

## مدل‌بندی تعداد ویزیت‌های پزشکی با استفاده از رگرسیون شبکه عصبی کانوی-مکسول-پواسون

ترنم ترابی نعمان<sup>۱</sup>، مهدی عمادی<sup>۲\*</sup> و محمد آرشی<sup>۳</sup>

۱،۲،۳ گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد، مشهد، ایران

تاریخ دریافت: ۱۴۰۴/۰۶/۲۲

تاریخ پذیرش: ۱۴۰۴/۰۹/۱۵

### چکیده:

از آنجایی که بسیاری از تحلیل‌های تجربی و علمی بر پایه داده‌های شمارشی صورت می‌گیرد، مدل‌بندی داده‌های شمارشی از اهمیت بسزایی برخوردار است. از جمله این تحلیل‌ها می‌توان به رگرسیون تعداد مراجعات به پزشک بر پایه یک سری متغیر پیشگو اشاره کرد. بدیهی است دسترسی به مدل رگرسیون پیشگوی مناسب می‌تواند تاثیر زیادی بر تصمیم‌گیری‌ها و برنامه‌ریزی‌ها داشته باشد. در این پژوهش در راستای بهبود دقت مدل‌بندی، رگرسیون پواسون با شبکه عصبی ترکیب شده است تا مدل‌سازی بهتری برای داده شمارشی داشته باشیم. در این راستا، مدل شبکه عصبی پواسون همراه با الگوریتم محاسباتی ارائه شده و به منظور ارزیابی کارایی آن از دیدگاه برآورد پارامتر و دقت پیشگویی، با دو مدل جنگل تصادفی و گرادیان تقویتی از روش‌های یادگیری گروهی، مقایسه شده است. برای بررسی پدیده بیش‌پراکنش از مدل شبکه عصبی کانوی-مکسول-پواسون (کام-پواسون) استفاده کرده و با شبکه عصبی پواسون مقایسه انجام شد. با پیاده‌سازی این مدل‌ها بر روی یک مجموعه داده واقعی پزشکی، نشان داده می‌شود که مدل رگرسیون شبکه عصبی کام-پواسون از عملکرد بهتری برخوردار است. **واژه‌های کلیدی:** رگرسیون پواسون، رگرسیون کانوی-مکسول-پواسون، شبکه‌های عصبی، پرسپترون چند لایه، داده شمارشی، جنگل تصادفی، گرادیان تقویتی.

### ۱ مقدمه

محدودیت‌هایی که روش‌های سنتی دارند، نیاز علم امروز استفاده از روش‌هایی است که با تولید خطای کمتر و دقت بیشتر، اعتماد بیشتری را از منظر دقت پیش‌گویی حاصل کند. روش‌های دیگر از جمله مدل رگرسیونی بعضی مواقع محدودیت‌های خطی دارند که مناسب نبوده و در برخی موارد دقت کم‌تری نسبت به حالت غیرخطی دارند، درحالی‌که شبکه عصبی به عنوان یک رویکرد غیرخطی این نقطه ضعف را برطرف می‌کند. فلاح و همکاران [۱۶] به بررسی این موضوع پرداخته‌اند. مورد دیگر استفاده از نسبت درستی‌نمایی به‌عنوان یک ابزار مهم برای تعیین خطا و تعیین پارامترها است. به منظور ارزیابی عملکرد مدل و مقایسه پیش‌گویی‌ها با داده‌های واقعی، می‌توان از نسبت درستی‌نمایی بهره گرفت و با استفاده از آن دریابیم که کدام مدل یا فرضیه برای داده‌ها مناسب‌تر است. ریزوی و پته [۱۸] نشان دادند با استفاده از تقریب نسبت درستی‌نمایی به وسیله شبکه عصبی می‌توان به نتایج بهتری رسید. اخیراً مونتسینوس-لوپز و سالازار [۱۳] نشان دادند که مدل شبکه عصبی عمیق پواسون عملکرد بهتری نسبت به روش‌های رگرسیون بیزی و

یکی از دلایل اصلی استفاده از روش‌های مبتنی بر یادگیری ماشین در تحلیل داده‌های شمارشی، توانایی این روش‌ها در مدل‌بندی روابط پیچیده و غیرخطی، به‌ویژه زمانی که با مشکلاتی مانند بیش‌پراکنش، کم‌پراکنش مواجه هستیم، است. در حالی که مدل‌های آماری سنتی، مانند مدل‌های خطی یا لگاریتمی، عمدتاً برای داده‌هایی با روابط خطی و ساختار ساده‌تر طراحی شده‌اند. این مدل‌ها در برابر داده‌های شمارشی با ویژگی‌های پیچیده عملکرد ضعیفی از خود نشان می‌دهند. در سال‌های اخیر، رویکردهای پیشرفته مبتنی بر یادگیری ماشین، مانند درخت تصمیم، جنگل تصادفی، ماشین بردار پشتیبان و گرادیان تقویتی به‌طور فزاینده‌ای استفاده می‌شوند. برای آگاهی بیشتر شو و یه [۲۶] را ببینید.

در مسئله اصلی مدل‌بندی داده‌های شمارشی علاقه‌مندیم شبکه‌ای منحصر به فرد برای اهداف مطلوب بسازیم. با توجه به بررسی

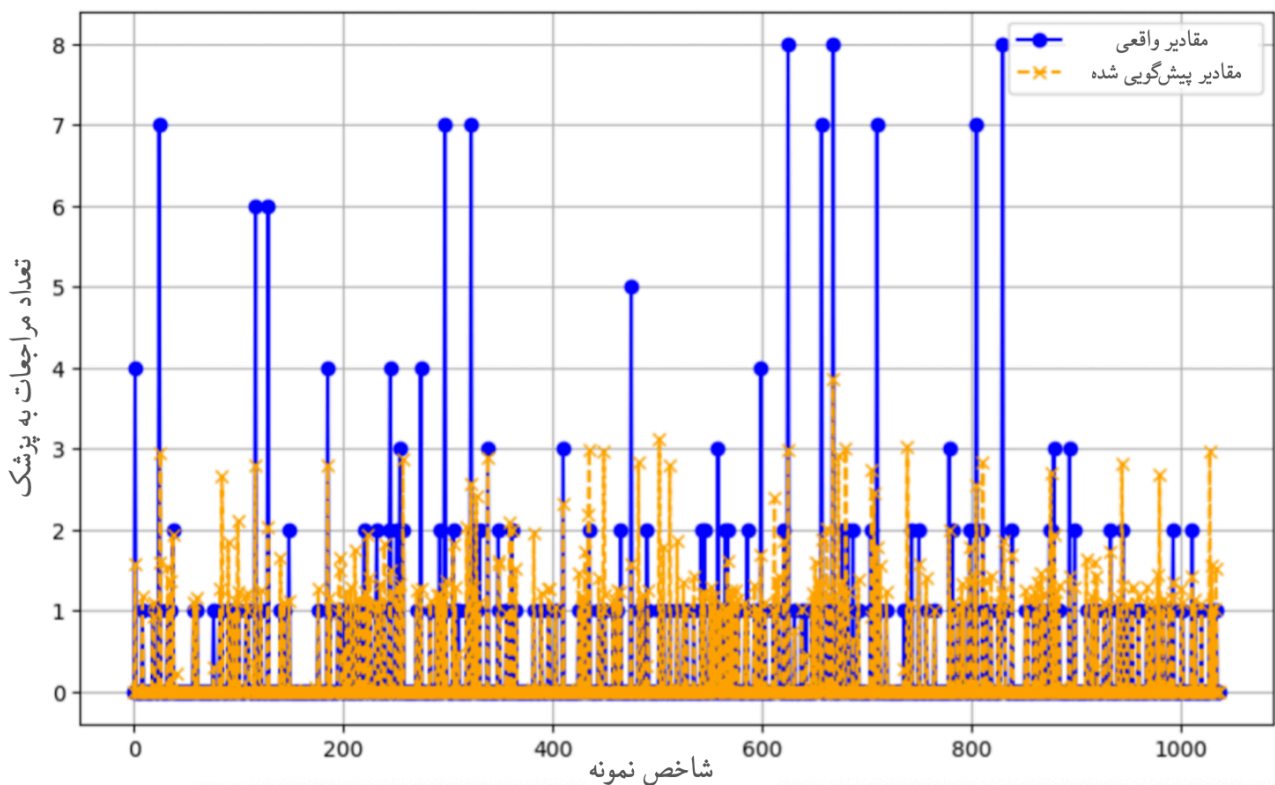
گرفته شود. برای این منظور، چالش اصلی در داده‌های مورد استفاده به عنوان یک مثال کاربردی در بخش بعدی ارائه می‌شود.

## ۱.۱ مثال انگیزشی

با هدف توسعه مدل برای پیش‌گویی دقیق‌تر تعداد ویزیت‌های پزشکی، از یک مجموعه داده در دسترس بهره گرفته‌ایم. این مجموعه داده شامل اطلاعات مربوط به تعداد ویزیت‌های پزشکی افراد در دوره یک ساله (۱۹۷۷ تا ۱۹۷۸) بوده که به عنوان منبعی برای مدل‌سازی داده‌های شمارشی استفاده می‌شود. در این مطالعه ۵۱۹۰ شرکت‌کننده مورد بررسی قرار گرفتند که قادر به پاسخ‌گویی و ارائه اطلاعات دقیق درباره وضعیت اقتصادی و سلامتی و دسترسی به خدمات بهداشتی بودند. شرکت‌کنندگان تعداد ویزیت‌های پزشکی خود را در طول دوره گزارش کردند که این متغیر به‌عنوان متغیر وابسته در نظر گرفته شد. شکل ۱ نمودار مقادیر واقعی در مقابل مقادیر پیش‌گویی شده را توسط یک مدل رگرسیون پواسون نشان می‌دهد.

مدل‌های رگرسیون پواسون دارد. برای مشاهده تحقیقات مشابه برپایه روش‌های یادگیری آماری-ماشینی و هوش مصنوعی، می‌توان به مددی و مترجم [۱] مراجعه کرد. مطالعاتی نظیر شاراف و تسکوس [۲۵] که در آن از شبکه‌های عصبی مصنوعی برای مدل‌سازی زمان بقا گسسته و داده‌های سانسور شده استفاده شده است، می‌تواند مورد توجه قرار گیرد. رودریگو [۲۰] در خصوص کاربرد مدل‌های شبکه‌های عصبی بیزی در حوزه سلامت منبع مناسبی برای مقایسه و گسترش دانش موجود در این چارچوب محسوب می‌شود. علاوه بر این، می‌توان به رودریگو و تسکوس [۱۹] که به مدل‌سازی بیزی غیرخطی داده‌های شمارشی با استفاده از شبکه‌های عصبی مصنوعی پرداخته است نیز اشاره کرد. در این راستا لی و ها [۱۲] در زمینه کاربرد شبکه‌های عمیق در مدل‌های شکستگی نیمه‌پارامتری نیز توسعه‌هایی ارائه داده‌اند.

هدف اصلی این مقاله توسعه رگرسیون پواسون به رگرسیون شبکه عصبی کانوی-مکسول-پواسون در راستای افزایش عملکرد مدل در مجموعه‌ای از داده‌های پزشکی است. در این راستا نشان داده می‌شود که این روش چگونه می‌تواند برای مدل‌بندی مجموعه داده پزشکی به کار



شکل ۱: مقایسه مقادیر واقعی و مقادیر پیش‌گویی شده در داده‌های تعداد مراجعات به پزشک

$$y_j = a_{out} \left( b_j + \sum_{m=1}^j w_{mj} \cdot a_{hid} \left( b_m + \sum_{i=1}^m w_{im} \cdot x_i \right) \right), \quad (1)$$

که در آن  $w_{mj}$  وزن بین لایه پنهان  $m$  و لایه خروجی  $j$ ، و  $w_{im}$  وزن بین لایه ورودی  $i$  و لایه پنهان  $m$  است. همچنین  $b_j$  اریبی در لایه خروجی و  $b_m$  اریبی در لایه پنهان است. وزن‌ها می‌توانند از طریق کمینه‌سازی مجموع توان دوم خطاهای متغیر پیش‌گویی یا از طریق بیشینه‌سازی درست‌نمایی داده‌ها تعیین شوند. در این راستا می‌توان از الگوریتم‌های بهینه‌سازی مشتقات نزولی یا مختصات نزولی استفاده کرد.

## ۱.۱.۲ رگرسیون شبکه عصبی پواسون

مدل رگرسیون شبکه عصبی پواسون یک توسعه از مدل پرسپترون چند لایه است که برای مدل‌سازی داده‌های شمارشی طراحی شده است. در این مدل، فرض می‌شود که متغیر وابسته از توزیع پواسون پیروی می‌کند. به عبارت دیگر، احتمال مشاهده یک مقدار خاص از متغیر وابسته ( $Y$ ) از تابع احتمال  $P(Y=y) = \frac{\lambda^y e^{-\lambda}}{y!}$ ،  $y = 0, 1, 2, \dots$  محاسبه می‌شود. که در آن  $\lambda > 0$  پارامتر نرخ توزیع پواسون است. در مدل شبکه عصبی پواسون،  $\lambda$  به عنوان خروجی شبکه عصبی محاسبه می‌شود. این پارامتر به صورت تابعی از ورودی‌ها و وزن‌های شبکه عصبی به صورت

$$\lambda = \exp(f(x_i; W))$$

تعریف می‌شود که در آن  $f(x_i; W)$  تابعی است که در آن  $x_i$  بردار ورودی و  $W$  پارامترهای شبکه هستند که توسط شبکه عصبی مدل‌سازی می‌شود که در آن  $W$  شامل تمام وزن‌ها و اریبی‌های شبکه است. بنابراین، مدل عملاً یک تابع غیرخطی از ویژگی‌های ورودی  $x_i$  با پارامتر پواسون  $\lambda_i$  تعریف می‌کند. از آنجایی که  $\lambda$  باید مثبت باشد، از تابع نمایی استفاده می‌شود. هدف از آموزش مدل، بیشینه‌سازی تابع درست‌نمایی داده‌ها است. لگاریتم تابع درست‌نمایی برای مجموعه داده:

$$(x_i, y_i), i = 1, 2, \dots, N$$

به صورت

$$\log(L(W)) = \sum_{i=1}^N [y_i \log(\lambda_i) - \lambda_i - \log(y_i!)]$$

محاسبه می‌شود. از آنجایی که جمله  $\log(y_i!)$  مستقل از پارامترها است، می‌توان آن را از تابع هدف حذف کرد. بنابراین، تابع هدف نهایی که باید بیشینه شود به صورت زیر است:

$$\mathcal{L}(W) = \sum_{i=1}^N [y_i \log(\lambda_i) - \lambda_i]. \quad (2)$$

همان‌طور که در شکل ۱ مشاهده می‌شود به دلیل قله‌های زیاد پیش‌گویی تعداد ویزیت‌ها با چالش‌هایی همراه بوده و توسعه مدل‌های دقیق‌تر توصیه می‌شود. بنابراین جهت دستیابی به الگوی رفتاری عمیق‌تری از بیماران با توجه به متغیرهای پیش‌گو، می‌توان از مدل‌های یادگیری ماشین، از جمله شبکه عصبی کام-پواسون استفاده کرد. ادامه مقاله به بررسی رویکردهای مختلف تحلیل داده‌های شمارشی با تأکید بر روش‌های یادگیری ماشین و مقایسه آن‌ها با یکدیگر اختصاص دارد که به شرح زیر سازمان‌دهی شده است. در بخش دوم، به بررسی نظریه و مبانی روش‌های یادگیری ماشین مانند شبکه عصبی، جنگل تصادفی و گرادیان تقویتی پرداخته شده است. بررسی کاربرد مدل بهینه و تحلیل داده‌های مربوط به مراجعات به پزشک در بخش سوم انجام شده است و در نهایت، نتایج حاصل از تحلیل داده‌ها ارائه شده و عملکرد مدل‌ها ارزیابی شده است. در پایان نتیجه مطالعات صورت گرفته به طور مختصر در بخش چهار ارائه شده است. از آنجایی که برای مدل‌بندی رگرسیون شبکه عصبی به یک الگوریتم عددی نیاز است، پیوست این مقاله در برگزیده الگوریتم بهینه‌سازی مورد نیاز است.

## ۲ مدل‌های یادگیری ماشین

در این بخش، به طور مختصر، به معرفی مدل رگرسیون شبکه عصبی پواسون می‌پردازیم. همچنین برای ارزیابی رفتار این مدل و مقایسه، دو مدل جنگل تصادفی و گرادیان تقویتی نیز همراه با الگوریتم‌های محاسباتی ارائه می‌شود.

### ۱.۲ مدل شبکه عصبی پواسون

مدل پرسپترون چند لایه (MLP) یکی از پرکاربردترین مدل‌های شبکه عصبی در حوزه پزشکی است که شامل یک مجموعه لایه پرسپترون (مجموعه نورون‌های موازی و کاملاً متصل) بوده که به صورت متوالی قرار گرفته‌اند. مدل MLP از سه نوع لایه ورودی، یک یا چند لایه پنهان و یک لایه خروجی، تشکیل شده است. هر نورون در این شبکه وزن ( $w$ ) و اریبی ( $b$ ) متفاوتی دارد. نورون‌های لایه پنهان، تابع فعال‌ساز را از نتایج لایه ورودی محاسبه کرده سپس این مقدار به نورون‌های لایه پنهان بعدی یا به لایه خروجی منتقل می‌شود.

با یک لایه پنهان و تابع فعال‌ساز  $a$  و ورودی  $x_i$  و خروجی  $y_j$  مدل MLP مشابه پژوهش فلاح و همکاران [۱۶] دارای رابطه رگرسیون غیرخطی زیر است:

این انعطاف‌پذیری، مدل کام-پواسون را به یک چارچوب جامع برای تحلیل داده‌های شمارشی تبدیل کرده است.

شبکه عصبی پیشنهادی شامل سه نوع لایه است؛ لایه ورودی شامل متغیرهای توضیحی  $x_i$  است، یک یا چند لایه پنهان با نورون‌هایی که در تابع فعال‌ساز ReLU را به‌کار می‌برند و لایه خروجی شامل دو نورون که به ترتیب مقادیر  $\lambda_i$  و  $v_i$  را برای هر نمونه  $i$  تولید می‌کنند.

برای تضمین مثبت بودن پارامترهای توزیع، از توابع فعال‌ساز مناسب استفاده می‌شود. برای  $\lambda_i$ ، تابع Softplus یا نمایی به‌کار می‌رود،

$$\lambda_i = \exp(f_\lambda(x_i; W_\lambda))$$

برای  $v_i$ ، نیز از تابع Softplus یا نمایی استفاده می‌شود تا  $v_i > 0$  تضمین شود، به طوری‌که

$$v_i = \exp(f_v(x_i; W_v)), i = 1, 2, \dots, N.$$

در اینجا،  $f_\lambda$  و  $f_v$  دو تابع غیرخطی هستند که توسط شبکه عصبی مدل می‌شوند و  $W = \{W_\lambda, W_v, b_\lambda, b_v\}$  مجموعه تمام پارامترهای قابل یادگیری شبکه است.

تابع درستنمایی برای یک مجموعه داده شامل  $N$  نمونه مستقل  $(x_i, y_i)$  به صورت

$$L(W) = \prod_{i=1}^N \frac{\lambda_i^{y_i}}{(y_i!)^{v_i} Z(\lambda_i, v_i)},$$

است. بنابراین، لگاریتم درستنمایی به صورت

$$\log L(W) = \sum_{i=1}^N [y_i \log \lambda_i - v_i \log(y_i!) - \log Z(\lambda_i, v_i)]$$

خواهد بود.

از آنجا که  $\log(y_i!)$  مستقل از پارامترهای مدل است، در بهینه‌سازی نقشی ندارد. با این حال، تابع نرمال‌سازی  $\log Z(\lambda_i, v_i)$  وابسته به پارامترهاست و باید در محاسبات در نظر گرفته شود. این تابع معمولاً به صورت عددی (مثلاً با استفاده از سری‌های همگرا یا تقریب‌های لگاریتمی) محاسبه می‌شود.

در نتیجه، تابع لگاریتم درستنمایی به صورت

$$\mathcal{L}(W) = \sum_{i=1}^N [y_i \log \lambda_i - v_i \log(y_i!) - \log Z(\lambda_i, v_i)]$$

است.

به عبارت دیگر، مدل شبکه عصبی پواسون با استفاده از الگوریتم‌های بهینه‌سازی آموزش داده می‌شود تا پارامترهای شبکه ( $W$ ) به گونه‌ای تنظیم شوند که تابع هدف (۲) بیشینه شود.

## ۲.۱.۲ محاسبه گرادیان خطا

پارامترهای مدل شبکه عصبی در این پژوهش مشابه با روش رامله‌ارت و هینتون [۲۱] با استفاده از الگوریتم پس‌انتشار محاسبه شده است که از قوانین مشتق‌گیری زنجیره‌ای برای به‌روزرسانی وزن‌ها و اریبی‌ها استفاده می‌کند.

در این مقاله، از تابع ReLU<sup>۱</sup> به عنوان تابع فعال‌سازی در لایه پنهان و از تابع Softplus به عنوان تابع فعال‌سازی در لایه خروجی استفاده شده است. مراحل محاسبه گرادیان خطا به صورت الگوریتمی در پیوست ارائه شده است.

## ۲.۲ مدل رگرسیون شبکه عصبی کانوی-مکسول-پواسون

مدل رگرسیون شبکه عصبی کانوی-مکسول-پواسون یک تعمیم از مدل پرسپترون چندلایه (MLP) است که برای مدل‌سازی داده‌های شمارشی با ساختار پراکنندگی عمومی طراحی شده است. در این مدل، فرض می‌شود که متغیر وابسته  $Y$  از توزیع کام-پواسون پیروی می‌کند. می‌گوییم متغیر تصادفی  $Y$  دارای توزیع کام-پواسون است اگر تابع جرم احتمال آن به صورت

$$P(Y=y) = \frac{\lambda^y}{(y!)^v Z(\lambda, v)}, \quad y = 0, 1, 2, \dots, \quad (3)$$

تعریف شود که در آن  $\lambda > 0$  پارامتر نرخ یا شدت وقوع رویداد است (مشابه پارامتر  $\lambda$  در توزیع پواسون)،  $v \geq 0$  پارامتر کنترل پراکنندگی است که تعیین می‌کند داده‌ها بیش‌پراکنش یا کم‌پراکنش دارند و  $Z(\lambda, v) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^v}$  تابع نرمال‌ساز است.

این ساختار به مدل این امکان را می‌دهد تا با تنظیم پارامتر  $v$ ، رفتارهای مختلفی از داده‌ها را پوشش دهد.

الف. اگر  $v = 1$ ، آنگاه مدل به توزیع پواسون استاندارد تقلیل می‌یابد.

ب. اگر  $v < 1$ ، آنگاه مدل قادر به توصیف بیش‌پراکنش است.

ج. اگر  $v > 1$ ، آنگاه کم‌پراکنشی را مدل‌سازی می‌کند.

<sup>1</sup>Rectified Linear Unit

## ۱.۲.۲ محاسبه گرادیان خطا

پارامترهای شبکه عصبی با استفاده از الگوریتم پسانتشار خطا و روش‌های بهینه‌سازی مبتنی بر مشتق، آموزش داده می‌شوند. برای محاسبه گرادیان‌های لازم، از قاعده زنجیره‌ای و مشتقات جزئی تابع هدف نسبت به خروجی‌های شبکه ( $\lambda_i$  و  $v_i$ ) و سپس نسبت به وزن‌ها و اربیبی‌ها استفاده می‌شود. در نتیجه:

مشتق لگاریتم درستنمایی نسبت به  $v_i$  و  $\lambda_i$  به صورت زیر به دست می‌آید:

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = \frac{y_i}{\lambda_i} - \frac{\partial \log Z(\lambda_i, v_i)}{\partial \lambda_i}, \quad \frac{\partial \mathcal{L}}{\partial v_i} = -\log(y_i!) - \frac{\partial \log Z(\lambda_i, v_i)}{\partial v_i}$$

محاسبه مشتقات  $\partial \log Z / \partial v$  و  $\partial \log Z / \partial \lambda$  نیازمند روش‌های عددی یا تقریب‌های مناسب است که در پژوهش‌های اخیر به‌ویژه در کانگ و همکاران [۱۱] با رویکردهای بیزی و تقریب‌های سریع مورد بررسی قرار گرفته‌اند.

## ۳.۲ مدل جنگل تصادفی

یکی از روش‌هایی که در یادگیری گروهی از آن استفاده می‌شود، روش پرکاربرد جنگل‌های تصادفی است. در این روش یک مجموعه درخت تصمیم با زیرمجموعه‌ای از داده‌های اصلی به صورت موازی آموزش دیده و در نهایت دقت پیش‌گویی هر درخت میانگین‌گیری یا رای‌گیری می‌شود. جنگل تصادفی یک روش یادگیری گروهی است که توسط بریمن [۷] معرفی شد. این روش عملکرد پایدارتری نسبت به روش درخت تصمیم ارائه می‌دهد چراکه از ترکیب نتایج چند درخت تصمیم حاصل شده است و برای هر مدل رگرسیون و طبقه‌بندی قابل استفاده است و حتی می‌توان به داده‌های شمارشی نیز تعمیم داد. که در موچاگالوسا و کاکای [۱۵] اشاره شده است. در ادامه، مراحل الگوریتم جنگل تصادفی به صورت کامل بیان می‌شود.

## ۴.۲ مدل گرادیان تقویتی

مدل گرادیان تقویتی به نقل از چن و گسترین [۱۱] از الگوریتم‌های یادگیری ماشین بر پایه درخت تصمیم بوده که از مدل اولیه ضعیف‌تر استفاده کرده و آن مدل را اصلاح می‌کند. با انجام این روش به تعداد مشخص، مثلاً  $M$  بار، مدل بهتری به دست می‌آید.

## ۱.۴.۲ مراحل اصلی الگوریتم گرادیان تقویتی

۱. مدل اولیه  $Y_0(x)$

مدل اولیه  $Y_0(x)$  یک مقدار ثابت است که به عنوان پیش‌گویی اولیه برای تمام نمونه‌ها انتخاب می‌شود. این مقدار معمولاً به صورت میانگین مقادیر واقعی تعیین می‌شود.

۲. محاسبه گرادیان‌ها و مشتق‌های دوم

برای هر نمونه  $i$ ، گرادیان ( $G_i$ ) و مشتق دوم ( $H_i$ ) تابع درستنمایی ( $L$ )، نسبت به پیش‌گویی‌های مدل ( $\hat{y}_i$ ) محاسبه می‌شود. مشتق اول تابع درستنمایی، نشان‌دهنده جهت و اندازه خطای مدل برای نمونه  $i$  است و مشتق دوم تابع درستنمایی، نشان‌دهنده انعطاف‌پذیری مدل در برابر تغییرات پیش‌گویی‌ها است.

۳. تعیین وزن‌های درخت‌ها ( $W_i^*$ )

وزن‌های درخت‌ها بر اساس گرادیان‌ها و مشتق‌های دوم محاسبه می‌شوند. این وزن‌ها با توجه به پارامتر تنظیم‌کننده وزن ( $\lambda$ ) و نرخ یادگیری ( $\eta$ ) تعیین می‌شوند. پارامتر تنظیم‌کننده وزن به منظور کاهش واریانس و جلوگیری از بیش‌برازش استفاده می‌شود و  $n$  تعداد نمونه‌های داده‌ها را نشان می‌دهد.

۴. آموزش درخت جدید  $Y_m(x)$

در هر مرحله یک درخت تصمیم جدید با استفاده از گرادیان‌ها و مشتق‌های دوم آموزش داده می‌شود. هدف آموزش درخت جدید این است که خطای مدل فعلی را کاهش دهد.

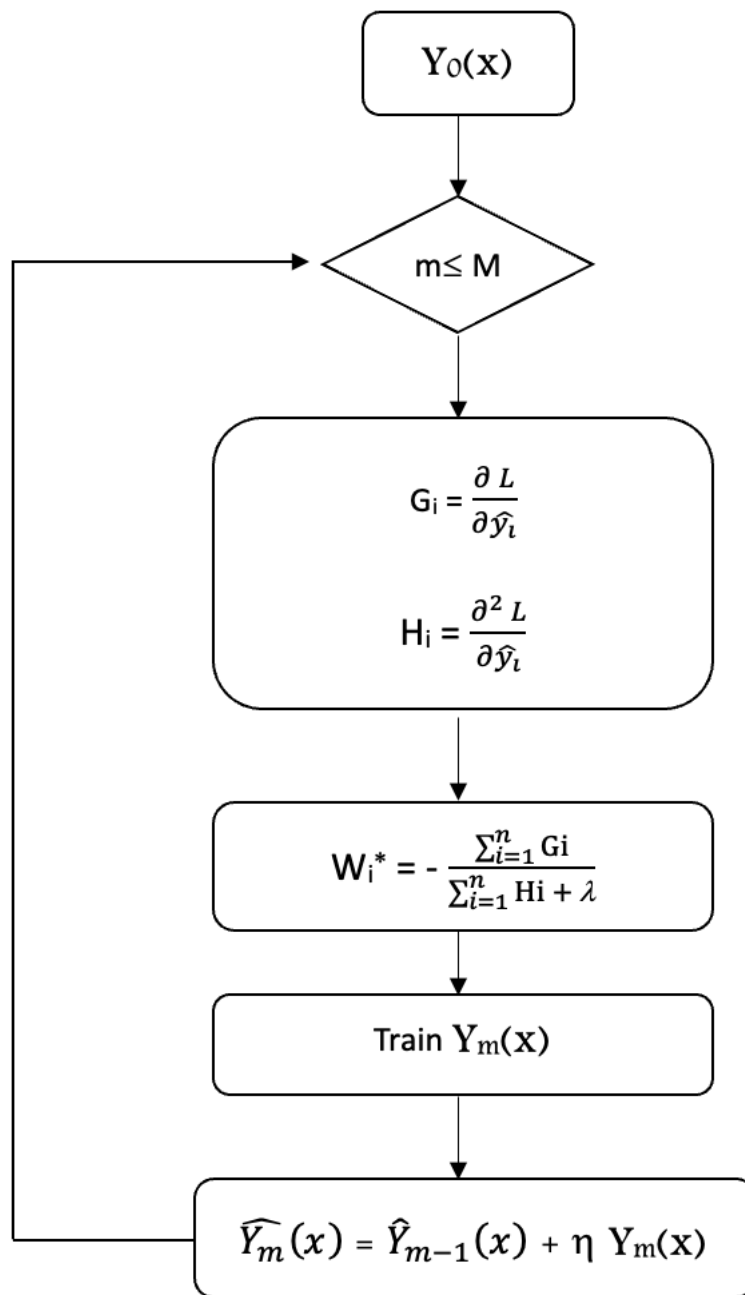
۵. به‌روزرسانی پیش‌گویی‌ها ( $\hat{Y}_m(x)$ )

پس از آموزش درخت جدید، پیش‌گویی‌ها به صورت تجمعی به‌روزرسانی می‌شوند. نرخ یادگیری ( $\eta$ ) تعیین می‌کند که در هر مرحله چقدر از پیش‌گویی‌های جدید استفاده می‌شود و  $\hat{Y}_m(x)$  پیش‌گویی نهایی مدل بعد از  $m$  مرحله را نشان می‌دهد.

۶. تکرار تا رسیدن به تعداد درخت‌های مطلوب ( $M$ )

این فرآیند تا زمانی که تعداد درخت‌های مورد نظر به دست آید، تکرار می‌شود. در هر مرحله، مدل فعلی با درخت جدید به‌روزرسانی می‌شود.

در شکل ۲، ساختار کلی الگوریتم گرادیان تقویتی نمایش داده شده است. این الگوریتم شامل مراحل اصلی مانند تعیین مدل اولیه  $Y_0(x)$ ، محاسبه گرادیان‌ها و مشتق‌های دوم، تعیین وزن‌های درخت‌ها، آموزش درخت جدید، به‌روزرسانی پیش‌گویی‌ها و تکرار تا رسیدن به تعداد درخت‌های مطلوب است.



شکل ۲: الگوریتم آموزش مدل گرادیان تقویتی

به بیمه پزشکی خصوصی و وضعیت شرایط مزمن (حاد و مزمن) هستند. متغیرهای توضیحی استاندارد شدند؛ به این صورت که میانگین از مقادیر خام کسر شد و سپس تفاوت به دست آمده بر انحراف معیار تقسیم شد. این استانداردسازی هم در مدل‌سازی با رگرسیون پواسون شبکه عصبی و هم در سایر روش‌های مدل‌سازی انجام شد. داده‌ها به‌طور تصادفی به دو بخش تقسیم شدند: مجموعه آموزشی (برای توسعه مدل) و مجموعه آزمایشی (برای ارزیابی مدل). این اقدام اطمینان حاصل

### ۳ مثال کاربردی

داده‌های مورد استفاده در این مطالعه از یک تحقیق در استرالیا به دست آمده که در بسته اقتصادسنجی کاربردی AER در نرم افزار R موجود است.

متغیرهای توضیحی شامل سن، جنسیت، درآمد خانوار، وضعیت سلامتی، تعداد بیماری‌ها، تعداد روزهای محدودیت فیزیکی، دسترسی

پیش‌گویی‌ها، با استفاده از رابطه

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2},$$

محاسبه شد.

همچنین مقایسه شبکه عصبی پواسون و شبکه عصبی کام-پواسون نیز با توجه به معیار MSE با رابطه  $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$  و معیار MAE با رابطه  $MAE = \left| \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \right|$  انجام شد.

### ۱.۳ تحلیل نتایج عددی

شبکه‌های عصبی پواسون و کام-پواسون به همراه دو مدل یادگیری ماشین شامل گرادیان تقویتی و مدل‌های جنگل تصادفی با استفاده از مجموعه آموزشی برازش داده شدند. عملکرد این مدل‌ها با استفاده از معیارهای RMSE (ریشه میانگین مربعات خطا) و  $R^2$  (ضریب تعیین) برای تحلیل داده‌های آزمایشی ارزیابی شد که نتایج در جدول ۱ و ۲ ارائه شده است.

می‌کند که مدل‌ها به درستی ارزیابی شده‌اند و قابلیت تعمیم به داده‌های جدید را دارند. در این پژوهش، برای پیاده‌سازی مدل شبکه عصبی پواسون از یک معماری با سه لایه استفاده شد. لایه ورودی شامل مقادیر متغیرهای توضیحی هستند که از این لایه ۶۴ نورون به لایه بعد منتقل می‌شود، لایه پنهان نیز با استفاده از تابع فعال‌ساز ReLU پیاده‌سازی شده است، این لایه اطلاعات را فشرده کرده به این معنا که ۳۲ نورون به لایه خروجی منتقل می‌شود. در نهایت لایه خروجی مدل تنها شامل یک نورون با تابع فعال‌ساز Softplus است.

مقایسه مدل‌ها:

برای هر سه روش مدل‌سازی استفاده شده، برای ارزیابی کلی عملکرد مدل از ضریب تعیین با فرمول

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

استفاده کردیم که در این معادله  $y_i$  مقدار واقعی و  $\hat{y}_i$  مقدار پیش‌گویی شده و  $\bar{y}$  میانگین مقادیر واقعی می‌باشد.

به علاوه، ریشه میانگین مربعات خطا را برای اندازه‌گیری دقت

جدول ۱: مقادیر  $R^2$  و RMSE برای داده‌های واقعی تعداد ویزیت‌های پزشکی

مدل	RMSE	$R^2$
گرادیان تقویتی	۰٫۷۰۶۸	۰٫۴۳۵۰
جنگل تصادفی	۰٫۶۵۷۱	۰٫۵۱۱۷
شبکه عصبی پواسون	۰٫۶۳۶۴	۰٫۵۴۲۰

جدول ۲: مقایسه عملکرد مدل‌ها بر اساس معیارهای پیش‌گویی

MSE	MAE	مدل
۰٫۶۰	۰٫۴۵	Poisson-NN
۰٫۴۲	۰٫۴۲	COM-Poisson-NN

دادند. این یافته‌ها بیانگر این هستند که شبکه عصبی پواسون به‌عنوان یک مدل پیشرفته، توانسته است عملکرد بهتری در مدل‌سازی داده‌های شمارشی نسبت به روش‌های سنتی و مدل‌های یادگیری ماشین ارائه دهد. به عبارتی هرچند گرادیان تقویتی در معیارهای دقت پیش‌گویی عملکرد قابل توجهی دارد، اما این مدل فرضیات توزیعی داده‌های شمارشی را به‌صورت صریح در نظر نمی‌گیرد و خروجی آن فاقد تفسیر

همان‌طور که در جدول ۱ مشاهده می‌شود، مدل شبکه عصبی پواسون عملکرد بهتری نسبت به سایر مدل‌ها داشت. این مدل با کمترین مقدار RMSE و بیشترین مقدار  $R^2$  توانست خطای پیش‌گویی را کاهش داده و واریانس داده‌ها را بهتر تبیین کند. مدل‌های جنگل تصادفی و گرادیان تقویتی نیز عملکرد قابل قبولی داشتند، اما در مقایسه با شبکه عصبی پواسون، دقت کم‌تری در پیش‌گویی متغیر وابسته نشان

رگرسیون شبکه عصبی کام-پواسون استفاده شد. این مدل قادر است هم پیش‌پراکنش و هم کم‌پراکنش را مدل‌سازی کند و بنابراین برای داده‌های شمارشی موردنظر با ساختار پراکندگی عمومی مناسب‌تر است.

در نهایت، مقایسه جامع عملکرد مدل‌ها نشان می‌دهد که مدل شبکه عصبی پواسون در معیارهای خطای پیش‌گویی شامل میانگین قدر مطلق خطا و میانگین مربعات خطا عملکرد ضعیف‌تری نسبت به مدل کام-پواسون دارد. از سوی دیگر، مدل شبکه عصبی پواسون عملکرد بهتری نسبت به مدل‌های یادگیری ماشینی مبتنی بر درخت تصمیم (از جمله گرادیان تقویتی و جنگل تصادفی) ارائه داد. با این حال، شبکه عصبی مبتنی بر توزیع کانوی-مکسول-پواسون نه تنها از مدل‌های شبکه عصبی پواسون و مدل‌های درختی پیشی گرفت، بلکه کمترین خطاهای پیش‌گویی را در بین تمام مدل‌های مورد ارزیابی کسب کرد. این یافته، اهمیت تطبیق ساختار آماری مدل با ویژگی‌های ذاتی داده‌ها را به‌وضوح تأیید می‌کند.

احتمالی مستقیم است. در مقابل، مدل‌های مبتنی بر توزیع مانند شبکه عصبی پواسون یا شبکه عصبی کام-پواسون ضمن تطبیق با ساختار آماری داده، امکان استنتاج احتمالی را فراهم می‌کنند. بنابراین، برتری یک مدل نسبت به دیگری باید با توجه به هدف تحلیل، پیش‌گویی نقطه‌ای یا مدل‌سازی احتمالی و نه صرفاً بر اساس خطای پیش‌گویی، قضاوت شود.

از طرفی تأکید ما بر این است که هرچند مدل شبکه عصبی پواسون انعطاف‌پذیری غیرخطی بالایی در مدل‌سازی رابطه بین متغیرهای ورودی و نرخ رویداد  $\lambda$  فراهم می‌آورد، اما همچنان تحت فرض بنیادین توزیع پواسون یعنی برابری میانگین و واریانس عمل می‌کند. بررسی داده‌های تعداد ویزیت‌های پزشکی در جدول ۳ نشان داد که میانگین برابر با ۰٫۳۰، واریانس برابر با ۰٫۶۳ و نسبت واریانس به میانگین برابر با ۲٫۱۱ است که به‌وضوح حاکی از وجود پیش‌پراکنش قابل توجه در داده‌های ما می‌باشد. در نتیجه منجر به استفاده از مدل جامع‌تر برای بررسی داده‌هایی با چالش پیش‌پراکنش شد. در نتیجه، از روش

جدول ۳: آمار توصیفی داده‌های شمارشی تعداد ویزیت‌های پزشکی

مقدار	شاخص
۰٫۳۰	میانگین
۰٫۶۳	واریانس
۲٫۱۱	نسبت واریانس به میانگین

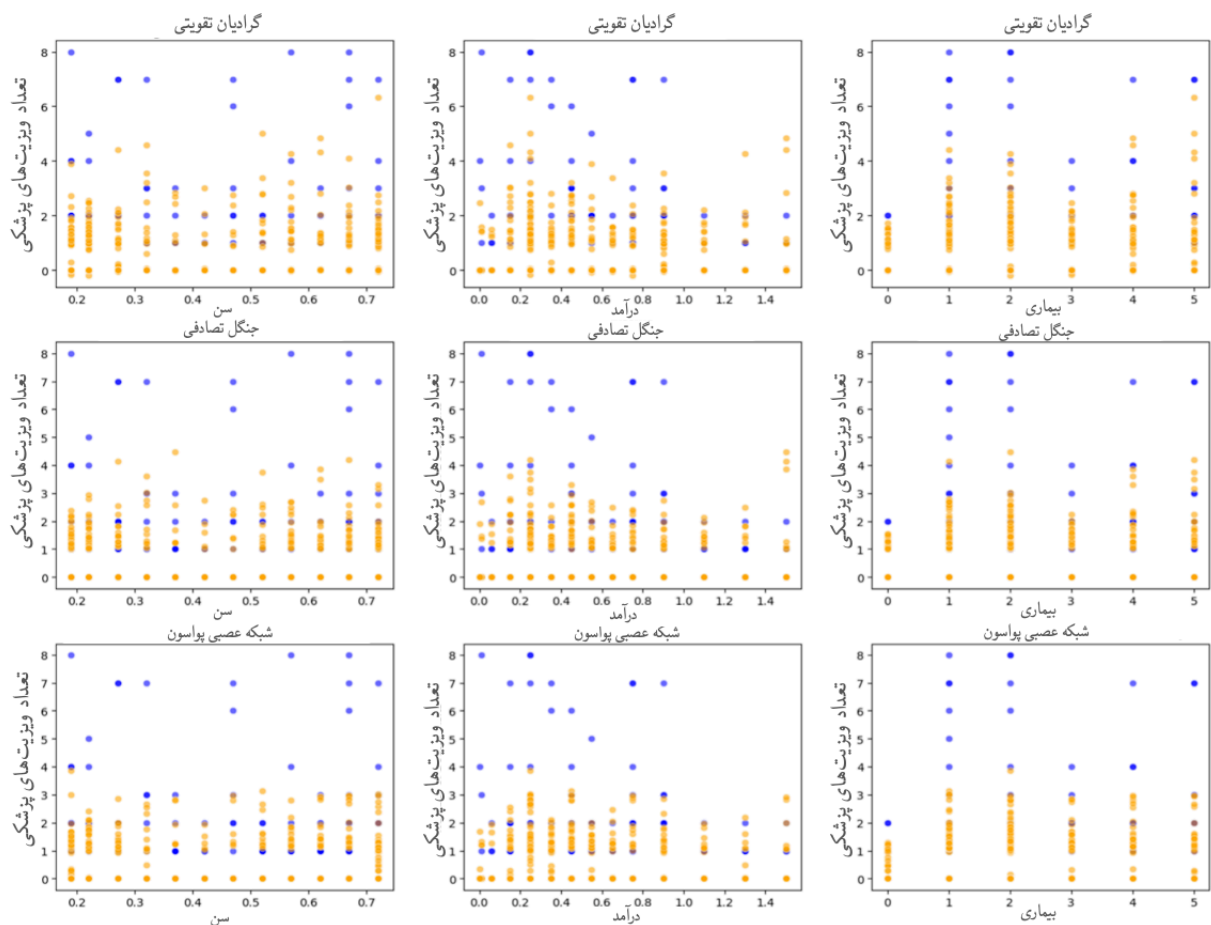
مدل خاص است و ستون‌ها نشان‌دهنده متغیر مستقل مشخص شده‌اند. این نمودارها نشان می‌دهند که چگونه مدل‌ها تعداد بازدید پزشکی را بر اساس هر یک از متغیرهای مستقل پیش‌بینی می‌کنند.

در شکل ۲، نمودارهای پراکندگی برای مقایسه روابط بین سه متغیر مستقل سن، درآمد و بیماری و تعداد بازدید پزشکی نشان داده شده است. این نمودارها برای سه الگوریتم، گرادیان تقویتی جنگل تصادفی و شبکه عصبی پواسون رسم شده‌اند. هر ردیف در شکل مربوط به یک

## ۴ بحث و نتیجه‌گیری

در این مطالعه، عملکرد شبکه‌های عصبی پواسون در مقایسه با دو مدل یادگیری ماشینی شامل گرادیان تقویتی و جنگل تصادفی برای پیش‌گویی تعداد ویزیت‌های پزشکی بررسی شد. نتایج نشان داد که شبکه عصبی پواسون با کم‌ترین مقدار ریشه میانگین مربعات خطا و بیش‌ترین مقدار  $R^2$  (ضریب تعیین)، عملکرد بهتری نسبت به سایر مدل‌ها داشت. از طرفی مشکل پراکندگی در داده‌ها با استفاده از

به صورت کلی همبستگی مثبتی در متغیر سن و بیماری دیده می‌شود یعنی با افزایش سن و داشتن بیماری تعداد مراجعات افراد به پزشک افزایش می‌یابد و همچنین همبستگی منفی در متغیر درآمد دیده می‌شود که تأیید می‌کند افراد با درآمد بالاتر کم‌تر به پزشک مراجعه می‌کنند.



شکل ۳: مقادیر پیش‌گویی تعداد ویزیت‌های پزشکی را در برابر سه متغیر مستقل نشان می‌دهند. هر ردیف به یک مدل خاص مربوط است، و هر ستون برای یک متغیر مستقل مشخص است.

رفتار آن روی داده‌های سری زمانی با وابستگی‌های زمانی و الگوهای فصلی بررسی نشده است که ممکن است منجر به خطاهای پیش‌بینی قابل توجه در این نوع داده‌ها شود. تعداد ویزیت‌های پزشکی می‌تواند به صورت داده‌های سری زمانی ثبت شود و بنابراین مدل سازی آن بدون در نظر گرفتن وابستگی‌ها و الگوهای زمانی ممکن است منجر به تخمین‌های نادرست و کاهش دقت پیش‌گویی شود. برای قضاوت دقیق‌تر لازم است مدل‌سازی در روی داده‌های متنوع نیز مورد بررسی قرار گیرد.

همچنین با توجه به اینکه داده‌های شمارشی گاهی دارای صفرهای اضافی هستند، فراوانی مشاهدات صفر نیز به عنوان آینده تحقیق مورد بررسی قرار خواهد گرفت. با این حال، نتایج نشان داد که میزان صفرها در داده‌های این مقاله با توزیع پواسون و با نرخ تخمین‌زده‌شده سازگار است و نیازی به افزودن مؤلفه‌های اضافی برای مدل‌سازی صفرهای اضافی احساس نشد.

مدل رگرسیون شبکه عصبی کانوی-مکسول-پواسون برطرف شده و در مقایسه با مدل شبکه عصبی پواسون عملکرد بهتر و خطای کمتری داشت. این یافته‌ها بیانگر قابلیت بالای شبکه عصبی کانوی-مکسول-پواسون در مدل‌سازی داده‌های شمارشی و توضیح واریانس متغیر وابسته است. همان‌طور که در نتایج مشاهده شد، متغیرهای سن و بیماری همبستگی مثبتی با تعداد ویزیت‌های پزشکی داشتند، درحالی‌که متغیر درآمد همبستگی منفی تایید می‌کند. این الگوها با یافته‌های مطالعات قبلی در حوزه سلامت مطابقت دارد. برای اطلاعات بیشتر می‌توان [۱۷] را بررسی کرد.

برای تحقیقات آینده پیشنهاد می‌شود فرآیند آموزش و تنظیم پارامترها در شبکه عصبی پواسون با الگوریتم‌های پیشرفته‌تری تنظیم شود، (مانند استفاده از روش‌های بهینه‌سازی بیزی)، همچنین می‌توان از معیارهای ارزیابی متفاوتی مانند  $L1$ -norm بهره گرفت. از طرفی مدل شبکه عصبی پواسون تنها روی داده‌های مستقل آزمایش شده و

در نهایت، این مطالعه نشان داد که شبکه عصبی پواسون می‌تواند مرحله ۲: پس انتشار به عنوان یک ابزار مناسب برای مدل‌سازی داده‌های شمارشی در حوزه سلامت و پزشکی مورد استفاده قرار گیرد.

خطای شبکه از لایه خروجی به سمت لایه ورودی منتشر می‌شود و گرادینت‌ها نسبت به وزن‌ها و اریبی‌ها محاسبه می‌شوند.

۱. خطای لایه خروجی

خطای لایه خروجی را از تابع درستنمایی مدل رگرسیون پواسون محاسبه می‌کنیم:

$$\delta^{(out)} = \frac{\partial L}{\partial Z^{(out)}} = \frac{\partial L}{\partial \hat{y}} f'(Z^{(out)})$$

در اینجا، تابع هدف مدل رگرسیون پواسون به صورت زیر تعریف می‌شود:

$$L = \sum_{i=1}^N [y_i \log(\hat{y}_i) - \hat{y}_i]$$

که در آن  $y_i$  مقدار واقعی و  $\hat{y}_i$  مقدار پیش‌گویی شده است. مشتق تابع درستنمایی نسبت به  $\hat{y}$  برابر است با:

$$\frac{\partial L}{\partial \hat{y}} = \frac{y}{\hat{y}} - 1$$

همچنین مشتق تابع فعال‌سازی Softplus( $f'(Z^{(out)})$ ) برابر است با:

$$f'(Z^{(out)}) = \frac{1}{1 + e^{-Z^{(out)}}}$$

بنابراین خطای لایه خروجی به صورت زیر محاسبه می‌شود:

$$\delta^{(out)} = \left(\frac{y}{\hat{y}} - 1\right) \left(\frac{1}{1 + e^{-Z^{(out)}}}\right)$$

۲. خطای لایه پنهان

برای هر لایه پنهان  $h$ ، خطای لایه  $(\delta^{(h)})$  از لایه بعدی  $(\delta^{(h+1)})$  به دست می‌آید:

$$\delta^{(h)} = \frac{\partial L}{\partial Z^{(h)}} = \frac{\partial L}{\partial Z^{(h+1)}} \cdot \frac{\partial Z^{(h+1)}}{\partial a^{(h)}} \cdot \frac{\partial a^{(h)}}{\partial Z^{(h)}}$$

مشتق تابع فعال‌سازی ReLU  $\left(\frac{\partial a^{(h)}}{\partial Z^{(h)}}\right)$  عبارت است از:

$$\frac{\partial a^{(h)}}{\partial Z^{(h)}} = \begin{cases} 1 & \text{if } Z^{(h)} > 0 \\ 0 & \text{if } Z^{(h)} \leq 0 \end{cases}$$

۳. بروزرسانی پارامترها

پس از محاسبه خطاهای هر لایه، گرادینت‌ها نسبت به وزن‌ها و اریبی‌ها محاسبه می‌شوند:

## تقدیر و تشکر

از داوران محترم که با دقت مقاله را ارزیابی و نظرات ارزشمندی ارائه دادند، صمیمانه تشکر می‌شود. پیشنهادات ایشان منجر به اصلاحات اساسی در متن شد و سهم بسزایی در ارتقای دقت، وضوح و کیفیت کلی این پژوهش داشت.

## پیوست

### الگوریتم محاسبه گرادینت خطا

مرحله ۱: انتشار پیشرو

۰.۱ برای هر لایه پنهان  $h$ :

خروجی خطی لایه  $h$  را محاسبه می‌کنیم:

$$Z^{(h)} = \sum_{i=1}^n (w_i^{(h)} \cdot x_i + b_i^{(h)})$$

خروجی فعال‌سازی لایه  $h$  را با استفاده از تابع فعال‌ساز ReLU

محاسبه می‌کنیم:

$$a^{(h)} = \text{ReLU}(Z^{(h)}) = \max(0, Z^{(h)})$$

۰.۲ برای لایه خروجی:

خروجی خطی لایه خروجی را محاسبه می‌کنیم:

$$Z^{(out)} = \sum_{i=1}^n (w_i^{(out)} \cdot a_i^{(h)} + b_i^{(out)})$$

خروجی فعال‌سازی لایه خروجی out را با استفاده از تابع

فعال‌سازی Softplus محاسبه می‌کنیم:

$$\hat{y} = \text{Softplus}(Z^{(out)}) = \log(1 + e^{Z^{(out)}})$$

$\hat{y}$  نشان‌دهنده مقدار پیش‌گویی شده توسط مدل است. به عبارت

دیگر،  $\hat{y}$  خروجی شبکه عصبی پس از اعمال تابع فعال‌سازی در لایه خروجی است که به عنوان تخمینی برای متغیر وابسته  $y$  در نظر گرفته می‌شود.

$$w^{(h)} \leftarrow w^{(h)} - \eta \frac{\partial L}{\partial w^{(h)}}$$

$$b^{(h)} \leftarrow b^{(h)} - \eta \frac{\partial L}{\partial b^{(h)}}$$

$$\frac{\partial L}{\partial w^{(h)}} = \delta^{(h)} \cdot a^{(h-1)}$$

$$\frac{\partial L}{\partial b^{(h)}} = \delta^{(h)}$$

سپس پارامترهای شبکه (وزن‌ها و اریبی‌ها) با استفاده از یک الگوریتم بهینه‌سازی به‌روزرسانی می‌شوند:

که در آن  $\eta$  نرخ یادگیری است.

## مراجع

- [۱] مددی، م.، و متوسلی، ک. (۱۴۰۳). تحلیل داده‌های بقا با روش‌های مختلف یادگیری آماری. *مجله علوم آماری*، ۱۸(۲)، ۴۱۵-۴۳۴.
- [2] Ashraf, M., Zaman, M., & Ahmed, M. (2020). An intelligent prediction system for educational data mining based on ensemble and filtering approaches. *Procedia Computer Science*, **167**, 1471-1483.
- [3] Breiman, L. (2001). Random forests. *Machine Learning*, **45**(1), 5-32.
- [4] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, August 13-17, pp. 785-794.
- [5] Eftekhar, B., Mohammad, K., Ardebili, H., Ghodsi, M., & Ketabchi, E. (2005). Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC Medical Informatics and Decision Making*, **5**, 3.
- [6] Fallah, N., Gu, H., Mohammad, K., Seyyedsalehi, S. A., Nourijelyani, K., & Eshraghian, M. R. (2009). Nonlinear Poisson regression using neural networks: A simulation study. *Neural Computing and Applications*, **18**(8), 945-954.
- [7] Fallah, N., Mitnitski, A., & Rockwood, K. (2011). Applying neural network Poisson regression to predict cognitive score changes. *Journal of Applied Statistics*, **38**(9), 2051-2062.
- [8] Fan, J., Ma, C., & Zhong, Y. (2019). A selective overview of deep learning. arXiv preprint arXiv:1904.06981.
- [9] Hamasaki, H. (2025). The relationship between socioeconomic status and health behaviors in older adults: A narrative review. *Healthcare*, **13**(14), 1669.
- [10] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Sanya, China, April 23-26, pp. 1026-1034.
- [11] Kang, B., Hughes, J., & Haran, M. (2025). Fast Bayesian inference for spatial mean-parameterized Conway–Maxwell–Poisson models. *Journal of Computational and Graphical Statistics*, **34**(2), 697-706.
- [12] Lee, H., Ha, I. D., & Lee, Y. (2023). Deep neural networks for semiparametric frailty models via H-likelihood. arXiv preprint arXiv:2307.06581.

- [13] Montesinos-Lopez, O. A., Montesinos-Lopez, J. C., Salazar, E., Barron, J. A., Montesinos-Lopez, A., Buenrostro-Mariscal, R., & Crossa, J. (2021). Application of a Poisson deep neural network model for the prediction of count data in genome-based prediction. *The Plant Genome*, **14**(2), e20118.
- [14] Montesinos-López, O. A., Montesinos-López, J. C., Crossa, J., Martini, J. F., Burgueño, J., & Singh, R. (2022). Comparing gradient boosting machine and Bayesian threshold BLUP for genome-based prediction of categorical traits in wheat breeding. *The Plant Genome*, **15**(1), e20214.
- [15] Mushagalusa, C. A., & Kakaï, R. G. (2022). Random forests in count data modelling: An analysis of the influence of data features and overdispersion on regression performance. *Journal of Probability and Statistics*, 2833537.
- [16] Nachman, B., & Thaler, J. (2021). Pluribus unum ex machina: Learning from many collider events at once. arXiv preprint arXiv:2104.03290.
- [17] Ragab, M., Abdel Aal, A. K., Jifri, A. O., & Omran, N. F. (2021). Enhancement of predicting students' performance model using ensemble approaches and educational data mining techniques. *Wireless Communications and Mobile Computing*, **2021**, 6241676.
- [18] Rizvi, S., Pettee, M., & Nachman, B. (2024). Learning likelihood ratios with neural network classifiers. *Journal of High Energy Physics*, **2024**(2), 136.
- [19] Rodrigo, H., & Tsokos, C. (2020). Bayesian modelling of nonlinear Poisson regression with artificial neural networks. *Journal of Applied Statistics*, **47**(5), 757-774.
- [20] Rodrigo, H. S. (2017). Bayesian artificial neural networks in health and cybersecurity. PhD thesis, Tampa, FL: University of South Florida.
- [21] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, **323**(6088), 533-536.
- [22] Sellers, K. F., Borle, S., & Shmueli, G. (2012). The COM-Poisson model for count data: A survey of methods and applications. *Applied Stochastic Models in Business and Industry*, **28**(2), 173-191.
- [23] Sellers, K. F., Swift, A. W., & Weems, K. S. (2017). A flexible distribution class for count data. *Journal of Statistical Distributions and Applications*, **4**(1), 22.
- [24] Sharaf, T. (2015). Statistical learning with artificial neural network applied to health and environmental data. PhD thesis, Tampa, FL: University of South Florida.
- [25] Sharaf, T., & Tsokos, C. P. (2015). Two artificial neural networks for modeling discrete survival time of censored data. *Advances in Artificial Intelligence*, **2015**, 270165.
- [26] Xu, X., Ye, T., Gao, J., et al. (2024). Generalized hurdle count data models based on interpretable machine learning with an application to health care demand. *Computing*, **106**, 295-325.

# **Modeling the Number of Medical Visits Using Conway–Maxwell–Poisson Neural Network Regression**

**T. Torabi Neman<sup>1\*</sup>, M. Emadi<sup>2</sup> and M. Arashi<sup>3</sup>**

<sup>1,2,3</sup> Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Iran

## **Abstract:**

Since many empirical and scientific analyses are based on count data, modeling such data is of great importance. Examples of these analyses include regression of the number of doctor visits based on a set of predictor variables. Obviously, access to an appropriate predictive regression model can have a significant impact on decision-making and planning. In this research, aiming to improve modeling accuracy, Poisson regression has been combined with an artificial neural network to provide a better modeling framework for count data. In this regard, a Poisson neural network regression model along with a computational algorithm is presented. To evaluate its performance in terms of parameter estimation and predictive accuracy, it is compared with two ensemble machine learning models: Random Forest and Gradient boosting. On the other hand, to investigate the overdispersion phenomenon, the Conway–Maxwell–Poisson neural network model was used and compared with the Poisson neural network. By implementing the aforementioned models on a real medical dataset, we demonstrate that the Conway–Maxwell–Poisson neural network regression model exhibits superior performance.

**Keywords:** Poisson Regression, Conway–Maxwell–Poisson Regression, Neural Networks, Multilayer Perceptron, Count Data, Random Forest, XGBoost.