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 به‌کار می رود، استفاده از شبکه‌های عصبی واحد ضربی تکاملی است. به‌منظور برآورد پارامترهای مدلی که بدین صورت به‌دست 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 به‌گونه‌ای 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

۱ مقدمه

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 به‌کار می‌رود. این روش توسط هستی و تیبشیرانی (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 حاصل‌ضرب متغیرهای تبیینی که هر کدام به یک توان حقیقی دلخواه رسیده‌اند، تعریف می‌شوند. به‌منظور 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 معنی‌دار 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

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 به‌دست 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 به‌گونه‌ای 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 به‌منظور 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 $X = (x_1, \cdots, x_k)$ 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 به‌صورت

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

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 این‌که متغیر پاسخ Y، J سطح داشته باشد، آن‌گاه بردار احتمالات $\pi(X) = (\pi_1(X), \cdots, \pi_J(X))$ که احتمال قرار گرفتن یک فرد با بردار متغیرهای توضیحی X در J سطح مختلف متغیر پاسخ نشان می‌دهد را می توان به‌صورت

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

مدل‌سازی 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 به‌صورت حاصل‌ضرب متغیرهای تبیینی که هر کدام به یک توان حقیقی دلخواه رسیده‌اند، تعریف می‌شوند. به عنوان مثال $x_1^{w_1}\cdots x_k^{w_k}$ یک تابع غیرخطی از متغیرهای تبیینی است. با وارد کردن m تابع غیرخطی به رابطه (۲) مدل

$$\ln\left[\frac{\pi_\ell(X)}{\pi_J(X)}\right] = \alpha_0^\ell + \sum_{i=1}^{k}\alpha_i^\ell x_i + \sum_{j=1}^{m}\beta_j^\ell\left(\prod_{i=1}^{k}x_i^{w_{ji}}\right), \quad \ell = 1,\cdots,J-1, \quad (۳)$$

به‌دست 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 $\{(y_i, X_i) : i = 1,\cdots,n\}$، منفی لگاریتم تابع درستنمایی به صورت

$$\ell^* = -\sum_{i=1}^{n}\ln\left[P(Y_i = y_i|X_i)\right] = \sum_{i=1}^{n}\{-\sum_{\ell=1}^{J}y_i^\ell f_\ell(X_i) + \ln[\sum_{\ell=1}^{J}f_\ell(X_i)]\} \quad (۴)$$

تعریف می‌شود، که در آن

$$f_\ell(X) = \alpha_0^\ell + \sum_{i=1}^{k}\alpha_i^\ell x_i + \sum_{j=1}^{m}\beta_j^\ell(\prod_{i=1}^{k}x_i^{w_{ji}}), \quad \ell = 1,\cdots,J-1, \quad (۵)$$

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

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

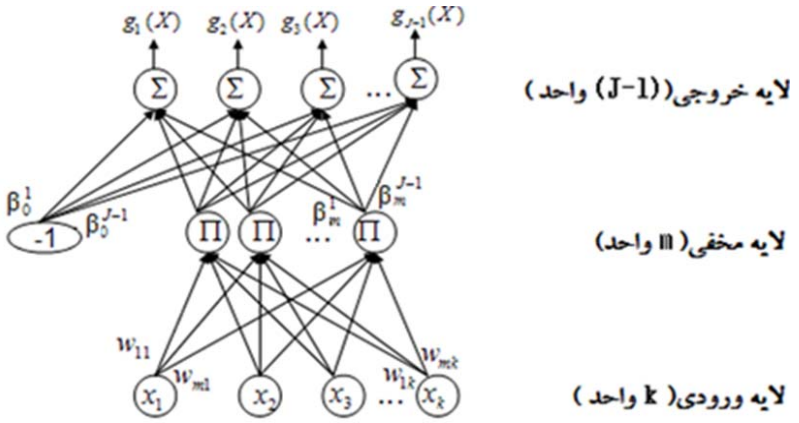
روش‌های بهینه‌ساز کلاسیک، مینیمم سراسری به‌دست نیاید. بنابراین به‌منظور برآورد پارامترها از یک روش ترکیبی استفاده می‌شود. این روش ترکیبی، از یک الگوریتم تکاملی و یک روش بهینه‌ساز کلاسیک تشکیل شده است، که در دو گام خلاصه می‌شود. در گام اول با استفاده از یک الگوریتم تکاملی تعداد توابع پایه‌ای که در مدل 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 به‌عنوان 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 $f_\ell(X)$ 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، کافی‌است 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

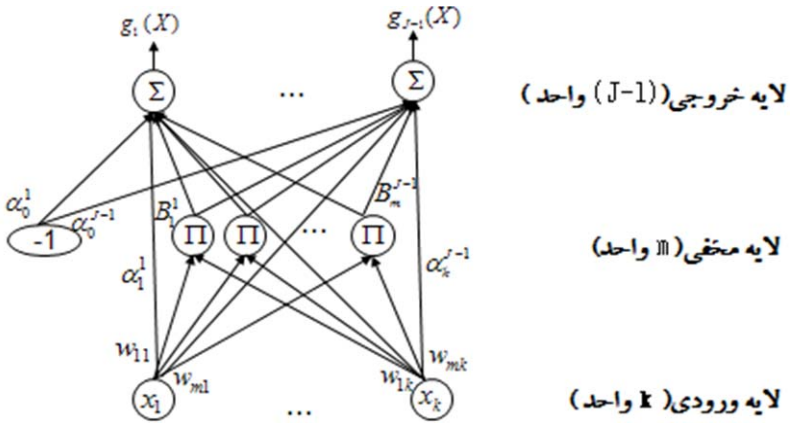
۵ Artifical Neural Network



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

جمعی گویند. هر واحد با دریافت سیگنال‌های ورودی و وزن‌های مربوط به آن‌ها، سیگنال خروجی را محاسبه می‌کند.

به عنوان مثال j-امین واحد در لایه مخفی، سیگنال‌های ورودی $x_1, \cdots, x_k$ را به‌ترتیب با وزن‌های $w_{j1}, \cdots, w_{jk}$ دریافت می‌کند و سیگنال خروجی $\prod_{i=1}^{k} x_i^{w_{ji}}$ را تولید می‌نماید. این سیگنال خروجی یک سیگنال ورودی برای واحدهای لایه خروجی محسوب می‌شود. ℓ-امین واحد از لایه خروجی، m سیگنال $\prod_{i=1}^{k} x_i^{w_{ji}}$ $(j = 1, \cdots, m)$ را به ترتیب با وزن‌های $\beta_1^\ell, \cdots, \beta_m^\ell$ و $(k+1)$ سیگنال $1, x_1, \cdots, x_k$ را به ترتیب با وزن‌های $\alpha_0^\ell, \alpha_1^\ell \cdots, \alpha_k^\ell$ دریافت می‌کند و سیگنال خروجی را به صورت $g_\ell(X) = \alpha_0^\ell + \sum_{i=1}^{k} \alpha_i^\ell x_i + \sum_{j=1}^{m} \beta_j^\ell (\prod_{i=1}^{k} x_i^{w_{ji}})$ محاسبه می‌نماید. ملاحظه می‌شود که در محاسبه سیگنال خروجی، با توجه به جمعی و ضربی بودن هر واحد، به ترتیب از حاصل‌جمع وزنی و حاصل‌ضرب توانی سیگنال‌های ورودی استفاده می‌شود. در این شبکه که یک شبکه عصبی واحد ضربی تکاملی است، تعداد واحدهای لایه مخفی و وزن‌های شبکه نامعلوم هستند. این پارامترها به‌گونه‌ای برآورد می‌شوند یا به طور معادل شبکه به‌گونه‌ای آموزش می‌بیند



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.

$$L^* = \sum_{i=1}^{n} \left\{ -\sum_{\ell=1}^{J-1} y_i^{(\ell)} g_\ell(X_i) + \ln\left(1 + \sum_{\ell=1}^{J-1} \exp\{g_\ell(X_i)\}\right)\right\} \qquad (۶)$$

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 اکسترمم‌یابی یک تابع، ممکن است به‌دلیل زیاد بودن اکسترمم‌های موضعی، رسیدن به اکسترمم فراموضعی[۶] با به‌کار گرفتن روش‌های کلاسیک امکان‌پذیر نباشد. به‌طور کلی در مواردی که فضای جستجو یک فضای بزرگ و پیچیده است، استفاده از الگوریتم‌های تکاملی به عنوان ابزارهای بهینه‌ساز 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 به‌کارگرفتن عملگرهای تکاملی، جمعیت بعد ساخته می‌شود، با این ایده که جمعیتی بهتر از

[۶] Global extremum

جمعیت فعلی به‌دست آید. این فرآیند تولید آن قدر تکرار می‌شود تا شرایط خاتمه محقق گردد. در نهایت، مناسب‌ترین فرد در آخرین جمعیت تولید شده به‌عنوان جواب مسأله معرفی می‌شود. (برای جزئیات بیشتر به انگلبرچ (۲۰۰۲) مراجعه شود).

به دلیل قسمت غیرخطی $\sum_{j=1}^{m} \beta_j^{\ell}(\prod_{i=1}^{k} x_i^{w_{ji}})$ که در g_ℓ ظاهرشده است، برای مینیمم کردن رابطه (۶) از یک الگوریتم تکاملی استفاده می‌شود. در این مسأله بهینه‌سازی، شبکه‌ای جستجو می‌شود که دارای تعداد بهینه واحد در لایه مخفی و بهترین وزن‌ها باشد. بنابراین هر جواب از مسأله را می‌توان به‌صورت یک شبکه در نظر گرفت. الگوریتم تکاملی در ابتدا با مجموعه‌ای از این شبکه‌ها آغاز می‌شود. به این مجموعه، جمعیت اولیه گویند و هر شبکه یک فرد نامیده می‌شود. برای تولید یک شبکه کافی است تا مقادیر پارامترهای شبکه $(m, W = (w_{11}, \cdots, w_{mk}), \beta = (\beta_1^1, \cdots, \beta_m^{J-1}), \alpha = (\alpha_0^1, \cdots, \alpha_k^{J-1}))$ به‌طور تصادفی از فضای پارامتری انتخاب شوند. در مرحله بعد مقدار برازندگی هر فرد از جمعیت اولیه با توجه به رابطه

$$A = \frac{1}{1 + L^*} \tag{۷}$$

محاسبه می‌شود. هدف الگوریتم تولید افراد با بالاترین برازندگی است. بنابراین با استفاده از جمعیت اولیه و عملگرهای تکاملی انتخاب و جهش، جمعیت بعد تولید می‌شود، با این ایده که جمعیتی بهتر از جمعیت فعلی به‌دست آید. (این عملگرها در ادامه این بخش مطرح می‌شوند.) این روند تولید آنقدر ادامه می‌یابد تا یکی از شرط‌های زیر حاصل شود.

- تعداد جمعیت‌های 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 می‌شود و به این صورت برآوردی برای پارامترهای شبکه به‌دست می‌آید. با برآورد شدن پارامترها، m تابع پایه

$z_j = \prod_{i=1}^{k} x_i^{w_{ji}},\ (j = 1, \cdots, m)$ حاصل می‌شود. در نهایت با استفاده از روش گام به گام و با توجه به متغیرهای $x_1, \cdots, x_k, z_1, \cdots, z_m$ مدل لوژستیک نهایی به‌دست می‌آید.

در الگوریتم تکاملی مورد نظر، عملگر انتخاب به این‌صورت است که پس از محاسبه برازندگی افراد، بهترین ۱۰% از جمعیت انتخاب می‌شوند و جایگزین بدترین ۱۰% جمعیت می‌گردند و عماگر جهش نیز به دو صورت پارامتری و ساختاری به شرح زیر انجام می‌شود.

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

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

ایجاد می‌کند، که در آن

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

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

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

واریانس $\gamma_r(t)$ در طول تکامل بهنگام می‌شود. روش‌های مختلفی به‌منظور بهنگام‌سازی وجود دارد. در این الگوریتم از قانون $\frac{1}{5}$ موفقیت ریچنبرگ استفاده

می‌شود (ریچنبرگ، ۱۹۷۳). طبق این قانون، نسبت جهش‌های موفقیت‌آمیز باید برابر با $\frac{1}{5}$ باشد. بنابراین اگر این نسبت بیشتر از $\frac{1}{5}$ شد، باید واریانس جهش افزایش یابد و در غیر این‌صورت واریانس جهش کم می‌شود. بنابراین

$$\gamma_r(t+s) = \begin{cases} (1+\lambda)\gamma_r(t) & s_s > \frac{1}{5} \\ (1-\lambda)\gamma_r(t) & s_s < \frac{1}{5} \\ \gamma_r(t) & s_s = \frac{1}{5} \end{cases}$$

که در آن $\lambda = 0/1$، $r = 1, 2$ و s_s نسبت جهش‌های موفقیت‌آمیز در s مرحله از تولید است. این بهنگام‌سازی باعث می‌شود که الگوریتم در یک مینیمم محلی قرار نگیرد و همچنین به فرایند تکامل سرعت می‌بخشد.

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

۱- **افزودن گره:** یک یا چند گره، با وزن‌های تصادفی به لایهٔ مخفی اضافه می‌شوند.

۲- **افزودن اتصال:** این جهش ابتدا در لایهٔ مخفی و پس از آن در لایهٔ خروجی صورت می‌گیرد. برای افزودن یک اتصال از لایهٔ ورودی به لایهٔ مخفی، یک گره به‌طور تصادفی از هر یک از این لایه‌ها انتخاب می‌شود، سپس این دو گره با یک وزن تصادفی به یکدیگر متصل می‌شوند. یک فرایند مشابه از لایهٔ مخفی به لایهٔ خروجی انجام می‌شود.

۳- **حذف گره:** یک یا چند گره از لایه مخفی به‌طور تصادفی انتخاب می‌شوند و همراه با تمام اتصالاتشان از شبکه حذف می‌گردند.

۴- **حذف اتصال:** مشابه افزودن اتصال، این جهش ابتدا در لایهٔ مخفی و سپس در لایهٔ خروجی انجام می‌شود. بدین صورت که به‌طور تصادفی دو گره به ترتیب از لایهٔ تحت جهش و لایهٔ قبل آن انتخاب می‌شوند، سپس اتصال بین این دو گره حذف می‌گردد.

۵- **ترکیب گره‌ها:** دو گره a و b به‌طور تصادفی از لایه‌ی مخفی انتخاب می‌شوند و

با گره c جایگزین می‌گردند. این گره ترکیبی از دو گره a و b است. وزن‌های مربوط به گره c، به‌صورت $\beta_c^\ell = \beta_a^\ell + \beta_b^\ell$ و $w_{cj} = (w_{aj} + w_{bj})/۲$ تعریف می‌شوند.

برای تعیین تعداد گره‌های تحت جهش، یعنی تعداد گره‌های اضافه شده به لایهٔ مخفی در جهش ساختاری افزودن گره، تعداد گره‌های حذف شده از لایهٔ 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 می‌شود، که در آن u عددی تصادفی از توزیع یکنواخت $(۰, ۱)$ است. Δ_{MIN} و Δ_{MAX} به ترتیب کمترین و بیشترین تعداد گره‌های تحت جهش را مشخص می‌نمایند و مقادیر متناظر با آنها توسط کاربر الگوریتم مشخص می‌شود. عبارت $۱ + u[\Delta_O n_O]$ تعداد اتصالات تحت جهش، بین لایهٔ مخفی و لایهٔ خروجی را تعیین می‌کند و از عبارت $۱ + u[\Delta_H n_H]$ برای تعیین تعداد اتصالات تحت جهش بین لایهٔ مخفی و لایهٔ ورودی استفاده می‌شود، که در آن‌ها u عددی تصادفی از توزیع یکنواخت $(۰, ۱)$، Δ_O و Δ_H به ترتیب نسبت تعداد اتصالات در لایهٔ خروجی و لایهٔ مخفی را مشخص می‌کنند و توسط کاربر الگوریتم تعیین می‌شوند. همچنین n_O و n_H به ترتیب تعداد اتصالات در لایهٔ خروجی و لایهٔ مخفی شبکهٔ موجود را نشان می‌دهند. 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 می‌کند. در هنگام انجام جهش همواره جهش‌های حذفی و ترکیبی قبل از جهش‌های افزایشی، مورد توجه قرار می‌گیرند 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 $T^۲(g)$ درنظر گرفته می‌شود. بر روی هر شبکه تنها یک جهش صورت می‌گیرد و اگر تحت احتمالات $T(g)$ و $T^۲(g)$ هیچ جهشی انتخاب نشد، یک جهش به‌طور تصادفی انتخاب می‌شود.

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 می‌شود. داده‌ها شامل ۵۰

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) به‌کار 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

$$\ln(\frac{\pi}{۱-\pi}) = -۶۷/۰۲ + ۱۴/۱۰w + ۴۳/۸۷YP + ۸/۲۰E + ۲۶/۰۸Si$$

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 این‌که n تعداد مشاهدات و p تعداد پارامترهای مجهول مدل را نشان دهد، از رابطه

$$AIC = -۲\log(L(\hat{\theta})) + ۲p + \frac{۲p(p+۱)}{n-p+۱} \qquad (۹)$$

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 به‌وسیله 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

$$z_۱ = w^{۰/۵}\ E^{۰/۲۶}\ c^{-۰/۲۶}\ Si^{-۰/۰۳}\ Mn^{۰/۷۲}\ Al^{۰/۹۹}$$

---

۷ Yield point

۸ Ultimate tensile strength

۹ Elongation

۱۰ Akaike's Information Criterion

با روش برآورد جدید به‌دست می‌آید. مدل لوژستیک ساخته شده به روش گام‌به‌گام، با توجه به متغیرهای تبیینی و تابع z_1 به‌صورت

$$\ln(\frac{\pi}{1-\pi}) = -32/85 + 30/34Yp + 30/23Si - 42/28Mn + 23/36z_1$$

است. معیار AIC برای این مدل برابر با ۴۰/۶۷ است که نسبت به مدل لوژستیک معمولی دو واحد کاهش در شاخص AIC ایجاد شده است. با به‌کار بردن روش ترکیبی دو تابع غیر خطی به صورت

$$z_1 = w^{-1/46}\, YP^{-0/17}\, UTS^{-0/13}\, E^{-0/6}\, c^{-0/85}\, Si^{-0/08}\, Mn^{-0/05}\, Al^{0/34}\, N^{-0/42}$$

$$z_2 = w^{-0/33}\, YP^{-0/38}\, UTS^{-0/39}\, E^{0/70}\, C^{-1/45}\, Si^{0/09}\, Mn^{1/13}\, Al^{0/72}\, N^{-1/12}$$

به مدل اضافه می‌شوند. مدل لوژستیک ساخته شده به روش گام‌به‌گام، با توجه به متغیرهای تبیینی و توابع z_1 و z_2 به صورت

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

است. معیار AIC برای این مدل برابر با ۳۷/۷۰ حاصل شده است. همان‌طور که ملاحظه می‌شود، 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 می‌کند.

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

روش معمولی	روش جدید	روش ترکیبی
۴۲/۸۵	۴۰/۶۷	۳۷/۷۰

بحث و نتیجه‌گیری

در این مقاله روش برآورد پارامترهای مدل لوژستیک ساخته شده با شبکه‌های 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 می‌کند.

تقدیر و تشکر

مولفین از داوران و هیئت تحریریهٔ محترم مجله علوم آماری قدردانی و تشکر می‌کنند.

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), *Gegeralized 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*, McGrow-Hill, New York.

Rechenberg, I., (1973), *Evolutionsstrategie: Optimierung Technischer Systeme Nach Prinzipien Der Biologischen Evolution*, Stuttgart Framman Holzboog Verlag.