https://www.ui.ac.ir

ریاضی و جامعه
شاپا (چاپی): ۲۳۴۵-۶۴۹۳، شاپا(الکترونیکی): ۲۳۴۵-۶۵۰۷
جلد –، (۱۴۰x) شماره –، صص. xx–xx

https://math-sci.ui.ac.ir

رویکرد یادگیری ماشین برای تقریب مارتینگل بازساز در مسائل مالی با ابعاد بالا

سعید سقایی طوقچی بهاره اختری *

چکیده. مسئله‌ی مدیریت ریسک و قیمت‌گذاری سبدهای مالی، از مسائل مهم در ریاضیات مالی به‌شمار می‌آیند. روش‌های رایج به محاسبه‌ی مستقیم مقدار مورد انتظار شرطی در هر گام زمانی (به عنوان مثال با استفاده از روش‌های مبتنی بر مونت‌کارلو) می‌پردازند. چالش اصلی در این رویکرد، زمانی است که بعد مسئله افزایش می‌یابد و روش‌های شبیه‌سازی از نوع مونت‌کارلو با هزینه‌های محاسباتی سنگین مواجه می‌شوند. در این مقاله با استفاده از مفهوم مارتینگل بازساز که فرآیند ارزش سبد را از طریق اطلاعات نهایی بیان می‌کند، به‌جای محاسبه‌ی مستقیم مقدار مورد انتظار شرطی در هر گام زمانی، ابتدا تابع ارزش نهایی سبد تقریب زده می‌شود و سپس با تکیه بر ویژگی مارتینگل، مقادیر میانی محاسبه می‌شوند. تقریب مذکور با استفاده از روش یادگیری نظارت‌شده و بر پایه‌ی داده‌های شبیه‌سازی‌شده انجام می‌شود. نتایج روش پیشنهادی در مدل‌های گاوسی، با مقایسه‌ی نتایج آن با روش‌های مونت‌کارلوی تودرتو و مونت‌کارلوی کمترین مربعات از نوع رگرسیون–اکنون نشان می‌دهند که از نظر دقت و کارایی محاسباتی عملکرد بهتری دارد.

۱. مقدمه

در سال‌های اخیر، مدیریت ریسک به یکی از ارکان اصلی فعالیت مؤسسات مالی، بانک‌ها و شرکت‌های بیمه تبدیل شده است. ارزش سبدهای مالی و تعهدات بیمه‌ای تحت تأثیر عوامل ریسکی متعددی از جمله ریسک بازار، ریسک اعتباری، ریسک نرخ بهره و ریسک طول عمر قرار دارد. از این‌رو، ارزیابی دقیق توزیع زیان و محاسبه سنجه‌های ریسک مستلزم توسعه روش‌های عددی کارآمد است. با افزایش پیچیدگی محصولات مالی و رشد ابعاد مدل‌های مورد استفاده در عمل، نیاز به روش‌هایی که بتوانند این کمیت‌ها را با دقت و هزینه محاسباتی مناسب برآورد کنند، بیش از پیش احساس می‌شود.

یکی از چالش‌های اصلی در مدیریت ریسک و قیمت‌گذاری محصولات مالی، محاسبه‌ی مقدار مورد انتظار شرطی است که در بسیاری از مسائل مالی به‌عنوان مؤلفه‌ی اصلی ارزش‌گذاری و اندازه‌گیری ریسک ظاهر می‌شود. یکی از متداول‌ترین رویکردها برای این منظور، روش مونت‌کارلوی تودرتو است که در آن برای هر سناریوی بیرونی، مجموعه‌ای از شبیه‌سازی‌های درونی تولید می‌شود. با وجود دقت مناسب، هزینه‌ی محاسباتی بالای این روش، به‌ویژه در مسائل با ابعاد بالا، استفاده از آن را در بسیاری از کاربردهای عملی محدود می‌کند [۲۰، ۱۰].

رده‌بندی موضوعی ریاضیات ۲۰۲۰: 91G20، 91G60، 97M30.
عبارات و کلمات کلیدی: یادگیری ماشین، مارتینگل بازساز، مونت‌کارلوی تودرتو، رگرسیون–اکنون، رگرسیون–آتی.
نوع مقاله: مروری
* نویسنده مسئول
دبیرتخصصی رابط: ——.
تاریخ دریافت: ۱۴۰x/xx/xx تاریخ پذیرش: ۱۴۰x/xx/xx تاریخ انتشار آنلاین: ۱۴۰x/xx/xx
ارجاع به مقاله: نام نویسندگان، عنوان مقاله، ریاضی و جامعه –، (۱۴۰x) شماره –، xx–xx. https://dx.doi.org/ . /msci.xxxx

به‌منظور کاهش این هزینه، روش‌های مونت‌کارلوی کمترین مربعات توسعه یافته‌اند که مقدار مورد انتظار شرطی را از طریق رگرسیون تقریب می‌زنند. ایده‌ی اصلی این روش‌ها نخست در ارزش‌گذاری اختیارمعامله‌های آمریکایی [۲۲] و در مسائل بیمه‌ی عمر به‌کار گرفته شد [۱، ۲]. در چارچوب روش‌های مونت‌کارلوی کمترین مربعات، دو رویکرد اصلی رگرسیون–اکنون و رگرسیون–آتی شکل گرفتند. در رگرسیون–اکنون، مقدار مورد انتظار شرطی مستقیماً بر حسب متغیرهای حالت در زمان فعلی تقریب زده می‌شود در حالی که در رگرسیون–آتی ابتدا پرداخت یا ارزش نهایی برآورد شده و سپس مقدار مورد انتظار شرطی به‌صورت تحلیلی محاسبه می‌شود. مطالعه‌ی [۹] نشان داد که رگرسیون–آتی، در شرایط یکسان، می‌تواند تقریب‌های دقیق‌تری نسبت به رگرسیون–اکنون تولید کند هرچند این مزیت معمولاً با افزایش پیچیدگی محاسباتی و نیاز به ارزیابی مقدار مورد انتظار شرطی توابع تقریب همراه است.

در سال‌های اخیر، پژوهشگران تلاش کرده‌اند با بهره‌گیری از ابزارهای یادگیری ماشین و روش‌های کاهش ابعاد، محدودیت‌های روش‌های مبتنی بر رگرسیون را برطرف کنند. در این راستا، مدل‌های مبتنی بر توابع پایه‌ای قطعه‌ای [۲۴]، توابع اسپلاین [۶]، شبکه‌های عصبی [۳]، انتخاب داده‌محور توابع پایه [۱۱]، روش‌های هسته‌ای [۱۷، ۴] و رگرسیون فرایند گاوسی [۱۴، ۲۶، ۲۱، ۱۳] پیشنهاد شده‌اند.

با وجود پیشرفت‌های حاصل‌شده، استفاده از روش‌های مبتنی بر رگرسیون در مسائل با ابعاد بالا همچنان با چالش‌های جدی روبه‌رو است. در بسیاری از کاربردهای عملی، افزایش تعداد عوامل ریسکی موجب رشد سریع تعداد توابع پایه مورد نیاز برای تقریب می‌شود. علاوه بر این، بسیاری از روش‌های موجود فاقد ساختاری هستند که امکان محاسبه فرم تحلیلی مقدار مورد انتظار شرطی را فراهم کند. در نتیجه، توسعه روش‌هایی که ضمن حفظ دقت، قابلیت استفاده در ابعاد بالا را داشته باشند، همچنان یک مسئله مهم محسوب می‌شود.

در این پژوهش، برای مواجهه با این چالش‌ها، از چارچوب مارتینگل بازساز استفاده می‌کنیم. ایده اصلی این رویکرد، تقریب فرآیند ارزش سبد مالی با استفاده از ترکیب خطی توابع پایه آموزش‌دیده از داده‌ها می‌باشد. در این چارچوب، از ابزارهای یادگیری ماشین برای آموزش توابع تقریب استفاده شده و به‌منظور مقابله با مسئله ابعاد بالا، کاهش ابعاد خطی مبتنی بر داده‌ها به‌کار گرفته می‌شود. این ساختار امکان محاسبه مقدار مورد انتظار شرطی را به فرم تحلیلی برای کلاس مهمی از توابع پایه فراهم می‌کند و در نتیجه کارایی رویکرد رگرسیون–آتی را به‌طور قابل توجهی افزایش می‌دهد.

رویکرد مارتینگل بازساز می‌تواند پلی میان روش‌های کلاسیک مونت‌کارلوی کمترین مربعات و رویکردهای نوین یادگیری ماشین ایجاد کند. این چارچوب، ضمن حفظ تفسیرپذیری ریاضی مدل، امکان بهره‌گیری از تقریب‌های داده‌محور را برای محاسبه‌ی کمیت‌های ریسکی در مسائل با ابعاد بالا فراهم می‌آورد. به‌ویژه نشان خواهیم داد که ترکیب کاهش ابعاد مبتنی بر داده‌ها با توابع پایه مناسب، نه‌تنها مشکل افزایش تعداد پارامترها را تا حد زیادی کنترل می‌کند، بلکه امکان محاسبه‌ی کارآمد مقدار مورد انتظار شرطی را نیز حفظ می‌نماید. برای ارزیابی رویکرد مزبور، عملکرد آن در دو مسئله‌ی شاخص شامل ارزش‌گذاری یک اختیار خرید اروپایی و محاسبه‌ی تعهدات یک محصول بیمه‌ی عمر وابسته به مسیر، بررسی شده و نتایج حاصل با روش‌های متداولی نظیر مونت‌کارلوی تودرتو و مونت‌کارلوی کمترین مربعات مبتنی بر رگرسیون–اکنون مقایسه می‌شود.

در ادامه، ابتدا صورت‌بندی ریاضی مسئله‌ی مارتینگل بازساز و چارچوب یادگیری ماشین ارائه می‌شود. سپس ساختار توابع تقریب، روش‌های عددی و معیارهای ارزیابی تشریح می‌گردند. در بخش مطالعات عددی، رفتار روش پیشنهادی در سناریوهای مالی و بیمه‌ای با ابعاد بالا مورد بررسی قرار گرفته و نقاط قوت و محدودیت‌های آن در مقایسه با رویکردهای مرسوم تحلیل می‌شود. این مسیر، از مبانی نظری تا نتایج محاسباتی، تصویری جامع از ظرفیت‌های رویکرد مارتینگل بازساز برای مدیریت ریسک و ارزش‌گذاری در مسائل پیچیده‌ی دنیای واقعی ارائه می‌کند. جزئیات نظری بیشتر و اثبات‌های مربوط به قضایا در [۷] ارائه شده است.

۲. مسئله‌ی مارتینگل بازساز

بردار رندم $X = (X_1, \dots, X_T)$، یک مولد سناریو اقتصادی با افق زمانی متناهی T است که در آن زمان در واحد سال اندازه‌گیری می‌شود. بردار X که در آن $X_t \in \mathbb{R}^d$ برای $t \in [0, T]$، یک فرآیند محرک تصادفی[1] است و مولفه‌های X_t از یکدیگر مستقل هستند. پالایه‌ی $\mathcal{F}_t := \sigma(X_1, \dots, X_t)$، جریان اطلاعات را نشان می‌دهد. فضای مسیر را به‌صورت $\Omega = \mathbb{R}^{dT}$ در نظر می‌گیریم و توزیع این فرآیند را با نماد $\mathbb{Q}$ نمایش می‌دهیم که آن را اندازه‌ی ریسک–خنثی متناظر می‌نامیم. در این چارچوب تمامی مقادیر مالی و جریان‌های نقدی[2] به‌صورت تنزیل‌شده نسبت به یک دارایی مبنا[3] بیان می‌شوند مگر آنکه خلاف آن تصریح شود.

سبدی از دارایی‌ها[4] و بدهی‌ها[5] که ارزش فعلی آن مجموع جریان‌های نقدی است، در پایان افق زمانی دارای ارزش نهایی ذیل است:

$$f(X) = \sum_{t=1}^{T} \zeta_t, \tag{۱}$$

که در آن $\zeta_t = \zeta_t(X_1, \dots, X_t)$ جریان نقدی تا زمان t بوده که $\mathcal{F}_t$–اندازه پذیر است. فرض می‌کنیم ζ_t و بنابراین تابع f، به‌صورت برون‌زا[6] و از پیش تعیین‌شده، عضو $L^2_{\mathbb{Q}}$ هستند. هدف، یافتن فرآیند ارزش سبد $V := (V_t)_{0 \le t \le T}$ است که به‌صورت ذیل تعریف می‌شود:

$$V_t := \mathbb{E}^{\mathbb{Q}}_t[f(X)] = \underbrace{\sum_{s=1}^{t} \zeta_s}_{\text{جریان نقدی انباشته‌شده تا زمان } t} + \underbrace{\mathbb{E}^{\mathbb{Q}}_t\left[\sum_{s=t+1}^{T} \zeta_s\right]}_{\text{ارزش لحظه‌ای در زمان } t}, \tag{۲}$$

که در آن $\mathbb{E}^{\mathbb{Q}}_t[\cdot] := \mathbb{E}^{\mathbb{Q}}[\cdot \mid \mathcal{F}_t]$ مقدار مورد انتظار شرطی نسبت به اطلاعات تا زمان t است. با استفاده از ویژگی شرطی مکرر داریم:

$$\mathbb{E}^{\mathbb{Q}}[V_{t+1} \mid \mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}}\big[\mathbb{E}^{\mathbb{Q}}[f(X) \mid \mathcal{F}_{t+1}] \,\big|\, \mathcal{F}_t\big] = \mathbb{E}^{\mathbb{Q}}[f(X) \mid \mathcal{F}_t] = V_t,$$

بنابراین V یک مارتینگل در فضای $L^2_{\mathbb{Q}}$ است که مقدار نهایی $f(X)$ را بازسازی می‌کند. در اغلب کاربردهای واقعی، مقدار V_t به‌صورت تحلیلی قابل محاسبه نیست و باید از طریق شبیه‌سازی تابع $f(X)$ تقریب زده شود. با توجه به اینکه برای هر مسیر شبیه‌سازی‌شده، نیاز به ارزیابی تابع داریم، این مرحله می‌تواند از نظر محاسباتی پرهزینه باشد.

یکی از کاربردهای مهم محاسبه‌ی فرآیند V در بحث مدیریت ریسک است. چارچوب‌های نظارتی توانگری مالی[7] در صنعت بیمه و بانکداری، محاسبه‌ی کفایت سرمایه[8] را بر مبنای اندازه‌گیری ریسک تغییرات ارزش سبد $\Delta V_t = V_t - V_{t-1}$ الزامی می‌دانند. در مقررات بیمه، ریسک معمولاً برای افق زمانی یک‌ساله سنجیده می‌شود و سرمایه اقتصادی موردنیاز به‌صورت $\rho[-\Delta V_1]$ محاسبه می‌شود که در آن ρ یکی از معیارهای ریسک است. با توجه به اینکه هدف مدیریت ریسک، اندازه‌گیری زیان است، عبارت $-\Delta V_1$ به‌عنوان زیان بالقوه در نظر گرفته می‌شود. رایج‌ترین معیارهای ریسک شامل ارزش در معرض خطر[9] و زیان مورد انتظار[10] است.

با وجود چارچوب نظری منسجم مارتینگل بازساز، در عمل محاسبه‌ی مقادیر مورد انتظار شرطی و ارزش‌گذاری‌های متناظر غالباً به‌صورت تحلیلی امکان‌پذیر نیست. ازاین‌رو، برای تقریب این مقادیر و برآورد معیارهای ریسک، نیاز به روش‌های عددی از جمله مونت‌کارلوی تودرتو[11] وجود دارد. در بخش بعد، به بررسی این روش و چالش‌های محاسباتی آن می‌پردازیم.

۳. مونت‌کارلو تودرتو و چالش‌های محاسباتی

در روش شبیه‌سازی مونت‌کارلوی تودرتو، هدف تخمین مقادیر مورد انتظار شرطی است که شامل دو مرحله شبیه‌سازی است:

[1] Stochastic Driver Process [2] Cash Account [3] Numeraire [4] Assets [5] Liabilities [6] Exogenously [7] Solvency Regulatory Frameworks [8] Capital Adequency [9] Value at Risk [10] Expected Shortfall [11] Nested Monte Carlo

https://dx.doi.org/ . /msci.xxxx

۱.۳. مرحله‌ی بیرونی. در مرحله‌ی بیرونی، ابتدا مجموعه‌ای از n_0 مسیر مستقل و هم‌توزیع از فرآیند تصادفی محرک X تا افق زمانی t تولید می‌شود. هر مسیر بیرونی را می‌توان به‌صورت ذیل نمایش داد:

$$X_{1:t}^{(i)} = (X_1^{(i)}, X_2^{(i)}, \ldots, X_t^{(i)}), \quad i = 1, \ldots, n_0,$$

۲.۳. مرحله‌ی درونی. برای هر مسیر بیرونی $X_{1:t}^{(i)}$، ارزش سبد مالی در زمان t باید محاسبه شود. بنابراین برای هر مسیر بیرونی، یک مرحله‌ی شبیه‌سازی درونی اجرا می‌شود. این مرحله شامل تولید n_1 مسیر مستقل از ادامه‌ی فرآیند تصادفی پس از زمان t و محاسبه‌ی میانگین نمونه‌ای خروجی تابع ارزش سبد مالی بر این مسیرهاست. به‌عبارت دیگر، اگر f تابع ارزش نهایی باشد آنگاه

$$V_t^{(i)} := \frac{1}{n_1} \sum_{j=1}^{n_1} f\big(X_{t+1:T}^{(i,j)} \mid X_{1:t}^{(i)}\big).$$

پس از تشکیل توزیع تجربی[12] بر اساس مقادیر $\{V_t^{(i)}\}_{i=1}^{n_0}$، زمانی که n_1 به اندازه‌ی کافی بزرگ باشد، افزایش n_0 موجب همگرایی توزیع تجربی به توزیع موردنظر می‌شود. در این حالت، بر اساس قضیه گلیونکو–کانتلی[13]، توزیع تجربی با افزایش تعداد نمونه‌های بیرونی به‌طور یکنواخت به تابع توزیع متناظر همگرا می‌شود. ازاین‌رو با انتخاب مناسب n_0 و n_1، می‌توان از توزیع تجربی به‌عنوان تقریب توزیع موردنظر استفاده کرد.

در این روش، دو منبع اصلی خطای شبیه‌سازی وجود دارد. نخست، خطای ناشی از شبیه‌سازی بیرونی است که به دلیل محدود بودن n_0 ایجاد می‌شود. با افزایش n_0، تعداد نمونه‌های مورد استفاده برای تشکیل توزیع تجربی افزایش یافته و طبق قانون اعداد بزرگ[14]، تقریب توزیع موردنظر دقیق‌تر می‌شود. دوم، برای هر مسیر بیرونی، مقدار $V_t^{(i)}$ از میانگین‌گیری روی n_1 شبیه‌سازی درونی به دست می‌آید. بنابراین این مقدار تنها یک تقریب از امید شرطی متناظر است. افزایش n_1 خطای شبیه‌سازی درونی را کاهش می‌دهد. در نتیجه، دقت نهایی روش مونت‌کارلوی تودرتو به هر دو تعداد n_0 و n_1 وابسته است.

در نهایت، معیار ریسک موردنظر با جایگزینی توزیع تجربی در تعریف نظری آن معیار تخمین زده می‌شود. اگرچه این روش از نظر پیاده‌سازی ساده است اما به‌کارگیری آن در چندین افق زمانی منجر به رشد نمایی تعداد کل شبیه‌سازی‌ها می‌شود. این ویژگی باعث می‌شود این روش تنها در موارد خاص یا برای اعتبارسنجی نتایج به‌کار گرفته شود.

۴. یادگیری ماشین

در سال‌های اخیر، پیشرفت‌های چشمگیر در توان محاسباتی و دسترسی به حجم عظیمی از داده‌ها، موجب شده است که روش‌های یادگیری ماشین به یکی از ابزارهای اصلی در حل مسائل پیچیده علمی، مهندسی و مالی تبدیل شوند. هدف اصلی یادگیری ماشین، استخراج الگوها و روابط موجود در داده‌ها و استفاده از این روابط برای پیش‌بینی، تقریب یا تصمیم‌گیری است. برخلاف روش‌های کلاسیک که در آن‌ها مدل ریاضی از پیش تعیین می‌شود، در یادگیری ماشین ساختار تابع تقریب‌زننده از داده‌ها آموخته می‌شود و مدل قادر است روابط پیچیده و غیرخطی میان متغیرها را شناسایی نماید.

از دیدگاه ریاضی، مسئله یادگیری ماشین را می‌توان به صورت تقریب یک نگاشت ناشناخته در نظر گرفت. فرض کنید $X \in \mathcal{X}$ یک متغیر ورودی و $Y \in \mathcal{Y}$ یک متغیر خروجی باشد. هدف، یافتن تابعی به صورت $f : \mathcal{X} \to \mathcal{Y}$ است که بتواند رابطه میان ورودی و خروجی را تا حد امکان بازسازی نماید. در عمل، این تابع ناشناخته بوده و تنها مجموعه‌ای از مشاهدات $\{(X^{(i)}, Y^{(i)})\}_{i=1}^{N}$ در اختیار قرار دارد. بر اساس این داده‌ها، یک خانواده از توابع انتخاب شده و پارامترهای آن به گونه‌ای تعیین می‌شوند که خروجی مدل با داده‌های مشاهده‌شده بیشترین سازگاری را داشته باشد.

[12]Empirical Distribution [13]Glivenko–Cantelli Theorem [14]Law of Large Numbers

یکی از رایج‌ترین شاخه‌های یادگیری ماشین، یادگیری نظارت‌شده[15]است. در این چارچوب، برای هر مشاهده ورودی، مقدار متناظر خروجی نیز در دسترس است و هدف یادگیری رابطه میان آن‌ها می‌باشد. در حوزه‌ی ریاضیات مالی، پیش‌بینی قیمت دارایی‌ها و ارزش‌گذاری ابزارهای مشتقه در این دسته قرار می‌گیرند.

۱.۴. صورت‌بندی رگرسیون. بسیاری از مسائل یادگیری ماشین را می‌توان به‌صورت یک مسئله رگرسیونی در فضای L^2 تفسیر کرد. فضای L^2 مجموعه‌ای از متغیرهای تصادفی است که دارای گشتاور مرتبه دوم متناهی باشند، یعنی $L^2 = \{X : \mathbb{E}[X^2] < \infty\}$. این فضا یک فضای هیلبرت[16] است و به کمک ضرب داخلی $\langle X, Y\rangle = \mathbb{E}[XY]$ ساختار هندسی مناسبی برای تعریف فاصله، تصویر عمودی و تقریب بهینه فراهم می‌کند. در این چارچوب، هدف یافتن تابعی است که بتواند متغیر هدف را بر اساس اطلاعات موجود در متغیرهای ورودی با کمترین خطای ممکن تقریب بزند. اگر X متغیر ورودی و Y متغیر هدف باشد، مسئله رگرسیون به‌صورت کمینه‌سازی میانگین مربعات خطا[17] تعریف می‌شود:

$$\hat{f} = \arg\min_{f \in L^2} \mathbb{E}\Big[(Y - f(X))^2\Big].$$

با استفاده از ویژگی‌های فضاهای هیلبرت می‌توان نشان داد که جواب این مسئله برابر است با $\hat{f}(X) = \mathbb{E}[Y \mid X]$. بنابراین بهترین تقریب در معنای L^2، همان امید شرطی متغیر هدف نسبت به اطلاعات موجود در متغیرهای ورودی است. از دیدگاه هندسی نیز امید شرطی را می‌توان تصویر عمودی متغیر تصادفی Y بر زیرفضای تولیدشده توسط اطلاعات X در فضای L^2 در نظر گرفت.

برای محاسبه کمیتی به فرم $V_t = \mathbb{E}^{\mathbb{Q}}[f(X_1, \ldots, X_T) \mid \mathcal{F}_t]$ که در آن f یک پرداخت یا کمیت نهایی، $\mathcal{F}_t = \sigma(X_1, \ldots, X_t)$ و X_t متغیرهای حالت در زمان t هستند، دو رویکرد اصلی وجود دارد:

در رگرسیون–اکنون، مقدار مورد انتظار شرطی مستقیماً بر حسب متغیرهای حالت در زمان فعلی تقریب زده می‌شود:

$$V_t(X_1, \ldots, X_t) \approx \sum_{k=1}^{K} \beta_k \phi_k(X_1, \ldots, X_t),$$

که در آن ϕ_k توابع پایه و β_k ضرایب مجهول هستند.

در مقابل، در رگرسیون–آتی ابتدا کمیت نهایی تقریب زده می‌شود:

$$f(X_1, \ldots, X_T) \approx \sum_{k=1}^{K} \beta_k \phi_k(X_1, \ldots, X_T),$$

و سپس با استفاده از خطی بودن امید شرطی خواهیم داشت:

$$V_t \approx \sum_{k=1}^{K} \beta_k \mathbb{E}^{\mathbb{Q}}[\phi_k(X_1, \ldots, X_T) \mid \mathcal{F}_t].$$

بنابراین در رگرسیون–اکنون، تقریب مستقیماً بر روی امید شرطی انجام می‌شود در حالی که در رگرسیون–آتی ابتدا کمیت نهایی تقریب زده شده و سپس ساختار شرطی مدل مورد استفاده قرار می‌گیرد. در این مقاله، محاسبه مارتینگل بازساز را می‌توان نمونه‌ای از رویکرد رگرسیون–آتی در نظر گرفت. زیرا ابتدا کمیت‌های نهایی تقریب زده شده و سپس ساختار شرطی مارتینگل برای بازسازی مقادیر زمانی مورد استفاده قرار می‌گیرد.

[15]Supervised Learning [16]Hilbert Space [17]Mean Squared Error

https://dx.doi.org/ . /msci.xxxx

۲.۴. شبکه‌های عصبی مصنوعی[۱۸]. شبکه‌های عصبی مصنوعی یکی از قدرتمندترین ابزارهای یادگیری نظارت‌شده برای تقریب توابع پیچیده و غیرخطی هستند. ایده اصلی این مدل‌ها از ساختار شبکه‌های عصبی زیستی الهام گرفته شده است، اگرچه مدل‌های ریاضی مورد استفاده در عمل بسیار ساده‌تر از سامانه عصبی واقعی هستند.

یک شبکه عصبی پیش‌خور[۱۹] را می‌توان به عنوان یک تابع پارامتری $f_\theta : \mathbb{R}^d \to \mathbb{R}^m$ در نظر گرفت که در آن θ مجموعه پارامترهای مدل شامل وزن‌ها و اریبی‌ها[۲۰] است. در ساده‌ترین حالت، اگر $x = (x_1, \ldots, x_d)^\top$ بردار ورودی باشد، خروجی اولین لایه پنهان[۲۱] به صورت ذیل تعریف می‌شود:

$$h^{(1)} = \phi\left(W^{(1)}x + b^{(1)}\right)$$

که در آن $W^{(1)}$ ماتریس وزن‌ها، $b^{(1)}$ بردار اریبی و ϕ تابع فعال‌سازی[۲۲] است. توابع فعال‌سازی امکان مدل‌سازی روابط غیرخطی را فراهم می‌کنند و نقش اساسی در توانایی تقریب شبکه‌های عصبی دارند.

در شبکه‌های چندلایه، خروجی هر لایه به عنوان ورودی لایه بعدی مورد استفاده قرار می‌گیرد. اگر شبکه شامل L لایه باشد، خروجی نهایی به صورت $f_\theta(x) = h^{(L)}$ به دست می‌آید که در آن

$$h^{(\ell)} = \phi_\ell\left(W^{(\ell)}h^{(\ell-1)} + b^{(\ell)}\right), \qquad \ell = 1, \ldots, L.$$

وجود چندین لایه پنهان موجب می‌شود که شبکه بتواند ساختارهای پیچیده موجود در داده‌ها شناسایی کند. شبکه‌های دارای تعداد زیادی لایه پنهان را شبکه‌های عصبی عمیق[۲۳] می‌نامند.

یکی از مهم‌ترین ویژگی‌های نظری شبکه‌های عصبی، توانایی آن‌ها در تقریب توابع غیرخطی است. بر اساس ویژگی تقریب‌پذیری سراسری، یک شبکه عصبی پیش‌خور با حداقل یک لایه پنهان و تعداد کافی نورون قادر است هر تابع پیوسته تعریف‌شده روی یک مجموعه فشرده را با دقت دلخواه تقریب زند. اهمیت این ویژگی در آن است که بسیاری از نگاشت‌های مورد نیاز در مسائل مالی، ساختار تحلیلی شناخته‌شده‌ای ندارند و معمولاً تنها از طریق شبیه‌سازی قابل محاسبه هستند. در چنین شرایطی، شبکه‌های عصبی می‌توانند به عنوان ابزارهای انعطاف‌پذیر برای تقریب این توابع مورد استفاده قرار گیرند.

۵. رویکرد یادگیری ماشین در مسئله‌ی مارتینگل بازساز

همان‌گونه که در بخش‌های پیشین مشاهده شد، چالش اصلی در محاسبه مارتینگل بازساز، ناشی از نیاز به برآورد امیدهای شرطی است که محاسبه آن‌ها غالباً به اجرای روش‌های مونت‌کارلوی تودرتو وابسته است. این موضوع به ویژه در مسائل با ابعاد بالا، هزینه محاسباتی قابل توجهی ایجاد می‌کند و استفاده از این روش‌ها را در کاربردهای عملی با محدودیت مواجه می‌سازد. از این رو، یافتن تقریب‌هایی کارآمد، گامی اساسی در جهت توسعه روش‌های محاسباتی به شمار می‌آید. برای دستیابی به این هدف، چارچوبی مبتنی بر یادگیری ماشین پیشنهاد می‌شود که ضمن حفظ دقت مناسب، می‌تواند هزینه محاسباتی مورد نیاز برای محاسبه مارتینگل بازساز را به طور قابل توجهی کاهش دهد.

در این چارچوب، مسئله به‌صورت یک مسئله‌ی رگرسیونی در فضای $L^2_{\mathbb{Q}}$ صورت‌بندی می‌شود. به طور مشخص، هدف یادگیری، تقریب نگاشت $X \mapsto \mathbb{E}_{\mathbb{Q}}[f(X_T) \mid X_t = X]$ است. ورودی مدل شامل نمونه‌هایی از متغیر حالت X در زمان‌های مختلف بوده و خروجی متناظر، مقادیر شبیه‌سازی‌شده‌ی تابع هدف است.

در این رویکرد، تابع ارزش نهایی f به‌صورت مستقیم با تصویر کردن آن بر روی یک زیرفضا با بعد متناهی در فضای $L^2_{\mathbb{Q}}$ تقریب زده می‌شود. این زیرفضا توسط مجموعه‌ای از توابع پایه تولید می‌گردد که همان نگاشت ویژگی[۲۴] را تشکیل می‌دهند. عملکرد این روش به انتخاب مناسب توابع پایه وابسته است. فرض می‌کنیم که مقادیر مورد انتظار شرطی توابع پایه، به‌صورت تحلیلی وجود دارد. این تقریب

[18] Artificial Neural Networks [19] Feedforward Neural Network [20] Biases [21] Hidden Layer [22] Activation Function [23] Deep Neural Networks [24] Feature Map

از طریق آموزش بر مبنای مجموعه‌ای متناهی از مشاهدات X حاصل می‌شود. همچنین مبتنی بر این مشاهدات، یک اندازه‌ی تجربی تعریف می‌شود که برآوردی از اندازه‌ی نظری مدل $\mathbb{Q}$ است که با افزایش تعداد مشاهدات، به اندازه‌ی واقعی مدل همگرا خواهد شد.

۱.۵. تقریب با بعد متناهی[۲۵]

در این روش مجموعه‌ای از توابع پایه به‌صورت $\phi_{\theta,i}: \mathbb{R}^{dT} \to \mathbb{R}$ برای $i = ۱, \dots, m$ و $\theta \in \Theta$ در فضای $L^۲_{\mathbb{Q}}$ تعریف می‌شود. این توابع، نگاشت ویژگی $\phi_\theta = (\phi_{\theta,۱}, \dots, \phi_{\theta,m})^\top$ را تشکیل می‌دهند که ساختار آن وابسته به پارامتر θ است. پارامتر θ معمولاً شامل وزن‌ها و اریبی‌های قابل یادگیری است که در فرآیند آموزش، برای دستیابی به بهترین تقریب انتخاب می‌شوند.

برای هر $\theta \in \Theta$، تصویر تابع f در فضای $L^۲_{\mathbb{Q}}$ در زیرفضای تولیدشده توسط $\{\phi_{\theta,۱}, \dots, \phi_{\theta,m}\}$ به‌صورت $f_\theta = \sum_{i=۱}^m \phi_{\theta,i}\, \beta_{\theta,i} = \phi_\theta^\top \beta_\theta$ قرار می‌گیرد که در آن $\beta_\theta \in \mathbb{R}^m$ بردار ضرایب بهینه است و از طریق حل مسئله‌ی کمینه‌سازی ذیل به‌دست می‌آید:

$$\min_{\beta \in \mathbb{R}^m} \left\| f - \phi_\theta^\top \beta \right\|_{L^۲_{\mathbb{Q}}}. \tag{۳}$$

فرض می‌کنیم که مقدار مورد انتظار شرطی $\mathbb{E}_t^{\mathbb{Q}}[\phi_\theta(X)]$ به‌صورت تحلیلی قابل محاسبه است. اکنون تابع $G_{\theta,t}$ را به‌صورت ذیل تعریف می‌کنیم:

$$\begin{aligned} G_{\theta,t} &: \mathbb{R}^{d\times t} \to \mathbb{R}^m, \\ G_{\theta,t}(x_۱, \dots, x_t) &= \mathbb{E}^{\mathbb{Q}}\big[\, \phi_\theta(x_۱, \dots, x_t, X_{t+۱}, \dots, X_T) \,\big] \\ &= \mathbb{E}_t^{\mathbb{Q}}[\phi_\theta(X)]. \end{aligned} \tag{۴}$$

در نتیجه، فرآیند ارزش سبد تقریب‌زده‌شده به‌صورت ذیل است:

$$\begin{aligned} V_{\theta,t} &= \mathbb{E}_t^{\mathbb{Q}}[f_\theta(X)] \\ &= \mathbb{E}_t^{\mathbb{Q}}[\phi_\theta^\top \beta_\theta] \\ &= G_{\theta,t}(X_۱, \dots, X_t)^\top \beta_\theta \approx V_t. \end{aligned} \tag{۵}$$

در ادامه، فرض می‌کنیم که اندازه‌ی احتمال واقعی $\mathbb{P}$ نسبت به اندازه‌ی ریسک–خنثی $\mathbb{Q}$ مطلقاً پیوسته باشد یعنی $\mathbb{P} \ll \mathbb{Q}$ و مشتق رادون–نیکودیم[۲۶] $\frac{d\mathbb{P}}{d\mathbb{Q}}$ در $L^۲(\mathbb{Q})$ قرار گیرد. اندازه‌ی $\mathbb{Q}$ برای قیمت‌گذاری و تقریب تابع هدف به‌کار می‌رود در حالی که اندازه‌ی $\mathbb{P}$ بیانگر توزیع واقعی بازار بوده و برای ارزیابی ریسک و خطای مسیر مورد استفاده قرار می‌گیرد. بنابراین با تغییر اندازه از $\mathbb{P}$ به $\mathbb{Q}$ و نامساوی کوشی–شوارتز[۲۷] می‌توان خطای مسیر را که در نرم $L^۱_{\mathbb{P}}$ سنجیده می‌شود، به خطای تقریب در نرم $L^۲_{\mathbb{Q}}$ مرتبط کرد. با توجه به اینکه فرآیندهای ارزش $(V_t)_{t=۰}^T$ و $(V_{\theta,t})_{t=۰}^T$ تحت اندازه‌ی $\mathbb{Q}$ مارتینگل هستند، فرآیند $M_t := V_t - V_{\theta,t}$ نیز تحت $\mathbb{Q}$ یک مارتینگل است. بنابراین، با استفاده از نابرابری دوب[۲۸] می‌توان کران بالایی برای خطای بیشینه مسیر به‌دست آورد:

$$\left\| \max_{t\le T} |V_t - V_{\theta,t}| \right\|_{L^۱_{\mathbb{P}}} \le \left\| \frac{d\mathbb{P}}{d\mathbb{Q}} \right\|_{L^۲_{\mathbb{Q}}} \left\| \max_{t\le T} |V_t - V_{\theta,t}| \right\|_{L^۲_{\mathbb{Q}}} \le ۲ \left\| \frac{d\mathbb{P}}{d\mathbb{Q}} \right\|_{L^۲_{\mathbb{Q}}} \|f - f_\theta\|_{L^۲_{\mathbb{Q}}}. \tag{۶}$$

بنابراین هرچه اختلاف بین تابع هدف واقعی f و تقریب آن f_θ کمتر باشد، خطای کل مسیر نیز کاهش می‌یابد. توجه داشته باشید که کمیت $\left\| \frac{d\mathbb{P}}{d\mathbb{Q}} \right\|_{L^۲_{\mathbb{Q}}}$ یک ضریب تغییر اندازه‌ی احتمال است که مقدار آن توسط مدل‌ساز و با توجه به ساختار تغییر اندازه‌ی ریسک قابل محاسبه

[25] Finite-Dimensional Approximation [26] Radon–Nikodym [27] Cauchy–Schwarz [28] Doob's Inequality

https://dx.doi.org/ . /msci.xxxx

است. همچنین عبارت $\|f - f_\theta\|_{L^2_\mathbb{Q}}$ در سمت راست نابرابری، همان تابع هدف مسأله‌ی یادگیری (۳) است. این کمیت را می‌توان با استفاده از شبیه‌سازی مونت‌کارلو و بر اساس داده‌های آموزشی که برای آموزش f_θ به‌کار رفته‌اند، برآورد کرد.

۲.۵. یادگیری ویژگی[۲۹]. در این بخش، علاوه‌بر ضرایب β، پارامتر θ نیز مستقیماً بر اساس داده‌ها برآورد می‌شود. برای این منظور، پارامتر θ به‌گونه‌ای انتخاب می‌شود که خطای تقریب تابع هدف کمینه گردد. در نتیجه، مسئله‌ی بهینه‌سازی (۳) به‌صورت ذیل بروزرسانی می‌شود:

$$\min_{(\theta,\beta)\in\Theta\times\mathbb{R}^m} \left\| f - \phi_\theta^\top \beta \right\|_{L^2_\mathbb{Q}}. \tag{۷}$$

معادله (۷) یک مسئله بهینه‌سازی غیرمحدب[۳۰]است. این در حالی است که برای پارامتر β مسأله همواره محدب باقی می‌ماند. در ادامه، فرضیات لازم برای وجود جواب ارائه می‌شود.

لم ۱.۵. فرض کنید

۱) Θ مجموعه‌ای فشرده است؛

۲) $\theta \mapsto \phi_{\theta,i} : \Theta \to L^2_\mathbb{Q}$ برای همه‌ی iها پیوسته است و

۳) مجموعه‌ی $\{\phi_{\theta,1}, \dots, \phi_{\theta,m}\}$ در $L^2_\mathbb{Q}$ برای همه‌ی $\theta \in \Theta$ مستقل خطی است.

آن‌گاه معادله (۷) دارای جواب است.

اکنون نگاشت ویژگی ϕ_θ را به‌صورت ذیل در نظر می‌گیریم:

$$\phi_{\theta,i}(x) := g_i(A^\top x + b), \quad i = 1, \dots, m. \tag{۸}$$

که در آن $g_i : \mathbb{R}^p \to \mathbb{R}$ توابعی معلوم و $\theta = (A, b)$ شامل یک ماتریس وزن $A \in \mathbb{R}^{dT\times p}$ با رتبه‌ی کامل و یک بردار اریبی $b \in \mathbb{R}^p$ است. به منظور انعکاس ساختار زمانی بردار محرک تصادفی $X \in \mathbb{R}^{dT}$، ماتریس A را به صورت بلوکی نمایش می‌دهیم:

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_T \end{pmatrix}, \qquad A_t \in \mathbb{R}^{d\times p},$$

که در آن هر بلوک A_t متناظر با اطلاعات زمان t است. بنابراین

$$\phi_{\theta,i}(X) = g_i\Big(\sum_{t=1}^{T} A_t^\top X_t + b\Big). \tag{۹}$$

مقادیر مورد انتظار معرفی‌شده در رابطه (۴) به‌صورت ذیل نوشته می‌شوند:

$$\begin{aligned} G_{\theta,t,i}(x_1, \dots, x_t) &= \mathbb{E}^\mathbb{Q}\Big[\, \phi_{\theta,i}(x_1, \dots, x_t, X_{t+1}, \dots, X_T) \,\Big] \\ &= \mathbb{E}^\mathbb{Q}\Big[\, g_i\Big(\sum_{s=1}^{t} A_s^\top x_s + \sum_{s=t+1}^{T} A_s^\top X_s + b\Big) \,\Big]. \end{aligned} \tag{۱۰}$$

در ادامه، سه ساختار مختلف برای نگاشت ویژگی در قالب رابطه‌ی (۸) مورد بررسی قرار خواهد گرفت: پایه چندجمله‌ای کامل[۳۱]، نگاشت چندجمله‌ای همراه با کاهش بعد خطی[۳۲] و شبکه‌ی عصبی کم‌عمق[۳۳].

[29] Feature Learning [30] Non-Convex Optimization [31] Full Polynomial Basis [32] Polynomial Feature Map with Linear Dimensionality Reduction [33] Shallow Neural Network

۳.۵. پایه‌ی چندجمله‌ای کامل. در این بخش، حالت ساده‌ایی را در نظر می‌گیریم که در آن هیچ‌گونه وزن و اریبی به‌عنوان پارامتر قابل یادگیری وجود ندارد. بنابراین مجموعه‌ی پارامترها به صورت $\Theta = \{(I_{dT}, \circ)\}$ تعریف می‌شوند. بنابراین، مسأله‌ی بهینه‌سازی (۷) معادل با مسأله‌ی بهینه‌سازی (۳) است. نگاشت ویژگی ϕ از پایه‌ای تشکیل می‌شود که متعلق به فضای تمام چندجمله‌ای‌های با درجه‌ی حداکثر δ روی $\mathbb{R}^{dT}$ است و به‌صورت ذیل تعریف می‌شود:

$$\mathrm{Pol}_\delta(\mathbb{R}^{dT}) = \mathrm{span}\left\{ x^\alpha \,\middle|\, \alpha \in \mathbb{N}_\circ^{dT},\ |\alpha| \le \delta \right\}, \tag{۱۱}$$

که در آن $\mathbb{N}_\circ^{dT}$ مجموعه‌ی همه‌ی بردارهای $\alpha = (\alpha_۱, \ldots, \alpha_{dT})$ با طول dT و مؤلفه‌های نامنفی است. برای هر α، تک‌عضوی پایه به‌شکل $x^\alpha = x_۱^{\alpha_۱} x_۲^{\alpha_۲} \ldots x_{dT}^{\alpha_{dT}}$ تعریف می‌شود که در آن $\sum_{i=۱}^{dT} \alpha_i \le \delta$. برای اطمینان از این‌که هر تابع پایه ϕ_i عضوی از $L_{\mathbb{Q}}^۲$ باشد، فرض می‌کنیم نرم بردار محرک تصادفی X در $L_{\mathbb{Q}}^{۲\delta}$ متناهی باشد:

$$\mathbb{E}^{\mathbb{Q}}\left[\|X\|^{۲\delta}\right] < \infty. \tag{۱۲}$$

در این مرحله کل مجموعه‌ی تک‌عضوی‌های چندجمله‌ای این فضا برای تقریب تابع هدف مورد استفاده قرار می‌گیرد. اگرچه این ساختار به استخراج فرم بسته‌ای برای کمیت $G_{t,i}$ در رابطه (۱۵) منجر می‌شود، اما تعداد جملات حاصل از بسط چندجمله‌ای با افزایش ابعاد مسئله به‌صورت نمایی رشد می‌کند؛ موضوعی که کارایی محاسباتی روش را در مسائل با ابعاد بالا به‌شدت محدود می‌سازد. به‌طور دقیق‌تر، تعداد توابع پایه از رابطه‌ی ذیل پیروی می‌کند:

$$m = \dim \mathrm{Pol}_\delta(\mathbb{R}^{dT}) = \binom{dT + \delta}{dT}. \tag{۱۳}$$

جدول ۱ نشان می‌دهد که بعد m از اندازه‌ی نمونه‌ی آموزشی در کاربردهای عملی فراتر می‌رود.

جدول ۱. بعد فضای $\mathrm{Pol}_\delta(\mathbb{R}^{dT})$ برای $\delta = ۳$.

T	$d = ۳$	$d = ۵$
۵	۸۱۶	۳۲۷۶
۴۰	۳۰۲۶۲۱	۱۳۷۳۷۰۱

فرم دقیق مقدار مورد انتظار شرطی در معادله (۴) وابسته به انتخاب نگاشت ویژگی ϕ و ساختار توزیع $\mathbb{Q} = \bigotimes_{t=۱}^{T} \mathbb{Q}_t$ برای بردار تصادفی X است. در این ساختار، توزیع کل به‌صورت حاصل‌ضرب مستقل توزیع‌های حاشیه‌ای $\mathbb{Q}_t$ تعریف می‌شود و فضای هیلبرت متناظر نیز به‌صورت ضرب تانسوری فضاهای تابعی تک‌دوره‌ای به‌صورت $L_{\mathbb{Q}}^۲ = \bigotimes_{t=۱}^{T} L_{\mathbb{Q}_t}^۲$ بیان می‌گردد. انتخاب مجموعه‌ای متعامد از چندجمله‌ای‌ها به‌عنوان پایه در این فضای تانسوری، ساختار محاسباتی مقدار مورد انتظار شرطی را به‌طور قابل ملاحظه‌ای ساده می‌کند.

بر اساس قضیه‌ی ۸.۲۵ از [۲۷]، خانواده‌ای از چندجمله‌ای‌های متعامد $\{g_\alpha \mid \alpha \in \mathbb{N}_\circ^{dT}\}$ بر روی فضای $\mathbb{R}^{dT}$ نسبت به اندازه‌ی احتمال $\mathbb{Q}$ وجود دارد. بنابراین یک پایه‌ی متعامد $\{g_\alpha\}$ در $\mathbb{R}^{dT}$ نسبت به اندازه‌ی $\mathbb{Q} = \bigotimes_{t=۱}^{T} \mathbb{Q}_t$ وجود دارد که ساختار آن به‌شکل ذیل تعریف می‌شود:

$$g_\alpha(X) = \prod_{t=۱}^{T} h_{t,\alpha_t}(x_t), \quad X = (x_۱, \ldots, x_T), \quad \alpha = (\alpha_۱, \ldots, \alpha_T), \tag{۱۴}$$

که در آن $x_t \in \mathbb{R}^d$ و $\alpha_t \in \mathbb{N}_\circ^d$. برای هر مؤلفه زمانی $t = ۱, \ldots, T$، $\{h_{t,\alpha_t}\}_{\alpha_t \in \mathbb{N}_\circ^d}$ یک خانواده‌ی متعامد از چندجمله‌ای‌های چندمتغیره روی $\mathbb{R}^d$ نسبت به اندازه‌ی احتمال $\mathbb{Q}_t$ است. در این ساختار، برای هر زمان t، $\alpha_t = (\alpha_{t,۱}, \ldots, \alpha_{t,d})$ یک بردار شاخص با طول d است که نشان می‌دهد در چندجمله‌ای‌های متناظر با زمان t، هر متغیر چه توانی دارد. به‌طور دقیق‌تر، h_{t,α_t} یک چندجمله‌ای

https://dx.doi.org/ . /msci.xxxx

چندمتغیره روی $\mathbb{R}^d$ است که به‌شکل $h_{t,\alpha_t}(x_t) = \prod_{j=1}^{d} p_{t,j}^{(\alpha_{t,j})}(x_{t,j})$ تعریف می‌شود که در آن $p_{t,j}^{(\alpha_{t,j})}$ پایه‌ی متعامد تک‌متغیره درجه $\alpha_{t,j}$ در بعد j و زمان t است. در نتیجه:

$$\deg(g_\alpha) = |\alpha| = \sum_{t=1}^{T} |\alpha_t|, \quad |\alpha_t| = \sum_{j=1}^{d} \alpha_{t,j}, \quad \deg(h_{t,\alpha_t}) = |\alpha_t|.$$

در این ساختار هر پایه متعامد g_α با ترکیب دقیق پایه‌های تک‌متغیره متعامد ساخته می‌شود. بنابراین m تابع پایه‌ی متعامد g_{α_i} $(\phi_i = g_{\alpha_i})$ از فضای $\mathrm{Pol}_\delta(\mathbb{R}^{dT})$ نسبت به اندازه‌ی $\mathbb{Q}$ در نظر گرفته می‌شود. مقادیر مورد انتظار شرطی در رابطه‌ی (۱۰) به‌واسطه‌ی متعامد بودن $\{h_{t,\alpha^i}\}_{\alpha_t}$ ها به‌صورت تحلیلی قابل محاسبه هستند. با توجه به اینکه $A = I_{dT}$، $b = 0$ و نظر به تعامد $\{h_{t,\alpha^i}\}_{\alpha_t}$، مقدار مورد انتظار شرطی هر تابع پایه ϕ_i نسبت به اطلاعات تا زمان t به‌صورت ذیل به‌دست می‌آید:

$$\begin{aligned} G_{t,i}(x_1, \ldots, x_t) &= \mathbb{E}^{\mathbb{Q}}\Big[g_{\alpha^i}\Big(\sum_{s=1}^{t} A_s^\top x_s + \sum_{s=t+1}^{T} A_s^\top X_s + b \Big) \Big] \\ &= \prod_{s=1}^{t} h_{s,\alpha_s^i}(x_s) \times \prod_{s=t+1}^{T} \mathbb{E}^{\mathbb{Q}}[\, h_{s,\alpha_s^i}(X_s)\,] \\ &= \prod_{s=1}^{t} h_{s,\alpha_s^i}(x_s) \cdot \prod_{s=t+1}^{T} \mathbf{1}_{\alpha_s^i = 0}, \end{aligned} \tag{۱۵}$$

که در آن به دلیل تعامد توابع پایه نسبت به اندازه‌ی احتمال $\mathbb{Q}$، هر تابع پایه‌ی غیرثابت نسبت به تابع ثابت متعامد است، بنابراین

$$\mathbb{E}^{\mathbb{Q}}[h_{s,\alpha_s^i}(X_s)] = 0, \qquad \alpha_s^i \neq 0.$$

به‌طور مشابه، مقدار مورد انتظار هر تابع پایه نیز به‌صورت $\mathbb{E}^{\mathbb{Q}}[\phi_i(X)] = \phi_i(0) \cdot \mathbf{1}_{\alpha^i = 0}$ محاسبه می‌شود.

مثال ۲.۵. فرض کنید $X \sim \mathcal{N}(0, I_{dT})$ یک متغیر نرمال چندمتغیره باشد. در این حالت، می‌توان پایه‌های کلی g_α را به‌صورت چندجمله‌ای‌های هرمیت چندمتغیره[۳۴] تعریف کرد که نسبت به توزیع نرمال استاندارد متعامد هستند. به‌ویژه، پایه‌ی کلی با شاخص $\alpha \in \mathbb{N}_0^{dT}$ روی $\mathbb{R}^{dT}$ ساخته می‌شود در حالی که پایه‌ی هر مؤلفه $h_{t,\alpha_t} \equiv h_{\alpha_t}$ یک چندجمله‌ای هرمیت چندمتغیره در $\mathbb{R}^d$ همراه با بردار شاخص $\alpha_t \in \mathbb{N}_0^d$ است.

۴.۵. نگاشت ویژگی چندجمله‌ای همراه با کاهش بعد خطی. همان‌طور که در بخش قبل مشاهده کردیم، استفاده از پایه‌ی کامل چندجمله‌ای باعث افزایش تعداد توابع پایه و در نتیجه ابعاد مسئله می‌شود. در این بخش به راهکارهای لازم جهت کاهش اثر این موضوع می‌پردازیم. فرض می‌کنیم $p \leq dT$ و $\{g_1, \ldots, g_m\}$ یک پایه از فضای $\mathrm{Pol}_\delta(\mathbb{R}^p)$ است. نگاشت‌های ویژگی از نوع معادله (۸) به‌صورت ذیل تعریف می‌شوند:

$$\phi_{(A,b),i}(x) = g_i(A^\top x + b), \quad A \in \mathbb{R}^{dT \times p}, \quad b \in \mathbb{R}^p.$$

همچنین فرض می‌کنیم $\phi_{(A,b),i} \in L^2_{\mathbb{Q}}$. نکته‌ی مهم این است که بدون از دست دادن کلیت می‌توانیم اریبی b را صفر در نظر بگیریم. دلیل این کار آن است که فضای چندجمله‌ای‌ها شامل جمله ثابت است و در نتیجه هر انتقال خطی b را می‌توان در ترکیب پایه‌های چندجمله‌ای پوشش داد. علاوه‌براین، ماتریس وزن A را طوری در نظر می‌گیریم که روی خمینه استیفل[۳۵]

[34] Multivariate Hermite Polynomials [35] Stiefel Manifold

$\mathbb{R}^{dT}$ قرار گیرد. در این حالت، ستون‌های A چارچوبی متعامد از مرتبه p در فضای $V_p(\mathbb{R}^{dT}) = \{ A \in \mathbb{R}^{dT\times p} \mid A^\top A = I_p\}$ است.

قضیه ۳.۵. برای هر ماتریس $A, \tilde{A} \in \mathbb{R}^{dT\times p}$ با رتبه‌ی کامل و بردار $b \in \mathbb{R}^p$، شرایط ذیل معادل‌اند:
۱) $\tilde{A} \in V_p(\mathbb{R}^{dT})$ و $\operatorname{span}\{\phi_{(A,b),1}, \dots, \phi_{(A,b),m}\} = \operatorname{span}\{\phi_{(\tilde{A},0),1}, \dots, \phi_{(\tilde{A},0),m}\}$.
۲) برای ماتریس متعامد $U_{p\times p}$ داریم: $\tilde{A} = A(A^\top A)^{-\frac{1}{2}}U$.

با توجه به قضیه ۳.۵، بدون کاستن از کلیت می‌توان مجموعه‌ی پارامترها را به‌صورت $\Theta = V_p(\mathbb{R}^{dT})$ در نظر گرفت. در واقع، دو ماتریس A و $\tilde{A}$ که تنها به‌وسیله‌ی یک تبدیل متعامد در فضای $\mathbb{R}^p$ به یکدیگر مرتبط‌اند، منجر به یک خانواده‌ی یکسان از توابع پایه‌ای می‌شوند. در این حالت، نگاشت ویژگی به فرم ساده‌ی $\phi_{(A,0),i}(x) := \phi_{A,i}(x) = g_i(A^\top x)$ بازنویسی می‌شود. قضیه بعد به بررسی وجود و یکتایی جواب برای مسأله‌ی بهینه‌سازی در چارچوب این ساختار پارامتری می‌پردازد.

قضیه ۴.۵. برای نگاشت ویژگی چندجمله‌ای، یک نقطه بهینه در مجموعه‌ی $\Theta = V_p(\mathbb{R}^{dT})$ برای مسئله بهینه‌سازی (۷) وجود دارد. در این حالت، یکتایی برقرار نیست. به عبارت دیگر زیرفضای بهینه‌ی $\operatorname{span}\{\phi_{A,1}, \dots, \phi_{A,m}\}$ به‌طور کلی یکتا نیست.

مسأله‌ی (۷) معادل یک مسئله‌ی کاهش بعد خطی روی خمینه‌ی $V_p(\mathbb{R}^{dT}) \times \mathbb{R}^m$ است. در این ساختار، نگاشت خطی $A : \mathbb{R}^{dT} \to \mathbb{R}^p$ صرفاً نقش کاهش‌دهنده‌ی بعد را ایفا می‌کند و هیچ محدودیتی بر فضای $\operatorname{Pol}_\delta(\mathbb{R}^p)$ اعمال نمی‌شود. بنابراین کل مجموعه‌ی پایه‌های $\{g_i\}_{i=1}^m$ در فضای کاهش‌یافته به‌کار گرفته می‌شود. در واقع، بعد کل فضای بهینه‌سازی در این چارچوب برابر است با

$$\dim V_p(\mathbb{R}^{dT}) = dTp - \frac{1}{2}p(p+1) \quad \text{و} \quad m = \dim \operatorname{Pol}_\delta(\mathbb{R}^p) = \binom{p+\delta}{p}.$$

جدول ۴.۵ نشان می‌دهد که بعد کل فضای بهینه‌سازی برای مقادیر مختلف T، d و p در مقایسه با جدول ۱ کمتر است.

جدول ۲. بعد کل: $\dim V_p(\mathbb{R}^{dT}) + \dim \operatorname{Pol}_\delta(\mathbb{R}^p)$ برای $\delta = 3$.

T	$d = 3, p = 3$	$d = 5, p = 10$
5	$39 + 20 = 59$	$195 + 286 = 481$
40	$354 + 20 = 374$	$1945 + 286 = 2231$

محاسبه‌ی مقدار مورد انتظار شرطی در معادله (۴) به سادگی فرم بسته‌ی معادله (۱۵) نیست. بلکه باید مقادیر مورد انتظار مورد بحث (مطابق معادله (۱۰)) به‌صورت ذیل محاسبه شود:

$$G_{A,t,i}(x_1, \dots, x_t) = \mathbb{E}^{\mathbb{Q}}\left[g_i\left(\sum_{s=1}^{t} A_s^\top x_s + \sum_{s=t+1}^{T} A_s^\top X_s \right)\right], \tag{۱۶}$$

که در آن $A \in V_p(\mathbb{R}^{dT})$. ارزیابی معادله (۱۶) در نهایت به محاسبه‌ی گشتاورهای چندمتغیره‌ی متغیر تصادفی $Y \in \mathbb{R}^p$ منتهی می‌شود که در آن $Y = \sum_{s=1}^{t} A_s^\top x_s + \sum_{s=t+1}^{T} A_s^\top X_s$. یکی از ابزارهای ترکیبیاتی مفید در محاسبه‌ی گشتاورهای چندمتغیره، رابطه‌ی بین تک‌جمله‌ای‌های چندمتغیره و ترکیبی از توان‌های جمع خطی متغیرها است که لم ذیل به آن می‌پردازد.

https://dx.doi.org/ . /msci.xxxx

گزاره ۵.۵. [۱۸]. فرض کنید $x_1,\dots,x_n \in \mathbb{R}$ ، $s_1,\dots,s_n \in \mathbb{N}_0$ و $s = s_1 + s_2 + \cdots + s_n$. در این صورت، داریم:

$$x_1^{s_1} x_2^{s_2} \cdots x_n^{s_n} = \frac{1}{s!} \sum_{q_1=0}^{s_1} \cdots \sum_{q_n=0}^{s_n} (-1)^{\sum_{i=1}^n q_i} \prod_{i=1}^n \binom{s_i}{q_i} \Big(\sum_{i=1}^n h_i x_i \Big)^s, \quad h_i = \frac{s_i}{2} - q_i.$$

که در آن عبارت $\Big(\sum_{i=1}^n h_i x_i \Big)^s$ نشان‌دهنده‌ی توان s–ام ترکیب خطی متغیرها است.

برای نمونه، اگر $n = 2$ و $s_1 = s_2 = 1$ باشد، داریم:

$$y_1 y_2 = \frac{1}{2!} \sum_{q_1=0}^{1} \sum_{q_2=0}^{1} (-1)^{q_1+q_2} \binom{1}{q_1} \binom{1}{q_2} (h_1 y_1 + h_2 y_2)^2, \qquad h_i = \tfrac{1}{2} - q_i.$$

با توجه به اینکه $\binom{1}{q} = 1$ برای $q \in \{0,1\}$، چهار حالت ممکن برای (q_1,q_2) را می‌نویسیم:

(q_1,q_2)	$(-1)^{q_1+q_2}$	(h_1,h_2)	$(h_1y_1+h_2y_2)^2$
$(0,0)$	$+1$	$(\frac{1}{2},\frac{1}{2})$	$(\frac{1}{2}y_1+\frac{1}{2}y_2)^2 = \frac{1}{4}(y_1+y_2)^2$
$(1,0)$	-1	$(-\frac{1}{2},\frac{1}{2})$	$(-\frac{1}{2}y_1+\frac{1}{2}y_2)^2 = \frac{1}{4}(y_2-y_1)^2$
$(0,1)$	-1	$(\frac{1}{2},-\frac{1}{2})$	$(\frac{1}{2}y_1-\frac{1}{2}y_2)^2 = \frac{1}{4}(y_1-y_2)^2$
$(1,1)$	$+1$	$(-\frac{1}{2},-\frac{1}{2})$	$(-\frac{1}{2}y_1-\frac{1}{2}y_2)^2 = \frac{1}{4}(y_1+y_2)^2$

$$\begin{aligned} y_1 y_2 &= \frac{1}{2}\Big[\tfrac{1}{4}(y_1+y_2)^2 - \tfrac{1}{4}(y_2-y_1)^2 - \tfrac{1}{4}(y_1-y_2)^2 + \tfrac{1}{4}(y_1+y_2)^2\Big] \\ &= \frac{1}{2}\Big[\tfrac{1}{2}(y_1+y_2)^2 - \tfrac{1}{2}(y_1-y_2)^2\Big] = \frac{1}{4}\Big[(y_1+y_2)^2 - (y_1-y_2)^2\Big], \end{aligned}$$

در ادامه تعمیمی از رابطه‌ی $4y_1y_2 = (y_1+y_2)^2 - (y_1-y_2)^2$ ارائه می‌دهیم.

$$y^\alpha \equiv y_1^{\alpha_1} \cdots y_p^{\alpha_p} = \frac{1}{|\alpha|!} \sum_{\nu=0}^{\alpha} (-1)^{|\nu|} \binom{\alpha_1}{\nu_1} \cdots \binom{\alpha_p}{\nu_p} \left(h_{\alpha,\nu}^\top y \right)^{|\alpha|}, \tag{۱۷}$$

که در آن $h_{\alpha,\nu} = \big(\frac{\alpha_1}{2} - \nu_1, \dots, \frac{\alpha_p}{2} - \nu_p\big)^\top$. تعداد جملات در رابطه‌ی (۱۷) را می‌توان از طریق شمارش مقادیر ممکن برای $\nu = (\nu_1,\dots,\nu_p)$ به‌دست آورد. از این‌رو، تعداد جملات مؤثر برابر با $\frac{(\alpha_1+1)(\alpha_2+1)\cdots(\alpha_p+1)}{2}$ است. بنابراین محاسبه‌ی $\mathbb{E}^{\mathbb{Q}}[Y^\alpha]$ در معادله (۱۶)، به محاسبه‌ی گشتاور مرتبه‌ی $|\alpha|$ متغیرهای تصادفی اسکالر $h_{\alpha,\nu}^\top Y$ در معادله (۱۷) کاهش می‌یابد. این گشتاورها برای توزیع‌های مختلف متغیر تصادفی X به‌صورت تحلیلی در دسترس هستند.

مثال ۶.۵. در حالت نرمال چندمتغیره، اگر $X \sim \mathcal{N}(0, I_{dT})$، آن‌گاه داریم:

$$h_{\alpha,\nu}^\top Y \sim \mathcal{N}\left(\sum_{s=1}^{t} h_{\alpha,\nu}^\top A_s^\top x_s, \sum_{s=t+1}^{T} \|A_s h_{\alpha,\nu}\|^2 \right).$$

این گشتاورهای تک‌متغیره همان‌گونه که در [۱۸] (گزاره ۲) بیان شده است؛ به‌صورت تحلیلی در دسترس هستند.

۵.۵. شبکه عصبی کم‌عمق. سومین ساختار نگاشت ویژگی، شبکه عصبی کم‌عمق (شامل یک لایه‌ی پنهان) با تابع فعال‌سازی رلو[۳۶] است. برخلاف نگاشت‌های چندجمله‌ای که دارای ساختار تابعی از پیش‌تعیین‌شده هستند، شبکه‌های عصبی امکان تقریب خودکار توابع غیرخطی پیچیده را مستقیماً از داده‌ها فراهم می‌کنند. فرض می‌کنیم $p = m$ که در آن p بعد فضای بهینه‌سازی و m تعداد نورون‌های لایه‌ی پنهان است. همچنین تابع g_i به فرم $g_i(y) := y_i^+ = \max\{۰, y_i\}$ برای $i = ۱, \dots, m$ تعریف می‌شود. در نتیجه، نگاشت ویژگی از معادله (۸) به صورت ذیل بازنویسی خواهد شد:

$$g_i(y) := y_i^+ = \max\{۰, y_i\}, \quad i = ۱, \dots, m.$$

بنابراین نگاشت ویژگی در معادله (۸) به شکل ذیل بازنویسی می‌شود:

$$\phi_{(A,b),i}(x) \equiv \phi_{(a_i,b_i)}(x) = (a_i^\top x + b_i)^+, \tag{۱۸}$$

که در آن ماتریس وزن $A = (a_۱, \dots, a_m) \in \mathbb{R}^{dT \times m}$ و بردار اریبی $b \in \mathbb{R}^m$. فرض می‌کنیم شرط (۱۲) برای $\delta = ۱$ برقرار است لذا $\phi_{(a_i,b_i)} \in L^۲_{\mathbb{Q}}$. با توجه به ویژگی همگنی مثبت[۳۷] تابع فعال‌سازی رلو نسبت به پارامترها، یعنی $\phi_{(\lambda a_i, \lambda b_i)} = \lambda \phi_{(a_i,b_i)}$ برای هر $\lambda > ۰$، بدون از دست دادن کلیت، فرض می‌کنیم که بردارهای پارامتر (a_i, b_i) روی کره واحد $\mathbb{S}_{dT} \subset \mathbb{R}^{dT+۱}$ قرار دارند؛ یعنی $\|(a_i, b_i)\| = ۱$. در نتیجه، فضای پارامترها را می‌توان به‌صورت خمینه $\Theta = (\mathbb{S}_{dT})^m$ در نظر گرفت که در آن $(\mathbb{S}_{dT})^m = \underbrace{\mathbb{S}_{dT} \times \cdots \times \mathbb{S}_{dT}}_{m \text{ حاصل‌ضرب}}$ حاصل‌ضرب به‌صورت دکارتی است.

در ارتباط با استقلال خطی توابع $\phi_{(a_i,b_i)}$، قضیه‌ای بنیادی در ادامه ارائه می‌شود که با وجود اهمیت نظری بالا، به‌نظر می‌رسد تاکنون در ادبیات علمی این حوزه کمتر مورد توجه قرار گرفته است.

قضیه ۷.۵. برای هر $(a_i, b_i), (\tilde{a}_i, \tilde{b}_i) \in \mathbb{S}_{dT}$ و $i = ۱, \dots, m$، گزاره‌های ذیل برقرارند:

(۱) اگر $\operatorname{span}\{\phi_{(a_۱,b_۱)}, \dots, \phi_{(a_m,b_m)}\} = \operatorname{span}\{\phi_{(\tilde{a}_۱,\tilde{b}_۱)}, \dots, \phi_{(\tilde{a}_m,\tilde{b}_m)}\}$ آن‌گاه $\{\pm(a_۱, b_۱), \dots, \pm(a_m, b_m)\} = \{\pm(\tilde{a}_۱, \tilde{b}_۱), \dots, \pm(\tilde{a}_m, \tilde{b}_m)\}$.

(۲) اگر برای هر $i \neq j$، $(a_i, b_i) \neq \pm(a_j, b_j)$ آن‌گاه $\{\phi_{(a_۱,b_۱)}, \dots, \phi_{(a_m,b_m)}\}$ مستقل خطی است.

توجه داشته باشید که حکم معکوس در قسمت ۱ از قضیه‌ی ۷.۵ برقرار نیست. علاوه‌براین، فرض قسمت ۲ از قضیه‌ی ۷.۵ را نمی‌توان صرفاً به نابرابری جفتی $(a_i, b_i) \neq (a_j, b_j)$ برای هر $i \neq j$ تقلیل داد.

قضیه ۸.۵. برای شبکه عصبی کم‌عمق با تابع فعال‌سازی رلو، در حالت کلی کمینه‌کننده‌ای برای مسئله‌ی بهینه‌سازی (۷) وجود ندارد. علاوه بر این در صورت وجود جواب نیز، یکتایی برقرار نیست. به این معنا که زیرفضای بهینه $\operatorname{span}\{\phi_{(a_۱,b_۱)}, \dots, \phi_{(a_m,b_m)}\}$ یکتا نیست.

مقدار مورد انتظار شرطی توابع رلو مطابق با رابطه‌ی (۱۰) برای $i = ۱ \dots m$، به‌صورت ذیل محاسبه می‌شود:

$$G_{(a_i,b_i),t}(x_۱, \dots, x_t) = \mathbb{E}^{\mathbb{Q}}\left[\left(b_i + \sum_{s=۱}^{t} a_{i,s}^\top x_s + \sum_{s=t+۱}^{T} a_{i,s}^\top X_s\right)^+\right], \tag{۱۹}$$

که در آن $a_i^\top = \left(a_{i,۱}^\top, a_{i,۲}^\top, \dots, a_{i,T}^\top\right)$ و $a_{i,s}$ بردار وزن مربوط به مرحله‌ی زمانی s است.

[36] ReLU [37] Positive Homogeneity

https://dx.doi.org/ . /msci.xxxx

ارزیابی رابطه‌ی (۱۹) معادل با محاسبه‌ی عبارت $\mathbb{E}^{\mathbb{Q}}[Y^+]$ است که متغیر تصادفی اسکالر Y به‌شکل $Y = b_i + \sum_{s=1}^{t} a_{i,s}^\top x_s + \sum_{s=t+1}^{T} a_{i,s}^\top X_s$ تعریف می‌شود. در برخی موارد، با توجه به توزیع متغیر تصادفی X، می‌توان این مقدار مورد انتظار را به‌صورت تحلیلی محاسبه کرد.

مثال ۹.۵. در حالت نرمال چندمتغیره، $X \sim \mathcal{N}(0, I_{dT})$، بخش تصادفی متغیر Y یعنی $\sum_{s=t+1}^{T} a_{i,s}^\top X_s$، یک ترکیب خطی از متغیرهای نرمال مستقل با میانگین صفر است ولذا خود نیز یک متغیر نرمال با میانگین صفر است. بنابراین متغیر Y دارای توزیع نرمال با پارامترهای ذیل خواهد بود:

$$Y \sim \mathcal{N}\Big(b_i + \sum_{s=1}^{t} a_{i,s}^\top x_s,\ \sum_{s=t+1}^{T} \|a_{i,s}\|^2\Big). \tag{۲۰}$$

در این صورت، با ترکیب معادله (۲۰) و فرمول قیمت‌گذاری اختیار خرید بشیلیه[۳۸] که برای متغیر تصادفی نرمال $Z \sim \mathcal{N}(\mu, \sigma^2)$ به‌صورت $\mathbb{E}^{\mathbb{Q}}[Z^+] = \mu\,\Phi\left(\frac{\mu}{\sigma}\right) + \sigma\,\Phi'\left(\frac{\mu}{\sigma}\right)$ است، می‌توان یک فرم تحلیلی برای $\mathbb{E}^{\mathbb{Q}}[Y^+]$ به‌صورت ذیل به‌دست آورد:

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}[Y^+] = &\left(b_i + \sum_{s=1}^{t} a_{i,s}^\top x_s\right) \Phi\left(\frac{b_i + \sum_{s=1}^{t} a_{i,s}^\top x_s}{\sqrt{\sum_{s=t+1}^{T} \|a_{i,s}\|^2}}\right) \\ &+ \sqrt{\sum_{s=t+1}^{T} \|a_{i,s}\|^2} \cdot \Phi'\left(\frac{b_i + \sum_{s=1}^{t} a_{i,s}^\top x_s}{\sqrt{\sum_{s=t+1}^{T} \|a_{i,s}\|^2}}\right),\end{aligned}$$

که در آن Φ تابع توزیع نرمال استاندارد و Φ' چگالی آن است [۵].

در مواردی که توزیع متغیر تصادفی پیچیده است اما تبدیل فوریه تعمیم‌یافته[۳۹] توزیع‌های حاشیه‌ای $\widehat{\mathbb{Q}}_t(u) = \mathbb{E}^{\mathbb{Q}}[e^{u^\top X_t}]$ به‌صورت تحلیلی برای دامنه‌ی مناسبی از بردارهای مختلط u در دسترس است، می‌توان برخی از توابع متغیر تصادفی را با استفاده از انتگرال‌گیری در حوزه‌ی مختلط تحلیل کرد. به‌ویژه تابع $Y^+ = \max(Y, 0)$ را می‌توان به‌صورت نمایش انتگرالی بر حسب توابع نمایی مختلط بازنویسی کرد. برای هر $w > 0$، تابع Y^+ دارای نمایش انتگرالی ذیل است:

$$Y^+ = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{(w+i\lambda)Y}}{(w+i\lambda)^2}\, d\lambda \tag{۲۱}$$

بنابراین، ارزیابی معادله (۱۹) به محاسبه‌ی انتگرال خطی ذیل تقلیل می‌یابد:

$$\begin{aligned} G_{(a_i,b_i),t}(x_1,\ldots,x_t) &= \mathbb{E}^{\mathbb{Q}}\left[Y^+\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[\frac{1}{2\pi}\int_{\mathbb{R}} e^{(w+i\lambda)Y} \cdot \frac{1}{(w+i\lambda)^2}\, d\lambda\right] \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \mathbb{E}^{\mathbb{Q}}\left[e^{(w+i\lambda)Y}\right] \cdot \frac{1}{(w+i\lambda)^2}\, d\lambda \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{F}_Y(w+i\lambda) \cdot \frac{1}{(w+i\lambda)^2}\, d\lambda \end{aligned} \tag{۲۲}$$

[38] Bachelier [39] extended Fourier transform

که در آن $\widehat{F}_Y(w+i\lambda)$ تبدیل فوریه تعمیم‌یافته متغیر Y است و فرم تحلیلی آن از ساختار خطی Y استخراج می‌شود:

$$\widehat{F}_Y(w+i\lambda) = \mathbb{E}^{\mathbb{Q}}\left[e^{(w+i\lambda)Y}\right] = e^{(w+i\lambda)\left(b_i+\sum_{s=1}^{t} a_{i,s}^{\top} x_s\right)} \prod_{s=t+1}^{T} \widehat{\mathbb{Q}}_s\left((w+i\lambda)a_{i,s}\right).$$

توجه داشته باشید که انتگرال‌هایی از نوع فوریه، در دنیای مالی به‌ویژه در مدل‌های مالی مدرن، از جمله مدل‌های با پرش[۴۰] یا مدل‌های افاین[۴۱]، برای قیمت‌گذاری اختیارمعامله بسیار کاربرد دارد.

۶.۵. برآورد با نمونه‌ی متناهی[۴۲]

علیرغم نتایج مربوط به عدم یکتایی و حتی عدم وجود جواب، در قضایای ۴.۵ و ۸.۵ برای نگاشت ویژگی چندجمله‌ای با کاهش بعد و شبکه‌های عصبی، در کاربردهای عملی می‌توان جواب‌های موضعی مسئله‌ی (۷) با استفاده از الگوریتم‌های شبه‌نیوتنی[۴۳] تقریب زد. در عمل به اندازه‌ی مدل $\mathbb{Q}$ که توزیع نظری زیربنای داده‌ها را توصیف می‌کند، دسترسی مستقیم نداریم. لذا آن را با توزیع تجربی $\widehat{\mathbb{Q}}$ تقریب می‌زنیم. این توزیع، مبتنی بر یک نمونه‌ی آموزشی متشکل از مشاهدات مستقل و هم‌توزیع $x^{(1)},\dots,x^{(n)} \sim \mathbb{Q}$ به‌صورت $\widehat{\mathbb{Q}} = \frac{1}{n}\sum_{i=1}^{n} \delta_{x^{(i)}}$ تعریف می‌شود که در آن $\delta_{x^{(i)}}$ تابع دلتای دیراک[۴۴] در نقطه‌ی $x^{(i)}$ است.

در روش تقریب با استفاده از پایه‌ی کامل چندجمله‌ای، مسئله‌ی (۷) به رابطه‌ی (۳) تقلیل می‌یابد و پارامتر بهینه‌ی تجربی به‌صورت بسته قابل محاسبه است:

$$\widehat{\beta} = \left(\frac{1}{n}\Phi^{\top}\Phi\right)^{-1}\left(\frac{1}{n}\Phi^{\top}y\right), \qquad (۲۳)$$

که در آن $\Phi \in \mathbb{R}^{n\times m}$ ماتریس ویژگی با مؤلفه‌های $\Phi_{ij} = \phi_j(x^{(i)})$ است.

فرض می‌کنیم نمونه‌های آموزشی $\{x^{(i)}\}_{i=1}^{n}$ مستقل و هم‌توزیع بوده و ماتریس گرام[۴۵] مثبت معین و در نتیجه وارون‌پذیر است و همچنین تعداد پارامترها (m) ثابت است. تحت این فرضیات و در حالت کلاسیک $m < n$، برآوردگر کمترین مربعات طبق قانون اعداد بزرگ همگرای در احتمال است یعنی $\widehat{\beta} \xrightarrow{P} \beta$. با فرض وجود گشتاورهای مرتبه‌ی دوم لازم، قضیه‌ی حد مرکزی[۴۶] بیان می‌دارد که $\sqrt{n}(\widehat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0,\Sigma)$ که در آن Σ ماتریس کوواریانس حدی است. لازم به ذکر است که این نتایج کلاسیک در چارچوب بعد پارامتر ثابت و با فرض $m < n$ بیان می‌شوند و نمی‌توان آن‌ها را به‌طور مستقیم به حالت‌هایی که تعداد توابع پایه از تعداد نمونه‌ها بیشتر است (یعنی $m > n$) تعمیم داد. در این حالت، به‌ویژه در صورت عدم وارون‌پذیری ماتریس $\Phi^{\top}\Phi$، از منظم‌سازی رگرسیون ریج[۴۷] در پیاده‌سازی عددی استفاده می‌کنیم.

در حالی که در مبنای کامل چندجمله‌ای پارامترهای مدل تنها شامل ضرایب خطی هستند و برآورد آن‌ها در فرم بسته قابل محاسبه است، در دو ساختار دیگر مسئله به یک بهینه‌سازی غیرخطی تبدیل می‌شود.

برای نگاشت ویژگی چندجمله‌ای همراه با کاهش بعد خطی، مسئله‌ی (۷) روی فضای پارامترهای $V_p(\mathbb{R}^{dT})\times\mathbb{R}^m$ تعریف می‌شود. مؤلفه‌ی اول شامل ماتریس‌هایی است که شرط تعامد ستون‌ها را برآورده می‌کنند. بنابراین به‌صورت یک خمینه ریمانی[۴۸] در نظر گرفته می‌شود. مؤلفه‌ی دوم فضای پارامترهای بردار وزن β را تشکیل می‌دهند. با توجه به اینکه فضای پارامتر غیرخطی و دارای ساختار خمینه است. برای بهینه‌سازی مؤثر در این فضا، از الگوریتم شبه‌نیوتنی RBFGS[۴۹] بهره می‌گیریم که نسخه‌ای تعمیم‌یافته از الگوریتم BFGS[۵۰] برای فضاهای ریمانی است. مقاله‌ی [۱۶] چارچوب نظری جامعی برای الگوریتم‌های شبه‌نیوتنی ریمانی ارائه داده است.

برای مدل شبکه‌ی عصبی کم‌عمق با تابع فعال‌سازی رلو، مسئله‌ی (۷) روی فضای پارامتر اقلیدسی $\mathbb{R}^{dT\times m}\times\mathbb{R}^m\times\mathbb{R}^m$ تعریف می‌شود. این فضا شامل وزن‌های بین ورودی و لایه‌ی پنهان، اریبی‌های لایه‌ی پنهان و وزن‌های خروجی است. این مدل با وجود ساختار ساده‌اش، به‌دلیل ماهیت غیرمحدب مسئله، بهینه‌سازی چالش‌برانگیزی دارد. از این‌رو، برای یافتن مینیمم موضعی، از الگوریتم شبه‌نیوتنی

[40] Lévy-type [41] Affine Models [42] Finite-Sample Estimation [43] Quasi-Newton [44] Dirac Delta Function [45] Gram Matrix [46] Central Limit Theorem [47] Ridge Regression [48] Riemannian Manifold [49] Riemannian Broyden–Fletcher–Goldfarb–Shanno [50] Broyden–Fletcher–Goldfarb–Shanno

https://dx.doi.org/ . /msci.xxxx

BFGS استفاده می‌کنیم که در کتابخانه‌ی Scikit-learn [۲۳] برای زبان برنامه‌نویسی پایتون پیاده‌سازی شده است. با توجه به نبود یکتایی و حتی وجود جواب برای نگاشت ویژگی چندجمله‌ای با کاهش بعد و شبکه‌ی عصبی رلو کم‌عمق، همچنان پرسشی باز در حوزه‌ی پژوهش باقی می‌ماند که آیا ویژگی همگرایی سازگار در این روش‌ها برقرار است و آیا می‌توان تضمین‌های نظری برای آن‌ها ارائه کرد یا خیر.

۶. نتایج عددی

در این بخش، معیارهای مورد استفاده برای ارزیابی عملکرد روش‌های مختلف معرفی می‌شوند. با توجه به اینکه در کاربردهای مدیریت ریسک، دقت مدل در نواحی بحرانی توزیع از اهمیت ویژه‌ای برخوردار است، کیفیت برازش در دم توزیع و در بدنه‌ی توزیع مورد بررسی قرار می‌گیرد.

برای سنجش عملکرد مدل در دم توزیع، از دو معیار متداول ارزش در معرض ریسک[۵۱] و زیان مورد انتظار شرطی[۵۲] استفاده می‌شود. برای یک متغیر تصادفی زیان L و سطح اطمینان $\alpha \in (۰, ۱)$، ارزش در معرض ریسک در سطح α به‌صورت ذیل تعریف می‌شود:

$$\mathrm{VaR}_\alpha(L) := \inf\{x \in \mathbb{R} : \mathbb{P}(L \le x) \ge \alpha\}.$$

این کمیت، صدک α-ام توزیع زیان را نشان می‌دهد.

همچنین زیان مورد انتظار شرطی در سطح α به‌صورت ذیل تعریف می‌شود:

$$\mathrm{ES}_\alpha(L) := \frac{۱}{۱-\alpha}\int_\alpha^۱ \mathrm{VaR}_u(L)\,du.$$

در واقع ES_α میانگین زیان‌های فراتر از آستانه‌ی VaR_α را اندازه‌گیری می‌کند.

در این چارچوب، مدل‌ها به‌عنوان برآوردگرهای آماری در نظر گرفته می‌شوند. ازاین‌رو برای محاسبه‌ی VaR_α و ES_α، با انجام R اجرای مستقل از کل زنجیره‌ی شبیه‌سازی، یک توزیع تجربی از نتایج به‌دست می‌آید. همچنین برای میانگین‌گیری این توزیع تجربی از معیار میانگین خطای مطلق درصدی[۵۳] استفاده می‌شود.

۱.۶. میانگین خطای مطلق درصدی (MApE).

فرض کنید مجموعه‌ای از n مشاهده‌ی مستقل از بردار تصادفی $X_{۱:t}$ در اختیار داریم. برای تابع f، با در نظر گرفتن R تکرار از برآوردگر آن، یعنی $\hat{f}$، مجموعه‌ای از توابع $\{\hat{f}_j\}_{j=۱}^R$ به دست می‌آوریم. برای هر تابع در این مجموعه می‌توان یک توزیع تجربی از $\hat{V}_t = \mathbb{E}_t^{\mathbb{Q}}[\hat{f}(X)]$، با استفاده از $X_{۱:t}$ تولید کرد. بنابراین مجموعه‌ای از مقادیر تجربی $\{\hat{V}_t^{(j)}\}_{j=۱}^R$ به دست خواهد آمد. در نهایت میانگین خطای مطلق درصدی برای زیان مورد انتظار شرطی و ارزش فعلی به‌صورت ذیل تعریف می‌شود:

$$\mathrm{MApE\ ES}_\alpha = \frac{۱}{R}\sum_{j=۱}^{R}\frac{\left|ES_\alpha[-\Delta\hat{V}_t^{(j)}] - ES_\alpha[-\Delta V_t]\right|}{ES_\alpha[-\Delta V_t]}. \tag{۲۴}$$

$$\mathrm{MApE\ PV} = \frac{۱}{R}\sum_{j=۱}^{R}\frac{\left|\hat{V}_۰^{(j)} - V_۰\right|}{V_۰}. \tag{۲۵}$$

[51] Value at Risk (VaR) [52] Expected Shortfall (ES) [53] Mean Absolute Percentage Error

۲.۶. میانگین خطای نسبی L^1. برای ارزیابی عملکرد مدل در بدنه‌ی توزیع از میانگین خطای نسبی L^1 استفاده می‌کنیم که به‌صورت ذیل تعریف می‌شود:

$$\frac{1}{R}\sum_{j=1}^{R}\frac{\mathbb{E}\left[\,|\hat{V}_t^{(j)}-V_t|\,\right]}{\mathbb{E}[\,|V_t|\,]}. \tag{۲۶}$$

۷. سبد مالی شامل اختیار خرید اروپایی[۵۴]

در این بخش، به‌عنوان نخستین مطالعه‌ی عددی، سبدی شامل یک اختیار خرید اروپایی بررسی می‌شود که دارایی پایه‌ی آن یک شاخص سهام[۵۵] است. در این تحلیل، فرض می‌شود دارنده‌ی اختیار در موقعیت فروش[۵۶] قرار دارد. بنابراین، توجه ما معطوف به ناحیه‌ی زیان‌آور توزیع فرآیند ارزش سبدمالی است، یعنی زمانی که مقدار شاخص سهام افزایش می‌یابد. اگرچه سررسید قراردادها برابر $T=۵$ و $T=۴۰$ سال است، معیار ریسک برای افق یک‌ساله تعریف و گزارش می‌شود. یعنی هدف، سنجش ریسک ارزش سبد در پایان سال اول است. برای اندازه‌گیری ریسک دم، معیار $\mathrm{ES}_{۹۹٪}(-\Delta V_۱)$ از توزیع تجربی سناریوهای شبیه‌سازی‌شده محاسبه می‌شود که در آن $-\Delta V_۱ := V_۰ - V_۱$ زیان یک‌ساله است. به‌منظور بازتولیدپذیری نتایج، داده‌ها و سناریوهای شبیه‌سازی‌شده‌ی محرک تصادفی X در [۸] و کدهای پایتون در Profile GitHub در دسترس هستند و می‌توان از آن‌ها برای اجرای مجدد محاسبات استفاده کرد.

برای شبیه‌سازی مسیرها یک مولد سناریو اقتصادی در نظر می‌گیریم. این مولد وظیفه دارد مسیرهای متغیرهای مالی را برای مدل‌سازی پویایی وضعیت اقتصادی در طول افق زمانی مورد نظر شبیه‌سازی کند. مولد سناریو اقتصادی، بردار تصادفی X_t را به یک بردار از عوامل اقتصادی نگاشت می‌کند که شامل مؤلفه‌هایی نظیر شاخص سهام S_t، نرخ بهره r_t و حساب نقدی C_t است.

برای محاسبه‌ی ارزش سبدمالی در سررسید T، پرداختی اختیار خرید اروپایی به‌صورت $\max(S_T-K,\,۰)$ تعریف می‌شود. دارنده سبدمالی در این مثال فروشنده‌ی اختیار است، بنابراین این پرداخت به‌عنوان خروجی نقدی با علامت منفی نمایش داده شده است. از طرفی، برای تبدیل این پرداخت اسمی به مقدار تنزیل‌شده، از حساب نقدی C_T به‌عنوان دارایی مبنا استفاده می‌شود:

$$f(X) = -\frac{\max(S_T-K,\,۰)}{C_T} \tag{۲۷}$$

در این مطالعه، قیمت اعمال برابر با مقدار اولیه‌ی شاخص سهام است $(K=S_۰=۱۰۰)$ و مدل شامل سه محرک تصادفی است $(d=۳)$.

شایان ذکر است که علامت منفی برای تاکید بر این نکته است که سودآوری نگهدارنده اختیار خرید از منظر فروشنده زیان خواهد بود. اما در عمل در همه محاسبات مربوط به V_t و $V_۰$، مقدار $f(X)$ به صورت مثبت لحاظ می‌شود.

شبیه‌سازی مونت‌کارلو تودرتو برای سبد مالی شامل اختیار خرید اروپایی. همانطور که در بخش ۲ شرح دادیم، مقدار ارزش سبد مالی در زمان t برابر با مقدار مورد انتظار شرطی آن تحت اندازه‌ی ریسک‌خنثی $\mathbb{Q}$ است $(V_t=\mathbb{E}_t^{\mathbb{Q}}[f(X)])$. شبیه‌سازی تودرتو به‌گونه‌ای طراحی می‌شود که در هر تکرار از شبیه‌سازی بیرونی، یک شبیه‌سازی درونی نیز جهت برآورد مقدار مورد انتظار شرطی انجام گیرد. در این مقاله، برای به‌دست آوردن مقدار مبنا، تکنیک شبیه‌سازی مونت‌کارلوی تودرتو در مقیاس بزرگ اجرا می‌شود تا برآورد دقیقی از مقدار مورد انتظار شرطی به‌دست آید. بدین منظور تعداد شبیه‌سازی‌های بیرونی را $n_۰=۱۰۰۰۰۰۰$ و شبیه‌سازی‌های درونی را $n_۱=۱۰۰۰۰۰$ قرار می‌دهیم. حال معیار $\mathrm{ES}_{۹۹٪}(-\Delta V_۱)$ را بر اساس توزیع تجربی داده‌های شبیه‌سازی‌شده محاسبه می‌کنیم. ابتدا داده‌های $-\Delta V_۱$ به‌صورت نزولی مرتب می‌شوند و سپس چندک-αام برای $\alpha=۰/۹۹$ استخراج می‌گردد. اگر تعداد کل سناریوهای شبیه‌سازی‌شده n باشد، تعداد داده‌های مربوط به دنباله زیان‌آور برابر با $n_{\text{tail}}=\lfloor(۱-\alpha)\times n\rfloor$ است. در نتیجه، تخمین تجربی ES_α

[54] European Call Option [55] Equity Index [56] Short Position

https://dx.doi.org/ . /msci.xxxx

به‌صورت ذیل محاسبه می‌شود:

$$\mathrm{ES}_{\alpha}(-\Delta V_{۱}) = \frac{۱}{n_{\mathrm{tail}}} \sum_{i=۱}^{n_{\mathrm{tail}}} \mathrm{Largest}_{i}(-\Delta V_{۱}),$$

که در آن $\mathrm{Largest}_{i}(-\Delta V_{۱})$ نشان‌دهنده i–اُمین مقدار بزرگ‌تر از نظر شدت زیان در بین داده‌های مرتب‌شده است.

پیش از بررسی کیفیت تخمین‌گر مونت‌کارلو تودرتو در شرایط بودجه‌ی محدود شبیه‌سازی، لازم است تصمیم بگیریم این بودجه بین مرحله‌ی بیرونی و درونی چگونه تقسیم شود. ترکیب‌های مختلفی از $n_{۰}$ و $n_{۱}$ وجود دارند که هر کدام منجر به تخمین متفاوتی می‌شوند. جدول ۳مقدار MAPE را برای ترکیب‌های مختلف از یک بودجه‌ی شبیه‌سازی ثابت $n_{۰} \times n_{۱} = ۵۰۰۰۰$ نشان می‌دهد. همان‌طور که انتظار می‌رود، با افزایش تعداد شبیه‌سازی‌های درونی $n_{۱}$ در شبیه‌سازی مونت‌کارلوی تودرتو، خطای MAPE کاهش یافته و کنترل می‌شود.

به‌منظور دستیابی به نتایجی خوش‌بینانه‌تر برای روش شبیه‌سازی مونت‌کارلوی تودرتو، بهترین ترکیب پارامترها بر اساس معیار کمینه‌سازی MAPE برای مقادیر برآوردشده‌ی زیان مورد انتظار انتخاب می‌شود. با این حال، این سوگیری احتمالی در جهت برآوردهای خوش‌بینانه، تأثیر معناداری بر تحلیل ندارد. زیرا روش پیشنهادی مبتنی بر مارتینگل بازساز در عمل دقت بالاتری نسبت به سایر رویکردها نشان می‌دهد.

جدول ۳. مقایسه‌ی $\mathrm{ES}_{۹۹٪}$ در روش مونت‌کارلوی تودرتو برای افرازهای مختلف مجموع شبیه‌سازی‌های ثابت $n_{۰} \times n_{۱} = ۵۰۰۰۰$ که در آن تعداد شبیه‌سازی‌های داخلی $n_{۱}$ در بازه ۱ تا ۵۰۰ تغییر می‌کند.

سررسید	مقدار مبنا	مونت کارلو تودرتو (خطای MAPE برای شبیه‌سازی‌های داخلی مختلف)							
		۱	۱۰	۲۵	۵۰	۱۰۰	۲۵۰	۴۰۰	۵۰۰
۵	۵۶٫۸۵۸۸	۲۳۳٫۴٪	۲۹٫۵٪	۱۱٫۸٪	۷٫۹٪	۹٫۳٪	۱۴٫۴٪	۱۹٫۰٪	۲۰٫۴٪
۴۰	۶۲٫۶۲۰۵	۲۰۲۷٫۰٪	۴۶۳٫۶٪	۲۳۲٫۱٪	۱۲۹٫۲٪	۶۶٫۶٪	۲۴٫۶٪	۲۱٫۴٪	۱۹٫۱٪

در جدول ۴ مشاهده می‌کنیم که با افزایش بودجه‌ی شبیه‌سازی، خطای برآورد زیان مورد انتظار $\mathrm{ES}_{۹۹٪}(-\Delta V_{۱})$ با سرعت بیشتری کاهش می‌یابد؛ در حالی‌که بهبود دقت ارزش فعلی $V_{۰}$ کندتر است. در شرایطی که هدف صرفاً برآورد مقدار اولیه $V_{۰}$ باشد، اختصاص کل بودجه‌ی محاسباتی به مرحله‌ی شبیه‌سازی بیرونی منطقی‌تر خواهد بود. برای نمونه، در حالتی که اندازه‌ی نمونه برابر با ۵۰۰۰۰ باشد، MAPE برای $V_{۰}$ در ساختار تک‌مرحله‌ای تنها حدود ۰٫۶٪ تا ۱٪ گزارش شده است؛ در حالی‌که با توزیع همین بودجه بین دو مرحله‌ی شبیه‌سازی بیرونی و درونی (در قالب شبیه‌سازی تودرتو)، این مقدار به حدود ۱٫۷٪ تا ۲٫۴٪ افزایش می‌یابد.

جدول ۴. مقایسه‌ی ارزش فعلی و زیان مورد انتظار در روش مونت‌کارلوی تودرتو با استفاده از خطای MAPE؛ با تقسیم بهینه‌ی بودجه شبیه‌سازی درونی‌بیرونی $(n_{۰} \times n_{۱})$ برای اندازه نمونه‌های مختلف. اعداد بر حسب درصد هستند.

تعداد نمونه‌ها $(n_{۰} \times n_{۱})$	ارزش فعلی $(V_{۰})$		زیان مورد انتظار (ES)	
	$T = ۵$	$T = ۴۰$	$T = ۵$	$T = ۴۰$
۱۰۰۰	۶٫۵	۷٫۵	۲۶٫۷	۴۰۳٫۶
۵۰۰۰	۴٫۰	۳٫۹	۱۵٫۹	۱۱۰٫۹
۱۰۰۰۰	۳٫۸	۳٫۱	۱۴٫۵	۵۶٫۶
۵۰۰۰۰	۱٫۷	۲٫۴	۷٫۹	۱۹٫۱

رویکرد یادگیری ماشین برای سبد مالی شامل اختیار خرید اروپایی. در این بخش، توابع پایه معرفی‌شده در بخش‌های ۳.۵ تا ۵.۵ را به‌کار می‌گیریم. برای هر یک از خانواده‌های توابع پایه، سنجه‌های ارزیابی معرفی‌شده در بخش ۷ به‌کار می‌روند و نتایج

روش‌های پیشنهادی با برآورد مبنا حاصل از شبیه‌سازی بزرگ مونت‌کارلوی تودرتو مقایسه می‌شوند. در ارزیابی روش‌های مبتنی بر رگرسیون، از میانگین خطای L_1 بر روی توزیع تجربی ΔV_1 نیز استفاده می‌شود.

در جداول ۵ و ۶ مشاهده می‌شود که روش رگرسیون–آتی (مارتینگل بازساز) در هر دو معیار ارزش فعلی و زیان مورد انتظار ۹۹٪ عملکرد بهتری دارد. برای مقایسه‌ی جامع‌تر، میانگین خطای L_1 در جدول ۷ نیز بررسی شده است. این موضوع، هم‌راستا با یافته‌های پیشین در حوزه روش‌های رگرسیون–آتی است. با این حال، جداول ۵ و ۶ نتایجی برای پایه‌ی کامل چندجمله‌ای تحت روش مارتینگل بازساز در افق زمانی $T = 40$ ارائه نمی‌کنند. دلیل این امر، افزایش غیرقابل کنترل تعداد توابع پایه با افزایش بعد مسئله است. این دشواری‌های محاسباتی در مسائل با بعد بالا، انگیزه‌ای برای استفاده از روش کاهش بعد خطی فراهم می‌سازد.

برای مثال اختیار خرید اروپایی، مقدار $p = 3$ انتخاب شده است. تحلیل حساسیت بر روی این پارامتر انجام گرفت و مشخص شد که مقادیر بزرگ‌تر گاهی و نه در همه موارد عملکرد بهتری دارند. در بخش ۹، تحلیلی از حساسیت نسبت به نقطه‌ی شروع بهینه‌سازی نیز ارائه می‌کنیم. با توجه به اینکه الگوریتم BFGS یک روش شبه‌نیوتنی است، در حالت کلی تضمینی برای یافتن کمینه‌ی سراسری وجود ندارد.

در ادامه، نتایج حاصل از مدل شبکه‌ی عصبی کم‌عمق معرفی‌شده را گزارش می‌کنیم. تعداد کل نورون‌های لایه‌ی پنهان برابر $m = 101$ در نظر گرفته شده است که یکی از آن‌ها صرفاً برای اریبی به‌کار می‌رود به‌طوری‌که $A_{101} = 0$. تعداد پارامترهای مدل شامل دو بخش است:

- وزن‌های بین ورودی و لایه‌ی پنهان: برای هر نورون فعال در لایه‌ی پنهان (یعنی $m - 1$ نورون) به تعداد ابعاد ورودی $(dT + 1)$ وزن وجود دارد. بنابراین تعداد وزن‌های بین ورودی و لایه پنهان برابر است با $(m - 1) \times (dT + 1)$.
- وزن‌های بین لایه‌ی پنهان و خروجی: با توجه به اینکه خروجی یک اسکالر است، برای هر یک از m نورون پنهان، یک وزن خروجی در نظر گرفته می‌شود. این بخش در مجموع m پارامتر خواهد داشت.

در نتیجه، تعداد کل پارامترها $(m - 1) \times (dT + 1) + m$ است. تعداد نورون‌ها و ساختار کلی مدل (شامل تعداد و نوع لایه‌ها، نحوه‌ی اتصال آن‌ها و توابع فعال‌سازی) از طریق فرآیند اعتبارسنجی متقابل انجام شده است. بهینه‌سازی شبکه‌ی عصبی با بهره‌گیری از الگوریتم BFGS و از طریق روش پس‌انتشار خطا[57] انجام شده است. نتایج جداول ۵ و ۶ نشان می‌دهد که مدل مارتینگل بازساز با شبکه‌ی عصبی در این مثال عملکرد بسیار خوبی دارد و در بیشتر موارد از سایر روش‌ها بهتر عمل می‌کند. اگرچه وقتی تعداد نمونه‌ها بسیار پایین است، همان‌گونه که در مدل چندجمله‌ای مشاهده شد، روش رگرسیون–آتی از رگرسیون–اکنون عملکرد بهتری دارد. مقایسه‌ی زمان اجرای روش‌های مختلف در بخش ۱۰ آمده است.

جدول ۵. مقایسه‌ی خطای MApE برای ارزش فعلی V_0 در مثال اختیار خرید اروپایی. اعداد بر حسب درصد هستند.

[57] Error Backpropagation

https://dx.doi.org/ . /msci.xxxx

نمونه‌ها	مونت کارلو تودرتو	پایه‌ی چندجمله‌ای کامل		کاهش بعد خطی	شبکه‌ی عصبی	
		رگرسیون–اکنون	رگرسیون–آتی	رگرسیون–آتی	رگرسیون–اکنون	رگرسیون–آتی
$T = ۵$						
۱۰۰۰	۶٫۵	۴٫۲	۱٫۲	۰٫۴	۴٫۳	۰٫۲
۵۰۰۰	۴٫۰	۱٫۶	۰٫۲	۰٫۲	۲٫۱	۰٫۱
۱۰۰۰۰	۳٫۸	۱٫۲	۰٫۱	۰٫۱	۱٫۳	۰٫۱
۵۰۰۰۰	۱٫۷	۰٫۶	۰٫۱	۰٫۱	۰٫۵	$< ۰٫۱$
$T = ۴۰$						
۱۰۰۰	۷٫۵	۶٫۸		۴٫۷	۸٫۲	۵٫۷
۵۰۰۰	۳٫۹	۳٫۴		۲٫۳	۳٫۸	۱٫۷
۱۰۰۰۰	۳٫۱	۲٫۳		۱٫۳	۲٫۲	۰٫۶
۵۰۰۰۰	۲٫۴	۱٫۰		۰٫۵	۰٫۹	۰٫۲

جدول ۶. مقایسه‌ی خطای (MApE) برای زیان مورد انتظار $\mathrm{ES}_{۹۹٪}$ در مثال اختیار خرید اروپایی. اعداد بر حسب درصد هستند.

نمونه‌ها	مونت کارلو تودرتو	پایه‌ی چندجمله‌ای کامل		کاهش بعد خطی	شبکه‌ی عصبی	
		رگرسیون–اکنون	رگرسیون–آتی	رگرسیون–آتی	رگرسیون–اکنون	رگرسیون–آتی
$T = ۵$						
۱۰۰۰	۲۶٫۷	۱۹٫۲	۳٫۱	۱٫۶	۱۶۹٫۱	۱٫۰
۵۰۰۰	۱۵٫۹	۸٫۵	۱٫۲	۱٫۰	۴۱٫۱	۰٫۲
۱۰۰۰۰	۱۴٫۵	۵٫۹	۰٫۸	۰٫۸	۱۸٫۴	۰٫۱
۵۰۰۰۰	۷٫۹	۲٫۸	۰٫۵	۰٫۶	۵٫۳	$< ۰٫۱$
$T = ۴۰$						
۱۰۰۰	۴۰۳٫۶	۱۴۶٫۰		۱۴٫۶	۷۶۸٫۸	۱۹٫۴
۵۰۰۰	۱۱۰٫۹	۴۰٫۲		۵٫۹	۲۸۱٫۲	۶٫۲
۱۰۰۰۰	۵۶٫۶	۲۴٫۲		۴٫۹	۱۷۳٫۰	۳٫۵
۵۰۰۰۰	۱۹٫۱	۱۰٫۳		۲٫۴	۴۸٫۸	۰٫۷

جدول ۷. مقایسه‌ی میانگین خطای نسبی $L^۱$ در مثال اختیار خرید اروپایی. اعداد بر حسب درصد هستند.

نمونه‌ها	پایه‌ی چندجمله‌ای کامل		کاهش بعد خطی	شبکه‌ی عصبی	
	رگرسیون–اکنون	رگرسیون–آتی	رگرسیون–آتی	رگرسیون–اکنون	رگرسیون–آتی
$T=۵$					
۱۰۰۰	۱۵٫۴	۴٫۵	۱٫۱	۸۳٫۸	۰٫۵
۵۰۰۰	۶٫۸	۰٫۹	۰٫۶	۳۴٫۶	۰٫۴
۱۰۰۰۰	۴٫۶	۰٫۷	۰٫۵	۲۳٫۶	۰٫۳
۵۰۰۰۰	۲٫۱	۰٫۵	۰٫۴	۱۱٫۰	۰٫۳
$T=۴۰$					
۱۰۰۰	۲۶٫۵		۵٫۲	۱۱۴٫۲	۷٫۵
۵۰۰۰	۱۱٫۳		۲٫۶	۵۴٫۰	۲٫۲
۱۰۰۰۰	۸٫۱		۱٫۹	۳۷٫۹	۱٫۲
۵۰۰۰۰	۳٫۷		۱٫۳	۱۷٫۲	۰٫۷

۸. بیمه عمر

در این بخش، بعد از بررسی کارایی یادگیری مارتینگل بازساز در قیمت‌گذاری اختیار خرید اروپایی، قرارداد بیمه‌ی عمر متغیر تضمین‌شده[۵۸] را بررسی می‌کنیم. برخلاف اختیار خرید اروپایی، این مسئله شامل جریان‌های نقدی وابسته به مسیر در چندین نقطه‌ی زمانی است و علاوه‌بر متغیرهای تصادفی بازار به یک مدل تصادفی مرگ‌ومیر[۵۹] نیز وابسته است. در این مقاله، داده‌های مربوط به بیمه‌گذاران واقعی در دسترس نبوده و به‌جای آن از یک جمعیت فرضی برای انجام محاسبات استفاده شده است.

قرارداد بیمه‌ی عمر متغیر تضمین‌شده، حساب سرمایه‌گذاری بیمه‌گذار با تضمین بازده در صورت فوت است. در هر دوره زمانی، بیمه‌گذاران حق‌بیمه پرداخت می‌کنند که برای خرید دارایی‌ها استفاده می‌شود. این دارایی‌ها در صندوق سرمایه‌گذاری، سپرده‌گذاری شده و بر اساس یک تخصیص دارایی ثابت (مطابق با تخصیص اولیه) میان دارایی‌ها تقسیم می‌شود. زمانی که بیمه‌گذار فوت کند، ارزش حساب سرمایه‌گذاری به ذی‌نفعان آن پرداخت می‌شود. در تاریخ سررسید، بیشینه‌ی میان ارزش حساب سرمایه‌گذاری و مبلغ تضمین‌شده به تمامی بازماندگان پرداخت می‌شود. یعنی اگر ارزش حساب سرمایه‌گذاری کمتر از مبلغ تضمین‌شده باشد، شرکت بیمه مابه‌التفاوت را پرداخت می‌کند. مبلغ تضمین‌شده برابر با مجموع تمامی حق‌بیمه‌های پرداخت‌شده است.

در ادامه یک سبد سرمایه‌گذاری مشخص در نظر می‌گیریم که شامل چهار دارایی است: اوراق قرضه بدون کوپن[۶۰] ۱۰ساله، اوراق قرضه بدون کوپن ۲۰ساله، شاخص سهام[۶۱] و شاخص املاک و مستغلات[۶۲]. در این مدل، اوراق قرضه به‌صورت سالانه بازتخصیص می‌یابند به‌گونه‌ای که سبد همواره شامل اوراقی با سررسید ثابت باقی بماند. تمامی متغیرهای مالی به‌صورت اسمی بیان شده‌اند مگر خلاف آن ذکر شود. جریان نقدی تنزیل‌شده که بدهی کل شرکت را نشان می‌دهد؛ به‌صورت ذیل تعریف می‌شود:

$$\zeta_t = \begin{cases} D_t \dfrac{\max(A_t, G_t)}{C_t}, & t < T, \\ L_{T-1} \dfrac{\max(A_T, G_T)}{C_T}, & t = T, \end{cases} \tag{۲۸}$$

برای $t = ۱, \dots, T$. در رابطه (۲۸)، D_t تعداد فوت‌شدگان در بازه $(t-۱, t]$، L_t تعداد بیمه‌گذاران زنده در زمان t، A_t ارزش دارایی‌ها در زمان t، G_t ارزش تضمین‌شده در زمان t و C_t ارزش حساب نقدی در زمان t است.

[58] Variable Annuity Guarantee [59] Mortality [60] Zero Coupon Bond [61] Equity Index [62] Real Estate Index

https://dx.doi.org/ . /msci.xxxx

ارزش سبد دارایی‌ها برابر است با

$$A_t = \sum_{i=۱}^{۴} U_t^i V_t^i = \sum_{i=۱}^{۴} U_{t-۱}^i V_t^{i-} + P_t, \quad t = ۰, \dots, T. \tag{۲۹}$$

با

$$V_t^i = \begin{cases} B(t, t+۱۰), & i = ۱, \\ B(t, t+۲۰), & i = ۲, \\ S_t, & i = ۳, \\ RE_t, & i = ۴, \end{cases} \qquad V_t^{i-} = \begin{cases} B(t, t+۹), & i = ۱, \\ B(t, t+۱۹), & i = ۲, \\ V_t^i, & i = ۳, ۴, \end{cases} \tag{۳۰}$$

که در آن V_t^i قیمت دارایی i–ام در زمان t است و V_t^{i-} بازتخصیص اوراق قرضه با سررسید ثابت را برای $i = ۱, ۲$ نمایش می‌دهد. همچنین $U_t^i = \frac{U_{t-۱}^i V_t^{i-} + M_i P_t}{V_t^i}$ تعداد واحدهای دارایی i–ام نگه‌داری‌شده در بازه $(t, t+۱]$ است ($U_{-۱}^i := ۰$). همچنین $B(t,s)$ ارزش اوراق قرضه در زمان t با سررسید s، S_t ارزش شاخص سهام در زمان t، RE_t ارزش شاخص املاک و مستغلات در زمان t، M_i نسبت تخصیص دارایی‌ها به‌صورت $M = \left(\frac{۱}{۳}, \frac{۱}{۳}, \frac{۱}{۵}, \frac{۲}{۱۵}\right)$ و P_t حق‌بیمه پرداختی در زمان t برای دوره $(t, t+۱]$ است. در ادامه قرار می‌دهیم $P_t = ۱۰۰$.

تضمین بیمه‌نامه به‌صورت ذیل تعریف می‌شود:

$$G_t = \begin{cases} ۰, & t = ۰, \\ G_{t-۱} + P_{t-۱}, & t \geq ۱. \end{cases} \tag{۳۱}$$

در مورد متغیرهای جمعیت‌شناختی[۶۳]، تعداد کل فوت‌شدگان و تعداد کل افراد زنده به‌صورت ذیل تعریف می‌شوند:

$$D_t = \sum_x D_t^x, \qquad L_t = \sum_x L_t^x, \tag{۳۲}$$

که در آن $D_t^x = L_{t-۱}^x q_x(t)$ تعداد فوت‌شدگان در سن x در بازه $(t-۱, t]$ وقتی $x \in (۳۰, ۷۰)$ است. همچنین $q_x(t)$ نرخ مرگ‌ومیر برای سن x در بازه $(t-۱, t]$ و $L_t^x = L_{t-۱}^x - D_t^x$ تعداد افراد زنده در سن x در زمان t با $L_۰^x = ۱۰۰۰$ است.

محرک تصادفی Xدارای $d = ۵$ مؤلفه است که دو مؤلفه مربوط به مدل نرخ بهره و دیگر مؤلفه‌ها در ارتباط با سهام، املاک و مستغلات و مرگ‌ومیر تصادفی است. مدل املاک و مستغلات از ساختاری مشابه مدل سهام پیروی می‌کند، با این تفاوت که از محرک تصادفی مستقل و نوسان‌پذیری کمتری استفاده می‌کند. همچنین، مدل مرگ‌ومیر تصادفی بر اساس مدل لی–کارتر[۶۴] تدوین شده است تا روند و نوسانات تصادفی را در طول زمان بازنمایی کند. نرخ مرگ‌ومیر به‌صورت ذیل مدل‌سازی می‌شوند:

$$q_x(t) = ۱ - e^{-m_x(t)}, \qquad m_x(t) = e^{a_x + b_x k(t)}, \tag{۳۳}$$

$$k(t) = k(t-۱) - ۰/۳۶۵ + \varepsilon_t, \quad \varepsilon_t = ۰/۶۲۱ X_t^{(lc)}, \quad k(۰) = -۱۱/۴۱, \tag{۳۴}$$

که در آن $m_x(t)$ شدت مرگ‌ومیر[۶۵] در زمان t برای سن x، همچنین $X_t^{(lc)}$ مؤلفه‌ی مربوط به مرگ‌ومیر از محرک تصادفی X و a_x و b_x پارامترهای مدل لی–کارتر هستند [۱۹].

نتایج حاصل از قیمت‌گذاری بیمه‌ی عمر، نتایج مربوط به اختیار خرید اروپایی را تأیید می‌کنند. روش مارتینگل بازساز عملکرد بسیار خوبی دارد. به‌ویژه مدل شبکه‌ی عصبی که در اکثر موارد بهترین نتایج را ارائه می‌دهد. اگرچه این قرارداد، برخی محدودیت‌های روش‌ها را بهتر نشان می‌دهد.

[63] Demographic Variables [64] Lee–Carter Model [65] Force Of Mortality

در برآورد ارزش فعلی، جدول ۸ نشان می‌دهد که روش مونت‌کارلوی تودرتو همچنان بسیار کارآمد است. اما روش‌های مبتنی بر رگرسیون، دقت بهتری دارند. مدل شبکه‌ی عصبی در حالتی که تعداد نمونه‌ها بسیار پایین (۱۰۰۰) و بعد بالا ($T = ۴۰$) باشد، عملکرد ضعیفی دارد و لذا بدترین نتایج به‌دست می‌آید. علت این امر بیش‌برازش است. مشابه مثال اختیار خرید اروپایی، از طریق اعتبارسنجی متقابل، $m = ۱۰۱$ نورون را انتخاب کرده‌ایم که یکی از آن‌ها صرفاً برای اریبی است به‌طوری‌که $A_{۱۰۱} = ۰$. بنابراین تعداد کل پارامترها برای $T = ۵$ برابر با $۲۶۰۰ + ۱۰۱ = ۲۷۰۱$ و برای $T = ۴۰$ برابر است با $۲۰۱۰۰ + ۱۰۱ = ۲۰۲۰۱$. همچنین مشاهده می‌کنیم که روش چندجمله‌ای کاهش‌بعد خطی (که با $p = ۱۰$ محاسبه شده است) برتری خود را نسبت به مبنای کامل چندجمله‌ای نشان می‌دهد. مبنای چندجمله‌ای کامل دارای MApE برابر ۶۱٪است. زیرا این مبنا شامل ۳۲۷۶ پارامتر می‌شود (جدول ۱)، در حالی‌که تنها ۱۰۰۰ نمونه در دسترس بوده است. در مقابل، روش کاهش بعد خطی چندجمله‌ای دارای MApE کمتر از ۰/۱٪ است. زیرا تنها شامل ۲۸۶ پارامتر است.

برخلاف مثال اختیار خرید اروپایی که در آن شبکه‌های عصبی بهترین نتایج را داشتند؛ در این مثال روش کاهش بعد خطی چندجمله‌ای در برخی موارد نتایج بهتری ارائه می‌دهد. در مجموع، شبکه‌های عصبی ظرفیت بیشتری برای بهبود از طریق انتخاب ابرپارامترهای جایگزین دارند و می‌توان انتظار داشت که در مثال بیمه‌ی عمر نتایج بهتری به‌دست آید. در برخی موارد، ناکافی بودن تعداد نمونه‌های آموزشی منجر به بیش‌برازش و نتایج ضعیف خارج از نمونه می‌شود. برای مثال، در مبنای کامل چندجمله‌ای با $T = ۵$، بهبود قابل توجهی در نتایج هنگام افزایش داده‌ی آموزشی از ۱۰۰۰ به ۵۰۰۰ نمونه مشاهده می‌شود. این مبنا دارای ۳۲۷۶پارامتر است (جدول ۱)، به این معنا که در حالت ۱۰۰۰ نمونه، تعداد پارامترها بیشتر از تعداد نمونه‌هاست. همین اثر در مدل شبکه‌ی عصبی مارتینگل بازساز با $T = ۴۰$ نیز دیده می‌شود زمانی‌که اندازه نمونه از ۱۰۰۰۰ به ۵۰۰۰۰ افزایش می‌یابد. این امر با توجه به تعداد بالای پارامترها (۲۰۲۰۱) قابل توضیح است.

در برخی حالت‌ها، به‌طور مثال در برآوردگر رگرسیون–آتی شبکه‌ی عصبی با $T = ۵$، مقدار MApE برای معیار زیان مورد انتظار، هنگام افزایش اندازه نمونه از ۵۰۰۰ به ۱۰۰۰۰ و ۵۰۰۰۰ نه تنها کاهش پیدا نمی‌کند، بلکه افزایش نیز می‌یابد (جدول ۹). این نتیجه در نگاه اول متناقض به نظر می‌رسد، زیرا انتظار می‌رود که با افزایش تعداد نمونه‌ها، دقت روش بهبود یابد. اما باید توجه داشت که معیاری که در فرآیند آموزش مدل کمینه می‌شود، با معیاری که در نهایت گزارش و مقایسه می‌شود یکسان نیست. در آموزش شبکه‌ی عصبی، تابع هدف به‌گونه‌ای طراحی شده است که خطا را بر روی کل توزیع جریان‌های نقدی کاهش دهد. یعنی مدل تلاش می‌کند تا در تمامی مقاطع توزیع، تقریب مناسبی ارائه دهد. در مقابل، خطای مورد استفاده در ارزیابی زیان مورد انتظار تنها بر بخش دم توزیع تمرکز دارد. به بیان دیگر، بهبود عملکرد مدل در کل توزیع لزوماً به معنای بهبود دقت در نواحی دم توزیع نیست. در واقع ممکن است مدل با افزایش نمونه‌ها در تقریب کلی بهتر عمل کند اما دقت آن در بخش‌های انتهایی توزیع کاهش یابد. زیرا این نواحی وزن کمتری در فرآیند بهینه‌سازی داشته‌اند. به همین دلیل است که در جدول ۹ با معیار MApE افزایش خطا مشاهده می‌شود، در حالی‌که در خطای میانگین نسبی $L^۱$ چنین رفتاری دیده نمی‌شود (جدول ۱۰). این پدیده نشان‌دهنده‌ی اهمیت انتخاب دقیق معیار خطای بهینه‌سازی در مسائل مدیریت ریسک است. زیرا با توجه به معیار انتخاب‌شده، عملکرد مدل در نواحی حساس توزیع می‌تواند بسیار متفاوت باشد.

جدول ۸. مقایسه‌ی خطای MApE برای ارزش فعلی در مثال بدهی بیمه. اعداد بر حسب درصد هستند.

https://dx.doi.org/ . /msci.xxxx

نمونه‌ها	مونت کارلو تودرتو	پایه‌ی چندجمله‌ای کامل		کاهش بعد خطی	شبکه‌ی عصبی	
		رگرسیون–اکنون	رگرسیون–آتی	رگرسیون–آتی	رگرسیون–اکنون	رگرسیون–آتی
$T = ۵$						
۱۰۰۰	۰٫۳	۰٫۲	۶۱٫۰	$< ۰٫۱$	۰٫۳	۰٫۱
۵۰۰۰	۰٫۳	۰٫۱	$< ۰٫۱$	$< ۰٫۱$	۰٫۱	$< ۰٫۱$
۱۰۰۰۰	۰٫۲	۰٫۱	$< ۰٫۱$	$< ۰٫۱$	۰٫۱	$< ۰٫۱$
۵۰۰۰۰	۰٫۱	$< ۰٫۱$	$< ۰٫۱$	$< ۰٫۱$	$< ۰٫۱$	$< ۰٫۱$
$T = ۴۰$						
۱۰۰۰	۰٫۵	۰٫۵		۰٫۲	۰٫۶	۶٫۱
۵۰۰۰	۰٫۳	۰٫۲		۰٫۱	۰٫۲	۰٫۳
۱۰۰۰۰	۰٫۳	۰٫۲		۰٫۱	۰٫۲	۰٫۴
۵۰۰۰۰	۰٫۲	۰٫۱		$< ۰٫۱$	۰٫۱	۰٫۱

جدول ۹. مقایسه‌ی خطای MApE برای زیان مورد انتظار در مثال بدهی بیمه. اعداد بر حسب درصد هستند.

نمونه‌ها	مونت کارلو تودرتو	پایه‌ی چندجمله‌ای کامل		کاهش بعد خطی	شبکه‌ی عصبی	
		رگرسیون–اکنون	رگرسیون–آتی	رگرسیون–آتی	رگرسیون–اکنون	رگرسیون–آتی
$T = ۵$						
۱۰۰۰	۳۰٫۷	۸۰٫۶	۶۱۳٫۹	۷٫۹	۱۹۸٫۱	۲٫۹
۵۰۰۰	۱۱٫۱	۱۸٫۶	۰٫۵	۵٫۳	۴۸٫۷	۰٫۵
۱۰۰۰۰	۹٫۲	۱۰٫۳	۰٫۳	۵٫۱	۲۵٫۸	۰٫۶
۵۰۰۰۰	۵٫۷	۳٫۳	۰٫۲	۳٫۷	۶٫۶	۰٫۸
$T = ۴۰$						
۱۰۰۰	۱۰۵٫۴	۲۲۶٫۴		۲۲٫۹	۵۲۰٫۷	۱۴٫۴
۵۰۰۰	۳۲٫۷	۶۴٫۹		۵٫۷	۱۴۷٫۳	۱۱٫۱
۱۰۰۰۰	۱۸٫۷	۳۵٫۹		۵٫۹	۸۴٫۴	۱۰٫۵
۵۰۰۰۰	۱۰٫۵	۱۰٫۰		۷٫۲	۱۳٫۲	۰٫۵

جدول ۱۰. مقایسه‌ی میانگین خطای L^1 در مثال بدهی بیمه. اعداد بر حسب درصد هستند.

نمونه‌ها	پایه‌ی چندجمله‌ای کامل		کاهش بعد خطی	شبکه‌ی عصبی	
	رگرسیون–اکنون	رگرسیون–آتی	رگرسیون–آتی	رگرسیون–اکنون	رگرسیون–آتی
$T = ۵$					
۱۰۰۰	۱٫۶	۶۱٫۰	۰٫۲	۴٫۷	۰٫۱
۵۰۰۰	۰٫۷	$< ۰٫۱$	۰٫۲	۱٫۶	۰٫۱
۱۰,۰۰۰	۰٫۵	$< ۰٫۱$	۰٫۲	۱٫۰	$< ۰٫۱$
۵۰,۰۰۰	۰٫۲	$< ۰٫۱$	۰٫۲	۰٫۳	$< ۰٫۱$
$T = ۴۰$					
۱۰۰۰	۳٫۳		۰٫۸	۱۰٫۰	۶٫۱
۵۰۰۰	۱٫۴		۰٫۳	۳٫۴	۰٫۶
۱۰,۰۰۰	۱٫۰		۰٫۴	۲٫۲	۰٫۵
۵۰,۰۰۰	۰٫۴		۰٫۴	۰٫۶	۰٫۱

۹. تحلیل حساسیت نسبت به ابرپارامترها

تحلیل حساسیت روش کاهش‌بعد خطی چندجمله‌ای. روش کاهش‌بعد خطی چندجمله‌ای معرفی‌شده در بخش ۴.۵ دارای دو ابرپارامتر اصلی است: بعد هدف(p) و درجه چندجمله‌ای(δ). علاوه‌براین، الگوریتم RBFGS که برای حل مسئله‌ی بهینه‌سازی (۷) در این روش به‌کار می‌رود، خود دارای پارامترهای متعددی است که مهم‌ترین آن‌ها نقطه‌ی شروع برای ماتریس $A_۰$ است. پارامتر درجه‌ی چندجمله‌ای δ همانند تمام تقریب‌های چندجمله‌ای، اثری قابل‌انتظار بر نتایج دارد. افزایش δ باعث غنی‌تر شدن فضای توابع تقریب می‌شود اما در مقابل تعداد پارامترهای مدل و هزینه‌ی محاسباتی افزایش می‌یابد. در ادامه، تمرکز بر دو پارامتر حساس‌تر روش کاهش‌بعد خطی چندجمله‌ای یعنی p و $A_۰$ قرار گرفته است.

انتخاب p. افزایش بعد هدف (p) نگاشت ویژگی ϕ_θ را غنی‌تر می‌کند و خطای تقریب را می‌کاهد. اما در عوض پارامترهای بیشتری را برای آموزش معرفی می‌کند و به داده‌ی آموزشی بیشتری نیاز دارد. افزایش p با تعداد داده‌ی آموزشی کم می‌تواند به افزایش خطا منجر شود. برای بررسی اثر این پارامتر در مثال بیمه عمر، یکی از روش‌های نقطه شروع که در ادامه با آن آشنا می‌شویم (روش قطری) و یک سررسید ثابت ($T = ۵$) انتخاب شده است. نتایج در جداول ۱۱و ۱۲ ارائه شده‌اند.

جدول ۱۱. مقایسه‌ی خطای MApE برای زیان مورد انتظار در مثال بدهی بیمه برای مقادیر مختلف p در مبنای کاهش‌بعد خطی چندجمله‌ای. اعداد بر حسب درصد هستند.

نمونه‌ها	$p = ۵$	$p = ۱۰$	$p = ۱۵$	$p = ۲۰$
$T = ۵$				
۱۰۰۰	۲۲٫۳	۱۴٫۹	۲۲۸٫۰	۷۶۷٫۹
۵۰۰۰	۲۹٫۳	۱۷٫۵	۸٫۲	۹٫۷
۱۰۰۰۰	۳۳٫۳	۲۳٫۶	۱۳٫۲	۹٫۰
۵۰۰۰۰	۳۷٫۵	۳۱٫۰	۲۱٫۰	۱۰٫۷

جدول ۱۲. مقایسه‌ی میانگین خطای $L_۱$ در مثال بدهی بیمه برای مقادیر مختلف p در مبنای کاهش‌بعد خطی چندجمله‌ای. اعداد بر حسب درصد هستند.

https://dx.doi.org/ . /msci.xxxx

نمونه‌ها	$p=۵$	$p=۱۰$	$p=۱۵$	$p=۲۰$
$T=۵$				
۱۰۰۰	۱٫۰	۱٫۰	۳٫۱	۳۷٫۳
۵۰۰۰	۱٫۰	۰٫۸	۰٫۶	۰٫۵
۱۰۰۰۰	۱٫۰	۰٫۹	۰٫۶	۰٫۶
۵۰۰۰۰	۱٫۰	۰٫۹	۰٫۷	۰٫۵

انتخاب نقطه‌ی شروع $A_۰$. در ادامه سه روش ذیل برای ساخت $A_۰$ بررسی می‌شوند.

(۱) نقطه‌ی شروع تصادفی

در ساده‌ترین حالت، نقطه‌ی شروع $A_۰$ به‌صورت تصادفی ساخته می‌شود. برای این منظور ابتدا dTp نمونه‌ی مستقل از توزیع نرمال استاندارد $\mathcal{N}(۰,۱)$ تولید کرده و آن‌ها را در یک ماتریس $B\in\mathbb{R}^{dT\times p}$ قرار می‌دهیم. سپس با نرمال‌سازی ستون‌ها، ماتریس شروع به‌صورت $A_۰=B\,(B^\top B)^{-۱/۲}$ تعریف می‌شود. این ساختار تضمین می‌کند که $A_۰$ روی خمینه استیفل $V_p(\mathbb{R}^{dT})$ قرار گیرد. هدف از این انتخاب آن است که الگوریتم RBFGS از نقطه‌ای خنثی و بدون سوگیری اولیه آغاز شود. با این حال، در عمل چنین نقطه‌ی شروعی اغلب عملکرد ضعیف‌تری دارد. زیرا در مسائل مالی، اهمیت زمانی جریان‌های نقدی یکسان نیست.

(۲) ماتریس قطری مستطیلی

در این روش، نقطه‌ی شروع $A_۰$ به‌گونه‌ای ساخته می‌شود که تضمین کند هر یک از dT بعد ورودی، دست‌کم در یکی از p بعد خروجی، وزنی غیرصفر داشته باشد. به‌طور دقیق، ماتریس $B:=A_۰$ تنظیم می‌شود که درآن

$$B_{ij}=\begin{cases}۱ & i=j\ \wedge\ i\neq p,\\ \dfrac{۱}{\sqrt{dT-p+۱}} & i\geq p,\\ ۰ & \text{در غیر این صورت.}\end{cases}$$

این نقطه‌ی شروع بازتاب‌دهنده‌ی این ایده است که ابعاد نزدیک‌تر به زمان حال همچنان به‌صورت جداگانه نمایان می‌شوند، در حالی‌که ابعاد دورتر به‌طور جمعی نمایش داده می‌شوند.

(۳) فشرده‌سازی زمانی[۶۶]

ایده همان تجمیع ابعاد دورتر در یک بعد خروجی است. با این تفاوت که تنها در امتداد زمان (بعد T) تجمیع انجام می‌شود، نه در ابعاد d. برای مثال اختیار خرید اروپایی $T=۵$، $d=۳$ و $p=۳$، یک ساختار نرمال‌شده به‌صورت ذیل است:

$$A_۰=\begin{bmatrix}\frac{۱}{\sqrt{T}} & ۰ & ۰\\ ۰ & \frac{۱}{\sqrt{T}} & ۰\\ ۰ & ۰ & \frac{۱}{\sqrt{T}}\\ \vdots & \vdots & \vdots\\ \frac{۱}{\sqrt{T}} & ۰ & ۰\\ ۰ & \frac{۱}{\sqrt{T}} & ۰\\ ۰ & ۰ & \frac{۱}{\sqrt{T}}\end{bmatrix}\qquad(۳۵)$$

[66] Folding

در مثال بیمه عمر، در سناریوی $T = ۵$، $d = ۵$ و $p = ۱۰$، یک نقطه شروع مناسب در نمایش بلوکی به‌صورت ذیل است:

$$A_۰ = \begin{bmatrix} \mathbf{I}_d & [۰]_{d\times(p-d)} \\ [۰]_{d\times d} & \frac{۱}{\sqrt{T-۱}}\mathbf{I}_{(p-d)} \\ \vdots & \vdots \end{bmatrix} \quad \text{(در } T-۱ \text{ سطر زمانی تکرار می‌شود)}, \tag{۳۶}$$

که در آن $\mathbf{I}_d$ ماتریس همانی و $[۰]$ ماتریس صفر با ابعاد مشخص است. این ساختار دو ویژگی مهم دارد: (۱) هر بعد ورودی دست‌کم یک ستون خروجی با وزن غیرصفر دارد که این یعنی پوشش کامل ابعاد. (۲) فشرده‌سازی فقط در بعد زمان انجام می‌شود و نرمال‌سازی باعث می‌شود که نرم این ستون‌های تجمیعی برابر ۱ باقی بماند.

در جداول ۱۳ و ۱۴ عملکرد سه نوع نقطه‌ی شروع را مقایسه کردیم. نتایج نشان می‌دهند که روش فشرده‌سازی زمانی بهترین عملکرد را در میان تمامی نقاط شروع دارد.

جدول ۱۳. مقایسه‌ی خطای MApE برای زیان مورد انتظار در مثال بدهی بیمه برای نقاط شروع مختلف در مبنای چندجمله‌ای با کاهش‌بعد خطی. اعداد بر حسب درصد هستند.

نمونه‌ها	فشرده‌سازی زمانی	قطری	تصادفی
$T = ۵$			
۱۰۰۰	۷٫۹	۱۴٫۹	۱۶٫۶
۵۰۰۰	۵٫۳	۱۷٫۵	۳۱٫۵
۱۰۰۰۰	۵٫۱	۲۳٫۶	۳۱٫۸
۵۰۰۰۰	۳٫۷	۳۱٫۰	۳۲٫۴
$T = ۴۰$			
۱۰۰۰	۲۲٫۹	۳۸٫۰	۶۳٫۹
۵۰۰۰	۵٫۷	۵۴٫۷	۶۹٫۴
۱۰۰۰۰	۵٫۹	۵۳٫۰	۶۸٫۲
۵۰۰۰۰	۷٫۲	۴۹٫۲	۶۷٫۰

جدول ۱۴. مقایسه‌ی میانگین خطای $L_۱$ برای نقاط شروع مختلف در مبنای چندجمله‌ای با کاهش‌بعد خطی در مثال بیمه عمر. اعداد بر حسب درصد هستند.

نمونه‌ها	فشرده‌سازی زمانی	قطری	تصادفی
$T = ۵$			
۱۰۰۰	۰٫۲	۱٫۰	۱٫۲
۵۰۰۰	۰٫۲	۰٫۸	۱٫۰
۱۰۰۰۰	۰٫۲	۰٫۹	۱٫۰
۵۰۰۰۰	۰٫۲	۰٫۹	۱٫۰
$T = ۴۰$			
۱۰۰۰	۰٫۸	۱٫۷	۱٫۷
۵۰۰۰	۰٫۳	۱٫۳	۱٫۵
۱۰۰۰۰	۰٫۴	۱٫۳	۱٫۵
۵۰۰۰۰	۰٫۴	۱٫۲	۱٫۵

https://dx.doi.org/ . /msci.xxxx

تحلیل حساسیت روش شبکه عصبی. روش شبکه عصبی دارای یک ابرپارامتر اصلی یعنی عرض لایه پنهان (h) است. در نتایج اصلی مثال بیمه عمر، تمامی حالت‌ها در لایه‌ی پنهان از پهنای ثابت $h = ۱۰۰$ استفاده شده است. این مقدار بر اساس یک آزمون حساسیت روی مقادیر مختلف (۱۰، ۵۰، ۱۰۰ و ۲۰۰) انتخاب شده است. انتخاب پهنای لایه می‌تواند به روش‌های متفاوتی صورت گیرد و در این بخش برخی از این روش‌ها در مثال بیمه عمر بررسی می‌شوند. خلاصه‌ی این نتایج در جداول ۱۵ و ۱۶ آمده است. نتایج نشان می‌دهند که برخی از این روش‌ها حتی عملکرد بهتری نسبت به انتخاب ما در نتایج اصلی داشته‌اند؛ این امر بیانگر آن است که مبنای شبکه‌ی عصبی، که در مقایسه‌های ما بهترین عملکرد را داشته، هنوز ظرفیت بهبود و ارتقای بیشتر را داراست.

اولین روش استفاده از پهنای حداقل نظری[۶۷] برای عرض لایه پنهان است که این روش در [۱۲] توصیف شده است. این روش عملکرد مناسبی نشان نمی‌دهد که این موضوع، تناقضی با حداقل نظری ندارد. زیرا تعریف حداقل نظری مبتنی بر شبکه‌هایی با عمق دلخواه است. روش دوم آن است که پهنای لایه‌ی شبکه را به‌گونه‌ای انتخاب کنیم که ابعاد فضای پارامترها با ابعاد متناظر در روش کاهش‌بعد خطی چندجمله‌ای یکسان شود. این روش در بعد بالا یعنی $T = ۴۰$ نتایج بسیار خوبی دارد. روش سوم آن است که پهنای لایه‌ی شبکه‌ی عصبی به‌گونه‌ای انتخاب شود که با تعداد توابع پایه m مطابقت داشته باشد. نتایج حاصل از این انتخاب مشابه نتایج دو مورد قبل است و بهبود چشمگیری نسبت به آن‌ها مشاهده نمی‌شود.

جدول ۱۵. مقایسه‌ی خطای MApE برای زیان مورد انتظار شرطی در مثال بدهی بیمه برای عرض‌های مختلف لایه پنهان شبکه‌ی عصبی. اعداد بر حسب درصد هستند.

نمونه‌ها	$h = ۱۰۰$ (ثابت)	حداقل عرض	برابر با ابعاد روش کاهش بعد خطی	ابعاد برابر m
$T = ۵$				
۱۰۰۰	۲٫۹	۲٫۲	۳٫۱	۳٫۹
۵۰۰۰	۰٫۵	۳٫۲	۴٫۱	۰٫۴
۱۰۰۰۰	۰٫۶	۳٫۳	۴٫۲	۰٫۲
۵۰۰۰۰	۰٫۸	۳٫۳	۴٫۲	۰٫۳
$T = ۴۰$				
۱۰۰۰	۱۴٫۴	۱۴٫۵	۱۷٫۵	۱۴٫۱
۵۰۰۰	۱۱٫۱	۱۲٫۵	۲٫۷	۱۲٫۸
۱۰۰۰۰	۱۰٫۵	۱۲٫۶	۲٫۰	۱۳٫۲
۵۰۰۰۰	۰٫۵	۱٫۴	۲٫۲	۲٫۴

جدول ۱۶. مقایسه‌ی میانگین خطای $L_۱$ در مثال بدهی بیمه برای پهنای لایه‌های مختلف در شبکه‌ی عصبی. اعداد بر حسب درصد هستند.

[67] Minimal Width

نمونه‌ها	$h = ۱۰۰$ (ثابت)	حداقل پهنا	برابر با ابعاد روش کاهش بعد خطی	ابعاد برابر m
$T = ۵$				
۱۰۰۰	۰٫۱	۰٫۱	۰٫۱	۰٫۱
۵۰۰۰	۰٫۱	۰٫۱	۰٫۱	۶٫۱
۱۰۰۰۰	۰٫۱	۰٫۱	۰٫۱	۶٫۴
۵۰۰۰۰	۰٫۱	۰٫۱	۰٫۱	۶٫۵
$T = ۴۰$				
۱۰۰۰	۰٫۱	$< ۰٫۱$	$< ۰٫۱$	۰٫۱
۵۰۰۰	۰٫۱	۰٫۱	۰٫۱	۰٫۶
۱۰۰۰۰	۰٫۱	۰٫۱	۰٫۱	۰٫۸
۵۰۰۰۰	۰٫۱	۰٫۱	۰٫۱	۰٫۵

۱۰. مقایسه زمان اجرا

در این بخش بررسی می‌کنیم که اجرای مراحل آموزش و پیش‌بینی برای هر مدل چه مدت زمان نیاز دارد. نتایج برای محاسبه $\hat{V}_۱(x_i)$ برای یک میلیون نمونه اعتبارسنجی است. این نتایج به‌وضوح اثر کاهش بعد را در هزینه محاسباتی روش مارتینگل بازساز آشکار می‌کند. همان‌طور که انتظار می‌رود، روش‌های رگرسیون–اکنون سریع‌تر از روش‌های رگرسیون–آتی هستند. کندترین روش، کاهش بعد خطی چندجمله‌ای است که در مسائل با بعد بالا می‌تواند تا ۲۵دقیقه زمان ببرد. زمان اجرای روش کاهش بعد خطی چندجمله‌ای به شدت به پارامتر p حساس است. به‌عنوان مثال، برای $T = ۴۰$ و اندازه نمونه ۱۰۰۰، حل مسئله با $p = ۵$ تنها ۳۳ ثانیه طول می‌کشد در حالی‌که با $p = ۱۰$ این زمان به ۲۷۹٫۱ ثانیه می‌رسد. این در حالیست که مدل شبکه عصبی را می‌توان نسبتاً سریع حل کرد.

جدول ۱۷. مقایسه‌ی زمان اجرا در مثال اختیار خرید اروپایی (بر حسب ثانیه).

نمونه‌ها	پایه‌ی چندجمله‌ای کامل		کاهش‌بعد خطی	شبکه‌ی عصبی	
	رگرسیون–اکنون	رگرسیون–آتی	رگرسیون–آتی	رگرسیون–اکنون	رگرسیون–آتی
$T = ۵$					
۱۰۰۰	۰٫۹	۱٫۵	۱٫۴	۹٫۵	۱۸٫۴
۵۰۰۰	۰٫۹	۲٫۳	۱٫۶	۱۶٫۱	۲۱٫۱
۱۰۰۰۰	۰٫۹	۳٫۰	۲٫۱	۲۴٫۰	۲۴٫۴
۵۰۰۰۰	۱٫۰	۱۳٫۴	۷٫۱	۹۲٫۶	۴۹٫۴
$T = ۴۰$					
۱۰۰۰	۰٫۹		۷٫۲	۱۰٫۰	۱۸٫۵
۵۰۰۰	۰٫۹		۲۲٫۳	۱۵٫۵	۲۴٫۴
۱۰۰۰۰	۱٫۰		۵۱٫۷	۲۴٫۰	۳۰٫۰
۵۰۰۰۰	۱٫۴		۲۷۱٫۷	۱۹٫۱	۷۶٫۱

جدول ۱۸. مقایسه‌ی زمان اجرا در مثال بدهی بیمه (بر حسب ثانیه).

https://dx.doi.org/ . /msci.xxxx

نمونه‌ها	پایه‌ی چندجمله‌ای کامل		کاهش‌بعد خطی	شبکه‌ی عصبی	
	رگرسیون–اکنون	رگرسیون–آتی	رگرسیون–آتی	رگرسیون–اکنون	رگرسیون–آتی
$T = ۵$					
۱۰۰۰	۰٫۸	۴٫۳	۳۰٫۴	۴٫۰	۵٫۵
۵۰۰۰	۰٫۹	۵۴٫۴	۱۱۳٫۲	۱۱٫۱	۸٫۲
۱۰۰۰۰	۱٫۱	۴۵٫۳	۴۹٫۶	۱۹٫۸	۱۱٫۹
۵۰۰۰۰	۳٫۰	۱۰۸٫۱	۲۳۸٫۶	۹۵٫۱	۴۲٫۷
$T = ۴۰$					
۱۰۰۰	۱٫۳		۲۷۹٫۱	۴٫۶	۶٫۳
۵۰۰۰	۲٫۶		۶۹۸٫۹	۱۲٫۹	۱۵٫۷
۱۰۰۰۰	۴٫۳		۱۸۱۵٫۱	۲۲٫۹	۲۶٫۵
۵۰۰۰۰	۱۷٫۹		۱۴۷۲٫۸	۳۱٫۷	۱۱۵٫۵

۱۱. نتیجه‌گیری

در چارچوب نیاز روزافزون به محاسبات دقیق و سریع برای قیمت‌گذاری سبدها و مدیریت ریسک، در این مقاله روشی داده‌محور برای ساخت مارتینگل‌های بازساز معرفی گردید. مدل پیشنهادی در چارچوب یادگیری نظارت‌شده، در مسائل با ابعاد بالا عملکرد مؤثری دارد. استفاده از مارتینگل بازساز مبتنی بر یادگیری ماشین، ضمن کاهش قابل توجه هزینه محاسباتی، دقتی قابل توجه و در بسیاری از موارد برتر از روش شبیه‌سازی مونت‌کارلوی تودرتوی استاندارد فراهم می‌کند.

نتایج نظری برای دو مثال کاربردی شامل یک اختیار خرید اروپایی و یک محصول بیمه‌ای با جریان‌های نقدی وابسته به مسیر مورد بررسی قرار گرفت. مارتینگل‌های بازساز در این دو مثال، نسبت به سایر روش‌های رایج در صنعت مالی، عملکرد بهتری از خود نشان دادند. این برتری با مقایسه‌های گسترده و تحلیل‌های حساسیت به‌خوبی نشان داده شد.

در نهایت، چالش نظری بنیادین در مدل پیشنهادی، تحلیل سازگاری آماری و رفتار همگرایی برآوردگرهای حاصل از نگاشت‌های کاهش‌بعد و ساختارهای مبتنی بر شبکه‌های عصبی است. این مسئله از آن جهت اهمیت دارد که در چنین ساختارهایی، نه‌تنها یکتایی جواب بهینه لزوماً برقرار نیست، بلکه حتی وجود جواب نیز قابل تضمین نخواهد بود. پرداختن به این پرسش‌ها می‌تواند زمینه‌ساز توسعه‌ی تضمین‌های نظری مستحکم‌تر و گسترش افق‌های پژوهشی جدید در این حوزه باشد.

مراجع

[1] G. Andreatta and S. Corradin, Valuing mortality risk in life insurance contracts, Working paper, 2003.

[2] J. F. Carriere, Valuation of the early-exercise price for options using simulations and nonparametric regression, *Insurance: Mathematics and Economics*, **19** (1996), 19–30.

[3] C. Castellani, P. De Angelis, M. Peracchi and M. Vellekoop, Solvency capital requirement: Neural network versus least-squares Monte Carlo, *Insurance: Mathematics and Economics*, **80** (2018), 147–161.

[4] J. De Spiegeleer, D. B. Madan, S. Reyners and W. Schoutens, Machine learning for quantitative finance: fast derivative pricing, hedging and fitting, *Quantitative Finance*, **18** (2018), 1635–1643.

[5] F. Delbaen and W. Schachermayer, *The Mathematics of Arbitrage*, Springer Finance, Springer-Verlag, Berlin, 2006.

[6] Q. D. Duong, Application of Bayesian penalized spline regression for internal modeling in life insurance, *European Actuarial Journal*, **9** (2019), 67–107.

[7] L. Fernandez-Arjona and D. Filipovi'c, A machine learning approach to portfolio pricing and risk management for high-dimensional problems, *Mathematical Finance*, **32** (2022), 982–1019.

[8] L. Fernandez-Arjona and D. Filipovi'c, Reproducibility data for tables in "A machine learning approach to portfolio pricing and risk management for high-dimensional problems", Dataset, Zenodo, 2022. https://zenodo.org/records/3837381

[9] P. Glasserman and B. Yu, Simulation for American options: Regression now or regression later, in *Monte Carlo and Quasi-Monte Carlo Methods 2002*, Springer, 2004, 213–226.

[10] M. B. Gordy and S. Juneja, Nested simulation in portfolio risk measurement, *Management Science*, **56** (2010), 1833–1848.

[11] H. Ha and D. Bauer, Least-squares Monte Carlo for insurance liability valuation with data-driven basis selection, *European Actuarial Journal*, **10** (2020), 521–548.

[12] B. Hanin and M. Sellke, Approximating continuous functions by ReLU nets of minimal width, arXiv preprint arXiv:1710.11278, 2017.

[13] D. Hainaut, American option pricing with model constrained Gaussian process regressions, *Applied Mathematics and Computation*, **512** (2026), 129747.

[14] L. J. Hong, Z. Hu and G. Liu, Monte Carlo methods for value-at-risk and conditional value-at-risk: A review, *ACM Transactions on Modeling and Computer Simulation*, **27** (2017), 1–37.

[15] K. Hornik, M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks*, **2** (1989), 359–366.

[16] W. Huang, K. A. Gallivan and P.-A. Absil, A Broyden class of quasi-Newton methods for Riemannian optimization, *SIAM Journal on Optimization*, **25** (2015), 1660–1685.

[17] L. Jiang and C. Xu, A new options pricing method: semi-stochastic kernel regression method with constraints, *International Journal of Computer Mathematics*, **100** (2023), 1809–1820.

[18] R. Kan, From moments of sum to moments of product, *Journal of Multivariate Analysis*, **99** (2008), 542–554.

[19] R. D. Lee and L. R. Carter, Modeling and forecasting U.S. mortality, *Journal of the American Statistical Association*, **87** (1992), 659–671.

[20] S. H. Lee and P. W. Glynn, Computing the distribution function of a conditional expectation via Monte Carlo: Discrete and continuous cases, in *Proceedings of the Winter Simulation Conference 2003*, (2003), 356–364.

[21] N. Lehdili, P. Oswald and H. D. Nguyen, Performance-Enhancing Market Risk Calculation Through Gaussian Process Regression and Multi-Fidelity Modeling, *Computation*, **13** (2025), 134.

[22] F. A. Longstaff and E. S. Schwartz, Valuing American options by simulation: A simple least-squares approach, *The Review of Financial Studies*, **14** (2001), 113–147.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, **12** (2011), 2825–2830.

[24] A. Pelsser and N. Schweizer, The difference between LSMC and replicating portfolio in insurance liability modelling, *European Actuarial Journal*, **6** (2016), 441–494.

[25] P. Petersen, M. Raslan and F. Voigtlaender, Topological properties of the set of functions generated by neural networks of fixed size, *Foundations of Computational Mathematics*, **21** (2021), 375–444.

[26] J. Risk and M. Ludkovski, Gaussian process regression for simulation metamodeling, *Operations Research*, **66** (2018), 802–823.

[27] T. J. Sullivan, *Introduction to Uncertainty Quantification*, Texts in Applied Mathematics **63**, Springer, Cham, 2015.

[28] Z. Zhang, S. Zohren and S. Roberts, Deep learning for portfolio optimization, *Journal of Financial Data Science*, **3** (2021), 8–20.

سعید سقایی طوقچی
اصفهان، خیابان هزار جریب، دانشگاه اصفهان، دانشکده ریاضی و آمار، گروه ریاضی کاربردی و علوم کامپیوتر
saeidsaghaei۱۰@gmail.com

سعید سقایی در سال ۱۳۹۷ وارد مقطع کارشناسی رشته ریاضیات و کاربردها در دانشگاه صنعتی اصفهان شد و در سال ۱۴۰۴ مدرک کارشناسی ارشد رشته ریاضیات مالی را از دانشگاه اصفهان اخذ کرده است.

بهاره اختری
اصفهان، خیابان هزار جریب، دانشگاه اصفهان، دانشکده ریاضی و آمار، گروه ریاضی کاربردی و علوم کامپیوتر
b.akhtari@mcs.ui.ac.ir
b.akhtari@ipm.ir

بهاره اختری عضو هیئت علمی گروه ریاضی مالی دانشکده ریاضی و آمار دانشگاه اصفهان است. وی همچنین پژوهشگر مقیم پژوهشگاه دانش های بنیادی است.