Today

- Quality of Huffman Codes
- Universal Coding
- Lempel Ziv Algorithm

Admin

- PS2 due tomorrow
- PS1 will be handed back Thursday

Review

\[ C : \Omega_{\{1, \ldots, n\}} \rightarrow D^*_{\text{often} D = \{0, 1\}} \]

- Kraft’s Inequality: \( l_i = |(C_i)| \), then \( \sum_{i=1}^{n} D^{-l_i} \leq 1 \) if code is uniquely decodable.
- if \( p_i \) is prob. of element \( i \), we would like to minimize \( E[l] = \sum_{i=1}^{n} p_i l_i \).
- Entropy inequality: \( \frac{H(p_1, \ldots, p_n)}{\log D} \leq E[L] \)

1. Kraft’s inequality is tight if \( l_1, \ldots, l_n \) satisfy \( \sum_{i=1}^{n} D^{-l_i} \leq 1 \) then \( \exists C : \{i, \ldots, n\} \rightarrow D^* \) s.t. \( |C(i)| = l_i \).

2. Shannon Coding Method
   - \( l_i = \lceil \log D \frac{1}{p_i} \rceil \leq \log D \frac{1}{p_i} + 1 \Rightarrow E[L] \leq \frac{H(X)}{\log D} + 1 \)
   - should use to compress \( \bar{X} = (X_1, \ldots, X_k) \) where \( k \rightarrow \infty, X_1, \ldots, X_k \) i.i.d. \( \sim X \)
   - (from here, \( D = 2 \)).

   - \( kH(X) = H(\bar{X}) \leq E[\text{length compressing } \bar{X}] \leq H(\bar{X}) + 1 = kH(X) + 1. \)
     \( \rightarrow \) loss becomes \( \frac{1}{k} \) per element.

Huffman Coding

- "optimal" prefix code for variable \( X \)
- \( C_{\text{Huffman}} : \{i, \ldots, n\} \rightarrow \{0, 1\}^* \)
- Huffman code \((p_1, \ldots, p_n)\)
  - if \( n \leq 2, \ldots \)
  - sort so that \( p_1 \geq p_2 \geq \cdots \geq p_n \)
- \( C' \leftarrow \text{Huffman Code}(p_1, p_2, \ldots, p_{n-2}, p_{n-1} + p_n) \)

- \[
C[i] = \begin{cases} 
C'[i] & \text{if } i \leq n-2, \\
C'[n-1]0 & \text{if } i = n-1, \\
C'[n-1]1 & \text{if } i = n.
\end{cases}
\]

Today

Claim: For any prefix-free code \( C : \{1, \ldots, n\} \rightarrow \{0, 1\}^* \) it is the case that \( \sum_{i=1}^{n} p_i|C(i)| \geq \sum_{i=1}^{n} p_i|C_{\text{Huff}}(i)|. \)

Prefix free:
- All codewords are leaves.
- in optimal tree, can always assume \( p_i < p_j \Rightarrow l_i \leq l_j \)
- in optimal tree, no nodes have only one child
- \( \exists 2 \) leaves at lowest level with the same parent and with the two lowest probabilities.
- \( E[\text{length}(p_1, \ldots, p_n)] \geq E[\text{length}(p_1, \ldots, p_{n-2}, p_{n-1} + p_n)] + (p_{n-1} + p_n)1 \)

\( X_1, X_2, \ldots, X_t, X_i \text{ i.i.d.} \sim X \) then compressing with Huffman/Shannon is more realistic.

Markovian Source (Hidden Markov Chain or Ergodic Source)
- Finite State Space \( \{1, \ldots, n\} \)
- Transition prob. matrix \( \{p_{ij}\}_{i,j=1,\ldots,n} \)
- \( (i, j) \rightarrow b_{ij} \in \{0, 1\} \)
- Build for English, but what happens if source switches to French?

Universal Coding

Goal: compress information produced by a Markovian Source
- must be efficient
- has no prior knowledge of source

Consider \( X \in \{1, \ldots, n\}, p(X = i) = p_i, X_1, \ldots, X_t \text{ i.i.d.} \sim X. \)
Compress \( (\bar{X} = (X_1, \ldots, X_t)) \)
- let \( t_i \) be the number of occurrences of \( i \) in \( \bar{X} \)
- send \( (t_1, \ldots, t_n) \)
- which of \( \binom{t}{t_1, \ldots, t_n} \) possible sequences was seen
- amount of communication \( \rightsquigarrow \) negligible +\( tH(X) \)
AEP for Ergodic Markovian Source

if \((X_1, \ldots, X_L)\) elements drawn from finite Markovian (ergodic) source then

\[
\frac{-\log \Pr(p(X_1, \ldots, X_L))}{L} \rightarrow H(X)
\]

entropy rate of process.

With probability \(1 - \delta\),
\[
2^{-H(X)L(1+\epsilon)} \leq p(X_1, \ldots, X_L) \leq 2^{-H(X)L(1-\epsilon)}
\]

Divide \(t\)-length sequences into blocks of length \(L\).

Compression idea \((L, k)\)

- \(X_1, \ldots, X_t \sim Y_1, \ldots, Y_{t'}, Y_i \in \{0, 1\}^L, t' = \frac{t}{L}\)

- (1) typical set: \(w \in \{0, 1\}^L\) s.t. \(w\) appears at least \(k\) times, send \(w \leq 2^L\) bits

- (2) for each block:
  - “0” (typical) and index into set of elements sent in step 1 \(\approx H(X)(L+1)t/L\) bits
  - “1” (nontypical) and \(w \in \{0, 1\}^L \approx \delta(L+1)t/L\) bits

- as \(t \rightarrow \infty\),
\[
2^L + H(X)(L+1)\frac{t}{L} + \delta(L+1)\frac{t}{L} \approx H(X)t.
\]