Lab 3 – Gaussian Process Regression

Short course on Statistical modelling for optimization

N. Durrande - J.C. Croix, Universidad Tecnológica de Pereira, 2017

The aim of this lab session is to obtain the best possible GPR model for the data that has been obtained with the catapult simulator.

GPR with GPy

GPy is a python package for Gaussian process models. If you have not already installed it on your computer, we advise that you download the developers version on github `https://github.com/SheffieldML/GPy/tree/devel` and follow the instructions.

Q1. Import the data (design of experiments and outputs from the simulator) you have generated.

Q2. The code for creating and optimizing a basic GP model is already given in the python script. Read it carefully to understand each line signification.

Q3. Compute the IMSE associated to the model based on a test set of 1000 points. What do you think about the model accuracy?

Q4. Plot the leave-one-out predictions against the actual observations. What do you think about the model accuracy?

Q5. Implement a function that returns the Q^2 criterion and compute its value on the LOO predictions. What component of the model are we testing here?

Q6. Compute the standardized LOO residuals and that compare them to the $\mathcal{N}(0,1)$ distribution. What does this test allows us to asses?

Q7. Try various models and select the best one. When building the models, you may consider changing:

- the kernel (try various ones and sums of kernels)
- the way kernel parameters are estimated (optimization staring point, boundaries, ...)
- possible rotations of the input space