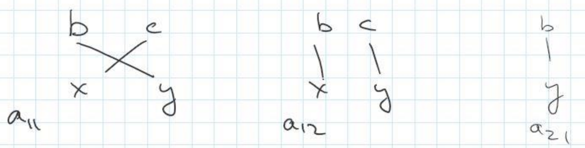
Lecture 5

- Continuing from last lecture (IBM Model 1)
  - Hidden variable a (alignment, a one-to-one function)
  - $p(f|e,a) = \prod_{j=1}^{n} p(f_j|e_j)$
  - $p(a|f,e,\theta) = \frac{p(a,f|e)}{\sum_{a'} p(a',f|e)} = \frac{p(a|e)p(f|a,e)}{\sum_{a'} p(a'|e)p(f|a',e)} = \frac{\prod p(f_j|e_{a_j})}{\sum p(f_j|e_{a_j})}$
    - Assume $p(a|e)$ uniform
  - 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:

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

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

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

M-step:

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

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

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

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

E-step:

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

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

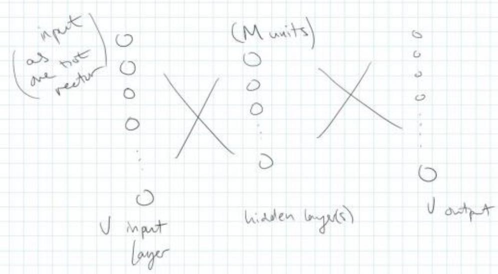
$$p(a_{21}, e_2\ f_2) = 1$$

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

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

- Neural Models (We will be studying embeddings)
    - Review of neural models – refer to 6.036 lecture notes on Stellar
    - Huge impact on computer vision…not so much in NLP. Why?
        - Neural models do not really take into count the structure of language
        - Compared to previous models, neural models also need much more data
- Neural Language Models
    - Start with bigram model $p(w_i|w_{i-1})$

Bigram Model:

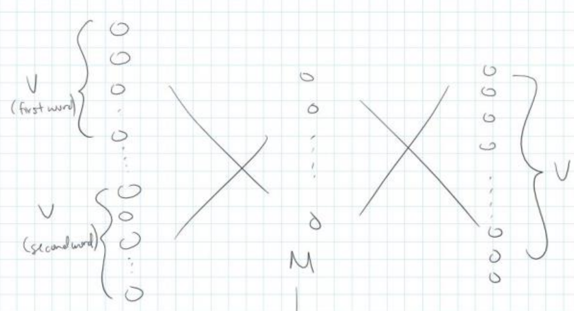


For full flexibility, need V units in the hidden layer (there are V^2 connections from input to hidden)

So how can we create embeddings?
↳ dense rather than sparse vector to represent words

1. For each word, look at its weights to the hidden layer
2. " " " " " " " from " " "

Trigram Model:



for full flexibility, need V^2 units

Recurrent Neural Models: