Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 7, n. 1, 2020.

Trabalho apresentado no XXXIX CNMAC, Uberlândia - MG, 2019.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Use of surrogate models in tumor growth modeling

João V. O. Silva[1]
Regina C. Almeida [2]
Renato S. Silva [3]
Laboratório Nacional de Computação Científica, LNCC, Petrópolis, RJ

1 Introduction

Cancer modeling is a class of diseases that span mechanisms occuring in different time and space scales. Current modeling framework combines continuum models and individual based models to better represent such heterogeneous multiscale dynamics. Due to inherent uncertainties involved in this kind of problem, Bayesian inference for parameters estimation [1] is an adequate methodology since it allows taking into account uncertanties due to the presence of error in the measurements as well as model inadequacies. It relies on simulating the forward model for many possible different configurations of the parameter set, which can lead to an overwhelming computational burden. Surrogate models have been developed in the literature to alleviate this issue. According to the source, this approach can be also named as a metamodel, reduced model, response surface, among others. The general idea is to build surrogates for the quantity of interest that can be cheaply and accurately evaluated without requiring to run expensive models. The accuracy of such strategy for Bayesian inference is still an open issue [3].

Surrogate models can be obtained through simplifications of the physical problem (hierarchical models), projections of the governing equations onto an appropriate reduced vector basis (projection models), or directly using the available data (data-driven models). Here we investigate a data-driven approach namely the Gaussian Process Regression (GPR) [1]. Unlike classical regression schemes that assume a simplified model to represent the input data and compute the function coefficients to minimize a choosen error measure (mean squared error, e.g.), GPR is known as a non-parametric approach. Specifically, given a data set $\mathbf{y} = \{y^{(1)}, \ldots, y^{(n)}\}$ at $\mathbf{x} = \{x^{(1)}, \ldots, x^{(n)}\}$ and assuming $y^{(i)} = h(x^{(i)}) + \sigma^{(i)}$, where $h(x^{(i)})$ is a candidate function evaluated at $x^{(i)}$ and $\sigma^{(i)} \sim \mathcal{N}(0, \sigma^2_{noise})$ is the noise, a desired prediction $\mathbf{h}_*$ and the data $\mathbf{y}$ form the following multivariate Gaussian distribution:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{h}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K + \sigma^2_{noise} I & K_*^T \\ K_* & K_{**} \end{bmatrix}\right),$$

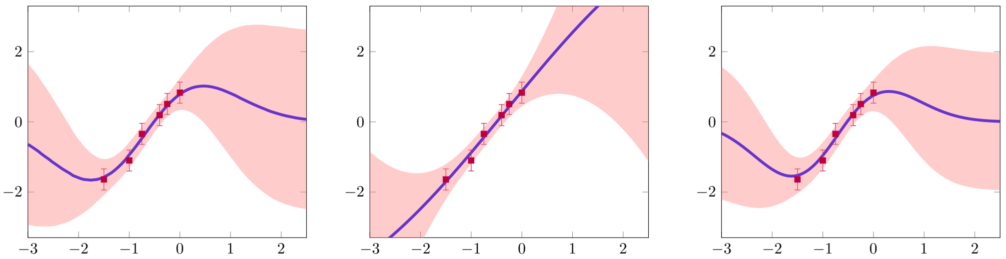
in which $K_{ij} = k(x^{(i)}, x^{(j)})$, $K_* = k(x^{(*)}, x^{(i)})$, $i, j = 1, \ldots, n$, and $K_{**} = k(x^{(*)}, x^{(*)})$. The choice of the kernel $k(x, x')$ plays a major role in this model approach. Typical choices

[1]jvos@lncc.br
[2]rcca@lncc.br
[3]rssr@lncc.br

are the squared exponential and rational quadratic kernels given by:

$$k_{\text{SE}}(x,x') = \sigma_f^2 \exp\left(\frac{-(x-x')^2}{2\lambda_f^2}\right), \quad k_{\text{RQ}}(x,x') = \sigma_e^2 \left(1 + \frac{-(x-x')^2}{2\alpha_r\lambda_e^2}\right)^{-\alpha_r}$$

in which and σ_f, λ_f, σ_e, α_r and λ_e are hyperparameters. The predictions $\mathbf{h}_*$ are obtained by sampling from the posterior distribution $\mathbf{h}_*|\mathbf{y} \sim \mathcal{N}(K_*(K+\sigma_{noise}^2 I)^{-1}\mathbf{y}, \quad K_{**} - K_*(K+\sigma_{noise}^2 I)^{-1}K_*^T)$. Some examples of GPR application are shown in Figure 1.



(a) Squared exponential kernel: $(\sigma_f, \lambda_f) = (1.3, 1.04)$

(b) Squared exponential kernel: $(\sigma_f, \lambda_f) = (13, 5)$

(c) Rational quadratic kernel: $(\sigma_e, \alpha_r, \lambda_e) = (1, 2.2\text{e}{+}7, 0.85)$

Figure 1: Gaussian process regression using an input data (red) with noise of $\sigma_{noise} = 0.3$. There are noticeable differences in the resulting mean function (blue) and 95% confidence interval (shadow) depending on the choices made for the kernel function and hyperparameters.

2 Objective

The underlying aspects of the GPR method will be analyzed in this work. The use of distinct kernel functions and the optimal selection of their hyperparameters will be investigated, as well as the numerical algorithms used in the calculations needed to compute the method's associated posterior distribution. By realizing this analysis, this work aims to detail the benefits and difficulties of using GPR as a surrogate metamodel in a tumor growth problem.

Acknowledgements

João V. O. Silva thanks CNPq for the finantial support.

References

[1] C. E. Rasmussen, C. K. I. Williams. Gaussian Process Regression for Machine Learning. *MIT Press*, 2006.

[2] M. E. Tipping. Bayesian Inference: An Introduction to Principles and Practice in Machine Learning. *Lecture Notes in Computer Science*, 41–62, 2004.

[3] L. Yan and Y. Zhang. Convergence analysis of surrogate-based methods for Bayesian inverse problems. *Inverse Problems*, 33(12) 125001, 2017.