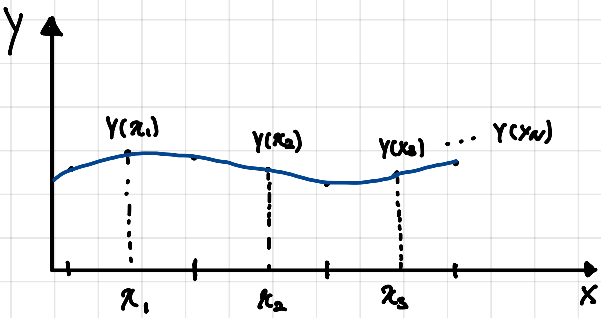
Review

Gaussian Process

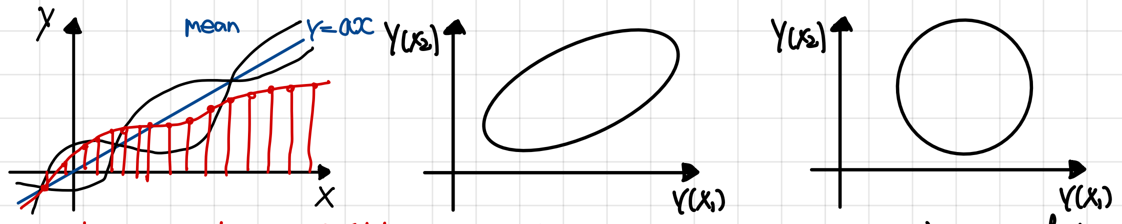
- Infinite dimentional Gaussian
- Subset of variables follow Gaussian
- Distribution of functions



- $Y(x_1), Y(x_2), Y(x_3)$ are jointly distribution
- $\mu(x) = E[Y(x)]$: mean function
- $\Sigma(x_1, x_2) = Cov(Y(x_1), Y(x_2))$
  → Covariance function

$GP(\mu, \Sigma)$: Gaussian Process of mean function $\mu(x)$ and Covariance function $\Sigma(x_1, x_2)$

EX) $m(x) = ax$, $k(x_1, x_2) = \exp\left(-\frac{|x_1 - x_2|^2}{\sigma^2}\right)$ variance에 $x_1 - x_2$ relation이 들어간다.



→ 이 함수 mean과 variance를 정한다. x_1 and x_2 close x_1 and x_2 are far

EX) $y = f(x) = w^T \phi(x)$ Random function

Sampling random w makes Random function!

Random!

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_D \end{bmatrix} \in \mathbb{R}^D, \quad \underline{w \sim \mathcal{N}(0, \alpha^{-1} I)}, \quad \phi = \begin{bmatrix} \phi_1(x) \\ \phi_2(x) \\ \vdots \\ \phi_D(x) \end{bmatrix} : \text{Basis function}$$

Gaussian Random vector

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} w^T\phi(x_1) \\ w^T\phi(x_2) \\ \vdots \\ w^T\phi(x_N) \end{bmatrix}, \quad \Phi = [\,\phi(x_1)\ \phi(x_2)\ \cdots\ \phi(x_N)\,] \in \mathbb{R}^{D \times N}$$

$Y = \Phi^T W$ (Note Y is a Gaussian)

$E[Y] = \Phi^T \underbrace{E[W]}_{0} = 0$, $Cov[Y] = E[YY^T] - \underbrace{E[Y]E[Y]^T}_{0}$ (Definition)

$= E[\Phi^T W W^T \Phi] = \Phi^T \underbrace{E[WW^T]}_{= \frac{1}{\alpha} I} \Phi$ (w's variance)

$= \frac{1}{\alpha} \Phi^T \Phi = K \in \mathbb{R}^{N \times N}$ $\left(K_{ij} = \underline{k(x_i, x_j)} = \frac{1}{\alpha} \Phi^T(x_i) \cdot \Phi(x_j)\right)$

Basis function에 의해 결정됨

Ex) $y(x) = w^T \phi(x)$

연관이 있다.

$$\bigcirc \to \underset{x_t}{\bigcirc} \to \underset{x_{t+1}}{\bigcirc} \to \bigcirc$$

$x_{t+1} = a x_t + \underset{\text{↳ Gaussian}}{\varepsilon_t}$ $\quad \varepsilon_t \sim N(0, \sigma^2)$ and $x_0 = 0$ $\quad (0 < a < 1)$

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

이 과정에서 x_t to x_0

scholar x Gaussian → Gaussian ⇒ x_t 또한 Gaussian 이다.

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

$$\begin{aligned} Cov(x_t, x_{t+\Delta t}) &= E[x_t x_{t+\Delta t}] - \underbrace{E[x_t]E[x_{t+\Delta t}]}_{0} \\ &= E\left[\left(\sum_{i=0}^{\infty} a^i \varepsilon_{t-1-i}\right)\left(\sum_{i=0}^{\infty} a^i \varepsilon_{t+\Delta t-1-i}\right)\right] \end{aligned}$$

⇒ $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = 0$

If ε_t and $\varepsilon_{t+\Delta t}$ are independent, $E[\varepsilon_t, \varepsilon_{t+\Delta t}] = 0$ for $t = t'$

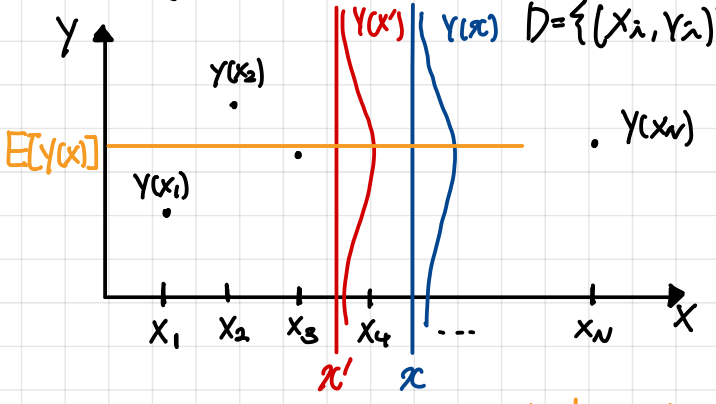
$$\left(\boxed{\varepsilon_{t-1}} \ \boxed{a\varepsilon_{t-2}} \ \boxed{a^2\varepsilon_{t-3}}\right) \times \left(\varepsilon_{t+\Delta t-1} \cdots \ \boxed{a^{\Delta t}\varepsilon_{t-1}} \ \boxed{a^{\Delta t+1}\varepsilon_{t-2}} \ \boxed{a^{\Delta t+2}\varepsilon_{t-3}}\right) \to a^{\Delta t+2i}\varepsilon_{t-1-i}$$

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

등비급수 → $\frac{a^{\Delta t}}{1-a^2}$

$\therefore m(t) = 0$, $k(t_1, t_2) = \frac{\sigma^2 a^{|t_1 - t_2|}}{1-a^2}$

GP regression



$D=\{(x_i, y_i)\}_{i=1}^{N}$

Jointly distribution!

$$y(x) \sim \mathcal{N}(\mu_x, \sigma_x^2) = P(y(x) \mid y(x_1), \cdots, y(x_N))$$

$$(y(x), y(x')) \sim \mathcal{N}(\mu_x, \sigma_x^2) = P(y(x), y(x') \mid y(x_1), \cdots, y(x_N))$$

Using $\begin{cases} \mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b) \\ \Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \end{cases}$ for $P(y(x), y(x') \mid y(x_1), \cdots, y(x_N))$

$m(x) = 0$, $k(x_1, x_2)$

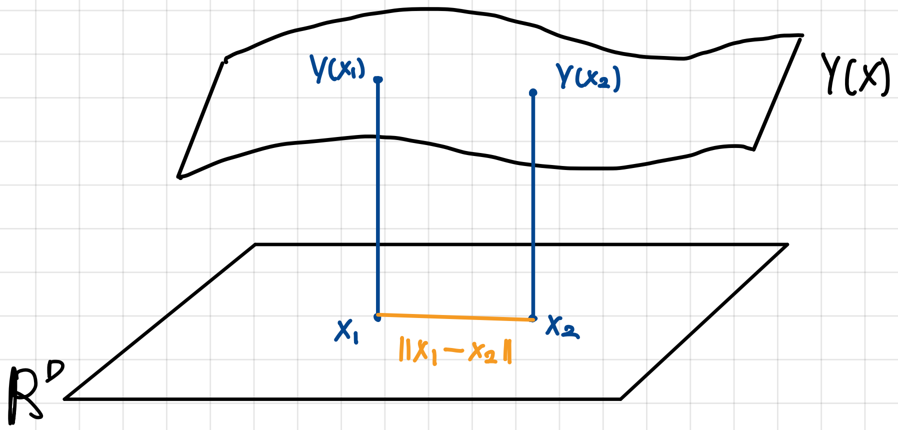
Def $\mathbf{k} \equiv \begin{pmatrix} k(x, x_1) \\ k(x, x_2) \\ \vdots \\ k(x, x_N) \end{pmatrix} \in \mathbb{R}^N$, $K \equiv \begin{pmatrix} k(x_1, x_2) & \cdots & k(x_1, x_N) \\ \vdots & & \vdots \\ k(x_N, x_1) & \cdots & k(x_N, x_N) \end{pmatrix} \in \mathbb{R}^{N \times D}$

Cov(y(x), y(x_1))

Cov(y(x_1), y(x_N))

$$m(y(x) \mid y(x_1), \cdots, y(x_N)) = 0 + \mathbf{k}^T K^{-1}(y - 0) = \underline{\mathbf{k}^T K^{-1} y}$$ Prediction

$$\mathrm{Var}(y(x) \mid y(x_1), \cdots, y(x_N)) = \underline{k(x, x) - \mathbf{k}^T K^{-1} \mathbf{k}}$$ Uncertainty



$k(x_i, x_j) = k(\|x_i - x_j\|)$

Similarity between inputs

Ex) $k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2r^2}\right)$

Ex) $k(x_i, x_j) = \frac{a^{\|x_i - x_j\|^2}}{1 - a^2} \cdot \sigma^2$

$(0 < a < 1)$