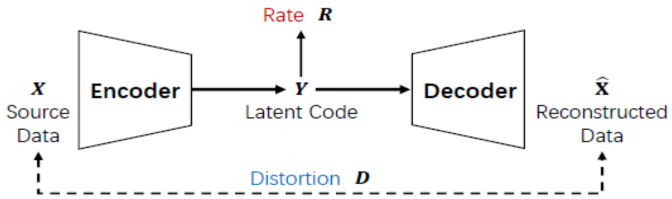
정보이론

손실압축

Background



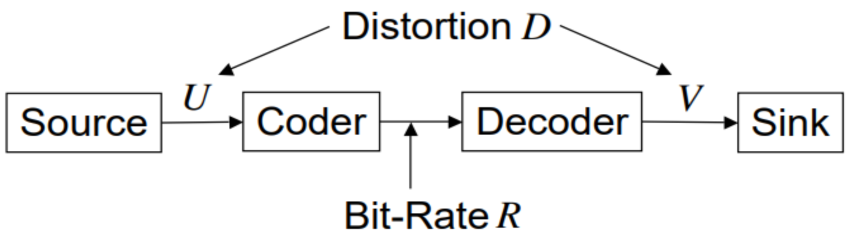
Rate = the bit length(size) of the latent code, Y
Distortion = the fidelity of reconstructed data.

1. Rate-Distortion Theory(부호율-변형 이론)

- 주어진 변형값(왜곡), D 를 넘지 않으면서 보낼 수 있는 최소의 부호율(bit per sample), R 을 알려주는 이론.
  - It addresses the problem of determining the minimal number of bit per symbol, as measured by the rate, R, that should be communicated over a channel, so that the source(input signal) can be approximately reconstructed at the receiver(output signal) without exceeding an expected distortion, D.
  - The minimum average transmission rate such that the average distortion is no larger than D.
- 즉, 정보를 최대한 잃지 않으면서 얼만큼 압축을 할 수 있느냐에 대한 문제를 다루고 있음.
  - Rate–distortion theory gives an analytical expression for how much compression can be achieved using lossy compression methods.
  - Theoretical discipline treating data compression from the viewpoint of information theory.

A. Rate-Distortion Trade offs

- 왜곡을 줄이기 위해서는 계속 bit 수(rate of bit) 를 늘려야 함.
- 왜곡(distortion) ↓ → bit rate(bit per sample, 부호율) ↑
  - The Rate-Distortion theory presents an analytical expression for the trade-off between the bit-rate and reconstruction quality of data compression.
  - Rate-distortion theory analyzes the fundamental tradeoff between the rate(bit per sample) used for representing samples from a data source $X \sim p_X$. And expected distortion uncured in decoding those samples from their compressed representations.
- 이러한 관계는 Rate-Distortion function 을 통해서 표현 할 수 있음.



<전송 시스템>

https://ko.wikipedia.org/w/index.php?title=%EB%B6%80%ED%98%B8%EC%9C%A8-%EB%B3%80%ED%98%95_%EC%9D%B4%EB%A1%A0&action=edit&section=2
https://en.wikipedia.org/wiki/Rate%E2%80%93distortion_theory
http://proceedings.mlr.press/v97/blau19a/blau19a.pdf
http://iphome.hhi.de/wiegand/assets/pdfs/DIC_rd_theory_quantization_07.pdf

Background

1. Rate-Distortion Theory(부호율-변형 이론)
   B. Lossy compression

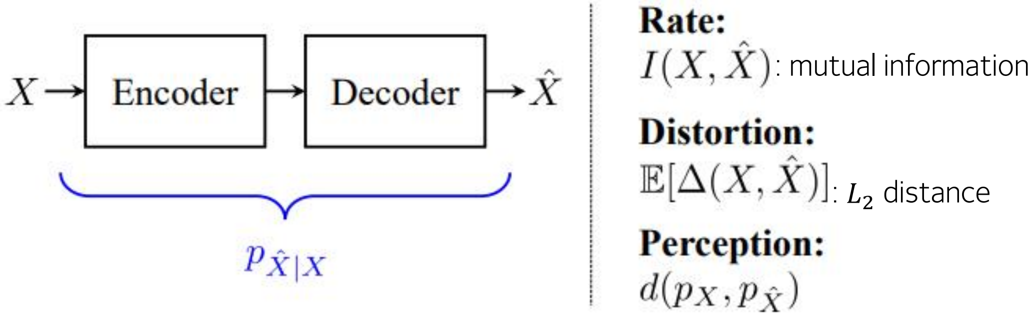


Figure 2. **Lossy compression.** A source signal $X \sim p_X$ is mapped into a coded sequence by an encoder and back into an estimated signal $\hat{X}$ by the decoder. Three desired properties are: (i) the coded sequence be compact (low bit rate); (ii) the reconstruction $\hat{X}$ be similar to the source X on average (low distortion); (iii) the distribution $p_{\hat{X}}$ be similar to p_X, so that decoded signals are perceived as genuine source signals (good perceptual quality).

- 기본이 되는 용어는 3가지 있음.

1. Rate
2. Distortion
3. Perception

Rate(압축률): 원본 신호 대비 얼마나 잘 압축되었는지를 mutual information 을 통해서 정의함.

- 압축률이 높을수록 서로 independent 함,

- mutual information
  - $I(A, B)$ =A 라는 확률 변수를 통해서 B 라는 확률 변수에 대하여 얻어진 정보량
  - 즉, 하나의 확률변수가 다른 하나의 확률 변수에 대하여 제공하는 정보의 양
  - Mutual information is one of many quantities that measures how much one random variables tells us about another.

http://proceedings.mlr.press/v97/blau19a/blau19a.pdf

Background

1. Rate-Distortion Theory(부호율-변형 이론)

   C. Rate-Distortion function

Distortion 과 bit rate 는 trade-off 즉, 반비례 관계에 있음.

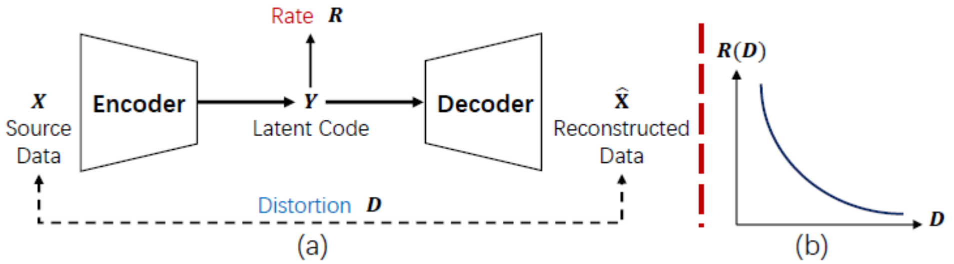


Figure 2: (a) Illustration of the data compression system and (b) Rate-Distortion trade-offs.

- 가우시안이고 i.i.d 인 신호 x 에 Rate-Distortion function 은 다음과 같음.
- if the expected distortion is bounded by D, then the lowest achievable rate, R is characterized by the(information) rate-distortion function

$$R(D) = min_{p_{\hat{x}|x}} I(x, \hat{x})\ s.t. \mathbb{E}[\Delta(x, \hat{x})] < D$$

의미] 주어진 변형값(왜곡), D 를 넘지 않으면서 보낼 수 있는 최소의 부호율(정보), R

$$R(D) = \begin{cases} \frac{1}{2}\log_2(\sigma_x^2/D), & \text{if } D \le \sigma_x^2 \\ 0, & \text{if } D > \sigma_x^2 \end{cases}$$

여기서, σ_x^2 는 신호 x 의 크기임.

의미] 1, distortion 의 양은 원신호의 전력보다 커질 수 없음.
2. Distortion 을 줄이기 위해서 계속 bit 수를 늘려야 함.

=> 결국, rate-distortion theory 는 신호의 압축률과 정보 손실 간의 trade-off 를 푸는 문제임.

*iid(independent and identically distributed, 독립 항등 분포)
- 두개 이상의 확률 변후를 함께 고려할 때, 이들의 확률적 특성이,
  1. 통계적 독립(independent) 이며,
  2. 동일한 확률분포(identically distributed) 를 가지고 있는 것을 말함.

http://proceedings.mlr.press/v97/blau19a/blau19a.pdf

Background

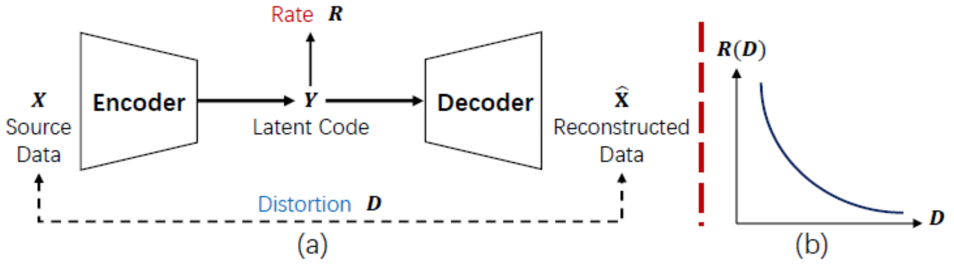


Figure 2: (a) Illustration of the data compression system and (b) Rate-Distortion trade-offs.

2. Information Bottleneck Theory

- Rate-distortion theory 을 distortion function 없이 확장한 이론.
  - Information Bottleneck theory further extends it without explicitly defining the distortion function.
  - Rate-distortion 에서 사용된 distortion function 을 활용하지 않은 이유.
    - Distortion 을 재는 것은 신호의 meaningful 혹은 relevant feature 를 알아야 하는데 이것이 쉽지 않음.
  - Rate-distortion theory 에 따른 압축률(rate) 를 정의할 때, signal rate 만 고려하면 유의미한 feature 혹은 정보가 대부분 소실 될 수 있음.
- 핵심 아이디어] latent code, y 를 활용함.
  - Positive joint distribution, $p(x,y)$ 가 존재한다는 가정 하에 (y 와 x 는 반드시 dependent) y 와의 relevant information 을 보존하는 것으로 distortion 을 간접적으로 측정함.
- 데이터, x 로 부터 관련 정보인 y 로 정보를 압축할 때 y 와의 관련성(accuracy)과 x 의 압축성(compression) 사이의 최고의 trade-off 를 정보량을 통해서 찾는 기법.
  - IB Formalized the problem that finding a short code for X that preserves the maximum information about Y through a 'bottleneck'
- 즉, y 에 대한 x 의 정보를 최대한 보존하면서, x 자체는 최대한 압축하는 기법.
- 따라서, 목적 함수는 다음과 같이 표현됨(슬라이드 뒷면).

https://parkgeonyeong.github.io/Information-Bottleneck-%EC%A0%95%EB%A6%AC/
https://lyusungwon.github.io/studies/2018/05/03/dvib/

Background

2. Information Bottleneck Theory

a. 목적 함수

$$R_{IB} = I(x;y) - \beta I(y;\hat{x})$$

1. $I(y;\hat{x})$
   - y 와의 관련성(accuracy)
   - The network first fits on the training data, where $I(y;\hat{x})$ increase.
2. $I(x;y)$
   - x 의 압축성(compression)
   - Then forgets minor information, where $I(x;y)$ decrease.

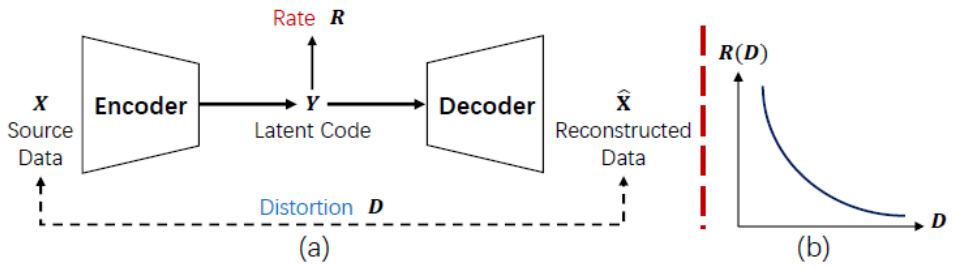


Figure 2: (a) Illustration of the data compression system and (b) Rate-Distortion trade-offs.

- Deep learning 관점에서 해석
  - 압축 과정에서 발생하는 정보의 손실이 있음.
  - This observation imply role of forgetting in learning
  - 중요한 공통된 패턴 정보를 학습.
    - Deep model primarily learns common patterns of the training data for reconstruction.
  - 반면, 중요하지 않은 영상에 특화된 디테일 정보는 학습하지 않음.
    - In contrast, infrequent pattern and image-specific details are typically forgettable trivialities for the trained models.

https://en.wikipedia.org/wiki/Information_bottleneck_method
https://parkgeonyeong.github.io/Information-Bottleneck-%EC%A0%95%EB%A6%AC/
https://lyusungwon.github.io/studies/2018/05/03/dvib/

Background

Rate-Distortion Theory. The Rate-Distortion theory (Shannon et al. 1959; Cover 1999; Blau and Michaeli 2019) presents an analytical expression for the trade-off between the bit-rate and reconstruction quality of data compression. Fig. 2 demonstrates a typical system of lossy data compression, where *rate* indicates the bit length (size) of the latent code Y while *distortion* reflects the fidelity of reconstructed data. The distortion D is measured by $D = \mathbb{E}[\Delta(X, \hat{X})]$, where Δ is ℓ_1 loss in our case. Given a maximum distortion D^*, the lower-bound for the bit-rate R is given by:

$$R(D^*) = \min_{D \leq D^*} \{I(X; \hat{X})\}, \quad (1)$$

where $I(\cdot)$ denotes the mutual information.

Information Bottleneck Theory. Rate-Distortion theory determines the level of inevitable distortion D with a specific rate R. Information Bottleneck theory (Tishby, Pereira, and Bialek 1999; Tishby and Zaslavsky 2015; Shwartz-Ziv and Tishby 2017) further extends it without explicitly defining the distortion function:

$$R = \min\{I(X;Y) - \beta I(Y;\hat{X})\}. \quad (2)$$

This theory further conjectures that the training process of deep models consists of two stages. The network first fits on the training data, where $I(Y;\hat{X})$ increases and then forgets minor information, where $I(X;Y)$ decreases. This observation implies the essential role of forgetting in learning, and the deep model thus primarily learns common patterns of the training data for reconstruction. In contrast, infrequent patterns and image-specific details are typically forgettable trivialities for the trained models.