CSE 250A: Assignment 6

Jiaxu Zhu A53094655

November 16, 2015

6.1 Viterbi algorithm

(b) a plot of the most likely sequence of hidden states versus time is displayed in Fig. 1.

6.2 Inference in HMMs

(a)

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

(b)

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

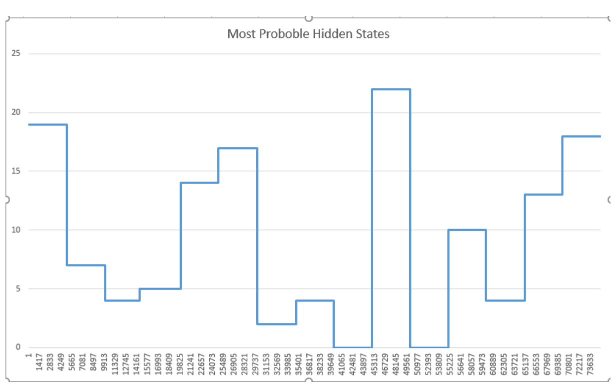

Figure 1: most likely sequence of hidden states

(c)

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

(d)

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

6.3 Belief updating

(a)

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

(b)

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

The reason why this real-time updating 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 a multivariate Gaussian. It is a product of two independent Gaussians because it can be written a

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

(c) The distribution $P(X_1, X_2, ..., X_T)$ is a multivariate Gaussian. It is a product of two independent Gaussians because it can be written as

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

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

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

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

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

(f) The distribution $P(X_1, X_2, ..., X_t|S_1, S_2, ..., S_t)$ is a multivariate Gaussian. It is a product of two independent Gaussians because it can be written a

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

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)

$$\begin{aligned}
P(y=1|\vec{x}) &= \frac{\pi_1 |\sum|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \sum^{-1} (\vec{x}-\vec{\mu}_1)}}{\sum_{i=0}^{1} \pi_i |\sum|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \sum^{-1} (\vec{x}-\vec{\mu}_i)}} \\
&= \frac{1}{1 + \frac{\pi_0}{\pi_1} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_0)^T \sum^{-1} (\vec{x}-\vec{\mu}_0) + \frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \sum^{-1} (\vec{x}-\vec{\mu}_1)}} \\
(\vec{x}-\vec{\mu}_0)^T \sum (\vec{x}-\vec{\mu}_0) - (\vec{x}-\vec{\mu}_1)^T \sum (\vec{x}-\vec{\mu}_1) &= \sum_{i=1}^{d} \sum_{j=1}^{d} [(x_i - \mu_{0i})(x_j - \mu_{0j}) - (x_i - \mu_{1i})(x_j - \mu_{1j})] \sum_{ij} \\
&= \sum_{i=1}^{d} \sum_{j=1}^{d} [(\mu_{1j} - \mu_{0j}) x_i + (\mu_{1i} - \mu_{0i}) x_j + \mu_{0i}\mu_{0j} - \mu_{1i}\mu_{1j}] \sum_{ij} \\
&= \sum_{i=1}^{d} [\sum_{j=1}^{d} (\sum_{ij} + \sum_{ji})(\mu_{1j} - \mu_{0j}) x_i] + \sum_{i=1}^{d} \sum_{j=1}^{d} (\mu_{0i}\mu_{0j} - \mu_{1i}\mu_{1j}) \sum_{ij} \\
w_i &= \frac{1}{2} \sum_{j=1}^{d} (\sum_{ij} + \sum_{ji})(\mu_{1j} - \mu_{0j}) \\
b &= \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} (\mu_{0i}\mu_{0j} - \mu_{1i}\mu_{1j}) \sum_{ij} + \log(\frac{\pi_0}{\pi_1})
\end{aligned}$$

(c)

$$\begin{aligned}
\frac{P(y=1|\vec{x})}{P(y=0|\vec{x})} &= k \\
P(y=1|\vec{x}) &= \frac{k}{k+1} \\
e^{-(\vec{w}\cdot\vec{x}+b)} &= \frac{1}{k} \\
\vec{w}\cdot\vec{x} &= \log(k) - b
\end{aligned}$$

That's the hyperplane.