Multiple-output Gaussian processes

Mauricio A. Álvarez, PhD

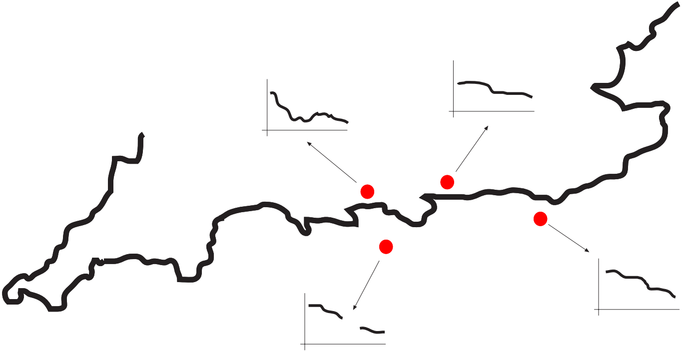
Curso de entrenamiento ArcelorMittal

Sensor Network

South Coast of England

Sensor Network

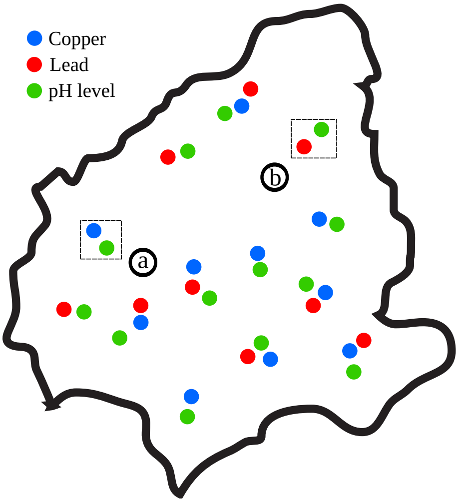
South Coast of England



Sensor location

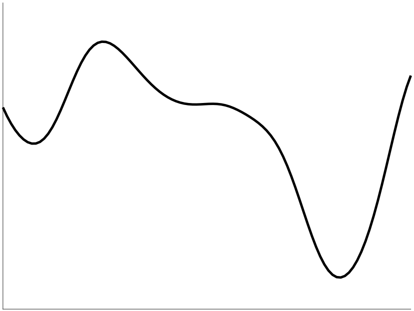
Jura Data Set

Region of
Swiss Jura

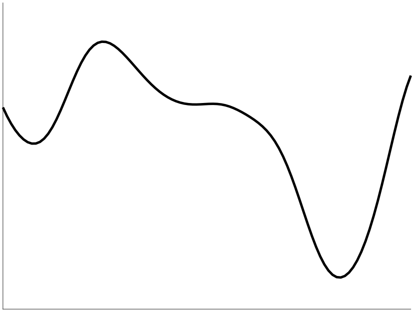


Contents

Single-output Gaussian process

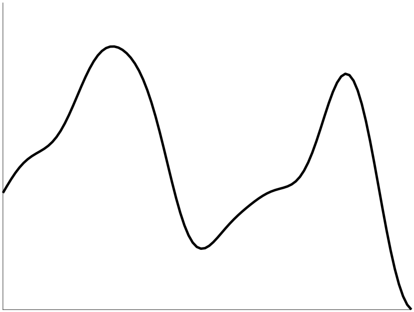


Single-output Gaussian process



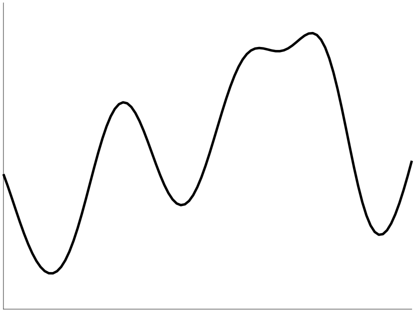
$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

Single-output Gaussian process



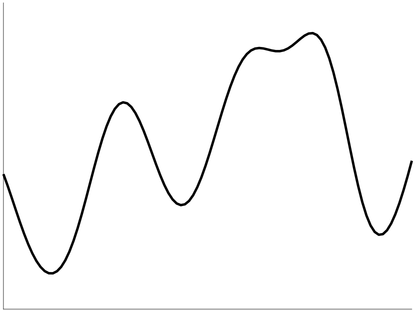
$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

Single-output Gaussian process



$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

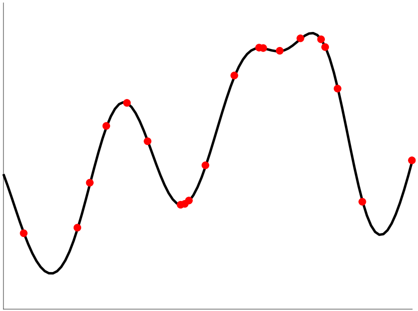
Single-output Gaussian process



$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D} = \{(\mathbf{x}_i, f(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

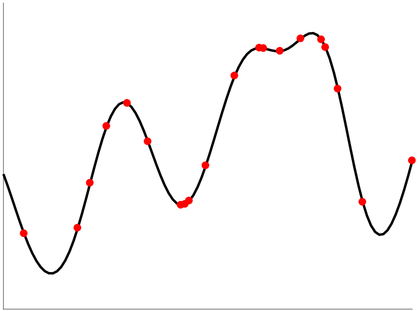
Single-output Gaussian process



$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D} = \{(\mathbf{x}_i, f(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

Single-output Gaussian process

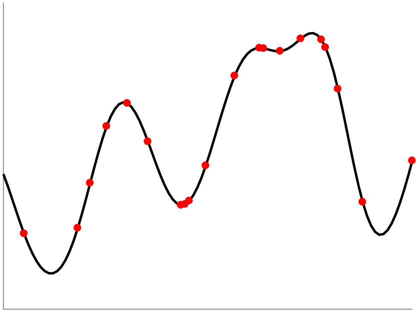


$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D} = \{(\mathbf{x}_i, f(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

$$\begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} \right)$$

Single-output Gaussian process

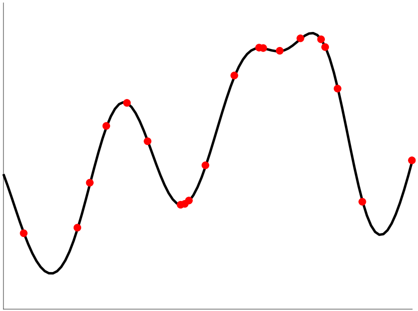


$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D} = \{(\mathbf{x}_i, f(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

$$\underbrace{\begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{bmatrix}}_{\mathbf{f}} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} \right)$$

Single-output Gaussian process

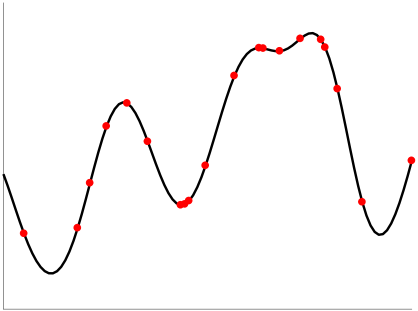


$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D} = \{(\mathbf{x}_i, f(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

$$\underbrace{\begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{bmatrix}}_{\mathbf{f}} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \underbrace{\begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}}_{\mathbf{K}} \right)$$

Single-output Gaussian process

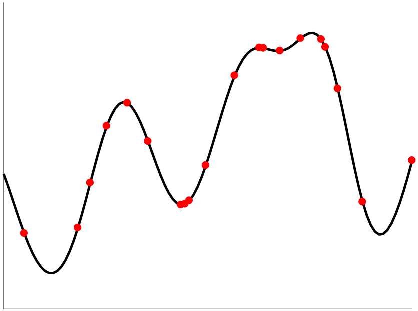


$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D} = \{(\mathbf{x}_i, f(\mathbf{x}_i)) | i = 1, \dots, N\}$$

$$\underbrace{\begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{bmatrix}}_{\mathbf{f}} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{0}}, \underbrace{\begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}}_{\mathbf{K}} \right)$$

Single-output Gaussian process



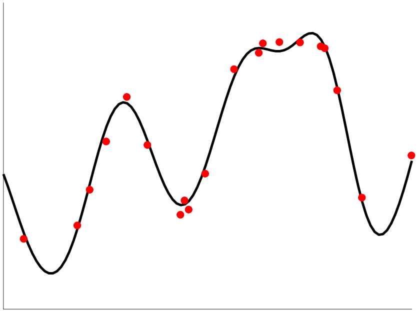
$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D} = \{(\mathbf{x}_i, f(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

$$\underbrace{\begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{bmatrix}}_{\mathbf{f}} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{0}}, \underbrace{\begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}}_{\mathbf{K}} \right)$$

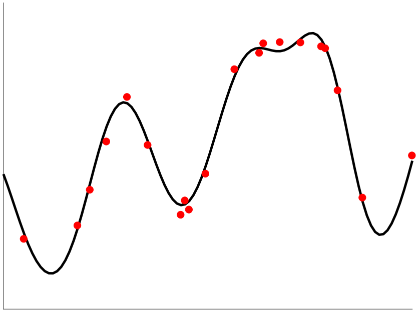
For prediction: $p(f(\mathbf{x}_*)|\mathbf{f})$

Single-output Gaussian process



$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

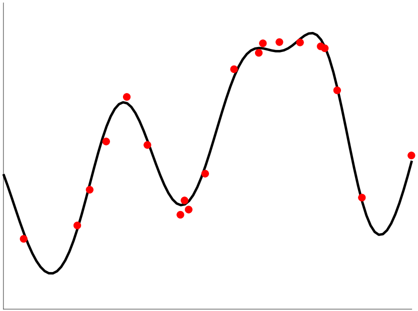
Single-output Gaussian process



$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

$$y(\mathbf{x}_i) = f(\mathbf{x}_i) + \epsilon_i$$

Single-output Gaussian process

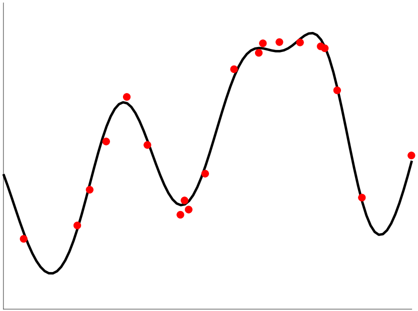


$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

$$y(\mathbf{x}_i) = f(\mathbf{x}_i) + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Single-output Gaussian process



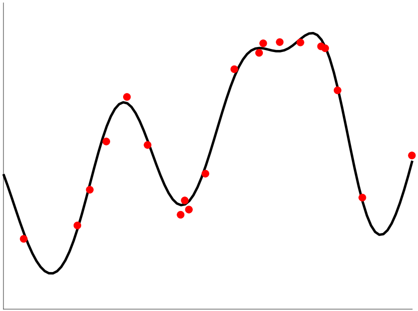
$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

$$y(\mathbf{x}_i) = f(\mathbf{x}_i) + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$\mathcal{D} = \{(\mathbf{x}_i, y(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

Single-output Gaussian process



$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

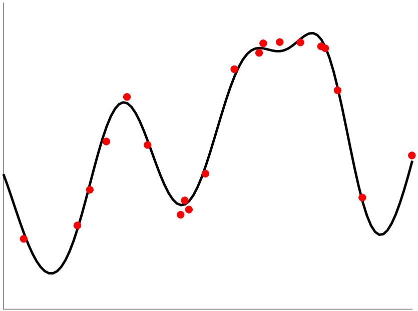
$$y(\mathbf{x}_i) = f(\mathbf{x}_i) + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$\mathcal{D} = \{(\mathbf{x}_i, y(\mathbf{x}_i))|i = 1, \ldots, N\}$$

$$\begin{bmatrix} y(\mathbf{x}_1) \\ \vdots \\ y(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} + \sigma^2 \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}\right)$$

Single-output Gaussian process



$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

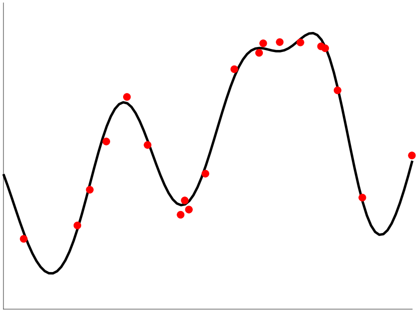
$$y(\mathbf{x}_i) = f(\mathbf{x}_i) + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$\mathcal{D} = \{(\mathbf{x}_i, y(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

$$\underbrace{\begin{bmatrix} y(\mathbf{x}_1) \\ \vdots \\ y(\mathbf{x}_N) \end{bmatrix}}_{\mathbf{y}} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} + \sigma^2 \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \right)$$

Single-output Gaussian process



$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

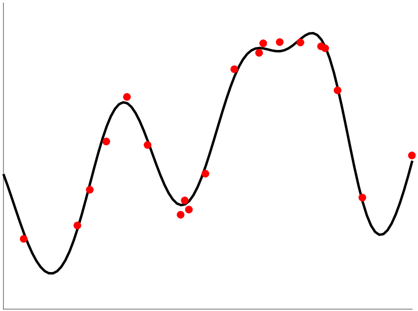
$$y(\mathbf{x}_i) = f(\mathbf{x}_i) + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$\mathcal{D} = \{(\mathbf{x}_i, y(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

$$\underbrace{\begin{bmatrix} y(\mathbf{x}_1) \\ \vdots \\ y(\mathbf{x}_N) \end{bmatrix}}_{\mathbf{y}} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \underbrace{\begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}}_{\mathbf{K}} + \sigma^2 \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \right)$$

Single-output Gaussian process



$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

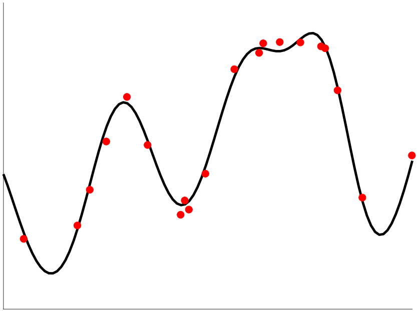
$$y(\mathbf{x}_i) = f(\mathbf{x}_i) + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$\mathcal{D} = \{(\mathbf{x}_i, y(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

$$\underbrace{\begin{bmatrix} y(\mathbf{x}_1) \\ \vdots \\ y(\mathbf{x}_N) \end{bmatrix}}_{\mathbf{y}} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \underbrace{\begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}}_{\mathbf{K}} \underbrace{+}_{+} \sigma^2 \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \right)$$

Single-output Gaussian process



$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

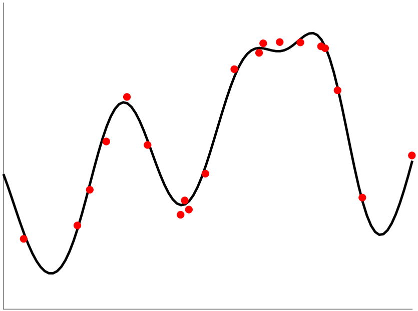
$$y(\mathbf{x}_i) = f(\mathbf{x}_i) + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$\mathcal{D} = \{(\mathbf{x}_i, y(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

$$\underbrace{\begin{bmatrix} y(\mathbf{x}_1) \\ \vdots \\ y(\mathbf{x}_N) \end{bmatrix}}_{\mathbf{y}} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \underbrace{\begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}}_{\mathbf{K}} \underbrace{+}_{+} \underbrace{\sigma^2 \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}}_{\sigma^2 \mathbf{I}} \right)$$

Single-output Gaussian process



$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

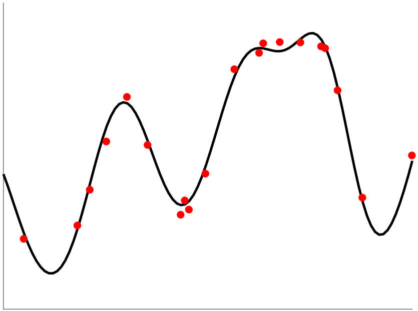
$$y(\mathbf{x}_i) = f(\mathbf{x}_i) + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$\mathcal{D} = \{(\mathbf{x}_i, y(\mathbf{x}_i)) | i = 1, \dots, N\}$$

$$\underbrace{\begin{bmatrix} y(\mathbf{x}_1) \\ \vdots \\ y(\mathbf{x}_N) \end{bmatrix}}_{\mathbf{y}} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{0}}, \underbrace{\begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}}_{\mathbf{K}} \underbrace{+}_{+} \underbrace{\sigma^2 \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}}_{\sigma^2 \mathbf{I}} \right)$$

Single-output Gaussian process



$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

$$y(\mathbf{x}_i) = f(\mathbf{x}_i) + \epsilon_i$$

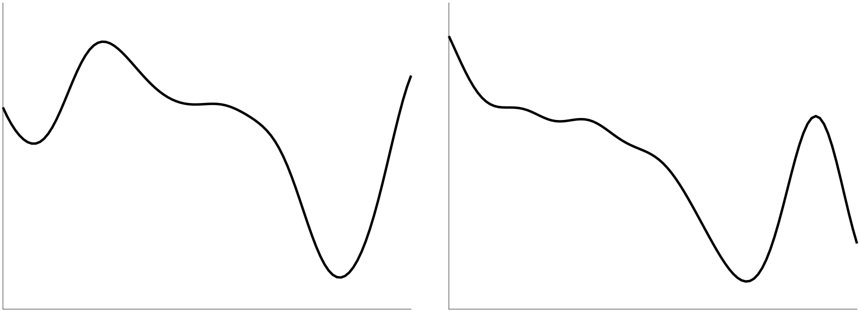
$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$\mathcal{D} = \{(\mathbf{x}_i, y(\mathbf{x}_i)) | i = 1, \dots, N\}$$

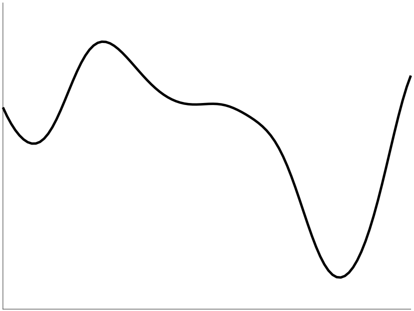
$$\underbrace{\begin{bmatrix} y(\mathbf{x}_1) \\ \vdots \\ y(\mathbf{x}_N) \end{bmatrix}}_{\mathbf{y}} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{0}}, \underbrace{\begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}}_{\mathbf{K}} \underbrace{+}_{+} \underbrace{\sigma^2 \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}}_{\sigma^2 \mathbf{I}} \right)$$

For prediction: $p(f(\mathbf{x}_*)|\mathbf{y})$

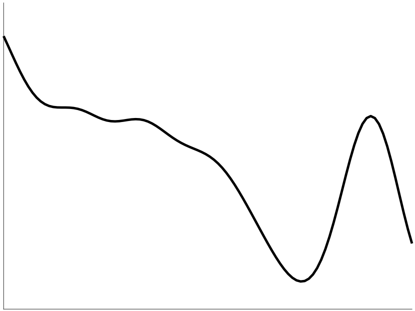
Multiple-output Gaussian process



Multiple-output Gaussian process

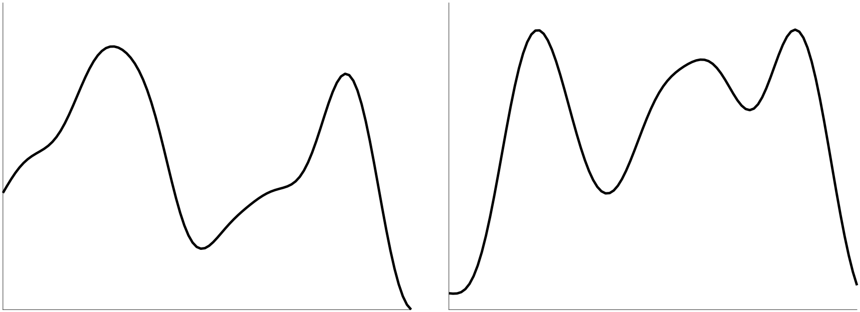


$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$



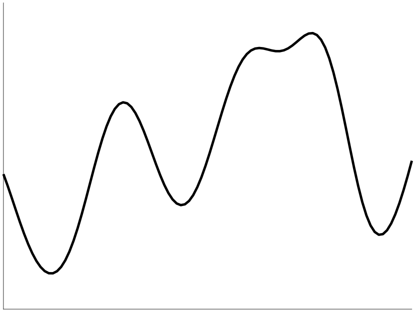
$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

Multiple-output Gaussian process



$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$ $f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

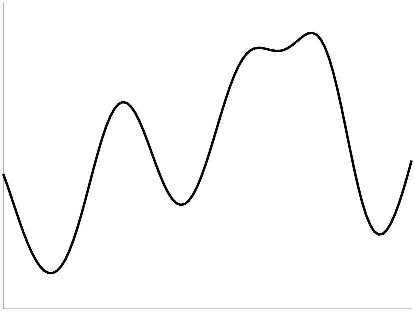
Multiple-output Gaussian process



$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$

$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

Multiple-output Gaussian process

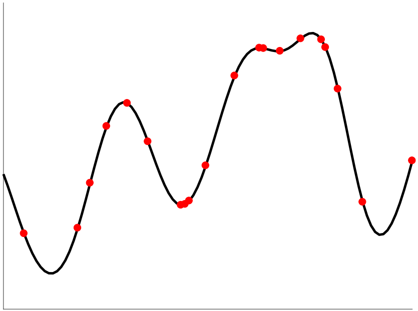


$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$

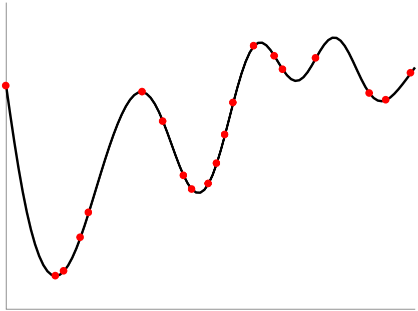
$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))|i = 1, \ldots, N_1\}$ $\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$

Multiple-output Gaussian process



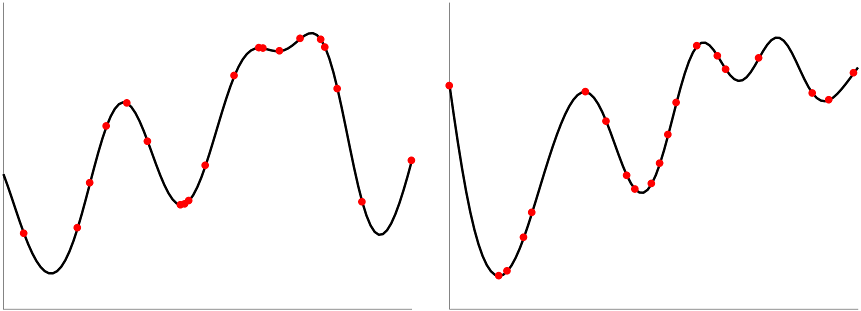
$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$



$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))|i = 1, \ldots, N_1\}$ $\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$

Multiple-output Gaussian process

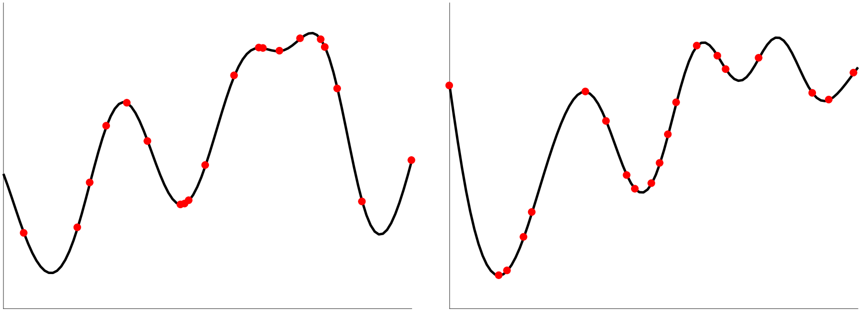


$$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}')) \qquad f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))|i = 1, \ldots, N_1\} \quad \mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$$

$$\mathbf{f}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_1) \qquad \mathbf{f}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_2)$$

Multiple-output Gaussian process



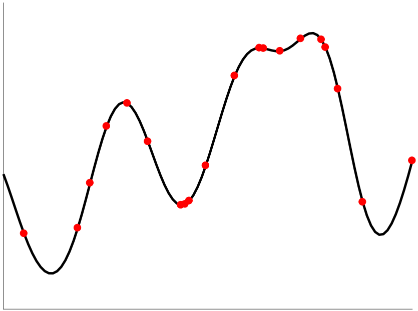
$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$ $\quad$ $f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

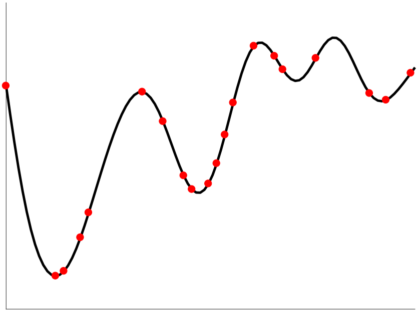
$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))|i = 1, \ldots, N_1\}$ $\quad$ $\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$

$\mathbf{f}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_1)$ $\quad$ $\mathbf{f}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_2)$

$$\begin{bmatrix}\mathbf{f}_1\\ \mathbf{f}_2\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{K}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{K}_2\end{bmatrix}\right)$$

Multiple-output Gaussian process





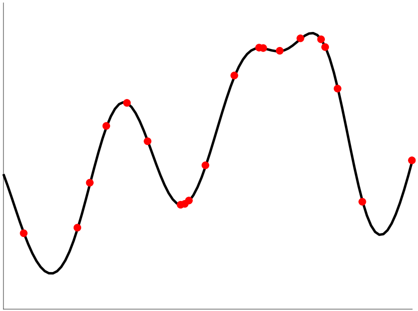
$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$ $\quad$ $f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

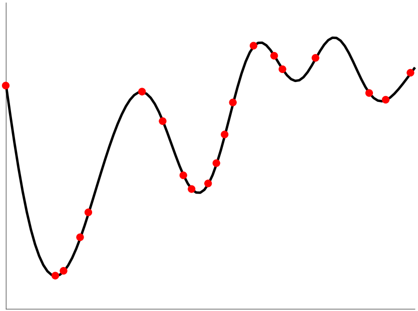
$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))|i = 1, \ldots, N_1\}$ $\quad$ $\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$

$\mathbf{f}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_1)$ $\quad$ $\mathbf{f}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_2)$

$$\underbrace{\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix}}_{\mathbf{f}} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{K}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_2 \end{bmatrix}\right)$$

Multiple-output Gaussian process





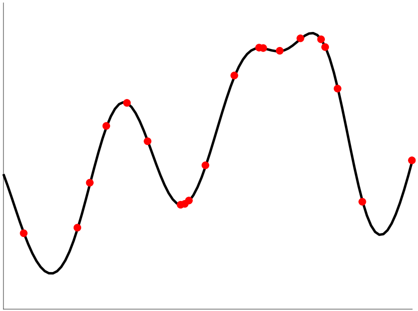
$$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}')) \qquad f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$$

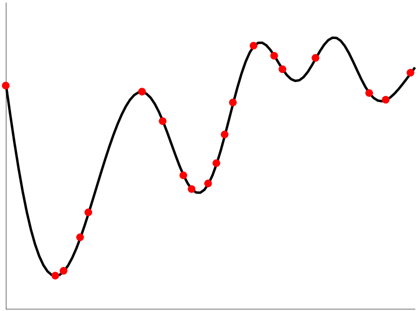
$$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))|i = 1, \ldots, N_1\} \quad \mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$$

$$\mathbf{f}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_1) \qquad \mathbf{f}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_2)$$

$$\underset{\mathbf{f}}{\begin{bmatrix}\mathbf{f}_1 \\ \mathbf{f}_2\end{bmatrix}} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0} \\ \mathbf{0}\end{bmatrix}, \underset{\mathbf{K}_{\mathbf{f},\mathbf{f}}}{\begin{bmatrix}\mathbf{K}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_2\end{bmatrix}}\right)$$

Multiple-output Gaussian process





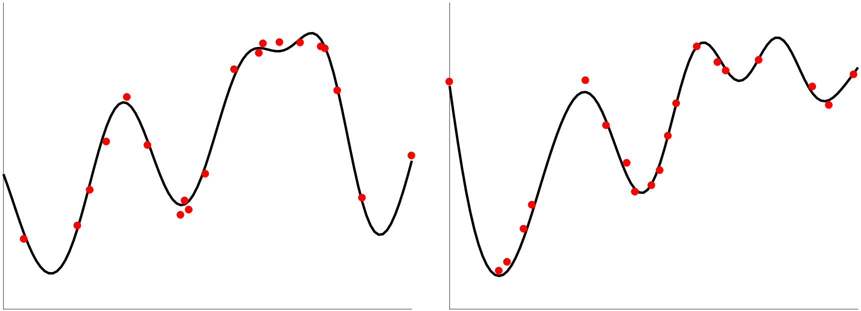
$$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}')) \qquad f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))|i = 1, \ldots, N_1\} \quad \mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$$

$$\mathbf{f}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_1) \qquad \mathbf{f}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_2)$$

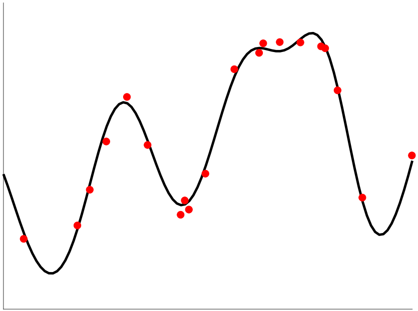
$$\underbrace{\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix}}_{\mathbf{f}} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}}_{\mathbf{0}}, \underbrace{\begin{bmatrix} \mathbf{K}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_2 \end{bmatrix}}_{\mathbf{K}_{\mathbf{f},\mathbf{f}}} \right)$$

Multiple-output Gaussian process

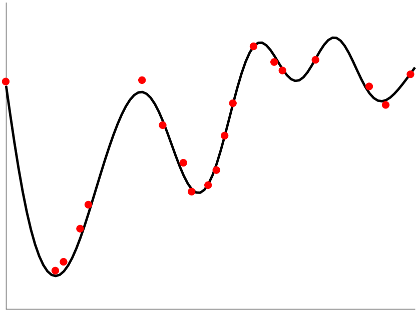


$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$ $f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

Multiple-output Gaussian process



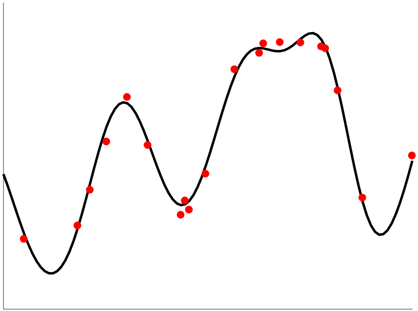
$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$

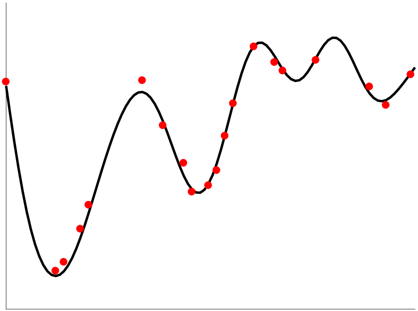


$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, y_1(\mathbf{x}_{i,2}))|i = 1, \ldots, N_1\}$ $\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, y_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$

Multiple-output Gaussian process



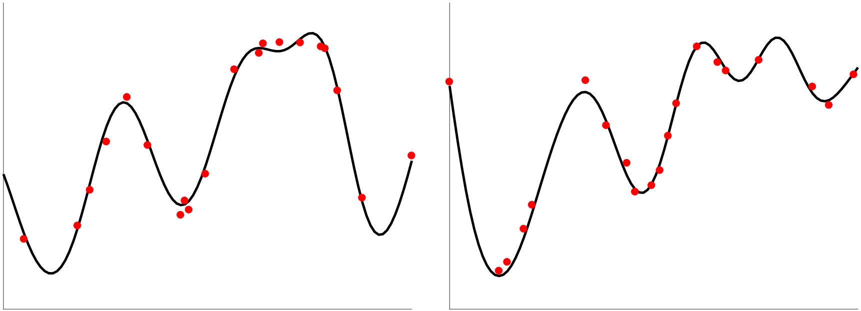


$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$ $\quad$ $f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, y_1(\mathbf{x}_{i,2}))|i = 1, \ldots, N_1\}$ $\quad$ $\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, y_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$

$\mathbf{y}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_1 + \sigma_1^2\mathbf{I})$ $\quad$ $\mathbf{y}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_2 + \sigma_2^2\mathbf{I})$

Multiple-output Gaussian process



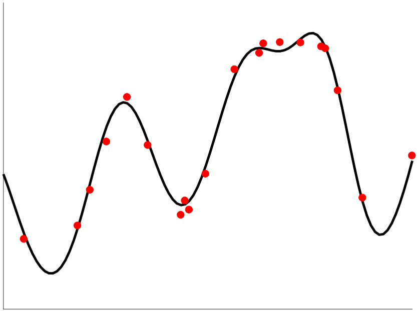
$$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}')) \qquad f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, y_1(\mathbf{x}_{i,2}))|i = 1, \ldots, N_1\} \quad \mathcal{D}_2 = \{(\mathbf{x}_{i,2}, y_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$$

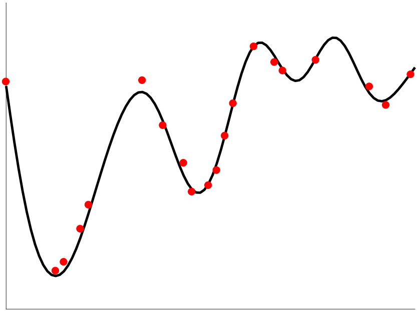
$$\mathbf{y}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_1 + \sigma_1^2\mathbf{I}) \qquad \mathbf{y}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_2 + \sigma_2^2\mathbf{I})$$

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{K}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{K}_2\end{bmatrix} + \begin{bmatrix}\sigma_1^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma_2^2\mathbf{I}\end{bmatrix}\right)$$

Multiple-output Gaussian process



$$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$$



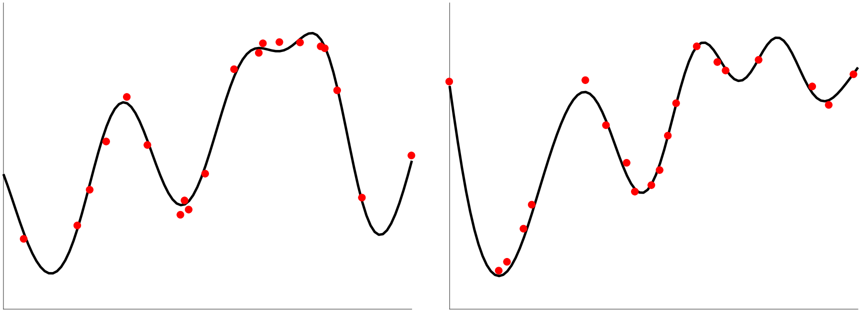
$$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, y_1(\mathbf{x}_{i,2}))|i = 1, \ldots, N_1\} \quad \mathcal{D}_2 = \{(\mathbf{x}_{i,2}, y_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$$

$$\mathbf{y}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_1 + \sigma_1^2\mathbf{I}) \qquad \mathbf{y}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_2 + \sigma_2^2\mathbf{I})$$

$$\underbrace{\begin{bmatrix}\mathbf{y}_1 \\ \mathbf{y}_2\end{bmatrix}}_{\mathbf{y}} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0} \\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{K}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_2\end{bmatrix} + \begin{bmatrix}\sigma_1^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\mathbf{I}\end{bmatrix}\right)$$

Multiple-output Gaussian process



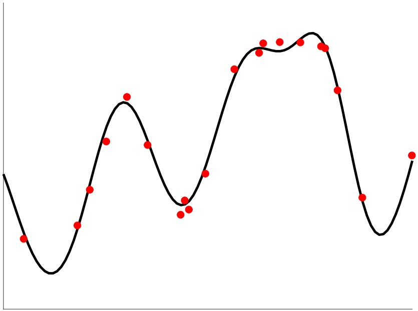
$$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}')) \qquad f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, y_1(\mathbf{x}_{i,2}))|i = 1, \dots, N_1\} \quad \mathcal{D}_2 = \{(\mathbf{x}_{i,2}, y_2(\mathbf{x}_{i,2}))|i = 1, \dots, N_2\}$$

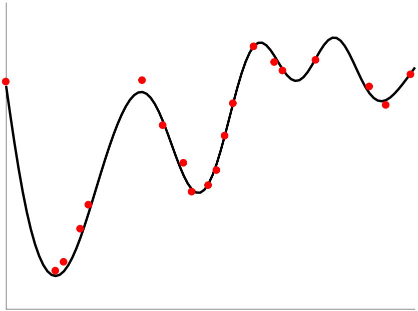
$$\mathbf{y}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_1 + \sigma_1^2\mathbf{I}) \qquad \mathbf{y}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_2 + \sigma_2^2\mathbf{I})$$

$$\underbrace{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}}_{\mathbf{y}} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \underbrace{\begin{bmatrix} \mathbf{K}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_2 \end{bmatrix}}_{\mathbf{K}_{\mathbf{f},\mathbf{f}}} + \begin{bmatrix} \sigma_1^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\mathbf{I} \end{bmatrix} \right)$$

Multiple-output Gaussian process



$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$



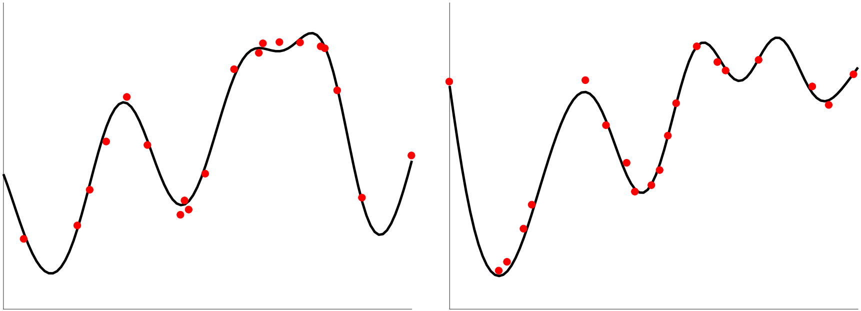
$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, y_1(\mathbf{x}_{i,2}))|i = 1, \ldots, N_1\}$ $\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, y_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$

$\mathbf{y}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_1 + \sigma_1^2\mathbf{I})$ $\mathbf{y}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_2 + \sigma_2^2\mathbf{I})$

$$\underbrace{\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}}_{\mathbf{y}} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \underbrace{\begin{bmatrix}\mathbf{K}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{K}_2\end{bmatrix}}_{\mathbf{K}_{\mathbf{f},\mathbf{f}}} + \underbrace{\begin{bmatrix}\sigma_1^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma_2^2\mathbf{I}\end{bmatrix}}_{\Sigma}\right)$$

Multiple-output Gaussian process



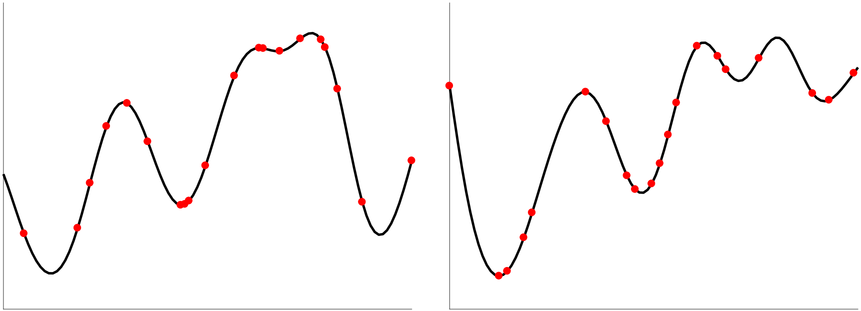
$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$ $\quad$ $f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, y_1(\mathbf{x}_{i,2}))|i = 1, \ldots, N_1\}$ $\quad$ $\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, y_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$

$\mathbf{y}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_1 + \sigma_1^2\mathbf{I})$ $\quad$ $\mathbf{y}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_2 + \sigma_2^2\mathbf{I})$

$$\underbrace{\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}}_{\mathbf{y}} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}}_{\mathbf{0}}, \underbrace{\begin{bmatrix}\mathbf{K}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{K}_2\end{bmatrix}}_{\mathbf{K}_{\mathbf{f},\mathbf{f}}} + \underbrace{\begin{bmatrix}\sigma_1^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma_2^2\mathbf{I}\end{bmatrix}}_{\Sigma}\right)$$

Kernels for multiple outputs

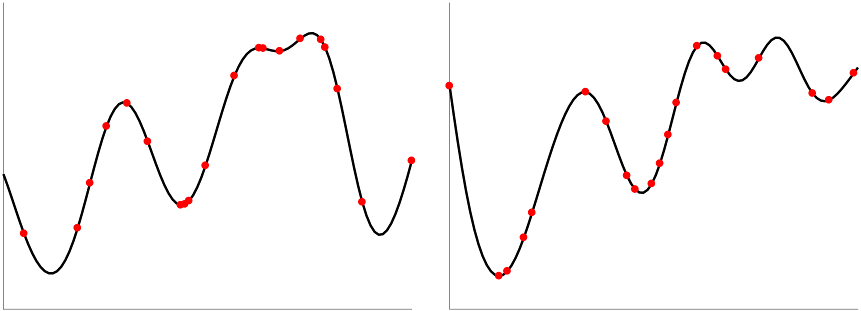


$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$

$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))|i = 1, \ldots, N_1\}$ $\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$

Kernels for multiple outputs

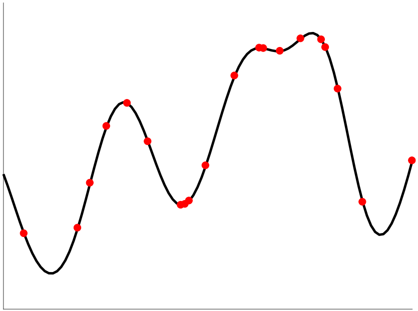


$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$ $\quad$ $f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

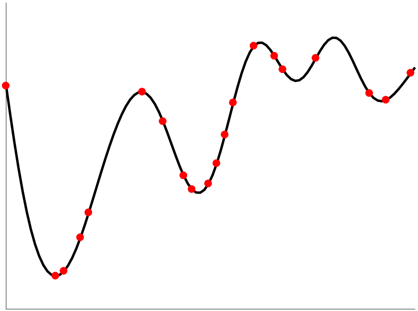
$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))|i = 1, \ldots, N_1\}$ $\quad$ $\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$

$$\underbrace{\begin{bmatrix}\mathbf{f}_1\\ \mathbf{f}_2\end{bmatrix}}_{\mathbf{f}} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}}_{\mathbf{0}}, \underbrace{\begin{bmatrix}\mathbf{K}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{K}_2\end{bmatrix}}_{\mathbf{K}_{\mathbf{f},\mathbf{f}}}\right)$$

Kernels for multiple outputs



$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$

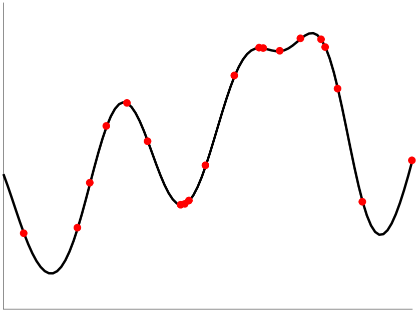


$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

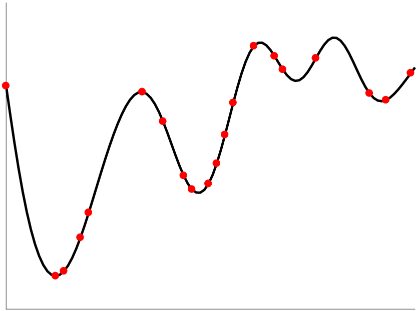
$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))|i = 1, \ldots, N_1\}$ $\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$

$$\mathbf{K}_{\mathbf{f},\mathbf{f}} = \begin{bmatrix} \mathbf{K}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_2 \end{bmatrix}$$

Kernels for multiple outputs



$$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$$

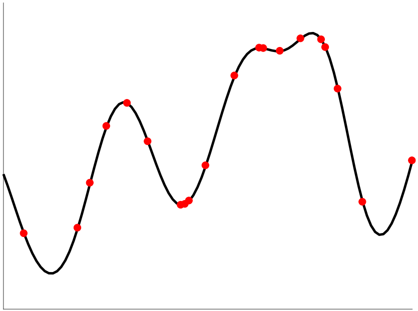


$$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$$

$$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))|i = 1, \ldots, N_1\} \quad \mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$$

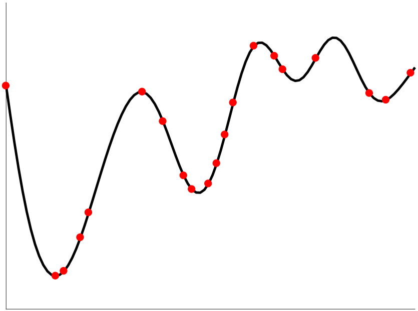
$$\mathbf{K}_{\mathbf{f},\mathbf{f}} = \begin{bmatrix} \mathbf{K}_1 & \mathbf{?} \\ \mathbf{?} & \mathbf{K}_2 \end{bmatrix}$$

Kernels for multiple outputs



$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$

$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))|i = 1, \ldots, N_1\}$



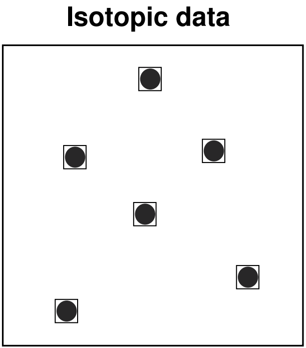
$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

$\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))|i = 1, \ldots, N_2\}$

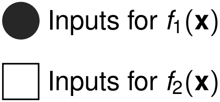
$$\mathbf{K}_{\mathbf{f},\mathbf{f}} = \begin{bmatrix} \mathbf{K}_1 & \textbf{?} \\ \textbf{?} & \mathbf{K}_2 \end{bmatrix}$$

Build a cross-covariance function $\mathrm{cov}[f_1(\mathbf{x}), f_2(\mathbf{x}')]$ such that $\mathbf{K}_{\mathbf{f},\mathbf{f}}$ is positive semi-definite.

Different input configurations of the data

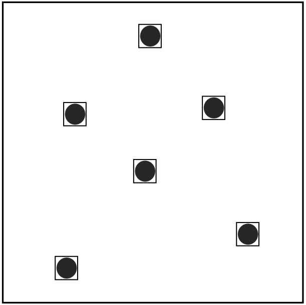


Sample sites are shared

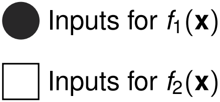


Different input configurations of the data

Isotopic data

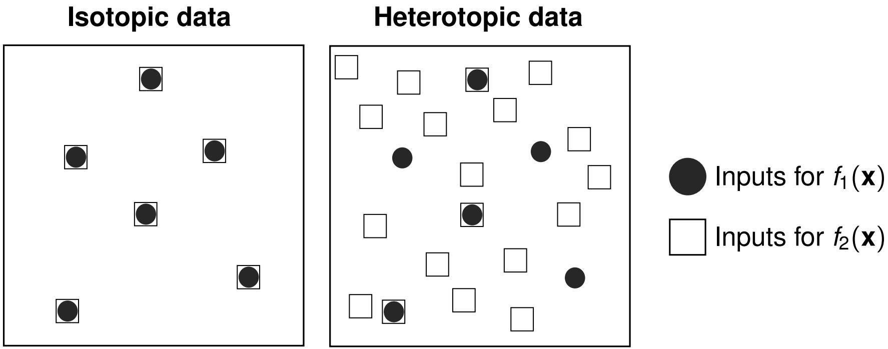


Sample sites are shared



$$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))_{i=1}^N\}$$
$$\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))_{i=1}^N\}$$

Different input configurations of the data

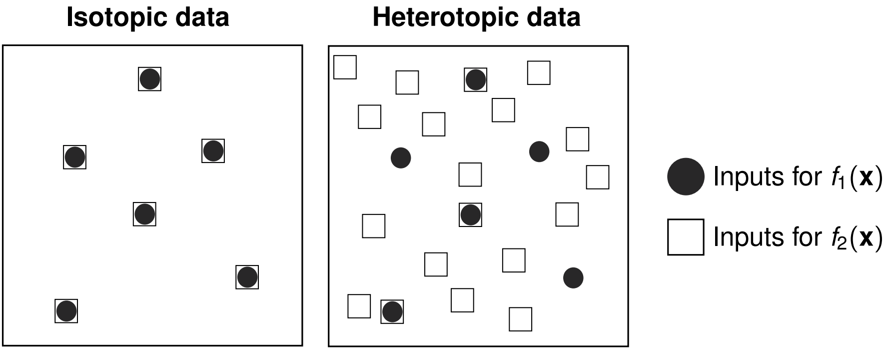


Sample sites are shared

Sample sites may be different

$$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))_{i=1}^N\}$$
$$\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))_{i=1}^N\}$$

Different input configurations of the data



Sample sites are shared

Sample sites may be different

$$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))_{i=1}^{N}\}$$
$$\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))_{i=1}^{N}\}$$

$$\mathcal{D}_1 = \{(\mathbf{x}_{i,1}, f_1(\mathbf{x}_{i,1}))_{i=1}^{N_1}\}$$
$$\mathcal{D}_2 = \{(\mathbf{x}_{i,2}, f_2(\mathbf{x}_{i,2}))_{i=1}^{N_2}\}$$

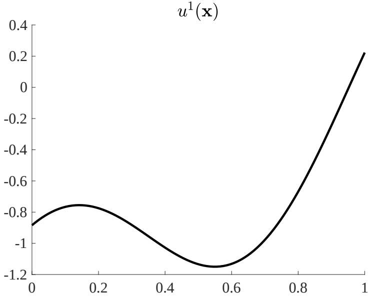
Contents

Intrinsic coregionalization model (ICM): two outputs

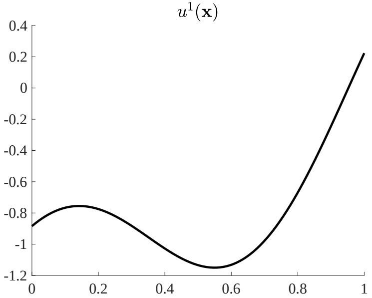
- Consider two outputs $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^p$.
- We assume the following generative model for the outputs
  1. Sample from a GP $u(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$ to obtain $u^1(\mathbf{x})$
  2. Obtain $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ by linearly transforming $u^1(\mathbf{x})$

$$f_1(\mathbf{x}) = a_1^1 u^1(\mathbf{x})$$
$$f_2(\mathbf{x}) = a_2^1 u^1(\mathbf{x})$$

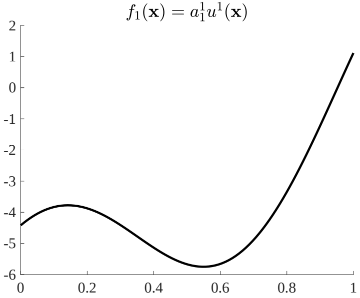
ICM: samples



ICM: samples

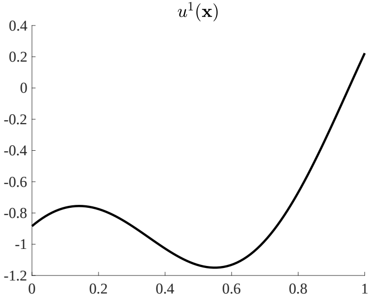


$u^1(\mathbf{x})$
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
0
0.2
0.4
0.6
0.8
1

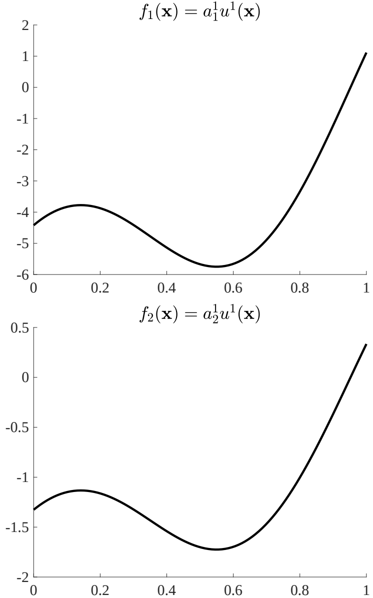


$f_1(\mathbf{x}) = a_1^1 u^1(\mathbf{x})$
2
1
0
-1
-2
-3
-4
-5
-6
0
0.2
0.4
0.6
0.8
1

ICM: samples

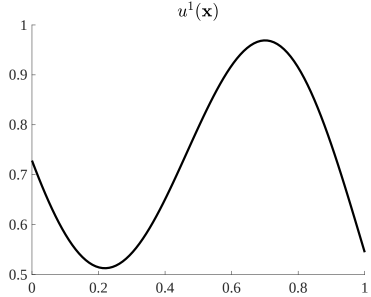


$u^1(\mathbf{x})$

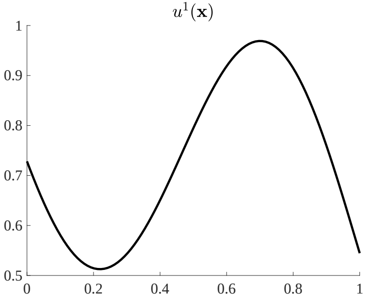


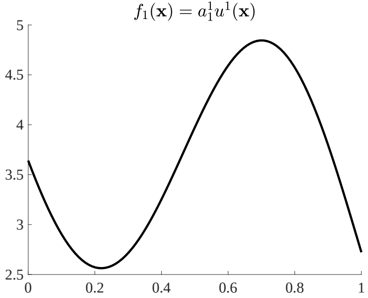
$f_1(\mathbf{x}) = a_1^1 u^1(\mathbf{x})$
$f_2(\mathbf{x}) = a_2^1 u^1(\mathbf{x})$

ICM: samples

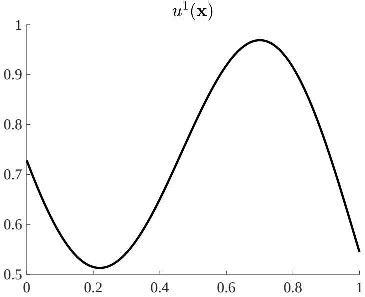


ICM: samples

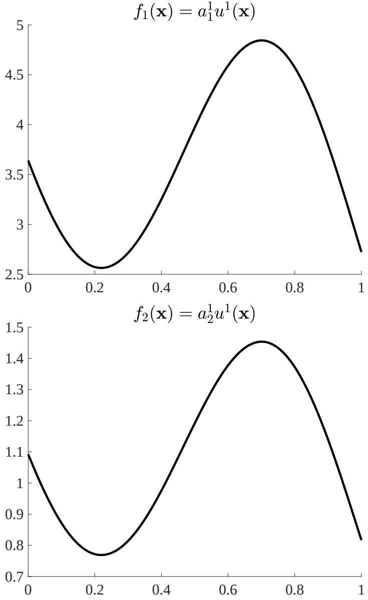

$u^1(\mathbf{x})$
1
0.9
0.8
0.7
0.6
0.5
0
0.2
0.4
0.6
0.8
1


$f_1(\mathbf{x}) = a_1^1 u^1(\mathbf{x})$
5
4.5
4
3.5
3
2.5
0
0.2
0.4
0.6
0.8
1

ICM: samples



$u^1(\mathbf{x})$



$f_1(\mathbf{x}) = a_1^1 u^1(\mathbf{x})$
$f_2(\mathbf{x}) = a_2^1 u^1(\mathbf{x})$

ICM: covariance (I)

- For a fixed value of $\mathbf{x}$, we can group $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ in a vector $\mathbf{f}(\mathbf{x})$

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix}$$

- We refer to this vector as a *vector-valued function*.

- The covariance for $\mathbf{f}(\mathbf{x})$ is computed as

$$\operatorname{cov}\left(\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')\right) = \mathbb{E}\left\{\mathbf{f}(\mathbf{x})[\mathbf{f}(\mathbf{x}')]^\top\right\} - \mathbb{E}\left\{\mathbf{f}(\mathbf{x})\right\}\left[\mathbb{E}\left\{\mathbf{f}(\mathbf{x}')\right\}\right]^\top.$$

- We compute first the term $\mathbb{E}\left\{\mathbf{f}(\mathbf{x})[\mathbf{f}(\mathbf{x}')]^\top\right\}$

$$\mathbb{E}\left\{\begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix}\begin{bmatrix} f_1(\mathbf{x}') & f_2(\mathbf{x}') \end{bmatrix}\right\} = \begin{bmatrix} \mathbb{E}\left\{f_1(\mathbf{x})f_1(\mathbf{x}')\right\} & \mathbb{E}\left\{f_1(\mathbf{x})f_2(\mathbf{x}')\right\} \\ \mathbb{E}\left\{f_2(\mathbf{x})f_1(\mathbf{x}')\right\} & \mathbb{E}\left\{f_2(\mathbf{x})f_2(\mathbf{x}')\right\} \end{bmatrix}$$

ICM: covariance (II)

- We compute the expected values as

$$\mathbb{E}\left\{f_1(\mathbf{x})f_1(\mathbf{x}')\right\} = \mathbb{E}\left\{a_1^1 u^1(\mathbf{x}) a_1^1 u^1(\mathbf{x}')\right\} = (a_1^1)^2 \mathbb{E}\left\{u^1(\mathbf{x})u^1(\mathbf{x}')\right\}$$

$$\mathbb{E}\left\{f_1(\mathbf{x})f_2(\mathbf{x}')\right\} = \mathbb{E}\left\{a_1^1 u^1(\mathbf{x}) a_2^1(\mathbf{x}')\right\} = a_1^1 a_2^1 \mathbb{E}\left\{u^1(\mathbf{x})u^1(\mathbf{x}')\right\}$$

$$\mathbb{E}\left\{f_2(\mathbf{x})f_2(\mathbf{x}')\right\} = \mathbb{E}\left\{a_2^1 u^1(\mathbf{x}) a_2^1 u^1(\mathbf{x}')\right\} = (a_2^1)^2 \mathbb{E}\left\{u^1(\mathbf{x})u^1(\mathbf{x}')\right\}$$

- The term $\mathbb{E}\left\{\mathbf{f}(\mathbf{x})[\mathbf{f}(\mathbf{x}')]^\top\right\}$ follows as

$$\begin{aligned}\mathbb{E}\left\{\mathbf{f}(\mathbf{x})[\mathbf{f}(\mathbf{x}')]^\top\right\} &= \begin{bmatrix}(a_1^1)^2\mathbb{E}\left\{u^1(\mathbf{x})u^1(\mathbf{x}')\right\} & a_1^1 a_2^1\mathbb{E}\left\{u^1(\mathbf{x})u^1(\mathbf{x}')\right\} \\ a^1 a^2\mathbb{E}\left\{u^1(\mathbf{x})u^1(\mathbf{x}')\right\} & (a_2^1)^2\mathbb{E}\left\{u^1(\mathbf{x})u^1(\mathbf{x}')\right\}\end{bmatrix} \\ &= \begin{bmatrix}(a_1^1)^2 & a_1^1 a_2^1 \\ a_1^1 a_2^1 & (a_2^1)^2\end{bmatrix}\mathbb{E}\left\{u^1(\mathbf{x})u^1(\mathbf{x}')\right\}\end{aligned}$$

- The term $\mathbb{E}\left\{\mathbf{f}(\mathbf{x})\right\}$ is computed as

$$\mathbb{E}\left\{\begin{bmatrix}f_1(\mathbf{x}) \\ f_2(\mathbf{x})\end{bmatrix}\right\} = \begin{bmatrix}\mathbb{E}\left\{f_1(\mathbf{x})\right\} \\ \mathbb{E}\left\{f_2(\mathbf{x})\right\}\end{bmatrix} = \begin{bmatrix}\mathbb{E}\left\{a_1^1 u^1(\mathbf{x})\right\} \\ \mathbb{E}\left\{a_2^1 u^1(\mathbf{x})\right\}\end{bmatrix} = \begin{bmatrix}a_1^1 \\ a_2^1\end{bmatrix}\mathbb{E}\left\{u^1(\mathbf{x})\right\}$$

ICM: covariance (III)

- Putting the terms together, the covariance for $\mathbf{f}(\mathbf{x}')$ follows as

$$\begin{bmatrix}(a_1^1)^2 & a_1^1 a_2^1 \\ a_1^1 a_2^1 & (a_2^1)^2\end{bmatrix} \mathbb{E}\left\{u^1(\mathbf{x})u^1(\mathbf{x}')\right\} - \begin{bmatrix}a_1^1 \\ a_2^1\end{bmatrix} \begin{bmatrix}a_1^1 & a_2^1\end{bmatrix} \mathbb{E}\left\{u^1(\mathbf{x})\right\} \mathbb{E}\left\{u^1(\mathbf{x}')\right\}$$

- Defining $\mathbf{a} = [a_1^1 \;\; a_2^1]^\top$,

$$\begin{aligned}\operatorname{cov}\left(\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')\right) &= \mathbf{a}\mathbf{a}^\top \mathbb{E}\left\{u^1(\mathbf{x})u^1(\mathbf{x}')\right\} - \mathbf{a}\mathbf{a}^\top \mathbb{E}\left\{u^1(\mathbf{x})\right\} \mathbb{E}\left\{u^1(\mathbf{x}')\right\} \\ &= \mathbf{a}\mathbf{a}^\top \underbrace{\left[\mathbb{E}\left\{u^1(\mathbf{x})u^1(\mathbf{x}')\right\} - \mathbb{E}\left\{u^1(\mathbf{x})\right\} \mathbb{E}\left\{u^1(\mathbf{x}')\right\}\right]}_{k(\mathbf{x},\mathbf{x}')} \\ &= \mathbf{a}\mathbf{a}^\top k(\mathbf{x}, \mathbf{x}')\end{aligned}$$

- We define $\mathbf{B} = \mathbf{a}\mathbf{a}^\top$, leading to

$$\operatorname{cov}\left(\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')\right) = \mathbf{B}k(\mathbf{x}, \mathbf{x}') = \begin{bmatrix}b_{11} & b_{12} \\ b_{21} & b_{22}\end{bmatrix} k(\mathbf{x}, \mathbf{x}')$$

- Notice that $\mathbf{B}$ has rank one.

ICM: two outputs and two latent samples

- We can introduce a bit more of complexity in the model before as follows.

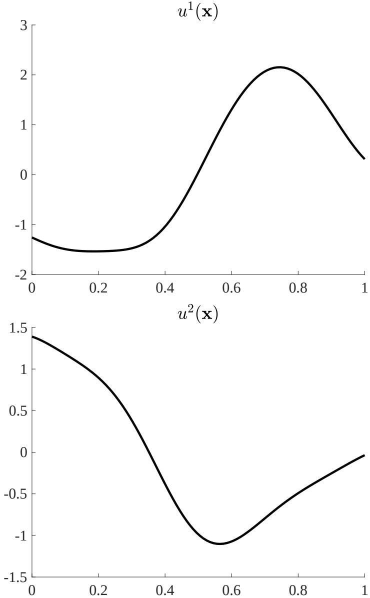
- Consider again two outputs $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^p$.

- We assume the following generative model for the outputs
  1. Sample **twice** from a GP $u(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$ to obtain $u^1(\mathbf{x})$ and $u^2(\mathbf{x})$
  2. Obtain $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ by adding a scaled transformation of $u^1(\mathbf{x})$ and $u^2(\mathbf{x})$

$$f_1(\mathbf{x}) = a_1^1 u^1(\mathbf{x}) + a_1^2 u^2(\mathbf{x})$$
$$f_2(\mathbf{x}) = a_2^1 u^1(\mathbf{x}) + a_2^2 u^2(\mathbf{x})$$

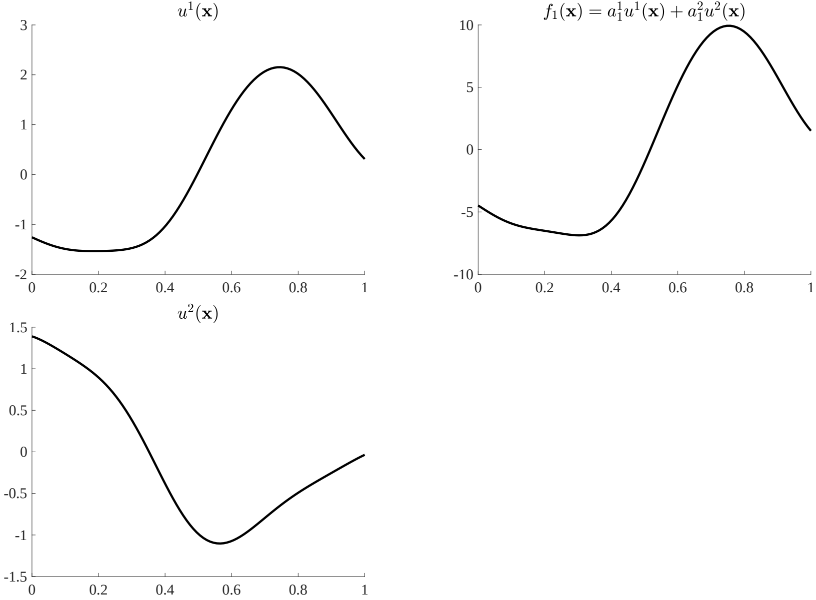
- Notice that $u^1(\mathbf{x})$ and $u^2(\mathbf{x})$ are independent, although they share the same covariance $k(\mathbf{x}, \mathbf{x}')$.

ICM: samples



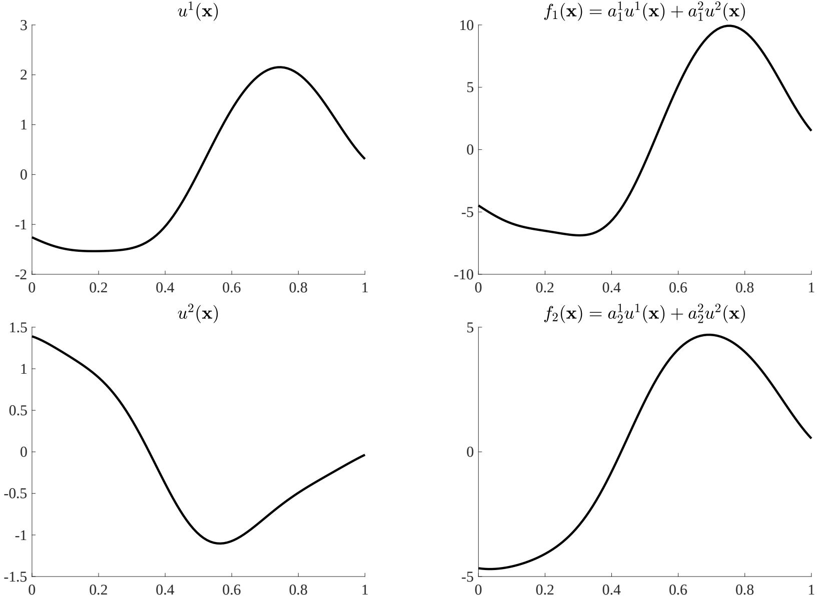
$u^1(\mathbf{x})$
3
2
1
0
-1
-2
0
0.2
0.4
0.6
0.8
1
$u^2(\mathbf{x})$
1.5
1
0.5
0
-0.5
-1
-1.5
0
0.2
0.4
0.6
0.8
1

ICM: samples



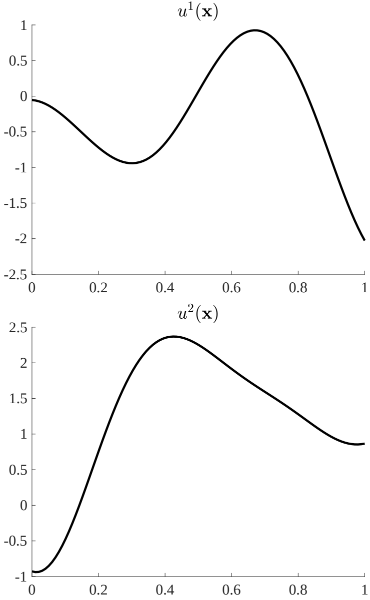
$u^1(\mathbf{x})$
$f_1(\mathbf{x}) = a_1^1 u^1(\mathbf{x}) + a_1^2 u^2(\mathbf{x})$
$u^2(\mathbf{x})$

ICM: samples



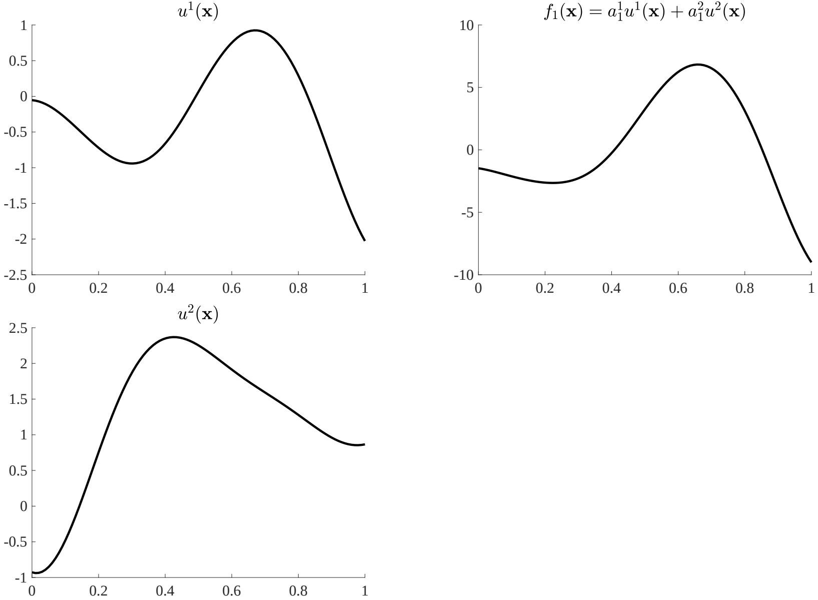
$u^1(\mathbf{x})$
$f_1(\mathbf{x}) = a_1^1 u^1(\mathbf{x}) + a_1^2 u^2(\mathbf{x})$
$u^2(\mathbf{x})$
$f_2(\mathbf{x}) = a_2^1 u^1(\mathbf{x}) + a_2^2 u^2(\mathbf{x})$
3
2
1
0
-1
-2
10
5
-5
-10
1.5
0.5
-0.5
-1.5
0.2
0.4
0.6
0.8

ICM: samples



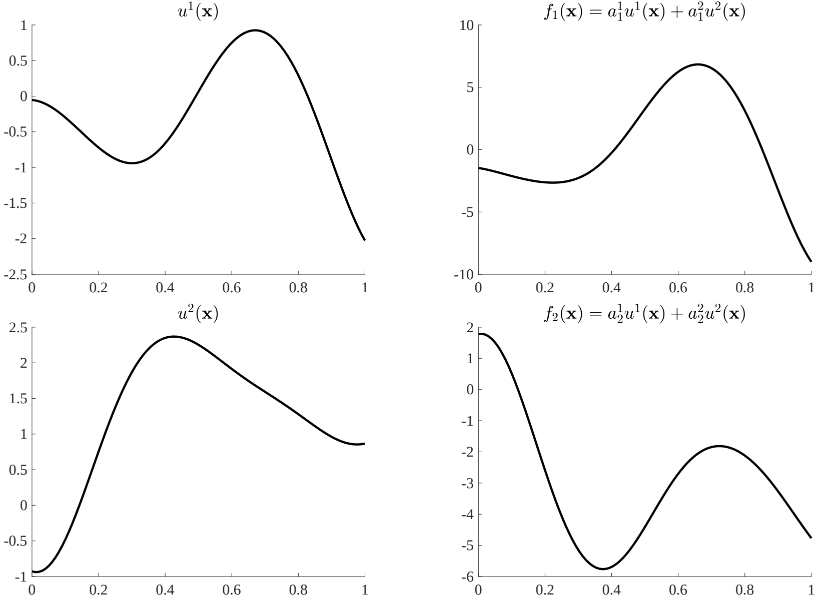
$u^1(\mathbf{x})$
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
0.2
0.4
0.6
0.8
1
$u^2(\mathbf{x})$
2.5
2
1.5
1
0.5
0
-0.5
-1
0
0.2
0.4
0.6
0.8
1

ICM: samples



$u^1(\mathbf{x})$
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
0.2
0.4
0.6
0.8
1
$f_1(\mathbf{x}) = a_1^1 u^1(\mathbf{x}) + a_1^2 u^2(\mathbf{x})$
10
5
0
-5
-10
0
0.2
0.4
0.6
0.8
1
$u^2(\mathbf{x})$
2.5
2
1.5
1
0.5
0
-0.5
-1
0
0.2
0.4
0.6
0.8
1

ICM: samples



$u^1(\mathbf{x})$
$f_1(\mathbf{x}) = a_1^1 u^1(\mathbf{x}) + a_1^2 u^2(\mathbf{x})$
$u^2(\mathbf{x})$
$f_2(\mathbf{x}) = a_2^1 u^1(\mathbf{x}) + a_2^2 u^2(\mathbf{x})$
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
10
5
-5
-10
2.5
2
1.5
-6
-4
-3
0.2
0.4
0.6
0.8

ICM: covariance

- The vector-valued function can be written as $\mathbf{f}(\mathbf{x})$

$$\mathbf{f}(\mathbf{x}) = \mathbf{a}^1 u^1(\mathbf{x}) + \mathbf{a}^2 u^2(\mathbf{x})$$

where $\mathbf{a}^1 = [a_1^1 \;\; a_2^1]^\top$ and $\mathbf{a}^2 = [a_1^2 \;\; a_2^2]^\top$.

- The covariance for $\mathbf{f}(\mathbf{x})$ is computed as

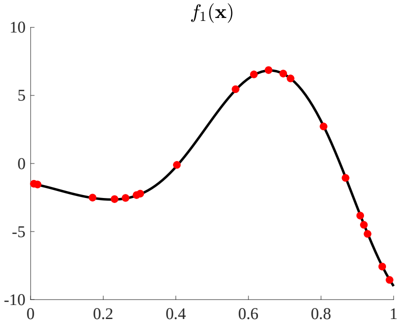
$$\begin{aligned}
\operatorname{cov}\left(\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')\right) &= \mathbf{a}^1(\mathbf{a}^1)^\top \operatorname{cov}(u^1(\mathbf{x}), u^1(\mathbf{x}')) + \mathbf{a}^2(\mathbf{a}^2)^\top \operatorname{cov}(u^2(\mathbf{x}), u^2(\mathbf{x}')) \\
&= \mathbf{a}^1(\mathbf{a}^1)^\top k(\mathbf{x}, \mathbf{x}') + \mathbf{a}^2(\mathbf{a}^2)^\top k(\mathbf{x}, \mathbf{x}') \\
&= \left[\mathbf{a}^1(\mathbf{a}^1)^\top + \mathbf{a}^2(\mathbf{a}^2)^\top\right] k(\mathbf{x}, \mathbf{x}')
\end{aligned}$$

- We define $\mathbf{B} = \mathbf{a}^1(\mathbf{a}^1)^\top + \mathbf{a}^2(\mathbf{a}^2)^\top$, leading to

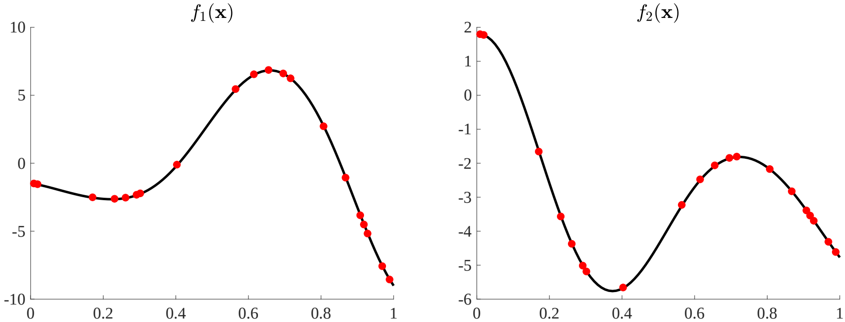
$$\operatorname{cov}\left(\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')\right) = \mathbf{B}k(\mathbf{x}, \mathbf{x}') = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} k(\mathbf{x}, \mathbf{x}')$$

- Notice that $\mathbf{B}$ has rank two.

ICM: observed data

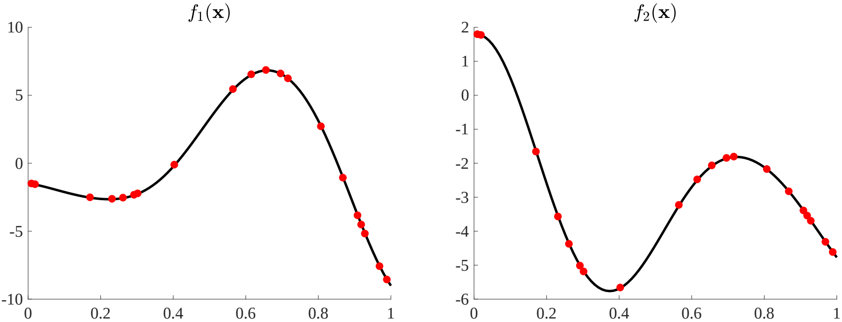


ICM: observed data



$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))|i = 1, \ldots, N\}$ $\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))|i = 1, \ldots, N\}$

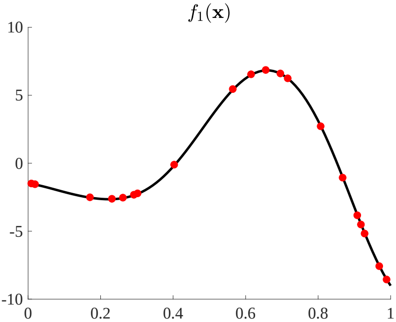
ICM: observed data

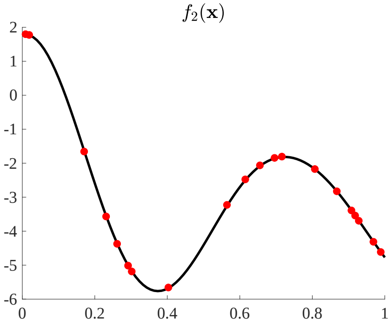


$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i)) | i = 1, \ldots, N\}$ $\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i)) | i = 1, \ldots, N\}$

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1) \\ \vdots \\ f_1(\mathbf{x}_N) \\ f_2(\mathbf{x}_1) \\ \vdots \\ f_2(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} b_{11}\mathbf{K} & b_{12}\mathbf{K} \\ b_{21}\mathbf{K} & b_{22}\mathbf{K} \end{bmatrix} \right)$$

ICM: observed data



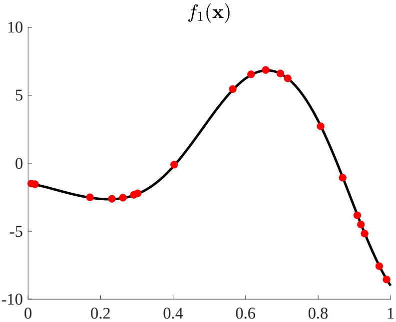


$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))|i = 1, \ldots, N\}$ $\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))|i = 1, \ldots, N\}$

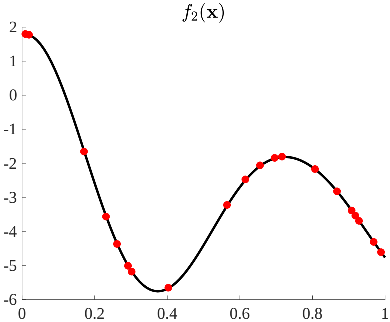
$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1) \\ \vdots \\ f_1(\mathbf{x}_N) \\ f_2(\mathbf{x}_1) \\ \vdots \\ f_2(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} b_{11}\mathbf{K} & b_{12}\mathbf{K} \\ b_{21}\mathbf{K} & b_{22}\mathbf{K} \end{bmatrix} \right)$$

The matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$ has elements $k(\mathbf{x}_i, \mathbf{x}_j)$.

ICM: observed data



$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))|i = 1, \ldots, N\}$

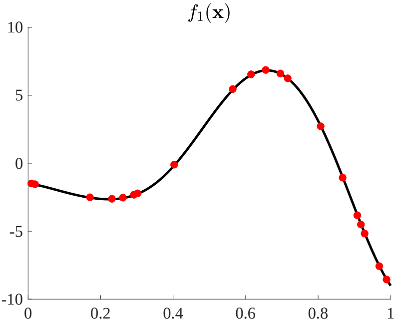


$\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))|i = 1, \ldots, N\}$

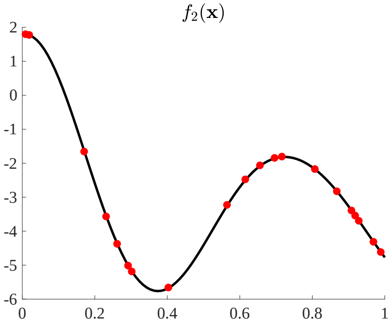
The Kronecker product between matrices $\mathbf{C} \in \mathbb{R}^{c_1 \times c_2}$ and $\mathbf{G} \in \mathbb{R}^{g_1 \times g_2}$ with

$$\mathbf{C} = \begin{bmatrix} c_{1,1} & \cdots & c_{1,c_2} \\ \vdots & \vdots & \vdots \\ c_{c_1,1} & \cdots & c_{c_1,c_2} \end{bmatrix} \quad \text{is} \quad \mathbf{C} \otimes \mathbf{G} = \begin{bmatrix} c_{1,1}\mathbf{G} & \cdots & c_{1,c_2}\mathbf{G} \\ \vdots & \vdots & \vdots \\ c_{c_1,1}\mathbf{G} & \cdots & c_{c_1,c_2}\mathbf{G} \end{bmatrix}$$

ICM: observed data



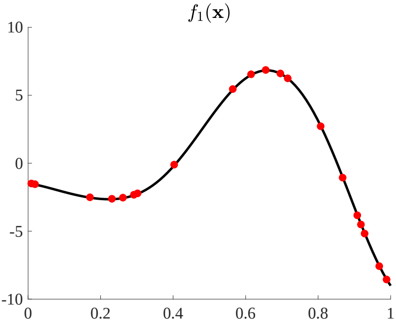
$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))|i = 1, \ldots, N\}$



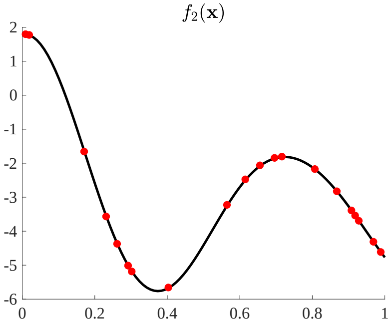
$\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))|i = 1, \ldots, N\}$

$$\begin{bmatrix}\mathbf{f}_1\\ \mathbf{f}_2\end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1)\\ \vdots \\ f_1(\mathbf{x}_N)\\ f_2(\mathbf{x}_1)\\ \vdots \\ f_2(\mathbf{x}_N)\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \mathbf{B}\otimes\mathbf{K}\right)$$

ICM: observed data



$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i)) | i = 1, \ldots, N\}$



$\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i)) | i = 1, \ldots, N\}$

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1) \\ \vdots \\ f_1(\mathbf{x}_N) \\ f_2(\mathbf{x}_1) \\ \vdots \\ f_2(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \mathbf{B} \otimes \mathbf{K} \right)$$

The matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$ has elements $k(\mathbf{x}_i, \mathbf{x}_j)$.

ICM: general case

- Consider a set of functions $\{f_d(\mathbf{x})\}_{d=1}^{D}$.

- In the ICM

$$f_d(\mathbf{x}) = \sum_{i=1}^{R} a_d^i u^i(\mathbf{x}),$$

  where the functions $u^i(\mathbf{x})$ are GPs sampled independently, and share the same covariance function $k(\mathbf{x}, \mathbf{x}')$.

- For $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}) \cdots f_D(\mathbf{x})]^\top$, the covariance $\text{cov}[\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')]$ is given as

$$\text{cov}[\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')] = \mathbf{A}\mathbf{A}^\top k(\mathbf{x}, \mathbf{x}') = \mathbf{B}\, k(\mathbf{x}, \mathbf{x}'),$$

  where $\mathbf{A} = [\mathbf{a}^1\ \mathbf{a}^2 \cdots \mathbf{a}^R]$.

- The rank of $\mathbf{B} \in \mathbb{R}^{D\times D}$ is given by R.

ICM: autokrigeability

- If the outputs are considered to be noise-free, prediction using the ICM under an isotopic data case is equivalent to independent prediction over each output.
- This circumstance is also known as autokrigeability.

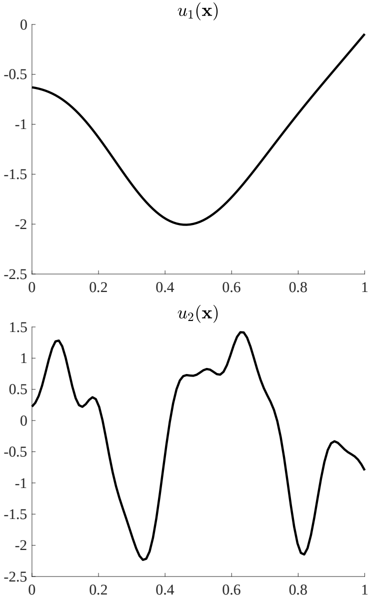
Contents

Semiparametric Latent Factor Model (SLFM)

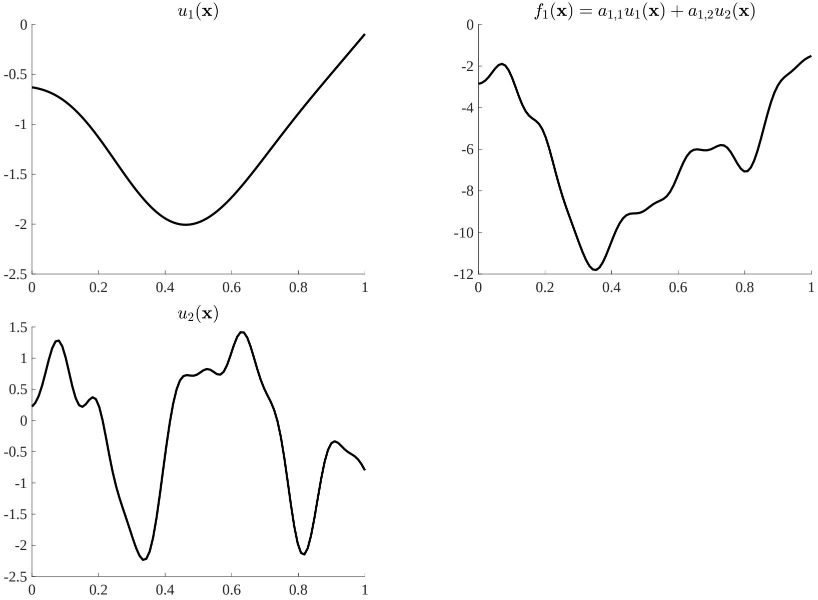
- ICM uses R samples $u^i(\mathbf{x})$ from $u(\mathbf{x})$ with the same covariance function.
- SLFM uses Q samples from $u_q(\mathbf{x})$ processes with different covariance functions.
- The SLFM with $Q = 1$ is the same to the ICM with $R = 1$.
- Consider two outputs $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^p$.
- Suppose we have $Q = 2$.
- We assume the following generative model for the outputs
  1. Sample from a GP $\mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$ to obtain $u_1(\mathbf{x})$.
  2. Sample from a GP $\mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$ to obtain $u_2(\mathbf{x})$.
  3. Obtain $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ by adding a scaled versions of $u_1(\mathbf{x})$ and $u_2(\mathbf{x})$

$$f_1(\mathbf{x}) = a_{1,1}u_1(\mathbf{x}) + a_{1,2}u_2(\mathbf{x})$$
$$f_2(\mathbf{x}) = a_{2,1}u_1(\mathbf{x}) + a_{2,2}u_2(\mathbf{x})$$

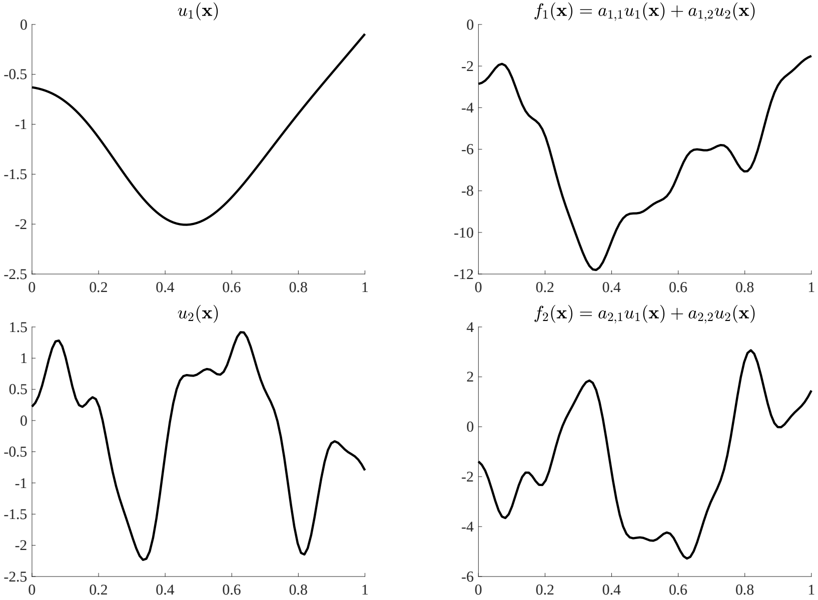
SLFM: samples



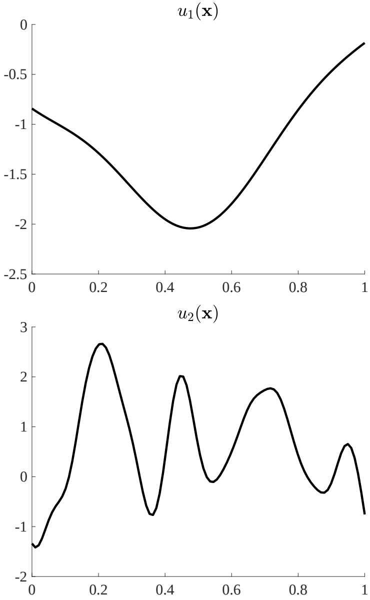
SLFM: samples



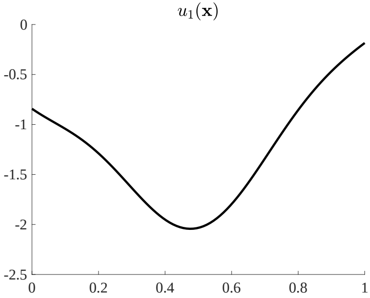
SLFM: samples

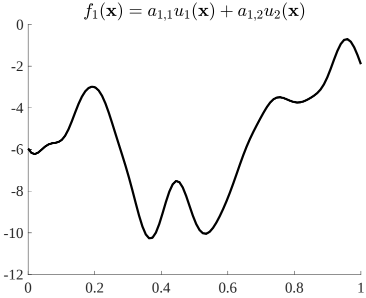


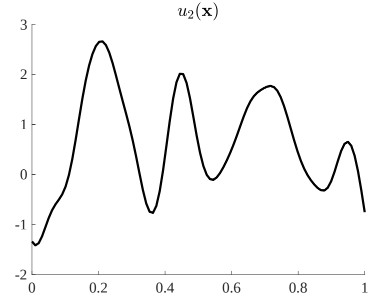
SLFM: samples



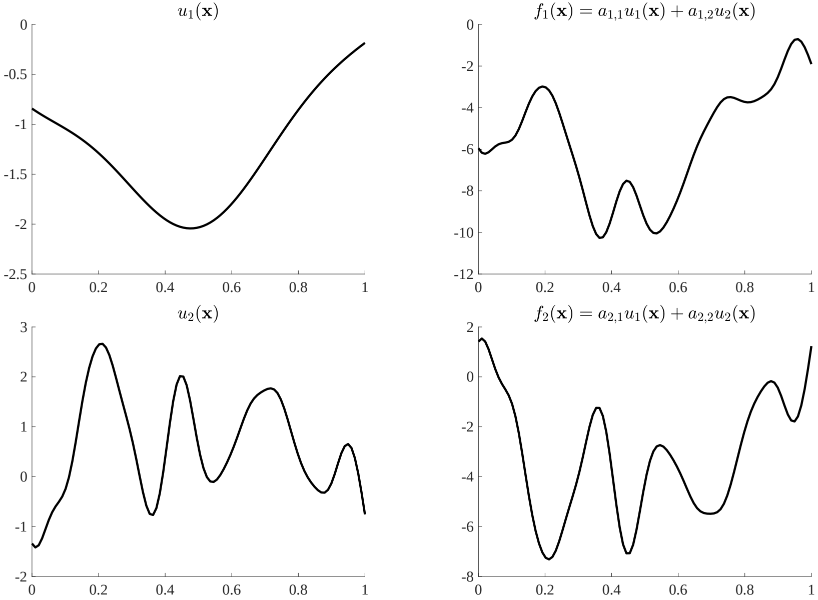
SLFM: samples







SLFM: samples



SLFM: covariance

- The vector-valued function can be written as $\mathbf{f}(\mathbf{x})$

$$\mathbf{f}(\mathbf{x}) = \mathbf{a}_1 u_1(\mathbf{x}) + \mathbf{a}_2 u_2(\mathbf{x})$$

  where $\mathbf{a}_1 = [a_{1,1} \;\; a_{2,1}]^\top$ and $\mathbf{a}_2 = [a_{1,2} \;\; a_{2,2}]^\top$.

- The covariance for $\mathbf{f}(\mathbf{x})$ is computed as

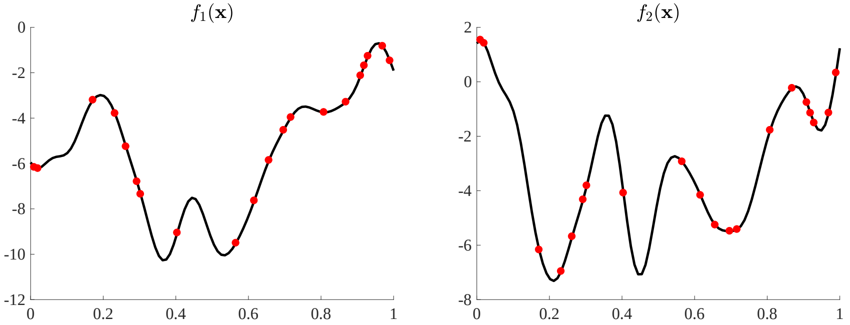
$$\begin{aligned}\operatorname{cov}\left(\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')\right) &= \mathbf{a}_1(\mathbf{a}_1)^\top \operatorname{cov}(u_1(\mathbf{x}), u_1(\mathbf{x}')) + \mathbf{a}_2(\mathbf{a}_2)^\top \operatorname{cov}(u_2(\mathbf{x}), u_2(\mathbf{x}')) \\ &= \mathbf{a}_1(\mathbf{a}_1)^\top k_1(\mathbf{x}, \mathbf{x}') + \mathbf{a}_2(\mathbf{a}_2)^\top k_2(\mathbf{x}, \mathbf{x}')\end{aligned}$$

- We define $\mathbf{B}_1 = \mathbf{a}_1(\mathbf{a}_1)^\top$ and $\mathbf{B}_2 = \mathbf{a}_2(\mathbf{a}_2)^\top$, leading to

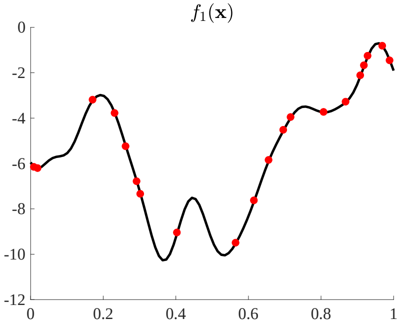
$$\operatorname{cov}\left(\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')\right) = \mathbf{B}_1 k_1(\mathbf{x}, \mathbf{x}') + \mathbf{B}_2 k_2(\mathbf{x}, \mathbf{x}')$$

- Notice that $\mathbf{B}_1$ and $\mathbf{B}_2$ have rank one.

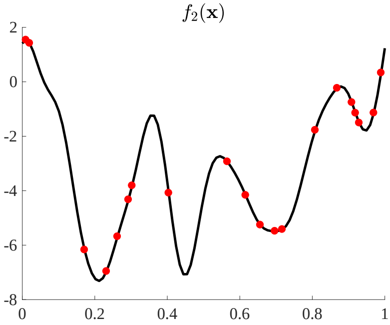
SLFM: observed data



SLFM: observed data

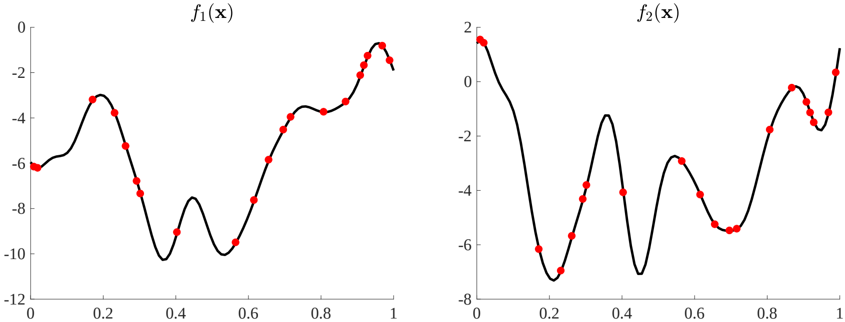


$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))|i = 1, \ldots, N\}$



$\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))|i = 1, \ldots, N\}$

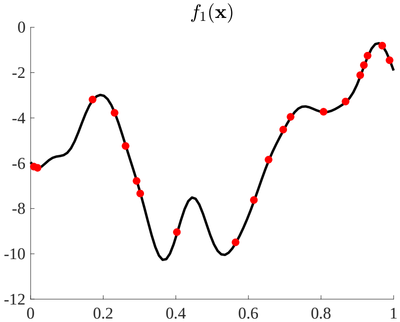
SLFM: observed data



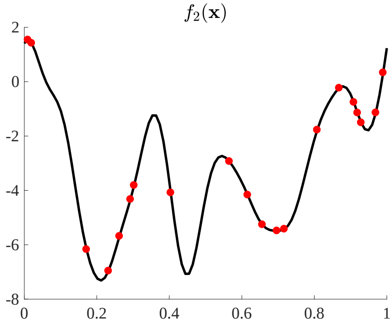
$$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))|i = 1, \dots, N\} \qquad \mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))|i = 1, \dots, N\}$$

$$\begin{bmatrix}\mathbf{f}_1\\ \mathbf{f}_2\end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1)\\ \vdots \\ f_1(\mathbf{x}_N)\\ f_2(\mathbf{x}_1)\\ \vdots \\ f_2(\mathbf{x}_N)\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \mathbf{B}_1 \otimes \mathbf{K}_1 + \mathbf{B}_2 \otimes \mathbf{K}_2\right)$$

SLFM: observed data



$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))|i = 1, \ldots, N\}$

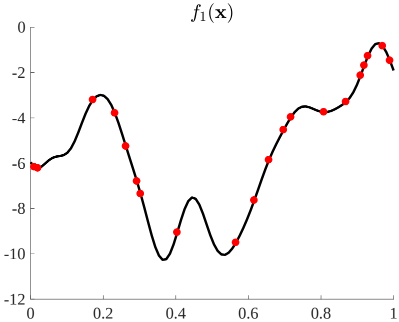


$\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))|i = 1, \ldots, N\}$

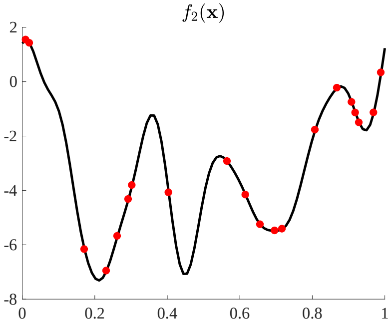
$$\begin{bmatrix}\mathbf{f}_1\\ \mathbf{f}_2\end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1)\\ \vdots \\ f_1(\mathbf{x}_N)\\ f_2(\mathbf{x}_1)\\ \vdots \\ f_2(\mathbf{x}_N)\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \mathbf{B}_1 \otimes \mathbf{K}_1 + \mathbf{B}_2 \otimes \mathbf{K}_2\right)$$

The matrix $\mathbf{K}_1 \in \mathbb{R}^{N\times N}$ has elements $k_1(\mathbf{x}_i, \mathbf{x}_j)$.

SLFM: observed data



$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))|i = 1, \dots, N\}$



$\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))|i = 1, \dots, N\}$

$$\begin{bmatrix}\mathbf{f}_1\\ \mathbf{f}_2\end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1)\\ \vdots \\ f_1(\mathbf{x}_N)\\ f_2(\mathbf{x}_1)\\ \vdots \\ f_2(\mathbf{x}_N)\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \mathbf{B}_1 \otimes \mathbf{K}_1 + \mathbf{B}_2 \otimes \mathbf{K}_2\right)$$

The matrix $\mathbf{K}_1 \in \mathbb{R}^{N\times N}$ has elements $k_1(\mathbf{x}_i, \mathbf{x}_j)$.

The matrix $\mathbf{K}_2 \in \mathbb{R}^{N\times N}$ has elements $k_2(\mathbf{x}_i, \mathbf{x}_j)$.

SLFM: general case

- Consider a set of functions $\{f_d(\mathbf{x})\}_{d=1}^{D}$.
- In the SLFM

$$f_d(\mathbf{x}) = \sum_{q=1}^{Q} a_{d,q} u_q(\mathbf{x}),$$

where the functions $u_q(\mathbf{x})$ are GPs with covariance functions $k_q(\mathbf{x}, \mathbf{x}')$.
- For $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}) \cdots f_D(\mathbf{x})]^\top$, the covariance $\mathrm{cov}[\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')]$ is given as

$$\mathrm{cov}[\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')] = \sum_{q=1}^{Q} \mathbf{A}_q \mathbf{A}_q^\top k_q(\mathbf{x}, \mathbf{x}') = \sum_{q=1}^{Q} \mathbf{B}_q k_q(\mathbf{x}, \mathbf{x}'),$$

where $\mathbf{A}_q = \mathbf{a}_q$.
- The rank of each $\mathbf{B}_q \in \mathbb{R}^{D \times D}$ is one.

Contents

Linear model of coregionalization (LMC)

- The LMC generalizes the ICM and the SLFM allowing several independent samples from GPs with different covariances.
- Consider a set of functions $\{f_d(\mathbf{x})\}_{d=1}^{D}$.
- In the LMC

$$f_d(\mathbf{x}) = \sum_{q=1}^{Q} \sum_{i=1}^{R_q} a_{d,q}^{i} u_q^{i}(\mathbf{x}),$$

where the functions $u_q^i(\mathbf{x})$ are GPs with zero means and covariance functions

$$\mathrm{cov}[u_q^i(\mathbf{x}), u_{q'}^{i'}(\mathbf{x}')] = k_q(\mathbf{x}, \mathbf{x}'),$$

if $i = i'$ and $q = q'$.

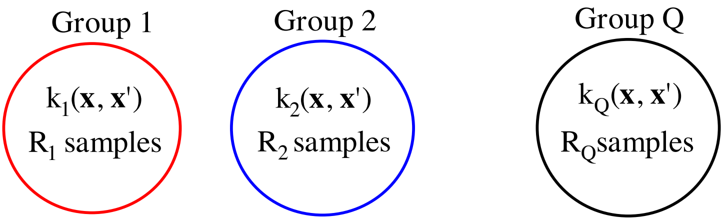
LMC: interpretation

- In the LMC

$$f_d(\mathbf{x}) = \sum_{q=1}^{Q} \sum_{i=1}^{R_q} a^i_{d,q} u^i_q(\mathbf{x}).$$

- There are Q groups of samples.

- For each group, there R_q samples obtained independently from the same GP with covariance $k_q(\mathbf{x}, \mathbf{x}')$.



LMC: example

- The LMC corresponds to the sum of Q ICMs.
- Suppose we have $D = 2$, $Q = 2$ and $R_q = 2$. According to the LMC

$$f_1(\mathbf{x}) = a^1_{1,1}u^1_1(\mathbf{x}) + a^2_{1,1}u^2_1(\mathbf{x}) + a^1_{1,2}u^1_2(\mathbf{x}) + a^2_{1,2}u^2_2(\mathbf{x}),$$
$$f_2(\mathbf{x}) = a^1_{2,1}u^1_1(\mathbf{x}) + a^2_{2,1}u^2_1(\mathbf{x}) + a^1_{2,2}u^1_2(\mathbf{x}) + a^2_{2,2}u^2_2(\mathbf{x}),$$

LMC: example

- The LMC corresponds to the sum of Q ICMs.

- Suppose we have $D = 2$, $Q = 2$ and $R_q = 2$. According to the LMC

$$
\begin{aligned}
f_1(\mathbf{x}) &= a^1_{1,1} u^1_1(\mathbf{x}) + a^2_{1,1} u^2_1(\mathbf{x}) + a^1_{1,2} u^1_2(\mathbf{x}) + a^2_{1,2} u^2_2(\mathbf{x}), \\
f_2(\mathbf{x}) &= a^1_{2,1} u^1_1(\mathbf{x}) + a^2_{2,1} u^2_1(\mathbf{x}) + a^1_{2,2} u^1_2(\mathbf{x}) + a^2_{2,2} u^2_2(\mathbf{x}),
\end{aligned}
$$

LMC: example

- The LMC corresponds to the sum of Q ICMs.

- Suppose we have $D = 2$, $Q = 2$ and $R_q = 2$. According to the LMC

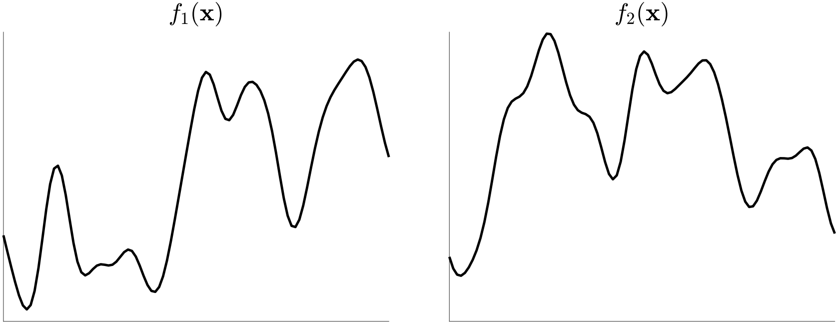
$$
\begin{aligned}
f_1(\mathbf{x}) &= a_{1,1}^1 u_1^1(\mathbf{x}) + a_{1,1}^2 u_1^2(\mathbf{x}) + a_{1,2}^1 u_2^1(\mathbf{x}) + a_{1,2}^2 u_2^2(\mathbf{x}),\\
f_2(\mathbf{x}) &= a_{2,1}^1 u_1^1(\mathbf{x}) + a_{2,1}^2 u_1^2(\mathbf{x}) + a_{2,2}^1 u_2^1(\mathbf{x}) + a_{2,2}^2 u_2^2(\mathbf{x}),
\end{aligned}
$$

LMC: example

- The LMC corresponds to the sum of Q ICMs.

- Suppose we have $D = 2$, $Q = 2$ and $R_q = 2$. According to the LMC

$$f_1(\mathbf{x}) = a_{1,1}^1 u_1^1(\mathbf{x}) + a_{1,1}^2 u_1^2(\mathbf{x}) + a_{1,2}^1 u_2^1(\mathbf{x}) + a_{1,2}^2 u_2^2(\mathbf{x}),$$
$$f_2(\mathbf{x}) = a_{2,1}^1 u_1^1(\mathbf{x}) + a_{2,1}^2 u_1^2(\mathbf{x}) + a_{2,2}^1 u_2^1(\mathbf{x}) + a_{2,2}^2 u_2^2(\mathbf{x}),$$

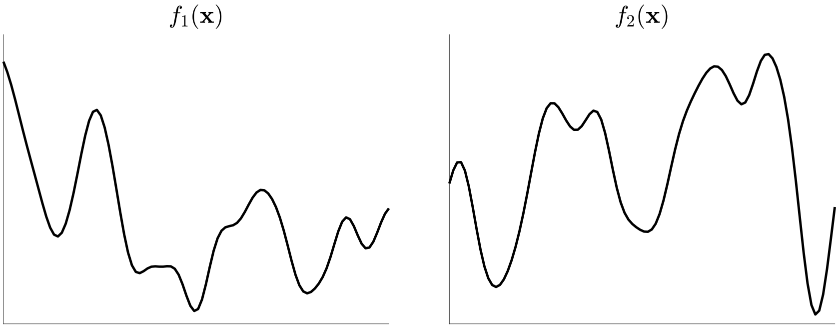


LMC: example

- The LMC corresponds to the sum of Q ICMs.

- Suppose we have $D = 2$, $Q = 2$ and $R_q = 2$. According to the LMC

$$f_1(\mathbf{x}) = a^1_{1,1}u^1_1(\mathbf{x}) + a^2_{1,1}u^2_1(\mathbf{x}) + a^1_{1,2}u^1_2(\mathbf{x}) + a^2_{1,2}u^2_2(\mathbf{x}),$$
$$f_2(\mathbf{x}) = a^1_{2,1}u^1_1(\mathbf{x}) + a^2_{2,1}u^2_1(\mathbf{x}) + a^1_{2,2}u^1_2(\mathbf{x}) + a^2_{2,2}u^2_2(\mathbf{x}),$$



LMC: covariance for $\mathbf{f}(\mathbf{x})$

- For $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}) \cdots f_D(\mathbf{x})]^\top$, the covariance $\text{cov}[\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')]$ is given as

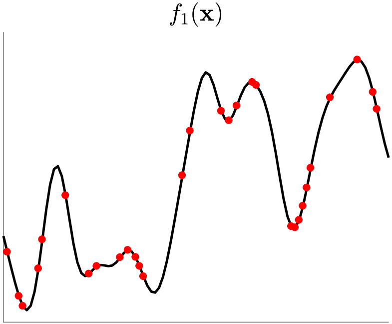
$$\text{cov}[\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')] = \sum_{q=1}^{Q} \mathbf{A}_q \mathbf{A}_q^\top k_q(\mathbf{x}, \mathbf{x}') = \sum_{q=1}^{Q} \mathbf{B}_q k_q(\mathbf{x}, \mathbf{x}'),$$

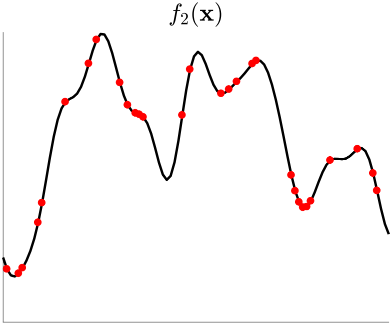
  where $\mathbf{A}_q = [\mathbf{a}_q^1 \ \mathbf{a}_q^2 \cdots \mathbf{a}_q^{R_q}]$.

- The rank of each $\mathbf{B}_q$ is R_q.

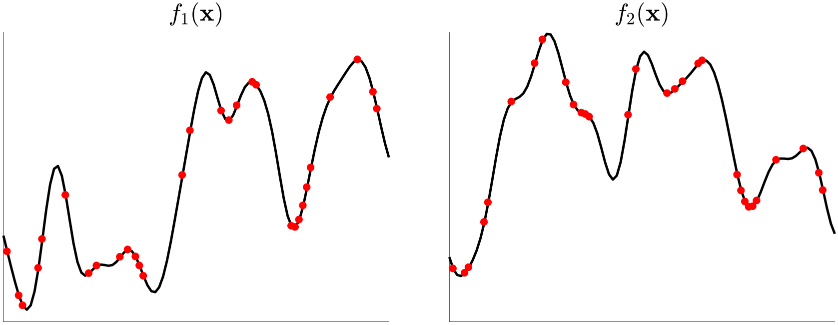
- The matrices $\mathbf{B}_q$ are known as the *coregionalization matrices*.

LMC: observed data



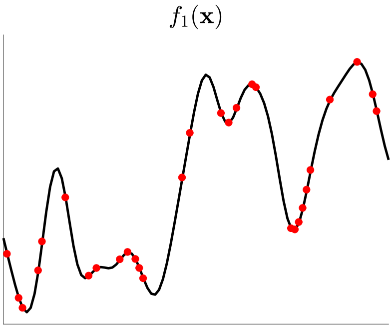


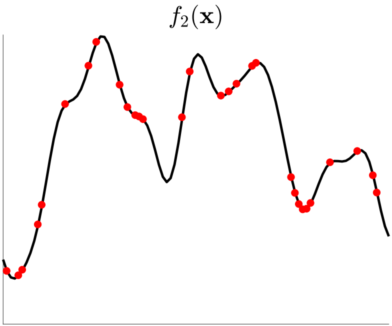
LMC: observed data



$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i)) | i = 1, \dots, N\}$ $\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i)) | i = 1, \dots, N\}$

LMC: observed data

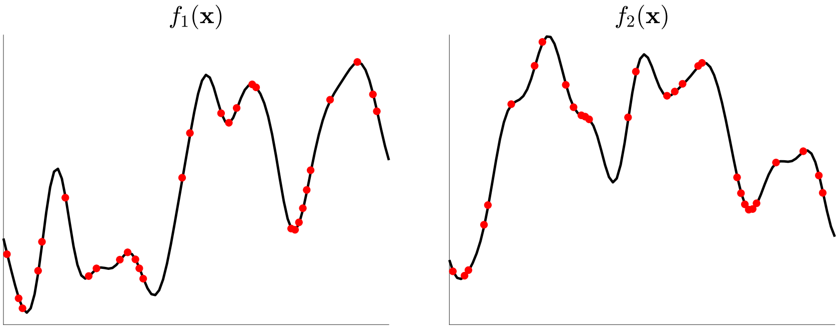




$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))|i = 1, \ldots, N\}$ $\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))|i = 1, \ldots, N\}$

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1) \\ \vdots \\ f_1(\mathbf{x}_N) \\ f_2(\mathbf{x}_1) \\ \vdots \\ f_2(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \sum_{q=1}^{Q} \mathbf{B}_q \otimes \mathbf{K}_q \right)$$

LMC: observed data

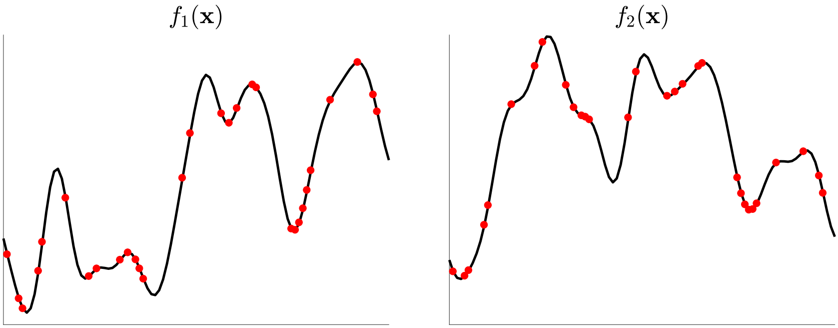


$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))|i = 1, \dots, N\}$ $\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))|i = 1, \dots, N\}$

$$\begin{bmatrix}\mathbf{f}_1\\ \mathbf{f}_2\end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1) \\ \vdots \\ f_1(\mathbf{x}_N) \\ f_2(\mathbf{x}_1) \\ \vdots \\ f_2(\mathbf{x}_N)\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \sum_{q=1}^{Q} \mathbf{B}_q \otimes \mathbf{K}_q\right)$$

The matrix $\mathbf{K}_q \in \mathbb{R}^{N\times N}$ has elements $k_q(\mathbf{x}_i, \mathbf{x}_j)$.

LMC: observed data



$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))|i = 1, \ldots, N\}$ $\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))|i = 1, \ldots, N\}$

$$\begin{bmatrix}\mathbf{f}_1\\ \mathbf{f}_2\end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1)\\ \vdots \\ f_1(\mathbf{x}_N)\\ f_2(\mathbf{x}_1)\\ \vdots\\ f_2(\mathbf{x}_N)\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \sum_{q=1}^{Q} \mathbf{B}_q \otimes \mathbf{K}_q\right)$$

The matrix $\mathbf{K}_q \in \mathbb{R}^{N\times N}$ has elements $k_q(\mathbf{x}_i, \mathbf{x}_j)$.

The matrix $\mathbf{B}_q \in \mathbb{R}^{D\times D}$ has elements b^q_{ij}.

Contents

Moving average function

- Consider again a set of D functions $\{f_d(\mathbf{x})\}_{d=1}^{D}$.

- Each function could be expressed through a convolution integral between a kernel, $\{G_d(\mathbf{x})\}_{d=1}^{D}$, and a function $u(\mathbf{x})$,

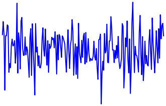
$$f_d(\mathbf{x}) = \int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z} = G_d(\mathbf{x}) * u(\mathbf{x}).$$

- For the integral to exist, it is assumed that the kernel $G_d(\mathbf{x})$ is a continuous function with compact support or square-integrable.

- The kernel $G_d(\mathbf{x})$ is also known as the moving average function or the smoothing kernel.

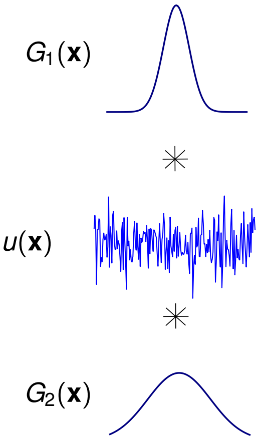
- In Dependet Gaussian processes (DGP) the latent function $u(\mathbf{x})$ is white Gaussian noise (WGN).

A pictorial representation

$u(\mathbf{x})$ 

$u(\mathbf{x})$: latent function.

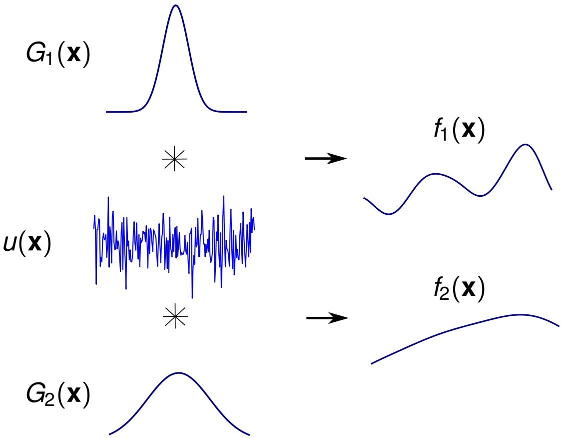
A pictorial representation



$u(\mathbf{x})$: latent function.

$G_1(\mathbf{x})$, $G_2(\mathbf{x})$: smoothing kernels.

A pictorial representation



$u(\mathbf{x})$: latent function.

$G_1(\mathbf{x})$, $G_2(\mathbf{x})$: smoothing kernels.

$f_1(\mathbf{x})$, $f_2(\mathbf{x})$: output functions.

Cross-covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x})$

- The cross-covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$, $\mathrm{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right]$, is

$$\mathbb{E}\left[\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z}\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')u(\mathbf{z}')\mathrm{d}\mathbf{z}'\right] -$$
$$\mathbb{E}\left[\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z}\right]\mathbb{E}\left[\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')u(\mathbf{z}')\mathrm{d}\mathbf{z}'\right]$$
$$= \int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')\,\mathrm{E}\left[u(\mathbf{z})u(\mathbf{z}')\right]\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z} -$$
$$\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})\mathbb{E}\left[u(\mathbf{z})\right]\mathrm{d}\mathbf{z}\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')\mathbb{E}\left[u(\mathbf{z}')\right]\mathrm{d}\mathbf{z}'$$
$$= \int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')\times$$
$$\left\{\mathrm{E}\left[u(\mathbf{z})u(\mathbf{z}')\right] - \mathbb{E}\left[u(\mathbf{z})\right]\mathbb{E}\left[u(\mathbf{z}')\right]\right\}\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{z}'$$
$$= \int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')k(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{z}'$$

- In the DGP $k(\mathbf{z},\mathbf{z}') = \sigma^2\delta(\mathbf{z}-\mathbf{z}')$.

Cross-covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x})$

- The cross-covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$, $\mathrm{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right]$, is

$$\mathbb{E}\left[\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z}\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')u(\mathbf{z}')\mathrm{d}\mathbf{z}'\right]-$$
$$\mathbb{E}\left[\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z}\right]\mathbb{E}\left[\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')u(\mathbf{z}')\mathrm{d}\mathbf{z}'\right]$$
$$=\int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')\,\mathrm{E}\left[u(\mathbf{z})u(\mathbf{z}')\right]\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}-$$
$$\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})\mathbb{E}\left[u(\mathbf{z})\right]\mathrm{d}\mathbf{z}\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')\mathbb{E}\left[u(\mathbf{z}')\right]\mathrm{d}\mathbf{z}'$$
$$=\int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')\times$$
$$\left\{\mathrm{E}\left[u(\mathbf{z})u(\mathbf{z}')\right]-\mathbb{E}\left[u(\mathbf{z})\right]\mathbb{E}\left[u(\mathbf{z}')\right]\right\}\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{z}'$$
$$=\int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')k(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{z}'$$

- In the DGP $k(\mathbf{z},\mathbf{z}')=\sigma^2\delta(\mathbf{z}-\mathbf{z}')$.

Cross-covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x})$

- The cross-covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$, $\operatorname{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right]$, is

$$
\begin{aligned}
&\mathbb{E}\left[\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z} \int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')u(\mathbf{z}')\mathrm{d}\mathbf{z}'\right] - \\
&\mathbb{E}\left[\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z}\right] \mathbb{E}\left[\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')u(\mathbf{z}')\mathrm{d}\mathbf{z}'\right] \\
&= \int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')\,\mathrm{E}\left[u(\mathbf{z})u(\mathbf{z}')\right]\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z} - \\
&\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})\mathbb{E}\left[u(\mathbf{z})\right]\mathrm{d}\mathbf{z} \int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')\mathbb{E}\left[u(\mathbf{z}')\right]\mathrm{d}\mathbf{z}' \\
&= \int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')\times \\
&\left\{\mathrm{E}\left[u(\mathbf{z})u(\mathbf{z}')\right] - \mathbb{E}\left[u(\mathbf{z})\right]\mathbb{E}\left[u(\mathbf{z}')\right]\right\}\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{z}' \\
&= \int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')k(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{z}'
\end{aligned}
$$

- In the DGP $k(\mathbf{z},\mathbf{z}') = \sigma^2\delta(\mathbf{z}-\mathbf{z}')$.

Cross-covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x})$

- The cross-covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$, $\operatorname{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right]$, is

$$\mathbb{E}\left[\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z}\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')u(\mathbf{z}')\mathrm{d}\mathbf{z}'\right]-$$

$$\mathbb{E}\left[\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z}\right]\mathbb{E}\left[\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')u(\mathbf{z}')\mathrm{d}\mathbf{z}'\right]$$

$$=\int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')\,\mathrm{E}\left[u(\mathbf{z})u(\mathbf{z}')\right]\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}-$$

$$\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})\mathbb{E}\left[u(\mathbf{z})\right]\mathrm{d}\mathbf{z}\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')\mathbb{E}\left[u(\mathbf{z}')\right]\mathrm{d}\mathbf{z}'$$

$$=\int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')\times$$

$$\left\{\mathrm{E}\left[u(\mathbf{z})u(\mathbf{z}')\right]-\mathbb{E}\left[u(\mathbf{z})\right]\mathbb{E}\left[u(\mathbf{z}')\right]\right\}\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{z}'$$

$$=\int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')k(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{z}'$$

- In the DGP $k(\mathbf{z},\mathbf{z}')=\sigma^2\delta(\mathbf{z}-\mathbf{z}')$.

Cross-covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x})$

- The cross-covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$, $\text{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right]$, is

$$
\begin{aligned}
&\mathbb{E}\left[\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z}\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')u(\mathbf{z}')\mathrm{d}\mathbf{z}'\right] - \\
&\mathbb{E}\left[\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z}\right]\mathbb{E}\left[\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')u(\mathbf{z}')\mathrm{d}\mathbf{z}'\right] \\
&= \int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')\,\mathrm{E}\left[u(\mathbf{z})u(\mathbf{z}')\right]\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z} - \\
&\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})\mathbb{E}\left[u(\mathbf{z})\right]\mathrm{d}\mathbf{z}\int_{\mathcal{X}} G_{d'}(\mathbf{x}'-\mathbf{z}')\mathbb{E}\left[u(\mathbf{z}')\right]\mathrm{d}\mathbf{z}' \\
&= \int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')\times \\
&\left\{\mathrm{E}\left[u(\mathbf{z})u(\mathbf{z}')\right] - \mathbb{E}\left[u(\mathbf{z})\right]\mathbb{E}\left[u(\mathbf{z}')\right]\right\}\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{z}' \\
&= \int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')k(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{z}'
\end{aligned}
$$

- In the DGP $k(\mathbf{z},\mathbf{z}') = \sigma^2\delta(\mathbf{z}-\mathbf{z}')$.

Example of $\text{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right]$ (I)

- The cross-covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$, $\text{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right]$, is

$$\text{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right] = \sigma^2 \int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z}) G_{d'}(\mathbf{x}'-\mathbf{z}) \mathrm{d}\mathbf{z}$$

- **Example**. Assume that the smoothing kernels follow a Gaussian form

$$G_d(\mathbf{x}-\mathbf{z}) = \frac{S_d |\mathbf{P}_d|^{1/2}}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{z})^\top \mathbf{P}_d(\mathbf{x}-\mathbf{z})\right],$$

- We use the identity of the product of two Gaussians

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_1, \mathbf{P}_1^{-1})\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_2, \mathbf{P}_2^{-1}) = \mathcal{N}(\boldsymbol{\mu}_1|\boldsymbol{\mu}_2, \mathbf{P}_1^{-1}+\mathbf{P}_2^{-1})\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_c, \mathbf{P}_c^{-1}),$$

where $\boldsymbol{\mu}_c = (\mathbf{P}_1+\mathbf{P}_2)^{-1}(\mathbf{P}_1\boldsymbol{\mu}_1 + \mathbf{P}_2\boldsymbol{\mu}_2)$ and $\mathbf{P}_c^{-1} = (\mathbf{P}_1+\mathbf{P}_2)^{-1}$.

Example of $\text{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right]$ (II)

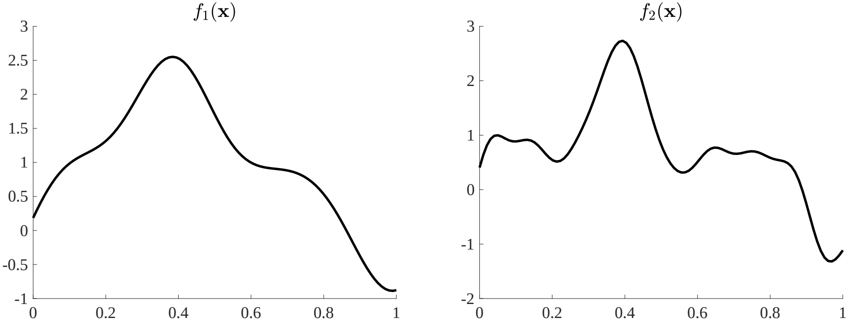
- The cross-covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$, $\text{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right]$, is

$$
\begin{aligned}
\text{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right] &= \sigma^2 \int_{\mathcal{X}} G_d(\mathbf{x} - \mathbf{z}) G_{d'}(\mathbf{x}' - \mathbf{z}) \mathrm{d}\mathbf{z} \\
&= \frac{\sigma^2 S_d S_{d'}}{(2\pi)^{p/2} |\mathbf{P}_{\text{eqv}}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^\top \mathbf{P}_{\text{eqv}}^{-1} (\mathbf{x} - \mathbf{x}')\right],
\end{aligned}
$$

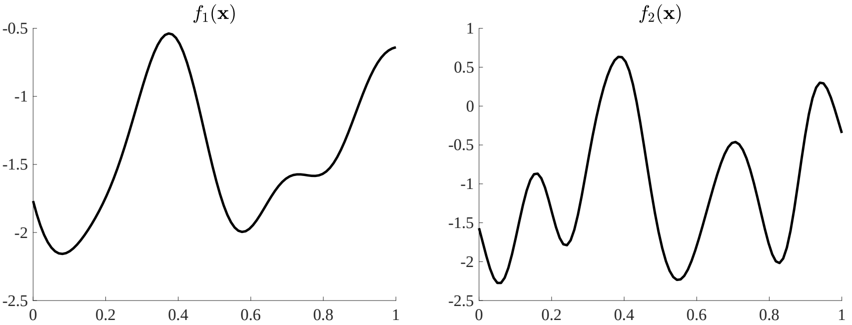
where $\mathbf{P}_{\text{eqv}} = \mathbf{P}_d^{-1} + \mathbf{P}_{d'}^{-1}$.

- **Exercise.** Show how to obtain the expression above

PC: samples

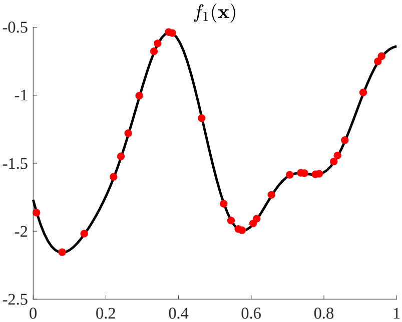

$f_1(\mathbf{x})$
3
2.5
2
1.5
1
0.5
0
-0.5
-1
0
0.2
0.4
0.6
0.8
1
$f_2(\mathbf{x})$
3
2
1
0
-1
-2
0
0.2
0.4
0.6
0.8
1

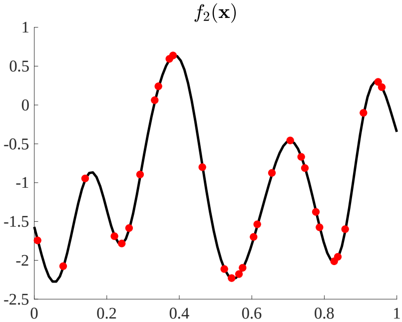
PC: samples



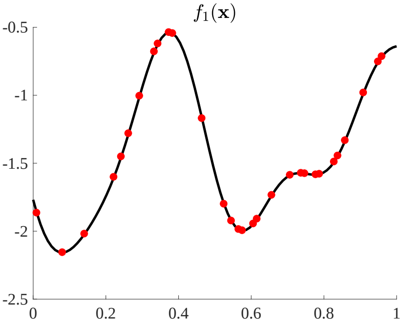
$f_1(\mathbf{x})$
-0.5
-1
-1.5
-2
-2.5
0
0.2
0.4
0.6
0.8
1
$f_2(\mathbf{x})$
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
0.2
0.4
0.6
0.8
1

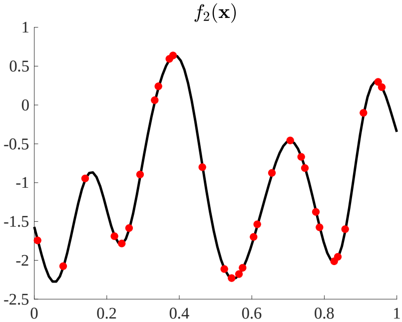
PC: observed data





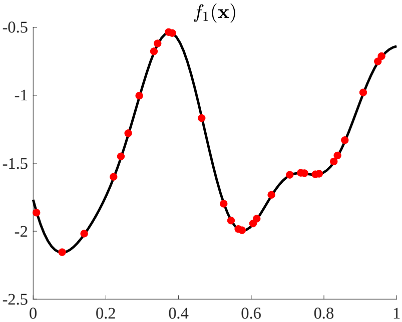
PC: observed data



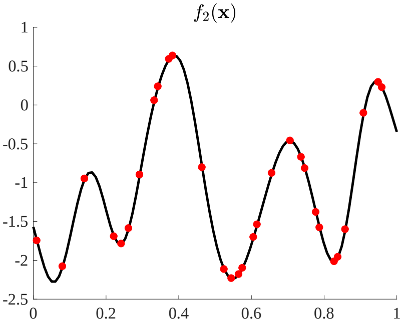


$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i)) | i = 1, \dots, N\}$ $\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i)) | i = 1, \dots, N\}$

PC: observed data



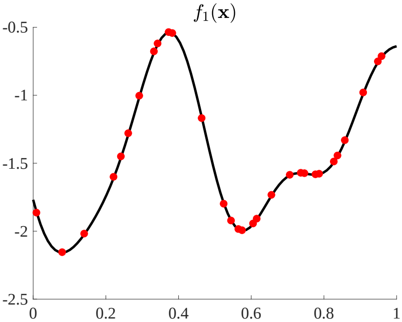
$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i))|i = 1, \dots, N\}$



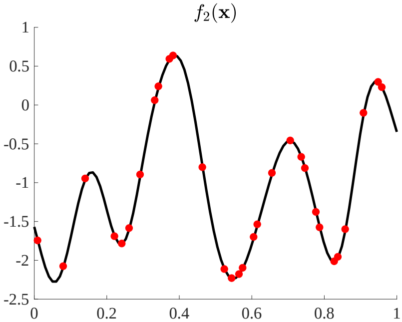
$\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i))|i = 1, \dots, N\}$

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1) \\ \vdots \\ f_1(\mathbf{x}_N) \\ f_2(\mathbf{x}_1) \\ \vdots \\ f_2(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_2} \\ \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_2} \end{bmatrix}, \right)$$

PC: observed data



$$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

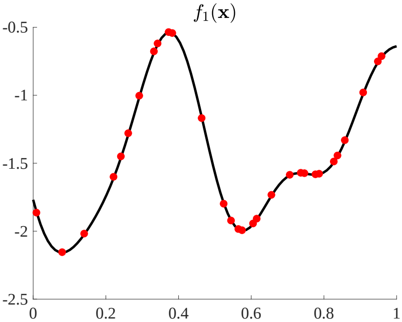


$$\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

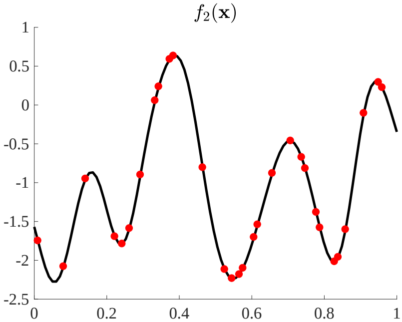
$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1) \\ \vdots \\ f_1(\mathbf{x}_N) \\ f_2(\mathbf{x}_1) \\ \vdots \\ f_2(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_2} \\ \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_2} \end{bmatrix}, \right)$$

The matrix $\mathbf{K}_{\mathbf{f}_d,\mathbf{f}_d} \in \mathbb{R}^{N\times N}$ has elements $\text{cov}\,[f_d(\mathbf{x}), f_d(\mathbf{x}')]$.

PC: observed data



$$\mathcal{D}_1 = \{(\mathbf{x}_i, f_1(\mathbf{x}_i)) | i = 1, \ldots, N\}$$



$$\mathcal{D}_2 = \{(\mathbf{x}_i, f_2(\mathbf{x}_i)) | i = 1, \ldots, N\}$$

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}_1) \\ \vdots \\ f_1(\mathbf{x}_N) \\ f_2(\mathbf{x}_1) \\ \vdots \\ f_2(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_2} \\ \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_2} \end{bmatrix}, \right)$$

The matrix $\mathbf{K}_{\mathbf{f}_d,\mathbf{f}_d} \in \mathbb{R}^{N \times N}$ has elements $\mathrm{cov}\,[f_d(\mathbf{x}), f_d(\mathbf{x}')]$.

The matrix $\mathbf{K}_{\mathbf{f}_d,\mathbf{f}_{d'}} \in \mathbb{R}^{N \times N}$ has elements $\mathrm{cov}\,[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')]$.

Beyond $u(\mathbf{x})$ as a white Gaussian noise

- Consider again a set of D functions $\{f_d(\mathbf{x})\}_{d=1}^{D}$.

- Each function could be expressed through a convolution integral between a kernel, $\{G_d(\mathbf{x})\}_{d=1}^{D}$, and a function $u(\mathbf{x})$,

$$f_d(\mathbf{x}) = \int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z} = G_d(\mathbf{x}) * u(\mathbf{x}).$$

- Assuming $u(\mathbf{x})$ is a GP with zero mean and covariance $k(\mathbf{x},\mathbf{x}')$.

- The cross-covariance is now given as

$$\operatorname{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right] = \int_{\mathcal{X}}\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z}')k(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{z}'$$

A process $u(\mathbf{x})$ with covariance $k(\mathbf{x}, \mathbf{x}')$

- The cross-covariance is

$$\operatorname{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right] = \int_{\mathcal{X}} \int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z}) G_{d'}(\mathbf{x}'-\mathbf{z}') k(\mathbf{z}, \mathbf{z}') \mathrm{d}\mathbf{z} \mathrm{d}\mathbf{z}'$$

- **Example.** Assume that the smoothing kernels and the covariance for $u(\mathbf{x})$ follow a Gaussian form

$$G_d(\mathbf{x}-\mathbf{z}) = \frac{S_d |\mathbf{P}_d|^{1/2}}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{z})^\top \mathbf{P}_d (\mathbf{x}-\mathbf{z})\right],$$

$$k(\mathbf{z}, \mathbf{z}') = \frac{|\Lambda|^{1/2}}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\mathbf{z}-\mathbf{z}')^\top \Lambda (\mathbf{z}-\mathbf{z}')\right],$$

- Using again the identities of products of two Gaussians, we get

$$\begin{aligned}\operatorname{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right] &= \int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z}) G_{d'}(\mathbf{x}'-\mathbf{z}) \mathrm{d}\mathbf{z} \\ &= \frac{S_d S_{d'}}{(2\pi)^{p/2} |\mathbf{P}_{\text{eqv}}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{x}')^\top \mathbf{P}_{\text{eqv}}^{-1} (\mathbf{x}-\mathbf{x}')\right],\end{aligned}$$

where $\mathbf{P}_{\text{eqv}} = \mathbf{P}_d^{-1} + \mathbf{P}_{d'}^{-1} + \Lambda^{-1}$.

A process $u(\mathbf{x})$ with covariance $k(\mathbf{x}, \mathbf{x}')$

- The cross-covariance is

$$\operatorname{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right] = \int_{\mathcal{X}} \int_{\mathcal{X}} G_d(\mathbf{x} - \mathbf{z}) G_{d'}(\mathbf{x}' - \mathbf{z}') k(\mathbf{z}, \mathbf{z}') \mathrm{d}\mathbf{z} \mathrm{d}\mathbf{z}'$$

- **Example.** Assume that the smoothing kernels and the covariance for $u(\mathbf{x})$ follow a Gaussian form

$$G_d(\mathbf{x} - \mathbf{z}) = \frac{S_d |\mathbf{P}_d|^{1/2}}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{z})^\top \mathbf{P}_d (\mathbf{x} - \mathbf{z})\right],$$

$$k(\mathbf{z}, \mathbf{z}') = \frac{|\boldsymbol{\Lambda}|^{1/2}}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{z}')^\top \boldsymbol{\Lambda} (\mathbf{z} - \mathbf{z}')\right],$$

- Using again the identities of products of two Gaussians, we get

$$\begin{aligned}\operatorname{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right] &= \int_{\mathcal{X}} G_d(\mathbf{x} - \mathbf{z}) G_{d'}(\mathbf{x}' - \mathbf{z}) \mathrm{d}\mathbf{z} \\ &= \frac{S_d S_{d'}}{(2\pi)^{p/2} |\mathbf{P}_{\text{eqv}}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^\top \mathbf{P}_{\text{eqv}}^{-1} (\mathbf{x} - \mathbf{x}')\right],\end{aligned}$$

where $\mathbf{P}_{\text{eqv}} = \mathbf{P}_d^{-1} + \mathbf{P}_{d'}^{-1} + \boldsymbol{\Lambda}^{-1}$.

A process $u(\mathbf{x})$ with covariance $k(\mathbf{x}, \mathbf{x}')$

- The cross-covariance is

$$\text{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right] = \int_{\mathcal{X}} \int_{\mathcal{X}} G_d(\mathbf{x} - \mathbf{z}) G_{d'}(\mathbf{x}' - \mathbf{z}') k(\mathbf{z}, \mathbf{z}') \mathrm{d}\mathbf{z} \mathrm{d}\mathbf{z}'$$

- **Example.** Assume that the smoothing kernels and the covariance for $u(\mathbf{x})$ follow a Gaussian form

$$G_d(\mathbf{x} - \mathbf{z}) = \frac{S_d |\mathbf{P}_d|^{1/2}}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{z})^\top \mathbf{P}_d (\mathbf{x} - \mathbf{z})\right],$$

$$k(\mathbf{z}, \mathbf{z}') = \frac{|\boldsymbol{\Lambda}|^{1/2}}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{z}')^\top \boldsymbol{\Lambda} (\mathbf{z} - \mathbf{z}')\right],$$

- Using again the identities of products of two Gaussians, we get

$$\begin{aligned} \text{cov}\left[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')\right] &= \int_{\mathcal{X}} G_d(\mathbf{x} - \mathbf{z}) G_{d'}(\mathbf{x}' - \mathbf{z}) \mathrm{d}\mathbf{z} \\ &= \frac{S_d S_{d'}}{(2\pi)^{p/2} |\mathbf{P}_{\text{eqv}}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^\top \mathbf{P}_{\text{eqv}}^{-1} (\mathbf{x} - \mathbf{x}')\right], \end{aligned}$$

where $\mathbf{P}_{\text{eqv}} = \mathbf{P}_d^{-1} + \mathbf{P}_{d'}^{-1} + \boldsymbol{\Lambda}^{-1}$.

More general process convolutions

- We can include more latent processes $u_1(\mathbf{x}), u_2(\mathbf{x}), \ldots, u_Q(\mathbf{x})$

$$f_d(\mathbf{x}) = \sum_{q=1}^{Q} \sum_{i=1}^{R_q} \int_{\mathcal{X}} G_{d,q}^i(\mathbf{x}-\mathbf{z}) u_q^i(\mathbf{z}) \mathrm{d}\mathbf{z},$$

where $\operatorname{cov}[u_q^i(\mathbf{z}), u_{q'}^{i'}(\mathbf{z}')] = k_q(\mathbf{z}, \mathbf{z}') \delta_{i,i'} \delta_{q,q'}$.

- A general expression for $\operatorname{cov}[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')]$ follows as

$$\sum_{q=1}^{Q} \sum_{i=1}^{R_q} \int_{\mathcal{X}} G_{d,q}^i(\mathbf{x}-\mathbf{z}) \int_{\mathcal{X}} G_{d',q}^i(\mathbf{x}'-\mathbf{z}') k_q(\mathbf{z}, \mathbf{z}') \mathrm{d}\mathbf{z}' \mathrm{d}\mathbf{z}.$$

More general process convolutions

- We can include more latent processes $u_1(\mathbf{x}), u_2(\mathbf{x}), \ldots, u_Q(\mathbf{x})$

$$f_d(\mathbf{x}) = \sum_{q=1}^{Q} \sum_{i=1}^{R_q} \int_{\mathcal{X}} G^i_{d,q}(\mathbf{x} - \mathbf{z}) u^i_q(\mathbf{z}) \mathrm{d}\mathbf{z},$$

where $\operatorname{cov}[u^i_q(\mathbf{z}), u^{i'}_{q'}(\mathbf{z}')] = k_q(\mathbf{z}, \mathbf{z}')\delta_{i,i'}\delta_{q,q'}$.

- A general expression for $\operatorname{cov}[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')]$ follows as

$$\sum_{q=1}^{Q} \sum_{i=1}^{R_q} \int_{\mathcal{X}} G^i_{d,q}(\mathbf{x} - \mathbf{z}) \int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}' - \mathbf{z}') k_q(\mathbf{z}, \mathbf{z}') \mathrm{d}\mathbf{z}' \mathrm{d}\mathbf{z}.$$

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G_{d,q}^{i}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G_{d',q}^{i}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Intrinsic Coregionalization Model [Goovaerts, 1997] or **Multi-task Gaussian Processes** [Bonilla et al., 2008]

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Intrinsic Coregionalization Model [Goovaerts, 1997] or **Multi-task Gaussian Processes** [Bonilla et al., 2008]

$$G^i_{d,q}(\mathbf{x}-\mathbf{z}) = a^i_{d,q}\delta(\mathbf{x}-\mathbf{z}), \quad Q=1, \quad R_q>1$$

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G_{d,q}^{i}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G_{d',q}^{i}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Intrinsic Coregionalization Model [Goovaerts, 1997] or **Multi-task Gaussian Processes** [Bonilla et al., 2008]

$$G_{d,q}^{i}(\mathbf{x}-\mathbf{z}) = a_{d,q}^{i}\delta(\mathbf{x}-\mathbf{z}), \quad Q=1, \quad R_q>1$$

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{R_1} a_{d,1}^{i}a_{d',1}^{i}k_1(\mathbf{x},\mathbf{x}').$$

Starting with the general expression we had before ...

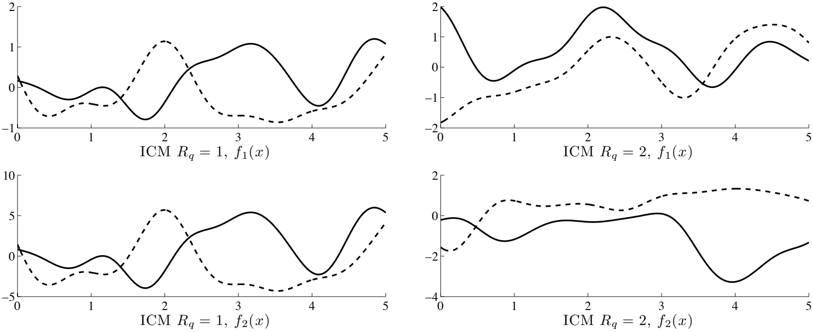
Intrinsic Coregionalization Model

$$\mathbf{K}_{\mathbf{f},\mathbf{f}} = \mathbf{B} \otimes \mathbf{K}$$

Starting with the general expression we had before ...

Intrinsic Coregionalization Model

$$\mathbf{K}_{\mathbf{f},\mathbf{f}} = \mathbf{B} \otimes \mathbf{K}$$



Rank 1

Rank 2

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows [Higdon, 2002, Boyle and Frean, 2005, Álvarez et al., 2012]

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows [Higdon, 2002, Boyle and Frean, 2005, Álvarez et al., 2012]

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^{i}_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^{i}_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Semiparametric Latent Factor Model [Teh et al., 2005]

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows [Higdon, 2002, Boyle and Frean, 2005, Álvarez et al., 2012]

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Semiparametric Latent Factor Model [Teh et al., 2005]

$$G^i_{d,q}(\mathbf{x}-\mathbf{z}) = a^i_{d,q}\delta(\mathbf{x}-\mathbf{z}), \quad R_q = 1, \quad Q > 1$$

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows [Higdon, 2002, Boyle and Frean, 2005, Álvarez et al., 2012]

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Semiparametric Latent Factor Model [Teh et al., 2005]

$$G^i_{d,q}(\mathbf{x}-\mathbf{z}) = a^i_{d,q}\delta(\mathbf{x}-\mathbf{z}), \quad R_q = 1, \quad Q > 1$$

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q} a^1_{d,q}a^1_{d',q}k_q(\mathbf{x},\mathbf{x}').$$

Starting with the general expression we had before ...

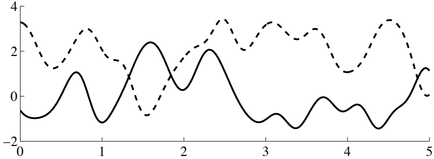
Semiparametric Latent Factor Model

$$\mathbf{K}_{\mathbf{f},\mathbf{f}} = \sum_{q=1}^{Q} \mathbf{a}_q \mathbf{a}_q^\top \otimes \mathbf{K}_q$$

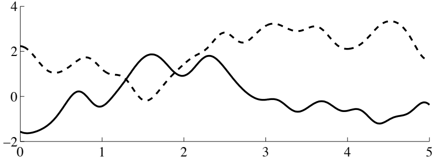
Starting with the general expression we had before ...

Semiparametric Latent Factor Model

$$\mathbf{K}_{\mathbf{f},\mathbf{f}} = \sum_{q=1}^{Q} \mathbf{a}_q \mathbf{a}_q^\top \otimes \mathbf{K}_q$$



LMC with $R_q = 1$ and $Q = 2$, $f_1(x)$



LMC with $R_q = 1$ and $Q = 2$, $f_2(x)$

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows [Higdon, 2002, Boyle and Frean, 2005, Álvarez et al., 2012]

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows [Higdon, 2002, Boyle and Frean, 2005, Álvarez et al., 2012]

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Linear Model of Coregionalization [Journel and Huijbregts, 1978, Goovaerts, 1997, Wackernagel, 2003].

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows [Higdon, 2002, Boyle and Frean, 2005, Álvarez et al., 2012]

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G_{d,q}^{i}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G_{d',q}^{i}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Linear Model of Coregionalization [Journel and Huijbregts, 1978, Goovaerts, 1997, Wackernagel, 2003].

$$G_{d,q}^{i}(\mathbf{x}-\mathbf{z}) = a_{d,q}^{i}\delta(\mathbf{x}-\mathbf{z}), \quad R_q > 1, \quad Q > 1,$$

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows [Higdon, 2002, Boyle and Frean, 2005, Álvarez et al., 2012]

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G_{d,q}^{i}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G_{d',q}^{i}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Linear Model of Coregionalization [Journel and Huijbregts, 1978, Goovaerts, 1997, Wackernagel, 2003].

$$G_{d,q}^{i}(\mathbf{x}-\mathbf{z}) = a_{d,q}^{i}\delta(\mathbf{x}-\mathbf{z}), \quad R_q > 1, \quad Q > 1,$$

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q} a_{d,q}^{i}a_{d',q}^{i}k_q(\mathbf{x},\mathbf{x}').$$

Starting with the general expression we had before ...

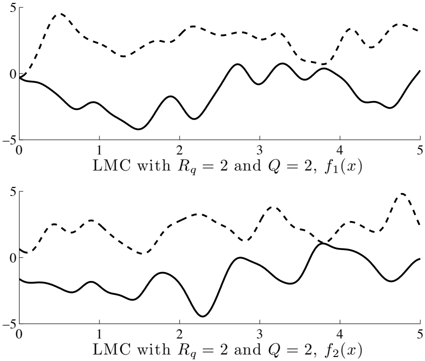
Linear Model of Coregionalization

$$\mathbf{K}_{\mathbf{f},\mathbf{f}} = \sum_{q=1}^{Q} \mathbf{B}_q \otimes \mathbf{K}_q$$

Starting with the general expression we had before ...

Linear Model of Coregionalization

$$\mathbf{K}_{\mathbf{f},\mathbf{f}} = \sum_{q=1}^{Q} \mathbf{B}_q \otimes \mathbf{K}_q$$



Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows [Higdon, 2002, Boyle and Frean, 2005, Álvarez et al., 2012]

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^{i}_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^{i}_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows [Higdon, 2002, Boyle and Frean, 2005, Álvarez et al., 2012]

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^{i}_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^{i}_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Dependent GPs [Higdon, 2002, Boyle and Frean, 2005]

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows [Higdon, 2002, Boyle and Frean, 2005, Álvarez et al., 2012]

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Dependent GPs [Higdon, 2002, Boyle and Frean, 2005]

$$Q=1, \quad R_q = 1 \quad k_1(\mathbf{z},\mathbf{z}') = \sigma^2\delta(\mathbf{z},\mathbf{z}'),$$

Starting with the general expression we had before ...

Assume we have D outputs, $\{f_d(\mathbf{x})\}_{d=1}^D$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows [Higdon, 2002, Boyle and Frean, 2005, Álvarez et al., 2012]

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

Dependent GPs [Higdon, 2002, Boyle and Frean, 2005]

$$Q = 1, \quad R_q = 1 \quad k_1(\mathbf{z},\mathbf{z}') = \sigma^2\delta(\mathbf{z},\mathbf{z}'),$$

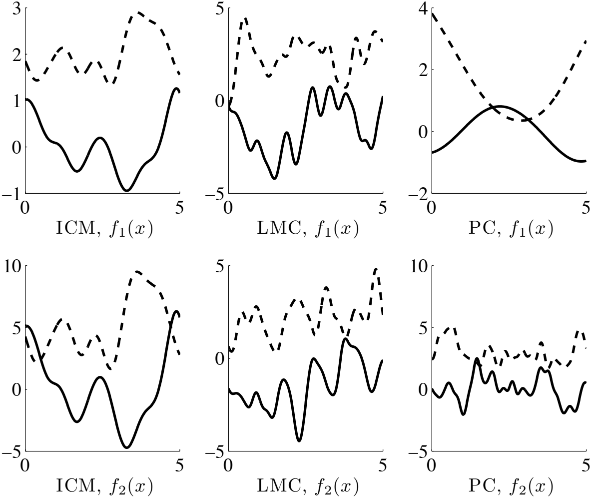
$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sigma^2\int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})G_{d'}(\mathbf{x}'-\mathbf{z})\mathrm{d}\mathbf{z}.$$

Starting with the general expression we had before ...

Comparison

Starting with the general expression we had before ...

Comparison



Kernels for vector-valued functions

Foundations and Trends® in
Machine Learning
Vol. 4, No. 3 (2011) 195–266

DOI: 10.1561/2200000036

the essence of knowledge

Kernels for Vector-Valued Functions: A Review

By Mauricio A. Álvarez,
Lorenzo Rosasco and Neil D. Lawrence

Contents

Gaussian process priors for vector-valued functions

- We saw a series of models for the set of outputs $\{f_d(\mathbf{x})\}_{d=1}^{D}$, that led to a valid covariance function for the vector $\mathbf{f}(\mathbf{x})$.

- For a finite number of inputs, $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^{N}$, the prior distribution over the vector $\mathbf{f} = [\mathbf{f}_1^\top, \ldots, \mathbf{f}_D^\top]^\top$ is given as

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_D \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_2} & \cdots & \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_D} \\ \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_2} & \cdots & \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_D} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{K}_{\mathbf{f}_D,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_D,\mathbf{f}_2} & \cdots & \mathbf{K}_{\mathbf{f}_D,\mathbf{f}_D} \end{bmatrix} \right).$$

Gaussian process priors for vector-valued functions

- We saw a series of models for the set of outputs $\{f_d(\mathbf{x})\}_{d=1}^{D}$, that led to a valid covariance function for the vector $\mathbf{f}(\mathbf{x})$.

- For a finite number of inputs, $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^{N}$, the prior distribution over the vector $\mathbf{f} = [\mathbf{f}_1^\top, \dots, \mathbf{f}_D^\top]^\top$ is given as

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_D \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_2} & \cdots & \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_D} \\ \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_2} & \cdots & \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_D} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{K}_{\mathbf{f}_D,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_D,\mathbf{f}_2} & \cdots & \mathbf{K}_{\mathbf{f}_D,\mathbf{f}_D} \end{bmatrix} \right).$$

Gaussian process priors for vector-valued functions

- We saw a series of models for the set of outputs $\{f_d(\mathbf{x})\}_{d=1}^{D}$, that led to a valid covariance function for the vector $\mathbf{f}(\mathbf{x})$.

- For a finite number of inputs, $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^{N}$, the prior distribution over the vector $\mathbf{f} = [\mathbf{f}_1^\top, \ldots, \mathbf{f}_D^\top]^\top$ is given as

$$\underbrace{\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_D \end{bmatrix}}_{\mathbf{f}} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \underbrace{\begin{bmatrix} \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_2} & \cdots & \mathbf{K}_{\mathbf{f}_1,\mathbf{f}_D} \\ \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_2} & \cdots & \mathbf{K}_{\mathbf{f}_2,\mathbf{f}_D} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{K}_{\mathbf{f}_D,\mathbf{f}_1} & \mathbf{K}_{\mathbf{f}_D,\mathbf{f}_2} & \cdots & \mathbf{K}_{\mathbf{f}_D,\mathbf{f}_D} \end{bmatrix}}_{\mathbf{K}_{\mathbf{f},\mathbf{f}}} \right).$$

Noisy observations

- In practice, we usually have access to noisy observations, so we model the outputs $\{y_d(\mathbf{x})\}_{d=1}^{D}$ using

$$y_d(\mathbf{x}) = f_d(\mathbf{x}) + \epsilon_d(\mathbf{x}),$$

where $\{\epsilon_d(\mathbf{x})\}_{d=1}^{D}$ are independent white Gaussian noise processes with variance σ_d^2.

- The marginal likelihood is given as

$$p(\mathbf{y}|\mathbf{X}, \theta) = \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}_{\mathbf{f},\mathbf{f}} + \boldsymbol{\Sigma}),$$

where $\mathbf{y} = \left[\mathbf{y}_1^\top, \mathbf{y}_2^\top \ldots, \mathbf{y}_D^\top\right]^\top$

- The vector θ refers to the hyperparameters and $\boldsymbol{\Sigma} = \Sigma \otimes \mathbf{I}_N$.

Noisy observations

- In practice, we usually have access to noisy observations, so we model the outputs $\{y_d(\mathbf{x})\}_{d=1}^{D}$ using

$$y_d(\mathbf{x}) = f_d(\mathbf{x}) + \epsilon_d(\mathbf{x}),$$

where $\{\epsilon_d(\mathbf{x})\}_{d=1}^{D}$ are independent white Gaussian noise processes with variance σ_d^2.

- The marginal likelihood is given as

$$p(\mathbf{y}|\mathbf{X}, \theta) = \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}_{\mathbf{f},\mathbf{f}} + \boldsymbol{\Sigma}),$$

where $\mathbf{y} = \left[\mathbf{y}_1^\top, \mathbf{y}_2^\top \ldots, \mathbf{y}_D^\top\right]^\top$

- The vector θ refers to the hyperparameters and $\boldsymbol{\Sigma} = \Sigma \otimes \mathbf{I}_N$.

Noisy observations

- In practice, we usually have access to noisy observations, so we model the outputs $\{y_d(\mathbf{x})\}_{d=1}^{D}$ using

$$y_d(\mathbf{x}) = f_d(\mathbf{x}) + \epsilon_d(\mathbf{x}),$$

where $\{\epsilon_d(\mathbf{x})\}_{d=1}^{D}$ are independent white Gaussian noise processes with variance σ_d^2.

- The marginal likelihood is given as

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}_{\mathbf{f},\mathbf{f}} + \boldsymbol{\Sigma}),$$

where $\mathbf{y} = \left[\mathbf{y}_1^\top, \mathbf{y}_2^\top \ldots, \mathbf{y}_D^\top\right]^\top$

- The vector $\boldsymbol{\theta}$ refers to the hyperparameters and $\boldsymbol{\Sigma} = \Sigma \otimes \mathbf{I}_N$.

Hyperparameter Learning

- Let $\mathcal{D} = \{\mathbf{X}_n, \mathbf{y}_n\}_{n=1}^{N}$ represents the data, and $\boldsymbol{\theta}$ represents the hyperparameters of the covariance function.

- The marginal likelihood for the outputs can be written as

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}_{\mathbf{f},\mathbf{f}} + \boldsymbol{\Sigma}),$$

  where $\mathbf{K}_{\mathbf{f},\mathbf{f}} \in \mathbb{R}^{ND \times ND}$ with each element given by $\operatorname{cov}[f_d(\mathbf{x}_n), f_{d'}(\mathbf{x}_{n'})]$.

- The matrix $\boldsymbol{\Sigma}$ represents the covariance associated with some independent processes.

- Hyperparameters are estimated by maximizing the logarithm of the marginal likelihood.

Predictive distribution

- Prediction for a set of test inputs $\mathbf{X}_*$ is done using standard Gaussian process regression techniques.

- The predictive distribution is given by

$$p(\mathbf{y}_*|\mathbf{y}, \mathbf{X}, \theta) = \mathcal{N}(\mathbf{y}_*|\boldsymbol{\mu}_*, \mathbf{K}_{\mathbf{y}_*,\mathbf{y}_*}),$$

with

$$\begin{aligned}
\boldsymbol{\mu}_* &= \mathbf{K}_{\mathbf{f}_*,\mathbf{f}} \left(\mathbf{K}_{\mathbf{f},\mathbf{f}} + \Sigma\right)^{-1} \mathbf{y}, \\
\mathbf{K}_{\mathbf{y}_*,\mathbf{y}_*} &= \mathbf{K}_{\mathbf{f}_*,\mathbf{f}_*} - \mathbf{K}_{\mathbf{f}_*,\mathbf{f}} \left(\mathbf{K}_{\mathbf{f},\mathbf{f}} + \Sigma\right)^{-1} \mathbf{K}_{\mathbf{f}_*,\mathbf{f}}^{\top} + \Sigma_*.
\end{aligned}$$

Can you prove autokrigeability?

- The predictive distribution is given by

$$p(\mathbf{y}_*|\mathbf{y}, \mathbf{X}, \theta) = \mathcal{N}(\mathbf{y}_*|\boldsymbol{\mu}_*, \mathbf{K}_{\mathbf{y}_*,\mathbf{y}_*}),$$

with

$$\boldsymbol{\mu}_* = \mathbf{K}_{\mathbf{f}_*,\mathbf{f}} \left(\mathbf{K}_{\mathbf{f},\mathbf{f}} + \boldsymbol{\Sigma}\right)^{-1} \mathbf{y},$$
$$\mathbf{K}_{\mathbf{y}_*,\mathbf{y}_*} = \mathbf{K}_{\mathbf{f}_*,\mathbf{f}_*} - \mathbf{K}_{\mathbf{f}_*,\mathbf{f}} \left(\mathbf{K}_{\mathbf{f},\mathbf{f}} + \boldsymbol{\Sigma}\right)^{-1} \mathbf{K}_{\mathbf{f}_*,\mathbf{f}}^\top + \boldsymbol{\Sigma}_*.$$

- **Exercise**: Prove that if the outputs are considered to be noise-free, prediction using the ICM under an isotopic data case is equivalent to independent prediction over each output.

Contents

The cokriging estimator

- In geostatistics, the framework that allows for optimal predictions in the multivariate case is known by the general name of *cokriging* [Goovaerts, 1997].

- In general, the output value for f_d evaluated at $\mathbf{x}_*$ is estimated as

$$\hat{f}_d(\mathbf{x}_*) - \mu_d(\mathbf{x}_*) = \sum_{s=1}^{D} \sum_{\alpha_s=1}^{n_s(\mathbf{x}_*)} \lambda_{\alpha_s}(\mathbf{x}_*) \left[f_s(\mathbf{x}_{\alpha_s}) - \mu_s(\mathbf{x}_{\alpha_s}) \right],$$

  where $\lambda_{\alpha_s}(\mathbf{x}_*)$ are the weights assigned to the output data $f_s(\mathbf{x}_{\alpha_s})$, $\mu_s(\mathbf{x}_{\alpha_s})$ are the expected values of $f_s(\mathbf{x}_{\alpha_s})$, and $n_s(\mathbf{x}_*) \leq N$.

- Cokriging estimators need to be unbiased ($E[f_d(\mathbf{x}_*) - \hat{f}_d(\mathbf{x}_*)] = 0$) and minimize the error variance σ_E^2,

$$\sigma_E^2(\mathbf{x}_*) = \operatorname{var}\left[f_d(\mathbf{x}_*) - \hat{f}_d(\mathbf{x}_*) \right].$$

Cokriging assumes a model for f_d

- Cogriking estimators differ in the form they assume for $f_d(\mathbf{x})$.

- In general, each output function is decomposed into a residual $R_d(\mathbf{x})$ and a trend $\mu_d(\mathbf{x})$,

$$f_d(\mathbf{x}) = R_d(\mathbf{x}) + \mu_d(\mathbf{x}), \quad \forall d$$

- Residuals are assumed to be Gaussian processes with zero mean.

- The covariance for the residuals is denoted as $k_{d,d}(\mathbf{x}, \mathbf{x}')$ and the cross-covariance between residuals as $k_{d,d'}(\mathbf{x}, \mathbf{x}')$.

Simple cokriging

- The simple cokriging estimator is given as

$$\hat{f}_d(\mathbf{x}_*) - \mu_d = \sum_{s=1}^{D} \sum_{\alpha_s=1}^{n_s(\mathbf{x}_*)} \lambda_{\alpha_s}(\mathbf{x}_*) \left[f_s(\mathbf{x}_{\alpha_s}) - \mu_s) \right].$$

- It can be shown that this is an unbiased estimator.

- Coefficients $\lambda_{\alpha_s}(\mathbf{x}_*)$ can be obtained by minimizing the variance $\sigma_E^2(\mathbf{x}_*)$, leading to

$$\begin{bmatrix} \boldsymbol{\lambda}_1(\mathbf{x}_*) \\ \vdots \\ \boldsymbol{\lambda}_D(\mathbf{x}_*) \end{bmatrix} = \left(\begin{bmatrix} \mathbf{K}_{1,1} & \cdots & \mathbf{K}_{1,D} \\ \vdots & \vdots & \vdots \\ \mathbf{K}_{D,1} & \cdots & \mathbf{K}_{D,D} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{k}_{1,1} \\ \vdots \\ \mathbf{k}_{D,1} \end{bmatrix}$$

where $\mathbf{K}_{d,d'} = \left[k_{d,d'}(\mathbf{x}_{\alpha_d}, \mathbf{x}_{\beta_{d'}}) \right]$ and $\mathbf{k}_{d,1} = [k_{d,1}(\mathbf{x}_{\alpha_d}, \mathbf{x}_*)]$.

- The predictor is then $\hat{f}_d(\mathbf{x}_*) = \boldsymbol{\lambda}^\top \mathbf{f}$.

Contents

Contents

Efficient approximations (I)

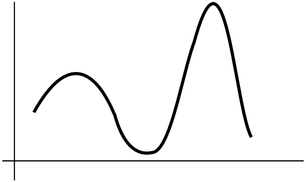
- Learning $\boldsymbol{\theta}$ through marginal likelihood maximization involves the inversion of the matrix $\mathbf{K}_{\mathbf{f},\mathbf{f}} + \boldsymbol{\Sigma}$.

- The inversion of this matrix scales as $\mathcal{O}(D^3N^3)$.

- If only a few number $K < N$ of values of $u(\mathbf{x})$ are known, then the set of outputs are uniquely determined.

Efficient approximations (II)

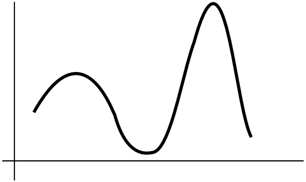
Sample from $p(u)$



$$f_d(\mathbf{x}) = \int_{\mathcal{X}} G_d(\mathbf{x} - \mathbf{z}) u(\mathbf{z}) \mathrm{d}\mathbf{z}$$

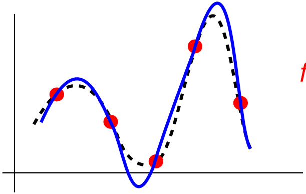
Efficient approximations (II)

Sample from $p(u)$



$$f_d(\mathbf{x}) = \int_{\mathcal{X}} G_d(\mathbf{x} - \mathbf{z}) u(\mathbf{z}) \mathrm{d}\mathbf{z}$$

Sample from $p(u|\mathbf{u})$



$$f_d(\mathbf{x}) \approx \int_{\mathcal{X}} G_d(\mathbf{x} - \mathbf{z}) \, \mathrm{E}\left[u(\mathbf{z})|\mathbf{u}\right] \mathrm{d}\mathbf{z}$$

Efficient approximations

Journal of Machine Learning Research 12 (2011) 1459-1500 Submitted 11/09; Revised 10/10; Published 5/11

Computationally Efficient Convolved Multiple Output Gaussian Processes

Mauricio A. Álvarez* MALVAREZ@UTP.EDU.CO
School of Computer Science
University of Manchester
Manchester, UK, M13 9PL

Neil D. Lawrence† N.LAWRENCE@SHEFFIELD.AC.UK
School of Computer Science
University of Sheffield
Sheffield, S1 4DP

Editor: Carl Edward Rasmussen

Contents

Cross-coregionalization matrices

- In the LMC

$$f_d(\mathbf{x}) = \sum_{q=1}^{Q} \sum_{i=1}^{R_q} a^i_{d,q} u^i_q(\mathbf{x}).$$

- The basic processes $u^i_q(\mathbf{x})$ [Guzmán et al., 2002] are assumed to be nonorthogonal, leading to the following covariance function

$$\mathrm{cov}[\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')] = \sum_{q=1}^{Q} \sum_{q'=1}^{Q} \mathbf{B}_{q,q'} k_{q,q'}(\mathbf{x}, \mathbf{x}'),$$

where $\mathbf{B}_{q,q'}$ are *cross-coregionalization* matrices. matrices.

Non-stationarity LMC

- We can write the vector-valued function $\mathbf{f}(\mathbf{x})$ as

$$\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{u}(\mathbf{x}),$$

where $\mathbf{A} = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_Q]$ and $\mathbf{u}(\mathbf{x}) = [u_1(\mathbf{x}) \ \cdots \ u_Q(\mathbf{x})]^\top$.

- A non-stationary version allows $\mathbf{A}$ to change with $\mathbf{x}$ [Gelfand et al., 2004, Wilson et al., 2012]

$$\mathbf{f}(\mathbf{x}) = \mathbf{A}(\mathbf{x})\mathbf{u}(\mathbf{x}).$$

Contents

Extensions [Calder and Cressie, 2007]

- A more general form

$$f_d(\mathbf{x}) = \int G_d(\mathbf{x}, \mathbf{z}) u(\mathbf{z}) \mathrm{d}\mathbf{z}$$

$$f_d(\mathbf{x}) = \sum_j G_d(\mathbf{x}, \mathbf{z}_j) u(\mathbf{z}_j)$$

- Non-stationary models

$$f_d(\mathbf{x}) = \int G_{d,\theta(\mathbf{x})}(\mathbf{x}, \mathbf{z}) u(\mathbf{z}) \mathrm{d}\mathbf{z},$$

$$f_d(\mathbf{x}) = \int G_d(\mathbf{x}, \mathbf{z}) u_{\theta(\mathbf{z})}(\mathbf{x}) \mathrm{d}\mathbf{z}$$

Latent force models [Álvarez et al., 2009]

- Mechanistically inspired kernel smoothing functions.

$$
\begin{array}{ll}
G_d(t,t') \propto \exp\left[-D_q\left(t-t'\right)\right] & \text{first ODE} \\
G_d(t,t') \propto \exp\left[-\alpha_q\left(t-t'\right)\right]\sin\left[\omega_q\left(t-t'\right)\right] & \text{second ODE} \\
G_d(\mathbf{x},\mathbf{x}') = \exp\left[-\sum_i \frac{(x_i-x_i')^2}{4C}\right] & \text{PDE}
\end{array}
$$

Contents

Summary

- We can do multi-task learning or transfer learning with GPs.
- Different ways to build meaningful cross-covariance functions.
- Once defined, we can do all the things we know to do with a single-output GP.
- Cokriging is just prediction with GPs (with a quadratic loss function).
- Several extensions of LMC and PCs.
- Current research: spectral representations for the joint covariance function.

References I

Mauricio A. Álvarez, David Luengo, and Neil D. Lawrence. Latent Force Models. In David van Dyk and Max Welling, editors, *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, pages 9–16, Clearwater Beach, Florida, 16-18 April 2009. JMLR W&CP 5.

Mauricio A. Álvarez, Lorenzo Rosasco, and Neil D. Lawrence. Kernels for vector-valued functions: a review. *Foundations and Trends® in Machine Learning*, 4(3):195–266, 2012.

Edwin V. Bonilla, Kian Ming Chai, and Christopher K. I. Williams. Multi-task Gaussian process prediction. In John C. Platt, Daphne Koller, Yoram Singer, and Sam Roweis, editors, *NIPS*, volume 20, Cambridge, MA, 2008. MIT Press.

Phillip Boyle and Marcus Frean. Dependent Gaussian processes. In Lawrence Saul, Yair Weiss, and Léon Bouttou, editors, *NIPS*, volume 17, pages 217–224, Cambridge, MA, 2005. MIT Press.

Catherine A. Calder and Noel Cressie. Some topics in convolution-based spatial modeling. In *Proceedings of the 56th Session of the International Statistics Institute*, August 2007.

Alan E. Gelfand, Alexandra M. Schmidt, Sudipto Banerjee, and C.F. Sirmans. Nonstationary multivariate process modeling through spatially varying coregionalization. *TEST*, 13(2):263–312, 2004.

Pierre Goovaerts. *Geostatistics For Natural Resources Evaluation*. Oxford University Press, USA, 1997.

J.A. Vargas Guzmán, A.W. Warrick, and D.E. Myers. Coregionalization by linear combination of nonorthogonal components. *Mathematical Geology*, 34 (4):405–419, 2002.

David M. Higdon. Space and space-time modelling using process convolutions. In C. Anderson, V. Barnett, P. Chatwin, and A. El-Shaarawi, editors, *Quantitative methods for current environmental issues*, pages 37–56. Springer-Verlag, 2002.

Andre G. Journel and Charles J. Huijbregts. *Mining Geostatistics*. Academic Press, London, 1978. ISBN 0-12391-050-1.

Yee Whye Teh, Matthias Seeger, and Michael I. Jordan. Semiparametric latent factor models. In Robert G. Cowell and Zoubin Ghahramani, editors, *AISTATS 10*, pages 333–340, Barbados, 6-8 January 2005. Society for Artificial Intelligence and Statistics.

Hans Wackernagel. *Multivariate Geostatistics*. Springer-Verlag Heidelberg New york, 2003.

Andrew Gordon Wilson, David A. Knowles, and Zoubin Ghahramani. Gaussian process regression networks. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, pages 1139–1146, 2012.