Exercise of Supervised Learning: Gaussian Processes 1

Yawei Li

yawei.li@stat.uni-muenchen.de

January 21, 2025

Exercise 1: Bayesian Linear Model

In the Bayesian linear model, we assume that the data follows the following law:

$$y = f(\mathbf{x}) + \epsilon = \boldsymbol{\theta}^\top \mathbf{x} + \epsilon,$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and independent of $\mathbf{x}$. On the data-level this corresponds to

$$y^{(i)} = f\left(\mathbf{x}^{(i)}\right) + \epsilon^{(i)} = \boldsymbol{\theta}^\top \mathbf{x}^{(i)} + \epsilon^{(i)}, \quad \text{for } i \in [n],$$

where $\epsilon^{(i)} \in \mathcal{N}(0, \sigma^2)$ are i.i.d. and all independent of $\mathbf{x}^{(i)}$'s. In the Bayesian perspective it is assumed that the parameter vector $\boldsymbol{\theta}$ is stochastic and follows a distribution. Assume we are interested in the so-called maximum a posteriori estimate of $\boldsymbol{\theta}$, which is defined by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} | \mathbf{X}, \mathbf{y}).$$

(a) Show that if we choose a **uniform distribution** over the parameter vector $\boldsymbol{\theta}$ as the prior belief, i.e., $q(\boldsymbol{\theta}) \propto 1$, then the maximum a posteriori estimate coincides with the **empirical risk minimizer for the L2-loss** (over linear models).

Exercise 1(b)

Show that if we choose a **Gaussian distribution** over the parameter vectors $\boldsymbol{\theta}$ as the prior belief, i.e.,

$$q(\boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2\tau^2}\boldsymbol{\theta}^\top\boldsymbol{\theta}\right], \quad \tau > 0,$$

then the maximum a posteriori eestimate coincides for a specific choice of τ with the **regularized** empirical risk miminizer for the L2-loss with L2 penalty (over the linear models), i.e., the Ridge regression.

Exercise 1(c)

Show that if we choose a **Laplace distribution** over the parameter vectors $\boldsymbol{\theta}$ as the prior belief, i.e.,

$$q(\boldsymbol{\theta}) \propto \exp\left[-\frac{\sum_i^p |\boldsymbol{\theta}_i|}{\tau}\right], \quad \tau > 0,$$

then the maximum a posteriori estimate coincides for a specific choice of τ with the regularized empirical risk minimizer for the L2-loss with L1 penalty (over the linear models), i.e., the Lasso regression.

Exercise 2: Covariance Functions

Consider the commonly used covariance functions mentioned in the lecture slides: constant, linear, polynomial, squared exponential, Matern, exponential covariance functions.

(a) Show that they are valid covariance functions. (**Proofs for Matern and exp. cov. functions are out of scope and omitted.**) You may use the following composition rules. In these rules we assume that $k_0(\cdot,\cdot)$ and $k_1(\cdot,\cdot)$ are valid covariance functions.

1. $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{x}'$ is a valid covariance function;
2. $k(\mathbf{x}, \mathbf{x}') = c \cdot k_0(\mathbf{x}, \mathbf{x}')$ is a valid covariance function if $c \geq 0$ is constant.
3. $k(\mathbf{x}, \mathbf{x}') = k_0(\mathbf{x}, \mathbf{x}') + k_1(\mathbf{x}, \mathbf{x}')$ is a valid covariance function;
4. $k(\mathbf{x}, \mathbf{x}') = k_0(\mathbf{x}, \mathbf{x}') \cdot k_1(\mathbf{x}, \mathbf{x}')$ is a valid covariance function;
5. $k(\mathbf{x}, \mathbf{x}') = g(k_0(\mathbf{x}, \mathbf{x}'))$ is a valid cov. func. if g is a polynomial function with **pos.** coefficients;
6. $k(\mathbf{x}, \mathbf{x}') = t(\mathbf{x}) \cdot k_0(\mathbf{x}, \mathbf{x}') \cdot t(\mathbf{x}')$ is a valid covariance function, where t is any function;
7. $k(\mathbf{x}, \mathbf{x}') = \exp(k_1(\mathbf{x}, \mathbf{x}'))$ is a valid covariance function;
8. $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{A} \mathbf{x}'$ is a valid covariance function if $\mathbf{A} \succeq 0$.

Exercise 2 (b)

(b): Are these covariance functions stationary or isotropic? Justify your answer.