

732A96/TDDE15 Advanced Machine Learning

Hidden Markov Models

Jose M. Peña
IDA, Linköping University, Sweden

Lecture 7: Autoregressive and Explicit-Duration Hidden Markov Models

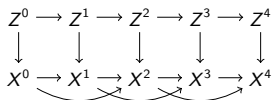
Contents

- ▶ Autoregressive Hidden Markov Models
 - ▶ Definition
 - ▶ Learning
 - ▶ Forward-Backward Algorithm
- ▶ Explicit-Duration Hidden Markov Models
 - ▶ Definition

- ▶ Main source
 - ▶ Chiappa, S. Explicit-Duration Markov Switching Models. *Foundations and Trends in Machine Learning* 7, 803-886, 2014. Sections 1-3.3.
- ▶ Additional source
 - ▶ Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer, 2006. Chapter 13.1-13.2.

Autoregressive Hidden Markov Models: Definition

- ▶ To overcome the poor modeling of long range correlations in HMMs, by allowing $Pa(X^t) \subseteq \{Z^t, X^{0:t}\}$.



- ▶ Hereinafter, we focus on the simplest AR-HMM, i.e. $Pa(X^t) = \{Z^t, X^{t-1}\}$.

Autoregressive Hidden Markov Models: Learning

- ▶ Almost the same as for HMMs.
- ▶ Recall that maximizing the log likelihood function over $x^{0:T}$ is inefficient (no closed form solution) and ineffective (multimodal).
- ▶ Consider maximizing its expectation

$$\begin{aligned} E[\log p(Z^{0:T}, x^{0:T})] &= \sum_{z^{0:T}} p(z^{0:T} | x^{0:T}) \log p(z^{0:T}, x^{0:T}) \\ &= \sum_{z^{0:T}} p(z^{0:T} | x^{0:T}) [\log \theta_{z^0} + \sum_{t=1}^{T-1} \log \theta_{z^{t+1}|z^t} + \sum_{t=1}^T \log \theta_{x^t|z^t, x^{t-1}}] \\ &= \sum_{z^0} p(z^0 | x^{0:T}) \log \theta_{z^0} + \sum_{t=1}^{T-1} \sum_{z^t} \sum_{z^{t+1}} p(z^t, z^{t+1} | x^{0:T}) \log \theta_{z^{t+1}|z^t} + \sum_{t=1}^T \sum_{z^t} p(z^t | x^{0:T}) \log \theta_{x^t|z^t, x^{t-1}} \end{aligned}$$

- ▶ Then

- ▶ $\theta_{z^0}^{ML} = \frac{p(z^0 | x^{0:T})}{\sum_{z^0} p(z^0 | x^{0:T})}$
- ▶ $\theta_{z^{t+1}|z^t}^{ML} = \frac{\sum_{t=1}^{T-1} p(z^t, z^{t+1} | x^{0:T})}{\sum_{t=1}^{T-1} \sum_{z^{t+1}} p(z^t, z^{t+1} | x^{0:T})}$
- ▶ $\theta_{x^t|z^t, x^{t-1}}^{ML} = \frac{\sum_{t=1}^T p(z^t | x^{0:T}) 1_{\{x^t, x^{t-1} \in x^{0:T}\}}}{\sum_{t=1}^T p(z^t | x^{0:T}) 1_{\{x^{t-1} \in x^{0:T}\}}}$

- ▶ Note that computing $p(z^0 | x^{0:T})$, $p(z^t, z^{t+1} | x^{0:T})$ and $p(z^t | x^{0:T})$ requires inference: Forward-backward algorithm.

Autoregressive Hidden Markov Models: Forward-Backward Algorithm

$$\begin{aligned}
 p(z^t | x^{0:T}) &= \frac{p(x^{0:t-1}, x^{t+1:T} | z^t, x^t) p(z^t, x^t)}{p(x^{0:T})} \\
 &= \frac{p(x^{0:t-1} | z^t, x^t) p(z^t, x^t) p(x^{t+1:T} | z^t, x^t)}{p(x^{0:T})} \text{ by } X^{0:t-1} \perp_p X^{t+1:T} | Z^t \cup X^t \\
 &= \frac{p(x^{0:t}, z^t) p(x^{t+1:T} | z^t, x^t)}{p(x^{0:T})} = \frac{\alpha(z^t) \beta(z^t)}{\sum_{z^t} \alpha(z^t) \beta(z^t)}
 \end{aligned}$$

$$\begin{aligned}
 p(z^t, z^{t+1} | x^{0:T}) &= \frac{p(x^{0:t-1}, x^{t+2:T} | z^t, z^{t+1}, x^t, x^{t+1}) p(z^t, z^{t+1}, x^t, x^{t+1})}{p(x^{0:T})} \\
 &= \frac{p(x^{0:t-1} | z^t, x^t) p(x^{t+2:T} | z^{t+1}, x^{t+1}) p(x^{t+1} | z^{t+1}, x^t) p(z^{t+1} | z^t) p(z^t, x^t)}{p(x^{0:T})} \\
 &\quad \text{by } X^{0:t-1} \perp_p X^{t+2:T} | Z^t \cup Z^{t+1} \cup X^t \cup X^{t+1} \\
 &\quad \quad X^{0:t-1} \perp_p Z^{t+1} \cup X^{t+1} | Z^t \cup X^t \\
 &\quad \quad X^{t+2:T} \perp_p Z^t \cup X^t | Z^{t+1} \cup X^{t+1} \\
 &\quad \quad X^{t+1} \perp_p Z^t | Z^{t+1} \cup X^t \\
 &\quad \quad Z^{t+1} \perp_p X^t | Z^t \\
 &= \frac{\alpha(z^t) \beta(z^{t+1}) p(x^{t+1} | z^{t+1}, x^t) p(z^{t+1} | z^t)}{\sum_{z^t} \sum_{z^{t+1}} \alpha(z^t) \beta(z^{t+1}) p(x^{t+1} | z^{t+1}, x^t) p(z^{t+1} | z^t)}
 \end{aligned}$$

Autoregressive Hidden Markov Models: Forward-Backward Algorithm

$$\begin{aligned}
 \alpha(\mathbf{z}^t) &= p(x^t | z^t, x^{t-1}) p(z^t, x^{t-1}) p(x^{0:t-2} | z^t, x^{t-1}) \text{ by } X^{0:t-2} \perp_p X^t | Z^t \cup X^{t-1} \\
 &= p(x^t | z^t, x^{t-1}) p(x^{0:t-1}, z^t) = p(x^t | z^t, x^{t-1}) \sum_{z^{t-1}} p(x^{0:t-1}, z^t | z^{t-1}) p(z^{t-1}) \\
 &= p(x^t | z^t, x^{t-1}) \sum_{z^{t-1}} p(x^{0:t-1} | z^{t-1}) p(z^t | z^{t-1}) p(z^{t-1}) \text{ by } X^{0:t-1} \perp_p Z^t | Z^{t-1} \\
 &= p(x^t | z^t, x^{t-1}) \sum_{z^{t-1}} p(x^{0:t-1}, z^{t-1}) p(z^t | z^{t-1}) \\
 &= p(x^t | z^t, x^{t-1}) \sum_{z^{t-1}} \alpha(\mathbf{z}^{t-1}) p(z^t | z^{t-1}) \text{ with } \alpha(\mathbf{z}^0) = p(x^0 | z^0) p(z^0)
 \end{aligned}$$

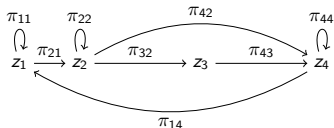
$$\begin{aligned}
 \beta(\mathbf{z}^t) &= \sum_{z^{t+1}} p(x^{t+1:T}, z^{t+1} | z^t, x^t) = \sum_{z^{t+1}} p(x^{t+1:T} | z^{t+1}, z^t, x^t) p(z^{t+1} | z^t, x^t) \\
 &= \sum_{z^{t+1}} p(x^{t+1:T} | z^{t+1}, x^t) p(z^{t+1} | z^t) \text{ by } X^{t+1:T} \perp_p Z^t | Z^{t+1} \cup X^t \text{ and } X^t \perp_p Z^{t+1} | Z^t \\
 &= \sum_{z^{t+1}} p(x^{t+2:T} | z^{t+1}, x^{t+1}) p(x^{t+1} | z^{t+1}, x^t) p(z^{t+1} | z^t) \text{ by } X^{t+2:T} \perp_p X^t | Z^{t+1} \cup X^{t+1} \\
 &= \sum_{z^{t+1}} \beta(\mathbf{z}^{t+1}) p(x^{t+1} | z^{t+1}, x^t) p(z^{t+1} | z^t)
 \end{aligned}$$

$$\beta(\mathbf{z}^T) = 1 \text{ by } p(z^T | x^{0:T}) = \frac{\alpha(\mathbf{z}^T) \beta(\mathbf{z}^T)}{p(x^{0:T})} = p(z^T | x^{0:T}) \beta(\mathbf{z}^T)$$

► **Exercise.** Derive the Viterbi algorithm for AR-HMMs.

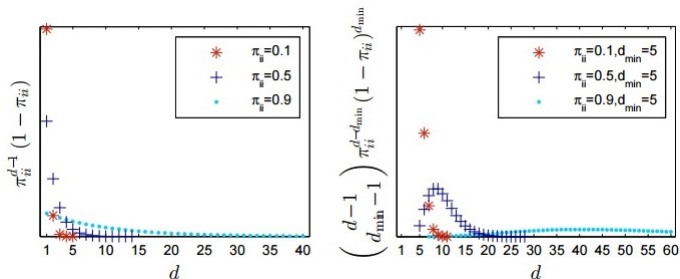
Explicit-Duration Hidden Markov Models: Definition

- ▶ To control the distribution of the number of time steps for which the HMM remains in a given state.
- ▶ Consider the following Markov chain over the states:



- ▶ The probability of remaining in state z_i for exactly $d - 1$ time steps is

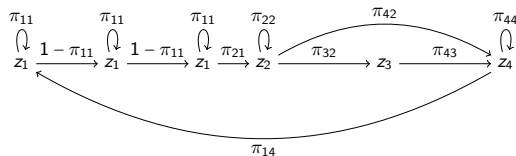
$$p(Z^{t+1:t+d-1} = z_i, Z^{t+d} \neq z_i | Z^t = z_i) = \pi_{ii}^{d-1} (1 - \pi_{ii}) = \text{Geometric}(d; 1 - \pi_{ii})$$



Source: Chiappa (2014).

Explicit-Duration Hidden Markov Models: Definition

- ▶ We can obtain a negative binomial distribution by imposing a minimum duration on the time spent in a state, and duplicating the corresponding observation model. For instance, when $d_{min}=3$



- ▶ The probability of remaining in state z_i for exactly $d - 1$ time steps is

$$\begin{aligned}
 p(Z^{t+1:t+d-1} = z_i, Z^{t+d} \neq z_i | Z^t = z_i) &= \binom{d-1}{d_{min}-1} \pi_{ii}^{d-d_{min}} (1 - \pi_{ii})^{d_{min}} \\
 &= \text{NegativeBinomial}(d; d_{min}, 1 - \pi_{ii})
 \end{aligned}$$

Explicit-Duration Hidden Markov Models: Decreasing

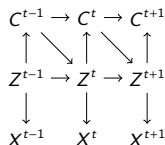
- ▶ The geometric and negative binomial cases define a segment duration distribution **implicitly**. We now define the segment duration distribution **explicitly**, by introducing auxiliary unobserved random variables.
- ▶ Specifically, let us use the random variables $C^{1:T}$ which take **decreasing** values within a segment, starting with the duration of the segment and ending with 1. Then

$$p(z_i^t | z_j^{t-1}, c^{t-1}) = \begin{cases} \pi_{ij} & \text{if } c^{t-1} = 1 \\ 1_{i=j} & \text{if } c^{t-1} > 1 \end{cases}$$

$$p(c^t | z^t, c^{t-1}) = \begin{cases} \rho(z^t, c^t) & \text{if } c^{t-1} = 1 \\ 1_{c^t=c^{t-1}-1} & \text{if } c^{t-1} > 1 \end{cases}$$

where $\rho(z^t, c^t)$ is the segment duration distribution.

- ▶ In graphical terms, we have



Explicit-Duration Hidden Markov Models: Increasing

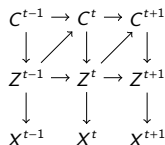
- ▶ Alternatively, we can use the random variables $C^{1:T}$ which take **increasing** values within a segment, starting 1 and ending with the duration of the segment. Then

$$p(z_i^t | z_j^{t-1}, c^t) = \begin{cases} \pi_{ij} & \text{if } c^t = 1 \\ 1_{i=j} & \text{if } c^t > 1 \end{cases}$$

$$p(c^t | z^{t-1}, c^{t-1}) = \begin{cases} 1 - \lambda(z^{t-1}, c^{t-1}) & \text{if } c^t = 1 \\ \lambda(z^{t-1}, c^{t-1}) & \text{if } c^t = c^{t-1} + 1 \end{cases}$$

where $\lambda(z^{t-1}, c^{t-1})$ is the probability of the segment continuing.

- ▶ In graphical terms, we have



Explicit-Duration Hidden Markov Models: Decreasing-Duration

- Alternatively, we can use the random variables $C^{1:T}$ which take **decreasing** values within a segment, starting with the duration of the segment and ending with 1. And the random variables $D^{1:T}$ which indicate the **duration** of the current segment. Then

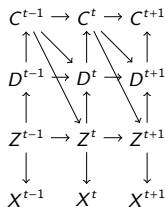
$$p(z_i^t | z_j^{t-1}, c^{t-1}) = \begin{cases} \pi_{ij} & \text{if } c^{t-1} = 1 \\ 1_{i=j} & \text{if } c^{t-1} > 1 \end{cases}$$

$$p(c^t | c^{t-1}, d^t) = \begin{cases} 1_{c^t=d^t} & \text{if } c^{t-1} = 1 \\ 1_{c^t=c^{t-1}-1} & \text{if } c^{t-1} > 1 \end{cases}$$

$$p(d^t | z^t, c^{t-1}, d^{t-1}) = \begin{cases} \rho(z^t, d^t) & \text{if } c^{t-1} = 1 \\ 1_{d^t=d^{t-1}} & \text{if } c^{t-1} > 1 \end{cases}$$

where $\rho(z^t, d^t)$ is the segment duration distribution.

- In graphical terms, we have



Explicit-Duration Hidden Markov Models: Increasing-Duration

- ▶ **Exercise.** Derive the transition model for the increasing-duration case.
- ▶ Relatively easy to adapt the FB and Viterbi algorithms to explicit-duration HMMs.

Contents

- ▶ Autoregressive Hidden Markov Models
 - ▶ Definition
 - ▶ Learning
 - ▶ Forward-Backward Algorithm
- ▶ Explicit-Duration Hidden Markov Models
 - ▶ Definition

Thank you