Multi Output Gaussian Processes

Rajbir Singh Nirwan (Raj)

June 01, 2021

Outline

- Single Output Gaussian Processes
- Multi Output Gaussian Processes
- Scalable Exact Inference in Multi Output GPs
  - Theory (short)
  - Experiments

Single Output Gaussian Processes

Any finite collection of function values at $x_1, x_2, \ldots, x_N$ is jointly Gaussian distributed

$$p\left(f\left(x_1\right), f\left(x_2\right), \ldots, f\left(x_N\right)\right) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K}) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{pmatrix}\right)$$

$$k_{ij} = k(x_i, x_j)$$

Single Output Gaussian Processes

Any finite collection of function values at $x_1, x_2, \ldots, x_N$ is jointly Gaussian distributed

$$p\left(f\left(x_1\right), f\left(x_2\right), \ldots, f\left(x_N\right)\right) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K}) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{pmatrix}\right)$$

$$k_{ij} = k(x_i, x_j)$$

Common Kernel Functions

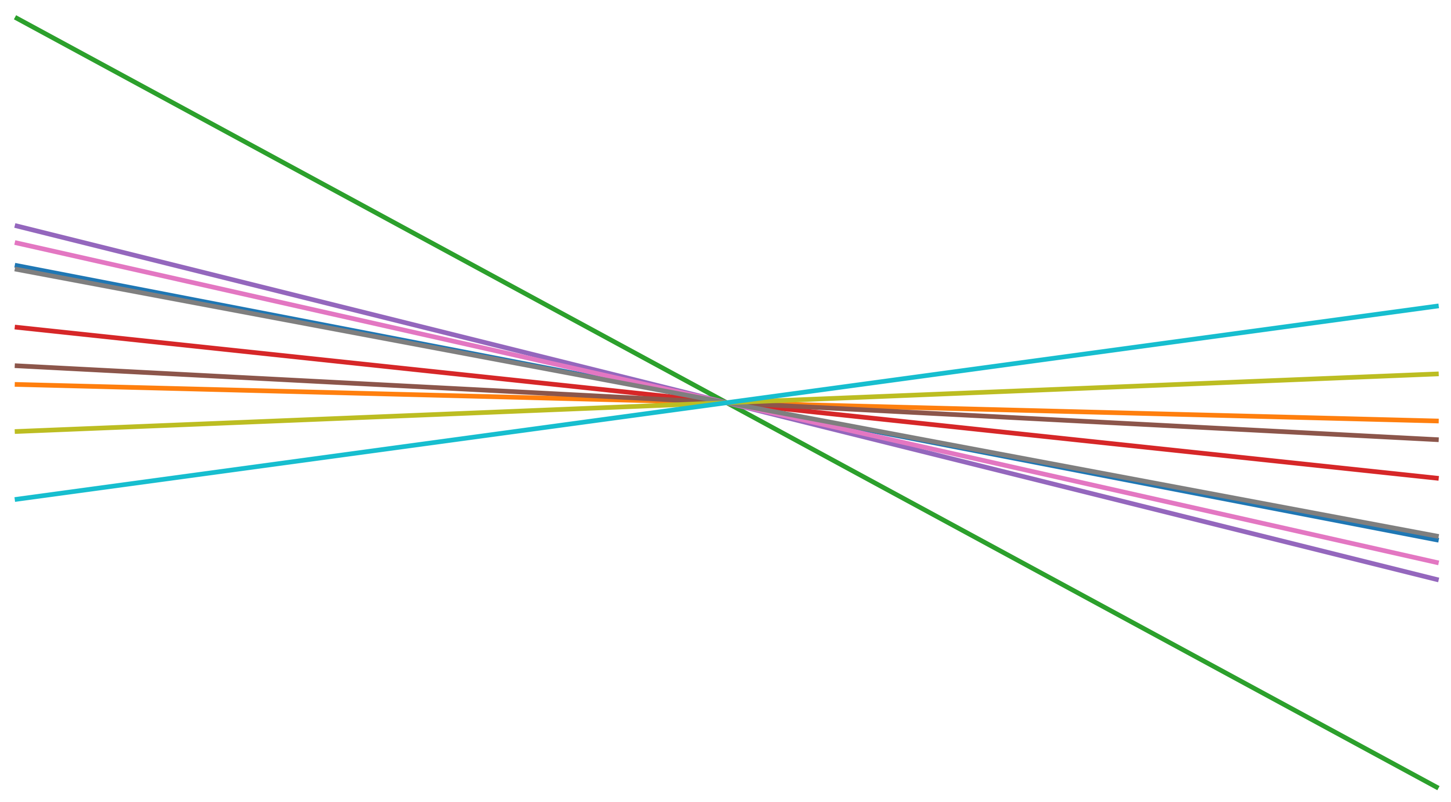
$$k_{linear}(x, x') = xx'$$

$$k_{rbf}(x, x') = \exp\left(-\frac{1}{2\ell^2}(x - x')^2\right)$$

$$k_{ou}(x, x') = \exp\left(-\frac{1}{\ell}|x - x'|\right)$$

$$k_{mat32}(x, x') = \left(1 + \frac{\sqrt{3}|x - x'|}{\ell}\right)\exp\left(-\frac{\sqrt{3}|x - x'|}{\ell}\right)$$

$$k_{periodic}(x, x') = \exp\left(-\frac{2}{\ell^2}\sin^2(|x - x'|)\right)$$



Single Output Gaussian Processes

Any finite collection of function values at $x_1, x_2, \ldots, x_N$ is jointly Gaussian distributed

$$p\left(f\left(x_1\right), f\left(x_2\right), \ldots, f\left(x_N\right)\right) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K}) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{pmatrix}\right) \qquad k_{ij} = k(x_i, x_j)$$

Common Kernel Functions

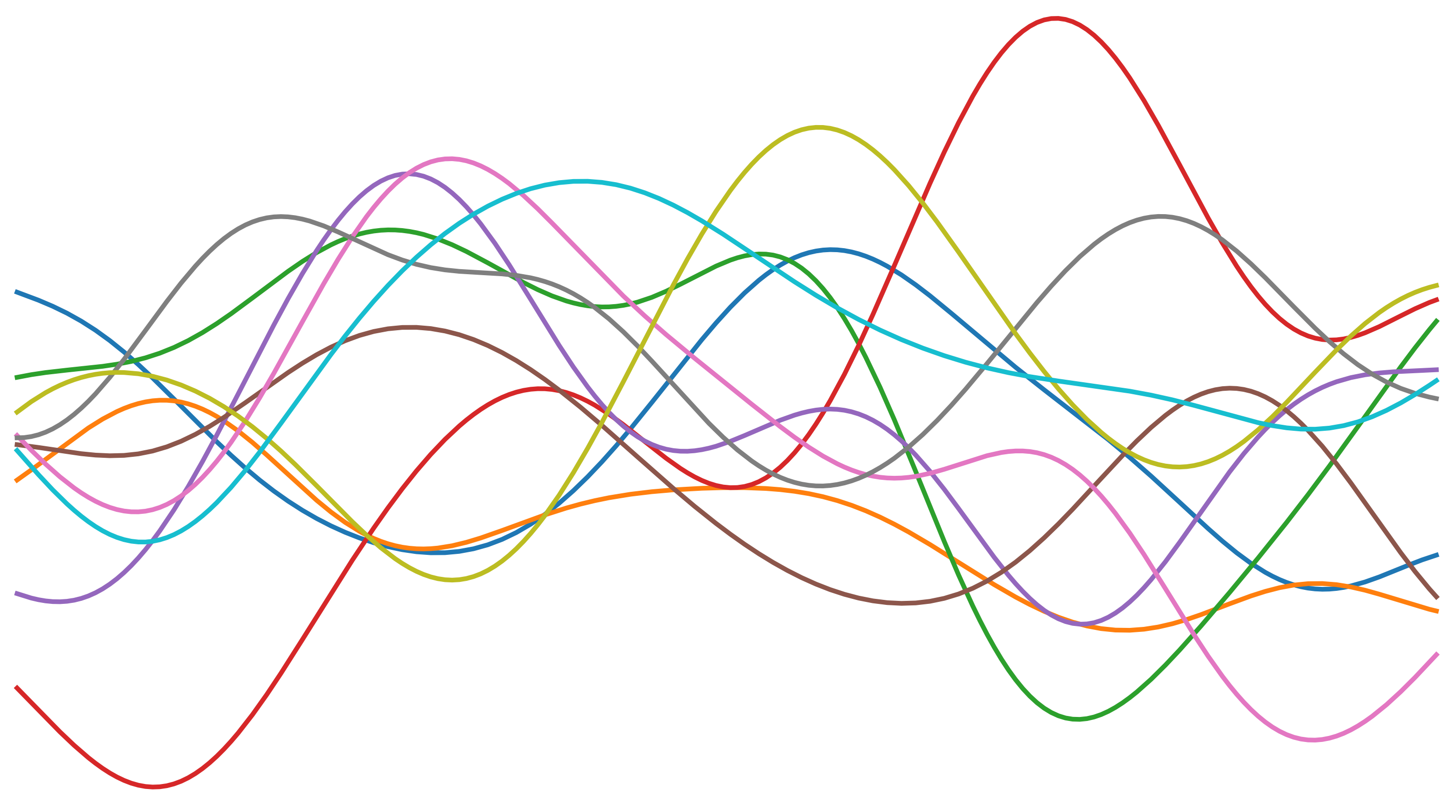
$$k_{linear}(x, x') = xx'$$

$$k_{rbf}(x, x') = \exp\left(-\frac{1}{2\ell^2}(x - x')^2\right)$$

$$k_{ou}(x, x') = \exp\left(-\frac{1}{\ell}|x - x'|\right)$$

$$k_{mat32}(x, x') = \left(1 + \frac{\sqrt{3}\,|x - x'|}{\ell}\right)\exp\left(-\frac{\sqrt{3}\,|x - x'|}{\ell}\right)$$

$$k_{periodic}(x, x') = \exp\left(-\frac{2}{\ell^2}\sin^2(|x - x'|)\right)$$



Single Output Gaussian Processes

Any finite collection of function values at $x_1, x_2, \ldots, x_N$ is jointly Gaussian distributed

$$p\left(f(x_1), f(x_2), \ldots, f(x_N)\right) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K}) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{pmatrix}\right)$$

$$k_{ij} = k(x_i, x_j)$$

Common Kernel Functions

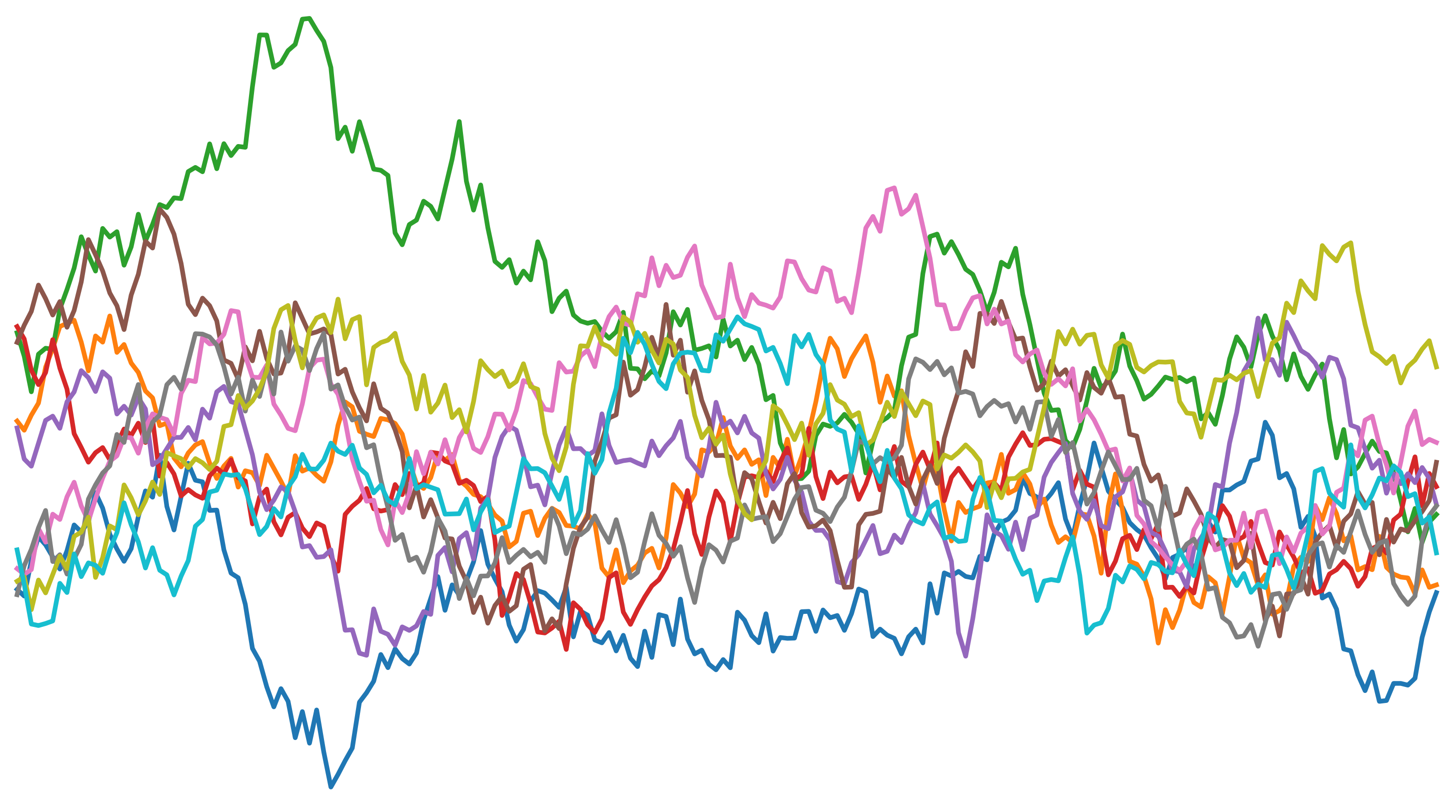
$$k_{linear}(x, x') = xx'$$

$$k_{rbf}(x, x') = \exp\left(-\frac{1}{2\ell^2}(x - x')^2\right)$$

$$k_{ou}(x, x') = \exp\left(-\frac{1}{\ell}|x - x'|\right)$$

$$k_{mat32}(x, x') = \left(1 + \frac{\sqrt{3}|x - x'|}{\ell}\right)\exp\left(-\frac{\sqrt{3}|x - x'|}{\ell}\right)$$

$$k_{periodic}(x, x') = \exp\left(-\frac{2}{\ell^2}\sin^2(|x - x'|)\right)$$



Single Output Gaussian Processes

Any finite collection of function values at $x_1, x_2, \ldots, x_N$ is jointly Gaussian distributed

$$p\left(f\left(x_1\right), f\left(x_2\right), \ldots, f\left(x_N\right)\right) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K}) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{pmatrix}\right)$$

$$k_{ij} = k(x_i, x_j)$$

Common Kernel Functions

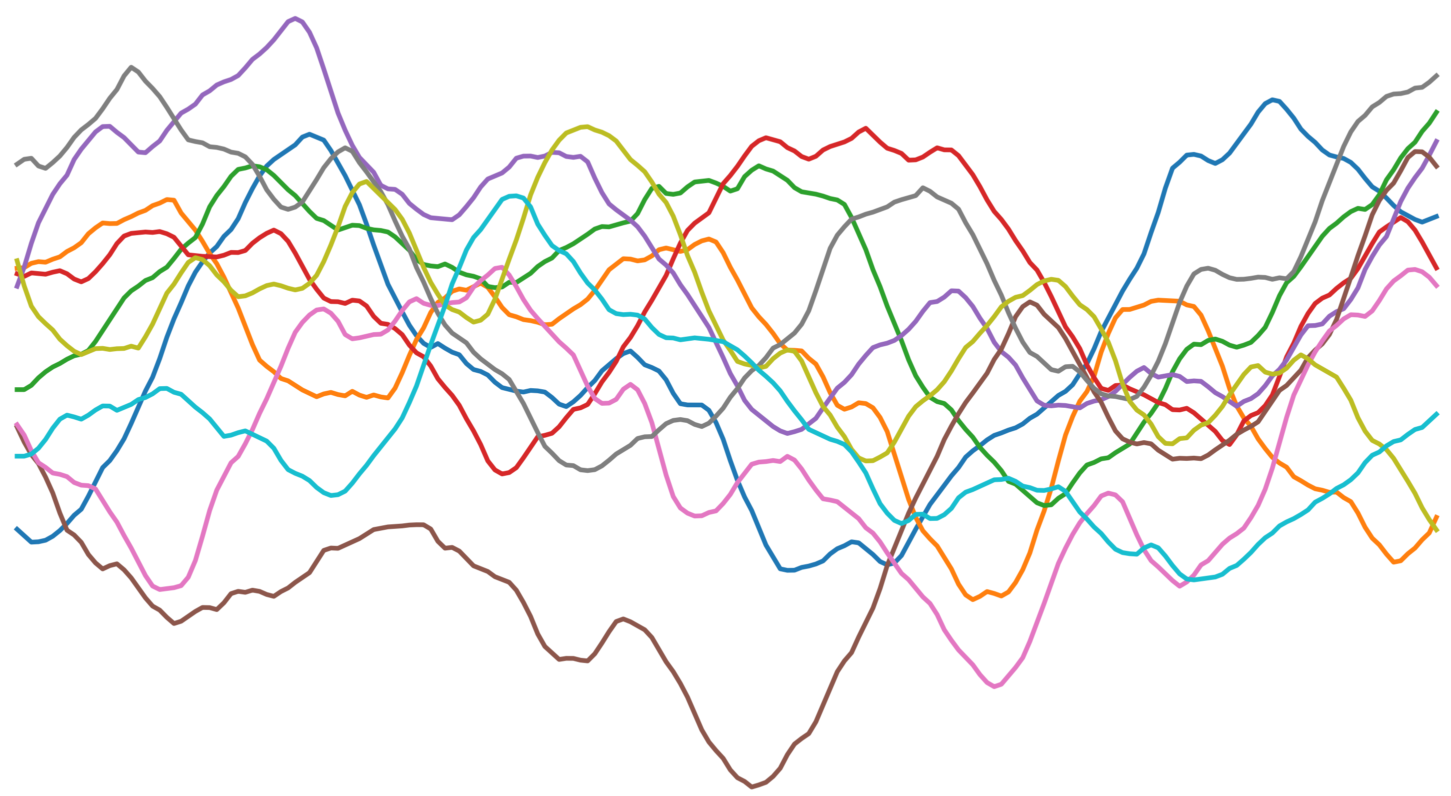
$$k_{linear}(x, x') = xx'$$

$$k_{rbf}(x, x') = \exp\left(-\frac{1}{2\ell^2}(x - x')^2\right)$$

$$k_{ou}(x, x') = \exp\left(-\frac{1}{\ell}|x - x'|\right)$$

$$k_{mat32}(x, x') = \left(1 + \frac{\sqrt{3}\,|x - x'|}{\ell}\right) \exp\left(-\frac{\sqrt{3}\,|x - x'|}{\ell}\right)$$

$$k_{periodic}(x, x') = \exp\left(-\frac{2}{\ell^2}\sin^2(|x - x'|)\right)$$



Single Output Gaussian Processes

Any finite collection of function values at $x_1, x_2, \ldots, x_N$ is jointly Gaussian distributed

$$p\left(f\left(x_1\right), f\left(x_2\right), \ldots, f\left(x_N\right)\right) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K}) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{pmatrix}\right) \qquad k_{ij} = k(x_i, x_j)$$

Common Kernel Functions

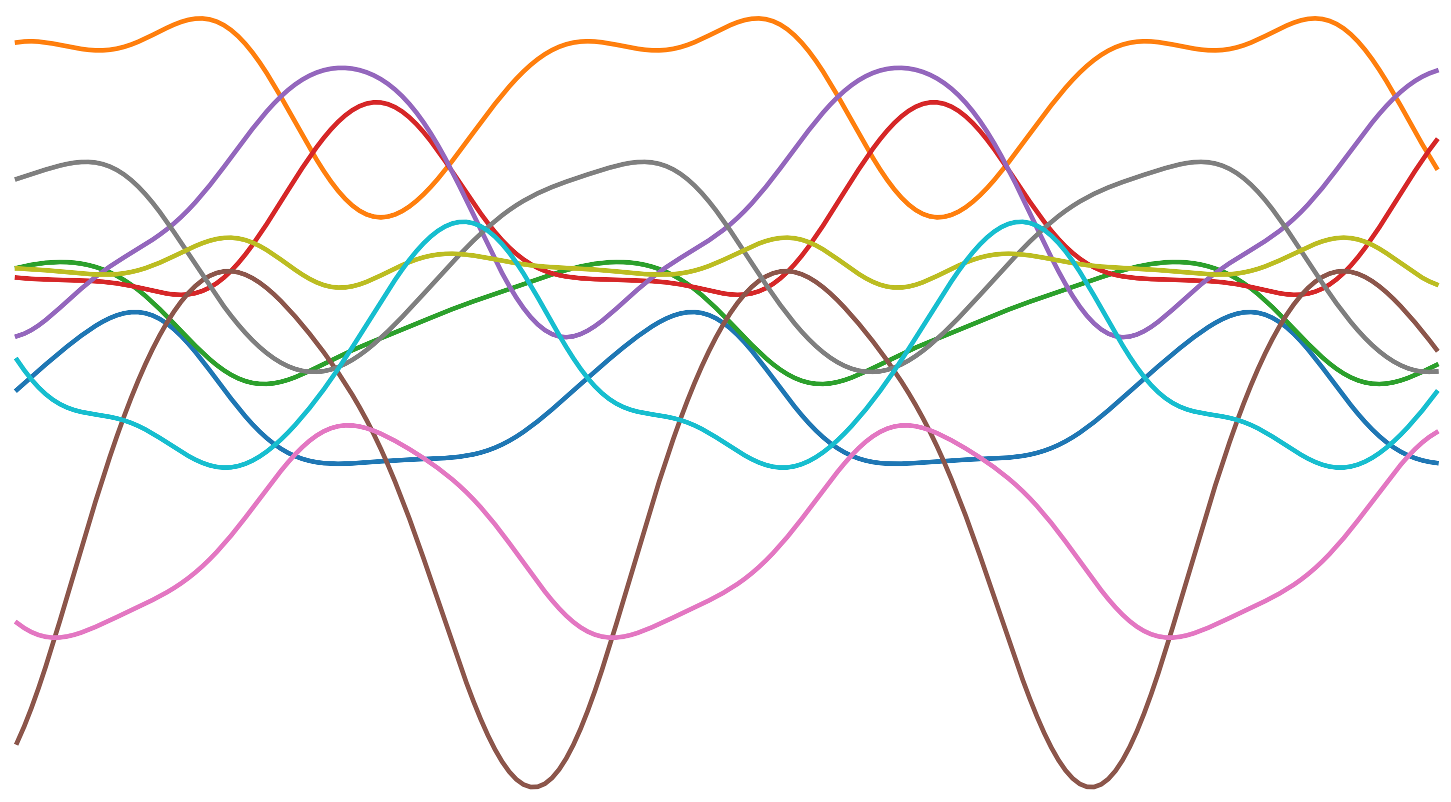
$$k_{linear}(x, x') = xx'$$

$$k_{rbf}(x, x') = \exp\left(-\frac{1}{2\ell^2}(x - x')^2\right)$$

$$k_{ou}(x, x') = \exp\left(-\frac{1}{\ell}|x - x'|\right)$$

$$k_{mat32}(x, x') = \left(1 + \frac{\sqrt{3}\,|x - x'|}{\ell}\right)\exp\left(-\frac{\sqrt{3}\,|x - x'|}{\ell}\right)$$

$$k_{periodic}(x, x') = \exp\left(-\frac{2}{\ell^2}\sin^2(|x - x'|)\right)$$



Learning (Inference)

Bayes Theorem

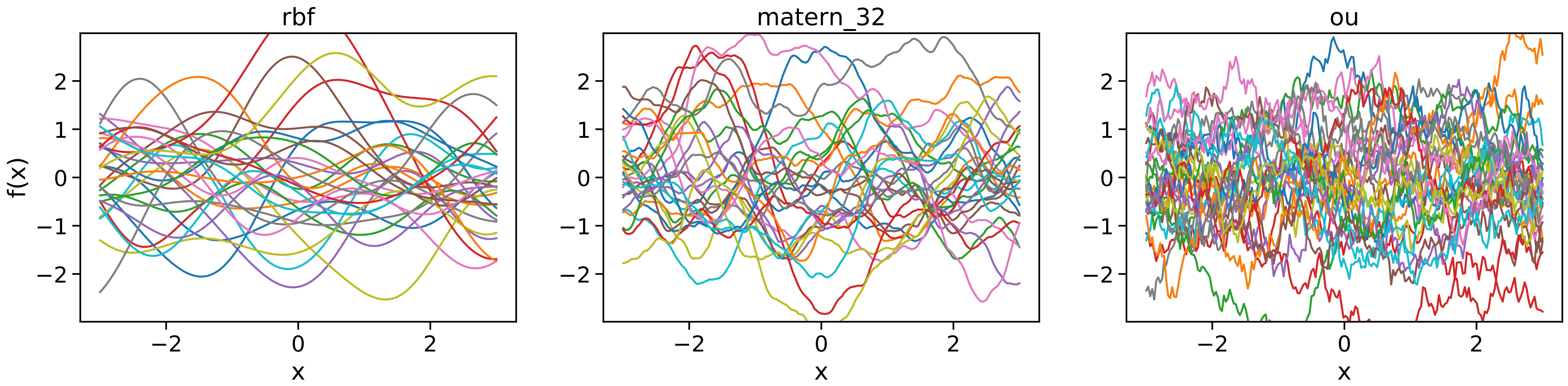
$$p(f \mid y) = \frac{p(y \mid f)p(f)}{p(y)}$$

Learning (Inference)

Bayes Theorem

$$p(f\,|\,y) = \frac{p(y\,|\,f)p(f)}{p(y)}$$

Prior

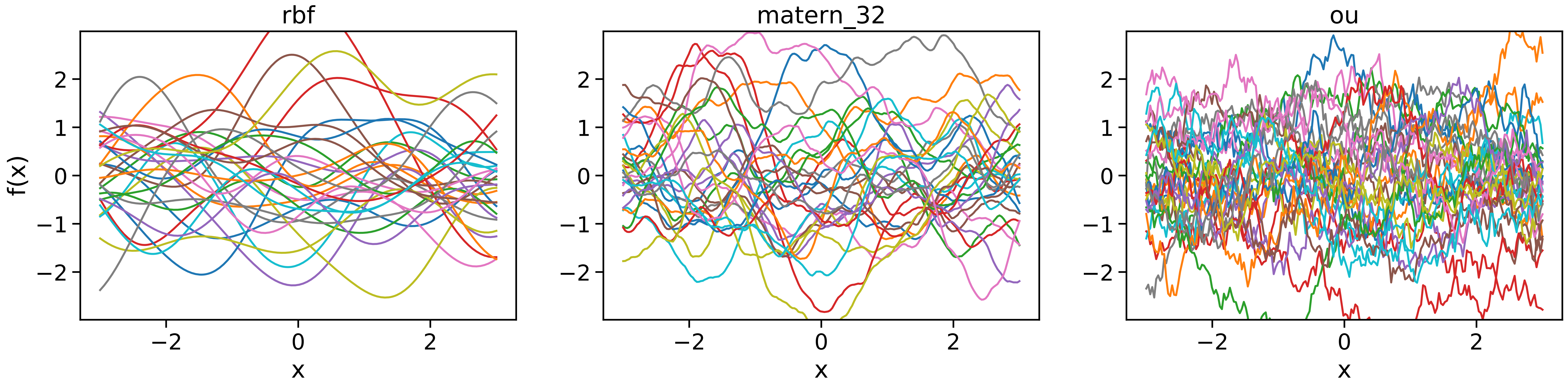


Learning (Inference)

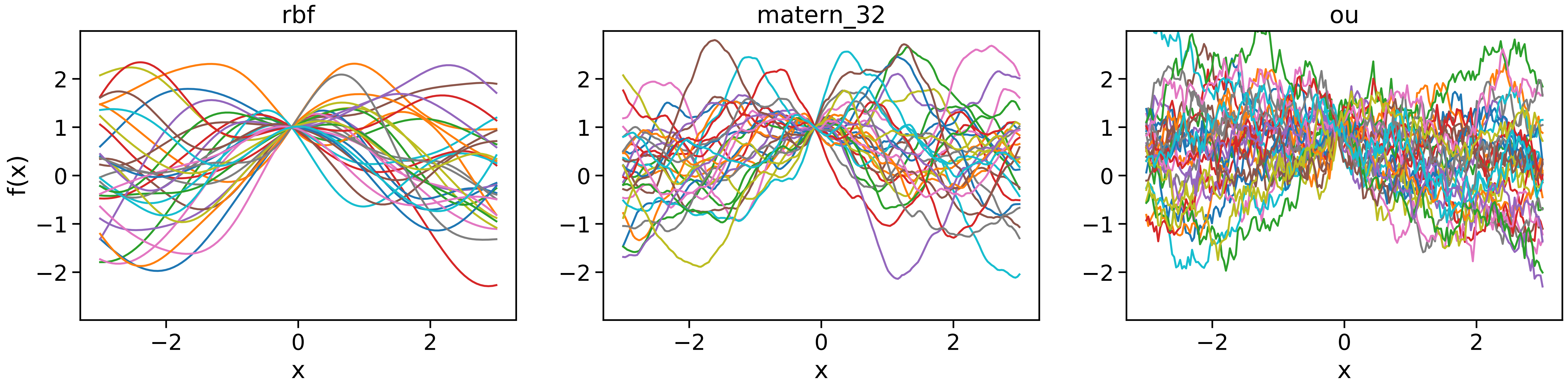
Bayes Theorem

$$p(f\,|\,y) = \frac{p(y\,|\,f)p(f)}{p(y)}$$

Prior



Posterior



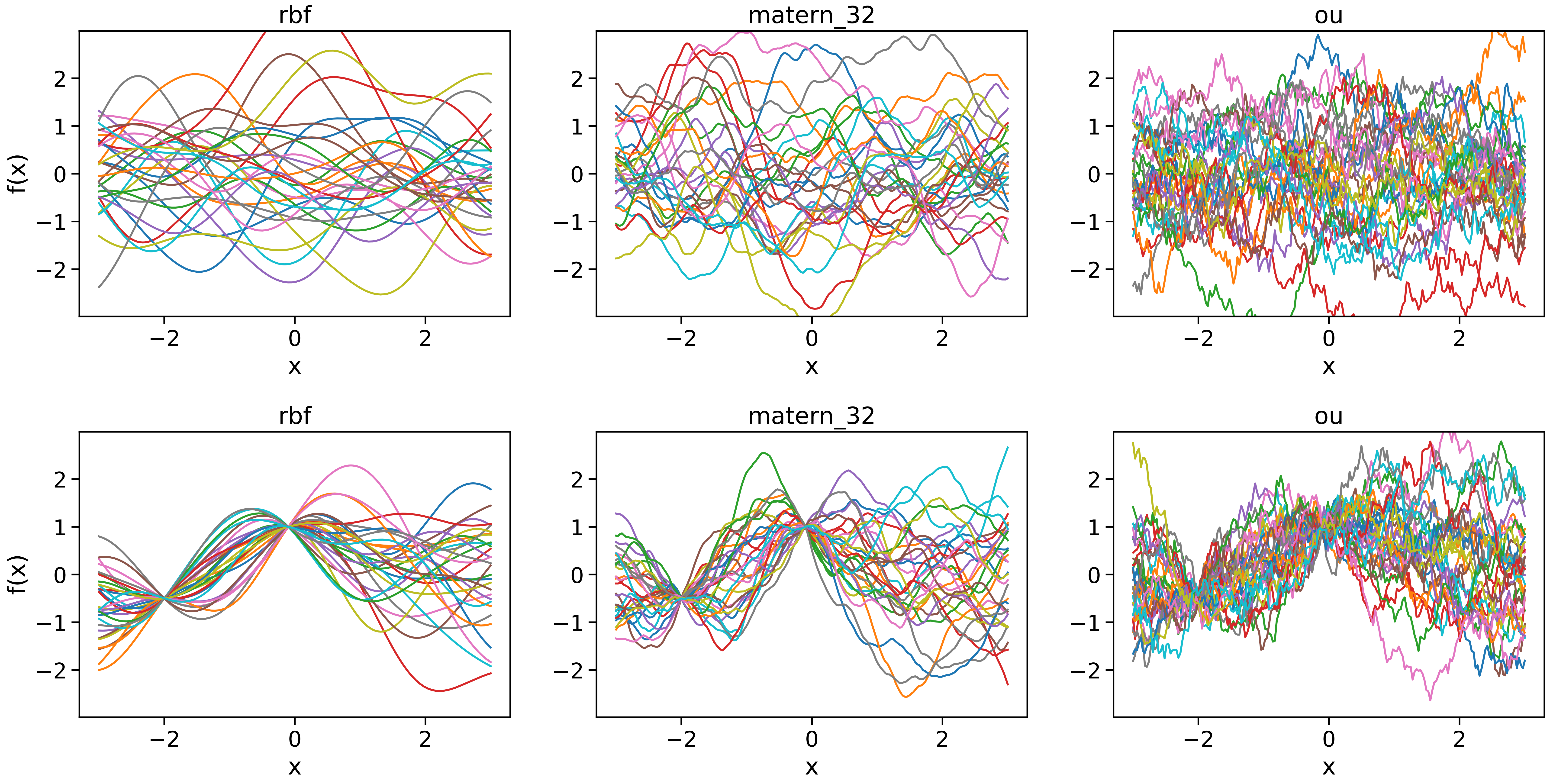
Learning (Inference)

Bayes Theorem

$$p(f\,|\,y) = \frac{p(y\,|\,f)p(f)}{p(y)}$$

Prior

Posterior



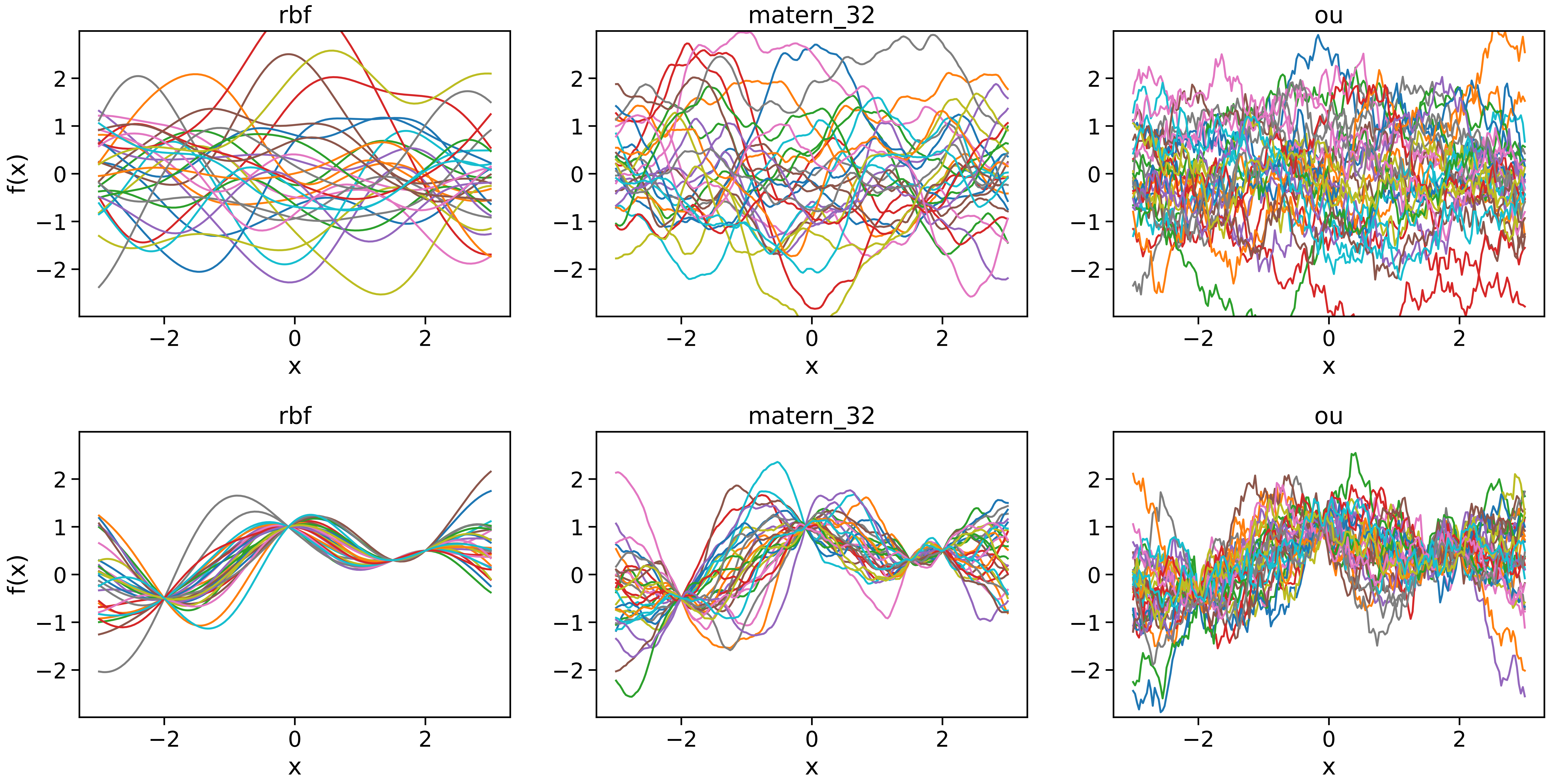
Learning (Inference)

Bayes Theorem

$$p(f\,|\,y) = \frac{p(y\,|\,f)p(f)}{p(y)}$$

Prior

Posterior



Learning (Inference)

Bayes Theorem

$$p(f \mid y) = \frac{p(y \mid f)p(f)}{p(y)}$$

Prior

$$\begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_N) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{pmatrix}\right)$$

Learning (Inference)

Bayes Theorem

$$p(f \mid y) = \frac{p(y \mid f)p(f)}{p(y)}$$

Prior

$$\begin{pmatrix} f\left(x_1\right) \\ f\left(x_2\right) \\ \vdots \\ f\left(x_N\right) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{pmatrix}\right)$$

Likelihood

$$y_i = f(x_i) + \epsilon$$

$$y_i \mid f_i \sim \mathcal{N}(f_i, \sigma^2)$$

$$\boldsymbol{y} \mid \boldsymbol{f} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{K} + \sigma^2 \boldsymbol{I})$$

Learning (Inference)

Bayes Theorem

$$p(f\,|\,y) = \frac{p(y\,|\,f)p(f)}{p(y)}$$

Prior

$$\begin{pmatrix} f\left(x_1\right) \\ f\left(x_2\right) \\ \vdots \\ f\left(x_N\right) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{pmatrix}\right)$$

Posterior

$$\boldsymbol{f}_*\,|\,\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}', \boldsymbol{K}')$$

$$\boldsymbol{\mu}' = \boldsymbol{K}_*(\boldsymbol{K} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{y}$$

$$\boldsymbol{K}' = \boldsymbol{K}_{**} - \boldsymbol{K}_*(\boldsymbol{K} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{K}_*{}^T$$

Likelihood

$$y_i = f(x_i) + \epsilon$$

$$y_i\,|\,f_i \sim \mathcal{N}(f_i, \boldsymbol{\sigma}^2)$$

$$\boldsymbol{y}\,|\,\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K} + \sigma^2\boldsymbol{I})$$

Learning (Inference)

Bayes Theorem

$$p(f\,|\,y) = \frac{p(y\,|\,f)p(f)}{p(y)}$$

Prior

$$\begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_N) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{pmatrix}\right)$$

Posterior

$$\boldsymbol{f}_*\,|\,\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}', \boldsymbol{K}')$$

$$\boldsymbol{\mu}' = \boldsymbol{K}_*(\boldsymbol{K} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{y}$$

$$\boldsymbol{K}' = \boldsymbol{K}_{**} - \boldsymbol{K}_*(\boldsymbol{K} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{K}_*{}^T$$

Likelihood

$$y_i = f(x_i) + \epsilon$$

$$y_i\,|\,f_i \sim \mathcal{N}(f_i, \sigma^2)$$

$$\boldsymbol{y}\,|\,\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K} + \sigma^2\boldsymbol{I})$$

Marginal Likelihood

$$\begin{aligned} p(\boldsymbol{y}) &= \int p(\boldsymbol{y}\,|\,\boldsymbol{f})p(\boldsymbol{f})d\boldsymbol{f} \\ &= -\frac{1}{2}\boldsymbol{y}^T(\boldsymbol{K} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{y} \\ &\quad -\frac{1}{2}\log|\boldsymbol{K} + \sigma^2\boldsymbol{I}| - \frac{N}{2}\log 2\pi \end{aligned}$$

Multi Output Gaussian Processes

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

Multi Output Gaussian Processes

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{N1} \\ y_{N2} \\ \vdots \\ y_{NT} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K})$$

$$\boldsymbol{\mu} \in \mathbb{R}^{NT}$$

$$\boldsymbol{K} \in \mathbb{R}^{NT \times NT}$$

Inference Time $\sim \mathcal{O}((NT)^3)$

Multi Output Gaussian Processes

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{N1} \\ y_{N2} \\ \vdots \\ y_{NT} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K})$$

$$\boldsymbol{\mu} \in \mathbb{R}^{NT}$$

$$\boldsymbol{K} \in \mathbb{R}^{NT \times NT}$$

Inference Time $\sim \mathcal{O}((NT)^3)$

Exploit Structure using Kronecker Products

$$\boldsymbol{A} \otimes \boldsymbol{B} \equiv \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} \end{pmatrix}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$$

Multi Output Gaussian Processes

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{N1} \\ y_{N2} \\ \vdots \\ y_{NT} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K})$$

$$\boldsymbol{\mu} \in \mathbb{R}^{NT}$$
$$\boldsymbol{K} \in \mathbb{R}^{NT \times NT}$$

Inference Time $\sim \mathcal{O}((NT)^3)$

Exploit Structure using Kronecker Products

$$\boldsymbol{A} \otimes \boldsymbol{B} \equiv \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} \end{pmatrix}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$$

rowVec$(\boldsymbol{Y}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}_{NN} \otimes \boldsymbol{K}_{TT})$

colVec$(\boldsymbol{Y}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}_{TT} \otimes \boldsymbol{K}_{NN})$

Inference Time $\sim \mathcal{O}(N^3 + T^3)$

Multi Output Gaussian Processes

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{N1} \\ y_{N2} \\ \vdots \\ y_{NT} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K})$$

$$\boldsymbol{\mu} \in \mathbb{R}^{NT}$$
$$\boldsymbol{K} \in \mathbb{R}^{NT \times NT}$$

Inference Time $\sim \mathcal{O}((NT)^3)$

Exploit Structure using Kronecker Products

$$\boldsymbol{A} \otimes \boldsymbol{B} \equiv \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} \end{pmatrix}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$$

rowVec$(\boldsymbol{Y}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}_{NN} \otimes \boldsymbol{K}_{TT})$
colVec$(\boldsymbol{Y}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}_{TT} \otimes \boldsymbol{K}_{NN})$

Inference Time $\sim \mathcal{O}(N^3 + T^3)$

N independent Processes

rowVec$(\boldsymbol{Y}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{NN} \otimes \boldsymbol{K}_{TT}) = \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{Y}_{n,:} \,|\, \boldsymbol{0}, \boldsymbol{K}_{TT})$

Inference Time $\sim \mathcal{O}(T^3)$

Multi Output Gaussian Processes

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{N1} \\ y_{N2} \\ \vdots \\ y_{NT} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K})$$

$$\boldsymbol{\mu} \in \mathbb{R}^{NT}$$
$$\boldsymbol{K} \in \mathbb{R}^{NT \times NT}$$

Inference Time $\sim \mathcal{O}((NT)^3)$

Exploit Structure using Kronecker Products

$$\boldsymbol{A} \otimes \boldsymbol{B} \equiv \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} \end{pmatrix}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$$

rowVec$(\boldsymbol{Y}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}_{NN} \otimes \boldsymbol{K}_{TT})$
colVec$(\boldsymbol{Y}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}_{TT} \otimes \boldsymbol{K}_{NN})$

Inference Time $\sim \mathcal{O}(N^3 + T^3)$

N independent Processes

rowVec$(\boldsymbol{Y}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{NN} \otimes \boldsymbol{K}_{TT}) = \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{Y}_{n,:} \mid \boldsymbol{0}, \boldsymbol{K}_{TT})$

Inference Time $\sim \mathcal{O}(T^3)$

Sparse Variational Gaussian Processes with P inducing points

Inference Time $\sim \mathcal{O}(TP^2)$ where $P \propto \log(T)$ is sufficient.

$\rightarrow$ total Inference Time $\sim \mathcal{O}(T \log T)$

Instantaneous Linear Mixing Model (ILMM)

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

$$y_{nt} = f_n(x_t) + \epsilon_{nt}$$

$$\boldsymbol{f}(x) = \boldsymbol{h}_1 u_1(x) + \boldsymbol{h}_2 u_2(x) + \ldots + \boldsymbol{h}_M u_M(x) = \boldsymbol{H}\boldsymbol{u}(x)$$

$$\rightarrow \boldsymbol{y} = \boldsymbol{H}\boldsymbol{u} + \boldsymbol{\epsilon} \qquad \boldsymbol{H} \in \mathbb{R}^{N \times M}$$

Instantaneous Linear Mixing Model (ILMM)

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

$$y_{nt} = f_n(x_t) + \epsilon_{nt}$$

$$\boldsymbol{f}(x) = \boldsymbol{h}_1 u_1(x) + \boldsymbol{h}_2 u_2(x) + \ldots + \boldsymbol{h}_M u_M(x) = \boldsymbol{H}\boldsymbol{u}(x)$$

$$\rightarrow \boldsymbol{y} = \boldsymbol{H}\boldsymbol{u} + \boldsymbol{\epsilon} \qquad \boldsymbol{H} \in \mathbb{R}^{N \times M}$$

Generative model

$\boldsymbol{u} \sim GP(\mathbf{0}, \boldsymbol{K}(x, x'))$	Prior
$\boldsymbol{f}(x)\,\vert\,\boldsymbol{H}, \boldsymbol{u}(x) = \boldsymbol{H}\boldsymbol{u}(x)$	Mixing
$\boldsymbol{y}(x)\,\vert\,\boldsymbol{f}(x) \sim GP(\boldsymbol{f}(x), \boldsymbol{\Sigma})$	Likelihood

Instantaneous Linear Mixing Model (ILMM)

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

$$y_{nt} = f_n(x_t) + \epsilon_{nt}$$

$$\boldsymbol{f}(x) = \boldsymbol{h}_1 u_1(x) + \boldsymbol{h}_2 u_2(x) + \ldots + \boldsymbol{h}_M u_M(x) = \boldsymbol{H}\boldsymbol{u}(x)$$

$$\rightarrow \boldsymbol{y} = \boldsymbol{H}\boldsymbol{u} + \boldsymbol{\epsilon} \qquad \boldsymbol{H} \in \mathbb{R}^{N \times M}$$

Generative model

$\boldsymbol{u} \sim GP(\boldsymbol{0}, \boldsymbol{K}(x, x'))$ Prior

$\boldsymbol{f}(x) \,|\, \boldsymbol{H}, \boldsymbol{u}(x) = \boldsymbol{H}\boldsymbol{u}(x)$ Mixing

$\boldsymbol{y}(x) \,|\, \boldsymbol{f}(x) \sim GP(\boldsymbol{f}(x), \boldsymbol{\Sigma})$ Likelihood

$$\rightarrow \; \boldsymbol{T}\boldsymbol{y} \,|\, \boldsymbol{u} \sim GP(\boldsymbol{u}, \boldsymbol{\Sigma}_T)$$

Instantaneous Linear Mixing Model (ILMM)

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

$$y_{nt} = f_n(x_t) + \epsilon_{nt}$$

$$\boldsymbol{f}(x) = \boldsymbol{h}_1 u_1(x) + \boldsymbol{h}_2 u_2(x) + \ldots + \boldsymbol{h}_M u_M(x) = \boldsymbol{H}\boldsymbol{u}(x)$$

$$\rightarrow \boldsymbol{y} = \boldsymbol{H}\boldsymbol{u} + \boldsymbol{\epsilon} \qquad \boldsymbol{H} \in \mathbb{R}^{N \times M}$$

Generative model

$\boldsymbol{u} \sim GP(\boldsymbol{0}, \boldsymbol{K}(x, x'))$ Prior

$\boldsymbol{f}(x) \,|\, \boldsymbol{H}, \boldsymbol{u}(x) = \boldsymbol{H}\boldsymbol{u}(x)$ Mixing

$\boldsymbol{y}(x) \,|\, \boldsymbol{f}(x) \sim GP(\boldsymbol{f}(x), \boldsymbol{\Sigma})$ Likelihood

$$\rightarrow \boldsymbol{T}\boldsymbol{y} \,|\, \boldsymbol{u} \sim GP(\boldsymbol{u}, \boldsymbol{\Sigma}_T)$$

Reduction from $(N \times T)$ to $(M \times T)$ $\qquad M \ll N$

Inference Time $\sim \mathcal{O}((MT)^3)$

Instantaneous Linear Mixing Model (ILMM)

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

$$y_{nt} = f_n(x_t) + \epsilon_{nt}$$

$$\boldsymbol{f}(x) = \boldsymbol{h}_1 u_1(x) + \boldsymbol{h}_2 u_2(x) + \ldots + \boldsymbol{h}_M u_M(x) = \boldsymbol{Hu}(x)$$

$$\rightarrow \boldsymbol{y} = \boldsymbol{Hu} + \boldsymbol{\epsilon} \qquad \boldsymbol{H} \in \mathbb{R}^{N \times M}$$

Generative model

$\boldsymbol{u} \sim GP(\boldsymbol{0}, \boldsymbol{K}(x, x'))$ Prior

$\boldsymbol{f}(x) \,|\, \boldsymbol{H}, \boldsymbol{u}(x) = \boldsymbol{Hu}(x)$ Mixing

$\boldsymbol{y}(x) \,|\, \boldsymbol{f}(x) \sim GP(\boldsymbol{f}(x), \boldsymbol{\Sigma})$ Likelihood

$$\rightarrow \; \boldsymbol{Ty} \,|\, \boldsymbol{u} \sim GP(\boldsymbol{u}, \boldsymbol{\Sigma}_T)$$

Reduction from $(N \times T)$ to $(M \times T)$ $\quad M \ll N$

Inference Time $\sim \mathcal{O}((MT)^3)$

Orthogonal ILMM (Decouples posterior latent processes)

$$\boldsymbol{K}(x, x') = k(x, x')\boldsymbol{I}_M$$

$$\boldsymbol{H} = \boldsymbol{US}^{\frac{1}{2}}$$

$$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_M + \boldsymbol{HDH}^T$$

Instantaneous Linear Mixing Model (ILMM)

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

$$y_{nt} = f_n(x_t) + \epsilon_{nt}$$

$$\boldsymbol{f}(x) = \boldsymbol{h}_1 u_1(x) + \boldsymbol{h}_2 u_2(x) + \ldots + \boldsymbol{h}_M u_M(x) = \boldsymbol{H}\boldsymbol{u}(x)$$

$$\rightarrow \boldsymbol{y} = \boldsymbol{H}\boldsymbol{u} + \boldsymbol{\epsilon} \qquad \boldsymbol{H} \in \mathbb{R}^{N \times M}$$

Generative model

$\boldsymbol{u} \sim GP(\boldsymbol{0}, \boldsymbol{K}(x, x'))$ Prior

$\boldsymbol{f}(x) \,|\, \boldsymbol{H}, \boldsymbol{u}(x) = \boldsymbol{H}\boldsymbol{u}(x)$ Mixing

$\boldsymbol{y}(x) \,|\, \boldsymbol{f}(x) \sim GP(\boldsymbol{f}(x), \boldsymbol{\Sigma})$ Likelihood

$$\rightarrow \ \boldsymbol{T}\boldsymbol{y} \,|\, \boldsymbol{u} \sim GP(\boldsymbol{u}, \boldsymbol{\Sigma}_T)$$

Reduction from $(N \times T)$ to $(M \times T)$ $\quad M \ll N$

Inference Time $\sim \mathcal{O}((MT)^3)$

Orthogonal ILMM (Decouples posterior latent processes)

$\boldsymbol{K}(x, x') = k(x, x')\boldsymbol{I}_M$

$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{S}^{\frac{1}{2}}$

Inference Time $\sim \mathcal{O}(MT^3)$

$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_M + \boldsymbol{H}\boldsymbol{D}\boldsymbol{H}^T$

sparse GPs $\rightarrow \mathcal{O}(MT)$