Covariance functions for libgp

February 20, 2019

1 Gaussian Process

We are given $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^n$ and $\boldsymbol{z} = (z_1, \ldots, z_n)$ with $\boldsymbol{x}_i \in \mathbb{R}^d$ and $z_i \in \mathbb{R}$. The joint distribution of $\boldsymbol{z}$ is given by,

$$p(\boldsymbol{z}) = \mathcal{N}\left(\boldsymbol{z} \,|\, 0, \boldsymbol{C}_n\right) \tag{1}$$

where $\boldsymbol{C}_n$ is a matrix with elements $(\boldsymbol{C}_n)_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the covariance function of the Gaussian process. It should hold that $k(\boldsymbol{x}, \boldsymbol{y}) = k(\boldsymbol{y}, \boldsymbol{x})$

Given a new point $(\boldsymbol{x}_{n+1}, z_{n+1})$, the conditional distribution of z_{n+1} on $\boldsymbol{X}$ and $\boldsymbol{z}$ is given by,

$$p(z_{n+1} \,|\, \boldsymbol{X}, \boldsymbol{z}, \boldsymbol{x}_{n+1}) = \mathcal{N}\left(z_{n+1} \,|\, \mu(\boldsymbol{x}_{n+1}), \sigma^2(\boldsymbol{x}_{n+1})\right), \tag{2}$$

where

$$\mu(\boldsymbol{x}_{n+1}) = \boldsymbol{k}^\top \boldsymbol{C}_n^{-1} \boldsymbol{y}, \tag{3}$$

and

$$\sigma^2(\boldsymbol{x}_{n+1}) = c - \boldsymbol{k}^\top \boldsymbol{C}_n^{-1} \boldsymbol{k}. \tag{4}$$

Here, $(\boldsymbol{k})_i = k(\boldsymbol{x}_i, \boldsymbol{x}_{n+1})$, $i = 1, \ldots, n$ and $c = k(\boldsymbol{x}_{n+1}, \boldsymbol{x}_{n+1})$ and satisfy,

$$\boldsymbol{C}_{n+1} = \begin{pmatrix} \boldsymbol{C}_n & \boldsymbol{k}^\top \\ \boldsymbol{k} & c \end{pmatrix}. \tag{5}$$

1.1 Derivatives of the Gaussian process w.r.t. state variables

We are interested in calculating derivatives of $\mu(\boldsymbol{x})$ w.r.t. the state variable $\boldsymbol{x}$. Thus

$$\frac{\mathrm{d}\mu(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} = \frac{\mathrm{d}\boldsymbol{k}^\top}{\mathrm{d}\boldsymbol{x}} \boldsymbol{C}_n \boldsymbol{z}, \tag{6}$$

where the matrix $\frac{\mathrm{d}\boldsymbol{k}^\top}{\mathrm{d}\boldsymbol{x}}$ has elements,

$$\left(\tfrac{\mathrm{d}\boldsymbol{k}^\top}{\mathrm{d}\boldsymbol{x}}\right)_{i,j} = \frac{\mathrm{d}k(\boldsymbol{x}, \boldsymbol{x}_i)}{\mathrm{d}x_j}, \tag{7}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, d$.

2 Covariance Functions

2.1 Linear Covariance

name	`cov_linear_one`
formula	$k(\boldsymbol{x}, \boldsymbol{y}) = \vartheta^2(1 + \boldsymbol{x} \cdot \boldsymbol{y})$
parameters	ϑ
log-parameters	$\varphi = -\log \vartheta$
$\frac{\mathrm{d}k}{\mathrm{d}\varphi}(\vartheta)$	$-2\vartheta^2(1 + \boldsymbol{x} \cdot \boldsymbol{y})$
$\frac{\mathrm{d}k(\boldsymbol{x}, \boldsymbol{x}_i)}{\mathrm{d}x_j}$	$x_{i,j}$

2.2 Linear Covariance with Automatic Relevance Determination

name	`cov_linear_ard`
formula	$k(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{d} \frac{x_i y_i}{\lambda_i}$
parameters	$\boldsymbol{\vartheta} = (\lambda_1, \ldots, \lambda_d)$
log-parameters	$\boldsymbol{\varphi} = \log \boldsymbol{\vartheta}$
$\frac{\mathrm{d}k}{\mathrm{d}\varphi_i}(\boldsymbol{\vartheta})$	$-2\frac{x_i y_i}{\lambda_i}$
$\frac{\mathrm{d}k(\boldsymbol{x}, \boldsymbol{x}_i)}{\mathrm{d}x_j}$	$\frac{x_{i,j}}{\lambda_j}$

2.3 Squared exponential covariance function with isotropic distance measure

name	`cov_se_iso`
formula	$k(\boldsymbol{x}, \boldsymbol{y}) = \alpha^2 \exp\left(-\frac{1}{2\lambda^2}\lVert\boldsymbol{x} - \boldsymbol{y}\rVert^2\right),$
parameters	$\boldsymbol{\vartheta} = (\lambda, \alpha)$
log-parameters	$\boldsymbol{\varphi} = \log \boldsymbol{\vartheta}$
$\frac{\mathrm{d}k}{\mathrm{d}\varphi_1}(\boldsymbol{\vartheta})$	$\frac{1}{\lambda^2}\lVert\boldsymbol{x} - \boldsymbol{y}\rVert^2 k(\boldsymbol{x}, \boldsymbol{y})$
$\frac{\mathrm{d}k}{\mathrm{d}\varphi_2}(\boldsymbol{\vartheta})$	$2k(\boldsymbol{x}, \boldsymbol{y})$
$\frac{\mathrm{d}k(\boldsymbol{x}, \boldsymbol{x}_i)}{\mathrm{d}x_j}$	$\lambda^{-2}(x_{i,j} - x_j)\, k(\boldsymbol{x}, \boldsymbol{x}_i)$

2.4 Squared exponential covariance function with Automatic Relevance Determination

name	`cov_se_ard`
formula	$k(\boldsymbol{x},\boldsymbol{y}) = \alpha^2 \exp\left(-\frac{1}{2}\sum_{i=1}^{d}\frac{(x_i-y_i)^2}{\lambda_i^2}\right),$
parameters	$\boldsymbol{\vartheta} = (\lambda_1,\dots,\lambda_d,\alpha)$
log-parameters	$\boldsymbol{\varphi} = \log\boldsymbol{\vartheta}$
$\frac{\mathrm{d}k}{\mathrm{d}\varphi_i}(\boldsymbol{\vartheta})$	$\frac{1}{\lambda_i^2}\,\lVert\boldsymbol{x}-\boldsymbol{y}\rVert^2\,k(\boldsymbol{x},\boldsymbol{y}), \quad i=1,\dots,d$
$\frac{\mathrm{d}k}{\mathrm{d}\varphi_{d+1}}(\boldsymbol{\vartheta})$	$2k(\boldsymbol{x},\boldsymbol{y})$

2.5 Independent covariance function (white noise)

name	`cov_noise`
formula	$k(\boldsymbol{x},\boldsymbol{y}) = \alpha^2\delta(\lVert\boldsymbol{x}-\boldsymbol{y}\rVert)$
parameters	$\vartheta = \alpha$
log-parameters	$\varphi = \log\vartheta$
$\frac{\mathrm{d}k}{\mathrm{d}\varphi}(\vartheta)$	$2\alpha^2\delta(\lVert\boldsymbol{x}-\boldsymbol{y}\rVert)$

2.6 Periodic covariance function

name	`cov_periodic`
formula	$k(\boldsymbol{x},\boldsymbol{y}) = \sigma_f^2\exp\left(-2\lambda^{-2}\,\sin^2\frac{\pi\lVert\boldsymbol{x}-\boldsymbol{y}\rVert}{T}\right)$
parameters	$\boldsymbol{\vartheta} = (\alpha,\sigma_f,T)$
log-parameters	$\boldsymbol{\varphi} = \log\boldsymbol{\vartheta}$
$\frac{\mathrm{d}k}{\mathrm{d}\varphi_1}(\boldsymbol{\vartheta})$	$4\,k(\boldsymbol{x},\boldsymbol{y})\,\lambda^{-2}\,\sin^2\frac{\pi\lVert\boldsymbol{x}-\boldsymbol{y}\rVert}{T}$
$\frac{\mathrm{d}k}{\mathrm{d}\varphi_2}(\boldsymbol{\vartheta})$	$2\,k(\boldsymbol{x},\boldsymbol{y})$
$\frac{\mathrm{d}k}{\mathrm{d}\varphi_3}(\boldsymbol{\vartheta})$	$4\,k(\boldsymbol{x},\boldsymbol{y})\,\lambda^{-1}\beta\,\frac{\sin\beta}{\lambda}\cos\beta, \quad \beta=\frac{\pi\lVert\boldsymbol{x}-\boldsymbol{y}\rVert}{T}$