11-711 Recitation

Nikita Srivatsan

2017-12-01

Initializing HMM Parameters - $\alpha\beta$

- Forward-backward scores α and β
  - α ≔¯_(ツ)_/¯
  - β ≔¯_(ツ)_/¯

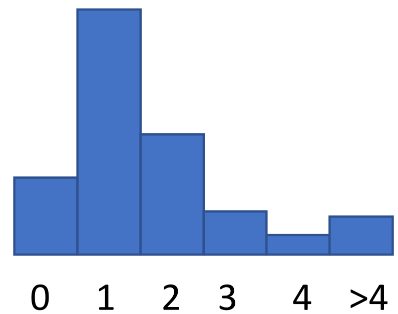
$$\alpha_j^i = \theta^{(t)}_{\mathbf{f}_j,\mathbf{e}_i} \cdot \sum_{i'=0}^{I-1} \alpha_{j-1}^{i'} \cdot p\left(\mathbf{a}_j = i \middle| \mathbf{a}_{j-1} = i', \mathbf{e}; \theta^{(t)}, \psi^{(t)}\right)$$

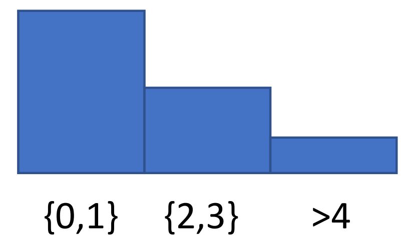
$$\beta_j^i = \sum_{i'=0}^{I-1} \theta^{(t)}_{\mathbf{f}_j,\mathbf{e}_{i'}} \cdot p\left(\mathbf{a}_{j+1} = i' \middle| \mathbf{a}_j = i, \mathbf{e}; \theta^{(t)}, \psi^{(t)}\right) \cdot \beta_{j+1}^{i'}$$

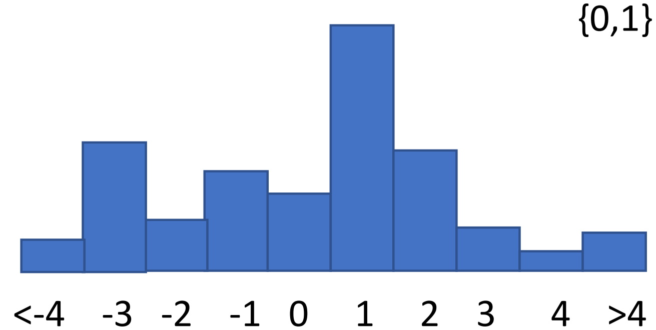
Initializing HMM Parameters - ψ

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

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



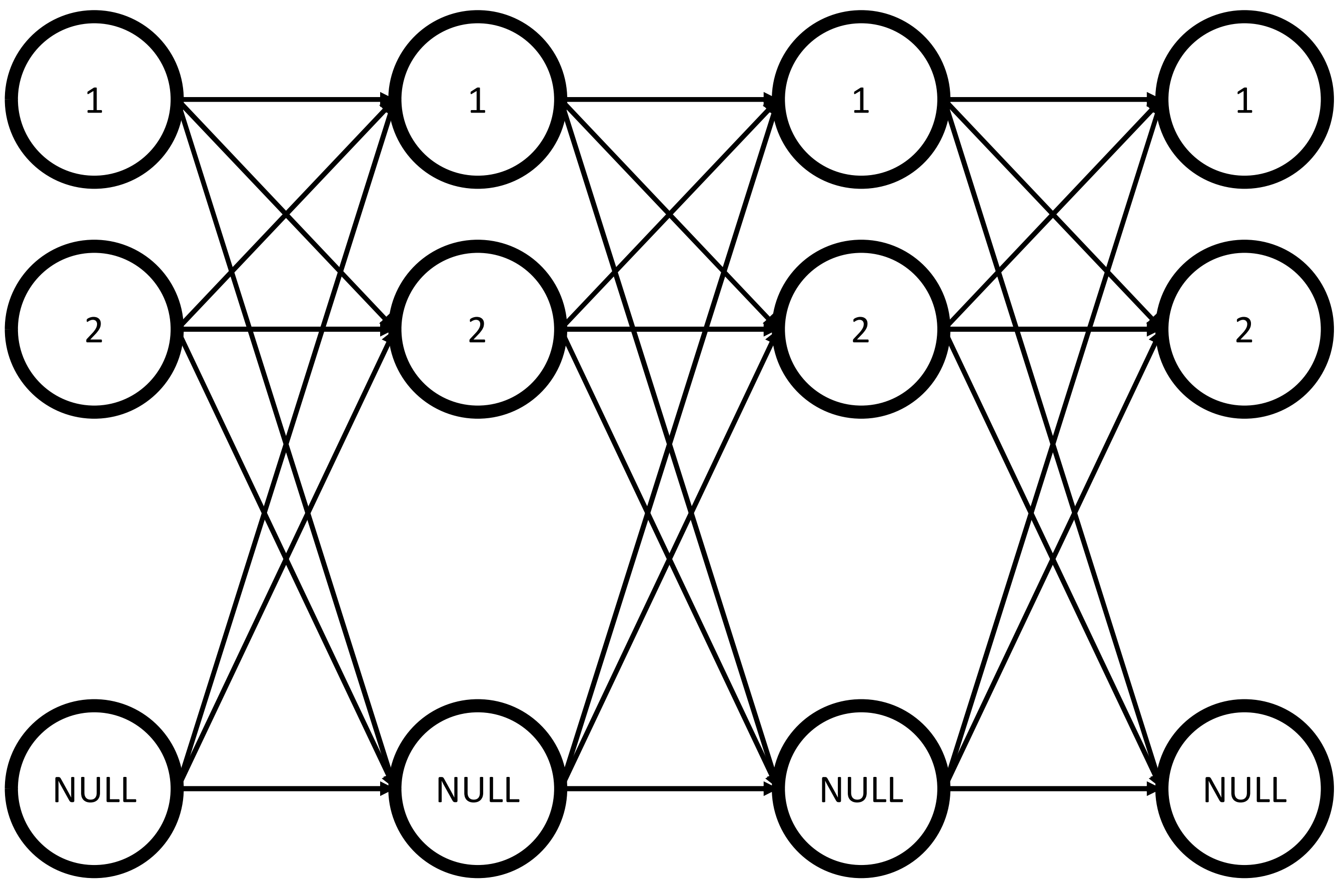




Initializing Translation Parameters - θ

- 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

$$\begin{aligned} p(i+I \mid i', I) &= p_0 \cdot \delta(i, i') \\ p(i+I \mid i'+I, I) &= p_0 \cdot \delta(i, i') \\ p(i \mid i'+I, I) &= p(i \mid i', I) \end{aligned}$$

[Och & Ney 2003]