



## مسئله ۱. کاربرد شبکه‌های بازگشتی (۳+۶ نمره)

الف) شبکه‌های عصبی بازگشتی در کاربردهای مختلفی در زمینه پردازش رشته‌ها به کار می‌روند. در یک دسته‌بندی کلی این کاربردها را به دسته‌های یک-به-چند (one-to-many)، چند-به-چند (many-to-one) و چند به چند (many-to-many) تقسیم می‌کنند. همچنین رشته‌های مورد استفاده در آخرین دسته ممکن است از نظر زمانی و ترتیبی، همگام یا غیرهمگام باشند. در مورد هر یک از این دسته‌ها توضیح دهید و از کاربردهای هریک لااقل یک مثال معرفی کنید. (۳ نمره)

ب) مسئله seq2seq در کدام یک از دسته‌های معرفی شده قرار می‌گیرد؟ معماری شبکه عصبی encoder-decoder که برای حل این مسئله به کار می‌رود را توضیح دهید. (۳ نمره)

ج) مکانیزم توجه (attention) چیست و چگونه می‌تواند به شبکه‌های encoder-decoder کمک کند. در این مورد توضیح دهید. (۳ نمره)

## مسئله ۲. آموزش شبکه‌های بازگشتی (۹ نمره)

الف) می‌دانیم که الگوریتم backpropagation به صورت گسترده برای آموزش شبکه‌های عصبی feed-forward استفاده می‌شود. در مورد شبکه‌های بازگشتی از الگوریتم مشابهی به نام backpropagation through time استفاده می‌شود. در مورد این الگوریتم توضیح دهید و تفاوت آن را با backpropagation عادی بیان کنید. (۳ نمره)

ب) یکی از مشکلات جدی در آموزش شبکه‌های عصبی بازگشتی مشکل vanishing gradient و explosion می‌باشد. این دو مشکل را شرح دهید و توضیح دهید که چرا این مشکلات در شبکه‌های عصبی بازگشتی شدیدتر هستند. (۳ نمره)

پ) توضیح دهید که معماری LSTM چگونه می‌تواند به عنوان راه حلی برای رفع دو مشکل ذکر شده به کار گرفته شود. (۳ نمره)

## مسئله ۳. Exposure Bias (۵ نمره)

شبکه‌های تکرار شونده به طور کلی با هدف مدل کردن یک دنباله آموزش می‌یابند. این شبکه‌ها معمولاً از دو روند متفاوت برای فاز آموزش و تست استفاده می‌کنند. به این صورت که در فاز آموزش، خروجی/برچسب درست در مرحله  $t-1$  از مجموعه دادگان گرفته می‌شود و به عنوان ورودی گام  $t$ ام استفاده می‌شود. به عبارت دیگر شبکه در هر گام، خروجی درست مرحله قبل را از مجموعه دادگان دریافت می‌کند. از طرفی در فاز تست از آن جایی که طبیعتاً مجموعه دادگانی در دسترس نیست، خروجی گرفته شده از خود مدل در مرحله  $t-1$  به عنوان ورودی گام  $t$  در نظر گرفته می‌شود. این تفاوت بین دو فاز باعث به وجود آمدن مشکلی به نام exposure bias می‌شود. این مشکل را توضیح دهید و یک مثال از حالتی که این مشکل می‌تواند کیفیت خروجی مدل را به شدت کاهش دهد بیاورید. به نظر شما چرا نمی‌شود از روندی که در فاز تست وجود دارد در فاز آموزش هم استفاده کنیم؟

#### مسئله ۴. شبکه مولد متخاصمی (۱۵ نمره)

تابع هدف شبکه مولد متخاصمی به صورت زیر است

$$\mathcal{V}(G, D) = \mathbb{E}_{x \sim p_{data}}[\log D(x)] + \mathbb{E}_{x \sim p_g}[\log(1 - D(x))] \quad (۱)$$

که در آن  $G$  شبکه مولد و  $D$  شبکه تمییزدهنده است و می‌خواهیم آن را نسبت به  $D$  بیشینه و نسبت به  $G$  کمینه کنیم. حال با توجه به توضیحات داده شده به سوالات زیر پاسخ دهید.

(الف) با فرض نامحدود بودن ظرفیت تمییزدهنده، تمییزدهنده بهینه ( $D^*$ ) را بدست آورید. (۵ نمره)

(ب) حال با فرض بهینه بودن تمییزدهنده، نشان دهید کمینه کردن تابع هزینه GAN معادل کمینه کردن فاصله Jensen Shannon بین توزیع دادگان آموزش و توزیع شبکه مولد است. (۵ نمره)

(ج) حال با توجه به قسمت قبل نشان دهید که  $p_g \approx p_{data}$  یک نقطه بهینه برای تابع هدف می‌باشد. (۵ نمره)

#### مسئله ۵. چسبندگی قله (۵ نمره)

آموزش شبکه‌های مولد متخاصمی با چالش‌هایی همراه است. یکی از این چالش‌ها مشکل Mode Collapse است. این مشکل را شرح دهید و توضیح دهید چرا با این مشکل در آموزش این شبکه‌ها مواجه هستیم.

#### مسئله ۶. اثر تابع هزینه متفاوت (۱۴ نمره)

در این قسمت می‌خواهیم تاثیر استفاده از توابع هدف متفاوت را برای آموزش ببینیم.

(۱) مشکل استفاده از  $\mathbb{E}_{x \sim p_g}[\log(1 - D(x))]$  به عنوان تابع هدف شبکه مولد که می‌خواهیم آن را کمینه کنیم را توضیح دهید و توضیح دهید چگونه استفاده از تابع هدف  $\mathbb{E}_{x \sim p_g}[\log(D(x))]$  که می‌خواهیم آن را بیشینه کنیم مشکل ذکر شده را برطرف می‌کند. (۵ نمره)

(۲) فرض کنید که دامنه توزیع اصلی و توزیع شبکه مولد با هم همپوشانی نداشته باشند و تمییزدهنده نیز نزدیک به تمییزدهنده بهینه باشد. با فرض‌های مطرح شده به سوال‌های زیر پاسخ دهید.

(الف) تمییزدهنده را به صورت  $\sigma(a)$  در نظر بگیرید که در آن  $a$ ، logits شبکه است. حال گرایان  $\log(1 - D(x))$  را نسبت به logits های شبکه  $D$  بدست آورید. (  $a$  را به صورت تابعی از ورودی یعنی  $f(x)$  در نظر بگیرید) (۳ نمره)

(ب) در این حالت چه گرایانی به شبکه مولد می‌رسد؟ چه مشکلی ایجاد می‌شود؟ (۳ نمره)

(ج) حال اگر از تابع  $-\log D(x)$  به عنوان هدف شبکه مولد استفاده شود آیا باز هم این مشکل وجود دارد؟ (۳ نمره)

#### مسئله ۷. مشکل تابع فاصله JS (۷ نمره)

(امتیازی) همانطور که می‌دانیم، در شبکه‌های مولد متخاصمی می‌خواهیم فاصله Jensen Shannon را کمینه کنیم و برای این کار از روش Gradient Descent استفاده می‌کنیم. حال توضیح دهید اگر توزیع دادگان آموزش و توزیع شبکه مولد چه ویژگی نسبت به همدیگر داشته باشند نمی‌توانیم از رویکرد بالا برای آموزش شبکه استفاده کنیم. برای رفع این مشکل چه کاری می‌تواند صورت گیرد؟

## مسئله ۸. کاربرد شبکه‌های مولد متخاصمی (۶ نمره)

یکی از کاربردهای شبکه‌های مولد متخاصمی کاربرد Super Resolution است. در این کاربرد ما می‌خواهیم تصاویر با رزولوشن پایین را تبدیل به تصاویر با رزولوشن بالا کنیم. برای این کار می‌توان از یک شبکه عصبی به این صورت استفاده کرد که به عنوان ورودی، تصاویر با رزولوشن پایین را بگیرد و در خروجی تصاویر با رزولوشن بالا را به ما برگرداند. برای این روش می‌توان از تابع هدف Mean Square Error (MSE) استفاده کرد. مشکلی که در این راهکار به وجود می‌آید آن است که تصاویر تولید شده با رزولوشن بالا، تاری می‌شوند. یکی از روش‌ها برای رفع این موضوع، استفاده از تابع هزینه GAN در کنار تابع هزینه MSE است. فرض کنید دیتاست ما به صورت جفت داده‌های  $(I^{LR}, I^{HR})$  است که شامل تصویر با رزولوشن پایین و تصویر با رزولوشن بالای متناظر با آن است. در مورد ساختار شبکه عصبی مورد استفاده برای این کار و تابع هزینه هر یک از شبکه‌های مولد و تمیزدهنده توضیح دهید.

موفق باشید