Prospective Learning: Learning for a Dynamic Future

Ashwin De Silva[1*], Rahul Ramesh[2*], Rubing Yang[2*],
Siyu Yu[1], Joshua T. Vogelstein[2,†], Pratik Chaudhari[2,†]

[1]Johns Hopkins University, [2]University of Pennsylvania

†, ∗ Equal Contribution

Arxiv: arxiv.org/abs/2411.00109

Code: github.com/neurodata/prolearn

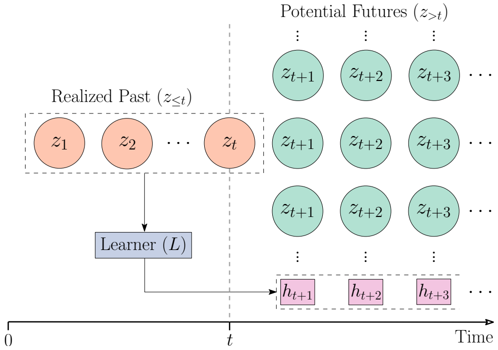
Johns Hopkins University

Penn
University of Pennsylvania

PAC-learning meets time

PAC learning assumes that the distribution of future samples is identical to the past.

But what if the distribution or goals change over time?



We propose **prospective learning**, a theoretical framework which defines learnability with respect to a stochastic process.

Prospective learning

Data. $z_t = (x_t, y_t)$ is the datum at time t. Data is drawn from a stochastic process $Z \equiv (Z_t)_{t \in \mathbb{N}}$.

Hypothesis class: A prospective learner selects an infinite sequence of hypotheses $h \equiv (h_1, \ldots, h_t, h_{t+1}, \ldots)$ where $h_t : \mathcal{X} \mapsto \mathcal{Y}$.

Prospective loss: Future loss incurred by a hypothesis h

$$\bar{\ell}_t(h, Z) = \limsup_{\tau \to \infty} \frac{1}{\tau} \sum_{s=t+1}^{t+\tau} \ell(s, h_s(X_s), Y_s)$$

Prospective risk: Prospective risk at time t is the expected future loss

$$R_t(h) = \mathbb{E}\left[\bar{\ell}_t(h, Z) \mid z_{\leq t}\right] = \int \bar{\ell}_t(h, Z) \, \mathrm{d}\mathbb{P}_{Z \mid z_{\leq t}}$$

Prospective learnability

Definition (Strong Prospective Learnability)

A family of stochastic processes is strongly prospectively learnable, if there exists a learner with the following property: there exists a time $t'(\epsilon, \delta)$ such that for any $\epsilon, \delta > 0$ and for any stochastic process Z from this family, the learner outputs a hypothesis h such that

$$\mathbb{P}\left[R_t(h) - R_t^* < \epsilon\right] \geq 1 - \delta,$$

for any $t > t'$.

Prospective learnability

Theorem (Prospective ERM is a strong prospective learner)

Consider a finite family of stochastic processes $\mathcal{Z}$. If we have (a) consistency, i.e., there exists an increasing sequence of hypothesis classes $\mathcal{H}_1 \subseteq \mathcal{H}_2 \subseteq \dots$ with each $\mathcal{H}_t \subseteq (\mathcal{Y}^{\mathcal{X}})^{\mathbb{N}}$ such that $\forall Z \in \mathcal{Z}$,

$$\lim_{t \to \infty} \mathbb{E}\left[\inf_{h \in \mathcal{H}_t} R_t(h) - R_t^*\right] = 0, \tag{1}$$

where $h \in \mathcal{H}_t$ is a random variable in $\sigma(Z_{\leq t})$, and (b) uniform concentration of the limsup, i.e., $\forall Z \in \mathcal{Z}$,

$$\mathbb{E}\left[\max_{h \in \mathcal{H}_t} \left| \bar{\ell}_t(h, Z) - \max_{u_t \leq m \leq t} \frac{1}{m} \sum_{s=1}^{m} \ell(s, h_s(x_s), y_s) \right| \right] \leq \gamma_t, \tag{2}$$

for some $\gamma_t \to 0$ and $u_t \to \infty$ with $u_t \leq t$ (all uniform over the family of stochastic processes), then there exists a sequence i_t that depends only on γ_t such that a learner that returns

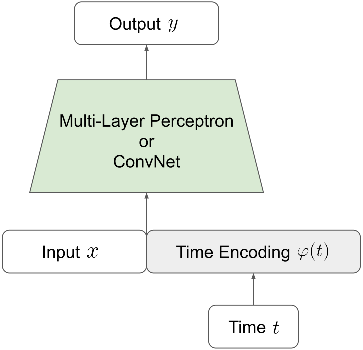
$$\hat{h} = \operatorname*{arg\,min}_{h \in \mathcal{H}_{i_t}} \max_{u_{i_t} \leq m \leq t} \frac{1}{m} \sum_{s=1}^{m} \ell(s, h_s(x_s), y_s), \tag{3}$$

is a strong prospective learner for this family. We define Prospective ERM as the learner that implements (3) given train data $z_{\leq t}$.

Implementing a prospective learner

We encode absolute time using sines and cosines

$$t \mapsto \varphi(t) = (\sin(\omega_1 t), \ldots, \sin(\omega_{d/2} t), \cos(\omega_1 t), \ldots, \cos(\omega_{d/2} t))$$



The neural network is a function of both absolute time t and input x.

The time encoding can be concatenated near the first few closer to the last few layers.

Experimental results