Ch 3 Nonparametric Regression and Spatial Smoothing

Tuesday, March 12, 2019 1:37 PM

(A) Suppose $y_i = f(x_i) + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$, for some unknown function f. Suppose the prior distribution for the unknown function is a mean-zero Gaussian process: $f \sim GP(0, C)$ for some covariance function C. Let $x_1, \dots, x_n$ denote the previously observed x points. Derive the posterior distribution $[f(x_1), \dots, f(x_N)]^T$ given the corresponding outcomes $y_1, \dots, y_N$, assuming that you know σ^2.

We have that $y_i = f(x_i) + \varepsilon_i$ so the distribution of y_i is $N(f(x_i), \sigma^2)$ and the distribution of $\underset{\sim}{y}$ given $f \sim MVN(f(\underset{\sim}{x}), \sigma^2 \underset{\sim}{I})$ because $f(x_{1:n}) \sim GP(0, C)$. Under this model, we can think of $f(x_i)$ kind of as a parameter of y. Therefore, we can find

$$p(f|y) \propto p(y|f)\, p(f)$$

Where $p(f|x_{1:n}) \sim MVN(0, C(x,x) = C)$ and $p(y|f) \sim MVN(f(\underset{\sim}{x}), \sigma^2 I)$. For notational ease, from here on out, x and y will be vectors.

$$p(f(x)|y) \propto \exp\left(-\tfrac{1}{2}(y - f(x))'(\sigma^2 \underset{\sim}{I})^{-1}(y - f(x)) - \tfrac{1}{2} f(x)' C^{-1} f(x)\right)$$

$$\propto \exp\left\{-\tfrac{1}{2}\left(f(x)'((\sigma^2 I)^{-1} + C^{-1}) f(x) - 2 f(x)'((\sigma^2 I)^{-1} \underset{\sim}{y})\right)\right\}$$ multiplying out, collecting like terms, and discarding constants

$$\sim N\left\{[(\sigma^2 I)^{-1} + C^{-1}]^{-1}(\sigma^2 I)^{-1}\underset{\sim}{y},\ ((\sigma^2 I)^{-1} + C^{-1})^{-1}\right\}$$ via complete the square

$$= N\left\{[\tfrac{1}{\sigma^2} I + C^{-1}]^{-1}(\tfrac{1}{\sigma^2} I) y,\ [\tfrac{1}{\sigma^2} I + C^{-1}]^{-1}\right\}$$

Another solution is to use lemma from part C and say that the joint of y & $f(x)$ is also multivariate normal. Then we can use the same argument as before to find $p(f|y)$.

$$\begin{bmatrix} y \\ f(x) \end{bmatrix} \qquad \begin{aligned} \theta &\sim N(m, V) &&\Rightarrow f(x) \sim N(0, C) \\ y(\theta) &\sim N(R\theta, \Sigma) &&\Rightarrow y(f(x)) \sim N(f(x), \sigma^2 \underset{\sim}{I}) \end{aligned}$$

b) As before, suppose we observed data $y_i = f(x_i) + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$, for $i = 1, \dots, N$. Now we wish to predict the value of the function $f(x^*)$ at some new points x^* where we haven't seen previous data. Suppose that f has a mean-zero Gaussian process prior, $f \sim GP(0, C)$. Show that the posterior mean $E\{f(x^*)|y_{1:N}\}$ is a linear smoother, and derive expressions for both the smoothing weights & the posterior variance of $f(x^*)$.

In part b of the Gaussian process exercises, we saw that when we know the true values of $f(x_{1:N})$, the joint posterior of the training and testing data was

$$\begin{bmatrix} f(x) \\ f(x^*) \end{bmatrix} \sim N\left(\begin{pmatrix} \mu(x) \\ \mu(x^*) \end{pmatrix}, \begin{bmatrix} C(x,x) & C(x,x^*) \\ C(x^*,x) & C(x^*,x^*) \end{bmatrix}\right).$$

However, now, we do not know the true values of $f(x_{1:N})$. Instead we have some noisy observed data $y_i = f(x_i) + \varepsilon_i$. Therefore, we need to amend the joint posterior to accomodate the extra variance introduced by the noise. In addition, we assume $f(x) \sim GP(0, C)$, so the distribution of y not given $f(x)$ is $y \sim (0, C(x,x) + \sigma^2 I)$. Thus, the amended joint is

Again, x and y are vectors,
$$\begin{bmatrix} y \\ f(x^*) \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} C(x,x) + \sigma^2 I & C(x,x^*) \\ C(x^*,x) & C(x^*,x^*) \end{bmatrix}\right).$$

Using the results from chapter 1 and part b of the previous section, it is easy to see that the conditional posterior mean

$$E[f(x^*)|y_{1:N}, x, x^*, \sigma^2] = \underset{1\times n}{C(x^*, x)}[\underset{n\times n}{C(x,x)} + \sigma^2 I]^{-1} \underset{n\times 1}{y}$$

$$E[f(x^*)|y_{1:N}, x, x^*, \sigma^2] = \sum_{j=1}^{n} \alpha_j y_j \text{, where } \alpha_j \text{ is the } j\text{th term of } C(x^*, x)[C(x,x)^{-1} + \sigma^2 I]^{-1}$$

Thus, the expectation of $f(x^*)$ is a weighted linear combination of y_i, aka a linear smoother. The variance of $f(x^*)$ can be similarly derived using previous results.

$$Var(f(x^*)|y_{1:N}, x, x^*, \sigma^2) = C(x^*, x^*) - C(x^*, x)[C(x,x) + \sigma^2 I]^{-1} C(x, x_*)$$

C) See code

Using $\sigma^2 = 1$, $\tau_1 = 10$, $b = 20$, and $\tau_2 = 0.0001$, I was able to generate the following result.