Exercise 1: Gaussian Processes

Assume your data follows the following law:

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

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

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

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

(a) Derive / define the prior distribution of $\boldsymbol{f}$.

(b) Derive the posterior distribution $\boldsymbol{f}|\boldsymbol{y}$.

(c) Derive the posterior predictive distribution $y_*|x_*, \boldsymbol{x}, \boldsymbol{y}$ 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 $(\boldsymbol{y}, \boldsymbol{x})$ and $\boldsymbol{\Psi} = \boldsymbol{I}$. Do this as efficiently as possible by explicitly calculating all expensive computations by hand. Do the same for the posterior predicitve distribution of y_*. Test your implementation using simulated data.