First Take on GPs

Cory Lanker

10/24/2019

Gaussian Processes: High-Level Summary

The goal in regression is to approximate the regression function. We aim for the "best" function $f(x)$ that maps the (multivariate) input space $X = x$ to the continuous (scalar) response y. A statistical model is a class of functions related by a set of parameters, and generally restricting the choice of $f(x)$ to a statistical model's family of functions is too restrictive, resulting in (potentially significant) estimator bias. As opposed to restricting the collection of functions to be considered for $f(x)$ and then selecting the one that matches the data best (i.e., maximizes the likelihood), we can instead consider all functions (without restriction) and place a prior on each function according to how likely we would want that function to be our chosen $f(x)$. For example, smoother functions would have much higher prior weight. A prior that accomplishes the task of assigning weights to functions is a *process*. This approach is not tractable, but an elegant solution is to discretize the space where the function is considered—and being even more restrictive, allow only a finite number of x to be considered.

Pursuing this further, take any $f(x)$. Evaluate $f(x)$ at the finite number of points under consideration: $x_1, x_2, \ldots, x_n$. These values have positive probability of coming from the following model:

$$MVN_n(h(x), \Sigma)$$

where $h(x)$ is some arbitrary defined function and Σ is a covariance method that is not singular. Though $f(x)$ itself cannot be generated with the MVN distribution (not unless $f(x)$ is itself from the same type of MVN), its n-point reduction can be approximated. The "prior" then is the combination of $h(x)$, likely from some demeaning operation using whatever statistical methods are applicable, and the covariance Σ, as different Σ will sharply control the probability of the different X drawn from this MVN distribution.