Derivatives of a Gaussian Process

Selwyn Gomes

April 8, 2021

Approach in 1-Dimension

For simplicity let us assume a 1-D function $f(\mathbf{x}) : \mathbb{R} \Rightarrow \mathbb{R}$ as our objective. The GP prior over the observations $\mathbf{y}$ at inputs $\mathbf{x}$ is defined as

$$f \sim \mathrm{GP}(\mu, \Sigma) \tag{1}$$

First, for ease of derivation let us take a zero mean ($\mu = 0$) Gaussian prior with a squared exponential kernal $k(\mathbf{x}, \mathbf{x}^{'})$. The kernal k is given as:

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

Where σ^2 and ℓ are the signal variance and lenghtscale respectively. As we already know, the derivatives of a GP is another GP which can be obtained using μ and Σ. Simply, the GP over f' would be given as follows.

$$f' \sim \mathrm{GP}\left(\frac{\partial \mu}{\partial x}, \frac{\partial \Sigma}{\partial x}\right) \tag{3}$$

where the derivative of the covariance matrix can be expressed in terms of the kernal $k(\mathbf{x}, \mathbf{x}^{'})$. Let us call the kernal function of the GP over f' as $k_{11}(\mathbf{x}, \mathbf{x}^{'})$ and is given as

$$k_{11}(\mathbf{x}, \mathbf{x}^{'}) = \mathrm{Cov}\left[\frac{\partial f(x)}{\partial x}, \frac{\partial f(x^{'})}{\partial x^{'}}\right]$$

by linearity of expectation

$$k_{11}(\mathbf{x}, \mathbf{x}^{'}) = \frac{\partial^2}{\partial x \, \partial x^{'}} \mathrm{Cov}\left[f(x), f(x^{'})\right]$$

$$k_{11}(\mathbf{x},\mathbf{x}^{'}) = \frac{\partial^2}{\partial x\,\partial x^{'}} k(\mathbf{x},\mathbf{x}^{'})$$

$$k_{11}(\mathbf{x},\mathbf{x}^{'}) = \frac{\sigma^2}{\ell^2}\left(1.0 - \frac{(\mathbf{x}-\mathbf{x}^{'})^2}{\ell^2}\right)\exp\left(-\frac{(\mathbf{x}-\mathbf{x}^{'})^2}{2\ell^2}\right) \tag{4}$$

The GP over f' will have zero mean and co-variance defined by kernal $k_{11}(\mathbf{x},\mathbf{x}^{'})$. However it is not possible to perform direct inference using this GP with only access to function values ($\mathbf{y}$,$\mathbf{x}$).To overcome this issue we create a joint distribution between $f(x)$ and $f'(x)$.This new multivariate normal distribution gives us access to the GP over the derivatives conditioned on observations $\mathbf{y}$ at inputs $\mathbf{x}$. To create the joint distribution we need to also calculate the co-variance between function and derivative values.

Let $k_{01}(\mathbf{x},\mathbf{x}^{'})$ be the co-variance between an observation at $\mathbf{x}$ and derivative at $\mathbf{x}^{'}$. Then

$$k_{01}(\mathbf{x},\mathbf{x}^{'}) = \mathrm{Cov}\left[\frac{\partial f(x^{'})}{\partial x^{'}}, f(x)\right]$$

by linearity of expectation

$$k_{01}(\mathbf{x},\mathbf{x}^{'}) = \frac{\partial}{\partial x^{'}}\mathrm{Cov}\left[f(x^{'}), f(x)\right]$$

$$k_{01}(\mathbf{x},\mathbf{x}^{'}) = \frac{\partial}{\partial x^{'}} k(\mathbf{x},\mathbf{x}^{'})$$

$$k_{01}(\mathbf{x},\mathbf{x}^{'}) = \frac{\sigma^2}{\ell^2}(\mathbf{x}-\mathbf{x}^{'})\exp\left(-\frac{(\mathbf{x}-\mathbf{x}^{'})^2}{2\ell^2}\right) \tag{5}$$

similarly, let $k_{10}(\mathbf{x},\mathbf{x}^{'})$ be the co-variance between a derivative at $\mathbf{x}$ and observation at $\mathbf{x}^{'}$. Then

$$k_{10}(\mathbf{x},\mathbf{x}^{'}) = \mathrm{Cov}\left[\frac{\partial f(x)}{\partial x}, f(x^{'})\right]$$

by linearity of expectation

$$k_{10}(\mathbf{x},\mathbf{x}^{'}) = \frac{\partial}{\partial x}\mathrm{Cov}\left[f(x), f(x^{'}x)\right]$$

$$k_{10}(\mathbf{x},\mathbf{x}^{'}) = \frac{\partial}{\partial x} k(\mathbf{x},\mathbf{x}^{'})$$

$$k_{10}(\mathbf{x},\mathbf{x}^{'}) = \frac{\sigma^2}{\ell^2}(\mathbf{x}^{'} - \mathbf{x}) \exp\left(-\frac{(\mathbf{x}-\mathbf{x}^{'})^2}{2\ell^2}\right) \tag{6}$$

The joint distribution would look like:

$$\begin{bmatrix} f(x) \\ f'(x) \end{bmatrix} \sim \mathbf{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} k(\mathbf{x},\mathbf{x}^{'}) & k_{01}(\mathbf{x},\mathbf{x}^{'}) \\ k_{10}(\mathbf{x},\mathbf{x}^{'}) & k_{11}(\mathbf{x},\mathbf{x}^{'}) \end{bmatrix}\right) \tag{7}$$

Given a set of observed function values $\mathcal{D} = \{\mathbf{x}_{1:N}, \mathbf{f}_{1:N}\}$, the rules for conditionalization of multivariate Gaussian densities allow us to derive the posterior distribution over f'. Let $\mathbf{x}^*$ denote an arbitrary collection of locations and $f^* = f'(\mathbf{x}^*)$ denote the corresponding derivative values. Then we can re-write the previous equation with new notations as

$$\begin{bmatrix} f_{1:N} \\ f^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} K & k^* \\ k^{*\top} & k_{11}(\mathbf{x}^*,\mathbf{x}^*) \end{bmatrix}\right) \tag{8}$$

where $K = k(\mathbf{x}_{1:N}, \mathbf{x}_{1:N})$ and $k^* = [k_{10}(\mathbf{x}^*,\mathbf{x}_1), k_{10}(\mathbf{x}^*,\mathbf{x}_2), ..., k_{10}(\mathbf{x}^*,\mathbf{x}_N)]^\top$. Thus we have the following posterior distribution over f^*:

$$f^*|\mathbf{x}_{1:N}, f_{1:N}, \mathbf{x}^* \sim \mathcal{N}(\mu(\mathbf{x}^*), \sigma^2(\mathbf{x}^*)) \tag{9}$$

where

$$\mu(\mathbf{x}^*) = k^{*\top} K^{-1} f_{1:N}$$

$$\sigma^2(\mathbf{x}^*) = k_{11}(\mathbf{x}^*,\mathbf{x}^*) - k^{*\top} K^{-1} k^*$$

where $\mu(\mathbf{x}^*)$ and $\sigma^2(\mathbf{x}^*)$ are posterior mean and covariance.

Approach in Higher Dimensions

Let $f(\mathbf{x})$ denote a D-dimensional function with a GP prior over the function values given by equation (1) with $\mu = 0$ and kernal:

$$k(\mathbf{x},\mathbf{x}^{'}) = \sigma^2 \exp\left(-\frac{||(\mathbf{x}-\mathbf{x}^{'})||_2^2}{2\ell^2}\right) \tag{10}$$

where instead of using difference in the 1-D case, we now use the euclidean distance. Let f'_d denote the function over the partial derivative w.r.t the d dimension where $d = \{1, 2....D\}$, then the GP over f'_d would be:

$$f'_d \sim \mathrm{GP}\left(\frac{\partial \mu}{\partial x_d}, \frac{\partial \Sigma}{\partial x_d}\right) \tag{11}$$

The kernal function for f_d' (between partial derivatives) is given as follows:

$$k_{11}(\mathbf{x},\mathbf{x}^{'}) = \frac{\partial^2}{\partial x_d \, \partial x_d^{'}} k(\mathbf{x},\mathbf{x}^{'})$$

$$k_{11}(\mathbf{x},\mathbf{x}^{'}) = \frac{\sigma^2}{\ell^2}\left(1.0 - \frac{(\mathbf{x}_d - \mathbf{x}_d^{'})^2}{\ell^2}\right)\exp\left(-\frac{||(\mathbf{x}-\mathbf{x}^{'})||_{\mathbf{2}}^{\mathbf{2}}}{2\ell^2}\right) \tag{12}$$

Let $k_{01}(\mathbf{x},\mathbf{x}^{'})$ be the co-variance between an observation at $\mathbf{x}$ and partial derivative at $\mathbf{x}^{'}$. Then

$$k_{01}(\mathbf{x},\mathbf{x}^{'}) = \frac{\partial}{\partial x_d^{'}} k(\mathbf{x},\mathbf{x}^{'})$$

$$k_{01}(\mathbf{x},\mathbf{x}^{'}) = \frac{\sigma^2}{\ell^2}(\mathbf{x}_d - \mathbf{x}_d^{'})\exp\left(-\frac{||(\mathbf{x}-\mathbf{x}^{'})||_{\mathbf{2}}^{\mathbf{2}}}{2\ell^2}\right) \tag{13}$$

similarly, let $k_{10}(\mathbf{x},\mathbf{x}^{'})$ be the co-variance between a partial derivative at $\mathbf{x}$ and observation at $\mathbf{x}^{'}$. Then

$$k_{10}(\mathbf{x},\mathbf{x}^{'}) = \frac{\partial}{\partial x_d} k(\mathbf{x},\mathbf{x}^{'})$$

$$k_{10}(\mathbf{x},\mathbf{x}^{'}) = \frac{\sigma^2}{\ell^2}(\mathbf{x}_d^{'} - \mathbf{x}_d)\exp\left(-\frac{||(\mathbf{x}-\mathbf{x}^{'})||_{\mathbf{2}}^{\mathbf{2}}}{2\ell^2}\right) \tag{14}$$

Given a set of observed function values $\mathcal{D} = \{\mathbf{X}, \mathbf{f}_{1:N}\}$, where matrix $\mathbf{X}$, of size (N,D) denotes all N training input vectors and matrix $\mathbf{f}_{1:N}$ of size (N,1) gives the function values, we need to estimate the GP prior over $f_d^{'}$ for all $d = \{1, 2....D\}$.

The rules for conditionalization of multivariate Gaussian densities allow us to derive the posterior distribution over f_d'. Let $\mathbf{x}^*$ denote an arbitrary collection of locations and $f^* = f's_d(\mathbf{x}^*)$ denote the corresponding partial derivative values. The joint process will given as

$$\begin{bmatrix} f_{1:N} \\ f^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} K & k^* \\ k^{*\top} & k_{11}(\mathbf{x}^*,\mathbf{x}^*) \end{bmatrix}\right) \tag{15}$$

where $K = k(\mathbf{X},\mathbf{X})$ and $k^* = [k_{10}(\mathbf{x}^*, X_1), k_{10}(\mathbf{x}^*, \mathbf{X}_2), ..., k_{10}(\mathbf{x}^*, \mathbf{X}_N)]^\top$. Thus we have the following posterior distribution over f^*:

$$f^*|\mathbf{X}, f_{1:N}, \mathbf{x}^* \sim \mathcal{N}(\mu_d(\mathbf{x}^*), \sigma_d^2(\mathbf{x}^*)) \tag{16}$$

where

$$\mu_d(\mathbf{x}^*) = k^{*\top} K^{-1} f_{1:N}$$
$$\sigma_d^2(\mathbf{x}^*) = k_{11}(\mathbf{x}^*, \mathbf{x}^*) - k^{*\top} K^{-1} k^*$$

where $\mu_d(\mathbf{x}^*)$ and $\sigma_d^2(\mathbf{x}^*)$ are posterior mean and covariance for $f_d^{'}(\mathbf{x}^*)$.

Additional Points

The co-variance between any two partial derivatives is given by:

$$k_{11}(\mathbf{x}_g, \mathbf{x}_h^{'}) = \frac{\sigma^2}{\ell^2}\left(\delta_{gh} - \frac{(\mathbf{x}_g - \mathbf{x}_g^{'})(\mathbf{x}_h - \mathbf{x}_h^{'})}{\ell^2}\right)\exp\left(-\frac{||(\mathbf{x} - \mathbf{x}^{'})||_{\mathbf{2}}^{\mathbf{2}}}{2\ell^2}\right) \tag{17}$$

where $\delta_{gh} = 1$ if $g = h$ and 0 otherwise.