변이형 오토인코더의 최적화 문제 연구

고영민[1], 고선우[2]
[1]전주대학교 인공지능연구소
[2]전주대학교 인공지능학과, 인공지능연구소
gjtrj55@naver.com, godfriend0@gmail.com

요약: 생성모델의 목적은 관심 있는 데이터 전체인 모집단, 즉 분포를 알아내어 그 분포로부터 데이터를 생성하는 것이다. 생성모델 중 하나인 변이형 오토인코더는 비선형차원축소 방법 중 하나인 오토인코더를 적용한 생성모델이다. 기존에 연구된 변이형 오토인코더의 목적함수는 관측 가능한 변수가 주어졌을 때, 잠재 변수의 확률인 사후분포를 근사화하는 변분추론에 기반한 ELBO(Evidence Lower BOund)을 사용하고 있다. 이 목적함수는 복원오차와 잠재 분포 오차를 동일한 비율로 반영한다. 본 연구는 생성모델의 목적에 맞는 변이형 오토인코더의 최적화 문제를 정의한다. 정의된 최적화 문제에서 유도된 라그랑지안 승수에 따라 복원오차와 잠재 분포 오차의 중요도가 달라진다. 그리고 정의된 최적화 문제를 실제로 계산하기 위한 구체적인 목적함수를 유도하였다.

주제어: 생성모델, 변이형 오토인코더, 라그랑지안 승수

A Study on the Optimization Problem of Variational AutoEncoder

Young-Min Ko[1], Sun-Woo Ko[2*]
[1]Jeonju University Artificial Intelligence Research Center
[2]Department of Artificial Intelligence and Artificial Intelligence Research Center
Jeonju University

1. 변이형 오토인코더의 최적화 문제 정의

생성모델의 목적은 우리가 관심 있는 데이터 전체인 모집단, 즉 분포를 알아내어 그 분포로부터 데이터를 생성하는 것이다.

생성모델 중 하나인 변이형 오토인코더는 비선형차원축소 방법인 오토인코더[1]를 적용한 방법이다. 먼저 오토인코더의 최적화 문제를 정의하면 다음 식 (1)과 같다.

$$\begin{aligned} &\min L(Z,\tilde{Z}) \\ &s.t.\ \tilde{Z}= f_d(f_e(Z)) \end{aligned} \tag{1}$$

식 (1)에서 Z은 관측된 데이터들을 의미하며 $\tilde{Z}$은 Z을 입력으로 받은 인코더 함수 f_e와 디

코더 함수 f_d을 통해 나온 복원된 데이터를 의미한다. 즉 오토인코더는 관측된 데이터와 f_e, f_d을 통해 나온 복원된 데이터로 정의되는 손실함수를 최소화하는 문제이다.

이런 오토인코더를 적용하여 생성모델로 사용하는 일반적인 변이형 오토인코더의 목적함수 식 (2)는 관측 가능한 확률변수 $\boldsymbol{z}$가 주어졌을 때 잠재변수 $\boldsymbol{z}^c$가 일어날 사후분포인 조건부 확률분포 $p(\boldsymbol{z}^c|\boldsymbol{z})$을 다루기 쉬운 분포 $q(\boldsymbol{z}^c)$으로 근사화하는 변분추론을 기반으로 한다[2].

$$D_{KL}\left[q(\boldsymbol{z}^c) \| p(\boldsymbol{z}^c|\boldsymbol{z})\right] \tag{2}$$

하지만 생성모델의 목적을 생각하여 변이형 오토인코더를 고려하면

1. 먼저 목적함수는 우리가 관심 있는 관측 가능한 확률변수 $\boldsymbol{z}$을 최대한 잘 복원하는 $\tilde{\boldsymbol{z}}$을 만들어야 한다. 이 경우 개념적으로 위험도 최소화(Risk minimization)를 해야 한다[3].

2. 첫 번째 제약 조건으로 오토인코더를 사용하여 $\tilde{\boldsymbol{z}}$을 만든다.

3. 두 번째 제약 조건으로 다루기 쉬운 잠재 분포에서 데이터를 샘플링하기 위해 잠재 분포의 차이를 비교하는 $D_{KL}\left[q(\boldsymbol{z}^c) \| p(\boldsymbol{z}^c|\boldsymbol{z})\right]$의 값이 어떤 상수 δ값 보다 작아야 한다.

위와 같이 세 가지를 고려한 변이형 오토인코더의 최적화 문제를 정의하면 다음 식 (3)으로 표현된다.

$$\begin{gathered} \min E\left[L(\boldsymbol{z}, \tilde{\boldsymbol{z}})\right] \\ s.t.\ \tilde{\boldsymbol{z}} = f_d\left(f_e(\boldsymbol{z})\right) \\ D_{KL}\left[q(\boldsymbol{z}^c) \| p(\boldsymbol{z}^c|\boldsymbol{z})\right] \le \delta \end{gathered} \tag{3}$$

식 (3)에서 $E\left[L(\boldsymbol{z}, \tilde{\boldsymbol{z}})\right]$은 $\boldsymbol{z}$에 대한 위험도 최소화를 나타내며 E은 $L(\boldsymbol{z}, \tilde{\boldsymbol{z}})$의 기댓값을 나타낸다. 위와 같이 제약조건이 있는 최적화 문제는 라그랑지안 승수법을 적용하여 제약조건이 없는 문제로 변환할 수 있으며 식 (4)와 같다.

$$\begin{gathered} \min E\left[L(\boldsymbol{z}, \tilde{\boldsymbol{z}})\right] + \alpha\left(D_{KL}\left[q(\boldsymbol{z}^c) \| p(\boldsymbol{z}^c|\boldsymbol{z})\right] - \delta\right) \\ s.t.\ \alpha \ge 0 \end{gathered} \tag{4}$$

식 (4)에서 α는 라그랑지안 승수를 나타내며 식 (3)의 제약조건에 의해 0과 같거나 커야 한다. 정리하여 개념적으로 변이형 오토인코더의 최적화 문제는 식 (4)으로 유도된다.

2. 실제 계산상 목적함수 유도

앞서 정의된 변이형 오토인코더의 최적화 문제를 실제로 계산할 때 두 가지 문제점이 발생한다. 첫 번째, $E\left[L(\boldsymbol{z}, \tilde{\boldsymbol{z}})\right]$에서 $L(\boldsymbol{z}, \tilde{\boldsymbol{z}})$의 기댓값을 구할 때 정확한 확률을 모른다. 이때 분포를 가정하거나 n개의 관측된 데이터 Z의 표본평균을 사용하여 근사화하는 몬테카를로 방법[4]이 있다. 몬테카를로 방법과 손실함수 L을 l2-norm을 사용할 경우 식 (5)로 표현할 수 있다. 식

(5)에서 $\|\ \|_F^2$은 프로베니우스 놈을 의미한다.

$$E[L(z,\tilde{z})] = \int L(z,\tilde{z})p(z)dz \approx \frac{1}{n}\| Z - \tilde{Z} \|_F^2 \tag{5}$$

두 번째, $D_{KL}[q(z^c)\|p(z^c|z)]$을 실질적으로 계산할 때 ELBO(Evidence Lower BOund)을 사용하며 Kingma와 Welling[2]이 유도한 식은 식 (6)과 같다.

$$ELBO = \frac{1}{n}\| Z - \tilde{Z} \|_F^2 + D_{KL}[N(\hat{u},\hat{\sigma})\|N(\mathbf{0},I)] \tag{6}$$

식 (6)에서 $(1/n)\| Z - \tilde{Z} \|_F^2$ 항은 복원오차를, $D_{KL}[N(\hat{u},\hat{\sigma})\|N(\mathbf{0},I)]$ 항은 잠재 분포의 차이를 의미한다. $N(\hat{u},\hat{\sigma})$는 $q(z^c)$을 정규분포로 가정하고 $\hat{u},\hat{\sigma}$은 Z가 주어졌을 때 인코더 함수를 통해 구한 추정량을 의미하며 $N(\mathbf{0},I)$은 사전분포 $p(z^c)$을 가정한 것이다.

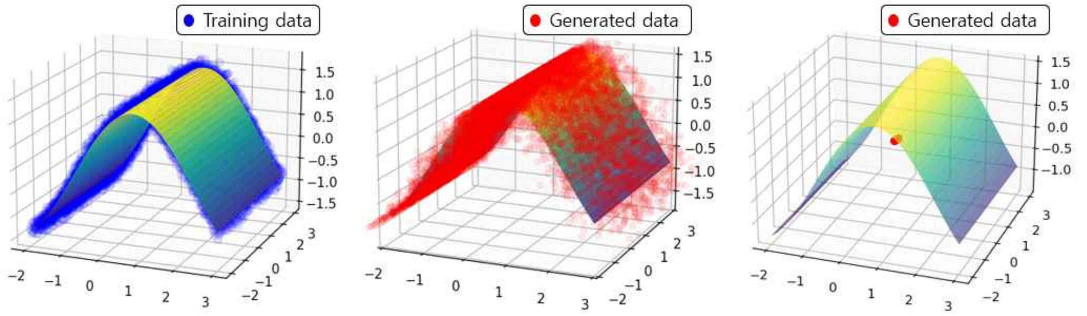
이제 실질적으로 계산해야하는 최적화 문제는 식 (7)과 같이 유도된다.

$$\begin{aligned} &\min\,(1+\alpha)\frac{1}{n}\| Z - \tilde{Z} \|_F^2 + \alpha\left(D_{KL}[N(\hat{u},\hat{\sigma})\|N(\mathbf{0},I)] - \delta\right) \\ &s.t.\ \alpha \geq 0 \end{aligned} \tag{7}$$

기존에 연구되고 있던 변이형 오토인코더의 목적함수는 ELBO(식 (6))이지만 생성모델의 목적에 맞춰 정의된 변이형 오토인코더의 최적화 문제는 식 (7)과 같이 정의할 수 있다. ELBO와의 주된 차이점은 복원오차항에 $(1+\alpha)$, 잠재 분포 차이항에 α 계수가 붙는다는 점이다. 이 의미는 복원오차와 잠재 분포 오차의 비율이 다르다는 것을 나타내며 만약 α값이 0에 가까울 경우 복원오차에 대한 중요성을 강조하는 것이고, α값이 충분히 크게 형성되면 복원오차와 잠재분포 차이의 비중을 비슷하게 보는 것으로 해석할 수 있다.

3. 실험

생성모델의 목적에 맞춰 정의된 식 (7)을 학습하여 생성모델의 목적에 부합하는지 보기 위해 기존의 ELBO(식 (6))를 최적화하는 목적함수와 비교하여 실험을 진행하였다.



(A) (B) (C)

Figure 1. (A) Data used in the experiment, (B) Data distribution generated by sampling in the latent space after learning the expression defined according to the purpose of the generative model, (C) Data distribution generated by sampling in the latent space after learning ELBO.

실험에 사용된 데이터는 3차원에 형성되어 있는 간단한 코사인 함수에서 구간을 설정하여 약간의 노이즈를 주어 10,000개 데이터를 샘플링하였다[Figure 1 (A)]. 신경망의 구조는 인코더와 디코더 모두 은닉층을 1개 가지고 그 노드 수가 32개, 파라메트릭 하이퍼볼릭 탄젠트 함수를 사용하고 2차원 잠재차원을 가지도록 설계하였다. 구체적으로 식 (7)을 학습할 때 상수 δ값은 적당히 작은 0.1을 주었고 신경망의 파라미터들과 α값은 동시에 학습하였다. α값이 음수가 되는 경우를 방지하기 위해 α값이 음수가 되는 경우, 기울기를 반대로 바꾸어 학습하는 방법을 사용하였다. 제안된 식 (7)과 ELBO를 200에폭으로 각 10회 학습 후 손실함수의 평균과 표준편차 결과를 Table 1에 나타내었다.

Table 1. Experiment result(mean±std)

	Reconstruction error	Latent distribution error	$(1+\alpha)\times$Reconstruction error	$\alpha\times$Latent distribution error	α(initial value=1)
Proposed objective function	0.009±0.002	56832±2740	0.009±0.002	10.53±13.9	0.00018±0.0002
ELBO	1.67±0.0005	0.028±0.025	X	X	X

Table 1의 ELBO의 결과는 복원오차 평균값 1.67보다 잠재 분포 오차 평균값 0.028이 훨씬 작은 것을 볼 수 있다. 그 중 ELBO 목적함수로 학습된 하나의 변이형 오토인코더에 대해 2차원 잠재공간에서 표준정규분포로 10,000개의 데이터를 샘플링하여 디코더를 통해 데이터를 생성한 결과는 Figure 1 (C)이며, 복원오차가 크므로 데이터 생성이 잘 되지 않은 것을 볼 수 있다. 반대로 식 (7)을 학습한 경우에 복원오차 평균값이 0.009로 ELBO의 복원오차보다 훨씬 작

은 것을 볼 수 있다. 특히 상당히 큰 잠재 분포 오차 값을 α값이 상쇄하여 $\alpha \times$잠재 분포 오차 평균값이 10.53으로 작아진 것을 확인할 수 있다. 마찬가지로 2차원 잠재공간에서 표준정규분포로 10,000개 데이터를 생성한 그림은 Figure 1 (B)이며 ELBO를 학습한 경우보다 원함수에 가까운 데이터를 생성한 것을 볼 수 있다.

REFERENCES

[1] H. Bourlard and Y. Kamp, "Auto-association by multilayer perceptrons and singular value decomposition", Biological Cybernetics, 59, pp.291-294, 1988.

[2] Diederik P. Kingma and Max Welling, "Auto-Encoding Variational Bayes", DOI: https://doi.org/10.48550/arXiv.1312.6114, 2014.

[3] Vladimir Vapnik, "Principles of risk minimization for learning theory", NIPS 1991.

[4] N. Metropolis, "The Beginning of the Monte Carlo Method," Los Alamos Science Special Issue, Vol. 15, pp.125-130, 1987.