دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده برق و کامپیوتر

Deep Generative Models

تمرین شماره ۲- قسمت دوم

طراحان:

علیرضا غفوری، پرهام زیلوچیان مقدم

فرشاد سنگری، علی هدایت نیا

زمان تحویل: **۱۴۰۲/۰۹/۲۰**

آذر ۱۴۰۲

فهرست

پرسش ۱ – Generative Adversarial Networks (GANs)

مقدمه

در این پرسش از تمرین قصد داریم تا با نحوه کارکرد یک مدل ساده از شبکه‌های GAN آشنا شویم. بدین منظور از شما می‌خواهیم تا یک شبکه GAN ساده که به شما معرفی می‌شود را طراحی کرده و سپس آن را روی یک دیتاست سبک آموزش دهید و نتایج بدست آمده را ارزیابی نمایید. در نهایت از شما خواسته می‌شود تا بر روی این شبکه اصلاحاتی را اعمال کنید تا عملکرد آن بهبود پیدا کند.

آموزش مدل GAN بر روی دیتاست MNIST

برای برطرف شدن مشکل ناپدید شدن گرادیان‌ها[1] در طول آموزش یک تابع خطای اشباع ناپذیر[2] پیشنهاد می‌شود که به صورت زیر تعریف می‌گردد:

$$L^{\text{ns}}_{\text{generator}}(\theta;\phi) = -\mathbb{E}_{\mathbf{z}\sim\mathcal{N}(0,I)}\Big[\log D_\phi(G_\theta(\mathbf{z}))\Big]$$

همچنین برای تخمین mini-batch، از تخمین مونت کارلو از هدف یادگیری[3] به صورت زیر بهره می‌بریم:

$$L_{\text{discriminator}}(\phi;\theta) \approx -\frac{1}{m}\sum_{i=1}^{m}\log D_\phi(\mathbf{x}^{(i)}) - \frac{1}{m}\sum_{i=1}^{m}\log\Big(1 - D_\phi(G_\theta(\mathbf{z}^{(i)}))\Big)$$

$$L^{\text{ns}}_{\text{generator}}(\phi;\theta) \approx -\frac{1}{m}\sum_{i=1}^{m}\log D_\phi(G_\theta(\mathbf{z}^{(i)}))$$

for batch-size m, and batches of *real-data* $\mathbf{x}^{(i)} \sim p_{\text{data}}(\mathbf{x})$ and *fake-data* $\mathbf{z}^{(i)} \sim \mathcal{N}(0, I)$

[1] Vanishing gradients
[2] non-saturating loss
[3] Learning objective

برای پیاده سازی generator می‌توانید از معماری زیر استفاده کنید:

	Sequential Blocks	In_Channels	Out_Channels	Batch Norm., Stride,Padding
1	Linear	64	512	BN
	ReLU	-	-	-
2	Linear	512	?!	BN
	ReLU	-	-	-
3	PixelShuffle	-	-	-
4	Conv 3*3	16	32	BN , p=1
	ReLU	-	-	-
5	PixelShuffle	-	-	-
6	Conv 3*3	8	?!	p=1

برای پیاده سازی discriminator نیز معماری زیر توصیه می‌شود:

	Sequential Blocks	In_Channels	Out_Channels	Stride & Padding
1	Conv 4*4	1	32	S=2 , p=1
	ReLU	_	_	_
2	Conv 4*4	32	64	S=2 , p=1
	ReLU	_	_	_
3	Linear	?!	512	_
4	Linear	512	1	_

در جداول فوق قسمتی که علامت سوال قرار داده شده را به سادگی می‌توانید به دست آورده و تکمیل کنید.

سوالات

با دنبال کردن جزئیات مطرح شده، به سوالات زیر پاسخ دهید:

A. یکی از عملگرهایی که در ساختار پیشنهادی generator به شما معرفی شد، PixelShuffle بود. در مورد این مفهوم جالب که به تازگی معرفی شده است، توضیح دهید. برای این منظور باید نحوه عملکرد و تاثیر آن را توضیح دهید. همچنین می‌توانید به خاستگاه این مفهوم اشاره کنید (این که این مفهوم اولین بار به چه علت و برای چه کاربردی معرفی شد). به نظر شما در شبکه کنونی معرفی شده در این سوال، این عملگر چه تاثیری می‌گذارد؟ **(۳ نمره)**

B. کلاس های مربوط به Generator و Discriminator را در فایلی که در اختیار شما گذاشته‌ایم، با توجه به ساختارهای معرفی شده تکمیل کنید. **(۶ نمره)**

C. تابع خطای مورد نیاز برای آموزش مدل را با توجه به فرمول معرفی شده پیاده سازی کنید. **(۴ نمره)**

D. در تابع آموزش قسمت‌های مشخص شده را تکمیل کنید تا مدل GAN طراحی شده شما بتواند آموزش ببیند. **(۳ نمره)**

E. اکنون به ارزیابی مدل آموزش دیده خود بپردازید. برای این منظور:

a. پس از آموزش مدل به میزانی که کیفیت مطلوب حاصل شود، ابتدا نمودار تغییرات loss در طول فرآیند آموزش را برای هردو قسمت generator و discriminator ترسیم کنید. **(۴ نمره)**

b. سپس خروجی نهایی مدل را به ازای عکس های تصادفی در قالب یک تصویر ۱۰ در۱۰ نمایش دهید. ساختار تصویر خروجی شما در گزارش باید مشابه شکل زیر باشد:

شکل 1: نمونه خروجی تولیدی به وسیله شبکه GAN بر روی دیتاست MNIST

با استفاده از کدی که در اختیار شما قرار گرفته است، خروجی مدل را به ازای ۳ اپاک مختلف ابتدایی، میانی و نهایی به شکلی که برای شما توضیح داده شد، گزارش کنید. **(۳ نمره)**

c. یکی از پرکاربردترین معیارهای ارزیابی برای مدل های تولیدی، معیار FID score می‌باشد. ابتدا مختصرا این معیار را معرفی کرده و سپس عملکرد مدل خود را با محاسبه این معیار مورد بررسی قرار دهید. تحلیل شما از عملکرد مدل چیست؟ **(۴ نمره)**

F. مدلی که پیاده سازی کردید، نسخه ابتدایی GAN بود. این مدل دارای نقاط ضعفی بود که با گذر زمان، محققان سعی کردند با معرفی نسخه های جدیدی از GAN، این مشکلات را برطرف کرده و عملکرد مدل اولیه را بهبود دهند. در این بخش از شما می‌خواهیم تعدادی از این نسخه‌های بهبود یافته را مختصرا بررسی کنید و توضیح دهید هرکدام از این نسخه ها چه مشکلاتی از نسخه اولیه را برطرف کرده و چگونه این کار را انجام دادند؟ خروجی شما می‌تواند در قالب یک جدول به صورت زیر باشد (اختیاری) اما پاسخ شما باید حتما جاهای خالی جدول را شامل شود (هرکدام از این بخش‌ها را در حداکثر یک پاراگراف توضیح دهید.) **(۸ نمره)**

پی نوشت: لینک مقالات مربوطه در جدول زیر قرار داده شده است.

جدول ۱: تعدادی از شبکه‌های GAN معروف به منظور بررسی

نام مدل GAN	چه مشکلی را برطرف می‌کند؟	چگونه این مشکل را مرتفع می‌کند؟
Wasserstein GAN (WGAN)		
PG-GAN		
Big GAN		
Style GAN		

G. **(امتیازی)** یکی از نسخه‌هایی که بررسی کردید، WGAN بود. حال: **(۵ نمره)**

a. با توجه به تابع خطای معرفی شده در مقاله مربوط به این نسخه، کد خود را اصلاح کرده، مدل را با تنظیمات جدید آموزش داده و قسمت b , c این پرسش را تکرار کنید.

b. به نظر شما برای مدل WGANی که پیاده سازی کردید، همچنان چه مشکلاتی پابرجاست؟ آنها را توضیح دهید و اگر راه حلی برای مرتفع کردن این مشکلات در نظر دارید، بیان کنید.

سوال ۲– Diffusion Model

مدل‌های Diffusion-based یک زنجیره مارکوف به طول T تشکیل می‌دهند که در این زنجیره مارکوف سعی می‌کنند، با افزودن نویز گوسی، داده را در T مرحله به نویز کاملا گوسی تبدیل کنند.

در مدل‌های Diffusion-based دو مسیر Forward و Backward داریم. در مسیر Forward به داده‌ها، نویز اضافه می‌کنیم تا در نهایت به نویز کاملا گوسی تبدیل شوند و در مسیر Backward سعی می‌کنیم که با تخمین مقدار نویز افزوده شده در هر گام، این نویز را حذف کنیم و به تصویر اولیه بازگردیم.

نکته حائز اهمیت در مدل‌های Diffusion-based این است که این مدل‌ها ما را قادر می‌سازند که از یک نویز کاملا گوسی، یک داده جدید را تولید کنیم.

مسیر Forward

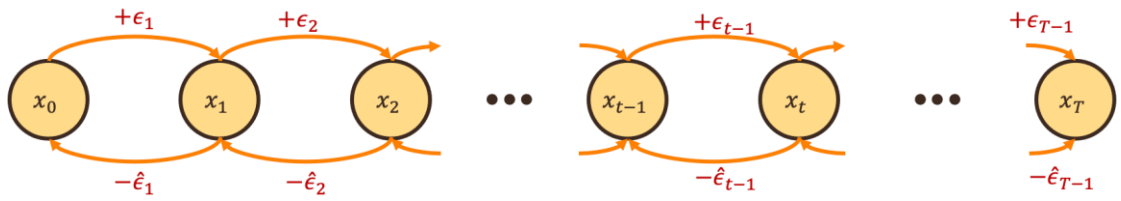
در مسیر forward در یک زنجیره مارکوف طی T-گام با طول $\{\beta_t \in (0,1)\}_{t=1}^{T}$، به تصویر نویز می‌افزاییم و نمونه‌های نویزی $x_1, x_2, \dots, x_T$ را بدست می‌آوریم. بنابراین، خواهیم داشت:

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \qquad (2.1)$$

$$q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}) \qquad (2.2)$$

مسیر Backward

اگر بتوانیم از توزیع $q(x_{t-1}|x_t)$ بتوانیم نمونه بگیریم، خواهیم توانست در یک فرآیند تکرار شونده از یک نویز گوسی $x_T \sim \mathcal{N}(0, I)$، یک نمونه واقعی بسازیم.



شکل ۲- زنجیره مارکوف مدل‌های Diffusion-based

سوالات تئوری

(سوال ۱) سه معیار مهم برای مقایسه مدل های مولد، عبارت اند از: کیفیت، تنوع و سرعت نمونه برداری. شبکه های مولدی که تا به حال آموخته‌اید (شبکه های VAE، GAN، Normalizing Flow و Diffusion-based) را به صورت **مختصر** بر اساس این معیار ها مقایسه کنید. **(۳ نمره)**

(سوال ۲) طبق مقاله DDPM (1)،همانطور که دیدیم در مسیر رو به جلو نیازی به اضافه کردن نویز به صورت تکرار شونده ($q(x_t|x_{t-1})$) نیست. در واقع می توان، با یک مرحله ($q(x_t|x_0)$) به هر کدام از بازنمایی های میانی رسید. با استفاده از خاصیت نویز گوسی، این فرایند را اثبات کنید. (راهنمایی: اگر X_1, X_2 مستقل باشند و $\begin{cases} X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \\ X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2) \end{cases}$ ، آنگاه: $X_3 = X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$) **(۵ نمره)**

(سوال ۳) یک فرض مهم در مسیر رو به عقب این است که توزیع $q(x_{t-1}|x_t)$ را گوسی فرض کنیم. در چه صورتی این فرض صادق است؟ با وجود اینکه این توزیع گوسی می باشد، مسیر رو به عقب را به کمک توزیع $q(x_{t-1}|x_t, x_0)$ تخمین می زنند. دلیل این امر را بررسی کنید. **(۵ نمره)**

در صورتی که توزیع تخمین زده شده را با $p_\theta(x_{t-1}|x_t)$ نمایش دهیم، خواهیم داشت:

$$p_\theta(x_{0:T}) = p_\theta(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t) \qquad (2.3)$$

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}\big(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)\big) \qquad (2.4)$$

(سوال ۴) قابل ذکر است که پارامترهای توزیع گوسی $q(x_{t-1}|x_t, x_0)$، به صورت زیر قابل محاسبه است:

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}\big(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}\, I\big) \qquad (2.5)$$

$$\tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \cdot \beta_t \qquad (2.6)$$

$$\tilde{\mu}_t(x_t, x_0) = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_t \right) \qquad (2.7)$$

از آنجا که setup مدل‌های Diffusion-base شبیه به VAE است، همانند VAE می‌توان با استفاده از روش تخمین توزیع Variational Inference، بجای Log-Likelihood، مقدار ELBO را بیشینه می‌کنیم و در این صورت تابع خطا به صورت زیر خواهد بود:

$$L_{VLB} = L_T + \sum_{t=1}^{T-1} L_t + L_0 \quad where \quad \begin{cases} L_T = D_{KL}\big(q(x_T|x_0) \parallel p_\theta(x_T)\big) \\ L_t = D_{KL}\big(q(x_{t-1}|x_t, x_0) \parallel p_\theta(x_{t-1}|x_t)\big) \quad 1 \le t < T \\ L_0 = -\log p_\theta(x_0|x_1) \end{cases} \qquad (2.8)$$

مفهوم هر کدام از ترم‌های تابع هزینه (2.8) را به صورت مختصر توضیح دهید. **(۶ نمره)**

(سوال ۵) همانطور که در رابطه (2.8) می بینیم، تابع هزینه اولیه دارای سه ترم می باشد. مقاله DDPM کدام ترم(ها) را در برای فرایند بهینه سازی در نظر نگرفته است؟ چرا؟ **(۳ نمره)**

(سوال ۶) در صورتی که به جای توزیع گوسی، یک توزیع پیچیده برای مسیر رو به عقب در نظر می گرفتیم، چه تاثیری در محاسبه ترم(های) تابع هزینه دارد؟ **(۲ نمره)**

(سوال ۷) تابع هزینه نهایی ارائه شده در این مقاله به صورت میانگین خطای مربعات می باشد. روند ساده سازی که مقاله DDPM پیش گرفته و به این تابع هزینه رسیده است را به صورت مختصر شرح دهید. **(۵ نمره)**

(سوال ۸-امتیازی) مقاله DDPM برای تخمین نویز از یک شبکه U-NET(2) استفاده می کند. این شبکه برای تمامی زمان ها یکسان می باشد. در نتیجه نیاز است تا پارامتر زمان(که یک پارامتر گسسته است) نیز به شبکه داده شود. این پارامتر(زمان) با چه تکنیکی و به کدام قسمت (های) شبکه داده می‌شود؟ دلیل این امر را بررسی کنید. **(۵ نمره)**

(سوال ۹- امتیازی) همانطور که مشاهده کردیم، یکی از چالش ها و مشکلات مدل Diffusion پایه این است که ابعاد را کاهش نمی دهد و بازنمایی های میانی دارای ابعاد یکسانی با تصویر اولیه می باشد. این مساله (مخصوصا در دیتاست های با ابعاد بالا) باعث هزینه های محاسباتی بالایی می شود. مقاله Latent Diffusion(4) برای این چالش، راه حلی را ارائه داده است. این راه حل را به صورت مختصر بررسی کنید. این مدل برای بهبود مدل‌های پیشین چه تغییری در ساختار U-Net اعمال می‌کند؟ همچنین این مدل در ساختار U-NET از تکنیک Cross-Attention استفاده می کند. این تکنیک را به صورت مختصر مطرح و بررسی کنید. **(۵ نمره)**

(سوال ۱۰) مدل DDIM(3) تعمیمی از DDPM است. این موضوع را بررسی کنید. در چه صورت DDIM همان DDPM می شود؟ **(۶ نمره)**

(سوال ۱۱-امتیازی) چگونه می‌توان از یک مدل Diffusion-based ، در مسئله Semantic Segmentation استفاده کرد؟ (هدف از این سوال، ارائه یک روش خلاقانه است.) **(۵ نمره)**

سوالات پیاده‌سازی

در این قسمت قصد داریم که یک مدل DDPM پایه را بر روی دادگان CIFAR10 آموزش دهیم. برای این منظور قسمت‌های TODO فایل diffusion.ipynb را پر کنید. لازم به ذکر است، در این فایل برخی موارد مانند شبکه U-NET قرار داده شده است که می توانید از آن استفاده کنید.

(سوال ۱۲) رابطه گام به گام ($q(x_t|x_{t-1})$) مسیر رو به جلو (2.1) را پیاده‌سازی کنید و تصاویر نویزی مربوط به یک تصویر دلخواه در چند گام مختلف نمایش دهید. **(۲ نمره)**

(سوال ۱۳) رابطه ای را که برای مسیر رو به جلو در یک مرحله است ($q(x_t|x_0)$) ، پیاده سازی کنید و تصاویر مربوط به گام‌های مختلف مسیر افزودن نویز را نمایش دهید. **(۲ نمره)**

(سوال ۱۴) بخش 2.4 فایل diffusion.ipynb را کامل کنید. ابرپارامترهای[۴] آموزش را به همراه نمودار تابع خطای دادگان آموزش و اعتبارسنجی در طی فرآیند آموزش گزارش کنید. روال کلی فرایند آموزش و نمونه برداری به صورت زیر می باشد. (راهنمایی: برای آنکه به نتیجه مطلوبی برسید، تعداد گام‌های فرآیند افزودن نویز را برابر 1000 در نظر بگیرید.) **(۲۰ نمره)**

(سوال ۱۵) بخش 2.5 فایل diffusion.ipynb را که مربوط به فرآیند تولید نمونه است، کامل کنید. به کمک مدل آموزش یافته، ۱۰۰ نمونه تصویر تولید کنید. تصاویر را به صورت یک grid 10x10 گزارش دهید. **(۳ نمره)**

(سوال ۱۶) برای مدلی که در قسمت های قبل آموزش داده اید، معیار FID را محاسبه کرده و نتیجه را گزارش دهید. برای محاسبه این معیار نیازی به پیاده سازی آن نیست و به سادگی می توانید از پکیج pytorch-fid استفاده کنید. **(۳ نمره)**

Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \text{Uniform}(\{1, \dots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$
6: **until** converged

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \dots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

شکل 3- فرآیند آموزش و نمونه برداری مدل DDPM پایه

[۴] Hyperparameters

مراجع

(1) Ho, J., Jain, A. and Abbeel, P., 2020. Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, pp.6840-6851.

(2) Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.

(3) Song, J., Meng, C. and Ermon, S., 2020. Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502.

(4) Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10684-10695).

نکات پیاده سازی و تحویل

- مهلت ارسال این تمرین تا پایان روز "دوشنبه ۲۰ آذر ماه" خواهد بود.
- این زمان قابل تمدید نیست و درصورت نیاز میتوانید از grace time استفاده کنید.
- پیاده سازی با زبان برنامه نویسی پایتون باید باشد و کدهای شما باید قابل اجرا بوده و به همراه گزارش آپلود شوند.
- انجام این تمرین به صورت یک نفره می‌باشد.
- در صورت مشاهده هر گونه تشابه در گزارش کار یا کدهای پیاده‌سازی، این امر به منزله تقلب برای طرفین در نظر گرفته خواهد شد.
- استفاده از کدهای آماده بدون ذکر منبع و بدون تغییر به منزله تقلب خواهد بود و نمره تمرین شما **صفر** در نظر گرفته می‌شود
- در صورت رعایت نکردن فرمت گزارش کار نمره گزارش به شما تعلق نخواهد گرفت.
- تحویل تمرین به صورت **دستنویس** قابل پذیرش نیست.
- تمامی تصاویر و جداول مورد استفاده در گزارش کار باید دارای توضیح (caption) و شماره باشند.
- بخش زیادی از نمره شما مربوط به گزارش کار و روند حل مسئله است.
- لطفا گزارش ، فایل کدها و سایر ضمائم مورد نیاز را با فرمت زیر در سامانه بارگذاری نمائید.

HW2_Part2_[Lastname]_[StudentNumber].zip

به طور مثال:

HW2_ Part2_Zilouchian_12345678.zip

- در صورت وجود سوال و یا ابهام میتوانید از طریق رایانامه زیر با موضوع DGM_HW2_part2 با دستیاران آموزشی در ارتباط باشید:
  - پرسش اول

    alirezaghafouri@ut.ac.ir یا تلگرام alirezaghafouri1@

    p.zilouchian@ut.ac.ir یا تلگرام parham_zm@
  - پرسش دوم

    farshads7778@gmail.com یا تلگرام frshdsn6@

    a.hedayat.m@gmail.com یا تلگرام ahedayat2@

شاد و سلامت باشید.