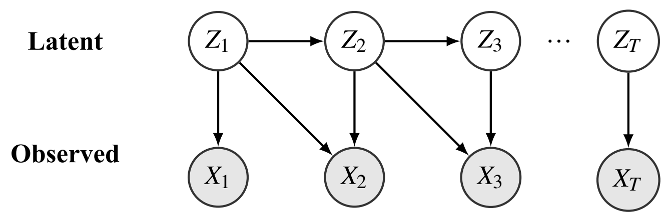
Notes

September 28

We tend to find the Forward-backward algorithm for the new model.

Suppose we have T observations and T hidden states. Let X_i denote the ith observation where $X_i \in \{v_1, v_2, \cdots, v_k\}$ (or just 1, $\cdots$k for convinience). Let Z_i denote the ith hidden states where $Z_i \in \{o_1, o_2, \cdots, o_m\}$ (or just 1, $\cdots$ m for convinience). The graphical model is below.



Generally we need to know $P(X_{1:T})$ for future computation. From property of HMM we know

$$P(X_{1:T}, Z_{1:T}) = P(Z_1)P(X_1|Z_1)\prod_{t=2}^{T} P(Z_t|Z_{t-1})P(X_t|Z_{t-1}, Z_t)$$

Thus we repeat same operation,

$$\begin{aligned}
P(X_{1:T}) &= \sum_{Z_{1:T}} P(X_{1:T}, Z_{1:T}) \\
&= \sum_{Z_{1:T}} \underbrace{P(Z_1)P(X_1|Z_1)}_{S_1(Z_1)} \prod_{t=2}^{T} P(Z_t|Z_{t-1})P(X_t|Z_{t-1}, Z_t) \\
&= \sum_{Z_{2:T}} \underbrace{\left(\sum_{Z_1} S_1 P(Z_2|Z_1)P(X_2|Z_1, Z_2)\right)}_{S_2(Z_2)} \prod_{i=3}^{T} P(Z_t|Z_{t-1})P(X_t|Z_{t-1}, Z_t) \\
&\cdots \\
&= \sum_{Z_{j+1:T}} \underbrace{\left(\sum_{Z_j} S_j P(Z_{j+1}|Z_j)P(X_{j+1}|Z_j, Z_{j+1})\right)}_{S_{j+1}(Z_{j+1})} \prod_{t=j+2}^{T} P(Z_t|Z_{t-1})P(X_t|Z_{t-1}, Z_t)
\end{aligned}$$

By same kind of trick, we also have

$$
\begin{aligned}
P(X_{1:T}) &= \sum_{Z_{1:T}} P(X_{1:T}, Z_{1:T}) \\
&= \sum_{Z_{1:T-1}} \underbrace{\sum_{Z_T} P(Z_T|Z_{T-1})P(X_T|Z_{T-1}, Z_T)}_{R_{T-1}(Z_{T-1})} P(Z_1)P(X_1|Z_1) \prod_{t=2}^{T-1} P(Z_t|Z_{t-1})P(X_t|Z_{t-1}, Z_t) \\
&= \sum_{Z_{1:T-1}} \underbrace{\Big(\sum_{Z_{T-1}} R_{T-1} P(Z_{T-1}|Z_{T-2})P(X_{T-1}|Z_{T-2}, Z_{T-1}) \Big)}_{R_{T-2}(Z_{T-2})} P(Z_1)P(X_1|Z_1) \prod_{i=2}^{T-2} P(Z_t|Z_{t-1})P(X_t|Z_{t-1}, Z_t) \\
&\cdots \\
&= \sum_{Z_{1:j}} \underbrace{\Big(\sum_{Z_j} R_j P(Z_j|Z_{j-1})P(X_j|Z_{j-1}, Z_j) \Big)}_{R_{j-1}(Z_{j-1})} P(Z_1)P(X_1|Z_1) \prod_{t=2}^{j-1} P(Z_t|Z_{t-1})P(X_t|Z_{t-1}, Z_t)
\end{aligned}
$$

Thus we decomposed $P(X_{1:T})$ as function of $P(Z_{j+1}|Z_j)$, $P(X_{j+1}|Z_j, Z_{j+1})$(or $P(X_1|Z_1)$) and $P(Z_1)$ in two ways. Then we set the parameter vector θ as

- $\pi = (\pi_1, \cdots, \pi_m)$, where $\pi_i = P(Z_1 = i)$
- $\phi_0 = \Big(b_{in}\Big)_{m\times k}$, where $b_{in} = P(X_1 = n|Z_1 = i)$
- $\phi = \big(b_{ijn}\big)_{m\times m\times k}$, where $b_{ijn} = P(X_t = n|Z_{t-1} = i, Z_t = j)$
- $T = \Big(t_{ij}\Big)_{m\times m}$, where $t_{ij} = P(Z_{t+1} = j|Z_{t=i})$
- $\theta = (\pi, \phi_0, \phi, T)$

And the two notations, S and R, have specific meanings. (If having priors θ, we may condition on it).

$$
\begin{aligned}
S_t(Z_t) &= P(X_1, \cdots, X_t, Z_t|\theta) \\
R_t(Z_t) &= P(X_T, \cdots, X_{t+1}|Z_t, \theta)
\end{aligned}
$$

They are easy to see by checking the first term and induction relations:

$$
\begin{aligned}
S_t(Z_t) &= \sum_{Z_{t-1}} S_{t-1} P(Z_t|Z_{t-1})P(X_t|Z_{t-1}, Z_t) \\
R_t(Z_t) &= \sum_{Z_{t+1}} R_{t+1} P(Z_{t+1}|Z_t)P(X_{t+1}|Z_t, Z_{t+1})
\end{aligned}
$$