Mattias Villani
Department of Statistics
Stockholm University
Dept of Computer and Information Science
Linköping University

Advanced Bayesian Learning

Computer Lab 1 - GP Regression and Classification

The labs are the only examination, so you should do the labs **individually**.
You can use any programming language you prefer, but do **submit the code**.
Submit a readable report in **PDF** (no Word documents!) or a **JuPyteR notebook**

---

1. *Heteroscedastic GP regression*

(a) Consider the following heteroscedastic GP regression

$$\begin{aligned} y_i &= f(\boldsymbol{x}_i) + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2_{n,i}\right) \\ f(\boldsymbol{x}) &\sim \mathrm{GP}\left(0, k(\boldsymbol{x}, \boldsymbol{x}')\right) \\ \log \sigma^2_{n,i} &= w_0 + \boldsymbol{w}_1^T \boldsymbol{x} \end{aligned}$$

Let $\boldsymbol{X}_*$ be a matrix of test inputs and $\boldsymbol{f}_*$ the corresponding mean function values. Implement an algorithm that samples from the joint posterior

$$p(\boldsymbol{f}_*, w_0, \boldsymbol{w}_1 | \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{X}_*) = p(\boldsymbol{f}_* | w_0, \boldsymbol{w}_1, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{X}_*) p(w_0, \boldsymbol{w}_1 | \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{X}_*).$$

Use the prior $(w_0, \boldsymbol{w}_1)^T \sim N(0, \tau^2 I)$, independent of f.
[Hint: $p(\boldsymbol{f}_* | w_0, \boldsymbol{w}_1, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{X}_*)$ is really close to the formulas (2.22) to (2.24) in the GPML book. The distribution $p(w_0, \boldsymbol{w}_1 | \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{X}_*)$ is available in closed form and its expression is close to an expression in the GPML book (if you think a little ...). However, $p(w_0, \boldsymbol{w}_1 | \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{X}_*)$ is not a standard (known) distribution, and you need to use use MCMC or HMC, or something else, to simulate from it. It is OK to use a package for MCMC and HMC.]

(b) Use your code in 1a) to analyze the Lidar data (available on the course web page) with the Distance variable as the only covariate/feature in both the mean and variance. Use a squared exponential kernel for f. Set the prior hyperparameters σ_f, ℓ and τ^2 to reasonable values.

2. *Poisson GP regression.*

   (a) Consider the following Poisson GP regression for count data:

   $$\begin{aligned} y_1, \ldots, y_n | f &\overset{iid}{\sim} \operatorname{Pois}\left(\exp(f(\boldsymbol{x}))\right) \\ f(\boldsymbol{x}) &\sim \operatorname{GP}\left(0, k_\theta(\boldsymbol{x}, \boldsymbol{x}')\right) \end{aligned}$$

   Derive the Laplace approximation of the posterior of f on the training data $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{X})$.

   (b) Derive the (saddlepoint) approximation of the log marginal likelihood for this model.

3. BONUS QUESTION, ONLY IF YOU HAVE THE TIME AND ENERGY. A more general heteroscedastic GP regression can be of the form

   $$\begin{aligned} y_i &= f(\boldsymbol{x}_i) + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \exp(g(\boldsymbol{x}_i))\right) \\ f(\boldsymbol{x}) &\sim \operatorname{GP}\left(0, k_{\theta_f}(\boldsymbol{x}, \boldsymbol{x}')\right) \\ g(\boldsymbol{x}) &\sim \operatorname{GP}\left(0, k_{\theta_g}(\boldsymbol{x}, \boldsymbol{x}')\right) \end{aligned}$$

   where θ_f are kernel hyperparameter for the mean function $f(\boldsymbol{x})$ and θ_g are hyperparameters for the (log) variance function $g(\boldsymbol{x})$. Note that noise is heteroscedastic and modelled nonparametrically using an additional GP. Let $\boldsymbol{X}_*$ be a matrix of test inputs and $\boldsymbol{f}_*$ the corresponding mean function values, just as in the GPML book. Moreover, let $\boldsymbol{g}_*$ be the corresponding (log) variances at the test inputs, and finally let $\boldsymbol{g}$ be the vector of (log) variances at the training inputs. Derive the joint posterior distribution of $\boldsymbol{f}_*$, $\boldsymbol{g}_*$ and $\boldsymbol{g}$ given some training data $\boldsymbol{y}$ and $\boldsymbol{X}$ in the form of the following decomposition

   $$p(\boldsymbol{f}_*, \boldsymbol{g}_*, \boldsymbol{g} | \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_*) = p(\boldsymbol{f}_* | \boldsymbol{g}_*, \boldsymbol{g}, \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_*) p(\boldsymbol{g}_* | \boldsymbol{g}, \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_*) p(\boldsymbol{g} | \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_*).$$

   [Hint: $p(\boldsymbol{f}_* | \boldsymbol{g}_*, \boldsymbol{g}, \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_*)$ is a standard distribution, and so is $p(\boldsymbol{g}_* | \boldsymbol{g}, \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_*)$. The last factor, $p(\boldsymbol{g} | \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_*)$ can be derived in closed form, but will not be a standard distribution].

HAVE FUN!