بسم الله الرحمن الرحيم

عنوان گزارش:
محاسبه حمل و نقل بهینه و مسائل مرتبط با پوشش ریسک از طریق جریمه‌دهی و شبکه‌های عصبی

محمد سوری – فرزانه حسینی

پروژه ی ریاضی مالی

اساتید: دکتر فتوحی و دکتر آسا

مرداد ۱۴۰۲

فهرست مطالب

۱ مقدمه

گزارش زیر خلاصه ای از مقاله‌ی "محاسبه حمل و نقل بهینه و مسائل مرتبط با پوشش ریسک از طریق جریمه‌دهی و شبکه‌های عصبی" است. این مقاله به صورت خاص روی یک مسئله بهینه سازی به شکل $\phi(f) = \sup_{\nu\in\mathbb{Q}} \int f\,d\nu$ که به در بسیاری از مسائل ظاهر می‌شود متمرکز است. در این گزارش ما در ابتدا مقدمه ای از مقاله و سپس خلاصه ای از کاری که قرار است انجام بدهیم را ارائه داده‌ایم. در ادامه برای نشان دادن همگرایی های لازم با استفاده از قضیه دنیل–استون غیرخطی و رادون–نیکودیم مواردی را اثبات کردیم و در انتها با معرفی شبکه های عصبی و توضیح مثال عددی حل شده و نشان دادن نتایج کد کار را به پایان رساندیم. در طول این گزارش تمامی قضایای مورد استفاده و نکات مربوطه برای حل مثال عددی به طور کامل توضیح داده شده است. به طور خاص ما از مرجع [۱۵] برای اطمینان از درستی حل آنالیتیک مسئله استفاده کردیم و تمامی نکات مربوط به حل آنالیتیک را نیز درون گزارش قرار داده ایم.

در این مقاله، یک روش جریمه‌دهی ارائه می‌شود که امکان محاسبه یک کلاس گسترده از مسائل بهینه‌سازی را با استفاده از شبکه‌های عصبی فراهم می‌کند. این مسائل بهینه‌سازی شامل مسائلی هستند که به شکل زیر قابل بیان هستند:

$$\phi(f) = \sup_{\nu\in\mathbb{Q}} \int f\,d\nu$$

مسئله بهینه‌سازی مورد نظر در این مقاله به صورت معمول برای مسئله حمل و نقل بهینه مطرح می‌شود. به طور کلی، این نوع توابع و مسائل بهینه‌سازی در بسیاری از حوزه‌ها و کاربردهای دیگر نیز دیده می‌شوند.

این توابع به طور خاص در ریاضیات مالی کاربردهای زیادی در زمینه محاسبه پوشش ریسک از بالا و بهینه سازی سبد سهام در حضور عدم قطعیت دارند.

برای حل مسئله اولیه $\phi(f)$، ما از فرمولاسیون دوگان آن (با استفاده از قضیه دنیل استون غیرخطی در بخش 4) استفاده خواهیم کرد و محدود به زیرکلاسی از مسائل بهینه‌سازی می‌شویم که به عنوان قیمت کمینه پوشش ریسک از بالا قابل دستیابی‌اند:

$$\phi(f) = \inf_{\substack{h\in\mathcal{H}:\\ h\geq f}} \int h d\mu_0.$$

که در آن $\mu_0 \in \mathcal{Q}$ و $\mathcal{H}$ مجموعه‌ای از توابع پیوسته و کران دار $h : \mathcal{X} \to \mathbb{R}$ است. رابطه ی میان $\mathcal{H}$ و $\mathcal{Q}$ در بخش های بعدی توضیح داده می شود.

با استناد به مرجع [۱] نیز می توان دید تحت شرایط نرمال‌سازی کافی، مقادیر مسئله اصلی $\phi(f) = \sup_{\nu\in\mathbb{Q}} \int f\,d\nu$ و مسئله دوگان آن $\phi(f) = \inf_{\substack{h\in\mathcal{H}:\\ h\geq f}} \int h d\mu_0$ می توانند با هم برابر باشند.

۱.۱ پیشنیه تاریخی(مسئله مونژ)

یک مثال معمول توصیف کانتوروویچ [۲] $(Kantorovich)$ یک روش ریاضی است که توسط ریاضیدان لئونید کانتوروویچ برای تخمین حل مسئله حمل و نقل بهینه مونژه توسعه داده شده است. مسئله حمل و نقل بهینه مونژه یک مسئله ریاضی است که در آن سعی می‌شود به صورت بهینه‌ترین روش برای انتقال جرم از یک توزیع به توزیع دیگر پیدا شود و هزینه حمل و نقل کمینه گردد.

در تخمین کانتوروویچ، مجموعه‌ای از اندازه های احتمال را در فضای $X = X_1 \times X_2$ در نظر می‌گیریم. فضاهای X_1 و X_2 هر کدام با اندازه‌احتمال μ_1 و μ_2 مشخص شده‌اند. مجموعه Q را به عنوان مجموعه اندازه احتمال های در فضای X با بازتاب‌های μ_1 و μ_2 در نظر می‌گیریم.

ایده این است که یک برنامه حمل و نقل را پیدا کنیم که به هر نقطه (x_1, x_2) در فضای X، اندازه احتمالی از مجموعه Q را اختصاص دهد که میزان جرمی که از x_1 به x_2 انتقال می‌یابد را نشان دهد.

برای حل این مسئله، مجموعه H را از تمام توابع پیوسته و کران دار $h(x_1, x_2)$ که به عنوان مجموعه دو تابع $h_1(x_1)$ و $h_2(x_2)$ قابل بیان هستند، معرفی می‌کنیم. در اینجا، h_1 به X_1 تعلق دارد و h_2 به X_2 تعلق دارد. به عبارت دیگر $h(x_1, x_2) = h_1(x_1) + h_2(x_2)$. این نمایش به ما امکان می‌دهد که مشارکت x_1 و x_2 را در هزینه کل یا تابع هدف اصلی، جدا کنیم.

هدف تخمین کانتوروویچ یافتن یک برنامه حمل و نقل است که هزینه کل را کمینه کند. هزینه کل به عنوان انتگرال $h(x_1, x_2)$ نسبت به اندازه احتمال مشترک μ_0 در فضای X تعریف می‌شود.

که میتوان نوشت : $\int h d\mu_0 = \int_{\mathcal{X}_1} h_1 d\mu_1 + \int_{\mathcal{X}_2} h_2 d\mu_2$

با فرمول‌بندی مسئله به این شکل، تخمین کانتوروویچ به ما امکان می‌دهد تا یک حل را پیدا کنیم که محاسبه آن نسبت به حل مستقیم مسئله حمل و نقل بهینه مونژه آسانتر باشد.

در تمامی مسائل زیر هم روش ارائه شده در مقاله پاسخگو خواهد بود:

مسائل حمل و نقل بهینه چند حاشیه ای و فواصل واسرشتین $(Wasserstein)$ ([۳] و [۴] و [۵]) ، مسائل حمل و نقل بهینه مارتینگل ([۶] و [۷] و [۸] و [۹])، مسائل ارزش در معرض ابهام در حضور عدم قطعیت ([۱۰] و [۱۱] و [۱۲]) ، محاسبه مقادیر بدترین حالت کوپولا و محدودیت‌های فرشه-هوفدینگ بهبودیافته ([۱۳] و [۱۴])و غیره.

۲ خلاصه ای از رویکرد مقاله

هدف از این روش، حل عددی $\phi(f)$ است. پیاده‌سازی این روش بر مبنای نمایش دوگانی $\phi(f)$ انجام می‌شود. ابتدا، مسئله را به یک فضای ابعاد متناهی تبدیل می‌کنیم. به جای کار با مجموعه $\mathcal{H}$، یک زیرمجموعه $\mathcal{H}^m$ را در نظر می‌گیریم. تابع $\phi(f)$ را به صورت $\phi^m(f)$ نمایش می‌دهیم که به صورت $\phi^m(f) = \inf_{h \in \mathcal{H}^m} \left\{ \int h d\mu_0 : h \geq f \right\}$ تعریف می‌شود. در این روش، یک دنباله $(\mathcal{H}^m)_{m \in \mathbb{N}}$ در نظر می‌گیریم به طوری که $\mathcal{H}^1 \subseteq \mathcal{H}^2 \subseteq \cdots \subseteq \mathcal{H}$ و $\mathcal{H}^\infty = \cup_{m \in \mathbb{N}} \mathcal{H}^m$ در فضای $\mathcal{H}$ یک چگالی به خصوص دارد.به طور مشخص‌تر، $\mathcal{H}^m$ می‌تواند

$$
\begin{tikzcd}[column sep=huge, row sep=huge]
\begin{matrix}\phi(f)\\ \textit{(initial problem)}\end{matrix} & \begin{matrix}\phi_{\theta,\gamma}(f)\\ \textit{(penalized version of } \phi(f)\textit{)}\end{matrix} \arrow[l, "\text{(Thm. 2.2)}"', "\gamma\to\infty"] \\
\begin{matrix}\phi^m(f)\\ \textit{(finite-dim. version of } \phi(f)\textit{)}\end{matrix} \arrow[u, "m\to\infty", "\begin{matrix}\text{(Prop. 2.3)}\\ \text{(Lemma 3.3)}\end{matrix}"'] & \begin{matrix}\phi^m_{\theta,\gamma}(f)\\ \textit{(numerically solved problem)}\end{matrix} \arrow[l, "\text{(Prop. 3.7)}"', "\gamma\to\infty"] \arrow[u, "m\to\infty", "\text{(Remark 3.5)}"'] \arrow[ul, "\text{(Prop. 2.4, Remark 3.5)}"', "\min\{\gamma,m\}\to\infty"]
\end{tikzcd}
$$

شکل ۱: مسائل مطرح شده و روابط آن ها

مجموعه‌ای از شبکه‌های عصبی با ساختار ثابت (اما مقادیر پارامترها نامشخص) باشد و m تعداد نورون‌ها در هر لایه را نشان می‌دهد.

برای به‌روزرسانی پارامترها در $\mathcal{H}^m$ (به عنوان مثال با استفاده از روش‌های نزول گرادیان)، محدودیت نامساوی $f \geq h$ جریمه می‌شود. به این منظور، یک اندازه احتمال مرجع θ را در فضای حالت $\mathcal{X}$ معرفی می‌کنیم. این اندازه برای نمونه‌برداری از نقاطی که در آن‌ها محدودیت نامساوی $f \geq h$ قابل بررسی است، استفاده می‌شود. علاوه بر این، یک تابع جریمه مشتق پذیر و غیرنزولی $\beta : \mathbb{R} \to \mathbb{R}_+$ معرفی می‌شود. این مسئله جریمه‌شده به صورت

$$\phi^m_{\theta,\beta}(f) = \inf_{h\in\mathcal{H}^m} \left\{ \int h d\mu_0 + \int \beta(f-h) d\theta \right\}$$

تعریف می‌شود. در این روش، مسئله

$$\phi_{\theta,\beta}(f) = \inf_{h\in\mathcal{H}} \left\{ \int h d\mu_0 + \int \beta(f-h) d\theta \right\}$$

را برای تحلیل نظری معرفی می‌کنیم. ما دوباره دنباله‌هایی از توابع جریمه $(\beta_\gamma)_{\gamma>0}$ را با پارامتر جریمه γ در نظر می‌گیریم و از نمادگذاری $\phi_{\theta,\gamma}(f) := \phi_{\theta,\beta_\gamma}(f)$ و $\phi^m_{\theta,\gamma}(f) := \phi^m_{\theta,\beta_\gamma}(f)$ استفاده می‌کنیم. در اینجا، یک عامل جریمه رو به افزایش را می‌توان به عنوان اعمال دقیق‌تر و دقیق‌تر محدودیت نامساوی $f \geq h$ در نظر گرفت. مسائل $\phi^m_{\theta,\gamma}(f)$ که به صورت عددی حل می‌شوند، همان مسائلی هستند که در نهایت پیاده‌سازی می‌شوند. گام نهایی، پیدا کردن یک راه حل عددی برای $\phi^m_{\theta,\gamma}(f)$ است، که در عمل به معنای یافتن پارامترهای بهینه شبکه $\mathcal{H}^m$ است. ما راه حل بهینه عددی را با $\hat{\phi}^m_{\theta,\gamma}(f)$ نشان می‌دهیم.

۳ تنظیم و تقریب تابع‌های پوشش ریسک

بگذارید $\mathcal{P}(\mathcal{X})$ را مجموعه‌ی همه‌ی اندازه‌های احتمال بورل (یک اندازه احتمال بورل یک اندازه احتمال است که بر روی فضای بورل قابل اعمال است.) روی یک فضای لهستانی (یک فضای توپولوژیک است که دارای خواص توپولوژیک مشخص (مانند جداسازی و شمولیت شمارا) و وجود متریک مناسب است که به صورت خاص برای بررسی خواص ریاضیاتی و احتمالی استفاده می‌شوند. فضای لهستانی، یک کلاس گسترده از فضاهای توپولوژیک را شامل می‌شود، از جمله فضاهای هیلبرت) $\mathcal{X}$ در نظر بگیریم. همچنین $C_b(\mathcal{X})$ را فضای خطی تمام توابع پیوسته و کران‌دار $f : \mathcal{X} \to \mathbb{R}$ در نظر بگیرید. ما تابع پوشش ریسک از بالا را در نظر می‌گیریم:

$$\phi(f) := \inf \left\{ \int h d\mu_0 : h \geq f \text{ ، } h \in \mathcal{H} \text{ برای برخی از } \right\} \tag{۱}$$

که در آن $f \in C_b(\mathcal{X})$، $\mu_0 \in \mathcal{P}(\mathcal{X})$ یک اندازه قیمت‌گذاری است و $\mathcal{H} \subseteq C_b(\mathcal{X})$ است. در طول این بخش، فرض می‌کنیم $\mathcal{H}$ یک فضای تابعی خطی است که شامل توابع ثابت (توابعی که تنها یک مقدار ثابت را دارند) است. به طور ساده، $\phi(f)$ حداقل هزینه‌ای است (حداقل مقدار تابع) که برای محافظت در برابر احتمال زیان تابع f از زیر با استفاده از تابعی مثل $h \in \mathcal{H}$ نیاز است. برای استنتاج یک نمایش دوگانی، فرض می‌کنیم که تابع ϕ از بالا پیوسته است، به عبارت دیگر برای هر دنباله (f_n) در $C_b(\mathcal{X})$ که $f^n \downarrow 0$ (به معنای اینکه دنباله f_n به طور نقطه‌ای به صفر کاهش می‌یابد)، داریم $\phi\left(f_n\right) \downarrow 0$. با استفاده از قضیه‌ی دنیل–استون غیرخطی، نمایش زیر را برای $\phi(f)$ به دست می‌آوریم:

$$\phi(f) = \max_{\mu \in \mathcal{Q}} \int f d\mu \tag{۲}$$

که در آن $\mathcal{Q}$ مجموعه‌ی غیرتهی از اندازه‌های احتمال بورل است، به طوری که برای هر $h \in \mathcal{H}$، $\int h d\mu = \int h d\mu_0$ برقرار است. به عبارت دیگر ، $\mathcal{Q}$ مجموعه‌ای از اندازه‌های احتمال است که در هر نقطه‌ی $\mathcal{H}$، مقدار تابع μ با مقدار تابع μ_0 برابر است. به خصوص، $\mu_0 \in \mathcal{Q}$. با این حال، به منظور خوانایی بیشتر، ما بر روی $C_b(\mathcal{X})$ تمرکز می‌کنیم.
مثال زیر تنظیمات پایه را توضیح می‌دهد:
در این مثال، $\mathcal{X}$ فضای $\mathbb{R}^d$ است و $\mathcal{P}(\mathbb{R}^d)$ مجموعه‌ی تمام اندازه‌های احتمال بورل بر روی فضای $\mathbb{R}^d$ را نشان می‌دهد. سپس، ما با استفاده از مجموعه $\Pi(\mu_1, \ldots, \mu d)$، که مجموعه‌ای از اندازه‌های احتمال با حاشیه‌های مشخص است، مجموعه $\mathcal{Q}$ را تعریف می‌کنیم. در این مثال، $\mathcal{Q}$ همه‌ی اندازه‌های احتمالی است که حاشیه اول آنها μ_1، حاشیه دوم آنها μ_2 و غیره است.
با فرض $\mathcal{Q} \neq \emptyset$، می‌توان اثبات کرد که تابع پوشش ریسک از بالا ϕ در این مثال پیوسته از بالا است:

حمل و نقل بهینه چند حاشیه ای:

$$\mathcal{Q} = \Pi(\mu_1, \dots, \mu_d)$$
$$\mathcal{H} = \{h \in C_b(\mathbb{R}^d) : h(x_1, \dots, x_d) = h_1(x_1) + \cdots + h_d(x_d) \text{ برای همه‌ی } (x_1, \dots, x_d) \in \mathbb{R}^d \text{ و برخی از } h_i \in C_b(\mathbb{R})\}$$

این مسئله به صورت تحقیقاتی در مراجع [۲] و [۵] مورد بررسی قرار گرفته است.

۴ تنظیم تابع پوشش ریسک از بالا با استفاده از جریمه

نشان می‌دهیم که نمایش مزدوج تابع $\phi_{\theta,\gamma}(f)$ به صورت زیر است:

$$\phi_{\theta,\gamma}(f) := \inf_{h \in C_b(\mathcal{X})} \{\phi(h) + \psi_{\theta,\gamma}(f-h)\}$$

$$= \inf_{h \in \mathcal{H}} \left\{ \int h d\mu_0 + \int \beta_\gamma(f-h) d\theta \right\} \tag{۳}$$

که در آن $\phi(h)$ تابع پوشش ریسک ($superhedging functional$) است که قبلاً از آن صحبت شده است و $\psi_{\theta,\gamma}(f)$ تابعی است که با همگشت گیری ($convolution$) تعریف می‌شود و به صورت زیر تعریف می‌شود:

$$\psi_{\theta,\gamma}(f) := \int \beta_\gamma(f) d\theta$$

در اینجا، $\beta_\gamma(x)$ یک تابع جریمه است که پارامتر $\gamma > 0$ را دارد و به صورت زیر تعریف می‌شود:

$$\beta_\gamma(x) := \frac{1}{\gamma}\beta(\gamma x)$$

تابع $\beta(x)$ یک تابع صعودی، مشتق‌پذیر و محدب است که از قبل فرض شده است. همچنین، فرض می‌شود که $\lim_{x\to\infty} \beta(x)/x = \infty$. حالا، به دنبال نمایش مزدوج تابع $\phi_{\theta,\gamma}(f)$ هستیم. برای این کار، تابع مزدوج محدب ($conjugate$) تابع $\beta_\gamma(x)$ را معرفی می کنیم:

$$\beta_\gamma^*(y) := \sup_{x \in \mathbb{R}} \{xy - \beta_\gamma(x)\} \text{ برای همه } y \in \mathbb{R}_+$$

که دارای خاصیت زیر است:

$$\beta_\gamma^*(y) = \frac{\beta^*(y)}{\gamma}$$

اثبات:

$$\beta^*_\gamma(y) := \sup_{x\in\mathbb{R}} \{xy - \beta_\gamma(x)\}$$
$$= \sup_{x\in\mathbb{R}} \left\{\frac{\gamma}{\gamma}(xy) - \frac{1}{\gamma}\beta(\gamma x)\right\}$$
$$= \sup_{x\in\mathbb{R}} \left\{\frac{1}{\gamma}((\gamma xy) - \beta(\gamma x))\right\}$$

با در نظر گرفتن $x' = \gamma x$ داریم:

$$= \sup_{x'\in\mathbb{R}} \left\{\frac{1}{\gamma}(x'y) - \beta(x')\right\}$$
$$\beta^*_\gamma(y) = \frac{\beta^*(y)}{\gamma}$$

مثال های از تابع $\beta(x)$ را مطرح می کنیم که تابع استفاده شده در نمونه عددی ما مثال دوم است:
(الف) تابع جریمه نمایی $\beta(x) = \exp(x-1)$ با مزدوج $\beta^*(y) = y\log(y)$
(ب) تابع جریمه L^p، $\beta(x) = \frac{1}{p}(\max\{0,x\})^p$ با مزدوج $\beta^*(y) = \frac{1}{q}y^q$ که در آن $q = \frac{p}{p-1}$ برای برخی $p > 1$ است.
در قضیه زیر دوگان $\phi_{\theta,\gamma}(f)$ را نشان می دهیم و و همگراییش به $\phi(f)$ را اثبات می کنیم. برای فهمیدن بهتر صورت قضیه و اثبات آن نیاز است که در ابتدا با دو قضیه ی رادون نیکودیم و نسخه غیرخطی قضیه ی دنیل استون آشنا شویم.

۱.۴ قضیه ی رادون نیکودیم

قضیه رادون–نیکودیم $(Radon - Nikodym)$ یک نتیجه بنیادی در نظریه اندازه است که شرایطی را برای وابستگی یک اندازه به اندازه دیگر فراهم می‌کند. این قضیه یک ارتباط بین دو اندازه را برقرار می‌کند و تابعی که دو اندازه را به هم تبدیل می‌کند را مشتق رادون–نیکودیم می‌خوانند که نرخ تغییر بین این دو اندازه را نشان می‌دهد.
فرض کنیم (X, Σ) یک فضای اندازه‌پذیر باشد که X همان فضای لهستانی مورد مطالعه و Σ یک سیگما–جبر روی آن است. حال فرض کنیم که دو سیگما–اندازه متناهی مثل μ و ν داریم که $\nu \ll \mu$ که این رابطه به این معناست که ν نسبت به μ مطلقا پیوسته است. معنای این گزاره هم این است که اگر برای هر زیرمجموعه اندازه‌پذیر مثل $A \subseteq X$ داشته باشیم که $\mu(A) = 0$ آنگاه داریم $\nu(A) = 0$. آنگاه تابع سیگما–اندازه‌پذیر f موجود است به طوری که برای هر $A \subseteq X$ که اندازه‌پذیر باشد داریم:

$$\nu(A) = \int_A f\, d\mu,$$

و به این f مشتق رادون-نیکودیم ν نسبت به μ گفته می‌شود و معمولا با $\frac{d\nu}{d\mu}$ نمایش داده می‌شود. یک راه برای مشتق گیری یک اندازه جدید از یک اندازه از پیش تعیین شده، اختصاص دادن چگالی به هر نقطه از فضا و سپس انتگرال گیری روی زیرمجموعه‌های قابل اندازه‌گیری مورد نظر است که با همان صورت‌بندی بالا بیان می‌شود. که اندازه جدیدی است که برای هر زیرمجموعه قابل اندازه‌گیری A تعریف می‌شود و تابع f چگالی در یک نقطه خاص است، یک خاصیت مفید درباره مشتق رادون-نیکودیم این است که اگر $\lambda \ll \mu$ و g یک تابع μ-انتگرال‌پذیر باشد آنگاه داریم:

$$\int_X g\, d\mu = \int_X g \frac{d\mu}{d\lambda}\, d\lambda.$$

۲.۴ قضیه ی دنیل-استون غیرخطی

فرض کنید $\mathcal{X}$ یک فضای لهستانی باشد. با توجه به یک تابع اندازه‌پذیر $\kappa : \mathcal{X} \to [1, \infty)$، با $C_\kappa(\mathcal{X})$ مجموعه تمام توابع پیوسته $f : \mathcal{X} \to \mathbb{R}$ را که $|f|/\kappa$ کران‌دار است، به عنوان شبکه برداری استون استون (یک شبکه برداری استون استون، یک فضای ریز $Riesz$ است که همچنین دیدکیند $Dedekind$کامل است. این به این معناست که هر زیرمجموعه محدود غیرتهی از این فضا یک سوپریمم دارد. به عبارت دیگر، برای هر زیرمجموعه S از فضا، اگر یک عنصر u وجود داشته باشد که برای هر $x \in S$، $x \leq u$ و برای هر حد بالای دیگر v از S، $u \leq v$، آنگاه u حد بالای S است.) نشان می‌دهیم. به عنوان مثال، اگر κ کران‌دار باشد، داریم $C_\kappa(\mathcal{X}) = C_b(\mathcal{X})$، یا اگر $\kappa(x) = 1 + |x|$ در $\mathcal{X} = \mathbb{R}^d$ باشد، فضای $C_\kappa\left(\mathbb{R}^d\right)$ شامل تمام توابع پیوسته $f : \mathbb{R}^d \to \mathbb{R}$ با رشد خطی است. علاوه بر این، با $ca_\kappa^+(\mathcal{X})$، مجموعه تمام اندازه‌های بورل μ روی $\mathcal{X}$ را نشان می‌دهیم که شرط $\int \kappa d\mu < \infty$ را برآورده می کنند.

گزاره الف) فرض کنید $\phi : C_\kappa(\mathcal{X}) \to \mathbb{R}$ یک تابع صعودی محدب است که پیوسته از بالا است، به این معنی که برای هر دنباله (f^n) که $f^n \downarrow 0$ باشد، داریم $\phi\left(f^n\right) \downarrow 0$. در این صورت، تابع ϕ یک نمایش دوگان دارد به شکل زیر:

$$\phi(f) = \max_{\mu \in ca_\kappa^+(\mathcal{X})} \left\{ \int f d\mu - \phi^*(\mu) \right\} \quad \text{برای همه}\; f \in C_\kappa(\mathcal{X}), \tag{۴}$$

که $\phi^* : ca_\kappa^+(\mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ تابع مزدوج محدب ϕ است که به شکل زیر تعریف می‌شود:

$$\phi^*(\mu) = \sup_{f \in C_\kappa(\mathcal{X})} \left\{ \int f d\mu - \phi(f) \right\}.$$

به عنوان یک کاربرد، تابع پوشش ریسک از بالا را در نظر بگیرید به شکل زیر:

$$\phi(f) := \inf \left\{ \int h d\mu_0 : h \geq f \quad \text{برای برخی}\; h \in \mathcal{H} \right\}$$

در $C_\kappa(\mathcal{X})$، که در آن $\mu_0 \in ca_\kappa^+(\mathcal{X})$ یک اندازه احتمال است و $\mathcal{H} \subseteq C_\kappa(\mathcal{X})$ یک مخروط محدب است که شامل κ است. بررسی ساده نشان می‌دهد که ϕ یک تابع صعودی و محدب با مقادیر حقیقی روی $C_\kappa(\mathcal{X})$ است. همچنین، اگر ϕ از بالا پیوسته باشد، طبق گزاره الف، نمایش دوگان (۴) را دارد. مزدوج محدب آن به شکل زیر تعریف می‌شود:

$$\phi^*(\mu) = \sup_{f \in C_\kappa(\mathcal{X})} \left\{ \int f d\mu - \inf_{h \in \mathcal{H}: h \geq f} \int h d\mu_0 \right\}$$

$$= \sup_{h \in \mathcal{H}} \sup_{f \in C_\kappa(\mathcal{X}): h \geq f} \left\{ \int f d\mu - \int h d\mu_0 \right\}$$

$$= \sup_{h \in \mathcal{H}} \left\{ \int h d\mu - \int h d\mu_0 \right\}. \tag{۵}$$

از آنجا که $\mathcal{H}$ یک مخروط محدب است که حاوی ثابت ها است، نتیجه می‌گیریم که $\phi^*(\mu) = 0$ هرگاه $\mu \in ca_\kappa^+(\mathcal{X})$ یک اندازه احتمال باشد به طوری که $\int h d\mu = \int h d\mu_0$ برای همه $h \in \mathcal{H}$، و در غیر این صورت $\phi^*(\mu) = +\infty$. به خصوص، در صورتی که $C_\kappa(\mathcal{X}) = C_b(\mathcal{X})$ باشد، نتیجه می‌شود که نمایش دوگان (۲) بدست می‌آید.

۳.۴ قضیه ۲.۲

حال پس از بیان پیشنیاز های لازم قضیه ی زیر را داریم:
فرض کنید $f \in C_b(\mathcal{X})$ باشد. فرض کنید $\pi \in \mathcal{Q}$ وجود داشته باشد به طوری که $\pi \ll \theta$ و $\int \beta^* \left(\frac{d\pi}{d\theta}\right) d\theta < \infty$. در این صورت، داریم:

$$\phi_{\theta,\gamma}(f) = \max_{\mu \in \mathcal{Q}} \left\{ \int f d\mu - \frac{1}{\gamma} \int \beta^* \left(\frac{d\mu}{d\theta} \right) d\theta \right\}. \tag{۶}$$

علاوه بر این، داریم:

$$\phi_{\theta,\gamma}(f) - \frac{\beta(0)}{\gamma} \leq \phi(f) \leq \phi_{\theta,\gamma}(f) + \frac{1}{\gamma} \int \beta^* \left(\frac{d\mu_\varepsilon}{d\theta} \right) d\theta + \varepsilon \tag{۷}$$

هرگاه $\mu_\varepsilon \in \mathcal{Q}$ یک ε–بهینه‌ساز (۲) باشد به طوری که $\mu_\varepsilon \ll \theta$ و $\int \beta_\gamma^* \left(\frac{d\mu_\varepsilon}{d\theta}\right) d\theta < \infty$ باشد. اگر $\hat{h} \in \mathcal{H}$ یک کمینه‌کننده (۳) باشد، آنگاه $\hat{\mu} \in \mathcal{P}(\mathcal{X})$ که به صورت زیر تعریف می‌شود، یک بیشینه‌کننده (۶) است:

$$\frac{d\hat{\mu}}{d\theta} := \beta_\gamma'(f - \hat{h}) \tag{۸}$$

برهان: ابتدا با اثبات این نشان می‌دهیم که $\phi_{\theta,\gamma}$ تابعی حقیقی مقدار و پیوسته از بالا روی $C_b(\mathcal{X})$ است. برای این منظور، همانطور که در معرفی قضیه دنیل-استون نشان داده شده است، داریم $\phi^*(\mu)=\sup_{h\in\mathcal{H}}\left(\int hd\mu-\int hd\mu_0\right)$ برای هر $\mu\in\mathcal{P}(\mathcal{X})$. بنابراین، داریم:

$$\mu\in\mathcal{Q} \text{ اگر و تنهااگر } \int hd\mu=\int hd\mu_0 \text{ برای همه } h\in\mathcal{H}.$$

از آنجا که $\beta_\gamma(x)\geq xy-\frac{1}{\gamma}\beta^*(y)$ برای همه $x\in\mathbb{R}$ و $y\in\mathbb{R}_+$ برقرار است با در نظر گرفتن $x=f-h$ و $y=\frac{d\pi}{d\theta}$، نتیجه می گیریم که:

$$\int\beta_\gamma(f-h)d\theta\geq\int f-hd\pi-\frac{1}{\gamma}\int\beta^*\left(\frac{d\pi}{d\theta}\right)d\theta$$

بنابراین، داریم:

$$\begin{aligned}\phi_{\theta,\gamma}(f)&=\inf_{h\in\mathcal{H}}\left\{\int hd\pi+\int\beta_\gamma(f-h)d\theta\right\}\\&\geq\int fd\pi-\frac{1}{\gamma}\int\beta^*\left(\frac{d\pi}{d\theta}\right)d\theta>-\infty\end{aligned}$$

برای هر $f\in C_b(\mathcal{X})$. این نشان می‌دهد که $\phi_{\theta,\gamma}$ بر روی $C_b(\mathcal{X})$ تابعی حقیقی مقدار است. بعلاوه، فرض کنید (f^k) یک دنباله در $C_b(\mathcal{X})$ باشد به طوری که $f^k\downarrow f$. برای هر $h\in\mathcal{H}$، قضیه همگرایی یکنوا نتیجه می‌دهد که $\int\beta_\gamma\left(f^k-h\right)d\theta\ \to\ \int\beta_\gamma(f-h)d\theta$، بنابراین $\phi_{\theta,\gamma}\left(f^k\right)\downarrow\phi_{\theta,\gamma}(f)$. بنابراین، از قضیه غیرخطی دنیل-استون نتیجه می‌شود که:

$$\phi_{\theta,\gamma}(f)=\max_{\mu\in\mathcal{P}(\mathcal{X})}\left\{\int fd\mu-\phi^*_{\theta,\gamma}(\mu)\right\} \text{ برای همه } f\in C_b(\mathcal{X}),$$

که مزدوج محدب همگشت $\inf_{f\in C_b(\mathcal{X})}\{\phi(f)+\psi_{\theta,\gamma}(\cdot-f)\}$ به عنوان مجموع مزدوج های محدب ϕ^* و $\psi^*_{\theta,\gamma}$ به دست می‌آید. با توجه به قسمت اول اثبات، اگر $\mu\in\mathcal{Q}$، آنگاه $\phi^*(\mu)=0$ و در غیر این صورت $\phi^*(\mu)=+\infty$ است. علاوه بر این،

$$\begin{aligned}\psi^*_{\theta,\gamma}(\mu)&=\sup_{f\in C_b(\mathcal{X})}\left\{\int fd\mu-\int\beta_\gamma(f)d\theta\right\}\\&=\sup_{f\in C_b(\mathcal{X})}\left\{\int f\frac{d\mu}{d\theta}-\beta_\gamma(f)d\theta\right\}=\int\beta^*_\gamma\left(\frac{d\mu}{d\theta}\right)d\theta\end{aligned}$$

اگر $\mu\ll\theta$ باشد و در غیر این صورت $\psi^*_{\theta,\gamma}(\mu)=+\infty$ است. بنابراین مزدوج محدب $\phi^*_{\theta,\gamma}(\mu)$ به شکل زیر تعریف می‌شود:

$$\phi^*_{\theta,\gamma}(\mu)=\phi^*(\mu)+\psi^*_{\theta,\gamma}(\mu)=\begin{cases}\frac{1}{\gamma}\int\beta^*\left(\frac{d\mu}{d\theta}\right)d\theta & \text{اگر } \mu\in\mathcal{Q} \text{ و } \mu\ll\theta\\ \infty & \text{در غیر این صورت}\end{cases}$$

حال می‌خواهیم که با مقایسه حد پایین و حد بالا برای تابع $\phi_{\theta,\gamma}(f)$، رابطه‌ای به شکل معادله ۷ را اثبات کنیم. از طرفی داریم:

$$\begin{aligned}\phi_{\theta,\gamma}(f) &= \inf_{h\in\mathcal{H}}\left\{\int hd\mu_0 + \int \beta_\gamma(f-h)d\theta\right\} \\ &\leq \inf_{\substack{h\in\mathcal{H}\\ h\geq f}} \int hd\mu_0 + \beta_\gamma(0) = \phi(f) + \frac{\beta(0)}{\gamma}.\end{aligned}$$

که با استفاده از ویژگی غیرنزولی بودن $\beta_\gamma(\cdot)$ و دانستن این نکته که سمت چپ تمام انتخاب‌های ممکن برای h را (شامل آن‌هایی که کوچکتر از f هستند) در نظر می‌گیرد، در حالی که سمت راست فقط انتخاب‌های بزرگتر یا مساوی f را در نظر می‌گیرد، قابل اثبات است.
همچنین با در نظر گرفتن یک ε-بهینه‌ساز $\mu_\varepsilon \in \mathcal{Q}$ برای ۲ به شرطی که $\mu_\varepsilon \ll \theta$، نامساوی زیر حاصل می‌شود:

$$\begin{aligned}\phi(f) &\leq \int fd\mu_\varepsilon + \varepsilon \leq \int fd\mu_\varepsilon - \phi^*_{\theta,\gamma}(\mu_\varepsilon) + \phi^*_{\theta,\gamma}(\mu_\varepsilon) \\ &+ \varepsilon \leq \phi_{\theta,\gamma}(f) + \frac{1}{\gamma}\int \beta^*\left(\frac{d\mu_\varepsilon}{d\theta}\right)d\theta + \varepsilon.\end{aligned}$$

توافق $-\infty + \infty = +\infty$ نیز در نظر گرفته شده است.
و در نهایت برای اثبات بخش آخر قضیه $\hat{h} \in \mathcal{H}$ را کمینه کننده معادله $\phi_{\theta,\gamma}(f) = \int \hat{h}d\mu_0 + \int \beta_\gamma(f-\hat{h})d\theta$ در نظر بگیرید.
با معرفی پارامتر λ و تعریف $h_\lambda := \hat{h} + \lambda h$ برای هر $h \in \mathcal{H}$، می‌توانیم شرط مرتبه اول(برای یک مسئله بهینه‌سازی، شرط مرتبه اول مشتق‌پذیری تابع هدف نسبت به متغیرهای مستقل را در نقطه بهینه بیان می‌کند و این شرط باید برآورده شود.) را مورد بررسی قرار دهیم.
شرط مرتبه اول بیان می‌کند که با گرفتن مشتق تابع هدف نسبت به پارامتر λ و قرار دادن $\lambda = 0$، باید مشتق در نقطه مذکور صفر شود. این عبارت به صورت زیر است:

$$\left.\frac{d}{d\lambda}\right|_{\lambda=0}\left(\int h_\lambda d\mu_0 + \int \beta_\gamma(f-h_\lambda)d\theta\right) = 0.$$

با مشتق‌گیری رابطه زیر به دست می‌آید:

$$\int hd\mu_0 - \int \beta'_\gamma(f-\hat{h})hd\theta = 0.$$

این معادله نشان می‌دهد که اندازه احتمال $\hat{\mu}$ با مشتق رادون-نیکودیم $\frac{d\hat{\mu}}{d\theta} := \beta'_\gamma(f-\hat{h})$، برای همه $h \in \mathcal{H}$، شرط $\int hd\mu_0 = \int hd\hat{\mu}$ را برآورده می‌کند. علاوه بر این، با توجه به ویژگی گفته شده در بخش اول اثبات، داریم $\hat{\mu} \in \mathcal{Q}$.

با انتگرال‌گیری اتحاد $\beta_\gamma(x) = x\beta'_\gamma(x) - \beta^*_\gamma\left(\beta'_\gamma(x)\right)$ (که در آن $y = \beta'_\gamma(x)$) با $x = f - \hat{h}$ نسبت به θ، به عبارت زیر می‌رسیم:

$$\int \beta_\gamma(f - \hat{h})d\theta = \int f - \hat{h}d\hat{\mu} - \int \beta^*_\gamma\left(\frac{d\hat{\mu}}{d\theta}\right)d\theta.$$

این معادله نشان می‌دهد که:

$$\phi_{\theta,\gamma}(f) = \int \hat{h}d\mu_0 + \int \beta_\gamma(f - \hat{h})d\theta = \int fd\hat{\mu} - \int \beta^*_\gamma\left(\frac{d\hat{\mu}}{d\theta}\right)d\theta.$$

بنابراین، $\hat{\mu} \in \mathcal{Q}$ به عنوان یک بیشینه‌ساز برای معادله ۶ عمل می‌کند.$\square$

۵ تقریب تابع پوشش ریسک کمینه از بالا

در این بخش، یک دنباله $\mathcal{H}^1 \subseteq \mathcal{H}^2 \subseteq \cdots$ از زیرمجموعه‌های $\mathcal{H}$ را در نظر می‌گیریم و $\mathcal{H}^\infty := \bigcup_{m\in\mathbb{N}} \mathcal{H}^m$ را تعریف می‌کنیم. برای هر $m \in \mathbb{N} \cup \{+\infty\}$، تابع پوشش ریسک کمینه از بالای تقریبی را به صورت زیر تعریف می‌کنیم:

$$\phi^m(f) := \inf\left\{\int hd\mu_0 : h \geq f \text{ برای برخی } h \in \mathcal{H}^m\right\}$$

برای تقریب $\phi(f)$ با $\phi^m(f)$، به شرط چگالی زیر روی مجموعه $\mathcal{H}^\infty$ نیاز داریم.
شرط (D): برای هر $\varepsilon > 0$ و $\mu \in \mathcal{P}(\mathcal{X})$ اعمال می‌شود:
(a) برای هر $h \in \mathcal{H}$، $h' \in \mathcal{H}^\infty$ موجود است به طوری که $\int |h - h'|\, d\mu \leq \varepsilon$.
(b) $h'' \in \mathcal{H}^\infty$ موجود است به طوری که $1_{K^c} \leq h''$ و $\int h''d\mu \leq \varepsilon$ برای برخی از زیرمجموعه‌های فشرده K از $\mathcal{X}$.
شرط (D) به ما می‌گوید که زیرمجموعه $\mathcal{H}^\infty$ باید به طور کامل در فضای احتمالات $\mathcal{P}(\mathcal{X})$ تکثیر شود. بخش (a) نشان می‌دهد که هر تابع $h \in \mathcal{H}$ می‌تواند به دقت به تابع $h' \in \mathcal{H}^\infty$ نزدیک شود، به این معنی که انتگرال مطلق اختلاف آنها نسبت به اندازه احتمالاتی μ کمتر از ε باشد. بخش (b) نشان می‌دهد که می‌توانیم یک تابع $h'' \in \mathcal{H}^\infty$ پیدا کنیم که دارای تابع مشخصه ای همچون (1_{K^c}) باشد و انتگرال آن نسبت به اندازه احتمالاتی μ کمتر از ε باشد، که در آن K یک زیرمجموعه فشرده از $\mathcal{X}$ است.
با اعمال شرط (D)، تضمین می‌شود که دنباله $\phi^m(f)$ به طور کامل به $\phi(f)$ همگرا می‌شود، به این معنی که هر چقدر که m بیشتر می‌شود، تقریب $\phi^m(f)$ به $\phi(f)$ بهتر می‌شود. در واقع، با گسترش زیرمجموعه‌های $\mathcal{H}^m$، تابع پوشش ریسک کمینه از بالا تقریبی $\phi^m(f)$ به تابع پوشش ریسک کمینه از بالا $\phi(f)$ نزدیک‌تر می‌شود. با استفاده از شرط (D) می‌توانیم قضیه زیر را اثبات کنیم.

گزاره ۳.۲: فرض کنید $\mathcal{H}^\infty$ یک فضای تابعی خطی است که حاوی توابع ثابت می‌باشد. با فرض شرط (D) داریم:

$$\lim_{m\to\infty} \phi^m(f) = \phi^\infty(f) = \phi(f)$$

برای همه $f \in C_b(\mathcal{X})$.
برهان: $f \in C_b(\mathcal{X})$ را ثابت در نظر می‌گیریم. اینکه $\lim_{m\to\infty} \phi^m(f) = \phi^\infty(f) \geq \phi(f)$ نتیجه‌ای از تعریف $\mathcal{H}^\infty$ است. علاوه بر این، برای هر $\varepsilon > 0$، شرط (D) تضمین می‌کند که $h \in \mathcal{H}^\infty$ و $K \subseteq \mathcal{X}$ به طوری که $h \leq 1_{K^c}$ و $\varepsilon \leq \int hd\mu_0$. بنابراین، $\phi^\infty(1_{K^c}) \leq$ $\varepsilon \leq \int hd\mu_0$ و لم $Dini$ نشان می‌دهد که ϕ^∞ در $C_b(\mathcal{X})$ از بالا پیوسته است. با استفاده از گزاره دنیل-استون غیرخطی به دست می‌آید:

$$\phi^\infty(f) = \max_{\mu\in\mathcal{P}(\mathcal{X})} \left\{ \int f d\mu - \phi^{\infty *}(\mu) \right\}$$

مشابه ۵، مزدوج محدب آن به صورت زیر تعریف می‌شود:

$$\phi^{\infty *}(\mu) = \sup_{h\in\mathcal{H}^\infty} \left(\int hd\mu - \int hd\mu_0 \right) \leq \sup_{h\in\mathcal{H}} \left(\int hd\mu - \int hd\mu_0 \right) = \phi^*(\mu)$$

باقی می‌ماند که نشان دهیم برای $h \in \mathcal{H}$ و $\mu \in \mathcal{P}(\mathcal{X})$ با $\int hd\mu - \int hd\mu_0 > 0$، وجود دارد $h' \in \mathcal{H}^\infty$ به طوری که $\int h'd\mu - \int h'd\mu_0 > 0$. اما این به طور مستقیم از بخش اول شرط (D) برای اندازه احتمال $\frac{1}{2}\mu + \frac{1}{2}\mu_0$ بدست می‌آید. در واقع، دنباله‌ای به نام (h^n) در $\mathcal{H}^\infty$ وجود دارد به طوری که $h^n \to h$ در $L^1(\mu)$ و $L^1(\mu_0)$، که نشان می‌دهد که $\int h^n d\mu - \int h^n d\mu_0 > 0$ برای n بزرگ کافی است.□
با فرض یک اندازه نمونه‌برداری θ و یک تابع جریمه پارامتری β_γ که در بخش قبلی تعریف شده است، ما نسخه تقریبی از تابع پوشش ریسک کمینه از بالا که نرمال سازی شده است را به صورت زیر تعریف می‌کنیم:

$$\phi^m_{\theta,\gamma}(f) = \inf_{h\in\mathcal{H}^m} \left\{ \int hd\mu_0 + \int \beta_\gamma(f-h)d\theta \right\}$$

برای همه $f \in C_b(\mathcal{X})$. به عنوان نتیجه‌ای از دو مرحله تقریبی $\phi_{\theta,\gamma}(f) \to \phi(f)$ برای $\gamma \to \infty$ در قضیه ۲.۲ و $\phi^m(f) \to \phi(f)$ برای $m \to \infty$ در گزاره ۳.۲، نتیجه زیر را به دست می‌آوریم.
گزاره ۴.۲: فرض کنید $\mathcal{H}^\infty$ شرط (D) را برآورده می‌کند و برای هر $\varepsilon > 0$ یک بهینه‌ساز ε-به نام μ_ε از (۶) وجود دارد که $\theta \gg \mu_\varepsilon$ و $\int \beta^* \left(\frac{d\mu_\varepsilon}{d\theta}\right) d\theta < +\infty$. سپس می‌خواهیم ثابت کنیم که برای هر تابع $f \in C_b(\mathcal{X})$، $\phi^m_{\theta,\gamma}(f)$ به $\phi(f)$ میل می‌کند هنگامی که $\min\{m,\gamma\} \to \infty$.
برهان: ابتدا توجه می‌کنیم که:

$$\phi^m(f) + \beta_\gamma(0) \geq \inf_{\substack{h\in\mathcal{H}^m \\ h\geq f}} \left\{ \int hd\mu_0 + \int \beta_\gamma(f-h)d\theta \right\} \geq \phi^m_{\theta,\gamma}(f) \geq \phi_{\theta,\gamma}(f)$$

که در آن اولین نامساوی از صعودی بودن تابع β_γ استفاده کرده‌ایم، دومین نامساوی فقط شرط $h \geq f$ را حذف کرده است و سومین نامساوی نتیجه‌ای از $\mathcal{H}^m \subseteq \mathcal{H}$ است.
حالا یک $\varepsilon > 0$ دلخواه در نظر بگیرید. با توجه به شرط (D) و قضیه ۲.۲، اعداد طبیعی m_0 و γ_0 وجود دارد به‌طوری که:

$$\phi^m(f) \leq \phi(f) + \varepsilon \quad \text{و} \quad \phi(f) \leq \phi_{\theta,\gamma}(f) + \varepsilon$$

برای همه $m \geq m_0$ و $\gamma \geq \gamma_0$. این نشان می‌دهد که:

$$\phi(f) + \varepsilon + \frac{\beta(0)}{\gamma} \geq \phi^m(f) + \beta_\gamma(0) \geq \phi^m_{\theta,\gamma}(f) \geq \phi_{\theta,\gamma}(f) \geq \phi(f) - \varepsilon$$

برای همه $m \geq m_0$ و $\gamma \geq \gamma_0$. این نتیجه می‌دهد که $\phi^m_{\theta,\gamma}(f)$ به $\phi(f)$ میل می‌کند هرگاه $\min\{m,\gamma\} \to \infty$. $\square$

۶ حل با شبکه‌های عصبی

در ادامه مقاله تلاش می‌کند حل مساله روی فضای H^m را به بهینه‌سازی یک شبکه عصبی مصنوعی پیش‌رونده تقلیل دهد. شبکه عصبی مورد بررسی، یک MLP به فرم زیر است:

$$\mathbb{R}^d \ni x \mapsto A_l \circ \underbrace{\varphi \circ A_{l-1}}_{(l-1)layer} \circ \cdots \circ \underbrace{\varphi \circ A_0}_{1.layer}(x)$$

که تمام تبدیل‌های A_l آفین هستند و تمام توابع ϕ توابع غیرخطی فعال‌سازی هستند که شرایط خاصی داشته باشند. بعد ورودی شبکه d است و بعد لایه‌های پنهان m است و خروجی شبکه در نهایت یک عدد حقیقی است که همان مقدار $\phi(f)$ در مساله اصلی خواهد بود. هر تبدیل آفین به فرم یک ضرب ماتریسی از وزن‌های شبکه عصبی به علاوه یک عبارت بایاس به فرم زیر قابل نمایش است:

$$A_j(x) = M_j x + b_j$$

تمامی این وزن‌ها و بایاس‌ها پارامترهای شبکه عصبی هستند که برای یک D طبیعی خاص می‌توان آن‌ها را به عنوان عضوی از R^D در نظر گرفت.

۱.۶ مقدمات

در ادامه، تمام پارامترهای ممکن برای شبکه عصبی با ساختاری ثابت را در مجموعه‌ای مثل $\Xi \subset \mathbb{R}^D$ تعریف می‌کنیم. برای یک $\xi \in \Xi$ خاص هم یک شبکه عصبی خاص را می‌توان با l لایه و بعد ورودی d و بعد لایه‌های پنهان m به فرم زیر نمایش داد:

$$N_{l,d,m}(\xi) = A_l \circ \varphi \circ A_{l-1} \circ \cdots \circ \varphi \circ A_0$$

و همچنین تمام شبکه‌های عصبی این چنینی را در یک مجموعه بزرگ‌تر به فرم زیر می‌ریزیم:

$$\mathfrak{N}_{l,d,m}(\Xi)$$

. مقاله فرض می‌کند که تابع فعالسازی پیوسته، غیرنزولی و در منفی بی‌نهایت به صفر و در بی‌نهایت به ۱ همگراست. ولی در ادامه بحث می‌کند که در عمل در بسیاری از موارد، این فرض ضروری نیست و می‌توان از توابع فعالسازی دیگری که شرایط ریلکس‌تری دارند مثل تابع $RelU$ نیز استفاده کرد.

۲.۶ مدلسازی

در ادامه تلاش می‌کنیم h و همچنین h^m را به کمک شبکه‌های عصبی مدلسازی کنیم.
فضای H نمایشی به فرم زیر دارد:

$$\mathcal{H} = \left\{ \sum_{j=1}^{J} e_j h_j \circ \pi_j + a : h_j \in C_b\left(\mathbb{R}^{d_j}\right), a \in \mathbb{R} \right\},$$

که در آن $e_j \in C_b(\mathcal{X})$ و $\pi_j : \mathcal{X} \to \mathbb{R}^{d_j}$ به ازای تمام j های مجاز توابعی پیوسته هستند.
همچنین داریم:

$$\mathcal{H}^{\infty} = \left\{ \sum_{j=1}^{J} e_j h_j \circ \pi_j + a : h_j \in \mathfrak{N}_{l_j,d_j}, a \in \mathbb{R} \right\}$$

و

$$\mathcal{H}^{m} = \left\{ \sum_{j=1}^{J} e_j h_j \circ \pi_j + a : h_j \in \mathfrak{N}_{l_j,d_j,m}\left(\Xi_{j,m}\right), a \in \mathbb{R} \right\}$$

بنابراین مساله اصلی قابل مدلسازی به فرم زیر است:

$$
\begin{aligned}
\phi_{\theta,\gamma}^{m}(f) &= \inf_{h\in\mathcal{H}^m}\left\{\int h d\mu_0 + \int \beta_\gamma(f-h)d\theta\right\}\\
&= \inf_{a\in\mathbb{R}}\inf_{h_j\in\mathfrak{N}_{l_j,d_j,m}(\Xi_{j,m})}\left\{\int\sum_{j=1}^{J} e_j h_j\circ\pi_j d\mu_0 + a\right.\\
&\quad\left. + \int\beta_\gamma\left(f-\sum_{j=1}^{J} e_j h_j\circ\pi_j - a\right)d\theta\right\}\\
&= \inf_{a\in\mathbb{R}}\inf_{\xi_j\in\Xi_{j,m}}\left\{\int\sum_{j=1}^{J} e_j N_{l_j,d_j,m}\left(\xi_j\right)\circ\pi_j d\mu_0 + a\right.\\
&\quad\left. + \int\beta_\gamma\left(f-\sum_{j=1}^{J} e_j N_{l_j,d_j,m}\left(\xi_j\right)\circ\pi_j - a\right)d\theta\right\}
\end{aligned}
$$

پس در واقع بهینه‌سازی مساله اصلی به بهینه‌سازی نسبت به پارامترهای شبکه عصبی تقلیل یافت. مقاله در ادامه ثابت می‌کند که این دو بهینه‌سازی به یکدیگر همگرا هستند و بنابراین با حل شبکه عصبی، حل مساله اصلی را نیز یافته‌ایم.

۷ مثال‌های عددی

در این بخش به معرفی کوتاهی از سایر مسائلی که توسط مقاله حل شده‌اند می‌پردازیم.

۱.۷ حمل و نقل بهینه و مرزهای فرشهت-هوفدینگ

در این مسئله، اثرات توابع جریمه مختلف، ضریب جریمه، اندازه دسته، و تعداد تکرارها را مورد بررسی قرار می‌دهیم. متغیرها را به شکل زیر تعریف می‌کنیم: $\mathcal{X} = [0,1]^d$ ، $\theta = \mathcal{U}\left([0,1]^d\right)$ (که $\mathcal{U}(\cdot)$ توزیع یکنواخت را نشان می‌دهد) و $\mathcal{Q} = \{v\in\mathcal{P}(\mathcal{X}) : v_i = \mathcal{U}([0,1])\}$ ، جایی که ν_i برای حاشیه i–ام از ν است. برای برخی نقطه ثابت $z\in[0,1]^d$، تابع $f : [0,1]^d \to \mathbb{R}_+$ را به شکل زیر تعریف می‌کنیم:

$$
f(x) = \begin{cases} 1, & \text{اگر } x_i \le z_i \text{ برای همه } i\in\{1,2,\ldots,d\}\\ 0, & \text{بقیه حالات}\end{cases}
$$

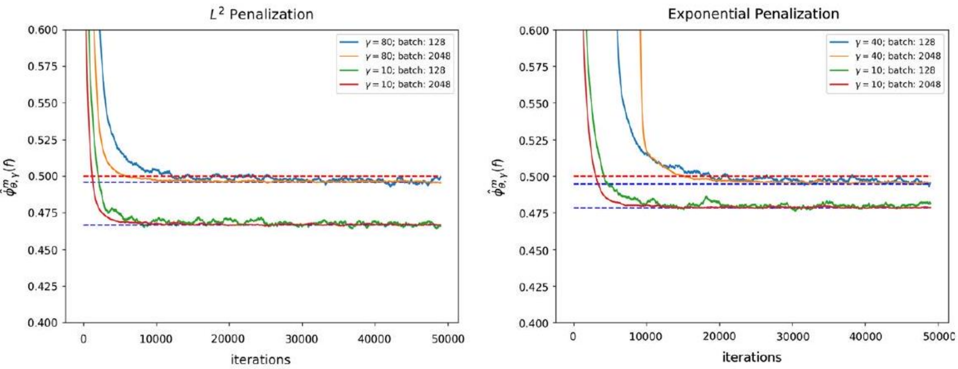
در اینجا، $z\in[0,1]^d$ نقطه‌ای ثابت در فضای d بعدی $\mathcal{X}$ است. تابع f یک نقطه x را به عنوان ورودی می‌گیرد و اگر هر جزء آن کوچکتر یا مساوی مقدار متناظر در z باشد، خروجی ۱ و گرنه خروجی ۰ می‌دهد. به عبارت دیگر، اگر x درون یا روی مکعب تعریف شده توسط z باشد، $f(x)$ برابر با ۱ است و اگر خارج از مکعب باشد، $f(x)$ برابر با ۰ است.

ما به دنبال بالاترین مقدار میانگین f بر روی همه توزیع‌های ممکن در $\mathcal{Q}$ هستیم. این ارزش ریاضی به صورت زیر تعریف می‌شود:

$$\phi(f) = \sup_{v \in \mathcal{Q}} \int f dv$$

ارزش $\phi(f)$ متناظر با حداکثر ارزش یک کوپولا d بعدی در نقطه z است. با استفاده از مرزهای فرشهت–هوفدینگ، ما یک راه‌حل تحلیلی برای این مسئله داریم که به صورت زیر است:

$$\phi(f) = \min_{i \in \{1,\ldots,d\}} z_i$$



شکل ۲: مرزهای فرشهت–هوفدینگ: $d = 2, z_1 = 0.5, z_2 = 0.75$. مقایسه تابع جریمه L^2 با $\beta_\gamma(x) = \gamma \max\{0, x\}^2$ و تابع جریمه نمایی با $\beta_\gamma(x) = \frac{\exp(\gamma x - 1)}{\gamma}$. ارزش‌ها نمودار شده، میانگین اجراها در طول ۱۰۰۰ تکرار آخر است. خط قرمز نقطه‌چین نمایش ارزش واقعی $\phi(f)$ است. خطوط آبی نقاط‌چین مرزهای پایینی $\phi_{\theta,\gamma}(f)$ هستند که با استفاده از روابط گفته شده در قضیه ۲.۲ برای انتخاب‌های متناسب از γ به دست می‌آیند.

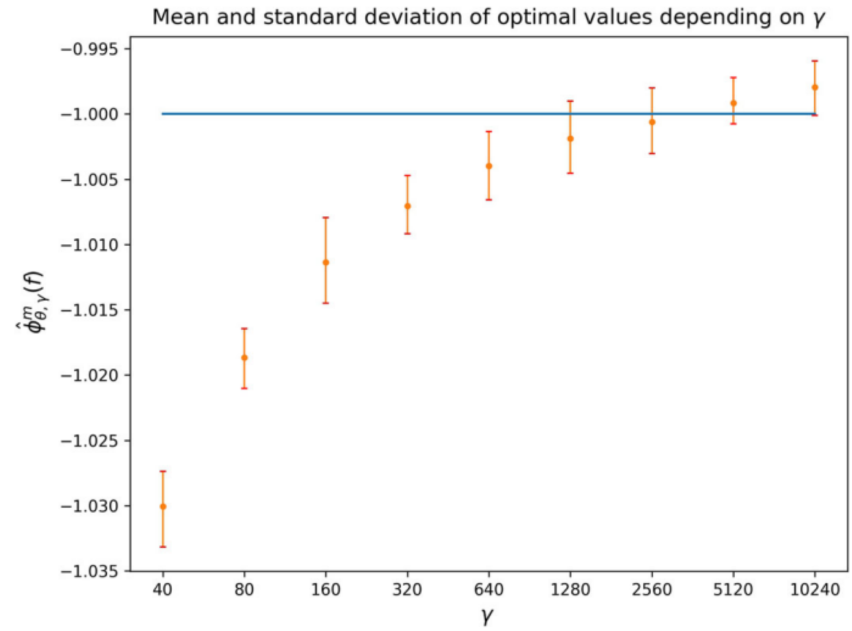
در شکل ۲، نحوه وابستگی $\hat{\phi}^m_{\theta,\gamma}(f)$ به تعداد تکرارهای بهینه‌ساز $Adam$ و اندازه دسته $(batchsize)$ نشان داده شده است. مشاهده می‌شود که در حالی که اندازه دسته‌های بزرگتر به مقدار $\hat{\phi}^m_{\theta,\gamma}(f)$ کمک می‌کنند، اما به سرعت همگرایی کمک زیادی نمی‌کند. مشاهده می‌شود که با افزایش اندازه دسته $(batchsize)$، همگرایی پایدارتری حاصل می‌شود. اما سرعت همگرایی به طور قوی به اندازه دسته مرتبط نیست. این نشان می‌دهد که افزایش اندازه دسته ممکن است به همگرایی سریع و در نهایت پایدار منجر شود.
به علت پایداری بیشتر، ما بیشتر از جریمه L^2 برای برنامه‌های بعدی استفاده خواهیم کرد. علاوه بر این، شکل نشان می‌دهد که راه‌حل‌های عددی به طور تقریبی حداقل‌های $\phi_{\theta,\gamma}(f)$ را به عنوان مشخص شده در روابط قضیه ۲.۲ به دست می‌آورند. به این معنا که تقریباً داریم $\phi(f) \approx \phi^m_{\theta,\gamma}(f) + \frac{1}{\gamma} \int \beta^* \left(\frac{d\hat{\mu}}{d\theta}\right) d\theta$ که در آن $\hat{\mu}$ یک بهینه‌ساز $\phi(f)$ است.

۲.۷ مسئله حمل و نقل بهینه مارتینگل

در حمل و نقل بهینه مارتینگل، مسئله حمل و نقل بهینه با تعیین یک محدودیت مارتینگل بر روی محدودیت‌های حاشیه‌ای گسترش می‌یابد. ابعاد در اینجا به عنوان مراحل زمانی گسسته در نظر گرفته می‌شوند و اندازه‌ها در $\mathcal{Q}$ توزیع‌های فرآیندهای تصادفی گسسته $(X_t)_{t=1,\dots,d}$ با توزیع‌های حاشیه‌ای ثابت همچنین شرطی که فرآیند یک مارتینگل است، هستند.
در اینجا، ما به یک مثال ساده با $d = 2$ توجه می‌کنیم، که یک راه‌حل آنالیتیکال برای آن شناخته شده است. این مثال از[۱۶] اقتباس شده است. مجموعه $\mathcal{X} := [-1,1]\times[-2,2], \theta := \mathcal{U}(\mathcal{X})$ را تعریف کنیم و مقادیر زیر را در نظر بگیریم:

$$\mathcal{Q} := \{v = v_1 \otimes K : v_1 = \mathcal{U}([-1,1]), v_2 = \mathcal{U}([-2,2]),$$
$$x = \int_{-2}^{2} yK(x,dy) \text{ همه‌ی جاهای بی‌شمار } v_1 \text{ نسبت به } \Big\}.$$

برای $f = -|x-y|^\rho$ مقدار $\phi(f)$ برابر با -1 برای همه $\rho > 2$ است. ما این مسئله را با $\rho = 2.3$ پیاده‌سازی می‌کنیم، جایی که از تابع جریمه L^2 برای مقادیر مختلف γ استفاده می‌کنیم. نتایج در شکل ۳ نمایش داده شده است. می‌توان دید که تا مقادیر γ تا حدود ۱۲۸۰، رفتار ارزش بهینه تقریباً به همان شکلی است که در روابط قضیه ۲.۲ پیش‌بینی می‌شود، به این معنا که اگر γ را دو برابر کنیم، خطا تقریباً دو برابر کم می‌شود. اما برای مقادیر بزرگ‌تر از γ، عدم پایداری‌های عددی رخ می‌دهد و بهینه‌یاب نمی‌تواند بهینه واقعی را پیدا کند. این نشان‌دهنده این است که مقدار $\hat{\phi}^m_{\theta,\gamma}(f)$ بالاتر از $\phi(f) = -1$ است.



شکل ۳: میانگین ارزش‌های عددی بهینه و مرزهای اطمینان ۹۵ درصد برای ۱۰۰ اجرای مستقل برای ارزش‌های مختلف γ (جریمه L^2). شبکه به مدت ۲۰،۰۰۰ تکرار با اندازه دسته ۱۰۲۴ آموزش داده شده است. ارزش بهینه واقعی مسئله بدون جریمه برابر با -1 است.

۸ مسئله پیاده‌سازی شده: بهینه‌سازی سبد سهام در حضور عدم قطعیت وابستگی دارایی‌ها

در این بخش، مثال عددی مورد بحث در مقاله را تشریح خواهیم کرد و نحوه اعمال روش مقاله روی این مثال و همچنین حل آنالیزی آن را مورد بررسی قرار خواهید داد. در پایان، کد پروژه تلاش خواهد کرد این مساله را با روش مقاله پیاده‌سازی کند و مدعیات ریاضیاتی مقاله را در واقعیت اعتبارسنجی کند.

۱.۸ تشریح مساله و حل ریاضیاتی

فرض کنیم دو دارایی داریم که از امید ریاضی بازده و نیز واریانس بازده هرکدام به تنهایی مطلع هستیم، ولی از کوواریانس بازده‌های آن‌ها خبری نداریم. در واقع مساله، بهینه‌سازی سبد سهام در حضور عدم قطعیت وابستگی است. در چنین مسائلی، معمولاً یک معیار ریسک تعریف می‌شود. مقاله‌ی [۱۵] معیار ریسک زیر را پیشنهاد داده است:

$$E(Y_x) - \lambda . Var(Y_x)$$

که در آن x در واقع وزن دارایی دوم در سبد است و Y_x سبد سهام تشکیل شده از وزن‌های x و $1-x$ است. متغیر λ نیز ضریب ریسک‌گریزی سرمایه‌گذار است. واضح است که هدف هر سرمایه‌گذاری فارغ از ضریب ریسک‌گریزی، بیشینه‌سازی عبارت بالاست. ولی چون از کوواریانس دو دارایی مطلع نیستیم، در محاسبه صریح واریانس به مشکل می‌خوریم. پس راه حل جایگزین، در نظر گرفتن بدترین حالت ممکن برای کوواریانس دو دارایی است.
فرض کنیم دارایی اول بازده μ_1 و دارایی دوم بازده μ_2 دارد و همچنین فرض کنیم که واریانس دو دارایی به تنهایی نیز به ترتیب σ_1^2 و σ_2^2 باشد.
آنگاه به وضوح داریم:

$$E(Y_x) = (1-x).\mu_1 + x.\mu_2$$

$$Var(Y_x) = (1-x)^2.\sigma_1^2 + x^2.\sigma_2^2 + 2.Cov(K_1, K_2)$$

از آنجا که از مقدار واقعی کوواریانس خبر نداریم، آن را C می‌نامیم و بدترین حالت ممکن را در نظر می‌گیریم. پس در واقع تحلیل ما به فرم زیر است:

$$min_C((1-x).\mu_1 + x.\mu_2 - \lambda.((1-x)^2.\sigma_1^2 + x^2.\sigma_2^2 + 2.x.(1-x).Cov(K_1, K_2)))$$

این مینیمم به وضوح زمانی رخ می‌دهد که مقدار C ماکزیمم باشد، که به معنی بیشینه همبستگی بین دو دارایی است. حال به عنوان مثالی خاص، فرض می‌کنیم U و V دو متغیر تصادفی از توزیع یکنواخت $(0,1)$ باشند و داشته باشیم $W = 2.V^2$ اگر فرض کنیم بازده دارایی اول از توزیع U و بازده دارایی دوم از توزیع W است آنگاه داریم:

$$Y_x = (1-x).U + 2.x.V^2$$

پس

$$E(Y_x) = 0.5 * (1-x) + \frac{2.x}{3}$$

$$Var(Y_x) = \frac{(1-x)^2}{12} + \frac{16.x^2}{45} + 2.x.(1-x).C$$

ولی بیشینه C ممکن در زمان وابستگی کامل رخ می‌دهد که معادل است با زمانی که توزیع اساسی دو متغیر تصادفی یکسان باشد. یعنی U و V^2 با هم برابر باشند و یک متغیر تصادفی باشند. در چنین حالتی

$$Cov(U, 2.V^2) = 2.Cov(U, V^2)$$

که چون واریانس U برابر است با $\frac{1}{12}$ ماکزیمم این کوواریانس هم برابر واریانس هرکدام و بنابراین همین مقدار است. پس

$$maxCov(U, 2.V^2) = 2 * \frac{1}{12} = \frac{1}{6}$$

و با جایگذاری و ساده‌سازی داریم:

$$E(Y_x) - \lambda.Var(Y_x) =$$
$$0.5 + \frac{x}{6} - \lambda.(\frac{1}{12} + \frac{x}{6} + \frac{19.x^2}{180})$$

ولی حالا صرفا معیاری برای بدترین حالت ممکن داریم. برای بیشینه‌سازی سبد سهام، باید در این بدترین حالت، معیار بالا را بیشینه کنیم. پس برای سبد بهین با فرض عدم فروش استقراضی داریم:

$$R = max_{x\in[0,1]}(0.5 + \frac{x}{6} - \lambda.(\frac{1}{12} + \frac{x}{6} + \frac{19.x^2}{180}))$$

برای یافتن مقدار بیشینه کافیست از طرفین نسبت به وزن دارایی دوم مشتق بگیریم و برابر صفر قرار دهیم که به معادله زیر خواهیم رسید:

$$\frac{1}{6} - \lambda.(\frac{1}{6} + \frac{19.x}{90}) = 0$$

$$x = \frac{15.(1-\lambda)}{19.\lambda}$$

و با جایگذاری مقدار بالا در معادله اصلی هم مقدار بهین به دست خواهد آمد. در کد پیاده‌سازی شده، توابع $analyticalportfolio$ و $analyticaloptimalvalue$ همین مقادیر را پیاده‌سازی کرده‌اند.

۲.۸ پیاده‌سازی روش مقاله

حال به تشریح ارتباط روش مقاله با این مساله خواهیم پرداخت. فرض کنیم که معیار ریسک خود را با تابعی نشان می‌دهیم:

$$f_x = E(Y_x) - \lambda.Var(Y_x)$$

هدف یافتن مقدار زیر است:

$$\max_{x\in[0,1]}(\inf_{v\in Q}(\int f_x dv))$$

اما به طور خاص

$$\inf_{v\in Q}(\int f_x) = -\sup_{v\in Q}(\int(-f_x dv)) = -\phi(-f_x)$$

پس مساله اصلی عبارت است از

$$\sup_{x\in[0,1]}(-\phi(-f_x))$$

و با جایگذاری تابع f_x طبق تعریفش مساله اصلی به فرم زیر هم قابل بازنویسی است

$$\begin{aligned}&\sup_{x\in[0,1]} -\phi\left(-f_x\right)\\ &:= \sup_{x\in[0,1]}\inf_{v\in\mathcal{Q}}\int(1-x)\xi_1 + x\xi_2\\ &\quad -\lambda\left((1-x)\xi_1 + x\xi_2 - (1-x)\int_0^1 \zeta_1\theta_1\left(d\zeta_1\right) - x\int_0^2 \zeta_2\theta_2\left(d\zeta_2\right)\right)^2 v(d\xi).\end{aligned}$$

که با توجه به نکات مطرح شده می‌توان عبارت بالا را به فرم زیر هم نوشت:

$$\begin{aligned}&\sup_{x\in[0,1]} -\phi\left(-f_x\right)\\ &:= -\inf_{x\in[0,1]}\sup_{v\in\mathcal{Q}} -\int(1-x)\xi_1 + x\xi_2\\ &\quad -\lambda\left((1-x)\xi_1 + x\xi_2 - (1-x)\int_0^1 \zeta_1\theta_1\left(d\zeta_1\right) - x\int_0^2 \zeta_2\theta_2\left(d\zeta_2\right)\right)^2 v(d\xi).\end{aligned}$$

با نوشتن فرم دوگان و جریمه‌سازی به عبارت زیر می‌رسیم:

$$-inf_{x\in[0,1]\ \&\ h\in H}\int hd\mu_0 + \int\beta(-f-h)d\theta$$

پس کافیست یک شبکه عصبی طراحی کنیم که علاوه بر پارامترهای مربوط به تابع h متغیر اضافی وزن x را هم در گرادیان در نظر بگیرد و نسبت به هردوی این پارامترها کمینه‌سازی کند.

تنها مساله باقیمانده، تعیین اندازه احتمال نمونه‌گیری θ است. به این منظور، مقاله دو مرجع نمونه‌گیری معرفی کرده است. مرجع اول، مشابه همان مثال اصلی است. فرض کنیم θ_1 همان توزیع احتمال یکنواخت روی ۰ و ۱ باشد و θ_2 مستقل از θ_1 از دو برابر مربع یک توزیع یکنواخت باشد. پس به عبارتی:

$$\theta_1 \sim \mathcal{U}(0,1)$$

و اگر

$$r \sim \mathcal{U}(0,1)$$

آنگاه

$$\theta_2 = 2.r^2$$

همچنین فرض می‌کنیم فضای لهستانی ما به فرم زیر است:

$$\mathcal{X} = [0,1] * [0,2]$$

در این صورت، مرجع نمونه‌گیری اول مورد بررسی به فرم زیر تعریف می‌شود: $\theta^{(1)} = \theta_1 \otimes \theta_2$ مرجع دوم نیز به فرم زیر خواهد بود: $\theta^{(2)} = 0.5\theta^{(1)} + 0.5\,(\mathcal{U}([0,1]) \circ (\mathrm{Id}, \varphi)^{-1})$ پس در واقع مرجع دوم، به صورت نیم مرجع اول(حاصل ضرب دکارتی) و نیم همبستگی مطلق است. پس به نوعی یک همبستگی در این مرجع شبیه‌سازی شده است. از آنجایی که جواب بهین در همبستگی کامل پیدا می‌شود، توقع داریم که مرجع دوم نتایج نزدیک‌تری به نتایج آنالیزی و واقعی بدهد. در ضمن به این نکته هم باید توجه داشت که برای H از فضای توابع دو متغیره تفکیک‌پذیر استفاده میکنیم. پس برای هر h عضو H داریم:

$$h(x_1, x_2) = h_1(x_1) + h_2(x_2)$$

پس دو شبکه عصبی مختلف خواهیم داشت که یکی برای تخمین h_1 و دیگری برای تخمین h_2 استفاده می‌شود و سپس این دو را با هم جمع زده و برای محاسبه انتگرال، از تمام نمونه‌های جمع‌آوری شده، میانگین می‌گیریم.

در کد پیاده‌سازی شده در پروژه، ابتدا در دو تابع، نقاطی به اندازه دلخواه از دو توزیع ذکر شده تولید می‌کنیم. سپس شبکه عصبی دولایه‌ای با تعداد ابعاد پنهان(نورون‌ها)ی دلخواه به عنوان ورودی تعریف می‌کنیم. سپس تابع $createmodels$ شبکه‌های عصبی مورد نیاز را می‌سازد. اگر قرار باشد بیشتر از یک سهام در سبد داشته باشیم، در این بخش لازم است که از تابع مدلساز اصلی، به تعداد دارایی‌های موجود در سبد شبکه عصبی بسازیم.

در ادامه، توابع اصلی روش مقاله پیاده‌سازی می‌شوند. تابع h به عنوان حاصل جمع خروجی شبکه‌های عصبی به ازای هر ورودی پیاده‌سازی می‌شود. به عبارتی هر ورودی $data$ که شامل اعضایی به فرم (x_1, x_2) است به ترتیب در MLP های h_1, h_2 به اصطلاح $forwardpass$ می‌شود و نتیجه‌ها با هم جمع شده به عنوان خروجی تابع h بازگردانده می‌شوند. تابع انتگرال‌گیری h از تمام خروجی‌های بالا در دیتا، میانگین می‌گیرد. تابع $portfolioreturn$ یک وزن را دریافت کرده و یک دیتاست برمی‌گرداند که وزن دریافتی را روی بازده دارایی دوم اعمال کرده است. پس از معرفی یک تابع برای محاسبه واریانس، تابع f مطابق تعریف بالا نوشته می‌شود که یک سبد را می‌گیرد و معیار ریسک را روی آن محاسبه می‌کند. برای تضمین اینکه فروش استقراضی وجود نداشته باشد و حتما وزن هر سبد در بازه $[0,1]$ باشد، یک تابع جریمه در

قالب تابع $noshortsellingpenalty$ معرفی شده است که از طریق تابع $Relu$ تضمین می کند همیشه وزن ورودی، نامنفی و کمتر مساوی ۱ باشد. در ادامه در یک تابع مستقیما انتگرال تابع جریمه بتا را تعریف کرده‌ایم. برای تابع جریمه از تابع L^2 استفاده کرده‌ایم که به فرم زیر تعریف می‌شود

$$\beta(x) = \frac{1}{2}. \max(0, x)^2$$

$$\beta_\gamma(x) = \frac{1}{\gamma}.\beta(\gamma x)$$

پس داریم:

$$\beta_\gamma(x) = \frac{1}{2\gamma}. \max(0, \gamma x)^2$$

و طبق بحث بالا و مساله دوگان، به عنوان ورودی به تابع بتا مقدار $-f-h$ را در هر نقطه داده‌ایم. سپس از تمام مقادیر میانگین گرفته‌ایم تا به تقریبی عددی از انتگرال بتا برسیم. در پایان، تابع هدف طبق روش مقاله، به عنوان مجموع انتگرال‌های توابع $h, \beta(-f-h)$ و همچنین در صورت عدم وجود فروش استقراضی، تابع جریمه فروش استقراضی تعریف شد. سپس، شبکه عصبی نسبت به پارامترها و همچنین متغیر وزن سبد دوم و برای کمینه‌سازی این تابع هدف آموزش داده شد. هایپرپارامترهایی که جواب مطلوب دادند به قرار زیر بودند:

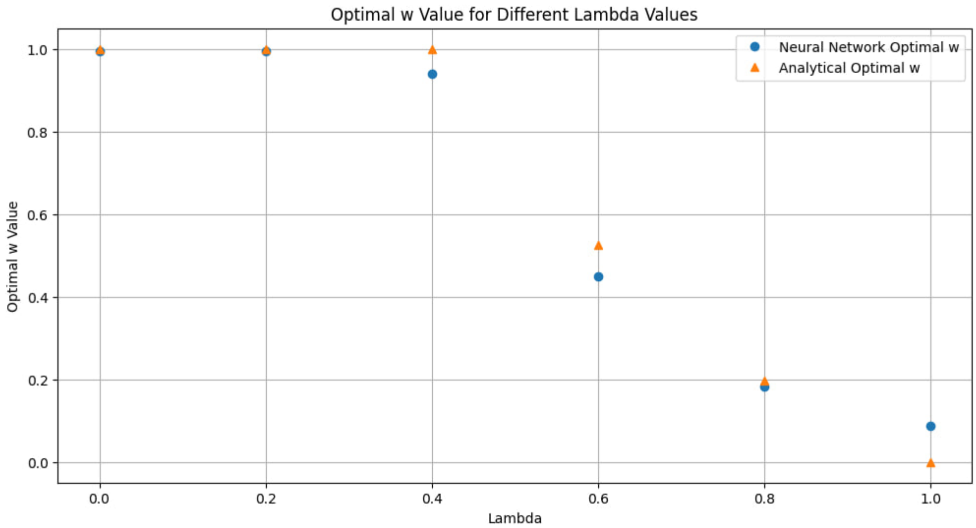
$$m = 32, \gamma = 160, learningrate = 0.001, \beta 1 = 0.99,$$

$$\beta 2 = 0.995, epochs = 10000, batchsize = 2^{13}$$

و به ازای مقادیر مختلف ضریب ریسک‌گریزی از ۰ تا ۱ با افزایش‌های ۲.۰ و همچنین از ۷.۰ تا ۸۷.۰ با افزایش‌های ۲.۰ مقادیر بهین به ازای دو توزیع مختلف محاسبه شدند و نمودارهایی ترسیم شدند. در پایان، می‌خواستیم تأثیرات تغییر پارامترهای m, γ را بسنجیم. ولی بودجه محاسباتی کافی برای آموزش کامل شبکه عصبی به ازای مقادیر مختلف نداشتیم. مشکل اصلی، اینرسی بالای متغیر w در مقابل تغییر در روند بهینه‌سازی بود که تکرارهای زیادی را می‌طلبید. پس از هیوریستیک مونته کارلو استفاده کردیم، به این شکل که مقادیری تصادفی از w انتخاب شدند، بهینه‌سازی فقط نسبت به پارامترهای شبکه عصبی انجام شد، سپس از پارامترهای شبکه‌های عصبی به دست آمده میانگین‌گیری شد و بعد فقط بهینه‌سازی نسبت به w انجام شد که میزان اینرسی را به شدت کاهش می‌دهد. نتایج برای مقدار $\lambda = 0.8$ بررسی شدند. نتایج گرچه دقیق نبودند، ولی نشان دادند که به ازای مقادیر مختلف m, γ هم همچنان جواب‌های به دست آمده در همسایگی معقولی از جواب واقعی هستند.

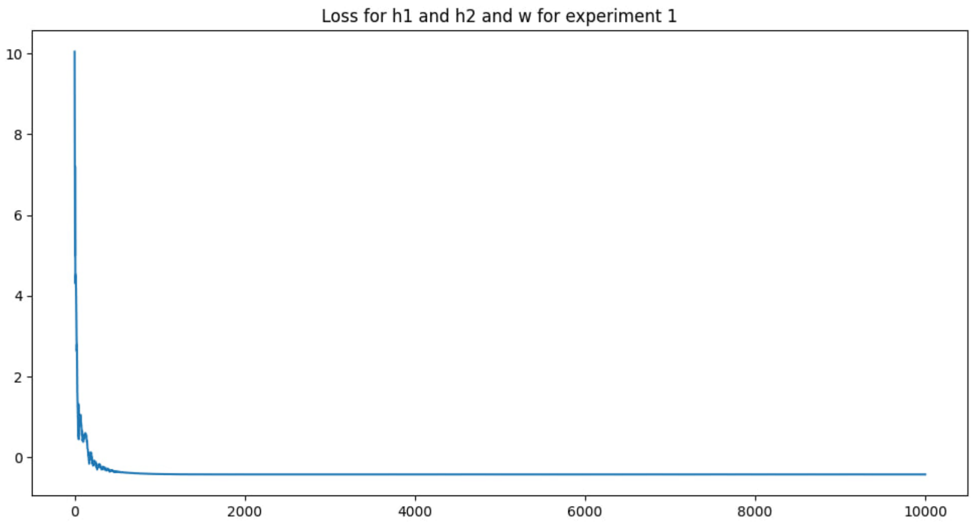
۳.۸ نتایج

در این قسمت گزارش به بیان نتایج به دست آمده از کد می‌پردازیم. نتایج زیر برای زمانی است که از $\theta^{(1)} = \theta_1 \otimes \theta_2$ به عنوان اندازه مرجع استفاده کردیم.



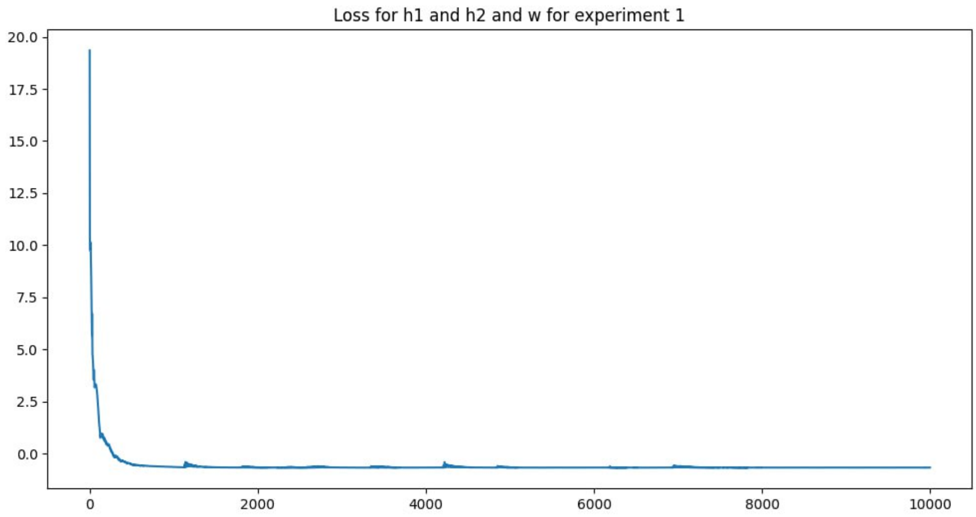
شکل ۴: مقادیر بهینه w برای مقادیر مختلف لاندا در حل آنالیتیکی و حل با توزیع $\theta^{(1)}$

$w = [0.99412537, 0.99498355, 0.93941647, 0.44836196, 0.18348294, 0.08750755]$

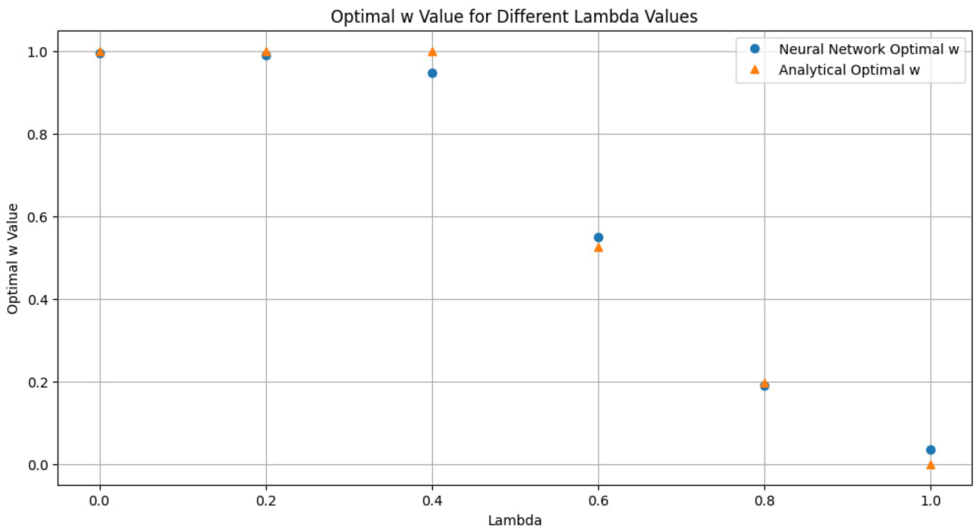


شکل ۵: نمونه تابع $loss$ حین آموزش با توزیع $\theta^{(1)}$

همچنین نتایج زیر حاصل در نظر گرفتن $\theta^{(2)} = 0.5\theta^{(1)} + 0.5\,(\mathcal{U}([0,1]) \circ (\mathrm{Id}, \varphi)^{-1})$ به عنوان اندازه مرجع می‌باشد.

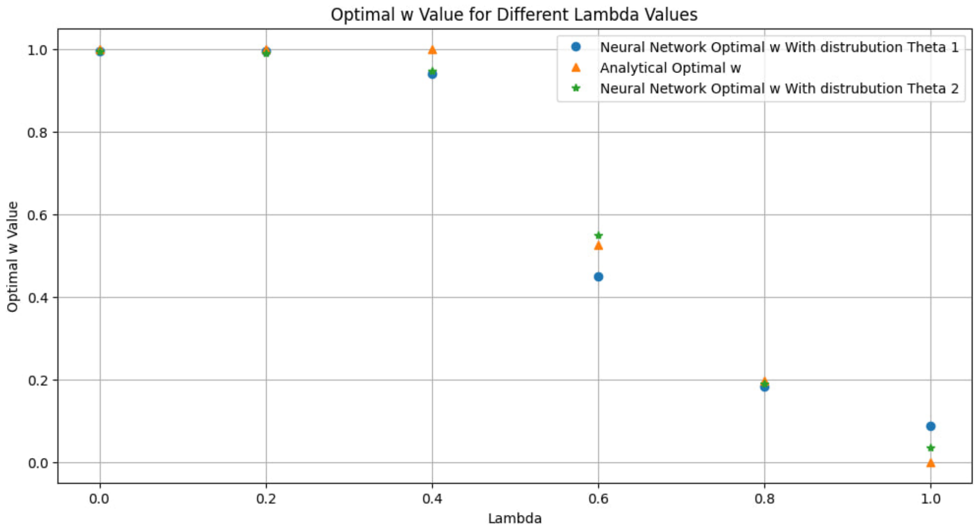


شکل ۶: نمونه تابع $loss$ برای $\theta^{(2)}$

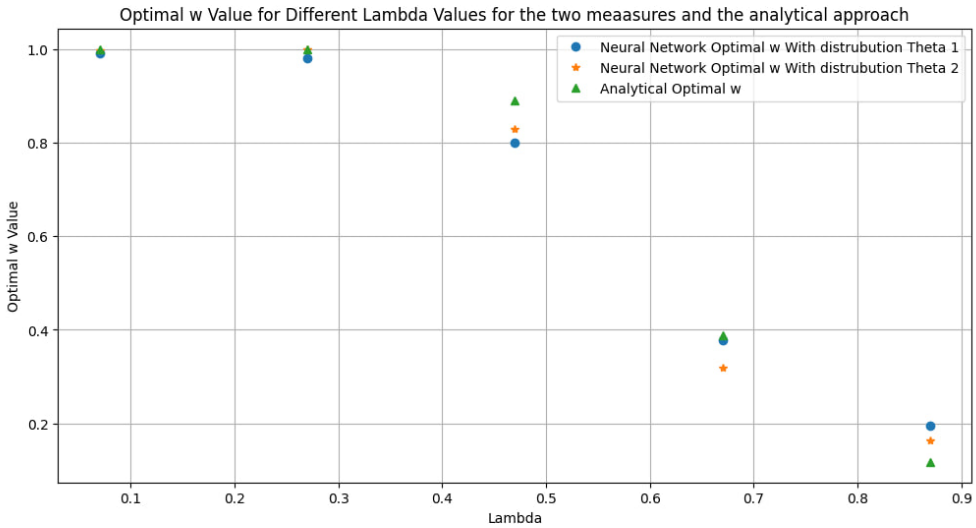


شکل ۷: مقادیر بهینه w برای مقادیر مختلف لاندا در حل آنالیتیکی مسئله و حل با توزیع $\theta^{(2)}$
$w = [0.99536186, 0.99050164, 0.947735, 0.55005884, 0.18909983, 0.034861833]$

در نهایت مقایسه نتایج هر دو توزیع در کنار هم:

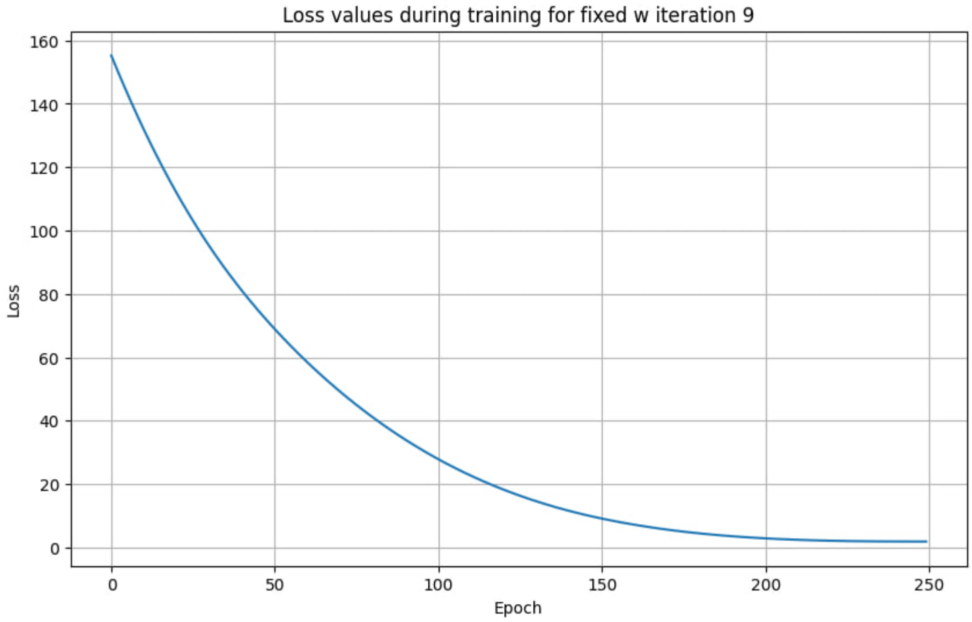


شکل ۸: مقایسه مقدار بهینه w در حل آنالیتیکی مسئله و حل با توزیع‌های $\theta^{(1)}$ و $\theta^{(2)}$

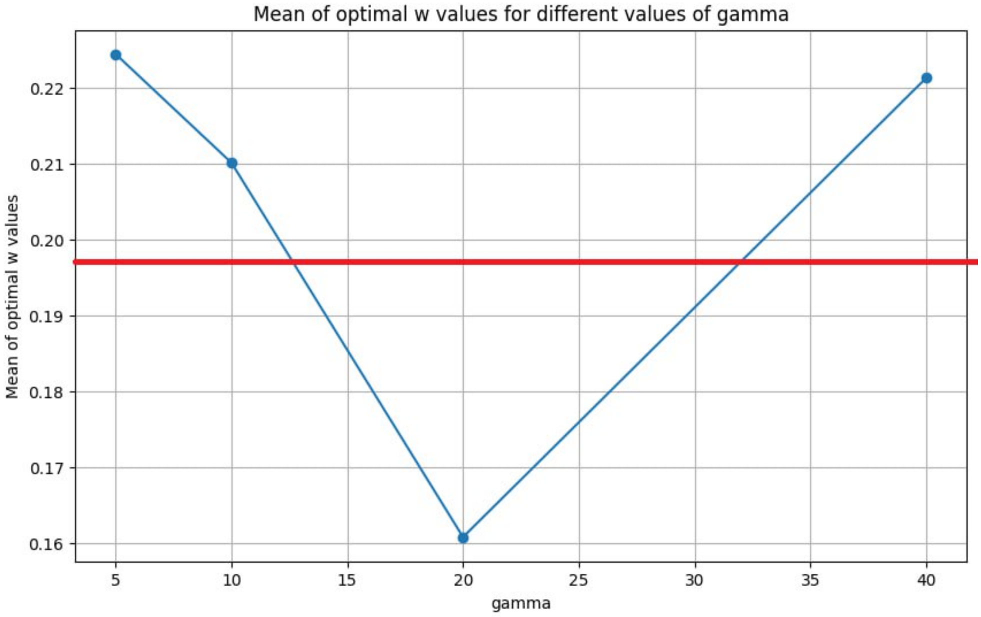


شکل ۹: نتایج شبیه‌سازی برای مقادیر $0.07 \leq \lambda \leq 0.87$ با فواصل 0.2 بدون فروش استقراضی، می توان دید که همچنان نتایج برای $\theta^{(2)}$ بهتر است جز در یک مورد که می‌تواند دلیل آن تعداد کم آزمایش‌ها باشد.

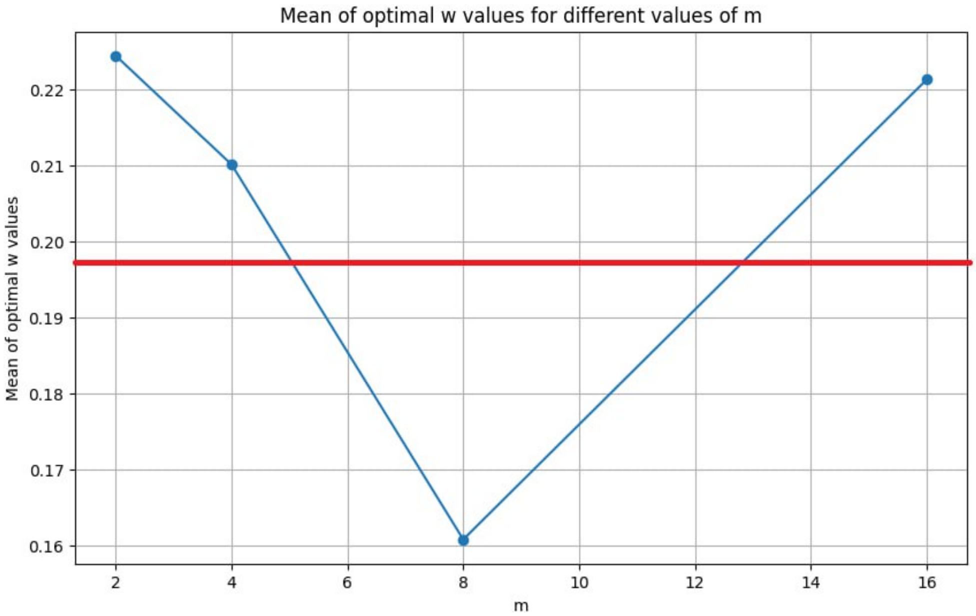
در پایان هم برای سنجش هایپرپارامترهای مختلف چند آزمایش با روش هیوریستیک انجام شدند که نتایج به قرار زیر هستند:



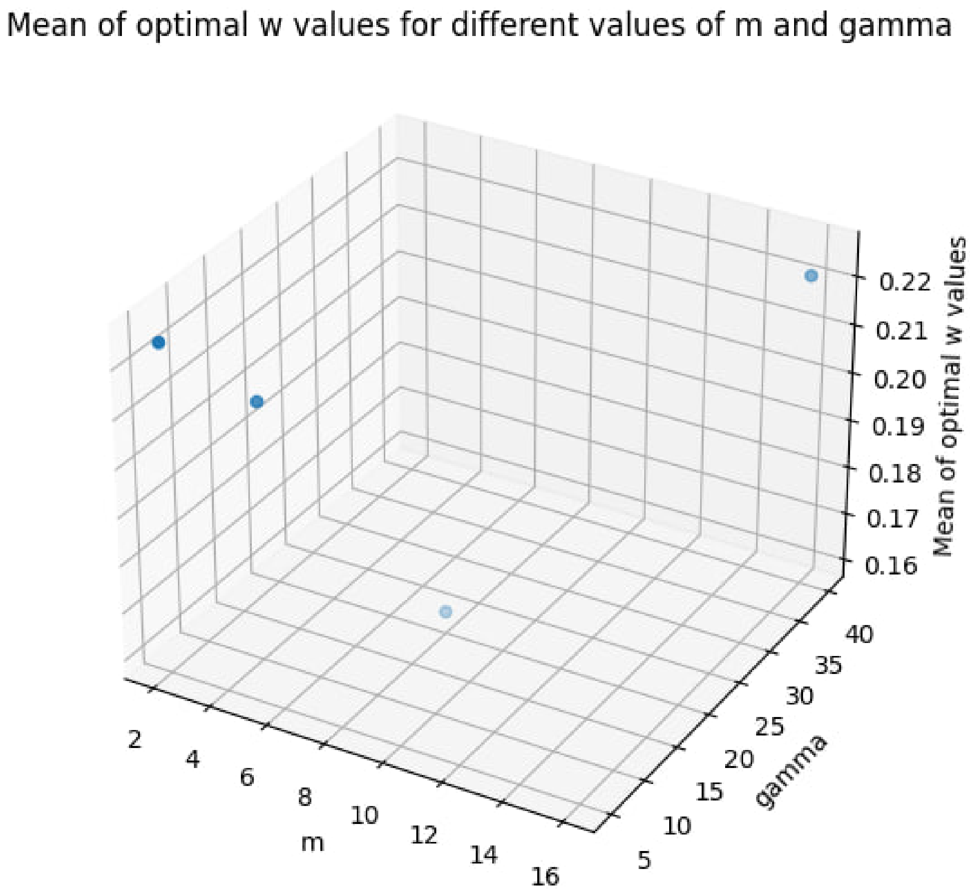
شکل ۱۰: نمونه تابع $loss$ برای $\theta^{(1)}$



شکل ۱۱: میانگین وزن‌های به دست آمده برای ۴ مقدار متفاوت گاما به ازای $\lambda = 0.8$.
خط قرمز، جواب بهین واقعی است.



شکل ۱۲: میانگین وزن‌های به دست آمده برای سبد دوم به ازای مقادیر مختلف m به ازای $\lambda = 0.8$.
خط قرمز، جواب بهین واقعی است.



شکل ۱۳: پراکندگی وزن‌های به دست آمده به ازای مقادیر مختلف گاما و m

به طور کلی دیده می‌شود گرچه روش هیوریستیک خطای بالایی دارد، اما به ازای حتی مقادیر کوچکی از m همچنان جوابی نزدیک به جواب واقعی می‌دهد.
همچنین دیده می‌شود که واقعا توزیع دوم که کمی همبستگی شبیه‌سازی می‌کند، عموما جواب‌های نزدیک‌تری به جواب واقعی می‌دهد.

۴.۸ سوپرهجینگ اختیار معامله خرید اروپایی

در این بخش، خیلی خلاصه به پیاده‌سازی سوپرهجینگ اختیار خرید اروپایی خواهیم پرداخت.
برای تابع مورد نظر داریم:

$$f = max(0, S_T - K)$$

که K قیمت استرایک سررسید است. حال برای سوپرهجینگ داریم:

$$Suph = \inf_h \int hdP$$

پس با جریمه‌سازی به فرم مقاله داریم:

$$Suph\phi^{m}_{\gamma,\theta} = \inf_{h\in H^m}(\int(hd\mu) + \int(\beta_\gamma(f-h)))$$

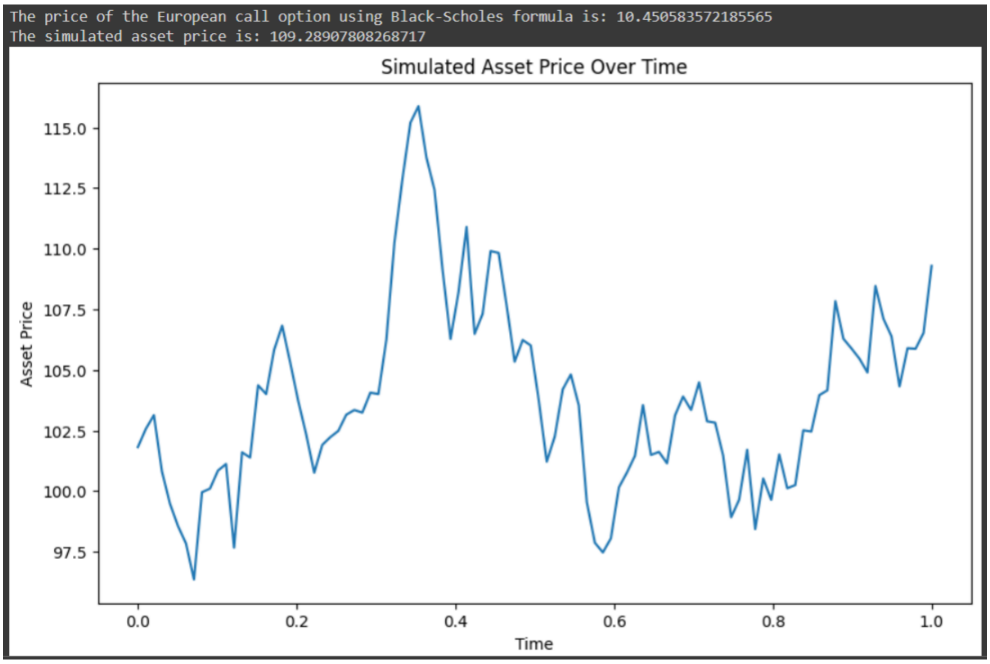
در ادامه در کد، ابتدا قیمت اختیار معامله خرید اروپایی با پارامترهای مشخص زیر با فرمول بلک شولز محاسبه شده:

$$S_0 = 100, K = 100, T = 1, r = 0.05.\sigma = 0.2$$

که قیمت برابر با مقدار زیر شد:

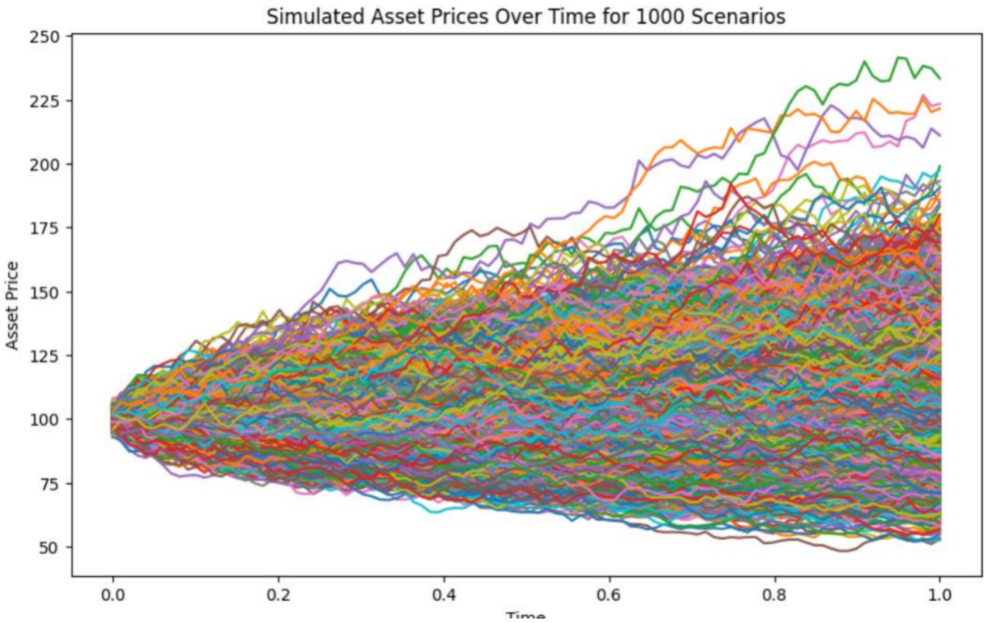
$$price \approx 10.450$$

حال با همین پارامترها و فرض ساده‌سازی میانگین=نرخ بدون ریسک، یک حرکت براونی هندسی شبیه‌سازی شد:



شکل ۱۴: شبیه سازی یک حرکت براونی هندسی با پارامترها

و در ادامه ۵ هزار حرکت براونی شبیه‌سازی شد:



شکل ۱۵: شبیه سازی ۵ هزار حرکت براونی هندسی با پارامترها

و سوپرهجینگ با روش مقاله انجام شد:

```
Epoch: 0 Loss for h 851028.9
Epoch: 100 Loss for h 769594.25
Epoch: 200 Loss for h 692834.7
Epoch: 300 Loss for h 621012.8
Epoch: 400 Loss for h 554118.75
Epoch: 500 Loss for h 491832.56
Epoch: 600 Loss for h 433976.1
Epoch: 700 Loss for h 380378.62
Epoch: 800 Loss for h 330887.94
Epoch: 900 Loss for h 285373.66
Epoch: 1000 Loss for h 243741.47
Epoch: 1100 Loss for h 205892.88
Epoch: 1200 Loss for h 171753.6
Epoch: 1300 Loss for h 141250.39
Epoch: 1400 Loss for h 114313.54
Epoch: 1500 Loss for h 90866.984
Epoch: 1600 Loss for h 70826.07
Epoch: 1700 Loss for h 54091.855
Epoch: 1800 Loss for h 40543.848
Epoch: 1900 Loss for h 30029.021
Epoch: 2000 Loss for h 22344.715
Epoch: 2100 Loss for h 17170.566
Epoch: 2200 Loss for h 13489.464
Epoch: 2300 Loss for h 10672.547
Epoch: 2400 Loss for h 8467.578
Epoch: 2500 Loss for h 6726.745
Epoch: 2600 Loss for h 5351.6006
Epoch: 2700 Loss for h 4267.769
Epoch: 2800 Loss for h 3416.1604
```

```
Epoch: 2800 Loss for h 3416.1604
Epoch: 2900 Loss for h 2750.262
Epoch: 3000 Loss for h 2237.8938
Epoch: 3100 Loss for h 1830.234
Epoch: 3200 Loss for h 1504.8729
Epoch: 3300 Loss for h 1244.354
Epoch: 3400 Loss for h 1034.088
Epoch: 3500 Loss for h 863.77344
Epoch: 3600 Loss for h 725.4606
Epoch: 3700 Loss for h 612.8631
Epoch: 3800 Loss for h 521.07654
Epoch: 3900 Loss for h 446.2518
Epoch: 4000 Loss for h 385.164
Epoch: 4100 Loss for h 335.31314
Epoch: 4200 Loss for h 294.60657
Epoch: 4300 Loss for h 261.37732
Epoch: 4400 Loss for h 234.27553
Epoch: 4500 Loss for h 212.20421
Epoch: 4600 Loss for h 194.2745
Epoch: 4700 Loss for h 179.78159
Epoch: 4800 Loss for h 168.146
Epoch: 4900 Loss for h 158.90019
Epoch: 5000 Loss for h 151.65065
Epoch: 5100 Loss for h 146.07065
Epoch: 5200 Loss for h 141.87808
Epoch: 5300 Loss for h 138.83307
Epoch: 5400 Loss for h 136.72092
Epoch: 5500 Loss for h 135.34929
Epoch: 5600 Loss for h 134.54085
```

شکل ۱۶: لاگ آموزش

دیده می‌شود که تابع فقدان نسبتا هموار می‌شود و قیمت‌ها واقعا بالاتر از قیمت واقعی هستند. با ادامه آموزش و همچنین انجام پوشش از پایین، می‌توان به یک بازه مناسب بازسازی برای قیمت اختیار معامله از روش مقاله رسید. این موضوع از این نظر حائز اهمیت است که می‌توان از این روش برای اختیارهای اگزاتیک‌تر و جالب‌تر استفاده کرد.

۹ نتیجه گیری

با مطالعه و بررسی مقاله و پیاده‌سازی مثال عددی مربوط به بهینه‌سازی سبد سهام در حضور عدم قطعیت، ما نویسندگان گزارش به این نتیجه رسیدیم که نتیجه همگرایی ریاضیاتی گرفته شده در مقاله در عمل هم در جهان واقعی برقرار است و می‌توان از این روش برای تعیین بدترین حالت معیارهای ریسک مختلف در عدم قطعیت، پوشش مالی کمینه از بالا(سوپر هجینگ) یا کران‌دار کردن توزیع توام متغیرهایی تصادفی استفاده کرد که همه کاربردهای زیادی در مالی دارند. نکته بارز و قابل توجه در شبکه عصبی طراحی شده در مقاله، سادگی ساختار آن است. شبکه عصبی مورد اشاره صرفا یک MLP ساده است و می‌توان برای کاربردهای اختصاصی مثل تحلیل سری‌های زمانی، با تعمیم روش مقاله به مدل‌های پیچیده‌تری مثل RNN ها یا استفاده

از چارچوب یادگیری تقویتی، مسائل حتی پیچیده‌تری را هم حل کرد. همچنین ما بر این باور هستیم که آزمایشات بیشتری برای تعیین ادعای همگرایی برای m و γ بدون روش هیوریستیک ما و با استفاده از آموزش کامل شبکه عصبی نیاز است، که متاسفانه ما بودجه محاسباتی این کار را نداشتیم. در صورت تأمین چنین بودجه محاسباتی، به نظر نویسندگان گزارش، میل دادن پارامترهای همگرایی به اعداد بزرگ و سنجش فاصله تا مقادیر آنالیزی واقعی، اطلاعات ارزشمندی درباره صادقانه بودن ادعاهای مقاله در کاربردهای واقعی، اعم از مالی و عمومی خواهد داشت. همچنین در پایان، یک کاربرد برای حل مسائل سوپرهجینگ در اختیار معامله اروپایی هم از روش مقاله بررسی شد این مثال عددی در مقاله اصلی موجود نبود و مساله به تشخیص نگارندگان مساله حل شده است.

مراجع

[1] Cheridito, P., Kupper, M., Tangpi, L.: Duality formulas for robust pricing and hedging in discrete time. SIAM J. Financ. Math. 8(1), 738–765 (2017)

[2] Kantorovich, L.V.: On the translocation of masses. Dokl. Akad. Nauk SSSR 37, 199–201 (1942)

[3] Bartl, D., Cheridito, P., Kupper,M., Tangpi, L.: Duality for increasing convex functionals with countably many marginal constraints. Banach J. Math. Anal. 11(1), 72–89 (2017)

[4] Vallender, S.: Calculation of the Wasserstein distance between probability distributions on the line. Theory Probab. Appl. 18(4), 784–786 (1974)

[5] Villani, C.: Optimal Transport: Old and New, vol. 338. Springer, New York (2008)

[6] Beiglböck, M., Henry-Labordère, P., Penkner, F.: Model-independent bounds for option prices: a mass transport approach. Financ. Stoch. 17(3), 477–501 (2013)

[7] Galichon, A., Henry-Labordere, P., Touzi, N.: A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. Ann. Appl. Probab. 24(1), 312–336 (2014)

[8] Guo, G., Obloj, J.: Computational methods for martingale optimal transport problems. arXiv preprint arXiv:1710.07911 (2017)

[9] Henry-Labordère, P.: Automated option pricing: numerical methods. Int. J. Theor. Appl. Financ. 16(08), 1350042 (2013)

[10] Bernard, C., Rüschendorf, L., Vanduffel, S., Yao, J.: How robust is the value-at-risk of credit risk portfolios? Eur. J. Financ. 23(6), 507–534 (2017)

[11] Embrechts, P., Puccetti, G., Rüschendorf, L.: Model uncertainty and VaR aggregation. J. Bank. Financ. 37(8), 2750–2764 (2013)

[12] Puccetti, G., Rüschendorf, L.: Computation of sharp bounds on the distribution of a function of dependent risks. J. Comput. Appl. Math. 236(7), 1833–1840 (2012)

[13] Bartl, D., Kupper, M., Lux, T., Papapantoleon, A.: Sharpness of improved Fréchet-Hoeffding bounds: an optimal transport approach. arXiv preprint arXiv:1709.00641 (2017)

[14] Lux, T., Papapantoleon, A.: Improved Fréchet-Hoeffding bounds on d-copulas and applications in model-free finance. Ann. Appl. Probab. 27(6), 3633–3671 (2017)

[15] Pflug, G.C., Pohl, M.: A review on ambiguity in stochastic portfolio optimization. Set-Valued Var. Anal. 23(1), 11–25 (2017)

[16] Alfonsi, A., Corbetta, J., Jourdain, B.: Sampling of probability measures in the convex order and approximation of martingale optimal transport problems (2017)