HMM

27/01/25

1) Evaluation
2) Decoding
3) Learning

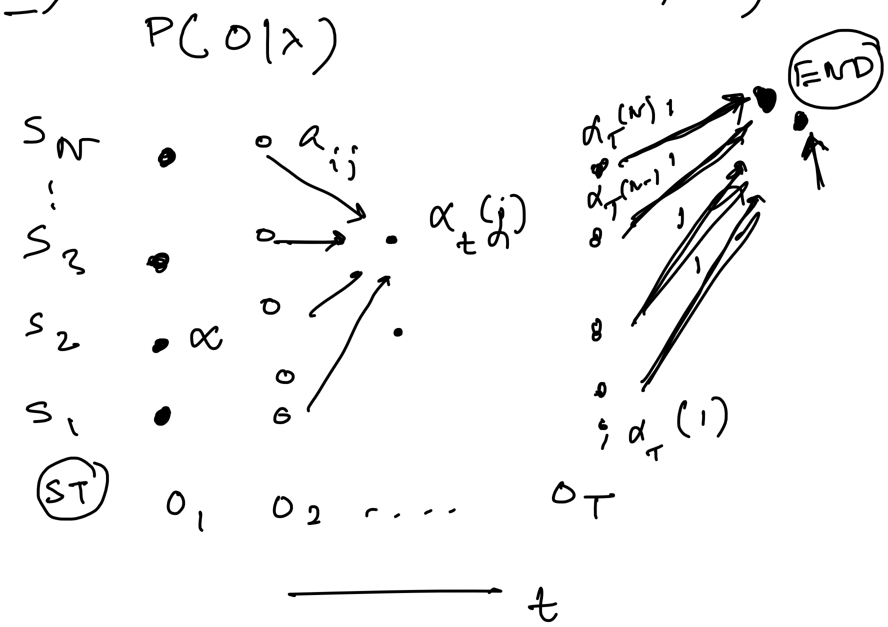
HMM: $\lambda(A, B)$ ← Given, $O = (O_1, O_2 \dots, O_T)$

a_{ij}

$P(O|\lambda)$

forward prob.

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i)\, a_{ij}\, b_j(O_t)$$



initialization

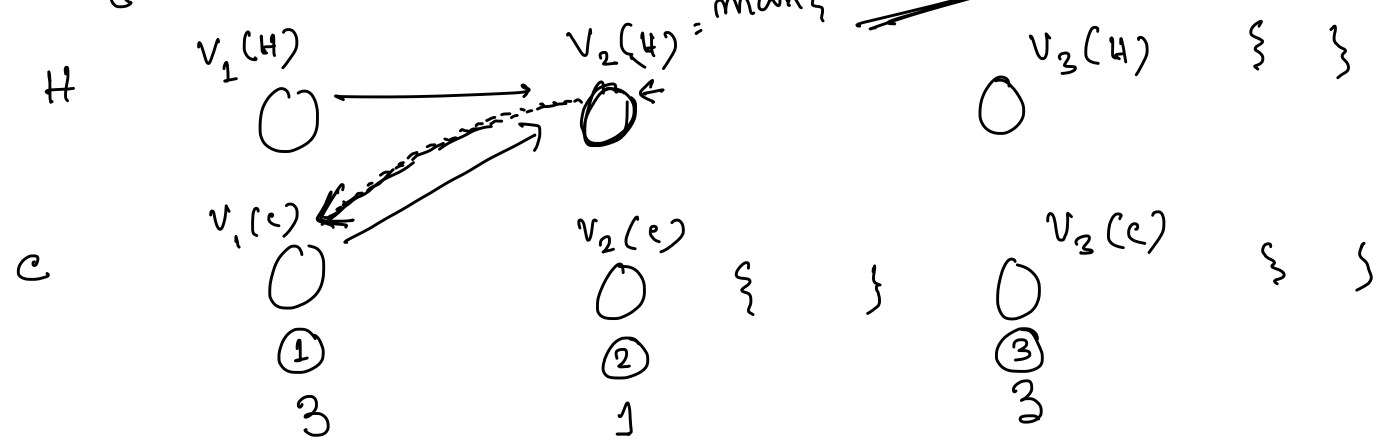
$$\alpha_1(j) = \pi_j\, b_j(O_1) : 1 \le j \le N$$

Termination

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

Decoding — / Viterbi Al

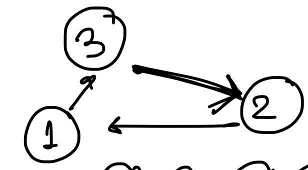
$V_2(H) = \max\{V_1(C) \times a_{CH},\ V_1(H) \times a_{HH}\}$



Learning $O: \{O_1, \ldots, O_T\}$ $\lambda = (A, B)$

Forward-Backward Algo

A $a_{ij} = \dfrac{\text{Cnt}(i \to j)}{\text{Cnt}(i \to x)}$



- 3 1 2 1 3 3
- 3 2 1 1 1 2
- 2 1 1 1 3 2
- .
- .

3 3 2
H H C

1 1 2
C C C

1 2 3
C H H E

$\pi_h = 1/3$ $\pi_c = 2/.$

π_3

π_2

$P(H|H) = 2/3$

a_{HH} $P(H|C) = 1/3$

A, B

$a_t(i)$ $b_j(O_t)$

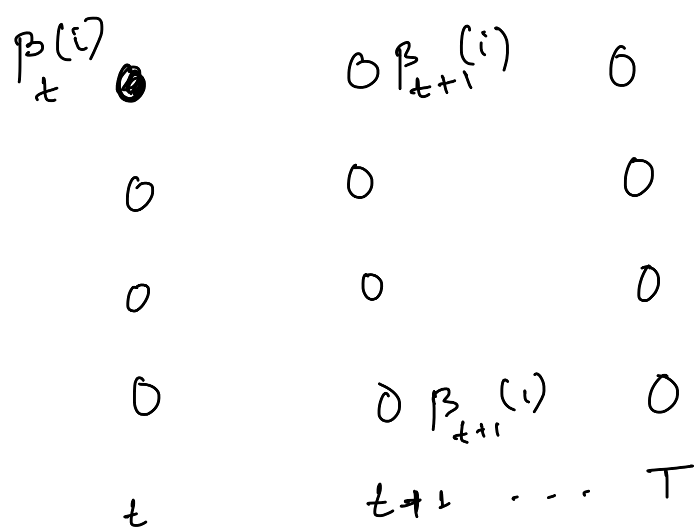
Iteratively :/ EM Algo

- Backward prob.

$\beta_t(i)$ = prob. of seeing the observation from $(t+1)$ to the end, given that we are in the state i at time t

initial:

$\beta_T(i) = 1 / a_{i,F}$



$$\beta_t(i) = \sum_{j=1}^{N} \beta_{t+1}(j)\, a_{ij}\, b_j(O_{t+1})$$

$\beta_0(1)$

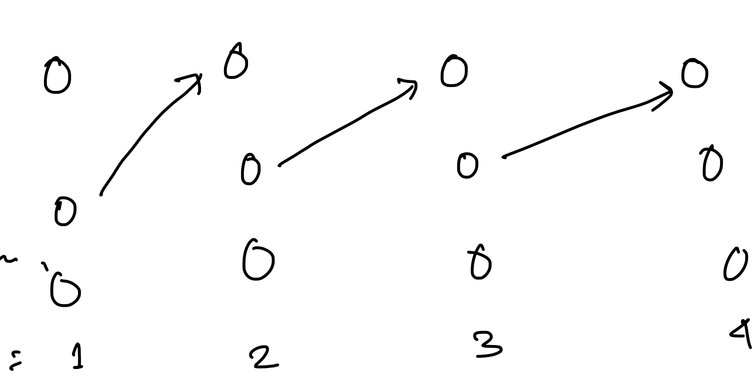
Termination

$$P(O|\lambda) = \sum_{j=1}^{N} \beta_1(j)\, \pi_j\, b_j(O_1)$$

$$\hat{a}_{ij} = \frac{\text{expected number of transitions from } i \to j}{\text{expected no of trans from } i}$$

t $i \to j$

Assume we had some estimate of the prob. that given transition $i \to j$ was taken at a particular time t in the seq of observation. If we knew this prob for each t, we could sum over all time t to estimate the total count for $i \to j$



$\xi_t(i,j)$ = prob of being in state i at time t and state j at time (t+1), given the observation seq

$$= P(q_t = i, q_{t+1} = j \mid O, \lambda)$$

$$P(X \mid Y, Z) = \frac{P(X, Y \mid Z)}{P(Y \mid Z)} = \frac{P(q_t = i, q_{t+1} = j, O \mid \lambda)}{P(O \mid \lambda)}$$

$$\xi(i,j) = \frac{\alpha_t(i)\, \beta_{t+1}(j)\, a_{ij}\, b_j(O_{t+1})}{P(O \mid \lambda)}$$

$$\boxed{\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \sum_{k=1}^{N} \xi_t(i,k)}}$$

$$\hat{b}_j(O_t) = \frac{\text{expd no of times in state } j \text{ and occurring } O_t}{\text{expd no of times in state } j}$$

$$\gamma_t(j) = P(q_t = j \mid O, \lambda)$$

$$= \frac{P(q_t = j, O \mid \lambda)}{P(O \mid \lambda)} = \frac{\alpha_t(j)\, \beta_t(j)}{P(O \mid \lambda)}$$

$$\boxed{\hat{b}_j(d_t) = \frac{\sum_{t=1 \text{ s.t. } O_t = d_t}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}}$$

initialize A, B

E - Step:

$\gamma_t(j)$

$\xi_t(i,j)$

M - Step

$\hat{a}_{ij}$

$\hat{b}_j(d_t)$

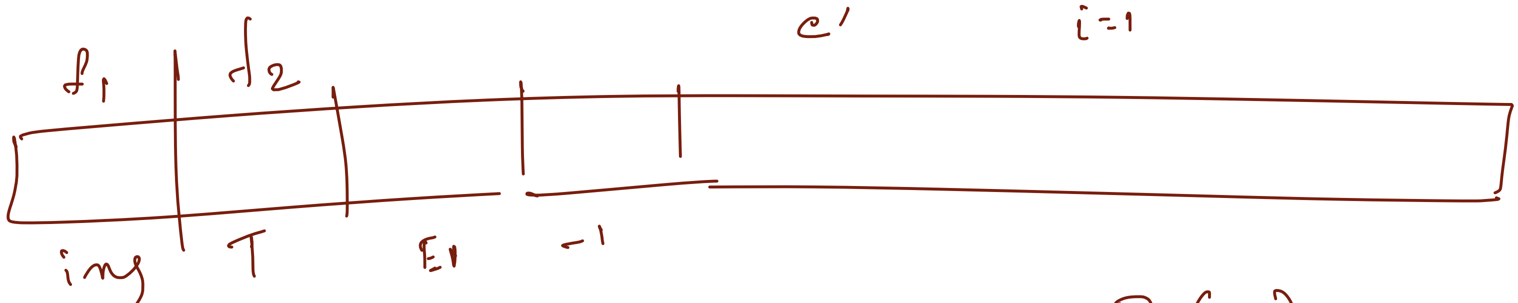
Maximum Entropy Modeling

f_i = ith feature of x.

$$P(c \mid x) = \frac{\exp \sum_{i=1}^{F} w_i f_i}{C}$$

$$= \frac{\exp \sum w_i f_i}{\sum_{c} \exp \sum_{i=1}^{F} w_i f_i}$$

$$= \frac{\exp \sum w_{ci} f_i}{\sum_{c'} \exp \sum_{i=1}^{F} w_{c'i} f_i}$$



$$\hat{w} = \arg\max \sum_{i} \log P(\underline{\qquad}) - \alpha \underline{R(W)}$$