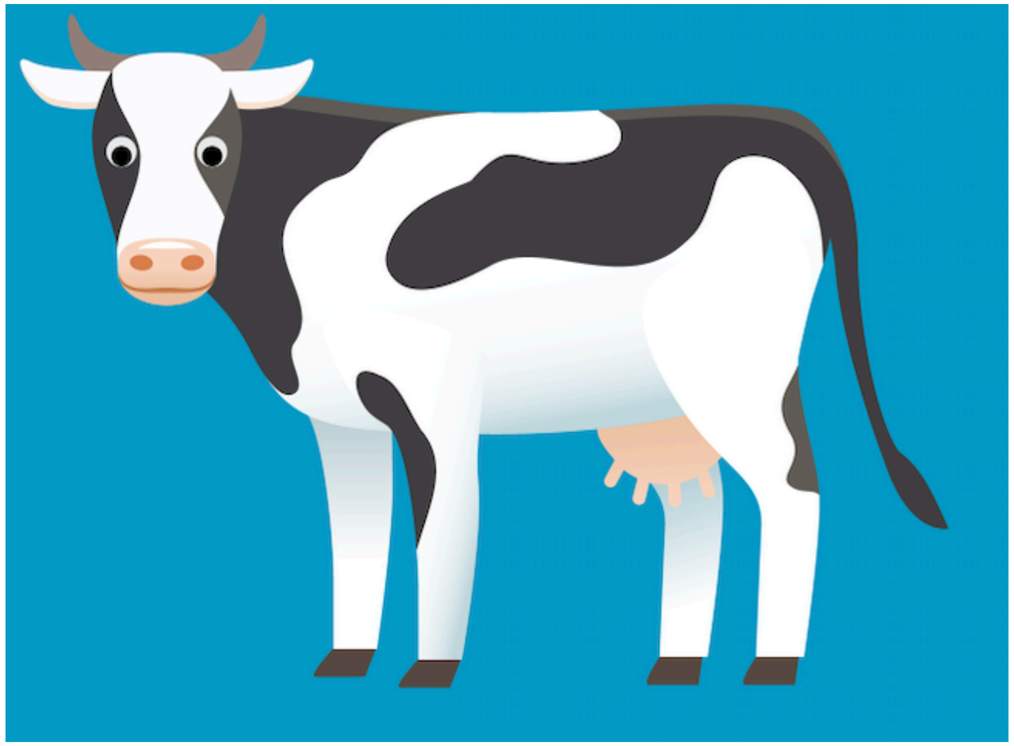
Linear Regression Games:
Convergence Guarantees to Approximate Out-of-Distribution Solutions

Kartik Ahuja

Failures due to spurious correlations

Usual cow: green background



Unusual cow: blue background

Deep neural networks exploit background color!

Motivation

- **Invariant risk minimization** (IRM) recently proposed to address out-of-distribution generalization

- Can IRM or subsequent methods **provably outperform** ERM even when data is generated from linear models?
  - No! due to **non-convex loss**

- How to build methods that provably outperform ERM?

Linear Regression Game

- Two environments, linear predictors, and square loss

- Environment 1 selects $\boldsymbol{\theta}_1 \in \Theta$ to minimize its risk

$$\min_{\boldsymbol{\theta}_1 \in \Theta} R_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbb{E}_{\mathbb{P}_1}\left[\left(Y_1 - \boldsymbol{\theta}_1^\top \boldsymbol{X}_1 - \boldsymbol{\theta}_2^\top \boldsymbol{X}_1\right)^2\right]$$

- Environment 2 selects $\boldsymbol{\theta}_2 \in \Theta$ to minimize its risk

$$\min_{\boldsymbol{\theta}_2 \in \Theta} R_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbb{E}_{\mathbb{P}_2}\left[\left(Y_2 - \boldsymbol{\theta}_1^\top \boldsymbol{X}_2 - \boldsymbol{\theta}_2^\top \boldsymbol{X}_2\right)^2\right]$$

- Environment 1 and 2 playing a **linear regression game (LRG)**

- Simultaneous minimization leads to a **Nash equilibrium (NE)** $(\boldsymbol{\theta}_1^\dagger, \boldsymbol{\theta}_2^\dagger)$

- **NE-based ensemble predictor** $\bar{\boldsymbol{\theta}}^\dagger = \boldsymbol{\theta}_1^\dagger + \boldsymbol{\theta}_2^\dagger$

Linear Regression Game: Key Results

- Θ is ℓ_∞ ball: NE of LRG closer to ideal OOD solution
- Θ is ℓ_∞ ball: best response dynamics learn the NE

Thank you!