Mattias Villani
Department of Statistics
Stockholm 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!)
- **JuPyteR/Quarto notebook compiled to PDF/HTML**

---

1. *Heteroscedastic GP regression*

   (a) Consider the following heteroscedastic GP regression

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

   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:

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

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.

HAVE FUN!