A. Overview

(1) a. Entropy
   b. Huffman encoding
   c. N-grams
   d. Cross-entropy
   e. Software

B. Entropy

(2) For comparison purposes, assume that all information is encoded with binary numbers: \textit{bits}.

(3) \(\{a, b, c, d\} = \{00, 01, 10, 11\}\)

(4) Entropy
   The number of digits needed in a binary encoding to represent all possible items.

C. Huffman encoding

(5) Why not represent \(\{a, b, c, d\}\) as \(\{0, 1, 10, 11\}\)?

(6) How is 1011 parsed?: \(1 - 0 - 1 - 1\) or \(10 - 1 - 1\) or \(1 - 0 - 11\) or \(10 - 11\)

(7) There are other ways of creating a parsable binary encoding for four items, e.g. \(\{1, 00, 010, 011\}\).

(8) A Huffman tree:
   a. Make a binary tree from the two lowest-frequency elements.
   b. Their parent node is assigned the sum of their probabilities.
   c. Keep doing this until all elements are in the tree.
   d. Assign 0 and 1 to each pair of nodes.
   e. Read the code off the tree.
(9) \[
\begin{array}{c}
\text{1} \\
.5 & .5 \\
.2 & .3 \\
.2 & .1
\end{array}
\Rightarrow
\begin{array}{c}
\text{0} \\
10 & 11 \\
10 & 110 & 111
\end{array}
\]

(10) \[
\begin{array}{c}
\text{1} \\
.5 & .5 \\
.25 & .25 & .25 & .25
\end{array}
\Rightarrow
\begin{array}{c}
\text{0} \\
10 & 11 \\
00 & 01 & 10 & 11
\end{array}
\]

D. Formulas

(11) Entropy: \( H(p) = - \sum p(x) \log p(x) \)

(12) Per-word entropy: \( H_{\text{rate}} = - \frac{1}{n} \sum p(x_{1n}) \log p(x_{1n}) \)

(13) Shannon-McMillan-Breiman theorem: \( H(L) = \lim_{n \to \infty} \frac{1}{n} \log p(w_1 w_2 \ldots w_n) \)

(14) Perplexity: \( 2^{H(L)} \)

E. Entropy & n-grams

(15) Unigram entropy of four texts:

<table>
<thead>
<tr>
<th>Text</th>
<th>Entropy</th>
<th>Perplexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>a b a b a b a b a b b</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>a b c d e a b c d e</td>
<td>2.32</td>
<td>5</td>
</tr>
<tr>
<td>a b a b a a a a a a a</td>
<td>0.72</td>
<td>1.65</td>
</tr>
<tr>
<td>a a a a a b b b b b</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

(16) Bigram entropy of the first and fourth texts:

<table>
<thead>
<tr>
<th>Text</th>
<th>Entropy</th>
<th>Perplexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>a b a b a b a b a b b</td>
<td>0.128</td>
<td>1.093</td>
</tr>
<tr>
<td>a a a a a b b b b b</td>
<td>0.489</td>
<td>1.404</td>
</tr>
</tbody>
</table>

F. Cross entropy

(17) Cross entropy:
\[ H(p, q) = - \sum p(x) \log q(x) \]

(18) By the Shannon-Breiman-McMillan theorem:
\[ H(p, q) = \lim_{n \to \infty} \frac{1}{n} \log q(w_1 w_2 \ldots w_n) \]
(19) Cross-entropy using the four texts in (15):

<table>
<thead>
<tr>
<th>Text</th>
<th>1</th>
<th>∞</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (a b a b a b a b a b)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 (a b c d e a b c d e)</td>
<td>2.32</td>
<td>2.32</td>
<td>2.32</td>
<td>2.32</td>
</tr>
<tr>
<td>3 (a b a b a a a a a a)</td>
<td>1.32</td>
<td></td>
<td>0.72</td>
<td>1.32</td>
</tr>
<tr>
<td>4 (a a a a a b b b b b)</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

(20) \(H(p) \leq H(p, q)\)

G. Software

(21) All software is invoked on the command-line as follows: perl program-name
    If arguments are required, a suitable error message is displayed.

(22) a. entropy.pl: computes the per-word entropy of a text using unigrams.
    b. entropy2.pl: calculates the per-word entropy of a text using bigrams

References
