

Exercise 6

Neural Networks

Program Analysis for System Security and Reliability 2018
ETH Zürich

April 24, 2018

Recall that the activation function $\text{ReLU}(x) = \begin{cases} x & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$

Let N_θ be a one input and one hidden layer neural network defined as

$$N_{\theta \in \mathbb{R}^3}(x) = -\text{ReLU}(\text{ReLU}(-2x - 1)\theta_1 + \text{ReLU}(2x - 1)\theta_2 + \theta_3)$$

We define the loss to be the identity on the output of this network: $\text{loss}_{N_\theta}(x) = N_\theta(x)$

Problem 1. Suppose $\theta = [1, 1, 1]$. Use T-FGSM on N_θ with $\epsilon = 1$ and $x = 0.5$

Problem 2. Apply a second T-FGSM again the new input you found.

Problem 3. Find values for θ where two iterations of T-FGSM with $\epsilon = 0.5$ find a value for x which is closer to 0.5 than the value discovered by one iteration with $\epsilon = 0.5$. Feel free to use a computer to help solve these.

Problem 4. Prove that a one-hidden-layer neural network with ReLU activations, $N_\theta : \mathbb{R} \rightarrow \mathbb{R}$ can approximate function B defined as follows:

$$B(x) = \begin{cases} 1 & x \in (0, 1] \\ 0 & \text{otherwise.} \end{cases}$$

To approximate B pointwise means that you can find a sequence $\theta_{i=1\dots}$ such that for any $x \in \mathbb{R}$ and for any $\epsilon > 0$, there is an $i \in \mathbb{N}$ such that:

$$\forall j > i. |N_{\theta_j}(x) - B(x)| \leq \epsilon.$$