

Machine Learning 4771: Homework 2

(15% of the grade)

Due on February 26, 2:40pm

Note: You are not allowed to submit an answer you cannot explain in person. Please take this seriously: You can be quizzed on your homework.

Late assignments will not be accepted.

Problem 1

(Regression) Given a set of training examples $(x_1, y_1), \dots, (x_N, y_N) \in \mathbf{R}^n \times \mathbf{R}$, let \mathbf{X} be a $N \times n$ matrix with row i corresponding to example x_i , and let $\mathbf{y} = [y_1, \dots, y_N]^T$ (column vector containing the N training labels). Consider the problem of finding $\mathbf{w} \in \mathbf{R}^n$ minimizing

$$\|\mathbf{X} \cdot \mathbf{w} - \mathbf{y}\|^2 + \|\mathbf{w}\|^2,$$

where $\|\cdot\|$ is the Euclidean norm.

Does the regularization term $\|\mathbf{w}\|^2$ force the solution vector \mathbf{w} to have a small number of non-zero entries? Explain why or why not.

Problem 2

Describe a concept class C for which the halving algorithm is not optimal, i.e., you would get a better worst-case mistake bound by *not* going with the majority vote among the concepts in C consistent with examples observed so far. Explain your answer.

Problem 3

Describe a concept class C where the worst-case mistake bound of the halving algorithm ($\log |C|$) is not tight. Explain your answer.

Bonus: You will receive bonus points for examples significantly different from the ones appearing in the following paper:

Nick Littlestone, Learning Quickly When Irrelevant Attributes Abound: A New Linear-Threshold Algorithm, *Machine Learning*, 2(4): 285–318, 1998.

Copying any text from the paper will deterministically lead to a quiz. You can use examples from the paper, but you have to explain them yourself.

Problem 4

Describe a concept class C where *no* randomized algorithm can do better than $(\log |C|)/2$ mistakes in expectation (over the randomness in the algorithm). Explain your answer.

Problem 5

An online learning algorithm is *lazy* if it changes its state only when it makes a mistake. Show that for any deterministic online learning algorithm A achieving a mistake bound M with respect to a concept class C , there exists a lazy algorithm A' that also achieves M with respect to C .

Recall the definition: Algorithm A has a mistake bound M with respect to a learning class C if A makes at most M mistakes on *any* sequence that is consistent with a function in C .

Problem 6

Prove mistake bounds for the following two modifications to WINNOW, for the class of monotone disjunctions over n boolean variables. (Such a disjunction is a boolean function of the form $\bigvee_{i \in S} x_i$ for some subset $S \subseteq \{1, \dots, n\}$. Here \bigvee denotes the boolean OR function.) Assume that all labels are consistent with a monotone disjunction.

Given an example $x = (x_1, \dots, x_n)$,

- If the algorithm makes a mistake on a negative (predicts 1 when the correct label of x is 0), then for each x_i equal to 1 set $w_i = 0$.
- Whenever the correct label is 0 (regardless of whether the algorithm made a mistake or not), set $w_i = 0$ for each x_i equal to 1.

A formal proof is expected.