0장

0. Introduction

※ Information Theory 2 Questions

By Shannon

1. What is the minimum representation of a probabilistic source?
   : Entropy (H)
2. What is the maximum transmission rate of a communication channel?
   : Channel Capacity (C) : 정보 손실 없이, 채널을 통해 최대 전달 data rate.

※ Information quantity (amount)

- Event에 따른 정보 양이 다름.
- p 확률로 발생하는 event의 information amount

$$= \log \frac{1}{p} = -\log p$$

$h(x) = -\log p(x)$

- 이벤트 = Random Variable.
  - x를 가질 확률 $p(x)$가 정의됨.
  - probability distribution (discrete or continuous)
- Entropy : Random Variable의 평균 정보량
  = RV에 의해 얻을 수 있는 평균 정보량

$$: H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

$\mathcal{X}$: 가질 수 있는 값들의 집합.
x와 상관 X. 오로지 $p(x)$에 상관 O.

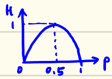
$$= E_x[-\log p(x)]$$

※ Entropy

- $$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

  $= 2$: bits
  $= e$: nets

$p(x) = 0$ 이면?
- $\log p(x) \to -\infty$ 이지만 $p(x) \to 0$의 속도가 더 빠름.
- $H(x) = 0$.



- RV에 의해 얻을 수 있는 평균 정보량

ex) Bernoulli RV : p, $1-p$

$\Rightarrow H(X) = -p \log p - (1-p) \log (1-p)$

$\triangleq H(p)$ — binary entropy

⊗ 정보 = RV를 표현하기 위해 필요한 number (bit)

- RV의 평균 Uncertainty.

- Entropy 예시
  - $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\} \rightarrow H(X) = 1.875$ — H 구하기 $= -\Sigma p \log p$ / 실제 필요 bit수 : 확률의 개수보다 큼, $2^{\wedge}$.
  - 3 bit중 평균 1.875 bits의 정보량 : 압축 가능
  - 압축 정도는 discrete < Correlation
  - Uniform RV : $p = \frac{1}{N}$ (동일)
    : 정보량 최대. 압축 불가.
  - Bernoulli (Binary) RV : $\mathcal{X} = \{0, 1\}$
    : $p(0) = p(1) = 0.5$시 최대. 1 bit.

⇒ Probability distribution에 따라 RV의 정보량 (Entropy)이 달라진다.
+ Uniform p.d에서 entropy는 최대.

* Data Compression

- Source coding : source에 의해 생성된 data의 representation length 줄이기.
- Source coding Theorem
  - iid RV X에 의해 만들어진 length n vector는 n이 증가할수록 $nH(X)$로 표현 가능.
  - $H(X)$: 압축 후에 symbol당 갖는 평균 정보량 의 최솟값.
    : Data Compression Limit

* Lossless Source coding

① Entropy coding
  - 딱 Entropy 만큼 coding 하는 것.
  - iid RV. 각 sequence에 $-n \log p$ 길이 codeword 할당.
  - 평균적으로 $nH(X)$ bits 표현.
  - Ex) Huffman coding, arithmetic coding, JPEG, mp3, --
  - 한계 : 실제로 RV는 iid한 확률이 아님. 서로 Correlation 존재.

② Dictionary coding
  - Data 기반 압축. data의 pattern을 분석하여 이를 기반으로 압축.
  - LZW coding

* Lossy source coding

- 원본과 압축의 복원 차이가 특정 크기 (distortion) 안에 있음. Allow ignorable error.
  - 특정 크기 distortion 을 minimize 하는게 관건.
- audio, image, video, ---
  - 실제 사용) image, audio → lossy (+ lossless)

    data → lossless : zip, tar

* Communication Problem

- Channel 통신
  - Transmitter → Channel → Receiver
  - 통신 과정에서 data의 유실 가능성이 있음.
- 유실된 data 복원?
  - Error correction.
  - Repetition (여러개 보내기) and majority decoding (가장 많이 나타난걸 채택)
- Rate - Reliability trade-off
  - 있을 것 같지만, Shannon says NO.
  - Shannon : data rate이 channel capacity 보다 낮으면, trade off 없다.

* Channel Coding Theorem

어떤 channel에서든

- Channel capacity 보다 낮은 data rate으로 통신하면, error-free 전송 가능하다. = achievability
  - data rate = 전체 전송하는 data 중 실제 보내고싶은 data 비율.
  - data rate이 낮다 = 보내는 bit가 많다. (parity bit 추가)
- converse : Channel capacity 보다 높은 data rate으로 통신할 땐, error-free 전송 불가하다.

✳ KL divergence

- 두 distribution 간의 거리 measure
- $D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$.
  - q라고 가정하고 인코딩, 전송 했을때 정보량
  - 실제(p) 분포에 따른 정보량.

✳ Channel capacity

- X, Y는 서로 correlated 된 Random Variable들.
- Conditional entropy
  - Y가 주어졌을 때, X의 entropy.
  - $H(X|Y) = -\sum_{x,y} p(x,y) \log p(x|y) = E_{x,y}[-\log p(x|y)]$ X=Y면 0임
- Mutual Information
  - 두 R.V X, Y 간의 correlation 에 의해 감소되는 정보량. X, Y 간의 dependence.
  - $I(X;Y) = H(X) - H(X|Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$

    → X, Y가 independent면 $p(x,y) = p(x)p(y)$ 이므로 $I(X;Y) = 0$. 나머지경우 >0.

    → $D(\underbrace{p(x,y)}_{X,Y\text{ 실제}} \| \underbrace{p(x)p(y)}_{\text{independent 라고 가정}})$

    ex) X: not noise, Y: X+ noise 면

    $I(X;Y) = I(Y;X) = \underbrace{H(Y)}_{Y\text{의 정보량}} - \underbrace{H(Y|X)}_{Y\text{중 noise에 해당하는 양}}$ = 보낼 수 있는 정보량.

  - $I(X;Y) = I(Y;X)$ and $I(X;Y) \geq 0$
- Channel capacity의 정의

  1) 수학적 정의
  - $C = \sup_{p(x)} I(X;Y)$. sup : supremum; 열린구간에서 최대값.
  - Mutual Information 의 최대값. — 감소되는 정보량이 최대면, 더 적은 bit으로도 표현 가능. X를 control 함.

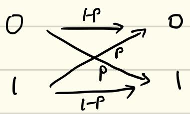
  2) 기능적 정의
  - Error를 0으로 보낼 수 있는 최대 전송 rate.
- Shannon : random code 로 channel capacity를 이룰수 있다.
  - : implementable 하지 않음.

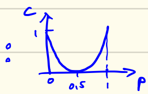
※ Channel Capacity 예시

ex1) Noiseless binary channel : 1 bit

- $0 \to 0$, $1 \to 1$ 로 전송.
- $C = \max I(X;Y) = H(Y) - H(Y|X) = 1 - 0 = 1$ bit.

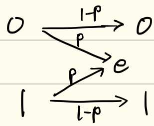
ex2) Binary Symmetric channel

- 



- $C = \max I(X;Y) = 1 + p\log p + (1-p)\log(1-p) = 1 - H(p)$

ex3) Binary erasure channel

- 

- $C = 1 - p$

ex4) AWGN channel

- $C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right)$ ($y = x + n$, n: real gaussian noise)

  P: power constraint $= E[x^2]$

  N: noise variance $= E[n^2]$

※ Turbo Codes (Claude Berrou, 1993)

- Shannon 의 channel code 문제 해결.
- First code to achieve theoretical limit.
- 3G standards.
  - Mobile 통신 : WCDMA, LTE, WiBro, ···

Gallager

※ LDPC Codes (1960 → 1996 by Mackay)

- turbo와 비교할만한 성능, 빠른 decoder.

※ Polar codes (Arikan 09)

- Achieve capacity, 증명 됨.