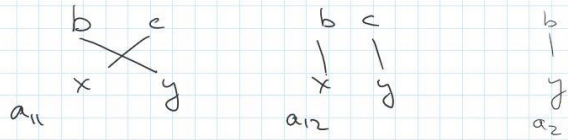


Lecture 5

- Continuing from last lecture (IBM Model 1)
 - o Hidden variable a (alignment, a one-to-one function)
 - o $p(f|e, a) = \prod_{j=1}^n p(f_j|e_j)$
 - o $p(a|f, e, \theta) = \frac{p(a, f|e)}{\sum_{a'} p(a', f|e)} = \frac{p(a|e)p(f|a, e)}{\sum_{a'} p(a'|e)p(f|a', e)} = \frac{\prod p(f_j|e_{a_j})}{\sum p(f_j|e_{a_j})}$
 - Assume $p(a|e)$ uniform
 - o Example
 - e_1 : "b c", e_2 : "b"
 - f_1 : "x y", f_2 : "y"
 - Initialize $p(x|b) = p(y|b) = p(x|c) = p(y|c) = \frac{1}{2}$

All alignments:



E-step:

$$p(a_{11}|e_1, f_1) = \frac{\frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}} = \frac{1}{2}$$

$$p(a_{12}|e_1, f_1) = \frac{1}{2}$$

$$p(a_{21}|e_2, f_2) = 1$$

M-step:

$$\hat{p}(y|b) = \frac{\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1}{\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1} = \frac{3/2}{2} = 3/4$$

$$\hat{p}(x|b) = 1/4$$

$$\hat{p}(y|c) = \frac{1/2}{1/2 + 1/2} = 1/2$$

$$\hat{p}(x|c) = 1/2$$

E-step:

$$p(a_{11}|e_1, f_1) = \frac{3/4 \cdot 1/2}{3/4 \cdot 1/2 + 1/4 \cdot 1/2} = 3/4$$

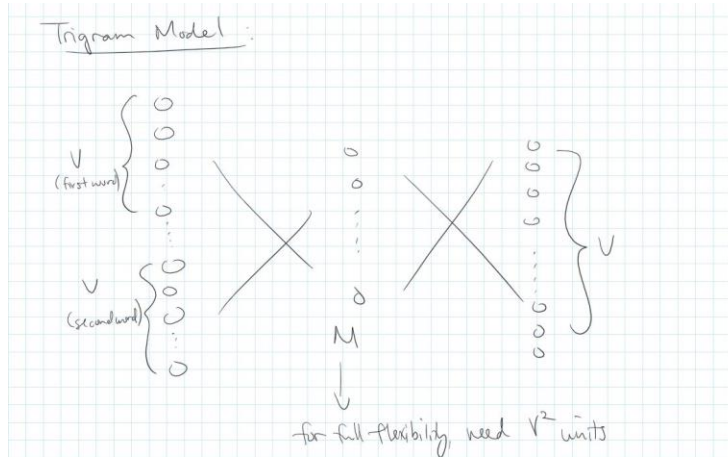
$$p(a_{12}|e_1, f_1) = 1/4$$

$$p(a_{21}|e_2, f_2) = 1$$

$$\text{M-step: } \hat{p}(x|c) = \frac{3/4}{3/4 + 1/4} = 3/4$$

$$\hat{p}(y|c) = 1/4$$

- Bigram Model
-
- The diagram illustrates a Bigram Model architecture. It consists of three layers: an input layer, a hidden layer, and an output layer. The input layer is labeled 'input (as two next tokens)' and has V input units. The hidden layer is labeled 'hidden layer(s)' and has M units. The output layer is labeled 'output' and has V output units. Connections are shown between the input and hidden layers, and between the hidden and output layers, with large 'X' marks indicating the flow of information. The input layer is connected to the hidden layer, and the hidden layer is connected to the output layer.
- For full flexibility, need V units in the hidden layer
(there are V^2 connections from input to hidden)
- So how can we create embeddings?
- ↳ dense rather than sparse vector to represent words
- ① For each word, look at its neighbors to the hidden layer
 - ② " " " " " " " from " " "



Handwritten diagram illustrating a relationship between two sets of vectors, v and h .

- On the left, a brace labeled v groups a vertical list of zeros: $\begin{Bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{Bmatrix}$.
- To the right of this, there is a large 'X' mark.
- Below the 'X', there is a brace labeled h grouping another vertical list of zeros: $\begin{Bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{Bmatrix}$.
- To the right of the h brace, there is another large 'X' mark.
- An arrow points from the 'X' mark associated with the h vectors towards the 'X' mark associated with the v vectors.