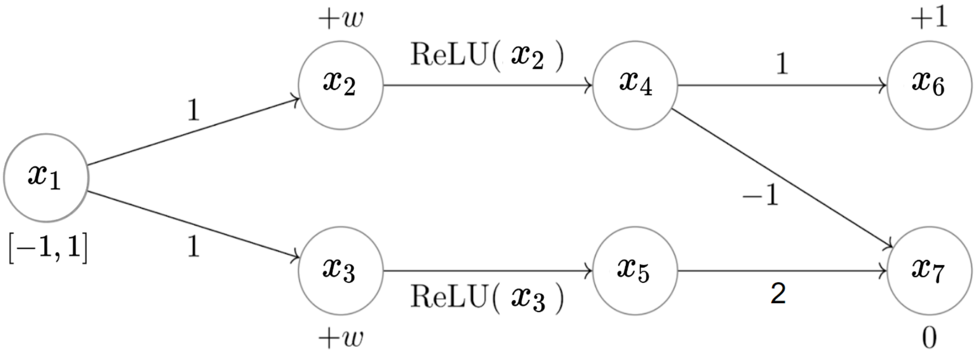
Exercise 05

Continuity of Certified Training

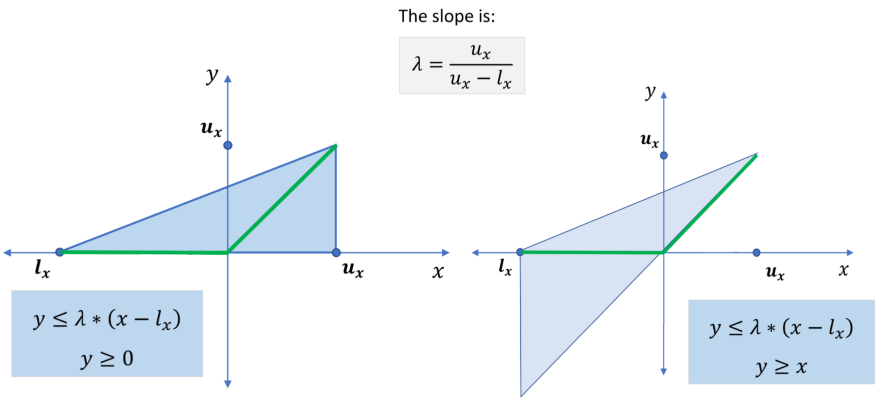
Reliable and Trustworthy Artificial Intelligence
ETH Zurich

Problem 1 (Continuity of DeepPoly in the Weight Space)**.** In this exercise, we demonstrate one common problem that arises during certified training with many precise convex relations, namely, that they are not continuous functions of the network weights. This phenomena in turn results in a much harder optimization problem that needs to be solved by the certified training algorithm. Consider the following neural network:



Here, the network has single input neuron $x_1 \in [-1, 1]$ and 1 parameter w such that $x_2 = x_1 + w$, $x_3 = x_1 + w$, $x_4 = \text{ReLU}(x_2)$, $x_5 = \text{ReLU}(x_3)$, $x_6 = x_4 + 1$ and $x_7 = 2x_5 - x_4$.

(a) Recall that in DeepPoly in order to compute the linear bounds of a ReLU node whose sign cannot be determined, we need to choose which of the two different convex relaxations for the ReLU, shown in (a) and (b) in Fig. 1, is applied. A common efficient heuristic to choose between the two relaxations is based on selecting the triangle with the smaller area in Fig. 1 which is equivalent to comparing u_x and $-l_x$ (see Exercise 03). Using this heuristic, compute the DeepPoly upper bound for x_7 as a function of the network parameter w. Is the function continuous? Why or why not?


The slope is:
$\lambda = \frac{u_x}{u_x - l_x}$
y
u_x
l_x
u_x
x
$y \le \lambda * (x - l_x)$
$y \ge 0$
y
u_x
l_x
u_x
x
$y \le \lambda * (x - l_x)$
$y \ge x$

(a) $u_x \le -l_x$ (b) $u_x > -l_x$

Figure 1: DeepPoly ReLU approximations

(b) Consider another heuristic for DeepPoly that always selects the relaxation from (a) in Fig. 1 regardless of the area. Compute the upper bound for x_7 as a function of the network parameter w using this simpler heuristic. Is the function continuous? Why or why not?

(c) Compute the upper bound for x_7 as a function of the network parameter w using the Box domain. Is the function continuous? Why or why not?

(d) Plot the functions for the different methods above (i.e the two DeepPoly heuristics and Box) on the same graph. What do you observe?