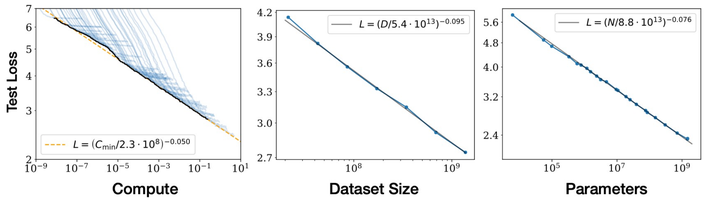
Improved Bayes Risk Can Yield Reduced Social Welfare Under Competition

Meena Jagadeesan, Michael I. Jordan, Jacob Steinhardt*, Nika Haghtalab* (UC Berkeley)

NeurIPS 2023

Scaling trends under competition

Increasing scale improves accuracy for an **isolated system** [4].



However, in digital marketplaces, model-providers often compete with each other for users.

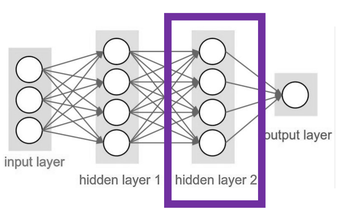
YouTube vs TikTok vs Google vs Spotify vs bing vs pandora vs Instagram

Main question: under competing model-providers, how does increasing scale impact equilibrium social welfare? We study this through the lens of data representations.

From scale to data representations

GPT3 Google BERT

Pretrained model



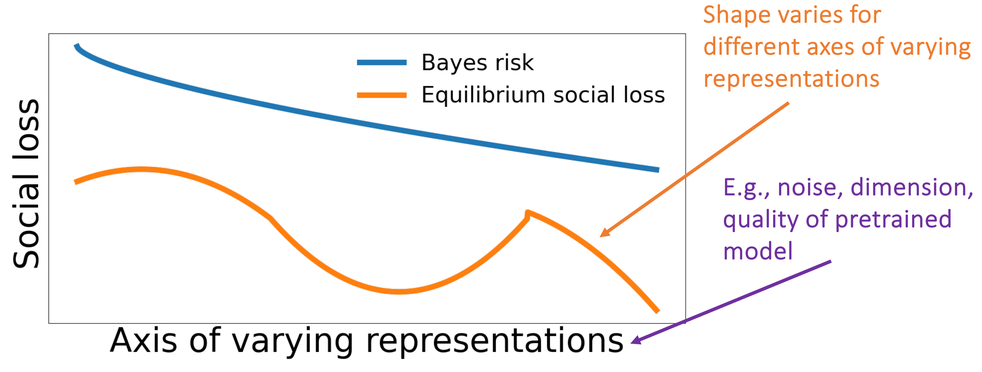
- Platform 1's finetuned model
- Platform 2's finetuned model
- Platform 3's finetuned model

A bird's eye view:

- Pretrained model = learns data representations that improve with scale (e.g., # of parameters)
- Finetuning = uses these data representations to learn a model that optimizes an objective (e.g., market share)

Non-monotonicity of the social welfare

Result (Informal): When model-providers compete for users, the equilibrium social welfare (i.e., overall predictive accuracy) for users can be **non-monotonic** in data representation quality (i.e., Bayes risk).



Consequence for scaling trends: Increasing "scale" may decrease social welfare under competition.

Theoretical characterization of non-monotonicity

Proposition (Informal):

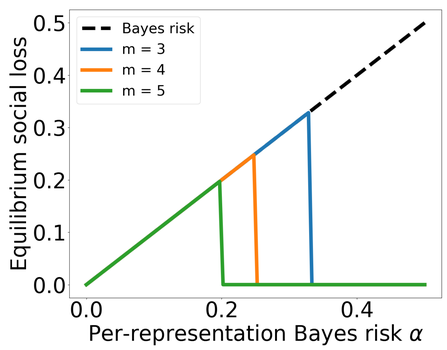
Consider binary classification where F = all deterministic functions. Let $f_1, \ldots, f_m$ be a pure strategy Nash equilibrium.

The equilibrium social loss is:

$SL(f_1, \ldots, f_m) = E[\alpha(x) * 1[\alpha(x) < 1/m]]$,

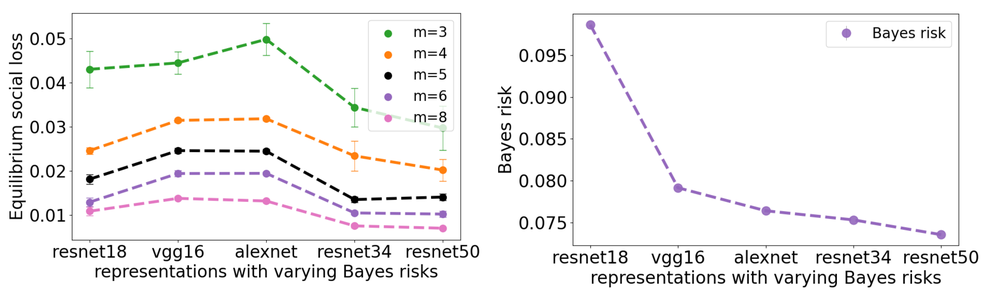
where $\alpha(x)$ is equal to:

$\min(P[Y = 1 | X = x], P[Y = 0 | X = x])$.



(See the paper for generalization to multi-class classification.)

Example experiments on CIFAR-10



Model

Task: classification over $(x, y) \sim D$ with model family F

Each model provider $j \in [m]$ chooses a predictor $f_j \in F$.

Each user (x, y) noisily chooses $j^*(x,y) \in [m]$ offering the best prediction: $\Pr[j^*(x,y) = j] \propto \exp(-\ell(f_j(x), y) \mid / c)$.

A model-provider's utility equals the market share:

$u(f_j; f_{-j}) = E_D[\Pr[j^*(x,y) = j]]$.

Our focus: pure strategy Nash equilibria in the game between m model-providers

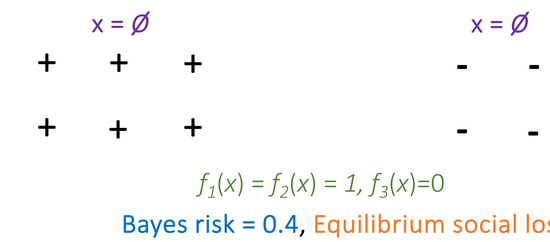
Equilibrium social loss = $E[\ell(f_{j^*(x,y)}(x), y)]$

Bayes risk = $\min_{f \in F} E[\ell(f(x), y)]$

Intuition for non-monotonicity

Intuition: Lower quality data representations lead to greater "disagreement" among the predictors chosen at equilibrium.

Setup 1: Worst data representation quality



$f_1(x) = f_2(x) = 1, f_3(x) = 0$

Bayes risk = 0.4, Equilibrium social loss = 0

Setup 2: Better data representation quality

$f_1(x_0) = f_2(x_0) = 0, f_3(x_0) = 1$ $\quad$ $f_1(x_1) = f_2(x_1) = f_3(x_1) = 1$

Bayes risk = 0.3, Equilibrium social loss = 0.1

References

[1] Ben-Porat, Tennenholtz. "Regression equilibrium". EC 2019.
[2] Kleinberg, Raghavan. "Algorithmic Monoculture". PNAS 2021.
[3] Feng et al. "Bias-variance games". EC 2022
[4] Kaplan et al. "Scaling laws for neural language models." arXiv 2020.