Continuation of Week 3

Huffman coding

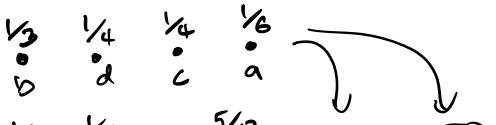
RecursiveHuffman:
Input: Forest f_{in}
Output: Forest f_{out}

if #trees in $f_{in} = 1$ then return $f_{out} = f_{in}$

else
create forest f' by joining two roots with smallest probabilities
return f_{out} = RecursiveHuffman(f')

probabilities	1/6,	1/3,	1/4,	1/4
letters	a	b	c	d

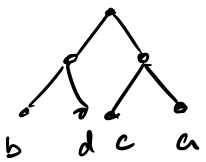
Step I :



Step II :

Step III :

Step IV :



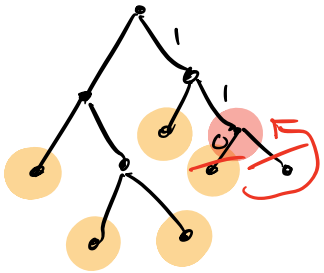
...

$\Rightarrow$ b → 00, d → 01
c → 10, a → 11

Lemma: There exists an optimal prefix code with the following property:

- The two longest codewords are of equal length and are siblings, and their respective symbols have lowest probability

Proof Any prefix code can be represented as a tree with codewords on leaves.



Claim:
← this cannot be of minimal expected length

⇒ optimal prefix codes have all leaves occupied

Theorem: (Optimality of Huffman codes)
Given any source X, any code constructed using the Huffman algorithm achieves the minimal expected codeword length for any prefix code.

Proof: By induction: n is number of symbols

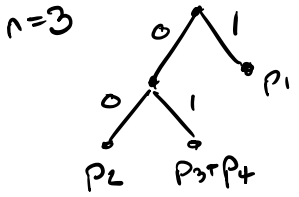
$n=2$: Huffman produces optimal code!

$n-1 \to n$: Assume $p_1 \geq p_2 \geq \dots \geq p_n$ are ordered,

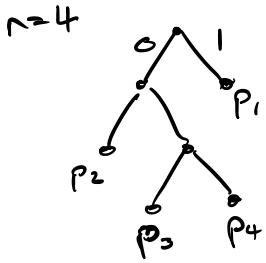
By induction Huffman for $(p_1, p_2, \dots, p_{n-1}+p_n)$ yields an optimal code for $n-1$ symbols.

⇒ expected length L^*_{n-1}

Huffman for $(p_1, p_2, \dots, p_{n-1}, p_n)$ will produce a tree with $L_n = L^*_{n-1} + p_{n-1} + p_n$ (1)



$$L_3 = p_1 + 2p_2 + 2(p_3+p_4)$$



$$L_4 = p_1 + 2p_2 + 3p_3 + 3p_4$$

$$\Rightarrow L_4 = L_3 + p_3 + p_4$$

Moreover, $L_{n-1}^* \leq L_n^* - p_{n-1} - p_n$ ②

①+②: $L_n \leq L_n^* \Rightarrow L_n = L_n^*$ is optimal. □

2.3 Fixed-length block codes

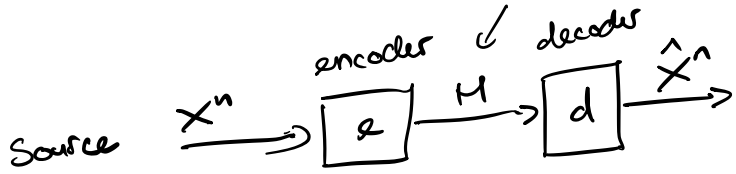
variable length
single symbol
lossless compression

⟷

Fixed-length
block encoding
lossy compression

Block coding

$M \in [|M|]$
codeword, fixed length



$$X^n = (X_1, X_2, \dots X_n)$$

If $P[X^n = \hat{X}^n] = 1$ then we have lossless compression

$$\dots |\text{supp}\{P_X\}|$$

$n=1$: $|M| = |\{x : P_X(x) > 0\}|$

$\Rightarrow$ codeword length: $\lceil \log |\operatorname{supp}\{P_X\}| \rceil$

$\lceil \log |\operatorname{supp}\{P_X\}| \rceil$ can be much larger than $H(X)$

general n: $|M| = |\{x : P_X(x) > 0\}|^n = |\operatorname{supp}\{P_X\}|^n$

$\Rightarrow$ codeword length $\lceil n \log |\operatorname{supp}\{P_X\}| \rceil$

Lossless compression with fixed-length codes is not very efficient!

Definition (Block code) An $(n, 2^L)$-code consists of

- an encoder $e : \mathcal{X}^n \to \{0,1\}^L$ and
- a decoder $d : \{0,1\}^L \to \mathcal{X}^n$

Definition: (Achievable rate) A rate R is achievable for a RVS X if there exist a sequence of $(n, 2^{\lfloor nR \rfloor})$-codes with encoders e_n and decoders d_n s.t.

$$\lim_{n\to\infty} P(\hat{X}^n \neq X^n) = 0$$

where $\hat{X}^n = d_n(M_n)$, $M_n = e(X^n)$, $X^n \sim P_X$

The rate is the ratio between codeword length (in bits) and the number of source symbols that are stored

$\Rightarrow$ number of bits per source symbol!

Def. (Optimal source coding rate)

$R^*(X) = \inf\{R : R \text{ is achievable on } X\}$

Theorem: For any DMS X with pmf P_X, we have

$$R^*(X) = H(X).$$

The proof has two parts:

a) Achievability: $R^*(X) \leq H(X)$
   to show this we need to find codes with rate arbitrarily close to $H(X)$

b) Converse: $R^*(X) \geq H(X)$
   this shows optimality of the above codes

The proof techniques required for a) and b) are very different. We will treat them separately.

2.3.2 Proof the converse using Fano's inequality

If $X = Y$ then $H(X|Y) = 0$
If $P_r[X=Y]$ is large, then $H(X|Y)$ is small
↳ Fano's inequality

Lemma: (Fano's inequality) Let X and Y be rv with joint pmf P_{XY} and let $\varepsilon = P[X \neq Y]$. Then,

$$H(X|Y) \leq h(\varepsilon) + \varepsilon(\log|\mathcal{X}| - 1) \leq 1 + \varepsilon \log|\mathcal{X}|$$

$h(\varepsilon)$: binary entropy

$\varepsilon(\log|\mathcal{X}| - 1)$: vanishes as $\varepsilon \to 0$

Proof: $H(X|Y) = \sum P_Y(y)\, H(X|Y=y)$, thus, we first bound $H(X|Y=y)$

Define $\varepsilon_y = 1 - P_{X|Y}(y|y)$, then

$$\sum_y P_Y(y)\,\varepsilon_y = 1 - \sum_y P_Y(y)\,P_{X|Y}(y,y)$$
$$= 1 - \sum_y P_{XY}(y,y) = P[X \neq Y] = \varepsilon$$

$$H(X|Y=y) = -\sum_x P_{X|Y}(x|y)\log P_{X|Y}(x|y)$$
$$= -(1-\varepsilon_y)\log(1-\varepsilon_y) - \sum_{x\neq y} P_{X|Y}(x|y)\log P_{X|Y}(x|y)$$
$$= -(1-\varepsilon_y)\log(1-\varepsilon_y) - \varepsilon_y \sum_{x\neq y} \frac{P_{X|Y}(x|y)}{\varepsilon_y}\log\frac{P_{X|Y}(x|y)}{\varepsilon_y}$$
$$- \varepsilon_y \underbrace{\sum_{x\neq y}\frac{P_{X|Y}(x|y)}{\varepsilon_y}}_{=1}\log\varepsilon_y$$
$$= h(\varepsilon_y) - \varepsilon_y \underbrace{\sum_{x\neq y}\frac{P_{X|Y}(x|y)}{\varepsilon_y}\log\frac{P_{X|Y}(x|y)}{\varepsilon_y}}_{\leq \log(|\mathcal{X}|-1)}$$
$$\leq h(\varepsilon_y) + \varepsilon_y \log(|\mathcal{X}|-1)$$

$$\sum_{x\neq y} P_{X|Y}(x|y) = 1 - P_{X|Y}(y|y) = \varepsilon_y$$

$$\Rightarrow H(X|Y) \leq \sum_y P_Y(y)\,h(\varepsilon_y) + \varepsilon\log(|\mathcal{X}|-1)$$

Jensen's $\leq h(\varepsilon) + \varepsilon\log(|\mathcal{X}|-1)$

□

Proof of Converse:

Consider now any sequence of $(n, 2^{\lfloor nR \rfloor})$-codes with encoders e_n decoders d_n s.t. the probability of error $\varepsilon_n := P[\hat{X}^n \neq X^n]$ satisfies

$$\varepsilon_n \longrightarrow 0 \text{ as } n \to \infty.$$

By Fano's inequality, we must have

$$H(X^n | \hat{X}^n) \leq 1 + \varepsilon_n \cdot \log |\mathcal{X}|^n = 1 + \varepsilon_n n \log |\mathcal{X}|$$

$$H(X^n | M) \overset{DPI}{\leq} H(X^n | \hat{X}^n) \leq 1 + \varepsilon_n n \log |\mathcal{X}|$$

← note that $\hat{X}^n = d_n(M)$

Furthermore, since $|M| = 2^{\lfloor nR \rfloor} \leq 2^{nR}$ we have

$$H(M) \leq \log |M| \leq nR$$

$$\Rightarrow \quad nR \geq H(M)$$

$$= H(M) - H(M|X^n) + H(M|X^n)$$

because $M = e_n(X^n)$ is deterministic given X.

$$= I(M : X^n)$$

$$= H(X^n) - H(X^n | M)$$

$$\geq n H(X) - 1 - \varepsilon_n n \log |\mathcal{X}|$$

$$\Rightarrow \quad R \geq H(X) - \frac{1}{n} - \varepsilon_n \log |\mathcal{X}|$$

Since this has to hold for all n and $\varepsilon_n \to 0$ as $n \to \infty$, we must have $R \geq H(X)$

$\Rightarrow R^*(X) \geq H(X)$ □