مدل‌های مولد

نیم‌سال اول ۱۴۰۳-۱۴۰۴

مدرس: دکتر سیدصالحی

دانشکده علوم ریاضی

موعد تحویل: ۵ دی

تمرین سری چهارم

- مهلت ارسال پاسخ تا ساعت ۵۹ : ۲۳ روز مشخص‌شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمارین تا سقف ۱۵ روز وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال‌شده با سیاست با تاخیر خواهند بود.
- 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.

مسئله‌ی ۱. تبدیل متغیر (۲۰ نمره)

اساس مدل Normalizing Flow تبدیل یک توزیع ساده به توزیعی پیچیده، به کمک دنباله‌ای از تبدیلات وارون‌پذیر است. مدل هایی نظیر Glow یا RealNVP از تبدیلات گسسته در زمان و مدل های Continuous Normalizing Flow از تبدیلات پیوسته در زمان بهره می‌برند. در این تمرین مثالی از یک تبدیل گسسته و یک تبدیل پیوسته در زمان را بررسی می‌کنیم.

الف (تبدیلات گسسته در زمان) بردار تصادفی Z را با توزیع نرمال دو متغیره در نظر بگیرید.

$$Z \sim \mathcal{N}(\begin{bmatrix} 0/0 \\ 0/0 \end{bmatrix}, I)$$

بردار تصادفی Y از اعمال یک تبدیل تک‌مرحله‌ای روی Z بدست آمده است. تابع چگالی این بردار تصادفی را بدست آورید.

$$Y = \begin{bmatrix} \frac{3}{5} Z_1(t) + \frac{2}{5} Z_2(t) \\ \frac{1}{5} Z_1(t) + \frac{4}{5} Z_2(t) \end{bmatrix}$$

ب (تبدیلات پیوسته در زمان) فرض کنید اینبار متغیر Z را تحت تبدیلی پیوسته در زمان تبدیل میکنیم که از دینامیک زیر تبعیت می‌کند:

$$\frac{d}{dt} \begin{bmatrix} Z_1(t) \\ Z_2(t) \end{bmatrix} = \begin{bmatrix} \tanh(Z_1(t)^3) \\ \tanh(Z_2(t)^3) \end{bmatrix}$$

$$Z(t=0) \sim \mathcal{N}(\begin{bmatrix} 0/0 \\ 0/0 \end{bmatrix}, I)$$

به عبارت دیگر میدانیم در لحظه‌ی $t = 0/0$ یک متغیر نرمال داریم و این متغیر را تحت دینامیک فوق تبدیل می‌کنیم. پس این متغیر در هر لحظه یک تابع چگالی متفاوت دارد.

ب ۱ تعدادی نمونه از توزیع اولیه تولید کرده و با حل عددی دینامیک فوق مختصات نقاط تبدیل شده را بدست آورید. با رسم یک هیستوگرام دوبعدی از این نقاط، تلاش کنید تابع چگالی توزیع در لحظه‌ی $t = 0/5$ را نمایش دهید.

ب ۲ به کمک قضیه‌ی تبدیل متغیر لحظه‌ای[۱] ، دینامیک لگاریتم تابع چگالی این متغیر را به فرم یک معادله‌ی دیفرانسیل بدست آورید:

$$\frac{\partial \log p(Z(t),t)}{\partial t} = ?$$

ب ۳ به کمک نتیجه‌ی بدست آمده در قسمت (**ب ۲**) مقدار تابع چگالی را در لحظات

$t \in \{۰/۰, ۰/۱, ۰/۲, ۰/۳, ۰/۴, ۰/۵\}$ برای نقاط $Z \in [-۴, ۴]^۲$ به صورت تقریبی محاسبه کرده و نمایش دهید. روش محاسبه را توضیح دهید. برای اطمینان از صحت محاسبات، مطمئن شوید نتیجه‌ی این قسمت با نتیجه حاصل از قسمت (**ب ۱**) مغایرت زیادی نداشته باشد. در هر لحظه باید یک تابع دومتغیر را نمایش دهید. برای اینکار از یک heatmap با رزولوشن کافی استفاده کنید.

ب ۴ تبدیلی که در قسمت های قبل استفاده کردیم، ممکن است دارای تعدادی پارامتر برای یادگیری مثل θ باشد. به صورت شهودی این مدل مانند یک شبکه‌ی عصبی با عمق پیوسته است و θ مثل یک وزن قابل یادگیری برای این شبکه است. مثلا تبدیل زیر را در نظر بگیرید:

$$\frac{d}{dt}\begin{bmatrix} Z_۱(t) \\ Z_۲(t) \end{bmatrix} = \begin{bmatrix} \tanh(\theta Z_۱(t)^۳) \\ \tanh(\theta Z_۲(t)^۳) \end{bmatrix}$$

چنین مدلی در نهایت یک تابع چگالی تعریف می‌کند که در قسمت های قبل مقدار آن را محاسبه کردیم. طبیعتا رفتار چگالی حاصل اکنون توسط پارامتر های مدل یعنی θ تعیین می‌شود. برای طراحی یک مدل Generative باید بتوانیم این پارامتر را به نحوی یاد بگیریم که مقدار تابع چگالی در نقاطی که در دیتاست داریم زیاد باشد. به عبارت دیگر قصد داریم θ را طوری پیدا کنیم که likelihood مدل تعریف شده روی داده ها افزایش یابد.

$$\theta^* = \operatorname{argmax}_\theta \mathbb{E}_{x\sim \text{Data}}[\log p(x;\theta)]$$

برای حل این مسئله‌ی بهینه‌سازی باید بتوانیم از متغیر $Z(t)$ در لحظه‌ی نهایی تبدیل نسبت به پارامتر θ مشتق بگیریم. در حقیقت باید به نحوی از حل کننده‌ی یک معادله‌ی دیفرانسیل مشتق بگیریم. با مطالعه‌ی روش الحاقی در مقاله‌ی Neural ODE روند محاسبه‌ی این مشتق را بدون پرداختن به جزئیات توصیف کنید.

ب ۵ اغلب روش های حل عددی معادلات دیفرانسیل مثل روش اویلر از عملیات های مشتق پذیر تشکیل شده اند. بر این اساس می‌شود به جای استفاده از روش الحاقی، مستقیما از چنین روش هایی مشتق گرفت! به صورت مختصر توضیح دهید چرا این ایده‌ی خوبی نیست.

نکته برای این تمرین می‌توانید از زبان و پکیج های دلخواه خود استفاده کنید. مثلا برای حل عددی معادلات خواسته شده در زبان Python می‌توانید از پکیج scipy یا torchdiffeq و در زبان Julia می‌توانید از پکیج DifferentialEquations استفاده کنید. برای قسمت های **ب ۱** و **ب ۳** نیازی به ارسال کد نیست، صرفا نمودار ها و توضیحات خواسته شده را ارسال کنید.

مسئله‌ی ۲. مدل انرژی (۱۰ نمره)

فرض کنید قصد داریم یک مدل انرژی مثل $p_\phi(x)$ را آموزش دهیم.

$$p_\phi(x) \propto e^{-E_\phi(x)}$$

در اینجا E یک شبکه‌ی عصبی و ϕ پارامتر های آن است. ثابت نرمالسازی این مدل را با Z_ϕ نشان می‌دهیم:

$$Z_\phi = \int e^{-E_\phi(x)} dx$$

[۱] به قضیه شماره ۱ در مقاله‌ی Neural ODE رجوع کنید

اگر قصد آموزش این مدل به کمک Maximum Likelihood Estimation را داشته باشیم، به گرادیان زیر احتیاج خواهیم داشت:

$$\nabla_\phi \log p_\phi(x) = -\nabla_\phi E_\phi(x) - \nabla_\phi \log Z_\phi$$

الف ثابت کنید

$$\nabla_\phi \log Z_\phi = \mathbb{E}_{x\sim p_\phi(x)}[-\nabla_\phi E_\phi(x)]$$

ب امیدریاضی مورد نیاز که در قسمت قبل بدست آمد را می‌توان با هر روش استاندارد مونت‌کارلو تقریب زد. به طور مختصر مشکل استفاده از این روش برای آموزش مدل انرژی را توضیح دهید .

مسئله‌ی ۳. سوال عملی اول (۳۵ نمره)

در این سوال ما یک مدل مبتنی بر انرژی ساده را بر روی مجموعه داده ارقام MNIST آموزش می‌دهیم و تصاویر را با استفاده از این مدل EBM تجسم می‌کنیم. اگر این کار برای شما دشوار بود، می‌توانید مجموعه داده را کوچکتر کنید. احتمالاً برای این مسئله به GPU نیاز خواهید داشت.

توضیحات پیاده‌سازی مرحله‌به‌مرحله در فایل نوت‌بوک سوال آورده شده است.

مسئله‌ی ۴. سوال عملی دوم (۳۵ نمره)

در این مسئله، ما یک مدل جریان‌های نرمال‌سازی ساده را آموزش می‌دهیم تا داده‌هایی مشابه مجموعه داده ماه‌ها از sklearn تولید کنیم. شما برای این مسئله نیازی به GPU ندارید.

توضیحات پیاده‌سازی مرحله‌به‌مرحله در فایل نوت‌بوک سوال آورده شده و نمونه ای از خروجی تکامل مدل در فایل تمرین ضمیمه‌شده است.

موفق باشید :)