشند [۳].

۱-۴ نُرم

تعریف ۱-۴-۱. تابع حقیقی $\|.\|$ تعریف شده بر فضای برداری V را نُرم نامیم اگر در سه خاصیت زیر صدق کند:

۱. به ازای هر $x \in V$، $\|x\| \geq ۰$ و $\|x\| = ۰$ اگر و فقط اگر $x = ۰$

۲. به ازای هر $x \in V$ و $\alpha \in \mathbb{R}$، $\|\alpha x\| = |\alpha|.\|x\|$

۳. به ازای هر $x, y \in V$ (نابرابری مثلثی) $\|x + y\| \leq \|x\| + \|y\|$

فضای برداری V مجهز به نرم $\|.\|$ را یک فضای برداری نرم‌دار می‌نامیم. از آنجایی که دامنه تعریف نرم، فضایی برداری است، بسته به اینکه اعضای فضای برداری چه باشند، نرم ممکن است برای بردار، ماتریس، یا تابع، تعریف شود. ورودی نرم،عضوهای فضای برداری و خروجی آن عدد حقیقی مثبتی است پس برد هر نرم، مجموعه اعداد حقیقی مثبت می‌باشد. می‌توانیم در مورد $\|x\|$ به عنوان طول یا اندازه بردار x فکر کنیم. یک نرم بر روی یک فضای برداری مفهوم قدر مطلق، $|r|$ ، برای یک عدد حقیقی یا مختلط را تعمیم می‌دهد.

۱-۴-۱ انواع نرم

معروف‌ترین نرم‌ها در $\mathbb{R}^n$ نرم اقلیدسی $\ell_۲$ است که به صورت زیر تعریف می‌شود:

$$\|x\|_۲ = \left(\sum_{i=۱}^{n} x_i^۲ \right)^{\frac{۱}{۲}}$$

که در آن $x = (x_۱, x_۲, \ldots, x_n)^T$. این نرمی است که متناظر با مفهوم ذاتی طول می‌باشد.
نوع دیگر نرم، نرم بی‌نهایت ℓ_∞ است که به صورت زیر تعریف می‌شود:

$$\|x\|_\infty = \max |x_i|$$

سومین نرم مهم در $\mathbb{R}^n$، نرم $\ell_۱$ نامیده می‌شود:

$$\|x\|_۱ = \sum_{i=۱}^{n} \|x_i\|$$

در حالت کلی داریم:

$$\|x\|_p = \left(\sum_{i=۱}^{n} x_i^p \right)^{\frac{۱}{p}}$$

۱-۵ نرم‌های ماتریسی

حال به تعریف نرم برای ماتریس‌ها می‌پردازیم. اگر چه با نرم‌های ماتریسی سروکار داریم که شرایط آنها، همان شرایط (۱) تا (۳) می‌باشند، اما معمولاً نرم ماتریسی را ترجیح می‌دهیم که رابطه معنی داری با یک نرم برداری داشته باشد و اگر یک نرم برداری $\|.\|$ مشخص شده باشد نرم ماتریسی به صورت زیر تعریف می‌گردد:

$$\|A\| = \sup\{\|Au\| : u \in \mathbb{R}^n, \|u\| = ۱\} \qquad (۱-۱)$$

این نرم ماتریسی وابسته به نرم برداری مفروض نیز نامیده می‌شود. در اینجا A یک ماتریس $n \times n$ است.

قضیه ۱-۵-۱. اگر $\|.\|$ نرمی در $\mathbb{R}^n$ باشد، آنگاه رابطه (۱-۱) یک نرم بر روی فضای خطی همه ماتریس‌های $n \times n$ تعریف می‌کند.

برای اثبات به مرجع [؟] مراجعه شود.

اگر نرم بردار $\|\cdot\|_\infty$ به صورت

$$\|x\|_\infty = \max |x_i|$$

تعریف شود، آنگاه نرم ماتریسی طبیعی آن به صورت زیر تعریف می‌شود:

$$\|A\|_\infty = \max \sum_{j=۱}^{n} |a_{ij}|$$

نرم ماتریسی مهم دیگر، نرم ماتریسی $\ell_۲$ (که نرم طیفی نیز نامیده می‌شود) است که به صورت زیر تعریف می‌شود:

$$\|A\|_۲ = \sup \|Ax\|_۲ = \max |\sigma_i|$$

که در آن σ_i ها مقادیر تکین A می‌باشند. در اینجا نرم برداری طبیعی، نرم اقلیدسی است.

نرم فربنیوس ماتریسی هم به صورت زیر تعریف می‌شود:

$$\|A\|_F = \left(\sum_{i=۱}^{m} \sum_{j=۱}^{n} |a_{ij}|^۲ \right)^{۱/۲}$$

تعریف ۱-۵-۲ (شرایط کروش- کان- تاکر). مسئله بهینه سازی زیر مفروض است :

$$\text{Min} f(x) \qquad (۲-۱)$$

$$st$$

$$g(x) \geq ۰$$

که $f(x)$ و $g(x)$ دو تابع مشتق پذیر هستند. تابع لاگرانژ [۱] به صورت زیر تعریف می شود:

$$L(x, \lambda) = f(x) - \lambda g(x) \qquad \lambda \geq ۰ \qquad (۳-۱)$$

جواب مسئله بهینه سازی ۲-۲۵ ، به وسیله بهینه سازی تابع لاگرانژ ۲-۴۳ نسبت به x و λ حاصل می شود به طوری که اولا

$$\nabla f(x) - \lambda \nabla g x(x) = ۰$$

و دوما شرایط زیر نیز برقرار باشد:

$$g(x) \geq ۰$$

$$\lambda \geq ۰$$

$$\lambda g(x) = ۰$$

این شرایط به عنوان شرایط کروش کان تاکر [۲] شناخته می شوند [۶].

تعریف ۱-۵-۳. شبه معکوس [۳]: فرض می کنیم ماتریس $A \in \mathbb{R}^{m \times n}$ باشد آن گاه ماتریس $A^{\dagger} \in \mathbb{R}^{n \times m}$ شبه معکوس ماتریس A است، اگر

$$AA^{\dagger}A = A$$

و ماتریس های $U \in \mathbb{R}^{n \times n}$ و $V \in \mathbb{R}^{m \times m}$ موجود باشند به طوری که

$$A^{\dagger} = UA^{T}$$

$$A^{\dagger} = A^{T}V$$

برقرار باشند.

[۱] Lagrange [۲] Karush Kuhn Tucker [۳] Pseudo-inverse

شرایط لازم $A^{\dagger} = UA^T = A^TV$ را می توان به صورت زیر تعبیر نمود.

هر سطر از ماتریس شبه معکوس $A^{\dagger}$ از A ترکیب خطی از سطرهای A^T است و هر ستون $A^{\dagger}$ ترکیب خطی از ستون های A^T است [۷].

۱-۶ مساله حداقل مربعات

دستگاه معادلات خطی ناسازگار زیر را در نظر بگیرید،

$$Ax = b$$

چون سیستم ناسازگار است $b \notin R(A)$ و برای هیچ مقدار x تساوی مذکور برقرار نیست. لذا داریم،

$$\epsilon = b - Ax \qquad (۴-۱)$$

که در آن ϵ بردار خطا [۱] می باشد. اندازه خطا با استفاده از نرم دو به صورت زیر تعریف می شود:

$$\| \epsilon \| = \| b - Ax \| \qquad (۵-۱)$$

برای یک سیستم ناسازگار $Ax = b$، هدف، یافتن برداری مانند $\hat{x}$ است به طوری که خطای محاسبه شده به صورت $\| \hat{\epsilon} \| = \| b - A\hat{x} \|$ کوچکترین مقدار خطای ممکن باشد در این صورت بردار $\hat{x}$ را جواب حداقل مربعات [۲] می گویند [۳].

تعریف ۱-۶-۱. خوشه بندی ،فرایند دسته بندی مجموعه ای از اشیاء به خوشه هایی هست ،که اعضا درونی هر خوشه بیشترین شباهت را به یکدیگر و کمترین شباهت را نسبت به اعضا سایر خوشه ها داشته باشند. از خوشه بندی در علوم مختلفی مانند مهندسی، پزشکی، علوم اجتماعی و بازاریابی استفاده می شود.

خوشه بندی فرایند تفکیک داده ها یا اشیا به زیرکلاس هایی به نام خوشه می باشد. در هر خوشه داده هایی قرار می گیرند که به نظر میرسد شباهت بیشتری به همدیگر دارندو داده هایی که به نظر می رسد شباهت کمتری به یک دیگر دارند، در خوشه های مختلف قرار می گیرند.

جهت تجزیه و تحلیل خوشه ها، اشیا با نقاطی در فضای N بعدی نمایش داده می شود که بردارها نشان دهنده ویژگی اشیا بوده و هدف مساله نیز دسته بندی N داده به K خوشه هست که اشیا هر خوشه بیشترین شباهت را به یکدیگر داشته باشند.

---

[۱] Eroor Vector [۲] The least square

تعریف ۱-۶-۲. برای خوشه بندی داده هایی با ابعاد بالا دو روش اصلی وجود دارد : روشهای خوشه بندی زیرفضا و روشهای کاهش ابعاد. روشهای خوشه بندی زیرفضا، در زیرفضاهایی از فضای اولیه و اصلی به دنبال خوشه ها می گردند. از انواع این روشها می توان به روشهای جستجوی زیرفضا، روشهای خوشه بندی مبتنی بر همبستگی و روشهای خوشه بندی دوگانه اشاره کرد. در روشهای کاهش ابعاد،فضای جدیدی با ابعاد کمتر ایجاد می شود و روش در این فضا به دنبال خوشه ها می گردد.

تعریف ۱-۶-۳. خوشه بندی طیفی: خوشه بندی طیفی یک روش کاهش ابعاد محسوب می شود. ایده ی کلی در این روشها ساخت ابعاد جدید با استفاده از یک ماتریس خویشاوندی است.

تعریف ۱-۶-۴. خوشه‌بندی انحصاری [۱] و خوشه‌بندی با هم‌پوشی [۲]

در روش خوشه‌بندی انحصاری پس از خوشه‌بندی هر داده دقیقا به یک خوشه تعلق می‌گیرد مانند روش خوشه‌بندی K-Means.

ولی در خوشه‌بندی با هم پوشی پس از خوشه‌بندی به هر داده یک درجه تعلق بازای هر خوشه نسبت داده می‌شود. به عبارتی یک داده می‌تواند با نسبتهای متفاوتی به چندین خوشه تعلق داشته باشد. نمونه‌ای از آن خوشه‌بندی فازی است.

۱-۶-۱ مقدارهای ویژه و بردارهای ویژه

یکی از مباحث اصلی در جبر خطی و آنالیز ماتریس، مقدارهای ویژه و بردارهای ویژه ماتریس های مربعی هستند که اطلاعات اساسی درباره ماتریس در اختیار ما قرار می دهند. مفهوم مرتبط برای ماتریس های مستطیلی مقدارهای منفرد و بردارهای منفرد نامیده می شوند که نقش تعیین کننده ای در تقریب رتبه پایین ماتریس دارند و مشخصات مهم ماتریس اصلی را به ما می دهند.

تعریف ۱-۶-۵. اسکالر $\lambda \in \mathbb{C}$ مقدار ویژه ماتریس $A \in \mathbb{C}^{n\times n}$ است، هر گاه بردار ناصفر $x \in \mathbb{C}^n$ موجود باشد به طوری که $Ax = \lambda x$. بردار x بردار ویژه متناظر به مقدار ویژه λ نامیده می شود.

یک ماتریس مربعی از مرتبه $n \times n$ با احتساب تعداد تکرار، دارای n مقدار ویژه است. مجموعه تمام مقدارهای ویژه با $\sigma(A)$ نشان داده می شود. شعاع طیفی A بیشینه مقدار قدر مطلق $\sigma(A)$ است و با $\rho(A)$ نشان داده می شود:

$$\rho(A) = \max\{|\lambda|; \lambda \in \sigma(A)\}.$$

یک ابزار بسیار مفید در تجزیه ماتریس، تجزیه مقدار منفرد است که در قضیه بعدی تعریف می شود:

[۱] Exclusive or Hard Clustering [۲] Overlapping or Soft Clustering

قضیه ۱-۶-۶. برای هر ماتریس $A \in \mathbb{R}^{m\times n}$ ، ماتریس های متعامد $U \in \mathbb{R}^{m\times m}$ و $V \in \mathbb{R}^{n\times n}$ چنان وجود دارند که

$$A = U\Sigma V^T \tag{۱-۶}$$

که در آن

$$\Sigma = \begin{pmatrix} \sigma_1 & & 0 & \\ & \ddots & & O_{r\times(n-r)} \\ 0 & & \sigma_r & \\ & O_{(m-r)\times r} & & O_{(m-r)\times(n-r)} \end{pmatrix} \tag{۱-۷}$$

و مقدارهای منفرد σ_i اسکالرهای حقیقی هستند به طوری که

$$\sigma_1 \geq \ldots \sigma_r > 0. \tag{۱-۸}$$

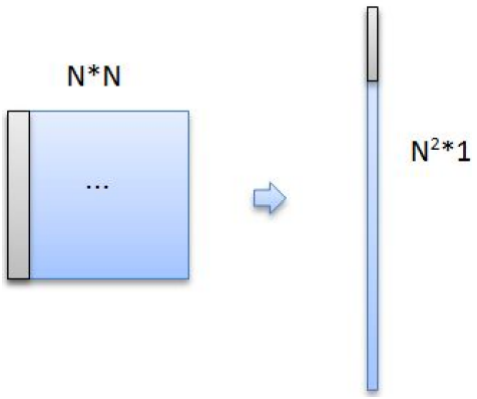
همچنین ستون های U و V به ترتیب بردارهای ویژه A^TA و AA^T هستند.

۱-۷ تجزیه ماتریسی

به طورکلی، تجزیه ماتریسی ابزاری برای تحلیل داده ها است. هرتجزیه تعبیرهای مختلفی را از ساختار ضمنی داده ها آشکار می سازد که البته این تعبیرها از نظر ریاضی هم ارز هستند. یک پایگاه داده را می توان ماتریسی با m سطر و n ستون در نظر گرفت. برای مثال، در ماتریس متناظر با یک پایگاه داده تصویری، هر ستون می تواند معرف یک تصویر باشد ۱-۱ . تجزیه ماتریسی باز نمایش ویژه و ساده ای از ماتریس داده اولیه از طریق تولید ماتریس های جدید است. لذا به کارگیری آن، به این دلیل که داده ها اغلب پیچیده هستند و منجر به ناکارآمدی روش های موجود در داده کاوی می شوند، معقول و مناسب خواهد بود. حتی می توان از آن برای پاکسازی داده ای [1] در شرایطی که داده ها آلوده به نویز باشند نیز استفاده کرد. تجزیهٔ مقدار تکین، تحلیل مؤلفهٔ اصلی، تجزیه LDA و تجزیه نامنفی ماتریسی از روش های شناخته شده در مبحث تجزیه ماتریسی هستند. در این پایان نامه سعی بر این است که چگونگی تجزیه ماتریس داده توسط NMF به تفصیل ارائه گردد. اما قبل از آن، سه تجزیه دیگر را مرور می کنیم.

پیشرفت های بوجود آمده در جمع آوری داده و قابلیت های ذخیره سازی در طی دهه های اخیر باعث شده در بسیاری از علوم با حجم بزرگی از اطلاعات روبرو شویم. محققان در زمینه های مختلف مانند مهندسی، ستاره شناسی، زیست

[1] data cleaning



شکل ۱-۱: تبدیل داده از از قالب ماتریس به بردار

شناسی و اقتصاد هر روز با مشاهدات بیشتر و بیشتری روبرو می شوند. در مقایسه با بسترهای داده ای قدیمی و کوچکتر، بسترهای داده ای امروزی چالش های جدیدی در تحلیل داده ها بوجود آورده اند. روش های آماری سنتی به دو دلیل امروزه کارائی خود را از دست داده اند. علت اول افزایش تعداد مشاهدات است، و علت دوم که از اهمیت بالاتری برخوردار است افزایش تعداد متغیرهای مربوط به یک مشاهده می باشد.

تعداد متغیرهایی که برای هر مشاهده باید اندازه گیری شود ابعاد داده نامیده می شود. عبارت "متغیر" بیشتر در آمار استفاده می شود در حالی که در علوم کامپیوتر و یادگیری ماشین بیشتر از عبارات "ویژگی" و یا "صفت" استفاده می گردد. بسترهای داده ای که دارای ابعاد زیادی هستند علیرغم فرصتهایی که به وجود می آورند، چالش های محاسباتی زیادی را ایجاد می کنند. یکی از مشکلات داده های با ابعاد زیاد اینست که در بیشتر مواقع تمام ویژگی های داده ها برای یافتن دانشی که در داده ها نهفته است مهم و حیاتی نیستند یا حتی وجودشان باعث دسته بندی نادرست می‌شود. به همین دلیل در بسیاری از زمینه ها کاهش ابعاد داده یکی از مباحث قابل توجه باقی مانده است.

۱-۷-۱ کاهش ابعاد

در مبحث کاهش ابعاد [۱] داده هدف اصلی کم کردن ویژگی های مساله است

روشهای کاهش ابعاد داده به دو دسته تقسیم می شوند:

۱. روش انتخاب ویژگی ها [۲]:

---

[۱] Feature Reduction [۲] Feature selection

در این روش یک زیرمجموعه از ویژگی های موجود انتخاب می کنیم.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \xrightarrow{\text{انتخاب ویژگی}} \begin{bmatrix} x_{i_1} \\ x_{i_2} \\ \vdots \\ x_{i_M} \end{bmatrix}$$

گاهی اوقات تحلیل های داده ای نظیر طبقه بندی برروی فضای کاهش یافته نسبت به فضای اصلی بهتر عمل می کند.

۲. روش استخراج ویژگی ها [1]:

این روش یک فضای چند بعدی را به یک فضای با ابعاد کمتر نگاشت می کند. در واقع با ترکیب مقادیر ویژگی های موجود، تعداد کمتری ویژگی بوجود می آورد بطوریکه این ویژگی ها دارای تمام (یا بخش اعظمی از) اطلاعات موجود در ویژگیهای اولیه باشند. این روش به دو دسته ی خطی و غیر خطی تقسیم می شود.

در این روش با ترکیب کردن (خطی یا غیر خطی) ویژگی های موجود، ویژگی های جدید را بدست می آوریم که بعد کمتری دارد. در واقع در این روش به دنبال تابعی (خطی یا غیر خطی) مانند $Y = f(X)$ هستیم که با استفاده از آن بردار ویژگی های جدید را بدست آوریم. اگر

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

بردار ویژگی های موجود باشد آنگاه می خواهیم بردار ویژگی های جدید

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}$$

[1] Feature exxtraction

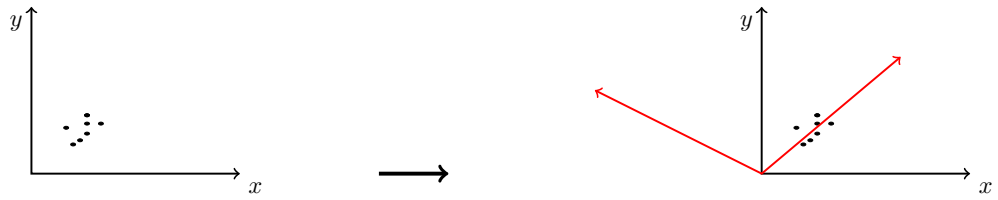
با بعد کمتر $(M < N)$ و به وسیله ی تابع f بدست می آوریم:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \xrightarrow{\text{استخراج ویژگی}} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}\right)$$

روش های خطی که ساده ترند و فهم آنها راحت تر است بدنبال یافتن یک زیرفضای عمومی هستند. اما روشهای غیرخطی که مشکلترند و تحلیل آنها سخت تر است بدنبال یافتن یک زیرفضای محلی می باشند. از روش های خطی می توان به روشهای PCA، LDA، SVD و NMF اشاره کرد [۸].

۱-۸ روش تجزیه PCA

تکنیک PCA بهترین روش برای کاهش ابعاد داده به صورت خطی می‌باشد. یعنی با حذف ضرایب کم‌اهمیت بدست آمده از این تبدیل، اطلاعات از دست رفته نسبت به روشهای دیگر کمتر است. البته کاربرد PCA محدود به کاهش ابعاد داده نمی‌شود و در زمینه‌های دیگری مانند شناسایی الگو و تشخیص چهره نیز مورد استفاده قرار می‌گیرد. در این روش محورهای مختصات جدیدی برای داده‌ها تعریف شده و داده‌ها براساس این محورهای مختصات جدید بیان می‌شوند. در واقع در اینجا PCA یک دوران روی محورهای مختصات انجام می دهد. اولین محور باید در جهتی قرار گیرد که واریانس داده‌ها ماکسیمم شود (یعنی در جهتی که پراکندگی داده‌ها بیشتر است). دومین محور باید عمود بر محور اول به گونه‌ای قرار گیرد که واریانس داده‌ها ماکسیمم شود. به همین ترتیب محورهای بعدی عمود بر تمامی محورهای قبلی به گونه‌ای قرار می‌گیرند که داده‌ها در آن جهت دارای بیشترین پراکندگی باشند. در شکل ۱-۲ زیر این مطلب برای داده‌های دو بعدی نشان داده شده است.



شکل ۱-۲: تبدیل PCA

با این تفاسیر در PCA به دنبال محوری هستیم که با استفاده از آن داده ها را بهتر نمایش دهیم.

مراحل انجام روش PCA

- تعیین ماتریس داده نرمال شدهA و ماتریس کواریانسCمتناظر با آن

- محاسبه بردارهای ویژه ماتریس C
- مرتب کردن بردارهای ویژه بر اساس کاهش مقادیر ویژه متناظرشان
- انتخاب r تا از بردارهای ویژه نرمال شده متناظر با مقادیر ویژه بزرگتر
- تشکیل تبدیل خطی $Q = [q_1 q_2 \dots q_n]$ (ماتریس پایه) با استفاده از این r بردار پایه

در روش تحلیل مولفه اصلی، داده های اولیه روی زیرفضای تولید شده توسط ماتریس پایه C تصویر می شوند که به این ترتیب، ماتریس متناظر $H = [h_1 h_2 \dots h_r] \in \mathbb{R}^{r \times n}$ (ماتریس اوزان) بدست می آید:

$$A \approx QH = \sum_{j=1}^{r} q_j h_j^T =: A_r. \tag{۹-۱}$$

در نتیجه هر یک از داده های اولیه، توسط r مولفه اصلی به صورت زیر بازسازی می شوند:

$$a_j \approx Qh_j = \sum_{j=1}^{r} q_i h_{ij}, \qquad j = 1, \dots, n. \tag{۱۰-۱}$$

هر چند استفاده از روش فوق فواید بسیاری دارد، اما نکاتی را نیز در به کارگیری آن برای یک مجموعه داده معین باید در نظر داشت. روش مذکور از این بابت که هیچ تابع توزیع احتمالی برای مشاهدات در نظر نمیگیرد، روشی آماری نیست. بنابراین دانستن این نکته که PCA تنها برای ارائه داده ها به شکلی ساده تر و با ابعادی کمتر به کار میرود مهم است. نکته دیگر اینکه اغلب تفسیر اقتصادی صحیح از مولفه هایی که ترکیب خطی از متغیرهای اصلی هستند مشکل است . علاوه بر این مولفه ها بر اساس داده های استاندارد شده استخراج میشوند که این امر تفسیر و کاربرد نتایج را برای مراحل بعدی مشکل میسازد [۸].

۹-۱ تحلیل تفکیک خطی LDA

تحلیل تفکیک خطی (یا تشخیص خطی فیشر) روش های آماری هستند که از جمله در یادگیری ماشین و شناسایی الگو برای پیدا کردن ترکیب خطی ویژگی هایی که به بهترین صورت دو یا چند کلاس از اشیا را از هم جدا می کند، استفاده می شوند. یک کاربرد عمده ی این روش، کاستن تعداد بعدهای داده است. LDA ارتباطی تنگاتنگ با تحلیل مولفه اصلی PCA دارد. چرا که هر دو متد به دنبال ترکیبی خطی از متغیرهایی هستند که به بهترین نحو داده ها را توصیف می کنند.

کاربرد اصلی LDA در مسائل طبقه بندی و دسته بندی است [۶] . نکات زیر را در مسائل دسته بندی داریم:

- در مسائل دسته بندی یک بردار ورودی X به یکی از K کلاس مجزای C_k اختصاص داده می شود.

- برای این کار فضای ورودی به نواحی تصمیم گیری تقسیم بندی می شود که مرزهای آنرا سطوح تصمیم گیری می نامند.
- سطوح تصمیم گیری مدل LDA از توابع خطی تشکیل شده است. برای جداسازی فضای ورودی D بعدی از ابرصفحه های $D-1$ بعدی استفاده می شود.

در مسائل دو کلاسی تابع هدف به صورت زیر است:

$$t \in \{0, 1\}$$

که مقدار $t = 0$ برای کلاس C_1 و مقدار $t = 1$ برای کلاس C_2 استفاده می شود.

۱-۹-۱ بررسی مدل LDA

فرض بر این است که بردار x که یک داده D بعدی است با استفاده از رابطه $y = W^T x$ به یک بعد نگاشت شود. این کار باعث می شود تا داده هایی که در D بعد به خوبی جداپذیر بودند در یک بعد در هم فرو رفته و جدا نشوند. اما می توان با انتخاب صحیح W افکنش [1] داده را طوری انجام داد که حداکثر جداسازی بدست آید. در یک مساله دوکلاسه اگر تعداد N_1 داده به کلاس C_1 و تعداد N_2 داده به کلاس C_2 تعلق داشته باشد در این صورت میانگین هر کلاس برابر است با:

$$\mu_1 = \frac{1}{N_1} \sum_{n \in C_1} x_n \qquad \mu_2 = \frac{1}{N_2} \sum_{n \in C_2} x_n$$

میانگین هر کلاس برای داده های تصویر شده به صورت زیر محاسبه می شود:

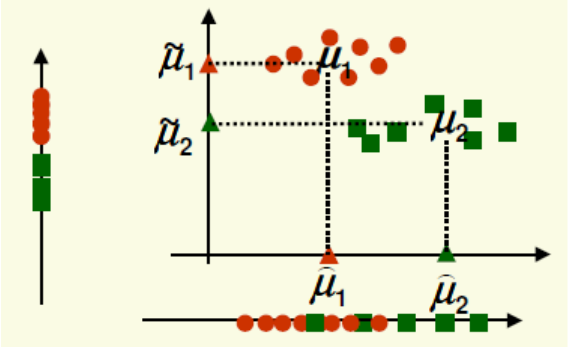
$$\hat{\mu_1} = \frac{1}{N_1} \sum_{y \in C_i} y = \frac{1}{N_1} \sum_{x \in C_i} W^T x = W^T \frac{1}{N_1} \sum_{x \in C_i} x = W^T \mu_1$$

$$\hat{\mu_2} = \frac{1}{N_2} \sum_{y \in C_i} y = \frac{1}{N_2} \sum_{x \in C_i} w^T x = W^T \frac{1}{N_2} \sum_{x \in C_i} x = W^T \mu_2$$

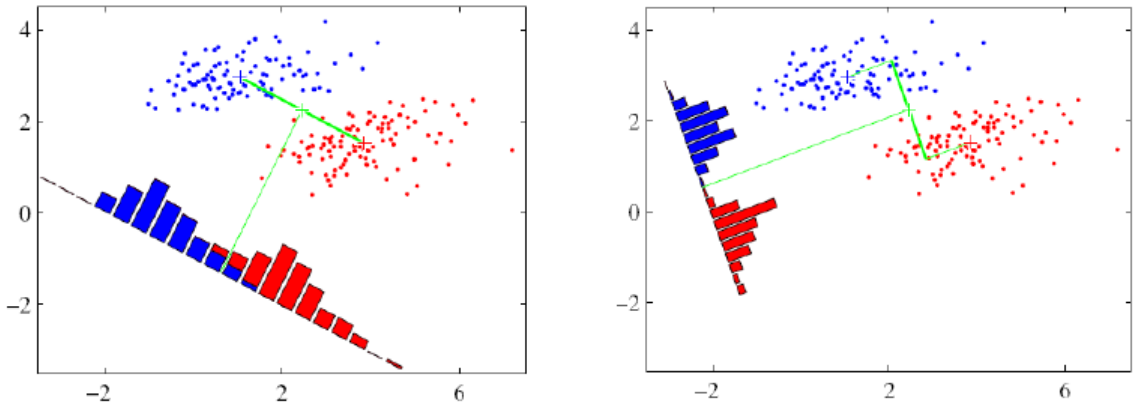
یک راه ساده برای جداسازی داده های نگاشت شده این است که فاصله بین میانگین این داده های نگاشت شده افزایش یابد. در این حالت می توان W را طوری انتخاب کرد که فاصله زیر افزایش یابد:

$$\hat{\mu_2} - \hat{\mu_1} = W^T(\mu_2 - \mu_1)$$

[1] projection



شکل ۱-۳: تفکیک کلاس ها



شکل ۱-۴: افکنش کلاس ها

تابع هدف به صورت زیر تعریف می شود:

$$J(W) = |\hat{\mu}_۲ - \hat{\mu}_۱| = |W^T\mu_۱ - W^T\mu_۲| = |W^T(\mu_۲ - \mu_۱)|$$

فاصله بین میانگین دو کلاس لزوما معیار خوبی برای جداپذیر ساختن دو کلاس نیست. به شکل ۱-۳ توجه کنید این مطلب اهمیت بررسی واریانس داخل کلاسی را روشن می کند.

داده های شکل ۱-۴ در فضای اصلی جداپذیر هستند اما وقتی به خطی که میانگین های آنها را به هم وصل می کند تصویر می شوند همپوشانی زیادی دارند [۳].

۱-۱۰ تجزیه مقدار منفرد SVD

در تجزیه مقدار منفرد یک ماتریس $A \in \mathbb{R}^{m\times n}$ ، به حاصلضرب سه ماتریس U، Σ ، V تبدیل می‌شود به طوری که

$$A = U\Sigma V^T = [u_1, u_2, \cdots, u_m] \begin{bmatrix} D & \vdots & O_1 \\ \cdots & \cdots & \cdots \\ O_2 & \vdots & O_3 \end{bmatrix} [v_1, v_2, \cdots, v_n]^T$$

که در آن U و V ماتریس‌های یکا متعامد و به ترتیب با ابعاد $m \times m$ و $n \times n$ می‌باشند و ماتریس Σ یک ماتریس شبه قطری $m \times n$ است. O_1 ، O_2 و O_3 ماتریس‌های صفر می‌باشند و D یک ماتریس قطری است که عناصر روی قطر آن مقادیر تکین ماتریس A هستند. علاوه بر این عناصر روی قطر D نامنفی هستند و داریم $D_{ii} = \sigma_i$ ، زیرا σ_i ها جذر مقادیر ویژه ماتریس‌های A^TA ویا AA^T هستند و نمی‌توانند مقدار منفی بگیرند. همچنین این مقادیر به ترتیب نزولی روی قطر اصلی ماتریس D مرتب شده‌اند و داریم:

$$\sigma_1 \geq \sigma_2 \geq \sigma \geq \cdots \geq \sigma_p \geq 0$$

که $p = \min(m, n)$.

ستون های U و V به عنوان بردارهای منفرد و درایه های روی قطر اصلی Σ یعنی σ_iها به عنوان مقادیر منفرد شناخته شده هستند.

با توجه به این که از تجزیه SVD داریم:

$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma V^T V\Sigma U^T = U\Sigma^2 U^T$$

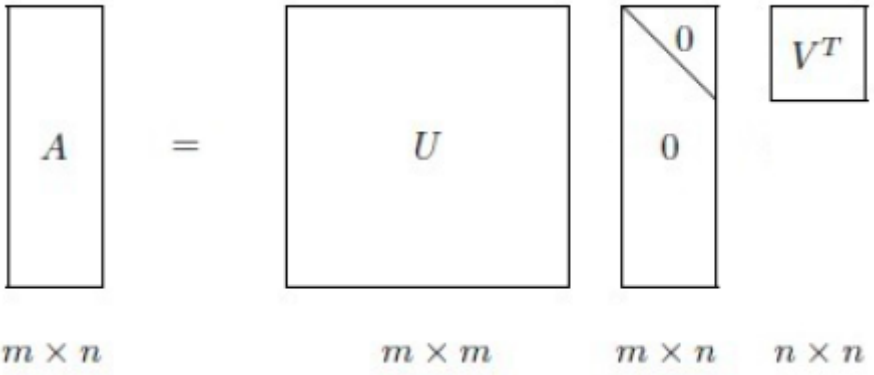
$$A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = (V\Sigma U^T U\Sigma V^T = V\Sigma^2 V^T$$

، می توان نتیجه گرفت، ستون های U و V، به ترتیب، بردارهای ویژه ماتریس های AA^T و A^TA است و مقادیر منفرد، جذر مقادیر ویژه ماتریس AA^T و A^TA است.

تجزیه SVD در علوم مختلف به نام های دیگری شناخته شده است، به عنوان مثال در آمار و آنالیز داده ها مفهوم بردارهای منفرد خیلی شبیه به آنالیز مؤلفه های اصلی [۱] و در پردازش تصویر با نام بسط کارهانن لوو [۲] معروف است. می توانیم تجزیه SVD را به صورت نمادین مانند شکل ۱-۵ نشان بدهیم.

---

[۱] principal component analysis [۲] Karhunen Loewe expansion



شکل ۱-۵: تجزیه SVD

مثال ۱-۱۰-۱. در این جا ماتریس زیر را با استفاده از نرم افزار Matlab تجزیه می کنیم.

$$A = \begin{pmatrix} ۱ & ۲ & ۳ \\ ۴ & ۵ & ۶ \\ ۷ & ۸ & ۹ \\ ۱۰ & ۱۱ & ۱۲ \end{pmatrix}$$

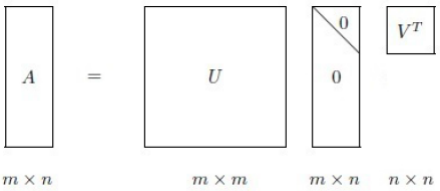
$$U = \begin{pmatrix} -۰/۱۴۰۹ & ۰/۸۲۴۷ & ۰/۵۴۵۶ & -۰/۰۴۷۸ \\ -۰/۳۴۳۹ & ۰/۴۲۶۳ & -۰/۶۹۱۹ & ۰/۴۷۰۴ \\ -۰/۵۴۷۰ & ۰/۰۲۷۸ & -۰/۲۵۳۱ & -۰/۷۹۷۵ \\ -۰/۷۵۰۱ & -۰/۳۷۰۶ & ۰/۳۹۹۴ & ۰/۳۷۴۸ \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} ۲۵/۴۶۲۴ & ۰ & ۰ \\ ۰ & ۱/۲۹۰۷ & ۰ \\ ۰ & ۰ & ۰/۰۰۰۰ \end{pmatrix}$$

$$V = \begin{pmatrix} -۰/۵۰۴۵ & -۰/۷۶۰۸ & -۰/۴۰۸۲ \\ -۰/۵۷۴۵ & -۰/۰۵۷۱ & ۰/۸۱۶۵ \\ -۰/۶۴۴۵ & ۰/۶۴۶۵ & -۰/۴۰۸۲ \end{pmatrix}$$

با قسمت کردن $U = (U_۱, U_۲)$ که $U_۱ \in \mathbb{R}^{m \times m}$ می توانیم SVD کوچک شده را به صورت زیر بدست آوریم

$$A = U_۱ \Sigma V^T,$$



شکل ۱-۶: تجزیه SVD کوچک شده

که به صورت نمادین شکل ۱-۶ داریم.

همچنین SVD را می توان به صورت بسط ماتریسی زیر نوشت:

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T,$$

که معمولا به فرم "ضرب خارجی" شناخته شده و این بسط از فرم کوچک شده SVD مشتق شده است [۸] :

$$A = U_1 \Sigma V^T = \begin{pmatrix} u_1, & u_2, & \dots, & u_n \end{pmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix} \begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{pmatrix}$$

$$= \begin{pmatrix} u_1 & u_2 & \dots & u_n \end{pmatrix} \begin{bmatrix} \sigma_1 v_1^T \\ \sigma_2 v_2^T \\ \vdots \\ \sigma_n v_n^T \end{bmatrix}$$

۱-۱۱ تقریب مرتبه پایین

مساله تقریب مرتبه پایین، تقریب زدن یک ماتریس $m \times n$ داده شده A با یک ماتریس $\hat{A}$ دیگر با رتبه مشخص r، به طوری است که r از m و n کوچک‌تر باشد. استفاده از تجزیه مقدار منفرد بهترین تقریب را برای این مساله نتیجه می‌دهد. تعریف ریاضی مساله در ادامه آورده شده است:

$$\text{minimize over}_{\hat{A}} \quad \|A - \hat{A}\|_F \quad \text{subject to} \quad \text{rank}(\hat{A}) \leq r \qquad (۱۱-۱)$$

فرض کنید

$$A = U\Sigma V^T \in \mathbb{R}^{m\times n}, \quad m \le n \tag{۱-۱۲}$$

تجزیه مقدار منفرد ماتریس A باشد و U، $\Sigma = \text{diag}(\sigma_۱, \ldots, \sigma_m)$ و V به صورت زیر افرازبندی شده باشند:

$$U = \begin{bmatrix} U_۱ & U_۲ \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_۱ & ۰ \\ ۰ & \Sigma_۲ \end{bmatrix}, \quad V = \begin{bmatrix} V_۱ & V_۲ \end{bmatrix}$$

جایی که $\Sigma_۱$ ، $r \times r$ باشد، $U_۱$ ، $m \times r$ و $V_۱$ ، $n \times r$ باشند. آنگاه ماتریس رتبه- r از ضرب زیر به دست می‌آید:

$$\hat{A} = U_۱\Sigma_۱V_۱^T$$

به گونه‌ای که

$$\|A - \hat{A}\|_F = \min_{\text{rank}(\hat{A})\le r} \|A - \hat{A}\|_F = \sqrt{\sigma_{r+۱}^۲ + \ldots + \sigma_m^۲} \tag{۱-۱۳}$$

$\hat{A}$ یکتاست اگر و تنها اگر $\sigma_{r+۱} \neq \sigma_r$.

از آنجایی که مساله اصلی مورد بحث در این پایان نامه یک مساله کمترین مربعات است، نرم فربنیوس به طور گسترده ای در آن به کار خواهد رفت [۹، ۱۰].

فصل ۲

تجزیه نامنفی ماتریس

۲-۱ مقدمه

امروزه پیشرفت فناوریِ رایانه ای منجر به افزایش حجم داده ها و نیز به وجود آمدن پایگاه های بزرگ داده ها شده است. در نتیجه روش های مختلفی برای کشف دانش از آنها، معرفی شده است و یا در حال معرفی هستند. در این راستا، یکی از شاخه های نسبتا جدید علمی موسوم به داده کاوی مورد توجه زیادی قرار گرفته است. ماتریس های داده در کاربردهای داده کاوی، غالبا نامنفی هستند که این ویژگی محدودیت هایی را در استفاده از روش های ماتریسی کلاسیک به همراه دارد. اگرچه به کارگیری این روش ها سبب کاهش بعد داده های بزرگ می شود، اما تعبیری صحیح از داده های نامنفی از آنها به دست نمی آید. اخیراً روش جدیدی با نام تجزیه نامنفی ماتریسی برای نمایش خطی داده های نامنفی پیشنهاد شده است که علاوه بر کاهش بعد داده ها، محدودیت روش های کلاسیک را ندارد. در این روش، ماتریس بزرگ متناظر با داده های نامنفی به دو ماتریس نامنفی کوچک تجزیه می شود. در این فصل ابتدا تجزیه نامنفی ماتریسی به طور کامل شرح داده شده و سپس نسخه های مختلف آن معرفی خواهد شد. در ادامه این فصل قابلیت های خوشه بندی NMF بیان خواهدشد.

۲-۱-۱ پیشینه تحقیقات انجام شده

تاکنون روش های زیادی از جمله روش های تحلیل مولفه اصلی ، تحلیل مولفه مستقل ، تجزیه نیمه گسسته ، تجزیه مقادیر تکین و تجزیه ماتریس نامنفی برای کاهش بعد داده ها معرفی شده است. در برخی از کاربردها به عنوان مثال تحلیل و داده کاوی داده های تصویر، لازم است به منظور کاهش بعد از روش هایی استفاده شود که شرایط نامنفی بودن ماتریس های کاهش یافته را حفظ کند که از بین تمام روش های مزبور تنها روش تجزیه ی ماتریس نامنفی این شرایط را حفظ می کند [۱۰]. تجزیه ماتریس نامنفی برای اولین بار در سال ۱۹۷۰ تحت عنوان "تجزیه منحنی های خود سازمانده " ارائه شد [۱۰]. پاآترو و تاپر در ۱۹۹۴ مساله NMF رامطرح کرده و الگوریتمی برای آن ارائه کردند که از طریق حل مساله کمترین مربعات

به صورت تناوبی W و H را می یافت. اما NMF زمانی معروفیت پیدا کرد که لی و سونگ در سال ۱۹۹۹ یک الگوریتم ضربی بسیار کارا برای آن ارائه نمودند [۱۱، ۱۲]. پس از انتشار مقاله آن ها تجزیه مزبور با نام تجزیه نامنفی ماتریس معرفی گردید. این تجزیه بسیاری از ساختارهای داده اصلی را حفظ و نامنفی بودن هر دو ماتریس پایه و ضرایب را ضمانت می کند [۱۰]. بعد از لی و سونگ الگوریتم های متعددی برای این مساله ارائه شده است که غیر از بهبودهایی که برای روش کمترین مربعات تناوبی ارائه شده است بقیه روش ها در زمره روش های کاهش گرادیان طبقه بندی می شوند [۱۱]. تجزیه ی غیر منفی ماتریس (NMF) یک تکنیک برای کاهش خطی ابعاد و تجزیه و تحلیل داده ها است که عناصر غیر منفی را برای داده های ورودی غیر منفی ارائه می دهد. در اصل، NMF یک الگوریتم یادگیری بدون ناظر از جبر خطی است که نه تنها ابعاد داده را کاهش می دهد، بلکه به طور همزمان خوشه بندی را نیز انجام می دهد.

این مجموعه طیف گسترده ای از برنامه های کاربردی از جمله تجزیه و تحلیل متن، خوشه بندی مستندات، تشخیص چهره / تصویر، مدل سازی زبان، پردازش گفتار و بسیاری دیگر را یافته است [۱۳].

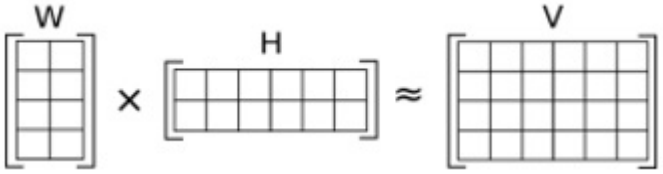
یک ماتریس داده غیر منفی حاوی مجموعه ای از اقلام با تعداد ثابت از ویژگی ها است، الگوریتم NMF ، تجزیه ماتریس داده ها را تولید می کند وعوامل پنهان جالب را که عامل تعامل بین اقلام و ویژگی های آنها هستند، نشان می دهد. می توان در نظر گرفت که هر آیتم را می توان با استفاده از این عوامل نهفته توصیف کرد و با استفاده از اطلاعات در مورد چگونگی بیان هر یک از عوامل نهفته برای هر آیتم، می توان مواردی را که ویژگی های پنهان یکسان دارند، خوشه بندی کرد.

۲-۲ بیان مسأله

تجزیه نامنفی ماتریس مجموعه ای از الگوریتم ها برای تجزیه ماتریس V به دو ماتریس نامنفی W و H یعنی $nmf(V) \rightarrow WH$. تجزیه ی ماتریس ها معمولا یکتا نیست و روشهای مختلفی برای انجام آن ارائه شده است مانند تحلیل مولفه های اصلی یا تجزیه مقادیر منفرد و غیره. تفاوت این روش با روش های دیگر در این است که در تجزیه نامنفی ماتریس این محدودیت را در نظر می گیریم که ماتریس های W و H باید ماتریس غیر منفی باشند.

تجزیه نامنفی ماتریس از مسائل بروز در زمینه جبر خطی است که برای تشخیص الگو، تحلیل داده ها و کاهش بعد به کار می رود. در این حوزه، هدف، تجزیه ماتریس V شامل داده های نامنفی به حاصل ضرب ماتریس پایه W و ماتریس ضرایب H^T با درایه های نامنفی است [۱۱].

تجزیه نامنفی ماتریس یعنی تقریب ماتریس نامنفی $V \in \mathbb{R}_+^{m\times n}$ به وسیله حاصل ضرب عوامل نامنفی $W \in \mathbb{R}_+^{m\times r}$ و $H \in \mathbb{R}_+^{n\times r}$ به ازای یک مقدار خاص r تبدیل به ابزار مفیدی در دامنه وسیعی از کاربردها از جمله در پردازش تصویر، پردازش متن، آوانویسی موسیقی، تحلیل تصاویر ویدئویی، بیوانفورماتیک، شیمی و غیره شده است. مقالات علمی و برنامه های نرم افزاری متعددی در این مورد و مسائل مرتبط با آن به سرعت ایجاد شده و گسترش یافته است. مجموعه $\mathbb{R}_+^{m\times n}$ به تمام ماتریس های نامنفی $m \times n$ دلالت دارد. هدف از تجزیه نامنفی ماتریس (NMF) علاوه بر کاهش مرتبه ای که در بسیاری از کاربردها مد نظر است، مدل بندی، تفسیر مساله و استخراج ویژگی ها به وسیله عوامل نامنفی است. بیان هر



شکل ۲-۱: ضرب ماتریس وزن با ماتریس ویژگی

یک از قابلیت های آن مستلزم تعریف یک فضای برداری برای موضوع مورد نظر می باشد، تا تحت فضای برداری مدل بتوان NMF را به عنوان یک تکنیک روی آن اعمال کرد [۱۴، ۱۵، ۱۶].

مساله NMF به صورت زیر بیان می شود:

اگر $V \in \mathbb{R}_{+}^{m \times n}$ در این صورت $W \in \mathbb{R}_{+}^{m \times r}$ و $H \in \mathbb{R}_{+}^{n \times r}$ را چنان بیابید که $V \cong WH^T$. که در آن r رتبه کاهش یافته، یک پارامتر وابسته به مساله است [۱۱، ۱۶، ۱۷].

تجزیه ماتریس نامنفی در حوزه جداسازی کور منابع و سنجش از دور نیز مورد استفاده قرار گرفته است مهم ترین مساله در به کارگیری الگوریتم NMF در جداسازی داده های سنجش از دور عدم وجود پاسخ یکتا به سبب وجود کمینه های محلی ناشی از غیر محدب بودن تابع هزینه است. از جمله راه کارهای کاهش اثر این نقص افزودن قیدهای کمکی مطابق با ویژگی داده ها به تابع هزینه است [۱۴].

این مساله را میتوان در قالب یک مسئله مینیمم سازی به شکل زیر بیان کرد:

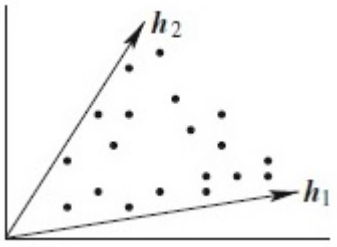
روش های متعددی برای بیان کیفیت تقریب در مساله تجزیه نامنفی ماتریس وجود دارد اما متری که بیشتر به کار می رود نرم فربنیوس است و هدف کمینه کردن معیار زیر است:

$$F(V, WH^T) = \frac{۱}{۲} \parallel V - WH^T \parallel_F^۲ \qquad (۱-۲)$$

یک مساله برنامه ریزی غیر محدب مقید است. از این رو برخی قضایای موجود برای وجود و یافتن مینیمم مطلق در آن صادق نمی باشد، بنابراین به دنبال پیدا کردن کمینه محلی هستیم [۱۱، ؟].

با توجه به مجموعه ای از بردارهای داده n بعدی چند متغیره، بردارها در ستون های ماتریس $n \times m$، V قرار می گیرند کهm تعداد نمونه ها در مجموعه داده است. این ماتریس تقریبا به ماتریس $n \times r$، (ماتریس وزن) و ماتریس $r \times m$، (ماتریس ویژگی)، جایی که r تعداد عوامل تعیین شده توسط کاربر است، تجزیه می شود. معمولا r کوچکتر از n یا m انتخاب می شود، به طوری که W وHکوچکتر از ماتریس اصلیV هستند. دلیل این که تجزیه ماتریس غیر منفی نامیده می شود این است که فاکتورها و مقادیر بدون مقادیر منفی را بازمی گرداند. این به این معنی است که تمام ویژگیهای ماتریس داده باید مقادیر مثبت داشته باشند [۱۳].

تجزیه ی ماتریس یا تجزیه و تحلیل عامل یک کار مهم است که در تجزیه و تحلیل اطلاعات دنیای واقعی در ابعاد بالا مفید است. SVD یک روش کلاسیک برای تجزیه ی ماتریس است، که تقریب کمترین رتبه بهینه به یک ماتریس با مقادیر



شکل ۲-۲: عدم انطباق NMF

واقعی، در اصطلاح مربعات خطا را ارائه می دهد.بسیاری از زمینه های کاربردی، از جمله بازیابی اطلاعات، تشخیص الگو و داده کاوی، نیاز به پردازش اطلاعات باینری دارند نه داده های واقعی. بسیاری از داده های واقعی زندگی و یا سیگنال های فیزیکی، مانند شدت پیکسل،نوسان طیفی، نوشته های متنی، بیان ژن، کیفیت هوا، اطلاعات و شمار تصادفات ، به طور طبیعی با ارقام غیر منفی نشان داده شده است.در تجزیه و تحلیل مخلوطی از این داده ها، نامنفی بودن اجزای منحصر به فرد یک محدودیت مناسب است.انواع تکنیک ها برای تجزیه و تحلیل داده ها مانند PCA غیر منفی، ICA غیر منفی و تجزیه ی ماتریس غیر منفی (NMF) در دسترس هستند. هدف از همه این تکنیک ها، بیان کردن داده های غیرمنفی داده شده به عنوان یک ترکیب خطی غیر منفی تضمین شده از مجموعه ای از پایه های غیرمنفی است. NMF ، همچنین به عنوان تقریب ماتریس غیر منفی یا تجزیه ی ماتریسی مثبت شناخته شده است، یک روش یادگیری بدون ناظر برای تجزیه ی یک ماتریس به عنوان یک حاصلضرب از دو ماتریس است که در آن تمام عناصر غیرمنفی هستند.

NMF یک روش مناسب و ثابت برای انجام کارهایی مانند پردازش تصاویر و سیگنال های غیرمنفی، بازیابی طیفی ، استخراج ویژگی، کاهش ابعاد، رده بندی و خوشه بندی ، مدل سازی زبان، داده کاوی، نوروبیولوژی (جداسازی ژن) و نمایه های بیان ژن دارد. NMF ممکن است به طور مستقیم تجزیه منحصر به فرد انجام ندهد. اگر چه یکتایی ممکن است با اعمال برخی محدودیت ها به عوامل بهبود یابد، اما هنوز هم یک مسئله چالش برانگیز است که به طور کلی منجر به شناسایی منابع می شود .شکل ۲-۲ مشکل عدم هماهنگی را در دو بعد نشان می دهد.فضای باز بین نقاط داده و محور مختصات وجود دارد. ما می توانیم بردارهای پایه h_1 و h_2 را در هر جایی از این فضای باز بین محور مختصات و داده ها انتخاب کنیم و هر نقطه داده را دقیقا با یک ترکیب خطی غیرمنفی از این بردارها بیان کنیم [۱۸].

۲-۳ کاربردها

NMF این توانایی را دارد که به طور خودکار عوامل فیزیکی و قابل تفسیر را استخراج کند. NMF در پردازش تصویر، پردازش(استخراج) متن و تصویربرداری فراطیفی مورد استفاده قرار می گیرد که در ادامه شرح داده خواهند شد. علاوه بر این ها کاربردهای دیگری نیز دارد که عبارتند از:

- کنترل انتشار هوا [۱۹]

- زیست شناسی محاسباتی [۲۰]
- جداسازی منابع کور [۲۱]
- جداسازی منابع تک کانال [۲۲]
- خوشه بندی [۱۵]
- تجزیه و تحلیل موسیقی [۲۳]
- تشخیص جامعه [۲۴]

۲-۳-۱ کاهش بعد

کاهش بعد به عنوان قابلیت اصلی NMF به این شکل صورت می پذیرد که اگر تحت فضای برداری مدل، داده های نامنفی در ماتریس V ذخیره شود، آنگاه بعد از اعمال تکنیک NMF به ازای مقداری از r بر روی ماتریس V و یافتن مدل های جدید W و H می توان به جای ذخیره ماتریس V ماتریس های W و H را ذخیره نمود. به این ترتیب داریم:

$$Size(V) = mn \tag{۲-۲}$$

$$Size(W) + Size(H) = r(m+n) \tag{۲-۳}$$

اگر $r < \frac{mn}{m+n}$ در این صورت توانسته ایم ماتریس V را به ماتریس های کوچک تر W و H ذخیره کنیم. در عین حال WH^T تقریبی از ماتریس اصلی است. در عمل هر گاه W ماتریس خیلی بزرگ و r کوچک باشد آنگاه کاهش بعد قابل ملاحظه ای روی می دهد. کاهش خطای تقریب رابطه مستقیمی با افزایش مقدار r دارد، لذا بایستی تعادلی بین افزایش خطا و کاهش مرتبه هنگام انتخاب مقدار r برقرار نمود. تحت فضای برداری مدل با فرض ماتریس V و تجزیه آن به ماتریس های W و H ، ستون های ماتریس W نقش استخراج کننده ویژگی های اصلی موضوع مورد بحث را خواهد داشت. به ماتریس W ماتریس پایه اطلاق می شود. دیده می شود که نامنفی بودن ماتریس ها نقش تعیین کننده ای در این موقعیت ایفا می کنند [۱۱].

۲-۳-۲ کاربرد در پردازش تصویر

در سال های اخیر پردازش تصاویر دیجیتالی به عنوان یک موضوع مهم در حوزه هایی مانند تشخیص چهره، تشخیص نوری کاراکتر، بازیابی تصاویر بر پایه محتوا و غیره مطرح گشته است. هر تصویر دیجیتالی تکرنگ، یک ماتریس مستطیلی از

پیکسل هاست و هر پیکسل به وسیله شدت روشنایی آن نمایش داده می شود. از آنجایی که شدت روشنایی به وسیله مقادیر نامنفی اندازه گیری می شود میتوانیم هر تصویر را به صورت یک ماتریس نامنفی نشان دهیم. در مورد تصاویر رنگی نیز به همین طریق ولی به وسیله ماتریس های نامنفی متعدد(معمولا ۳ ماتریس برای رنگ های قرمز و سبز و آبی) عمل می شود [۱۱]. در پردازش تصویر هر ستون ماتریس داده $X \in \mathbb{R}_+^{p\times n}$ یک تصویر سطح خاکستری از چهره است. ورودی (i,j) ام ماتریس X نشان دهنده شدت روشنایی پیکسل i ام از چهره j ام است. NMF دو عامل (W,H) را طوری تولید میکند که هر تصویر $X(:,j)$ با استفاده از ترکیب خطی از ستونهای W تقریب زده شود.

$$\overbrace{X(:,j)}^{\text{تصویر صورت j ام}} \approx \sum_{k=1}^{r} \overbrace{W(:,k)}^{\text{ویژگیهای چهره}} \overbrace{H(k,j)}^{\text{اهمیت ویژگیها در تصویر j ام}} = \overbrace{WH(:,j)}^{\text{تقریب تصویر j ام}}$$

از آنجایی که W نامنفی است، ستونهای W می توانند به عنوان یک تصویر تفسیر شوند.

وزنها در ترکیب های خطی غیرمنفی هستند $(H \geq 0)$ این تصاویر پایه تنها می توانند به منظور بازسازی هر تصویر اصلی جمع بندی شوند [؟].

۲-۳-۳ کاربرد در پردازش متن

فرض بر این است که هر ستون ماتریس داده X یک سند و هر سطر یک کلمه باشد ورودی (i,j) ام از ماتریس داده X میتواند تعداد دفعاتی باشد که کلمه i ام در سند j ام ظاهر شده است و در این صورت هر ستون X یک بردار شمارش کلمات یک سند است.

هر سند به یک مجموعه از کلمات با وزنهای مختلف وابسته است در حالی که ترتیب کلمات در سند در نظر گرفته نمی شود.

$$\overbrace{X(:,j)}^{\text{سند j ام}} \approx \sum_{k=1}^{r} \overbrace{W(:,k)}^{\text{کلمه k ام}} \overbrace{H(k,j)}^{\text{اهمیت کلمه k ام در سند j ام}} \qquad W \geq 0, H \geq 0$$

این تجزیه را می توان به این شکل تفسیر کرد:

۱. چون W نامنفی است، هر ستون W می تواند به عنوان مجموعه ای از کلمات در نظر گرفته شود.

۲. چون وزنها در ترکیب خطی غیرمنفی هستند $(H \geq 0)$، تنها می توان الحاق مجموعه ای از کلمات ستونهای W را برای بازسازی تمامی اسناد اصلی بدست آورد [؟].

۲-۳-۴ کاربرد در پردازش تصاویر فراطیفی

ستونهای ماتریس داده غیرمنفی X، امضاهای طیفی از پیکسل ها در صحنه ای که در حال تصویربرداری اند، هستند. امضاهای طیفی یک پیکسل کسری از نور روی داد است که توسط این پیکسل در طول موج های مختلف بازتاب می شود و بنابراین غیرمنفی است. برای تصویر فراطیفی معمولا بین ۱۰۰ تا ۲۰۰ باندنشان داده شده با طول موج وجود دارد که در طیف وسیعی از نور مرئی دیده می شود که این مسئله تجزیه و تحلیل دقیق تر از صحنه مورد مطالعه را بدست می آورد. دو هدف این روش:

۱. شناسایی مواد تشکیل دهنده تصویر شامل چمن، جاده ها یا سطوح فلزی است که این موارد به عنوان اجزای پایانی معرفی شده اند.

۲. شناسایی پیکسل ها یعنی این که کدام پیکسل شامل عناصر پایانی است و کدام شباهت (درجه) است.

ساده ترین مدل مورد استفاده برای حل این مسئله، مدل مخلوط خطی است. در این مدل فرض بر این است که امضای طیفی یک پیکسل نتیجه حاصل از ترکیب خطی امضای طیفی آن دسته از اجزاست. وزن ها در ترکیب خطی مربوط به فراوانی این عناصر پایانی در آن پیکسل است.

مثلا اگر یک پیکسل حاوی ۳۰ درصد چمن و ۷۰ درصد سطح جاده باشد تحت مدل مخلوط خطی، امضای طیفی آن ۰.۳ برابر امواج طیفی چمن به علاوه ۰.۷ برابر امضای طیفی سطح جاده خواهد بود.

مدل NMF به این شکل است: امضاهای طیفی عناصر پایانی، عناصر پایه هستند یعنی ستونهای W، در حالی که فراوانی عناصر پایانی در هر پیکسل وزنها هستند یعنی ستونهای H. همین طور رتبه بندی فاکتور r با تعداد عناصر پایانی در تصویر فراطیفی مطابقت دارد.

$$\overbrace{X(:,j)}^{\text{امضای طیفی پیکسل j ام}} \approx \sum_{k=1}^{r} \overbrace{W(:,k)}^{\text{امضای طیفی عنصر پایانی k ام}} \overbrace{H(k,j)}^{\text{فراوانی عنصر پایانی k ام در پیکسل j ام}}$$

[؟]. در ادامه یک مثال برای تجزیه نامنفی ماتریس ارائه شده است:

مثال ۲-۳-۱. فرض کنید مجموعه ای از سخنرانی ها را داریم و می خواهیم موضوعات مشترک این سخنان را تعیین کنیم. برای شروع، باید ماتریس داده ها را تعریف کنیم. برای مدل سازی موضوع از ماتریس سند مدت استفاده می کنیم که تعداد رخدادهای هر واژه را در هر سند توصیف می کند. برای نمایش ساده تر، ماتریس را فقط ۴×۵ ساختیم. در سناریوی واقعی، تعداد ستون ها و ردیف ها بسیار بالاتر خواهد بود. ماتریس داده را طوری ساختیم که ردیف ها به اسناد و ستون ها به کلمات متصل می شوند:

جدول ۲-۱: ماتریس ورودی V

	کار	انرژی	بازار	اشتغال
سخنرانی اول	۳۶	۳	۴۵	۵۴
سخنرانی دوم	۴	۳۴	۲۳	۳۱
سخنرانی سوم	۹	۶۵	۱۱	۰
سخنرانی چهارم	۱۷	۳	۳	۰
سخنرانی پنجم	۰	۱۴	۷	۴

حالا که ماتریس داده ها را تعریف کردیم، به اجرای الگوریتم ادامه می دهیم. الگوریتم نیاز به یک پارامتر تعریف شده توسط کاربر، تعداد عوامل مورد نیاز برای استخراج یعنی r(در بخش قبلی) دارد. فرض کنید می خواهیم ۲ عامل را از داده های مان استخراج کنیم. در این حالت، پس از اجرای الگوریتم، ازماتریس های W (وزن) وH(عوامل) نتایج را دریافت می کنیم. ماتریسW(ماتریس وزن) سند با ابعاد ۲ × ۵:

جدول ۲-۲: ماتریس W حاصل از تجزیه ماتریس V

	عامل اول	عامل دوم
سخنرانی اول	۰.۰۲۱۱۳۵۲۱۸	۰.۶۳۴۱۱۵۴۲
سخنرانی دوم	۰.۲۶۸۹۳۵۷۸	۰.۲۴۲۴۸۵۴۴
سخنرانی سوم	۰.۵۶۵۲۱۰۶۱	۲.۲۲۰۴e-۱۶
سخنرانی چهارم	۰.۰۲۸۲۵۶۲۷۴	۰.۰۸۸۳۳۲۷۷۵
سخنرانی پنجم	۰.۱۱۶۶۶۲۲۳	۰.۰۳۵۰۶۶۳۶۵

ماتریسH(ماتریس عامل) یک ماتریس کلمه است، با ابعاد ۴ × ۲:

جدول ۲-۳: ماتریس H حاصل از تجزیه ماتریس V

	کار	انرژی	بازار	اشتغال
عامل اول	۱۰.۹۷۵۱۲۸	۱۱۸.۱۶۵۰۳	۲۱.۲۴۶۲۵۹	۲.۲۲۰۴e-۱۶
عامل دوم	۵۵.۰۲۴۸۷۲	۰.۸۳۴۹۶۷۸۲	۶۷.۷۵۳۷۴۱	۸۹

است. اول، نگاهی به ماتریس تجزیه می کنیم. آنچه که می خواهیم در اینجا مشخص کنیم این است که کدام کلمات برای هر یک از ویژگی ها مهمترین وزن را دارند.

برای عامل اول، کلمه "انرژی" دارای بالاترین وزن کلی است و به دنبال آن "بازار" است. دو کلمه دیگر وزن کمتری دارند. عامل دوم وزن بالایی برای کلمات "اشتغال"، "بازار" و "کار" دارد، در حالیکه وزن آن برای کلمه "انرژی" کم است. با در نظر داشتن این موضوع می توان گفت که عامل اول مجموعه ای از کلمات مربوط به موضوع انرژی را نشان می دهد، در حالی که عامل دوم مجموعه ای از کلمات مربوط به موضوع کار را نشان می دهد. مرحله نهایی این است که به ماتریس Wبرسیم. با یافتن بالاترین وزن برای هر سخنرانی می توان نتیجه گرفت که موضوع اصلی سخنرانی چیست. به عنوان مثال، موضوع اصلی سند اول موضوع کار بود، در حالی که موضوع اصلی گفتار ۳ سند سوم موضوع انرژی بود. حال تجزیه SVD را روی این ماتریس اعمال می کنیم و همانطور که مشاهده می شود اطلاعاتی که از تجزیه NMF بدست آمد از این تجزیه حاصل نمی شود.

$$U = \begin{pmatrix} -۰/۷۱۰۹ & ۰/۶۰۹۳ & -۰/۱۷۸۴ & ۰/۲۱۷۶ & -۰/۲۱۰۱ \\ -۰/۵۱۱۵ & -۰/۱۳۷۰ & ۰/۶۰۹۴ & -۰/۵۲۱۶ & ۰/۲۷۵۸ \\ -۰/۴۵۲۴ & -۰/۷۶۹۵ & -۰/۲۹۹۴ & ۰/۱۳۲۰ & -۰/۳۱۰۱ \\ -۰/۰۹۴۲ & ۰/۰۳۲۳ & -۰/۷۰۲۰ & -۰/۴۴۸۶ & ۰/۵۴۴۱ \\ -۰/۱۳۹۳ & -۰/۱۲۹۵ & ۰/۱۱۹۶ & ۰/۶۷۹۷ & ۰/۶۹۸۲ \end{pmatrix}$$

$$S = \begin{pmatrix} ۹۶/۰۷۲۲ & ۰ & ۰ & ۰ \\ ۰ & ۶۴/۹۵۵۳ & ۰ & ۰ \\ ۰ & ۰ & ۲۱/۰۵۷۰ & ۰ \\ ۰ & ۰ & ۰ & ۳/۲۴۸۱ \\ ۰ & ۰ & ۰ & ۰ \end{pmatrix}$$

$$V = \begin{pmatrix} -۰/۳۴۶۷ & ۰/۲۳۱۱ & -۰/۸۸۳۹ & -۰/۲۱۲۲ \\ -۰/۵۳۲۵ & -۰/۸۴ & ۰/۰۱۳۹ & -۰/۱۰۲۷ \\ -۰/۵۲۰۳ & ۰/۲۳۰۸ & ۰/۰۶۷۸ & ۰/۸۱۹۴ \\ -۰/۵۷۰۵ & ۰/۴۳۳۲ & ۰/۴۶۲۵ & -۰/۵۲۲۶ \end{pmatrix}$$

۴-۲ الگوریتم هایی برای تجزیه NMF

مسائل بهینه سازی NMF معمولا غیر محدب هستند. NMF معمولا با روش متناوب کاهش گرادیان انجام می شود که مربع به فاصله اقلیدسی اعمال می شود. از جمله مهمترین الگوریتم های NMF می توان به الگوریتم قواعد به روز رسانی ضربی و الگوریتم کمترین مربعات غیر منفی متناوب [۱] اشاره کرد که در ادامه به طور مختصر شرح خواهیم داد و در فصل سوم به طور کامل به آنها خواهیم پرداخت.

[۱] ANLS

روشهای گرادیان پیش بینی شده در حل مسائل مینیمم سازی محدب در مقیاس بزرگ با توجه به محدودیت های خطی بسیار کارآمد هستند و الگوریتم های گرادیان پیش بینی شده NMF برای حل مسائل مینیمم سازی در مقیاس بزرگ با توجه به محدودیت های غیرمنفی و تنکی بسیار کارآمد هستند. [۱۸].

۲-۴-۱ الگوریتم قوانین به روز رسانی ضربی و کمترین مربعات غیرمنفی

همان طور که قبلا اشاره شد مسئله NMF به شکل زیر فرمولبندی می شود

$$\min_{W,H} \| V - WH \|_F, \tag{۴-۲}$$

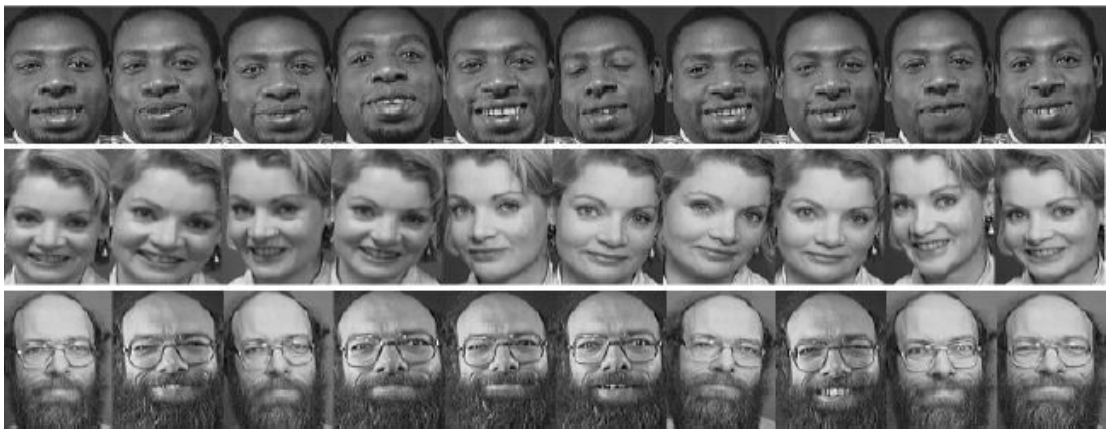
با توجه به این محدودیت که تمام عناصر W و H غیر منفی هستند. قوانین به روز رسانی ضربی برای NMF با رابطه زیر ارائه می شود:

$$H \leftarrow V \otimes \frac{W^T V}{W^T W H}, \tag{۵-۲}$$

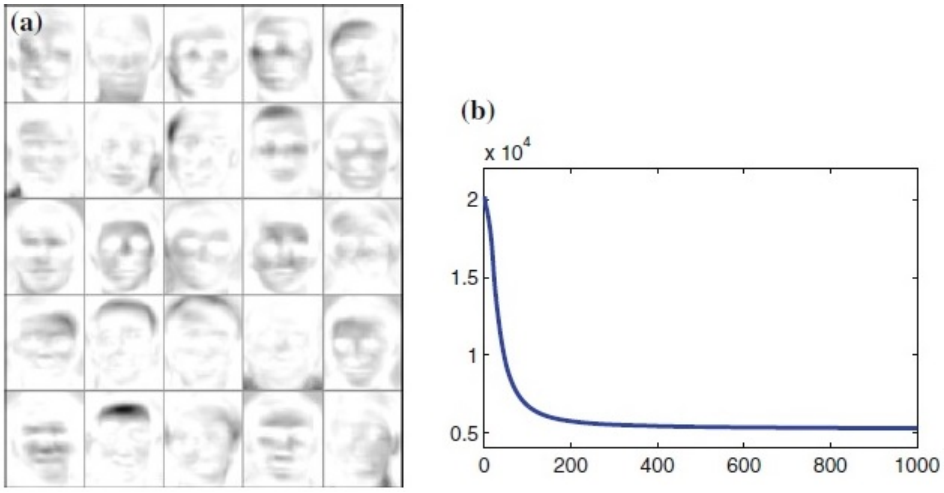
$$W \leftarrow W \otimes \frac{V H^T}{W H H^T}, \tag{۶-۲}$$

که در آن $\otimes$ و / به ترتیب به علامت ضرب و تقسیم علامتگذاری شده است.ماتریس های W و H با مقادیر تصادفی مثبت مقدار دهی می شوند و این معادلات تکرار می شوند.

پس از یادگیری بردارهای پایه W در NMF ، داده های جدید در ماتریس V به فضایی با ابعاد r نگاشت می شوند W را ثابت نگه میداریم و پس از آن به صورت تصادفی H را مقداردهی اولیه می کنیم و این عمل را تا رسیدن به همگرایی تکرار می کنیم. یا H را ثابت نگه داشته و سپس مسئله ی $V^T = WH^T$ را برای W با استفاده از شبه معکوس حل می کنیم. راه حل کمترین مربعات می تواند مقادیر منفی H را تولید کند. می توان از طریق جایگذاری مقادیر منفی به صفر و یا با استفاده از کمترین مربعات غیر منفی ،ماتریس های نامنفی W و H را بدست آورد. جایگذاری مقادیر منفی به صفر، محاسباتی بسیار ساده تر از حل کمترین مربعات با محدودیت های غیر منفی است، اما بعضی از اطلاعات پس از صفر شدن از دست می رود. الگوریتم به روز رسانی ضربی ممکن است به یک نقطه ثابت همگرا شود. در یک استراتژی اصلاح شده ، اگر یک ستون کامل از W صفر باشد، آن ستون و همچنین ردیف مربوطه در H بدون تغییر هستند. همگرایی این استراتژی اثبات شده است [۱۸] شکل ۴-۱ یک نمونه از تصاویر پایگاه داده ORL است. این پایگاه داده شامل تصاویر چهل شخص است و از هر شخص ۱۰ تصویر متفاوت موجود می‌باشد الگوریتم به روز رسانی ضربی با استفاده از تابع هدف اقلیدسی برای نمایش بر مبنای بخشهایی از پایگاه داده تصویر صورت ORL پیاده سازی شده است. برای پایگاه داده ORL ، $m = ۹۲, r = ۲۵$



شکل ۲-۳: نمونه تصاویر بانک ORL



شکل ۲-۴: NMF به پایگاه داده تصویر صورت ORL اعمال می شود. a تصاویر پایه برای ماتریس W و b روند تغییرات تابع هدف.

و $n = ۴۰۰$. تصاویر پایه حاصل از پایگاه داده تصویر صورت ORL نشان داده شده است و همچنین تغییرات تابع هدف در طی مسیر بهینه سازی در شکل ۲-۲ نشان داده شده است. همگرایی الگوریتم به روز رسانی ضربی بسیار آهسته است. تابع هدف الگوریتم حداقل مربعات متناوب غیرمنفی، به شکل زیر است:

$$\min_{W,H} F(W,H) = \frac{۱}{۲} \parallel V - WH \parallel_F^۲, \qquad (۷-۲)$$

با توجه به این محدودیت که تمام عناصر W و VH غیر منفی هستند. الگوریتم را می توان به عنوان دو مسئله بهینه سازی محدب متناوب اجرا کرد.

$$W_{r+۱} = \arg\min_{W \geq ۰} F(W, H_r),$$

$$H_{r+1} = \underset{H \geq 0}{\operatorname{argmin}} F(W_{r+1}, H), \tag{۸-۲}$$

این روش می تواند بسیار سریع باشد، در عمل خوب کارکند و همگرایی سریع داشته باشد.

از آنجایی که یک الگوریتم گرادیان برای یادگیری NMF استاندارد از همگرایی آهسته برای مسائل بزرگ در مقیاس بزرگ است، برخی از تلاش ها با استفاده از پیشنهاد روش گرادیان، پیشنهاد روش متناوب LS به NMF برای سرعت بخشیدن به همگرایی آن اعمال شده است. همانطور که گفته شد NMF ممکن است تجزیه منحصر به فرد انجام ندهد ولی برخی از اصلاحات NMF برای تجزیه گزارش شده اند که با اعمال محدودیت تنکی در ماتریس مختلط یا ماتریس منبع یا هر دو منحصر به فرد هستند. یک چارچوب الگوریتمی بهبود یافته برای مسئله کمترین مربعات NMF، بسیاری از نقایص روش های مبتنی برکاهش گرادیان شامل الگوریتم به روز رسانی ضربی و الگوریتم کمترین مربعات متناوب را برطرف می کند.

۲-۵ تجزیه های مربوط به NMF

در این بخش برخی از روش‌های تجزیه نامنفی ماتریس به اختصار بیان خواهد شد یک مرور کلی در مورد روش های تقسیم بندی ماتریس ارائه می‌شود در این بخش X ماتریس ورودی است و ماتریس‌های F و G ماتریس‌های حاصل از تجزیه هستند:

۱. SVD: تجزیه ماتریسی کلاسیک، تجزیه و تحلیل مولفه اصلی (PCA) با استفاده از تجزیه مقادیر منفرد (تکین) است.

$$X \approx U\Sigma V^T$$

فرض کنیم U، V دارای علائم مخلوط باشند داده های ورودی می توانند علائم مخلوط داشته باشند:

$$SVD: \quad X_{\pm} \approx U_{\pm} V_{\pm}. \tag{۹-۲}$$

۲. NMF استاندارد: NMF استاندارد تجزیه یک ماتریس غیر منفی $X \in \mathbb{R}^{m \times n}$ به دو ماتریس نامنفی $F \in \mathbb{R}^{m \times r}$ و $G \in \mathbb{R}^{r \times n}$ ($r < min\{m, n\}$) است که

$$X_+ = F_+ G_+ + E \tag{۱۰-۲}$$

که E خطا (یا باقی مانده) است و M_+ نشان می دهد ماتریس M نامنفی است. بهینه سازی آن در فضای اقلیدسی به صورت زیر فرمول بندی می شود:

$$\min_{F,G} \frac{1}{2} \parallel X - FG \parallel_F^2 \qquad st. \qquad F, G \geq 0 \tag{۱۱-۲}$$

بحث آماری این فرمول از تابع درست نمایی لگاریتم با فرض خطای گوسی بدست آمده است اگر نقاط داده چند متغیره در ستون X مرتب شوند آنگاه F به عنوان ماتریس پایه و G به عنوان ماتریس ضرایب خواهد بود در نتیجه هر ستون F یک بردار پایه است و هر نقطه داده ترکیب خطی غیر منفی از بردارهای پایه است. کاملا پیداست که هدف بهینه سازی یک مسئله ی بهینه سازی غیر محدب است. در نتیجه قوانین به روز رسانی ضربی برای حل معادله معرفی شدند هرچند که پیاده سازی آن ساده است اما همگرایی به یک نقطه ثابت را تضمین نمی کند [۲۵].

۳. Semi-NMF: NMF استاندارد فقط برای داده های غیر منفی که برنامه های کاربردی آن محدودیت دارند کار می کند. آن را به Semi-NMF تعمیم میدهیم که محدودیت های غیرمنفی روی داده های X و ماتریس پایه F حذف شده است که با معادله زیر آن را بیان می کنیم:

$$\min_{F,G} \frac{1}{2} \parallel X - FG \parallel_F^2 \qquad st. \qquad G \geq 0 \tag{۱۲-۲}$$

Semi-NMF را می توان برای ماتریسی از نشانه های مخلوط شده به کار برد به همین دلیل NMF به زمینه های بسیاری گسترش می یابد اگرچه قوانین به روز رسانی بر اساس کاهش گرادیان سرعت همگرایی پایینی را ارائه می دهد. با ثابت نگه داشتن G بروزرسانی F یک مسئله حداقل مربعات است که یک راه حل تحلیلی دارد.

$$F = XG^T(GG^T)^{-1} = XG_+ \tag{۱۳-۲}$$

که $G_+ = G^T(GG^T)^{-1}$ یک شبه معکوس است. به روزرسانی G تا زمانی که F ثابت شود یک مسئله اساسی NNLS است [۲۵].

۴. NMF: هنگامی که داده های ورودی غیرمنفی هستند، محدودیت غیرمنفی را برای F و G در نظر می گیریم. میتوان NMF استاندارد را به شکل زیر نوشت:

$$NMF: \quad X_+ \approx F_+ G_+, \tag{۱۴-۲}$$

با استفاده از محدودیت $X, F, G \geq 0$.

۵. NMF افکنشی: یک مدل NMF پیش بینی شده برای یافتن یک طرح تقریبی ماتریس ورودی است. می تواند به صورت زیر فرموله شود.

$$X \approx FF^T X. \tag{۱۵-۲}$$

۶. NMF محدب: به طور کلی بردارهای پایه $F = (f_1, \ldots, f_k) \in \mathbb{R}_+^{n \times k}$ یک فضای بسیار بزرگتر از فضای ورودی ستونهای $X = (x_1, \ldots, x_n)$ است. در حالی که طبق قضیه ۲-۸-۲ ، F به معنی مراکز خوشه هاست. برای تاکید این معنی هندسی، F را به عنوان ترکیبی محدب از نقاط داده ورودی محدود می کنیم ، یعنی قرار گرفتن F در فضای ورودی، $f_1 = w_{1l}x_1 + \cdots + w_{nl}x_n = Xw_l$ یا $F = XW, w_{il} \geq 0$. این فرم محدود کردن تجزیه را NMF محدب می نامند. NMF -محدب برای هر دو داده های ورودی مخلوط و غیرمنفی اعمال می شود:

$$X_\pm \approx X_\pm W_+ G_+^T. \tag{۱۶-۲}$$

۷. NMF کرنلی : یک نگاشت برداری، که در ماشین های بردار پشتیبان استفاده می شود،

$$x_i \to \phi(x_i), \quad or, \quad X \to \phi(X) = (\phi(x_1), \ldots, \phi(x_n))$$

. همانند مفهوم پیشنهاد شده برای NMF محدب ، F را به ترکیبی محدب از نقاط ورودی تبدیل شده محدود می کنیم:

$$\phi(X) \simeq \phi(X)WG^T, \tag{۱۷-۲}$$

مینیمم سازی تابع هدف به فرم زیر خواهد بود:

$$\begin{aligned} &\| \phi(X) - \phi(X)WG^T \|^2 \\ &= Tr[\phi(X)^T\phi(X) - 2G^T\phi^T(X)\phi(X)W + W^T\phi^T(X)\phi(X)WG^TG], \end{aligned}$$

که فقط به کرنل $\kappa = \phi^T(X)\phi(X)$ وابسته است. این گسترش NMF کرنلی شبیه PCA کرنلی و K-means کرنلی است.

۸. NMF چند لایه ای : در NMF چند لایه، ماتریس عامل اصلی با مجموعه ای از ماتریس های آبشاری با استفاده از تجزیه متوالی جایگزین می شود. با توجه به یک ماتریس ورودی، X اولین بار به $F_{(۱)}G_{(۱)}^T$ تجزیه می شود. سپس در مرحله دوم $G_{(۱)}$ به $F_{(۲)}G_{(۲)}^T$ تجزیه می شود و فرآیند چندین بار تکرار می شود. مدل NMF چند لایه را می توان به صورت زیر نوشت:

$$X \approx F_{(۱)}F_{(۲)}\dots F_{(t)}G_{(t)}^T = FG^T. \qquad (۱۸-۲)$$

NMF چند لایه می تواند عملکرد بسیاری از الگوریتم های NMF را بهبود بخشد و مسئله مینیمم های محلی را با توجه به ماهیت توزیع شده و چند مرحله ای و تجزیه ترتیبی با شرایط اولیه متفاوت مورد توجه قرار دهد.

به طور خلاصه، فرمولبندی های NMF جدید پیشنهاد شده در مجموع به شرح زیر خلاصه می شوند:

$$SVD: \qquad X_{\pm} \approx U_{\pm}V_{\pm}$$

$$NMF: \qquad X_{+} \approx F_{+}G_{+}^T$$

$$ProjectiveNMF: \qquad X \approx FF^TX$$

$$Semi-NMF: \qquad X_{\pm} \approx F_{\pm}G_{+}^T$$

$$Convex-NMF: \qquad X_{\pm} \approx X_{\pm}W_{+}G_{+}^T$$

$$Convex-hullNMF: \qquad X \approx CG^T, \quad c_i \in conv(X)$$

$$Kernel-NMF: \qquad \phi(X_{\pm}) \approx \phi(X_{\pm})W_{+}G_{+}^T$$

$$Symetric-NMF: \qquad W_{+} \approx H_{+}S_{+}H_{+}^T$$

$$Multi-layerNMF: \qquad X \approx F_{(۱)}F_{(۲)}\dots F_{(t)}G_{(t)}^T$$

البته فرمول بندی های NMF دیگری نیز وجود دارد که در بحث فوق وجود ندارد مانند NMF پیچیده برای مجموعه ای از ماتریس های غیر منفی و NMF بیزی با ترکیب روش های بیزی برای NMF و... [۲۶، ۱۸].

۲-۶ NMF برای خوشه بندی

تجزیه های استاندارد یک ماتریس داده با استفاده از مقادیر ویژه تجزیه (SVD) به طور گسترده ای در تجزیه و تحلیل مولفه اصلی (PCA) استفاده شده است. یک تجزیه مانند SVD حاوی مقادیر منفی است و در نتیجه تفسیر آن دشوار است.

تجزیه نامنفی ماتریس NMF مزایای بسیاری نسبت به تجزیه های استاندارد PCA , SVD دارد.

فرض کنیم

$$X = (x_1, x_2, ..., x_n) \in R_+^{p \times n}$$

و داده های ماتریس مقادیر نامنفی باشد. در پردازش تصویر هر ستون یک تصویر سطح خاکستری دو بعدی از پیکسل هاست. در متن کاوی، هر ستون یک سند است. در NMF، ماتریس ورودی X به دو ماتریس نامنفی تجزیه می شود:

$$X \approx FG^T \qquad (۱۹-۲)$$

که $F = (f_1, f_2, ..., f_k) \in R_+^{p \times k}$ ، $G = (g_1, g_2, ..., g_k) \in R_+^{n \times k}$ و k یک پارامتر از پیش تعیین شده است. تجزیه ها توسط مینیمم سازی کمترین مربعات بدست می آیند.

در این بخش NMF جهت خوشه بندی داده ها مورد مطالعه قرار می‌گیرد و همینطور نشان داده می‌شود که NMF با خوشه بندی K-means کرنلی و خوشه بندی طیفی مبتنی بر لاپلاسین معادل است.

۲-۶-۱ خوشه بندی NMF و K-means

الگوریتم خوشه بندی K-means یکی از محبوب ترین روش های خوشه بندی داده است. فرض کنید $X = (x_1, \ldots, x_n)$ با n نقطه داده که آن را به k خوشه ی متقابلا غیر مجزا تقسیم می کنیم. تابع هدف خوشه بندی K-means را می توان به صورت زیر نوشت

$$J_{Kmeans} = \sum_{i=1}^{n} \min_{1 \le K \le k} \| x_i - f_k \|^2 = \sum_{k=1}^{k} \sum_{i \in C_k} \| x_i - f_k \|^2$$

قضیه زیر نشان می دهد که NMF به طور ذاتی مربوط به الگوریتم خوشه بندی K-means است [۱۵، ۱۸].

قضیه ۲-۶-۱. $G\,NMF$- متعامد

$$\min_{F,\geq 0,G\geq 0} \| X - FG^T \|^2, \quad s.t. \quad G^TG = I. \tag{۲۰-۲}$$

معادل خوشه بندی *K-means* است. رابطه ۳-۸ حتی اگر X و F دارای ورودی های علامت مثبت و منفی باشند نیز برقرار است.

برای توضیح بیشتر فرض کنیم $C = (c_1, \ldots, c_k)$ مراکز خوشه هاست که از طریق خوشه بندی K-means بدست می آید. همینطور فرض کنید H شاخص های خوشه ای باشد:

برای مثال ۱ $h_{ki} = 1$ اگر x_i متعلق به خوشهc_k باشد در غیر این صورت $h_{ki} = 0$ می توان تابع هدف خوشه بندی K-means را به این صورت نوشت:

$$J = \sum_{i=1}^{n}\sum_{k=1}^{k} h_{ik} \| x_i - c_k \|^2 = \| X - CH^T \|^2$$

که معادل است با NMF، که F به معنای مراکز خوشه ها و G نشانگر شاخص خوشه بندی است. بنابراین K-means و NMF دارای یک هدف هستند، اما با محدودیت های مختلف. در حقیقت، تابع هدف K-means می تواند این گونه بیان شود: اگر از محدودیت غیرمنفی چشم پوشی کنیم مادامی که محدودیت تعامد را حفظ کنیم ، مولفه اصلی (PCA) راه حل است. از سوی دیگر، اگر در حالی که تعامد را نادیده میگیریم نامنفی بودن را حفظ کنیم، (NMF) راه حل است.

۲-۷ خوشه بندی K-means کرنلی و NMF متقارن

خوشه بندی K-means یکی از روشهای خوشه بندی است که به طور گسترده استفاده می شود. در اینجا ابتدا به طور خلاصه K-means را معرفی می کنیم. این اطلاعات پس زمینه های لازم، نمادها و نشانه ها را برای رویکرد تجزیه نامنفی ماتریس فراهم می کند.

K-means از نمونه های اولیه k ، از مراکز خوشه ها، برای توصیف داده ها استفاده می کند.

تابع هدف مساله، مینیمم سازی مجموع مربعات خطاست که به شکل زیر است:

$$J_K = \sum_{k=1}^{k}\sum_{i\in c_k} \| X_i - m_k \|^2 = C_2 - \sum_k \frac{1}{n_k}\sum_{i,j\in c_k} X_i^T X_j, \tag{۲۱-۲}$$

که $X = (x_1, x_2, \ldots, x_n)$ ماتریس داده است، $m_k = \sum_{i\in c_k} \frac{x_i}{n_k}$ مرکز خوشه c_k از n_k نقطه است و $c_2 = \sum_i \| x_i \|^2$. راه حل خوشه بندی این است که به وسیله k تا بردار شاخص نامنفی نمایش دهیم.

$$H = (h_1, h_2, \ldots, h_n), \qquad h_k^T h_l = \delta_{kl} \tag{۲۲-۲}$$

که

$$h_k = (0, \ldots, 0, \overbrace{1, \ldots, 1}^{n_k}, 0, \ldots, 0)^T / {n_k}^{\frac{1}{2}} \tag{۲۳-۲}$$

حال معادله ۳-۸ بدست می آید. $J_k = Tr(X^T X) - Tr(H^T X^T X H)$. عبارت اول یک مقدار ثابت است فرض می کنیم $W = X^T X$ بنابراین مینیمم J_k بدست می آید:

$$\max_{H^T H = I, H \geq 0} J_w(H) = Tr(H^T W H). \tag{۲۴-۲}$$

ماتریس دو به دو مشابه $W = X^T X$ ضرب داخلی استاندارد ماتریس کرنلی خطی است.که می توان به هر کرنل دیگری تعمیم داد این کار با یک تبدیل غیر خطی(نگاشت) به فضای با ابعاد بالاتر انجام می شود.

$$X_i \longrightarrow \phi(X_i)$$

خوشه بندی تابع هدف تحت این نگاشت، به کمک معادله ۳-۸ می تواند به شکل زیر نوشته شود.

$$\min J_k(\phi) = \sum_i \| \phi(X_i) \|^2 - \sum_k \frac{1}{n_k} \sum_{i,j \in c_k} \phi(X_i)^T \phi(X_j). \tag{۲۵-۲}$$

عبارت اول یک ثابت برای یک تابع نگاشت داده شده $\phi(0)$ است و می توان نادیده گرفت. فرض می کنیم ماتریس کرنل $W_{ij} = \phi(X_i)^T \phi(X_j)$ با استفاده از شاخص خوشه H خوشه بندی K-means کرنلی، با کاهش معادله ۳-۱۱ معادل است.

تابع هدف در معادله ۳-۱۱ می تواند به طور نمادین به صورت زیر نوشته شود.

$$\min J_W = \sum_k \frac{1}{n_k} \sum_{i,j \in c_k} W_{ij} = Tr(H^T W H) \tag{۲۶-۲}$$

هدف K-means کرنلی به حداکثر رساندن شباهت درون خوشه هاست. استفاده از K-means کرنلی می تواند توزیع داده های پیچیده تر از توزیع Gausian را توصیف کند [۱۵].

۲-۷-۱ تجزیه نامنفی K-means کرنلی

در این بخش ثابت می‌شود که بهینه‌سازی معادله ۳-۱۱ می تواند به وسیله تجزیه ماتریس حل شود.

$$W \simeq HH^T, \qquad H \geq ۰ \tag{۲۷-۲}$$

یک تابع هدف مناسب به شکل زیر است:

$$\min_{H\geq ۰} J_l = \| W - HH^T \|^۲, \tag{۲۸-۲}$$

که نرم ماتریس $\| A \|^۲ = \sum_{ij} a_{ij}^۲$ نرم فربینیوس است.

قضیه ۲-۷-۱. تجزیه $W = HH^T$ معادل است با خوشه بندی *K-means* کرنلی با تعامد قوی رابطه معادله ۳-۹

برهان. ماکزیمم سازی معادله ۳-۱۱ را به صورت زیر می توان نوشت:

$$H = \arg \min_{H^T H=I, H\geq ۰} -۲Tr(H^T W H)$$

$$= \arg \min_{H^T H=I, H\geq ۰} \| W \|^۲ -۲Tr(H^T W H) + \| H^T H \|^۲$$

$$= \arg \min_{H^T H=I, H\geq ۰} \| W - HH^T \|^۲ . \tag{۲۹-۲}$$

به راحتی با تعامد $H^T H = I$ اثبات کامل می شود. □

اگر وضعیت نامنفی بودن نادیده گرفته شود، راه حل برای H ، بردارهای ویژه K با بزرگترین مقادیر ویژه هستند وتعامد نیز حفظ شده است. در حال حاضر نامنفی بودن H را حفظ می کنیم. آیا تعامدش را از دست داده است؟

قضیه ۲-۷-۲. تجزیه $W = HH^T$ تقریبا تعامد H را حفظ می کند.

برهان. میتوان دید که $\min_{J_l} = \| W - HH^T \|$ معادل است با

$$\max_{H\geq ۰} Tr(H^T W H), \tag{۳۰-۲}$$

$$\min_{H\geq 0} \parallel H^T H \parallel^2 . \qquad (۳۱-۲)$$

هدف اول بهبود بهینه سازی اصلی هدف معادله ۳-۱۱ است. روی ۲ عبارت تمرکز می کنیم توجه داشته باشید که:

$$\parallel H^T H \parallel^2 = \sum_{\ell k} (H^T H)^2_{\ell k} = \sum_{\ell \neq k} (h_\ell^T h_k)^2 + \sum_k (h_k^T h_k)^2$$

مینیمم سازی عبارت اول معادل است با اجرای تعامد در

$$h_\ell : h_\ell^T h_{ell} \approx 0.$$

مینیمم سازی عبارت دوم معادل است با

$$\min \parallel h_1 \parallel^4 + \ldots + \parallel h_k \parallel^4 \qquad (۳۲-۲)$$

با این حال، H نمی تواند کاملا صفر باشد در غیر این صورت باید داشته باشیم $Tr(H^T W H) = 0$. به صورت خلاصه تر چون $W \simeq HH^T$،

$$\sum_{ij} W_{ij} \simeq \sum_{ij} (HH^T)_{ij} = \sum_{k_i j} h_{ik} h_{jk} = \sum_k \mid h_k \mid^2, \qquad (۳۳-۲)$$

که L_1 نرم $\mid h \mid = \sum_i \mid h_i \mid = \sum_i h_i$ بردار h است. این یعنی $\parallel h_\ell \parallel > 0$. بنابراین معادله ۲-۳۱ با محدودیت غیر صفر معادله ۲-۳۳ بهینه سازی می شود در ضمن H نزدیکترین ستونهای متعامد را دارد برای مثال

$$h_\ell^T h_k \approx \begin{cases} 0 & if L \neq k \\ \parallel h_k \parallel^2 & if L = k \end{cases} \qquad (۳۴-۲)$$

علاوه بر این به حداقل رساندن معادله ۲-۳۲ با استفاده از محدودیت غیر صفر معادله ۲-۳۳ منجر به محدودیت برابری ستونها می شود

$$\parallel h_1 \parallel \simeq \parallel h_2 \parallel \simeq \ldots \simeq \parallel h_k \parallel. \qquad (۳۵-۲)$$

این یک تعادل تقریبی از اندازه خوشه ها را برای ما تضمین می کند. □

نزدیکترین تعامد ستونهای H برای خوشه بندی داده ها مهم است یک تعامد دقیق نشان می دهد که هر سطر از H می

تواند تنها یک عنصر غیر صفر داشته باشد که نشان می دهد هر شی داده تنها به یک خوشه تعلق دارد.این یک خوشه بندی انحصاری است مانند خوشه بندی که در K-means بود. شرایط نزدیکترین تعامد، به عنوان مثال هر شی داده می تواند به شکل کسری (به طور جزئی) به بیش از یک خوشه تعلق داشته باشد. این روش خوشه بندی هم پوشی است. به طور کلی غیر متعامد بودن در میان ستونهای H تفسیر روشن (تعبیر واضحی) از خوشه بندی ندارد [۱۵].

۲-۸ خوشه بندی طیفی و NMF

در سالهای اخیر خوشه بندی طیفی با استفاده از گراف لاپلاسین با یک رویکرد قوی برای خوشه بندی داده ها ظهور کرده است. در اینجا روی خوشه بندی طیفی توابع هدف تمرکز می کنیم. سه هدف وجود دارد: برش نسبت [۹] و برش نرمال [۹] و برش min max چند روش خوشه بندی توابع هدف عبارتند از:

$$J = \sum_{۱\leq p<q\leq k} \frac{S(C_p, C_۲)}{\rho(G_p)} + \frac{S(C_p, C_q)}{\rho(C_q)} = \sum_{k=۱}^{k} \frac{S(C_k, C_k)}{\rho C_k} \qquad (۳۶-۲)$$

$$\rho(C_k) = \begin{cases} |C_k| & for \quad RatioCut \\ \sum_{i\in C_k} d_i & for \quad NormalizeCut \\ S(C_k, C_k) & for \quad MinmaxCut \end{cases} \qquad (۳۷-۲)$$

که c_k مکمل زیرمجموعه c_k در گراف G است،

$$d_i = \sum_J W_{ij}$$

و

$$S(A, B) = \sum_{i\in A} \sum_{j\in B} W_{ij}$$

در این جا نشان می دهیم که به حداقل رساندن این توابع هدف می تواند با انجام روشی غیر از روش تجزیه نامنفی ماتریس معادل باشد. اثبات آن از چند راه طیفی ساده از برش نرمال و برش min max [۹] پیروی می کند.ما روی برش نرمال تمرکز می کنیم.

قضیه ۲-۸-۱. برش نرمال با استفاده از شباهت های دو به دو ماتریس W با خوشه بندی *K-means* کرنلی با ماتریس

کرنل معادل است.

$$\widetilde{W} = D^{\frac{-1}{2}} W D^{\frac{-1}{2}} \tag{۳۸-۲}$$

که $D = (d_1, \ldots, d_n)$

نتیجه ۱. برش نرمال با استفاده از شباهت W با تجزیه نامنفی ماتریس معادل است.

$$\min_{H \geq 0} J_3 = \| \widetilde{W} - HH^T \|^2 \tag{۳۹-۲}$$

برهان. فرض کنیم h_k شاخص خوشه باشد در معادله ۳-۱۰ به راحتی می توان دید که

$$S(C_k, \bar{C}_k) = \sum_{i \in C_k} \sum_{j \in \bar{C}_k} W_{ij} = h_\ell^T (D - W) h_\ell \tag{۴۰-۲}$$

و

$$\sum_{i \in C_k} d_i = h_\ell^T D h_\ell$$

. برای بردار شاخص خوشه یک مقیاس تعریف می کنیم. $z_\ell = D^{1/2} h_\ell / \| D^{1/2} h_\ell \|$ که از شرایط تعامد اطاعت می کند. $z_\ell^T z_k = \delta_{\ell k}$، یا $Z^T Z = I$ که $Z = (z_1, z_2, ..., z_k)$. برای جایگزینی به تابع هدف برش نرمال داریم:

$$J_{NC} = \sum_{\ell=1}^{k} \frac{h_\ell^T (D - W) h_\ell}{h_\ell^T D h_\ell} = \sum_{\ell=1}^{k} Z_\ell^T (I - \widetilde{W}) z_\ell$$

عبارت اول یک ثابت است بنابراین به حداقل رساندن آن دشوار است.

$$\max_{Z^T Z = I, Z \geq 0} Tr(Z^T \widetilde{W} Z) \tag{۴۱-۲}$$

این مطابق است با خوشه بندی K-means کرنلی معادله ۳-۱۱ هنگامی که معادله Z بدست آمد، ما میتوانیم H را با بهینه سازی تابع زیر بهبود دهیم.

$$\min_{H \geq 0} \sum_{\ell} \| \hat{Z}_\ell - \frac{D^{1/2} - h_\ell}{\| D^{1/2} h_\ell \|} \|^2 \tag{۴۲-۲}$$

راه حل های دقیق به صورت زیر هستند:

$$h_k = D^{-1/2} \bar{Z}_k$$

و یا

$$H = D^{-1/2}Z$$

. بنابراین سطر i از Z در یک ثابت $d_i^{-1/2}$ ضرب شده است. وزن نسبی درون خوشه ای متفاوت ، در همان سطر باقی می ماند. بنابراین H نشان دهنده خوشه بندی براساس Z است. □

قضیه ؟؟ نشان می دهد خوشه بندی طیفی به طور مستقیم به خوشه بندی K-means کرنلی مربوط است که طبق قضیه ؟؟ هم ارز با NMF است. بنابراین NMF ، خوشه بندی K-means کرنلی و خوشه بندی طیفی در یک روش متحد هستند: آنها نسخه های مختلف از همان مساله با محدودیت های کمی متفاوت هستند [۱۵].

۲-۸-۱ قضیه محدودیت NMF

تجزیه NMF با تجزیه SVD (تجزیه مقادیر منفرد) به علت عدم حذف علامتهای مثبت و منفی متفاوت است. اما اهمیت اساسی این عدم حذف چیست؟ این ویژگی یک محدودیت است. قضیه محدودیت، مبنای علمی برای نرمال سازی F و G را در $X = FG^T$ ارائه می دهد. ماتریس A محدود شده است اگر $۰ \leq A_{ij} \leq ۱$ توجه داشته باشید که برای هر ماتریس ورودی غیر منفی، می توانیم آن را به مقیاس کوچکتری محدود کنیم.

ویژگی محدودیتی NMF : اگر X محدود شود، تجزیه ماتریس های تجزیه ی F، G نیز باید محدود شوند. دقیق تر،

قضیه ۲-۸-۲ (قضیه محدودیت). فرض کنید $۰ \leq X \leq ۱$ ماتریس داده ورودی باشد. F و G ماتریس های غیرمنفی رضایت بخش هستند $X = FG^T$: یک ماتریس مورب D وجود دارد به طوری که وقتی $X = FG^T = F^*(G^*)^T$ $(FD)(GD^{-1})^T$ را ترسیم می کنیم:

$$۰ \leq F^*_{ij}, G^*_{ij}$$

. D به این صورت ساخته شده است:

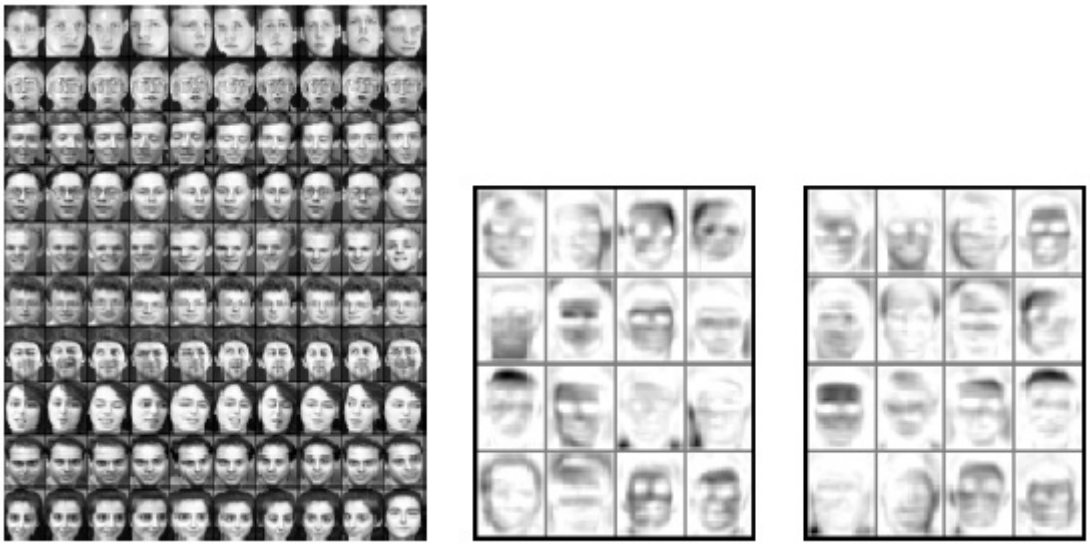
$$D = D_F^{1/2} D_G^{-1/2}, D_F = diag(f_1, \ldots, f_k), f_k = max_p F_{kp}$$

و

$$D_G = diag(g_1, \ldots, g_k), g_k = max_p G_{kp}$$

. این زمانی بدست می آید که X متقارن باشد.

$$X = HH^T, ۰ \leq H_{ij} \leq ۱$$



شکل ۲-۵: مجموعه داده های تصویری ORL شامل ۴۰۰ تصویر از ۴۰ نفر است. میانه / راست: تجزیه ی NMF محاسبه شده $F = (f_1, \ldots, f_{16})$ برای دو اجرا با شروع اولیه تصادفی.

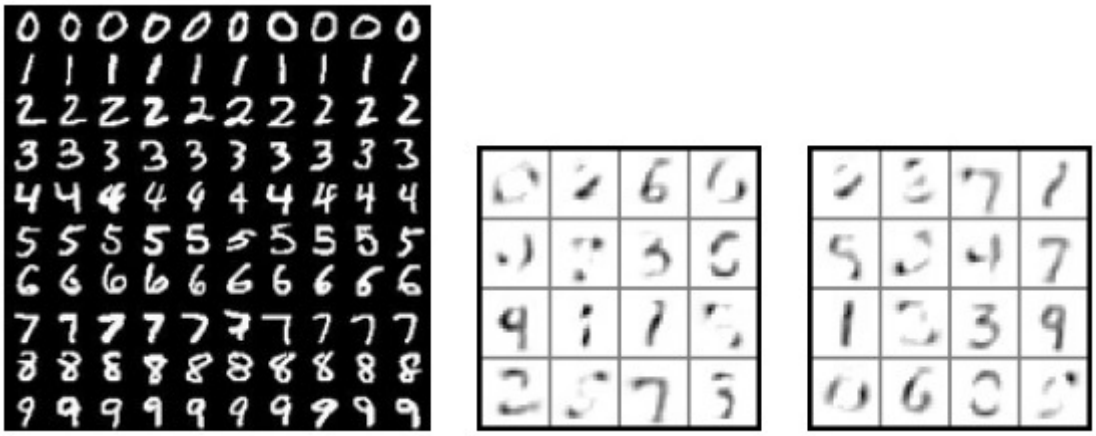
.

این قضیه وجود یک مقیاس مناسب را تأیید می کند که هم W و هم H محدود هستند، عناصر آنها نمی توانند از اندازه ماتریس داده ورودی تجاوز کند. توجه کنید که تجزیه SVD دارای ویژگی محدودیت نیست. در این مورد، حتی اگر داده های ورودی در محدوده $0 \leq X_{ij} \leq 1$، باشد، نمی توانیم تضمین کنیم که برای همه i,j و $|F^*_{ij}| \leq 1$و $|V^*_{ij}| \leq 1$ [۲۶].

۲-۹ قابلیت های خوشه بندی NMF

۲-۹-۱ مثالها

در ادامه قابلیت های خوشه بندی NMF با استفاده از چند نمونه ارائه شده است. شکل ۲-۵ و ۲-۶ عوامل موجود در F (تصاویر) را بر روی دو مجموعه داده محاسبه کرده است: ORL مجموعه داده های تصویر چهره و مجموعه داده تصویرعددی. در اینجا هر تصویر به عنوان یک بردار از مقادیر پیکسل خاکستری نمایش داده می شود. توجه داشته باشید که با استفاده از قضیه ؟؟ ، این عوامل تجزیه F مراکز خوشه بندی(نمایندگان خوشه ها) هستند. به طور مستقیم، این تصاویر نمایندگان خوشه های تصاویر اصلی هستند. مثال ها نشان می دهد که NMF یک نمای کلی از مجموعه داده ها را فراهم می کند.



شکل ۲-۶: چپ:مجموعه داده های تصویری عددی. میانه / راست:تجزیه ی NMF محاسبه شده $F = (f_1, \ldots, f_{16})$ برای دو اجرا با شروع اولیه تصادفی.

۲-۹-۲ تجزیه و تحلیل

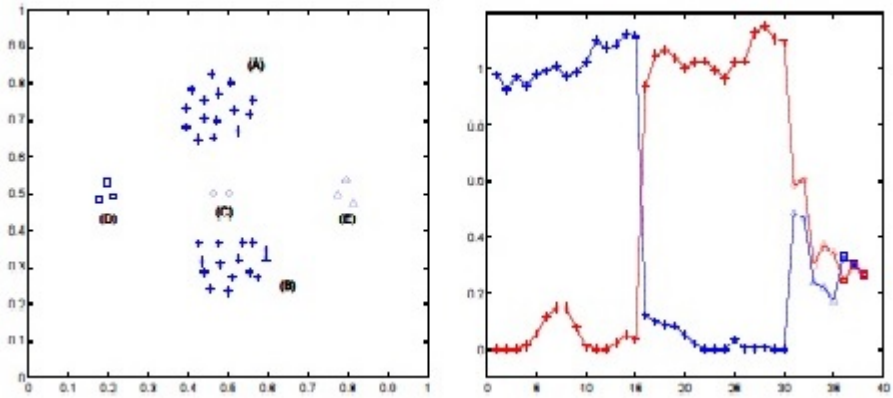
به طور کلی، مزایای NMF در سرتا سر روش های یادگیری بدون ناظر موجود را می توان در زیر خلاصه کرد. NMF می تواند به طور گسترده ای متغیر توزیع داده ها را به دلیل انعطاف پذیری تجزیه ماتریس به منظور بدست آوردن تابع هدف خوشه بندی k-means طراحی(مدل) کند. یکی دیگر از مزایای NMF این است که NMF می تواند به طور همزمان خوشه بندی انحصاری و هم پوشی را انجام دهد [۲۶].

شکل ۲-۷ یک مثال ارائه می دهد. مجموعه داده ی دو بعدی در زیر فریم سمت چپ شامل ۳۸ نقطه داده است و مقادیر G در زیر فریم سمت راست نشان داده شده است. همانطور که در شکل دیده می شود، مقادیر G برای نقاط در مناطق $\{C, D, E\}$ نشان می دهد که آنها به صورت کسر به خوشه های مختلف اختصاص داده می شوند.در نتیجه، NMF قادر به انجام خوشه بندی نرم است. [۱۸]

مزیت سوم این است که NMF قادر به انجام همزمان خوشه بندی ردیفها (ویژگیها) و ستونها ی (نقاط داده) یک ماتریس داده ورودی است. اهداف NMF زیر را با محدودیت های متعامد نرمال در تجزیه ماتریس ها محاسبه کنید:

$$J_{orth} = \| X - FG^T \|^2, \quad s.t. \quad F \geq 0, G \geq 0, F^T F = I, G^T G = I. \qquad (۴۳-۲)$$

محدودیت های متعامد نرمال و محدودیت های غیرمنفی در معادله ۲-۴۳ نتیجه گیری تقریب F و G نتایج خوشه بندی k-means در هر دو ویژگی و نقاط داده شده است. مزیت چهارم این است که بسیاری دیگر از مسائل داده کاوی و یادگیری ماشین را می توان به عنوان یک مسئله NMF اصلاح کرد.



شکل ۲-۷: زیر فریم سمت چپ: مجموعه داده دوبعدی از ۳۸ نقطه داده .زیر فریم سمت راست مقادیر $G = (g_1, g_2)$ به صورت منحنی های آبی و قرمز نشان داده شده است. نقاط در نواحی $\{A, B\}$ به ترتیب ۳۰-۱، جایی که مقادیر g_1 و g_2 به طور قابل توجهی متفاوت است، نشان می دهد که این نقاط به صورت خوشه ای به هر یک از دو خوشه A یا خوشه B (خوشه بندی انحصاری) خوشه بندی می شوند. نقاط درنواحی $\{C, D, E\}$ شماره های ۳۸-۳۱ هستند که مقادیر g_1 و g_2 محصور هستند، نشان می دهد که این نقاط به صورت کسر به خوشه های A و B (خوشه بندی هم پوشی) خوشه بندی می شوند.

فصل ۳

برخی الگوریتم های مربوط به NMF

۳-۱ مقدمه

معرفی الگوریتم های تجزیه ماتریس معمولا روش های به روز رسانی تکراری است: به روز رسانی یک عامل در حالی که عوامل دیگر را ثابت نگه میداریم. الگوریتم های موجود برای NMF شامل به روز رسانی های ضربی از لی و سانگ [۲۷، ۱۸، ۲۸] ، الگوریتم های کمترین مربعات (ALS) و ، حداقل مربعات غیر منفی متناوب می باشد. در این فصل ابتدا مسائل کاربردی در الگوریتم NMF را شرح می دهیم و سپس برخی از مهمترین الگوریتم های مربوط به NMF را بیان خواهیم کرد [۲۹، ۲۸].

۳-۲ مسائل کاربردی در الگوریتم NMF

۳-۲-۱ مقدار دهی اولیه

همانند خوشه بندی K-means ، مقداردهی اولیه نیز هنگام استفاده از NMF برای خوشه بندی نقش مهمی را ایفا می کند، زیرا تابع هدف ممکن است دارای مینیمم های محلی زیادی باشد. مینیمم سازی متناوب ذاتی در الگوریتم های NMF غیرقابل نفوذ است، حتی اگر تابع هدف با توجه به مجموعه ای از متغیرها محدب باشد.

یک مقداردهی تصادفی ساده که ماتریسهای عامل را به صورت ماتریسهای تصادفی آغاز می کنند، به طور کلی مؤثر نیست زیرا اغلب منجر به همگرایی آهسته به مینیمم محلی می شود. بسیاری از تکنیک های مختلف برای بهبود مقداردهی اولیه به صورت تصادفی توسعه داده شده است:

۱. ارزیابی های چندگانه: ایده اصلی این است که اجرای چندگانه را با استفاده از مقدار دهی اولیه مختلف تصادفی

انجام دهیم و بهترین تخمین ها را از چند اجرا انتخاب کنیم. این روش نیاز به انجام چندین بار NMF دارد و به صورت محاسباتی پرهزینه است.

۲. مقدار دهی اولیه مبتنی بر تجزیه: NMF یک تجزیه ماتریس رتبه پایین محدود شده است، بنابراین می توان نتایج حاصل از روش های تجزیه رتبه پایین متناوب را به عنوان مقدار اولیه استفاده کرد. به عنوان مثال می توان به SVD مبتنی بر مقداردهی اولیه اشاره کرد [۲۸].

۳. خوشه بندی مبتنی برمقداردهی اولیه: اگر NMF به عنوان یک فرایند خوشه بندی در نظر گرفته شود، می توان استراتژی مقداردهی اولیه را بر اساس نتایج بهینه از سایر الگوریتم های خوشه بندی جستجو کرد (به عنوان مثال، K-means محدب).

هر دو روش تجزیه مبتنی بر مقداردهی اولیه و خوشه بندی مبتنی بر مقداردهی اولیه، می تواند به کاهش سریع خطا و همگرایی سریع تر منجر شود [۱۲].

۳-۲-۲ معیارهای توقف

علاوه بر تعداد تکرارهای از پیش تعیین شده، معیارهای دیگری نیز برای توقف الگوریتم های تکراری NMF استفاده شود که عبارتند از:

۱. کاهش تابع هدف به یک آستانه معین؛

۲. وجود تغییرات کمی در ماتریس عامل یا تابع هدف بین تکرارهای پیوسته

Lin از شرایط توقف محدودیت بهینه سازی شده برای روش های تکراری NMF استفاده کرده است و کیم و پارک، معیار همگرایی ترکیبی را با استفاده از شرایط مطلوب KKT (Karush-Kuhn-Tucker) ، وشرایط بهینگی و همگرایی موقعیت های بزرگترین عناصر در ماتریسهای عامل مورد آزمایش قرار دادند.

بر اساس قضیه ۲-۸-۲ ، پس از اتمام الگوریتم های NMF، G مربوط به شاخص خوشه است. بنابراین تخصیص خوشه بندی می تواند توسط بزرگترین عنصر از هر ردیف در G تعیین شود. تخصیص خوشه بندی همچنین می تواند با پیدا کردن یک تابع عضویت خوشه بندی مجزا که نزدیک به G حاصل شده است(به G محدود شده است)، تعیین شود [۱۲].

الگوریتم ۳-۱ الگوریتم کاهش گرادیان

1: begin Initialize a, criterionθ, $\eta(0), k = 0$
2: do $k \leftarrow k + 1$
3: $a \leftarrow a - \eta(k)\nabla J(a)$
4: until $\eta(k)\nabla J(a) < \theta$
5: return a
6: end

۳-۲-۳ روشهای تعیین طول گام

۳-۲-۳-۱ روش کاهش گرادیان

رویکردی که برای یافتن راه حل برای مجموعه ای از نابرابری های خطی $a^t y > ۰$ خواهیم داشت، تعریف یک تابع معیار $J(a)$ است که اگر a یک بردار راه حل باشد، حداقل می شود. این مسئله اغلب میتواند توسط روش کاهش گرادیان حل شود پایه ی روش کاهش گرادیان بسیار ساده است با انتخاب بردار وزنی دلخواه $a(۱)$ شروع میکنیم و بردار گرادیان $\nabla J(a(۱))$ را محاسبه می کنیم. مقدار بعدی $a(۲)$ با حرکت $a(۱)$ در جهت سریعترین شیب، یعنی در امتداد منفی گرادیان بدست می آید. به طور کلی $a(k+۱)$ از $a(k)$ با استفاده از معادله زیر بدست می آید:

$$a(k+۱) = a(k) - \eta(k)\nabla J(a(k)) \qquad (۱-۳)$$

که η نرخ یادگیری است که طول گام را تعیین می کند. الگوریتم این روش در ۳-۱ آمده است. مسئله مهم در این جا انتخاب نرخ یادگیری $\eta(k)$ است. اگر $\eta(k)$ خیلی کوچک انتخاب شود همگرایی آهسته است در حالی که اگر $\eta(k)$ خیلی بزرگ انتخاب شود ممکن است واگرا شود.

یک روش اصولی برای تعیین نرخ یادگیری به این صورت است :

فرض بر این است که تابع معیار می تواند توسط بسط مرتبه دوم حول یک نقطه $a(k)$ تقریب زده شود:

$$J(a) \approx J(a(k)) + \nabla J^t(a - a(k) + \frac{۱}{۲}(a - a(k)^t H(a - a(k)) \qquad (۲-۳)$$

که H ماتریس هسین از مشتق جزئی مرتبه دوم $\frac{\partial^۲ J}{\partial a_i \partial a_j}$ است که در هسین $a(k)$ ارزیابی می شود. سپس $a(k+۱)$ معادله ۳-۱ را در معادله ۳-۲ جایگزین می کنیم.

$$J(a(k+۱)) \approx J(a(k)) - \eta(k) \parallel \nabla J \parallel^۲ + \frac{۱}{۲}\eta^۲(k)\nabla J^t H \nabla J$$

الگوریتم ۳-۲ الگوریتم نیوتن

1: begin Initialize a, criterionθ
2: do $k \leftarrow k + 1$
3: $a \leftarrow a - H^{-1}\nabla J(a)$
4: until $H^{-1}\nabla J(a) < \theta$
5: return a
6: end

که $J(a(k+1))$ می تواند با انتخاب

$$\eta(k) = \frac{\| \nabla J \|^2}{\nabla^t H \nabla J} \tag{۳-۳}$$

به حداقل برسد که H به a وابسته است و بنابراین به طور غیر مستقیم به k. این یک انتخاب بهینه برای $\eta(k)$ ارائه می دهد [۳۰].

۳-۲-۳-۲ روش نیوتن

یک روش جایگزین که با نادیده گرفتن معادله ۳-۱ و انتخاب $a(k+1)$ برای حداقل سازی بسط مرتبه دوم بدست آمده، الگوریتم نیوتن است که خط سوم الگوریتم ۳-۱ جایگزین شده با:

$$a(k+1) = a(k) - H^{-1}\nabla J \tag{۴-۳}$$

که به الگوریتم ۳-۲ منجر می شود. به طور کلی الگوریتم نیوتن در هر مرحله بهتر از الگوریتم کاهش گرادیان است، حتی با مقدار بهینه $\eta(k)$ را بهبود می بخشد با این حال الگوریتم نیوتن اگر ماتریس هسین H منحصر به فرد باشد قابل اجرا نیست. علاوه بر این حتی وقتی که H غیر منفرد باشد زمان مورد نیاز برای محاسبه معکوس ماتریس در هر تکرار $O(d^3)$ است که به سادگی می تواند مزیت نسبی را جبران کند و از معایب مهم این روش به شمار می رود [۳۰] . در الگوریتم های به کار رفته در این فصل از الگوریتم کاهش گرادیان برای محاسبه طول گام استفاده شده است.

۳-۳ الگوریتم های مربوط به NMF

۳-۳-۱ قواعد به هنگام ضربی

پرکاربرد ترین رهیافت برای مساله

$$\min_{W\geq ۰,H\geq ۰} f(W,H) = \frac{۱}{۲} \parallel A - WH \parallel_F^۲ \qquad (۵-۳)$$

را می توان قواعد به هنگام ضربی [۲۸، ۲۷] دانست که توسط لی و سانگ در سال ۲۰۰۱ پیشنهاد گردید. تابع هزینه مسئله NMF نامحدب است و لذا نمی توان نقطه کمینه سراسری را پیدا کرد. اما روشهایی در بهینه سازی عددی وجود دارند که برای پیدا کردن کمینه ی موضعی به کار می روند. شاید روشهای گرادیان از ساده ترین آنها باشند. به طور کلی، الگوریتم لی و سانگ را می توان یک روش کاهش گرادیان دانست که در قالب یک فرم متناوب برای بهنگام کردن مولفه های عوامل ماتریسی از قواعد

$$W_{ia} \leftarrow W_{ia} - \eta_{ia}(\nabla_W f(W,H))_{ia}, H_{bj} \leftarrow H_{bj} - \mu_{bj}(\nabla_H f(W,H))_{bj}, \qquad (۶-۳)$$

در خلاف جهت گرادیان استفاده می کند. در اینجا برای اینکه شرط غیر منفی بودن تضمین شود داریم:

$$W_{ia} - \eta_{ia}(WHH^T)_{ia} = ۰$$

در نتیجه

$$\eta_{ia} = \frac{W_{ia}}{(WHH^T)_{ia}}$$

به عنوان طول گام برای بروزرسانی ماتریس W و همینطور

$$H_{bj} - \mu_{bj}(W^TWH)_{bj} = ۰$$

در نتیجه

$$\mu_{bj} = \frac{H_{bj}}{(W^TWH)_{bj}}$$

به عنوان طول گام برای بروزرسانی ماتریس H در نظر گرفته می شود [۲۸]. با محاسبه گرادیان رابطه (۳-۵) نسبت به W و H داریم:

$$f = (A^T - H^T W^T)(A - WH) = A^T A - A^T WH - H^T W^T A + H^T W^T WH$$

$$\nabla_H = -W^T A - W^T A + ۲W^T WH = ۲W^T WH - ۲W^T A = W^T WH - W^T A$$

$$\nabla_W = -AH^T - AH^T + ۲WHH^T = ۲WHH^T - ۲AH^T = WHH^T - AH^T$$

حال با جایگذاری در رابطه (۳-۶) داریم:

$$W_{ia} \leftarrow W_{ia} - \frac{W_{ia}}{(WHH^T)_{ia}}(WHH^T - AH^T)_{ia}$$

$$W_{ia} = \frac{W_{ia}(WHH^T)_{ia} - W_{ia}(WHH^T)_{ia} + W_{ia}(AH^T)_{ia}}{(WHH^T)_{ia}}$$

$$H_{bj} \leftarrow H_{bj} - \frac{H_{bj}}{(W^T WH)_{bj}}(W^T WH - W^T A)_{bj}$$

$$H_{bj} = \frac{H_{bj}(W^T WH)_{bj} - H_{bj}(W^T WH)_{bj} + H_{bj}(W^T A)_{bj}}{(WTWH)_{bj}}$$

در نهایت W_{ia} و H_{bj} از رابطه زیر بدست می آیند:

$$W_{ia} \leftarrow W_{ia}\frac{(AH^T)_{ia}}{(WHH^T)_{ia}}, H_{bj} \leftarrow H_{bj}\frac{(W^T A)_{bj}}{(W^T WH)_{bj}} \qquad (۳-۷)$$

همگرایی روشهای مبتنی بر گرادیان، به انتخاب طول گام حساس است. لی و سانگ برای کمینه کردن نرم فروبنیوس $\| A - WH \|_F^۲$ ، قواعد به هنگام ضربی (۳-۷) را پیشنهاد کردند. روشن است که وقتی $A = WH$ ، هر یک از ضرایب این قواعد برابر با یک خواهند شد؛ به این معنی که $\nabla_W f(W,H) = ۰$ و $\nabla_H f(W,H) = ۰$. تحت این شرایط لی و سانگ ثابت کردند که نرم فروبنیوس $\| A - WH \|_F^۲$ تحت قواعد (۳-۷) ناافزایشی است.البته نسخه اصلی این قضیه شامل قسمت دومی نیز هست که بعدا معلوم شد لزوما درست نیست. لی و سانگ در قسمت دوم نسخه اصلی قضیه ادعا کردند که این نرم تحت قواعد (۳-۷) ، در یک نقطه ایستا تغییر نمی کند به این دلیل که الگوریتم لی و سانگ، یک روش نقطه ثابت است و لذا اگر با شروع از $H > ۰$ به نقطه ایستای $H^* > ۰$ همگرا شود، آنگاه $\nabla_H f(W^*, H^*) = ۰$ اما علی رغم مثبت بودن عامل ماتریسی در همه تکرارها، ممکن است برخی مولفه های آن به صفر میل کنند، زیرا ما با دستگاه هایی سر و کار داریم که اعداد در آنها دارای دقت محدود هستند و در نتیجه اعداد بسیار کوچک، به صفر گرد می شوند. به عبارت دیگر، مولفه های صفر به هنگام نمی شوند. حال اگر در یکی از تکرارها، مولفه ای مانند H_{bj} صفر شود، آن گاه

الگوریتم ۳-۳ الگوریتم Update Multiplicative

1: Initialize: $W_{ia}^{1} > 0, H_{bj}^{1} > 0, \forall i, a, b, j$

2: For $K = 1, 2, \ldots$

$$W_{ia}^{k+1} = W_{ia}^{k} \frac{(V(H^{k})^{T})_{ia}}{(W^{k}H^{k}(H^{k})^{T})_{ia}}, \forall i, a,$$

$$H_{bj}^{k+1} = H_{bj}^{k} \frac{((W^{k+1})^{T}V)_{bj}}{((W^{k+1})^{T}W^{k+1}H^{k})_{bj}}, \forall b, j.$$

$H^{*bj} = ۰$. با توجه به شرط $\nabla_H f(W^*, H^*)_{bj} \geq ۰$ ، معلوم نیست قاعده به هنگام چگونه باید عمل کند، زیرا وقتی $\nabla_H f(W^*, H^*) \neq ۰$ ، روش باید نقطه فعلی را به هنگام کند که در این صورت حتما بهینه بودن جواب فعلی نقض خواهد شد. حتی گاهی ممکن است با صفر شدن مولفه ای، مقدار گرادیان منفی شود. کلیه ی استدلال های بیان شده، برای قاعده دوم نیز برقرار است. با دلایل ذکر شده، الگوریتم لی و سانگ دارای همگرایی قوی نیست. معمولا برای جلوگیری از مشکلات عددی، عدد کوچک مثبتی مانند $\epsilon = ۱۰^{-۹}$ به مخرج کسر هر یک از قواعد اضافه می کنند. الگوریتم لی و سانگ [۲۸] برای تابع هزینه نرم دار به این صورت معرفی می شود که ابتدا ماتریس‌های $W^k \geq ۰$ و $H^k \geq ۰$ به صورت تصادفی مقداردهی اولیه می‌شوند و مقدار نهایی این ماتریس‌ها با چندین تکرار قوانین به روزرسانی لی و سانگ با افزودن مقدار ϵ به مخرج کسرها بدست می‌آید. الگوریتم کامل این روش در ۳-۳ آمده است.

۳-۳-۲ روش کمترین مربعات متناوب

با توجه به نامحدب بودن تابع هزینه در مسئله NMF تبدیل این مسئله به دو زیرمسئله محدب و در نتیجه استفاده از ویژگی های بهینه سازی محدب منطقی به نظر می رسد. در اینجا با ثابت نگه داشتن یکی از عوامل ماتریسی، می توان عامل ماتریسی دیگر را از یک مسئله کمترین مربعات به دست آورد. به این ترتیب در روش کمترین مربعات متناوب با قید نامنفی (ANLS) مسئله ی ۳-۵ به دو مسئله ی کمترین مربعات با قید نامنفی به صورت

$$\min_{H \geq ۰} \frac{۱}{۲} \parallel A - WH \parallel_F^۲ \qquad (۸-۳)$$

با عامل ثابت W و

$$\min_{W \geq ۰} \frac{۱}{۲} \parallel A^T - H^T W^T \parallel_F^۲ \qquad (۹-۳)$$

با عامل ثابت $(H^T)H$ بازنویسی می شود و تا برقراری شرط همگرایی، به طور تناوبی به دنبال هم می آیند. از آنجا که عامل سمت راست در زیرمسئله ی ۳-۸و۳-۹ یک ماتریس است، در ادامه از نام NLS ماتریسی برای آن ها استفاده می شود [۲۸].

در سال ۲۰۰۷ ، بری و همکاران وی الگوریتم پایه ی ALS را برای NMF بیان کردند. در این الگوریتم، هریک از زیرمسئله های ۳-۸ و ۳-۹ به صورت LS نامقیدحل می شود و در صورت به وجود آمدن درآیه های منفی در عامل ماتریسی، مقدار صفر جایگزین آن درآیه ها خواهد شد. این عمل، ضمن برقراری شرط نامنفی بودن عوامل ماتریسی تولیدشده، به تنکی آنها نیز کمک می کند. برای پیاده‌سازی روش NNLS ابتدا ماتریس $W^k \geq ۰$ مقداردهی اولیه می‌شود. سپس طی چندین تکرار در یک حلقه مسئله ۳-۸ به صورت نامقید برای عامل ماتریسی H حل شده، درایه های منفی ماتریس H با مقدار صفر(عمل تصویر کردن) جایگزین می‌شوند سپس مسئله ۳-۹ به صورت نامقید برای عامل ماتریسی W حل شده و درایه های منفی ماتریس W با مقدار صفر(عمل تصویر کردن) تعویض می‌شوند و این حلقه تا برقراری شرط توقف تکرار می‌شود پیاده سازی کامل این روش در الگوریتم ۳-۴ به طور کامل ارائه شده است. اگرچه این الگوریتم موجبِ تسریع محاسبات می شود، اما تحلیل همگرایی برای آن سخت خواهد بود، زیرا قیدهای نامنفی در دو زیرمسئلهٔ محدب دقیقاً برقرار نمی شوند. در حالت کلی، این الگوریتم به نقطهٔ ایستا همگرا نیست و از آن به طور مستقیم استفاده نمی شود. معمولا این الگوریتم به عنوان یک راه انداز برای الگوریتم های دیگر به کار می رود.هر NLS ماتریسی به ترتیب با ثابت در نظر گرفتن W و ،H با مسائل

$$\min_{H_{:j}\geq ۰} \frac{۱}{۲} \sum_j \| A_{:j} - WH_{:j} \|_۲^۲, \qquad (۱۰-۳)$$

و

$$\min_{W_{:j}^T\geq ۰} \frac{۱}{۲} \sum_j \| A_{:j}^T - H^T W_{:j}^T \|_F^۲, \qquad (۱۱-۳)$$

معادل است. از این رو زیرمسئله‌های ۳-۸ و ۳-۹ به ترتیب به n و m مسئله NLS برداری تبدیل می شود. نماد $X_{:j}$ را برای نشان دادن ستون j ام ماتریس X به کار می بریم. برای حل هر NLS برداری، می توان از الگوریتم NNLS [۳۱] با نام کاربردی مجموعه فعال استفاده کرد.

ناحیه پذیرفتنی یک LS مقید با قید نامنفی، همان قید مفروض است. با این فرض هر جواب مسئله NLS برداری باید در ناحیه پذیرفتنی باشد و هر جواب پذیرفتنی دارای دو رده از متغیرها است که توسط آنها شناسایی می شود. رده ی اول، شامل متغیرهایی است که در درون ناحیه پذیرفتنی و رده ی دوم، شامل متغیرهایی است که بر روی مرز ناحیه پذیرفتنی قرار دارند. مجموعه غیرفعال، توسط اندیس NLS متناظر با متغیرهای رده ی اول و مجموعه فعال توسط اندیس متناظر با متغیرهای رده ی دوم شناخته می شوند. سازوکار الگوریتم مجموعه فعال در حل مسئله برداری این گونه است که برای هر جواب پذیرفتنی آغازین، متغیرها به طور تکراری بین مجموعه فعال و غیرفعال جابه جا می شوند. در چنین حالتی، تابع هزینه به طور پیوسته

الگوریتم ۳-۴ الگوریتم NNLS

1: $P = \phi$
2: $R = \{1, 2, \ldots, M\}$
3: $d = 0$
4: $w = Z^T(x - Zd)$
B. Main loop
5: Proceed if $R \neq \phi \wedge [\max_{n \in R}(w_n) > tolerance]$
6: $m = argmax_{n \in R}(w_n)$
7: Include the index m in P and remove it from R
8: $s^p = [(Z^p)^T Z^p]^{-1}(Z^p)^T X$
C. Inner loop
9: Proceed $min(s)^p \leq 0$
10: $\alpha = -\min_{n \in P}[d_n/(d_n - s_n)]$
11: $d := d + \alpha(s - d)$
12: Update R and P
13: $s^p = [(Z^p)^T Z^p]^{-1}(Z^p)^T X$
14: $s_R = 0$
end C
15: $d = s$
16: $w = Z^T(x - Zd)$
end B

و یکنواخت کاهش پیدا می کند در حالی که پذیرفتنی بودنِ جواب، محفوظ می ماند.

روشن است که استفاده از الگوریتم مجموعه فعال برای همه NLS های برداری، حتی با انجام محاسبات خیلی سریع، منجر به اجرایی کند و غیر قابل قبول می شود. به منظور ارائه جایگزینی که برای حل زیرمسائل ۳-۱۰ و ۳-۱۱ مناسب باشد، می توان به اقدامات کیم و پارک در استفاده از روش مجموعه فعال ترکیبیاتی FC-NNLS و نیز اقدامات لین در استفاده از روش تصویر گرادیان اشاره کرد.

الگوریتم FC-NNLS توسط بنتم و کنین برای حل NLS ماتریسی پیشنهاد گردید. در حقیقت، این الگوریتم استفاده مستقیم از الگوریتم مجموعه فعال است. همانطور که اشاره شد هر NLS ماتریسی را می توان به چندین NLS برداری تبدیل کرد. تعداد NLS های برداری به تعداد ستون های ماتریس ضرایب زیرمسائل بستگی دارد. لذا می توان به جای اجرای الگوریتم تکراری مجموعه فعال برای هر NLS برداری، فقط یک تکرار از این الگوریتم را برای یک NLS ماتریسی به کار برد. در این استراتژی ترکیبیاتی، ستون هایی از ماتریس ضرایب که دارای مجموعه های غیرفعال یکسان اند با هم در نظر گرفته می شوند و مسئله به ازای متغیرهای متناظر با مجموعه غیرفعال، حل می شود. ساختار دقیق این الگوریتم همراه با شبه کد متلب آن در ادامه آمده است [۲۸].

۳-۳-۳ الگوریتم FC-NNLS

در یک سطح بسیار اساسی، روش مجموعه فعال / غیرفعال، مسئله کمترین مربعات را به عنوان دنباله ای از مسائل محدود شده برابری حل می کند. الگوریتم کلی NNLS مسئول تعریف این توالی است. این سرعت راه حل را کنترل می کند و جزئیات هر یک از مسائل فردی را با انتخاب متغیرهای خاص غیر فعال که باید در هر مرحله از توالی تخمین زده شود، تعریف می کند. در ستون موازی، رویکرد ترکیبی، ستون های K که مجموعه های غیر فعال آنها یکسان هستند، شناسایی می شود، پشته های مناسب محاسبه شده و برای متغیرهای غیرفعال حل می شود. در حالی که نمونه های پیشین در مورد شرایط عملکرد تعداد محاسبات شبه معکوس مورد بحث قرار گرفته اند، قطعا هزینه هایی برای شناسایی و دسته بندی مجموعه های غیر فعال یکسان وجود دارد. انجام کارآیی این فعالیت برای دستیابی به عملکرد خوب الگوریتم FC-NNLS کلی ضروری است. [۲۹].

در مرحله اولیه ی الگوریتم FC-NNLS ، بخش هایی از شبه معکوس را که قبلا در طول محاسبات ثابت باقی می ماند، پیش پردازش می شود. سپس از آنها برای حل مسئله کمترین مربع بدون محدودیت، RHS ضربی استفاده می شود. این توصیف عمومی است، از آنجا که بسیاری از روش های موجود برای انجام کمینه سازی (به عنوان مثال، معادلات نرمال و تجزیه ی متعامد) وجود دارد از راه حل بدون محدودیت برای مقداردهی الگوریتم استفاده می شود. این متغیرهای مجموعه اولیه غیر فعال را به سادگی با شناسایی مکان های مقادیر غیر منفی در Kو ساخت یک مجموعه منفعل برای هر ستون تعیین می کند.

همچنین ستون های Kرا که راه حل های آنها مطلوب نیست، به طور خاص، ستون هایی با مقادیر منفی نشان می دهد. این شاخص های ستون در مجموعه Fقرار می گیرند که در طول الگوریتم لیست فعلی شاخص های ستون برای راه حل های غیرقطعی را حفظ خواهد کرد. مجموعه کمکی مجموعه H یک زیر مجموعه ای از Fاست که حاوی شاخص هایی برای راه حل هایی است که در حال حاضر غیر قابل اجرا هستند. باقی مانده ی الگوریتم، یک ساختار حلقه اصلی / حلقه داخلی دارد که مشابه الگوریتم اصلی NNLS لاوسون و هانسون است [۲۹].

حلقه اصلی برای تولید محاسبات کوچکترین مربعات با توجه به مجموعه غیر فعال فعلی و آزمایش آنها برای بهینه سازی کار می کند. در صورتی که یک راه حل بهینه نیست، یک متغیر را که از مجموعه فعال به مجموعه غیر فعال برای تکرار بعدی منتقل می شود، انتخاب می کند.

حلقه درونی برای این واقعیت است که راه حل کمترین مربعات بدون محدودیت برای یک متغیر مجموعه غیر فعال ممکن است در برخی موارد منفی شود و مکانیزمی برای انتقال متغیرها از مجموعه غیر فعال به مجموعه فعال فراهم می کند. مجموعه H نشان دهنده راه حل های نامناسب است و حلقه داخلی هنگامی که H خالی می شود پایان می یابد.

معیار بهینه و همچنین جزئیات انتخاب متغیر و استراتژی های جنبش در جاهای دیگر کاملا توضیح داده شده است. هنگامی که یک ستون حاوی یک راه حل بهینه است، شاخص آن از مجموعه F حذف می شود. الگوریتم هنگامی که F خالی می شود، پایان می یابد.

الگوریتم FC-NNLS شامل یک زیرمجموعه ای به نام الگوریتم CSSLS (کمترین مربع زیر فضای ترکیبی) است که استراتژی ترکیبی را برای به دست آوردن راه حل برای متغیر مجموعه غیر فعال استفاده می کند. جزئیات الگوریتم CSSLS به زودی بعد از الگوریتم FC-NNLS دنبال می شود.

نوآوری عمده ای که به صراحت در الگوریتم تکرار منفرد الگوریتم FC-NNLS برای تمام راه حل های غیرقطعی نشان داده شده توسط مجموعه ی F قبل از شروع تکرارهای بعدی تکمیل می شود. جزئیات در مراحل ۱۲، ۱۵ و ۱۷ را که آزمایش بهینه سازی و نحوه انتقال متغیرها بین مجموعه های فعال و غیر فعال را توضیح می دهد حذف شده است. این جزئیات، با این که مهم هستند، منحصرا مربوط به نوآوری الگوریتم جدید نیست [۲۹].

الگوریتم CSSLS موتور FC-NNLS است که ، مجموعه ای از مجموعه های غیر فعال یکسان را شناسایی و پیدا می کند. سپس، راه حل کوچکترین مربع اجرا می شود به گونه ای که محدودیت برابر صفر روی متغیرهای مجموعه فعال اعمال شود. زیرمجموعه های ماتریس K را که می خواهیم در معیار کمترین مربعات در مراحل ۹ و ۱۳ الگوریتم FC-NNLS برآورد کنیم در نظر بگیرید. این کار را با استفاده از زیرمجموعه های اندازه گیری شده با جذب در مجموعه های A و مجموعه غیرفعال در P و محدودیتهای C انجام خواهیم داد برای وضوح، برای لحظه ای، شاخص های مجموعه ای را در K و A و P قرار می دهیم. پس از مقدار دهی اولیه K به صفر، یک مجموعه جدید M که حاوی شاخص های ستون K است تعریف می کنیم و آن را به زیر مجموعه های k مربوط به k مجموعه ی غیر فعال منفرد در P تقسیم می کنیم.

سپس برای هر مجموعه غیر فعال منحصر به فرد، یک راه حل کوچکترین مربعات را برای متغیرهای مجموعه غیر فعال با استفاده از زیر مسائل ساخته شده از ستون های C و ردیف K مطابق با متغیرهای مجموعه غیر فعال انجام می دهیم، و ستون های K و A مربوط به همه شاخص ستونهایی که یک مجموعه غیر فعال دارند. روال راه حل بدون محدودیت چندین مرتبه به عنوان مجموعه های غیر فعال منحصر به فرد در P اجرا می شود [۲۹].

الگوریتم ۳-۵ الگوریتم FC-NNLS

1: Index the columns (RHS) of the solution matrix: $M := \{1, 2, \ldots, p\}$
2: Index the rows of the solution matrix: $N := \{1, 2, \ldots, L\}$
3: Precompute constant parts of the pseudoinverse, e.g., $C^T C$ and $C^T A$
4: Compute $\min_k \parallel CK - A \parallel_F$ the unconstrained estimate for K
5: Initialize the set of passive sets: $P = \{p_1 p_2 \ldots p_p\}$, where $P_j = \{x \in N : xK_j > 0\}$
6: Find the set of columns that is yet to be optimized: $F = \{j \in M : p_j \neq N\}$
7: Complete the overwriting solution: $\forall j \in F, \forall x \notin p_j, xK_j = 0$
Main loop
8: if $F = \phi$,go to step 19, you have finished
9: Compute $\min_{k_F} \parallel CK_F - A_F \parallel$ using algorithm CSSLS and P_F
Inner loop
10: Put indices of columns with negative variables into set $H : H = \{j \in F : \min_{x \in p_j}(xK_j < 0)\}$
11: if $H = \phi$, go to step 15
12: $\forall i \in H$, select the variables to move out of the passive sets P_H
13: Compute $\min_{k_H} \parallel CK_H - A_H \parallel$ using algorithm CSSLS and P_H
14: Go to step 10
15: $\forall i \in F$, test solution K_i for optimality
16: Remove from F indices of columns whose solutions are optimal
17: $\forall i \in F$, select the variables tomove into the passive sets P_F
18: Go to step 8
19: End of computation

الگوریتم ۳-۶ الگوریتم CSSLS

1: Initialize: $K = 0$
2: Index the columns of $K : M = \{1, 2, \ldots, p\}$
3: Find the set of k unique passive sets: $U = \{U_1 U_2 \ldots U_k\} = \text{unique}(P)$
4: Index the columns of K with identical passive sets into ε $\varepsilon_j = \{i \in M : P_i = U_j\}$
5: $\forall j \in \{1, 2, \ldots, k\}$, solve: $\min_{u_j K_{\varepsilon_j}} \parallel C_{U_j U_j} K_{\varepsilon_j} - A_{\varepsilon_j} \parallel$
6: End computation

فصل ۴

آزمایشات ونتایج

در این فصل ابتدا یک روش NMF جدید را معرفی می کنیم. سپس برخی روشهای نمونه برداری را به اختصار بیان می کنیم و در نهایت روشهای NMF و NMF جدید و روش SVM را روی مجموعه داده ORL و مجموعه داده گلهای زنبق (iris) با هم مقایسه می کنیم.

۴-۱ روش NMF پیشنهادی

در این بخش یک الگوریتم جدید NMF با سرعت همگرایی سریع و عملکرد بالا را معرفی خواهیم کرد. در این روش، با توجه به روش کاهش گرادیان، $\tilde{\rho}_k$ را به عنوان طول گام در نظر می گیریم که نسبت به طول گام در روش قواعد به هنگام ضربی معرفی شده در فصل قبل محدودتر است. فرمولهای بروزرسانی غیرمنفی پیشنهادی به فرم درجه دوم هستند و به صورت زیر بیان می شوند:

$$H^{k+1} = H_k \odot (1 - \epsilon_k) Q_H^k + \epsilon_k (Q_H^k)^2, \tag{۴-۱}$$

$$W^{k+1} = W_k \odot ((1 - \epsilon_k) Q_W^k + \epsilon_k (Q_W^k)^2), \tag{۴-۲}$$

به طوری که

$$Q_H^k = \frac{(W_k)^T V}{(W^k)^T W^k H^k}, \quad Q_W^k = \frac{V (H^{k+1})^T}{W^k H^{k+1} (H^{k+1})^T}$$

که $A^{۲}$ به $A \odot A$ اشاره می کند و ϵ پارامتری است در بازه $[۰, ۱]$ و به صورت:

$$\epsilon_k = ۱ - ۱/k$$

محاسبه می شود که متغیر k معرف تکرار kام الگوریتم برای تجزیه NMF است.

۴-۱-۱ تعیین طول گام

برای تعیین طول گام دو شرط زیر باید برقرار باشد:

۱. $\tilde{\rho_k} \geq \rho_k$

۲. محدودیت غیرمنفی باید حفظ شود.

$\tilde{\rho_k}$ را با بازسازی معادله زیر بدست می آوریم:

$$H^k - \rho_k \odot (W^k)^T W^k H^k = ۰$$

در نتیجه $\tilde{\rho_k}$ با استفاده از رابطه زیر محاسبه می شود:

$$\tilde{\rho_k} = \rho_k + \epsilon_k \frac{(W^k)^T V \odot H^k}{((W^k)^T W^k H^k)^{۲}}$$

اثبات به طور کامل در مرجع [۳۲] آمده است. در نتیجه

$$\tilde{\rho_k} \geq \rho_k$$

این بدان معناست که NMF پیشنهادی نرخ همگرایی سریع تری دارد.

۴-۲ انتخاب مقدار r

تاکنون هیچ روش مشخص و ثابتی برای محاسبه مقدار r معرفی نشده است. دو رویکرد رایج عبارتند از: آزمایش و خطا که در این روش مقادیر مختلف r آزمایش می شود و مقداری انتخاب می شود که به ازای آن مسئله کمترین میزان خطا را داشته باشد. روش دیگری که می توان به کار برد برآورد با استفاده از تجزیه svd است. به این صورت که ابتدا تجزیه svd را

برای مسئله مورد نظر بدست آورده سپس با توجه به مقادیر ویژه ماتریس S، تعداد بزرگترین مقادیر ویژه را به عنوان r در نظر گرفت. این روش خصوصا برای ماتریسهای کوچک خوب جواب می دهد از این روش انتخاب r در تجزیه داده های iris استفاده شده است [۳۳].

۳-۴ نمونه برداری

نمونه برداری یکی از روشهای اصلی برای انتخاب داده هاست. در این روش از میان تمامی داده هایی که در مجموعه داده وجود دارند، با توجه به اندازه مجموعه داده نمونه [۱] تعدادی از داده ها انتخاب خواهند شد. هر چه اندازه مجموعه داده نمونه عدد بزرگتری باشد، الگوها و نظم های موجود بهتر شناسایی می شوند.

به این دلیل که پردازش کل داده های موجود در مجموعه داده برای الگوریتم های داده کاوی بسیار زمان گیر است، از نمونه برداری استفاده می شود. یک زیرمجموعه داده نمونه برداری شده باید خصوصیات مهم و مورد توجه را در مجموعه داده اولیه داشته و نماینده مناسبی برای کل مجموعه داده باشد. انواع الگوریتم های نمونه برداری از دو بعد قابل بررسی هستند. این ابعاد عبارتند از: توزیع دسته ها و ملاحظات جایگذاری.

از بعد توزیع دسته ها انواع الگوریتم های نمونه برداری عبارتند از:

- نمونه برداری تصادفی [۲] : در این روش به صورت تصادفی و بر اساس تعداد نمونه های موجود و بر اساس اندازه مجموعه داده نمونه، عملیات نمونه برداری انجام می شود. در این روش احتمال انتخاب هر رکورد برای مجموعه داده نمونه نهایی یکسان است.

- نمونه برداری متوازن [۳] : در این روش کل مجموعه داده طبقه بندی می شود و سعی خواهد شد تا توازن تعداد نمونه ها در قطعات مختلف پیش و پس از عملیات نمونه برداری حفظ شود. از این روش به ویژه در مسائل دسته بندی که توزیع یا فراوانی دسته ها تفاوت چشمگیری با یکدیگر دارند استفاده می شود [؟].

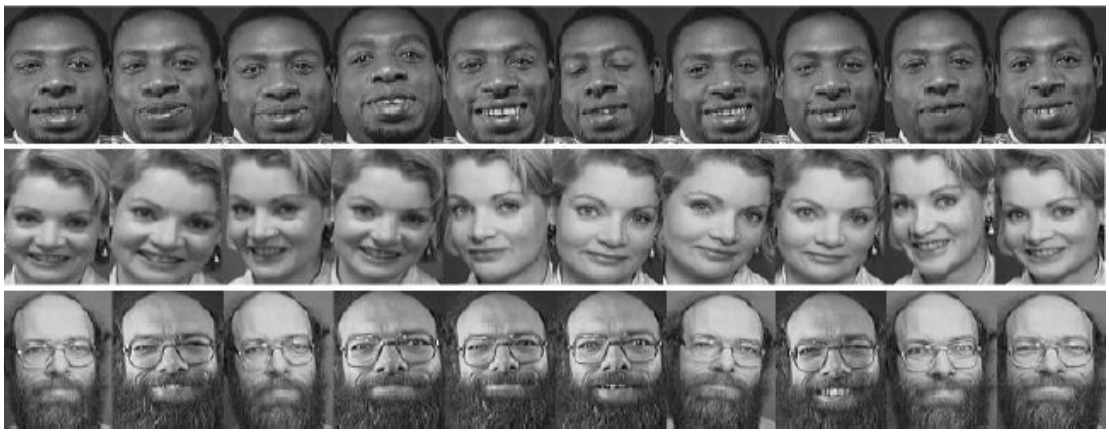
از بعد ملاحظات جایگذاری انواع الگوریتم های نمونه برداری عبارتند از:

- نمونه برداری بدون جایگذاری [۴] : در این روش هر رکوردی که برای مجموعه داده نمونه نهایی انتخاب می شود، از مجموعه داده اصلی حذف شده و احتمال انتخاب مجدد آن رکورد وجود نخواهد داشت. اگر نسبت تعدادکل رکوردها به اندازه مجموعه داده نمونه عدد بزرگی باشد(اندازه مجموعه داده نمونه کوچک باشد) بهتر است از روش نمونه برداری بدون جایگذاری استفاده شود.

- نمونه برداری با جاگذاری [۵] : در این روش هنگامی که یک رکورد برای مجموعه داده نمونه نهایی انتخاب شد، از مجموعه داده اصلی حذف نمی شود و احتمال انتخاب مجدد آن رکورد وجود دارد. اگر نسبت تعداد کل رکوردها به

---

[۱] Sample Size [۲] Simple Random Sampling [۳] Stratified Sampling [۴] Sampling without Replacement [۵] Sampling with Replacement



شکل ۴-۱: نمونه تصاویر بانک ORL

مجموعه داده نمونه عدد کوچکی باشد (اندازه مجموعه داده نمونه بزرگ باشد) بهتر است از روش نمونه برداری با جایگذاری استفاده شود [؟].

۴-۴ مجموعه داده های مورد استفاده

در این فصل از دو مجموعه داده ORL و iris استفاده شده است که هر یک را به اختصار شرح می دهیم.

۴-۴-۱ بانک داده ORL

بانک داده ORL بین سالهای ۱۹۹۲ تا ۱۹۹۴ میلادی تهیه شده است. تصاویر این بانک داده تغییراتی از قبیل شدت روشنایی، حالتهای چهره (چشمان باز و بسته، لبخند)، تصاویر با عینک و بدون عینک، مقیاس و فاصله و چرخش زاویه سر(اندک) را شامل می شود. اگرچه تهیه تصاویر به صورت سیستماتیکی صورت نپذیرفته است اما از این بانک داده تقریبا در تمامی پژوهشهای شناسایی چهره استفاده شده است. نمونه هایی از تصاویر این بانک داده در شکل ۴-۱ نشان داده شده است. در جدول ۴-۱ اطلاعات عمومی این بانک داده شامل تعداد افراد، شرایط تصویربرداری، ابعاد و تعداد تصاویر و آدرس اینترنتی آورده شده است.

جدول ۴-۱: مشخصات بانک داده ORL

تعداد افراد	ابعاد تصاویر	شرایط تصویر برداری	تعداد تصاویر
۴۰	۶۴*۶۴	pose	۴۰۰

۴-۴-۲ مجموعه داده iris

این مجموعه در واقع مجموعه‌ای از داده‌ها می‌باشد که شامل سه نمونه گل زنبق است؛ که توسط فیشر در سال ۱۹۳۶، برای نشان دادن تکنیک‌های خطی تفکیک پذیر معرفی گردید. از این رو به نام مجموعه داده گل زنبق فیشر نیز خوانده می‌شود. از طرف دیگر، به دلیل اینکه ادگار اندرسون نیز این مجموعه را به دلیل کیفیت تنوع جغرافیایی در شبه جزیره گاسپه، گردآوری کرده است، به مجموعه داده زنبق اندرسون نیز مشهور می‌باشد. این مجموعه شامل ۵۰ نمونه از سه نوع گل زنبق با نام‌های Setosa ، Virginica و Versicolor می‌باشد. شاخصه‌هایی که در این مجموعه جهت اندازه گیری مدنظر می‌باشند؛ پهنای گلبرگ، طول گلبرگ، پهنای کاسبرگ و طول کاسبرگ می باشد.

۴-۵ آزمایشات و نتایج

در این بخش آزمایشات انجام شده و نتایج آنها، ارائه شده و با روش SVM مقایسه شده است.

۴-۵-۱ مقایسه روش SVM و روش NMF روی پایگاه داده ORL

در این بخش روشهای NMF و SVM روی مجموعه تصاویر ORL جهت طبقه بندی تصاویر اعمال شده است. آزمایشات انجام شده برای هر نمونه داده چندین بار تکرار شده و در نهایت میانگین آنها در جداول ۴-۳ و ۴-۲ وارد شده است. در هر مرحله تعدادی از تصاویر هر شخص به عنوان مجموعه داده آموزشی (تعدادنمونه ها در جداول)، و بقیه تصاویر به عنوان مجموعه داده آزمایشی در نظر گرفته شده است. روش NMF برای مقادیر مختلف $r = ۴۰, ۱۰۰, ۱۵۰$ آزمایش شده است. همانطور که مشاهده می کنید روش NMF دقت بالاتری نسبت به روش SVM برای طبقه بندی داده های ORL دارد و همچنین این کار را در مدت زمان بسیار کمتری انجام می دهد.

جدول ۴-۲: روش SVM روی پایگاه داده ORL

تعداد نمونه ها	۱	۲	۳	۴	۵	۶	۷
میزان صحت	۴۶.۶۶	۵۸.۴۴	۶۶.۴۳	۷۷.۵۰	۸۴	۸۶.۲۵	۸۸.۳۳
زمان اجرا	۴۷.۹۹	۴۳.۰۵	۶۱.۲۲۷	۷۳.۷۵۴	۸۶.۸۵۰	۱۱۳.۵۵۹	۱۳۷.۳۴۴

جدول ۴-۳: روش NMF اولیه روی پایگاه داده ORL

تعداد نمونه ها	۱	۲	۳	۴	۵	۶	۷
میزان صحت-۴۰	۵۶.۸۶	۶۸.۴۶	۷۱.۸۸	۷۹.۴۱	۸۲.۲۱	۸۳.۵۷	۶۸.۵۳
زمان اجرا	۲.۹۴	۳.۴۷	۴.۱۷	۴.۸۳	۵.۷۳	۶.۲۶	۷.۲۲
میزان صحت-۱۰۰	۵۹.۹۶	۷۱.۷۴	۷۳.۷۳	۸۲.۳۸	۸۵.۷	۸۵.۴۲	۸۸.۹۲
زمان اجرا	۶.۵۵	۷.۸۴	۹.۱۰	۹.۹۹	۱۱.۰۲	۱۲.۶۹	۱۴.۰۸
میزان صحت-۱۵۰	۶۱.۱۱	۷۳.۷۱	۷۴.۸۵	۸۱.۷۴	۸۵.۸۶	۸۶.۸۸	۹۰
زمان اجرا	۱۱.۵۸	۱۳.۵۱	۱۵.۱	۱۶.۴۴	۱۷.۹۵	۱۹.۴۲	۲۰.۹۴

۴-۵-۲ مقایسه روش SVM و روش NMF روی پایگاه داده iris

در روش SVM روی پایگاه داده iris از روش نمونه برداری تصادفی با جایگذاری استفاده شده است. در روش NMF اولیه روی پایگاه داده iris از روش نمونه برداری متوازن و با جایگذاری استفاده شده است در این آزمایش به طور تصادفی، تعداد مساوی نمونه از هر دسته انتخاب شده است. همانطور که در جدول ۴-۴ مشاهده می کنید روش SVM کار طبقه بندی داده های iris را به مراتب بهتر از روش NMF انجام می دهد. با مقایسه نتایج این بخش و بخش قبل می توان نتیجه گرفت که روش NM F طبقه بندی داده های تصویر را بهتر از داده های عددی انجام می دهد.

جدول ۴-۴: روش SVM روی پایگاه داده iris

تعداد نمونه ها	۳	۶	۹	۱۲	۱۵	۱۸	۲۱
میزان صحت	۸۱.۷۳	۷۲.۲۶	۸۰.۴۰	۸۴.۶۶	۸۱.۷۳	۸۳.۲	۷۹.۶
زمان اجرا	۲.۴۲	۲.۳۶	۲.۴۲	۲.۴۱	۲.۴۱	۲.۴۱	۲.۴۹

جدول ۴-۵: روش NMF اولیه روی پایگاه داده iris

تعداد نمونه ها	۱	۲	۳	۴	۵	۶	۷
میزان صحت	۶۱.۹۰	۳۸.۱۹	۲۱.۱۹	۴۲.۰۳	۷۶.۳۰	۷۴.۲۴	۴۰.۳۱
زمان اجرا	۰.۰۰۹	۰.۰۰۹	۰.۰۰۹	۰.۰۰۹	۰.۰۰۹	۰.۰۰۹	۰.۰۰۹

۴-۵-۳ روش NMF پیشنهادی

در این بخش روش NMF پیشنهادی که در ابتدای فصل معرفی شد روی مجموعه داده ORL اعمال شده و نتایج در جدول ۴-۶ وارد شده است. با مقایسه این جدول و جدول ۴-۳ مشاهده می شود که روش NMF پیشنهادی بهتر از روش NMF اولیه عمل می کند.

جدول ۴-۶: روش NMF پیشنهادی روی پایگاه داده ORL

تعداد نمونه ها	۱	۲	۳	۴	۵	۶	۷
میزان صحت-۴۰	۵۸.۱۷	۶۸.۵۸	۷۲.۶۴	۸۰.۲۱	۸۲.۵	۸۳.۷۵	۸۷.۵
زمان اجرا	۲.۸۴	۳.۶	۴.۲	۴.۹۳	۵.۶۹	۶.۰۸	۶.۷۷
میزان صحت-۱۰۰	۶۰.۵۵	۷۲.۴۰	۷۴.۸۶	۸۲.۶۴	۸۵.۳۳	۸۵.۶۳	۹۰.۲۱
زمان اجرا	۶.۸۰	۷.۶۲	۸.۵۷	۱۰.۷۱	۱۱.۸۲	۱۳.۶۲	۱۴.۴۳
میزان صحت-۱۵۰	۶۱.۵۲	۷۳.۹۰	۷۵.۰۰	۸۳.۳۳	۸۶.۰۰	۸۷.۶۸	۹۰.۲۷
زمان اجرا	۱۱.۵۷	۱۳.۵۸	۱۵.۲۶	۱۶.۵۳	۱۸.۴۵	۲۰.۸۹	۲۱.۱۹

۴-۵-۴ روشی دیگر برای تجزیه NMF

با اعمال تغییرات جزئی در روش NMF پیشنهادی روشی دیگر را معرفی می کنیم که داده های تصویر ORL را با دقت بیشتری نسبت به روشهای پیشین طبقه بندی می کند. در این روش به جای ϵ_k معرفی شده در روش NMF پیشنهادی مقدار جدید:

$$\epsilon_k = ۱ - ۱/(k * k)$$

را قرار دادیم.

تاکنون در همه روش های NMF در این پایان نامه برای مقایسه داده های آزمایشی با داده های آموزشی از معیار فاصله فربنیوس استفاده شده است. در این آزمایش از معیار correlation [۳۴] برای این مقایسه استفاده شده است. همانطور که در جدول ۴-۷ مشاهده می کنید استفاده از این معیار فاصله نتایجی به مراتب بهتر از روشهای قبلی ارائه می دهد.

جدول ۴-۷: روش NMF جدید روی پایگاه داده ORL

تعداد نمونه ها	۱	۲	۳	۴	۵	۶	۷
میزان صحت-۴۰	۶۳.۰۸	۷۵.۲۷	۷۵.۶۱	۸۱.۴۲	۸۴.۳۵	۸۵.۷۲	۸۸.۶۹
زمان اجرا	۲.۹۸	۳.۶۵	۴.۳۲	۴.۹۲	۵.۷۰	۶.۳۲	۷.۱۱
میزان صحت-۱۰۰	۶۶.۸۲	۷۸.۲۲	۷۹.۲۴	۸۴.۷۰	۸۷.۵۷	۸۸.۹۳	۹۰.۴۶
زمان اجرا	۶.۷۴	۷.۵۴	۸.۱۹	۸.۹۶	۱۰.۸۶	۱۱.۴۱	۱۳.۳۳
میزان صحت-۱۵۰	۶۷.۲۱	۷۸.۹۳	۸۰.۶۶	۸۵.۷۷	۸۸.۵۰	۸۹.۳۷	۹۰.۸۳
زمان اجرا	۱۱.۸۸	۱۲.۱۷	۱۴.۸۸	۱۵.۴۷	۱۷.۴۶	۱۸.۴۵	۱۹.۲۰

فهرست منابع

[۱] ثابتی، کامران. تجزیه نامنفی ماتریس و کاربرد آن در خوشه بندی. پایان‌نامه کارشناسی‌ارشد، دانشکده علوم پایه، دانشگاه شاهد.، دی ۱۳۸۸.

[2] Hefferon, Jim. *An Introduction to Optimization*. Virginia Commonwealth University Mathematics, 2006.

[۳] پورصدیق، سمیه. شناسایی فریم های زمینه ویدئو با استفاده از تجزیه .qr پایان‌نامه کارشناسی‌ارشد، سبزوار: دانشکده ریاضی و علوم کامپیوتر.، بهمن ۱۳۹۴.

[4] Watkins, David S. *Fundamentals of matrix computations*. Pure and applied mathematics. Wiley-Interscience, New York, 2010.

[۵] قاسمی، مجتبی. تجزیه ماتریس ها و کاربردهای آن در تشخیص الگو و پردازش تصویر. پایان‌نامه کارشناسی‌ارشد، سبزوار: دانشکده ریاضی و علوم کامپیوتر دانشگاه حکیم سبزواری.، اردیبهشت ۱۳۹۱.

[6] Bishop, Christopher M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1 ed. , 2007.

[7] Chong, E.K.P., Zak S.H. and Żak, S.H. *An Introduction to Optimization*. Wiley Series in Discrete Mathematics and Optimization, 2013.

[۸] سونیا، کمالی،. روش های جبر خطی عددی در تشخیص چهره. پایان‌نامه کارشناسی‌ارشد، دانشکده علوم، دانشگاه شیراز.، شهریور ۱۳۹۱.

[9] Golub, G.H and Van Loan, C.F. *Matrix Computations*. Johns Hopkins University Press, Baltimore, 2nd ed., 1989.

[۱۰] .فاطمیون، فاطمه سادات. تحلیل خوشه بندی برای داده بیان ژن با استفاده از تجزیه نامنفی ماتریس. پایان‌نامه کارشناسی‌ارشد، یزد: دانشکده ریاضی، دانشگاه یزد.، مهر ۱۳۹۰.

[۱۱] رحمانی، م. و مومنی ماسوله ح.، و ثابتی ک. تجزی نامنفی ماتریس و کاربردهای آن. در دومین کنفرانس بین المللی تحقیق در عملیات ایران، صفحات ۲۶۶-۲۶۹، بابلسر: جهاد دانشگاهی صنعتی شریف.، ۱۳۸۸.

[12] lee, D. and seung, H. Algorithms for non-negative factorization. *IN NIPS*, p. 7 pages, 2000.

[13] Luka vazic. http://vazic.me/.

[۱۴] علی زاده، ح و قاسمیان، ح. جداسازی طیفی و مکانی تصاویر با استفاده از semi-nmf و تبدیل .pca پردازش علائم و داده ها، صفحات ۵۷-۷۰، آبان ۱۳۹۳.

[15] Ding, C., and He X. and D.simon, H. On the equivalence of nonnegative matrix factorization and spectral clustering. in *SIAM international conference on Data Mining, SDM*, pp. 606–610, 2005.

[16] Babaee, M., and Wolf T. and Rigoll, G. Toward semantic attributes in dictionary learning and non-negative matrix factorization. *Pattern Recognition Letters,*, pp. 172–178, July 2016.

[17] Pompili, F., and Gillis N. and Absil P.-A. and Glineur F. Two algorithms for orthogonal nonnegative matrix factorization. *Neurocomputing 141,*, pp. 15–25, February 2014.

[18] Du, K.-L. and Swamy., M. N. S. Nonnegative matrix factorization. *Neural Networks and Statistical Learning,*, pp. 407–417, 2014.

[19] P, Paatero and U, Tapper. Positive matrix factorization: A nonnegative factor model with optimal utilization of error estimates of data values. *Environmetrics 5*, pp. 111– 126, June 1994.

[20] K, Devarajan. Nonnegative matrix factorization: An analytical and interpretive tool in computational biology. *PLoS Computational Biology*, July 2008.

[21] Chan T.H, Ma W.K, Chi C.Y and Y, Wang. A convex analysis framework for blind separation of nonnegative sources. *IEEE Transactions on Signal Processing*, pp. 5120– 5134, Oct 2008.

[22] A, Lefevre. Dictionary learning methods and single-channel source separation, November 2012.

[23] Fevotte C, Bertin N and J.L, Durrieu. Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis. *Neural Computation*, pp. 793–830, March 2009.

[24] Wang F, Li T, Wang X Zhu S and C, Ding. Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery*, pp. 493–521, May 2011.

[25] Yifeng, Li., and Alioune Ngom. The non-negative matrix factorization toolbox for biological data mining. *Source Code for Biology and Medicine,*, April 2013.

[26] Charu C. Aggarwal, Chandan K. Reddy. *Data Clustering: Algorithms and Applications*. CRC Press, 2013., illustrated ed. , 2013.

[27] Lin., Chih-Jen. Projected gradient methods for nonnegative matrix factorization. *Neural Computation,*, 19:2756 – 2779, 2007.

[۲۸] یوسفی، مهسا و رزقی، منصور. تجزیهٔ نامنفی ماتریسی: روشی برای تحلیل داده های نامنفی. فرهنگ و اندیشه ریاضی،، صفحات ۷۱-۹۰، ۱۳۹۵.

[29] Benthem, Mark .H.Van and R.Keenan., Michael. Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems. *Journal of Chemometrics,*, pp. 441–450, January 2005.

[30] Duda, Richard O., Hart, Peter E., and Stork, David G. *Pattern Classification (2nd Ed)*. Wiley, 2001.

[31] Rasmus, Bro and Sijmen, De.Jong. A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics,*, 11:393–401, 1997.

[32] Yugao. Li, Wensheng. Chen, Binbin Pan Yang. Zhao and Chen., Bo. An efficient non-negative matrix factorization with its application to face recognition. in *Lecture Notes in Computer Science*, pp. 112–119, Chinese Conference on Biometric Recognition, 1388.

[33] Gillis., Nicolas. The why and how of nonnegative matrix factorization. *Machine Learning and Pattern Recognition Series*, pp. 257 – 291, Mar 2014.

[34] David.Guillamet, Jordi.Vitria. Non-negative matrix factorization for face recognition. in *Lecture Notes in Computer Science*, pp. 336–344, in Artificial Intelligence, 2002.

پیوست آ

برنامه های MATLAB روش های مطرح شده در پایان نامه

آ-۱ روش NMF

برنامه آ-۱: روش NMF روی داده های iris

```
load fisheriris
X=meas;
rng(1);                         % For reproducibility
n = size(meas,1);
m=1;% Sample size
qIdx1 = randsample(50,m); % Indices of query data
qIdx2= randsample(50,m);
qIdx2=qIdx2+50;
qIdx3 = randsample(50,m);
qIdx3=qIdx3+100;
qIdx=[qIdx1 qIdx2 qIdx3]';
% qIdx = randsample(n,m);
tIdx = ~ismember(1:n,qIdx); % Indices of training data
A1 = meas(qIdx,:);
A1=A1';
A2 = meas(tIdx,:);
```

```
A2=A2';
r=1;
Step=1;
R=50-m;
c=3;
[true,err,Rc,Norm1,Norm2] = NMF_Classification_fun(A1,A2,r,Step,R,c);
```

برنامه آ-۲: روش NMF اولیه روی داده های ORL

```
function [true,err,Rc,Norm1,Norm2] = NMF_Classification_fun(A1,A2,r,
    Step,R,c)

[m1,n1]=size(A1);
W1=abs(rand(m1,r));
H1=abs(rand(r,n1));
Maxitter=120;
eps=10^-9;
for i=1: Maxitter
    D_w=(A1*H1')./((W1*H1*H1')+eps);
    W1=max(0,(D_w.*W1));
    D_h=(W1'*A1)./((W1'*W1*H1)+eps);
    H1=max(0,(D_h.*H1));
end
[m2,n2]=size(A2);
% [W1,H1]=nnmf(A1,r);
[H2] = kfcnnls(W1, A2);
Norm1=norm(A1-W1*H1,2)/sqrt(n1*m1);
Norm2=norm(A2-W1*H2,2)/sqrt(n2*m2);

k=1;
[l,b]=size(H1);
for i=1:Step:b
    H_m(:,k)=mean(H1(:,i:i+(Step-1)),2);
    k=k+1;
```

```
end


B = pdist2(H_m',H2');
[Min,Rc] =min(B);


idxT = 1:c;
idxT = idxT(ones(R,1),:);
idxT = idxT(:);
err = sum(Rc ~= idxT')/ numel(idxT);
Rc=reshape(Rc,[R,c]);
true=1-err;
```

برنامه آ-۳: روش NMF پیشنهادی روی داده های ORL

```
function [true,err,Rc,Norm1,Norm2] = NMF_Classification_fun(A1,A2,r,
    Step,R,c)


[m1,n1]=size(A1);
W1=abs(rand(m1,r));
H1=abs(rand(r,n1));
Maxitter=120;
% eps=0.9;
for i=1: Maxitter
    eps=1-1/i;
    Q_H=(W1'*A1)./(W1'*W1*H1)+10^(-9);
    H1=max(0,(H1.*((1-eps)*Q_H+eps*(Q_H).^2)));


    Q_W=(A1*H1')./(W1*H1*H1')+10^(-9);
    W1=max(0,(W1.*((1-eps)*Q_W+eps*(Q_W).^2)));
end


[m2,n2]=size(A2);


[H2] = kfcnnls(W1, A2);
Norm1=norm(A1-W1*H1,2)/sqrt(n1*m1);
```

```
Norm2=norm(A2-W1*H2,2)/sqrt(n2*m2);
%%

k=1;
[l,b]=size(H1);
for i=1:Step:b
    H_m(:,k)=mean(H1(:,i:i+(Step-1)),2);
    k=k+1;
end

B = pdist2(H_m',H2');
[Min,Rc] =min(B);

idxT = 1:c;
idxT = idxT(ones(R,1),:);
idxT = idxT(:);
err = sum(Rc ~= idxT')/ numel(idxT);
Rc=reshape(Rc,[R,c]);
true=1-err;
```

برنامه آ-۴: روش NMF جدید روی داده های ORL

```
function [true,err,Rc,Norm1,Norm2] = NMF_Classification_fun(A1,A2,r,
    Step,R,c)

[m1,n1]=size(A1);
W1=abs(rand(m1,r));
H1=abs(rand(r,n1));
Maxitter=120;
% eps=0.9;
for i=1: Maxitter
    eps=1-1/(i*i);
    Q_H=(W1'*A1)./(W1'*W1*H1)+10^(-9);
    H1=max(0,(H1.*((1-eps)*Q_H+eps*(Q_H).^2)));
```

```
    Q_W=(A1*H1')./(W1*H1*H1')+10^(-9);
    W1=max(0,(W1.*((1-eps)*Q_W+eps*(Q_W).^2)));
end
[m2,n2]=size(A2);

[H2] = kfcnnls(W1, A2);
Norm1=norm(A1-W1*H1,2)/sqrt(n1*m1);
Norm2=norm(A2-W1*H2,2)/sqrt(n2*m2);

k=1;
[l,b]=size(H1);
for i=1:Step:b
    H_m(:,k)=mean(H1(:,i:i+(Step-1)),2);
    k=k+1;
end

B = pdist2(H_m',H2','correlation');
% B = pdist2(H_m',H2');
[Min,Rc] =min(B);

idxT = 1:c;
idxT = idxT(ones(R,1),:);
idxT = idxT(:);
err = sum(Rc ~= idxT')/ numel(idxT);
Rc=reshape(Rc,[R,c]);
true=1-err;
```

برنامه آ-۵: روش NMF روی داده های iris

```
load fisheriris
X=meas;
rng(1);                         % For reproducibility
n = size(meas,1);
m=1;% Sample size
qIdx1 = randsample(50,m); % Indices of query data
```

```
qIdx2= randsample(50,m);
qIdx2=qIdx2+50;
qIdx3 = randsample(50,m);
qIdx3=qIdx3+100;
qIdx=[qIdx1 qIdx2 qIdx3]';
% qIdx = randsample(n,m);
tIdx = ~ismember(1:n,qIdx); % Indices of training data
A1 = meas(qIdx,:);
A1=A1';
A2 = meas(tIdx,:);
A2=A2';
r=1;
Step=1;
R=50-m;
c=3;
[true,err,Rc,Norm1,Norm2] = NMF_Classification_fun(A1,A2,r,Step,R,c);
```

برنامه آ-۶: روش NMF روی داده های ORL

```
load fisheriris
X=meas;
rng(1);                        % For reproducibility
n = size(meas,1);
m=1;% Sample size
qIdx1 = randsample(50,m); % Indices of query data
qIdx2= randsample(50,m);
qIdx2=qIdx2+50;
qIdx3 = randsample(50,m);
qIdx3=qIdx3+100;
qIdx=[qIdx1 qIdx2 qIdx3]';
% qIdx = randsample(n,m);
tIdx = ~ismember(1:n,qIdx); % Indices of training data
A1 = meas(qIdx,:);
A1=A1';
A2 = meas(tIdx,:);
```

```
A2=A2';
r=1;
Step=1;
R=50-m;
c=3;
[true,err,Rc,Norm1,Norm2] = NMF_Classification_fun(A1,A2,r,Step,R,c);
```

آ-۲ روش SVM

برنامه آ-۷: روش SVM روی داده های lriris

```
load fisheriris
n = size(meas,1);
% tIdx=[40,90,132,127,136,122,131,121];
m=4;
tIdx =randsample(n,m);% Indices of query data
% r=2;
% m=3*r;
% tIdx1 = randsample(50,r); % Indices of query data
% tIdx2= randsample(50,r);
% tIdx2=tIdx2+50;
% tIdx3 = randsample(50,r);
% tIdx3=tIdx3+100;
% tIdx=[tIdx1 tIdx2 tIdx3]';
qIdx = ~ismember(1:n,tIdx); % Indices of training data
Xnew = meas(qIdx,:);
xdata = meas(tIdx,:);

group(1:50)=1;
group(50:100)=2;
group(101:150)=3;
```

```
Mdl = fitcecoc(xdata,group(tIdx));
Ynew = predict(Mdl,Xnew);

yq=group(qIdx);

count=0;
for i=1:length(qIdx)-m
  if Ynew(i)==yq(i)'
   count=count+1;
  end
end
result=count/length(qIdx)*100;
```

برنامه آ-۸: روش SVM روی داده های ORL

```
load('olivettiFaces.mat');
X=faces'; clear faces;
Y = repmat((1:40),10,1); Y = Y(:);
[s d] = size(X);

h = 64; w = 64;
name = 'faces';
m = 10;
n = 40;
A = reshape(1:m*n,m,n);
numTrain = 1;
tIdx = A(1:numTrain,:);
tIdx = tIdx(:)';
xdata = X(tIdx,:);
qIdx = setdiff(1:s,tIdx);

Xnew = X(qIdx,:);
for i=1:40
group(i*10-9:i*10)=i;
```

```
end
Mdl = fitcecoc(xdata,group(tIdx));
Ynew = predict(Mdl,Xnew);
yq=group(qIdx);
count=0;
for i=1:length(qIdx)
  if Ynew(i)==yq(i)'
   count=count+1;
  end
end
result=count/length(qIdx)*100;
```

واژه‌نامه فارسی به انگلیسی

استخراج ویژگی . . . . . Feature extraction
بردار ویژه . . . . . Eigenvector
بعد . . . . . Dimension
پردازش تصویر . . . . . Image Processing
تجزیه غیر منفی ماتریس . . . . . Non-negative matrix factorization
تقریب مرتبه پایین . . . . . Low order approximation
خوشه بندی . . . . . Clustering
داده کاوی . . . . . Data mining
رده بندی . . . . . Classification
زیرفضای برداری . . . . . Vector subSpace
شبه معکوس . . . . . Pseudo-reverse
ضرایب . . . . . Coefficients
طول گام . . . . . Step length
کاهش ابعاد . . . . . Dimension reduction
کمترین مربعات . . . . . Least squares
متعامد . . . . . Orthogonal
محدودیت . . . . . Limitation
مقدار ویژه . . . . . Eigenvalue
نرم . . . . . Norm
همبستگی . . . . . Correlation

واژه‌نامه انگلیسی به فارسی

رده بندی . . . . . Classification

خوشه بندی . . . . . Clustering

ضرایب . . . . . Coefficients

همبستگی . . . . . Correlation

داده کاوی . . . . . Data mining

بعد . . . . . Dimension

کاهش ابعاد . . . . . Dimension reduction

مقدار ویژه . . . . . Eigenvalue

بردار ویژه . . . . . Eigenvector

استخراج ویژگی . . . . . Feature extraction

پردازش تصویر . . . . . Image Processing

کمترین مربعات . . . . . Least squares

محدودیت . . . . . Limitation

تقریب مرتبه پایین . . . . . Low order approximation

تجزیه غیر منفی ماتریس . . . . . Non-negative matrix factorization

نرم . . . . . Norm

متعامد . . . . . Orthogonal

شبه معکوس . . . . . Pseudo-reverse

طول گام . . . . . Step length

زیرفضای برداری . . . . . Vector subSpace

Hakim Sabzevari University

An Outline of MSc. Thesis

Surname:Sanchooli	Name:Elham	Student No.:9413137051

Supervisor: Dr. Mahmood AminToosi

Advisor: Dr. Amin Rafiei

Faculty of Mathematics and Computer Science

Program: Decision Science and Knowledge Engineering

Title of thesis: Non-negative matrixFactorization for Data Clustering

Keywords: reduce dimension, Nonnegative Matrix Factorization, extraction of the feature, Clustering, NMF

Abstract: The Nonnegative Matrix Factorization (NMF) is one of the most up-to-date topics in the field of linear algebra, which is used to identify patterns, analyze data, and reduction dimension. In this domain, the purpose of decomposing a matrix A including nonnegative data is obtained by multiplying the base matrix U and the matrix of coefficients V^T with non-negative elements. In general, the need for the analysis and extraction of the feature has led us to reduce dimension on and compression. This action is desirable to reduce the cost of storing information.

In this dissertation, various approaches to NMF have been investigated. Using it in a vector space model maintains many of the features of the original data and ensures that both the matrix of the base and the matrix of the coefficients remain non-negative. Particularly in the case of people's images, the NMF collective model is used to represent the main features of the face, such as eyes, cheeks and lips.

Hakim Sabzevari University
Faculty of Mathematics and Computer Science

A Thesis Submitted in Partial Fulfilment of the Requirement for the Degree of Master of Science in Decision Science and Knowledge Engineering

Non-negative matrixFactorization for Data Clustering

Supervisor:
Dr. Mahmood AminToosi

Advisor:
Dr. Amin Rafiei

By:
Elham Sanchooli

August 2017