AutoML: Gaussian Processes

Covariance Functions for GPs - Advanced

Bernd Bischl Frank Hutter Lars Kotthoff
Marius Lindauer Joaquin Vanschoren

MS-Continuity and Differentiability I

We wish to describe a Gaussian process in terms of its smoothness. There are several notions of continuity for random variables. One is the continuity/differentiability in mean square (MS).

Definition

A Gaussian process $f(\mathbf{x})$ is said to be **MS continuous** at $\mathbf{x}_*$, if

$$\mathbb{E}[|f(\mathbf{x}^{(k)}) - f(\mathbf{x}_*)|^2] \overset{k\to\infty}{\longrightarrow} 0 \text{ for all converging sequences } \mathbf{x}^{(k)} \overset{k\to\infty}{\longrightarrow} \mathbf{x}_*.$$

A Gaussian process $f(\mathbf{x})$ is said to be **MS differentiable** along the i direction, if the following limit exists, with $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$ being the unit vector along the i-th axis.

$$\lim_{h\to 0} \mathbb{E}[|\frac{f(\mathbf{x} + h\,\boldsymbol{e}_i) - f(\mathbf{x})}{h}|]$$

MS-Continuity and Differentiability II

- The MS continuity/differentiability do not necessarily lead to the continuity/differentiability of sampled functions!
- The MS continuity/differentiability of a Gaussian process can be derived from the smoothness properties of the kernel.
- The GP is continuous in MS iff the covariance function $k(\mathbf{x}, \mathbf{x}')$ is continuous.
- The MS derivative of a Gaussian process exists iff the second derivative $\frac{\partial^2 k(\mathbf{x},\mathbf{x}')}{\partial \mathbf{x} \partial \mathbf{x}'}$ exists.

Squared Exponential Covariance Function

The squared exponential function is one of the most commonly used covariance functions.

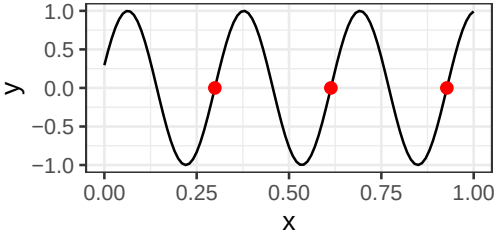
$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\ell^2}\right).$$

Properties:

- It depends merely on the distance $r = \|\mathbf{x} - \mathbf{x}'\| \rightarrow$ isotropic and stationary.
- Infinitely differentiable $\rightarrow$ the corresponding GP is too smooth.
- It utilizes strong smoothness assumptions $\rightarrow$ unrealistic for modeling most of the physical processes.

Upcrossing Rate and Characteristic Length-Scale I

- Another way to describe a Gaussian process is the expected number of up-crossings at level-0 on the unit interval, which we denote by N_0.



- For an isotropic covariance function $k(r)$, the expected number of up-crossings can be calculated explicitly:

$$\mathbb{E}[N_0] = \frac{1}{2\pi}\sqrt{\frac{-k''(0)}{k(0)}}.$$

Upcrossing Rate and Characteristic Length-Scale II

Example (squared exponential):

$$
\begin{aligned}
k(r) &= \exp\left(-\frac{r^2}{2\ell^2}\right) \\
k'(r) &= -k(r) \cdot \frac{r}{\ell^2} \\
k''(r) &= k(r) \cdot \frac{r^2}{\ell^4} - k(r) \cdot \frac{1}{\ell^2}
\end{aligned}
$$

The expected number of upcrossings at level-0 is

$$
\mathbb{E}[N_0] = \frac{1}{2\pi}\sqrt{\frac{-k''(0)}{k(0)}} = \frac{1}{2\pi}\sqrt{\frac{1}{\ell^2}} = (2\pi\ell)^{-1}.
$$