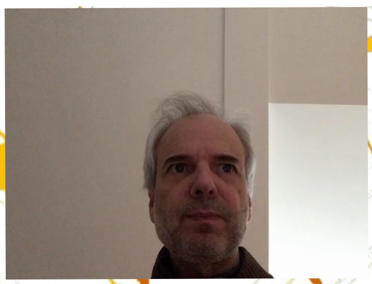


TIME SERIES ANALYSIS & RECURRENT NEURAL NETWORKS #5

- State Space Models & Evidence Lower Bound
- Expectation-Maximization Algorithm
- Parameter & state inference in linear SSM

Main lecture: Daniel Durstewitz

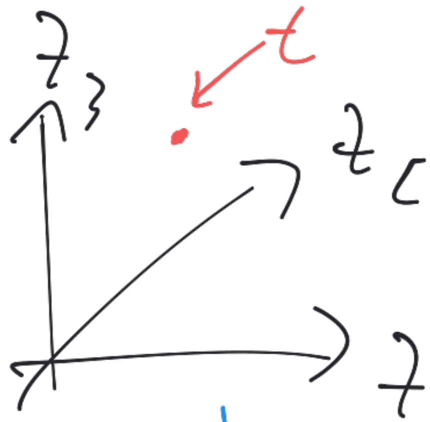
Exercises: Leonard Bereska, Manuel Brenner, Daniel Kramer, Georgia Koppe

Heidelberg University

State Space Models

$$\dot{z} = f(z, t), \quad z = (z_1(t), \dots z_n(t))$$

$$z_t = F(z_{t-1}, \varepsilon_t)$$



$$p(x_t, x_{t'} | z_t, z_{t'}) = p(x_t | z_t)\, p(x_{t'} | z_{t'}) \neq p(x_t, x_{t'}) \neq p(x_t)\, p(x_{t'})$$

$\{x_t\}_{ob}$, $t = 1 \dots T$, $x_t \sim p_{\theta_{obs}}(g(z_t))$ in general

Linear state space model / linear DS

$\{x_t\}, t=1\dots T, \; x_t = (x_{1t} \dots x_{Nt})^T$

obs. eq. $\underset{N\times 1}{x_t} = \underset{N\times M}{B}\,\underset{M\times 1}{z_t} + \underset{N\times 1}{\eta_t}, \quad \eta_t \sim N(0, \underset{N\times N}{\Gamma})$

$x_t \sim N(Bz_t, \Gamma), \quad x_t \sim p_{\theta_{obs}}(g(z_t))$

$\theta_{obs} = \{B, \Gamma\}$

latent process eq. $z_t = \underset{M\times M}{A} z_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma)$ → VAR(1)

transition matrix

$z_t \sim N(Az_{t-1}, \Sigma)$ $\quad \theta_{lat} = \{A, \Sigma, \mu_0, \Sigma_0\}$

initial cond. $z_1 \sim N(\mu_0, \Sigma_0)$ $\quad \Sigma_0 = \Sigma$

$Z = \{z_t\}, \; t = 1 \dots T$

EM algo.

$p(x_{11} \dots x_{M1} \dots x_{1T} \dots x_{NT}, z_{11} \dots z_{M1} \dots z_{MT})$

M-step: max. $E_q[\log p(X, Z)]$ w.r.t. $\theta = \{B, \Gamma, A, \Sigma, \mu_0\}$

$$\boxed{\begin{aligned} p(X,Z) &= p(X|Z)\,p(Z) \\ &= p(z_1)\,p(x_1|z_1) \prod_{t=2}^{T} p(x_t|z_t)\,p(z_t|z_{t-1}) \end{aligned}}$$

$$E_q[\log p(X,Z)] = E_q[\log p(z_1)] + E_q\left[\sum_{t=2}^{T} \log p(z_t|z_{t-1})\right] + E_q\left[\sum_{t=1}^{T} \log p(x_t|z_t)\right]$$

initial cond.

trans. eq.

obs. eq.

$z_1 \sim N(\mu_0, \Sigma)$

$$= E_q\left[\cancel{-\frac{M}{2}\log(2\pi)} - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(z_1-\mu_0)^T \Sigma^{-1} (z_1 - \mu_0)\right] +$$

$z_t \sim N(Az_{t-1}, \Sigma)$

$$E_q\left[\cancel{-\frac{M(T-1)}{2}\log(2\pi)} - \frac{T-1}{2}\log|\Sigma| - \frac{1}{2}\sum_{t=2}^{T}(z_t - Az_{t-1})^T \Sigma^{-1} (z_t - Az_{t-1})\right] +$$

$x_t \sim N(Bz_t, \Gamma)$

$$E_q\left[\cancel{-\frac{NT}{2}\log(2\pi)} - \frac{T}{2}\log|\Gamma| - \frac{1}{2}\sum_{t=1}^{T}(x_t - Bz_t)^T \Gamma^{-1} (x_t - Bz_t)\right]$$

$$= -\frac{T-1}{2}\log|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}\Big[E[z_t^T \Sigma^{-1} z_t] - E[z_t^T \Sigma^{-1} A z_{t-1}] - E[z_{t-1}^T A^T \Sigma^{-1} z_t] + E[z_{t-1}^T A^T \Sigma^{-1} A z_{t-1}]\Big] + \text{const.}$$

$$x^T A y = \mathrm{Tr}[A y x^T]$$

$$\Rightarrow E[z_t^T \Sigma^{-1} z_t] = \mathrm{Tr}(\Sigma^{-1} E[z_t z_t^T])$$

$$E[z_t^T \Sigma^{-1} A z_{t-1}] = \mathrm{Tr}(\Sigma^{-1} A\, E[z_{t-1} z_t^T])$$

$$\mathrm{Tr}(A^T \Sigma^{-1} A\, E[z_{t-1} z_{t-1}^T])$$

$\rightarrow E[z_t], \; E[z_t z_t^T], \; E[z_{t-1} z_t^T]$

performing max. ELBO w.r.t. θ require 1st and 2nd moments $p(Z|X)$

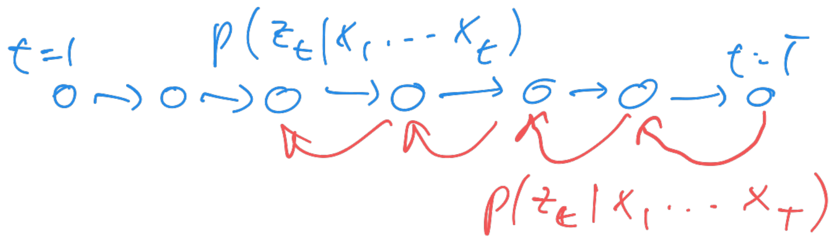
E-step: $\arg\min_q KL[q(z|x) \| p_\theta(z|x)]$ given fixed θ

$\rightarrow E[z_t], E[z_t z_{t-1}^T], \ldots$

$q(z|x) = p_\theta(z|x)$

"Kalman filter-smoother recursions"

$\rightarrow$ linear in T



"filter / forward loop"

"smoothing / backward loop"

$\rightarrow p(z_t, z_{t-1} | x_1 \ldots x_T) \rightarrow E[z_t z_{t-1}^T]$

$$p(z_t|x_1 \ldots x_T) = \frac{p(z_t, x_1 \ldots x_t, x_{t+1} \ldots x_T)}{p(x_1 \ldots x_T)}$$

- Bayes' rule
- Markov prop.
- condit. indep.

$$= \underbrace{\frac{p(z_t, x_1 \ldots x_t)}{p(x_1 \ldots x_t)}}_{\leq t} \times \underbrace{\frac{p(x_{t+1} \ldots x_T | z_t, \cancel{x_1 \ldots x_t})}{p(x_{t+1} \ldots x_T | x_1 \ldots x_t)}}_{> t}$$

$$= p(z_t | x_1 \ldots x_t) = \frac{p(z_t, x_1 \ldots x_t)}{p(x_t|x_1 \ldots x_{t-1}) \, p(x_1 \ldots x_{t-1})}$$

obs. model

$$= \frac{p(x_t | z_t, \cancel{x_1 \ldots x_{t-1}}) \, p(z_t | x_1 \ldots x_{t-1}) \, \cancel{p(x_1 \ldots x_{t-1})}}{p(x_t | x_1 \ldots x_{t-1}) \, \cancel{p(x_1 \ldots x_{t-1})}}$$

$$= \frac{p(x_t|z_t) \int_{z_{t-1}} p(z_t, z_{t-1} | x_1 \ldots x_{t-1}) \, dz_{t-1}}{p(x_t | x_1 \ldots x_{t-1})}$$

transition model

$$= \frac{p(x_t|z_t) \int_{z_{t-1}} p(z_t|z_{t-1}) \, p(z_{t-1} | x_1 \ldots x_{t-1}) \, dz_{t-1}}{p(x_t | x_1 \ldots x_{t-1})}$$

$\rightarrow$ recursion in time!