

به نام خدا



دانشگاه تهران  
پردیس دانشکده‌های فنی  
دانشکده برق و کامپیوتر



درس مدل‌های مولد عمیق

مدرس: دکتر مصطفی توسلی‌پور

تمرین شماره ۲

آبان ماه ۱۴۰۳

سوال اول – مدل های مبتنی بر جریان.....	۴
بخش اول.....	۴
زیربخش اول.....	۴
زیربخش دوم (۲۰ نمره).....	۵
زیربخش سوم (۳ نمره).....	۵
زیربخش چهارم (۱۲ نمره).....	۵
زیربخش پنجم (امتیازی) (۵ نمره).....	۶
بخش دوم.....	۶
زیربخش اول (۱۰ نمره).....	۷
زیربخش دوم (۵ نمره).....	۷
سوال دوم – مدل های مولد متخاصم.....	۸
بخش اول – GAN.....	۸
زیربخش اول (۵ نمره).....	۹
زیربخش دوم (۵ نمره).....	۹
زیربخش سوم (۳ نمره).....	۹
بخش دوم – wasserstein gan.....	۱۰
زیربخش اول (۵ نمره).....	۱۰
زیربخش دوم (۴ نمره).....	۱۱
زیربخش سوم (۴ نمره).....	۱۱
زیربخش چهارم (۴ نمره).....	۱۱
زیر بخش سوم – پیاده سازی مدل های gan.....	۱۱

۱۱	زیربخش اول (۴ نمره).....
۱۳	زیربخش دوم (۳ نمره).....
۱۳	زیربخش سوم (۶ نمره).....
۱۴	زیربخش چهارم ( ۴ نمره ) .....
۱۴	زیربخش پنجم (۳ نمره).....
۱۴	زیربخش ششم (۵ نمره).....
۱۵	نکات تحویل.....

## سوال اول – مدل های مبتنی بر جریان

در این سؤال، هدف آشنایی بیشتر با مدل های مبتنی بر جریان برای متغیرهای تصادفی پیوسته است. این مدل ها از طریق بیشینه سازی لگاریتم درستنمایی<sup>۱</sup> دادگان آموزش می بینند. در ادامه، به بررسی یکی از کاربردهای مهم این مدل ها می پردازیم:

مدل های مولد در یادگیری ماشین ابزارهای محبوبی برای شبیه سازی توزیع داده ها هستند و نقش مهمی در تشخیص داده های خارج از توزیع<sup>۲</sup> دارند. یکی از این روش ها، استفاده از مدل های جریان نرمال کننده<sup>۳</sup> است که به دلیل توانایی در مدل سازی توزیع های پیچیده، مورد توجه قرار گرفته اند. مدل هایی مانند RealNVP در این دسته قرار می گیرند و قادرند توزیع های ساده را به توزیع های پیچیده تر تبدیل کنند، در حالی که احتمال های دقیق را حفظ می کنند. به طور خاص، این مدل ها با تخمین لگاریتم درستنمایی داده های ورودی می توانند احتمال داده های داخل توزیع<sup>۴</sup> و خارج از توزیع را محاسبه کنند.

با این حال، تحقیقات نشان داده اند که استفاده از مدل های مبتنی بر جریان در شناسایی داده های خارج از توزیع با چالش هایی همراه است.

### بخش اول

#### زیربخش اول

در این قسمت، هدف آشنایی بیشتر با مدل جریان نرمال RealNVP است. ابتدا مقاله ای که در این [لینک](#) قرار دارد را مطالعه کنید.

---

<sup>۱</sup> log-likelihood

<sup>۲</sup> Out-of-Distribution

<sup>۳</sup> Normalizing Flow

<sup>۴</sup> In-Distribution

---

### زیربخش دوم (۲۰ نمره)

در این بخش، هدف پیاده‌سازی مدل RealNVP بر روی پایگاه داده تصویری FashionMNIST است. برای این کار لازم است که تبدیل کوپلینگ را پیاده‌سازی کنید. در این تبدیل، ورودی را به دو قسمت مساوی تقسیم کرده و پس از هر تبدیل کوپلینگ، از یک تبدیل جایگشت برای تغییر ترتیب خروجی‌ها استفاده کنید (بخش ۳,۵ مقاله اصلی).

نکات مهم در پیاده‌سازی:

۱. پایگاه داده را به دو زیرمجموعه آموزش، ارزیابی تقسیم کنید.
  ۲. انتخاب معماری شبکه‌های  $S$  و  $t$  به اختیار شما است.
  ۳. برای بهبود عملکرد مدل، از حداقل ۴ جفت تبدیل (کوپلینگ و جایگشت) پشت سر هم استفاده نمایید.
- در پایان، نمودار مقدار خطای آموزش و تست را بر حسب اپاک‌ها رسم کرده و ۹ تصویر تصادفی از مدل آموزش‌دیده را تولید کرده و نمایش دهید.

---

### زیربخش سوم (۳ نمره)

یکی از روش‌های تشخیص داده‌های خارج از توزیع، استفاده از روش‌های مبتنی بر احتمال<sup>۱</sup> است. در این زمینه تحقیق کنید، یکی از روش‌های موجود در این دسته را توضیح دهید و یکی از دلایل ناکامی این مدل‌ها در تشخیص داده‌های خارج از توزیع را بیان کنید.

---

### زیربخش چهارم (۱۲ نمره)

---

<sup>۱</sup> Likelihood-based OOD Detection

یکی از کاربردهای مهم مدل‌های مولد، شناسایی داده‌های خارج از توزیع از طریق محاسبه مقدار لگاریتم درست‌نمایی است. برای بررسی این قابلیت، از پایگاه داده‌های MNIST و یک دیتاست دیگر به انتخاب خود استفاده کرده و مقادیر log-likelihood را برای این دو دیتاست محاسبه کنید. علاوه بر این، همین مقادیر را برای داده‌های آموزشی و ارزیابی از داده‌های FashionMNIST نیز به دست آورید.

با تحلیل مقادیر به دست آمده، بررسی کنید که آیا امکان شناسایی داده‌های خارج از توزیع به طور مؤثر فراهم است یا خیر. یافته‌ها و نتیجه‌گیری خود را در قالب تحلیلی دقیق ارائه دهید.

---

### زیربخش پنجم (امتیازی) (۵ نمره)

آزمایش زیربخش چهارم را در فضای نهان<sup>۱</sup> تکرار کنید؛ به این صورت که ابتدا داده‌ها را با یک انکودر به فضای نهان انتقال داده و جریان RealNVP را با همان تنظیمات قبلی در این فضا اجرا کنید. پس از محاسبه مجدد مقادیر log-likelihood برای هر چهار دیتاست گفته شده در سوال قبل (MNIST, FashionMNIST-train, FashionMNIST-test, Custom dataset)، مشاهدات خود را تحلیل کرده و نتایج این بخش را با بخش قبل مقایسه کنید. در صورت مشاهده بهبود، علت آن را توضیح دهید.

### بخش دوم

در این سؤال، هدف آشنایی بیشتر با مدل‌های جریان باقیمانده<sup>۲</sup> و ساختار آنهاست که به عنوان یکی از مدل‌های پیشرفته در یادگیری جریان‌های نرمال‌کننده شناخته می‌شوند. این مدل‌ها با معرفی توابع تبدیل خاصی، سعی در حفظ دقت و معکوس‌پذیری دارند، که این ویژگی‌ها در مدل‌سازی توزیع‌های پیچیده و ارزیابی احتمالات دقیق اهمیت زیادی دارند. برخلاف برخی مدل‌های دیگر، جریان‌های باقیمانده از تکنیک‌های نوآورانه‌ای برای کاهش پیچیدگی محاسباتی استفاده می‌کنند.

---

<sup>۱</sup> latent space

<sup>۲</sup> Residual flow

در ادامه، به بررسی ساختار این مدل‌ها و یکی از روش‌های کلیدی محاسباتی آنها می‌پردازیم تا نقش آنها در بهبود دقت و کارایی مدل‌های مبتنی بر جریان را بهتر درک کنیم.

---

#### زیربخش اول (۱۰ نمره)

با مراجعه به مقاله اصلی، ساختار این مدل‌ها و معماری توابع تبدیل آنها را توضیح دهید و نشان دهید این توابع معکوس پذیر بوده و چگونگی محاسبه معکوس آنها را شرح دهید.

---

#### زیربخش دوم (۵ نمره)

یکی از مقادیری که در کار با مدل‌های مبتنی بر جریان نیاز داریم، دترمینان ژاکوبین توابع تبدیل است. با مراجعه به [مقاله](#) Residual Flows، نحوه محاسبه دترمینان ژاکوبین توابع تبدیل آن را یافته و تکنیک به کار رفته در آن (Hutchinson trace estimator) را تحقیق کرده و به اختصار توضیح دهید.

## سوال دوم - مدل های مولد متخصص

### بخش اول- GAN

در شبکه های مولد متخصصانه (Generative Adversarial Network) ما شبکه ای را طراحی می کنیم که یک توزیع داده  $p_{data}(x)$  با بعد بالا را مدل سازی می کند .

برای این هدف، یک شبکه مولد تعریف می شود :

$$G_{\theta} : R^k \rightarrow R^n$$

نمونه ی تولید شده توسط مدل مولد را ابتدا با نمونه برداری از یک توزیع نرمال با بعد  $k$  ،  $z \sim N(0, I)$  و سپس استفاده از مدل مولد بر روی نویز  $G(z)$  به دست می آوریم. همچنین یک متمایز کننده تعریف می شود :

$$D_{\phi} : R^n \rightarrow (0, 1)$$

وظیفه این متمایز کننده ، تمایز بین تصاویر واقعی و تصاویر تولیدی توسط شبکه مولد است . به دلیل اینکه خروجی شبکه متمایز کننده به عنوان یک احتمال تفسیر می شود ، آخرین لایه این شبکه اغلب تابع sigmoid است .

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

لایه ی قبل از sigmoid را به عنوان logit های شبکه متمایز کننده در نظر بگیرید که خروجی این لایه با  $h(x)$  نمایش داده می شود .

متداول ترین روش آموزش شبکه های مولد به این صورت است که در آن به طور متناوب یک گام نزول گرادیان بر روی شبکه متمایز کننده با توجه به شبکه مولد و سپس یک گام نزول گرادیان بر روی شبکه مولد با توجه به شبکه متمایز کننده است .



$$\min_{\emptyset} L_D(\emptyset ; \theta)$$

$$\min_{\theta} L_G(\emptyset ; \theta)$$

تابع هزینه شبکه متمایز کننده به صورت زیر تعریف می شود :

$$L_D(\emptyset ; \theta) = -E_{x \sim p_{data}(x)}[\log D_{\emptyset}(x)] - E_{z \sim N(0,1)}[\log(1 - D_{\emptyset}(G_{\theta}(z)))]$$

تابع هزینه شبکه مولد نیز به صورت زیر تعیین می شود :

$$L_G(\emptyset ; \theta) = E_{z \sim N(0,1)}[\log(1 - D_{\emptyset}(G_{\theta}(z)))]$$

#### زیربخش اول (۵ نمره)

این نوع تابع هزینه که برای شبکه مولد تعریف شده است، دچار مشکل محو شدن گرادیان می شود.

نشان دهید مشتق تابع هزینه شبکه مولد نسبت به  $\theta$  به طور تقریبی برابر صفر است اگر در شبکه متمایز کننده داشته باشیم :  $D(G_{\theta}(z)) \approx 0$  یا به عبارت دیگر اگر داشته باشیم :  $h_{\emptyset}(G_{\theta}(z)) \ll 0$ .

#### زیربخش دوم (۵ نمره)

توزیع تولید شده توسط شبکه ی مولد را به صورت  $p_{\theta}(x)$  نمایش می دهیم . با استفاده از این توزیع داده می توان تابع هزینه شبکه متمایز کننده را به صورت زیر نمایش داد :

$$L_D(\emptyset ; \theta) = -E_{x \sim p_{data}(x)}[\log D_{\emptyset}(x)] - E_{x \sim p_{\theta}(x)}[\log(1 - D_{\emptyset}(x))]$$

کمترین مقدار تابع هزینه متمایز کننده زمانی رخ می دهد که داشته باشیم  $D_{\emptyset} = D^*$ ، که مقدار  $D^*$  به صورت زیر تعریف می شود :

$$D^*(x) = \frac{p_{data}(x)}{p_{\theta}(x) + p_{data}(x)}$$

#### زیربخش سوم (۳ نمره)

طبق تعریف برای شبکه متمایز کننده می دانیم که داریم :  $D_{\theta}(x) = \sigma(h_{\theta}(x))$  . نشان دهید که اگر داشته باشیم  $D^* = D_{\theta}$  , آنگاه در  $h_{\theta}(x)$  , در شبکه متمایز کننده رابطه زیر برقرار است :

$$h_{\theta}(x) = \log \frac{p_{data}(x)}{p_{\theta}(x)}$$

## بخش دوم – WASSERSTEIN GAN

در مدل های Wasserstein GAN(WGAN) برای حل مشکلات GAN, از معیار فاصله Wasserstein استفاده می کند.

این فاصله روشی برای اندازه گیری میزان تفاوت بین دو توزیع داده های واقعی و داده های تولید شده است. در این مدل، به جای اینکه شبکه متمایز کننده به سادگی واقعی یا جعلی بودن داده ها را پیش بینی کند، مدل WGAN تلاش می کند تا فاصله بین توزیع واقعی و توزیع داده های تولید شده را تخمین بزند.

فاصله ی Wasserstein بین دو توزیع به صورت زیر تعریف می شود :

$$W(P_r, P_{\theta}) = \sup_{\|f\|_L \leq 1} [E_{x \sim P_r}[f(x)] - E_{x \sim P_{\theta}}[f(x)]]$$

در معیار فاصله ی Wasserstein نمایش داده شده , منظور از  $p_r$  توزیع احتمال داده های واقعی و همچنین منظور از  $p_{\theta}$  , توزیع احتمال داده های تولید شده توسط شبکه مولد است.

### زیربخش اول ( ۵ نمره )

مدل WGAN چه مشکلاتی را که در GAN ها وجود دارد، برطرف کرده است؟ این مشکلات چگونه توسط WGAN حل شده اند؟

با توضیح مفاهیم و تکنیک های مورد استفاده در WGAN ، بیان کنید که این مدل چه تفاوت های اساسی با GAN های معمولی دارد که به بهبود نتایج منجر می شود.

---

#### زیربخش دوم (۴ نمره)

در مدل WGAN، عملکرد و نقش شبکه متمایزکننده چیست و چگونه با نقش آن در GAN ها تفاوت دارد؟

---

#### زیربخش سوم (۴ نمره)

در مدل WGAN، به جای استفاده از معیارهای Kullback-Leibler (KL) و Jensen-Shannon (JS) برای اندازه‌گیری تفاوت بین توزیع داده‌های واقعی و تولیدشده، از معیار فاصله Wasserstein استفاده می‌شود. دلیل استفاده از فاصله Wasserstein در WGAN چیست و این معیار چگونه نسبت به معیارهای KL و JS عملکرد بهتری ارائه می‌دهد؟

در پاسخ خود به مزایا و محدودیت‌های هر یک از این معیارها و تأثیر آن‌ها بر فرآیند آموزش مدل اشاره کنید.

---

#### زیربخش چهارم (۴ نمره)

چه معیارهایی برای ارزیابی همگرایی شبکه مولد WGAN وجود دارد و چگونه می‌توان اطمینان حاصل کرد که این شبکه به طور مؤثری به سمت توزیع مطلوب همگرا می‌شود؟

### زیر بخش سوم – پیاده سازی مدل های GAN

در این بخش، قصد داریم یک شبکه GAN را پیاده سازی کرده و بر روی داده های Fashion MNIST شبکه را آموزش دهیم.

---

#### زیربخش اول (۴ نمره)

یک کلاس برای پیاده سازی شبکه تعریف کنید و مدل مولد و مدل متمایز کننده را مطابق با ساختارهای ذکر شده در جداول پیاده سازی نمایید.

### معماری شبکه مولد

	layer	Input size	stride	padding	Kernel_size
1	ConvTranspose2d	(batch_size,latent_dim,1,1)	1	0	7
2	BatchNorm2d	(batch_size,128,7,7)	-	-	-
3	ReLU	(batch_size,128,7,7)	-	-	-
4	ConvTranspose2d	(batch_size,128,7,7)	2	1	4
5	BatchNorm2d	(batch_size,64,14,14)	-	-	-
6	ReLU	(batch_size,64,14,14)	-	-	-
7	ConvTranspose2d	(batch_size,64,14,14)	2	1	4
8	Sigmoid	(batch_size,1,28,28)	-	-	-

### معماری شبکه متمایز کننده

	layer	Input_size	Stride	padding	Kernel_size
1	Conv2d	(batch_size,1,28,28)	2	1	4
2	Leaky ReLU (0.2)	(batch_size,64,14,14)	-	-	-
3	Conv2d	(batch_size,64,14,14)	2	1	4
4	BatchNorm2d	(batch_size,128,7,7)	-	-	-

5	Leaky ReLU (0.2)	(batch_size,128,7,7)	-	-	-
6	Conv2d	(batch_size,128,7,7)	2	1	4
7	BatchNorm2d	(batch_size,256,3,3)	-	-	-
8	Leaky ReLU (0.2)	(batch_size,256,3,3)	-	-	-
9	Conv2d	(batch_size,256,3,3)	1	0	3
10	BatchNorm2d	(batch_size,512,1,1)	-	-	-
11	Leaky ReLU (0.2)	(batch_size,512,1,1)	-	-	-
12	Flatten	(batch_size,512,1,1)	-	-	-
13	Linear	(batch_size,512)	-	-	-
14	Sigmoid	(batch_size,1)	-	-	-

### زیربخش دوم (۳ نمره)

در مدل مولد از لایه ی ConvTranspose2d استفاده شده است . این لایه در شبکه های عصبی چه نقشی دارد؟ با ذکر جزئیات توضیح دهید که این لایه چگونه ابعاد داده ها را تغییر می دهد و پارامتر های آن چگونه بر خروجی تاثیر می گذارند .  
همچنین رابطه محاسبه ابعاد خروجی این لایه را بیان کنید.

### زیربخش سوم (۶ نمره)

شبکه را آموزش داده و در سه مرحله از آموزش ، ۲۵ نمونه تصویر تولید شده را نمایش دهید:

۱ - ایپاک ابتدایی

۲ - ایپاک میانی

۳ - ایپاک پایانی

سپس، تصاویر تولید شده در مراحل فوق را با تصاویر دادگان واقعی مقایسه کنید و تحلیل کنید که چگونه کیفیت تصاویر تولید شده با پیشرفت آموزش بهبود می یابد. برای آموزش شبکه از مقادیر زیر می توانید استفاده کنید:

<b>Latent dimension</b>	100
<b>Learning rate</b>	0.0002
<b>Batch size</b>	64
<b>epochs</b>	100

---

#### زیربخش چهارم ( ۴ نمره )

نمودارهای تابع هزینه شبکه مولد و شبکه متمایزکننده را در طول فرایند آموزش رسم کنید و نمایش دهید.

سپس، روند تغییرات این توابع هزینه را تحلیل کرده و توضیح دهید که چگونه این تغییرات نشان دهنده یادگیری و بهبود عملکرد شبکه ها هستند.

---

#### زیربخش پنجم ( ۳ نمره )

معیار Frechet Inception Distance (FID) چیست و چگونه برای ارزیابی کیفیت تصاویر تولید شده توسط GAN ها استفاده می شود؟

فرمول ریاضی FID را بیان کنید و توضیح دهید که چگونه این معیار فاصله بین توزیع های ویژگی تصاویر واقعی و تصاویر تولید شده را اندازه گیری می کند.

---

#### زیربخش ششم ( ۵ نمره )

معیار FID را برای ارزیابی کیفیت تصاویر تولید شده توسط شبکه مولد خود در ایپاک های مختلف (ایپاک های ابتدایی، میانی و پایانی) محاسبه کنید.

روند تغییرات FID را نمایش داده و تحلیل کنید که چگونه تغییرات این معیار نشان‌دهنده بهبود یا افت کیفیت تصاویر تولیدشده در طول آموزش است.



نمونه خروجی تولیدی به وسیله شبکه GAN بر روی داده های Fashion MNIST

## نکات تحویل

- مهلت ارسال این تمرین تا پایان روز "سه شنبه ۲۹ آبان 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.
- گزارش بایستی به زبان فارسی باشد.
- 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 ضمیمات مورد نیاز را با فرمت زیر در سامانه بارگذاری نمائید.
- HW2\_[Lastname]\_[StudentNumber].zip
- در صورت وجود سوال و یا ابهام می‌توانید از طریق رایانامه زیر با موضوع TAI\_HW2 با دستیاران آموزشی در ارتباط باشید:
  - سوال اول – بخش اول  
[fatemehnadir@gmail.com](mailto:fatemehnadir@gmail.com)
  - سوال اول – بخش دوم  
[erfanasgari21@gmail.com](mailto:erfanasgari21@gmail.com)
  - سوال دوم  
[amjadinasrin77@gmail.com](mailto:amjadinasrin77@gmail.com)

با آرزوی سلامتی و موفقیت روزافزون.