Notetaker: Andrew Kwok

SANGMAN KIM.

Lecture 11

9 May: Hidden Markov Models

Compute: $P^\lambda(\mathcal{O})$, $P^\lambda(q, \mathcal{O})$

$\mathcal{O} = (o_0, o_1, \ldots, o_{T-1})$

# of operations to compute $P^\lambda(\mathcal{O})$ "blindly" is $\sim N^T T$
(N is # of states in hidden MC)

Forward Algorithm (relates to training, recognition)
(dynamic programming idea)

Fix $\left|\begin{array}{l}\lambda = (a, b, \pi) \\ \mathcal{O}^* = (o_0^*, \ldots, o_{T-1}^*)\end{array}\right.$

Write s_i as i for simplicity.

Let $\alpha_t^\lambda(i) = P^\lambda(q_t = i,\ o_t = O_t^*, \ldots, o_0 = O_0^*)$

$t = 0, 1, \ldots, T-1 \qquad i = 1, \ldots, N$

(i) $\alpha_0^\lambda(i) = P^\lambda(q_0 = i,\ o_0 = O_0^*) = \pi_i b_i(O_0^*) \qquad (i = 1, \ldots, N)$

Recursion (ii) $\alpha_{t+1}^\lambda(j) = P^\lambda(q_{t+1} = j,\ o_{t+1} = O_{t+1}^*,\ o_t = O_t^*, \ldots, o_0 = O_0^*)$

$$= \sum_{i=1}^N P^\lambda(q_{t+1} = j,\ o_{t+1} = O_{t+1}^*,\ q_t = i,\ o_t = O_t^*, \ldots, o_0 = O_0^*)$$

$$= \sum_{i=1}^N P^\lambda(q_{t+1} = j,\ o_{t+1} = O_{t+1}^* \mid q_t = i,\ o_t = O_t^*, \ldots, o_0 = O_0^*) \cdot P^\lambda(q_t = i,\ o_t = O_t^*, \ldots, o_0 = O_0^*)$$

Markov Prop of HMM $\longrightarrow$ $= \sum_{i=1}^N P^\lambda(q_{t+1} = j,\ o_{t+1} = O_{t+1}^* \mid q_t = i)\, \alpha_t^\lambda(i)$

$$= \sum_{i=1}^N a_{ij}\, b_j(O_{t+1}^*)\, \alpha_t^\lambda(i)$$

so, $P^\lambda(O^*) = \sum_{i=1}^N P^\lambda(q_{T-1} = i,\ o|_{T-1} = O^*) = \sum_{i=1}^N \alpha_{T-1}^\lambda(i)$

# of operations to compute $P^\lambda(O^*)$ using forward algorithm

$$= \underbrace{(N-1)}_{\text{additions}} + (T-1)\cdot[N(\underbrace{N-1}_{\text{adds}} + \underbrace{2N}_{\text{mults}})] + \underbrace{N \cdot 1}_{\text{mults}} = \mathcal{O}(N^2 T)$$

ie, for $N=5$; $T=100$, $\begin{cases}\text{fwd alg} \sim 3000 \text{ ops} \\ \text{"Blind" method} \sim 10^{72}\end{cases}$!!!

Backward Algorithm

Fix $\left| \begin{array}{l} \lambda = (a, b, \pi) \\ \mathcal{O}^* = (o_0^*, \ldots, o_{T-1}^*) \end{array} \right.$

Let $\beta_t^\lambda(i) = P^\lambda(o_{t+1} = o_{t+1}^*, \ldots, o_{T-1} = o_{T-1}^* \mid q_t = i)$ $\quad i = 1, \ldots, N$, $\quad t = 0, 1, \ldots, T-1$

(i) $\beta_{T-1}^\lambda(i) = 1$ for $i = 1, \ldots, N$

Recursion (backwards) (ii) $\beta_t^\lambda(j) = \sum_{k=1}^N P^\lambda(o_{t+1} = o_{t+1}^*, \ldots, o_{T-1} = o_{T-1}^*, q_{t+1} = k \mid q_t = j)$ $\quad t = 0, 1, \ldots, T-2$, $\quad j = 1, \ldots, N$

$$= \sum_{k=1}^N P^\lambda(o_{t+2} = o_{t+2}^*, \ldots, o_{T-1} = o_{T-1}^* \mid q_{t+1} = k, o_{t+1} = o_{t+1}^*, q_t = j) \cdot P^\lambda(q_{t+1} = k, o_{t+1} = o_{t+1}^* \mid q_t = j)$$

by MP of HMM $\longrightarrow$ $$= \sum_{k=1}^N \underbrace{P^\lambda(o_{t+2} = o_{t+2}^*, \ldots, o_{T-1} = o_{T-1}^* \mid q_{t+1} = k)}_{\beta_{t+1}^\lambda(k)} a_{jk} \, b_k(o_{t+1}^*)$$

$$= \sum_{k=1}^N \beta_{t+1}^\lambda(k) \, a_{jk} \, b_k(o_{t+1}^*)$$

Now: $\beta_0^\lambda(i) = P^\lambda(o_1 = o_1^*, \ldots, o_{T-1} = o_{T-1}^* \mid q_0 = i)$

$$P^\lambda(\mathcal{O}) = P^\lambda(o_0 = o_0^*, \ldots, o_{T-1} = o_{T-1}^*)$$

$$= \sum_{i=1}^N P^\lambda(q_0 = i, o_0 = o_0^*, \ldots, o_{T-1} = o_{T-1}^*)$$

$$= \sum_{i=1}^N P^\lambda(o_1 = o_1^*, \ldots, o_{T-1} = o_{T-1}^* \mid q_0 = i, o_0 = o_0^*) \, P^\lambda(q_0 = i, o_0 = o_0^*)$$

MP of HMM $\rightarrow$ $$= \sum_{i=1}^N P^\lambda(o_1 = o_1^*, \ldots, o_{T-1} = o_{T-1}^* \mid q_0 = i) \, \pi_i \, b_i(o_0^*)$$

$$= \sum_{i=1}^N \beta_0^\lambda(i) \, \pi_i \, b_i(o_0^*)$$

similar operation count as forward algorithm.

Viterbi Algorithm (Decoding)

Given: $\mathcal{O}^* = (o_0^*, \ldots, o_{T-1}^*)$

Fix: $\lambda = (a, b, \pi)$

Value function: $V^\lambda(\mathcal{O}^*) = \max_q P^\lambda(\mathcal{O}^*, q)$ $\quad q = (q_0, \ldots, q_{T-1})$

$q^* = \operatorname{argmax}_q P^\lambda(\mathcal{O}^*, q)$ = a value of q that achieves $V^\lambda(\mathcal{O}^*)$

Let $\delta_t^\lambda(i) = \max\limits_{\hat{q}=(q_0,\ldots,q_{t-1})} P^\lambda(q_t = i,\ q|_{t-1} = \hat{q},\ O_t = O_t^*, \ldots, O_0 = O_0^*)$ $\quad t = 0, 1, \ldots, T-1$, $\quad i = 1, \ldots, N$

(i) $\delta_0^\lambda(i) = P^\lambda(q_0 = i,\ O_0 = O_0^*) = \pi_i\ b_i(O_0^*)$

Recursion (ii)

$$\delta_{t+1}^\lambda(j) = \max_{\hat{q}=(q_0,\ldots,q_t)} P^\lambda(q_{t+1} = j,\ q|_t = \hat{q},\ O|_{t+1} = O^*|_{t+1})$$

$$= \max_{i=1}^{N} \max_{\tilde{q}=(q_0,\ldots,q_{t-1})} P^\lambda(q_{t+1} = j,\ q_t = i,\ q|_{t-1} = \tilde{q},\ O_{t+1} = O_{t+1}^*,\ O|_t = O^*|_t)$$

$$= \max_{i=1}^{N} \max_{\tilde{q}} P^\lambda(q_{t+1} = j,\ O_{t+1} = O_{t+1}^* \mid q_t = i,\ q|_{t-1} = \tilde{q},\ O|_t = O^*|_t) \cdot P^\lambda(q_t = i,\ q|_{t-1} = \tilde{q},\ O|_t = O^*|_t)$$

MP $\longrightarrow$

$$= \max_{i=1}^{N} \max_{\tilde{q}} \underbrace{P^\lambda(q_{t+1} = j,\ O_{t+1} = O_{t+1}^* \mid q_t = i)}_{a_{ij}\ b_j(O_{t+1}^*)} \cdot P^\lambda(q_t = i,\ q|_{t-1} = \tilde{q},\ O|_t = O^*|_t)$$

$$= \max_{i=1}^{N} \left[a_{ij}\ b_j(O_{t+1}^*)\ \max_{\tilde{q}} P^\lambda(q_t = i,\ q|_{t-1} = \tilde{q},\ O|_t = O^*|_t)\right]$$

$$= \max_{i=1}^{N} \left[a_{ij}\ b_j(O_{t+1}^*)\ \delta_t^\lambda(i)\right]$$

Now let $\psi_{t+1}^\lambda(j) = \operatorname*{argmax}_{i=1}^{N} \left(a_{ij}\ b_j(O_{t+1}^*)\ \delta_t^\lambda(i)\right)$

Finally, $V^\lambda(O^*) = \max\limits_{i=1}^{N} \delta_{T-1}^\lambda(i)$

and, $\delta_{T-1}^\lambda(i) = \max\limits_{\tilde{q}=(q_0,\ldots,q_{T-2})} P^\lambda(q_{T-1} = i,\ q|_{T-2} = \tilde{q},\ O|_{T-1} = O^*|_{T-1})$

Backtrack to find the optimal q:

$$q_{T-1}^* = \operatorname*{argmax}_{i=1}^{N} \delta_{T-1}^\lambda(i)$$

$\vdots$

$$q_t^* = \psi_{t+1}^\lambda(q_{t+1}^*)$$

$\vdots$

operation count is $\sim N^2 T$