Computational Semantics for Natural Language Processing

Alexis Tabin
16-821-803

Assignment 1

Hierarchical Softmax & Huffmann Coding

D, document $\qquad$ V, vocabulary $\qquad$ $p(w) = \frac{\# w}{|D|}$, word frequency

$|D|$, # of words $\qquad$ $|V|$, voc size

$\# w$: # of times w appears in D

$$P_w(c) = \frac{\exp\{(s(w,c))\}}{\sum_{c \in V} \exp\{(s(w,c))\}}, \quad O(|V|) \text{ for each pair } (w,c)$$

approximated by

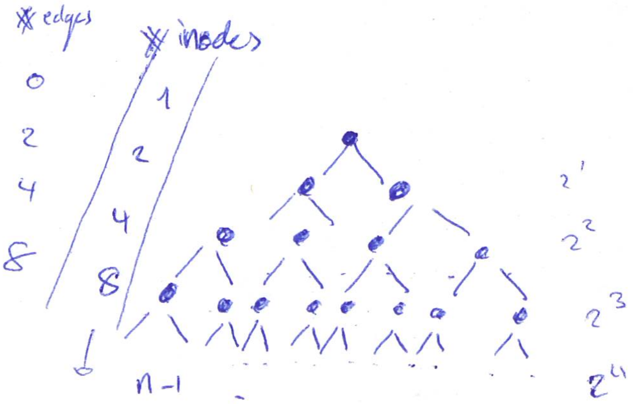
$$P_w(c) = \prod_{j=1}^{L(c)-1} P_w((c,j) \to (c,j+1))$$

$L(c)$: depth of node c in tree ξ

(c,j): node of depth j on road to leaf c

$P_w((c,j) \to (c,j+1))$: prob of transition

3.1 binary tree, $|V| = 2^{10} = 1024$



To compute $P_w(c)$, we need to go all the way down to the leaf, the way is composed of 9 edges here.

General case:

$$O(\log|V|\xi - 1) = O(\log|V|)$$

nbr nodes: $\sum_{i=0}^{n-1} 2^i$, here $n = 10$

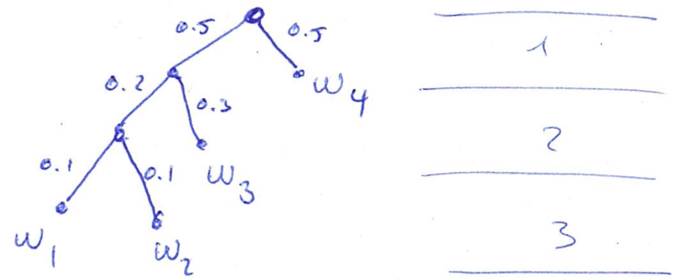
$\Rightarrow \boxed{\#\text{ inner nodes} = 1023}$

Total # nodes = inner nodes + leaf nodes
$= 1023 + 1024 = 2047$

# of edge $= 2046$

3.2 | $|V| = 4$

Intuitively, we have :



$$\mathbb{E}[L(c)] = 3 \cdot 0.1 + 3 \cdot 0.1 + 2 \cdot 0.3 + 1 \cdot 0.5 = 1{,}7$$

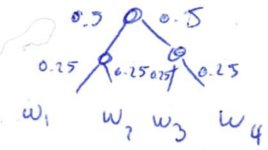
• $\cancel{\mathbb{E}[L(c)] \geq \mathbb{E}[-\log(p(w_i))]}$

$$L(w_i) \geq -\log(p(w_i))$$

no clue

• worst case : let take the last example.

the worst case is this one :



: $\mathbb{E}[L(c)] = 4 \cdot (0.25 \cdot 2)$

$= 2 = \log_2 4$

general case : $\mathbb{E}[L(c)] = O(\log|V|)$