BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

Department of Computer Science and Information Systems

II SEMESTER 2022-2023

BITS F453 – Computational Learning Theory

Part-A (Closed Book)

18th May 2023 | Comprehensive Exam | Weightage: 20%

Note: Each question carries 2 marks. Minimum time is 60 mins. and Maximum time is 90 mins.

1. Give 4 examples of infinite hypothesis classes which are PAC learnable. Justify your answer. No marks without proper justification.

2. What is Regularized Loss Minimization (RLM) learning paradigm? Apply RLM rule with Tikhonov regularization to formulate the ridge regression problem.

3. How can we work out the VC dimension of a neural network? Give a few ideas.

4. How can we characterize online learning? What is the notion of ERM in context of online learning? How effective is ERM? Explain through an online learning algorithm which uses ERM.

5. Define VC and Littlestone dimensions. For a given hypothesis class, what is the relationship between the two dimensions. Give an example of an infinite hypothesis class for which the difference between the two types of dimensions is maximum.

6. Explain how VC dimension and Littlestone dimension (in both realizable & unrealizable settings) can be worked out for a hypothesis class using game playing.

7. Compare Consistent, Halving, and Standard Optimal algorithms in terms of mistake bound and Littlestone dimensions.

8. *For* $i = 1, \ldots, r$, *let* $f_i : \mathbb{R}^d \to \mathbb{R}$ *be a convex function. Show that the following functions from* $\mathbb{R}^d$ *to* $\mathbb{R}$ *are also convex.*
   - $g(x) = \max_{i \in [r]} f_i(x)$
   - $g(x) = \sum_{i=1}^{r} w_i f_i(x)$, *where for all* i, $w_i \geq 0$.

9. Find the Lipschitzness and Smoothness of the following functions:
   (a) $f(x) = \log(1+\exp(x))$
   (b) $f(x) = x^2$

10. Explain how the concepts learned in the course can be applied to practical Machine/Deep Learning.

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

Department of Computer Science and Information Systems

II SEMESTER 2022-2023

BITS F453 – Computational Learning Theory

Part-B (Open Book)

18th May 2023 | Comprehensive Exam | Weightage: 20%

1. Prove that under realizability assumption the class of all axis parallel rectangles in two dimension is PAC learnable. An axis parallel rectangle labels all points lying on or inside it as positive class and all other points as negative class. Derive an expression for sample complexity for learning this class.

[4]

2. Show that the Natarajan Dimension degenerates to VC dimension when there are exactly two classes.

[4]

3. Show that Hyperplanes, Half-spaces, & Euclidean balls are convex:

Hyperplanes: $\{x \mid a^T x = b\}$ $(a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0)$

Halfspaces: $\{x \mid a^T x \leq b\}$ $(a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0)$

Euclidean balls: $\{x \mid ||x - x_c|| \leq r\}$ $(x_c \in \mathbb{R}^n, r \in \mathbb{R}, ||.||$ 2-norm$)$

[1+1+2]

4. For any constant $c > 0$, a polynomial kernel of degree $d \in \mathbb{N}$ is the kernel defined over $\mathbb{R}^n$ by:

$$\forall\, \boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^n, \quad \boldsymbol{k}(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x}.\boldsymbol{x}' + c)^{\boldsymbol{d}}$$

Use an appropriate polynomial kernel to solve the XOR classification problem.

Hint: Polynomial kernels map the input space to a higher-dimensional space of dimension $\binom{n+d}{d}$.

[4]

5. Explain each of the following cases with regard to the value of the hyperparameter C in case of soft SVM for mis-classification of points. Also, which of these cases is the optimal one. Compare the size of the hypotheses class in each of these cases with respect to each other. [4]