

بسمه تعالی

امتحان میان ترم

استاد درس: فاطمه سیدصالحی

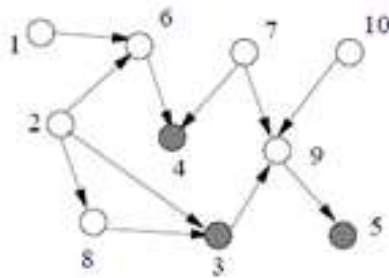
دانشکده علوم ریاضی

زمان: ۱۰۰ دقیقه

مدل های مولد

مدل های گرافی احتمالاتی

۱. برای گراف زیر وجود و عدم وجود شرایط استقلال زیر را مشخص کنید: (۶ نمره)



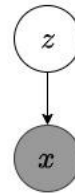
$$X_9 \perp X_8 | \{X_3, X_4, X_5\}$$

$$X_7 \perp X_{10} | \{X_3, X_4, X_5\}$$

$$X_2 \perp X_8 | \{X_3, X_4, X_5\}$$

۲. شبکه زیر با توزیع های احتمالاتی شرطی روی آن را در نظر بگیرید. (۹ نمره)

$$p(z) = z e^{-\frac{z^2}{2}} \quad z \geq 0$$
$$p(x|z) = \mathcal{N}(x|z, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-z)^2}$$



می خواهیم به کمک variational inference توزیع احتمال پسین $p(z|x)$ را تقریب بزنیم. برای این کار توزیع پسین را با یک توزیع رایلی با پارامتر α به صورت زیر تقریب می زنیم:

$$q(z) = \frac{z}{\alpha} e^{-\frac{z^2}{2\alpha}}$$

مقدار بهینه پارامتر α را محاسبه کنید.

راهنمایی: برای یک متغیر Z با توزیع رایلی با پارامتر α داریم:

$$E[Z] = \sqrt{\frac{\alpha\pi}{2}} \quad E[Z^2] = 2\alpha$$

مدل‌های مبتنی بر معماری Autoregressive

۱. معماری Masked Autoencoder for Distribution Estimation ترکیبی از معماری AE و Autoregressive برای تخمین توزیع فاکتورایز شده با chain rule را به یاد آورید. در میان دو شکل زیر توضیح دهید که آیا هیچ یک می‌تواند به عنوان یک MADE باشد یا خیر؟ (۱۰ نمره)



الف) در حالت کلی یک MADE با یک لایه نهان با بعد $D-1$ در نظر بگیرید که توزیع داده در آن D بعدی است. تابع لایه‌های این شبکه را به شکل زیر فرض کنید:

$$h(\mathbf{x}) = g(\mathbf{b} + (\mathbf{W} \odot \mathbf{M}^W)\mathbf{x})$$

$$\hat{\mathbf{x}} = \text{sigmoid}(\mathbf{c} + (\mathbf{V} \odot \mathbf{M}^V)h(\mathbf{x}))$$

که M^W و M^V ماتریس‌های masking هستند. برای هر کدام از شکل‌های بالا که MADE معتبر هستند ماتریس‌های masking را بنویسید.

ب) ماتریس $M^V M^W = M$ را در نظر بگیرید. ماکسیمم تعداد خانه‌های غیر صفر ماتریس M که در آن صورت ویژگی autoregressive گراف همچنان حفظ می‌شود را محاسبه کنید؟

۲. توضیح دهید کدام یک از دو معماری transformer encoder و transformer decoder به عنوان یک مدل مولد قابل استفاده هستند؟ چرا؟ (۵ نمره)

۳. فرض کنید تعداد زیادی داده برای زدن برچسب به محتوای یک متن کوتاه در اختیار داریم. برای این کار دو رویکرد زیر در نظر گرفته شده است: (۵ نمره)

- توسعه و آموزش یک مدل کوچک discriminative بر روی مجموعه داده برای هدف مشخص

- استفاده از مدل زبانی GPT4 و دادن داده‌ها به شکل few-shot در prompt

آیا در رویکرد دوم هم آموزش رخ می‌دهد؟ برای هر یک از دو رویکرد یک مزیت بر دیگری ارائه دهید.

مدل‌های مبتنی بر معماری VAE

۱. دو معماری AE و VAE را در نظر بگیرید. (۱۰ نمره)

الف) چرا AE یک مدل مولد نمی‌تواند باشد و VAE چگونه این مساله را حل کرده است که به عنوان مدل مولد قابل استفاده است؟

ب) توضیح دهید چرا در بازسازی یک تصویر (وارد کردن تصویر به معماری و گرفتن تصویر بازسازی شده) VAE عملکرد ضعیف‌تری نسبت به AE دارد؟ معماری $\beta - VAE$ چگونه این مشکل را حل کرده است؟

۲. تابع هزینه معماری VAE را در نظر بگیرید: (۲۰ نمره)

$$E_{q_\phi(z|x)}[\log p(z, x; \theta) - \log q_\phi(z|x)]$$

چالش این تابع در حین بهینه‌سازی به دلیل وجود امید ریاضی نسبت به توزیع q_ϕ است که پارامترهای آن در عملکرد گرادینان وجود دارد.

الف) به کمک روش importance sampling تساوی زیر را ثابت کنید:

$$\nabla_\phi E_{q_\phi(z)}[f(z)] = \frac{1}{n} \sum_{i=1}^n f(z^i) \nabla_{q_\phi(z^i)} \log q_\phi(z^i)$$

ب) چگونه می‌توان از روش بالا برای حل مساله بهینه‌سازی VAE استفاده کرد؟ این تخمین چه مشکلی دارد؟

ج) تکنیک reparameterization چه رویکردی را برای حل بهینه‌سازی بالا پیشنهاد می‌دهد؟

د) مختصر توضیح دهید که رویکرد RL چگونه می‌تواند در بهینه‌سازی عبارت بالا مورد استفاده قرار گیرد؟

مدل‌های مبتنی بر معماری GAN

۱. در معماری Goodfellow GAN، برای یک generator ثابت، discriminator بهینه را بیابید. حال نشان دهید که تابع بهینه‌سازی این معماری در هر گام به دنبال کمینه کردن فاصله Jensen-Shannon میان توزیع واقعی داده و توزیع داده‌های generator است. (۱۰ نمره)

۲. آموزش یک مدل GAN به کمک خانواده f-divergence چه چالش‌هایی دارد؟ برای حل این چالش چه جایگزینی را برای تخمین فاصله دو توزیع احتمالاتی پیشنهاد می‌کنید؟ (۵ نمره)

۳. چرا در معماری GAN ممکن است دچار مشکل Mode collapse شویم؟ آیا اگر راهی برای بیرون آوردن generator از یک modality داشته باشیم و آن را به سمت یادگیری modality های دیگر فضا ببریم، می‌توانیم مطمئن باشیم که generator کل فضای توزیع را در طول زمان به یاد می‌آورد؟ چرا؟ (۱۰ نمره)

۴. در معماری Gaussian VAE توزیع variational به شکل یک توزیع نرمال در نظر گرفته شد. در اینجا می‌خواهیم محدودیتی روی فرم توزیع variational و فرم توزیع نهایی داده‌ها به شکل مستقیم نگذاریم و آن را به شکل ضمنی و به کمک معماری GAN مدل کنیم. (۱۵ نمره)

فرم زیر از تابع هزینه VAE را در نظر بگیرید:

$$E_{q_{\phi}(z|x)}[\log p(x|z; \theta)] - D_{KL}(q_{\phi}(z|x) || p(z))$$

الف) نشان دهید عبارت اول را می‌توان به شکل زیر بازنویسی کرد:

$$E_q \left[\log \frac{p(x|z; \theta)}{p_{data}(x)} \right] + E_q [\log p_{data}(x)]$$

ب) معماری جدیدی پیشنهاد دهید که در آن

- خروجی شبکه variational توسط یک discriminator با توزیع پیشین $p(z)$ مقایسه می‌شود.

- خروجی شبکه مولد (دیکدر) توسط یک discriminator دیگر با توزیع اصلی داده مقایسه می‌شود:

معماری جدید را بکشید و تابع هزینه آن را با ذکر توضیحات محاسبه کنید. تابع هزینه باید بر روی پارامترهای شبکه دیکدر، انکدر و دو discriminator تعریف شود.