Robust Certification for Laplace Learning on Geometric Graphs

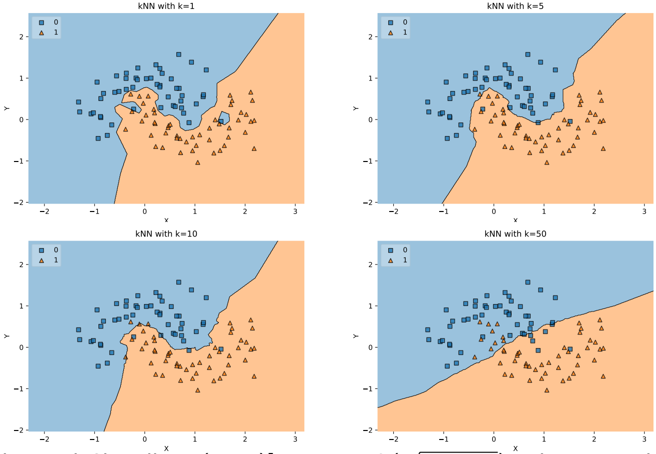
Bao Wang
Department of Mathematics
Scientific Computing and Imaging Institute
University of Utah

Joint work with Matthew Thorpe, University of Manchester

Matthew Thorpe and Bao Wang, Robust Certification for Laplace Learning on Geometric Graphs, arXiv:2104.10837, 2021. (MSML, accepted)

Partially supported by DOE, NSF, and Univ of Utah

Robustness of the kNN Classifier



Theorem. [Wang, Jha, and Chaudhuri (2018)] For $k = \Omega(\sqrt{dn \log n})$, where d is the data dimension and n is the sample size, then the robustness region of kNN classifier approaches that of the Bayes Optimal classifier in the large sample limit.

Semi-supervised Laplace Learning

kNN did not fully utilize the underlying geometry of the data, can geometry of the data improve robustness of the classifier?

We consider semi-supervised Laplace learning, which can be formulated as: Let $\Omega_N := \{\mathrm{x}\}_{i=1}^N \subset \mathbb{R}^d$ be a set of feature vectors with a subset of $\Gamma_N := \{\mathrm{x}_i\}_{i \in Z_N} \subset [N]$, and we denote the label be $\ell_N(\mathrm{x}) := \ell|_{\Gamma_N}$. We can construct a graph $\mathrm{W}_N := (\mathrm{W}_{\mathrm{x},\mathrm{y}})_{\mathrm{x},\mathrm{y} \in \Omega_N}$. We solve the following constrained minimization problem to get the prediction

$$\min_{u(\mathrm{x})} \sum_{\mathrm{x},\mathrm{y} \in \Omega_N} \mathrm{W}_{\mathrm{x},\mathrm{y}} (u(\mathrm{x}) - u(\mathrm{y}))^2 \quad \text{subject to} \quad u(\mathrm{x}) = \ell_N(\mathrm{x}).$$

In particular, we consider the *Geometric Random graphs*, in which $\mathrm{W}_{\mathrm{x},\mathrm{y}} = \mathrm{W}_{\epsilon,\mathrm{x},\mathrm{y}} = \eta_\epsilon(|\mathrm{x} - \mathrm{y}|)$ and $\eta_\epsilon = \frac{1}{\epsilon^d}\eta(\cdot/\epsilon)$ and $\eta : [0, +\infty) \to [0, +\infty)$ is non-increasing, positive, $\eta(t) \geq 1$ for all $t \leq 1$ and $\eta(t) = 0$ for all $t \geq 2$. In addition, wither η is Lipschitz continuous, or $\eta(t) = 1_{t \leq 1}$.

How Laplace learning improves robustness of the classifier?

Theoretical Result

δ-Robustness Radius Definition: Let D_n be a dataset of n feature vectors of which the fraction $\beta \in [0,1]$ are labelled, $\hat{D}_n$ be any dataset built by perturbing the feature vectors at most r (but keping the same labels), and $D_n \mapsto u(\cdot; D_n)$ a solution to the semi-supervised learning problem. The δ-robustness radius $\mathcal{R}_\delta(D_n)$ is the largest r such that

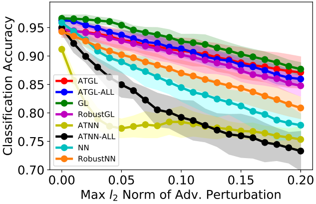
$$\sup_x |u(x; D_n) - u(\hat{x}; \hat{D}_n| \leq \delta.$$

Approximate Statement of Theorem: Let u be the Laplacian Regularisation solution. For ϵ small enough and $\beta \in [\epsilon^2, 1]$ there exists constants $C > c > 0$ such that for all $r \in (0, c\sqrt{\beta}\epsilon)$ with probability $1 - Cne^{-cn\beta\epsilon^d}$ the δ-robustness radius is greater than r for

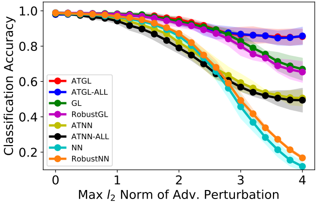
$$\delta = \frac{C\epsilon}{\sqrt{\beta}} \log\left(\frac{\sqrt{\beta}}{\epsilon}\right).$$

Classifier	k	Assumption on r	Reference
kNN	$\Omega(\sqrt{n \log n})$	None	Wang, Jha, and Chaudhuri (2018)
GL	$\Omega\left(\frac{\log n}{1-\beta}\right)$	$r \leq c\sqrt{1-\beta}\left(\frac{\log n}{n(1-\beta)}\right)^{\frac{1}{d}}$	**This Work**

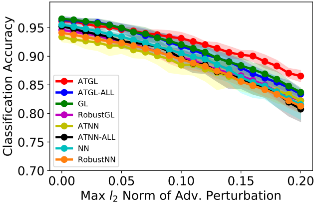
Numerical Results



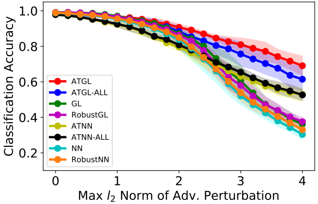
Halfmoon, Direct Attack



MNIST 1v7, Direct Attack



Halfmoon, Kernel Sub Attack



MNIST 1v7, Kernel Sub Attack

Figure: Robust accuracies of GL vs. kNN classifiers for three datasets classification under WB attacks with different maximum perturbation measured in ℓ_2-norm. GL-based classifiers are consistently more accurate than kNN-based classifiers.