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 the effect of competition on choice of prediction algorithms?
- prediction algorithms optimal in isolation still optimal in competitive environments?
- 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
argmin_β \sum_i (\beta \cdot x_i - y_i)^2 - \lambda \|\beta\|_2
where \lambda controls bias-variance tradeoff, selects by planner
- higher \lambda \Rightarrow higher bias, lower variance

General Framework of Statistical Decision Theory
- Prior distribution \pi over pairs (x, f(x))
  - x: feature vector
  - \pi(x, f(x)) = \omega x + \varepsilon; label
- Learning algorithm \mathcal{A} takes training data \mathcal{D} = \{(x^i, y^i)\} as input, and produces estimator \hat{f}_D as output
  - Loss of \mathcal{A} on vector x: \mathcal{L}(\mathcal{A}, x) = (\hat{f}(x) - f(x))^2
  - Risk of \mathcal{A} on vector x: R(\mathcal{A}, x) = \mathbb{E}_D[\mathcal{L}(\mathcal{A}, x)]
    \equiv \mathbb{E}_D[(\hat{f}(x) - f(x))^2 + \text{Var}[f(x)]]
  - \mathcal{A} is empirical risk minimization if \mathcal{A} solves \hat{f}_D = \arg\min_\theta \mathbb{E}_D[\mathcal{L}(\mathcal{A}, x)]
    \cdot \text{bias}
    \cdot \text{variance}

  - 3-stage timeline:
    - ex ante stage: \hat{f}_D is trained by picking \lambda
    - interim stage: new data point (x, f(x)) are realized
    - only feature x is observed
    - ex post stage: estimate \hat{f}(x) is produced

Main Result

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

Theorem.
Suppose for all \lambda, (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 \lambda; 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 \pi
- Let \lambda^* be the optimal parameter minimizing \mathbb{E}_x[\mathcal{L}(\mathcal{A}, x)]

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

(*) see the formal definition of assumptions A1-A5 in Section 5 of paper
(**) assumptions are numerically verified in empirical experiment

Bias-Variance Game
- Research Question: Fixing J, how should planners decide \lambda 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 \lambda_i
  - for each new data point (x, f(x)), player i gains payoff
    \mathcal{L}(\mathcal{A}_i, x) \leq \mathcal{L}(\mathcal{A}_{-i}, x)
  - other payoff functions are also considered in this paper

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

Theorem.
Suppose for all \lambda, (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 \lambda; 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 \pi
- Let \lambda^* be the optimal parameter minimizing \mathbb{E}_x[\mathcal{L}(\mathcal{A}, x)]

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

(*) see the formal definition of assumptions A1-A5 in Section 5 of paper
(**) assumptions are numerically verified in empirical experiment

Running Example: Ridge Regression on Price Prediction

Given California housing prices data, train ridge regression to predict housing prices
argmin_\beta \sum_i (\beta \cdot x_i - y_i)^2 - \lambda \|\beta\|_2
where \lambda controls bias-variance tradeoff, selects by planner
- higher \lambda \Rightarrow higher bias, lower variance

Bias-Variance Game
- Research Question: Fixing J, how should planners decide \lambda 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 \lambda_i
  - for each new data point (x, f(x)), player i gains payoff
    \mathcal{L}(\mathcal{A}_i, x) \leq \mathcal{L}(\mathcal{A}_{-i}, x)
  - other payoff functions are also considered in this paper

arXiv preprint: 1909.03618