Computational Learning Theory Mid Term Examination (35 points)
03/11/2022—— 9AM-10:30AM—— CS F453
Instructor: Snehanshu Saha Name: _____

Handwritten/printed notes and calculator are allowed. Laptop, mobile phone and any other form of electronic gadget is NOT allowed. Any violation will be interpreted as unfair means and disciplinary action will be taken. Read all directions carefully and write your answers in the space provided.
ANSWER ALL QUESTIONS.

NOTE: Use supplementary pages provided to do the scratch work.

1. (10 points) Mark the following statements as True (**T**) or False (**T**). No explanation is necessary. Each question carries 2 points.
   (a) Let $S = [(x_1, y_1); ::: (x_n, y_n)]$ be n linearly separable points by a separator through the origin in R^d. Let S_1 be generated from S as: $S_1 = [(cx_1, cy_1); ::: (cx_n, cy_n)]$, where $c > 1$ is a constant. The error bound of perceptron on S_1 is larger than the error bound on S.
   (b) Suppose the VC dimension of a hypothesis class H is d. Suppose we are given samples $[(x_i, y_i); i = 1..n]$ where x_i's are distinct and $n \leq d$. Then there always exists a hypothesis in H which perfectly classifies the samples.
   (c) To show that the VC-dimension of a concept class H (containing functions from X to (0, 1) is d, it is sufficient to show that there exists a subset of X with size d that can be labeled by H in all possible 2^d ways.
   (d) The true error of a hypothesis h can be lower than its training error on the sample S.
   (e) PAC Learning paradigm is easier to learn compared to the agnostic-PAC learning counterpart.

2. (10 points) Consider the sum-of-squared-error decision function $J_s(a) = \sum_{i=1}^{n}(a^t y_i - b_i)^2$. Assume a is in the neighborhood of a local minima of $J(a)$. Show that, for a positive learning rate $\eta(k)$, J indeed attains minima at the optimal $\eta(k)$. This shows that we can compute optimal learning rate for a quadratic decision function!

   Hint: Use a second-order Taylor series expansion of the decision function at the point $a(k)$: $J(a) = J(a(k)) + \nabla J^t(a(k))(a - a(k)) + 1/2 * (a - a(k))^t * H(a(k))(a - a(k))$ and the update rule: $a(k+1) = a(k) - \eta(k)\nabla J(a(k))$. Note H is the Hessian matrix.

3. (10 points) An axis aligned rectangle classifier in the plane is a classifier that assigns the value 1 to a point if and only if it is inside a certain rectangle. Formally, given real numbers $a_1 \leq b_1; a_2 \leq b_2$, the classifier returns 1 if $a_1 \leq x_1 \leq; b_1; a_2 \leq x_2 \leq b_2$, else it returns 0. The hypothesis class, $H(|H| = \infty)$ is the collection of all the classifiers defined above.
   Let A be the algorithm that returns the smallest rectangle enclosing all positive examples in the training set. Show that A is an ERM. Furthermore, show that if A receives a training set of size $\geq \frac{4log(4/\delta)}{\epsilon}$, then, with probability of at least $1 - \delta$, it returns a hypothesis with error of at most ϵ.

4. (5 points) Given a sample of m bounded points $X = [(x_1, x_2, ..., x_m), i, |x_i| \leq M]$, define the function $f(X) = \frac{\sum_{i=1}^{n} x_i}{m}$. Furnish a bound on the probability $Pr[|f(X)E[f(X)] \geq \epsilon]$.