4 Exercises, week 4

The exercise solutions have to be returned at the latest on Sunday April 9'th.

- Pen and paper exercises: you can scan the solutions and combile them into pdf or write them with Latex/word/... Compile all the answers into one pdf file
- Computer exercises: Report the answers to no-coding parts of exercises (if any) in pdf and compile with answers to pen and paper exercises. Additionally, send also the code used to solve the exercises. Note!
  - Only code should be returned. **Do not send data files**!
  - Write and comment the code so that it can be run by using your code only and the data provided in the course web pages.
  - If the lecturer cannot understand or run your code you will not get points from coding part even if the results were correct.
- zip all files into one folder to reduce the size of submission.

Send the zipped files to jarno.vanhatalo@helsinki.fi.

For basic properties and results concerning Gaussian distributions and processes see e.g.
`https://en.wikipedia.org/wiki/Multivariate_normal_distribution`
`http://www.gaussianprocess.org/gpml/chapters/`

4.1 Priors for transformed parameters and sampling for restricted parameters, 2 points

Consider you have a parameter $\theta \in \Re_+$ and you want to give it a prior so that

$$1/\theta \sim \text{Student}-t_+(\nu = 4, \mu = 0, s2 = 1). \tag{13}$$

That is, you want to define a prior implicitly so that the prior for the inverse of the parameter follows Student-t distribution. Solve the probability density function for θ. Write a pseudo code on how would you implemet this distribution in STAN.

Consider you have a parameter $\theta \in \Re_+$ with a posterior distribution $p(\theta|y)$ that is not analytically trackable. Hence, you want to use MCMC to sample from $p(\theta|y)$. However, you have a device to sample only from distributions that are defined in $\Re$. Hence, you want to make a tranformation $w = log(\theta)$ and sample from the distribution of $p(w|y)$. Solve the posterior distribution $p(w|\theta)$. Write a pseudo code on how would you implement the sampling from $p(\theta|y)$ in STAN using this result.

Hint! You may use the results concerning the dimension preserving transformation of random variables. Let $p_u(u)$ be the probability density of the vector u. We transform to $v = f(u)$

where v has the same number of components as u. If $p_u(u)$ is a continuous distribution and $v = f(u)$ is a one-to-one transformation, then the joint density of the transformed vector is

$$p_v(v) = |J| p_u(f^{-1}(v))$$

where $|J|$ is the determinant of the Jacobian of the transformation $u = f^{-1}(v)$.

4.2 Computer: Additive Gaussian process, 3 points

In this exercise we will analyse the Mauna Loa Co2 data[1] discussed also by Rasmussen and Williams (2006, Chapter 5). Here we will implement three different models which will later be used to demonstrate model comparison.

Load the Mauna Loa data from course web page. We will analyze the data with the model

$$y(t_i) = f(t_i) + \epsilon_i \tag{14}$$

where the error terms are i.i.d. Gaussian distributed, $\epsilon_i \sim N(0, \sigma_\epsilon)$ and the latent function follows one of the three alternatives

$$\begin{aligned}
&1) \quad f(t) \sim GP(0, k_{\text{exp}}(t, t')) \\
&2) \quad f(t) \sim GP(0, k_{\text{exp}}(t, t') + k_{\text{sexp}}(t, t')) \\
&3) \quad f(t) \sim GP(\alpha + t\beta, k_{\text{exp}}(t, t') + k_{\text{sexp}}(t, t'))
\end{aligned}$$

where $\alpha \sim N(0, \sigma_\alpha^2)$ and $\beta \sim N(0, \sigma_\beta^2)$, $k_{\text{exp}}(\cdot, \cdot)$ is the exponential covariance function and $k_{\text{sexp}}(\cdot, \cdot)$ is the squared exponential covariance function. In model 2) the assumption is that the exponential covariance function corresponds to a GP with short (monthly) time scale changes whereas the squared exponential corresponds to a GP with long term (decadal) changes. In model 3) the linear model for the mean can be transformed to covariance function (see lecture notes) and the linear part is assumed to model any trends in the data. The priors for the model parameters should be weakly informative so that

$$\begin{aligned}
\sigma_\epsilon^2 &\sim \text{Student}-t(\nu = 1, \mu = 0, s = 0.1) \\
\sigma_{\text{exp}}^2 &\sim \text{Student}-t(\nu = 1, \mu = 0, s = 1) \\
\sigma_{\text{sexp}}^2 &\sim \text{Student}-t(\nu = 1, \mu = 0, s = 1) \\
l_{\text{exp}} &\sim \text{Student}-t(\nu = 4, \mu = 0, s = 1) \\
1/l_{\text{sexp}} &\sim \text{Student}-t(\nu = 4, \mu = 0, s = 1)
\end{aligned}$$

The variances of the intercept and linear term should be fixed to a large number, for example, $\sigma_\alpha^2 = \sigma_\beta^2 = 10$, so that the linear model corresponds to "fixed effects".

Notice that the prior for the length-scale of the squared exponential is given through its inverse. The reason for this is that we want *a priori* to favor squared exponential to capture long term changes (=long length-scale). There is no direct way to implement prior for inverse of a parameter in STAN. Hence, you either need to define the model using the inverse length-scale or solve the induced prior for l_{sexp} (see exercise 4.1).

Your tasks are the following:

[1] `http://cdiac.esd.ornl.gov/ftp/trends/co2/maunaloa.co2`

a) Implement each of the models 1-3) and infer the posterior distribution of their hyperparameters with STAN, conditional to the full data. Report the posterior distributions of the hyperparameters as histograms and plot the predicted function $f(t)$. (1 point)

b) With models 2-3) calculate the posterior predictive distribution of the alternative model components. That is in 2) visualize the posterior distribution of the function related to the exponential and squared exponential parts (e.g. plot the mean and 95% credible interval). In 3) plot the linear part $a + t\beta$ and the squared exponential and the exponential covariance function parts. In addition, calculate the posterior of α and β (see the solutions to exercises of week 3). (1 point)

c) Redo the posterior fitting of part a) with data only from years 1958-1970 and 1980-2000 and use the model to predict the whole time span from 1958-2008. Plot the prediction and data to the same figure. What are the differences between the models? Which model looks most reasonable from the predictive point of view? (1 point)

Hint! You should use your earlier code or earlier exercise results as the base code to solve this exercise. Additionally, see Section 3.4 from the lecture notes and earlier exercises for hints related to additive GPs and linear model in GP framework.