The Optimal Strategy for a Linear Regression Game

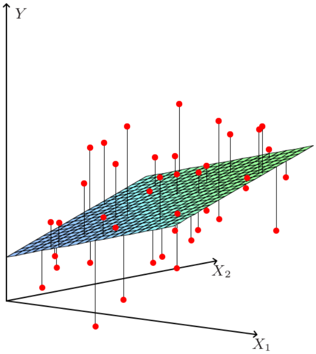
Peter Bartlett

Computer Science and Statistics
University of California at Berkeley

Mathematical Sciences
Queensland University of Technology

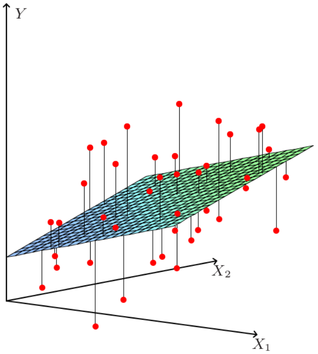
23 June, 2017
Dagstuhl

Online fixed design linear regression



(B., Koolen, Malek, Takimoto, Warmuth, 2015)

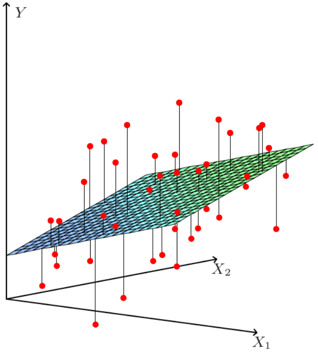
Online fixed design linear regression



Protocol

(B., Koolen, Malek, Takimoto, Warmuth, 2015)

Online fixed design linear regression

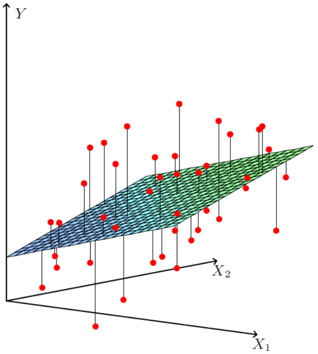


Protocol

Given: T;

(B., Koolen, Malek, Takimoto, Warmuth, 2015)

Online fixed design linear regression

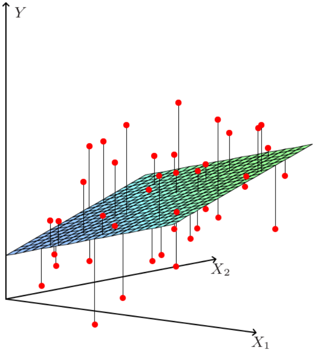


Protocol

Given: T; $x_1, \ldots, x_T \in \mathbb{R}^p$;

(B., Koolen, Malek, Takimoto, Warmuth, 2015)

Online fixed design linear regression

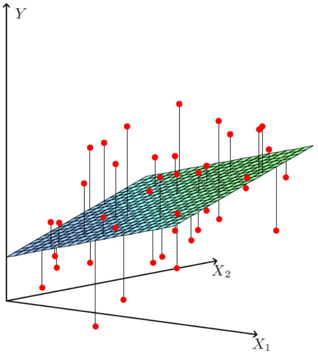


Protocol

Given: T; $x_1, \ldots, x_T \in \mathbb{R}^p$; $\mathcal{Y}^T \subset \mathbb{R}^T$.

(B., Koolen, Malek, Takimoto, Warmuth, 2015)

Online fixed design linear regression

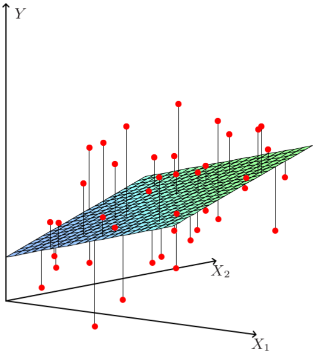


Protocol

Given: T; $x_1, \ldots, x_T \in \mathbb{R}^p$; $\mathcal{Y}^T \subset \mathbb{R}^T$.
For $t = 1, 2, \ldots, T$:

(B., Koolen, Malek, Takimoto, Warmuth, 2015)

Online fixed design linear regression



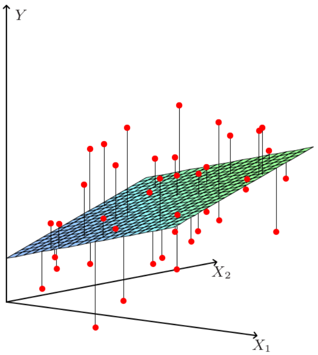
Protocol

Given: T; $x_1, \ldots, x_T \in \mathbb{R}^p$; $\mathcal{Y}^T \subset \mathbb{R}^T$.
For $t = 1, 2, \ldots, T$:

- Learner predicts $\hat{y}_t \in \mathbb{R}$

(B., Koolen, Malek, Takimoto, Warmuth, 2015)

Online fixed design linear regression



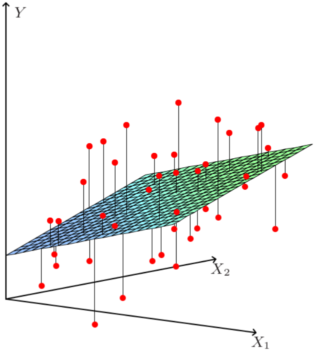
Protocol

Given: T; $x_1, \ldots, x_T \in \mathbb{R}^p$; $\mathcal{Y}^T \subset \mathbb{R}^T$.
For $t = 1, 2, \ldots, T$:

- Learner predicts $\hat{y}_t \in \mathbb{R}$
- Adversary reveals $y_t \in \mathbb{R}$

(B., Koolen, Malek, Takimoto, Warmuth, 2015)

Online fixed design linear regression



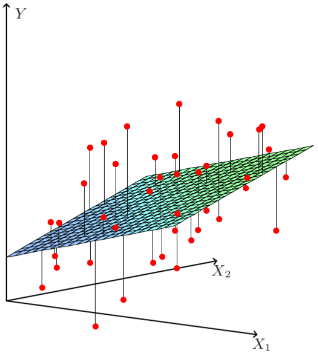
Protocol

Given: T; $x_1, \ldots, x_T \in \mathbb{R}^p$; $\mathcal{Y}^T \subset \mathbb{R}^T$.
For $t = 1, 2, \ldots, T$:

- Learner predicts $\hat{y}_t \in \mathbb{R}$
- Adversary reveals $y_t \in \mathbb{R}$ $\quad (y_1^T \in \mathcal{Y}^T)$

(B., Koolen, Malek, Takimoto, Warmuth, 2015)

Online fixed design linear regression



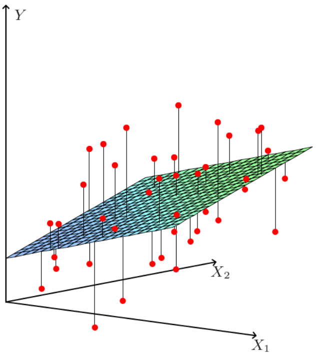
Protocol

Given: T; $x_1, \ldots, x_T \in \mathbb{R}^p$; $\mathcal{Y}^T \subset \mathbb{R}^T$.
For $t = 1, 2, \ldots, T$:

- Learner predicts $\hat{y}_t \in \mathbb{R}$
- Adversary reveals $y_t \in \mathbb{R}$ $\quad (y_1^T \in \mathcal{Y}^T)$
- Learner incurs loss $(\hat{y}_t - y_t)^2$.

(B., Koolen, Malek, Takimoto, Warmuth, 2015)

Online fixed design linear regression



Protocol

Given: T; $x_1, \dots, x_T \in \mathbb{R}^p$; $\mathcal{Y}^T \subset \mathbb{R}^T$.
For $t = 1, 2, \dots, T$:

- Learner predicts $\hat{y}_t \in \mathbb{R}$
- Adversary reveals $y_t \in \mathbb{R} \quad (y_1^T \in \mathcal{Y}^T)$
- Learner incurs loss $(\hat{y}_t - y_t)^2$.

$$\text{Regret} = \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2.$$

(B., Koolen, Malek, Takimoto, Warmuth, 2015)

Online fixed design linear regression

Online linear regression: previous work

- (Foster, 1991): ℓ_2-regularized least squares.
- (Cesa-Bianchi et al, 1996): ℓ_2-constrained least squares.
- (Kivinen and Warmuth, 1997): exponentiated gradient (relative entropy regularization).
- (Vovk, 1998): aggregating algorithm.
- (Forster, 1999; Azoury and Warmuth, 2001): aggregating algorithm is last-step minimax.

Online Linear Regression

Regret

$$\sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2$$

Online Linear Regression

Minimax Regret

$$\sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2$$

Online Linear Regression

Minimax Regret

$$\min_{\hat{y}_1} \qquad \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2$$

Online Linear Regression

Minimax Regret

$$\min_{\hat{y}_1} \max_{y_1 \in \mathcal{Y}} \qquad \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2$$

Online Linear Regression

Minimax Regret

$$\min_{\hat{y}_1} \max_{y_1 \in \mathcal{Y}} \cdots \min_{\hat{y}_T} \qquad \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t \right)^2$$

Online Linear Regression

Minimax Regret

$$\min_{\hat{y}_1} \max_{y_1 \in \mathcal{Y}} \cdots \min_{\hat{y}_T} \max_{y_T \in \mathcal{Y}} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2$$

Online Linear Regression

The value of the game: Minimax Regret

$$V_T(\mathcal{Y}) = \min_{\hat{y}_1} \max_{y_1 \in \mathcal{Y}} \cdots \min_{\hat{y}_T} \max_{y_T \in \mathcal{Y}} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2$$

Online Linear Regression

The value of the game: Minimax Regret

$$V_T(\mathcal{Y}) = \min_{\hat{y}_1} \max_{y_1 \in \mathcal{Y}} \cdots \min_{\hat{y}_T} \max_{y_T \in \mathcal{Y}} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2$$

Strategy:

$$S : \bigcup_{t=0}^{T} \mathcal{Y}^t \to \mathbb{R}.$$

Online Linear Regression

The value of the game: Minimax Regret

$$V_T(\mathcal{Y}) = \min_{\hat{y}_1} \max_{y_1 \in \mathcal{Y}} \cdots \min_{\hat{y}_T} \max_{y_T \in \mathcal{Y}} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2$$

Strategy:

$$S : \bigcup_{t=0}^{T} \mathcal{Y}^t \to \mathbb{R}.$$

$$V_T(\mathcal{Y}) = \min_{S} \max_{y_1^T \in \mathcal{Y}^T} \left(\sum_{t=1}^{T} \left(S\left(y_1^{t-1}\right) - y_t\right)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2 \right)$$

Online Linear Regression

The value of the game: Minimax Regret

$$V_T(\mathcal{Y}) = \min_{\hat{y}_1} \max_{y_1 \in \mathcal{Y}} \cdots \min_{\hat{y}_T} \max_{y_T \in \mathcal{Y}} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2$$

Minimax Optimal Strategy:

$$S^* : \bigcup_{t=0}^{T} \mathcal{Y}^t \to \mathbb{R}.$$

$$V_T(\mathcal{Y}) = \min_{S} \max_{y_1^T \in \mathcal{Y}^T} \left(\sum_{t=1}^{T} \left(S\left(y_1^{t-1}\right) - y_t\right)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2 \right)$$

$$= \max_{y_1^T \in \mathcal{Y}^T} \left(\sum_{t=1}^{T} \left(S^*\left(y_1^{t-1}\right) - y_t\right)^2 - \min_{\beta} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2 \right).$$

Online Prediction Games

Questions

Online Prediction Games

Questions

- Minimax regret?

Online Prediction Games

Questions

- Minimax regret?
- Optimal player strategy?

Online Prediction Games

Questions

- Minimax regret?
- Optimal player strategy?
- Efficient algorithms?

Online Prediction Games

Questions

- Minimax regret?
- Optimal player strategy?
- Efficient algorithms?
- Horizon-free?

Online Prediction Games

Questions

- Minimax regret?
- Optimal player strategy?
- Efficient algorithms?
- Horizon-free?
- Optimal adversary strategy?

Outline

- Fixed design.
- Minimax strategy is regularized least squares.
- Box and ellipsoid constraints.
- Adversarial covariates.

Linear regression in a probabilistic setting

Ordinary least squares (linear model, uncorrelated errors)

Given $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^p \times \mathbb{R}$,

Linear regression in a probabilistic setting

Ordinary least squares (linear model, uncorrelated errors)

Given $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^p \times \mathbb{R}$, choose

$$\hat{\beta} = \left(\sum_{t=1}^{n} x_t x_t^\top \right)^{-1} \sum_{t=1}^{n} x_t y_t,$$

Linear regression in a probabilistic setting

Ordinary least squares (linear model, uncorrelated errors)

Given $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^p \times \mathbb{R}$, choose

$$\hat{\beta} = \left(\sum_{t=1}^{n} x_t x_t^\top \right)^{-1} \sum_{t=1}^{n} x_t y_t,$$

and for a subsequent $x \in \mathbb{R}^p$, predict

$$\hat{y} = x^\top \hat{\beta} = x^\top \left(\sum_{t=1}^{n} x_t x_t^\top \right)^{-1} \sum_{t=1}^{n} x_t y_t,$$

Linear regression in a probabilistic setting

Ordinary least squares (linear model, uncorrelated errors)

Given $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^p \times \mathbb{R}$, choose

$$\hat{\beta} = \left(\sum_{t=1}^{n} x_t x_t^\top \right)^{-1} \sum_{t=1}^{n} x_t y_t,$$

and for a subsequent $x \in \mathbb{R}^p$, predict

$$\hat{y} = x^\top \hat{\beta} = x^\top \left(\sum_{t=1}^{n} x_t x_t^\top \right)^{-1} \sum_{t=1}^{n} x_t y_t,$$

A sequential version of OLS?

$$\hat{y}_{n+1} := x_{n+1}^\top \left(\sum_{t=1}^{n} x_t x_t^\top \right)^{-1} \sum_{t=1}^{n} x_t y_t.$$

Linear regression in a probabilistic setting

Ordinary least squares (linear model, uncorrelated errors)

Given $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^p \times \mathbb{R}$, choose

$$\hat{\beta} = \left(\sum_{t=1}^{n} x_t x_t^\top \right)^{-1} \sum_{t=1}^{n} x_t y_t,$$

and for a subsequent $x \in \mathbb{R}^p$, predict

$$\hat{y} = x^\top \hat{\beta} = x^\top \left(\sum_{t=1}^{n} x_t x_t^\top \right)^{-1} \sum_{t=1}^{n} x_t y_t,$$

A sequential version of ridge regression

$$\hat{y}_{n+1} := x_{n+1}^\top \left(\sum_{t=1}^{n} x_t x_t^\top + \lambda I \right)^{-1} \sum_{t=1}^{n} x_t y_t.$$

Online fixed design linear regression

Fix $x_1, \ldots, x_T \in \mathbb{R}^p$.

Online fixed design linear regression

Fix $x_1, \ldots, x_T \in \mathbb{R}^p$. $\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_t| \leq B_t\}$.

Online fixed design linear regression

Sufficient statistics

Fix $x_1, \ldots, x_T \in \mathbb{R}^p$. $\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_t| \leq B_t\}$.
Maintain statistics: $s_n = \sum_{t=1}^{n} y_t x_t$

Online fixed design linear regression

Sufficient statistics

Fix $x_1, \ldots, x_T \in \mathbb{R}^p$. $\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_t| \leq B_t\}$.

Maintain statistics: $s_n = \sum_{t=1}^n y_t x_t$

Minimax* strategy: linear

$$\hat{y}^*_{n+1} = x_{n+1}^\top P_{n+1} s_n.$$

Online fixed design linear regression

Sufficient statistics

Fix $x_1, \ldots, x_T \in \mathbb{R}^p$. $\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_t| \le B_t\}$.
Maintain statistics: $s_n = \sum_{t=1}^n y_t x_t$

Minimax* strategy: linear

$$\hat{y}^*_{n+1} = x_{n+1}^\top P_{n+1} s_n.$$

$$P_n^{-1} = \sum_{t=1}^n x_t x_t^\top +$$

Online fixed design linear regression

Sufficient statistics

Fix $x_1, \ldots, x_T \in \mathbb{R}^p$. $\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_t| \leq B_t\}$.

Maintain statistics: $s_n = \sum_{t=1}^{n} y_t x_t$

Minimax* strategy: linear

$$\hat{y}^*_{n+1} = x_{n+1}^\top P_{n+1} s_n.$$

$$P_n^{-1} = \sum_{t=1}^{n} x_t x_t^\top + \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

Online fixed design linear regression

Sufficient statistics

Fix $x_1, \ldots, x_T \in \mathbb{R}^p$.

$\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_t| \le B_t\}$.

Maintain statistics: $s_n = \sum_{t=1}^n y_t x_t$

Minimax* strategy: linear

$$\hat{y}^*_{n+1} = x_{n+1}^\top P_{n+1} s_n.$$

Maximin distribution:

$$\Pr(\pm B_{n+1}) = \frac{1}{2} \pm \frac{x_{n+1}^\top P_{n+1} s_n}{2 B_{n+1}}$$

$$P_n^{-1} = \sum_{t=1}^n x_t x_t^\top + \sum_{t=n+1}^T \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

Online fixed design linear regression

Sufficient statistics

Fix $x_1, \ldots, x_T \in \mathbb{R}^p$. $\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_t| \le B_t\}$.
Maintain statistics: $s_n = \sum_{t=1}^n y_t x_t$

Value-to-go: quadratic

$$s_n^\top P_n s_n - \sigma_n^2 + \sum_{t=n+1}^{T} B_t^2 x_t^\top P_t x_t.$$

Minimax* strategy: linear

$$\hat{y}_{n+1}^* = x_{n+1}^\top P_{n+1} s_n.$$

Maximin distribution:

$$\Pr\left(\pm B_{n+1}\right) = \frac{1}{2} \pm \frac{x_{n+1}^\top P_{n+1} s_n}{2 B_{n+1}}$$

$$P_n^{-1} = \sum_{t=1}^{n} x_t x_t^\top + \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

Online fixed design linear regression

Sufficient statistics

Fix $x_1, \dots, x_T \in \mathbb{R}^p$. $\mathcal{Y}^T = \{(y_1, \dots, y_T) : |y_t| \le B_t\}$.
Maintain statistics: $s_n = \sum_{t=1}^n y_t x_t$, $\sigma_n^2 = \sum_{t=1}^n y_t^2$.

Value-to-go: quadratic

$$s_n^\top P_n s_n - \sigma_n^2 + \sum_{t=n+1}^{T} B_t^2 x_t^\top P_t x_t.$$

Minimax* strategy: linear

$$\hat{y}_{n+1}^* = x_{n+1}^\top P_{n+1} s_n.$$

Maximin distribution:

$$\Pr\left(\pm B_{n+1}\right) = \frac{1}{2} \pm \frac{x_{n+1}^\top P_{n+1} s_n}{2 B_{n+1}}$$

$$P_n^{-1} = \sum_{t=1}^{n} x_t x_t^\top + \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

Online fixed design linear regression

Sufficient statistics

Fix $x_1, \dots, x_T \in \mathbb{R}^p$. $\mathcal{Y}^T = \{(y_1, \dots, y_T) : |y_t| \le B_t\}$.
Maintain statistics: $s_n = \sum_{t=1}^n y_t x_t$, $\sigma_n^2 = \sum_{t=1}^n y_t^2$.

Value-to-go: quadratic

$$s_n^\top P_n s_n - \sigma_n^2 + \sum_{t=n+1}^{T} B_t^2 x_t^\top P_t x_t.$$

$*$ provided: $$B_n \ge \sum_{t=1}^{n-1} \left| x_n^\top P_n x_t \right| B_t.$$

Minimax* strategy: linear

$$\hat{y}_{n+1}^* = x_{n+1}^\top P_{n+1} s_n.$$

Maximin distribution:

$$\Pr(\pm B_{n+1}) = \frac{1}{2} \pm \frac{x_{n+1}^\top P_{n+1} s_n}{2B_{n+1}}$$

$$P_n^{-1} = \sum_{t=1}^{n} x_t x_t^\top + \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

Linear regression

Box constraints

$$\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_n| \leq B_n\} \qquad B_n \geq \sum_{t=1}^{n-1} \left| x_n^\top P_n x_t \right| B_t.$$

Linear regression

Box constraints

$$\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_n| \le B_n\} \qquad B_n \ge \sum_{t=1}^{n-1} \left| x_n^\top P_n x_t \right| B_t.$$

Minimax strategy: linear

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}.$$

Linear regression

Box constraints

$$\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_n| \leq B_n\} \qquad B_n \geq \sum_{t=1}^{n-1} \left| x_n^\top P_n x_t \right| B_t.$$

Minimax strategy: linear

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}.$$

$$P_n^{-1} = \sum_{t=1}^{n} x_t x_t^\top + \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

Linear regression

Box constraints

$$\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_n| \leq B_n\} \qquad B_n \geq \sum_{t=1}^{n-1} \left| x_n^\top P_n x_t \right| B_t.$$

$$\text{Regret} = \sum_{t=1}^{T} B_t^2 x_t^\top P_t x_t.$$

Minimax strategy: linear

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}.$$

$$P_n^{-1} = \sum_{t=1}^{n} x_t x_t^\top + \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

Linear regression

Box constraints

$$\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_n| \leq B_n\} \qquad B_n \geq \sum_{t=1}^{n-1} \left| x_n^\top P_n x_t \right| B_t.$$

$$\text{Regret} = \sum_{t=1}^{T} B_t^2 x_t^\top P_t x_t.$$

Minimax strategy: linear

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}.$$

$$P_n^{-1} = \sum_{t=1}^{n} x_t x_t^\top + \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

c.f. ridge regression: $\displaystyle\sum_{t=1}^{n} x_t x_t^\top + \lambda I.$

Linear regression

Box constraints

$$\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_n| \leq B_n\} \qquad B_n \geq \sum_{t=1}^{n-1} \left| x_n^\top P_n x_t \right| B_t.$$

$$\text{Regret} = \sum_{t=1}^{T} B_t^2 x_t^\top P_t x_t.$$

Minimax strategy: linear

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}.$$

Optimal shrinkage

$$P_n^{-1} = \sum_{t=1}^{n} x_t x_t^\top + \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

c.f. ridge regression:

$$\sum_{t=1}^{n} x_t x_t^\top + \lambda I.$$

Linear regression

Box constraints

$$\mathcal{Y}^T = \{(y_1, \ldots, y_T) : |y_n| \le B_n\} \qquad B_n \ge \sum_{t=1}^{n-1} \left| x_n^\top P_n x_t \right| B_t.$$

$$\text{Regret} = \sum_{t=1}^{T} B_t^2 x_t^\top P_t x_t.$$

Minimax strategy: linear

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}.$$

Optimal shrinkage

$$P_n^{-1} = \sum_{t=1}^{n} x_t x_t^\top + \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

c.f. ridge regression:

$$\sum_{t=1}^{n} x_t x_t^\top + \lambda I.$$

Linear regression: Proof idea

Offline optimal:

$$V(s_T, \sigma_T^2, T) = -\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2$$

Linear regression: Proof idea

Offline optimal:

$$V(s_T, \sigma_T^2, T) = -\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2 \qquad \beta_T = P_T s_T,$$

Linear regression: Proof idea

Offline optimal:

$$V(s_T, \sigma_T^2, T) = -\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2 \qquad \beta_T = P_T s_T,$$

$$P_T = \left(\sum_{t=1}^{T} x_t x_t^\top\right)^{-1},$$

Linear regression: Proof idea

Offline optimal:

$$V(s_T, \sigma_T^2, T) = -\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2 \qquad \beta_T = P_T s_T,$$

$$P_T = \left(\sum_{t=1}^{T} x_t x_t^\top\right)^{-1},$$

$$s_T = \sum_{t=1}^{T} y_t x_t,$$

Linear regression: Proof idea

Offline optimal:

$$V(s_T, \sigma_T^2, T) = -\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2 \qquad \beta_T = P_T s_T,$$

$$= -\sigma_T^2 + \beta_T^\top s_T \qquad P_T = \left(\sum_{t=1}^{T} x_t x_t^\top\right)^{-1},$$

$$s_T = \sum_{t=1}^{T} y_t x_t,$$

Linear regression: Proof idea

Offline optimal:

$$V(s_T, \sigma_T^2, T) = -\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2 = -\sigma_T^2 + \beta_T^\top s_T$$

$$\beta_T = P_T s_T,$$

$$P_T = \left(\sum_{t=1}^{T} x_t x_t^\top\right)^{-1},$$

$$s_T = \sum_{t=1}^{T} y_t x_t,$$

$$\sigma_T^2 = \sum_{t=1}^{T} y_t^2.$$

Linear regression: Proof idea

Offline optimal:

$$
\begin{aligned}
V(s_T, \sigma_T^2, T) &= -\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2 \\
&= -\sigma_T^2 + \beta_T^\top s_T \\
&= -\sigma_T^2 + s_T^\top P_T s_T .
\end{aligned}
$$

$$
\begin{aligned}
\beta_T &= P_T s_T, \\
P_T &= \left(\sum_{t=1}^{T} x_t x_t^\top\right)^{-1}, \\
s_T &= \sum_{t=1}^{T} y_t x_t, \\
\sigma_T^2 &= \sum_{t=1}^{T} y_t^2 .
\end{aligned}
$$

Linear regression: Proof idea

Value to go

We show by induction that

$$V(s_t, \sigma_t^2, t) = s_t^\top P_t s_t - \sigma_t^2 + \gamma_t.$$

Linear regression: Proof idea

Value to go

We show by induction that

$$V(s_t, \sigma_t^2, t) = s_t^\top P_t s_t - \sigma_t^2 + \gamma_t.$$

It's true for $T = t$ with $\gamma_T = 0$.

Linear regression: Proof idea

Value to go

We show by induction that

$$V(s_t, \sigma_t^2, t) = s_t^\top P_t s_t - \sigma_t^2 + \gamma_t.$$

It's true for $T = t$ with $\gamma_T = 0$.
Provided the problem is not too constrained, i.e., $B \geq |x_{t+1}^\top P_{t+1} s_t|$ (otherwise, the player should clip $\hat{y}_{t+1}$ to B or $-B$),

$$\begin{aligned} V(s_t, \sigma_t^2, t) = s_t^\top &\left(P_{t+1} x_{t+1} x_{t+1}^\top P_{t+1} + P_{t+1} \right) s_t \\ &- \sigma_t^2 + \gamma_{t+1} + B^2 x_{t+1}^\top P_{t+1} x_{t+1}. \end{aligned}$$

Linear regression: Proof idea

Optimal predictions

$$\hat{y}_{n+1} = x_{n+1}^\top P_{n+1} s_n, \qquad s_n = \sum_{t=1}^{n} y_t x_t,$$

$$P_T = \left(\sum_{t=1}^{T} x_t x_t^\top \right)^{-1}, \qquad P_n = P_{n+1} x_{n+1} x_{n+1}^\top P_{n+1} + P_{n+1}.$$

Linear regression: Proof idea

An alternative recurrence

$$P_n^{-1} = \sum_{t=1}^{n} x_t x_t^\top + \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top .$$

Proof: Sherman-Morrison.

Linear regression: Regret

$$\text{Regret} = \sum_{t=1}^{T} B_t^2 x_t^\top P_t x_t.$$

Theorem

$$\max_{x_1,\dots x_T} \sum_{t=1}^{T} x_t^\top P_t x_t \le p\left(1 + 2\ln\left(1 + \frac{T}{2}\right)\right).$$

Outline

- Fixed design.
- Minimax strategy is regularized least squares.
- **Box and ellipsoid constraints.**
- Adversarial covariates.

Linear regression: Alternative constraints

Ellipsoid constraints (weighted 2-norm)

$$\mathcal{Y}_R^T = \left\{ (y_1, \ldots, y_T) : \sum_{t=1}^{T} y_t^2 x_t^\top P_t x_t \leq R \right\}.$$

Linear regression: Alternative constraints

Ellipsoid constraints (weighted 2-norm)

$$\mathcal{Y}_R^T = \left\{ (y_1, \ldots, y_T) : \sum_{t=1}^{T} y_t^2 x_t^\top P_t x_t \leq R \right\}.$$

Minimax strategy: linear

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}.$$

Linear regression: Alternative constraints

Ellipsoid constraints (weighted 2-norm)

$$\mathcal{Y}_R^T = \left\{ (y_1, \ldots, y_T) : \sum_{t=1}^{T} y_t^2 x_t^\top P_t x_t \leq R \right\}.$$

Minimax regret $= R$.

Minimax strategy: linear

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}.$$

Linear regression: Alternative constraints

Ellipsoid constraints (weighted 2-norm)

$$\mathcal{Y}_R^T = \left\{ (y_1, \ldots, y_T) : \sum_{t=1}^{T} y_t^2 x_t^\top P_t x_t \leq R \right\}.$$

Minimax regret $= R$.

Minimax strategy: linear

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}. \qquad \text{(MM)}$$

Equalizer property

For all $y_1, \ldots, y_T$,

$$\text{Regret of (MM)} := \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t \right)^2$$

Linear regression: Alternative constraints

Ellipsoid constraints (weighted 2-norm)

$$\mathcal{Y}_R^T = \left\{ (y_1, \ldots, y_T) : \sum_{t=1}^{T} y_t^2 x_t^\top P_t x_t \leq R \right\}.$$

Minimax regret $= R$.

Minimax strategy: linear

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}. \qquad \text{(MM)}$$

Equalizer property

For all $y_1, \ldots, y_T$,

$$\text{Regret of (MM)} := \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t \right)^2$$
$$= \sum_{t=1}^{T} y_t^2 x_t^\top P_t x_t.$$

Linear regression: Alternative constraints

Equalizer property

For all $y_1, \ldots, y_T$,

$$\text{Regret of } (\text{MM}) := \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t \right)^2$$
$$= \sum_{t=1}^{T} y_t^2 x_t^\top P_t x_t.$$

(Proof is by induction.)

Linear regression: Alternative constraints

Equalizer property

For all $y_1, \ldots, y_T$,

$$\text{Regret of } (\text{MM}) := \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t \right)^2$$
$$= \sum_{t=1}^{T} y_t^2 x_t^\top P_t x_t.$$

(Proof is by induction.)

Corollary

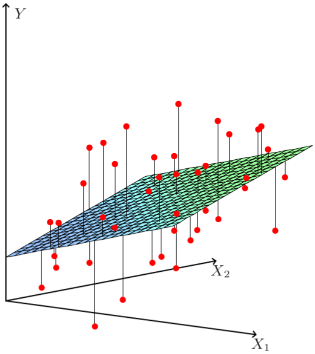
For every R, (MM) is minimax optimal on

$$\mathcal{Y}_R^T = \left\{ (y_1, \ldots, y_T) : \sum_{t=1}^{T} y_t^2 x_t^\top P_t x_t \leq R \right\}.$$

Outline

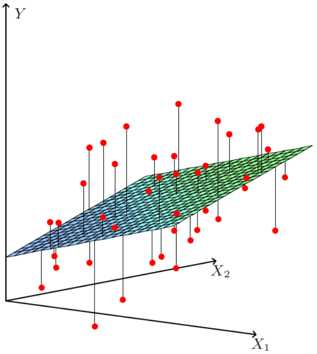
- Fixed design.
- Minimax strategy is regularized least squares.
- Box and ellipsoid constraints.
- **Adversarial covariates.**

Linear regression: Adversarial covariates



(B., Malek, 2017)

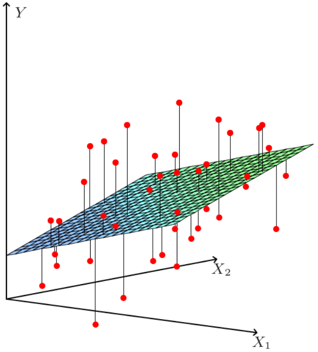
Linear regression: Adversarial covariates



Protocol

(B., Malek, 2017)

Linear regression: Adversarial covariates

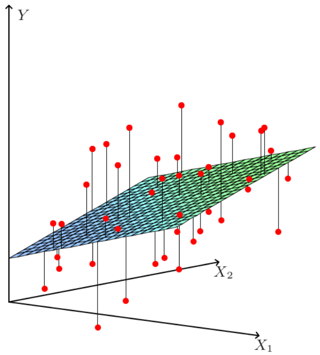


Protocol

Given: T;

(B., Malek, 2017)

Linear regression: Adversarial covariates

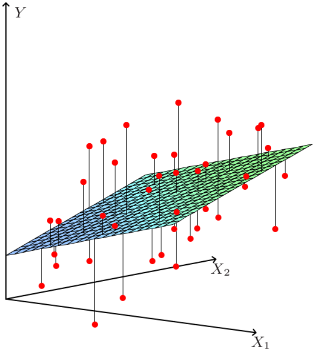


Protocol

Given: T; $\mathcal{X}^T \subset (\mathbb{R}^p)^T$;

(B., Malek, 2017)

Linear regression: Adversarial covariates

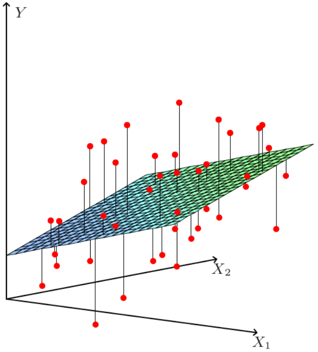


Protocol

Given: T; $\mathcal{X}^T \subset (\mathbb{R}^p)^T$; $\mathcal{Y}^T \subset \mathbb{R}^T$.

(B., Malek, 2017)

Linear regression: Adversarial covariates

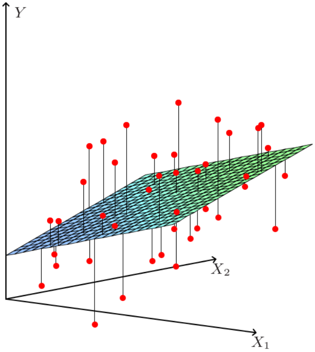


Protocol

Given: T; $\mathcal{X}^T \subset (\mathbb{R}^p)^T$; $\mathcal{Y}^T \subset \mathbb{R}^T$.
For $t = 1, 2, \ldots, T$:

(B., Malek, 2017)

Linear regression: Adversarial covariates



Protocol

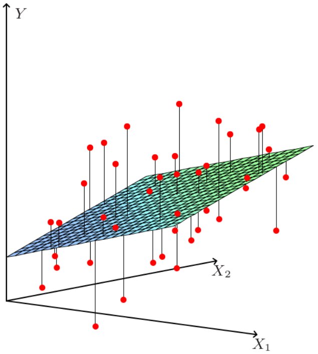
Given: T; $\mathcal{X}^T \subset (\mathbb{R}^p)^T$; $\mathcal{Y}^T \subset \mathbb{R}^T$.

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

- Adversary reveals $x_t \in \mathbb{R}^p$

(B., Malek, 2017)

Linear regression: Adversarial covariates



Protocol

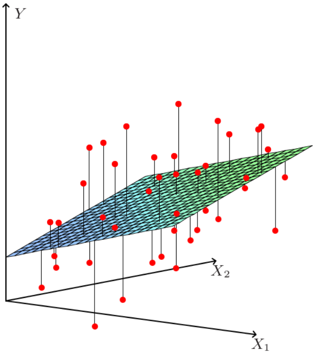
Given: T; $\mathcal{X}^T \subset (\mathbb{R}^p)^T$; $\mathcal{Y}^T \subset \mathbb{R}^T$.

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

- Adversary reveals $x_t \in \mathbb{R}^p$ $(x_1^T \in \mathcal{X}^T)$

(B., Malek, 2017)

Linear regression: Adversarial covariates



Protocol

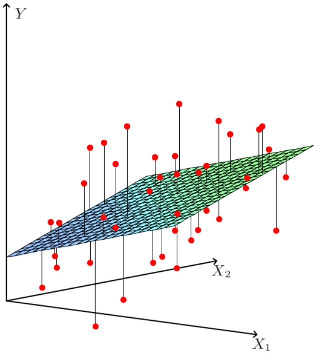
Given: T; $\mathcal{X}^T \subset (\mathbb{R}^p)^T$; $\mathcal{Y}^T \subset \mathbb{R}^T$.

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

- Adversary reveals $x_t \in \mathbb{R}^p$ ($x_1^T \in \mathcal{X}^T$)
- Learner predicts $\hat{y}_t \in \mathbb{R}$

(B., Malek, 2017)

Linear regression: Adversarial covariates



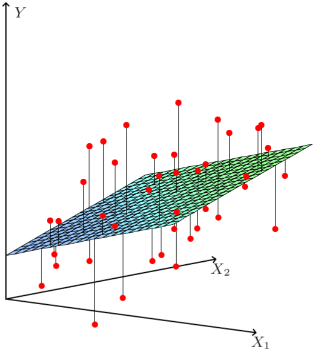
Protocol

Given: T; $\mathcal{X}^T \subset (\mathbb{R}^p)^T$; $\mathcal{Y}^T \subset \mathbb{R}^T$.
For $t = 1, 2, \ldots, T$:

- Adversary reveals $x_t \in \mathbb{R}^p$ $(x_1^T \in \mathcal{X}^T)$
- Learner predicts $\hat{y}_t \in \mathbb{R}$
- Adversary reveals $y_t \in \mathbb{R}$

(B., Malek, 2017)

Linear regression: Adversarial covariates



Protocol

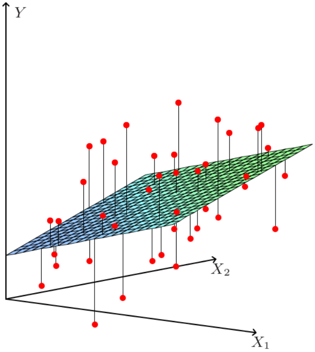
Given: T; $\mathcal{X}^T \subset (\mathbb{R}^p)^T$; $\mathcal{Y}^T \subset \mathbb{R}^T$.

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

- Adversary reveals $x_t \in \mathbb{R}^p$ $(x_1^T \in \mathcal{X}^T)$
- Learner predicts $\hat{y}_t \in \mathbb{R}$
- Adversary reveals $y_t \in \mathbb{R}$ $(y_1^T \in \mathcal{Y}^T)$

(B., Malek, 2017)

Linear regression: Adversarial covariates



Protocol

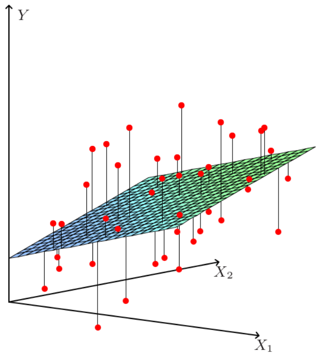
Given: T; $\mathcal{X}^T \subset (\mathbb{R}^p)^T$; $\mathcal{Y}^T \subset \mathbb{R}^T$.

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

- Adversary reveals $x_t \in \mathbb{R}^p$ $(x_1^T \in \mathcal{X}^T)$
- Learner predicts $\hat{y}_t \in \mathbb{R}$
- Adversary reveals $y_t \in \mathbb{R}$ $(y_1^T \in \mathcal{Y}^T)$
- Learner incurs loss $(\hat{y}_t - y_t)^2$.

(B., Malek, 2017)

Linear regression: Adversarial covariates



Protocol

Given: T; $\mathcal{X}^T \subset (\mathbb{R}^p)^T$; $\mathcal{Y}^T \subset \mathbb{R}^T$.
For $t = 1, 2, \ldots, T$:

- Adversary reveals $x_t \in \mathbb{R}^p$ $(x_1^T \in \mathcal{X}^T)$
- Learner predicts $\hat{y}_t \in \mathbb{R}$
- Adversary reveals $y_t \in \mathbb{R}$ $(y_1^T \in \mathcal{Y}^T)$
- Learner incurs loss $(\hat{y}_t - y_t)^2$.

$$\text{Regret} = \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t\right)^2.$$

(B., Malek, 2017)

Linear regression: Adversarial covariates

A covariance budget

Recall:

$$P_T^{-1} = \sum_{t=1}^{T} x_t x_t^\top,$$

$$P_n = P_{n+1} x_{n+1} x_{n+1}^\top P_{n+1} + P_{n+1}.$$

Linear regression: Adversarial covariates

A covariance budget

Recall:

$$P_T^{-1} = \sum_{t=1}^{T} x_t x_t^\top,$$

$$P_n = P_{n+1} x_{n+1} x_{n+1}^\top P_{n+1} + P_{n+1}.$$

Equivalently,

$$P_n^{-1} = \sum_{t=1}^{n} x_t x_t^\top + \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

Linear regression: Adversarial covariates

A covariance budget

Recall:

$$P_T^{-1} = \sum_{t=1}^{T} x_t x_t^\top,$$

$$P_n = P_{n+1} x_{n+1} x_{n+1}^\top P_{n+1} + P_{n+1}.$$

Equivalently,

$$P_n^{-1} = \sum_{t=1}^{n} x_t x_t^\top + \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

Define

$$P_0^{-1} = \sum_{q=1}^{T} \frac{x_q^\top P_q x_q}{1 + x_q^\top P_q x_q} x_q x_q^\top \succeq 0.$$

Linear regression

A reformulation

$$P_0^{-1} = \sum_{q=1}^{T} \frac{x_q^\top P_q x_q}{1 + x_q^\top P_q x_q} x_q x_q^\top \succeq 0.$$

Linear regression

A reformulation

$$P_0^{-1} = \sum_{q=1}^{T} \frac{x_q^\top P_q x_q}{1 + x_q^\top P_q x_q} x_q x_q^\top \succeq 0.$$

Theorem

$$P_{t+1} = P_t - \frac{a_t}{b_t^2} P_t x_{t+1} x_{t+1}^\top P_t,$$

where
$$a_t = \frac{\sqrt{4b_t^2 + 1} - 1}{\sqrt{4b_t^2 + 1} + 1},$$
$$b_t^2 = x_{t+1}^\top P_t x_{t+1}.$$

Linear regression

Legal covariate sequences

For any $t \geq 0$, any $x_1, \ldots, x_t$ and any P_t, the following two conditions are equivalent.

Legal covariate sequences

For any $t \geq 0$, any $x_1, \ldots, x_t$ and any P_t, the following two conditions are equivalent.

1. There is a $T \geq t$ and a sequence $x_{t+1}, \ldots, x_T$ such that, under the forward iteration,
$$P_T^{-1} = \sum_{q=1}^{T} x_q x_q^\top .$$

Linear regression

Legal covariate sequences

For any $t \geq 0$, any $x_1, \ldots, x_t$ and any P_t, the following two conditions are equivalent.

1. There is a $T \geq t$ and a sequence $x_{t+1}, \ldots, x_T$ such that, under the forward iteration,
$$P_T^{-1} = \sum_{q=1}^{T} x_q x_q^\top .$$
2. $P_t^{-1} \succeq \sum_{q=1}^{t} x_q x_q^\top$.

Linear regression

Legal covariate sequences

For any $t \geq 0$, any $x_1, \ldots, x_t$ and any P_t, the following two conditions are equivalent.

1. There is a $T \geq t$ and a sequence $x_{t+1}, \ldots, x_T$ such that, under the forward iteration,
$$P_T^{-1} = \sum_{q=1}^{T} x_q x_q^\top .$$
2. $P_t^{-1} \succeq \sum_{q=1}^{t} x_q x_q^\top$.

Adversarial covariates

Thus, each $P_0 \succeq 0$ (a 'covariance budget') defines a set of sequences $x_1, \ldots, x_T$.

Linear regression

Legal covariate sequences

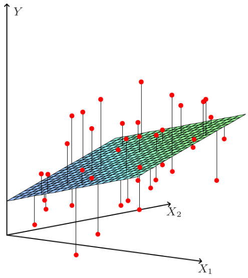
For any $t \geq 0$, any $x_1, \ldots, x_t$ and any P_t, the following two conditions are equivalent.

1. There is a $T \geq t$ and a sequence $x_{t+1}, \ldots, x_T$ such that, under the forward iteration,
$$P_T^{-1} = \sum_{q=1}^{T} x_q x_q^\top.$$
2. $P_t^{-1} \succeq \sum_{q=1}^{t} x_q x_q^\top$.

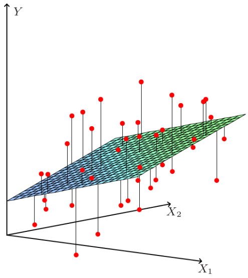
Adversarial covariates

Thus, each $P_0 \succeq 0$ (a 'covariance budget') defines a set of sequences $x_1, \ldots, x_T$.
The same strategy is optimal for each of these sequences.

Linear regression: Adversarial covariates; horizon-free

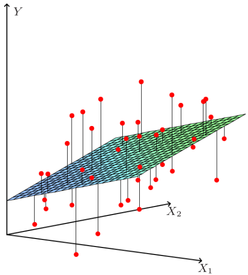


Linear regression: Adversarial covariates; horizon-free



Protocol

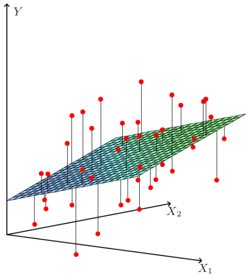
Linear regression: Adversarial covariates; horizon-free



Protocol

Given:

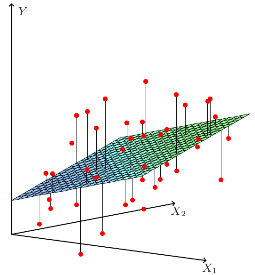
Linear regression: Adversarial covariates; horizon-free



Protocol

Given: $\mathcal{Z} \subset \bigcup_{T \geq 1} (\mathbb{R}^p \times \mathbb{R})^T$.

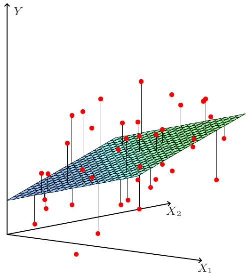
Linear regression: Adversarial covariates; horizon-free



Protocol

Given: $\mathcal{Z} \subset \bigcup_{T \geq 1} (\mathbb{R}^p \times \mathbb{R})^T$.
For $t = 1, 2, 3, \ldots$

Linear regression: Adversarial covariates; horizon-free

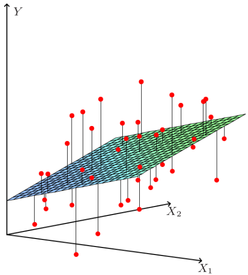


Protocol

Given: $\mathcal{Z} \subset \bigcup_{T \geq 1} (\mathbb{R}^p \times \mathbb{R})^T$.
For $t = 1, 2, 3, \ldots$

- Adversary reveals $x_t \in \mathbb{R}^p$

Linear regression: Adversarial covariates; horizon-free

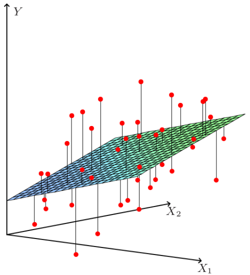


Protocol

Given: $\mathcal{Z} \subset \bigcup_{T \geq 1} (\mathbb{R}^p \times \mathbb{R})^T$.
For $t = 1, 2, 3, \ldots$

- Adversary reveals $x_t \in \mathbb{R}^p$
- Learner predicts $\hat{y}_t \in \mathbb{R}$

Linear regression: Adversarial covariates; horizon-free

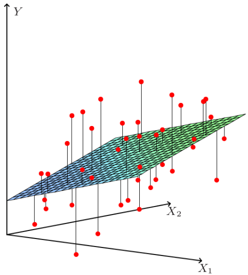


Protocol

Given: $\mathcal{Z} \subset \bigcup_{T \geq 1} (\mathbb{R}^p \times \mathbb{R})^T$.
For $t = 1, 2, 3, \ldots$

- Adversary reveals $x_t \in \mathbb{R}^p$
- Learner predicts $\hat{y}_t \in \mathbb{R}$
- Adversary reveals $y_t \in \mathbb{R}$

Linear regression: Adversarial covariates; horizon-free

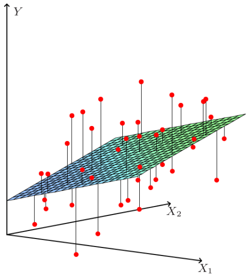


Protocol

Given: $\mathcal{Z} \subset \bigcup_{T \geq 1} (\mathbb{R}^p \times \mathbb{R})^T$.
For $t = 1, 2, 3, \ldots$

- Adversary reveals $x_t \in \mathbb{R}^p$
- Learner predicts $\hat{y}_t \in \mathbb{R}$
- Adversary reveals $y_t \in \mathbb{R}$ $\qquad (x_1^T, y_1^T \in \mathcal{Z})$

Linear regression: Adversarial covariates; horizon-free

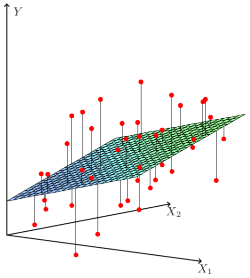


Protocol

Given: $\mathcal{Z} \subset \bigcup_{T \geq 1} (\mathbb{R}^p \times \mathbb{R})^T$.
For $t = 1, 2, 3, \ldots$

- Adversary reveals $x_t \in \mathbb{R}^p$
- Learner predicts $\hat{y}_t \in \mathbb{R}$
- Adversary reveals $y_t \in \mathbb{R}$ $\qquad (x_1^T, y_1^T \in \mathcal{Z})$
- Learner incurs loss $(\hat{y}_t - y_t)^2$.

Linear regression: Adversarial covariates; horizon-free



Protocol

Given: $\mathcal{Z} \subset \bigcup_{T \geq 1} (\mathbb{R}^p \times \mathbb{R})^T$.

For $t = 1, 2, 3, \ldots$

- Adversary reveals $x_t \in \mathbb{R}^p$
- Learner predicts $\hat{y}_t \in \mathbb{R}$
- Adversary reveals $y_t \in \mathbb{R}$ $\quad (x_1^T, y_1^T \in \mathcal{Z})$
- Learner incurs loss $(\hat{y}_t - y_t)^2$.

$$\text{Regret} = \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - \min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(\beta^\top x_t - y_t \right)^2.$$

Linear regression

Constraints on y_ts

1. *Box constraints*: $\mathcal{B}(B) := \{y_1^T : |y_t| \leq B_t\}$, for $B_1, B_2, \ldots > 0$.

Linear regression

Constraints on y_ts

1. *Box constraints*: $\mathcal{B}(B) := \{y_1^T : |y_t| \le B_t\}$, for $B_1, B_2, \ldots > 0$.
2. *Ellipsoidal constraints*:

$$\mathcal{E}(x_1^T, R) := \left\{ y_1^T : \sum_{t=1}^{T} y_t^2 x_t^\top P_t x_t \le R \right\}.$$

Linear regression

Constraints on x_ts

1. *Compatibility constraints*:

$$\mathcal{X}(B) = \left\{ x_1^T : B_t \geq \sum_{s=1}^{t-1} \left| x_t^\top P_t x_s \right| B_s \text{ for } 2 \leq t \leq T \right\}.$$

Linear regression

Constraints on x_ts

1. *Compatibility constraints*:

$$\mathcal{X}(B) = \left\{ x_1^T : B_t \geq \sum_{s=1}^{t-1} \left| x_t^\top P_t x_s \right| B_s \text{ for } 2 \leq t \leq T \right\}.$$

2. *Covariance constraints*:

$$\overline{\mathcal{X}}(\Sigma) = \left\{ x_1^T : \text{for } P_0, \ldots, P_T \text{ defined by } x_1^T,\ P_0^{-1} = \Sigma \right\}.$$

Linear regression

Theorem

For all positive semidefinite Σ; $B_1, B_2, \ldots > 0$
the forward strategy s^* is horizon-independent minimax optimal,

Theorem

For all positive semidefinite Σ; $B_1, B_2, \ldots > 0$
the forward strategy s^* is horizon-independent minimax optimal,

$$\sup_T \sup_{x_1^T \in \mathcal{X}} \left(\sup_{y_1^T \in \mathcal{Y}(x_1^T)} R_T(s^*, x_1^T, y_1^T) - \min_s \sup_{y_1^T \in \mathcal{Y}(x_1^T)} R_T(s, x_1^T, y_1^T) \right) = 0.$$

with respect to the following $(\mathcal{X}, \mathcal{Y}(x_1^t))$:

$$(\mathcal{X}(B_1^T) \cap \overline{\mathcal{X}}(\Sigma), \mathcal{B}(B_1^T)),$$

Linear regression

Theorem

For all positive semidefinite Σ; $B_1, B_2, \ldots > 0$; and $R > 0$, the forward strategy s^* is horizon-independent minimax optimal,

$$\sup_T \sup_{x_1^T \in \mathcal{X}} \left(\sup_{y_1^T \in \mathcal{Y}(x_1^T)} R_T(s^*, x_1^T, y_1^T) - \min_s \sup_{y_1^T \in \mathcal{Y}(x_1^T)} R_T(s, x_1^T, y_1^T) \right) = 0.$$

with respect to the following $(\mathcal{X}, \mathcal{Y}(x_1^t))$:

$$(\mathcal{X}(B_1^T) \cap \overline{\mathcal{X}}(\Sigma), \mathcal{B}(B_1^T)), \qquad (\overline{\mathcal{X}}(\Sigma), \mathcal{E}(x_1^T, R)).$$

Linear regression

Theorem

For all positive semidefinite Σ; $B_1, B_2, \ldots > 0$; and $R > 0$, the forward strategy s^* is horizon-independent minimax optimal,

$$\sup_T \sup_{x_1^T \in \mathcal{X}} \left(\sup_{y_1^T \in \mathcal{Y}(x_1^T)} R_T(s^*, x_1^T, y_1^T) - \min_s \sup_{y_1^T \in \mathcal{Y}(x_1^T)} R_T(s, x_1^T, y_1^T) \right) = 0.$$

with respect to the following $(\mathcal{X}, \mathcal{Y}(x_1^t))$:

$$(\mathcal{X}(B_1^T) \cap \overline{\mathcal{X}}(\Sigma), \mathcal{B}(B_1^T)), \qquad (\overline{\mathcal{X}}(\Sigma), \mathcal{E}(x_1^T, R)).$$

That is, s^* performs as well as the best strategy that sees the covariate sequence.

The minimax strategy as regularized least squares

The minimax strategy predicts $\hat{y}_n = \hat{\theta}_n^\top x_n$, where $\hat{\theta}_n$ is the solution to

$$\min_\theta \sum_{t=1}^{n-1} (\theta^\top x_t - y_t)^2 + \theta^\top R_n \theta,$$

The minimax strategy as regularized least squares

The minimax strategy predicts $\hat{y}_n = \hat{\theta}_n^\top x_n$, where $\hat{\theta}_n$ is the solution to

$$\min_\theta \sum_{t=1}^{n-1} (\theta^\top x_t - y_t)^2 + \theta^\top R_n \theta,$$

$$R_n := \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top .$$

The minimax strategy as regularized least squares

The minimax strategy predicts $\hat{y}_n = \hat{\theta}_n^\top x_n$, where $\hat{\theta}_n$ is the solution to

$$\min_\theta \sum_{t=1}^{n-1} (\theta^\top x_t - y_t)^2 + \theta^\top R_n \theta,$$

$$R_n := \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

Indeed,

$$\hat{\theta}_n = \left(\sum_{t=1}^{n} x_t x_t^\top + R_n \right)^{-1} s_{n-1}$$

The minimax strategy as regularized least squares

The minimax strategy predicts $\hat{y}_n = \hat{\theta}_n^\top x_n$, where $\hat{\theta}_n$ is the solution to

$$\min_\theta \sum_{t=1}^{n-1} (\theta^\top x_t - y_t)^2 + \theta^\top R_n \theta,$$

$$R_n := \sum_{t=n+1}^{T} \frac{x_t^\top P_t x_t}{1 + x_t^\top P_t x_t} x_t x_t^\top.$$

Indeed,

$$\hat{\theta}_n = \left(\sum_{t=1}^{n} x_t x_t^\top + R_n \right)^{-1} s_{n-1} = P_n^{-1} s_{n-1}.$$

Linear regression

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}$$

- Minimax optimal for two families of label constraints: box constraints and problem-weighted ℓ_2 norm constraints.

Linear regression

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}$$

- Minimax optimal for two families of label constraints:
  box constraints and problem-weighted ℓ_2 norm constraints.
- Strategy does not need to know the constraints.

Linear regression

$$\hat{y}_n^* = x_n^\top P_n s_{n-1}$$

- Minimax optimal for two families of label constraints: box constraints and problem-weighted ℓ_2 norm constraints.
- Strategy does not need to know the constraints.
- Same strategy is optimal for covariate sequences consistent with some 'covariance budget' P_0.