EE5604: Introduction to Statistical Learning Theory, Fall 2022

Indian Institute of Technology Hyderabad
HW 0, Assigned: Sunday 25.09.2022. 25 points.
Due: Saturday 01.10.2022 at 11:59 pm.

1. Find the VC dimension of the set of linear indicator functions $Q(z,\alpha) = \theta\{\sum_{i=1}^{n} \alpha_i z_i + \alpha_0\}$. $\theta(.)$ is the indicator function and z is a n-dimensional vector. (5)

2. For the binary classification problem, given an example of a class of functions with infinite VC dimension. Clearly justify your example. (5)

3. State and prove the consistency of the ERM principle for the set Λ with finite cardinality. Clearly explain your work. (5)

4. A crucial step in the proof of the consistency of the ERM principle for the set Λ with infinite cardinality is

$$\mathbb{P}(\sup_{\alpha\in\Lambda}\frac{1}{n}|\sum_{i=1}^{n}\sigma_i(\mathbf{1}(f(x_i,\alpha)\neq y_i)-\mathbf{1}(f(x_i^{'},\alpha)\neq y_i^{'}))|\geq\frac{\epsilon}{2}) \leq \mathbb{P}(\sup_{\alpha\in\Lambda}\frac{1}{n}|\sum_{i=1}^{n}\sigma_i(\mathbf{1}(f(x_i,\alpha)\neq y_i)|\geq\frac{\epsilon}{4}$$
$$\text{OR } \sup_{\alpha\in\Lambda}\frac{1}{n}|\sum_{i=1}^{n}\sigma_i(\mathbf{1}(f(x_i^{'},\alpha)\neq y_i^{'}))|\geq\frac{\epsilon}{4})$$

   Show this. (5)

5. Implement a multi-layer perceptron to model the XOR gate. Generate a dataset containing $N = 1000$ datapoints. You can generate these points by adding noise to the input of the XOR truth table.

   (a) Experiment with the train-test split and plot the training and test errors versus epochs for each split. Specifically, go from 0:100 to 100:0 in steps of 20 and train the model for 25 epochs. For the cases with zero training/test points report only the testing/training error respectively. (3)

   (b) Pick a couple of train-test splits and experiment with N. Specifically, choose $N = 1000, 5000, 10000$. For each case, plot training and test errors versus epochs. Note down your observations in terms of how the errors vary with N as well as how the train and test error values compare for a given train-test split. Are your observations in line with the ERM theory? (2)