CS 229r - IT ∈ CS

LECTURE 4

2/7/19

TODAY

- COMPRESSION
  - SINGLE SHOT: SHANNON + HUFFMAN
  - UNIVERSAL: (Start Lempel-Ziv).
  - MARKOV SOURCES + HIDDEN MARKOV MODELS

UP TO NOW: Asymptotic Compression

Compress n i.i.d. variable $X_1 \ldots X_n$

SINGLE-SHOT:

- Sender + Reciever know P_X ; design E, D
- Sender gets $X \sim P_X$
- Must send $E(X) \in \{0,1\}^*$ to receiver
- Receiver decodes $\hat{X} = D(E(X))$.
- Want $\hat{X} = X$ & minimize $\mathbb{E}_{X \sim P_X} [|E(X)|]$

Two Solutions

1. ~~Se~~ [Shannon] (Note this is in contrast to his solution for asymptotic compression)

~~Sender~~. let $l_i \triangleq \left\lceil \log \frac{1}{p_i} \right\rceil$ $\quad \Omega = \{1 \ldots m\}$

- Design (by Kraft sufficiency), prefix-free $E : \Omega \to \{0,1\}^*$

  s.t. $E(i) \in \{0,1\}^{l_i}$

- send $E(X)$.

Performance: $\underset{X \sim P_X}{\mathbb{E}} \left[|E(X)| \right] = \underset{X \sim P_X}{\mathbb{E}} \left[\left\lceil \log \frac{1}{P_X(X)} \right\rceil \right]$

$$\leq \underset{X \sim P_X}{\mathbb{E}} \left[\log \frac{1}{P(X)} + 1 \right]$$

$$= 1 + \underset{X \sim P_X}{\mathbb{E}} \left[\log \frac{1}{P(X)} \right]$$

$$= 1 + H(X)$$

Optimal to within 1 bit!

[Note: follows from Shannon's asymptotic theorem that no prefix-free scheme achieves $< H(X)$.

HUFFMAN

Recursive algorithm $(P_1 \ldots P_n)$

- Sort $P_1 \geq P_2 \geq \ldots P_n$
- Combine P_{n-1} & P_n to form new element & get $Q_1, Q_2, \ldots Q_{n-1}$ with $Q_i = P_i$ except $Q_{n-1} = P_{n-1} + P_n$
- $E' \leftarrow$ Huffman-Code $(Q_1 \ldots Q_n)$;
- $E(i) = E'(i)$ except $E(n-1) = 0 \cdot E'(n-1)$ & $E(n) = 1 \cdot E'(n-1)$

Theorem: Huffman is optimal $\forall P$.

Proof:

Ingredients ①: if $P_1 \geq P_2 \geq \ldots P_n$ then $l_1 \leq l_2 \leq \ldots l_n$

Proof: ~~Sw~~ Else swap l_i, l_j & get cost that is no more ☒

② $l_{n-1} = l_n$: Proof look at tree. ~~Child~~ Leaf at depth l_n must have sibling;

~~3~~ (2.5) W.l.o.g P_{n-1} & P_n are siblings.

(3) "Optimal Substructure":

$$q_1 = P_1 \;\ldots\; q_{n-2} = P_{n-2} \quad q_{n-1} = P_{n-1} + P_n$$

Optimal tree for $P_1 \ldots P_n$ with siblings P_{n-1} & P_n

is also optimal tree for $q_1 \ldots q_{n-1}$ with parent(P_n) as leaf.

(4) Induction $\boxtimes$

UNIVERSAL CODING: How do gzip, zip, ... work?

- Single file to compress; No distribution?; Produces something smaller!

- Central Algorithm: Lempel-Ziv + variations
  - works well empirically
  - has a "Theorem"!

Algorithm: Input $w \in \Sigma^{*n}$ $\quad \Sigma$ = finite alphabet

- Break w into $S_0 \circ S_1 \circ S_2 \circ S_3 \ldots \circ S_m$
  - where $\circ$ denotes concatenation of strings
  - S_0 = empty string; $S_i = S_{j_i} \circ b_i$ where $j_i < i$ & $b_i \in \Sigma$;
  - Compression of w = Encoding of $j_1, j_2 \ldots j_m$ & $b_1 \ldots b_m$

in any reasonable prefix-free

De"compressing": $(j_1, \dots j_m)$
$b_1 \dots b_m$

- Can compute for $i = 1$ to m

$$S_i = S_{j_i} \circ b_i$$

- output $S_1 \circ S_2 \circ \dots S_m$

How well does this perform?

- Case 1: $W_1 \dots W_n$ i.i.d. over P distribution over Σ.

Claim: Expected length of compression $= (1 + o(1)) \cdot H(P) n$

(Proof later)

- More Impressive

Case 2: $W_1 \dots W_n$ drawn from "Hidden Markov Model"

$\Rightarrow$ Expected length of Compression $= (1 + o(1)) \cdot$ "$H(M)$" $\cdot n$

entropy of chain

Need to Define: ① "Hidden Markov Model"

& ② Entropy $H(M)$ of Hidden Markov Model.

Markov Chain

- $Z_1, Z_2, \dots, Z_n, \dots$ form a (time-invariant) Markov Chain

  if $\Pr[Z_i \mid Z_1 \dots Z_{i-1} = \alpha_1 \dots \alpha_{i-1}] = \Pr[Z_i \mid Z_{i-1} = \alpha_{i-1}]$

  $= \Pr[Z_j \mid Z_{j-1} = \alpha_{i-1}]$

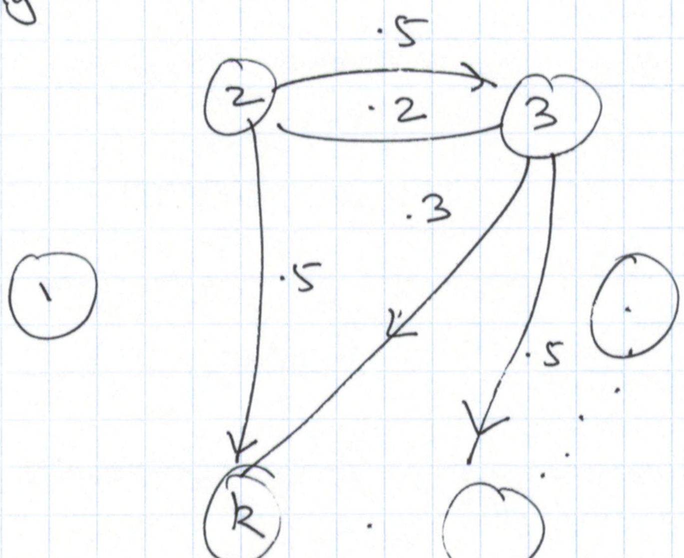
- Conditioned on last "state" Z_{i-1}, distribution of Z_i doesn't depend on past.

- if $Z_i \in \Omega = \{1, \dots k\}$

  - Markov chain specified by matrix: $M \in \mathbb{R}^{k \times k}$

    $M_{ij} = \Pr[Z_2 = i \mid Z_1 = j]$.

  - Pictorially



etc.

Entropy (Rate) of Markov Chain : M

$$= \lim_{n\to\infty} H(Z_n \mid Z_{n-1}, \dots Z_1)$$

$$= \lim_{n\to\infty} H(Z_n \mid Z_{n-1})$$

$$= H(Z_2 \mid Z_1).$$

[Corr. to # bits needed to describe Z_n given $Z_1 \dots Z_{n-1}$]

Simplifying setting : • M is irreducible (path from every state to every other state)

& a periodic (gcd (cycle lengths) = 1).

• ~~Assume~~ if so MC has a (unique) stationary distribution $\pi = \pi(M)$ s.t. if $Z_i \sim \pi$ then $Z_{i+1} \sim \pi$.

• Assume $Z_1 \sim \pi$.

• Theorem [Shannon] : Can compute Entropy rate of Markov Chain $H(M)$ ~~given~~ given M.

Hidden Markov Models

Specified by (1) Markov Chain M on k states $= \Gamma$

(2) Distributions $\{P_\gamma\}_{\gamma \in \Gamma}$ supported on Ω

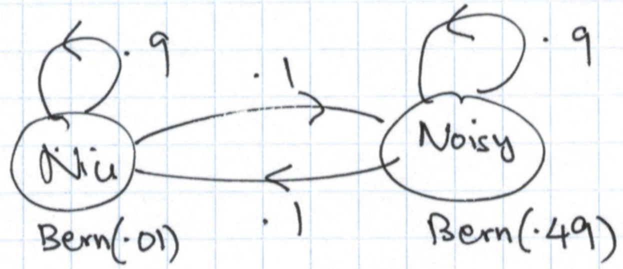
Generates sequence $X_1 \dots X_n$ as follows

(1) Pick $Z_1 \sim \pi$ (stationary prob. of M)

(2) Generate $Z_i \mid Z_{i-1}$ according to M

(3) Generate $X_i \sim P_{Z_i}$

Example:



- Entropy of HMM M

$$H(M) = \lim_{n \to \infty} \{ H(X_n \mid X_{n-1} \cdots X_1) \}$$

- Limit exists! $\forall n \quad H(X_n \mid X_{n-1} \cdots X_1) \leq H(X_{n-1} \mid X_{n-2} \cdots X_1)$

NEXT LECTURE: Analysis of L-2 on HMM.