

## پیش بینی ماهانه جریان با استفاده از ماشین بردار پشتیبان بر مبنای آنالیز مؤلفه اصلی

اشکان فرخ‌نیا<sup>۱</sup>

مجید دهقانی<sup>۲</sup>

امیر خاکپور<sup>۲</sup>

روح‌اله نوری<sup>۱</sup>

(دریافت ۸۸/۷/۲۹ پذیرش ۸۹/۵/۲۰)

### چکیده

هدف اصلی این تحقیق بررسی تأثیر انتخاب متغیرهای ورودی با استفاده از آنالیز مؤلفه اصلی (PCA) بر عملکرد مدل ماشین بردار پشتیبان (SVM) برای پیش‌بینی ماهانه دبی رودخانه بود. به این منظور ابتدا با استفاده از ۱۸ متغیر ورودی به مدل SVM، دبی جریان ماهانه پیش‌بینی شد. سپس با استفاده از PCA تعداد متغیرهای ورودی به مدل SVM از ۱۸ متغیر به ۵ مؤلفه کاهش یافت. در نهایت با استفاده از آماره توسعه یافته توسط نویسندگان مقاله، عملکرد مدل‌های ارائه شده (SVM و PCA-SVM) مورد ارزیابی قرار گرفت. یافته‌های این تحقیق نشان داد که پیش‌پردازش متغیرهای ورودی به مدل SVM با استفاده از PCA، بهبود عملکرد مدل SVM را به همراه داشته است.

**واژه‌های کلیدی:** ماشین بردار پشتیبان، آنالیز مؤلفه اصلی، جریان ماهانه

## Monthly Stream Flow Prediction Using Support Vector Machine Based on Principal Component Analysis

Roohollah Noori<sup>1</sup>

Amir Khakpour<sup>2</sup>

Majid Dehghani<sup>3</sup>

Ashkan Farokhnia<sup>4</sup>

(Received Oct. 21, 2009 Accepted Aug. 11, 2010)

### Abstract

The main goal of this research is to evaluate the role of input selection by Principal Component Analysis (PCA) on Support Vector Machine (SVM) performance for monthly stream flow prediction. For this purpose, SVM is used to predict monthly flow as a function of 18 input variables. PCA is subsequently employed to reduce the number of input variables from 18 to 5 PCs which are finally fed into the SVM model. SVM and PCA-SVM models are evaluated in terms of their performance using a developed statistic by the authors. Findings show that preprocessing of input variables by PCA improved SVM performance.

**Keywords:** Support Vector Machine, Principal Component Analysis, Monthly Stream Flow.

1. Member of Water Research Institute, Ministry of Energy, Ph.D Student of Environmental Eng., Graduate Faculty of Environment, University of Tehran (Corresponding Author) (+98 21) 7700305 roohollahnoori@gmail.com
2. Management Director of CELCO, and Ph.D Student of Environmental Eng., Graduate Faculty of Environment, University of Tehran
3. Ph.D Student of Hydraulic Eng., Dept. of Eng., Islamic Azad University, Science and Research Branch, Tehran
4. Member of Water Research Institute, Ministry of Energy, Ph.D. Student of Hydraulic Structure, College of Agriculture, Tarbiyat Modarres University, Tehran

- ۱- کارشناس مؤسسه تحقیقات آب وزارت نیرو، دانشجوی دکتری مهندسی محیط زیست، دانشکده تحصیلات تکمیلی محیط زیست، دانشگاه تهران (نویسنده مسئول) roohollahnoori@gmail.com (۰۲۱)۷۷۰۰۳۰۵
- ۲- مدیر عامل شرکت عمران زیست آزما (CELCO)، دانشجوی دکتری مهندسی محیط زیست، دانشکده تحصیلات تکمیلی محیط زیست، دانشگاه تهران
- ۳- دانشجوی دکتری مهندسی عمران آب، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران
- ۴- کارشناس مؤسسه تحقیقات آب وزارت نیرو، دانشجوی دکتری سازه‌های آبی، دانشکده کشاورزی، دانشگاه تربیت مدرس

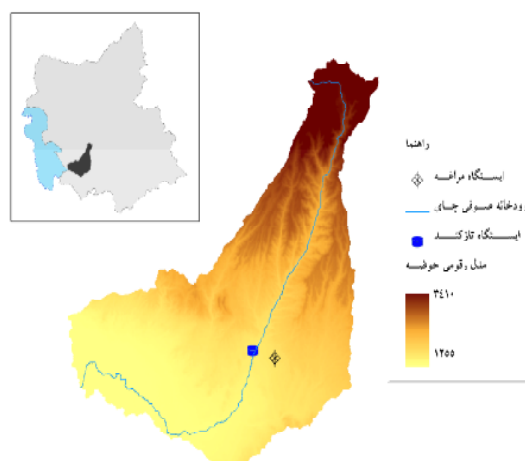
[۱۱ و ۱۲]. نوری و همکاران در تحقیقی به بررسی نقش PCA بر عملکرد شبکه عصبی مصنوعی<sup>۹</sup> برای پیش‌بینی جریان ماهانه پرداختند و بهبود عملکرد مدل ترکیبی PCA و ANN را گزارش نمودند [۱۳]. نوری و همکاران برای کاهش متغیرهای ورودی به مدل SVM در پیش‌بینی میزان زباله تولیدی در شهر مشهد از PCA استفاده نمودند و با مقایسه مدل ترکیبی PCA-SVM با مدل SVM برتری مدل PCA-SVM را نتیجه گرفتند [۱۴].

با توجه به مطالب مذکور، در این مقاله توسعه مدل SVM به منظور پیش‌بینی جریان ماهانه، انتخاب ورودی‌ها برای مدل SVM با استفاده از PCA و بررسی نقش پیش‌پردازش متغیرهای ورودی بر عملکرد مدل SVM مورد بررسی قرار گرفت.

## ۲- مواد و روشها

### ۲-۱- منطقه مورد مطالعه و داده‌ها

رودخانه صوفی‌چای از رودخانه‌های بزرگ حوزه آبریز دریاچه ارومیه است که ورودی اصلی به سد علویان را تشکیل می‌دهد. اطلاعات بارش، تابش خورشیدی و حداقل، حداکثر و میانگین دمای هوا از ایستگاه سینوپتیک مراغه و متوسط دبی رودخانه صوفی‌چای در محل ایستگاه هیدرومتری تازکند در بالادست سد علویان به مدت ۱۸ سال در مقیاس ماهانه (از اواسط سال ۱۳۶۲ تا اواخر ۱۳۸۳) در این تحقیق مورد استفاده قرار گرفت. لذا در این مطالعه، اطلاعات ذکر شده هر کدام با سه تأخیر زمانی که مجموعاً ۱۸ متغیر را تشکیل می‌دهند، به عنوان گزینه‌های ورودی مدل‌ها برای پیش‌بینی دبی یک ماه بعد مدنظر قرار گرفتند. شکل ۱ موقعیت این حوزه آبریز و ایستگاه مربوطه را در استان آذربایجان شرقی نشان می‌دهد.



شکل ۱- موقعیت منطقه مورد مطالعه و ایستگاه مورد استفاده

پیش‌بینی دبی جریان به عنوان یکی از مطرح‌ترین چالش‌های مدیریت منابع آب در دهه‌های اخیر بوده به طوری که محققان روشهای متفاوتی را برای این امر در مقالات مختلف ارائه نموده و به کار گرفته‌اند. این روشها عمدتاً شامل تکنیک‌های رگرسیونی، مدل‌های مفهومی و روشهای پیچیده‌تر بر پایه شبکه‌های عصبی مصنوعی، نروفازی و اخیراً مدل‌های ماشین بردار پشتیبان<sup>۱</sup> است [۷-۱]. در این مطالعه نقش پیش‌پردازش متغیرهای ورودی به مدل SVM به منظور پیش‌بینی ماهانه دبی جریان مورد بررسی قرار گرفت.

ساختار مدل SVM توسط وپنیک<sup>۲</sup> که یک ریاضیدان روسی در دهه ۱۹۶۰ بود، ارائه شد [۸]. مدل‌های SVM که بر پایه تئوری آموزش احتمال قرار دارند، یک کلاس جدید از مدل‌هایی هستند که به منظور طبقه‌بندی و پیش‌بینی در علوم مختلف استفاده می‌شوند. اخیراً این مدل‌ها در یک گستره وسیع از مسائل هیدرولوژیکی مورد استفاده قرار گرفته‌اند. آسفا و همکاران<sup>۳</sup> روش مناسبی برای پیش‌بینی دبی جریان فصلی و ساعتی با استفاده از مدل SVM ارائه نمودند. در تحقیق مذکور با استفاده از میزان آب معادل برف و حجم جریان در دوره‌های قبل، میزان حجم جریان برای مقیاس‌های زمانی شش ماهه و ۲۴ ساعته پیش‌بینی گردید که نتایج، رضایت‌بخش گزارش شد [۷]. یو و همکاران<sup>۴</sup> در مطالعه‌ای با استفاده از مدل SVM به پیش‌بینی سطح رقوم سیل در رودخانه لن-یانگ<sup>۵</sup> تا یوان پرداختند. به منظور کالیبره کردن پارامترهای مدل، از روش بهینه‌سازی جستجوی شبکه<sup>۶</sup> استفاده شد. نتایج نشان داد که مدل به خوبی سطح آب رودخانه را برای ۱ تا ۶ ساعت بعد پیش‌بینی می‌کند [۹]. در تحقیقی دیگر یو و لینگ<sup>۷</sup> از مدل SVM به منظور پیش‌بینی سری زمانی متغیرهای هیدرولوژیکی استفاده نمودند [۱۰].

به هر حال مشابه با هر مدل ریاضی و آماری، مدل‌های SVM نیز دارای معایبی هستند. وجود تعداد زیاد متغیر ورودی، اصلی‌ترین مشکل در توسعه این مدل‌هاست. تعداد زیاد متغیرهای ورودی ممکن است مانع یافتن مدل بهینه توسط SVM شود. راههای متعددی برای کاهش تعداد متغیرهای ورودی پیشنهاد شده است که از آن جمله می‌توان به روش آنالیز مؤلفه اصلی<sup>۸</sup> اشاره نمود

<sup>1</sup> Support Vector Machine (SVM)

<sup>2</sup> Vapnik.

<sup>3</sup> Asefa et al.

<sup>4</sup> Yu et al.

<sup>5</sup> Lan-Yang

<sup>6</sup> Grid Search

<sup>7</sup> Liang

<sup>8</sup> Principal Component Analysis (PCA)

<sup>9</sup> Artificial Neural Network (ANN)

## ۲-۲- آنالیز مؤلفه‌های اصلی (PCA)

PCA از روشهای آماری چندمتغیره است که می‌توان از آن برای کاهش پیچیدگی تحلیل متغیرها و تفسیر بهتر اطلاعات در مواردی که با حجم زیادی از اطلاعات روبرو هستیم استفاده نمود [۱۵-۱۷]. با اعمال این روش، متغیرهای اولیه به مؤلفه‌های اصلی<sup>۱</sup> و مستقل از یکدیگر تبدیل می‌شوند. هر مؤلفه اصلی می‌تواند با دنباله زیر مشخص شود

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad (۱)$$

که در این رابطه

$Z_i$  مؤلفه مورد نظر،  $a_{ij}$  ضرایب مربوط به متغیرهای اولیه و  $X_i$  نیز متغیر اولیه است. ضرایب مربوط به متغیرهای اولیه از حل معادله زیر به دست می‌آید [۱۸]

$$|R - \lambda I| = 0 \quad (۲)$$

که در آن

$I$  ماتریس واحد،  $R$  ماتریس همبستگی بین متغیرهای اولیه و  $\lambda$  مقادیر ویژه است. از این مقادیر ویژه، بردارهای ویژه به دست می‌آیند. جزئیات بیشتر در مورد PCA در منابع دیگر ارائه شده است [۱۹ و ۲۰ و ۲۱].

## ۲-۳- ماشین بردار پشتیبان

در مورد تئوری مدل‌های مطرح شده در این مقاله، تحقیقات زیادی انجام شده است [۲۲-۲۴]. لذا در این مقاله تنها به شرح خلاصه‌ای از مدل مورد استفاده SVM به نام  $\varepsilon$ -SVM پرداخته شد. در یک مدل رگرسیونی SVM لازم است وابستگی تابعی متغیر وابسته  $y$  به مجموعه‌ای از متغیرهای مستقل  $x$  تخمین زده شود. فرض بر این است که مانند دیگر مسائل رگرسیونی، رابطه بین متغیرهای وابسته و مستقل توسط یک تابع معین  $f$  به علاوه یک مقدار اضافی نویز<sup>۲</sup> مشخص می‌شود

$$y = f(x) + \text{noise} \quad (۳)$$

بنابراین موضوع اصلی پیدا کردن فرم تابع  $f$  است که بتواند به صورت صحیح، موارد جدیدی را که SVM تاکنون تجربه نکرده است پیش‌بینی کند. این تابع به وسیله آموزش مدل SVM بر روی یک مجموعه داده به عنوان مجموعه آموزش که شامل فرآیندی به منظور بهینه‌سازی دائمی تابع خطا است، قابل دسترسی است. بر مبنای تعریف این تابع خطا، دو نمونه از مدل‌های SVM شناخته شده است که عبارتند از الف) مدل‌های رگرسیونی SVM نوع اول

که به مدل‌های SVM  $v$ - مشهورند ب) مدل‌های رگرسیونی SVM نوع دوم که به مدل‌های  $\varepsilon$ -SVM مشهوراند. در این مطالعه مدل  $\varepsilon$ -SVM به دلیل کاربرد گسترده آن در مسائل رگرسیونی استفاده گردید. برای این مدل، تابع خطا به صورت زیر تعریف می‌شود

$$\frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \quad (۴)$$

تابع خطای مذکور لازم است که با توجه به محدودیت‌های زیر کمینه گردد

$$\begin{aligned} w^T \phi(x_i) + b - y_i &\leq \varepsilon + \xi_i^* \\ y_i - w^T \phi(x_i) - b &\leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* &\geq 0, \quad i = 1, \dots, N \end{aligned} \quad (۵)$$

که در این روابط

$C$  ثابت گنجایش<sup>۳</sup>،  $w$  بردار ضرایب،  $w^T$  ترانزاده بردار ضرایب،  $\xi_i$  و  $\xi_i^*$  ضرایب کمبود<sup>۴</sup>،  $b$  ضریبی ثابت،  $N$  الگوی آموزش مدل و  $\phi$  تابع کرنل است. کرنل تابع پایه شعاعی<sup>۵</sup> به عنوان بهترین انتخاب از بین دیگر توابع کرنل گزارش شده است [۱۴، ۲۵ و ۲۶]. بنابراین در این تحقیق تابع مذکور که توسط  $\gamma$  در رابطه ۶ تعریف شده است، به کار گرفته شد

$$K(x_i, x) = \exp\left(-\gamma |x_i - x|^2\right) \quad (۶)$$

با توجه به روابط ۴ تا ۶ لازم است برای پیش‌بینی دبی ماهانه توسط مدل SVM سه پارامتر  $\gamma$ ،  $\varepsilon$  و  $C$  بهینه شوند.

## ۳- نتایج و بحث

### ۳-۱- توسعه مدل SVM

برای مدل‌سازی جریان توسط مدل SVM با توجه به نوع تابع کرنل مورد استفاده در این تحقیق (تابع کرنل RBF) لازم است سه پارامتر  $C$ ،  $\varepsilon$  و  $\gamma$  بهینه شوند. برای بهینه‌سازی سه پارامتر مذکور از روش جستجوی شبکه استفاده شد [۲۷]. با روش بهینه‌سازی جستجوی شبکه، مقادیر  $C$ ،  $\varepsilon$  و  $\gamma$  به ترتیب ۰.۳، ۰.۰۰۲ و ۴/۵ به دست آمد. همچنین برای ارزیابی نتایج مدل‌ها از سه آماره ضریب تعیین ( $R^2$ )، میانگین قدرمطلق خطا<sup>۶</sup> و جذر میانگین مربعات خطا<sup>۷</sup> استفاده شد که نتایج مراحل آموزش و آزمایش مدل SVM برای پیش‌بینی جریان ماهانه بر مبنای این سه آماره در جدول ۱ نشان

<sup>3</sup> Capacity Constant

<sup>4</sup> Slack Variables

<sup>5</sup> Radial Base Function (RBF)

<sup>6</sup> Mean Absolute Error (MAE)

<sup>7</sup> Root Mean Square Error (RMSE)

<sup>1</sup> Principal Components (PCs)

<sup>2</sup> Noise

داده شده است. نتایج مرحله آزمایش مدل منتخب SVM نیز در شکل ۲ آمده است.

جدول ۱- نتایج مراحل آموزش و آزمایش مدل های SVM ، PCA-SVM

شاخص آماری	مرحله آموزش		مرحله آزمایش	
	SVM	PCA-SVM	SVM	PCA-SVM
R <sup>2</sup>	۱	۰/۹۲	۰/۸۸	۰/۸۲
MAE	۰/۰۳	۰/۷۹	۰/۸۴	۰/۹۹
RMSE	۰/۰۴	۱/۴۲	۱/۳۶	۱/۶۹

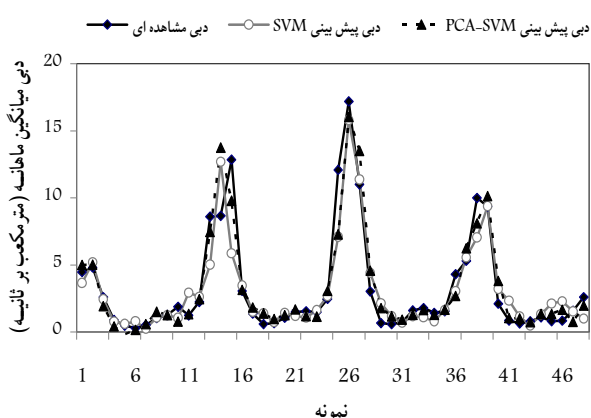
### ۳-۳- آماره نسبت تفاوت توسعه یافته<sup>۱</sup> (DDR)

مطابق جدول ۱ مدل PCA-SVM نتایج بهتری نسبت به مدل SVM دارد و تمام شاخصهای آماری، نتایج بهتری را برای این مدل نشان می‌دهند. ذکر این نکته لازم است که آماره‌های R<sup>2</sup>، MAE و RMSE بیانگر معیار متوسطی از خطا در مدل بوده و از روی آنها نمی‌توان قضاوتی صحیح در مورد عملکرد مدل و نحوه توزیع خطا انجام داد. علاوه بر این مقادیر R<sup>2</sup> و RMSE به شدت تحت تأثیر تعداد الگوهای مورد استفاده و همچنین داده‌های پرت است. بنابراین معرفی آماره‌ای مناسب و استفاده از آن در کنار آماره‌های مذکور می‌تواند در تعیین مدل برتر کمک زیادی به محقق نماید. در این راستا و به منظور دستیابی به آماره‌ای مناسب به منظور بررسی عملکرد مدل، تلاشهای متعددی صورت گرفته است که از جمله می‌توان به آماره ارائه شده توسط جین و ایندروتی و همچنین نوری و همکاران اشاره نمود [۲۸-۳۱]. با توجه به مطالب مذکور در این قسمت از تحقیق برای ارزیابی بهتر این دو مدل از نسبت تفاوت آماری توسعه یافته DDR، استفاده شد (رابطه ۷). جزئیات بیشتر در مورد این آماره در مراجع مربوطه در دسترس است [۲۹-۳۱]

$$DDR = \left( \frac{\text{Predicted Value}}{\text{Observed Value}} \right) - 1 \quad (7)$$

مطابق رابطه ۷ اگر DDR=0، آنگاه مقادیر پیش‌بینی شده معادل مقادیر اندازه‌گیری شده می‌شود. اگر DDR>0، مقدار پیش‌بینی شده از مقدار مشاهداتی بیشتر و اگر DDR<0، مقدار پیش‌بینی شده از مقدار مشاهداتی کمتر می‌گردد. برای قضاوت بهتر در این مورد می‌توان تابع گوسی مقادیر DDR، را محاسبه و به صورت یک توزیع نرمال استاندارد رسم نمود. برای این منظور ابتدا لازم است مقادیر DDR استاندارد شده و سپس با استفاده از تابع گوسی مقادیر نرمال شده DDR (Q<sub>DDR</sub>) محاسبه گردد. توزیع نرمال استاندارد برای هر یک از مدل‌های SVM و PCA-SVM در شکل ۳ آمده است.

در شکل ۳ تمایل بیشتر گراف به خط مرکزی و بزرگ‌تر بودن مقدار ماکسیمم Q<sub>DDR</sub> برای هر مدل نشان دهنده دقت بیشتر مدل



شکل ۲- نتایج مرحله آزمایش مدل‌های SVM و PCA-SVM

### ۳-۲- توسعه مدل PCA-SVM

مقدار آماره KMO=۰/۶۹۴ نیاز به استفاده از PCA بر روی متغیرهای ورودی را تأیید کرد [۱۷]. به منظور استفاده از PCA از استانداردسازی متغیرهای ورودی، ماتریس متقارن همبستگی R از مرتبه ۱۸ تشکیل شد. پس از حل معادله ۲، ۱۸ مقدار ویژه و برای هر مقدار ویژه ۱۸ بردار ویژه به دست آمد. با استفاده از این مقادیر، ۱۸ مؤلفه اصلی از متغیرهای ورودی تشکیل شد. در این مقاله براساس توصیه نوری و همکاران در مورد ارزیابی پیش‌پردازش متغیرهای ورودی به ANN برای پیش‌بینی دبی جریان ماهانه با استفاده از PCA، پنج مؤلفه اول به عنوان ورودی به مدل SVM استفاده گردید [۱۳]. پس از آن همانند مرحله قبل، مدل SVM مناسب برای پیش‌بینی جریان توسعه داده شد. در این مرحله نیز مقادیر بهینه سه پارامتر C، ε و γ به ترتیب معادل ۱/۶، ۰/۰۲۷ و ۳/۲ به دست آمد. نتایج مراحل آموزش و آزمایش این مدل (PCA-SVM) در جدول ۱ آمده است. مطابق جدول ۱ مشخص است که هر سه آماره مورد استفاده در مرحله آزمایش مدل‌ها برای

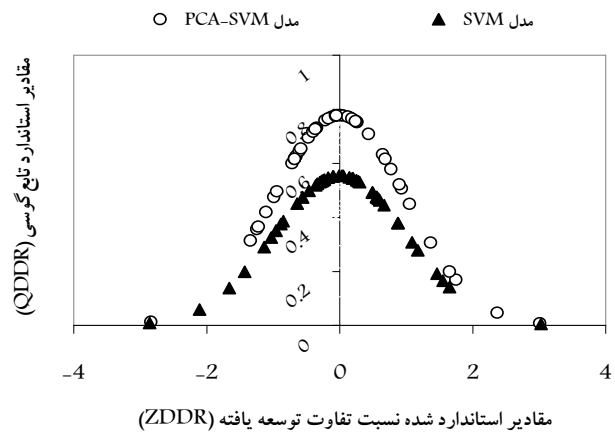
<sup>1</sup> Developed Discrepancy Ratio (DDR)

به نتایج به دست آمده در این تحقیق مدل PCA-SVM به عنوان بهترین مدل برای پیش بینی دبی ماهانه جریان انتخاب شد.

#### ۴- نتیجه گیری

در این تحقیق یک روش پیش پردازش داده های ورودی به مدل SVM یعنی آنالیز مؤلفه اصلی مورد استفاده قرار گرفت و تأثیر آن بر نتایج به دست آمده از مدل SVM بررسی شد. نتایج بررسی ها نشان داد که روش پیشنهادی می تواند عملکرد بهتر و دقیق تر مدل SVM را برای پیش بینی دبی ماهانه فراهم آورد. در حالت کلی از این تحقیق مطالب زیر نتیجه گیری می شود:

- ۱- پیش پردازش متغیرهای ورودی با استفاده از PCA تأثیر مثبتی بر عملکرد مدل SVM دارد.
- ۲- نسبت تفاوت آماری توسعه داده شده، یک شاخص آماری مناسب برای کنترل عملکرد مدل ها بوده و نتایج آن، هماهنگی مناسبی با دیگر آماره های مرسوم دارد.
- ۳- در نهایت پیش پردازش داده ها برای افزایش عملکرد مدل SVM به خصوص در مواردی که اطلاعات کافی در مورد متغیرهای ورودی وجود ندارد و تعداد ورودی ها زیاد نیست.



شکل ۳- نمودار توزیع نرمال استاندارد شده مقادیر DDR برای مدل های SVM و PCA-SVM

است. ماکسیمم مقدار  $Q_{DDR}$  برای مدل های SVM و PCA-SVM به ترتیب ۰/۵۵ و ۰/۷۸ است. با توجه به نتایج به دست آمده از این شکل واضح است که پیش پردازش متغیرهای ورودی با استفاده از PCA می تواند عملکرد مدل SVM را بهبود بخشد. بنابراین با توجه

#### ۵- مراجع

- 1- Kuligowski, R., and Barros, A. P. (1998). "Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks." *Weather and Forecasting*, 13 (40), 1195-1205.
- 2- Adeloye, A. J., and Munari, A. D. (2006). "Artificial neural network based generalized storage-yield-reliability models using the Levenberg-Marquardt algorithm." *J. of Hydrology*, 362 (1-4), 215-230.
- 3- Zhao, R. J. (1992). "The Xinanjiang model applied in China." *J. of Hydrology*, 135 (1-4), 371-381.
- 4- Jain, A., and Srinivasulu, S. (2006). "Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques." *J. of Hydrology*, 317 (3-4), 291-306.
- 5- Xiong, L. H., Shamseldin, A. Y., and O'Connor, K. M. (2001). "A nonlinear combination of the forecasts of rainfall-runoff models by the first order Takagi-Sugeno fuzzy system." *J. of Hydrology*, 245 (1-4), 196-217.
- 6- Muller-Wohlfeil, D.I., Xu, C. Y., and Iversen, H. L. (2003). "Estimation of monthly river discharge from Danish catchments." *Nordic Hydrology*, 34 (4), 295-320.
- 7- Asefa, T., Kemblowski, M., McKee, M., and Khalil, A. (2005). "Multi-time scale stream flow predictions: The support vector machines approach." *J. of Hydrology*, 318 (1-4), 7-16.
- 8- Vapnik, V. N. (1995). *The nature of statistical learning theory*, 1<sup>st</sup> Ed., Springer-Verlag, New York.
- 9- Yu, P. S., Chen, S. T., and Chang, I. F. (2006). "Support vector regression for real-time flood stage forecasting." *J. of Hydrology*, 328 (3-4), 704-716.
- 10- Yu, X., and Liong, S. Y. (2006). "Forecasting of hydrologic time series with ridge regression in feature space." *J. of Hydrology*, 332 (3-4), 290-302.
- 11- Zhang, Y. X. (2007). "Artificial neural networks based on principal component analysis input selection for clinical pattern recognition analysis." *Talanta*, 73 (1), 68-75.
- 12- Zhang, Y., Li, H., Hou, A., and Havel, J. (2006). "Artificial neural networks based on principal component analysis input selection for quantification in overlapped capillary electrophoresis peaks." *Chemometrics and Intelligent Laboratory Systems*, 82 (1-2), 165-175.

- 13- Noori, R., Farrokhnia, A., Morid, S., and Riyahi-Madvar, H. (2008). "Effect of input variables preprocessing in artificial network on monthly flow prediction by PCA and wavelet transformation." *J. of Water and Wastewater*, 69, 13-22. (in Persian)
- 14- Noori, R., Abdoli, M. A., Ameri A., and Jalili-Ghazizade, M. (2008). "Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: A case study of Mashhad." *Environmental Progress and Sustainable Energy*, 28 (2), 249-258.
- 15- Camdevyren, H., Demyr, N., Kanik, A., and Kesdyn, S. (2005). "Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs." *Ecol. Model*, 181 (4), 581-589.
- 16- Lu, W. Z., Wang, W. J., Wang, X. K., Xu, Z. B., and Leung, A. Y. T. (2003). "Using improved neural network to analyze RSP, NO<sub>x</sub> and NO<sub>2</sub> levels in urban air in Mong Kok, Hong Kong." *Environmental Monitoring and Assessment*, 87 (3), 235-254.
- 17- Noori, R., Kerachian, R., Khodadadi, A., and Shakibinia, A. (2007). "Assessment of importance of water quality monitoring stations using principal component and factor analysis: A case study of the karoon river." *J. of Water and Wastewater*, 63, 60-69. (In Persian)
- 18- Manly, B. F. J. (1986). *Multivariate statistical methods: A primer*, 2<sup>nd</sup> Ed., Chapman and Hall, London.
- 19- Tabachnick, B. G., and Fidell, L. S. (2001). *Using multivariate statistics*, 3<sup>rd</sup> Ed., Allyn and Bacon, Boston, London.
- 20- Ouyang, Y. (2005). "Evaluation of river water quality monitoring stations by principal component analysis." *Water Research*, 39 (12), 2621-2635.
- 21- Noori, R., Ashrafi, Kh., and Ajdarpour, A. (2008). "Comparison of ANN and PCA based multivariate linear regression applied to predict the daily average concentration of Co: A case study of Tehran." *J. of Physics Earth Space*, 34 (1), 135-152.
- 22- Vapnik, V. N. (1998). *Statistical learning theory*, 1<sup>st</sup> Ed., Wiley, New York.
- 23- Cristianini, N., and Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*, 1<sup>st</sup> Ed., Cambridge University Press, Cambridge.
- 24- Chen, S. T., and Yu, P. S. (2007). "Real-time probabilistic forecasting of flood stages." *J. of Hydrology*, 340 (1-2), 63-77.
- 25- Noori, R., Karbassi, A., Farokhnia, A., and Dehghani, M. (2009). "Predicting the longitudinal dispersion coefficient using support vector machine and adaptive neuro-fuzzy inference system techniques." *Environmental Engineering Science*, 26 (10), 1503-1510.
- 26- Dibike, Y. B., Velickov, S., Solomatine, D. P., and Abbott, M. B. (2001). "Model induction with support vector machines: Introduction and applications." *J. of Computing in Civil Eng.*, 15 (3), 208-216.
- 27- Hsu, C. W., Chang, C. C., and Lin, C. J. (2003). "A practical guide to support vector classification." <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>, (4 Mar. 2009)
- 28- Jain, A., and Indurthy, S. K. V. P. (2003). "Comparative analysis of event based rainfall-runoff modeling techniques-deterministic, statistical, and artificial neural network." *J. of Hydrologic Engineering*, 8(2), 93-98.
- 29- Noori, R., Hoshiyaripour, G. A., Ashrafi, K., and Araabi, B. N. (2010). "Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration." *Atmospheric Environment*, 44 (4), 476-482.
- 30- Noori, R., Khakpour, A., Omidvar, B., and Farokhnia, A. (2010). "Comparison of ANN and principal component analysis multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statiste." *Expert Systems with Applications*, 37 (8), 5856-5862.
- 31- Noori, R., Karbassi, A. R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M. H., Forokhnial, A., and Ghafari-Goushesh, M. (2011). "Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction." *J. of Hydrology*, (In Press).