AutoML: Gaussian Processes

Covariance Functions for GPs

Bernd Bischl Frank Hutter Lars Kotthoff
Marius Lindauer Joaquin Vanschoren

Covariance function of a GP I

The marginalization property of the Gaussian process implies that for any set of input values, the corresponding vector of function values is Gaussian:

$$\boldsymbol{f} = \left[f\left(\mathbf{x}^{(1)}\right), \ldots, f\left(\mathbf{x}^{(n)}\right)\right] \sim \mathcal{N}\left(\boldsymbol{m}, \boldsymbol{K}\right).$$

- The covariance matrix $\boldsymbol{K}$ is constructed according to the chosen inputs $\left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\right\}$.
- Each entry $\boldsymbol{K}_{ij}$ is computed by $k\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right)$.
- Technically, to be a valid covariance matrix, $\boldsymbol{K}$ needs to be positive semi-definite for **every** choice of inputs $\left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\right\}$.
- A function $k(\cdot, \cdot)$ that satisfies this condition is called **positive definite**.

Covariance function of a GP II

- Recall that the purpose of the covariance function is to control to which degree the following condition is fulfilled:

If $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ are close in the $\mathcal{X}$-space, their function values $f(\mathbf{x}^{(i)})$ and $f(\mathbf{x}^{(j)})$ should be close in $\mathcal{Y}$-space.

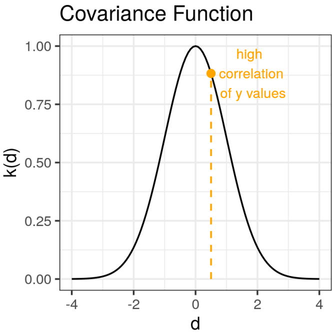
- Closeness of $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ in the input space $\mathcal{X}$ is measured by $\boldsymbol{d} = \mathbf{x}^{(i)} - \mathbf{x}^{(j)}$.

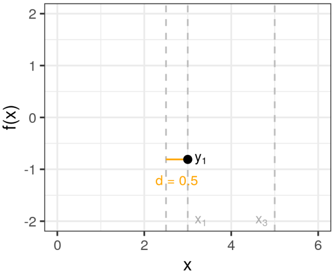
- $\boldsymbol{K}_{ij}$ is the covariance of $f(\mathbf{x}^{(i)})$ and $f(\mathbf{x}^{(j)})$, and **stationary** covariance functions are those in which the following holds:

$$k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = k(\boldsymbol{d})$$

Covariance function of a GP: Example I

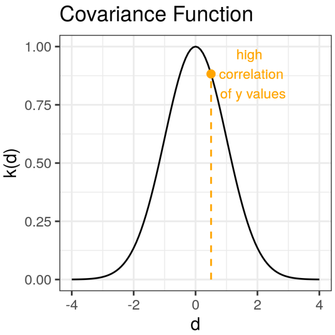
- Let $f(\mathbf{x})$ be a GP with $k(\mathbf{x}, \mathbf{x}') = \exp(-\frac{1}{2}\|\boldsymbol{d}\|^2)$ where $\boldsymbol{d} = \mathbf{x} - \mathbf{x}'$.
- Consider two points $\mathbf{x}^{(1)} = 3$ and $\mathbf{x}^{(2)} = 2.5$. To investigate how correlated their function values are, compute their correlation!

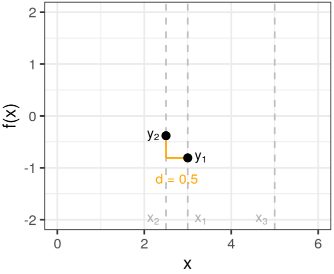




Covariance function of a GP: Example II

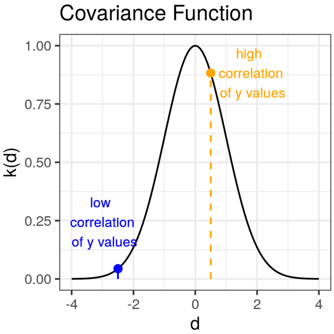
- Assume that we observe a value of $y^{(1)} = -0.8$. Under the said assumption for the Gaussian process, the value of $y^{(2)}$ should be close to $y^{(1)}$.

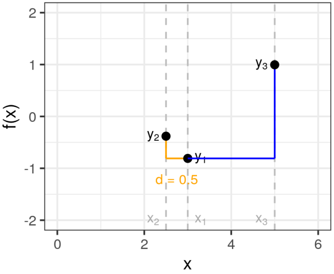




Covariance function of a GP: Example III

- Now, let us take a new point $\mathbf{x}^{(3)}$ which is not too close to $\mathbf{x}^{(1)}$.
- Their function values should not be so correlated. That is, $y^{(1)}$ and $y^{(3)}$ are probably far away from each other.

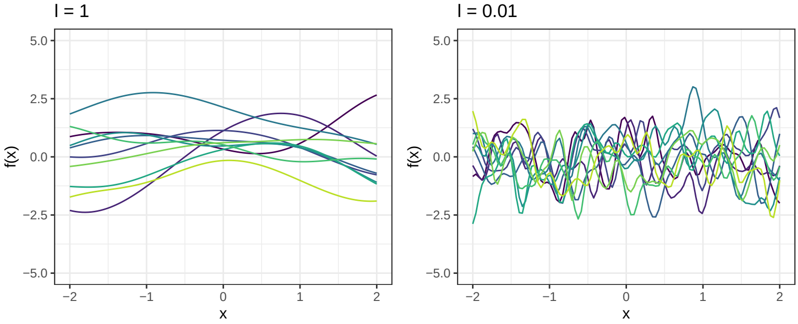




Sampling from a GP: Covariance Function

Let us draw 10 functions from a Gaussian process prior with the squared exponential covariance function but with two different values of ℓ, also called the characteristic length-scale.

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2\ell^2}\|\mathbf{x} - \mathbf{x}'\|^2\right)$$



Covariance Functions

Three types of properties are commonly used in covariance functions:

- k is **stationary** if its returning values depend on $\boldsymbol{d} = \mathbf{x} - \mathbf{x}'$ and is denoted by $k(\boldsymbol{d})$.
- k is **isotropic** if its returning values depend on $r = \|\mathbf{x} - \mathbf{x}'\|$and is denoted by $k(r)$.
- k is a **dot product** if its returning values depend on $\mathbf{x}^T\mathbf{x}'$.

- Isotropy implies stationarity.
- Isotropic functions are rotationally invariant.
- Stationary functions are translationally invariant:

$$k(\mathbf{x}, \mathbf{x} + \boldsymbol{d}) = k(\mathbf{0}, \boldsymbol{d}) = k(\boldsymbol{d})$$

Commonly Used Covariance Functions I

Name	$k(\mathbf{x}, \mathbf{x}')$
constant	σ_0^2
linear	$\sigma_0^2 + \mathbf{x}^T \mathbf{x}'$
polynomial	$(\sigma_0^2 + \mathbf{x}^T \mathbf{x}')^p$
squared exponential	$\exp(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\ell^2})$
Matérn	$\frac{1}{2^\nu \Gamma(\nu)} \left(\frac{\sqrt{2\nu}}{\ell} \|\mathbf{x} - \mathbf{x}'\| \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}}{\ell} \|\mathbf{x} - \mathbf{x}'\| \right)$
exponential	$\exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|}{\ell}\right)$

$K_\nu(\cdot)$ is the modified Bessel function of the second kind.

Commonly Used Covariance Functions II

- Some random functions drawn from Gaussian processes with a Squared Exponential Kernel (left), Polynomial Kernel (middle), and a Matérn Kernel (right, $\ell = 1$).
- The choice of the hyperparameter determines the "wiggliness" of the function.