COMS30035, Machine learning: Sequential Data 3: EM for HMMs

Edwin Simpson
edwin.simpson@bristol.ac.uk

Department of Computer Science, SCEEM
University of Bristol

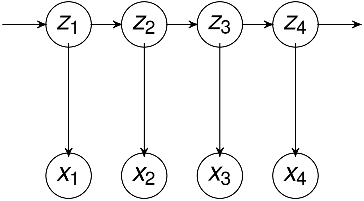
November 3, 2020

Agenda

- Markov Models
- Hidden Markov Models
- EM for HMMs
- Linear Dynamical Systems
- Bayesian Timeseries Modelling with Gaussian Processes

Generalising the Hidden Markov Models

- HMM assumes discrete latent states.
- Linear dynamical systems (LDS) assume states have continuous values.
- Both have the same graphical model:



- Inference has the same form as for an HMM, but when marginalising $\boldsymbol{z}_{n-1}$ and $\boldsymbol{z}_{n+1}$, we take integrals instead of sums.

Motivations for LDS

- Noisy sensors: inferring the true sequence of states from observations with Gaussian noise.
- Tracking: predicting the next movement and tracing the path from noisy observations.

Transition and Emission Distributions for LDS

- $p(\boldsymbol{z}_1) = \mathcal{N}(\boldsymbol{z}_1|\boldsymbol{\mu}_0, \boldsymbol{V}_0)$;
- $p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1}) = \mathcal{N}(\boldsymbol{z}_n|\boldsymbol{A}\boldsymbol{z}_{n-1}, \Gamma)$;
- $p(\boldsymbol{x}_n|\boldsymbol{z}_n) = \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{C}\boldsymbol{z}_n, \Sigma)$.
- Note that the means of both distributions are *linear* functions of the latent states.
- This choice of distributions ensures that the posteriors are also Gaussians with updated parameters
- This means that $\mathcal{O}(N)$ inference can still be performed using the sum-product algorithm.

Inference for an LDS

- *Kalman filter* = forward pass of sum-product for LDS.
- *Kalman smoother* = backward pass of sum-product for LDS.
- No need for an analogue of Viterbi: the most likely sequence is given by the individually most states, so we get this from the Kalman equations.

Forward Inference (Kalman Filter) for an LDS

- $$\alpha(\boldsymbol{z}_n) = \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{C}\boldsymbol{z}_n, \Sigma) \int \mathcal{N}(\boldsymbol{z}_n | \boldsymbol{A}\boldsymbol{z}_{n-1}, \Gamma)\alpha(\boldsymbol{z}_{n-1})\mathrm{d}\boldsymbol{z}_{n-1} \tag{1}$$
- Normalising results in a Gaussian-distributed variable, whose parameters can be computed efficiently: $\hat{\alpha}(\boldsymbol{z}_n) = p(\boldsymbol{z}_n | \boldsymbol{x}_1, ..., \boldsymbol{x}_N) = \mathcal{N}(\boldsymbol{z}_n | \boldsymbol{\mu}_n, \boldsymbol{V}_n)$, where
  - $\boldsymbol{\mu}_n$ is a function of $\boldsymbol{\mu}_{n-1}$, $\boldsymbol{x}_n$, $\boldsymbol{A}$ and $\boldsymbol{C}$.
  - $\boldsymbol{V}_n$ is a function of $\boldsymbol{V}_{n-1}$, Σ, $\boldsymbol{A}$, Γ and $\boldsymbol{C}$.
- We can view each forward step as predicting $\boldsymbol{z}_n$ based on the distribution over $\boldsymbol{z}_{n-1}$, then correcting that prediction given the new observation $\boldsymbol{x}_n$.
- For details, see Bishop (2006), Section 13.3.1

Backward Inference (Kalman Smoother) for an LDS

- Backward pass also follows that of the HMM: messages are passed from the final state to the start of the sequence.
- The backward messages contain information about future states that affects the posterior distribution at each step n.
- Since the transition and emission probabilities are all Gaussian, the posterior *responsibilities* are also Gaussian, as are the *state pair* expectations.
- For details, see Bishop (2006), Section 13.3.1

Learning the Parameters of LDS

- Kalman filter/smoother are analogous to the forward-backward algorithm for HMMs.
- Remember that this algorithm is used for the *E step* of EM.
- The parameters are optimised in the *M step* as before, by using the responsibilities $\mathbb{E}[\boldsymbol{z}_n]$, $\mathbb{E}[\boldsymbol{z}_n \boldsymbol{z}_n^T]$ and state pair expectations $\mathbb{E}[\boldsymbol{z}_n \boldsymbol{z}_{n-1}^T]$.
- For details, see Bishop (2006), Section 13.3.2

Now do the quiz!

Please do the quiz for this lecture on Blackboard.