Data Compression

Nathanael Bosch (nathanael.bosch@gmail.com) 28. Mai 2015

In the following X will be a random variable with range $\mathcal{X}$.
Without loss of generality, we can assume that the $|A|$-ary alphabet is
$A = \{0, 1, \ldots, |A| - 1\}$.

Definition:

- $A^+ := \bigcup_{n \in \mathbf{N}} A^{\times n}$ is the set of finite non-empty words in A.
- $A^* := A^+ \cup \emptyset$ is the set of all words in A.
- $C : \mathcal{X} \to A^+$ is called a symbol code for X. $C(\mathcal{X})$ is the set of codewords, $l_x \in \mathbf{N}$ the respective length and $L(C) := \sum_{x \in \mathcal{X}} p(x) l_x$ the average length.
- A symbol code C is called
  - "non-singular" iff C is injective
  - "uniquely decodable" iff the map $\mathcal{X}^+ \to A^+$ defined by $x_1 \ldots x_n \mapsto C(x_1) \ldots C(x_n)$ is injective
  - "prefix-free" (or a "prefix code") iff there is no pair $x \neq x'$ so that $C(x) = C(x')a$ for some $a \in A$

Theorem (Kraft inequality)**.** *For any prefix-free code $C : \mathcal{X} \to A^+$ the codewprd lengths $l_1, l_2, \ldots, l_{|\mathcal{X}|}$ must satisfy the inequality*

$$\sum_i |A|^{-l_i} \leq 1$$

Conversely, given a set of codeword lengths that satisfy this inequality, there exists a prefix-free code with these word lengths.

Theorem. *Let $\mathcal{X}$ be countable infinite. Every decodable code satisfies Kraft's inequality.*
Conversely, if Kraft's inequality holds for some $l \in \mathbf{N}^{\mathcal{X}}$, then there exists a prefix-free code with these codeword lengths $(l_x)_{x \in \mathcal{X}}$.

Definition: A symbol code $C : \mathcal{X} \to A^+$ is called "optimal" w.r.t. a random variable X iff the average codeword length $L(C)$ is minimal among all symbol codes in $(A^+)^{\mathcal{X}}$.

Hauptseminar Entropie und Informationstheorie, SS2015, TU München

Theorem. *For any instantaneous $|A|$-ary code for a random variable X holds:*

$$L \geq H_{|A|}(X),$$

with equality if and only if $|A|^{-l_i} = p_i$.

Theorem. *Let $l_1^*, l_2^*, \ldots, l_m^*$ be optimal codeword lengths for a source distribution p and an alphabet A, and let L^* be the associated expected length of an optimal code ($L^* = \sum p_i l_i^*$). Then*

$$H_{|A|}(X) \leq L^* < H_{|A|}(X) + 1.$$

Theorem. *The minimum expected codeword length per symbol satisfies*

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq L_n^* < \frac{H(X_1, X_2, \ldots, X_n)}{n} + \frac{1}{n}.$$

Moreover, if $X_1, X_2, \ldots, X_n$is a stationary stochastic process,

$$L_n^* \to H(\mathcal{X}),$$

where $H(\mathcal{X})$ is the entropy rate of the process.

Construction (Huffman code)**:** Recursive construction of a "Huffmann code" $C : \mathcal{X} \to A^+$ by building an $|A|$-ary tree from its leaves:

- step 0) assign each symbol from $\mathcal{X}$ to a leaf,
- step 1) assign the $|A|$ least probable symbols to leaves with a common vertex,
- step 2) 'combine' these symbols to a single one whose probability equals the sum of the $|A|$ probabilities.

Then iterate 1)→2)→1)→... until there is only one symbol left (the root).

Theorem. *Huffman coding is optimal; that is, if C^* is a Huffman code and C' is any other uniquely decodable code, $L(C^*) \leq L(C')$.*

References

[1] Thomas M. Cover and Joy A. Thomas, *Elements of information theory*, Wiley-Interscience; second edition (2006).

[2] Michael Wolf: `http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MA5103_2012W/lecture4.pdf`, (2012).

[3] Michael Wolf: `http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MA5103_2012W/lecture5.pdf`, (2012).