Resum Primer Parcial (5 primers temes)

Informació i Seguretat (Universitat Autònoma de Barcelona)

RESUM 1r PARCIAL – INFORMACIÓ I SEGURETAT

TEMA 2: CONCEPTES BÀSICS DE LA TEORIA DE LA INFORMACIÓ

QUANTITAT INFORMACIÓ = QUANTITAT INCERTESA

Mesura de Hartley (incertesa sobre n resultats possibles i equiprob.)

$$I(n) = \log(n)$$

⇒ PROBLEMA: no es tenen en compte les prob. de cada resultat.

Mesura de Shannon

$$I(A) = \log \frac{1}{p(A)} = -\log p(A)$$

⇒ Incertesa/informació d'un esdeveniment A amb probab. $p(A)$.

Entropia, informació d'una font

$$H(X) = \sum (p(a_i) \cdot I(a_i)) = \sum \left(p(a_i) \cdot \log \cdot \frac{1}{p(a_i)}\right)$$

↳ Unitats de l'entropia: bits/resultat o bé bits/missatge

NOTA: $\log_2$ → perquè volem els resultats en bits.

$$H(X) = -\sum (p_i \cdot \log p_i)$$

Quantitat d'informació mitjana que té una font d'informació.

PROPIETATS:

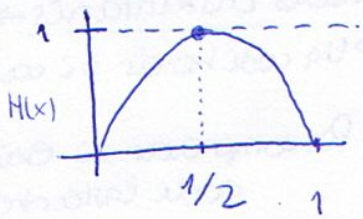
- $H(X) \geq 0$
- $H(X) = 0$ si hi ha una $p_i = 1$.
- $H(X) \leq \log n$
- $H(X) = \log n$ si $p_i = \frac{1}{n}$ $\forall i$

Entropia binària

Si tenim 2 esdeveniments: $S = \{a_1, a_2\}$ $\begin{pmatrix} P(a_1) = p \\ P(a_2) = 1-p \end{pmatrix}$

$$H(X) = p \cdot \log \frac{1}{p} + (1-p) \cdot \log \frac{1}{(1-p)}$$

↳ Com a màxim serà 1 bit.



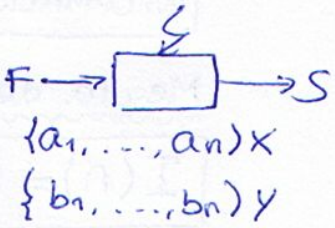
↳ si són equiprobab.

Entropia conjunta

$$H(X,Y) = -\sum\sum p(x_i, y_i) \cdot \log p(x_i, y_i)$$

Entropia condicionada

$$H(X|Y) = \sum\sum p(x_i, y_i) \cdot \log p(x_i|y_i)$$



Informació mútua

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

Capacitat del canal

$$C = \max\{I(X,Y)\} = \max\{H(X) - H(Y|X)\}$$

TEMA 3: CODIFICACIÓ DE LA FONT

$S = \{a_1, \ldots, a_n\}$ CONJUNT MISSATGES

$A = \{\sigma_1, \ldots, \sigma_n\}$ ALFABET D-ARI DEL CODI

Exemple:
$A = \{0, 1\}$
ALFABET 2-ARI O BINARI

- CODIS DE LONGITUD FIXA

Exemples: ASCII, UNICODE...

$$L \geq \frac{\log n}{\log D} \qquad D^L \geq n$$

L → longitud
D → símbols
n → paraules-codi

Si codifiquem seq. de símbols: $D^L \geq n^m$ $\quad L \geq \frac{m \cdot \log n}{\log D}$

- CODIS DE LONGITUD VARIABLE

Exemples: MORSE, , Zip.

Només treballem amb codis de descodificació única.

- $\bar{L} = \sum p_i L_i$ Longitud mitjana.
- Codis instantanis → tots són de descodif. única
  → Un codi ho és si cap paraula-codi és prefix d'una altra.
  → Per comprovar si existeix un codi instantani ⇒ KRAFT $\sum D^{-l_i} \leq 1$

- 1r TEOREMA SHANNON $\boxed{\bar{L} \geq \frac{H(S)}{\log D}}$ Mínim valor de $\bar{L}$ que podrem obtenir.

- Eficiència d'un codi D-ari amb longitud mitjana $\bar{L}$ per a un conjunt de missatges S: $\boxed{\eta = \frac{H(S)}{\bar{L} \cdot \log D}}$ (≤ 1)

- Codi <u>òptim</u>. → Un codi és òptim si per el mateixa conjunt de missatge només existeixen altres codis amb longitud mitjana <u>igual o superior</u>.
  - ↳ Si $\eta = 1$ → òptim
  - ↳ Si comparem codis D-aris pel mateix S → comparem les $\bar{L}$.
    Si són codis amb diferent alfabet → comparem eficiències.
- Construcció codis òptims → MÈTODE DE HUFFMAN

TEMA 4 : COMPRESSIÓ DE DADES

TAXA COMPRESSIÓ $\boxed{R = \frac{|C|}{|M|}}$ → Fitxer comprimit / → Fitxer original

% compressió $\boxed{(1-R)\cdot 100\%}$

BITRATE $\boxed{BR = \frac{|C| \text{ bits}}{|M| \text{ símbols}}}$

MÈTODES DE COMPRESSIÓ:

- Per taxa de compressió:
  - Sense pèrdua (lossless).
  - Amb pèrdua (lossy). P.e. MP3, JPEG...
- Per model:
  - Estàtics (no adaptatius).
  - Dinàmics (adaptatius).

TÈCNIQUES DE COMPRESSIÓ

- Basats en repetició. RLE (els dibuixos amb 0s i 1s.
- Mètodes estadístics. Huffman, aritmètic
- Basats en diccionari. LZ, LZW
- Basats en transformades. DCT, wavelets, JPEG

COMPRESSIÓ D'IMATGES

- Mètodes més usuals.
  - Sense Compressió. BMP, RAW, PGM, PPM.
  - Sense pèrdua. PNG, GIF, TIFF...
  - JPEG.
  - JPEG2000.
  - MPEG.

MAJOR QUALITAT
MSE ↓
PSNR ↑

- Mesures de distorsió. ($\hat{I}$ imat. orig. / I imat comp.)
  - Màxim error.
    $d(I,\hat{I}) = \max_{ij} |I_{ij} - \hat{I}_{ij}|$
  - MSE.
    $MSE = \frac{1}{N_x} \cdot \frac{1}{N_y} \sum\sum (I_{ij} - \hat{I}_{ij})^2$
  - PSNR.
    $PSNR = 10 \log_{10} \frac{(\max_{ij} I_{ij})^2}{MSE}$

TEMA 5: CODIFICACIÓ DEL CANAL

TIPUS DE CANALS

Canal sense pèrdua
- $H(B|A) = 0 \Rightarrow I(A,B) = H(A)$
- $n \nleq m$
- cada columna de Π té un únic element $\neq 0$.
- $C = \max_{P(i)} H(A) = \log n$

Canal determinista
- $H(B|A) = 0 \Rightarrow I(A,B) = H(B)$
- $m \leq n$
- cada fila de Π hi ha un únic elem. $\neq 0$
- $C = \max_{P(i)} H(B) = \log m$

Canal sense soroll
- Si és sense pèrdua i determinista, $I(A,B) = H(A) = H(B)$
- A cada columna hi ha un 1 i la resta 0's.
- $n = m$
- $C = \log n = \log m$

Canal totalment simètric
- les files de Π són iguals excepte que canvia l'ordre, a les columnes, igual.
- $C = \log m - \underline{\underline{H}}$
  On $H = H(B|A)$

Canal amb entrada i sortida independents
- $H(A|B) = H(A)$; $H(B|A) = H(B)$
  ↳ $I(A,B) = 0$
- No serveix per transmetre informació.
- les files de Π són idèntiques.
- $C = 0$