TINF - MASSOVNE

① H - entropija, količina inf. } [bit/simbol]
C - kapacitet } [bit/simbol]

entropija

→ na izlazu iz kanala

$H(Y|X)$ - šum

I - transinformacija, kol. borena inf.

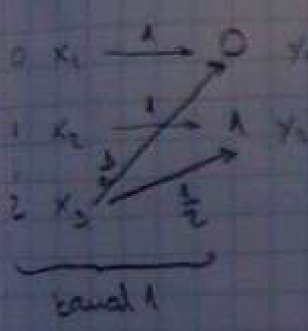
$C = \max\{I\}$ → kad ulaz, izlaz ovisni

$H(X,Y)$

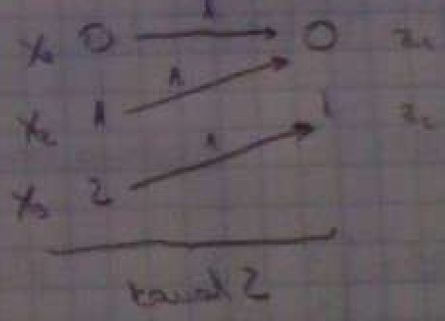
nezavisni, onda je H(X|Y) max a C=0 /

1.4 $X = \{0, 1, 2\}$

$Y = Z = \{0, 1\}$



(vjer. da 0 prede u 0)



a) $H(X) = ?$

$p(x=0) = 0.25$

$p(x=1) = 0.25$

$p(x=2) = 0.5$

$$H(X) = -\sum_{i=1}^{3} p(x_i) \log_2 [p(x_i)] =$$

$= 1.5$ bit/simbol

b) $H(Y) = -\sum p(y_i) \log_2 [p(y_i)]$

iz kanala → matrica jer nam treba $p(y_i)$

$$p(Y|X) = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \; \text{izlaz} \\ \begin{array}{c} 0 \\ \text{ulaz} \quad 1 \\ 2 \end{array} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{array}$$

da ovo dobijemo

$$p(X,Y) = p(x_i) \cdot p(Y|X)$$

→ jer $p(x)$

$$0.25 \cdot 1 \quad = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} \begin{array}{l} - p(x_1) \\ p(x_2) \\ p(x_3) \end{array}$$

$\quad p(y_1) \quad p(y_2)$

$$p(Y_1) = \frac{1}{4} + 0 + \frac{1}{4} = \frac{1}{2}$$

$$p(Y_2) = 0 + 2 \cdot \frac{1}{4} = \frac{1}{2}$$

$H(Y) = 1$ bit/simbol

c) $H(Z) = ?$ gledamo kanal 2 → tražimo $p(z|x)$

$$p(z) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad p(x,z) = \begin{bmatrix} \frac{1}{4} & 0 \\ \frac{1}{4} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{array}{l} - p(x_1) \\ p(x_2) \\ p(x_3) \end{array}$$

$\quad p(z_1) \quad p(z_2)$

$$p(z_1) = p(z_2) = \frac{1}{2}$$

$H(Z) = 1$ bit/simbol.

d) $H(Y,Z) = ?$ iz kanala vidimo da Y i Z ne ovise jedan o drugom $\Rightarrow H(Y,Z) = H(Y) + H(Z) = 2$ bit/simbol

e) $I(X;Y) = ?$ → gledamo matrice

* može se gledati i kao $I = H(Y) - H(Y|X)$ (korištenje dijagrama)

$$I(X;Y) = -\sum_{i=1}^{3}\sum_{j=1}^{3} p(x_i, y_j)\log_2\left[\frac{p(x_i, y_j)}{p(x_i)\,p(y_j)}\right] =$$

→ po matricama

$=$

$= 0.5$ bit/simbol

f) $I(X,Z) = ?$ → isto po matricama gledati

$= 1$ bit/simbol

17. X, Y su neovisni (X je ulaz, Y izlaz)

$$\begin{array}{l} \\ p(X,Y) = \\ \\ \\ \end{array}
\begin{array}{c}
\begin{bmatrix} p_1 & p_1 & \cdots & p_1 \\ p_2 & p_2 & \cdots & p_2 \\ \vdots & & & \\ p_n & p_n & \cdots & p_n \end{bmatrix} \\
p(y_1) \qquad\qquad p(y_m)
\end{array}
\begin{array}{c} p(x_1) \\ p(x_2) \\ \\ p(x_n) \\ \hline 1 \end{array}$$

$\left.\begin{array}{l} p(x_i),\ i=1 \ldots n \\ p(y_j),\ j=1 \ldots m \end{array}\right\}$ matrica n redova, m stupaca

a) $\underline{H(X) = ?}$ preko varijabli p_i, m

$p(x_1) = m\,p_1$ (jer imamo m stupaca)

$p(x_2) = m\,p_2$

$p(x_i) = m\,p_i$

$$H(X) = -\sum_{i=1}^{n} m \cdot p_i \log_2 [m \cdot p_i]$$

b) $\underline{H(Y)} = ?$

$p(y_i) = \frac{1}{m}$ (m stupaca, a zbroj svih $p(y)$ mora biti $= 1$

$$H(Y) = -\sum_{i=1}^{m} \frac{1}{m} \log_2\left[\frac{1}{m}\right] = -\frac{1}{m} \sum_{i=1}^{m} \log_2\left[\frac{1}{m}\right] = -\frac{1}{m} \cdot \underbrace{m}_{\text{jer zbrajamo } m \text{ puta}} \log_2\left[\frac{1}{m}\right] =$$

$$= \log_2[m]$$

c) $H(Y|X) = ?$

znamo da su neovisni → 

→ upravo zato što su neovisni,

dakle $H(Y|X) = H(Y)$

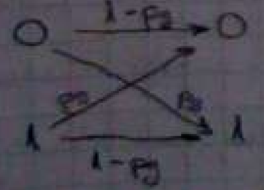
* tić ono zeza ☺

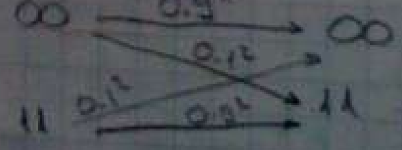
1.8 (prebitovano)

$x_1 = 00$ (simbol kojeg smo kodirali s 00)

$x_2 = 11$

$p_g = 0.1$ prenosimo x_1 i x_2 na izlaz, možemo dobiti:





01, 10 } još se i ovo može dobiti

$\underline{p(x_1) = p(x_2) = \frac{1}{2}}$

$$p(Y|X) = \begin{array}{c|cccc} & 00 & 11 & 01 & 10 \\ \hline 00 & 0.9^2 & 0.1^2 & 0.1 \cdot 0.9 & 0.09 \\ 11 & 0.1^2 & 0.9^2 & 0.1 \cdot 0.9 & 0.09 \end{array}$$

traži se iznos donesene inf: $I = ?$ treba nam matrica $p(X,Y)$

$$p(X,Y) = \begin{bmatrix} \text{prvi red matrice } p(Y|X) \text{ množimo s } p(x_1) \\ \text{drugi red } p(Y|X) \cdot p(x_2) \end{bmatrix}$$

I = po formuli = 0.762 bit/simbol   [inače nam treba $p(x)$, $p(x,y)$ pa sve možemo izračunati]

1.28 $C = ?$

$$p(Y|X) = \begin{array}{c} p_1 \\ p_2 \\ p_3 \end{array} \begin{bmatrix} \boxed{\frac{1}{2}} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \boxed{\frac{1}{2}} \end{bmatrix}$$

množimo p_1, p_2, p_3 s ovom matricom da dobijemo $p(x,y)$

1. povučemo pravac simetrično i uspoređujemo simetrične brojeve, ako su jednaki onda je kanal simetričan
2. ako ne bude iz prve dijagonale, probat drugu: $\begin{bmatrix} \diagdown & \bigcirc & \square \\ \bigcirc & \diagdown & \bigcirc \\ \square & \bigcirc & \diagdown \end{bmatrix}$

napomena: ako to gledamo

$[A] = [A^T]$ → to vrijedi samo za kvadratne,

a npr. $\begin{bmatrix} \boxed{\frac{1}{2}} & 0 \\ \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \\ 0 & \boxed{\frac{1}{2}} \end{bmatrix}$ neće točno ispasti tako

+ nešto o permutacijama u svakom redu i svakom stupcu ste brojke ($\frac{1}{2}$, $\frac{1}{3}$ i $\frac{1}{6}$) u našem primjeru

→ ako kanal simetričan onda vrijedi $p_1 = p_2 = p_3$

u našem slučaju $p_i = \frac{1}{3}$ → možemo onda naći $p(x,y)$

$$p(x,y) = \begin{bmatrix} \frac{1}{6} & \frac{1}{9} & \frac{1}{18} \\ \frac{1}{18} & \frac{1}{6} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{18} & \frac{1}{6} \end{bmatrix} \begin{array}{l} = \frac{1}{3} \\ = \frac{1}{3} \\ = \frac{1}{3} \end{array}$$

$$\begin{array}{ccc} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}$$

i sada $C = I = 0,126$ bit/simbol

1.74

$$x_1 \xrightarrow{\frac{1}{2}} y_1, \quad x_1 \xrightarrow{\frac{1}{2}} y_2, \quad x_2 \xrightarrow{1} y_3$$

$$p(Y|X) \; \begin{array}{c} p \\ 1-p \end{array} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

<u>C=?</u>

kanal ~~nije~~ simetričan:

ako kanal nesimetričan

* napomena: 99,99% da će biti zadana z ulaza čije vjerojatnosti označavamo s p i (1-p) da možemo derivirati

$\boxed{\frac{dI}{dp} = 0}$ da nađemo maximum što nam je onda = $\underline{C}$

$$I = \sum\sum p(x,y) \log_2 \left[\frac{p(x,y)}{p(x) \cdot p(y)}\right]$$

$$p(x,y) = \begin{array}{c} \begin{bmatrix} \frac{1}{2}p & \frac{1}{2}p & 0 \\ 0 & 0 & 1-p \end{bmatrix} \begin{array}{c} p \\ (1-p) \end{array} \\ \frac{1}{2}p \quad \frac{1}{2}p \quad 1-p \end{array}$$

$$I = 2 \cdot \frac{1}{2}p \, \log_2 \left[\frac{\frac{1}{2}p}{\frac{1}{2}p^2}\right] + (1-p) \log_2 \left[\frac{(1-p)}{(1-p)^2}\right] = 0$$

radi prva 2 člana u prvom redu

$$\left[\log_2(x)\right]' = \frac{1}{x} \cdot \frac{1}{\ln 2}$$

→ kad deriviramo i izj. s 0 onda ispada $p = \frac{1}{2}$, $p(x_1) = p(x_2) = \frac{1}{2}$

$C = 1$ bit/simbol

→ dakle ili kanal nesimetričan gdje deriviramo i dobivamo p ili simetričan pa znamo da su sve vjerojatnosti jednake ($p_1+p_2+p_3=1$, $p_1=p_2=p_3$)

1.29

X - ulaz $\qquad Z=\{0,1\}$ $\qquad X=\{1,2,\dots,n\}$

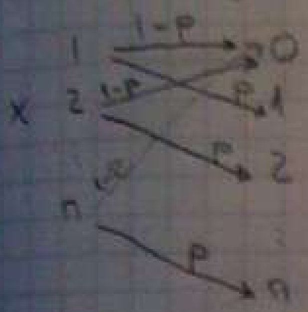
Y - izlaz $\qquad p(Z=0)=1-p$ $\qquad q_X=[q_1, q_2, \dots, q_n]$

Z - šum $\qquad p(Z=1)=p$

$Y=XZ$

$I(X,Y)=?$ → izlaz može biti $\{0,1,\dots,n\}$

→ radi Z-a



opet množimo s $p(x)$

$$p(Y|X)=\begin{array}{c|c} & \begin{array}{ccccc}0 & 1 & 2 & \dots & n\end{array} \\ \begin{array}{c}1\\2\\ \vdots\\ n\end{array} & \begin{bmatrix}1-p & p & 0 & & 0\\ 1-p & 0 & p & & 0\\ \vdots & & & & \\ 1-p & 0 & 0 & & p\end{bmatrix}\begin{array}{c}p(x_1)\\ \\ \\ p(x_n)\end{array} \\ & \begin{array}{ccc}p(x_1) & p(x_2) & p(x_n)\end{array}\end{array} \Rightarrow$$

$$p(x,y)=\begin{array}{c|c} & \begin{array}{ccccc}0 & 1 & 2 & & n\end{array} \\ \begin{array}{c}1\\2\\ \vdots\\ n\end{array} & \begin{bmatrix}(1-p)q_1 & pq_1 & 0 & & 0\\ (1-p)q_2 & 0 & pq_2 & & \\ & & & & \\ (1-p)q_n & & & & pq_n\end{bmatrix}\begin{array}{c}q_1\\ q_2\\ \\ q_n\end{array} = p(x_i) \\ & \begin{array}{ccccc}1-p & pq_1 & pq_2 & \dots & pq_n\end{array} = p(y_j)\end{array}$$

n puta broj

$$I=(1-p)q_1 \underbrace{\log_2\left[\frac{\overbrace{(1-p)q_1}^{=1}}{(1-p)q_1}\right]}_{0} +$$

→ i tako za cijeli prvi stupac, $\log(1)=0$, cijeli izraz $=0$

$$+\,p q_1 \log_2\left[\frac{p q_1}{p q_1^2}\right] + \dots + p q_n \log_2\left[\frac{p q_n}{p q_n^2}\right] =$$

$$=\sum_{i=1}^{n} p\, q_i \log_2\left[\frac{1}{q_i}\right] = p \sum_{i=1}^{n} q_i \log_2\left[\frac{1}{q_i}\right]$$

* u rješenju u zb se vidi da je $q_1=q_2=\dots=q_n$ ali nam to nigdje zadali.

② Entropijsko kodiranje

$\boxed{\text{sËum}} \rightarrow 0101$ (kodiramo sekvenca, najčešće baza = 2)

$L \rightarrow$ srednja dulj. kodne riječi , $L = \sum n_i l_i$

$H(X) \leq L < H(X) + 1$

$\varepsilon = \frac{H}{L} \rightarrow$ efikasnost

mi radimo o prefiksnim kodovima koji zadovoljavaju Kraftovu nejednakost.

$$\boxed{\sum_{i=1}^{n} \underbrace{d}_{\text{baza}}^{-l_i} \leq 1}$$

npr.: $x_1 = 01$
$x_2 = 10$
$x_3 = 11$
$x_4 = 00$

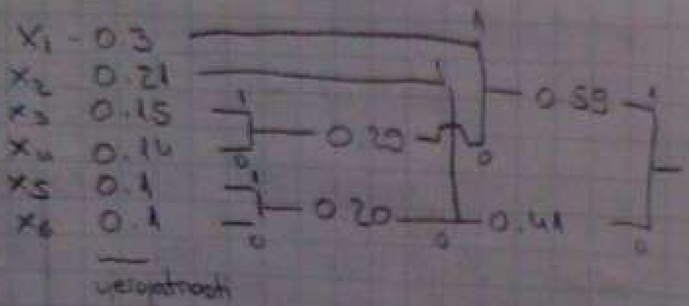
$\sum = 1 \rightarrow$ kod prefiksan

( ) HUFFMANOVO KODIRANJE

Z.9

$d = 3$

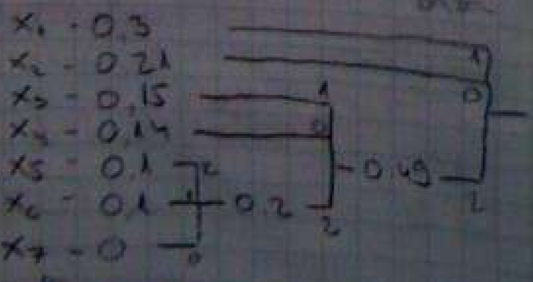
za $b = 2$



x_1 - 11
x_2 - 01
x_3 - 101
x_4 - 100
x_5 - 001
x_6 - 000

1. poredamo vjer. od najveće prema najmanjoj
2. zbrajamo po dvije najmanje, kodiramo ih od manje prema većoj s 0, 1
3. samo očitamo zdesna brojeve i tako su nam kodirani simboli

za $b \neq 2$, gledamo $k = \left\lceil \frac{n-1}{b-1} \right\rceil = \frac{5}{2} = 3$ (n – broj simbola, b – baza), $N' = (b-1) \cdot k + 1 = 7$ (broj simbola za kodiranje)



$x_1 = 1$
$x_2 = 0$
$x_3 = 21$
$x_4 = 20$
$x_5 = 222$
$x_6 = 221$

$L = \sum n_i l_i = 1.69$ tribit/simbol

$H = 2.47$ bit/simbol

$\varepsilon = \frac{H}{L}$ NE!!

$L_2 = L_3 \log_2 3 = 2.67$ bit/simbol

$\varepsilon = \frac{H}{L} = \frac{2.47}{2.67} = 0.925$

b) $H(\infty)$

...012010...

$p(0), p(1), p(2) = ?$ → da bismo izračunali entropiju [illegible] poruke

$p'(0)$ gledamo koliko 0 imamo gdje i koliko se puta pojavljuju u tom simbolu i množimo s vjer.

$p'(0) = (1 \cdot \overset{p(x_1)}{0.21} + 1 \cdot \overset{p(x_4)}{0.14}) \cdot k = 0.35k$

$p'(1) = 0.55k$ [jedinica nam se pojavljuje u x_1, x_3, x_6]

$p'(2) = (1 \cdot \overset{x_3}{0.15} + 1 \cdot \overset{x_4}{0.14} + 3 \cdot \overset{x_5}{0.1} + 2 \cdot \overset{x_6}{0.1})\, k = 0.79k$

$\Sigma = 1 \Rightarrow 0.35k + 0.55k + 0.79k = 1$

$\Rightarrow k = 0.5917,$

$p(0) = p'(0) = 0.35 \cdot 0.5917 = 0.207$

$p(1) = 0.325$

$p(2) = 0.468$

$H(\infty) = -\sum p_i \underline{\underline{\log_2}} [p_i] = 1.91$ bit/simbol

NE napisati $\log_3$!! jer smo se maknuli od tribita! pretvorili

2.17

0, 1

$p(0) = 2\,p(1)$

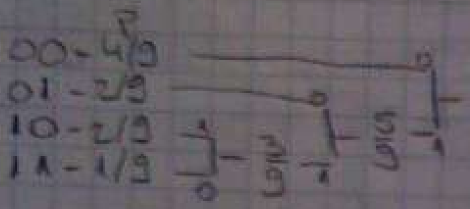
znamo da $p(0)+p(1)=1$ dakle

$p(0) = \frac{2}{3}$

$p(1) = \frac{1}{3}$

uvjet u zad: $L \leq 0,94$ bit/simbol

x_1 0
x_2 1 } onda nam je $L=1$ → ne valja → grupiramo simbole



00 kodiramo s 1 bitom
01 s 2 bita
10 s 3 bita
11 s 3 bita

gledamo $\frac{4}{9}\cdot 1 + \frac{2}{9}\cdot 2 + \frac{2}{9}\cdot 3 + \frac{1}{9}\cdot 3 = 1,888$, ali ovo nije $= L$

→ imamo parove dakle dijelimo s 2 → $\frac{1,888}{2} = 0,944$

još uvijek preveće

→ moramo raditi trojke

Uvod

140 $=N$ simbola na ulazu jednake vjer. $p(x_i) = \frac{1}{140}$, a koristimo kod Huffmana

$\frac{E=?}{L=?}$

$H = -\sum p(x_i) \log_2 [p(x_i)]$

X simbola se kodira s k bitova

Y s (k+1)

$N = X + Y = Z$

$$\boxed{\begin{aligned} &X = 2^{k+1} - N \quad , k \in Z \, , \, X \geq 0 \\ &Y = Z - X \end{aligned}}$$

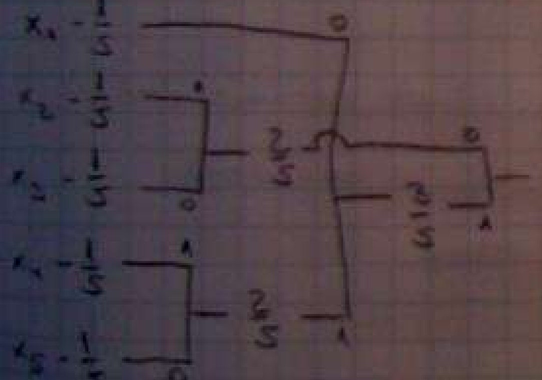
→ formula ako se trebaju zadatak [illegible] papiru :)

I biramo X → $X = 2^{k+1} - 140$ tako da dobijemo najmanji k

ako $k=7$ → $X = 116$ → kodiramo ih s 7 bitova

$Y = 24$ smo kodirali s 8 bitova

npr $N=5$ [ovo je bilo samo da vidimo da formula „radi"]



→ s 2 bita smo kodirali x_1

2

2

3

3

$X = 2^{k+1} - N$, ako $k = 2$

$Y = 2$

$X = 3$

LZ77 - šablona, kao i aritmetičko

LZW

poruka: 0 1 0 4 5

a b a aa aaa

2 - ab
3 - ba

4 - aa
5 - aaa

rječnik: 0 - a
1 - b

pr: abaaaa:

a imamo
ab nemamo => ab
ba - 3
4 - aa
5 - aaa

abbaaa
0 1 0 4 0

ovo znači da ga do kodirati s 2

→ gledamo po rječniku što imamo, a što fali, i popunjavamo rječnik

→ došli do problema, do 4 → to nam može biti samo „aa" jer smo ab već iskodirali

→ ono što „uvijek" znamo jest da je 0-a a 1-b, kako se krećemo po poruci uzimamo u obzir i unazad da se poruka ne bi krivo protumačila te kodiramo po redu sve mogućnosti

→ u zadatku smo znali da je 0=a, 1=b, odmah moramo iskodirati „ab", zatim imamo 0, koju znamo da je „a", pa smo dobili novu comb. ba, sve po redu da bude pravilje.

poruka nam je 01045

1. 0 = a, 1 = b } 01045 ab
2. 2 = ab → 01045 a ba
3. 3 = ba → 01045 aa → gledamo što nam 4 može značiti, a prije njega imamo a → ili aa ili ab - već imamo
4. 4 = aa → 01045 aa → prije petice imamo „aa", stoga 5 može biti samo aaa