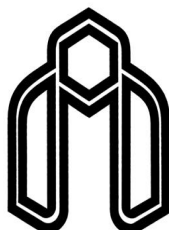


بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه صنعتی شاهرود

دانشکده علوم ریاضی

گروه آمار

پایان نامه کارشناسی ارشد

تعیین نقطه تغییر در واریانس فرآیندهای چندمتغیره توسط طرح ترکیبی دو مرحله‌ای

امیرهاشمی

استاد راهنما

داود شاهسونی

آذر ۱۳۹۴

ادایه

تقدیم به مادر مهربانم برای همه‌ی دعا‌های خیرش، کسی که پری وار به من
معنای واژه صبر را آموخت.

تقدیم به پدر بزرگوارم برای همه‌ی حمایت‌های بی‌دریغش.

تقدیم به همسر مهربانم، کسی که به من معنای واژه امید را آموخت.

تعمدنامه

اینجانب امیر هاشمی دانشجوی کارشناسی ارشد رشته آمار ریاضی دانشکده علوم ریاضی دانشگاه شاهرود، نویسنده پایان‌نامه با عنوان تعیین نقطه تغییر در واریانس فرآیندهای چندمتغیره توسط طرح ترکیبی دو مرحله‌ای، تحت راهنمایی داود شاهسونی متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان‌نامه، تاکنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه شاهرود متعلق دارد، و مقالات مستخرج با نام “ دانشگاه شاهرود “ یا “ Shahrood University “ به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده) شده است، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

امیر هاشمی
آذر ۱۳۹۴

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه شاهرود می‌باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان‌نامه بدون ذکر منبع مجاز نمی‌باشد.

چکیده

شناسایی موثر نقطه تغییر در یک فرآیند چندمتغیره یکی از مسائل مهمی است که محققان با آن روبرو هستند. اکثر مطالعات موجود برای این مهم، توسط روش درستنمایی ماکسیمم یا روش یادگیری ماشین انجام گرفته است. مشکلی که در روش درستنمایی ماکسیمم وجود دارد، فرض مشخص بودن توزیع فرآیند می باشد که معمولاً در عمل، اطلاعاتی در مورد توزیع فرآیند نداریم؛ همچنین روش یادگیری ماشین نیز هنگامی که تعداد متغیرهای ورودی زیاد باشد، هزینه محاسباتی زیادی خواهد داشت. در این مطالعه، با در نظر گرفتن دو گونه مختلف از ترکیب روش های مدل سازی آماری و یادگیری ماشین، سعی داریم که مشکل مذکور را تا حدی مرتفع کنیم. گونه اول، ترکیب روش رگرسیون لجستیک با روش ماشین بردار پشتیبان است؛ به طوری که بعد متغیرهای ورودی، ابتدا توسط روش رگرسیون لجستیک کاهش داده شده و این مجموعه کاهش یافته به عنوان ورودی روش ماشین بردار پشتیبان در نظر گرفته می شوند. سرانجام با استفاده از خروجی روش ماشین بردار پشتیبان، رهیافت شناسایی نقطه تغییر در فرآیندهای چندمتغیره مطرح می شود. در گونه دوم، روش رگرسیون چندمتغیره سازوار اسپلاین، جایگزین روش رگرسیون لجستیک شده و بقیه مراحل عیناً اجرا می شود. نتایج شبیه سازی حاکی از آن است که مدل ترکیبی مطرح شده به طور موثری قادر به شناسایی نقطه تغییر در فرآیندهای چندمتغیره می باشد و این مدل عملکرد بهتری نسبت به نمودار کنترل شوهارت و همچنین مدل تکی ماشین بردار پشتیبان دارد.

کلمات کلیدی: نقطه تغییر، فرآیندهای چندمتغیره، طرح ترکیبی دو مرحله ای، رگرسیون لجستیک، رگرسیون

چندمتغیره سازوار اسپلاین، ماشین بردار پشتیبان.

لیست مقالات مستخرج از پایان نامه

۱. هاشمی ا.، شاهسونی د. (۱۳۹۴)، ”تعیین نقطه تغییر در واریانس فرآیندهای چندمتغیره توسط ترکیب روش‌های آماری و یادگیری ماشین”، نخستین کنفرانس ملی محاسبات نرم، دانشگاه گیلان، گیلان، ایران.

فهرست مطالب

خ	فهرست تصاویر
۱	فهرست جداول
۳	۱ مقدمه و الگوریتم کلی تحقیق
۳	۱.۱ مقدمه
۴	۲.۱ زمینه‌های بالقوه تغییر
۵	۱.۲.۱ انواع تغییرات در پارامتر فرآیند
۱۰	۳.۱ بیان مسئله
۱۱	۴.۱ تاریخچه تحقیق
۱۲	۵.۱ ضرورت انجام تحقیق
۱۵	۲ روش‌های رده‌بندی مورد بحث
۱۵	۱.۲ رگرسیون لجستیک چندمتغیره
۱۷	۱.۱.۲ برآورد پارامترها
۲۰	۲.۱.۲ انتخاب متغیرهای مهم
۲۱	۲.۲ رگرسیون چندمتغیره سازوار اسپلاین
۲۳	۱.۲.۲ انتخاب توابع پایه‌ای مهم
۲۴	۳.۲ ماشین بردار پشتیبان
۲۴	۱.۳.۲ ماشین بردار پشتیبان خطی:
۳۰	۲.۳.۲ ماشین بردار پشتیبان در سیستم‌های خطی جداناپذیر
۳۲	۳.۳.۲ ماشین بردار پشتیبان در سیستم‌های غیرخطی
۳۳	۴.۳.۲ هسته
۳۷	۳ روش شناسی
۳۷	۱.۳ ساختار فرآیند و تغییر در واریانس
۳۹	۲.۳ نمودار کنترل $ S $
۴۱	۱.۲.۳ هشدار غلط

۴۲	تعریف متغیرهای جدید Z و Y	۲.۲.۳
۴۴	راهبرد شناسایی نقطه تغییر	۳.۳
۵۱		مطالعات شبیه‌سازی	۴
۵۲	مدل SVM بهینه	۱.۴
۵۳	تعیین حد آستانه	۲.۴
۵۳	تعیین زمان هشدار نمودار کنترل $ S $	۳.۴
۵۶	رهیافت SVM	۴.۴
۵۷	رهیافت ترکیبی $LR - SVM$	۵.۴
۵۷	رهیافت ترکیبی $MARS - SVM$	۶.۴
۵۷	تعیین متغیرهای مهم	۷.۴
۵۸	تعیین نقطه تغییر	۸.۴
۶۱	نتیجه‌گیری	۹.۴
۶۳	آینده تحقیق	۱.۹.۴
۶۵		آ ارزیابی همخطی چندگانه	
۶۷		ب روش نیوتن-رافسون	
۶۹		پ آماره آزمون نمره و والد	
۷۱		ت قضیه تصویر	
۷۳		ث شرایط KKT	
۷۵		ج کدهای نرم‌افزار R	
۷۵	شبیه‌سازی مجموعه داده‌های آموزشی، آزمون، حد آستانه و تعیین متغیرهای مهم	۱.ج
۷۹	بهینه‌سازی پارامترهای مدل SVM	۲.ج
۸۳	تعیین حد آستانه	۳.ج
۸۸	تعیین نقطه تغییر	۴.ج
۹۵		مراجع	
۹۹		واژه‌نامه فارسی به انگلیسی	
۱۰۱		واژه‌نامه انگلیسی به فارسی	
۱۰۲		نمایه	

فهرست تصاویر

۴	پایش یک فرآیند چندمتغیره توسط نمودار کنترل S	۱.۱
۵	سه نوع تغییر بالقوه در فرآیند (پیزدک، ۲۰۰۳)	۲.۱
۵	تاخیر نمودار کنترل در اعلام زمان تغییر فرآیند (فهمی و ال سید، ۲۰۰۶)	۳.۱
۶	نمایش شماتیکی تغییر پله‌ای (آتشگر، ۱۳۹۳)	۴.۱
۶	نمودار کنترل شوهارت حاوی تغییر پله‌ای یگانه در میانگین فرآیند (امیری و اللهیاری، ۲۰۱۲)	۵.۱
۷	نمایش شماتیکی تغییر پله‌ای چندگانه در پارامتر فرآیند (آتشگر، ۱۳۹۳)	۶.۱
۷	نمودار کنترل شوهارت حاوی تغییر پله‌ای چندگانه در میانگین فرآیند (امیری و اللهیاری، ۲۰۱۲)	۷.۱
۸	نمایش شماتیکی تغییر با روند خطی در پارامتر فرآیند (آتشگر، ۱۳۹۳)	۸.۱
۸	نمودار کنترل شوهارت حاوی تغییر با روند خطی در میانگین فرآیند (امیری و اللهیاری، ۲۰۱۲)	۹.۱
۹	تغییر هم‌نوا در پارامتر فرآیند (آتشگر، ۱۳۹۳)	۱۰.۱
۱۰	نمودار کنترل شوهارت حاوی تغییر هم‌نوا در میانگین فرآیند (امیری و اللهیاری، ۲۰۱۲)	۱۱.۱
۱۶	نمودار پراکنش متغیر پاسخ در مقابل متغیر سن (هاسمر و همکاران، ۲۰۰۴)	۱.۲
۱۶	نمودار پراکنش متغیر پاسخ در مقابل متغیر گروه سنی (هاسمر و همکاران، ۲۰۰۴)	۲.۲
۲۱	انتخاب متغیرهای مهم به روش رویه پیشرو گام به گام والد	۳.۲
۲۵	داده‌های آموزشی که به صورت خطی جدا پذیرند	۴.۲
۲۵	تعدادی از ابرصفحه‌هایی که رده‌بندی را انجام داده‌اند	۵.۲
۲۶	اب‌صفحه‌های مختلف همراه ابرصفحه‌های حاشیه‌ای	۶.۲
۲۷	معادلات ابرصفحه‌های حاشیه‌ای ومجاز‌ساز بهینه	۷.۲
۲۷	بیشترین فاصله بین ابرصفحه‌های حاشیه‌ای	۸.۲
۳۰	بردارهای پشتیبان رده ۱ و ۰ بر روی ابرصفحه‌های حاشیه‌ای	۹.۲
۳۰	سیستم‌های خطی جداناپذیر با میزان خطا ϵ	۱۰.۲
۳۳	انتقال داده‌ها از فضای ورودی (سمت چپ) به فضای ویژگی (سمت راست)	۱۱.۲
۴۲	نمودار کنترل S	۱.۳
۴۲	هشدارهای غلط اعلام شده توسط نمودار کنترل	۲.۳

۴۸	نمودار کنترل $ S $ همراه نقطه تغییر واقعی و تعیین مدل سازی SVM	۳.۳
۴۸	رهیافت های ترکیبی $LR - SVM$ و $MARS - SVM$	۴.۳
۵۴	نمودار کنترل $ S $ برای اولین مجموعه داده که اولین هشدار در زمان $T = ۲۱۴$ رخ داده است.	۱.۴
۵۴	نمودار کنترل $ S $ برای دومین مجموعه داده که اولین هشدار در زمان $T = ۲۲۲$ رخ داده است.	۲.۴
۵۴	نمودار کنترل $ S $ برای سومین مجموعه داده که اولین هشدار در زمان $T = ۲۱۰$ رخ داده است.	۳.۴
۵۵	نمودار کنترل $ S $ برای چهارمین مجموعه داده که اولین هشدار در زمان $T = ۲۳۶$ رخ داده است.	۴.۴
۵۵	نمودار آریبی برآورد نقاط تغییر به ازای مقادیر مختلف δ	۵.۴
۶۲	نمودار MSE برآورد نقاط تغییر به ازای مقادیر مختلف δ	۶.۴
۷۱	تصویر بردار a روی بردار b	۱.ت

فهرست جداول

۱.۲	توابع هسته، $K(\underline{Z}_i, \underline{Z})$ ، پارامتر مقایس و مقادیر مثبت را اختیار می‌کند، $a, b \geq$	۳۴
۱.۳	جدول مشاهدات متغیر جدید Z	۴۳
۲.۳	تعیین نقطه تغییر فرآیند توسط خروجی مدل رده‌بندی	۴۵
۳.۳	محاسبه احتمال تجمعی توزیع دو جمله‌ای برای زمان‌های 1 تا $T - 1$	۴۷
۱.۴	حد آستانه به‌ازای تغییرات δ و رهیافت‌های SVM ، $LR - SVM$ و $MARS$	۵۳
۲.۴	برآورد نقطه تغییر توسط رهیافت نمودار کنترل $ S $ برای 10° مجموعه داده، همراه میانگین و خطای معیار آن به‌ازای تغییرات δ	۵۶
۳.۴	دقت رده‌بندی و متغیرهای معنی‌دار (با اهمیت) مدل LR به‌ازای مقادیر مختلف δ	۵۷
۴.۴	متغیرهای معنی‌دار (مهم) مدل $MARS$ به‌ازای مقادیر مختلف δ	۵۸
۵.۴	خروجی رهیافت ترکیبی $LR - SVM$ برای هفتمین مجموعه داده به‌ازای $\delta = 1/15$	۵۹
۶.۴	برآورد نقاط تغییر توسط رهیافت SVM برای 10° مجموعه داده آزمون همراه با تغییرات δ	۶۰
۷.۴	برآورد نقاط تغییر توسط رهیافت ترکیبی $LR - SVM$ برای 10° مجموعه داده آزمون همراه با تغییرات δ	۶۰
۸.۴	برآورد نقاط تغییر توسط رهیافت ترکیبی $MARS - SVM$ برای 10° مجموعه داده آزمون همراه با تغییرات δ	۶۱
۹.۴	میانگین و خطای معیار (داخل پرانتز) برآورد نقاط تغییر توسط رهیافت‌های نمودار کنترل $ S $ ، SVM ، ترکیبی $LR - SVM$ و ترکیبی $MARS - SVM$ به‌ازای مقادیر مختلف δ	۶۱

فصل ۱

مقدمه و الگوریتم کلی تحقیق

۱.۱ مقدمه

یکی از عوامل مهم خرید یک محصول، کیفیت آن است؛ بنابراین، تولید محصولی با کیفیت، عامل کلیدی برای رسیدن به موفقیت در دنیای تجارت است. این محصول می‌تواند اتومبیل، کامپیوتر، پوشاک یا هر نوع محصولی باشد که دارای خط تولید است و روزانه تیراژ بالایی از آن تولید می‌شود. در این حین، ممکن است اختلالی در خط تولید نفوذ کند و باعث تولید محصولی نامرغوب شود. برای تولید یک محصول با کیفیت، اختلال به وجود آمده باید هرچه سریع‌تر کشف و از فرآیند حذف گردد.

یکی از عوامل بهبود کیفیت، کنترل فرآیند آماری^۱ (SPC) تولید محصول می‌باشد. متداول‌ترین ابزار SPC، نمودار کنترل است که در سال ۱۹۳۱ توسط شوهارت^۲ معرفی شده است. این ابزار، بیش از ۸۰ سال در عرصه بهبود کیفیت کالا، موثر واقع شده است.

نمودارهای شوهارت، تغییر یا اختلال را در فرآیندهای یک‌متغیره یا چندمتغیره تشخیص می‌دهند. برای مثال، در فرآیندهای یک‌متغیره از نمودار کنترل \bar{X} برای شناسایی اختلال در میانگین فرآیند و از نمودارهای کنترل S و R برای شناسایی اختلال در واریانس فرآیند استفاده می‌شود. در فرآیندهای چندمتغیره نیز از نمودارهای کنترل T^2 هتلینگ^۳ و $|S|$ به ترتیب به منظور پایش فرآیند و کشف زمان تغییر در بردار میانگین و در ساختار کوواریانس (واریانس) استفاده می‌شود. این نمودارها، دارای دو حد کنترلی بالایی^۴ (UCL) و پایینی^۵ (LCL) و یک خط میانی^۶ (CL) هستند که در آن، مقادیر مربوط به شاخص اندازه‌گیری، نسبت به زمان رسم می‌شوند. معیار تصمیم‌گیری در این نمودارها به این‌گونه است که اگر نقاط، مابین UCL و LCL قرار گیرند، آن‌گاه فرآیند تحت کنترل می‌باشد؛ در غیر این صورت

^۱ Statistical Process Control

^۲ Shewhart

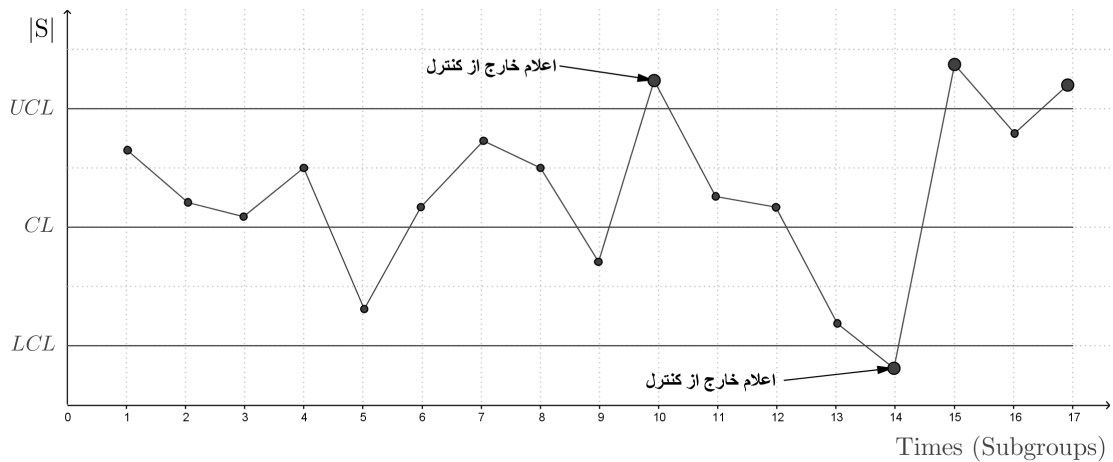
^۳ Hotelling T^2

^۴ Upper control limit

^۵ Lower control limit

^۶ Center line

فرآیند خارج از کنترل خواهد بود. شکل ۱.۱ نمونه‌ای از نمودار کنترل $|S|$ را نشان می‌دهد.



شکل ۱.۱: پایش یک فرآیند چندمتغیره توسط نمودار کنترل $|S|$

تغییر در فرآیند، ممکن است در پارامترهای مختلف ایجاد شود؛ لذا، زمینه‌های بالقوه تغییر (رسولی، ۱۳۹۳) را مطرح می‌کنیم.

۲.۱ زمینه‌های بالقوه تغییر

فرض کنید توزیع متغیر تصادفی X در یک فرآیند چندمتغیره وابسته به پارامتر θ باشد. اگر تابع احتمال X ، قبل از تغییر به صورت $f_0(x, \theta_0)$ و پس از تغییر به صورت $f_1(x, \theta_1)$ نمادگذاری شود، آنگاه تغییر ایجاد شده در فرآیند، ممکن است به دو گونه زیر باشد (آتشگر، ۱۳۹۳):

۱. تابع احتمال فرآیند، پس از نقطه تغییر، ثابت باقی می‌ماند ولی پارامتر یا پارامترهای فرآیند دچار تغییر می‌شوند؛ به طوری که $f_1 = f_0$ و $\theta_1 \neq \theta_0$.

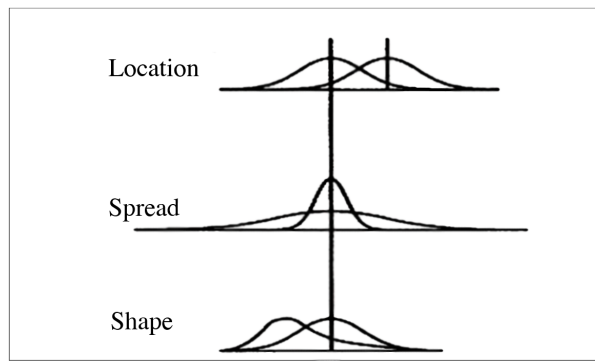
۲. تابع احتمال فرآیند پس از نقطه تغییر، تغییر می‌کند و پارامترهای فرآیند با توجه به توزیع آماری جدید، تعیین می‌شوند؛ به طوری که $f_1 \neq f_0$.

در هر توزیع آماری سه پارامتر کلیدی و مهم وجود دارد (پیزدک، ۲۰۰۳): مکانی^۱، پراکندگی^۲ و شکل^۳ توزیع. این پارامترها که می‌توان آن‌ها را به عنوان زمینه‌های بالقوه تغییر فرآیند در نظر گرفت، در شکل ۲.۱ آمده‌اند.

^۱Location

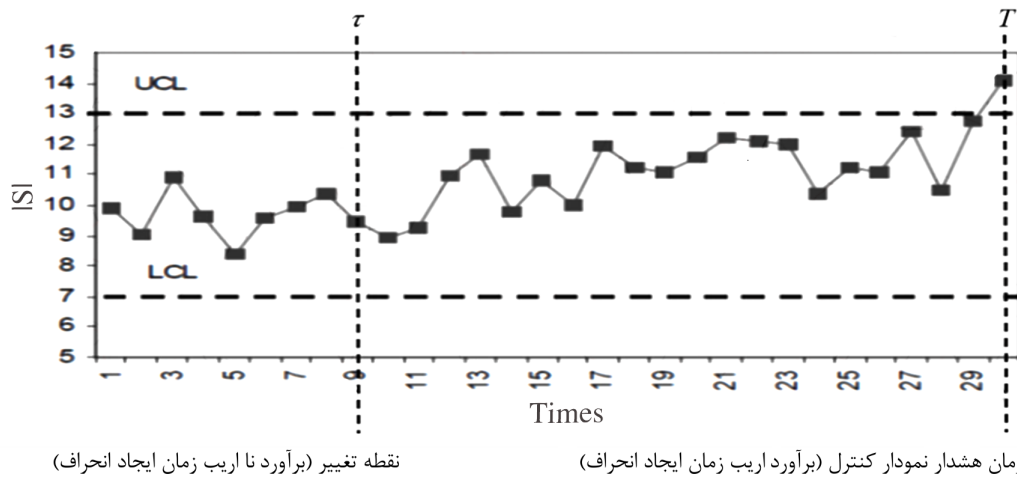
^۲Spread

^۳Shape



شکل ۲.۱: سه نوع تغییر بالقوه در فرآیند (پیزدک، ۲۰۰۳)

با توجه به توضیحات ذکر شده، می‌توان تغییر در فرآیند را به یکی از مشخصه‌های اصلی فرآیند (مکانی، پراکندگی و شکل) نسبت داد و نوع ایجاد تغییرات را به صورت تغییر در پارامتر یا توزیع فرآیند در نظر گرفت. همچنین زمان ایجاد تغییر را نیز می‌توان از طریق پایش فرآیند به وسیله نمودارهای کنترل و برآوردکننده‌های نارایب نقطه تغییر، برآورد کرد. نمونه‌ای از برآوردکننده اریب (زمان هشدار) و برآوردکننده نارایب (نقطه تغییر) زمان ایجاد انحرافات بادلیل در نمودارهای کنترل، در شکل ۳.۱ نشان داده شده است.



نقطه تغییر (برآورد نارایب زمان ایجاد انحراف)

زمان هشدار نمودار کنترل (برآورد اریب زمان ایجاد انحراف)

شکل ۳.۱: تاخیر نمودار کنترل در اعلام زمان تغییر فرآیند (فهمی و ال سید، ۲۰۰۶)

۱.۲.۱ انواع تغییرات در پارامتر فرآیند

با توجه به شرایطی که بر فرآیند تأثیر می‌گذارند، تغییر ممکن است با الگوهای مختلفی در پارامتر فرآیند رخ دهد. در اینجا انواع تغییر در فرآیند را دسته‌بندی و تشریح می‌کنیم.

تغییر پله‌ای یگانه

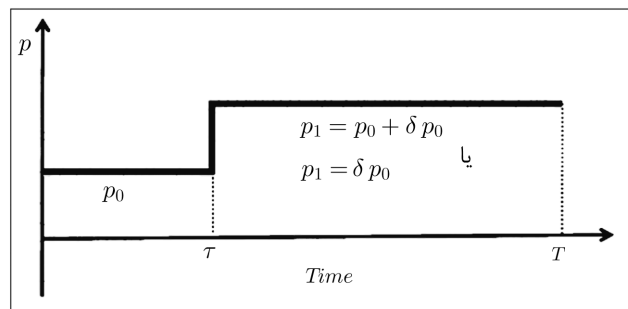
تغییر پله‌ای یگانه^۱ به نوعی از تغییر گویند که پارامتر فرآیند تحت کنترل، در نقطه نامعلومی از زمان، به میزان نامشخصی تغییر کرده و به حالت خارج از کنترل در آید. پارامتر تا زمانی در همین سطح جدید باقی می‌ماند که علل ایجاد انحراف شناسایی و حذف گردند. این‌گونه تغییر می‌تواند عللی نظیر شکستن ابزار، تعویض اپراتور، تغییر در تأمین‌کننده قطعات و علل مشابه داشته باشد. می‌توان رابطه این تغییر را به صورت زیر نشان داد.

$$p_1 = p_0 + \delta p_0, \quad (1.1)$$

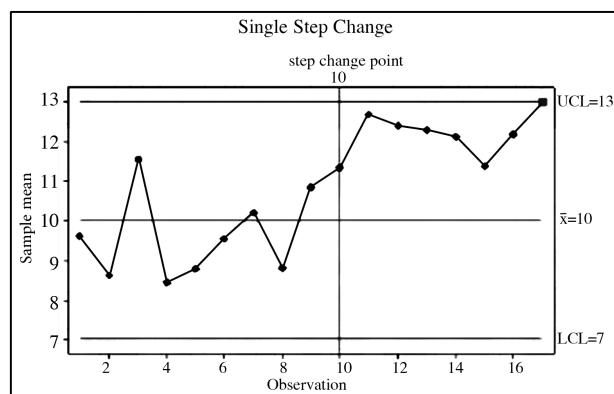
$$p_1 = \delta p_0, \quad (2.1)$$

که در آن δ بیانگر اندازه تغییر، p_0 و p_1 به ترتیب معرف میانگین یا واریانس فرآیند، وقتی که فرآیند تحت کنترل و خارج از کنترل باشد، هستند.

تغییر از نوع (۱.۱) در پارامتر مکانی و از نوع (۲.۱) در پارامتر پراکندگی فرآیند صورت می‌گیرد. نمایش شماتیک این تغییر در شکل ۴.۱ و نمونه عملی از آن در شکل ۵.۱ نشان داده شده است.



شکل ۴.۱: نمایش شماتیکی تغییر پله‌ای (آتشگر، ۱۳۹۳)



شکل ۵.۱: نمودار کنترل شوهارت حاوی تغییر پله‌ای یگانه در میانگین فرآیند (امیری و الله‌یاری، ۲۰۱۲)

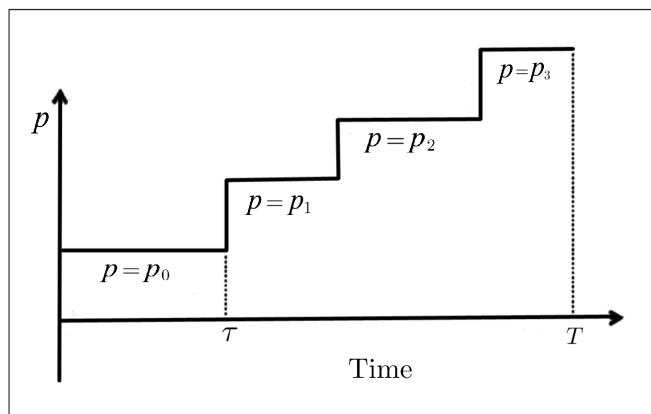
همان‌طور که مشخص است، دامنه تغییرات نمودار به سطح بالاتری افزایش یافته است.

^۱Single step change

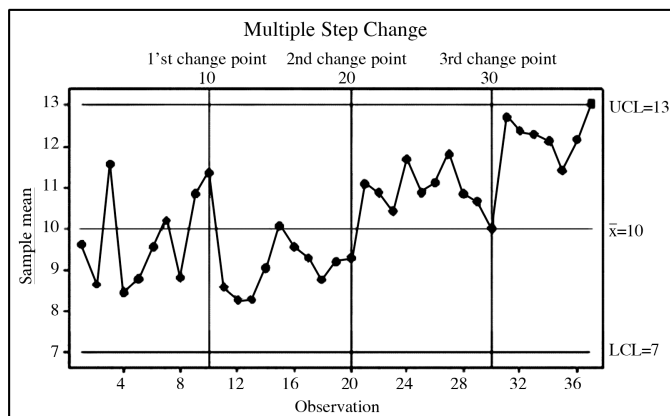
تغییر پله‌ای چندگانه

در تغییر پله‌ای چندگانه^۱، قبل از اعلام هشدار توسط نمودار کنترل، مبنی بر خارج از کنترل بودن فرآیند، سطح پارامتر در چند نقطه زمانی و در هر بار به میزان نامعلومی تغییر می‌کند. این نوع تغییر می‌تواند به دلیل تغییر یک یا چند متغیر موثر فرآیند در زمان‌های مختلف باشد. در این جا لازم است که تعداد تغییرات رخ داده و نیز زمان هریک از آنها برآورد شود. نمودار شماتیک این نوع تغییر در شکل ۶.۱ و نمونه عملی از این نوع تغییر، در شکل ۷.۱ نشان داده شده است.

همان‌طور که ملاحظه می‌شود، دامنه تغییرات زمان ۱۰ تا ۲۰ در سطح پایین‌تری (دارای میانگین کمتری) از سایر رده‌های زمانی دیگر واقع شده است و این دامنه در زمان‌های ۳۰ تا ۳۸ به بالاترین سطح خود رسیده است.



شکل ۶.۱: نمایش شماتیکی تغییر پله‌ای چندگانه در پارامتر فرآیند (آتشگر، ۱۳۹۳)

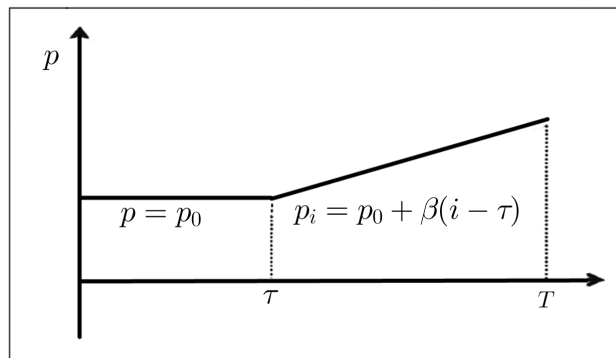


شکل ۷.۱: نمودار کنترل شوهارت حاوی تغییر پله‌ای چندگانه در میانگین فرآیند (امیری و الهیاری، ۲۰۱۲)

^۱Multiple step change

تغییر با روند خطی

در عمل، پارامترهای یک فرآیند ممکن است به تدریج و با شیب^۱ نامشخص تغییر کنند (شکل ۸.۱ را ببینید). در تغییر پارامتر با روند خطی^۲، پارامتر فرآیند از لحظه‌ای نامعین شروع به تغییر کرده و تغییر مطابق با یک معادله خطی وابسته به زمان، تا اعلام هشدار توسط نمودار کنترل ادامه می‌یابد. در این حالت، زمان شروع تغییر و نیز شیب خط تغییر نامعلوم بوده و باید با روش مناسبی برآورد شوند. معمولاً این نوع تغییر در اثر عللی نظیر فرسودگی تدریجی ابزار، خستگی اپراتور و سایر علل مشابه اتفاق می‌افتد.

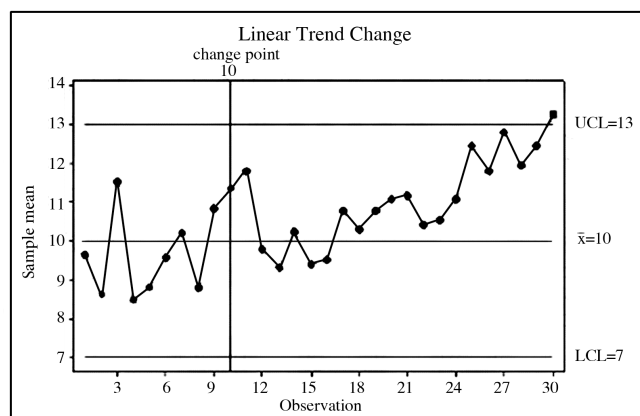


شکل ۸.۱: نمایش شماتیکی تغییر با روند خطی در پارامتر فرآیند (آتشگر، ۱۳۹۳)

این نوع تغییر را می‌توان به صورت زیر بیان کرد:

$$p_i = p_0 + \beta(i - \tau)$$

که در آن، β شیب تغییر، p_i مقدار پارامتر در زمان i و τ نقطه تغییر واقعی فرآیند را نشان می‌دهد. نمونه عملی از این نوع تغییر در شکل ۹.۱ آمده است.



شکل ۹.۱: نمودار کنترل شوهارت حاوی تغییر با روند خطی در میانگین فرآیند (امیری و الهیاری، ۲۰۱۲)

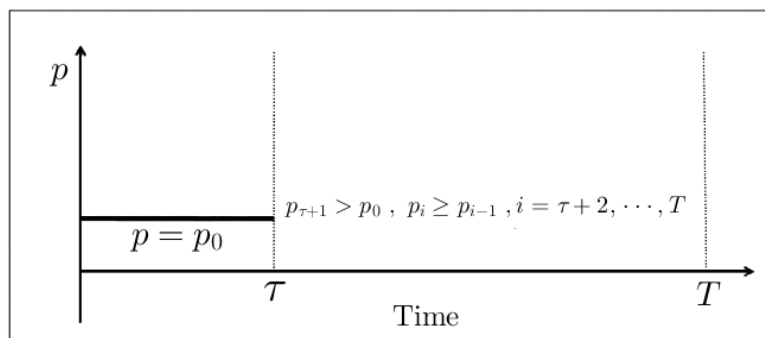
^۱ Slope

^۲ Linear trend

تغییر یکنوا

هر سه نوع تغییر بررسی شده مذکور، به گونه‌ای بودند که نوع آنها معین در نظر گرفته شده است. در حالی که، در واقعیت امر، معمولاً حتی نوع تغییر نیز برای ما شناخته شده نیست و باید بدون دانستن نوع تغییر، به برآورد نقطه تغییر بپردازیم. تغییرات یکنوا^۱، آن دسته از تغییراتی هستند که نحوه تأثیر آنها بر پارامتر فرآیند از قبل برای ما معلوم نیست ولی راستای تأثیر آنها مشخص است. یعنی می‌دانیم که پارامتر فرآیند قرار است افزایش یا کاهش یابد. در حقیقت یک تغییر یکنوا می‌تواند به صورت تغییر پله‌ای غیرکاهشی (غیرافزایشی)، تغییر پله‌ای چندگانه غیرکاهشی (غیرافزایشی)، تغییر با روند خطی یا غیرخطی غیرکاهشی (غیرافزایشی) و یا ترکیبی از سایر انواع تغییرات یکنوا باشد.

تغییرات یکنوا مجموعه‌ای از تغییرات هم‌نوا^۲ و ناهم‌نوا^۳ می‌باشند. زمانی که پارامتر فرآیند به صورت غیرکاهشی تغییر کند، به این نوع تغییر، هم‌نوا گوئیم. تغییرات هم‌نوا می‌توانند بر اثر تغییرات چند متغیر مؤثر فرآیند، هر یک به صورت‌های مختلف و به شکل افزایشی ایجاد شوند. شکل ۱۰.۱ تغییر هم‌نوا در پارامتر فرآیند را به صورت شهودی نشان می‌دهد. همان‌طور که ملاحظه می‌شود از زمان $\tau + 1$ تا زمان T ، رفتار فرآیند نامعلوم و سطح پارامتر فرآیند افزایش یافته است. همچنین نمونه‌ای از این نوع تغییر در نمودار کنترل شوهارت در شکل ۱۱.۱ نشان داده شده است.



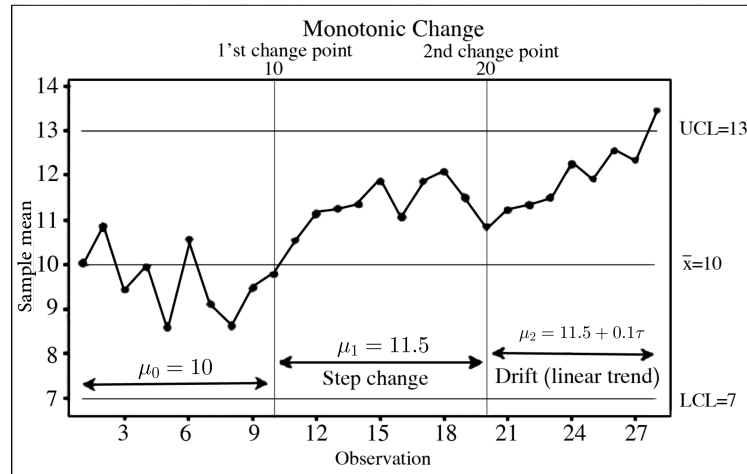
شکل ۱۰.۱: تغییر هم‌نوا در پارامتر فرآیند (آتشگر، ۱۳۹۳)

نوع دیگری از تغییرات یکنوا حالتی است که پارامترهای فرآیند به صورت غیرافزایشی تغییر کنند. به این نوع تغییر، تغییر ناهم‌نوا گوئیم. تغییرات ناهم‌نوا حالت متقابل تغییرات هم‌نوا هستند و می‌توانند بر اثر تغییرات چند متغیر مؤثر فرآیند، هر یک به صورت‌های مختلف و به شکل کاهش ایجاد شوند.

^۱ Monotonic change

^۲ Isotonic

^۳ Antitonic



شکل ۱۱.۱: نمودار کنترل شوهارت حاوی تغییر همنا در میانگین فرآیند (امیری و الهیاری، ۲۰۱۲)

۳.۱ بیان مسئله

زمانی که یک نمودار کنترل، هشدار خارج از کنترل می‌دهد، این هشدار فقط بیان‌گر وجود انحرافات بادلیل در فرآیند است؛ اما زمان واقعی رخداد این انحرافات ممکن است قبل از زمان هشدار باشند که آن را نقطه تغییر^۱ می‌نامند. برآورد نقطه تغییر^۲، جستجو برای عوامل ایجادکننده خطاهای بادلیل را کاهش داده و در نتیجه، مهندسان با محدود کردن عوامل تغییر در فرآیند و به‌کار بردن روش‌های مختلف، می‌توانند علل ایجاد تغییر را شناسایی کنند. در نمودارهای کنترلی که نسبت به تغییر حساس‌تر هستند، زمان هشدار^۳ به نقطه تغییر نزدیک‌تر و در آنهایی که از حساسیت کمتری برخوردارند، فاصله زمان هشدار از نقطه تغییر بیشتر است. از آنجا که زمان هشدار نمودارهای کنترل با نقطه تغییر واقعی فرآیند فاصله دارد، می‌توان زمان هشدار را یک برآوردکننده اریب^۴ برای نقطه تغییر دانست (ساموئل و همکاران، ۱۹۹۸).

اهمیت شناسایی نقطه تغییر در فرآیندهای مختلف باعث شده است که برآورد آن در دو دهه اخیر مورد توجه محققان قرار گیرد. برخی از روش‌های برآورد نقطه تغییر عبارتند از:

- روش برآورد درست‌نمایی ماکزیمم^۵ (MLE)

- روش‌های یادگیری ماشین^۶ (ML)

^۱ Change point

^۲ Change point estimation

^۳ Signal time

^۴ Biased estimator

^۵ Maximum likelihood estimation

^۶ Machine learning

۴.۱ تاریخچه تحقیق

در دو دهه اخیر، مطالعات زیادی به منظور تعیین نقطه تغییر فرآیند انجام شده است. برخی از آن‌ها به شرح زیر می‌باشند:

ساموئل و پیگناتیلو (۱۹۹۸)، یک فرآیند با توزیع پواسن را توسط نمودارهای کنترل تعداد نقص در واحد بازرسی (C) و تعداد نقص در واحد محصول (U) تحت نظارت قرار دادند. سپس روش MLE را به منظور شناسایی نقطه تغییر به کار بردند.

ساموئل و همکاران (۱۹۹۸)، نمودار کنترل \bar{X} را برای نمایان سازی تغییرات میانگین فرآیند یک متغیره به کار بردند. به هنگام اعلام خارج از کنترل توسط این نمودار، روش MLE را به منظور تعیین نقطه تغییر فرآیند، مورد استفاده قرار دادند. سپس توسط شبیه‌سازی‌های مونت کارلویی، رویکرد مطرح شده را مورد ارزیابی قرار دادند.

در مطالعه دیگری، نمودار کنترل S، برای نمایان سازی تغییرات واریانس مورد استفاده قرار گرفت. پس از پایش فرآیند توسط این نمودار و اعلام خارج از کنترل، روش MLE برای تعیین نقطه تغییر، به کار رفت. در نهایت عملکرد این رویکرد، توسط شبیه‌سازی‌های مونت کارلویی مورد بررسی قرار گرفت (ساموئل و همکاران، ۱۹۹۸).

پیگناتیلو و ساموئل (۲۰۰۱)، نمودارهای کنترل مجموع تجمعی^۱ (CUSUM) و میانگین متحرک موزون نمایی^۲ (EWMA) را به منظور پایش فرآیند تولید، به کار بردند. سپس رویکردی را بر اساس تابع درستنمایی برای تعیین نقطه تغییر فرآیند، توسط برآورد یک فاصله اطمینان مطرح کردند. مطالعه دیگری برای تعیین موثر نقطه تغییر، حاکی از آن است که ابتدا فرآیند تولید توسط نمودار کنترل تعداد اقلام معیوب (np) تحت نظارت قرار گرفته و سپس روش MLE برای شناسایی نقطه تغییر به کار رفته است. سرانجام نتایج حاصل، بیانگر کشف موثر نقطه تغییر می‌باشد (پیگناتیلو و ساموئل، ۲۰۰۱).

شائو و همکاران (۲۰۰۶) با استفاده از نمودار کنترل S فرآیندی با توزیع گاما را تحت پایش قرار دادند و پس از اعلام خارج از کنترل توسط آن نمودار، روش MLE را برای تعیین موثر نقطه تغییر به کار بردند. آن‌ها برای دستیابی به UCL و LCL، از روش‌های مونت کارلویی استفاده نمودند.

در پژوهشی که شائو و هئو (۲۰۰۶) به منظور کشف و شناسایی اختلالات نفوذ کرده در فرآیندی با توزیع گاما انجام دادند، نمودار کنترل \bar{X} مورد استفاده قرار گرفت و سپس روش MLE برای تعیین نقطه تغییر مورد استفاده قرار گرفت. سرانجام توسط شبیه‌سازی‌های مونت کارلویی، عملکرد این رویکرد را مورد بررسی قرار دادند.

در پژوهش‌های صورت گرفته، باری دیگر شائو و هئو (۲۰۰۹)، مطالعه‌ای را برای تعیین نقطه تغییر فرآیند انجام دادند که در آن، نمودار کنترل EWMA به منظور پایش فرآیندی با توزیع یکنواخت به کار رفت؛ سپس روش MLE را برای تعیین موثر نقطه تغییر مورد استفاده قرار دادند.

نورالسنا و همکاران (۲۰۰۹)، یک فرآیند چندمتغیره با توزیع هندسی را توسط نمودار کنترل هندسی

^۱Cumulative sum control chart

^۲Exponentially weighted moving average

تحت نظارت قرار دادند. همچنین هشدارهای غلط^۱ که ممکن است توسط نمودار کنترل صورت گیرد را نیز مطرح کردند و سرانجام برای تعیین موثر نقطه تغییر، روش MLE را مورد استفاده قرار دادند. هئو و همکاران (۲۰۱۱)، با تحت نظارت قرار دادن فرآیندی چندمتغیره توسط نمودار کنترل |S| برای کشف و شناسایی اختلال ایجاد شده در ساختار کوواریانس فرآیند، روش MLE را به کار بردند. شائو و لین (۲۰۱۱)، نمودار کنترل T^2 هتلینگ را به منظور پایش یک فرآیند چندمتغیره که اختلال در بردار میانگین آن، نفوذ کرده بود، به کار بردند. سپس با روش ماشین بردار پشتیبان، راهبردی برای برآورد نقطه تغییر فرآیند معرفی کردند. شائو و هئو (۲۰۱۱) برای تعیین نقطه تغییر یک فرآیند با توزیع گاما، ترکیبی از نمودار کنترل EWMA و روش MLE را مورد بررسی قرار دادند.

۵.۱ ضرورت انجام تحقیق

در همه فرآیندهایی که شناسایی نقطه تغییر توسط روش MLE مورد مطالعه قرار گرفته است، لازم است که توزیع فرآیند مشخص باشد. با توجه به این که در دنیای واقعی، اکثر اوقات شناختی از توزیع فرآیند نداریم؛ بنابراین استفاده از روش MLE برای شناسایی نقطه تغییر می تواند کارا نباشد. همچنین، گاهی ممکن است پدیده بیش برآزش^۲ (مدلی که برای پیشگویی با داده های جدید کارایی ندارد) رخ دهد (شائو و همکاران، ۲۰۱۱). در فرآیندهای چندمتغیره، مشتقات تابع درستنمایی نیز پیچیده تر از حالت یک متغیره خواهند بود.

چون اکثر مواقع شناختی از توزیع فرآیند نداریم لذا محققان برای برآورد نقطه تغییر، از روش ML بهره بردند (شائو و همکاران، ۲۰۱۱؛ شائو، ۲۰۰۹)؛ اما تعداد متغیرهای ورودی در این مطالعات خیلی کم بوده است. از طرفی تحقیقات خیلی کمی در مورد مدیریت تعداد ورودی ها در روش ML صورت گرفته است. در واقع یکی از مهمترین مشکلات روش ML، هزینه زمانی بالا به دلیل تعداد زیاد متغیرهای ورودی می باشد (کراون و شالویک، ۱۹۹۷؛ چئو و همکاران، ۲۰۰۴).

همانطور که ملاحظه می شود، هر دو روش MLE و ML مشکلات خاص خود را دارند؛ بنابراین انگیزه ای برای حل مشکلات فوق (نامعلوم بودن توزیع فرآیند و تعداد بالای متغیرهای ورودی) و تعیین نقطه تغییر در فرآیندهای چندمتغیره به وجود آمد. به منظور بهبود فرآیند تولید و کاهش محاسبات زمانی، سیستم های تکاملی ترکیبی^۳، بطور رایج مورد استفاده قرار گرفته اند (چان و همکاران، ۲۰۱۰؛ چان و همکاران، ۲۰۱۱). تعدادی از مطالعات، رهیافت ترکیبی را برای SPC به کار برده اند (بیشوف و میلر، ۲۰۰۲؛ لو و همکاران، ۲۰۱۱)؛ اما این رهیافت، به ندرت در تعیین نقطه تغییر مورد استفاده قرار گرفته است.

^۱ False alarms

^۲ Overfit

^۳ Hybrid evolutionary systems

در این پایان‌نامه، دو رهیافت ترکیبی دو مرحله‌ای را با یک روش موثر در شناسایی نقطه تغییر فرآیند، ادغام می‌کنیم؛ به طوری که این رهیافت‌های ترکیبی، شامل ترکیب رگرسیون لجستیک^۱ با ماشین بردار پشتیبان^۲ (LR-SVM) و ترکیب رگرسیون چندمتغیره سازوار اسپلاین^۳ با ماشین بردار پشتیبان (MARS-SVM) می‌باشند. در این جا فرآیند تولید، یک فرآیند چندمتغیره است که ساختار کواریانسی آن (واریانس) تغییر کرده است.

هنگامی که با تعداد زیادی از متغیرهای توضیحی روبرو هستیم، ابتدا در مرحله اول، بعد فضای متغیرهای توضیحی را توسط دو روش مطرح شده LR و MARS کاهش می‌دهیم. سپس در مرحله دوم، مجموعه متغیرهای کاهش یافته را به عنوان ورودی مدل SVM در نظر می‌گیریم. وظیفه اصلی مدل SVM، رده‌بندی فرآیندی چندمتغیره به رده‌ی تحت کنترل و رده‌ی خارج از کنترل است. بعد از اجرای رده‌بندی مدل SVM، یک راهبرد شناسایی مفید برای تعیین نقطه تغییر معرفی می‌کنیم. با توجه به ماهیت فرآیند تولید، می‌توان فرآیند را در زمان‌های قبل از نقطه تغییر، به صورت تحت کنترل و در زمان‌های بعد از این نقطه، به صورت خارج از کنترل تلقی کرد؛ لذا دو حالتی بودن وضعیت مذکور، شرایط وجودی یک متغیر پاسخ کیفی با مقادیر به ترتیب ۰ و ۱ را ایجاد می‌کند. همچنین با در نظر گرفتن تغییر در ساختار کواریانسی فرآیند چندمتغیره، متغیرهای جدید توضیحی ایجاد می‌شوند که می‌توان با این مجموعه متغیرهای توضیحی و متغیر پاسخ، یک مدل رده‌بندی را ایجاد نمود؛ بدین منظور:

- در فصل دوم، روش‌های رده‌بندی آماری (رگرسیون لجستیک چندمتغیره و رگرسیون چندمتغیره سازوار اسپلاین) و یادگیری ماشین (ماشین بردار پشتیبان) مورد مطالعه قرار گرفته‌اند.
- در فصل سوم، چون تغییر در ساختار کواریانسی (واریانس) فرآیند فرض شده است، متغیرهای جدیدی تحت عنوان Z با ماهیت کواریانسی تشکیل می‌شوند؛ این متغیرها به عنوان متغیرهای توضیحی (ورودی) مدل رده‌بندی قرار می‌گیرند. با استفاده از خروجی مدل و یک راهبرد مؤثر، نقطه تغییر فرآیند شناسایی می‌گردد؛ همچنین در این فصل، طرح ترکیبی دو مرحله‌ای مطرح می‌گردد.
- سرانجام در فصل چهارم، مطالعات شبیه‌سازی و نتایج ارائه می‌گردد. نتایج حاکی از آن است که طرح ترکیبی دو مرحله‌ای در تعیین نقطه تغییر، مؤثر واقع شده است.

^۱Regression logistic

^۲Support vector machine

^۳Multivariate adaptive regression splines

فصل ۲

روش‌های رده‌بندی مورد بحث

۱.۲ رگرسیون لجستیک چندمتغیره

در مسائل رگرسیونی، هنگامی که متغیر پاسخ از نوع کمی است، کمیت کلیدی، میانگین شرطی متغیر پاسخ Y به شرط داشتن متغیر توضیحی Z است. این میانگین شرطی که به صورت $E(Y|z)$ نشان داده می‌شود را می‌توان به صورت یک تابع خطی از Z در نظر گرفت.

$$E(Y|z) = \beta_0 + \beta_1 z,$$

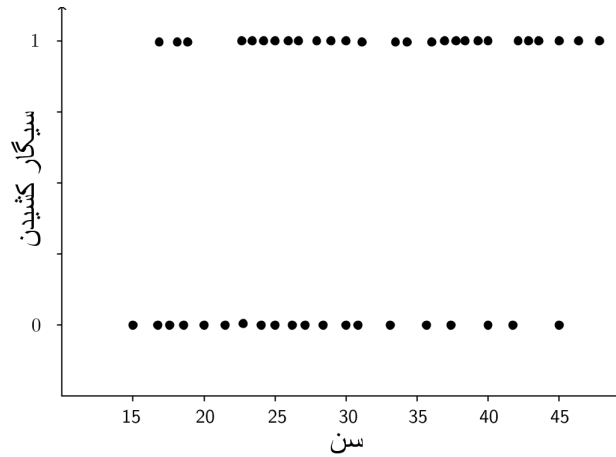
که برد آن بین $-\infty$ و ∞ است.

اگر متغیر پاسخ از نوع کیفی دو حالت با مقادیر 0 و 1 باشد، آنگاه

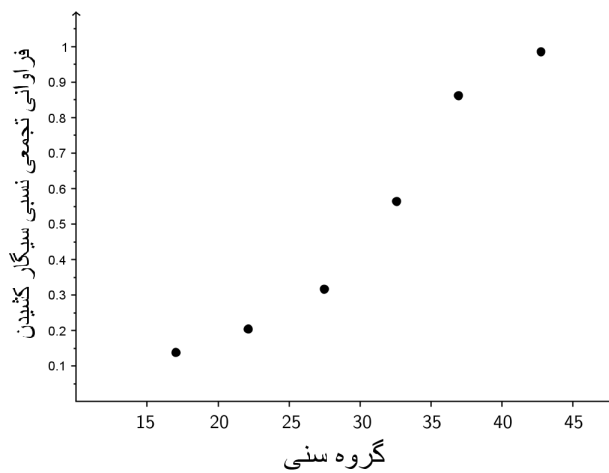
$$P(Y = 1|Z = z) = E(Y|Z = z)$$

خواهد بود، لذا این میانگین شرطی مقادیری بین 0 و 1 را اختیار می‌کند. از آنجا که مقادیر متغیر پاسخ 0 و 1 است لذا نمودار پراکنش متغیر پاسخ در برابر متغیر توضیحی، تفسیرپذیر نخواهد بود. برای مثال اگر متغیر پاسخ، سیگاری بودن یا نبودن افراد و متغیر توضیحی، سن افراد باشند، نمودار پراکنش مربوطه به صورت شکل ۱.۲ خواهد بود.

برای رفع مشکل تفسیرپذیری نمودار، می‌توان با تبدیل متغیر کمی سن به متغیر کیفی گروه سنی، این مشکل را رفع کرد. نمودار پراکنش متغیر پاسخ در برابر متغیر گروه سنی در شکل ۲.۲ آمده است که شکل S را در ذهن تداعی می‌کند. از آنجا که توابع توزیع تجمعی متغیرهای تصادفی تقریباً S شکل هستند، لذا مدل‌سازی متغیر پاسخ می‌تواند مطابق صورت تابعی یکی از این توابع تجمعی باشد. در رگرسیون لجستیک فرض شده است که این تابع همان تابع توزیع تجمعی متغیر تصادفی لجستیک است.



شکل ۱.۲: نمودار پراکنش متغیر پاسخ در مقابل متغیر سن (هاسمر و همکاران، ۲۰۰۴)



شکل ۲.۲: نمودار پراکنش متغیر پاسخ در مقابل متغیر گروه سنی (هاسمر و همکاران، ۲۰۰۴)

فرض کنید T معرف تعداد مشاهدات، N معرف تعداد متغیرهای توضیحی، Y بردار متغیر پاسخ، Q ماتریس طرح متغیرهای توضیحی با $T \times (N + 1)$ درآیه و بردار پارامترها، $\underline{\beta}$ برداری ستونی با $N + 1$ مولفه باشند.

$$Q = \begin{bmatrix} 1 & z_{1,1} & \cdots & z_{1,N} \\ 1 & z_{2,1} & \cdots & z_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{T,1} & \cdots & z_{T,N} \end{bmatrix}_{T \times (N+1)}, \quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}_{(N+1) \times 1}, \quad \underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}_{T \times 1}.$$

همچنین اگر $\underline{Z} = (Z_1, Z_2, \dots, Z_N)'$ و $P(Y = 1 | \underline{Z} = \underline{z}) = \pi(\underline{z})$ احتمال شرطی متغیر پاسخ به شرط داشتن \underline{z} باشند، مدل رگرسیون لجستیک چندمتغیره به صورت زیر تعریف می‌شود:

$$\pi(\underline{z}) = P(Y = 1 | \underline{Z} = \underline{z}) = \frac{e^{g(\underline{z})}}{1 + e^{g(\underline{z})}} .$$

که در آن، $g(\underline{z})$ تبدیل لجیت بوده و به صورت زیر محاسبه می شود:

$$g(\underline{z}) = \text{Ln} \left(\frac{\pi(\underline{z})}{1 - \pi(\underline{z})} \right) = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_N z_N . \quad (1.2)$$

۱.۱.۲ برآورد پارامترها

در رگرسیون لجستیک، پارامترها از روش درستنمایی ماکسیم برآورد می شوند و چون صورت بسته ای برای برآوردیابی این مدل وجود ندارد، توسط روش های عددی، پارامترهای مدل برآورد می شوند. رابطه (۱.۲) را برای هر یک از مشاهدات، می توان به صورت زیر نوشت:

$$g_i(\underline{z}) = \text{Ln} \left(\frac{\pi_i(\underline{z})}{1 - \pi_i(\underline{z})} \right) = \sum_{k=0}^N z_{i,k} \beta_k, \quad i = 1, \dots, T, \quad (2.2)$$

که در آن برای هر $i = 1, \dots, T$ ، داریم $Z_{i,0} = 1$. چون متغیر پاسخ از نوع کیفی دو حالتی است؛ بنابراین متغیر پاسخ دارای توزیع برنولی خواهد بود، یعنی

$$Y_i \sim B(1, \pi_i(\underline{z})), \quad i = 1, \dots, T .$$

برآورد پارامترهای مدل، توسط روش برآورد درستنمایی ماکسیم به دست می آیند. تابع درستنمایی به صورت زیر بیان می شود:

$$L(\underline{\beta} | \underline{y}) = f(\underline{y} | \underline{\beta}) = \prod_{i=1}^T \pi_i(\underline{z})^{y_i} (1 - \pi_i(\underline{z}))^{1-y_i} .$$

با در نظر گرفتن $\pi = \pi(\underline{z})$ داریم:

$$L(\underline{\beta} | \underline{y}) = \prod_{i=1}^T \left(\frac{\pi_i}{1 - \pi_i} \right)^{y_i} (1 - \pi_i) . \quad (3.2)$$

اما بنا به رابطه (۲.۲) داریم:

$$\left(\frac{\pi_i}{1 - \pi_i} \right) = e^{\sum_{k=0}^N z_{i,k} \beta_k} . \quad (4.2)$$

که پس از ساده کردن به صورت زیر خواهد بود:

$$\pi_i = \frac{e^{\sum_{k=0}^N z_{i,k} \beta_k}}{1 + e^{\sum_{k=0}^N z_{i,k} \beta_k}} . \quad (5.2)$$

با جایگذاری رابطه (۴.۲) و (۵.۲) در (۳.۲) داریم:

$$\begin{aligned} L(\underline{\beta}|\underline{y}) &= \prod_{i=1}^T \left(e^{\sum_{k=0}^N z_{i,k}\beta_k} \right)^{y_i} \cdot \left(1 - \frac{e^{\sum_{k=0}^N z_{i,k}\beta_k}}{1 + e^{\sum_{k=0}^N z_{i,k}\beta_k}} \right), \\ &= \prod_{i=1}^T \left(e^{y_i \sum_{k=0}^N z_{i,k}\beta_k} \right) \cdot \left(1 + e^{\sum_{k=0}^N z_{i,k}\beta_k} \right)^{-1}. \end{aligned} \quad (6.2)$$

رابطه (۶.۲)، هسته تابع درستنمایی برای ماکسیم شدن است. با توجه به این‌که تابع لگاریتم یک تابع یکنوا است و ماکسیم کردن تابع درستنمایی معادل با ماکسیم کردن لگاریتم آن می‌باشد؛ پس می‌توان نوشت:

$$l(\underline{\beta}) = \sum_{i=1}^T y_i \left(\sum_{k=0}^N z_{i,k}\beta_k \right) - \text{Ln} \left(1 + e^{\sum_{k=0}^N z_{i,k}\beta_k} \right). \quad (7.2)$$

با مشتق گرفتن از (۷.۲) نسبت به β_k داریم:

$$\begin{aligned} \frac{\partial l(\underline{\beta})}{\partial \beta_k} &= \sum_{i=1}^T y_i z_{i,k} - \frac{1}{1 + e^{\sum_{k=0}^N z_{i,k}\beta_k}} \cdot \frac{\partial}{\partial \beta_k} \left(1 + e^{\sum_{k=0}^N z_{i,k}\beta_k} \right), \\ &= \sum_{i=1}^T y_i z_{i,k} - \frac{e^{\sum_{k=0}^N z_{i,k}\beta_k}}{1 + e^{\sum_{k=0}^N z_{i,k}\beta_k}} \cdot z_{i,k}, \\ &= \sum_{i=1}^T y_i z_{i,k} - \pi_i \cdot z_{i,k}. \end{aligned} \quad (8.2)$$

برآورد درستنمایی ماکسیم بردار β با حل $N + 1$ دستگاه معادله

$$\frac{\partial l(\underline{\beta})}{\partial \beta_k} = 0, \quad k = 0, \dots, N.$$

نتیجه می‌شود. لازم به ذکر است که حل این دستگاه به روش‌های عددی امکان پذیر است (پیوست ب را ببینید).

حال مشتق دوم رابطه (۶.۲) را بررسی می‌کنیم. برای اینکه نقاط به‌دست آمده ماکسیم کننده باشند، باید ماتریس بدست آمده از مشتق دوم، معین منفی باشد، یعنی همه عناصر روی قطر اصلی آن، کوچک‌تر از صفر باشند.

مشتق دوم رابطه (۶.۲) به‌صورت زیر به‌دست می‌آید:

$$\begin{aligned} \frac{\partial^2 l(\beta)}{\partial \beta_k \partial \beta_{k'}} &= \frac{\partial}{\partial \beta_{k'}} \sum_{i=1}^T y_i z_{i,k} - z_{i,k} \pi_i, \\ &= \frac{\partial}{\partial \beta_{k'}} \sum_{i=1}^T -z_{i,k} \pi_i, \\ &= - \sum_{i=1}^T z_{i,k} \frac{\partial}{\partial \beta_{k'}} \left(\frac{e^{\sum_{k=0}^N z_{i,k} \beta_k}}{1 + e^{\sum_{k=0}^N z_{i,k} \beta_k}} \right). \end{aligned} \quad (9.2)$$

با در نظر گرفتن $u(z) = \sum_{k=0}^N z_{i,k} \beta_k$, $i = 1, \dots, T$ داریم:

$$\begin{aligned} \frac{d}{dz} \frac{e^{u(z)}}{1 + e^{u(z)}} &= \frac{(1 + e^{u(z)}) \cdot e^{u(z)} \cdot \frac{d}{dz} u(z) - e^{u(z)} \cdot e^{u(z)} \cdot \frac{d}{dz} u(z)}{(1 + e^{u(z)})^2}, \\ &= \frac{e^{u(z)} \cdot \frac{d}{dz} u(z)}{(1 + e^{u(z)})^2}, \\ &= \frac{e^{u(z)}}{1 + e^{u(z)}} \cdot \frac{1}{1 + e^{u(z)}} \cdot \frac{d}{dz} u(z). \end{aligned}$$

در نتیجه رابطه (۹.۲) را می‌توان به صورت زیر نوشت:

$$\frac{\partial^2 l(\beta)}{\partial \beta_k \partial \beta_{k'}} = - \sum_{i=1}^T z_{i,k} \pi_i (1 - \pi_i) z_{i,k'} < 0.$$

بنابراین نقاط به دست آمده از رابطه (۸.۲) ماکسیم کننده لگاریتم تابع درستنمایی مذکور هستند. اگر رابطه (۸.۲) را با $l'(\beta)$ نشان دهیم و $\beta^{(0)}$ مقادیر اولیه دلخواه برای هر بردار β_k باشند، آنگاه بنا به روش نیوتن-رافسون داریم:

$$\beta^{(1)} = \beta^{(0)} + [-l''(\beta^{(0)})]^{-1} \cdot l'(\beta^{(0)}). \quad (10.2)$$

برای بازنویسی و فهم آسان‌تر رابطه (۱۰.۲)، فرض کنید که μ یک بردار ستونی با T مولفه‌ی μ_i برای $E(y_i | z_i) = \pi_i$ باشد، لذا مشتق اول لگاریتم تابع درستنمایی را می‌توان به صورت ماتریسی نوشت:

$$l'(\beta) = Q^T (y - \mu),$$

که یک بردار ستونی با $N + 1$ مولفه‌ی $\frac{\partial l(\beta)}{\partial \beta_k}$ می‌باشد و اگر W یک ماتریس قطری با مولفه‌های $\pi_i(1 - \pi_i)$ باشد، مشتق دوم لگاریتم تابع درستنمایی را می‌توان به صورت ماتریسی نوشت:

$$l''(\beta) = -Q^T W Q,$$

که یک ماتریس مربعی $(N + 1) \times (N + 1)$ با مولفه‌های $\frac{\partial^2 l(\beta)}{\partial \beta_k \partial \beta_{k'}}$ می‌باشد؛ بنابراین رابطه (۱۰.۲) را به صورت زیر می‌توان نوشت:

$$\beta^{(1)} = \beta^{(0)} + [Q^T W Q]^{-1} \cdot Q^T (y - \mu) . \quad (11.2)$$

رابطه (۱۱.۲) تا زمانی که مولفه‌های بردار β به همگرایی برسند، تکرار می‌شود.

۲.۱.۲ انتخاب متغیرهای مهم

انتخاب متغیرهای مهم در رگرسیون لجستیک توسط یکی از روش‌های انتخاب متغیر، موسوم به روش پیشرو والد^۱، انجام می‌گیرد. در این رویه ابتدا متغیرها توسط آزمون نمره^۲ وارد مدل شده و سپس توسط آزمون والد^۳ (پیوست پ را ببینید) از مدل خارج می‌شوند که شرح دقیق‌تر این رویه، در الگوریتم زیر آمده است:

الگوریتم انتخاب متغیر در رویه پیشرو گام به گام والد

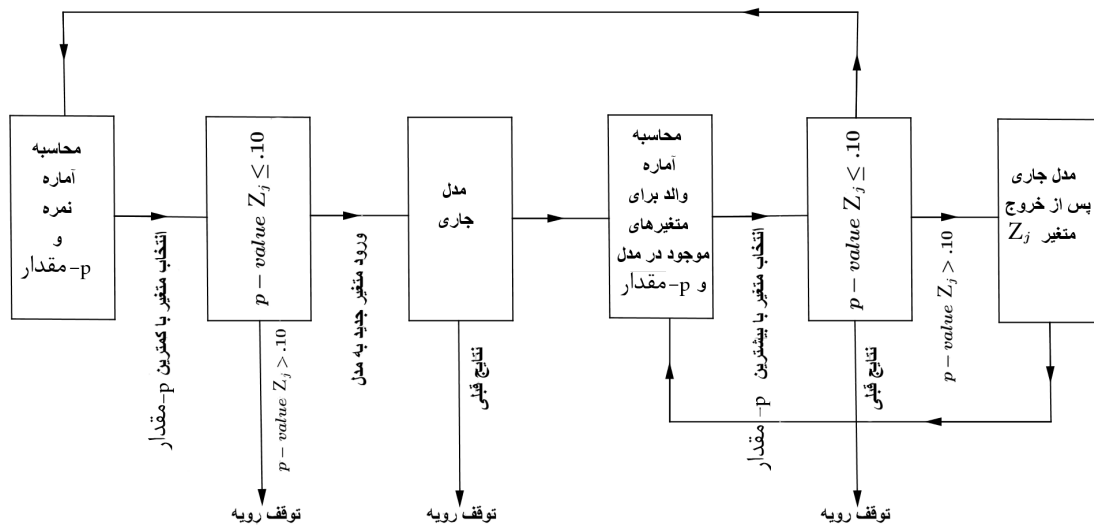
۱. محاسبه آماره نمره و p -مقدار برای هر متغیر (متغیرهای خارج از مدل).
۲. انتخاب متغیر با کمترین p -مقدار به طوری که اگر p -مقدار کوچک‌تر از $\alpha = 0.1$ باشد، مرحله ۳ اجرا شود، در غیر این صورت رویه متوقف شود.
۳. مدل فعلی با افزایش یک متغیر جدید، به روز شده و اگر نتایج این مدل، شبیه نتایج مدل قبل باشد، رویه متوقف شود؛ در غیر این صورت مرحله بعد اجرا شود.
۴. محاسبه آماره والد به همراه p -مقدار برای هر متغیر موجود در مدل.
۵. انتخاب متغیر با بیشترین p -مقدار به طوری که اگر p -مقدار کوچک‌تر از $\alpha = 0.1$ شد، به مرحله ۱ برگرد، اگر مدل جاری با حذف متغیر، شبیه مدل قبلی شد، رویه متوقف شود، در غیر این صورت مرحله بعد اجرا شود.
۶. مدل جاری با خارج کردن متغیر از مدل، به روز شده و پس از برآورد پارامترهای مدل، به مرحله ۴ برمی‌گردد.

^۱Wald forward method

^۲Score test

^۳Wald test

شکل ۳.۲ انتخاب متغیرهای مهم را نشان می‌دهد.



شکل ۳.۲: انتخاب متغیرهای مهم به روش رویه پیشرو گام به گام والد

۲.۲ رگرسیون چندمتغیره سازوار اسپلاین

رگرسیون چند متغیره سازوار اسپلاین^۱ MARS، ترکیبی از رگرسیون اسپلاین^۲، برازش مدل به روش گام به گام^۳ و افراز بازگشتی^۴ است. این روش، هم در مسائل رگرسیونی و هم در مسائل رده‌بندی به‌کار می‌رود؛ به عبارتی دیگر، اگر متغیر پاسخ از نوع کمی باشد، با مسئله رگرسیونی و اگر از نوع کیفی باشد، با مسئله رده‌بندی مواجه هستیم. این روش هنگامی که تعداد متغیرهای توضیحی زیاد است، عملکرد خوبی را از خود نشان می‌دهد و به‌طور موثر در شناسایی اثرات متقابل عمل می‌کند.

روش MARS، به‌جای استفاده مستقیم از متغیرها، توابعی از آنها را به‌منظور مدل‌سازی به‌کار می‌برد؛ این توابع را توابع پایه‌ای^۵ گویند. برای هر یک از متغیرهای Z_1, Z_2, \dots, Z_N توابع پایه‌ای h_1

^۱Multivariate Adaptive Regression Spline

^۲Spline regression

^۳Stepwise model fitting

^۴Recursive partitioning

^۵Basis functions

و h_2 را در هر گره t به صورت زیر تعریف می‌کنند:

$$h_1(Z_k) = (Z_k - t)_+ = \max(0, Z_k - t), \quad k = 1, \dots, N,$$

$$= \begin{cases} Z_k - t & Z_k > t \\ 0 & \text{otherwise} \end{cases},$$

$$h_2(Z_k) = (t - Z_k)_+ = \max(0, t - Z_k), \quad k = 1, \dots, N,$$

$$= \begin{cases} t - Z_k & Z_k < t \\ 0 & \text{otherwise} \end{cases},$$

که

$$t \in \{z_{1,k}, z_{2,k}, \dots, z_{T,k}\}, \quad k = 1, 2, \dots, N,$$

و T نمایانگر تعداد مشاهدات است. مجموعه

$$C = \{(Z_k - t)_+, (t - Z_k)_+; k = 1, \dots, N\}.$$

را جفت توابع پایه‌ای گویند. در روش MARS مقادیر بهینه گره‌ها در مکان‌هایی تعیین می‌شوند که مشاهدات دارای بیشترین تغییر هستند یا به عبارت دیگر دارای تغییر رفتار هستند. این ویژگی روش MARS، هزینه زمانی محاسبات را به شدت کاهش می‌دهد. تعداد توابع پایه‌ای مدل با در نظر داشتن N متغیر و T مشاهده، حداکثر برابر $2TN$ است ولی پس از رویه حذف پسرو، تعداد این توابع به $G \leq 2TN$ کاهش خواهد یافت. لازم به ذکر است اگر رویه حذف پسرو انجام نگیرد، پدیده بیش برآزش صورت می‌گیرد که در این صورت مدلی ناکارا جهت پیشگویی خواهیم داشت. در این جا چون متغیر پاسخ از نوع کیفی دوحالته است، لذا مدل MARS را به صورت زیر داریم:

$$\ln \left(\frac{P(Y_i = 1 | z)}{1 - P(Y_i = 1 | z)} \right) = f_i(z), \quad i = 1, \dots, T,$$

$$= \beta_0 + \sum_{j=1}^G \beta_j h_j(Z).$$

و در نتیجه مدل برآورد شده به صورت زیر حاصل می‌شود:

$$\hat{f}_i(z) = \hat{\beta}_0 + \sum_{j=1}^G \hat{\beta}_j h_j(Z).$$

که در آن، $h_j(Z)$ توابع پایه‌ای هستند که عضو مجموعه C یا اثرات متقابل مرتبه دو یا مراتب بالاتر توابع پایه‌ای هستند. پارامترهای مدل $(\beta_1, \dots, \beta_G)$ توسط روش MLE، برآورد می‌شوند؛ ضریب ثابت نیز توسط $\bar{f}(z) = \sum_{i=1}^T \frac{1}{T} f_i(z)$ به دست می‌آید.

¹Knot

۱.۲.۲ انتخاب توابع پایه‌ای مهم

توابع پایه‌ای در مرحله پیشرو، به صورت جفت وارد مدل می‌شوند ولی در مرحله پسرو، تک تک حذف می‌شوند. در مرحله پیشرو، ابتدا برآورد ضریب ثابت، $\hat{\beta}_0 = \overline{f(z)}$ وارد مدل می‌شود، در مرحله بعد، چون با افزایش جفت توابع پایه‌ای به مدل، مجموع توان‌های دوم خطا^۱ (SSE) مدل کاهش می‌یابد، آن جفت تابع پایه‌ای وارد مدل می‌شوند که بیشترین کاهش را در SSE مدل ایجاد کنند.

برای مثال، فرض کنید در مرحله اول مقدار $SSE_1 = 100$ باشد و اکنون در مرحله دوم مرحله پیشرو قرار داریم. سه جفت تابع پایه‌ای، نامزد ورود به مدل هستند؛ اگر مدل M_A ، شامل برآورد ضریب ثابت و جفت تابع پایه‌ای اول، مدل M_B ، شامل برآورد ضریب ثابت و جفت تابع پایه‌ای دوم و مدل M_C ، شامل برآورد ضریب ثابت و جفت تابع پایه‌ای سوم باشند و $SSE_A = 95$ ، $SSE_B = 80$ و $SSE_C = 85$ ، آنگاه چون جفت تابع پایه‌ای دوم بیشترین کاهش را در SSE مدل ایجاد کرده‌اند، وارد مدل می‌شوند. در مرحله سوم، اگر مدل M_D ، شامل مدل جاری (مدلی که شامل برآورد ضریب ثابت و جفت تابع پایه‌ای دوم است) به علاوه جفت تابع پایه‌ای اول و مدل M_E ، شامل مدل جاری به علاوه جفت تابع پایه‌ای سوم باشند و مقادیر $SSE_D = 65$ و $SSE_E = 73$ باشند، آنگاه بیشترین کاهش SSE مدل، توسط ورود جفت تابع پایه‌ای اول به مدل ایجاد شده است و در مرحله آخر جفت تابع پایه‌ای سوم وارد مدل می‌شوند. اثرات متقابل دو تایی یا بیشتر این جفت توابع پایه‌ای نیز ممکن است وارد مدل شوند که در این صورت مرحله فوق با در نظر گرفتن این اثرات متقابل انجام می‌گیرد. همچنین در شرایط زیر مرحله پیشرو متوقف خواهد شد:

۱. تعداد توابع پایه‌ای توسط کاربر مشخص گردد.
 ۲. با افزایش جفت توابع پایه‌ای جدید، مقدار R^2 مدل کمتر از 0.001 تغییر کند.
 ۳. مقدار R^2 مدل بزرگتر مساوی 0.999 باشد.
 ۴. مقدار $GRsq$ کمتر از 10 - باشد.
- که در آن

$$SSE = \sum_{i=1}^T (f_i(z) - \hat{f}_i(z))^2 .$$

$$R^2 = \frac{\sum_{i=1}^T (\hat{f}_i(z) - \overline{f(z)})^2}{\sum_{i=1}^T (f_i(z) - \overline{f(z)})^2} .$$

$$GRsq = 1 - \frac{GCV}{GCV.null} .$$

$$GCV = \frac{SSE}{\left(1 - \frac{\nu_m + 1}{T}\right)^2} .$$

که $GCV.null$ ، مقدار GCV مدلی است که تنها شامل ضریب ثابت و m تعداد توابع پایه‌ای است. ν پارامتر هموارساز می‌باشد که بایستی بهینه شود؛ این پارامتر معمولاً بین ۲ تا ۴ انتخاب می‌شود.

^۱Sum of squared error

پس از اتمام مرحله پیشرو، مرحله حذف پسرو کار خود را شروع می‌کند. فرض کنید f_g ، معرف مدل جاری (مدلی که پس از اتمام مرحله پیشرو به دست آمده است) باشد که شامل g تابع پایه‌ای و ضریب ثابت است، در هر مرحله از حذف پسرو، یک تابع پایه‌ای از مدل حذف خواهد شد. این روند تا زمانی که تمام توابع پایه‌ای از مدل خارج شوند تکرار خواهد شد. لذا مدل‌های $f_0, f_1, \dots, f_{g-2}, f_{g-1}$ را خواهیم داشت. سوالی که پیش می‌آید آن است که این مرحله چگونه توابع پایه‌ای را حذف می‌کند؟ در جواب می‌توان گفت:

فرض کنید h_1, h_2, \dots, h_g توابع پایه‌ای مدل f_g باشند، از آن جا که با حذف توابع پایه‌ای از مدل، SSE مدل افزایش می‌یابد، لذا آن تابع پایه‌ای از مدل f_g حذف می‌شود که کمترین افزایش را در SSE_{g-1} ایجاد کند.

مرحله حذف پسرو نیز همانند مرحله پیشرو است؛ تفاوت این آن در این است که در هر مرحله تنها یکی از توابع پایه‌ای از مدل حذف می‌شود و معیار حذف این تابع پایه‌ای از مدل، دستیابی به کمترین مقدار SSE بعد از حذف این تابع پایه‌ای از مدل است. در نتیجه همه این مدل‌ها، $f_0, f_1, \dots, f_{g-2}, f_{g-1}$ می‌توانند مدل نهایی MARS باشند، مدل بهینه توسط معیار GCV تعیین می‌گردد که مدل بهینه دارای کمترین مقدار GCV است.

در این جا معیار GCV ، بیان کننده خطای رده‌بندی مدل MARS و معیاری جهت تعیین دقت این مدل است؛ لذا مدل بهینه، کمترین مقدار GCV را اختیار می‌کند.

۳.۲ ماشین بردار پشتیبان

ماشین بردار پشتیبان (SVM) یک روش رگرسیونی و رده بندی است که اولین بار در سال ۱۹۹۵ توسط وپنیک^۱ و همکارانش در آزمایشگاه AT & Bell پیشنهاد شد. در حال حاضر یکی از بهترین روش‌های رده‌بندی، SVM می‌باشد. در واقع این روش یک روش رده‌بند دوتایی است که توانایی رده‌بندی به چند رده را نیز دارد. همچنین برای مسائل خطی و غیرخطی مورد استفاده قرار می‌گیرد. روش SVM با به‌کارگیری یک ابرصفحه مجزا ساز بهینه، رده‌بندی را ایفا می‌کند.

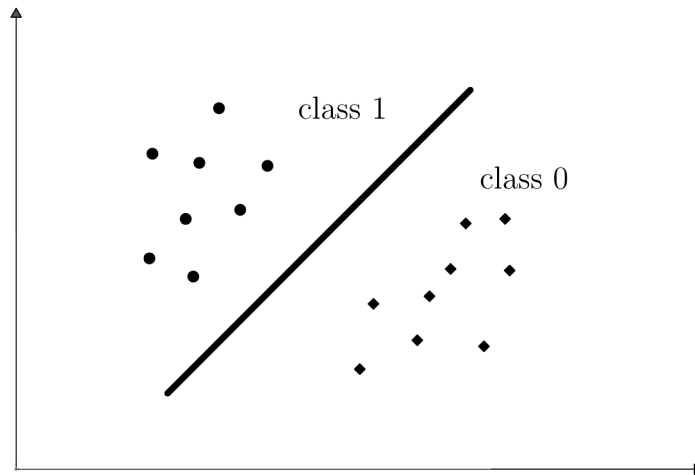
۱.۳.۲ ماشین بردار پشتیبان خطی:

فرض کنید یک مسئله رده‌بندی با دو رده را در اختیار داریم، $Z_i = (Z_{i,1}, \dots, Z_{i,N})$ ، $\{(Z_i, Y_i)\}_{i=1}^M$ و $Y_i = \{-1, 1\}$ مجموعه داده‌های آموزشی (مجموعه داده‌ای که مدل‌سازی با آن انجام می‌گیرد) باشند که به طور خطی از هم مجزا هستند؛ M و N به ترتیب معرف تعداد مشاهدات و تعداد متغیرهای توضیحی هستند. شکل ۴.۲ این مسئله را نشان می‌دهد.

همانطور که بیان شد، هدف این روش، یافتن ابرصفحه بهینه با شرط رده‌بندی صحیح داده‌های آموزشی است. اگر این ابرصفحه را به صورت

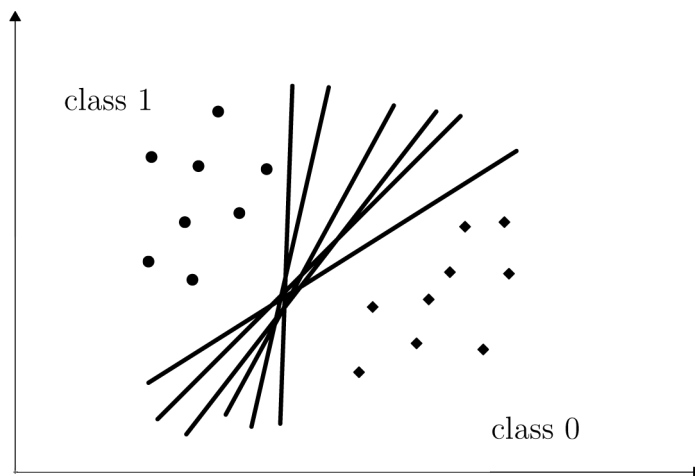
$$W^T Z + b = 0,$$

^۱Vapnik



شکل ۴.۲: داده های آموزشی که به صورت خطی جدا پذیرند

در نظر بگیریم که در آن برداری N بعدی و b ضریب ثابت است، آن گاه به ازای مقادیر مختلف W و b می توان تعداد زیادی از ابرصفحه ها را یافت که توانایی رده بندی داده های آموزشی را دارند. شکل ۵.۲ گویای این مسئله است. از بین این ابرصفحه ها، تنها آن ابرصفحه ای بهینه است که دارای بیشترین فاصله از نقاط حاشیه ای دو رده باشد. به این ابرصفحه، ابرصفحه مجزاساز بهینه^۱ می گویند.



شکل ۵.۲: تعدادی از ابرصفحه هایی که رده بندی را انجام داده اند

فرض کنید داده های آموزشی به طور خطی جداپذیر باشند، متغیر پاسخ Y_i مقادیر $+1$ و -1 را اختیار می کند و به ترتیب رده های 1 و 0 را نشان می دهند.

$$Y_i = \begin{cases} +1 & \text{if } Z_i \in 1 \text{ رده} \\ -1 & \text{if } Z_i \in 0 \text{ رده} \end{cases}$$

^۱Optimal separating hyperplane

این ابرصفحه‌ها، با در نظر گرفتن مقادیر مختلف برای جهت ابرصفحه (W) و مکان آن (b) در فضا تعیین می‌گردند.

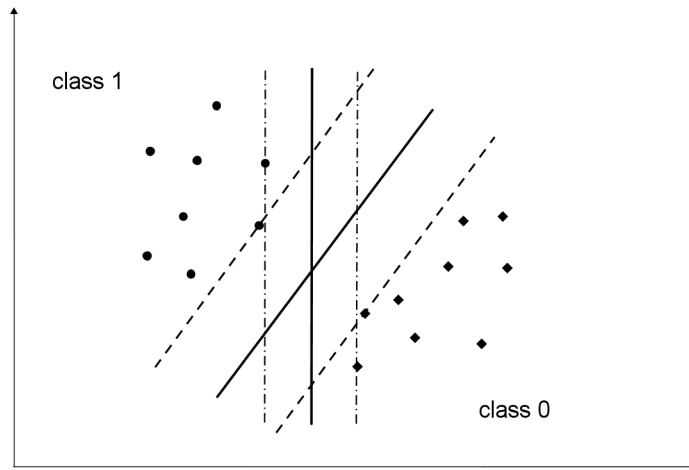
تابع تصمیم‌گیری به صورت زیر تعریف می‌شود:

$$d(\underline{Z}) = \text{sign}(W^T \underline{Z} + b) . \quad (12.2)$$

یعنی:

$$\begin{cases} (W^T \underline{Z}_i + b) > 0 & \text{if } y_i = +1 \\ (W^T \underline{Z}_i + b) < 0 & \text{if } y_i = -1 \end{cases} . \quad (13.2)$$

ابرفصفحه مجزاساز بهینه، دارای دو ویژگی خاص است؛ ویژگی اول این است که کمترین خطای ممکن را در رده‌بندی داده‌ها ایجاد کند و ویژگی دوم آن است که از حاشیه دو رده دارای بیشترین فاصله باشد؛ بنابراین پارامترهای W و b طوری تعیین می‌گردند که شروط مذکور برقرار باشند. شکل ۶.۲ ابرصفحه‌های مجزاساز را نشان می‌دهد.

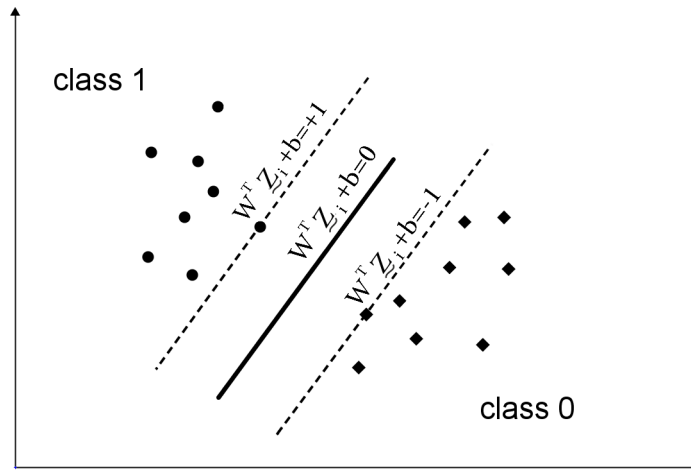


شکل ۶.۲: ابرصفحه‌های مختلف همراه ابرصفحه‌های حاشیه‌ای

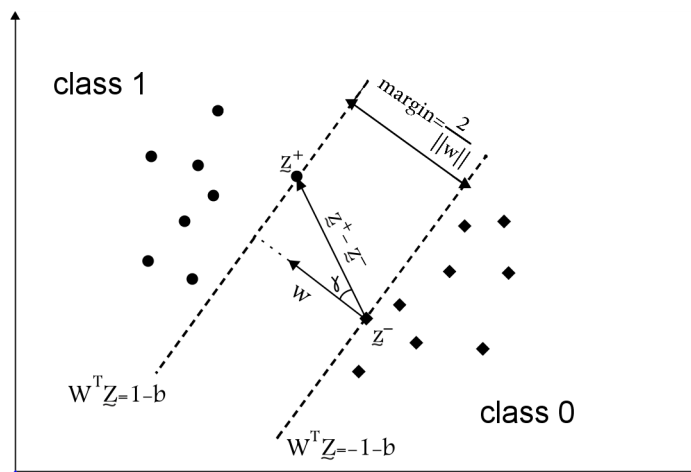
اگر فرض کنیم معادله‌ی ابرصفحه مجزاساز بهینه، به صورت $W^T \underline{Z}_i + b = 0$ باشد، آنگاه معادله ابرصفحه‌های حاشیه‌ای $W^T \underline{Z}_i + b = +1$ و $W^T \underline{Z}_i + b = -1$ خواهد بود. شکل ۷.۲ معادلات ابرصفحه‌ها را نشان می‌دهد. فاصله بین ابرصفحه‌های $W^T \underline{Z}_i + b = +1$ و $W^T \underline{Z}_i + b = -1$ را توسط قضیه تصویر (پیوست ت را ببینید) و شکل ۷.۲ داریم:

$$\begin{aligned} \text{Margin} &= \|\underline{Z}^+ - \underline{Z}^-\| \cdot \cos \gamma = \frac{W \cdot (\underline{Z}^+ - \underline{Z}^-)}{\|W\|} = \frac{(W \underline{Z}^+ - W \underline{Z}^-)}{\|W\|}, \\ &= \frac{(1 - b) - (-1 - b)}{\|W\|} = \frac{2}{\|W\|}, \end{aligned}$$

که در آن Margin فاصله بین دو ابرصفحه حاشیه‌ای، \underline{Z}^+ و \underline{Z}^- به ترتیب داده آموزشی روی ابرصفحه‌های حاشیه‌ای $W^T + \underline{Z}_i + b = +1$ و $W^T + \underline{Z}_i + b = -1$ هستند.



شکل ۷.۲: معادلات ابرصفحه‌های حاشیه‌ای و مجزاساز بهینه



شکل ۸.۲: بیشترین فاصله بین ابرصفحه‌های حاشیه‌ای

چون فرض بر این است که داده‌های آموزشی جداپذیر خطی هستند، لذا هیچ یک از داده‌ها روی ابرصفحه $W^T Z_i + b = 0$ قرار نمی‌گیرند؛ بنابراین معادله (۱۳.۲) را می‌توان به صورت زیر نوشت:

$$(W^T Z_i + b) \begin{cases} \geq 1 & \text{if } y_i = +1 \\ \leq -1 & \text{if } y_i = -1 \end{cases}$$

که به طور ساده‌تر داریم:

$$y_i(W^T Z_i + b) \geq 1, \quad i = 1, 2, \dots, M. \quad (14.2)$$

برای دستیابی به ابرصفحه مجزاساز که بیشترین حاشیه را ایجاد کند، می‌بایست فاصله بین ابرصفحه‌های حاشیه‌ای متناظر با آن، ماکسیمم شود؛ لذا $\frac{1}{\|w\|}$ باید ماکسیمم شود. برای ساده‌تر شدن محاسبات، به جای ماکسیمم کردن $\frac{1}{\|w\|}$ ، می‌توان مربع مخرج آن، $\|w\|^2$ ، را مینیمم کرد.

از آن‌جا که مینیمم کردن $\|W\|^2$ هم‌ارز با مینیمم کردن عبارت $\frac{1}{2}W^T W$ است، می‌توان مسئله بهینه‌سازی را به صورت زیر که یک مسئله درجه دو^۱ می‌باشد، نوشت.

$$\begin{aligned} \text{Minimize } J(W) &\equiv \frac{1}{2}W^T W, \\ \text{Subject to } &y_i(W^T Z_i + b) \geq 1, \quad i = 1, 2, \dots, M. \end{aligned} \quad (15.2)$$

می‌دانیم که اگر یک مسئله درجه دو، دارای شرط نامساوی باشد، آنگاه تابع هدف^۲ منحصر به فردی خواهد داشت؛ به عبارت دیگر فقط یک نقطه اکسترمم دارد. معادله (۱۵.۲)، دارای این ویژگی است که ویژگی مهم روش SVM، محسوب می‌شود. این ویژگی، این روش را نسبت به روش شبکه‌های عصبی چندلایه^۳ برتر ساخته است، چرا که روش شبکه‌های عصبی چندلایه، دارای تعداد زیادی مینیمم موضعی است. برای مینیمم سازی مسئله درجه دو رابطه (۱۵.۲) و دستیابی به مقادیر بهینه W و b با توجه به قید نامساوی مطرح شده، می‌توان از ضریب لاگرانژ^۴ استفاده کرد.

ضریب لاگرانژ برای شناسایی نقاط خاصی از تابع که دارای چندین متغیر و محدودیت است، به کار می‌رود. به عبارت دیگر اگر هدف مینیمم سازی تابع $f(x)$ با توجه به محدودیت $g(x) \leq 0$ باشد، آنگاه تابع لاگرانژ، $L(x, \alpha)$ ، باید مینیمم شود که به صورت زیر تعریف می‌شود:

$$L(x, \alpha) = f(x) + \alpha g(x),$$

که در آن، $\alpha \geq 0$ ، ضریب لاگرانژ نامیده می‌شود.

با توجه به تابع لاگرانژ، رابطه (۱۵.۲) به صورت بدون قید رابطه (۱۶.۲) تبدیل می‌شود که به آن تابع لاگرانژ اولیه^۵ گویند.

$$L_p(W, b, \alpha) = \frac{1}{2}W^T W - \sum_{i=1}^M \alpha_i [y_i(W^T Z_i + b) - 1], \quad (16.2)$$

که در آن

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_M)^T \geq 0.$$

برای مینیمم کردن رابطه (۱۶.۲)، باید نقاط مانای^۶ (نقاط بحرانی) تابع لاگرانژ را به دست آورد؛ برای دستیابی به این نقاط، شرایط KKT^۷ (پیوست ۳ را ببینید) بررسی می‌شود که به صورت زیر می‌باشد:

^۱Quadratic problem

^۲Objective function

^۳Multi-layer neural networks

^۴Lagrange coefficient

^۵Primal lagrange function

^۶Stationary points

^۷Karush-Kuhn-Tucker conditions

$$\begin{aligned}\frac{\partial L_p(W, b, \alpha)}{\partial W} &= W - \sum_{i=1}^M \alpha_i y_i \underline{Z}_i = 0, \\ \frac{\partial L_p(W, b, \alpha)}{\partial b} &= 0.\end{aligned}\quad (17.2)$$

نتیجه رابطه (۱۷.۲) به صورت زیر است:

$$\begin{aligned}W &= \sum_{i=1}^M \alpha_i y_i \underline{Z}_i, \\ - \sum_{i=1}^M \alpha_i y_i &= 0.\end{aligned}\quad (18.2)$$

با قرار دادن مقدار W به دست آمده از رابطه (۱۸.۲) (W بهینه) در رابطه (۱۶.۲)، معادله اساسی ماشین‌های بردار به دست می‌آید که این معادله، صورت دوگان معادله لاگرانژ^۱ است. معادله اساسی ماشین‌های بردار به صورت زیر است:

$$\begin{aligned}Maximize L_d(\alpha) &= \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=1}^M y_i y_j \alpha_i \alpha_j \underline{Z}_i^T \underline{Z}_j, \\ Subject\ to &\begin{cases} \alpha_i \geq 0 \\ \sum_{i=1}^M \alpha_i y_i = 0 \end{cases}.\end{aligned}\quad (19.2)$$

در جواب (۱۹.۲) مقادیر ضرایب لاگرانژ، α ، برای هر یک از بردار داده‌های آموزشی به دست خواهند آمد و با توجه به محدودیت بیان شده در رابطه (۱۹.۲)، مقداری بزرگ‌تر یا مساوی صفر خواهند داشت. از بین بردار داده‌های آموزشی رده ۱ و ۰، آن‌هایی که مقدار α متناظر با آن‌ها، بزرگ‌تر از صفر باشد، بردارهای پشتیبان^۲ نامیده می‌شوند.

در عمل برای به دست آوردن ابرصفحه بهینه، فقط بردارهای پشتیبان به کار می‌روند. لذا تا حد زیادی هزینه محاسبات کاهش می‌یابد. بردارهای پشتیبان دقیقاً بر روی ابرصفحه‌های $\pm 1 = W^T \underline{Z} + b$ قرار می‌گیرند. در شکل ۹.۲ بردارهای پشتیبان به وضوح مشخص شده اند. مقدار بهینه b نیز از رابطه زیر به دست می‌آید:

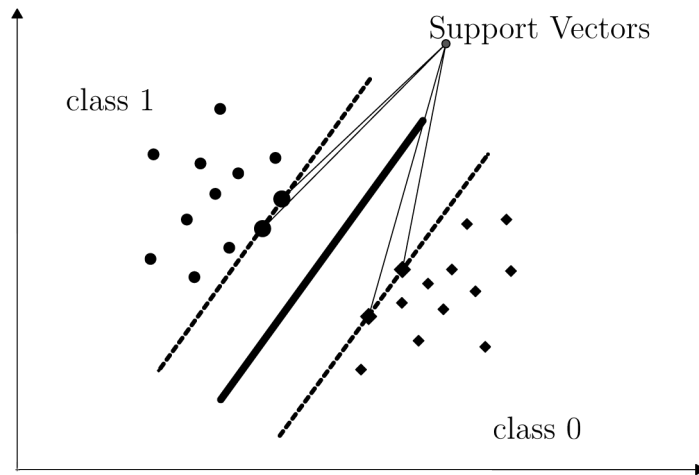
$$b = \frac{1}{SV} \sum_{i=1}^{SV} (y_i - W^T \underline{Z}_i), \quad (20.2)$$

که در آن SV تعداد بردارهای پشتیبان است.

حال با توجه به مقادیر بهینه W و b و با قرار دادن آن‌ها در رابطه (۱۲.۲)، SVM داده جدید \underline{Z} را به رده ۱، اگر مقدار تابع تصمیم بزرگ‌تر از صفر باشد یا رده ۰، اگر مقدار تابع تصمیم کوچک‌تر از صفر باشد، تخصیص می‌دهد.

^۱Dual lagrange function

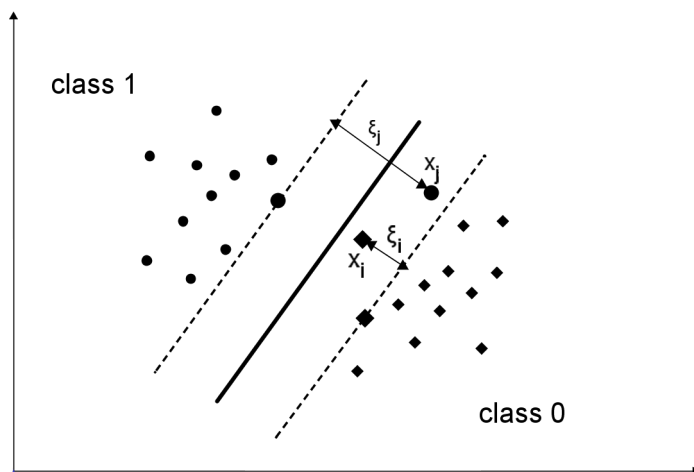
^۲Support Vectors



شکل ۹.۲: بردارهای پشتیبان رده ۱ و ۰ بر روی ابرصفحه‌های حاشیه‌ای

۲.۳.۲ ماشین بردار پشتیبان در سیستم‌های خطی جداناپذیر

در سیستم‌های خطی، گاهی اوقات داده‌ها طوری قرار گرفته‌اند که به‌طور خطی جداناپذیر نیستند، با این حال می‌توان با قبول مقدار کمی خطا، این داده‌ها را به‌طور خطی رده‌بندی کرد. در این شرایط، تعدادی از داده‌ها به غلط به رده دیگری تخصیص می‌یابند. در شکل ۱۰.۲ ملاحظه می‌شود که با یک رده‌بند خطی نمی‌توان داده‌ها را کاملاً از هم جدا کرد لذا در این شرایط، نمی‌توان ابرصفحه‌ی خطی یافت که عمل جداسازی را به خوبی انجام دهد، مگر این‌که داده‌هایی، درون حاشیه جداسازی قرار گیرند.



شکل ۱۰.۲: سیستم‌های خطی جداناپذیر با میزان خطا ξ

برای حل این مشکل، متغیر $\xi \geq 0$ به نام متغیر کساد^۱ تعریف می‌شود که به نوعی خطای رده‌بندی را نشان می‌دهد.

سه حالت برای بردار داده‌ها به وجود خواهد آمد:

۱. داده‌هایی که خارج از حاشیه قرار می‌گیرند؛ بنابراین به‌طور صحیح رده‌بندی شده‌اند و شرایط رابطه

^۱Slack variable

(۱۴.۲) را برقرار می‌سازند.

۲. داده‌هایی که درون حاشیه قرار می‌گیرند و به درستی رده‌بندی شده‌اند و در شرایط زیر صدق می‌کنند:

$$0 \leq y_i(W^T Z_i + b) < 1.$$

۳. داده‌هایی که به غلط رده‌بندی شده‌اند و در شرط زیر صدق می‌کنند:

$$y_i(W^T Z_i + b) < 0.$$

سه حالت فوق را می‌توان با اضافه کردن یک محدودیت جدید به قید نامساوی رابطه (۱۴.۲) به صورت زیر نوشت:

$$y_i(W^T Z_i + b) \geq 1 - \xi_i.$$

که حالت اول با در نظر گرفتن $\xi_i = 0$ و حالت دوم با $0 < \xi_i \leq 1$ و حالت سوم با $\xi_i > 1$ ایجاد خواهند شد؛ بنابراین هدف، رسیدن به یک حالت موازنه^۱ بین ماکسیم شدن حاشیه و مینیم شدن خطا رده‌بندی می‌باشد. صورت ریاضی آن برابر با مینیم کردن تابع هزینه زیر است:

$$J(W, b, \xi) = \frac{1}{\eta} \|W\|^2 + C \sum_{i=1}^M l(\xi_i). \quad (21.2)$$

که در آن ξ بردار پارامترهای ξ_i است. C ، پارامتر موازنه بین فاصله حاشیه و میزان خطا رده‌بندی می‌باشد که توسط کاربر تعیین می‌گردد.

تابع $l(\xi_i)$ به صورت زیر تعریف می‌شود:

$$l(\xi_i) = \begin{cases} 1 & \xi_i > 0 \\ 0 & \xi_i = 0 \end{cases}.$$

چون رابطه (۲۱.۲) شامل تابع ناپیوسته $l(\cdot)$ است، لذا بهینه سازی آن مشکل است؛ بنابراین صورت اولیه به صورت زیر تغییر می‌کند:

$$\begin{aligned} \text{Minimize } J(W, b, \xi) &= \frac{1}{\eta} \|W\|^2 + C \sum_{i=1}^M \xi_i, \\ \text{Subject to } \begin{cases} y_i[W^T Z_i + b] \geq 1 - \xi_i & , i = 1, 2, \dots, M \\ \xi_i \geq 0 & , i = 1, 2, \dots, M \end{cases}. \end{aligned} \quad (22.2)$$

برای حل رابطه (۲۲.۲) دوباره ضرایب لاگرانژ را به کار می‌بریم و به معادله لاگرانژ اولیه زیر دست می‌یابیم:

$$L_p(W, b, \xi, \alpha, \beta) = \frac{1}{\eta} W^T W + C \sum_{i=1}^M \xi_i - \sum_{i=1}^M \alpha_i [y_i(W^T Z_i + b) - 1 + \xi_i] - \sum_{i=1}^M \beta_i \xi_i. \quad (23.2)$$

^۱Trade off

که در آن α و β بردارهای ضرایب لاگرانژ هستند. همچنین برای حل رابطه (۲۳.۲) شرایط KKT را به صورت زیر داریم:

$$\begin{aligned}\frac{\partial L_p(W, b, \xi, \alpha, \beta)}{\partial W} &= 0, \\ \frac{\partial L_p(W, b, \xi, \alpha, \beta)}{\partial b} &= 0, \\ \frac{\partial L_p(W, b, \xi, \alpha, \beta)}{\partial \xi} &= 0.\end{aligned}\quad (24.2)$$

از حل معادلات (۲۴.۲) نتایج زیر به دست می‌آید:

$$W = \sum_{i=1}^M \alpha_i y_i Z_i,$$

$$\sum_{i=1}^M \alpha_i y_i = 0,$$

$$\alpha_i + \beta_i = C. \quad (25.2)$$

با جایگذاری روابط (۲۵.۲) در رابطه (۲۳.۲)، به صورت دوگان معادله لاگرانژ دست می‌یابیم که معادله اساسی SVM در حالت خطی جداناپذیر می‌باشد و به صورت زیر است:

$$\begin{aligned}Maximize L_d(\alpha) &= \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M y_i y_j \alpha_i \alpha_j Z_i^T Z_j, \\ Subject to &\begin{cases} 0 \leq \alpha_i \leq C \\ \sum_{i=1}^M \alpha_i y_i = 0 \end{cases}.\end{aligned}\quad (26.2)$$

بنابراین صورت نهایی این مسئله، مشابه سیستم‌های خطی جداناپذیر می‌باشد و تنها در کران بالای ضرایب لاگرانژ تفاوت دارد.

۳.۳.۲ ماشین بردار پشتیبان در سیستم‌های غیرخطی

در بخش قبل ماشین بردار پشتیبان خطی را معرفی کردیم، در این بخش حالت غیرخطی و همچنین تعمیم‌پذیری ماشین بردار پشتیبان خطی به ماشین بردار پشتیبان غیرخطی را معرفی خواهیم کرد. ایده ماشین بردار پشتیبان غیرخطی زمانی به وجود آمد که در بسیاری از مسائل رگرسیونی و رده‌بندی، ماشین بردار خطی ناکارا بود. چرا که در این سیستم، داده‌ها باید به‌طور خطی جداناپذیر باشند. به این ترتیب مطالعات زیادی برای حل این مشکل انجام گرفت.

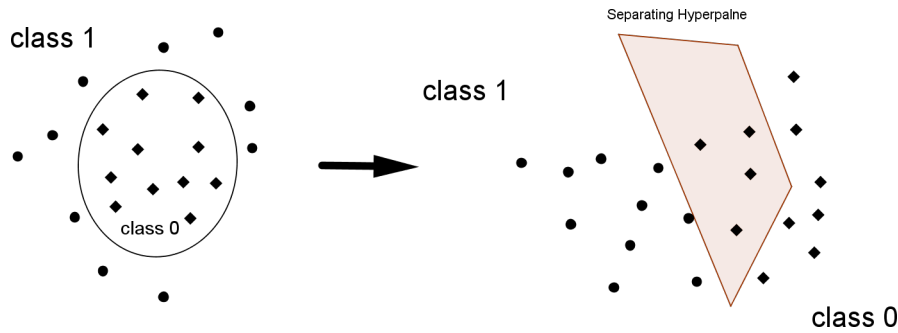
سرانجام نظریه مرسر^۱ برای حل این مشکل ارائه شد و SVM برای مسائل غیرخطی مورد استفاده قرار گرفت (مرسر، ۱۹۰۹). ایده اصلی این نظریه، انتقال برداری نظیر Z از فضای محدود (فضای ورودی) به فضای بالاتر (فضای ویژگی یا فضای هیلبرت^۲) با استفاده از تبدیل هیلبرت و رده‌بندی آن

^۱Mercer

^۲Hilbert

در فضای بالا بود.

شکل ۱۱.۲ فضای ورودی و فضای ویژگی را نشان می‌دهد. در فضای ویژگی، بردار \underline{Z} را با $\varphi(\underline{Z})$ نشان می‌دهند، لذا با جایگذاری $\varphi(\underline{Z})$ به جای \underline{Z} در روابط (۱۸.۲) و (۲۰.۲)، W و b بهینه ماشین بردار پشتیبان غیرخطی به صورت زیر خواهد بود:



شکل ۱۱.۲: انتقال داده‌ها از فضای ورودی (سمت چپ) به فضای ویژگی (سمت راست)

$$W = \sum_{i=1}^M \alpha_i \varphi(\underline{Z}_i) y_i . \quad (27.2)$$

$$b = \frac{1}{SV} \sum_{i=1}^{SV} (y_i - W^T \varphi(\underline{Z}_i)) .$$

همچنین تابع تصمیم ماشین بردار پشتیبان غیرخطی به صورت زیر تعریف می‌شود:

$$d(\underline{Z}) = \text{sign}(W^T \varphi(\underline{Z}) + b) .$$

که با جایگذاری مقدار بهینه W در رابطه فوق، تابع تصمیم به صورت زیر خواهد بود:

$$d(\underline{Z}) = \text{sign} \left(\sum_{i=1}^{SV} \alpha_i y_i \varphi(\underline{Z}_i)^T \varphi(\underline{Z}) + b \right) .$$

در این صورت

$$\underline{Z} \in \begin{cases} \text{رده ۱} & , d(\underline{Z}) > 0 \\ \text{رده ۰} & , d(\underline{Z}) < 0 \end{cases} ,$$

و اگر $d(\underline{Z}) = 0$ باشد، آن‌گاه مشاهده \underline{Z} غیر قابل رده‌بندی خواهد بود.

۴.۳.۲ هسته

در سیستم‌های غیرخطی همانطور که بیان شد به جای \underline{Z} از $\varphi(\underline{Z})$ استفاده می‌شود و برای محاسبه تابع تصمیم آن، نیاز به محاسبه $\varphi(\underline{Z})$ و ضرب داخلی W و φ می‌باشد. از طرفی می‌دانیم که $\varphi(\underline{Z})$ در فضای بعد بالا هیلبرت قرار دارد و محاسبه آن پیچیده و زمان‌بر است؛ بنابراین مقدار W را نیز از رابطه (۲۷.۲) نمی‌توان به دست آورد. این مشکل، برای اولین بار توسط کورتس^۱ و وپنیک (۱۹۹۵) با معرفی

^۱Cortes

ایده ترفند هسته^۱ حل شد. در واقع در این ترفند، به جای ضرب داخلی $\varphi(\underline{Z}_i)^T \varphi(\underline{Z})$ که بسیار پیچیده است، می‌توان از تابع هسته^۲ استفاده کرد. با توجه به این ترفند داریم:

$$K(\underline{Z}_i, \underline{Z}) = \varphi(\underline{Z}_i)^T \varphi(\underline{Z}) .$$

توابع هسته این ویژگی را دارند که به جای استفاده از $\varphi(\underline{Z})$ که در فضای بعد بالا هیلبرت قرار دارد، از داده‌های ورودی \underline{Z} استفاده کنند. این ویژگی مهم باعث رفع مشکل محاسبه توابع در بعد بالا شده است. تابع تصمیم b و بهینه SVM در حالت کلی به صورت زیر تعریف می‌گردد:

$$d(\underline{Z}) = \text{sign} \left(\sum_{i=1}^{SV} \alpha_i y_i K(\underline{Z}_i, \underline{Z}) + b \right),$$

که در آن

$$b = \frac{1}{SV} \sum_{i=1}^{SV} (y_i - \alpha_i y_i K(\underline{Z}_i, \underline{Z})) .$$

در جدول ۱.۲ تابع هسته‌های مختلف معرفی شده‌اند.

به طور خاص، تابع هسته سیگموئید، یک تابع هسته نیست، (اما تحت مقادیر مطمئن a و b در شرایط مرسر صدق می‌کند) با این حال این تابع بسیار محبوب در بعضی شرایط همچون شبکه‌های عصبی دو لایه^۳ است. توابع هسته گوسی و لاپلاسی از نوع هسته‌های مانا می‌باشند که صورت کلی آن‌ها به صورت $K(\underline{Z}_i, \underline{Z}) = K(\underline{Z}_i - \underline{Z})$ هسته چند جمله‌ای نیز از نوع هسته غیر مانا می‌باشد.

جدول ۱.۲: توابع هسته، $K(\underline{Z}_i, \underline{Z})$ ، پارامتر مقایس و مقادیر مثبت را اختیار می‌کند، $a, b \geq 0$ و d عدد صحیح و $|\underline{Z}|^2 = \underline{Z}^T \underline{Z}$ نرم اقلیدسی می‌باشد.

تایع هسته	نوع هسته
$\underline{Z}_i^T \underline{Z}$	خطی
$(\underline{Z}_i^T \underline{Z} + 1)^d$	چند جمله‌ای درجه d
$\exp \left(- \frac{\ \underline{Z}_i - \underline{Z}\ ^2}{2\sigma^2} \right)$	تابع پایه‌ای شعاعی گوسی
$\exp \left(- \frac{\ \underline{Z}_i - \underline{Z}\ }{\sigma} \right)$	لاپلاسی
$\tanh(a \underline{Z}_i^T \underline{Z} + b)$	سیگموئید

باید توجه داشت که به طور حتم مشخص نیست که از کدام تابع هسته باید استفاده کرد و معمولاً انتخاب نوع تابع هسته توسط دانش و مطالعات پیشین انجام می‌گیرد. با این حال زمانی که اطلاعاتی را در اختیار نداریم، توصیه شده است که از تابع هسته تابع پایه‌ای شعاعی گوسی^۴ (RBF) یک پارامتری

^۱Kernel trick

^۲Kernel function

^۳Two-layers neural networks

^۴Gaussian radial basis function

(σ) یا تابع هسته چندجمله‌ای درجه پایین ($d = 1$ یا 2) استفاده شود. در صورت لزوم، می‌توان تابع هسته‌های پیچیده دیگر را نیز به‌کار برد و نتایج را مقایسه کرد. لازم به ذکر است، برخی از کتب یادگیری ماشین، پارامتر تابع هسته RBF را با γ نمایش می‌دهند که معادل با $1/2\sigma^2$ است.

فصل ۳

روش شناسی

۱.۳ ساختار فرآیند و تغییر در واریانس

فرض کنید فرآیند چندمتغیره، ابتدا تحت کنترل بوده و متغیرهای توضیحی از توزیع نرمال با بردار میانگین μ و ماتریس واریانس کوواریانس Σ_0 پیروی کنند که این ساختار کوواریانسی، بعد از زمان نامعلوم $\tau + 1$ به Σ_1 تغییر می‌یابد. همچنین، فرض کنید که تغییر ایجاد شده در ساختار کوواریانسی فرآیند، از نوع پله‌ای یگانه^۱ است، لذا ماتریس واریانس کوواریانس، زمانی که فرآیند خارج از کنترل باشد، به صورت زیر تعریف می‌شود:

$$\Sigma_1 = \delta \Sigma_0 .$$

که در آن δ یک مقدار ثابت و Σ_0 دارای ساختار زیر است:

$$\Sigma_0 = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \dots & \sigma_{1,p} \\ \sigma_{2,1} & \sigma_{2,2} & \dots & \dots & \sigma_{2,p} \\ \vdots & \vdots & \sigma_{3,3} & \dots & \sigma_{3,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{p,1} & \sigma_{p,2} & \dots & \dots & \sigma_{p,p} \end{bmatrix} .$$

در حالت کلی، یک نمونه تصادفی n تایی چندمتغیره را با X_1, \dots, X_n نمایش می‌دهند؛ اما برای کنترل یک فرآیند، بایستی در هر یک از زمان‌های بازبینی، یک نمونه تصادفی n تایی گرفته شود؛ لذا به یک اندیس اضافی برای نماد گذاری نیاز داریم. برای مثال، می‌توان نمونه تصادفی چندمتغیره را برای زمان i به صورت $X_{i,1}, \dots, X_{i,n}$ بیان کرد. اگر نمونه تصادفی از توزیع نرمال p متغیره تبعیت کند، می‌توان

^۱Single step

نوشت:

$$\begin{aligned}
\underline{X}_{1,1}, \dots, \underline{X}_{1,n} &\stackrel{iid}{\sim} N_p(\underline{\mu}, \Sigma_0), \\
\underline{X}_{2,1}, \dots, \underline{X}_{2,n} &\stackrel{iid}{\sim} N_p(\underline{\mu}, \Sigma_0), \\
&\vdots \\
\underline{X}_{\tau,1}, \dots, \underline{X}_{\tau,n} &\stackrel{iid}{\sim} N_p(\underline{\mu}, \Sigma_0), \\
\underline{X}_{\tau+1,1}, \dots, \underline{X}_{\tau+1,n} &\stackrel{iid}{\sim} N_p(\underline{\mu}, \Sigma_1), \\
&\vdots \\
\underline{X}_{T,1}, \dots, \underline{X}_{T,n} &\stackrel{iid}{\sim} N_p(\underline{\mu}, \Sigma_1),
\end{aligned}$$

که در آن $N_p(\cdot)$ نماد توزیع نرمال p متغیره و iid به معنای مستقل و هم توزیع بودن متغیرها است. همچنین

$$\underline{X}_{i,j} = \begin{pmatrix} X_{i,j,1} \\ X_{i,j,2} \\ \vdots \\ X_{i,j,p} \end{pmatrix}_{p \times 1}.$$

یک بردار p تایی از متغیرهاست که ماتریس مشاهدات برای زمان i را می‌توان به صورت

$$\begin{pmatrix} X'_{i,1} \\ X'_{i,2} \\ \vdots \\ X'_{i,n} \end{pmatrix} = \begin{pmatrix} X_{i,1,1} & X_{i,1,2} & \cdots & X_{i,1,p} \\ X_{i,2,1} & X_{i,2,2} & \cdots & X_{i,2,p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{i,n,1} & X_{i,n,2} & \cdots & X_{i,n,p} \end{pmatrix}.$$

نمایش داد. ساختار کوواریانس فرآیند از زمان نامعلوم $\tau + 1$ ، از Σ_0 به Σ_1 تغییر کرده است. زمان T ، اولین زمانی است که نمودار کنترل $|S|$ ، هشدار خارج از کنترل را اعلام می‌کند. از آن جا که Σ توسط ماتریس واریانس کوواریانس نمونه‌ای S برآورد می‌شود، لذا برای هر یک از زمان‌های بازبینی، نظیر زمان i ام، می‌توان نوشت:

$$\begin{aligned}
S_i &= \frac{1}{n-1} \sum_{j=1}^n (\underline{X}_{i,j} - \bar{\underline{X}}_i)(\underline{X}_{i,j} - \bar{\underline{X}}_i)', \quad i = 1, 2, \dots, T, \\
&= \begin{bmatrix} S_{i,1,1} & S_{i,1,2} & \cdots & \cdots & S_{i,1,p} \\ S_{i,2,1} & S_{i,2,2} & \cdots & \cdots & S_{i,2,p} \\ \vdots & \vdots & S_{i,3,3} & \cdots & S_{i,3,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{i,p,1} & S_{i,p,2} & \cdots & \cdots & S_{i,p,p} \end{bmatrix}. \quad (1.3)
\end{aligned}$$

که در آن

$$\begin{aligned}\bar{X}_i &= \begin{pmatrix} \frac{1}{n} \sum_{j=1}^n X_{i,j,1} \\ \frac{1}{n} \sum_{j=1}^n X_{i,j,2} \\ \vdots \\ \frac{1}{n} \sum_{j=1}^n X_{i,j,p} \end{pmatrix}, \\ &= \begin{pmatrix} \bar{X}_{i,1} \\ \bar{X}_{i,2} \\ \vdots \\ \bar{X}_{i,p} \end{pmatrix}.\end{aligned}$$

۲.۳ نمودار کنترل $|S|$

همانطور که بیان شد، مقدار T با توجه به نمودار کنترل $|S|$ به دست می‌آید. بدین منظور نمودار کنترل $|S|$ را تشریح می‌کنیم.

نمودار کنترل $|S|$ ، یکی از ابزارهای SPC می‌باشد و در مواقعی که تغییرپذیری واریانس فرآیند، مورد بحث است، به کار می‌رود. در این نمودارها، به طور معمول از سطح اطمینان $\alpha = 0.01$ برای تعیین حدود کنترل استفاده می‌شود؛ به عبارت دیگر، فاصله اطمینان 99% را معیار به دست آوردن حدود کنترل خود قرار می‌دهند؛ برای این منظور، یک فاصله اطمینان 99% برای پارامتر θ به صورت زیر تعریف می‌گردد.

$$\hat{\theta} \pm 3\sqrt{\text{var}(\hat{\theta})}.$$

در حالت خاص، برای نمودار کنترل $|S|$ ، حدود کنترل عبارتند از:

$$E(|S|) \pm 3\sqrt{\text{var}(|S|)}.$$

توزیع دترمینان ماتریس واریانس کواریانس نمونه را می‌توان به صورت زیر معرفی کرد (اندرسون، ۱۹۸۴):

$$|S| \sim \frac{|\Sigma|}{(n-1)^p} U_1 U_2 \dots U_p. \quad (2.3)$$

که در آن متغیرهای U_k ، $k = 1, 2, \dots, p$ مستقل از هم و دارای توزیع کی-دو با $n - k$ درجه آزادی هستند.

با توجه به رابطه (۲.۳)، گشتاور مرتبه r دترمینان ماتریس واریانس کواریانس نمونه به صورت زیر است

(اندرسون، ۱۹۸۴):

$$\alpha_r = E(|S|^r) = \left(\frac{\nu}{n-1}\right)^{rp} |\Sigma|^r \prod_{k=1}^p \frac{\Gamma\left(r + \frac{n-k}{\nu}\right)}{\Gamma\left(\frac{n-k}{\nu}\right)}.$$

لذا با توجه به این که $\Gamma(\beta + 1) = \beta \Gamma(\beta)$ ، گشتاورهای مراتب اول و دوم عبارتند از:

$$\begin{aligned} \alpha_1 = E(|S|) &= \left(\frac{\nu}{n-1}\right)^p |\Sigma| \prod_{k=1}^p \frac{\Gamma\left(1 + \frac{n-k}{\nu}\right)}{\Gamma\left(\frac{n-k}{\nu}\right)}, \\ &= \left(\frac{\nu}{n-1}\right)^p |\Sigma| \prod_{k=1}^p \frac{\left(\frac{n-k}{\nu}\right) \Gamma\left(\frac{n-k}{\nu}\right)}{\Gamma\left(\frac{n-k}{\nu}\right)}, \\ &= \frac{|\Sigma|}{(n-1)^p} \prod_{k=1}^p (n-k). \end{aligned}$$

$$\begin{aligned} \alpha_\nu = E(|S|^\nu) &= \left(\frac{\nu}{n-1}\right)^{\nu p} |\Sigma|^\nu \prod_{k=1}^p \frac{\Gamma\left(\nu + \frac{n-k}{\nu}\right)}{\Gamma\left(\frac{n-k}{\nu}\right)}, \\ &= \left(\frac{\nu}{n-1}\right)^{\nu p} |\Sigma|^\nu \prod_{k=1}^p \frac{\left(\frac{n-k}{\nu} + 1\right) \left(\frac{n-k}{\nu}\right) \Gamma\left(\frac{n-k}{\nu}\right)}{\Gamma\left(\frac{n-k}{\nu}\right)}, \\ &= \frac{|\Sigma|^\nu}{(n-1)^{\nu p}} \prod_{k=1}^p (n-k + \nu)(n-k). \end{aligned}$$

بنابراین می‌توان واریانس دترمینان ماتریس واریانس کواریانس را به صورت زیر نوشت:

$$\begin{aligned} \text{var}(|S|) &= \alpha_2 - \alpha_1^2, \\ &= \left[\frac{|\Sigma|^2}{(n-1)^{2p}} \prod_{k=1}^p (n-k+2)(n-k) \right] - \left[\frac{|\Sigma|^2}{(n-1)^{2p}} \prod_{k=1}^p (n-k)^2 \right], \\ &= \frac{|\Sigma|^2}{(n-1)^{2p}} \prod_{k=1}^p (n-k) \left(\prod_{k=1}^p (n-k+2) - \prod_{k=1}^p (n-k) \right). \end{aligned}$$

اگر b_1 و b_2 را به صورت زیر تعریف کنیم:

$$\begin{aligned} b_1 &= \frac{1}{(n-1)^p} \prod_{k=1}^p (n-k), \\ b_2 &= \frac{1}{(n-1)^{2p}} \prod_{k=1}^p (n-k) \left(\prod_{k=1}^p (n-k+2) - \prod_{k=1}^p (n-k) \right). \end{aligned}$$

و $|\Sigma_0|$ ، دترمینان ماتریس واریانس کواریانس فرآیند، زمانی که تحت کنترل است، باشد، آنگاه حدود کنترل نمودار $|S|$ عبارتند از:

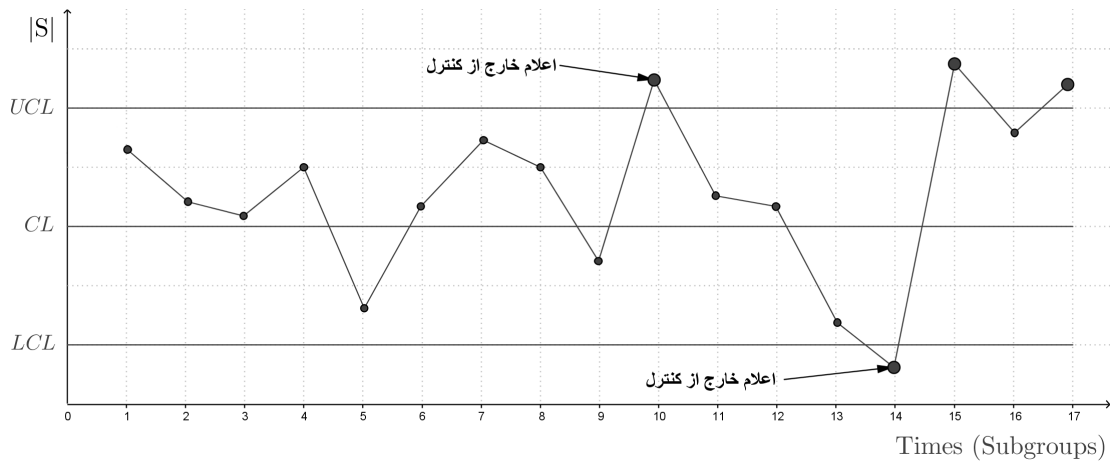
$$\begin{aligned} CL &= |\Sigma_0| b_1, \\ UCL &= |\Sigma_0| (b_1 + 3\sqrt{b_2}), \\ LCL &= |\Sigma_0| (b_1 - 3\sqrt{b_2}). \end{aligned}$$

که در آن CL ، UCL و LCL به ترتیب، خط میانی، حد کنترل بالایی و حد کنترل پایینی نمودار هستند. در عمل $|\Sigma_0|$ مجهول است و به جای آن از برآوردگر نارایب $|\hat{S}|/b_1$ استفاده می‌شود. در این صورت اگر مقدار $|S|$ در هر یک از زمان‌های مورد بررسی خارج از کنترل واقع شود، به معنای آن است که فرآیند، خارج از کنترل می‌باشد. شکل ۱.۳ نمایش شماتیکی نمودار $|S|$ را نشان می‌دهد که در آن هر یک از نقاط واقع در نمودار، مقدار دترمینان ماتریس واریانس کواریانس نمونه‌ای به صورت (۱.۳) هستند.

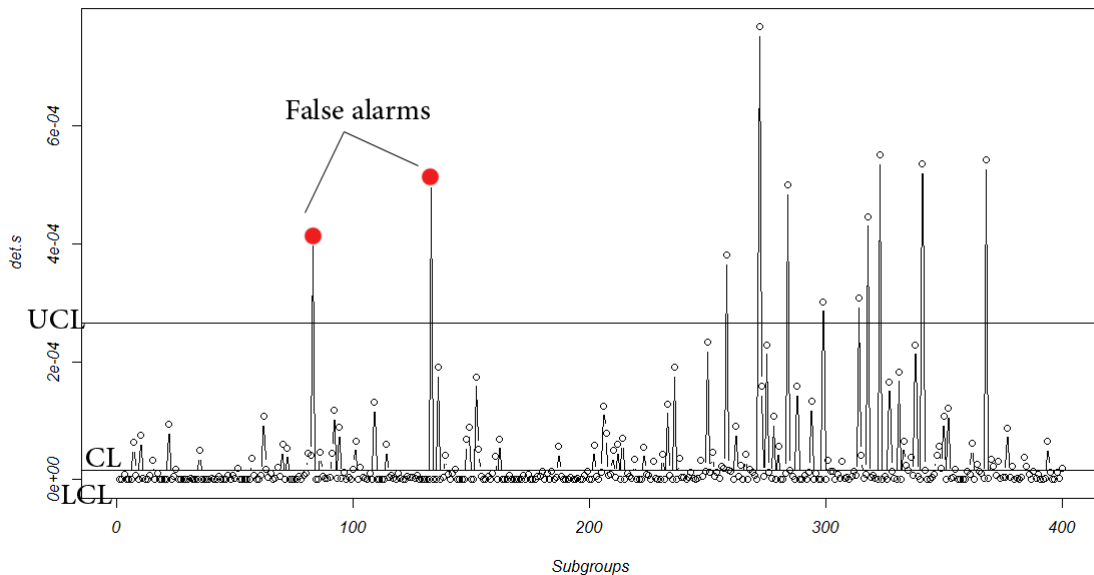
۱.۲.۳ هشدار غلط

گاهی اوقات این امکان وجود دارد که نمودار کنترل، خارج از کنترل بودن را زودتر از نقطه تغییر واقعی، اعلام کند. به این نوع اعلام، هشدار غلط^۱ گویند. این هشدارها به طور معمول توسط تجربه مهندسين فرآیند تشخیص داده می‌شوند. برای مثال شکل ۲.۳ که نتیجه یک شبیه‌سازی با $n = 12$ ، $p = 10$ و نقطه تغییر $201 = \tau + 1$ است را در نظر بگیرید که قطر اصلی Σ_0 برابر ۱ و بقیه عناصر برابر 0.5 و $\Sigma_1 = 1/1 \Sigma_0$ است. همان‌طور که مشخص است، نقاط خارج از حد کنترل بالایی در زمان‌های قبل از $201 = \tau + 1$ هشدار غلط هستند.

^۱ False alarm



شکل ۱.۳: نمودار کنترل $|S|$



شکل ۲.۳: هشدارهای غلط اعلام شده توسط نمودار کنترل

۲.۲.۳ تعریف متغیرهای جدید Z و Y

از آنجا که هدف ما، برآورد زمان تغییر در واریانس فرآیند چندمتغیره است، لذا راهبرد روش پیشنهادی آن است که از مؤلفه‌های ساختار کوواریانسی در هر زمان، برای تعیین (برآورد) نقطه تغییر استفاده گردد. از آنجا که ماتریس واریانس کوواریانس متقارن است؛ بنابراین فقط از عناصر روی قطر اصلی و بالای قطر استفاده می‌کنیم؛ این عناصر را به‌عنوان متغیرهای جدید Z که ماهیت کوواریانسی دارند برمی‌گزینیم. برای فهم بهتر، متغیرهای Z را در حالت $p = 3$ تعریف می‌کنیم؛ سپس این تعریف را برای حالت کلی تعمیم می‌دهیم.

برای بردار متغیر توضیحی زیر ($p = 3$)

$$\underline{X}_{i,j} = \begin{pmatrix} X_{i,j,1} \\ X_{i,j,2} \\ X_{i,j,3} \end{pmatrix} .$$

مجموعه همهی ماتریس‌های واریانس کوواریانس S_i , $i = 1, \dots, T$ را به صورت زیر در نظر بگیرد:

$$S_1 = \begin{bmatrix} s_{1,1,1} & s_{1,1,2} & s_{1,1,3} \\ s_{1,2,1} & s_{1,2,2} & s_{1,2,3} \\ s_{1,3,1} & s_{1,3,2} & s_{1,3,3} \end{bmatrix}, S_2 = \begin{bmatrix} s_{2,1,1} & s_{2,1,2} & s_{2,1,3} \\ s_{2,2,1} & s_{2,2,2} & s_{2,2,3} \\ s_{2,3,1} & s_{2,3,2} & s_{2,3,3} \end{bmatrix}, \dots, S_T = \begin{bmatrix} s_{T,1,1} & s_{T,1,2} & s_{T,1,3} \\ s_{T,2,1} & s_{T,2,2} & s_{T,2,3} \\ s_{T,3,1} & s_{T,3,2} & s_{T,3,3} \end{bmatrix} .$$

متغیرهای جدید Z را به صورت زیر تعریف می‌کنیم:

درآیه‌های (۱, ۱) همهی این ماتریس‌ها، مقادیر متغیر Z_1 ,

درآیه‌های (۱, ۲) همهی این ماتریس‌ها، مقادیر متغیر Z_2 ,

درآیه‌های (۱, ۳) همهی این ماتریس‌ها، مقادیر متغیر Z_3 ,

درآیه‌های (۲, ۲) همهی این ماتریس‌ها، مقادیر متغیر Z_4 ,

درآیه‌های (۲, ۳) همهی این ماتریس‌ها، مقادیر متغیر Z_5 ,

درآیه‌های (۳, ۳) همهی این ماتریس‌ها، مقادیر متغیر Z_6 ,

و درمیان همهی این ماتریس‌ها، مقادیر متغیر Z_7 را تشکیل می‌دهند.

نمادگذاری جدید را می‌توان در قالب جدول زیر خلاصه کرد:

جدول ۱۰۳: جدول مشاهدات متغیر جدید Z

Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7
$z_{1,1} = s_{1,1,1}$	$z_{1,2} = s_{1,1,2}$	$z_{1,3} = s_{1,1,3}$	$z_{1,4} = s_{1,2,2}$	$z_{1,5} = s_{1,2,3}$	$z_{1,6} = s_{1,3,3}$	$z_{1,7} = s_1 $
$z_{2,1} = s_{2,1,1}$	$z_{2,2} = s_{2,1,2}$	$z_{2,3} = s_{2,1,3}$	$z_{2,4} = s_{2,2,2}$	$z_{2,5} = s_{2,2,3}$	$z_{2,6} = s_{2,3,3}$	$z_{2,7} = s_2 $
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$z_{T,1} = s_{T,1,1}$	$z_{T,2} = s_{T,1,2}$	$z_{T,3} = s_{T,1,3}$	$z_{T,4} = s_{T,2,2}$	$z_{T,5} = s_{T,2,3}$	$z_{T,6} = s_{T,3,3}$	$z_{T,7} = s_T $

با تعمیم این تعریف، می‌توان متغیرهای Z را برای حالت p متغیره نیز تعریف نمود:

$$Z_{i,1} = s_{i,1,1}, \quad Z_{i,2} = s_{i,1,2}, \quad \dots, \quad Z_{i,p} = s_{i,1,p}, \quad Z_{i,p+1} = s_{i,2,2},$$

$$Z_{i,p+2} = s_{i,2,3}, \quad \dots, \quad Z_{i,N-1} = s_{i,p,p}, \quad Z_{i,N} = |S_i|, \quad i = 1, \dots, T .$$

که $N = \frac{p(p+1)}{2} + 1$ یکی بیشتر از تعداد عناصر روی قطر اصلی و بالای قطر ماتریس S_i است. طبق

تعریف فوق، تعداد متغیرهای Z برابر $\frac{p(p+1)}{2} + 1$ می‌باشد. مشاهدات این متغیرها نیز همانند حالت

۳ متغیره تعریف می‌گردند. لذا:

درآیه‌های (۱, ۱) ماتریس‌های S_i , $i = 1, \dots, T$, مقادیر متغیر Z_1 ,

درآیه‌های (۱, ۲) ماتریس‌های S_i , $i = 1, \dots, T$, مقادیر متغیر Z_2 ,

\vdots

\vdots
 درآیه‌های $(1, p)$ ماتریس‌های $T, i = 1, \dots, T$ ، S_i ، مقادیر متغیر Z_p ،
 درآیه‌های $(2, 2)$ ماتریس‌های $T, i = 1, \dots, T$ ، S_i ، مقادیر متغیر Z_{p+1} ،
 درآیه‌های $(2, 3)$ ماتریس‌های $T, i = 1, \dots, T$ ، S_i ، مقادیر متغیر Z_{p+2} ،
 \vdots
 \vdots
 درآیه‌های (p, p) ماتریس‌های $T, i = 1, \dots, T$ ، S_i ، مقادیر متغیر Z_{N-1} ،
 و درمیان ماتریس‌های $T, i = 1, \dots, T$ ، S_i ، مقادیر متغیر Z_N را تشکیل می‌دهند.
 اکنون فرض کنید، نقطه تغییر واقعی در $\tau + 1$ رخ داده است؛ لذا متغیر پاسخ Y را به صورت زیر
 تعریف می‌کنیم.

$$y = \begin{pmatrix} \circ \\ \vdots \\ \circ \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{T \times 1}$$

که مقادیر آن تا زمان τ برابر صفر (فرآیند تحت کنترل) و از زمان $\tau + 1$ تا زمان T ، برابر ۱ (فرآیند خارج از کنترل) هستند.

با در نظر داشتن مجموعه Z_1, \dots, Z_N به عنوان متغیرهای توضیحی و Y به عنوان متغیر پاسخ، موضوع مورد بررسی ما تبدیل به یک مسئله رده‌بندی خواهد شد. بدین منظور مدل‌های مختلف رده‌بندی آماری و یادگیری ماشین می‌توانند مورد استفاده قرار گیرند. نتیجه اعمال هر مدل رده‌بندی، بردار برآورد \hat{y} را در پی خواهد داشت که از آن‌ها برای تعیین نقطه تغییر فرآیند توسط راهبرد زیر بهره می‌جوییم.

از طرفی باید توجه داشت که برای اعتبار بخشی به نتایج حاصل از برخی روش‌های رده‌بندی و رگرسیونی، مانند روش‌های رگرسیون معمولی و رگرسیون لجستیک، نیاز است که قبل از اعمال این‌گونه روش‌ها، ارزیابی هم‌خطی چندگانه روی متغیرهای توضیحی صورت گیرد.

۳.۳ راهبرد شناسایی نقطه تغییر

فرض کنید نمودار کنترل $|S|$ ، نقطه زمانی T را به عنوان اولین زمان خارج از کنترل بودن فرآیند، نشان دهد که در واقع، نقطه تغییر واقعی ممکن است در T یا زمان‌های قبل از آن باشد. به کار بردن هر مدل رده‌بندی برای زمان‌های 1 تا $T - 1$ (به صورت پس‌رو) می‌تواند وضعیت فرآیند را در خصوص تحت کنترل بودن یا خارج از کنترل بودن بیان کند. به طور کلی اگر خروجی مدل را برای زمان‌های $T - 1$ تا مثلاً $T - C$ ($1 \leq C \leq T - 1$) برابر یک باشد، با قطعیت می‌توان گفت که $T - C$ نقطه تغییر واقعی است؛ اما اگر در بعضی از زمان‌های $T - 1, T - 2, T - 3, \dots$ و $T - C$ خروجی مدل

برابر صفر باشد (جدول ۲.۳)، چه تصمیمی باید گرفت؟

جدول ۲.۳: تعیین نقطه تغییر فرآیند توسط خروجی مدل رده‌بندی

خروجی مدل	زمان
۰	۱
۰	۲
⋮	⋮
۰	$T - C - ۱$
۱	$T - C$
۰	$T - C + ۱$
۱	$T - C + ۲$
⋮	⋮
۱	$T - ۳$
۰	$T - ۲$
۰	$T - ۱$

راهبردی برای حل این مشکل و تعیین نقطه تغییر فرآیند مطرح شده است که به قرار زیر است: هر یک از زمان‌های ۱، ۲، ۳، ...، $T - ۱$ می‌توانند کاندیدی برای برآورد نقطه تغییر محسوب شوند، به‌عنوان مثال، برای پاسخ به این سوال که آیا زمان $T - ۳$ ، اولین زمان ایجاد تغییر در فرآیند است، خروجی مدل رده‌بندی را در زمان‌های $T - ۱$ ، $T - ۲$ و $T - ۳$ در نظر می‌گیریم. از آن‌جا که متغیر پاسخ دو رده‌ای است، لذا نتیجه رده‌بندی در هر یک از زمان‌های فوق، می‌تواند به‌صورت یک آزمایش برنولی تلقی شود که موفقیت به معنای "خارج از کنترل بودن" (خروجی ۱ در مدل رده‌بندی) و p_s معرف احتمال موفقیت است که می‌توان آن را معادل با دقت رده‌بندی مدل دانست؛ در نتیجه تعداد آزمایش‌های برنولی مستقل از هم، برابر $M_3 = ۳$ است.

با تعریف متغیر تصادفی X_3^* که معرف تعداد موفقیت در M_3 آزمایش برنولی مستقل از هم است، می‌توان نوشت:

$$X_3^* \sim B(M_3, p_s),$$

به عبارت دیگر X_3^* ، تعداد دفعاتی است که خروجی مدل رده‌بندی برابر ۱ است. اگر مقدار تجربی X_3^* را با نماد n_3^* نشان دهیم که $n_3^* \in \{0, 1, 2, 3\}$ ، می‌توان گفت که هرچه تعداد ۱‌ها در این ۳ آزمایش بیشتر باشند، با اطمینان بیشتری می‌توان پذیرفت که $T - ۳$ نقطه تغییر است. این مطلب معادل آن است که هرچه احتمال تجمعی

$$P(X_3^* \leq n_3^*) = \sum_{x_3^*=0}^{n_3^*} \binom{M_3}{x_3^*} P_s^{x_3^*} (1 - P_s)^{M_3 - x_3^*}.$$

بیشتر باشد، اعتماد پذیرش نقطه $T - ۳$ به‌عنوان نقطه تغییر بیشتر است.

به طور کلی، برای آزمون این که زمان $T - i$ ، نقطه تغییر است، نتیجه دودویی خروجی مدل رده بندی را از زمان $T - 1$ تا $T - i$ ($i = 1, \dots, T - 1$) به عنوان $M_i = i$ آزمایش برنولی مستقل فرض می کنیم. متغیر تصادفی X_i^* را تعداد موفقیت در M_i آزمایش برنولی مستقل در نظر می گیریم ($X_i^* \sim B(M_i, p_s)$). با علم این که تکیه گاه این متغیر تصادفی برابر با $\{0, \dots, M_i\}$ و n_i^* معرف تعداد موفقیت تجربی در M_i آزمایش برنولی مستقل باشند؛ آنگاه بزرگ بودن احتمال تجمعی زیر بیان گر آن است که اعتماد ما برای آن که زمان $T - i$ نقطه تغییر باشد، بیشتر است.

$$P(X_i^* \leq n_i^*) = \sum_{x_i^*=0}^{n_i^*} \binom{M_i}{x_i^*} P_s^{x_i^*} (1 - P_s)^{M_i - x_i^*}, \quad i = 1, \dots, T - 1. \quad (3.3)$$

اما "بزرگ بودن" نسبی است، لذا به یک حد آستانه ای L برای تشخیص بزرگی نیاز داریم؛ بنابراین قاعده تصمیم برای این که زمان $T - i$ ، نقطه تغییر باشد، عبارت است:

$$P(X_i^* \leq n_i^*) \geq L, \quad i = 1, 2, \dots, T - 1. \quad (4.3)$$

جدول ۳.۳ مراحل شکل گیری احتمالات تجمعی را برای زمان های $T - 1$ تا 1 نشان می دهد. از آن جا که ممکن است قاعده فوق برای زمان های متعددی صدق کند، لذا نقطه تغییر را اولین آن زمان ها در نظر می گیریم؛ برای مثال اگر زمان های $T - 8$ ، $T - 10$ و $T - 13$ در رابطه فوق صدق کنند، نقطه تغییر فرآیند را زمان $T - 13$ اعلام می کنیم.

از آن جایی که کارایی این راهبرد، به انتخاب مدل رده بندی خوب وابسته است، لذا انتخاب مدل رده بندی مناسب بسیار مهم است.

حال این سوال پیش می آید که حد آستانه چگونه تعریف می گردد؟ چون هیچ نظریه برای محاسبه حد آستانه وجود ندارد، بنابراین با استفاده از تجربیات و آزمایشات شبیه سازی شده توسط محققان، حد آستانه را به صورت زیر تعریف کرده اند (شائو و ک-شان، ۲۰۱۵).

تعیین حد آستانه

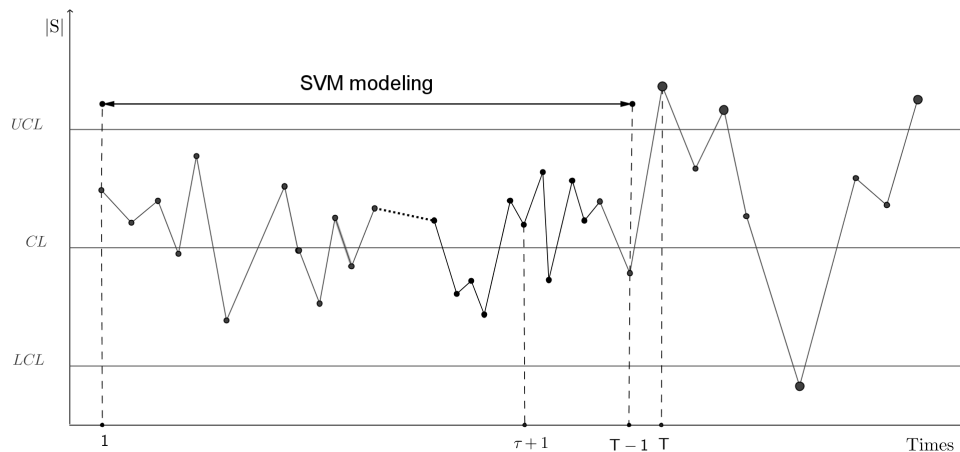
برای تعیین حد آستانه، ابتدا مدل رده بندی با $M \geq T - 1$ مشاهده اجرا می شود، سپس خروجی این مدل (\hat{Y})، به منظور تعیین حد آستانه، استفاده می گردد. چون خروجی متغیر پاسخ دو رده ای است، لذا آن را یک آزمایش برنولی در نظر می گیرند. با فرض این که متغیر تصادفی X_D به عنوان تعداد موفقیت (خارج از کنترل بودن فرآیند) در M آزمایش برنولی مستقل است، $E(X_D) = M \times p_s$ که به عنوان احتمال موفقیت و $n_s = [M \times p_s]$ یک عدد صحیح می باشند. در این صورت، احتمال تجمعی زیر به عنوان حد آستانه تعریف می گردد:

$$L = P(X_D \leq n_s) = \sum_{x_D=0}^{n_s} \binom{M}{x_D} P_s^{x_D} (1 - P_s)^{M - x_D}. \quad (5.3)$$

سوالی که در ذهن باقی می ماند آن است که چه مدل رده بندی ای مناسب است. یکی از مشهورترین مدل های رده بندی، مدل SVM می باشد؛ بنابراین در این پایان نامه، این مدل را برای تحت کنترل یا خارج از کنترل بودن فرآیند، به کار می بریم (شکل ۳.۳).

جدول ۳.۳: محاسبه احتمال تجمعی توزیع دوجمله‌ای برای زمان‌های ۱- تا T

آیا زمان ۱، برآورد نقطه تغییر است؟	آیا زمان ۳-، برآورد نقطه تغییر است؟	آیا زمان ۱-، برآورد نقطه تغییر است؟
۱	$T-3$	$T-2$	$T-1$	زمان‌های شامل شده
۲	$T-2$	$T-1$		
⋮	$T-1$			
$T-1$				
تعداد آزمایش‌های برنولی= $T-1$	تعداد آزمایش‌های برنولی= 3	تعداد آزمایش‌های برنولی= 2	تعداد آزمایش‌های برنولی= 1	تعبیر خروجی مدل رده‌بندی نمادها
$M_{T-1} = T-1$	$M_3 = 3$	$M_2 = 2$	$M_1 = 1$	
متغیر تصادفی = X_{T-1}^*	متغیر تصادفی = X_3^*	متغیر تصادفی = X_2^*	متغیر تصادفی = X_1^*	
تکیه‌گاه = $\{0, 1, \dots, T-1\}$	تکیه‌گاه = $\{0, 1, 2, 3\}$	تکیه‌گاه = $\{0, 1, 2\}$	تکیه‌گاه = $\{0, 1\}$	
تعداد موفقیت‌های = n_{T-1}^*	تعداد موفقیت‌های = n_3^*	تعداد موفقیت‌های = n_2^*	تعداد موفقیت‌های = n_1^*	
مشاهده شده	مشاهده شده	مشاهده شده	مشاهده شده	
احتمال موفقیت = p_s	احتمال موفقیت = p_s	احتمال موفقیت = p_s	احتمال موفقیت = p_s	
$X_{T-1}^* \sim B(T-1, p_s)$	$X_3^* \sim B(3, p_s)$	$X_2^* \sim B(2, p_s)$	$X_1^* \sim B(1, p_s)$	توزیع احتمال
$P(X_{T-1}^* \leq n_{T-1}^*)$	$P(X_3^* \leq n_3^*)$	$P(X_2^* \leq n_2^*)$	$P(X_1^* \leq n_1^*)$	احتمال تجمعی مربوطه

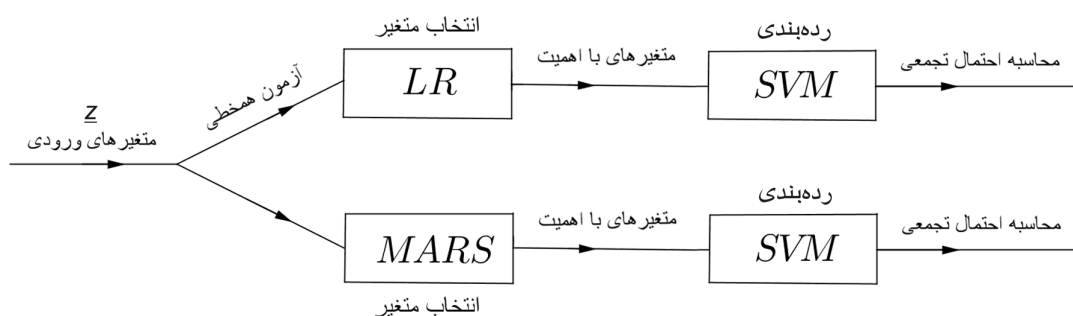


شکل ۳.۳: نمودار کنترل $|S|$ همراه نقطه تغییر واقعی و تعیین مدل سازی SVM

همان طور که در فصل اول بیان شد، مدل SVM یا دیگر مدل های ML، زمانی که تعداد متغیرهای توضیحی زیاد باشد، با کاهش کارایی مواجه خواهد شد؛ لذا بهتر است، متغیرهای توضیحی که تاثیر کمتری بر متغیر پاسخ می گذارند را حذف کنیم و سپس مدل SVM را اجرا کنیم. از این رو، رهیافت های ترکیبی برای حل این مشکل معرفی شده اند.

در رهیافت ترکیبی اول، به منظور تشخیص و انتخاب متغیرهای مهم، مدل آماری رگرسیون لجستیک (LR) پیشنهاد شده است که پس از مشخص شدن متغیرهای مهم، این متغیرها به عنوان متغیرهای ورودی مدل SVM در نظر گرفته می شوند.

اما برای ارزیابی این رهیافت ترکیبی، رهیافت دیگری را به طور موازی با رهیافت اول، در نظر می گیریم که در آن روش MARS جایگزین LR می شود. در شکل ۴.۳ ترکیب این روش ها به طور شماتیکی نشان داده شده است که آن ها را به ترتیب LR-SVM و MARS-SVM نام گذاری می کنیم.



شکل ۴.۳: رهیافت های ترکیبی $MARS - SVM$ و $LR - SVM$

بنابراین می توان چهار رهیافت نمودار کنترل $|S|$ ، روش SVM، روش های ترکیبی LR-SVM و MARS-SVM را به منظور تعیین (برآورد) نقطه تغییر، مورد ارزیابی قرار داده و در آخر، توسط نمودارهای اریبی^۱ و میانگین مربعات خطا^۲ MSE، این چهار رهیافت را مقایسه می کنیم.

^۱Bias

^۲Mean square error

معیار اریبی و MSE به صورت زیر تعریف می‌شوند:

$$\text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta,$$

$$\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = (E(\hat{\theta}) - \theta)^2 + \text{Var}(\hat{\theta}) = (\text{bias}(\hat{\theta}))^2 + \text{Var}(\hat{\theta}). \quad (۶.۳)$$

که در آن، $\theta = \tau + 1$ نقطه تغییر واقعی فرآیند است.

در نهایت برای ارزیابی عملکرد رهیافت‌های مطرح شده، تعداد متناهی مجموعه داده جدید تولید می‌کنیم. پس از تعیین نقطه تغییر توسط رهیافت‌های مذکور، میانگین و خطای معیار برآورد نقاط تغییر را ثبت می‌کنیم. همچنین به منظور تحلیل حساسیت رهیافت‌ها نسبت به شدت تغییر در فرآیند، مقادیر مختلفی را برای δ ($\Sigma_1 = \delta \Sigma_0$) در نظر می‌گیریم.

فصل ۴

مطالعات شبیه‌سازی

در این فصل می‌خواهیم با انجام شبیه‌سازی، عملکرد رهیافت‌های مطرح شده در فصل سوم را برای شناسایی نقطه تغییر بررسی کنیم. بدین منظور، حجم نمونه در هر زمان بازرسی را برابر ۱۲ ($n = 12$)، تعداد متغیرهای توضیحی نرمال چندمتغیره X را برابر ۱۰ ($p = 10$)، بردار میانگین را برابر صفر ($\mu = 0$)، ماتریس واریانس کوواریانس فرآیند، زمانی که تحت کنترل است را برابر

$$\Sigma_0 = \begin{bmatrix} 1 & 0.5 & \dots & \dots & 0.5 \\ 0.5 & 1 & 0.5 & \dots & 0.5 \\ \vdots & 0.5 & 1 & 0.5 & 0.5 \\ \vdots & \vdots & 0.5 & \ddots & 0.5 \\ 0.5 & \dots & \dots & 0.5 & 1 \end{bmatrix} .$$

و نقطه تغییر واقعی فرآیند را برابر ۲۰۱ ($\tau + 1 = 201$) در نظر می‌گیریم (بخش ۱.۳ را ببینید). چون تغییر ایجاد شده در فرآیند از نوع پله‌ای یگانه (بخش ۱.۲.۱) است، لذا ماتریس واریانس کوواریانس فرآیند، زمانی که خارج از کنترل باشد، به صورت زیر تعریف می‌شود:

$$\Sigma_1 = \delta \Sigma_0 .$$

از آنجا که نمودارهای کنترل شوهارت، زمانی که شدت تغییر در فرآیند کم است، در تعیین نقطه تغییر ناکارا هستند، لذا برای تحلیل حساسیت تغییرات پله‌ای فرآیند، مقادیر کوچک $\delta = 1/1$ ، $\delta = 1/125$ ، $\delta = 1/15$ و $\delta = 1/175$ را در نظر می‌گیریم. لازم به ذکر است که بر اساس شبیه‌سازی‌های انجام شده، به این نتیجه رسیدیم که اگر شدت تغییر زیاد باشد (مقدار δ بزرگ باشد) نمودارهای کنترل کارا بوده و به خوبی می‌توانند اختلال نفوذ کرده در فرآیند را تشخیص دهند.

برای انتخاب متغیرهای مهم در مدل رگرسیون لجستیک، نرم‌افزار SPSS 22 و روش پیشرو گام‌به‌گام والد به کار بردیم. همچنین برای مدل‌سازی روش MARS و SVM و رسم نمودار کنترل $|S|$ ، نرم‌افزار R را مورد استفاده قرار دادیم.

مراحل انجام بررسی عملکرد رهیافت‌های پیشنهادی در این فصل را می‌توان به ترتیب زیر خلاصه

نمود:

۱. تعیین مدل SVM بهینه.
۲. تعیین حد آستانه.
۳. تعیین زمان هشدار نمودار کنترل $|S|$.
۴. رهیافت SVM .
۵. رهیافت ترکیبی $LR - SVM$.
۶. رهیافت ترکیبی $MARS - SVM$.
۷. تعیین متغیرهای مهم.
۸. تعیین نقطه تغییر.
۹. نتیجه‌گیری.

۱.۴ مدل SVM بهینه

ابتدا برای به‌دست آوردن مدل SVM بهینه، ۱۰۰۰ نمونه تصادفی به حجم ۱۲ که ۵۰۰ نمونه اول، تحت کنترل (با بردار میانگین صفر و ماتریس واریانس کوواریانس Σ_0) و ۵۰۰ نمونه دوم، خارج از کنترل (با بردار میانگین صفر و ماتریس واریانس کوواریانس Σ_1) است را از توزیع نرمال چندمتغیره تولید می‌کنیم (کد مربوطه در پیوست ج صفحه ۷۵ موجود است). سپس برای به‌دست آمدن متغیرهای $Z = (Z_1, \dots, Z_N)'$ ، ماتریس واریانس کوواریانس ۱۰۰۰ نمونه ۱۲ تایی را محاسبه می‌کنیم. پس از ایجاد متغیرهای توضیحی Z (متغیرهای ورودی) و متغیر پاسخ (۵۰۰ مشاهده اول برابر با ۱- و ۵۰۰ نمونه دوم برابر با ۱)، مدل SVM را با این مجموعه داده (مجموعه داده آموزشی) اجرا می‌کنیم. همان‌طور که در فصل دوم (بخش ۲.۳.۲) بیان شد، پارامترهای مدل SVM می‌بایست بهینه شوند. لذا با استفاده از تابع هسته RBF و روش جستجوی شبکه‌ای^۱ به ازای مقادیر مختلفی از پارامترهای مدل $(\gamma = 2 \times 10^{-15}, 2 \times 10^{-13}, \dots, 2 \times 10^3, C = 2 \times 10^{-3}, 2 \times 10^{-1}, \dots, 2 \times 10^{15})$ ، پارامترهای بهینه را به‌دست می‌آوریم. بدین ترتیب برای مجموعه داده مذکور، با توجه به آزمایشات فراوانی که صورت گرفت، زوج $(\gamma = 2 \times 10^{-5}, C = 2 \times 10^5)$ در تابع هسته RBF ، به‌عنوان پارامترهای بهینه مدل SVM مشخص شدند (کد مربوطه در پیوست ج صفحه ۷۹ موجود است). یافتن پارامترهای بهینه، سبب افزایش دقت رده‌بندی مدل، هم در مجموعه داده آموزشی و هم در مجموعه داده آزمون می‌شود. لازم به توضیح است که داده آزمون، داده‌ای است که در ساخت مدل دخیل نبوده و صرفاً برای ارزیابی دقت مدل ساخته شده توسط داده آموزشی، مورد استفاده قرار می‌گیرد.

^۱Grid search method

۲.۴ تعیین حد آستانه

برای تعیین حد آستانه، مدل SVM بهینه را با ۱۰۰۰ مشاهده جدید از متغیرهای \underline{Z} و Y که با روندی مشابه بخش قبل شبیه‌سازی شده‌اند، اجرا می‌کنیم؛ پس از اجرای مدل SVM، خروجی این مدل، به‌منظور تعیین حد آستانه در نظر گرفته می‌شود (کد مربوطه در پیوست ج صفحه ۸۳ موجود است). نتایج حد آستانه را به ازای مقادیر مختلف δ و رهیافت‌های مذکور در جدول زیر داریم:

جدول ۱.۴: حد آستانه به‌ازای تغییرات δ و رهیافت‌های SVM، LR-SVM و MARS-SVM

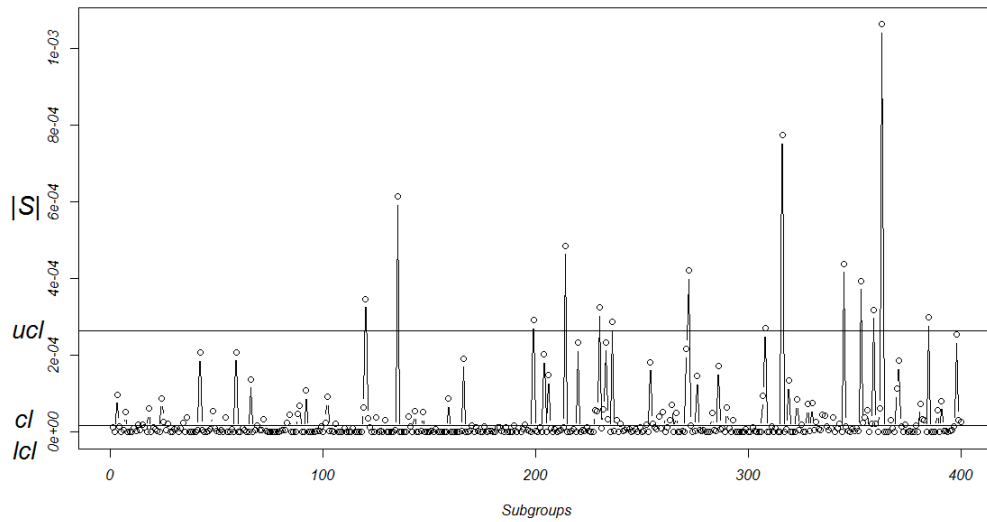
رهیافت‌ها و δ	SVM	LR-SVM	MARS-SVM
۱/۱	۰/۵۱۲۰۰	۰/۵۱۲۲۶	۰/۵۱۲۳۸
۱/۱۲۵	۰/۵۱۲۱۵	۰/۵۱۲۱۹	۰/۵۱۲۲۴
۱/۱۵	۰/۵۱۱۹۱	۰/۵۱۲۰۶	۰/۵۱۲۰۳
۱/۱۷۵	۰/۵۱۱۸۹	۰/۵۱۲۲۳	۰/۵۱۲۱۸

همان‌طور که ملاحظه می‌شود، مقدار حد آستانه برای تمام حالت‌ها برابر ۰/۵۱ به‌دست آمد و تنها در اعشار سوم به بعد اختلاف دارند.

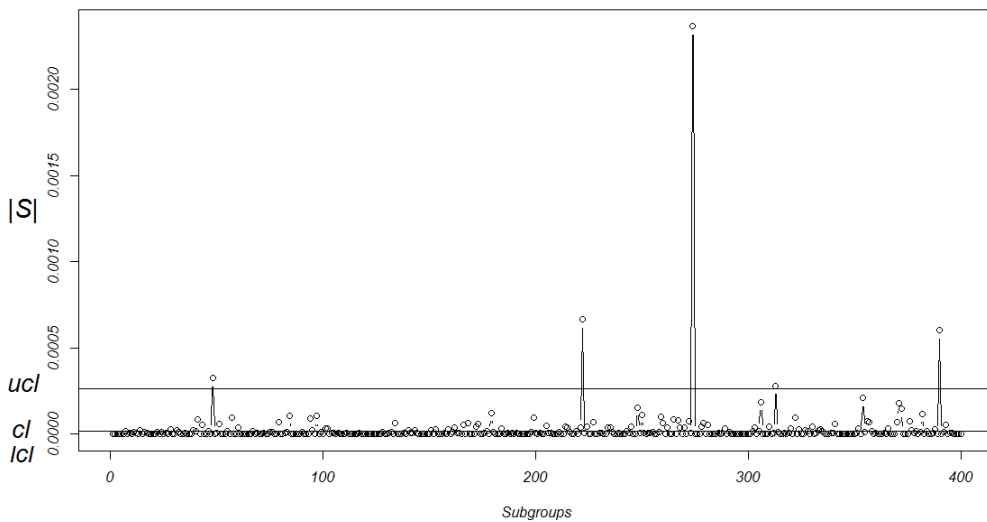
۳.۴ تعیین زمان هشدار نمودار کنترل $|S|$

به‌منظور تشکیل مجموعه داده جدید (داده آزمون، به‌هنگام تعیین نقطه تغییر)، با فرض این‌که زمان تغییر واقعی فرآیند برابر ۲۰۱ است، ۴۰۰ نمونه تصادفی به حجم ۱۲ را از توزیع نرمال چندمتغیره تولید می‌کنیم که ۲۰۰ نمونه اول آن، با بردار میانگین صفر و ماتریس واریانس کوواریانس Σ و ۲۰۰ نمونه بعدی با بردار میانگین صفر و ماتریس واریانس کوواریانس Σ_1 تولید شده‌اند (کد مربوطه در پیوست ج صفحه ۷۵ موجود است). با رسم نمودار کنترل $|S|$ ، زمان T که اولین هشدار خارج از کنترل توسط این نمودار است، مشخص می‌گردد (کد مربوطه در پیوست ج صفحه ۸۸ موجود است). برای اعتبارسنجی رهیافت‌های مطرح‌شده، ۱۰ مجموعه داده جدید، تولید می‌کنیم؛ سپس میانگین و خطای معیار برآورد نقاط تغییر این ۱۰ مجموعه داده را ثبت می‌کنیم.

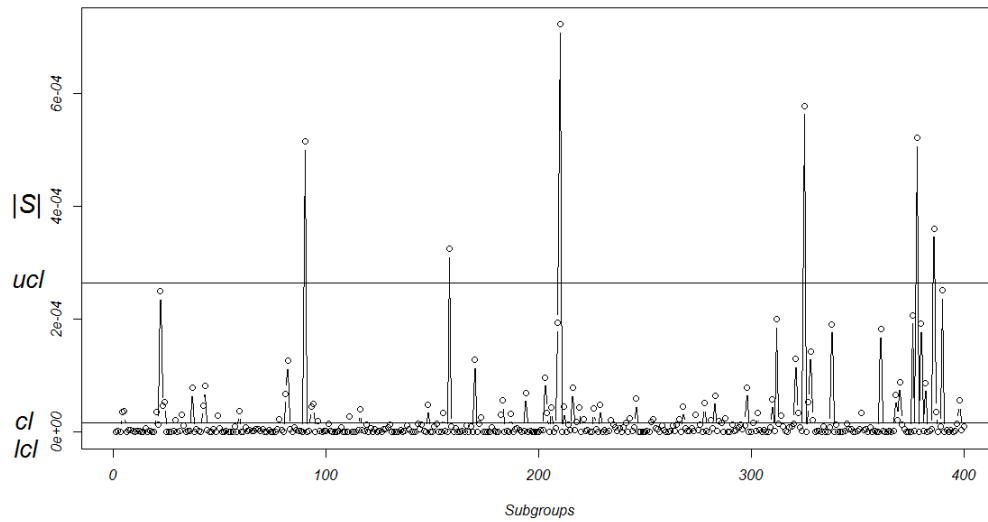
شکل‌های ۱.۴ تا ۴.۴، نمودارهای کنترل $|S|$ را برای برخی از مجموعه داده‌های جدید به‌ازای $\delta = 1/1$ نشان می‌دهند. همان‌طور که در این نمودارها مشخص است، نقاط خارج از بازه‌های کنترل را می‌توان به دو دسته تقسیم کرد: نقاط قبل از نقطه زمانی ۲۰۱ و نقاط بعد از این نقطه زمانی که هشدارهای غلط در دسته اول قرار می‌گیرند و اولین نقطه در دسته دوم به عنوان مقدار T در نظر گرفته می‌شود.



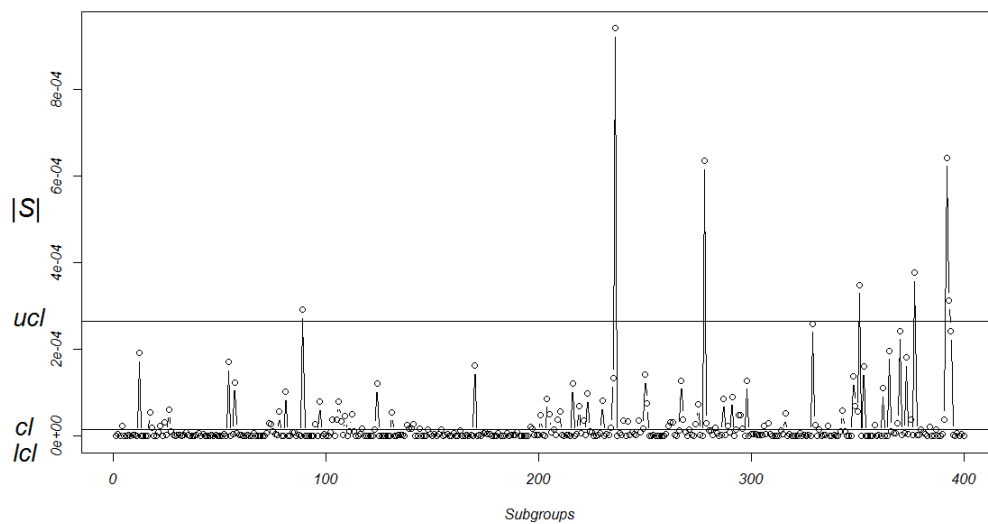
شکل ۱.۴: نمودار کنترل $|S|$ برای اولین مجموعه داده که هشدار در زمان $T = ۲۱۴$ رخ داده است.



شکل ۲.۴: نمودار کنترل $|S|$ برای دومین مجموعه داده که هشدار در زمان $T = ۲۲۲$ رخ داده است.



شکل ۳.۴: نمودار کنترل $|S|$ برای سومین مجموعه داده که اولین هشدار در زمان $T = 210$ رخ داده است.



شکل ۴.۴: نمودار کنترل $|S|$ برای چهارمین مجموعه داده که اولین هشدار در زمان $T = 236$ رخ داده است.

جدول ۲.۴، برآورد نقطه تغییر، میانگین و خطای معیار این نقاط را برای ۱۰ مجموعه داده آزمون به ازای مقادیر مختلف δ نشان می‌دهد. ملاحظه می‌شود که با افزایش مقادیر δ برآورد نقطه تغییر توسط این رهیافت، دقت بالاتری دارد. همان‌طور که قبلاً گفته شد، این رهیافت زمانی که شدت تغییر ایجاد شده در فرآیند کم است، نمی‌تواند عملکرد خوبی از خود نشان دهد و با افزایش شدت تغییر، کارایی خود را به دست می‌آورد.

جدول ۲.۴: برآورد نقطه تغییر توسط رهیافت نمودار کنترل $|S|$ برای ۱۰ مجموعه داده، همراه میانگین و خطای معیار آن به ازای تغییرات δ .

$\delta = 1/175$	$\delta = 1/15$	$\delta = 1/125$	$\delta = 1/1$	
۲۰۴	۲۰۴	۲۱۴	۲۱۴	مجموعه داده اول
۲۲۲	۲۲۲	۲۲۲	۲۲۲	مجموعه داده دوم
۲۰۹	۲۰۹	۲۱۰	۲۱۰	مجموعه داده سوم
۲۳۶	۲۳۶	۲۳۶	۲۳۶	مجموعه داده چهارم
۲۰۳	۲۰۳	۲۰۳	۲۱۷	مجموعه داده پنجم
۲۰۷	۲۰۷	۲۰۷	۲۰۷	مجموعه داده ششم
۲۲۲	۲۲۲	۲۲۲	۲۲۲	مجموعه داده هفتم
۲۱۱	۲۱۱	۲۳۰	۲۳۰	مجموعه داده هشتم
۲۳۳	۲۳۳	۲۳۳	۲۳۳	مجموعه داده نهم
۲۰۹	۲۰۹	۲۰۹	۲۰۹	مجموعه داده دهم
۲۱۵/۶	۲۱۵/۶	۲۱۸/۶	۲۲۰	میانگین نقاط تغییر
۳/۷۶۶	۳/۷۶۶	۳/۶۹۴	۳/۲۸۰	خطای معیار نقاط تغییر

۴.۴ رهیافت SVM

پس از تشکیل مجموعه داده جدید (داده آزمون) و با در نظر گرفتن مقدار T که در بخش قبل به آن اشاره شد، ماتریس واریانس کوواریانس $T - 1$ نمونه را محاسبه می‌کنیم و پس از ایجاد متغیرهای جدید Z (با فرض $p = 10$ ، شامل ۵۶ متغیر توضیحی) و تشکیل متغیر پاسخ Y ، مدل بهینه SVM را برای این مجموعه داده، به منظور رده‌بندی فرآیند اجرا می‌کنیم. سپس برای تعیین نقطه تغییر، راهبرد شناسایی نقطه تغییر در بخش ۳.۳ را اعمال می‌کنیم. این روش تعیین نقطه تغییر را رهیافت SVM می‌نامیم (کد مربوطه در پیوست ج صفحه ۸۸ موجود است).

۵.۴ رهیافت ترکیبی LR – SVM

برای تعیین نقطه تغییر توسط رهیافت ترکیبی LR-SVM، نیاز به تشکیل مجموعه داده جدیدی برای تعیین متغیرهای مهم داریم؛ این مجموعه داده، شامل ۴۰۰ نمونه تصادفی است که ۲۰۰ نمونه‌ی اول آن، تحت کنترل و ۲۰۰ نمونه‌ی بعدی، خارج از کنترل می‌باشد. برای تولید چنین مجموعه داده‌ای، توزیع نرمال چندمتغیره را مورد استفاده قرار دادیم (کد مربوطه در پیوست ج صفحه ۷۵ موجود است). سپس با در نظر گرفتن مقدار T که در بخش قبل به آن اشاره شد، ماتریس واریانس کوواریانس زمان‌های ۱ تا T را محاسبه می‌کنیم؛ از این رو، متغیرهای جدید Z به دست می‌آیند. با اعمال مدل LR، متغیرهای مهم تعیین می‌شوند؛ سپس این متغیرهای مهم را به منظور رده‌بندی فرآیند و تعیین نقطه تغییر، وارد مدل SVM بهینه می‌کنیم. سرانجام راهبرد شناسایی نقطه تغییر را اعمال می‌کنیم (کد مربوطه در پیوست ج صفحه ۸۸ موجود است).

۶.۴ رهیافت ترکیبی MARS – SVM

این رهیافت همانند رهیافت ترکیبی LR-SVM است؛ با این تفاوت که به جای مدل LR، مدل MARS را جایگزین می‌کنیم.

۷.۴ تعیین متغیرهای مهم

رگرسیون لجستیک

ابتدا نتایج ارزیابی هم‌خطی چندگانه بین متغیرهای توضیحی Z را مطالعه می‌کنیم. پس از بررسی هم‌خطی چندگانه بین ۵۶ متغیر توضیحی Z ، به‌ازای مقادیر مختلف δ (۱/۱، ۱/۱۲۵، ۱/۱۵، ۱/۱۷۵)، نتایج حاکی از آن بود که هیچ یک از کمیت‌های VIF بزرگ‌تر از ۱۰ نیستند؛ بنابراین تمام متغیرهای توضیحی، صلاحیت ورود به مدل رگرسیون لجستیک را خواهند داشت.

نتایج حاصل، به‌طور خلاصه در جدول ۳.۴ آمده است. نتایج جدول ۳.۴ نشان می‌دهد که با افزایش مقدار δ ، دقت رده‌بندی افزایش یافته است. پس از به‌دست آمدن متغیرهای مهم به‌ازای مقادیر مختلف δ ، مدل SVM بهینه را با این متغیرهای توضیحی برازش می‌دهیم.

جدول ۳.۴: دقت رده‌بندی و متغیرهای معنی‌دار (با اهمیت) مدل LR به‌ازای مقادیر مختلف δ .

δ	متغیرهای معنی‌دار	دقت رده‌بندی
۱/۱	Z11, Z16, Z26, Z28, Z32, Z43	٪ ۸۱/۲
۱/۱۲۵	Z6, Z14, Z20, Z26, Z35, Z40, Z48, Z50, Z55	٪ ۹۱/۱
۱/۱۵	Z11, Z20, Z26, Z45, Z51, Z53	٪ ۹۳/۲
۱/۱۷۵	Z18, Z30, Z32, Z44, Z49	٪ ۹۹

رگرسیون چندمتغیره سازوار اسپلاین

متغیرهای مهم تشخیص داده شده توسط مدل MARS به ازای مقادیر مختلف δ در جدول ۴.۴ آمده است.

جدول ۴.۴: متغیرهای معنی‌دار (مهم) مدل MARS به ازای مقادیر مختلف δ .

δ	متغیرهای معنی‌دار
۱/۱	Z2, Z8, Z15, Z24, Z25, Z35, Z38, Z55
۱/۱۲۵	Z7, Z12, Z16, Z35, Z40, Z41, Z50
۱/۱۵	Z1, Z5, Z11, Z20, Z26, Z45, Z53, Z56
۱/۱۷۵	Z6, Z24, Z28, Z29, Z30, Z44, Z47

به منظور رده‌بندی وضعیت فرآیند و تعیین نقطه تغییر، مدل SVM بهینه را با این متغیرهای مهم اجرا می‌کنیم.

۸.۴ تعیین نقطه تغییر

با اعمال مدل SVM بهینه بر کل مجموعه متغیرهای توضیحی (رهیافت SVM) و همچنین بر مجموعه متغیرهای توضیحی کاهش یافته حاصل از دو روش LR و MARS، می‌توان نقطه تغییر را بنا بر راهبرد شناسایی نقطه تغییر (بخش ۳.۳ را ببینید) تعیین کرد.

نتایج رده‌بندی رهیافت ترکیبی LR-SVM، برای هفتمین مجموعه داده به ازای $\delta = 1/15$ در جدول ۵.۴ آمده است. در این جدول ستون *Time*، به عنوان زمان بازبینی فرآیند چندمتغیره، M_i تعداد آزمایش‌های دو جمله‌ای، *SVM output* معرف خروجی مدل SVM، n_i^* تعداد موفقیت تجربی در M_i آزمایش برنولی و ستون آخر $P(X_i^* \leq n_i^*)$ احتمال تجمعی توزیع دو جمله‌ای هستند.

جدول ۵.۴: خروجی رهیافت ترکیبی $LR - SVM$ برای هفتمین مجموعه داده به ازای $\delta = 1/15$

<i>Time</i>	M_i	<i>SVM output</i>	n_i^*	$P(X_i^* \leq n_i^*)$
1	221	-1	78	1.65e-19
2	220	-1	78	3.09e-19
3	219	-1	78	5.74e-19
4	218	-1	78	1.06e-18
5	217	-1	78	1.96e-18
.
.
.
200	22	1	12	0.203
201	21	-1	11	0.158
202	20	-1	11	0.232
203	19	-1	11	0.329
204	18	-1	11	0.445
205	17	1	11	0.574
206	16	-1	10	0.504
207	15	1	10	0.643
208	14	-1	9	0.572
209	13	1	9	0.717
210	12	-1	8	0.649
211	11	1	8	0.796
212	10	1	7	0.734
213	9	1	6	0.659
214	8	1	5	0.568
215	7	-1	4	0.464
216	6	1	4	0.677
217	5	1	3	0.568
218	4	1	2	0.434
219	3	-1	1	0.279
220	2	-1	1	0.575
221	1	1	1	1

برآورد نقطه تغییر در این مجموعه داده، اولین زمانی است که احتمال تجمعی $P(X_i^* \leq n_i^*)$ (ستون آخر) بزرگ‌تر یا مساوی 0.51206 (حد آستانه) باشد. لذا برآورد نقطه تغییر با شرایط مذکور، زمان ۲۰۵ است. به همین ترتیب، برای مجموعه داده‌های دیگر عمل می‌کنیم. در نهایت برآورد نقاط تغییر توسط سه رهیافت SVM، ترکیبی LR-SVM و ترکیبی MARS-SVM را طی جداول ۶.۴، ۷.۴ و ۸.۴ به‌ازای مقادیر مختلف δ داریم:

جدول ۶.۴: برآورد نقاط تغییر توسط رهیافت *SVM* برای ۱۰ مجموعه داده آزمون همراه با تغییرات δ .

$\delta = 1/175$	$\delta = 1/15$	$\delta = 1/125$	$\delta = 1/1$	برآورد نقطه تغییر و تغییرات δ
۲۰۲	۱۹۴	۲۱۲	۲۰۸	مجموعه داده اول
۲۰۲	۲۱۱	۲۱۹	۱۹۶	مجموعه داده دوم
۲۰۳	۲۰۱	۲۰۸	۲۰۸	مجموعه داده سوم
۱۹۸	۲۲۸	۲۳۳	۲۰۳	مجموعه داده چهارم
۱۹۸	۲۰۰	۲۰۱	۲۱۴	مجموعه داده پنجم
۲۰۱	۱۹۹	۱۹۶	۲۰۰	مجموعه داده ششم
۲۲۱	۲۱۹	۲۰۵	۲۰۸	مجموعه داده هفتم
۲۰۱	۲۰۶	۲۲۱	۲۲۸	مجموعه داده هشتم
۲۲۰	۲۲۱	۲۲۴	۲۲۴	مجموعه داده نهم
۲۰۸	۲۰۲	۲۰۸	۲۰۷	مجموعه داده دهم
۲۰۸/۸	۲۰۸/۱	۲۱۲/۷	۲۰۹/۶	میانگین نقاط تغییر
۳/۶۱۷	۳/۵۴۲	۳/۶۰۳	۳/۱۶۳	خطای معیار نقاط تغییر

جدول ۷.۴: برآورد نقاط تغییر توسط رهیافت ترکیبی *LR-SVM* برای ۱۰ مجموعه داده آزمون همراه با تغییرات δ .

$\delta = 1/175$	$\delta = 1/15$	$\delta = 1/125$	$\delta = 1/1$	برآورد نقطه تغییر و تغییرات δ
۲۰۱	۱۹۴	۲۰۹	۱۹۴	مجموعه داده اول
۲۰۳	۲۱۵	۲۲۰	۲۱۲	مجموعه داده دوم
۱۹۹	۲۰۰	۱۹۶	۲۰۸	مجموعه داده سوم
۲۲۱	۱۹۹	۲۲۳	۲۲۸	مجموعه داده چهارم
۲۰۷	۲۰۲	۲۰۱	۲۱۵	مجموعه داده پنجم
۱۸۹	۲۰۰	۱۹۸	۱۹۲	مجموعه داده ششم
۲۲۰	۲۰۵	۲۱۷	۲۰۳	مجموعه داده هفتم
۲۰۰	۱۹۸	۲۲۲	۲۰۸	مجموعه داده هشتم
۲۱۸	۲۱۹	۲۱۸	۲۲۱	مجموعه داده نهم
۱۹۹	۱۹۸	۱۹۹	۱۸۸	مجموعه داده دهم
۲۰۴/۷	۲۰۳	۲۱۰/۳	۲۰۶/۹	میانگین نقاط تغییر
۳/۴۷۴	۲/۵۱۷	۳/۴۴۵	۴/۰۸۱	خطای معیار نقاط تغییر

جدول ۸.۴: برآورد نقاط تغییر توسط رهیافت ترکیبی $MARS - SVM$ برای 10° مجموعه داده آزمون همراه با تغییرات δ .

$\delta = 1/175$	$\delta = 1/15$	$\delta = 1/125$	$\delta = 1/1$	برآورد نقطه تغییر و تغییرات δ
۲۰۲	۲۰۲	۲۱۲	۲۱۲	مجموعه داده اول
۲۲۰	۲۱۴	۲۱۵	۲۲۰	مجموعه داده دوم
۲۰۳	۲۰۰	۱۸۸	۱۹۹	مجموعه داده سوم
۲۲۸	۲۲۱	۲۱۹	۲۳۴	مجموعه داده چهارم
۱۹۸	۲۰۲	۲۰۲	۲۱۵	مجموعه داده پنجم
۲۰۱	۲۰۰	۱۹۸	۱۹۹	مجموعه داده ششم
۲۰۵	۲۲۱	۲۱۶	۲۱۱	مجموعه داده هفتم
۲۰۷	۲۰۴	۱۹۵	۲۱۳	مجموعه داده هشتم
۲۰۱	۲۲۱	۲۲۰	۲۲۶	مجموعه داده نهم
۲۰۸	۱۹۹	۲۰۳	۱۹۱	مجموعه داده دهم
۲۰۷/۳	۲۰۸/۴	۲۰۶/۸	۲۱۲	میانگین نقاط تغییر
۲/۹۹۶	۳/۰۵۲	۳/۵۱۱	۴/۱۲۸	خطای معیار نقاط تغییر

۹.۴ نتیجه‌گیری

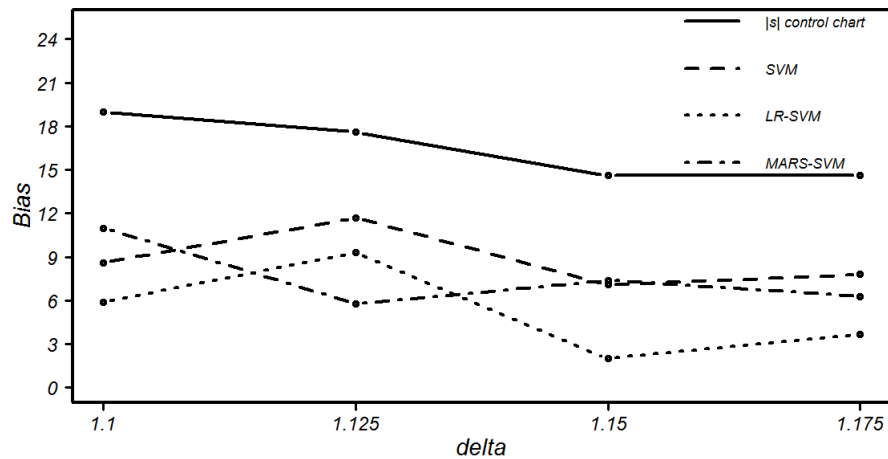
اگر نقطه تغییر را با θ نشان دهیم، میانگین و خطای معیار برآورد نقاط تغییر ($\hat{\theta}$) توسط رهیافت‌های مذکور به ازای مقادیر مختلف δ ، در جدول ۹.۴ آمده است.

جدول ۹.۴: میانگین و خطای معیار (داخل پرانتز) برآورد نقاط تغییر توسط رهیافت‌های نمودار کنترل $|S|$ ، SVM ، ترکیبی $LR-SVM$ و ترکیبی $MARS-SVM$ به ازای مقادیر مختلف δ

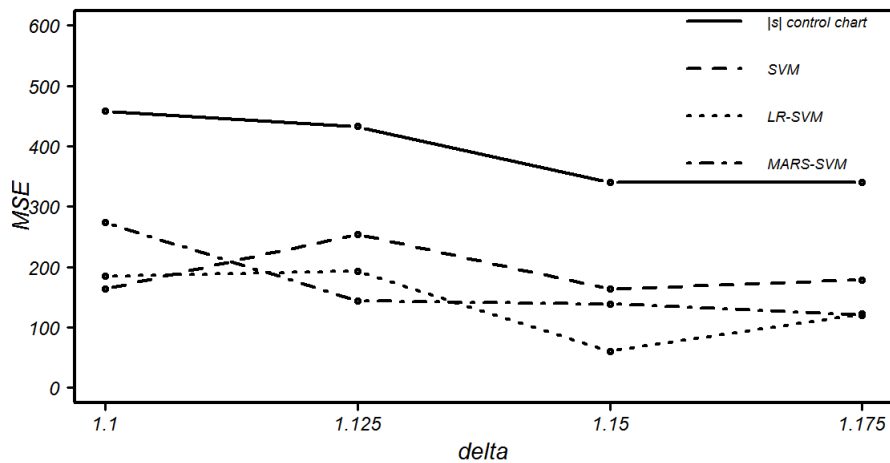
$\hat{\theta}_{ S }$	$\hat{\theta}_{SVM}$	$\hat{\theta}_{LR-SVM}$	$\hat{\theta}_{MARS-SVM}$	δ
۲۲۰(۳/۲۸۰)	۲۰۹/۶(۳/۱۶۳)	۲۰۶/۹(۴/۰۸۱)	۲۱۲(۴/۱۲۸)	۱/۱
۲۱۸/۶(۳/۶۹۴)	۲۱۲/۷(۳/۶۰۳)	۲۱۰/۳(۳/۴۴۵)	۲۰۶/۸(۳/۵۱۱)	۱/۱۲۵
۲۱۵/۶(۳/۷۶۶)	۲۰۸/۱(۳/۵۴۲)	۲۰۳(۲/۵۱۷)	۲۰۸/۴(۳/۰۵۲)	۱/۱۵
۲۱۵/۶(۳/۷۶۶)	۲۰۸/۸(۳/۶۱۷)	۲۰۴/۷(۳/۴۷۴)	۲۰۷/۳(۲/۹۹۶)	۱/۱۷۵

با در نظر گرفتن مقادیر مختلف δ ، می‌توان گفت، رهیافت‌های ترکیبی مطرح شده نسبت به نمودار کنترل $|S|$ و رهیافت SVM ، دارای خطای معیار کمتری هستند و لذا موثرتر می‌باشند. همچنین میانگین نقاط تغییر در رهیافت‌های ترکیبی، نزدیک‌تر به نقطه تغییر واقعی است.

برای اعتبار بخشی به نتایج حاصل، معیارهای دیگری را به منظور ارزیابی عملکرد رهیافت‌های مذکور، مورد استفاده قرار می‌دهیم. بدین منظور، نمودارهای اریبی^۱ و میانگین توان‌های دوم خطا (MSE) برآورد نقاط تغییر را به ازای مقادیر مختلف δ رسم می‌کنیم. این نمودارها در شکل‌های ۵.۴ و ۶.۴ آمده‌اند.



شکل ۵.۴: نمودار اریبی برآورد نقاط تغییر به ازای مقادیر مختلف δ



شکل ۶.۴: نمودار MSE برآورد نقاط تغییر به ازای مقادیر مختلف δ

^۱Bias

با توجه به نمودارهای اریبی و MSE برآورد نقاط تغییر، رهیافت‌های ترکیبی مطرح شده، به‌طور موثر و دقیق‌تر از رهیافت‌های قبلی قادر به شناسایی نقطه تغییر فرآیند می‌باشند. به عبارت دیگر، رهیافت ترکیبی LR-SVM در برآورد نقطه تغییر فرآیند دارای اریبی کمتر از رهیافت‌های دیگر است. همچنین رهیافت ترکیبی MARS-SVM تنها در $\delta = 1/1$ دارای اریبی بیشتر از رهیافت SVM است و در سایر مقادیر δ ، برتری خود را نسبت به رهیافت‌های SVM و نمودار کنترل $|S|$ حفظ می‌کند. از طرفی با در نظر داشتن این‌که در $\delta = 1/1$ مقدار MSE رهیافت SVM از سایر رهیافت‌ها کمتر است اما در سایر مقادیر δ ، نمودار MSE نتایج نمودار اریبی را تایید می‌کند.

۱.۹.۴ آینده تحقیق

پیشنهاد می‌گردد طرح ترکیبی دو مرحله‌ای به‌منظور کشف تغییرات ایجاد شده در سایر پارامترهای فرآیند چندمتغیره، همچون بردار میانگین و همبستگی انجام گیرد. همچنین با بررسی سایر روش‌های رده‌بندی و انتخاب متغیر از قبیل روش‌های تحلیل ممیزی خطی^۱، تحلیل ممیزی درجه دو^۲، درخت رگرسیونی و رده‌بندی^۳ (CART)، SVM و شبکه‌های عصبی مصنوعی^۴ (ANN) می‌توان رهیافت‌های ترکیبی دیگر را برای تعیین نقطه تغییر فرآیند تشکیل داد و سپس نتایج را مقایسه نمود.

^۱Linear discriminant analysis

^۲Quadratic discriminant analysis

^۳Classification and regression tree

^۴Artificial neural networks

پیوست آ

ارزیابی همخطی چندگانه

اگر متغیرهای توضیحی دارای همخطی^۱ باشند، یا به عبارت دیگر ترکیبات خطی شدیدی از یکدیگر باشند، آنگاه نتایج حاصل از مدل‌های رگرسیون معمولی و رگرسیون لجستیک، از اعتبار ساقط خواهند بود. لذا قبل از انجام این نوع مدل‌سازی‌ها، ارزیابی اثر همخطی مهم است.

یکی از کمیت‌های مهم در ارزیابی همخطی چندگانه، کمیت عامل تورم واریانس^۲ VIF می‌باشد. اگر VIF مربوط به یک متغیر برابر ۱ شود، می‌توان گفت این متغیر با سایر متغیرها وابستگی خطی ندارد. کمیت VIF برای متغیر توضیحی Z_i به صورت زیر است:

$$VIF = \frac{1}{1 - R_i^2}, \quad i = 1, 2, \dots, N.$$

که در آن، N تعداد متغیرهای توضیحی و R_i^2 ضریب تعیین رگرسیون متغیر Z_i بر روی سایر متغیرهای توضیحی است.

معیار تصمیم‌گیری برای میزان همخطی به این‌گونه است که اگر $VIF > 10$ باشد، آنگاه متغیر Z_i با سایر متغیرهای دیگر دارای همخطی چندگانه است، در غیر این صورت ($VIF \leq 10$) همخطی چندگانه متغیر Z_i با سایر متغیرها قابل اغماض است. لذا اگر مقدار کمیت VIF برای متغیر Z_i بزرگ‌تر از ۱۰ باشد، آنگاه این متغیر، صلاحیت ورود به مدل رگرسیونی یا رده‌بندی را نخواهد داشت.

^۱Multicollinearity

^۲Variance Inflation Factor

پیوست ب

روش نیوتن-رافسون

روش متداول برای این موضوع، نیوتن-رافسون است. این روش با یک حدس اولیه برای رسیدن به جواب، شروع به کار می‌کند؛ در ادامه دو جمله اول بسط تیلور را در نقطه اولیه محاسبه می‌کند. جواب این بسط، جواب جدیدی می‌باشد که به جواب واقعی نزدیکتر است. در نهایت تا زمانی که به جواب واقعی همگرا شود، نقطه جدید را در بسط تیلور جایگزین می‌کند.

بسط چند جمله‌ای مرتبه n تیلور، برای تابع f در نقطه $x = x_0$ ، با فرض موجود بودن مشتق n ام تابع f در نقطه $x = x_0$ به صورت زیر تعریف می‌شود:

$$\sum_{i=0}^n \frac{f^{(i)}(x_0)}{i!} (x - x_0)^i .$$

که در آن، $f^{(i)}(x_0)$ مشتق مرتبه i تابع f در نقطه x_0 است. برای دستیابی به خط مماس روی تابع f در نقطه $(x_0, f(x_0))$ ، بسط تیلور مرتبه یک را برابر صفر قرار می‌دهیم.

$$f(x_0) + f'(x_0) \cdot (x - x_0) = 0 .$$

با ساده کردن، جواب را داریم:

$$x = x_0 - \frac{f(x_0)}{f'(x_0)} .$$

با قرار دادن $x = x_1$ و با تکرار این روند مقادیر x_2, x_3, \dots به دست می‌آیند، این عمل تا زمانی تکرار خواهد شد که مقادیر x_i ، $i = 1, 2, \dots$ بسیار نزدیک هم شوند و به یک عدد خاص همگرا شوند؛ این عدد خاص، جواب تقریبی خواهد بود.

پیوست پ

آماره آزمون نمره و والد

آماره نمره

مشتق جزئی لگاریتم تابع درست‌نمایی ماکسیم نسبت به پارامتر β را نمره یا ضریب نمره گویند. چون این کمیت، تابعی از β و مشاهدات، Z ، است، لذا نمی‌تواند آماره باشد. با این حال اگر بتوان مقدار ثابتی را برای پارامتر β در نظر گرفت، مثلاً $\beta = \beta_0$ ، آنگاه این کمیت دیگر به پارامتر وابسته نیست، لذا یک آماره خواهد بود. همان‌طور که بیان شد، تابع نمره به صورت زیر است:

$$U = U(\beta, Z) = \frac{\partial}{\partial \beta} \log L(\beta|Z = z) = \frac{1}{L(\beta|Z = z)} \frac{\partial L(\beta|Z = z)}{\partial \beta},$$

که در آن $L(\beta|Z = z)$ تابع درست‌نمایی ماکسیم است.

آماره آزمون نمره

این آزمون به نام‌های دیگری همچون آزمون نمره راتو و آزمون ضریب لاگرانژ^۱ معروف است. این آزمون یک آزمون آماری با فرضیه صفر ساده است. مزیت اصلی این آزمون آن است که نیازی به اطلاعات فرضیه مقابل ندارد.

برای محاسبه آماره آزمون نمره، ابتدا می‌بایست معیار اطلاع فیشر را محاسبه کرد، لذا آن را به صورت زیر داریم:

$$I(\beta) = -E \left[\frac{\partial^2}{\partial \beta^2} \log L(\beta|Z = z) | \beta \right].$$

برای آزمون $H_0: \beta = \beta_0$ ، آماره نمره را به صورت زیر داریم:

$$S(\beta_0) = \frac{U(\beta_0)^2}{I(\beta_0)}.$$

این آماره آزمون دارای توزیع مجانبی کی-دو با یک درجه آزادی است.

^۱Lagrange multiplier test

آماره آزمون والد

برای آزمون $H_0: \beta = \beta_0$ ، آماره والد را به صورت زیر تعریف می‌کنیم:

$$W = \frac{\hat{\beta}^2}{\text{var}(\hat{\beta})},$$

که $\beta_0 = 0$ و $\hat{\beta}$ برآورد درست‌نمایی ماکسیمم پارامتر β است. این آماره آزمون دارای توزیع مجانبی کی-دو با یک درجه آزادی می‌باشد.

پیوست

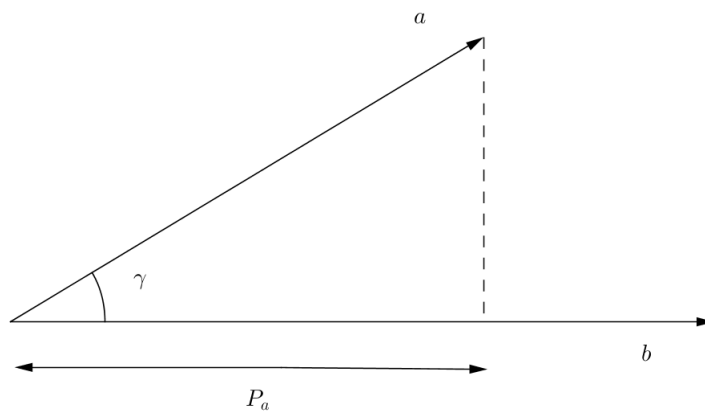
قضیه تصویر

قضیه تصویر

اگر \underline{a} و \underline{b} بردارهایی در فضای \mathbb{R}^N ، با زاویه بین γ باشند، آنگاه P_a را تصویر \underline{a} روی \underline{b} نامند و به صورت زیر تعریف می‌شود:

$$P_a = \|\underline{a}\| \cdot \cos \gamma = \frac{\underline{a} \cdot \underline{b}}{\|\underline{b}\|}, \quad (\text{ت. ۱})$$

که در آن، $\|\underline{a}\| = \sqrt{\underline{a} \cdot \underline{a}} = \sum_{i=1}^N |a_i|^2$ بیانگر اندازه بردار است.



شکل ت. ۱: تصویر بردار a روی بردار b

در واقع P_a مبین آن است که چه مقدار از \underline{a} همسو در راستای \underline{b} است (افشاری، ۱۳۹۱).

پیوست

شرایط KKT

مسئله بهینه سازی زیر را در نظر بگیرید.

$$\begin{aligned} & \text{minimize} && J(\theta), \\ & \text{subject to} && f_i(\theta) \geq 0, \quad i = 1, 2, \dots, M. \end{aligned}$$

که در آن $J(\theta)$ تابع هدف و $f_i(\theta)$ محدودیت مسئله هستند. اگر جواب بهینه این مسئله، به ازای θ^* رخ دهد، آنگاه مجموعه‌ای از شرایط لازم تحت عنوان شرایط KKT وجود دارند که θ^* می‌بایست در آن صدق کند. این شرایط عبارت اند از:

$$\begin{aligned} \frac{\partial}{\partial \theta} L(\theta^*, \alpha) &= 0, \\ \alpha_i &\geq 0, \quad i = 1, 2, \dots, M, \\ \alpha_i f_i(\theta^*) &= 0 \quad i = 1, 2, \dots, M. \end{aligned}$$

که در آن α_i ضرایب لاگرانژ و $L(\theta^*, \alpha)$ معادله‌ی لاگرانژ این مسئله می‌باشند (بازارا و همکاران، ۲۰۱۰).

پیوست ج

کدهای نرم افزار R

ج.۱ شبیه سازی مجموعه داده های آموزشی، آزمون، حد آستانه و تعیین متغیرهای مهم

```
library(MASS)
library(Matrix)
library(abind)
library(e1071)
library(earth)
#---making training & test & dim & threshold data sets
#generate sigma0 and other sigma-----
p=10
n=12
mu=rep(0,p)
s0=matrix(rep(0.5,p*p),c(p,p))
colnames(s0)=c("x1","x2","x3","x4","x5","x6","x7","x8","x9","x10")
rownames(s0)=c("x1","x2","x3","x4","x5","x6","x7","x8","x9","x10")
diag(s0)=c(rep(1,p))
s0
sigma1=1.1*s0
sigma2=1.125*s0
sigma3=1.15*s0
sigma4=1.175*s0
```

```

#simulate data from multivariate normal distribution
set.seed(1)
options(max.print=1000000000)
in.control=replicate(3000,mvrnorm(n, mu, s0))
out.control=replicate(3000,mvrnorm(n, mu, sigma4))

#-----making model.set=training set-----
set.seed(2)
modelindex=sample(3000,500,replace=F)
model1=in.control[, ,modelindex]
in.control=in.control[, , -modelindex]
model2=out.control[, ,modelindex]
out.control=out.control[, , -modelindex]
model.set=abind(model1,model2)
#-----making faz1.set=threshold set-----
set.seed(12)
trainindex=sample(2500,500,replace=F)
train1=in.control[, ,trainindex]
in.control=in.control[, , -trainindex]
train2=out.control[, ,trainindex]
out.control=out.control[, , -trainindex]
faz1.set=abind(train1,train2)
#---- making--test set 1-----
set.seed(13)
testindex=sample(2000,200,replace=F)
test1=in.control[, ,testindex]
in.control=in.control[, , -testindex]
test2=out.control[, ,testindex]
out.control=out.control[, , -testindex]
test.set1=abind(test1,test2)
#---- making--test set 2-----
set.seed(14)
testindex=sample(1800,200,replace=F)
test1=in.control[, ,testindex]

```

```
in.control=in.control[,,-testindex]
test2=out.control[, ,testindex]
out.control=out.control[,,-testindex]
test.set2=abind(test1,test2)
#---- making--test set 3-----
set.seed(15)
testindex=sample(1600,200,replace=F)
test1=in.control[, ,testindex]
in.control=in.control[,,-testindex]
test2=out.control[, ,testindex]
out.control=out.control[,,-testindex]
test.set3=abind(test1,test2)
#---- making--test set 4-----
set.seed(16)
testindex=sample(1400,200,replace=F)
test1=in.control[, ,testindex]
in.control=in.control[,,-testindex]
test2=out.control[, ,testindex]
out.control=out.control[,,-testindex]
test.set4=abind(test1,test2)
#---- making--test set5-----
set.seed(17)
testindex=sample(1200,200,replace=F)
test1=in.control[, ,testindex]
in.control=in.control[,,-testindex]
test2=out.control[, ,testindex]
out.control=out.control[,,-testindex]
test.set5=abind(test1,test2)
#---- making--test set 6-----
set.seed(35)
testindex=sample(1000,200,replace=F)
test1=in.control[, ,testindex]
in.control=in.control[,,-testindex]
test2=out.control[, ,testindex]
```

```

out.control=out.control[,,-testindex]
test.set6=abind(test1,test2)
#---- making--test set 7-----
set.seed(19)
testindex=sample(800,200,replace=F)
test1=in.control[, ,testindex]
in.control=in.control[,,-testindex]
test2=out.control[, ,testindex]
out.control=out.control[,,-testindex]
test.set7=abind(test1,test2)
#---- making--test set 8-----
set.seed(37)
testindex=sample(600,200,replace=F)
test1=in.control[, ,testindex]
in.control=in.control[,,-testindex]
test2=out.control[, ,testindex]
out.control=out.control[,,-testindex]
test.set8=abind(test1,test2)
#---- making--test set 9-----
set.seed(38)
testindex=sample(400,200,replace=F)
test1=in.control[, ,testindex]
in.control=in.control[,,-testindex]
test2=out.control[, ,testindex]
out.control=out.control[,,-testindex]
test.set9=abind(test1,test2)
#---- making--test set 10-----
test.set10=abind(in.control,out.control)
#--auxiliary(dim) sets for MARS & Logistic models--
set.seed(10)
in.dim=replicate(200,mvrnorm(n, mu, s0))
out.dim=replicate(200,mvrnorm(n, mu, sigma4))
dim.set=abind(in.dim,out.dim)

```


ج.۲. بهینه‌سازی پارامترهای مدل SVM

برای مثال کد مربوط به بهینه‌سازی پارامترهای مدل SVM با متغیرهای توضیحی کاهش یافته توسط مدل MARS زمانی که $\delta = 1/1$ است را قرار می‌دهیم.

```
#compute covariance matrix when the undrlying process in control
#for training data set
covmat0=array(rep(0,50000),c(p,p,500))
for(i in 1:500){

  covmat0[, ,i]=var(model.set[, ,i])
}
##compute covariance matrix for all time
covmat=array(rep(0,100000),c(p,p,1000))
for(i in 1:1000){

  covmat[, ,i]=var(model.set[, ,i])
}
m=(p*(p+1)/2)+1
z.model=array(rep(0,1000*m),c(1000,m))
for(i in 1:1000){
  l=1
  for(u in 1:p){
    for(o in u:p){
      z.model[i,l]=covmat[u,o,i]
      l=l+1
    }
  }
  z.model[i,((p*(p+1)/2))+1]=det(covmat[, ,i])
}

#adding response variable in matrix z.model
y0.model=rep(-1,500)
y1.model=rep(1,500)
y.model=append(y0.model,y1.model)
```

```

#z.model.training variables with T observation
z.model=cbind(z.model,y.model)
colnames(z.model)=
  c("z1","Z2","Z3","Z4","Z5","Z6","Z7","Z8","Z9","Z10",
    "Z11","Z12","Z13","Z14","Z15","Z16","Z17","Z18","Z19","Z20",
    "Z21","Z22","Z23","Z24","Z25","Z26","Z27","Z28","Z29","Z30",
    "Z31","Z32","Z33","Z34","Z35","Z36","Z37","Z38","Z39","Z40",
    "Z41","Z42","Z43","Z44","Z45","Z46","Z47","Z48","Z49","Z50",
    "Z51","Z52","Z53","Z54","Z55","Z56","y.model")
#-----choosing importance variables with mars model----
z.mars=z.model[,2]
z.mars=cbind(z.mars,z.model[,8])
z.mars=cbind(z.mars,z.model[,15])
z.mars=cbind(z.mars,z.model[,24])
z.mars=cbind(z.mars,z.model[,25])
z.mars=cbind(z.mars,z.model[,35])
z.mars=cbind(z.mars,z.model[,38])
z.mars=cbind(z.mars,z.model[,55])
z.mars=cbind(z.mars,z.model[,57])
colnames(z.mars)=c("Z2","Z8","Z15","Z24","Z25","Z35","Z38","Z55","y")

# data for faz1 part-----

#compute covariance matrix when the undrlying process in control
#for threshold data set
covmat0=array(rep(0,50000),c(p,p,500))
for(i in 1:500){

  covmat0[, ,i]=var(faz1.set[, ,i])
}
##compute covariance matrix for all time
covmat=array(rep(0,100000),c(p,p,1000))
for(i in 1:1000){

```

```

covmat[, , i]=var(faz1.set[, , i])
}
m=(p*(p+1)/2)+1
z.faz1=array(rep(0,1000*m),c(1000,m))
for(i in 1:1000){
  l=1
  for(u in 1:p){
    for(o in u:p){
      z.faz1[i,l]=covmat[u,o,i]
      l=l+1
    }
  }
  z.faz1[i,((p*(p+1)/2))+1]=det(covmat[, , i])
}

#adding response variable in matrix z.faz1
y0.faz1=rep(-1,500)
y1.faz1=rep(1,500)
y.faz1=append(y0.faz1,y1.faz1)
#z.faz1.training variables with T obesrvation
z.faz1=cbind(z.faz1,y.faz1)
colnames(z.faz1)=
  c("z1", "Z2", "Z3", "Z4", "Z5", "Z6", "Z7", "Z8", "Z9", "Z10",
    "Z11", "Z12", "Z13", "Z14", "Z15", "Z16", "Z17", "Z18", "Z19", "Z20",
    "Z21", "Z22", "Z23", "Z24", "Z25", "Z26", "Z27", "Z28", "Z29", "Z30",
    "Z31", "Z32", "Z33", "Z34", "Z35", "Z36", "Z37", "Z38", "Z39", "Z40",
    "Z41", "Z42", "Z43", "Z44", "Z45", "Z46", "Z47", "Z48", "Z49", "Z50",
    "Z51", "Z52", "Z53", "Z54", "Z55", "Z56", "y.faz1")
#-----choosing importance variables with mars model-----
z.threshold=z.faz1[,2]
z.threshold=cbind(z.threshold,z.faz1[,8])
z.threshold=cbind(z.threshold,z.faz1[,15])
z.threshold=cbind(z.threshold,z.faz1[,24])
z.threshold=cbind(z.threshold,z.faz1[,25])

```

```

z.threshold=cbind(z.threshold,z.faz1[,35])
z.threshold=cbind(z.threshold,z.faz1[,38])
z.threshold=cbind(z.threshold,z.faz1[,55])
z.threshold=cbind(z.threshold,z.faz1[,57])
colnames(z.threshold)=c("Z2","Z8","Z15","Z24","Z25",
                        "Z35","Z38","Z55","y")
new.var=rbind(z.mars,z.threshold)

u=11
m=12
a=2e-5
b=2e15
d=2e-15
e=2e-13

ps.train=array(0,c(m,u))
ps.threshold=array(0,c(m,u))
c=rep(0,m)
c[1]=2e-7
gama=rep(0,u)
gama[1]=2e-17
for(i in 2:m){
  #c[i]=a+((b-a)/m)*i
  c[i]=100*c[i-1]
  for(j in 2:u){
    #gama[j]=d+((e-d)/u)*j
    gama[j]=100*gama[j-1]
    svm.fit=svm(new.var[1:1000,-10],new.var[1:1000,10],kernel="r",
               type="C-classification",gamma=gama[j],cost=c[i])
    svm.pred=predict(svm.fit)
    ps.train[i,j]=mean(svm.pred==new.var[1:1000,10])
    svm.pred=predict(svm.fit,new.var[1001:nrow(new.var),-10])
    ps.threshold[i,j]=mean(svm.pred==new.var[1001:nrow(new.var),10])
    j=j+1
  }
}

```

```

}
i=i+1
}

AIR.obtim=list(ps.train,ps.threshold)
#-----excluding z variables-----
write.table(AIR.obtim, file=paste("E:/obtim MARS+SVM(sigma1).csv")
, append = FALSE, quote = TRUE, sep = " ",
eol = "\n", na = "NA", dec = ".", row.names =FALSE,
col.names = TRUE, qmethod = c("escape", "double"),
fileEncoding = "")

```

ج.۳ تعیین حد آستانه

برای مثال کد روش ترکیبی *MARS - SVM* به ازای $\delta = 1/1$ را قرار می‌دهیم:

```

#compute covariance matrix when the undrlying process in control
#for training data set
covmat0=array(rep(0,50000),c(p,p,500))
for(i in 1:500){

    covmat0[, ,i]=var(model.set[, ,i])
}
##compute covariance matrix for all time
covmat=array(rep(0,100000),c(p,p,1000))
for(i in 1:1000){

    covmat[, ,i]=var(model.set[, ,i])
}
m=(p*(p+1)/2)+1
z.model=array(rep(0,1000*m),c(1000,m))
for(i in 1:1000){
    l=1

```

```

for(u in 1:p){
  for(o in u:p){
    z.model[i,l]=covmat[u,o,i]
    l=l+1
  }
}
z.model[i,((p*(p+1)/2))+1]=det(covmat[, ,i])
}

#adding response variable in matrix z.model
y0.model=rep(-1,500)
y1.model=rep(1,500)
y.model=append(y0.model,y1.model)
#z.model.training variables with T obesrvation
z.model=cbind(z.model,y.model)
colnames(z.model)=
  c("z1","Z2","Z3","Z4","Z5","Z6","Z7","Z8","Z9","Z10",
    "Z11","Z12","Z13","Z14","Z15","Z16","Z17","Z18","Z19","Z20",
    "Z21","Z22","Z23","Z24","Z25","Z26","Z27","Z28","Z29","Z30",
    "Z31","Z32","Z33","Z34","Z35","Z36","Z37","Z38","Z39","Z40",
    "Z41","Z42","Z43","Z44","Z45","Z46","Z47","Z48","Z49","Z50",
    "Z51","Z52","Z53","Z54","Z55","Z56","y.model")
#-----choosing importance variables with mars model----
z.mars=z.model[,2]
z.mars=cbind(z.mars,z.model[,8])
z.mars=cbind(z.mars,z.model[,15])
z.mars=cbind(z.mars,z.model[,24])
z.mars=cbind(z.mars,z.model[,25])
z.mars=cbind(z.mars,z.model[,35])
z.mars=cbind(z.mars,z.model[,38])
z.mars=cbind(z.mars,z.model[,55])
z.mars=cbind(z.mars,z.model[,57])
colnames(z.mars)=c("Z2","Z8","Z15","Z24","Z25",
  "Z35","Z38","Z55","y")

```

```

# data for threshold determination-----

#compute covariance matrix when the underlying process in control
#for threshold data set
covmat0=array(rep(0,50000),c(p,p,500))
for(i in 1:500){

  covmat0[, ,i]=var(faz1.set[, ,i])
}
##compute covariance matrix for all time
covmat=array(rep(0,100000),c(p,p,1000))
for(i in 1:1000){

  covmat[, ,i]=var(faz1.set[, ,i])
}
m=(p*(p+1)/2)+1
z.faz1=array(rep(0,1000*m),c(1000,m))
for(i in 1:1000){
  l=1
  for(u in 1:p){
    for(o in u:p){
      z.faz1[i,l]=covmat[u,o,i]
      l=l+1
    }
  }
  z.faz1[i,((p*(p+1)/2))+1]=det(covmat[, ,i])
}

#adding response variable in matrix z.faz1
y0.faz1=rep(-1,500)
y1.faz1=rep(1,500)
y.faz1=append(y0.faz1,y1.faz1)
#z.faz1.training variables with T observation

```

```

z.faz1=cbind(z.faz1,y.faz1)
colnames(z.faz1)=
  c("z1","Z2","Z3","Z4","Z5","Z6","Z7","Z8","Z9","Z10",
    "Z11","Z12","Z13","Z14","Z15","Z16","Z17","Z18","Z19","Z20",
    "Z21","Z22","Z23","Z24","Z25","Z26","Z27","Z28","Z29","Z30",
    "Z31","Z32","Z33","Z34","Z35","Z36","Z37","Z38","Z39","Z40",
    "Z41","Z42","Z43","Z44","Z45","Z46","Z47","Z48","Z49","Z50",
    "Z51","Z52","Z53","Z54","Z55","Z56","y.faz1")
#-----choosing importance variables with mars model-----
z.threshold=z.faz1[,2]
z.threshold=cbind(z.threshold,z.faz1[,8])
z.threshold=cbind(z.threshold,z.faz1[,15])
z.threshold=cbind(z.threshold,z.faz1[,24])
z.threshold=cbind(z.threshold,z.faz1[,25])
z.threshold=cbind(z.threshold,z.faz1[,35])
z.threshold=cbind(z.threshold,z.faz1[,38])
z.threshold=cbind(z.threshold,z.faz1[,55])
z.threshold=cbind(z.threshold,z.faz1[,57])
colnames(z.threshold)=c("Z2","Z8","Z15","Z24","Z25",
  "Z35","Z38","Z55","y")

#-----SVM modeling with training set and threshold data set--
# threshold determination-----
z.new=rbind(z.mars,z.threshold)
#svm parameters all set-----
svm.fit=svm(z.new[1:1000,-9],z.new[1:1000,9],kernel="r",
  type="C-classification",gamma=2e-05,cost=2e+5)
svm.pred1=predict(svm.fit)
ps.train=mean(svm.pred1==z.new[1:1000,9])
# use svm model that estimated parameters in faz1
svm.pred2=predict(svm.fit,z.new[1001:nrow(z.new),-9])
ps.test=mean(svm.pred2==z.new[1001:nrow(z.new),9])
ps=ps.test
astane4=1000

```



```

svm.pred=svm.pred2
v=as.integer(svm.pred==1)
probb=rep(0,(astane4))
n_s=floor(1000*ps)
thre=pbinom(n_s,1000,ps)
ns=rep(0,(astane4))
ns[1]=v[astane4]
probb[1]=pbinom(ns[1],1,ps)
sample.number=rep(1,(astane4))
for(i in 2:(astane4)){
  n=i
  sample.number[i]=i
  ns[i]=v[astane4+1-i]+ns[i-1]
  probb[i]=pbinom(ns[i],n,ps)
}

#which(probb>=max(probb))
backward.number=rep(0,(astane4))
probb2=rep(0,(astane4))
ns2=rep(0,(astane4))
for(i in 1:(astane4)){
  backward.number[astane4+1-i]=i
  probb2[astane4+1-i]=probb[i]
  ns2[astane4+1-i]=ns[i]
}
svm.mars=list(sample.number,backward.number,v,ns2,probb2,thre)
names(svm.mars)=c("sample number","backward number","SVM output"
  ,"Cumulative sum of 1(ns)",
  "cumulative probability(P)","Threshold")

#-----excluding table-----
write.table(svm.mars, file=paste("E:/1000 modeled&1000 test.csv")
  , append = FALSE, quote = TRUE, sep = " ",
  eol = "\n", na = "NA", dec = ".", row.names =FALSE,

```

```
col.names = TRUE, qmethod = c("escape", "double"),
fileEncoding = "")
```

ج.۴ تعیین نقطه تغییر

برای مثال کد مربوط به رهیافت ترکیبی *MARS - SVM* برای پنجمین مجموعه داده آزمون به ازای $\delta = 1/1$ را قرار می دهیم.

```
#compute covariance matrix when the underlying process in control
#for modeling data set
covmat0=array(rep(0,50000),c(p,p,500))
for(i in 1:500){

  covmat0[, ,i]=var(model.set[, ,i])
}
##compute covariance matrix for all time
covmat=array(rep(0,100000),c(p,p,1000))
for(i in 1:1000){

  covmat[, ,i]=var(model.set[, ,i])
}
m=(p*(p+1)/2)+1
z.model=array(rep(0,1000*m),c(1000,m))
for(i in 1:1000){
  l=1
  for(u in 1:p){
    for(o in u:p){
      z.model[i,l]=covmat[u,o,i]
      l=l+1
    }
  }
  z.model[i,((p*(p+1)/2))+1]=det(covmat[, ,i])
}
```

```

#adding response variable in matrix z.model
y0.model=rep(-1,500)
y1.model=rep(1,500)
y.model=append(y0.model,y1.model)
#z.model.training variables with T obesrvation
z.model=cbind(z.model,y.model)
colnames(z.model)=
  c("z1","Z2","Z3","Z4","Z5","Z6","Z7","Z8","Z9","Z10",
    "Z11","Z12","Z13","Z14","Z15","Z16","Z17","Z18","Z19","Z20",
    "Z21","Z22","Z23","Z24","Z25","Z26","Z27","Z28","Z29","Z30",
    "Z31","Z32","Z33","Z34","Z35","Z36","Z37","Z38","Z39","Z40",
    "Z41","Z42","Z43","Z44","Z45","Z46","Z47","Z48","Z49","Z50",
    "Z51","Z52","Z53","Z54","Z55","Z56","y.model")
#-----choosing importance variables with mars model----
z.mars=z.model[,2]
z.mars=cbind(z.mars,z.model[,8])
z.mars=cbind(z.mars,z.model[,15])
z.mars=cbind(z.mars,z.model[,24])
z.mars=cbind(z.mars,z.model[,25])
z.mars=cbind(z.mars,z.model[,35])
z.mars=cbind(z.mars,z.model[,38])
z.mars=cbind(z.mars,z.model[,55])
z.mars=cbind(z.mars,z.model[,57])
colnames(z.mars)=c("Z2","Z8","Z15","Z24","Z25","Z35","Z38","Z55","y")
# data for test part-----

#compute covariance matrix when the undrlying process in control
#(for test data set)
covmat0=array(rep(0,20000),c(p,p,200))
for(i in 1:200){
  covmat0[,i]=var(test.set5[,i])
}
##compute covariance matrix for all time

```

```

covmat=array(rep(0,40000),c(p,p,400))
for(i in 1:400){

  covmat[, ,i]=var(test.set5[, ,i])
}
m=(p*(p+1)/2)+1
z.test=array(rep(0,400*m),c(400,m))
for(i in 1:400){
  l=1
  for(u in 1:p){
    for(o in u:p){
      z.test[i,l]=covmat[u,o,i]
      l=l+1
    }
  }
  z.test[i,((p*(p+1)/2))+1]=det(covmat[, ,i])
}

#adding response variable in matrix z.test
y0.test=rep(-1,200)
y1.test=rep(1,200)
y.test=append(y0.test,y1.test)
#z.test variables with T observation
z.test=cbind(z.test,y.test)
colnames(z.test)=
  c("z1","z2","z3","z4","z5","z6","z7","z8","z9","z10",
    "z11","z12","z13","z14","z15","z16","z17","z18","z19","z20",
    "z21","z22","z23","z24","z25","z26","z27","z28","z29","z30",
    "z31","z32","z33","z34","z35","z36","z37","z38","z39","z40",
    "z41","z42","z43","z44","z45","z46","z47","z48","z49","z50",
    "z51","z52","z53","z54","z55","z56","y.test")

# compute b1 and b2 for generalize sample variance chart (SPC)
zarb1=1

```

```

for(q in 1:p){
  zarb1=(n-q)*zarb1
}
b1=(1/(n-1)^p)*zarb1
zarb2=1
for(w in 1:p){
  zarb2=(n-w+2)*zarb2
}
b2=(1/(n-1)^(2*p))*zarb1*(zarb2-zarb1)
#-----
det.s0=rep(0,200)
for(i in 1:200){
  det.s0[i]=det(covmat0[, ,i])
}
det.s=rep(0,400)
for(i in 1:400){
  det.s[i]=det(covmat[, ,i])
}
cl=det(s0)*b1
lcl=det(s0)*(b1-(3*sqrt(b2)))
ucl=det(s0)*(b1+3*sqrt(b2))
det.s
plot(det.s,type="b",xlab="Subgroups")
abline(a = ucl, b = 0, h =ucl ,
       coef = NULL, untf = FALSE)
abline(a = cl, b = 0, h =cl ,
       coef = NULL, untf = FALSE)
which(det.s>ucl)
astane.test=which(det.s>ucl)
astane.test2=which(astane.test>200)
astane.test3=min(astane.test2)
astane.test4=min(astane.test[astane.test3]:astane.test[length(astane.test)])
z.test=z.test[1:(astane.test4),]
#-----choosing importance variables with mars model----

```

```

z.mars.test=z.test[,2]
z.mars.test=cbind(z.mars.test,z.test[,8])
z.mars.test=cbind(z.mars.test,z.test[,15])
z.mars.test=cbind(z.mars.test,z.test[,24])
z.mars.test=cbind(z.mars.test,z.test[,25])
z.mars.test=cbind(z.mars.test,z.test[,35])
z.mars.test=cbind(z.mars.test,z.test[,38])
z.mars.test=cbind(z.mars.test,z.test[,55])
z.mars.test=cbind(z.mars.test,z.test[,57])
colnames(z.mars.test)=c("Z2","Z8","Z15","Z24","Z25",
                        "Z35","Z38","Z55","y")
#---SVM modeling with training data set and test set----
# making table and detecting change point-----
z.new=rbind(z.mars,z.mars.test)
#svm parameters all set-----
svm.fit=svm(z.new[1:1000,-9],z.new[1:1000,9],kernel="r",
            type="C-classification",gamma=2e-05,cost=2e+5)
svm.pred1=predict(svm.fit)
ps.train=mean(svm.pred1==z.new[1:1000,9])
# use svm model that estimated parameters in faz1
svm.pred2=predict(svm.fit,z.new[1002:nrow(z.new)-1,-9])
ps.test=mean(svm.pred2==z.new[1002:nrow(z.new)-1,9])
ps=ps.test
astane4=astane.test4
svm.pred=svm.pred2
v=as.integer(svm.pred==1)
probb=rep(0,(astane4-1))
nA=rep(0,(astane4-1))
nA[1]=v[astane4-1]
probb[1]=pbinom(nA[1],1,ps)
sample.number=rep(1,(astane4-1))
for(i in 2:(astane4-1)){
  n=i
  sample.number[i]=i
}

```

```

nA[i]=v[astane4-i]+nA[i-1]
probb[i]=pbinom(nA[i],n,ps)
}

#which(probb>=max(probb))
backward.number=rep(0,(astane4-1))
probb2=rep(0,(astane4-1))
nA2=rep(0,(astane4-1))
for(i in 1:(astane4-1)){
  backward.number[astane4-i]=i
  probb2[astane4-i]=probb[i]
  nA2[astane4-i]=nA[i]
}

svm.mars=list(sample.number,backward.number,v,nA2,probb2)
names(svm.mars)=c("sample number","backward number","SVM output"
  ,"Cumulative sum of 1(nA)","cumulative probability(P)")

#-----excluding table-----
write.table(svm.mars, file=paste("E:/(test set5+mars).csv")
  , append = FALSE, quote = TRUE, sep = " ",
  eol = "\n", na = "NA", dec = ".", row.names =FALSE,
  col.names = TRUE, qmethod = c("escape", "double"),
  fileEncoding = "")

```


مراجع

- [۱] آتشگرک. (۱۳۹۳)، "کنترل کیفیت آماری پیشرفته"، موسسه آموزشی و تحقیقاتی صنایع دفاعی.
- [۲] افشاری م. (۱۳۹۱)، "کاربرد روش‌های مختلف یادگیری ماشین در استخراج پارامترهای کیفی آب از داده‌های تشعشع طیفی"، پایان‌نامه کارشناسی ارشد، دانشگاه شاهرود، دانشکده علوم ریاضی.
- [۳] رسولی م. (۱۳۹۳)، "شناسایی و تخمین نقطه تغییر در بردار پارامتر فرآیندهای پربازده با چند مشخصه کیفی وصفی"، پایان‌نامه کارشناسی ارشد، دانشگاه صنعتی خواجه نصیرالدین طوسی، دانشکده مهندسی صنایع.
- [۴] غفوری ح. (۱۳۹۱)، "رویکردی به فضاهاى هیلبرت فازی"، پایان‌نامه کارشناسی ارشد، دانشگاه فرودسی مشهد، دانشکده علوم ریاضی.
- [5] Abe S. (2005), "Support vector machines for pattern classification", Vol. 2. London: Springer.
- [6] Anderson T.W. (1984), "An Introduction to Multivariate Statistical Analysis", Wiley.
- [7] Amiri A., and Allahyari S. (2012), "Change point estimation methods for control chart postsignal diagnostics: a literature review", *Quality and Reliability Engineering International* 28, 673-685.
- [8] Bazaraa M.S., Jarvis J.J., and Sherali H.D. (2010), "Linear Programming and Network Flows", John Wiley Sons, Mathematics - 748 pages.
- [9] Bischoff W., and Miller F. (2002), "A minimax two-stage procedure for comparing treatments: looking at a hybrid test and estimation problem as a whole", *Statistica Sinica* 12, 1133-1144.
- [10] Chan K.Y., Dillon T.S., and Kwong C.K. (2011), "Modeling of a liquid epoxy molding process using a particle swarm optimization-based fuzzy regression approach", *IEEE Transactions on Industrial Informatics* 7 (1), 148-158.
- [11] Chan K.Y., Kwong C.K., and Tsim Y.C. (2010), "Modelling and optimization of fluid dispensing for electronic packaging using neural fuzzy networks and genetic algorithms", *Engineering Applications of Artificial Intelligence* 23, 18-26.

- [12] Chou S.M., Lee T.S., Shao Y.E., and Chen I.F. (2004), "Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines", *Expert Systems with Applications* 27, 133–142.
- [13] Craven M.W., and Shavlik J.W. (1997), "Using neural networks for data mining", *Future Generation Computer Systems* 13, 221–229.
- [14] Fahmy H. M., and Elsayed E. A. (2006), "Detection of linear trends in process mean", *International journal of production research* 44, 487-504.
- [15] Friedman, Jerome H. (1991), "Multivariate adaptive regression splines", *The annals of statistics* 1-67.
- [16] Hastie T., Tibshirani R., and Friedman J. (2001), "The Elements of Statistical Learning", *Springer Series in Statistics* Springer New York Inc., New York, NY, USA.
- [17] Hosmer Jr, David W., and Lemeshow S. (2004), "Applied logistic regression", *John Wiley Sons*.
- [18] Hou C.D., Shao Y.E., and Huang S. (2011), "Estimation of the change point of a multivariate normal process using the $|S|$ control chart and MLE method", *ICIC Express Letters – An International Journal of Research and Surveys* 5, 1845–1849.
- [19] IBM SPSS statistics 22 Algorithms.
- [20] Izenman A.J. (2008), "Modern multivariate statistical techniques", Vol. 1. New York: Springer.
- [21] Lu C.J., Shao Y.E., and Li P.H. (2011), "Mixture control chart patterns recognition using independent component analysis and support vector machine", *Neurocomputing* 74, 1908–1914.
- [22] Mercer J. (1909), "Functions of positive and negative type and their connection with the theory of integral equations", *Philosophical Transactions of the Royal Society, London, Series A*, 209, 415–446.
- [23] Milborrow S. (2015), "Notes on the earth package".
- [24] Montgomery D.C. (2007), "Introduction to statistical quality control", *John Wiley Sons*.
- [25] Noorossana R., Saghaei A., Paynabar K., and Abdi S. (2009), "Identifying the period of a step change in high-yield processes", *Quality and Reliability Engineering International* 25, 875–883.
- [26] Pignatiello J.J., and Samuel T.R. (2001), "Estimation of the change point of a normal process mean in SPC applications", *Journal of Quality Technology* 33, 82–95.

- [27] Pignatiello J.J., and Samuel T.R. (2001), "Identifying the time of a step change in the process fraction nonconforming", *Quality Engineering* 13, 357–365.
- [28] Pyzdek T. (2003), "The Six Sigma Handbook", Revised and Expanded, McGraw-hill.
- [29] Samuel T.R., and Pignatiello J.J. (1998), "Identifying the time of a change in a Poisson rate parameter", *Quality Engineering* 10, 673–681.
- [30] Samuel T.R., Pignatiello J.J., and Calvin J.A. (1998), "Identifying the time of a step change in a normal process variance", *Quality Engineering* 10, 529–538.
- [31] Samuel T.R., Pignatiello J.J., and Calvin J.A. (1998), "Identifying the time of a step change with control charts", *Quality Engineering* 10, 521–527.
- [32] Shao Y.E. (2009), "An integrated neural networks and SPC approach to identify the starting time of a process disturbance", *ICIC Express Letters – An International Journal of Research and Surveys* 3, 319–324.
- [33] Shao Y.E., and Hou C.D. (2011), "A combined MLE and EWMA chart approach to estimate the change point of a gamma process with individual observations", *International Journal of Innovative Computing, Information and Control* 7, 2109–2122.
- [34] Shao Y.E., and Hou C.D. (2009), "Estimation of the change point of a uniform process using the EWMA chart and MLE", *ICIC Express Letters – An International Journal of Research and Surveys* 3, 451–456.
- [35] Shao Y.E., and Hou C.D. (2006), "Estimation of the starting time of a step change disturbance in a gamma process", *Journal of the Chinese Institute of Industrial Engineers* 23 319–327.
- [36] Shao Y.E., Hou C.D., and Wang H.J. (2006), "Estimation of the change point of a gamma process by using the S control chart and MLE", *Journal of the Chinese Institute of Industrial Engineers* 23, 207–214.
- [37] Shao Y.E., Huang H.Y., Hou C.D., Lin K.S., and Tsai J.E. (2011), "Change point determination for an attribute process", *ICIC Express Letters – An International Journal of Research and Surveys* 5, 3117–3122.
- [38] Shao Y.E., and Ke-Shan L. (2015), "Change Point Determination for an Attribute Process Using an Artificial Neural Network-Based Approach", *Discrete Dynamics in Nature and Society*.
- [39] Shao Y.E., and Lin Y. (2011), "Determination of Starting time of a multivariate disturbance using support vector machine and hotelling T^2 control chart", *ICIC Express Letters – An International Journal of Research and Surveys* 5, 1757–1761.

واژه‌نامه فارسی به انگلیسی

Statistical process control	کنترل کیفیت آماری
Shewhart control chart	نمودار کنترل شوهارت
Subgroup	زیرگروه
Upper control limit	حد کنترل بالایی
Lower control limit	حد کنترل پایینی
Change point	نقطه تغییر
Signal time	زمان هشدار
Single step change	تغییر پله‌ای یگانه
Multiple step change	تغییر پله‌ای چندگانه
Drift	شیب
Linear trend	روند خطی
Monotonic change	تغییر یکنوا
Isotonic change	تغییر هم‌نوا
Antitonic change	تغییر ناهم‌نوا
Maximum likelihood estimator	برآورد درست‌نمایی ماکزیمم
Machine learning method	روش یادگیری ماشین
False alarm	هشدار غلط
Multivariate logistic regression	رگرسیون لجستیک چندمتغیره
Multivariate adaptive regression spline	رگرسیون چندمتغیره سازوار اسپلاین
knot	گره
Support vector machine	ماشین بردار پشتیبان
Multicollinearity	همخطی چندگانه
Variance inflation factor	عامل تورم واریانس
Wald forward method	روش پیشرو والد
Score statistic	آماره نمره

Lagrange multiplier test	آزمون افزایشنده لاگرانژ
Basis function	توابع پایه‌ای
Recursive partitioning	قسمت‌بندی بازگشتی
Quadratic problem	مسئله درجه دو
Objective function	تابع هدف
Lagrange coefficient	ضریب لاگرانژ
Stationary points	نقاط ایستادن
Support vectors	بردارهای پشتیبان
Slack variable	متغیر کساد
Trade off	موازنه
Feature space	فضای ویژگی
Kernel trick	ترفند کرنل
Kernel function	تابع کرنل
Vector space	فضای برداری
Banach space	فضای باناخ
Cauchy sequence	دنباله کوچی
Normed space	فضای نرم‌دار
Inner product space	فضای ضرب داخلی
Hilbert space	فضای هیلبرت

واژه‌نامه انگلیسی به فارسی

Statistical process control	کنترل کیفیت آماری
Subgroup	زیرگروه
Change point	نقطه تغییر
Signal time	زمان هشدار
Single step change	تغییر پله‌ای یگانه
Multiple step change	تغییر پله‌ای چندگانه
Drift	شیب
Linear trend	روند خطی
Monotonic change	تغییر یکنوا
Isotonic change	تغییر هم‌نوا
Antitonic change	تغییر ناهم‌نوا
False alarm	هشدار غلط
Multivariate logistic regression	رگرسیون لجستیک چندمتغیره
Multivariate adaptive regression spline	رگرسیون چندمتغیره سازوار اسپلاین
knot	گره
Support vector machine	ماشین بردار پشتیبان
Multicollinearity	همخطی چندگانه
Variance inflation factor	عامل تورم واریانس
Wald forward method	روش پیشرو والد
Lagrange multiplier test	آزمون افزایشنده لاگرانژ
Basis function	توابع پایه‌ای
Recursive partitioning	قسمت‌بندی بازگشتی
Slack variable	متغیر کساد
Trade off	موازنه

نمایه

ا	آماره نمره، ۷۹
	آماره والد، ۸۰
ب	بردارهای پشتیبان، ۳۰
ت	ترفند کرنل، ۳۴
	تغییر پله‌ای یگانه، ۷
	توزیع دترمیان ماتریس واریانس کواریانس، ۳۹
ر	روش پیشرو والد، ۲۰
ش	شرایط KKT ، ۸۳
ض	ضریب لاگرانژ، ۲۹
ع	عامل تورم واریانس، ۷۵
ف	فضای باناخ، ۸۲
	فضای برداری، ۸۱
	فضای ضرب داخلی، ۸۲
	فضای نرم‌دار، ۸۲
	فضای هیلبرت، ۸۲
گ	گره، ۲۲
م	متغیر کساد، ۳۱
ن	نرم، ۸۱

Abstract

Effective detection of change point is one of the important issues that researchers, in the field of quality control, are dealing with. Most of these studies have been done by maximum likelihood estimation (MLE) or machine learning (ML) methods. Deficiency of MLE method is that, the distribution of the underlying process must be known, whereas it is not so in practice. Also, deficiency of ML method is that, when we have more input variables, computational cost is too much. In this study, we use a two-stage hybrid scheme to detect the change point of a multivariate process. In the first stage, statistical models are used to reduce the dimension of variable space, and in the second stage, a ML method is used to provide a basis for detection of change point. The statistical models are logistic regression (LR) and multivariate adaptive regression spline (MARS). These two methods provide a reduced set of inputs for ML method of support vector machine (SVM). Thereafter, an effective approach is used on the SVM output to detect the change point. Our achievements show that the hybrid scheme has much better performance than the case where the SVM is applied on the whole set of variables.

Keywords: Change point, Multivariate process, Logistic regression, MARS, SVM, Quality control.



Shahrood University of Technology

Faculty of Mathematical Sciences

**Change point detection in variance of
multivariate processes using a two-stage
hybrid scheme**

Amir Hashemi

Supervisor

Davood Shahsavani

December 2015