EE660 Homework 2

Assigned: 9/18/2025, Due: 10/2/2025

Instructions: You may collaborate with others on this problem set, but each student must independently write their own solutions. We highly encourage you to use LaTeX to typeset your solutions. We will accept handwritten assignments, however if the solution to a problem is too illegible for the grader to read, then they may use their discretion and consider the problem incomplete. Solutions are due by 11:59pm Pacific Time on the due date, and are only to be submitted on DEN. Do not email the course staff with your assignment.

Template: The link `https://www.overleaf.com/read/hjgknqqhryqy` contains a basic LaTeX template that you may use. Note, however, that you are not required to use this template.

Note 1: When a problem mentions a "universal constant", this just means a number that does not depend on any of the problem parameters.

1. Problem 1

In this problem, we will prove that the Gaussian RBF kernel cannot be expressed as the inner product of a finite-dimensional feature map. For simplicity, we will work in one dimension (i.e., $d = 1$), and consider the bandwidth $\sigma = 1$, so that the Gaussian RBF kernel is simply:

$$k : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}, \quad k(x, y) = \exp(-(x-y)^2/2).$$

(a) Define the matrix $E_n \in \mathbb{R}^{n \times n}$ with $(E_n)_{i,j} := \exp(i \cdot j)$ for $i, j \in [n]$. Show that E_n is invertible for all $n \in \mathbb{N}_+$.

Hint: A $n \times n$ Vandermonde matrix $V(x_1, \ldots, x_n)$ is defined to have entries $V(x_1, \ldots, x_n)_{i,j} = x_i^{j-1}$ for $i, j \in [n]$. You may use without proof the formula:

$$\det(V(x_1, \ldots, x_n)) = \prod_{1 \leqslant i < j \leqslant n} (x_j - x_i).$$

You may also use without proof the fact that $\det(AB) = \det(A)\det(B)$ for two size-conforming square matrices A, B.

(b) Consider the data points $\{x_i\}_{i=1}^n$ with $x_i = i$ for $i \in [n]$. Let $K_n \in \mathbb{R}^{n \times n}$ be the kernel matrix associated with $\{x_i\}_{i=1}^n$, i.e., $(K_n)_{i,j} = k(x_i, x_j)$ for $i, j \in [n]$. Use part (a) to show that K_n is invertible for all $n \in \mathbb{N}_+$.

(c) Use part (b) to conclude that there does *not* exist any feature map $\Phi : \mathbb{R} \mapsto \mathbb{R}^p$ with $p \in \mathbb{N}_+$ such that $k(x, y) = \langle \Phi(x), \Phi(y) \rangle$ for all $x, y \in \mathbb{R}$.

2. Problem 2

Let ℓ be the zero-one loss, and let $\mathscr{F}$ be a countably infinite function class of the form:

$$\mathscr{F} = \{f_\kappa \mid \kappa \in \mathbb{N}_+\}.$$

Let $\delta \in (0,1)$ and $n \in \mathbb{N}_+$. Show that with probability at least $1-\delta$, we have:

$$\forall \kappa \in \mathbb{N}_+,\ L[f_\kappa] - \hat{L}_n[f_\kappa] \leqslant c_0\sqrt{\frac{\log(c_1\kappa/\delta)}{n}},$$

where c_0, $c_1 > 1$ are universal positive constants.

Hint: You may use without proof the solution to the Basel problem: $\sum_{i=1}^\infty i^{-2} = \pi^2/6$.

Hint: Given an countable sequence of events $\{E_i\}_{i=1}^\infty$, recall that probability measures satisfy the *countable sub-additivity* property: $\mathbb{P}(\cup_{i=1}^\infty E_i) \leqslant \sum_{i=1}^\infty \mathbb{P}(E_i)$.

Note: You may use without proof the following bound established in the proof of Proposition 1.21 from the lecture notes:

$$\mathbb{P}\{L[f] - \hat{L}_n[f] \geqslant t\} \leqslant \exp(-nt^2/2), \quad t > 0,\ f \in \mathscr{F}.$$

3. Problem 3

Let $x_i \in \mathbb{R}^d$, $i = 1, \ldots, n$ be drawn iid from an underlying distribution on $\mathbb{R}^d$. Let $Q \in \mathbb{R}^{d\times d}$ be a positive definite matrix.[1] For any $d \times d$ positive definite matrix M, let $\|\cdot\|_M$ be the weighted Euclidean norm $\|x\|_M := \sqrt{x^\mathsf{T} M x}$ for all $x \in \mathbb{R}^d$. Consider the following linear function class:

$$\mathscr{F}_Q := \{x \mapsto \langle x, w\rangle \mid \|w\|_{Q^{-1}} \leqslant 1\}.$$

(a) Show that:

$$\mathcal{R}_n(\mathscr{F}_Q) \leqslant \sqrt{\frac{\mathbb{E}\|x_1\|_Q^2}{n}}.$$

(b) Suppose furthermore that the x_i's are drawn iid from $N(0, I_d)$. Show that:

$$\mathcal{R}_n(\mathscr{F}_Q) \leqslant \sqrt{\frac{\operatorname{tr}(Q)}{n}},$$

where the trace of Q, denoted $\operatorname{tr}(Q) = \sum_{i=1}^d Q_{ii}$, is the sum of the diagonal entries of Q. Note that by setting $Q = I_d$, this result implies that $\mathcal{R}_n(\mathscr{F}_2) \leqslant \sqrt{d/n}$ as we saw in lecture.

Hint: You may use without proof that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for any size conforming matrices A, B.

4. Problem 4

Suppose that $x_1, \ldots, x_n$ are not necessarily independent. Show that the bound

$$\mathbb{E} \sup_{f\in\mathscr{F}} \frac{1}{n}\sum_{i=1}^n (\mathbb{E}[f(x_i)] - f(x_i)) \leqslant 2\mathcal{R}_n(\mathscr{F})$$

no longer holds in general.

[1] Recall that a $d \times d$ matrix M is *positive definite* if it is symmetric (i.e., $M = M^\mathsf{T}$) and all its d eigenvalues are positive.

5. Problem 5

Consider the alternative definition of Rademacher complexity of a function class $\mathscr{F}$:

$$\bar{\mathcal{R}}_n(\mathscr{F}) := \mathbb{E} \sup_{f \in \mathscr{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(x_i) \right|.$$

Clearly, we have that this alternative definition dominates the definition of Rademacher complexity presented in lecture, i.e., $\mathcal{R}_n(\mathscr{F}) \leqslant \bar{\mathcal{R}}_n(\mathscr{F})$. On the other hand, let $\mathscr{F}$ be a set of functions mapping $\mathsf{X} \mapsto [-1, 1]$. Show a reverse inequality:

$$\bar{\mathcal{R}}_n(\mathscr{F}) \leqslant 2\mathcal{R}_n(\mathscr{F}) + \sqrt{\frac{1}{n}}.$$