



به نام خدا
درس یادگیری عمیق
تمرین سری ششم
استاد درس : دکتر محمدرضا محمدی
دستیاران : مهدی خورش، سید محمد موسوی،
امیرحسین نمازی
دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر
نیمسال دوم تحصیلی ۱۴۰۳ - ۱۴۰۴

مهلت تحویل : ۱۴۰۴/۰۳/۲۰

لطفاً به نکات موجود در سند قوانین انجام و تحویل تمرین ها دقت فرمایید.

سوالات تئوری



۱. در این درس با مدل های انتشار نویز زدایی^۱ آشنا شدید.

اکنون مدل های امتیازی شرطی شده با نویز^۲ را مطالعه کرده و به پرسش های زیر پاسخ دهید (۱۵ نمره):

- (آ) روش را کوتاه توضیح داده و بگویید Langevin dynamics چیست و چه کارکردی در آن دارد؟
- (ب) چگونه مدل های انتشار چالش فرض چندگانه و چگالی داده کم را برطرف می کنند؟
- (ج) اندازه نویز افزوده شده و گام زمانی چه تاثیری در مدل های انتشار دارد؟ برای مثال اگر به جای ۴۰ مرحله نویز تنها ۳ گام نویز ولی با اندازه بیشتری افزوده شود یا برعکس ۸۰ گام نویز با اندازه کمتری افزوده شود، هر کدام چه برتری و کاستی هایی دارد؟



۲. WGAN، که در این مقاله معرفی شد، یکی از اولین گام های بزرگ به سوی پایداری سازی آموزش GAN بود. با چند تغییر، نویسندگان توانستند نشان دهند چگونه می توان GAN هایی را آموزش داد که دارای دو ویژگی زیر باشند:

¹DDPM

²NCSN

- یک معیار ضرر معنادار که با همگرایی generator و کیفیت نمونه‌ها همبستگی دارد.
- بهبود پایداری فرآیند بهینه‌سازی. به طور خاص، این مقاله تابع ضرر واسرشتاین را برای هر دو discriminator و generator معرفی می‌کند. استفاده از این تابع ضرر به جای binary cross-entropy منجر به همگرایی پایدارتر GAN می‌شود.

در بسیاری از موارد، الگوریتم GAN می‌تواند به عنوان کمینه‌سازی انحراف بین توزیع داده‌ها p_{data} و توزیع مدل p_g در نظر گرفته شود. در این مسئله، ما یک مشکل با انحراف‌های مختلف (مثلاً انحراف Jensen-Shannon و انحراف KL) و یک راه‌حل بالقوه برای رفع آن (Wasserstein Distance) را بررسی خواهیم کرد (۲۰ نمره).

الف) فرض کنید $p_{data} \sim \mathcal{N}(\theta_0, \epsilon^2)$ و $p_g \sim \mathcal{N}(\theta, \epsilon^2)$ توزیع‌های نرمال با انحراف معیار ϵ باشند که به ترتیب حول $\theta_0 \in \mathbb{R}$ و $\theta \in \mathbb{R}$ مرکزیت دارند. نشان دهید که:

$$D_{KL}(p_g \| p_{data}) = \frac{(\theta - \theta_0)^2}{2\epsilon^2}$$

ب) فرض کنید p_g و p_{data} هر دو جرم احتمال را فقط در یک بخش بسیار کوچک از دامنه داشته باشند؛ یعنی، حد $\epsilon \rightarrow 0$. چه اتفاقی برای $D_{KL}(p_g \| p_{data})$ و مشتق آن نسبت به θ با فرض اینکه $\theta \neq \theta_0$ می‌افتد؟

ج) آیا این امر مشکلی برای یک GAN که با تابع ضرر تعریف‌شده‌ی زیر آموزش داده شده باشد، ایجاد می‌کند؟ چرا؟

$$L_G(\theta; \phi) = \mathbb{E}_{x \sim p_{\theta}(x)}[\log(1 - D_{\phi}(x))] - \mathbb{E}_{x \sim p_{\theta}(x)}[\log D_{\phi}(x)]$$

د) تحت همان شرایط (ب)، انحراف KL، انحراف JS و فاصله واسرشتاین را مقایسه کنید.



۳. در ارتباط با DDPM به سوالات زیر پاسخ دهید (۲۰ نمره):

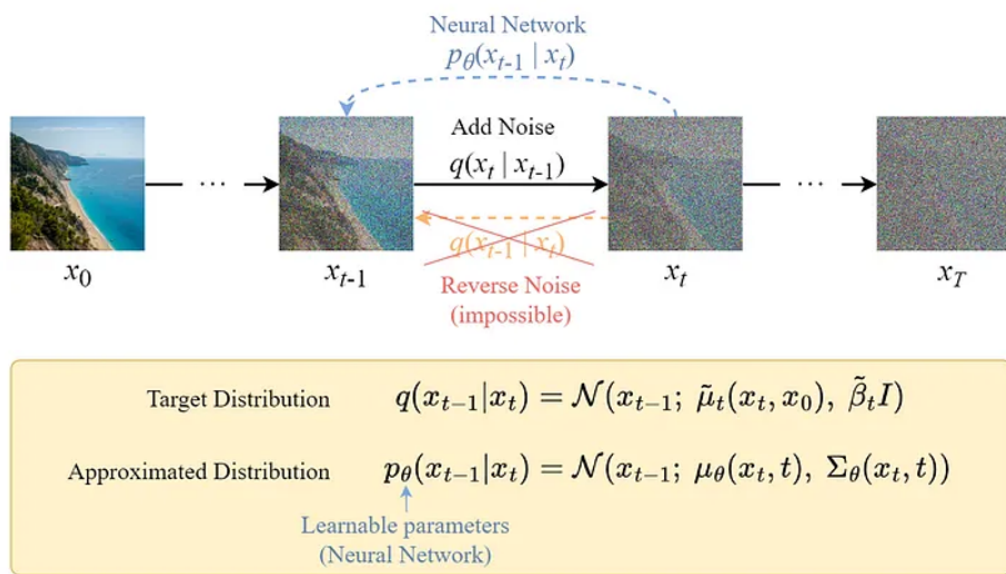
(آ) فرض کنید x_t متغیر تصادفی (یک نمونه داده مثلاً یک تصویر) ای است که در فرایند Forward Diffusion Process با اضافه شدن مقداری نویز به x_{t-1} بدست می‌آید. بر این اساس عبارت زیر را توضیح دهید:

$$x_t = \sqrt{1 - \beta_t} x_{t-1} + \sqrt{\beta_t} \epsilon$$

(ب) با استفاده از ترند تغییر پارامتر^۳ نشان دهید عبارت قسمت (آ) را می‌توان بصورت زیر نوشت:

$$x_t = \sqrt{\alpha_t} x_0 + \sqrt{1 - \alpha_t} \epsilon$$

(ج) با توجه به شکل ۱ توضیح دهید که چرا به هنگام حذف نویز از داده در فرایند Reverse Diffusion Process در هر مرحله نمی‌توانیم به طور مستقیم نویز را حذف کنیم (از x_t به x_{t-1} برسیم) و چرا این عمل برایمان غیرقابل حل^۴ می‌باشد؟ (با توجه به شکل ۱ این امر موجب می‌شود تا از روش‌های تخمین تابع مانند شبکه‌های عصبی استفاده کنیم).



شکل ۱: تابع توزیع تخمین‌زده‌شده توسط شبکه‌ی عصبی

(د) تحقیق کنید که چه ویژگی‌هایی از مدل U-Net موجب شد که نویسندگان مقاله DDPM از آن برای معماری کار خود استفاده کنند؟

سوالات عملی

۴. به این **لینک گیت‌هاب** رفته و به پرسش‌های زیر پاسخ دهید (۲۵ نمره).

(آ) درباره هدایت بدون دسته‌بند مطالعه کرده و بگویید در کجای کد از آن استفاده شده است؟

³Reparameterization trick

⁴intractable

- (ب) شرط در مدل انتشار شرطی به چه شکل اعمال شده است؟ آیا تنها به همین روش می‌توان شرط را اعمال کرد؟ اگر خیر، روش‌های دیگر چیست؟
- (ج) فرض کنید شرط ما متن باشد و راهنمای بدون طبقه‌بندی را به شکل زیر استفاده کنیم:

$$text_embeddings = text_encoder(["\text{ }", prompt])$$

کد را بر این اساس تغییر بدهید و بگویید کدام ماژول‌ها تغییر می‌کنند.

- (د) stable diffusion چیست؟ کد را براساس آن تغییر دهید و بگویید کدام ماژول‌ها تغییر می‌کنند.



۵. در این تمرین، شما قرار است دو مدل مولد مهم در یادگیری ماشین را روی مجموعه داده‌ی MNIST پیاده‌سازی کنید: خودرمزگذار متغیر (VAE) و شبکه مولد تخاصمی (GAN) (۲۰ نمره). مدل‌های VAE یک متغیر پنهان احتمالاتی را از داده‌های ورودی یاد می‌گیرند، سپس از روی این توزیع نمونه‌برداری کرده و داده‌های جدیدی تولید می‌کنند. مدل‌های GAN از یک شبکه مولد برای تولید تصاویر استفاده می‌کنند که توزیع آن‌ها به توزیع داده‌های واقعی نزدیک است.

نوتبوک GAN-VAE.ipynb حاوی کدهایی است که برخی بخش‌های آن با برچسب TODO مشخص شده‌اند. شما باید این بخش‌ها را با دقت تکمیل کنید تا مدل‌ها به درستی پیاده‌سازی شوند. پیشنهاد می‌شود پیش از نوشتن کد، تمامی توضیحات و سلول‌ها را به دقت مطالعه کنید تا درک کاملی از ساختار مدل‌ها و نحوه پیاده‌سازی آن‌ها داشته باشید.



۶. در این تمرین، هدف شما پیاده‌سازی یک مدل احتمالی انتشار نویز است. برای درک بهتر مفاهیم، توصیه می‌شود مقاله اصلی مربوط به **DDPM** را مطالعه کنید (۲۰ نمره). شما باید نوتبوک DDPM.ipynb را تکمیل کرده و تمام سلول‌های آن را اجرا کنید. بخش‌هایی که نیاز به تکمیل دارند، با برچسب TODO در داخل بلوک‌های کد مشخص شده‌اند. پیش از شروع به نوشتن کد، تمام توضیحات متنی و کدهای داده‌شده را با دقت بخوانید. این نوت‌بوک با استفاده از محیط‌های رایگان Google Colab و Kaggle آزمایش شده است؛ می‌توانید از این پلتفرم‌ها برای اجرای کدهای خود استفاده کنید. اطمینان حاصل کنید که تمامی سلول‌ها بدون خطا اجرا می‌شوند و عملکرد مورد انتظار را ارائه می‌دهند.