مجله تخصصی اپیدمیولوژی ایران؛ ۱۳۸۹، دوره ٦، شماره ٤: صفحات ۳۹-۲۸.

مقاله پژوهشی

طراحی شبکه عصبی مصنوعی برای مدل‌بندی پاسخ‌های دو متغیره آمیخته و کاربرد آن در داده‌های پزشکی

مرتضی سدهی[۱]، یدالله محرابی[۲]،انوشیروان کاظم نژاد[۳]، وحید جوهری مجد[٤]، فرزاد حدائق[٥]

[۱] دانشجوی دکتری آمار زیستی، دانشکده پزشکی دانشگاه تربیت مدرس، تهران، ایران

[۲] استاد آمار زیستی، دانشکده پزشکی دانشگاه تربیت مدرس، تهران، ایران

[۳] دانشیار گروه مهندسی کنترل، دانشکده مهندسی برق دانشگاه تربیت مدرس، تهران، ایران

[۴] دانشیار گروه مهندسی کنترل، دانشکده مهندسی برق دانشگاه تربیت مدرس، تهران، ایران

[۵] دانشیار و فوق تخصص غدد، پژوهشکده علوم غدد درون ریز و متابولیسم، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

نویسنده رابط: یدالله محرابی، نشانی: تهران،اوین، دانشگاه علوم پزشکی شهید بهشتی، دانشکده بهداشت،گروه اپیدمیولوژی وامارزیستی. تلفن: ۲۲۴۳۲۰۴۱ ، نمابر: ۲۲۴۳۲۰۳۷ ،

پست الکترونیک: ymehrabi@gmail.com

تاریخ دریافت: ۱۳۸۸/۱۱/۱۴؛ پذیرش: ۱۳۸۹/۴/۱۲

مقدمه و اهداف: زمانی که در یک مطالعه متغیرهای پاسخ دارای مقیاس اندازه‌گیری متفاوت باشند، پاسخ‌ها را چندمتغیره آمیخته می‌گویند. با توجه به محدودیت‌ها و برقرارنبودن برخی پیش فرض‌ها، روش‌های کلاسیک آماری برای مدل‌بندی و پیش‌بینی این نوع پاسخ‌ها کارایی ندارند. هدف این مطالعه، طراحی شبکه عصبی مصنوعی برای مدل‌بندی و پیش‌بینی پاسخ‌های دومتغیره آمیخته است.

روش کار: این مطالعه شامل سه مرحله طراحی مدل، شبیه‌سازی و برازش مدل بر داده‌های واقعی است. پس از طراحی مدل، با در نظر گرفتن پارامترهای مختلف، دو مجموعه داده شبیه‌سازی و مدل‌های یک متغیره و دومتغیره مورد ارزیابی قرار گرفتند. برای آموزش شبکه از الگوریتم شیب توام مقیاس شده و برای تعیین مناسب‌ترین مدل از معیار صحت پیش‌بینی استفاده شد. مدل پیشنهادی برای پیش‌بینی توام سندرم متابولیک و شاخص مقاومت به انسولین در مطالعه قند و لیپید تهران به‌کار گرفته شد. برنامه‌های رایانه‌ای در نرم‌افزارهای R 2.9.0 و MATLAB 7.6 طراحی و اجرا گردید.

نتایج: در مجموعه شبیه‌سازی اول، صحت پیش‌بینی در مدل‌های یک متغیره و دومتغیره تقریبا یکسان ولی در مجموعه شبیه‌سازی دوم، در مدل‌های دو متغیره نسبت به مدل‌های یک متغیره بیشتر است. در مدل‌های دو متغیره، صحت پیش‌بینی مدل با افزایش همبستگی متغیرهای پاسخ، بیشتر می‌شود. در داده‌های واقعی مدل با ۱۰ گره در لایه میانی دارای بیشترین صحت پیش‌بینی است.

نتیجه‌گیری: تحقیق نشان داد، در حالتی‌که دو متغیر پاسخ با متغیرهای کمکی ارتباط دارند مدل دومتغیره نسبت به یک متغیره مناسب‌تر است و با افزایش همبستگی، صحت پیش‌بینی افزایش می‌یابد.

واژگان کلیدی: پاسخ‌های آمیخته - شبکه عصبی مصنوعی، مدل‌های دومتغیره، مطالعه قندولیپید تهران

مقدمه

در مطالعات اپیدمیولوژی و پزشکی در هنگام تحلیل داده‌ها اغلب با مواردی مواجه می‌شویم که لازم است دو متغیر پاسخ[۱] به‌صورت توام (همزمان) از روی تعدادی متغیر کمکی[۲] پیش‌بینی گردند. در ادبیات آماری این‌گونه مدل‌ها را مدل‌های دومتغیره[۳] می‌نامند. زمانی که متغیرهای پاسخ، هردوکمی یا هردو کیفی باشند، از روش‌های استاندارد آماری مانند رگرسیون دومتغیره[۴] یا رگرسیون لجستیک دومتغیره[۵] برای مدل‌بندی روابط بین متغیرهای پاسخ و کمکی استفاده می‌گردد. مبنای استنباط در این روش‌ها، در نظر گرفتن توزیع مناسب برای بردار پاسخ است. اما زمانی که یکی از متغیرهای پاسخ کمی و دیگری کیفی باشد،

[۱] Outcome

[۲] Covariate

[۳] Bivariate Models

[۴] Bivariate Regression

[۵] Bivariate Logistic Regression

[Downloaded from irje.tums.ac.ir on 2026-06-22]

اصطلاحاً این‌گونه پاسخ‌ها را پاسخ‌های آمیخته[۱] یا پاسخ‌های نامتناسب[۲] می‌گویند. به عنوان مثال در مطالعه حاضر متغیرهای پاسخ، سندرم متابولیک (کیفی) و شاخص مقاومت به انسولین (کمی) بوده که باید توسط متغیرهای تبیینی به‌صورت توام پیش‌بینی شوند.

در این حالت ساده‌ترین شیوه تحلیل، استفاده از روش‌های یک‌متغیره برای هر یک از پاسخ‌های کمی و کیفی به‌صورت جداگانه است. از آنجا که این روش، در مدل‌سازی و پیش‌بینی، همبستگی بین پاسخ‌ها را نادیده می‌گیرد، از کارایی لازم برخوردار نیست (۱). از طرفی زمانی که با پاسخ‌های آمیخته سروکار داریم، با توجه به محدودیتی که برای در نظر گرفتن توزیع احتمال مناسب برای بردار پاسخ در این‌گونه موارد وجود دارد، روش‌های دو متغیره معمول در آمار کلاسیک نیز، کارایی لازم را ندارند.

در سال‌های اخیر برای مدل‌سازی و پیش‌بینی پاسخ‌های دو-متغیره آمیخته، تحقیقاتی در آمار کلاسیک انجام شده است. راهبردهای موجود در این تحقیقات برای مدل‌بندی پاسخ‌های دو متغیره آمیخته را می‌توان به چند دسته تقسیم‌بندی کرد: روش‌های مبتنی بر فاکتورگیری توزیع توام پاسخ‌ها (۳-۱)، روش‌های مبتنی برمعرفی یک متغیر پنهان برای مدل‌سازی همبستگی بین متغیرهای پاسخ (۱۰-۴) و روش برآورد معادلات تعمیم یافته (۱۳-۱۱).

در روش‌های کلاسیک آماری، محدودیت‌هایی مانند لزوم در نظر گرفتن یک توزیع احتمال برای هر یک از متغیرهای پاسخ یا توزیع توام آن‌ها، خطی بودن رابطه بین متغیرهای وابسته و متغیرهای مستقل، یکسان بودن واریانس خطاها، عدم وجود هم خطی چند گانه واثرات متقابل مرتبه بالا، می‌تواند باعث ایجادخطا در مدل‌ها، برآوردها و پیش‌بینی‌های آن‌ها گردد (۱۴).

از آنجا که درتحقیقات علوم پزشکی و اپیدمیولوژی غالبا مسئله سلامت انسان مطرح است، پیش‌بینی صحیح نتایج اهمیت بیشتری می‌یابد، لذا لازم است حتی‌الامکان از روش‌هایی استفاده کرد که پیش‌بینی بر اساس آن‌ها دارای حداقل خطا و بیشترین اطمینان باشد. با توجه به اینکه مسائل مرتبط با پاسخ‌های آمیخته به وفور در مطالعات پزشکی مشاهده می‌شود و با در نظر گرفتن اینکه روش‌های موجود در آمار کلاسیک برای مدل‌بندی و پیش-بینی، به خاطر محدودیت‌هایشان در عمل کارایی چندانی ندارند، ارائه روش‌هایی که بتواند راهگشای این‌گونه مسائل باشد، بسیار مفید و ارزنده به نظر می‌رسد.

یکی از مناسب‌ترین روش‌ها استفاده از مدل‌های شبکه عصبی مصنوعی[۳] است که دارای محدودیت‌های روش‌های کلاسیک آماری نیست. شبکه عصبی هیچ فرض اولیه‌ای بر توزیع داده‌ها تحمیل نمی‌کند، ضمن اینکه هیچ محدودیتی نیز برای شکل تابعی رابطه بین متغیرهای مستقل و وابسته اعمال نمی‌کند بلکه شبکه عصبی، خود، این رابطه تابعی را کشف می‌کند، که لزوماً، یک رابطه خطی نیست.

مبحث شبکه عصبی مصنوعی مربوط به شبیه‌سازی قوه یادگیری در مغز انسان و پیاده‌سازی آن به صورت الگوریتم‌های کامپیوتری است. شبکه‌های عصبی از عناصر عملیاتی ساده‌ای به صورت موازی ساخته می‌شوند. این عناصر از سیستم‌های عصبی زیستی الهام گرفته شده‌اند (۱۵).

شبکه عصبی در کاربرد از جهاتی همانند یک «ذهن زنده» عمل می‌کند. به این معنا که از مشاهدات انتزاعی خود به قضاوت می‌پردازد. لذا، شبکه عصبی مدتی را صرف آموزش کرده و سپس به‌صورت عملیاتی به‌کار گرفته می‌شود. شبکه عصبی، آنچه را مشاهده می‌کند در قالب پارامترهای درونی خود به‌خاطر می‌سپارد. در واقع، تکرار هر یک از مشاهدات موجب تغییر پارامترهای درونی شبکه در جهت حفظ روابط حاکم بر مشاهدات است. آنچه در ذهن شبکه عصبی در قالب پارامترهای شبکه نگهداری می‌شود، نه تک‌تک مشاهدات بلکه الگو و برداشت کلی از مشاهدات است، بنابراین معمولاً این استواری و ثبات در عمل را دارد که در برخورد با عموم نمونه‌های مشابه، عملکردی مناسب و همراه با خطای قابل اغماض داشته باشد (۱۵).

مشاهدات یا همان نمونه‌های آموزشی شبکه عصبی، می‌تواند همراه با یک پیش قضاوت اولیه و یا بدون قضاوت اولیه باشد. به عبارتی آموزش شبکه عصبی می‌تواند با ناظر[۴] یا بدون ناظر[۵] باشد. شبکه عصبی، بسته به ساختار درونی‌اش در برخورد با مسائل مختلف، عملکرد متفاوتی دارد، لذا، انتخاب ساختار شبکه متناسب با مسئله مورد بررسی از اهمیت بسیار بالایی برخوردار است. از طرفی انتخاب مناسب مقادیر اولیه پارامترهای شبکه در نتیجه آموزش آن بسیار مؤثر خواهد بود.

---

[۱] Mixed Responses

[۲] Non - Commensurate Responses

[۳] Artificial Neural Networks

[۴] Supervised Learning

[۵] Unsupervised Learning

[Downloaded from irje.tums.ac.ir on 2026-06-22]

شبکه‌های عصبی از جنبه‌های توپولوژی، ساختاری و روش‌های یادگیری به انواع مختلفی تقسیم می‌شوند و هر یک در کاربردهای خاصی عملکرد مناسبی از خود نشان می‌دهند. شبکه عصبی چندلایه پرسپترون(MLP)[۱] با روش یادگیری پس انتشار[۲] یکی از متداول‌ترین شبکه‌های کاربردی است. در مباحث نظری اثبات شده که شبکه MLP در صورت انتخاب صحیح ساختار مناسب داخلی، قادر است هرگونه سیستم غیرخطی را مدل کرده و شبیه-سازی نماید (۱۵).

ساختار شبکه MLP شامل تعدادی گره[۳] با تابع فعالیت (محرک)[۴] مشخص است که در لایه‌های مجزا قرار دارند. هر گره، به واسطه ضرایب وزنی خود، خروجی تمامی گره‌های لایه قبلی را تجمیع کرده و از طریق تابع فعالیت به لایه بعدی ارسال می‌کند. شبکه عصبی MLP دارای یک لایه ورودی[۵]، یک لایه خروجی[۶] و حداقل یک لایه پنهانی[۷](میانی) است، تعداد گره در هر لایه متفاوت و بستگی به ساختار شبکه و مسئله مورد بررسی دارد (۱۵).

در یک شبکه MLP با یک لایه پنهانی، مقدار خروجی واحد iام از رابطه زیر به دست می‌آید:

(۱)

$$\hat{y}_i = \Phi_2\left(b_0 + \sum_{j=1}^{M} w_j \Phi_1\left(b_{j0} + \sum_{s=1}^{p} x_{is} w_{js}\right)\right) \qquad i = 1,2,.....,n$$

در این فرمول، n تعداد الگوها(مشاهدات)، M تعداد گره‌های لایه پنهانی، p تعداد گره‌های لایه ورودی (معادل تعداد متغیرهای کمکی)، w_{js} وزن مربوط به ورودی x_{is} در گره j ام، w_j وزن مربوط به گره j ام، b_{j0} و b_0 به ترتیب مقادیر اریبی گره‌های لایه میانی و لایه خروجی و Φ_2 و Φ_1 به ترتیب توابع فعالیت لایه خروجی و لایه میانی شبکه هستند. توابع فعالیت در مدل شبکه عصبی همانند توابع ربط در مدل‌های خطی تعمیم یافته است که به عنوان نمونه می‌توان به تابع خطی، تابع سیگموئید و تابع تانژانت هایپربولیک اشاره کرد.

یکی از روش‌های پیدا کردن وزن‌ها در یک شبکه MLP، الگوریتم پس انتشار خطا است. فرض کنیم مجموعه‌ای از الگوهای ورودی (مشاهدات) در دسترس باشند که خروجی‌های هر یک (پاسخ‌ها) مشخص و هدف، آموزش دادن شبکه به صورت باناظر باشد. پس از آموزش، چنین شبکه‌ای مانند مدل‌های رگرسیونی می‌تواند برای پیش‌بینی پاسخی متناظر با یک الگوی ورودی جدید مورد استفاده قرار بگیرد. برای اجرای آموزش، ضرایب وزنی شبکه در جهت حداقل کردن تابع هدف شبکه، که معمولا میانگین مربعات خطا (MSE) است، تغییر می‌کنند. برخی روش‌های محاسباتی آموزش پس انتشار نیز با هدف رسیدن به حداکثر میزان یادگیری با توجه ویژه به تعمیم‌پذیری شبکه در مواجهه با نمونه‌های جدید ارائه شده است. در این روش‌ها شبکه عصبی مجموعه آموزشی را به دو زیر مجموعه آموزش و آزمون تفکیک می‌کند. در پایان هر گام آموزش، بلافاصله عملکرد شبکه در مورد زیرمجموعه آزمون را اندازه‌گیری کرده و آموزش تا رسیدن به یک عملکرد مناسب در مورد نمونه‌های آزمون ادامه می‌یابد. اعتبار شبکه‌های عصبی مصنوعی با استفاده از داده‌های آزمون و یا اعتبار مقطعی[۸] مورد بررسی قرار می‌گیرد (۱۶).

تفسیر اپیدمیولوژیک شبکه‌های عصبی مصنوعی در مقایسه با مدل‌های آماری مرسوم پیچیده‌تر است، با این وجود، این‌گونه مدل‌ها در زمینه‌های گوناگون علوم پزشکی از جمله اپیدمیولوژی (۱۷)، پیش‌بینی سرطان پروستات (۱۸)، پیش‌بینی حاملگی ناخواسته (۱۹) و پیش‌بینی مرگ پس از جراحی قلب باز (۲۰) به کار گرفته شده‌اند.

در این مقاله، هدف، معرفی یک روش جدید بر مبنای مدل شبکه عصبی مصنوعی برای مدل‌سازی پاسخ‌های دو متغیره آمیخته با فرض دو حالتی بودن متغیر کیفی است. مطالعه شبیه‌سازی برای مقایسه صحت مدل‌های یک متغیره و دو متغیره در شرایط مختلف استفاده گردید. همچنین به منظور ارزیابی مدل دو متغیره آمیخته در داده‌های واقعی، مدل شبکه عصبی برای پیش‌بینی توام سندرم متابولیک و شاخص مقاومت به انسولین(HOMA-IR) در گروهی از افراد شرکت کننده در مطالعه قند و لیپید تهران استفاده شد.

روش کار

مطالعه حاضر در سه مرحله انجام شد که شامل طراحی مدل،

[۱] Multi-Layer Perceptron
[۲] Back-Propagation
[۳] Nodes
[۴] Activation Function
[۵] Input Layer
[۶] Output Layer
[۷] Hidden Layer
[۸] Cross Validation

[Downloaded from irje.tums.ac.ir on 2026-06-22]

مطالعه شبیه‌سازی و کاربرد مدل در داده‌های واقعی است.

طراحی مدل

در این مرحله توپولوژی مناسب مدل شبکه عصبی مصنوعی شامل تعیین نوع شبکه عصبی مورد استفاده، تعداد لایه‌های میانی شبکه، تعداد گره‌ها در هر یک از لایه‌های میانی، تعیین تابع محرک در هر لایه و مشخص نمودن روش آموزش شبکه بر مبنای مسئله مورد بررسی مشخص گردید. برای مدل‌بندی پاسخ‌های دو متغیره آمیخته از یک شبکه عصبی MLP با یک لایه میانی استفاده شد.

با توجه به این‌که در مسئله مورد بررسی، بردار پاسخ دومتغیره است، خروجی متغیر k ام از واحد i در شبکه MLP از رابطه زیر به‌دست می‌آید:

(۲)

$$\hat{y}_{ik}=\Phi_2\left(b_{0k}+\sum_{j=1}^{M}w_{jk}\Phi_1\left(b_{j0}+\sum_{s=1}^{p}x_{is}w_{js}\right)\right) \qquad i=1,....,n \quad k=1,2$$

نمادهای استفاده شده در فرمول (۲) مشابه فرمول (۱) است.

برای آموزش شبکه از الگوریتم پس انتشار استفاده گردید. الگوریتم BP دو مسیر محاسباتی دارد، مسیر اول پیشخور یا مسیر رفت و مسیر دوم پسخور یا مسیر برگشت نامیده می‌شود. برای ورودی k ام، معادلات در مسیر رفت، به شکل زیر است:

$$\hat{y}^{0}=x(k)$$
$$\hat{y}^{l+1}(k)=\hat{f}^{l+1}\left(\hat{w}^{l+1}(k).\hat{y}^{l}+\hat{b}^{l+1}(k)\right) \qquad ,l=1,2,...,L-1$$
$$\hat{y}=\hat{y}^{L}(k)$$

در این مسیر پارامترهای شبکه در خلال اجرای محاسبات تغییر نمی‌کنند و توابع فعالیت روی تک‌تک نرون‌ها عمل می‌نمایند، یعنی:

$$\hat{f}^{l+1}\left(\hat{n}(k)\right)=\left[f^{l+1}\left(\hat{n}_1(k)\right),....,f^{l+1}\left(\hat{n}_{S_{l+1}}(k)\right)\right]^T$$

در مسیر برگشت، ماتریس‌های حساسیت از لایه آخر به لایه اول برگشت داده می‌شوند. معادلات زیر پویایی مسیر برگشت را بیان می‌کنند:

$$\hat{\delta}^{L}(k)=-2\dot{\hat{f}}^{L}(k)\hat{e}(k)$$
$$\hat{\delta}^{l}(k)=\dot{\hat{f}}^{L}(\hat{n}^{l}).\left(\hat{w}^{l+1}\right)^{T}\hat{\delta}^{l+1} \qquad ,l=L-1,.....,1$$
$$\hat{e}(k)=t(k)-\hat{y}(k)$$

درمسیر برگشت، کار از لایه خروجی یعنی جایی که بردار خطا در اختیار است، شروع می‌شود. سپس بردار خطا از لایه آخر به لایه اول توزیع می‌شود و گرادیان محلی، نرون به نرون با الگوریتم بازگشتی محاسبه می‌شود. در این مسیر هم پارامترهای شبکه تغییر نخواهند کرد.

سرانجام ماتریس‌های وزن و اریبی با روابط زیر تنظیم می‌شوند:

$$\hat{w}^{l}(k+1)=\hat{w}^{l}(k)-\alpha.\hat{\delta}^{l}(k)\left(\hat{y}^{l-1}(k)\right)^{T}$$
$$\hat{b}^{l}(k+1)=\hat{b}^{l}(k)-\alpha.\hat{\delta}^{l}(k) \qquad ,l=1,2,....,L$$

در فرمول‌های اخیر L تعداد لایه‌های شبکه، f تابع فعالیت، $\hat{n}^{l}$ خروجی شبکه در لایه پنهانی، α نرخ یادگیری شبکه و $\hat{\delta}^{l}$ گرادیان تبدیل در لایه lام است. $t(k)$ مقدارخروجی واقعی نمونه k ام و $\hat{e}(k)$ خطای برآورد شده برای نمونه kام است. بحث تفصیلی چگونگی تنظیم وزن و اریبی در الگوریتم پس انتشار در ضمیمه آمده است.

در این مطالعه از الگوریتم آموزش شیب توام مقیاس شده (SCG)[1] استفاده گردید. این الگوریتم سرعت وکارایی بالایی برای آموزش دارد. جزییات کامل در مورد الگوریتم SCG در (۲۱) آمده است.

داده‌های این مقاله شامل دو بخش است. در بخش اول داده‌های عددی شبیه‌سازی شد و سپس مدل‌ها بر اساس داده‌های شبیه‌سازی در حالت‌های مختلف مورد ارزیابی قرار گرفت. در بخش دوم، مدل بر روی داده‌های واقعی مطالعه قند و لیپید تهران مورد ارزیابی قرار گرفت. در زیر به جزئیات هر یک پرداخته می‌شود.

مطالعه شبیه‌سازی

در این مطالعه دو مجموعه متفاوت داده شبیه‌سازی گردید تا مدل‌های مورد بررسی با استفاده از آن‌ها مورد ارزیابی قرار گیرند.

در مجموعه اول شبیه‌سازی، حالتی در نظر گرفته شد که همه متغیرهای کمکی با هر دو متغیر پاسخ ارتباط داشته باشند. اندازه اثر[2] متفاوت متغیر کمی روی متغیرهای پاسخ، در دو حالت اثر کم و اثر زیاد شبیه سازی گردید. ابتدا داده‌ها با استفاده از توزیع نرمال دو متغیره با پارامترهای زیر تولید گردید:

(۳)

$$\begin{pmatrix} y_{bi}^{*} \\ y_{ci} \end{pmatrix} \sim MVN\left(\begin{pmatrix} -0.5+\beta_{b1}x_i \\ 5+\beta_{c1}x_i \end{pmatrix},\begin{pmatrix} 1 & 6\rho \\ & 36 \end{pmatrix}\right)$$

[1] Scaled Conjugate Gradient

[2] Effect size

[Downloaded from irje.tums.ac.ir on 2026-06-22]

سپس متغیر دوحالتی y_{bi} به‌صورت زیر تعریف شد:

(۴)

$$y_{bi}=\begin{cases}0 & y_{bi}^{*}\le 0\\ 1 & y_{bi}^{*}>0\end{cases}$$

در عبارت بالا $X \sim Bernoulli\,(0.5)$ است. در شبیه‌سازی اول، بردار ضرایب مرتبط با متغیر تبیینی به‌صورت $(\beta_{b1},\beta_{c1})=(0.2,2)$، (به ترتیب معادل $\frac{۱}{۵}$ و $\frac{۱}{۳}$ انحراف معیار متغیر مربوط) انتخاب گردید که معادل اثر کم متغیر تبیینی روی هر دو متغیر پاسخ است. در شبیه‌سازی دوم، بردار ضرایب مرتبط با متغیر تبیینی به‌صورت $(\beta_{b1},\beta_{c1})=(1,2)$، (به ترتیب معادل ۱ و $\frac{۱}{۳}$ انحراف معیار متغیر مربوط) انتخاب گردید که معادل اثر بالای متغیر تبیینی روی متغیر پاسخ دوحالتی و اثر کم متغیر کمکی روی متغیر پاسخ کمی است. در شبیه‌سازی سوم، بردار ضرایب مرتبط با متغیر کمکی به‌صورت $(\beta_{b1},\beta_{c1})=(1,6)$، (معادل ۱ انحراف معیار متغیر مربوط) انتخاب گردید که معادل اثر زیاد متغیر کمکی روی هر دو متغیر پاسخ است.

در مجموعه دوم شبیه‌سازی، حالتی در نظر گرفته شد که برخی متغیرهای کمکی تنها با یکی از متغیرهای پاسخ پیوسته یا دوحالتی ارتباط داشته باشند. در این حالت داده‌ها به‌صورت زیر شبیه‌سازی گردید:

(۵)

$$\begin{pmatrix}y_{bi}^{*}\\ y_{ci}\end{pmatrix}\sim MVN\left(\begin{pmatrix}-1+\beta_{b1}x_{i}+\beta_{b2}x_{bi}\\ 5+\beta_{c1}x_{i}+\beta_{c2}x_{ci}\end{pmatrix},\begin{pmatrix}1 & 6\rho\\ & 36\end{pmatrix}\right)$$

که در آن $X \sim Bernoulli\,(0.5)$، $X_{b}\sim N\,(0,1)$ و $X_{c}\sim N\,(0,4)$. برای مجموعه شبیه‌سازی دوم، بردار ضرایب (β_{b1},β_{c1}) برای تعیین اندازه اثر متغیر کمکی X روی هریک از متغیرهای پاسخ همانند مجموعه شبیه‌سازی اول در نظر گرفته شد و بردار ضرایب $(\beta_{b2},\beta_{c2})=(1,1)$ در همه حالات ثابت در نظر گرفته شد.

در هر مجموعه شبیه‌سازی، از رابطه (۴) برای ساختن متغیر دوحالتی y_{bi} از روی y_{bi}^{*} استفاده گردید. متغیرهای X، X_{b} و X_{c} به‌گونه‌ای انتخاب شدند که متغیرهای کمکی هم شامل متغیرهای کمی و هم شامل متغیرهای دوحالتی باشند. در اینجا X یک متغیر کیفی دوحالتی و X_{b} و X_{c} متغیرهای کمی پیوسته هستند.

پارامتر کلیدی برای مطالعه شبیه‌سازی، میزان همبستگی بین متغیرهای پاسخ است، چون لحاظ نمودن این پارامتر در مدل‌بندی داده‌ها معیار تمایز مدل‌های دومتغیره از مدل‌های یک‌متغیره است و ممکن است روی میزان صحت پیش‌بینی مدل‌ها مؤثر باشد. بنابراین مجموعه داده‌ها با سطوح مختلف ضریب همبستگی $(\rho=۰/۳,۰/۶,۰/۹)$ تولید گردید. همبستگی بین متغیرهای پاسخ y_{bi} و y_{ci} وابسته به مقادیر متغیرهای کمکی است. برای $x_{i}=0$ (و برای $x_{bi}=x_{ci}=0$)، همبستگی متغیرهای پاسخ y_{bi} و y_{ci} متناظر با $(\rho=۰/۳,۰/۶,۰/۹)$ به‌ترتیب برابر ۰/۲، ۰/۵ و ۰/۹ به‌دست می‌آید. با توجه به مقادیر در نظر گرفته شده برای پارامترهای β_{b1}، β_{c1} و ضرایب همبستگی، در مجموع ۱۸ سناریوی شبیه‌سازی در نظر گرفته شد و برای هر سناریو ۵۰۰ نمونه مستقل هر کدام با حجم ۳۰۰ رکورد تولید گردید.

با استفاده از داده‌های دو مجموعه شبیه‌سازی، مدل‌های یک متغیره و دومتغیره بر داده‌ها برازش داده شد. در مدل‌های یک متغیره، از یک شبکه MLP با یک لایه میانی و تابع فعالیت خطی در لایه خروجی استفاده گردید. برای متغیر پاسخ دوحالتی، از تابع فعالیت سیگموئید و برای متغیر پاسخ پیوسته از تابع فعالیت تانژانت هایپربولیک در لایه میانی استفاده گردید. در مدل دومتغیره شبکه MLP با یک لایه میانی با تابع فعالیت تانژانت هایپربولیک در لایه میانی و لایه خروجی طراحی شد. مدل‌ها با تعداد گره‌های میانی ۲، ۳، ۴، ۵ و ۱۰ برازش گردید. برای تعیین تعداد گره‌های مناسب برای لایه میانی در هر یک از مدل‌ها از معیار میانگین مربعات خطا (MSE) استفاده به عمل آمد. برای آموزش شبکه در تمامی مدل‌ها از الگوریتم SCG استفاده شد. برای هر نمونه ۳۰۰ تایی، ۱۵۰ نمونه برای آموزش، ۵۰ نمونه برای آزمون و ۱۰۰ نمونه برای اعتبار سنجی در نظر گرفته شد. برای هر مدل معیار صحت پیش‌بینی[۱] به‌صورت درصد موارد پیش‌بینی صحیح تعریف و محاسبه شد. مقدار پیش‌بینی شده توسط شبکه برای متغیر دوحالتی عددی بین صفر و یک است که با در نظر

[۱] Predictive Accuracy

[Downloaded from irje.tums.ac.ir on 2026-06-22]

گرفتن نقطه برش ۰/۵ به یک متغیر دوحالتی با مقادیر صفر و یک تبدیل گردید. پیش‌بینی برای هر مورد زمانی صحیح تلقی شد که مقدار پیش‌بینی شده متغیرپاسخ پیوسته حداکثر در فاصله یک انحراف معیار از مقدار واقعی قرار داشته باشد و مقدار پیش‌بینی متغیر دوحالتی پس از تبدیل با مقدار واقعی برابر باشد. میانگین مقادیر صحت پیش‌بینی ۵۰۰ مدل حاصل از نمونه‌های مستقل برای داده‌های آزمون و داده‌های اعتبارسنجی برای مدل‌های مختلف محاسبه گردید. برای مقایسه نهایی مدل‌های تک متغیره و دومتغیره از معیار صحت پیش‌بینی استفاده شد.

کاربرد مدل در داده‌های واقعی

برای استفاده از مدل در داه‌های واقعی از اطلاعات مربوط به مطالعه قند و لیپید تهران استفاده شد. مطالعه آینده‌نگر قند و لیپید تهران بر روی جمعیت نمونه‌ای از ساکنین منطقه ۱۳ تهران، با هدف تخمین میزان شیوع اختلالات متابولیک و شناسایی عوامل خطر ساز بیماری‌های قلبی عروقی صورت گرفته است. جزئیات مربوط به این مطالعه در(۲۲) آمده است. در تحقیق حاضر، ۳۴۷ نفر از افراد مطالعه فوق که در زمان بررسی اولیه در فاز یک مطالعه فاقد سندرم متابولیک بودند، مورد بررسی قرار گرفته‌اند. از این تعداد، ۱۲۲ نفر بر مبنای تعریف ATPIII در فاز دوم مطالعه، یعنی حدود سه سال پس از بررسی اولیه، به سندرم متابولیک مبتلا شده‌اند. متغیرهای مستقل مورد بررسی در این مطالعه عبارتند از: سن، جنس، وضعیت تاهل، سابقه بیماری قلبی عروقی، نمایه توده بدن (BMI)، LDL ،HDL، کلسترول تام، تری‌گلسیرید، قند خون ناشتا، قند خون دو ساعته، مصرف سیگار(هرگز، گاهی، همیشه)، فشار خون سیستولیک، فشار خون دیاستولیک و دور کمر که مقدار همه آن‌ها در فاز یک مطالعه مورد بررسی قرار گرفت. متغیردو حالتی سندرم متابولیک و متغیر کمی HOMA-IR در فاز دوم مطالعه به عنوان متغیرهای وابسته (پاسخ) در نظر گرفته شدند.

در این مطالعه از یک شبکه MLP با تابع فعالیت تانژانت هایپربولیک در لایه میانی و تابع فعالیت خطی در لایه خروجی استفاده شد. ابتدا داده‌ها به دو بخش تقسیم شد. دربخش اول ۱۷۳ نفر(۵۰٪) برای آموزش شبکه و ۷۴ نفر(۲۱٪) برای آزمون شبکه و در بخش دوم تعداد ۱۰۰ نفر(۲۹٪) برای اعتبار سنجی مدل مورد استفاده قرار گرفت. قبل از برازش مدل، داده‌ها نرمال-سازی شدند تا در استفاده از تابع فعالیت تانژانت هایپربولیک مشکل ایجاد نشود. تعداد گره‌های لایه ورودی معادل تعداد متغیرهای مستقل یعنی ۱۵ در نظر گرفته شد. تعداد گره‌های لایه میانی از ۷ تا ۱۲ گره در نظر گرفته شد. از الگوریتم SCG برای آموزش شبکه استفاده گردید.

برای پایان هر الگوریتم، همچنین تعیین اندازه بهینه برای تعداد تکرار و ضریب یادگیری از معیار میانگین مجذور خطاها (MSE) استفاده گردید. برای تعیین مناسب‌ترین مدل در مرحله نهایی از معیار صحت پیش‌بینی که در قسمت قبلی تعریف شداستفاده گردید.

برنامه‌های رایانه‌ای در نرم افزارهای R 2.9.0 و MATLAB 7.6 طراحی و اجرا گردید.

یافته‌ها

جداول ۱ تا ۳ صحت پیش‌بینی مدل‌های یک متغیره و دو متغیره را بر اساس اندازه اثر متغیر کمکی مشترک و سطح همبستگی بین متغیرهای پاسخ در داده‌های شبیه‌سازی ارائه می‌دهد. هر یک از اعداد داخل جداول، میانگین شاخص صحت پیش‌بینی ۵۰۰ نمونه مستقل شبیه‌سازی شده هر کدام به حجم ۳۰۰ است. نتایج نشان می‌دهد در مجموعه شبیه‌سازی اول یعنی حالتی که دو متغیر پاسخ با متغیر کمکی مشترک رابطه دارند، صحت پیش‌بینی در مدل‌های یک متغیره و دو متغیره تقریبا یکسان است، اما در مجموعه شبیه‌سازی دوم (حالتی‌که هر کدام از دو متغیر پاسخ با یکی از متغیرهای کمکی ارتباط دارند، صحت پیش‌بینی در مدل‌های دو متغیره نسبت به مدل‌های یک متغیره بیشتر است و در مدل‌های دو متغیره با افزایش همبستگی متغیرهای پاسخ صحت پیش‌بینی افزایش می‌یابد. همچنین در هردو مجموعه شبیه‌سازی با افزایش اندازه اثر متغیر کمکی مشترک صحت پیش‌بینی نیز افزایش می‌یابد.

جدول ۴ درصد صحت پیش‌بینی توام سندرم متابولیک و شاخص مقاومت به انسولین را در داده‌های واقعی نشان می‌دهد. در این جدول مدل‌های تک متغیره و دومتغیره در ۷۴ نفر مجموعه آزمون و ۱۰۰ نفر مجموعه اعتبار سنجی بر اساس تعدادگره‌های لایه میانی ارائه شده‌است. نتایج نشان می‌دهد مدل با ۱۰ گره در لایه میانی دارای بیشترین صحت پیش‌بینی است. تعدادگره‌های لایه میانی از این جهت بسیار مهم است که اگر تعداد آن‌ها کم باشد شبکه برای حل مسائل غیر خطی و پیچیده با کمبود منابع یادگیری مواجه می‌شود و اگر زیاد باشد باعث ایجاد دو مشکل خواهد شد، اول آنکه زمان آموزش شبکه افزایش می‌یابد و دوم آنکه ممکن است شبکه خطاهای موجود در داده‌ها را نیز یاد بگیرد

[Downloaded from irje.tums.ac.ir on 2026-06-22]

جدول شماره ۱- درصد صحت پیش‌بینی مطالعه شبیه‌سازی برای بردار ضرایب $(\beta_{b1}, \beta_{c1}) = (0.2, 2)$

مدل	مجموعه شبیه‌سازی اول			مجموعه شبیه‌سازی دوم		
	ضریب همبستگی			ضریب همبستگی		
	۰/۳	۰/۶	۰/۹	۰/۳	۰/۶	۰/۹
مدل یک متغیره						
داده‌های آزمون	۵۳/۸۳	۵۸/۰۲	۵۷/۲۸	۶۱/۵۹	۶۳/۴۳	۶۷/۰۵
داده‌های اعتبارسنجی	۵۲/۹۵	۵۷/۹۵	۶۱/۸۵	۶۴/۱۵	۶۵/۴۳	۶۵/۷۱
مدل دومتغیره						
داده‌های آزمون	۵۱/۹۳	۵۷/۷۰	۶۱/۷۶	۶۴/۰۸	۷۴/۰۴	۷۶/۸۷
داده‌های اعتبارسنجی	۵۲/۲۸	۵۷/۹۵	۶۰/۹۵	۶۴/۶۴	۷۳/۰۰	۷۵/۸۲

جدول شماره ۲- درصد صحت پیش‌بینی مطالعه شبیه‌سازی برای بردار ضرایب $(\beta_{b1}, \beta_{c1}) = (1, 2)$

مدل	مجموعه شبیه‌سازی اول			مجموعه شبیه‌سازی دوم		
	ضریب همبستگی			ضریب همبستگی		
	۰/۳	۰/۶	۰/۹	۰/۳	۰/۶	۰/۹
مدل یک متغیره						
داده‌های آزمون	۵۳/۷۴	۵۳/۱۷	۵۳/۴۴	۶۴/۰۰	۶۰/۷۲	۶۵/۵۷
داده‌های اعتبارسنجی	۵۵/۳۰	۵۵/۰۱	۵۵/۰۰	۶۳/۴۲	۶۳/۴۴	۶۷
مدل دومتغیره						
داده‌های آزمون	۵۳/۵۴	۵۸/۲۳	۵۹/۸۲	۶۷/۵۰	۷۶/۷۳	۷۹/۸۳
داده‌های اعتبارسنجی	۵۶/۹۸	۵۷/۶۸	۶۱/۱۰	۶۷/۷۹	۷۶/۵۵	۸۰/۰۰

جدول شماره ۳- درصد صحت پیش‌بینی مطالعه شبیه‌سازی برای بردار ضرایب $(\beta_{b1}, \beta_{c1}) = (1, 6)$

مدل	مجموعه شبیه‌سازی اول			مجموعه شبیه‌سازی دوم		
	ضریب همبستگی			ضریب همبستگی		
	۰/۳	۰/۶	۰/۹	۰/۳	۰/۶	۰/۹
مدل یک متغیره						
داده‌های آزمون	۵۹/۳۷	۵۹/۳۴	۶۳/۱۹	۶۷/۸۶	۶۹/۵۲	۶۸/۴۰
داده‌های اعتبارسنجی	۶۰/۹۲	۶۰/۹۸	۶۳/۹۰	۶۷/۸۶	۶۸/۲۹	۶۷/۱۴
مدل دومتغیره						
داده‌های آزمون	۵۹/۶۵	۶۱/۱۴	۶۳/۶۴	۷۳/۳۳	۷۸/۲۸	۸۲/۰۷
داده‌های اعتبارسنجی	۵۹/۷۱	۶۱/۴۹	۶۳/۲۵	۷۴/۳۷	۷۷/۳۴	۷۹/۸۱

و در پیش‌بینی ضعیف عمل کند (۱۶).

بحث

در این تحقیق مدل شبکه عصبی مصنوعی برای مدل‌بندی پاسخ‌های دو متغیره آمیخته با فرض این‌که یکی از متغیرها کمی پیوسته و دیگری کیفی دوحالتی باشد طراحی شد. سپس یک مطالعه شبیه‌سازی برای مقایسه صحت پیش‌بینی مدل‌های یک متغیره و دومتغیره در حالت‌های مختلف انجام شد. علاوه بر آن، مدل پیشنهادی برای پیش‌بینی سندرم متابولیک و شاخص مقاومت به انسولین در نمونه‌ای از داده‌های واقعی افراد شرکت کننده در مطالعه قند و لیپید تهران، استفاده گردید.

نتایج مطالعه شبیه‌سازی نشان داد در حالتی که دو متغیر پاسخ با متغیر(های) تبیینی مشترک ارتباط دارند صحت پیش‌بینی مدل یک متغیره و مدل دومتغیره تقریباً یکسان است. در عوض، در حالتی که متغیرهای پاسخ با متغیر(های) تبیینی مختلفی ارتباط دارند صحت پیش‌بینی مدل دو متغیره از مدل یک متغیره بیشتر

[Downloaded from irje.tums.ac.ir on 2026-06-22]

جدول شماره ٤- درصدصحت پیش‌بینی سندرم متابولیک و شاخص مقاومت به انسولین بر اساس تعداد مختلف گره‌های لایه میانی بااستفاده از الگوریتم آموزشSCG

	تعداد گره‌های لایه میانی					
مدل	۷	۸	۹	۱۰	۱۱	۱۲
مدل یک متغیره						
داده‌های آزمون	۵۵/۴۰	۶۲/۱۶	۶۴/۸۶	۷۰/۲۷	۶۳/۵۱	۶۰/۸۱
داده‌های اعتبارسنجی	۵۴	۶۳	۶۳	۷۱	۶۱	۵۹
مدل دومتغیره						
داده‌های آزمون	۵۶/۷۵	۶۳/۵۱	۶۷/۵۶	۷۸/۳۷	۶۴/۸۶	۶۲/۱۶
داده‌های اعتبارسنجی	۵۴	۶۱	۶۵	۷۸	۶۲	۶۰

است. همچنین، در مجموعه شبیه‌سازی دوم، در مدل دو متغیره با افزایش ضریب همبستگی بین متغیرهای پاسخ، صحت پیش‌بینی مدل افزایش می‌یابد. در این حالت مدل دو متغیره نسبت به مدل یک متغیره مناسب‌تر است. در ارتباط با تعمیم پذیری مدل، نتایج حاصل از داده‌های اعتبارسنجی در هر دو مجموعه شبیه‌سازی نشان می‌دهد که مدل از نظر تعمیم‌پذیری عملکرد مطلوبی دارد.

در داده‌های واقعی، ضریب همبستگی بین سندرم متابولیک و شاخص مقاومت به انسولین برابر ۰/۲۱۸ است. نتایج نشان می‌دهد اختلاف زیادی بین صحت پیش‌بینی مدل یک متغیره و دو متغیره وجود ندارد، این موضوع با توجه به پایین بودن همبستگی بین متغیرهای پاسخ، قابل توجیه است.

روش‌های موجود ارائه شده در آمار کلاسیک برای مدل‌بندی پاسخ‌های دومتغیره آمیخته (۱۴-۱) علاوه بر محدودیت‌های معمول روش‌های کلاسیک، دارای مشکلاتی از قبیل مفروضات اولیه زیاد برای طراحی مدل، پیچیدگی‌های محاسباتی و عدم قابلیت اجرای این مدل‌ها توسط نرم‌افزارهای استاندارد آماری بوده که باعث می‌شود در بسیاری موارد کاربرد آن‌ها در عمل امکان‌پذیر نباشد.

از آنجا که مدل شبکه عصبی ارائه شده در این پژوهش محدودیت‌های روش‌های کلاسیک را ندارد، می‌تواند به عنوان یک روش مناسب برای مدل‌بندی و پیش‌بینی پاسخ‌های آمیخته مورد استفاده قرار گیرد.

علی‌رغم تمامی مزایایی که مدل‌های شبکه عصبی دارند، این مدل‌ها دارای محدودیت‌هایی نیز می‌باشند، از جمله این‌که در مدل‌های شبکه عصبی با توجه به اینکه توزیع پارامترهای شبکه مشخص نمی‌باشد، امکان انجام استنباط آماری برای پارامترها نیز وجود ندارد. از معایب دیگر مدل شبکه عصبی این است که برخلاف مدل‌های رگرسیون، در مدل شبکه عصبی کلاسیک امکان تعیین میزان تأثیر هریک از متغیرهای مستقل در پیش‌بینی متغیرهای پاسخ وجود ندارد، مگر اینکه از روش‌های بهینه سازی مانند الگوریتم ژنتیک استفاده گردد.

مدل‌های ارائه شده در شبکه عصبی مصنوعی برای حالت‌های دو متغیره و چند متغیره که تاکنون ارائه گردیده است، مبتنی بر پاسخ‌های دو حالتی بوده (۱۸،۱۹) و در زمینه پاسخ‌های آمیخته مطالعه‌ای انجام نشده است. بنابراین مقایسه یافته‌های تحقیق حاضر با سایر تحقیقات امکان پذیر نشد.

نتیجه‌گیری

با توجه به نتایج مطالعه می‌توان نتیجه گرفت، زمانی‌که همبستگی بین متغیرهای پاسخ زیاد باشد، استفاده از مدل دو متغیره ارجح است، ولی زمانی‌که همبستگی بین پاسخ‌ها کم باشد، مدل‌های یک متغیره و دومتغیره نتایج نسبتاً مشابهی دارند.

پیشنهادات: استفاده از روش‌های بهینه‌سازی شبکه‌های عصبی مانند الگوریتم ژنتیک یا استفاده از مدل‌های شبکه عصبی بیزی از مواردی است که علاقمندان می‌توانند در این زمینه مطالعاتی را انجام دهند. همچنین تعمیم این روش به حالت چند متغیره و یا وضعیتی که متغیر کیفی بیش از دو حالت دارد از موارد دیگری است که توسط علاقمندان می‌تواند مورد بررسی قرار گیرد.

تشکر و قدردانی

از پژوهشکده علوم غدد درون‌ریز و متابولیسم دانشگاه علوم پزشکی شهید بهشتی که داده‌های این تحقیق را در اختیار قرار داده‌اند سپاسگزاری می‌نماید. همچنین از همکاری صمیمانه آقای اکبر بیگلریان دانشجوی دکتری آمارزیستی دانشگاه تربیت مدرس و سرکار خانم مریم صفرخانی کارشناس ارشد واحد آمار

[Downloaded from irje.tums.ac.ir on 2026-06-22]

پژوهشکده علوم غدد درون ریز و متابولیسم کمال تشکر و قدردانی را به عمل می‌آوریم.

منابع

1- Cox DR, Wermuth N. Response models for mixed binary and quantitative variables, Biometrika .1992; 79:441 –61.
2- Fitzmaurice G, Larid N. Regression models for a bivariate discrete and continuous outcome with clustering, JASA .1995; 90, 845 –52.
3- Catalano PJ, Ryan LM. Bivariate latent variable models for clustered discrete and continuous outcomes. Journal of the American Statistical Association.1992; 87: 651–8.
4- Sammel MD, Ryan LM, Legler JM. Latent variable models for mixed discrete and continuous outcomes. Journal of the Royal Statistical Society, Series B: Methodological.1997; 59: 667–78.
5- Sammel M, Lin X, Ryan L. Multivariate linear mixed models for multiple outcomes. Statistics in Medicine.1999; 18: 2479–92.
6- Arminger G, Kusters U. Latent trait models with indicators of mixed measurement level. Latent Trait and Latent Class Models. 1988; Plenum Press: New York, U.S.A.
7- Dunson DB. Bayesian latent variable models for clustered mixed outcomes. Journal of the Royal Statistical Society, Series B: Statistical Methodology.2000; 62: 355–66.
8- Dunson DB, Chen Z, Harry J. A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. Biometrics.2003; 59: 521–30.
9- Lin X, Ryan L, Sammel M, Zhang D, Padungtod C, Xu X. A scaled linear mixed model for multiple outcomes. Biometrics.2000; 56: 593–601
10- Armando T.P, Saron-Lise T.N. Correlated bivariate continuous and binary outcomes: Issues and applications. Statist. Med. 2009; 28: 1753-73.
11- Liang K, Zeger S. Longitudinal data analysis using generalized linear models. Biometrika. 1986; 73: 13–22.
12- Prentice RL, Zhao LP. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. Biometrics.1991; 47: 825–39.
13- Zhao LP, Prentice RL, Self SG. Multivariate mean parameter estimation by using a partly exponential model. Journal of the Royal Statistical Society, Series B.1992; 54: 805–11.
14- Teixeira-Pinto A, Normand T. Correlated Bivariate Continuous and Binary Outcomes: Issues and Applications: Statistics in Medicine.2009; 28: 1753-73.
15- Anderson A . An introduction to neural network , Cambridge , MA : MIT press; 1995; p.795-851.
16- Wang , S. An insight into the standard back–propagation neural network model for regression analysis . J. Mgmt. Sci.1998; 26, 133 –40.
17- Duh M.S, Walker A.M, Ayania J.Z. Epidemiologic interpretation of artificial neural networks, Am. Epidemiol. 1998; 147: 1112-22.
18- Chakraborty, S. Bayesian neural networks for bivariate binary data: An application to prostate cancer study , Statistics in Medicine. 2005; 24 , 3645 –62.
19- Sadat Hashemi SM, Kazemnejad A, Lucas C. Architect of artificial neural network for modeling of multivariate binary responses and its application to predicting type of unwanted pregnancy. PhD thesis, 2003, Tarbiat Modares University.
20- Biglarian A, Babaei GHR. Application of ANN in determining important predictors of inpatients mortality after coronary artery bypass graft surgery and its comparison with logistic regression. Modares, 2005, 1: 21-30.
21- Moller MF. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks.1993; 6: 525-33.
22- Azizi F, Ghanbarian A, Momanan A, Hadaegh F, Mirmiran P, Hedayati M, Mehrabi Y et al. Prevention of non-communicable disease in a population in nutrition transition: Tehran Lipid and Glucose Study (Phase II). Trials. 2009; 10.

پیوست

الگوریتم یادگیری پس انتشار

این روش بر کمینه‌سازی تابع مربع خطا استوار شده است. در این روش کمیت‌های پاسخ از لایه آخر به لایه اول فرستاده می‌شود تا پارامترهای شبکه اصلاح شوند. اگر داده های یادگیری را به صورت زیر نشان دهیم:

$$\{(p_1,t_1),(p_2,t_2),.....,(p_Q,t_Q)\}$$

پس از اعمال ورودی $p(k)$ (که یکی از p_i ها در k امین محاسبه است) و خروجی $t(k)$ (که پاسخ مطلوب برای ورودی p_i است) سیگنال خطا در خروجی نرون j ام از لایه خروجی در k امین تکرار به صورت زیر است:

$$e_j(k)=t_j(k)-a_j(k) \qquad (۱)$$

از رابطه فوق، مربع خطا $e_j^2(k)$ و سپس مجموع مربعات خطا برای نرون‌های لایه خروجی محاسبه می‌شود:

$$F(k)=\sum_{j=1}^{S_L}e_j^2(k) \qquad (۲)$$

$F(k)$ شاخص اجرایی شبکه و S_L تعداد نرون در لایه L ام(لایه خروجی) می باشد. برای تعداد Q الگو (زوج مرتب داده یادگیری)، میانگین مربع خطاها برابر است با:

(۳)

[Downloaded from irje.tums.ac.ir on 2026-06-22]

$$F_{av}(k)=\frac{1}{Q}\sum_{l=1}^{Q}F(l)=\frac{1}{Q}\sum_{l=1}^{Q}e^{T}(l)e(l)$$

F_{av} نشان می دهد شبکه تا چه حد داده‌های یادگیری را آموخته است. تغییرات مربوط به پارامترهای شبکه را می‌توانیم به صورت زیر بنویسیم:

(۴)

$$w_{ij}^{l}(k+1)=w_{ij}^{l}(k)-\alpha\frac{\partial F(k)}{\partial w_{ij}^{l}(k)}$$

(۵)

$$b_{i}^{l}(k+1)=b_{i}^{l}(k)-\alpha\frac{\partial F(k)}{\partial b_{i}^{l}(k)}$$

اندیس i ، معرف i امین عنصر از بردار ورودی k ام به لایه l ام، و اندیس i به نرون i ام از لایه l ام مربوط می‌شود. بنابراین هدف اولیه، محاسبه جملات:

$$\frac{\partial F(k)}{\partial b_{i}^{l}(k)} \quad \text{و} \quad \frac{\partial F(k)}{\partial w_{ij}^{l}(k)}$$

برای نرون‌های لایه خروجی و لایه‌های میانی است. معادلات زیر رفتار نرون را تعیین می‌کند:

(۶)

$$n_{j}^{L}(k)=\sum_{i=1}^{S_{L-1}}a_{i}^{L-1}(k)w_{ji}^{L}(k)+b_{j}^{L}(k)$$

$$a_{j}^{L}(k)=f^{L}\left(n_{j}^{L}(k)\right) \qquad j=1,2,.....,S_{L}$$

$$e_{j}=t_{j}(k)-a_{j}^{L}(k)$$

که در آن S_{L-1} تعداد نرون‌ها درلایه L-1 و S_{L} تعداد نرون‌های لایه خروجی می‌باشد. مطابق قانون زنجیره‌ای در مشتق‌گیری می‌توانیم بنویسیم:

(۷)

$$\frac{\partial F(k)}{\partial x}=\frac{\partial F(k)}{\partial e_{j}^{L}}\cdot\frac{\partial e_{j}(k)}{\partial a_{j}^{L}(k)}\cdot\frac{\partial a_{j}^{L}(k)}{\partial n_{j}^{L}(k)}\cdot\frac{\partial n_{j}^{L}(k)}{\partial x} \qquad x=\begin{cases}b_{j}^{L}\\ w_{ji}^{L}\end{cases}$$

رابطه فوق، حساسیت F نسبت به تغییرات x را نشان می‌دهد. با توجه به معادلات (۲) و (۶) داریم:

(۸)

$$\frac{\partial F(k)}{\partial e_{j}(k)}=2e_{j}(k)$$

$$\frac{\partial e_{j}(k)}{\partial a_{j}^{L}(k)}=-1$$

(۹)

$$\frac{\partial a_{j}^{L}(k)}{\partial n_{j}^{L}(k)}=\dot{f}\left(n_{j}^{L}(k)\right)=\frac{df\left(n_{j}^{L}(k)\right)}{n_{j}^{L}(k)}$$

(۱۰)

$$\frac{\partial n_{j}^{L}(k)}{\partial x}=\begin{cases}a_{i}^{L-1} & x=w_{ji}^{L}(k)\\ 1 & x=b_{j}^{L}\end{cases}$$

با قرار دادن مقادیر فوق در رابطه (۷) خواهیم داشت:

$$\frac{\partial F(k)}{\partial x}=\begin{cases}-2e_{j}(k)\dot{f}^{L}\left(n_{j}^{L}(k)\right)a_{i}^{L-1}(k) & , \quad x=w_{ji}^{L}(k)\\ -2e_{j}(k)\dot{f}^{L}\left(n_{j}^{L}(k)\right) & , \quad x=b_{j}^{L}(k)\end{cases}$$

اکنون می‌توانیم جملات اصلاحی وزن و اریبی را حساب کنیم:

(۱۱)

$$\Delta w_{ji}^{L}(k)=2\alpha e_{j}(k)\dot{f}^{L}\left(n_{j}^{L}(k)\right).a_{i}^{L-1}(k)$$

$$\Delta b_{j}^{L}(k)=2\alpha e_{j}(k)\dot{f}^{L}\left(n_{j}^{L}(k)\right)$$

α را اصطلاحا نرخ یادگیری می‌گویند. معادلات فوق را می‌توانیم به شکل خلاصه‌تر زیر بنویسیم:

(۱۲)

$$\Delta w_{ji}^{L}(k)=-\alpha.\delta_{j}^{L}a_{i}^{L-1}(k)$$

$$\Delta b_{j}^{L}(k)=-\alpha.\delta_{j}^{L}$$

که در آن:

(۱۳)

$$\delta_{j}^{L}=-2e_{j}(k)\dot{f}^{L}\left(n_{j}^{L}(k)\right)$$

اکنون می‌توانیم معادله (۷) را به‌صورت زیر بنویسیم:

(۱۴)

$$\frac{\partial F(k)}{\partial x}=\delta_{j}^{L}.\frac{\partial n_{j}^{L}(k)}{\partial x} \qquad , \quad x=\begin{cases}b_{j}^{L}\\ w_{ji}^{L}\end{cases}$$

عبارت دوم سمت راست بالا، عددی ثابت بوده و لذا $\delta_{j}(k)$ حساسیت رفتار شبکه را به‌طور محلی اندازه می‌گیرد. $\delta_{j}(k)$ را می‌توان به شکل زیر نوشت:

(۱۵)

$$\delta_{j}(k)=\frac{\partial F(k)}{\partial n_{j}^{L}}$$

با توجه به معادلات (۱۲) و (۱۳) برای تنظیم پارامترهای شبکه

[Downloaded from irje.tums.ac.ir on 2026-06-22]

متناظر با نرون i ام از لایه خروجی مقدار $e_j(k)$ سیگنال خطای مربوط به نرون i ام بایستی در اختیار باشد. این مقدار برای نرون‌های خروجی به خاطر قابل رویت بودنشان در دسترس است، اما این سیگنال‌ها برای نرون‌های لایه میانی که نرون‌های پنهان نامیده می‌شوند، به علت این‌که $a_j^l(k)$ برای $1<l<L$ قابل رویت نیستند، قابل اندازه‌گیری نمی‌باشند. از طرفی می‌دانیم تمام نرون‌های پنهان در مقدار سیگنال خطا که در لایه خروجی شبکه ظاهر می‌شود، سهیم هستند، در حالی‌که مقدار خروجی آن‌ها در دسترس نیست. بنابراین باید راهی باشد که توسط آن نرون‌ها بسته به میزان توزیع‌شان در برابر خطا، بتوتنند طبق سیاست تشویق و تنبیه، پارامترهای خود را تنظیم کنند. الگوریتم BP اساسا چنین راهی را مدل‌سازی نموده و به این منظور ابداع شده است.

تنظیم پارامترهای لایه میانی

جهت سهولت فرض می‌کنیم شبکه دارای یک لایه میانی باشد. معادلات زیر را می‌توان برای این شبکه نوشت:

$$n_j^1(k)=\sum_{i=1}^{R}p_i(k).w_{ji}^1(k)+b_j^1(k) \qquad ,a_j^1=f^1\left(n_j^1(k)\right)$$

$$n_m^2(k)=\sum_{j=1}^{S_1}a_j^1.w_{mj}^2(k)+b_m^2(k) \qquad ,a_m^2=f^2\left(n_m^2(k)\right)$$

$$e_m(k)=t_m(k)-a_m^2(k) \qquad (۱۶)$$

اندیس j معرف شماره نرون در لایه اول، و اندیس m معرف شماره نرون در لایه دوم است. همچنین اندیس i نشان دهنده i امین ورودی، اندیس k معرف تکرار k ام در الگوریتم BP است. شاخص اجرایی را برای لایه دوم در نظر می‌گیریم:

$$F(k)=\sum_{m=1}^{S_2}e_m^2(k)$$

اکنون جملات اصلاحی Δw و Δb را با استفاده از معادلات (۴)، (۵) و (۷) تعیین می‌کنیم:

$$\frac{\partial F(k)}{\partial e_m(k)}=2e_m(k) \qquad ,\frac{\partial e_m(k)}{\partial a_m^2(k)}=-1$$

$$\frac{\partial F(k)}{\partial a_j^L(k)}=2\sum_{m=1}^{S_2}e_m(k).\frac{\partial e_m(k)}{\partial a_j^L(k)}$$

$$\frac{\partial e_m(k)}{\partial a_j^1(k)}=\frac{\partial e_m(k)}{\partial n_m^2(k)}.\frac{\partial n_m^2(k)}{\partial a_j^1(k)}$$

$$\frac{\partial e_m(k)}{\partial n_m^2(k)}=\frac{\partial e_m(k)}{\partial a_m^2(k)}.\frac{\partial a_m^2(k)}{\partial n_m^2(k)}=-f^{\cdot 2}\left(n_m^2(k)\right)$$

$$\frac{\partial n_m^2(k)}{\partial a_j^1(k)}=w_{mj}^2(k)$$

بر اساس معادله (۷) داریم:

$$\frac{\partial F(k)}{\partial x(k)}=\frac{\partial F(k)}{\partial n_j^1(k)}.\frac{\partial n_j^1(k)}{\partial x(k)}$$

که در آن:

$$\frac{\partial F(k)}{\partial n_j^1(k)}=\frac{\partial F(k)}{\partial n_m^1(k)}.\frac{\partial n_m^1(k)}{\partial n_j^1(k)}=-2\sum_m\left[e_m(k).f^{\cdot}\left(n_m^2(k)\right)w_{mj}^2(k)\right].f^{\cdot 1}\left(n_j^1(k)\right)$$

با توجه به رابطه (۱۳) برای $L=2$ داریم:

(۱۷)

$$\frac{\partial F(k)}{\partial n_j^1}=-2\left[\sum_{m=1}^{S_2}\delta_m^2(k).w_{mj}^2(k)\right].f^{\cdot 1}\left(n_j^1(k)\right)$$

سمت چپ معادله فوق گرادیان محلی برای نرون i ام از لایه میانی اول است، و سمت راست رابطه فوق نشان می‌دهد که حساسیت رفتار شبکه چگونه توسط δ_m^2 ها (حساسیت رفتار شبکه در لایه خروجی) می‌تواند بیان شود. زیرا با توجه به روابط

$$\delta_j^1(k)=\frac{\partial F(k)}{\partial n_j^1}$$

قبلی گرادیان محلی، به عبارت $f^{\cdot 1}\left(n_j^1(k)\right)$ بستگی دارد. مقادیر δ_m^2 ها با استفاده از (۱۳) قابل محاسبه هستند، زیرا سیگنال‌های خطا در لایه خروجی در اختیار هستند و می‌توانیم پارامترهای نرون i ام را مطابق فرمول زیر تنظیم کنیم:

$$\frac{\partial F(k)}{\partial x(k)}=\frac{\partial F(k)}{\partial n_j^1(k)}.\frac{\partial n_j^1(k)}{\partial x(k)}$$

$$\frac{\partial F(k)}{\partial x(k)}=\begin{cases}\delta_j^1(k).p_i(k) & if \quad x=w_{ji}^1(k)\\ \delta_j^1(k) & if \quad x=b_j^1\end{cases}$$

در این رابطه δ_j^1 حساسیت F نسبت به تغییرات ورودی خالص نرون i ام است. در حالت کلی برای نرون i ام از لایه l ام رابطه زیر را می‌توان نوشت:

$$\frac{\partial F(k)}{\partial x}=\begin{cases}\delta_j^l(k).a_i^{l-1}(k) & if \quad x=w_{ji}^l(k)\\ \delta_j^l(k) & if \quad x=b_j^l\end{cases}$$

بنابراین روابط بازگشتی زیر را می‌توانیم بنویسیم:

[Downloaded from irje.tums.ac.ir on 2026-06-22]

$$w_{ji}^{l}(k+1)=w_{ji}^{l}(k)-\alpha.\delta_{j}^{l}(k).a_{i}^{l-1}(k)$$
$$b_{j}^{l}(k+1)=b_{j}^{l}(k)-\alpha.\delta_{j}^{l}(k) \qquad j=1,2,.....,S_{l} \;,\; i=1,2,...,S_{l-1}$$
(۱۸)

اگر ماتریس‌های زیر را تعریف کنیم:

$$\hat{w}^{l}(k)=\left[w_{ji}^{l}(k)\right]_{S_l\times S_{l-1}}$$
$$\hat{\delta}^{l}(k)=\left[\frac{\partial F(k)}{\partial n_1^l},....,\frac{\partial F(k)}{\partial n_{S_l}^l}\right]^T$$
$$\hat{a}^{l-1}=\left[a_i^{l-1}\right]^T_{S_{l-1}\times 1}$$
$$\hat{b}^{l-1}=\left[b(k)\right]_{S_l\times 1}$$

شکل ماتریسی معادلات (۱۸) به‌صورت زیر است:
(۱۹)

$$\left[w_{ji}^{l}(k+1)\right]_{S_l\times S_{l-1}}=\left[w_{ji}^{l}(k)\right]_{S_l\times S_{l-1}}-\alpha\left[\frac{\partial F(k)}{\partial n_1^l},....,\frac{\partial F(k)}{\partial n_{S_l}^l}\right]^T\left[a_i^{l-1}\right]^T_{S_{l-1}\times 1}$$
$$\left[b_j(k+1)\right]_{S_l\times 1}=\left[b_j(k)\right]_{S_l\times 1}-\alpha\left[\frac{\partial F(k)}{\partial n_1^l},....,\frac{\partial F(k)}{\partial n_{S_l}^l}\right]^T$$

پس انتشار حساسیت‌ها

برای نرون‌های لایه L-1 به مقادیر δ_j^{L-1} احتیاج داریم. این مقادیر را می‌توان با توجه به رابطه (۱۷) و مطابق فرمول زیر به‌دست آورد:
(۲۰)

$$\delta_j^{L-1}=-2\left[\sum_{m=1}^{S_L}\delta_m^L(k).w_{mj}^{L}(k)\right].\dot{f}^{L-1}\left(n_j^{L-1}(k)\right)$$

در نهایت، با تکرار رابطه فوق می‌توانیم δ_j^1 را از روی δ_j^2 محاسبه کنیم.

برای ارائه شکل ماتریسی رابطه فوق از ماتریس ژاکوبین استفاده می‌کنیم:

$$\begin{bmatrix} \frac{\partial n_1^l}{\partial n_1^{l-1}} & \frac{\partial n_1^l}{\partial n_2^{l-1}} & .... & \frac{\partial n_1^l}{\partial n_{S_{l-1}}^{l-1}} \\ . & & & \\ . & & & \\ . & & & \\ \frac{\partial n_{S_1}^l}{\partial n_1^{l-1}} & \frac{\partial n_{S_2}^l}{\partial n_2^{l-1}} & .... & \frac{\partial n_{S_1}^l}{\partial n_{S_{l-1}}^{l-1}} \end{bmatrix}_{S_l\times S_{l-1}}$$ (۲۱)

درایه (i,j) ام ماتریس فوق را می‌توان به شکل زیر نوشت:
(۲۲)

$$\frac{\partial n_i^l(k)}{\partial n_j^{l-1}(k)}=\frac{\partial\left[\sum_{m=1}^{S_{l-1}}w_{im}^{l}(k).a_m^{l-1}(k)+b_i^l(k)\right]}{\partial n_j^{l-1}(k)}=w_{ij}^{l}(k).\frac{\partial a_j^{l-1}(k)}{\partial n_j^{l-1}(k)}=w_{ij}^{l}(k).\dot{f}^{l-1}\left(n_j^{l-1}(k)\right)$$

اگر قرار دهیم:

$$\hat{f}^{l-1}\left(\hat{n}^{l-1}(k)\right)=diag\left(\left[\dot{f}^{l-1}\left(n_1^{l-1}(k)\right),....,\dot{f}^{l-1}\left(n_{S_{l-1}}^{l-1}(k)\right)\right]\right)$$

که در آن diag ماتریسی قطری را تعریف می‌کند، ماتریس ژاکوبین (۲۱) به‌صورت زیر نوشته خواهد شد:

$$\frac{\partial\hat{n}^{l}}{\partial\hat{n}^{l-1}}=\left[w_{ji}^{l}(k)\right]_{S_l\times S_{l-1}}.\hat{\dot{f}}^{l-1}\left(\hat{n}^{l-1}(k)\right)$$

با استفاده از قاعده مشتقات زنجیره‌ای، رابطه بازگشتی مطلوب را به‌دست می‌آوریم:
(۲۳)

$$\hat{\delta}^{l-1}(k)=\frac{\partial\hat{F}(k)}{\partial\hat{n}^{l-1}(k)}=\left[\frac{\partial\hat{n}^{l}(k)}{\partial\hat{n}^{l-1}(k)}\right]^T.\frac{\partial\hat{F}(k)}{\partial\hat{n}^{l}(k)}=\hat{\dot{f}}^{l-1}\left(\hat{n}^{l-1}(k)\right).\left(\hat{w}^{l}(k)\right).\hat{\delta}^{l}(k)$$

از معادله (۲۳) مشاهده می‌شود که انتخاب نام پس انتشار برای این الگوریتم مناسب بوده است، زیرا بر اساس معادلات فوق حساسیت‌ها از لایه‌های بالاتر به لایه‌های میانی پایین‌تر در شبکه انتشار می‌یابند، به‌طوری‌که:

$$\hat{\delta}^{L}\rightarrow\hat{\delta}^{L-1}\rightarrow\hat{\delta}^{L-2}\rightarrow....\rightarrow\hat{\delta}^{1}$$

بنابراین لازم است ابتدا $\hat{\delta}^{L}$ در لایه خروجی محاسبه شود. برای این کار از رابطه (۱۳) استفاده می‌کنیم و رابطه برداری زیر را خواهیم داشت:

$$\hat{\delta}^{L}(k)=-2\hat{\dot{f}}^{L}\left(\hat{n}^{L}(k)\right)\hat{e}(k)$$

که در آن:

$$\hat{e}(k)=\hat{t}(k)-\hat{a}^{L}\left(\hat{p}(k),k\right)$$

[Downloaded from irje.tums.ac.ir on 2026-06-22]