Vector-valued Gaussian Processes

Andrew Moore, Grady Wright (Department of Mathematics)

Keywords — Gaussian Processes, Statistics, Velocity Fields

I. Overview of Gaussian Processes

Gaussian processes are generalizations of the multivariate Gaussian distribution. Rather than characterizing a probability distributions of vectors, Gaussian processes can be used to describe a probability distribution over families of functions.

i. Multivariate Gaussian (Multivariate Normal) Distribution

The multivariate normal distribution is used to model *random vectors* (vectors whose elements are jointly distributed random variables). This distribution is parameterized by a *mean vector* $\boldsymbol{\mu}$ and *covariance matrix* $\boldsymbol{\Sigma}$. Suppose $\boldsymbol{x} \in \mathbb{R}^N$ is drawn from a multivariate Gaussian distribution. Then, we can write the following:

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

ii. Gaussian Processes (GPs)

Formally, a *Gaussian process* (GP) is an uncountably infinite collection of random variables, with any finite sample from the process sharing a joint multivariate Gaussian distribution. GPs are fully specified by a *mean function* m and *covariance (kernel) function* k. For notational convenience, it's often assumed that the mean function is $\mathbf{0}$, but this is not required. The kernel function must produce a positive semi-definite matrix when evaluated on a set of input points (or vectors).

Positive semi-definite matrix.
Let $M \in \mathbb{R}^{N\times N}$ be a symmetric matrix. We say that M is positive semi-definite if, for all vectors $x \in \mathbb{R}^N$ not equal to $\mathbf{0}$, the following holds: $x^\top M x \geq 0$.

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

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

where $\|\cdot\|$ is the Euclidean norm. The squared exponential kernel has two *hyperparameters*, α and ρ, which control the variance scale and length scale of functions drawn from the GP. Typically, α is set to 1, and ρ is specified based on domain knowledge or is estimated from observed data. This choice of kernel function reflects the assumption that the covariance between two points (or vectors) x and x' decays exponentially based on the distance between them.

Euclidean (L2) Norm.
Let $\boldsymbol{x} \in \mathbb{R}^N$. The Euclidean Norm of $\boldsymbol{x}$, written $\|\boldsymbol{x}\|$, is defined as

$$\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^{N} x_i} = \sqrt{x_1 + x_2 + \cdots + x_N}. \tag{3}$$

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 $\boldsymbol{y}$:

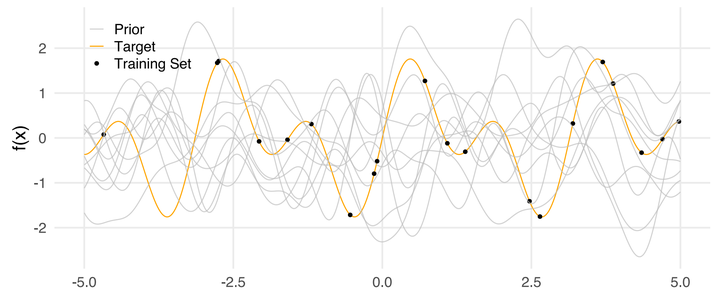
$$\begin{aligned} \boldsymbol{y} &= f(\boldsymbol{x}) \quad \text{or} \\ \boldsymbol{y} &= f(\boldsymbol{x}) + \varepsilon \quad (\text{with additive noise}). \end{aligned} \tag{4}$$

Once f has been estimated, it can then be used to predict the values of future or test points. Gaussian Process Regression can be considered a Bayesian method for learning f. As an overview, we'll consider an example using simulated data where $x_i, y_i \in \mathbb{R}$.

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:

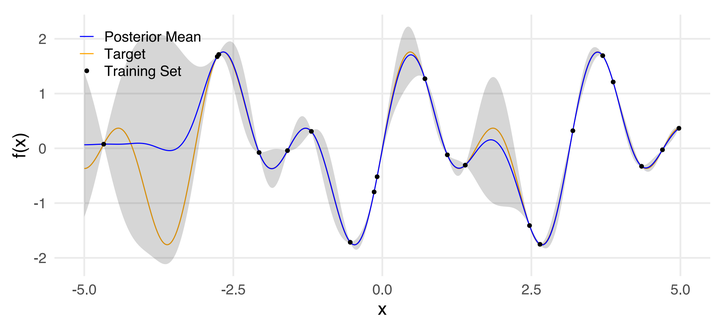
$$\begin{aligned} f &\sim \mathcal{N}_M(\mathbf{0}, k(\boldsymbol{x}_*, \boldsymbol{x}_*)) \\ k(\boldsymbol{x}_*, \boldsymbol{x}_*) \in \mathbb{R}^{M\times M} &= \big[k(\boldsymbol{x}_{*i}, \boldsymbol{x}_{*j})\big]_{i,j}^M = \begin{pmatrix} k(x_{*1}, x_{*1}) & k(x_{*1}, x_{*2}) & \dots & k(x_{*1}, x_{*M}) \\ k(x_{*2}, x_{*1}) & k(x_{*2}, x_{*2}) & \dots & k(x_{*2}, x_{*M}) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_{*M}, x_{*1}) & k(x_{*M}, x_{*2}) & \dots & k(x_{*M}, x_{*M}) \end{pmatrix}. \end{aligned} \tag{5}$$

Draws from the prior distribution (shown in grey) don't necessarily agree with our data points. They represent our state of knowledge before observing the training data. In this instance, our kernel function ensures that draws are smooth curves– matching what we might believe is a property of the target function.



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

$$\begin{aligned} \boldsymbol{y}_* \mid \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}_* &\sim \mathcal{N}_M\big(\hat{\mu}, \hat{\Sigma}\big) \\ \hat{\mu} \in \mathbb{R}^M &= k(\boldsymbol{x}_*, \boldsymbol{x})(k(\boldsymbol{x}, \boldsymbol{x}))^{-1}\boldsymbol{y} \\ \hat{\Sigma} \in \mathbb{R}^{M\times M} &= k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{x}_*, \boldsymbol{x})(k(\boldsymbol{x}, \boldsymbol{x}))^{-1}k(\boldsymbol{x}_*, \boldsymbol{x})^\top \\ k(\boldsymbol{x}, \boldsymbol{x}) \in \mathbb{R}^{N\times N} &= \big[k(x_i, x_j)\big]_{i,j}^N \\ k(\boldsymbol{x}_*, \boldsymbol{x}) \in \mathbb{R}^{M\times N} &= \big[k(x_{*i}, x_j)\big]_{i,j}^{M,N}. \end{aligned} \tag{6}$$



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.

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.

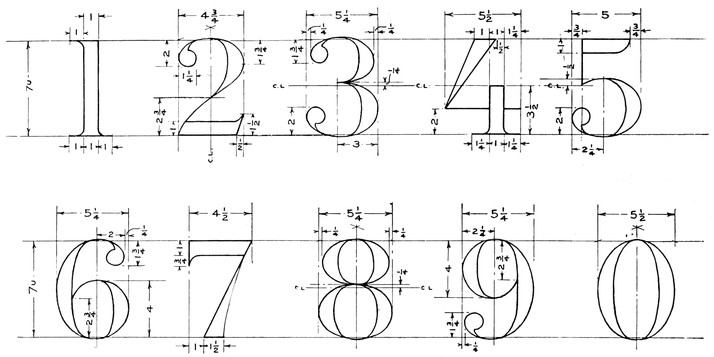


Figure 1: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.