CSC401/2511 Assignment 3 Q&A

Charlie Chenyu Zhang

Agenda

- Due day reminder & Common questions on Piazza
- Levenshtein distance (examples in docstring)
- Q&A

Agenda

- Due day reminder & Common questions on Piazza
- Levenshtein distance (examples in docstring)
- Q&A

A3 Due: Friday, 5 April 2019 at 19h00, electronically

Q: underflow/small numbers, divided by zero?

A: from scipy.special import logsumexp

Q: Do we have to use preComputedForM?

A: No, you don't have to.

$$\log b_m(\vec{x_t}) = -\sum_{n=1}^{d}\left(\frac{1}{2}\vec{x_t}[n]^2\vec{\sigma_m}^{-2}[n] - \vec{\mu_m}[n]\vec{x_t}[n]\vec{\sigma_m}^{-2}[n]\right) - \left(\sum_{n=1}^{d}\frac{\vec{\mu_m}[n]^2}{2\vec{\sigma_m}^2[n]} + \frac{d}{2}\log 2\pi + \frac{1}{2}\log\prod_{n=1}^{d}\vec{\sigma_m}^2[n]\right)$$

Q: Can I add to the header of this function log_p_m_x using log_Bs and/or preComputedForM?

A: Please don’t modify the function headers in the starter code. You are welcome to define your own helper functions for an efficient implementation.

Agenda

- Due day reminder & Common questions on Piazza
- Levenshtein distance (examples in docstring)
- Q&A

Levenshtein Distance Pseudocode

Modification

Input: REF: reference array of words
Input: HYP: hypothesis array of words
begin
  $n \leftarrow$ The number of words in REF
  $m \leftarrow$ The number of words in HYP
  $R \leftarrow zeros(n+1, m+1)$ `// Matrix of distances`
  $B \leftarrow zeros(n+1, m+1)$ `// Backtracking matrix`
  For all i, j s.t. $i = 0$ or $j = 0$, ~~set $R[i,j] \leftarrow \infty$, except $R[0,0] \leftarrow 0$~~
  for $i = 1..n$ **do**
    for $j = 1..m$ **do**
      $del \leftarrow R[i-1, j] + 1$
      $sub \leftarrow R[i-1, j-1] + (REF[i] == HYP[j])?0 : 1$
      $ins \leftarrow R[i, j-1] + 1$
      $R[i, j] \leftarrow$ `Min` (del, sub, ins)
      if $R[i, j] == del$ **then**
        $B[i, j] \leftarrow$ 'up'
      end
      else if $R[i, j] == ins$ **then**
        $B[i, j] \leftarrow$ 'left'
      end
      else
        $B[i, j] \leftarrow$ 'up-left'
      end
    end
  end
  Return ~~100~~ $R[n, m]/n$
end

Set R[i,j] <- 0, 1, 2, ...

Algorithm 1: Computation of Levenshtein distance, with backtracking.

Levenshtein Distance Example

REF ← ["who", "is", "there"]

HYP ← ["is", "there"]

n ← 3

m ← 2

R ←

		is	there
	0	1	2
who	1		
is	2		
there	3		

B ←

		is	there
	0	left	left
who	up		
is	up		
there	up		

Levenshtein Distance Example

R

REF \ HYP		is	there
	0	1	2
who	1	1	2
is	2	1	2
there	3	2	1

tie, choose either

B

REF \ HYP		is	there
		left	left
who	up	up-left	left
is	up	up-left*	left
there	up	up	up-left*

\# up ≡ # deletion ①

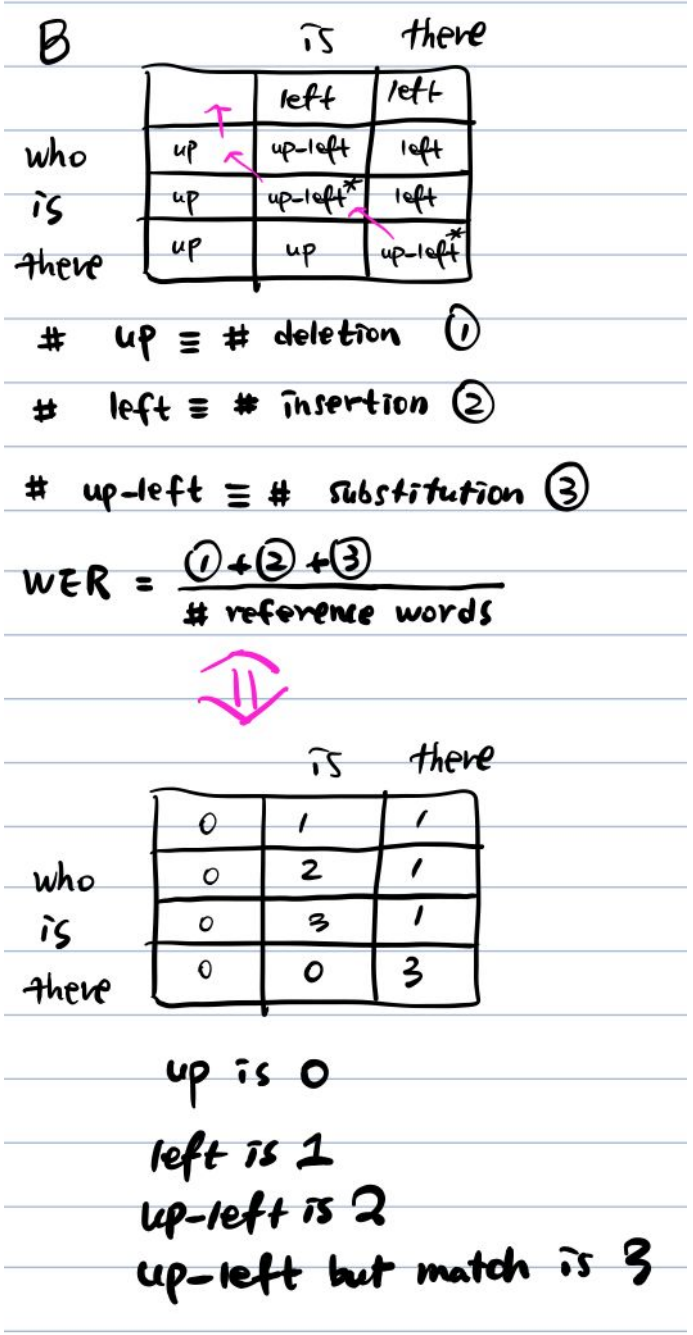
\# left ≡ # insertion ②

\# up-left ≡ # substitution ③

$$WER = \frac{① + ② + ③}{\#\ \text{reference words}}$$

Levenshtein Distance Example

You can either store strings in B, or you can store integers in B



Agenda

- Due day reminder & Common questions on Piazza
- Levenshtein distance (examples in docstring)
- Q&A

CSC401/2511 Assignment 3 Q&A