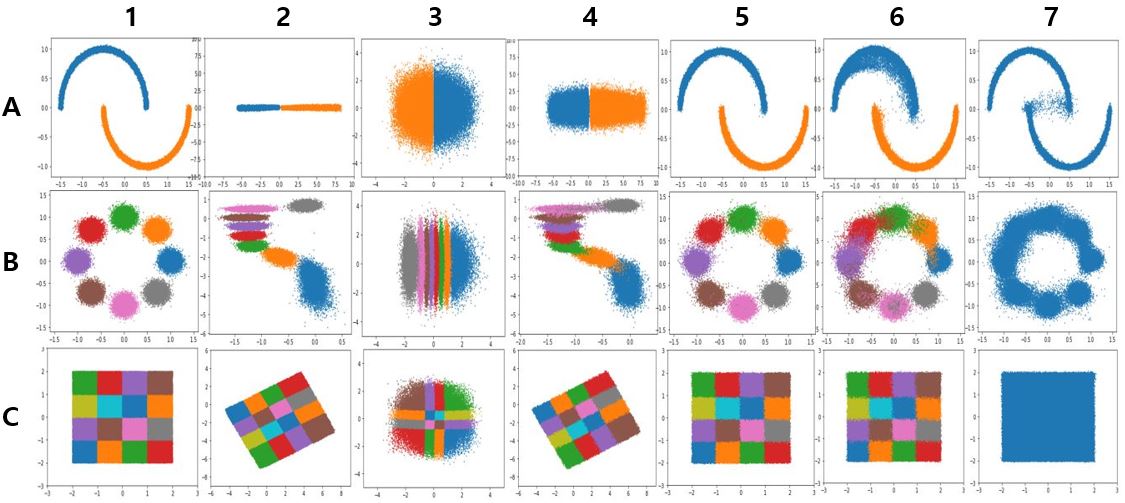
**실험**

**실험 1**

실험 1은 관측변수 X, 중간변수 M, 잠재변수 Z의 변분 분포와 실제 분포를 시각화 하는 실험입니다. 이 실험에서는 서로 다른 세개의 데이터셋이 사용되었습니다. 세 데이터셋 모두 2차원의 크기를 가졌으며, 각각의 데이터집합에 서로 다른 모델이 사용되었습니다. 각각의 모델은 시각화를 위해 중간변수 M, 잠재변수 Z 모두 2차원의 크기를 갖도록 설정하였습니다.



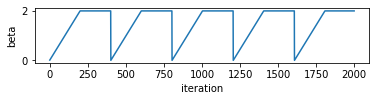
[그림 3]

[그림 3]은 실험 1에 대한 그림입니다. 각각의 행은 서로 다른 데이터 집합의 결과를 나타냅니다. 데이터들의 색상은 위치를 식별하기 위해 시각화 과정에서 부여한 것일 뿐 모델의 학습에 이용된 정보는 아닙니다. 1열은 데이터 셋, 2열은 외부 인코더의 결과의 표본 , 3열은 내부 인코더의 결과의 표본 , 4열은 내부 인코더로부터 입력을 받았을 때의 외부 인코더의 결과의 표본 를 나타냅니다. 5, 6, 7열은 모두 외부 디코더의 결과물입니다. 5열은 외부 인코더로 입력을 받았을 때의 외부 디코더의 결과의 표본 , 6열은 내부 인코더의 으로부터 내부 디코더를 거친 외부 디코더의 결과의 표본 , 7열은 실제 사전분포 로부터 추출된 모델이 학습한 실제 분포의 표본 를 나타냅니다.

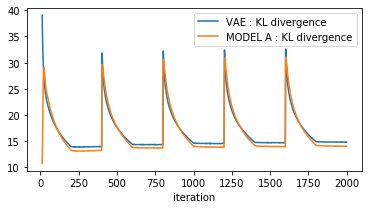
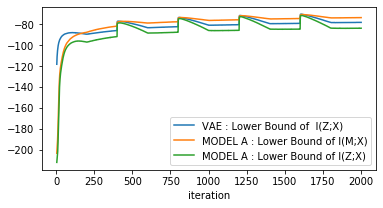
3열을 통해 본 모델이 분리 표현 학습을 잘 수행하는 것을 확인할 수 있습니다. 2열과 4열을 보면 와 ​가 서로 매칭되며, 특히 A행에서는 앞서 언급한 쿨백-라이블러 발산항의 특징이 잘 나타나는 것을 관찰할 수 있습니다.

**실험 2**

실험 2는 학습 과정에서 모델A의 음의 복원 손실항과 잠재채널의 실제 사전분포와의 쿨백-라이블러 발산 항의 변화 양상을 보입니다. 비교를 위해 같은 조건에서 학습된 β-VAE의 모델의 결과를 같이 나타냈습니다. 실험은 MNIST 데이터셋을 이용하였고, 각각의 잠재벡터의 크기는 8, 모델A의 중간벡터의 크기는 8을 설정하였습니다.



[그림 4]



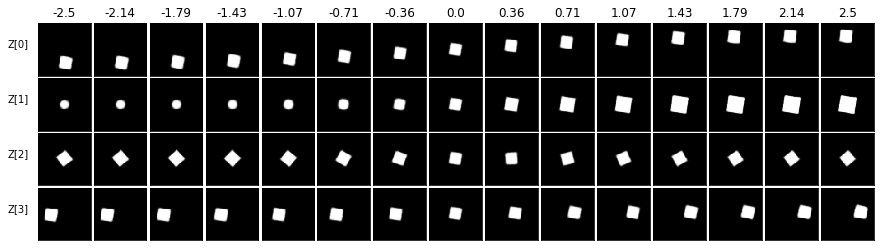
[그림 5]

[그림 4]는 두 모델의 학습에 사용된 β 순회 어닐링 스케줄입니다. [그림 5]의 왼쪽 그림은 두 모델의 복원 손실항을 나타냅니다. 연두색 선은 모델 A의 학습 과정에서 계산된 항, 주황색 선은 항의 평균을 나타냅니다. 하늘색 선은 비교를 위해 따로 학습된 β-VAE모델의 학습 과정에서 계산된 항의 평균을 나타냅니다. 주황색 선은 외부 인코더의 차원 축소 학습의 성능을 나타낸다고 할 수 있습니다. 실험 결과 이는 실제로 기존의 β-VAE의 복원성능보다 조금 더 좋은 성능을 보였습니다. 반면 모델A의 실제 복원 능력에 해당한다고 할 수 있는 연두색 선은 기존의 β-VAE보다 조금 뒤떨어지는 것으로 확인되었습니다. 그러나 복원 손실항이 높다는 것이 분리 표현 능력이 뒤떨어진다는 것을 의미하는 것은 아니며 오른쪽 그림에서 확인할 수 있듯 잠재채널과 실제 사전분포와의 쿨백-라이블러 발산 항이 모델A가 근소하게 더 작다는 점에서 모델A가 기존의 β-VAE에 비하여 뒤떨어진다고 볼 수는 없을 것입니다.

**실험 3**

실험 3은 두 가지 데이터셋(dSprites, 3D chairs)을 학습한 모델의 생성 모델의 결과를 보이는 실험입니다.

**실험 3-1:**



[그림 5]

실험 3-1은 dSprites 데이터 셋에 대한 실험입니다. dSprites 데이터 셋은 x, y축 위치, 물체의 크기, 물체의 회전 각도, 물체의 종류(사각형, 타원, 하트) 총 다섯 가지의 생성 요인이 있습니다. 본 실험에서는 사각형 데이터만을 이용하였습니다. 중간벡터와 잠재벡터의 크기는 4, β는 0부터 10까지 순회 어닐링 스케줄링을 사용하였습니다. [그림 5]는 모델A의 생성 결과를 나타냅니다. 각각의 행은 나머지 잠재변수들이 고정된 채 해당 행의 왼쪽에 표시된 잠재변수의 변화에 따른 생성물의 결과를 나타냅니다. 각 열의 위쪽 값은 변화되는 잠재변수의 값을 나타냅니다. 3번째, 0번째 잠재변수는 각각 x, y축의 위치를 표현하는 것으로 보이며, 1번째 잠재변수는 크기, 2번째 잠재변수는 회전 각도를 표현한 것으로 보입니다.

**실험 3-2:**

**결론**

ㅂ회전 각도, 물체의 종류(사각형, 타원, 하트) 총 다섯가지의 생성 요인이 있습니다. 본 실험에서는 사각형 데이터만을 이용하였습니다. 중간벡터와 잠재벡터의 크기는 4, β는 0 부터 10까지 순회 어닐링 스케줄링을 사용하였습니다

**참고 문헌**

[1] Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep variational information bottleneck. International Conference on Learning Representations, 2016.

[2] Christopher P. Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins and Alexander Lerchner. Understanding disentangling in -VAE, 2017.

[3] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed and Alexander Lerchner. -VAE: Learning Basic Visual Concepts with a Constrained Variational Framework, 2017.

[4] Alexander A. Alemi, Ben Poole, Ian Fischer, Joshua V. Dillon, Rif A. Saurous and Kevin Murphy. Fixing a Broken ELBO, 2017.

[5] Ricky T. Q. Chen, Xuechen Li, Roger Grosses and David Duvenaud. Isolating Sources of Disentanglement in VAEs, 2018.

[6] Matthew D. Hoffman and Matthew J. Johnson. ELBO surgery: yet another way to carve up the variational evidence lower bound, 2016.

[7] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes, 2013.

[8] Hao Fu, Chunyuan Li, Xiaodong Liu, Jianfeng Gao, Asli Celikyilmaz and Lawrence Carin. Cyclical Annealing Schedule: A Simple Approach to Mitigating KL Vanishing, 2019.

[9] Adji B. Dieng, Yoon Kim, Alexander M. Rush and David M. Blei. Avoiding Latent Variable Collapse With Generative Skip Models, 2019.

**부록**

**A:**

(A-1)

(A-2)

(A-3)

(A-4)

(A-1)의 항은 데이터 분포 고유의 엔트로피입니다. 따라서 상수 취급하여 무시됩니다. (A-2)의 부등식은 쿨백-라이블러 발산함수는 언제나 0 이상이라는 성질(non-negativity)에 근거하여 식 로부터 유도됩니다.

**B:**

N은 데이터 집합의 크기입니다. K는 데이터 집합의 인덱스 변수입니다. 를 만족하는 이산 균일 분포를 따르며 각각의 확률 변수들과 의 관계를 갖습니다.

(B-1)

(B-2)

(B-3)

(B-4)

(B-5)

= (B-6)

여기서 k에 의해 x가 특정된다는 점에서 =라는 점이 활용됩니다. (B-2)의 부등식은 젠센 부등식(Jensen inequality)으로 부터 유도됩니다.

A, B의 증명은 논문 [1]을 참고하였습니다.

**C:**

(

(

(

(

(

(

**D:**

(

(

(

(

(

**E:**

E는나중

**F:**