Lec7 Gaussian Process

Gaussian Process

$X(t)$ Continuous Time Continuous States

$\forall n, \forall t_1, \cdots, t_n. \ (X(t_1), \cdots, X(t_n))^T = X \Rightarrow X \sim N(\mu, \Sigma)$ $\boxed{\text{Joint Gaussian}}$ 联合高斯分布

$n=1, \ X \sim N(\mu, \sigma^2), \ f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

$n=2, \ (X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$,

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right)$$

$$X \in \mathbb{R}^n, \ X \sim N(\mu, \Sigma), \ f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\underbrace{(x-\mu)^T}_{1\times n}\underbrace{\Sigma^{-1}}_{n\times n}\underbrace{(x-\mu)}_{n\times 1}\right)$$

$E(X) = \mu, \ \Sigma = E(X-\mu)(X-\mu)^T$ (Covariance Matrix)

①

$f_X(x) \geq 0$

对角阵 $\Rightarrow -\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)$

$dx_1, \cdots, dx_n$

$$② \int_{\mathbb{R}^n} f_X(x)\,dx = \frac{1}{(2\pi)^{\frac{n}{2}}(\det\Sigma)^{\frac{1}{2}}} \int_R \exp\left(\underline{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}\right)dx$$

$\Sigma = \Sigma^T \geq 0, \ \Sigma = U^T\Lambda U, \ \Lambda = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_k^2), \ UU^T = U^TU = I$

$\Sigma^{-1} = U^T\Lambda^{-1}U, \ \Lambda^{-1} = \mathrm{diag}\left(\frac{1}{\sigma_1^2}, \cdots, \frac{1}{\sigma_k^2}\right), \ y = U\cdot(x-\mu)$

$dx \to dy \qquad x = U^Ty + \mu$

$dx = J \cdot dy$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}(\det\Sigma)^{\frac{1}{2}}} \int \exp\left(-\frac{1}{2}y^T\Lambda^{-1}y\right)dy$$

$(\sqrt{2\pi})^n \prod_{k=1}^{n}\sigma_k \qquad \sum_{k=1}^{n}\frac{y_k^2}{2\sigma_k^2}$

$$= \prod_{k=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma_k}\int_{-\infty}^{+\infty}\exp\left(-\frac{y_k^2}{2\sigma_k^2}\right)dy_k\right)$$

Characteristic Functions (特征函数) $\phi_X(\omega) = E(\exp(i\omega X))$ $\sum_{k=1}^{n}\omega_k X_k$

$$\phi_X(\omega) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det\Sigma)^{\frac{1}{2}}}\int_R \exp(i\omega^T x)\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)dx$$ 配方

$$-\frac{1}{2\sigma^2}(x-\mu)^2 + i\omega x = -\frac{1}{2\sigma^2}\left[x-(\mu+i\sigma^2\omega)\right]^2 + i\omega\mu - \frac{1}{2}\sigma^2\omega^2$$

$$\Rightarrow -\frac{1}{2}(x-\mu-i\Sigma\omega)^T\Sigma^{-1}(x-\mu-i\Sigma\omega) + i\omega^T\mu - \frac{1}{2}\omega^T\Sigma\omega$$

$\sim N(\mu + i\Sigma\omega, \sigma^2)$

$$= \exp\left(i\omega^T\mu - \frac{1}{2}\omega^T\Sigma\omega\right)\underbrace{\frac{1}{(2\pi)^{\frac{n}{2}}(\det\Sigma)^{\frac{1}{2}}}\int_{R^n}\exp\left(-\frac{1}{2}(x-\mu-i\Sigma\omega)^T\Sigma^{-1}(x-\mu-i\Sigma\omega)\right)dx}_{1}$$

$$\boxed{\phi_X(\omega) = \exp\left(i\omega^T\mu - \frac{1}{2}\omega^T\Sigma\omega\right)}$$ ☆

Linearity $X \in R^n$ $X \sim N(\mu, \Sigma)$, $A \in \mathbb{R}^{m\times n}$,

$$Y = AX \Rightarrow Y \sim N(A\mu, A\Sigma A^T)$$

$$\phi_Y(w) = E[\exp(iw^T Y)] = E\left(\exp(\underbrace{iw^T A x}_{\overset{i(A^Tw)^Tx}{\shortparallel}})\right) = E\left(\exp(i(A^Tw)^Tx)\right) = \phi_x(A^Tw)$$

$$= \exp\left[i(A^Tw)^T\mu - \frac{1}{2}(A^Tw)^T\Sigma(A^Tw)\right]$$

$$= \exp\left(iw^T\boxed{A\mu} - \frac{1}{2}w^T\boxed{A\Sigma A^T}w\right)$$

联合分布是 Gaussian 分布，边缘分布一定是 Gaussian 分布

但边缘分布是 Gaussian，反之联合不一定是 Gaussian

$(X_1, \cdots, X_n)^T \sim N$，$\forall \{n_1, \cdots, n_k\} \subseteq \{1, \cdots, n\}$，$\{X_{n_1}, \cdots, X_{n_k}\} \sim N$

$$\begin{pmatrix} X_{n_1} \\ \vdots \\ X_{n_k} \end{pmatrix} = \begin{pmatrix} 0 \cdots n_1, \cdots 0 \\ 0 \cdots n_2 \cdots 0 \\ \\ 0 \cdots\cdots n_k \cdot 0 \end{pmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

Boundary of Gaussian must be Gaussian

保持边缘 Gaussian，破坏联合 Gaussian

(X_1, X_2) $\quad f_{X_1,X_2}(X_1, X_2) = \frac{1}{2\pi}\exp\left(-\frac{X_1^2+X_2^2}{2}\right) + C(x,y) \cdot (\sin x \sin y + 1)$

$\int_R C(x,y)\,dx = \int_R C(x,y)\,dy = 0$, $C(x,y) = \sin x \sin y$

Sufficient

① $X_1, \cdots, X_n$ independent $X_k \sim N(\mu_k, \sigma_k^2) \Rightarrow (X_1, \cdots, X_n) \sim N(\mu, \Sigma)$

$$\mu = (\mu_1, \cdots, \mu_n)^T, \quad \Sigma = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_n^2)$$

② $X_1, \cdots, X_n \sim N \iff \boxed{\forall}\alpha \in \mathbb{R}^n$, $\alpha^T X \sim N$，(独立条件有些强) (一维高斯)

"$\Leftarrow$" $\forall \alpha$, $\phi_{\alpha^TX}(w) = \exp\left(iw\mu_{\alpha^TX} - \frac{1}{2}w^2\sigma_{\alpha^TX}^2\right)$ $\quad \alpha^TX$ 特征函数 $\quad \phi_X(\alpha) = E[\exp(i\alpha^TX)]$

$$= E[\exp(i\alpha^TXw)] \xrightarrow{\text{Let } w=1} E[\exp(i\alpha^TX)]$$

$\phi_X(\alpha) = \phi_{\alpha^TX}(1) = \exp\left(i\mu_{\alpha^TX} - \frac{1}{2}\sigma_{\alpha^TX}^2\right)$, $E(X) = \mu$

$$\mu_{\alpha^TX} = E(\alpha^TX) = \alpha^T\mu$$

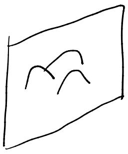
$$\sigma_{\alpha^TX}^2 = E(\alpha^TX - \alpha^T\mu)(\alpha^TX - \alpha^T\mu)^T = \underbrace{\alpha^T E[(X-\mu)(X-\mu)^T]\alpha}_{\alpha^T\Sigma\alpha}$$

$$\Rightarrow \Phi_X(\alpha) = \exp\left(i\alpha^T\mu - \frac{1}{2}\alpha^T\Sigma\alpha \right)$$

DDPM (2015) Denoise Diffusion Probabilistic Model

Sampling
⇓
Simpler $\boxed{\text{Random Number Generator}}$

Computer Drawing ⟺ Random Number Sampling



如何产生一座山? 连续加噪声的过程 (Direct Diffusion)

$$\xrightarrow{+N_1} x_1 \xrightarrow{+N_2} x_2 \xrightarrow{+N_3} x_3 \longrightarrow \cdots \longrightarrow x_T \sim N(0, \sigma^2 I)$$

$$x_0 \cdots\cdots \xleftarrow[NN]{} x_{t-1} \xleftarrow[NN]{} x_t \xleftarrow[NN]{} \cdots \xleftarrow[NN]{} x_T \sim N(0, \sigma^2 I)$$

NN (神经网络)

Reverse inference

Parameter Trick

① $x_t = \boxed{\sqrt{\alpha_t}}\, x_{t-1} + \boxed{\sqrt{(1-\alpha_t)}}\, \varepsilon \qquad \varepsilon \sim N(0, I)$

从 t-1 到 t:

$$\boxed{x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{(1-\alpha_t)}\, \varepsilon} = \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}\, x_{t-2} + \sqrt{(1-\alpha_{t-1})}\, \varepsilon_1\right) + \sqrt{(1-\alpha_t)}\, \varepsilon_2$$

(*)

$$= \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \underbrace{\sqrt{\alpha_t} \cdot \sqrt{1-\alpha_{t-1}}\, \varepsilon_1 + \sqrt{1-\alpha_t}\, \varepsilon_2}_{\text{噪声项}}$$

$\alpha_t(1-\alpha_{t-1}) + 1 - \alpha_t$

$= 1 - \alpha_t \alpha_{t-1}$

高斯过程做线性变换

$$= \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + (1 - \alpha_t \alpha_{t-1})\, \varepsilon$$

$$= \sqrt{\alpha_t \alpha_{t-1} \alpha_{t-2}}\, x_{t-3} + (1 - \alpha_t \alpha_{t-1} \alpha_{t-2})\, \varepsilon$$

$$= \cdots$$

$$= \sqrt{\prod_{k=1}^{t} \alpha_k}\, x_0 + \sqrt{1 - \prod_{k=1}^{t} \alpha_k}\, \varepsilon \qquad \bar{\alpha}_t = \prod_{k=1}^{t} \alpha_k$$

$$\Rightarrow \boxed{x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1 - \bar{\alpha}_t}\, \varepsilon} \quad (\star)$$

$\Rightarrow$ 噪声(Noise) ε Can be expressed by x_0 / x_t

Denoise 反向过程

三个分布都知道

② $$P(x_{t-1} \mid x_t, x_0) = P(x_t \mid x_{t-1}, x_0) \cdot \frac{P(x_{t-1} \mid x_0)}{P(x_t \mid x_0)}$$

主变量 作为参数出现 (Markov 与 x_0 无关)

不能直接从 t 到 t-1:

忽略 x_0 $\quad P(x_{t-1} \mid x_t) = P(x_t \mid x_{t-1}) \cdot \frac{P(x_{t-1})}{P(x_t)}$

$$P(x_t \mid x_{t-1}) = \frac{P(x_t\, x_{t-1})}{P(x_{t-1})}$$

[Bayes formula]
贝叶斯公式

$P(x_t \mid x_{t-1}, x_0) \sim N(\sqrt{\alpha_t}\, x_{t-1}, (1-\alpha_t) I)$ 由上式(*)得

$P(x_t \mid x_0) \sim N(\sqrt{\bar{\alpha}_t}\, x_0, (1-\bar{\alpha}_t) I)$ 由上式(☆)得

$P(x_{t-1} \mid x_0) \sim N(\sqrt{\bar{\alpha}_{t-1}}\, x_0, (1-\bar{\alpha}_{t-1}) I)$

Focus on 指数上方二次形:

$$-\frac{1}{2}\left(\frac{(x_t - \sqrt{\alpha_t}\, x_{t-1})^2}{1-\alpha_t} - \frac{(x_t - \sqrt{\bar{\alpha}_t}\, x_0)^2}{1-\bar{\alpha}_t} + \frac{(x_t - \sqrt{\bar{\alpha}_{t-1}}\, x_0)^2}{1-\bar{\alpha}_{t-1}} \right)$$

$$= -\frac{1}{2}\left(\frac{x_t^2 - 2\sqrt{\alpha_t}\, x_{t-1} x_t + \alpha_t x_{t-1}^2}{1-\alpha_t} - \frac{x_t^2 - 2\sqrt{\bar{\alpha}_t}\, x_0 x_t + \bar{\alpha}_t x_0^2}{1-\bar{\alpha}_t} + \frac{x_{t-1}^2 - 2\sqrt{\bar{\alpha}_{t-1}}\, x_0 x_{t-1} + \bar{\alpha}_{t-1} x_0^2}{1-\bar{\alpha}_{t-1}} \right)$$

留下与 x_{t-1} 主变量有关的

$$= -\frac{1}{2}\left(\left(\frac{\alpha_t}{1-\alpha_t} + \frac{1}{1-\bar{\alpha}_{t-1}} \right) x_{t-1}^2 - 2x_{t-1}\left(\frac{\sqrt{\alpha_t}\, x_t}{1-\alpha_t} + \frac{\sqrt{\bar{\alpha}_{t-1}}\, x_0}{1-\bar{\alpha}_{t-1}} \right) + C(x_0, x_t) \right)$$

高斯精神：只关心二次项（配方思想）

$\bar{\alpha}_{t-1} = \prod_{k=1}^{t-1} \alpha_k \Rightarrow \bar{\alpha}_{t-1} \cdot \alpha_t = \bar{\alpha}_t$

$$Var = \left(\frac{\alpha_t}{1-\alpha_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right)^{-1} = \left(\frac{\alpha_t(1-\bar{\alpha}_{t-1}) + (1-\alpha_t)}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}\right)^{-1} = \left(\frac{1-\bar{\alpha}_{t-1}\alpha_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}\right)^{-1} = \left(\frac{1-\bar{\alpha}_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}\right)^{-1} = (1-\alpha_t)\text{(某个因子)}$$

$$mean = \frac{\dfrac{\sqrt{\alpha_t}\,x_t}{1-\alpha_t} x_t + \dfrac{\sqrt{\bar{\alpha}_{t-1}}\,x_0}{1-\bar{\alpha}_{t-1}}}{\dfrac{1-\bar{\alpha}_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t} \underline{x_0}$$

Ground truth

$\boxed{x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\,\varepsilon}$ 替换 x_0

$$② \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}\left(\frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \sqrt{1-\bar{\alpha}_t}\,\varepsilon\right)\right)$$

$$= \left[\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}}\right] \underline{x_t} - \left(\frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t} \cdot \sqrt{1-\bar{\alpha}_t}\right) \underline{\varepsilon}$$

$\alpha_t \cdot \sqrt{\bar{\alpha}_{t-1}}$

$$= \left(\frac{\sqrt{\alpha_t}\sqrt{\bar{\alpha}_t}(1-\bar{\alpha}_{t-1}) + \sqrt{\bar{\alpha}_{t-1}} - \alpha_t\sqrt{\bar{\alpha}_{t-1}}}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}}\right) x_t - \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{\sqrt{1-\bar{\alpha}_t}}\,\varepsilon$$

$$-\alpha_t\sqrt{\bar{\alpha}_{t-1}}\,\bar{\alpha}_{t-1} + \sqrt{\bar{\alpha}_{t-1}} = -\sqrt{\bar{\alpha}_{t-1}}\,\bar{\alpha}_t + \sqrt{\bar{\alpha}_{t-1}}$$

$$\Rightarrow \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\bar{\alpha}_t)}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}} = \frac{1}{\sqrt{\alpha_t}}$$

$$= \boxed{\frac{1}{\sqrt{\alpha_t}} \cdot x_t - \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{\sqrt{1-\bar{\alpha}_t}} \cdot \varepsilon}$$

mean