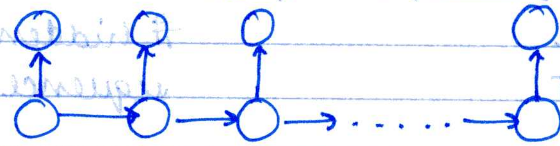
Review

* Hidden Markov Models



... observations $O_t \in \{1,2,\dots m\}$

... states $S_t \in \{1,2,\dots n\}$

* Parameters

$a_{ij} = P(S_{t+1}=j / S_t = i)$ transition matrix

$b_{ik} = P(O_t = k / S_t = i)$ emission matrix

$\pi_i = P(S_1 = i)$ initial state distribution

* Key questions

1) How to compute likelihood $P(O_1, O_2, \dots O_T)$?

2) How to decode $\underset{S_1, S_2, \dots S_T}{\operatorname{argmax}} P(S_1, S_2, \dots S_T / O_1, O_2, \dots O_T)$?

$\boxed{\text{inference}}$

3) How to monitor $P(S_t = i / O_1, O_2, \dots O_t)$ in real-time?

4) How to estimate $\{\pi_i, a_{ij}, b_{ik}\}$?

$\boxed{\text{learning}}$

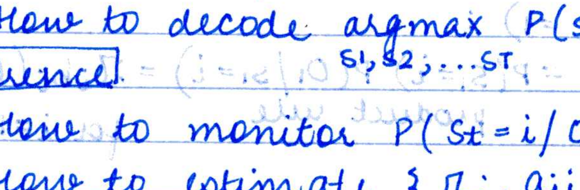
(1) Computing likelihood $P(O_1, O_2, \dots O_T)$

* Efficient recursion

$$P(O_1, O_2, \dots O_t, O_{t+1}, S_{t+1} = j) = \underbrace{\sum_{i=1}^{n} P(O_1, O_2, \dots O_t, O_{t+1}, S_{t+1}=j, S_t = i)}_{\text{marginalization}}$$

$$= \sum_{i=1}^{n} P(O_1, O_2, \dots O_t, S_t = i)\, P(S_{t+1} = j, O_{t+1} / S_t = i, O_1, O_2, \dots O_t)$$

$$= \sum_{i=1}^{n} P(O_1, O_2, \dots O_t, S_t = i)\, P(S_{t+1} = j, O_{t+1} / S_t = i)$$ conditional independence



Product rule + CI

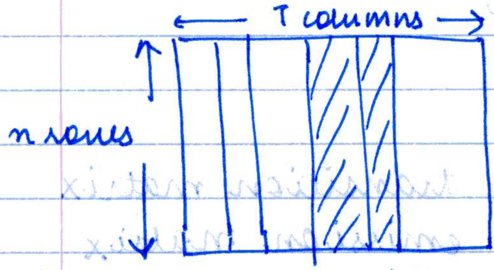
$$= \sum_{i=1}^{n} \underbrace{P(O_1, O_2, \dots O_t, S_t = i)}_{\text{recursive instance}}\, \underbrace{P(S_{t+1} = j / S_t = i)}_{\text{transition matrix}}\, \underbrace{P(O_{t+1} / S_{t+1} = j, \cancel{S_t = i})}_{\text{emission matrix}}$$

* Shorthand notation

$$\alpha_{it} = P(O_1, O_2, \ldots O_t, S_t = i)$$

$i = 1, 2, \ldots n$ # hidden states

$t = 1, 2, \ldots T$ sequence length




* Forward algorithm

2) Recursive step:

$$\alpha_{j,t+1} = \sum_{i=1}^{n} \alpha_{it}\, a_{ij}\, b_j(O_{t+1})$$ same as $b_{j,O_{t+1}}$

1) Initial column ($t = 1$):

$$\alpha_{i1} = P(O_1, S_1 = i) = P(S_1 = i)\, P(O_1 / S_1 = i) = \pi_i b_i(O_1)$$

product rule

for $i = 1, 2, \ldots n$

* Likelihood

$$P(O_1, O_2, \ldots O_{T-1}, O_T) = \sum_{i=1}^{n} P(O_1, O_2, \ldots O_T, S_T = i)$$

marginalization

$$= \sum_{i=1}^{n} \alpha_{iT}$$

Scales as $O(n^2 T)$

Linear, not exponential, in sequence length.
Quadratic in # hidden states.

* Warning: more calculations will underflow for long sequences $T \gg 100$.

(2) How to compute most likely hidden state sequence?

$$S^* = \{S_1^*, S_2^*, \ldots S_T^*\}$$

$$= \operatorname{argmax} P(S_1, S_2, \ldots S_T / O_1, O_2, \ldots O_T)$$

$$= \underset{S}{\operatorname{argmax}} \; P(s_1, s_2, \ldots s_T / O_1, O_2, \ldots O_T)$$

$$= \underset{S}{\operatorname{argmax}} \left[ \frac{P(s_1, s_2, \ldots s_T, O_1, O_2, \ldots O_T)}{P(O_1, O_2, \ldots O_T)} \right]$$ product rule

$$= \underset{S}{\operatorname{argmax}} \; P(s_1, s_2, \ldots s_T, O_1, \ldots O_T)$$ b/c denominator is constant w.r.t. S

How to compute $S^* = \underset{S}{\operatorname{argmax}} \; \log P(s_1, \ldots s_T, O_1, \ldots O_T)$?
↑ monotonic function

Define: $l_{it}^*$ - log probability of most likely t-step state sequence that accounts for first t observations and that ends in state i at time t.

$$= \max_{\{s_1, s_2, \ldots s_{t-1}\}} \log P(s_1, s_2, \ldots s_{t-1}, s_t = i, O_1, O_2, \ldots O_t)$$

Recursion

1) Base case $(t=1)$

$$l_{i1}^* = \log P(s_1 = i, O_1) = \log [P(s_1 = i) P(O_1 / s_1 = i)]$$ product rule

$$= \log [\pi_i \, b_i(O_1)]$$

$$= \log \pi_i + \log b_i(O_1)$$

2) from time t to time t+1

$$l_{j,t+1}^* = \max_{s_1, s_2, \ldots s_t} \log P(s_1, s_2, \ldots s_t, s_{t+1} = j, O_1, O_2, \ldots O_t, O_{t+1})$$

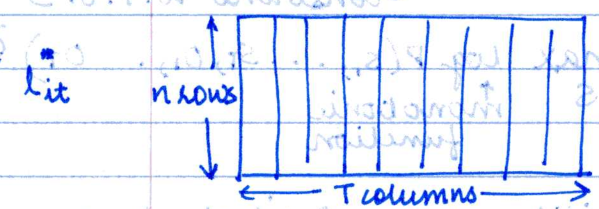
$$= \max_{s_1, s_2, \ldots s_{t-1}} \max_i \log P(s_1, s_2, \ldots s_t = i, s_{t+1} = j, O_1, \ldots O_t, O_{t+1})$$

$$= \max_{s_1, s_2, \ldots s_{t-1}} \max_i \log [P(s_1, s_2, \ldots s_{t-1}, s_t = i, O_1, O_2, \ldots O_t)$$
$$P(s_{t+1} = j / \cancel{s_1, s_2, \ldots} s_t = i, \cancel{O_1, \ldots O_t})$$ CI
$$P(O_{t+1} / \cancel{s_1, s_2, \ldots s_t = i}, s_{t+1} = j, \cancel{O_1, O_2, \ldots O_t})]$$
product rule

$$= \max_{s_1, s_2, \ldots s_{t-1}} \max_i \log [P(s_1, \ldots s_{t-1}, s_t = i, O_1, \ldots O_t) P(s_{t+1} = j / s_t = i) P(O_{t+1} / s_{t+1} = j)]$$

$$l^*_{j,t+1} = \max_i \Big[ \max_{s_1,s_2,\dots s_{t-1}} \log P(s_1,s_2,\dots s_t=i, O_1,O_2,\dots O_t) + \log P(s_{t+1}=j / s_t=i) + \log P(O_{t+1} / s_{t+1}=j) \Big]$$

$$l^*_{j,t+1} = \max_i \left[ l^*_{it} + \log a_{ij} \right] + \log b_j(O_{t+1})$$




fill in $l^*$ column-by-column

* How to derive $S^*$ from $l^*$?

Record most likely transitions

$\Phi_{t+1}(j)$ = most likely state at time t given that we land at time t+1 in state $S_{t+1}=j$.
(with observations $O_1, O_2, \dots O_{t+1}$)

$$\Phi_{t+1}(j) = \operatorname*{argmax}_i \left[ l^*_{it} + \log a_{ij} \right]$$

* Compute $S^*$ by back-tracking;

end of sequence, last hidden state $\rightarrow$ $S^*_T = \operatorname*{argmax}_i [l^*_{iT}]$

for $t = T-1$ to $1$
$S^*_t = \Phi_{t+1}(S^*_{t+1})$ } "Backward" pass

$S^*$ is known as "Viterbi" path (state sequence)
Viterbi algorithm is instance of dynamic programming

(3) Belief updating

How to update beliefs in real time based on incoming evidence?

$q_{it} = P(S_t=i / O_1, O_2, \dots O_t)$ cond. prob that $S_t=i$ based on evidence only upto time t.

$$q_{it} = \frac{P(S_t=i, O_t / O_1, O_2, \dots O_{t-1})}{P(O_t / O_1, O_2, \dots O_{t-1})}$$ product rule