Lab 3 – Gaussian Process Regression

Short course on Statistical modelling for optimization

N. Durrande - J.C. Croix, Universidad Tecnológica de Pereira, 2016

The aim of this lab session is to obtain the best possible GPR model for the data that has been collected ~~yesterday~~ during lunch time.

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 you have generated and choose your favourite parametrization. You may also rescale the data.

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. Since you have two (or three) observations for each design point, does-it make sense to use leave-one-out to asses the prediction quality ? Modify the `leaveOneOut` function accordingly.

Q4. Write a function that computes the standardised LOO residuals and that compare them to the $\mathcal{N}(0,1)$ distribution.

Q5. Compute the Q^2 and look at the standardized residuals of your first model. Is it convincing ?

Q6. 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, ...)
- the way you take the noise into account (fixed, estimated)
- ...