مدل های تولیدی

نیم‌سال اول ۱۴۰۳-۱۴۰۴

مدرس: دکتر سیدصالحی

دانشکده‌ی علوم کامپیوتر

موعد تحویل: ۲۸ دی

تمرین سری پنجم

- مهلت ارسال پاسخ تا ساعت ۵۹ : ۲۳ روز مشخص‌شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمارین (به استثنای هفته‌ی امتحان میانترم) تا سقف پانزده روز وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال‌شده پذیرفته نخواهندبود.
- هم‌کاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.
- **نمرات برای بخش‌های مختلف هر سوال باهم یکسان هستند.**

مسئله‌ی ۱. ۲۵ نمره

در فرآیند کار با مدل‌های انتشار (diffusion models)، به‌جای محاسبه مستقیم تابع هدف معمولاً به دنبال یافتن کران‌هایی برای آن‌ها هستیم. این کران‌ها به ما کمک می‌کنند که رفتار مدل را بهتر درک کرده و به طور غیرمستقیم از آن‌ها برای بهبود عملکرد مدل استفاده کنیم. یکی از کران‌های مهم، تابع درست‌نمایی (Likelihood Function) است که حتی برخی از مدل‌های مولد، تابع هدف بر حسب لگاریتم درست‌نمایی و یا تقریب‌هایی از آن نوشته می‌شود. در مقاله (Maximum Likelihood Training of Score-Based Diffusion Models(Yang Song, 2021 که در پیوست تمرین قرار دارد، نشان داده شده است با در نظر گیری یک وزن خاص در تابع هدف مدل انتشار مبتنی بر امتیاز (Score Based Diffusion Models) می‌توان کرانی از جنس درست‌نمایی برای این مدل یافت. در ادامه با استفاده از قضایای مطرح شده در این مقاله به‌دنبال یافتن این کران برای ارزیابی مدل در یک نقطه هستیم.
فرض کنید $p(x)$ نمایانگر توزیع اولیهٔ داده‌ها باشد که در فرایند انتشار رو به جلو به توزیع نویزی تبدیل می‌شود. توزیع‌های میانی در این فرایند، درزمان $t \in (۰, T]$ با $p_t(x)$ و توزیع انتقال بین x. و x_t با $p_{۰,t}(x'|x)$ نشان داده می‌شوند. در فرایند رو به جلو در زمان t=T به دنبال رسیدن به توزیعی مشخص مانند $\pi(x)$ هستیم یعنی انتظار داریم $p_T(x) \cong \pi(x)$.
فرض کنید فرایند پس‌انتشار از نقطه $\hat{x}_\theta(T) \sim \pi$ آغاز شده و به $\hat{x}_\theta(۰) \sim p_\theta^{SDE}$ ختم می‌شود. از آن‌جایی‌که توزیع p_θ^{SDE} در تعیین عملکرد مدل بسیار پراهمیت است این مقاله به دنبال یافتن کران‌هایی از جنس درست‌نمایی این توزیع برای تابع هدف اصلی است. در قضیه یک این مقاله، با درنظرگیری $\lambda(t) = g(t)^۲$ ثابت می‌شود:

$$D_{KL}(p \| p_\theta^{SDE}) \leqslant \mathcal{J}_{SM}(\theta; g(\cdot)^۲) + D_{KL}(p_T \| \pi). \qquad (۱)$$

حال توضیح دهید:
الف) این کران چگونه به درست‌نمایی توزیع p_θ^{SDE} ربط پیدا می‌کند؟ مفهوم و برداشت خود را از این کران نوشته و بگویید این کران و تابع هزینه $\mathcal{J}_{SM}$ از نظر عملکردی (کیفی و کمی) همسو هستند یا خیر؟ برای اطلاعات بیشتر به (Corollary 1) در مقاله مراجعه نمایید.)
از آن‌جایی‌که در ارزیابی مدل، هدف این است که میزان دقت مدل برای پیش‌بینی برخی داده‌ها محاسبه شود، استفاده از تابع هدف آموزش که معمولا معیاری روی کل دادگان است و توانایی ارزیابی عملکرد مدل برای یک نمونه را ندارد کافی نیست، در این قسمت سعی داریم کرانی برای تابع هدف در یک نقطه محاسبه کنیم،گزاره زیر را در نظر بگیرید:

گزاره: اگر $\mathcal{H}(p)$ نشانگر آنتروپی توزیع p باشد، آن‌گاه

$$\mathcal{H}(p)=\mathcal{H}(p_T(x))-\frac{۱}{۲}\int_{۰}^{T}\mathbb{E}_{p_t(x)}\left[۲\nabla\cdot f(x,t)+g(t)^{۲}\|\nabla_x\log p_t(x)\|_{۲}^{۲}\right]dt. \tag{۲}$$

حال با استفاده از قضیه یک و گزاره بالا به سوالات زیر پاسخ دهید.
ب) اثبات کنید:

$$\begin{aligned}-\mathbb{E}_{p(x)}[\log p_\theta^{\mathrm{SDE}}(x)]\leqslant &-\mathbb{E}_{p_T(x)}[\log\pi(x)]\\&+\frac{۱}{۲}\int_{۰}^{T}\mathbb{E}_{p_{۰t}(x'|x)p(x)}\left[g(t)^{۲}\|s_\theta(x',t)-\nabla_{x'}\log p_{۰t}(x'|x)\|_{۲}^{۲}\right.\\&\left.-g(t)^{۲}\|\nabla_{x'}\log p_{۰t}(x'|x)\|_{۲}^{۲}-۲\nabla\cdot f(x',t)\right]dt.\end{aligned} \tag{۳}$$

با استفاده از قسمت ب میتوان کرانی از جنس تابع درست نمایی برای ارزیابی عملکرد مدل در یک نقطه یافت.
پ) نامساوی زیر را اثبات کنید: (این بخش اثبات قسمتی از تئوری سه در مقاله است.)

$$-\log p_\theta^{\mathrm{SDE}}(x)\leqslant\mathcal{L}_\theta^{\mathrm{DSM}}(x) \tag{۴}$$

که در آن

$$\begin{aligned}\mathcal{L}_\theta^{\mathrm{DSM}}(x):=-\mathbb{E}_{p_{۰T}(x'|x)}[\log\pi(x')]&+\frac{۱}{۲}\int_{۰}^{T}\mathbb{E}_{p_{۰t}(x'|x)}\left[g(t)^{۲}\|s_\theta(x',t)-\nabla_{x'}\log p_{۰t}(x'|x)\|_{۲}^{۲}\right]dt\\&-\frac{۱}{۲}\int_{۰}^{T}\mathbb{E}_{p_{۰t}(x'|x)}\left[g(t)^{۲}\|\nabla_{x'}\log p_{۰t}(x'|x)\|_{۲}^{۲}+۲\nabla_{x'}\cdot f(x',t)\right]dt.\end{aligned} \tag{۵}$$

راهنمایی: در فرایند اثبات‌ها از رابطه $\int p(x')p_{۰t}(x'|x)dx'=p_t(x)$ کمک بگیرید، همچنین راهنمایی و توضیحات بیشتر برای تمامی بخش‌های سوال در مقاله موجود هستند.

مسئله‌ی ۲. شبیه‌سازی دینامیک لانجوین(۲۵ نمره)

می‌دانیم مدل های Neural ODE و Continuous Normalizing Flow بر اساس دینامیک زیر تعریف می‌شوند:

$$dX=f(X,t)dt$$

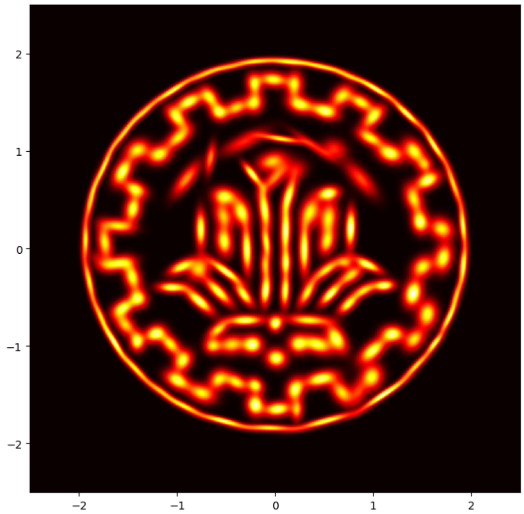
همانطور که میدانید، دینامیک فوق را می‌توان به عنوان یک تبدیل پیوسته در زمان روی نمونه ها تفسیر کرد و همچنین میدانیم که تابع چگالی احتمال در هر لحظه را میتوان به کمک قضیه‌ی تبدیل متغیر لحظه‌ای (مقاله Neural ODE) محاسبه کرد. البته اگر تبدیلی که روی نمونه ها انجام می‌شود یک تبدیل تصادفی باشد، دیگر نمی‌توانیم از قضیه‌ی تغییر متغیر لحظه‌ای استفاده کنیم و به قضایای کلی‌تری برای بدست آوردن تابع چگالی نیاز داریم. به عنوان یک تلاش برای تعمیم، دینامیک زیر را در نظر بگیرید که مولفه‌ی اول آن مشابه قبلی و مولفه‌ی دیگر آن یک فرایند انتشار است.

$$dX=f(X,t)dt+g(X,t)dW \tag{۶}$$

دینامیک لانجوین یک حالت خاص از دینامیک فوق است که برای نمونه‌برداری از یک چگالی احتمال مثل π به صورت زیر تعریف می‌شود.

$$dX=\nabla\log\pi(X)dt+\sqrt{۲}dW$$

در ادامه، از تابع چگالی دومتغیره π که در پایین رسم شده است، برای تعدادی آزمایش استفاده میکنیم:



این تابع چگالی درواقع یک Gaussian Mixture Model مرکب از ۱۵۰ توزیع نرمال دومتغیره است:

$$\pi(x) = \sum_{k}^{۱۵۰} w_k \pi_k(x; \mu_k, \Sigma_k)$$

پارامتر های این مدل شامل وزن مولفه ها w_k، بردار های میانگین μ_k و ماتریس های کوواریانس Σ_k در فایل همراه تمرین موجود است. در نوت بوک زیر هم نحوه‌ی خواندن پارامتر ها نشان داده شده است:

https://colab.research.google.com/drive/1F-Xg-30uoSWiO8PXp_qSgbDU9cIqL3EQ?usp=sharing

قصد داریم از توزیع π نمونه‌برداری کنیم اما به‌جای استفاده از روش سرراست نمونه‌برداری از مدل های GMM میخواهیم از دینامیک لانجوین استفاده کنیم.

با این انگیزه، شبیه‌سازی دینامیک لانجوین را با شرایط خواسته شده هر قسمت انجام دهید. برای هر قسمت یک scatter plot از نمونه ها در قدم پایانی رسم کنید (چون تعداد نمونه ها زیاد است اندازه نقاط را به طور مناسب انتخاب کنید).

الف ابتدا تعداد $۱۰^۴$ نمونه از یک توزیع نرمال دومتغیره $\mathcal{N}(۰, I)$ تولید کنید. این نمونه ها را با $X_۰$ نشان میدهیم. برای این قسمت فرایند انتشار در دینامیک لانجوین را نادیده گرفته و دینامیک گسسته زیر را روی هر یک از نمونه ها اجرا کنید:

$$X_{t+۱} = X_t + \tau \nabla \log \pi(X_t) \tag{۷}$$

طول قدم τ را مقدار ثابت $۱۰^{-۳}$ فرض کرده و برای ۵۰۰ قدم دینامیک را شبیه سازی کنید.

ب اینبار مولفه‌ی انتشار در دینامیک لانجوین را هم اضافه کرده و با شرایط مشابه قسمت قبل دینامیک زیر را اجرا کنید:

$$X_{t+۱} = X_t + \tau \nabla \log \pi(X_t) + \sqrt{۲\tau}\zeta_t \tag{۸}$$

در اینجا ζ_t برداری تصادفی از یک توزیع $\mathcal{N}(۰, I)$ دومتغیره است. توزیع اولیه‌ی نمونه ها را مشابه قسمت قبل فرض کنید و نمونه ها در آخرین قدم را رسم کنید. این نتیجه را با نتیجه‌ی قسمت قبل مقایسه کنید.

پ شبیه‌سازی قسمت قبل را با طول قدم $\tau = ۱۰^{-۴}$ تکرار کنید. تغییر طول قدم τ چه تاثیری بر نتیجه دارد؟

ت شبیه‌سازی قسمت **ب** را تکرار اما اینبار برای شبیه‌سازی نویز ζ_t از یک توزیع نرمال دومتغیره با ماتریس کوواریانس

$$I = \begin{bmatrix} ۲/۰ & ۰/۰ \\ ۰/۰ & ۰/۱ \end{bmatrix}$$

استفاده کنید.

نمودار های خواسته شده و تفسیر مشاهدات خود را ارسال کنید. کد شبیه‌سازی را هم به گزارش خود پیوست کنید.

نکته برای این تمرین، میتوانید از زبان دلخواه خود استفاده کنید. با توجه به تعداد زیاد مولفه های مدل GMM مراقب خطاهای عددی باشید!

مسئله‌ی ۳. مدل‌های انتشار برای ترمیم داده سری زمانی (۱۵ نمره)

در این سوال، قصد داریم نحوه استفاده از مدل‌های انتشار برای کاربردهای به‌جز تولید داده، بررسی کنیم. یکی از مهم‌ترین مسائل مهم که در آن از خاصیت مولد مدل‌های انتشار میتوان استفاده کرد، برای ترمیم داده‌های ناقص سری زمانی می‌باشد. در این بخش، مقاله CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation که در تمرین به پیوست آمده است را مورد مطالعه قرار می‌دهیم. باتوجه به آن، به سوالات زیر پاسخ دهید:

الف) ایده کلی مطرح‌شده در این مقاله و روش برای آموزش مدل را به صورت دقیق توضیح دهید؟

ب) استراتژی‌های مطرح‌شده برای ایجاد داده آموزشی برای مدل (بخش ۴.۳) را به‌اختصار توضیح دهید و آنها را با یکدیگر مقایسه کنید؟ به‌نظر شما کدام یک از آنها بهبود بیشتری بر مدل نهایی می‌توانند داشته باشند؟

ج) متریک‌های ارزیابی برای قدرت ترمیم داده سری زمانی شرح دهید.

د) در مخزن گیتهاب پروژه این مقاله، وزن مدل از پیش آموزش دیده قرار گرفته است. با استفاده از این مدل و کدهای در این مخزن، این مدل را بر یک مجموعه داده سری زمانی به‌غیر از مجموعه داده‌های استفاده‌شده در مقاله بازتنظیم (finetune) و دقت ترمیم آن را با استفاده از همان متریک‌ها گزارش کنید. برای این بخش، پاسخ خود در یک نوتبوک به‌همراه توضیحات مختصر هربخش آن تحویل دهید. برای یافتن مجموعه داده، شما آزاد به انتخاب مجموعه داده از اینترنت و سایت kaggle هستید.

مسئله‌ی ۴. بخش عملی (۳۵ نمره)

.برای بخش عملی، نوتبوک پیوست‌شده را تکمیل کنید

موفق باشید :)