Es Informazione / 06-05

1)

X, Y indip. $\quad H(X) = 11 \quad H(Y) = 2$

$$H(X,Y) = 11 + 2 = 13$$

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

mutua informaz. " perche' indip., Y non dice niente "

in assenza di rumore e' tutto l'informazione

$H(X|Y) = 0$ sempre se $X\ Y$ indip

2) $X = \{a, b, c, d, e, f\}$ $\quad p(a) = p(b) = p(c) = p(d) = \frac{1}{8}$

$p(e) = p(f) = \frac{1}{4}$

- $H(X)$?

$$H(X) = 4 \cdot \frac{1}{8} \cdot 3 + 2 \cdot \frac{1}{4} \cdot 2 = \frac{5}{2}$$

- L media di C (date le codifiche)

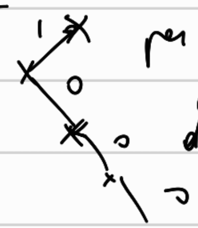
$$L(C) = \sum p \cdot L(C_i) = 4 \cdot \frac{1}{8} \cdot 3 + 2 \cdot \frac{1}{4} \cdot 2 = \frac{5}{2}$$

$\nexists$ codifica più efficiente perche' $L(C) = H(X)$ ho raggiunto il lim teorico

- Discuti decifrabilità e istantaneità

↳ non ambiguo (cioe' f iniettiva, no stesse codifiche per diversi input = seq)

la sola decifrabilità si può osservare che non ho albero

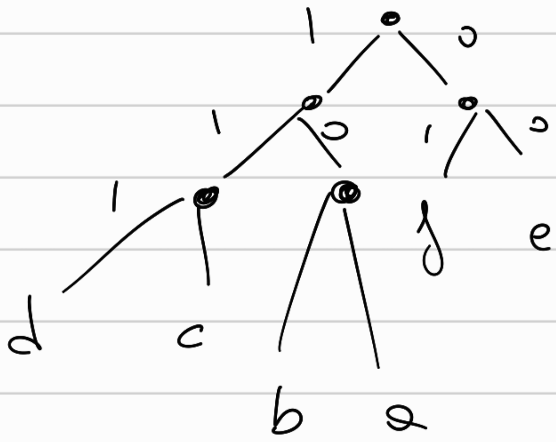


per C: 1, 10, 100, 1000 cod diverse

dove in x ho i simboli è dec. ma non istantanea

inoltre istantaneità $\longrightarrow$ decifrabilità

↳ trovo subito le foglie (vedi sotto)

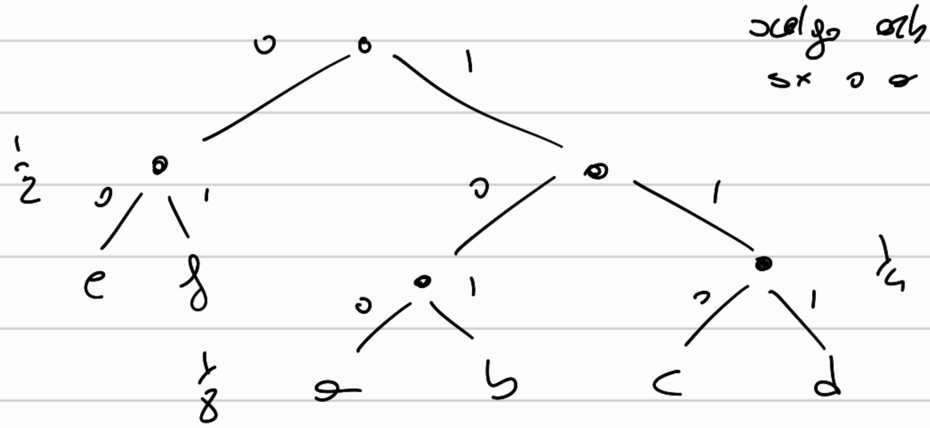


nessuna codifica è prefisso di un'altra → istantaneità

(lo si osserva anche dall'albero: ci assicura che non ho codifiche come prefissi)

3) Calcola codifica di Huffman (quella di sopra lo è, perché codifica ottimale)
(non è unica)

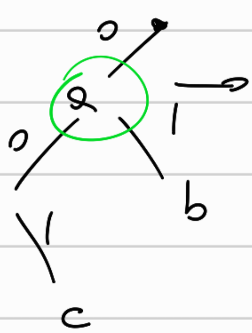
Parto dalla meno prob. (dalle foglie, verso l'alto)



scelgo arbitrariamente sx 0 o 1

4) $X : \{a, b, c\}$ $C(a) = 0$, $C(b) = 01$, $C(c) = 001$

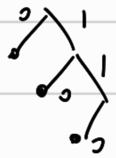
1) è univoc. dec?
2) è istantanea?
3) soddisfa Kraft McMillan
4) Se X sono i simboli: cosa si può dire su $H(X)$ e $L(C)$



non è istantanea perché posizionato su a attende ulteriori bit (ho un simbolo come nodo intermedio)

ma allora non è istantanea: se non dec non può essere istantanea (sarebbe dec) no univoc dec perché le seq. 001 può essere c ma anche ab

- $2^{-1} + 2^{-2} + 2^{-3} \le 1$ ✓ $\rightarrow$ quindi posso costruire una codifica ottimale dec. e istantanea



con 3 codifiche una lunga 1
" 2
" 3

- senza probabilità $H(X)$ non posso calcolarlo e poiché non sono univoc. dec. non posso affermare che $H(X) \le L(X)$

5) $X = \{x_1, x_2, x_3, x_4\}$ $\quad \mathcal{I}_{int} = \{l_1, l_2, l_3, l_4\}$

$l_1 = 1 \quad l_2 = l_3 = 2 \quad l_4 = 3$

Perché $\nexists$ cod istantanea che abbia come lunghezze di codifica gli interi sopra?

ma allora vale disug. K-McMillen

$$\underset{\text{attesa}}{L(C)} = \sum 2^{-l_i} = 2^{-1} + 2^{-2} + 2^{-2} + 2^{-3}$$

$$= \underbrace{\frac{1}{2} + \underbrace{\frac{1}{4} + \frac{1}{4}}}_{=1} + \frac{1}{8}$$

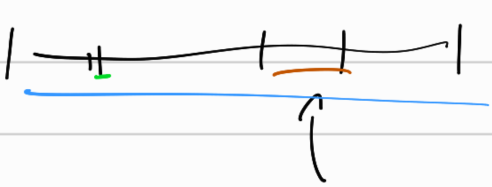
scoprendo che $1 + \frac{1}{8} \not\le 1$ superando il limite teorico posto da K-McMillen

6) Perché per sequenze più grandi corrispondono codifiche più brevi?

Questo perché la lunghezza della codifica è tale per cui se $x = \underbrace{x_1}_{\text{simboli più probabili}} x_2 \, x_3 \dots x_N$

$$L_N = p(i_1)\, p(i_2) \dots p(i_N)$$

Serve meno precisione per rappresentare un intervallo più grande



$L_c = 1/p$

n° cifre necessarie

pescare rispetto da codifica più piccola (intervallo più grande) perché ho più probabilità di cadere lì rispetto su dove p di ricadere è più piccola (L più grande)