Machine Learning from Data – IDC – 2023

HW5 – Theory + SVM

1. Kernels and mapping functions (30 pts)
   a. (10 pts) Consider two kernels K_1 and K_2, with the mappings φ_1 and φ_2 respectively. Show that $K = 5K_1 + 4K_2$ is also a kernel and find its corresponding φ.
   b. (10 pts) Consider a kernel K_1 and its corresponding mapping φ_1 that maps from the lower space R^n to a higher space R^m (m > n). We know that the data in the higher space R^m, is separable by a linear classifier with the weights vector w.
   Given a different kernel K_2 and its corresponding mapping φ_2, we create a kernel $K = 5K_1 + 4K_2$ as in section a above. Can you find a linear classifier in the higher space to which φ, the mapping corresponding to the kernel K, is mapping?
   If YES, find the linear classifier weight vector.
   If NO, prove why not.
   c. (10 pts) Consider the space $S = \{1,2, \dots N\}$ for some finite N (each instance in the space is a 1-dimension vector and the possible values are 1, 2, …, N) and the function $K(x, y) = 9 \cdot f(x, y)$ for every $x, y \in S$.
   Prove that K is a valid kernel by finding a mapping φ such that:
   $$\varphi(x) \cdot \varphi(y) = 9 \min(x, y) = K(x, y)$$
   For example, if the instances are $x = 4, y = 8$, for some $N \geq 8$, then:
   $$\varphi(x) \cdot \varphi(y) = \varphi(4) \cdot \varphi(8) = 9 \cdot \min(4,8) = 36$$

2. Lagrange multipliers (20 pts)

Suppose you are running a factory, producing some sort of widget that requires steel as a raw material. Your costs are predominantly human labor, which is \$20 per hour for your workers, and the steel itself, which runs for \$170 per ton.
Suppose your revenue R is modeled by the following equation:

$$R(h, s) = 200 \cdot h^{\frac{2}{3}} \cdot s^{\frac{1}{3}}$$

Where:

- h represents hours of labor
- s represents tons of steel

If your budget is \$20,000, what is the maximum possible revenue?

3. <u>PAC Learning and VC dimension (30 pts)</u>

Let $X = \mathbb{R}^2$. Let

$C = H = \left\{ h(r_1, r_2) = \left\{ (x_1, x_2) \middle| \begin{matrix} x_1^2 + x_2^2 \geq r_1 \\ x_1^2 + x_2^2 \leq r_2 \end{matrix} \right\} \right\}$, for $0 \leq r_1 \leq r_2$,

the set of all origin-centered rings.

a. (8 pts) What is the $VC(H)$? Prove your answer.
b. (14 pts) Describe a polynomial sample complexity algorithm L that learns C using H. State the time complexity and the sample complexity of your suggested algorithm. Prove all your steps.

In class we saw a bound on the sample complexity when H is finite.

$$m \geq \frac{1}{\varepsilon}\left(\ln|H| + \ln\frac{1}{\delta}\right)$$

When $|H|$ is infinite, we have a different bound:

$$m \geq \frac{1}{\varepsilon}\left(4\log_2\frac{2}{\delta} + 8VC(H)\log_2\frac{13}{\varepsilon}\right)$$

c. (8 pts) You want to get with 95% confidence a hypothesis with at most 5% error. Calculate the sample complexity with the bound that you found in b and the above bound for infinite $|H|$. In which one did you get a smaller m? Explain.

4. <u>VC dimension (20 pts)</u>

Let $X = \mathbb{R}$ and $n \in \mathbb{N}$.

Define "x-node decision tree" for any $x = 2^n - 1$ to be a full binary decision tree with x nodes (including the leaves).

Let H_m be the hypothesis space of all "x-node decision tree" with $n \leq m$.

a. (5 pts) What is the $VC(H_3)$? Prove your answer.
b. (15 pts) What is the $VC(H_m)$? Prove your answer.