

# VQ-VAE

## 벡터 vector quantization

양자화

\* 양자화

정보를 이산적인 단위로 변환

\* 벡터 양자화

연속적인 벡터를 유한한 개수의

이산적인 벡터로 맵핑

복수

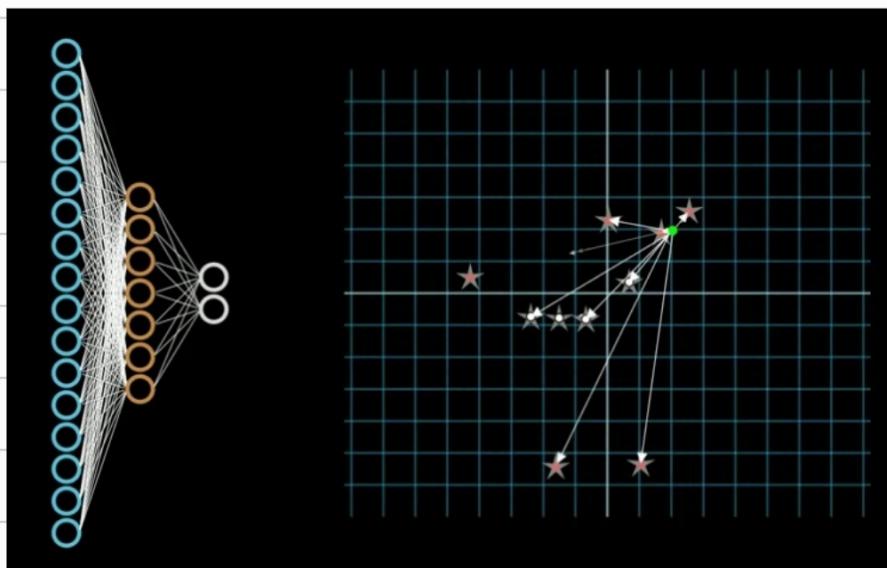
Encoder: 원본 데이터를 더 적은 개수의 숫자로 표현 (n)  $\rightarrow$  (e) 이때 다양한 개정과 일정한 복원이 요구됨.

AE: 그 개정을 찾는다.

이때  $(z^n \in \mathbb{R}^{14}, z^e \in \mathbb{R}^4)$

데이터의 특징, 패턴

VQ: 잠재 공간에서  $z^e$ 를  $q^e$ 로 맵핑 하며 벡터의 움직임을 제한



\*  $\rightarrow$  유한한 이산적인 벡터  $q^e$

●  $\rightarrow$  인코딩된 벡터  $z^e$

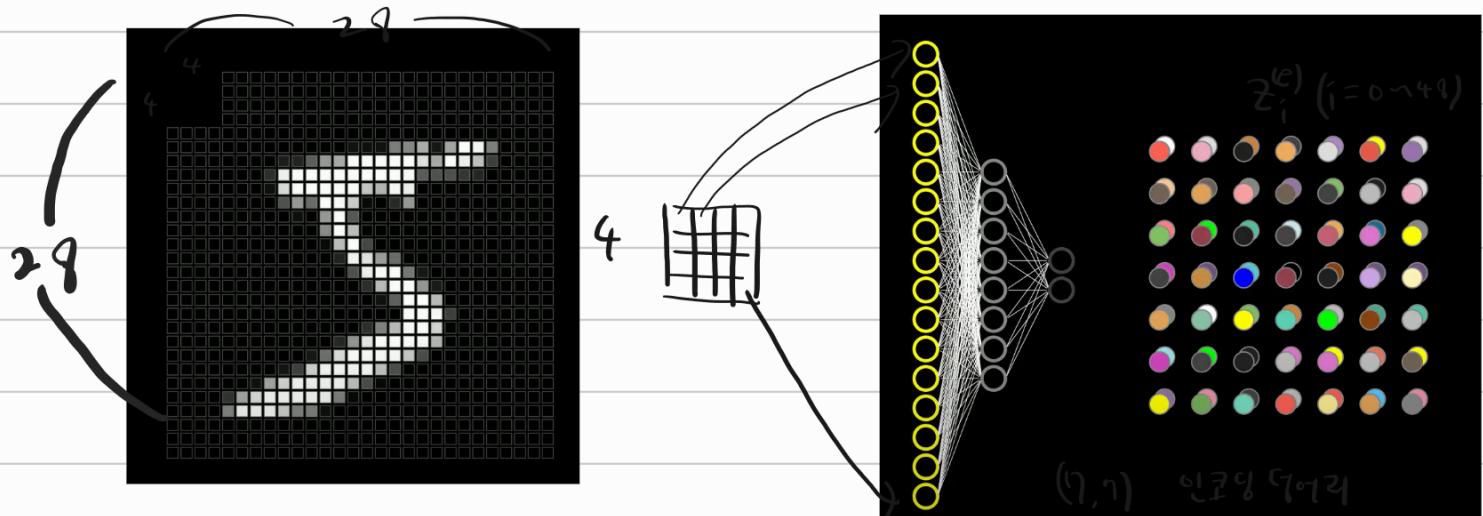
$z^e$ 를 가장 가까운  $q^e$ 로 대응

Input vector  $z^n \rightarrow z^e \rightarrow q^e$

차원도 줄이는데  $q^e$ 로 대응까지?

정보 손실 + 데이터 왜곡

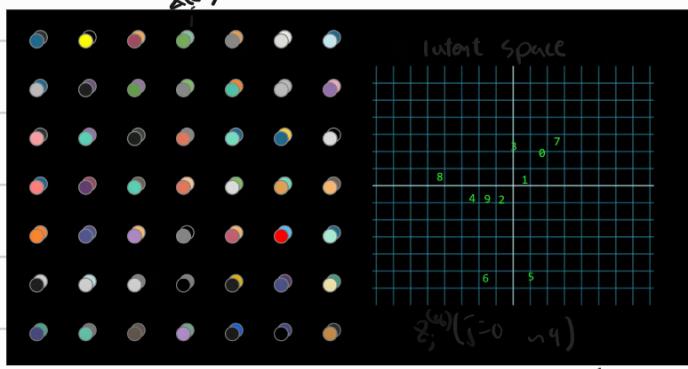
왜 저작한 잠재 벡터  $q^e$ 를 사용하지 않고 미리 가정한 이산적인  $q^e$ 에 대응시킬까?



앞서 말한 것처럼 인코딩 개수가 유한하다면? = 이산적이라면?

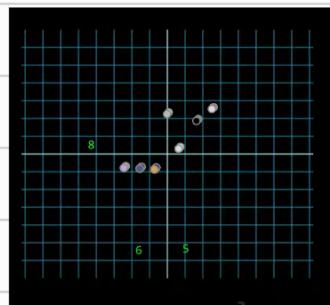
이때 인코딩=저작한 잠재 공간의 벡터  $z^e$ 가 유한하다?  $\rightarrow$  각  $z^e$ 에 정수인덱스를 부여 할 수 있음

2(6)



2차원 벡터는

점(6)에 대응



그럼 여기서 왜 인코딩 벡터를 정수 시퀀스로 만들까?

위에서 만든 인코딩 벡터는 continuous하지 인코딩하기 때문이

기 때문에 Loss (MSE, L1...) 같은 거리가 비슷한 연산을 사용하는 순서를

2	2	9	4	4	9	9
2	9	1	0	7	3	9
2	2	1	1	4	4	2
2	9	9	0	1	2	2
2	2	4	2	7	2	2
9	1	7	0	2	9	2
2	2	2	4	2	2	2

정수 숫자의 나열을

구할 수 있음

↳ 언어 모델

[GPT Transformer]

VQ-Transformer

정수 숫자의 시퀀스를 잘 맞추는 것

근데 애초에 인코딩 벡터는 점(6)과 같은

여기서 모델은 인코딩 벡터는 모르지만

정수 시퀀스의 패턴은 파악하는 있음

모델은 정수 시퀀스를 잘 맞추는 것

근데 애초에 인코딩 벡터는 점(6)과 같은

벡터 양자화

Vector Quantization : 무한한 점(6)을 유한한 점(6)으로 통하는 거

무한한 차원을 2차원이라고 가정할 때 2개의 실수로 이루어진 점(6)은  $\mathbb{R}^2$ 에서 무한한 경계의 수를 가짐

ex)  $\pi = 3.141592\ldots$ 을 계산하는 명령을 위해 양자화를 하면 3으로 나타낼 수 있음. 이때 0.141592... 만큼 오차가 생김

그럼 2차원의 무한차원을 양자화한다면 3.1로 나타낼 수 있음

## Residual V Q

↳ 계층적 벡터 양자화 아님

그럼 해석의 인코딩 벡터 점(6)을 나타낼 때 여러번 양자화를 통해 오차를 줄인다면  
모델이 학습해야 할 정수 시퀀스 점(6)도 많아짐

→ 더 큰 모델 필요

근데 애초에 어떻게 학습할까?

+ 이 벡터 양자화 개념은 GMM에서 어떻게 사용되는지

VQ	Residual VQ
2 2 9 4 4 9 9	2 9 1 2 8 4 7 9 9 0
2 9 1 0 7 3 9	2 6 9 2 1 8 0 5 7 9 3 1 9 5
2 2 1 1 4 4 2	2 2 3 1 7 4 4 4 7 3
2 9 9 0 1 2 2	2 1 9 5 9 0 4 3 2 2 8
2 2 4 2 7 2 2	2 2 4 6 5 7 2 8 6
9 1 7 0 2 9 2	9 3 1 2 7 0 2 0 5 2 1
2 2 2 4 2 2 2	2 6 5 8 4 2 9 2 2 0

① VQ-VAE는 점(6)을 정수 시퀀스 점(6)로 어떻게 표현할지 학습

② 별도의 생성 모델 (Transformer, PixelCNN 등)을 가지거나 VQ-VAE에서 학습한 여러 정수 시퀀스들의 패턴과 규칙을 학습

③ 빼놓은 시퀀스 생성 후 VQ-VAE 디코더로 복원