

به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



درس مدل های مولد

مدرس: دکتر مصطفی توسلی پور

تمرین شماره ۳

آذر ماه ۱۴۰۳

۲	فهرست.....
۳	سوال اول Diffusion Model.....
۳	سوالات تئوری (25 نمره).....
۴	سوالات پیاده سازی (30 نمره).....
۶	سوال دوم - Score-based models.....
۶	سوالات تئوری (25 نمره).....
۷	سوالات پیاده سازی (30 نمره).....
۹	نکات تحویل.....

سوال اول DIFFUSION MODEL

مدل‌های based-Diffusion یک زنجیره مارکوف به طول T تشکیل می‌دهند. در این زنجیره مارکوف با افزودن نویز گوسی سعی می‌کنیم تا داده را در T مرحله به نویز کاملاً گوسی تبدیل کنیم.

در مدل‌های based-Diffusion دو مسیر Forward و Backward داریم. در مسیر Forward به داده‌ها، نویز اضافه می‌کنیم تا در نهایت به نویز کاملاً گوسی تبدیل شوند. هم‌چنین 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 based-Diffusion این است که این مدل‌ها ما را قادر می‌سازند که از یک نویز کاملاً گوسی، یک داده جدید را تولید کنیم.

سوالات تئوری (25 نمره)

۱. طبق مقاله **DDPM** در مسیر رو به جلو نیازی به اضافه کردن نویز به صورت تکرار شونده نیست. در واقع می‌توان با یک مرحله به هر کدام از بازنمایی‌های میانی رسید. با استفاده از خاصیت نویز گوسی، این فرآیند را اثبات کنید. (راهنمایی: اگر X_1 و X_2 دو متغیر نرمال مستقل باشند، آنگاه مجموع آن دو نیز یک متغیر نرمال خواهد بود). **(5 نمره)**
۲. یک فرض مهم در مسیر روبه عقب این است که توزیع $q(x_{t-1} \vee x_t)$ را گوسی فرض کنیم. در چه صورتی این فرض صادق است؟ با وجود اینکه این توزیع گوسی می‌باشد، مسیر رو به عقب را به کمک توزیع $q(x_{t-1} \vee x_t, x_0)$ تخمین می‌زنند. دلیل این امر را بررسی کنید. **(5 نمره)**
۳. تابع هزینه معرفی شده در مقاله دارای سه ترم است. مفهوم هر ترم را به صورت مختصر توضیح داده و بیان کنید مقاله DDPM از کدام ترم(ها) در فرآیند بهینه‌سازی استفاده نکرده است؟ چرا؟ **(5 نمره)**
۴. مدل **DDIM** تعمیمی از DDPM است. در چه صورت DDIM همان DDPM می‌شود؟ **(3 نمره)**
۵. همانطور که می‌دانید یکی از مشکلات مدل‌های Diffusion کندی سرعت است. به دلیل اینکه در مسیر backward باید گام به گام denoising انجام شود، فرآیند نمونه‌برداری کند است. از این رو، بسیاری از مقالات و پژوهش‌ها در تلاش هستند تا این مشکل را برطرف کنند و فرآیند

نمونه‌برداری را تسریع بخشند. مقاله [ES-DDPM](#) را مطالعه کرده و توضیح دهید این مقاله چه رویکردی برای افزایش سرعت اتخاذ کرده است. (7 نمره)

سوالات پیاده سازی (30 نمره)

در این قسمت قصد داریم که مدل DDPM و DDIM را با هم مقایسه کنیم. بدین منظور یک مدل DDPM پایه را بر روی مجموعه داده [sprites](#) آموزش می‌دهیم. نمونه‌ای از تصاویر این مجموعه داده که مربوط به شخصیت‌های بازی‌های کامپیوتری می‌باشد، مشاهده می‌کنید.



شکل ۱- تعدادی از تصاویر مجموعه داده sprites

در طی این تمرین، مرحله نمونه‌گیری را به کمک هر دو الگوریتم DDPM و DDIM پیاده سازی کرده و نتایج را مقایسه می‌کنیم.

برای سهولت می‌توانید از نوت بوک پیوست شده در تمرین استفاده کنید. قسمت dataset و مدل Unet پیاده سازی شده است و شما باید سایر قسمت‌ها را که با TODO# مشخص شده است، تکمیل کنید (اجباری بر استفاده از این نوت بوک نیست).

۱. برای فرآیند forward نیاز به یک noise scheduler داریم تا فرآیند اضافه کردن نویز را به کمک آن انجام دهیم. بدین منظور linear scheduler را پیاده سازی کنید و برای پیاده سازی از مقادیر $\beta_1 \square \beta_2$ که در قسمت hyperparameters نوت بوک مقدار دهی شده‌اند، استفاده کنید. پس از تعریف α_t ، β_t را نیز تعریف کنید. (2 نمره)

۲. رابطه‌ای را که برای مسیر رو به جلو در یک مرحله است، پیاده سازی کنید. این پیاده سازی را در تابع "perturb_input" انجام دهید. (5 نمره)

۳. قسمت "Train Diffusion" در فایل نوت بوک که مربوط به آموزش مدل می‌باشد، تکمیل کنید. نمودار تابع خطای دادگان آموزش در طی فرآیند گزارش کنید. ابرپارامترهای موردنیاز در قسمت *hyperparameters* داده شده است. در صورت لزوم می‌توانید آن‌ها را تغییر دهید. (5 نمره)

۴. قسمت *DDPM sampling* را تکمیل کنید. برای پیاده سازی از الگوریتم زیر که در مقاله اصلی آمده است، استفاده کنید. توضیح دهید، $\delta_t z$ که در واقع یک نویز نرمال است که به تصویر *denoise* شده اضافه می‌گردد، به چه منظور است؟

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:    $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \epsilon_{\theta}(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$ 
5: end for
6: return  $\mathbf{x}_0$ 

```

شکل ۲- الگوریتم *sampling* برای DDPM

سپس 50 نمونه داده تولید کرده و به کمک "plot_sample" در قسمت *utils* نمایش دهید. (تابع نمایش تصویر را می‌توانید خودتان پیاده سازی کنید و لزومی بر استفاده از این تابع نمی‌باشد). (5 نمره)

۵. حال تابع "sample_ddim" را که مربوط به نمونه‌گیری به روش DDIM است، تکمیل کنید.

سپس 50 نمونه داده تولید کرده و نمایش دهید. (5 نمره)

۶. به کمک کد زیر در بلاک آخر نوت بوک، زمان نمونه‌گیری دو مدل DDPM و DDIM را محاسبه کرده و با یکدیگر مقایسه کنید. (3 نمره)

```

1. %timeit -r 1 sample_ddim(50)
2. %timeit -r 1 sample_ddpm(50)

```

۷. برای هر دو مدل DDPM و DDIM که در قسمت های قبل آموزش داده‌اید، معیار FID را محاسبه کرده و نتایج را مقایسه کنید. برای محاسبه این معیار نیازی به پیاده سازی آن نیست و به سادگی می‌توانید از پکیج *pytorch-fid* استفاده کنید. (5 نمره)

سوال دوم - SCORE-BASED MODELS

مدل های Score-based برای مدلسازی توزیع احتمال، به جای تخمین خود توزیع به یادگیری تابع Score می پردازند. این مدل ها شباهت زیادی به مدل های Diffusion-based دارند و می توان آنها را خاستگاه مدل های Diffusion امروزی دانست و بوسیله آنها به درک بهتری نسبت به تئوری پشت این مدل ها دست یافت.

سوالات تئوری (25 نمره)

۱. هر توزیع احتمالی را می توانیم به صورت زیر پارامتریزه کنیم. توضیح دهید محاسبه و یادگیری کدام بخش از این نمایش مشکل ساز است؟ در مدل های Score-based چه تابعی به جای آن آموخته می شود و استفاده از آن چگونه این مسئله را حل می کند؟ (4 نمره)

$$p_{\theta}(x) = \frac{e^{-f_{\theta}(x)}}{Z_{\theta}}$$

۲. با مطالعه بخش های ابتدایی مقاله [Sliced Score Matching](#)، روش ارائه شده در این مقاله برای کاهش سختی محاسبات تابع هزینه مدل های Score-based را به اختصار توضیح دهید. (7 نمره)

$$L(\theta) = \frac{1}{2} E_{p(x)}[s_{\theta}(x)^2] + E_{p(x)}[\nabla_x s_{\theta}(x)]$$

۳. اضافه کردن نویز به داده ها در زمان آموزش (Noise perturbation) به چه دلیل انجام می شود؟ توضیح دهید این نویز با چه توزیع احتمالی مدل می شود و پارامترهای آن در هنگام آموزش به چه صورت در نظر گرفته می شوند. همچنین با مقایسه آن با رابطه اضافه کردن نویز در مدل DDPM، تفاوت اصلی این مدل ها در هنگام آموزش را تحلیل کنید. (5 نمره)
۴. فرآیند نمونه گیری از یک مدل Score-based را توضیح دهید. چرا در این فرآیند از روش Langevin Dynamics Sampling استفاده می کنیم؟ (4 نمره)

۵. با مراجعه به [این مقاله](#)، روش Annealed Langevin Dynamics را به اختصار توضیح داده و علت استفاده از آن در مقابل روش سوال قبل را توضیح دهید. (5 نمره)

سوالات پیاده سازی (30 نمره)

در این قسمت قصد داریم آموخته های خود از مدل های Score-based را در قالب یک تمرین پیاده سازی به صورت تصویری مشاهده کنیم تا نسبت به مفاهیم و روش های مورد استفاده در آنها شناخت بهتری پیدا کنیم.

برای پیاده سازی این بخش از نوت بوک پیوست شده در تمرین استفاده کنید. با وارد کردن شماره دانشجویی خود در ابتدای نوتبوک، توزیع و دیتاست خود را بسازید و در ادامه قسمت هایی را که با TODO# مشخص شده است را تکمیل کنید.

۱. توزیع مورد بررسی در این بخش «مخلوط 2 گوسی» یا Mixture of Two Gaussians است. با استفاده از رابطه pdf این توزیع چند متغیره، score function آن را به صورت تحلیلی محاسبه کرده و سپس آن را در قسمت [TODO 1] کد وارد کنید و نمودار ترسیم شده را گزارش و تحلیل کنید. (7 نمره)

$$p(x) = w_1 N(x|\mu_1, \Sigma_1) + w_2 N(x|\mu_2, \Sigma_2)$$

۲. فرآیند نویزی کردن، آموزش مدل و تابع هزینه زیر را برای score matching with noise perturbation در قسمت [TODO 2] پیاده سازی کنید.

$$L(\theta) = \frac{1}{2} E_{xp(x), \mathbb{x} \sim p_\sigma(\mathbb{x}|x)} \left[\left\| s_\theta(\mathbb{x}) + \frac{1}{\sigma^2} \epsilon \right\|_2^2 \right]$$

$$\mathbb{x} = x + \epsilon; \epsilon \sim N(\mu = 0, \Sigma = \sigma^2 I) \rightarrow p_\sigma(\mathbb{x}|x) = N(\mathbb{x} \vee \mu = x, \Sigma = \sigma^2 I)$$

مدل تعریف شده را روی این تابع هزینه به ازای $\sigma \in \{1, 3, 7\}$ آموزش دهید و نمودارهای حاصل را مقایسه و تحلیل کنید. (7 نمره)

3. توابع نمونه گیری غیرتصادفی و همچنین نمونه گیری به روش Langevin dynamics را در قسمت [TODO 3] تکمیل نمایید و نمودار نمونه گیری را به ازای چند نقطه شروع متفاوت گزارش و تفسیر کنید. (5 نمره)

4. با اجرای هرکدام از روش های سوال قبل به ازای 1000 نقطه شروع و رسم scatter plot مربوط به هر روش، کیفیت تخمین توزیع اصلی توسط هر روش را مقایسه کنید. (4 نمره)

۵. به کمک توابع توسعه داده شده در بخش های قبل، مدلی آموزش دهید که به جای انحراف معیار ثابت، مقادیر مختلفی از انحراف معیار (شدت نویز) را در هنگام آموزش دیده باشد:

$$\sigma \in [1, 20]$$

سپس برای نمونه گیری از آن، σ_t را با مقادیر بزرگ شروع کرده و به مراتب کاهش دهید.
همچنین نمودارهای score field به ازای مقادیر مختلف σ و توزیع تخمین زده شده به روش Langevin dynamics را رسم کنید و با بخش های قبل مقایسه کنید. (7 نمره)

نکات تحویل

- مهلت ارسال این تمرین تا پایان روز "جمعه ۳۰ آذر ماه" خواهد بود.
- این زمان قابل تمدید نیست و در صورت نیاز می‌توانید از grace time استفاده کنید.
- در نظر داشته باشید که حداکثر مهلت آپلود تمرین در سامانه تا ۷ روز پس مهلت تحویل است و پس از آن سامانه بسته خواهد شد.
- 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 ضامین مورد نیاز را با فرمت زیر در سامانه بارگذاری نمایید.
- HW1_[Lastname]_[StudentNumber].zip
- در صورت وجود سوال و یا ابهام می‌توانید از طریق رایانامه زیر با موضوع TAI_HW1 با دستیاران آموزشی در ارتباط باشید:

○ سوال اول

یاسمن حق‌بین hbn.yasaman@gmail.com

○ سوال دوم

عرفان عسگری erfanasgari21@gmail.com

با آرزوی سلامتی و موفقیت روزافزون.