



- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- 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.

## VAE (۹۵ نمره)

### مسئله ۱. (۱۰ نمره)

(آ) یکی از راه‌های معمول برای تخمین گرادیان یک امید ریاضی، استفاده از رابطه‌ی زیر است:

$$\nabla_{\theta} \mathbb{E}_{z \sim q_{\theta}(z)} [f(z)] = \frac{1}{N} \sum_{i=1}^N f(z^i) \cdot \nabla_{\theta} \ln q_{\theta}(z^i)$$

که در آن هر  $z^i$  نمونه‌ی مستقلی از توزیع  $q_{\theta}(z)$  می باشد. درستی این رابطه را نشان دهید و بیان کنید که چگونه می توانیم از آن در VAE استفاده کنیم. با مراجعه به این مقاله مشکلی که در استفاده از این روش وجود دارد را بیان کنید.

(ب) به صورت شهودی بیان کنید که روش Reparameterization چگونه می تواند این مشکل را حل کند؟

(ج) در بسیاری از موارد تابع خطای رمزگشای VAE را خطای MSE در نظر میگیریم. این در حالی است که هدف ما بیشینه کردن تابع  $\mathbb{E}_{z \sim q_{\theta}(z|x)} \ln p_{\theta}(x|z)$  می باشد. در چه صورتی و با چه فرض هایی این دو کار معادل یکدیگر هستند؟

### مسئله ۲. (۱۰ نمره)

(آ) در یک VAE اگر داده ورودی از نوع باینری باشد (تصویری با پیکسل های ۰ و ۱)، می توان 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 تابع فعال سازی آخرین لایه کدگشا را sigmoid در نظر بگیرید. اثبات کنید که در تابع هزینه این شبکه یک جمله ای به شکل Binary Cross Entropy ظاهر می شود.

(ب) تکنیک Reparameterization روی بسیاری از توزیع های پیوسته قابل اعمال است. تحقیق کنید که چگونه می توان از این تکنیک برای یک توزیع categorical استفاده کرد؟

(ج) یکی از نسخه های تغییر یافته VAE مقاله مربوط به beta-VAE می باشد. در این روش یک ضریب beta ای پشت یکی از توابع هزینه قرار میگیرد. درباره این روش تحقیق کنید و بیان کنید:

- (۱) با انجام چه روندی از محاسبات، این ضریب در تابع هدف این روش ظاهر می‌شود؟
- (۲) هدف از افزودن ضریب beta چیست و اضافه کردن آن چه تاثیری روی ویژگی‌های فضای نهان یادگرفته شده توسط مدل دارد؟

## GAN (۹۵ نمره)

### مسئله‌ی ۳. (۲۰ نمره)

برای متغیرهای تصادفی دلخواه  $X$  و  $Y$  عبارات زیر را اثبات نمایید:

$$(آ) \quad \mathbb{E}[X|Y=y] = \mathbb{E}[X] : Y \text{ و } X \text{ با فرض مستقل بودن}$$

$$(ب) \quad \mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$$

$$(ج) \quad \text{var}(X) = \mathbb{E}[\text{var}(X|Y)] + \text{var}[\mathbb{E}(X|Y)]$$

$$(د) \quad \mathbb{E}[XY] = \mathbb{E}[Y\mathbb{E}[X|Y]]$$

### مسئله‌ی ۴. (۲۰ نمره)

با فرض اینکه  $A$  یک ماتریس  $n \times n$  معکوس پذیر می باشد، موارد زیر را ثابت کنید.

(آ) اگر  $\lambda$  یکی از مقدار ویژه های ماتریس  $A$  باشد،  $\frac{1}{\lambda}$  نیز یکی از مقدار ویژه های ماتریس  $A^{-1}$  با همان بردار ویژه خواهد بود.

(ب) اگر  $A$  یک ماتریس مثلثی باشد، عناصر روی قطر ماتریس مقادیر ویژه آن خواهند بود.

(ج) ضرب مقدار ویژه های ماتریس  $A$  برابر دترمینان ماتریس  $A$  خواهد بود.

(د) جمع مقدار ویژه های ماتریس  $A$  برابر  $\text{trace}(A)$  خواهد بود.

### مسئله‌ی ۵. (۱۵ نمره)

با فرض اینکه  $a$  و  $x$  بردارهای ستونی و  $A$  ماتریس مربعی باشد، روابط زیر را برای مشتقات ماتریسی اثبات کنید.

$$(آ) \quad \frac{da^T x}{dx} = a^T$$

$$(ب) \quad \frac{dx^T A x}{dx} = x^T (A + A^T)$$

$$(ج) \quad \frac{d(x^T a)^2}{dx} = 2x^T a a^T$$

### مسئله‌ی ۶. (۲۰ نمره)

در این سوال، هدف اصلی محاسبه MLE و MAP برای میانگین یک توزیع نرمال تک متغیره است. در این سوال ما فرض می کنیم که  $N$  نمونه  $x_1, \dots, x_N$  به طور مستقل از یک توزیع نرمال با واریانس معلوم  $\sigma^2$  و میانگین نامعلوم  $\mu$  گرفته شده اند.

۱. تخمین گر MLE برای  $\mu$  را محاسبه کنید. تمامی محاسبات مورد استفاده را به طور دقیق ذکر کنید.
۲. حال تخمین گر MAP را برای  $\mu$  بدست آورید. برای این بخش فرض کنید که توزیع پیشین (Prior) برای  $\mu$  خود یک توزیع نرمال با میانگین  $\nu$  و واریانس  $\beta^2$  باشد. تمامی محاسبات خود را به طور دقیق ذکر کنید. راهنمایی: می توانید از این نکته استفاده کنید که:

$$\beta^2 \left( \sum_{i=1}^N (x_i - \mu)^2 \right) + \sigma^2 (\mu - \nu)^2$$

$$= \left[ \mu \sqrt{N\beta^2 + \sigma^2} - \frac{\sigma^2 \nu + \beta^2 \sum_{i=1}^N x_i}{\sqrt{N\beta^2 + \sigma^2}} \right]^2 - \frac{[\sigma^2 \nu + \beta^2 \sum_{i=1}^N x_i]^2}{N\beta^2 + \sigma^2} + \beta^2 \left( \sum_{i=1}^N x_i^2 \right) + \sigma^2 \nu^2$$

البته توجه کنید که بدون استفاده از این راهنمایی هم امکان حل سوال وجود دارد.

## سوالات عملی (۱۵ نمره)

### مسئله ۷. (۱۵ نمره)

در این سوال می خواهیم تاثیر تعداد داده های نمونه برداری شده را بر مقادیر بدست آمده توسط تخمینگرهای MLE و MAP بررسی کنیم.

یک توزیع به فرم گاوسی تک متغیره با پارامترهای  $x \sim N(\mu = 2, \sigma^2 = 5)$  را در نظر بگیرید که داده های موجود از این توزیع به صورت i.i.d نمونه برداری می شوند. همچنین فرض کنید توزیع پیشین  $\mu$  یک توزیع گاوسی با پارامترهای  $N(1, 2)$  می باشد.

به ترتیب ۱۰، ۱۰۰ و ۱۰۰۰ نمونه از توزیع گاوسی اولیه نمونه بردارید. در هر حالت با کمک داده های نمونه برداری شده و توزیع پیشین بالا و روابط بدست آمده در مساله قبل، با استفاده از تخمین گرهای MLE و MAP مقدار  $\mu$  را تخمین بزنید.

نتایج بدست آمده را تفسیر کنید. (به دلیل ذات تصادفی بودن نمونه های برداشته شده، برای تفسیر دقیق تر توصیه می شود چندبار عمل بالا را تکرار نمایید.)

موفق باشید (:)