hw2

October 10, 2024

CS 521: Fall 2024

Homework 2

By Alex Marzban

Visit the GitHub repository to access the code for this assignment: https://github.com/marzbana/CS521_HWs/hw2

0.1 Problem 1: MILP Encoding for Maxpool

In the lecture, we learned about the Mixed Integer Linear Programming (MILP) based encoding of the ReLU operation. In this exercise, your goal is to design an encoding of the Maxpool operation using MILP.

0.1.1 (a)

Design a MILP encoding for the Maxpool operation:

$$y := \max(x_1, x_2)$$

where the input bounds for x_1 and x_2 are $[a_1, b_1]$ and $[a_2, b_2]$ with a_1, a_2, b_1, $b_2 \in \mathbb{R}$.

Answer

$$y = \max(x_1, x_2)$$

$$y \geq x_1 \tag{1}$$
$$y \geq x_2 \tag{2}$$
$$y \leq x_1 + \max(b_1 + b_2) \cdot (1 - a) \tag{3}$$
$$y \leq x_2 + \max(b_1 + b_2) \cdot a \tag{4}$$
$$a \in \{0, 1\} \tag{5}$$

0.1.2 (b)

Now, consider the neural network shown in Fig. 1. The neural network has two inputs (x_1, x_2) and two outputs (x_9, x_{10}) neurons and consists of two layers with affine transformations (edges colored blue) and one layer with the maxpool operation (edges colored green). The transformations in the network are given as:

$$
\begin{aligned}
x_3 &:= x_1 + x_2 \\
x_4 &:= x_1 - 2 \\
x_5 &:= x_1 - x_2 \\
x_6 &:= x_2 \\
x_7 &:= \max(x_3, x_4) \\
x_8 &:= \max(x_5, x_6) \\
x_9 &:= x_7 \\
x_{10} &:= -x_7 + x_8 - 0.5
\end{aligned}
$$

Use the MILP encoding for ReLU and Maxpool to verify the property that for all values of $x_1, x_2 \in [0, 1]$, the output satisfies $x_9 > x_{10}$. Can the MILP analysis prove this property? Show your work.

Answer

Objective

$$\text{Minimize } x_9 - x_{10}$$

Subject to

Box Bounds

$$
\begin{aligned}
0 \le x_1 \le 1 \\
0 \le x_2 \le 1 \\
0 \le x_3 \le 2 \\
-2 \le x_4 \le -1 \\
-1 \le x_5 \le 1 \\
0 \le x_6 \le 1 \\
0 \le x_7 \le 2 \\
0 \le x_8 \le 1 \\
0 \le x_9 \le 2 \\
-2.5 \le x_{10} \le 0.5
\end{aligned}
$$

Affine

$$
\begin{aligned}
x_3 &= x_1 + x_2 \\
x_4 &= x_1 - 2 \\
x_5 &= x_1 - x_2 \\
x_6 &= x_2 \\
x_9 &= x_7 \\
x_{10} &= -x_7 + x_8 - 0.5
\end{aligned}
$$

Maxpool Constraints

For $x_7 = \max(x_3, x_4)$ **:**

$$
\begin{aligned}
& x_7 \geq x_3 && (1) \\
& x_7 \geq x_4 && (2) \\
& x_7 \leq x_3 + 2(1 - a_7) && (3) \\
& x_7 \leq x_4 + 2a_7 && (4) \\
& a_7 \in \{0, 1\} && (5)
\end{aligned}
$$

For $x_8 = \max(x_5, x_6)$ **:**

$$
\begin{aligned}
& x_8 \geq x_5 && (6) \\
& x_8 \geq x_6 && (7) \\
& x_8 \leq x_5 + 2(1 - a_8) && (8) \\
& x_8 \leq x_6 + 2a_8 && (9) \\
& a_8 \in \{0, 1\} && (10)
\end{aligned}
$$

Property Verification To verify:

$$x_9 > x_{10}, \forall x_1, x_2 \in [0, 1]$$

We must solve the MLP:

1. **Objective:** Minimize $x_9 - x_{10}$
2. **Property Holds If:** The minimum value of $x_9 - x_{10}$ is greater than 0.

Solving the MILP

```
Property Verification:
x9 - x10 = 0.5 > 0
Property holds: x9 > x10 for all inputs in [0,1] x [0,1].
```

0.2 Problem 2: Programming

In this problem, you will implement interval analysis for a simple neural network.

0.2.1 Network Description

Implement a fully connected neural network consisting of **3 layers** (each *layer* here is a linear layer followed by a ReLU), each of size **50 neurons**. Use cross-entropy loss and train your network on MNIST. You can use the dataloaders from previous assignments (`https://github.com/ishcha/CS521_HWs/blob/main/hw1/adversarial_training.ipynb`) if you want.

0.2.2 Task

Implement interval analysis for your network. Use this to measure the robustness of your network for **10 L-infinity neighborhoods**, sized evenly between **0.01** and $\mathbf{0.1^2}$ (inclusive). As usual, present your observations clearly along with experimental evidence (numbers, images, etc.) in your submission.

0.2.3 Solution Requirements

You should present your solution for this in the form of a Jupyter notebook. We recommend using **Google Colab** since we can interact with your solution easily, but you can also just upload the notebook to your **GitHub repo**.

0.3 Problem 3: Interpreting Neural Network Robustness Proofs

We have discussed various neural network verification techniques in the lectures. The proofs of robustness of neural networks generated by these techniques could, however, be uninterpretable for humans. Recent work PROFIT addresses this issue and provides insights into the parts of neural networks important for robustness proofs. In this problem, we ask you to read the aforementioned paper and write a critical, conference-style review for it including the following parts:

1. **Summary:**
   Summarize the paper in your own words.
2. **Strengths (at least 3):**
   Elaborate on the strengths you see in the paper.
3. **Weaknesses (at least 3):**
   Mention the weaknesses of this paper and suggest possible fixes for the same.
4. **Extensions (at least 1):**
   Although such a section is not included in conference reviews, mention possible extensions of this work in the context of contemporary research.

Fun fact: This paper was the result of a course project in CS 521 - Fall 2022.