Sandia National Laboratories

Exceptional service in the national interest

SCIENTIFIC MACHINE LEARNING AND TENSORFLOW TUTORIAL

Gaussian Processes

Ravi G Patel

Scientific Machine Learning Department

February 1 – 2, 2024

Numerical PDEs: Analysis, Algorithms, and Data Challenges

ICERM

Brown University

U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND2024-00936O

GAUSSIAN PROCESSES IS A GENERALIZATION OF THE MULTIVARIATE NORMAL

- We previously looked at fitting general linear models
- Leveraging conjugacy for Gaussian likelihoods and priors, we can easily compute
  - Posterior
  - Posterior predictive
  - Hyperparameters
- Let's look at piecewise constant Bayesian fits to data
  - Increasing number of intervals

PROPERTIES OF MULTIVARIATE NORMAL DISTRIBUTIONS

- A multivariate normal is parameterized by a mean and positive definite covariance,

$$y \sim MvN(\mu, \Sigma)$$

- Sampling a MvN,

$$y_i = \mu + L z_i$$
$$z_i \sim MvN(0, I), \quad L = \text{Choleskey}(\Sigma)$$

- Marginal distribution,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \sim MvN\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix}\right) \rightarrow \begin{array}{l} y_1 \sim MvN(\mu_1, \Sigma_{11}) \\ y_2 \sim MvN(\mu_2, \Sigma_{22}) \end{array}$$

- Conditional distribution

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \sim MvN\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix}\right) \rightarrow \begin{array}{l} y_1 | y_2 \sim MvN(\mu_{1|2}, \Sigma_{1|2}) \\ \mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2) \\ \Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T \end{array}$$

- Linear transformations of MvN RV's are also MvN RV's

$$y \sim MvN(\mu, \Sigma) \rightarrow Ly \sim MvN(L\mu, L\Sigma L^T)$$

MVN FOR FUNCTION EVALUATIONS

- Model function evaluations at a discrete set of points as a multivariate normal,
  - For $y_1 = y(x_1), y_2 = y(x_2) \rightarrow p\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = MvN\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right)$
  - For $\boldsymbol{y} = \{y_1, y_2, \ldots\} \rightarrow p(\boldsymbol{y}) = MvN(\boldsymbol{\mu}, \Sigma)$
- Covariance controls how smooth the function is and it’s magnitude

DEFINITION OF A GAUSSIAN PROCESS

- A Gaussian process (GP) is the infinite dimensional analog of a MvN
- We let the mean be a function and replace the covariance with a kernel,

$$y \sim \mathcal{GP}[\mu, K]$$

$$\mu : [0,1] \to \mathbb{R} \qquad K : [0,1] \times [0,1] \to \mathbb{R} \quad \text{where } K(x_i, x_j) = K(x_j, x_i) \text{ and } \sum_{ij} c_i c_j K(x_i, x_j) \geq 0$$

- To implement on a computer, we can sample a function by marginalizing and recovering a MvN for its function evaluations at a finite set of points,

$$\begin{bmatrix} y_1 \\ y_2 \\ y \backslash \{y_1, y_2\} \end{bmatrix} \sim \mathcal{GP}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu \backslash \{\mu_1, \mu_2\} \end{bmatrix}, \begin{bmatrix} K_{11} & K_{12} & \\ K_{12}^T & K_{22} & K_{12,*} \\ K_{12,*}^T & & K_{*,*} \end{bmatrix} \right) \to \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \sim MvN\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} K_{11} & K_{12} \\ K_{12}^T & K_{22} \end{bmatrix} \right)$$

  - We write, $y_i = y(x_i), K_{ij} = K(x_i, \tilde{x}_j)$
  - For sets of positions, $K(x, \tilde{x}) = \begin{bmatrix} K_{11} & K_{12} & \dots \\ K_{12} & K_{22} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$

REGRESSION WITH GAUSSIAN PROCESSES

- We perform regression by finding the distribution of new function evaluations, y_* at x_* , conditioned on data, y_d

$$y_*|y_d \sim MvN(\mu_{*|d}, K_{*|d})$$
$$\mu_{*|d} = \mu_* + K_{*d}K_{dd}^{-1}(y_d - \mu_d)$$
$$K_{*|d} = K_{**} - K_{*d}K_d^{-1}K_{d*}$$

- We can compute all covariances, since we have the covariance kernel
- The square exponential is a common kernel,

$$K(x, \tilde{x}) = a\exp(-l(x - \tilde{x})^2))$$

  - where a, l are hyperparameters

CONSTRAINTS WITH GAUSSIAN PROCESS

- We can similarly add constraints by, e.g., boundary conditions, by conditioning

$$y_* \Big| \begin{bmatrix} y_d \\ y_c \end{bmatrix} \sim MvN(\mu_{*|dc}, K_{*|dc})$$

$$\mu_{*|dc} = \mu_* + K_{*dc} K_{dc,dc}^{-1} (y_{dc} - \mu_{dc})$$

$$K_{*|dc} = K_{**} - K_{*dc} K_{dc,dc}^{-1} K_{*dc}^T$$

TYPE II LIKELIHOOD MAXIMIZATION

- As with MvN, the hyperparameters in GP's can be tuned by maximizing the Type II likelihood,

$$\min_{a,l} \; y_d^T K_{dd}(a,l)^{-1} y_d + \log\det(K_{dd}(a,l))$$

LINEAR TRANSFORMATIONS OF A GAUSSIAN PROCESS

- Just as for MvN's, a linear transform of a GP is a GP,

$$y \sim \mathcal{GP}(\mu, K) \to Ly \sim \mathcal{GP}(\nu, G)$$
$$G = L[L[K(\cdot, \tilde{x})](x, \cdot)] = LKL^T$$
$$\nu = L\mu$$

- E.g. scalar multiplication, $L(y) = ay \to G = a^2 K$

- E.g., the derivative operator, $L = \partial_x \to G = \partial_x \partial_{\tilde{x}} K$

- E.g., the integral operator, $(Ly)(x) = \int_0^x y(z)dz \to G = \int_0^{\tilde{x}} \int_0^x K(z, \tilde{z}) dz d\tilde{z}$

INVERSE PROBLEM FOR PDE'S

- Linear transformations of GPs can be parameterized,

$$u \sim \mathcal{GP}[\mu, K] \rightarrow L_\phi u \sim \mathcal{GP}[L_\phi \mu, L_\phi[L_\phi[K(\cdot, \tilde{x})](x, \cdot)]]$$

- Using this, we can infer a PDE, e.g.,

$$\phi \Delta u = f$$
$$L_\phi = \phi \Delta$$

- To infer parameterized linear operators, optimize the type II likelihood for the joint distribution

$$K_\phi = L_\phi[K(\cdot, \tilde{x})]$$
$$K_{\phi\phi} = L_\phi[K_\phi](x, \cdot)]$$
$$G = \begin{bmatrix} K & K_\phi \\ K_\phi^T & K_{\phi\phi} \end{bmatrix}$$
$$\min_{a,l,\phi} \begin{bmatrix} u_d \\ f_d \end{bmatrix}^T G^{-1} \begin{bmatrix} u_d \\ f_d \end{bmatrix} + \log\det(G)$$

M. Raissi and G.E. Karniadakis, arXiv:1701.02440v1 (2017)

INFERENCE FOR TIME EVOLVING SYSTEMS

- Linear PDE's can similarly be inferred
- The first order Euler update is a linear operator,

$$\partial_t u = Hu$$

$$u^{n+1} = u^n + \Delta t H u^n = (I + \Delta t H) u^n$$

- If $u^n \sim \mathcal{GP}[\mu, K]$

for $L_\phi = (I + \Delta t H_\phi)$ we have $u^{n+1} \sim \mathcal{GP}[L_\phi \mu, L K L_\phi^T]$

- As before, we can maximize the type II likelihood for the joint distribution of $\begin{bmatrix} u^n \\ u^{n+1} \end{bmatrix}$

- This works for other integration schemes with linear update operators, e.g., backwards Euler, exponential, etc.

M. Raissi and G.E. Karniadakis, arXiv:1701.02440v1 (2017)

EXERCISE

- Use the provided data to infer the inhomogenous conductivity in the Poisson equation,

$$
\begin{aligned}
-\nabla \cdot k \nabla u &= f \qquad x \in (-1, 1) \\
u &= 0 \qquad x \in \{-1, 1\}
\end{aligned}
$$

- using Legendre polynomials for $k(x)$