

Exercise 1: Gaussian Processes

Assume your data follows the following law:

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}),$$

with $\mathbf{f} = f(\mathbf{x}) \in \mathbb{R}^n$ being a realization of a Gaussian process (GP), for which we a priori assume

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')).$$

\mathbf{x} here only consists of 1 feature that is observed for n data points.

- (a) Derive / define the prior distribution of \mathbf{f} .
- (b) Derive the posterior distribution $\mathbf{f}|\mathbf{y}$.
- (c) Derive the posterior predictive distribution $y_*|x_*, \mathbf{x}, \mathbf{y}$ for a new sample x_* from the same data-generating process.
- (d) Implement the GP with squared exponential kernel, zero mean function and $\ell = 1$ from scratch for $n = 2$ observations (\mathbf{y}, \mathbf{x}) . Do this as efficiently as possible by explicitly calculating all expensive computations by hand. Do the same for the posterior predictive distribution of y_* . Test your implementation using simulated data.