

مجله علوم آماری، بهار و تابستان ۱۳۸۸

جلد ۳، شماره ۱، ص ۱۷-۳۰

برآورد پارامترهای مدل رگرسیون لوژستیک ساخته شده با شبکه‌های عصبی واحد ضربی تکاملی

مریم ترک‌زاده ماهانی، سروش علیمرادی

دانشکده ریاضی، دانشگاه صنعتی اصفهان

تاریخ دریافت: ۱۳۸۸/۲/۱۶ تاریخ آخرین بازنگری: ۱۳۸۸/۷/۲۰

چکیده: یکی از ابزارهایی که برای تعیین اثرات غیرخطی و اثرات متقابل بین متغیرهای تبیینی در یک مدل رگرسیون لوژستیک به کار می‌رود، استفاده از شبکه‌های عصبی واحد ضربی تکاملی است. به منظور برآورد پارامترهای مدلی که بدین صورت به دست می‌آید، یک روش ترکیبی مورد استفاده قرار می‌گیرد؛ این روش از ترکیب دو ابزار بهینه‌ساز کلاسیک و الگوریتم تکاملی ساخته می‌شود. در این مقاله ساختار شبکه‌های عصبی به گونه‌ای تغییر داده می‌شود که تمام پارامترهای مدل با یک الگوریتم تکاملی قابل برآورد باشند. سپس دو روش برآورد مورد مقایسه قرار گرفته و نتایج نشان می‌دهد که برآورد پارامترها با الگوریتم‌های تکاملی منجر به مدلی می‌شود که از نظر معیار اطلاع آکاییک نسبت به مدل لوژستیک معمولی دقیق‌تر است، اما استفاده از روش ترکیبی، مدل بهتری را نتیجه می‌دهد.

واژه‌های کلیدی: رگرسیون لوژستیک، شبکه‌های عصبی، الگوریتم‌های تکاملی.

آدرس الکترونیک مسئول مقاله: مریم ترک‌زاده ماهانی، mtmiut@gmail.com
کد موضوع‌بندی ریاضی (۲۰۰۰): ۶۲J۱۲

فرض خطی بودن اثرات متغیرهای تبیینی، یکی از فرض‌هایی است که در ساختن مدل‌های رگرسیونی از جمله مدل لوژستیک در نظر گرفته می‌شود. با این حال ممکن است در عمل با مواردی مواجه شویم که بین متغیرهای تبیینی اثر متقابل وجود داشته باشد یا متغیرهای تبیینی دارای اثرات غیرخطی باشند. در این موارد می‌توان توابعی غیرخطی از متغیرهای تبیینی را در مدل وارد نمود که توابع پایه نامیده می‌شوند. تا کنون روش‌های مختلفی معرفی شده‌اند که از توابع پایه متفاوتی استفاده کرده‌اند. به عنوان مثال استفاده از مدل‌های خطی تعمیم یافته یکی از روش‌هایی است که برای تعیین اثرات غیرخطی متغیرهای تبیینی به کار می‌رود. این روش توسط هستی و تیشیرانی (۱۹۹۰) معرفی شده است. برای وارد کردن اثرات متقابل بین متغیرهای تبیینی در مدل، کوپربرگ و همکاران (۱۹۹۷) روش رگرسیون اسپلاین سازوار چند متغیره^۱ (MARS) چندگانه را معرفی کرده‌اند که حالت ترکیبی روش MARS معرفی شده توسط فریدمن (۱۹۹۱) است. مهم‌ترین مشکل این روش‌ها نامعلوم بودن تعداد بهینه توابع پایه است. هرواس-مارتینز و همکاران (۲۰۰۸) روشی را معرفی نمودند که علاوه بر وارد کردن اثرات غیرخطی و اثرات متقابل در مدل، تعداد بهینه توابع پایه را نیز مشخص می‌کند. در این روش توابع پایه به صورت حاصل ضرب متغیرهای تبیینی که هر کدام به یک توان حقیقی دلخواه رسیده‌اند، تعریف می‌شوند. به منظور تعیین تعداد بهینه توابع پایه و برآورد توان‌های مربوط به متغیرهای تبیینی در هر تابع پایه، از شبکه‌های عصبی^۲ و الگوریتم‌های تکاملی^۳ استفاده می‌شود. در واقع شبکه‌های عصبی پیشنهاد می‌کنند که چه اثراتی در مدل وارد شوند، سپس با استفاده از روش گام به گام^۴ اثراتی که حضورشان در مدل معنی دار نیست، از مدل حذف می‌شوند. هرواس-مارتینز و همکاران (۲۰۰۸) برای برآورد پارامترهای مدل از یک روش ترکیبی استفاده کرده‌اند، که در مرحله اول

^۱ Multivariate Adaptive Regression Spline

^۲ Neural network

^۳ Evolutionary algorithm

^۴ Stepwise method

م. ترکزاده، س. علیمزادی: برآورد پارامترهای مدل رگرسیون لوژستیک ۱۹

آن تعداد بهینه توابع پایه و توان‌های مربوط به متغیرهای تبیینی با روش بهینه‌ساز سراسری الگوریتم تکاملی تعیین می‌شوند و در مرحله دوم ضرایب مدل با روش بهینه‌ساز کلاسیک نیوتن-رافسون به دست می‌آیند. شبکه عصبی که در روش برآورد ترکیبی مورد استفاده قرار می‌گیرد، یک شبکه سه لایه‌ای است که هیچ اتصالی بین لایه اول و لایه سوم آن وجود ندارد و تنها بر قسمت غیرخطی مدل رگرسیون لوژستیک منطبق می‌شود. در این مقاله با افزودن یک سری اتصالات بین لایه اول و لایه سوم، شبکه عصبی به گونه‌ای طراحی می‌شود که علاوه بر قسمت غیرخطی، قسمت خطی مدل لوژستیک را نیز شامل شود و تمامی پارامترهای مدل با استفاده از یک الگوریتم تکاملی قابل برآورد باشند. سپس کارایی روش جدید، بر اساس معیار اطلاع آکاییک، با روش ترکیبی هرواس-مارتینز و همکاران (۲۰۰۸) مورد مقایسه قرار می‌گیرد. در این راستا و به منظور انجام مقایسه‌های تجربی از یک مجموعه داده واقعی مربوط به یک مجتمع تولیدی بزرگ فولاد استفاده می‌شود.

۲ رگرسیون لوژستیک با شبکه‌های عصبی واحد ضربی تکاملی

فرض کنید $X = (x_1, \dots, x_k)$ بردار حاصل از k متغیر تبیینی و Y متغیر پاسخ دودویی باشد که دو سطح آن با مقادیر صفر و یک کدگذاری شده‌اند. مدل لوژستیک به صورت

$$\text{logit} [\pi(X)] = \ln \left[\frac{\pi(X)}{1 - \pi(X)} \right] = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k, \quad (1)$$

تعریف می‌شود، که در آن $\pi(X) = P(Y = 1|X)$ است. در صورتی که متغیر پاسخ بیشتر از دو سطح داشته باشد، می‌توان از رگرسیون لوژستیک چندگانه استفاده نمود. با فرض این که متغیر پاسخ Y ، J سطح داشته باشد، آن‌گاه بردار احتمالات $(\pi_1(X), \dots, \pi_J(X))$ که احتمال قرار گرفتن یک فرد با بردار متغیرهای توضیحی X در J سطح مختلف متغیر پاسخ نشان می‌دهد را می‌توان به صورت

$$\ln \left[\frac{\pi_l(X)}{\pi_J(X)} \right] = \alpha_0^l + \alpha_1^l x_1 + \dots + \alpha_k^l x_k, \quad l = 1, \dots, J - 1, \quad (2)$$

۲۰..... مجله علوم آماری، بهار و تابستان ۱۳۸۸، جلد ۳، شماره ۱، ص ۱۷-۳۰

مدل سازی کرد. جزئیات بیشتر در آگرستی (۲۰۰۷) مطرح شده است. در روابط (۱) و (۲) تنها اثرات خطی متغیرهای تبیینی در مدل وارد شده‌اند. به منظور وارد کردن اثرات غیرخطی و اثرات متقابل، m تابع غیرخطی از متغیرهای تبیینی به مدل اضافه می‌شود. در اینجا m مجهول است و توابع غیرخطی به صورت حاصل ضرب متغیرهای تبیینی که هر کدام به یک توان حقیقی دلخواه رسیده‌اند، تعریف می‌شوند. به عنوان مثال $x_1^{w_1} \dots x_k^{w_k}$ یک تابع غیرخطی از متغیرهای تبیینی است. با وارد کردن m تابع غیرخطی به رابطه (۲) مدل

$$\ln \left[\frac{\pi_\ell(X)}{\pi_J(X)} \right] = \alpha_\ell + \sum_{i=1}^k \alpha_i^\ell x_i + \sum_{j=1}^m \beta_j^\ell \left(\prod_{i=1}^k x_i^{w_{ji}} \right), \quad \ell = 1, \dots, J-1, \quad (3)$$

به دست می‌آید. معمولاً پارامترهای مدل (۳) با روش ماکسیمم درست‌نمایی برآورد می‌شوند. همچنین در مسیر برآوردیابی از روش‌های بهینه‌ساز کلاسیک مانند نیوتن-رافسون استفاده می‌شود. با فرض J -سطحی بودن متغیر پاسخ و داشتن مجموعه مشاهدات $\{(y_i, X_i) : i = 1, \dots, n\}$ منفی لگاریتم تابع درست‌نمایی به صورت

$$\ell^* = - \sum_{i=1}^n \ln [P(Y_i = y_i | X_i)] = \sum_{i=1}^n \left\{ - \sum_{\ell=1}^J y_i^\ell f_\ell(X_i) + \ln \left[\sum_{\ell=1}^J f_\ell(X_i) \right] \right\} \quad (4)$$

تعریف می‌شود، که در آن

$$f_\ell(X) = \alpha_\ell + \sum_{i=1}^k \alpha_i^\ell x_i + \sum_{j=1}^m \beta_j^\ell \left(\prod_{i=1}^k x_i^{w_{ji}} \right), \quad \ell = 1, \dots, J-1, \quad (5)$$

و $f_J(X) = 0$. در مجموعه مشاهدات $\{(y_i, X_i) : i = 1, \dots, n\}$ بردار X_i بردار k -بعدی متغیرهای تبیینی مربوط به i -امین مشاهده است و بردار J -بعدی $y_i = (y_i^1, \dots, y_i^J)$ سطح مربوط به متغیر پاسخ i -امین مشاهده را مشخص می‌کند، به این صورت که اگر سطح متغیر پاسخ برابر با r باشد ($r = 1, \dots, J$)، r -امین مؤلفه این بردار برابر با یک و بقیه مؤلفه‌ها برابر با صفر در نظر گرفته می‌شوند.

تابع ℓ^* در رابطه (۴) به دلیل قسمت غیرخطی $\left(\prod_{i=1}^k x_i^{w_{ji}} \right)$ که به مدل اضافه شده است، دارای مینیمم‌های موضعی زیادی است و ممکن است با

م. ترکزاده، س. علیمردی: برآورد پارامترهای مدل رگرسیون لوژستیک ۲۱

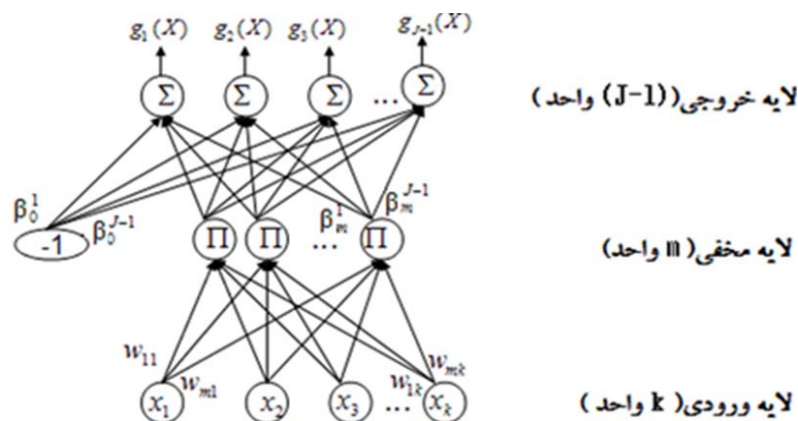
روش‌های بهینه‌ساز کلاسیک، مینیمم سراسری به دست نیاید. بنابراین به منظور برآورد پارامترها از یک روش ترکیبی استفاده می‌شود. این روش ترکیبی، از یک الگوریتم تکاملی و یک روش بهینه‌ساز کلاسیک تشکیل شده است، که در دو گام خلاصه می‌شود. در گام اول با استفاده از یک الگوریتم تکاملی تعداد توابع پایه‌ای که در مدل وارد می‌شوند و توان‌های مربوط به هر تابع پایه مشخص می‌شوند. در گام دوم توابع پایه حاصل از گام اول در رابطه (۳) قرار می‌گیرند. چون مدل حاصل نسبت به پارامترها خطی است، می‌توان با یک روش بهینه‌ساز کلاسیک بقیه پارامترهای مدل را برآورد کرد. (برای جزئیات بیشتر به هرواس-مارتینز و همکاران (۲۰۰۸) مراجعه شود).

۳ برآورد پارامترهای مدل

شبکه عصبی مصنوعی^۵ (ANN) یک روش عمومی و کاربردی است که با استفاده از یک مجموعه داده، توابع برداری پیوسته و گسسته را تقریب می‌زند. ANN از اتصال یک سری واحد ساده ایجاد می‌شود. به این واحدها گره نیز می‌گویند. هر واحد یک بردار ورودی با مقادیر حقیقی دریافت می‌کند که می‌تواند خروجی واحدهای قبل باشد و یک خروجی با مقدار حقیقی تولید می‌نماید که می‌تواند به عنوان ورودی واحدهای دیگر مورد استفاده قرار گیرد. (برای آشنایی بیشتر به میشل (۱۹۹۷) مراجعه شود).

شبکه عصبی که در روش ترکیبی مورد استفاده قرار می‌گیرد، تنها قسمت غیرخطی $f_l(X)$ در رابطه (۵) را شامل می‌شود. شکل ۱ این شبکه را نمایش می‌دهد. برای اینکه تمام پارامترهای مدل با استفاده از یک الگوریتم تکاملی قابل برآورد باشند، کافی است تا شبکه عصبی را به فرم شکل ۲ تغییر داد تا کل مدل در ساختار یک شبکه عصبی ظاهر شود. این شبکه از سه لایه ورودی، مخفی و خروجی تشکیل شده است. در هر لایه تعدادی واحد وجود دارد. به واحدهای لایه مخفی با نماد Π ، واحدهای ضربی و به واحدهای لایه خروجی با نماد Σ ، واحدهای

^۵ Artificial Neural Network

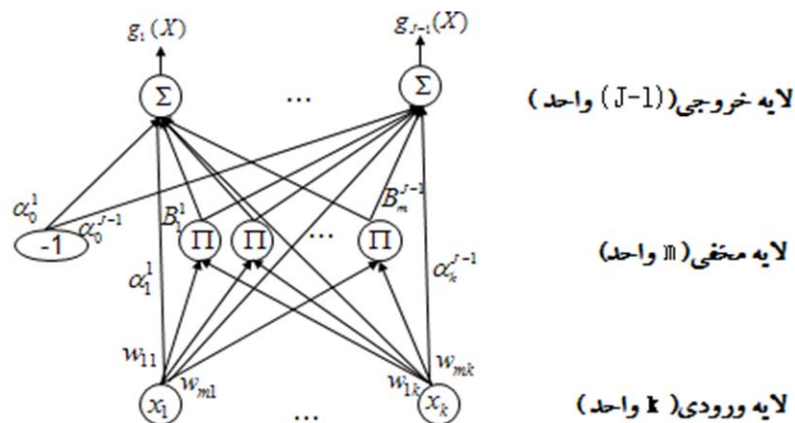


شکل ۱: شبکه عصبی متناظر با قسمت غیر خطی مدل (۴)

جمع می گویند. هر واحد با دریافت سیگنال های ورودی و وزن های مربوط به آنها، سیگنال خروجی را محاسبه می کند.

به عنوان مثال j -امین واحد در لایه مخفی، سیگنال های ورودی x_1, \dots, x_k را به ترتیب با وزن های w_{j1}, \dots, w_{jk} دریافت می کند و سیگنال خروجی $\prod_{i=1}^k x_i^{w_{ji}}$ را تولید می نماید. این سیگنال خروجی یک سیگنال ورودی برای واحدهای لایه خروجی محسوب می شود. l -امین واحد از لایه خروجی، m سیگنال $\prod_{i=1}^k x_i^{w_{ji}}$ ($j = 1, \dots, m$) را به ترتیب با وزن های $\beta_1^l, \dots, \beta_m^l$ و $(k+1)$ سیگنال $1, x_1, \dots, x_k$ را به ترتیب با وزن های $\alpha_0^l, \alpha_1^l, \dots, \alpha_k^l$ دریافت می کند و سیگنال خروجی را به صورت $g_l(X) = \alpha_0^l + \sum_{i=1}^k \alpha_i^l x_i + \sum_{j=1}^m \beta_j^l (\prod_{i=1}^k x_i^{w_{ji}})$ محاسبه می نماید. ملاحظه می شود که در محاسبه سیگنال خروجی، با توجه به جمع و ضربی بودن هر واحد، به ترتیب از حاصل جمع وزنی و حاصل ضرب توانی سیگنال های ورودی استفاده می شود. در این شبکه که یک شبکه عصبی واحد ضربی تکاملی است، تعداد واحدهای لایه مخفی و وزن های شبکه نامعلوم هستند. این پارامترها به گونه ای برآورد می شوند یا به طور معادل شبکه به گونه ای آموزش می بیند

م. ترکزاده، س. علیمردی: برآورد پارامترهای مدل رگرسیون لوژستیک ۲۳



شکل ۲: ظاهر شدن کل مدل در ساختار یک شبکه عصبی

که منفی لگاریتم تابع درستنمایی متناظر با شبکه عصبی در رابطه (۶) مینیمم شود.

$$L^* = \sum_{i=1}^n \left\{ - \sum_{\ell=1}^{J-1} y_i^{(\ell)} g_{\ell}(X_i) + \ln \left(1 + \sum_{\ell=1}^{J-1} \exp \{g_{\ell}(X_i)\} \right) \right\} \quad (6)$$

برای حل یک مسأله‌ی بهینه‌سازی، لازم است تا جواب بهینه در بین تمامی جواب‌های امکان‌پذیر جستجو شود. در برخی مسائل، فضای جستجو (فضای حاصل از تمامی جواب‌های امکان‌پذیر)، فضای بزرگی است و بررسی تمام جواب‌های امکان‌پذیر و یافتن جواب بهینه، نیاز به صرف زمان بسیار دارد.

در مسأله اکسترم‌یابی یک تابع، ممکن است به دلیل زیاد بودن اکسترم‌های موضعی، رسیدن به اکسترم فراموضعی^۶ با به کار گرفتن روش‌های کلاسیک امکان‌پذیر نباشد. به طور کلی در مواردی که فضای جستجو یک فضای بزرگ و پیچیده است، استفاده از الگوریتم‌های تکاملی به عنوان ابزارهای بهینه‌سازی تصادفی توصیه می‌شود.

یک الگوریتم تکاملی با مجموعه‌ای از جواب‌های تصادفی آغاز می‌شود. این جواب‌ها جمعیت اولیه را می‌سازند. با استفاده از جمعیت اولیه و با به کار گرفتن عملگرهای تکاملی، جمعیت بعد ساخته می‌شود، با این ایده که جمعیتی بهتر از

^۶ Global extremum

۲۴ مجله علوم آماری، بهار و تابستان ۱۳۸۸، جلد ۳، شماره ۱، ص ۱۷-۳۰

جمعیت فعلی به دست آید. این فرآیند تولید آن قدر تکرار می شود تا شرایط خاتمه محقق گردد. در نهایت، مناسب ترین فرد در آخرین جمعیت تولید شده به عنوان جواب مسأله معرفی می شود. (برای جزئیات بیشتر به انگلیسچ (۲۰۰۲) مراجعه شود).

به دلیل قسمت غیرخطی $\sum_{j=1}^m \beta_j^l (\prod_{i=1}^k x_i^{w_{ji}})$ که در g_l ظاهر شده است، برای مینیم کردن رابطه (۶) از یک الگوریتم تکاملی استفاده می شود. در این مسأله بهینه سازی، شبکه ای جستجو می شود که دارای تعداد بهینه واحد در لایه مخفی و بهترین وزن ها باشد. بنابراین هر جواب از مسأله را می توان به صورت یک شبکه در نظر گرفت. الگوریتم تکاملی در ابتدا با مجموعه ای از این شبکه ها آغاز می شود. به این مجموعه، جمعیت اولیه گویند و هر شبکه یک فرد نامیده می شود. برای تولید یک شبکه کافی است تا مقادیر پارامترهای شبکه $(m, W = (w_{11}, \dots, w_{mk}), \beta = (\beta_1^1, \dots, \beta_m^{J-1}), \alpha = (\alpha_0^1, \dots, \alpha_k^{J-1}))$ به طور تصادفی از فضای پارامتری انتخاب شوند. در مرحله بعد مقدار برازندگی هر فرد از جمعیت اولیه با توجه به رابطه

$$A = \frac{1}{1 + L^*} \quad (7)$$

محاسبه می شود. هدف الگوریتم تولید افراد با بالاترین برازندگی است. بنابراین با استفاده از جمعیت اولیه و عملگرهای تکاملی انتخاب و جهش، جمعیت بعد تولید می شود، با این ایده که جمعیتی بهتر از جمعیت فعلی به دست آید. (این عملگرها در ادامه این بخش مطرح می شوند.) این روند تولید آنقدر ادامه می یابد تا یکی از شرط های زیر حاصل شود.

- تعداد جمعیت های تولید شده به یک عدد معین مانند ۲۰۰ برسد.
- واریانس مربوط به مقدار برازندگی بهترین ۱۰٪ افراد جمعیت کمتر از یک عدد معین مانند 10^{-4} شود.

پس از توقف الگوریتم بهترین فرد در آخرین جمعیت تولید شده به عنوان جواب مسأله در نظر گرفته می شود و به این صورت برآوردی برای پارامترهای شبکه به دست می آید. با برآورد شدن پارامترها، m تابع پایه

م. ترکزاده، س. علیمادی: برآورد پارامترهای مدل رگرسیون لوژیستیک ۲۵.....

گام به گام و با توجه به متغیرهای $x_1, \dots, x_k, z_1, \dots, z_m$ مدل لوژیستیک نهایی به دست می آید.

در الگوریتم تکاملی مورد نظر، عملگر انتخاب به این صورت است که پس از محاسبه برانزندی افراد، بهترین ۱۰٪ از جمعیت انتخاب می شوند و جایگزین بدترین ۱۰٪ جمعیت می گردند و عموماً جهش نیز به دو صورت پارامتری و ساختاری به شرح زیر انجام می شود.

الف- جهش پارامتری: جهش پارامتری تغییری در مقدار وزن های شبکه شکل ۲ به صورت

$$\begin{aligned} w_{ji}(t+1) &= w_{ji}(t) + \varepsilon_1(t), \quad j = 1, \dots, m, \quad i = 1, \dots, k, \\ \beta_j^\ell(t+1) &= \beta_j^\ell(t) + \varepsilon_2(t), \quad j = 1, \dots, m, \quad \ell = 1, \dots, (J-1), \\ \alpha_j^\ell(t+1) &= \alpha_j^\ell(t) + \varepsilon_2(t), \quad j = 0, \dots, k, \quad \ell = 1, \dots, (J-1), \end{aligned}$$

ایجاد می کند، که در آن

$$\varepsilon_r(t) \sim N(0, \gamma_r(t)), \quad r = 1, 2$$

باید دقت کرد که مقدار واریانس $\gamma_1(t)$ بسیار کمتر از مقدار واریانس $\gamma_2(t)$ است، زیرا این واریانس مقادیر مربوط به توان ها را تغییر می دهد. بنابراین مقدار آن باید کوچک باشد تا منجر به یک تغییر بزرگ نشود. پس از آنکه جهش صورت گرفت، مقدار برانزندی فرد جهش یافته محاسبه می شود. ΔA اختلاف مقدار برانزندی، قبل و بعد از جهش را نشان می دهد. اگر $\Delta A \geq 0$ ، این گام تصادفی با احتمال یک و در غیر این صورت با احتمال زیر پذیرفته می شود.

$$p = \exp\left(\frac{\Delta A}{T(g)}\right) \quad T(g) = 1 - A(g) \quad 0 < T(g) \leq 1, \quad (8)$$

واریانس $\gamma_r(t)$ در طول تکامل بهنگام می شود. روش های مختلفی به منظور بهنگام سازی وجود دارد. در این الگوریتم از قانون $\frac{1}{8}$ موفقیت ریچنبرگ استفاده

۲۶ مجله علوم آماری، بهار و تابستان ۱۳۸۸، جلد ۳، شماره ۱، ص ۱۷-۳۰

می شود (ریچنبرگ، ۱۹۷۳). طبق این قانون، نسبت جهش های موفقیت آمیز باید برابر با $\frac{1}{\delta}$ باشد. بنابراین اگر این نسبت بیشتر از $\frac{1}{\delta}$ شد، باید واریانس جهش افزایش یابد و در غیر این صورت واریانس جهش کم می شود. بنابراین

$$\gamma_r(t+s) = \begin{cases} (1+\lambda)\gamma_r(t) & s_s > \frac{1}{\delta} \\ (1-\lambda)\gamma_r(t) & s_s < \frac{1}{\delta} \\ \gamma_r(t) & s_s = \frac{1}{\delta} \end{cases}$$

که در آن $\lambda = 0/1$ ، $r = 1, 2$ و s_s نسبت جهش های موفقیت آمیز در s مرحله از تولید است. این بهنگام سازی باعث می شود که الگوریتم در یک مینیمم محلی قرار نگیرد و همچنین به فرایند تکامل سرعت می بخشد.

ب- جهش ساختاری: جهش ساختاری شبکه یک تغییر در ساختار شبکه ایجاد می کند و امکان جستجو در نواحی جدید را فراهم می نماید. همچنین به متنوع بودن جمعیت کمک می کند. این جهش را می توان به پنج صورت زیر انجام داد:

۱- **افزودن گره:** یک یا چند گره، با وزن های تصادفی به لایه مخفی اضافه می شوند.
 ۲- **افزودن اتصال:** این جهش ابتدا در لایه مخفی و پس از آن در لایه خروجی صورت می گیرد. برای افزودن یک اتصال از لایه ورودی به لایه مخفی، یک گره به طور تصادفی از هر یک از این لایه ها انتخاب می شود، سپس این دو گره با یک وزن تصادفی به یکدیگر متصل می شوند. یک فرایند مشابه از لایه مخفی به لایه خروجی انجام می شود.

۳- **حذف گره:** یک یا چند گره از لایه مخفی به طور تصادفی انتخاب می شوند و همراه با تمام اتصالاتشان از شبکه حذف می گردند.

۴- **حذف اتصال:** مشابه افزودن اتصال، این جهش ابتدا در لایه مخفی و سپس در لایه خروجی انجام می شود. بدین صورت که به طور تصادفی دو گره به ترتیب از لایه تحت جهش و لایه قبل آن انتخاب می شوند، سپس اتصال بین این دو گره حذف می گردد.

۵- **ترکیب گره ها:** دو گره a و b به طور تصادفی از لایه مخفی انتخاب می شوند و

م. ترکزاده، س. علیمرادی: برآورد پارامترهای مدل رگرسیون لوژستیک ۲۷

با گره c جایگزین می گردند. این گره ترکیبی از دو گره a و b است. وزنهای مربوط به گره c ، به صورت $\beta_c^l = \beta_a^l + \beta_b^l$ و $w_{cj} = (w_{aj} + w_{bj})/2$ تعریف می شوند. برای تعیین تعداد گره‌های تحت جهش، یعنی تعداد گره‌های اضافه شده به لایه مخفی در جهش ساختاری افزودن گره، تعداد گره‌های حذف شده از لایه مخفی در جهش ساختاری حذف گره و تعداد زوج گره‌های ترکیب شده در جهش ساختاری ترکیب گره از عبارت $\Delta_{MIN} + uT(g)[\Delta_{MAX} - \Delta_{MIN}]$ استفاده می شود، که در آن u عددی تصادفی از توزیع یکنواخت $(0, 1)$ است. Δ_{MAX} و Δ_{MIN} به ترتیب کمترین و بیشترین تعداد گره‌های تحت جهش را مشخص می نمایند و مقادیر متناظر با آنها توسط کاربر الگوریتم مشخص می شود. عبارت $1 + u[\Delta_{ONO}]$ تعداد اتصالات تحت جهش، بین لایه مخفی و لایه خروجی را تعیین می کند و از عبارت $1 + u[\Delta_{HNH}]$ برای تعیین تعداد اتصالات تحت جهش بین لایه مخفی و لایه ورودی استفاده می شود، که در آن‌ها u عددی تصادفی از توزیع یکنواخت $(0, 1)$ ، Δ_H و Δ_O به ترتیب نسبت تعداد اتصالات در لایه خروجی و لایه مخفی را مشخص می کنند و توسط کاربر الگوریتم تعیین می شوند. همچنین n_H و n_O به ترتیب تعداد اتصالات در لایه خروجی و لایه مخفی شبکه موجود را نشان می دهند. در اینجا تعداد اتصالات تحت جهش به تعداد اتصالات افزوده شده در جهش ساختاری افزودن اتصال و تعداد اتصالات حذف شده در جهش ساختاری حذف اتصال اشاره می کند. در هنگام انجام جهش همواره جهش‌های حذفی و ترکیبی قبل از جهش‌های افزایشی، مورد توجه قرار می گیرند و در واقع احتمال مربوط به جهش‌های حذفی و ترکیبی برابر با $T(g)$ در رابطه (۸) است که بیشتر از احتمال مربوط به جهش‌های افزایشی یعنی $T^2(g)$ در نظر گرفته می شود. بر روی هر شبکه تنها یک جهش صورت می گیرد و اگر تحت احتمالات $T(g)$ و $T^2(g)$ هیچ جهشی انتخاب نشد، یک جهش به طور تصادفی انتخاب می شود.

۴ بررسی تجربی

در این بخش از داده‌های مربوط به یک مجتمع تولیدی بزرگ فولاد برای ارزیابی روش ارائه شده و مقایسه آن با روش ترکیبی استفاده می شود. داده‌ها شامل ۵۰

۲۸ مجله علوم آماری، بهار و تابستان ۱۳۸۸، جلد ۳، شماره ۱، ص ۱۷-۳۰

نمونه از ورق‌های فولاد تولید شده هستند. برای هر نمونه سالم یا معیوب بودن آن تعیین می‌گردد. منظور از ورق معیوب ورقی است که مصرف کننده نتواند از آن استفاده کند. وزن (w) و خواص مکانیکی نقطه تسلیم^۷ (YP)، استحکام کشش نهایی^۸ (UTS)، ازدیاد طول^۹ (E)، میزان کربن (C)، سیلیسیم (Si)، منگنز (Mn)، آلومینیم (Al) و نیتروژن (N) به کار رفته در هر نمونه اندازه‌گیری می‌شود. برای ساختن مدل لوژستیک به روش گام‌به‌گام از فرمان proc logistic در نرم افزار SAS استفاده می‌شود. مدل حاصل به صورت

$$\ln\left(\frac{\pi}{1-\pi}\right) = -67/02 + 14/10w + 43/87YP + 8/20E + 26/08Si$$

است، که در آن π احتمال معیوب بودن نمونه را نشان می‌دهد. معیار اطلاع آکاییک^{۱۰} (AIC) برای این مدل برابر با ۴۲/۸۵ است. برای محاسبه معیار AIC با فرض این که n تعداد مشاهدات و p تعداد پارامترهای مجهول مدل را نشان دهد، از رابطه

$$AIC = -2 \log(L(\hat{\theta})) + 2p + \frac{2p(p+1)}{n-p+1} \quad (9)$$

استفاده می‌شود، که در آن $L(\hat{\theta})$ مقدار تابع درست‌نمایی در نقطه‌ی $\hat{\theta}$ است. طبق این معیار روشی مناسب‌تر است که AIC کمتری داشته باشد.

در ساختن مدل لوژستیک به وسیله شبکه‌های عصبی روند کار مشابه مدل لوژستیک معمولی است، با این تفاوت که علاوه بر متغیرهای تبیینی، یک سری توابع غیرخطی از آن‌ها نیز در مدل وارد می‌شوند. این توابع نشان‌دهنده‌ی اثرات غیرخطی متغیرهای تبیینی در مدل لوژستیک هستند. در اینجا تابع غیرخطی

$$z_1 = w^{0/5} E^{0/26} c^{-0/26} Si^{-0/03} Mn^{0/72} Al^{0/99}$$

^۷ Yield point

^۸ Ultimate tensile strength

^۹ Elongation

^{۱۰} Akaike's Information Criterion

م. ترکزاده، س. علیمردی: برآورد پارامترهای مدل رگرسیون لوژستیک ۲۹

با روش برآورد جدید به دست می آید. مدل لوژستیک ساخته شده به روش گام به گام، با توجه به متغیرهای تبیینی و تابع z_1 به صورت

$$\ln\left(\frac{\pi}{1-\pi}\right) = -32/85 + 30/24YP + 30/23Si - 42/28Mn + 23/36z_1$$

است. معیار AIC برای این مدل برابر با $40/67$ است که نسبت به مدل لوژستیک معمولی دو واحد کاهش در شاخص AIC ایجاد شده است. با به کار بردن روش ترکیبی دو تابع غیر خطی به صورت

$$z_1 = w^{-1/46} YP^{-0/17} UTS^{-0/12} E^{-0/16} c^{-0/85} Si^{-0/08} Mn^{-0/05} Al^{0/24} N^{-0/42}$$

$$z_2 = w^{-0/23} YP^{-0/28} UTS^{-0/29} E^{0/70} C^{-1/45} Si^{0/09} Mn^{1/12} Al^{0/22} N^{-1/12}$$

به مدل اضافه می شوند. مدل لوژستیک ساخته شده به روش گام به گام، با توجه به متغیرهای تبیینی و توابع z_1 و z_2 به صورت

$$\ln\left(\frac{\pi}{1-\pi}\right) = -9/74 + 22/25YP - 4/57z_1 - 5/66z_2$$

است. معیار AIC برای این مدل برابر با $37/70$ حاصل شده است. همان طور که ملاحظه می شود، با روش ترکیبی معیار AIC نسبت به مدل لوژستیک معمولی پنج واحد کاهش یافته است و در واقع نسبت به روش جدید، مدل بهتری را پیشنهاد می کند.

جدول ۱: مقادیر معیار AIC سه روش برآورد مدل لوژستیک داده های فولاد

روش معمولی	روش جدید	روش ترکیبی
42/85	40/67	37/70

بحث و نتیجه گیری

در این مقاله روش برآورد پارامترهای مدل لوژستیک ساخته شده با شبکه های عصبی تکاملی واحد ضربی ارائه شد. مقادیر معیار AIC مندرج در جدول ۱ بیانگر آن است که مدل حاصل از روش ارائه شده نسبت به مدل لوژستیک معمولی بهتر است، اما روش ترکیبی نسبت به روش برآورد جدید مدل بهتری را ارائه می کند.

۳۰ مجله علوم آماری، بهار و تابستان ۱۳۸۸، جلد ۳، شماره ۱، ص ۱۷-۳۰

تقدیر و تشکر

مؤلفین از داوران و هیئت تحریریه محترم مجله علوم آماری قدردانی و تشکر می کنند.

مراجع

- Agresti, A., (2007), An Introduction to Categorical Data Analysis, *John Wiley*, New York.
- Engelbrecht, A. P., (2002), Computational Intelligence, An Introduction, *John Wiley*, New York.
- Friedman, J. (1991), Multivariate Adaptive Regression Splines, *The Annals of Statistics*, **19**, 1-141.
- Hastie, T. J., and Tibshirani, R. J., (1990), *Generalized Additive Models*, Chapman and Hall, London.
- Hervas-Martinez, C., Martinez-Estudillo, F. J., and Carbonero-Ruz M. (2008), Multilogistic Regression by Means of Evolutionary Product-Unit Neural Networks, *Neural Networks*, **21**, 951-961.
- Kooperberg, C., Bose, S., and Stone, C. J. (1997), Polychotomous Regression, *Journal of the American Statistical Association*, **92**, 117-127.
- Mitchell, T. P., (1997), *Machine Learning*, McGraw-Hill, New York.
- Rechenberg, I., (1973), *Evolutionsstrategie: Optimierung Technischer Systeme Nach Prinzipien Der Biologischen Evolution*, Stuttgart Franman Holzboog Verlag.