Minimax Fixed-Design Linear Regression

Peter L. Bartlett, Wouter M. Koolen, Alan Malek, Eiji Takimoto, Manfred Warmuth

Conference on Learning Theory
Paris, France
July 5th, 2015
Context: Linear regression

- We have data \((x_1, y_1), \ldots, (x_T, y_T)\)
- Offline linear regression: predict \(\hat{y} = \theta^T x\), where
  \[
  \theta_T = (X^T X)^{-1} X^T Y.
  \]
- Online linear regression:
  1. We see \(x_1, \ldots, x_T\) before hand
  2. Need to predict \(\hat{y}_t\) before seeing \(y_t\)
Protocol

Given: \( \mathbf{x}_1, \ldots, \mathbf{x}_T \in \mathbb{R}^d \)
For \( t = 1, 2, \ldots, T \):

- Learner predicts \( \hat{y}_t \in \mathbb{R} \),
- Adversary reveals \( y_t \in \mathbb{R} \),
- Learner incurs loss \( (\hat{y}_t - y_t)^2 \).

Figure: Fixed-design protocol
Our goal is to find a strategy that achieves the minimax regret:

\[
\min_{\hat{y}_1} \max_{y_1} \cdots \min_{\hat{y}_T} \max_{y_T} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\theta \in \mathbb{R}^d} \sum_{t=1}^{T} (\theta^T x_t - y_t)^2
\]
The Minimax Strategy

- Is linear

\[
\hat{y}_t = s_{t-1}^T P_t x_t \quad \text{where} \quad s_t = \sum_{q=1}^{t} x_q y_q,
\]

- with coefficients:

\[
P_t^{-1} = \sum_{q=1}^{t} x_q x_q^T + \sum_{q=t+1}^{T} \frac{x_q P_q x_q}{1 + x_q P_q x_q} x_q x_q^T.
\]

  - least squares
  - re-weighted future instances

- Cheap recursive calculation, can be done before seeing \(y_t\)s.
- Minimax under alignment condition and \(|y_t| \leq B\)
Guarantees

▶ If the adversary plays \( y_t \) with

\[
\sum_{t=1}^{T} y_t^2 x_t^\top P_t x_t = R,
\]

we are minimax against all \( y_t \)s in this set

▶ Minimax strategy does not depend on \( R \)

▶ We achieve regret exactly \( R = O(\log T) \)

▶ Visit us at the poster session!