STA 941: Bayesian Nonparametrics
HW 4

1. The polynomial spline kernel of smoothness $\nu \in \mathbb{N}$ is given by

$$k(s,t) = \sum_{i=0}^{\nu-1} \frac{s^i t^i}{(i!)^2} + \int_0^1 \frac{(s-u)_+^{\nu-1}(t-u)_+^{\nu-1}}{\{(\nu-1)!\}^2} du, \quad s,t, \in [0,1],$$

where $x_+ = \max(x, 0)$.

   (a) Show that k indeed is a kernel.

   (b) Show that k is the reproducing kernel for the Hilbert space $\mathcal{H}$ containing all functions on $[0,1]$ with $(\nu-1)$ continuous derivatives and an ν-th derivative with a finite L_2 norm, with inner product $(f,g)_{\mathcal{H}} = \sum_{i=0}^{\nu-1} f^{(i)}(0)g^{(i)}(0) + \int_0^1 f^{(\nu)}(u)g^{(\nu)}(u)du$.
   [Hint: Taylor's theorem in the integral form says for any f in $\mathcal{H}$, and any $t \in [0,1]$,

$$f(t) = \sum_{i=0}^{\nu-1} \frac{f^{(i)}(0)}{i!} t^i + \int_0^1 \frac{(t-u)_+^{\nu-1} f^{(\nu)}(u)}{(\nu-1)!} du.$$

   Use this, and the fact that $g(x) = x_+^p/p!$ is $(p-1)$-times continuously differentiable with $g^{(i)} = x_+^{p-i}/(p-i)!$]

   (c) Use the kernel representation theorem to show that for observations $\{(x_j, y_j) \in [0,1], \mathbb{R} : i = j, \ldots, n\}$, the minimizer $\hat{f}$ of $\mathrm{PSS}(f) = \sum_{j=1}^{n}\{y_j - f(x_j)\}^2 + \lambda \|f\|_{\mathcal{H}}^2$ is a polynomial spline of order $m = 2\nu - 1$ with knots at the observed x_j's, and, $\hat{f}^{(i)}(t) = 0$ for $t > \max(x_1, \ldots, x_n)$, for all $j = \nu, \ldots, 2\nu - 1$.
   [NOTE: This is not the natural smoothing spline solution! The penalty used above is $\|f\|_{\mathcal{H}}^2 = \sum_{i=0}^{\nu-1}\{f^{(i)}(0)\}^2 + \int_0^1 \{f^{(\nu)}(u)\}^2 du$ whereas in natural smoothing spline the penalty is only $\int_0^1 \{f^{(\nu)}(u)\}^2 du$. Also, the natural smoothing spline solution $\hat{f}$ has $\hat{f}^{(i)}(t) = 0$ for $i \geq \nu$ on both $t > \max(x_1, \ldots, x_n)$ and $t < \min(x_1, \ldots, x_n)$. This could be established by carrying out the Taylor's expansion around $\xi_1 = \min(x_1, \ldots, x_n)$ instead of 0 as was done above.]

2. Let Z_n, Z'_n, $n = 1, 2, \ldots$, be two independent sequences of independent $N(0,1)$ random variables. For each $N \in \mathbb{N}$ define the random function

$$Y^N(t) = \pi^{-1/2} \sum_{n=1}^{N} \frac{1}{n}\{Z_n \cos(2^n t) + Z'_n \sin(2^n t)\}, \; t \in [-\pi, \pi].$$

   (a) Show that there exists a centered, Gaussian element $Y \in L_2([-\pi,\pi])$ such that $\lim_{N\to\infty} \|Y^N - Y\|_{L_2} = 0$ with probability one. [Hint: show that with probability one, $(Y^N : N \geq 1)$ form a Cauchy sequence in $L_2([-\pi,\pi])$. Why is Y Gaussian?].

   (b) Calculate $\rho(s,t) = [\mathbb{E}\{Y(s) - Y(t)\}^2]^{1/2}$ and argue that $Y(t)$ is stationary.

3. A (separable) stationary Gaussian process $Y(t)$ on $t \in [0,1]$ has sample paths that are almost surely discontinuous if and only if the sample paths are almost surely unbounded ("if" is trivial, "only if" is Belyaev's theorem (Belyaev, 1961)). A use of Szidon's Lemma[1] shows that the $Y(t)$ in #2 above is unbounded with probability one – hence $Y(t)$ does not have continuous sample paths. Let us probe what happens to the sufficient condition we have seen for continuity of sample paths.

   (a) Show that for $u = \pi 2^{-M}/3$, $M \in \mathbb{N}$, $r(u) := \rho(0,u) \geq \sqrt{3/\pi}\{\log_2(\pi/(3u)+1\}^{-1/2}$.

   (b) Assuming $r(u) \geq c\{\log(1/u)\}^{-1/2}$ holds generally for all small $u \in (0,1)$ and for some constant $c > 0$, show that $\int_0^1 \{\log N(r, [-\pi,\pi], \rho)\}^{1/2} dr = \infty$.

4. Consider the regression model

$$
\begin{aligned}
y_i &= f(x_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \\
f &\sim \mathrm{GP}(0, C^{\mathrm{SE}}_{\tau,\ell})
\end{aligned}
$$

   where $C^{\mathrm{SE}}_{\tau,\ell}(s,t) = \tau^2 \exp\{-(s-t)^2/(2\ell^2)\}$ is the squared-exponential kernel with variance τ^2 and characteristic length-scale ℓ. Assume all $x_i \in [0,1]$.

   (a) Simulate and plot[2] one realization f_0 of f from the prior with $\tau = 1$, $\ell = 0.08$.

   (b) Fix $n = 20$ and generate data $\{(x_i, y_i) : i = 1, \ldots, n\}$ by drawing x_i's uniformly from $[0,1]$ and then drawing $y_i \sim N(f_0(x_i), \sigma^2)$ with $\sigma = 0.1$. Here f_0 is the function you simulated in part (a). Note that the x_i's may not be on the grid of x-values you used for simulation. How do you extend your simulation of f_0 to these new points?

   (c) Now onto estimation of f based on the data you generated in part (b). Assume $\tau = 1$ and $\sigma = 0.1$ are given, whereas ℓ and f are unknown. Suppose further we restrict ℓ to the set of values $\mathcal{L} = \{0.16/r : r \in \{0.1, 0.5, 1, 2, 3, 4, 5\}\}$. For each such value of ℓ, evaluate the log marginal likelihood $\log p(y \mid x, \ell)$ and tabulate these. How informative is the data about ℓ? Also, for each choice of ℓ make plots of the posterior mean and 95% credible band of $f(x)$ for x over the uniform grid used in part (a). Overlay the data points and the true f_0 on your plot. Comment on the qualitative differences between these plots, keeping in mind the concepts of smoothing and local learning.

   (d) Under a (discrete) uniform prior on ℓ over the finite set $\mathcal{L}$ above, visually summarize the posterior mean and 95% credible bands for f.

References

Belyaev, Y. K. (1961). Continuity and hölder?s conditions for sample functions of stationary gaussian processes. In *Proc. 4th Berkeley Symp. Math. Statist. and Prob*, Volume 2, pp. 22–33.

[1] if $\sum_n (b_n^2 + c_n^2) < \infty$, $\sum(|b_n| + |c_n|) = \infty$ and $\theta_{n+1}/\theta_n > \lambda > 1$ for all n then the $L_2[-\pi, pi]$ function $f(t) := \sum_n \{b_n \cos(\theta_n t) + c_n \sin(\theta_n t)\}$, $t \in [-\pi, \pi]$ is unbounded.

[2] Of course you will have to do this over some discrete grid of x values. You may use the uniform grid $\{0, 0.01, \ldots, 1.00\}$ of mesh size 0.01 for both simulation and graphing.