Probabilistic Machine Learning

Exercise Sheet #5

GP regression

1. **Exam-Type Question — Linear Splines** Consider an unknown function $f : \mathbb{R}_+ \to \mathbb{R}$ assumed to be drawn from the Gaussian process prior $p(f) = \mathcal{GP}(f; 0, k)$ with the (Wiener process) kernel $k(a, b) = \min(a, b)$. Assume that you are given two evaluations $\boldsymbol{y} = [y_1, y_2$ at $X = [x_1 = 1, x_2 = 2]$, measured according to the likelihood $p(\boldsymbol{y} \mid f) = \mathcal{N}(\boldsymbol{y}; f_X, \sigma^2 I)$.

   (a) The posterior arising from this generative model is also a Gaussian process. What, in the usual shorthand with the kernel Gram matrix k_{XX}, is its mean and covariance function?

   (b) Now Write down the *explicit* form of the posterior mean and and covariance functions, by filling in the entries of k_{XX} and explicitly inverting it. Hint: The inverse of a 2×2 matrix is given by

   $$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ab - bc} \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$$

   In doing so, show that the posterior mean is a piecewise *linear spline*.

2. **Theory Question — Stochastic Differential Equations** In Lecture 12, the LTI stochastic differential equation

   $$dx(t) = Fx\,dt + L\,d\omega \qquad \text{with} \qquad x(t_0) = x_0, F \in \mathbb{R}^{d \times d}, L \in \mathbb{R}^d \tag{1}$$

   was defined as a reformulation of the Gaussian process over the function $x : t \in \mathbb{R} \mapsto x(t) \in \mathbb{R}^d$ with mean and covariance (i.e. kernel) function

   $$m(t) = e^{F(t-t_0)} x_0 \qquad k(t_a, t_b) = \int_{t_0}^{\min(t_a,t_b)} e^{F(t_a-\tau)} L L^\intercal e^{F^\intercal (t_b-\tau)}\,d\tau \tag{2}$$

   where e^X is the matrix exponential function

   $$e^X := \sum_{n=0}^{\infty} \frac{X^n}{n!} \qquad \text{and} \qquad X^n := \underbrace{X \cdot X \cdots X}_{n \text{ times}} \tag{3}$$

   Consider the choices (partly discussed briefly in the lecture already)

(a)	$F = 0$	$L = \theta$	(Wiener process)
(b)	$F = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$	$L = \begin{bmatrix} 0 \\ \theta \end{bmatrix}$	(integrated Wiener process)
(c)	$F = -\xi$	$L = \theta$	(Ornstein-Uhlenbeck)
(d)	$F = \begin{bmatrix} 0 & 1 \\ -\xi^2 & -2\xi \end{bmatrix}$	$L = \begin{bmatrix} 0 \\ \theta \end{bmatrix}.$	(Matérn 3/2)

   These four choices identify stochastic processes whose names are printed on the right. If you are interested in more details, see overleaf.

   Use Eq. (2) to find explicit forms for the mean function $m(t)$ and kernel $k(t_a, t_b)$.

Some hints:
For (b), note that $F^2 = 0$ (and thus $F^k = 0 \; \forall k \geq 2$ — the matrix is *nilpotent*). This property can be used in Eq. (3) to compute $\exp(Ft)$.
For (d), note that if a matrix X has eigenvalue decomposition $X = VDV^{-1}$, then its matrix exponential can be written as $e^X = Ve^DV^{-1}$, where e^D is a diagonal matrix containing the (scalar) exponentials of the eigenvalues of X. So you can just compute the eigenvalue decomposition of this 2×2 matrix F to find $\exp F$. This can be done manually for such small matrices (linear algebra reminder: just solve the characteristic polynomial $\det(\lambda I - F)$ for the eigenvalues λ, then find the eigenvectors by solving $(F - \lambda I)v = 0$ for v).

3. **Practical Question — Modeling CO_2** Lecture 11 provided a detailed example for modeling a structured physical process through time. In this exercise, you get to do the same, but with much more interesting data: The CO_2 concentration in Earth's atmosphere. You already know this dataset from the very first sheet. In week 1, you got to try out your skills in standard deep learning models. Now, after 5 weeks of learning about probabilistic models, try your hand again to see if you can now build a better model. More information can be found in the jupyter notebook.

Background Information on Ex. 2 (only for those interested, not necessary to solve the exercise)

The matrix F in (d) is what is known as a *companion matrix*: It is of the form

$$F = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -c_{q-1} \end{bmatrix}$$

Such a matrix is the "companion" of the polynomial

$$p(t) = c_0 + c_1 t + \cdots + c_{q-1} t^{q-1} + t^q,$$

because the characteristic polynomial of F is equal to p. The Matérn family (for $q \in \mathbb{N}$)

$$k_{q+1/2}(r := |t_a - t_b|) = \theta^2 \frac{\Gamma(q+1)}{\Gamma(2q+1)} \sum_{i=0}^{q} \frac{(q+i)!}{i!(q-i)!} \left(\sqrt{8\nu}\frac{r}{\lambda}\right)^{q-i} \cdot \exp\left(-\sqrt{2\nu}\frac{r}{\lambda}\right).$$

is the family of kernels associated with SDEs of the form as in (c), (d), with F the companion matrix of the polynomial $(\xi + it)^{-(q+1)}$ with $\xi = \sqrt{2\nu}\lambda$. Thus, F only has a single, degenerate, eigenvalue $\lambda = -\xi$. That is, they correspond to state space models $x(t) = [x_0, x_1, \dots, x_q]$ where the state consists of the first q derivatives of x_0 (due to the 1's in F). Thus, the Matérn class provides the basic GP model for q-times continuously differentiable and stationary functions. The base case $q = 0$ (c) is known as the *Ornstein-Uhlenbeck (OU)*process and was already mentioned several times. Similar to how the Wiener process models Brownian motion of free particles in an ideal gas, the OU process models the velocity of such particles when bound by a harmonic potential.