



دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

یادگیری عمیق

نیم سال اول ۱۴۰۲-۰۲

استاد: حمید بیگی

گردآورندگان: حسن حمیدی، امیرحسین عاملی، مهدی کافی

بررسی و بازبینی: علی قاری زاده

مهلت ارسال: ۲۵ دی

شبکه های مولد

تمرین سری پنجم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- امکان ارسال تمرین با ۷ روز تاخیر مجاز وجود دارد. همچنین به ازای هر روز تاخیر غیر مجاز ۱۰ درصد از نمره تمرین به صورت ساعتی کسر خواهد شد.
- 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.
- پاسخ تمامی سوالات (تئوری و عملی) را در یک فایل فشرده به صورت `DL_HW5_[firstName]_[lastName]_[StudentId]` نامگذاری کرده و ارسال کنید.

سوالات نظری (۶۵ نمره)

۱. معیارهای ارزیابی مدل های تولید کننده: (نمره ۱۵)

مهمترین ویژگی ای که از مدل های تولید کننده انتظار می رود تولید داده های معنادار و با توزیع پراکنده است. برای مثال تولید تصاویر بدون نویز و با ویژگی های معنادار از مزایای یک مدل تولید کننده می باشد.

(آ) برای سنجش مدل های تولید کننده معیارهای FID و IS استفاده می شوند. مختصرا نحوه ی کارکرد هر یک از این معیارها را توضیح دهید.

(ب) مشهورترین معیار در ادبیات مدل های تولید کننده در حوزه ی تصویر معیار FID است. با ارائه ی مثال معایب این معیار را توضیح دهید.

(ج) در راستای سنجش دقیقتر عملکرد مدل های تولید کننده، اخیرا معیارهای جدیدتری در مقالات معرفی شده اند. با مراجعه به این مقاله معیار Density and Coverage را توضیح دهید. این مقاله چه ویژگی ای از مدل های تولید کننده را ارزیابی می کند و نحوه ی عملکرد الگوریتم آن را توضیح دهید.

۲. بازبینی تابع هزینه ی خودکدگذارها: (نمره ۱۵)

می توانیم تابع هزینه ی ELBO را به صورت زیر بازنویسی کنیم.

$$E_{x \sim p_{data}(x)} [E_{q_{\phi}(z|x)} [\ln p_{\theta}(x|z) + \ln p_{\lambda}(z) - \ln q_{\phi}(z|x)]]$$

در حالیکه $p_{data}(x) = \frac{1}{N} \sum_{n=1}^N \delta(x - x_n)$ است.

(آ) بخش زیر از تابع هزینه ی بالا، عبارت منظم ساز بین کدگذار و توزیع پیشین است.

$$\Omega = E_{x \sim p_{data}(x)} [E_{q_{\phi}(z|x)} [\ln p_{\lambda}(z) - \ln q_{\phi}(z|x)]]$$

با محاسبات ریاضی نشان دهید که این عبارت را می توان به صورت ترکیبی از cross-entropy بین توزیع پیشین و پسین یعنی $CE[q_{\phi}(z) || p_{\lambda}(z)]$ و آنتروپی شرطی $q_{\phi}(z|x)$ یعنی $H[q_{\phi}(z|x)]$ نوشت.

۳. می‌دانیم تابع هدف خودکدگذار وردشی به صورت زیر تعریف می‌شود: (نمره ۱۵)

$$\mathcal{L}_{ELBO}(\phi, \theta) = \mathbb{E}_{x' \sim p_{Data}} [\mathbb{E}_{z' \sim q_{\phi}(z|x')} [\log p_{\theta}(x'|z')] - KL(q_{\phi}(z|x') || p(z))]$$

$$\approx \frac{1}{n} \sum_{i=1}^n \left(\mathbb{E}_{z' \sim q_{\phi}(z|x^{(i)})} [\log p_{\theta}(x^{(i)}|z')] - KL(q_{\phi}(z|x^{(i)}) || p(z)) \right)$$

در این رابطه، $q_{\phi}(z|x^{(i)})$ توزیع کدگذار روی فضای نهان به ازای ورودی $x^{(i)}$ است. $p_{\theta}(x'|z')$ نیز توزیع کدگشا روی فضای خروجی به ازای ورودی z' است و $p_{\theta}(x^{(i)}|z')$ مقدار احتمال آن توزیع به ازای $x^{(i)}$. همچنین توزیع $p(z)$ یک توزیع گاوسی استاندارد (میانگین صفر و ماتریس کوواریانس I) است.

(آ) اگر توزیع خروجی کدگشا را به صورت گاوسی چندمتغیره $\mathcal{N}(\mu, I)$ فرض کنیم که در آن بردار μ خروجی آخرین لایه‌ی کدگشا (به ازای ورودی z') باشد، جمله اول از رابطه بالا $\mathbb{E}_{z' \sim q_{\phi}(z|x')} [\log p_{\theta}(x'|z')]$ را محاسبه کنید. در این حالت جمله‌ای به شکل Mean Squared Error (MSE) در تابع هدف ظاهر می‌شود.

(ب) اگر داده‌ی ورودی از نوع باینری باشد (مثلاً تصویر Black and White)، می‌توان به جای توزیع گاوسی چندمتغیره روی خروجی کدگشا، از توزیع برنولی چندمتغیره استفاده کرد. توزیع خروجی کدگشا را به شکل برنولی چندمتغیره در نظر بگیرید و جمله‌ی اول رابطه‌ی بالا را محاسبه کنید. در روابط خود، خروجی آخرین لایه از کدگشا (پس از اعمال sigmoid) را a بنامید. در این حالت جمله‌ای به شکل Binary Cross Entropy (BCE) در تابع هدف ظاهر می‌شود.

(ج) در پیاده سازی خودکدگذار وردشی، گاهی اوقات که ورودی باینری نیست نیز با normalize کردن ورودی به بازه‌ی $[0, 1]$ ، از BCE در تابع هدف استفاده می‌کنند. توضیح دهید که استفاده از BCE در قیاس با MSE چه پیامدهایی در پی خواهد داشت؟

۴. با فرض دو توزیع احتمال زیر: (نمره ۱۰)

$$\forall (x, y) \sim P, x = 0, y \sim \text{Uniform}(0, 1)$$

$$\forall (x, y) \sim Q, x = \theta (0 \leq \theta \leq 1), y \sim \text{Uniform}(0, 1)$$

(آ) مقادیر $KL(P||Q)$, $KL(Q||P)$, $JSD(P, Q)$, $W(P, Q)$ را محاسبه کنید.

(ب) دو توزیع احتمال P, Q دیگر پیشنهاد دهید که $JSD(P, Q)$ نسبت به پارامترهای Q یا P مشتق پذیر نباشند.

(ج) از مثال های بالا برای توضیح مشکلات JSD در ساختار GAN پایه استفاده کنید.

۵. به سوالات زیر در مورد $WGAN$ پاسخ دهید. (نمره ۱۰)

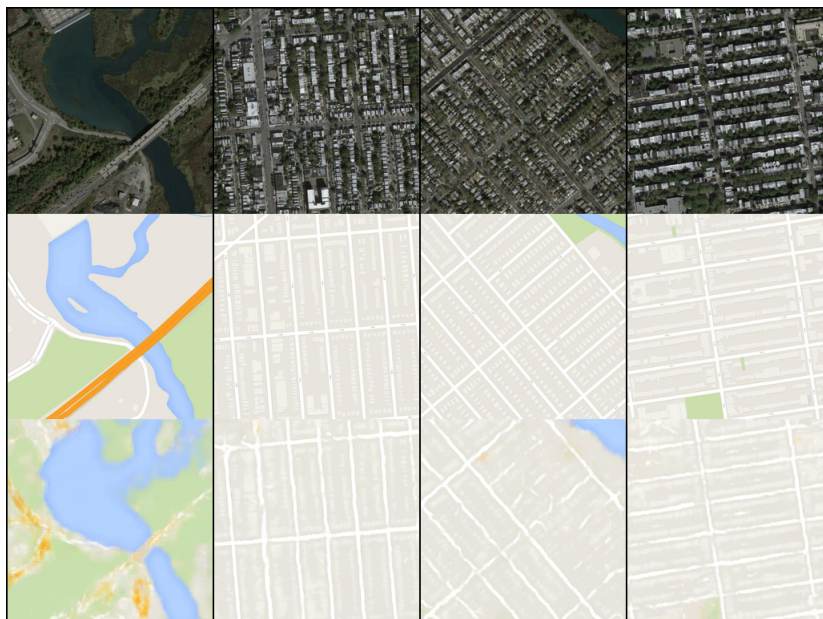
(آ) تفاوت $Critic$ در این مدل و $Discriminator$ در مدل GAN پایه را شرح دهید.

(ب) چرا از تابع فعال سازی مانند $Sigmoid$ در لایه آخر $Critic$ استفاده نمی‌شود؟

(ج) چرا برای آموزش بهتر مولد باید آموزش $Critic$ به تعداد بیشتری نسبت به مولد انجام شود؟ چه محدودیتی بر روی آپدیت وزن های این قسمت از مدل داریم دو روش برای اعمال این محدودیت را نام ببرید.

سوالات عملی (۳۵ نمره)

۱. فایل نوت بوک `pix2pix.ipynb` حاوی یک پیاده سازی ناکامل از **مقاله** می باشد. ابتدا مقاله و جزئیات پیاده سازی آن را خوانده سپس این نوت بوک را کامل کنید. آموزش بر روی وظیفه ی ترجمه تصویر به تصویر تصاویر ماهواره ای می باشد. به دلیل محدودیت های سخت افزاری نتیجه مانند نمونه زیر قابل قبول است. برای نتیجه مانند تصویر زیر باید مدل را حدود ۱۰۰ اپیاک آموزش دهید این مقدار آموزش بر روی گوگل کولب حدود یک ساعت زمان خواهد برد.



شکل ۱: ردیف اول ورودی مدل، ردیف دوم خروجی مرجع، ردیف سوم خروجی مدل بعد از ۶۰ اپیاک

۲. در این تمرین عملی قصد داریم یک VAE را بر روی داده گان MNIST آموزش دهیم. در فایل `DL_HW5_VAE.ipynb` به ترتیب بخش های مشخص شده با TODO را تکمیل کنید.