Vector-valued Gaussian Processes

Andrew Moore, Grady Wright (Department of Mathematics)

Keywords — Gaussian Processes, Statistics, Velocity Fields

I. Overview of Gaussian Processes

Gaussian processes generalize the multivariate Gaussian distribution and can be used to describe a probability distribution over families of functions.

i. Multivariate Gaussian Distribution

The multivariate normal distribution is used to model *random vectors* (vectors of jointly distributed random variables).

$$\begin{aligned} \boldsymbol{x} \in \mathbb{R}^N &\sim \mathcal{N}_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ \boldsymbol{\mu} \in \mathbb{R}^N &= (\mu_1, \mu_2, ..., \mu_N)^\top = (\mathbb{E}(x_1), \mathbb{E}(x_2), ..., \mathbb{E}(x_N))^\top \\ \boldsymbol{\Sigma} \in \mathbb{R}^{N\times N} &= \mathbb{E}\Big((\boldsymbol{x}-\mu)(\boldsymbol{x}-\mu)^\top\Big) = \big[\mathrm{cov}(x_i, x_j)\big]_{i,j}^N \\ x_i &\sim \mathcal{N}(\mu_i, \boldsymbol{\Sigma}_{ii}) \end{aligned} \tag{1}$$

ii. Gaussian Processes (GPs)

- *Gaussian process* (GP): an uncountably infinite collection of random variables; any finite sample is a draw from a MV Gaussian distribution.
- GPs are fully specified by a *mean function* m and *covariance (kernel) function* k.
- The kernel function must produce a positive semi-definite matrix when evaluated on a set of input points (or vectors).

We focus on the *squared exponential kernel* $k : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$, defined as:

$$k(x, x') = \alpha^2 \exp\left(-\frac{1}{2\rho^2}\,\|x - x'\|^2\right). \tag{2}$$

- $\|\cdot\|$ is the Euclidean Norm: $\|\boldsymbol{x}\| = \sqrt{x_1 + x_2 + \cdots + x_N}$
- α and ρ are *hyperparameters* (chosen, or estimated from data)

II. Gaussian Process Regression – Univariate $\mathbf{y}$

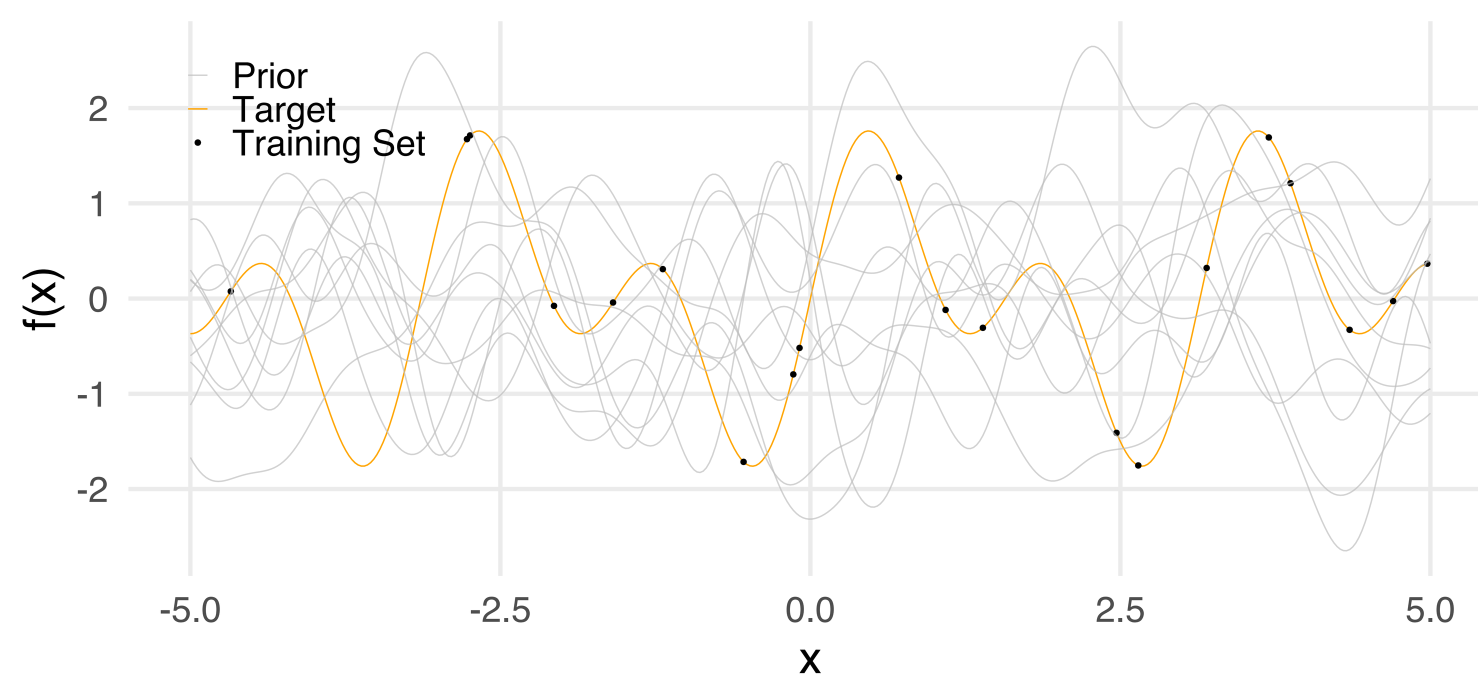
In practice, Gaussian Processes are often brought to bear on *regression problems*, in which an analyst has collected a dataset $S = (\boldsymbol{x}, \boldsymbol{y}) = \{(x_i, y_i) : x_i \in \mathbb{R}^p, y_i \in \mathbb{R}^d, i \in 1, 2, ..., N\}$ with the goal of learning the relationship f between $\boldsymbol{x}$ and $\boldsymbol{y}$.

Let $f(x) = \sin(2x) + \sin(4x)$ be an unknown function that an analyst is attempting to model using sampled data. Let $N = 20$ and $M = 400 - N$. We wish to use our sample S to predict the values $\boldsymbol{y}_* = f(\boldsymbol{x}_*)$ for test data $\boldsymbol{x}_* \in \mathbb{R}^M$.

We can draw samples from the prior distribution:

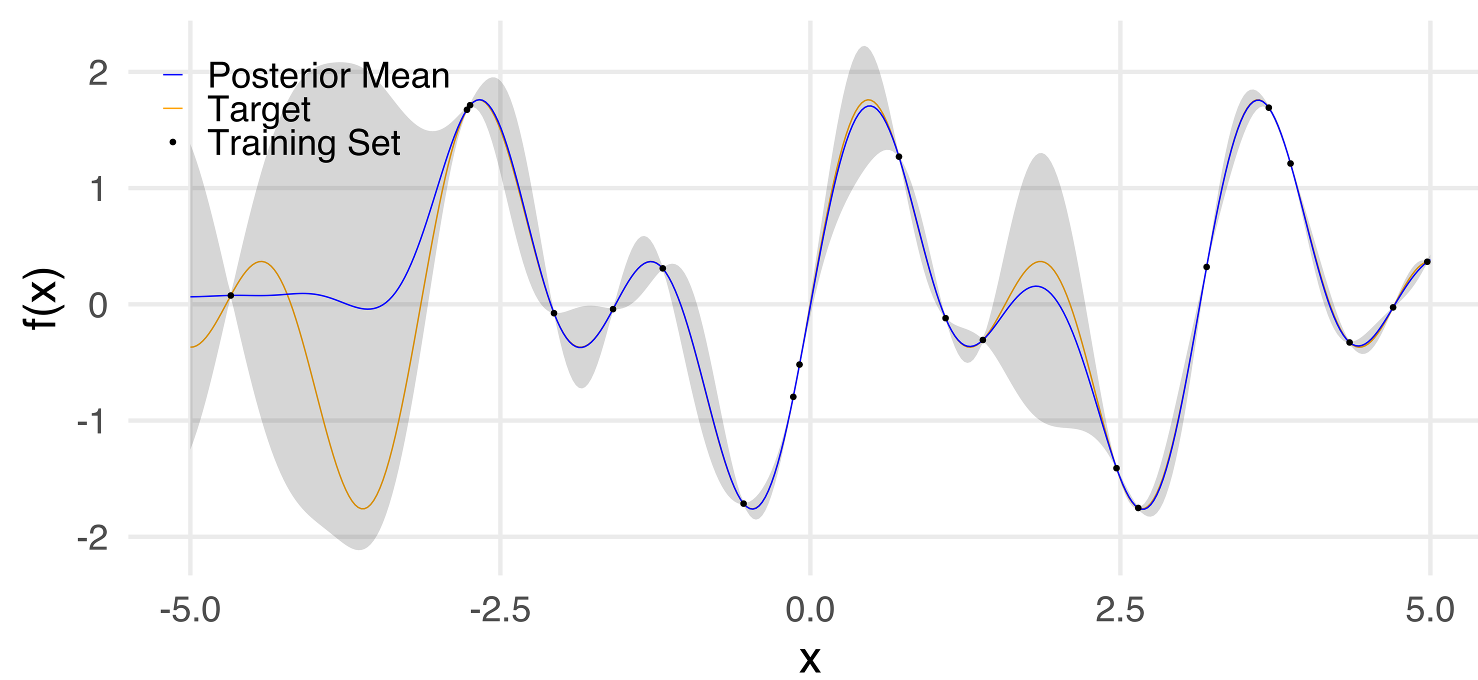
$$\begin{aligned} f &\sim \mathcal{GP}(\boldsymbol{0}, k) \\ \boldsymbol{y}_* &\sim \mathcal{N}_M(\boldsymbol{0}, k(\boldsymbol{x}_*, \boldsymbol{x}_*)) \\ k(\boldsymbol{x}_*, \boldsymbol{x}_*) \in \mathbb{R}^{M\times M} &= \big[k(\boldsymbol{x}_{*i}, \boldsymbol{x}_{*j})\big]_{i,j}^M. \end{aligned} \tag{3}$$

- Draws from the prior distribution (shown in grey) don't necessarily agree with our data points.
- Kernel choice determines properties of f (e.g., smoothness)



Our prior model for f and our observed data S can be combined to form a *posterior* distribution:

$$\begin{aligned} \boldsymbol{y}_* \mid \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}_* &\sim \mathcal{N}_M\big(\hat{\mu}, \hat{\Sigma}\big) \\ \hat{\mu} \in \mathbb{R}^M &= k(\boldsymbol{x}_*, \boldsymbol{x})(k(\boldsymbol{x}, \boldsymbol{x}))^{-1}\boldsymbol{y} \\ \hat{\Sigma} \in \mathbb{R}^{M\times M} &= k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{x}_*, \boldsymbol{x})(k(\boldsymbol{x}, \boldsymbol{x}))^{-1}k(\boldsymbol{x}_*, \boldsymbol{x})^\top \\ k(\boldsymbol{x}, \boldsymbol{x}) \in \mathbb{R}^{N\times N} &= \big[k(x_i, x_j)\big]_{i,j}^N \\ k(\boldsymbol{x}_*, \boldsymbol{x}) \in \mathbb{R}^{M\times N} &= \big[k(x_{*i}, x_j)\big]_{i,j}^{M,N}. \end{aligned} \tag{4}$$



III. Multioutput GPR – Vector-valued $\mathbf{y}$

1. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.
2. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.
3. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

Table 1: Lorem ipsum dolor sit amet

Lorem ipsum dolor sit.	Lorem ipsum.	Lorem ipsum.
Lorem ipsum dolor.	Lorem ipsum.	α
Lorem ipsum.	Lorem ipsum.	β
Lorem.	Lorem.	γ
Lorem ipsum dolor.	Lorem.	θ

Lorem Table 1 ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et.

i. Lorem ipsum dolor sit amet

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre.