CS272 - Theoretical Foundations of Learning, Decisions, and Games

Lecture 24: Multi-distribution Learning II

25th November, 2025

Lecturer: Nika Haghtalab Readings: [Nguyen and Zakynthinou, 2018, Haghtalab et al., 2022]
Scribe: Rhys Gould

1 Motivation

Last time, we introduced multi-distribution learning framework of Haghtalab et al. [2022]. Given distributions $D_1, D_2, \ldots, D_k$, we want to find a hypothesis $h \in \mathcal{H}$ such that

$$\max_{i\in[k]} L_{D_i}(h) \leq \min_{h^*\in\mathcal{H}} \max_{i\in[k]} L_{D_i}(h^*) + \epsilon.$$

for a desired sub-optimality gap $\epsilon > 0$. We stated the following sample complexity bounds:

Setting	**Single Distribution**	**Multi-Distribution**
Realizable	$\frac{d+\ln(1/\delta)}{\epsilon}$	$\frac{(d+k)\ln(k/\delta)}{\epsilon}$
Agnostic	$\frac{d+\ln(1/\delta)}{\epsilon^2}$	$\frac{(d+k)\ln(k/\delta)}{\epsilon^2}$

Last time, we only proved a bound of $\frac{(d+k)\ln(k/\delta)}{\epsilon^4}$ for the multi-distribution realizable PAC case as a warmup. Today, we will prove the bounds that are stated in the table.

2 The MinMax Formulation

Recall that as a technique we view multi-distribution learning as a minmax equilibrium and use no-regret dynamics to obtain a tighter bound.

The multi-distribution learning objective can be written as a two-player game:

$$\max_{i\in[k]} L_{D_i}(h) \leq \underbrace{\min_{h^*\in\mathcal{H}}}_{\substack{\text{minimizing}\\\text{player}}} \underbrace{\max_{i\in[k]}}_{\substack{\text{maximizing}\\\text{player}}} L_{D_i}(h^*) + \epsilon.$$

Different combinations of strategies for the two players yield different sample complexities:

Minimizing Player	Maximizing Player	Sample Complexity
Generic Best Response	Generic No-Regret	$\frac{(d+k)\ln(k/\delta)}{\epsilon^4}$
PAC Best Response	Specialized No-Regret	$\frac{(d+k)\ln(k/\delta)}{\epsilon}$ (realizable)
Generic No-Regret	Generic No-Regret	$\frac{(\log(\lvert\mathcal{H}\rvert)+k)\ln(k/\delta)}{\epsilon^2}$ (agnostic)

3 Improved Bounds for the Realizable PAC setting

In the last lecture, we proved the first row, with the minimizing player performing ERM over a sample set while the maximizing player used the randomized weighted majority algorithm. In the following, we will perform a tighter analysis for the realizable setting based on *multiplicative concentration* and no-regret learning via weighted majority. We will consider the following algorithm for the realizable case. This algorithm is due to Nguyen and Zakynthinou [2018].

Algorithm 1 Multi-Distribution Learning (Realizable Case)

Initialize $w_i^{(1)} = 1$ for all $i \in [k]$
for $t = 1, \ldots, T = O(\log(k/\delta))$ **do**
  $W^{(t)} \leftarrow \sum_{i=1}^k w_i^{(t)}$
  $P^{(t)} \leftarrow \frac{1}{W^{(t)}} \sum_{i=1}^k w_i^{(t)} D_i$
  Sample $S^{(t)} \sim P^{(t)}$ of size $m_{\epsilon'/16,\delta'}$
  $h^{(t)} \leftarrow \text{ERM}_{\mathcal{H}}(S^{(t)})$ ▷ Approximating the best response
  for $i = 1, \ldots, k$ **do** ▷ Loop for updating weights
    Sample $T_i \sim D_i$ of size $O(\ln(k/\delta)/\epsilon)$
    if $L_{T_i}(h^{(t)}) \geq \frac{3\epsilon'}{4}$ **then**
      $w_i^{(t+1)} \leftarrow 2w_i^{(t)}$
    else
      $w_i^{(t+1)} \leftarrow w_i^{(t)}$
    end if
  end for
end for
Return $\bar{h} = \text{Majority}(h^{(1)}, h^{(2)}, \ldots, h^{(T)})$

A large weight $w_i^{(t)}$ indicates that the distribution D_i has not been well-learned and deserves more focus. If a distribution has been fully learned, we would still like to keep $w_i^{(t)}$ non-zero in order to avoid gradually forgetting D_i over time. The sampling step $S^{(t)} \sim P^{(t)}$ allows us to approximate the best response (ideally we would want $L_{D_i}(h^{(t)}) = 0$ for all i, but this is impossible without infinite samples, so we approximate via the ERM).

3.1 Algorithmic Guarantees

We now establish several claims about Algorithm 1's correctness. We skip the proofs of the next two lemmas, as they are standard applications of sample complexity bounds and concentration inequalities that by now you can do on your own.

Lemma 3.1. *With probability $1 - \delta' T$, for all $t \in [T]$:*

$$L_{P^{(t)}}(h^{(t)}) \leq \frac{\epsilon'}{16}.$$

Proof (outline). This follows from a union bound over all T rounds, using the fact that ERM on $S^{(t)}$ achieves a low loss on $P^{(t)}$ in the realizable setting. □

Definition 3.2. Define the "bad set" at time t as:

$$\text{Bad}^{(t)} = \{i : w_i^{(t+1)} = 2w_i^{(t)}\},$$

i.e., the set of distributions whose weight was doubled at step t.

Lemma 3.3. *With probability $1 - \delta' T$, for all $i \in [k]$ and $t \in [T]$:*

- *If $L_{D_i}(h^{(t)}) \geq \epsilon'$, then $i \in$ Bad$^{(t)}$.*
- *If $L_{D_i}(h^{(t)}) \leq \epsilon'/2$, then $i \notin$ Bad$^{(t)}$.*

Proof (outline). This follows from a multiplicative Chernoff bound on the empirical loss estimates. □

Note that not being in $\text{Bad}^{(t)}$ is definitely good, but being in $\text{Bad}^{(t)}$ is not necessarily bad: the true loss may lie in the interval $(\epsilon'/2, \epsilon')$. From now on, we will say "with high probability" instead of "with probability $1 - \delta' T$" for brevity while accounting that the good events in previous lemmas have happened.

Lemma 3.4. *With high probability:*

$$W^{(t)}_{\textit{Bad}^{(t)}} := \sum_{i \in \textit{Bad}^{(t)}} w_i^{(t)} \leq \frac{1}{8} W^{(t)}.$$

Proof. Given that $L_{P^{(t)}}(h^{(t)}) \leq \epsilon'/16$, the total weight on distributions D_i for which $L_{D_i}(h^{(t)}) \geq \epsilon'/2$ is at most $1/8$ of the total weight by Markov's inequality. The proof follows by Lemma 3.3. □

An implication of Theorem 3.4 is that, at every timestep t, we have:

$$W^{(t+1)} = (W^{(t)} - W^{(t)}_{\text{Bad}^{(t)}}) + 2W^{(t)}_{\text{Bad}^{(t)}} = W^{(t)} + W^{(t)}_{\text{Bad}^{(t)}} \leq \frac{9}{8} W^{(t)}.$$

which we will apply recursively in the following lemma.

Lemma 3.5. *With high probability, for each i, the number of timesteps t where $i \in$ Bad$^{(t)}$ is at most $0.4T$.*

Proof. For a particular i:

$$w_i^{(T)} = 2^{|\{t:i\in \text{Bad}^{(t)}\}|} \leq W^{(T)} \leq \left(\frac{9}{8}\right)^T \cdot k.$$

Taking logarithms:

$$|\{t : i \in \text{Bad}^{(t)}\}| \leq T \log_2(9/8) + \log_2(k) \leq 0.4T,$$

for appropriately large T as we defined in the algorithm. □

An implication of Theorem 3.5 is that each distribution i is in the bad set less than half the time, which suggests that majority voting should work, as we will now verify.

We now analyze the majority vote on any D_i i.e., $\bar{h}$: Take any D_i and an $x \in \mathcal{X}$ from it. Supposed that x is mislabeled by Majority$(h^{(1)}, \ldots, h^{(T)})$. Then x was mislabeled by at least $0.5T$ of the hypotheses, but according to Lemma 3.5, at most $0.4T$ of these can come from "bad" rounds. Therefore, x was mislabeled by at least $0.1T$ of the leftover $0.6T$ functions $h^{(t)}$ that were not bad. As a consequence, the majority's error on D_i satisfies:

$$L_{D_i}(\bar{h}) \leq \frac{0.6}{0.1} \times \epsilon' \leq \epsilon.$$

4 Overview of the Approach for the Agnostic Setting

Next, we turn our attention to the agnostic case. We want to improve the warm up guarantees we proved in the last lecture that led to a $(d + k)/\epsilon^4$ rate. We introduce an algorithm by Haghtalab et al. [2022].

Recall that the previous lecture's approach used a no-regret algorithm against a best-response algorithm. The choice of a best-response algorithm was appropriate for ensuring that a near-optimal classifier was learned at every step for the corresponding distribution. However, this required taking $O(d/\epsilon^2)$ samples per round (over $T = \log(k)/\epsilon^2$ rounds).

We now take an alternative perspective: instead of enforcing per-step accuracy (which cumulatively leads to ϵT total error and as result excessive sample complexity), we consider randomized strategies whose per-step estimates are unbiased. After T timesteps, the total error does not grow additively; instead, it concentrates. In other words, we choose to allow high-variance but unbiased and independent errors that accumulate over many timesteps, whereas in the previous approach we chose low-variance but correlated estimators.

4.1 Setup

Minimizing player: Plays mixed strategy $\alpha^{(t)} \in \Delta(\mathcal{H})$ which then takes hypothesis $h^{(t)} \sim \alpha^{(t)}$. The sequence $\alpha^{(1)}, \ldots, \alpha^{(T)}$ satisfies no-regret over the sequence $D_{i_1}, \ldots, D_{i_T}$ chosen by the maximizing player (described below).

Crucially, the choice of $h^{(t)}$ is trained only on on history: $(x^{(1)}, y^{(1)}), \ldots, (x^{(t-1)}, y^{(t-1)})$ sampled from $D_{i_1}, \ldots, D_{i_{t-1}}$ respectively. That is, it's independent of the choice of D_{i_t}.

Evaluation of $h^{(t)}$ is done on $(x^{(t)}, y^{(t)}) \sim D_{i_t}$ The minimizing player has *full feedback* during evaluation: for all $h \in \mathcal{H}$, they have an unbiased estimator of $L_{P^{(t)}}(h)$.

Maximizing player: Mixed strategy $\beta^{(t)} \in \Delta_k$, with distribution index $i_t \sim \beta^{(t)}$. The sequence $i_1, \ldots, i_T$ satisfies no-regret over $h^{(1)}, \ldots, h^{(T)}$.

Note that i_t only observes $h^{(1)}, \ldots, h^{(t-1)}$ (and not $h^{(t)}$), so only sees $(x^{(1)}, y^{(1)}), \ldots, (x^{(t-1)}, y^{(t-1)})$.

The maximizing player has *bandit feedback*: only an unbiased estimator for the specific arm i_t that was pulled.

Notation. Let $z_{i_t}^{(t)} := (x_{i_t}^{(t)}, y_{i_t}^{(t)}) \sim D_{i_t}$, and $\ell(h, (x, y)) := \mathbf{1}\{h(x) \neq y\}$. For brevity we will define the collections $\alpha = \{\alpha^{(t)}\}_{t=1}^T$, $\beta = \{\beta^{(t)}\}_{t=1}^T$, and $z = \{z_{i_t}^{(t)}\}_{t=1}^T$ for $i_t \sim \beta^{(t)}$.

4.2 Regret Guarantees

Lemma 4.1. *For $i_t \sim \beta^{(t)}$, define the empirical regrets:*

$$\widehat{\textit{Regret}}_{\min}(\alpha, \beta, z) := \sum_{t=1}^{T} \ell(\alpha^{(t)}, z_{i_t}^{(t)}) - \min_{h^* \in \mathcal{H}} \sum_{t=1}^{T} \ell(h^*, z_{i_t}^{(t)}),$$

$$\widehat{\textit{Regret}}_{\max}(\alpha, \beta, z) := \max_{i^* \in [k]} \sum_{t=1}^{T} \ell(\alpha^{(t)}, z_{i^*}^{(t)}) - \sum_{t=1}^{T} \ell(\alpha^{(t)}, z_{i_t}^{(t)}).$$

Then with high probability:

$$\widehat{\textit{Regret}}_{\min}(\alpha, \beta, z) \leq O\left(\sqrt{T \ln(|\mathcal{H}|/\delta)}\right),$$

$$\widehat{\textit{Regret}}_{\max}(\alpha, \beta, z) \leq O\left(\sqrt{Tk \ln(1/\delta)}\right).$$

Proof sketch. We will not prove Lemma 4.1 in detail, but the guarantees follow directly from simple observations we have already made. The minimizing player operates in a full-information setting and therefore achieves expected regret $O\left(\sqrt{T \log |\mathcal{H}|}\right)$. The maximizing player, in contrast, operates in a bandit-information setting and as a result obtains expected regret $O\left(\sqrt{Tk}\right)$ as well. The parameter δ and the dependence of the regret bounds on it account for the high-probability guarantees rather than merely the expected guarantees.

□

The key insight is that we do not need ℓ to approximate L_{D_i} well at every time step t, as we did in last lecture. Instead, we only need **the empirical regrets to approximate the true regrets well**.

If one knew $L_{D_i}(h)$ for every h and i, we could consider the *true regrets*:

$$\text{Regret}_{\min}(\alpha, \beta) := \sum_{t=1}^{T} L_{\beta^{(t)}}(\alpha^{(t)}) - \min_{h^* \in \mathcal{H}} \sum_{t=1}^{T} L_{\beta^{(t)}}(h^*),$$

$$\text{Regret}_{\max}(\alpha, \beta) := \max_{i^* \in [k]} \sum_{t=1}^{T} L_{D_{i^*}}(\alpha^{(t)}) - \sum_{t=1}^{T} L_{\beta^{(t)}}(\alpha^{(t)}).$$

Lemma 4.2. *With probability* $1 - \delta$*:*

$$\left| \widehat{\textit{Regret}}_{\max}(\alpha, \beta, z) - \textit{Regret}_{\max}(\alpha, \beta) \right| \leq \sqrt{T \ln(k/\delta)},$$

$$\left| \widehat{\textit{Regret}}_{\min}(\alpha, \beta, z) - \textit{Regret}_{\min}(\alpha, \beta) \right| \leq \sqrt{T \ln(|\mathcal{H}|/\delta)}.$$

Proof sketch. Note that $\ell(\alpha^{(t)}, z_{i_t}^{(t)})$ is an unbiased estimator of $L_{\beta^{(t)}}(\alpha^{(t)})$. This holds because $\alpha^{(t)}$ is independent of $\beta^{(t)}$, since $\alpha^{(t)}$ was determined only by $\beta^{(1)}, \ldots, \beta^{(t-1)}$. This is a key property of playing two no-regret algorithm against each other instead of playing best-response algorithm which would have picked α_t by looking at β_t. Consequently, $\alpha^{(t)}$ is independent of $z_{i_t}^{(t)}$, conditioned on history.

This allows us to use martingales to bound the two regrets:

$$\left| \sum_{t=1}^{T} \ell(\alpha^{(t)}, z_{i_t}^{(t)}) - \sum_{t=1}^{T} L_{P^{(t)}}(\alpha^{(t)}) \right| \leq \sqrt{T \ln(1/\delta)}.$$

For the maximizing player: once $\alpha^{(1)}, \ldots, \alpha^{(T)}$ is fixed, for any distribution D_{i^*}, with high probability:

$$\max_{i^* \in [k]} \left| \sum_{t=1}^{T} \ell(\alpha^{(t)}, z_{i^*}^{(t)}) - \sum_{t=1}^{T} L_{D_{i^*}}(\alpha^{(t)}) \right| \leq \sqrt{T \ln(k/\delta)}.$$

An analogous bound holds for the minimizing player with $|\mathcal{H}|$ in place of k. □

References

Nika Haghtalab, Michael Jordan, and Eric Zhao. On-demand sampling: Learning optimally from multiple distributions. *Advances in Neural Information Processing Systems*, 35:406–419, 2022.

Huy Nguyen and Lydia Zakynthinou. Improved algorithms for collaborative pac learning. *Advances in Neural Information Processing Systems*, 31, 2018.