

MCS 548 – Mathematical Theory of Artificial Intelligence
Fall 2025
Problem Set 2

Lev Reyzin

Due: 10/24/25 at the beginning of class

Instructions: You may consult outside references, but cite all the resources used (e.g. which people or resources on the internet you consulted). You must, however, write up the answers on your own. All problems in this assignment require proof.

Problems

1. [10 pts.] Show that for AdaBoost $\prod_t Z_t$ is a monotonically decreasing function of t .
2. [10 pts.] Assume that the weak learning assumption of AdaBoost holds. Let h_t be the base learner selected at round t . Show that the base learner h_{t+1} selected by AdaBoost at round $t + 1$ must be different from h_t .
3. [10 pts.] Suppose that in boosting, the weak learning condition is guaranteed to hold so that $\epsilon_t \leq \frac{1}{2} - \gamma$ for some $\gamma > 0$ that is known before boosting begins. Describe a modified version of AdaBoost whose final classifier is an *unweighted* majority vote and whose training error is at most $(1 - 4\gamma^2)^{T/2}$.
4. [20 pts.] Assume we are given m examples $(x_1, y_1), \dots, (x_m, y_m)$ with $y_i \in \{-1, 1\}$ and that all x_i s are distinct. Consider training an SVM on this dataset with the kernel: $K(x, x') = \{1 \text{ if } x = x' \text{ and } 0 \text{ otherwise}\}$.
 - a. [5 pts.] Prove that K (as defined above) is positive definite symmetric. (This implies there is a function ϕ mapping each x to a high dimensional space so that $K(x, x') = \phi(x) \cdot \phi(x')$.)
 - b. [5 pts.] Explicitly compute the α_i s (as a function of the other specified quantities) that would be found by an SVM using this kernel to (implicitly) make the weight vector $\mathbf{w} = \sum_i \alpha_i y_i \phi(x_i)$.
 - c. [5 pts.] How many support vectors will this SVM have? What margin $\rho = 1/\|\mathbf{w}\|$ will this SVM achieve?
 - d. [5 pts.] What will this SVM do on unseen data? Do you expect this SVM to perform well? Give both an intuitive and a mathematical explanation.
5. [10 pts.] Suppose you have two coins, one perfectly fair, and one with bias toward H of $1/2 + \epsilon$ for some $\epsilon > 0$. It is known that to tell which coin is biased (with probability $> 3/4$) one needs to perform at least c/ϵ^2 coin flips ($c > 0$ is some constant). Show that this implies that asymptotic regret of $O(T^{1/2})$ achieved by Randomized Weighted Majority cannot be improved to $O(T^{1/2-\delta})$ for any constant $\delta > 0$.
6. [10 pts.] Imagine that an online learning algorithm A that runs in T rounds and has an expected regret bound of $\epsilon + T/\epsilon$, where ϵ is set by the algorithm. Clearly, the optimal setting is $\epsilon = \sqrt{T}$. The problem is that sometimes T is not known in advance. How do we fix this issue? We can have an algorithm A' that

does the following: A' starts with a parameter ϵ_1 and runs A for T_1 rounds, then adjusts the parameter to ϵ_2 and runs A for T_2 rounds, and so on. Construct a schedule of (ϵ_i, T_i) that asymptotically achieves the \sqrt{T} expected regret bound without knowing T in advance.