EE6327: Statistical Learning Theory 2025

Exam 4: November 2025

Instructor: Shashank Vatedka

Instructions: *This is a closed-book exam. You are not permitted to refer to any material or discuss the problem with anyone. Malpractice will be severely punished. Please mention your ROLL Number and name clearly in the answer sheet.*

Justify all your statements clearly. You may use any result proved in class (but clearly state which results you are using), but everything else needs to be proved.

Question 4.1 (5+2+8 pts). Let $\mathcal{B}$ be an arbitrary (base) hypothesis class. In the lectures, we saw two different ways (bagging and boosting) of combining base classifiers to give an improved classifier.

Let k be a *fixed* positive integer. Let us consider all possible ways of combining k base classifiers to give a richer hypothesis class:

$$\mathcal{H} = \left\{h(x) = g(h_1^B(x), \ldots, h_k^B(x)) : h_i^B \in \mathcal{B}, g : \{-1,1\}^k \to \{-1,1\}\right\}$$

Assuming that the VC dimension of $\mathcal{B}$ is d, show that the VC dimension of $\mathcal{H}$ (denoted $d_{\mathcal{H}}$) must satisfy (8pts)

$$d_{\mathcal{H}} \leqslant d \log_2 d_{\mathcal{H}} + d \log_2 \frac{e}{d} + 2^k.$$

Hint: Let $\mathcal{B}'$ denote the set of all functions from $\mathcal{X}' = \{-1,1\}^k$ to $\{-1,1\}$. What is the growth function (5pts) and VC dimension (2pts) of $\mathcal{B}'$?

Question 4.2. Find whether each of the following kernels on $\mathcal{X} = \mathbb{R}^d$ is PDS. If yes, prove that it is PDS. If not, give an example of $x_1, \ldots, x_n$ for which the gram matrix is not PSD (you may assume a specific d to give the example).

1. $K(x, x') = \|x + x'\|_2^2$ (3pts)
2. $K(x, x') = \sin(\|x + x'\|_2)$ (3pts)
3. $K(x, x') = 1_{\{\|x - x'\|_2^2 \leqslant 1\}}$ (3pts)

Question 4.3. Let us consider a "compressive sensing" problem for binary sequences. The goal is to demonstrate the existence of a binary $k \times d$ matrix H such that we can recover any s-sparse (number of 1's is at most s) binary vector $\underline{x}$ from

$$\underline{w} = H\underline{x}.$$

In the above equation, we assume that the operations are over the binary field. In other words, addition and multiplication are performed modulo-2 (or equivalently, multiplication is the AND operation and addition is the XOR operation).

Suppose that $s = dp$ for some $0 < p < 1/2$, where you may assume that dp is a positive integer. If $\mathcal{A}_p$ denotes the set of all dp-sparse sequences, then show that (5pts)

$$|\mathcal{A}_p| \leqslant C 2^{dH_2(p)}(1 + o(1)),$$

for some universal constant $C > 0$ where $o(1) \to 0$ as $d \to \infty$ and

$$H_2(p) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}.$$

You may use Stirling's approximation without proof:

$$n! \leqslant C\sqrt{n}\left(\frac{n}{e}\right)^n (1+o(1))$$

for some universal constant C.

Now show that if H is chosen to be a random matrix with iid Bernoulli(1/2) entries, then (10pts)

$$\Pr[H\underline{x} = H\underline{x}' \text{ for some } \underline{x} \neq \underline{x}', \text{ where } \underline{x}, \underline{x}' \in A_p] \to 0$$

as $d \to \infty$ if we select $k = dR$, where

$$R > H(2p).$$