The Good, the Bad, and the Sampled: a No-Regret Approach to Safe Online Classification

Tavor Baharav*, Spyros Dragazis*, Aldo Pacchiano

AISTATS 2026

May 4, 2026



Problem formulation

Safe Online Classification protocol

Input: error tolerance α, confidence level δ

For $t = 1, 2, \ldots, T$:

1. $\xleftarrow{\mathbf{X}_t \in \mathbb{R}^d}$, $\mathbf{X}_t \overset{iid}{\sim} P$
2. $\xrightarrow[\text{test/predict}]{Z_t \in \{0,1\}}$

If "predict" $(Z_t == 0)$:

$\xrightarrow{\hat{Y}_t}$, no feedback Y_t

Else "test":

$\xrightarrow{\hat{Y}_t = Y_t}$, observe Y_t

Output: Action, prediction sequence $\{Z_t, \hat{Y}_t\}_{t=1}^{T}$

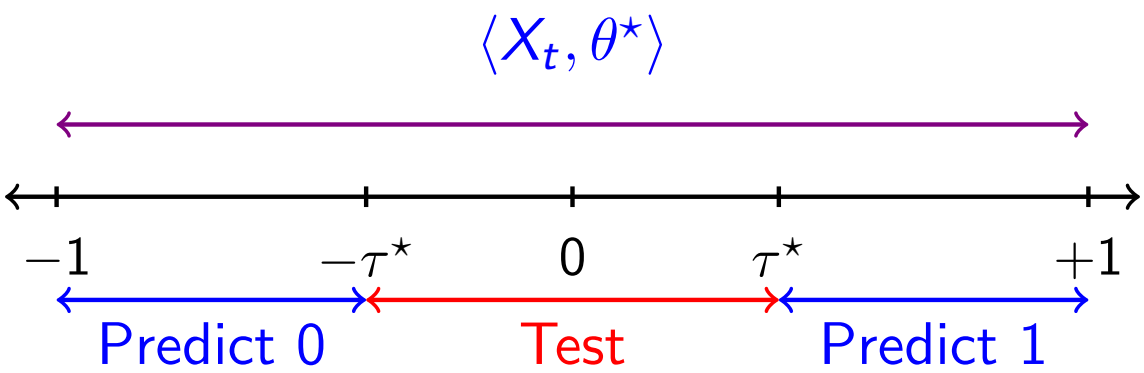
Logistic model

$$\mathbb{P}(Y = 1 \mid X_t) = \sigma\left(\langle X_t, \theta^\star \rangle\right)$$

Assumptions

- Contexts are i.i.d.
- P is smooth
- Logistic link model
- Bounded features / parameters

Baseline policy and safe regret



Knowing P and $\theta^\star$, an oracle policy can compute the threshold $\tau^\star$ such that

$$\mathbb{P}(\text{missclassification when testing with threshold } \tau^\star) = \alpha.$$

This induces the oracle testing frequency

$$p^\star = \mathbb{P}_{X \sim P}\left(\left|X^\top \theta^\star\right| \leq \tau^\star\right).$$

We define the safe regret of a policy by

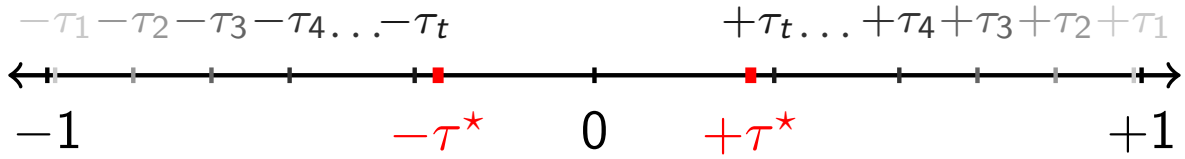
$$\texttt{Regret}_{\text{safe}}(T) = \mathbb{E}\left[\sum_{t=1}^{T}(Z_t - p^\star)\right],$$

subject to **safety constraint**

$$\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\left\{\widehat{Y}_t \neq Y_t\right\} \leq \alpha\right) \geq 1 - \delta.$$

Algorithm SCOUT, Result

- SCOUT computes the MLE as $\hat{\theta}_t$, tests when $|\langle X_t, \hat{\theta}_t \rangle| \leq \tau_t$



Theorem 1 (Informal)

When $p^\star > 0$ the regret of SCOUT is upper bounded by

$$\mathtt{Regret}_{SCOUT} \leq \mathcal{O}\left(\sqrt{\frac{dT\log(T/\delta)}{p^\star \lambda_0}}\right),$$

and achieves anytime (α, δ)-safety.

where $\lambda_0 \overset{\texttt{Lemma 1}}{>} 0$ is the minimum eigenvalue of the oracle baseline policy.

Technical Highlights

Pessimism: SCOUT tests whenever the baseline policy tests $\rightarrow$ **safety**.

$\hat{\theta}$ **converges to** $\theta^\star$: The minimum eigenvalue of the empirical covariance matrix grows linearly in t, i.e. $\lambda_{\min}^t \geq \frac{p^\star t \lambda_0}{12} \overset{\text{Lemma 1}}{>} 0$.

Stability with respect to unknown parameters $P, \theta^\star$. We show that:

$$\underbrace{\mathbb{P}_{\hat{P}}\left(|\langle X, \hat{\theta}\rangle| > \hat{\tau}\right)}_{\texttt{SCOUT tests}} \approx \underbrace{\mathbb{P}_{P}\left(|\langle X, \theta^\star\rangle| > \tau^\star\right)}_{\texttt{when baseline tests}} \quad \texttt{low regret.}$$

References

[1] Marc Abeille, Louis Faury, and Clément Calauzènes. Instance-wise minimax-optimal algorithms for logistic bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 3691–3699. PMLR, 2021.

[2] Francesco Orabona, Nicolo Cesa-Bianchi, et al. Better algorithms for selective sampling. In *Proceedings of the 28th international conference on machine learning: Bellevue, Washington, USA*, pages 433–440. Omnipress, 2011.

[3] Ayush Sekhari, Karthik Sridharan, Wen Sun, and Runzhe Wu. Selective sampling and imitation learning via online regression. *Advances in Neural Information Processing Systems*, 36:67213–67268, 2023.

Online Safe Classification

Key takeaway

We can simultaneously learn and maintain safety in online classification.

- Safety **constraint** with censored feedback is more medically practical
- Ongoing work: application to Malaria screening

Come find us: Poster 193

Contact

tavorb@mit.edu

dragazis@bu.edu

pacchian@bu.edu

arXiv

GitHub