Hierarchies of Relaxations for Online Prediction Problems with Evolving Constraints

Alexander Rakhlin
University of Pennsylvania

Karthik Sridharan
Cornell University
Problem Setup

Online prediction protocol:

For $t = 1$ to $V$

- Side information $x_t \in X$
- Set $C_t$ of constraints is revealed

Learner predicts $\hat{y}_t \in \mathbb{Y}$

Label $y_t \in \mathbb{Y}$ is revealed

End

Goal: minimize regret

$$\text{Reg}_V := V \sum_{t=1}^{V} \mathbb{1}\{\hat{y}_t \neq y_t\} - \inf_{f \in F} \sum_{t=1}^{V} \mathbb{1}\{f(x_t) \neq y_t\} \approx \times \text{OPT}$$
Online prediction protocol:

For $t = 1$ to $V$

Side information $x_t \in X$ and set $C_t$ of constraints is revealed

Learner predicts $\hat{y}_t \in \mathbb{Y}$

Label $y_t \in \mathbb{Y}$ is revealed

End

Goal: minimize regret

$$\text{Reg}_V := \min_{t=1}^{V} \{ \hat{y}_t \neq y_t \} - \inf_{f \in F} \{ \text{data} \} \min_{t=1}^{V} \{ f(x_t) \neq y_t \} \leq \times \text{OPT} x_t$$
Problem Setup

Online prediction protocol:

For $t = 1$ to $V$

- Side information $x_t \in X$
- Set $C_t$ of constraints is revealed

Learner predicts $\hat{y}_t \in [\mathcal{Y}]$

Label $y_t \in [\mathcal{Y}]$ is revealed

End

Goal: minimize regret

$$\text{Reg}_V : = V \sum_{t=1}^{V} 1 \{ \hat{y}_t \neq y_t \} - \inf_{f \in F} \sum_{t=1}^{V} 1 \{ f(x_t) \neq y_t \} \leq \frac{1}{2} \times \text{OPT}$$
Problem Setup

Online prediction protocol:

For $t = 1$ to $V$

Side information $x_t \in X$

and set $C_t$ of constraints is revealed

Learner predicts $\hat{y}_t \in \mathbb{Y}$

Label $y_t \in \mathbb{Y}$ is revealed

End

Goal: minimize regret

$\text{Reg}_V := V \sum_{t=1}^{V} 1 \{ \hat{y}_t \neq y_t \} - \inf_{f \in F}[\text{data}] V \sum_{t=1}^{V} 1 \{ f(x_t) \neq y_t \}$

$\times \text{OPT}$
Online prediction protocol:
For $t = 1$ to $V$
Side information $x_t \in X$ and set $C_t$ of constraints is revealed
Learner predicts $\hat{y}_t \in \mathbb{Y}$
Label $y_t \in \mathbb{Y}$ is revealed
End

Goal: minimize regret

\[
\text{Reg} := \sum_{t=1}^{V} \mathbf{1}\{\hat{y}_t \neq y_t\} - \inf_{f \in \mathcal{F}[\text{data}]} \sum_{t=1}^{V} \mathbf{1}\{f(x_t) \neq y_t\} \leq 1 \times \text{OPT}
\]
Problem Setup

Online prediction protocol:

For \( t = 1 \) to \( V \)

- Side information \( x_t \in \mathcal{X}_t \) and set \( \mathcal{C}_t \) of constraints is revealed
- Learner predicts \( \hat{y}_t \in [\kappa] \)
- Label \( y_t \in [\kappa] \) is revealed

End

Goal: minimize regret

\[
\text{Reg} := \sum_{t=1}^{V} \mathbf{1}\{\hat{y}_t \neq y_t\} - \inf_{f \in \mathcal{F}[\text{data}]} \sum_{t=1}^{V} \mathbf{1}\{f(x_t) \neq y_t\} \\
\leq 1 \times \text{OPT}
\]
Each \( c \in \mathcal{C}_t \) is represented by pair \((S_c, R_c)\) where \( S_c \subseteq V \) and \( R_c : [K]^{S_c} \rightarrow \mathbb{R}_{\geq 0} \).

We assume we know the stochastic model generating constraints and side information, ie. \( p(\mathcal{C}_t, x_t | \mathcal{C}_{1:t-1}, x_{1:t-1}) \).

Benchmark based on evolving constraint:

\[
\mathcal{F}[\text{data}] = \mathcal{F}_K [((\mathcal{C}, x)_{1:V} = \left\{ f \in \mathcal{F} : \sum_{c \in U \mathcal{C}_t} c(f(x_1, \ldots, x_V)) \leq K \right\}
\]

Example: binary node classification on a graph \( G = (V, E) \), given labeling \( g \in \{1, 2\}^V \), constraint \( c \) for each edge \((u, v) \in E\) represented as

\( S_c = (u, v) \) and \( R_c(g_u, g_v) = 1_{\{g_u \neq g_v\}} \).
Minimize regret (in expectation):

$$\mathbb{E}[\text{Reg}] = \sum_{t=1}^{V} \mathbb{E}[\mathbf{1}\{\hat{y}_t \neq y_t\}] - \inf_{f \in \mathcal{F}_K[(\mathcal{C}, x)_{1:V}]} \sum_{t=1}^{V} \mathbf{1}\{f(x_t) \neq y_t\}$$

- Even when we are given all the data, often finding best 
  $$f \in \mathcal{F}_K[(\mathcal{C}, x)_{1:V}]$$ is NP hard!

- Can we still predict as well as this model efficiently?
  Yes, via improper learning.

In statistical learning: [Kearns et al’94], [Srebro, Shraibman’05], [Shalev-Shwartz et al’11], [Shalev-Shwartz, Birnbaum’12], [Daniely et al’13,14]. In online learning: [Hazan et al’12], [Christiano’14]
Online Node Classification
Online Node Classification
Online Node Classification
Online Node Classification

$F_K[\bigcup_t \mathcal{C}_t, x_{1:V}]$
Online Node Classification

\[ \mathcal{F}_K \left[ \bigcup_t \mathcal{C}_t, x_{1:V} \right] \]

SDP Relax
Theorem

For $r^{th}$ level Lasserre hierarchy and the randomized strategy described above,

$$\mathbb{E}[\text{Reg}] \leq 2 \text{gap}(r) \mathbb{E}[\text{Rad}(\mathcal{F}_K[(C, x)_{1:V}])]$$

- We prove lower bound showing that the Rademacher rate is tight
- Approximation factor multiplies regret bound as opposed to $\text{OPT}$

Proof outline:

- We compete with a larger class (the SDP)
- But for final regret bound, Rademacher complexity can be viewed as original optimization problem with extra linear constraint
SUMMARY

• Examples: binary node classification with bounded disagreements, unique games prediction, metric labeling, …

• Trade off: computation Vs regret bound (level $r$ of Lasserre)

• Prediction algorithm only needs value of SDP

• Rounding not needed for prediction

• Existence of rounding (integrality gap) provides regret bound

• Lower bound matches up to integrality gap