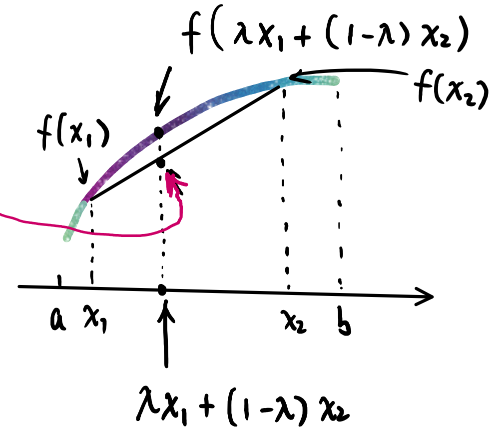
Lecture 11

Jensen's Inequality (cont'd)

- If f(x) is a concave function over an interval (a,b), then for every x_1 and x_2 inside the interval, and $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1-\lambda) x_2) \geq \lambda f(x_1) + (1-\lambda) f(x_2)$$



Show that the KL Divergence is always **non-negative**.

- KL Divergence (KL Distance) between two distributions

Kullback - Leibler

P(i) and Q(i)

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

$$\Downarrow$$

$$-D_{KL}(P \| Q) = \sum_i P(i) \underbrace{\log_2}_{f(\cdot)} \underbrace{\left[\frac{Q(i)}{P(i)} \right]}_{Y} \triangleq E[f(Y)], \text{ where } f(\ ) \text{ is a concave function}$$

By Jensen's Inequality,

$$E[\log_2(X)] \leq \log_2 E[X]$$

Thus

$$-D_{KL}(P \| Q) = E[\log_2 f(Y)] \leq \log_2 E[Y] = \log_2 \sum_i \left[P(i) \cdot \frac{Q(i)}{P(i)} \right] = \log_2 1 = 0$$

$$\boxed{D_{KL}(P\|Q) \geq 0}$$

Previously,

Cross Entropy: $H_c(A,B) = H(A) + D_{KL}(A\|B) \geq H(A)$

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

$I(X;Y) = D_{KL}\{P(x,y) \mid\mid P(x)P(y)\} \geq 0$

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

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

Where $I(X;Y) = I(Y;X) = H(X) - H(X|Y) = H(Y) - $ **$H(Y|X) \Rightarrow H(Y|X) = H(Y) - I(X;Y)$**

- Coding
  Assigning binary sequences to elements of an alphabet.

  Alphabet: set of symbols (called "letters")

  Code: The set of binary sequences, where the individual members are called the codewords.

  Rate of the Code: Average number of bits/symbol $\underbrace{\text{(avg. codeword length)}}_{\text{ACL}}$

Example: Alphabet = $\{a_1, a_2, a_3, a_4\}$

Letter	Probability	Code
a_1	0.5	00
a_2	0.25	01
a_3	0.125	10
a_4	0.125	11

$$ACL = 0.5 \times 2 + 0.25 \times 2 + 0.125 \times 2 + 0.125 \times 2$$

$$= 2 \text{ bits/symbol}$$

Alternatively,

Example: Alphabet = $\{a_1, a_2, a_3, a_4\}$

Letter	Probability	Code
a_1	0.5	1110
a_2	0.25	110
a_3	0.125	10
a_4	0.125	0

$$ACL = 0.5 \times 4 + 0.25 \times 3 + 0.125 \times 2 + 0.125 \times 1$$

$$= 2 + 0.75 + 0.25 + 0.125$$

$$= 3.125 \text{ bits/symbol}$$

Or, **shorter code for more likely letters**:

Example: Alphabet = $\{a_1, a_2, a_3, a_4\}$

Letter	Probability	Code
a_1	0.5	0
a_2	0.25	10
a_3	0.125	110
a_4	0.125	111

ACL = 0.5*1+ 0.25*2+ 3*0.125+ 3*0.125 = 1.75 bits/symbol

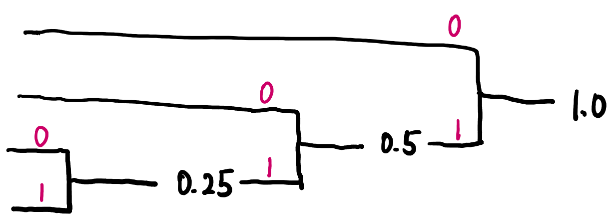
Unique decodable?

$$\begin{array}{cccc} a_1, & a_3, & a_2, & a_4 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 0 & 110 & 10 & 111 \end{array} \xrightarrow{\text{Encoder}} 0\,110\,10\,111 \xrightarrow{\text{Decoder}} a_1, a_3, a_2, a_4$$

$$\begin{array}{cccc} a_1, & a_3, & a_4, & a_2 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 0 & 110 & 111 & 10 \end{array} \rightarrow 0110111 10 \xrightarrow{\text{Decoder}} a_1, a_3, a_4, a_2$$

Huffman Coding

Code	Letter	Prob
0	a_1	0.5
10	a_2	0.25
110	a_3	0.125
111	a_4	0.125



$$H(0.5, 0.25, 0.125, 0.125) = -0.5\log_2 0.5 - 0.25\log_2 0.25 - 2 \times 0.125 \times \log_2 0.125$$

$$H(0.5, 0.25, 0.125, 0.125) = -0.5\log_2 0.5 - 0.25\log_2 0.25 - 2\times 0.125\times \log_2 0.125$$

$$= 0.5 + 0.25\times 2 + 0.25\times 3 = 1.75 \text{ bits/symbol}$$

ACL = 0.5*1+ 0.25*2+ 3*0.125+ 3*0.125 = 1.75 bits/symbol