MATH 462 HW 5
VERSION November 29, 2022

ADAM M. OBERMAN

Instructions. Refer to notes and references on `https://adam-oberman.github.io/Math462/`. Submit your solutions on MyCourses course page. Math exercises should be *handwritten*. You can get help from other students, but you should do the write up yourself. Coding exercises: export a PDF of the plots required.

This HW is not to be turned in. Do the exercises to study for the final.

5.1. **Exercises: Hilbert Space.** Reference for these questions: M. J. Wainwright (2019) High-dimensional statistics: A non-asymptotic viewpoint. Section 12.1 Hilbert Space Basics.

For $a < b$ define the vector space of functions,

$$C([a,b]) = \{f : [a,b] \to \mathbb{R}, f \text{ continuous}\}.$$

along with the inner product $\langle f, g, \rangle_{L^2([a,b])} = \int_a^b f(x)g(x)dx$ and norm $\|f\|_{L^2} = \langle f, f, \rangle_{L^2([a,b])}$.

Exercise 5.1. *In this exercise we show that $C([a,b])$ is not complete. We can define $L^2([a,b])$ to be the completion: any $f \in L^2$ is the limit of a Cauchy sequence in C (up to points of measure zero).*

(a) Show that $V = C([a,b])$ is a vector space of functions.
(b) Show that $\langle f, g, \rangle_{L^2([a,b])}$ defines an inner product on V, making V an inner product space. Conclude that $\|f\|_{L^2}$ is defines a norm using the fact that each inner product defines a norm.
(c) Set $[a,b] = [-1,1]$ and consider the sequence of functions $f_n(x) = \tanh(nx)$. Find the pointwise limit $f(x)$, which satisfies $f_n(x) \to f(x)$ for all $x \in [-1,1]$.
(d) Show that $f_n(x)$ is a Cauchy sequence in the norm $\|f\|_{L^2}$. Is the limit a function in V? Determine if V is complete.
(e) Prove that $C([0,1]) \subset L^2([0,1])$.

Exercise 5.2. *Consider $\ell^2(\mathbb{N}) = \{\theta = (\theta_1, \dots) \mid \sum_{i=1}^\infty \theta_i < \infty\}$. Define $\langle \theta, w, \rangle_{\ell^2} = \sum_{i=1}^\infty \theta_i w_i$*

(a) Show that $\ell^2(\mathbb{N})$ is an inner product space.
(b) Show that $\ell^2(\mathbb{N})$ is complete: every Cauchy sequence in $\ell^2(\mathbb{N})$ has a limit in $\ell^2(\mathbb{N})$.
(c) Show that $\ell^2(\mathbb{N})$ is an RKHS, and find the representer, the vector $R_i \in \ell^2(\mathbb{N})$ which satisfies $\langle \theta, R_i, \rangle_{\ell^2} = \theta_i$

Exercise 5.3. *In the exercise we show that the evaluation functional on $C([a,b])$ is unbounded. This means that L^2 is not an RKHS, since it is the completion of C.*

(a) Define $R_0 : C([-1,1]) \to \mathbb{R}$ by $R_0(f) = f(0)$, to be evaluation at zero. Prove that R_0 is a linear operator.
(b) Let $f \in C([-1,1])$. Define $A_\epsilon(f) = \langle f, f_\epsilon \rangle$ where

$$f_\epsilon(x) = \begin{cases} \frac{1}{2\epsilon}, & |x| \le \epsilon \\ 0, & \text{otherwise} \end{cases}$$

Show that $A_\epsilon(f)$ corresponds to the average of f on the interval $-\epsilon, \epsilon$. Show that $f(0) = \lim_{\epsilon \to 0} A_\epsilon(f)$.
(c) Interpret the last result as saying that $R_0 = \lim_{\epsilon \to 0} f_\epsilon$ for all $f \in C([-1.,1])$.

Date: November 29, 2022.

(d) Find the norm of A_ϵ. Conclude that R_0 is an unbounded linear operator on $C([-1.,1])$.

Exercise 5.4. *Consider the Hilbert space, $\mathbb{H}$. Prove the Cauchy-Schwartz inequality,*

$$\langle f, g\rangle \le \|f\|\|g\|, \qquad \text{for all} f, g \in \mathbb{H}$$

Exercise 5.5. *Let $\mathbb{H}$ be an RKHS of functions $f : \mathcal{X} \to \mathbb{R}$. Let $\{f_n\}_{n=1}^{\infty}, f, g, \in \mathbb{H}$,*

(a) Given $x \in \mathcal{X}$, Prove that there is a $C = C(x)$ such that $|f(x) - g(x)| \le C\|f - g\|_{\mathbb{H}}$. Express C in terms of the evaluation functional.

(b) Suppose $\lim_{n\to\infty} \|f_n - f\|_{\mathbb{H}} = 0$. Prove pointwise convergence: for all $x \in \mathcal{X}$, $\lim_{n\to\infty} |f_n(x) - f(x)| = 0$.

(c) Why is pointwise convergence important for machine learning functions?

5.2. **Coding.** Refer to the feature regression and kernel regression code `https://colab.research.google.com/drive/15sKQFCfgoMKE-fZOkVVWBmvWH9Sm-TF5?usp=sharing`

Let $f(x) \in \mathbb{R}^d$ be a feature vector, and define $K(x,z) = f(x) \cdot f(z)$. Given the dataset $S^m = \{(x_i, y_i)\}_{i=1}^m$, consider the regression function $h_w(x) = w \cdot f(x)$ along with the regularized the feature regression problem

$$\min_w \widehat{L}(h_w) + \lambda\|w\|^2$$

Consider also the kernel function $h_a(x) = \sum_{i=1}^m a_i K(x, x_i)$ and the Kernel Ridge Regression (KRR) problem

$$\min_a \widehat{L}(h_a) + \lambda\|h_a\|_{\mathbb{H}}^2$$

Exercise 5.6. *Using the code provided, compare the solutions of the two problems, using polynomial features. Consider the cases $d = m$ as well as the cases $d \neq m$. When $\lambda = 0$ do the functions overfit? How similar are the two solutions?*