Learning to Answer from Correct Demonstrations

Anonymous Author(s)
Affiliation
Address
email

Abstract

We study the problem of learning to generate an answer (or completion) to a question (or prompt), where there could be multiple correct answers, any one of which is acceptable at test time. Learning is based on demonstrations of some correct answer to each training question, as in Supervised Fine Tuning (SFT). We formalize the problem as apprenticeship learning (i.e., imitation learning) in contextual bandits, with offline demonstrations from some expert (optimal, or very good) policy, without explicitly observed rewards. In contrast to prior work, which assumes the demonstrator policy belongs to a low-complexity class, we propose relying only on the underlying reward model (i.e., specifying which answers are correct) being in a low-cardinality class, which we argue is a weaker assumption. We show that likelihood-maximization methods can fail in this setting, and instead present an approach that learns to answer nearly as well as the demonstrator, with sample complexity logarithmic in the cardinality of the reward class. Our method is similar to Syed and Schapire [2007], when adapted to a contextual bandit (i.e., single step) setup, but is a simple one-pass online approach that enjoys an "optimistic rate" (i.e., $1/\varepsilon$ when the demonstrator is optimal, versus $1/\varepsilon^2$ in Syed and Schapire), and works even with arbitrarily adaptive demonstrations.

Please refer to arxiv.2510.15464 for the most up to date version.

1 Introduction

Large Language Models (LLMs) are typically pretrained using maximum likelihood estimation (MLE) to model the conditional distributions of the next token in text Vaswani et al. [2017], Radford et al. [2019]. MLE has long been regarded as the "default" approach to density estimation, and it underpins much of modern machine learning Wald [1949], Cramér [1946], Wilks [1938], Neyman and Pearson [1928], Lehmann and Casella [1998]. However, many downstream applications, such as question answering and code completion, only require the LLM to produce only *a single valid completion* per prompt and *not* an accurate reproduction of the conditional distribution of completions. Thus, there is an apparent disparity between how LLMs are trained and their predominant use case. Motivated by this, we formalize a new learning objective called Precise Completion: finding a high-quality predictor that generates completions which lie in the support of ground-truth good responses.

The name Precise Completion is inspired by the foundational metric of *precision* in machine learning systems Manning [2008]. The *precision* metric is often paired with the *recall* metric, which roughly speaking measures how well the learner "covers" the entire support of valid outputs. Achieving both high precision and recall is challenging and often at odds: we have both theoretical Cohen et al. [2024], Charikar and Pabbaraju [2024], Kalavasis et al. [2025] and empirical Bronnec et al. [2024] evidence of this. Furthermore, in the standard training pipeline, LLMs undergo extensive alignment/post-training, and it is unclear whether the resulting models maintain

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

any guarantee of coverage over correct responses, i.e., high recall. Moreover, this exactly matches real-world usage of deployed LLMs such as GPT-4 OpenAI [2023], Claude Anthropic [2024], Gemini DeepMind [2023], DeepSeek DeepSeekAI [2025], and Meta's LLaMA Touvron et al. [2023], etc., where the feedback signal to the learner is based solely on the quality of the one output shown to the user. This motivates the study of the precision-only objective – which focuses on return any single valid completion – as a fundamental and important problem in its own right. We ask the following question:

What are the statistical limits of Precise Completion*, and what algorithms achieve them?*

Our Framework. We study the prompt-completion formulation for LLMs Ouyang et al. [2022], Rafailov et al. [2023], Huang et al. [2025], where the goal is to produce a high quality response $\widehat{y}$ for a prompt x. Formally, let $\mathcal{X}$ be the set of all possible prompts (questions, instructions, etc.) and $\mathcal{Y}$ be the set of all possible completions (answers, responses, etc.). There is an unknown ground truth $\sigma_\star : \mathcal{X} \to 2^{\mathcal{Y}}$ support function, which maps every $x \in \mathcal{X}$ to a support "good responses" $\sigma_\star(x) \subseteq \mathcal{Y}$. The learner observes $(x_i, y_i) \sim_{iid} \mathcal{D} \times \widetilde{\pi}$, from an unknown joint distribution supported on good responses, i.e. $x_i \sim \mathcal{D}$ and $y_i \sim \widetilde{\pi}(\cdot \mid x_i)$ where $\operatorname{supp} \widetilde{\pi}(\cdot \mid x) \subseteq \sigma_\star(x)$ for every prompt $x \in \mathcal{X}$. The goal of the learner is to output a (possibly stochastic) predictor $\widehat{\pi}$ whose loss is measured as

$$\textsf{Precise Completion}\text{ Loss:} \quad L_{\mathcal{D},\sigma_\star}(\widehat{\pi}) = \mathbb{E}_{x\sim\mathcal{D},\widehat{y}\sim\widehat{\pi}(\cdot|x)}\left[\mathbb{1}\{\widehat{y} \notin \sigma_\star(x)\}\right] . \tag{1}$$

This loss only captures the probability of the event of failing to output a good response, which is the criterion models are evaluated on in practice. It does not capture any distributional distance between the distributions over responses of learner ($\widehat{\pi}$) and demonstrator ($\widetilde{\pi}$), which models are never directly evaluated for, nor do we believe such evaluation is even possible.

We are interested in learning algorithms utilizing only *in-support observations*, that are available during pretraining or supervised fine tuning (SFT), without any other type of feedback available during post-training. The typical approach is to do density estimation of good responses via MLE. Our goal is to understand potential drawbacks of this approach where the objective of interest is Precise Completion, and find optimal statistical limits for this objective from good demonstrations.

1.1 Contributions

We study the Precise Completion problem from a learning-theoretic perspective. We distinguish between two types of function approximations to model the demonstrator (Section 2): (a) the class of conditional densities Π, with an unknown $\widetilde{\pi} \in \Pi$, and (b) the class of support functions $\mathcal{S}$, with a ground-truth $\sigma_\star \in \mathcal{S}$ and some $\widetilde{\pi}$ supported on $\sigma_\star$. We ask what are the optimal statistical estimators under standard *cardinality-based capacity controls* for these two natural types of function classes, for our Precise Completion problem. Our inquiry reveals an interesting algorithmic landscape as well as gaps (see also Table 1).

- **Capacity control via $|\Pi| = d$:** MLE is minimax optimal with respect to the cardinality parameter d for the Precise Completion problem (Section 3). In fact, it can achieve both precision and recall Cohen et al. [2024], or more generally optimize any bounded reward objective Foster et al. [2021]. In practice, however, MLE is often observed to fail at producing even precise responses, indicating that this coarse picture fails to capture the practical shortcomings of MLE. This motivates our next question: what performance guarantees can be established for MLE under a weaker form of capacity control, as captured below?
- **Capacity control via $|\mathcal{S}| = d$:** We show that MLE spectacularly fails at Precise Completion (Section 4.1), even though there is enough statistical information in the samples to start generating Precise Completion (Section 4.2). We also introduce a new way to utilize these samples, yielding a learner that achieves the optimal dependence on $|\mathcal{S}| = d$ in its sample complexity (Section 4.3). Interestingly, at this optimal limit, MLE can achieve a hallucinated overlap with the good responses (Remark 1), providing a theoretical support to a common empirical observation Ji et al. [2023].

See also the discussion in Section 6 for a broader contribution in relation to prior work.

Learning Rule	$\lvert\Pi\rvert = d < \infty$	$\lvert\mathcal{S}\rvert = d < \infty$
MLE	$\log d$ (Section 3)	May not be learnable (Section 4.1) (just overlap with $\log d$, Remark 1)
COMMON-INTERSECTION	MLE is optimal	d (Section 4.2)
Sample Efficient Leaner	MLE is optimal	$\log d$ (Section 4.3)

Table 1: Comparison of learning rules under two types of capacity control.

2 Setting

Let $\mathcal{X}$ and $\mathcal{Y}$ respectively be any countable spaces of all possible prompts and completions when learning from good demonstrations. Recall the setup introduced in Section 1 of a ground-truth support function $\sigma_\star : \mathcal{X} \to 2^{\mathcal{Y}}$, a marginal distribution $\mathcal{D} \in \Delta(\mathcal{X})$ and conditional distribution $\widetilde{\pi} : \mathcal{X} \to \Delta(\mathcal{Y})$. We use $(x, y) \sim \mathcal{D} \times \pi$ to denote a joint distribution where $x \sim \mathcal{D}$ and $y \sim \pi(\cdot \mid x)$. We observe m i.i.d. samples $S = \{(x_i, y_i) \sim_{iid} (\mathcal{D} \times \widetilde{\pi}) : i \in [m]\}$, and the goal of the learner is to start generating a *single correct responses* from samples, on new unseen prompts x (to be formalized in Definition 1), in the support $\sigma_*(x)$.

Two Function Approximations. It is important to note that our setup makes a distinction between the *support* of good responses $\sigma_\star$ which is a deterministic set-valued function, and the *conditional distribution* of the demonstrator $\widetilde{\pi}$. This forms a basis of our next investigation. We take the learning-theoretic view of Precise Completion problem, and model the expert demonstrator with a hypothesis class. Keeping the distinction between $\sigma_\star$ and $\widetilde{\pi}$ in mind, there are two types of function approximations one can consider even in the realizable setting:

- **Support Function Approximation.** There is a class of support functions $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ and the unknown $\sigma_* \in \mathcal{S}$. The demonstrator shows examples according to some unknown conditional distribution supported on $\sigma_\star$ (i.e. $\operatorname{supp} \widetilde{\pi}(\cdot \mid x) \subseteq \sigma_\star(x)$).
- **Conditional Density Approximation.** There is a class of conditional distributions $\Pi \subseteq (\Delta(\mathcal{Y}))^{\mathcal{X}}$ such that the unknown $\widetilde{\pi} \in \Pi$. One can consider $\sigma_\star(x) = \sigma_{\widetilde{\pi}}(x) := \operatorname{supp} \widetilde{\pi}(\cdot \mid x)$ as the reference ground-truth support function for evaluating the loss (Eq.(1)).

Contrasting views. While both the views of the support hypothesis class $\mathcal{S}$ and the conditional distribution hypothesis class Π are closely related—they agree on the realizability assumption and can be converted into one another—for every $\mathcal{S}$, one can consider the conditional density class $\Pi_{\mathcal{S}} := \bigcup_{\sigma \in \mathcal{S}} \Pi_\sigma$, where $\Pi_\sigma := \{\text{Any } \pi : \mathcal{X} \to \Delta(\mathcal{Y}) \text{ s.t. } \operatorname{supp} \pi(\cdot \mid x) \subseteq \sigma(x)\,,\ \forall x \in \mathcal{X}\}$, and for every Π one can consider the associated support class $\mathcal{S}_\Pi := \bigcup_{\pi \in \Pi} \{\sigma_\pi \mid \sigma_\pi(x) = \operatorname{supp} \pi(\cdot \mid x)\, \forall x \in \mathcal{X}\}$[1]—they differ philosophically in what they treat as the primary object of function approximation. The support function approximation directly targets the set of good completions, which is the natural object for our Precise Completion loss (1). In contrast, the conditional desnsity approximation posits a latent density over completions, which naturally motivates learning via density estimation. This difference also mirrors contemporary LLM practices: training a reward model is analogous to estimating $\sigma_\star$ (or a scoring function over $\mathcal{Y}$ for each x), while training a policy model corresponds to fitting for $\widetilde{\pi} \in \Pi$.

Learning from In-support Demonstrations. In either cases, the goal is to design a (possibly) stochastic predictor $\widehat{\pi}(S) : \mathcal{X} \to \Delta(\mathcal{Y})$ from i.i.d. samples $S = \{(x_i, y_i) \sim_{iid} (\mathcal{D} \times \widetilde{\pi}) : i \in [m]\}$ in order to minimize the Precise Completion loss according to Eq.(1). More formally, we will work under the Probably Approximately Correct (PAC) framework.

Definition 1 (Probably Approximately Precise Completion)**.** *Let* $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ *(respectively* $\Pi \subseteq (\Delta(\mathcal{Y}))^{\mathcal{X}}$*) be the hypothesis classes as defined in the above setups. We say that* $\mathcal{S}$ *(respectively* Π*) is Probably Approximately Precise completable from in-support demonstrations by an estimator* $\widehat{\pi} : (\mathcal{X} \times \mathcal{Y})^* \to (\Delta(\mathcal{Y}))^{\mathcal{X}}$ *with sample complexity* $m_{\mathcal{S},\widehat{\pi}} : (0,1) \times (0,1) \to \mathbb{N}$ *(respectively* $m_{\Pi,\widehat{\pi}}$*),*

[1] Note that for every $\pi \in \Pi$, there is unique σ_π, however, for every $\sigma \in \mathcal{S}$, the class of conditional law Π_σ supported on σ may be huge.

if for any $\varepsilon, \delta \in (0,1)$, for any sample size $m \geq m_{\mathcal{S},\widehat{\pi}}(\varepsilon,\delta)$ (respectively $m_{\Pi,\widehat{\pi}}$), for any choice of $\mathcal{D}, \sigma_\star, \widetilde{\pi}$ (respectively $\mathcal{D}, \widetilde{\pi}$), we have

$$\mathbb{P}_{S\sim(\mathcal{D}\times\widetilde{\pi})^m}\left[L_{\mathcal{D},\sigma_\star}(\widehat{\pi}(S)) \leq \varepsilon\right] \geq 1-\delta\,,$$

where $\widehat{\pi}(S)$ is the (possibly stochastic) predictor output by the estimator $\widehat{\pi}$ on the input S.

We note that our problem also corresponds to *imitation learning* for a simple *contextual bandit* problem with *binary rewards*[2], where the context space is $\mathcal{X}$ and the action space is $\mathcal{Y}$, however, in the regime where $\mathcal{X}$ and $\mathcal{Y}$ spaces are huge or possibly countably infinite.

3 Warm Up: Bounded Density Class Cardinality

The front where the above two types of function approximations introduced in Section 2 differ significantly is when one makes the capacity control assumption on the two (e.g. $|\Pi| < \infty$ is a much stronger control than $|\mathcal{S}| < \infty$). A natural starting point is to consider the finite cardinality conditional density class $|\Pi| = d < \infty$, similar to recent works in the literature Cohen et al. [2024], Foster et al. [2024] and beyond imitation learning Yun et al. [2025], Zhan et al. [2023], Xie et al. [2024], Zhang et al. [2025], Agarwal et al. [2025], Huang et al. [2024]. This also ensures the capacity control of the associated support class $|\mathcal{S}_\Pi| \leq |\Pi|$. The conditional density class naturally leads us to the density estimation based approach via maximum likelihood estimation:

$$\mathrm{MLE}_\Pi(S) = \arg\max_{\pi\in\Pi} \prod_{i=1}^{m} \pi(y_i \mid x_i) = \arg\min_{\pi\in\Pi} -\sum_{i=1}^{m} \log \pi(y_i \mid x_i)\,. \tag{MLE}$$

For the MLE, we have $D_{\mathsf{TV}}(\widehat{\pi}_{\mathrm{mle}}, \widetilde{\pi}) \to 0$ (also implying $L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_{\mathrm{mle}}) \to 0$) as $m \to 0$ giving us consistency. One can ask whether MLE is also minimax optimal among the family of finite classes of size d and what is the sample complexity of learning in the following sense:

$$\sup_{|\Pi|=d<\infty} \inf_{\widehat{\pi}} m_{\Pi,\widehat{\pi}}(\varepsilon,\delta)\,. \tag{2}$$

Indeed, the MLE turns out to be also optimal in this sense (up to universal constants).

Proposition 1. *Consider any hypothesis class $\Pi \subseteq (\Delta(\mathcal{Y}))^{\mathcal{X}}$ of conditional densities with $|\Pi| = d < \infty$. Then for any unknown marginal distribution $\mathcal{D}$, conditional law $\widetilde{\pi} \in \Pi$, and any $\varepsilon, \delta \in (0,1)$, with probability $1-\delta$ over $S \sim (\mathcal{D}\times\widetilde{\pi})^m$, for any $\widehat{\pi}_{\mathrm{mle}}(S) \in \mathrm{MLE}_\Pi(S)$:*

$$L_{\mathcal{D},\sigma_{\widetilde{\pi}}}(\widehat{\pi}_{\mathrm{mle}}(S)) \leq \frac{6\log(2d/\delta)}{m}\,,$$

Thus Π is learnable with $\widehat{\pi}_{\mathrm{mle}}$ (cf. Definition 1) with sample complexity: $m_{\Pi,\widehat{\pi}_{\mathrm{mle}}}(\varepsilon,\delta) = \frac{6}{\varepsilon}\log(2d/\delta)\,.$

This is a strong guarantee in that it enjoys no dependence on the $|\mathcal{X}|, |\mathcal{Y}|$, or even $\sup_{\pi,x} |\mathrm{supp}\pi(\cdot \mid x)|$, which in our case can be huge. We provide an intuition for why it can enjoy such a guarantee for the special cases of Π, without relying on the black-box of density estimation that uses the convergence in the Hellinger distance of $\widehat{\pi}_{\mathrm{mle}}$ to $\widetilde{\pi}$ in Section B, followed by the proof of any general Π that relies on it. The proof uses the ideas from Foster et al. [2024]; one can first use the standard guarantees in density estimation to establish the convergence in the squared Hellinger distance for $D^2_{\mathsf{H}}(\widehat{\pi}_{\mathrm{mle}}, \widetilde{\pi})$, followed by controlling the Precise Completion loss (Eq. (1)) in terms of $D^2_{\mathsf{H}}(\widehat{\pi}_{\mathrm{mle}}, \widetilde{\pi})$.

It is not too difficult to establish that $\Omega(\log d/\varepsilon)$ is samples are also necessary in the worst-case for any estimator, in order to get the expected error of at most ε. The construction simply follows from the lower bound for the standard supervised binary classification problem, which is a special case of our problem. I.e. there exists an instance Π of size d such for any estimator $\widehat{\pi}$:

$$\inf\ m \ \text{ s.t. } \sup_{\mathcal{D},\widetilde{\pi}\in\Pi} \mathbb{E}_{S\sim(\mathcal{D}\times\widetilde{\pi})^m}[L_{\mathcal{D},\sigma_\star}(\widehat{\pi}(S))] \leq \varepsilon\,, \text{ is } \ \Omega(\tfrac{\log d}{\varepsilon})\,.$$

[2]To some extent, our results can be generalized to general bounded non-binary rewards (see Remark 3).

This establishes MLE as the minimax optimal estimator for the family of finite classes of size d in the sense of (2). In fact, MLE at this sample complexity not only achieves a good performance according to our Precise Completion loss. It also achieves both precision and recall Cohen et al. [2024], or in fact, a good performance for any bounded reward function Foster et al. [2024], because MLE just converges in the Hellinger distance as discussed in the proof sketch.

However, in practice, pretrained or supervised fine-tuned models are not even known to reliably produce precise responses. This indicates that the finite-class viewpoint $|\Pi| < \infty$ and the associated optimality of MLE may be too crude to capture their behavior under the Precise Completion (Eq. (1)) loss. We are thus ask: how does MLE perform once we impose capacity control on the support class instead?

4 Main Results: Bounded Support Class Cardinality

Recall the setup of the support hypothesis class $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$. We control its capacity by assuming finite cardinality $|\mathcal{S}| = d < \infty$. Nature selects a ground-truth set-valued function $\sigma_\star \in \mathcal{S}$ that specifies the valid responses, and provides i.i.d. examples drawn from some unknown joint distribution $(\mathcal{D} \times \widetilde{\pi})$ over in-support pairs (i.e., $\mathrm{supp}(\widetilde{\pi}(\cdot \mid x)) \subseteq \sigma_\star(x)$, or equivalently $\widetilde{\pi} \in \Pi_{\sigma_\star}$). The learner's objective is to output in-support completions on unseen instances (cf. Definition 1).

This can occur in practice: when creating QA datasets, practitioners often hand-pick good responses for each prompt, without any attempt to enumerate all valid completions or generate from the full set of valid responses according to a fixed apriori distribution. Thus, imposing any further distributional assumptions on this sampling may be overly restrictive.

4.1 Simple Failures of natural choices of MLE

The problem is defined by the support class $\mathcal{S}$ and the promise that samples come from some $\widetilde{\pi}$ supported on the unknown $\sigma_\star \in \mathcal{S}$, without specifying a conditional density class Π. A natural idea is to perform density estimation over

$$\Pi_{\mathcal{S}} := \bigcup_{\sigma \in \mathcal{S}} \Pi_\sigma$$

via MLE, i.e. $\mathrm{MLE}_{\Pi_{\mathcal{S}}}(S) = \arg\max_{\pi \in \Pi_{\mathcal{S}}} \prod_{(x_i, y_i) \in S} \pi(y_i \mid x_i)$. However, this approach already fails on a simple instance.

Theorem 1 (MLE Failure 1)**.** *There exists $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ with $|\mathcal{S}| = |\mathcal{Y}| = 2$ and $\mathcal{X} = \mathbb{N}$ and a choice of $(\sigma_\star, \widetilde{\pi})$ such that for every sample size m and $\gamma \in (0, 1)$, there exists a marginal distribution $\mathcal{D}$ such that for $S \sim (\mathcal{D} \times \widetilde{\pi})^m$, some*[3] *$\widehat{\pi}_{\mathrm{mle}}(S) \in \mathrm{MLE}_{\Pi_{\mathcal{S}}}(S)$ has the following guarantee:*

$$\mathbb{P}_{S \sim (\mathcal{D} \times \widetilde{\pi})^m} \left(L_{\mathcal{D}, \sigma_\star}(\widehat{\pi}_{\mathrm{mle}}(S)) \geq 1 - \gamma \right) = 1 \,.$$

Consider $\mathcal{S} = \{\sigma_0, \sigma_{01}\}$, where $\sigma_0(x) = \{0\}$ and $\sigma_{01}(x) = \{0, 1\}$ for all x. If the true hypothesis is $\sigma_\star = \sigma_0$, all labels are 0. But since σ_{01} is unconstrained on unseen inputs, the estimator that outputs 0 on seen examples in the training set and 1 on unseen examples is a valid MLE. By choosing the marginal distribution so that at least a $1 - \gamma$ fraction of the probability mass lies on unseen inputs, the error is driven entirely by this missing mass, yielding loss at least $1 - \gamma$.[4]

In the previous example, MLE essentially *overfits*: it achieves zero error on observed data but fails to generalize, since the total class $\Pi_{\mathcal{S}}$ is too rich. Another nature is to restrict to a smaller class $\overline{\Pi}_{\mathcal{S}}$ of size $|\mathcal{S}|$, obtained by selecting a single representative $\overline{\pi}_r$ for each $\sigma \in \mathcal{S}$, where $\overline{\pi}_r(\cdot \mid x) =$

[3] Note that this ensures the failure of MLE after breaking ties arbitrarily. It is impossible to show that every $\widehat{\pi} \in \mathrm{MLE}_{\Pi_{\mathcal{S}}}(S)$ fails in the realizable setting; this is because the density $\widehat{\pi}$ that produces according to empirical distribution of observed examples and from the support of $\sigma_\star$ on unseen examples is always a valid MLE. However, it is unclear how to find this information theoretically from the training set S.

[4] Alternatively, even a large enough finite domain $|\mathcal{X}| = m/\gamma$ is enough if we allow the domain to depend on m, γ. More importantly, note that any $\widehat{\pi} \in \mathrm{MLE}_{\Pi_{\mathcal{S}}}(S)$ assigns the empirical distribution (i.e. memorizing) on any observed x. Thus, any $\widehat{\pi} \in \mathrm{MLE}_{\Pi_{\mathcal{S}}}(S)$ has a zero empirical error and the sample complexity is at most $|\mathcal{X}|/\varepsilon$. This implies that for any countable domain, we have consistency. But we are of course interested in going beyond memorization, and so sample complexities that don't depend, certainly not linearly, on $|\mathcal{X}|$.

$\mathrm{Unif}(\sigma(x))$. MLE can then be performed over

$$\overline{\Pi}_{\mathcal{S}} := \{\overline{\pi}_r : \sigma \in \mathcal{S}\}.$$

However, the true conditional $\widetilde{\pi}$ supported on $\sigma_\star$ need not coincide with the canonical choice $\pi_{\mathrm{unif},\sigma_\star}$, so $\overline{\Pi}_{\mathcal{S}}$ is misspecified. This mismatch suffices to make MLE fail again, even on seen examples, despite the capacity control.

Theorem 2 (MLE Failure 2)**.** *Fix $\gamma \in (0,1)$. There exists $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ with $|\mathcal{S}| = 2, |\mathcal{X}| = 1, |\mathcal{Y}| = 2\lceil 1/\gamma \rceil$, such that for some choice of $(\mathcal{D}, \sigma_\star \in \mathcal{S}, \widetilde{\pi} \in \Pi_{\sigma_\star})$, for every sample size m, for $S \sim (\mathcal{D} \times \widetilde{\pi})^m$, the unique $\widehat{\pi}_{\mathrm{mle}}(S) \in \mathrm{MLE}_{\overline{\Pi}_{\mathcal{S}}}(S)$ has the following performance guarantee:*

$$\mathbb{P}_{S\sim(\mathcal{D}\times\widetilde{\pi})^m}\left(L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_{\mathrm{mle}}(S)) \geq 1-\gamma\right) = 1\,.$$

See Section E for the proofs of these theorems.

Remark 1 (MLE achieves overlap)**.** These failures highlight a fundamental limitation of likelihood-based approaches for the Precise Completion objective. Interestingly, MLE over the restricted class $\overline{\Pi}_{\mathcal{S}} = \{\overline{\pi}_r : \sigma \in \mathcal{S}\}$ achieves a hallucinated overlap guarantee of with any sample size $m \geq \frac{1}{\varepsilon}\big(\log|\mathcal{S}| + \log(1/\delta)\big)$, For any $\widehat{\pi}_{\mathrm{mle}}(S) \in \mathrm{MLE}_{\overline{\Pi}_{\mathcal{S}}}(S)$, we have

$$\mathbb{P}_{S\sim(\mathcal{D}\times\widetilde{\pi})^m}\left(\mathbb{P}_{x\sim\mathcal{D}}[\,\mathrm{supp}\,\widehat{\pi}_{\mathrm{mle}}(S)(\cdot\mid x)\ \cap \sigma_\star(x) = \emptyset\,]\ \leq\ \varepsilon\right) \geq 1-\delta\,.$$

Thus, its predictions overlap with the ground-truth responses on all but an ε-fraction of inputs, though it may still output responses outside the support with nontrivial probability. See Section C.1 for further discussion on this.

Since overlap is not the objective of interest, these failures raise the natural question of whether finite cardinality $|\mathcal{S}| < \infty$ suffices for learnability according to Precise Completion (cf. Definition 1) loss. In the next section, we show that the answer is yes.

4.2 Learnability

Our rule is simple: output from the common intersection of consistent hypotheses if it is non-empty, and otherwise output any y within the support of some consistent σ. This suffices to ensure learnability.

Input: Sample $S = \{(x_i, y_i) : i \in [m]\}$ and a finite support hypothesis class $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$.

- Let $V(S) := \{\sigma \in \mathcal{S} : y_i \in \sigma(x_i), \forall (x_i, y_i) \in S\}$
- Return the predictor COMMON-INTERSECTION$(S) = \widehat{\pi}_{\mathrm{CI}}(S) : \mathcal{X} \to \mathcal{Y}$ as follows:

$$\widehat{\pi}_{\mathrm{CI}}(S)(x) = \begin{cases} y \in \bigcap_{\sigma\in V(S)} \sigma(x)\,, & \text{if } \bigcap_{\sigma\in V(S)} \sigma(x) \neq \emptyset\,; \\ \text{arbitrary } y & \text{otherwise.} \end{cases}$$

Theorem 3. *Any $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ with $|\mathcal{S}| = d < \infty$ is learnable according to Definition 1 using the rule $\widehat{\pi}_{\mathrm{CI}}$ with sample complexity $m_{\mathcal{S},\widehat{\pi}_{\mathrm{CI}}}(\varepsilon,\delta) = \varepsilon^{-1}\, d\,(\log d + \log(1/\delta))$.*

Note that the rule COMMON-INTERSECTION is *deterministic* and also *proper* in the following sense, when in the case when common intersection is empty, we output y that always belongs to $\sigma(x)$ for some fixed $\sigma \in V(S)$.[5]

Definition 2 (Proper Learning)**.** *We call a learning rule $\widehat{\pi} : (\mathcal{X} \times \mathcal{Y})^* \to (\Delta(\mathcal{Y}))^{\mathcal{X}}$ proper if for any $S \in (\mathcal{X} \times \mathcal{Y})^*$, the stochastic predictor $\widehat{\pi}(S)$ is supported on σ for some $\sigma \in \mathcal{S}$, i.e.* $\mathrm{supp}(\widehat{\pi}(S)(\cdot \mid x)) \subseteq \sigma(x)$ *for all $x \in \mathcal{X}$.*

The dependence on d in Theorem 3 is $\tilde{O}(d)$, in contrast to the logarithmic dependence in standard supervised learning. We show this dependence on d is tight (up to a log factor) for this rule, and even for a seemingly stronger variant that outputs by majority vote over consistent hypotheses (cf. Theorem 10 in Section D).

[5]The hypothesis class contains set-valued functions σ, whereas the prediction is a single label. Thus we define a notion of proper learning in Definition 2 that is natural for our problem.

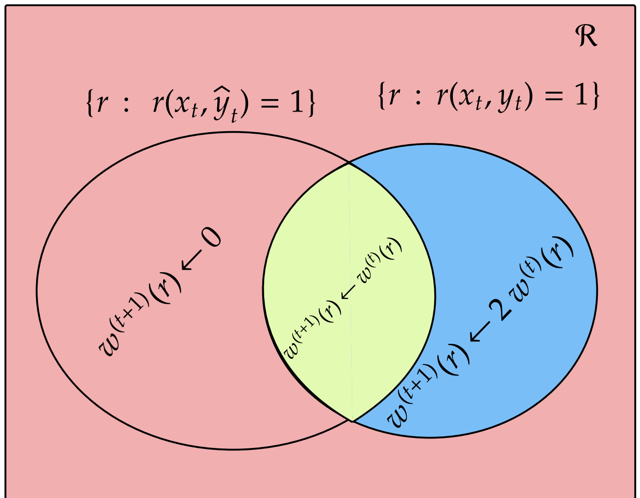

Figure 1: A visualization of the update rule of Algorithm 1 during t^{th} round. The version space shrinks to $V_{t+1} = A^t_{y_t}$. The hypotheses in the blue region are doubled their weights, and in the green region (and also white) regions are unchanged.

4.3 Exponential Improvement is Possible

While the rules in Section 4.2 guarantee learnability, we now ask for the optimal sample complexity for finite classes $\mathcal{S}$ of size d, as in (2) but now for $|\mathcal{S}| = d$. The main result of this section is that a $\log d$ dependence is in fact achievable. To achieve this, we first turn our attention to the even more challenging online version.

Online version: The adversary chooses $\sigma_\star \in \mathcal{S}$. In each round t:

- The adversary chooses $x_t \in \mathcal{X}$. The learner predicts $\widehat{y}_t \in \mathcal{Y}$.
- The adversary shows some $y_t \in \sigma_\star(x_t)$.
  (Importantly, the feedback does not inform the learner whether $\widehat{y}_t$ was a mistake or not.)

We will first design a new algorithm that utilizes the in-support cleverly and establish the mistake bound of $\log_2 |\mathcal{S}|$. The statistical estimator with logarithmic dependence on $|\mathcal{S}|$ will be designed by doing online to batch conversion. The algorithm maintains weight function $w^{(t)} : \mathcal{S} \to \mathbb{R}$ in each round; for any subset $\mathcal{S}' \subseteq \mathcal{S}$, we define $w^{(t)}(\mathcal{S}') := \sum_{\sigma \in \mathcal{S}'} w^{(t)}(\sigma)$.

Algorithm 1 Online rule based on weighted update for improved mistake bound

Input: Hypothesis class $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ with $|\mathcal{S}| < \infty$.

- Initialize $w^{(1)}(\sigma) = 1$ for all $\sigma \in \mathcal{S}$ and $V_1 = \mathcal{S}$.
- In every round, receiving x_t:
  1. Form $A^t_y = \{\sigma \in V_t : y \in \sigma(x_t)\}$ for each $y \in \mathcal{Y}$.
  2. Output $\widehat{y}_t = \arg\max_{y \in \mathcal{Y}} w^{(t)}(A^t_y)$.
  3. On receiving y_t, update the version space $V_{t+1} \leftarrow A^t_{y_t}$.
  4. Update $w^{(t+1)}(\sigma) \leftarrow 2\, w^{(t)}(\sigma)$ for all $\sigma \in A^t_{y_t} \setminus A^t_{\widehat{y}_t}$.

Theorem 4 (Online Guarantee)**.** *On any sequence $((x_t, y_t))_{t\in\mathbb{N}}$ realizable by some $\sigma_\star \in \mathcal{S}$, Algorithm 1 makes at most $\log_2 |\mathcal{S}|$ mistakes.*

Proof. Letting $W_{t+1} = w^{(t+1)}(V_{t+1})$ be the total weight in of the hypothesis leftover in the version space after completion of t rounds, we first note that the sequence $\{W_t\}_t$ is non-increasing. This is because of the property of the algorithm that, during every round t, the weight added to the system is at most the weight eliminated from the version space. Formally,

$$W_{t+1} = 2w^{(t)}(A^t_{y_t} \setminus A^t_{\widehat{y}_t}) + w^{(t)}(A^t_{y_t} \cap A^t_{\widehat{y}_t}) \le w^{(t)}(A^t_{y_t} \cup A^t_{\widehat{y}_t}) \le W_t\,,$$

where the first inequality follows from the property of the algorithm that it always chooses $\widehat{y}_t = \arg\max_{y\in\mathcal{Y}} w^{(t)}(A^t_y)$ (see also Figure 1). Now if the algorithm made M mistake on a realizable sequence for some $\sigma_\star \in \mathcal{S}$ at the end some t number of rounds, then it must be that

$$w^{(t+1)}(\sigma_\star) = 2^M \le W_{t+1} \le W_1 = |\mathcal{S}|\,,\ \text{which implies } M \le \log_2 |\mathcal{S}|\,.$$

□

Using the standard online-to-batch conversion—based on a randomized predictor that samples uniformly from all round predictor, we obtain a statistical estimator with the following performance.

Input: Sample $S = \{(x_i, y_i) : i \in [m]\}$ and a finite hypothesis class $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$.

- Run Algorithm 1 once over S, and record $(V_t, w^{(t)})$ before each round. Define

$$\widehat{\pi}_t(x) \;=\; \arg\max_{y\in\mathcal{Y}} \sum_{\sigma\in V_t} w^{(t)}(\sigma)\, \mathbf{1}\{y \in \sigma(x)\}. \tag{3}$$

- On a test $x \in \mathcal{X}$, sample $I \sim \mathrm{Unif}\{1, \dots, m\}$, and return $\widehat{\pi}_{\mathrm{o2b}}(S)(x) := \widehat{\pi}_I(x)$. I.e.

$$\widehat{\pi}_{\mathrm{o2b}}(S)(x) = \frac{1}{m}\sum_{t=1}^{m} \widehat{\pi}_t(x). \tag{4}$$

Theorem 5 (Statistical Guarantee)**.** *The estimator $\widehat{\pi}_{\mathrm{o2b}}$ in Eq.* (4) *achieves the following guarantee for any finite hypothesis class $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$, and an unknown joint distribution $(\mathcal{D} \times \widetilde{\pi})$ on $\sigma_\star \in \mathcal{S}$.*

$$\mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_{\mathrm{o2b}}(S))] \le \frac{\log_2 |\mathcal{S}|}{m}\,,$$

and, for any $\delta \in (0, 1)$, with probability at least $1 - \delta$,

$$L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_{\mathrm{o2b}}(S)) \;\le\; \frac{1 + 2\log|\mathcal{S}| + 12\log\left(\frac{\log m}{\delta}\right)}{m}\,.$$

This implies that $\mathcal{S}$ is learnable (cf. Definition 1) using the estimator $\widehat{\pi}_{\mathrm{o2b}}$ with sample complexity $m_{\mathcal{S},\widehat{\pi}_{\mathrm{o2b}}}(\varepsilon, \delta) = O\left(\varepsilon^{-1}(\log|\mathcal{S}| + \log(1/\varepsilon\delta)\right)$.

Remarkably, even under weak capacity control on $\mathcal{S}$ and with no assumptions on conditional densities, we obtain sample complexity proportional to $\log|\mathcal{S}|$, independent of $|\mathcal{X}|$, $|\mathcal{Y}|$, or $\sup_{\sigma,\,x} |\sigma(x)|$. The proof follows from concentration for martingale difference sequences Cesa-Bianchi et al. [2004], Tewari and Kakade [2008]. We obtain a sharper dependence on $1/\varepsilon$ in the realizable case via Freedman's inequality [Li et al., 2021, Theorem 3] (see Section C).

Remark 2 (Properties of the learning rule)**.** We note that our learning rule is neither (1) deterministic, (2) proper, (3) with zero empirical error. This is in contrast to COMMON-INTERSECTION rule which satisfies all the three properties. It remains an interesting question whether we can achieve any of the two properties simultaneously while having $\log|\mathcal{S}|$ dependence.

5 k-pass Error Minimization

In modern practice, pass-k accuracy is often used as a benchmark. This relaxes the original goal by allowing a stochastic predictor $\widehat{\mu} : \mathcal{X} \to \Delta(\mathcal{Y}^k)$, with loss

$$L_{\mathcal{D},\sigma_\star}(\widehat{\mu}) = \mathbb{E}_{x\sim\mathcal{D}}, \mathbb{E}_{\boldsymbol{y}=(y^{(1)},\dots,y^{(k)})\sim\widehat{\mu}(\cdot|x)} \left[\mathbb{1}\{y^{(i)} \notin \sigma_\star(x); \forall i \in [k]\}\right]. \tag{5}$$

Note that the above allows for any joint distribution over the set of k responses that the estimator may design. This allows for adaptive sampling, and does not restrict the learner to output from a product distribution (repeated independent sampling from a stochastic predictor $\widehat{\pi}$). Our goal is to understand how the parameter k affects sample complexity. With this relaxation, the complexity improves only by a $\log k$ factor in the cardinality parameter. The upper bound follows by extending Algorithm 1 and applying online-to-batch conversion, while the lower bound comes from a worst-case construction (see Section F).

Theorem 6 (Informal: k-pass loss)**.** *The minimax mistake bound as well as sample complexity bound in online and statistical settings respectively are $\Theta(\log_k d)$ for the family of finite classes of size d.*

Remark 3 (General bounded reward classes)**.** Both our estimators ($\widehat{\pi}_{\mathrm{o2b}}$ as well as its k-pass variant in Section F) can be generalized to a more general setting of imitation learning for bounded (possibly non-binary reward) function classes, under the promise that the expert demonstrator $\widetilde{\pi}$ shows only maximum reward examples. I.e. consider the reward class $\mathcal{R}$ containing functions $r : (\mathcal{X} \times \mathcal{Y}) \to [0, 1]$, and the promise here is that $\mathrm{supp}(\widetilde{\pi}(\cdot \mid x)) \subseteq \arg\max_{y \in \mathcal{Y}} r_\star(x, y)$ for some $r_\star \in \mathcal{R}$. We can design estimators that have $V(\widehat{\pi}) \geq V(\widetilde{\pi}) - \varepsilon$ with high probability. We leave it for future investigation to obtain multiplicative approximation $V(\widehat{\pi}) \geq (1-\varepsilon)V(\widetilde{\pi})$ or whether the assumption can be removed.

6 Discussion

Distinctions from prior work. Our first key contribution is to carefully separate between the capacity-control viewpoint on the density class Π and on support class $\mathcal{S}$, for the imitation learning problem. While still coarse, this distinction allows us to highlight a central limitation of MLE: its failure to generalize in producing valid completions on unseen prompts.

A second implication of our analysis concerns another important issue of so-called *hallucinations*. Our explanation is novel–grounded specifically in the failure of MLE–departing from other recent theoretical attempts [Kalai and Vempala, 2024, Kalavasis et al., 2025]. In particular, we show that hallucinations arise naturally in the *prompted completion* setting when learning is carried out by density estimation via MLE, which does not align with the goal of Precise Completion. To some extent, this is most closely related to a very recent work [Kalai et al., 2025, Section 3.2], where the prompted setting is also considered. However, their Theorem 3.1 establishes rates in terms of "calibration", which captures the model's coverage over good responses. This provides further evidence of a fundamental tradeoff between precision (validity) and recall (coverage/calibration). Yet their work treats calibration–encouraged by the density-estimation–as a desirable notion of generalization. By contrast, our work goes further: we argue that when both precision and recall cannot be simultaneously achieved, one should at least prioritize precision, which is the important objective on its own right in the prompted completion and matches the practical use case of LLMs. We then design new estimators that have strong performance guarantee with only Precise Completion.

Open questions. None of our learning rules are readily implementable, and show only statistical possibility. An intriguing open direction is to design practical surrogates for MLE that better aligns with the inductive biases required for the Precise Completion. While our results with finite classes $|\mathcal{S}| < \infty$ already demonstrate an interesting algorithmic landscape, we believe this problem may admit an even richer picture for infinite classes in terms of combinatorial dimensions. A special case of our problem is the multiclass setting where all support functions are of size one, which has a simple picture for finite classes, but has a much more intriguing landscape for infinite classes with combinatorial dimensions [Shalev-Shwartz and Ben-David, 2014, Daniely and Shalev-Shwartz, 2014, Brukhim et al., 2022]. The k-pass variant, in turn, corresponds to the list learnability problem Brukhim et al. [2022], Charikar and Pabbaraju [2023].

References

A. Agarwal, C. Dann, and T. V. Marinov. Design considerations in offline preference-based rl. *arXiv preprint arXiv:2502.06861*, 2025.

Anthropic. Claude 3 model family. `https://www.anthropic.com/news/claude-3`, 2024.

F. L. Bronnec, A. Verine, B. Negrevergne, Y. Chevaleyre, and A. Allauzen. Exploring precision and recall to assess the quality and diversity of llms. *arXiv preprint arXiv:2402.10693*, 2024.

N. Brukhim, D. Carmon, I. Dinur, S. Moran, and A. Yehudayoff. A characterization of multiclass learnability. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 943–955. IEEE, 2022.

N. Cesa-Bianchi, A. Conconi, and C. Gentile. On the generalization ability of on-line learning algorithms. *IEEE Transactions on Information Theory*, 50(9):2050–2057, 2004.

M. Charikar and C. Pabbaraju. A characterization of list learnability. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1713–1726, 2023.

M. Charikar and C. Pabbaraju. Exploring facets of language generation in the limit. *arXiv preprint arXiv:2411.15364*, 2024.

L. Cohen, Y. Mansour, S. Moran, and H. Shao. Probably approximately precision and recall learning. *arXiv preprint arXiv:2411.13029*, 2024.

H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, 1946.

A. Daniely and S. Shalev-Shwartz. Optimal learners for multiclass problems. In *Conference on Learning Theory*, pages 287–316. PMLR, 2014.

G. DeepMind. Introducing gemini: Our largest and most capable ai model. `https://blog.google/technology/ai/google-gemini-ai/`, 2023.

DeepSeekAI. Deepseek models overview. `https://www.deepseek.com`, 2025.

D. J. Foster, S. M. Kakade, J. Qian, and A. Rakhlin. The statistical complexity of interactive decision making. *arXiv preprint arXiv:2112.13487*, 2021.

D. J. Foster, A. Block, and D. Misra. Is behavior cloning all you need? understanding horizon in imitation learning. *Advances in Neural Information Processing Systems*, 37:120602–120666, 2024.

A. Huang, W. Zhan, T. Xie, J. D. Lee, W. Sun, A. Krishnamurthy, and D. J. Foster. Correcting the mythos of kl-regularization: Direct alignment without overoptimization via chi-squared preference optimization. *arXiv preprint arXiv:2407.13399*, 2024.

A. Huang, A. Block, Q. Liu, N. Jiang, A. Krishnamurthy, and D. J. Foster. Is best-of-n the best of them? coverage, scaling, and optimality in inference-time alignment. *arXiv preprint arXiv:2503.21878*, 2025.

Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. Bang, A. Madotto, and P. Fung. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38, 2023.

A. T. Kalai and S. S. Vempala. Calibrated language models must hallucinate. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 160–171, 2024.

A. T. Kalai, O. Nachum, S. S. Vempala, and E. Zhang. Why language models hallucinate. Technical report, OpenAI, Sept. 2025.

A. Kalavasis, A. Mehrotra, and G. Velegkas. On the limits of language generation: Trade-offs between hallucination and mode-collapse. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1732–1743, 2025.

E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, 1998.

G. Li, L. Shi, Y. Chen, Y. Gu, and Y. Chi. Breaking the sample complexity barrier to regret-optimal model-free reinforcement learning. *Advances in Neural Information Processing Systems*, 34: 17762–17776, 2021.

C. D. Manning. *Introduction to information retrieval*. Syngress Publishing,, 2008.

J. Neyman and E. S. Pearson. On the use and interpretation of certain test criteria for purposes of statistical inference: Part i. *Biometrika*, 20A(1/2):175–240, 1928.

OpenAI. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

R. Rafailov, A. Sharma, E. Mitchell, C. D. Manning, S. Ermon, and C. Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023.

S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.

U. Syed and R. E. Schapire. A game-theoretic approach to apprenticeship learning. *Advances in neural information processing systems*, 20, 2007.

A. Tewari and S. Kakade. Online-to-batch conversions. Lecture notes, CMSC 35900: Learning Theory, Toyota Technological Institute at Chicago, 2008. `https://home.ttic.edu/~tewari/lectures/lecture13.pdf`.

H. Touvron et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

A. Wald. Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics*, 20(4):595–601, 1949.

S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9(1):60–62, 1938.

T. Xie, D. J. Foster, A. Krishnamurthy, C. Rosset, A. Awadallah, and A. Rakhlin. Exploratory preference optimization: Harnessing implicit q*-approximation for sample-efficient rlhf. *arXiv preprint arXiv:2405.21046*, 2024.

J. Yun, J. Kim, J. Park, J. Kim, J. J. Ryu, J. Cho, and K.-S. Jun. Alignment as distribution learning: Your preference model is explicitly a language model. *arXiv preprint arXiv:2506.01523*, 2025.

W. Zhan, M. Uehara, N. Kallus, J. D. Lee, and W. Sun. Provable offline preference-based reinforcement learning. *arXiv preprint arXiv:2305.14816*, 2023.

Y. Zhang, L. Wang, M. Fang, Y. Du, C. Huang, J. Wang, Q. Lin, M. Pechenizkiy, D. Zhang, S. Rajmohan, et al. Distill not only data but also rewards: Can smaller language models surpass larger ones?, 2025. *URL https://arxiv. org/abs/2502.19557*, 2025.

A Technical Preliminary Lemmas

We start by a technical lemma about the one-sided change of measure bound on an expectation of a bounded function in terms of the Hellinger distance (e.g. [Foster et al., 2021, Lemma A.11]). We will use the exact variant from [Foster et al., 2024, Lemma 3.11].

Lemma 1 (Change-of-measure bound via Hellinger distance,Foster et al. [2024])**.** *Let $(\mathcal{Z}, \mathcal{F})$ be a measurable space and let $\mathbb{P}, \mathbb{Q}$ be probability measures on it. For every measurable function $h : \mathcal{Z} \to \mathbb{R}$:*

$$\big|\mathbb{E}_{\mathbb{P}}[h] - \mathbb{E}_{\mathbb{Q}}[h]\big| \;\le\; \sqrt{\frac{\mathbb{E}_{\mathbb{P}}\big[h^2\big] + \mathbb{E}_{\mathbb{Q}}\big[h^2\big]}{2}}\; D_{\mathsf{H}}(\mathbb{P}, \mathbb{Q}). \tag{6}$$

In particular for $h : \mathcal{Z} \to [0, R]$,

$$\mathbb{E}_{\mathbb{P}}[h] \;\le\; 2\,\mathbb{E}_{\mathbb{Q}}[h] \;+\; R\,D^2_{\mathsf{H}}(\mathbb{P}, \mathbb{Q})\;. \tag{7}$$

We now specify Freedman's inequality that provides us with a non-asymptotic bound on the sum of martingale difference sequence.

Lemma 2 (Freedman's inequality, Theorem 3 from Li et al. [2021])**.** *Consider a filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$, and write $\mathbb{E}_i[\cdot] := \mathbb{E}[\cdot \mid \mathcal{F}_i]$. Let*

$$Y_m = \sum_{i=1}^{m} X_i$$

where (X_i) is a real-valued scalar sequence satisfying:

$$|X_i| \leq R, \qquad \mathbb{E}_{i-1}[X_i] = 0 \quad \textit{for all } i \geq 1,$$

for some constant $R < \infty$. Define the predictable variance process

$$W_m := \sum_{i=1}^{m} \mathbb{E}_{i-1}[X_i^2],$$

and assume deterministically that $W_m \leq \sigma^2$ for some constant $\sigma^2 < \infty$. Then for any integer $n \geq 1$, with probability at least $1 - \delta$,

$$\big|Y_m\big| \;\leq\; \sqrt{8\,\max\big\{W_m,\, \frac{\sigma^2}{2n}\big\}\,\log\left(\frac{2n}{\delta}\right)} \;+\; \frac{4}{3} R \log\left(\frac{2n}{\delta}\right).$$

Maximum Likelihood Estimation for Density Estimation. We now state guarantee for the maximum likelihood estimator (MLE) for density estimation, exactly similar to [Foster et al., 2024, Section B.4]. Given a class of candidate densities $\mathcal{G}$ and i.i.d. samples $z_1, \ldots, z_m \sim g_*$ (possibly not in $\mathcal{G}$), we define the empirical negative log-likelihood (log-loss) of $g \in \mathcal{G}$ as

$$L_{\log}(g) \;=\; -\sum_{i=1}^{m} \log g(z_i).$$

The maximum likelihood estimator is then

$$\hat{g}_{\mathrm{mle}} \;\in\; \operatorname*{arg\,min}_{g \in \mathcal{G}} \; L_{\log}(g). \tag{8}$$

Definition 3 (Log-loss covering number)**.** *For a class $\mathcal{G} \subseteq \Delta(\mathcal{Z})$, we say that a subset $\mathcal{G}' \subseteq \Delta(\mathcal{Z})$ is an* ε-cover with respect to the log-loss *if for all $g \in \mathcal{G}$ there exists $g' \in \mathcal{G}'$ such that $\sup_{z \in \mathcal{Z}} \; \log(g(z)/g'(z)) \;\leq\; \varepsilon$. We denote the size of the smallest such cover by $\mathcal{N}_{\log}(\mathcal{G}, \varepsilon)$.*

We have the following property of MLE's convergence in the squared Hellinger distance with high probability.

Proposition 2. *With probability $1 - \delta$ over m i.i.d. samples from any $g_* \in \mathcal{G}$,*

$$D_{\mathsf{H}}^2(g_*, \hat{g}_{\mathrm{mle}}) \;\leq\; \inf_{\varepsilon > 0} \left\{ \frac{6 \log\big(2\,\mathcal{N}_{\log}(\mathcal{G}, \varepsilon)/\delta\big)}{m} + 4\varepsilon \right\} \;+\; 2 \inf_{g \in \mathcal{G}} \log\big(1 + D_{\chi^2}(g_* \,\|\, g)\big) \;+\; 2\,\varepsilon_{\mathrm{opt}}.$$

In particular, if $\mathcal{G}$ is finite and $\varepsilon_{opt} = 0$, the maximum likelihood estimator satisfies

$$D_{\mathsf{H}}^2(g_*, \hat{g}_{\mathrm{mle}}) \;\leq\; \frac{6 \log\big(2\,|\mathcal{G}|/\delta\big)}{m} \;+\; 2 \inf_{g \in \mathcal{G}} \log\big(1 + D_{\chi^2}(g_* \,\|\, g)\big)\;.$$

Note that the term $\inf_{g \in \mathcal{G}} \log\big(1 + D_{\chi^2}(g_* \,\|\, g)\big)$ corresponds to the misspecification error, and is zero if $g_* \in \mathcal{G}$.

We note that the proof of [Foster et al., 2024, Proposition B.1] contains a couple of minor typographical errors. Namely, in Eq.(20) therein, the authors aim to compare $\widetilde{g}$ and $\widehat{g}$, but ended up comparing $\widetilde{g}$ and g_*. A similar mistake is repeated a a couple of more times without affecting the correctness of the argument.

B Proof of Proposition 1

Intuition with the special cases of Π: We provide a more transparent and direct proof for the special case when for every $\pi \in \Pi, x \in \mathcal{X}$, the conditional density $\pi(\cdot \mid x)$ puts a uniform distribution over exactly s members of $\mathcal{Y}$ for some large but finite integer s. First, observe that in this special case we have a dichotomy; any hypothesis that does not contradict the data has the same likelihood as any other, so any $\pi \in \Pi$ that does not contradict with the data is MLE. For the unknown $\mathcal{D} \times \widetilde{\pi}$, we now consider any π such that

$$L_{\mathcal{D},\sigma_\star}(\pi) = \mathbb{P}_{x\sim\mathcal{D},\widehat{y}\sim\pi(\cdot|x)}\left(\widehat{y} \notin \sigma_{\widetilde{\pi}}(x)\right) > \varepsilon\,.$$

Then, due to the symmetry of the loss in the special case where each $\pi \in \Pi$ puts a uniform distribution on exactly s items, we have

$$\mathbb{P}_{x\sim\mathcal{D},\widehat{y}\sim\pi(\cdot|x)}\Big(\widehat{y} \notin \sigma_{\widetilde{\pi}}(x)\Big) \;=\; \mathbb{P}_{x\sim\mathcal{D},y\sim\widetilde{\pi}(\cdot|x)}\Big(y \notin \sigma_{\pi}(x)\Big) > \varepsilon,$$

where the key fact used is the ability to change the order of randomness between $\widehat{y} \sim \pi(\cdot \mid x)$ and $y \sim \widetilde{\pi}(\cdot \mid x)$.

This shows that when we sample $(x, y) \sim \mathcal{D} \times \widetilde{\pi}$, the probability that (x, y) does not fall in the support $\sigma_\pi(x)$ exceeds ε. Hence, for any fixed $\pi \in \Pi$, after m i.i.d. draws

$$\mathbb{P}_S(\pi \text{ survives}) \le (1-\varepsilon)^m \le e^{-\varepsilon m}.$$

Therefore, by a standard union bound,

$$\mathbb{P}_S\big(\exists \text{ bad } \pi \in \Pi \text{ that survives}\big) \;\le\; |\Pi|\, e^{-\varepsilon m}.$$

The proposition follows by choosing $m \ge m_{\Pi,\widehat{\pi}_{\mathrm{mle}}}(\varepsilon,\delta) = O\Big(\frac{\log|\Pi| + \log(1/\delta)}{\varepsilon}\Big)$.

Proof for any general Π: Consider any unknown but fixed marginal distribution $\mathcal{D} \in \Delta(\mathcal{X})$. For any conditional law $\pi : \mathcal{X} \to \Delta(\mathcal{Y})$, let $\mathbb{P}_{(\mathcal{D},\pi)}$ denote the joint law over $(\mathcal{X} \times \mathcal{Y})$ given by the marginal distribution $\mathcal{D}$ and the conditional law $\pi(\cdot \mid x)$. First observe that for any $S \in (\mathcal{X} \times \mathcal{Y})^*$, the joint law $\mathbb{P}_{(\mathcal{D},\widehat{\pi}_{\mathrm{mle}}(S))}$ is the MLE of among all joint distribution $\{\mathbb{P}_{(\mathcal{D},\pi)} : \pi \in \Pi\}$. Using Proposition 2, for $S \sim (\mathcal{D} \times \widetilde{\pi})^m$

$$\mathbb{P}_S\left(D_{\mathsf{H}}^2\left(\mathbb{P}_{(\mathcal{D},\widetilde{\pi})}, \mathbb{P}_{\mathcal{D},\widehat{\pi}_{\mathrm{mle}}(S)}\right) \le \frac{6\log(2|\Pi|/\delta)}{m}\right) \ge 1-\delta\,. \tag{9}$$

Now let $\sigma_\star : \mathcal{X} \to 2^{\mathcal{Y}}$ be the associated support set valued function of valid responses for $\sigma_\star(x) \supseteq \mathrm{supp}(\widetilde{\pi}(\cdot \mid x))$. Let us define a function $\mathrm{err} : (\mathcal{X} \times \mathcal{Y}) \to \{0, 1\}$ as

$$\mathrm{err}(x,y) = \begin{cases} 1 & \text{if } y \notin \sigma_\star(x), \\ 0 & \text{otherwise.} \end{cases}$$

Then using Lemma 1, we have for any conditional law $\pi : \mathcal{X} \to \Delta(\mathcal{Y})$

$$L_{\mathcal{D},\sigma_\star}(\pi) = \mathbb{E}_{\mathbb{P}_{(\mathcal{D},\pi)}}[\mathrm{err}] \le D_{\mathsf{H}}^2(\mathbb{P}_{(\mathcal{D},\widetilde{\pi})}, \mathbb{P}_{(\mathcal{D},\pi)})\,,$$

where we used the fact that $L_{\mathcal{D},\sigma_\star}(\widetilde{\pi}) = \mathbb{E}_{\mathbb{P}_{\mathcal{D},\widetilde{\pi}}}[\mathrm{err}] = 0$ and that err is a bounded function in $[0, 1]$. Combining this with (9), we obtain that with probability at least $1-\delta$ over $S \sim (\mathcal{D} \times \widetilde{\pi})^m$,

$$L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_{\mathrm{mle}}(S)) \le \frac{6\log(2|\Pi|/\delta)}{m}\,.$$

C Proof from Section 4.3

We have an online learning algorithm (Algorithm 1) that makes at most $\log_2|\mathcal{S}|$ mistakes (Theorem 4). We now show, how using online-to-batch conversion via the estimator $\widehat{\pi}_{\mathrm{o2b}}$ (Eq. 4), we can enjoy a similar sample complexity.

Proof of Theorem 5. Let $\ell_t = \mathbb{1}\{\widehat{\pi}_t(x_t) \notin \sigma_\star(x_t)\}$. Because $\widehat{\pi}_t$ is a deterministic function of $S_{<t} = \{(x_i, y_i) : i < t\}$, we have

$$\mathbb{E}[\ell_t \mid S_{<t}] = L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_t)\,.$$

Hence

$$\mathbb{E}_S\left[L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_{\mathrm{o2b}}(S))\right] = \mathbb{E}_S\left[\frac{1}{m}\sum_{t=1}^{m} L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_t)\right] = \mathbb{E}_S\left[\frac{1}{m}\sum_{t=1}^{m}\ell_t\right] \le \frac{\log_2|\mathcal{S}|}{m}\,,$$

where in the last inequality we used Theorem 4, which guarantees $\sum_{t=1}^{m}\ell_t \le \log_2|\mathcal{S}|$.

For the high-probability statement, define the martingale differences

$$Z_t := L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_t) - \ell_t, \qquad \text{where } |Z_t| \le 1 \text{ almost surely.}$$

Then $\mathbb{E}[Z_t \mid S_{<t}] = 0$, and

$$\mathbb{E}[Z_t^2 \mid S_{<t}] = \mathbb{E}[(L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_t) - \ell_t)^2 \mid S_{<t}] = \mathrm{Var}(\ell_t \mid S_{<t}) = L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_t)(1 - L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_t)) \;\le\; L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_t)\,.$$

And taking $W_m = \sum_{t=1}^{m} L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_t)$ and $\sigma^2 = m$ suffices, thus, using Lemma 2 with $n = \log m$ inequality gives us with probability $1-\delta$

$$\begin{aligned}\sum_{t=1}^{m} Z_t &\le \sqrt{8\left(1 + \sum_{t=1}^{m} L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_t)\right)\log\left(\frac{\log m}{\delta}\right)} + \frac{4}{3}\log\left(\frac{\log m}{\delta}\right) \\ &\le \frac{1}{2}\left(1 + \sum_{t=1}^{m} L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_t)\right) + 4\log\left(\frac{\log m}{\delta}\right) + \frac{4}{3}\log\left(\frac{\log m}{\delta}\right) \qquad (\mathrm{GM} \le \mathrm{AM})\end{aligned}$$

Substituting Z_t and rearranging terms,

$$\sum_{t=1}^{m} L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_t) \le 1 + 2\sum_{t=1}^{m}\ell_t + 12\log\left(\frac{\log m}{\delta}\right)$$

Finally noting that $L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_{\mathrm{o2b}}(S)) = \frac{1}{m}\sum_{t=1}^{m} L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_t)$ and that $\sum_{t=1}^{m}\ell_t \le \log_2|\mathcal{S}|$ (by Theorem 4), we obtain that with probability $1-\delta$,

$$L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_{\mathrm{o2b}}(S)) \le \frac{1 + 2\log_2|\mathcal{S}| + 12\log\left(\frac{\log m}{\delta}\right)}{m}$$

□

C.1 Overlap of MLE

Interestingly, MLE attains a hallucinated overlap at the statistical limit of Theorem 5, though its failure to directly optimize the Precise Completion objective of interest (cf Section 4.1). Consider the class $\overline{\Pi}_{\mathcal{S}} = \bigcup_{\sigma\in\mathcal{S}}\{\overline{\pi}_r\}$.

Theorem 7. *For any class $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ and an unknown joint realizable distribution $\mathcal{D} \times \widetilde{\pi}$, where $\widetilde{\pi}$ is supported on some $\sigma_\star \in \mathcal{S}$, for any estimator $\widehat{\pi}_{\mathrm{mle}}(S) \in \mathrm{MLE}_{\overline{\Pi}_{\mathcal{S}}}(S)$, we have the following guarantee: for any sample size $m \ge \varepsilon^{-1}(\log|\mathcal{S}| + \log(1/\delta))$, we have*

$$\mathbb{P}_{S\sim(\mathcal{D}\times\widetilde{\pi})^m}\left(\mathbb{P}_{x\sim\mathcal{D}}\left(\mathrm{supp}(\widehat{\pi}_{\mathrm{mle}}(S)(\cdot \mid x)) \cap \sigma_\star(x) = \emptyset\right) \le \varepsilon\right) \ge 1-\delta\,.$$

We now show the proof of Theorem 7 that MLE over the restricted class achieves overlap with the ground-truth. The proof is simple.

Proof of Theorem 7. First note that $\pi_{\mathrm{unif},\sigma_\star}$ has non-zero likelihood. Therefore, any policy in the set $\mathrm{MLE}_{\overline{\Pi}_{\mathcal{S}}}(S)$ must have non-zero likelihood. Thus, for any σ for which $\overline{\pi}_r \in \mathrm{MLE}_{\overline{\Pi}_{\mathcal{S}}}(S)$, we must have that $\sigma \in V(S) := \{\sigma \in \mathcal{S} : y_i \in \sigma(x_i)\,\forall(x_i, y_i) \in S\}$. Therefore, in order to establish

$$\mathbb{P}_{S\sim(\mathcal{D}\times\widetilde{\pi})^m}\left(\mathbb{P}_{x\sim\mathcal{D}}\left(\mathrm{supp}(\widehat{\pi}_{\mathrm{mle}}(S)(x)) \cap \sigma_\star(x) = \emptyset\right) \le \varepsilon\right) \ge 1-\delta\,,$$

it suffices to establish

$$\mathbb{P}_S\left(\forall \sigma \in V(S) : \mathbb{P}_{x \sim \mathcal{D}}\left(\sigma(x) \cap \sigma_\star(x) = \emptyset\right) \leq \varepsilon\right) \geq 1 - \delta\,. \tag{10}$$

Consider any bad $\sigma \in \mathcal{S}$ such that $\mathbb{P}_{x\sim\mathcal{D}}(\sigma(x) \cap \sigma_\star(x) = \emptyset) > \varepsilon$. With each draw $(x_i, y_i) \sim (\mathcal{D} \times \widetilde{\pi})$, we have that σ gets knocked-out of version space with probability at least ε, i.e. $\mathbb{P}_{(x_i,y_i)\sim(\mathcal{D}\times\widetilde{\pi})}(y_i \notin \sigma(x_i)) > \varepsilon$. Therefore, for any fixed σ, after sample $S \sim (\mathcal{D} \times \widetilde{\pi})^m$

$$\mathbb{P}_S(\sigma \in V(S)) \leq (1 - \varepsilon)^m \leq e^{-\varepsilon m}.$$

Therefore, by a standard union bound,

$$\mathbb{P}_S\big(\exists \text{ bad } \sigma \in V(S) \text{ that survives}\big) \;\leq\; |\mathcal{S}|\, e^{-\varepsilon m} \leq |\mathcal{S}|\, 2^{-\varepsilon m}.$$

The theorem follows by noting that the $|\mathcal{S}|\, 2^{-\varepsilon m} \leq \delta$ for any $m \geq \frac{\log|\mathcal{S}| + \log(1/\delta)}{\varepsilon}$. □

Remark 4 (Comparison with multi-class classification)**.** Note that for multiclass classification when $\mathcal{S} \subseteq \mathcal{Y}^{\mathcal{X}}$ (i.e. all $|\sigma(x)| = 1$, the guarantee captured in (10) is enough to ensure learnability by just outputting a single predictor from $\widehat{\sigma} \in V(S)$ (i.e. consistent / ERM). This happens because the overlap implies that that labels are the same and so no error. However, for our problem despite this overlap, it is unclear how to output a single label so that it belongs to the support of $\sigma_\star$. What would be sufficient for our problem is the following guarantee, where the quantifier $\forall \sigma \in V(S)$ is taken inside the randomness of test point sampling $x \sim \mathcal{D}$:

$$\mathbb{P}_S\left(\mathbb{P}_{x\sim\mathcal{D}}\left(\forall \sigma \in V(S) : \sigma(x) \cap \sigma_\star(x) = \emptyset\right) \leq \varepsilon\right) \geq 1 - \delta\,.$$

However, we know that this provably requires the sample size where there is $\Omega(|\mathcal{S}|)$ dependence on cardinality—see the lower bound for COMMON-INTERSECTION estimator (Theorem 10).

Theorem 7 says that the MLE over the restricted class $\overline{\Pi}_{\mathcal{S}}$ at least achieves an overlap on most of the unseen examples with high probability, at the optimal sample complexity. However, it does not guarantee that the response generated from $\widehat{\pi}_{\text{mle}}$ will be in the support. In particular, it may produce responses outside, with a decent probability depending on the amount of overlap the $\text{supp}(\widehat{\pi}_{\text{mle}}(S)(\cdot \mid x))$ has with $\sigma_\star$. It may be possible to turn this into a predictor that directly starts to produce good responses, depending on the overlap among hypotheses and other types of feedback available in post-training (e.g., whether a generated response is good or not). This overlap can be captured by a parameter that reflects the need for repeated sampling and the number of feedback that must be queried, which in turn allows for a more quantitative understanding of how many feedbacks are required to guarantee performance in terms of this parameter. For example, this parameter would be maximum in the case of multi-class classification (Remark 4) and no additional feedback is required. However, we leave it open to formulate an interesting setup that enables a study of both types of feedbacks together for our problem, and we do not attempt to investigate this any further.

D Proofs for Section 4.2: Common Intersection and Majority

Online Mistake and Statistical Sample Complexity bounds for COMMON-INTERSECTION. We start by analyzing the COMMON-INTERSECTION rule in the more difficult online setting which helps for the intuition for the statistical setting.

Theorem 8 (Online Guarantee for COMMON-INTERSECTION)**.** *On any sequence $((x_t, y_t))_{t\in\mathbb{N}}$ realizable by some $\sigma_\star \in \mathcal{S}$, the rule* COMMON-INTERSECTION *(applied to the sequence seen so far) makes at most $|\mathcal{S}| - 1$ mistakes.*

Proof of Theorem 8. Consider any round t in which there was a mistake made by the rule. It must be that the set of consistent hypothesis V_t in that round, it must be that there was no *common intersection* in that round, i.e. $\bigcap_{\sigma\in V_t} \sigma(x_t) = \emptyset$. That means even though we would not know whether we made a mistake in that round, observing y_t will eliminate at least one hypothesis from the version space (i.e. $|V_{t+1}| \leq |V_t| - 1$). Therefore, the rule cannot make $|V_1| - 1 = |\mathcal{S}| - 1$ mistakes on any realizable sequence. □

We now analyze the performance of this rule in the statistical version.

Proof of Theorem 3. Partition the m examples into $K := |\mathcal{S}|$ consecutive blocks $B_1, \ldots, B_K$, each of length $n \geq \frac{1}{\varepsilon}(\log|\mathcal{S}| + \log(1/\delta))$. Let V_t denote the version space just before block B_t begins. I.e. define the restricted dataset $S_t = B_1 \cup \cdots \cup B_{t-1}$ and

$$V_t = \{\sigma \in \mathcal{S} : y_i \in \sigma(x_i)\ \forall (x_i, y_i) \in S_t\}$$

with $V_1 = \mathcal{S}$. Define the region of x, where we do not have a common intersection among V_t.

$$A_t := \Big\{x \in \mathcal{X} : \bigcap_{\sigma \in V_t} \sigma(x) = \emptyset\Big\}.$$

Note that $A_{t+1} \subseteq A_t$ for all $t \in [K]$ because $V_{t+1} \subseteq V_t$. Moreover, we never make an error when outputting from common-intersection region. Now, using these facts, we have

$$\begin{aligned}
\mathbb{P}_S\left(L_{\mathcal{D},\sigma_*}(\widehat{\pi}_{\mathrm{CI}}(S)) > \varepsilon\right) &\leq \mathbb{P}_S\left(\mathbb{P}_{x\sim\mathcal{D}}(A_{K+1}) > \varepsilon\right) \\
&= \mathbb{P}_S\left(V_{K+1} \neq \emptyset \cap \mathbb{P}_{x\sim\mathcal{D}}(A_{K+1}) > \varepsilon\right) \quad (V_{K+1} \neq \emptyset \text{ always happens}) \\
&\leq \mathbb{P}_S\left(\exists t \in [K],\, V_{t+1} = V_t \ \cap\ \mathbb{P}_{x\sim\mathcal{D}}(A_{K+1}) > \varepsilon\right) \\
&\leq \mathbb{P}_S\left(\exists\, t \in [K], V_{t+1} = V_t \ \cap\ \mathbb{P}_{x\sim\mathcal{D}}(A_t) > \varepsilon\right) \\
&\leq \sum_{t=1}^{K} \mathbb{P}_{B_t}\left(\exists\, t \in [K], V_{t+1} = V_t \ \mid \mathbb{P}_{x\sim\mathcal{D}}(A_t) > \varepsilon\right) \\
&\leq \sum_{t=1}^{K} (1-\varepsilon)^{|B_t|} = K(1-\varepsilon)^n \\
&\leq |\mathcal{S}|\, 2^{-\varepsilon n} \leq |\mathcal{S}| \cdot \frac{\delta}{|\mathcal{S}|} = \delta\,.
\end{aligned}$$

The above calculation formalizes the following argument. There are two cases to consider:
Case 1. If $\mathbb{P}_{x\sim\mathcal{D}}(A_t) > \varepsilon$, then the probability that no $x \in A_t$ appears in block B_t is at most $(1-\varepsilon)^n \leq e^{-\varepsilon n} \leq 2^{-\varepsilon n} \leq \delta/|\mathcal{S}|$. Otherwise, some $x \in A_t$ appears; since $\bigcap_{\sigma\in V_t} \sigma(x) = \emptyset$, the observed label $y \in \sigma_\star(x)$ excludes at least one $\sigma \in V_t$, so $|V_{t+1}| \leq |V_t| - 1$.

Case 2. If $\mathbb{P}_{x\sim\mathcal{D}}(A_t) \leq \varepsilon$, then on A_t^c the intersection is nonempty, and because $\sigma_\star \in V_t$ it follows that the COMMON-INTERSECTION prediction is always correct there. Moreover, since $V_{t+1} \subseteq V_t$, the intersections can only grow and hence $A_{t+1} \subseteq A_t$. Therefore, once Case 2 holds, the final error remains below ε.

$$L_{\mathcal{D},\sigma_\star}(\widehat{\pi}_{\mathrm{CI}}(S)) \leq \mathbb{P}_{\mathcal{D}}(A_t) < \varepsilon\,.$$

Putting these together, with probability at least $1 - K \cdot (\delta/|\mathcal{S}|) \geq 1 - \delta$, each block in Case 1 eliminates at least one hypothesis, and there are at most $K = |\mathcal{S}|$ such eliminations are even possible. Hence either Case 2 occurs in some block (giving final error $< \varepsilon$), or Case 1 occurs in all K blocks, which is not possible in the realizable setting, arriving at a contradiction. ☐

D.1 Lower Bounds

For the lower bound, we will consider an even seemingly stronger rule where output is a response that belongs to the support of most number of consistent hypothesis.

> **Input:** Sample $S = \{(x_i, y_i) : i \in [m]\}$ and a finite support hypothesis class $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$.
> - Let $V(S) := \{\sigma \in \mathcal{S} : y_i \in \sigma(x_i), \forall (x_i, y_i) \in S\}$
> - Return the predictor MAJORITY$(S) = \widehat{\pi}_{\mathrm{Maj}}(S) : \mathcal{X} \to \mathcal{Y}$ defined as follows:
>
> $$\widehat{\pi}_{\mathrm{Maj}}(S)(x) = \arg\max_{y\in\mathcal{Y}} |\{\sigma \in V(S) : y \in \sigma(x)\}|$$

The lower bounds will hold for the following instance of the of the class.

Description of the class. Fix $d \in \mathbb{N}$. Let

$$\mathcal{Y} = \{0, 1\}, \qquad q := \left\lfloor \tfrac{d-1}{2} \right\rfloor, \qquad \mathcal{X} := \{1, 2, \ldots, q\}.$$

We define a hypothesis class $\mathcal{S} = \{\sigma_1, \sigma_2, \ldots, \sigma_d\} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ as follows.

- **Distinguished hypothesis.** Set $\sigma_1(x) = \{1\}$ for every $x \in \mathcal{X}$. This will serve as the ground-truth hypothesis.
- **Adversarial hypotheses.** For each $i \geq 2$, require that $0 \in \sigma_i(x)$ for all $x \in \mathcal{X}$. Moreover, for each $t \in \{1, \dots, q\}$ we designate a *pair* of hypotheses, $\sigma_{2t}, \sigma_{2t+1}$, that both exclude label 1 at coordinate t:
$$1 \notin \sigma_{2t}(t), \qquad 1 \notin \sigma_{2t+1}(t).$$
For all other coordinates $x \neq t$, these hypotheses include both labels, e.g.
$$\sigma_{2t}(x) = \sigma_{2t+1}(x) = \{0, 1\} \quad \text{for } x \neq t.$$
If $(d-1$ is odd, then there is one remaining index pairing. In that case, define $\sigma_d(x) = \{0, 1\}$ for all $x \in \mathcal{X}$.

Thus $\mathcal{S}$ has size exactly d, uses $2q \leq d-1$ adversarial hypotheses to plant two "anti–1" voters at each coordinate $t \in \mathcal{X}$, and possibly one additional "neutral" hypothesis if $d-1$ is odd. We are now ready to show the online lower bound.

Theorem 9 (Online Lower Bounds for COMMON-INTERSECTIONand MAJORITY)**.** *For every d, there exists a hypothesis class $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ with $|\mathcal{S}| = d, |\mathcal{X}| = \lfloor (d-1)/2 \rfloor$ and $|\mathcal{Y}| = 2$ such that both the rules make $|\mathcal{X}|$ mistakes (i.e. a mistake on every round).*

Proof of Theorem 9. Note that it suffices to show the lower bound of simply the MAJORITY rule, which also implies the lower bound on COMMON-INTERSECTION. Consider the hypothesis class $\mathcal{S}$ constructed above, and let the ground truth be $\sigma_* = \sigma_1$. Present the sequence of instances $x_t = t$ for $t = 1, \dots, q = |\mathcal{X}|$. Then $y_t = 1$ for all t under σ_*.

At each round t, the version space V_{t-1} contains σ_1 together with all adversarial hypotheses that have not yet been eliminated. By construction, every adversarial hypothesis other than σ_1 always includes 0, while at coordinate t at least two of them exclude 1. Hence
$$N_0(x_t; V_{t-1}) = |V_{t-1}| - 1 \quad \text{and} \quad N_1(x_t; V_{t-1}) \leq |V_{t-1}| - 2,$$
so the majority rule predicts 0 (which is an error according to $\sigma_\star = \sigma_1$) and errs, therefore the rule makes an error on every round, completing the proof. $\square$

Remark 5. Note that the rule COMMON-INTERSECTION and MAJORITY respectively recover the textbook rules Consistent and Halving in the standard realizable online classification Shalev-Shwartz and Ben-David [2014]. However, both the rules have a mistake bound of $\Omega(|\mathcal{S}|)$ in our setup even when the labels are binary in the worst case (cf. Theorem 9). This is in sharp contrast with the standard classification where Halving enjoys $\log_2 |\mathcal{H}|$ mistake bound. This failure is due to the set-valued nature of the support hypothesis.

We now show the statistical lower bound in a similar spirit.

Theorem 10 (Statistical Lower Bounds for COMMON-INTERSECTION and MAJORITY)**.** *For every d, there exists a problem instance $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ with $|\mathcal{S}| = d, |\mathcal{X}| = \lfloor (d-1)/2 \rfloor, |\mathcal{Y}| = 2$ and some choice of realizable joint distribution $(\mathcal{D} \times \widetilde{\pi})$ where $\widetilde{\pi}$ is supported on $\sigma_\star \in \mathcal{S}$ such that for any sample size $m \leq |\mathcal{X}|/2$, letting $\widehat{\pi}(S)$ to be either* COMMON-INTERSECTION *or* MAJORITY *have the following guarantee:*
$$\mathbb{P}_{S \sim \mathcal{D}^m} \left(L_{\mathcal{D}, \sigma_\star}(\widehat{\pi}(S)) \geq 1/2 \right) = 1.$$

Proof of Theorem 10. It suffices to prove the claim for MAJORITY again; the bound for COMMON-INTERSECTION then follows since on every point where MAJORITY errs in our construction, the common intersection is empty, forcing COMMON-INTERSECTION to use a fixed default and err as well.

Consider the same class $\mathcal{S}$ constructed above with $|\mathcal{X}| = q$ and ground truth $\sigma_* = \sigma_1$ (so the realizable label is always 1). Let $\mathcal{D}$ be the uniform distribution on $\mathcal{X}$. Take $\widetilde{\pi}$ to be the only conditional distribution supported on σ_*, so the joint $(\mathcal{D} \times \widetilde{\pi})$ is realizable.

Fix any sample size $m \leq q/2$ and draw $S \sim \mathcal{D}^m$. Let $S_{\text{unseen}} \subseteq \mathcal{X}$ be the set of coordinates *unseen* in S; then $|S_{\text{unseen}}| \geq q - m \geq q/2$. Let $V(S) \subseteq \mathcal{S}$ be the version space of hypotheses consistent with S (with respect to the labels of σ_* label, which are always 1).

Again, by construction, for each $t \in S_{\text{unseen}}$ there are two designated adversarial hypotheses in $V(S)$. At such a point $t \in S_{\text{unseen}}$:

- Every hypothesis in $V(S)$ includes label 0, except $\sigma_\star$, so

$$N_0(t; V(S)) = |V(S)| - 1.$$

- Every hypothesis in $V(S)$ includes label 1, except $\sigma_{2t}, \sigma_{2t+1}$, so

$$N_1(t; V(S)) \le |V(S)| - 2.$$

Thus $N_0(t; V(S)) > N_1(t; V(S))$, and the majority rule outputs 0 (which is an error according to $\sigma_\star$). Thus, $\sigma_* = \sigma_1$ is $y_t = 1$, MAJORITY errs on every unseen $t \in S_{\text{unseen}}$.

With $\mathcal{D}$ uniform on $\mathcal{X}$,

$$L_{\mathcal{D},\sigma_\star}(\widehat{\pi}(S)) \ \ge\ \mathbb{P}_{x\sim\mathcal{D}}\big[x \in S_{\text{unseen}}\big] \ =\ \frac{|S_{\text{unseen}}|}{q} \ \ge\ 1 - \frac{m}{q} \ \ge\ \frac{1}{2}.$$

Since this lower bound holds for every realization of S with $m \le q/2$, we have $\mathbb{P}_{S\sim(\mathcal{D},\widetilde{\pi})^m}\big(L_{\mathcal{D},\sigma_\star}(\widehat{\pi}(S)) \ge \frac{1}{2}\big) = 1$. This proves the theorem. $\square$

E Proofs for Section 4.1: Failures of MLE

We first show that there is a simple instance of a support class, where some MLE over the entire class $\Pi_{\mathcal{S}} := \bigcup_\sigma \Pi_\sigma$ fails.

Proof of Theorem 1. Fix any $\gamma \in (0,1)$. Let $\mathcal{Y} = \{0,1\}$ and $\mathcal{X} = \mathbb{N}$. Define a support class $\mathcal{S} = \{\sigma_0, \sigma_{01}\} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ by

$$\sigma_0(x) = \{0\} \qquad \text{and} \qquad \sigma_{01}(x) = \{0,1\} \quad \forall x \in \mathcal{X}.$$

Choose the ground-truth support $\sigma_* = \sigma_0$ and let the data-generating conditional be the point mass $\widetilde{\pi}(\cdot \mid x) = \delta_0(\cdot)$ for all x; thus every observed label equals 0. Let $\Pi_{\mathcal{S}}$ be any class of conditionals π that is compatible with $\mathcal{S}$.

Now fix a sample size $m \in \mathbb{N}$. Set

$$q \ :=\ \left\lceil \frac{m}{\gamma} \right\rceil,$$

and define the marginal $\mathcal{D}$ to be the uniform distribution on $[q] = \{1, 2, \ldots, q\}$, i.e., $\mathcal{D}(\{x\}) = 1/q$ for $x \in [q]$ and 0 otherwise.

For any dataset $S = \{(x_i, y_i)\}_{i=1}^m \sim (\mathcal{D} \times \widetilde{\pi})^m$, write $S_{\text{dis}} := \{x_i : i \in [m]\}$ for the set of distinct unlabeled inputs in S (so $S_{\text{dis}} \le m$). Consider the predictor $\widehat{\pi}$ defined by

$$\widehat{\pi}(\cdot \mid x) = \begin{cases} \delta_0(\cdot), & x \in S_{\text{dis}}, \\ \delta_1(\cdot), & x \notin S_{\text{dis}}. \end{cases} \tag{11}$$

We claim that $\widehat{\pi} \in \mathrm{MLE}_{\Pi_{\mathcal{S}}}(S)$. Indeed, the log-likelihood is

$$\ell_{\log}(\pi; S) = \sum_{i=1}^m \log \pi(0 \mid x_i) = \sum_{x \in S_{\text{dis}}} N_x(S)\, \log \pi(0 \mid x),$$

where $N_x(S) := |\{i : x_i = x\}|$. This expression depends only on the values $\pi(0 \mid x)$ for $x \in S_{\text{dis}}$ and is maximized by setting $\pi(0 \mid x) = 1$ for every $x \in S_{\text{dis}}$. For $x \notin S_{\text{dis}}$ the likelihood does not constrain $\pi(\cdot \mid x)$, so any choice is a (tie-breaking) maximizer; in particular, (11) yields a valid MLE in $\Pi_{\mathcal{S}}$.

We next evaluate its population loss against the support σ_*. Since $\sigma_*(x) = \{0\}$ for all x,

$$L_{\mathcal{D},\sigma_*}(\widehat{\pi}) = \mathbb{P}_{x\sim\mathcal{D},\, \widehat{y}\sim\widehat{\pi}(\cdot|x)}\big(\widehat{y} \notin \sigma_*(x)\big) = \mathbb{P}_{x\sim\mathcal{D}}\big(x \notin S_{\text{dis}}\big) = 1 - \frac{|S_{\text{dis}}|}{q}.$$

Using $|S_{\text{dis}}| \le m$ and $q \ge S_{\text{dis}}/\gamma$,

$$L_{\mathcal{D},\sigma_*}(\widehat{\pi}) \ \ge\ 1 - \frac{m}{q} \ \ge\ 1 - \gamma.$$

The bound holds deterministically for every sample S, hence

$$\mathbb{P}_{S\sim(\mathcal{D}\times\widetilde{\pi})^m}\left(L_{\mathcal{D},\sigma_*}(\widehat{\pi}) \ge 1 - \gamma\right) = 1\,.$$

$\square$

We now show Theorem 2 that the natural choice of restricting the capacity of the class by considering $\overline{\Pi}_{\mathcal{S}} = \bigcup_{\sigma\in\mathcal{S}}\{\overline{\pi}_r\}$ also does not work, when the expert demonstrations $\widetilde{\pi}$ does not necessarily follow the distribution $\pi_{\mathrm{unif},\sigma_\star}$ while still showing examples from $\sigma_\star$.

Proof of Theorem 2. Fix $\gamma \in (0,1)$ and let $s := \lceil 1/\gamma \rceil$. Take $\mathcal{X} = \{x\}$ and

$$\mathcal{Y} = \{y^\star\} \,\cup\, \{a_1,\dots,a_{s-1}\} \,\cup\, \{b_1,\dots,b_s\},$$

so $|\mathcal{Y}| = 1 + (s-1) + s = 2s = 2\lceil 1/\gamma \rceil$. Define $\sigma_1, \sigma_2 \in (2^{\mathcal{Y}})^{\mathcal{X}}$ by

$$\sigma_1(x) = \{y^\star, a_1, \dots, a_{s-1}\} \quad \text{(size } s\text{)}, \qquad \sigma_2(x) = \{y^\star, b_1, \dots, b_s\} \quad \text{(size } s{+}1\text{)},$$

so $\sigma_1(x) \cap \sigma_2(x) = \{y^\star\}$ and they are otherwise disjoint. Let $\mathcal{S} = \{\sigma_1, \sigma_2\}$ and

$$\overline{\Pi}_{\mathcal{S}} := \{\overline{\pi}_r : \sigma \in \mathcal{S}\}, \qquad \overline{\pi}_r(y \mid x) = \begin{cases} \frac{1}{|\sigma(x)|}, & y \in \sigma(x), \\ 0, & \text{otherwise.} \end{cases}$$

Set $\mathcal{D}$ to be the point mass at x and choose the ground-truth support $\sigma_* = \sigma_2$ with data-generating conditional $\widetilde{\pi} = \delta_{y^\star}$ (always emit $y^\star$). For any m, every dataset $S \sim (\mathcal{D} \times \widetilde{\pi})^m$ equals $\{(x, y^\star)\}^m$.

It is simple to see that $\pi_{\mathrm{unif},\sigma_1} \in \mathrm{MLE}_{\overline{\Pi}_{\mathcal{S}}}(S)$ is the unique maximum likelihood estimator. This is because

$$\prod_{i=1}^m \pi_{\mathrm{unif},\sigma_1}(y_i \mid x_i) = \left(\tfrac{1}{s}\right)^m, \qquad \prod_{i=1}^m \pi_{\mathrm{unif},\sigma_2}(y_i \mid x_i) = \left(\tfrac{1}{s+1}\right)^m.$$

However, with $\sigma_*(x) = \sigma_2(x)$ and the estimator $\widehat{\pi}_{\mathrm{mle}}(S) = \pi_{\mathrm{unif},\sigma_1}$ has the error

$$L_{\mathcal{D},\sigma_*}(\widehat{\pi}_{\mathrm{mle}}(S)) = \mathbb{P}_{\widehat{y}\sim\pi_{\mathrm{unif},\sigma_1}(\cdot|x)}\big(\widehat{y} \notin \sigma_2(x)\big) = 1 - \pi_{\mathrm{unif},\sigma_1}(y^\star \mid x) = 1 - \frac{1}{s} \;=\; 1 - \frac{1}{\lceil 1/\gamma \rceil} \;\geq\; 1 - \gamma.$$

All bounds are deterministic given S, hence

$$\mathbb{P}_{S\sim(\mathcal{D}\times\widetilde{\pi})^m}(L_{\mathcal{D},\sigma_*}(\widehat{\pi}_{\mathrm{mle}}(S)) \geq 1 - \gamma) = 1\,,$$

for every m, completing the proof. □

F Algorithms and Proofs for k-pass Error

In this appendix, we provide our guarantees for k-pass error (and formalize Theorem 6). We first start by describing an online rule for that.

Theorem 11 (Online pass-k guarantee). *On any sequence $((x_t, y_t))_{t\in\mathbb{N}}$ realizable by some $\sigma_\star \in \mathcal{S}$, Algorithm 2 makes at most $\log_{k+1}|\mathcal{S}|$ mistakes (i.e., rounds with $\{\widehat{y}_t^{(1)}, \dots, \widehat{y}_t^{(k)}\} \cap \sigma_\star(x_t) = \emptyset$).*

Proof. Let V_t be the version space at the start of round t, and as in the algorithm write $A_y^t = \{\sigma \in V_t : y \in \sigma(x_t)\}$ and $U_t := \bigcup_{i=1}^k A^t_{\widehat{y}_t^{(i)}}$. Define the potential $W_t := w^{(t)}(V_t) = \sum_{\sigma\in V_t} w^{(t)}(\sigma)$, then we again have $\{W_t\}_t$ is non-increasing.

$$\begin{aligned} W_{t+1} = (k+1)w^{(t)}(A^t_{y_t} \setminus U_t) + w^{(t)}(A^t_{y_t}) &= k\, w^{(t)}(A^t_{y_t} \setminus U_t) + w^{(t)}(A^t_{y_t} \setminus U_t) + w^{(t)}(A^t_{y_t} \cap U_t) \\ &\leq w^{(t)}(U_t \setminus A^t_{y_t}) + w^{(t)}(A^t_{y_t} \setminus U_t) + w^{(t)}(A^t_{y_t} \cap U_t) \\ &= w^{(t)}(U_t \cup A^t_{y_t}) \leq W_t\,, \end{aligned}$$

where in the first inequality we used Lemma 3. Now suppose the algorithm makes M pass-k mistakes by the end of round t. On each mistake round we must have $\sigma_\star \in A^t_{y_t} \setminus U_t$, so its weight is multiplied by $(k{+}1)$. Therefore

$$w^{(t+1)}(\sigma_\star) = (k{+}1)^M \;\leq\; W_{t+1} \;\leq\; W_1 = |\mathcal{S}|,$$

which yields $M \leq \log_{k+1}|\mathcal{S}|$. □

Algorithm 2 Online pass-k rule with greedy weighted selection

Input: Hypothesis class $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ with $|\mathcal{S}| < \infty$, parameter $k \in \mathbb{N}$.

- Initialize $V_1 = \mathcal{S}$ and $w^{(1)}(\sigma) = 1$ for all $\sigma \in \mathcal{S}$.
- In every round, receiving x_t:
  1. For each $y \in \mathcal{Y}$, form the slice $A_y^t = \{\sigma \in V_t : y \in \sigma(x_t)\}$.
  2. *(Greedy top-k selection).* Let $\mathcal{Y}_0 = \emptyset$. For $i = 1, 2, \ldots, k$ set
  $$\widehat{y}_t^{(i)} \in \arg\max_{y \in \mathcal{Y} \setminus \mathcal{Y}_{i-1}} w^{(t)}\Big(A_y^t \setminus \bigcup_{z \in \mathcal{Y}_{i-1}} A_z^t\Big), \qquad \mathcal{Y}_i \leftarrow \mathcal{Y}_{i-1} \cup \{\widehat{y}_t^{(i)}\}.$$
  (Break ties arbitrarily.)
  3. Output the k labels $\widehat{y}_t^{(1)}, \ldots, \widehat{y}_t^{(k)}$.
  4. On receiving the realized label y_t, update the version space: $V_{t+1} \leftarrow A_{y_t}^t$.
  5. *(Weight update).* Let $U_t := \bigcup_{i=1}^k A_{\widehat{y}_t^{(i)}}^t$.
  $$w^{(t+1)}(\sigma) \leftarrow (k+1)\, w^{(t)}(\sigma), \text{ for } \sigma \in A_{y_t}^t \setminus U_t,$$
  the weights of hypotheses in $A_{y_t}^t \cap U_t$ are not updated.

Again the heart of the proof is to show that the potential function is non-increasing, which is captured in the following lemma.

Lemma 3 (Removed weight is at least as much as added)**.** *Fix a round t. Let V_t be the current version space with weights $w^{(t)}(\cdot)$. For $y \in \mathcal{Y}$, define the slice $A_y := \{\sigma \in V_t : y \in \sigma(x_t)\}$ and let $A := A_{y_t}$ for the realized label y_t. Let $\widehat{y}_t^{(1)}, \ldots, \widehat{y}_t^{(k)}$ be the greedy top-k labels selected as in Algorithm 2 maximize uncovered weight at each step, and set $U := \bigcup_{i=1}^k A_{\widehat{y}_t^{(i)}}$. Then*

$$w^{(t)}(U \setminus A) \geq k\, w^{(t)}(A \setminus U)\,.$$

Proof. Define the uncovered mass in A after selecting first i labels greedily as:

$$a_i := w^{(t)}\Big(A \setminus \bigcup_{z \in \mathcal{Y}_i} A_z\Big) \quad \text{so that} \quad a_0 = w^{(t)}(A), \ \ a_k = w^{(t)}(A \setminus U), \ \text{ and } a_0 \geq a_1 \geq \cdots \geq a_k.$$

Also, define the uncovered weight for which $\widehat{y}_t^{(i)}$ got picked:

$$m_i := w^{(t)}\Big(A_{\widehat{y}_t^{(i)}} \setminus \bigcup_{z \in \mathcal{Y}_{i-1}} A_z\Big), \quad \text{and define} \quad s_i := a_{i-1} - a_i \ = \ w^{(t)}\Big(\big(A_{\widehat{y}_t^{(i)}} \cap A\big) \setminus \bigcup_{z \in \mathcal{Y}_{i-1}} A_z\Big).$$

By maximality of the greedy choice, we have

$$m_i \ \geq \ a_{i-1} \qquad \text{for all } i \in [k]. \tag{12}$$

The *new* mass *outside* of A at step i that will be removed from the version space is

$$w^{(t)}(A_{\widehat{y}_t^{(i)}} \setminus A \cup A_{\widehat{y}_t^{(1)}} \cdots \cup A_{\widehat{y}_t^{(i-1)}}) = w^{(t)}(A_{\widehat{y}_t^{(i)}} \setminus \bigcup_{z \in \mathcal{Y}_{i-1}} A_z) - w^{(t)}\Big(\big(A_{\widehat{y}_t^{(i)}} \cap A\big) \setminus \bigcup_{z \in \mathcal{Y}_{i-1}} A_z\Big) = m_i - s_i\,.$$

Using (12)

$$m_i - s_i \ \geq \ a_{i-1} - (a_{i-1} - a_i) \ = \ a_i.$$

Summing over $i = 1, \ldots, k$ gives

$$w^{(t)}(U \setminus A) = \sum_{i=1}^k w^{(t)}(A_{\widehat{y}_t^{(i)}} \setminus A \cup A_{\widehat{y}_t^{(1)}} \cdots \cup A_{\widehat{y}_t^{(i-1)}}) = \ \sum_{i=1}^k (m_i - s_i) \ \geq \ \sum_{i=1}^k a_i \ \geq \ k\, a_k \ = \ k\, a,$$

because $(a_i)_i$ is non-increasing and $a_k = a$. This proves the claim. $\square$

F.1 Statistical Upper Bound

Input: Sample $S = \{(x_i, y_i) : i \in [m]\}$ and a finite hypothesis class $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$.

- Run Algorithm 1 (pass-k version) once over S, recording $(V_t, w^{(t)})$ at the *start* of each round $t \in [m]$.
- Find the deterministic predictor $\widehat{\mu}_t : \mathcal{X} \to \mathcal{Y}^k$ used by the online algorithm from the snapshot. I.e. for any $x \in \mathcal{X}$ and any $t \in [m]$, define slices (with respect to V_t)
$$A_y^t(x) \coloneqq \{\sigma \in V_t : y \in \sigma(x)\}.$$
And greedily pick top k labels according to the rule described in Algorithm 2.
Let $\mathcal{Y}_0(x) = \emptyset$. For $i = 1, \ldots, k$ set
$$\widehat{y}_t^{(i)}(x) \in \arg\max_{y \in \mathcal{Y} \setminus \mathcal{Y}_{i-1}(x)} w^{(t)}\Big(A_y^t(x) \setminus \bigcup_{z \in \mathcal{Y}_{i-1}(x)} A_z^t(x)\Big)$$
$$\mathcal{Y}_i(x) \leftarrow \mathcal{Y}_{i-1}(x) \cup \{\widehat{y}_t^{(i)}(x)\},$$
breaking ties by arbitrary fixed rule.
- Then the deterministic predictor $\widehat{\mu}_t : \mathcal{X} \to \mathcal{Y}^k$ is given by:
$$\widehat{\mu}_t(x) \coloneqq (\widehat{y}_t^{(1)}(x), \ldots, \widehat{y}_t^{(k)}(x)).$$
- **Final batch predictor.** On a test input x, draw $I \sim \mathrm{Unif}\{1, \ldots, m\}$ and output
$$\widehat{\mu}_{\mathrm{o2b}}(S)(x) \coloneqq \widehat{\mu}_I(x). \tag{13}$$
(Equivalently: $\widehat{\mu}_{\mathrm{o2b}}$ is the uniform mixture over $\{\widehat{\mu}_t\}_{t=1}^m$.)

Below is the statistical guarantee for this estimator in similar spirit to Theorem 5.

Theorem 12 (Statistical Guarantee for pass-k)**.** *The estimator $\widehat{\mu}_{\mathrm{o2b}}$ in Eq.* (13) *achieves the following guarantee for any finite hypothesis class $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$, and an unknown joint distribution $(\mathcal{D} \times \widetilde{\pi})$ on $\sigma_\star \in \mathcal{S}$.*
$$\mathbb{E}_{S \sim \mathcal{D}^m}\big[L_{\mathcal{D},\sigma_\star}\big(\widehat{\mu}_{\mathrm{o2b}}(S)\big)\big] \le \frac{\log_{k+1}|\mathcal{S}|}{m},$$
and, for any $\delta \in (0,1)$, with probability at least $1-\delta$,
$$L_{\mathcal{D},\sigma_\star}\big(\widehat{\mu}_{\mathrm{o2b}}(S)\big) \le \frac{1 + 2\log_{k+1}|\mathcal{S}| + 12\log\Big(\frac{\log m}{\delta}\Big)}{m}.$$
This implies that $\mathcal{S}$ is learnable (cf. Definition 1) using the estimator $\widehat{\mu}_{\mathrm{o2b}}$ with sample complexity
$$m_{\mathcal{S},\widehat{\mu}_{\mathrm{o2b}}}(\varepsilon,\delta) = O\bigg(\frac{1}{\varepsilon}\Big(\log_{k+1}|\mathcal{S}| + \log\big(\tfrac{1}{\varepsilon\delta}\big)\Big)\bigg).$$

Proof of Theorem 12. The proof is exactly similar to that of Theorem 5 and given for completeness. Let $\ell_t = \mathbb{1}\{\widehat{y}_t^{(i)}(x_t) \notin \sigma_\star(x_t), \forall\, \widehat{y}_t^{(i)}(x_t) \in \widehat{\mu}_t(x_t)\}$. Because $\widehat{\mu}_t$ is a deterministic function of $S_{<t} = \{(x_i, y_i) : i < t\}$, we have
$$\mathbb{E}[\ell_t \mid S_{<t}] = L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_t).$$
Hence
$$\mathbb{E}_S\left[L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_{\mathrm{o2b}}(S))\right] = \mathbb{E}_S\left[\frac{1}{m}\sum_{t=1}^m L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_t)\right] = \mathbb{E}_S\left[\frac{1}{m}\sum_{t=1}^m \ell_t\right] \le \frac{\log_{k+1}|\mathcal{S}|}{m},$$
where in the last inequality we used Theorem 11, which guarantees $\sum_{t=1}^m \ell_t \le \log_{k+1}|\mathcal{S}|$.

For the high-probability statement, define the martingale differences
$$Z_t := L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_t) - \ell_t, \qquad \text{where } |Z_t| \le 1 \text{ almost surely.}$$

Then $\mathbb{E}[Z_t \mid S_{<t}] = 0$, and

$$\mathbb{E}[Z_t^2 \mid S_{<t}] = \mathbb{E}\big[(L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_t) - \ell_t)^2 \mid S_{<t}\big] = \mathrm{Var}(\ell_t \mid S_{<t}) = L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_t)\big(1 - L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_t)\big) \ \leq\ L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_t)\,.$$

Taking $W_m = \sum_{t=1}^m L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_t)$ and $\sigma^2 = m$ suffices; thus, using Lemma 2 with $n = \log m$ gives, with probability $1-\delta$,

$$\begin{aligned}\sum_{t=1}^m Z_t &\leq \sqrt{8\left(1 + \sum_{t=1}^m L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_t)\right)\log\left(\frac{\log m}{\delta}\right)} + \frac{4}{3}\log\left(\frac{\log m}{\delta}\right) \\ &\leq \frac{1}{2}\left(1 + \sum_{t=1}^m L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_t)\right) + 4\log\left(\frac{\log m}{\delta}\right) + \frac{4}{3}\log\left(\frac{\log m}{\delta}\right) \qquad (\text{GM} \leq \text{AM})\end{aligned}$$

Substituting Z_t and rearranging terms,

$$\sum_{t=1}^m L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_t) \ \leq\ 1 + 2\sum_{t=1}^m \ell_t + 12\log\left(\frac{\log m}{\delta}\right).$$

Finally, noting that $L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_{\mathrm{o2b}}(S)) = \frac{1}{m}\sum_{t=1}^m L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_t)$ and that $\sum_{t=1}^m \ell_t \leq \log_{k+1}|\mathcal{S}|$ (by Theorem 11), we obtain that with probability $1-\delta$,

$$L_{\mathcal{D},\sigma_\star}(\widehat{\mu}_{\mathrm{o2b}}(S)) \leq \frac{1 + 2\log_{k+1}|\mathcal{S}| + 12\log\left(\frac{\log m}{\delta}\right)}{m}.$$

□

F.2 Lower Bounds for Online and Statistical Settings for k-pass Error

We next provide a lower bound that, information-theoretically, this dependence cannot be improved and we only gain a factor of $1/\log k$ in sample complexity as well as mistake bound in the worst-case.

Theorem 13 (Online $\Omega(\log_{k+1}|\mathcal{S}|)$ mistake lower bound in multiclass when k outputs allowed)**.** *Fix integers $k \geq 1$ and $d \geq 2$. There exists a problem instance $\mathcal{S} \subseteq \mathcal{Y}^{\mathcal{X}}$ with $|\mathcal{S}| \leq d, |\mathcal{Y}| = k+1, |\mathcal{X}| = \lfloor \log_{k+1} d \rfloor$ such that for any deterministic online learning algorithm that outputs at most k labels, there exists a sequence $(x_t, y_t)_{t\in[|\mathcal{X}|]}$ realizable by some $\sigma_\star \in \mathcal{S}$ such that the algorithm makes mistake on every round.*

Note that our instance is an instance of multiclass classification problem $\sigma : \mathcal{X} \to \mathcal{Y}$. This is isomorphic to an instance $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$, where $|\sigma(x)| = 1$ for all $x \in \mathcal{X}, \sigma \in \mathcal{S}$.

Proof of Theorem 13. Let $m := \lfloor \log_{k+1} d \rfloor$ and take $\mathcal{X} = \{1, \ldots, m\}$ and $\mathcal{Y} = \{1, \ldots, k+1\}$. Consider the full product class $\mathcal{S} = \mathcal{Y}^{\mathcal{X}}$, which has size $(k+1)^m \leq d$. First of all, observe that in any round in which y_t does not belong to the list of $(\widehat{y}_t^{(1)}, \ldots, \widehat{y}_t^{(k)})$, the mistake is made because we are in the multiclass classification setting.

For rounds $t \in [m]$, present a fresh coordinate $x_t = t$. Since $|\mathcal{Y}| = k+1$, there exists a label $y_t \in \mathcal{Y}$ that the learner failed to output in the set; $y_t \neq \widehat{y}_t^{(i)}$ for all $i \in [k]$. Reveal this y_t. This forces a mistake on every round. Moreover, this sequence is realizable since $\mathcal{S} = \mathcal{Y}^{\mathcal{X}}$ contains all functions from $\mathcal{X}$ to $\mathcal{Y}$. □

Theorem 14 (Statistical lower bound of $\Omega(\log_k|\mathcal{S}|)$)**.** *Fix integers $k \geq 1$ and $q \geq 1$. Let $\mathcal{X} = \{1, \ldots, q\}$, $\mathcal{Y} = \{1, \ldots, 2k\}$, and take the hypothesis class $\mathcal{S} = \mathcal{Y}^{\mathcal{X}}$ (all multiclass functions), so its cardinality is $d := |\mathcal{S}| = (2k)^q$. Let $\mathcal{D}$ be the uniform distribution on $\mathcal{X}$. Then for any estimator $\widehat{\mu} : (\mathcal{X} \times \mathcal{Y})^* \to \Delta(\mathcal{Y}^k)^{\mathcal{X}}$*

$$\inf_{\widehat{\mu}} \sup_{\sigma\in\mathcal{S}} \ \mathbb{E}_{S\sim(\mathcal{D}\times\sigma)^m}\,\mathbb{P}_{x\sim\mathcal{D}}\mathbb{P}_{\widehat{\boldsymbol{y}}(x)\sim\widehat{\mu}(\cdot|x)}\big[\,\sigma(x) \notin \widehat{\boldsymbol{y}}(x)\,\big] \ \ \geq\ \ \tfrac{1}{2}\Big(1 - \tfrac{1}{q}\Big)^m.$$

In particular, to ensure expected error at most $0 < \varepsilon < \frac{1}{2}$ for all $\sigma \in \mathcal{S}$, one needs

$$m \ \geq\ \frac{\ln(1/(2\varepsilon))}{-\ln(1 - 1/q)} \ \ \geq\ \ q\,\ln\Big(\frac{1}{2\varepsilon}\Big).$$

Proof. Fix any (possibly randomized) estimator $\widehat{\mu}$. Let $S = \{(x_i, y_i)\}_{i=1}^m$ be the training sample drawn i.i.d. from $(\mathcal{D} \times \sigma)$ for $\sigma \sim \mathrm{Unif}(\mathcal{S})$, and let $U_S = \{x_i : 1 \le i \le m\} \subseteq \mathcal{X}$ be the set of distinct inputs seen in S. Draw $x \sim \mathcal{D}$ independently of S and then $\widehat{\boldsymbol{y}}(x) \sim \widehat{\mu}(S)(\cdot \mid x) \in \mathcal{Y}^k$.

On any $x \notin U_S$, under the prior where σ is uniform over $\mathcal{S}$, for any (possibly randomized) k-list $\widehat{\boldsymbol{y}}(x) \sim \widehat{\mu}(\cdot \mid x)$,

$$\mathbb{P}_\sigma\big[\sigma(x) \in \widehat{\boldsymbol{y}}(x) \mid S,\, x \notin U_S\big] = \mathbb{E}\left[\frac{\#\text{ of distinct labels in }\widehat{\boldsymbol{y}}(x)}{|\mathcal{Y}|} \;\middle|\; S,\, x \notin U_S\right] \le \frac{k}{2k} = \frac{1}{2},$$

so $\mathbb{P}_\sigma[\sigma(x) \notin \widehat{\boldsymbol{y}}(x) \mid S,\, x \notin U_S] \ge \frac{1}{2}$. (Allowing duplicates in $\widehat{\boldsymbol{y}}(x)$ cannot increase coverage.)

If $x \in U_S$, the learner can always include the observed label and incur zero error on that x. Therefore, for any estimator $\widehat{\mu}$,

$$\mathbb{P}_{\sigma, x, \widehat{\boldsymbol{y}}(x) \sim \widehat{\mu}(\cdot|x)}\big[\sigma(x) \notin \widehat{\boldsymbol{y}}(x) \mid S\big] \;\ge\; \frac{1}{2} \cdot \mathbb{P}[x \notin U_S].$$

Taking expectation over S and using $\mathcal{D} = \mathrm{Unif}(\mathcal{X})$ yields

$$\mathbb{E}_S\, \mathbb{P}_{\sigma, x, \widehat{\boldsymbol{y}}(x) \sim \widehat{\mu}(\cdot|x)}\big[\sigma(x) \notin \widehat{\boldsymbol{y}}(x)\big] \;\ge\; \frac{1}{2}\, \mathbb{E}_S\big[1 - |U_S|/q\big] \;=\; \tfrac{1}{2}\Big(1 - \tfrac{1}{q}\Big)^m,$$

since $\mathbb{E}[|U_S|] = q\big(1 - (1 - \frac{1}{q})^m\big)$. Finally, by minimax principle

$$\begin{aligned}\inf_{\widehat{\mu}} \sup_{\sigma \in \mathcal{S}} \mathbb{E}_S\, \mathbb{P}_{x \sim \mathcal{D}} \mathbb{P}_{\widehat{\boldsymbol{y}}(x) \sim \widehat{\mu}(\cdot|x)}\big[\,\sigma(x) \notin \widehat{\boldsymbol{y}}(x)\,\big] &\ge \inf_{\widehat{\mu}}\; \mathbb{E}_{\sigma \sim \mathrm{Unif}(\mathcal{S})}\, \mathbb{E}_S\, \mathbb{P}_{x \sim \mathcal{D}} \mathbb{P}_{\widehat{\boldsymbol{y}}(x) \sim \widehat{\mu}(\cdot|x)}\big[\,\sigma(x) \notin \widehat{\boldsymbol{y}}(x)\,\big] \\ &\ge\; \tfrac{1}{2}\Big(1 - \tfrac{1}{q}\Big)^m.\end{aligned}$$

For the sample-complexity bound, solve $\frac{1}{2}(1 - \frac{1}{q})^m \le \varepsilon$ for m and use $-\ln(1 - 1/q) \le 1/q$. $\square$

Because $d = (2k)^q$, we have $q = \log_{2k} d$, so the bound implies $m = \Omega\big(\log_k d\big)$ under $\mathcal{D} = \mathrm{Unif}(\mathcal{X})$.

Remark 6. Both online (Theorem 13) and statistical lower bounds (Theorem 13) for k-pass essentially demonstrate that one cannot do better than memorization below $\Omega(\log_k d)$ barrier in the worst-case, even for the special case of the problem of multiclass classification $\mathcal{S} \subseteq \mathcal{Y}^{\mathcal{X}}$ which is isomorphic to $\mathcal{S} \subseteq (2^{\mathcal{Y}})^{\mathcal{X}}$ with $|\sigma(x)| = 1$.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: **[TODO]**

Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.

- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: **[TODO]**

    Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: **[TODO]**

Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.