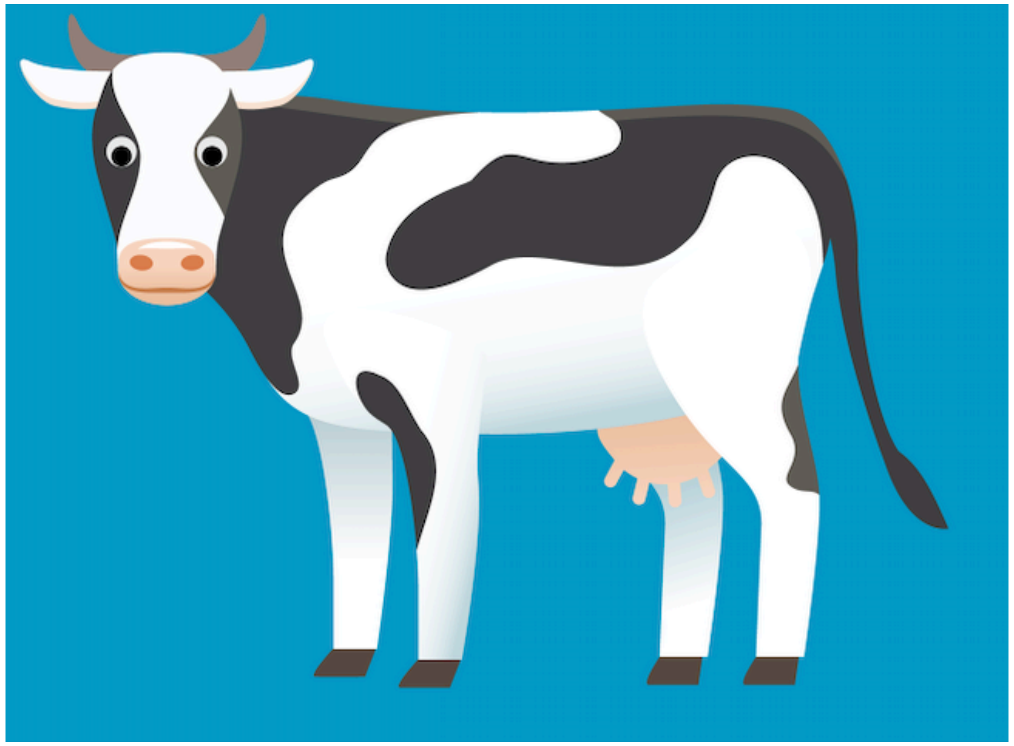
Empirical or Invariant Risk Minimization? A Sample Complexity Perspective

Kartik Ahuja

Failures due to spurious correlation

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

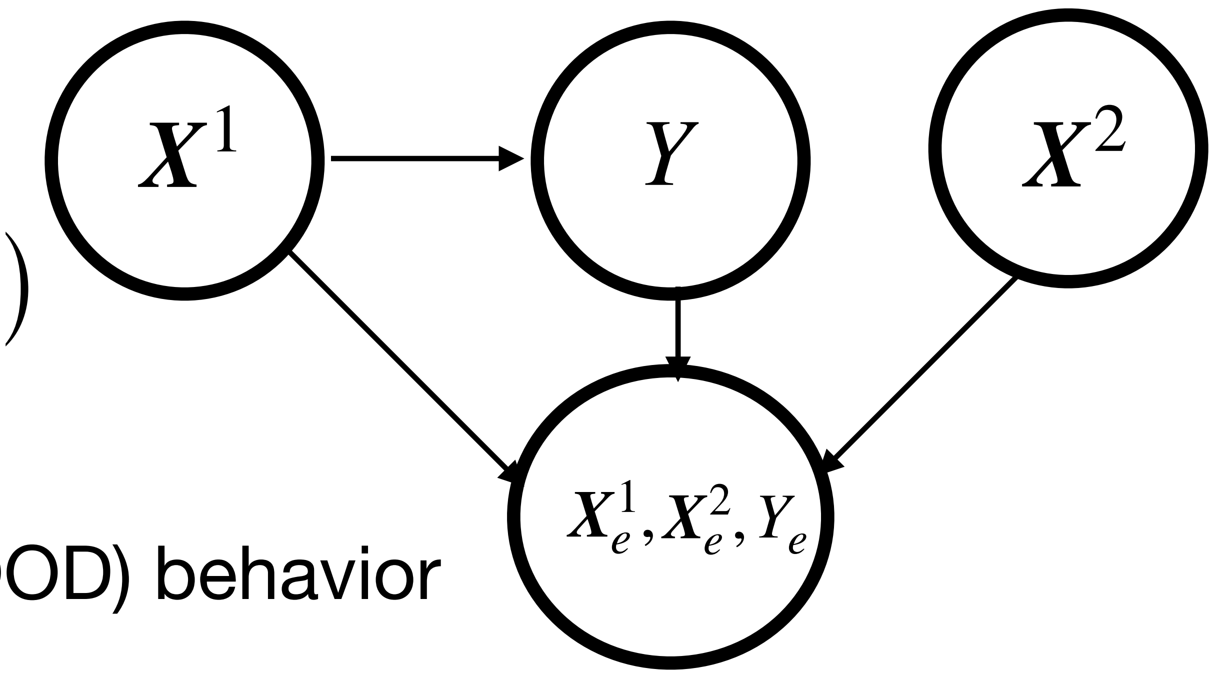
- Recent works [Gulrajani et al.] ERM continues to be SOTA

- A systematic **sample complexity based comparison** of ERM and IRM

IRM vs ERM under selection bias

$$Y \leftarrow f(\boldsymbol{X}^1)$$

$$(Y_e, \boldsymbol{X}_e^1, \boldsymbol{X}_e^2) \leftarrow \text{select}\big((Y, \boldsymbol{X}^1, \boldsymbol{X}^2)\big)$$



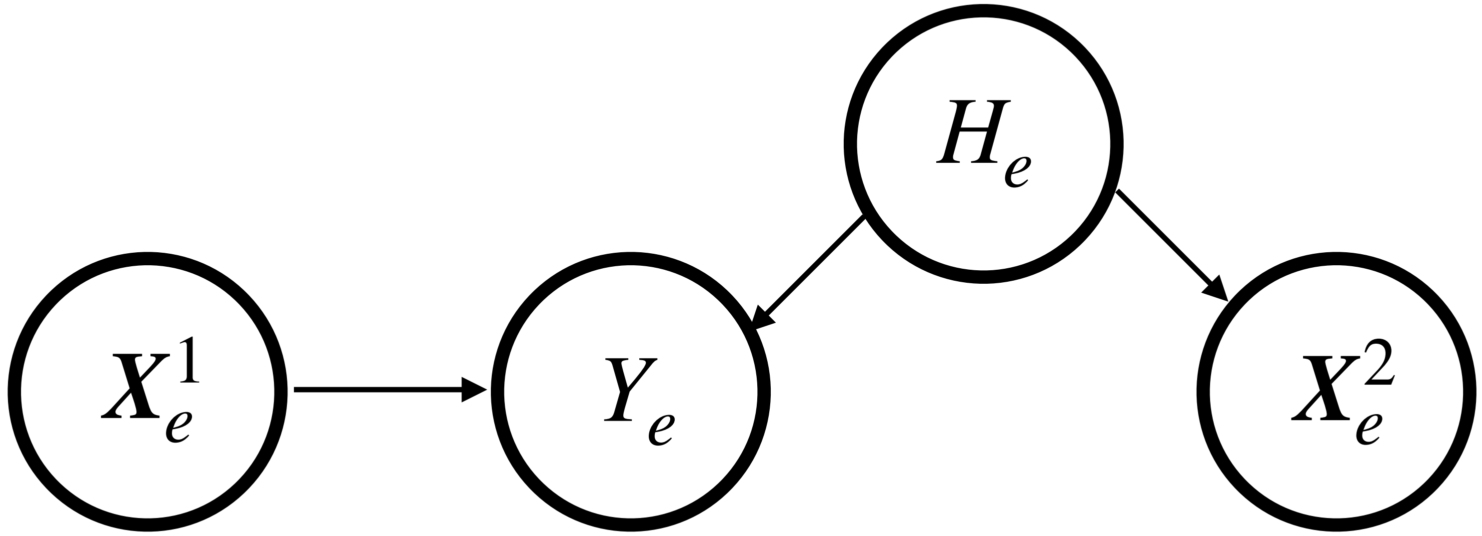
f : Exhibits ideal out-of-distribution (OOD) behavior

Bias	Method	Sample Complexity	OOD
Selection bias	ERM	$\frac{8L^2}{\nu^2}\log\left(2\frac{\lvert\mathcal{H}_\Phi\rvert}{\delta}\right)$	Yes
Selection bias	IRM	$\max\left\{\frac{8L^2}{\nu^2}\log\left(4\frac{\lvert\mathcal{H}_\Phi\rvert}{\delta}\right), \frac{16L'^4}{\epsilon^2}\log\left(\frac{2}{\delta}\right)\right\}$	Yes

IRM vs ERM under confounding bias

$$Y_e \leftarrow f(\boldsymbol{X}_e^1) + g_e(\boldsymbol{H}_e) + \varepsilon_e$$

$$\boldsymbol{X}_e^2 \leftarrow q_e(\boldsymbol{H}_e) + \zeta_e$$



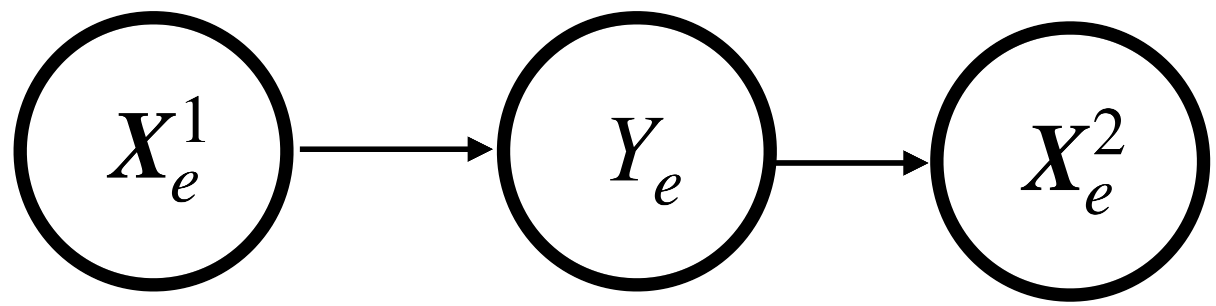
f : Exhibits ideal out-of-distribution behavior

Bias	Method	Sample Complexity	OOD
Confounding	ERM	$\frac{8L^2}{\nu^2}\log\left(2\frac{\|\mathscr{H}_\Phi\|}{\delta}\right)$	No
Confounding	IRM	$\frac{16L'^4}{\epsilon^2}\log\left(\frac{2\|\mathscr{H}_\Phi\|}{\delta}\right)$	Yes

IRM vs ERM under anti-causal bias

$$Y_e \leftarrow f(X_e^1) + \varepsilon_e$$

$$X_e^2 \leftarrow c_e(Y_e) + \zeta_e$$



$f:$ Exhibits ideal out-of-distribution behavior

Bias	Method	Sample Complexity	OOD
Anti-causal	ERM	$\frac{8L^2}{\nu^2}\log\left(2\frac{\lvert\mathscr{H}_\Phi\rvert}{\delta}\right)$	No
Anti-causal	IRM	$\frac{16L'^4}{\epsilon^2}\log\left(\frac{2\lvert\mathscr{H}_\Phi\rvert}{\delta}\right)$	Yes

Thank you!