مجله تحقیق در عملیات در کاربردهای آن، سال بیست و یکم، شماره چهارم (پیاپی ۸۳)،زمستان ۰۳، ص ص ۱۴-۱
شاپا ۷۲۸۶-۲۲۵۱

الگوریتم کاهش گروهی در مدل جمعی ناپارامتری تاوانیده

محمد کاظمی[1*]
استادیار، گروه آمار، دانشکده علوم ریاضی، دانشگاه گیلان، رشت، ایران

رسید مقاله: ۲ تیر ۱۴۰۳
پذیرش مقاله: ۲۴ آبان ۱۴۰۳

چکیده

مدل جمعی ناپارامتری یکی از مدل‌های رایج برای مدل‌سازی رابطه بین متغیرها است. در این مقاله، مدل جمعی ناپارامتری بعد بالا را در نظر می‌گیریم که در آن تعداد متغیرهای توضیحی می‌تواند از تعداد مشاهدات بیشتر باشد، اما تعداد متغیرهای توضیحی موثر بر پاسخ نسبت به تعداد مشاهدات، کوچک است. وقتی تعداد متغیرهای توضیحی مدل زیاد باشد، تفسیر مدل مشکل‌تر و هزینه محاسبات افزایش می‌یابد. لذا، شناسایی متغیرهای توضیحی موثر بر پاسخ یا مولفه‌های جمعی غیرصفر در این مدل بسیار مهم است. بدین منظور، ابتدا مولفه‌های جمعی را با استفاده از پایه‌های B - اسپلاین تقریب می‌زنیم. با به‌کارگیری این تقریب، مساله انتخاب متغیر به انتخاب گروه‌هایی از ضرایب غیرصفر تبدیل می‌شود. سپس از تابع‌های تاوان گروهی برای انتخاب ضرایب غیرصفر استفاده می‌کنیم. این امر معمولاً با مینیمم‌کردن مجموع توان‌های دوم خطا تحت یک شرط محدودکننده انجام می‌شود. مینیمم‌کردن این تابع هدف تاوانیده مستلزم استفاده از روش‌های بهینه‌سازی است. در این مقاله از الگوریتم کاهش گروهی برای حل مساله مینیمم‌سازی فوق استفاده می‌شود. در پایان، عملکرد این الگوریتم تحت سه تابع تاوان مختلف، با مطالعات شبیه‌سازی و تحلیل یک مجموعه داده واقعی بررسی می‌شود.

کلمات کلیدی: انتخاب متغیر گروهی، بهینه‌سازی، داده‌های بعد بالا، مدل جمعی ناپارامتری.

۱ مقدمه

در مطالعه رابطه بین متغیرهای توضیحی و متغیر پاسخ، اطلاعات پیشینی که نشان‌دهنده وجود یک رابطه خطی باشد، به‌ندرت در دسترس است. بنابراین منطقی به نظر می‌رسد که به‌جای مدل خطی از مدل‌هایی با انعطاف‌پذیری بیشتری استفاده کنیم. یکی از این مدل‌ها، مدل جمعی ناپارامتری است که توسط هستی و تیبشیرانی در سال ۱۹۹۰ معرفی شد [۱]. در این مدل هیچ فرض محدودکننده قوی برای شکل تابع‌های تعیین‌کننده اثر متغیرهای توضیحی بر پاسخ در نظر گرفته نمی‌شود؛ بنابراین برای برآورد مولفه‌های جمعی نامعلوم باید از روش‌های ناپارامتری استفاده شود.

[1] عهده دار مکاتبات
آدرس الکترونیکی: m.kazemi@guilan.ac.ir

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

از طرفی، ظهور فناوری‌های نوین در دهه اخیر و امکان ذخیره‌سازی داده‌ها در ابعاد بزرگ موجب شده است که مساله مدل‌سازی پدیده‌ها، با تعداد صدها یا هزاران متغیر توضیحی مواجه باشد. اگرچه حضور تعداد زیادی از متغیرهای توضیحی انعطاف‌پذیری مدل را افزایش می‌دهد، اما باعث مشکل شدن تفسیر مدل می‌شود. علاوه‌براین، در مدل‌های با بعد بالا، مساله همخطی بسیار شدید است، یعنی هر یک از متغیرهای توضیحی را می‌توان به صورت ترکیب خطی از سایر متغیرها نوشت [۲]. این موضوع منجر به کاهش توان پیشگویی مدل می‌شود. روش پیشنهادی در این حالت، مدل‌سازی با زیرمجموعه‌ای از متغیرهای توضیحی است. در این رهیافت علاقه‌مندیم زیرمجموعه‌ای از متغیرها را طوری انتخاب کنیم که رابطه بین متغیر پاسخ و متغیرهای توضیحی را به خوبی توصیف کند. یافتن زیرمجموعه‌ای مناسب از متغیرهای توضیحی برای ورود به مدل انتخاب متغیر نامیده می‌شود.

برای درک بهتر موضوع انتخاب متغیر، مدل رگرسیون خطی چندگانه $Y=\beta_0+\beta_1X_1+....+\beta_pX_p+\varepsilon$ را در نظر بگیرید. در این مدل، برای برآورد متغیر پاسخ Y برحسب متغیرهای توضیحی $X_1,...,X_p$، به‌طور معمول از روش معروف مینیمم‌سازی مجموع توان‌های دوم خطا برای یافتن بهینه ضرایب رگرسیونی β_jها استفاده می‌شود. به عبارت دیگر

$$\hat{\boldsymbol{\beta}}=\arg\min_{\boldsymbol{\beta}}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_{i1}-....-\beta_px_{ip})^2 \qquad (1)$$

که در آن $\boldsymbol{\beta}=(\beta_0,\beta_1,...,\beta_p)'$ بردار ضرایب رگرسیونی است. به منظور انتخاب متغیرهای مهم، باید ضرایب رگرسیونی را طوری برآورد کنیم که برآورد ضرایب متغیرهای توضیحی زاید دقیقاً برابر صفر باشند، یعنی باید مساله مینیمم‌سازی مذکور، به‌طور مقید انجام شود و بدین‌وسیله تشخیص متغیرهای مهم و همچنین پیش‌بینی پاسخ در گرو حل مسایل بهینه‌سازی باشد. این قیود بسیار متنوع بوده و برحسب نوع ساختار آن، برنامه‌ریزی خطی، درجه دوم، محدب و غیرمحدب کاربرد پیدا می‌کند. پس از حل این مساله بهینه‌سازی، ضرایب برخی از متغیرهای توضیحی دقیقا برابر صفر برآورد می‌شوند که منجر به حذف متغیرهای توضیحی متناظر یا به‌عبارت‌دیگر، انتخاب متغیر می‌شود. در حالت کلی، برآوردگر کمترین توان‌های دوم تاوانیده با مینیمم‌کردن مجموع توان‌های دوم خطا تحت قید $\sum_{j=1}^{p}p(|\beta_j|)\le t$ به دست می‌آید که در آن $p(.)$ تابع تاوان نامیده می‌شود. این مساله مینیمم‌سازی را با استفاده از لاگرانژ می‌توان به صورت زیر نوشت

$$\hat{\boldsymbol{\beta}}=\arg\min_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}\beta_jx_{ij})^2+\lambda\sum_{j=1}^{p}p(|\beta_j|)\right\} \qquad (2)$$

که در آن λ پارامتر تنظیم‌کننده نامیده می‌شود. به عنوان نمونه، می‌توان تابع تاوان را به شکل $p(|\beta_j|)=|\beta_j|^d$ در نظر گرفت که به آن، تاوان بریج گویند [۳]. در ادبیات آماری، این تابع تاوان به ازای $d=1$ موسوم به تاوان لاسو[1] است که در این صورت، حل مساله مینیمم‌سازی (۲) یک مساله بهینه‌سازی محدب است [۴]. در این وضعیت، با

[1] Least absolute shrinkage and selection operator

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

تغییر مقدار پارامتر λ ممکن است برخی از ضرایب β_j برابر صفر شوند که بدین ترتیب متغیر متناظر آن، یعنی X_j، از مدل حذف‌شده و با حذف متغیرهای غیر مرتبط با متغیر پاسخ، انتخاب متغیر انجام می‌شود.

تابع تاوان لاسو با اینکه در بسیاری از مسایل انتخاب متغیر عملکرد خوبی از خود نشان داده، اما دارای برخی محدودیت‌ها است. به‌عنوان مثال، لاسو برای یک مدل رگرسیونی خطی با p متغیر پیشگو و n مشاهده، حداکثر n متغیر را انتخاب می‌کند. بنابراین اگر تعداد متغیرهای توضیحی معنی‌دار در مدل بیشتر از n باشد، برخی از آن‌ها توسط لاسو انتخاب نمی‌شوند [۵]. برای غلبه بر این محدودیت‌ها، تابع‌های تاوان متعددی به‌منظور برآورد و انتخاب متغیر معرفی‌شده‌اند. فن و لی در سال ۲۰۰۱ یک تابع تاوان غیر محدب به نام SCAD [1] را معرفی کردند و نشان دادند که این تاوان از سه ویژگی مطلوب نا اریبی، تنکی و پیوستگی برخوردار است [۶]. سپس زو در سال ۲۰۰۶ نشان داد برآوردگر لاسو از ویژگی پیشگویی [2] برخوردار نیست [۷]. او برای رفع این مشکل، تاوان لاسوی تطبیقی را معرفی کرد که دارای این ویژگی است. همچنین ژانگ در سال ۲۰۱۰ تابع تاوان غیر محدب دیگری به نام MCP [3] را معرفی کرد که تاوان لاسو را به عنوان یک حالت خاص در بر می‌گیرد [۸].

در این مقاله، با هدف نشان دادن نقش به‌سزای مسایل بهینه‌سازی در پیش‌بینی دقیق‌تر مدل‌های رگرسیونی، مساله انتخاب متغیرهای مهم را در یک مدل جمعی ناپارامتری با تابع‌های تاوان مختلف در نظر می‌گیریم. ابتدا به دلیل نامعلوم بودن مولفه‌های جمعی ناپارامتری، از پایه‌های B- اسپلاین برای تقریب آنها استفاده می‌کنیم. با تقریب مولفه‌های جمعی توسط پایه‌های B-اسپلاین، مدل جمعی ناپارامتری به یک مدل خطی با ساختار گروهی کاهش می‌یابد. بنابراین، مساله انتخاب متغیر به انتخاب گروه‌هایی از ضرایب غیرصفر تبدیل می شود و لذا باید از تابع‌های تاوان گروهی استفاده کرد که قابلیت انتخاب گروهی را دارند. در نتیجه، انتخاب متغیرهای مهم مستلزم حل یک مساله بهینه‌سازی با ساختار گروهی است.

با یک تابع تاوان محدب، مانند لاسو، مجموع توان‌های دوم خطا یک تابع محدب است و لذا برآورد کمترین توان‌های دوم تاوانیده را می‌توان با روش‌های بهینه‌سازی محدب به‌دست آورد. اما اغلب تابع‌های تاوان که دارای سه ویژگی مطلوب تنکی، نااریبی و پیوستگی می‌باشند، تابع‌هایی غیر محدب بوده و این موضوع باعث می‌شود که تابع هدف (۲) محدب نباشد. محدب نبودن این تابع، مساله مینیمم‌سازی را با مشکل مواجه می‌کند. در این راستا، دو رویکرد کلی وجود دارد: در رویکرد اول، ابتدا تابع تاوان غیر محدب را با یک تابع محدب تقریب زده و سپس مساله بهینه‌سازی غیر محدب را با استفاده از الگوریتم‌های بهینه‌سازی محدب حل می‌کنند. رویکرد دوم، استفاده از الگوریتم‌های کاهش مختصات است. از جمله الگوریتم‌های موجود برای تقریب تاوان‌های غیر محدب می‌توان به الگوریتم‌های تقریب درجه دوم موضعی و تقریب خطی موضعی اشاره کرد ([۶]، [۹]). اما هر یک از این الگوریتم‌ها دارای برخی نقاط ضعف هستند. ایراد وارد بر الگوریتم تقریب درجه دوم موضعی این است که اگر یک متغیر در هر تکرار حذف شود، در تکرار بعدی وارد مدل نمی‌شود و در نتیجه از مدل نهایی حذف می‌شود. بنابراین، این روش مشابه روش انتخاب پیشرو عمل می کند. برای غلبه بر این مشکل، زو و لی در سال ۲۰۰۸

[1] Smoothly Clipped Absolute Deviation
[2] Oracle property
[3] Minimax concave penalty

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

الگوریتم تقریب خطی موضعی را برای تقریب تابع‌های تاوان غیر محدب پیشنهاد دادند و برتری این الگوریتم را نسبت به الگوریتم تقریب درجه دوم موضعی نشان دادند [۹]. اما هیچ یک از این دو الگوریتم در مدل‌های رگرسیونی بعد بالا کارایی لازم را ندارند. در مقابل، رویکرد کاهش مختصات برای برازش مدل‌های رگرسیونی تاوانیده بعد بالا بسیار مناسب است. تاکنون انواع الگوریتم‌های کاهش مختصات در مدل‌های تاوانیده مورد استفاده قرار گرفته است که از جمله آنها می‌توان به [۱۴-۱۰] اشاره کرد. این الگوریتم‌ها در ابتدا برای تابع‌های تاوان محدب مانند لاسو معرفی شدند، اما سپس تعمیم‌هایی از آنها برای تاوان‌های SCAD و MCP نیز معرفی شدند. به عنوان نمونه، برهنی و هوآنگ در سال ۲۰۱۱ با به‌کارگیری رویکرد کاهش مختصات، الگوریتم تقریب خطی موضعی را بهبود دادند و نشان دادند که چگونه این رویکرد می‌تواند به طورکارا در برازش مدل‌های رگرسیونی با تاوان‌های غیرمحدب مانند SCAD و MCP موثر بوده و در مسایل با بعد بسیار بالا نیز قابل استفاده است [۱۵]. سپس برهنی و هوآنگ در سال ۲۰۱۵ این الگوریتم را به حالت گروهی تعمیم داده و الگوریتم کاهش گروهی را برای حل مسایل بهینه‌سازی در مدل‌های با ساختار گروهی معرفی کردند [۱۶].

مهم‌ترین مزیت الگوریتم ارایه‌شده در [۱۶] کارایی محاسباتی آن است، زیرا اجرای آن نیاز به هیچ بهینه‌سازی عددی پیچیده یا محاسبات ماتریسی دشوار مانند وارون‌گیری ندارد و تنها شامل تعداد اندکی عملیات حسابی ساده است. لازم به ذکر است که پیش از این الگوریتم دیگری توسط شی در سال ۲۰۱۲ برای انتخاب متغیر گروهی ارایه شده بود [۱۷] که مراحل آن بسیار شبیه الگوریتم کاهش گروهی در [۱۶] است. با این وجود، برتری الگوریتم ارایه شده در [۱۶] این است که برای حالت تک گروهی، این الگوریتم در هر مرحله جواب‌های دقیقی را تولید می‌کند. در مقابل، روش ارایه‌شده در شی [۱۷] نیاز به چندین تکرار برای همگرا شدن به همان جواب دارد، به عبارت دیگر، این الگوریتم پایدار نیست.

در ادامه، ساختار مقاله به صورت زیر است: در بخش ۲ مدل جمعی ناپارامتری معرفی می‌شود. در بخش ۳ به موضوع انتخاب متغیرگروهی و معرفی روش‌های متداول انتخاب گروهی پرداخته می‌شود. در بخش ۴ با معرفی الگوریتم کاهش گروهی به حل مساله بهینه‌سازی برای انتخاب متغیر می‌پردازیم. در بخش ۵ با مطالعات عددی، عملکرد الگوریتم مذکور در انتخاب متغیرهای مهم را تحت تابع‌های تاوان مختلف بررسی می‌کنیم. در بخش ۶ تحلیل یک مجموعه داده واقعی ارایه می‌شود.

۲ مدل جمعی ناپارامتری

مدل جمعی ناپارامتری به‌صورت زیر تعریف می‌شود:

$$Y = \sum_{j=1}^{p} f_j(X_j) + \varepsilon \qquad (۳)$$

که Y متغیر پاسخ مرکزی شده، $\mathbf{X} = (X_1,...,X_p)^T$ متغیرهای توضیحی، f_j ها توابع یک متغیره نامعلوم هموار با $E\left[f_j(X_j)\right] = 0$ و ε خطای تصادفی با میانگین صفر است. همچنین فرض می‌شود که برخی از f_j ها برابر صفر هستند. هدف، پیداکردن متغیرهای توضیحی مهم یا به‌عبارت‌دیگر، یافتن مولفه‌های غیرصفر در مدل فوق و

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

برآورد آنها است. از آنجا که برخلاف مدل خطی، تابع‌های f_j نامعلوم هستند، ابتدا باید این تابع‌ها را با استفاده از پایه‌های B-اسپلاین تقریب بزنیم. در این مقاله، از اسپلاین‌های چندجمله‌ای برای تقریب f_jها استفاده می‌شود. بدون از دست دادن کلیت، فرض می‌کنیم که X_j مقادیر خود را در بازه $[۰,۱]$ اختیار می‌کند و $\tau_۰ = ۰ < \tau_۱ < ... < \tau_{K'} < ۱ = \tau_{K'+۱}$ افرازی از $[۰,۱]$ به زیر بازه‌های $[\tau_k, \tau_{k+۱})$ با K' گره داخلی با فواصل یکسان باشد. یک اسپلاین چندجمله‌ای از درجه q با گره‌های $\tau_k, k = ۱,...,K'$، یک‌چند جمله‌ای تکه‌ای از درجه $q-۱$ است که مشتق آن تا مرتبه $q-۲$ موجود و پیوسته باشد. مجموعه اسپلاین‌ها با یک دنباله ثابت از گره‌ها، دارای پایه‌های B-اسپلاین نرمال شده $\{B_۱(x), B_۲(x), ..., B_{K'+q}(x)\}$ می‌باشند. برای اطلاعات بیشتر درباره توابع B-اسپلاین به [۱۸] و [۱۹] مراجعه کنید. بنابراین می‌توانیم مولفه‌های f_j را به‌صورت زیر تقریب بزنیم

$$f_j(x) \approx \sum_{k=۱}^{K} \beta_{jk} B_{jk}(x) = \boldsymbol{\beta}_j{}^T \mathbf{B}_j(x), \qquad j = ۱,..., p$$

که $\mathbf{B}_j(x) = (B_{j۱}(x),...,B_{jK}(x))^T$ پایه‌های B-اسپلاین، $\boldsymbol{\beta}_j = (\beta_{j۱},...,\beta_{jK})^T$ بردار ضرایب نامعلوم و $K = K' + q$ تعداد پایه‌ها می‌باشند. فرض کنید $(\mathbf{x}_۱, y_۱),...,(\mathbf{x}_n, y_n)$، n مشاهده از $(\mathbf{X}, Y)$ باشند که $\mathbf{x}_i = (x_{i۱},...,x_{ip})^T$ و $\|\mathbf{a}\| = \left(\sum_{i=۱}^{m} |a_i|^۲\right)^{۱/۲} = (\mathbf{a}'\mathbf{a})^{۱/۲}$ نرم اقلیدسی بردار $\mathbf{a} \in \mathbb{R}^m$ باشد. بنابراین، تابع هدف تاوانیده را می‌توانیم به‌صورت زیر بنویسیم

$$Q(\boldsymbol{\beta}) = \frac{۱}{۲n} \sum_{i=۱}^{n} \left(y_i - \sum_{j=۱}^{p} \boldsymbol{\beta}_j{}^T \mathbf{B}_j(x_{ij}) \right)^۲ + \sum_{j=۱}^{p} p_\lambda(\|\boldsymbol{\beta}_j\|) \qquad (۴)$$

که $p_\lambda(.)$ تابع جریمه و λ پارامتر تنظیم‌کننده است که میزان انقباض ضرایب رگرسیونی را کنترل می‌کند. با استفاده از نمایش ماتریسی می‌توان تابع $Q(\boldsymbol{\beta})$ را به‌صورت زیر ساده کرد

$$Q(\boldsymbol{\beta}) = \frac{۱}{۲n} \|\mathbf{y} - \mathbf{Z}\boldsymbol{\beta}\|^۲ + \sum_{j=۱}^{p} p_\lambda(\|\boldsymbol{\beta}_j\|) \qquad (۵)$$

که $\mathbf{Z}_{(n \times pK)} = (\mathbf{Z}_۱,...,\mathbf{Z}_p)$، $\boldsymbol{\beta} = (\boldsymbol{\beta}_۱{}^T,...,\boldsymbol{\beta}_p{}^T)^T$، $\mathbf{y} = (y_۱,...,y_n)^T$ است و $\mathbf{Z}_j$ به‌صورت زیر است:

$$\mathbf{Z}_j = \begin{pmatrix} B_{j۱}(x_{۱j}) & B_{j۲}(x_{۱j}) & ... & B_{jK}(x_{۱j}) \\ B_{j۱}(x_{۲j}) & B_{j۲}(x_{۲j}) & ... & B_{jK}(x_{۲j}) \\ \vdots & \vdots & \vdots & \vdots \\ B_{j۱}(x_{nj}) & B_{j۲}(x_{nj}) & ... & B_{jK}(x_{nj}) \end{pmatrix}.$$

با تقریب f_jها توسط پایه‌های B-اسپلاین، ماتریس طرح $\mathbf{Z}$ ساختار گروهی پیدا می‌کند، یعنی K تابع پایه‌ای اول برای تقریب $f_۱$، K تابع پایه‌ای دوم برای تقریب $f_۲$ و به همین ترتیب، K تابع پایه‌ای pام برای تقریب f_p به کار می‌روند. بنابراین برای برقراری $f_j(x_j) = ۰$، باید همه مؤلفه‌های بردار $\boldsymbol{\beta}_j = (\beta_{j۱},...,\beta_{jK})^T$ صفر شوند، یعنی باید از روش‌های انتخاب متغیر گروهی استفاده کرد. از جمله تابع‌های تاوان رایج برای انتخاب متغیر

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

گروهی می‌توان تاوان لاسوی گروهی، تاوان SCAD گروهی و MCP گروهی را نام برد. این تابع‌های تاوان در [۲۳-۲۰، ۱۱] مورد بررسی قرار گرفته‌اند. همچنین یک روش بیزی موسوم به SSGL[1] برای انتخاب متغیر گروهی براساس تابع تاوان لاسوی گروهی توسط بای و همکاران در سال ۲۰۲۲ ارایه شده است [۲۴]. در بخش بعد ابتدا به معرفی این تابع‌ها و سپس الگوریتم کاهش گروهی برای حل مساله بهینه‌سازی می‌پردازیم.

۳ روش‌های انتخاب متغیر گروهی

در رویکرد انتخاب متغیر گروهی، هنگام برآورد و انتخاب متغیرهای مهم، هر گروه از متغیرها را به عنوان یک واحد در نظر می‌گیرند و این عمل فرآیند انتخاب متغیر را تسهیل می‌کند. در نتیجه متغیرهای یک گروه، یا همگی از مدل حذف‌شده و یا همگی در مدل باقی می‌مانند. ساختار گروهی متغیرها می‌تواند دلایل و اهداف متفاوتی داشته باشد. به عنوان مثال، در تحلیل داده‌های بیان ژنی، ژن‌های متعلق به یک مسیر بیولوژیکی یکسان می‌توانند در یک گروه قرار بگیرند. همچنین در مطالعات پیوند ژنتیکی، نشانگرهای ژنتیکی از ژن یکسان می‌توانند به عنوان یک گروه در نظر گرفته شوند. مطلوب است ساختار گروهی را در تحلیل این نوع داده‌ها مورد توجه قرار دهیم.

ابتدا تابع تاوان لاسوی گروهی توسط یوآن و لین [۱۱] به صورت زیر معرفی شد

$$P(\boldsymbol{\beta}) = \lambda \sum_{j=1}^{p} c_j \left\| \boldsymbol{\beta}_j \right\|$$

که λ پارامتر جریمه و c_jها برای تعدیل اندازه گروه‌ها استفاده می‌شوند. یک انتخاب مناسب $c_j = \sqrt{d_j}$ است که d_j اندازه گروه jام می‌باشد. با توجه به یکسان بودن اندازه تمام گروه‌ها، $c_j = \sqrt{K}$ در نظر گرفته شده است.

همان‌طور که در بخش ۱ بیان شد، تاوان لاسو از ویژگی پیشگویی برخوردار نیست. دلیل این مشکل آن است که تاوان لاسو برای تمام ضرایب مدل میزان انقباض یکسانی را در نظر می‌گیرد. این امر باعث می‌شود که مدل انتخاب‌شده توسط لاسو معمولا بزرگ‌تر از مدل واقعی باشد، زیرا قادر به تشخیص ضرایب کوچک از ضرایب صفر نیست. تابع تاوان لاسوی گروهی هم رفتاری شبیه لاسو داد و گروه‌های بی‌اهمیت زیادی را وارد مدل می‌کند. لذا، منطقی به نظر می‌رسد که از تاوان‌های مناسب‌تری مانند SCAD گروهی و MCP گروهی استفاده کنیم. این دو تابع تاوان همه ویژگی‌های مطلوب یک تابع تاوان مناسب را دارند. در صورت استفاده از تاوان‌های SCAD گروهی و MCP گروهی، عبارت تاوان در رابطه (۵) دارای شکل کلی زیر است:

$$P(\boldsymbol{\beta}) = \sum_{j=1}^{p} \rho_{\gamma}(\left\| \boldsymbol{\beta}_j \right\|; c_j \lambda)$$

که در آن، برای تابع تاوان MCP گروهی، $\rho_{\gamma}(x;\lambda) = \lambda \int_{0}^{|x|} (1 - \frac{t}{\lambda\gamma})_{+} dt,\ \gamma > 1$ و $x_{+} = xI\{x \geq 0\}$.

هنگامی‌که $\gamma \to \infty$، MCP گروهی به لاسوی گروهی همگرا می‌شود. همچنین، برای تابع تاوان SCAD گروهی $\rho_{\gamma}(x;\lambda) = \lambda \int_{0}^{|x|} \min\{1, \frac{(\gamma - \frac{t}{\lambda})_{+}}{\gamma - 1}\}_{+} dt,\ \gamma > 2$ است.

[1] Spike-and-Slab Group Lasso

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

٤ حل مساله بهینه‌سازی

برای مینیمم‌کردن تابع هدف (۵)، ابتدا یک حالت خاص از این مساله را در نظر می‌گیریم، یعنی فرض می‌کنیم که ماتریس‌های $\mathbf{Z}_j$ متعامد باشند. تحت این فرض برآوردگرهای صریحی برای ضرایب به‌دست می‌آید. در ادامه با معرفی الگوریتم کاهش گروهی، ازآنجاکه این الگوریتم هر بار تابع هدف را نسبت به یک گروه تکی بهینه می‌کند، این برآوردگرها در الگوریتم کاهش گروهی برای بهینه‌سازی استفاده خواهند شد.

٤-١ گروه‌های متعامد

در این بخش، فرض می‌کنیم که گروه‌ها متعامد هستند به طوری که $\mathbf{Z}_j'\mathbf{Z}_j = n\mathbf{I}_K, \mathbf{Z}_j'\mathbf{Z}_k = ۰,\ \mathrm{k} \neq j$. در این حالت مساله به برآورد p مدل رگرسیونی K متغیره تک گروهی به‌صورت $y = Z_j\beta_j + \varepsilon,\ j = ۱,...,p$ کاهش می‌یابد. فرض کنید $\tilde{\boldsymbol{\beta}}_j = \frac{۱}{n}\mathbf{Z}_j'\mathbf{y}$ برآورد کمترین توان‌های دوم معمولی برای بردار ضرایب $\boldsymbol{\beta}_j$ باشد. بدون از دست دادن کلیت فرض می‌کنیم $c_j = ۱$. از طرفی چون $\mathbf{Z}_j'\mathbf{Z}_j = n\mathbf{I}_K$، بنابراین

$$\frac{۱}{n}\left\|\mathbf{y} - \mathbf{Z}_j\boldsymbol{\beta}_j\right\|^۲ = \left\|\tilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\right\|^۲ + \frac{۱}{n}\|\mathbf{y}\|^۲ - \left\|\tilde{\boldsymbol{\beta}}_j\right\|^۲.$$

در نتیجه، تابع هدف تاوانیده برای مدل $y = \mathbf{Z}_j\boldsymbol{\beta}_j + \varepsilon$ را می‌توان به‌صورت زیر نوشت

$$\frac{۱}{۲n}\left\|\mathbf{y} - \mathbf{Z}_j\boldsymbol{\beta}_j\right\|^۲ + \rho(\left\|\boldsymbol{\beta}_j\right\|;\lambda,\gamma) = \frac{۱}{۲}\left\|\tilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\right\|^۲ + \rho(\left\|\boldsymbol{\beta}_j\right\|;\lambda,\gamma).$$

برای برآورد کمترین توان‌های دوم تاوانیده بردار $\boldsymbol{\beta}_j$، ابتدا فرض کنید $H(\tilde{\boldsymbol{\beta}}_j;\lambda) = u\left(\left\|\tilde{\boldsymbol{\beta}}_j\right\|;\lambda\right)\frac{\tilde{\boldsymbol{\beta}}_j}{\left\|\tilde{\boldsymbol{\beta}}_j\right\|}$ که در آن $u(.,.)$ به صورت زیر تعریف می‌شود

$$u(t;\lambda) = \begin{cases} t-\lambda & t > \lambda \\ ۰ & |t| \leq \lambda \\ t+\lambda & t < -\lambda. \end{cases}$$

اکنون با در نظر گرفتن $\rho(.)$ برابر تابع‌های تاوان لاسو، SCAD و MCP، می‌توان برآوردگرهای لاسوی گروهی، SCAD گروهی و MCP گروهی را به‌صورت زیر به‌دست آورد

$$Group\ LASSO:\ \hat{\beta}_j = H(\tilde{\beta}_j;\lambda) \qquad (۶)$$

$$Group\ SCAD:\ \hat{\beta}_j = F_S(\tilde{\beta}_j;\lambda,\gamma) = \begin{cases} \frac{\gamma}{\gamma - ۱}H(\tilde{\beta}_j;\lambda) & \left\|\tilde{\beta}_j\right\| \leq \gamma\lambda \\ \tilde{\beta}_j & \left\|\tilde{\beta}_j\right\| > \gamma\lambda \end{cases} \qquad (۷)$$

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

$$Group\ MCP:\ \ \widehat{\beta}_j = F_M(\tilde{\beta}_j;\lambda,\gamma) = \begin{cases} H(\tilde{\beta}_j;\lambda) & \left\|\tilde{\beta}_j\right\| \le ۲ \\ \frac{\gamma-۱}{\gamma-۲} H(\tilde{\beta}_j;\frac{\gamma\lambda}{\gamma-۱}), & ۲\lambda < \left\|\tilde{\beta}_j\right\| \le \gamma\lambda \\ \tilde{\beta}_j & \left\|\tilde{\beta}_j\right\| > \gamma\lambda. \end{cases} \quad (۸)$$

برآوردگرهای فوق به‌عنوان بلوک‌های ساختمان الگوریتم کاهش گروهی در بخش ۴-۲ به کار می‌روند.

٤-۲ الگوریتم کاهش گروهی

الگوریتم‌های کاهش هر بار تابع هدف را نسبت به یک پارامتر تکی بهینه می‌کنند و این چرخه به‌طور مکرر برای سایر پارامترها تا رسیدن به همگرایی ادامه می‌یابد. به‌طور مشابه، الگوریتم کاهش گروهی هر بار تابع هدف را نسبت به یک گروه تکی بهینه می‌کند و این عمل بهینه‌سازی برای سایر گروه‌ها تا رسیدن به همگرایی تکرار می‌شود. این الگوریتم به‌ویژه برای برازش مدل‌های لاسوی گروهی، SCAD گروهی و MCP گروهی مناسب می‌باشند، چون همان‌طور که دیدیم، هر سه برآوردگر مذکور در یک مدل تک-گروهی دارای شکل بسته می‌باشند.

هر مرحله از الگوریتم کاهش گروهی شامل بهینه‌سازی مجموع توان‌های دوم خطا نسبت به ضرایب گروه j می‌باشد. لذا تعریف کنید

$$Q_j(\boldsymbol{\beta}_j) = \frac{۱}{۲n}\left\|\mathbf{y} - \sum_{k\neq j}\mathbf{Z}_k\tilde{\boldsymbol{\beta}}_k - \mathbf{Z}_j\boldsymbol{\beta}_j\right\|^۲ + \rho(\left\|\boldsymbol{\beta}_j\right\|;c_j\lambda,\gamma)$$

که $\tilde{\boldsymbol{\beta}}_k$ آخرین مقدار به روز شده $\boldsymbol{\beta}_k$ است. حال $\tilde{\mathbf{y}}_j = \sum_{k\neq j}\mathbf{Z}_k\tilde{\boldsymbol{\beta}}_k$ و $\tilde{\mathbf{s}}_j = \frac{1}{n}\mathbf{Z}_j'(\mathbf{y}-\tilde{\mathbf{y}}_j)$ را در نظر بگیرید. دقت کنید که $\tilde{\mathbf{y}}_j$ همان مقادیر برازش شده بدون در نظر گرفتن گروه jام و $\tilde{\mathbf{s}}_j$ برآورد کمترین توان‌های دوم معمولی $\boldsymbol{\beta}_j$ پس از تعدیل اثر سایر گروه‌ها روی متغیر پاسخ است. بنابراین، مشابه رگرسیون کمترین توان‌های دوم معمولی، مقدار $\boldsymbol{\beta}_j$ که مینیمم‌کننده $Q_j(\boldsymbol{\beta}_j)$ باشد، برابر با مقدار $\boldsymbol{\beta}_j$ ای است که از رگرسیون کردن $\mathbf{Z}_j$ روی باقیمانده‌های جزئی، یعنی $\mathbf{y}-\tilde{\mathbf{y}}_j$، به‌دست می‌آید. با توجه به بخش ۴-۱ می‌دانیم که مینیمم‌کننده $Q_j(\boldsymbol{\beta}_j)$ برابر $F(\tilde{\mathbf{s}}_j;\lambda,\gamma)$ است که $F(.)$ با توجه به تابع تاوان، یکی از جواب‌های (۶) تا (۸) می‌باشد.

فرض کنید $\tilde{\boldsymbol{\beta}}^{(\cdot)} = (\tilde{\boldsymbol{\beta}}_۱^{(\cdot)'},...,\tilde{\boldsymbol{\beta}}_p^{(\cdot)'})^T$ مقادیر اولیه ضرایب و t نشان‌دهنده تکرار باشد. الگوریتم کاهش گروهی در الگوریتم ۱ خلاصه شده است. در این الگوریتم با قرار دادن $F(.;\lambda,\gamma)$ برابر یکی از جواب‌های $\widehat{\boldsymbol{\beta}}_j$ در روابط (۶) تا (۸)، برآوردهای لاسوی گروهی، SCAD گروهی و MCP گروهی را برای بردار ضرایب $\boldsymbol{\beta}$ در مدل (۵) به‌دست می‌آوریم. در مرحله ۴ الگوریتم ۱، همگرایی زمانی رخ می‌دهد که $\left\|\tilde{\boldsymbol{\beta}}_j^{(t+۱)} - \tilde{\boldsymbol{\beta}}_j^{(t)}\right\| < \varepsilon$.

در این الگوریتم با قرار دادن $F(.;\lambda,\gamma)$ برابر یکی از جواب‌های $\widehat{\boldsymbol{\beta}}_j$ در روابط (۶) تا (۸)، برآوردهای لاسوی گروهی، SCAD گروهی و MCP گروهی را برای بردار ضرایب $\boldsymbol{\beta}$ در مدل(۵) به‌دست می‌آوریم. در مرحله ۴ الگوریتم ۱، همگرایی زمانی رخ می‌دهد که $\left\|\tilde{\boldsymbol{\beta}}_j^{(t+۱)} - \tilde{\boldsymbol{\beta}}_j^{(t)}\right\| < \varepsilon$.

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

الگوریتم ۱: الگوریتم کاهش گروهی

- *مرحله ۱:* قرار دهید $t=0$ و بردار اولیه باقیمانده‌ها را به‌صورت $\mathbf{r}=\mathbf{y}-\tilde{\mathbf{y}}$ در نظر بگیرید که $\tilde{\mathbf{y}}=\sum_{j=1}^{p}\mathbf{Z}_j\tilde{\boldsymbol{\beta}}_j^{(0)}$.
- *مرحله ۲:* برای $j=1,\ldots,p$ محاسبات زیر را انجام دهید:

  آ. $\tilde{\mathbf{s}}_j=\frac{1}{n}\mathbf{Z}_j'\mathbf{r}+\tilde{\boldsymbol{\beta}}_j^{(t)}$

  ب. $\tilde{\boldsymbol{\beta}}_j^{(t+1)}\leftarrow F(\tilde{\mathbf{s}}_j;\lambda,\gamma)$

  پ. $\mathbf{r}\leftarrow\mathbf{r}-\mathbf{Z}_j(\tilde{\boldsymbol{\beta}}_j^{(t+1)}-\tilde{\boldsymbol{\beta}}_j^{(t)})$
- *مرحله ۳:* به‌روزرسانی $t+1\leftarrow t$.
- *مرحله ۴:* مراحل۲ و ۳ را تا رسیدن به همگرایی تکرار کنید.

این الگوریتم از دو مزیت عمده برخوردار است: الف) از نظر هزینه محاسباتی بسیار کارا است. ب) چون در هر مرحله به‌روزرسانی، تابع هدف نسبت به $\boldsymbol{\beta}_j$مینیمم می‌شود، این الگوریتم دارای ویژگی نزولی[1] است; یعنی الگوریتم با هر تکرار، تابع هدف را کاهش می‌دهد. از طرفی، برهنی و هوآنگ [۱۶] نشان دادند که $Q_j(\boldsymbol{\beta}_j)$ یک تابع اکیداً محدب نسبت به $\boldsymbol{\beta}_j$ است. بنابراین، ویژگی نزولی این الگوریتم با توجه به اکیداً محدب بودن $Q_j(\boldsymbol{\beta}_j)$ ، همگرایی الگوریتم را نتیجه می‌دهد.

لازم به ذکر است که در صورت استفاده از تاوان لاسوی گروهی، که تابع هدف تاوانیده (۵) محدب است، قضیه فوق همگرایی الگوریتم به مینیمم سراسری را ثابت می‌کند. اما برای SCAD گروهی و MCP گروهی که تابع هدف (۵) مجموع دو تابع محدب و غیر محدب است، همگرایی به یک مینیمم موضعی امکان‌پذیر است.

۵ مطالعات شبیه‌سازی

در این بخش، با استفاده از شبیه‌سازی، عملکرد الگوریتم کاهش گروهی را در حل مساله بهینه‌سازی مربوط به انتخاب متغیرهای مهم تحت سه تابع تاوان مختلف بررسی می‌کنیم. در این مثال فرض می‌کنیم که حجم نمونه برابر $n=200$ و تعداد متغیرهای توضیحی برابر $p=100,500,1000$ باشند. همچنین X_i ها متغیرهای تصادفی مستقل یکنواخت $(0,1)$ و خطاها دارای توزیع نرمال استاندارد فرض شده‌اند. دو مدل جمعی ناپارامتری زیر را در نظر بگیرید:

- مدل (۱): داده‌ها از مدل $Y=\sum_{j=1}^{4}f_j(X_j)+\sqrt{1.74}\varepsilon$ تولید شده‌اندکه در آن

$$f_1(x)=5x,\quad f_2(x)=3(2x-1)^2,\quad f_3(x)=\frac{4sin(2\pi x)}{2-sin(2\pi x)},$$
$$f_4(x)=6(0/1sin(2\pi x)+0/2cos(2\pi x)+0/3sin(2\pi x)^2+0/4cos(2\pi x)^3+0/5sin(2\pi x)^3).$$

[1] Descent property

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

- مدل (۲): داده ها از مدل $Y=\sum_{j=1}^{6} f_j(X_j)+\varepsilon$ تولید شده‌اند که در آن

$$f_1(x)=\frac{2(e^{-10x}-e^{-10})}{1-e^{-10}}-1,\qquad f_2(x)=\frac{-2(e^{-10x}-e^{-10})}{1-e^{-10}}+1,\qquad f_3(x)=2x^3-1,$$
$$f_4(x)=-2x^3+1,\qquad f_5(x)=8(x-0/5)^2-1,\qquad f_6(x)=-8(x-0/5)^2+1.$$

در مدل (۱)، متغیرهای (X_1,X_2,X_3,X_4)و در مدل (۲)، متغیرهای $(X_1,X_2,X_3,X_4,X_5,X_6)$ متغیرهای توضیحی موثر بر پاسخ هستند و بقیه متغیرهای توضیحی در پیشگویی پاسخ هیچ کمکی نمی‌کنند. هدف، پیدا کردن متغیرهای توضیحی مهم، یا بطور معادل شناسایی مولفه‌های جمعی غیر صفر است.

ابتدا برای تقریب f_j ها، از پایه‌های B- اسپلاین مکعبی با ۴ گره، با فاصله‌های یکسان، استفاده می‌شود. برای تابع تاوان MCP گروهی مقدار پارامتر $\gamma=3$ ، برای SCAD گروهی مقدار $\gamma=4$ و برای انتخاب پارامتر λ از روش اعتبارسنجی متقابل تعمیم‌یافته (GCV) استفاده‌شده است. معیار نتایج شبیه‌سازی پس از ۵۰۰ بار تکرار در جدول۱ خلاصه شده است. در این جدول، NS میانگین تعداد مولفه‌های غیرصفر انتخاب‌شده، NST میانگین تعداد مولفه‌های غیرصفر انتخاب شده که در مدل واقعی غیرصفر هستند، IN درصد دفعاتی که همه مولفه‌های غیر صفر در مدل واقعی، توسط این روش به عنوان مولفه‌های غیرصفر تشخیص داده شده‌اند و RMSE ریشه میانگین مربع خطا در ۵۰۰ بار تکرار، گزارش‌شده است. مقدار RMSE از رابطه زیر محاسبه می‌شود:

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mu_i-\hat{\mu}_i)^2}$$

که در آن μ_i میانگین واقعی و $\hat{\mu}_i$ میانگین برآورد شده Y_i به‌شرط معلوم بودن X_i می‌باشد.

در جدول ۱ مشاهده می‌شود که الگورتیم کاهش گروهی در حل مساله بهینه‌سازی (۵) تحت تابع‌های تاوان لاسوی گروهی، SCAD گروهی و MCP گروهی عملکرد مناسبی دارد. در مدل ۱، تحت هر یک از سه تابع تاوان، در ۱۰۰٪ دفعات متغیرهای مهم (X_1,X_2,X_3,X_4) به درستی انتخاب می‌شوند. با این وجود، تابع تاوان لاسویگروهی در مقایسه با دو تابع تاوان دیگر، تمایل به انتخاب مدل‌های بزرگ‌تر دارد. براساس معیار RMSE می‌توان دید که مقدار خطای مدل برازش شده با تابع تاوان لاسوی گروهی در مقایسه با دو مدل دیگر اندکی بیشتر، و خطای متناظر با تابع تاوان MCP گروهی نسبت به دو مدل دیگر کمتر است، اما تفاوت خطای این مدل‌ها بسیار ناچیز است.

در مدل (۲)، تحت هر یک از تابع‌های تاوان، متوسط تعداد متغیرهای مهم انتخاب‌شده نزدیک به مقدار واقعی، یعنی ۶ است. در این مدل نیز تابع تاوان لاسویگروهی گرایش به انتخاب مدل‌های با اندازه بزرگ‌تر را دارد و در مقابل، تابع تاوان MCP گروهی مدل کوچک‌تری را نتیجه می‌دهد. اگرچه برای $p=100$ مناسبت مدل های برازش شده با هریک از سه تابع تاوان را می‌توان نتیجه گرفت، اما با افزایش p، کفایت مدل جمعی ناپارامتری برازش شده به کمک الگوریتم کاهش گروهی کاهش می‌یابد.

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

جدول۱. نتایج شبیه‌سازی برای بررسی عملکرد الگوریتم کاهش گروهی در برازش مدل جمعی ناپارامتری با تابع‌های تاوان لاسوی گروهی، SCAD گروهی و MCP گروهی. NS میانگین تعداد مولفه‌های غیر صفر انتخاب‌شده، NST میانگین تعداد مولفه‌های غیرصفر انتخاب‌شده که به درستی انتخاب شده‌اند، IN درصد دفعاتی که همه مولفه‌های غیر صفر در مدل واقعی، به درستی تشخیص داده شده‌اند و RMSE خطای مدل در ۵۰۰ بار تکرار. (اعداد داخل پرانتز انحراف معیار را نشان می‌دهند.)

مدل	p	LASSO				SCAD				MCP			
		RMSE	IN	NST	NS	RMSE	IN	NST	NS	RMSE	IN	NST	NS
مدل ۱	۱۰۰	۵/۱۳	۱۰۰	۴ (۰)	۲۳/۶۲ (۹/۲۴)	۵/۰۷	۱۰۰	۴ (۰)	۱۳/۷۳ (۷/۷۷)	۵/۹۰	۱۰۰	۴ (۰)	۵/۷۴ (۲ /۳۲)
	۵۰۰	۵/۱۸	۱۰۰	۴ (۰)	۳۵/۴۶ (۱۶/۸۵)	۵/۱۰	۱۰۰	۴ (۰)	۲۱/۰۵ (۱۳ /۷۳)	۵/۰۸	۱۰۰	۴ (۰)	۷/۰۲ (۳ /۵۹)
	۱۰۰۰	۵/۱۵	۱۰۰	۴ (۰)	۴۰/۸۲ (۱۷/۵۸)	۵/۰۹	۱۰۰	۴ (۰)	۳۲/۳۷ (۱۵ /۷۵)	۵/۰۷	۱۰۰	۴ (۰)	۷/۶۸ (۳ /۷۰)
مدل ۲	۱۰۰	۰/۷۴	۹۷	۵/۹۷ (۰ /۱۸)	۲۷/۹۷ (۱۰/۳۷)	۰/۶۴	۹۷	۵/۹۷ (۰ /۱۷)	۲۵/۶۷ (۶/۴۷)	۰/۶۱	۹۱	۵/۹۱ (۰ /۳۱)	۱۰/۵ (۲ /۶۴)
	۵۰۰	۰/۸۳	۷۹	۵/۷۷ (۰/۴۵)	۴۱/۰۹ (۱۷/۳۷)	۰/۷۱	۹۱	۵/۸۹ (۰/۳۳)	۴۵/۱۸ (۸/۶۶)	۰/۶۷	۶۸	۵/۶۳ (۰ /۵۷)	۱۲/۸۵ (۳ /۳۷)
	۱۰۰۰	۰/۹۰	۶۷	۵/۶۰ (۰/۶۴)	۴۵/۸۶ (۲۲ /۴۰)	۰/۷۶	۷۹	۵/۷۴ (۰/۵۲)	۵۰/۵۱ (۱۱ /۰۶)	0/71	۵۵	۵/۴۰ (۰ /۷۷)	۱۴/۱۷ (۳ /۲۵)

٦ مثال کاربردی

در این بخش، برای بررسی کارایی الگوریتم کاهش گروهی در حل مساله بهینه‌سازی مدل‌بندی داده‌های بعد بالا، مجموعه داده‌های مربوط به تولید ریبوفلاوین (ویتامین2B) در باسیلوس سوبتیلیس را در نظر می‌گیریم که توسط بالمن و همکاران در سال ۲۰۱۴ مورد استفاده قرار گرفته [۲۵] و در بسته *"hdi"* نرم افزار R در دسترس است. در این داده‌ها، لگاریتم نرخ تولید ریبوفلاوین به عنوان متغیر پاسخ و لگاریتم سطح بیان $p = ۴۰۸۸$ به عنوان متغیرهای توضیحی در نظر گرفته می‌شوند. این مجموعه داده شامل اطلاعات مربوط به $n = ۷۱$ است. هدف، پیدا کردن ژن‌های مؤثر در پیش بینی نرخ تولید ریبوفلاوین است.

برای مدل‌سازی رابطه بین نرخ تولید ریبوفلاوین و ۴۰۸۸ ژن، از مدل جمعی ناپارامتری (۳) استفاده می‌شود. برای برازش این مدل، تابع‌های تاوان لاسوی گروهی، SCAD گروهی و MCP گروهی را به‌کار می‌بریم. همچنین متغیرهای توضیحی را طوری مقیاس‌بندی می‌کنیم که مقادیر آنها بین ۰ و ۱ باشند و از اسپلاین مکعبی با ۴ گره برای برآورد مؤلفه‌های جمعی استفاده می‌شود. جدول ۲ تعداد ژن‌های انتخاب شده، مجموع توان‌های دوم باقیمانده‌ها (RSS) و ضریب تعیین $(R^۲)$ متناظر با هر روش را نشان می‌دهدکه این مقادیر، به ترتیب، به صورت $RSS = \sum_{i=۱}^{n}(y_i - \hat{y}_i)^۲$ و $R^۲ = ۱ - \frac{RSS}{S_{yy}}$ محاسبه می‌شوند. برای یافتن مقدار بهینه پارامتر λ از معیارهای اعتبارسنجی متقابل تعمیم یافته (GCV)، معیار اطلاع بیزی (BIC) و معیار اطلاع آکائیک (AIC) استفاده شده است. این معیارها به صورت زیر محاسبه می‌شوند:

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

$$AIC = L + ۲df, \qquad BIC = ۲L + log(n)df, \qquad GCV = \frac{۲L}{(۱-\frac{df}{n})^۲},$$

که L مجموع توان‌های دوم خطا (یا لگاریتم تابع درستنمایی)، n تعداد مشاهدات و df تعداد ضرایب غیرصفر مدل به ازای λ داده شده است

با توجه به جدول ۲، برای تابع تاوان MCP، مقدار بهینه λ براساس معیارهای مختلف، مدل یکسانی را نتیجه می‌دهد. ژن‌های انتخاب شـده در این مدل عبارت از XKDI_at،YCKE_at،YFHI_at،YJZB_at،YTGD_at، YURN_at،YVFK_at YWRE_at هستند. اگرچه مقادیر RSS و $R^۲$ مناسبت هر یک از مدل‌های برازش‌شده با هر سـه تابع تاوان را تایید می‌کنند، اما تابع تاوان لاسـوی گروهی تعداد ژن‌های بیشـتری را انتخاب می‌کند و تابع تاوان MCP گروهی با انتخاب ۸ ژن، مدل مقرون به صـرفه‌تری را نتیجه می‌دهد. به طور خلاصـه، با توجه به اندازه مدل و معیارهای نیکویی برازش، تابع تاوان MCP گروهی مدل بهتری را به داده‌ها برازش می‌دهد.

جدول ۲. نتایج تحلیل داده‌های واقعی. NS تعداد ژن‌های انتخاب‌شده، RSS مجموع توان‌های دوم باقیمانده‌ها، $R^۲$ ضریب تعیین

	MCP			SCAD			LASSO		
	NS	RSS	$R^۲$	NS	RSS	$R^۲$	NS	RSS	$R^۲$
GCV	۸	**۰/۱۰۲**	**۹۹/۸۲**	۲۱	۰/۳۲۸	**۹۹/۴۴**	۳۴	۱/۵۷	**۹۷/۳۴**
BIC	۸	**۰/۱۰۲**	**۹۹/۸۲**	۱۹	۰/۴۶۶	**۹۹/۲۱**	۱۴	۱۲/۴۰	**۷۹/۰۸**
AIC	۸	**۰/۱۰۲**	**۹۹/۸۲**	۲۱	۰/۳۲۸	**۹۹/۴۴**	۳۸	۰/۷۳۹	**۹۸/۷۵**

۷ بحث و نتیجه‌گیری

در این مقاله، موضوع برآورد و انتخاب متغیر در مدل جمعی ناپارامتری بعد بالا با استفاده از رگرسیون تاوانیده مورد بررسی قرار گرفت. تحقق این امر مستلزم استفاده از روش‌های بهینه‌سازی برای مینیمم کردن مجموع توان‌های دوم خطای تاوانیده است. در این راستا، از الگوریتم کاهش گروهی برای حل مساله بهینه‌سازی استفاده شد. در نهایت عملکرد این الگورتیم در قالب یک مثال شبیه‌سازی و تحلیل یک مجموعه داده واقعی با سه تابع تاوان لاسوی گروهی، SCAD گروهی و MCP گروهی مورد ارزیابی قرار گرفت. نتایج شبیه‌سازی عملکرد مناسب الگوریتم کاهش را در حل این مساله بهینه‌سازی تحت سه تابع تاوان مذکور نشان می‌دهد. در خصوص مقایسه مدل‌های برازش داده شده با این سه تابع تاوان گروهی، اگرچه اختلاف خطای این مدل‌ها ناچیز است، اما تابع تاوان لاسوی گروهی متغیرهای توضیحی بیشتری را وارد مدل می‌کند. این در حالی است که تاوان‌های SCAD گروهی و MCP گروهی منتج به مدل‌های کوچک‌تری می‌شوند. لذا می‌توان نتیجه گرفت که مدل جمعی ناپارامتری برازش‌شده با تابع تاوان MCP گروهی و حل مساله بهینه‌سازی به کمک الگوریتم کاهش گروهی، با انتخاب متغیرهای توضیحی کمتر، مدل مقرون به صرفه‌تری را نتیجه می‌دهد.

همان‌طور که در مثال شبیه‌سازی دیدیم، با افزایش تعداد متغیرهای توضیحی p نسبت به اندازه نمونه n، عملکرد روش‌های پیشنهادی تحت تاثیر قرار می‌گیرد. هنگامی‌که p نسبت به n بسیار بزرگ باشد، توصیه می‌شود

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

که ابتدا با استفاده از یک روش غربالگری مناسب بعد مساله را کاهش داده و سپس از رگرسیون تاوانیده برای برآورد و انتخاب متغیر استفاده شود. برای اطلاعات بیشتر در مورد انواع روش‌های غربالگری به فن [۳۲-۲۶] مراجعه شود.

تقدیر و تشکر

نویسنده مقاله از داوران محترم که نظرات ارزشمند ایشان باعث بهبود مطالب ارایه‌شده در این مقاله گردید، کمال تشکر و قدردانی را دارد. این پژوهش در قالب طرح پژوهشی به شماره ۸۲۰۵۴-۱۵ پ و با استفاده از اعتبارات پژوهش و فناوری دانشگاه گیلان انجام شده است.

منابع

[1] Hastie, T., Tibshirani, R. (1990). Generalized Additive Models, Chapman and Hall, NewYork.
[2] James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning, New York: springer.
[3] Frank, L. E., Friedman, J. H. (1993). A statistical view of some chemometrics regression tools, Technometrics, 35(2), 109-135.
[4] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.
[5] Zou, H., Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B, 67(2), 301-320.
[6] Fan, J., Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of American Statistical Association, 96, 1348-1360.
[7] Zou, H. (2006). The adaptive lasso and its oracle properties, Journal of the American Statistical Association, 101, 1418–1429.
[8] Zhang, C. H. (2010). Nearly unbiased variable selection under the minimax concave penalty. The Annals of Statistics, 38, 894–942.
[9] Zou, H., Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. Annals of Statistics**, 36**, 1509–1533.
[10] Fu, W. (1998). Penalized regressions: the bridge versus the lasso. Journal of Computational and Graphical Statistics, 7, 397-416.
[11] Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society Series B, 68, 49-67.
[12] Friedman, J., Hastie, T., Hofling, H., Tibshirani, R. (2007). Pathwise coordinate optimization. Annals of Applied Statistics, 1, 302-332.
[13] Wu, T., Lange, K., (2008). Coordinate descent algorithms for lasso penalized regression. Annals of Applied Statistics, 2 224-244.
[14] Friedman, J., Hastie, T., Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 33 1-22.
[15] Breheny, P., Huang, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. Annals of Applied Statistics, 5, 232-253.
[16] Breheny, P., Huang, J. (2015). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. Statistics and Computing, 25(2), 173-187.
[17] She, Y. (2012). An iterative algorithm for fitting nonconvex penalized generalized linear models with grouped predictors. Computational Statistics & Data Analysis, 56(10) 2976-2990.
[18] De Boor, C. (2001). A Practical Guide to Splines. Revised edition. Applied Mathematical Sciences. New York: Spring-Verlag.
[19] Ruppert, D., Wand, M. P., Carroll, R. J. (2003). Semiparametric regression, Cambridge University Press, Cambridge.
[20] Bakin, S. (1999). Adaptive regression and model selection in data mining problems. PhD Thesis. Australian National University, Canberra.

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]

[21] Wang, L., Chen, G. and Li, H. (2007). Group SCAD regression analysis for microarray time course gene expression data. Bioinformatics, 23(12), 1494-1486.
[22] Wang, L., Li, H., Huang, J. Z. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. Journal of the American Statistical Association, 103, 1556-1569.
[23] Huang, J., Breheny, P., Ma, S. (2012). A selective review of group selection in high-dimensional models. Statistical Science, 27 481-499.
[24] Bai, R., Moran, G. E., Antonelli, J. L., Chen, Y., Boland, M. R. (2022). Spike-and-slab group lassos for grouped regression and sparse generalized additive models. Journal of the American Statistical Association, 117(537) 184-197.
[25] Bühlmann, P., Kalisch, M., Meier, L. (2014). High-dimensional statistics with a view towards applications in biology. Annual Review of Statistics and its Applications, 1, 255-278.
[26] Fan, J., Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. Journal of Machine Learning Research, 70 (5), 849-911.
[27] Li, R.Z., Zhong, W., Zhu, L.P. (2012). Feature screening via distance correlation learning. Journal of the American Statistical Association, 107, 1129-1139.
[28] Li, R., Zhong, W., Zhu, L. (2012). Feature screening via distance correlation learning. Journal of the American Statistical Association, 107(499), 1129-1139.
[29] Kazemi, M., Shahsavani, D., Arashi, M. (2018). A sure independence screening procedure for ultra-high dimensional partially linear additive models. Journal of Applied Statistics, 1-19.
[30] Kazemi, M. (2020). Partial correlation screening for varying coefficient models. Journal of Mathematical Modeling, 8(4), 363-376.
[31] Kazemi, M. (2024). A Robust variable screening approach with application to gene expression data, REVSTAT-Statistical Journal. Retrieved from https://revstat.ine.pt/index.php/REVSTAT/article/view/613
[32] Kazemi, M. (2024). Support vector machine in ultrahigh-dimensional feature space. Journal of Statistical Computation and Simulation, 94(3), 517-535.

[Downloaded from jorlu.ir on 2026-11-30]

[DOI: 10.71773/ijorlu-2024-1-2163]