کاربست تحلیل مولفه اصلی برای داده‌های هواشناختی در انتخاب ورودی شبکه عصبی مصنوعی

مهشید کاویانی و سید مجید میرزکنی *

گروه فیزیک فضا، دانشکده فیزیک، دانشگاه یزد، ایران

(تاریخ دریافت: ۱۳۹۲/۱۱/۲۱، تاریخ پذیرش: ۱۳۹۳/۹/۴)

چکیده

«انتخاب ورودی» صحیح، هوشمند و متناسب با هدف به‌کمک روش مناسب، اولین گام در طراحی شبکه عصبی مصنوعی (Artificial Neural Network) با نماد ANN به‌منظور پیش‌بینی است. دو رهیافت به‌منظور انتخاب داده‌های ورودی مناسب وجود دارد؛ در رهیافت اول سری زمانی پراسنج (پارامتر) موردنظر، یعنی هدف ANN، در سال‌های گذشته و در رهیافت دوم پراسنج‌هایی که با هدف رابطه خطی یا غیرخطی دارند، مورد استفاده قرار می‌گیرند. در اثنای فرایند انتخاب ورودی برمبنای رهیافت دوم، به علت حجم زیاد داده ورودی، خطاهای اندازه‌گیری، حضور داده‌های غیرمعمول و همبستگی بین متغیرهای ورودی، خطا افزایش و دقت پیش‌بینی کاهش می‌یابد. در اکثر موارد به علت فقدان اطلاعات دقیق مربوط به جزئیات داده به‌ناچار از روش سعی و خطا برای انتخاب ترکیب مناسبی از داده‌های ورودی و حذف داده‌های غیرمعمول استفاده می‌شود. در پژوهش حاضر، به‌منظور اجتناب از کاربرد روش سعی و خطا از تحلیل مولفه اصلی برای تعیین اطلاعات دقیق مربوط به جزئیات داده استفاده شده است. تحلیل مولفه اصلی توانایی‌های متعددی از جمله؛ کاهش ابعاد داده، استخراج مُدهای تغییرپذیری متغیرهای ورودی، از بین بردن همبستگی بین داده‌های خام و حذف داده‌های غیرمعمول دارد. در پژوهش حاضر، پراسنج‌های هواشناختی مرتبط با دما، به‌منظور تعیین داده‌های ورودی مناسب برای پیش‌بینی میانگین دمای روزانه سال ۲۰۰۹ در ایستگاه همدیدی شهرستان یزد در یک دوره آماری ۲۹ ساله (۱۹۸۰ تا ۲۰۰۸) مورد بررسی قرار گرفته‌اند. سپس با کاربست تحلیل مولفه اصلی، ضمن حذف همبستگی بین متغیرهای ورودی، داده‌های غیرمعمول شناسایی و حذف می‌شوند. درنهایت با بررسی مدهای تغییرپذیری هریک از پراسنج‌ها و مقایسه آنها با هدف ANN، داده‌های ورودی مناسب انتخاب می‌شود. نتایج نشان می‌دهد که با استفاده از توانایی‌های متعدد تحلیل مولفه اصلی می‌توان انتخاب ورودی صحیح، هوشمند و متناسب با هدف ANN را بدون استفاده از روش سعی و خطا عملی ساخت.

واژه‌های کلیدی: انتخاب ورودی، شبکه عصبی مصنوعی، پیش‌بینی، کاهش ابعاد، پراسنج‌های هواشناختی، مدهای تغییرپذیری

Application of principal component analysis to meteorological data in ANN input selection

Mahshid Kaviani and Seyed Majid MirRokni*

Space Physics Group, Department of Physics, Yazd University, Iran

(Received: 10 February 2014, accepted: 25 November 2014)

Summary

Properly intelligent "input selection" tailored to the target using an appropriate method is the first step in the design of Artificial Neural Network (ANN) for prediction. ANN architecture is not predetermined; the weights are determined based on input data during

* نگارنده رابط: mirrokni@yazd.ac.ir :*Corresponding author

the training process. Therefore, when input data is richer, ANN will be better trained and will have a better performance in predicting meteorological parameters. To solve nonlinear equations governing atmospheric motions, for which no general solutions are known, meteorologists have to use the appropriate approximations for prediction. Using the ability of the ANN to consider nonlinear effects, meteorologists will be able to predict most of meteorological parameters without considering the nonlinear equations governing atmospheric motions. There are two approaches for selecting the appropriate input data. In the first approach, time series of the desired parameter, ANN target, such as temperature, relative humidity, pressure, and wind speed from the previous years are used, while in the second approach, the parameters that have a nonlinear or linear relationship with the ANN target are used. Due to a large volume of input data, errors of measurement, the presence of unusual data and correlation between input variables, in the case of second approach, error increases and prediction accuracy decreases. In most cases, due to the lack of detailed information concerning the data, the trial and error method has to be used to select a proper combination of input data and elimination of unusual data. The trial and error method is one of the easiest methods for solving problems. Since in this method the governing relationships among the parameters are not considered, the solutions may not fit in the physical situation. In the present research, to avoid using the trial and error method, we use Principal Component Analysis (PCA) in order to determine the detailed information concerning the details of the input data. PCA has several abilities such as reduction of dimensions of data, extraction of variability modes of data, eliminating the correlation between raw data, and deletion of unusual data. These abilities can be used in various applications. For example, it is possible to reduce dimension data using PCA when we deal with a large volume of raw data. In fact, we achieve simultaneously three targets by using PCA. The first target is reduction of dimensions of the data; therefore, the training process of ANN performs better than the case when we use raw data. The second and the third targets are the extraction of variability modes and deletion of unusual data; thus ANN does not deviate and overtraining does not occur. In our accompanying research presented at the First Computational Physics Conference in 20−22 January 2014, we demonstrated that these abilities of PCA were very important in properly intelligent input selection tailored to the target. Meteorological parameters associated with temperature to determine appropriate parameters for predicting the average daily temperature in 2009 in Yazd synoptic station in a 29-year period (1980 to 2008) has been analyzed by using PCA. The results showed that using numerous capabilities of PCA, a correct, intelligent input selection appropriate for the ANN target without using the trial and error methods is possible.

Keywords: Input selection, artificial neural network, forecasting, reduction of dimensions, meteorological parameters

۱ مقدمه

انتخاب ورودی صحیح، هوشمندانه و متناسب با هدف، در طراحی ANN برای پیش‌بینی حائز اهمیت است (کستلانو و فانلی، ۲۰۰۰؛ بک و ترپنبرگ، ۲۰۰۱؛ چن و همکاران، ۲۰۰۹). فرایندهای دینامیکی و ترمودینامیکی جوّ، بی‌نظمی‌هایی در سامانه‌های جوّی ایجاد می‌کنند که ماهیت غیرخطی دارند. حل معادلات غیرخطی حاکم بر جوّ و یافتن روابط بین پراسنج‌های (پارامترهای) هواشناختی بسیار دشوار و پیچیده است (هلتون، ۲۰۰۴). علی‌رغم ماهیت آشوبی سامانه‌های جوّی، هریک از پراسنج‌های

هواشناختی که ظاهراً در زمان به‌صورت تصادفی متغیرند، اطلاعاتی تحت این فرایندها دارند (پدلاسکی، ۱۹۸۶). با این رویکرد، توانایی ANN، در تقریب توابع غیرخطی و اجتناب از پیچیدگی‌های معادلات حاکم بر جوّ، حائز اهمیت است (چرنساید و همکاران، ۱۹۹۴؛ گاردنر و دورلینگ، ۱۹۹۸؛ فابین و همکاران، ۲۰۰۷؛ اردیل و ارکاکلیوگلو، ۲۰۱۳).

ساختار ANN یک برنامه از قبل تعیین شده نیست و وزن‌های آن در فرایند آموزش براساس داده‌های ورودی تعیین می‌شود. بنابراین هرچه داده‌های ورودی غنی‌تر باشد شبکه بهتر آموزش می‌بیند و در پیش‌بینی داده‌های جدید عملکرد مناسب‌تری دارد (ترسپ و همکاران، ۱۹۹۵؛ پاسینی و همکاران، ۲۰۰۱). این رهیافت، بدون در نظر گرفتن و حل معادلات حاکم بر سامانه، به‌دنبال یافتن رابطه‌های بین متغیرهای ورودی برای رسیدن به هدف موردنظر است. برای پیش‌بینی ضروری است که داده‌های ورودی قادر به توصیف ویژگی‌های گوناگون سامانه مورد بررسی باشد. هرچه اطلاعات دقیق‌تری از ویژگی‌های سامانه در اختیار ANN قرار بگیرد، آثار پیچیدگی روابط غیرخطی بیشتر نمایان می‌شود و ANN با دقت بیشتری توانایی تقریب تابع حاکم بر مجموعه داده‌ها را دارد (هاسون، ۱۹۹۵؛ مهروترا، ۱۹۹۷؛ گرپ، ۲۰۰۷). در سامانه‌های آشوبی، مانند جوّ، تعیین و بررسی متغیرهایی که روابط خطی یا غیرخطی با داده‌های هدف دارند، حائز اهمیت است. شناخت داده‌های هدف و سامانه مورد بررسی و داشتن مهارت لازم در انتخاب هوشمندانه ورودی، نقش موثری در تفهیم روابط موجود بین داده‌های ورودی، برای ANN دارد.

پس از شناسایی داده‌های مرتبط با هدف، به‌منظور کسب آگاهی از اطلاعات جزئی‌تر از این مجموعه، از روش‌های پیش‌پردازش استفاده می‌شود (جانیتا و برین، ۲۰۰۸؛ آزاده و همکاران، ۱۳۹۰؛ قبائی سوق و همکاران، ۱۳۸۹). در مجموعه ورودی‌های منتخب، داده‌های نامأنوس و غیرمعمول وجود دارند. بخشی از این داده‌ها به‌علت خطای اندازه‌گیری ناشی از عواملی مانند ضعف در سامانه ثبت اطلاعات، خطای دستگاه‌های اندازه‌گیری، خطاهای انسانی و موارد دیگر، باعث انحراف روند آموزش شبکه می‌شوند (فانگ، ۲۰۱۰؛ مشاری و همکاران، ۱۳۸۸). همچنین، شرایط جوّی خاص نیز با ایجاد تغییر در الگوهای وضع هوا در یک زمان خاص، باعث می‌شوند پراسنج هواشناختی موردنظر از الگوی طبیعی خود پیروی نکند و جزء داده‌های غیرمعمول به‌شمار آید (مانندِ؛ مشاری و همکاران، ۱۳۸۸). به‌علاوه، متغیرهای خامی که برای آموزش ANN مورد استفاده قرار می‌گیرند، شامل اطلاعاتی هستند که با یکدیگر مرتبط‌اند (جانیتا و برین، ۲۰۰۸). این همبستگی در سراسر فرایند آموزش می‌تواند سبب گمراهی شبکه در رسیدن به هدف موردنظر شود و توانایی عمومیت ANN را در پیش‌بینی داده‌های جدید کاهش دهد (کرمر، ۱۹۹۱؛ جانیتا و برین، ۲۰۰۸؛ کاویانی و همکاران، ۱۳۹۲). پیش‌پردازش داده‌ها سبب شناخت مهم‌ترین پراسنج‌های موثر بر پدیده موردنظر برای مدل‌سازی با استفاده از روش‌های هوشمند است (قبائی سوق، ۱۳۸۹).

پیش‌پردازش داده‌ها به‌کمک روش تحلیل مولفه اصلی (Principal Component Analysis) با نماد PCA، یک روش چندمتغیره آماری است که با توجه به ساختار داده‌ها، برای شناسایی عوامل ایجاد تغییرات و داده‌های نامأنوس از آن استفاده می‌شود (فانگ، ۲۰۱۰؛ نوری قیداری، ۱۳۸۹). در این روش، طبقه‌بندی داده‌ها به‌کمک یک تبدیل خطی، داده‌ها را برحسب بیشترین مقادیر واریانس برمی‌گرداند (مانندِ؛ ایلین و ریکو، ۲۰۱۰). این تبدیل، بردارها را به دستگاه مختصات جدیدی می‌برد که بزرگ‌ترین واریانس داده روی اولین محور مختصات قرار می‌گیرد، دومین واریانس روی محور دوم مختصات واقع

می‌شود و روند به همین صورت برای سایر داده‌ها ادامه می‌یابد. با این رهیافت، این روش برای کاهش ابعاد نیز کاربرد دارد (مانندِ؛ همت‌پور و هاشمی، ۱۳۹۰). بدین‌ترتیب، مولفه‌هایی که بیشترین سهم واریانس را دارند، می‌توانند در تعیین روند تغییرات مورد توجه قرار بگیرند. روش PCA برای تحلیل و طبقه‌بندی داده‌هایی با حجم زیاد نیز کاربرد دارد، چراکه بین تعداد زیادی از متغیرهایی که به ظاهر ارتباط ندارند، رابطه خاصی را به‌صورت یک مدل برقرار می‌سازد (مانندِ؛ فتاحی و بهیار، ۱۳۹۰). همچنین این رهیافت، یکی از روش‌های انتخاب ویژگی است، که در فضایی با ابعاد کمتر و ساده‌تر، به بررسی ویژگی‌ها می‌پردازد (مانندِ؛ همت‌پور و هاشمی، ۱۳۹۰). در بررسی داده‌های هواشناختی و اقلیم‌شناختی، PCA را می‌توان برای انتخاب ترکیب درستی از مجموعه داده‌ها به‌منظور تحلیل ویژگی آنها ، حذف داده‌های نامأنوس، تعیین روند تغییرات، پهنه‌بندی‌های اقلیمی و به‌دست آوردن شاخص‌های اقلیمی در منطقه‌ای خاص به‌کار برد.

لامارک و فریتس (۱۹۷۱) به‌کمک PCA، الگوهای بی‌هنجاری مقادیر بارش ماهانه غرب امریکا را در یک دوره آماری ۳۵ ساله به‌دست آوردند و نتایج این تحقیق رابطه متناظر این الگوها با رشد درختان در آن ناحیه را نمایش داد. مونیاتی (۲۰۰۴) به‌منظور ذخیره ساختن اطلاعات مکانی و زمانی داده‌های تصاویر پردازش شده سنجش از دور چندنواری، در یک مجموعه با کمترین همبستگی و بررسی روند تغییرات، روش PCA را به‌کار برد. قنواتی و همکاران (۱۳۶۸) برای تعیین تغییرات مورفودینامیکی حوضه آبخیز طالقان، بعد از پردازش تصاویر ماهواره‌ای، روش‌های PCA و منطق فازی را به‌کار بردند و درنهایت با تحلیل نتایج محل بیشترین و کمترین تغییرات را در این منطقه ارزیابی کردند. موسوی و همکاران (۱۳۹۱) برای تعیین مولفه‌های اقلیمی موثر بر استان‌های کرمانشاه و کردستان با کاربست PCA روی ۲۲ متغیر اقلیمی، ۶ مولفه؛ دما، بارش، میزان رطوبت و تغییرات دمایی، ساعات آفتابی، روزهای توفانی و روزهای همراه با گردوخاک را موثرترین پراسنج‌های اقلیمی حاکم بر آن منطقه معرفی کردند. پوراصغر و همکاران (۱۳۹۲) رژیم بارش سالانه نیمه جنوبی ایران را با کاربست PCA و خوشه‌بندی به چهار زیرمنطقه همگن تقسیم‌بندی کردند و نتایج این پژوهش نشان داد ۵ مولفه اصلی اول، با گردش جوّی کنترل می‌شود و بارش سالانه این ناحیه متأثر از کم‌فشار سودانی و مدیترانه است.

هدف از پژوهش حاضر، انتخاب مجموعه داده ورودی مناسب برای ANN، به‌کمک روش PCA و بدون استفاده از روش سعی و خطا، به‌منظور پیش‌بینی میانگین روزانه دمای ایستگاه همدیدی شهرستان یزد (ایستگاه منتخب) در ۲۰۰۹ است. در بخش بعدی، ضمن معرفی دو رهیافت کلی در گام نخست انتخاب داده‌های ورودی، داده‌های مورد استفاده در این پژوهش مورد ملاحظه قرار می‌گیرد. سپس در بخش روش‌ها با استفاده از PCA جزئیات مجموعه داده منتخب تحلیل می‌شود و در بخش آخر نتایج پژوهش مورد بحث قرار می‌گیرد.

۲ داده‌ها و روش تحقیق

۲-۱ داده‌ها

در انتخاب ورودی مناسب، برای پیش‌بینی پراسنج‌های هواشناختی، دو رهیافت کلی را می‌توان دنبال کرد؛ رهیافت اول- استفاده از مقادیر سری زمانی هدف موردنظر در سال‌های گذشته و رهیافت دوم- یافتن و به‌کارگیری پراسنج‌هایی است که با هدف، رابطه خطی یا غیرخطی دارند (کائور و ساین، ۲۰۱۱؛ سنتاموریس و همکاران، ۲۰۱۲؛ اردیل و ارکاکلیوگلو، ۲۰۱۳؛ حیاتی و محبی، ۱۳۸۶؛ قبائی سوق و همکاران، ۱۳۸۹). جمع‌آوری اطلاعات هریک از مجموعه داده‌های پیش‌گفته، با استفاده از داده‌های بازتحلیل، ایستگاهی، ماهواره‌ای، رادار و مانند

آن، در یک دوره آماری معین و یک منطقه خاص، امکان‌پذیر است (چرنساید و همکاران، ۱۹۹۴؛ رحیمی‌خوب، ۱۳۸۶). با استفاده از رهیافت اول، تغییرات پراسنج موردنظر را در مقیاس‌های زمانی متفاوت در نظر می‌گیرند و با انتخاب هوشمندانه ورودی‌هایی که به بازه‌های زمانی خاصی از گذشته این مقادیر نسبت داده می‌شوند، دینامیک حاکم بر دوره آماری موردنظر شناسایی می‌شود. سپس با آموزش ANN، هدف را می‌توان پیش‌بینی کرد. در رهیافت دوم، مقادیر پراسنج‌های مرتبط با هدف، ضمن تعیین وزن‌های بهینه در فرایند آموزشی ANN، می‌توانند معرف شاخص‌های دینامیکی و ترمودینامیکی جوّ باشند و این خصوصیات را در حین فرایند آموزش، در یک دوره آماری، برای ANN تفهیم کنند. بنابراین براساس داده‌های در دسترس و رهیافتی که برای پیش‌بینی هدف در نظر گرفته شده است، می‌توان هریک از این روش‌ها را به‌کار برد.

در پژوهش حاضر، از داده‌های ایستگاه منتخب، در یک دوره آماری ۲۹ ساله (۱۹۸۰ تا ۲۰۰۹) استفاده شده است. به‌منظور پیش‌بینی مقادیر میانگین روزانه دما در سال ۲۰۰۹، پراسنج‌هایی که با دما رابطه دارند، ازجمله؛ داده‌های ساعات آفتابی روزانه، میانگین روزانه تندی باد، میانگین روزانه بارش، میانگین رطوبت نسبی، میانگین روزانه فشار، میانگین روزانه تبخیر، میانگین روزانه تابش، میانگین روزانه فشار بخار آب و میانگین روزانه دما در سال‌های قبل از ۲۰۰۹، در این دوره آماری جمع‌آوری شد. با توجه به ساختار ANN و اهمیت داده‌های ورودی در فرایند آموزشی، هرچه اطلاعات غنی‌تر باشد، ANN در پیش‌بینی هدف موردنظر عملکرد مناسب‌تری دارد. براین‌اساس، داده‌های میانگین روزانه تابش، میانگین روزانه تبخیر و ساعات آفتابی روزانه، به علت اطلاعات مفقود در بسیاری از سال‌ها در این دوره آماری، حذف و سایر پراسنج‌ها برای عملیات پیش‌پردازش در محیط نرم‌افزار مَتلَب فراخوانی شد.

۲-۲ روش تحقیق

با فراخوانی سری زمانی شش پراسنج ورودی منتخب در دوره آماری ۱۹۸۰ تا ۲۰۰۸، هرکدام یک ماتریس ۲۹×۳۶۵ و هدف ANN یک بردار ۱×۳۶۵ در محیط نرم‌افزار مَتلَب ایجاد می‌کند. هر ماتریس x از مرتبه $n \times p$، ماتریس کوواریانس نامنفی S دارد، که از رابطه $S = \frac{X^T X}{n-1}$ به‌دست می‌آید. در این رابطه $X = x - \bar{x}h$ است که $\bar{x}$ میانگین روی سطرهای ماتریس x و h ماتریس $1 \times n$ با مقادیر یک را نشان می‌دهد. با حل معادله $|S - \lambda I| = 0$ مقادیر ویژه آن محاسبه می‌شود. با داشتن مقادیر ویژه $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ و بردارهای ویژه متناظر $a_1, a_2, \ldots a_p$ و کاربست PCA روی بردار X، متغیرهای $\gamma_1, \gamma_2, \ldots, \gamma_p$، یعنی مولفه‌های X به‌دست می‌آید (مانند؛ نوری قیداری، ۱۳۸۹):

$$\gamma_1 = a_{11}X_1 + a_{21}X_2 + \ldots + a_{p1}X_p$$
$$\gamma_2 = a_{12}X_1 + a_{22}X_2 + \ldots + a_{p2}X_p$$

و درنهایت

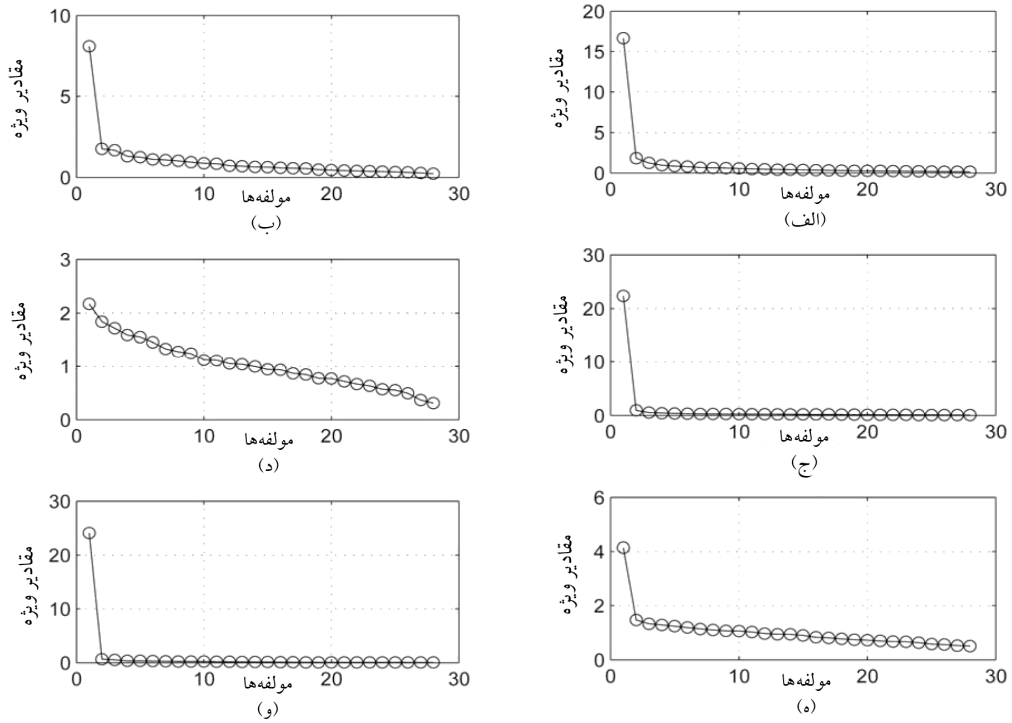
$$\gamma_p = a_{1p}X_1 + a_{2p}X_2 + \ldots + a_{pp}X_p$$

که $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)$ بردار مولفه‌ها است. واریانس مولفه‌های γ، متناظر با همان مقادیر ویژه ماتریس کوواریانس است. یعنی واریانس «اولین مولفه»، یعنی مولفه اصلی یا مُد پیشرو γ_1 ، برابر λ_1 است. پس مُد پیشرو بیشترین واریانس را دارد و نشان‌دهنده زیاد بودن توانایی آن در شناسایی تغییرات پراسنج موردنظر در دوره آماری پیش‌گفته است. با این رویکرد واریانس و واریانس تجمعی مولفه‌های هریک از پراسنج‌ها به‌دست می‌آید و براساس معیارهای آزمون اسکری (scree test)، ارزش ویژه و

مقدار واریانس تجمعی، می‌توان به جزئیات روند تغییرات و پراکندگی هر پراسنج در این دوره آماری دست یافت. در آزمون اسکری، نمودار مقادیر ویژه برحسب مولفه‌ها رسم شده است و محل شکستگی نمودار (در صورت وجود) فاصله بین مُد پیشرو و سایر مولفه‌ها را نشان می‌دهد (مانندِ؛ موسوی و همکاران، ۱۳۹۱). معیار ارزش ویژه، مولفه‌هایی با مقدار ویژه بزرگ‌تر از یک را در نظر می گیرد و از سایر مولفه‌ها صرف‌نظر می‌کند. درنهایت معیار واریانس تجمعی، مولفه‌هایی که درصد تجمعی بیشتری دارند را برای ادامه کار لحاظ می‌کند (مجموعه پژوهش‌های اقتصادی، ۱۳۸۹). به‌این‌ترتیب انتظار می‌رود، ضمن حذف همبستگی بین متغیرهای ورودی، جزئیات روند تغییرات و پراکندگی داده‌های هریک از پراسنج‌ها برای مراحل بعدی مشخص شود. گام بعدی حذف داده‌های نامأنوس است. با رسم نمودارهای مولفه اول برحسب مولفه دوم و بار دیگر برحسب مولفه سوم داده‌های نا مأنوس آشکار می‌شوند (مانندِ؛ نوری قیداری، ۱۳۸۹). پس از شناسایی و حذف داده‌های نامأنوس می‌توان جای این داده‌ها را با مقادیر میانگین روز قبل و بعد و یا روش‌های درون‌یابی پر کرد. روش PCA، برای جایگزین کردن داده‌های مفقود نیز کاربرد دارد (مانندِ؛ ایلین و ریکو، ۲۰۱۰). پس از بررسی روند تغییرات، ضریب همبستگی مولفه‌های اصلی هریک از پراسنج‌ها با هدف ANN محاسبه می‌شود و سپس با تحلیل دینامیکی و ترمودینامیکی، ورودی مناسب آن تعیین می‌شود.

۳ بحث

در پژوهش حاضر، با کاربست PCA و با استفاده از مقادیر ویژه ماتریس کوواریانس، داده‌های میانگین روزانه تندی

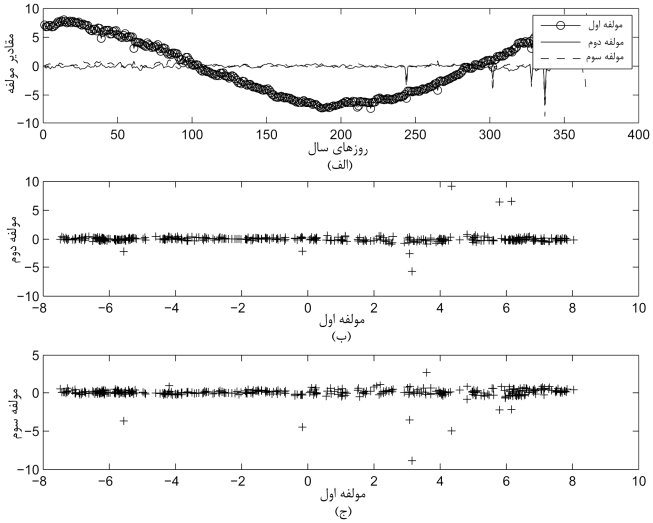


شکل ۱. نتایج آزمون اسکری با کاربست **PCA** برای مقادیر میانگین روزانه (الف) رطوبت نسبی (ب) فشار بخار آب (ج) فشار (د) بارش (ه) تندی باد و (و) دما در دوره آماری ۲۹ ساله (۱۹۸۰ تا ۲۰۰۸) برای ایستگاه منتخب.

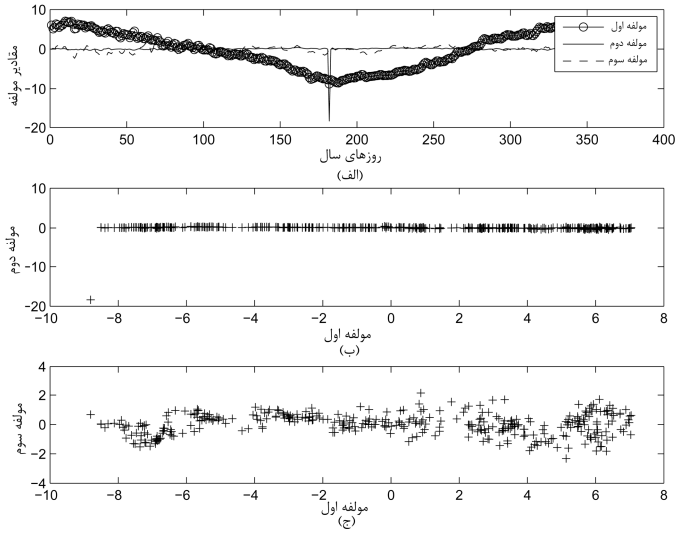
باد، میانگین روزانه بارش، میانگین رطوبت نسبی، میانگین روزانه فشار، میانگین روزانه فشار بخار آب و میانگین وزانه دما در دوره آماری ۱۹۸۰ تا ۲۰۰۸ ایستگاه همدیدی شهرستان یزد مورد بررسی قرار گرفته است. در روش PCA به‌جای استفاده از حجم زیادی از داده‌های خام، مولفه‌هایی که مقادیر ویژه بزرگ‌تری دارند برای توصیف روند کلی در هریک از پراسنج‌ها مورد توجه قرار می‌گیرند و داده‌های غیرمعمول که روند تغییرات را از مسیر اصلی منحرف می‌کنند به مولفه‌هایی با واریانس کمتر برده می‌شوند. به‌این‌ترتیب اهمیت هریک از داده‌ها در فرایند وزن‌دهی آموزش برای ANN مشخص می‌شود.

شکل ۱ نتایج آزمون اسکری را برای پراسنج‌های الف- رطوبت نسبی، ب- فشار بخار آب، ج- فشار، د- بارش، ه- تندی باد و و-دما نشان می‌دهد. در این شکل، محور افقی تعداد مولفه‌ها و محور قائم مقادیر ویژه ماتریس کوواریانس برای هریک از پراسنج‌ها است. پراسنج‌هایی که در نمودار آنها شکستگی بین مقدار ویژه مُد پیشرو و سایر مولفه‌ها بزرگ باشد، در این دوره آماری مُد غالب دارند که به‌کمک مُد پیشرو مربوط تعیین می‌شود. هرچه این تفاوت بزرگ‌تر باشد مُد غالب قوی‌تر است و تغییرات عمده سالانه برای پراسنجی با این ویژگی، در طول دوره آماری تعیین شده، پراکندگی کمتری دارد. نتایج حاصل از PCA نشان می‌دهد که در این دوره آماری، در شهرستان یزد، پراسنج‌های میانگین روزانه؛ رطوبت نسبی، فشار بخار آب، فشار و دما دارای مُد غالب هستند (شکل‌های ۱- الف، ۱- ب، ۱- ج و ۱- و). مقادیر میانگین روزانه پراسنج بارش در این دوره آماری، به علت اقلیم گرم و خشک شهرستان یزد و کمبود بارندگی، تغییرات قابل‌ملاحظه‌ای ندارد، بنابراین بین مقدار ویژه مُد پیشرو و سایر مولفه‌ها در نمودار شکستگی ایجاد نمی‌شود (شکل۱- د). در مورد پراسنج میانگین روزانه تندی باد، مُد پیشرو به‌علت تغییرات پراکنده تندی باد در روزهای سال و عوامل دینامیکی حاکم بر جوّ در این شهرستان، درطی این دوره آماری مُد غالب نیست (شکل ۱- ه).

جدول ۱ مقادیر ویژه و واریانس تجمعی شش پراسنج مورد بررسی را نشان می‌دهد. معیار مقادیر ویژه و مقدار واریانس تجمعی برای هریک از پراسنج‌ها، مولفه‌هایی با



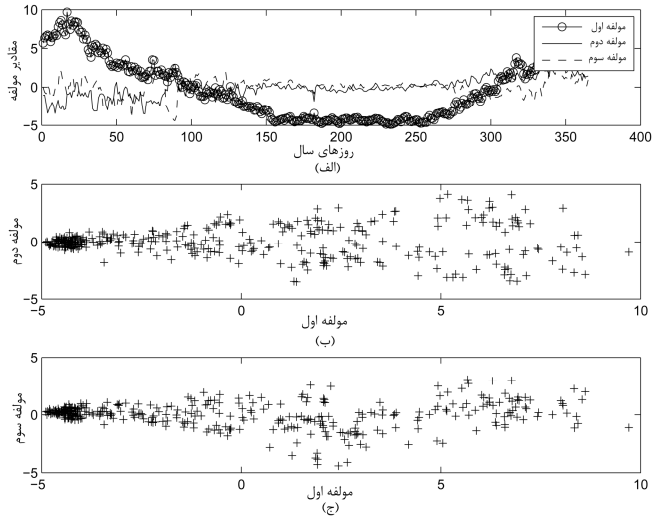
شکل ۲. مدهای تغییرپذیری میانگین روزانه دما (الف) مولفه‌های اول، دوم و سوم در روزهای سال (ب) مولفه اول برحسب مولفه دوم و (ج) مولفه اول برحسب مولفه سوم در دوره آماری ۲۹ ساله (۱۹۸۰ تا ۲۰۰۸) برای ایستگاه منتخب.



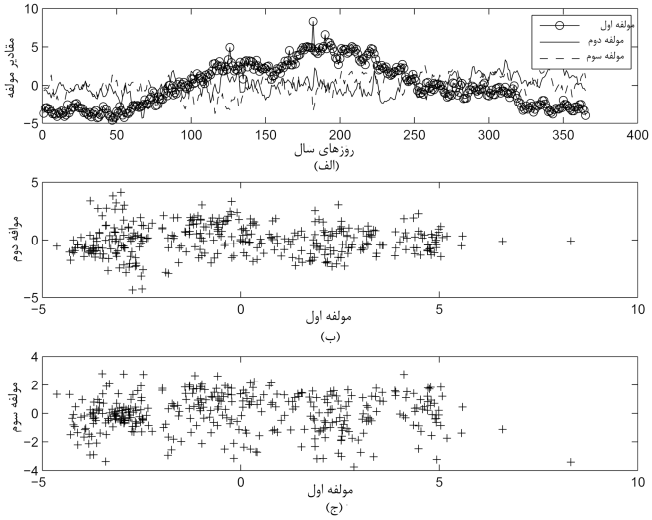
شکل ۳. مانند شکل ۲ ولی برای میانگین روزانه فشار.

مقدار ویژه بزرگ‌تر از یک را نگه می‌دارند و درصد تجمعی هریک از مولفه‌های باقی‌مانده نشان‌دهنده توانایی آن مولفه در نمایش تغییرات پراسنج در این دوره آماری است. نتایج جدول ۱ نشان می‌دهد که هشت مولفه اول فشار بخار آب، سه مولفه اول رطوبت نسبی، سیزده مولفه اول بارش، اولین مولفه فشار و دما و یازده مولفه اول تندی باد، شرایط معیارهای ذکر شده را دارند. هرچه تعداد این مولفه‌ها کمتر باشد، مُد پیشرو به مُد غالب نزدیک‌تر و پراکندگی داده‌ها کمتر است.

در شکل‌های ۲ تا ۷، نمودارهای قسمت الف، مقادیر سه مولفه اول را در روزهای سال، به‌منظور تعیین روند تغییرات، بررسی مُد پیشرو و در صورت وجود، مُد غالب نشان می‌دهد. مولفه اول بیشترین تغییرات را در این دوره



شکل ۴. مانند شکل ۲ ولی برای میانگین روزانه رطوبت نسبی.

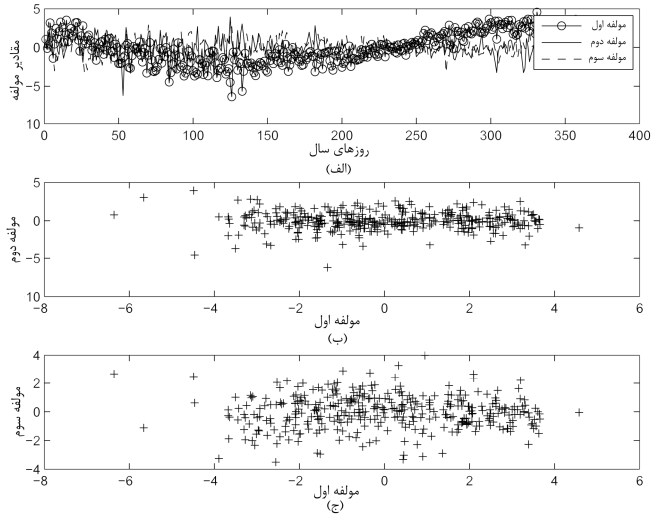


شکل ۵. مانند شکل ۲ ولی برای میانگین روزانه فشار بخار آب.

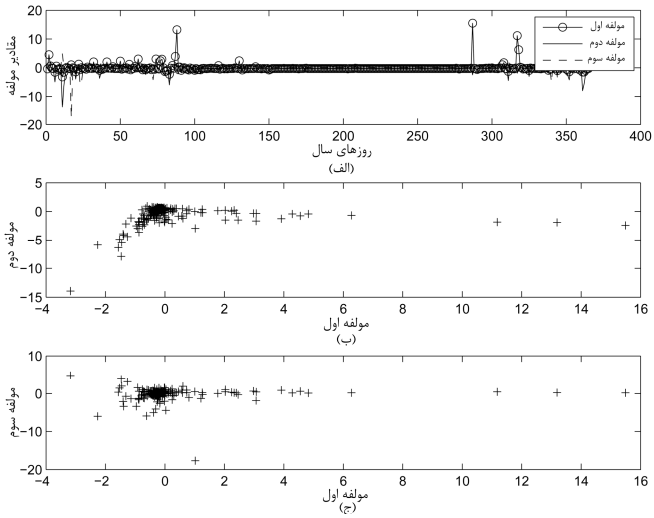
آماری تعیین می‌کند. بنابر ساخت ریاضی PCA هیچ‌یک از مولفه‌ها با یکدیگر همبستگی ندارند. بنابراین مولفه دوم، آن دسته از تغییرات جزئی‌تر که مُد پیشرو در نظر نگرفته است را نشان می‌دهد. همانند نتایج آزمون اسکری (شکل ۱)، مولفه‌های اول پراسنج‌های میانگین روزانه؛ دما، فشار، رطوبت نسبی و فشار بخار آب، از مولفه‌های دوم و سوم به‌وضوح متمایز است (شکل‌های ۲- الف، ۳- الف، ۴- الف و ۵ الف).

در نمودارهای قسمت ب و ج شکل‌های ۲ تا ۷ به‌منظور شناسایی داده‌های غیرمعمول، مولفه اول به‌ترتیب برحسب مولفه‌های دوم و سوم رسم شده است. با رسم این نمودارها داده‌های نامأنوس، که به علت‌های گوناگون وارد مجموعه داده‌های ورودی شده‌اند، مشخص می‌شوند. در هریک از نمودارها، اکثر داده‌ها در یک ناحیه متمرکز شده‌اند.

داده‌هایی که فاصله بیشتری نسبت به سایر داده‌ها دارند



شکل ۶. مانند شکل ۲ ولی برای میانگین روزانه تندی باد.



شکل۷. مانند شکل ۲ ولی برای میانگین روزانه بارش.

درحکم داده نامأنوس تلقی می‌شوند. این داده‌ها، به‌ویژه برای پراسنج‌هایی که مُد غالب دارند، ممکن است به سبب خطا در ثبت داده‌ها به علت‌های گوناگون و یا قرار گرفتن برخی از روزهای سال در شرایط جوّی خاص یا دوره‌های تغییر اقلیم، ایجاد شده باشند. اگر تعداد داده‌های نامأنوس زیاد باشد، روند آموزش ANN را تحت‌تاثیر قرار می‌دهد و باعث کاهش دقت و عمومیت آن می‌شود. در پژوهش حاضر، تعداد داده‌های نامأنوس اندک است و داده‌ها با میانگین‌گیری از روزهای قبل و بعد، جایگزین شده‌اند. جدول ۲، واریانس تجمعی و ضرایب همبستگی مقادیر ویژه مُد پیشرو هریک از پراسنج‌ها را با هدف ANN نشان می‌دهد. در پژوهش حاضر، هدف ANN، میانگین روزانه دما است. مقدار واریانس تجمعی مُد پیشرو دما ۸۲ درصد از پراکندگی دما را در یک دوره ۲۹ ساله نمایش می‌دهد و مُد غالب دارد.

بنابراین برای پیش‌بینی دما، انتخاب سری زمانی پراسنج‌هایی که دارای مُد غالب و ضریب همبستگی زیاد با هدف ANN هستند، مناسب به‌نظر می‌رسد. توجه به جدول ۲، نشان می‌دهد که مُد پیشرو دما و فشار با واریانس تجمعی به‌ترتیب ۸۲ و ۷۷ درصد، نماینده دوره آماری این دو پراسنج هستند. همچنین ضریب همبستگی بزرگی با هدف دارند (به‌ترتیب ۹۵ و ۹۳ درصد). واریانس تجمعی مُد پیشرو رطوبت نسبی و فشار بخار آب، مقادیر کوچک‌تری نسبت به دو پراسنج قبلی دارند که این نشان می‌دهد داده‌های خام این دو پراسنج پراکنده‌تر هستند. بااین‌حال مولفه اول آنها که به‌ترتیب نماینده ۵۵ و ۲۸ درصد از تغییرات دوره آماری این پراسنج‌ها است، ضریب همبستگی به‌نسبت بزرگی با هدف دارد (به‌ترتیب ۹۲ و ۸۶ درصد). مولفه اول پراسنج تندی باد به علت پراکندگی شدید داده‌ها در طی روزهای سال و در دوره آماری پیش‌گفته، قادر به نمایش ۱۴ درصد از روند تغییرات و جزئیات پراکندگی این پراسنج در سال است. با توجه به اینکه هدف ANN مُد غالب دارد، ضریب همبستگی آن با پراسنجی که مُد غالب ندارد، می‌باید مقدار کوچکی باشد. با توجه به جدول ۲، ۴۲ درصد همبستگی بین مولفه اول تندی باد و هدف ANN، این موضوع را تایید می‌کند. مقادیر داده‌های خام بارش به

جدول ۱. مقادیر ویژه و واریانس تجمعی پراسنج‌های تندی باد، دما، فشار، بارش، رطوبت نسبی و فشار بخار آب به‌منظور ارزیابی معیارهای ارزش ویژه و واریانس تجمعی.

شماره مولفه	تندی باد		دما		فشار		بارش		رطوبت نسبی		فشار بخار آب	
	مقدار ویژه	واریانس تجمعی	مقدار ویژه	واریانس تجمعی	مقدار ویژه	واریانس تجمعی	مقدار ویژه	واریانس تجمعی	مقدار ویژه	واریانس تجمعی	مقدار ویژه	واریانس تجمعی
۱	۴/۱۳۸۹	۰/۱۴۲۷	۲۴/۰۴۰۶	۰/۸۲۸۹	۲۲/۳۶۴۷	۰/۷۷۱۱	۲/۱۷۲۷	۰/۰۷۴۹	۱۶/۶۵۰۰	۰/۵۵۵۳	۸/۰۸۹۶	۰/۲۸۸۹
۲	۱/۴۶۹۵	۰/۰۵۰۶	۰/۶۶۸۴	۰/۰۲۳۰	۰/۹۵۳۸	۰/۰۳۲۸	۱/۸۴۲۰	۰/۰۶۳۵	۱/۸۳۰۰	۰/۰۶۱۰	۱/۷۶۱۴	۰/۰۶۲۹
۳	۱/۳۳۳۰	۰/۰۴۵۹	۰/۵۴۳۶	۰/۰۱۸۷	۰/۵۲۸۴	۰/۰۱۸۲	۱/۷۱۹۲	۰/۰۵۹۲	۱/۲۲۹۲	۰/۰۴۰۹	۱/۶۷۹۳	۰/۰۵۹۹
۴	۱/۲۸۴۹	۰/۰۴۴۳	۰/۳۴۹۱	۰/۰۱۲۰	۰/۴۱۳۴	۰/۰۱۴۲	۱/۵۹۰۶	۰/۰۴۵۶	۰/۹۶۷۴	۰/۰۳۲۲	۱/۳۰۲۴	۰/۰۴۶۵
۵	۱/۲۴۲۸	۰/۰۴۲۸	۰/۳۲۳۷	۰/۰۱۱۱	۰/۳۸۶۰	۰/۰۱۳۳	۱/۵۵۷۹	۰/۰۱۰۷	۰/۸۴۹۴	۰/۰۲۸۳	۱/۲۵۱۳	۰/۰۴۴۶
۶	۱/۱۸۹۱	۰/۰۱۵۲	۰/۳۰۷۸	۰/۰۱۰۶	۰/۳۳۴۳	۰/۰۱۱۵	۱/۴۵۹۹	۰/۰۱۲۷	۰/۷۸۱۲	۰/۰۲۶۰	۱/۱۱۱۲	۰/۰۳۹۶
۷	۱/۱۴۱۵	۰/۰۴۱۰	۰/۲۹۱۲	۰/۰۱۰۰	۰/۳۰۹۸	۰/۰۱۰۶	۱/۳۳۸۶	۰/۰۵۴۸	۰/۶۶۲۷	۰/۰۲۲۰	۱/۰۶۸۸	۰/۰۳۸۱
۸	۱/۱۰۶۵	۰/۰۳۹۳	۰/۲۵۸۴	۰/۰۰۸۹	۰/۲۹۱۱	۰/۰۱۰۰	۱/۲۸۴۴	۰/۰۵۳۷	۰/۶۲۲۳	۰/۰۲۰۷	۱/۰۲۴۵	۰/۰۳۶۵
۹	۱/۰۶۸۵	۰/۰۱۷۶	۰/۲۵۲۵	۰/۰۰۸۷	۰/۲۷۵۰	۰/۰۰۹۴	۱/۲۴۹۰	۰/۰۵۰۳	۰/۶۰۴۷	۰/۰۲۰۱	۰/۹۵۰۳	۰/۰۳۳۹
۱۰	۱/۰۶۳۰	۰/۰۱۸۱	۰/۲۲۴۸	۰/۰۰۷۷	۰/۲۶۰۱	۰/۰۰۸۹	۱/۱۲۶۵	۰/۰۱۷۰	۰/۵۶۵۶	۰/۰۱۸۸	۰/۸۶۸۶	۰/۰۳۱۰
۱۱	۱/۰۲۸۵	۰/۰۳۸۱	۰/۱۹۶۴	۰/۰۰۶۷	۰/۲۵۸۰	۰/۰۰۸۸	۱/۱۱۷۴	۰/۰۱۹۱	۰/۴۸۵۷	۰/۰۱۶۱	۰/۸۵۰۵	۰/۰۳۰۳
۱۲	۰/۹۷۱۲	۰/۰۳۶۶	۰/۱۸۴۰	۰/۰۰۶۳	۰/۲۳۶۴	۰/۰۰۲۶	۱/۰۵۱۷	۰/۰۱۹۵	۰/۴۵۱۵	۰/۰۱۵۰	۰/۷۲۲۵	۰/۰۰۸۳
۱۳	۰/۹۴۶۸	۰/۰۳۶۸	۰/۱۷۲۶	۰/۰۰۵۹	۰/۲۲۴۲	۰/۰۰۲۵	۱/۰۳۷۹	۰/۰۴۶۱	۰/۴۲۶۱	۰/۰۱۴۲	۰/۶۹۴۵	۰/۰۰۹۴
۱۴	۰/۹۳۸۱	۰/۰۳۵۴	۰/۱۲۹۹	۰/۰۰۴۴	۰/۲۱۷۱	۰/۰۰۸۱	۰/۹۹۶۳	۰/۰۴۴۲	۰/۴۰۹۳	۰/۰۱۳۶	۰/۶۴۵۳	۰/۰۲۵۸
۱۵	۰/۹۰۱۸	۰/۰۱۹۳	۰/۱۱۶۹	۰/۰۰۴۰	۰/۱۹۵۸	۰/۰۰۷۷	۰/۹۴۱۰	۰/۰۴۳۰	۰/۳۸۲۵	۰/۰۱۲۷	۰/۶۳۸۱	۰/۰۲۴۸
۱۶	۰/۸۲۸۹	۰/۰۲۰۰	۰/۱۰۴۱	۰/۰۰۱۰	۰/۱۸۰۷	۰/۰۰۷۴	۰/۹۳۵۱	۰/۰۲۱۸	۰/۳۳۴۳	۰/۰۰۳۳	۰/۵۸۹۲	۰/۰۱۱۱
۱۷	۰/۸۱۴۸	۰/۰۳۳۴	۰/۰۹۱۵	۰/۰۰۳۵	۰/۱۷۷۳	۰/۰۰۳۰	۰/۸۷۰۷	۰/۰۲۳۰	۰/۳۱۹۴	۰/۰۱۱۱	۰/۵۶۶۱	۰/۰۱۱۸
۱۸	۰/۷۷۱۶	۰/۰۳۲۶	۰/۰۸۶۹	۰/۰۰۱۴	۰/۱۵۹۰	۰/۰۰۶۷	۰/۸۴۴۸	۰/۰۲۴۷	۰/۲۸۲۵	۰/۰۱۰۶	۰/۵۴۹۸	۰/۰۱۲۷
۱۹	۰/۷۴۰۹	۰/۰۳۲۳	۰/۰۸۲۱	۰/۰۰۱۷	۰/۱۴۶۹	۰/۰۰۶۲	۰/۷۷۶۶	۰/۰۲۶۷	۰/۲۷۱۴	۰/۰۰۴۶	۰/۴۸۲۸	۰/۰۱۳۹
۲۰	۰/۷۲۷۶	۰/۰۳۱۰	۰/۰۷۶۲	۰/۰۰۱۶	۰/۱۳۷۵	۰/۰۰۶۱	۰/۷۶۹۸	۰/۰۲۶۵	۰/۲۶۴۶	۰/۰۰۴۸	۰/۴۵۴۹	۰/۰۱۳۷
۲۱	۰/۶۹۷۳	۰/۰۲۱۷	۰/۰۷۳۳	۰/۰۰۱۹	۰/۱۳۵۸	۰/۰۰۳۴	۰/۷۱۸۰	۰/۰۲۹۱	۰/۲۵۳۵	۰/۰۰۵۵	۰/۴۲۴۵	۰/۰۱۵۱
۲۲	۰/۶۷۱۵	۰/۰۲۸۵	۰/۰۶۸۹	۰/۰۰۳۱	۰/۱۲۸۷	۰/۰۰۵۴	۰/۶۶۹۶	۰/۰۳۰۰	۰/۲۱۸۲	۰/۰۰۵۹	۰/۳۹۰۱	۰/۰۱۶۲
۲۳	۰/۶۶۸۵	۰/۰۲۸۰	۰/۰۶۷۱	۰/۰۰۲۹	۰/۱۲۴۹	۰/۰۰۳۷	۰/۶۳۴۸	۰/۰۳۸۵	۰/۲۰۲۶	۰/۰۰۶۴	۰/۳۸۵۵	۰/۰۱۷۲
۲۴	۰/۶۳۰۶	۰/۰۲۳۰	۰/۰۶۱۶	۰/۰۰۲۱	۰/۱۱۲۲	۰/۰۰۳۸	۰/۵۶۸۲	۰/۰۳۸۸	۰/۱۹۴۳	۰/۰۰۶۷	۰/۳۵۶۳	۰/۰۲۳۰
۲۵	۰/۵۸۲۴	۰/۰۲۳۱	۰/۰۵۶۳	۰/۰۰۲۸	۰/۱۰۸۶	۰/۰۰۵۰	۰/۵۵۴۳	۰/۰۳۲۲	۰/۱۷۷۴	۰/۰۰۷۲	۰/۳۳۱۳	۰/۰۲۲۷
۲۶	۰/۵۶۱۳	۰/۰۲۶۶	۰/۰۵۰۳	۰/۰۰۲۶	۰/۰۹۸۵	۰/۰۰۴۳	۰/۴۹۳۲	۰/۰۳۲۴	۰/۱۶۶۲	۰/۰۰۹۴	۰/۳۱۲۷	۰/۰۲۱۰
۲۷	۰/۵۲۶۷	۰/۰۲۴۰	۰/۰۴۸۸	۰/۰۰۲۵	۰/۰۸۹۷	۰/۰۰۴۴	۰/۳۶۸۳	۰/۰۳۴۳	۰/۱۴۵۳	۰/۰۰۸۴	۰/۲۶۴۶	۰/۰۱۹۶
۲۸	۰/۵۱۰۴	۰/۰۲۵۰	۰/۰۴۰۷	۰/۰۰۲۳	۰/۰۷۵۶	۰/۰۰۴۶	۰/۳۱۰۴	۰/۰۳۶۲	۰/۱۴۰۸	۰/۰۰۹۰	۰/۲۳۲۵	۰/۰۲۰۲
۲۹	۰/۴۹۲۴	۰/۰۲۴۶	۰/۰۳۳۲	۰/۰۰۲۰	۰/۰۷۰۱	۰/۰۰۴۵	۰/۲۹۸۲	۰/۰۳۵۵	۰/۱۳۸۳	۰/۰۰۷۱	۰/۲۱۷۸	۰/۰۱۶۸

علت اقلیم گرم و خشک شهرستان یزد، تقریباً در اکثر روزهای سال برابر صفر است در نتیجه طبیعی است که مولفه اول آن در یک دوره ۲۹ ساله نیز واریانس تجمعی نزدیک به صفر و ضریب همبستگی آن با هدف شبکه، کوچک باشد.

۴ نتیجه‌گیری

همان‌طور که پیش‌تر بیان شد، در گام نخست انتخاب ورودی ANN دو رهیافت کلی را می‌توان دنبال کرد؛ رهیافت اول- استفاده از مقادیر سری زمانی هدف موردنظر در سال‌های گذشته و رهیافت دوم- یافتن و به‌کارگیری

پراسنج‌هایی که با هدف، رابطه خطی یا غیرخطی دارند. در پژوهش حاضر، با گزینش رهیافت دوم و مجموعه داده‌هایی از پراسنج‌های مرتبط با هدف ANN (دمای سال ۲۰۰۹)، در یک دوره آماری ۲۹ ساله (۱۹۸۰ تا ۲۰۰۸) و پس از کاربست PCA، جزئیات داده‌های منتخب مورد بررسی قرار گرفت. در نگاه اول ممکن است، انتخاب داده‌های دما در سال‌های قبل از ۲۰۰۹ در این دوره آماری، با داشتن بیشترین مقدار ضریب همبستگی، در پیش‌بینی هدف، بهترین گزینه به‌نظر برسد اما در بررسی جزئیات روند تغییرات که به علت فرایندهای دینامیکی و ترمودینامیکی ایجاد می‌شوند، این داده‌ها دقت کافی ندارند؛ مگر اینکه در کنار مولفه‌های اصلی سایر پراسنج‌هایی که مُد پیشرو آنها ضریب همبستگی بزرگی با دما دارد، قرار بگیرند. بنابراین علاوه بر داده‌های دما، داده‌های فشار، رطوبت نسبی و فشار بخار آب در سال‌های ۱۹۸۰ تا ۲۰۰۸ برای ورودی مناسب هستند.

پراسنج‌های باد و بارش ضریب همبستگی کوچکی با هدف ANN دارند اما به علت رابطه غیرخطی این پراسنج‌ها با دما نمی‌توان از آنها صرف‌نظر کرد (هلتون، ۲۰۰۴؛ پدلاسکی، ۱۹۸۶). این پراسنج‌ها به همراه سایر پراسنج‌های هواشناختی در فرایندهای دینامیکی و ترمودینامیکی جوّی با دما ارتباط پیدا می‌کنند و در یک‌سال بر تغییرات روزانه، ماهانه و فصلی آن اثر می‌گذارند. مقادیر بارش در یک‌سال مفهوم فصل تر و خشک را برای ANN تفهیم می‌نماید. همچنین تغییرات در مقادیر تندی باد، دینامیک حاکم بر جوّ را در این منطقه بدون توجه به معادله‌های پیچیده برای ANN تعریف می‌کند تا ANN طی فرایند آموزشی، تابع بین این پراسنج‌ها را با دما تقریب بزند. ازطرفی استفاده از داده‌های خام یک دوره آماری دو پراسنج اخیر در فرایند آموزشی، می‌تواند ANN را از مسیر خود در پیش‌بینی هدف منحرف کند. همچنین با توجه به نتایج حاصل از PCA، پراسنج‌های بارش و تندی باد، مُد غالب ندارند. بنابراین استفاده از یک دوره آماری آنها برای پیش‌بینی روزانه دما در یک‌سال صحیح به‌نظر نمی‌رسد. برای رفع این مشکل می‌توان از مقادیر این پراسنج‌ها در سال ۲۰۰۹ استفاده کرد که به‌کمک رهیافت اول در گام نخست انتخاب ورودی با دقت قابل‌قبولی امکان‌پذیر است. برای پراسنج بارش که مقادیر کوچکی در طی سال دارد، با جای‌گذاری مقادیر صفر و یک، می‌توان روزهای خشک و تر را تعیین و این متغیر جدید را به همراه مقادیر تندی باد در سال ۲۰۰۹ برای ورودی ANN انتخاب کرد.

ازآنجاکه کاربست PCA سبب کاهش ابعاد داده، استخراج مُدهای تغییرپذیری متغیرهای ورودی، از بین بردن همبستگی بین داده‌های خام و حذف داده‌های غیرمعمول می‌شود، بنابراین انتظار می‌رود، بتوان به‌کمک PCA و بدون استفاده از روش سعی و خطا بهترین داده‌های ورودی را برای ANN انتخاب کرد.

جدول۲. مقادیر واریانس تجمعی و ضریب همبستگی مُد پیشرو پراسنج‌ها با هدف شبکه عصبی.

پراسنج / معیار ارزیابی	فشار بخار آب	رطوبت نسبی	بارش	فشار	دما	تندی باد
واریانس تجمعی	۰/۲۸	۰/۵۵	۰/۰۷	۰/۷۷	۰/۸۲	۰/۱۴
ضریب همبستگی	۰/۸۶	-۰/۹۲	-۰/۰۷	۰/۹۳	۰/۹۵	-۰/۴۲

منابع

آزاده، آ.، شیخ‌علیشاهی، م.، تابش، م.، و نگهبان، الف.، ۱۳۹۰، تاثیر روش‌های پیش‌پردازش بر بهبود عمومیت شبکه‌های عصبی مصنوعی: مجله علوم پایه و کاربردی، **۶**، ۵۷۰-۵۸۰.

پوراصغر، ف.، جهانبخش، س.، ساری صراف، ب.، قائمی، ه.، و تدینی، م.، ۱۳۹۲، پهنه‌بندی رژیم بارش در نیمه جنوبی ایران: نشریه جغرافیا و برنامه‌ریزی، **۴۴**، ۲۷-۴۶.

حیاتی، م.، و محبی، ز.، ۱۳۸۶، پیش‌بینی دما براساس رهیافت شبکه عصبی: مجله علوم کاربردی جهانی، **۲**(۶)، ۶۱۳-۶۲۰.

رحیمی‌خوب، ع.، بهبهانی، م. ر.، و نظری‌فر، م.، ۱۳۸۶، پیش‌بینی دمای هوای استان خوزستان براساس داده‌های ماهواره نوا و مدل شبکه عصبی: مجله علوم و فنون کشاورزی و منابع طبیعی، **۴۲**، ۳۵۷-۳۶۴.

فتاحی، ا.، و بهیار، م. ب.، ۱۳۹۰، بررسی الگوهای سینوپتیکی خشکسالی‌های فراگیر در استان چهارمحال و بختیاری: فصلنامه تحقیقات جغرافیایی، **۲**، ۱۷۰۷۱-۱۷۰۹۳.

قبائی سوق، م.، مساعدی، الف.، حسام، م.، و هزارجریبی، الف.، ، ۱۳۸۹، ارزیابی تاثیر پیش‌پردازش پارامترهای ورودی به شبکه عصبی مصنوعی (ANN) با استفاده از روش‌های رگرسیون گام به گام و گاماتست به‌منظور تخمین سریع‌تر تبخیر و تعرق روزانه: نشریه آب و خاک، **۳**، ۶۱۰-۶۲۴.

قنواتی، ع.، ضیائیان، پ.، سردشتی، م.، و جنگی، ع. الف.، ۱۳۸۶، آشکارسازی تغییرات مورفودینامیک با استفاده از داده‌های سنجش از دور و تحلیل مولفه‌های اصلی (PCA) و منطق فازی (Fuzzy-Logic)، مطالعه موردی، حوضه آبخیز طالقان: پژوهش‌های جغرافیایی، **۶۲**، ۴۱-۵۳.

کاویانی، م.، میرکنی، س. م.، و معماریان، م. ح.، ۱۳۹۲، تأثیر روش‌های پیش‌پردازش داده‌های ورودی بر خروجی شبکه عصبی مصنوعی: ارائه شده در اولین کنفرانس فیزیک محاسباتی ایران، ۳۰ دی – ۲ بهمن، دانشگاه تربیت دبیر شهید رجایی، تهران.

محدث، ف.، ۱۳۸۹، روش‌های تحلیل مولفه‌های اساسی و بررسی عوامل، مطالعه موردی: استخراج شاخص قیمت دارایی‌ها، مجموعه پژوهش‌های اقتصادی، ۲-۷.

مشاری، ا.، ابراهیمی، الف.، صدری، س.، و ابراهیمی، م.، ، ۱۳۸۸، پالایش داده‌های آموزشی شبکه عصبی و بررسی تاثیر آن در کاهش خطای پیش‌بینی کوتاه‌مدت بار سیستم‌های قدرت: روش‌های عددی در مهندسی، **۲**، ۶۷-۷۹.

موسوی، س. ح.، آبختی گروسی، ه.، و خائفی، ن.، ۱۳۹۱، رفتارشناسی اقلیمی بر مبنای تحلیل مولفه‌های اصلی، مطالعه موردی، استان‌های کردستان و کرمانشاه: فصلنامه علمی-پژوهشی فضای جغرافیایی، سال دوازدهم، **۳۹**، ۱۵۷-۱۷۱.

نوری قیداری، م. ح.، ۱۳۸۹، شناسایی داده‌های پرت در آنالیز منطقه‌ای سیلاب به روش آنالیز مولفه‌های اصلی: پنجمین کنگره ملی مهندسی عمران، دانشگاه فردوسی مشهد، ایران.

همت پور، س. و هاشمی، ح.، ۱۳۹۰، استفاده از تکنیک‌های کاهش بعد تحلیل مولفه‌های اصلی (PCA) و تحلیل جداکننده‌های پارامتری‌شده (RDA) در رتبه‌بندی نشانگرهای لرزه‌ای: مجله فیزیک زمین و فضا، **۳۷**(۴)، ۲۱۷-۲۲۷.

Back, A. D., and Trappenberg, T. P., 2001, Selecting inputs for modeling using normalized higher order statistics and independent component analysis: IEEE Transaction on Neural Networks, **12**, 612-617.

Castellano, G., and Fanelli, A. M., 2000, Variable selection using neural-network models: Neurocomputing, **31**, 1-13.

component analysis on matlab: International Journal of The Computer, **16**, 1-8.

Kramer, M. A., 1991, Nonlinear principal component analysis using auto associative neural networks: AIChE Journal., **37**, 223-243.

Lamarche, V. C., Jr, and Fritts, H. C., 1971, Anomaly patterns of climate over the Western United States**,** 1700-1930, derived from prinicipal component analysis of tree-ring data: Mon. Wea. Rev., **99.2**, 138-142.

Mohandes, M. A., Rehman, S., and Halawani, T. O., 1998, Neural networks approach for wind speed prediction: Renewable Energy, **13**, 345-354.

Mehrotra, K., Mohan, C. K., and Ranka, S., 1997, Elements of Artificial Neural Network: MIT Press.

Munyati, C., 2004, Use of principal component analysis (PCA) of remote sensing images in wetland change detection on the Kafue flats, Zambia: Geocarto International, **19**, 11-22.

Pasini, A., Pelino, V., and Potesta, S., 2001, A neural network model for visibility nowcasting from surface observations: Results and sensitivity to physical input variables: J. Geophys. Res., **106**, 14951-14959.

Pedlosky, J., 1986, Geophysical Fluid Dynamics: Springer.

Santamouris, G. M., and Asimakopoulos, D., 2012, Modeling ambient air temperature time series using neural network: Journal of Geophysical Research:Atmospheres, **103**, 19509-19517.

Tresp, V., Neuneier, R., and Ahmad, S., 1995, Efficient Methods for Dealing With Missing Data, in Supervised Learning: MIT Press**.**

Chen, T., Zhang, C., Chen, X., and LI, L.Q., 2009, An input variable selection method for the artificial neural network of shear stiffness of worsted fabrics: Statical Analysis and Data Mining, **1.5**, 287-295.

Churnside, J. H., Stermitz, T. A., and Schroeder, J. A., 1994, Temperature profiling with neural network inversion of microwave radiometer data: Atmos. and Oceanic Tech., **11**, 105-109.

Erdil, A., and Arcaklioglu, E., 2013, The prediction of meteorological variables using artificial neural network: Neural Comput. & Applic., **22**, 1677-1683.

Fabian, D., De Dear, R., and Lellyett, S., 2007, Application of artificial neural network forecasting to predict fog at Canberra International Airport: Wea. Forecasting, **22**, 372-381.

Fang, Y., Omitaomu, O. A., and Ganguly, A. R., 2010, Incremental Anomaly Detection Approach: Springer.

Hassoun, M. H., 1995, Fundamentals of Artificial Neural Networks: Cambridge Univ. Press.

Gardner, M. W., and Dorling, S. R., 1998, Artifical neural networks (the multilayer perceptron)-A review of applications in the atmospheric science: Atmos. Environ., **32**, 2627-2636.

Graupe, D., 2007, Principles of Artificial Neural Networks: World Scientifics Publishing.

Hollton, J. R., 2004, An Introduction to Dynamics Meteorology: 4th ed., Academic Press.

Ilin, A., and Raiko, T., 2010, Practical approaches to principal component analysis in the presence of missing values: Journal of Machine Learning Research, **11**, 1957-2000.

Junita, M. S., and Hoyle, B. S., 2008, Improved neural network performance using principal