Bias-Variance Games

Yiding Feng (MSR) Ronen Gradwohl (Ariel) Jason Hartline (NWU) Aleck Johnsen (NWU) Denis Nekipelov (UVA)

Motivation

- In many applications, prediction algorithms have been used in competitive environments
  - recommendations, targeted advertising, pricing
  - Zillow vs. Trulia vs. Redfin
- What is effect of competitive on choice of prediction algorithms
  - are prediction algorithms optimal in isolation still optimal in competitive environment?
  - if not, what algorithms are better/best under competition?

Contributions and Takeaways

Bias-Variance Games: A game-theoretic model focusing on *"bias-variance tradeoff"*

- prediction algorithms optimal in isolation may no longer optimal under competition
- bias is more harmful than variance

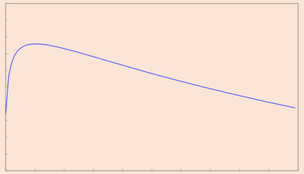
Running Example: Ridge Regression on Price Prediction

- Given *California housing prices data*, train ridge regression to predict housing prices

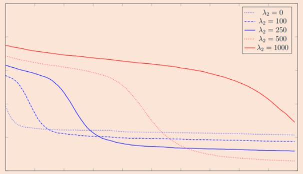
$$\text{argmin}_\beta \sum(\beta \cdot x_i - y_i)^2 - \lambda||\beta||_2$$

where λ controls bias-variance tradeoff, selects by planner

- higher $\lambda \Rightarrow$ higher bias, lower variance



one player: $\lambda \approx 100$ is optimal



two players: $\lambda = 0$ is dominant strategy

General Framework of Statistical Decision Theory

- Prior distribution π over pairs $(x, f(x))$
  - x: feature vector
  - $f(x) = \omega^T x + \varepsilon$: label
- Learning algorithm $\mathcal{A}$ takes training data $D = \{(x^i, y^i)\}$ as input, and produces estimator $\hat{f}_D$ as output
  - Loss of $\mathcal{A}$ on vector x: $L(\mathcal{A}, x) = \left(\hat{f}_D(x) - f(x)\right)^2$
  - Risk of $\mathcal{A}$ on vector x: $R(\mathcal{A}, x) = \mathbb{E}_D[L(\mathcal{A}, x)]$

$$\equiv \left(\mathbb{E}_D[\hat{f}_D(x)] - f(x)\right)^2 + \text{Var}[\hat{f}_D(x)] + \text{Var}[f(x)]$$

-bias- -variance-

- $\mathcal{A}$ is empirical risk minimization if

$$\mathcal{A} \text{ solves } \hat{f}_D = \text{argmin}_{\hat{f}} \frac{1}{|D|} \sum\left(\hat{f}(x^i) - y^i\right)^2 + \lambda \mathcal{J}(\hat{f})$$

  - $\mathcal{J}$: penalty function
  - λ: control bias-variance tradeoff and then minimize $\mathbb{E}_x[R(\mathcal{A}, x)]$
- 3-stage timeline: (higher $\lambda \Rightarrow$ higher bias lower variance)
  - ex ante stage: $\hat{f}_D$ is trained by picking $\mathcal{J}, \lambda$
  - interim stage: new data point $(x, f(x))$ are realized

    only feature x is observed
  - ex post stage: estimate $\hat{f}_D(x)$ is produced

Bias-Variance Game

- **Research Question**: Fixing $\mathcal{J}$, how should planners decide λ under competition (comparing to in isolation)?
- Two players (i.e., planners) $i \in \{1,2\}$
  - each player i design learning algorithm $\mathcal{A}_i$ by deciding her λ_i
  - for each new data point $(x, f(x))$, player i gains payoff

$$\left(1 - L(\mathcal{A}_i, x)\right) \cdot \mathbb{I}\{L(\mathcal{A}_i, x) \leq L(\mathcal{A}_{-i}, x)\}$$

other payoff functions are also considered in this paper

Main Result

The Interim Game

- Suppose $(x, f(x))$ is fixed.
- Observation: if bias^2 + variance is constant for all λ, in isolation (i.e., only a single player), the player is indifferent among all λ

Theorem.

Suppose for all λ, (i) bias^2 + variance is 1; and (ii) error $\hat{f}_D(x) - f(x)$ is normally distributed. Then players' utilities are strictly decreasing in λ; and thus $\lambda = 0$ is the *dominant strategy*.

result is qualitatively robust – various robustness checks for other error distributions, other payoff functions, non-constant tradeoff

- Additional results on reducing bias/variance while holding the other fixed
  - lowering variance may be harmful
  - lowering bias may be harmful
  - lowering bias is beneficial under "natural error distributions"

The Ex Ante Game

- Suppose $(x, f(x))$ is sampled from π
- Let λ^* be the optimal parameter minimizing $\mathbb{E}_x[R(\mathcal{A}, x)]$

Theorem.

Under assumptions A1-A5, suppose the other player j uses λ^*, player i strictly prefers $\lambda^* - \epsilon$ than λ^*, i.e., the derivative of utility is negative at λ^*.

(*) see the formal definition of assumptions A1-A5 in Section 5 of paper

(**) assumptions are numerically verified in empirical experiment

arXiv preprint: 1909.03618