# Neural Network Verification

# Neural Network Verification

Neural network f

Scalar output z = f($\mathbf{x}$)

E.g. in binary classification, z = s(y*;$\mathbf{x}$) – s(y;$\mathbf{x}$) for y ≠ y*

Property: f($\mathbf{x}$) > 0 for all $\mathbf{x} \in X$

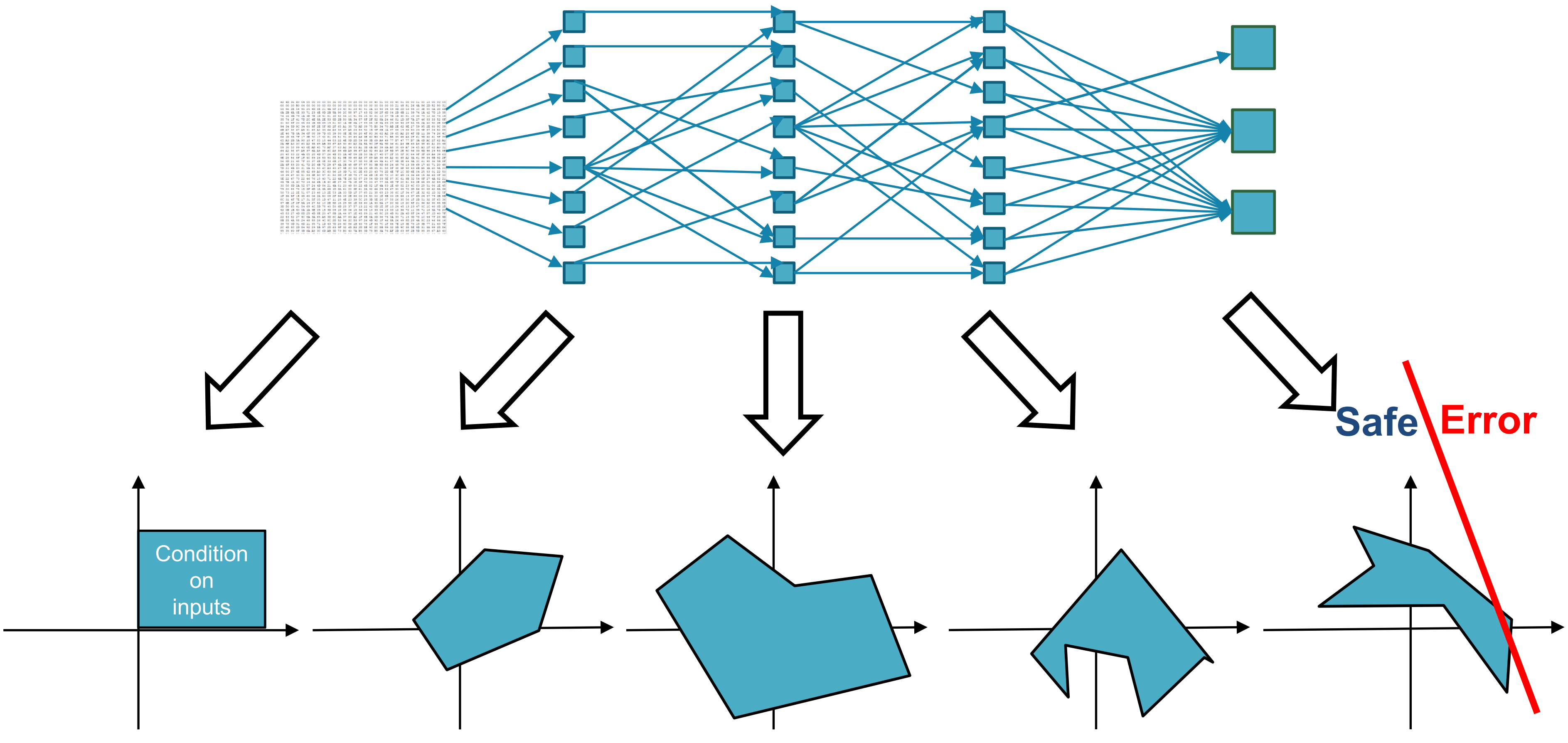
# Outline

- Incomplete Verification
  - Overview
  - Example: Interval Bound Propagation
  - Example: Linear Programming Relaxation

- Complete Verification
  - Branch and Bound
  - Application to verification

# Neural Network Verification

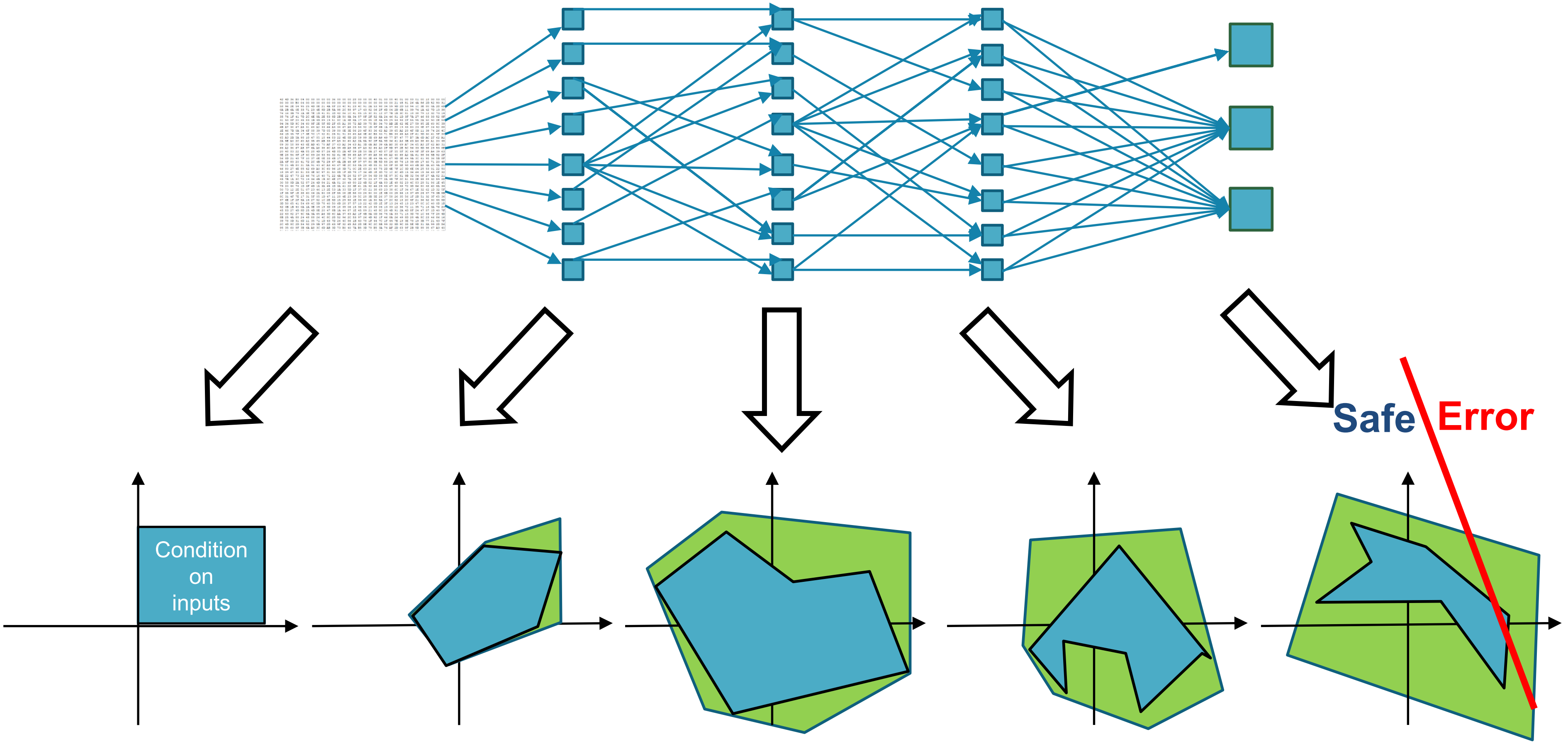
Is there an erroneous output?




Non-convexity makes the problem NP-hard

# Incomplete Verification

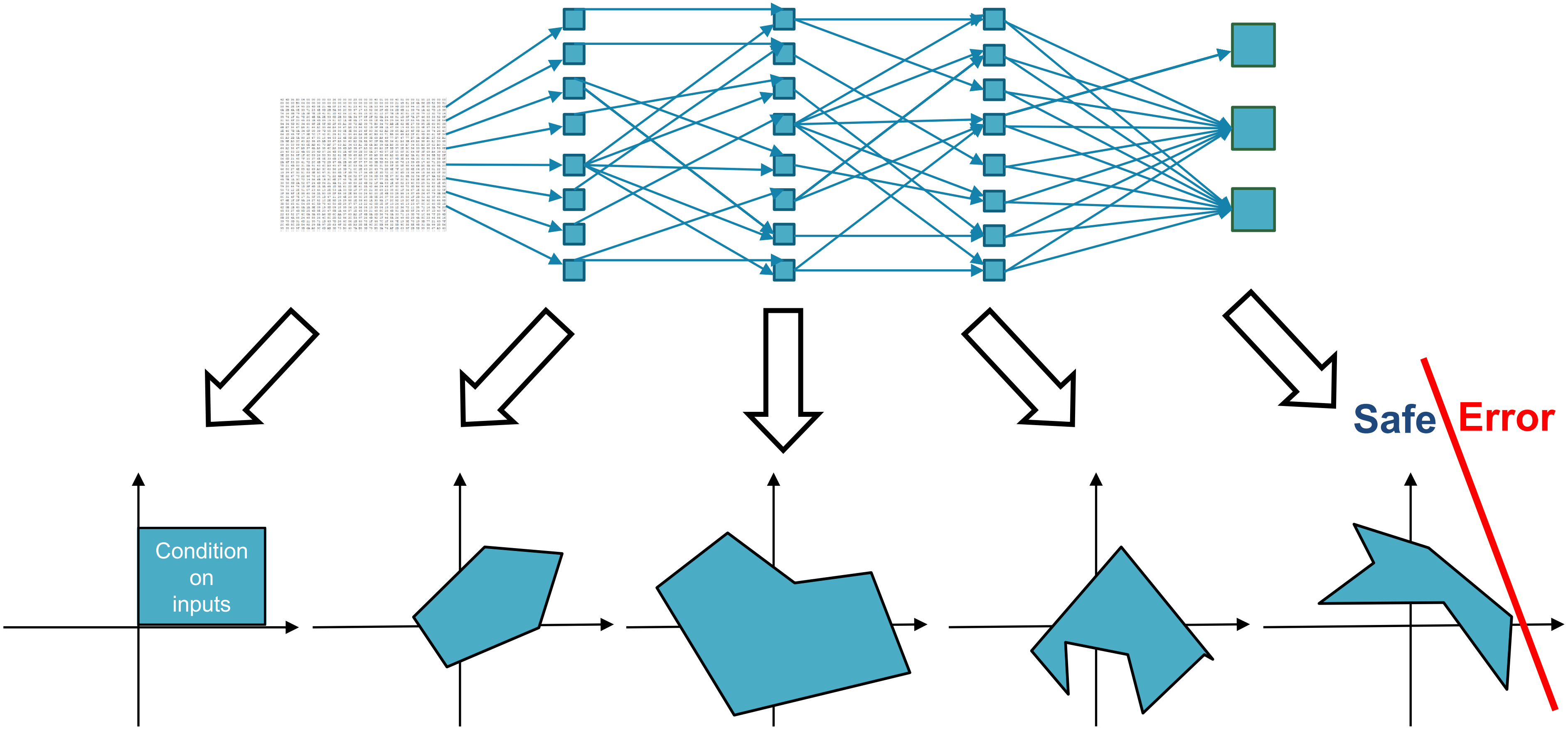
Is there an erroneous output?




Replace by a convex superset

# Incomplete Verification

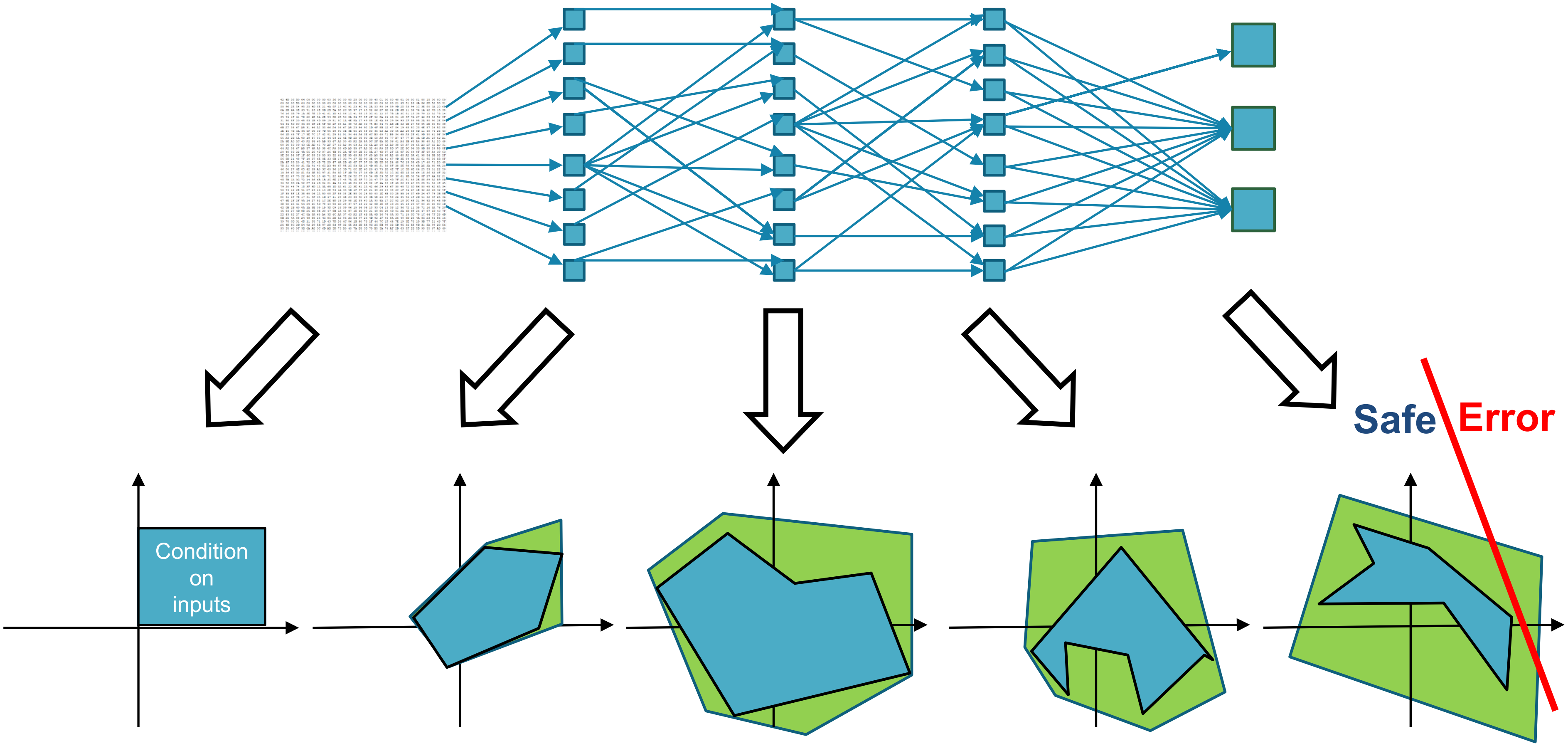
Is there an erroneous output?




Say, non-convex set has no erroneous output

# Incomplete Verification

Is there an erroneous output?




Convex superset might give incorrect answer

# Incomplete Verification

- Useful in practice
- Verifiably robust training
- Key part of complete verification
- How do we construct convex superset?

# Outline

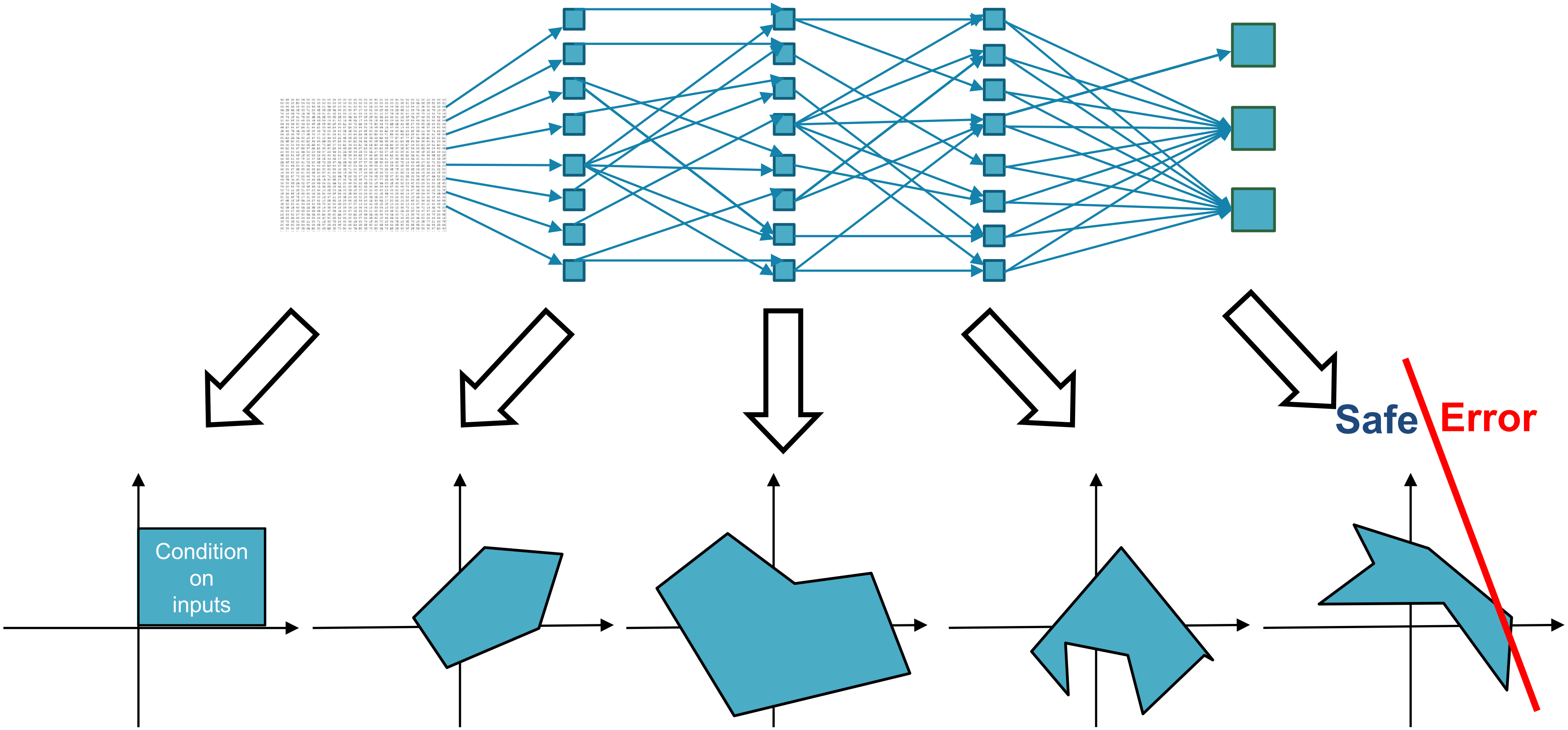
- Incomplete Verification
  - Overview
  - **Example: Interval Bound Propagation**
  - Example: Linear Programming Relaxation

- Complete Verification
  - Branch and Bound
  - Application to verification

Mirman et al., 2018; Gowal et al., 2018

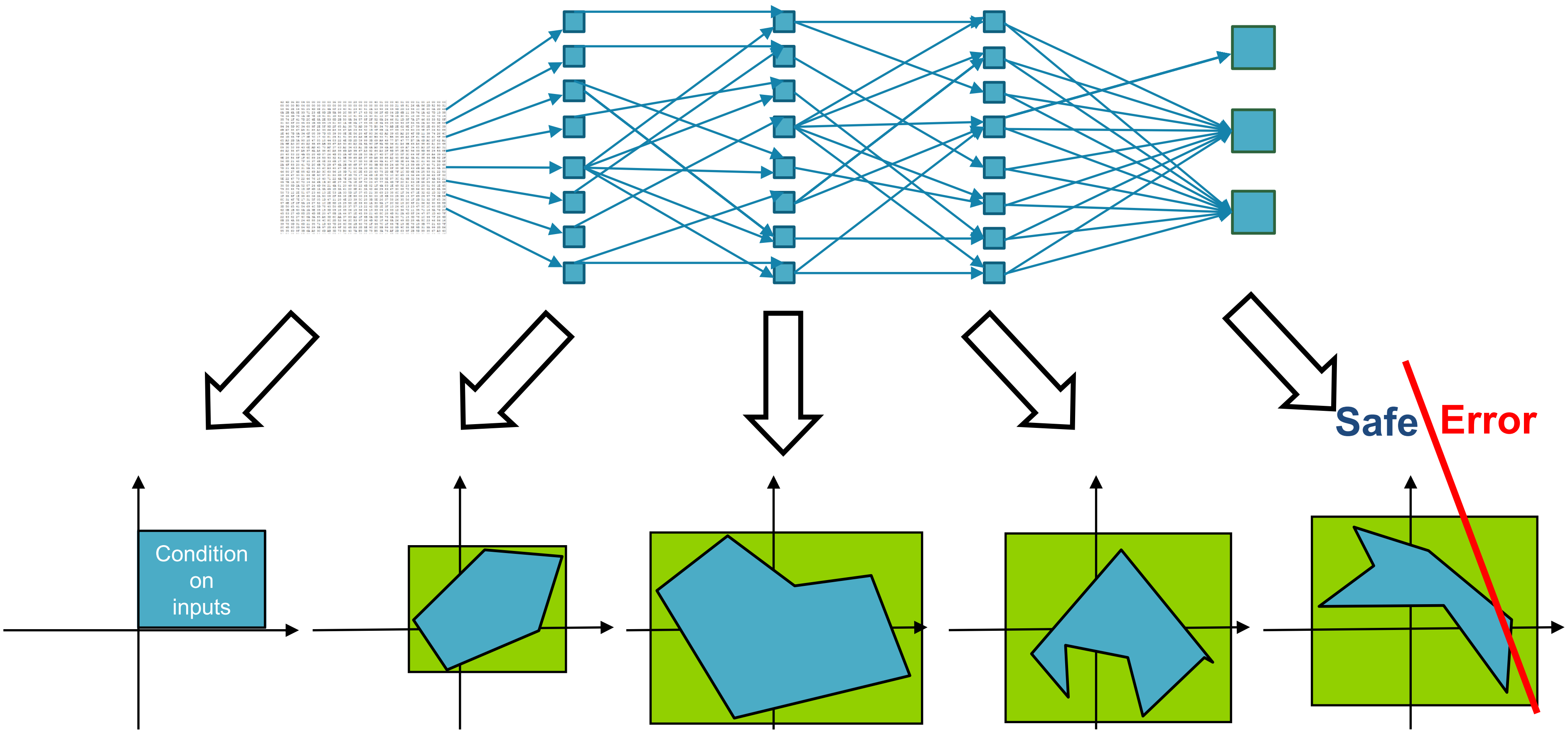
# Neural Network Verification

Is there an erroneous output?



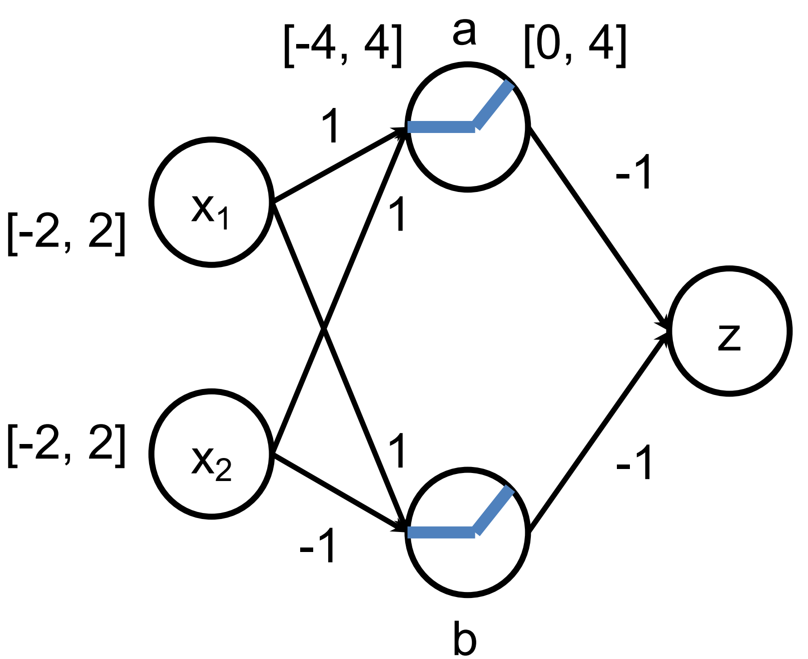
# Inteval Bound Propagation

Is there an erroneous output?




Axis aligned convex superset

# Example




$-2 \le x_1 \le 2$

$-2 \le x_2 \le 2$

$a_{in} = x_1 + x_2$

$a_{out} = \max\{a_{in}, 0\}$

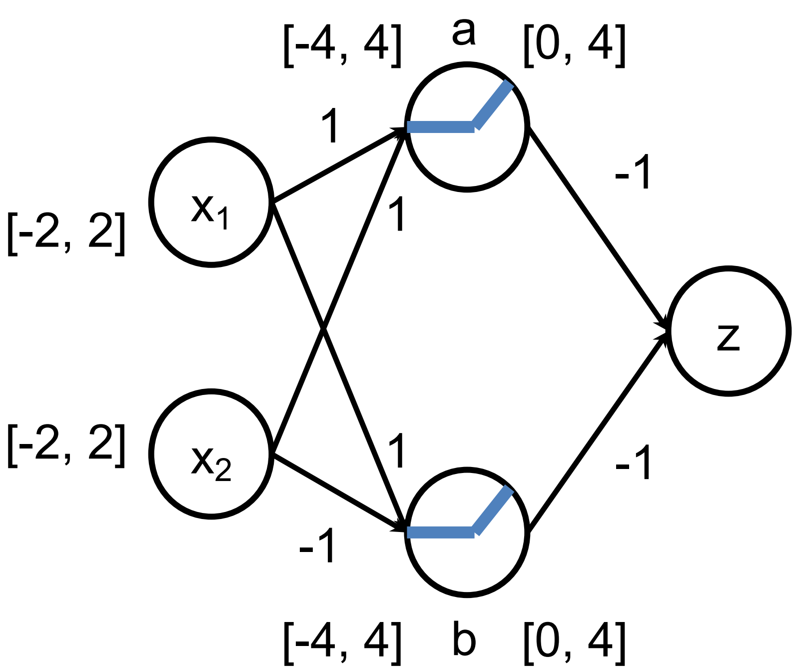
Minimum value of $a_{in}$? -4

Maximum value of $a_{in}$? 4

Minimum value of $a_{out}$? 0

Maximum value of $a_{out}$? 4

# Example




$-2 \leq x_1 \leq 2$

$-2 \leq x_2 \leq 2$

$b_{in} = x_1 - x_2$

$b_{out} = \max\{b_{in}, 0\}$

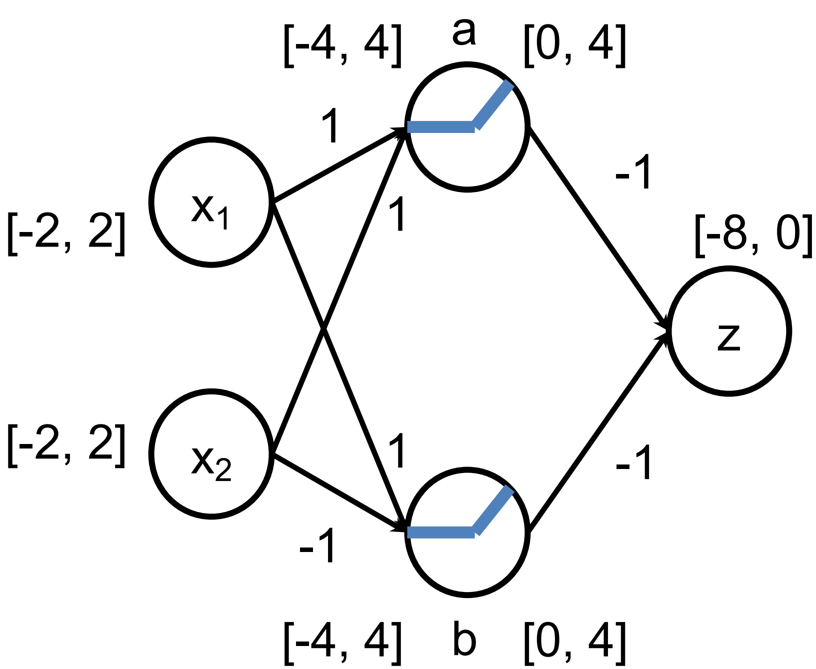
Minimum value of $b_{in}$? -4

Minimum value of $b_{out}$? 0

Maximum value of $b_{in}$? 4

Maximum value of $b_{out}$? 4

# Example




$-2 \leq x_1 \leq 2$

$-2 \leq x_2 \leq 2$

$b_{in} = x_1 - x_2$

$b_{out} = \max\{b_{in}, 0\}$

$z = -a_{out} - b_{out}$

Minimum value of z? -8

Maximum value of z? 0

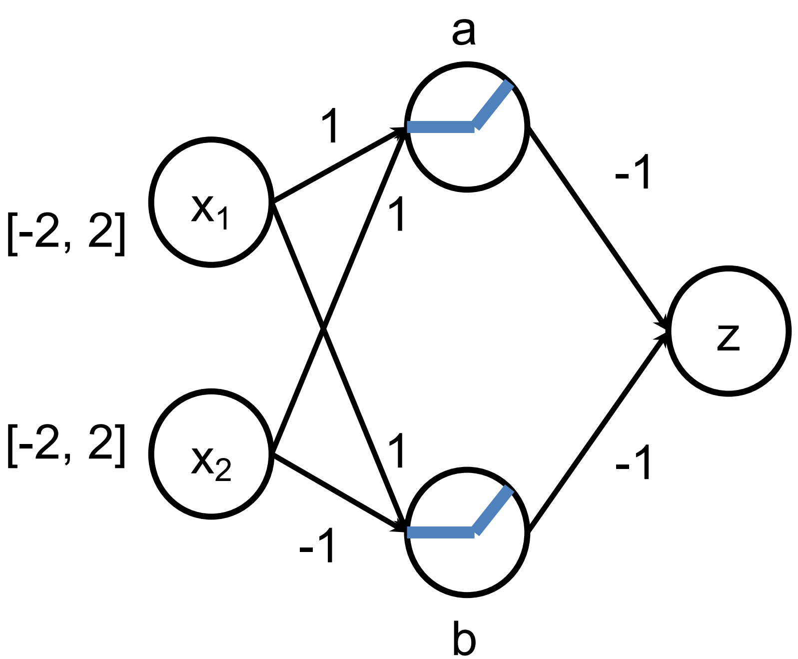
# Outline

- Incomplete Verification
  - Overview
  - Example: Interval Bound Propagation
  - **Example: Linear Programming Relaxation**

- Complete Verification
  - Branch and Bound
  - Application to verification

Ehlers 2017; Wong and Kolter, 2018

# Example




$$\min \quad z$$

$$\text{s.t.} \quad -2 \le x_1 \le 2$$

$$-2 \le x_2 \le 2$$

$$a_{in} = x_1 + x_2$$

$$b_{in} = x_1 - x_2$$

$$a_{out} = \max\{a_{in}, 0\}$$

$$b_{out} = \max\{b_{in}, 0\}$$

$$z = - a_{out} - b_{out}$$

# Example

$$\min \quad z$$

**Linear constraints**

Easy to handle

$$\text{s.t.} \quad -2 \leq x_1 \leq 2$$

$$-2 \leq x_2 \leq 2$$

$$a_{in} = x_1 + x_2$$

$$b_{in} = x_1 - x_2$$

$$a_{out} = \max\{a_{in},0\}$$

$$b_{out} = \max\{b_{in},0\}$$

$$z = - a_{out} - b_{out}$$

# Example

$$\min \quad z$$

$$\text{s.t.} \quad -2 \leq x_1 \leq 2$$

$$-2 \leq x_2 \leq 2$$

$$a_{in} = x_1 + x_2$$

$$b_{in} = x_1 - x_2$$

$$\mathbf{a_{out} = max\{a_{in},0\}}$$

**Non-linear constraints**

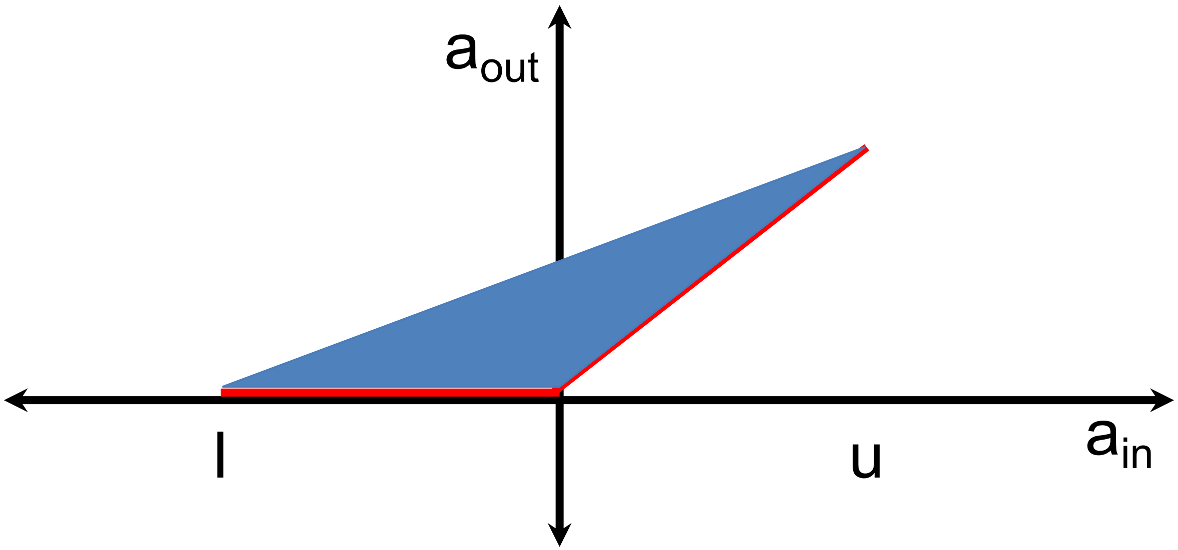
$$\mathbf{b_{out} = max\{b_{in},0\}}$$

**NP-hard problem**

$$z = - a_{out} - b_{out}$$

# Relaxation

$a_{out} = \max\{a_{in}, 0\}$ $\qquad$ $a_{in} \in [l, u]$




Ehlers 2017

Replace with convex superset

# Example

$$
\begin{array}{ll}
\min & z \\
\text{s.t.} & -2 \leq x_1 \leq 2 \\
& -2 \leq x_2 \leq 2 \\
& a_{in} = x_1 + x_2 \\
& b_{in} = x_1 - x_2 \\
& \mathbf{a_{out} = \max\{a_{in}, 0\}} \\
& \mathbf{b_{out} = \max\{b_{in}, 0\}} \\
& z = - a_{out} - b_{out}
\end{array}
$$

# Example

**Linear Program**

Several **"efficient"** solvers

$$\min \quad z$$

$$\text{s.t.} \quad -2 \leq x_1 \leq 2$$

$$-2 \leq x_2 \leq 2$$

$$a_{in} = x_1 + x_2$$

$$b_{in} = x_1 - x_2$$

$$a_{out} \geq 0,\ a_{out} \geq a_{in},\ a_{out} \leq 0.5a_{in} + 2$$

$$b_{out} \geq 0,\ b_{out} \geq b_{in},\ b_{out} \leq 0.5b_{in} + 2$$

$$z = -\ a_{out} - b_{out}$$

# Outline

- Incomplete Verification
  - Overview
  - Example: Interval Bound Propagation
  - Example: Linear Programming Relaxation

- **Complete Verification**
  - Branch and Bound
  - Application to verification

# Neural Network Verification

Neural network f

Scalar output $z = f(\mathbf{x})$

E.g. in binary classification, $z = s(y^*;\mathbf{x}) - s(y;\mathbf{x})$ for $y \neq y^*$

Property: $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in X$

Complete methods try to disprove the property

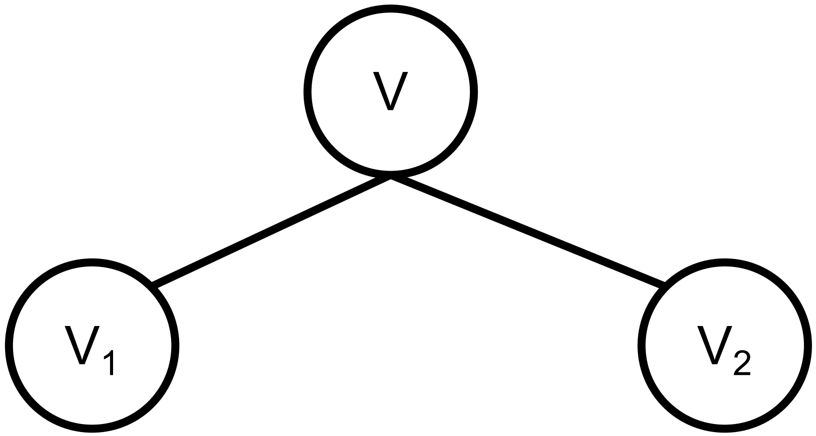
# Outline

- Incomplete Verification
  - Overview
  - Example: Interval Bound Propagation
  - Example: Linear Programming Relaxation

- Complete Verification
  - **Branch and Bound**
  - Application to verification

# Branch and Bound

Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$

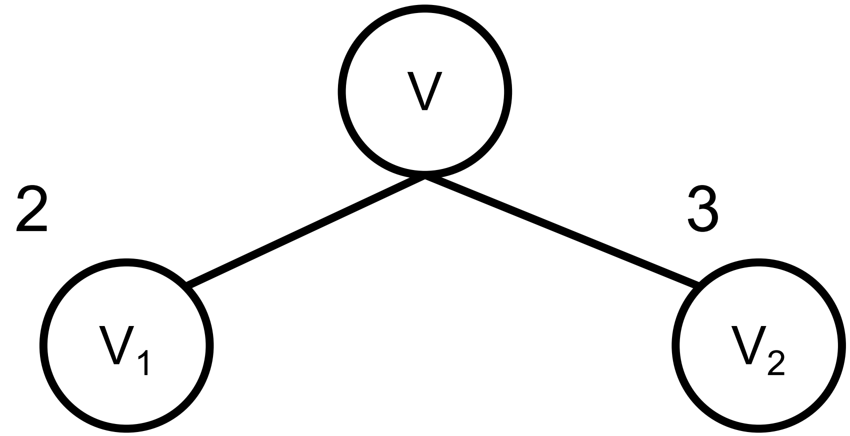



BRANCH: Split the feasible set

2 or more usually disjoint subsets

# Branch and Bound

Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$

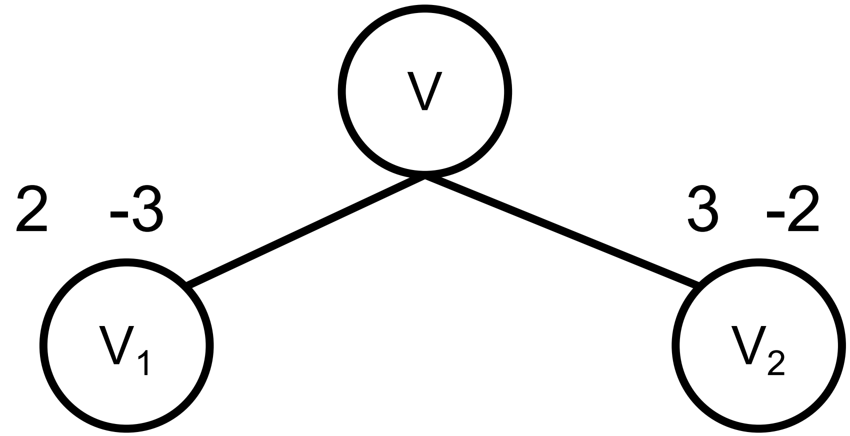



BOUND: Compute upper bounds for each branch

$h(\mathbf{v})$ for any feasible $\mathbf{v}$ (adversarial attacks)

# Branch and Bound

Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$

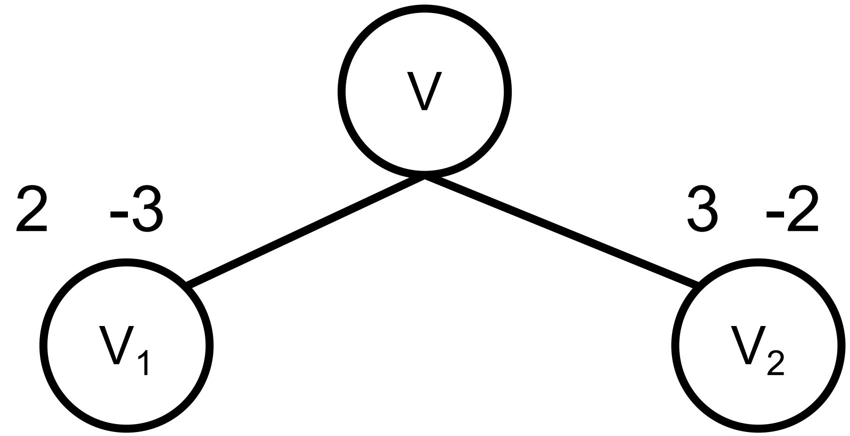



BOUND: Compute lower bounds for each branch

Convex relaxations (incomplete methods)

# Branch and Bound

Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$

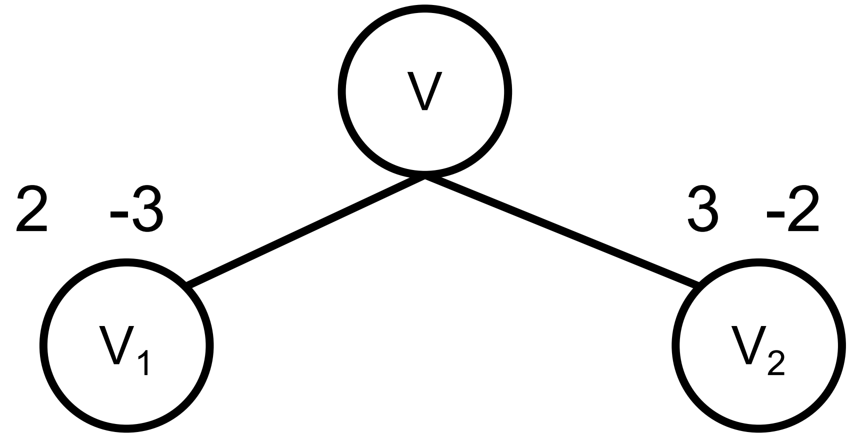



PRUNE: Any lower bounds greater than 0?

NO

# Branch and Bound

Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$

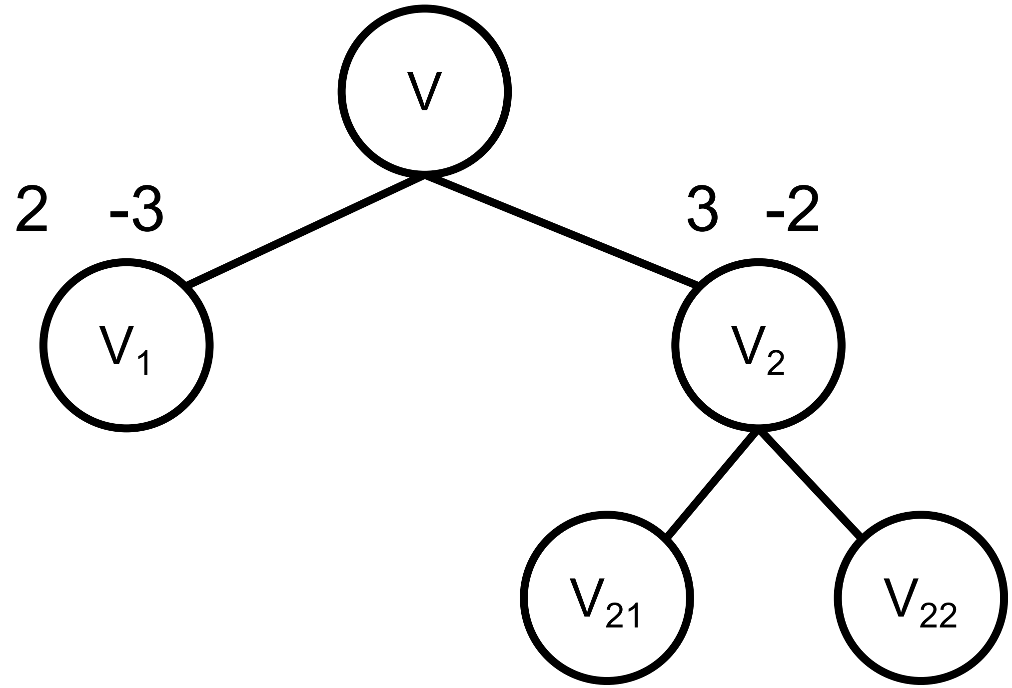



SELECT: Choose a subproblem

Say, we choose $V_2$

# Branch and Bound

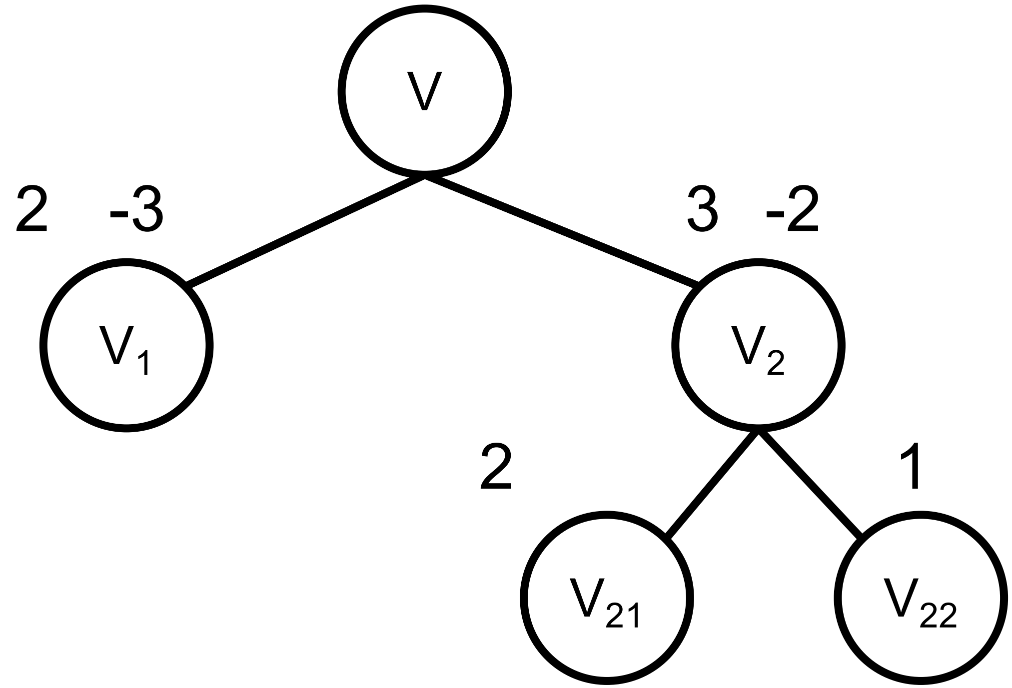
Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$




BRANCH: Split the feasible set

# Branch and Bound

Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$

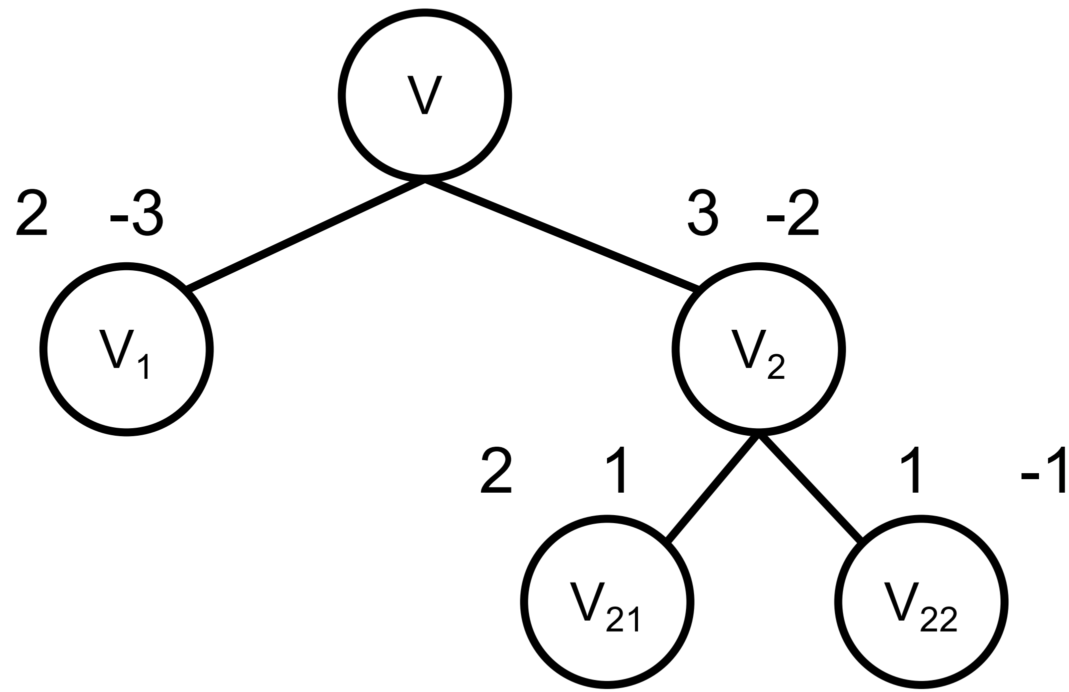



BOUND: Compute upper bounds

Upper bounds of children are smaller than the parent

# Branch and Bound

Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$

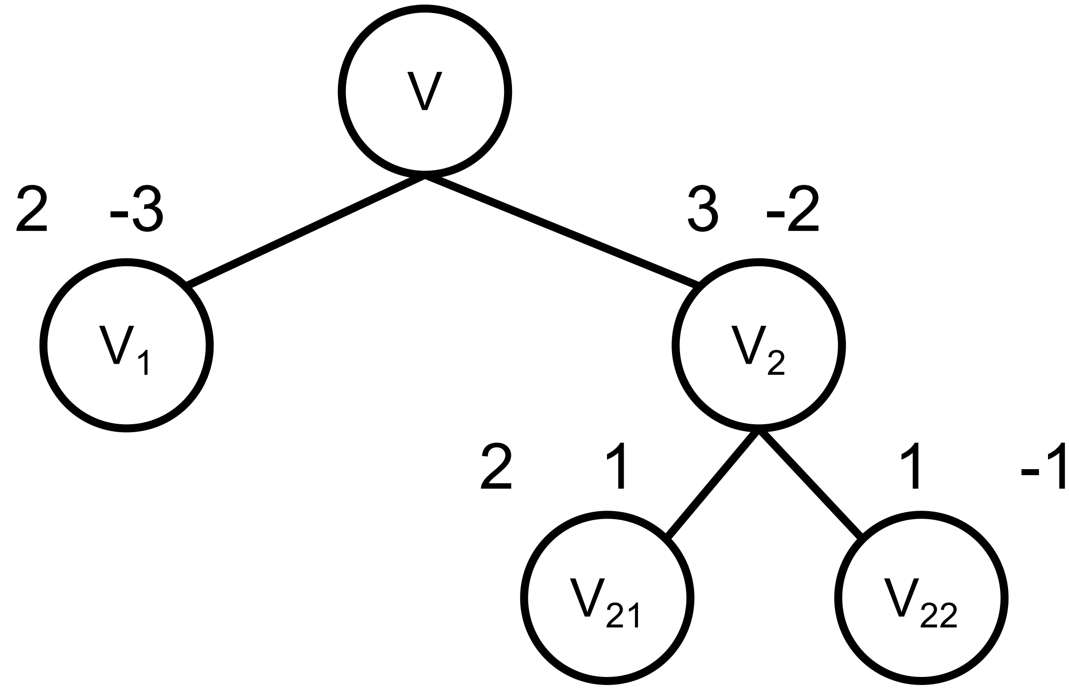



BOUND: Compute lower bounds

Lower bounds of children are greater than the parent

# Branch and Bound

Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$

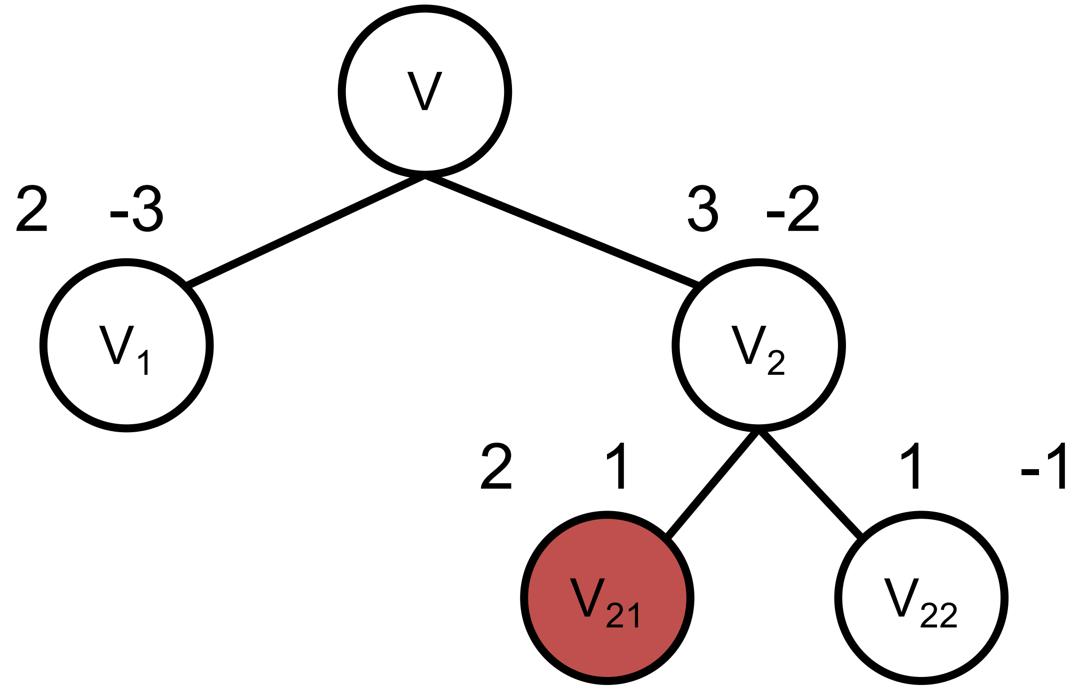



PRUNE: Any lower bounds greater than 0?

YES

# Branch and Bound

Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$

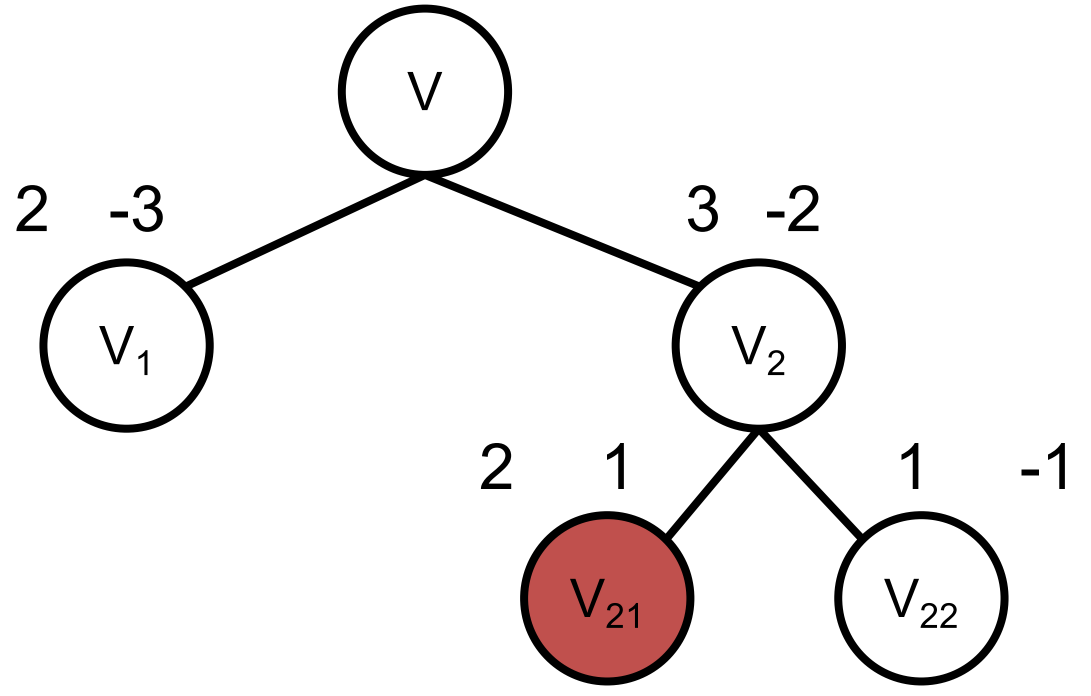



PRUNE: Any lower bounds greater than 0?

YES

# Branch and Bound

Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$

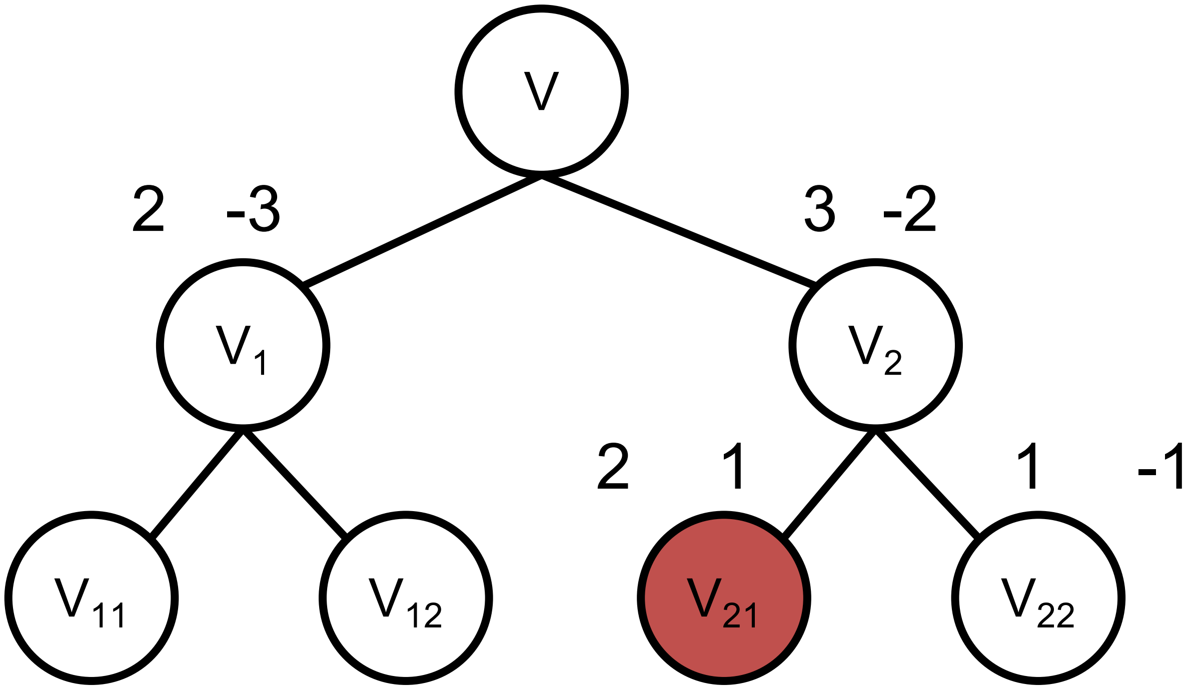



SELECT: Choose a subproblem

Say, we choose $V_1$

# Branch and Bound

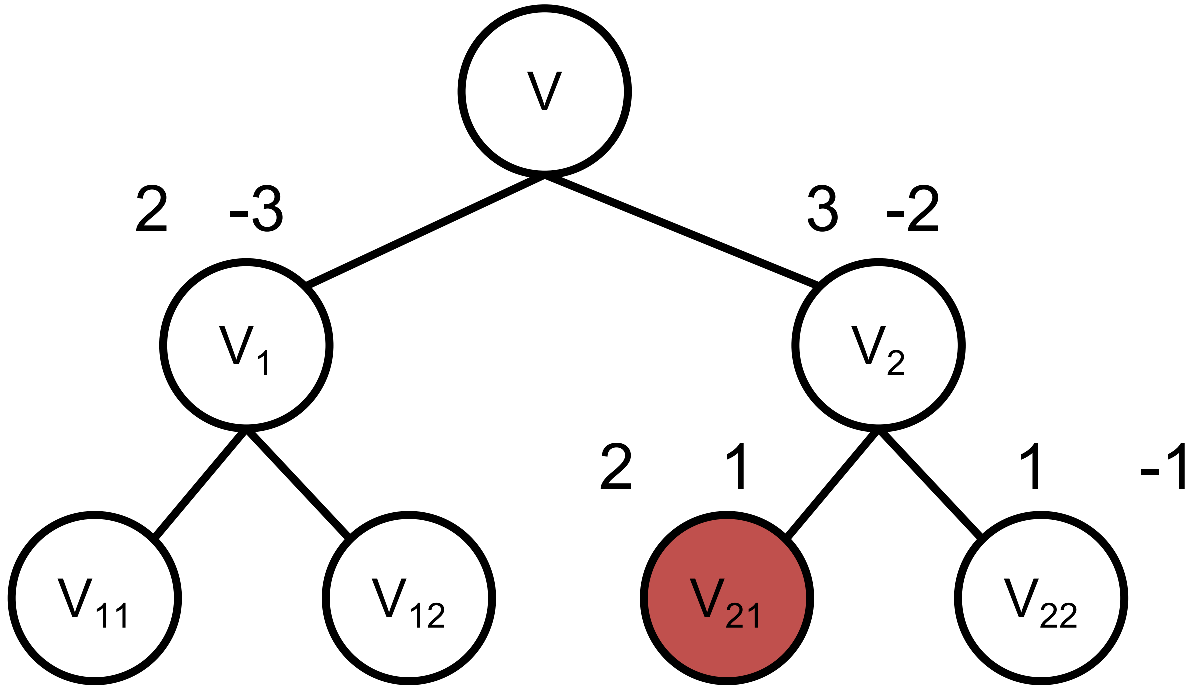
Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$




BRANCH: Split the feasible set

# Termination – Case I

Find **v** ∈ V such that h(**v**) ≤ 0

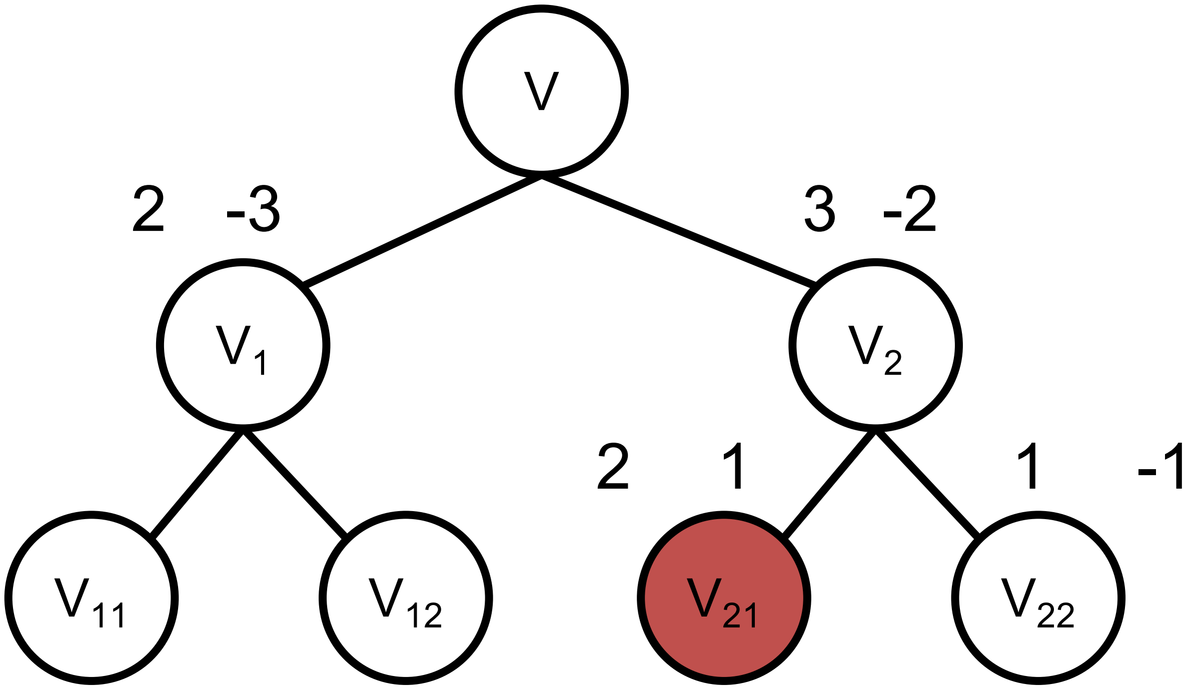



We find a counter-example

An upper bound that is less than 0

# Termination – Case II

Find $\mathbf{v} \in V$ such that $h(\mathbf{v}) \leq 0$




We prove there does not exist $\mathbf{v} \in V$ s.t. $h(\mathbf{v}) \leq 0$

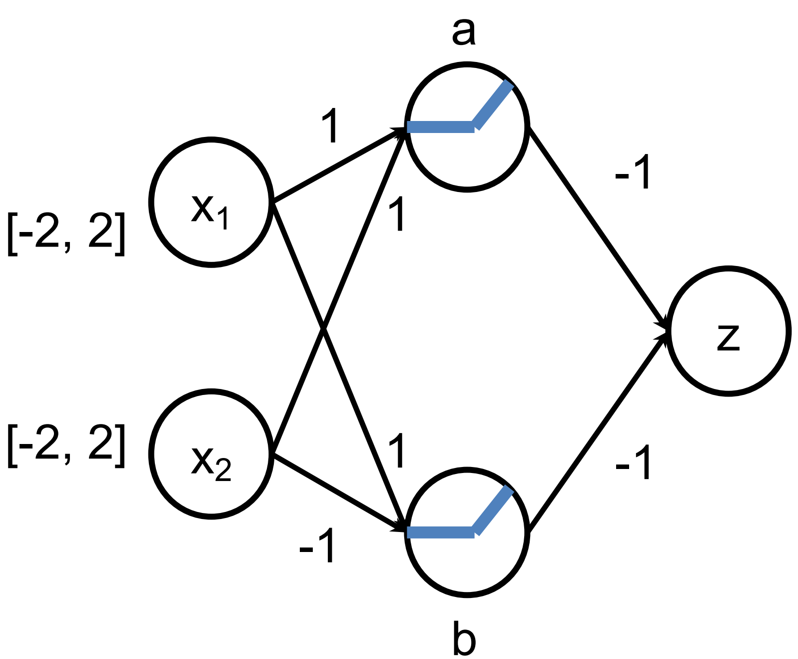
All leaf nodes have lower bound > 0

# Outline

- Incomplete Verification
  - Overview
  - Example: Interval Bound Propagation
  - Example: Linear Programming Relaxation

- Complete Verification
  - Branch and Bound
  - **Application to verification**

# Example




Prove that z > -5

$-2 \le x_1 \le 2$

$-2 \le x_2 \le 2$

$a_{in} = x_1 + x_2$

$b_{in} = x_1 - x_2$

$a_{out} = \max\{a_{in}, 0\}$

$b_{out} = \max\{b_{in}, 0\}$

$z = -a_{out} - b_{out}$

# Bounding

$-2 \leq x_1 \leq 2$

$-2 \leq x_2 \leq 2$

$a_{in} = x_1 + x_2$

$b_{in} = x_1 - x_2$

$a_{out} = \max\{a_{in},0\}$

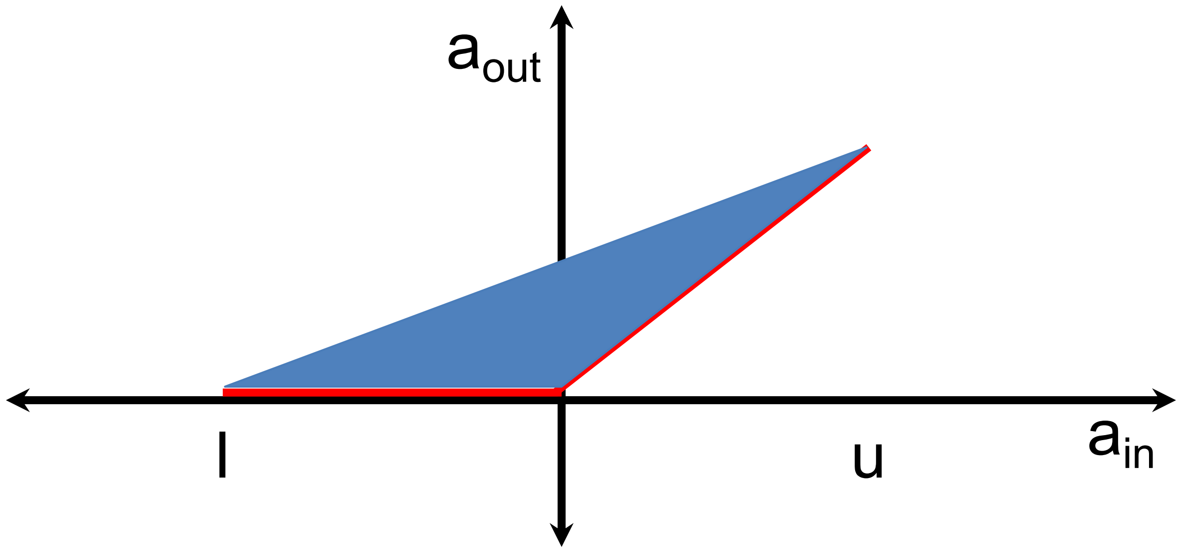
$b_{out} = \max\{b_{in},0\}$

$z = - a_{out} - b_{out}$

Relax all non-linearities

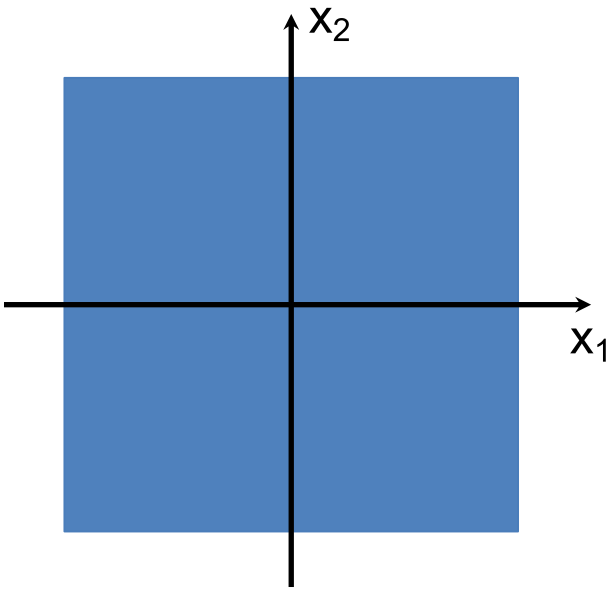
# Relaxation

$a_{out} = \max\{a_{in}, 0\}$ $\quad\quad$ $a_{in} \in [l,u]$




Replace with convex superset

# Bounding




$-2 \leq x_1 \leq 2$

$-2 \leq x_2 \leq 2$

$a_{in} = x_1 + x_2$

$b_{in} = x_1 - x_2$

$a_{out} \geq a_{in}, a_{out} \geq 0, a_{out} \leq a_{in}/2+2$

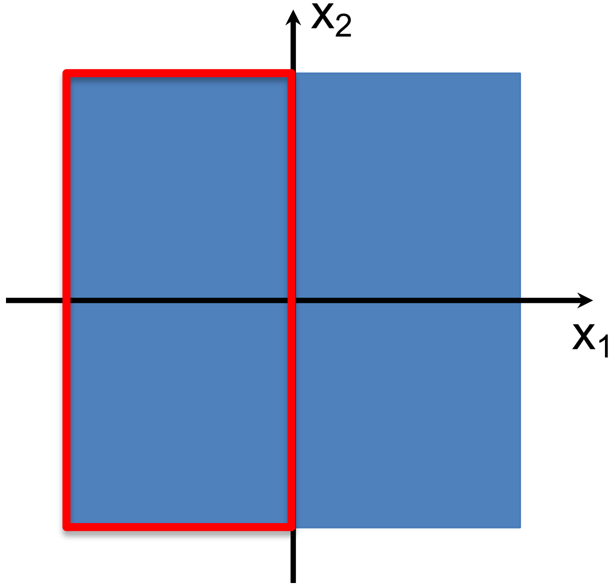
$b_{out} \geq b_{in}, b_{out} \geq 0, b_{out} \leq b_{in}/2+2$

$z = - a_{out} - b_{out}$

$z_{min} = -6$

min z

# Bounding




$-2 \leq x_1 \leq 2$

$-2 \leq x_2 \leq 2$

$a_{in} = x_1 + x_2$

$b_{in} = x_1 - x_2$

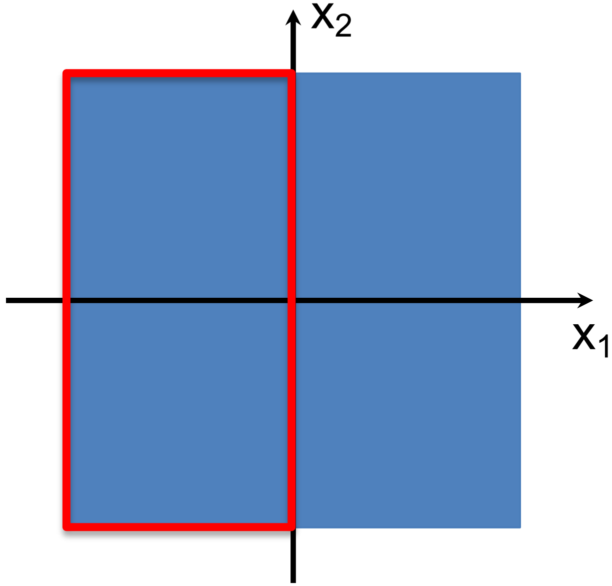
$a_{out} \geq a_{in}$, $a_{out} \geq 0$, $a_{out} \leq a_{in}/2+2$

$b_{out} \geq b_{in}$, $b_{out} \geq 0$, $b_{out} \leq b_{in}/2+2$

$z = - a_{out} - b_{out}$

$\min z$

# Bounding




$-2 \leq x_1 \leq 0$

$-2 \leq x_2 \leq 2$

$a_{in} = x_1 + x_2$

$b_{in} = x_1 - x_2$

$a_{out} \geq a_{in}, a_{out} \geq 0, a_{out} \leq a_{in}/3+4/3$

$b_{out} \geq b_{in}, b_{out} \geq 0, b_{out} \leq b_{in}/3+4/3$

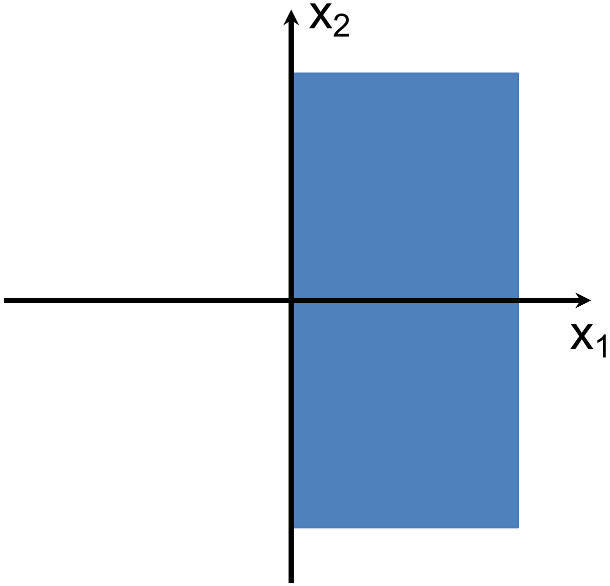
Prune away

$z = - a_{out} - b_{out}$

$z_{min} = -2.66$

min z

# Bounding




$0 \le x_1 \le 2$

$-2 \le x_2 \le 2$

$a_{in} = x_1 + x_2$

$b_{in} = x_1 - x_2$

$a_{out} \ge a_{in}$, $a_{out} \ge 0$, $a_{out} \le 2a_{in}/3+4/3$

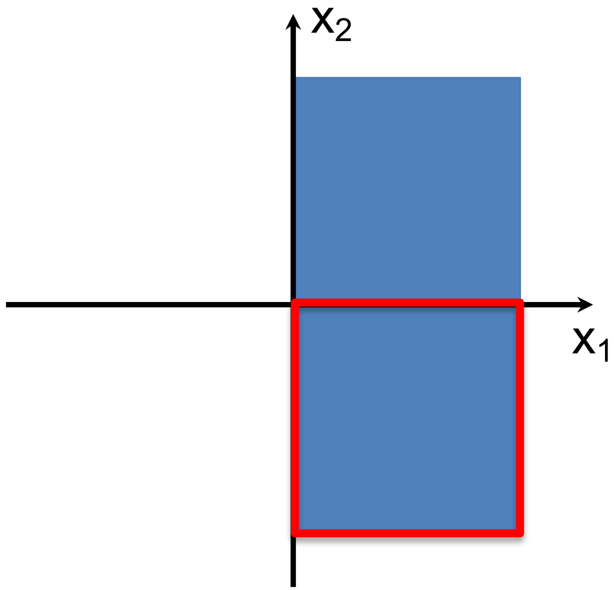
$b_{out} \ge b_{in}$, $b_{out} \ge 0$, $b_{out} \le 2b_{in}/3+4/3$

$z = - a_{out} - b_{out}$

$z_{min} = -5.33$

min $z$

# Bounding




$0 \leq x_1 \leq 2$

$-2 \leq x_2 \leq 2$

$a_{in} = x_1 + x_2$

$b_{in} = x_1 - x_2$

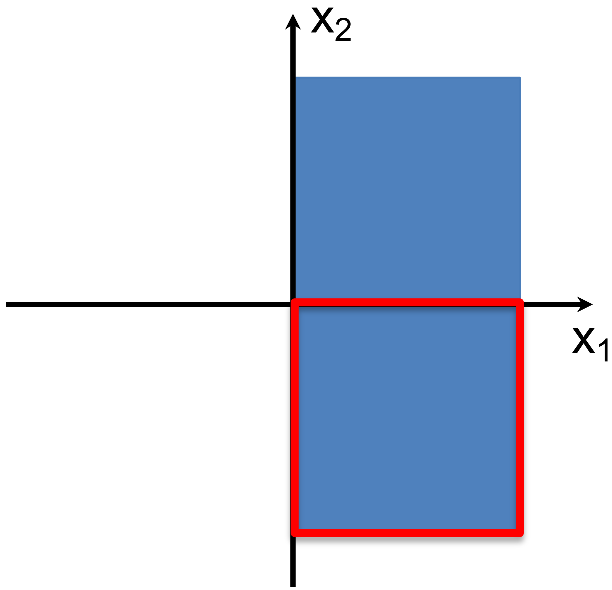
$a_{out} \geq a_{in}$, $a_{out} \geq 0$, $a_{out} \leq 2a_{in}/3+4/3$

$b_{out} \geq b_{in}$, $b_{out} \geq 0$, $b_{out} \leq 2b_{in}/3+4/3$

$z = - a_{out} - b_{out}$

min $z$

# Bounding




$0 \leq x_1 \leq 2$

$-2 \leq x_2 \leq 0$

$a_{in} = x_1 + x_2$

$b_{in} = x_1 - x_2$

$a_{out} \geq a_{in},\ a_{out} \geq 0,\ a_{out} \leq a_{in}/2+1$

$b_{out} \geq b_{in},\ b_{out} \geq 0,\ b_{out} \leq b_{in}$

$z = -\ a_{out} - b_{out}$

Continue until termination

$\min z$

# Branch and Bound

- Unified framework for complete verification
- Different bounds and bounding algorithms
  - Bound propagation (e.g. $\beta$-CROWN)
  - Tight LP relaxations (e.g. disjunctive programming)
  - Efficient solvers (e.g. Stagewise, Active sets)
- Different branching
  - Hand-designed heuristics (e.g. BaBSR)
  - Learning based heuristics (e.g. NN Branching)

# Questions?

Jax code for verification:
https://github.com/deepmind/jax_verify