Exercise Day 2: Data-Driven (Multi-Objective) Bayesian Optimisation

Tinkle Chugh

The aim of this workshop is to familiarise you with Gaussian processes through two main tasks. The first task involves calculating a covariance matrix analytically (using pen and paper). The second task focuses on applying a Gaussian process regression model to a real-world dataset.

1. Given the kernel function

$$k(x, x') = \sigma_f^2 \exp\left(-0.5\frac{(x - x')^2}{l^2}\right)$$

with hyperparameters $l = 2$ and $\sigma_f = 1$, compute the covariance matrix for the following three data points: $x_1 = 1$, $x_2 = 2$, and $x_3 = 4$.

2. **Regression on a Real-World Dataset:**
You have been provided with a dataset containing average atmospheric CO_2 concentrations over time. Use Gaussian process regression to predict CO_2 concentrations based on the year. Because this is a one-dimensional dataset, you should easily be able to visualise your model's predictions over time.

Experiment with selecting an appropriate kernel for this specific task. Try to compare the performance of your Gaussian process model against other machine learning techniques, such as neural networks and random forests.

You may use any library of your choice for the Gaussian process implementation. Popular options include `sklearn`, `GPy`, `GPflow`, or `GPyTorch` in Python, and `gpfit` in MATLAB.