TIME SERIES ANALYSIS & RECURRENT NEURAL NETWORKS #6

- Kalman filter & smoother
- Poisson state space models

Main lecture: Daniel Durstewitz

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

Heidelberg University

Linear state space models

obs. eqn. $x_t \sim N(Bz_t, \Gamma)$

$X = \{x_t\}, t = 1, \dots T$

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

ini. cond. $z_1 \sim N(\mu_0, \Sigma)$

goal: max. $\log p(X|\theta)$ w.r.t.

$\theta = \{B, \Gamma, A, \Sigma, \mu_0\}$

$\rightarrow$ ELBO $\leq \log p(X|\theta)$

EM algorithm

M-step: max. $E_{q(z|x)}[\log p(x,z)]$ w.r.t. θ

$\rightarrow E[z_t], E[z_t z_t^T], E[z_t z_{t-1}^T]$ from $p(z|x)$

$\rightarrow$ quad. eqn. for lin. Gaussian models

$\rightarrow$ linear regr. problem

E-step: min. $KL[q(z|x) \| p_\theta(z|x)]$

$q(z|x) = p_\theta(z|x)$ given θ^*

$p(z_t|x_1 \dots x_T), p(z_t, z_{t-1}|x_1 \dots x_T)$

$$p(z_t|x_1 \dots x_T) = \underbrace{\boxed{\frac{p(z_t, x_1 \dots x_t)}{p(x_1 \dots x_t)}}}_{\text{filtering loop}} \times \underbrace{\frac{p(x_{t+1} \dots x_T | z_t)}{p(x_{t+1} \dots x_T | x_1 \dots x_t)}}_{\text{smooth loop}}$$

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

$$\underbrace{p(z_t|x_1 \dots x_t)}_{N(\mu_t, V_t)} = \frac{\overbrace{p(x_t|z_t)}^{\text{obs. eqn.}} \int_{z_{t-1}} \overbrace{p(z_t|z_{t-1})}^{\text{trans. model}} \underbrace{p(z_{t-1}|x_1 \dots x_{t-1})}_{N(\mu_{t-1}, V_{t-1})} \, dz_{t-1}}{p(x_t|x_1 \dots x_{t-1})}$$

$$p(z_t|z_{t-1}) = N(Az_{t-1}, \Sigma)$$

$$p(z_{t-1}|x_1 \dots x_{t-1}) = N(\mu_{t-1}, V_{t-1})$$

$$z_t = (z_{1t} \dots z_{Mt})^T$$

$$\int_{z_{t-1}} (2\pi)^{-M/2} |\Sigma|^{-1/2} e^{-\frac{1}{2}(z_t - Az_{t-1})^T \Sigma^{-1} (z_t - Az_{t-1})} \times (2\pi)^{-M/2} |V_{t-1}|^{-\frac{1}{2}} e^{-\frac{1}{2}(z_{t-1} - \mu_{t-1})^T V_{t-1}^{-1} (z_{t-1} - \mu_{t-1})} dz_{t-1}$$

focus on exponent: $-\frac{1}{2}(z_{t-1} - m)^T H^{-1} (z_{t-1} - m)$

$$-\frac{1}{2}\left[(z_t - Az_{t-1})^T \Sigma^{-1} (z_t - Az_{t-1}) + (z_{t-1} - \mu_{t-1})^T V_{t-1}^{-1} (z_{t-1} - \mu_{t-1})\right]$$

$$= -\frac{1}{2}\Big[z_{t-1}^T \underbrace{(A^T \Sigma^{-1} A + V_{t-1}^{-1})}_{=: H^{-1}} z_{t-1} - (z_t^T \Sigma^{-1} A + \mu_{t-1}^T V_{t-1}^{-1}) z_{t-1}$$

$$- z_{t-1}^T \underbrace{(A^T \Sigma^{-1} z_t + V_{t-1}^{-1} \mu_{t-1})}_{H^{-1} m} + z_t^T \Sigma^{-1} z_t + \mu_{t-1}^T V_{t-1}^{-1} \mu_{t-1}\Big]$$

$$(A^T \Sigma^{-1} z_t + V_{t-1}^{-1} \mu_{t-1}) = H^{-1} m$$

$$\Rightarrow m = (A^T \Sigma^{-1} A + V_{t-1}^{-1})^{-1} (A^T \Sigma^{-1} z_t + V_{t-1}^{-1} \mu_{t-1})$$

$$= -\frac{1}{2}\Big[z_{t-1}^T H^{-1} z_{t-1} - m^T H^{-1} z_{t-1} - z_{t-1}^T H^{-1} m$$

$$+ m^T H^{-1} m - m^T H^{-1} m + z_t^T \Sigma^{-1} z_t + \mu_{t-1}^T V_{t-1}^{-1} \mu_{t-1}\Big]$$

$$c := (2\pi)^{-M/2} |\Sigma|^{-1/2} (2\pi)^{-M/2} |V_{t-1}|^{-1/2}$$

$$\Rightarrow c \cdot \int_{z_{t-1}} e^{-\frac{1}{2}(z_{t-1} - m)^T H^{-1} (z_{t-1} - m)} e^{-\frac{1}{2}(-m^T H^{-1} m + z_t^T \Sigma^{-1} z_t + \mu_{t-1}^T V_{t-1}^{-1} \mu_{t-1})} dz_{t-1}$$

$$(2\pi)^{-M/2} |H|^{-1/2} \cdot |H|^{1/2}$$

$$\tilde{c} := (2\pi)^{-M/2} |\Sigma|^{-1/2} |V_{t-1}|^{-1/2} \cdot |H|^{1/2}$$

$$\Rightarrow \tilde{c} e^{-\frac{1}{2}[-m^T H^{-1} m + z_t^T \Sigma^{-1} z_t + \mu_{t-1}^T V_{t-1}^{-1} \mu_{t}]} \times$$

$$\underbrace{\int_{z_{t-1}} (2\pi)^{-M/2} |H|^{-1/2} e^{-\frac{1}{2}(z_{t-1} - m)^T H^{-1} (z_{t-1} - m)} dz_{t-1}}_{1}$$

$H^{-1} := (A^T\Sigma^{-1}A + V_{t-1}^{-1})$

$m = (A^T\Sigma^{-1}A + V_{t-1}^{-1})^{-1}(A^T\Sigma^{-1}z_t - V_{t-1}^{-1}\mu_{t-1})$

$-\frac{1}{2}\left[-m^T H^{-1} m + z_t^T\Sigma^{-1}z_t + \mu_{t-1}^T V_{t-1}^{-1}\mu_{t-1}\right]$

$= -\frac{1}{2}\Big[z_t^T\Sigma^{-1}z_t + \mu_{t-1}V_{t-1}^{-1}\mu_{t-1} - (A^T\Sigma^{-1}z_t - V_{t-1}^{-1}\mu_{t-1})^T \cancel{(A^T\Sigma^{-1}A+V_{t-1}^{-1})^{-1}(A^T\Sigma^{-1}A+V_{t-1}^{-1})}(A^T\Sigma^{-1}A+V_{t-1}^{-1})^{-1}(A^T\Sigma^{-1}z_t - V_{t-1}^{-1}\mu_{t-1})\Big]$

$= -\frac{1}{2}\Big[z_t^T \underbrace{\left(\Sigma^{-1} - \Sigma^{-1}A(A^T\Sigma^{-1}A + V_{t-1}^{-1})^{-1}A^T\Sigma^{-1}\right)}_{=(AV_{t-1}A^T+\Sigma)^{-1}} z_t$ Woodbury's identities

$- z_t^T\left(\Sigma^{-1}A(A^T\Sigma^{-1}A + V_{t-1}^{-1})^{-1}V_{t-1}^{-1}\right)\mu_{t-1}$

$- \mu_{t-1}^T\left(V_{t-1}^{-1}(A^T\Sigma^{-1}A + V_{t-1}^{-1})^{-1}\Sigma^{-1}A^T\right)z_t$

$+ \mu_{t-1}^T \underbrace{\left(V_{t-1}^{-1} - V_{t-1}^{-1}(A^T\Sigma^{-1}A + V_{t-1}^{-1})^{-1}V_{t-1}^{-1}\right)}_{=A^T(AV_{t-1}A^T+\Sigma)^{-1}A}\mu_{t-1}\Big]$

$(AB)^{-1} = B^{-1}A^{-1}$

$\left[\cancel{(A^T\Sigma^{-1})}^{-1}(\cancel{A^T\Sigma^{-1}}A + \cancel{V_{t-1}^{-1}})\cancel{V_{t-1}}\right]^{-1}$

$= \underbrace{(A^{-1}A)}_{I}{}^T\left[AV_{t-1} + (A^T\Sigma^{-1})^{-1}\right]^{-1}$

$= A^TA^{-T}\left[AV_{t-1} + \Sigma A^{-T}\right]^{-1}$

$= A^T\left[AV_{t-1}A^T + \Sigma\right]^{-1}$

$-\frac{1}{2}\Big[z_t^T(AV_{t-1}A^T+\Sigma)^{-1}z_t - z_t^T \underbrace{(AV_{t-1}A^T+\Sigma)^{-1}}_{=:L_{t-1}^{-1}} A\mu_{t-1}$

$- \mu_{t-1}^T A^T(AV_{t-1}A^T+\Sigma)^{-1}z_t + \mu_{t-1}^T A^T(AV_{t-1}A^T+\Sigma^{-1})A\mu_{t-1}\Big]$

$= -\frac{1}{2}(z_t - A\mu_{t-1})^T L_{t-1}^{-1}(z_t - A\mu_{t-1})$

$\Rightarrow \int_{z_{t-1}} p(z_t|z_{t-1})\underbrace{p(z_{t-1}|x_1\dots x_{t-1})}_{N(\mu_{t-1},V_{t-1})}\,dz_{t-1}$ Kolmogorov-Chapman

$= N(A\mu_{t-1}, AV_{t-1}A^T + \Sigma^{-1}) = \underbrace{p(z_t|x_1\dots x_{t-1})}_{\text{1-step forward density}}$

$$p(z_t|x_1\dots x_t) = \frac{\overbrace{N(Bz_{t-1},\Gamma)}^{p(x_t|z_t)}\,\overbrace{N(A\mu_{t-1},L_{t-1})}^{p(z_t|x_1\dots x_{t-1})}}{p(x_t|x_1\dots x_{t-1})}$$

$= N(\mu_t, V_t)$

μ_{t-1} μ_t μ_{t+1}

Kalman filter recursions:

$$\boxed{\begin{aligned} &\mu_t = A\mu_{t-1} + K_t(x_t - \underbrace{B\underbrace{A\mu_{t-1}}_{\hat{z}_t}}_{\hat{x}_t}) \\ &\text{Kalman gain mtx. } K_t := L_{t-1}B^T(BL_{t-1}B^T + \Gamma)^{-1} \\ &V_t = \left[(AV_{t-1}A^T + \Sigma)^{-1} + B^T\Gamma^{-1}B\right]^{-1} \end{aligned}}$$

$p(z_t|x_1\dots x_t)$, $p(z_t|x_1\dots x_T)$

→ smoother step

$$\underbrace{p(z_t|x_1\dots x_T)}_{=:\gamma_t = \alpha_t\beta_t} = \underbrace{p(z_t|x_1\dots x_t)}_{=:\alpha_t = N(\mu_t,V_t)} \times \underbrace{\frac{p(x_{t+1:T}|z_t)}{p(x_{t+1:T}|x_1\dots x_t)}}_{=:\beta_t}$$

$\alpha_1 \quad \alpha_2 \qquad\qquad \alpha_T$

$\circ \to \circ \to \dots \to \circ$

$\circ \leftarrow \circ \leftarrow \circ \dots \circ \leftarrow \circ$

$\gamma_1 \qquad\qquad\qquad \gamma_T$

$p(z_T|x_1\dots x_T)$

$\gamma_T = \alpha_T$

$$\gamma_t = \alpha_t\beta_t = \alpha_t\underbrace{\frac{p(x_{t+1}\dots x_T|z_t)}{p(x_{t+1}\dots x_T|x_1\dots x_t)}}_{=:\beta_t}$$

$$= \alpha_t\frac{\int_{z_{t+1}} p(z_{t+1}, x_{t+1}, x_{t+2},\dots x_T|z_t)\,dz_{t+1}}{p(x_{t+1}\dots x_T|x_1\dots x_t)}$$ $p(x_{t+1}|z_{t+1})p(z_{t+1}|z_t)$

$$= \alpha_t\frac{\int_{z_{t+1}} p(x_{t+2}\dots x_T|z_{t+1},\cancel{z_t})\,\overbrace{p(x_{t+1},z_{t+1}|z_t)}\,dz_{t+1}}{p(x_{t+2}\dots x_T|x_1\dots x_{t+1})\,p(x_{t+1}|x_1\dots x_t)}$$

$$= \alpha_t\frac{\int_{z_{t+1}} p(x_{t+2}\dots x_T|z_{t+1})\,p(x_{t+1}|z_{t+1})\,p(z_{t+1}|z_t)\,dz_{t+1}}{p(x_{t+2}\dots x_T|x_1\dots x_{t+1})\,p(x_{t+1}|x_1\dots x_t)}$$ obs. dens. trans. dens.

$= \beta_{t+1} = \alpha_{t+1}^{-1}\gamma_{t+1}$

$$= \alpha_t\frac{\int_{z_{t+1}} \alpha_{t+1}^{-1}\gamma_{t+1}\,p(x_{t+1}|z_{t+1})\,p(z_{t+1}|z_t)\,dz_{t+1}}{p(x_{t+1}|x_1\dots x_t)}$$

$$\boxed{\begin{aligned} &p(z_t|x_1\dots x_T) \to E[z_t] =: \tilde{\mu}_t \\ &\qquad\qquad\qquad \mathrm{cov}[z_t] =: \tilde{V}_t \\ &E[z_tz_t^T] = \mathrm{cov}[z_t] + E[z_t]E[z_t^T] \\ &E[z_tz_{t-1}^T] \end{aligned}}$$

$E[z_t|x_1\dots x_T]$

$$p(z_t, z_{t-1}|x_1\dots x_T) = \frac{p(x_1\dots x_T|z_t,z_{t-1})\,p(z_{t-1})\,p(z_t|z_{t-1})}{p(x_1\dots x_T)}$$

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

trans. obs.

$$= \underbrace{\frac{p(x_1\dots x_{t-1}, z_{t-1})}{p(x_1\dots x_{t-1})}}_{\alpha_{t-1}} \times \underbrace{\frac{p(z_t|z_{t-1})\,p(x_t|z_t)}{p(x_t|x_1\dots x_{t-1})}}_{\text{normalizing KF}} \times \underbrace{\frac{p(x_{t+1}\dots x_T|z_t)}{p(x_{t+1}\dots x_T|x_1\dots x_t)}}_{\beta_t = \alpha_t^{-1}\gamma_t}$$