Information Theory & Coding

Oct 26th, 2020

Last week:

"C" of a channel: given a $P(y|x)$

$$C(P) = \max_{p_X} I(X;Y)$$

- seen that $p_X \mapsto I(X;Y)$ is a concave $\cap$ function.

- Convex set $S \subseteq \mathbb{R}^k$: $(x,y \in S \quad 0 \le \lambda \le 1)$

$$\Rightarrow (\lambda x + (1-\lambda) y \in S)$$

- convex / concave function $f : S \mapsto \mathbb{R}$
  $\cup$ $\quad$ $\cap$

$$f(\lambda x + (1-\lambda) y) \underset{\geq}{\leq} \lambda f(x) + (1-\lambda) f(y)$$

- $f : (a,b) \mapsto \mathbb{R}$ is $\cup / \cap$ $\quad f'' \underset{\leq}{\geq} 0$

$p_X \mapsto I(X;Y)$ is $\cap$

$$p_X \equiv \underbrace{(p_X(1), p_X(2) \ldots, p_X(k))}_{\in \mathbb{R}^k}$$

p_X is distribution on $\{1, \ldots, k\}$

with $p_X(k) \geq 0$, $\sum_{k=1}^{K} p_X(k) = 1$

S_k : simplex-k $= \{(p_1 \ldots p_k) : p_i \geq 0, \sum_i p_i = 1\}$

S_k is a convex set.

Properties of vectors which maximize

$$p \longrightarrow f(p)$$
$$\cap \qquad \cap$$
$$S_k \qquad \mathbb{R}$$

<u>Thm</u>: given $f: S_k \longrightarrow \mathbb{R}$, smooth (differentiable)

and given $p = (p_1 .., p_k) \in S_k$. If p maximizes f in the simplex, then

$$\left[\exists \mu \text{ s.t } \quad \frac{\partial f}{\partial p_i} \begin{array}{l} \leq \mu \quad \forall i \\ = \mu \quad \forall i \ p_i > 0. \end{array} \right] \quad \text{KKT conditions}$$

<u>Pf</u>: If p maximizes f, p must also be a local maximum. So lets take a look at a $q \in S_k$

s.t $\quad q_i = p_i$ except $\quad q_1 = p_1 + \varepsilon \qquad \varepsilon > 0$
$$q_2 = p_2 - \varepsilon$$

assume $p_2 > 0$

$$\underline{f(q)} = \underline{f(p)} + \left(\varepsilon \frac{\partial f}{\partial p_1} - \varepsilon \frac{\partial f}{\partial p_2} \right) + \underline{o(\varepsilon)}$$

Since $f(q) \leq f(p)$ it must be the case that

$$\frac{\partial f}{\partial p_1} \leq \frac{\partial f}{\partial p_2}$$

Consequently for every $[i, j \text{ s.t } p_j > 0]$ we must have

$$\frac{\partial f}{\partial p_i} \leq \frac{\partial f}{\partial p_j}.$$

$\Rightarrow$ there exits a constant μ s.t

$$\frac{\partial f}{\partial p_i} = \mu \quad \forall i \text{ s.t } p_i > 0$$

$$\frac{\partial f}{\partial p_i} \leq \mu \quad \forall i$$

//

Thm: Suppose $f : S_K \to \mathbb{R}$ is concave $\cap$. if $S_K \ni p$ satisfies the KKT condition then

$$f(p) \geq f(q) \quad \forall q \in S_K .$$

Pf: Suppose f is concave, $p \in S_K$ satisfying KKT and $q \in S_K$. (we need to prove $f(q) \leq f(p)$.)

take $\lambda \in (0,1)$ and consider

$$f((1-\lambda)p + \lambda q) \geq (1-\lambda) f(p) + \lambda f(q)$$

So $$f(q) \leq \frac{1}{\lambda}\left[f((1-\lambda)p + \lambda q) - (1-\lambda) f(p)\right]$$

$$= f(p) + \frac{1}{\lambda}\left[f(p + \lambda(q-p)) - f(p)\right]$$

In particular

$$f(q) - f(p) \leq \lim_{\lambda \to 0} \frac{1}{\lambda}\left[f(p + \lambda(q-p)) - f(p)\right]$$

By taylor' therem

$$f(p + \lambda(q-p)) = f(p) + \sum_i \lambda(q_i - p_i)\frac{\partial f}{\partial p_i} + o(\lambda^2)$$

So $\dfrac{f(p+\lambda(q-p))-f(p)}{\lambda} = \sum_i (q_i-p_i)\dfrac{\partial f}{\partial p_i} + o(\lambda)$

$\Rightarrow \underbrace{\lim_{\lambda\to 0} \quad "}_{} = \underbrace{\sum_i (q_i-p_i)\dfrac{\partial f}{\partial p_i}}_{}$

if we show $\underbrace{\sum_i (q_i-p_i)\dfrac{\partial f}{\partial p_i}}_{} \leq 0$ we will be done.

To that end, remember that p satisfies KKT so

that $\exists \mu$ s.t. $\dfrac{\partial f}{\partial p_i} \leq \mu \quad \forall i$

$\qquad\qquad\qquad\quad = \mu \quad \forall i$ s.t. $p_i > 0$.

$$\sum_i (q_i-p_i)\frac{\partial f}{\partial p_i} = \sum_{i:\, p_i>0} (q_i-p_i)\mu + \sum_{i:\, p_i=0} \underbrace{(q_i-0)}_{\geq 0}\underbrace{\frac{\partial f}{\partial p_i}}_{\leq \mu}$$

$$\leq \sum_i (q_i-p_i)\mu$$

$$= \mu - \mu \qquad \left(\text{because } \sum_i q_i = 1 = \sum_i p_i\right)$$

$$= 0. \quad //.$$

Conclusion: for $f: S_k \to \mathbb{R}$ concave $\cap$,

KKT is necessary & sufficient for maximization.

Examples: consider the 3-simplex, and

$f(p_1, p_2, p_3) = p_1 p_2 p_3$

$\log f(p_1, p_2, p_3) = \log p_1 + \log p_2 + \log p_3$ is $\cap$

to maximize f $\equiv$ to maximize $\log f$

KKT conditions:

$$\frac{\partial \log f}{\partial p_i} = \mu \quad \forall i$$

$$\frac{1}{p_1} = \mu, \quad \frac{1}{p_2} = \mu, \quad \frac{1}{p_3} = \mu$$

$\Rightarrow$ $p_1 = p_2 = p_3 = \frac{1}{3}$ is the only maximizer.

Ex':

$f(p_1, p_2, p_3) = p_1^2 p_2 p_3^3$

$\log f = 2\log p_1 + \log p_2 + 3\log p_3$ is $\cap$

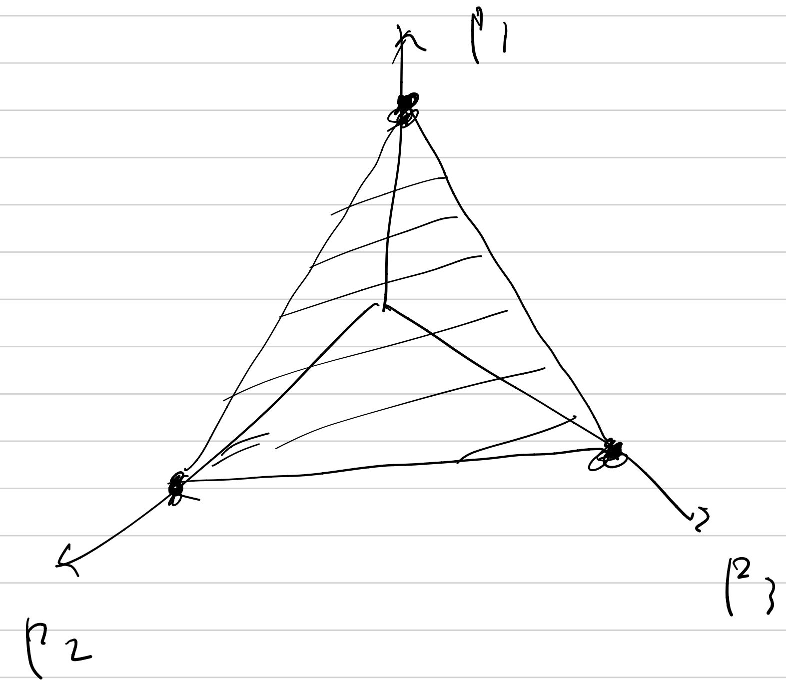
$$\Rightarrow \frac{2}{p_1} = \mu, \quad \frac{1}{p_2} = \mu, \quad \frac{3}{p_3} = \mu$$

$\Rightarrow$ $p_1 = 2p_2$, $p_3 = 3p_2$

$$(p_1, p_2, p_3) = \frac{1}{6}(2, 1, 3) = \left(\frac{1}{3}, \frac{1}{6}, \frac{1}{2}\right)$$

is the maximizer.

3 - simplex :



K-simplex has

$$\left.\begin{array}{l}(1,0,\dots,0)\\(0,1,0,\dots)\\ \vdots \\ (0,\dots,0,1)\end{array}\right\}$$

K standard unit-vectors as its "corner" points.

Let us now specialize the KKT conditions to the problem of maximizing $p_X \longrightarrow f(p_X) = I(X;Y)$

Suppose $\mathcal{X} = \{1,\dots,K\}$
$\mathcal{Y} = \{1,\dots,J\}$ then

$$f(p_X) = f(\underbrace{p_1,\dots,p_K}_{\in S_K}) = \sum_{i,j} p_i P_{j|i} \log_e \frac{P_{j|i}}{\sum_{k=1}^{K} p_k P_{j|k}}$$

$$\left(= \sum_{x,y} p(x)\, p(y|x) \log_e \frac{p(y|x)}{p(y)} \right)$$

$$\frac{\partial f}{\partial p_1} = \sum_{i,j} \frac{\partial}{\partial p_1}\left(p_i P_{j|i} \log_e \frac{P_{j|i}}{\sum_k p_k P_{j|k}} \right)$$

$$= \sum_{i,j} \mathbb{1}\{i=1\} P_{j|i} \log_e \frac{P_{j|i}}{\sum_k p_k P_{j|k}} - p_i P_{j|i} \frac{P_{j|1}}{\sum_k p_k P_{j|k}}$$

$$= \sum_j P_{j|1} \log_e \frac{P_{j|1}}{\sum_k p_k P_{j|k}} - \sum_j P_{j|1} =$$

$$= \sum_j P_{j|2} \log_e \frac{P_{j|2}}{\sum_k p_k P_{j|k}} - 1$$

$$\Rightarrow \frac{\partial f}{\partial p_i} = \sum_j P_{j|i} \log \frac{P_{j|i}}{\sum_k p_k P_{j|k}} - \log e$$

KKT conditions become: p_X maximizes $I(X;Y)$ iff

$$\exists \mu \text{ s.t. } \sum_y P(y|x) \log \frac{P(y|x)}{p_Y(y)} \leq \mu \quad \forall x$$

$$= \mu \quad \forall x \; p(x) > 0$$

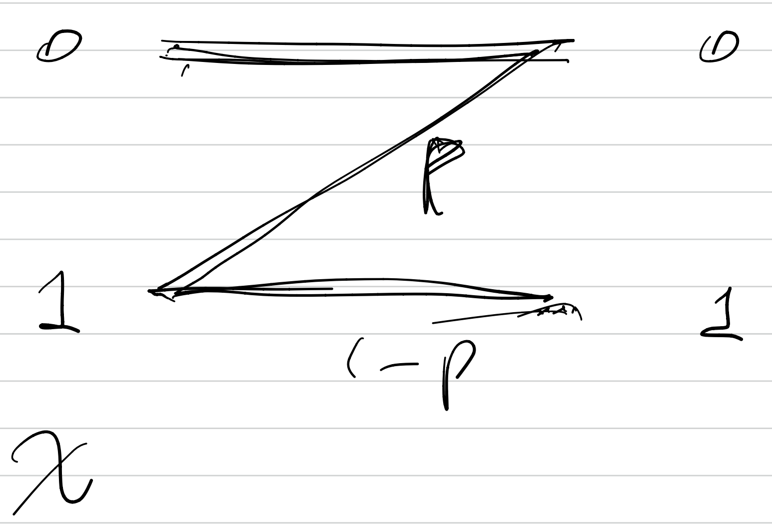
if p_X satisfies KKT then

$$p_X(x) \sum_y P(y|x) \log \frac{P(y|x)}{p_Y(y)} = \mu \, p_X(x) \quad \forall x.$$

$$\Rightarrow \underbrace{\sum_x \quad ''}_{I(X;Y) = C(P)} = \mu$$

So the "μ" in the KKT is $= C(P)$.

Example : P : channel Z-channel (p)



$p(0|0) = 1$

$p(1|0) = 0$

$p(0|1) = p$

$p(1|1) = 1-p$

KKT conditions :

$x=0$

$$1 \log \frac{1}{p_Y(0)} = C$$

$x=1$

$$p \log \frac{p}{p_Y(0)} + (1-p) \log \frac{1-p}{p_Y(1)} = C$$

$$\Rightarrow (1-p) \log \frac{1-p}{p_Y(1)} = \log \frac{1}{p_Y(0)} - p \log \frac{p}{p_Y(0)}$$

$$\Rightarrow (1-p) \log \frac{p_Y(0)}{p_Y(1)} = h_2(p)$$

$$\log \frac{p_Y(0)}{p_Y(1)} = \frac{h_2(p)}{1-p} \Rightarrow \boxed{\frac{p_Y(0)}{p_Y(1)} = 2^{\frac{h_2(p)}{1-p}}}$$

$$\left(p_Y(0), p_Y(1)\right) = \frac{\left(2^{h_2(p)/(1-p)}, 1\right)}{1 + 2^{h_2(p)/(1-p)}}$$

$$C = \log_2\left(1 + 2^{-h_2(p)/(1-p)}\right)$$

The KKT conditions for $p_X \mapsto I(X;Y)$ are useful in finding the $C(\ )$ of channels constructed from other channels:

Ex: $P_1(y_1|x_1)$ $\quad x_1 \in \mathcal{X}_1,\ y_1 \in \mathcal{Y}_1$

given $P_2(y_2|x_2)$ $\quad x_2 \in \mathcal{X}_2,\ y_2 \in \mathcal{Y}_2$.

$C(P_1)$, $C(P_2)$ are already known.

a new channel is defined as follows

$$P(y_1 y_2 | x_1 x_2) = P_1(y_1|x_1)\, P_2(y_2|x_2)$$

we can prove

$C(P) = C(P_1) + C(P_2)$. via KKT conditions

We have now seen some structural aspects of the computation of $C(P)$.

We also have our "bad news" theorem.

$sH > C(P) \Rightarrow$ reliable communication not possible

amount of source entropy to be conveyed / channel use

i.e., there is a $p = f(sH, C(P))$ s.t.

any comm. system has symbol error probability $\geq p$.

We shall embark on a program to show that

at communication rates $R < C(P)$

← # of bits of information / channel use

we can design arbitrarily reliable comm. systems (enc, dec)

namely

Thm: given a channel P (memoryless, stationary)

$R < C(P)$, $\varepsilon > 0$ $\exists$ enc, dec s.t.

$\hat{P}_e(\text{enc}, P, \text{dec}) < \varepsilon$

$\text{rate}(\text{enc}) \geq R$.