**به نام خدا**

**دانشگاه صنعتی شریف**
**دانشکده‌ی مهندسی برق**

# یادگیری عمیق

پاییز ۱۴۰۱
استاد: دکتر فاطمی‌زاده

گردآورندگان: علی مجلسی، امید جفائی، علیرضا عباسیان

تمرین پنجم | VAE & AE | مهلت ارسال: سه شنبه ۲۵ دی ماه

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص‌شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمارین تا سقف ۵ روز و در مجموع ۱۸ روز، وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال‌شده پذیرفته نخواهند بود. همچنین، به ازای هر روز تأخیر 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. (دقت کنید در صورت تشخیص مشابهت غیرعادی برخورد جدی صورت خواهد گرفت.)
- 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.
- نتایج و پاسخ های خود را در یک فایل با فرمت zip به نام HW۵-Name-StudentNumber در سایت CW قرار دهید. برای بخش عملی تمرین نیز در صورتی که کد تمرین و نتایج خود را در گیت‌هاب بارگذاری می‌کنید، لینک مخزن مربوطه (repository) را در پاسخنامه خود قرار بدهید. دقت کنید هر سه فایل نوتبوک تکمیل شده بخش عملی را در گیت‌هاب قرار دهید. همچنین لازم است تا دسترسی های لازم را به دستیاران آموزشی مربوط به این تمرین بدهید.
- لطفا تمامی سوالات خود را از طریق صفحه درس در سایت Quera مطرح کنید (برای اینکه تمامی دانشجویان به پاسخ‌های مطرح شده به سوالات دسترسی داشته باشند و جلوی سوالات تکراری گرفته شود، به سوالات در بسترهای دیگر پاسخ داده نخواهد شد).
- دقت کنید کدهای شما باید قابلیت اجرای دوباره داشته باشند، در صورت دادن خطا هنگام اجرای کدتان، حتی اگه خطا بدلیل اشتباه تایپی باشد، نمره صفر به آن بخش تعلق خواهد گرفت.

## سوالات نظری (۱۰۰ نمره)

۱. (۲۰ نمره) در این سوال قصد داریم به تفاوت VAE و AE بپردازیم.

(آ) گمان کنید برای تولید دیتای مشابه دیتاست می‌خواهیم از یک AE معمولی استفاده کنیم. یک AE را آموزش داده‌ایم و یک نقطه رندوم (با توزیع یونیفرم) در فضای نهان انتخاب کرده و آن را وارد ماژول دیکودر آموزش دیده می‌کنیم، به نظر شما احتمال اینکه خروجی دیکودر، شبیه به دیتاست باشد بیشتر است یا احتمال اینکه یک نقطه تصادفی در فضای دیتا انتخاب کنیم و شبیه دیتاست بشود؟ چرا؟ (۵ نمره)

(ب) حداقل سه اشکال روش قسمت (آ) برای تولید دیتا شبیه به دیتاست را بیان کنید و بگویید VAE چگونه این مشکلات را رفع می‌کنند. (۵ نمره)

(ج) فرض کنید در حین فرایند آموزش AE به خروجی انکودر، یک نویز گوسی با میانگین صفر و واریانس $R \times ۰/۰۵$ اضافه کنیم. منظور از $R$ میانگین مربعات فاصله نقاط فضای نهان از مرکز آن نقاط است که در هر گام آموزش به روزرسانی می‌شود. آیا دیکودر آموزش دیده در این روش، عملکرد بهتری نسبت به دیکودر AE معمولی دارد؟ منظور این است که اگر یک نقطه از فضای نهان به صورت تصادفی انتخاب شود، خروجی کدام یک محتمل‌تر است که شبیه دیتاست باشد. (۵ نمره)

(د) آیا VAE مزیتی نسبت به روش مطرح شده در (ج) دارد؟ تفاوت کلیدی این دو روش چیست؟ (۵ نمره)

۲. (۳۰ نمره)

در این تمرین قصد داریم با تخمین ML و ارتباط آن با VAE بیشتر آشنا شویم.

(آ) فرض کنید دیتاست ما به صورت $D = \{x_۱, x_۲, \cdots, x_n\}$ است. درباره تخمین maximum likelihood مطالعه کنید و توضیح دهید چرا پارامترهای توزیع باید به گونه‌ای باشد که رابطه زیر را بیشینه کند.

$$\sum_{i=۱}^{n} log(p_\theta(x_i))$$

توجه کنید که منظور از $p_\theta(x_i)$ این است که احتمال اینکه در خروجی $x_i$ را ببینیم تابعی از پارامترهای $\theta$ است. (۵ نمره)

(ب) هم ارزی کمینه کردن خطای cross entropy و تخمین ML را نشان دهید. (۵ نمره)

(ج) می‌دانیم که هدف نهایی از VAE این است که مدل مولدی داشته باشیم که توزیع خروجی آن شبیه به توزیع دیتاست باشد. در VAE شبیه به شبکه های معمولی میخواهیم از stochastic gradient descent استفاده کنیم! بنابراین در فرایند یادگیری به جای آنکه کل دیتاست را یک‌جا ببینیم و $\sum_{i=۱}^{n} log(p_\theta(x_i))$ را بیشینه کنیم، بعد از اعمال هر ورودی، سعی در تغییر پارامترهای شبکه داریم به گونه ای که ذره ای $log(p_\theta(x_i))$ بیشتر شود. یعنی در اعمال هر ورودی، هدف این است که ذره ای احتمال تولید خروجی‌ای شبیه به آن ورودی، بیشتر شود. در طی درس دیدید که این لگاریتم احتمال در معادله ELBO صدق می‌کند که در زیر آورده شده است.

$$\log p_\theta(x_i) - D_{KL}[q_\phi(z|x_i)||p_\theta(z|x_i)] = \mathbb{E}_z\left[\log p_\theta(x_i|z)\right] - D_{KL}[q_\phi(z|x_i)||p_\theta(z)] \quad (۱)$$

در رابطه (۱) منظور از $\theta$ پارامترهای دیکودر و منظور از $\phi$ پارامترهای انکودر است.

i. ثابت کنید که فاصله KL نامنفی است و بدین ترتیب کران پایینی برای لگاریتم احتمال بیابید. (۵ نمره)

ii. توجیه کنید که چرا در بسیاری از پیاده‌سازی‌های VAE نظیر عبارت $\mathbb{E}_z\left[\log p_\theta(x_i|z)\right]$ خطای cross entropy بین تصویر دیتاست و تصویر خروجی دیکودر را کمینه می‌کنند؟ (۱۵ نمره)

۳. (۲۰ نمره) چرا در VAE فرض میکنند توزیع فضای نهان گوسی است؟ (به مواردی به جز ساده شدن محاسبات اشاره کنید) تحقیق کنید که آیا در عمل به جز گوسی، از توزیع های دیگری نیز استفاده می‌کنند؟

۴. (۳۰ نمره) مقاله $\beta$ - VAE [۱] را مطالعه کنید و به سوالات زیر پاسخ دهید.

(آ) به صورت خلاصه ایده‌ی $\beta$ - VAE را توضیح دهید و تفاوت آن را با VAE بیان کنید. (۱۵ نمره)

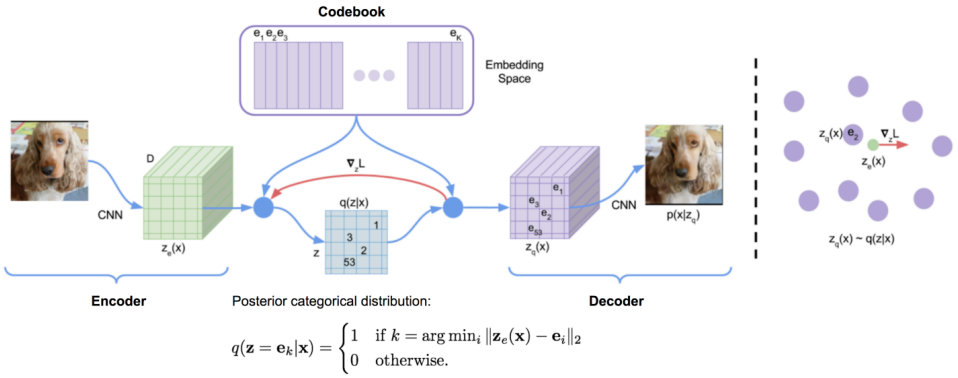
(ب) بر اساس اطلاعات موجود در Section 2 این مقاله، اهمیت و کارکرد disentanglement metric را شرح دهید. (۱۵ نمره)

[۱] Irina Higgins, Lo¨ ıc Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. ''beta-vae: Learning basic visual concepts with a constrained variational framework.'' In 5th International Conference on Learning Representations, ICLR 2017, 2017.

## سوالات عملی (۲۰۰ نمره)

۱. (۱۰۰ نمره) لطفا نوتبوک Q1_VAE_CVAE را کامل کنید. در این نوتبوک دو ساختار Variational Auto Encoder (VAE) و Conditional Variational Auto Encoder (CVAE) پیاده سازی می شوند.

۲. (۱۰۰ نمره) فایل نوتبوکی در اختیار شما قرار داده شده است که شامل راهنمایی‌های لازم برای انجام تمرین می‌باشد. در این تمرین، هدف شما این است که مدل VQ-Vecor Quantized Variational Autoencoder (VQ-VAE) را با استفاده از ساختار و مراحل مطرح شده در مقاله‌ی Neural Discrete Representation Learning تکمیل کنید. پیش از تکمیل نوت بوک لطفا به پرسش های زیر پاسخ دهید. هچنین در ویدیو این لینک توضیحات خوبی راجع به این مقاله داده شده است.




شکل ۱: ساختار کلی VQ-VAE

(آ) تابع هزینه VQ-VAE در معاله ۲ نشان داده شده است. توضیح دهید هر کدام از سه ترم این تابع هزینه چه معنایی دارد.

$$\mathcal{L} = \underbrace{\log(p(x \mid z_q(x)))}_{(۱)} + \underbrace{\|z_e(x) \cdot detach() - e\|}_{(۲)} + \beta \underbrace{\|z_e(x) - e \cdot detach()\|}_{(۳)} \quad (۲)$$

**توضیح اضافه راجع به تابع detach** : در مقاله گفته شده است که در فرآیند یادگیری، گرادیان های بعد از کدبوک مستقیما به قبل از کدبوک کپی می شوند و نیازی به محاسبه مشتق فرآیند (تابع) خود کدبوک نیست. می توان این فرآیند کپی کردن گرادیان ها را با رابطه زیر نوشت:

$$B = A + (f(A) - A).detach() \quad (۳)$$

متد detach() باعث می شود که متغیر مورد نظر به هنگام گرفتن مشتق، عدد ثابت حساب شود که یعنی در فرآیند back-propagation تاثیری نداشته‌باشد. در رابطه بالا A ورودی قبل از کد بوک، B خروجی کد بوک و تابع f(.) خود فرآیند (تابع) کدبوک یا همان Quantization Vector است.

(ب) چگونه در این مدل بردار ها کوانتیزه می شوند. راجع به codebook و نحوه یاگیری آن به صورت مختصر توضیح دهید. همچنین توضیح دهید بردار های codebook به صورت شهودی چه چیزی را نشان می دهند.