Probabilistic Machine Learning

Exercise Sheet #5

1. **EXAMple: Kernels**
   Remember that a symmetric $N \times N$ matrix K is positive definite if and only if $\boldsymbol{v}^T K \boldsymbol{v} > 0$ for all vectors $\boldsymbol{v} \in \mathbb{R}^N \setminus \mathbf{0}$.

   (a) For $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^D$, prove or disprove that the function $k(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a}^\intercal \boldsymbol{b}$ is a Mercer (i.e. positive definite) kernel.

   (b) Suppose k is a Mercer kernel. Prove or disprove that $\alpha^2 k$ is also a Mercer kernel, for $\alpha \in \mathbb{R}$

   (c) Suppose k and h are Mercer kernels. Prove or disprove that $k + h$ is also a Mercer kernel.

2. **Theory Question: Parametric and nonparametric regression**
   Consider the random function $f : \mathbb{R} \rightarrow \mathbb{R}$ from the centered (i.e. zero-mean) GP prior $p(f) = \mathcal{GP}(f; 0, k)$ with the "square-exponential" kernel

$$k(a, b) = \theta^2 \exp\left(-\frac{(a-b)^2}{2\lambda^2}\right) \quad \text{with parameters } \theta, \lambda \in \mathbb{R}_+,$$

   and, independently, the random parametric function $g(x) = \boldsymbol{\phi}_x^\intercal \boldsymbol{w}$ with $\boldsymbol{\phi}_x, \boldsymbol{w} \in \mathbb{R}^F$ induced by the prior $p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}; \boldsymbol{\mu}, \Sigma)$.

   (a) What is the prior distribution over the function $h(x) = f(x) + g(x)$?

   (b) Consider observations $\boldsymbol{y} \in \mathbb{R}^N$ observed with the likelihood $p(\boldsymbol{y} \mid h) = \mathcal{N}(\boldsymbol{y}; h_X, \sigma^2 I)$ at locations $X = [x_1, \dots, x_n]$. Given these observations, what is the posterior distribution over $\boldsymbol{w}$?

3. **Practical Question**

   Last time you created a simple probabilistic model to estimate the mortality rate of COVID-19 given time-series data. In this week's exercise you will estimate the number of people that will be infected in the United States over time. You will build a probabilistic version of a SIR model and fit it to infection data. Inference will be performed using a probabilistic programming language. On ILIAS you can find a `jupyter` notebook that contains the detailed instructions.

   Note: Although the exercise uses real data, it does not aspire to satisfy the standards of epidemiology or public policy making. It is deliberately simple and designed to be feasible within the scope of this lecture course. Do not mistake it for a thorough scientific analysis, and do not draw overly confident conclusions from it.