CHENNAI MATHEMATICAL INSTITUTE

Midsemester Exam - TFML

Date: Oct 6, 2024

(1) (3 marks) Using the concavity of the $\ln(x)$ function, show that $\ln(n!) \leq \int_1^n \ln(x)dx + ln(n)/2$. Use this to give an upper bound on $n!$.
(2 marks) Also prove a reasonable lower bound. '

(2) (5 marks) We are learning a concept class from samples which are labelled 0,1. We believe there are bunch of $d$ data features and if any one of them is true, the data point should be classified as 1. Suppose the data is generated from a concept $h^*$, which is a disjunction of these features. Design a learning algorithm whose empirical risk is zero. For given $\epsilon, \delta$, how many samples will you need? Assume there are $d$ features.

(3) (4 marks) Let $\mathcal{H}_1, \mathcal{H}_2$ be two hypothesis classes each with VC dimension $d$. Suppose we take the union of these hypothesis classes as our new hypothesis class $\mathcal{H}$. Show that the size of the largest set that can be shattered by $\mathcal{H}$ is at most $2d + 1$

(4) (4 marks) Given a hypothesis class $\mathcal{H}$, we produce the hypothesis class $\mathcal{H}'$ as follows. The hypothesis in $\mathcal{H}'$ are tuples $(h_1, h_2, \ldots, h_{2k+1})$ with $h_i \in \mathcal{H}$. On the given data set assume the classification by $h' := (h_1, \ldots, h_{2k+1}) \in \mathcal{H}'$ is $h'(x) = 1$ if for majority of the $i$'s, $h_i(x) = 1$, it is 0 otherwise.
Assuming the VC dimension of $\mathcal{H}$ is $d$ what is the VC dimension of $\mathcal{H}'$?

(5) (4 marks) Let $\ell_1, \ell_2, \ldots, \ell_k$ be non negative integers such that $\sum_i 3^{-\ell_i} \leq 1$. Show that the following procedure constructs a prefix free ternary code $\mathcal{C}$ with $k$ codewords of lengths $\ell_i$, $1 \leq i \leq k$

- W.l.o.g, $\ell_1 \leq \ell_2 \leq \cdots \leq \ell)k$. On a complete ternary tree of depth $\ell_k$, take a node at depth $\ell_1$. The path from root to node gives a codeword of length $\ell_1$ which we add to $\mathcal{C}$. Now remove all nodes below that and proceed to $\ell_2$ and continue the same.

(6) (1 mark) Recall that in the minimum description length paradigm, given a sample $S$, we return a hypothesis $h_S$ satisfying the conditions given below.

$$h \in argmin_{h \in \mathcal{H}}[L_S(h) + \sqrt{\frac{|h| + \ln(2/\delta)}{2m}}]$$

Here $m$ is the number of samples and $|h|$ is the description length of $h$ - what is the bias variance tradeoff in this paradigm as compared to ERM?
(2 marks) Now assume the data comes from a distribution $\mathcal{D}$. Also suppose for a constant $B$ we consider hypothesis class $\mathcal{H}_B$ containing

only hypothesis with description length at most $B$. Set $h_B^*$ to be a hypothesis in $\mathcal{H}_B$ which minimizes total risk. Compute a bound on $L_{\mathcal{D}}(h_S) - L_{\mathcal{D}}(h_B^*)$.

(1 mark) What should be the sample complexity if we wish to to output a hypothesis which is at most $\epsilon$-away from $L_{\mathcal{D}}(h_B^*)$ with probability $1-\delta$?

(7) (2 marks) Compute the distance of a point $x$ to the hyperplane $w^T x + b = 0$, in terms of quantities involving $w, x, b$.

(2 marks) In the linear separable case, we wish to find a hyperplane $w$ and $b$ which maximizes the minimum of the distances of the points $x_i$ to the hyperplane $w^t x + b = 0$. Show that the following is a valid formulation to find such a separating hyperplane:

$$min_w \frac{1}{2}||w||^2, \; such\ that\ \forall i, y_i(w^t x_i + b) \geq 1$$