GP as an Infinite Dimensional Gaussians

Ex) $x_{t+1} = a x_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2)$

$E[\varepsilon_t] = 0$

$E[\varepsilon_t^2] = \sigma^2$

$E[\varepsilon_t \varepsilon_{t'}] = 0 \quad t \neq t'$

$$
\begin{aligned}
x_t &= a x_{t-1} + \varepsilon_{t-1} \\
&= a(a x_{t-2} + \varepsilon_{t-2}) + \varepsilon_{t-1} \\
&= a^2 x_{t-2} + a\varepsilon_{t-2} + \varepsilon_{t-1} \\
&= a^3 x_{t-3} + a^2 \varepsilon_{t-3} + a\varepsilon_{t-2} + \varepsilon_{t-1} \\
&= \cdots \\
&= \sum_{i=0}^{\infty} a^i \varepsilon_{t-1-i}
\end{aligned}
$$

$$E[x_t] = \sum_{i=0}^{\infty} a^i E[\varepsilon_{t-1-i}] = 0$$

$$E[x_t x_{t+\Delta t}] = E\left(\sum_{i=0}^{\infty} a^i \varepsilon_{t-1-i}\right)\left(\sum_{j=0}^{\infty} a^j \varepsilon_{t+\Delta t-1-j}\right)$$

$\cancel{t}-\cancel{1}-i = \cancel{t}+\Delta t-\cancel{1}-j \Rightarrow j = i+\Delta t$

$$\begin{pmatrix} \cdots & a^2\varepsilon_{t-3} & a\varepsilon_{t-2} & \varepsilon_{t-1} & & & \\ \cdots & a^{\Delta t+2}\varepsilon_{t-3} & a^{\Delta t+1}\varepsilon_{t-2} & a^{\Delta t}\varepsilon_{t-1} & \cdots & a\varepsilon_{t+\Delta t-2} & \varepsilon_{t+\Delta t-1} \end{pmatrix}$$

$$= \sum_{i=0}^{\infty} a^i a^{i+\Delta t} E[\varepsilon_{t-1-i}^2] = \sum_{i=0} a^{2i+\Delta t} \sigma^2 = \underline{\underline{\frac{a^{\Delta t}\sigma^2}{1-a^2}}}$$

$$m(t) = 0, \quad k(t,t') = \frac{a^{|t-t'|}\sigma^2}{1-a^2}$$

$$k(t_1,t_2) = \frac{a^{|t_1-t_2|}\sigma^2}{1-a^2} = \begin{array}{r}\\ \\ \\ \\ t_1 \rightarrow \\ t_1-1 \rightarrow \\ t_1-2 \rightarrow \\ t_1-3 \rightarrow \\ \\ \end{array}\begin{pmatrix} \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \sigma \\ a\sigma \\ a^2\sigma \\ \vdots \end{pmatrix}^T \begin{pmatrix} \vdots \\ 0 \\ 0 \\ \sigma \\ a\sigma \\ \vdots \\ a^{\Delta t}\sigma \\ a^{\Delta t+1}\sigma \\ a^{\Delta t+2}\sigma \\ \vdots \end{pmatrix}\begin{array}{l} \\ \\ \leftarrow t_2 \\ \leftarrow t_2-1 \\ \leftarrow t_2-2 \\ \vdots \\ \leftarrow t_2-\Delta t = t_1-1 \\ \leftarrow t_2-\Delta t-1 \\ \leftarrow t_2-\Delta t-2 \\ \\ \end{array}$$

$$= \phi(t_1)^T \phi(t_2)$$

$$\phi(t) \equiv \begin{pmatrix} \vdots \\ 0 \\ 0 \\ \sigma \\ a\sigma \\ a^2\sigma \\ \vdots \end{pmatrix}\begin{array}{l} \vdots \\ \\ \leftarrow t \\ \leftarrow t-1 \\ \leftarrow t-2 \\ \vdots \\ \\ \end{array}$$

Covariance

= Inner product of two vectors in ϕ-space.

<u>Note:</u>

Inner product matrix is a p.d. matrix

$$K \equiv \begin{pmatrix} \phi(x_1)^T\phi(x_1) & \phi(x_1)^T\phi(x_2) & \cdots & \phi(x_1)^T\phi(x_N) \\ \vdots & & & \\ \phi(x_N)^T\phi(x_1) & \cdots & & \phi(x_N)^T\phi(x_N) \end{pmatrix}$$

For a nonzero $\vec{c} \in \mathbb{R}^N$,

$$\vec{c}^T K \vec{c} = \sum_{i,j=1}^{N} c_i K_{ij} c_j = \sum_{i,j=1}^{N} c_i \phi(x_i)^T \phi(x_j) c_j$$

$$= \left(\sum_{i=1}^{N} c_i \phi(x_i) \right)^T \left(\sum_{j=1}^{N} c_j \phi(x_j) \right) > 0$$

$\therefore$ Inner product matrix K is p.d.

$$E[y \mid x, \mathcal{D}] = x^T \left(XX^T + \frac{\sigma^2}{\sigma_0^2} I\right)^{-1} X y \quad \cdots ①$$

$$\underbrace{\left(XX^T + \frac{\sigma^2}{\sigma_0^2} I\right)}_{A} X = XX^T X + \frac{\sigma^2}{\sigma_0^2} X$$

$$= X \underbrace{\left(X^T X + \frac{\sigma^2}{\sigma_0^2} I\right)}_{=B}$$

$$AX = XB$$

$$\cancel{A^{-1}A} X B^{-1} = A^{-1} X \cancel{B B^{-1}}$$

$$X B^{-1} = A^{-1} X$$

$$\Rightarrow X\left(X^T X + \frac{\sigma^2}{\sigma_0^2} I\right)^{-1} = \left(XX^T + \frac{\sigma^2}{\sigma_0^2} I\right)^{-1} X$$

$$\therefore ① = x^T X \left(X^T X + \frac{\sigma^2}{\sigma_0^2} I\right)^{-1} y$$

$$= k^T \left(K + \frac{\sigma^2}{\sigma_0^2} I\right)^{-1} y \quad \Leftarrow \text{Same as GP mean}$$

If $K_{ij} = x_i^T x_j$, $k_i = x^T x_i$.

When we consider GP with covariance

$$k(t_1, t_2) = \frac{a^{|t_1 - t_2|} \sigma^2}{1 - a^2}, \quad \text{GP mean after}$$

observation is the predictive mean of the linear Bayesian model in the $\phi(t)$ space, with $y = w^T \phi(x)$ and Gaussian prior on w.

<u>Note:</u> Inner product of two $\phi(x_1), \phi(x_2)$ is the covariance between $y(x_1)$ and $y(x_2)$,

$$k(x_1, x_2) = E[y(x_1) y(x_2)] - E[y(x_1)] E[y(x_2)]$$

Ex) $k(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right)$

$$\left(\exp(t) = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots \right)$$

$$= \exp\left(-\frac{1}{2\sigma^2}\{x_1^2 + x_2^2 - 2x_1^T x_2\}\right)$$

$$= \frac{\exp\left(\frac{x_1^T x_2}{\sigma^2}\right)}{\exp\left(\frac{x_1^2}{\sigma^2}\right)^{\frac{1}{2}} \exp\left(\frac{x_2^2}{\sigma^2}\right)^{\frac{1}{2}}}$$

$$= \frac{1 + \frac{x_1^T x_2}{\sigma^2} + \frac{1}{2!}\left(\frac{x_1^T x_2}{\sigma^2}\right)^2 + \cdots}{\sqrt{1 + \frac{x_1^2}{\sigma^2} + \frac{1}{2!}\left(\frac{x_1^2}{\sigma^2}\right)^2 + \cdots}\sqrt{1 + \frac{x_2^2}{\sigma^2} + \frac{1}{2!}\left(\frac{x_2^2}{\sigma^2}\right)^2 + \cdots}}$$

$$\phi(x) = \begin{pmatrix} 1 \\ | \\ x \\ | \\ x_1^2 \\ x_2^2 \\ \vdots \\ \sqrt{2}x_1x_2 \\ \sqrt{2}x_1x_3 \\ \vdots \\ \vdots \end{pmatrix}$$

$$= \frac{\phi(x_1)^T \phi(x_2)}{\|\phi(x_1)\| \, \|\phi(x_2)\|}$$

$$k(x_1, x_2) = \frac{\phi(x_1)}{\|\phi(x_1)\|}^T \frac{\phi(x_2)}{\|\phi(x_2)\|}$$

$$\mathrm{Var}[y \mid x, \mathcal{D}] = \sigma^2 + \sigma_0^2 k(x,x) - \sigma_0^2 k^T \left(K + \frac{\sigma^2}{\sigma_0^2} I\right)^{-1} k$$

Variance의 이것만
GP와 다름.

GP에서 observation noise를
줘서 $k(x,x') \Leftarrow k(x,x') + \frac{\sigma^2}{\sigma_0^2}\delta_{x,x'}$
으로 수정해서 사용하면
완전히 GP와 동일.