

مدل رگرسیون پواسون صفر آماسیده به کمک یادگیری عمیق

نویسنده : سپهر نادری

استاد راهنما : دکتر زهرا صابری



۱. داده‌های شمارشی.....	۵
۱.۱ تعریف و ویژگی‌ها.....	۵
۱.۲ توزیع پواسون.....	۵
۱.۳ رگرسیون پواسون.....	۵
۱.۴ بیش‌پراکنش و صفرهای اضافی.....	۶
۲. شبکه‌های عصبی.....	۶
۲.۱ تعریف.....	۷
۲.۲ توان تقریب.....	۷
۲.۳ شبکه‌ی توزیعی و مسئله‌ی جعبه‌ی سیاه.....	۷
۳. پرسپترون چندلایه (MLP).....	۷
۳.۱ معماری.....	۷
۳.۲ توابع فعال‌سازی.....	۸
۳.۳ آموزش و پس‌انتشار خطا.....	۸
۳.۴ مشتقات مرتبه‌ی دوم و عملگر.....	۹
۴. رگرسیون پواسون آماسیده در صفر (ZIP).....	۹
۴.۱ پیشینه.....	۹
۴.۲ ساختار آمیخته.....	۹
۴.۳ گشتاورها.....	۹
۴.۴ رگرسیون و برآورد.....	۱۰
۴.۵ ZIP و مدل مانع.....	۱۰
۴.۶ محدودیت.....	۱۱
۵. شبکه عصبی پواسون آماسیده در صفر (ZIP-NN).....	۱۱
۵.۱ تعریف.....	۱۱

۵.۲	تابع زیان.....	۱۱
۵.۳	گرادیان و ارتباط با EM.....	۱۲
۵.۴	پیش‌بینی.....	۱۲
۵.۵	آموزش و ارزیابی.....	۱۲
۵.۶	نتایج.....	۱۲
۶	چارچوب آماری دقیق مدل NN-ZIP و بازسازی ریاضی آن.....	۱۳
۶.۱	امید ریاضی، واریانس و شاخص پراکندگی.....	۱۴
۶.۲	ساختار شبکه و تعریف NN-ZIP.....	۱۶
۶.۳	تابع درست‌نمایی و تابع زیان.....	۱۷
۶.۴	گرادیان‌های صریح و رابطه با EM.....	۱۸
۶.۵	شناسایی پارامترها و مسائل پایداری.....	۱۹
۷	طراحی پیشنهادی مطالعه و پروتکل ارزیابی.....	۲۰
۷.۱	مدل‌های مبنا و مقایسه‌ی منصفانه.....	۲۱
۷.۲	معیارهای ارزیابی: از MAE و RMSE تا معیارهای توزیعی.....	۲۲
۸	تحلیل نتایج موجود و نقد آماری آن‌ها.....	۲۲
۸.۱	یک نکته‌ی مهم درباره‌ی داده‌ی واقعی مقالات دانشجویان دکتری.....	۲۵
۹	تفسیرپذیری NN-ZIP و خروج از جعبه‌ی سیاه.....	۲۵
۹.۱	تحلیل اثر حاشیه‌ای در NN-ZIP.....	۲۶
۱۰	کنترل بیش‌برازش و انتخاب ابرپارامترها.....	۲۶
۱۱	الگوریتم پیشنهادی برای آموزش NN-ZIP.....	۲۷
۱۲	آزمون‌های تکمیلی.....	۲۸
۱۲.۱	سنجش اهمیت صفرآماسیدگی.....	۲۸
۱۳	مقایسه‌ی ZIP، ZINB و Hurdle: چه زمانی NN-ZIP انتخاب درستی است؟.....	۲۹
۱۴	پیوست A: فرمول‌های کلیدی.....	۲۹
	مراجع.....	۳۰

مبانی نظری و مفاهیم بنیادی

این فصل یک زنجیره‌ی استدلالی را دنبال می‌کند. داده‌های شمارشی را پواسون مدل می‌کند. پواسون فرض هم‌پراکنشی سخت‌گیرانه‌ای دارد که در حضور صفرهای اضافی نقض می‌شود. مدل پواسون آماسیده در صفر (ZIP) این نقض را با ساختاری آمیخته جبران می‌کند، اما پیش‌گوهایش خطی‌اند. شبکه‌های عصبی این محدودیت را برمی‌دارند، اما در حالت خالص چارچوب احتمالی ندارند. شبکه عصبی پواسون آماسیده در صفر (ZIP-NN) این دو را یکجا گرد می‌آورد.

۱. داده‌های شمارشی

۱.۱. تعریف و ویژگی‌ها. متغیر پاسخ را «شمارشی» می‌نامیم هرگاه تعداد وقوع یک رویداد گسسته را در یک بازه‌ی زمانی، ناحیه‌ی مکانی یا واحد آماری معین اندازه بگیرد، یعنی. این داده‌ها در اقتصاد، اپیدمیولوژی، مهندسی و علوم اجتماعی فراوان‌اند، مانند تعداد مقالات منتشرشده‌ی دانشجویان دکتری. سه ویژگی مدل‌سازی آن‌ها با رگرسیون خطی گاوسی را نامناسب می‌کند. مقادیر گسسته و نامنفی‌اند و پیش‌بینی منفی بی‌معناست. توزیعشان معمولاً چولگی به راست دارد. واریانس هم با میانگین افزایش می‌یابد.

۱.۲. توزیع پواسون. توزیع پایه برای داده‌های شمارشی، پواسون با پارامتر است:

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, 2, \dots, \quad \mathbb{E}(Y) = \text{Var}(Y) = \lambda.$$

برابری میانگین و واریانس («هم‌پراکنشی»، Equidispersion) ویژگی تعیین‌کننده‌ی پواسون و درعین حال ضعیف‌ترین حلقه‌ی آن در کاربردهای واقعی است. احتمال صفر نیز کاملاً به میانگین گره خورده است، $P(Y = 0) = e^{-\lambda}$. برای مثال، برای $\lambda = 2$ فقط حدود ۱۳/۵٪ مشاهدات صفر انتظار می‌رود.

۱.۳. رگرسیون پواسون. در رگرسیون پواسون (Poisson Regression)، و میانگین از طریق پیوند لگاریتمی (پیوند کانونی) به پیش‌گوها مرتبط می‌شود:

$$\ln \lambda_i = x_i^T \beta \Rightarrow \mathbb{E}[Y_i | x_i] = \lambda_i = e^{x_i^T \beta}.$$

پیوند لگاریتمی مثبت بودن میانگین را تضمین و اثرها را ضربی می‌کند. e^{β_j} نسبت نرخ (Rate Ratio) به ازای یک واحد افزایش x_{ij} است. لگاریتم درست‌نمایی $\ell(\beta) = \sum_i [y_i x_i^T \beta - e^{x_i^T \beta} - \ln y_i!]$ در صورت رتبه کامل بودن ماتریس طرح اکیداً مقعر است. معادلات نمره‌ی $\sum_i (y_i - \lambda_i) x_i = 0$ معمولاً با حداقل مربعات باوزن‌دهی شده تکراری (IRLS) حل می‌شوند (Cameron & Trivedi, ۲۰۱۳).

۱.۴. بیش‌پراکنش و صفرهای اضافی. اگر باشد، بیش‌پراکنش (Overdispersion) داریم. دو منبع رایج آن ناهمگنی مشاهده‌نشده (که به مدل دو جمله‌ای منفی می‌انجامد) و سهم بزرگ صفرهاست. «صفر اضافی» همواره نسبت به مدل برازش شده تعریف می‌شود، نه به صورت مطلق. ۳۰٪ صفر برای با پواسون سازگار است (۰)، ولی برای (احتمال صفر حدود ۵٪) نشانه‌ی صفر اضافی است.

صفرها دو منشأ متفاوت دارند:

- (الف) **صفر ساختاری:** واحدهایی که به دلایل بنیادی، عدم تمایل یا عدم دسترسی اساساً در معرض رویداد نیستند و صفرشان قطعی است.
 - (ب) **صفر تصادفی (نمونه‌ای):** واحدهایی که در معرض رویداد بوده‌اند اما در بازه‌ی مشاهده اتفاقاً رویدادی ثبت نکرده‌اند. این صفرها از خود فرایند پواسون می‌آیند.
- رگرسیون پواسون این دو را تفکیک نمی‌کند و همه‌ی صفرها را به یک فرایند نسبت می‌دهد. پیامد آن سه‌گانه است:

۱. احتمال صفر کمتر از مقدار واقعی برآورد می‌شود و توزیع پیش‌بینی شده با داده همخوان نیست.
۲. بیش‌پراکنش ناشی از صفرهای اضافی خطاهای معیار را کوچک‌نمایی می‌کند و فاصله‌های اطمینان و آزمون‌ها را نامعتبر می‌سازد.
۳. برآوردگر (شبه) درست‌نمایی پواسون فقط وقتی سازگار است که میانگین شرطی درست مشخص شده باشد. در حضور صفر اضافی $\mathbb{E}[Y | x] = (1 - \pi(x))\lambda(x)$ است و اگر π به x وابسته باشد، این تابع شکل $e^{x^T \beta}$ ندارد. در نتیجه برآورد به مقداری میل می‌کند که با پارامتر واقعی فرایند شمارش فرق دارد (تورش). اگر π ثابت باشد، فقط عرض از مبدأ جابه‌جا می‌شود و شیب‌ها سازگار می‌مانند.

۲. شبکه‌های عصبی

۲.۱. تعریف. شبکه‌ی عصبی پیشخور خانواده‌ی پارامتری از توابع است که از ترکیب نگاشت‌های آفین و توابع فعال‌سازی غیرخطی ساخته می‌شود. پارامترهای آن (θ) با کمینه‌سازی یک تابع زیان و به روش‌های مبتنی بر گرادیان از روی داده برآورد می‌گردند. از منظر آماری، شبکه‌ی عصبی یک مدل رگرسیونی غیرخطی و بسیار پرپارامتر است که مدل‌های خطی تعمیم‌یافته را به‌عنوان حالت خاص در بر می‌گیرد. شبکه‌ای بدون لایه‌ی پنهان که خروجی‌اش از وارون تابع پیوند عبور کند (برای پواسون، تابع نمایی) دقیقاً یک مدل خطی تعمیم‌یافته (GLM) است.

۲.۲. توان تقریب. انعطاف‌پذیری شبکه‌ها پشتوانه‌ی نظری دارد. یک شبکه‌ی پیشخور با یک لایه‌ی پنهان و تابع فعال‌سازی غیرچندجمله‌ای، با تعداد کافی نرون، می‌تواند هر تابع پیوسته روی یک مجموعه‌ی فشرده را با دقت دلخواه تقریب بزند (Cybenko, ۱۹۸۹; Hornik et al., ۱۹۸۹; Leshno et al., ۱۹۹۳). این قضیه وجودی است. نه می‌گوید چند نرون لازم است، نه تضمین می‌کند که بهینه‌سازی روی نمونه‌ی متناهی به آن تقریب برسد. به همین دلیل ارزیابی بیرون‌نمونه و کنترل پیچیدگی ناگزیر است.

۲.۳. شبکه‌ی توزیعی و مسئله‌ی جعبه‌ی سیاه. شبکه‌ی خالص را به‌دلیل ابهام پارامترها «جعبه‌ی سیاه» می‌نامند. در ZIP-NN شبکه پاسخ را مستقیماً پیش‌بینی نمی‌کند، بلکه پارامترهای یک توزیع شرطی را خروجی می‌دهد و آموزش با بیشینه‌سازی درست‌نمایی همان توزیع انجام می‌شود. این ایده هم‌خانواده‌ی شبکه‌های چگالی آمیخته (Bishop, ۱۹۹۴) است.

آنچه از این راه حفظ می‌شود تفسیرپذیری در سطح توزیع است. π_i هم‌چنان احتمال صفر ساختاری و λ_i هم‌چنان شدت فرایند شمارش است. این به‌معنای تفسیرپذیری ضرایب نیست. نگاشت $(\pi, \lambda) \mapsto x$ غیرخطی است و وزن‌های شبکه به‌دلیل تقارن‌های جایگشتی نرون‌های پنهان یکتا نیستند. بنابراین استنباط روی وزن‌ها بی‌معناست و تفسیر باید روی توابع $\pi(x)$ ، $\lambda(x)$ و $\mu(x) = (1 - \pi(x))\lambda(x)$ انجام شود، مثلاً با نمودارهای وابستگی جزئی.

۳. پرسپترون چندلایه (MLP)

۳.۱. معماری. پرسپترون چندلایه (MLP) یک شبکه‌ی پیشخور کاملاً متصل با لایه‌ی پنهان است:

$$h^{(0)} = x, \quad u^{(k)} = W^{(k)} h^{(k-1)} + b^{(k)}, \quad h^{(k)} = \phi(u^{(k)}), \quad k = 1, \dots, K,$$

که در آن $W^{(k)}$ ماتریس وزن‌ها، $b^{(k)}$ بردار بایاس (عرض از مبدأ) و ϕ تابع فعال‌سازی نقطه‌به‌نقطه است. در این متن «بایاس» برای پارامتر شبکه و «تورش» برای مفهوم آماری به کار رفته تا خلط نشود. لایه‌ی خروجی پیش‌ساخت‌های خطی $u^{(o)} = W^{(o)}h^{(K)} + b^{(o)}$ را می‌سازد و تابع فعال‌سازی خروجی آن‌ها را به دامنه‌ی مجاز پارامتر هدف می‌نگارد.

۳.۲. توابع فعال‌سازی.

تابع ضابطه برد نقش در مدل

واحد خطی اصلاح‌شده (ReLU)	$\max(0, u)$	$[0, \infty)$	لایه‌ی پنهان
تانژانت هایپربولیک (tanh)	$\tanh u$	$(-1, 1)$	لایه‌ی پنهان
تابع سیگموئید	$\sigma(u) = \frac{1}{1 + e^{-u}}$	$(0, 1)$	خروجی π
تابع سافت‌پلاس (softplus)	$\zeta(u) = \ln(1 + e^u)$	$(0, \infty)$	خروجی λ
تابع نمایی	e^u	$(0, \infty)$	خروجی λ (جایگزین)

softplus تقریبی هموار از ReLU است و $\zeta'(u) = \sigma(u)$. تابع نمایی همان پیوند لگاریتمی رگرسیون پواسون است و در حالت بدون لایه‌ی پنهان مدل کلاسیک را بازتولید می‌کند. رشد سریع آن می‌تواند ناپایداری عددی بسازد، در حالی که softplus برای مقادیر بزرگ خطی رشد می‌کند و پایدارتر است.

۳.۳. آموزش و پس‌انتشار خطا. آموزش دو گام دارد. انتشار پیشرو خروجی و زیان را می‌دهد و پس‌انتشار خطا (Rumelhart et al., ۱۹۸۶) گرادیان‌ها را با قاعده‌ی زنجیره‌ای محاسبه می‌کند. با تعریف و قرارداد و داریم:

$$\delta^{(k)} = (W^{(k+1)\top} \delta^{(k+1)}) \odot \phi'(u^{(k)}), \quad \frac{\partial \mathcal{L}}{\partial W^{(k)}} = \delta^{(k)} h^{(k-1)\top}, \quad \frac{\partial \mathcal{L}}{\partial b^{(k)}} = \delta^{(k)}.$$

به‌روزرسانی با گرادیان کاهشی انجام می‌شود که نرخ یادگیری پارامتر آن است. در پیاده‌سازی‌های معاصر NN-ZIP، استفاده از بهینه‌سازی‌های گرادیان مبنا مانند Nadam یا Adam رایج است؛ در مطالعه‌ی Jose و همکاران (۲۰۲۴) از Nadam همراه با توقف زود هنگام و dropout استفاده شده است. در این قسمت نیز Nadam به‌عنوان انتخاب مرجع روش‌شناختی گزارش می‌شود. از آنجا که سطح زیان نامحدب است، مقداردی‌اولیه، نرخ یادگیری و تنظیم‌سازی می‌توانند بر جواب نهایی اثر بگذارند و کمینه‌ی سراسری تضمین نمی‌شود.

۳.۴. مشتقات مرتبه‌ی دوم و عملگر. ماتریس هسین برای شبکه‌ای با پارامتر است و ساخت، ذخیره و وارونه‌سازی صریح آن در ابعاد بالا عملی نیست. عملگر

$$\mathcal{R}_v[f(\theta)] = \frac{\partial}{\partial \varepsilon} f(\theta + \varepsilon v) \Big|_{\varepsilon=0}$$

عملگر هسین-بردار HVP از نظر محاسباتی مهم است، در پیاده‌سازی‌های مورد استناد، تابع هدف با منفی لگاریتم درست‌نمایی و روش‌های مرتبه‌ی اول مانند Nadam کمینه می‌شود. بنابراین HVP در این پژوهش بیشتر به‌عنوان ابزار تحلیل انحنا، gradient-check پیشرفته و مسیر پژوهشی برای آموزش مرتبه‌ی دوم معرفی می‌شود، نه به‌عنوان یک مرحله‌ی الزامی در آموزش مدل.

۴. رگرسیون پواسون آماسیده در صفر (ZIP)

۴.۱. پیشینه. مدل پواسون آماسیده در صفر (Zero-Inflated Poisson; ZIP)، یعنی مدل شمارشی آمیخته‌ای که در آن بخشی از صفرها ساختاری‌اند را Lambert (۱۹۹۲) برای تحلیل نقص‌های تولید صنعتی معرفی کرد. پس از آن Böhning و همکاران (۱۹۹۹) آن را در اپیدمیولوژی دندان، Hall (۲۰۰۰) با افزودن اثرهای تصادفی، و Ridout و همکاران (۲۰۰۱) در آزمون مدل ZIP در برابر مدل دوجمله‌ای منفی آماسیده در صفر (ZINB) گسترش دادند.

۴.۲. ساختار آمیخته. توزیع پواسون آماسیده در صفر (ZIP) بر پایه‌ی دو فرایند مستقل است. متغیر پنهان را در نظر بگیرید که یعنی واحد صفر ساختاری است. اگر باشد، قطعی است. اگر باشد، است و می‌تواند صفر تصادفی یا مقدار مثبت بدهد. در نتیجه:

$$P(Y_i = y | x_i) = \begin{cases} \pi_i + (1 - \pi_i)e^{-\lambda_i}, & y = 0 \\ (1 - \pi_i)\frac{e^{-\lambda_i}\lambda_i^y}{y!}, & y = 1, 2, \dots \end{cases}$$

۴.۳. گشتاورها. با محاسبه‌ی مستقیم:

$$\mathbb{E}(Y_i) = (1 - \pi_i)\lambda_i, \quad \text{Var}(Y_i) = (1 - \pi_i)\lambda_i(1 + \pi_i\lambda_i).$$

پس شاخص پراکندگی $\text{Var}/\mathbb{E} = 1 + \pi_i\lambda_i$ است و برای $\pi_i > 0$ همواره بزرگ‌تر از یک است. ZIP ذاتاً بیش‌پراکنده است. افزون بر این، چون $t \mapsto e^{-t}$ محدب است، نامساوی ینسن می‌دهد

$$e^{-(1-\pi_i)\lambda_i} \leq \pi_i + (1 - \pi_i)e^{-\lambda_i}.$$

یعنی احتمال صفر در ZIP هیچ‌گاه کمتر از احتمال صفر پواسونِ هم‌میانگین نیست و فقط برای $\pi_i = 0$ برابر می‌شود. این همان معنای دقیق «صفر اضافی» است.

۴.۴. رگرسیون و برآورد. در ZIP کلاسیک هر دو پارامتر با پیش‌گوی خطی مدل می‌شوند:

$$\ln \lambda_i = x_i^T \beta, \quad \text{logit}(\pi_i) = \ln \frac{\pi_i}{1 - \pi_i} = x_i^T \gamma.$$

در حالت کلی می‌توان برای دو جزء مجموعه‌ی متفاوتی از متغیرها به کار برد. γ تعیین می‌کند چه واحدی صفر ساختاری است و β شدت شمارش را در واحدهای فعال تعیین می‌کند. لگاریتم درست‌نمایی برابر است با

$$\begin{aligned} \ell(\beta, \gamma) &= \sum_i \mathbb{I}_{y_i = 0} [\ln(\pi_i + (1 - \pi_i)e^{-\lambda_i})] + \sum_i \mathbb{I}_{y_i > 0} [\ln(1 - \pi_i) - \lambda_i + y_i \ln \lambda_i - \ln y_i!]. \end{aligned}$$

به‌دلیل لگاریتمِ مجموع در جمله‌ی صفرها، این تابع عموماً مقعر نیست و ممکن است بیش از یک بیشینه‌ی موضعی داشته باشد. برآورد معمولاً با الگوریتم EM یا الگوریتم امید-بیشینه‌سازی (Expectation-Maximization; Dempster et al., ۱۹۷۷) انجام می‌شود. اگر Z_i مشاهده‌پذیر بود، لگاریتم درست‌نمایی کامل به دو جزء جداگانه تفکیک می‌شد:

$$\ell_c = \sum_i \{z_i \ln \pi_i + (1 - z_i) \ln(1 - \pi_i)\} + \sum_i (1 - z_i) [y_i \ln \lambda_i - \lambda_i - \ln y_i!].$$

گام E مقدار Z_i را با امید شرطی‌اش جایگزین می‌کند:

$$\hat{z}_i = \mathbb{E}[Z_i | y_i] = \begin{cases} \frac{\pi_i}{\pi_i + (1 - \pi_i)e^{-\lambda_i}}, & y_i = 0 \\ 0, & y_i > 0. \end{cases}$$

در گام M، جزء π یک رگرسیون لجستیک با پاسخ‌های کسری $\hat{Z}_i$ و جزء λ یک رگرسیون پواسون با وزن‌های $1 - \hat{Z}_i$ است.

۴.۵. ZIP و مدل مانع. در مدل مانع (Mullahy, ۱۹۸۶; Hurdle) همه‌ی صفرها از جزء دوتایی می‌آیند و مقادیر مثبت از پواسون صفر-بریده. در مدل ZIP صفرها هم از جزء ساختاری و هم از جزء پواسون می‌آیند. انتخاب میان این دو باید بر پایه‌ی معنای فرایند تولید داده باشد.

۴.۶. محدودیت. محدودیت اصلی ZIP کلاسیک خطی بودن پیش‌گوها در مقیاس پیوند است. هر اثر غیرخطی یا برهم‌کنش باید دستی مشخص شود (چندجمله‌ای، اسپلاین، جمله‌ی برهم‌کنش)، و در غیر این صورت خطای مشخص‌سازی رخ می‌دهد. وقتی ابعاد بالا باشد یا شکل رابطه ناشناخته باشد، این کار غیرعملی است. ZIP-NN همین شکاف را هدف می‌گیرد.

۵. شبکه عصبی پواسون آماسیده در صفر (ZIP-NN)

۵.۱. تعریف. ZIP-NN مدلی است که پارامترهای ZIP را یک شبکه‌ی عصبی خروجی می‌دهد:

$$\lambda_i = \lambda(x_i; \theta), \quad \pi_i = \pi(x_i; \theta), \quad Y_i | x_i \sim \text{ZIP}(\pi_i, \lambda_i).$$

مدل ZIP کلاسیک حالت خاصی از این مدل است. اگر لایه‌ی پنهان حذف شود، دو سر خروجی مستقیماً به x وصل باشند و خروجی λ نمایی و خروجی π سیگموئید باشد، دقیقاً ZIP کلاسیک به‌دست می‌آید. پس ZIP-NN تعمیمی آشیانه‌ای است و برتری احتمالی آن ناشی از انعطاف غیرخطی است، نه تغییر چارچوب احتمالی. «پواسون عصبی» (شبکه با درست‌نمایی پواسون) هم مدلی جداست و نباید با ZIP-NN یکی شمرده شود.

۵.۲. تابع زیان. آموزش با کمینه‌سازی منفی لگاریتم درست‌نمایی انجام می‌شود:

$$\begin{aligned} \mathcal{L}(\theta) &= -\sum_i \log p(y_i | x_i; \theta) \\ &= 1^n \left[I(y_i = 0) \ln(\pi_i + (1 - \pi_i)e^{-\lambda_i}) \right. \\ &\quad \left. + I(y_i > 0) \ln \left((1 - \pi_i) \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \right) \right], \end{aligned}$$

که در آن $I(\cdot)$ تابع نشانگر است. پس برآورد حاصل، بیشینه‌درست‌نمایی (یا در صورت جریمه/توقف زودهنگام (Early Stopping)، بیشینه‌درست‌نمایی تنظیم‌شده) در پارامتریزاسیون شبکه است. جمله‌ی $\ln y_i!$ به θ وابسته نیست و در بهینه‌سازی حذف می‌شود، اما برای گزارش درست‌نمایی و معیار اطلاعات آکائیکه (AIC) لازم است. برای پایداری عددی، $\ln(\pi + (1 - \pi)e^{-\lambda})$ را با تابع logaddexp و بر پایه‌ی لوجیت $a = \text{logit}(\pi)$ محاسبه کنید، با $\ln \pi = -\zeta(-a)$ و $\ln(1 - \pi) = -\zeta(a)$.

۵.۳. گرادیان و ارتباط با EM. فرض کنید و پیش ساخت باشد (با) و مانند بخش ۴-۴ تعریف شود. مشتق زیان هر مشاهده چنین است:

$$\frac{\partial \mathcal{L}_i}{\partial a_i} = \pi_i - \hat{z}_i, \quad \frac{\partial \mathcal{L}_i}{\partial c_i} = (1 - \hat{z}_i)(\lambda_i - y_i).$$

اگر خروجی λ با تابع softplus (سافت پلاس) ساخته شود (یعنی $\lambda_i = \zeta(c_i)$ ، مشتق دوم می شود $(1 - \hat{z}_i)(1 - y_i/\lambda_i)\sigma(c_i)$. این ها نمره ی رگرسیون لجستیک با برچسب نرم $\hat{z}_i$ و نمره ی رگرسیون پواسون با وزن $1 - \hat{z}_i$ هستند. بنابراین شبکه با گرادیان مستقیم همان کاری را می کند که EM در مدل ZIP کلاسیک با گام های متناوب انجام می دهد (هویت فیشر). این روابط را می توان با gradient check عددی هم راستی آزمایی کرد.

۵.۴. پیش بینی. مدل یک توزیع کامل می دهد:

$$\hat{\mu}_i = \mathbb{E}[Y_i | x_i] = (1 - \hat{\pi}_i)\hat{\lambda}_i, \quad \hat{p}_{0i} = \hat{\pi}_i + (1 - \hat{\pi}_i)e^{-\hat{\lambda}_i}.$$

سنجه هایی مانند میانگین قدرمطلق خطا (MAE)، میانگین مربعات خطا (MSE) و ریشه میانگین مربعات خطا (RMSE) فقط یک پیش بینی نقطه ای را ارزیابی می کنند. در این قسمت $\hat{y}_i = \hat{\mu}_i$ است.

۵.۵. آموزش و ارزیابی. داده ها به بخش آموزش (۸۰٪) و اعتبارسنجی (۲۰٪) تقسیم شدند. برای مهار بیش برآش (کاهش زیان آموزش همراه با افزایش زیان بیرون نمونه) از توقف زود هنگام (Early Stopping) استفاده شد. در این روش آموزش وقتی متوقف می شود که زیان اعتبارسنجی دیگر بهبود نیابد، و این خود نوعی تنظیم سازی ضمنی است. دقت پیش بینی روی مشاهده ی مجموعه ی ارزیابی با این سه معیار سنجیده شد:

$$\text{MAE} = \frac{1}{n} \sum_i |y_i - \hat{y}_i|, \quad \text{MSE} = \frac{1}{n} \sum_i (y_i - \hat{y}_i)^2, \quad \text{RMSE} = \sqrt{\text{MSE}}.$$

۵.۶. نتایج.

داده ی شبیه سازی شده ($p = 3, n = 2000$, ۶۵/۹۵٪ صفر):

مدل		MAE	RMSE
پواسون کلاسیک		۰.۸۲۰	۱.۰۹۰

مدل	RMSE	MAE
پواسون عصبی	۰.۸۴۰	۱.۱۱۰
ZIP کلاسیک	۰.۸۳۸	۱.۱۰۸
ZIP-NN	۰.۷۷۲	۱.۰۴۹

داده‌ی واقعی مقالات دانشجویان دکتری ($N = 915$, 30.5% صفر):

مدل	RMSE	MSE	MAE
پواسون کلاسیک	۱.۳۰	۳.۰۵	۱.۷۵
ZIP کلاسیک	۱.۳۰	۳.۲۸	۱.۸۱
ZIP-NN	۱.۲۲	۲.۹۵	۱.۷۲

ZIP-NN در هر دو مجموعه داده کمترین خطا را دارد. نسبت به بهترین رقیب، کاهش نسبی خطا (محاسبه‌ی خودم از اعداد جدول) در داده‌ی شبیه‌سازی شده حدود $5/9\%$ برای MAE و $3/8\%$ برای RMSE است. در داده‌ی واقعی حدود $6/2\%$ برای MAE، $3/3\%$ برای MSE و $1/7\%$ برای RMSE است. دو الگو نیاز به توضیح دارند:

- در شبیه‌سازی، پواسون عصبی از پواسون کلاسیک بدتر است. یعنی صرفاً افزودن غیرخطی بودن، بدون ساختار صفر-آماسیده، مزیتی نیاورده است. علت را بررسی کنید، مثلاً بیش‌برازش یا خطی بودن مکانیسم تولید داده.
- در داده‌ی واقعی، ZIP کلاسیک از پواسون هم بدتر است (RMSE برابر ۱.۸۱ در برابر ۱.۷۵). این با انتظار نظری همخوان نیست. یک توضیح ممکن ناهمخوانی تابع هدف (درست‌نمایی) با سنجه‌ی ارزیابی (خطای مربعی) یا برآورد ناپایدار است.

۶. چارچوب آماری دقیق مدل NN-ZIP و بازسازی ریاضی آن

در نسخهٔ اولیه، زنجیرهٔ مفهومی از پواسون به ZIP و سپس به NN-ZIP به‌درستی ترسیم شده است؛ در این بخش، NN-ZIP به‌عنوان یک مدل احتمالی شمارشی تعریف می‌شود که شبکهٔ عصبی به جای ضرایب ثابت،

توابع انعطاف‌پذیر از متغیرهای توضیحی را برای دو مؤلفه ZIP تولید می‌کند. این تفکیک مهم است، زیرا مزیت شبکه در اینجا صرفاً پیش‌بینی یک مقدار نقطه‌ای نیست؛ بلکه یادگیری یک توزیع شرطی کامل برای متغیر پاسخ است.

فرض کنید برای $i = 1, \dots, n$ ، متغیر پاسخ $Y_i \in \{0, 1, 2, \dots\}$ و بردار متغیرهای توضیحی $X_i \in R^p$ مشاهده شده باشد. یک متغیر پنهان Z_i معرفی می‌کنیم که نشان می‌دهد مشاهده از حالت صفر ساختاری آمده است یا از فرایند شمارش. در صورتی که $Z_i = 1$ ، آنگاه $Y_i = 0$ با احتمال یک؛ و در صورتی که $Z_i = 0$ ، متغیر Y_i از یک توزیع پواسون با پارامتر λ_i پیروی می‌کند. احتمال قرار گرفتن در مؤلفه صفر را با π_i نشان می‌دهیم. در نتیجه، صفر مشاهده‌شده می‌تواند دو منشأ داشته باشد و همین نکته منبع اصلی انعطاف ZIP است.

$$P(Y_i = 0 | X_i) = \pi_i + (1 - \pi_i)e^{-\lambda_i}$$

$$P(Y_i = y | X_i) = (1 - \pi_i) e^{-\lambda_i} \lambda_i^y / y!, \quad y = 1, 2, \dots$$

دو پارامتر مدل باید در دامنه‌های مناسب قرار گیرند. بنابراین شبکه عصبی مستقیماً λ_i و π_i را بدون قید تولید نمی‌کند؛ بلکه دو پیش‌سازه $\eta_{\lambda,i}$ و $\eta_{\pi,i}$ را تولید کرده و سپس نگاشت‌های تضمین‌کننده اعتبار پارامتر را اعمال می‌کند. انتخاب کلاسیک و بسیار مناسب برای پارامتر نرخ، تابع نمایی و برای احتمال، تابع سیگموئید است:

$$\lambda_i = \exp(\eta_{\lambda,i}), \quad \pi_i = 1 / (1 + \exp(-\eta_{\pi,i}))$$

اگر برای هر مشاهده یک میزان مواجهه یا Exposure وجود داشته باشد، آن متغیر بهتر است در مؤلفه شمارش به صورت offset وارد شود، نه به عنوان یک متغیر توضیحی معمولی. در این حالت می‌توان نوشت $\eta_{\lambda,i} = \log(o_i) + f_{\lambda}(X_i)$ ، که در آن o_i میزان مواجهه یا مدت زمان در معرض خطر است. این پارامتردهی از نظر تفسیر آماری یکسان‌سازی نرخ‌ها را ممکن می‌کند و از آنجا که در مدل‌های کاربردی ZIPNN نیز استفاده از exposure به عنوان offset گزارش شده است.

نقش	تعریف	کمیت
پاسخ شمارشی	تعداد رخداد مشاهده‌شده	Y_i
ورودی شبکه	برداری متغیرهای توضیحی	X_i
احتمال صفر ساختاری	$\Pr(Z_i = 1 X_i)$	π_i
شدت فرایند شمارش	میانگین مؤلفه پواسون	λ_i

خروجی صفر-آماسیده	پیش‌سازۀ سیگموئید	$\eta_{\pi,i}$
خروجی نرخ پواسون	پیش‌سازۀ لگاریتمی	$\eta_{\lambda,i}$

۶.۱. امید ریاضی، واریانس و شاخص پراکندگی

ZIP صرفاً «توزیع با صفر زیاد» نیست؛ بلکه مشخصات گشتاوری آن دقیقاً قابل استخراج است. با استفاده از ساختار آمیخته، برای میانگین شرطی داریم:

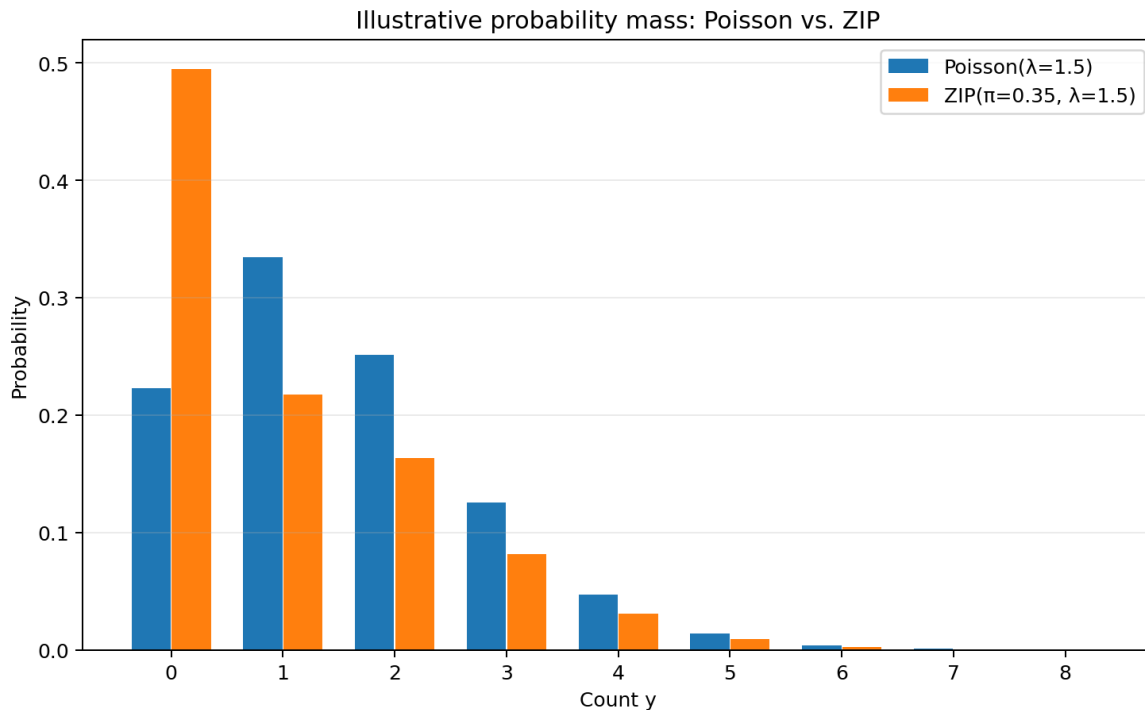
$$E(Y_i|X_i) = (1 - \pi_i) \lambda_i$$

از طرف دیگر، واریانس را می‌توان با قانون واریانس کل محاسبه کرد. شرط بر Z_i ، در حالت ساختاری و پواسونی، واریانس برابر صفر و λ_i است. بنابراین واریانس کل برابر می‌شود با حاصل جمع امید واریانس شرطی و واریانس امید شرطی. نتیجه مهم چنین است:

$$\text{Var}(Y_i|X_i) = (1 - \pi_i) \lambda_i [1 + \pi_i \lambda_i]$$

در نتیجه، شاخص پراکندگی شرطی یعنی $\text{Var}(Y_i|X_i)/E(Y_i|X_i)$ ، در حالت $\pi_i > 0$ ، برابر $1 + \pi_i \lambda_i$ است و از یک بزرگ‌تر می‌شود. این نتیجه توضیح می‌دهد که حتی اگر مؤلفه شمارش خودش پواسونی باشد، ترکیب با صفر ساختاری می‌تواند بیش‌پراکندگی ایجاد کند. از این منظر، فرض هم‌پراکندگی پواسون در داده‌های صفرآماسیده یک فرض قوی و بالقوه ناسازگار با سازوکار تولید داده است.

برای احتمال صفر، رابطه $P(Y_i = 0|X_i) = \pi_i + (1 - \pi_i)e^{-\lambda_i}$ نشان می‌دهد که صفر مشاهده‌شده ترکیبی از صفر ساختاری و صفر نمونه‌ای است. بنابراین اگر تنها نسبت صفرهای داده را با $e^{-\lambda}$ مقایسه کنیم، هنوز تشخیص قطعی صفر ساختاری انجام نداده‌ایم. آزمون و تشخیص باید در سطح مدل و همراه با بررسی سایر مقادیر شمارشی و پراکندگی انجام شود.



شکل ۱. مقایسهٔ تابع جرم احتمال پواسون و ZIP در یک مثال تحلیلی با $\lambda = 1.5$ و $\pi = 0.35$. این شکل صرفاً آموزشی است

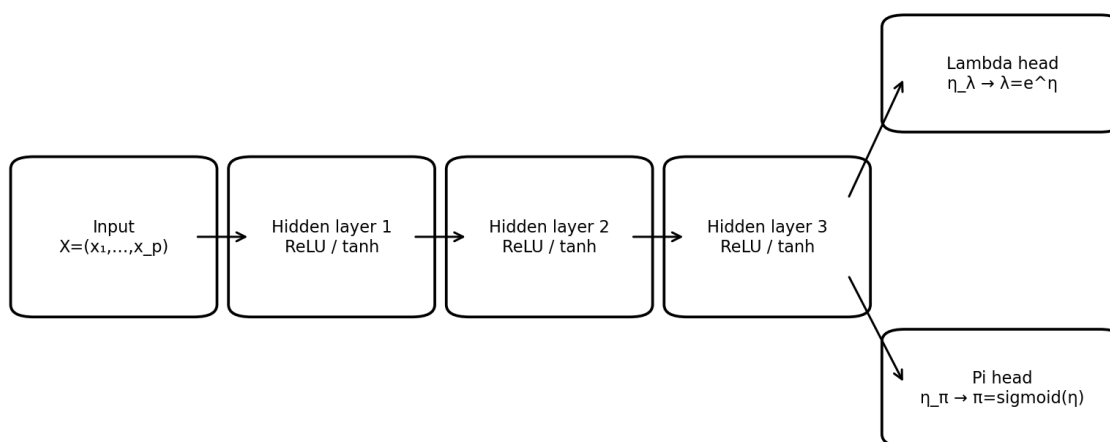
۶.۲. ساختار شبکه و تعریف NN-ZIP

مدل NN-ZIP را می‌توان یک مدل رگرسیونی توزیعی دانست. ورودی X_i از چند لایهٔ پنهان عبور می‌کند و در خروجی، دو عدد واقعی تولید می‌شوند. این دو عدد سپس با نگاشت‌های مناسب به λ_i و π_i تبدیل شده و یک توزیع ZIP شرطی را مشخص می‌کنند. این ساختار با شبکه‌ای که فقط $\hat{y}_i$ را پیش‌بینی می‌کند تفاوت بنیادی دارد، زیرا عدم قطعیت aleatoric مربوط به فرایند شمارش در خود خروجی مدل کدگذاری می‌شود.

در معماری مرجع گزارش شده توسط Jose و همکاران (۲۰۲۴)، تنهٔ پیشخور سه لایهٔ پنهان ۲۰، ۱۵ و ۱۰ نورونی دارد و یک لایهٔ چگال دو نورونی پیش از لایهٔ توزیعی قرار می‌گیرد. استفاده از یک distribution layer باعث می‌شود شبکه به جای خروجی خام، مستقیماً پارامترهای یک توزیع صفر-آماسیده را تولید کند. استفاده از همین معماری به عنوان baseline، و مقایسهٔ آن با معماری‌های کم‌عمق‌تر و عمیق‌تر، روش دفاع‌پذیرتری برای مطالعهٔ اثر پیچیدگی شبکه است.

نکتهٔ مهم در طراحی این مدل آن است که بهتر است صفر-آماسازی به دو خروجی مستقل و منعطف متصل باشد. اگر از یک شبکهٔ مشترک برای هر دو جزء استفاده شود، مزیت انتقال ویژگی بین مؤلفهٔ صفر و مؤلفهٔ شمارش حاصل می‌شود؛ اما همین اشتراک می‌تواند باعث coupling بیش‌ازحد شود. یک توسعهٔ مفید، آزمون سه معماری است: (الف) تنهٔ مشترک با دو خروجی، (ب) دو زیرشبکهٔ مستقل، و (ج) تنهٔ مشترک به‌علاوهٔ skip connection. مقایسهٔ این سه ساختار می‌تواند مستقیماً یک سؤال پژوهشی قابل انتشار بسازد.

NN-ZIP / ZIPNN architecture schematic



Distribution layer: ZIP(π_i, λ_i) $\rightarrow$ $E(Y_i|x_i)=(1-\pi_i)\lambda_i$

شکل ۲. معماری مفهومی NN-ZIP: تنهٔ چندلایهٔ مشترک، دو پیش‌سازهٔ توزیعی و لایهٔ خروجی ZIP. ساختار بر اساس معماری‌های ZIPNN گزارش‌شده در ادبیات بازطراحی شده است.

۶.۳. تابع درستنمایی و تابع زیان

اگر $u_i = I(Y_i = 0)$ نشانگر صفر بودن مشاهده باشد، لگاریتم درستنمایی هر مشاهده به‌صورت قطعه‌ای نوشته می‌شود، زیرا دقیقاً همان تابع هدفی است که شبکه را از یک رگرسیون معمولی به یک مدل آماری احتمالی تبدیل می‌کند.

$$\begin{aligned} \ell_i = & u_i \log[\pi_i + (1 - \pi_i)e^{-\lambda_i}] \\ & + (1 - u_i)\{\log(1 - \pi_i) - \lambda_i + y_i \log(\lambda_i) - \log(y_i!)\} \end{aligned}$$

در نتیجه، تابع زیان قابل کمینه‌سازی برای شبکه، منفی میانگین یا منفی مجموع این مقدار است. استفاده از NLL مزیتی دارد که MAE و MSE به‌تنهایی ندارند NLL: تمام اطلاعات مربوط به توزیع پیش‌بینی‌شده را ارزیابی می‌کند و شبکه را مجبور می‌سازد به مشاهده واقعی احتمال مناسب اختصاص دهد. اگر فقط MSE روی میانگین پیش‌بینی‌شده کمینه شود، مدل ممکن است میانگین مناسبی تولید کند اما احتمال صفر یا پراکندگی شرطی را بد برآورد کند.

در محاسبه عددی بهتر است عبارت مربوط به صفرها به شکل پایدار محاسبه شود، زیرا جمع $\pi_i + (1 - \pi_i)e^{-\lambda_i}$ می‌تواند در برخی نواحی پارامتری به مقادیر بسیار کوچک یا نزدیک صفر برسد. استفاده از پارامتردهی logit و توابعی مانند logaddexp، یا محاسبات معادل در چارچوب TensorFlow Probability، از underflow و overflow جلوگیری می‌کند. این مسئله برای شبکه‌هایی با عمق بیشتر یا نرخ‌های بزرگ‌تر λ اهمیت بیشتری پیدا می‌کند.

۶.۴. گرادیان‌های صریح و رابطه با الگوریتم EM

یکی از نقاط قوت نسخه حاضر این است که می‌توان رابطه آموزش مستقیم شبکه و الگوریتم EM برای ZIP کلاسیک را به‌طور ریاضی روشن کرد. اگر $\eta_\lambda = \log$ و $\eta_\pi = \text{logit}$ باشد، برای $y_i > 0$ مشتق لگاریتم درست‌نمایی نسبت به η_λ و η_π ساده است:

$$\partial \ell_i / \partial \eta_\lambda = y_i - \lambda_i, \quad \partial \ell_i / \partial \eta_\pi = -\pi_i, \quad (y_i > 0)$$

برای $y_i = 0$ ، اگر $q_i = \pi_i + (1 - \pi_i)e^{-\lambda_i}$ تعریف شود، خواهیم داشت:

$$\partial \ell_i / \partial \eta_\lambda = -\lambda_i(1 - \pi_i)e^{-\lambda_i} / q_i$$

$$\partial \ell_i / \partial \eta_\pi = \pi_i(1 - \pi_i)(1 - e^{-\lambda_i}) / q_i$$

این مشتقات نشان می‌دهند چرا صفرها برای آموزش مدل پیچیده‌تر از مقادیر مثبت‌اند: یک صفر هم‌زمان درباره مؤلفه ساختاری و مؤلفه پواسون اطلاعات می‌دهد و گرادیان آن نتیجه ترکیب این دو مسیر است. در EM کلاسیک، برای یک صفر مشاهده‌شده احتمال پسین ساختاری بودن صفر برابر است با $\tau_i = \pi_i / [\pi_i + (1 - \pi_i)e^{-\lambda_i}]$. در گام E این مقدار محاسبه و در گام M به‌صورت وزن وارد رگرسیون دو مؤلفه می‌شود. شبکه عصبی می‌تواند بدون تشکیل صریح τ_i و تنها با گرادیان مستقیم تابع NLL، همین تخصیص نرم

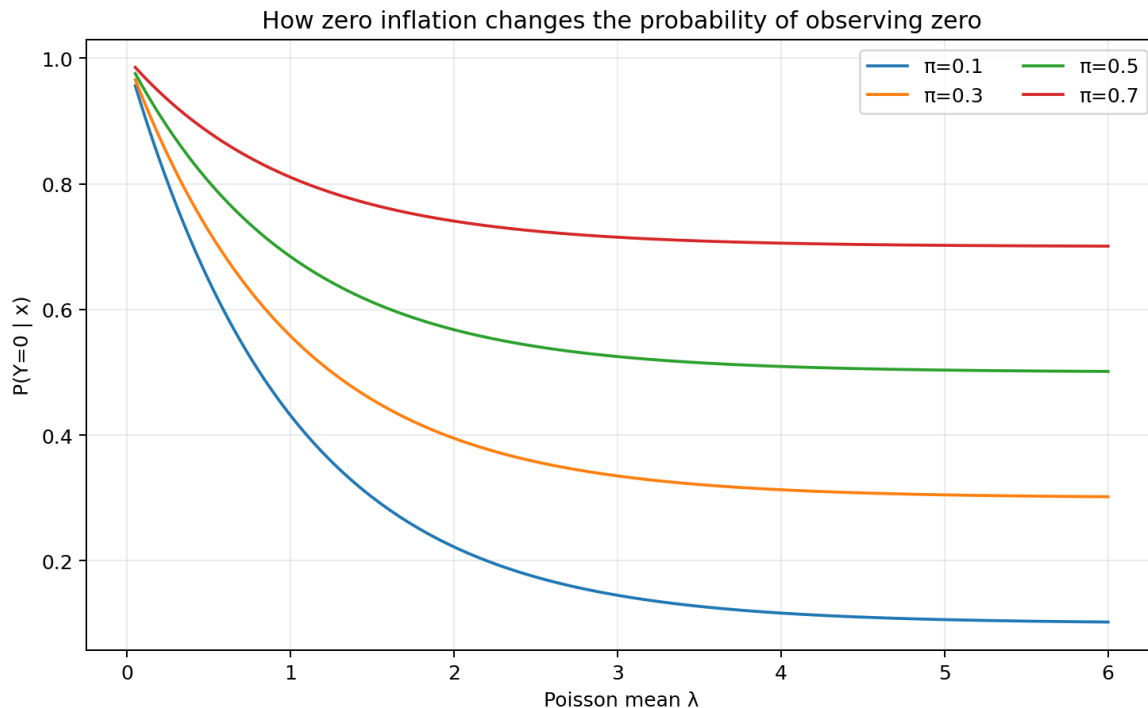
را به صورت ضمنی یاد بگیرد. از این نظر، رابطه مفهومی بین EM و گرادیان شبکه وجود دارد، اما این دو الگوریتم از نظر اجرا یکسان نیستند.

۶.۵. شناسایی پارامترها و مسائل پایداری

صفر آماسیده بودن به تنهایی تضمین نمی کند که π_i و λ_i به طور پایدار شناسایی شوند. اگر λ_i بسیار کوچک باشد، مؤلفه پواسون نیز احتمال صفر بالایی تولید می کند و تفکیک صفر ساختاری از صفر تصادفی دشوار می شود. این مسئله به ویژه در نمونه های کوچک یا زمانی که هر دو مؤلفه دقیقاً از یک مجموعه بزرگ متغیرها استفاده می کنند تشدید می شود. بنابراین برای مطالعات تجربی باید علاوه بر معیار عملکرد، حساسیت تخمین ها به مقداردهی اولیه و ساختار شبکه نیز گزارش شود.

دو راهکار عملی عبارت اند از: نخست، استفاده از تنظیم سازی، dropout یا توقف زود هنگام برای کاهش پیچیدگی مؤثر شبکه؛ و دوم، محدود کردن معماری یا مجموعه ویژگی های مؤلفه صفر در صورتی که منطق فرایند تولید داده چنین چیزی را توجیه کند. در مدل های کلاسیک ZIP نیز استفاده از متغیرهای متفاوت در دو مؤلفه می تواند به تفسیر بهتر کمک کند. در مدل عصبی، همین ایده به تفکیک فضا های ویژگی یا حتی دو زیر شبکه جدا تعمیم می یابد.

از دیدگاه علمی، نباید صرفاً به کاهش MAE برای انتخاب ساختار شبکه بسنده کرد. اگر NN-ZIP نسبت به ZIP کلاسیک MAE کمتری داشته باشد ولی احتمال صفر را به شدت بد کالیبره کند، از نظر آماری ممکن است مدل بهتری نباشد. بنابراین کالیبراسیون احتمال صفر، NLL و شکل کامل توزیع پیش بینی شده باید بخشی از ارزیابی اصلی باشد.



شکل ۳. اثر π بر احتمال صفر پیش‌بینی‌شده در ZIP. مقادیر شکل تحلیلی‌اند و برای توضیح رفتار مدل ساخته شده‌اند.

۷. طراحی پیشنهادی مطالعه و پروتکل ارزیابی

روند تحلیل باید از پیش مشخص و قابل بازتولید باشد. که در سه سطح تفکیک شود: مجموعه آموزش برای برآورد وزن‌ها، مجموعه اعتبارسنجی برای انتخاب ابرپارامترها و توقف زودهنگام، و مجموعه آزمون که تنها یک‌بار در انتهای فرایند برای گزارش عملکرد استفاده می‌شود. استفاده از یک تفکیک ثابت یا seed مشخص اهمیت دارد، زیرا تغییرات تصادفی در شبکه‌های عصبی می‌تواند باعث نوسان محسوس در نتایج شود.

اگر داده واقعی مورد استفاده همان مجموعه Doctoral Publications باشد، این داده شامل ۹۱۵ مشاهده و متغیر پاسخ تعداد مقالات منتشرشده در سه سال پایانی دوره دکتری است. مستندات این داده، متغیرهای gender، married، kids، prestige و mentor را به‌عنوان متغیرهای توضیحی معرفی می‌کنند و منبع آن را مکمل‌های Long می‌دانند

مرحله	هدف	خطر اصلی	اقدام پیشنهادی
تقسیم داده	تفکیک train/validation/test	نشت اطلاعات	train استانداردسازی فقط روی

آموزش	NLL کمینه سازی	بیش برزش	early stopping + dropout
تنظیم	depth, width, lr انتخاب	test انتخاب بر اساس	validation فقط
آزمون	برآورد عملکرد تعمیم	seed نوسان ناشی از	و گزارش seed چند میانگین/انحراف معیار
تشخیص	کنترل صفر و پراکندگی	اتکا به یک معیار	NLL + zero calibration + rootogram

در پیش پردازش متغیرهای عددی، استانداردسازی یا نرمال سازی باید فقط از آمار مجموعه آموزش استفاده کند. متغیرهای طبقه ای نیز باید با one-hot encoding یا embedding به شکل سازگار تبدیل شوند. در مورد dataset کوچک مانند ۹۱۵ مشاهده، افزایش بی رویه عرض شبکه به سرعت تعداد پارامترها را بالا می برد. بنابراین معماری های خیلی عمیق بدون شواهد تجربی، انتخاب خوبی نیستند.

۷.۱. مدل های مبنا و مقایسه منصفانه

حداقل مقایسه باید شامل پنج مدل باشد: پواسون کلاسیک، پواسون عصبی، ZIP کلاسیک، NN-ZIP و در صورت وجود بیش پراکندگی شدید ZINB یا NB. این مقایسه اجازه می دهد سهم هر ایده جدا شود. مقایسه پواسون با NN-ZIP به تنهایی کافی نیست، زیرا دو تغییر هم زمان رخ داده است: توزیع از پواسون به ZIP تغییر کرده و رابطه خطی به غیرخطی تبدیل شده است. بنابراین افزودن دو baseline میانی یعنی Neural Poisson و ZIP کلاسیک برای نسبت دادن بهبود به علت درست ضروری است.

مدل	توزیع شرطی	رابطه ورودی-پارامتر	مزیت اصلی	محدودیت
Poisson-GLM	Poisson	log link + خطی	ساده و تفسیرپذیر	هم پراکندگی و صفرهای اضافی
Poisson-NN	Poisson	غیرخطی	انعطاف پذیر	عدم مدل سازی صفر ساختاری
ZIP-GLM	ZIP	دو پیشگوی خطی	مدل سازی صریح صفرهای اضافی	خطی بودن هر دو مؤلفه
NN-ZIP	ZIP	دو پیشگوی غیرخطی	انعطاف + توزیع کامل	بیش برزش و تفسیر دشوار
ZINB	ZINB	خطی یا غیرخطی	صفر آماسیدگی + پراکندگی بیشتر	پارامترگذاری پیچیده تر

۷.۲. معیارهای ارزیابی: از MAE و RMSE تا معیارهای توزیعی

در نسخهٔ اولیه، MAE، MSE و RMSE گزارش شده‌اند. این معیارها برای ارزیابی پیش‌بینی نقطه‌ای لازم‌اند، ولی برای NN-ZIP کافی نیستند؛ زیرا خروجی مدل یک توزیع کامل است. معیارهای اصلی پیشنهادی در نسخهٔ نهایی عبارت‌اند از NLL یا Log Score، MAE، RMSE، خطای احتمال صفر، و در صورت امکان یک scoring rule مناسب برای پیش‌بینی‌های احتمالاتی. ادبیات ارزیابی count data نشان می‌دهد که scoring rule درستی، مدل را تشویق می‌کنند توزیع پیش‌بینی صادقانه و کالیبره‌ای ارائه کند.

$$MAE = (1/n) \sum |y_i - \hat{y}_i|$$

$$MSE = (1/n) \sum (y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{MSE}$$

$$NLL = -(1/n) \sum \log P(Y_i = y_i | X_i)$$

برای احتمال صفر می‌توان یک معیار سادهٔ کالیبراسیون تعریف کرد: در هر مشاهده، مدل احتمال $\hat{p}_{0,i} = P(Y_i = 0 | X_i)$ را ارائه می‌دهد و با نشانگر $Z_i = I(Y_i = 0)$ مقایسه می‌شود. خطای مربعی احتمال صفر به شکل Brier-like score یعنی میانگین $(\hat{p}_{0,i} - Z_i)^2$ ، شاخص مناسبی برای بررسی جداگانهٔ مؤلفهٔ صفر است. علاوه بر مقدار عددی، رسم reliability curve برای احتمال صفر می‌تواند نشان دهد که آیا پیش‌بینی ۰.۸ واقعاً حدود ۸۰ درصد صفر در میان مشاهدات مشابه ایجاد کرده است یا خیر.

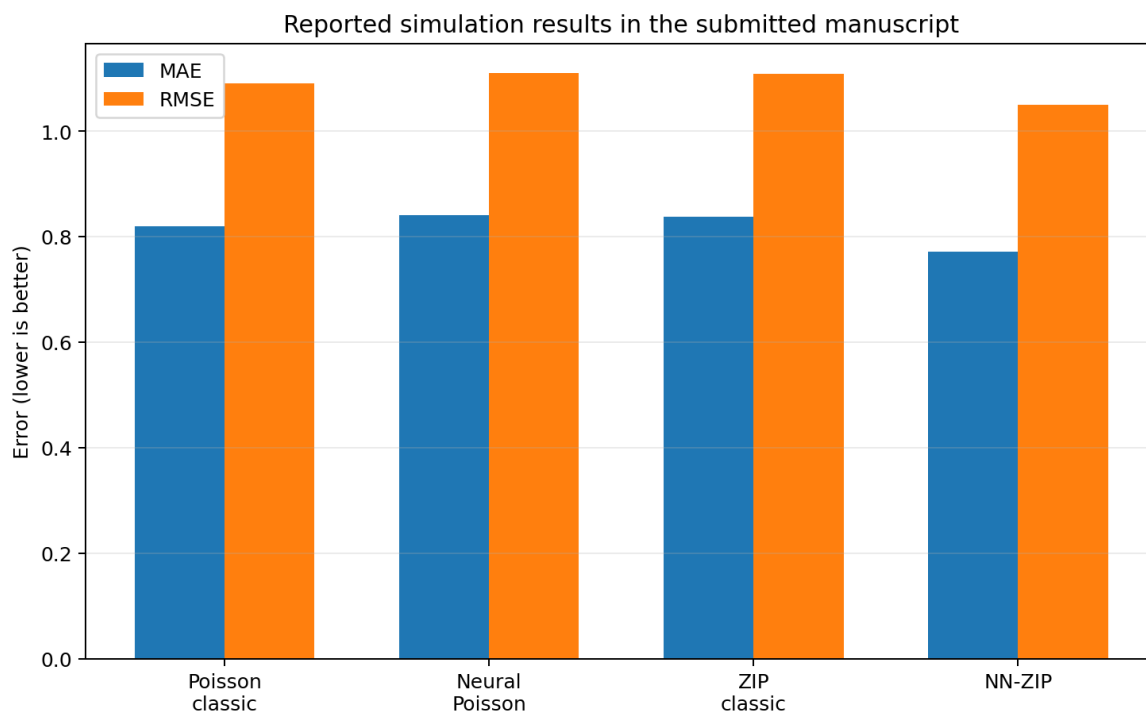
یک ابزار تشخیصی مناسب دیگر rootogram است که اختلاف بین فراوانی مشاهده‌شده و مورد انتظار را در مقیاس ریشهٔ دوم نشان می‌دهد. این نمودار در تشخیص کم‌برازش در بخش صفر و tail توزیع بسیار مفید است و برای count regression و مدل‌های hurdle/zero-inflated توصیه شده است

۸. تحلیل نتایج موجود و نقد آماری آن‌ها

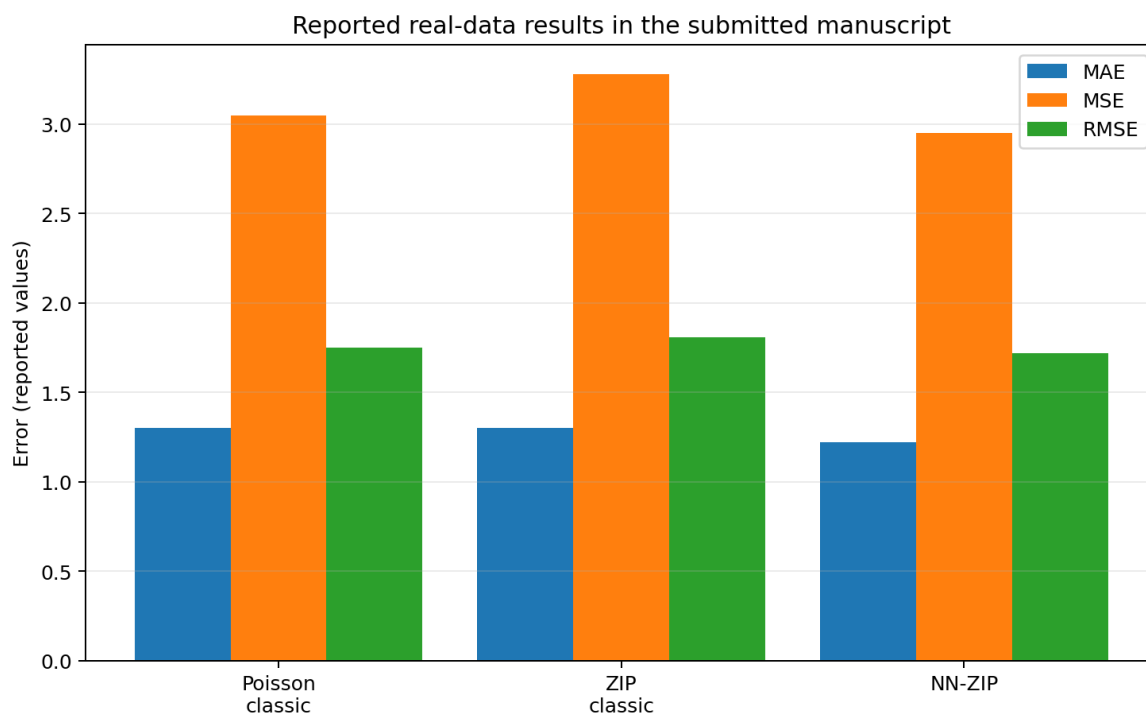
نتایج نسخهٔ اولیهٔ نشان می‌دهند که در شبیه‌سازی با $n = 2000$ و حدود ۶۵.۹۵ درصد صفر، NN-ZIP کمترین MAE و RMSE را میان مدل‌های گزارش شده به دست آورده است: $MAE = 0.772$ و $RMSE = 1.049$. این الگو با منطق مسئله سازگار است، زیرا داده شبیه‌سازی شده ساختاری صفرآماسیده دارد و مدل NN-ZIP هم‌زمان شکل غیرخطی و حضور صفرهای ساختاری را مدل می‌کند. با این حال، برای تبدیل این نتیجه به یک ادعای علمی قوی، باید معلوم شود فرایند تولید دادهٔ شبیه‌سازی دقیقاً چه بوده، π و λ چگونه به X وابسته بوده‌اند، و آیا شبکه در شرایطی که رابطه واقعاً غیرخطی نیست نیز برتری دارد یا نه.

در دادهٔ واقعی مقالات دانشجویان دکتری نیز NN-ZIP دارای $MAE = 1.22$ ، $MSE = 2.95$ و $RMSE = 1.72$ گزارش شده است، در حالی که Poisson کلاسیک و ZIP کلاسیک به ترتیب $RMSE$ برابر ۱.۷۵ و ۱.۸۱ دارند. این نتیجه جالب است، اما به تنهایی کافی نیست تا نتیجه بگیریم ZIP کلاسیک «بدتر از پواسون» است. ممکن است معیار هدف، نمونهٔ آزمون، تفاوت در تنظیمات برآورد یا ناپایداری بهینه‌سازی باعث این ترتیب شده باشد.

همچنین مقایسهٔ درصد کاهش خطا باید از یک baseline دقیق تعریف شود. در دادهٔ شبیه‌سازی شده، کاهش MAE NN-ZIP نسبت به بهترین رقیب حدود ۵.۹ درصد و کاهش RMSE حدود ۳.۸ درصد گزارش شده است. چنین بهبودی کوچک اما واقعی است و نیاز به interval یا آزمون تکرارپذیری دارد. پیشنهاد می‌شود شبیه‌سازی حداقل در ۳۰ تا ۱۰۰ تکرار با seedهای متفاوت اجرا شود و برای هر معیار میانگین، انحراف معیار و فاصلهٔ اطمینان بوت‌استرپی یا Monte Carlo گزارش شود.



شکل ۴. بازنمایی گرافیکی مقادیر **MAE** و **RMSE** گزارش شده برای دادهٔ شبیه‌سازی شده در نسخهٔ اولیه. اعداد از جدول استخراج شده‌اند.



شکل ۵. بازنمایی گرافیکی نتایج گزارش شده برای داده واقعی مقالات دانشجویان دکتری. این نمودار تنها نتایج نسخه اولیه را بازنمایی می کند و به معنای بازتحلیل خام داده نیست.

۸.۱. یک نکته مهم درباره داده واقعی مقالات دانشجویان دکتری

داده مورد اشاره با مشخصات مجموعه PhD Publications در بسته AER همخوانی بسیار بالایی دارد: ۹۱۵ مشاهده، متغیر پاسخ تعداد مقالات و متغیرهایی مانند جنسیت، تأهل، تعداد فرزندان کم سن، پرستش برنامه دکتری و تعداد مقالات استاد راهنما. مستندات این مجموعه همچنین نمونه هایی برای برازش Poisson، Negative Binomial، ZIP و ZINB ارائه می کنند. بنابراین استفاده از آن برای مطالعه count regression انتخاب مناسبی است.

۹. تفسیرپذیری NN-ZIP و خروج از جعبه سیاه

نقطه ضعف طبیعی شبکه عصبی این است که ضرایب وزن ها معادل ضرایب یک مدل خطی نیستند. حتی اگر شبکه دارای یک ورودی و یک لایه پنهان باشد، وزن ها به دلیل تقارن های جایگشتی بین نورون ها شناسه یکتایی ندارند. بنابراین تفسیر «افزایش یک واحد در وزن نورون شماره ۴» از نظر علمی معنای مستقیمی ندارد. تفسیر باید در سطح تابع پیش بینی، پارامترهای توزیع و تغییر در احتمال صفر یا نرخ شمارش انجام شود.

یک مسیر جدی استفاده از LocalGLMnet است. در این رویکرد، شبکه به جای یک بردار ضرایب ثابت، ضرایب وابسته به ویژگی $\beta(x)$ را تولید می کند. این ضرایب می توانند برای نمایش سهم محلی متغیرها و ساخت توضیحاتی شبیه GLM به کار روند. در ادبیات LocalGLMnet، ZIPNN دقیقاً برای تفسیر مدل صفر-آماسیده به کار گرفته شده است.

از طرف دیگر، روش های SHAP و partial dependence نیز می توانند برای تفسیر شبکه استفاده شوند. با این حال باید روشن شود که این روش ها تفسیر پسینی هستند و لزوماً پارامتر structural-zero را به یک رابطه علی تبدیل نمی کنند. مخصوصاً اگر متغیرهای ورودی با هم هم بسته باشند، اهمیت feature ممکن است بین آن ها پخش شود. به همین دلیل گزارش اثرگذاری متغیرها باید همراه با نمودار وابستگی و تحلیل حساسیت باشد.

سطح تفسیر	چه چیزی گزارش می شود؟	مناسب برای
سطح توزیع	$\pi(x), \lambda(x), P(Y = 0 x)$	تحلیل آماری و ریسک

سطح پیش‌بینی	$E(Y x)$, MAE/RMSE	پیش‌بینی
سطح ویژگی	$\beta_j(x)$ PDP/ICE/SHAP	توضیح اثر متغیر
سطح پایداری	و معماری seed تغییر نتایج با	اعتبار روش

۹.۱. تحلیل اثر حاشیه‌ای در NN-ZIP

در مدل NN-ZIP دو نوع اثر حاشیه‌ای باید از هم تفکیک شود. نخست اثر X_j بر λ ، که بر شدت شمارش در واحدهای فعال اثر می‌گذارد؛ دوم اثر X_j بر π ، که احتمال قرار گرفتن در حالت صفر ساختاری را تغییر می‌دهد. اثر نهایی X_j بر میانگین شرطی $E(Y|X) = (1 - \pi)\lambda$ ترکیبی از هر دو مسیر است. بنابراین ممکن است یک متغیر λ را افزایش دهد اما هم‌زمان π را نیز افزایش دهد و اثر نهایی بر میانگین کوچک یا حتی منفی شود.

$$\partial E(Y|X)/\partial x_j = (1 - \pi) \partial \lambda / \partial x_j - \lambda \partial \pi / \partial x_j$$

رابطه بالا از نظر تفسیری بسیار مهم است. اگر متغیر x_j با ورود به مؤلفه صفر ساختاری احتمال عدم وقوع را بالا ببرد، ولی در واحدهای در معرض وقوع شدت شمارش را هم افزایش دهد، یک ضریب واحد برای خلاصه کردن این رفتار کافی نیست. همین مسئله یکی از بهترین توجیه‌های علمی برای استفاده از NN-ZIP در حضور روابط پیچیده و ناهمگون است.

۱۰. کنترل بیش‌برازش و انتخاب ابرپارامترها

شبکه عصبی انعطاف‌پذیر است و این مزیت در نمونه‌های کوچک می‌تواند به ضعف تبدیل شود. در dataset با ۹۱۵ مشاهده، شبکه‌ای با صدها یا هزاران پارامتر ممکن است به راحتی الگوهای تصادفی را یاد بگیرد. بنابراین استفاده از early stopping باید بخشی از طراحی مطالعه باشد، نه یک اقدام اصلاحی پس از مشاهده عملکرد. در ادبیات ZIPNN نیز early stopping و dropout برای کنترل بیش‌برازش به کار رفته‌اند.

ابریارامترهایی که باید مورد جست‌وجو قرار گیرند شامل تعداد لایه‌ها، تعداد نوروها در هر لایه، تابع فعال‌سازی، نرخ یادگیری، نرخ dropout، اندازه batch و تعداد epoch است. اما جست‌وجوی کور تمام فضای پارامترها برای یک dataset کوچک توجیهی ندارد. راه بهتر استفاده از مجموعه محدودی از معماری‌های از پیش تعریف‌شده و گزارش شفاف آن‌هاست. مثلاً معماری‌های ۸-۱۰، ۲۰-۱۵-۱۰ و ۳۲-۱۶-۸ را می‌توان با دو یا سه نرخ یادگیری مقایسه کرد.

ابزارامتر	مقادیر پیشنهادی برای مطالعه	منطق انتخاب
Hidden layers	۱ تا ۳	کنترل ظرفیت مدل
Neurons	۸-۳۲ در هر لایه	$n = 915$ متناسب با
Activation	ReLU / tanh	مقایسه اشباع و غیرخطی
Optimizer	Nadam / Adam	گرادیان مرتبه اول
Learning rate	$1e-4$ تا $5e-3$	حساسیت همگرایی
Dropout	۰ تا ۰.۳۰	کنترل بیش‌برازش
Early stopping	patience = 20 – 50	validation NLL توقف بر اساس

در زمان گزارش نتایج باید مشخص شود که بهترین مدل بر اساس کدام تابع هدف انتخاب شده است. انتخاب بر اساس کمترین RMSE و سپس گزارش NLL بدون اینکه NLL در انتخاب دخالت داشته باشد، ممکن است ساختاری نامتوازن ایجاد کند. پیشنهاد می‌شود معیار انتخاب اصلی NLL باشد و MAE/RMSE و zero calibration به‌عنوان معیارهای ثانویه گزارش شوند؛ زیرا هدف مدل، پیش‌بینی یک توزیع شرطی است.

۱۱. الگوریتم پیشنهادی برای آموزش NN-ZIP.

Train / Validation / Test به D گام ۱: تقسیم

Train فقط روی encoder و scaler گام ۲: برازش

η_λ و η_π گام ۳: عبور X از شبکه و تولید

$\pi = \text{sigmoid}(\eta_\pi)$ و $\lambda = \exp(\eta_\lambda)$ گام ۴:

برای هر مشاهده ZIP مدل NLL گام ۵: محاسبه

Nadam/Adam گام ۶: محاسبه گرادیان‌ها و به‌روزرسانی وزن‌ها با

Early Stopping و اجرای Validation NLL گام ۷: پایش

Test و ارزیابی یک‌باره روی checkpoint گام ۸: انتخاب بهترین

برای کنترل ناپایداری ناشی از مقداردهی اولیه، توصیه می‌شود کل روند حداقل با چند seed مستقل تکرار شود. گزارش «بهترین نتیجه از ۱۰ اجرا» از نظر علمی بسیار ضعیف‌تر از گزارش میانگین $\pm$ انحراف معیار ۱۰ اجراست، مگر اینکه از پیش معیار انتخاب را تعریف کرده باشید. همین نکته باید در بخش محدودیت‌ها نیز صریحاً مطرح شود.

۱۲. آزمون‌های تکمیلی

مجموعه نتایج فعلی هنوز کافی نیست. حداقل چهار تحلیل تکمیلی پیشنهاد می‌شود. نخست، مطالعه شبیه‌سازی چندسطحی که در آن درصد صفرهای اضافی از ۱۰، ۳۰، ۵۰ تا ۷۰ درصد تغییر کند. دوم، کنترل غیرخطی بودن رابطه X با پارامترها؛ یعنی سناریویی که در آن رابطه واقعاً خطی است تا معلوم شود شبکه آیا مزیت اضافی ایجاد می‌کند یا صرفاً ظرفیت غیرضروری مصرف می‌کند. سوم، مقایسه NN-ZIP با ZINB و در صورت لزوم ZINB-NN. چهارم، بررسی کالبراسیون و خطای احتمال صفر در کنار MAE و RMSE.

در مطالعه شبیه‌سازی می‌توان مثلاً چنین مدل مولدی تعریف کرد: $\lambda_i = 0.4 + 0.6x_1 - \log$ و $\pi_i = -0.8 + 0.7x_1 - 0.4x_4 \logit$ و $\sin(x_3)$ $0.5x_2^2 + 0.3$ سپس X ها از توزیع‌های مشخص نمونه‌گیری شوند و Y از ZIP حاصل شود. هدف این نیست که این ضرایب «نتیجه واقعی» تلقی شوند؛ بلکه چنین سناریویی آزمون می‌کند که آیا NN-ZIP می‌تواند روابط غیرخطی شناخته‌شده را بازایی کند.

در هر سناریو، بهتر است علاوه بر خطای پیش‌بینی، خطای پارامترهای توزیعی نیز ارزیابی شود؛ مثلاً میانگین خطای مطلق بین π واقعی و π پیش‌بینی‌شده، و بین λ واقعی و λ پیش‌بینی‌شده. این ارزیابی مشخص می‌کند آیا مدل فقط میانگین نهایی را خوب می‌زند یا واقعاً دو مؤلفه ZIP را به‌درستی یاد گرفته است.

۱۲.۱. سنجش اهمیت صفرآماسیدگی

برای مطالعه حساسیت، شاخص صفرآماسیدگی را می‌توان به صورت اختلاف بین نسبت صفر مشاهده‌شده و نسبت صفر مورد انتظار تحت پواسون برازش‌شده تعریف کرد. البته این شاخص فقط یک معیار توصیفی است و به‌تنهایی آزمون رسمی محسوب نمی‌شود. برای نتیجه‌گیری باید احتمال صفر تحت مدل برازش‌شده، پراکندگی و سایر بخش‌های توزیع هم بررسی شوند.

$$ZI - gap = \hat{P}(Y = 0)_{\text{observed}} - (1/n) \sum \exp(-\hat{\lambda}_i)$$

اگر Zi-gap بزرگ باشد، احتمال می‌رود مدل پواسون سهم صفرها را کم‌برآورد کند؛ با این حال اگر λ ناهمگن یا مشخصات mean نادرست باشد، بخشی از این فاصله می‌تواند ناشی از خطای مدل میانگین باشد. این دقیقاً دلیل دیگری است که صفرآماسیدگی نباید فقط بر اساس شمارش صفرها قضاوت شود.

۱۳. مقایسه ZIP، ZINB، و Hurdle: چه زمانی NN-ZIP انتخاب درستی است؟

ZIP زمانی انتخاب طبیعی است که مسئله اصلی، وجود صفرهای اضافی نسبت به پواسون باشد و مؤلفه شمارش در حالت فعال تقریباً پواسونی باقی بماند. اگر داده بیش‌پراکندگی شدیدی داشته باشد که حتی پس از لحاظ صفرهای اضافی باقی بماند، ZINB می‌تواند مناسب‌تر باشد. اگر منطق فرایند تولید بگویند که صفرها همگی از مرحله «عدم وقوع» ناشی می‌شوند و پس از عبور از مانع همه مقادیر مثبت از یک توزیع بریده می‌آیند، مدل hurdle از نظر معنایی مناسب‌تر است.

مدل	صفر از کجا می‌آید؟	بخش مثبت	کاربرد مناسب
Poisson	یک منبع	Poisson	داده شمارشی ساده
NB	یک منبع	Negative Binomial	overdispersion
ZIP	دو منبع	Poisson	excess zeros
ZINB	دو منبع	Negative Binomial	excess zeros + overdispersion
Hurdle	همه صفرها از مانع	zero-truncated count	فرایند دو مرحله‌ای

نکته کلیدی این است که NN-ZIP فقط در صورتی انتخاب برتر است که هم ساختار ZIP مناسب باشد و هم روابط غیرخطی واقعاً وجود داشته باشند. استفاده از شبکه صرفاً به دلیل مدرن بودن، از نظر روش‌شناختی دفاع‌پذیر نیست. اگر ZIP خطی تقریب کافی ارائه دهد و حجم نمونه کوچک باشد، شبکه بزرگ ممکن است فقط نوسان و واریانس برآورد را افزایش دهد. بنابراین باید از یک مقایسه طبقه‌بندی شده استفاده شود: «آیا توزیع مناسب است؟» و پس از آن «آیا رابطه مناسب غیرخطی است؟».

۱۴. پیوست A: فرمول‌های کلیدی

$$E(Y|X) = (1 - \pi)\lambda$$

$$\text{Var}(Y|X) = (1 - \pi)\lambda(1 + \pi\lambda)$$

$$P(Y = 0|X) = \pi + (1 - \pi)e^{-\lambda}$$

$$P(Y = y|X) = (1 - \pi)\text{Pois}(y; \lambda), \quad y \geq 1$$

$$\eta_\lambda = f_\lambda(X) + \text{offset}, \quad \lambda = \exp(\eta_\lambda)$$

$$\eta_\pi = f_\pi(X), \quad \pi = \text{sigmoid}(\eta_\pi)$$

$$\text{NLL} = -\sum [I(y = 0) \log q_0 + I(y > 0) \{\log(1 - \pi) - \lambda + y \log \lambda - \log(y!)\}]$$

یادداشت درباره نتایج و شکل‌های افزوده شده:

شکل‌های ۱ و ۳ کاملاً تحلیلی و آموزشی هستند و بر اساس فرمول ZIP تولید شده‌اند. شکل‌های ۴ و ۵ فقط اعداد گزارش شده در نسخه اولیه را بازنمایی می‌کنند و ادعای باز اجرای مدل یا دسترسی به داده خام ندارند. بنابراین در صورت آماده بودن کد و داده اصلی، پیشنهاد می‌شود نمودارهای تجربی نهایی با خروجی واقعی اجرای مجدد جایگزین شوند. این تفکیک برای جلوگیری از تولید شواهد ظاهری بدون پشتوانه داده‌ای ضروری است.

مراجع

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Akademiai Kiado.

Bishop, C. M. (1994). Mixture density networks. Technical Report NCRG/94/004, Aston University.

Böhning, D., Dietz, E., Schlattmann, P., Mendonça, L., & Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society: Series A*, 162(2), 195–209.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression Analysis of Count Data* (2nd ed.). Cambridge University Press.

- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4), 303–314.
- Czado, C., Gneiting, T., & Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65(4), 1254–1261.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1), 1–38.
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of AISTATS*, 9, 249–256.
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, 56(4), 1030–1039.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366.
- Jose, A., Macdonald, A. S., Tzougas, G., & Streftaris, G. (2024). Interpretable zero-inflated neural network models for predicting admission counts. *Annals of Actuarial Science*, 18(3), 644–674. <https://doi.org/10.1017/S1748499524000058>
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.
- Kleiber, C., & Zeileis, A. (2016). Visualizing count data regressions using rootograms. *The American Statistician*, 70(3), 296–303. <https://doi.org/10.1080/00031305.2016.1173590>
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1–14. <https://doi.org/10.1080/00401706.1992.10485228>
- Leshno, M., Lin, V. Y., Pinkus, A., & Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6), 861–867.

- Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Sage.
- Martens, J. (2010). Deep learning via Hessian-free optimization. *Proceedings of ICML*.
- Martin, J., & Hall, D. B. (2016). R^2 measures for zero-inflated regression models for count data with excess zeros. *Journal of Statistical Computation and Simulation*, 86(18), 3777–3790.
- Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33(3), 341–365.
- Pearlmutter, B. A. (1994). Fast exact multiplication by the Hessian. *Neural Computation*, 6(1), 147–160.
- Ridout, M., Hinde, J., & Demétrio, C. G. B. (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, 57(1), 219–223.
- Richman, R., & Wüthrich, M. V. (2023). LocalGLMnet: Interpretable deep learning for tabular data. *Scandinavian Actuarial Journal*, 2023(1), 71–95. <https://doi.org/10.1080/03461238.2022.2081816>
- Rodrigo, H., & Tsokos, C. P. (2020). Bayesian modelling of nonlinear Poisson regression with artificial neural networks. *Journal of Applied Statistics*, 47(5), 757–774. <https://doi.org/10.1080/02664763.2019.1653268>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, 27(8), 1–25. <https://doi.org/10.18637/jss.v027.i08>
- Young, S., Jenkins, P., Da, L., Dotson, J., & Wei, H. (2025). Fully heteroscedastic count regression with deep double Poisson networks.

Proceedings of the 42nd International Conference on Machine Learning, 267, 72786–72818.

Jun, S. (2026). Zero-inflated data analysis using graph neural networks with convolution. *Computers*, 15(2), 104.

<https://doi.org/10.3390/computers15020104>

Lee, H., Ha, I. D., Hwang, C., & Lee, Y. (2026). A robust neural network with random effects for subject-specific prediction of clustered count data. *Statistical Methods in Medical Research*, 35(8).

<https://doi.org/10.1177/09622802261459873>

Workie, M. S., & Yi, X. (2025). Robust Bayesian inference in the multilevel zero-inflated generalized Poisson model. *Statistics in Medicine*.

<https://doi.org/10.1002/sim.70173>

Jose, A., Macdonald, A. S., Tzougas, G., & Streftaris, G. (2022). A combined neural network approach for the prediction of admission rates related to respiratory diseases. *Risks*, 10(11), 217.