Exact Generalization Guarantees
For (Regularized) Wasserstein Distributionally Robust Models

Waïss Azizian, Franck Iutzeler, Jérôme Malick

NeurIPS In Paris 2023

Standard and robust models in ML

- $f_\theta(\xi)$ the loss induced by a model parametrized by θ on a sample $\xi = (x, y)$
- $\hat{P}_n$ empirical distribution coming from true distribution P

Empirical risk minimization

$$\text{minimize} \quad \mathbb{E}_{\xi \sim \hat{P}_n}[f_\theta(\xi)] \quad \text{empirical risk}$$

Generalization guarantees:

$$\text{relate} \quad \mathbb{E}_{\xi \sim \hat{P}_n}[f_\theta(\xi)] \quad \text{empirical risk} \quad \text{to} \quad \mathbb{E}_{\xi \sim P}[f_\theta(\xi)] \quad \text{true risk}$$

$\rightarrow$ Only approximate, ERM can lead to overconfident decisions, sensitive to distribution shifts

Standard and robust models in ML

- $f_\theta(\xi)$ the loss induced by a model parametrized by θ on a sample $\xi = (x, y)$
- $\hat{P}_n$ empirical distribution coming from true distribution P

Empirical risk minimization

$$\text{minimize} \quad \mathbb{E}_{\xi \sim \hat{P}_n}[f_\theta(\xi)] \quad \text{empirical risk}$$

Generalization guarantees:

$$\text{relate} \quad \mathbb{E}_{\xi \sim \hat{P}_n}[f_\theta(\xi)] \quad \text{empirical risk} \quad \text{to} \quad \mathbb{E}_{\xi \sim P}[f_\theta(\xi)] \quad \text{true risk}$$

$\rightarrow$ Only approximate, ERM can lead to overconfident decisions, sensitive to distribution shifts

Wasserstein distributionally robust optimization

$$\text{minimize} \quad \sup_{Q : W_2(\hat{P}_n, Q) \leq \rho} \mathbb{E}_{\xi \sim Q}[f_\theta(\xi)] \quad \text{empirical robust risk}$$

where the $\sup$ is over the Wasserstein ball of radius ρ around $\hat{P}_n$

Main Contribution: Exact Generalization for WDRO

Our Theorem (Informal)

Under compactness and smoothness assumptions, for $\delta \in (0,1)$, for ρ small enough and for any n, if

$$\rho \geq \mathcal{O}\left(\sqrt{\frac{\log 1/\delta}{n}}\right)$$

Generalization guarantee: w.p. $1-\delta$, for all $\theta \in \Theta$,

$$\text{empirical robust risk} \quad \sup_{Q: W_2(\hat{P}_n, Q) \leq \rho} \mathbb{E}_{\xi \sim Q}[f_\theta(\xi)] \quad \geq \quad \mathbb{E}_{\xi \sim P}[f_\theta(\xi)] \quad \text{true risk}$$

- Covers many examples: logistic regression, smooth kernels, smooth neural networks,...
- No curse of dimensionality for ρ
- Improves upon existing works [Esfahani and Kuhn, 2018; An and Gao, 2021; Blanchet et al., 2021;...]
- Extensions: distributions shifts, not overly pessimistic, entropic regularization...