Master Thesis:

Is Gaussian process regression robust under weakly perturbed data?

Course of Study: Mathematics, Machine Learning, Computational Statistics

Topic

Stochastic differential equations (SDEs) provide an efficient framework for modeling and describing complex systems in various applications. In a very basic form their solutions are given by a continuous-time Markov process X satisfying the equation

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t), \quad t > 0, \quad X(0) = x_0,$$

where x_0 is an initial point and W is a standard Brownian motion. Let us, for simplicity, consider the case where X takes values in $\mathbb{R}$. A fine equidistant time-discretization with the Euler-Maruyama scheme yields the discrete model equation

$$\Delta x_i = b(x_i)\Delta t + \sigma(x_i)\Delta W_i, \quad x_i := X(i\Delta t), \quad \Delta t > 0,$$

where $\Delta x_i := x_{i+1} - x_i, \Delta W_i := W_{i+1} - W_i \sim \mathcal{N}(0, \Delta t)$. This model equation lends itself to a regression framework of the form

$$y_i = f(x_i) + \epsilon_i, \quad \epsilon_i \overset{ind.}{\sim} \mathcal{N}(0, \sigma^2(x_i)\Delta t).$$

The goal now is to estimate the functions b and σ given the observations $\mathbf{x} = (x_i)_{i=0,\ldots,N}$ through solving an appropriate Bayesian inverse problem with Gaussian processes as priors for the functions b and σ and using the resulting posterior means as estimators.

Tasks

While the presented method is fairly established in the recent literature of the statistical estimation for SDEs, there seems to be no exact investigation of robustness properties of these nonparametric estimators under so-called weakly perturbed data; this kind of data is described through another SDE whose solution X_ϵ converges weakly to X as $\epsilon \to 0$. So in this project there will be a particular focus on (mostly) numerical exploration of how the nonparametric estimators behave under realizations of the process X_ϵ when ϵ is close to zero.

Other tasks include numerical implementation in any preferred language of choice, a proper review of the necessary mathematical background, and, if time and interest permits, improving computational tractability of the method, e.g. through sparse Gaussian processes with spherical harmonic features, or diving into theoretical considerations of statistical properties of the estimators.

What we offer

A master thesis project close to cutting-edge research with the opportunity to bring in own ideas and a potential prospect of publication in a journal or a likewise scientific contribution.

Contact: Sebastian Krumscheid, Jaroslav Borodavka

E-mail: sebastian.krumscheid@kit.edu, jaroslav.borodavka@kit.edu