Information Theory &
Coding 13th oct 2020

Yesterday:

M : I.L. F.S.M

LZ: Lempel-Ziv method

$u_1 u_2 \ldots$: any sequence $\mathcal{U}^\infty$

$\Rightarrow \quad \rho(M, u_1 u_2 \ldots) \geq \rho(LZ, u_1 u_2 \ldots)$

if $\ldots U_1 U_2 U_3 \ldots$ is an ergodic stoch process.

Ergodic process: for any function $f : \mathcal{U}^n \to \mathbb{R}$

$$\underbrace{\lim_{N\to\infty} \underbrace{\frac{1}{2N+1} \sum_{i=-N}^{N} f(U_{i+1} \ldots U_{i+n})}_{\text{time-average}}} = E[f(U_1 \ldots U_n)]$$

w.p. 1

then $\rho(LZ, U_1 U_2 U_3 \ldots) \underset{\text{w.p. 1}}{\leq} \mathcal{H}(\{U\}) = \lim_{n\to\infty} \frac{1}{n} H(U_1 \ldots U_n)$

Why: take M to be the FSM that implements the Huffman code for $(U_1 \ldots U_n)$

$$\underbrace{U_1 \ldots U_n}_{\downarrow \text{ binary str}} \; \underbrace{U_{n+1} \ldots U_{2n}}_{\downarrow \text{ bin}} \ldots$$

$$\frac{1}{n} E\left[\text{length } c_n(U_1 \ldots U_n)\right] \le \frac{1}{n}\left[H(U_1 \ldots U_n) + 1\right] \to \mathcal{H}(\,)$$

as n suff. large

$$\rho(M, \underbrace{U_1 \ldots U_n} \underbrace{U_{n+1} \ldots U_{2n}} \underbrace{U_{2n+1} \ldots U_{3n}}, \ldots)$$

$$= \frac{1}{n} E\left(\text{length } C_n(U_1 \ldots U_n)\right)$$

ergodicity

$$\Rightarrow \rho(LZ, U_1 U_2 \ldots) \le \frac{1}{n}\left(H(U_1 \ldots U_n) + 1\right) \quad \forall n$$

$$\Rightarrow \quad '' \quad \le \mathcal{H}.$$

$\ge$ free because any invertible repr. needs $\mathcal{H}$ bits/letter

$$\Rightarrow \rho(LZ, \underline{U_1 U_2 \ldots}) = \mathcal{H}(\{U_i\}) \text{ if}$$

$U_1 U_2 \ldots$ ergodic.

---

we have seen that LZ performs well.

$$\Rightarrow q(u_1 \ldots u_n) = 2^{-\text{length } LZ(u_1 \ldots u_n)}$$

$$\frac{1}{n} D(\underline{p(u_1 \ldots u_n)} \| \underline{q}(u_1 \ldots u_n)) \to 0$$

for any stationary process p.

$$D(p_n \| q_n) = \sum_{u_1 \ldots u_n} p(u_1 \ldots u_n) \log \frac{p(u_1 \ldots u_n)}{q(u_1 \ldots u_n)}$$

write $p(u_1 \ldots u_n) = p(u_1)\, p(u_2 | u_1)\, p(u_3 | u_1 u_2) \ldots p(u_n | u_1 \ldots u_{n-1})$

$q(\quad) = q(\;)\, q(\quad)\, q(\quad) \quad q(\quad)$

$$\underbrace{\frac{1}{n} D(p_n \| q_n)}_{< \varepsilon} = \frac{1}{n}\sum_{i=1}^{n} \sum_{u_1 \dots u_n} \underline{p(u_1 \dots u_n)} \underbrace{\log \frac{p(u_i | u_1 \dots u_{i-1})}{q(u_i | u_1 \dots u_{i-1})}}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \sum_{u_1 \dots u_i} p(u_1 \dots u_{i-1})\, p(u_i | u_1 \dots u_{i-1}) \underbrace{\log \frac{p(u_i | u_1 \dots u_{i-1})}{q(u_i | u_1 \dots u_{i-1})}}$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}}_{\text{averaging over } i} \underbrace{\left(\sum_{u_1 \dots u_{i-1}} p(u_1 \dots u_{i-1})\right.}_{\text{averaging over } (u_1 \dots u_{i-1})} \underbrace{D\left(p(\cdot | u_1 \dots u_{i-1})_{u_i | u_1 \dots u_{i-1}} \| q(\cdot | u_1 \dots u_{i-1})_{(u_i | u_1 \dots u_{i-1})} \right)}_{\geq 0 \text{ number}} \left. \right)$$

$\Rightarrow$ for all i except $n\sqrt{\varepsilon}$ i values between $1, \dots, n$

$$\sum_{u_1 \dots u_{i-1}} p(u_1 \dots u_{i-1})\, D(p_{u_i | u_1 \dots u_{i-1}} \| q_{u_i | u_1 \dots u_{i-1}}) < \sqrt{\varepsilon}$$

if $\frac{a_1 + \dots + a_n}{n} < \varepsilon$ $\quad a_i \geq 0$

~~then~~ $a_1 + \dots + a_n < \varepsilon n$

if $n\sqrt{\varepsilon}$ of the a_i's were $\geq \sqrt{\varepsilon}$ then would be false

$\Rightarrow$ for all $u_1 \dots u_{i-1}$ except a set with

$p(\quad)$ probability $< \sqrt{\sqrt{\varepsilon}}$ the value of

$$D(p(\cdot | u_1 \dots u_{i-1}) \| q(\cdot | u_1 \dots u_{i-1})) < \sqrt{\sqrt{\varepsilon}}$$

Moral: LZ "learns" the true statistics of

p(nextletter | past letters).

Shortcomings:

$$\#\text{ of bits } LZ(u_1 \ldots u_n) \leq \overbrace{q(u_1 \ldots u_n)}^{\#\text{ of distinct words}} \log q(u_1 \ldots u_n) + q(u_1 \ldots u_n) \log (|\mathcal{U}|^2)$$

$$\text{" } M(\quad) \geq q(\quad) \log(\quad) - q(u_1 \ldots u_n) \log (8n^2)$$

$\Rightarrow$ LZ may need a "large n" to compress well if $|\mathcal{U}|$ is large or the statistics of the source is very complex.

In the chat: order of KL:

$$D(p\|q) = \sum_x{}' p(x) \log \frac{p(x)}{q(x)} = \begin{cases} \infty & \text{if } \exists x \text{ s.t. } p(x) > 0 \text{ \& } q(x) = 0. \end{cases}$$

$$p \ll q \equiv (\forall x \quad q(x) = 0 \Rightarrow p(x) = 0)$$

if $p \ll q \Rightarrow D(\ ) < \infty$

else $D(\ ) = +\infty$.

Formal Properties of Information Measures

Recall:

distribution the RV $\bar{U}$

$$H(U) = \sum_u p(u) \log \frac{1}{p(u)} = E\left[\log_2 \frac{1}{p(\bar{U})}\right]$$

we have seen

$D(p \| \text{uniform}) \geq 0$

Thm: $0 \leq H(U) \leq \log |\mathcal{U}|$

each term in $\sum p(u) \log \frac{1}{p(u)}$ is ≥ 0.

$=$ iff $\exists u_0$ s.t. $\Pr(U = u_0) = 1$

$\equiv$ U is deterministic

equality iff $\bar{U}$ is uniformly distributed on $\mathcal{U}$

$$H(U|V) = \sum_{u,v} p(uv) \log \frac{1}{p(u|v)} = \sum_v p(v) \underbrace{\sum_u p(u|v) \log \frac{1}{p(u|v)}}_{H(U|V=v)}$$

Thm $0 \leq H(U|V) \leq H(U)$

$=$ iff U & V indep

$\forall v$ with $p(v) > 0$ $\exists u_0(v)$ s.t. $\Pr(U = u_0(v) | V = v) = 1$.

$\equiv$ $U = u_0(V)$ with probability 1.

Seen: Chain Rule:

$$H(U_1 \ldots U_n) = \sum_{i=1}^{n} H(U_i | U_1 \ldots U_{i-1})$$

$$H((U_1 \ldots U_n) | (V)) = \sum_{i=1}^{n} H(U_i | U_1 \ldots U_{i-1} V)$$

$$H(A,B|C,D) = H((A,B)|(C,D))$$

$$\cancel{\neq H(A,(B|C),D)}$$ ~~syntax error~~

Recall

$$I(U;V) = H(U) - H(U|V)$$
$$= H(V) - H(V|U)$$
$$= H(U) + H(V) - H(UV)$$
$$= D(P_{UV} \| P_U \cdot P_V) \geq 0.$$

equality iff independence

$$I(U;V) \leq \min\{H(U), H(V)\}$$

Recall Also

$$I(U;V|W) = H(U|W) - H(U|VW)$$
$$= H(V|W) - H(V|UW)$$
$$= H(U|W) + H(V|W) - H(UV|W)$$
$$\geq 0$$

equality iff $U - W - V$

Thm: $I(U_1 \ldots U_n; V)$

$$= \sum_{i=1}^{n} I(U_i; V | U_1 \ldots U_{i-1}).$$

Pf: $I(U_1 \ldots U_n; V) = H(U_1 \ldots U_n) - H(U_1 \ldots U_n | V)$

$$= \sum_i \left[H(U_i | U_1 \ldots U_{i-1}) - H(U_i | U_1 \ldots U_{i-1} V) \right]$$

$$= \sum_i I(U_i ; V | U_1 \ldots U_{i-1}) \quad //$$

Also $I(U_1 \ldots U_n ; V | W)$

$$= \sum I(U_i ; V | U_1 \ldots U_{i-1}, W).$$

Thm: Data Processing:

if $U - V - W$, then

$I(U;W) \leq I(U;V)$ and $I(U;W) \leq I(V;W)$

chain rule

Pf: $I(U; VW) = I(U;V) + \cancel{I(U;W|V)}^{\to 0}$

$= I(U;W) + I(U;V|W)$

chain rule

So: $I(U;V) = I(U;VW) = I(U;W) + (\text{something} \geq 0)$

$\geq I(U;W)$. //

Theorem (Fano's Inequality):

Suppose U and V are RVs taking values on the same alphabet $\mathcal{U}$. Then,

$$H(U|V) \le h_2(p) + p\log(|\mathcal{U}|-1)$$

with $p = \Pr(U \neq V)$

$$h_2(p) = p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}$$ (Binary entropy function).

Pf: let $W = \begin{cases} 1 & \text{if } U \neq V \\ 0 & \text{if } U = V \end{cases}$

$\Pr(W=1) = p \quad \Pr(W=0) = 1-p$

$H(W) = h_2(p)$.

$$H(UW|V) = H(U|V) + \underbrace{\cancel{H(W|UV)}}_{\to 0} = H(U|V)$$

so $H(U|V) = H(UW|V)$

$= H(W|V) + H(U|WV)$

$\le H(W) + H(U|WV)$

$= h_2(p) + H(U|WV)$

$$H(U|WV) = \sum_{w,v}' p(w,v)\, H(U|W=w, V=v)$$

the term with $w=0$

$W=0 \equiv U=V$

$W=1 \equiv U \neq V$

$$= \sum_v' p(W=1, v)\, \underbrace{H(U|W=1, V=v)}_{\le \log_2(|\mathcal{U}|-1)}$$

$$\leq \log_2(|\mathcal{U}|-1) \sum_{v} p(W=1, V=v)$$

$$= \log_2(|\mathcal{U}|-1) \Pr(W=1)$$

$$= p\log_2(|\mathcal{U}|-1).$$

//