به نام خدا

دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

مدل‌های مولد عمیق

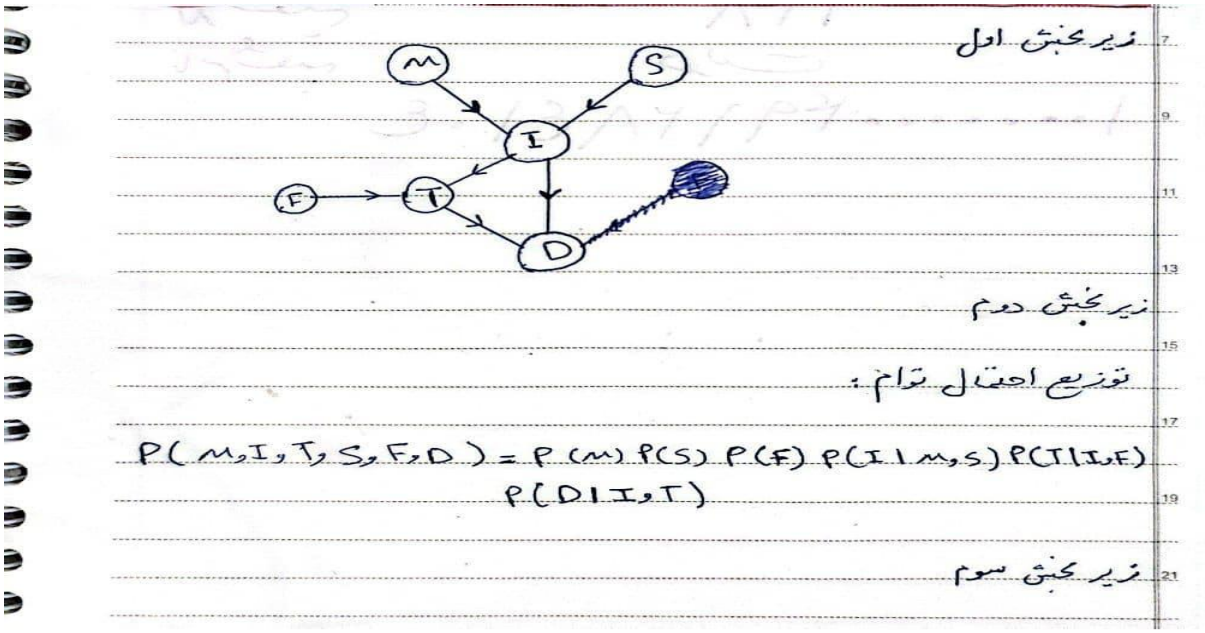
تمرین اول

نام و نام خانوادگی	محمد متقی
شماره دانشجویی	159403018
تاریخ ارسال گزارش	

فهرست

سؤال اول

بخش اول



A) نادرست، میان Fو Dمسیر فعالی وجود دارد. $F \to T \to D$:چون شرطی بر T نیست، مسیر باز است و این دو متغیر وابسته‌اند.

B) **درست،** مسیرهای S به I به D و S به I به T به D همگی به‌خاطر شرط بر I بسته می‌شوند چون I گره‌ای غیر برخوردکننده در میانه مسیر است؛پس S و D (فصل و مرگ با دانستن شدت بیماری) به شرط I مستقل‌اند.

C) **درست،** میان M و F فقط مسیرهایی وجود دارد که شامل برخوردکننده‌ها هستند، مانند M به I به T از یک طرف و F از طرف دیگر که در T به هم می‌رسند. از آن‌جا که برخوردکننده‌ها تا وقتی رویشان شرط نگذاریم مسیر را می‌بندند، این دو متغیر مستقل‌اند.

D) **نادرست،** وقتی روی برخوردکننده T شرط می‌گذاریم، مسیر M به I به T از یک سو و F از سوی دیگر باز می‌شود و وابستگی القا می‌گردد؛ پس استقلال از بین می‌رود..

E) **درست،** شرطگذاری همزمان بر D و I باعث بسته شدن همه مسیرهای میان M و T می‌شود؛ مسیر مستقیم M به I به T با شرط بر I بسته است و مسیر غیرمستقیم M به I به D به T نیز به‌دلیل وجود همان شرط‌ها بسته می‌ماند. به این ترتیب در این حالت M و T مستقل‌اند.

بخش دوم

زیربخش اول

در شبکه‌های بیزی، هر متغیر فقط به والدهای خودش وابسته است پس توزیع توأم همه متغیرها حاصل‌ضرب احتمال هر متغیر مشروط به والدهای خودش است.

$$p(O,S,A,T,B,M,C)=p(O)p(A)p(C)p(S|O)p(T|O,A)p(B|S)p(M|A,T,B)$$

زیربخش دوم

Markov blanket یک گره مجموعه‌ای از گره‌هاست که اگر مقدارشان را بدانی،آن گره از تمام گره‌های دیگر شبکه مستقل می‌شود. یعنی فرزاندان، والدین و سایر والدین فرزندان آن گره.

$$MB(T)=\{O,A,M,B\}$$

زیربخش سوم

خیر، گراف مارکوف حاصل از این شبکه‌ی بیزی Perfect I-Map نیست چون در فرایند moralization ، والدهای مشترک هر گره به هم وصل می‌شوند و ساختارهای برخوردکننده مثل $O \rightarrow T \leftarrow A$ یا $A \rightarrow M \leftarrow T$ از بین می‌روند. این باعث می‌شود بعضی استقلال‌هایی که در شبکه‌ی بیزی وجود داشت (مثل استقلال بین Oو A یا بین Aو B هنگام شرطگذاری) دیگر در گراف مارکوف برقرار نباشد. در نتیجه گراف فقط یک I-Map معمولی است، نه Perfect I-Map، چون بخشی از استقلال‌های اصلی مدل جهت‌دار را از دست می‌دهد.

زیربخش چهارم

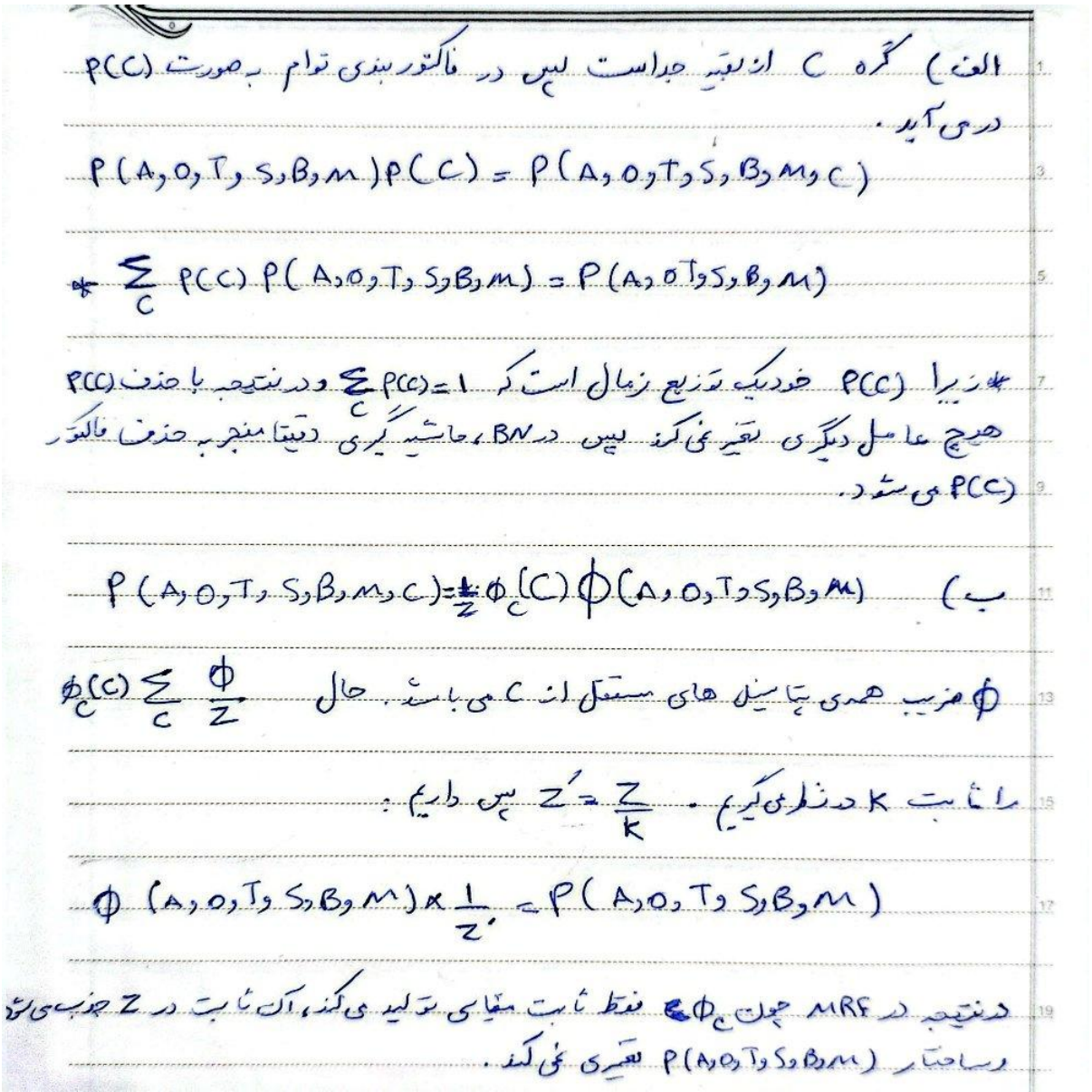
خیر، گراف مارکوف حاصل chordal نیست .در تعریف، گرافی chordal است که هیچ چرخه‌ای با چهار گره یا بیشتر بدون یال میان‌بر نداشته باشد. در گراف ما، پس از مارال‌سازی، چرخه‌هایی مانند $O-S-B-T-O$ یا $A-T-M-B-A$ وجود دارند که بین رأس‌های غیرمتوالی مثل O و B یا Sو Tیالی وجود ندارد. نبود این یال‌ها یعنی چرخه‌های بدون chord داریم، بنابراین گراف غیرchordal است و نمی‌توان آن را دقیقاً با یک شبکه‌ی بیزی معادل بازسازی کرد.

زیربخش پنجم

کلیک‌های ماکسیمال دقیقاً $\{A,O,T\}$، $\{A,T,M\}$، $\{T,B,M\}$، $\{O,S\}$و $\{S,B\}$هستند. مجموعه $\{O,S,B\}$کلیک نیست چون یال‌های OB و ST وجود ندارند و بنابراین $\{O,S\}$و $\{S,B\}$هر کدام به‌تنهایی ماکسیمال باقی می‌مانند. در نتیجه توزیع توأم به‌صورت گیبس روی کلیک‌های ماکسیمال فاکتور می‌شود.

$$p(A,O,T,S,B,M,C) \propto \phi AOT(A,O,T)\phi ATM(A,T,M)\phi TBM(T,B,M)\phi OS(O,S)\phi SB(S,B)\phi C(C)$$

زیربخش ششم


الف) گره C از بقیه جداست پس در فاکتوربندی توام به صورت $P(C)$ در می‌آید.

$$P(A,O,T,S,B,M)P(C) = P(A,O,T,S,B,M,C)$$

$$* \sum_C P(C)\,P(A,O,T,S,B,M) = P(A,O,T,S,B,M)$$

*زیرا $P(C)$ خودیک توزیع زمال است که $\sum_C P(C)=1$ و در نتیجه با حذف $P(C)$ هیچ عامل دیگری تغییر نمی‌کند پس در BN، حاشیه‌گیری دقیقا منجر به حذف فاکتور $P(C)$ می‌شود.

ب) $P(A,O,T,S,B,M,C)=\frac{1}{Z}\phi_C(C)\phi(A,O,T,S,B,M)$

$\phi_C(C)$ ضریب همه پتانسیل های مستقل از C می‌باشد. حال $\sum_C \frac{\phi}{Z}$

با ثابت K در نظر می‌گیریم. $Z'=\frac{Z}{K}$ پس داریم:

$$\phi(A,O,T,S,B,M) \propto \frac{1}{Z'} = P(A,O,T,S,B,M)$$

در نتیجه در MRF چون ϕ_C فقط ثابت متقابلی تولید می‌کند، آن ثابت در Z جذب می‌شود و ساختار $P(A,O,T,S,B,M)$ تغییری نمی‌کند.

بخش سوم

زیربخش اول

کلیک‌های ماکسیمال این‌ها هستند $\{A,B,C\}$، $\{B,C,D\}$، $\{C,D,F\}$، $\{B,E\}$، $\{E,F\}$، $\{E,G\}$.

$$p(A,B,C,D,E,F,G) \propto$$

$$\phi ABC(A,B,C)\phi BCD(B,C,D)\phi CDF(C,D,F)\phi BE(B,E)\phi EF(E,F)\phi EG(E,G)$$

زیربخش دوم

الف)نادرست، G و A مستقل نیستند زیرا مسیر A B E G باز است و وابستگی بین آن‌ها برقرار می‌شود.

ب) نادرست، F و A با داشتن مجموعه D و C مستقل نیستند. مسیرهای A C F و A B D F با شرط‌های C و D بسته می‌شوند اما مسیر A B E F همچنان باز است چون روی B و E شرط نکرده‌ایم و این مسیر وابستگی را حفظ می‌کند.

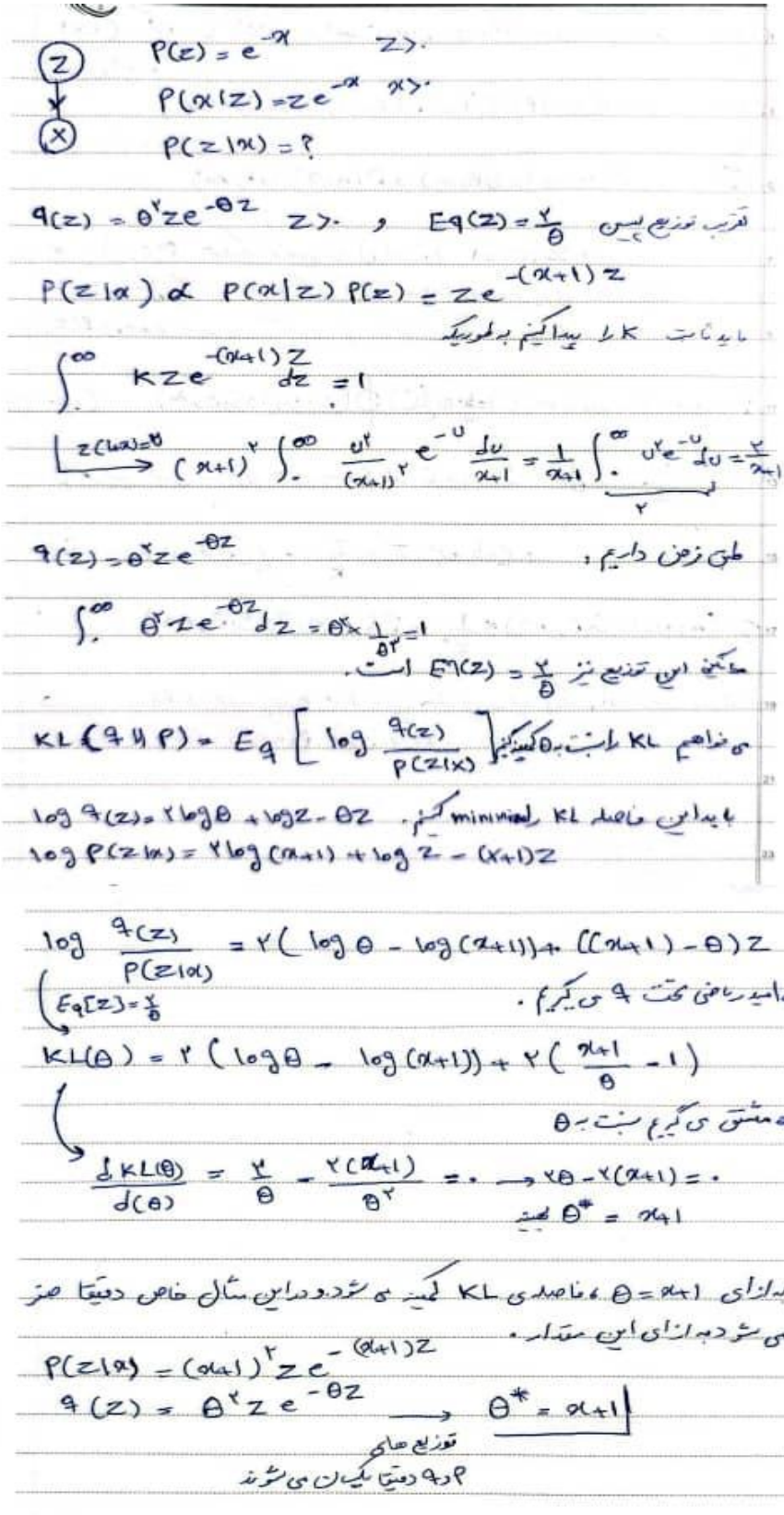
ج) درست، G و C با داشتن E مستقل هستند زیرا تمام مسیرهای میان آن‌ها مانند C F E G و C B E G از E عبور می‌کنند و شرط بر E این مسیرها را می‌بندد.

د) درست، شرط دانستن B و C برای A کافی است و افزودن E اطلاعات جدیدی نمی‌دهد زیرا در گراف بدون جهت، A با شرط همسایه‌هایش B و C از سایر گره‌ها مستقل می‌شود.

زیربخش سوم

همانطور که در بخش قبل ثابت کردیم، ضرب یک ثابت مثبت در پتانسیل ثابت نرمال‌سازی Z را مقیاس می‌کند و شکل توزیع پس از نرمال‌سازی تغییر نمی‌کند؛ توزیع نهایی روی $(A,\dots,G)$ همان قبلی است.

بخش چهارم


Z
X
$P(z) = e^{-z}$ $z>0$
$P(x|z) = ze^{-xz}$ $x>0$
$P(z|x) = ?$
$q(z) = \theta^2 z e^{-\theta z}$ $z>0$, $Eq(z) = \frac{2}{\theta}$ تقریب توزیع پسین
$P(z|x) \propto P(x|z)P(z) = ze^{-(x+1)z}$
باید ثابت K را پیدا کنیم به طوریکه
$\int_0^\infty Kze^{-(x+1)z} dz = 1$
$z(x+1)=u$ $\frac{1}{(x+1)^2}\int_0^\infty u^2 e^{-u}\frac{du}{x+1} = \frac{1}{x+1}\underbrace{\int_0^\infty u^2 e^{-u} du}_{2} = \frac{2}{x+1}$
$q(z) = \theta^2 z e^{-\theta z}$
طبق فرض داریم:
$\int_0^\infty \theta^2 z e^{-\theta z} dz = \theta^2 \times \frac{1}{\theta^2} = 1$
که نشان میدهد این توزیع نیز $E_q(z) = \frac{2}{\theta}$ است.
$KL(q \| p) = E_q\left[\log \frac{q(z)}{p(z|x)}\right]$ می خواهیم KL را نسبت به θ کمینه کنیم
$\log q(z) = 2\log\theta + \log z - \theta z$ با پیدا این فاصله KL minimized کنیم.
$\log p(z|x) = 2\log(x+1) + \log z - (x+1)z$
$\log \frac{q(z)}{p(z|x)} = 2(\log\theta - \log(x+1)) + ((x+1)-\theta)z$
$E_q[z] = \frac{2}{\theta}$ با امید ریاضی تحت q می گیریم.
$KL(\theta) = 2(\log\theta - \log(x+1)) + 2\left(\frac{x+1}{\theta} - 1\right)$
مشتق می گیریم نسبت به θ
$\frac{dKL(\theta)}{d(\theta)} = \frac{2}{\theta} - \frac{2(x+1)}{\theta^2} = 0 \rightarrow 2\theta - 2(x+1) = 0$
یعنی $\theta^* = x+1$
پس از ای $\theta = x+1$ فاصله ی KL کمینه شده و در این مثال خاص دقیقا صفر می شود به ازای این مقدار.
$P(z|x) = (x+1)^2 z e^{-(x+1)z}$
$q(z) = \theta^2 z e^{-\theta z} \rightarrow \theta^* = x+1$
توزیع های q و p دقیقا یکسان می شوند

سؤال دوم

بخش اول

زیربخش اول

به طور کلی می توان گفت در مدل‌های مولد نهفته، محاسبهٔ مستقیم احتمال داده‌ها $\log\ p_\theta(x)$ ممکن نیست و برای رفع این مشکل از یک توزیع تقریبی $q_\phi(z \mid x)$استفاده می‌کنیم و با نامساوی ینسن (Jensen)کران پایینی به‌دست می‌آوریم که به آن ELBO می‌گویند.

این کران جمع دو مؤلفه است. مؤلفه اول کیفیت بازسازی را می‌سنجد. مؤلفه دوم میزان انحراف توزیع کمکی از پرایر را جریمه می‌کند. هر چه کران بالاتر باشد مدل بهتر داده را توضیح می‌دهد.

$$E_{q(z|x)}[logp(x|z)] - D_{KL}(q(z|x)||p(z))$$

The first term = reconstruction, the second term = regularization

زیربخش دوم

با توجه به توضیحات گیت هاب، dSprites یک دیتاست دوبعدی است که برای ارزیابی میزان Disentanglementدر مدل‌های Unsupervised Representation Learning طراحی شده است. هر تصویر شامل یک شکل ساده در پس‌زمینه سیاه با ابعاد 64×64 پیکسل است و همه تصاویر به صورت رویه‌ای ساخته شده‌اند.

این تصاویر بر اساس 6 عامل نهفته مستقل (Independent Latent Factors) تولید شده‌اند که در دیتاست dSprites تمام ترکیب‌های ممکن از این 6 عامل دقیقاً یک بار ظاهر می‌شوند، بنابراین تعداد کل تصاویر برابر با 737,280نمونه می شود.

فاکتور ها: color, shape, scale, orientation, position X, and position Y

هدف از ایجاد dSprites این بوده که مانند یک تست استاندارد برای ارزیابی Disentanglement عمل کند و بسنجد که آیا مدل‌های یادگیری بدون نظارت قادرند عوامل نهفته واقعی را به‌درستی از داده استخراج کنند یا نه و در مقاله معروف β-VAE: Learning basic visual concepts with a constrained variational framework منتشرشده در کنفرانس ICLR 2017، از این دیتاست برای محاسبه‌ی متریک‌های سنجش Disentanglement استفاده شده است.

شکل 1 : سه نمونه تصادفی از دیتاست dsprites

زیربخش سوم

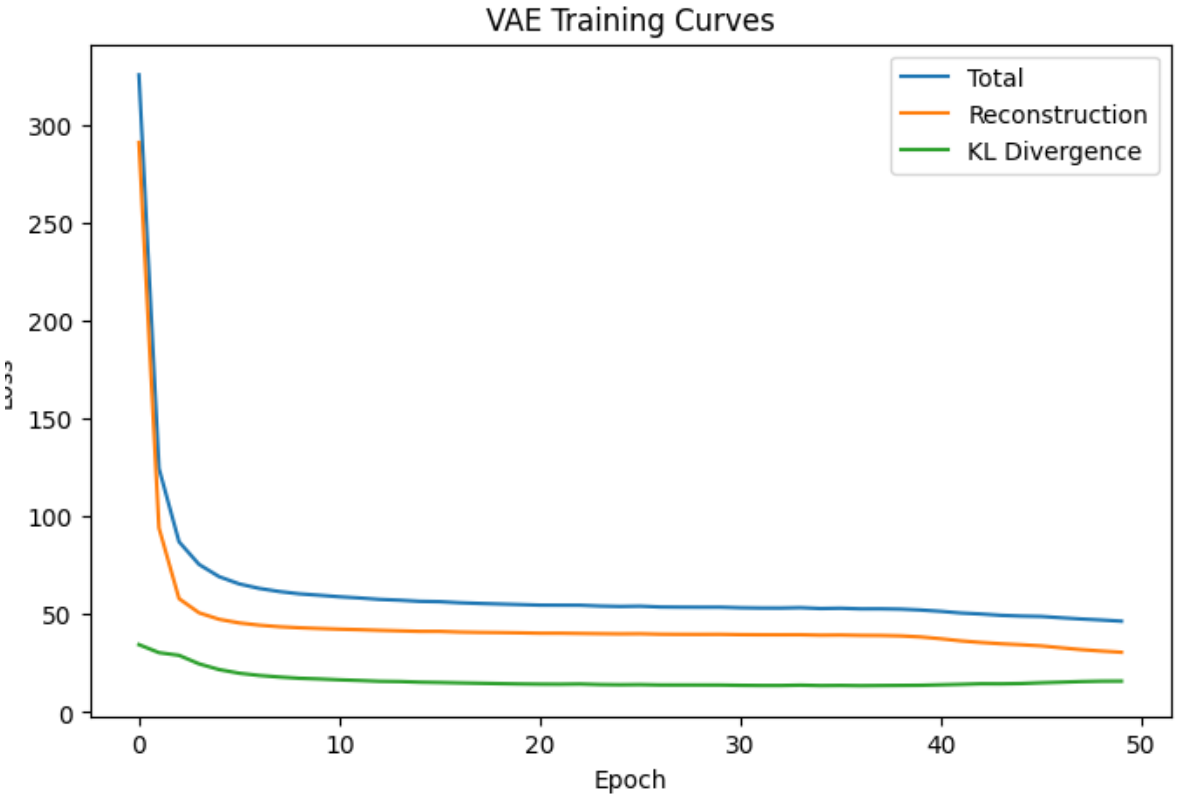
در مدل VAE برخلاف یک Autoencoder معمولی، خروجی بخش Encoder یک بردار ثابت نیست، بلکه یک توزیع احتمالی روی فضای نهفته است که معمولاً نرمال در نظر گرفته می‌شود در واقع، به جای اینکه Encoder مستقیماً مقدار z را برگرداند، دو بردار میانگین و انحراف معیار تولید می کند و سپس برای هر نمونه، یک z تصادفی از این توزیع انتخاب می‌کنیم.

اما این عمل تصادفی (Sampling) درون شبکه باعث می‌شود مسیر گرادیان‌ها از Decoder به Encoder قطع شود، زیرا عمل نمونه‌گیری قابل مشتق‌گیری نیست و به همین دلیل آموزش مدل با Backpropagation دیگر ممکن نیست.

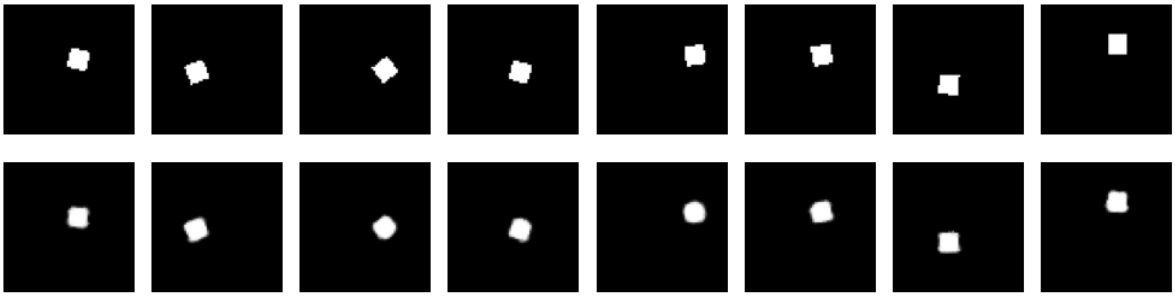
راه‌حل این مشکل استفاده از Reparameterization Trick است. در این روش، به جای آن‌که مستقیماً از توزیع نرمال نمونه بگیریم، یک نویز استاندارد را از توزیع نرمال استاندارد N(0, I) می‌کشیم و سپس z را به صورت تابعی از سه پارامتر μ، σ و ε بازنویسی می‌کنیم:

$$z = \mu + \sigma \odot \varepsilon, \varepsilon \sim \mathcal{N}(0, I)$$

به این ترتیب، تمام بخش‌های μ و σ که از شبکه می‌آیند، در مسیر گرادیان قرار می‌گیرند و تنها ε یک نویز ثابت است. این بازنویسی باعث می‌شود عملیات نمونه‌گیری به شکل یک عمل مشتق‌پذیر دربیاید و بتوان از Backpropagation استفاده کرد. در نتیجه، مشکل قطع شدن گرادیان‌ها حل می‌شود و Encoder و Decoder هر دو می‌توانند به طور همزمان با روش گرادیان نزولی آموزش ببینند.



شکل 2 آموزش VAE با استفاده از 10000 داده در Epoch 50



شکل 3 نمونه های تصادفی و بازسازی شده

نتایج آموزش مدل VAE نشان می‌دهد که فرآیند یادگیری به‌صورت پایدار و مؤثر انجام شده است. در نمودار تابع هزینه، مشاهده می‌شود که Total Loss در چند epoch نخست به‌شدت کاهش یافته و سپس به مقدار تقریباً ثابت حدود ۴۶ رسیده است، که نشان‌دهنده‌ی همگرایی مدل است. کاهش سریع در مراحل ابتدایی به این معناست که مدل در همان دوره‌های اولیه ساختار کلی داده‌ها را یاد گرفته و در ادامه فقط جزئیات بازسازی را بهبود داده است. همچنین، Reconstruction Loss به حدود ۳۰ کاهش یافته که حاکی از بازسازی مناسب شکل‌ها و موقعیت اجسام در تصاویر dSprites است. در مقابل، مقدار KL Divergence در حدود ۱۵ مانده که نشان می‌دهد فضای نهان منظم و نزدیک به توزیع نرمال استاندارد باقی مانده است.

از نظر کیفی، بازسازی‌های نهایی بسیار مشابه ورودی‌ها هستند و مدل موفق شده ویژگی‌های اصلی نظیر شکل و موقعیت را به‌درستی حفظ کند. در تصاویر بازسازی‌شده تفاوت چشمگیری با نمونه‌های اصلی دیده نمی‌شود و تنها مقدار کمی افت جزئیات در لبه‌ها وجود دارد، که درVAE های پایه طبیعی است. در مجموع، مدل تعادل مناسبی میان دقت بازسازی و منظم‌بودن فضای نهان برقرار کرده است.

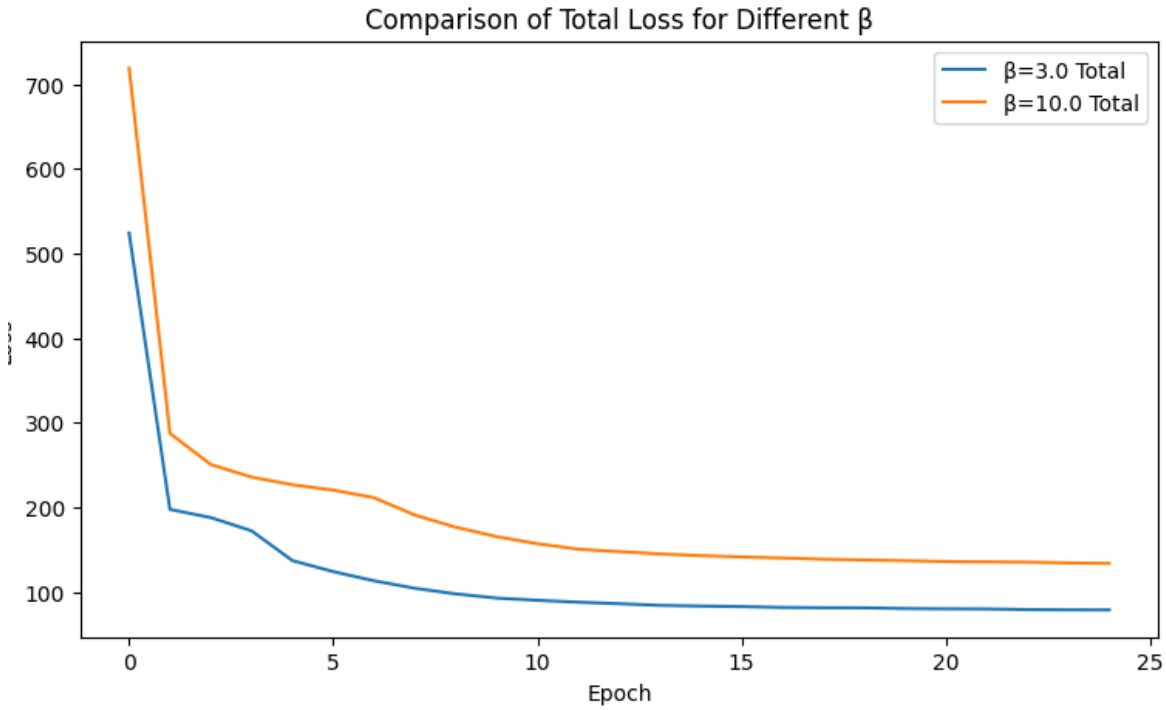
زیربخش پنجم

ترم اول یعنی بازسازی (Reconstruction Term) تلاش می‌کند خروجی مدل تا حد ممکن شبیه ورودی باشد .ترم دوم یعنی KL Divergence باعث می‌شود توزیع نهفته $q(z|x)$ خیلی از توزیع پرایر $p(z)$ معمولاً Normal(0,1) دور نشود.

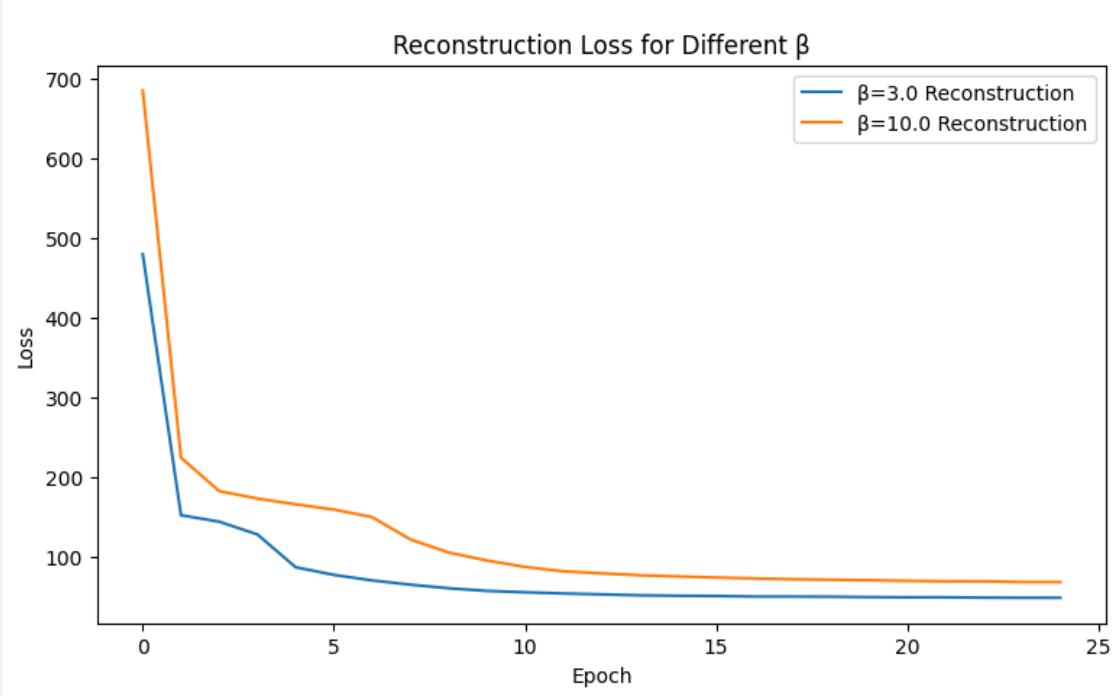
اگر $\beta = 1$ باشد، مدل دقیقاً همان VAE کلاسیک است ولی وقتی $\beta > 1$باشد، مقدار ترم KL در تابع هدف بزرگ‌تر می‌شود و مدل مجازات بیشتری برای تفاوت بین $q(z|x)$ و $p(z)$ در نظر می‌گیرد .به زبان ساده مدل را مجبور می‌کنیم که متغیرهای نهفته‌اش (z) تا حد ممکن شبیه توزیع نرمال استاندارد و از هم مستقل باشند.

β-VAEهمان VAE است که در آن با بزرگ‌تر کردنβ ، به مدل فشار بیشتری وارد می‌کنیم تا فضای نهفته‌اش منظم‌تر و مستقل‌تر باشد. این باعث می‌شود عوامل پنهان داده‌ها به صورت تفکیک‌شده در متغیرهای z نمایش داده شوند.

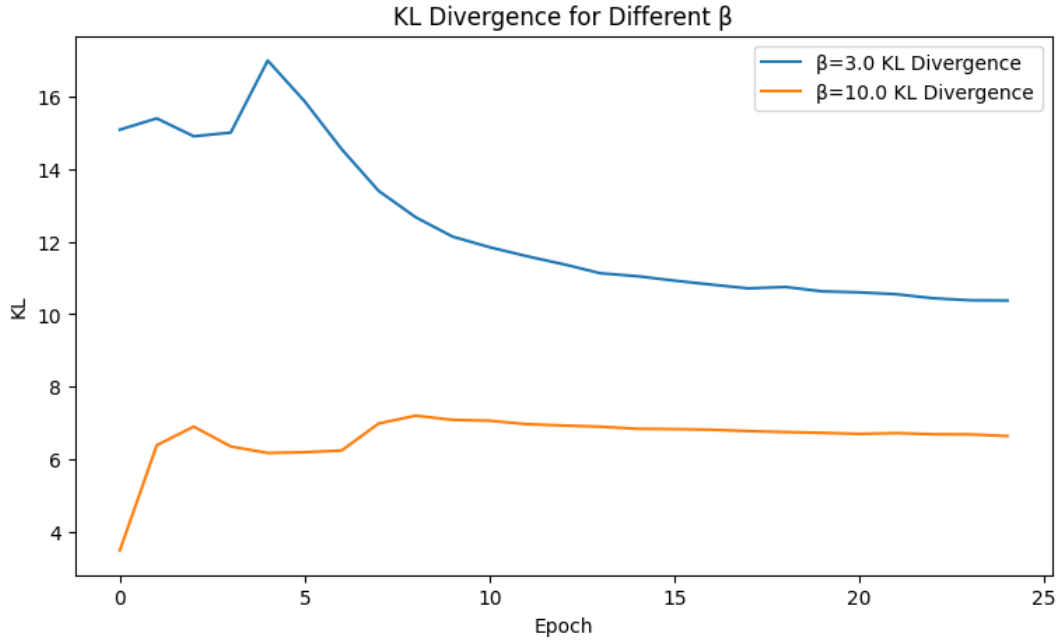
وقتی **β>1** باشد، مدل بیشتر وادار می‌شود که $q(z \mid x)$ به $p(z)$ نزدیک‌تر شود.



شکل 4 مقایسه loss بین دو مقدار beta

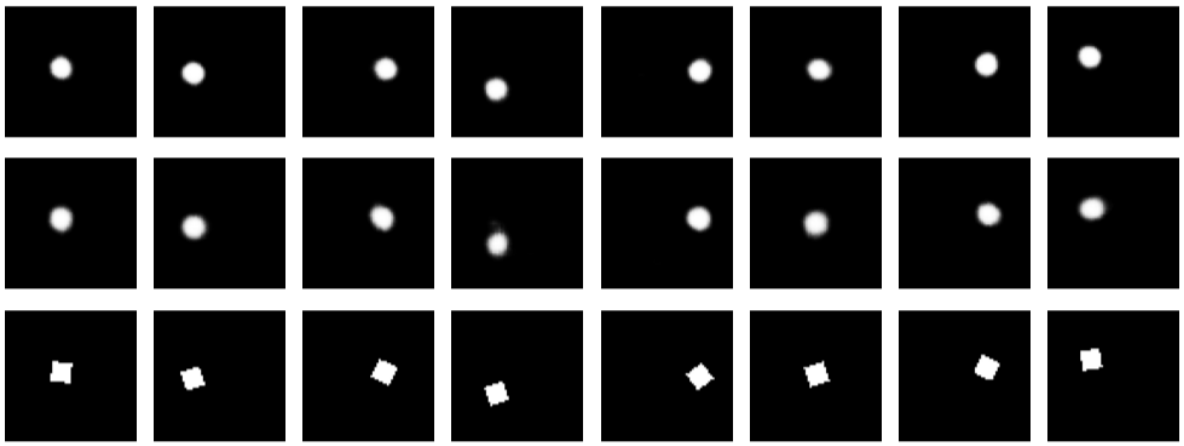


شکل 5 مقایسه loss بازسازی



شکل 6 مقایسه KL فاصله به ازای دو مقدار beta

می توان از نمودار های بالا دریافت که بر اساس منحنی‌های به‌دست‌آمده، β=3 بازسازی بهتری ارائه می‌دهد ولی فضای نهان کمتر منظم است؛β=10 بازسازی ساده‌تر اما فضای نهان منظم‌تر و مناسب‌تر برایdisentanglement می‌سازد. انتخاب مقدار β باید بر اساس هدف نهایی صورت گیرد: برایdSprites اگر نمایش جدایش‌پذیر معیار اصلی است، β=10 انتخاب ترجیحی است؛ اگر کیفیت بازسازی اولویت دارد، β=3بهتر عمل می‌کند.



شکل 7 مقایسه اشکال بازی سازی شده با دو مقدار beta 3 و 10 و عکس اصلی در پایین

همان‌طور که در شکل مشاهده می‌شود، مدل با β=3 بازسازی‌های واضح‌تری ارائه می‌دهد، در حالی‌که β=10 تصاویر نرم‌تر و کم‌جزئیات‌تر تولید کرده است. این رفتار مطابق انتظار VAE است با افزایشβ ، مدل فشار بیشتری برای منظم‌سازی فضای نهان اعمال می‌کند، در نتیجه دقت بازسازی کاهش یافته ولی نمایش نهان جدایی‌پذیرتر می‌شود.

به طوری کلی می توان اینگونه جمع بندی کرد که نتایج به‌دست‌آمده از اجرای مدل‌های β-VAE با مقادیر مختلف β نشان می‌دهد که این پارامتر نقش کلیدی در تنظیم توازن میان دقت بازسازی و منظم‌بودن فضای نهان دارد. با مقدار β=3 مدل توانست شکل‌ها و موقعیت اجسام را با وضوح بالا و خطای بازسازی کمتر بازتولید کند، در حالی‌که افزایش β به 10 باعث شد تصاویر خروجی نرم‌تر و ساده‌تر شوند اما فضای نهان ساختاریافته‌تر و منظم‌تر گردد. همان‌طور که در نمودارهای آموزشی دیده می‌شود، در β بزرگ‌تر، مقدار Reconstruction Loss افزایش و KL Divergence کاهش یافته است، که بیانگر محدودتر شدن اطلاعات انتقال‌یافته از ورودی به فضای نهان است. این رفتار مطابق با تئوری β-VAE است که با اعمال ضریب بزرگ‌تر بر ترمKL ، مدل را وادار می‌کند بازنمایی‌های فشرده‌تر و مستقل‌تری بیاموزد. در نتیجه، مدل با β کوچک‌تر در بازسازی دقیق‌تر عمل می‌کند، ولی مدل با β بزرگ‌تر نمایش نهانی فراهم می‌سازد که برای disentanglement و شناسایی عوامل پنهان در داده‌ها مناسب‌تر است.

زیربخش هفتم

معیار MIG (Mutual Information Gap)یکی از پرکاربردترین سنجه‌ها برای ارزیابی میزان جدایش‌پذیری نمایش نهان (disentanglement) در مدل‌های مولد مانند β-VAE است. ایده‌ی اصلی این معیار بر پایه‌ی اندازه‌گیری ارتباط میان هر فاکتور پنهان واقعی داده مثل موقعیت، مقیاس یا چرخش در dSprites و هر بعد از فضای نهان z است.

برای هر فاکتور، ابتدا اطلاعات متقابل بین آن فاکتور و تمام ابعاد z محاسبه می‌شود، سپس دو مقدار بیشینه انتخاب می‌شوند: بیشترین اطلاعات متقابل و دومین بیشترین مقدار. اختلاف بین این دو مقدار بیان می‌کند که آیا یک فاکتور عمدتاً توسط یک بُعد خاص در z توضیح داده می‌شود یا بین چند بُعد پخش شده است MIG به‌صورت میانگین نرمال‌شده‌ی این اختلاف‌ها بر کل فاکتورها تعریف می‌شود.

اگر MIG بالا باشد، یعنی هر ویژگی پنهان داده با یک بُعد از فضای نهان متناظر است (disentangled)، در حالی‌که MIG پایین یا نزدیک صفر نشان می‌دهد ابعاد نهان درهم و غیرمستقل‌اند. بنابراین، این معیار سنجه‌ای کمی از میزان تفکیک‌پذیری عوامل مولد در فضای نمایش مدل فراهم می‌کند.

```
β=1   MIG=0.0108  {'shape': 0.0, 'scale': 0.0, 'orientation': np.float64(0.0013), 'posX': np.float64(0.0195), 'posY': np.float64(0.0331)}
β=3   MIG=0.0198  {'shape': 0.0, 'scale': 0.0, 'orientation': np.float64(0.001), 'posX': np.float64(0.0502), 'posY': np.float64(0.048)}
β=10  MIG=0.0138 {'shape': 0.0, 'scale': 0.0, 'orientation': np.float64(0.0), 'posX': np.float64(0.0492), 'posY': np.float64(0.0196)}
```

شکل 8 محاسبه MIG به ازای beta های مختلف

در این بخش، عملکرد مدل‌های β-VAE با مقادیر مختلف β مورد بررسی قرار گرفت تا میزان جدایش متغیرهای نهان (disentanglement) با استفاده از شاخص Mutual Information Gap (MIG) ارزیابی شود. سه مدل با βهای 1، 3 و 10 آموزش داده شدند و نتایج عددی به‌صورت زیر به‌دست آمدند:

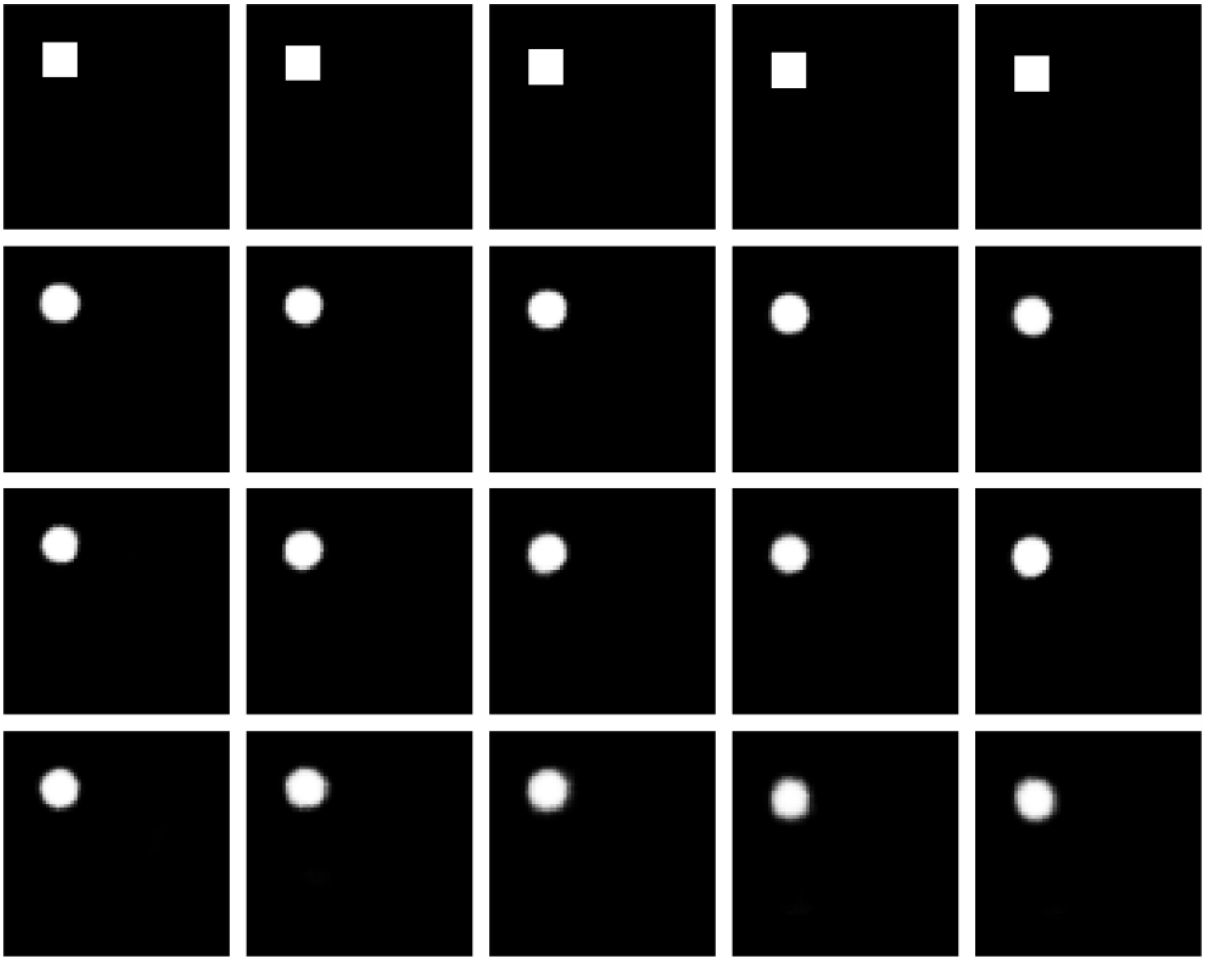
$\beta=1 \rightarrow \text{MIG}=0.0108$ ، $\beta=3 \rightarrow \text{MIG}=0.0198$ و $\beta=10 \rightarrow \text{MIG}=0.0138$.

این نتایج نشان می‌دهد که با افزایش β از 1 به 3، مقدار MIG افزایش یافته است، که بیانگر بهبود در جدایش متغیرهای پنهان و افزایش منظم‌سازی فضای نهان است. با این حال، افزایش بیش از حد β تا مقدار 10 باعث کاهش MIG می‌شود، زیرا مدل بیش از حد محدود شده و بخشی از توانایی بازسازی دقیق تصویر را از دست می‌دهد.

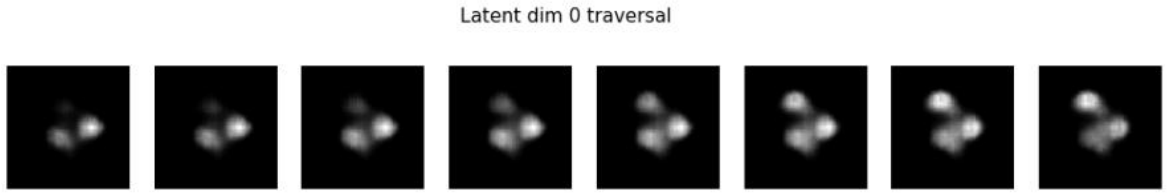
بررسی تفکیکی عوامل (per-factor) نیز نشان داد که بیشترین اطلاعات در ابعاد مربوط به موقعیت مکانی (posX و posY) متمرکز است، در حالی که متغیرهای مربوط به شکل و اندازه سهم ناچیزی دارند. این نتیجه منطقی است زیرا در داده‌های dSprites، تنوع شکل‌ها نسبت به تغییرات مکانی کمتر است.

در مجموع، می‌توان نتیجه گرفت که مدل با β=3 بهترین تعادل میان بازسازی مناسب و جدایش مؤثر ویژگی‌های نهان را برقرار کرده است و مقدار MIG در این حالت بیشترین مقدار مثبت را دارد، که با پیش‌بینی‌های نظری مقاله‌ی β-VAE سازگار است..

زیربخش هشتم



شکل 9 مقایسه‌ی بازسازی تصاویر ورودی در مدل‌های β-VAE با مقادیر مختلف β.



شکل 10 تغییر تدریجی در بُعد صفر فضای نهان (Latent traversal) برای مدل β-VAE با β=10.

در این شکل، مقدار یکی از ابعاد فضای نهان (z_0) از مقدار منفی به مثبت تغییر کرده است. مشاهده می‌شود که با تغییر این بُعد، تنها یک ویژگی از تصویر (مانند موقعیت یا شکل جسم) تغییر می‌کند در حالی که سایر ویژگی‌ها ثابت می‌مانند. این رفتار نشان‌دهنده‌ی جدایش‌پذیری (disentanglement)در نمایش نهانی مدل β-VAE است.

برای بررسی کیفی عملکرد مدل‌ها، از دو روش استفاده شد: بازسازی تصاویر ورودی و مشاهده‌ی traversal فضای نهان. در بازسازی‌ها مشاهده شد که برای$\beta=1$ ، خروجی‌ها از لحاظ بصری واضح‌تر و نزدیک‌تر به داده‌های واقعی هستند اما مدل درک تفکیک‌پذیری از فاکتورهای مختلف را ندارد. در$\beta=3$ ، بازسازی اندکی ساده‌تر شده ولی مدل ساختار منظم‌تری از ویژگی‌ها در فضای پنهان یافته است. در$\beta=10$ ، بازسازی‌ها نرم‌تر و انتزاعی‌تر هستند و برخی جزئیات از بین رفته‌اند، اما هر بُعد از فضای نهان مسئول یک ویژگی خاص تصویر (مثل موقعیت یا چرخش) شده است.

نتایج traversal نیز این الگو را تأیید می‌کنند: با تغییر تدریجی مقدار یک بُعد از z در مدل$\beta=10$ ، تنها یک ویژگی مشخص در خروجی تغییر می‌کند در حالی که سایر ویژگی‌ها ثابت می‌مانند. این رفتار نشانه‌ی واضحی از یادگیری بازنمایی disentangled است. در مجموع، هرچند افزایش β باعث افت جزئی در کیفیت بازسازی می‌شود، ولی منجر به نمایش‌های پنهان منظم‌تر و قابل‌تجزیه‌تری می‌شود که ویژگی‌های مستقل داده را به‌طور جداگانه در فضای نهان مدل می‌نمایانند.

بخش دوم

زیربخش اول

مدل VQ-VAE (Vector Quantized VAE)یکی از نسخه‌های پیشرفته VAE است که فضای نهان را از حالت پیوسته به Discrete latent spaceتبدیل می‌کند. در این روش، خروجی Encoder به‌جای یک بردار پیوسته، به نزدیک‌ترین بردار از میان مجموعه‌ای از codebook vectorsنگاشت می‌شود. به این ترتیب، مدل یاد می‌گیرد که نمایش داده‌ها را در قالب کدهای گسسته ذخیره کند.
در فرآیند آموزش، چون عمل quantization مشتق‌پذیر نیست، از تکنیک Straight-Through Estimator برای عبور گرادیان‌ها استفاده می‌شود.
مزیت اصلی VQ-VAE این است که از posterior collapse جلوگیری می‌کند و باعث می‌شود نمایش‌های نهفته معنادارتر و فشرده‌تر یاد گرفته شوند.
این ویژگی باعث شده VQ-VAE در داده‌هایی مثل گفتار و ویدئو عملکرد بهتری داشته باشد، زیرا ساختارهای سطح بالا را بهتر حفظ می‌کند. در عوض، به دلیل گسسته بودن فضای نهان، فرآیند آموزش پیچیده‌تر و حساس‌تر به انتخاب اندازه‌ی codebook است.

زیربخش دوم

مدل VampPrior توسط Tomczak و Welling برای بهبود توزیع پیشین در VAE معرفی شد.
در VAE معمولی، پرایر معمولاً یک توزیع نرمال استاندارد است $p(z) = \mathcal{N}(0, I)$که ممکن است

بیش از حد محدودکننده باشد و منجر به غیرفعال شدن برخی ابعاد نهان (Inactive latent units) شود.
درVampPrior ، پرایر به‌صورت ترکیب چند توزیع پسین تعریف می‌شود که در آن u_kها ورودی‌های مصنوعی (pseudo-inputs) قابل یادگیری‌اند.

$$p_\lambda(\mathbf{z}) = \frac{1}{K}\sum_{k=1}^{K} q_\phi(\mathbf{z}|\mathbf{u}_k),$$

شکل 11 فرمول prior در مدل vampPrior

فرمول بالا تعریف اصلی پرایر در مدل VampPriorاست و نشان می‌دهد که برخلاف VAE معمولی که از پرایر ثابت و ساده‌ای مثل نرمال استاندارد $\mathcal{N}(0, I)$استفاده می‌کند، در VampPrior پرایر به‌صورت میانگینی از چند توزیع پسین تعریف می‌شود. در این مدل، مجموعه‌ای از pseudo-inputs با نماد u_kدر نظر گرفته می‌شود که مدل در طول آموزش آن‌ها را یاد می‌گیرد. سپس برای هر u_k، انکودر توزیع $q_\phi(z \mid u_k)$را تولید می‌کند و پرایر نهایی برابر با میانگین این توزیع‌ها است.

این طراحی باعث می‌شود پرایر بتواند چندحالته (multi-modal) و منطبق‌تر با داده‌های واقعی باشد.
در نتیجه، مدل ظرفیت بیشتری برای بازنمایی ساختارهای پیچیده پیدا می‌کند، ابعاد نهان بیشتری فعال می‌شوند و بازسازی‌ها کیفیت بالاتری دارند.
البته در مقابل، مدل VampPrior از نظر محاسباتی سنگین‌تر است و نیاز به حافظه بیشتری برای ذخیره‌ی pseudo-inputs دارد.

زیربخش سوم

مدل SC-VAE ترکیبی است از اصول sparse coding وVAE ؛ یعنی فرض می‌کند که کد نهانی می‌تواند به صورت ترکیبی خطی از تعداد کمی atoms زده شود، و به کمک نسخهٔ یادگرفتنی از ISTA (Iterative Shrinkage-Thresholding Algorithm) این ترکیب را حل می‌کند .نقش ISTA در این مدل آن است که فرآیندی تکراری برای یافتن ترکیبی از اتم‌ها با مقدار آستانه‌گذاری دارد و با یادگیری پارامترهای آن، کد نهانی منظم‌تر، فشرده‌تر و قابل تفسیرتر می‌شود. مزیت اصلی SC-VAE نسبت به مدل پایه VAE این است که بازسازی دقیق‌تر و با وضوح بالاتر ارائه می‌دهد و نمایشی تنک

از داده ایجاد می‌کند که مفاهیم زیرین را بهتر جدا می‌کند. از سوی دیگر، معایب ممکن است پیچیدگی محاسباتی بیشتر، نیاز به تنظیم دقیق آستانه‌ها و اتصالات الگوریتمی بیشتر باشد.

مراجع

1. Chen, R. T. Q., Li, X., Grosse, R. B., & Duvenaud, D. (2018). Isolating sources of disentanglement in variational autoencoders. In Advances in Neural Information Processing Systems (NeurIPS 2018).

2. van den Oord, A., Vinyals, O., & Kavukcuoglu, K. (2017). Neural discrete representation learning. In Advances in Neural Information Processing Systems (NeurIPS 2017).

3. Tomczak, J. M., & Welling, M. (2018). VAE with a VampPrior. In Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics (AISTATS 2018) (Vol. 84, pp. 1214-1223). PMLR.

4. Xiao, P., Qiu, P., Ha, S., Bani, A., Zhou, S., & Sotiras, A. (2024). Sparse Coding based Variational Autoencoder with Learned ISTA (SC-VAE). Pattern Recognition, 161, Article 111187.