Homework 5

AMATH 563, Spring 2025

Due on June 9, 2025 at midnight.

Directions, Reminders and Policies

- You must upload two pdf files, one for theory questions and one for your computational report 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

Let $(\mathcal{H}, \|\cdot\|, \langle\cdot,\cdot\rangle)$ be an RKHS of functions from a set $\mathcal{X} \to \mathbb{R}$, with kernel K.

1. Let $\varphi \in \mathcal{H}^*$, be a bounded linear functional on $\mathcal{H}$. Show that the function $K\varphi : x \mapsto \varphi(K(\cdot, x))$ (this simply means the functional φ is applied to $K(\cdot, x)$ as a function from $\mathcal{X} \to \mathbb{R}$ for each fixed x) is the Riesz representer of φ with respect to the RKHS inner product. That is,

$$\varphi(f) = \langle K\varphi, f\rangle. \tag{1}$$

Hint: (i) First consider the case of f belonging to the pre-Hilbert space $\mathcal{H}_0$; (ii) Observe that by taking φ to be the pointwise evaluation functional (1) is nothing more than the reproducing property.

2. Let $\varphi_1, \ldots, \varphi_m \in \mathcal{H}^*$ be a sequence of bounded linear functionals on $\mathcal{H}$. Show that the orthogonal complement of $\operatorname{span}\{K\varphi_1, \ldots, K\varphi_m\}$ is precisely the subspace

$$\{f \in \mathcal{H} \mid \varphi_j(f) = 0, \quad j = 1, \ldots, m\}.$$

3. Define the bounded linear operator

$$\boldsymbol{\varphi} : \mathcal{H} \to \mathbb{R}^m, \qquad \boldsymbol{\varphi}(f) := (\varphi_1(f), \ldots, \varphi_m(f))^T,$$

and consider the (generalized) interpolation problem

$$\begin{cases} \text{minimize}_{u \in \mathcal{H}} & \|u\| \\ \text{subject to} & \boldsymbol{\varphi}(u) = \mathbf{y}. \end{cases}$$

Consider the matrix $\Theta \in \mathbb{R}^{m \times m}$ with entries $\Theta_{ij} = \varphi_i(K\varphi_j)$. Prove that whenever Θ is invertible then the minimizer $u^\star$ is given by the formula

$$u^\star = \sum_{j=1}^m \alpha_j^\star K\varphi_j, \qquad \text{where} \qquad \boldsymbol{\alpha}^\star = \Theta^{-1}\mathbf{y}. \tag{2}$$

Hint: Observe that if the $\varphi_j = \delta_{x_j}$ were pointwise evaluation functionals at a set of points $X = \{x_1, \ldots, x_m\}$ then the above result coincides with usual interpolation formula in RKHS from HW4 problem 1a.

4. Consider the optimization problem

$$u^\star = \operatorname*{arg\,min}_{u \in \mathcal{H}} L(\varphi_1(u), \ldots, \varphi_m(u)) + \frac{\lambda}{2}\|u\|^2,$$

for a continuous and non-negative loss function $L : \mathbb{R}^m \to \mathbb{R}$. Show that every minimizer $u^\star$ of this problem is of the form

$$u^\star = \sum_{j=1}^m \alpha_j^\star K\varphi_j, \qquad \boldsymbol{\alpha}^\star = \Theta^{-1}\boldsymbol{z}^\star$$

where $\boldsymbol{z}^\star \in \mathbb{R}^m$ solves

$$\boldsymbol{z}^\star = \operatorname*{arg\,min}_{\boldsymbol{z} \in \mathbb{R}^m} L(\boldsymbol{z}) + \frac{\lambda}{2}\boldsymbol{z}^T\Theta^{-1}\boldsymbol{z}.$$

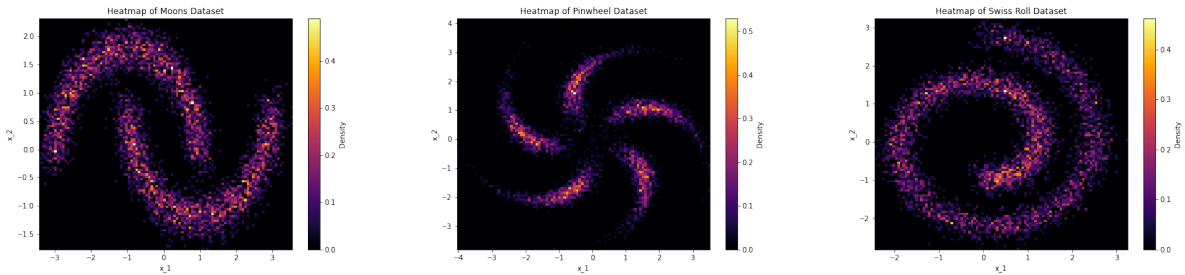

Figure 1 Heatmap of empirical samples from target distributions. From left to right: Moons, Pinwheel, and Swiss Roll data distributions.

Computational Problems

Solve the computational problem below. Perform the required tasks and hand in a report of a **maximum of 10 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.**

In this problem you will design, implement, and test a kernel-based algorithm for transporting a standard normal distribution to a set of complex target distributions in 2D. The target distributions of interest are benchmarks taken from the seminal paper of Grathwohl et al, "FFJORD: Free-form Continuous Dynamics for Scalable Reversible Generative Models". The `HW5-Setup.ipynb` notebook contains the functions for generating and visualizing empirical samples from the the target distributions shown in Figure 1.

Throughout this assingment, $\eta = N(0, I)$ will denote the standard normal distribution in $\mathbb{R}^2$ which will be used as our reference distribution. We will write ν for the target distribution, which will be one of the three distributions above. We will also use N to denote the number of empirical samples from both η and ν.

Hint: Don't forget to normalize your data sets!

1. Write a function that computes the MMD distance between the empirical samples from η and ν for $N = 5000$ samples. Provide a table of the empirical MMD distances between η, ν for each of the three target distributions for the RBF kernel, the Laplacian kernel, and the polynomial kernel with degree $d = 2$. For RBF and Laplacian kernels use the median heuristic to find an appropriate lengthscale.
2. Design and implmenet a minimum MMD transport model for transporting η to ν by parameterizing the transport map as an RKHS function.
   - Clearly explain your mathematical formulation and implementation of the model in a theory or algorithm section of your report.
   - Start with the RBF kernel for both the MMD and the transport map. You will need to tune the lengthscale parameter for the transport map carefully.
   - When implementing the algorithm start with a small N (e.g. $N = 500$ or 1000) to be able to run your code quickly and debug. Once you are confident in the correctness of your code try increasing N.
   - You will have to solve a non-convex but smooth optimization problem after applying the representer theorem for the transport map. You can solve this problem using gradient descent, quasi-Newton

algorithms like L-BFGS, or any other optimization method of your choice. Clearly explain the algorithm you choose and describe how you computed the required gradients or other pertinent quantities and parameters for the optimization algorithm. Also explain how you diagnosed convergence of the optimization algorithm.

Note: You may also use automatic differentiation libraries if you would like. JAX is a good option.

3. Once you have developed your transport model from Part 2. Proceed to train your transport map using the RBF kernel, Laplace kernel, and the polynomial kernel of degree $1, 2, 4$. For each choice of the kernel for the transport map, pick your best results after tuning hyper parameters such as lengthscales and regularization parameters and present a figure showing a heatmap of 5000 empirical samples generated by your transport map in the style of Figure 1.
4. Discuss how the performance of different kernels compares in terms of the MMD distance between the transported samples and the target distribution as well as the visual quality of the samples.
5. Explore the effect of the parameters of the model such as number of training samples N as well as various lengthscale parameters in the MMD loss or the RKHS of the map as well as the regularization parameter. Discuss your findings as you see fit.