

# CSE 250A: Assignment 6

Jiaxu Zhu A53094655

November 14, 2015

## 6.1

## 6.2 Inference in HMMs

(a)

$$\begin{aligned}
 P(S_t = i | S_{t+1} = j, o_1, o_2, \dots, o_T) &= P(S_t = i | S_{t+1} = j, o_1, o_2, \dots, o_t) \quad (d - \text{separation } I) \\
 &= \frac{P(S_{t+1} = j | S_t = i, o_1, o_2, \dots, o_t) P(S_t = i | o_1, o_2, \dots, o_t)}{P(S_{t+1} = j | o_1, o_2, \dots, o_t)} \\
 &= \frac{P(S_{t+1} = j | S_t = i) P(S_t = i, o_1, o_2, \dots, o_t)}{P(S_{t+1} = j, o_1, o_2, \dots, o_t)} \quad (d - \text{separation } II) \\
 &= \frac{P(S_{t+1} = j | S_t = i) P(S_t = i, o_1, o_2, \dots, o_t)}{\sum_k P(S_{t+1} = j, S_t = k, o_1, o_2, \dots, o_t)} \\
 &= \frac{P(S_{t+1} = j | S_t = i) P(S_t = i, o_1, o_2, \dots, o_t)}{\sum_k P(S_{t+1} = j | S_t = k) P(S_t = k, o_1, o_2, \dots, o_t)} \\
 &= \frac{a_{ij} \alpha_{it}}{\sum_k a_{kj} \alpha_{kt}}
 \end{aligned}$$

(b)

$$\begin{aligned}
 P(S_{t+1} = j, | S_t = i, o_1, o_2, \dots, o_T) &= P(S_{t+1} = j, | S_t = i, o_{t+1}, o_{t+2}, \dots, o_T) \quad (d - \text{separation } I \& II) \\
 &= \frac{P(o_{t+1}, o_{t+2}, \dots, o_T | S_{t+1} = j, S_t = i) P(S_{t+1} = j, | S_t = i)}{P(o_{t+1}, o_{t+2}, \dots, o_T, | S_t = i)} \\
 &= \frac{P(o_{t+1} | S_{t+1} = j) P(o_{t+2}, \dots, o_T | S_{t+1} = j) P(S_{t+1} = j, | S_t = i)}{P(o_{t+1}, o_{t+2}, \dots, o_T, | S_t = i)} \quad (d - \text{separation } I \& II) \\
 &= \frac{\beta_{j(t+1)} a_{ij} b_j(o_{t+1})}{\beta_{it}}
 \end{aligned}$$

(c)

$$\begin{aligned}
 P(S_{t+1} = k, | S_{t-1} = i, S_t = j, o_1, o_2, \dots, o_T) &= P(S_{t+1} = k, | S_t = j, o_{t+1}, o_{t+2}, \dots, o_T) \quad (d - \text{separation } I \& II) \\
 &= \frac{\beta_{k(t+1)} a_{jk} b_k(o_{t+1})}{\beta_{jt}}
 \end{aligned}$$

(d)

$$\begin{aligned}
 P(S_{t+1} = k, | S_{t-1} = i, o_1, o_2, \dots, o_T) &= \sum_j P(S_{t+1} = k, S_t = j | S_{t-1} = i, o_1, o_2, \dots, o_T) \\
 &= \sum_j P(S_{t+1} = k | S_t = j, S_{t-1} = i, o_1, o_2, \dots, o_T) P(S_t = j | S_{t-1} = i, o_1, o_2, \dots, o_T) \\
 &= \sum_j \frac{\beta_{k(t+1)} a_{jk} b_k(o_{t+1})}{\beta_{jt}} \frac{\beta_{jt} a_{ij} b_j(o_t)}{\beta_{i(t-1)}}
 \end{aligned}$$

### 6.3 Belief updating

(a)

$$\begin{aligned}
q_{jt} &= P(S_t = j | o_1, o_2, \dots, o_t) \\
&= \frac{P(S_t = j, o_1, o_2, \dots, o_t)}{P(o_1, o_2, \dots, o_t)} \quad (\text{product rule}) \\
&= \frac{\sum_i P(S_t = j, S_{t-1} = i, o_1, o_2, \dots, o_t)}{\sum_{ik} P(S_t = k, S_{t-1} = i, o_1, o_2, \dots, o_t)} \quad (\text{marginalization}) \\
&= \frac{\sum_i P(o_t | S_t = j) P(S_t = j | S_{t-1} = i) P(S_{t-1} = i | o_1, o_2, \dots, o_{t-1}) P(o_1, o_2, \dots, o_{t-1})}{\sum_{ik} P(o_t | S_t = k) P(S_t = k | S_{t-1} = i) P(S_{t-1} = i | o_1, o_2, \dots, o_{t-1}) P(o_1, o_2, \dots, o_{t-1})} \quad (\text{product + d - separation I}) \\
&= \frac{b_j(o_t) \sum_i a_{ij} q_{it-1}}{\sum_{ik} b_j(o_t) a_{ik} q_{it-1}}
\end{aligned}$$

(b)

$$\begin{aligned}
P(x_t | y_1, y_2, \dots, y_t) &= \frac{P(x_t, y_1, y_2, \dots, y_t)}{P(y_1, y_2, \dots, y_t)} \quad (\text{product rule}) \\
&= \frac{\int dx_{t-1} P(x_t, x_{t-1}, y_1, y_2, \dots, y_t)}{\int dx_t \int dx_{t-1} P(x_t, x_{t-1}, y_1, y_2, \dots, y_t)} \quad (\text{marginalization}) \\
&= \frac{\int dx_{t-1} P(y_t | x_t) P(x_t = j | x_{t-1}) P(x_{t-1} | y_1, y_2, \dots, y_{t-1}) P(y_1, y_2, \dots, y_{t-1})}{\int dx_t \int dx_{t-1} P(y_t | x_t) P(x_t = j | x_{t-1}) P(x_{t-1} | y_1, y_2, \dots, y_{t-1}) P(y_1, y_2, \dots, y_{t-1})} \quad (\text{product + d - separation I}) \\
&= \frac{P(y_t | x_t) \int dx_{t-1} P(x_t | x_{t-1}) P(x_{t-1} | y_1, y_2, \dots, y_{t-1})}{\int dx_t P(y_t | x_t) \int dx_{t-1} P(x_t | x_{t-1}) P(x_{t-1} | y_1, y_2, \dots, y_{t-1})}
\end{aligned}$$

The reason why this real-time updating is difficult for all but Gaussian random variables is that

### 6.4 Continuous density HMM

(a) The distribution  $P(X_t | S_{t-1})$  is a mixture of univariate Gaussians. It contains  $n$  mixture components because it can be written as

$$\begin{aligned}
P(X_t | S_{t-1}) &= \sum_{i=1}^n P(X_t, S_t = i | S_{t-1}) \\
&= \sum_{i=1}^n P(X_t | S_t = i) P(S_t = i | S_{t-1})
\end{aligned}$$

(b) The distribution  $P(X_t, X_{t'} | S_t, S_{t'})$  is not a mixture of univariate Gaussians. It is a product of Gaussians because it can be written as

$$\begin{aligned}
P(X_t, X_{t'} | S_t, S_{t'}) &= P(X_t | S_t, S_{t'}) P(X_{t'} | S_t, S_{t'}) \\
&= P(X_t | S_t) P(X_{t'} | S_{t'})
\end{aligned}$$

(c) The distribution  $P(X_1, X_2, \dots, X_T)$  is not a mixture of univariate Gaussians. It is a product of Gaussians because it can be written as

$$\begin{aligned}
P(X_1, X_2, \dots, X_T) &= \sum_{S'} P(S', X_1, X_2, \dots, X_T) \\
&= \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_T=1}^n P(S_1 = i_1) \prod_{t=2}^T P(S_t = i_t | S_{t-1} = i_{t-1}) \prod_{t=1}^T P(S_t = i_t | X_t)
\end{aligned}$$

(d) The distribution  $P(X_t|X_1, X_2, \dots, X_{t-1})$  is a mixture of univariate Gaussians. It contains  $n^2$  mixture components because it can be written as

$$\begin{aligned}
P(X_t|X_1, X_2, \dots, X_{t-1}) &= \sum_{i=1}^n \sum_{j=1}^n P(X_t, S_t = i, S_{t-1} = j|X_1, X_2, \dots, X_{t-1}) \\
&= \sum_{i=1}^n \sum_{j=1}^n P(X_t|S_t = i, S_{t-1} = j)P(S_t = i, S_{t-1} = j|X_1, X_2, \dots, X_{t-1}) \\
&= \sum_{i=1}^n \sum_{j=1}^n P(X_t|S_t = i)P(S_t = i|S_{t-1} = j)P(S_{t-1} = j|X_1, X_2, \dots, X_{t-1}) \\
&= \sum_{i=1}^n \sum_{j=1}^n P(X_t|S_t = i)P(S_t = i|S_{t-1} = j) \frac{P(X_{t-1}|S_{t-1} = j, X_1, X_2, \dots)}{den} P(S_{t-1} = j|X_1, X_2, \dots,
\end{aligned}$$

(e) The distribution  $P(X_t|X_1, X_2, \dots, X_{t-1})$  is a mixture of univariate Gaussians. It contains  $n^T$  mixture components because it can be written as

$$\begin{aligned}
P(X_T) &= \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_T=1}^n P(X_T, S_1, S_2, \dots, S_T) \\
&= \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_T=1}^n P(X_T|S_T = i_T)P(S_T = i_T|S_{T-1} = i_{T-1}) \dots P(S_2 = i_2|S_1 = i_1)P(S_1 = i_1)
\end{aligned}$$

(f) The distribution  $P(X_1, X_2, \dots, X_t|S_1, S_2, \dots, S_t)$  is not a mixture of univariate Gaussians. It is a product of Gaussians because it can be written as

$$\begin{aligned}
P(X_1, X_2, \dots, X_t|S_1, S_2, \dots, S_t) &= \frac{P(X_1, X_2, \dots, X_t, S_1, S_2, \dots, S_t)}{P(S_1, S_2, \dots, S_t)} \\
&= \frac{P(S_1) \prod_{t=2}^T P(S_t|S_{t-1}) \prod_{t=1}^T P(S_t|X_t)}{P(S_1) \prod_{t=2}^T P(S_t|S_{t-1})} \\
&= \prod_{t=1}^T P(S_t|X_t)
\end{aligned}$$

## 6.5 Mixture model decision boundary

(a)

$$\begin{aligned}
P(y = 1|\vec{x}) &= \frac{P(\vec{x}|y = 1)P(y = 1)}{P(\vec{x})} \\
&= \frac{P(\vec{x}|y = 1)P(y = 1)}{\sum_{i=0}^1 P(\vec{x}|y = i)P(y = i)} \\
&= \frac{\pi_1 (2\pi)^{-\frac{d}{2}} |\sum_1|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \sum_1^{-1}(\vec{x}-\vec{\mu}_1)}}{\sum_{i=0}^1 \pi_i (2\pi)^{-\frac{d}{2}} |\sum_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \sum_i^{-1}(\vec{x}-\vec{\mu}_i)}} \\
&= \frac{\pi_1 |\sum_1|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \sum_1^{-1}(\vec{x}-\vec{\mu}_1)}}{\sum_{i=0}^1 \pi_i |\sum_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \sum_i^{-1}(\vec{x}-\vec{\mu}_i)}}
\end{aligned}$$

(b)

$$\begin{aligned}
P(y = 1|\vec{x}) &= \frac{\pi_1 |\sum|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \Sigma^{-1}(\vec{x}-\vec{\mu}_1)}}{\sum_{i=0}^1 \pi_i |\sum|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma^{-1}(\vec{x}-\vec{\mu}_i)}} \\
&= \frac{1}{1 + \frac{\pi_0}{\pi_1} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_0)^T \Sigma^{-1}(\vec{x}-\vec{\mu}_0) + \frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \Sigma^{-1}(\vec{x}-\vec{\mu}_1)}}
\end{aligned}$$

Since  $(\vec{x} - \vec{\mu}_0)^T \Sigma^{-1}(\vec{x} - \vec{\mu}_0) = \sum_{i=1}^d \sum_{j=1}^d (x_i - \mu_{0i})(x_j - \mu_{0j}) \sum_{ij}$

$$P(y = 1|\vec{x}) = \frac{1}{1 + \frac{\pi_0}{\pi_1} e^{-\frac{1}{2}()}}$$