Noise Injection Node Regularization for Robust learning

Noam Levi* Itay Bloch* Marat Freytsis Tomer Volansky

* Denotes equal contribution

Overview

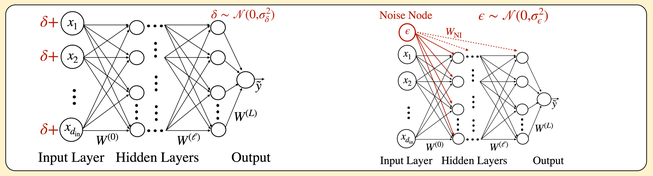
- We **introduce** a new mechanism for probing deep learning dynamics via adaptive *Noise Injection Nodes (NINs)*.
- **We show:** the system undergoes distinct phases of learning, depending on the scale of injected noise.
- Training with NINs results in an **implicit regularization scheme**: *Noise Injection Node Regularization (NINR)*.

Noise Injection Nodes & Dynamics

Inspired by Zhang et al. (2017) & Arpit et al. (2017)

"Deep neural networks manage to capture the correlations in noisy training data while fitting the noisy part by brute force"

Standard input corruption: Noise Injection Node:



- Approximate SGD equations for the weights including NINs :

NIWs: $W_{\rm NI}^{(t+1)} \simeq \left(1 - \frac{1}{2}\eta\sigma_\epsilon^2 \mathscr{H}_{\ell_{\rm NI}}\right) W_{\rm NI}^{(t)} - \eta \frac{\sigma_\epsilon}{\sqrt{\mathscr{B}}} g_{\ell_{\rm NI}}^{(t)}$

Exponential factor — Random walk

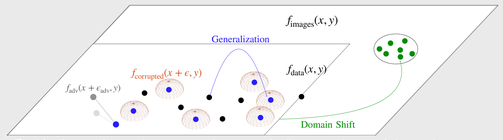
Data weights: $W^{(t+1)} \simeq W^{(t)} - \eta \frac{\partial L_{\epsilon=0}}{\partial W^{(t)}} - \eta W_{\rm NI}^{(t)} \frac{\sigma_\epsilon}{\sqrt{\mathscr{B}}} \frac{\partial g_{\ell_{\rm NI}}^{(t)}}{\partial W^{(t)}} - \frac{1}{2}\eta\sigma_\epsilon^2 W_{\rm NI}^{(t)} \frac{\partial \mathscr{H}_{\ell_{\rm NI}}}{\partial W^{(t)}} W_{\rm NI}^{(t)}$

Objective gradient — Random walk — Exponential factor

- **For small noise, the NIWs decay and standard learning proceeds.**
- **Dynamics change with the local curvature:** $|1 - \eta\sigma_\epsilon^2 \mathscr{H}_{\ell_{\rm NI}}| \geq 1.$
- **At large noise, the network begins by learning on random data, leading to de-noising or failure.**

Noise Injection Node Regularization

- Distributional shifts are ubiquitous in ML deployment



*Red indicates input data corruption, gray indicates adversarial attacks

- Implicit regularization can be seen by expanding the noisy average loss function in the large batch size $|\mathscr{B}|$ limit:

$$L(\theta, W_{\rm NI}) = \frac{1}{|\mathscr{B}|} \sum_{x,y,\epsilon \in \mathscr{B}} \mathscr{L}(\theta, W_{\rm NI}; x, y, \epsilon) \simeq L(\theta) + \frac{1}{2}\sigma_\epsilon^2 W_{\rm NI}^T \mathscr{H}_{\ell_{\rm NI}} W_{\rm NI}$$

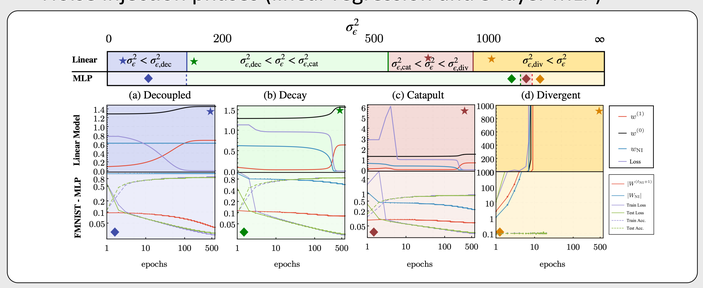
Regularizes the local Hessian

Method:

Train with NINs produces implicit regularization (best at large noise) → Set the NIN to zero, test on corrupted inputs → Observe substantial increase in robustness to input noise

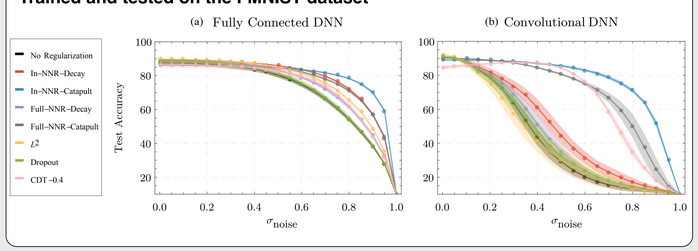
Results

- Noise injection phases (linear regression and 3-layer MLP)



- Robustness against input corruption (NINR vs. other methods)

***Trained and tested on the FMNIST dataset**



Comparisons

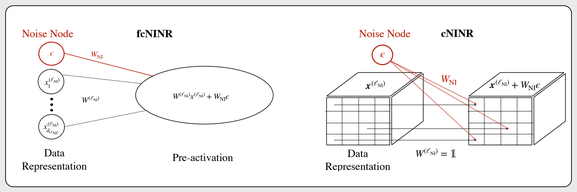
•Test accuracy on clean data (FMNIST)

	None	L_2	Dropout	in-NINR (Decay)	in-NINR (Catapult)	full-NINR (Decay)	full-NINR (Catapult)	CDT (0.2)	CDT (0.4)
FC(%)	87.7 ± 2.6	89.9 ± 3.4	88.3 ± 0.5	89.1 ± 0.7	*86.2 ± 0.8*	88.5 ± 1.8	88.2 ± 0.6	*86.6 ± 0.9*	*85.5 ± 3.8*
CNN(%)	91.0 ± 1.0	92.2 ± 0.7	91.0 ± 1.1	91.0 ± 1.2	*89.0 ± 0.6*	91.0 ± 0.8	90.0 ± 0.2	*84.6 ± 2.6*	*84.1 ± 6.4*

***CDT indicates training with corrupted inputs**
***in-NINR indicates connecting a NIN only at the input, while full-NINR indicates connecting a NIN at every layer**

- As expected, training with simple corrupted inputs lead to degradation in clean test accuracy

Dense & CNN Implementations

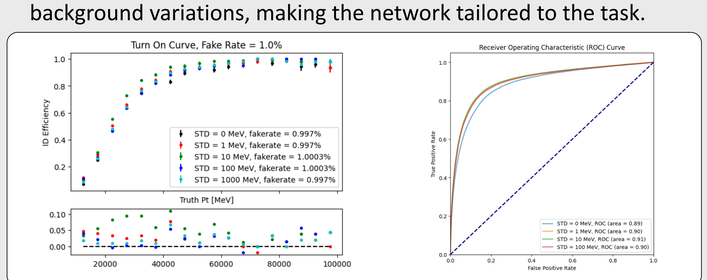


Applications to Detector Simulations

Preliminary Collaboration with Yuval Frid and Liron Barak

We train a ResNET18 to distinguish prompt from QCD fragmented photons in events generated by PYTHIA8 and GEANT4 (COCOA).

- We generate full events for both prompt and QCD fragments and train the networks with/without NINR.
- We observe at some of the working points, a definite increase in detection ability on real events for networks trained with NINR, even though the injected noise has a simple gaussian distribution.
- We plan to introduce structured noise to better model the background variations, making the network tailored to the task.



RUTGERS THE STATE UNIVERSITY OF NEW JERSEY — BERKELEY LAB Lawrence Berkeley National Laboratory — אוניברסיטת תל-אביב TEL AVIV UNIVERSITY — EPFL

Check Out Our Papers: