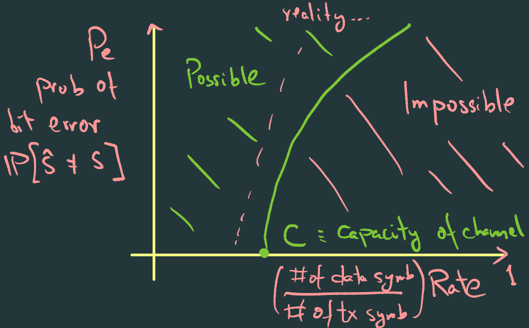
ORIE 4742 - Info Theory and Bayesian ML

Chapter 5: Channel Coding

March 1, 2021

Sid Banerjee, ORIE, Cornell



dependent rv and info content

entropy: basic properties

rv X taking values $\mathcal{X} = \{a_1, a_2, \ldots, a_k\}$, with pmf $\mathbb{P}[X = a_i] = p_i$

Shannon's entropy function

- outcome $X = a_i$ has *information content*: $h(a_i) = \log_2 \left(\frac{1}{p_i}\right)$
- random variable X has *entropy*: $H(X) = \mathbb{E}[h(X)] = \sum_{i=1}^{k} p_i \log_2 \left(\frac{1}{p_i}\right)$

- only depends on distribution of X (i.e., $H(X) = H(p_1, p_2, \ldots, p_k)$)
- $H(X) \geq 0$ for all X
- if $X \sim$ uniform on $\mathcal{X}$, then $H(X) = \log_2 |\mathcal{X}|$; else, $H(X) \leq \log_2 |\mathcal{X}|$
- if $X \perp\!\!\!\perp Y$, then $H(X, Y) = H(X) + H(Y)$

  where joint entropy $H(X, Y) \triangleq \sum_{(x,y)} \underbrace{p(x, y)}_{p(x)p(y)} \log_2 1/\underbrace{p(x, y)}_{p(x)p(y)}$

conditional entropy

conditional entropy

for any rvs X, Y: $H(X|Y) = \sum_{y \in \mathcal{Y}} p(y) H(X|Y=y)$
$$= \sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{X}} p(x|y) \log_2(1/p(x|y))$$

Consider $X \sim (\underbrace{a_1, a_2}_{Y=1}, \underbrace{a_3, a_4}_{Y=0})$ wp (p_1, p_2, p_3, p_4), $Y = \mathbb{1}\{x \in \{a_1, a_2\}\}$

$Y \sim \text{Bin}(p_1 + p_2)$

$Y=1 \Rightarrow X \in (a_1, a_2)$ wp $\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$

$$H(X) = H(Y) + \underbrace{\boxed{(p_1+p_2) H_2\left(\frac{p_1}{p_1+p_2}\right) + (p_3+p_4) H_2\left(\frac{p_3}{p_3+p_4}\right)}}_{H(X|Y)}$$

$Y=0 \Rightarrow X \in (a_3, a_4)$ wp $\left(\frac{p_3}{p_3+p_4}, \frac{p_4}{p_3+p_4}\right)$

the chain rule

the chain rule (information content)

for any rvs X, Y and realizations x, y:

$$\underbrace{h(x,y)}_{\log_2(1/p(x,y))} = h(x) + h(y|x) = \underbrace{h(y)}_{\log_2(1/p(y))} + \underbrace{h(x|y)}_{\log_2(1/p(x|y))}$$

$$\log_2(p(x,y)) = \log_2(p(x)\,p(y|x))$$

$$= \log_2(p(x)) + \log_2(p(y|x))$$

the chain rule

the chain rule (entropy)

for any rvs X, Y:

$$\underbrace{H(X,Y)}_{\sum_{(x,y)} p(x,y)\log_2 \underbrace{\frac{1}{p(x,y)}}_{\frac{1}{p(y)p(x|y)}}} = \left[H(X) + H(Y|X) = \underbrace{H(Y)}_{\sum p(y)\log_2 \frac{1}{p(y)}} + \underbrace{H(X|Y)}_{\sum p(x,y)\log_2 \frac{1}{p(x|y)}} \right]$$

Note: $H(X) - H(X|Y) = H(Y) - H(Y|X)$

mutual information (information gain ...)

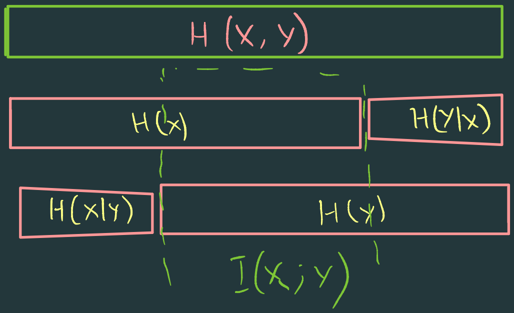
mutual information

for any rvs X, Y:

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

moreover, given any other conditioning rv Z

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(Y|Z) - H(Y|X,Z)$$



blank

Alt defn.

$$I(X;Y) = D_{KL}\Big(\underset{\substack{\uparrow\\ \text{want to}\\ \text{encode}}}{P(X,Y)} \,\Big\|\, \underset{\substack{\uparrow\\ \text{use to}\\ \text{encode}}}{P(X)P(Y)}\Big)$$

$$\geq 0 \quad \text{(by Gibbs' Ineq)}$$

Pf. RHS $=$

$$\sum_{(x,y)} p(x,y)\log_2\left(\frac{p(x,y)}{p(x)p(y)}\right) = \sum_{x,y} p(y)p(x|y)\log_2\left(\frac{\cancel{p(x)}p(y|x)}{\cancel{p(x)}p(y)}\right)$$

$$= \underbrace{\sum p(y)\log_2\left(\frac{1}{p(y)}\right)} - \underbrace{\sum p(x,y)\log_2\left(\frac{1}{p(y|x)}\right)}$$

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

example

$P(x,y)$		x = 1	2	3	4	$P(y)$
y	1	$1/8$	$1/16$	$1/32$	$1/32$	$1/4$
	2	$1/16$	$1/8$	$1/32$	$1/32$	$1/4$
	3	$1/16$	$1/16$	$1/16$	$1/16$	$1/4$
	4	$1/4$	0	0	0	$1/4$
$P(x)$		$1/2$	$1/4$	$1/8$	$1/8$	

$$
P(x|y):\quad
\begin{array}{cc|cccc}
 & & & x & & \\
 & & 1 & 2 & 3 & 4 \\
\hline
 & 1 & 1/2 & 1/4 & 1/8 & 1/8 \\
y & 2 & 1/4 & 1/2 & 1/8 & 1/8 \\
 & 3 & 1/4 & 1/4 & 1/4 & 1/4 \\
 & 4 & 1 & 0 & 0 & 0
\end{array}
\qquad
P(y|x):\quad
\begin{array}{cc|cccc}
 & & & x & & \\
 & & 1 & 2 & 3 & 4 \\
\hline
 & 1 & 1/4 & 1/4 & 1/4 & 1/4 \\
y & 2 & 1/8 & 1/2 & 1/4 & 1/4 \\
 & 3 & 1/8 & 1/4 & 1/2 & 1/2 \\
 & 4 & 1/2 & 0 & 0 & 0
\end{array}
$$

$h(x)\quad 1 \quad 2 \quad 3 \quad 3, \quad H(x) = 7/4$

$h(y)\quad 2 \quad 2 \quad 2 \quad 2, \quad H(y) = 2$

$$H(X,Y) = 2\cdot\left(\frac{1}{4}\right) + 3\left(\frac{2}{8}\right) + 4\left(\frac{6}{16}\right) + 5\left(\frac{4}{32}\right) = \frac{27}{8}$$

$$H(X)+H(Y) = \frac{30}{8} > H(X,Y)$$

$$H(X|Y) = \frac{1}{4}\cdot\left(\frac{7}{4} + \frac{7}{4} + 2 + 0\right) = \frac{11}{8}$$

$$H(Y|X) = \frac{1}{2}\cdot\left(\frac{7}{4}\right) + \frac{1}{4}\left(\frac{3}{2}\right) + \frac{1}{8}\left(\frac{3}{2}\right) + \frac{1}{8}\left(\frac{3}{2}\right) = \frac{13}{8}$$

$$\left.\begin{aligned} H(X)+H(Y|X) &= \frac{7}{4} + \frac{13}{8} = \frac{27}{8} \\ H(Y)+H(X|Y) &= 2 + \frac{11}{8} = \frac{27}{8} \end{aligned}\right\} = H(X,Y)$$

$$\left.\begin{aligned} H(X)-H(X|Y) &= \frac{7}{4} - \frac{11}{8} = \frac{3}{8} \\ H(Y)-H(Y|X) &= 2 - \frac{13}{8} = \frac{3}{8} \end{aligned}\right\} = I(X;Y)$$

mutual information and KL divergence

mutual information

for any rvs X, Y: $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$

$$= D_{KL}\left(P(x,y) \,||\, P(x)P(y)\right)$$

= loss in encoding length when we approximate $P(x,y)$ by $P(x)P(y)$

Some properties

- Suppose $X \perp\!\!\!\perp Y$

$H(X|Y) = H(X)$, $H(Y|X) = H(Y)$ ← (check...)

$\Rightarrow I(X;Y) = 0$

- Suppose $Y = f(X)$ (deterministic fn)

$H(Y|X) = 0$ ($H(X|Y) = ?$ 0 if f invertible)

$\Rightarrow I(X;Y) = H(Y)$ ($= H(X)$ if f invertible)

Data Processing Inequality

Markov Chain

$$X \longrightarrow Y \longrightarrow Z$$

Source | data | processed data

X and Y are conditionally indep given Z

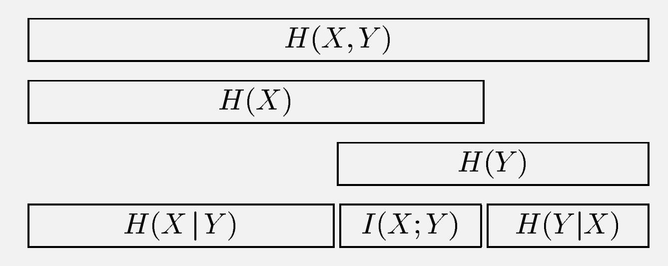
Data proc^{g} ineq

$$\boxed{I(X;Z) \leq I(X;Y)}$$

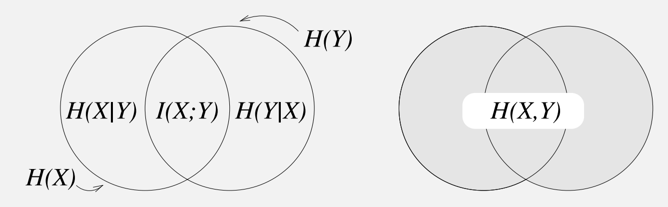
= iff Z is a sufficient stats

- $X=Y,\ Z=0 \Rightarrow I(X;Z)=0,\ I(X;Y)=H(X)$
- $Y \perp\!\!\!\perp X,\ Z=Y \Rightarrow I(X;Y)=0,\ I(X;Z)=0$

visualizing mutual information



Correct picture



misleading