INFORMATION THEORY BASICS

EE6317 - Channel Coding

Information Theory

- Source coding / data compression
- Channel coding / reliably communicating in the presence of noise.
- Hypothesis testing / classification with a prior.

Information Theory

- Source coding $\quad \underline{H(X)} = -\sum_{x \in \mathcal{X}} p(x) \log_2 \frac{1}{p(x)}$

- Channel coding $\quad \underline{I(X;Y)} = \sum_{xy} p(xy) \log_2 \frac{p(xy)}{p(x)p(y)}$

- Hypothesis testing $\quad \underline{D(p\|q)} = \sum_{x} p(x) \log \frac{p(x)}{q(x)}$

Source coding

raw file — compressed file — decompressed

$$X^n \longrightarrow \boxed{\text{ENCODER}} \longrightarrow C^k \longrightarrow \boxed{\text{DECODER}} \longrightarrow \hat{X}^n$$

$X^n \in \mathcal{X}^n$

$C^k \in \{0,1\}^k$

probability of error

$$P_e = \Pr[X^n \neq \hat{X}^n]$$

$X^n \sim \text{iid}(p_X)$

$$\boxed{R = \frac{k}{n}}$$

- fixed-length — k is fixed (X^n)
- variable-length — k depend on X^n

(zip, gzip, 7z)

$\mathcal{X}$: input alphabet (words, letters, ASCII chars)

The source coding theorem (Discrete memoryless source iid)

Theorem (Shannon, 1948). For any $\epsilon > 0$,

① $\exists$ compression algorithms that achieve

$$\lim_{n\to\infty} P_e = 0 \quad \&$$

$$\lim_{n\to\infty} R = H(X) + \epsilon$$

② For any compression algorithm with

$$\lim_{n\to\infty} R \le H(X) - \epsilon,$$

$$\lim_{n\to\infty} P_e > 0.$$

not possible to compress below H(X)

General fixed-length compressor

ENC DEC

$x^n(1) \longrightarrow 00\cdots0 \longrightarrow x^n(1)$

$x^n(2) \longrightarrow 00\cdots01 \longrightarrow x^n(2)$

$\vdots \qquad\qquad \vdots$

$x^n(2^{nR}) \longrightarrow 11\cdots1 \longrightarrow x^n(2^{nR})$

$x^n(2^{nR}+1)$

$\vdots$

$x^n(|\mathcal{X}|^n)$

$c_1 c_2 \cdots c_k$

$k = nR$

reproducible R

always in error

Analysis

Rate $\quad R = \frac{1}{n} \log_2 |\mathcal{R}| = \frac{k}{n}$

Probability of error $\quad P_e = \sum_{x^n \in \mathcal{R}^c} P_{X^n}(x^n)$

$= \Pr[X^n \notin \mathcal{R}]$

High probability set $\mathcal{A}$: $\quad \Pr[X^n \notin \mathcal{A}] \to 0$ as $n \to \infty$

Source compression: find $\underline{\underline{\mathcal{A}}}$ of min. size

(Weakly) Typical set (Cover & Thomas)

$$T_{\epsilon,w}^{(n)}(p_X) = \Big\{ x^n \in \mathcal{X}^n : H(X) - \epsilon \le \frac{1}{n}\log_2 \frac{1}{p_{X^n}(x^n)} \le H(X) + \epsilon$$

OR

$$2^{-n(H(X)+\epsilon)} \le p_{X^n}(x^n) \le 2^{-n(H(X)-\epsilon)} \Big\}$$

$$R = T_{\epsilon,w}^{(n)}(p_X) \longrightarrow \text{one compression}$$

Properties

1. $\Pr[X^n \notin T_{\epsilon,w}^{(n)}(p_X)] \to 0$
   as $n \to \infty$

High prob set!

2. $|T_{\epsilon,w}^{(n)}(p_X)| \leq 2^{n(H(X)+\epsilon)}$

$R \leq H(X)+\epsilon$

3. $|T_{\epsilon,w}^{(n)}(p_X)| \geq 2^{n(H(X)-\epsilon)}(1-\epsilon)$
   for all suff. large n.

$$Y(x^n) = \frac{1}{n} \log \frac{1}{p_{X^n}(x^n)} = \frac{1}{n} \log \frac{1}{\prod_{i=1}^{n} p_X(x_i)}$$

$$= \frac{\sum_{i=1}^{n} \log \frac{1}{p_X(x_i)}}{n}$$

$$Y(X^n) = \frac{\sum_{i=1}^{n} \log \frac{1}{p_X(X_i)}}{n} \rightarrow \text{sum of iid RVs!}$$

$$\mathbb{E}\, Y(X^n) = H(X)$$

$$\Pr\left[\, |Y(X^n) - H(X)| > \epsilon \,\right] \rightarrow 0$$

$$\text{as } n \rightarrow \infty$$

$$\Pr[X^n \in T_{\epsilon,w}^{(n)}(p_x)] \quad \begin{matrix} \geq \frac{1}{2} \text{ for large } n \\ \leq 1 \end{matrix}$$

$$\sum_{x^n \in T_\epsilon^{(n)}(p_x)} p_{X^n}(x^n) \quad \begin{matrix} \geq \frac{1}{2} \\ \leq 1 \end{matrix}$$

$$\quad \vee|$$

$$p_{X^n}(x^n) \geq 2^{-n(H(X)+\epsilon)}$$

$$\sum_{x^n \in T_{\epsilon,w}^{(n)}(p_x)} \underline{2^{-n(H(X)\mp\epsilon)}} \quad \begin{matrix} \geq \frac{1}{2} \\ \leq 1 \end{matrix}$$

$$2^{-n(H(X)+\epsilon)} \quad |T_{\epsilon,w}^{(n)}(p_x)| \quad \begin{matrix} \geq \frac{1}{2} \\ \leq 1 \end{matrix}$$

$$|T_{\epsilon,w}^{(n)}(p_x)| \leq 2^{n(H(p_x)+\epsilon)}$$

$$|T_{\epsilon,w}^{(n)}(p_x)| \geq 2^{n(H(X)-\epsilon)} \quad \text{for large } n$$

Chapter 3 (C&T)

(Strongly) typical set

Type of x^n $\mu_{x^n}(a) = \dfrac{\#\text{ occurrences of } a \text{ in } x^n}{n}$

$$T_{\epsilon,s}^{(n)}(p_X) = \left\{ x^n \in \mathcal{X}^n : \left| \mu_{x^n}(a) - p_X(a) \right| \leq \epsilon\, p_X(a) \quad \forall a \in \mathcal{X} \right\}$$

Eg

$$\underline{a} \quad \underline{b} \quad \underline{a} \quad \underline{a} \quad c \quad \underline{b} \quad \underline{b} \quad \underline{a} \quad \underline{a} \quad = x^9$$

$$\mu_{x^n}(a) = \frac{5}{9}$$

$$\mu_{x^n}(b) = \frac{3}{9}$$

$$\mu_{x^n}(c) = \frac{1}{9}$$

$$\mu_{x^n} = \begin{bmatrix} 5/9 \\ 3/9 \\ 1/9 \end{bmatrix}$$

$T_{\epsilon,s}^{(n)}(p_X)$ is a high probability set

$$\mu_{x^n}(a) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i = a\}}$$

$$\mathbb{1}_{\{x_i = a\}} = \begin{cases} 1 & \text{if } x_i = a \\ 0 & \text{if } x_i \neq a \end{cases}$$

$$\mu_{X^n}(a) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i = a\}}$$

$$\mathbb{E}\,\mu_{X^n}(a) = \frac{1}{n}\sum_{i=1}^{n} \left(1 \times p_X(a) + 0 \times (1 - p_X(a))\right)$$

$$= p_X(a)$$

By WLLN, $\Pr[X^n \notin T_{\epsilon,s}^{(n)}(p_X)] \to 0$ as $n \to \infty$

Properties of entropy

① $H(X) \geq 0 \quad \forall p_X$

$H(X) = 0$ iff X is deterministic

② $H(X)$ is a concave function of p_X.

③ $H(X_1, X_2) = H(X_1) + H(X_2 | X_1)$

Mutual information

1. $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$
   $= I(Y;X)$

2. $I(X;Y) \geq 0 \;\forall\; p_{XY}$ equality iff $X \perp\!\!\!\perp Y$

3. $I(X;Y)$ - convex in $p_{Y|X}$ for fixed p_X
   $f(p_X, p_{Y|X})$ concave in p_X for fixed $p_{Y|X}$

4. $H(X) \geq H(X|Y)$

5. $I(X;YZ) = I(X;Y) + I(X;Z|Y)$

Data processing inequality

$X \leftrightarrow Y \leftrightarrow Z$ Markov chain

$\Rightarrow \quad I(X;Y) \geq I(X;Z)$

$X \longrightarrow Y \longrightarrow Z$

$\uparrow$ unknown $\quad \uparrow$ observable $\quad \uparrow$ processed

Fano's inequality

unknown observable

For any Markov chain $X - Y - \hat{X}$, $\rightarrow$ estimate

$$H(X|Y) \leq H(X|\hat{X}) \leq H_2(P_e) + P_e \log_2 |\mathcal{X}|$$

where

$$P_e = \Pr[X \neq \hat{X}]$$

$$H_2(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

(Chapter 2 of C&T)