Lecture 8

Overall,

$$\boxed{I(X;Y) = I(Y;X) = H(X) - H(X|Y) = H(Y) - H(Y|X)}$$

$H(X) = \frac{7}{4}$ bits

$I(X;Y) = \frac{3}{8}$ bits

$H(X|Y) = \frac{11}{8}$ bits

$H(Y) = 2$ bits

$I(Y;X) = \frac{3}{8}$ bits

$H(Y|X) = \frac{13}{8}$ bits

Conditioning can help achieve better compression! Gain = I(X;Y)

- Joint Entropy and Mutual Information

Joint Entropy: Uncertainty about the joint sources X and Y.

$$H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 \underbrace{p(x,y)}$$

$p(x,y) = p(y)\cdot p(x|y)$, by Bayes' Rule

$$= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 p(y) - \sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 p(x|y)$$

$$= H(Y) + H(X|Y)$$

Similarly,

$$H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 \underbrace{p(x,y)}$$

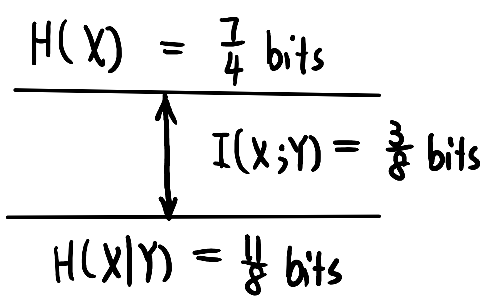
$p(x,y) = p(x)\cdot p(y|x)$, by Bayes' Rule

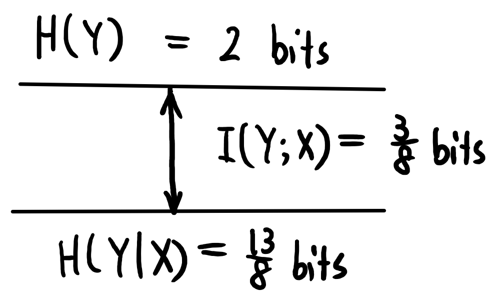
$$= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 p(x) - \sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 p(y|x)$$

$$= H(X) + H(Y|X)$$

In summary,
H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) = 7/4 + 13/8 = **27/8 bits** = 2 + 11/8

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





Compare: H(X,Y) = 27/8 bits < H(X) + H(Y) = 7/4 + 2 = 15/4 bits,
H(X) + H(Y) - H(X,Y) = 15/4 - 27/8 = 3/8 bit = I(X;Y)

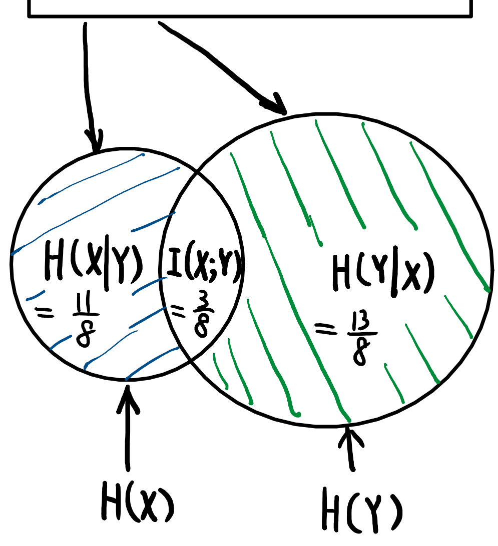
Blocking can help achieve better compression! Gain = I(X;Y)

Go back to the relation between joint entropy and mutual information:

$\boxed{\text{H(X,Y) = H(X) + H(Y|X) = H(X) + H(Y) - I(X;Y)}}$

Where I(X;Y) = I(Y;X) = H(X) - H(X|Y) = H(Y) - **H(Y|X) => H(Y|X) = H(Y) - I(X;Y)**

$\boxed{\text{H(X,Y) = H(X) + H(Y) - I(X;Y)}}$



$H'(X)$ $\qquad$ $H'(Y)$

$= \frac{7}{4}$ $\qquad$ $= 2$

If X and Y are independent, then $P(x,y) = P(x)\cdot P(y)$, and $P(x|y) = P(x)$

$P(x,y) = P(x|y)\cdot P(y)$ in general

- Mutual Information I(X;Y)

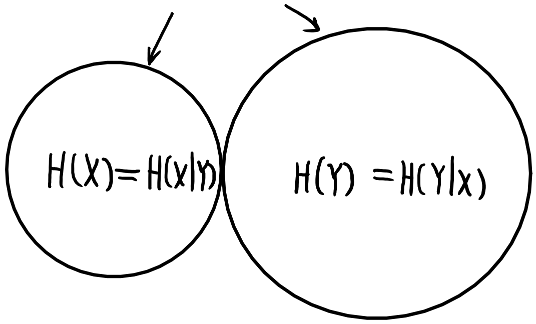
$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} P(x,y)\log_2 \left[\underbrace{\frac{P(x|y)}{P(x)}}_{\frac{P(x)}{P(x)}=1}\right] = \sum_{x\in X}\sum_{y\in Y} P(x,y)\log_2 1 = 0$$

$\boxed{H(X,Y) = H(X) + H(Y|X) = H(X) + H(Y)}$

And $H(Y) = H(Y|X)$

Similarly, $H(X) = H(X|Y)$

$H(X,Y) = H(X) + H(Y)$



- KL Divergence (KL Distance) between two distributions

Kullback - Leibler

P(i) and Q(i)

$$D_{KL}(P||Q) = \sum_i \left\{ P(i)\log_2\left[\frac{P(i)}{Q(i)}\right]\right\},\quad \text{where } i = 1, 2, \cdots,$$

$D_{KL}(X||Y)$

$$= \sum^{4} P(X=i)\log_2 \frac{P(X=i)}{P(Y\ \ldots}$$

Joint Distribution P(X,Y):

Y \ X	1	2	3	4	Σ
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$	$P(Y=1) = \frac{1}{4}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$	$P(Y=2) = \frac{1}{4}$

$$= \sum_{i=1}^{4} P(X=i) \log_2 \frac{P(X=i)}{P(Y=i)}$$

$$= \frac{1}{2}\log_2 \frac{\frac{1}{2}}{\frac{1}{4}} + \frac{1}{4}\log_2 \frac{\frac{1}{4}}{\frac{1}{4}} + \frac{1}{8}\log_2 \frac{\frac{1}{8}}{\frac{1}{4}} + \frac{1}{8}\log_2 \frac{\frac{1}{8}}{\frac{1}{4}}$$

$$= \frac{1}{2} + 0 - \frac{1}{8} - \frac{1}{8} = \frac{1}{4}$$

1	$\overline{8}$	$\overline{16}$	$\overline{32}$	$\overline{32}$	$P(Y=1) = \overline{4}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$	$P(Y=2) = \frac{1}{4}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$P(Y=3) = \frac{1}{4}$
4	$\frac{1}{4}$	0	0	0	$P(Y=4) = \frac{1}{4}$
Σ	$P(X=1)$	$P(X=2)$	$P(X=3)$	$P(X=4)$	
	$\Downarrow$	$\Downarrow$	$\Downarrow$	$\Downarrow$	
	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	

$$D_{KL}(Q\|P) = \sum_i \left\{ Q(i) \log_2 \left[\frac{Q(i)}{P(i)}\right] \right\}, \quad \text{where } i = 1, 2, \cdots,$$

How about $D_{KL}(Y\|X) = ?$

In general, $D_{KL}(P\|Q) \neq D_{KL}(Q\|P)$

- Look back at the Mutual Information I(X;Y)

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2\left[\frac{p(x|y)}{p(x)}\right] = \sum_{x\in X}\sum_{y\in Y}\boxed{p(x,y)}\log_2\left[\frac{\boxed{p(x,y)}}{p(x)\,p(y)}\right]$$

Thus, the mutual info. Is the KL distance between the Joint PMF and Product of marginal PMFs.

I(X;Y) = D_{KL}\{P(x,y) || P(x)P(y)\}

$$= \sum_{x\in X}\sum_{y\in Y} P(x,y)\log_2\frac{P(x,y)}{P(x)P(y)}$$

$$= P(X=1,Y=1)\log_2\frac{P(X=1,Y=1)}{P(X=1)\cdot P(Y=1)}$$

$+\cdots$

16 terms in total

$= \frac{3}{8}$ bit (Verify this!)

Joint Distribution P(X,Y):

Y \ X	1	2	3	4	Σ
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$	$P(Y=1) = \frac{1}{4}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$	$P(Y=2) = \frac{1}{4}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$P(Y=3) = \frac{1}{4}$
4	$\frac{1}{4}$	0	0	0	$P(Y=4) = \frac{1}{4}$
Σ	$P(X=1) \Downarrow \frac{1}{2}$	$P(X=2) \Downarrow \frac{1}{4}$	$P(X=3) \Downarrow \frac{1}{8}$	$P(X=4) \Downarrow \frac{1}{8}$	

In X and Y are independent, P(x,y) = P(x)P(y), thus P(x,y)/P(x)P(y) = 1,
I(X;Y) = D_{KL}\{P(x,y) || P(x)P(y)\} = 0.