

# بهبود برازش رگرسیون با استفاده از خوشه بندی شبکه‌ی عصبی به منظور پیش بینی اوج مصرف انرژی الکتریکی ماهانه

علیرضا سروش (دانشجوی دکتری)

رضا بوادران کاظم زاده (دانشیار)

اردشیر بحرینی نژاد (استادیار)

دانشکده‌ی فنی و مهندسی، بخش مهندسی صنایع، دانشگاه تربیت مدرس

در سیستم قدرت الکتریکی، پیش بینی اوج بار به منظور بهینه سازی اقتصادی، برنامه ریزی خط مشی، تخصیص بودجه و مدیریت سوخت نیروگاه‌ها نقش بسیار مهمی بازی می‌کند. دقت روش پیش بینی در توسعه‌ی استراتژی تأمین قدرت، و طرح توسعه به‌ویژه برای کشورهای درحال توسعه بسیار مهم است. زیرا در این کشورها نرخ رشد تقاضا پویا و زیاد است. در این نوشتار برای اولین بار یک مدل رگرسیون تلفیقی به منظور پیش بینی اوج بار الکتریکی ماهانه با استفاده از روش خوشه بندی و آنالیز مؤلفه‌های اصلی<sup>۱</sup> (PCA) ارائه شده است. بدین منظور داده‌های تقاضای اوج بار ماهانه‌ی ۱۴ سال گذشته در شبکه‌ی قدرت الکتریکی سراسری ایران مورد تجزیه و تحلیل قرار می‌گیرد. مدل ارائه شده امکان پیش بینی انرژی الکتریکی مصرفی یا اوج بار ماهانه را برای ماه آینده فراهم می‌کند. بدین ترتیب که ابتدا برای بهبود برازش رگرسیون، نقشه‌ی خودسازمان دهی<sup>۲</sup> (SOM) به منظور خوشه بندی اوج بار مصرفی ماهانه طراحی شده است و ماه‌های مشابه با یکدیگر در گروه‌های یکسان قرار داده شده‌اند. سپس با به کارگیری شاخص دیویس - بولدین بهترین حالت خوشه بندی تعیین شده است. همچنین به منظور کاهش ابعاد ورودی‌ها و بهبود نتایج از آنالیز مؤلفه‌های اصلی استفاده شده است. به منظور پیش بینی اوج بار ماهانه برای هر خوشه نیز از روش رگرسیون استفاده شده است. آنالیز و مقایسه‌ی نتایج پیش بینی با روش رگرسیون بدون انجام خوشه بندی نشان می‌دهد که دقت عمل مدل ارائه شده بسیار خوب است و خوشه بندی داده‌ها موجب بهبود پیش بینی شده است.

a.soroush@modares.ac.ir  
rkazem@modares.ac.ir  
bahreininejad@modares.ac.ir

واژگان کلیدی: پیش بینی، اوج بار الکتریکی ماهانه، خوشه بندی، نقشه‌ی خودسازمان ده، رگرسیون، شاخص دیویس - بولدین.

## ۱. مقدمه

تأمین انرژی الکتریکی مورد نیاز مصرف کنندگان به اقتصادی ترین وجه ممکن در شرایط شبکه، و فراتر از آن مبنای برنامه‌های توسعه‌ی ظرفیت‌های تأسیساتی شبکه‌ی برق رسانی کشور است و نهایتاً در کل برنامه‌ی تأمین نیاز جامعه به حامل‌های مختلف انرژی نقش اساسی دارد.

به طور کلی بار الکتریکی در طی روز، هفته، ماه و حتی سال تغییر می‌کند اما آنچه که در یک جامعه‌ی معین از نظر مصرف بار الکتریکی ثابت می‌ماند، روند تغییرات بار است. هنر پیش بینی بار آن است که با مطالعه و تحلیل روند گذشته‌ی بار و کسب شناخت کافی از عوامل مؤثر و میزان اثرگذاری هر یک از این عوامل، از میان روش‌ها و مدل‌های مختلف پیش بینی بار، مناسب ترین و نزدیک ترین مدلی را که با واقعیت موجود در شبکه مطابقت دارد انتخاب، و بدین سان بار دوره‌های زمانی

در سالیان گذشته پیش بینی بار الکتریکی به دلیل نقش مهم آن در عملیات مؤثر و اقتصادی استفاده از برق مورد توجه دانشگاهیان و محققان صنعتی قرار گرفته است. برنامه‌ی تولید نیروگاه‌ها در دوره‌های مختلف زمانی براساس تغییرات بار طی ساعت، روز، هفته، ماه و سال تنظیم می‌شود. انجام این برنامه ریزی نیاز به آگاهی یا پیش بینی بار برای این دوره‌های زمانی دارد. از طرف دیگر، برنامه‌های توسعه‌ی نیروگاه‌ها و ظرفیت‌های تولیدی نیز به همین صورت و با پیش بینی اوج بار در دوره‌های برنامه ریزی تدوین می‌شود. بنابراین، پیش بینی بار مبنای بهره برداری از ظرفیت نیروگاه‌ها و شبکه‌ی برق رسانی و تخصیص بهینه‌ی ظرفیت‌ها به منظور

تاریخ: دریافت ۱۳۸۶/۶/۱۱، داوری ۱۳۸۶/۹/۱۹، پذیرش ۱۳۸۶/۱۲/۸.

مختلف مورد نیاز شبکه را با تخمینی قابل قبول پیش‌بینی کند. به دلیل وجود رفتار تصادفی بار، همواره درصدی خطا در پیش‌بینی بار وجود دارد اما این خطا نباید از حد پذیرفته شده و قابل تحمل بیشتر باشد. دقت نسبی در پیش‌بینی بار در صنعت برق از اهمیت خاصی برخوردار است. به خصوص هنگامی که پیش‌بینی بار مبنای برنامه‌ریزی توسعه‌ی شبکه و ظرفیت نیروگاهی باشد زیرا هرگونه پیش‌بینی مازاد مصرف، موجب سرمایه‌گذاری اضافی و عاطل ماندن ظرفیت تأسیساتی می‌شود؛ به عکس هرگونه پیش‌بینی کم‌تر از نیاز واقعی، شبکه را با کمبود تولید و صدمه به تجهیزات بر اثر بار اضافی مواجه خواهد کرد.

هدف از پیش‌بینی اوج بار ماهانه، پیش‌بینی تقاضای الکتریکی آینده مبتنی بر داده‌های اوج بار مصرفی و شرایط آب و هوایی است. ایجاد یک متدولوژی میان‌مدت دقیق، سریع و قوی برای پیش‌بینی بار ماهانه موجب برآورد دقیق‌تر میزان مصرف الکتریسیته خواهد شد. پیش‌بینی اوج بار الکتریکی ماه بعد، کاربردی اساسی در عملیات و برنامه‌ریزی سیستم قدرت الکتریکی دارد.

بنابراین، روش‌های پیش‌بینی مختلفی در این حوزه پیشنهاد شده‌اند. برخی از محققین در مقالات از مدل‌های سری زمانی و رگرسیون استفاده کرده‌اند. با استفاده از این روش‌ها رابطه‌ی بین متغیرهای مستقل و متغیر وابسته (پیش‌بینی) قابل درک و ساده است. همچنین، در دو دهه‌ی گذشته تحقیقات گسترده‌ی در خصوص کاربرد روش‌های هوش مصنوعی<sup>۳</sup> (AI) برای مسئله‌ی پیش‌بینی بار انجام شده است. در میان این روش‌ها، شبکه‌های عصبی مصنوعی<sup>۴</sup> (ANN) بیشترین توجه را به سوی خود جلب کرده‌اند. اولین گزارش‌ها در مورد کاربردشان در مسئله‌ی پیش‌بینی بار در اواخر دهه‌ی ۱۹۸۰ و اوایل دهه‌ی ۱۹۹۰ منتشر شده بود.<sup>[۱]</sup>

در تمامی مقالاتی که درباره‌ی پیش‌بینی اوج بار ماهانه به چاپ رسیده است، نویسندگان کلیه ماه‌های سال را در یک دسته قرار داده و پیش‌بینی اوج بار را برای یک یا چند ماه بعد ارائه کرده‌اند، به طوری که داده‌های اوج بار ماه‌های مشابه و غیرمشابه در یک دسته فرض شده‌اند که این موضوع موجب افزایش خطای پیش‌بینی خواهد شد. در این نوشتار پس از مرور کارهای انجام شده در این زمینه، مدل رگرسیون تلفیقی ارائه می‌شود که ابتدا از یک شبکه‌ی نقشه‌ی خودسازمان‌ده (SOM) برای خوشه‌بندی ماه‌های سال استفاده می‌کند و پس از کاهش ابعاد ورودی‌ها با استفاده از PCA، پیش‌بینی اوج بار ماه بعد به روش رگرسیون انجام می‌شود. همچنین، این پیش‌بینی با استفاده از روش رگرسیون بدون انجام خوشه‌بندی نیز صورت می‌گیرد و در انتها با مقایسه‌ی نتایج به دست آمده نتیجه‌گیری خواهیم کرد.

## ۲. مروری بر ادبیات موضوع

چنان که پیش‌تر اشاره شد، در خصوص پیش‌بینی اوج بار ماهانه مقالات چندانی به چاپ نرسیده است اما در این بخش برخی از مقالات چاپ شده در این خصوص را مورد بررسی قرار می‌دهیم. در سال ۲۰۰۵ محققان به منظور پیش‌بینی مصرف الکتریسیته یا بار ماهانه‌ی میان‌مدت از شبکه‌ی عصبی استفاده کردند؛ آنان برای انجام این تحقیق بارهای ماهانه‌ی شش ماه قبل را به عنوان متغیرهای ورودی برگزیدند و نتایج پیش‌بینی را با روش میانگین متحرک مقایسه کرده‌اند.<sup>[۲]</sup> در سال ۲۰۰۶ نیز تعداد دیگری از محققین، با استفاده از روش میانگین متحرک یک پارچه‌ی اتورگرسیون فصلی (SARIMA) نسبت به پیش‌بینی تقاضای اوج بار ماهانه در منطقه‌ی سلیمانیه عراق اقدام کرده‌اند.<sup>[۳]</sup> این عده در پایان تحقیق مورد نظر به این نتیجه رسیدند که مدل‌های میانگین متحرک یک پارچه‌ی اتورگرسیون (ARIMA) برای

پیش‌بینی‌های کوتاه‌مدت مناسب‌ترند، زیرا این مدل‌ها بر داده‌های گذشته‌ی که اخیراً به وقوع پیوسته‌اند تأکید می‌کنند.

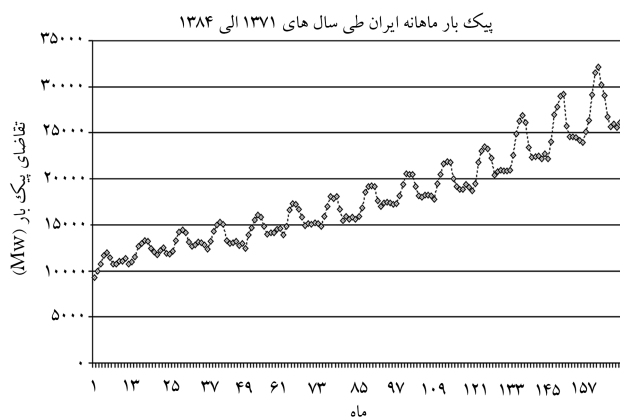
پیش‌تر، در سال ۲۰۰۴، مدل‌هایی برای پیش‌بینی اوج بار ماهانه‌ی هر سال و پیش‌بینی انرژی، با استفاده از شبکه‌های عصبی پیش‌خوراند و با شیوه‌ی یادگیری پس‌انتشار ارائه شده بود.<sup>[۴]</sup> ارائه‌ی این مدل با استفاده از دو روش پیش‌بینی مجزا برای پیش‌بینی اوج بار ماهانه، شامل مدل‌های تک‌تک<sup>۵</sup> و فزاینده<sup>۶</sup> صورت گرفت. قابلیت این مدل‌ها بر روی شبکه‌ی قدرت الکتریکی مصر مورد آزمایش قرار گرفته است.<sup>[۶]</sup>

پس از مطالعه و مرور ادبیات مربوط به پیش‌بینی اوج بار ماهانه متوجه می‌شویم که تمامی محققین داده‌های اوج بار ماهانه را در قالب یک مجموعه در نظر گرفته و با یکی از ابزارها به پیش‌بینی آن پرداخته‌اند؛ درحالی که لحاظ داده‌های مشابه و غیرمشابه در یک مجموعه موجب افزایش خطای پیش‌بینی خواهد شد. از این رو، در ادامه مدلی ارائه خواهیم کرد که با خوشه‌بندی داده‌های اوج بار ماهانه دقت پیش‌بینی در آن به میزان قابل توجهی افزایش خواهد یافت.

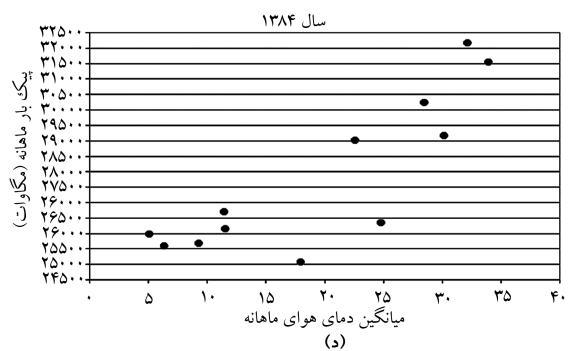
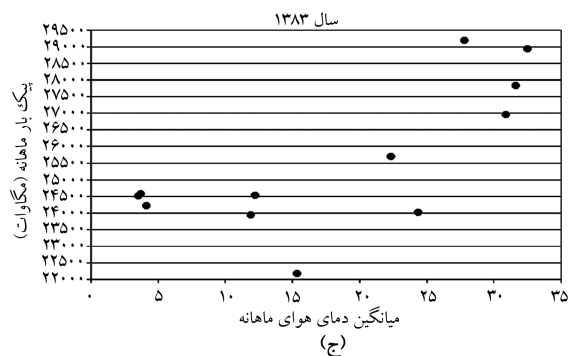
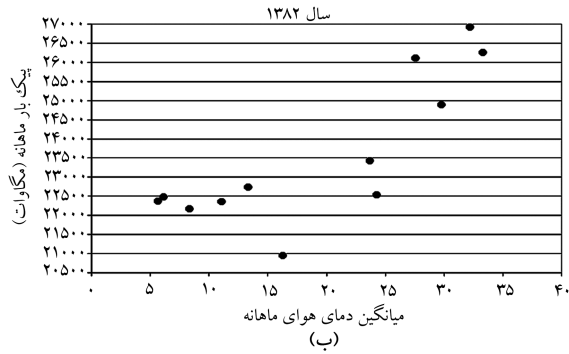
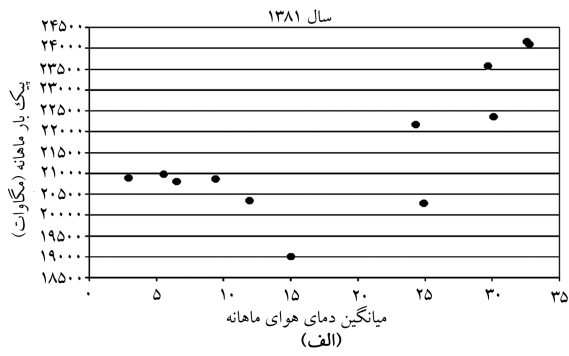
## ۳. تحلیل داده‌ها

در این تحقیق از داده‌های مصرف اوج بار ماهانه در شبکه سراسری ایران در سال‌های ۱۳۷۱ الی ۱۳۸۴، که جمعاً معادل ۱۶۸ ماه است، برای پیش‌بینی اوج بار ماهانه استفاده می‌شود. لازم به ذکر است که در بیشتر سال‌ها، اوج بار سالانه در تیر یا مردادماه رخ داده است. همچنین، براساس همین بازه زمانی، اطلاعات مربوط به متوسط دمای هوای ماهیانه‌ی ۳ شهر تهران (با هوای معتدل)، تبریز (با هوای سرد) و اهواز (با هوای گرم) از سازمان هواشناسی تهیه شده است. دمای هوا یکی از عوامل تأثیرگذاری است که می‌تواند به پیش‌بینی دقیق‌تر اوج بار ماهانه کمک کند. شکل ۱ روند سری زمانی تقاضای اوج بار ماهانه‌ی شبکه‌ی سراسری ایران طی ۱۴ سال متوالی را نمایش می‌دهد.

چنان که مشاهده می‌شود مصرف بار در هر سال، نسبت به همان مقطع زمانی در سال قبل، روندی روبه افزایش نشان می‌دهد. در فروردین ماه اوج بار در کم‌ترین میزان خود قرار دارد، اما با گرم‌تر شدن هوا مصرف برق افزایش یافته، و با کاهش میزان گرما اوج بار کاهش می‌یابد؛ مجدداً با سرد شدن هوا اوج بار (نه به اندازه‌ی ماه‌های گرم سال) افزایش می‌یابد. این شیوه‌ی تحلیل صرفاً نوعی بررسی دیداری است و ممکن است در آن خطا وارد شده باشد، از این رو، با توجه به این‌که تغییرات



شکل ۱. سری زمانی داده‌های تقاضای اوج بار ماهانه در شبکه‌ی سراسری ایران.



شکل ۳. رابطه‌ی میان اوج بار و میانگین وزنی متوسط دمای هوای ماهانه طی سال‌های ۱۳۸۱ تا ۱۳۸۴.

#### ۴. خوشه‌بندی با استفاده از شبکه‌ی عصبی نقشه

##### خودسازمان ده

هیچ‌کدام از تحقیقات انجام‌شده در زمینه‌ی پیش‌بینی اوج بار ماهانه برای انجام پیش‌بینی، داده‌های اوج بار مصرفی را دسته‌بندی نکرده‌اند. [۲-۶] درحالی‌که استفاده

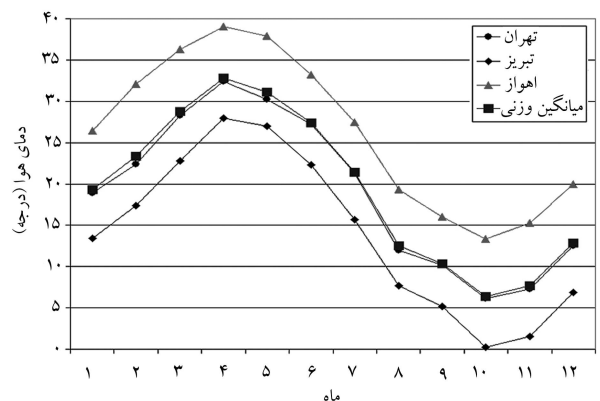
فصلی یکی از دلایل بسیار مهم برای انجام خوشه‌بندی است، لزوم استفاده از ابزاری برای شناسایی و طبقه‌بندی دقیق‌ترین الگوهای مصرف در دسته‌های یکسان بیشتر نمایان می‌شود، زیرا طبقه‌بندی داده‌ها موجب کاهش آشفتگی و افزایش در دقت پیش‌بینی اوج بار ماهانه شود.

#### ۱.۳. تأثیر دمای هوا بر الگوی مصرف بار

یکی از مهم‌ترین عوامل مؤثر بر الگوی مصرف میزان درجه حرارت است. بارهای خانگی و تجاری در ماه‌های گرم سال به دلیل استفاده از وسایل سردکننده - مانند کولر، فن و ... - و در ماه‌های سرد سال به دلیل استفاده از وسایل گرم‌کننده الکتریکی تحت تأثیر درجه حرارت‌اند.

به همین منظور، دمای سه شهر تهران، اهواز و تبریز به عنوان شاخصی برای سه نوع منطقه‌ی آب و هوایی (معتدل، گرم و سرد) از میان مناطق آب و هوایی مختلف کشور در نظر گرفته شده است و با توجه به ثقل بار مصرفی هر یک از مناطق مزبور ضرایب ۰/۶، ۰/۲، ۰/۲ در ماه‌های اردیبهشت، خرداد، تیر، مرداد، شهریور، مهر و ضرایب ۰/۶۵، ۰/۱۵، ۰/۲ در ماه‌های آبان، آذر، دی، بهمن، اسفند و فروردین به ترتیب به دمای تهران، اهواز و تبریز اختصاص داده شده است. این ضرایب با تقسیم جداگانه‌ی مجموع بار مصرفی استان‌هایی که دارای آب و هوای معتدل، گرم و سرد هستند بر مجموع بار مصرفی کلیه استان‌ها در هر یک از ماه‌های سال حاصل شده است. شکل ۲ متوسط دمای هوای ماهانه سه شهر تهران، اهواز و تبریز و نیز میانگین وزنی آنها را در سال ۱۳۸۴ براساس وزن‌های مذکور نمایش می‌دهد.

چنان که در شکل ۱ نشان داده شد، از آنجا که در برخی از ماه‌های سال با مصرف اوج بار تقریباً یکسان در دمای هوای متفاوت مواجه هستیم، و صرفاً استفاده از مصرف اوج بار برای دسته‌بندی ماه‌های سال نمی‌تواند آنها را از یکدیگر تفکیک کند و احتمال ایجاد آشفتگی در زمان پیش‌بینی افزایش می‌یابد. به همین دلیل از متغیر دما برای تفکیک اوج بار این ماه‌ها استفاده می‌شود. باید به این نکته اشاره کرد که رابطه‌ی میان مصرف اوج بار ماهانه و دمای هوا در هر سال تقریباً دارای شکل مشابهی است. به عنوان مثال، شکل ۳ رابطه‌ی میان مصرف اوج بار ماهانه و میانگین وزنی متوسط دمای هوا را طی سال‌های ۱۳۸۱ تا ۱۳۸۴ نشان می‌دهد.



شکل ۲. متوسط دمای هوای ماهانه‌ی شهرهای تهران، اهواز، تبریز و میانگین وزنی آنها در سال ۱۳۸۴.

از دسته‌بندی مناسب می‌تواند به کاهش خطای پیش‌بینی کمک کند. از این رو، در این تحقیق از نقشه خودسازمان‌ده (SOM) به عنوان یکی از ابزارهای بسیار مناسب برای خوشه‌بندی استفاده شده است.

نقشه خودسازمان‌ده در سال‌های اخیر کاربردهای فراوانی در زمینه‌های علوم مهندسی، پزشکی، زیست‌شناسی و اقتصاد در میان محققان پیدا کرده است. طی سال‌های ۱۹۹۸ تا ۲۰۰۲ تعداد ۲۰۹۲ مقاله علمی به چاپ رسیده که به نوعی از این الگوریتم در تدوین آنها استفاده شده است.<sup>[۷]</sup>

خوشه‌بندی  $Q$ ، به معنای افزایش مجموعه‌یی از داده‌ها به مجموعه‌یی از خوشه‌های  $Q_i, i = 1, \dots, C$  است. یک تعریف قابل قبول از خوشه‌بندی بهینه، افزایش است که فواصل بین نمونه‌های داخلی را کمینه، و فواصل بین خوشه‌ها را بیشینه کند.<sup>[۸]</sup> در شکل ۴ مفهوم پایه‌یی نقشه خودسازمان‌ده ارائه شده است. در این شکل،  $X_i$  نمایش‌گر بردار ورودی  $i$ ام و  $d_j$  نمایش‌گر مختصات مرکز نرون خروجی  $j$ ام است.  $W_{ij}$  نیز نشان‌دهنده وزن متصل کننده بردار ورودی  $i$ ام به نرون  $j$ ام است. همچنین، الگوریتم نقشه خودسازمان‌ده به ترتیبی که در شکل ۵ توصیف شده است، قابل اجرا است.

چنان‌که مشاهده می‌شود در الگوریتم نقشه خودسازمان‌ده که در شکل ۵ توصیف شده است،  $N_m(t)$  نمایش‌گر پارامتر آموزشی فاصله‌ی همسایگی است که حاکی از فاصله‌ی همه‌ی نرون‌های قرار گرفته در یک شعاع مشخص از نرون برنده  $m$  است.  $\alpha(t)$  نمایش‌گر نرخ یادگیری فاز ترتیب است که مقدار اولیه‌ی آن توسط کاربر تعیین می‌شود؛ به بیان دیگر، پارامتری آموزشی است که میزان تغییرات وزن و بایاس را در طول یادگیری کنترل می‌کند. این دو پارامتر آموزشی در دو فاز ترتیب و تنظیم تغییر می‌کنند.  $\sigma^2(t)$  نیز پارامتر آموزشی نمایش‌گر پهنای تابع همسایگی<sup>[۷]</sup> است که همچون نرخ یادگیری بصورت یکنواخت کاهش می‌یابد.

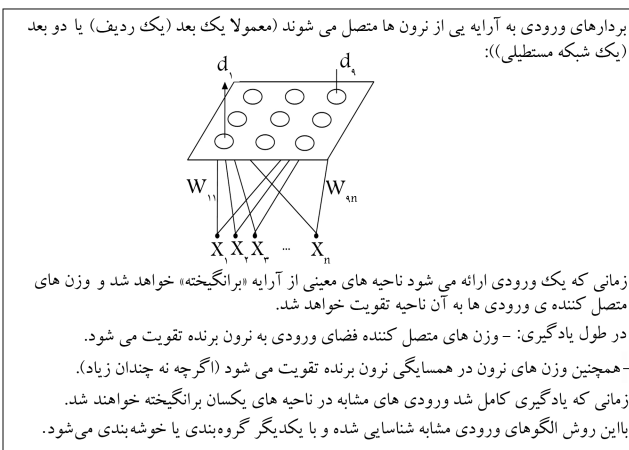
از طرف دیگر در ادبیات الگوریتم خودسازمان‌ده قاعده‌یی برای تعیین تعداد خوشه‌ها وجود ندارد و فقط در سال ۲۰۰۰ قاعده‌یی از ۲ تا  $\sqrt{N}$  به منظور تعیین دامنه‌ی تعداد خوشه‌ها پیشنهاد شد ( $N$  تعداد نمونه‌ها در مجموعه داده‌هاست).<sup>[۸]</sup> الگوریتم نقشه خودسازمان‌ده که در این تحقیق مورد استفاده قرار گرفته است تابع خطای رابطه ۱ را کمینه می‌کند:

$$E = \sum_{j=1}^C \sum_{x \in Q_j} h_m \|x - c_j\|^2 \quad (1)$$

که در آن  $C$  تعداد خوشه‌ها،  $h_m$  هسته همسایگی در واحد برنده  $m$ ، و  $c_j$  مرکز خوشه‌ی  $j$ ام است. همچنین، در این مقاله برای انتخاب بهترین خوشه‌بندی از شاخص دیویس - بولدین<sup>۸</sup> استفاده می‌شود که در آن  $S_c$  فاصله‌ی درون خوشه‌یی (مجموع فاصله بین کلیه‌ی بردارهای ورودی قرار گرفته در یک خوشه از مرکز همان خوشه) و  $d_{ce}$  فاصله‌ی بین خوشه‌یی (مجموع فاصله بین مراکز کلیه‌ی خوشه‌ها) را نمایش می‌دهد. براساس شاخص اعتبارسنجی دیویس - بولدین، بهترین خوشه‌بندی رابطه‌ی ۲ را کمینه می‌کند:

$$\frac{1}{C} \sum_{j=1}^C \max_{l \neq j} \left\{ \frac{S_c(Q_j) + S_c(Q_l)}{d_{ce}(Q_j, Q_l)} \right\} \quad (2)$$

این شاخص که فاصله‌ی درون خوشه‌یی و نیز فاصله‌ی بین خوشه‌یی را هنگام ارزیابی خوشه‌بندی حاصله مورد ارزیابی قرار می‌دهد، شاخص مناسبی برای سنجش



شکل ۴. مفهوم پایه‌یی نقشه خودسازمان‌ده.

۱. ابتدا  $w_{ij}$ ها با مقادیر تصادفی وزندهی می‌شوند. اندازه همسایگی  $N_m(t)$  باید بزرگ (اما کم‌تر از تعداد گره‌ها در یک بعد آرایه) باشد. تابع پارامتر  $\alpha(t)$  و  $\sigma^2(t)$  بین ۱ و ۰ است.
  ۲. در هر مرحله آموزش، یک الگوی ورودی  $x$  به صورت تصادفی از مجموعه لایه ورودی انتخاب می‌شود. شباهت (فاصله)  $d$  این ورودی با وزن‌های هر گره  $j$  از طریق رابطه زیر محاسبه می‌شوند.
 
$$d_j = \|x - w_j\| = \sqrt{\sum_{i=1}^n (x_i - w_{ij})^2}$$
  ۳. گره با حداقل فاصله را به عنوان  $m$  برنده انتخاب می‌شود.
  ۴. سپس وزن‌های متصل کننده‌ی لایه ورودی به نرون برنده و گره‌های همسایه‌اش بر اساس قاعده یادگیری به روز می‌شوند:
 
$$w_{ij}(t+1) = w_{ij}(t) + c[x_i - w_{ij}(t)]$$

که  $c = \alpha(t) \exp(-\|r_i - r_m\|/\sigma^2(t))$  در  $N_m(t)$  برای تمامی گره‌های  $j$  در  $N_m(t)$  که فاصله فیزیکی (تعداد گره‌ها) بین گره  $i$  و گره برنده  $m$  است.
  ۵. به همین ترتیب الگوریتم از مرحله ۲ با تعداد دوره‌های آموزشی مختلف ادامه داده می‌شود؛  $t$  تا یک افزایش می‌یابد، سپس فاصله همسایگی،  $\alpha(t)$  و  $\sigma^2(t)$  کاهش می‌یابد، به طوری‌که:
 
$$\alpha(t) = \alpha(0) N_m(t) / N_m(0)$$
- الگوریتم تکرار می‌شود تا وزن‌ها به ثبات برسند.

شکل ۵. الگوریتم نقشه خودسازمان‌ده.

تقسیم‌بندی الگوریتم خودسازمان‌ده نیز هست، زیرا هر قدر مقدار رابطه‌ی ۲ کم تر باشد، نشان‌دهنده‌ی خوب بودن نتایج خوشه‌بندی به لحاظ کروی بودن خوشه‌ها<sup>۹</sup> است. [۸]

#### ۱.۴. هنجارسازی داده‌های ورودی

از آنجاکه حداقل و حداکثر اوج بار مصرفی ماهانه در هر سال نسبت به سال قبل رو به افزایش است، خوشه‌بندی داده‌های هر سال به صورت جداگانه و با استفاده از الگوریتم خودسازمان‌ده ضرورت می‌یابد؛ زیرا اوج بار ماهانه در یک ماه از سال، نسبت به سال قبل خود بیشتر است. این مسئله موجب می‌شود که هنگام خوشه‌بندی، ماه‌های مشابه در دو سال متفاوت در خوشه‌های مختلف قرار بگیرند و به همین دلیل بردارهای ورودی سال‌های ۱۳۷۱ تا ۱۳۸۴ به صورت مجزا برای خوشه‌بندی ماه‌های سال آموزش داده می‌شوند (شامل سال‌های ۱۳۷۱ تا ۱۳۸۱ به عنوان سال‌های مورد استفاده برای برآورد ضرایب رگرسیونی، و سال‌های ۱۳۸۲ تا ۱۳۸۴ به عنوان سال‌های مورد استفاده برای پیش‌بینی)، و پس از تعیین بهترین تعداد خوشه در هر سال، وجه اشتراک بین آنها مشخص می‌شود. به این ترتیب بهترین حالت خوشه‌بندی ماه‌های سال براساس اوج بار ماهانه و میانگین وزنی متوسط دمای هوای ماهانه تثبیت می‌شود. این کار موجب افزایش اطمینان ما نسبت به ثبات خوشه‌بندی حاصله، و امکان تعمیم آن برای سال‌های آتی خواهد شد.

همچنین پس از تعیین محدوده‌ی ورودی‌های مشخص می‌شود که مقدار اوج بار ماهانه نسبت به دمای هوا عددی بزرگ‌تری است. این مسئله باعث می‌شود که در زمان خوشه‌بندی عامل دما تأثیر چندانی در تقسیم‌بندی نداشته باشد و خوشه‌بندی فقط براساس اوج بار مصرفی ماهانه انجام شود. برای جلوگیری از وقوع این مشکل باید ورودی‌ها را هنجارسازی کرد. به این منظور، از هنجارسازی استاندارد با میانگین صفر و واریانس واحد استفاده می‌شود. میانگین و انحراف معیار داده‌های ورودی با استفاده از رابطه‌ی ۳ محاسبه می‌شوند:

$$y_{new} = \frac{y_{old} - mean}{std} \quad (3)$$

که در آن  $y_{old}$  مقدار اصلی،  $y_{new}$  مقدار جدید، و  $mean$  و  $std$  به ترتیب میانگین و انحراف معیار داده‌های اصلی‌اند.

#### ۲.۴. الگوی آرایه‌ی شبکه و نرم فاصله

چنان‌که پیش‌تر نیز اشاره شد، برای تعیین دامنه‌ی تعداد خوشه‌های مورد استفاده برای دسته‌بندی داده‌ها از قاعده‌ی ۲ تا  $\sqrt{N}$  استفاده می‌شود. با توجه به این‌که تعداد نمونه‌های ما همان اوج بار مصرفی در هر یک از ماه‌های سال و برابر ۱۲ است، بنابراین تعداد خوشه‌ها باید بین ۲ تا ۳٫۴۶ باشد و به همین دلیل برای تعیین بهترین خوشه‌بندی باید بردارهای ورودی در حالت‌های ۲ و ۳ نرون و الگوهای آرایه‌ی (۱،۲) و (۱،۳) مورد آموزش قرار گیرند. بردارهای ورودی شامل دو عنصر اوج بار ماهانه و میانگین وزنی متوسط دمای هوای ماهانه است. عدد اول داخل پرانتز نمایش‌گر تعداد دسته براساس مشخصه‌ی اوج بار ماهانه و عدد دوم تعداد دسته براساس مشخصه‌ی میانگین وزنی متوسط دمای هوای ماهانه است و از ضرب این دو عدد، تعداد نرون مورد استفاده برای آموزش شبکه به دست می‌آید. همچنین برای سنجش فواصل درون خوشه‌ی بی و بین خوشه‌ی بی از معیار فاصله‌ی اقلیدسی استفاده می‌شود. بر این اساس، با فرض این‌که ماتریس داده‌ی  $y$  شامل  $(k^*n)$  شامل  $k$  ( $1^*n$ ) بردار ورودی  $y_1, y_2, \dots, y_n$  باشد، فواصل مختلف بین بردار  $y_r$  و  $y_s$  به عنوان

فاصله‌ی اقلیدسی با استفاده از رابطه‌ی ۴ محاسبه می‌شود:

$$d_{rs} = \sqrt{((y_r - y_s).^2)} \quad (4)$$

#### ۳.۴. آموزش الگوریتم خودسازمان‌ده (SOM)

الگوریتم خودسازمان‌ده (SOM) با استفاده از الگوریتم آموزشی افزایشی تصادفی<sup>۱۰</sup> برای تقسیم‌بندی مجموعه داده‌های هر سال آموزش داده شده است. در این الگوریتم در هر دوره‌ی آموزشی، هریک از بردارهای آموزشی (یا توالی‌ها) یک‌بار به صورت تصادفی به شبکه ارائه می‌شوند و مقادیر وزنی و باپاس بعد از ارائه‌ی هر بردار به روز می‌شوند. آموزش زمانی متوقف می‌شود که: تعداد دوره‌های آموزشی بیشینه شود؛ به کم‌ترین خطا دست‌یابیم؛ یا به بیشترین مقدار زمان برسیم.

لازم به ذکر است که نرخ یادگیری و فاصله‌ی همسایگی که در شکل ۵ در مورد آنها بحث شد، در قالب دو مرحله‌ی «ترتیب» و «تنظیم» تغییر داده می‌شوند. در این تحقیق، مرحله‌ی ترتیب با انجام ۱۰۰۰ گام پایان می‌پذیرد. در طول این مرحله، نرخ یادگیری از ۰٫۹ به ۰٫۲ آغاز و تا ۰٫۲۰۲ تعدیل می‌شود و فاصله‌ی همسایگی نیز از بیشینه فاصله‌ی نرونی تا ۱ تعدیل می‌شود. در طول این مرحله انتظار می‌رود که وزن‌های نرونی خودشان را در فضای ورودی سازگار با موقعیت‌های نرونی هم‌بسته مرتب کنند. در طول مرحله‌ی تنظیم نرخ یادگیری از ۰٫۲ به کم‌تری کاهش می‌یابد و فاصله‌ی همسایگی همیشه برابر ۱ است. انتظار می‌رود در طول این مرحله وزن‌ها تقریباً به صورت تصادفی در کل فضای ورودی پراکنده شوند در حالی‌که نظم توپولوژیکی‌شان در مرحله‌ی ترتیب حفظ می‌شود.

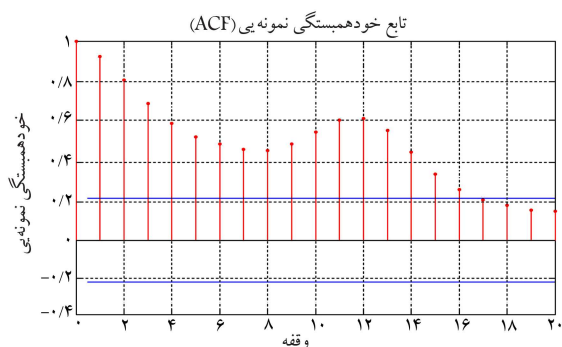
به این ترتیب، نقشه‌های ویژگی<sup>۱۱</sup> هم‌زمان با یادگیری گروه‌بندی ورودی‌های‌شان، توپولوژی و توزیع ورودی‌های‌شان را نیز یاد می‌گیرند.<sup>[۹]</sup> همچنین، هریک از شبکه‌های عصبی SOM در دو حالت ۵۰۰ و ۱۰۰۰ دوره آموزشی مورد آموزش قرار می‌گیرند.

#### ۴.۴. خروجی خوشه‌بندی

همان‌طورکه پیش از این بیان شد، شبکه‌ی عصبی SOM با دو آرایه‌ی مختلف و با استفاده از الگوریتم آموزشی افزایشی تصادفی و در دو مرحله ترتیب و تنظیم آموزش داده شد. بردارهای ورودی هریک از سال‌های ۱۳۷۱ تا ۱۳۸۴ به صورت جداگانه آموزش داده شده و نتایج خوشه‌بندی حاصل شد. سپس به منظور انتخاب بهترین خوشه‌بندی از میان دو آرایه‌ی ممکن، شاخص اعتبارسنجی دیویس - بولدین با استفاده از معیار فاصله‌ی اقلیدسی محاسبه شده است. جالب توجه است که در تمامی سال‌های مذکور خروجی خوشه‌بندی مشابه بوده، به طوری‌که در حالت دوخوشه ماه‌های اردیبهشت، خرداد، تیر، مرداد، شهریور و مهر در یک خوشه و ماه‌های آبان، آذر، دی، بهمن، اسفند و فروردین در خوشه‌ی دیگر قرار گرفته‌اند. در حالت سه‌خوشه نیز ماه‌های فروردین، آبان، آذر، دی، بهمن و اسفند در یک خوشه و ماه‌های اردیبهشت و مهر در خوشه‌ی دوم و ماه‌های خرداد، تیر، مرداد و شهریور در خوشه‌ی سوم واقع شده‌اند. جدول ۱ شاخص دیویس - بولدین به دست آمده برای هر یک از سال‌های ۱۳۷۱ تا ۱۳۸۴ را در حالت‌های ۲ و ۳ خوشه نشان می‌دهد. چنان‌که در جدول ۱ مشاهده می‌شود، کم‌ترین مقدار شاخص دیویس - بولدین با استفاده از سه نرون یا خوشه به دست می‌آید. لذا آرایه‌ی انتخاب شده (۱،۳) است و پیش‌بینی برای هر ماه در قالب یکی از این سه خوشه به صورت جداگانه انجام می‌شود. خروجی و نتیجه‌ی خوشه‌بندی نهایی ماه‌های سال به صورت زیر است: خوشه‌ی ۱: ماه‌های فروردین، آبان، آذر، دی، بهمن و اسفند؛

جدول ۱. مقادیر شاخص دیویس - بولدین برای تعداد خوشه‌های مختلف طی سال‌های ۱۳۷۱-۱۳۸۴.

سال		۱۳۷۷		۱۳۷۶		۱۳۷۵		۱۳۷۴		۱۳۷۳		۱۳۷۲		۱۳۷۱	
تعداد خوشه		۲	۳	۲	۳	۲	۳	۲	۳	۲	۳	۲	۳	۲	۳
مقدار شاخص دیویس - بولدین		۰٫۹۱	۲٫۴۳	۰٫۹۵	۲٫۴۲	۱٫۰۸	۲٫۴۷	۰٫۸۶	۲٫۵۹	۱٫۲۸	۲٫۶۸	۱٫۱	۲٫۵۴	۱٫۰۵	۲٫۵۱
سال		۱۳۸۴		۱۳۸۳		۱۳۸۲		۱۳۸۱		۱۳۸۰		۱۳۷۹		۱۳۷۸	
تعداد خوشه		۲	۳	۲	۳	۲	۳	۲	۳	۲	۳	۲	۳	۲	۳
مقدار شاخص دیویس - بولدین		۱٫۰۳	۲٫۴۱	۰٫۹۴	۲٫۲۵	۰٫۸۵	۲٫۳۹	۱٫۲۵	۲٫۴۹	۰٫۸۵	۲٫۴۱	۱٫۰۲	۲٫۵۲	۰٫۹۴	۲٫۵



شکل ۶. هم‌بستگی سری زمانی داده‌های اوج بار ماهانه طی سال‌های ۱۳۷۱ تا ۱۳۸۴.

به‌گونه‌یی است که از وقفه‌ی اول تا هشتم کاهش می‌یابد، با تغییر شرایط فصلی از وقفه‌ی نهم مجدداً افزایش می‌یابد، و در وقفه‌ی دوازدهم (همان ماه در سال گذشته) بیشینه می‌شود. براساس این نمودار، وقفه‌هایی که دارای بیش از ۰٫۶ هم‌بستگی هستند به‌عنوان تعدادی از عوامل تأثیرگذار بر مصرف اوج بار ماهانه در نظر گرفته شده‌اند. بدین ترتیب، متغیرهای ورودی برای هر یک از سه خوشه عبارت‌اند از: شماره‌ی ماه ( $M$ )، شماره‌ی سال ( $Y$ )، میانگین وزنی متوسط دمای هوای ماهانه‌ی سه شهر تهران، تبریز و اهواز در ماه جاری ( $T_0$ )؛ اوج بار ماهانه ( $L_1$ ) و میانگین وزنی متوسط دمای هوای ماهانه‌ی سه شهر تهران، تبریز و اهواز در ماه گذشته ( $T_1$ )؛ اوج بار ماهانه ( $L_2$ ) و متوسط دمای هوای ماهانه‌ی سه شهر تهران، تبریز و اهواز در دو ماه گذشته ( $T_2$ )؛ اوج بار ماهانه ( $L_3$ ) و میانگین وزنی متوسط دمای هوای ماهانه‌ی سه شهر تهران، تبریز و اهواز در ماه گذشته ( $T_3$ )؛ اوج بار ماهانه ( $L_4$ ) و متوسط دمای هوای ماهانه‌ی سه شهر تهران، تبریز و اهواز در سال گذشته ( $T_4$ ). در نتیجه معادله‌ی رگرسیونی هر یک از آنها چنین نوشته می‌شود:

$$L_0 = \beta_0 + \beta_1 M + \beta_2 Y + \beta_3 T_0 + \beta_4 L_1 + \beta_5 T_1 + \beta_6 L_2 + \beta_7 T_2 + \beta_8 L_3 + \beta_9 T_3 + \beta_{10} L_4 + \beta_{11} T_4$$

همچنین  $L_0$  نشان‌گر مقدار برآورد شده برای اوج بار ماه بعد است. برای پیش‌بینی اوج بار ماهانه در هر یک از خوشه‌ها می‌توان از معادله‌ی بالا استفاده کرد، اما بهتر است به‌جای استفاده از این معادله دو عمل پیش‌پردازشی را بر روی داده‌ها انجام دهیم تا پیش‌بینی‌ها بهبود یابند:

۱. هنجارسازی داده‌ها که موجب یکسان‌سازی مقادیر متغیرها شده و امکان برآورد دقیق‌تر پارامترها را فراهم می‌کند.

۲. کاهش ابعاد مسئله با استفاده از آنالیز مؤلفه‌های اصلی ( $PCA$ ) که از طریق

خوشه‌ی ۲: ماه‌های اردیبهشت و مهر؛

خوشه‌ی ۳: ماه‌های خرداد، تیر، مرداد و شهریور.

یادآوری مجدد این نکته ضرورت دارد که استفاده از شبکه‌ی عصبی  $SOM$  موجب کاهش خطای پیش‌بینی از طریق دسته‌بندی بهینه‌ی داده‌های ورودی نزدیک به یکدیگر در یک خوشه می‌شود. همچنین برای اطمینان از عملکرد  $SOM$ ، الگوریتم  $K - means$  نیز مورد آزمون قرار گرفت. مقایسه‌ی این دو روش با یکدیگر نشان داد که  $SOM$  به‌عنوان ورودی برای بخش پیش‌بینی جواب‌های بهتری ارائه می‌کند. دلایل این امر آن است که  $K - means$  مقادیر اولیه‌ی تصادفی را به مراکز خوشه‌ها تخصیص می‌دهد و در هر اجرا خروجی متفاوتی ارائه می‌کند، اما با به‌کارگیری  $SOM$  یک خروجی حاصل می‌شود. از طرف دیگر،  $K - means$  واریانس بین خوشه‌یی را بیشینه می‌کند اما تضمین نمی‌کند که نتیجه یک کمینه واریانس کلی داشته باشد.<sup>[۱۰]</sup> در ادامه در خصوص شیوه‌ی انجام پیش‌بینی بحث خواهیم کرد.

## ۵. پیش‌بینی با روش رگرسیون

روش رگرسیون به مطالعه‌ی وابستگی یک متغیر وابسته به یک یا چند متغیر مستقل می‌پردازد که با تخمین یا پیش‌بینی میانگین مقادیر متغیر وابسته در حالتی که مقادیر متغیرهای مستقل معلوم یا معین شده باشند، صورت می‌پذیرد. اصطلاح «خطی بودن» رگرسیون همیشه به رگرسیونی اطلاق می‌شود که از نظر پارامترها خطی باشد (یعنی پارامترها توان ۱ داشته باشند) که در این حالت ممکن است مدل از نظر متغیرهای مستقل، خطی یا غیرخطی باشد.<sup>[۱۱]</sup> در این نوشتار، مانند بخش خوشه‌بندی داده‌های ۱۱ سال (سال‌های ۱۳۷۱ تا ۱۳۸۱) برای تعیین پارامترها یا ضرایب متغیرهای مدل داده‌های سه سال (سال‌های ۱۳۸۲ تا ۱۳۸۴) به‌منظور پیش‌بینی و اعتبارسنجی مدل مورد استفاده قرار می‌گیرد. در زیربخش‌های بعدی به معرفی مشخصات مدل پیش‌بینی خواهیم پرداخت.

### ۱.۵. متغیرهای ورودی

وجود هم‌بستگی میان مقادیر متوالی داده‌ها یک ابزار کلیدی برای شناسایی الگوی پایه‌یی و تعیین مدلی مناسب با توجه به سری داده‌ها است. در مجموعه‌یی از اعداد کاملاً تصادفی، هم‌بستگی میان مقادیر متوالی نزدیک یا برابر با صفر خواهد بود، در حالی که مقادیر داده‌یی با مشخصه‌ی فصلی یا دوره‌یی به‌شدت هم‌بستگی خواهند داشت. شکل ۶ هم‌بستگی سری زمانی داده‌های اوج بار ماهانه طی سال‌های ۱۳۷۱ تا ۱۳۸۴ را نشان می‌دهد.

همان‌طور که در شکل ۶ مشاهده می‌کنید، الگوی فصلی و دوره‌یی میان مقادیر اوج بار ماهانه وجود داشته و این مقادیر به‌شدت با یکدیگر هم‌بسته‌اند. این هم‌بستگی

الف) اولین مؤلفه بیشینه‌ی واریانس داده‌ها را لحاظ کرده و دومین مؤلفه بیشینه‌ی واریانس لحاظ نشده توسط اولین مؤلفه را لحاظ می‌کند و الی آخر.

$$w_{k1} + w_{k2} + \dots + w_{kp} = 1 \quad k = 1, 2, \dots, p \quad (ب)$$

$$w_{k1}w_{z1} + w_{k2}w_{z2} + \dots + w_{kp}w_{zp} = 0 \quad i \neq j \quad (ج)$$

بدین ترتیب، پس از استفاده از  $PCA$  برای کاهش تعداد متغیرهای اصلی مدل رگرسیون هر خوشه، تعداد متغیرهای اصلی مدل رگرسیون در خوشه‌ی ۱ از ۱۱ متغیر به ۵ متغیر جدید، تعداد متغیرهای اصلی خوشه‌ی ۲ از ۱۱ متغیر به ۴ متغیر جدید، تعداد متغیرهای اصلی خوشه‌ی ۳ از ۱۱ متغیر به ۳ متغیر جدید و تعداد متغیرهای اصلی بدون استفاده از خوشه‌بندی از ۱۱ متغیر به ۴ متغیر جدید کاهش می‌یابد. شکل ۷ ماتریس مؤلفه‌های دوران داده‌شده<sup>۱۲</sup> مجموعه داده‌های هر سه خوشه به صورت جداگانه پس از انجام خوشه‌بندی و همچنین ماتریس مؤلفه‌های دوران داده‌شده کلیه داده‌ها بدون خوشه‌بندی را نشان می‌دهد.

چنان که در شکل ۷ مشاهده می‌شود، ۴ متغیر جدید برای کلیه داده‌ها حاصل می‌شود که با استفاده از آن تقریباً ۹۸٫۸ درصد کل واریانس مجموعه‌ی داده‌ها محاسبه می‌شود. به طوری که متغیرهای اصلی  $(Y, L1, L2, L3, L12)$  در شکل‌گیری مؤلفه‌ی اول، متغیرهای اصلی  $(T, T1, T12)$  در شکل‌گیری مؤلفه‌ی دوم، متغیرهای اصلی  $(T2, T3)$  در شکل‌گیری مؤلفه سوم، و متغیر اصلی  $M$  در شکل‌گیری مؤلفه چهارم تأثیر بیشتری دارند. همچنین مقدار کلی ۰٫۸۰۱ برای سنجی کیفیت نمونه‌گیری کایسر - میر - اولکین<sup>۱۳</sup>  $(KMO)$  نشان می‌دهد که ماتریس هم‌بستگی برای فاکتورگیری مناسب است.<sup>۱۲</sup>

در خصوص خوشه‌ی ۱ نیز با در نظر گرفتن ۵ متغیر جدید تقریباً ۹۸٫۹ درصد کل واریانس مجموعه داده‌ها محاسبه می‌شود. به طوری که متغیرهای اصلی  $(Y, L1, L2, L3, L12)$  در شکل‌گیری مؤلفه‌ی اول، متغیرهای اصلی  $(T, T12)$  در شکل‌گیری مؤلفه‌ی دوم، متغیرهای اصلی  $(T1, T2, T3)$  در شکل‌گیری مؤلفه‌ی سوم، و متغیر اصلی  $M$  در شکل‌گیری مؤلفه‌ی چهارم تأثیر بیشتری دارند. همچنین مقدار کلی ۰٫۷۸۹ برای سنجی  $KMO$  نشان می‌دهد که ماتریس هم‌بستگی برای فاکتورگیری نسبتاً مناسب است. به همین ترتیب در خصوص خوشه‌ی ۲ نیز با در نظر گرفتن ۴ متغیر جدید تقریباً ۹۹ درصد کل واریانس مجموعه داده‌ها و در مورد خوشه‌ی ۳ نیز با در نظر گرفتن ۳ متغیر جدید برای کلیه داده‌ها، تقریباً ۹۸ درصد کل واریانس مجموعه داده‌ها محاسبه می‌شود. مقدار کلی ۰٫۸۳۳ و ۰٫۷۹۵ نیز برای سنجی  $KMO$  به ترتیب برای خوشه‌های ۲ و ۳ نتیجه می‌دهد که ماتریس هم‌بستگی برای فاکتورگیری مناسب است.

به این ترتیب معادلات رگرسیونی کلیه داده‌ها بدون خوشه‌بندی و هر یک از سه خوشه بر اساس ماتریس مؤلفه‌های دوران داده شده چنین نوشته می‌شود:

$$L_a = \beta_0 + \sum_{p=1}^4 \beta_p C_p \quad (\text{معادله کلیه داده‌ها})$$

$$L_{c1} = \beta_0 + \sum_{p=1}^5 \beta_p C_p \quad (\text{معادله خوشه ۱})$$

$$L_{c2} = \beta_0 + \sum_{p=1}^4 \beta_p C_p \quad (\text{معادله خوشه ۲})$$

$$L_{c3} = \beta_0 + \sum_{p=1}^3 \beta_p C_p \quad (\text{معادله خوشه ۳})$$

ایجاد متغیرهای جدیدی که ترکیبی از متغیرهای اصلی هستند موجب از بین رفتن هم‌بستگی میان متغیرهای اصلی در مدل رگرسیون می‌شود. در بخش بعدی در خصوص پیش‌پردازش داده‌ها و نتایج آن بحث خواهیم کرد.

## ۲.۵. پیش‌پردازش داده‌ها

اگر مراحل پیش‌پردازش معینی بر روی ورودی‌ها و مقادیر واقعی خروجی - که همان متغیرهای ما هستند - صورت گیرد ضرایب (پارامترهای) دقیق‌تری برای این متغیرها به منظور پیش‌بینی در آینده حاصل می‌شود. از این رو، پیش از برآورد ضرایب هم‌مقیاس‌سازی ورودی‌ها و خروجی‌ها انجام می‌شود. پیش‌پردازش داده‌ها سبب می‌شود داده‌ها در دامنه‌ی معینی فشرده شوند. این فشرده‌سازی قادر است نتایج تخمین رگرسیونی را بهبود بخشد. در این تحقیق نیز به دلیل متفاوت بودن مقیاس داده‌های اوج بار ماهانه و میانگین وزنی متوسط دمای هوای ماهانه، استفاده از هنجارسازی ضروری به نظر می‌رسد. از این رو هنجارسازی داده‌های هر دسته با استفاده از روش هنجارسازی استاندارد با میانگین صفر و واریانس واحد انجام می‌شود. همچنین گاهی ابعاد بردار ورودی بزرگ است اما مؤلفه‌های بردارهای ورودی به شدت هم‌بسته‌اند. در چنین شرایطی کاهش ابعاد بردارهای ورودی بسیار مفید است. آنالیز مؤلفه‌های اصلی ( $PCA$ ) ابزاری کارا برای انجام این کار است.<sup>۱۴</sup> از این رو، بردارهای ورودی هنجار شده با میانگین صفر و واحد مورد آنالیز مؤلفه‌های اصلی قرار می‌گیرند. از آنالیز مؤلفه‌های اصلی به منظور کاهش ابعاد بردارهای ورودی و کاهش پیچیدگی مسئله در هر یک از دسته‌ها، و نیز برای حذف عواملی که تأثیرگذاری کم‌تری بر متغیر خروجی دارند، استفاده می‌کنیم.  $PCA$  مؤلفه‌های اصلی را که درصد اثرگذاری‌شان در کل تغییرات مجموعه داده‌ها کم‌تر از ۱٪ است حذف می‌کند. با استفاده از آنالیز مؤلفه‌های اصلی، با تعداد متغیرهای کم‌تر و معماری ساده‌تری سروکار داشته و نتایج پیش‌بینی بهتری حاصل خواهد شد.<sup>۱۴</sup>

$PCA$  یک روش آماری برای تعریف متغیرهای جدید برحسب ترکیب خطی از متغیرهای اولیه است. متغیرهای جدید نسبت به هم مستقل‌اند و واریانس آنها دارای روند نزولی است. هدف از  $PCA$ ، استخراج مؤلفه‌های اصلی مجموعه‌ی داده‌ها است. بر این اساس، اولین مؤلفه‌ی استخراجی دارای بیشترین واریانس مجموعه‌ی اولیه داده‌ها است و مؤلفه‌های استخراجی بعدی واریانس کم‌تری از مجموعه اولیه داده‌ها را لحاظ کرده و مجموع واریانس مؤلفه‌های استخراجی مساوی واریانس کل داده‌ها است. همچنین مؤلفه‌های استخراجی شده نسبت به هم مستقل‌اند. به طور کلی در روش آنالیز مؤلفه‌ی اصلی از  $P$  متغیر اولیه می‌توان  $m$  مؤلفه‌ی اصلی مستقل ( $m < P$ ) استخراج کرد. از طریق مؤلفه‌های استخراجی فضای  $P$  بعدی اطلاعات به فضای  $m$  بعدی قابل کاهش است. از نظر ریاضی مؤلفه‌های اصلی  $P$  متغیر اولیه چنین نمایش داده می‌شوند<sup>۱۴</sup>

$$\xi_1 = w_{11}z_1 + w_{12}z_2 + \dots + w_{1p}z_p$$

$$\xi_2 = w_{21}z_1 + w_{22}z_2 + \dots + w_{2p}z_p$$

⋮

$$\xi_p = w_{p1}z_1 + w_{p2}z_2 + \dots + w_{pp}z_p$$

در این روابط  $\xi_1, \xi_2, \dots, \xi_p$  مؤلفه‌ی اصلی، و  $w_{kz}$  ضریب  $z$  امین متغیر در  $k$  امین مؤلفه است. هدف از آنالیز مؤلفه‌های اصلی تخمین ضرایب است  $w_{kz}$  با این شرط که:

ماتریس مؤلفه‌های دوران داده شده کلیه داده‌ها					ماتریس مؤلفه‌های دوران داده‌شده خوشه ۱					
متغیرهای اصلی	مؤلفه				متغیرهای اصلی	مؤلفه				
	۱	۲	۳	۴		۱	۲	۳	۴	۵
M	۲,۷۱E-۰۲	-۰,۶۶۴	,۱۵۵	,۷۳۰	M	۳,۹۰E-۰۲	-۰,۶۹۲	-۴,۶۰E-۰۲	,۷۱۹	-۴,۹۱E-۰۴
Y	,۹۸۰	۲,۹۴۶E-۰۲	-۷,۲۲E-۰۲	-۵,۶۰E-۰۲	Y	,۹۸۲	۷,۸۲E-۰۲	-۲,۵۲E-۰۲	-۴,۹۶E-۰۲	۲,۴۷E-۰۲
T	۴,۷۰E-۰۲	,۹۸۷	۶,۳۴E-۰۲	-۰,۱۰۱	T	۲,۲۸E-۰۲	,۹۵۸	-۶,۸۹E-۰۲	-۸,۹۲E-۰۲	,۲۴۸
L۱	,۹۷۴	,۱۳۷	,۱۳۴	۵,۸۶E-۰۲	L۱	,۹۹۴	۲,۴۱E-۰۲	۷,۰۴E-۰۲	۳,۲۴E-۰۲	-۵,۷۵E-۰۳
T۱	۹,۵۱E-۰۲	,۸۱۸	,۵۵۵	-۵,۶۷E-۰۲	T۱	,۱۱۴	,۴۵۷	,۸۶۶	-۱,۳۲E-۰۲	۴,۴۸E-۰۲
L۲	,۹۷۵	۴,۹۲۷E-۰۳	,۱۹۰	۱,۰۰E-۰۲	L۲	,۹۷۴	۴,۴۱E-۰۲	,۱۹۵	۳,۹۸E-۰۲	۶,۲۳E-۰۳
T۲	,۱۲۹	,۴۲۴	,۸۹۳	۷,۱۰E-۰۳	T۲	,۱۱۴	-۹,۱۴E-۰۲	,۹۸۲	-۱,۸۹E-۰۲	-۲,۴۰E-۰۲
L۳	,۹۶۶	-۰,۱۳۶	,۱۸۲	-۳,۹۷E-۰۲	L۳	,۹۶۴	-۳,۴۳E-۰۲	,۲۴۲	-۹,۸۸E-۰۳	۸,۶۶E-۰۳
T۳	,۱۳۶	-۷,۲۹۰E-۰۲	,۹۷۹	۹,۳۷E-۰۲	T۳	,۱۱۴	-۰,۳۷۶	,۹۱۰	۳,۱۴E-۰۳	-۵,۱۷E-۰۳
L۱۲	,۹۶۵	,۲۱۱	۳,۴۵E-۰۲	۶,۲۱E-۰۲	L۱۲	,۹۹۶	-۲,۵۲E-۰۲	۵,۰۶E-۰۳	۲,۵۶E-۰۲	۱,۱۸E-۰۲
T۱۲	۶,۴۲۶E-۰۲	,۹۸۵	۶,۴۲۵E-۰۲	-۹,۸۸۸E-۰۲	T۱۲	۵,۲۹E-۰۲	,۹۶۱	-۶,۷۸E-۰۲	-۷,۹۲E-۰۲	-۰,۲۴۸

ماتریس مؤلفه‌های دوران داده‌شده خوشه ۲					ماتریس مؤلفه‌های دوران داده‌شده خوشه ۳			
متغیرهای اصلی	مؤلفه				متغیرهای اصلی	مؤلفه		
	۱	۲	۳	۴		۱	۲	۳
M	,۹۸۶	,۱۴۷	-۱,۷۴E-۰۲	۵,۹۴E-۰۳	M	۷,۲۹E-۰۲	,۹۶۱	-۰,۲۵۲
Y	-۰,۱۴۱	,۹۷۲	۸,۰۷E-۰۲	۳,۵۵E-۰۲	Y	,۹۷۸	-۳,۸۶E-۰۲	,۱۲۳
T	-۰,۸۷۴	,۱۲۵	,۱۵۹	,۴۳۸	T	۵,۵۳E-۰۲	-۹,۷۸E-۰۲	,۹۶۹
L۱	,۲۲۲	,۹۷۱	۵,۲۴E-۰۳	۲,۳۳E-۰۲	L۱	,۹۷۹	,۱۶۵	۸,۴۱E-۰۲
T۱	,۹۵۶	,۲۰۷	۲,۷۰E-۰۲	,۱۳۷	T۱	,۱۱۵	,۹۶۴	,۱۷۸
L۲	,۲۳۳	,۹۶۷	۲,۴۹E-۰۲	۱,۹۵E-۰۲	L۲	,۹۷۰	,۲۲۷	۱,۴۸E-۰۲
T۲	,۹۷۶	,۱۹۹	۳,۷۰E-۰۳	-۱,۱۵E-۰۲	T۲	,۱۱۰	,۹۸۳	-۶,۷۲E-۰۲
L۳	,۲۰۹	,۹۷۳	۱,۴۳E-۰۲	۱,۹۱E-۰۲	L۳	,۹۸۷	,۱۳۶	-۵,۵۴E-۰۲
T۳	,۹۸۰	,۱۷۴	۱,۵۹E-۰۳	-۲,۵۳E-۰۳	T۳	,۱۲۷	,۹۶۴	-۰,۱۸۲
L۱۲	-۲,۶۷E-۰۲	,۹۹۵	۴,۷۳E-۰۲	۷,۲۱E-۰۳	L۱۲	,۹۸۸	۳,۶۸E-۰۲	,۱۲۳
T۱۲	-۰,۸۱۳	,۲۵۱	,۵۱۶	۹,۵۱E-۰۲	T۱۲	,۱۱۷	-۰,۱۰۱	,۹۶۵

شکل ۷. ماتریس‌های مؤلفه‌های دوران داده شده در دو حالت پس از خوشه‌بندی و بدون خوشه‌بندی.

پیش‌بینی اوج بار ماه بعد پس از انجام خوشه‌بندی، و بدون استفاده از خوشه‌بندی مورد مقایسه قرار خواهد گرفت.

### ۳.۵. معیارهای عملکرد

در بیشتر نوشتارها، تنها میانگین درصد قدرمطلق خطا<sup>۱۵</sup> ( $MAPE$ ) به‌عنوان شاخصی برای سنجش میزان خطای پیش‌بینی ارائه شده است اما  $MAPE$  یک استاندارد در صنعت تأمین الکتریسیته است که به‌وضوح کفایت نمی‌کند. در زمینه‌ی انتخاب معیارهای خطا برای کمک به مقایسه‌ی روش‌های پیش‌بینی بحث‌های بسیاری انجام شده است. بیشتر محققین با این موضوع موافق‌اند که باید از تابع جریمه‌ی هم‌بسته با خطاهای پیش‌بینی در ارزیابی یک روش استفاده شود. اگر تابع خسارت<sup>۱۶</sup>

که در آن  $C_p$  نمایش‌گر متغیر جدید یا مؤلفه‌ی  $p$ ام و  $\beta_p$  نمایش‌گر ضریب متغیر  $p$ ام و  $L$  مقدار برآورد شده برای اوج بار ماه بعد است.

پیش از انجام پیش‌بینی با استفاده از معادلات رگرسیونی حاصله، بررسی مناسب بودن مدل‌های رگرسیونی با استفاده از تحلیل باقی‌مانده‌ها<sup>۱۴</sup> ضروری به نظر می‌رسد. شکل ۸ تحلیل باقی‌مانده‌ها برای کلیه‌ی داده‌ها و برای هر یک از سه خوشه را نمایش می‌دهد.

همان‌طور که در شکل ۸ مشاهده می‌کنید، پراکندگی باقی‌مانده‌ها تصادفی بوده و میزان انحراف از مدل رگرسیونی قابل قبول است. همچنین می‌توان متوجه شد که باقی‌مانده دارای توزیع نرمال است و جملات خطا واریانس ثابت دارند. بنابراین، از معادلات ذکرشده برای برآورد ضرایب مدل‌های رگرسیونی به‌ترتیبی که در بخش نتایج پیش‌بینی توضیح داده خواهد شد، استفاده می‌شود و نتایج مدل در دو حالت



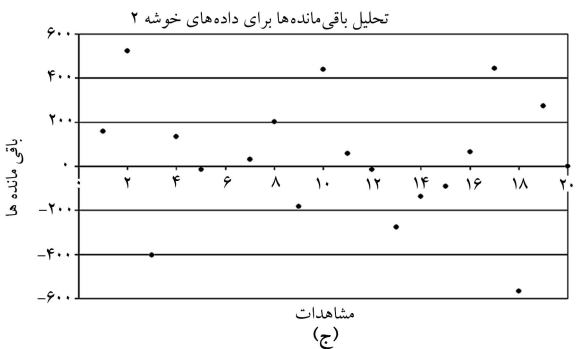
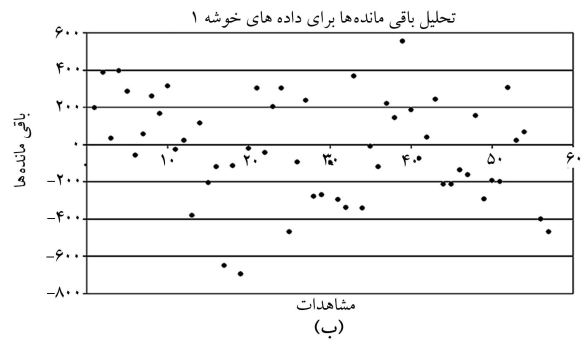
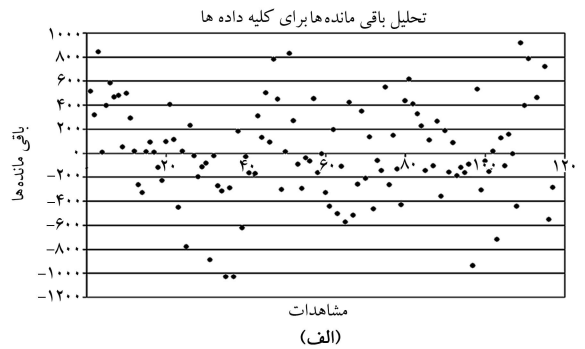
در این تحقیق برای سنجش بهترین پیش‌بینی اوج بار ماهانه از معیارهای میانگین قدرمطلق خطا<sup>۱۷</sup> ( $MAE$ )، میانگین درصد قدرمطلق خطا ( $MAPE$ )، بیشینه‌ی قدرمطلق خطا ( $MAXAE$ ) و بیشینه‌ی درصد قدرمطلق خطا ( $MAXAPE$ ) برای برآورد خطا به صورت عادی استفاده می‌شود. همچنین از معیار میانگین مربعات خطا<sup>۱۸</sup> ( $MSE$ ) و بیشینه مربعات خطا ( $MAXSE$ ) در سنجش خطاهای بزرگ بهره می‌گیرند. برای سنجش هم‌بستگی بین مقادیر خروجی واقعی و مقادیر پیش‌بینی شده توسط مدل، معیار  $R$  به کار گرفته می‌شود. لازم به ذکر است که با کاهش شش معیار اول و افزایش معیار  $R$  مدل بهبود می‌یابد.

#### ۴.۵. نتایج پیش‌بینی

همان‌طور که پیش‌تر بیان شد، پیش‌بینی اوج بار ماه بعد برای هر یک از سه خوشه با استفاده از معادلات رگرسیونی ارائه شده به صورت جداگانه انجام شد، به طوری که داده‌های ۱۱ سال برای تعیین پارامترهای مدل‌ها و داده‌های سه سال برای پیش‌بینی (اعتبارسنجی) مدل‌های رگرسیونی حاصل شده برای هر خوشه به کار گرفته شد. معیارهای یادشده در بخش قبل نیز عملکرد هر یک از خوشه‌ها را نشان می‌دهند. جدول ۲ بهترین مقادیر حاصل از پیش‌بینی با استفاده از مدل‌های رگرسیون تلفیقی برای هر یک از سه خوشه حاصل از خوشه‌بندی را نمایش می‌دهد. همچنین نتایج مدل ارائه شده بدون استفاده از  $PCA$  قابل مقایسه است.

چنان‌که مشاهده می‌کنید در خوشه‌های اول و دوم تمامی معیارهای عملکرد در خوشه‌ی سوم معیارهای مرتبط با میانگین خطا با به کارگیری مدل رگرسیون تلفیقی نسبت به مدل رگرسیون تلفیقی بدون استفاده از  $PCA$  جواب‌های بسیار بهتری ارائه می‌کنند. به طوری که در خوشه‌ی ۱ مدل رگرسیون تلفیقی پیش‌بینی ماه بعد به طور متوسط با ۱٫۱۱ درصد خطا برآورد می‌شود، در حالی که اگر از  $PCA$  برای پیش‌پردازش داده‌ها استفاده نکنیم، این مقدار برابر ۱٫۳۸ درصد خواهد بود. در مورد دو خوشه‌ی ۲ و ۳ نیز شاهد وضعیت مشابه خواهیم بود، به طوری که میزان برآورد خطا با استفاده از مدل رگرسیون تلفیقی برای خوشه‌ی ۲ برابر ۳٫۰۷ درصد و برای خوشه‌ی ۳ برابر ۱٫۶۵ درصد است. یادآور می‌شود که یکی از دلایلی که موجب بالا بودن میانگین خطای خوشه‌ی ۲ نسبت به خوشه‌های ۱ و ۳ شده است، کم بودن تعداد داده‌های مورد استفاده برای برآورد دقیق‌تر ضرایب معادله‌ی رگرسیونی است. همچنین می‌توان به نوسانات زیاد دمای هوا در ماه‌های اردیبهشت و مهر که در این خوشه قرار گرفته‌اند اشاره کرد که تأثیر مستقیم و متفاوتی بر اوج بار ماهانه طی سال‌های مورد نظر داشته‌اند.

در خصوص بالاتر بودن مقادیر معیارهای بیشینه خطای مدل رگرسیون تلفیقی نسبت به همین مدل بدون استفاده از  $PCA$  باید بیان کرد، چون بیشینه خطای هر سه خوشه باید در حالت کلی برای ارزیابی عملکرد مدل پیشنهادی مورد استفاده قرار گیرد. با توجه به این‌که بیشینه خطا در بین سه خوشه‌ی مدل رگرسیون تلفیقی مربوط به خوشه‌ی دوم با مقدار ۵٫۲۶ درصد است و بیشینه خطا در مدل رگرسیون تلفیقی بدون استفاده از  $PCA$  نیز مربوط به خوشه‌ی دوم بوده و برابر با ۷٫۶۳ درصد است. بنابراین نتیجه می‌گیریم که به طور کلی بالاتر بودن معیارهای بیشینه در خوشه‌ی سوم نقطه ضعف مدل پیشنهادی نیست. در ادامه برای نمایش میزان تأثیرگذاری خوشه‌بندی ماه‌های سال در بهبود پیش‌بینی اوج بار ماه بعد، در شکل ۹ نمودارهای مقادیر واقعی و پیش‌بینی شده اوج بار ماهانه طی سال‌های ۱۳۸۲ تا ۱۳۸۴ (سال‌های مورد استفاده برای اعتبارسنجی مدل) در دو حالت پس از انجام خوشه‌بندی و بدون انجام خوشه‌بندی با یکدیگر مقایسه می‌شوند.

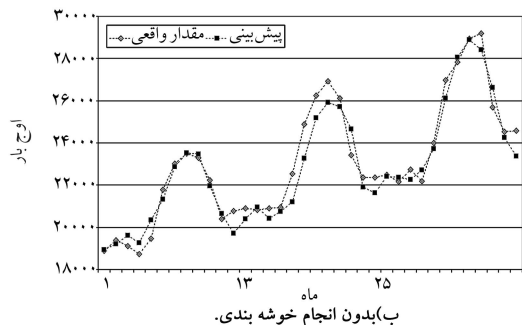
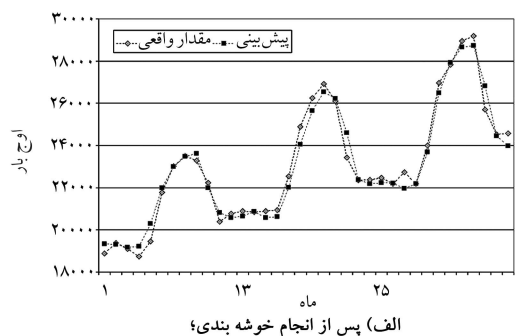


شکل ۸. تحلیل باقی مانده‌ها در دو حالت.

خطی باشد (خطی در درصد، نه در قدرمطلق خطا)،  $MAPE$  برای سنجش میزان خطا کافی خواهد بود - اگرچه مطالعات اخیر و تجربه‌ی اپراتورهای سیستم نشان می‌دهد که تابع خسارت در مسئله‌ی پیش‌بینی بار به وضوح غیرخطی است و در ضمن خطاهای بزرگ ممکن است اثرات خطرناکی برای استفاده‌ی همگانی داشته باشد. به همین دلیل است که اندازه‌گیری مبتنی بر مربعات خطا توصیه می‌شود، چون آنها خطاهای بزرگ را بیشتر جریمه می‌کنند.<sup>[۱]</sup>

جدول ۲. نتایج حاصل از پیش‌بینی با استفاده از مدل رگرسیون تلفیقی و مقایسه آن با حالت بدون استفاده از PCA.

خوشه‌ها	معیار عملکرد	مدل رگرسیون تلفیقی بدون استفاده از PCA	مدل رگرسیون تلفیقی
خوشه ۱	MAE	۲۹۰٫۴	۲۳۵٫۵
	MAPE	۱٫۳۸	۱٫۱۱
	MSE	۱۲۵۹۳۹	۸۶۹۶۰
	MAXAE	۸۷۵٫۹	۸۱۰٫۲
	MAXAPE	۳٫۸۵	۳٫۵۶
	MAXSE	۷۶۷۱۵۷	۶۵۶۴۳۷
خوشه ۲	R	۰٫۹۷۹	۰٫۹۸۷
	MAE	۸۶۱٫۵	۶۹۱٫۴
	MAPE	۳٫۶۶	۳٫۰۷
	MSE	۱۱۶۷۰۵۵	۵۷۲۸۱۸
	MAXAE	۱۹۶۰٫۹	۱۰۸۰٫۲
	MAXAPE	۷٫۶۳	۵٫۲۶
خوشه ۳	MAXSE	۳۸۴۵۲۴۴	۱۱۶۶۹۰۰
	R	۰٫۹۳۳	۰٫۹۳۶
	MAE	۴۳۸٫۳	۴۲۳٫۹
	MAPE	۱٫۶۹	۱٫۶۵
	MSE	۲۵۴۶۸۰	۲۳۲۳۸۵
	MAXAE	۹۳۱٫۸	۹۴۳٫۳
	MAXAPE	۳٫۳۵	۳٫۷۹
	MAXSE	۸۶۸۱۷۸	۸۸۹۸۷۰
	R	۰٫۹۷۷	۰٫۹۸۸



شکل ۹. نمودار مقایسه مقادیر واقعی و پیش‌بینی شده اوج بار ماهانه طی سال‌های ۱۳۸۲-۱۳۸۴ در دو حالت.

جدول ۳. نتایج کلی عملکرد دو حالت پیش‌بینی.

معیار عملکرد	مدل پیش‌بینی	مدل رگرسیون تلفیقی	مدل رگرسیون
MAE	۳۷۴٫۳	۵۴۳٫۳	
MAPE	۱٫۶۲	۲٫۳۴	
MSE	۲۱۶۴۱۲	۴۶۴۱۰۸	
MAXAE	۱۰۸۰٫۲	۱۶۰۸٫۴	
MAXAPE	۵٫۲۶	۶٫۴۶	
MAXSE	۱۱۶۶۹۰۰	۲۵۸۶۹۱۸	
R	۰٫۹۸۷	۰٫۹۷۳	

با خطایی معادل ۱٫۶۲ درصد خطا انجام دهد و در هر پیش‌بینی حدود ۳۷۴٫۳ مگاوات خطا وجود دارد. این درحالی است که با به‌کارگیری مدل رگرسیون بدون استفاده از خوشه‌بندی می‌توان به‌طور متوسط اوج بار مصرفی ماه بعد را با خطایی معادل ۲٫۳۴ درصد پیش‌بینی کرد؛ در هر پیش‌بینی حدود ۵۴۳٫۳ مگاوات خطا وجود دارد. این مقایسه نشان می‌دهد که خوشه‌بندی ماه‌های سال با بهبود برازش رگرسیونی موجب کاهش خطای پیش‌بینی به میزان قابل توجهی شده است.

## ۶. نتیجه‌گیری

در این نوشتار یک مدل رگرسیون تلفیقی برای پیش‌بینی اوج بار ماه بعد ارائه شد. در این تحقیق، ابتدا از یک نقشه خودسازمان‌ده برای تعیین بهترین خوشه‌بندی به‌منظور تقسیم‌بندی ماه‌های سال براساس دو فاکتور اوج بار مصرفی ماهانه و میانگین وزنی متوسط دمای هوای ماهانه سه شهر تهران، تبریز و اهواز استفاده شد. به‌دلیل اختلاف

مقایسه‌ی دو نمودار الف و ب در شکل ۹ بیان‌گر این نکته است که در نمودار الف (یعنی پس از استفاده از خوشه‌بندی) تطابق بیشتری بین مقادیر واقعی و پیش‌بینی‌شده‌ی اوج بار ماهانه نسبت به نمودار ب وجود دارد. این نکته مؤید بهبود نتایج پیش‌بینی با استفاده از مدل ارائه‌شده است.

همچنین، به‌منظور مقایسه‌ی نتایج حاصل از مدل رگرسیون تلفیقی و مدل رگرسیون بدون استفاده از خوشه‌بندی، میانگین وزنی نتایج حاصله برای هر یک از سه خوشه محاسبه شده است. برای این کار، و به‌منظور محاسبه‌ی میانگین مقادیر حاصله برای سه معیار عملکرد اول برحسب تعداد نمونه‌های موجود در هر خوشه، یک ضریب وزنی مطابق رابطه‌ی زیر در نظر گرفته شده است:

میانگین سه خوشه =  $۱۸ * \text{مقدار معیار عملکرد خوشه اول} + ۶ * \text{مقدار معیار عملکرد خوشه دوم} + ۱۲ * \text{مقدار معیار عملکرد خوشه سوم}$ .

از طرف دیگر، برای سه معیار عملکرد بعدی نیز بیشترین مقادیر سه خوشه لحاظ شده است. یادآور می‌شود که، به‌عنوان مثال، ضریب وزنی ۱۸ از طریق ضرب مدت سه‌سال در شش‌ماهی که در خوشه‌ی ۱ قرار گرفته‌اند، قابل محاسبه است. درخصوص خوشه‌های ۲ و ۳ نیز ضریب وزنی به‌همین ترتیب محاسبه می‌شود. در جدول ۳ نتایج حاصل از پیش‌بینی با و بدون استفاده از خوشه‌بندی مقایسه شده است.

با توجه به جدول ۳، مدل رگرسیون تلفیقی می‌تواند به‌طور میانگین هر پیش‌بینی

روش رگرسیون برای پیش‌بینی اوج بار ماهانه هر یک از خوشه‌ها به صورت جداگانه مورد استفاده قرار گرفت. با مقایسه‌ی نتایج حاصل از روش مدل رگرسیون تلفیقی و روش رگرسیون نتیجه می‌گیریم که خوشه‌بندی ماه‌های سال با بهبود برازش رگرسیونی موجب کاهش خطای پیش‌بینی شده است.

مقیاس این دو فاکتور، از هنجارسازی داده‌ها پیش از آغاز عملیات خوشه‌بندی استفاده شد. برای تعیین بهترین خوشه‌بندی، با استفاده از شاخص دیویس - بولدین در نهایت سه خوشه انتخاب شد. سپس برای کاهش ابعاد متغیرهای ورودی و بهبود نتایج پیش‌بینی روش *PCA* پیش از تعیین مدل رگرسیونی به‌کار گرفته شد. در ادامه،

## پانویس

1. principal component analysis
2. self-organizing map
3. artificial intelligence
4. artificial neural networks
5. multiplicative
6. additive
7. width of the neighborhood function
8. davies-bouldin index
9. spherical clusters
10. random order incremental training
11. feature maps
12. rotated component matrix
13. kaiser-meyer-olkin measure of sampling adequacy
14. residuals
15. mean absolute percentage error
16. loss function
17. mean absolute error
18. mean of square error

## منابع

1. Hippert, H.S.; Pedreira, C.E., and Castro, S.R. "Neural networks for short-term load forecasting: a review and evaluation", *IEEE Transactions on Power Syst.* **16**(1), pp. 44-55 (2001).
2. Yalcinoz, T., and Eminoglu, U. "Short term and medium term power distribution load forecasting by neural networks", *Energy Conversion and Management*, **46**, pp. 1393-1405, (2005).
3. Kareem, Y.H., and Majeed, A.R. "Monthly peak-load demand forecasting for sulaimany governorate using SARIMA", *IEEE*, pp. 1-5 (2006).
4. Phimpachan, S.; Chamnonghai, K.; Kumhom, P.; Jit-tiwarangkul, N., and Sangswang, A. "Energy and peak load forecast models using neural network for fast developing area", International Symposium on Communications and Information Technologies, Sapporo, Japan, October 26-29 (2004).
5. Phimpachanh, S.; Chamnonghai, K.; Kumhom, P., and Sangswang, A. "Using neural network for long term peak load forecasting in vientiane municipality", *IEEE*, pp. 319-322 (2004).
6. Temraz, H.K.; Salama, M.A., and Quintana, V.H., "Application of the decomposition technique for forecasting the load of a large electric power network", *IEEE Proc-Cener. Transm. Distrib.*, **143**(1), (January 1996).
7. Oja, M.; Kaski, S., and Kohonen, T. "Bibliography of self-organizing map (SOM) papers: 1998-2001 addendum", *Neural Computing Surveys*, **3**, pp. 1-156, (2002).
8. Vesanto, J., and Alhoniemi, E. "Clustering of the self-organizing map", *IEEE Transactions on Neural Networks*, **11**(3), pp. 586-600 (2000).
9. Demuth, H., and Beale, M. *MATLAB 6.5/ Neural Network Toolbox*, version 4, The MathWorks, Inc.,(CD-ROM),1-840 (2002).
10. Bishop, C.M. *Neural Networks for Pattern Recognition*, First Edition, Oxford University Press (1995).
11. Gujarati, D. *Basic Econometrics*, Third edition, McGraw-Hill (1995).
12. Sharma, S. *Applied Multivariate Techniques*, First Edition, John Wiley & Sons, Inc. (1996).

