

1. Let the instance space be the set of real numbers. Consider on-line learning algorithms that do the following in each trial:
 - an instance $x \in X$ is received
 - the algorithm predicts a real number y
 - as a reinforcement the algorithm is told whether its prediction was correct and if it was wrong it is told whether it was too high or too low.

Now assume that you have N such algorithms. Let S be a sequence of instances/reinforcements. Let m_i be the number of mistakes of the i -th algorithm on S and let m_* be the number of mistakes of the best algorithm.

Design a master algorithm that combines the predictions of the algorithm which makes at most $O(\log N + m_*)$ many mistakes that holds for arbitrary sequences S .

Hint: The master algorithm uses exponential weights and predicts based on the Weighted Median. Prove a bound of

$$M_{\text{Alg}}(S) \leq \frac{\ln N + \eta m_i}{\ln \frac{2}{1+\beta}}.$$

Hint: The proof is very similar to the mistake bound for the Weighted Majority Algorithm.

2. Assume we have n experts whose predictions are binary and assume that the predictions of the master algorithm and the labels are binary as well. Give an adversary argument that for any **deterministic** master algorithm algorithm produces a sequence of predictions of the experts and labels on which the master makes $\geq \lfloor \log_2 n \rfloor$ mistakes and one of the n experts is consistent with the binary labels.
3. As above but now force any **deterministic** master algorithm to make $\geq 2k + \lfloor \log_2 n \rfloor$ mistakes while one of the experts makes $\leq k$ mistakes.

Hint: Prove the $2k + 1$ lower bound for the case of $n = 2$. Then use the previous problem to get the $2k + \lfloor \log_2 n \rfloor$ lower bound.

4. EXTRA CREDIT

On page 8 of the class notes 3 I proved a relative loss bound of

$$\sum_{t=1}^T \mathbf{w}_t \cdot \mathbf{L}_t \leq \min_{\sum_{i=1}^n u_i=1} \frac{\Delta(\mathbf{u}, \mathbf{w}_1) - \Delta(\mathbf{u}, \mathbf{w}_{T+1}) + \eta \mathbf{u} \cdot \mathbf{L}_{\leq T}}{1 - e^{-\eta}}$$

for the decision theoretic setting of expert algorithm that uses exponential weights¹ Then on page 10 I proved a seemingly weaker bound using the potential method

$$\sum_{t=1}^T \mathbf{w}_t \cdot \mathbf{L}_t \leq \min_{\sum_{i=1}^n u_i=1} \frac{\Delta(\mathbf{u}, \mathbf{w}_1) + \eta \mathbf{u} \cdot \mathbf{L}_{\leq T}}{1 - e^{-\eta}}$$

What is happening?

Show that the two bounds are the same, i.e.

$$\min_{\sum_{i=1}^n u_i=1} (\Delta(\mathbf{u}, \mathbf{w}_1) - \Delta(\mathbf{u}, \mathbf{w}_{T+1}) + \eta \mathbf{u} \cdot \mathbf{L}_{\leq T}) = \min_{\sum_{i=1}^n u_i=1} (\Delta(\mathbf{u}, \mathbf{w}_1) + \eta \mathbf{u} \cdot \mathbf{L}_{\leq T}).$$

Hint: Plug the exponential weights $w_{T+1,i} = \frac{w_{1,i} e^{-\eta \mathbf{L}_{\leq T,i}}}{Z_T}$ into the argument of the first min above.

¹ $\mathbf{L}_{\leq T} := \sum_{t=1}^T \mathbf{L}_t$