Infokód. előadás

Forráskódolás

Emlékezetnélküli

IT forrás →

$x \to \{x_1, \dots, x_N\}$

$p(x) \to p(x_1), \dots, p(x_N)$

$c(x) \to \vdots$

$l(x) \to$

$$L(x) = \sum_x p(x)\, l(x)$$

$$C_{opt}: \min_C L(x)$$

1.) egyértelmű visszaállítás

prefix kód $\boxed{\sum_{i=1}^{N} 2^{-l(x_i)} \le 1}$

Kraft-egyenlőtlenség

2.) teoretikus korlát: $L(x) \ge H(x)$ biz. nem

3.) kódkonstrukciók nem

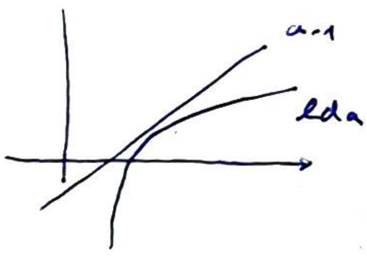
Kullback–Leibler távolság: $D(p\|q) = \sum_x p(x)\, \text{ld}\, \frac{p(x)}{q(x)}$

Tul: $D(p\|q) \ge 0$

azonos val. tér

Biz. $D(p\|q) = \sum_x p(x)\, \text{ld}\, \frac{p(x)}{q(x)}$

$a-1 \ge \text{ld}\, a$



$$-D(p\|q) = \sum_x p(x)\, \text{ld}\, \frac{p(x)}{q(x)} \le \sum_x p(x)\left(\frac{q(x)}{p(x)} - 1\right) =$$

$$= \underbrace{\sum_x q(x)}_{=1} - \underbrace{\sum_x p(x)}_{=1} = 0 \Rightarrow ✓$$

Entrópia

$$H(x) = \sum_x p(x)\, \text{ld}\, \frac{1}{p(x)}$$

Tul. $0 \le H(x) \le \text{ld}\, N$

Biz: $\sum_x \underbrace{p(x)}_{\ge 0}\, \text{ld}\, \underbrace{\frac{1}{p(x)}}_{\ge 1} \ge 0$

2. $D(p\|q) = \sum_x p(x)\,\mathrm{ld}\,\frac{p(x)}{1/N} = \sum_x p(x)\,\mathrm{ld}\,(p(x)\cdot N) = \sum_x p(x)\cdot \mathrm{ld}\,p(x) +$

$q(x) = \frac{1}{N}$

$+ \sum_x p(x)\,\mathrm{ld}\,N = -H(x) + \mathrm{ld}\,N \geq 0$ ✓

KöV: $p(x) = \frac{1}{N} \rightarrow H(x) = \mathrm{ld}\,N$

maximum itt van az entrópiának

Tétel: $L(x) \geq H(x)$ "forráskódolás alaptétele"

Bizonyítás

$p(x) \rightarrow$ valódi eloszlás

$q(x) \rightarrow$ virtuális eloszlás $\quad q(x) = \frac{2^{-l(x)}}{\sum_y 2^{-l(y)}}$

$0 \leq q(x) \leq 1$ ✓

$$\sum_x q(x) = \frac{\sum_x 2^{-l(x)}}{\sum_y 2^{-l(y)}} = 1$$

$$0 \leq D(p\|q) = \sum_x p(x)\,\mathrm{ld}\,\frac{p(x)}{q(x)} = \sum_x p(x)\,\mathrm{ld}\,\frac{p(x)}{\frac{2^{-l(x)}}{\sum_y 2^{-l(y)}}} =$$

$$= \sum_x p(x)\,\mathrm{ld}\,\frac{p(x)\sum_y 2^{-l(y)}}{2^{-l(x)}} \leq \sum_x p(x)\,\mathrm{ld}\,\frac{p(x)}{2^{-l(x)}} =$$

Kraft-e miatt

$$= \sum_x p(x)\,\mathrm{ld}(p(x)) - \sum_x p(x)\,\mathrm{ld}\,2^{-l(x)} = -\sum_x p(x)\cdot \mathrm{ld}\,\frac{1}{p(x)} + \sum_x p(x)\,l(x) =$$

$$= -H(x) + L(x)$$

$\Downarrow$

$\boxed{H(x) \leq L(x)}$

Shannon-Fano (SF) kód

Adott $x_1, \ldots, x_N \longrightarrow l_{SF}(x) = \left\lceil \mathrm{ld}\,\frac{1}{p(x)} \right\rceil$

$p(x_1), \ldots, p(x_N)$ $\qquad a \leq \lceil a \rceil \leq a+1$

Tul ① prefix kóda

$$\sum_{i=1}^{N} 2^{-l(x_i)} \le 1 \qquad \sum_{x} 2^{-\mathrm{ld}\frac{1}{p(x)}}$$

Biz.

$$\sum_{i=1}^{N} 2^{-\lceil \mathrm{ld}\frac{1}{p(x_i)}\rceil} \le \sum_{i=1}^{N} 2^{-\mathrm{ld}\frac{1}{p(x)}} = \sum_{i=1}^{N} p(x_i) = 1$$

Q.E.D.

érvényes fa kiterjesztésre

Huffman kód

Tétel: optimális kód: 1. $l_{N-1} = l_N$

$p_1 > \ldots > p_N$ 2. $\boxed{p_i > p_j} \to l_i < l_j$ *

Biz.: ellenhipotézis $l_i > l_j$

$$L(x) = \sum_{x \ne i,j} p(x)\, l(x) + p_i + p_i \cdot l_i + p_j \cdot l_j$$

$$L'(x) = \sum_{x \ne i,j} p(x)\, l(x) + p_i \cdot l_j + p_j \cdot l_i$$

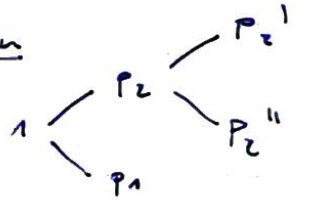
$$L'(x) \not> L(x)$$

$$\cancel{\sum_x p(x)\, l(x)} + p_i \cdot l_i + p_j \cdot l_j \not> \cancel{\sum_x p(x)\, l(x)} + p_i \cdot l_j + p_j \cdot l_i$$

$$p_i (l_j \cancel{- l_i}) \not> p_j (l_j \cancel{- l_i})$$

$p_i < p_j$ ↯ ellentmondás

mohó algoritmus



példa: (Huffman kód konstr.)

x_1	x_2	x_3	x_4	x_5
$p_1 = 0,25$	$p_2 = 0,15$	$p_3 = 0,2$	$p_4 = 0,25$	$p_5 = 0,15$

x_1'	x_2'	x_3'	x_4'	x_5'
$p_1 = 0,25$	$p_2 = 0,25$	$p_3 = 0,0,2$	$p_4 = 0,15$	$p_5 = 0,15$

$c_1 = (11)$

x_1' 0,25 — 0,25 — 0,25 $\xrightarrow{1}$ 0,55 $\xrightarrow{1}$ 1

$c_2 = (01)$

x_2' 0,25 — 0,25 $\xrightarrow{0}$ 0,45 $\xrightarrow{0}$ 0,45 $\xrightarrow{0}$

x_3' 0,2 — 0,2

$c_4 = (101)$ $c_3 = (000)$

x_4' 0,15 $\to$ 0,3 — 0,3

x_5' 0,15

$c_5 = (100)$

$L(x) = 0,25 \cdot 2 + 0,25 \cdot 2 + 0,2 \cdot 2 + 0,15 \cdot 3 + 0,15 \cdot 3 = 2,1$

$H(x) = \sum_x p(x) \operatorname{ld} \frac{1}{p(x)} = \dots$

$Hat = \frac{H(x)}{L(x)} = \dots$

Shannon - Fano - Elias (SFE) kód

(aritmetikai kód)

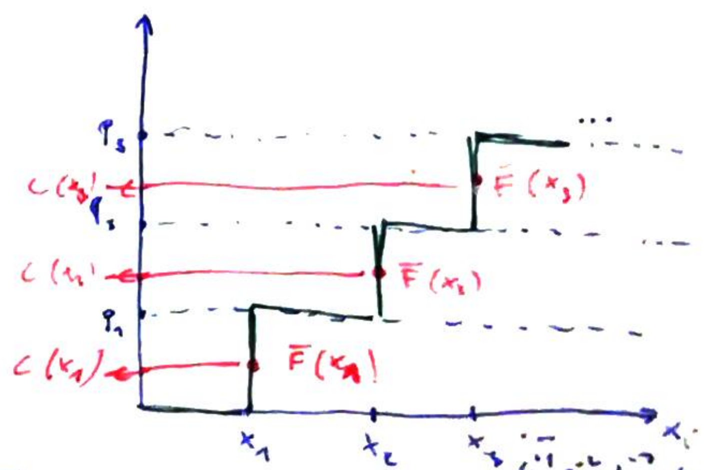
$0 < F(x) < 1$

$F(x) = \sum_{x_i < x} p(x_i)$

$\bar{F}(x) = \sum_{x_i < x} p(x) + \frac{1}{2} p(x)$

$\downarrow$

$C(x) = \left[\text{binconv}\left(\bar{F}(x)\right) \right]_{l(x)}$



ahol $l(x) = \lceil \operatorname{ld} \frac{1}{p(x)} \rceil + 1$ pl: 0,235... $\to$ 0, 0 0 1 1 ...

2^{-1} 2^{-2} 2^{-3} 2^{-4}

$l(x)$

$c = 0011$

$l(x) - 1 = \lceil \log \frac{1}{p(x)} \rceil$

$$\boxed{\frac{1}{2} p(x) \geq 2^{-l(x)}}$$

$0 < a < 1$

$$\sum_{i=1}^{\infty} a_i 2^{-i} - \sum_{i=1}^{l} a_i 2^{-i} \leq 2^{-l}$$

$$\bar{F}(x) - \lfloor \bar{F}(x) \rfloor_{l(x)} \leq 2^{-l} \leq \frac{1}{2} p(x) = \bar{F}(x_i) - \bar{F}(x_{i-1})$$