CSE 575: Statistical Machine Learning (Spring 2021)

Instructor: Nupur Thakur

Graphical Models

Arizona State University

Table of contents

Problems in HMM

- For a given HMM $\Lambda=\{\Theta,\Omega,A,B,\pi\}$
  - Estimation of model parameters
  - Given an observation sequence **O**= $\{o^1,o^2, \ldots ,o^k\}$, what is the most likely state sequence $S = \{s^1, s^2, \ldots ,s^k\}$ that has produced O?
  - How likely is an observation O?

HMM Parameter Estimation

- Given labeled data - state and observation

A

$$t(S_i|S_j) = \frac{\text{number of } (s^t = S_i, s^{t-1} = S_j)}{\text{number of } S_j}$$

B

$$e(o_r|S_j) = \frac{\text{number of } (o^t = o_r, s^t = S_j)}{\text{number of } S_j}$$

- Given observation sequences -

Forward - Backward Algorithm - Expectation - Maximization Approach.

HMM State Estimation — Decoding

$$P(s_1|s^*) \quad P(o_1|s_1)$$

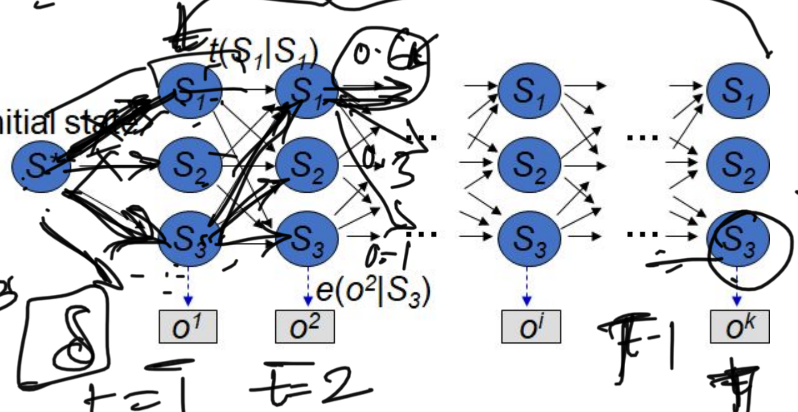
$$P(s_1,o_1) = P(s_1|s^*)\,P(o_1|s_1)\,P(s^*)$$

- Given an observation (sequence) O= {o_1, o_2, … ,o_k}, what is the most likely state sequence S={s_1, s_2, … ,s_k} that has produced O?

$$\underset{S}{\operatorname{argmax}}\ P(S|O) = \frac{P(S,O)}{P(O)}$$

Length of sequence T

No. of states



Trellis

$\psi = s^*, s_3$

δ

t=1 t=2 T-1 T

HMM State Estimation

$$P(s_1) \quad s_1 \xrightarrow{P(s_2|s_1)} s_2 \xrightarrow{P(s_3|s_2)} s_3$$

$$\{s_1, s_2, s_3\}$$

$$P(S,O) = P(O|S) \cdot P(S)$$

$$= P(o^1 \dots o^k \mid s_1 \dots s^k) \prod_{j=1}^{k} P(s^j \mid s^1 \dots s^k)$$

$$= \prod_{i=1}^{k} P(o^i \mid o^1 \dots o^k, s^1 \dots s^k) \prod_{j=1}^{k} P(s^j \mid s^1 \dots s^k)$$

$$= \prod_{i=1}^{k} P(o^i \mid s^i) \qquad \prod_{j=1}^{k} P(s^j \mid s^{j-1})$$

Emission probability

$$P(S,O) = \prod_{i=1}^{k} P(o^i \mid s^i) \prod_{j=1}^{k} P(s^j \mid s^{j-1})$$

Transition probability

HMM State Estimation - Viterbi Algorithm

Initialization $\delta_{S_i}(1) = t(S_i|s^*)e(o^1|S_i), \quad \forall S_i \in \Theta$

Induction:

for 2≤*t*≤*k*, do

$$\delta_{S_i}(t) = \max_{S_j} t(S_i|S_j)e(o^t|S_i)\delta_{S_j}(t-1)$$

$$\psi_{S_i}(t) = \operatorname*{argmax}_{S_j} t(S_i|S_j)e(o^t|S_i)\delta_{S_j}(t-1)$$

Termination:

- The probability of the best state sequence: $\max_{S_j} \delta_{S_j}(k)$
- The best last state: $\hat{s}^k = \operatorname*{argmax}_{S_j} \delta_{S_j}(k)$
- Back trace to get other states:

$$\hat{s}^t = \psi_{\hat{s}^{t+1}}(t), \text{ for } t = k-1, \ldots, 1.$$

HMM - Evaluate P(O)

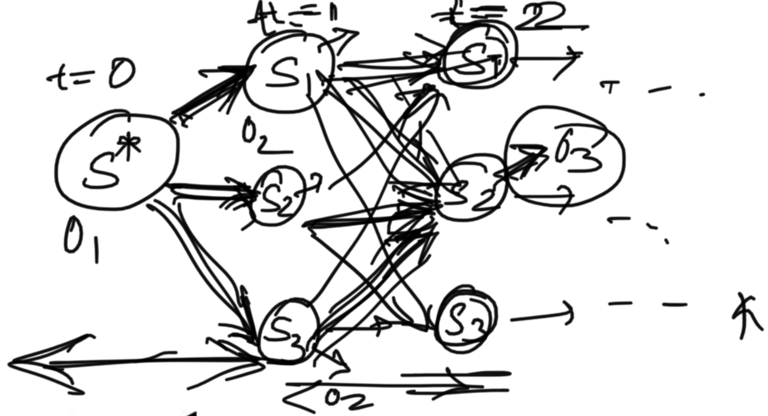
- How likely is an observation O? $O = \{O_1 \ldots O_K\}$

$$P(O) = \sum_S P(S, O)$$

— Forward Algorithm — α $(t-1)$ at t

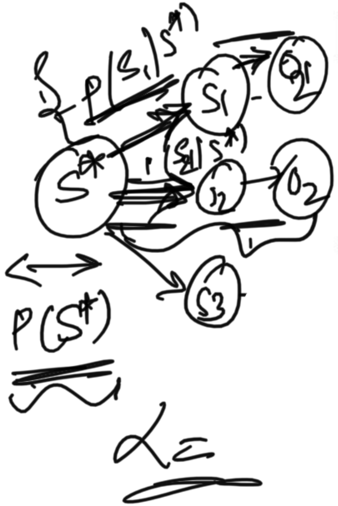
— Backward Algorithm — β $(t+1)$ to T at t^{th} step

HMM - Evaluate P(O)



- Forward Algorithm

Initialization: $\alpha_{S_i}(1) = t(S_i|s^*)e(o^1|S_i), \quad \forall S_i \in \Theta$



Induction:
for 2≤*t*≤*k*, do

$$\alpha_{S_i}(t) = \sum_{S_j} t(S_i|S_j)e(o^t|S_i)\alpha_{S_j}(t-1)$$

Termination:

$$P(\boldsymbol{O}) = \sum_{S_j} \alpha_{S_j}(k)$$

Questions?