



به نام خدا



دانشگاه تهران
دانشکده مهندسی برق و کامپیوتر
مدل‌های مولد عمیق

تمرین اول

نام و نام خانوادگی	سید محمد جزایری
شماره دانشجویی	810101399
تاریخ ارسال گزارش	1404/08/12

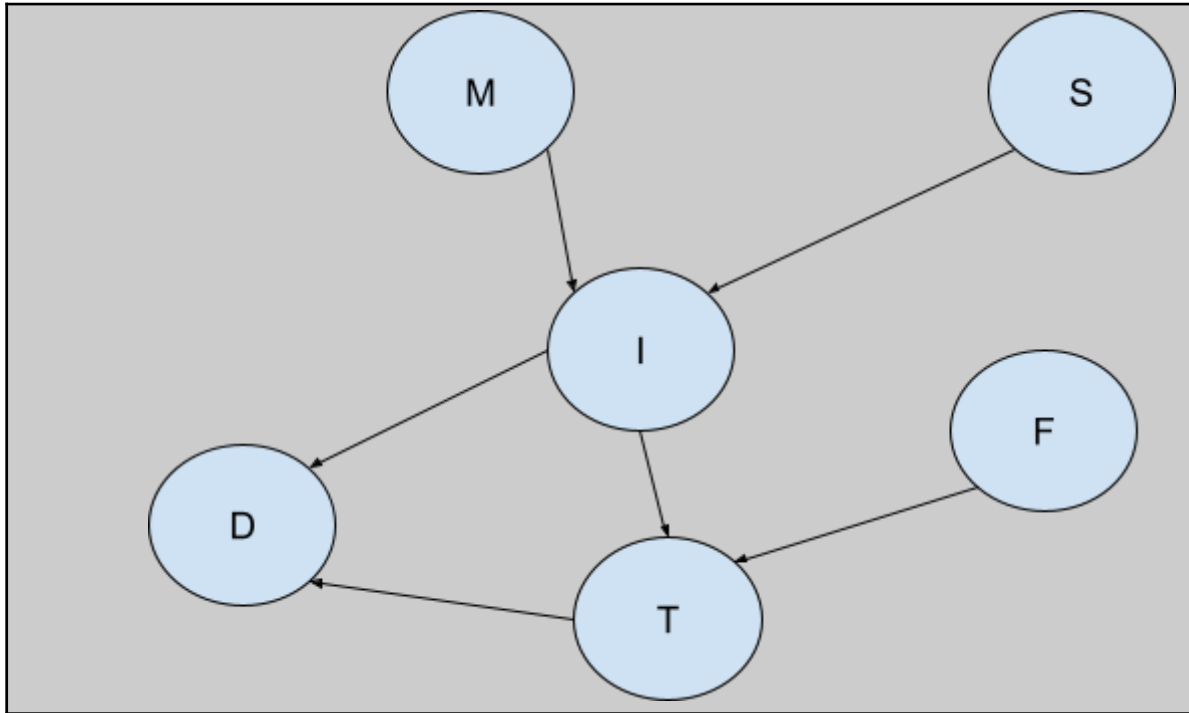
فهرست

3.....	سؤال اول.....
3.....	بخش اول.....
3.....	زیر بخش اول.....
4.....	زیر بخش دوم.....
4.....	زیر بخش سوم.....
4.....	بخش دوم.....
4.....	زیر بخش اول.....
4.....	زیر بخش دوم.....
4.....	زیر بخش سوم.....
4.....	زیر بخش چهارم.....
5.....	زیر بخش پنجم.....
5.....	زیر بخش ششم.....
5.....	بخش سوم.....
5.....	زیر بخش اول.....
5.....	زیر بخش دوم.....
5.....	زیر بخش سوم.....
5.....	زیر بخش چهارم.....
5.....	بخش چهارم.....
6.....	سؤال دوم.....
6.....	بخش اول.....
6.....	زیر بخش اول.....
6.....	زیر بخش دوم.....
6.....	زیر بخش سوم.....
6.....	زیر بخش چهارم.....
8.....	زیر بخش پنجم.....
8.....	زیر بخش ششم.....
10.....	زیر بخش هفتم.....
10.....	زیر بخش هشتم.....
11.....	بخش دوم.....
11.....	زیر بخش اول.....
11.....	زیر بخش دوم.....
12.....	زیر بخش سوم.....
13.....	مراجع.....

سؤال اول

بخش اول

زیر بخش اول



زیر بخش دوم)

$$P(M, S, I, F, T, D) = P(M)P(S)P(F)P(I|M, S)P(T|F)P(D|I, T)$$

زیر بخش سوم)

- a. نادرست. F والد T است و T والد D پس یک مسیر بین آن ها وجود دارد.
- b. درست. هر مسیری که S را به D وصل می کند حتما از I می گذرد.
- c. درست. چرا که در مسیر M-I-T-F یک V-structure در راس T وجود دارد.
- d. نادرست. استدلال مانند بخش قبل است با معلوم شدن T بهم وابسته می شوند.
- e. نادرست. در مسیر M-I-D-T یک V-structure در راس I وجود دارد.

بخش دوم

زیر بخش اول)

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

زیر بخش دوم)

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

زیر بخش سوم)

خیر چرا که $O \perp A$ در آن برقرار نیست ولی در گراف اصلی هست.

زیر بخش چهارم)

خیر چرا که حلقه S-O-T-B یک چهارتایی است که بین هیچ دو راس غیر مجاورش هیچ یالی نیست.

زیر بخش پنجم)

$$P(C, O, A, S, T, B, M) = \frac{1}{Z} \phi_1(C) \phi_2(O, S) \phi_3(B, S) \phi_4(O, T, A) \phi_5(A, T, M) \phi_6(B, T, M)$$

زیر بخش ششم)

• درست است.

$$P(O, A, S, T, B, M) = \int P(C)P(O)P(A)P(S|O)P(T|O, A)P(B|S)P(M|B, T, A)dc$$

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

چرا که هیچ عبارتی به جز $P(C)$ به C وابسته نیست و انتگرال برابر 1 است.

- درست است چرا که در حالت مارکوف توابع پتانسیل توزیع احتمال نیستند چرا که انتگرالشان یک نیست و این امر با تقسیم بر Z جبران می شود و اگر انتگرال یک تابع پتانسیل یک باشد با حذف مقدار Z هم عوض نمی شود.

بخش سوم

زیر بخش اول)

$$P(A, B, C, D, E, F, G) = \frac{1}{Z} \phi_1(A, B, C) \phi_2(B, C, D) \phi_3(C, D, F) \phi_4(F, E) \phi_5(B, E) \phi_6(E, G)$$

زیر بخش دوم)

- نا درست. این دو راس در گراف بهم مسیر دارند پس مستقل نیستند.
- نا درست. با حذف D, C همچنان مسیر $F-E-B-A$ وجود دارد.
- درست. با حذف E راس G دیگر به هیچ کدام از رئوس دسترسی ندارد.
- درست است. چرا که B, C برابر با Markov Blanket راس A هستند.

*زیر بخش سوم)

با انجام این کار مقدار احتمال نرمالایز نشده 5 برابر می شود و به همین دلیل Z هم 5 برابر می شود تا اثر آن را خنثی کند پس توزیع احتمال (ضرب توابع پتانسیل نرمالایز شده) هیچ فرقی نخواهد کرد.

بخش چهارم

با توجه به شبکه بیزی که داریم:

$$P(X, Z) = P(Z)P(X|Z)$$

$$q(Z) = KL(q(z) || P(X, Z)) = E_q[\log \frac{q(Z)}{P(Z, X)}] = E_q[\log \frac{\theta^2 z e^{-\theta z}}{z e^{-z(x+1)}}]$$

$$q(Z) = E_q[2 \log \theta - \theta z + z(x+1)] = 2E_q[\log \theta] - E_q[\theta z] + E_q[z(x+1)]$$

از آنجایی که امید ریاضی روی q است پس به جز z بقیه متغیر ها ثابت هستند. پس:

$$q(Z) = 2\log\theta - 2 * \theta * \frac{1}{\theta} + 2\frac{x+1}{\theta}$$

حالا برای یافتن θ بهینه مشتق می گیریم و مساوی صفر قرار می دهیم.

$$\frac{\partial q}{\partial \theta} = \frac{2}{\theta} - \frac{2x+2}{\theta^2} = 0 \Rightarrow \theta^* = x + 1$$

سؤال دوم

بخش اول

زیر بخش اول)

در مدل VAE ما توزیع احتمال داده ها را در اختیار نداریم بلکه توزیع توام را در اختیار داریم و محاسبه احتمال داده ها از لحاظ عملی نشدنی است. ترم دوم که فاصله فاصله توزیع فضای نهان از توزیع نرمال استاندارد را نشان می دهد و به دلیل خواص KL-Divergence مقدار کمینه اش صفر است همچنین فرض کرده ایم که z از توزیع نرمال با میانگین z می آید پس در توان این توزیع عبارت $\|x - D(z)\|^2$ قرار دارد که این همان reconstruction loss است و سعی می کند که فاصله خروجی درست را با خروجی دیکودر کم کند پس اگر این عبارت کم شود مدل بهتر عمل می کند.

زیر بخش دوم)

این دیتاست به عنوان یک محک برای مدل های مولدی که می خواهند disentanglement انجام دهند به کار می رود چرا که هر تصویر 6 بعد دارد و با تغییر مقادیر آن ها شکل ها عوض می شوند و شکل ها یا قلابند یا مستطیل یا بیضی یا دایره.

Random Sample of 8 Dsprites Images

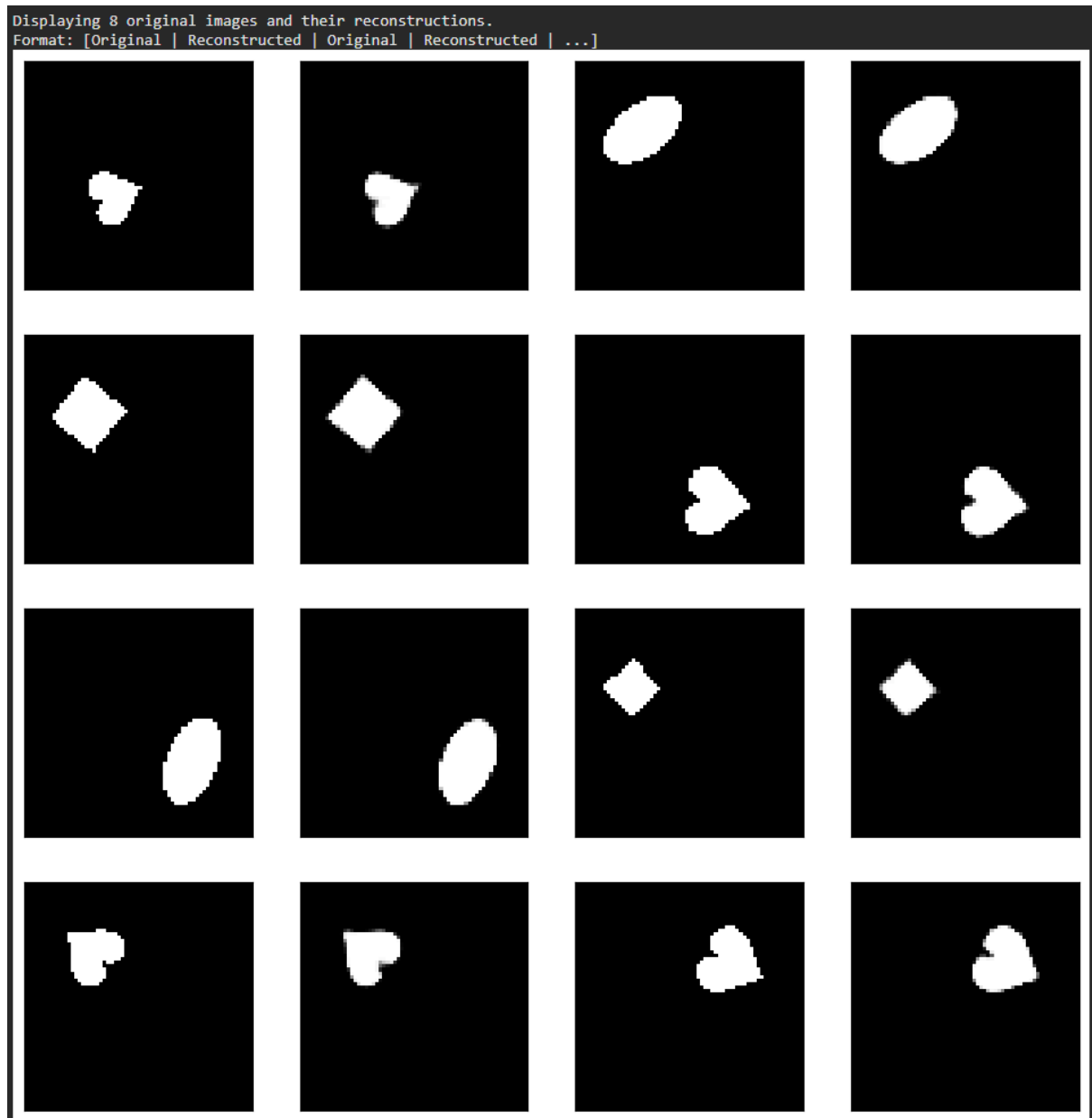


زیر بخش سوم)

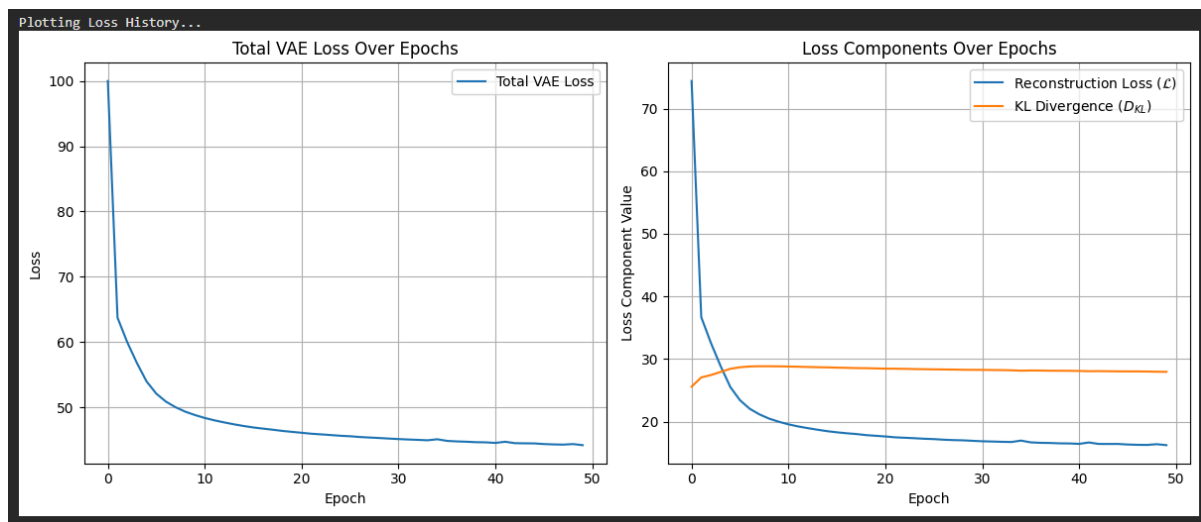
برای این کار از توزیع نرمال استاندارد یک نمونه تصادفی به نام ϵ می گیریم و آن را در انحراف معیار ضرب و با میانگین جمع می کنیم که هر دو هم خروجی های انکودر هستند اینگونه می توانیم مشتق بگیریم.

زیر بخش چهارم)

نمونه ای از تصاویر تولید شده به وسیله این مدل



نمودار هزینه

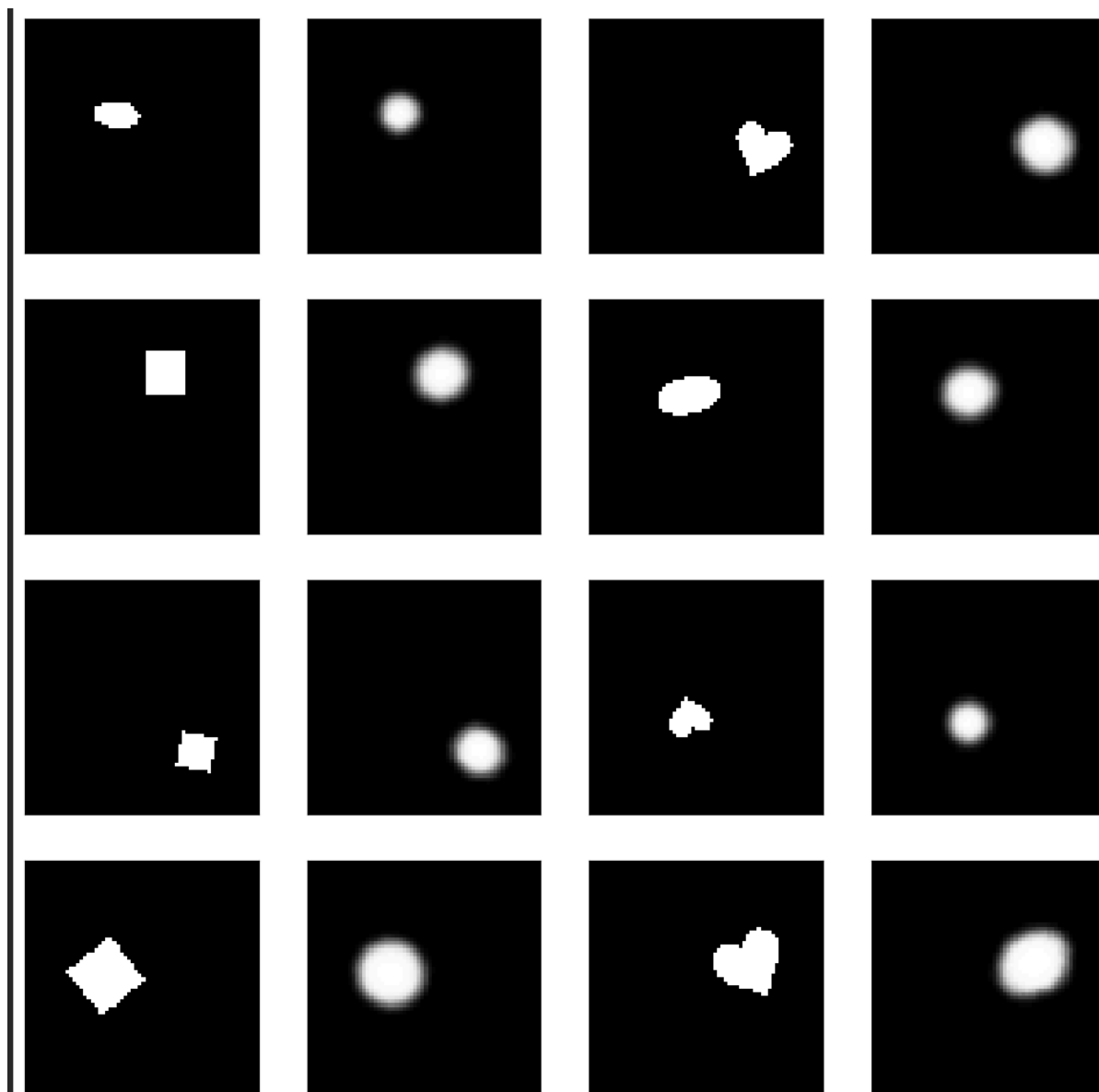


زیر بخش پنجم)

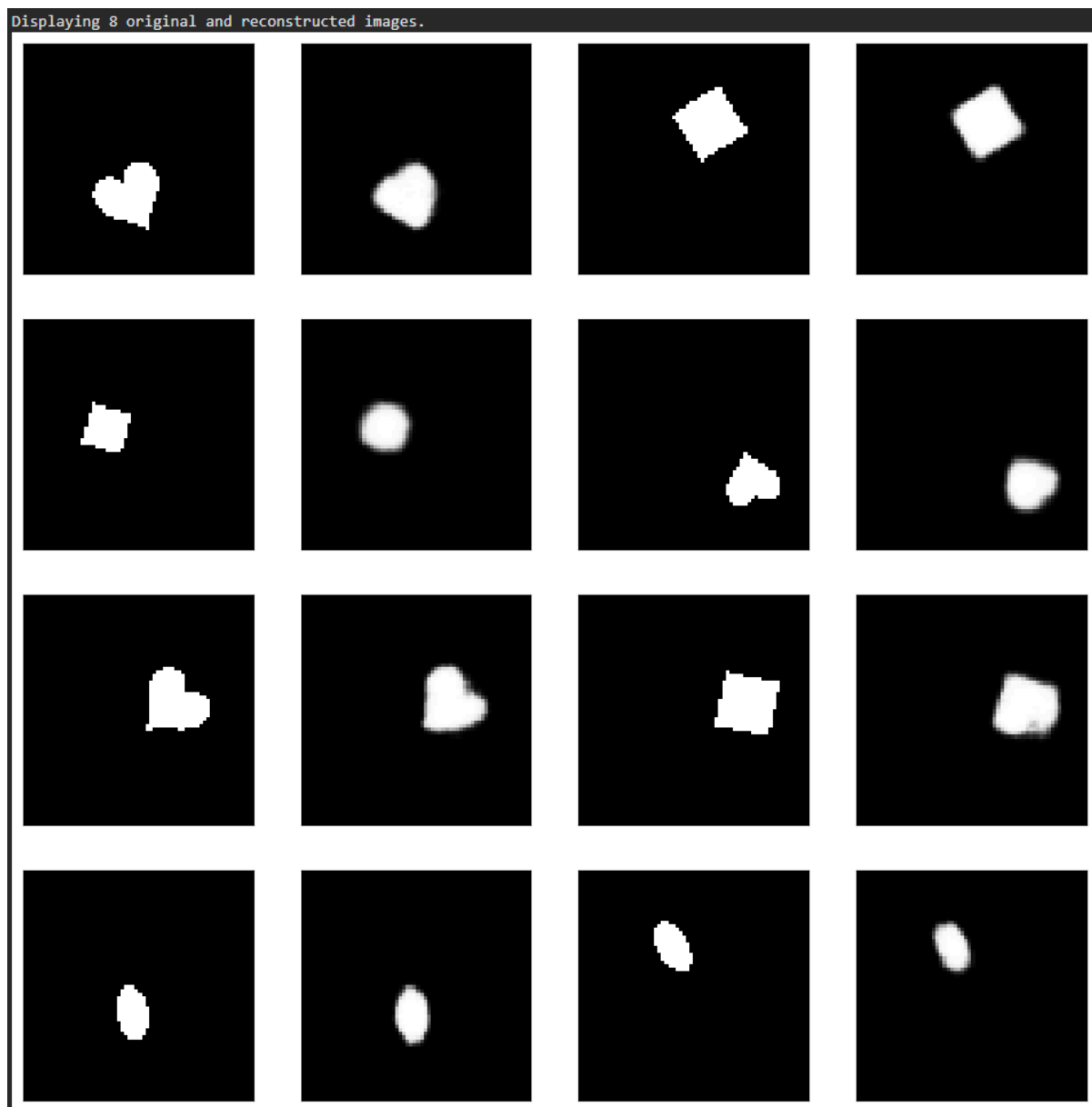
در این مدل سعی بر این است که ابعاد توزیع نهان معنا دار شوند و با اضافه کردن β در واقع بر روی نرمال شدن توزیع تاکید می کنم چرا که نرمال استاندارد نبودن آن پهنالی سنگینی خواهد داشت، حداقل سنگین تر از مدل عادی VAE.

زیر بخش ششم)

بنای 20



بنای 8



زیر بخش هفتم)

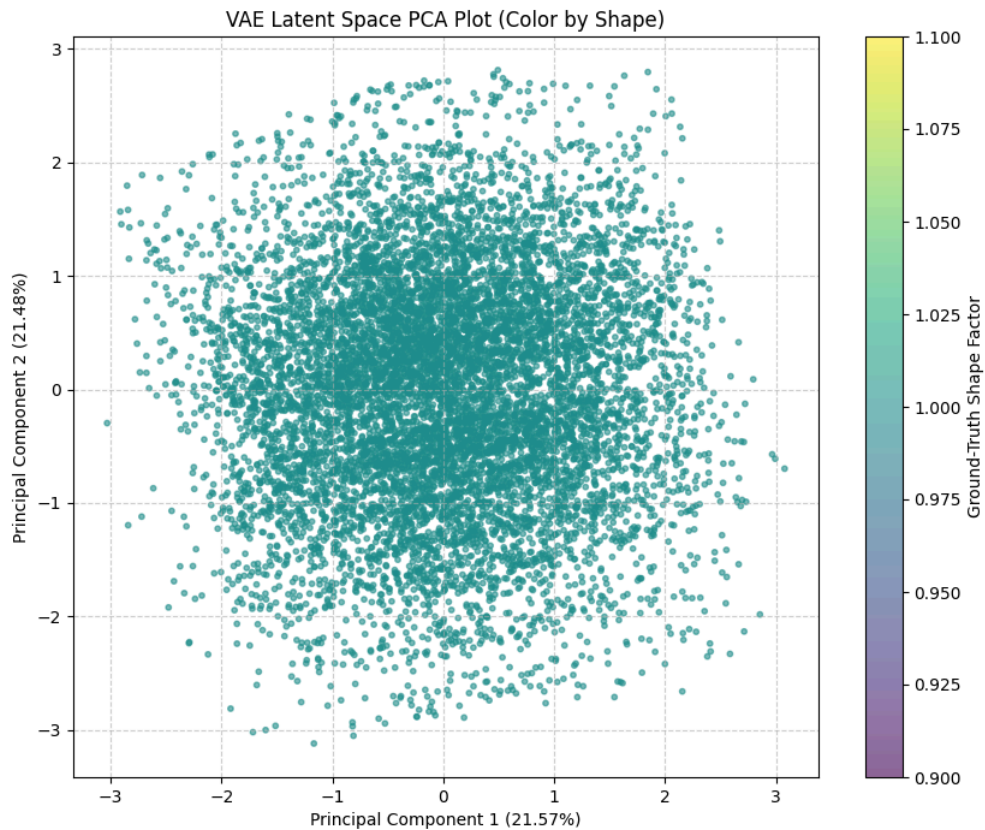
برای اندازه گیری استقلال ابعاد نهان از یکدیگر از این متریک استفاده می کنیم. معیار MIG ارزیابی می کند که یک بعد مشخص از فضای نهان (z_j) تا چه اندازه به طور انحصاری با یک عامل تغییر زمینه ای (Ground-Truth Factor) مانند شکل، مقیاس یا چرخش (v_k) هم راستا شده است. این معیار برای تعیین درجه در هم تنیدگی به دست آمده توسط یک مدل مولد طراحی شده است.

بتا برابر با 1: 0.001

بتا برابر با 8: 0.1431

بتا برابر با 20: 0.3764

زیر بخش هشتم)



همانگونه که در تصویر برای بتای 8 مشاهده می شود نمونه ها تقریباً شبیه به توزیع نرمال استاندارد در وسط پخش شده اند و همانگونه که گفته شد MIG این مدل پایین است و هر بعد تنها حدود 20 درصد داده ها را دارد و این یعنی هیچ بعد غالبی وجود ندارد پس disentanglement خیلی خوبی حاصل نشده است.

بخش دوم

زیر بخش اول)

تفاوت اصلی VQ-VAE با VAE استاندارد این است که VQ-VAE یک فضای نهان گسسته (Discrete Latent Space) را معرفی می کند در حالی که VAE استاندارد از فضای نهان پیوسته (مانند توزیع گوسی) استفاده می نماید.

این فرآیند، نقش نگاشت خروجی پیوسته رمزگذار ($z_e(x)$) را به نزدیکترین بردار (z_q) در یک مجموعه متناهی و قابل یادگیری از بردارها ایفا می کند که به آن **دفترچه کد** (Codebook) یا **فضای امبدینگ** گفته می شود.

مزایای این روش توزیع نهان ساده تر و نگاشت راحت تر خواص به ابعاد مختلف است.

زیر بخش دوم)

این روش پیشنهاد می کند که به جای توزیع نرمال استاندارد از یک mixture of Gaussians استفاده کنیم تا مدل انعطاف و قدرت مدل سازی بیشتری بگیرد پس توزیع پیشین تبدیل می شود به:

$$P(z) = \frac{1}{K} \sum_{k=1}^K P(z | u_k)$$

که u_k ورودی‌های کاذب قابل یادگیری هستند و مولفه‌های توزیع ما حول آن‌ها متمرکزند.

توزیع پیشین نرمال استاندارد اغلب برای تطبیق با توزیع پیچیده متغیرهای نهان رمزگذاری‌شده بسیار ساده است.

این عدم تطابق باعث می‌شود که جمله واگرایی KL اغلب غالب شود و نمایش‌های رمزگذاری‌شده را به سمت فروپاشی در یک منطقه کوچک از فضای نهان سوق دهد.

با استفاده از VampPrior، مدل به یک توزیع پیشین وابسته به داده دست می‌یابد که انعطاف‌پذیری بیشتری داشته و بهتر می‌تواند با پسین تجمعی واقعی مطابقت داشته باشد.

این تطابق بهتر محدودیت جمله واگرایی KL را کاهش می‌دهد و به رمزگذار اجازه می‌دهد تا فضای نهان را به‌طور موثر تر و گسترده‌تر مورد استفاده قرار دهد، که در نهایت منجر به بهبود کیفیت نمونه‌های تولیدی نسبت به توزیع پیشین گوسی استاندارد می‌شود.

زیر بخش سوم)

SC-VAE با هدف معرفی مفهوم **کدگذاری تنک (Sparse Coding - SC)** در چارچوب VAE طراحی شده است. کدگذاری تنک فرض می‌کند که داده‌ها می‌توانند به‌عنوان ترکیبی خطی از تعداد کمی از بردارهای پایه نمایش داده شوند که ضرایب این ترکیب اکثراً صفر هستند (تنک).

الگوریتم آستانه‌گذاری انقباضی تکراری (ISTA) یک الگوریتم متداول برای حل مسئله کدگذاری تنک (یافتن ضرایب تنک) است.

در SC-VAE، نویسندگان شبکه رمزگذار را به‌گونه‌ای طراحی می‌کنند که مراحل تکراری ISTA را تقلید کند. به‌طور خاص، شبکه استنتاج به‌عنوان یک شبکه ISTA (**L-ISTA** یادگرفته شده) ساخته می‌شود.

این طراحی، رمزگذار را مجبور می‌کند تا نمایش‌های نهان تولید کند که ذاتاً **تنک** هستند.

نمایش‌های تنک اغلب قابل تفسیرتر هستند، زیرا تنها تعداد کمی از ویژگی‌ها برای هر ورودی فعال می‌شوند، که درک اینکه کدام عوامل نهان، بازسازی را هدایت می‌کنند، آسان‌تر می‌سازد.

کدهای تنک اغلب در برابر نویز و ویژگی‌های نامربوط در داده‌های ورودی قوی‌تر هستند.

