PATTERN RECOGNITION AND MACHINE LEARNING

CHAPTER 13: SEQUENTIAL DATA

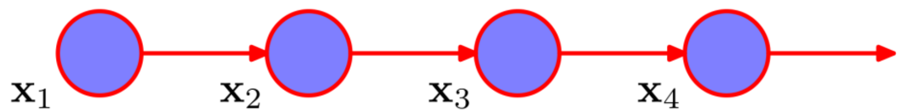
Outlines

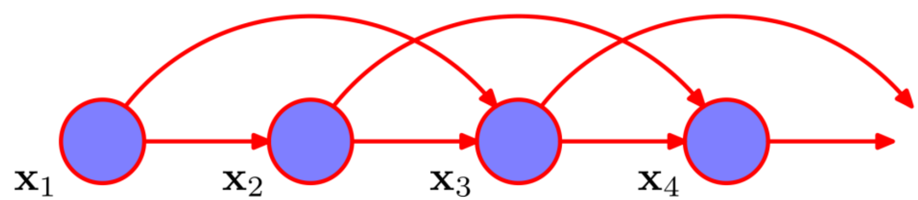
- Hidden Markov Models
- Maximum Likelihood and EM for HMM
- Forward-Backward and Sum-Product
- Viterbi Algorithm
- Linear Dynamics Systems
- Kalman and Particle Filters

Markov Models

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_N) = p(\mathbf{x}_1) \prod_{n=2}^{N} p(\mathbf{x}_n | \mathbf{x}_{n-1})$$

$$p(\mathbf{x}_n | \mathbf{x}_1, \ldots, \mathbf{x}_{n-1}) = p(\mathbf{x}_n | \mathbf{x}_{n-1})$$





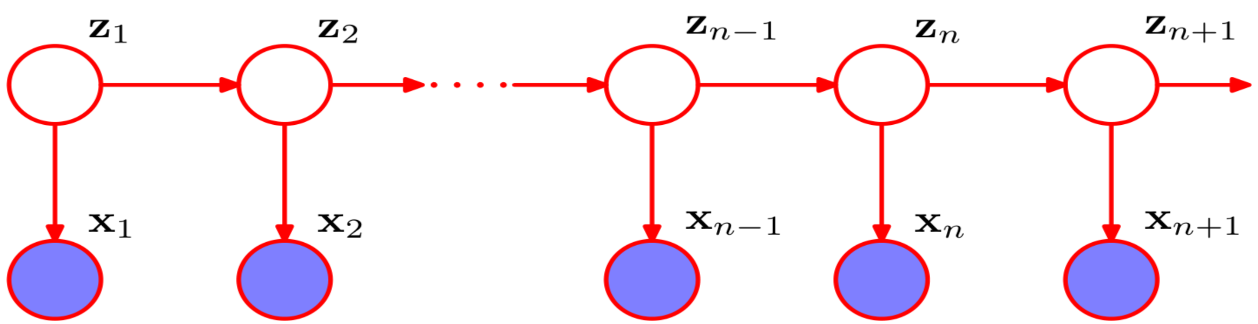
A second-order Markov chain

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_N) = p(\mathbf{x}_1)p(\mathbf{x}_2|\mathbf{x}_1) \prod_{n=3}^{N} p(\mathbf{x}_n | \mathbf{x}_{n-1}, \mathbf{x}_{n-2})$$

Markov Models

Using a Markov chain of latent variables

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{z}_1, \ldots, \mathbf{z}_N) = p(\mathbf{z}_1) \left[\prod_{n=2}^{N} p(\mathbf{z}_n | \mathbf{z}_{n-1}) \right] \prod_{n=1}^{N} p(\mathbf{x}_n | \mathbf{z}_n)$$



For continuous variables, we can use linear-Gaussian conditional distributions

Hidden Markov Models

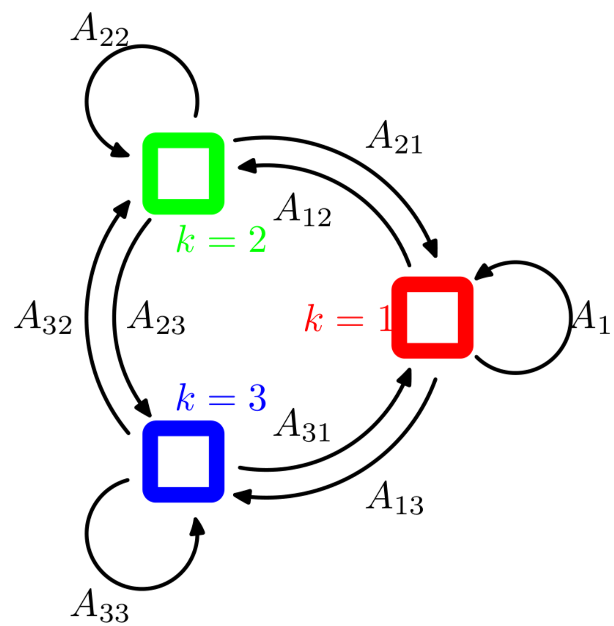
Conditional distribution for latent variable

$$p(\mathbf{z}_n|\mathbf{z}_{n-1,\mathbf{A}}) = \prod_{k=1}^{K}\prod_{j=1}^{K} A_{jk}^{z_{n-1,j}z_{nk}}$$

$$p(\mathbf{z}_1|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{z_{1k}}$$

$$\sum_k \pi_k = 1$$

A means transition probabilities



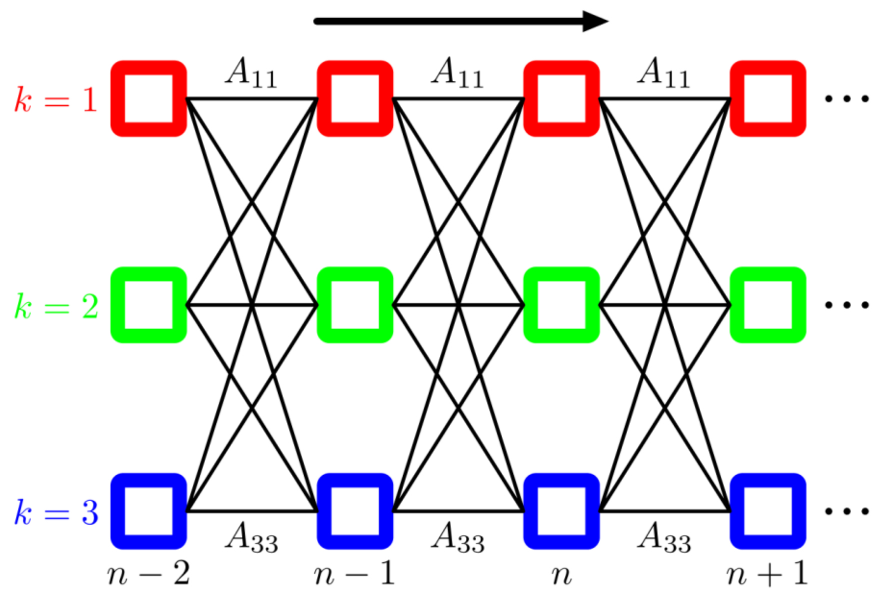
As in the case of a standard mixture model, the latent variables are the discrete multinomial variables zn using 1-of-K coding scheme

a model whose latent variables have three possible states corresponding to the three boxes. The black lines denote the elements of the transition matrix Ajk

Hidden Markov Models

Latent states lattice:
Representing the transitions between latent states

Each column of this diagram corresponds to one of the latent variables zn



Hidden Markov Models

Getting emission probabilities from latent variable

$$p(\mathbf{x}_n|\mathbf{z}_n, \phi) = \prod_{k=1}^{K} p(\mathbf{x}_n|\phi_k)^{z_{nk}}$$

$$p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) = p(\mathbf{z}_1|\boldsymbol{\pi}) \left[\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A})\right] \prod_{m=1}^{N} p(\mathbf{x}_m|\mathbf{z}_m, \phi)$$

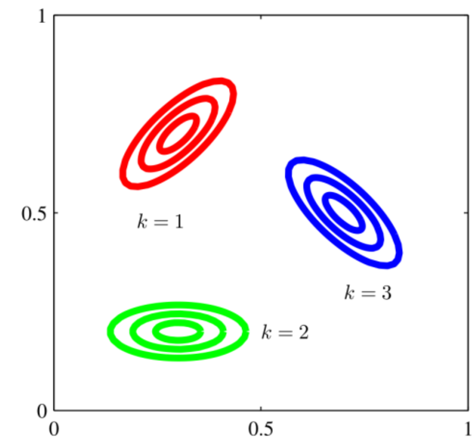
$$\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$$

$$\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$$

$$\boldsymbol{\theta} = \{\boldsymbol{\pi}, \mathbf{A}, \phi\}$$

Hidden Markov Models

a better understanding of the hidden Markov model

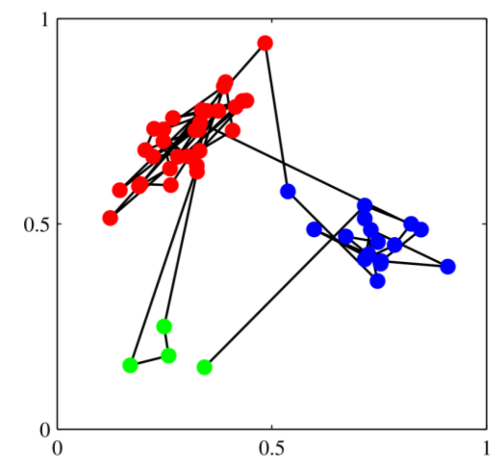


First choose the initial latent variable $\mathbf{z}_1$ $\mathbf{x}_1$ π_k

choose the state of the variable $\mathbf{z}_2$ $p(\mathbf{z}_2|\mathbf{z}_1)$

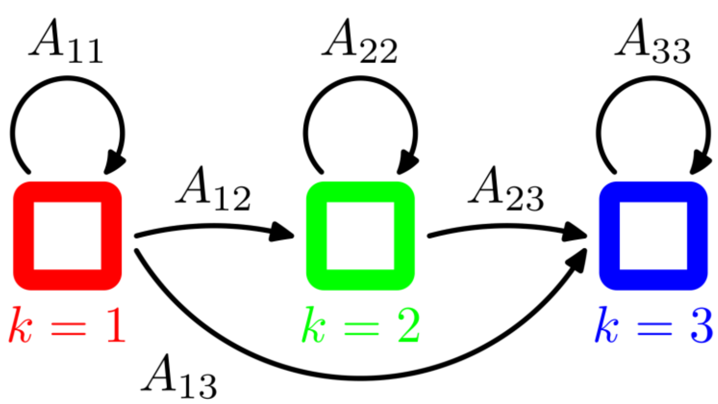
choose the state k A_{jk}

$k = 1, \ldots, K$

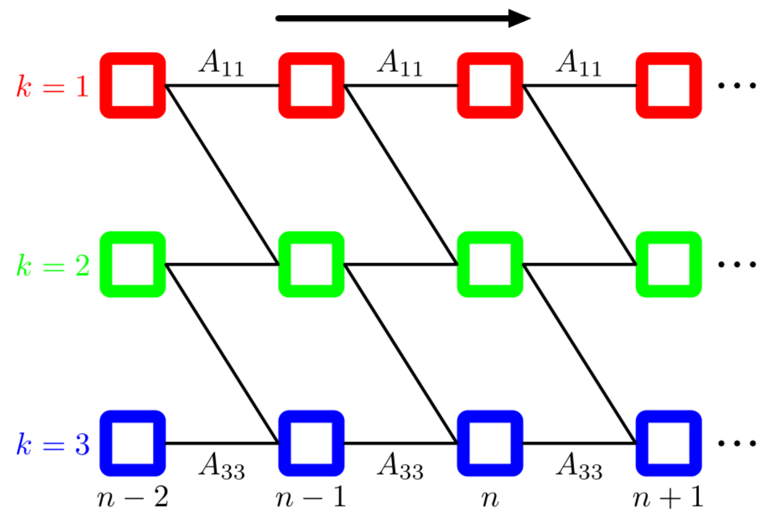


Hidden Markov Models

One Example of variants of the standard HMM model



3-state left-to-right hidden Markov model



Lattice diagram for a 3-state left- to-right HMM

Outlines

- Hidden Markov Models
- Maximum Likelihood and EM for HMM
- Forward-Backward and Sum-Product
- Viterbi Algorithm
- Linear Dynamics Systems
- Kalman and Particle Filters

Maximum Likelihood of HMM

Maximum likelihood for the HMM

$$p(\mathbf{X}|\boldsymbol{\theta}) = \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$$

EM algorithm to find an efficient framework

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$$

$$\Downarrow$$

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) \quad = \quad & \sum_{k=1}^{K} \gamma(z_{1k}) \ln \pi_k + \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \ln A_{jk} \\ & + \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \ln p(\mathbf{x}_n|\boldsymbol{\phi}_k). \qquad (\end{aligned}$$

EM of HMM

EM algorithm

$$\gamma(\mathbf{z}_n) = p(\mathbf{z}_n|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$$

$$\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) = p(\mathbf{z}_{n-1}, \mathbf{z}_n|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$$

$$\gamma(z_{nk}) = \mathbb{E}[z_{nk}] = \sum \gamma(\mathbf{z}) z_{nk}$$

$$\xi(z_{n-1,j}, z_{nk}) = \mathbb{E}[z_{n-1,j} z_{nk}] = \sum_{\mathbf{z}} \gamma(\mathbf{z}) z_{n-1,j} z_{nk}$$

$$\pi_k = \frac{\gamma(z_{1k})}{\sum_{j=1}^{K} \gamma(z_{1j})}$$

$$A_{jk} = \frac{\sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nk})}{\sum_{l=1}^{K} \sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nl})}$$

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

$$\boldsymbol{\Sigma}_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}}}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

Outlines

- Hidden Markov Models
- Maximum Likelihood and EM for HMM
- Forward-Backward and Sum-Product
- Viterbi Algorithm
- Linear Dynamics Systems
- Kalman and Particle Filters

Forward-Backward Algorithm

The forward-backward algorithm

The method of evaluating the quantities of $\gamma(z_{nk})$ $\xi(z_{n-1,j}, z_{nk})$

$$\gamma(\mathbf{z}_n) = p(\mathbf{z}_n|\mathbf{X}) = \frac{p(\mathbf{X}|\mathbf{z}_n)p(\mathbf{z}_n)}{p(\mathbf{X})}$$

$$= \frac{p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n)p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n)}{p(\mathbf{X})} = \frac{\alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)}{p(\mathbf{X})}$$

$$\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) = p(\mathbf{z}_{n-1}, \mathbf{z}_n|\mathbf{X})$$

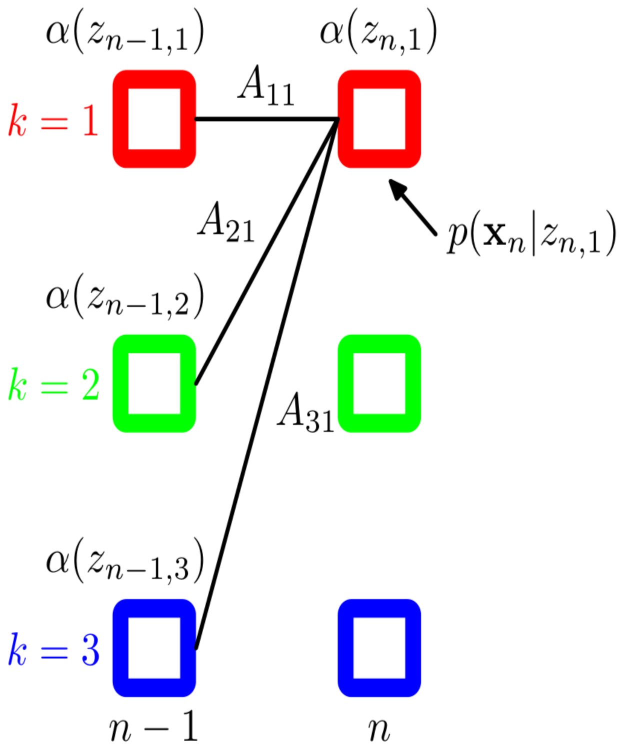
$$= \frac{\alpha(\mathbf{z}_{n-1})p(\mathbf{x}_n|\mathbf{z}_n)p(\mathbf{z}_n|\mathbf{z}_{n-1})\beta(\mathbf{z}_n)}{p(\mathbf{X})}$$

$$\begin{aligned}\alpha(\mathbf{z}_n) &\equiv p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n)\\ \beta(\mathbf{z}_n) &\equiv p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n)\end{aligned}$$

Forward Recursion

Forward recursion

$$\begin{aligned}
p(\mathbf{X}|\mathbf{z}_n) &= p(\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{z}_n) \\
&\quad\; p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n) \\
p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{x}_n,\mathbf{z}_n) &= p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_n) \\
p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_{n-1},\mathbf{z}_n) &= p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_{n-1}) \\
p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n,\mathbf{z}_{n+1}) &= p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_{n+1}) \\
p(\mathbf{x}_{n+2},\ldots,\mathbf{x}_N|\mathbf{z}_{n+1},\mathbf{x}_{n+1}) &= p(\mathbf{x}_{n+2},\ldots,\mathbf{x}_N|\mathbf{z}_{n+1}) \\
p(\mathbf{X}|\mathbf{z}_{n-1},\mathbf{z}_n) &= p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_{n-1}) \\
&\quad\; p(\mathbf{x}_n|\mathbf{z}_n)p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n) \\
p(\mathbf{x}_{N+1}|\mathbf{X},\mathbf{z}_{N+1}) &= p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \\
p(\mathbf{z}_{N+1}|\mathbf{z}_N,\mathbf{X}) &= p(\mathbf{z}_{N+1}|\mathbf{z}_N)
\end{aligned}$$



$$\alpha(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n)\sum_{\mathbf{z}_{n-1}}\alpha(\mathbf{z}_{n-1})p(\mathbf{z}_n|\mathbf{z}_{n-1})$$

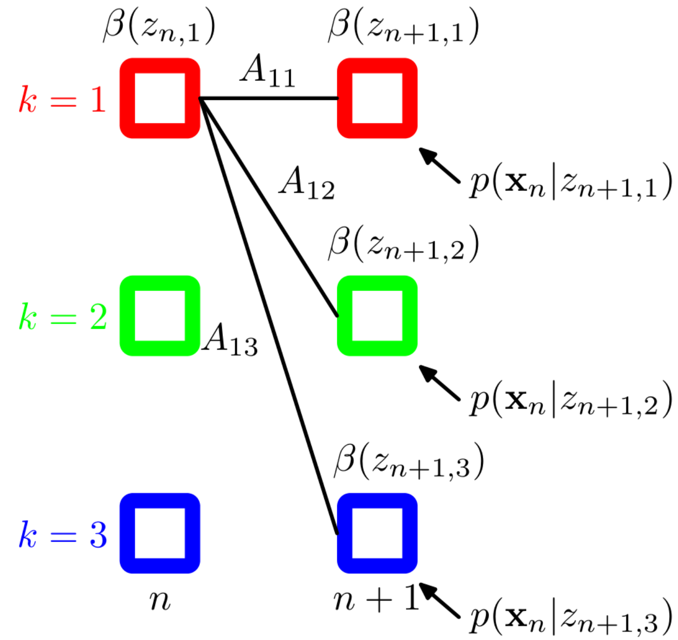
Backward Recursion

Backward recursion

$$\beta(\mathbf{z}_n) = p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)$$

$$= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+2},\ldots,\mathbf{x}_N|\mathbf{z}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n)$$

$$= \sum_{\mathbf{z}_{n+1}} \beta(\mathbf{z}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n)$$



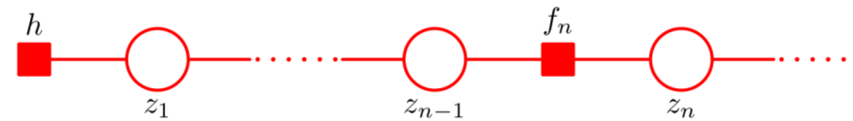
$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{n}\gamma(z_{nk})\mathbf{x}_n}{\sum_{n=1}^{n}\gamma(z_{nk})} = \frac{\sum_{n=1}^{n}\alpha(z_{nk})\beta(z_{nk})\mathbf{x}_n}{\sum_{n=1}^{n}\alpha(z_{nk})\beta(z_{nk})}$$

Sum-Product Algorithm

The sum-product algorithm for the HMM

transforming the directed into a factor graph

$$
\begin{aligned}
h(\mathbf{z}_1) &= p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1) \\
f_n(\mathbf{z}_{n-1}, \mathbf{z}_n) &= p(\mathbf{z}_n|\mathbf{z}_{n-1})p(\mathbf{x}_n|\mathbf{z}_n)
\end{aligned}
$$



f → z messages

$$\mu_{f_n \to \mathbf{z}_n}(\mathbf{z}_n) = \sum_{\mathbf{z}_{n-1}} f_n(\mathbf{z}_{n-1}, \mathbf{z}_n)\mu_{f_{n-1} \to \mathbf{z}_{n-1}}(\mathbf{z}_{n-1})$$

messages that are propagated from the root node back to the leaf node

$$\mu_{f_{n+1} \to f_n}(\mathbf{z}_n) = \sum_{\mathbf{z}_{n+1}} f_{n+1}(\mathbf{z}_n, \mathbf{z}_{n+1})\mu_{f_{n+2} \to f_{n+1}}(\mathbf{z}_{n+1})$$

$$p(\mathbf{z}_n, \mathbf{X}) = \mu_{f_n \to \mathbf{z}_n}(\mathbf{z}_n)\mu_{f_{n+1} \to \mathbf{z}_n}(\mathbf{z}_n) = \alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)$$

$$\gamma(\mathbf{z}_n) = \frac{p(\mathbf{z}_n, \mathbf{X})}{p(\mathbf{X})} = \frac{\alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)}{p(\mathbf{X})}$$

Scaling of HMM

Scaling factors

$$\alpha(\mathbf{z}_n) = p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n)$$

$$\widehat{\alpha}(\mathbf{z}_n) = p(\mathbf{z}_n|\mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{\alpha(\mathbf{z}_n)}{p(\mathbf{x}_1, \ldots, \mathbf{x}_n)}$$

scaling factors defined by conditional distributions over the observed variables

$$c_n = p(\mathbf{x}_n|\mathbf{x}_1, \ldots, \mathbf{x}_{n-1})$$

$$c_n\widehat{\alpha}(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n)\sum_{\mathbf{z}_{n-1}}\widehat{\alpha}(\mathbf{z}_{n-1})p(\mathbf{z}_n|\mathbf{z}_{n-1})$$

similarly define re-scaled variables

$$\widehat{\beta}(\mathbf{z}_n) = \frac{p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n)}{p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{x}_1, \ldots, \mathbf{x}_n)}$$

$$c_{n+1}\widehat{\beta}(\mathbf{z}_n) = \sum_{\mathbf{z}_{n+1}}\widehat{\beta}(\mathbf{z}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n)$$

$$\begin{aligned}\gamma(\mathbf{z}_n) &= \widehat{\alpha}(\mathbf{z}_n)\widehat{\beta}(\mathbf{z}_n)\\ \xi(\mathbf{z}_{n-1}, \mathbf{z}_n) &= c_n\widehat{\alpha}(\mathbf{z}_{n-1})p(\mathbf{x}_n|\mathbf{z}_n)p(\mathbf{z}_n|\mathbf{z}_{-1})\widehat{\beta}(\mathbf{z}_n)\end{aligned}$$

Outlines

- Hidden Markov Models
- Maximum Likelihood and EM for HMM
- Forward-Backward and Sum-Product
- Viterbi Algorithm
- Linear Dynamics Systems
- Kalman and Particle Filters

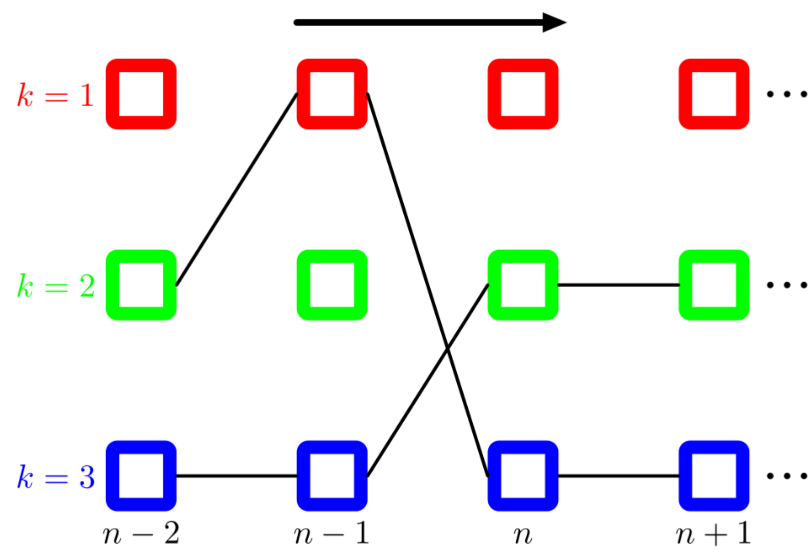
Viterbi Algorithm

The Viterbi algorithm

To find the most probable sequence of hidden states for a given observation sequence

$$\mu_{\mathbf{z}_n \to f_{n+1}}(\mathbf{z}_n) = \mu_{f_n \to \mathbf{z}_n}(\mathbf{z}_n)$$

$$\mu_{f_{n+1} \to \mathbf{z}_{n+1}}(\mathbf{z}_{n+1}) = \max_{\mathbf{z}_n} \left\{ \ln f_{n+1}(\mathbf{z}_n, \mathbf{z}_{n+1}) + \mu_{\mathbf{z}_n \to f_{n+1}}(\mathbf{z}_n) \right\}$$



f → z messages

$$\omega(\mathbf{z}_{n+1}) = \ln p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1}) + \max_{\mathbf{z}_n} \{\ln p(\mathbf{x}_{+1}|\mathbf{z}_n) + \omega(\mathbf{z}_n)\}$$

$$\omega(\mathbf{z}_1) = \ln p(\mathbf{z}_1) + \ln p(\mathbf{x}_1|\mathbf{z}_1)$$

$$\omega(\mathbf{z}_n) = \max_{\mathbf{z}_1,\ldots,\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_1, \ldots, \mathbf{z}_n)$$

Extensions of Hidden Markov Models

Extensions of the hidden Markov model

Using discriminative rather than maximum likelihood techniques

optimize the cross-entropy

$$\sum_{r=1}^{R} \ln p(m_r|\mathbf{X}_r) \qquad \sum_{r=1}^{R} \ln \left\{ \frac{p(\mathbf{X}_r|\boldsymbol{\theta}_r)p(m_r)}{\sum_{l=1}^{M} p(\mathbf{X}_r|\boldsymbol{\theta}_l)p(l_r)} \right\}$$

weakness of the hidden Markov model:
distribution of times for which the system remains in a given state

HMM is poor at capturing long-rangecorrelations between the observed variables

autoregressive HMM appears as a natural extension of the standard HMM when viewed as a graphical model

Outlines

- Hidden Markov Models
- Maximum Likelihood and EM for HMM
- Forward-Backward and Sum-Product
- Viterbi Algorithm
- Linear Dynamics Systems
- Kalman and Particle Filters

A Stochastic Linear Dynamical System

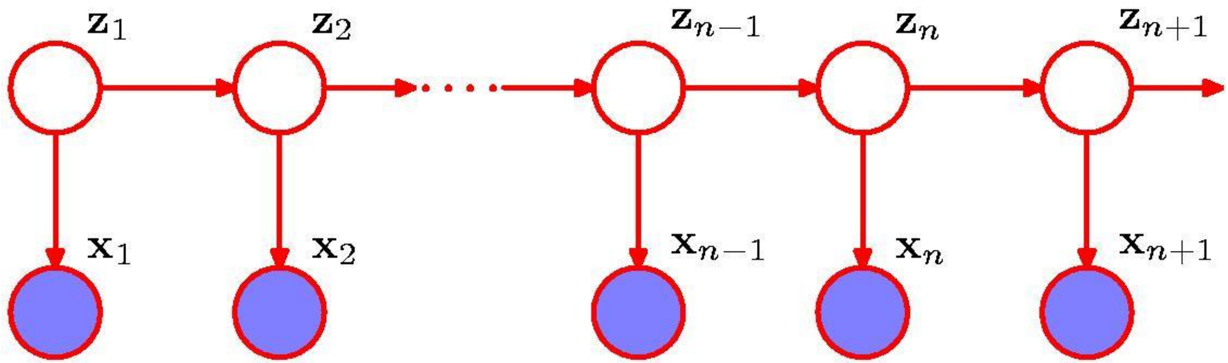


Fig. 13.15

$$\mathbf{z}_n = \mathbf{A}\mathbf{z}_{n-1} + \mathbf{w}_n \quad (13.78) \qquad \mathbf{w} \sim \mathcal{N}(\mathbf{w}|\mathbf{0},\mathbf{\Gamma}) \quad (13.81)$$

$$\mathbf{x}_n = \mathbf{C}\mathbf{z}_n + \mathbf{v}_n \quad (13.79) \qquad \mathbf{v} \sim \mathcal{N}(\mathbf{v}|\mathbf{0},\mathbf{\Sigma}) \quad (13.82)$$

$$\mathbf{z}_1 = \boldsymbol{\mu}_0 + \mathbf{u} \quad (13.80) \qquad \mathbf{u} \sim \mathcal{N}(\mathbf{u}|\mathbf{0},\mathbf{V}_0). \quad (13.83)$$

$$p(\mathbf{z}_n|\mathbf{z}_{n-1}) = \mathcal{N}(\mathbf{z}_n|\mathbf{A}\mathbf{z}_{n-1},\mathbf{\Gamma}) \quad (13.75)$$

$$p(\mathbf{x}_n|\mathbf{z}_n) = \mathcal{N}(\mathbf{x}_n|\mathbf{C}\mathbf{z}_n,\mathbf{\Sigma}). \quad (13.76)$$

$$p(\mathbf{z}_1) = \mathcal{N}(\mathbf{z}_1|\boldsymbol{\mu}_0,\mathbf{V}_0). \quad (13.77)$$

$$\theta = \{\mathbf{A},\mathbf{\Gamma},\mathbf{C},\mathbf{\Sigma},\boldsymbol{\mu}_0,\mathbf{V}_0\}$$

Inference Problem

- Finding the marginal distributions for the latent variables conditional on the observation sequence.

$$\widehat{\alpha}(\mathbf{z}_n) \equiv p(\mathbf{z}_n | \mathbf{x}_1, \ldots, \mathbf{x}_n)$$

$$\gamma(\mathbf{z}_n) \equiv p(\mathbf{z}_n | \mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}, \ldots, \mathbf{x}_N)$$

Outlines

- Hidden Markov Models
- Maximum Likelihood and EM for HMM
- Forward-Backward and Sum-Product
- Viterbi Algorithm
- Linear Dynamics Systems
- Kalman and Particle Filters

An Example Application: Tracking an Moving Object

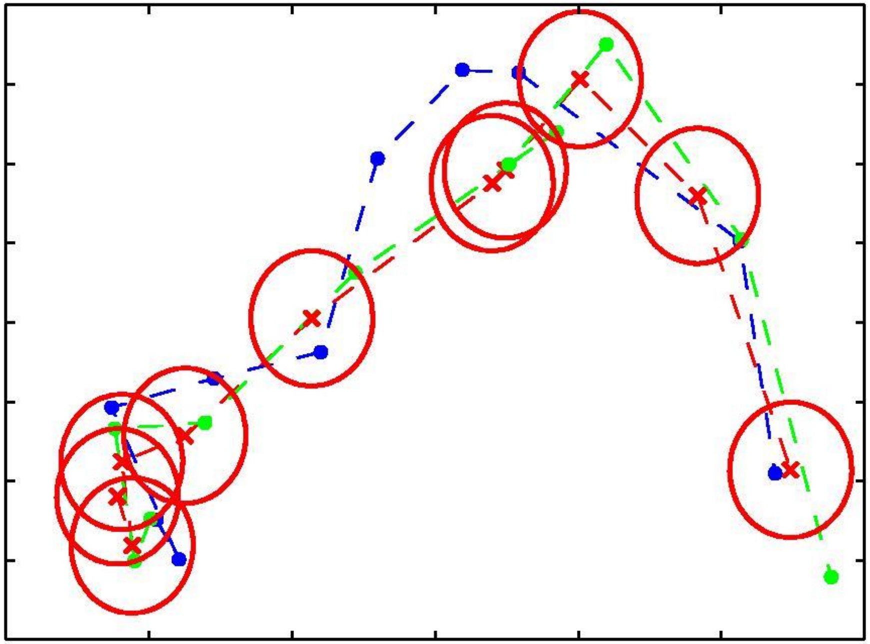


Fig. 13.22 An illustration of a linear dynamical system being used to track a moving object.

Blue: Z_n.
Green: X_n.
Red: The inferred.

One of the most important application of the Kalman filter.

Mean and Variance

Kalman filter equations

$$\boldsymbol{\mu}_n = \mathbf{A}\boldsymbol{\mu}_{n-1} + \mathbf{K}_n(\mathbf{x}_n - \mathbf{C}\mathbf{A}\boldsymbol{\mu}_{n-1}) \quad (13.89)$$

$$\mathbf{V}_n = (\mathbf{I} - \mathbf{K}_n\mathbf{C})\mathbf{P}_{n-1} \quad (13.90)$$

$$\mathbf{K}_n = \mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}}\left(\mathbf{C}\mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}} + \boldsymbol{\Sigma}\right)^{-1}. \quad (13.92)$$

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \mathbf{K}_1(\mathbf{x}_1 - \mathbf{C}\boldsymbol{\mu}_0) \quad (13.94)$$

$$\mathbf{V}_1 = (\mathbf{I} - \mathbf{K}_1\mathbf{C})\mathbf{V}_0 \quad (13.95)$$

$$\mathbf{K}_1 = \mathbf{V}_0\mathbf{C}^{\mathrm{T}}\left(\mathbf{C}\mathbf{V}_0\mathbf{C}^{\mathrm{T}} + \boldsymbol{\Sigma}\right)^{-1}. \quad (13.97)$$

$$\mathbf{P}_{n-1} = \mathbf{A}\mathbf{V}_{n-1}\mathbf{A}^{\mathrm{T}} + \boldsymbol{\Gamma}. \quad (13.88)$$

Interpretation of the Steps Involved

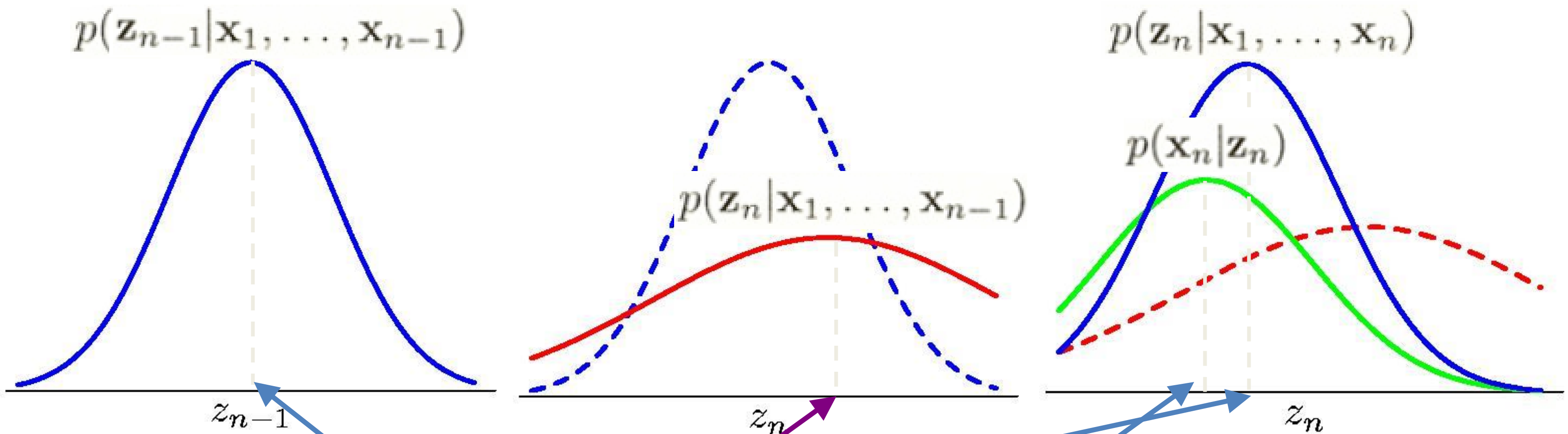


Fig. 13.21

$$\boldsymbol{\mu}_n = \mathbf{A}\boldsymbol{\mu}_{n-1} + \mathbf{K}_n(\mathbf{x}_n - \mathbf{C}\mathbf{A}\boldsymbol{\mu}_{n-1}) \tag{13.89}$$

Kalman filter as a process of

Making successive *predictions* and then

Correcting the predictions using the new observations.

Derivation of Eq.13.89, 13.90, 13.94, 13.95

$$c_n \widehat{\alpha}(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n) \int \widehat{\alpha}(\mathbf{z}_{n-1}) p(\mathbf{z}_n|\mathbf{z}_{n-1}) \, \mathrm{d}\mathbf{z}_{n-1}.$$

(13.85), (13.59) forward recursion

$$\widehat{\alpha}(\mathbf{z}_n) \equiv p(\mathbf{z}_n|\mathbf{x}_1, \ldots, \mathbf{x}_n)$$
$$c_n \equiv p(\mathbf{x}_n|\mathbf{x}_1, \ldots, \mathbf{x}_{n-1})$$

$$\widehat{\alpha}(\mathbf{z}_n) = \mathcal{N}(\mathbf{z}_n|\boldsymbol{\mu}_n, \mathbf{V}_n).$$ (13.84)

$$c_n \mathcal{N}(\mathbf{z}_n|\boldsymbol{\mu}_n, \mathbf{V}_n) = \mathcal{N}(\mathbf{x}_n|\mathbf{C}\mathbf{z}_n, \boldsymbol{\Sigma}) \int \mathcal{N}(\mathbf{z}_n|\mathbf{A}\mathbf{z}_{n-1}, \boldsymbol{\Gamma}) \mathcal{N}(\mathbf{z}_{n-1}|\boldsymbol{\mu}_{n-1}, \mathbf{V}_{n-1}) \, \mathrm{d}\mathbf{z}_{n-1}$$
$$= \mathcal{N}(\mathbf{x}_n|\mathbf{C}\mathbf{z}_n, \boldsymbol{\Sigma}) \mathcal{N}(\mathbf{z}_n|\mathbf{A}\boldsymbol{\mu}_{n-1}, \mathbf{P}_{n-1})$$

(13.87)

$$c_1 \widehat{\alpha}(\mathbf{z}_1) = p(\mathbf{z}_1) p(\mathbf{x}_1|\mathbf{z}_1). \tag{13.93}$$

Mean and Variance of $p(z_n | x_1, \ldots, x_n, x_{n+1}, \ldots, x_N)$

$$\gamma(\mathbf{z}_n) = \widehat{\alpha}(\mathbf{z}_n)\widehat{\beta}(\mathbf{z}_n) = \mathcal{N}(\mathbf{z}_n|\widehat{\boldsymbol{\mu}}_n, \widehat{\mathbf{V}}_n) \quad (13.84)$$

$$\gamma(\mathbf{z}_n) \equiv p(\mathbf{z}_n|\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}, \ldots, \mathbf{x}_N)$$

$$\widehat{\alpha}(\mathbf{z}_n) \equiv p(\mathbf{z}_n|\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

$$\widehat{\beta}(\mathbf{z}_n) \equiv \frac{p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n)}{p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{x}_1, \ldots, \mathbf{x}_n)}$$

Kalman smoother equations.

$$\widehat{\boldsymbol{\mu}}_n = \boldsymbol{\mu}_n + \mathbf{J}_n\left(\widehat{\boldsymbol{\mu}}_{n+1} - \mathbf{A}\boldsymbol{\mu}_N\right) \quad (13.100)$$

$$\widehat{\mathbf{V}}_n = \mathbf{V}_n + \mathbf{J}_n\left(\widehat{\mathbf{V}}_{n+1} - \mathbf{P}_n\right)\mathbf{J}_n^{\mathrm{T}} \quad (13.101)$$

$$\mathbf{J}_n = \mathbf{V}_n\mathbf{A}^{\mathrm{T}}\left(\mathbf{P}_n\right)^{-1} \quad (13.102)$$

Derivation of Eq.13.100, 13.101

$$c_{n+1}\widehat{\beta}(\mathbf{z}_n) = \int \widehat{\beta}(\mathbf{z}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n)\,\mathrm{d}\mathbf{z}_{n+1}. \qquad (13.99)$$

(13.62)

backward recursion

$$c_{n+1}\widehat{\alpha}(\mathbf{z}_n)\widehat{\beta}(\mathbf{z}_n) = \int \widehat{\alpha}(\mathbf{z}_n)\,\widehat{\beta}(\mathbf{z}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n)\,\mathrm{d}\mathbf{z}_{n+1}$$

$$= \int \frac{\widehat{\alpha}(\mathbf{z}_n)\widehat{\alpha}(\mathbf{z}_{n+1})\widehat{\beta}(\mathbf{z}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n)}{\widehat{\alpha}(\mathbf{z}_{n+1})}\,\mathrm{d}\mathbf{z}_{n+1}$$

$$\gamma(\mathbf{z}_n) = \int \frac{\widehat{\alpha}(\mathbf{z}_n)\,\gamma(\mathbf{z}_{n+1})\,p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n)}{\widehat{\alpha}(\mathbf{z}_{n+1})\,c_{n+1}}\,\mathrm{d}\mathbf{z}_{n+1}$$

the forward-backward algorithm

$$\mathcal{N}(\mathbf{z}_n|\widehat{\boldsymbol{\mu}}_n, \widehat{\mathbf{V}}_n) = \int \frac{\mathcal{N}(\mathbf{z}_n|\boldsymbol{\mu}_n, \mathbf{V}_n)\mathcal{N}(\mathbf{z}_{n+1}|\widehat{\boldsymbol{\mu}}_{n+1}, \widehat{\mathbf{V}}_{n+1})\mathcal{N}(\mathbf{x}_{n+1}|\mathbf{C}\mathbf{z}_{n+1}, \boldsymbol{\Sigma})\mathcal{N}(\mathbf{z}_{n+1}|\mathbf{A}\mathbf{z}_n, \boldsymbol{\Gamma})}{\mathcal{N}(\mathbf{z}_{n+1}|\boldsymbol{\mu}_{n+1}, \mathbf{V}_{n+1})\,\mathcal{N}(\mathbf{x}_{n+1}|\mathbf{C}\mathbf{A}\boldsymbol{\mu}_n, \mathbf{C}\mathbf{P}_n\mathbf{C}^{\mathrm{T}} + \boldsymbol{\Sigma})}\,\mathrm{d}\mathbf{z}_{n+1}$$

Learning Problem

- Determining the parameters $\vartheta = \{A, \Gamma, C, \Sigma, \mu_0, V_0\}$ using the *EM algorithm*.

Expectation of Log Likelihood Function

The complete data ({X, Z}) log likelihood function

$$\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) = \ln p(\mathbf{z}_1|\boldsymbol{\mu}_0, \mathbf{V}_0) + \sum_{n=2}^{N} \ln p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}, \boldsymbol{\Gamma}) + \sum_{n=1}^{N} \ln p(\mathbf{x}_n|\mathbf{z}_n, \mathbf{C}, \boldsymbol{\Sigma}) \qquad (13.108)$$

- The expectation of the log likelihood function with respect to $p(Z \mid X, \theta^{\text{old}})$

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) &= \mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})] \qquad (13.109) \\ &= \mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}[\ln p(\mathbf{z}_1|\boldsymbol{\mu}_0, \mathbf{V}_0)] + \mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}\left[\sum_{n=2}^{N} \ln p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}, \boldsymbol{\Gamma})\right] \\ &\quad + \mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}\left[\sum_{n=1}^{N} \ln p(\mathbf{x}_n|\mathbf{z}_n, \mathbf{C}, \boldsymbol{\Sigma})\right] \end{aligned}$$

Maximizing the Expectation

$$\mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}\left[\ln p(\mathbf{z}_1|\boldsymbol{\mu}_0, \mathbf{V}_0)\right] = -\frac{1}{2}\ln|\mathbf{V}_0| - \mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}\left[\frac{1}{2}(\mathbf{z}_1 - \boldsymbol{\mu}_0)^{\mathrm{T}}\mathbf{V}_0^{-1}(\mathbf{z}_1 - \boldsymbol{\mu}_0)\right] + \text{const}$$

$$\begin{aligned}\mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}\left[\sum_{n=2}^{N}\ln p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}, \boldsymbol{\Gamma})\right] = Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) &= -\frac{N-1}{2}\ln|\boldsymbol{\Gamma}| \\ &-\mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}\left[\frac{1}{2}\sum_{n=2}^{N}(\mathbf{z}_n - \mathbf{A}\mathbf{z}_{n-1})^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}(\mathbf{z}_n - \mathbf{A}\mathbf{z}_{n-1})\right] + \text{const} \qquad (13.112)\end{aligned}$$

$$\begin{aligned}\mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}\left[\sum_{n=1}^{N}\ln p(\mathbf{x}_n|\mathbf{z}_n, \mathbf{C}, \boldsymbol{\Sigma})\right] &= -\frac{N}{2}\ln|\boldsymbol{\Sigma}| \\ &-\mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}\left[\frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \mathbf{C}\mathbf{z}_n)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \mathbf{C}\mathbf{z}_n)\right] + \text{const}\end{aligned}$$

Maximizing each term with respect to the parameters. (Section 2.3.4)

Evaluated ϑ^{new}

$$\boldsymbol{\mu}_0^{\text{new}} = \mathbb{E}[\mathbf{z}_1] \tag{13.110}$$

$$\mathbf{V}_0^{\text{new}} = \mathbb{E}[\mathbf{z}_1\mathbf{z}_1^{\mathrm{T}}] - \mathbb{E}[\mathbf{z}_1]\mathbb{E}[\mathbf{z}_1^{\mathrm{T}}]. \tag{13.111}$$

$$\mathbf{C}^{\text{new}} = \left(\sum_{n=1}^{N} \mathbf{x}_n \mathbb{E}\left[\mathbf{z}_n^{\mathrm{T}}\right]\right)\left(\sum_{n=1}^{N} \mathbb{E}\left[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}\right]\right)^{-1} \tag{13.115}$$

$$\boldsymbol{\Sigma}^{\text{new}} = \frac{1}{N}\sum_{n=1}^{N}\left\{\mathbf{x}_n\mathbf{x}_n^{\mathrm{T}} - \mathbf{C}^{\text{new}}\mathbb{E}\left[\mathbf{z}_n\right]\mathbf{x}_n^{\mathrm{T}} - \mathbf{x}_n\mathbb{E}\left[\mathbf{z}_n^{\mathrm{T}}\right]\mathbf{C}^{\text{new}} + \mathbf{C}^{\text{new}}\mathbb{E}\left[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}\right]\mathbf{C}^{\text{new}}\right\}. \tag{13.116}$$

$$\mathbf{A}^{\text{new}} = \left(\sum_{n=2}^{N} \mathbb{E}\left[\mathbf{z}_n\mathbf{z}_{n-1}^{\mathrm{T}}\right]\right)\left(\sum_{n=2}^{N} \mathbb{E}\left[\mathbf{z}_{n-1}\mathbf{z}_{n-1}^{\mathrm{T}}\right]\right)^{-1} \tag{13.113}$$

$$\boldsymbol{\Gamma}^{\text{new}} = \frac{1}{N-1}\sum_{n=2}^{N}\left\{\mathbb{E}\left[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}\right] - \mathbf{A}^{\text{new}}\mathbb{E}\left[\mathbf{z}_{n-1}\mathbf{z}_n^{\mathrm{T}}\right] - \mathbb{E}\left[\mathbf{z}_n\mathbf{z}_{n-1}^{\mathrm{T}}\right]\mathbf{A}^{\text{new}} + \mathbf{A}^{\text{new}}\mathbb{E}\left[\mathbf{z}_{n-1}\mathbf{z}_{n-1}^{\mathrm{T}}\right]\left(\mathbf{A}^{\text{new}}\right)^{\mathrm{T}}\right\}. \tag{13.114}$$

$$\mathbb{E}\left[\mathbf{z}_n\right] = \widehat{\boldsymbol{\mu}}_n \tag{13.105}$$

$$\mathbb{E}\left[\mathbf{z}_n\mathbf{z}_{n-1}^{\mathrm{T}}\right] = \mathbf{J}_{n-1}\widehat{\mathbf{V}}_n + \widehat{\boldsymbol{\mu}}_n\widehat{\boldsymbol{\mu}}_{n-1}^{\mathrm{T}} \tag{13.106}$$

$$\mathbb{E}\left[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}\right] = \widehat{\mathbf{V}}_n + \widehat{\boldsymbol{\mu}}_n\widehat{\boldsymbol{\mu}}_n^{\mathrm{T}} \tag{13.107}$$

Extensions of LDS

Problem: Beyond the linear-Gaussian assumption.

- Considerable interest in *extending the basic linear dynamical system* in order to increase its capabilities.
- Gaussian $p(z_n \mid x_n)$ – A significant limitation.

Some extensions

- Gaussian mixture $p(z_n)$.
- Gaussian mixture $p(x_n \mid z_n)$ – Impractical.
- *The extended Kalman filter*.
- *The switching state space model / the switching hidden Markov model*.

Particle filters

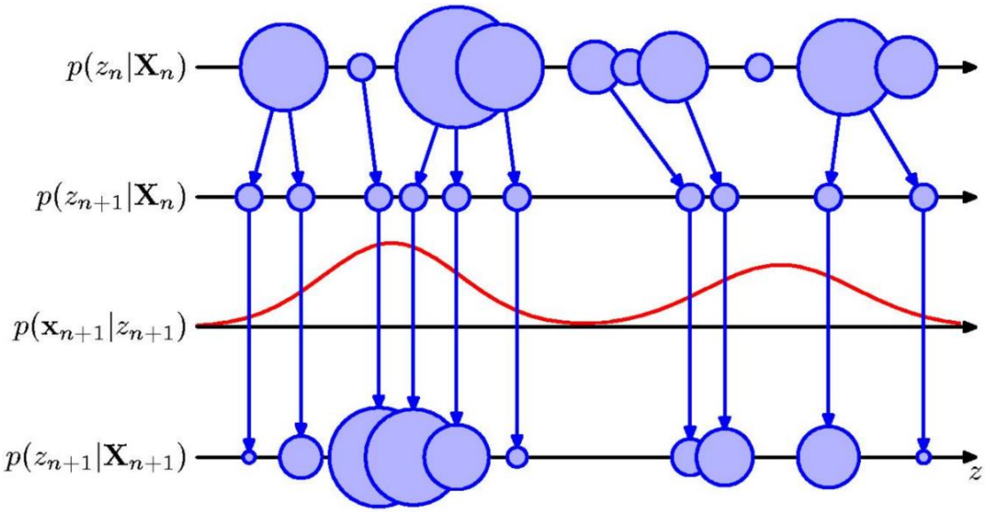
Non-Gaussian emission density $p(x_n \mid z_n)$

→ non-Gaussian $p(z_n \mid x_1, \ldots, x_n)$

→ mathematically intractable integral

$$c_n \widehat{\alpha}(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n) \int \widehat{\alpha}(\mathbf{z}_{n-1}) p(\mathbf{z}_n|\mathbf{z}_{n-1}) \, \mathrm{d}\mathbf{z}_{n-1}.$$

- Sampling-importance-resampling



HW13

HMM： 13.4 13.5 13.8 13.12

LDS: 13.19 13.24 13.25