Reading Assignment

symbol codes (Mackay chapter 5)

symbol codes

expected length of symbol code

let $X \sim \{p(x)\}_{x \in \mathcal{X}}$, and consider code $C(\cdot)$, and let $\ell(x) = |C(x)|$

the expected length of C is $\mathbb{E}[L(C, X)] = \sum_x p(x)\ell(x)$

what we want from symbol code C:

- unique decodability: $\forall\, x_1 x_2 \dots x_n \neq y_1 y_2 \dots y_n$, we have $C(x_1)C(x_2)\dots C(x_n) \neq C(y_1)C(y_2)\dots C(y_n)$ $\Leftrightarrow$ lossless
- easy to decode
- small $\mathbb{E}[L(C, X)]$

types of symbol codes

$H(X) = 1.75$

$\overbrace{\phantom{\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{8}}}$

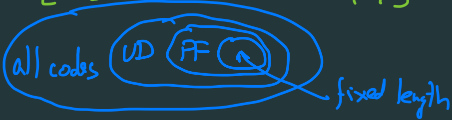
1 2 3 3

consider source producing $X \sim \{a, b, c, d\}$ with prob $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$

Symbols	info content ($h(x)$)	Code 1 (one-hot encoding)	$l(x)$	Code 2 (binary encoding)	$l(x)$
a	1	1000	4	00	2
b	2	0100	4	01	2
c	3	0010	4	10	2
d	3	0001	4	11	2
	$H(X) = 1.75$		$E[l(x)] = 4$		2

prefix codes (variable length, greedy encoding/decoding)

Symbol	info content $h(x)$	Code 3 (prefix-free)	$\ell(x)$	Code 4 (uniquely decodable)	$\ell(x)$
a	1	0	1	0	1
b	2	10	2	01	2
c	3	110	3	011	3
d	3	111	3	111	3
	$H(x) = 1.75$		$E[\ell(x)] = 1.75$		1.75



the limits of unique decodability

Kraft-McMillan inequality (conservation laws)

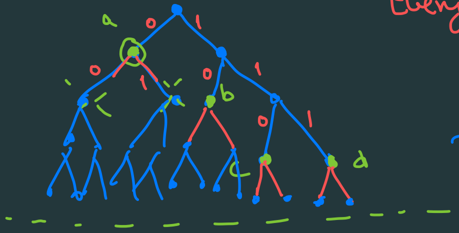
for any $C \equiv$ uniquely decodable binary code over $\mathcal{X}$, with $\ell(x) = |C(x)|$

$$\sum_{x \in \mathcal{X}} \underbrace{2^{-\ell(x)}}_{\text{fraction of leaf nodes in partition } x} \leq 1$$

moreover, for any $\{\ell(x)\}$ satisfying this, we can find a prefix code

- (Kraft's Ineq) Special case for prefix-free

Every prefix free code $\equiv$ subset of nodes in a binary tree



L_{max}

$2^{L_{max}}$ 'leaf node' partitioned among $|\mathcal{X}|$ symbols

Kraft's symbol-code supermarket

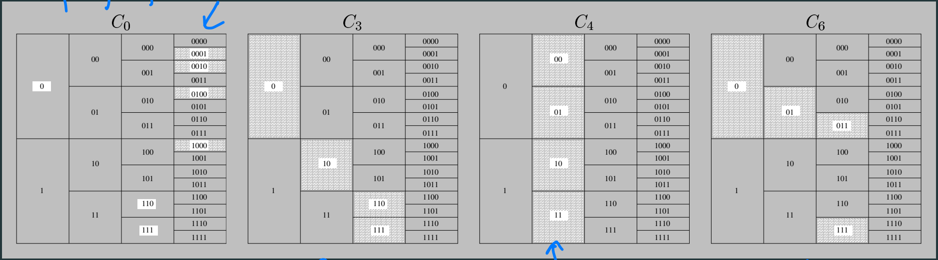
Kraft's inequality: for prefix codes $\sum_{x \in \mathcal{X}} 2^{-\ell(x)} \leq 1$

Costs

1/2	1/4	1/8	1/16	
$\boxed{0}$	00	000	0000	The total symbol code budget
			0001	
		001	0010	
			0011	
	01	010	0100	
			0101	
		011	0110	
			0111	
1	$\boxed{10}$	100	1000	
			1001	
		101	1010	
			1011	
	11	$\boxed{110}$	1100	
			1101	
		111	1110	
			$\boxed{1111}$	

Kraft's symbol-code supermarket

not prefix free! but obeys KM



prefix-free

fixed length

not prefix-code

defn - If $\sum_{x \in \mathcal{X}} 2^{-l(x)} = 1 \implies$ complete code

optimizing expected code length

- entropy of X: $H(X) = \sum_{i \in \mathcal{X}} p_i \log_2 \left(\frac{1}{p_i}\right)$
- Kraft-McMillan inequality: UD code $\{\ell_i\}_{\{i \in \mathcal{X}\}}$ satisfies $\sum_{i \in \mathcal{X}} 2^{-\ell_i} \leq 1$

$$\min \sum_{x \in \mathcal{X}} p(x)\, \ell(x) \quad \text{s.t} \quad \sum_{x \in \mathcal{X}} 2^{-\ell(x)} \leq 1, \quad \ell(x) \in \{1, 2, \ldots\}$$

Kraft-McMillan Ineq

- Idea 1 - $q(x) = 2^{-\ell(x)} / Z$, where $Z = \sum_{x \in \mathcal{X}} 2^{-\ell(x)} \leq 1$

$$\Rightarrow \min \sum_{x \in \mathcal{X}} p(x) \log_2 \left(\frac{1}{Z\, q(x)} \cdot \frac{p(x)}{p(x)} \right)$$

$$= H(X) + \sum_{x \in \mathcal{X}} p(x) \log_2 \left(\frac{p(x)}{q(x)} \right) + \overbrace{\log\left(\frac{1}{Z}\right)}^{\geq 0}$$

optimizing expected code length

$\left(q(x) = 2^{-\ell(x)}/Z \right)$

let $X \sim \{p(x)\}_{x \in \mathcal{X}}$, and consider code $C(\cdot)$, and let $\ell(x) = |C(x)|$
the expected length of C is $\mathbb{E}[L(C, X)] = \sum_x p(x)\ell(x)$

For any 0-error (uniquely decodable) code

$$\mathbb{E}[\ell(x)] = H(x) + \underbrace{D_{KL}(P \| Q)}_{\text{'misrepresentation loss'}} + \underbrace{\log_2(1/Z)}_{\text{'rounding loss'}}$$

- $Z = \sum_{x \in \mathcal{X}} 2^{-\ell(x)} \leq 1$ $\left(\Rightarrow \log_2(1/Z) \geq 0 \right.$ $\left. = 0 \text{ for complete codes} \right)$

- $D_{KL}(P \| Q) = \sum_x p(x) \log_2\left(\frac{p(x)}{q(x)}\right) = \underbrace{H_p(q)}_{\text{cross-entropy}} - H(p)$

$\geq 0 \ \forall p, q$ (Gibbs Ineq)

cross-entropy $\sum_x p(x) \log \frac{1}{q(x)}$

relative entropy and Gibb's inequality (≈ distance between distⁿ)

relative entropy (or Kullback-Leibler (KL) divergence)

the relative entropy $D_{KL}(p||q)$ between two distributions $p(x)$ and $q(x)$ defined over alphabet $\mathcal{X}$ is

$$D_{KL}(P||Q) = \sum_{x \in \mathcal{X}} p(x) \ln \left(\frac{p(x)}{q(x)} \right) \geq 0 \quad \forall P, Q$$

(Gibb's Ineq)

KL div ≈ 'distance' between 2 dist p, q

- Note - Not symmetric, $D(P||Q) \neq D(Q||P)$
- $D(P||P) = 0$
- Note - (usually) P and Q have same support

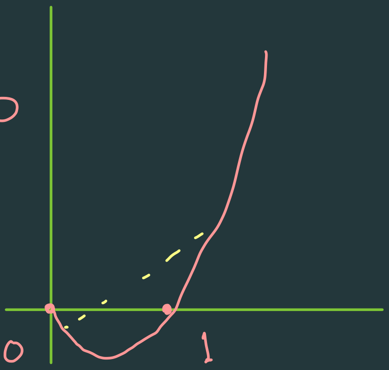
the function $\phi(x) = x \ln x$ $= \ln x / (1/x)$

$\phi'(x) = 1 + \ln x$

$\phi''(x) = \frac{1}{x} > 0 \ \forall x > 0$

$\Rightarrow \phi(x)$ is convex

$\Rightarrow \mathbb{E}[\phi(X)] \geq \phi(\mathbb{E}[X])$



relative entropy and Gibb's inequality

the relative entropy $D_{KL}(p||q) = \sum_{x \in \mathcal{X}} p(x) \ln \left(\frac{p(x)}{q(x)}\right) \geq 0$ for all p, q

$$= \sum_{x \in \mathcal{X}} q(x) \left(\frac{p(x)}{q(x)}\right) \ln \left(\frac{p(x)}{q(x)}\right)$$

$$= \mathbb{E}\left[Y \ln Y \right] \quad \text{where } Y = \frac{p(x)}{q(x)} \text{ w.p. } q(x)$$

$$\geq \mathbb{E}[Y] \ln \mathbb{E}[Y], \quad \mathbb{E}[Y] = \sum_{x} \left(\frac{p(x)}{q(x)}\right) q(x) = 1$$

$$= 0$$

optimizing expected code length

from before

$$\mathbb{E}[L(X)] \geq H(X) + \underbrace{D(P \| Q)}_{\geq 0} + \underbrace{\log_2\left(\frac{1}{Z}\right)}_{\geq 0}$$

$Q \sim 2^{-\ell(x)}/Z$

(and $= 0$ if $P = Q$)

by KM

($= 0$ for complete codes)

$$\Rightarrow \boxed{\mathbb{E}[L(X)] \geq H(X)} \quad \forall \text{ lossless codes}$$

optimizing expected code length

$\left(q(x) = 2^{-\ell(x)} / Z \right)$

To get $\mathbb{E}[L(X)] = H(X)$, need

1) Choose $\ell(x)$ s.t. $2^{-\ell(x)} = Z\, p(x) \;\forall x$

ie, $\underbrace{\ell(x) = \log_2 \frac{1}{p(x)} = h(x)}$

length of codeword = info content of each symbol

2) Set $Z = \sum_{x \in \mathcal{X}} 2^{-\ell(x)} = 1$

(what if we don't know $h(x)$ exactly ?)

aside: cross entropy

$\hat{l}(x)$ for some code

the cross entropy of p given q: $H_p(q) = \sum_{x \in \mathcal{X}} p(x) \underbrace{\ln\left(\frac{1}{q(x)}\right)}$
– avg length of message from if 'p mis-estimated as q'

$$D(P\|Q) = \underbrace{H_p(Q)}_{\geq 0} - \underbrace{H(P)}_{\geq 0} \geq 0$$

= Increase in length when source P is encoded based on distr Q

how good is the best symbol code?

i.e. - How good is the Huffman Code?

$$\mathbb{E}\left[L_{Huff}(X)\right] = \sum_{x \in \mathcal{X}} p(x) \left\lceil \log_2 \frac{1}{p(x)} \right\rceil$$

$$\leq \sum_{x \in \mathcal{X}} p(x) \left[\log_2\left(\frac{1}{p(x)}\right) + 1 \right]$$

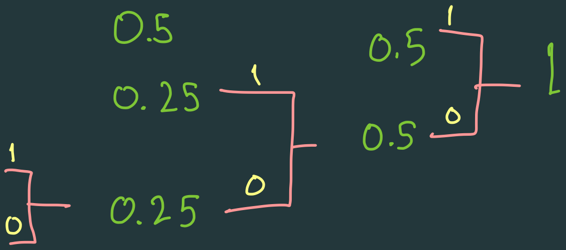
$$= H(X) + 1$$

Huffman code

consider $X \sim \{a, b, c, d\}$ with prob $\{0.5, 0.25, 0.125, 0.125\}$

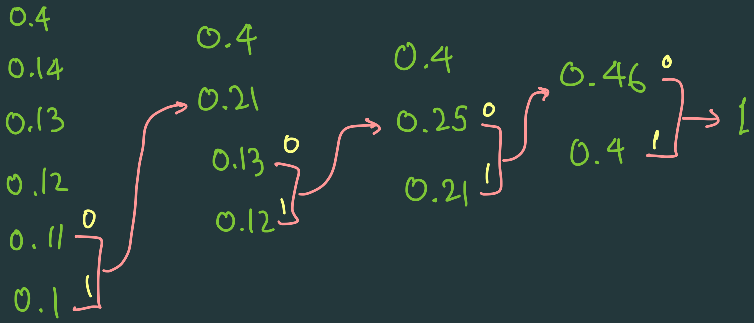
Idea - Solve max $\sum p(x) \, l(x)$ s.t $l(x) \equiv$ prefix-free

Code		
1	a	0.5
01	b	0.25
001	c	0.125
000	d	0.125



Huffman code

consider $X \sim \{a, b, c, d, e, f\}$ with prob $\{0.4, 0.14, 0.13, 0.12, 0.11, 0.10\}$



aside: information content in a perfect code

$\sum 2^{-\ell(x)} = 1$

let C be a perfect code for X, and given database $X_1 X_2 \ldots X_n$, suppose we pick one bit at random from the encoded sequence $C(X_1)C(X_2)\ldots C(X_n)$. what is the probability this bit is a 1?

$$\sum p(x) f_1(x) = \frac{1}{2}\cdot 0 + \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{8}\cdot\frac{2}{3} + \frac{1}{8}\cdot 1$$

$$= \frac{1}{3} \times \left(\text{sampling bias}\right)$$

X	Prob	Code	fraction of 1
a	1/2	0	0
b	1/4	10	1/2
c	1/8	110	2/3
d	1/8	111	1

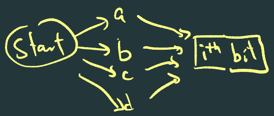
Note - complete code!

D = a a b a c b d c a a ...

L(D) = 0 0 10 0 110 10 111 110 0 0

$\mathbb{P}[B_i = 0]$ vs $\mathbb{P}[B_i = 1]$

Correct way - $\mathbb{P}[B_i = 1] = \dfrac{\sum p_i \cdot l_i \cdot f_i}{\sum p_i \cdot l_i}$



$$= \frac{0 + \frac{1}{4} + \frac{1}{8}\cdot 2 + \frac{1}{8}\cdot 3}{1.75}$$

$$= \frac{1}{2}$$