Metamodeling – Lab 1
Gaussian process regression

INSA Toulouse - 2021

For this lab session, you will use the *R* language using *RStudio* editor. We give here several valid covariance functions, or kernels, on $\mathbb{R} \times \mathbb{R}$:

$$
\begin{aligned}
\text{squared exp.} \quad & k(x,y) = \sigma^2 \exp\left(-\frac{(x-y)^2}{2\theta^2}\right) \\
\text{Matérn } 5/2 \quad & k(x,y) = \sigma^2 \left(1 + \frac{\sqrt{5}|x-y|}{\theta} + \frac{5|x-y|^2}{3\theta^2}\right) \exp\left(-\frac{\sqrt{5}|x-y|}{\theta}\right) \\
\text{Matérn } 3/2 \quad & k(x,y) = \sigma^2 \left(1 + \frac{\sqrt{3}|x-y|}{\theta}\right) \exp\left(-\frac{\sqrt{3}|x-y|}{\theta}\right) \\
\text{exponential} \quad & k(x,y) = \sigma^2 \exp\left(-\frac{|x-y|}{\theta}\right) \\
\text{Brownian} \quad & k(x,y) = \sigma^2 \min(x,y) \\
\text{white noise} \quad & k(x,y) = \sigma^2 \delta_{x,y} \\
\text{constant} \quad & k(x,y) = \sigma^2 \\
\text{linear} \quad & k(x,y) = \sigma^2 xy \\
\text{cosine} \quad & k(x,y) = \sigma^2 \cos\left(\frac{x-y}{\theta}\right) \\
\text{sinc} \quad & k(x,y) = \sigma^2 \frac{\theta}{x-y} \sin\left(\frac{x-y}{\theta}\right)
\end{aligned}
$$

Sampling from a GP

1. The script `kernFun.R` contains the implementations of the following type of kernels: linear (`linKern`), cosine (`cosKern`), and exponential (`expKern`). Each function takes as input the vectors `x`, `y` and `param` and that returns the matrix with general term $k(x_i, y_j)$. Using a similar structure, implement the functions for the Matérn 5/2 (`mat5_2Kern`) kernel.

2. Create a design of experiments X as a regular sequence of $n = 100$ points on $[0,1]$, and compute the covariance matrix K at X for the Matérn 5/2 kernel. How can you simulate zero-mean Gaussian samples based on this matrix? The function `mvrnorm()` from package MASS can be useful here.

3. What is the interpretation of θ? Hint: Consider the change of variable $x \mapsto x/\theta$. Check your proposition by drawing sample paths for different values of θ.

4. Choose $\theta = 1/2$ (for instance). Compare the sample paths obtained with the kernels Matérn ν, with $\nu = +\infty$ (squared exponential kernel), $\nu = 5/2$, $\nu = 3/2$ and $\nu = 1/2$ (exponential kernel). What is the interpretation of ν?

Gaussian process regression

From now on, let us choose the Matérn 5/2 kernel. We want to approximate the test function

$$f : \ x \in [0,1] \mapsto x + \sin(4\pi x) \tag{1}$$

with Gaussian process regression. In the no-trend case ($\mu = 0$, 'simple kriging'), the conditional mean and covariance are given by:

$$\begin{aligned} m(x) &= k(x,X)k(X,X)^{-1}Y \\ c(x,x') &= k(x,x') - k(x,X)k(X,X)^{-1}k(X,x') \end{aligned}$$

5. Create a design of experiments X composed of 15 points in the input space (regularly spaced for instance) and compute the vector of observations $Y = f(X)$.

6. Write two functions `condMean` and `condCov` that return the conditional mean and covariance. These functions take as inputs the scalar/vector of prediction point(s) `x`, the DoE vector `X`, the vector of responses `Y`, a kernel function `kern`, and the vector `param`.

7. Draw on the same graph $f(x)$, $m(x)$ and 95% prediction intervals: $m(x) \pm 1.96\sqrt{c(x,x)}$.

8. Generate samples from the conditional process.

9. What can you say about the size of the prediction intervals at x? What's happening when we replace Y by another vector Y'?

10. Draw the conditional mean for the Brownian kernel. Conclusion?

Making new from old (bonus)

Implement a kernel such that the sample paths are smooth and odd functions (i.e. such that $f(x) = -f(-x)$ for all $x \in \mathbb{R}$). How does it improve the approximation on the test function 1 on the interval $[-1,1]$? (by using the same design points X as before)