



- مهلت ارسال پاسخ تا ساعت ۵۹: ۲۳ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمارین تا سقف ۱۵ روز وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال‌شده با سیاست با تاخیر خواهند بود.
- 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.

مسئله‌ی ۱. (برآورد گرادیان تصادفی) (۱۵ نمره)

معمولا برای آموزش VAE از روشی به نام reparameterization trick برای برآورد یک گرادیان تصادفی استفاده می‌شود. این روش که گاهی اوقات pathwise derivative هم نامیده می‌شود، محدود به آموزش VAE نیست و در هر مسئله‌ای که نیاز به برآورد گرادیان یک امیدریاضی داشته باشیم می‌تواند مفید باشد. در واقع، مسئله برآورد گرادیان به علت ظاهر شدن در مسائل مختلف، یکی از مسائل بنیادی یادگیری تلقی می‌شود. در این تمرین قصد داریم با دو روش دیگر برای انجام اینکار آشنا شویم و واریانس گرادیان حاصل از این روش‌ها را باهم مقایسه کنیم. بدیهی است که هر قدر واریانس این برآوردگر کمتر باشد، آموزش VAE با batch های کوچکتری قابل انجام خواهد بود. با این انگیزه، مسئله‌ی برآورد گرادیان زیر را در نظر میگیریم:

$$\eta(\mu) = \nabla_{\mu} \mathbb{E}_{z \sim \mathcal{N}(\mu, 1/\sigma)} [f(z)] \quad (1)$$

در این نمادگذاری، امید نسبت به متغیر z محاسبه شده که این متغیر از توزیع نرمال $\mathcal{N}(\mu, 1/\sigma)$ پیروی میکند و $\mu \in \mathbb{R}$ پارامتری است که بنا داریم گرادیان یک تابع هزینه را نسبت به آن محاسبه کنیم. توجه کنید که چون گرادیان ∇_{μ} نسبت به μ یعنی یکی از پارامترهای توزیع z محاسبه می‌شود، نمی‌توانیم آن را داخل \mathbb{E}_z ببریم. همچنین دقت کنید سیستم‌های مشتق‌گیری خودکار کلاسیک نمیتوانند از یک تولیدکننده عدد تصادفی مشتق بگیرند.

الف با استفاده از reparameterization trick یک برآوردگر مونت‌کارلو برای η بدست آورده و آن را α بنامید. این کار را برای یکبار به ازای $f(z) = mz$ و یکبار برای $f(z) = (mz)^2$ انجام دهید.

ب برآوردگر دیگری به نام β به صورت زیر تعریف کنید:

$$\beta(\mu) = \frac{1}{N} \sum_{i=1}^N f(z_i) \nabla_{\mu} \log p(z_i; \mu) \quad z_i \sim \mathcal{N}(\mu, 1/\sigma)$$

در عبارت بالا p تابع چگالی متغیر $\mathcal{N}(\mu, 1/\sigma)$ است. این برآوردگر با نام های score function estimator یا REINFORCE شناخته می‌شود. این برآوردگر را از معادله‌ی ۱ بدست آورید.

پ برآوردگر سوم به نام γ را به صورت زیر تعریف میکنیم. این برآوردگر با نام های measure-valued estimator یا weak derivative شناخته می‌شود و می‌توان نشان داد که یک برآوردگر نارایب از η است، اما در این تمرین به دنبال

اثبات این نکته نیستیم.

$$\gamma(\mu) = \frac{1}{N\sqrt{2\pi}} \sum_{i=1}^N [f(\mu + y_i^+) - f(\mu - y_i^-)]$$

در اینجا متغیرهای y_i^+, y_i^- متغیرهای مستقل با توزیع Weibull با پارامترهای $\lambda = \sqrt{2}, k = 2$ هستند. به کمک شبیه سازی، مقدار این سه برآوردگر را به ازای مقادیر $m \in (-1/10, 1/10)$ و $\mu = 0.5$ یکبار برای $f(z) = mz$ و یکبار برای $f(z) = (mz)^2$ رسم و واریانس آن ها را در هر حالت با یکدیگر مقایسه کنید (محور افقی نمودار m و محور عمودی واریانس را نشان دهد). براساس این مشاهدات، به نظر شما کدام برآوردگر مناسب تر است؟ برای تولید یک نمونه مثل Y از توزیع Weibull میتوانید ابتدا یک متغیر $U \sim \text{Uniform}(0, 1)$ تولید و سپس از تبدیل $Y = \lambda(-\ln(U))^{1/k}$ استفاده کنید. پیاده سازی های آماده این توزیع در کتابخانه هایی مثل PyTorch وجود دارد، میتوانید در صورت ترجیح از آن ها استفاده کنید ولی مراقب تفاوت در قرارداد پارامترهای آن باشید.

نکته برای بخش سوم این تمرین، می توانید از زبان و پکیج های دلخواه خود استفاده کنید، نیازی به ارسال کد نیست. نمودار مقایسه و نتیجه گیری خود را ارسال کنید. اگر به درستی کد خود شک داشتید، مطمئن شوید میانگین این سه برآوردگر با یکدیگر متفاوت نباشد.

مسئله ۲. (یک واگرایی متفاوت) (۱۰ نمره)

یکی از عوامل موثر در کیفیت نتیجه ی VAE انتخاب مناسب توزیع پیشین $p(z)$ و توزیع وردشی $q(z|x)$ برای تخمین توزیع پسین است. توزیع گاوسی تنها یک مد دارد و ممکن است برای برخی از مسائل، استفاده از توزیع های منعطف تر منجر به بهبود کیفیت نتیجه شود. مثلاً خیلی خوب می شد اگر می توانستیم از یک Gaussian Mixture Model برای تعریف $p(z)$ یا $q(z|x)$ استفاده کنیم. مشکلی که وجود دارد این است که هنگام محاسبه ی ELBO، یافتن یک فرم بسته از واگرایی KL برای دو GMM ممکن نیست.

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

در این تمرین میخواهیم با یک واگرایی دیگر آشنا شویم که در این مورد بخصوص قادر به حل این مشکل است. واگرایی D_{CS} بین دو توزیع q و p را به صورت زیر تعریف میکنیم:

$$D_{\text{CS}}(p||q) = -\log \frac{\int p(x)q(x)dx}{\sqrt{\int p(x)^2 dx \int q(x)^2 dx}}$$

با وجود اینکه ظاهر آن پیچیده تر از D_{KL} است، استفاده از آن برای توزیع های گاوسی سخت نیست. الف حکم زیر در حالتی که p و q توزیع های نرمال چند متغیره باشند برقرار است.

$$D_{\text{CS}}(p||q) \leq \min\{D_{\text{KL}}(p||q), D_{\text{KL}}(q||p)\}$$

مقدار $D_{\text{CS}}(p||q)$ را در حالتی که هر دو توزیع q و p نرمال تک متغیره با واریانس یکسان σ^2 و میانگین های μ_1, μ_2 باشند بدست آورده و درستی نابرابری بالا را در این مورد تحقیق کنید.

با یک محاسبه سراسر میتوان نشان داد در حالتی که q و p هر دو GMM باشند، برخلاف $D_{\text{KL}}(q||p)$ یافتن فرم بسته برای $D_{\text{CS}}(p||q)$ امکان پذیر است که نیازی نیست آن را نشان دهید. با این انگیزه، میخواهیم با یک جایگزین برای ELBO آشنا شویم. بدین منظور تابع هدف زیر را در نظر بگیرید:

$$\mathcal{L}_{\text{CS}}(x; \theta, \phi) = \mathbb{E}_{q_\phi}[\log p_\theta(x|z)] - \lambda D_{\text{CS}}(q_\phi(z|x)||p(z))$$

در اینجا λ یک ضریب حقیقی است.

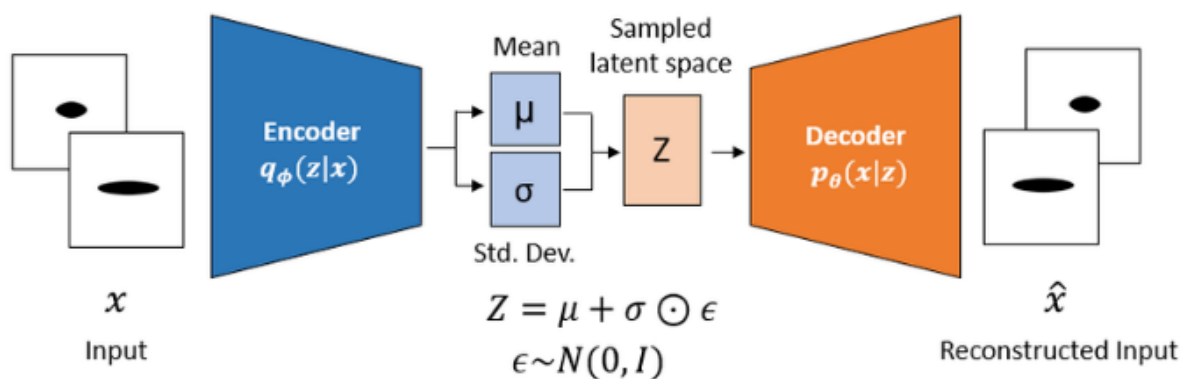
ب نشان دهید می توان \mathcal{L}_{CS} را به صورت زیر تجزیه کرد:

$$\mathcal{L}_{\text{CS}} = \log p(x) - D_{\text{KL}}(q(z|x)||p(z|x)) + D_{\text{KL}}(q(z|x)||p(z)) - \lambda D_{\text{CS}}(q(z|x)||p(z))$$

معنی هر یک از چهار مولفه‌ی این تجزیه را در مسئله VAE و اتفاقی که بر اثر بیشینه کردن \mathcal{L}_{CS} برای آن‌ها رخ می‌دهد را به صورت شهودی توضیح دهید. افزایش λ چه اثری دارد؟

مسئله‌ی ۳. (پیاده سازی β -VAE) (۲۵ نمره)

مدل β -VAE یک توسعه از مدل VAE است که در زمینه یادگیری ماشین برای یادگیری ویژگی‌های نهفته در داده‌ها استفاده می‌شود. هدف اصلی مدل بهبود جدا سازی در فضای نهفته است، به طوری که هر بعد از فضای نهفته نشان دهنده یک ویژگی مستقل در داده‌ها باشد. در مدل، یک پارامتر جدید به نام معرفی می‌شود که در تابع هزینه مدل استفاده می‌شود. این پارامتر به عنوان ضریب اصطلاح واگرایی KL -Divergence عمل می‌کند. با افزایش مقدار، مدل در فرایند بهینه سازی وزن‌ها، مجبور می‌شود تا توازن بین بازسازی داده‌ها و ساده سازی توزیع نهفته را تغییر دهد. این امر منجر به فشردگی بیشتر و جداسازی بهتر ویژگی‌ها می‌شود. در این تمرین قصد داریم یک شبکه مولد تصاویر fashion mnist آموزش دهیم.



پیش از ادامه‌ی سوال، مقاله‌ی β -VAE را مطالعه کنید.

(الف) یک کلاس طراحی کنید و مدل encoder و decoder را مطابق با جدول زیر پیاده سازی نمایید.

مدل encoder :

Padding	Stride	Kernel Size	Output Size	Type	Layer
۱	۲	۳	(size batch, ۸, ۱۴, ۱۴)	Conv2d	۱
-	-	-	(size batch, ۸, ۱۴, ۱۴)	relu	۲
۲	۱	۲	(size batch, ۱۶, ۷, ۷)	Conv2d	۳
-	-	-	(size batch, ۱۶, ۷, ۷)	relu	۴
-	-	-	(size batch, ۱۶, ۷, ۷)	Batchnorm2d	۵
۰	۲	۳	(size batch, ۳۲, ۳, ۳)	Conv2d	۶
-	-	-	(size batch, ۳۲, ۳, ۳)	relu	۷
-	-	-	(size batch, ۳۲ \times ۳ \times ۳)	Flatten	۸
-	-	-	(size batch, ۱۲۸)	Fully connected	۹

مدل decoder :

Padding	Stride	Kernel Size	Output Size	Type	Layer
-	-	-	(size batch, ۱۲۸)	Fully connected	۱
-	-	-	(size batch, ۱۲۸)	relu	۲
-	-	-	(size batch, ۳۲ × ۳ × ۳)	Fully connected	۳
-	-	-	(size batch, ۳۲ × ۳ × ۳)	relu	۴
۰	۲	۳	(size batch, ۱۶, ۷, ۷)	Convtranspose۲d	۵
-	-	-	(size batch, ۱۶, ۷, ۷)	relu	۶
-	-	-	(size batch, ۱۶, ۷, ۷)	Batchnorm۲d	۷
۱	۲	۳	(size batch, ۸, ۱۴, ۱۴)	Convtranspose۲d	۸
-	-	-	(size batch, ۸, ۱۴, ۱۴)	relu	۹
۱	۲	۳	(size batch, ۱, ۲۸, ۲۸)	Convtranspose۲d	۱۰
-	-	-	(size batch, ۱, ۲۸, ۲۸)	sigmoid	۱۱

(ب) تکنیک trick reparameterization را در کلاس مدل اضافه کرده و برای استفاده از آن، لایه های مورد نیاز را به encoder و decoder اضافه کنید. بعد فضای نهفته را برابر با ۱۰ در نظر بگیرید.

(پ) در مدل از لایه ی *ConvTranspose۲d* 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 پارامترهای آن چگونه بر خروجی تاثیر می گذارند. همچنین، رابطه محاسبه ابعاد خروجی این لایه را بیان کنید.

(ج) ابتدا تابع هزینه مورد استفاده در شبکه های VAE را پیاده سازی کنید. سپس با استفاده از این تابع هزینه، شبکه VAE را آموزش دهید. هاپیرپارامترها را به گونه ای تنظیم کنید که مدل آموزش ببیند. معیار خود برای انتخاب پارامتر β را توضیح دهید.

(د) پس از پیاده سازی و آموزش مدل VAE، ۱۰۰ نمونه از مجموعه داده های تست را به مدل وارد کرده و نمونه های بازسازی شده را نمایش دهید. علاوه بر این، ۱۰۰ نویز تصادفی را به دیکودر مدل اعمال کرده و نمونه های تولید شده را به نمایش بگذارید. از لحاظ بصری، تفاوت های موجود بین این دو مجموعه نمونه های بازسازی شده و تولید شده را تحلیل و توضیح دهید تا قابلیت های مدل در بازسازی و تولید داده های جدید به طور جامع تری روشن شود.

(ر) به طور جامع توضیح دهید که معیار FID چگونه محاسبه می شود و چه اطلاعاتی را درباره تصاویر تولید شده توسط مدل های VAE ارائه می دهد؟

(ه) پس از پیاده سازی و آموزش مدل VAE، مجموعه داده های تست را به مدل وارد کرده و نمونه های بازسازی شده را استخراج نمایید. سپس معیار FID را بین دیتاست تست و نمونه های بازسازی شده محاسبه کنید تا کیفیت بازسازی مدل را ارزیابی نمایید. علاوه بر این، به تعداد داده های موجود در مجموعه تست، نویز تصادفی را به دیکودر مدل ارائه داده و نمونه های تولیدی را تولید کنید. معیار FID را نیز بین داده های تولید شده و داده های تست محاسبه نمایید تا قابلیت مدل در تولید داده های جدید مورد سنجش قرار گیرد.

(ع) چه معیارهای جایگزینی می توانند برای ارزیابی مدل های VAE استفاده شوند و چگونه با FID مقایسه می شوند؟

مسئله ۴. تصدانه دو تصدانه (pix2pix) (۲۰ نمره)

(الف) در مورد معماری شبکه های conditional generative adversarial network (CGAN) توضیح دهید.

(ب) image translation یک حوزه مهم از بینایی ماشین است که در آن یک نگاشت از تصویر اولیه به تصویر خروجی مدل در فضاهای مختلف بدست می آوریم. سه مورد از کاربردهای image translation را توضیح دهید و برای هر کاربرد حداقل به یک مقاله که در آن حوزه از image translation استفاده کرده اند ارجاع دهید.

ج) در حوزه شبکه های مولد از دو مدل Pix2Pix و Cycle GAN می توان به عنوان دو مدل مطرح برای im-age translation نام برد در مورد این دو مدل توضیح داده و تفاوت های آن ها را در نوع کاربرد، generator و discriminator آنها توضیح دهید.

د) در این قسمت می خواهیم به بررسی دقیق تر مدل Pix2Pix بپردازیم برای این کار ابتدا **مقاله اصلی این مدل** را مطالعه کرده و در ادامه به سوالات مطرح شده پاسخ دهید

ه) بیان کنید که تفاوت های اصلی این مدل با یک مدل GAN ساده در چیست؟ (از منظر مفهومی و نه جزئیات معماری)

و) تفاوت بخش Discriminator این مدل را با Discriminator یک شبکه GAN عادی را به همراه شکل بیان کنید همچنین بیان کنید این تفاوت چگونه می تواند در راستای بهبود کلی عملکرد مدل عمل کند؟

ز) بخش Generator این مدل از معماری U-Net استفاده می کند همراه با شکل جزئیات این بخش را شرح دهید (نحوه کارکرد، اجزای مختلف و ابعاد ورود و خروجی هر بلاک) همچنین بیان کنید که کارکرد DropOut در تنوع بخشیدن به نتایج چیست؟

ح) تابع هزینه نهایی این معماری را بر اساس مقاله تفسیر کنید و توضیح دهید که با توجه به تابع هزینه بروزرسانی وزن های هر زیربخش از معماری کلی چطور انجام می گیرد؟

مسئله ۵. تصدانه دو تصدانه (pix2pix) در پیاده سازی (۳۰ نمره)

در این قسمت می خواهیم به پیاده سازی مدل Pix2Pix پرداخته و از آن برای تبدیل تصاویر بخش بندی شده (seg-mented image) که توسط خودروهای خودران تهیه شده به تصاویر حقیقی از محیط استفاده کنیم یک نمونه از این تناظر را می توانید در شکل ۱ مشاهده کنید



شکل ۱: تصویر نمونه برای تناظر ورودی و خروجی مدل

شما می توانید مثال و خروجی ای تعاملی از مدل Pix2Pix را **در این لینک** ببینید

نکته ۱: برای پیاده سازی این تمرین توصیه می شود از Kaggle یا Google Colab استفاده کنید این دو سامانه با تخصیص منابع پردازشی مورد نیاز نظیر GPU برای پردازش های مورد نیاز این سوال کافی بوده و در ارزیابی انجام شده محدودیت پردازشی ای برای حل این سوال نخواهید داشت.

نکته ۲: استفاده مدل های زبانی بزرگ و هوشمند در این تمرین مجاز بوده اما می بایست قسمت هایی که از مدل زبانی بزرگ استفاده شده به انضمام دستور (Prompt) استفاده شده در گزارش ذکر گردد. لازم به ذکر است بخش اصلی ایده و پیاده سازی می بایست حاصل تفکر و تلاش خود شما باشد و در صورتی که به صورت کامل از مدل های زبانی بزرگ برای پاسخ دهی استفاده کنید درصد پایینی از نمره سوال به شما تعلق می گیرد.

نکته ۳: برای کسب نمره کامل در قسمت های مختلف این سوال نیاز است تا notebook شامل کد و نتایج خود را آپلود نمایید اما توجه فرمایید تنها کدنویسی به تنهایی کافی نبوده و نمره اصلی سوالات به نتایج و تحلیل های شما در گزارش تعلق می گیرد.

الف) ابتدا دیتاست مد نظر این سوال را از طریق **این لینک** دریافت و به صورت کامل (هر دو پوشه train و val) را بارگذاری و مجموعه دادگان آموزش و اعتبار سنجی را بسازید و سپس سه تصویر از مجموع آموزشی (هم تصویر بخش بندی شده و هم تصویر واقعی متناظر) را نمایش دهید.

ب) با توجه به بخش (و) در سوال قبل کلاس Discriminator خود را تعریف کنید. همچنین با توجه به قسمت (ز) سوال قبل کلاس های DownSample و UpSample و در انتها کلاس Generator که خود متشکل از دو جزء قبلی است را پیاده سازی کنید.

ج) با توجه به بخش (ح) سوال قبل، توابع هزینه مدل را پیاده سازی کنید.

د) مدل را ۵۰ دوره (Epoch) آموزش دهید و پس از هر ده دوره آموزش سه مجموعه خروجی متفاوت از تصویر انتزاعی (بخش بندی شده - Segmented) تصویر تولید شده توسط مدل و تصویر حقیقی متناظر با تصویر انتزاعی را نمایش دهید و در نهایت نتیجه نهایی پس از آموزش کامل را در گزارش خود بیاورید.

ه) نمودار توابع هزینه Discriminator و Generator را رسم و تحلیل نمایید.

و) مقادیر ابرپارامترهای (Hyper-Parameters) مورد استفاده در مدل را ذکر نمایید.

ز) (نمره امتیازی) قسمت اصلی این سوال بابت ارسال notebook و تحلیل های شما در گزارش است

اما در این قسمت برای دریافت نمره امتیازی می توانید کد نوشته شده را به صورت Public **در قسمت کد** های مربوط به دیتاست در وب سایت Kaggle به اشتراک بگذارید و لینک و تصویر این فعالیت را در گزارش خود قرار دهید.

همچنین نیاز است تا کد مربوطه را در **صفحه مربوط به مقاله مورد بررسی در وب سایت** paperswithcode نیز به اشتراک بگذارید

در نهایت نیز می توانید به عنوان گزارش نهایی در وب سایت Medium گزارش کامل فعالیت خود را (به زبان انگلیسی) قرار دهید و به جای گزارش می توانید لینک توضیحات و گزارش کامل خود در Medium را در گزارش قرار دهید. برای دریافت نمره امتیازی این بخش انجام هر سه کار لازم است و می بایست لینک و تصویر فعالیت های انجام شده در گزارش قرار بگیرند. (انجام این قسمت علاوه بر دریافت نمره امتیازی برای تبدیل این تمرین به یک فعالیت علمی معتبر و ماندگار نیز توصیه می گردد.)

موفق باشید (:)