STAT215: Assignment 4

Your Name

Due: March 13, 2020

Problem 1: Consider a Gaussian linear dynamical system (LDS),

$$p(x_{1:T}, y_{1:T}) = \mathcal{N}(x_1 \mid 0, q^2) \left[\prod_{t=2}^{T} \mathcal{N}(x_t \mid a x_{t-1} + b, q^2) \right] \left[\prod_{t=1}^{T} \mathcal{N}(y_t \mid x_t, r^2) \right],$$

for $x_t, y_t \in \mathbb{R}$ for all t, and parameters $a, b \in \mathbb{R}$ and $q^2, r^2 \in \mathbb{R}_+$. Compute the forward filtered distribution $p(x_t \mid y_{1:t})$ in terms of the model parameters and the filtered distribution $p(x_{t-1} \mid y_{1:t-1})$. Solve for the base case $p(x_1 \mid y_1)$. For reference, consult the state space modeling chapters of either the Bishop or the Murphy textbook.

A directed graph of linear Gaussians is equivalent to a joint Gaussian distribution over all of variables, so the marginal distribution of each variable is also Gaussian. Similarly, in this linear Gaussian dynamical system, the posterior marginals of latent variables will be Gaussian. So, we know that the forward filtered distributions take the form

$$\begin{aligned} p(x_t \mid y_{1:t}) &= \mathcal{N}(x_t \,|\, m_t, v_t) \\ p(x_{t-1} \mid y_{1:t-1}) &= \mathcal{N}(x_{t-1} \,|\, m_{t-1}, v_{t-1}) \end{aligned}$$

We will first derive the forward filtered distribution $p(x_t \mid y_{1:t})$ by evaluating the posterior marginal and conditional distributions of a linear Gaussian model (method A), and then confirm with the derived Kalman filter equations (method B).

Method A

Given the model parameters and the filtered distribution from the previous timestep $p(x_{t-1} \mid y_{1:t-1}) \propto \mathcal{N}(x_{t-1} \,|\, m_{t-1}, v_{t-1})$, we have

$$\begin{aligned} p(x_t \mid y_{1:t}) &\propto p(y_{1:t} \,|\, x_t)\, p(x_t) \\ &\propto p(y_t \,|\, x_t)\, p(y_{1:t-1} \,|\, x_t)\, p(x_t) \\ &\propto p(y_t \,|\, x_t) \int p(y_{1:t-1} \,|\, x_t, x_{t-1})\, p(x_t \,|\, x_{t-1})\, \mathrm{d}x_{t-1} \\ &\propto p(y_t \,|\, x_t) \int p(x_{t-1} \,|\, y_{1:t-1})\, p(x_t \,|\, x_{t-1})\, \mathrm{d}x_{t-1} \\ &\propto \mathcal{N}(y_t \,|\, x_t, r^2) \int \mathcal{N}(x_{t-1} \,|\, m_{t-1}, v_{t-1})\, \mathcal{N}(x_t \,|\, a x_{t-1} + b, q^2)\, \mathrm{d}x_{t-1}. \qquad (1) \end{aligned}$$

Using the posterior marginal distribution results for linear Gaussian models (Bishop 2.115), we can simplify the integral as follows

$$\begin{aligned} p(x_t) &= \int p(x_{t-1})\, p(x_t \,|\, x_{t-1})\, \mathrm{d}x_{t-1} \\ &= \int \mathcal{N}(x_{t-1} \,|\, m_{t-1}, V_{t-1})\, \mathcal{N}(x_t \,|\, a x_{t-1} + b, q^2)\, \mathrm{d}x_{t-1} \\ &= \int \mathcal{N}(x_t \,|\, a m_{t-1}, p_{t-1})\, \mathrm{d}x_{t-1} = \mathcal{N}(x_t \,|\, a m_{t-1}, p_{t-1}) \end{aligned}$$

where $p_{t-1} = q^2 + a^2 v_{t-1}$. Then, continuing from Eqn. (1), we again make use of Bishop Eqns. (2.115) and (2.116) to identify,

Bishop Eqns. 2.113-2.117	This problem
$p(x) = \mathcal{N}(x \,\|\, \tilde{\mu}, \Lambda^{-1})$	$p(x_t) = \mathcal{N}(x_t \,\|\, a \cdot m_{t-1}, p_{t-1})$
$p(y \,\|\, x) = \mathcal{N}(y \,\|\, \tilde{A}x + \tilde{b}, L^{-1})$	$p(y_t \,\|\, x_t) = \mathcal{N}(y_t \,\|\, x_t, r^2)$
$p(y) = \mathcal{N}(y \,\|\, \tilde{A}\mu + b, \tilde{L}^{-1} + \tilde{A}\Lambda\tilde{A}^\top)$	$p(y_t) = \mathcal{N}(y_t \,\|\, a \cdot m_{t-1}, r^2 + p_{t-1}^{-1})$
$p(x \,\|\, y) = \mathcal{N}(x \,\|\, \tilde{\Sigma}(\tilde{A}^\top L(y - \tilde{b}) + \Lambda\tilde{\mu}), \tilde{\Sigma})$	$p(x_t \,\|\, y_t) = \mathcal{N}(x_t \,\|\, \tilde{\Sigma}(r^{-2} y + p_{t-1}^{-1}(a \cdot m_{t-1})), \tilde{\Sigma})$
for $\tilde{\Sigma} = (\Lambda + A^\top L A)^{-1}$	for $\tilde{\Sigma} = (p_{t-1}^{-1} + r^{-2})^{-1}$

where $\tilde{\mu} = a \cdot m_{t-1}$, $\Lambda^{-1} = p_{t-1}$, $\tilde{A} = 1$, $\tilde{b} = 0$, and $L^{-1} = r^2$. So $p(x_t \mid y_{1:t}) = \mathcal{N}(x_t \,|\, m_t, v_t)$ for

$$m_t = \frac{p_{t-1}\, y_t + a\, r^2\, m_{t-1}}{r^2 + p_{t-1}} \tag{2}$$

$$v_t = \frac{r^2\, p_{t-1}}{r^2 + p_{t-1}} \tag{3}$$

Method B

Bishop LDS Model	This problem
$p(x_n \,\|\, x_{n-1}) = \mathcal{N}(x_n \,\|\, A x_{n-1}, \Gamma)$	$p(x_t \,\|\, x_{t-1}) = \mathcal{N}(x_n \,\|\, a x_{t-1}, q^2)$
$p(y_n \,\|\, x_n) = \mathcal{N}(x_t \,\|\, C x_n, \Sigma)$	$p(y_t \,\|\, x_t) = \mathcal{N}(y_t \,\|\, x_t, r^2)$

Bishop Eqns. 13.89-13.92	This problem
$\mu_n = A\mu_{n-1} + K_n(x_n - CA\mu_{n-1})$	$m_t = a \cdot m_{t-1} + k_t(y_t - a \cdot m_{t-1})$
$V_n = (I - K_n C) P_{n-1}$	$v_t = (1 - k_t) p_{t-1}$
$c_n = \mathcal{N}(x_n \,\|\, CA\mu_{n-1}, CP_{n-1}C^\top + \Sigma)$	$p(y_t) = \mathcal{N}(y_t \,\|\, a \cdot m_{t-1}, p_{t-1} + r^2)$
$K_n = P_{n-1} C^\top (CP_{n-1}C^\top + \Sigma)^{-1}$	$k_t = p_{t-1}(p_{t-1} + r^2)^{-1}$

If we expand out the expressions for k_t, m_t, and v_t, we find that they agree with Eqn. (2) and (3) which were found via the first method:

$$
\begin{aligned}
k_t &= \frac{p_{t-1}}{p_{t-1} + r^2} \\
m_t &= a \cdot m_{t-1} + k_t (y_t - a \cdot m_{t-1}) \\
&= \frac{a \cdot m_{t-1} (p_{t-1} + r^2)}{p_{t-1} + r^2} + \frac{p_{t-1} (y_t - a \cdot m_{t-1})}{p_{t-1} + r^2} \\
&= \frac{a\, p_{t-1}\, m_{t-1} + a\, r^2\, m_{t-1} + p_{t-1}\, y_t - a\, p_{t-1}\, m_{t-1})}{p_{t-1} + r^2} \\
&= \frac{a\, r^2\, m_{t-1} + p_{t-1}\, y_t}{p_{t-1} + r^2} \qquad (4) \\
v_t &= (1 - k_t)\, p_{t-1} \\
&= \left(1 - \frac{p_{t-1}}{p_{t-1} + r^2}\right) p_{t-1} \\
&= \frac{p_{t-1} + r^2 - p_{t-1}}{p_{t-1} + r^2} p_{t-1} = \frac{p_{t-1}\, r^2 - p_{t-1}}{p_{t-1} + r^2} \qquad (5)
\end{aligned}
$$

The base case $p(x_1 \,|\, y_1)$ is then given by Bishop Eqns. (13.94-13.97):

$$
\begin{aligned}
m_1 &= k_1\, y_1 \\
v_1 &= (1 - k_1)\, q^2 \\
c_1 &= \mathcal{N}(y_1 \,|\, 0, q^2 + r^2) \\
k_1 &= \frac{q^2}{q^2 + r^2}
\end{aligned}
$$

In the Kalman filtering equations, we see how the new model mean m_t is a weighted average of the evolved model mean $a \cdot m_{t-1}$ and the error betwen observation and evolved mean.

Problem 2: Sample a time series of length $T = 30$ from the Gaussian LDS in Problem 1 with parameters $a = 1, b = 0, q = 0.1, r = 0.3$. Plot the sample of $x_{1:T}$ as a solid line, and plot the observed $y_{1:T}$ as $+$'s. Write code to compute the filtered distribution $p(x_t \mid y_{1:t})$ you derived in Problem 1. Then plot the mean of the filtered distribution $\mathbb{E}[x_t \mid y_{1:t}]$ over time as a solid line, and plot a shaded region encompassing the mean ± 2 standard deviations of the filtered distribution. All plots should be on the same axis. Include a legend.

Write your code in a Colab notebook and include a PDF printout of your notebook as well as the raw .ipynb file.

Problem 3: Reproduce Figure 2.5 of Rasmussen and Williams, *Gaussian Processes for Machine Learning*, available at http://www.gaussianprocess.org/gpml/chapters/RW2.pdf. Use a randomly generated dataset as described in the figure caption and surrounding text.

Write your code in a Colab notebook and include a PDF printout of your notebook as well as the raw .ipynb file.