فصلنامــه علمی پژوهشی دانش سرمایه‌گــذاری
سال چهارم/ شماره چهاردهم/ تابستان ۱۳۹٤

پیش‌بینی قیمت هفتگی نفت خام از طریق مدل ترکیبی ماشین بردار پشتیبان و خودرگرسیو میانگین متحرک انباشته

شاپور محمدی
دانشیار، مدیریت مالی، دانشگاه تهران، دانشکده‌ی مدیریت
shmohammadi@ut.ac.ir

رضا راعی
استاد، مدیریت مالی، دانشگاه تهران، دانشکده‌ی مدیریت
Raei@ ut.ac.ir

حسین کرمی
کارشناسی ارشد، مدیریت مالی، دانشگاه تهران، دانشکده‌ی مدیریت (مسئول مکاتبات)
hkfinance85 @ yahoo.com

تاریخ دریافت: ۹۳/۱۰/۱۶ تاریخ پذیرش: ۹۴/۱/۲۹

چکیده

همواره پیش‌بینی روند قیمت و نوسانات یکی از چالش‌های پیش‌روی معامله‌گران در بازارهای بورس نفت بوده و پیش‌بینی قیمت‌ها به عنوان یک امر ضروری وکاربردی مطرح می‌شود ولیکن باید پیش‌بینی را مورد توجه قرار داد که با دقت بیشتری صورت گیرد و نسبت به نتایج واقعی مشاهده شده خطای کمتری داشته باشد. به منظور پیش‌بینی قیمت هفتگی نفت خام برنت به عنوان یک نفت شاخص با توجه به دشوار بودن شناسایی دقیق الگوهای خطی و غیرخطی در سری‌های زمانی اقتصادی و مالی از ترکیب مدل‌های خودرگرسیو میانگین متحرک انباشته(ARIMA) با این پیش فرض که سری‌زمانی دارای الگوی خطی می‌باشد و ماشین بردار پشتیبان(SVM)که توانایی بالایی در مدل‌سازی الگوهای غیرخطی دارد به منظور افزایش دقت پیش‌بینی استفاده شده است. با توجه به آزمون مقایسه زوجی معیارهای ارزیابی عملکرد ریشه میانگین مجذور خطا(RMSE) و میانگین قدرمطلق درصد خطا(MAPE) که حاصل مقادیر پیش بینی شده و مقادیر واقعی هر یک از مدل‌ها هستند، نتایج بیانگر این موضوع بودند که در بیشتر موارد مدل ترکیبی خطای کمتری در پیش بینی قیمت نفت خام نسبت به کاربرد مجزای مدل‌های خودرگرسیو میانگین متحرک انباشته و ماشین بردار پشتیبان دارد.

واژه‌های کلیدی: خودرگرسیو میانگین متحرک انباشته (ARIMA)[۱]، ماشین بردار پشتیبان(SVM)[۲]، مدل ترکیبی[۳]، مانایی[۴].

۱- مقدمه

پیش‌بینی روند قیمت و نوسانات یکی از چالش‌های معامله‌گران و سرمایه‌گذاران در بازارهای بورس نفت بوده وپیش‌بینی قیمت‌ها به عنوان یک امر ضروری وکاربردی مطرح می‌شود ولیکن باید پیش‌بینی را مورد توجه قـرار داد که با دقت بیشتری صورت گیرد ونسبت به نتایج واقعی مشاهده شده خطای کمتری داشـته باشـد.در پـیش بینی سری‌های زمانی از مشاهدات گذشته‌ی یک متغیر به منظور توسعه‌ی مدل وپیش‌بینی درآینده استفاده می گردد. مدل‌های سری زمانی در طول چنـد دهـه گذشـته توسـعه بسـیاری یافتـه‌انـد امـا یکـی ازسـاده‌تـرین و پرکاربردترین آن‌ها مدل خودرگرسیو میانگین متحرک انباشته (ARIMA)[۵] می‌باشد که تحت عنوان روش باکس و جنکینز[۶] شناخته می‌شود. با وجود انعطاف پذیری بالای مدل مذکور در مدل‌سازی الگوهای خطی به‌دلیـل دارا بودن اجزای خودرگرسیو و میانگین متحرک،این مدل نمی‌تواند الگوهای غیرخطی را به طور دقیـق مـدل‌سـازی کند. شبکه‌های عصبی که به طور گسترده در پیش‌بینی سری‌های زمانی مورد استفاده قرار می‌گیرنـد در مـدل-سازی الگوهای غیرخطی بسیار انعطاف‌پذیر می‌باشند. الگوریتم ماشین بردار پشتیبان بـر اسـاس اصـول تئـوری یادگیری آماری و حداقل‌سازی ریسک ساختاری پایه‌گذاری شده است. این الگوریتم در ابتدای سـال‌هـای دهـه ۱۹۹۰ توسط واپنیک[۷] و همکارانش برای حل مسایل طبقه‌بندی طراحی شد و پس از آن با پیشـرفت‌هـایی کـه توسط هابر[۸] و دیگران صورت گرفت از ماشـین بـردار پشـتیبان در پـیش‌بینـی سـری‌هـای زمـانی بـه خصـوص مدل‌سازی سری‌های زمانی غیرخطی مورد استفاده قرار گرفت. به دلیل اینکه شناسایی الگوی داده‌ها در واقعیت اغلب پیچیده می‌باشد، مطالعات زیادی به منظور ترکیب مدل‌های خطی و غیرخطی انجام پذیرفته است.

۲- مبانی نظری و مروری بر پیشینه پژوهش

در زمینه مدل‌سازی سری‌های زمانی، روش‌های متفاوتی مورد استفاده قرار می‌گیرد که می‌توان به مدل‌های سنتی میانگین متحرک وخودرگرسیو میانگین متحرک انباشته که الگوهای خطی را مدل‌سازی می‌کنند و به دلیل سادگی در فهم و کاربرد در دهه‌های اخیر بسیار مورد استفاده قرار گرفته است، اشاره کرد. به دلیل مشاهده الگوهای غیرخطی در دنیای واقعی یک سری از مدل‌های غیرخطی مانند[۹]ARCH،[۱۰]GARCH و[۱۱] TGARCHمطرح شدند ولی همگی این مدل‌های غیرخطی الگوهای بخصوصی را توضیح می‌دهند. کار الگوریتم ماشین بردار پشتیبان در واقع یافتن قوانین حاکم بر داده‌ها و نیز روابط نهفته میان آن‌ها می‌باشد. این مدل‌های یادگیری، همه بر پایه این فرض اولیه استوارند که تجربیات کسب شده و آموخته‌های ماشین یادگیری برای پیش‌بینی و تصمیم‌گیری در آینده مناسب می‌باشند. واضح است که این خود فرض ثابت بودن مدل احتمالی تولید کننده داده‌ها را می‌طلبد. در SVM ها نیز هدف الگوریتم یادگیری بدست آوردن توانایی تعیین نسبتاً دقیق کلاس داده‌های دیده نشده پس از پایان مرحله آموزش می‌باشد. (آبه ،۲۰۰۵). با وجود مزایای متعدد، شبکه های عصبی با محدودیت‌هایی همراه هستندکه از جمله آنها می‌توان به ارائه نتایج پیچیده در مدل‌های خطی اشاره کرد و همچنین فرایند یادگیری نیازمند شمار بالای نمونه برای آموزش می‌باشد.

به دلیل اینکه تشخیص الگوی سری‌های زمانی مبنی بر خطی یا غیرخطی بودن روابط دشوار است و از طرفی به ندرت روابط کاملاً خطی و یا کاملاً غیرخطی مشاهده می‌شود (اغلب ترکیبی از دو الگو را دارا می‌باشند)، بنابراین ترکیبی از مدل‌های ARIMA و SVM می‌تواند به افزایش دقت پیش‌بینی کمک کند.

۲-۱- خودرگسیو میانگین متحرک انباشته(ARIMA)

مدل‌های خودرگرسیو میانگین متحرک انباشته را معمولاً به صورت ARIMA(p, d, q) نمایش می‌دهند و در این مدل‌ها برای پیش بینی مقادیر آتی متغیر از ترکیب خطی مقادیر گذشته متغیر و اجزای اخلال استفاده می‌شود. که در آن p مرتبه مورد نیاز برای خود رگرسیو،d میزان تفاضل مورد نیاز برای مانا کردن سری زمانی وq میزان مرتبه‌ی مورد نیاز برای میانگین متحرک است.

$$Y_t = \theta_\circ + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} - \cdots + \theta_q \phi_{t-q} + U_t \qquad (۱)$$

Y_t = *مقدار واقعی متغیر*

U_t = *خطای تصادفی در زمان* t

θ_j, ϕ_i = *ضرایب خودرگرسیو و میانگین متحرک*

p, q = *مرتبه های خودرگرسیو ومیانگین متحرک*

به طور مثال مدل ARIMA(1, 0, 1) به صورت زیر نمایش داده می‌شود.

$$Y_t = \theta_\circ + \phi_1 Y_{t-1} + \emptyset_1 u_{t-1} + U_t \qquad (۲)$$

شرط بهره‌گیری از مدل‌های خودرگرسیو میانگین متحرک انباشته، مانا بودن سری‌های زمانی می‌باشد. برای مانا کردن سری‌های زمانی مالی و اقتصادی اغلب از روش تفاضل استفاده می‌شود. جزء انباشتگی در مدل خودرگرسیو میانگین متحرک انباشته بیانگر میزان تفاضل مورد نیاز برای مانا کردن سری زمانی است. مدل خودرگرسیو میانگین متحرک انباشته بهینه، مدلی است‌که پسماندهای آن همبستگی خطی نداشته باشند، اما به دلیل اینکه تحلیل‌های پسماند قادر به شناسایی الگوهای غیرخطی نمی‌باشد، بنابرای وجود روابط غیرخطی در پسماندها می‌تواند محدودیتی برای مدل ARIMA به حساب آید.

مرحله اول: به منظور به کارگیری مدل خود رگرسیو میانگین متحرک انباشته جهت پیش بینی ، سری زمانی را از طریق روش تفاضل مانا می کنیم.

مرحله دوم: تشخیص وشناسایی مدل به معنی تعیین مرتبه مدل می‌باشد. مهمترین ابزار برای تشخیص مرتبه مدل، استفاده از تابع خود همبستگی (ACF)[۱۲] وتابع خود همبستگی جزئی (PACF)[۱۳] است.

مرحله سوم: برآورد مدل عمدتاً وابسته به تشخیص مدل در مرحله دوم است. مدل را می توان با روش OLS یا روش‌های دیگر مانند حداکثر درست نمایی برآورد نمود.

مرحله چهارم: در این مرحله با توجه به توابع خودهمبستگی وخودهمبستگی جزئی وآماره‌ی باکس والژانگ، مستقل بودن سری پسماندها مورد آزمون قرار می گیرد. مدل خودرگرسیو میانگین متحرک انباشته بهینه، مدلی است که پسماند های‌آن فاقد همبستگی خطی باشند.

مرحله پنجم: در این مرحله با استفاده از مدل خود رگرسـیو میـانگین متحـرک انباشـته ایـی کـه بـین سـری پسماندهای آن همبستگی وجود ندارد به پیش بینی ارزش متغیر درآینده می پردازیم.

۲-۲- ماشین بردار پشتیبان(SVM)

پایه‌های SVM در سال ۱۹۹۵ توسط ولادیمیر واپنیک[۱۴] گذارده شد.این ماشین‌های یادگیری بـه‌دلیـل دارا بودن ویژگی‌های خاص و کارایی بالا در عمل به سرعت به شهرت بالایی دست یافتند و بسیاری از الگوریتم‌هـای یادگیری پیش از خود را کم اعتبار نمودند.

فرمول‌بندی SVMمینیمم کردن ریسک ساختاری (SRM)[۱۵] را به کار مـی‌گیـرد کـه برتـری آن نسـبت بـه مینیمم کردن ریسک تجربی (ERM)[۱۶] که در شبکه‌های عصبی رواج دارد ثابت شده اسـت.(گان[۱۷] ،۱۹۹۸) ایـن در واقع تفاوتی است که به SVMتوانایی تعمیم بالاتری می‌دهد و هدف یادگیری آماری نیز جـز ایـن نیسـت. از آنچه گفته شد می‌توان نتیجه گرفت که SVMها قادرند خطای طبقه‌بندی را هم در آموزش و هـم در رویـارویی با داده‌های نادیده مینیمم کنند. این در حالی است که در شبکه‌های عصبی سه لایه مرزهـای تصـمیم مسـتقیماً تعیین نمی‌شوند و کارایی شبکه در فاز تست وابسته به تعداد داده‌های آموزش، روش آمـوزش و وزن‌هـای اولیـه دارد. این موضوع در صورتی که تعداد داده‌ها کم باشد به شدت قابلیت تعمـیم را تنـزل مـی‌دهد.(آبـه[۱۸] ،۲۰۰۵) واضح است که SVMها از این حیث نیز از سایر طبقه‌بندها برترند و حتی با وجود یک مشاهده برای هـر کـلاس نیز به‌خوبی کار می‌کنند.

در یادگیری و به‌طور دقیق‌تر در یادگیری از روی نمونه‌ها، یک مجموعه از مشـاهدات در اختیـار یادگیرنـده قرار می‌گیرد و سیستم یادگیرنده با یک الگوریتم مناسب از روی این داده‌ها یک مدل می‌سازد. این مدل پـس از تکمیل قادر خواهد بود که نماینده تمام داده‌ها باشد. در واقع ثمره فرآینـد یـادگیری را در ایـن مـدل مـی‌تـوان خلاصه کرد و دیگر به وجود داده‌ها نیازی نخواهد بود. چنین مدلی شامل یک یا چند تابع تصمیم خواهد بود.

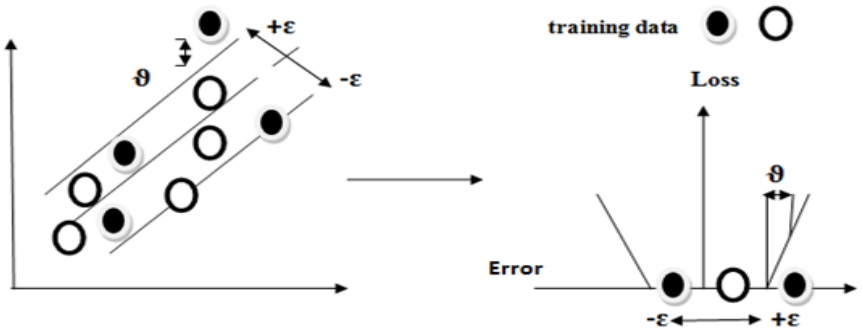
توابع تصمیم را می‌توان به الگوهای آموخته شده از مشاهدات موجود توسط ماشین یـادگیری تعبیـر نمـود. بنابراین در فاز تست، توابع تصمیم که نقش آموخته‌ها و تجربیات یادگیرنده را بازی می‌کنند تعیین کننده پاسـخ خروجی ماشین یادگیری به ازای داده‌های تست می‌باشند.

طبقه‌بندهای مبتنی بر شبکه‌های عصبی چندلایه نسبت به حضور داده‌های خـارجی آسـیب‌پذیرنـد زیـرا از حاصل جمع مربعات خطاها استفاده می‌کنند. بنابراین، برای جلوگیری از تأثیر داده‌های خارجی بایستی پـیش از آموزش حذف گردند. و یا مکانیزمی برای کم اثر کردن داده‌های خارجی باید در آموزش اتخاذ شود. درSVMهـا، پارامتر حاشیهC خطای طبقه‌بندی را کنترل می‌کند. اگر مقدار آن را بزرگ انتخاب کنیم، خطای طبقه‌بندی بـه کمینگی می‌گراید و اگر مقدار کوچکی برای آن برگزینیم، در واقع پذیرفته‌ایم که داده‌هایی کـه از مراکـز تجمـع دور هستند اشتباه طبقه‌بندی شوند. پس با انتخاب درست C، داده‌های خارجی کم‌اثر خواهند بود.

در آموزش یک طبقه‌بند مبتنی بر شبکه‌های عصبی چندلایه، حاصل جمع مربعات خطاهـای موجـود میـان ورودی‌ها و خروجی‌های مطلوب کمینه می‌شود. بنابراین، مرزهای کلاس و در نتیجه قابلیت تعمیم به مقدار قابل

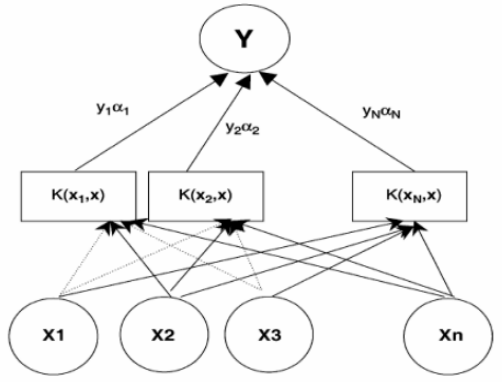
ملاحظه‌ای تقلیل می‌یابد. اما از آنجا که در SVMبنای آموزش بر بیشینه کردن حاشیه است، قابلیت تعمیم حتی در شرایط خاص مذکور هم آنچنان دست‌خوش تغییر نمی‌شود (آبه،۲۰۰۱).

ماشین بردار پشتیبان، تابع رگرسیون را با به کارگیری یک تابع خطی تخمین می‌زند و عملیات رگرسیون را با تابعی که انحراف از مقدار واقعی در آن به میزان کمتر از ε مجاز است انجام می دهد(تابع ضرر)[۱۹]، ماشین بردار پشتیبان با کمینه کردن ریسک ساختاری جواب بهینه را ارائه می دهد.در رگرسیون نیز SVM ها پس ازنگاشت داده‌ها به فضای ویژگی، مسائل پیچیده را به صورت رگرسیون خطی حل می‌نمایند.



شکل ۱- تابع ضرر وپنیک

مساله رگرسیون خطی در SVM به آسانی قابل گسترش به رگرسیون غیر خطی است. بدین منظور از توابع کرنل استفاده می‌شود و تاکنون کرنل های گوناگونی از جمله کرنل های چند جمله‌ای و پایه شعاعی(RBF) معرفی شده اند.



شکل ۲- فرایند ماشین بردار پشتیبان

۲-۳- مدل ترکیبی (hybrid model)

در مدل ترکیبی سعی بر این داریم که روابط غیرخطی در پسماندها را با استفاده از SVM مدل‌سازی کرده و پیش‌بینی پسماندهای حاصل از ماشین بردار پشتیبان را به پیش‌بینی خودرگرسیو میانگین متحرک انباشته اضافه کنیم. با این ترکیبب الگوهای خطی بوسیله‌یARIMA و الگوهای غیرخطی بوسیله‌ی SVM مدل‌سازی خواهند شد. از مدل ترکیبی به منظور بالابردن دقت پیش‌بینی قیمت هفتگی نفت خام برنت که به عنوان نفت شاخص معرفی می‌شود و مبنای قیمت گذاری دو سوم معاملات در بازارهای بورس نفت و فرآورده‌های نفتی است استفاده خواهیم کرد و فرضیه های پژوهش را مورد آزمون قرار می‌دهیم.

۲-۳-۱ فرآیند مدل ترکیبی

سری زمانی Y_t را می توان به دو بخش خطی و غیر خطی تقسیم کرد.

$$Y_t = L_t + N_t \qquad (۳)$$

در این مدل L_t به عنوان بخش خطی وN_t به عنوان بخش غیرخطی مدل ترکیبی معرفی می‌شود. در معادله زیر $\widehat{L}_t$بیانگر پیش‌بینی بخش خطی مدل از طریق مدل خودرگرسیو میانگین متحرک انباشته در زمان t وe_t نشان دهنده پسماند در زمان t است که از طریق مدل خودرگرسیو میانگین متحرک انباشته محاسبه شده است.

$$e_t = Y_t - \hat{L}_t \qquad (۴)$$

پسماندها از طریق ماشین بردار پشتیبان مدل سازی شده و به صورت زیر نمایش داده می‌شوند.

$$e_t = f(e_{t-1}, e_{t-2}, ..., e_{t-n}) + \varepsilon_t \qquad (۵)$$

در معادله بالا f بیانگر تابع غیرخطی که از طریق ماشین دار پشتیبان برازش می‌شود و ε_t به عنوان جزء اخلال معرفی می‌شود. در معادله زیر $\widehat{N}_t$ بیانگر پیش بینی پسماندها یا همان جزء غیرخطی است که از طریق ماشین بردار پشتیبان مدل سازی می‌شود و به بخش خطی مدل اضافه می‌گردد.

$$\hat{e}_t = \hat{N}_t \qquad (۶)$$

$$\hat{Y}_t = \hat{L}_t + \hat{N}_t \qquad (۷)$$

$\hat{Y}_t$: پیش بینی مدل ترکیبی

$\hat{L}_t$: پیش بینی مدل خودرگرسیو میانگین متحرک انباشته

$\widehat{N}_t$:پیش بینی پسماندها با ماشین بردار پشتیبان

۲-۴- پیشینه پژوهش

درتحقیقات داخلی وخارجی مطالعات زیادی انجام شده است که برتری مدل شبکه‌های عصبی مصنوعی را بر روش‌های سنتی در پیش‌بینی سری‌های زمانی نشان داده‌اند. اما در زمینه روش‌های ترکیبی ماشین بردار پشتیبان و روش‌های سنتی در تحقیقات داخلی مطالعات کمی صورت گرفته است، از این رو نمونه هایی از

تحقیقات انجام شده به وسیله روش‌های سنتی و مدل شبکه‌های عصبی مصنوعی که برای پیش‌بینی مورد استفاده قرار گرفته است را ذکر خواهیم کرد.

هونگ ژان و همکاران(۲۰۱۲)[۲۰] برای پیش بینی بارگذاری کوتاه‌مدت الکتریسیته با ترکیب مدل‌های خودرگرسیو میانگین متحرک انباشته و ماشین بردار پشتیبان به پیش‌بینی مدیریت کارای سیستم نیرو پرداختند و بیان نمودند که پیش‌بینی سری‌های زمانی از طریق مدل ترکیبی خودرگرسیو میانگین متحرک انباشته و ماشین بردارپشتیبان در مقایسه با کاربرد مجزای مدل‌هایARIMA و SVM ازدقت بیشتری برخودار است و معیارهای ارزیابیRMSE وMAPE کمتری را نشان می‌دهد.گاوریشچاکا و بانرجی(۲۰۰۶)[۲۱] با توجه به داده‌های مربوط به ۷۵۰روز شاخصS&P500 بین سال های ۱۹۹۹تا ۲۰۰۳،۵۰۰ داده را به عنوان مجموعه آموزش و۲۵۰ داده بعدی را برای مجموعه آزمون مورد استفاده قرار دادند.همچنین آنها از چند تابع‌کرنل درتحقیق خود استفاده‌کردند و به این نتیجه دست یافتند که کرنل RBF[۲۲] جهت پیش بینی بهترعمل می‌کند وپژوهش آنها نشان داد کهSVM بر مدل‌های EGARCH و GARCH برتری دارد.پینگ و چین (۲۰۰۵)[۲۳] با استفاده از مجموعه داده‌های مربوط به قیمت واقعی سهام ده شرکت با توجه به پیش‌بینی یک گام به جلو بیان نمودندکه مدل‌ترکیبی سری‌های زمانی قیمت سهام در مقایسه با مدل‌های خودرگرسیو میانگین متحرک انباشته و ماشین بردار پشتیبان، معیارهای ارزیابی MAPE وRMSE کمتری را نشان می‌دهد و دقت پیش بینی مدل ترکیبی نسبت به پیش بینی مجزا به وسیله هر یک از مدل‌های ماشین بردار پشتیبان وخودرگرسیو میانگین متحرک انباشته بیشتر است.

یو،وانگ ولای(۲۰۰۵)[۲۴] با توجه به مقادیر شاخصS&P و چند نماگر تحلیل تکنیکال به عنوان ورودی و انتخاب متغیرها به وسیله الگوریتم ژنتیک و طراحی مدل ماشین بردار پشتیبان با توجه به تابعRBF جهت پیش بینی روند بازار، در نهایت مدل ترکیبی خود را با شبکه عصبی پس‌انتشار خطا،ARIMA وSVM مقایسه نمودند.پژوهش آنها نشان داد که روش ترکیبی ماشین بردار پشتیبان با الگوریتم ژنتیک بر دیگر روش‌ها برتری دارد، در ضمن سریع‌تر از سایر روش‌ها عمل می‌کند. گوک (۲۰۰۲)[۲۵] در مطالعه مقایسه‌ای از شبکه‌ی‌عصبی و مدل باکس وجنکینز در پیش‌بینی سری‌های زمانی به منظور برازش مدل شبکه‌ی‌عصبی پیش‌خور چندلایه و مدل خودرگرسیو میانگین متحرک انباشته و مقایسه‌ی عملکرد این مدل‌ها درپیش بینی تعداد دفعات خرابی در سیستم کمپرسور، از سری زمانی۹۰خرابی در یک کارخانه نروژی در طی سال های ۱۹۶۸ تا ۱۹۸۹ استفاده کردند. نتایج تحقیق آنها نشان دادکه عملکرد مدل‌های شبکه‌ی‌عصبی بازگشتی و خودرگرسیو میانگین متحرک انباشته، از نظر دو مقیاس ارزیابی میانگین مجذورخطا(MSE) و میانگین قدرمطلق خطا(MAE) هم درکوتاه‌مدت و هم دربلندمدت نسبت به شبکه‌ی‌عصبی پیش‌خور چندلایه دارای خطای کمتری می‌باشند.

طبسی(۱۳۹۱) باتوجه به برآوردهای صورت گرفته برای ارزش در معرض ریسک با استفاده از مدل‌های سنتی پیش‌بینی نوسانات چون GARCH وEGARCH و انجام پس‌آزمون لوپز و پس‌آزمون مبتنی بر ریزش مورد انتظارکه رویکردهایی مقایسه‌ای برای سنجش اعتبار ارزش در معرض ریسک هستند با عملکرد مدل ترکیبی مورد مقایسه قرار گرفتند و نتایج حاصل از این پژوهش نشان داد که مدل ترکیبی نسبت به سایر مدل‌های مورد

استفاده در این تحقیق، به طور معناداری از عملکرد بهتری برخوردار است. راعی و فلاح‌پور(۱۳۸۷) اذعان داشتند که بر اساس نتیجه بررسی‌ها، مدل SVM می‌توانـد جانشـین بسـیار قدرتمنـدی بـرای مـدل آمـاری رگرسـیون لجستیک باشد.در واقع یافته‌های این تحقیق نشان داد که دقت کلی مدل ماشین بـردار پشـتیبان از دقـت کلـی مدل رگرسیون لجستیک به طور معناداری بیشتر است. به ویژه آنکه قدرت تعمـیم پـذیری مـدل ماشـین بـردار پشتیبان بیشتر از مدل رگرسیون لجستیک می‌باشد.زراءنژاد وحمیـد(۱۳۸۸) جهـت پـیش بینـی نـرخ تـورم در اقتصاد ایران با استفاده از شبکه‌های عصبی پویا(دیدگاه سری زمانی) از طریق شبکه‌های عصبی مصنوعی در هـر مرحله برای انتخاب بهترین شبکه ، خطاها و میزان پراکندگی جواب‌ها محاسبه گردید تا مدل بهینه برای پـیش بینی نرخ تورم انتخاب گردد. از آنجا که در دیدگاه سری زمانی وظیفه پیش‌بینی متغیرهای اقتصـادی بـه عهـده خود سری زمانی گذاشته می‌شود، به محقق اجازه تعیین سهم سایر عوامل موثر در تغییرات متغیر مورد نظر داده نمی‌شود و مدل طراحی شده بر اساس این دیدگاه تنها به فراگیری و مطالعه گذشته متغیر خواهد پرداخـت. لـذا اگر خطایی در داده‌های گذشته باشد یا رخ دهد، مدل براساس ورودی، مقادیر آتی را پیش بینی خواهـد کـرد و خطاهای مقادیر گذشته را به مقادیر آینده منتقل می‌کند. نوری، خاکپور، دهقانی و فرخ نیا(۱۳۸۹) بـرای پـیش بینی ماهانه جریان آب با استفاده از ماشین بردار پشتیبان بر مبنای آنالیز مولفه اصلی[۲۶](PCA) با هـدف بررسـی تاثیر انتخاب متغیرهای ورودی بر اساس آنالیز مولفه اصلی بر عملکرد مدل ماشین بردار پشتیبان انجام شـد. بـه این منظور ابتدا با استفاده از۱۸ متغیر به عنوان ورودی به مدلSVM، دبی جریان ماهانه آب پیش‌بینی گردید و سپس با استفاده از آنالیز مولفه اصلی تعداد متغیرهای ورودی به مدل ماشین بردار پشتیبان از ۱۸ متغیـر بـه ۵ مولفه کاهش یافت. در نهایت با استفاده ازآماره توسعه یافته توسط نویسنده‌گان مقاله، عملکـرد مـدل‌هـای ارائـه شده (SVM وPCA-SVM) مورد ارزیابی قرارگرفتند ویافته های تحقیق نشان داد که پیش پـردازش متغیرهـای ورودی به مدلSVM با استفاده از PCA، بهبود عملکرد SVMرا به همراه داشته است.

منجمی،ابرزی وشوازی(۱۳۸۸) به طراحی وارائه یک مدل پیش بینی قیمت سهام با استفاده از شبکه عصبی مصنوعی فازی و الگوریتم‌های ژنتیک جهت کاهش خطای پیش بینی قیمت سـهام بـا اسـتفاده ازآن نسـبت بـه استفاده از تکنیک شبکه عصبی مصنوعی به صورت منفرد پرداختند و پس از طراحی و پیاده سازی مـدل‌هـا بـا استفاده از چهار معیار سنجش خطا (R^2, MAPE, NMSE, MSE) دو مدل مورد مقایسـه قرارگرفـت. مقایسـه‌ی نتایج حاصل از مدل شبکه عصبی با نتایج حاصل از مدل شبکه‌ی عصبی فازی والگوریتم ژنتیک نشان داد کـه از نقطه نظر معیارهای ارزیابی عملکرد، پیش بینی قیمت سهام روز بعد توسط مدل ترکیبـی شـبکه عصـبی فـازی والگوریتم ژنتیک دقیق‌تر از شبکه عصبی است، به عبارتی دیگر پیش بینـی قیمـت سـهام بـا اسـتفاده از شـبکه عصبی فازی والگوریتم‌های ژنتیک، خطای برآورد قیمت سهام را نسبت به تکنیک شبکه عصبی مصنوعی کاهش می‌دهد. رمضانی(۱۳۸۳) در پژوهش خود به این نتیجه دست یافت که در پـیش‌بینـی میـزان تقاضـای ماهیانـه مصرف برق با توجه به داده‌های مربوط به سال‌های ۱۳۷۰تا سال۱۳۸۱و با توجه به پیش‌بینی‌های صورت گرفتـه توسط مدل‌های بهینه، شبکه عصبی میزان MSE و MAPE کمتری را در پیش بینی انجام گرفته نسبت به مدل خودرگرسیو میانگین متحرک انباشته داشته است و با توجه به معیارهای ارزیابی فوق مشخص اسـت کـه شـبکه

های عصبی مصنوعی عملکرد بهتری از مدل‌های خودرگرسیو میانگین انباشته درخصوص پیش‌بینی مقادیر سری‌زمانی داشته‌اند.

۳- بیان مسئله و فرضیه‌های پژوهش

مساله اصلی پژوهش و مبنای فرضیه های پژوهش این موضوع است که آیا مدل ترکیبی خطای کمتری در پیش‌بینی قیمت هفتگی نفت خام نسبت به کاربرد مدل‌های خودرگرسیو میانگین متحرک انباشته و ماشین بردار پشتیبان به طور مجزا دارد؟

فرضیه اول: بین دو جامعه ریشه میانگین مجذور خطا مدل ترکیبی و مدلARIMA اختلاف معناداری وجود دارد.

فرضیه دوم: بین دو جامعه ریشه میانگین مجذور خطا مدل ترکیبی و مدل SVM اختلاف معناداری وجود دارد.

فرضیه سوم: بین دو جامعه میانگین قدرمطلق درصد خطا مدل ترکیبی و مدلARIMA اختلاف معناداری وجود دارد.

فرضیه چهارم: بین دو جامعه میانگین قدرمطلق درصد خطا مدل ترکیبی و مدل SVM اختلاف معناداری وجود دارد.

۴- روش‌شناسی پژوهش

این پژوهش از لحاظ روش، توصیفی و از لحاظ هدف، کاربردی محسوب می‌شود چرا که می‌تواند در فرآیند تصمیم‌گیری مورد استفاده قرار گیرد و هدف تحقیق کاربردی، توسعه دانش کاربردی در یک زمینه خاص می‌باشد. مبانی نظری پژوهش از طریق مطالعات کتابخانه‌ای یعنی مطالعه مقالات وکتب منتشر شده به دست آمده و بر پایه آن فرضیه‌های پژوهش تدوین گردیده است. جهت برازش هر یک از مدل های خود رگرسیو میانگین متحرک انباشته ، ماشین بردار پشتیبان و مدل ترکیبی و مقایسه عملکرد مدل ها با یگدیگر از قیمت هفتگی نفت خام برنت از ابتدای سال ۲۰۰۶ تا سپتامبر ۲۰۱۲ برای پیش بینی متغیر مورد نظر تحقیق که قیمت هفتگی نفت خام برنت می باشد، استفاده شده است. با توجه به اینکه در انجام این تحقیق هدف را بر مبنای مقایسه عملکرد مدل ترکیبی نسبت به هر یک از مدل های خودرگرسیو میانگین متحرک انباشته وماشین بردار پشتیبان برای پیش بینی قیمت پایانی برای یک مرتبه پیش رو (one step ahead) قرار دادیم، بنابراین در مرحله اول با توجه به مطالب مطرح شده در بخش مبانی نظری از(328) مشاهده به منظور برازش هریک از مدل‌ها و پیش بینی مشاهده(۳۲۹ام) برای قیمت پایانی استفاده شده و در مرحله بعد مشاهده واقعی(۳۲۹ام)که مربوط به اولین هفته‌ای است که پیش‌بینی در مورد آن صورت گرفته است را به منظور برازش مجدد مدل به سری زمانی اضافه کرده و قیمت پایانی مرتبه (هفته) بعد را که مشاهده(۳۳۰ام) می‌باشد را پیش بینی می‌کنیم و مراحل فوق رابرای بیست مرتبه (هفته) تا پیش بینی مشاهده پایانی تکرار می‌کنیم. با توجه به پیش بینی های

انجام شده برای بیست هفته پیش‌رو معیارهای ارزیابی عملکرد ریشه میانگین مجذور خطـا (RMSE) و میـانگین قدر مطلق درصد خطا(MAPE) که نحوه محاسبه آن در زیر آمده است برای هـر هفتـه و بـرای هـر سـه مـدل محاسبه می‌شود. در مرحله پایانی با توجه به معیارهای ارزیابی عملکرد و بـا اسـتفاده از آزمـون مقایسـه زوجـی فرضیه‌های تحقیق را مورد آزمون قرار خواهیم داد.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{(i)} - F_{(i)})^2} \qquad (۱۸)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_{(i)} - F_{(i)}}{Y_{(i)}}\right| \times 100 \qquad (۱۹)$$

$Y_{(i)}$: مقدار واقعی متغیر

$F_{(i)}$: مقدار پیش‌بینی متغیر

n:تعداد اجزاء خارج نمونه‌ای

۵- نتایج پژوهش

در این بخش با توجه به مدل‌های کاربردی مطرح شده در مبانی نظری ، معیارهای ارزیابی عملکرد به همراه نتایج آزمون مقایسه زوجی برای بررسی فرضیه‌های پژوهش مورد بررسی قرار گرفته است.

جدول ۱- ریشه‌ی میانگین مربع خطا

مرتبه (هفته پیش رو)	RMSE(ریشه میانگین مربع خطا)		
	ARIMA	SVM	Hybrid
۱	0.2601	0. 5052	0.0656
۲	0.9000	0.9685	0.2259
۳	0.6118	0.5911	0.4156
۴	0.0586	0.3524	0.0449
۵	2.9778	2.8570	2.4986
۶	1.6346	1.4235	0.9985
۷	2.1659	3.2673	2.0678
۸	1.4192	1.8413	1.3162
۹	0.1966	0.5291	0.1693
۱۰	1.5749	1.8110	1.5250

مرتبه (هفته پیش رو)	RMSE(ریشه میانگین مربع خطا)		
	ARIMA	SVM	Hybrid
۱۱	0.1152	1.8759	0.0820
۱۲	0.8354	2.3406	0.2879
۱۳	0.1874	1.2011	0.6840
۱۴	2.5938	1.0010	0.8041
۱۵	1.7121	1.5631	1.2764
۱۶	0.8455	0.8291	0.8506
۱۷	1.6332	0.1436	1.2835
۱۸	2.3142	1.6021	2.3262
۱۹	0.8389	0.8458	0.7467
۲۰	0.1057	0.3062	0.1083
میانگین	1.1490	1.2928	0.8889

فرضیه اول:

H_0: بین دو جامعه ریشه میانگین مجذور خطا مدل ترکیبی و مدلARIMA اختلاف معناداری وجود ندارد.

$H_۱$: بین دو جامعه ریشه میانگین مجذور خطا مدل ترکیبی و مدلARIMA اختلاف معناداری وجود دارد.

فرضیه دوم:

H_0: بین دو جامعه ریشه میانگین مجذور خطا مدل ترکیبی و مدلSVM اختلاف معناداری وجود ندارد.

H_1: بین دو جامعه ریشه میانگین مجذور خطا مدل ترکیبی و مدل SVM اختلاف معناداری وجود دارد.

جدول ۲ - آزمون مقایسه زوجی ریشه میانگین مجذور خطا

RMSE				
	– ARIMA Hybrid		Hybrid – SVM	
Mean	Hybrid	ARIMA	Hybrid	SVM
	۰/۸۸۸۸	۱/۱۴۹۰	۰/۸۸۸۸	۱/۲۹۲۷
Observations	۲۰		۲۰	
Df	۱۹		۱۹	
t Stat	-۲,۵۶		-۲,۵۴	
P(T<=t) one-tail	۰,۰۰۹۴		۰,۰۰۹۷	
t Critical one-tail	۱,۷۲		۱,۷۲	
P(T<=t) two-tail	۰,۰۱۸۸		۰,۰۱۹۵	
t Critical two-tail	۲,۰۹۳		۲,۰۹۳	

با توجه به جدول فوق بین میانگین دو جامعه‌ی RMSE از مدل ترکیبی در مقایسه با مدل خودرگرسیو میانگین متحرک انباشته (ARIMA) و مدل ماشین بردار پشتیبان اختلاف معناداری وجود دارد و میانگین RMSE مدل ترکیبی در نمونه های مورد بررسی از دو مدل دیگر کمتر است. بنابراین، از نظر معیار ارزیابی RMSE، عملکرد مدل ترکیبی به طور معناداری در پیش‌بینی قیمت پایانی نفت خام از دو مدل خودرگرسیو میانگین متحرک انباشته و ماشین بردار پشتیبان بهتر عمل می‌کند.

جدول ۳- میانگین قدرمطلق درصد خطا

مرتبه (هفته پیش رو)	میانگین قدرمطلق درصد خطا (MAPE)		
	ARIMA	SVM	Hybrid
۱	0.002536	0.004926	0.000639
۲	0.008725	0.009389	0.00219
۳	0.005895	0.005696	0.004004
۴	0.000566	0.003405	0.000434
۵	0.031982	0.030685	0.026835
۶	0.017987	0.015663	0.010987
۷	0.024879	0.03753	0.023752
۸	0.016809	0.021809	0.01559
۹	0.00236	0.006355	0.002034
۱۰	0.019417	0.022328	0.018802
۱۱	0.001317	0.003816	0.00135
۱۲	0.001343	0.021869	0.000956
۱۳	0.009399	0.026334	0.003239
۱۴	0.002104	0.01348	0.007676
۱۵	0.027468	0.010601	0.008515
۱۶	0.017794	0.016245	0.013265
۱۷	0.008837	0.008665	0.00889
۱۸	0.017069	0.001501	0.013414
۱۹	0.023721	0.016422	0.023844
۲۰	0.009176	0.009252	0.008168
میانگین	0.012469	0.014299	0.009729

فرضیه سوم

H_0: بین دو جامعه میانگین قدرمطلق درصد خطا مدل ترکیبی و مدلARIMA اختلاف معناداری وجود ندارد.

H_1: بین دو جامعه خطای میانگین قدرمطلق درصد خطا مدل ترکیبی و مدل ARIMA اختلاف معناداری وجود دارد.

فرضیه چهارم

H0: بین دو جامعه میانگین قدرمطلق درصد خطا مدل ترکیبی و مدلSVM اختلاف معناداری وجود ندارد.

H1: بین دو جامعه میانگین قدرمطلق درصد خطا مدل ترکیبی و مدل SVM اختلاف معناداری وجود دارد.

با توجه به جدول فوق بین میانگین دو جامعه‌ی MAPE از مدل ترکیبی و مدل خودرگرسیو میانگین متحرک انباشته و دو جامعه‌ی میانگین قدرمطلق درصد خطا از مدل ترکیبی و مدل ماشین بردار پشتیبان اختلاف معنا-داری وجود دارد و میانگین MAPE برای مدل ترکیبی در نمونه‌ی مورد بررسی از دو مدل دیگر کمتر است. بنابراین، از نظر معیار ارزیابی میانگین قدر مطلق درصد خطا، مدل ترکیبی به طور معناداری در پیش‌بینی قیمت پایانی از دو مدل خودرگرسیو میانگین متحرک انباشته و ماشین بردار پشتیبان بهتر عمل می‌کند.

جدول ۴- آزمون مقایسه زوجی میانگین قدرمطلق درصد خطا

MAPE				
	Hybrid – ARIMA		Hybrid – SVM	
mean	Hybrid	ARIMA	Hybrid	SVM
	۰/۰۰۹۷	۰/۰۱۲۴۷	۰/۰۰۹۷	۰/۰۱۴۳۰
Observations	۲۰		۲۰	
Df	۱۹		۱۹	
t Stat	-۲,۵۹۰		-۲,۵۴۱	
P(T<=t) one-tail	۰,۰۰۸۹		۰,۰۰۹۹	
t Critical one-tail	۱,۷۲		۱,۷۲	
P(T<=t) two-tail	۰,۰۱۷۹		۰,۰۱۹۹	
t Critical two-tail	۲,۰۹۳		۲,۰۹۳	

۶- نتیجه‌گیری و بحث

با توجه به محاسبه معیارهای ارزیابی عملکرد ریشه میانگین مجذور خطا و میانگین قدرمطلق درصد خطا برای پیش بینی قیمت‌ها برای هر یک از سه مدل و انجام آزمون فرضیه‌ها، نتایج بیانگر این بودند که مدل ترکیبی در پیش بینی قیمت در مقایسه با مدل‌های خودرگرسیو میانگین متحرک انباشته و مدل ماشین بردار پشتیبان دارای خطای کمتری می‌باشد و فرضیه‌های تحقیق مبنی بر کمتر بودن خطای پیش بینی مدل ترکیبی

نسبت به مدل ARIMA و SVM رد نشدند .در مقایسه با نتیجـه گیـری‌هـای حاصـل شـده از ایـن پـژوهش بـا پژوهش‌های داخلی و خارجی انجام شده در زمینه کاربرد مدل‌های ترکیبی برای پـیش بینـی، دقـت مـدل‌هـای ترکیبی نسبت به کاربرد مجزای مدل های مورد استفاده برای پیش بینی مورد تایید قرار گرفت.

در تحقیق حاضر به مدل سازی الگوهای غیرخطی پسماندها از طریق ماشین بردار پشتیبان و اضافه کردن آنها به بخش پیش‌بینی خطی خودرگرسیو میانگین متحرک انباشته پرداختیم ولی برای بررسی کارایی و مقایسه نتـایج می‌توان این پسماندها را به مدل شبکه‌های عصبی از جمله شبکه عصبی پیش‌خورداد تـا دقـت آن را نسـبت بـه ماشین بردار پشتیبان مورد سنجش قرار داد و دقت پیش بینی ترکیب شـبکه عصـبی و ARIMA را نسـبت بـه کاربرد مجزای هر یک از مدل‌ها مورد بررسی قرار داد.

فهرست منابع

* راعی،رضا و فلاح‌پور،سعید(۱۳۸۷).کاربرد ماشین بردار پشتیبان در پیش بینی درماندگی مالی شرکت‌هـا بـا استفاده از نسبت های مالی. بررسی های حسابداری وحسابرسی،۴، ۳۴-۱۷.
* رمضانی،رضا(۱۳۸۳). مقایسه عملکرد پیش‌بینی‌در مدل‌های شبکه عصبی مصنوعی و ARIMA(موردمطالعه : تقاضای ماهیانه برق در ایران)،پایان نامه کارشناسی ارشد، دانشکده مدیریت دانشگاه تهران.
* زراءنژاد،منصور و حمید، شهرام(۱۳۸۸). پیش بینی نرخ تورم در اقتصاد ایران با استفاده از شبکه های عصبی پویا(دیدگاه سری زمانی)،فصلنامه اقتصاد مقداری.۱،۶۷-۹۸.
* سوری، علی(۱۳۹۱).اقتصاد سنجی همراه با کاربردEviews7،تهران:انتشارات چاپ الغدیر.
* صادقی، مهدی و شوال‌پور، سعید(۱۳۸۶). اقتصادسنجی سری‌های زمانی با رویکرد کاربردی،تهران: انتشارات دانشگاه امام صادق(ع).
* صادقی،آ(۱۳۸۹).پیش بینی شاخص قیمت سهام در بـورس اوراق بهـادار تهـران بـا رویکـرد ماشـین بـردار پشتیبان. پایان‌نامه کارشناسی ارشد، دانشکده مدیریت دانشگاه تهران.
* طبسی،م(۱۳۹۱). برآورد ارزش در معرض ریسک با استفاده از مدل ترکیبی ماشین بـردار پشـتیبان وگـارچ. پایان‌نامه کارشناسی‌ارشد،دانشکده مدیریت دانشگاه تهران.
* منجمی، سید امیرحسین، ابرزی، مهدی و شوازی، علیرضا(۱۳۸۸). پیش بینی قیمت سـهام در بـورس اوراق بهادار با استفاده از شبکه عصبی فازی و الگوریتم‌های ژنتیک ومقایسـه‌ی آن بـا شـبکه‌ی عصـبی مصـنوعی. فصل نامه اقتصاد مقداری،۲۲، ۲۶-۱.
* نوری، روح اله، خاکپو، امیر،دهقانی، مجید وفرخ نیا، اشکان (۱۳۹۰). پیش بینی ماهانه جریان آب با استفاده از ماشین بردار پشتیبان بر مبنای آنالیز مولفه اصلی. فصـلنامه علمـی وپژوهشـی آب وفاضـلاب، ۷۷، ۱۱۸-۱۲۳.

* Abe,S.(2005)support vector machines for pattern classification, springer, londen.
* Brooks,C (2008).Introductory Econometrics for Finance, UK Cambridge university press.
* Enders,W(1995). Applied Econometric Time Series.

* Gavrishchka,v & Banerjee, s (2006).Support vector machines an efficient for stock market volatility forecasting, springer, 3, 147-160
* Gunn, S (1998).Support Vector Machines for Classification and Regression, university of south Ampton.147-165.
* Gok, T (2002).comparative study of neural networks and Box-Jenkins ARIMA modeling in time series prediction, Computer & Industrial Engineering jurnal, 371-375.
* Hongzhan, N, Guohui, L, Xiaoman & Yong ,W (2012).Hybrid of ARIMA and SVMs for Short-Term Load Forecasting, 16,1455-1460.
* Huang, Z, Chen,h, jung, c, hwa, w & soushan, w(2004).credit rating analysis with support vector machines and neural network a market comparative study, Decision support system,37,543-558.
* Lipo,W(2005).Support Vector Machines:Theory and Applications, Springer.2,38-96.
* McKenzie, E(1988)A note on using the integrated form of ARIMA forecasts, International Journal of Forecasting,4, 117-134.
* PackD, J (1990). In defense of ARIMA modeling, International Journal of Forecasting, 6, 20-36.
* Ping, P & Chin, L (2005).a hybrid ARIMA and support vector machines model in stock price forecasting, omega,33,497-505.
* Sanchez, D (2003).Advanced Support Vector Machines and Kernel methods, Nero computing,48, 5-20.
* Tsay, R (2002). Analysis of Financial Time Series, John Wiley & sons.2,70-89.
* Vapnik,V(1995).The nature of statistical learning theory, Springer.2,12-68

یادداشت‌ها

1 - Autoregressive Integrated Moving Average
2 - Support Vector Machine
3 - hybrid model
4 - stationary
5- Autoregressive Integrated Moving Average
6 -Box-Jenkins
7- vapnik
8 -Huber
9 - Autoregressive conditional heteroskedasticity
10 - Generalized autoregressive conditional heteroskedasticity
11 - Threshold generalized autoregressive conditional heteroskedasticity
12-Auto Correlation Function
13-Partial Auto Correlation Function
14- Vapnik
15- Structural Risk Minimization
16- Empirical Risk Minimization
17- Gunn
18- Abe
19-Loss function
20- Hongzhan
21 - Gavrishchka & Banerjee
22 - Radial Basis Function
23 - Ping & Chin
24 - Yu, Vanng & Lay
25 - Gok
26 - Principal component analysis