Assignment 3 Question 2 MAST90125: Bayesian Statistical Learning

Due: Friday 25 October 2019

There are places in this assignment where R code will be required. Therefore set the random seed so assignment is reproducible.

```
set.seed(234567)  #Please change random seed to your student id number.
```

Question Two: (13 marks)

Researchers are interested in phone call duration. The available data consisted of 100 standardised phone call durations, and which hour ($t = 1, \dots, 10$) the call was initiated. The researchers assumed the following model,

$$\begin{aligned} p(y_i|\mu(t)) &= \mathcal{N}(\mu(t), \sigma^2) \\ p(\mu(t)) &= \mathcal{N}(0, k(t,t)) \end{aligned}$$

such that the covariance function $k(x, x')$ is squared exponential,

$$k(x, x') = \sigma_K^2 e^{-l \times (x-x')^2},$$

with σ_K^2 fixed to 1.85 and l fixed to 0.05. The data can be downloaded from LMS as `call.csv`.

a) The researchers are interested in making predictions of phone call duration, $\tilde{\boldsymbol{\mu}}(t)$ for hours $t = 1, \dots, 12$. As an initial step, they assumed σ^2 was 0.95. Assume the data $\mathbf{y}$ is conditioned on a single realisation of the Gaussian process prior. Do the following:

- Based on the information provided, determine the joint distribution of data $\mathbf{y}$ and predictions $\tilde{\boldsymbol{\mu}}(t)$.
- Determine the distribution of $\tilde{\boldsymbol{\mu}}(t)$ conditional on $\mathbf{y}$, σ^2, σ_K^2 and l. Show working
- Plot the predictions with 95 % and 99 % credible intervals along with the observed data. Comment, in a Bayesian language, on the behaviour of predictions where no data was observed.

Answer: The joint distribution of the data $\boldsymbol{y}$ and predictions is $\mu(\tilde{\mathbf{x}})$ is For lecture it was shown that

$$p\begin{pmatrix} \mathbf{y} \\ \mu(\tilde{\mathbf{x}}) \end{pmatrix} = \mathcal{N}\left(\begin{pmatrix} \boldsymbol{m}(\mathbf{x}) \\ \boldsymbol{m}(\tilde{\mathbf{x}}) \end{pmatrix}, \begin{pmatrix} \boldsymbol{k}(\mathbf{x},\mathbf{x}) + \boldsymbol{\Sigma} & \boldsymbol{k}(\mathbf{x},\tilde{\mathbf{x}}) \\ \boldsymbol{k}(\tilde{\mathbf{x}},\mathbf{x}) & \boldsymbol{k}(\tilde{\mathbf{x}},\tilde{\mathbf{x}}) \end{pmatrix}\right)$$

For this question joint distribution of the data $\mathbf{y}$ and predictions is $\mu(\tilde{\mathbf{t}})$ is

For lecture it was shown that

$$p\begin{pmatrix} \mathbf{y} \\ \mu(\tilde{\mathbf{t}}) \end{pmatrix} = \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{k}(\mathbf{t},\mathbf{t}) + \boldsymbol{\Sigma} & \boldsymbol{k}(\mathbf{t},\tilde{\mathbf{t}}) \\ \boldsymbol{k}(\tilde{\mathbf{t}},\mathbf{t}) & \boldsymbol{k}(\tilde{\mathbf{t}},\tilde{\mathbf{t}}) \end{pmatrix}\right)$$

where $k\left(\mathbf{t}, \tilde{\mathbf{t}}\right) = \sigma_K^2 e^{-l \times \left(\mathbf{t}-\tilde{\mathbf{t}}\right)^2}$ and $\Sigma = \sigma^2 \boldsymbol{I}$

Now we will find $\mu(\tilde{\mathbf{t}})$ conditional on $\mathbf{y}, \sigma^2, \sigma_K^2, l$.

For lectures, the noisless case had the follow distribution.

$$\boldsymbol{\mu}(\tilde{\mathbf{t}})|. \sim \mathcal{N}\big(\boldsymbol{m}(\tilde{\mathbf{t}}) + \boldsymbol{k}(\tilde{\mathbf{t}}, \mathbf{t})\boldsymbol{k}(\mathbf{t}, \mathbf{t}) - \mathbf{m}(\mathbf{t})), \boldsymbol{k}(\tilde{\mathbf{t}}, \tilde{\mathbf{t}}) - \boldsymbol{k}(\tilde{\mathbf{t}}, \mathbf{t})\boldsymbol{k}(\mathbf{t}, \mathbf{t})^{-1}\boldsymbol{k}(\mathbf{t}, \tilde{\mathbf{t}})\big)$$

for this question $\boldsymbol{m}(\tilde{\mathbf{t}}) = \mathbf{0}$, we must add Σ to the first element of covariance block matrix. In other words $\boldsymbol{k}(\mathbf{t}, \mathbf{t}) \rightarrow \boldsymbol{k}(\mathbf{t}, \mathbf{t}) + \Sigma$. Then we have that following

$$\boldsymbol{\mu}(\tilde{\mathbf{t}})|\mathbf{y}, \sigma^2, \sigma_K^2, l \sim \mathcal{N}(\boldsymbol{k}(\tilde{\mathbf{t}}, \mathbf{t}) \left[\boldsymbol{k}(\mathbf{t}, \mathbf{t}) + \Sigma\right]), \boldsymbol{k}(\tilde{\mathbf{t}}, \tilde{\mathbf{t}}) - \boldsymbol{k}(\tilde{\mathbf{t}}, \mathbf{t}) \left[\boldsymbol{k}(\mathbf{t}, \mathbf{t}) + \Sigma\right]^{-1} \boldsymbol{k}(\mathbf{t}, \tilde{\mathbf{t}})\Big)$$

b) Was the inference performed in part a) fully Bayesian? If not, how would you make the analysis fully Bayesian, noting any particular difficulties that arise by assuming a Gaussian process prior. Your answer should not exceed one page of writing.