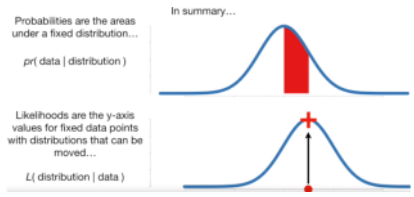
개인정리-VAE

2024년 10월 31일 목요일 오후 1:49

1. 가능도(Likelihood)

■확률 : 주어진 확률분포에서 관측값이 분포 안에서 얼마의 확률로 존재하는 가
-확률 = P(관측값 X | 확률분포 D)
-확률은 확률분포가 주어졌을 때 관측값 x에 대한 확률을 구하는 것

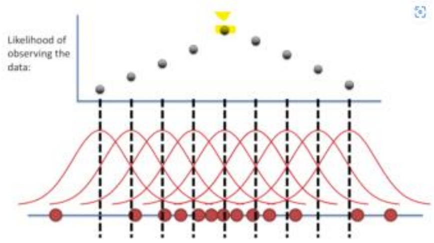
■가능도: 어떤 관측값이 어떤 확률분포로 부터 나왔는지에 대한 확률
-고정되는 값은 확률분포 D가 아닌 관측값 X임
-가능도 = L(확률분포D | 관측값 X)



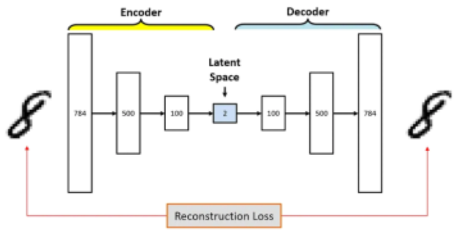
■최대 가능도 추정(Maximum Likelihood Estimation
-각 관측값 x에 대한 총 가능도(모든 가능도의 곱)이 최대가 되게하는 확률 분포
-즉 모든 관측값들이 있을 수 있는 확률분포를 구하는 것
-MLE 시 먼저 임의의 확률분포를 가정해야 함

-예를 들어 이러한 관측값이 존재할 경우, 이렇게 관측될 가능성이 가장 높은 확률분포를 찾는다고 하면

-이런식으로 총 가능도가 가장 높은 확률분포가 최대 가능도가 될 것이다.
-또한 이 때 모든 관측값의 평균은 MLE를 통해 찾은 평균과 같다.



2. AE, AutoEncoder



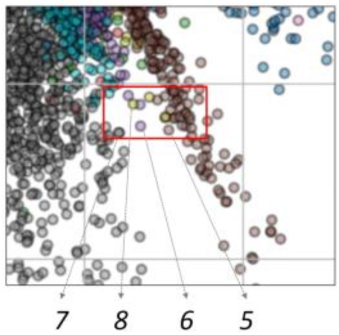
Autoencoder Architecture (Image by the author)

■기본 구조
-입력 x가 들어가면 x가 그대로 나오게 하는 것이 AE의 목적
-Encoder + Latent space + Decoder

★ 인코더에서는 입력레이어가 매우 작은 차원으로 프로젝션 되도록 함
-이 과정에서 고차원의 데이터가 작은 데이터로 모델링됨 (차원 축소)

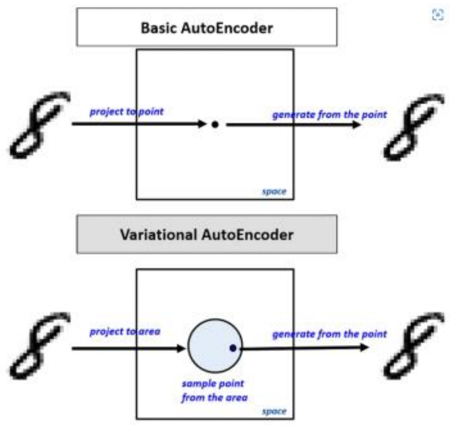
3. VAE, Variational Autoencoder

■AE의 문제점
a. 잠재공간의 해석이 어려움
b. 잠재공간이 연속적이지 않음
c. 생성된 데이터의 품질이 낮음



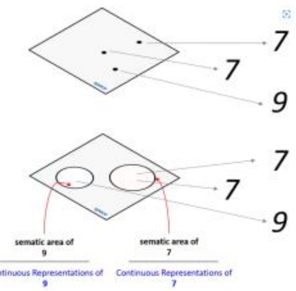
> 위 그림은 MNIST를 학습한 AE의 잠재공간을 2D 산점도로 표현한 것
> 보다 싶히 각 레이블이 even하게 군집화되지 않았고, 잘 scatter되지 않았음

■VAE의 등장: 점에서 영역으로



Differences in latent space construction between an autoencoder and a variational autoencoder (image by the author)

-기존의 AE는 인코더가 학습으로 배워낸 잠재공간이 개별 포인트로 고정됨(latent vector)
(예를 들면, MNIST의 8을 인코더가 학습하면 디코더는 그 특정 포인트를 학습하는 것)
-VAE는 이 잠재공간을 사람이 해석하기 좋은 형태로 가정
 -> 같은 숫자들의 데이터 포인트들은 같은 공간에 모여서 영역을 형성한다

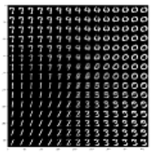


■가우시안 분포를 사용하기
-인간이 가장 해석하기 좋은 영역은 가우시안 분포(정규 분포)
-VAE는 가우시안 분포를 통해 잠재공간을 나누어 데이터 포인트들을 연속적 확률 분포로 나타냄

a. 가우시안은 **평균**과 **표준편차**를 통해서 확률분포를 그릴 수 있음
> 확률분포 그래프는 2차원 처럼 보이지만 실제로는 2차원이 아닌 1차원 데이터의 확률밀도함수(PDF)임
> 1차원 PDF는 x축에 값, y축에 그 값의 확률 밀도를 나타냄

■VAE의 예시 : MNIST

a. 각 숫자 이미지에 대해서 인코더는 평균과 표준편차를 구하여 가우시안 분포를 생성
b. 생성된 가우시안 분포는 잠재공간(2차원 가우시안 분포를 따름) 위에 매핑됨
 > 이 때의 분포는 너무 작아서 점 수준
c. 각각의 이미지들의 분포가 잠재공간에 올려져서 영역을 형성함
 > 비슷한 이미지들은 비슷한 위치에 모여서 영역을 이룰 것임
d. 생성된 잠재영역에서 랜덤한 일부 부분을 샘플링 하면



이렇게 되는 것임
e. 학습된 평균과 분산을 기반으로 하되, 그 분포 내에서 랜덤하게 샘플링

■Log Variance 사용

$$\sigma = \exp(\log(\sigma)) = \exp\left(\frac{2\log(\sigma)}{2}\right)$$
$$= \exp\left(\frac{\log(\sigma^2)}{2}\right) = \exp(0.5 * \text{log\_variance})$$

■VAE의 손실함
reconstruction error + KL Divergence

$$\mathcal{L}_{\text{VAE}}^{\text{CE}} = \mathcal{L}_{\text{recon}}^{\text{CE}} + \beta\mathcal{L}_{\text{KL}}$$

1) 재구성 오차
-인코더의 입력과 디코더가 재구성한 데이터의 차이를 나타내는 손실함수
-MSE나 Binary Crossentropy 같은 거 사용

$$\mathcal{L}_{\text{recon}}^{\text{CE}} = -\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right]$$

2) KL Divergence

$$\text{KL Divergence} = D_{KL}\left[q(z|x) \| p(z)\right]$$

-하나의 확률분포 q 가 다른 확률분포 p로 변하기 위해 필요한 양(혹은 차이)
-q(z|x)를 p(z)로 맞추기 위해서 필요한 정보

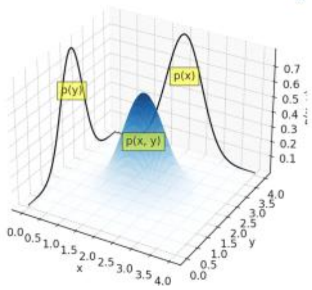
-인코더가 출력한 잠재변수의 분포 : q(z|x)
-표준 정규 분포: p(z)

-즉 VAE의 잠재 공간이 표준정규분포로부터 얼마나 떨어져있는가를 나타내는 지표

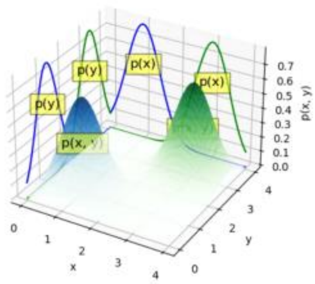
b. 이를 2차원 영역으로 모델링 하려면 x축과 y축의 어느 점에 영역이 그려질지를 결정해야 함

> 2차원 확률분포의 중심을 구하려면 x축의 평균과 y축의 평균이 모두 필요
> 2차원 확률분포가 얼마나 퍼지는지를 구하려면 x축의 분산과 y축의 분산이 모두 필요
>>> 즉 x_평균, y_평균, x_표준편차, y_표준편차 4개의 파라미터가 필요함

c. 위의 파라미터를 통해 그려진 2차원 확률분포의 그림은 다음과 같다



d. 2차원 잠재공간 위에 2개의 의미영역을 디자인

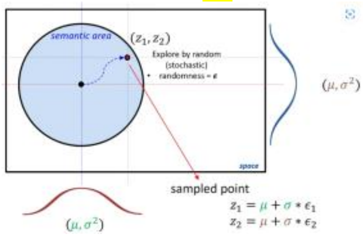


> 잠재공간의 차원수와 모양을 특정하지 않기 때문에 그냥 M개의 정규분포를 이용하여 잠재공간을 디자인
> **VAE의 가장 큰 장점은 연속적인 잠재공간을 갖게 된다는 점**

■가우시안 잠재공간과 샘플링

-AE에서는 Vector를 직접 연산해냄
-VAE는 확률분포 를 결정하는 **평균**과 **표준편차**를 학습함

-하지만 디코더는 확률분포가 아닌, 의미가 있는 표현인 벡터가 있어야 디코딩이 가능함
-확률 분포로부터 의미있는 벡터를 얻기 위해 **샘플링**을 수행



- 샘플링을 통해서 하나의 표현벡터를 얻어내고, 이 값이 디코더로 입력됨

※ 문제점
- 역전파 과정에서는 샘플링 문제가 발생
- 샘플링 과정은 랜덤성이 포함되어 있어서 인코더 파트까지 역전파가 되지 않음
>> 이를 해결하기 위해서 **재파라미터화 트릭**을 사용

■재파라미터화 트릭(Reparameterization Trick)

-인코더와 디코더의 성능 향상을 위해서 VAE는 역전파가 필수적
-역전파는 손실함수에 대한 각 매개변수의 기울기를 계산

-VAE에서 잠재변수 z는 인코더에서 얻은 평균과 표준편차를 바탕으로 무작위로 샘플링 됨
$z = \mu + \sigma \cdot \epsilon$
(ε는 표준정규분포에서 샘플링된 값이며 **무작위성**을 가짐)
>>> 이때 무작위성을 가지는 경우에는 평균과 표준편차의 변화에 따른 z값의 변화(기울기)를 알수가 없음 , 즉 미분 불가능

-재매개변수화 기법 : 인코더의 출력을 직접적으로 사용하여 z를 생성함
>>> z를 생성할 때 사용된 평균과 표준편차의 변화가 어떤 영향을 미치는지 추적 가능