

# 11-711 Recitation

Akshay Srivatsan

2017-12-01

# Initializing HMM Parameters - $\alpha\beta$

- Forward-backward scores  $\alpha$  and  $\beta$ 
  - $\alpha := \_(\_)\_$
  - $\beta := \_(\_)\_$

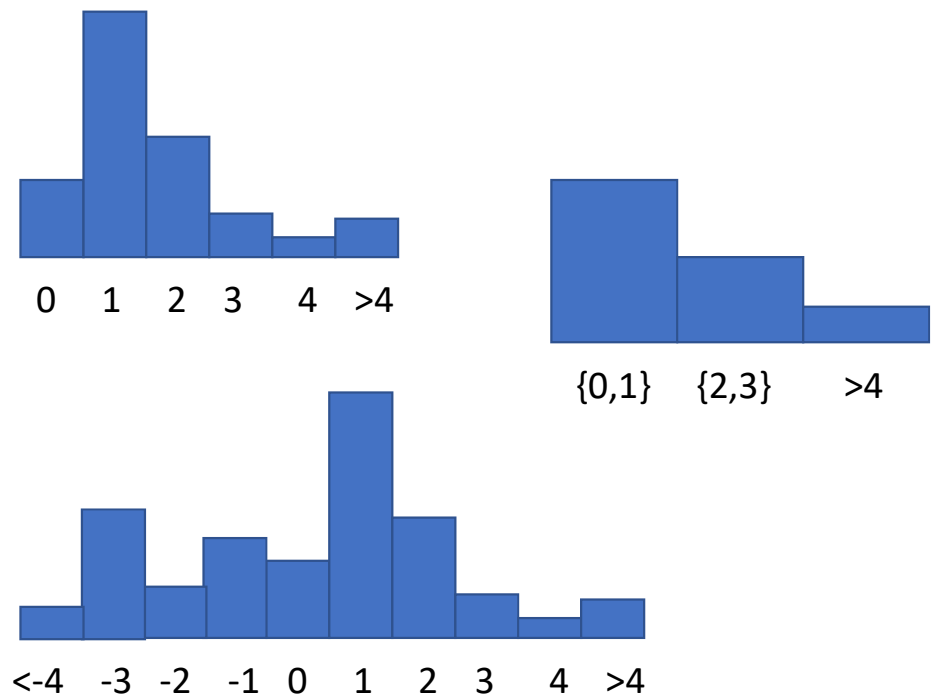
$$\alpha_j^i = \theta_{\mathbf{f}_j, \mathbf{e}_i}^{(t)} \cdot \sum_{i'=0}^{I-1} \alpha_{j-1}^{i'} \cdot p(\mathbf{a}_j = i \mid \mathbf{a}_{j-1} = i', \mathbf{e}; \theta^{(t)}, \psi^{(t)})$$

$$\beta_j^i = \sum_{i'=0}^{I-1} \theta_{\mathbf{f}_j, \mathbf{e}_{i'}}^{(t)} \cdot p(\mathbf{a}_{j+1} = i' \mid \mathbf{a}_j = i, \mathbf{e}; \theta^{(t)}, \psi^{(t)}) \cdot \beta_{j+1}^{i'}$$

# Initializing HMM Parameters - $\psi$

- Fixed transition scores
  - $\psi_k \propto \exp\{-\lambda(k - 1)\}$
- Symmetrical distribution
  - $\psi_{|i-i'|} = P(a_j = i | a_{j-1} = i')$
  - Initialize heavy on 1
- Asymmetrical distribution
  - $\psi_{i-i'} = P(a_j = i | a_{j-1} = i')$
- Prior on first alignment

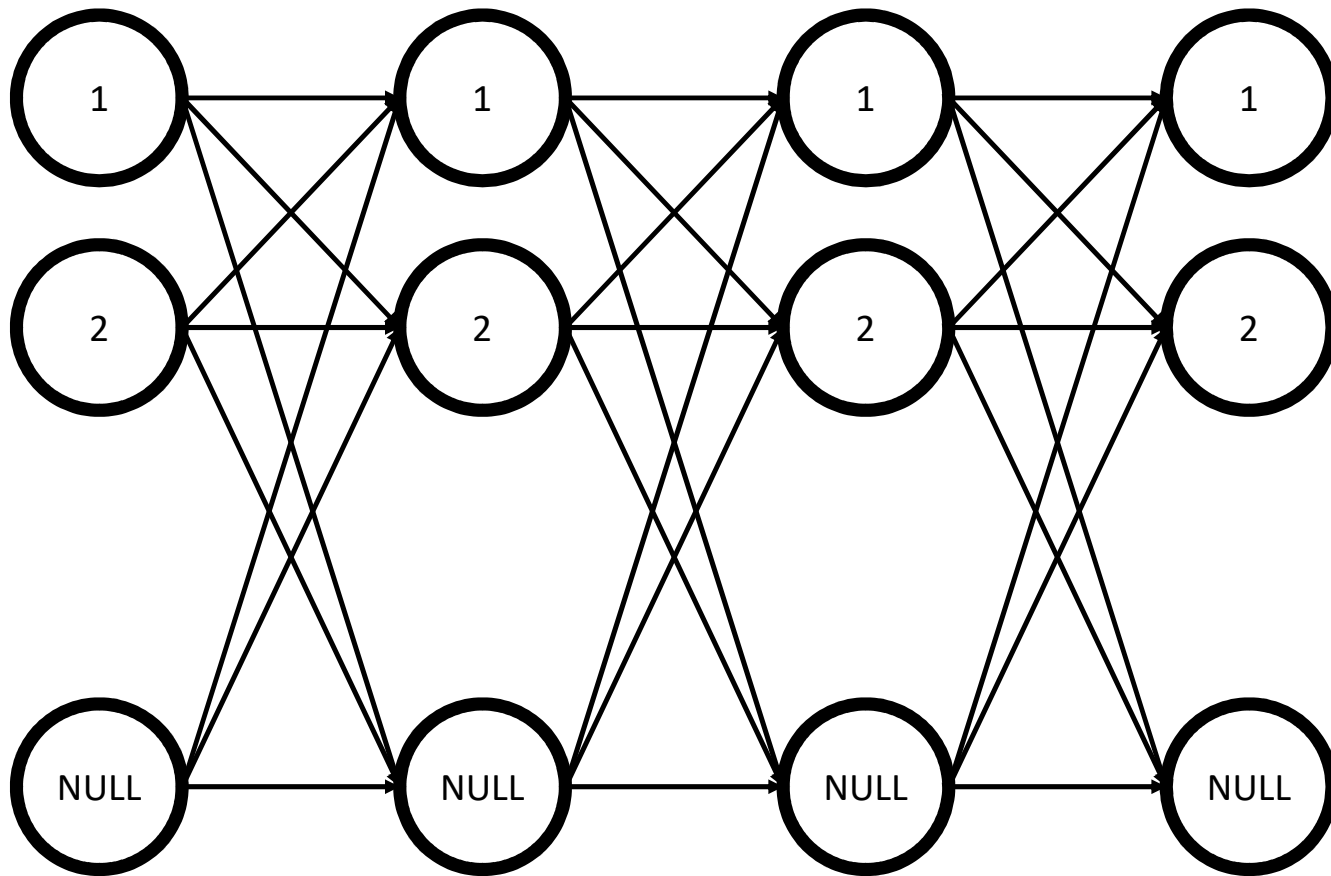
$$p(a_0 | \mathbf{e}) \stackrel{\text{def}}{=} \begin{cases} \varepsilon & \text{if } a_0 \text{ is null} \\ \frac{(1 - \varepsilon) \cdot \psi_{a_0}}{\sum_{i=0}^{I-1} \psi_i} & \text{otherwise} \end{cases}$$



# Initializing Translation Parameters - $\theta$

- Uniform over entire vocabulary
- Uniform over co-occurrences
- Normalized co-occurrence counts
- Pre-training via other models

# Null Transitions



# Parametrizing Null

- Fixed probability of entering null

$$p(a_j = \text{null} | a_{j-1} = i') = \epsilon$$

$$p(a_j = i | a_{j-1} = i') = (1 - \epsilon)\psi_{|i-i'|}$$

- Uniform probability of leaving null

$$p(a_j = i | a_{j-1} = \text{null}) = \frac{1 - \epsilon}{I}$$

# Splitting Null State

- Each null corresponds to a location in the target sentence
  - $e_{i+I}$  is the null equivalent of  $e_i$
- Transition scores involving null are proportional to distance

$$p(i + I | i', I) = p_0 \cdot \delta(i, i')$$

$$p(i + I | i' + I, I) = p_0 \cdot \delta(i, i')$$

$$p(i | i' + I, I) = p(i | i', I)$$

[Och & Ney 2003]