Homework 3

AMATH 563, Spring 2025

Due on May 12, 2025 at midnight.

Directions, Reminders and Policies

- You must upload two pdf files, one for theory questions and one for computational tasks to Gradescope.
- Make sure your solutions are well-written, complete, and readable for the theory question. I suggest you use Latex for your theory problems as well as your computational reports.
- For the computational tasks submit a pdf file typeset in Latex or any other typesetting software that can render equations appropriately. Please use the latex template provided on Canvas.
- I encourage collaborations and working with your colleagues to solve HW problems but you should only hand in your own work. We have a zero tolerance policy when it comes to academic misconduct and dishonesty including: Cheating; Falsification; Plagiarism; Engaging in prohibited behavior; Submitting the same work for separate courses without the permission of the instructor(s); Taking deliberate action to destroy or damage another person's academic work. **Such behavior will be reported to the UW Academic Misconduct office without warning**.

Theory Problems

1. Suppsoe $\Gamma : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a PDS kernel. Prove that $\forall x, x' \in \mathcal{X}$ it holds that $|\Gamma(x,x')|^2 \le \Gamma(x,x)\Gamma(x',x')$. Use this to show that the bilinear form
$$\langle f, g\rangle_0 = c^T K(X,Y) b,$$
for functions $f = c^T K(X, \cdot)$ and $g = b^T K(Y, \cdot)$ is an inner product.

2. Given a kernel K on $\mathcal{X}$ define its normalized version as
$$\bar{K}(x,x') = \begin{cases} 0 & \text{if} \quad K(x,x) = 0 \text{ or } K(x',x') = 0 \\ \dfrac{K(x,x')}{\sqrt{K(x,x)}\sqrt{K(x',x')}} & \text{Otherwise.} \end{cases}$$
Show that if K is PDS then so is $\bar{K}$.

3. Show that the following kernels on $\mathbb{R}^d$ are PDS:
   - Polynomial kernel: $K(x,x') = \left(x^T x' + c\right)^\alpha$ for $c > 0$ and $\alpha \in \mathbb{N}$.
   - Exponential kernel: $K(x,x') = \exp(x^T x')$.
   - RBF kernel: $K(x,x') = \exp(-\gamma^2 \|x - x'\|_2^2)$.

4. Let $\Omega \subseteq \mathbb{R}^d$ and let $\{\psi_j\}_{j=1}^n$ be a sequence of continuous functions on Ω and $\{\lambda_j\}_{j=1}^n$ a sequence of non-negative numbers. Show that $K(x,x') = \sum_{j=1}^n \lambda_j \psi_j(x)\psi_j(x')$ is a PDS kernel on Ω.

Computational Problems

Solve the computational problem below. Perform the required tasks and hand in a report of a maximum of six pages outlining your methodology, results, and findings. Use the computational report template provided on Canvas. You don't have to use Latex but it is highly recommended. Make sure that your figures and tables are readable, they are of good size, and labeled appropriately. Part of your grade will be assigned to the general tidiness and style of the report. **You don't need to hand in your code.**

Your goal in this problem is to use kernel ridge regression (KRR) to learn a system of ODEs from limited observations of its state. Before starting this problem be sure to download the starter Jupyter notebook which helps you with the setup of the problem.

In this problem we consider the Lotka-Volterra (LV) predator prey model
$$\begin{aligned} \frac{dp_1}{dt} &= \alpha p_1 - \beta p_1 p_2 \\ \frac{dp_2}{dt} &= -\gamma p_2 + \delta p_1 p_2 \end{aligned} \tag{1}$$
where p_1 and p_2 are the populations of the prey and predator respectively. The parameters $\alpha, \beta, \gamma, \delta$ are positive constants. The system of ODEs models the evolution of predator and prey models with α and γ being the growth rates of the prey and predator populations, and β and δ being the interaction rates between the two populations.

The Figure 1 below, taken from the notebook, shows a trajectory of this system. The true trajectories shown with the solid lines, are computed by solving the ODE system with a prescribed value of $(\alpha, \beta, \gamma, \delta) = (1, 0.1, 0.075, 1.5)$.

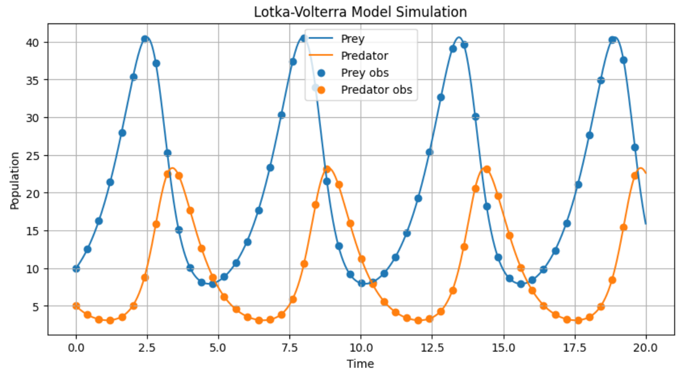

Figure 1 A trajectory of the Lotka-Volterra model. Solid lines show the true states of p_1, p_2 while the dots denote observations.

Now suppose that we do not have knowledge of (1), in fact, we do not even know that the system of ODEs is governed by four parameters. All that we know is that the system is of the form

$$\frac{dp_i}{dt} = f_i(p_1, p_2) \quad i = 1, 2 \tag{2}$$

where f_i are unknown functions and our goal is to learn the $f_i's$ from a set of observations of the system

$$\mathbb{R}^2 \ni y(t_n) = (p_1(t_n), p_2(t_n)), \quad n = 0, \dots, 49, \quad t_n = 0.4\, n.$$

You will solve this problem in three steps

- Step 1: Use KRR to learn the functions $\hat{p}_i \approx p_i$ from the data $Y = (y(t_n))_{n=0}^{49}$. Use the RBF kernel for this task and use cross validation or hand tuning to find the optimal lengthscales and regularization parameters to obtain a good fit.

  Present a plot of the error $|\hat{p}_i - p_i|$ to show the quality of your fit on a dense uniform grid on the time interval $[0, 20]$, e.g. the `t` vector in the notebook.

- Step 2: Consider a second more dense uniform grid $\tilde{t}_1, \dots \tilde{t}_{100}$ on the time interval $[0, 20]$ and compute the time derivatives of the functions $\hat{p}_i$ on this dense grid using the repsenter formula for KRR. It is a good idea to compute the derivatives of the RBF kernel by hand for this step and code them up as functions in your notebook.

  Present a plot of the error $|\frac{d}{dt}\hat{p}_i - \frac{d}{dt}p_i|$ on the $\tilde{t}_n$ grid to show the quality of your fit. You may use finite differences with small step size to compute the derivatives of the true functions $\frac{d}{dt}p_i$.

- Step 3: With $\hat{p}_i(\tilde{t}_n)$ and $\frac{d}{dt}\hat{p}_i(\tilde{t}_n)$ in hand, use KRR to compute approximations $\hat{f}_i$ to the functions f_i in (2). For this step you may Solve the optimization problems

$$\hat{f}_i = \arg\min_{h_i \in \mathcal{H}_i} \sum_{n=1}^{100} \left(\frac{d}{dt}\hat{p}_i(\tilde{t}_n) - h_i(\hat{p}_1(\tilde{t}_n), \hat{p}_2(\tilde{t}_n)) \right)^2 + \lambda_i \|h_i\|_{\mathcal{H}_i}^2,$$

  where $\mathcal{H}_i$ are the RKHSs associated with the kernels H_i.

  Take the H_i kernels to be RBF and second order polynomial kernels and compute $\hat{f}_i$; note you have to tune lengthscales as well as the λ_i's here as well. For each kernel present a 2D contour plot of $|f_i - \hat{f}_i|$ over the domain $[0, 600] \times [0, 60]$. Discuss your findings.

- Use the $\hat{f}_i$ to simulate the trajectories of your learned ODEs with a few new initial conditions and compare them to the true LV trajectories. Discuss your findings.

- Continue to play around with your code and explore problem. You can investigate how the resolution of the t_n and $\tilde{t}_n$ grids affects your solution. You may also investiate the performance of other kernels in Step 1 and Step 3.