On the Value of Stochastic Side Information in Online Learning

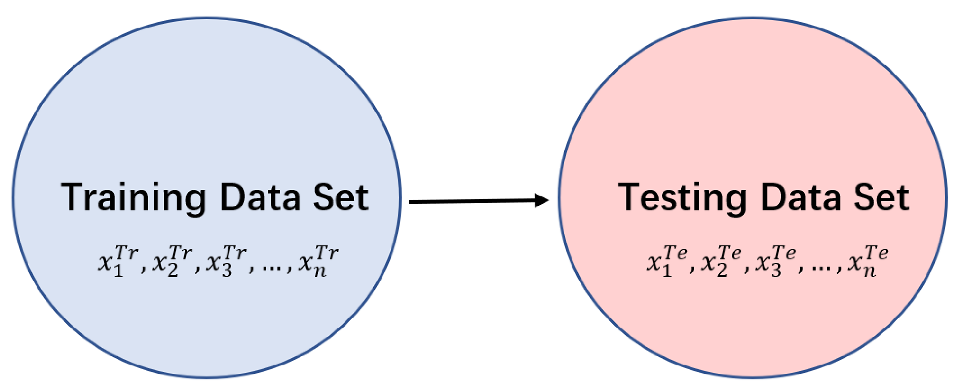
Junzhang Jia, Xuetong Wu, Jamie Evans, Jingge Zhu

Department of Electrical and Electronic Engineering, The University of Melbourne

Introduction to Online learning

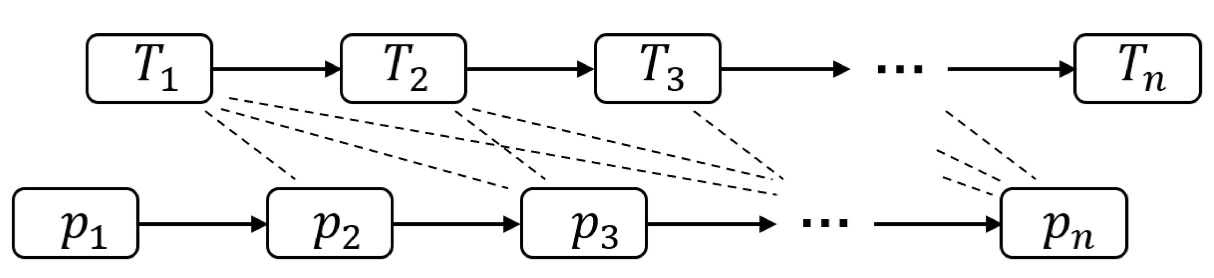
Traditional Machine Learning:

- All training data are known initially.



Online Learning:

- The data arrive sequentially.



Experts and Side Information

Target		x_1^T	x_2^T	x_3^T	$\cdots$	x_n^T
Experts	$\theta = 1$	f_1^1	f_2^1	f_3^1	$\cdots$	f_n^1
	$\theta = 2$	f_1^2	f_2^2	f_3^2	$\cdots$	f_n^2
	$\vdots$					$\vdots$
	$\theta = N$	f_1^N	f_2^N	f_3^N	$\cdots$	f_n^N
Sequential Side Information		s_1	s_2	s_3	$\cdots$	s_n

A Simple Motivation
Weather Forecaster System

Target:
Local weather

Experts:
Predictions from different models

Side information:
The weather from a nearby area

Minimax Expected Regret

Maximum over all possible target sequences

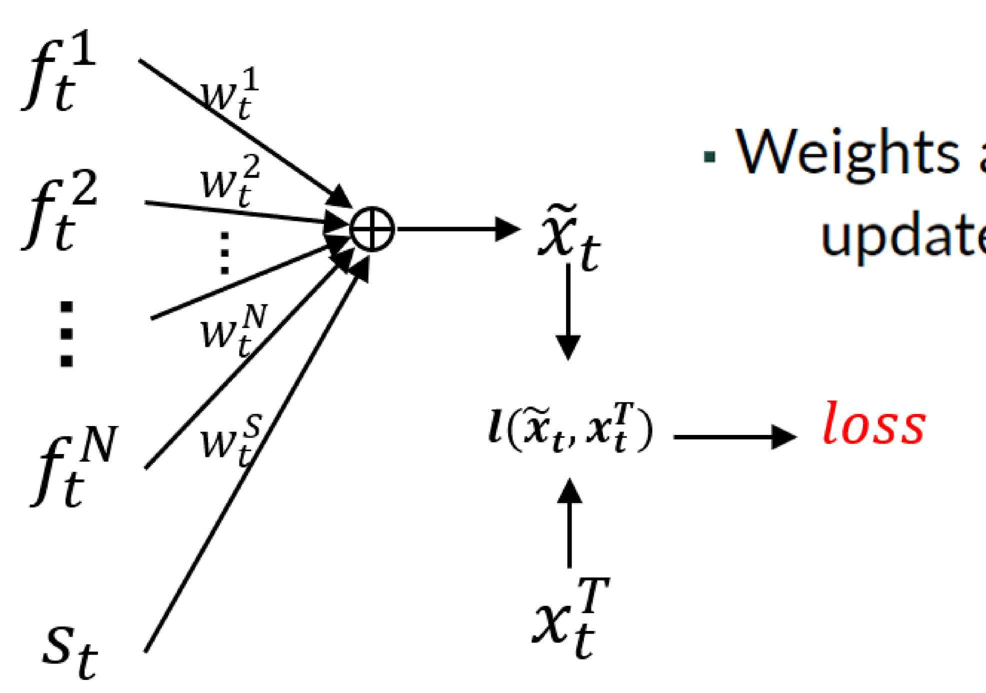
Comparing to the best expert

$$R := \min_{\tilde{x}_1,\ldots,\tilde{x}_n} \max_{x_1^T,\ldots,x_n^T} \mathbb{E}_{s_1,\ldots,s_n} \left\{ \sum_{t=1}^{n} \ell(\tilde{x}_n(s_1,\ldots,s_t), x_t^T) - \min_{\theta} \sum_{t=1}^{n} \ell(f_t^\theta, x_t^T) \right\}$$

Expectation over the stochastic side information

Minimum over all possible algorithms

Weighted Average Algorithm



- Weights are exponentially updated by their induced loss.

Main Contribution

This work mainly contributes to three aspects:

- **A Novel Setup:** We proposed a novel online learning problem formulation with stochastic side information.
- **Negative Regret:** We showed that introducing SSI can improve the typical learning rate under some possible cases.

Results

Theorem 1(Upper Bound): The minimax expected regret for predicting the target sequence with length n:

$$R \leq \sqrt{\frac{n}{2}\log(N+1)} + \min\{F(n), 0\},$$

where $F(n)$ depends on the problems' specifications.

Theorem 2(Lower Bound): Consider $\mathcal{X} = \{0,1\}$ and $\mathcal{D} = [0,1]$, and ℓ is the absolute loss $\ell(x,y) = |x-y|$, then

$$R \geq \sqrt{\frac{n}{2}\log(N+1)} + \left(\xi^* - \frac{1}{2}\right) n,$$

where ξ^* is a constant related to side information.

Both upper and lower bound could be negative!

Examples

SSI via a binary symmetric channel

$$R \leq \sqrt{\frac{n}{2}\log(N+1)} + \min\{0, nc_1\}$$
$$R \geq \sqrt{\frac{n}{2}\log(N+1)} + nc_2.$$

SSI via a Zero-mean Gaussian Channel

$$R \leq \sqrt{\frac{n}{2}\log(N+1)} + \min\{0, nc_3\}$$
$$R \geq \sqrt{\frac{n}{2}\log(N+1)} + nc_4.$$