RKHS of Gaussian processes

Debdeep Pati
Florida State University

October 18, 2016

Gaussian process: why flexible

- Realizations of a GP

$$\{g : g(x) = \sum_{k=1}^{K} w_k C(x, x_k), (x_1, \ldots, x_k) \subset \mathbb{X}, k \in \mathbb{N}, w_k \in \mathbb{R}\}$$

- Heuristics: We want to approximate an arbitrary function $f_0 : \mathbb{X} \to \mathbb{R}$. Setting $c(x, x') = \phi_\sigma(x - x')$, $w_k = f_0(x_k)$

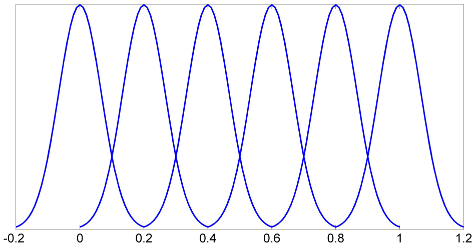
$$\sum_{k=1}^{K} w_k \phi_\sigma(x - x_k) = \sum_{k=1}^{K} f_0(x_k)\phi_\sigma(x - x_k) \approx \phi_\sigma \star f_0 \to f_0 \; as \; \sigma \to 0.$$

RKHS of Gaussian processes

- The RKHS $\mathbb{H}$ is the completion of the linear space

$$f(t) = \sum_{h=1}^{m} a_h C(s_h, t),\ s_h \in [0,1],\ a_h \in \mathbb{R}$$

- Illustration with the squared exponential kernel
  $C(s,t) = \exp(-\kappa|s-t|^2)$

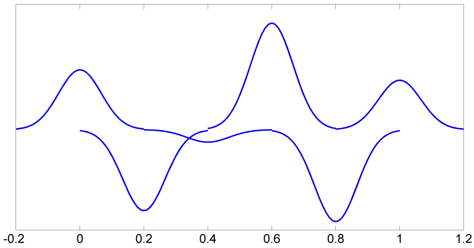


RKHS of Gaussian processes

- The RKHS $\mathbb{H}$ is the completion of the linear space

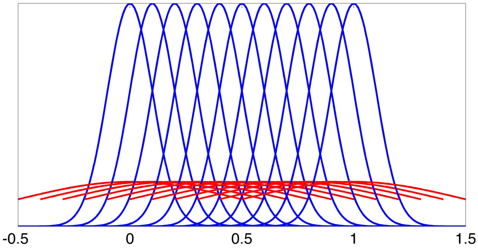
$$f(t) = \sum_{h=1}^{m} a_h C(s_h, t), \; s_h \in [0,1], \; a_h \in \mathbb{R}$$

- Illustration with the squared exponential kernel $C(s,t) = \exp(-\kappa|s-t|^2)$



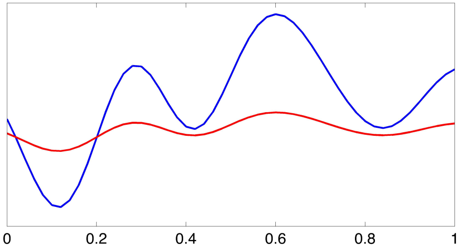
Why scaling works

- A or κ plays the role of an inverse-bandwidth
- Large A implies more peaked kernels
- Stretching the sample paths



Why scaling works

- A or κ plays the role of an inverse-bandwidth
- Large A enables approximation of rougher functions from the RKHS



- van der Vaart & van Zanten (2008): If $A^D \sim \text{gamma}(a, b)$, optimal rate of convergence adaptively over $C^\alpha[0,1]^D$ for any $\alpha > 0$

Theory for Gaussian random element

- Want mechanism to produce random (continuous) functions.
- A random vector $X : (\Omega, \mathcal{E}, P) \to \Re^k$ is Gaussian if $a^\mathrm{T} X$ is Gaussian for any $a \in \Re^k$
- Let $X : (\Omega, \mathcal{E}, P) \to (\mathcal{C}[0,1], || \cdot ||_\infty)$ be measurable
- X is called Gaussian if $L(X)$ is Gaussian for any linear functional L
- For example, $L(f) = f(1/2)$, $L(f) = 2f(1/3) - f(3/4)$, ...
- Clearly, for any $(t_1, \ldots, t_m)$, $\sum_{i=1}^m a_i X(t_i)$ is Gaussian for any $a \in \Re^m$
- $(X_{t_1}, \ldots, X_{t_m})$ is MVN

Covariance kernel approach

- Specify a joint Gaussian for $(X_{t_1}, \dots, X_{t_m})$ consistently
- Let $C(t,s)$ be a positive definite covariance kernel, i.e., $\mathbf{C} = \big(C(t_i, t_j)\big)$ is positive definite for any $t_1, \dots, t_m$
- $(X_{t_1}, \dots, X_{t_m}) \sim N(0, \mathbf{C})$, so that $C(s,t) = \mathrm{cov}(X_s, X_t)$
- Common examples: $C(t,s) = \min(t,s)$, $C(t,s) = \exp(-\kappa|t-s|)$, $C(t,s) = \exp(-\kappa|t-s|^2)$ etc

Series expansion approach

- Mercer's theorem: There exists a sequence of eigenvalues $\lambda_h \downarrow 0$ and an orthonormal system of eigenfunctions ϕ_h, such that

$$C(s,t) = \sum_{h=1}^{\infty} \lambda_h \phi_h(s) \phi_h(t)$$

- Define $\tilde{X}(t) = \sum_{h=1}^{\infty} \lambda_h^{1/2}\ Z_h\ \phi_h(t)$, where Z_h i.i.d. $\mathrm{N}(0,1)$
- $\mathrm{cov}(\tilde{X}_s, \tilde{X}_t) = \sum_{h=1}^{\infty} \lambda_h \phi_h(s) \phi_h(t) = C(s,t)$
- We can start with a series representation by choosing λ_h and ϕ_h. Different choices lead to splines, neural networks, wavelets, etc

RKHS of Gaussian processes

- In np Bayes, want priors to place positive probability around arbitrary neighborhoods of a large class of parameter values (large support property)
- The prior concentration plays a key role in determining the rate of posterior contraction
- The reproducing kernel Hilbert space (RKHS) of a Gaussian process determines the prior support and concentration
- Intuitively, a space of functions that are similar to the covariance kernel in terms of smoothness

RKHS of Gaussian processes

- Let X be a zero mean Gaussian process on $([0,1], ||_\infty)$ with covariance kernel $C(s,t) = E(X_s X_t)$
- The RKHS $\mathbb{H}$ is the completion of the linear space

$$f(t) = \sum_{h=1}^{m} a_h C(s_h, t),\ s_h \in [0,1],\ a_h \in \Re.$$

under the norm induced by the inner product,

$$\left\langle \sum_{i=1}^{k} \alpha_i C(s_i, \cdot), \sum_{j=1}^{l} \beta_j C(t_j, \cdot) \right\rangle_{\mathbb{H}} = \sum_{i=1}^{k} \sum_{j=1}^{l} \alpha_i \beta_j C(s_i, t_j)$$

- The completion can be identified with a set of functions $f : [0,1] \to \Re$ though the reproducing formula, $f(t) = \langle f, C(t, \cdot) \rangle$