

Generalised Lipschitz Regularisation Equals Distributional Robustness

Zac Cranko^{*1} Zhan Shi^{*2} Xinhua Zhang² Richard Nock³ Simon Kornblith³

¹ Universität Tübingen

² University of Illinois at Chicago

³ Google Brain

* Equal contribution

Contributions

- Demonstrate that

Distributional Robustness = Lipschitz Regularization

under **generalized** conditions and **novel** characterization of equality

- Polytime estimation of Lipschitz constant for universal function spaces
 - $O\left(\frac{1}{\epsilon^2}\right)$ sample complexity
 - Reproducing kernel Hilbert space (RKHS) of product kernels
 - Applied to robust SVM training

Distributional Robust Risk

- Motivation: data samples deviate from the true distribution
- Given a distribution μ on $X \times Y$, minimise the worst-case risk

$$DRR(f) := \sup_{\nu} \{ E_{(x,y) \sim \nu} [\text{loss}(f(x), y)] : \text{cost}_c(\mu, \nu) \leq \mathbf{r} \}$$

where $\text{cost}_c(\mu, \nu) = \inf_{\pi} \{ \int c \, d\pi : \pi \text{ couples } \mu \text{ and } \nu \}$

- Standard duality result using Lipschitz constant of f

$$DRR(f) \leq E_{(x,y) \sim \mu} [\text{loss}(f(x), y)] + \mathbf{r} \cdot \text{lip}_c(f)$$

Contribution 1: duality characterization

- Duality result using the **Lipschitz** constant of f

$$DRR(f) \leq E_{(x,y) \sim \mu} [\text{loss}(f(x), y)] + r \cdot \text{lip}_c(f)$$

 onerous assumptions ruling out many ML problems

 loose conditions for equality



 we generalise and improve upon existing results

 we tightly characterize equality

Reference	relation	f	c	μ	X
(Shafieezadeh-Abadeh et al., 2019, Thm. 14)	=	convex Lipschitz margin loss with linear classifier	norm	empirical dist.	\mathbb{R}^d
(Kuhn et al., 2019, Thm. 5)	\leq	upper semicontinuous	norm	empirical dist.	\mathbb{R}^d
(Kuhn et al., 2019, Thm. 10)	=	convex, Lipschitz	norm	empirical dist.	\mathbb{R}^d
(Gao & Kleywegt, 2016, Cor. 2 (iv))	\leq	similar to generalised Lipschitz	p-metric	empirical dist.	\mathbb{R}^d
Theorem 1 (this paper)	\leq =	- convex, generalised Lipschitz	- convex, k-positively homogeneous	probability measure	separable Banach space

Contribution 2: polytime Lipschitz constant

- Enforcing Lipschitz by $\|\nabla f(x_i)\|$ needs exponentially many x_i
- We show Lipschitz constant can be found in polynomial time
 - For a universal function space (RKHS of Gaussian kernel)
 - Product kernel in general $k(x, y) = \prod_{j=1}^d k_0(x_j, y_j)$
 - Method based on Nystrom approximation for $\partial_{y_1} k_0(x_1, y_1)$
 - Draw samples from a Borel measure on X
 - Sample complexity for ϵ error and $1 - \delta$ probability of Lipschitz constant

$$\tilde{\Theta}\left(\frac{1}{\epsilon^2} N_\epsilon^2 M_\epsilon^2 Q_\epsilon^2 \log \frac{dN_\epsilon}{\delta}\right)$$

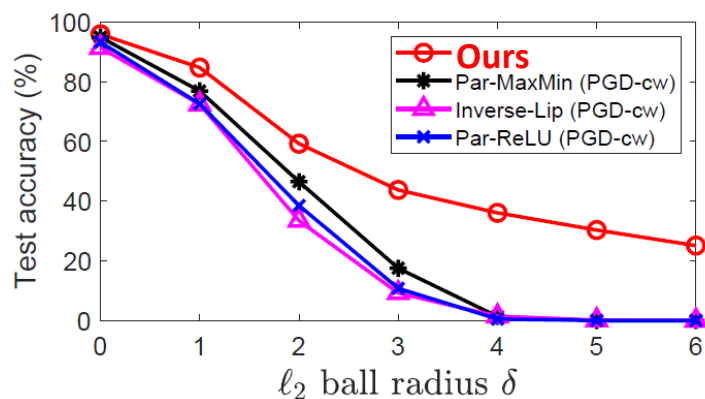
- Logarithmic in dimensionality
- $N_\epsilon, M_\epsilon, Q_\epsilon$ depend on kernel spectrum: universal constant for Gaussian and periodical kernel

Experiment

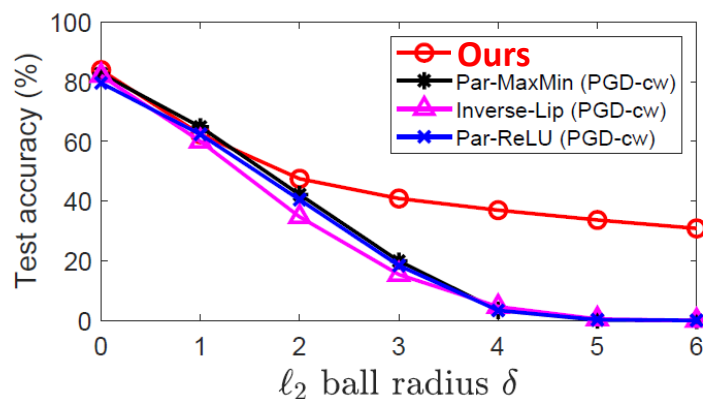
Test accuracy under PGD attacks on the C&W approximation

∞ -norm
bound

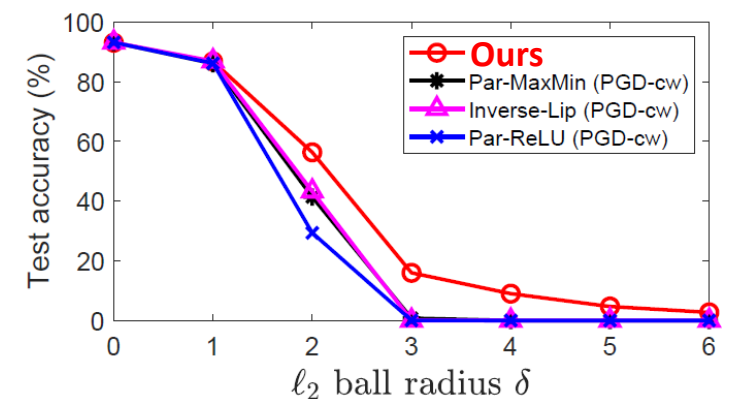
MNIST



Fashion-MNIST



CIFAR10



2-norm
bound

