

مدل‌بندی رگرسیون غیرخطی با شبکه‌های عصبی بیزی و مقایسه رویکردهای نمونه‌گیری و واریاسیونی

امید کریمی^{۱*} و فاطمه حسینی^۲

^{۱،۲} گروه آمار، دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه سمنان

تاریخ دریافت: ۱۴۰۴/۱۱/۱۰

تاریخ پذیرش: ۱۴۰۵/۰۴/۰۷

چکیده

مدل‌سازی داده‌های غیرخطی، به دلیل وجود روابط پیچیده و ساختارهای پنهان، مستلزم بهره‌گیری از رهیافت‌هایی فراتر از مدل‌های خطی کلاسیک است. در این پژوهش، به‌منظور ارزیابی عملکرد روش‌های مختلف استنباط بیزی در مدل‌سازی غیرخطی، از چارچوب شبکه‌های عصبی بیزی استفاده شده است. سه روش پرکاربرد در این حوزه شامل الگوریتم همیلتونی مونت‌کارلو، نمونه‌گیری تطبیقی بدون بازگشت و استنباط واریاسیونی، در قالب یک مدل رگرسیون غیرخطی پیاده‌سازی و با یکدیگر مقایسه شده‌اند. تحلیل‌ها بر روی داده‌های شبیه‌سازی‌شده و همچنین یک مجموعه داده واقعی انجام شده و عملکرد روش‌ها بر اساس معیارهایی همچون دقت پیش‌بینی، کیفیت برآورد توزیع پسین و زمان محاسباتی مورد ارزیابی قرار گرفته است. نتایج نشان می‌دهد که روش‌های مبتنی بر نمونه‌گیری، به‌ویژه در تقریب دقیق‌تر توزیع پسین، عملکرد مطلوب‌تری دارند و در مقابل، استنباط واریاسیونی با وجود اندکی افت در دقت نسبت به روش‌های نمونه‌گیری، به دلیل سرعت محاسباتی بالا و همگرایی سریع، برآوردهای قابل قبولی ارائه می‌دهد.

واژه‌های کلیدی: شبکه‌های عصبی بیزی، الگوریتم همیلتونی مونت‌کارلو، نمونه‌گیری بدون بازگشت، استنباط واریاسیونی.
رده‌بندی ریاضی: ۶۲M۴۰، ۶۲M۳۰.

۱ مقدمه

دقیق و مبتنی بر شواهد، حیاتی است [۷، ۱۲]. در آمار کلاسیک، مدل‌هایی مانند رگرسیون خطی و مدل‌های تعمیم‌یافته خطی [۱۴] برای تحلیل داده‌ها توسعه یافته‌اند. این مدل‌ها روابط میان متغیرها را خطی فرض کرده و غالباً بر توزیع نرمال یا اعضای خانواده نمایی برای خطاها تکیه دارند. اگرچه این فرض‌ها در بسیاری از موارد سودمند هستند، اما هنگام مواجهه با روابط غیرخطی یا داده‌های با ساختار پیچیده و نویز بالا، کارایی خود را از دست می‌دهند. برای رفع این محدودیت‌ها، رویکردهای انعطاف‌پذیرتری مانند مدل‌های افزایشی تعمیم‌یافته [۹]، فرآیندهای نرمال [۱۶] و جنگل‌های تصادفی [۴] پیشنهاد شده‌اند. با این حال، این روش‌ها نیز در مقیاس‌های

داده‌های غیرخطی در بسیاری از حوزه‌های علمی و کاربردی، از جمله علوم داده، یادگیری ماشین، اقتصاد، مهندسی زیستی و مدل‌سازی، نقشی اساسی ایفا می‌کنند [۱۶، ۱۵، ۳]. این داده‌ها معمولاً دارای روابط پیچیده، درهم‌کنش‌های غیرخطی و ساختارهای پنهان هستند که مدل‌سازی آن‌ها با استفاده از روش‌های سنتی دشوار است. از سوی دیگر، در بسیاری از کاربردهای عملی علاوه بر پیش‌بینی، برآورد میزان عدم قطعیت نیز اهمیت ویژه‌ای دارد؛ به‌ویژه در حوزه‌هایی مانند پزشکی، امور مالی و سامانه‌های ایمنی حساس که تصمیم‌گیری

الگوریتم همیلتونی مونت‌کارلو، نمونه‌گیری بدون بازگشت و استنباط واریاسیونی پیاده‌سازی می‌شود. هدف اصلی این مطالعه، روشن‌سازی نقاط قوت و ضعف هر یک از این روش‌ها و فراهم‌آوردن تحلیلی تجربی برای بهبود مدل‌سازی در مسائل غیرخطی است. ساختار مقاله به این صورت است که در بخش دوم، مبانی شبکه‌های عصبی و شبکه‌های عصبی بیزی معرفی می‌شود. در بخش سوم، مدل رگرسیون غیرخطی بیزی و چارچوب ریاضی روش‌های استنباطی مورد استفاده شامل الگوریتم‌های NUTS HMC^۱ و استنباط واریاسیونی ارائه می‌گردد. در بخش چهارم، نتایج مطالعات شبیه‌سازی و مقایسه عملکرد روش‌ها بر اساس معیارهای دقت پیش‌بینی، پوشش بازه‌های باورپذیر و هزینه محاسباتی بررسی می‌شود. سپس در بخش پنجم، عملکرد مدل بر روی داده‌های واقعی ارزیابی شده و در نهایت، جمع‌بندی و پیشنهادهایی برای پژوهش‌های آینده ارائه می‌شود.

۲ شبکه‌های عصبی بیزی

شبکه‌های عصبی مصنوعی^۵ (ANN) یکی از مهم‌ترین مدل‌های یادگیری ماشین هستند که با الهام از سازوکار پردازش اطلاعات در مغز انسان طراحی شده‌اند. این شبکه‌ها از مجموعه‌ای از گره‌ها یا نورون‌ها تشکیل شده‌اند که در لایه‌های متوالی سازمان می‌یابند. هر نورون، سیگنال‌های ورودی را دریافت کرده و پس از ترکیب آن‌ها با وزن‌ها و مقدار اریبی متناظر، خروجی خود را از طریق یک تابع فعال‌سازی به لایه بعد ارسال می‌کند. ترکیب و تکرار این لایه‌ها باعث می‌شود شبکه قادر به استخراج الگوهای پیچیده از داده‌ها باشد و در مسائل گوناگون مانند طبقه‌بندی، پیش‌بینی و بازشناسی الگو عملکرد موفقی داشته باشد. در شکل ۱ ساختار کلی یک شبکه عصبی کلاسیک نشان داده شده است. در سمت چپ، ورودی‌ها قرار دارند که مقادیر داده اولیه را به شبکه منتقل می‌کنند. این ورودی‌ها پس از وزن‌دهی و جمع با اریبی‌ها، وارد لایه پنهان می‌شوند. نورون‌های لایه پنهان وظیفه استخراج ویژگی‌ها و روابط غیرخطی میان داده‌ها را بر عهده دارند. در نهایت، سیگنال پردازش شده به لایه خروجی منتقل می‌شود تا مقدار یا دسته نهایی پیش‌بینی گردد. شبکه‌های عصبی کلاسیک به‌طور معمول وزن‌ها و اریبی‌ها را به‌عنوان مقادیر قطعی در نظر می‌گیرند. چنین فرضی موجب می‌شود که این

بزرگ یا هنگام وجود وابستگی‌های پیچیده و چندمتغیره با چالش‌هایی نظیر هزینه محاسباتی سنگین یا محدودیت در قابلیت تعمیم مواجه می‌شوند. در دهه‌های اخیر، شبکه‌های عصبی به دلیل توانایی بالای خود در مدل‌سازی روابط غیرخطی و یادگیری نگاشت‌های پیچیده، جایگاه برجسته‌ای در الگوریتم‌های یادگیری ماشین یافته‌اند [۸]. با وجود این، رویکردهای کلاسیک آموزش شبکه‌های عصبی، مانند بهینه‌سازی تابع زیان با استفاده از گرادیان کاهشی، معمولاً نقطه‌ای و قطعی هستند و اطلاعاتی درباره عدم قطعیت در برآوردها ارائه نمی‌دهند. این کاستی در کاربردهای حساس، از جمله تشخیص پزشکی و سامانه‌های ایمنی، می‌تواند پیامدهای جدی به همراه داشته باشد [۵]. در این راستا، شبکه‌های عصبی بیزی^۱ (BNNs) معرفی شده‌اند [۱۵، ۶، ۲]. این شبکه‌ها با تعریف توزیع‌های پیشین روی وزن‌ها و اریبی‌ها استنباط آماری و مدل‌سازی عدم قطعیت را ممکن می‌سازند. به عبارت دیگر، BNNs امکان نمونه‌گیری از توزیع‌های پسین را فراهم می‌کنند که نه تنها برآوردی از مقدار هدف، بلکه اطلاعاتی در مورد عدم اطمینان آن نیز ارائه می‌دهد. این ویژگی، آن‌ها را به ابزاری قدرتمند برای مدل‌سازی داده‌های پیچیده تبدیل کرده است [۷، ۱۲]. با این حال، به دلیل پیچیدگی و ابعاد بالای توزیع پسین در شبکه‌های عصبی، استنباط تحلیلی در عمل امکان‌پذیر نیست و باید از روش‌های عددی یا تقریبی استفاده کرد. در این مقاله، سه رویکرد پیشنهاد شده است، روش‌های مبتنی بر نمونه‌گیری مونت‌کارلو مانند الگوریتم همیلتونی مونت‌کارلو^۲ (HMC) [۱۵]، که با استفاده از مشتق لگاریتم پسین حرکت مؤثری در فضای پارامتر ایجاد می‌کنند؛ نسخه پیشرفته و تطبیقی آن یعنی نمونه‌گیری بدون بازگشت^۳ (NUTS) [۱۰]، که به‌طور خودکار پارامترهای حرکتی را تنظیم کرده و نیاز به تعیین دستی طول مسیر ندارد و در نهایت روش‌های استنباط واریاسیونی^۴ (VI) [۱۸، ۱]، که با تبدیل مسأله استنباط به یک مسأله بهینه‌سازی، از نظر محاسباتی بسیار سریع‌تر عمل می‌کنند، هرچند در تقریب توزیع‌های پیچیده یا چندوجهی ممکن است دقت کمتری داشته باشند. با توجه به گسترش روزافزون کاربردهای شبکه‌های عصبی بیزی و توسعه الگوریتم‌های استنباطی، ارزیابی نظام‌مند عملکرد روش‌های مختلف در شرایط غیرخطی واقعی یا شبیه‌سازی شده ضرورتی انکارناپذیر دارد. لذا در این پژوهش، یک مدل رگرسیون غیرخطی مبتنی بر شبکه عصبی بیزی ارائه می‌شود و در چارچوب این مدل، سه

¹Bayesian Neural Networks

²Hamiltonian Monte Carlo

³No-U-Turn Sampler

⁴Variational Inference

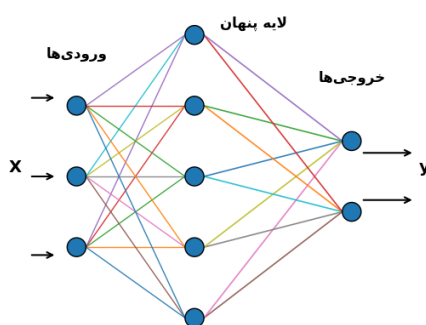
⁵Artificial Neural Networks

تعریف و محاسبه می‌گردد. در شکل ۲، ساختار یک شبکه عصبی بیزی نمایش داده شده است. ورودی‌ها $(x_1, \dots, x_n)$ در سمت چپ قرار دارند که با رنگ زرد مشخص شده‌اند و پس از عبور از چندین لایه پنهان (نقاط آبی) به خروجی y می‌رسند. تفاوت اصلی این ساختار با شبکه‌های عصبی کلاسیک آن است که وزن‌های هر لایه به جای مقادیر ثابت، به صورت متغیرهای تصادفی با توزیع پیشین نرمال $\mathcal{N}(\mu, \sigma^2)$ مدل‌سازی می‌شوند. این امر امکان مدل‌سازی عدم قطعیت را در فرآیند یادگیری فراهم می‌آورد. در سمت چپ شکل، پیشین نرمال نشان داده شده که بیانگر دانش اولیه نسبت به پارامترها پیش از مشاهده داده‌هاست. پس از عبور داده‌ها از لایه‌ها و به‌روزرسانی توزیع پارامترها با استفاده از داده‌های مشاهده‌شده، توزیع پسین در سمت راست شکل حاصل می‌شود که بیانگر دانش به‌روز شده از وزن‌هاست. در بالای شکل مسیر انتقال رو به جلو نمایش داده شده که فرآیند محاسبه خروجی بر اساس ورودی‌ها و وزن‌های احتمالاتی را نشان می‌دهد. در نهایت، این چارچوب بیزی نه تنها مقدار پیش‌بینی را فراهم می‌کند بلکه امکان کمی‌سازی عدم قطعیت در برآوردها را نیز به‌طور طبیعی به همراه دارد.

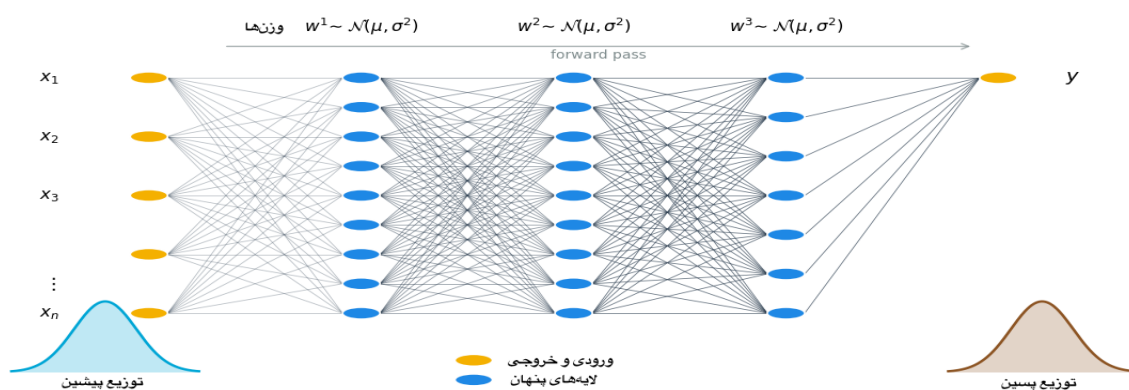
مدل‌ها در شرایطی که داده‌ها با عدم قطعیت و نویز همراه هستند، دچار خطا شده و به‌ویژه در مجموعه داده‌های کوچک یا دارای پراکندگی، با پدیده‌ی بیش‌برازش مواجه شوند. این محدودیت، انگیزه‌ای اساسی برای توسعه‌ی رویکردهای پیشرفته‌تر مانند شبکه‌های عصبی بیزی فراهم کرده است که امکان در نظر گرفتن عدم قطعیت در پارامترها و پیش‌بینی‌ها را به مدل می‌دهند.

۱.۲ شبکه‌های عصبی بیزی

شبکه‌های عصبی بیزی رویکردی هستند که در آن به جای استفاده از مقادیر ثابت برای وزن‌ها و اریبی‌ها، برای این پارامترها توزیع پیشین تعریف می‌شود. به این ترتیب، مدل قادر است حالت‌هایی از پارامترها را نیز در نظر بگیرد که هنوز در داده‌های مشاهده‌شده وجود نداشته‌اند. این ویژگی موجب افزایش توان تعمیم مدل و کاهش احتمال بیش‌برازش در مجموعه داده‌های مختلف می‌شود. در این چارچوب، تمرکز اصلی بر برآورد توزیع پسین پارامترهای شبکه قرار دارد که به صورت رسمی



شکل ۱: نمای شماتیک یک شبکه عصبی مصنوعی (ANN) شامل لایه ورودی، یک لایه پنهان و لایه خروجی.



شکل ۲: نمای شماتیک یک شبکه عصبی بیزی (BNN) با ورودی‌ها، لایه‌های پنهان و خروجی.

۳ تعریف مدل

استنباط دقیق غیرممکن بوده و نیازمند روش‌های عددی یا تقریبی است. در این پژوهش، برای تقریب توزیع پسین از دو رویکرد مبتنی بر نمونه‌گیری (همیلتونی مونت‌کارلو و نسخه‌ی تطبیقی آن، نمونه‌گیری بدون بازگشت) و یک رویکرد مبتنی بر تقریب بهینه‌سازی (استنباط واریاسیونی) استفاده شده است، [۱۷، ۱۱]. برای ارزیابی تجربی، یک مطالعه شبیه‌سازی با ساختار غیرخطی تولید شد و مدل پیشنهادی شبکه عصبی بیزی با استفاده از سه الگوریتم فوق بررسی شد. مقایسه مدل‌ها بر اساس معیارهایی نظیر دقت پیش‌بینی، پوشش فواصل باورمندی و کارایی محاسباتی انجام گرفت.

۱.۳ الگوریتم مونت‌کارلو همیلتونی

یکی از چالش‌های اصلی در استنباط بیزی برای مدل‌های پیچیده‌ای مانند شبکه‌های عصبی بیزی، محاسبه توزیع پسین $p(\theta | D)$ است، که در آن θ شامل پارامترهای مدل شبکه عصبی بیزی (وزن‌ها و اریبی‌ها) بوده و $D = \{(x_i, y_i)\}_{i=1}^n$ مجموعه داده‌های مشاهده‌شده است. از آن‌جا که این توزیع غالباً فرم تحلیلی بسته ندارد و فضای پارامترها نیز با ابعاد بالا و ساختاری پیچیده همراه است، استفاده از روش‌های مبتنی بر زنجیر مارکوفی (MCMC) اجتناب‌ناپذیر است. در میان این روش‌ها، الگوریتم HMC جایگاه ویژه‌ای دارد. این الگوریتم با الهام از مفاهیم مکانیک همیلتونی، حرکت در فضای پارامترها را به‌گونه‌ای طراحی می‌کند که نمونه‌ها بتوانند با سرعت بیشتری به نواحی با احتمال بالا دست یابند. برخلاف روش‌های ساده‌تری چون متروپولیس-هاستینگس که گام‌های تصادفی کوتاه‌برد دارند و منجر به تولید نمونه‌های با همبستگی بالا می‌شوند، HMC با استفاده از گرادیان لگاریتم پسین، مسیریابی پیوسته و هدفمند را طی می‌کند و بدین ترتیب همگرایی سریع‌تر و نمونه‌های کم‌همبسته‌تری به‌دست می‌دهد [۱۵].

در HMC برای هر پارامتر مدل θ_j یک متغیر کمکی موسوم به تکانه^۶ r_j در نظر گرفته می‌شود که معمولاً از یک توزیع نرمال با میانگین صفر و کوواریانس M نمونه‌گیری می‌شود. ترکیب متغیرهای (θ, r) یک فضای حالت همیلتونی ایجاد می‌کند که در آن انرژی کل برابر مجموع انرژی پتانسیل $U(\theta) = -\log p(\theta | D)$ و انرژی جنبشی $K(r) = \frac{1}{2} r^T M^{-1} r$ است. شبیه‌سازی این دینامیک به کمک انتگرال‌گیر لپ‌فراگ^۷ امکان تولید نمونه‌هایی را فراهم می‌کند که به‌طور تقریبی انرژی کل را حفظ می‌کنند. نقطه‌ی پیشنهادی حاصل، سپس با استفاده از معیار پذیرش متروپولیس ارزیابی می‌شود و در صورت

در این بخش رابطه غیرخطی میان متغیرهای ورودی و پاسخ با استفاده از چارچوب شبکه‌های عصبی بیزی مدل‌سازی می‌شود. رویکرد پیشنهادی بر مبنای رگرسیون بیزی قرار دارد، به‌گونه‌ای که نگاشت غیرخطی میان ورودی‌ها و خروجی‌ها از طریق یک شبکه عصبی با پارامترهای تصادفی مدل‌سازی می‌شود. این ساختار علاوه بر انعطاف‌پذیری در مدل‌سازی روابط پیچیده، قابلیت برآورد عدم قطعیت در پیش‌بینی‌ها را نیز فراهم می‌سازد [۱۳]. فرض کنیم داده‌های مشاهده‌شده $D = \{(x_i, y_i)\}_{i=1}^n$ شامل n جفت ورودی-پاسخ باشند، که در آن $x_i \in \mathbb{R}^d$ بردار ویژگی‌ها و $y_i \in \mathbb{R}$ پاسخ متناظر است. مدل پایه به صورت $y_i = f(x_i; \theta) + \varepsilon_i$ ، $i = 1, \dots, n$ تعریف می‌شود، که در آن تابع $f(x_i; \theta)$ خروجی شبکه عصبی بیزی با پارامترهای تصادفی $\theta = \{W^{(\ell)}, b^{(\ell)}\}_{\ell=1}^L$ است، که در آن $W^{(\ell)}$ ماتریس وزن‌های لایه ℓ ، $b^{(\ell)}$ بردار اریبی همان لایه و L تعداد لایه‌های شبکه عصبی است. وزن‌ها شدت و نحوه ارتباط میان نورون‌های لایه‌های متوالی را تعیین می‌کنند، در حالی که اریبی‌ها موجب انتقال و تنظیم تابع فعال‌سازی نورون‌ها می‌شوند. در چارچوب بیزی، این پارامترها به‌صورت متغیرهای تصادفی مدل‌سازی شده و برای آن‌ها توزیع پیشین در نظر گرفته می‌شود. ε_i ها خطاهای تصادفی مستقل و نرمال با میانگین صفر و واریانس ثابت σ^2 هستند. شبکه عصبی بیزی به عنوان یک مدل احتمالاتی برای نگاشت $f: \mathbb{R}^d \rightarrow \mathbb{R}$ تعریف می‌شود. در این چارچوب، توزیع پیشین بر پارامترهای شبکه معمولاً به صورت توزیع نرمال با میانگین صفر و کوواریانس همانی به‌صورت $\theta \sim \mathcal{N}(0, \sigma_\theta^2 I)$ در نظر گرفته می‌شود و مدل بیزی کامل به شکل

$$(y_i | x_i, \theta) \sim \mathcal{N}(f(x_i; \theta), \sigma^2), \quad \theta \sim \mathcal{N}(0, \sigma_\theta^2 I),$$

بیان می‌شود. تابع چگالی پسین پارامترها، $p(\theta | D) = \frac{p(D|\theta)p(\theta)}{p(D)}$ است، که در آن $p(D | \theta)$ تابع درست‌نمایی داده‌ها، $p(\theta)$ توزیع پیشین پارامترها و $p(D)$ توزیع حاشیه‌ای داده‌ها است،

$$p(D) = \int p(D | \theta) p(\theta) d\theta.$$

درست‌نمایی داده‌ها به صورت $p(D | \theta) = \prod_{i=1}^n \psi(y_i | f(x_i; \theta), \sigma^2)$ می‌باشد، که در آن ψ تابع چگالی نرمال با میانگین $f(x_i; \theta)$ و واریانس σ^2 است. از آن‌جا که تابع $f(x; \theta)$ توسط یک شبکه عصبی تعریف شده و توزیع پسین $p(\theta | D)$ به‌صورت تحلیلی قابل محاسبه نیست،

^۶momentum

^۷leapfrog integrator

تطبیقی طول مسیر حرکت را تنظیم کند و در نتیجه همگرایی بهینه‌تر و نمونه‌هایی با خودهمبستگی کمتر نسبت به HMC تولید نماید. در برخی کتابخانه‌های پایتون مانند Pyro، Stan و NumPyro این روش، به‌عنوان روش پیش‌فرض پیاده‌سازی می‌شود [۱۰]. مانند HMC این الگوریتم بر شبیه‌سازی حرکت یک ذره در فضای حالت (θ, r) متکی است، با این تفاوت که در NUTS تعداد گام‌ها به‌طور خودکار تعیین می‌شود. این فرآیند از طریق ساخت یک درخت دودویی از حالات مختلف پارامتر انجام می‌گیرد و گسترش آن تا زمانی ادامه می‌یابد که مسیر حرکت به نقطه بازگشت (U-turn) برسد. در این حالت الگوریتم مسیر را متوقف می‌کند تا از ایجاد نمونه‌های غیرکارآمد جلوگیری کند. بدین ترتیب، NUTS قادر است بدون نیاز به تنظیم دستی پارامترها، به‌طور تطبیقی حرکت در فضای پارامترها را کنترل نماید و همگرایی سریع‌تری نسبت به HMC فراهم آورد. در مدل شبکه عصبی بیزی مورد استفاده در این مطالعه، لگاریتم توزیع پسین به شکل

$$\log p(\theta | \mathcal{D}) = \sum_{i=1}^n \log[\psi(y_i | f(x_i, \theta), \sigma^2)] - \frac{1}{2\sigma_\theta^2} \|\theta\|^2 + C,$$

تعریف می‌شود که بخش اول تابع درست‌نمایی داده‌ها را نشان می‌دهد و بخش دوم لگاریتم توزیع پیشین نرمال است و در آن $\|\theta\|^2$ نرم اقلیدسی مربع‌شده بردار پارامترها است که از رابطه $\|\theta\|^2 = \theta^\top \theta = \sum_{j=1}^p \theta_j^2$ به‌دست می‌آید. مقدار $\sigma^2 = 0.1$ متناسب با سطح نویز افزوده‌شده در فرآیند تولید داده‌های شبیه‌سازی انتخاب شد. همچنین، $\sigma_\theta^2 = 1$ به‌عنوان یک پیشین گاوسی با پراکندگی متوسط برای پارامترهای شبکه در نظر گرفته شد تا نوعی منظم‌سازی ملایم بر وزن‌ها اعمال گردد و از بیش‌برازش جلوگیری شود. در پیاده‌سازی این الگوریتم برای شبکه‌های عصبی بیزی، چند ملاحظه عملی نقش کلیدی دارند. نخست آنکه، به دلیل وجود تعداد زیاد پارامترها در شبکه، محاسبه دستی گرادیان‌ها غیرعملی است؛ از این رو از قابلیت مشتق‌گیری خودکار در کتابخانه‌هایی مانند Pyro استفاده شد که امکان محاسبه دقیق $\nabla_\theta \log p(\theta | \mathcal{D})$ را فراهم می‌سازد. دوم، برای تضمین پایداری و همگرایی، چند صد نمونه ابتدایی صرف مرحله گرم‌سازی^۸ می‌شوند، (گرم‌سازی به مرحله‌ای اطلاق می‌شود که در آن الگوریتم علاوه بر حذف نمونه‌های اولیه، پارامترهای حرکتی مثل طول گام ϵ و ماتریس جرم را هم به‌صورت تطبیقی تنظیم می‌کند). الگوریتم ۲، نحوه پیاده‌سازی الگوریتم NUTS برای مدل شبکه عصبی بیزی است.

پذیرش به زنجیر اضافه می‌گردد. بنابراین انرژی کل سیستم همیلتونی پذیرش به زنجیر اضافه می‌گردد. بنابراین انرژی کل سیستم همیلتونی $H(\theta, r) = U(\theta) + K(r) = -\log p(\theta | \mathcal{D}) + \frac{1}{2} r^\top M^{-1} r$ تعریف می‌شود، که در آن $U(\theta)$ انرژی پتانسیل (منفی لگاریتم توزیع پسین) و $K(r)$ انرژی جنبشی است. الگوریتم با استفاده از روش عددی لپ‌فراگ مسیر حرکت را گسسته‌سازی کرده و با حفظ تقریبی انرژی کل، یک نقطه پیشنهادی (θ^*, r^*) ایجاد می‌کند که با احتمال متروپولیس پذیرفته یا رد می‌شود. در پیاده‌سازی الگوریتم HMC برای مدل شبکه عصبی بیزی محاسبه گرادیان $\nabla_\theta \log p(\theta | \mathcal{D})$ برای پارامترهای بسیار زیاد شبکه (خصوصاً در چندین لایه) به‌صورت دستی غیرعملی است. به همین دلیل، از کتابخانه‌های PyTorch یا Pyro استفاده شده که محاسبه دقیق و خودکار گرادیان‌ها را انجام می‌دهند. همچنین در الگوریتم HMC، انتخاب صحیح گام زمانی ϵ و تعداد گام‌های لپ‌فراگ L در HMC بسیار حساس است. گام کوچک باعث کاهش جابه‌جایی و گام بزرگ ممکن است منجر به واگرایی شود. برای وزن‌ها و اریبی‌ها شبکه، توزیع نرمال با میانگین صفر و واریانس یک انتخاب شده است. به‌منظور حفظ پایداری عددی در اجرای الگوریتم لپ‌فراگ، دقت محاسباتی مناسبی در تنظیم ϵ و استفاده از داده‌های نرمال‌سازی شده به کارگرفته شده است. الگوریتم ۱، نحوه پیاده‌سازی الگوریتم مونت‌کارلوی همیلتونی برای مدل شبکه عصبی بیزی است. ورودی‌ها، داده‌های $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$ ، تابع $f(x, \theta)$ ، توزیع پیشین $p(\theta)$ و در این مطالعه گام زمانی $\epsilon = 0.02$ ، تعداد گام‌های لپ‌فراگ $L = 25$ ، تعداد نمونه $N = 3000$ و ماتریس جرم $M = I$ هستند. خروجی، نمونه‌هایی از توزیع پسین $p(\theta | \mathcal{D})$ خواهد بود. الگوریتم مونت‌کارلوی همیلتونی یکی از روش‌های پایه‌ای و مؤثر در نمونه‌گیری از توزیع‌های پسین در مدل‌های با ابعاد بالا مانند شبکه‌های عصبی بیزی است. اما در عمل، به دلیل حساسیت بالای الگوریتم HMC به پارامترهایی مانند گام زمانی ϵ و تعداد گام‌های لپ‌فراگ L ، استفاده از نسخه‌ی تطبیقی آن یعنی الگوریتم NUTS رایج‌تر است. بنابراین، در این مطالعه برای مقایسه الگوریتم‌های نمونه‌گیری، از نسخه‌ی بهینه‌شده HMC، یعنی NUTS نیز استفاده شده است. این دو الگوریتم از نظر تئوری مشابه هستند، اما NUTS با انتخاب خودکار پارامترهای حرکتی، بهره‌وری بیشتری در مسائل واقعی دارد [۱۰]. الگوریتم NUTS توانسته جایگاه ویژه‌ای در استنباط بیزی برای مدل‌های پیچیده‌ای مانند شبکه‌های عصبی بیزی پیدا کند. مهم‌ترین ویژگی این الگوریتم، حذف نیاز به تعیین دستی پارامترهای حساس HMC یعنی گام زمانی ϵ و تعداد گام‌های لپ‌فراگ L است. این قابلیت موجب می‌شود که NUTS به‌صورت

الگوریتم ۱ استنباط HMC برای شبکه عصبی بیزی

- مقدار دهی اولیه: $\theta^{(0)}$ ، طول گام ϵ ، تعداد گام‌ها L و ماتریس جرم $M = I$.

- برای هر تکرار $n = 1 \dots N$:

۱. نمونه‌گیری تکانه: $r^{(n)} \sim \mathcal{N}(0, M)$.

۲. مقداردهی وضعیت: $(\theta, r) \leftarrow (\theta^{(n-1)}, r^{(n)})$.

۳. نیم‌گام اول برای تکانه:

$$r \leftarrow r + \frac{\epsilon}{2} \nabla_{\theta} \log p(\theta | \mathcal{D}),$$

که در آن $p(\theta | \mathcal{D}) \propto p(\mathcal{D} | \theta)p(\theta)$.

۴. برای $l = 1 \dots L$:

$$\theta \leftarrow \theta + \epsilon M^{-1} r \bullet$$

• اگر $l < L$: $r \leftarrow r + \epsilon \nabla_{\theta} \log p(\theta | \mathcal{D})$.

۵. نیم‌گام نهایی برای تکانه:

$$r \leftarrow r + \frac{\epsilon}{2} \nabla_{\theta} \log p(\theta | \mathcal{D}).$$

۶. محاسبه انرژی اولیه (H_{init}) و نهایی (H_{prop}):

$$H(\theta, r) = -\log p(\theta | \mathcal{D}) + \frac{1}{2} r^{\top} M^{-1} r.$$

۷. پذیرش نمونه پیشنهادی با احتمال:

$$\alpha = \min(1, \exp(H_{\text{init}} - H_{\text{prop}})).$$

۸. به‌روزرسانی وضعیت:

$$\theta^{(n)} = \begin{cases} \theta, & \text{با احتمال } \alpha \\ \theta^{(n-1)}, & \text{در غیر این صورت} \end{cases}$$

- در مرحله گرم‌سازی: تنظیم ϵ و M و ادامه نمونه‌گیری اصلی.

الگوریتم ۲ استنباط NUTS برای شبکه عصبی بیزی

- مقداردهی اولیه: $\theta^{(0)}$ ، طول گام ϵ ، ماتریس جرم اولیه $M = I$.

- برای $n = 1$ تا N :

۱. نمونه‌گیری تکانه: $r^{(n)} \sim \mathcal{N}(0, M)$.

۲. مقداردهی اولیه: $(\theta, r) \leftarrow (\theta^{(n-1)}, r^{(n)})$.

۳. ساخت درخت دودویی با شبیه‌سازی گام‌های لپ‌فراگ در دو جهت (جلو و عقب) به‌صورت بازگشتی.

۴. گسترش درخت تا زمانی که شرط U-turn برقرار شود:

• در هر مرحله تعداد گره‌های درخت دو برابر می‌شود.

• شرط توقف (U-turn):

$$(\theta_{\text{new}} - \theta_{\text{old}})^{\top} r_{\text{new}} < 0.$$

۵. انتخاب یکنواخت یک نمونه θ^* از میان حالات معتبر در درخت.

۶. پذیرش نمونه پیشنهادی با احتمال متروپولیس:

$$\alpha = \min(1, \exp(H(\theta^{(n-1)}, r^{(n)}) - H(\theta^*, r^*))),$$

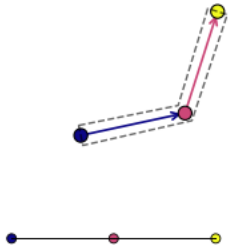
که در آن $H(\theta, r) = U(\theta) + K(r)$.

۷. به‌روزرسانی وضعیت:

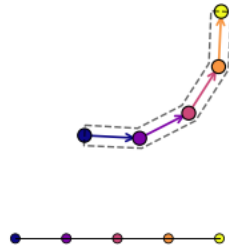
$$\theta^{(n)} = \begin{cases} \theta^*, & \text{با احتمال } \alpha \\ \theta^{(n-1)}, & \text{در غیر این صورت} \end{cases}$$

- در مرحله گرم‌سازی: تنظیم ϵ و تطبیق ماتریس جرم M برای بهبود کارایی نمونه‌گیری.

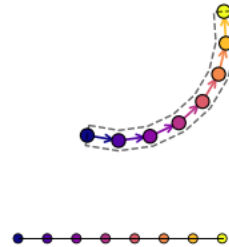
مرحله ۰ (گام‌ها = ۲)



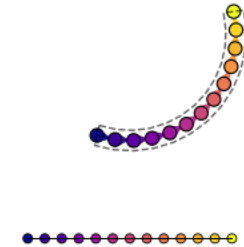
مرحله ۱ (گام‌ها = ۴)



مرحله ۲ (گام‌ها = ۷)



مرحله ۳ (گام‌ها = ۱۲)

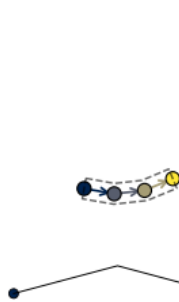


شکل ۳: نمای شماتیک مسیر نمونه‌گیری در الگوریتم HMC

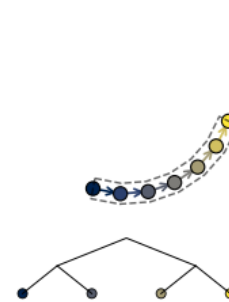
مرحله ۰



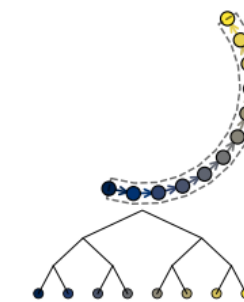
مرحله ۱



مرحله ۲



مرحله ۳



شکل ۴: نمایی از گسترش مسیر در الگوریتم NUTS

از حد بلند باشد، مسیر دچار بازگشت یا U-turn می‌شود که علاوه بر افزایش هزینه محاسباتی، کارایی نمونه‌گیری را کاهش می‌دهد. در مقابل، الگوریتم NUTS این مشکل را برطرف می‌کند. به این روش با گسترش تطبیقی مسیر و استفاده از ساختار درختی، به محض اینکه بازگشت محتمل شود مسیر را متوقف می‌کند. به این ترتیب NUTS نه تنها نیاز به تنظیم دستی طول مسیر ندارد، بلکه نمونه‌گیری کارآمدتر و پایدارتر از توزیع پسین فراهم می‌آورد.

۴ روش استنباط واریاسیونی

استنباط واریاسیونی (VI) یکی از رویکردهای پرکاربرد در تقریب توزیع‌های پسین پیچیده است که به‌ویژه در مدل‌هایی با ابعاد بالا، مانند شبکه‌های عصبی بیزی، اهمیت ویژه‌ای دارد. در حالی که روش‌های نمونه‌گیری نظیر HMC و NUTS به‌طور مستقیم از توزیع پسین نمونه‌برداری می‌کنند، در VI ایده اصلی آن است که این توزیع با یک

در شکل‌های ۳ و ۴ تفاوت اصلی میان HMC و NUTS نمایش داده شده است. در HMC هر مرحله طول مسیر ثابت است، اما تعداد گام‌های انتگرال‌گیری لپ‌فراگ متفاوت در نظر گرفته شده است. پندل‌های بالا مسیر حرکت در فضای حالت را نشان می‌دهند و ردیف پایین توالی حالات متناظر را به‌صورت یک زنجیر خطی نمایش می‌دهد. انتخاب نامناسب اندازه گام یا تعداد گام‌ها می‌تواند به حرکت‌های زائد یا بازگشت منجر شود و کارایی نمونه‌گیری را کاهش دهد. در NUTS هر مرحله طول مسیر با قاعده‌ی دوبرابرشدن افزایش می‌یابد و به‌طور همزمان یک درخت دودویی ساخته می‌شود. مرز خط‌چین محدوده‌ی نقاط پیشنهادی در آن مرحله را نمایش می‌دهد. الگوریتم هنگام تشخیص بازگشت (U-turn) مسیر را متوقف کرده و نمونه‌ی معتبر را انتخاب می‌کند؛ بدین‌ترتیب طول مسیر بدون نیاز به تنظیم دستی، به‌صورت تطبیقی تعیین می‌شود. یعنی در روش HMC، کاربر باید از ابتدا طول مسیر نمونه‌گیری را مشخص کند. اگر این طول خیلی کوتاه باشد، فضای پارامترها به‌طور کامل کاوش نمی‌شود و اگر بیش

الگوریتم ۳ استنباط واریاسیونی برای شبکه عصبی بیزی

۱. مقداردهی اولیه پارامترهای توزیع تقریبی:

$$\sigma = \log(1 + \exp(\rho)) \quad \text{به طوری که } \mu^{(0)}, \rho^{(0)}$$

۲. برای T تا $t = 1$:

- محاسبه انحراف معیارها:

$$\sigma^{(t)} = \log(1 + \exp(\rho^{(t-1)}))$$

- نمونه‌گیری از پارامترهای مدل با بازپارامتردهی:

$$\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad \theta^{(t)} = \mu^{(t-1)} + \sigma^{(t)} \odot \epsilon$$

- محاسبه گرادیان کران پایین شواهد (ELBO) با

استفاده از مشتق‌گیری خودکار $\nabla_{\phi} \mathcal{L}(\phi)$ برابر:

$$\nabla_{\phi} \left(\mathbb{E}_{q_{\phi}(\theta)} [\log p(\mathcal{D}|\theta)] - D_{KL}(q_{\phi}(\theta) \| p(\theta)) \right)$$

- به‌روزرسانی پارامترها با الگوریتم آدام:

$$\phi^{(t)} \leftarrow \phi^{(t-1)} + \eta \cdot \text{Adam}(\nabla_{\phi} \mathcal{L}(\phi))$$

۳. بازگرداندن $\phi^{(T)}$ به عنوان پارامترهای توزیع پسین تقریبی نهایی.

توزیع ساده‌تر و پارامتری $q_{\phi}(\theta)$ تقریب زده شود و سپس پارامترهای ϕ به گونه‌ای تنظیم شوند که فاصله بین q و توزیع پسین واقعی $p(\theta | \mathcal{D})$ حداقل گردد. معیار رایج برای اندازه‌گیری این فاصله، واگرایی کولبک لیبلر^۹ است که کمینه‌سازی آن اساس الگوریتم‌های واریاسیونی را تشکیل می‌دهد [۱، ۱۸]. در چارچوب شبکه‌های عصبی بیزی، پارامترهای مدل شامل وزن‌ها و اریبی‌ها هستند که برای آن‌ها پیشین نرمال چندمتغیره $\theta \sim \mathcal{N}(\mathbf{0}, \sigma_{\theta}^2 \mathbf{I})$ در نظر گرفته می‌شود. برای تقریب توزیع پسین نیز از تقریب میانگین-میدان^{۱۰} استفاده می‌شود. در این رهیافت، توزیع تقریبی پسین به صورت حاصل ضرب توزیع‌های مستقل برای هر پارامتر مدل فرض می‌شود، یعنی $q_{\phi}(\theta) = \prod_{i=1}^p q_{\phi_i}(\theta_i)$. در این مطالعه، هر مؤلفه $q_{\phi_i}(\theta_i)$ به صورت توزیع نرمال با میانگین μ_i و واریانس σ_i^2 مدل‌سازی شده است، به منظور تضمین مثبت بودن انحراف معیار $\sigma_i = \log(1 + \exp(\rho_i))$ در نظر گرفته می‌شود که یک نگاشت نرم و پایدار عددی محسوب می‌شود و در نهایت با توجه به این فرضیات $\phi = (\mu, \rho)$ است. بهینه‌سازی در VI با استفاده از کران پایین شواهد^{۱۱} (ELBO) انجام می‌شود که به شکل

$$\text{ELBO}(\phi) = \mathbb{E}_{q_{\phi}(\theta)} [\log p(\mathcal{D} | \theta)] - D_{KL}(q_{\phi}(\theta) \| p(\theta)) \quad (۱)$$

تعریف می‌شود. معیار $D_{KL}(\cdot \| \cdot)$ واگرایی کولبک-لیبلر بین دو توزیع احتمالی را نشان می‌دهد و نماد $\|$ جداکننده دو توزیع در تعریف واگرایی است. بخش اول رابطه (۱) بیانگر امیدریاضی لگاریتم درست‌نمایی داده‌ها تحت توزیع تقریب است و بخش دوم واگرایی KL میان توزیع تقریب و پیشین را نشان می‌دهد که در این حالت فرم تحلیلی بسته دارد و به راحتی قابل محاسبه است. امیدریاضی در جمله اول از طریق ترفند بازپارامتردهی^{۱۲} و گرادیان‌های تصادفی برآورد می‌شود. در این ترفند متغیر تصادفی θ به صورت تابعی قطعی از پارامترهای قابل یادگیری و یک متغیر تصادفی کمکی به صورت $\theta = \mu + \sigma \odot \epsilon$ بازنویسی می‌شود، که در آن $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ و نماد $\odot$ بیانگر ضرب مؤلفه به مؤلفه^{۱۳} بین بردارها است. این بازنویسی باعث می‌شود وابستگی تصادفی مستقیماً به ϵ منتقل شده و امکان محاسبه گرادیان نسبت به پارامترهای μ و σ به صورت پایدار و قابل مشتق‌گیری فراهم گردد. الگوریتم ۳ نحوه پیاده‌سازی الگوریتم VI برای مدل شبکه عصبی بیزی است. در شکل ۵ نمایی شماتیک از روند اجرای استنباط واریاسیونی در شبکه‌های عصبی بیزی نمایش داده شده است.

در این فرآیند، ابتدا یک توزیع تقریبی برای پارامترهای مدل در نظر گرفته می‌شود و سپس با بهینه‌سازی تابع هدف، این توزیع به تدریج به توزیع پسین واقعی نزدیک می‌شود. همچنین، شکل ۶ روند تدریجی نزدیک شدن توزیع تقریبی به توزیع پسین هدف را در چند مرحله متوالی نشان می‌دهد. خطوط هم‌تراز خاکستری بیانگر پسین هدف و بیضی‌ها نمایش‌دهنده توزیع تقریبی در مراحل مختلف هستند. با تکرار فرآیند بهینه‌سازی، توزیع تقریبی به صورت تدریجی به ناحیه‌های با احتمال بالاتر نزدیک می‌شود.

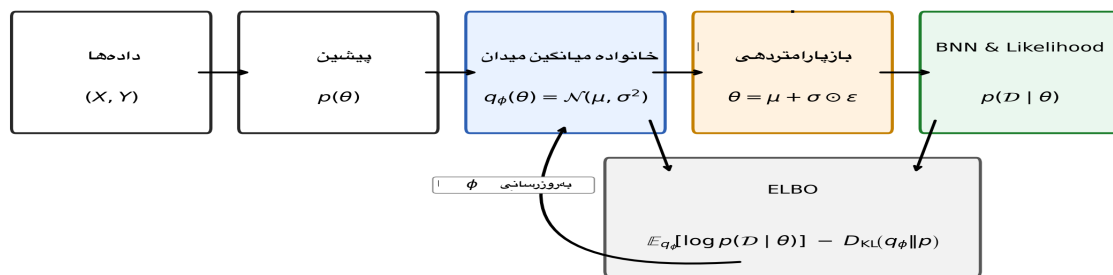
⁹Kullback-Leibler Divergence

¹⁰Mean-Field Approximation

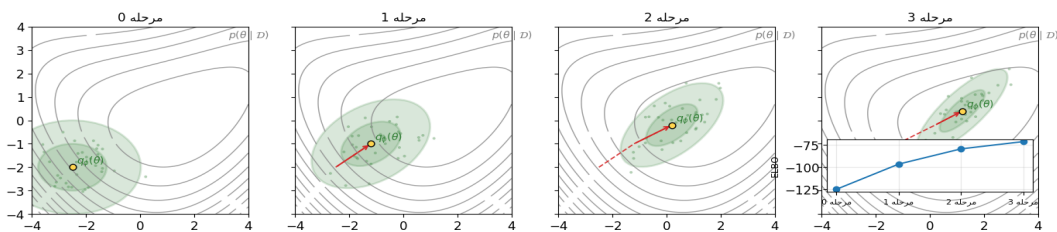
¹¹Evidence Lower Bound

¹²Reparameterization Trick

¹³Hadamard product



شکل ۵: نمای شماتیک روند استنباط واریاسیونی در شبکه‌های عصبی بیزی.



شکل ۶: فرایند استنباط واریاسیونی در چند مرحله متوالی.

استفاده شده که به دلیل داشتن ساختار غیرخطی و پسین قابل نمونه‌گیری، بستر مناسبی برای مقایسه اولیه، MAP، NUTS HMC و VI فراهم می‌کند. در این مدل، $\theta = (\theta_0, \theta_1, \theta_2)^\top$ به جای مقادیر ثابت، به صورت تصادفی با توزیع پیشین نرمال چندمتغیره $\theta \sim \mathcal{N}(0, \mathbf{I}_3)$ مدل‌سازی شدند. ابتدا برآورد ماکسیمم پسینی^{۱۴} (MAP) محاسبه شد. این برآورد، نقطه‌ای از فضای پارامترها را برمی‌گرداند که بیشترین احتمال پسین را دارد:

$$\hat{\theta}_{MAP} = \arg \max_{\theta} p(\theta | D) = \arg \max_{\theta} (p(D | \theta) p(\theta)).$$

این روش را می‌توان تعمیمی از برآورد بیشینه درست‌نمایی دانست که علاوه بر داده‌ها، دانش پیشین را نیز در نظر می‌گیرد. در این مطالعه از الگوریتم^{۱۵} (L-BFGS) برای یافتن نقطه‌ی MAP استفاده شد. این الگوریتم که یکی از کارآمدترین روش‌های شبه‌نیوتنی است، با بهره‌گیری از گرادیان و ذخیره‌سازی محدود اطلاعات ماتریس هسین، امکان بهینه‌سازی در مسائل با ابعاد بالا را فراهم می‌سازد. مزیت اصلی MAP در سرعت و سادگی محاسبات است و می‌تواند به سرعت به یک جواب همگرا شود. با این حال، محدودیت آن در این است که تنها یک

۵ شبیه‌سازی اول: رگرسیون غیرخطی

برای بررسی و مقایسه عملکرد روش‌های مختلف استنباط بیزی در مدل‌سازی روابط غیرخطی، یک مطالعه شبیه‌سازی طراحی شد. داده‌های مصنوعی بر اساس یک تابع سینوسی با نویز نرمال تولید گردیدند به این صورت که ورودی‌ها از بازه $[-3, 3]$ به صورت یکنواخت نمونه‌گیری شدند و خروجی‌ها مطابق رابطه

$$y_i = \sin(x_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 0.01)$$

محاسبه شدند. تعداد نمونه‌ها برابر با $n = 100$ در نظر گرفته شد. شکل ۴ (الف)، داده‌های شبیه‌سازی شده به همراه تابع واقعی را نشان می‌دهد. همان‌طور که دیده می‌شود، داده‌ها ساختار سینوسی را با انحراف ناشی از نویز دنبال می‌کنند و برای ارزیابی روش‌های بیزی در برآورد عدم قطعیت مناسب است. در مطالعه نخست، به منظور ارزیابی کنترل‌شده عملکرد الگوریتم‌های استنباط بیزی، از یک مدل غیرخطی بیزی کم‌بعد با فرم

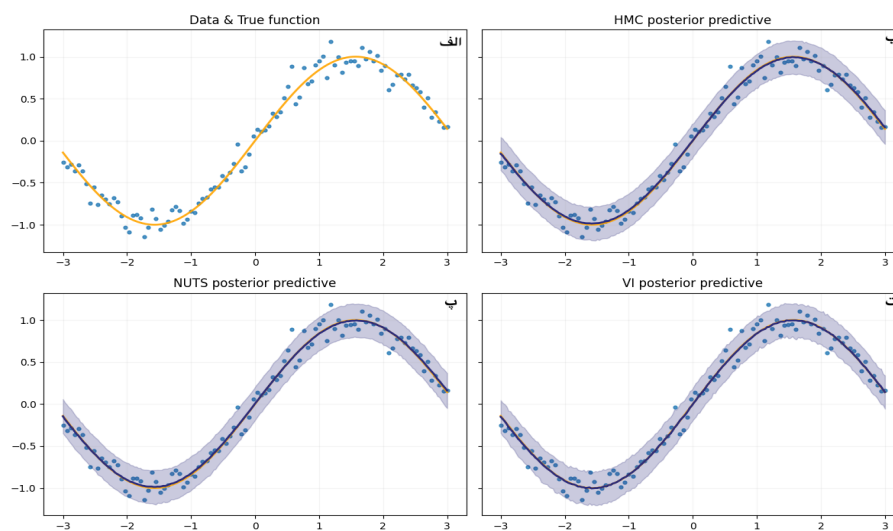
$$f(x; \theta) = \theta_0 + \theta_1 \sin(\theta_2 x)$$

¹⁴Maximum A Posteriori

¹⁵Limited-memory Broyden Fletcher Goldfarb Shanno

و VI بالاتر است. در نهایت، روش استنباط واریاسیونی (VI) برای دستیابی به تقریب سریع‌تر و مقیاس‌پذیرتر به کار گرفته شد. در این رویکرد، توزیع تقریبی $q(\theta)$ به صورت میانگین-میدان نرمال مدل شد و با استفاده از بازپارامتردهی و بهینه‌سازی ELBO برآورد شد. الگوریتم آدام با نرخ یادگیری $\eta = 0.1$ و 3000 تکرار برای بهینه‌سازی ELBO به کار گرفته شد. روش VI توانست به سرعت به توزیعی نزدیک به پسین واقعی برسد و پیش‌بینی‌هایی همراه با نوارهای عدم قطعیت معقول ارائه دهد. شکل ۴ نتایج حاصل از هر سه روش را نشان می‌دهند.

برآورد نقطه‌ای از توزیع پسین ارائه می‌دهد و قادر به نمایش عدم قطعیت پارامتری نیست. در این پژوهش، MAP به عنوان مقداردهی اولیه برای الگوریتم‌های نمونه‌گیری پیچیده‌تر به کار گرفته شد. در مرحله بعد ابتدا از الگوریتم HMC و سپس الگوریتم NUTS برای تقریب کامل توزیع پسین استفاده شد. در این مطالعه، پس از ۵۰۰ گام گرم‌سازی، تعداد ۳۰۰۰ نمونه از توزیع پسین استخراج گردید. نتایج نشان دادند که NUTS قادر به بازنمایی دقیق ساختار توزیع پسین و تولید فواصل باورمند معتبر است، هرچند زمان محاسباتی آن در مقایسه با MAP



شکل ۷: پیش‌بینی پسین شبکه عصبی بیزی با سه رویکرد مختلف، نقاط آبی داده‌های مشاهده‌شده، خط نارنجی تابع واقعی $\sin(x)$ و خطوط آبی تیره میانگین پیش‌بینی هر روش را نشان می‌دهند. ناحیه سایه‌زده بازه باورمندی ۹۵ درصد است.

ساده و داده‌های کافی، حتی روش‌های تقریبی مانند MAP می‌توانند عملکردی نزدیک به روش‌های دقیق‌تر نظیر HMC و NUTS و VI داشته باشند. با این حال، تفاوت اصلی در ماهیت عدم قطعیت است: MAP به صورت ضمنی از نویز مشاهده برای ایجاد بازه پیش‌بینی استفاده می‌کند و فاقد عدم قطعیت پارامتری واقعی است، در حالی که HMC و NUTS برآورد کاملی از پسین ارائه داده و قابلیت اعتماد بالاتری برای تحلیل‌های دقیق دارند. روش VI نیز با هزینه‌ی محاسباتی کمتر نسبت به HMC و NUTS، برآوردهایی نزدیک به آن‌ها به دست داده و بنابراین برای داده‌های بزرگ مقیاس انتخاب مناسبی محسوب می‌شود.

این شکل بیانگر آن است که هر سه روش مورد استفاده توانسته‌اند تابع واقعی $\sin(x)$ را با دقت بالا بازسازی کنند. خطوط پیش‌بینی میانگین تقریباً بر تابع هدف منطبق شده و نوارهای عدم قطعیت نیز در تمامی روش‌ها بسیار مشابه هستند. این نوارها نشان‌دهنده‌ی فاصله‌ی باورمند ۹۵ درصد برای پیش‌بینی هستند که عملاً معادل حدود دو انحراف معیار حول میانگین پیش‌بینی محاسبه شده است. مشاهده می‌شود که داده‌های واقعی به‌طور مناسبی درون این بازه‌ها قرار گرفته‌اند، که این امر نشان‌دهنده‌ی اعتبار برآوردهای بیزی و توانایی روش‌ها در نمایش عدم قطعیت پیش‌بینی است.

این همگرایی نتایج بیانگر آن است که در مسئله‌های با ساختار

جدول ۱: مقایسه عملکرد روش‌های مختلف در برآورد شبکه عصبی بیزی بر اساس داده‌های شبیه‌سازی

روش	MSE	پوشش مشاهدات (۹۵٪)	میانگین پهنای بازه (۹۵٪)	زمان (ثانیه)
HMC	۰/۰۰۰۱۰	۰/۹۵	۰/۳۹۶۳	۵۰/۳۹
NUTS	۰/۰۰۰۱۰	۰/۹۵	۰/۳۹۷۵	۶۰/۹۹
VI	۰/۰۰۰۰۱	۰/۹۶	۰/۳۹۶۴	۱۸/۱۸

یاد گرفته می‌شوند و تنها عدم قطعیت مشاهده (نویز گاوسی در خروجی) مدل می‌شود.

• **BNN-VI**: یک شبکه عصبی بیزی (BNN) که در آن برای وزن‌ها و اریبی‌ها توزیع پیشین گاوسی تعریف شده و توزیع پسین تقریباً با استفاده از استنباط واریاسیونی (VI) برآورد می‌شود. این رهیافت علاوه بر عدم قطعیت مشاهده، عدم قطعیت پارامتری را نیز در نظر می‌گیرد.

یک سناریوی رگرسیون با خطای ناهمواریانس در نظر گرفته شد. ورودی‌ها از بازه $[-3, 3]$ به صورت یکنواخت تولید شدند و خروجی‌ها براساس رابطه

$$y_i = \sin(x_i) + \sigma(x_i) \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1), \quad \sigma(x) = 0.1 + 0.15|x|$$

محاسبه گردید. در این مطالعه نسبت به مطالعه اول، چارچوب شبکه عصبی بیزی کامل‌تر با دو خروجی برای میانگین و انحراف معیار شرطی به‌کار گرفته شد تا هم میانگین و هم ناهمسان‌واریانس داده‌ها مدل‌سازی شوند. برای ارزیابی، یک شبکه نمایشی شامل ۳۰۰ نقطه در بازه $[-3, 3]$ ساخته شد که میانگین پیش‌بینی و باند اطمینان ۹۵ درصد بر روی آن محاسبه گردید.

شبکه طراحی‌شده دارای معماری $2 \rightarrow 20 \rightarrow 20 \rightarrow 1$ است. ورودی شبکه یک بعدی بوده و دو لایه پنهان با ۲۰ نورون و تابع فعال‌سازی $\tanh$ برای استخراج روابط غیرخطی در نظر گرفته شد. لایه خروجی شامل دو نورون است که به ترتیب میانگین $\mu(x)$ و انحراف معیار $\sigma(x)$ را برمی‌گردانند. این معماری امکان مدل‌سازی همزمان میانگین و واریانس مشروط را فراهم می‌کند و برای تحلیل داده‌های ناهمواریانس ضروری است. در مقابل، در شبکه عصبی بیزی وزن‌ها و اریبی‌ها به صورت متغیرهای تصادفی در نظر گرفته شده و برای آن‌ها پیشین نرمال استاندارد تعریف می‌شود. بنابراین، آموزش مدل شامل برآورد توزیع پسین این پارامترها بر اساس داده‌هاست و خروجی شبکه نیز به صورت یک توزیع احتمالی بیان می‌شود. بدین ترتیب، علاوه بر میانگین پیش‌بینی، عدم قطعیت ناشی از پارامترها (عدم قطعیت پارامتری) نیز در نظر گرفته می‌شود.

نتایج به‌دست‌آمده از شبیه‌سازی در جدول (۱) نشان می‌دهد که هر سه روش مورد بررسی توانسته‌اند تابع سینوسی هدف را با دقت بسیار بالا مدل‌سازی کنند، به‌طوری‌که مقدار MSE برای هر سه روش بسیار کوچک به‌دست آمده است. از نظر پوشش بازه‌های باورمندی برای هر سه روش تقریباً در سطح ۹۵ درصد قرار گرفته‌اند، البته VI با پوشش ۹۶ درصد اندکی محافظه‌کارانه‌تر عمل کرده و بازه‌های کمی پهن‌تر ایجاد کرده است. میانگین پهنای بازه برای همه روش‌ها بسیار نزدیک به هم و در حدود ۰/۳۹۶ است که نشان می‌دهد برآوردهای آن‌ها از عدم قطعیت خروجی تفاوت محسوسی با یکدیگر ندارد. تفاوت اصلی در زمان محاسبات مشاهده شده‌است. MAP در کسری از ثانیه اجرا می‌شود و عملاً سریع‌ترین روش است، در حالی که HMC و NUTS به‌ترتیب حدود ۵۰ و ۶۱ ثانیه زمان برده‌اند. روش VI تعادلی مناسب برقرار کرده و در حدود ۱۸ ثانیه به نتیجه رسیده است، در حالی که از نظر دقت و پوشش اختلاف چندانی با HMC و NUTS ندارد. بنابراین می‌توان نتیجه گرفت که برای داده‌های کوچک، HMC و NUTS معتبرترین برآوردها را ارائه می‌دهند، اما برای مسائل بزرگ‌مقیاس یا کاربردهایی که محدودیت زمانی وجود دارد، استنباط VI با توجه به سرعت بالاتر و دقت قابل قبول انتخابی عملی و کارآمد است، در حالی که MAP بیشتر به‌عنوان روشی سریع برای برآورد اولیه پارامترها قابل توصیه است.

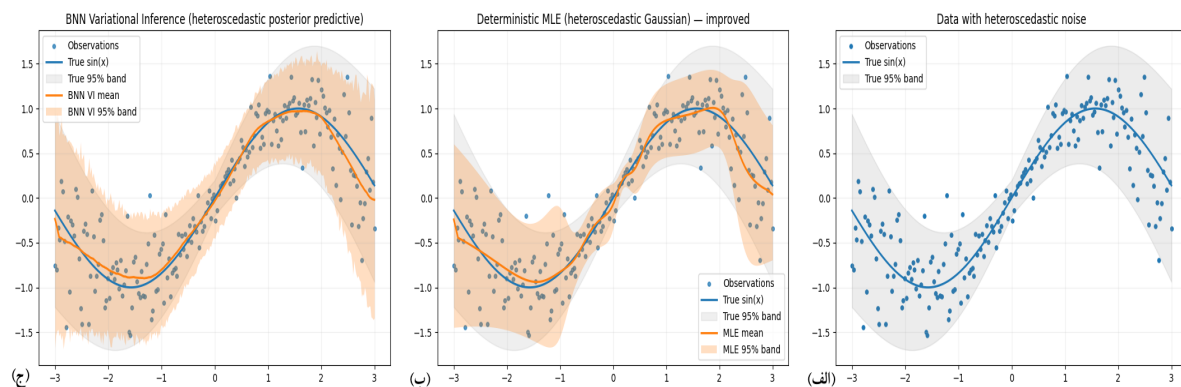
۶ شبیه‌سازی دوم: رگرسیون غیرخطی ناهمسان‌واریانس

در این بخش به منظور بررسی تأثیر مدل‌سازی بیزی در شرایط نویز وابسته به ورودی، یک مطالعه شبیه‌سازی با ساختار ناهمسان‌واریانس طراحی شد. دو رهیافت اصلی مورد مقایسه قرار گرفتند:

• **ANN-MLE**: یک شبکه عصبی کلاسیک (ANN) که با برآورد ماکسیمم درست‌نمایی (MLE) آموزش داده شده است. در این رهیافت، وزن‌ها و اریبی‌ها به صورت پارامترهای ثابت

شکل ۸: نتایج این شبیه‌سازی را نشان می‌دهد. در پنل (الف) داده‌های شبیه‌سازی شده به همراه باند واقعی ناهمواریانسی نمایش داده شده‌اند. پنل (ب) برازش ANN با روش MLE را نشان می‌دهد که اگرچه میانگین تابع سینوسی را به خوبی دنبال می‌کند، اما باندهای پیش‌بینی آن عمدتاً خوش‌بینانه بوده و پهنایی کمتر از پراکنش واقعی داده‌ها دارند. در مقابل، پنل (ج) نتایج BNN با استنباط واریاسیونی را نشان می‌دهد که به دلیل لحاظ کردن عدم قطعیت پارامتری، باندها اندکی وسیع‌تر و واقع‌گرایانه‌تر هستند، به ویژه در نواحی با $|x|$ بزرگ که نویز واقعی افزایش می‌یابد.

تحلیل کمی در جدول ۲ ارائه شده است. مقادیر خطای میانگین مربعات (MSE) نشان می‌دهد هر دو روش قادر به بازسازی دقیق تابع واقعی هستند، اما BNN-VI اندکی دقیق‌تر عمل کرده است. از



شکل ۸: شبیه‌سازی دوم (ناهمواریانسی): مقایسه سه سناریو شامل (الف) داده و باند واقعی، (ب) برازش شبکه عصبی مصنوعی با MLE و (ج) برازش شبکه عصبی بیزی با VI

جدول ۲: مقایسه عملکرد دو روش در سناریوی رگرسیون گوسی ناهمسان‌واریانس

روش	MSE	پوشش مشاهدات (۹۵٪)	میانگین پهنای بازه (۹۵٪)	زمان (ثانیه)
ANN-MLE	۰/۰۱۰۶	۰/۹۵	۱/۱۲۴۳	۱۲/۸۴
BNN-VI	۰/۰۰۹۳	۰/۹۷	۱/۳۵۶۷	۱۱۳۶/۴۰

و آزمایشگاهی بیماران دیابتی را دربر می‌گیرد. متغیر پاسخ، یک شاخص کمی مرتبط با میزان پیشرفت بیماری دیابت است. متغیرهای توضیحی شامل سن، جنسیت، شاخص توده بدنی، میانگین فشار خون، کلسترول کل، LDL، HDL، نسبت کلسترول کل به HDL، سطح لگاریتمی تری‌گلیسیرید سرم و سطح قند خون هستند، این داده‌ها

۷ ارزیابی مدل بر روی داده‌های واقعی

به‌منظور بررسی عملکرد روش‌های استنباط بیزی در عمل، مجموعه داده Diabetes مورد استفاده قرار گرفت. این مجموعه داده شامل ۴۴۲ مشاهده و ۱۰ متغیر توضیحی است که اطلاعات فیزیولوژیک

معیارهای ارزیابی، عملکرد نسبی دو رهیافت HMC و VI را از نظر دقت پیش‌بینی و هزینه محاسباتی به‌صورت شهودی نمایش می‌دهد. لازم به ذکر است در شکل ۹ معیارهای ارزیابی پس از نرمال‌سازی نمایش داده شده‌اند تا امکان مقایسه همزمان معیارهایی با مقیاس‌های متفاوت فراهم شود. در این فرایند، مقدار هر معیار بر بیشینه همان معیار تقسیم شده است، بنابراین مقادیر نمودار در بازه صفر تا یک قرار دارند و صرفاً برای مقایسه نسبی عملکرد روش‌ها استفاده می‌شوند. نتایج حاصل از داده واقعی نشان می‌دهد که الگوی کلی مشاهده‌شده در مطالعه شبیه‌سازی‌شده در یک مجموعه داده واقعی نیز برقرار است. به طور کلی، روش‌های مبتنی بر نمونه‌گیری مانند HMC در صورت تنظیم مناسب می‌توانند دقت بیشتری در تقریب توزیع پسین داشته باشند، در حالی‌که روش‌های واریاسیونی با کاهش هزینه محاسباتی، امکان اجرای سریع‌تر مدل‌های شبکه عصبی بیزی را فراهم می‌کنند. این یافته‌ها نشان می‌دهد که انتخاب روش استنباطی باید متناسب با هدف تحلیل، محدودیت محاسباتی و میزان حساسیت مسئله نسبت به عدم قطعیت انجام شود. برخلاف مدل‌های خطی کلاسیک که روابط میان متغیرها را به صورت خطی فرض می‌کنند، شبکه عصبی بیزی قادر است الگوها و وابستگی‌های غیرخطی موجود میان متغیرهایی مانند شاخص توده بدنی، فشار خون، کلسترول و قند خون با شدت پیشرفت بیماری را یاد بگیرد و در پیش‌بینی مقادیر پاسخ عملکرد مناسبی ارائه دهد. لذا از دیدگاه کاربردی، نتایج نشان می‌دهد که شبکه‌های عصبی بیزی می‌توانند ابزاری مناسب برای تحلیل داده‌هایی مانند داده‌های پزشکی و پیش‌بینی روند بیماری باشند زیرا علاوه بر پیش‌بینی مقدار پاسخ، میزان عدم قطعیت پیش‌بینی را نیز ارائه می‌کنند. این ویژگی باعث می‌شود مدل نه تنها شدت تقریبی پیشرفت بیماری را تخمین بزند، بلکه میزان اطمینان نسبت به هر پیش‌بینی را نیز مشخص کند. در نتیجه، استفاده از مدل‌های رگرسیون غیرخطی بیزی می‌تواند در مسائل پزشکی و تصمیم‌گیری‌های مبتنی بر داده، به‌ویژه در شرایطی که عدم قطعیت داده‌ها اهمیت دارد، مفید و قابل اتکا باشد.

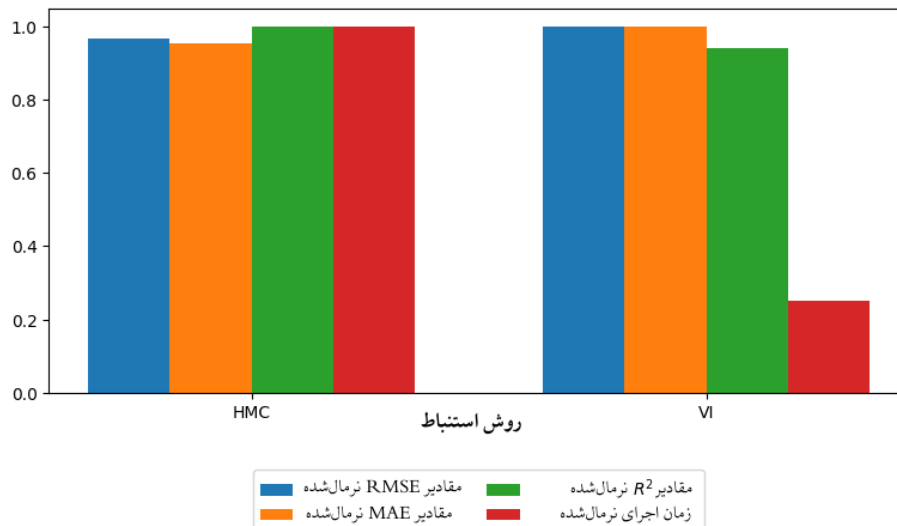
بحث و نتیجه‌گیری

در این پژوهش به ارزیابی سه رویکرد استنباط بیزی در چارچوب شبکه‌های عصبی برای مدل‌های غیرخطی پرداخته شد. برای این منظور دو مطالعه شبیه‌سازی طراحی گردید، نخست رگرسیون غیرخطی با واریانس ثابت و سپس رگرسیون غیرخطی ناهمواریانس بررسی شد.

در کتابخانه scikit-learn در زبان برنامه‌نویسی پایتون و نسخه اصلی داده‌ها در پایگاه Stanford Diabetes Dataset در دسترس قرار دارد. پیش از آموزش مدل، متغیرهای ورودی و متغیر پاسخ نرمال‌سازی شدند تا پایداری عددی الگوریتم‌های استنباطی افزایش یابد. زمان‌های گزارش‌شده در جدول، مربوط به یک اجرای کامل هر الگوریتم بر روی مجموعه داده است و شامل فرآیند آموزش مدل، استنباط پارامترها و تولید نمونه‌های پسین می‌باشد. تمامی آزمایش‌ها تحت شرایط یکسان محاسباتی و با معماری ثابت شبکه عصبی انجام شدند تا مقایسه میان روش‌های استنباطی منصفانه و قابل تفسیر باشد. با توجه به ماهیت تصادفی روش‌های مبتنی بر نمونه‌گیری و مقداردهی اولیه پارامترها، ممکن است زمان اجرا و مقادیر معیارهای خطا در اجراهای مختلف تغییرات جزئی داشته باشند، اما روند کلی نتایج و مقایسه نسبی روش‌ها پایدار باقی می‌ماند. معماری شبکه عصبی بیزی مورد استفاده در این مطالعه شامل یک لایه ورودی با ۱۰ متغیر توضیحی، یک لایه پنهان با ۳ نورون و یک لایه خروجی تک‌متغیره بود. در لایه پنهان از تابع فعال‌ساز غیرخطی tanh استفاده شد تا مدل بتواند روابط غیرخطی میان متغیرهای ورودی و پاسخ را یاد بگیرد. همچنین برای وزن‌ها و اریبی‌های شبکه توزیع‌های پیشین نرمال در نظر گرفته شد و استنباط پارامترها از طریق روش‌های HMC و VI انجام گرفت. انتخاب این معماری با هدف حفظ تعادل میان توان مدل‌سازی غیرخطی و هزینه محاسباتی انجام شد تا اجرای الگوریتم‌های بیزی، به‌ویژه روش‌های مبتنی بر نمونه‌گیری، امکان‌پذیر و پایدار باقی بماند. عملکرد دو روش استنباطی HMC و استنباط واریاسیونی VI مقایسه شد. معیارهای ارزیابی شامل ریشه میانگین مربعات خطا RMSE، میانگین قدرمطلق خطا MAE و زمان اجرا بودند. همان‌گونه که در جدول ۳ مشاهده می‌شود، روش HMC از نظر معیارهای خطای پیش‌بینی عملکرد اندکی بهتر از روش VI داشته است. مقدار کمتر RMSE و MAE در روش HMC نشان می‌دهد که این روش، به دلیل نمونه‌گیری مستقیم‌تر از توزیع پسین، توانسته است تقریب دقیق‌تری از عدم قطعیت پارامترهای شبکه عصبی بیزی ارائه دهد. با این حال، این بهبود دقت با افزایش هزینه محاسباتی همراه بوده است، به‌طوری‌که زمان اجرای HMC به طور محسوسی بیشتر از روش واریاسیونی بوده است. در مقابل، روش VI با وجود دقت اندکی کمتر، از نظر زمان اجرا عملکرد مناسب‌تری داشته است. بنابراین، در شرایطی که دقت بالاتر و تقریب کامل‌تر توزیع پسین اهمیت بیشتری داشته باشد، استفاده از HMC مناسب‌تر است اما در کاربردهایی که سرعت اجرا و مقیاس‌پذیری اهمیت بیشتری دارد، روش VI می‌تواند گزینه عملی‌تری باشد. شکل ۹ مقایسه نرمال‌شده

جدول ۳: مقایسه عملکرد روش‌های HMC و VI بر روی مجموعه داده واقعی Diabetes

روش	RMSE	MAE	زمان اجرا (ثانیه)
HMC	۵۲٫۳۹	۴۱٫۶۱	۸٫۴۳
VI	۵۳٫۰۱	۴۲٫۵۶	۲٫۱۴



شکل ۹: مقایسه نرمال‌شده عملکرد روش‌های HMC و VI بر روی مجموعه داده واقعی Diabetes. در این شکل، معیارهای MAE، RMSE، ضریب تعیین R^2 و زمان اجرا پس از نرمال‌سازی نمایش داده شده‌اند تا امکان مقایسه هم‌زمان معیارهای با مقیاس متفاوت فراهم شود.

واقع‌گرایانه‌تر ارائه دهد که هم عدم قطعیت مشاهده و هم عدم قطعیت پارامتری را در برمی‌گیرد. اگرچه هزینه زمانی اجرای VI بسیار بیشتر از MLE بود، اما دقت بالاتر در پوشش بازه‌ها نشان داد که این روش برای کاربردهایی که مدل‌سازی عدم قطعیت اهمیت ویژه دارد، انتخاب مناسب‌تری محسوب می‌شود.

به‌طور کلی می‌توان نتیجه گرفت که در مسائل ساده و با داده‌های کافی، حتی روش‌های سریع و تقریبی مانند MAP یا MLE می‌توانند برآوردهای قابل قبول ارائه دهند، اما در سناریوهای پیچیده‌تر یا در کاربردهای حساس که عدم قطعیت نقشی اساسی دارد، استفاده از رهیافت‌های بیزی کامل‌تر همچون HMC، NUTS و یا VI ضرورت پیدا می‌کند. انتخاب روش مناسب بستگی مستقیم به موازنه میان دقت، تفسیرپذیری و هزینه محاسباتی دارد.

با توجه به نتایج این مطالعه، پیشنهاد می‌شود در پژوهش‌های آینده به توسعه و ترکیب این رویکردها پرداخته شود، برای مثال استفاده از VI

در مطالعه نخست نشان داده شد که هر سه روش HMC، MAP و NUTS قادرند تابع سینوسی هدف را با دقت بالا بازسازی کنند و باندهای پیش‌بینی مشابهی ایجاد نمایند. تفاوت اصلی در این بود که MAP تنها یک برآورد نقطه‌ای فراهم می‌کند و عدم قطعیت پارامتری را لحاظ نمی‌کند، در حالی که روش‌های نمونه‌گیری، به‌ویژه NUTS تصویری کامل‌تر و معتبرتر از پسین ارائه دادند. روش استنباط واریاسیونی (VI) نیز با وجود تقریب بهینه‌سازی، توانست نتایجی نزدیک به نمونه‌گیری ارائه کند و در عین حال از مزیت سرعت بالاتر برخوردار بود.

در شبیه‌سازی دوم، سناریوی ناهمسان‌وارپانس طراحی شد تا توانایی مدل‌ها در شرایط پیچیده‌تر آزموده شود. در این حالت، رهیافت کلاسیک ANN-MLE گرچه میانگین تابع را به‌درستی بازسازی کرد، اما باندهای پیش‌بینی آن خوش‌بینانه و باریک‌تر از پراکندگی واقعی بودند. در مقابل، شبکه عصبی بیزی با VI توانست باندهایی پهن‌تر و

تقدیر و تشکر

نویسندگان مراتب تقدیر و سپاس خود را به سردبیر محترم، دبیر گرامی و داوران فرزانه‌ای تقدیم می‌دارند که با نظرات دقیق و رهنمودهای ارزشمند خود، موجب غنای هرچه بیشتر این پژوهش گردیدند. از سایر عزیزانی که در انجام این پژوهش همکاری داشته‌اند نیز صمیمانه تشکر و قدردانی می‌شود.

به‌عنوان مقداره‌ی اولیه و سپس بهره‌گیری از الگوریتم‌های همیلتونی برای بهبود دقت نمونه‌گیری. همچنین تعمیم چارچوب ارائه‌شده به مدل‌های پیچیده‌تر، به‌ویژه رگرسیون فضایی بیزی، می‌تواند زمینه‌ای ارزشمند برای بررسی عدم قطعیت در داده‌های مکانی و فضایی فراهم آورد. چنین گسترشی علاوه بر اهمیت نظری، کاربردهای فراوانی در حوزه‌هایی مانند اپیدمیولوژی، علوم محیطی و اقتصاد منطقه‌ای خواهد داشت.

مراجع

- [1] Blei, D.M., Kucukelbir, A., and McAuliffe, J.D. (2017). Variational inference: A review for statisticians, *Journal of the American Statistical Association*, **112**(518): 859-877.
- [2] Blundell, C., Cornebise, J., Kavukcuoglu, K., and Wierstra, D. (2015). Weight Uncertainty in Neural Networks, *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, **37**, 1613-1622.
- [3] Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*, Springer.
- [4] Breiman, L. (2001). Random forests, *Machine Learning*, **45**(1): 5-32.
- [5] Gawlikowski, J., Tassi, C.R.N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., Sheikh, A., Yang, W., Bamler, R., and Ng, A.Y. (2023). A survey of uncertainty in deep learning, *Machine Learning*, **112**: 745-811, Springer.
- [6] Gal, Y., and Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, *International Conference on Machine Learning (ICML)*.
- [7] Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence, *Nature*, **521**(7553): 452-459.
- [8] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*, MIT Press.
- [9] Hastie, T., and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman & Hall/CRC.
- [10] Hoffman, M.D., and Gelman, A. (2014). The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo, *Journal of Machine Learning Research*, **15**(1): 1593-1623.
- [11] Hubin, A., and Storvik, G. (2023). Variational inference for Bayesian neural networks under model and parameter uncertainty, *arXiv preprint arXiv:2305.00934*.
- [12] Kendall, A., and Gal, Y. (2017). What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?, *Advances in Neural Information Processing Systems (NeurIPS)*, **30**, 5574-5584. Curran Associates, Inc.
- [13] Magris, M., and Iosifidis, A. (2023). Bayesian learning for neural networks: an algorithmic survey, *Artificial Intelligence Review*, **56**(10): 11773-11823.

- [14] McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed., Chapman & Hall.
- [15] Neal, R.M. (1996). *Bayesian Learning for Neural Networks*, Springer, Lecture Notes in Statistics.
- [16] Rasmussen, C.E., Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*, MIT Press.
- [17] Rudner, T. G. J., Chen, Z., Teh, Y. W., and Gal, Y. (2023). Tractable function-space variational inference in Bayesian neural networks, *arXiv preprint arXiv:2312.17199*.
- [18] Zhang, C., Butepage, J., Kjellstrom, H., and Mandt, S. (2018). Advances in variational inference, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **41**(8): 2008-2026.

Nonlinear Regression Modeling Using Bayesian Neural Networks and Comparative Analysis of Sampling and Variational Inference

Omid Karimi^{1*}, Fatemeh Hosseini²

^{1,2} *Department of Statistics, Faculty of Mathematics, Statistics and Computer Science, Semnan University, Semnan, Iran*

Received: 2026/01/30

Accepted: 2026/06/28

Abstract

Modeling nonlinear data, due to the presence of complex relationships and latent structures, requires approaches that go beyond classical linear models. In this study, the Bayesian neural network framework is employed to evaluate the performance of different Bayesian inference methods for nonlinear modeling. Three widely used approaches in this area, namely Hamiltonian Monte Carlo, the adaptive No-U-Turn Sampler, and Variational Inference, are implemented and compared within a nonlinear regression framework. The analysis is conducted using both simulated and real-world datasets, and the methods are evaluated based on predictive accuracy, quality of posterior approximation, and computational efficiency. The results indicate that sampling-based methods provide more accurate approximations of the posterior distribution, particularly in terms of uncertainty quantification, whereas variational inference, despite a slight reduction in accuracy, offers competitive performance due to its high computational efficiency and rapid convergence.

Keywords: Bayesian Neural Networks, Hamiltonian Monte Carlo, No-U-Turn Sampler, Variational Inference.

Mathematics Subject Classification: 62M30, 62M40.

*Corresponding author, omid.karimi@semnan.ac.ir