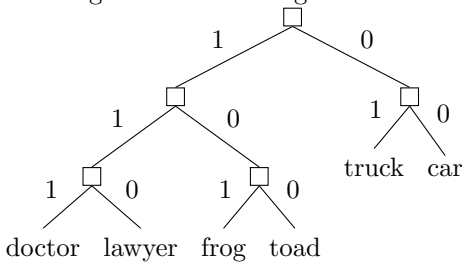
1 Word Cluster Representations

1.1 Question (time: 8:35, slide: 3)

Say we are given the following word cluster hierarchy



Which words are in the "10" cluster?

(a) doctor

(b) frog

(c) truck

(d) toad

2 The Brown Clustering Algorithm (Part 1)

2.1 Question (time: 11:50, slide: 7)

Say we are given a deterministic class function C with

$C(\text{dog}) = 1 \quad C(\text{man}) = 2 \quad C(\text{woman}) = 2 \quad C(\text{walk}) = 3$

Which of the following are possible definitions of e for a model with this class function?

(a) $e(\text{dog} \mid 1) = 1$, $e(\text{man} \mid 2) = 0.5$, $e(\text{woman} \mid 2) = 0.5$, $e(\text{walk} \mid 3) = 1$

(b) $e(\text{dog} \mid 1) = 1$, $e(\text{man} \mid 2) = 1$, $e(\text{woman} \mid 3) = 0.5$, $e(\text{walk} \mid 3) = 0.5$

(c) $e(\text{dog} \mid 1) = 1$, $e(\text{man} \mid 2) = 0.5$, $e(\text{woman} \mid 2) = 0.5$, $e(\text{man} \mid 3) = 0.5$, $e(\text{walk} \mid 3) = 0.5$

(d) $e(\text{dog} \mid 1) = 1$, $e(\text{man} \mid 2) = 0.9$, $e(\text{woman} \mid 2) = 0.1$, $e(\text{walk} \mid 3) = 1$

3 The Brown Clustering Algorithm (Part 2)

3.1 Question (time: 8:30, slide: 9)

Say we have estimated a clustering with two classes ($k = 2$) and that we have counts n such that

$n(1) = 30$ $\quad$ $n(2) = 10$

$n(1, 1) = 25$ $\quad$ $n(2, 2) = 5$ $\quad$ $n(1, 2) = 5$ $\quad$ $n(2, 1) = 5$

We now want to compute the quality of our clustering. What is the value of the expression $\sum_{c=1}^{k} \sum_{c'=1}^{k} p(c, c') \log \frac{p(c, c')}{p(c)p(c')}$?

A Answers

- (b) (d)

  Starting from the top of the tree we traverse the left edge (1) and then the right edge (0). The two words underneath this node are "frog" and "toad".

- (a) (d)

  Since the class function is deterministic, words should have zero probability of being emitted from a different class, e.g. $e(\text{man} \mid 3)$ must be 0.

- -0.341

  The answer is -0.341. First note that $p(1) = 3/4$, $p(2) = 1/4$, $p(1,1) = 0.625$, and $p(1,2) = p(2,1) = p(2,2) = 0.125$. So the sum is $0.625 \log \frac{0.625}{9/16} + 2 \times 0.125 \log \frac{0.125}{3/16} + 0.125 \log \frac{0.125}{1/16} = -0.341$.