Exercise 6

Neural Networks

Program Analysis for System Security and Reliability 2018
ETH Zürich

April 24, 2018

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

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 $x = 0.5$

Solution. $x' = 1.5$ and $N_\theta(x') = 3$

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

Solution. $x'' = 2.5$ and $N_\theta(x'') = 5$

Problem 3. Find values for θ where two iterations of 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.

Solution. $\theta = (-1, -1, 0)$

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

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

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

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

Solution. Let us define our neural network with a single parameter θ used multiple times.

$$N_\theta(x) = \text{ReLU}(x\theta) - \text{ReLU}(x\theta - 1) - \text{ReLU}(x\theta - \theta) + \text{ReLU}(x\theta - \theta - 1)$$

We can then define our sequence as $\theta_i = i$.

Let $\epsilon > 0$, and $x \in \mathbb{R}$. We can break this down into the following cases:

- $x \leq 0$: In this case, no matter what $i \in \mathbb{N}$ is chosen, $N_{\theta_i}(x) = 0 = B(x)$.
- $x \in (0, 1]$: In this case, let us pick $i = \max\{\lceil \frac{1-\epsilon}{x} \rceil, 1\}$. Suppose $x \leq \frac{1}{i}$. Then $N_{\theta_i}(x) = xi$ and $B(x) = 1$ then

  $$|N_{\theta_j}(x) - B(x)| = |xi - 1| = 1 - xi \leq 1 - x\frac{1-\epsilon}{x} = \epsilon.$$

  Finally, suppose $x > \frac{1}{i}$ then $N_{\theta_i}(x) = 1 = B(x)$
- $x > 1$: One can see the similarities from the proof of the previous case.