PyMCon Web Series

An introduction to Multi-output Gaussian processes using PyMC

Danh Phan

Feb 2023

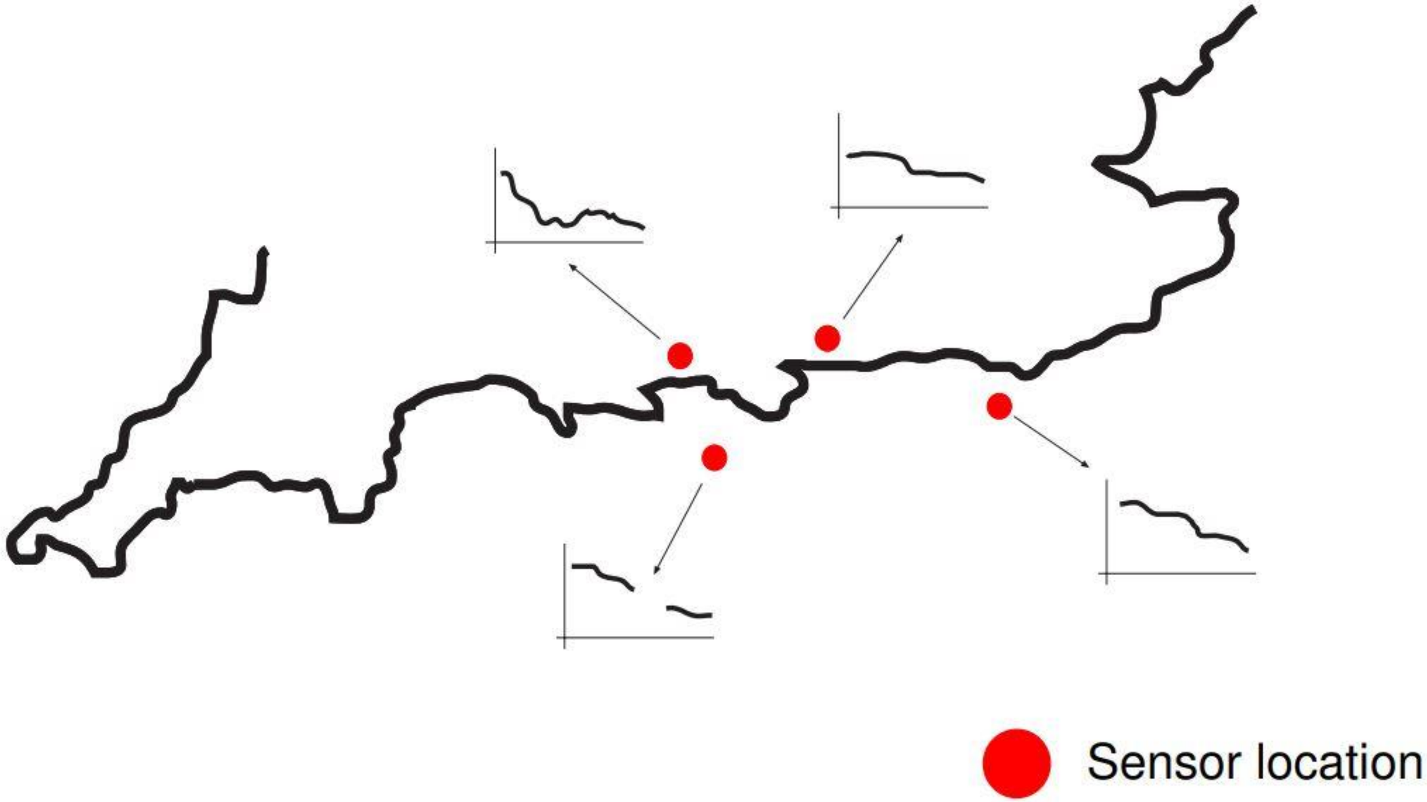
1. Why multi-output Gaussian Processes?

There are many cases where several outputs are affected by the same uncertainty.

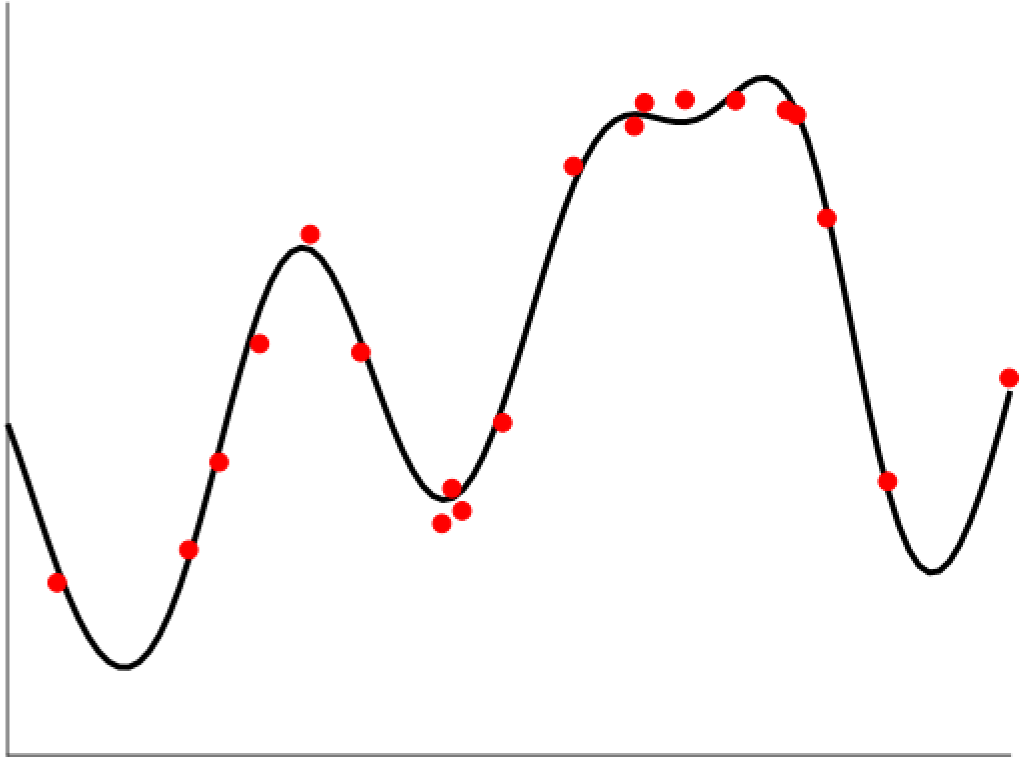
❖ For examples:

- The temperatures or wind speed in nearby locations
- Stock prices of tech companies
- The currency exchange with respect to USD dollars
- The EEG recordings from human neonates on human brains
- The spatial variability of over one risk factor for cancer across a geographical map.
- the continuous-space multi-crime dataset
- Spatial prediction of soil pollutants

The Sensor Network from South Coast of England



Single-output Gaussian Process



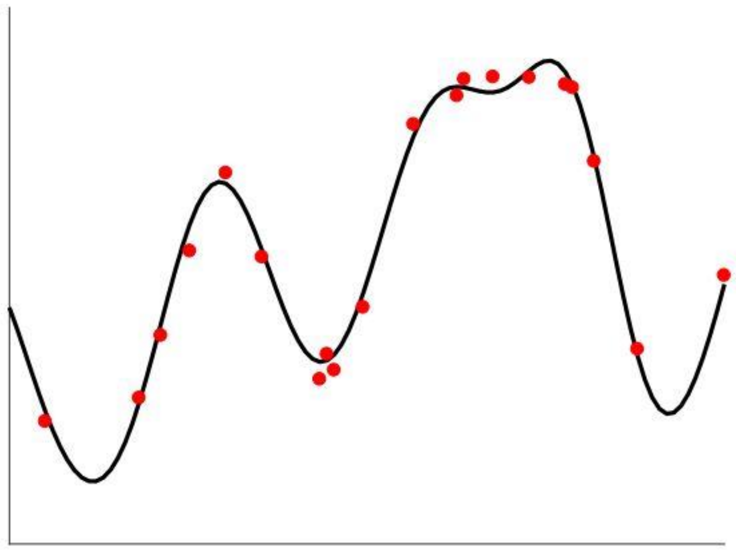
$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$

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

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

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

Single-output Gaussian Process



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

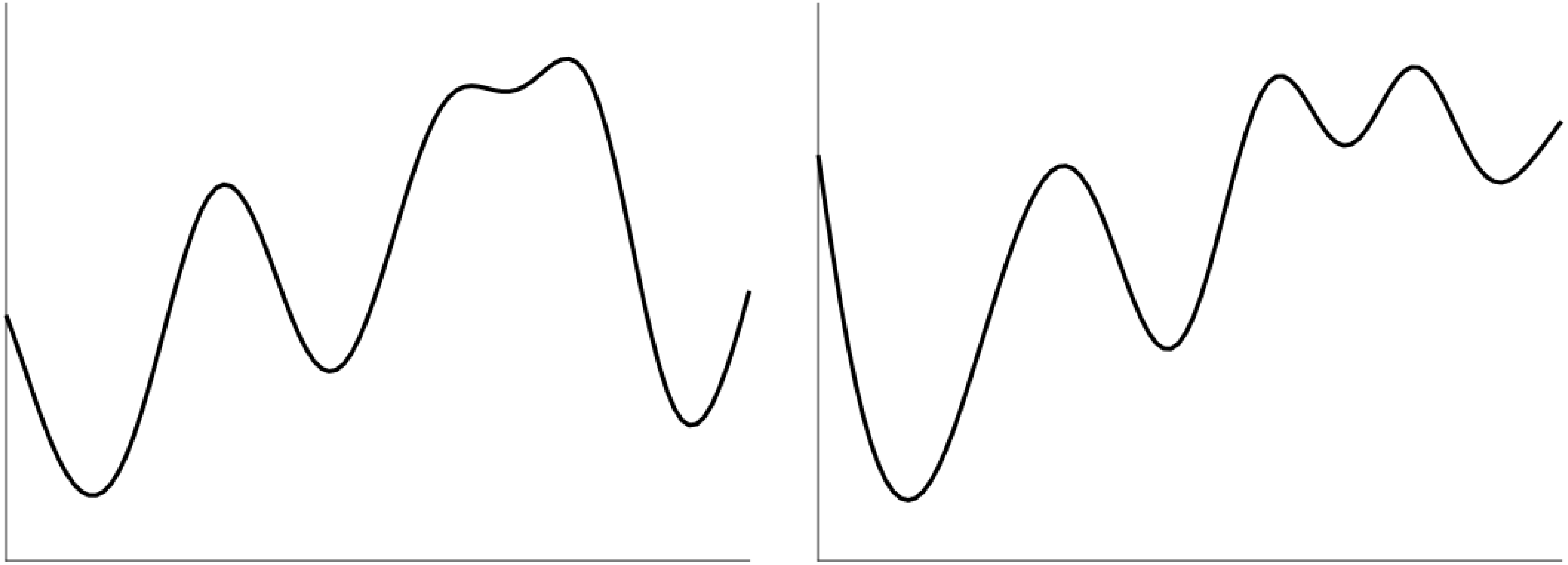
$$y(\mathbf{x}_i) = f(\mathbf{x}_i) + \epsilon_i$$

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

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

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

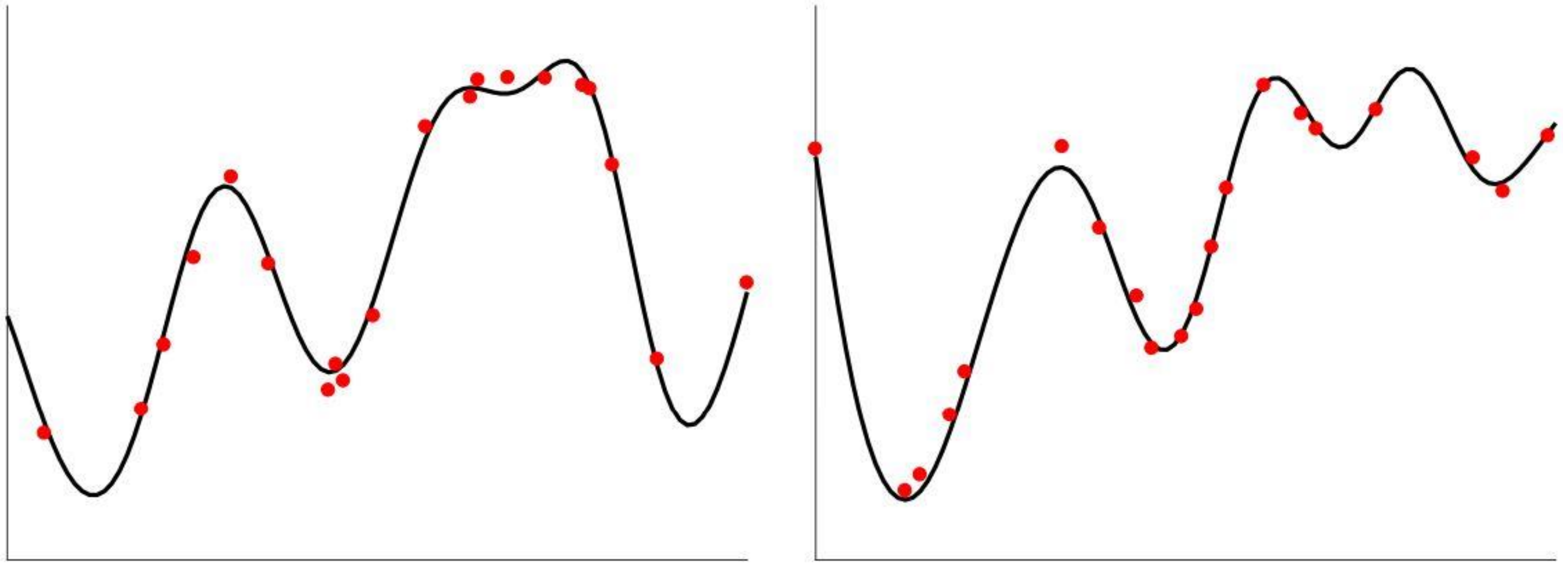
Multiple-output Gaussian process



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

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

Multiple-output Gaussian process



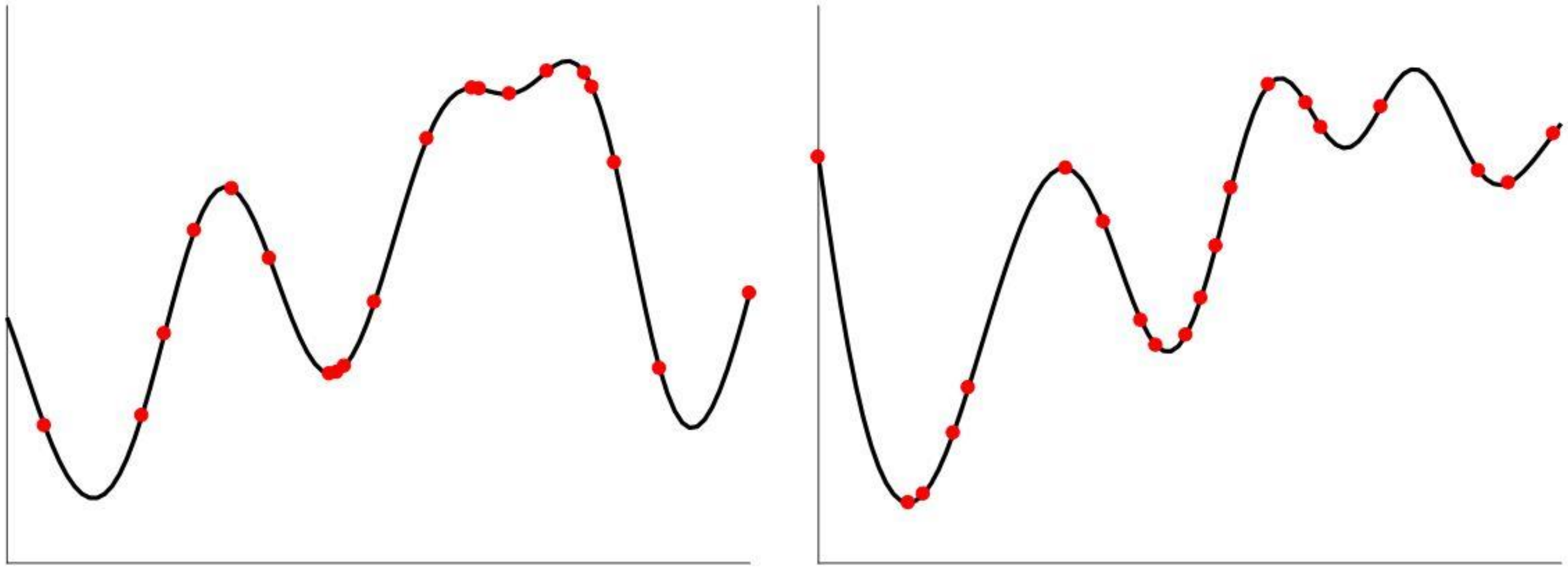
$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}'))$ $\quad$ $f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}'))$

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

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

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

Multiple-output Gaussian process



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

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

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

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.

2. Intrinsic coregionalization model (ICM)

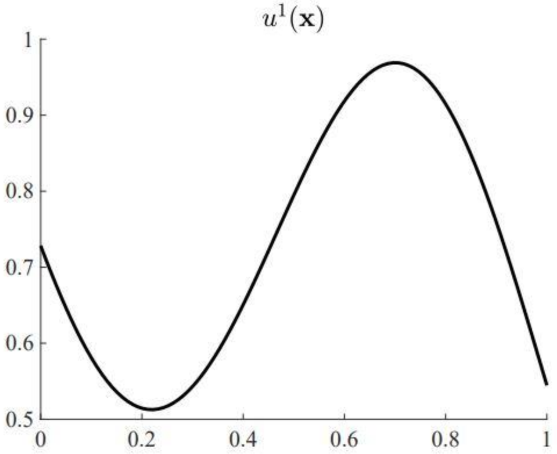
Two outputs and one latent sample

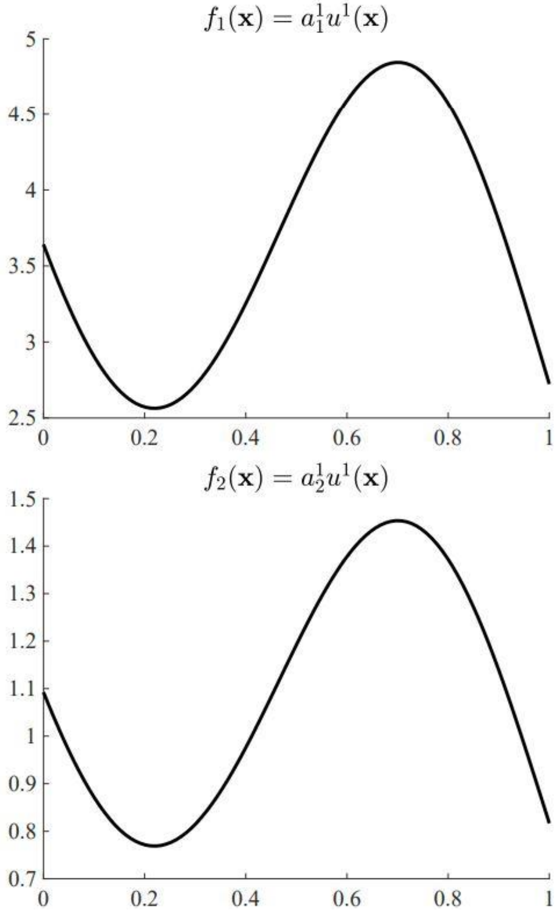
- Consider two outputs $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^p$.

- 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


$u^1(\mathbf{x})$
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
$f_2(\mathbf{x}) = a_2^1 u^1(\mathbf{x})$
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.2
0.4
0.6
0.8
1

ICM: covariance

- 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

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

ICM: covariance

- 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

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

ICM: two outputs and two latent samples

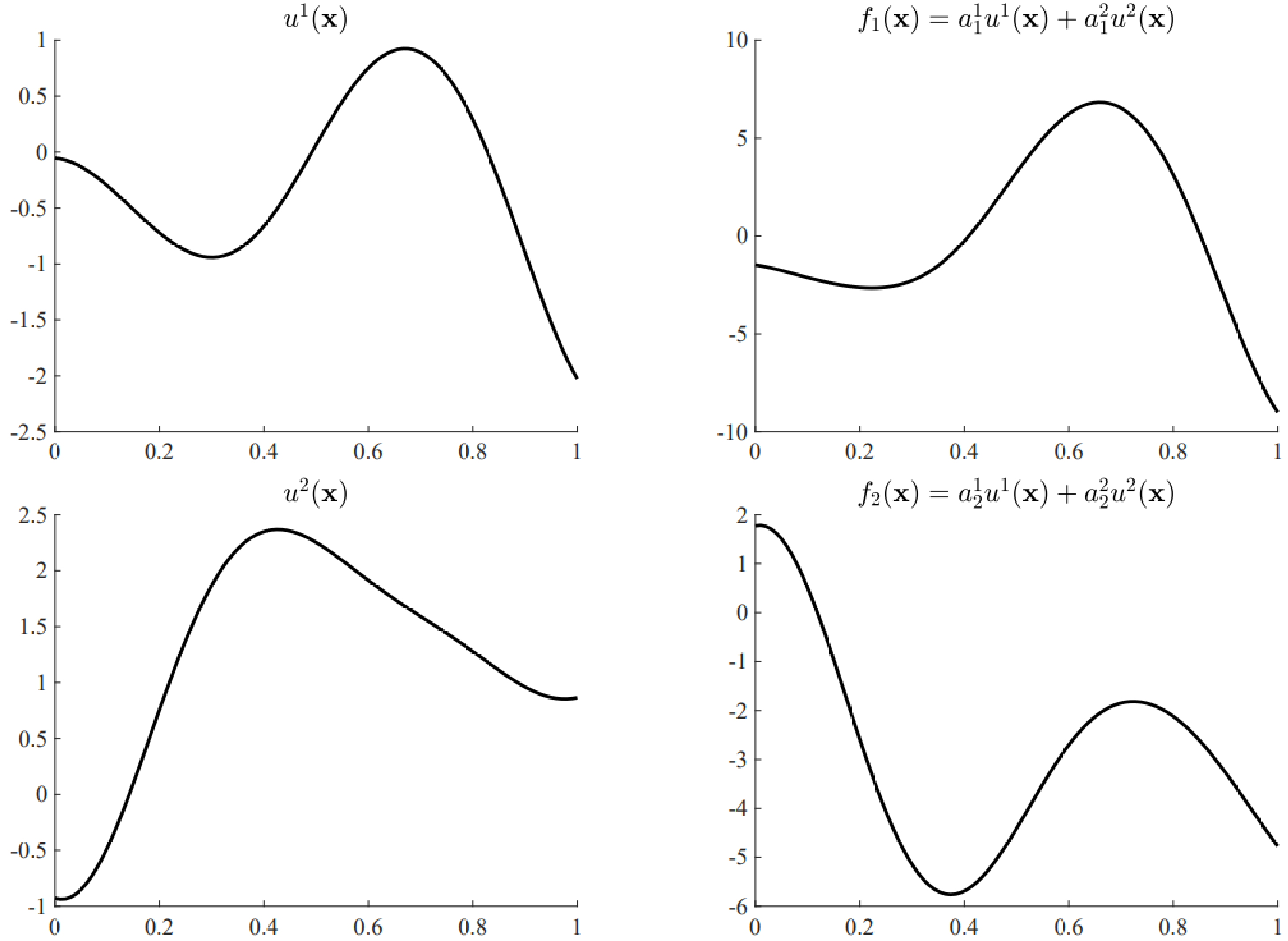
- We can introduce a bit more of complexity in the model before as follows.

- 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 $u^2(\mathbf{x})$

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

ICM: samples



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

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

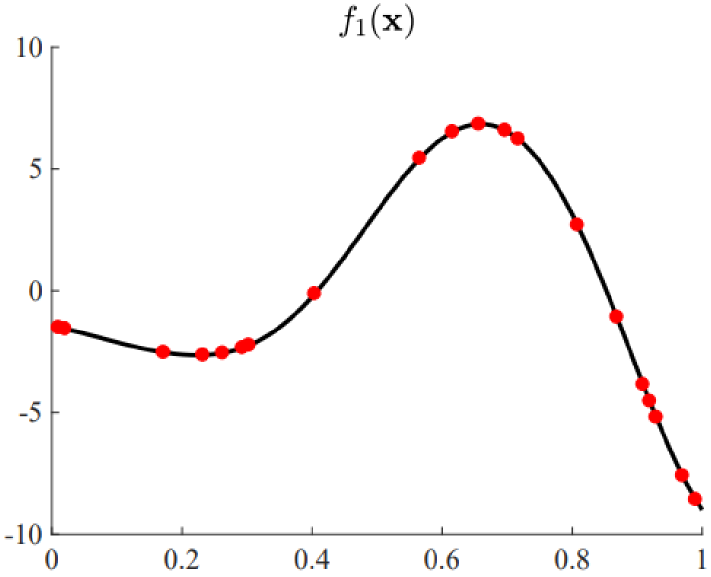
- The covariance for $\mathbf{f}(\mathbf{x})$ is computed as

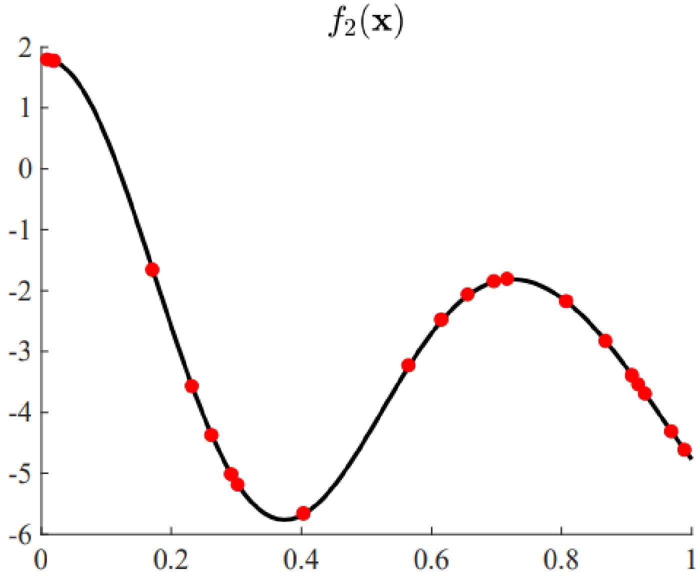
$$\begin{aligned}
\operatorname{cov}\left(\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')\right) &= \mathbf{a}^1(\mathbf{a}^1)^\top \operatorname{cov}(u^1(\mathbf{x}), u^1(\mathbf{x}')) + \mathbf{a}^2(\mathbf{a}^2)^\top \operatorname{cov}(u^2(\mathbf{x}), u^2(\mathbf{x}')) \\
&= \mathbf{a}^1(\mathbf{a}^1)^\top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

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

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

ICM: covariance





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

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

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

The Kronecker product between matrices

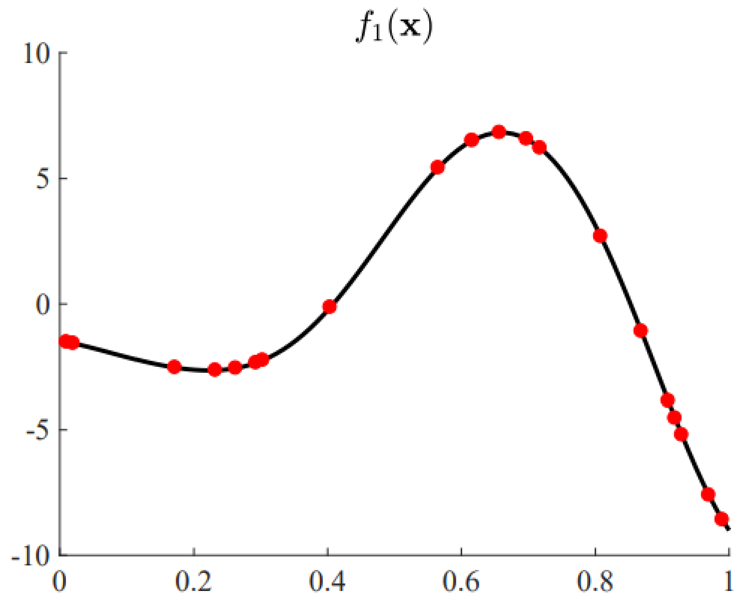
- If **A** is an $m \times n$ matrix and **B** is a $p \times q$ matrix, then the Kronecker product **A** ⊗ **B** is the $pm \times qn$ block matrix

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}$$

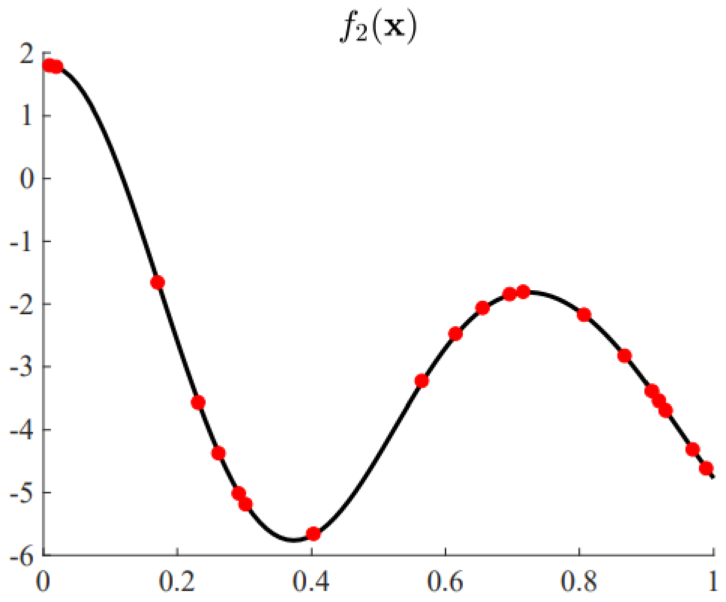
- Example

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \otimes \begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} = \begin{bmatrix} 1\begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} & 2\begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} \\ 3\begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} & 4\begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 1\times 0 & 1\times 5 & 2\times 0 & 2\times 5 \\ 1\times 6 & 1\times 7 & 2\times 6 & 2\times 7 \\ 3\times 0 & 3\times 5 & 4\times 0 & 4\times 5 \\ 3\times 6 & 3\times 7 & 4\times 6 & 4\times 7 \end{bmatrix} = \begin{bmatrix} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ 0 & 15 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{bmatrix}$$

ICM: covariance



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



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

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

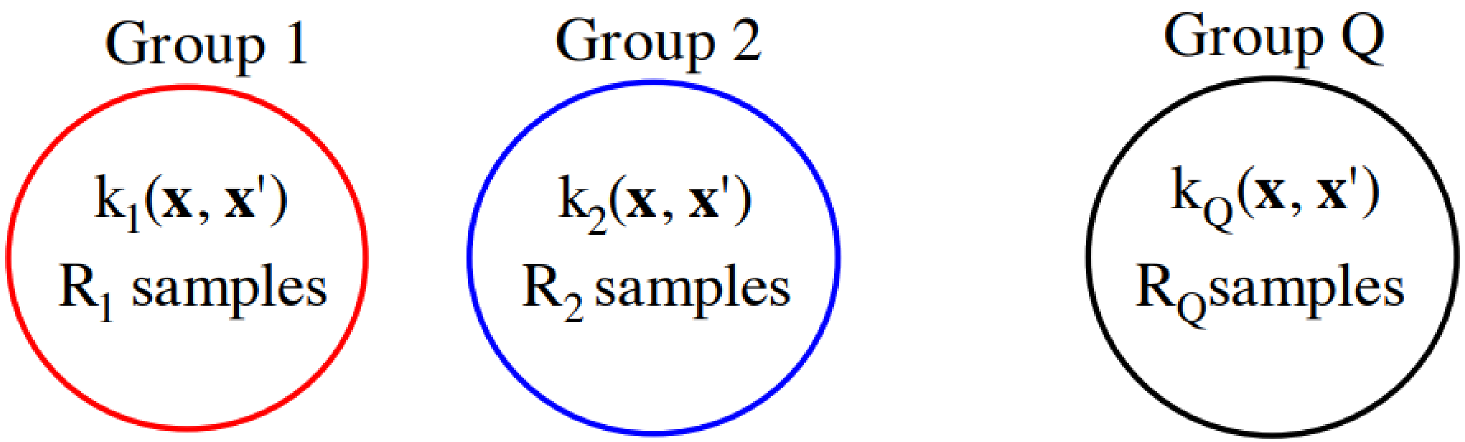
The matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$ has elements $k(\mathbf{x}_i, \mathbf{x}_j)$.

3. Linear model of coregionalization (LMC)

- The LMC corresponds to the sum of Q ICMs.

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

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



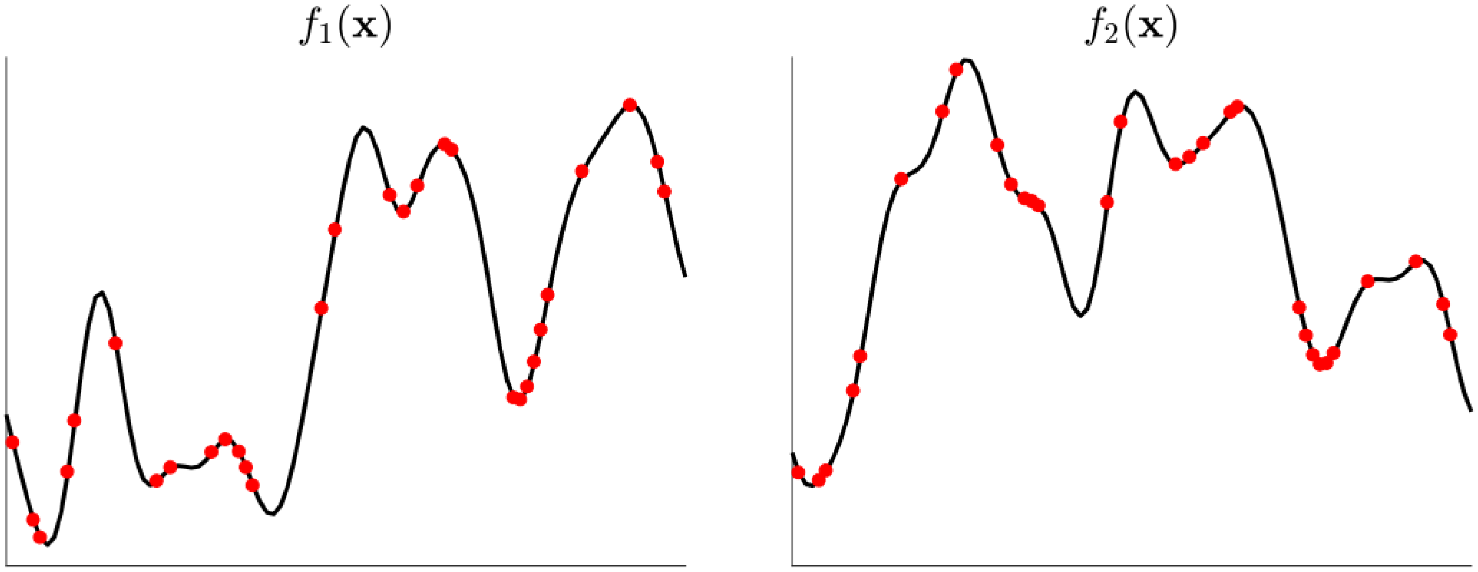
LCM: covariance

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

LCM: covariance



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

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

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_{ij}^q.

References

- This slides are a short version of [“Multiple-output Gaussian processes” presentation] from Mauricio A. Alvarez
- Kronecker product: https://en.wikipedia.org/wiki/Kronecker_product