LUND UNIVERSITY

Information Theory

Lecture 4
Huffman coding

STEFAN HÖST

Optimal codeword length

Theorem

Given a random variable X, the average codeword length for an optimal D-ary prefix code satisfies

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

with equality to the left if and only if $\ell_x = \log_D p(x)$.

Huffman code

Algorithm for binary Huffman code ($D = 2$)

To construct the code tree:

1. Sort the symbols according to their probabilities.
2. Let x_i and x_j, with probabilities p_i and p_j, respectively, be the two least probable symbols
   - Remove x_i and x_j from the list and connect them in a binary tree.
   - Add the root node $\{x_i, x_j\}$ as one symbol with prabaility $p_{ij} = p_i + p_j$ to the list.
3. If one symbol in the list

   STOP

   Else

   GOTO 2

Optimal codes

Three observations for an optimal prefix code

- In an optimal code codewords with high proabability should be shorter than codewords with low probability, i.e.

$$p_i < p_j \quad \Rightarrow \quad \ell_i \geq \ell_j$$

- In the tree for an optimal code there are no unused leaves.
- The two least probable codewords are of equal length, and can be constructed such that they differ only in the last bit.

Optimal codes construction

Construction of an optimal code

- Consider a code with k codewords.
- Assume x_k and x_{k-1}, with probablities p_k and p_{k-1}, are the two least probable codewords, and that their codewords differ only in the last bit.
- Define a reduced code with $k-1$ codewords where x_k and x_{k-1} are replaced by their parent node $\tilde{x}_{k-1}$, with probability $\tilde{p}_{k-1} = p_k + p_{k-1}$.

The first code is optimal if the reduced code is optimal.

Iterate until two codewords. Then an optimal code is given by $y_1 = 0$ and $y_2 = 1$.

Optimality of Huffman codes

Theorem

A binary Huffman code is an optimal prefix code.

That is, if $\mathcal{C}$ is a binary Huffman code, and $\mathcal{C}'$ is any other prefix code for the same set, then

$$L(\mathcal{C}) \leq L(\mathcal{C}')$$

Huffman code of vectors

Let $\boldsymbol{X} = (X_1, \dots, X_N)$ be a vector of N i.i.d. random variables, X_i. Encode the vectors with a Huffman code,

$$\boldsymbol{x} = (x_1, \dots, x_N) \xrightarrow{\text{Huffman}} \boldsymbol{y} = (y_1, \dots, y_{\ell_{\boldsymbol{x}}^{(N)}})$$

The average codeword length is

$$H(X_1, \dots, X_N) \leq L^{(N)} < H(X_1, \dots, X_N) + 1$$

The average codeword length per symbol, $L = \frac{1}{N} L^{(N)}$,

$$H(X) \leq L < H(X) + \frac{1}{N}$$

In the limit $N \to \infty$, the average codeword length is $L = H(X)$

LUND UNIVERSITY

Huffman code of vectors

Example of codeword length

A vector of N i.i.d. symbols is encoded using a Huffman code. For each symbol $p(0) = \frac{1}{4}$ and $p(1) = \frac{3}{4}$. The average codeword length per symbol is