Information Theory & Coding

Dec 8th, 2020

Rate-Distortion Theory:

$U_1 U_2 \ldots$ iid $\sim p_U$ (Source)

enc: $\mathcal{U}^n \to \{1 \ldots, M\}$ rate $= \frac{\log_2 M}{n}$

dec: $\{1 \ldots M\} \to \mathcal{V}^n$

distortion measure: $d: \mathcal{U} \times \mathcal{V} \to \mathbb{R}$

$E\left[\frac{1}{n} \sum_{i=1}^{n} d(U_i, V_i) \right] = D$, $U^n \to \text{enc} \to \text{dec} \to V^n$

(Bad news): Thm: given $p_U, d(,)$ then any enc/dec that has distortion $= D$ then rate $\geq R(D)$

where $R(D) = \min \{ I(U;V) : p_{V|U}, E[d(U,V)] = D \}$.

(Good news): given $p_U, d(,)$, and D, $\varepsilon > 0$ there exists an enc/dec. s.t.

① $|\text{distortion} - D| < \varepsilon$

② rate $\leq R(D) + \varepsilon$

Fact : $I(U;V)$ is a function of $p_U, p_{V|U}$,

$f(p_U, \underbrace{p_{V|U}})$. f is concave in p_U. $\cap$

f is convex in $p_{V|U}$ $\cap$.

Pf: we need to show that for every $0 \le \lambda \le 1$,

$$\left.\begin{aligned} &\lambda f(\underline{p_U}, p_1) + (1-\lambda) f(\underline{p_U}, p_2) \\ &\quad \ge f(\underline{p_U}, \underbrace{\lambda p_1 + (1-\lambda) p_2}). \end{aligned}\right\}$$

To do this consider the following collection of random variables: W, U, V with the following distribution,

$W \in \{1, 2\}$ $\quad U \sim p_U$

p: $\lambda, 1-\lambda$

$$p(w, u, v) = \underline{p(w)\, p(u)\, p_w(v|u)}$$

$$\Rightarrow \begin{cases} p(u, v | w=1) = p(u)\, p_1(v|u) \\ p(u, v | w=2) = p(u)\, p_2(v|u) \end{cases}$$

$$\underbrace{p(u,v)} = \lambda p(u) p_1(v|u) + (1-\lambda) p(u) p_2(v|u)$$

$$= \underbrace{p(u)} \underbrace{[\lambda p_1(v|u) + (1-\lambda) p_2(v|u)]}.$$

$$\underline{I(U;V)} = \underline{f(p_U, \lambda p_1 + (1-\lambda) p_2)}$$

$$\underline{I(U;V|W)} = \underbrace{\lambda f(p_U, p_1) + (1-\lambda) f(p_U, p_2)}.$$

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

$\qquad\qquad\qquad = I(U;V) + \underline{I(U;W|V)}$

$I(U;V|W) = I(U;VW) \geq I(U;V)$ //.

<u>Pf of the "Bad News" Theorem:</u>

Suppose we have:

$$U^n \to \text{enc} \xrightarrow[\{1 \ldots M\}]{} \text{dec} \to V^n.$$

suppose $D = \frac{1}{n}\sum_{i=1}^{n} E(d(U_i;V_i))$.

Let $p_i(u,v)$ be the distribution of (U_i,V_i)

$\qquad \parallel$

$\quad p(u)\, p_i(v|u)$.

So: $E(d(U_i,V_i)) = \sum_{u,v} p_i(u,v)\, d(u,v)$

$$D = \sum_{u,v} \underbrace{p(u) \left(\frac{1}{n} \sum_{i=1}^{n} p_i(v|u)\right)}_{p(u)\,p(v|u)} d(u,v)$$

$$= \underbrace{\sum_{u,v} p(u)p(v|u)\, d(u,v)}$$

with $p(v|u) = \frac{1}{n} \sum_{i=1}^{n} p_i(v|u)$. Note:

$$I(U_i; V_i) = f(p_u, p_i)$$

$$\underbrace{\frac{1}{n} \sum_{i=1}^{n} I(U_i; V_i)}\geq f\left(p_u, \frac{1}{n} \sum_{i=1}^{n} p_i\right)$$

$$= f(p_u, p_{v|u}) = \underbrace{I(U;V)}$$

Computed for the distribution $p(u,v)$ above

For this $p(u,v)$ we also have $E(d(U,V)) = D$.

To summarize:

$$\frac{1}{n} \log M \geq \frac{1}{n} I(U^n; V^n) = \frac{1}{n} \left(H(U^n) - H(U^n|V^n)\right)$$

data proc.

$$= \frac{1}{n} \left[\sum_{i=1}^{n} H(U_i) - \sum_{i=1}^{n} H(U_i | U^{i-1} V^n)\right]$$

$$\geq \frac{1}{n} \sum_{i=1}^{n} \left(H(U_i) - H(U_i|V_i)\right) = \frac{1}{n} \sum_{i=1}^{n} I(U_i; V_i)$$

$$\geq I(U;V) \geq R(D).$$

//.

For the "good news" counterpart we need to show the existence of enc/dec with certain properties. We will do so with a "random coding" ~~argument~~ method: idea:

Given P_U, $d(\cdot,\cdot)$, D, $R > R(D)$, $\varepsilon > 0$.

we will choose n, $M = 2^{nR}$, we will construct the decoder dec randomly in the following way:

Since $R > R(D)$ there exists $P_{V|U}$ s.t. $R > I(U;V)$ & $D = E(d(U,V))$. Fix this $P_{V|U}$, consider the joint distribution P_{UV}:

$$\mathrm{dec}(m) = V_1(m) \dots V_n(m)$$

$$\{V_i(m) : i = 1 \dots n,\ m = 1 \dots M\} \text{ iid } \sim P_V.$$

The encoder will work as follows: given u^n to encoder, we will search for an m s.t.

$$(u^n, \mathrm{dec}(m)) \in T(n, P_{UV}, \delta)$$ to be chosen.

if we find such an m we will set

$\text{enc}(u^n) = m$. ①

if no such m is found, then let

$\text{enc}(u^n) = 1$. ②

We will now show that the expected distortion of such a system is $\approx D$.

To analyze our constructed enc/dec, let us start with the value of

$\frac{1}{n}\sum_{i=1}^{n} d(u_i, v_i)$ where, $u^n \to \text{enc} \to \text{dec} \to v^n$

and ① happens, i.e,

$(u^n, v^n) \in T(p_{uv}, n, \delta)$.

$$\frac{1}{n}\sum_{i=1}^{n} d(u_i, v_i) = \sum_{u,v} \underbrace{\frac{1}{n}\#\{i : (u_i, v_i) = (u,v)\}}_{p(u,v)(1\pm\delta)} d(u,v)$$

$$= \sum_{u,v} p(u,v) d(u,v) \pm \delta \left(\sum_{u,v} p(u,v) \, |d(u,v)| \right)$$

$$= \underbrace{D \pm \delta \max_{u,v} |d(u,v)|}$$

$= D \pm \varepsilon/2$ if δ is small enough.

What about

$\frac{1}{n}\sum_{i=1}^{n} d(u_i, v_i)$ when ② happens.

$\left| \text{the values this} \right| \leq \max_{u,v} |d(u,v)|$.

What we will show is $\Pr(①) \geq 1-\delta'$

$\Pr(②) \leq \delta'$.

So $\left| E(\text{distortion}) - D \right|$

$= \left| D \Pr(①) \pm \frac{\varepsilon}{2} \Pr(①) - D + D_{max} \Pr(②) \right|$

$\leq |D|\delta' + D_{max}\delta' + \frac{\varepsilon}{2} \leq \varepsilon$ if δ' is small enough.

We have to show that

$\Pr(2)$ is small.

2: for $U^n \sim p_U$ and

$V^n(1), \ldots, V^n(M)$ iid $\sim p_V$. the

event that $(U^n, V^n(m)) \notin T(p_{UV}, n, \delta)$

for every m.

we will upper bound $\Pr(2)$ by considering

two cases:

A $\quad U^n \notin T(p_U, n, \delta/3)$

B $\quad U^n \in T(p_U, n, \delta/3)$.

$$P(2) = \Pr(A \,\&\, 2) + \Pr(B \,\&\, 2).$$

$$\leq \underbrace{\Pr(A)}_{\text{small } \checkmark} + \underbrace{\Pr(B \,\&\, 2)}_{\text{small} \Leftarrow \text{ will show.}}$$

We will now show that for any element

$u^n \in T(p_U, n, \delta/3)$

The probability that ② happens is small:

for such a u^n, we claim that for each

$m = 1, \ldots, M$

$$\Pr\left((u^n, V^n(m)) \in T(P_{UV}, \delta, n)\right) \geq \underbrace{2^{-n(\underbrace{I(U;V) + \delta''}_{\alpha})}}$$

δ'' small if n is large

if we take this claim, then

$$\Pr(②) = \prod_{m=1}^{M} \Pr\left((u^n, V^n(m)) \notin T\right)$$

$$\leq \prod_{m=1}^{M} \left(1 - 2^{-n\alpha}\right) \leq \prod_{1}^{M} e^{-2^{-n\alpha}}$$

$1 - x \leq e^{-x}$

$$= e^{-M 2^{-n\alpha}}$$

$$= e^{-2^{n(R-\alpha)}} \to 0 \quad \text{because}$$

$$R - \alpha = \left(R - I(U,V) - \delta''\right) > 0$$

$R > I(U,V)$

if δ'' is sufficiently small

So, all we have to prove is the claim:

if $\underbrace{u^n \in T(P_u, \delta/3, n)}$ and

V^n iid $\sim P_V$ then

$$\Pr\left(\underbrace{(u^n, V^n) \in T(P_{uv}, \delta, n)}\right) \text{ is } \geq 2^{-n(I(U;V)+\delta'')}$$

To prove this consider first when $\mathcal{U} = \{0,1\}$.

$$u^n = \overbrace{0\,0\,0\,0\,0\,0}^{k}\ \overbrace{1\,1\,1\,1\,1\,1\,1\,1}^{n-k}$$

$\underbrace{np(0)(1 \pm \delta/3)}_{0's}$ $\quad$ $np(1)(1 \pm \delta/3)$ 1's.

$V^n = V_1 \ldots V \quad\quad V_n$

if $V_1 \ldots V_k \in T(P_{V|U=0}, \delta/3, k)$ $\ast$

& $V_{k+1} \ldots V_n \in T(P_{V|U=1}, \delta/3, k)$ $\ast$

then $(u^n, V^n) \in T(P_{uv}, \delta, n)$.

how many $(0,v)$'s do we have?

$$k(p(v|0) \pm \delta/3) = np(0)(1 \pm \delta/3)\, p(v|0)(1 \pm \delta/3)$$

$$= np(0,v)(1 \pm \delta)$$

similarly # of $(1,v)$ in (u^n, V^n) will also be

$$= np(1,v)(1 \pm \delta).$$

So $\Pr((u^n, v^n) \in T(p_{uv}, \delta, n))$

$\geq \Pr(v^k \in T(p_{v|0}, \frac{\delta}{3}, k)) \cdot \Pr((v_{k+1} \ldots v_n) \in T(p_{v|1}, \frac{\delta}{3}, n-k))$

$\geq 2^{-k\left[D(p_{v|0} \| p_v) + \delta''\right]} \; 2^{-(n-k)\left[D(p_{v|1} \| p_v) + \delta''\right]}$

$= 2^{-n\left[\frac{k}{n} D(p_{v|0} \| p_v) + \frac{n-k}{n} D(p_{v|1} \| p_v) + \delta''\right]}$

$= \left[p(0) D(p_{v|0} \| p_v) + p(1) D(p_{v|1} \| p_v) + \delta'''\right]$

$p(0) \sum_v p(v|0) \log \frac{p(v|0)}{p(v)} + p(1) \sum_v p(v|1) \log \frac{p(v|1)}{p(v)}$

$\sum_{u,v} p(u,v) \log \frac{p(v|u)}{p(v)} = I(u;v)$

$= 2^{-n\left[I(u;v) + \delta'''\right]}$

//

Aside: in general for any $\delta_1 < \delta_2$ we will have

for any $u^n \in T(p_U, \delta_1, n)$, V^n iid $\sim p_V$.

$$\Pr\left((u^n, V^n) \in T(p_{UV}, \delta_2, n) \right) \geq 2^{-n[I(U;V)+\delta']}$$

Remarks:

fixed $p_{V|U}$

$$C(\beta) = \max \{ I(U;V) : p_U, \ E[b(U)] = \beta \}$$

$$R(D) = \min \{ I(U;V) : p_{V|U}, \ E(d(U,V)) = D \}$$

given p_U

Examples of $R(D)$ computation:

1. $\mathcal{U} = \{0,1\} = \mathcal{V}$, $d(u,v) = \mathbb{1}\{u \neq v\}$

✓ for $U \sim \text{Bern}(\frac{1}{2})$

2. $\mathcal{U} = \mathcal{V} = \mathbb{R}$ $\quad U \sim N(0, \sigma^2)$ $\quad d(u,v) = (u-v)^2$

Claim: $R(D) = \left(\frac{1}{2} \log \frac{\sigma^2}{D} \right)^+$ for $D \leq \sigma^2$.

$$\cancel{= \begin{cases} \frac{1}{2} \log \frac{\sigma^2}{D} & D \leq \sigma^2 \\ 0 & D > \sigma^2 \end{cases}}$$

Pf: $\min\limits_{\substack{P_{V|U}:\\ E[(U-V)^2]=D}} I(U;V)$

$$I(U;V) = h(U) - h(U|V)$$
$$= h(U) - h(U-V|V)$$
$$\geq h(U) - h(U-V)$$

since $E((U-V)^2) = D$ $\quad h(U-V) \leq \frac{1}{2}\log(2\pi e D)$

$$h(U) = \frac{1}{2}\log 2\pi e \sigma^2$$

$$\Rightarrow I(U;V) \geq \frac{1}{2}\log \sigma^2/D.$$

Can the two inequalities be made equality?

(2nd ineq. requires $U-V$ to be Gaussian, 0-mean.
1st " " $U-V$ to be $\perp$ of V.)

$$U = (U-V) + V$$

$N(0,\sigma^2)$ $\quad$ $N(0,D)$ $\quad$ indep $\quad$ $N(0,\sigma^2-D)$

as long as $\sigma^2 \geq D$ we can indeed find a distribution $f_{UV}(u,v)$ for which $f_U = N(0,\sigma^2)$

$E[(U-V)^2] = D$, $\quad I(U;V) = \frac{1}{2}\log \sigma^2/D$. //