Advanced Topics in Artificial Intelligence: EE6180
Indian Institute of Technology Madras — October 19, 2020
Instructor: Abhishek Sinha

Mid-Term Exam.

• The mid-term paper will be due on **October 28, 2020** in the class.
• Each problem carries 10 points.
• No collaboration among the students allowed. Any two or more identical or nearly-identical solutions will automatically receive zero points each.

---

1. **(Gaussian Complexity of ℓ_0-"balls")** Sparsity plays an important role in many classes of high-dimensional statistical models. In this problem, we will compute the Gaussian complexity of an s-sparse ℓ_0-ball intersected with a unit ℓ_2-ball. Consider the set

$$T^d(s) = \{\theta \in \mathbb{R}^d : ||\theta_0|| \leq s, ||\theta||_2 \leq 1.\}$$

corresponding to all s-sparse vectors contained within the Euclidean unit ball. Recall that the Gaussian Complexity of a set $V \subset \mathbb{R}^d$ is defined as

$$\mathcal{G}(V) = \mathbb{E}\big[\max_{\boldsymbol{v} \in V} \boldsymbol{v}^T \boldsymbol{w}\big],$$

where $w_i \sim_{\text{i.i.d.}} \mathcal{N}(0,1), \forall i$

In this problem, we prove that the Gaussian complexity of $T^d(s)$ is upper bounded as

$$\mathcal{G}(T^d(s)) \leq \sqrt{s} + \sqrt{2s \ln\left(\frac{ed}{s}\right)}. \quad (1)$$

(a) First show that $\mathcal{G}(T^d(s)) \leq \mathbb{E}\big[\max_{|S|=s} ||w_S||_2\big]$, where $w_S \in \mathbb{R}^{|S|}$ denotes the sub-vector of $(w_1, w_2, \ldots, w_d)$ indexed by the subset $S \subseteq \{1, 2, \ldots, d\}$.
(b) Next show that for any fixed subset S of cardinality s:

$$\mathbb{P}\big[||w_S||_2 \geq \sqrt{s} + \delta\big] \leq e^{-\delta^2/2}.$$

(c) Use the preceding parts to establish the bound (1).

2. **(Hedge with Many Good Experts)** Consider running the `Hedge` algorithm (with learning rate $\eta > 0$) in the standard expert's setting as discussed in the class. Show that for all $T \geq 1$ and any $L > 0$,

$$\hat{L}_T^{\texttt{Hedge}} \leq L + \frac{1}{\eta} \ln \frac{N}{N_L} + \eta T,$$

where $N_L = |1 \leq i \leq N : L_{i,T} \leq L|$.

3. **(Arbitrarily Small Training Error)** Let $\mathcal{Z} = \{(x_i, f(x_i)), 1 \leq i \leq N\}$ be a set of N training samples, where $\mathcal{X} = \{x_i, 1 \leq i \leq N\}$ is the set of N corresponding feature vectors and $f : \mathcal{X} \to \{\pm 1\}$ is some unknown target function. Suppose that we have a hypothesis class $\mathcal{H} \subseteq \{h : \mathcal{X} \to \{\pm 1\}\}$, such that for any distribution μ on $\mathcal{X}$, there exists an $h \in \mathcal{H}$, such that the classification error (w.r.t. μ) is at most $\frac{1}{2} - \gamma$, for some $\gamma > 0$, *i.e.,*

$$\mathbb{P}_{x \sim \mu}(h(x) \neq f(x)) \leq \frac{1}{2} - \gamma.$$

Let $\texttt{WM}_n(\mathcal{H})$ be the class of weighted majority vote functions consisting of n hypotheses, *i.e.,*

$$\texttt{WM}_n(\mathcal{H}) = \{w(x) = \texttt{sign}(\sum_{i=1}^{n} \alpha_i h_i(x)).$$

where $x \in \mathcal{X}, h_i \in \mathcal{H}, \alpha_i \geq 0, \sum_i \alpha_i = 1$. Prove that there exists a hypothesis in the class $\texttt{WM}_T(\mathcal{H})$ with $T = O(\frac{1}{\gamma^2} \log(\frac{1}{\epsilon}))$, which misclassifies at most an ϵ fraction of the training set $\mathcal{Z}$.

HINT: Recall the reduction of `Boosting` to `Online Learning` as discussed in the class. Use `Hedge` as your particular online learning algorithm. Result from Problem 2 could be useful.

4. **(`Group Testing` Lower Bounds)** In the ongoing COVID-19 pandemic, when the testing kits are short in supply, `Group Testing` is an effective method to carry out a large number of tests with a limited number of kits. Check out the following expository article to understand how `Group Testing` is being carried out in India and other parts of the world: `https://www.nature.com/articles/d41586-020-02053-6`.

The above article describes four possible methods of detecting COVID-19 via Group Testing. In this problem, we investigate the fundamental limits of all such testing procedures.

Formally, our goal is to figure out which of the k locations in an n-dimensional binary vector b are non-zero (say, equal to 1). One can query some subset of the dimensions, *i.e.,* a query vector ϕ is binary with 1's in the dimensions you want to query. The outcome of the query is binary and equals $\vee_i \phi_i b_i$, which is 1 iff at least one of the queried dimensions is 1 in the noiseless case and flipped independently with probability $q \leq 0.5$ in the noisy case. We say that an error occurs when the inferred locations of 1's differ in at least one coordinate from the original locations of 1's in the vector b. Show that[1],

(a) in the noiseless case, any group testing algorithm requires at least $(1-\epsilon)k\log(n/k) - 1$ queries to have probability of error at most ϵ.

(b) in the noisy case with noise probability $q \in [0, 1/2)$, any group testing algorithm requires at least $\frac{(1-\epsilon)k\log(n/k)-1}{1-h(q)}$ queries to have probability of error at most ϵ, where $h(q)$ is the usual binary entropy function.

[1] The base of all logarithms in this problem is 2.