
Lab 3 – Gaussian Process Regression

Short course on Statistical modelling for optimization

N. Durrande, Universidad Tecnológica de Pereira, 2015

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

1 Models 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 of github: <https://github.com/SheffieldML/GPy/tree/devel> (there is a link 'Download ZIP' on the right). The installation steps are: 1. unzip the file; 2. Open a terminal (for example the Anaconda terminal) and go to the unzipped folder; 3. Run the command `python setup.py install`. You should then be able to import the GPy library.

Q1. Import the data you have generated yesterday. If your data is in a csv file where the first 4 columns are the inputs and the last ones the outputs, you should just have to change the csv file name in the script.

Q2. Write a function that takes a model as input and that returns the leave-one-out predicted values and their variance.

Q3. 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 (starting point for optim, boundaries, ...)
- the way you take the noise into account (fixed, estimated)
- ...

Regarding the choice of the best model, you should consider at least the Q^2 criterion and an histogram of standardized residuals.