ICMU Mini-course on Gaussian Processes

Homework 2: Inference for Multivariate Normal Models; Gaussian Processes

Please upload your completed homework using this link by the end of the day on Thursday, February 11. Please email me if you have questions or would like to request an extension. The document name should include your last name. Scans or clear photographs of hand-written answers are acceptable if they are uploaded as a single document. If you upload multiple documents with the same name (e.g., if you wish to update a previously submitted document), we will grade the latest version.

1. Let $Y_1, \ldots, Y_n$ be a random sample of size n from the univariate normal distribution

$$Y_1, \ldots, Y_n \sim_{ind} N(\mu, \sigma^2),$$

where the population mean $\mu \in \mathbb{R}$ is unknown and the variance $\sigma^2 > 0$ is known. Next, assume that an appropriate prior probability model for the unknown mean μ is

$$\mu \sim N(m_0, \nu_0^2)$$

where m_0 and ν_0^2 are the prior mean and prior variance, respectively (these are chosen by the data analyst to model information about μ before the data is observed).

(a) Show that the posterior distribution of $\mu \mid Y_1 = y_1, \ldots, Y_n = y_n$ is univariate normal. *Hint: Recall that the posterior distribution describes our uncertainty about μ after the data is observed. Its probability density function (pdf) is,*

$$f(\mu \mid y_1, \ldots, y_n) \propto f(y_1, \ldots, y_n \mid \mu) f(\mu), \tag{1}$$

where $f(y_1, \ldots, y_n \mid \mu)$ is the likelihood (joint density of the data) and $f(\mu)$ is the prior pdf. Consider rearranging the right-hand-side to obtain the pdf of another normal distribution.

(b) Write down the expressions for the posterior mean m_n and the posterior variance ν_n^2 from part (a). Express the posterior mean m_n as a weighted average of the prior mean m_0 and the sample mean $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. If we take the posterior mean as an estimator of the unknown parameter μ, what does its form imply when the sample size n is small versus when it is large?

2. A random sample of n students is drawn from a large population, and their weights are measured. The average weight of the n sampled students is $\overline{y} = 150$ pounds. Assume the weights in the population are normally distributed with unknown mean μ and known variance of 400. Suppose your prior distribution for μ is normal with mean 180 and a variance of 1600.

(a) Write down the posterior distribution for μ as a function of n.

(b) For $n = 10$, give a 95% posterior credible interval for μ. *Hint: The endpoints of this interval are the 0.025 and the 0.975 quantiles of the posterior distribution. Numerical quantiles of the normal distribution may be obtained in R using the function qnorm().*

(c) For $n = 100$, give a 95% posterior credible interval for μ.

(d) Briefly describe how your answers in parts (b) and (c) differ.

3. Define the Gaussian process $X \sim \mathcal{GP}\left(m(t), C(t,s)\right),\ t \in [0,1]$ where $m(t) = 0$ and $C(t,s)$ is the squared exponential covariance with length-scale $\ell = 0.2$.

(a) What is the distribution of the random vector $\boldsymbol{X} = (X(0.1), X(0.5), X(0.9))^\top$? Calculate the numerical values of the mean vector and covariance matrix.

(b) Generate and plot 5 sample paths of the Gaussian process X over the domain $t \in [0, 1]$ over an equally-spaced discrete grid with $n = 200$ knots. Please include the R script and figures in your answer. *Hint: One way to sample such a realization is to compute $\boldsymbol{x} = \boldsymbol{L}\boldsymbol{g}$, where $\boldsymbol{g}$ is a random sample from the standard multivariate normal distribution $N_n(\boldsymbol{0}_n, \boldsymbol{I}_{n\times n})$ and $\boldsymbol{L}$ is the Cholesky factor of the covariance matrix $\boldsymbol{C}$ (where $\boldsymbol{C} = \boldsymbol{L}\boldsymbol{L}^T$). Adding a small jitter $\epsilon\boldsymbol{I}_{n\times n}$ (e.g., $\epsilon = 10^{-8}$) to the diagonal of $\boldsymbol{C}$ is recommended for numerical stability.*

(c) Repeat part (b) with $\ell = 0.1$.

(d) Briefly describe the difference between the plot in part (b) and (c).

4. Modify the R script `gaussian_process_realizations.R` from the course GitHub directory to sample realizations from a Gaussian process on the domain $D = [0, 1] \times [0, 1]$ with zero mean function and Matérn covariance with smoothness ν and length-scale ℓ defined as follows. Please include the R script and figures in your answer.

   (a) Set $\nu = 0.5$ and $\ell = 1$.

   (b) Set $\nu = 1.5$ and $\ell = 1$.

   (c) Briefly describe the difference between the samples in part (a) and part (b).