HMMs: Parameter Estimation

Labeled data $\left(\bar{x}^{(i)}, \bar{y}^{(i)}\right)_{i=1}^{D}$

Maximize $\sum_i \log P\left(\bar{y}^{(i)}, \bar{x}^{(i)}\right)$ generative (joint) likelihood

$$= \sum_i \log P\left(y_1^{(i)}\right) + \underbrace{\sum_i}_{\text{training data}} \underbrace{\sum_j}_{\text{sent index}} \log P\left(x_j^{(i)} \mid y_j^{(i)}\right) + \sum_i \sum_j \log P\left(y_j^{(i)} \mid y_{j-1}^{(i)}\right)$$

MLE with frequency counts: biased coin w/prob p of H

HHHT $\quad 3/4 = \arg\max_p 3 \log p + \log(1-p)$

HMM param estimation: Count + normalizing

Example

$\mathcal{T} = \{N, V, STOP\}$ $\qquad$ $\mathcal{V} = \{they, can, fish\}$

Data

N V STOP
they can

N V STOP
they fish

$$S = \begin{array}{c} N \\ V \end{array}\begin{bmatrix} 2 \\ 0 \end{bmatrix} \;\downarrow\; \begin{array}{c} N \\ V \end{array}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$T = \begin{array}{c|ccc} & N & V & STOP \\ \hline N & 0 & 2 & 0 \\ V & 0 & 0 & 2 \end{array} \;\downarrow\; \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$E = \begin{array}{c|ccc} & they & can & fish \\ \hline N & 2 & & \\ V & & 1 & 1 \end{array} \;\downarrow\; \begin{array}{c|ccc} & t & c & f \\ \hline N & 1 & 0 & 0 \\ V & 0 & 1/2 & 1/2 \end{array}$$

NOT $P(V \mid can)$

$P(can \mid V)$

Smoothing add counts to avoid 0s

+1 everywhere

$$T \to \begin{bmatrix} 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix} = \begin{matrix} 1/5 & 3/5 & 1/5 \\ 1/5 & 1/5 & 3/5 \end{matrix}$$

V-V $\searrow$

Ex N V V
they can fish

$$\Rightarrow P(\bar{y}, \bar{x}) = \begin{matrix} 1 & 3/5 & 1/5 & 3/5 \\ 1 & 1/2 & 1/2 & \end{matrix}$$