

Reduced Products of Abstract Domains for Fairness Certification of Neural Networks

Denis Mazzucato and Caterina Urban
École Normale Supérieure | Inria

COVID Time Papers In Person
SPLASH 2022 - December 7th, 2022



The image shows a collage of news snippets and a Google Translate interface, all centered around the theme of machine learning's social impact.

Top News Snippets:

- WIRED** - [In 2019, predictive algorithms will start to make banking fair for all](#) (October 10, 2018)
- WIRED** - [Amazon scraps secret AI recruiting tool that showed bias against women](#) (March 25, 2019)
- The Telegraph** - [AI used for first time in job interviews in UK to find best applicants](#) (September 27, 2019)
- WIRED** - [The AI Doctor Will See You Now](#) (December 21, 2019)
- Google Translate** interface showing English to French translation of "A nurse" and "A doctor".

Center Article:

Social Impact of Machine Learning

nature NEWS · 24 OCTOBER 2019

UPDATE 26 OCTOBER 2019

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

Bottom Right Article:

Machine Bias

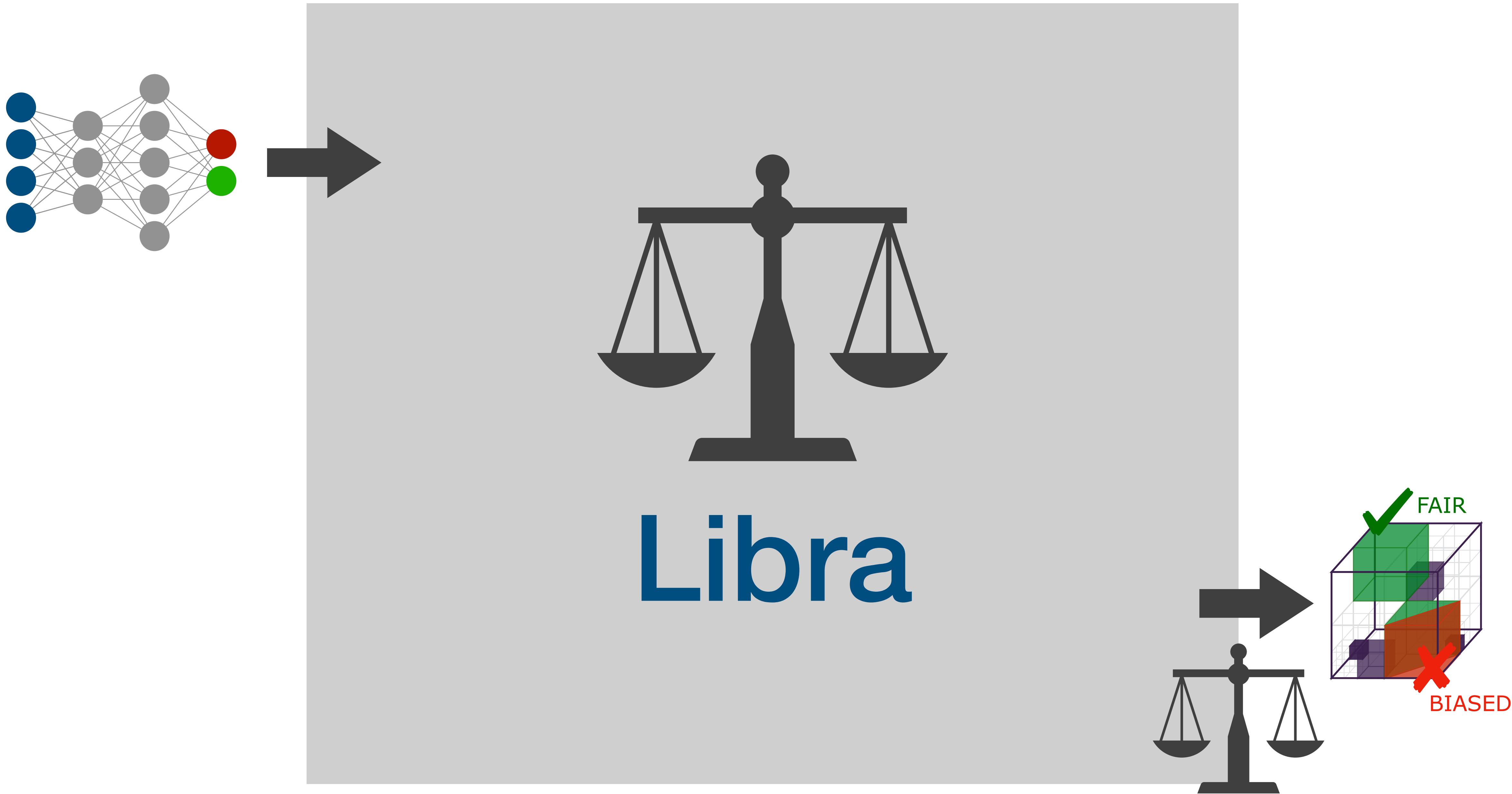
There's software used across the country to predict future criminals. And it's biased against blacks.

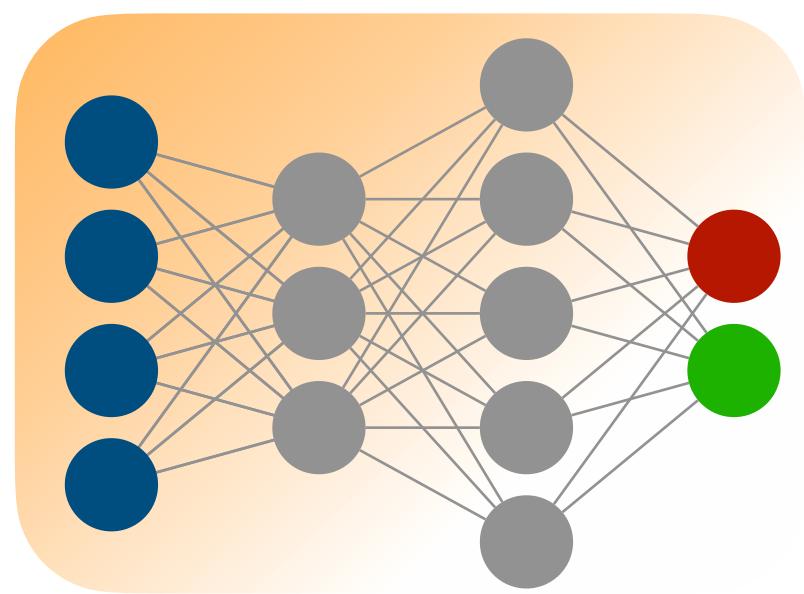
by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

Artificial Intelligence Act

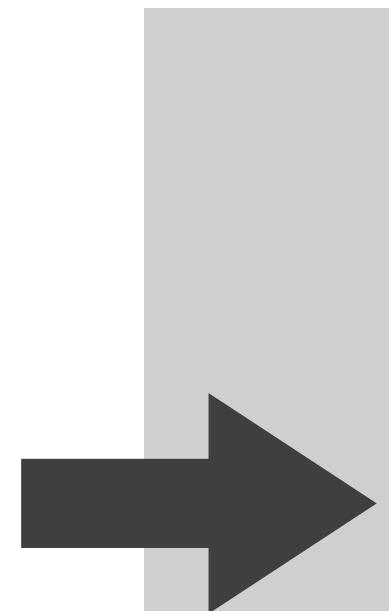
April 2021



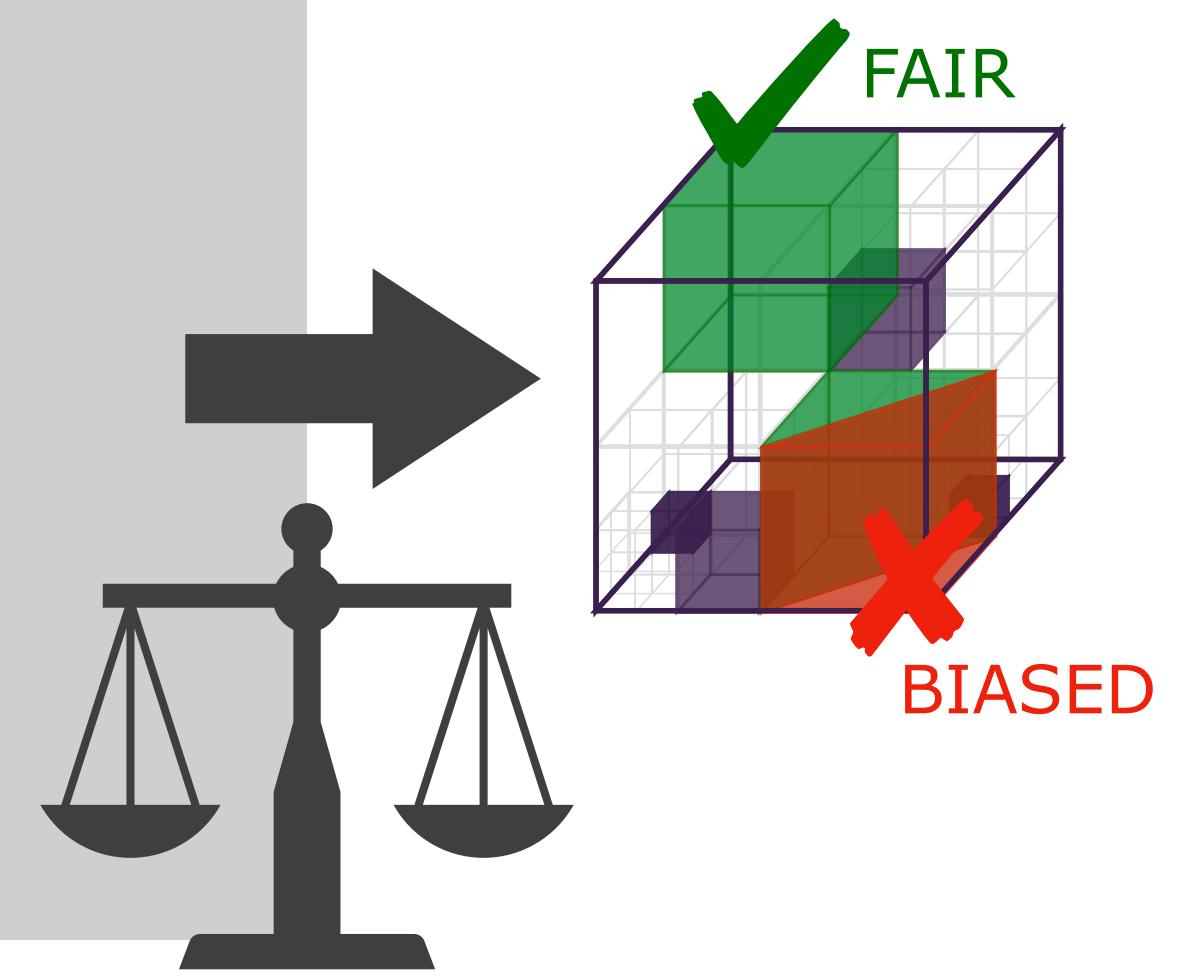




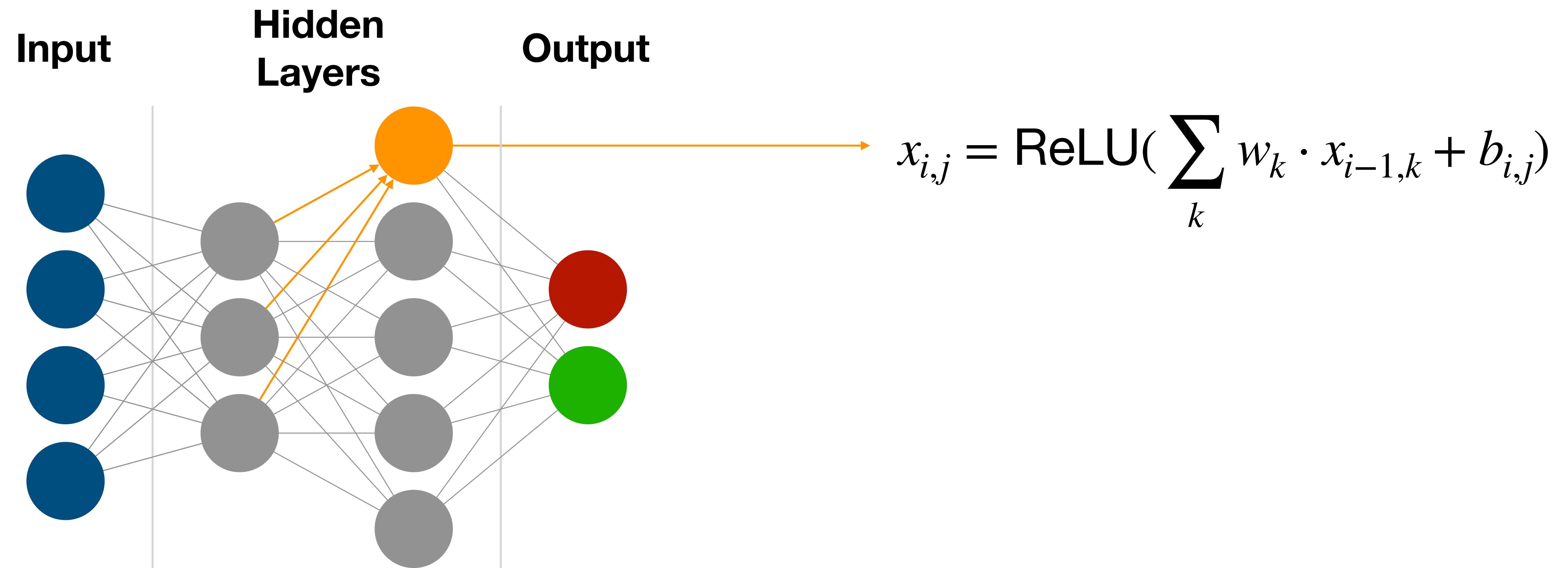
Neural Network



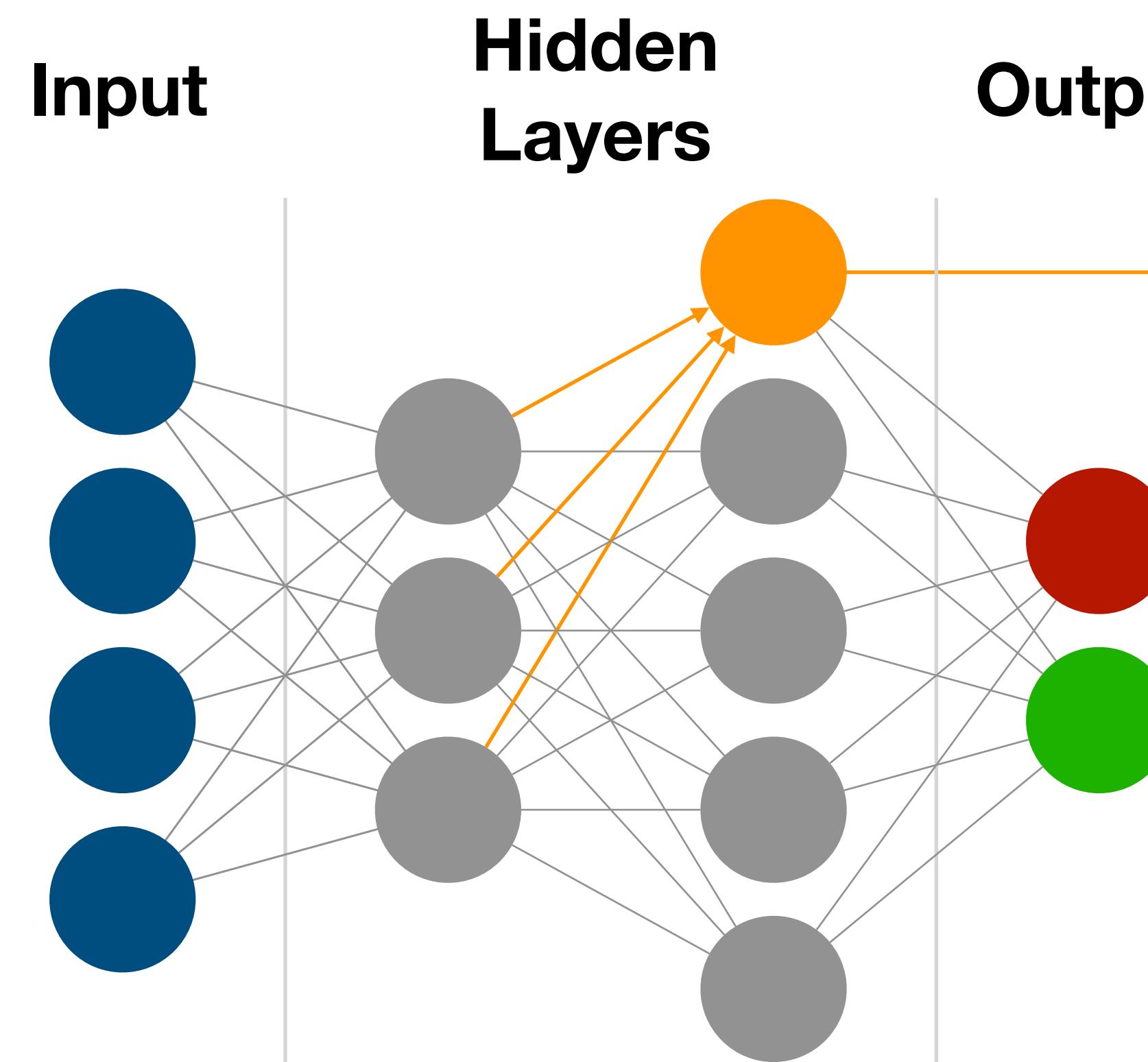
Libra



Feed-Forward Neural Networks with ReLU Activations

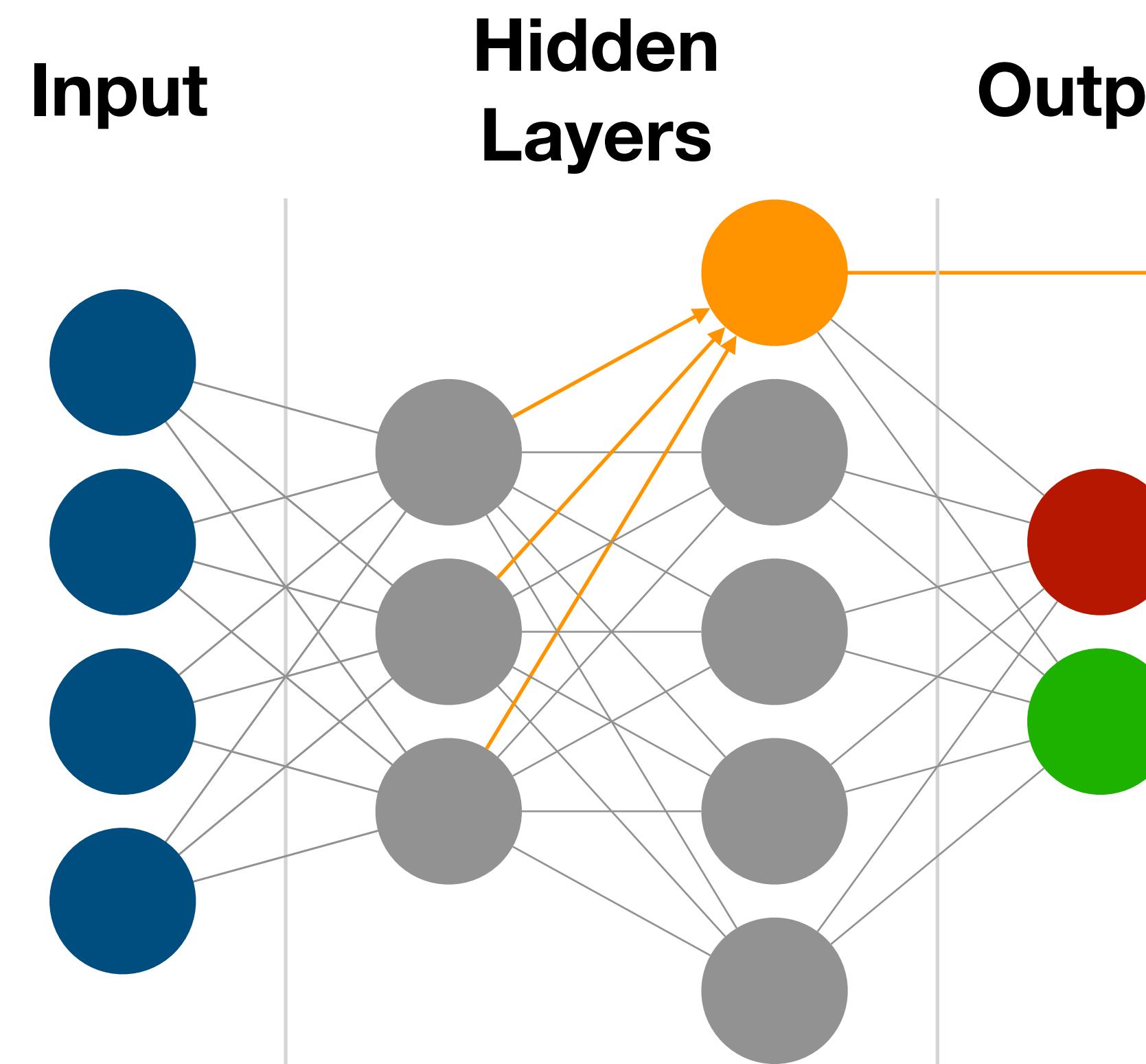


Feed-Forward Neural Networks with ReLU Activations



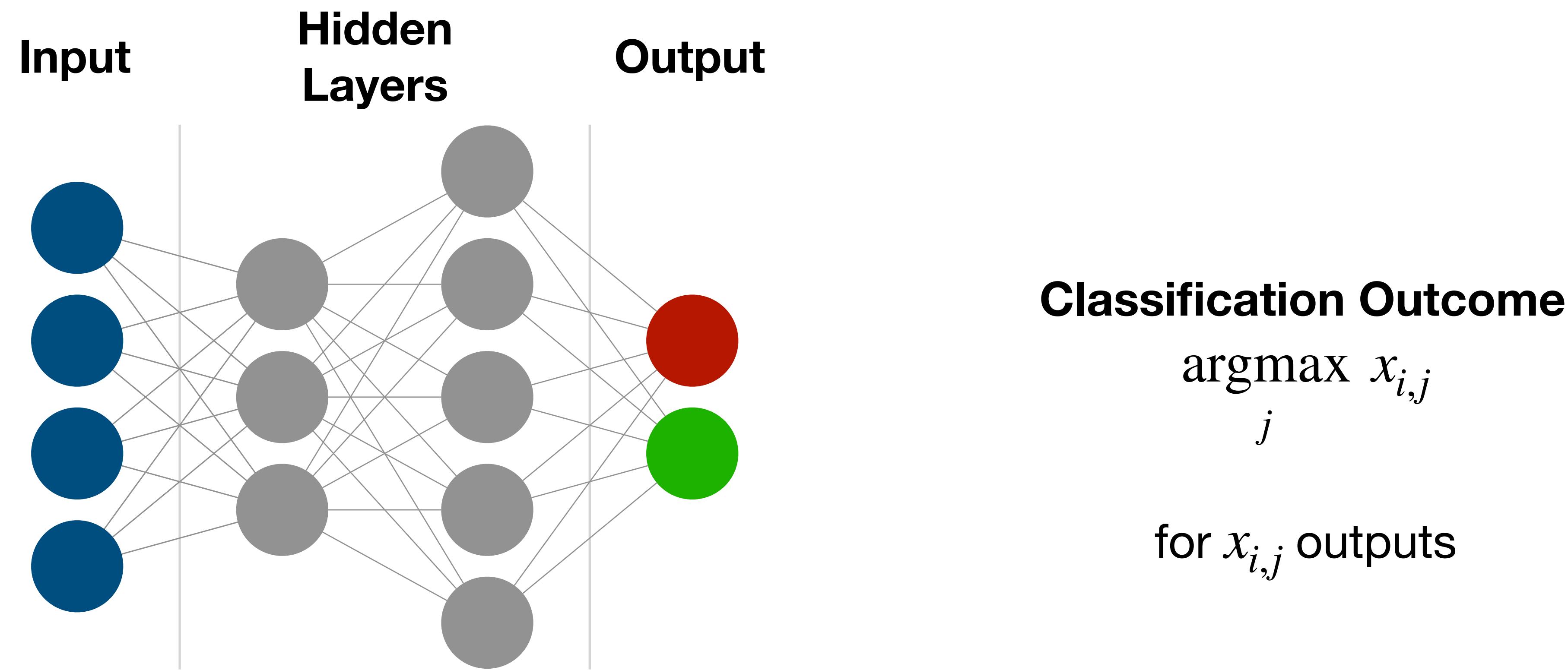
$$\begin{aligned}x_{i,j} &= \text{ReLU}\left(\sum_k w_k \cdot x_{i-1,k} + b_{i,j}\right) \\&= \max\left(\sum_k w_k \cdot x_{i-1,k} + b_{i,j}, 0\right)\end{aligned}$$

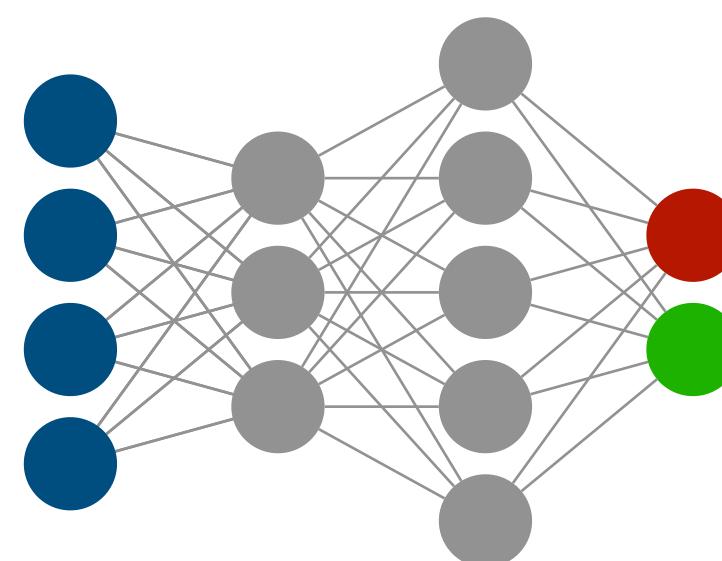
Feed-Forward Neural Networks with ReLU Activations



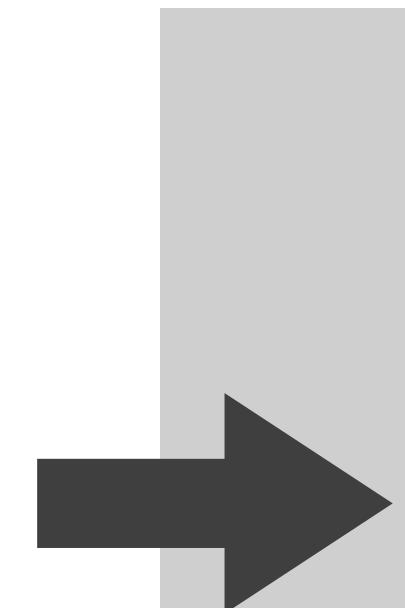
$$\begin{aligned}x_{i,j} &= \text{ReLU}\left(\sum_k w_k \cdot x_{i-1,k} + b_{i,j}\right) \\&= \max\left(\underbrace{\sum_k w_k \cdot x_{i-1,k} + b_{i,j}}_{\hat{x}_{i,j}}, 0\right) \\ \hat{x}_{i,j} &= \sum_k w_k \cdot x_{i-1,k} + b_{i,j}\end{aligned}$$

Feed-Forward Neural Networks with ReLU Activations

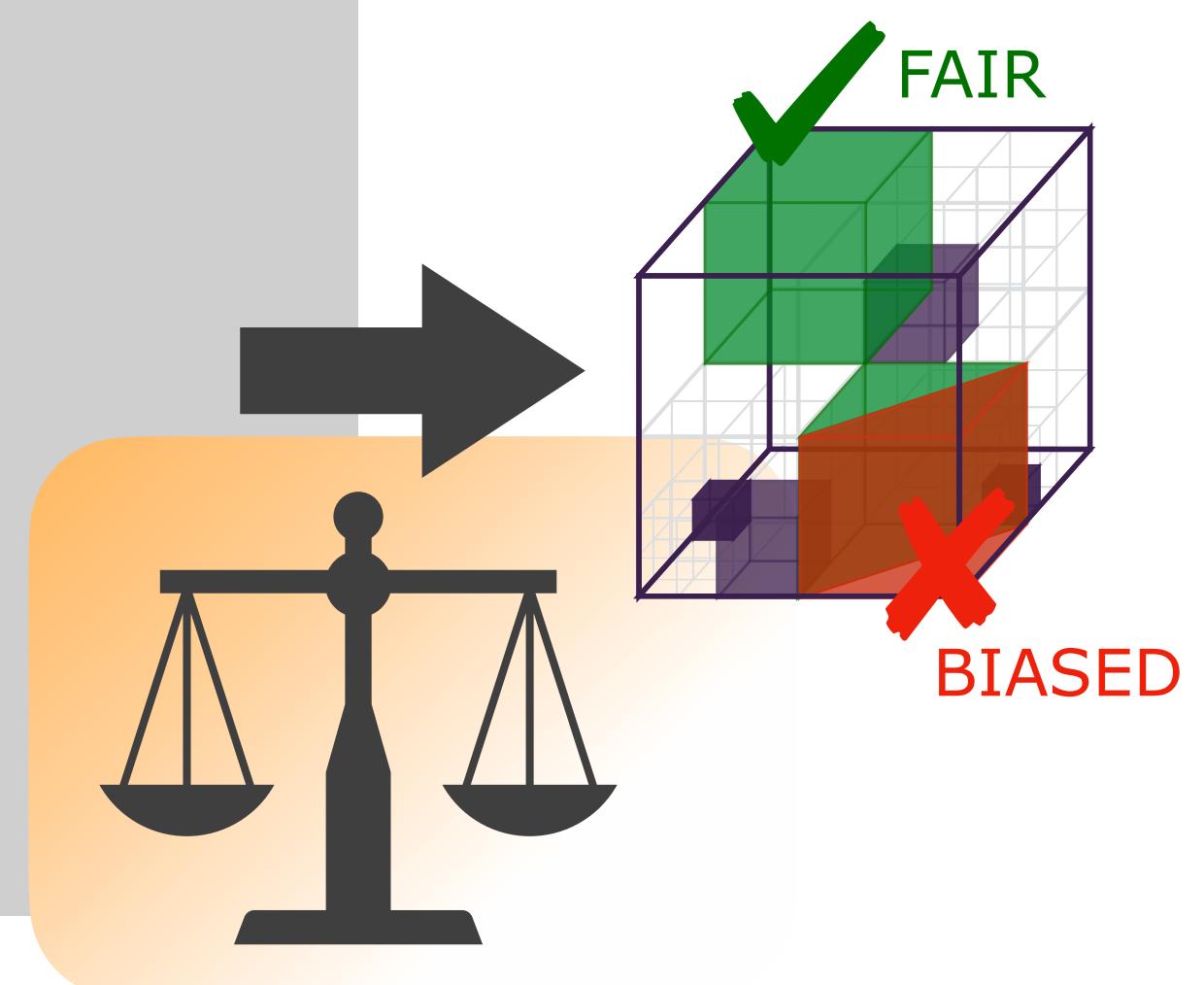




Neural Network



Libra



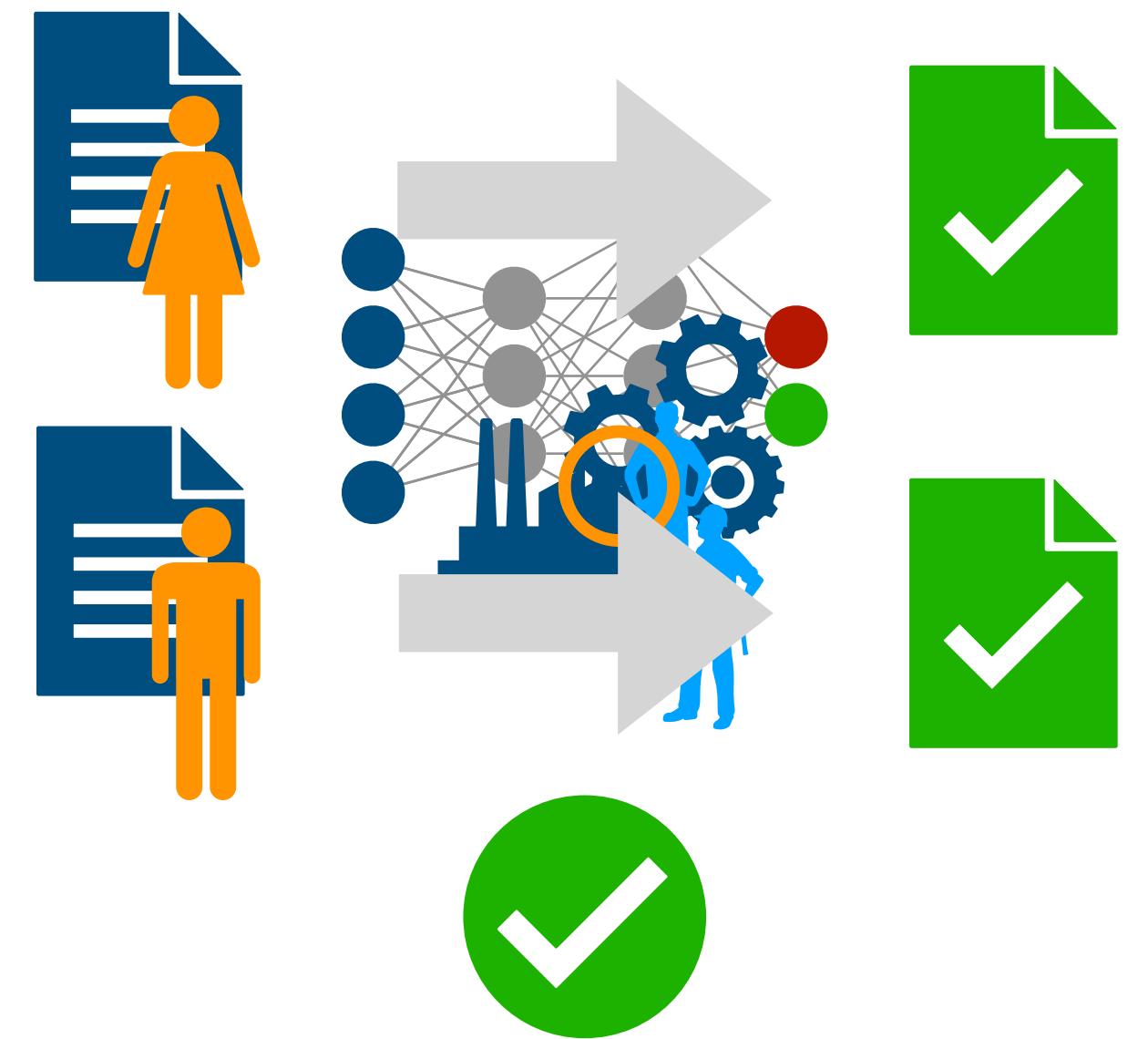
Dependency **Fairness**

The classification outcome is
Independent on the
Sensitive Features

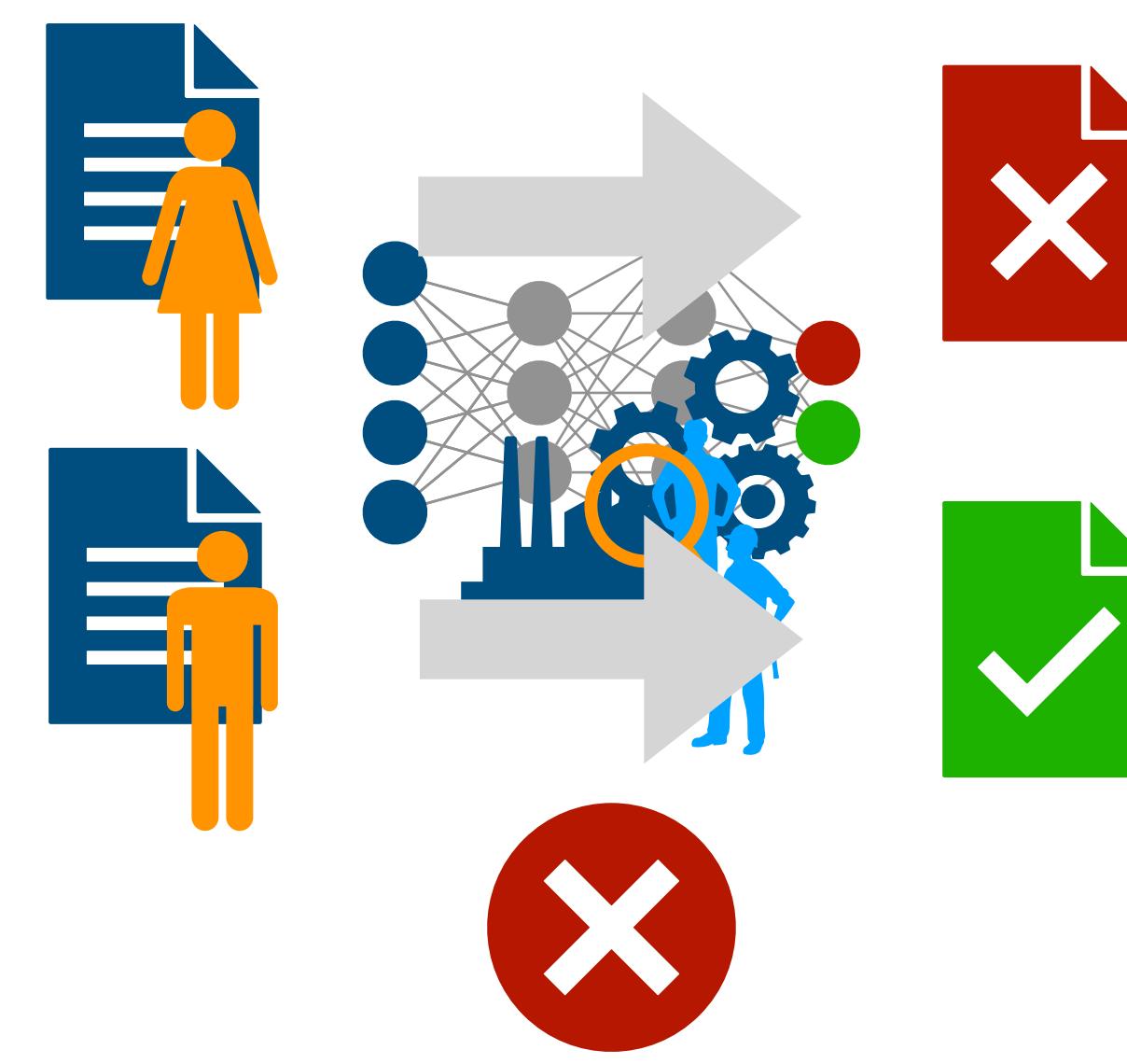
Recruiting Process



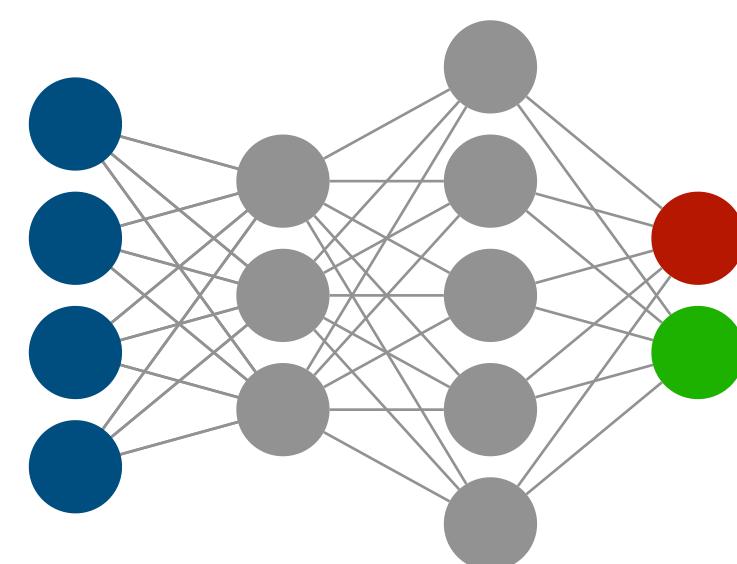
Recruiting Process



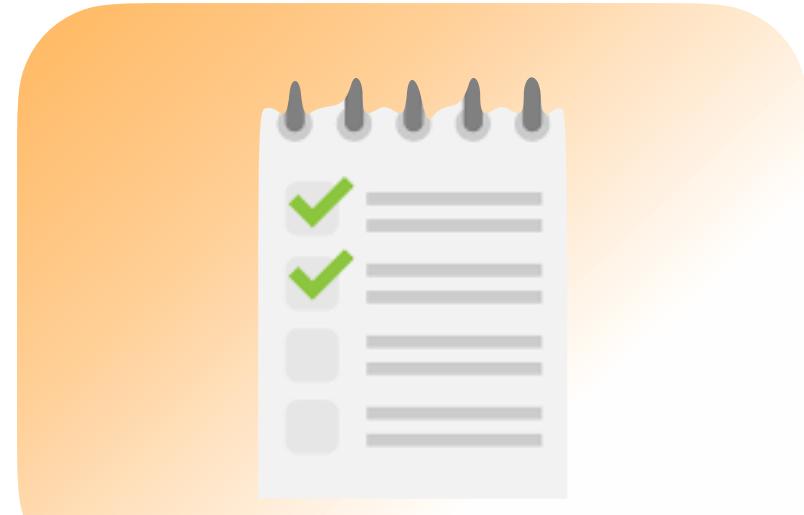
Fair



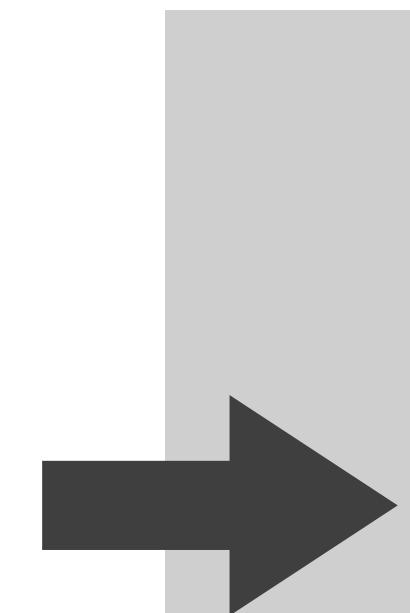
Unfair



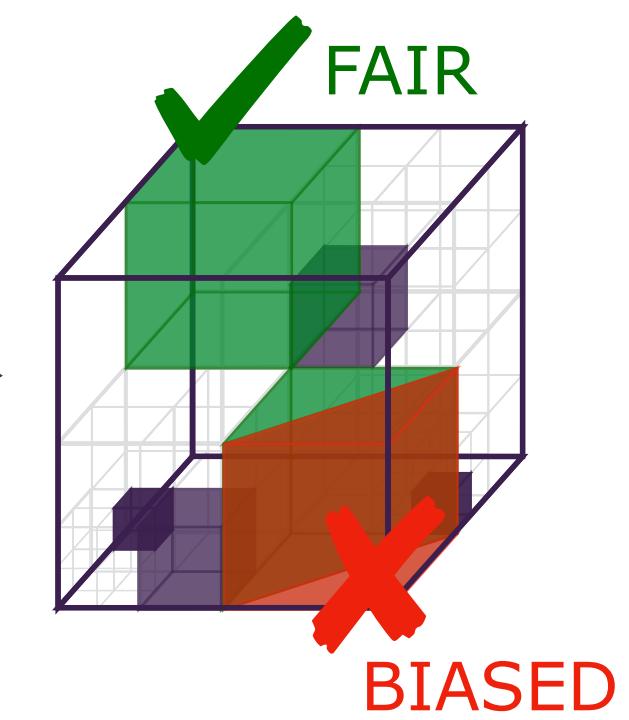
Neural Network



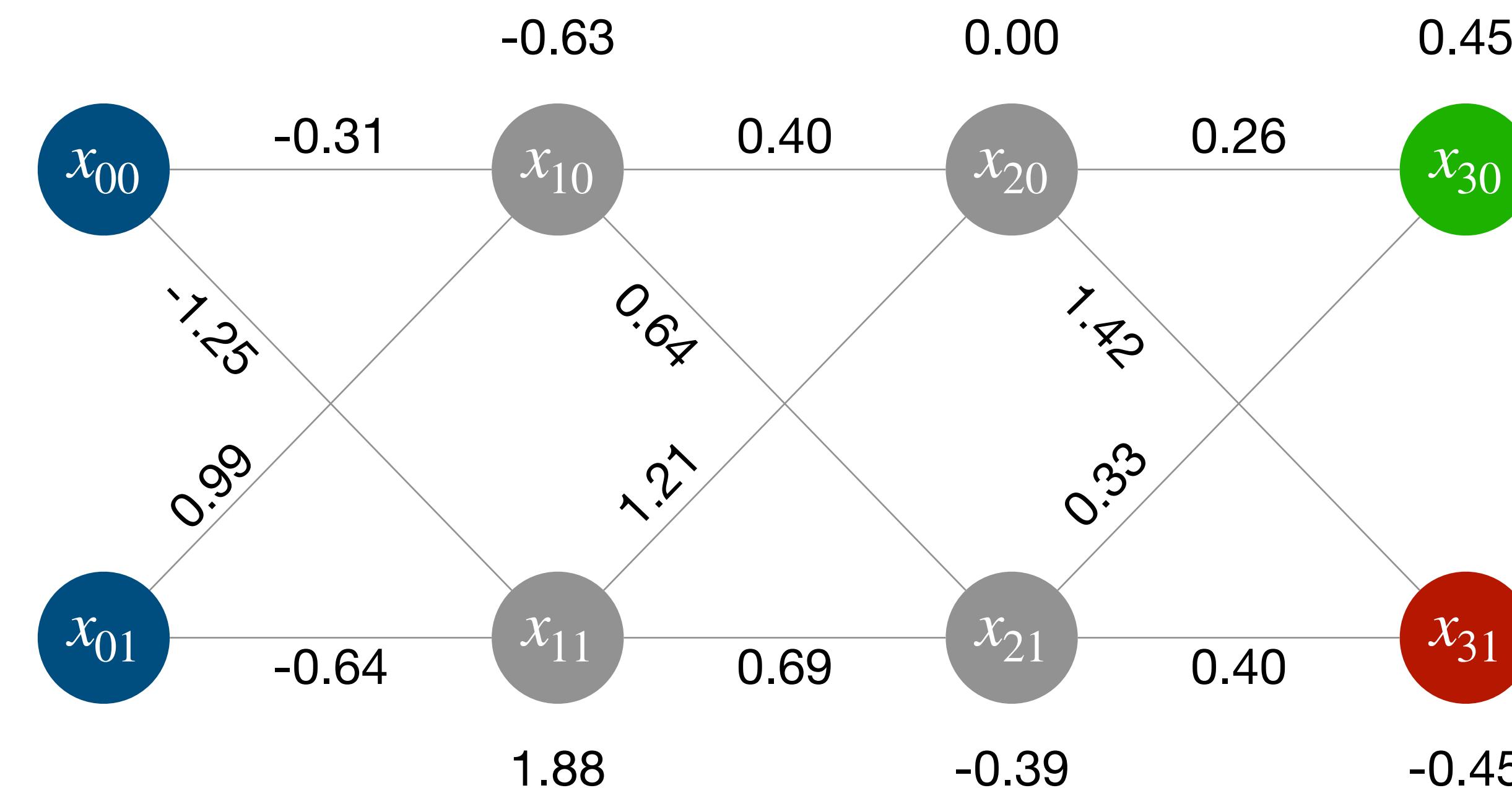
Specification



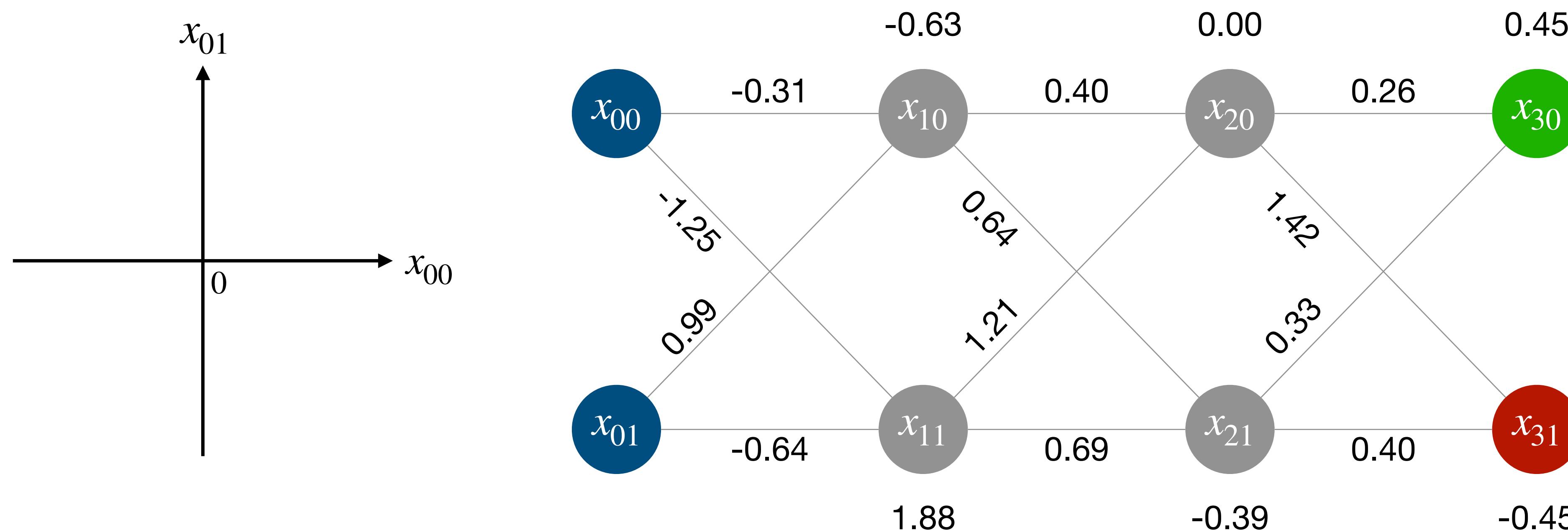
Libra



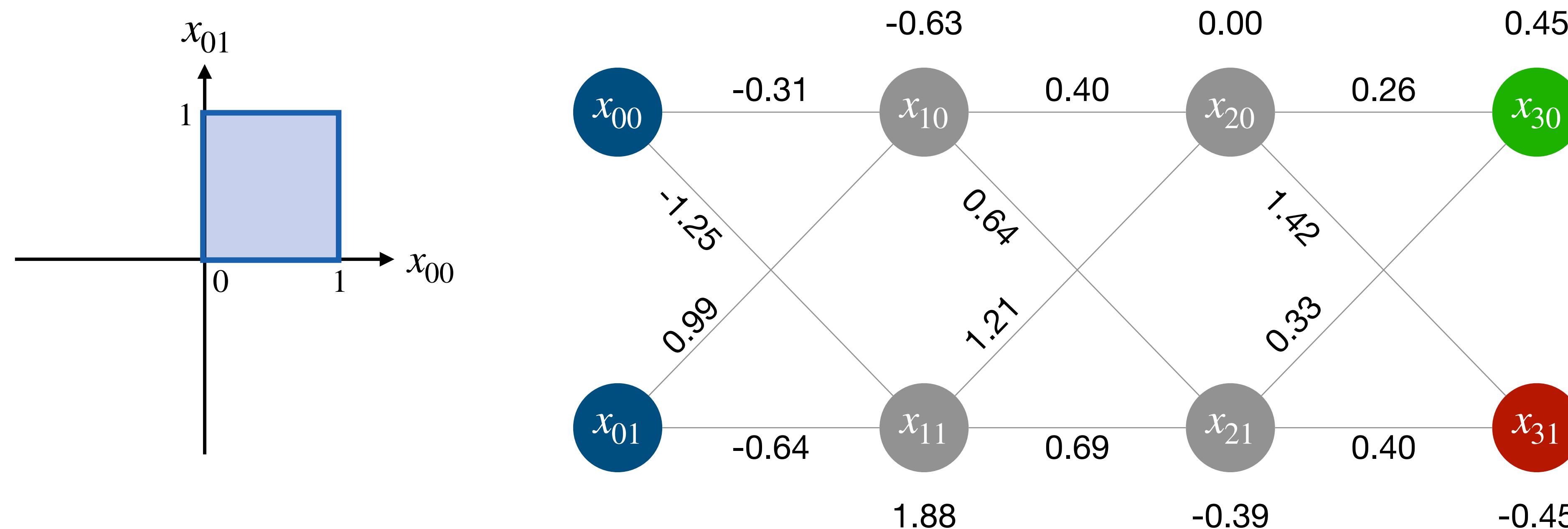
Specification



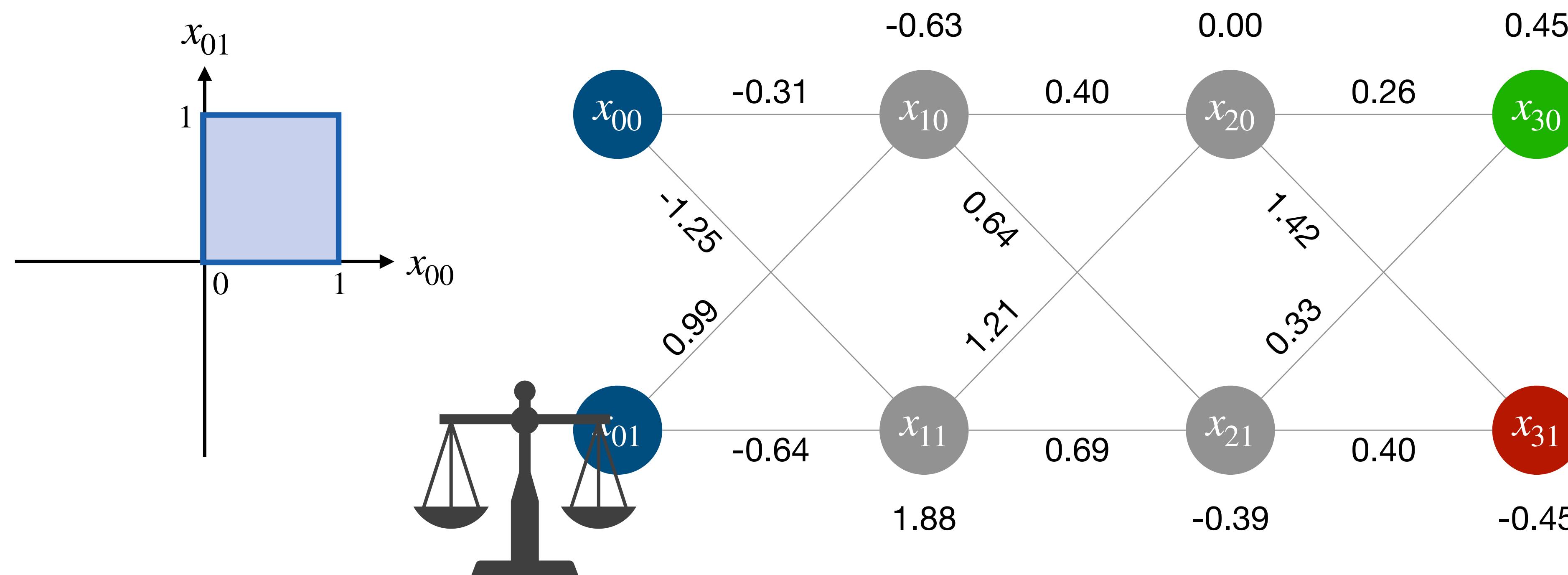
Specification

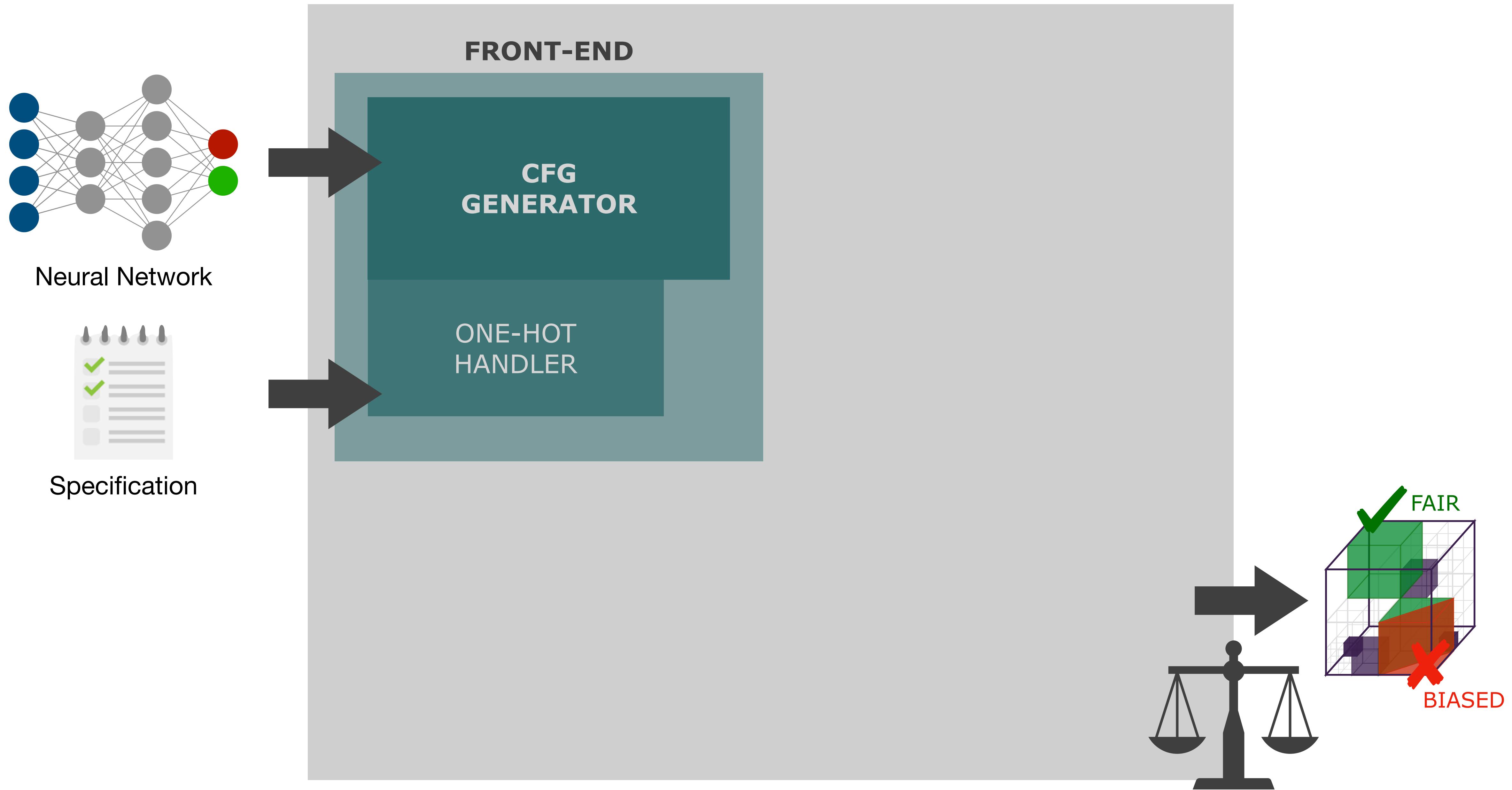


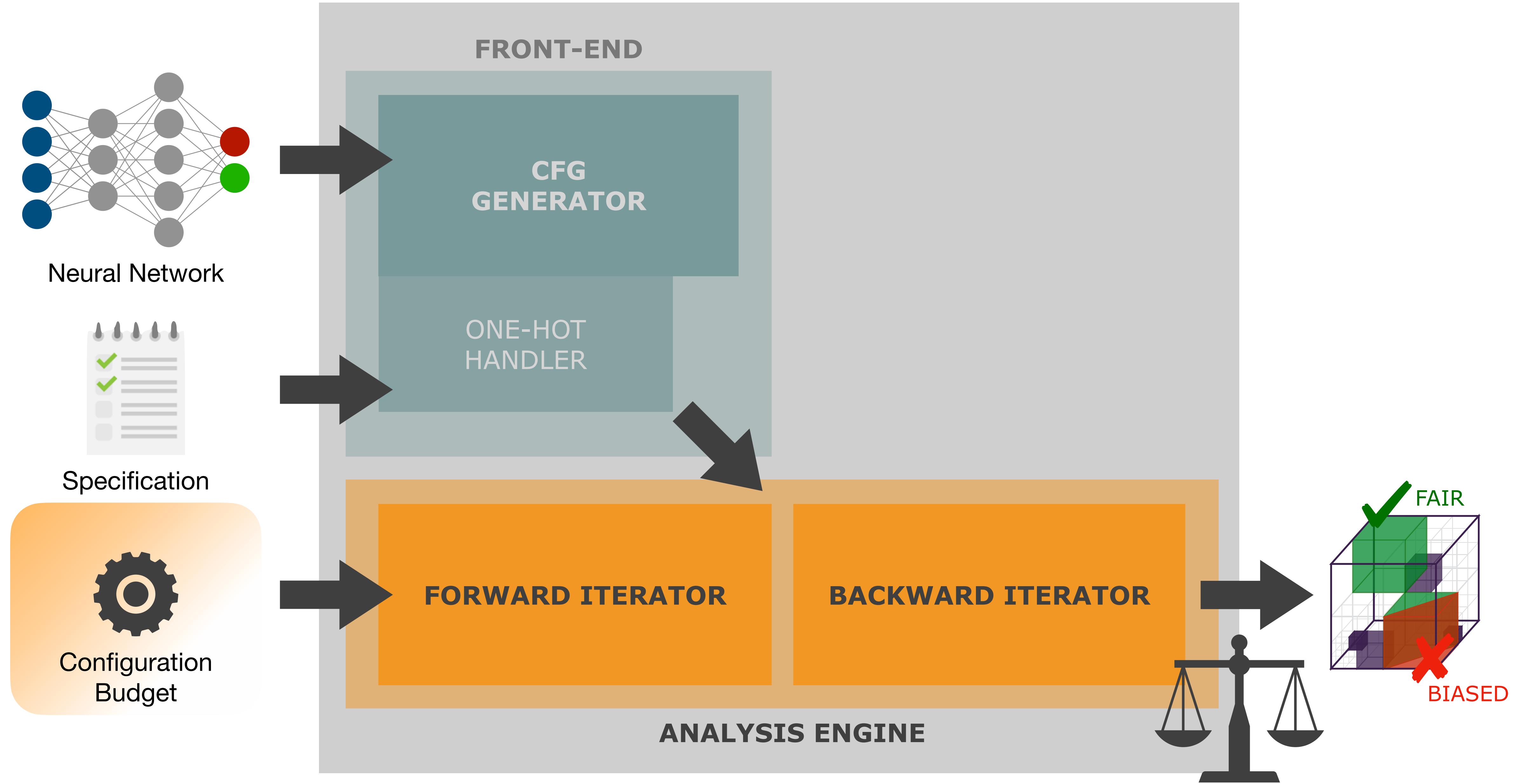
Specification



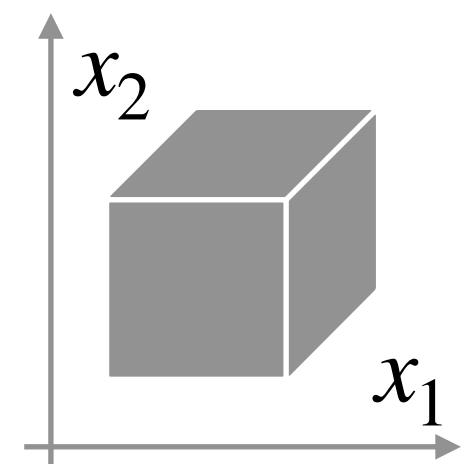
Specification



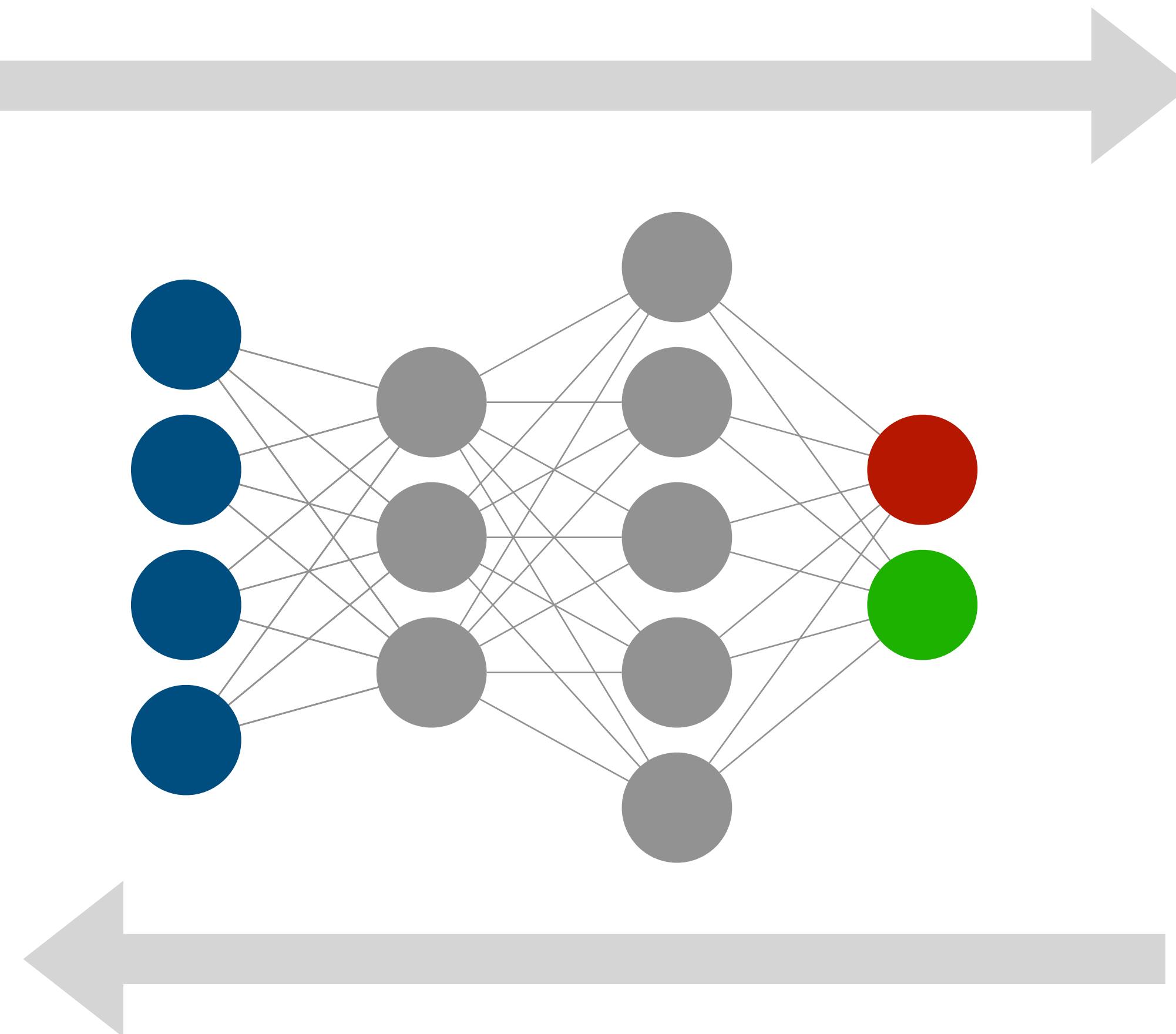
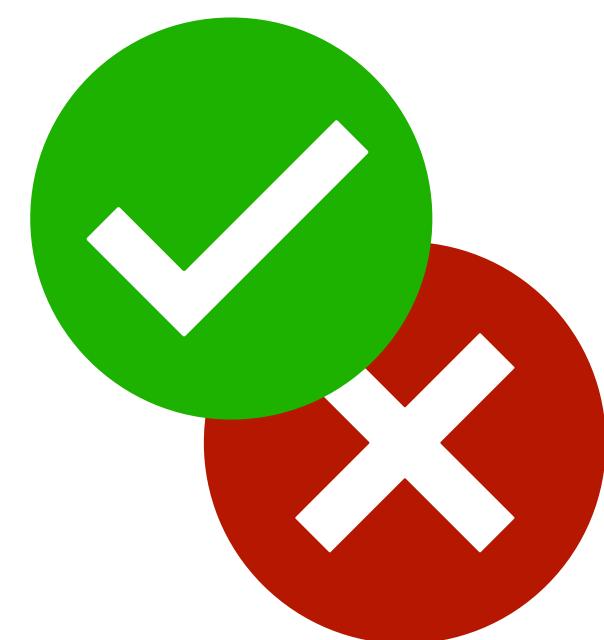




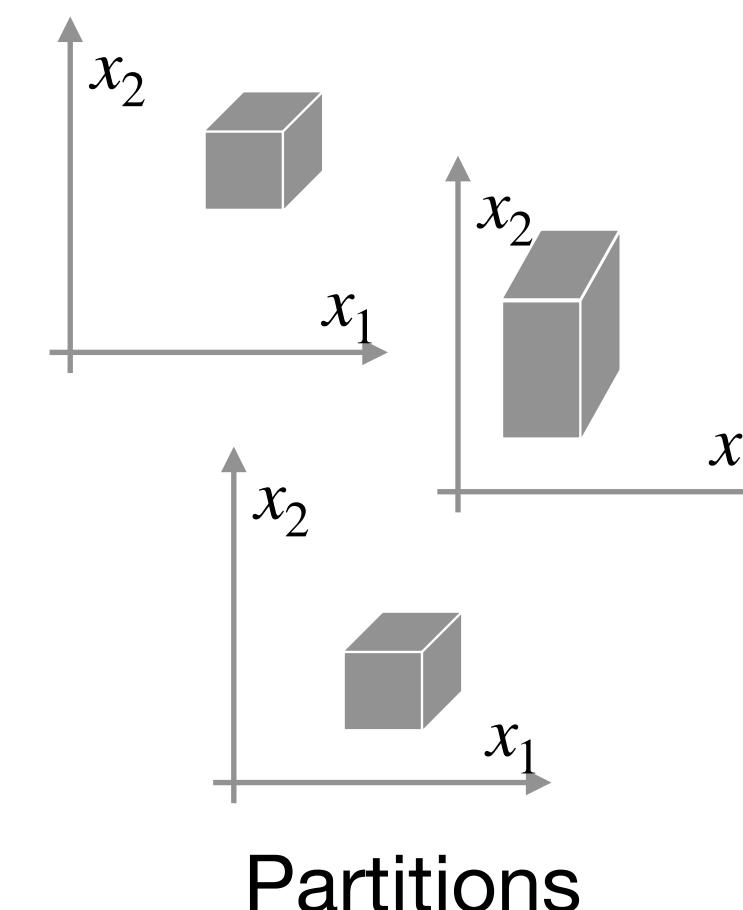
Cheap Forward Pre-Analysis



Input Space

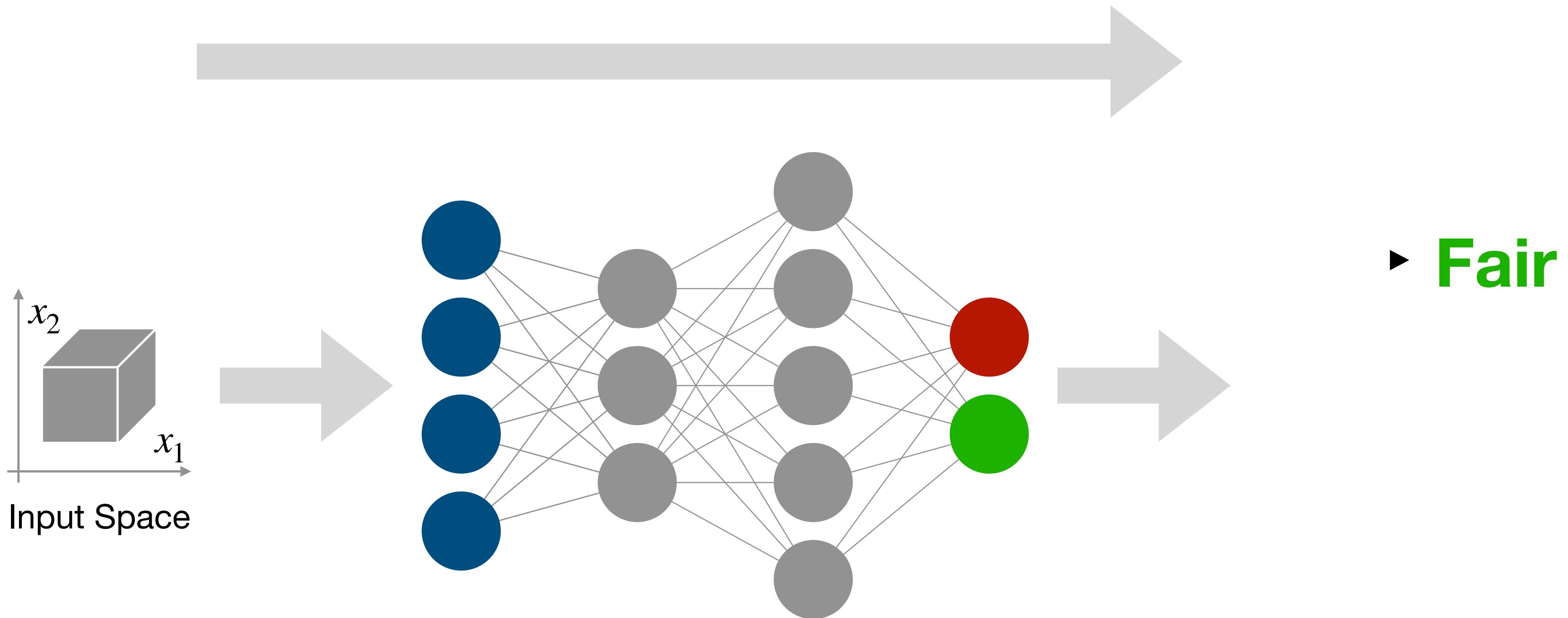


**Exact Backward Analysis
using Polyhedra**



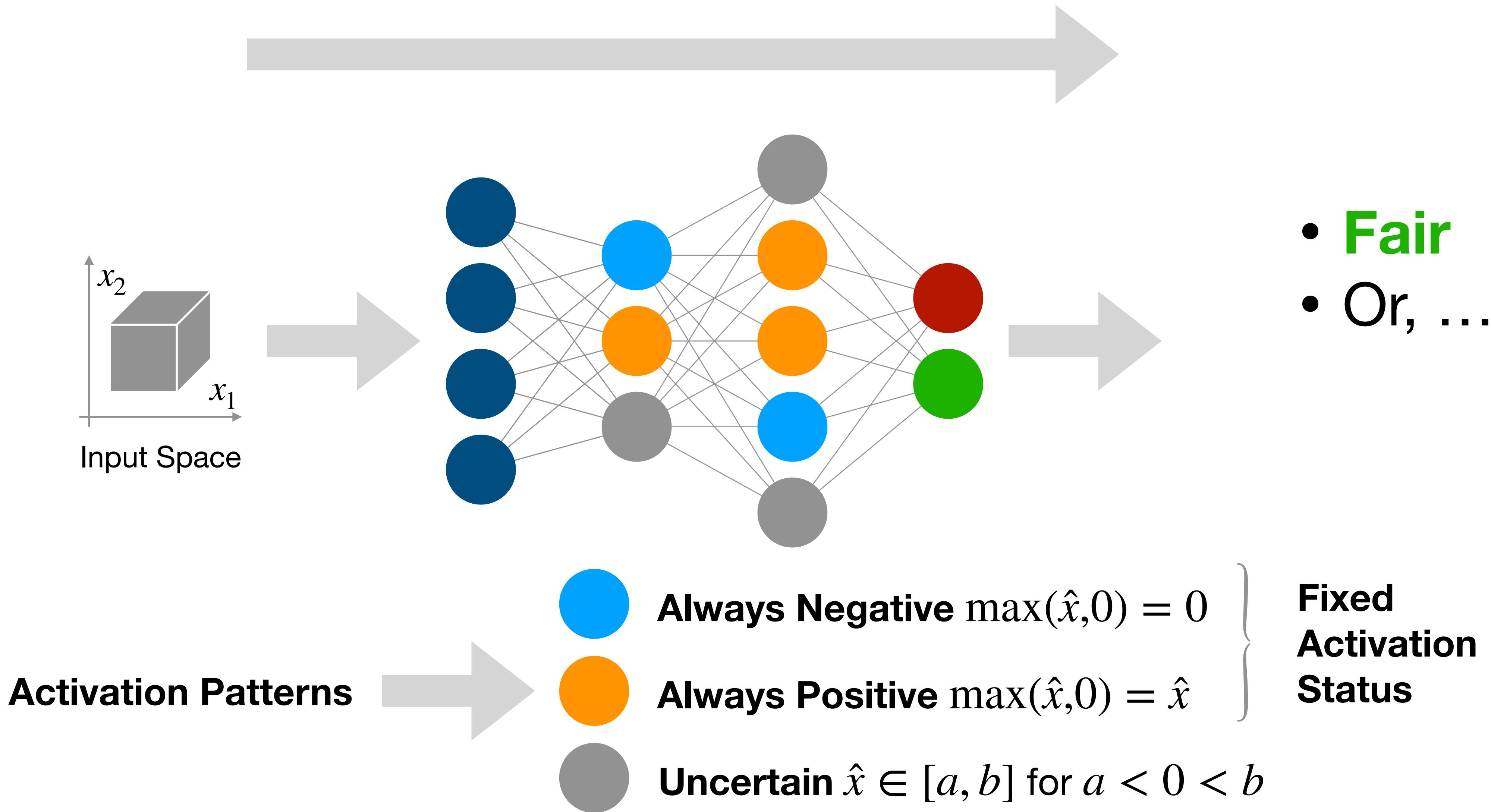
Partitions

Cheap Forward Pre-Analysis

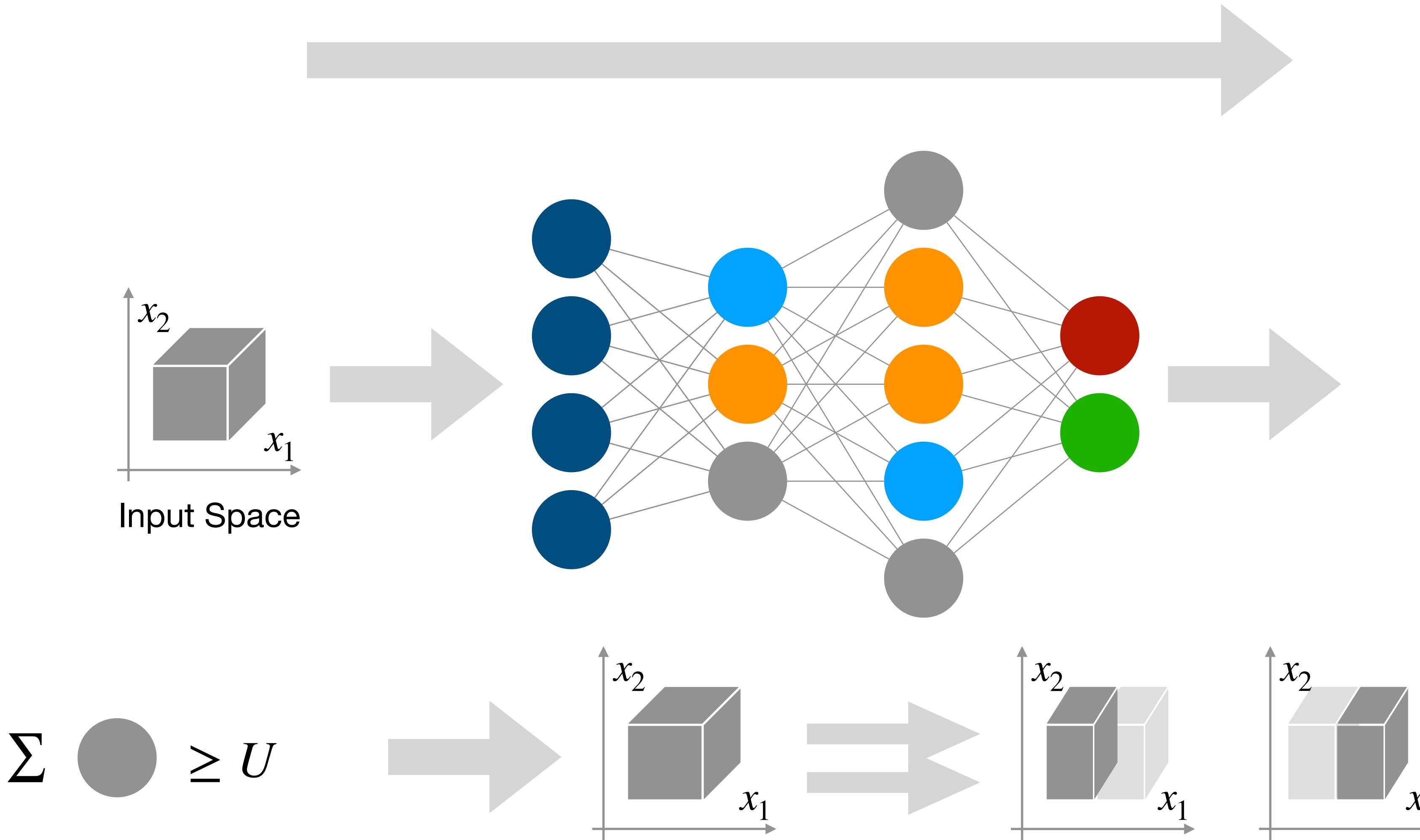


Propagate the partition through the network via **abstract domains**

Cheap Forward Pre-Analysis



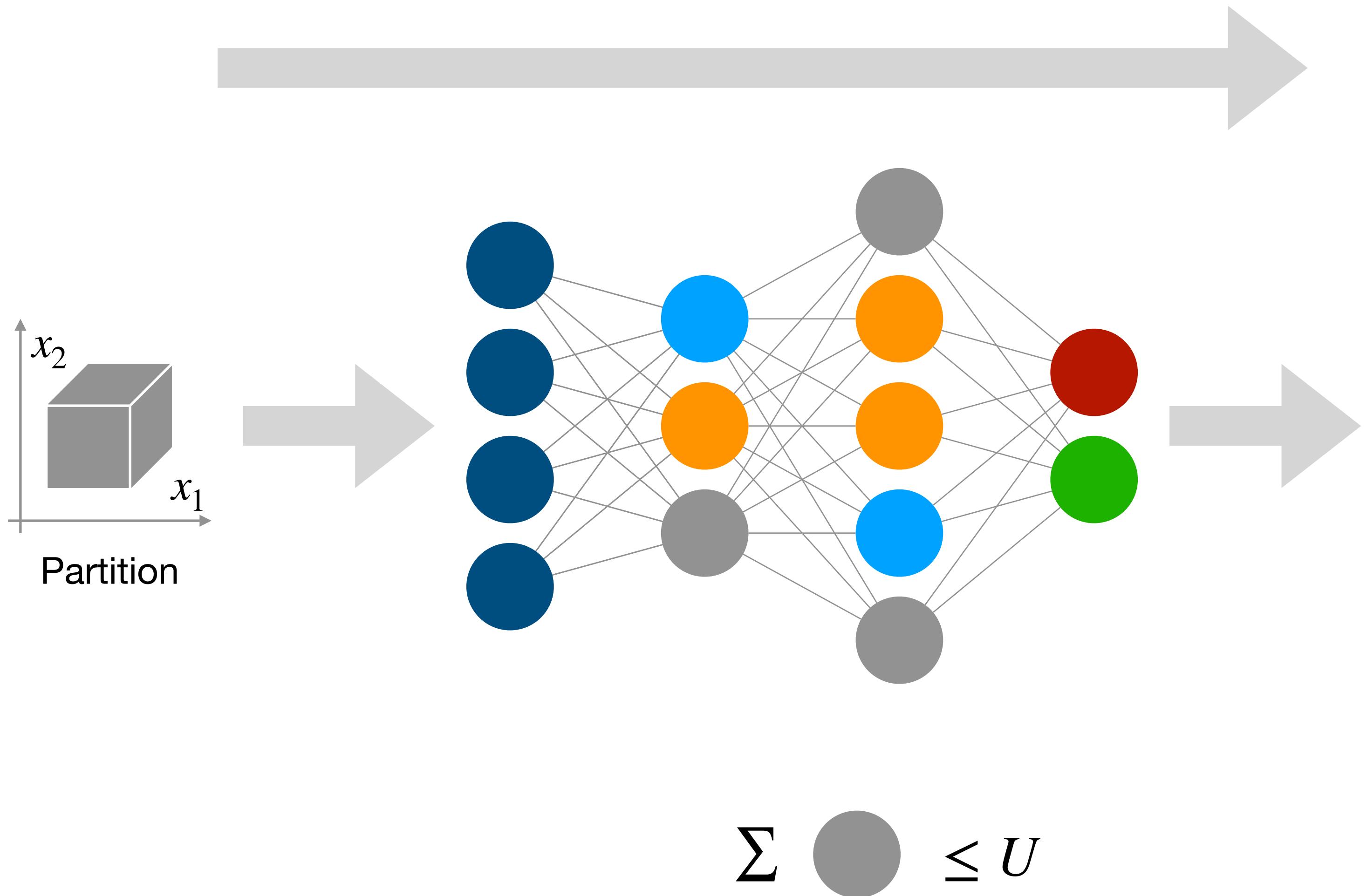
Cheap Forward Pre-Analysis



- Fair
- Partitioned

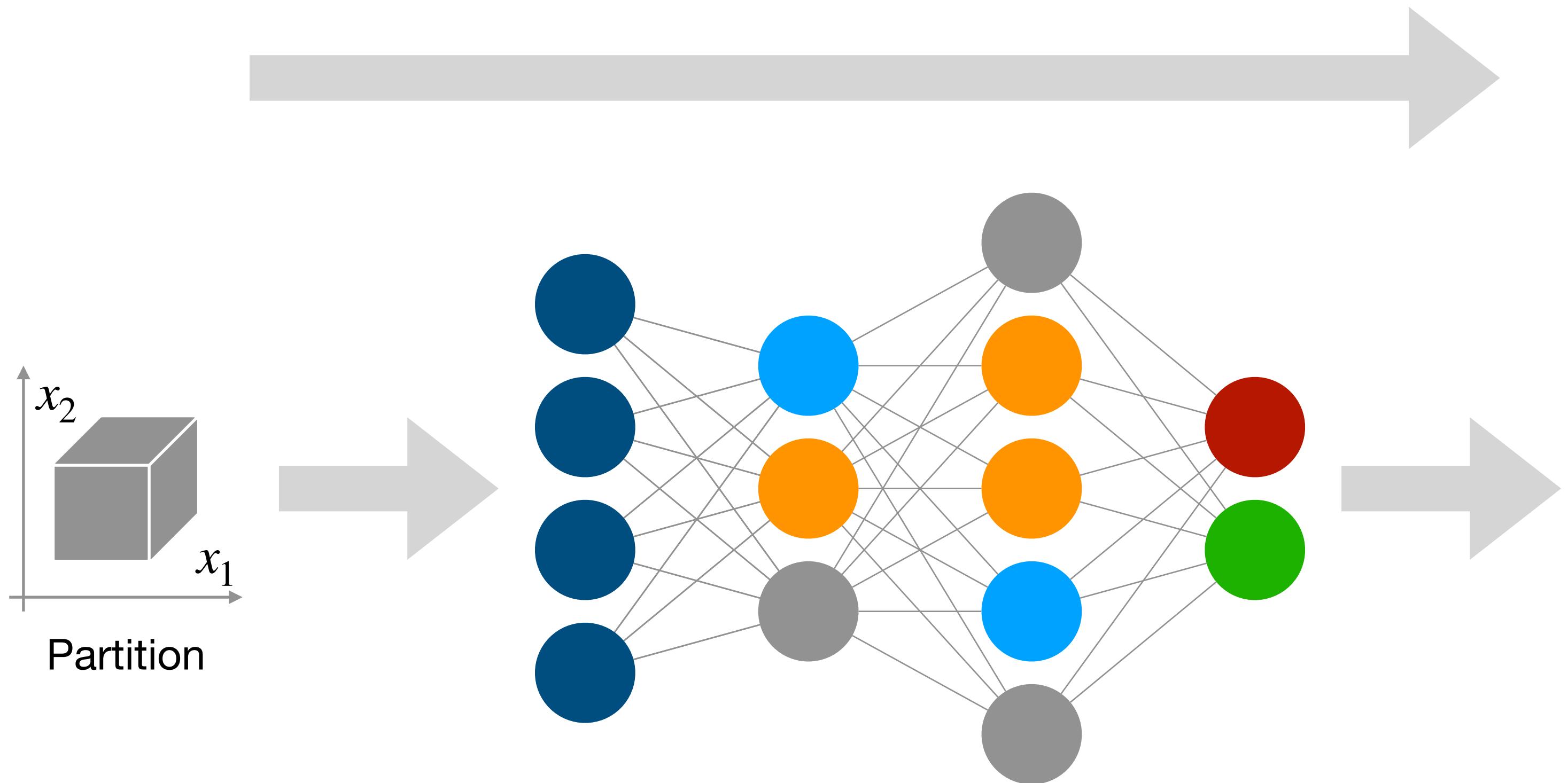
Along non-sensitive
features only

Cheap Forward Pre-Analysis



- Fair
- Partitioned
- Feasible

Cheap Forward Pre-Analysis

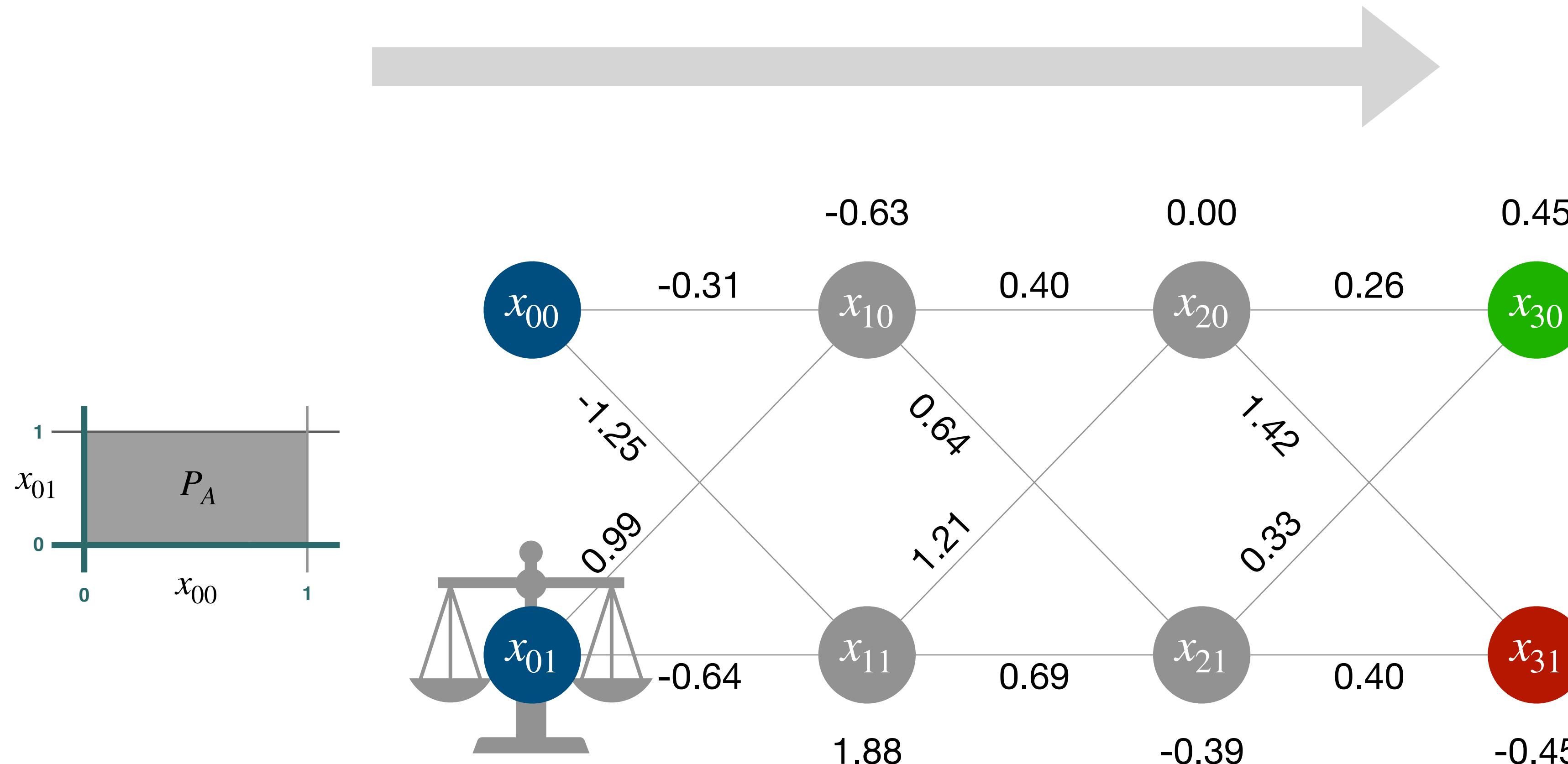


- Fair
- Partitioned
- Feasible
- Excluded

$\sum \bullet \geq U$, and the partition becomes smaller than L

Forward Analysis

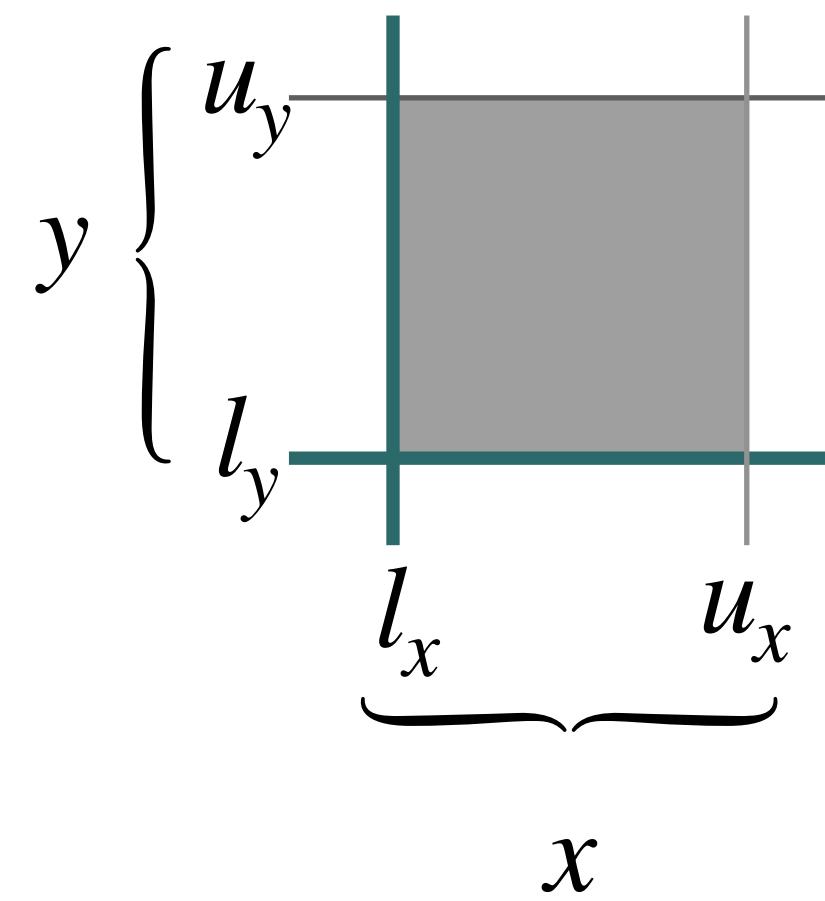
$L = 0.25, U = 2$



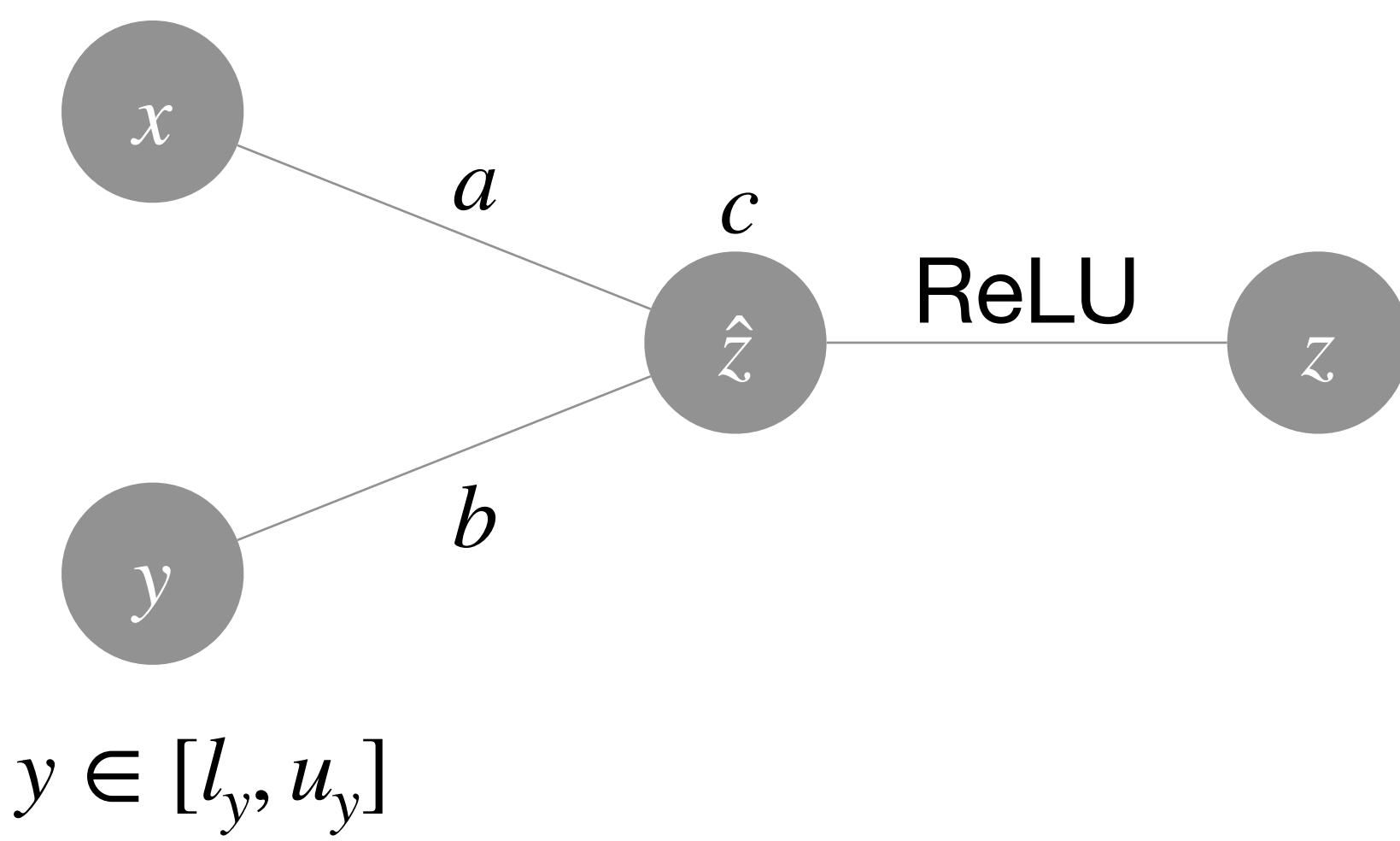
Forward Analysis



Boxes domain



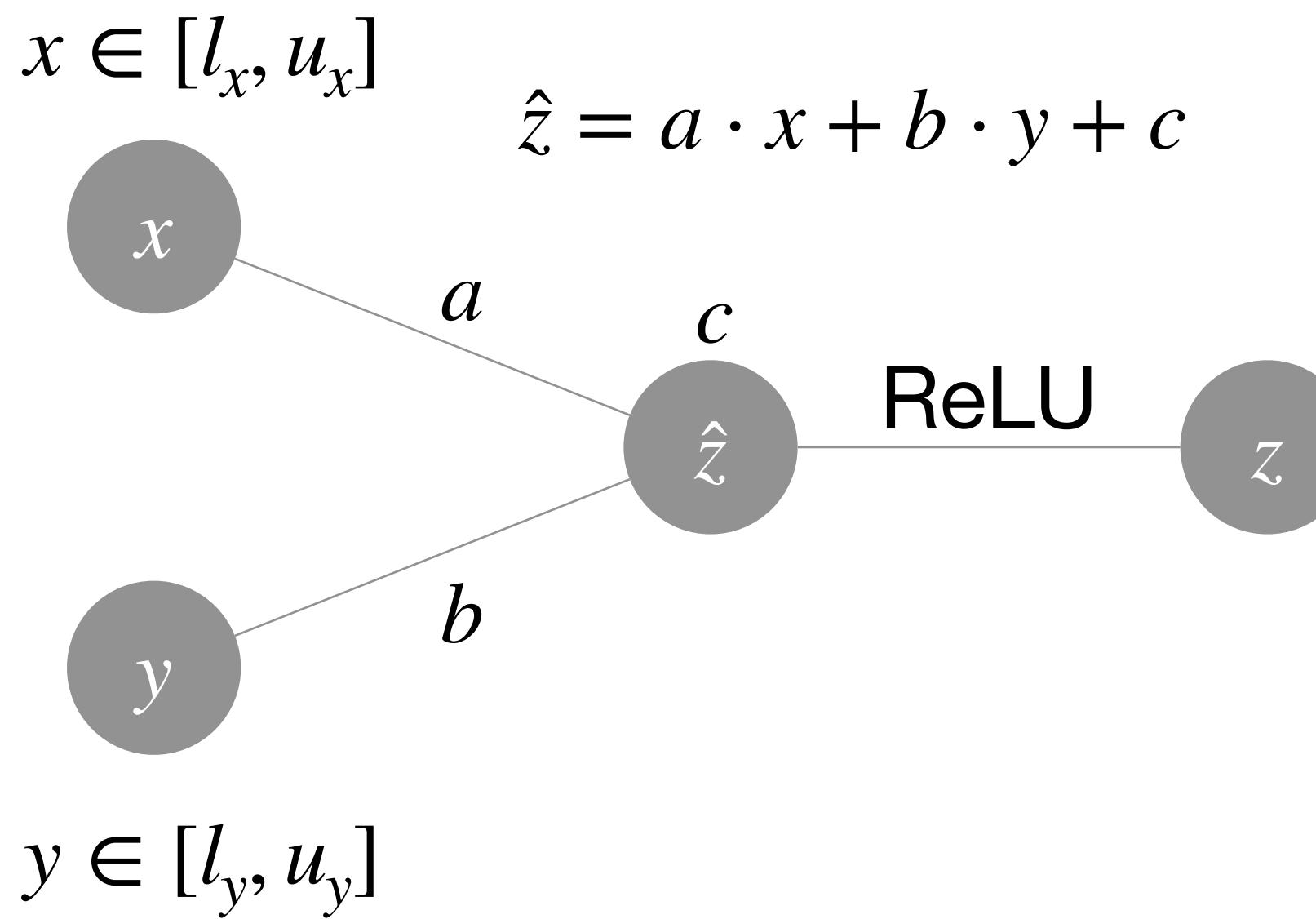
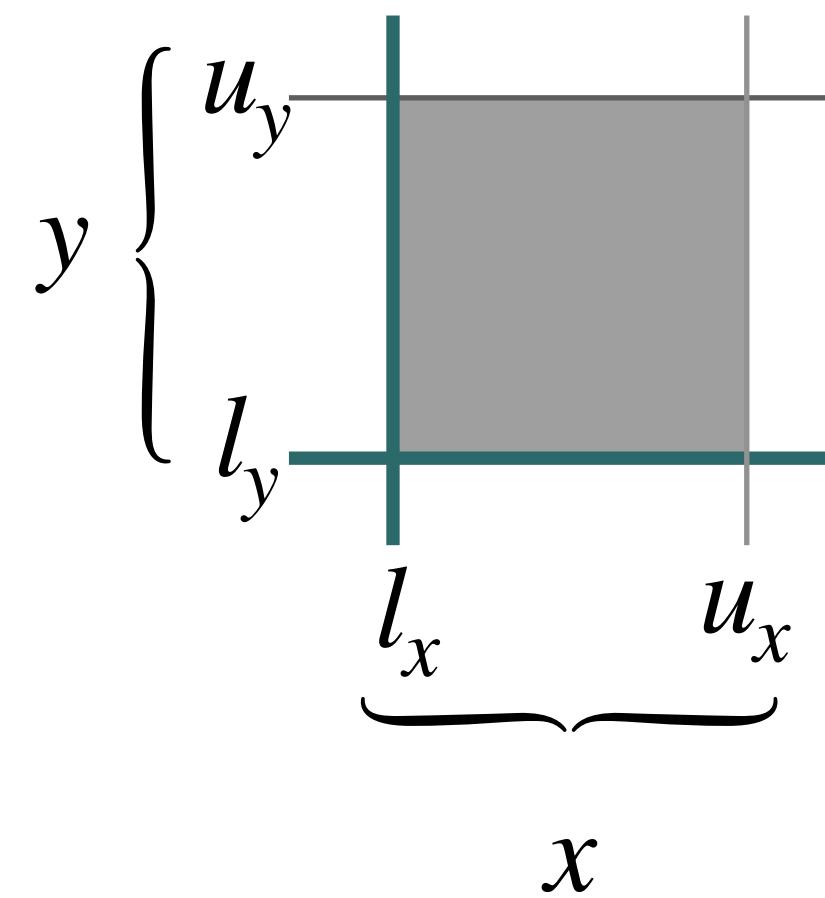
$$x \in [l_x, u_x]$$



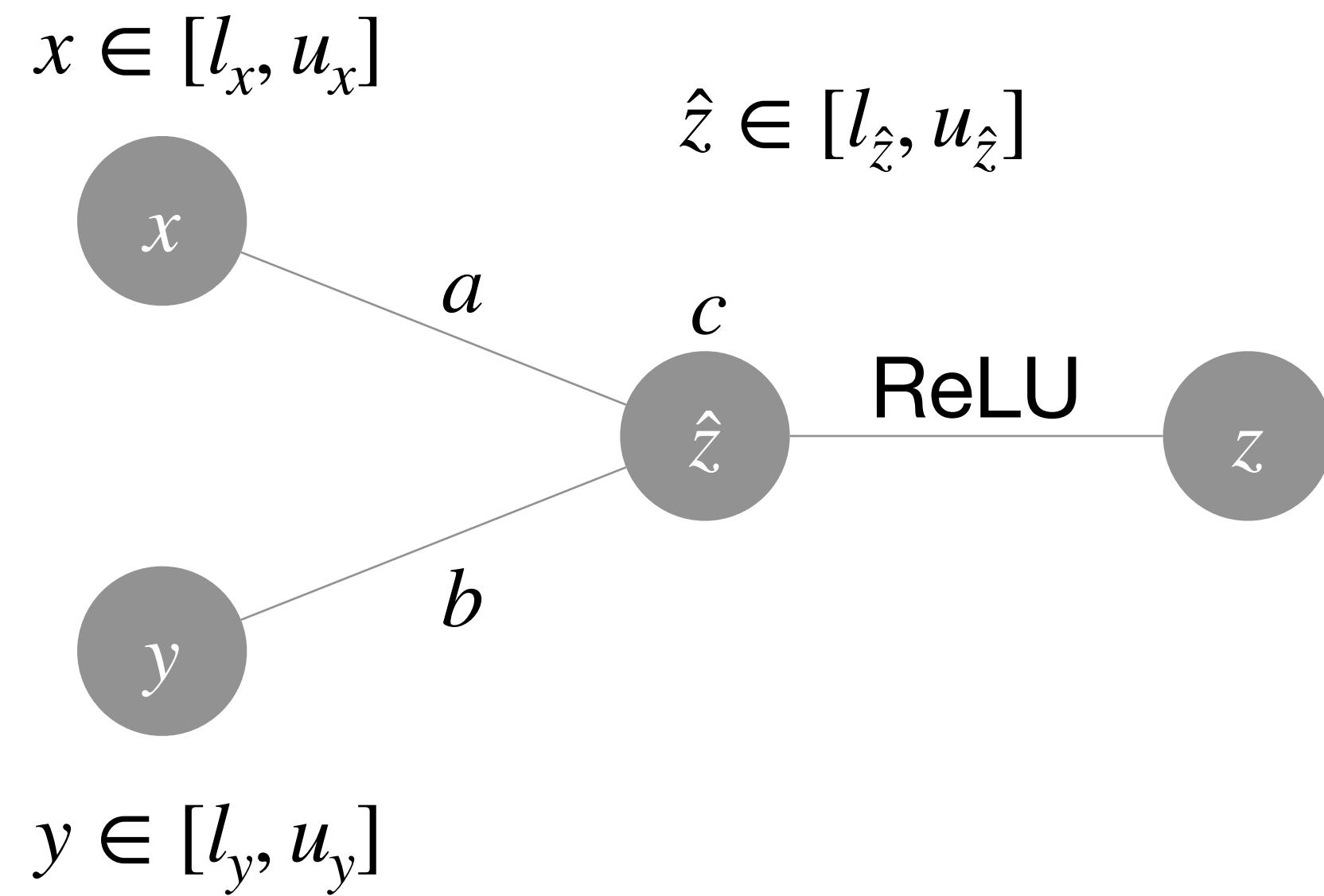
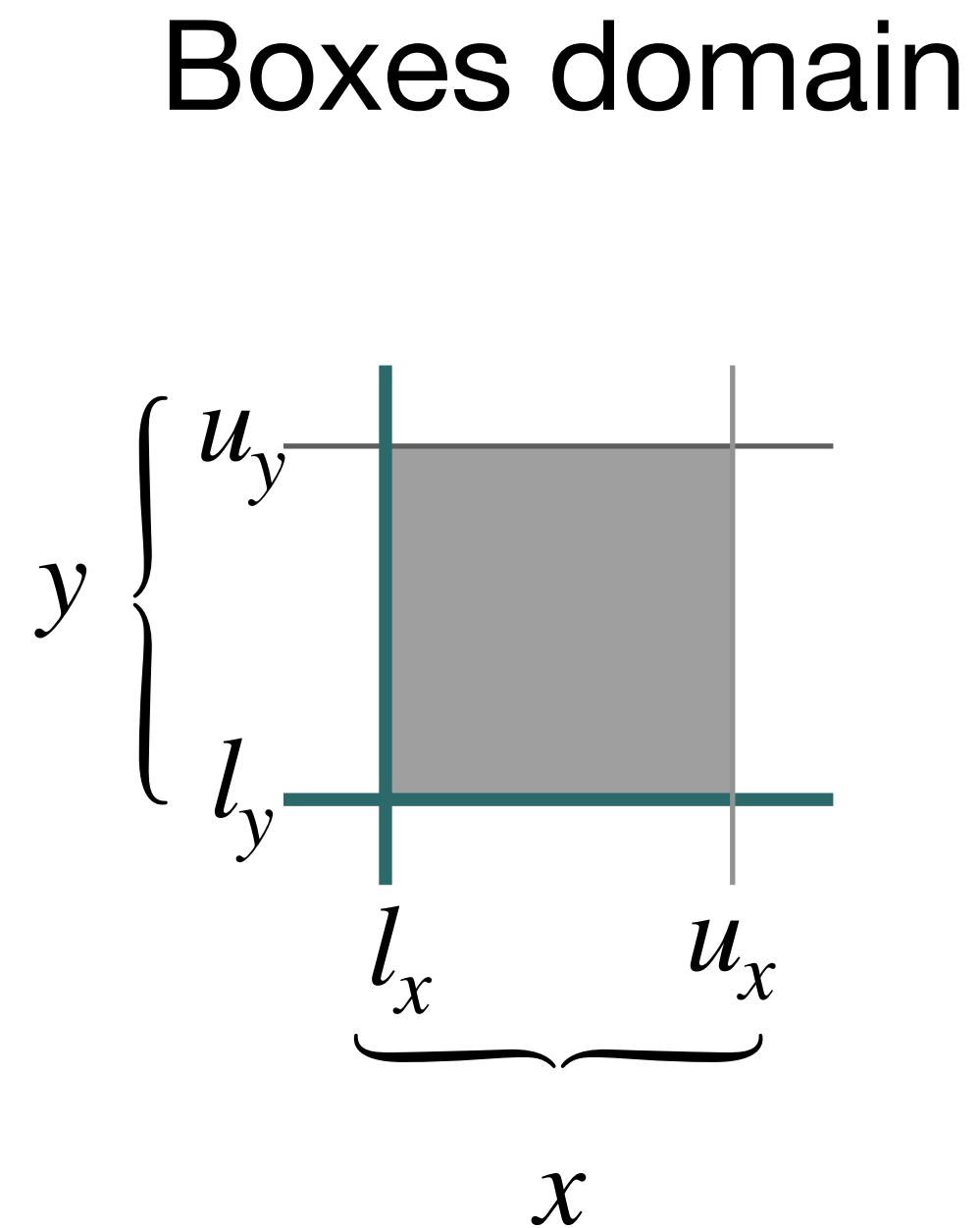
Forward Analysis



Boxes domain



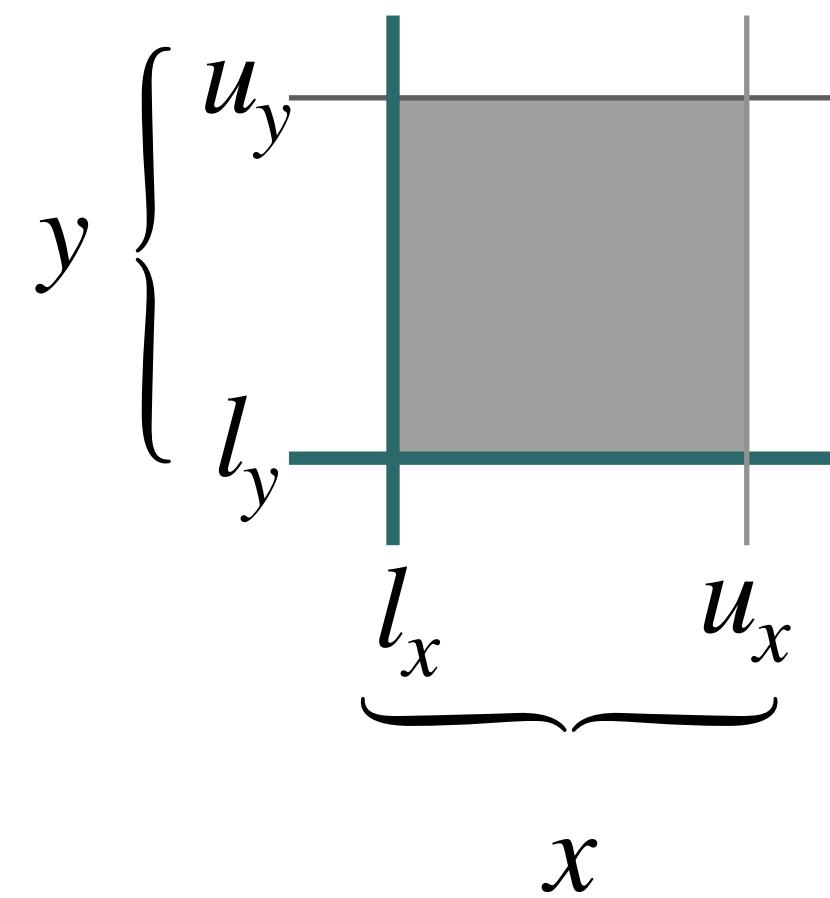
Forward Analysis



Forward Analysis

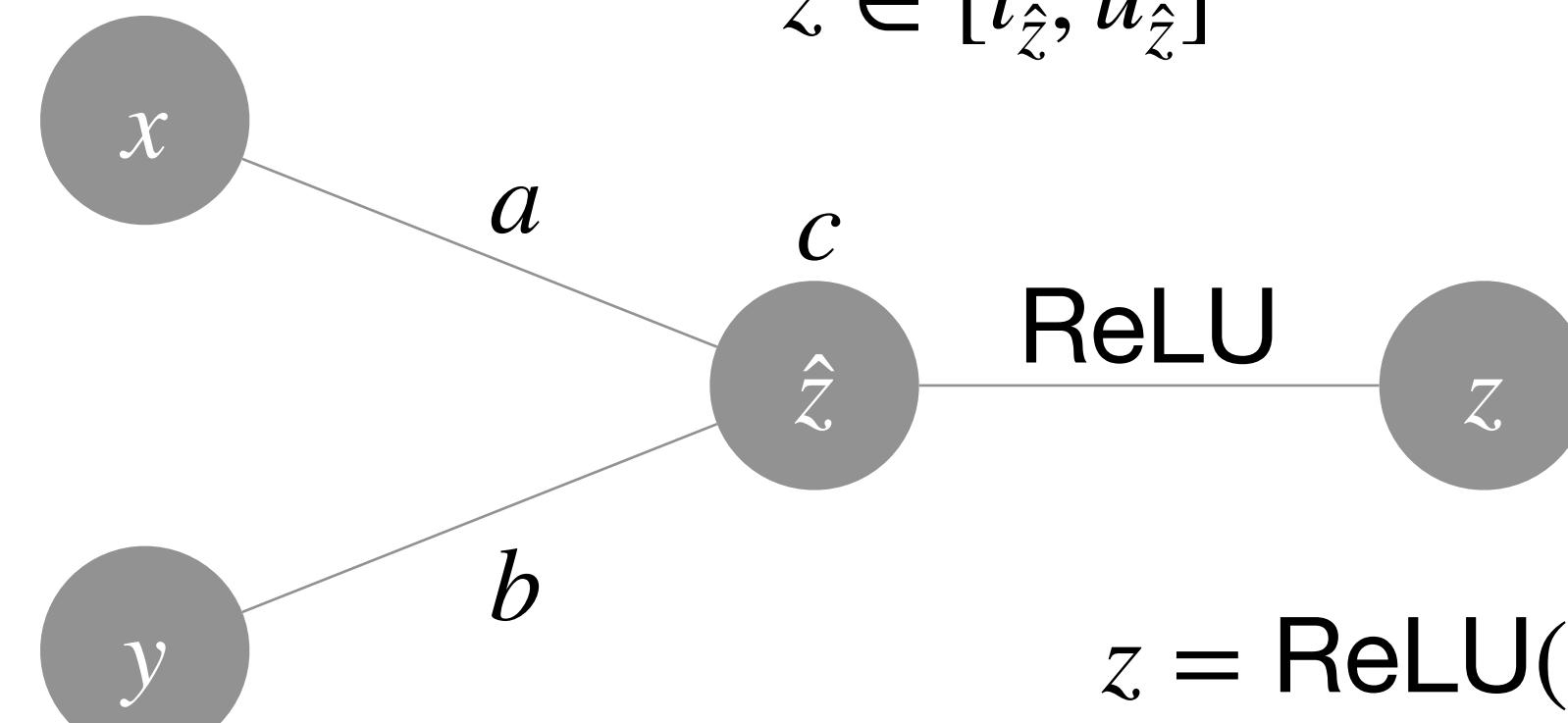


Boxes domain



$$x \in [l_x, u_x]$$

$$y \in [l_y, u_y]$$

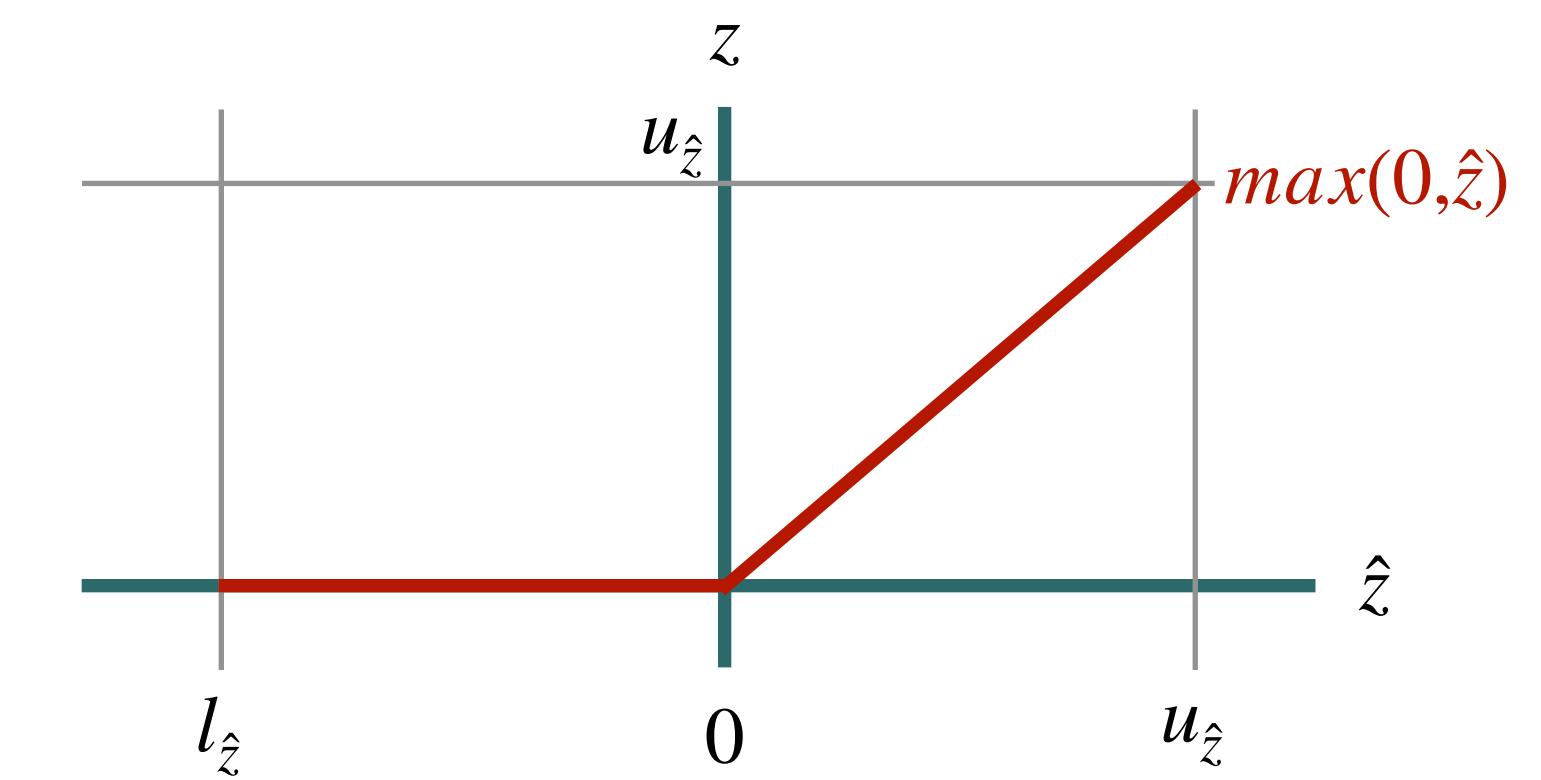
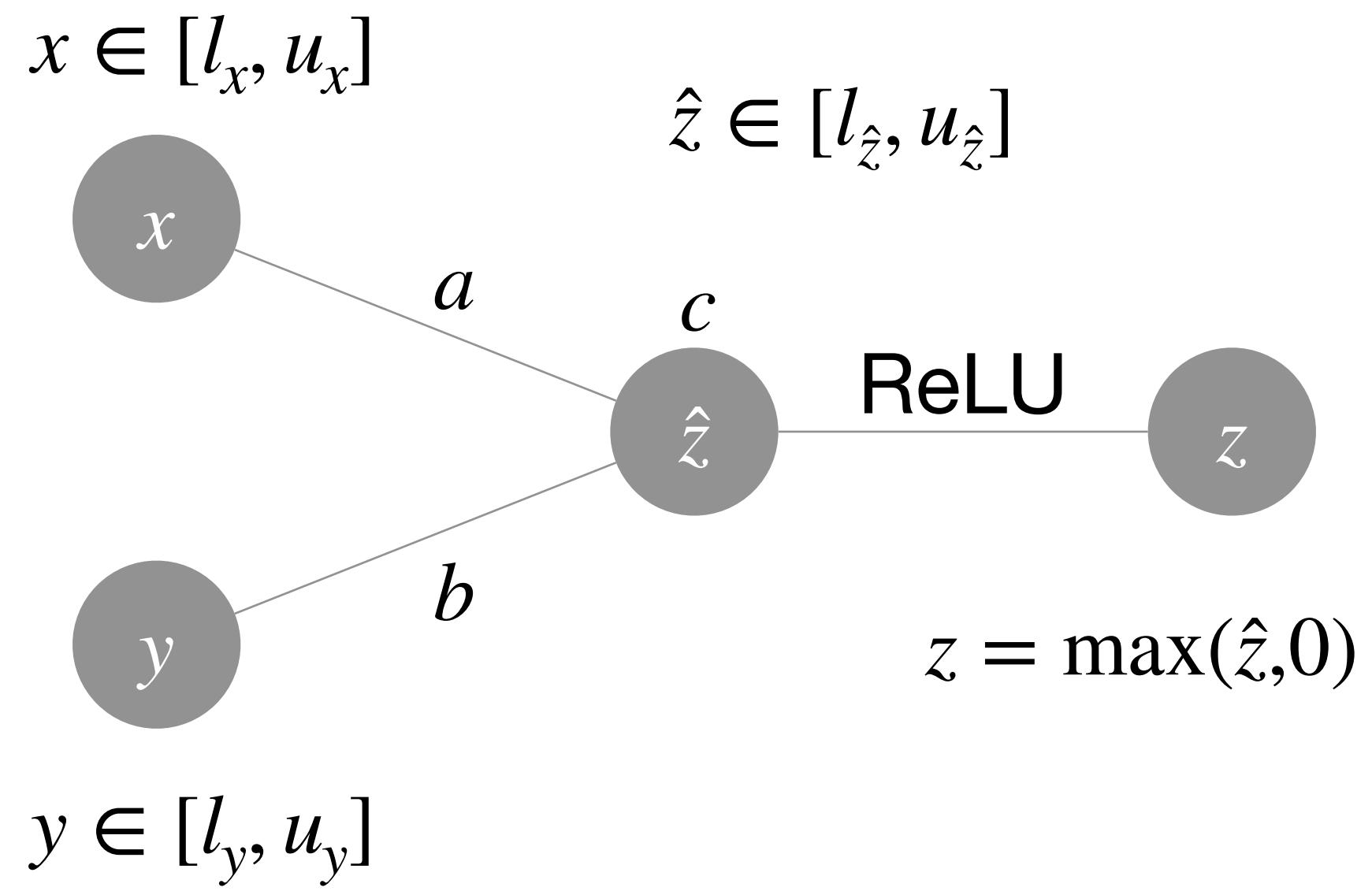
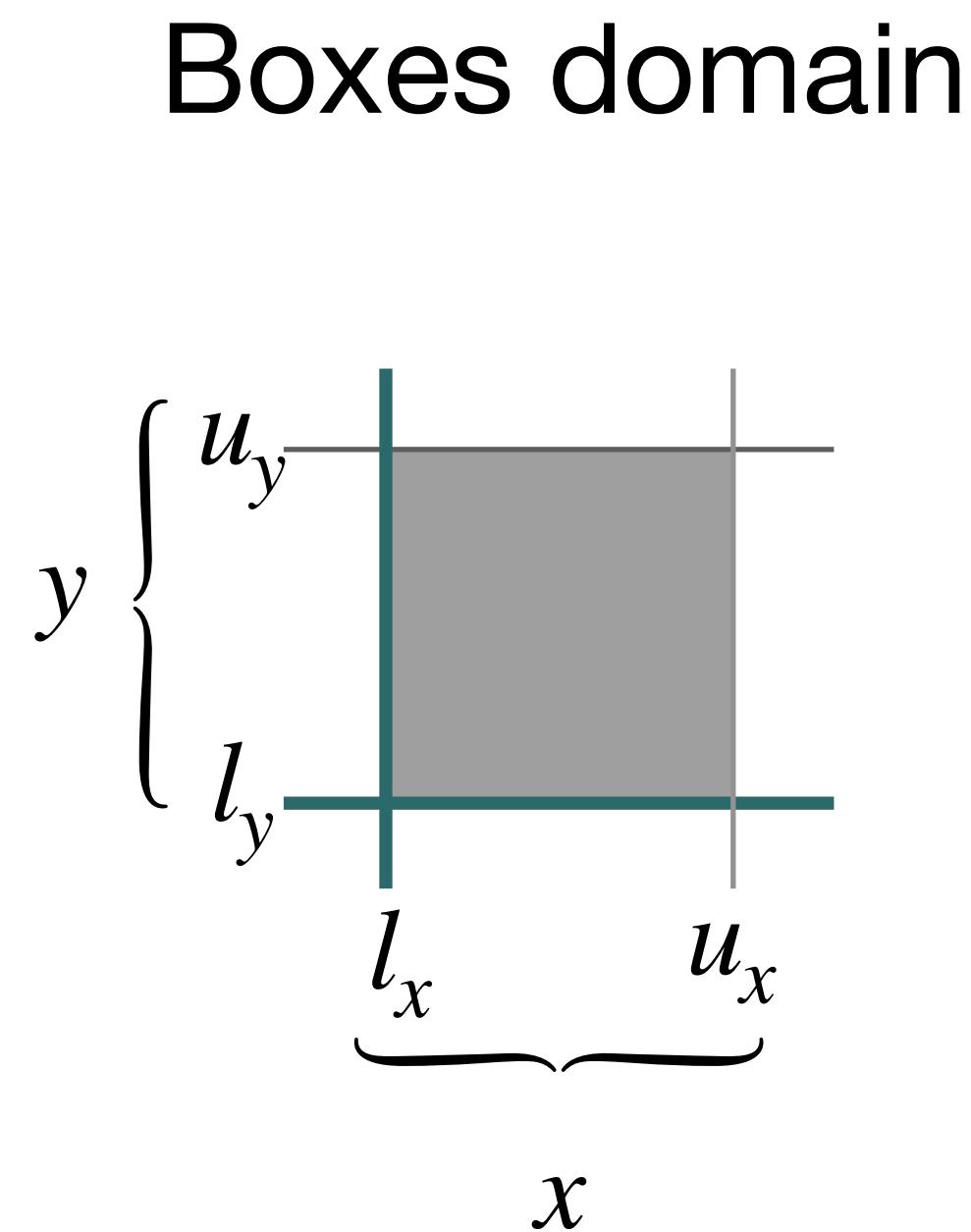


$$\hat{z} \in [l_{\hat{z}}, u_{\hat{z}}]$$

ReLU

$$z = \text{ReLU}(\hat{z}) = \max(\hat{z}, 0)$$

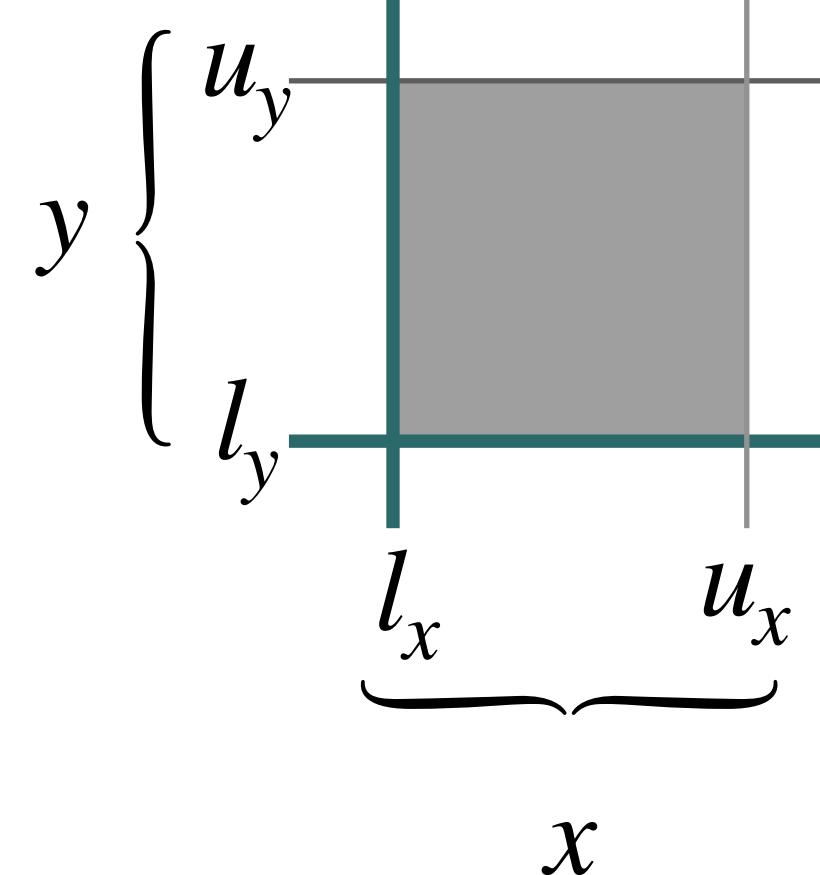
Forward Analysis



Forward Analysis

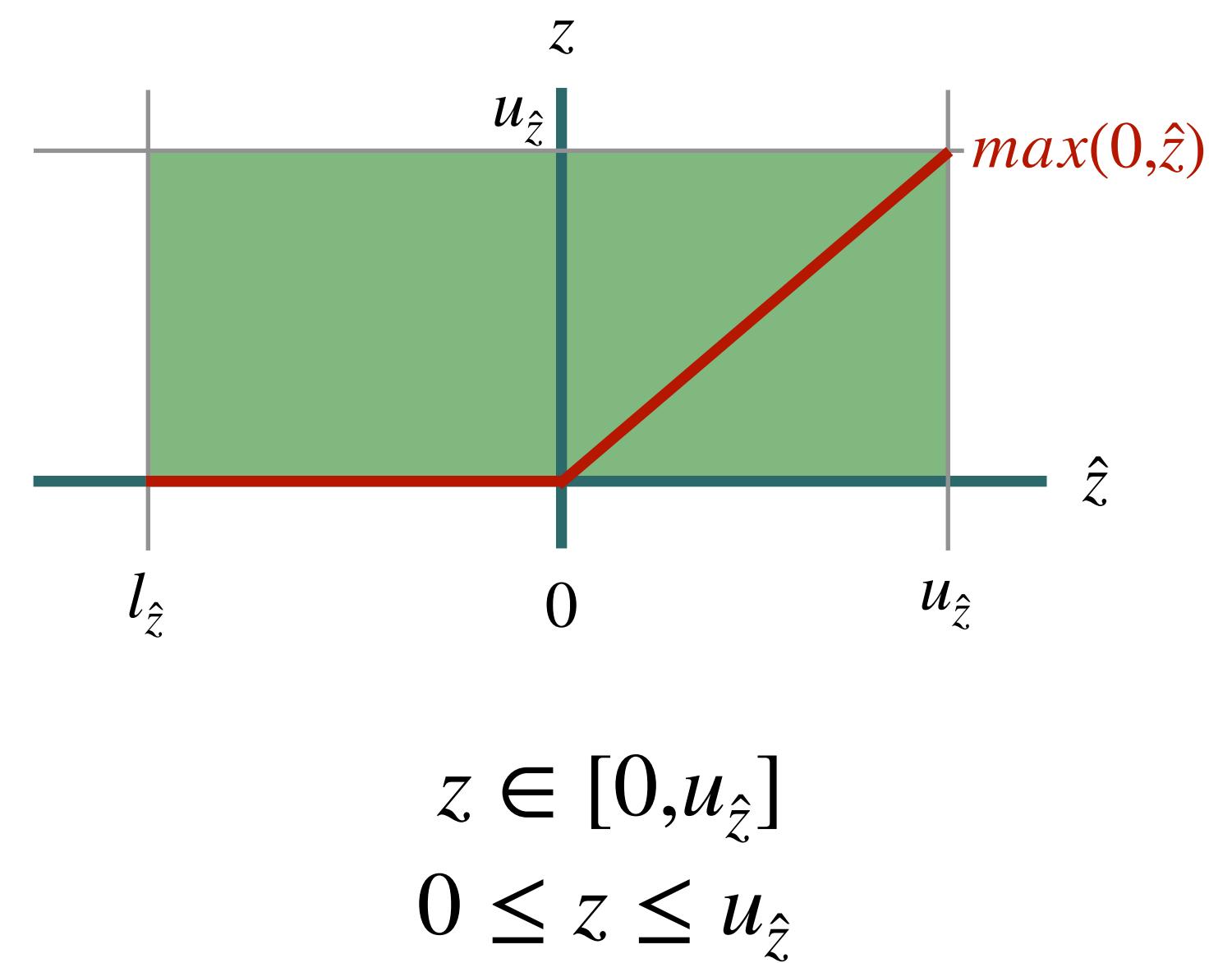
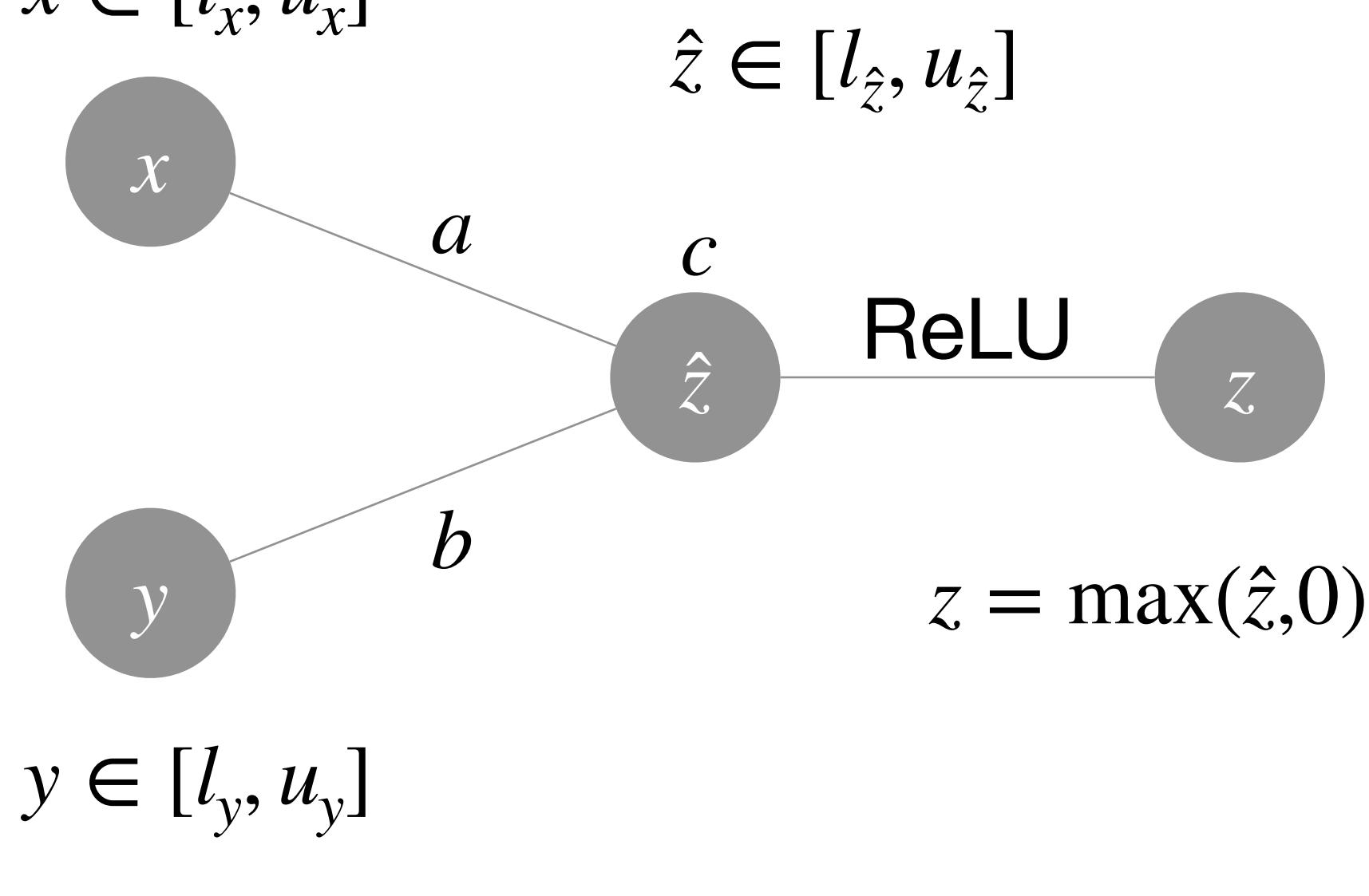


Boxes domain

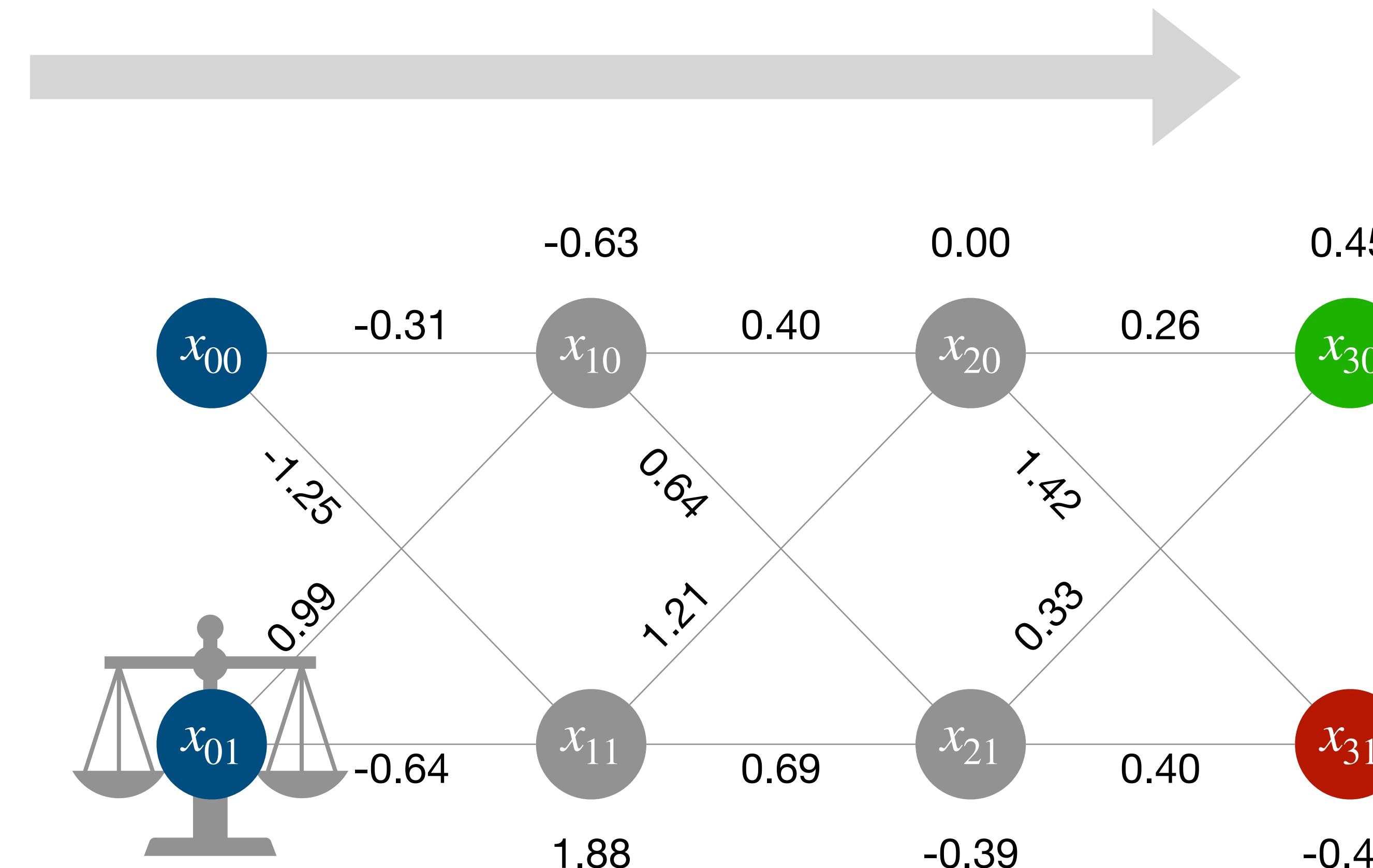


$$x \in [l_x, u_x]$$

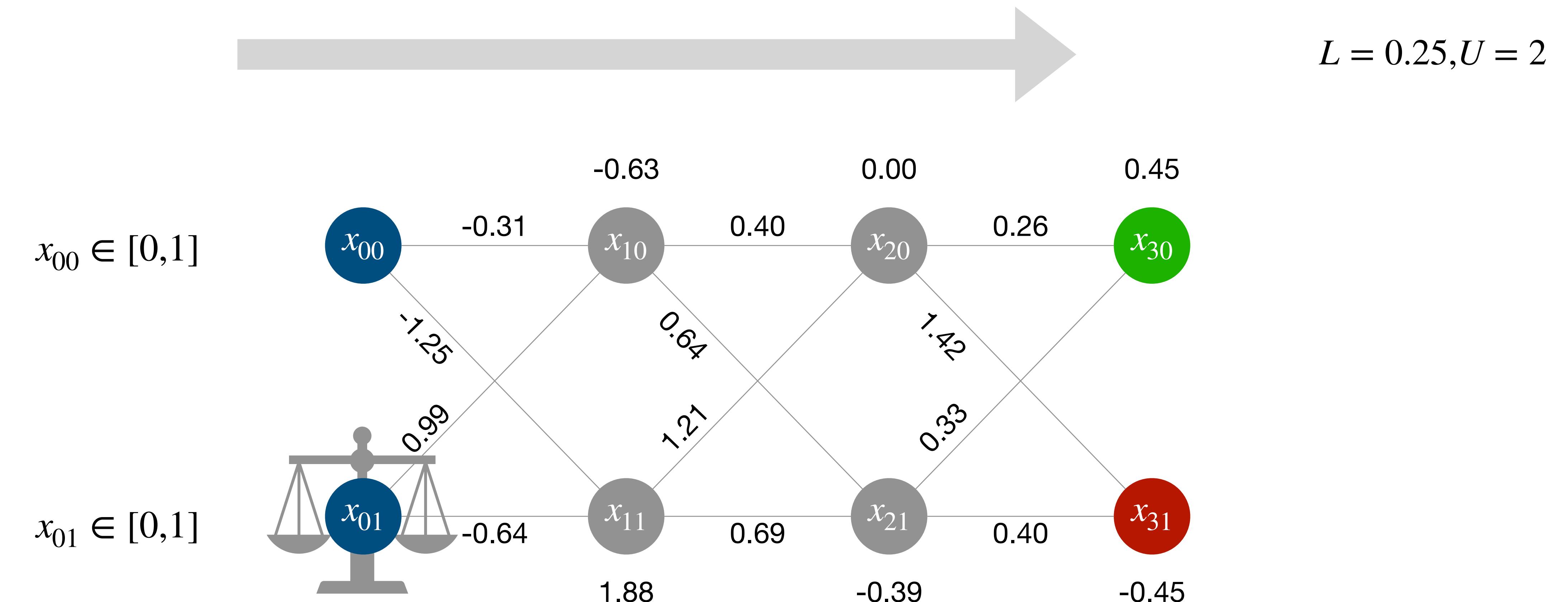
$$y \in [l_y, u_y]$$



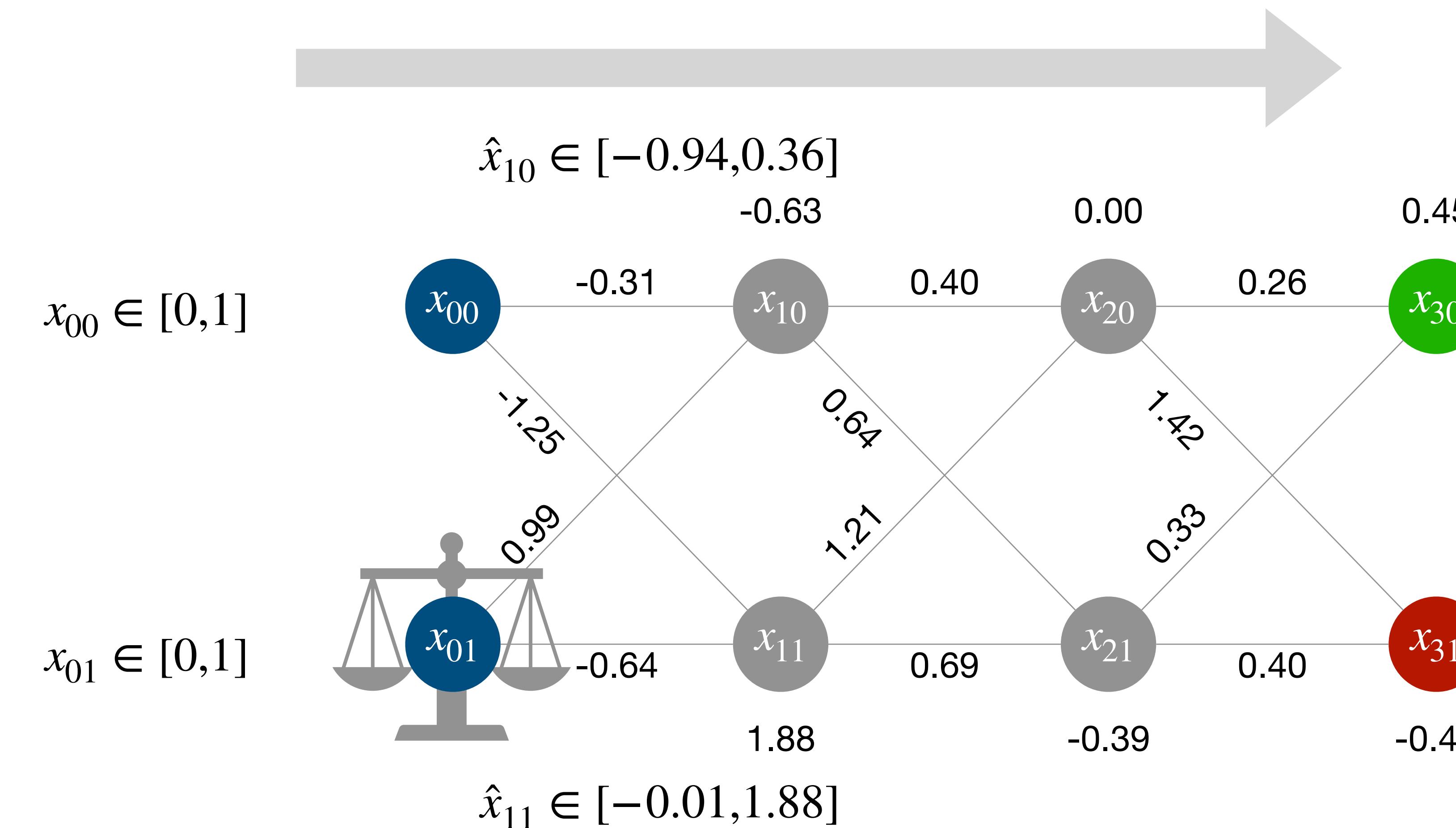
Forward Analysis



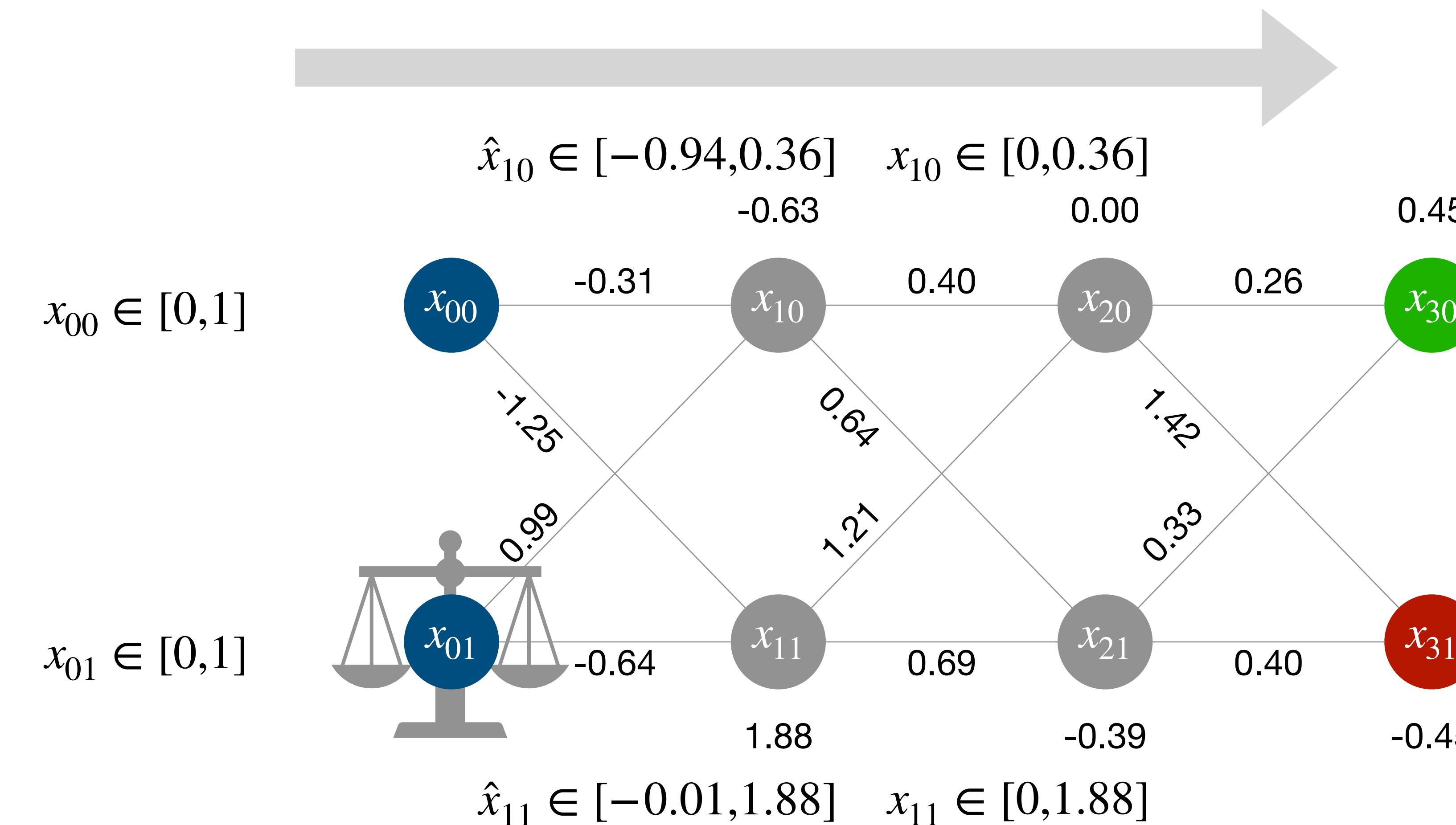
Forward Analysis



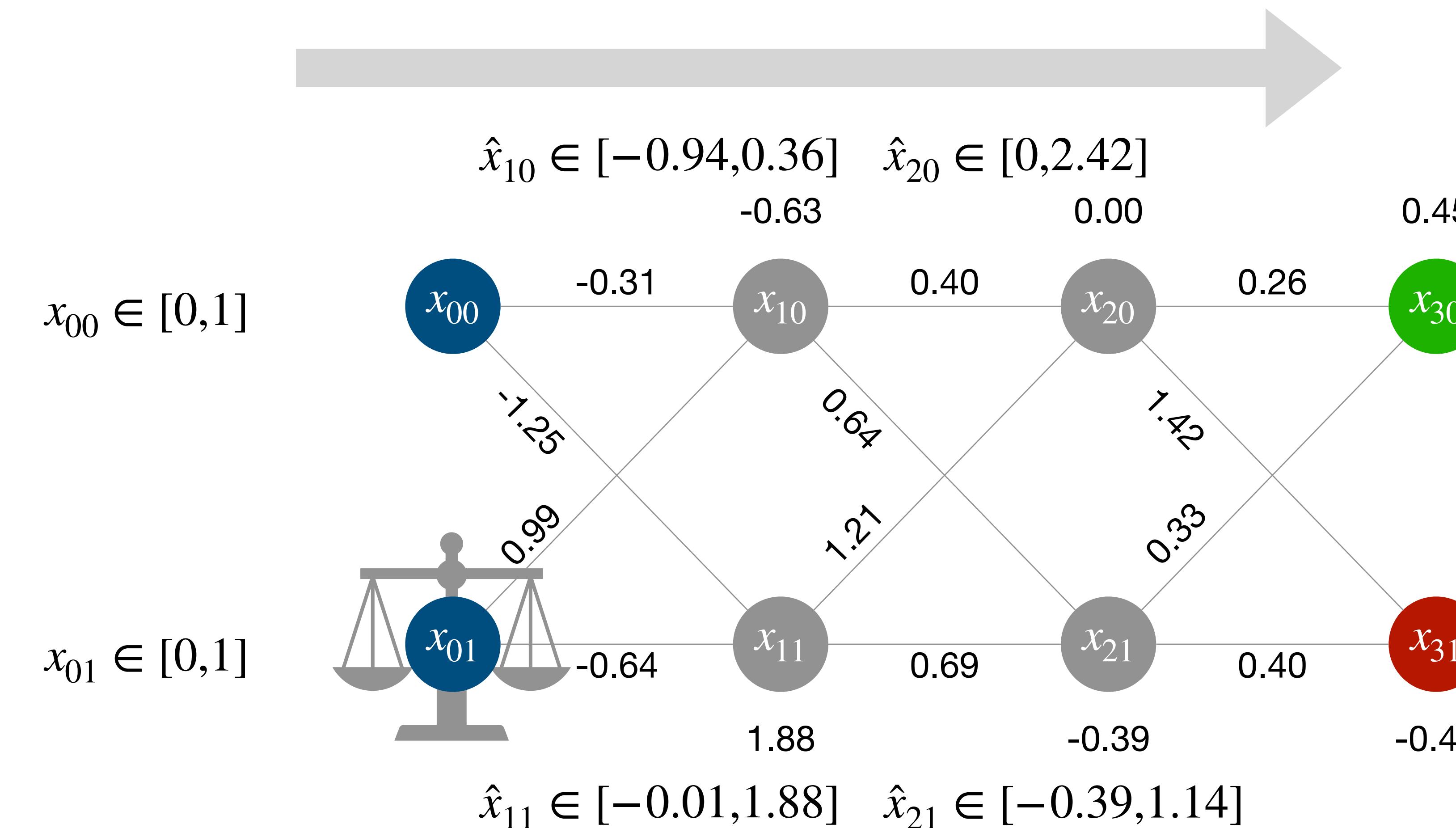
Forward Analysis



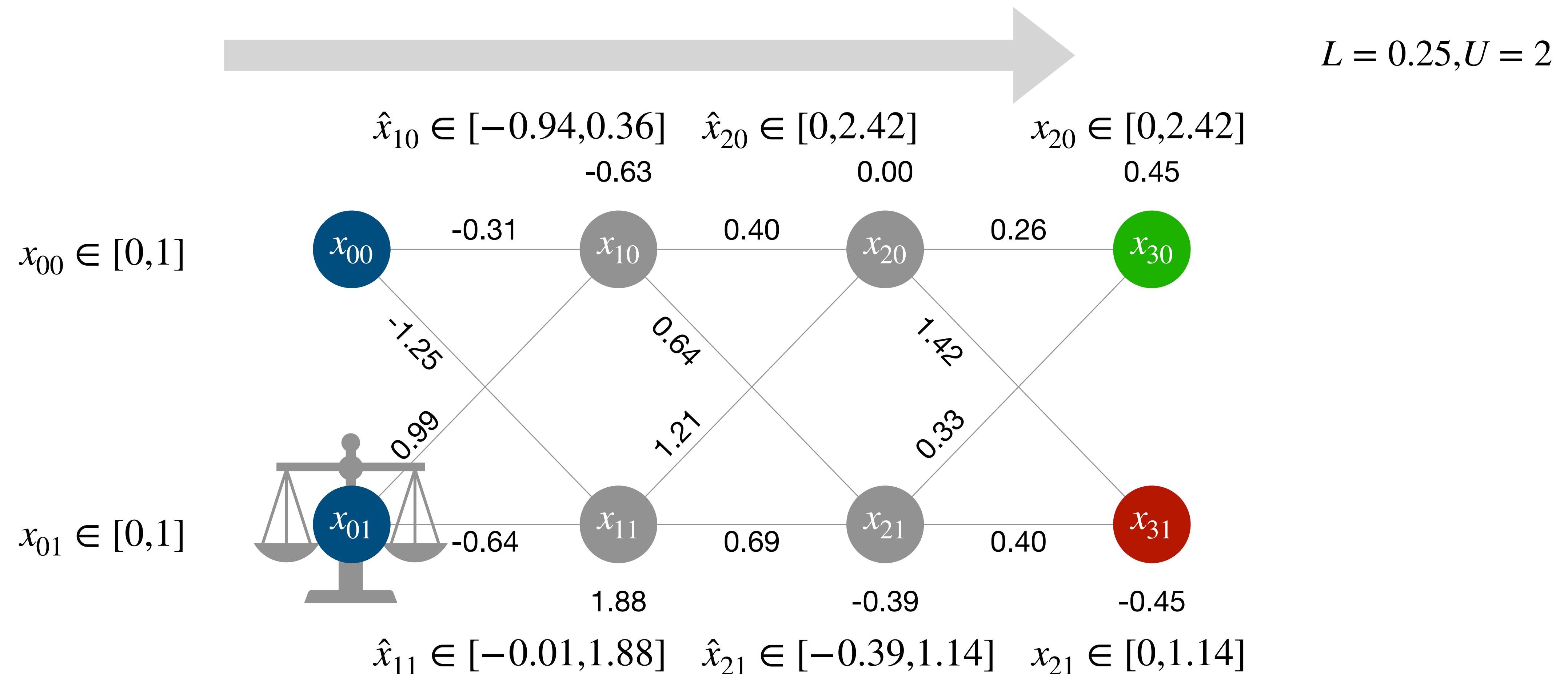
Forward Analysis



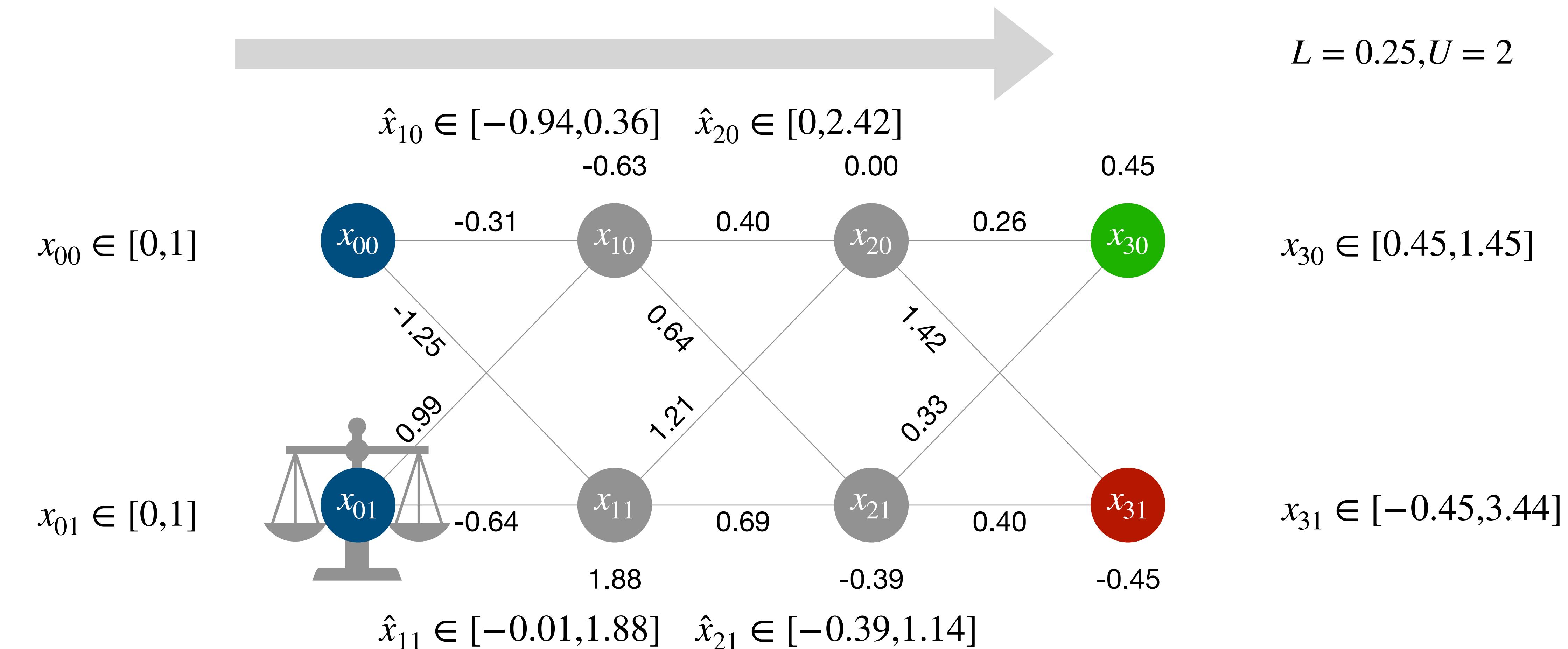
Forward Analysis



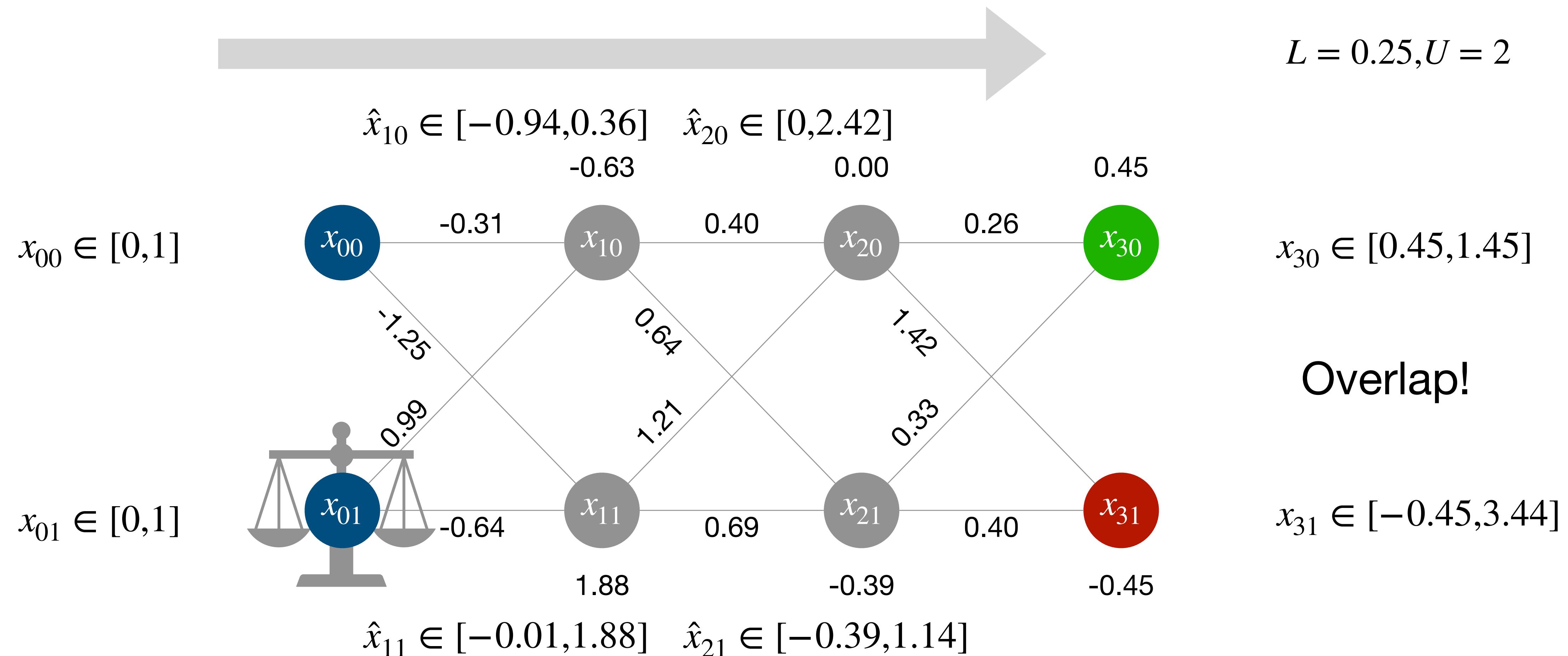
Forward Analysis



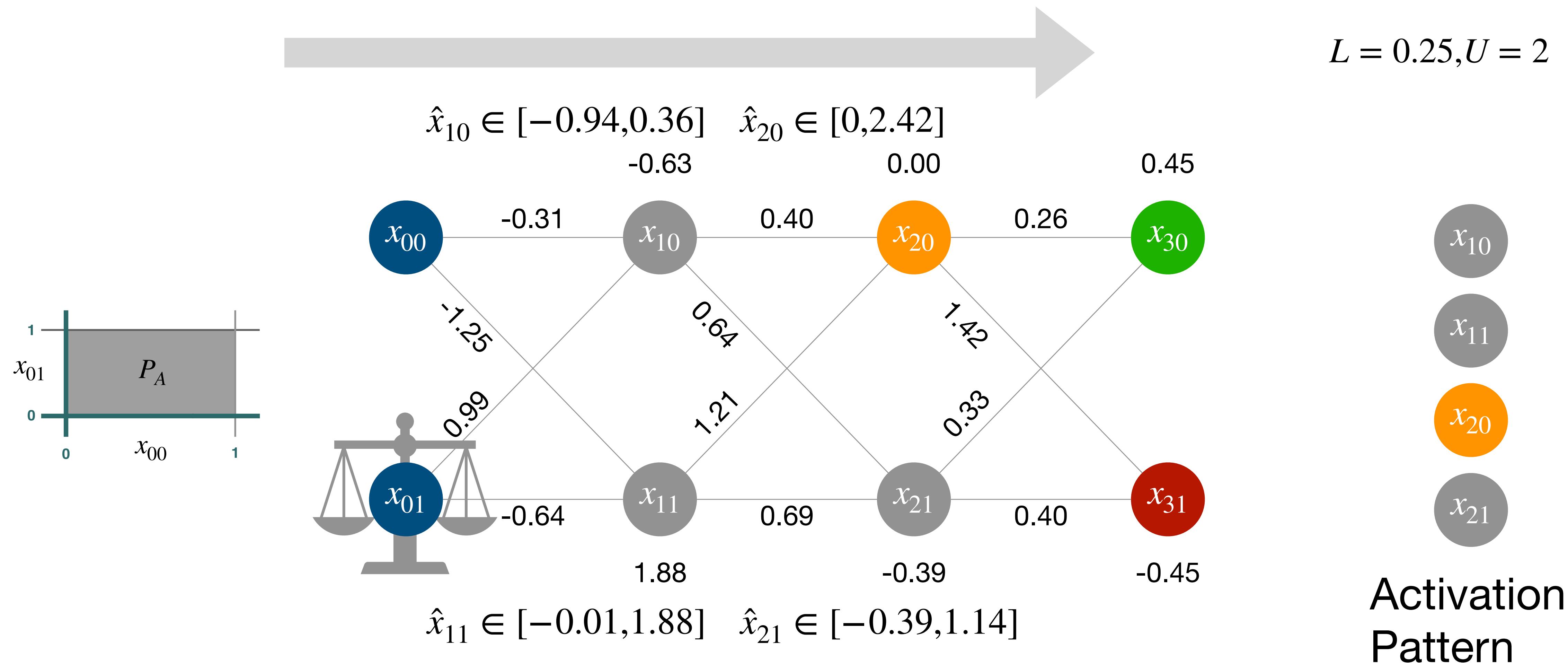
Forward Analysis



Forward Analysis



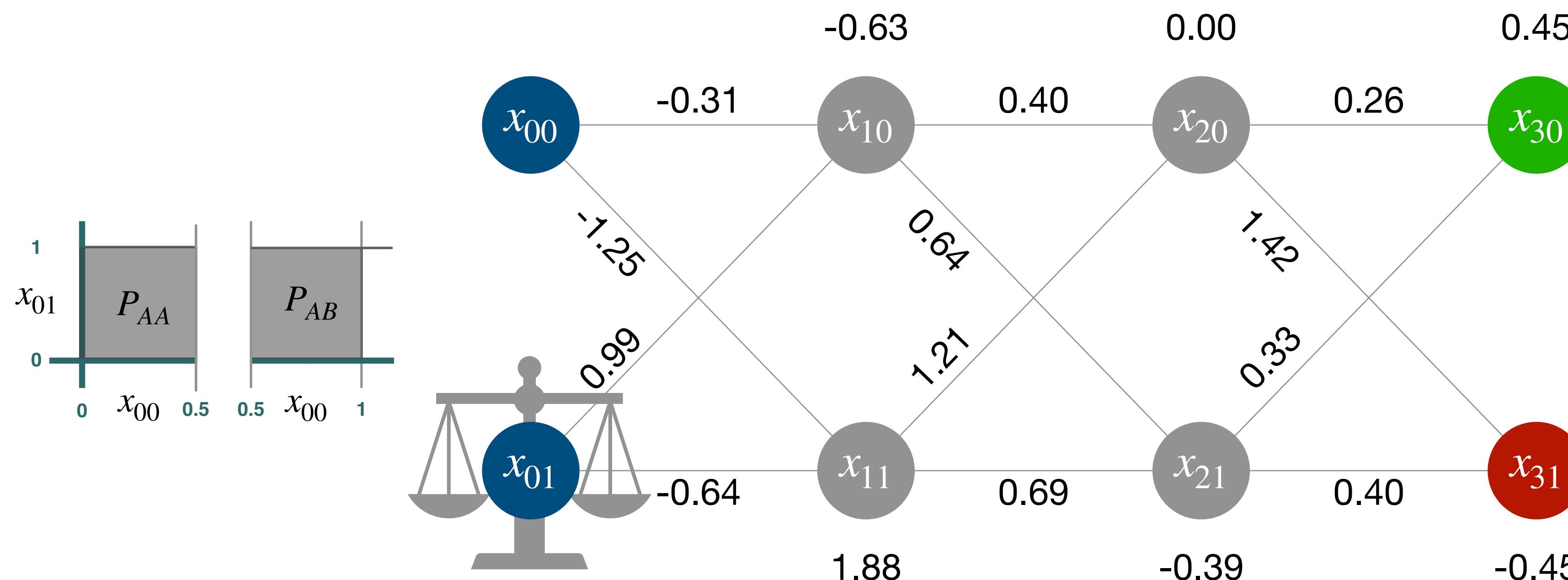
Forward Analysis



Forward Analysis

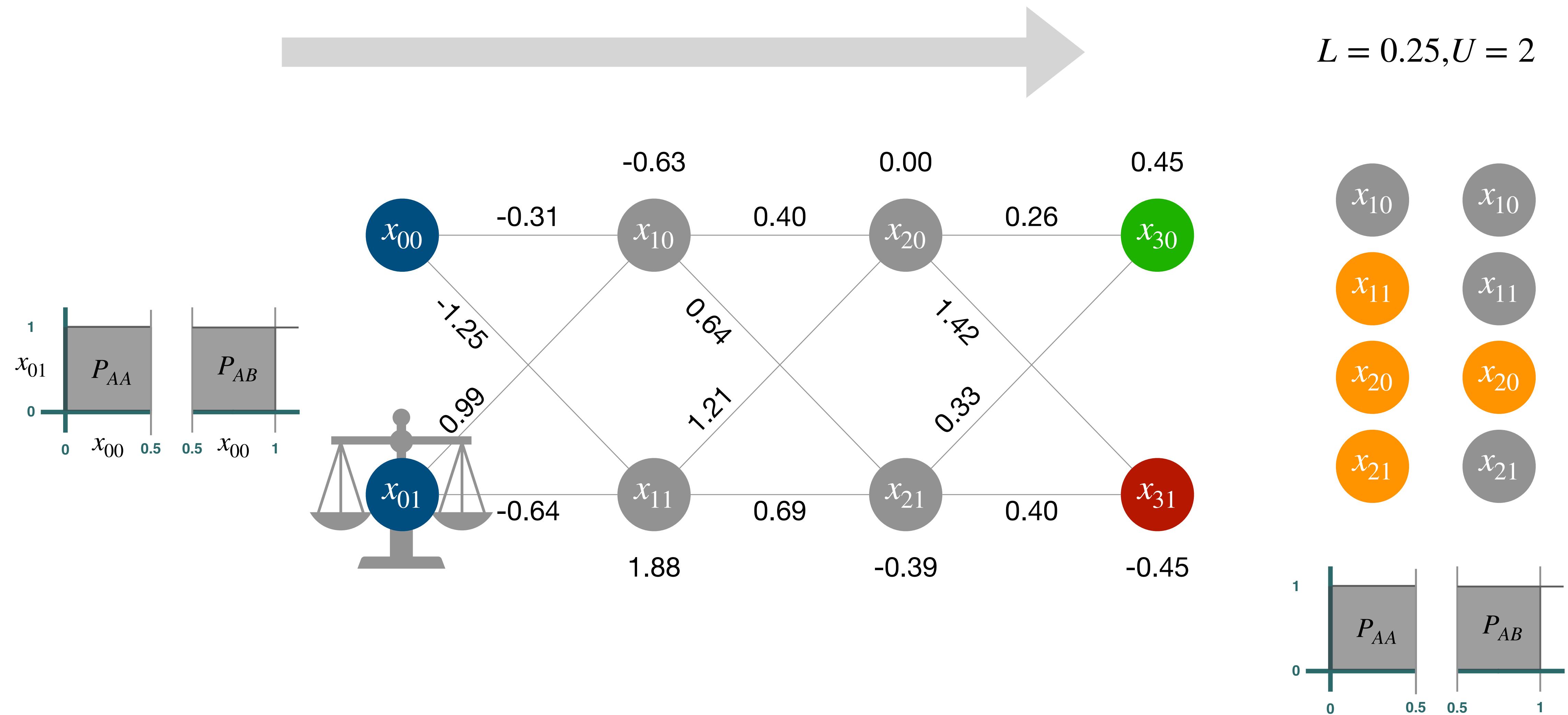


$L = 0.25, U = 2$



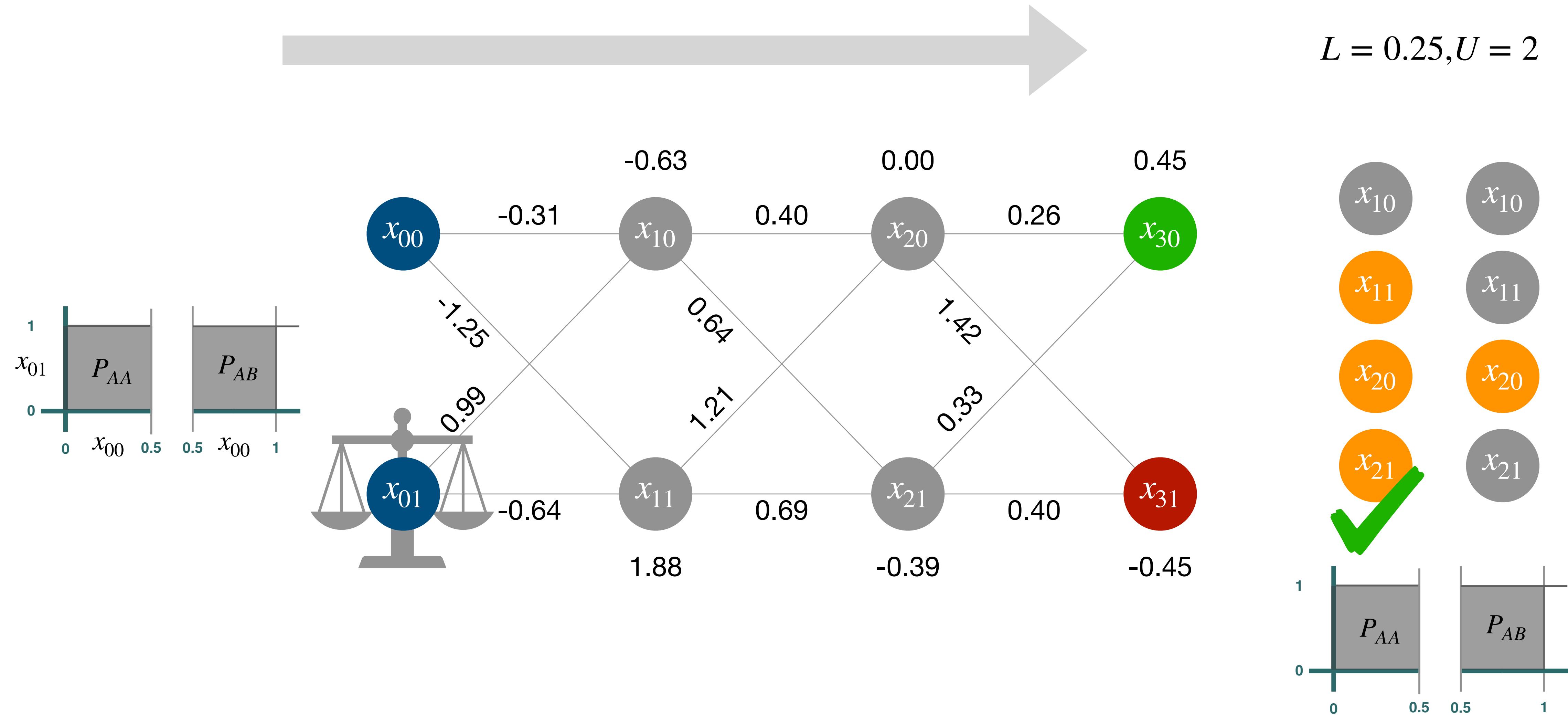
Forward Analysis

$L = 0.25, U = 2$

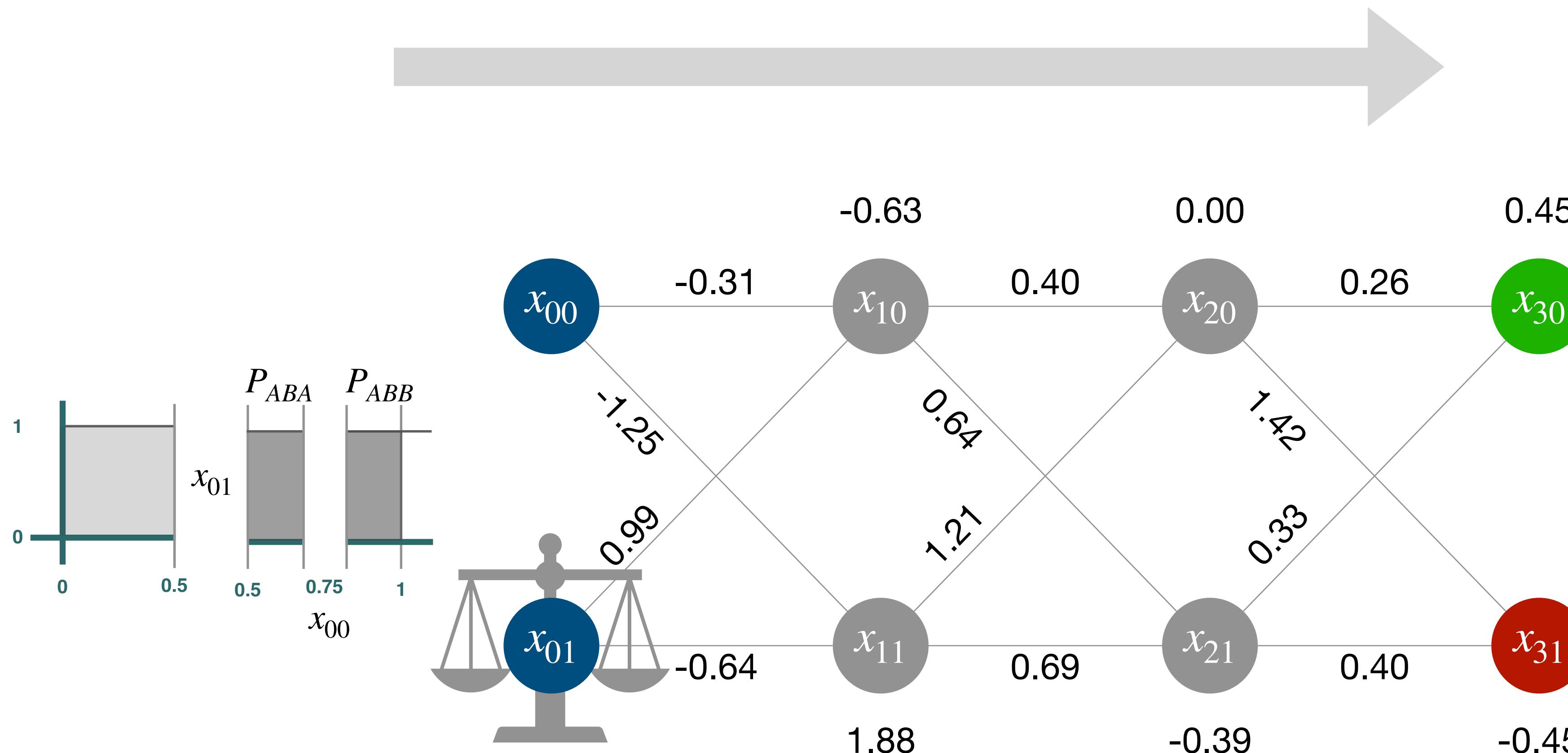


Forward Analysis

$L = 0.25, U = 2$

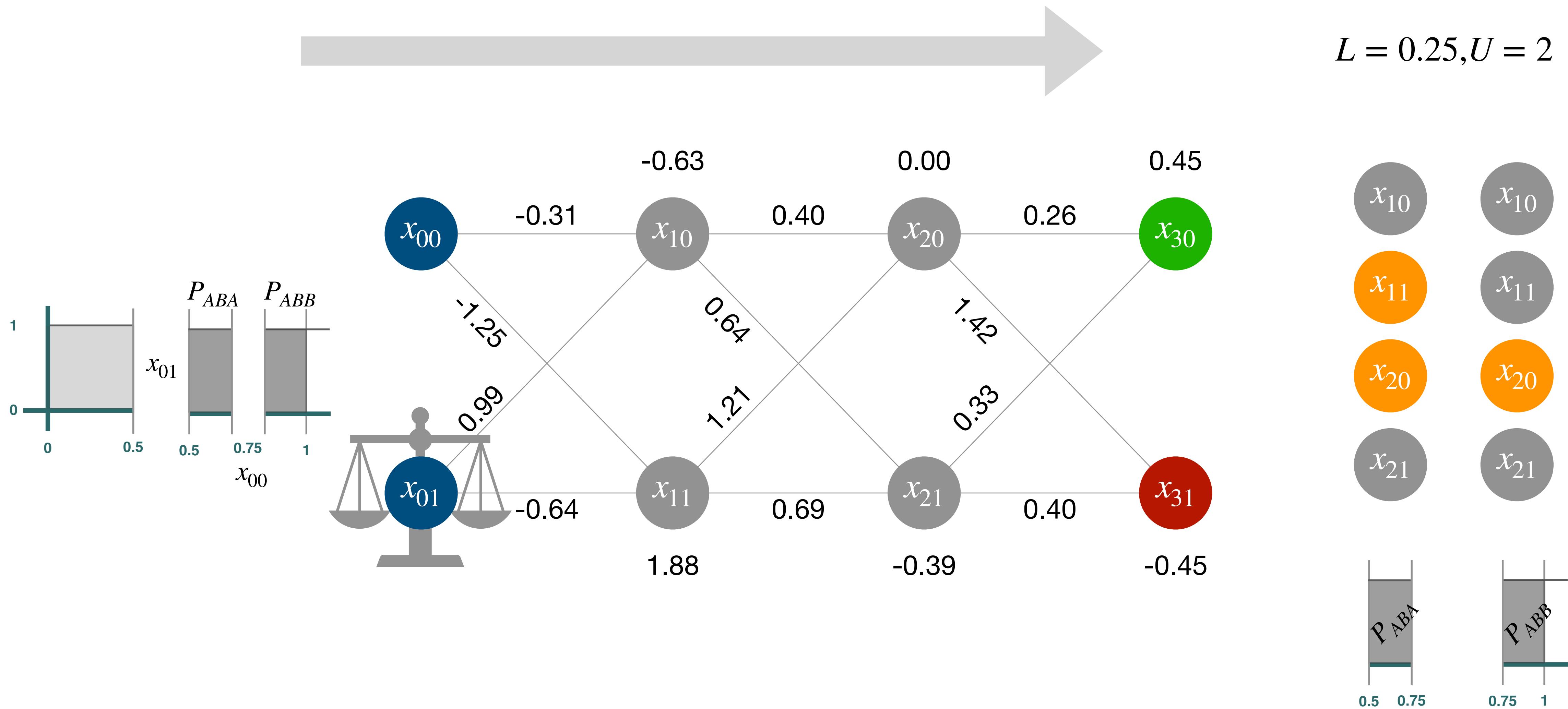


Forward Analysis



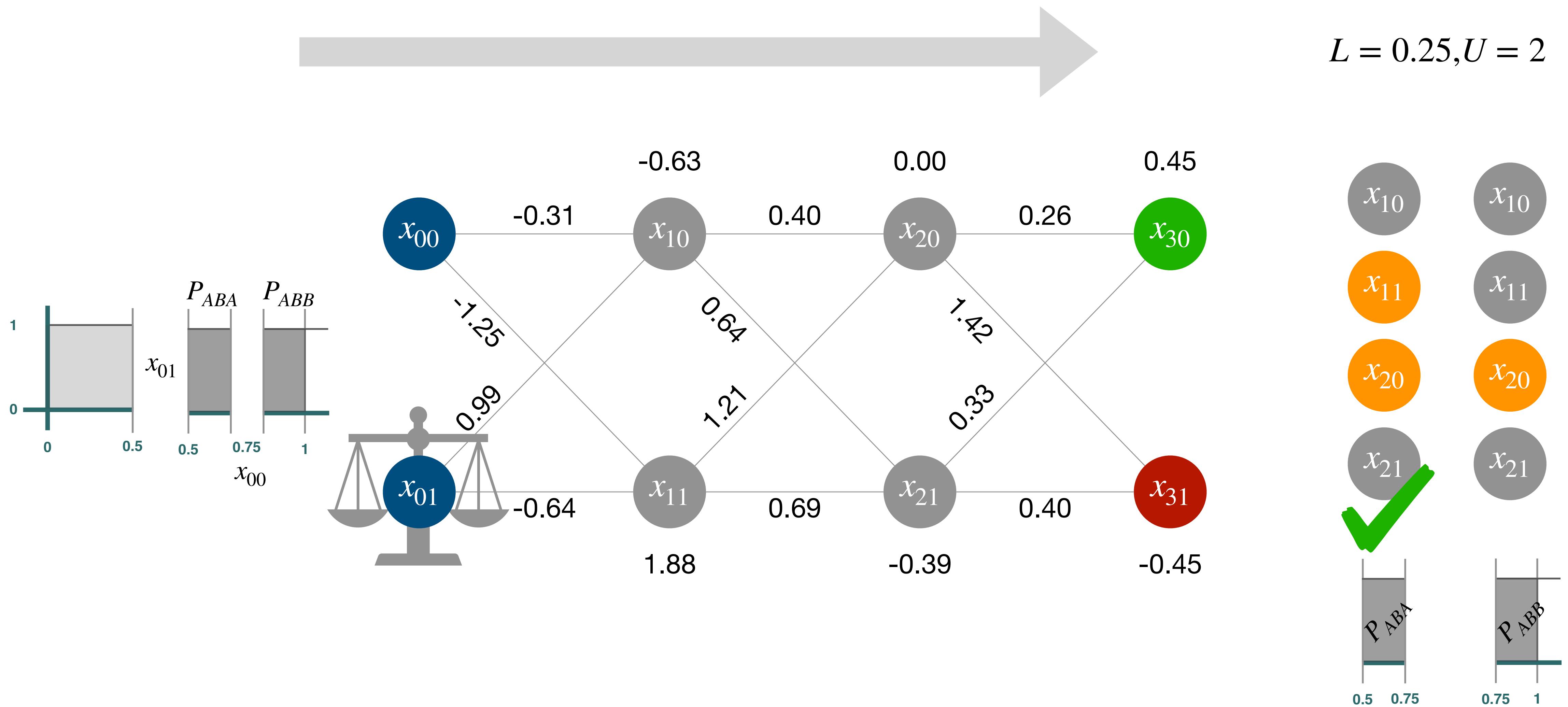
$L = 0.25, U = 2$

Forward Analysis



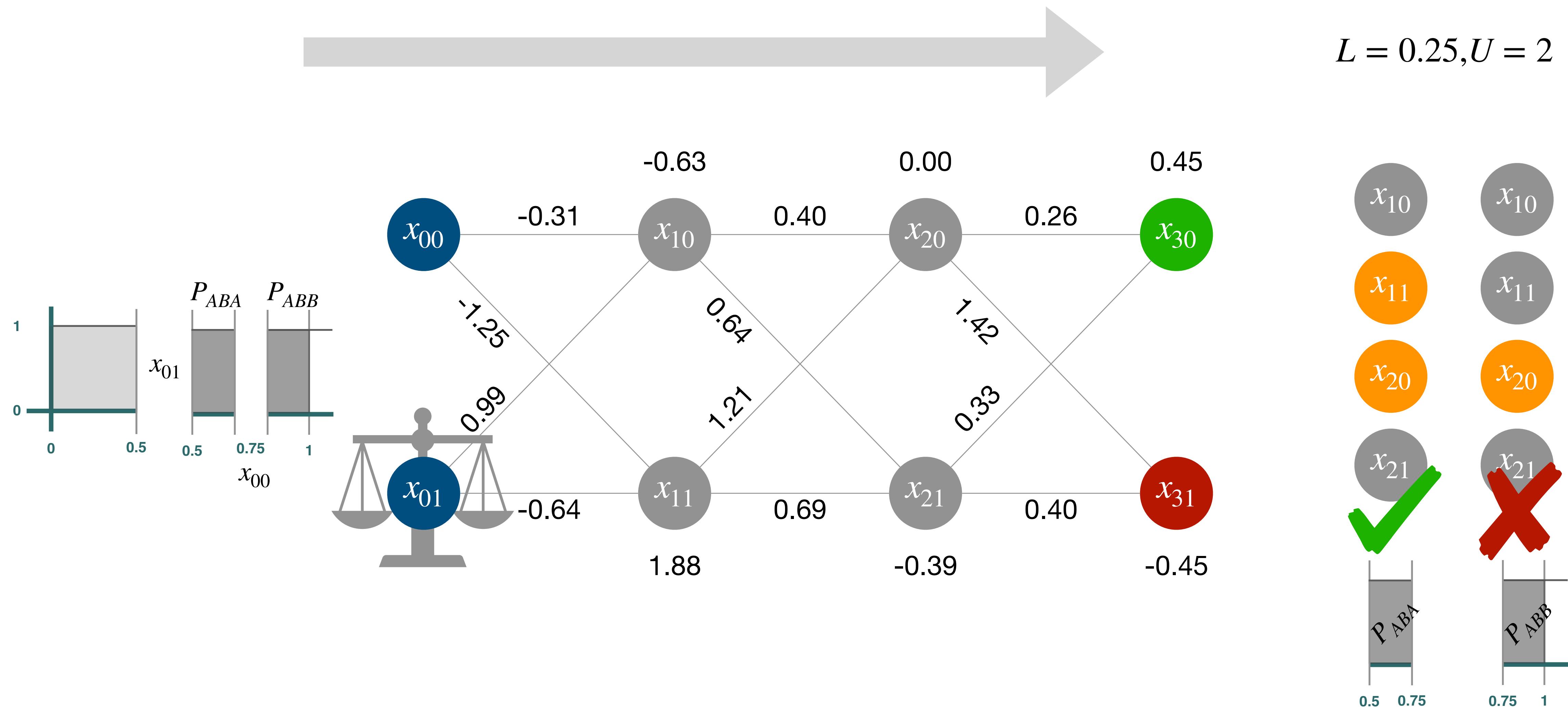
Forward Analysis

$L = 0.25, U = 2$

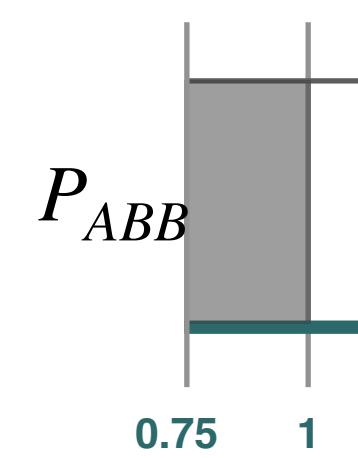
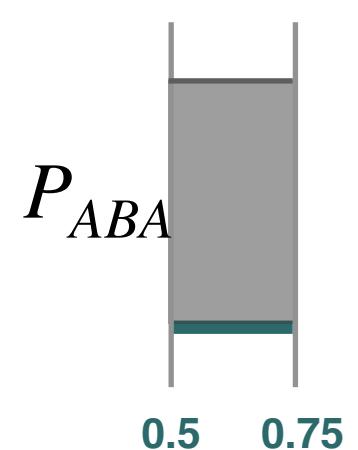
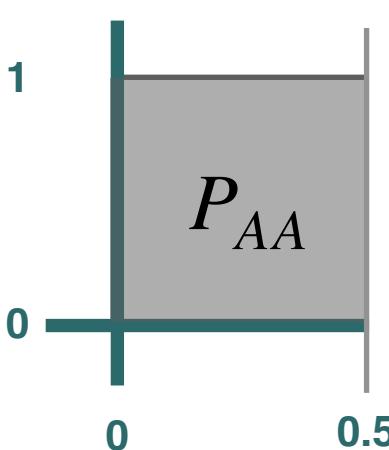


Forward Analysis

$L = 0.25, U = 2$



Forward Analysis



Forward Analysis



x_{10}

x_{11}

x_{20}

x_{21}

x_{10}

x_{11}

x_{20}

x_{21}

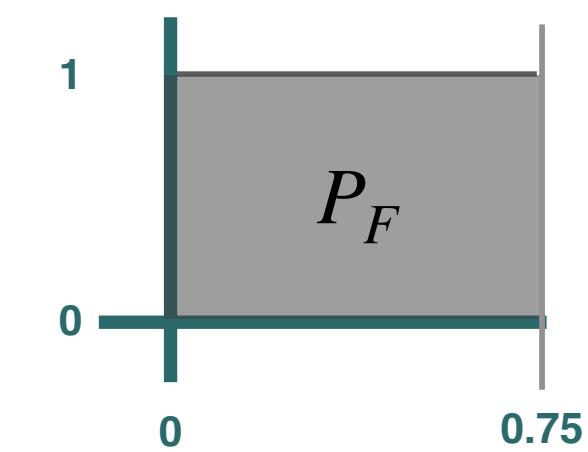
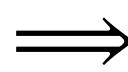
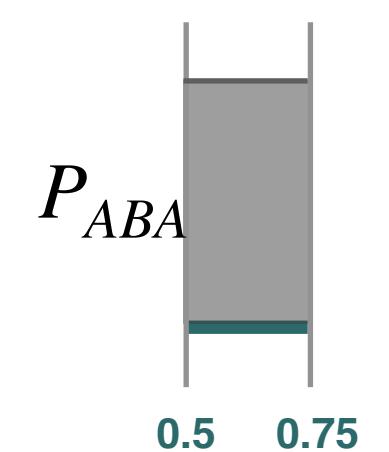
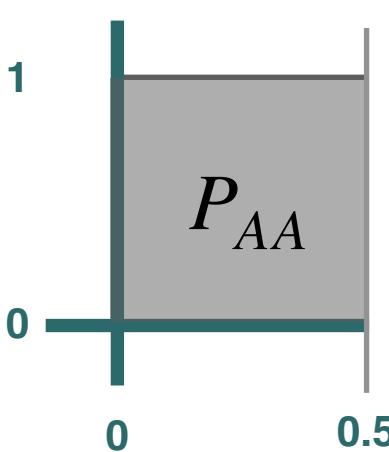


x_{10}

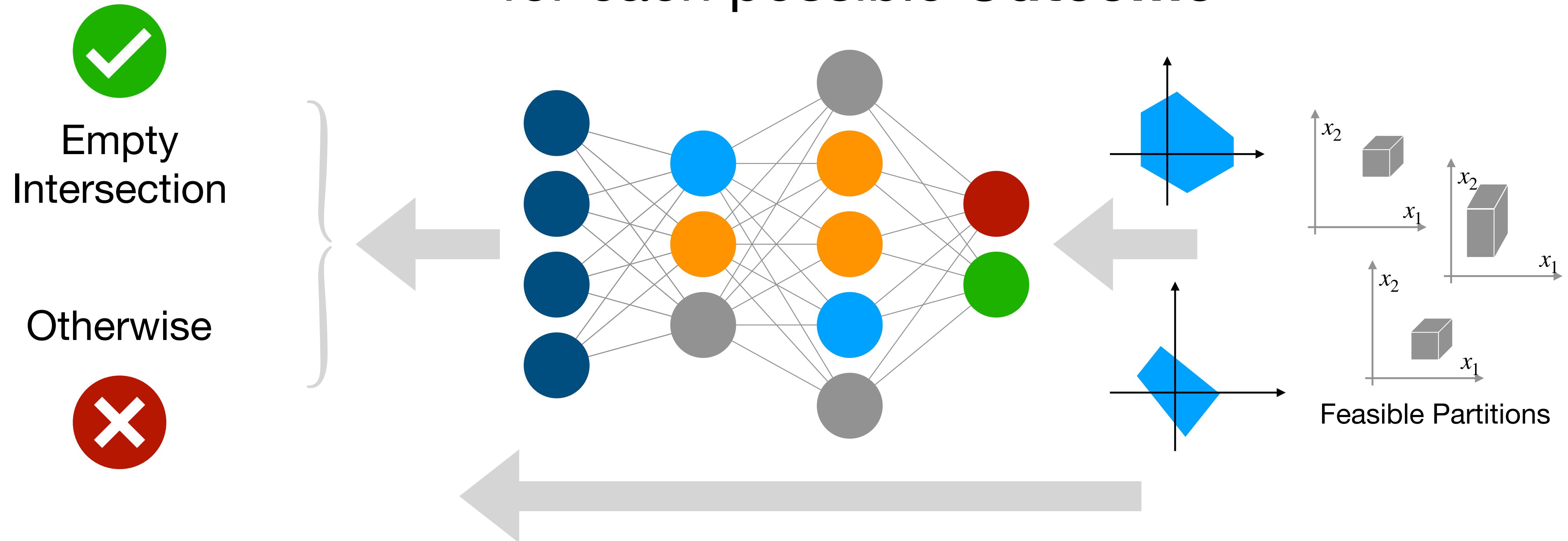
x_{11}

x_{20}

x_{21}

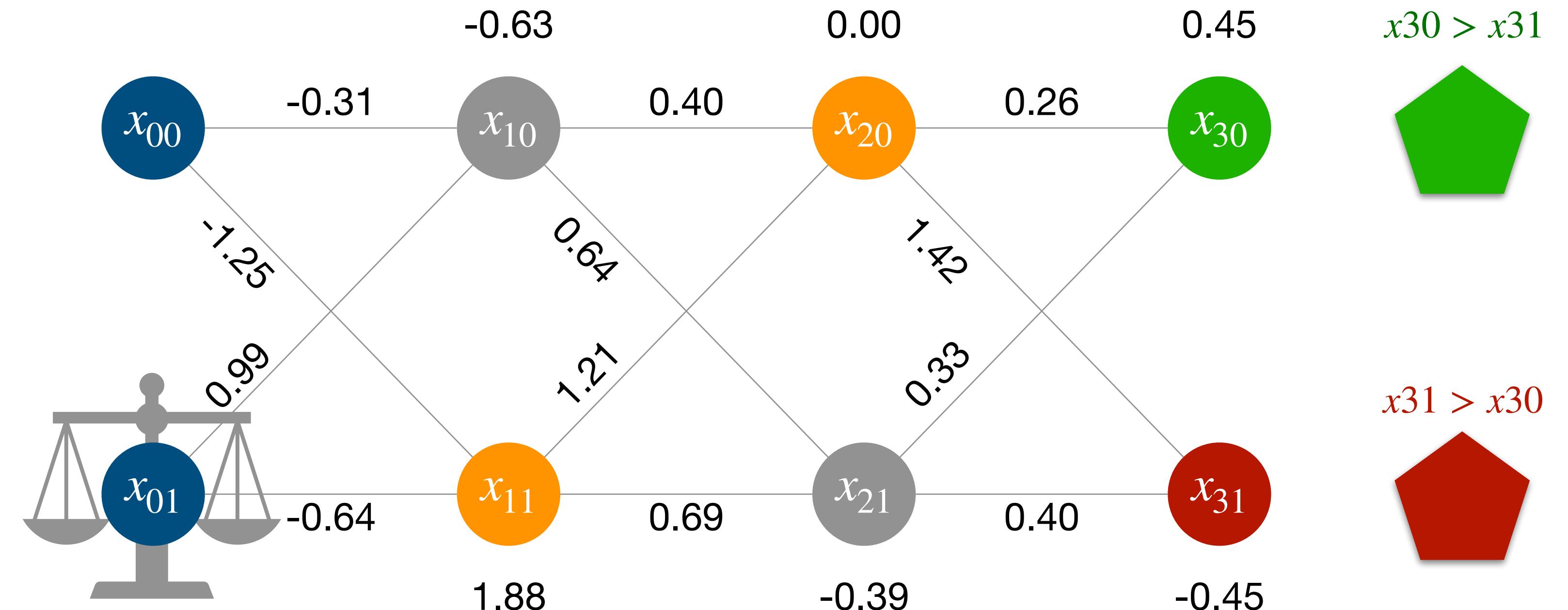
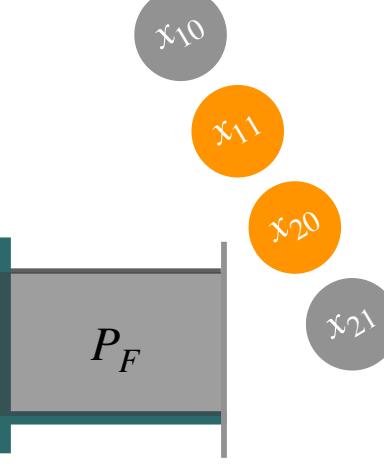


Proceed Backwards
for each **Feasible** partitions
for each possible **Outcome**

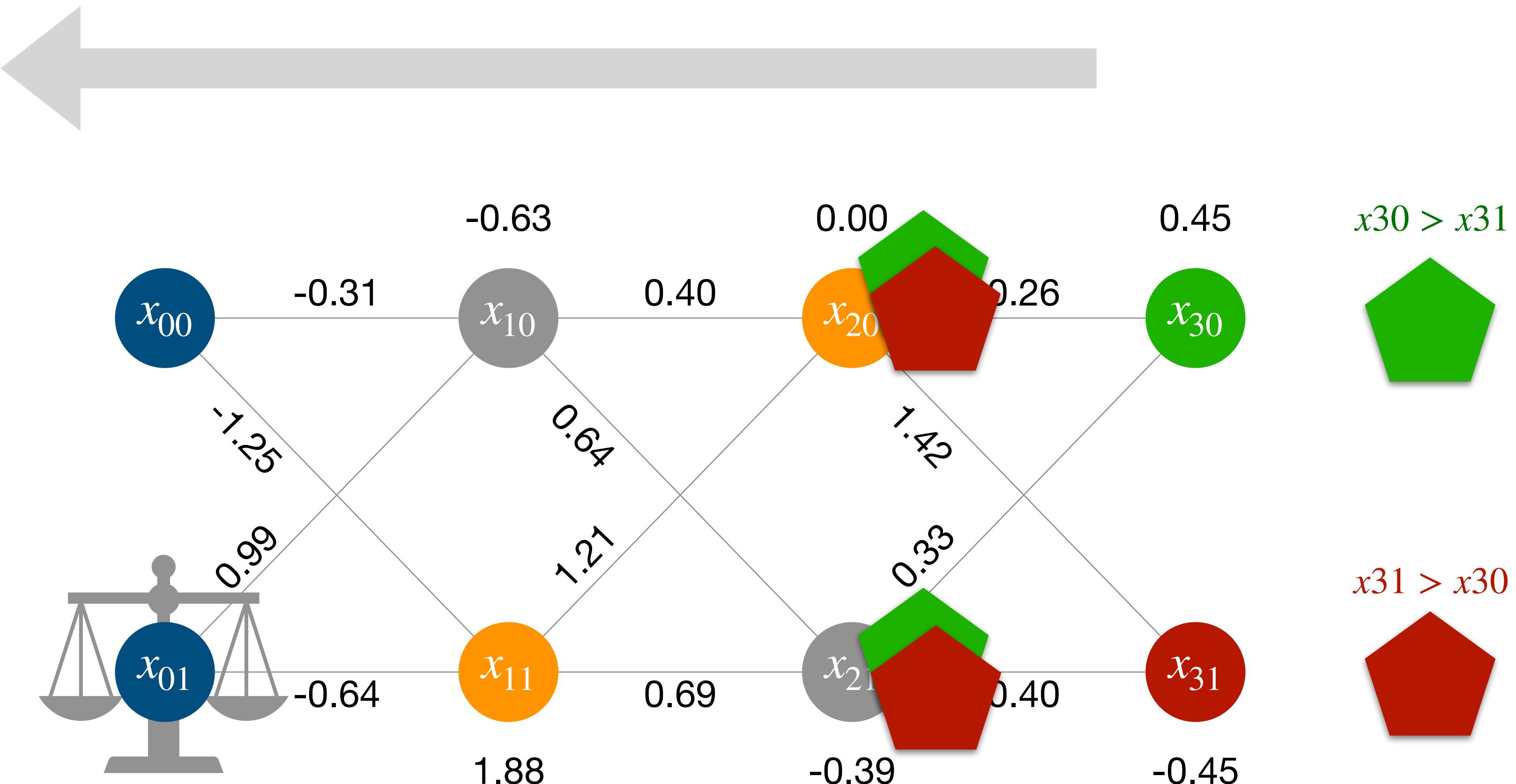


Backward Analysis

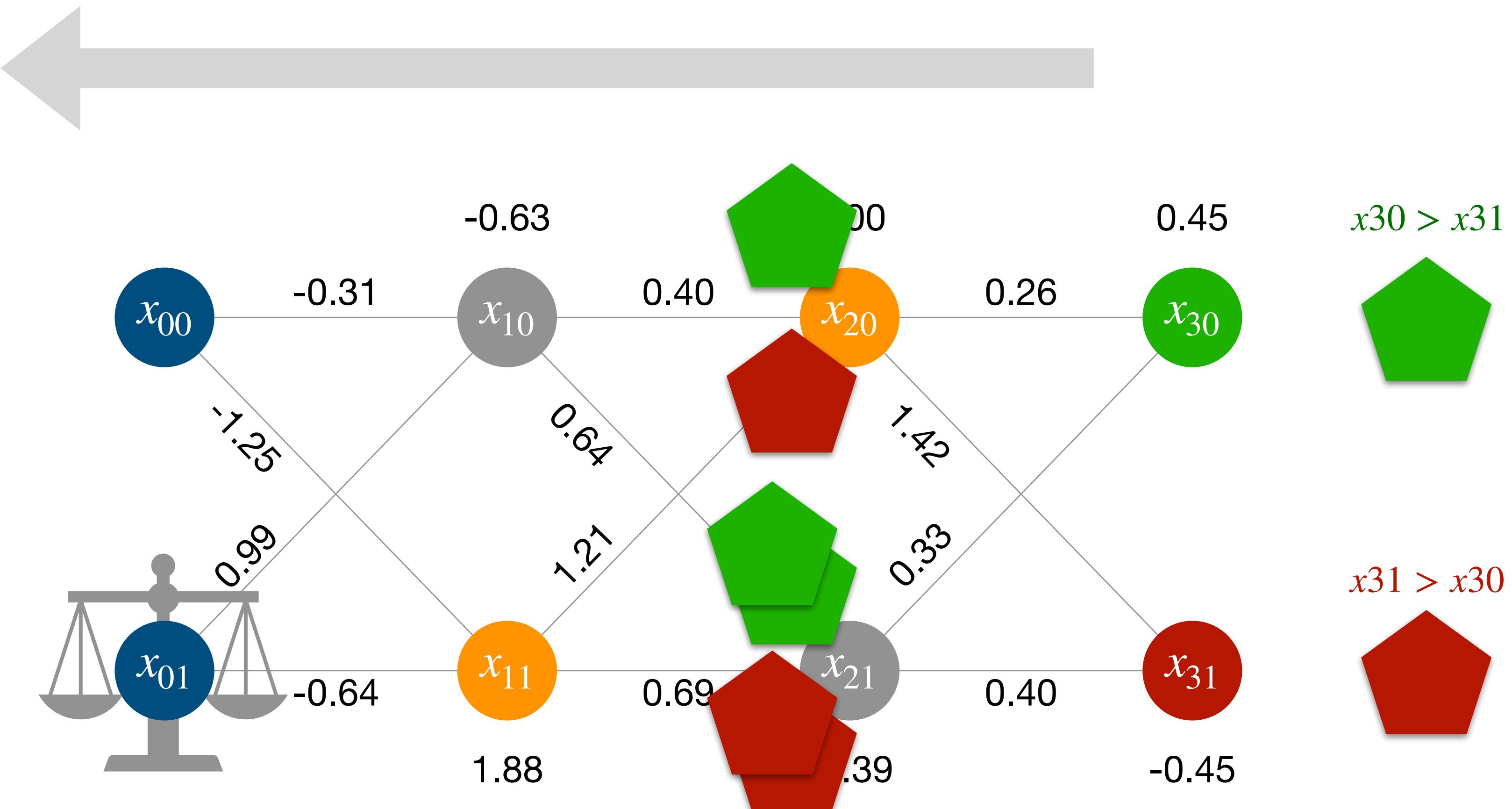
From the forward analysis



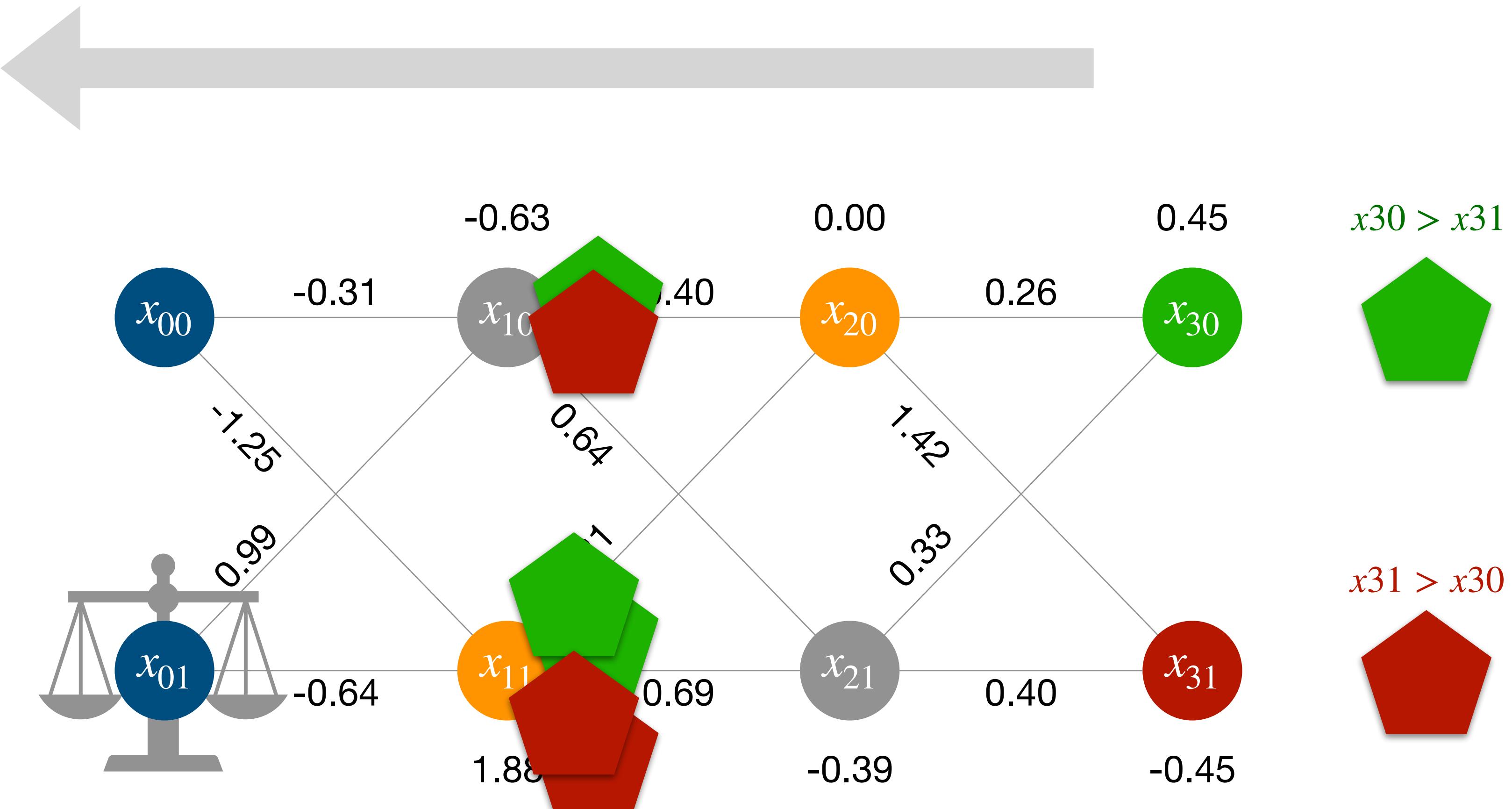
Backward Analysis



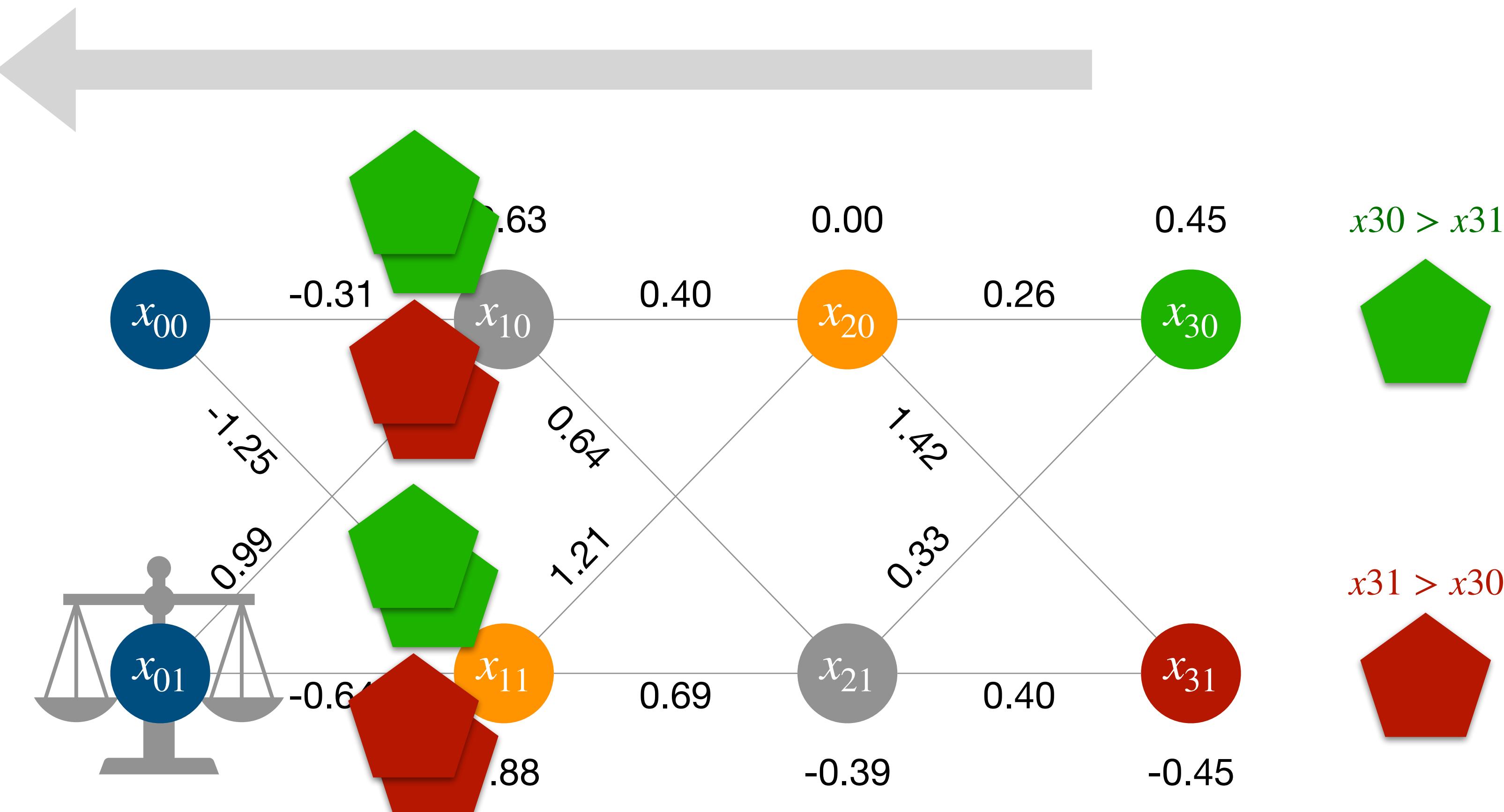
Backward Analysis



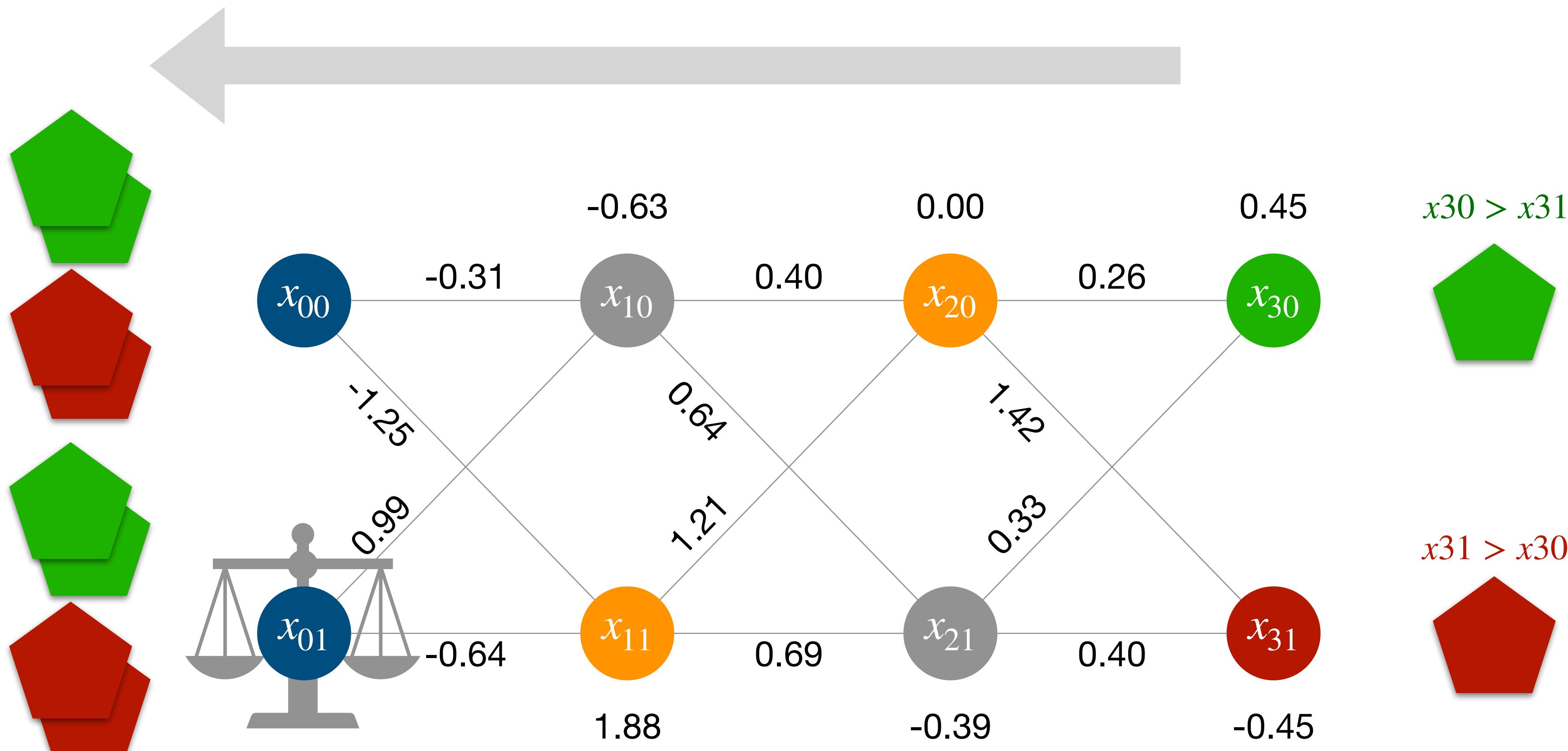
Backward Analysis



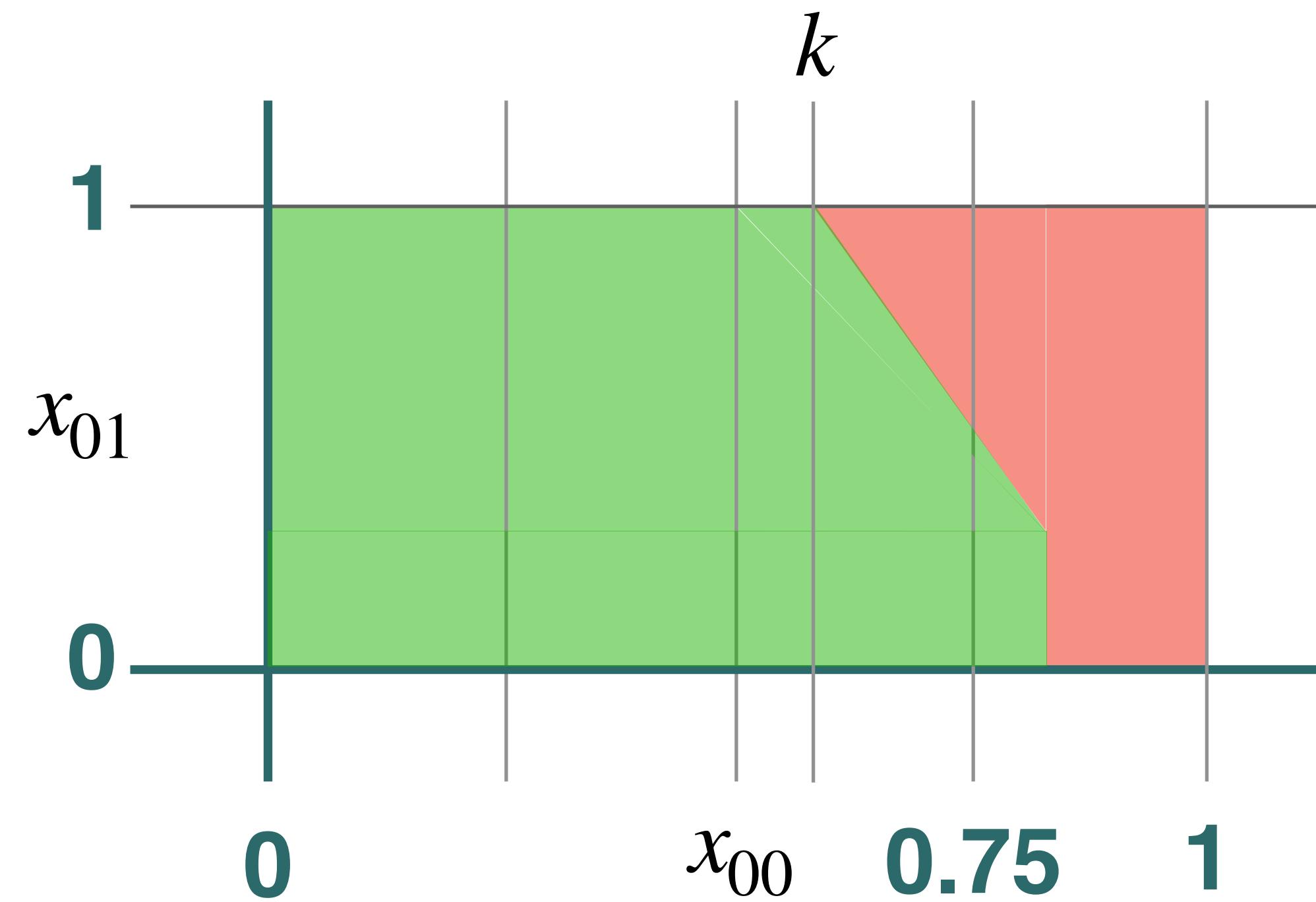
Backward Analysis



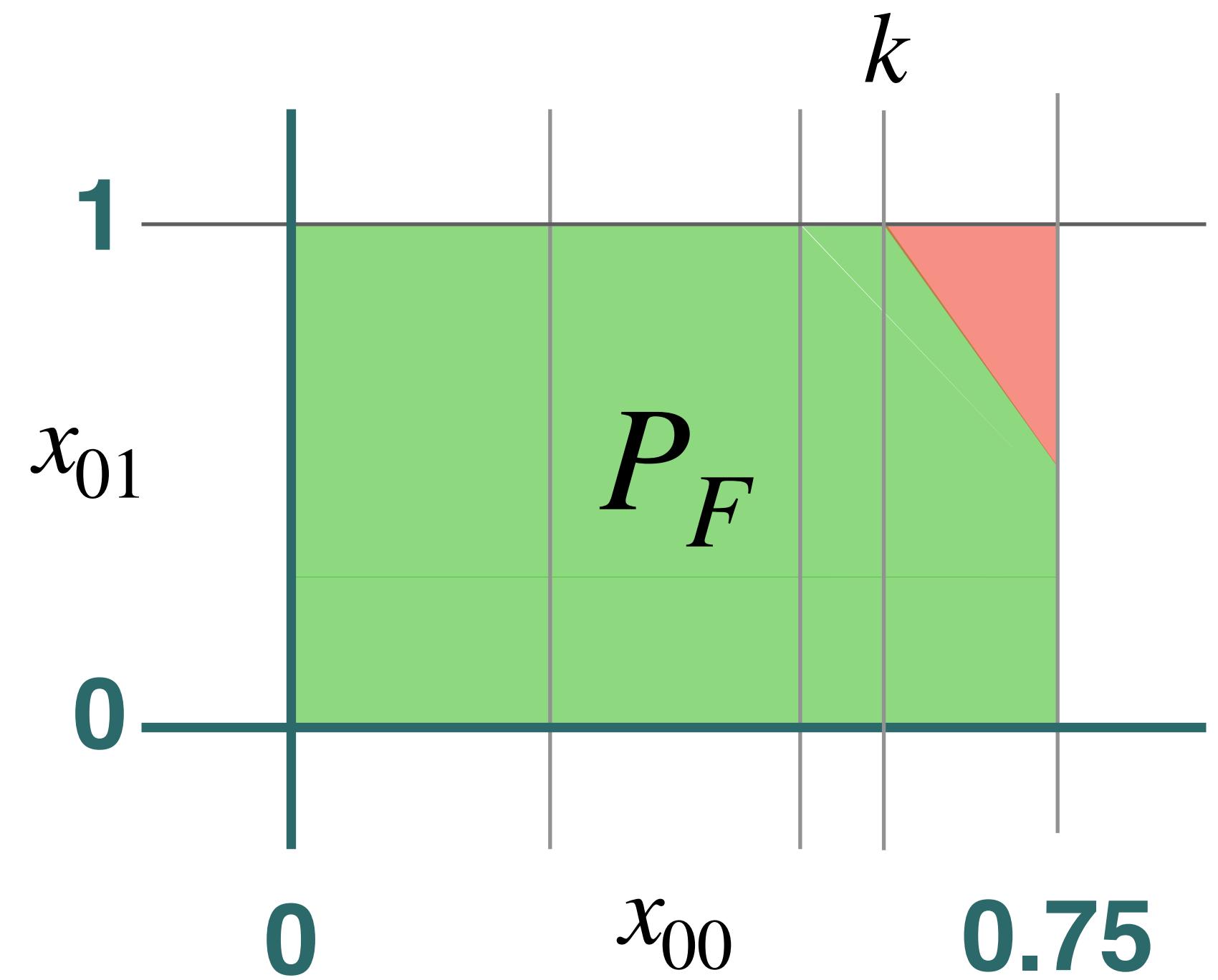
Backward Analysis



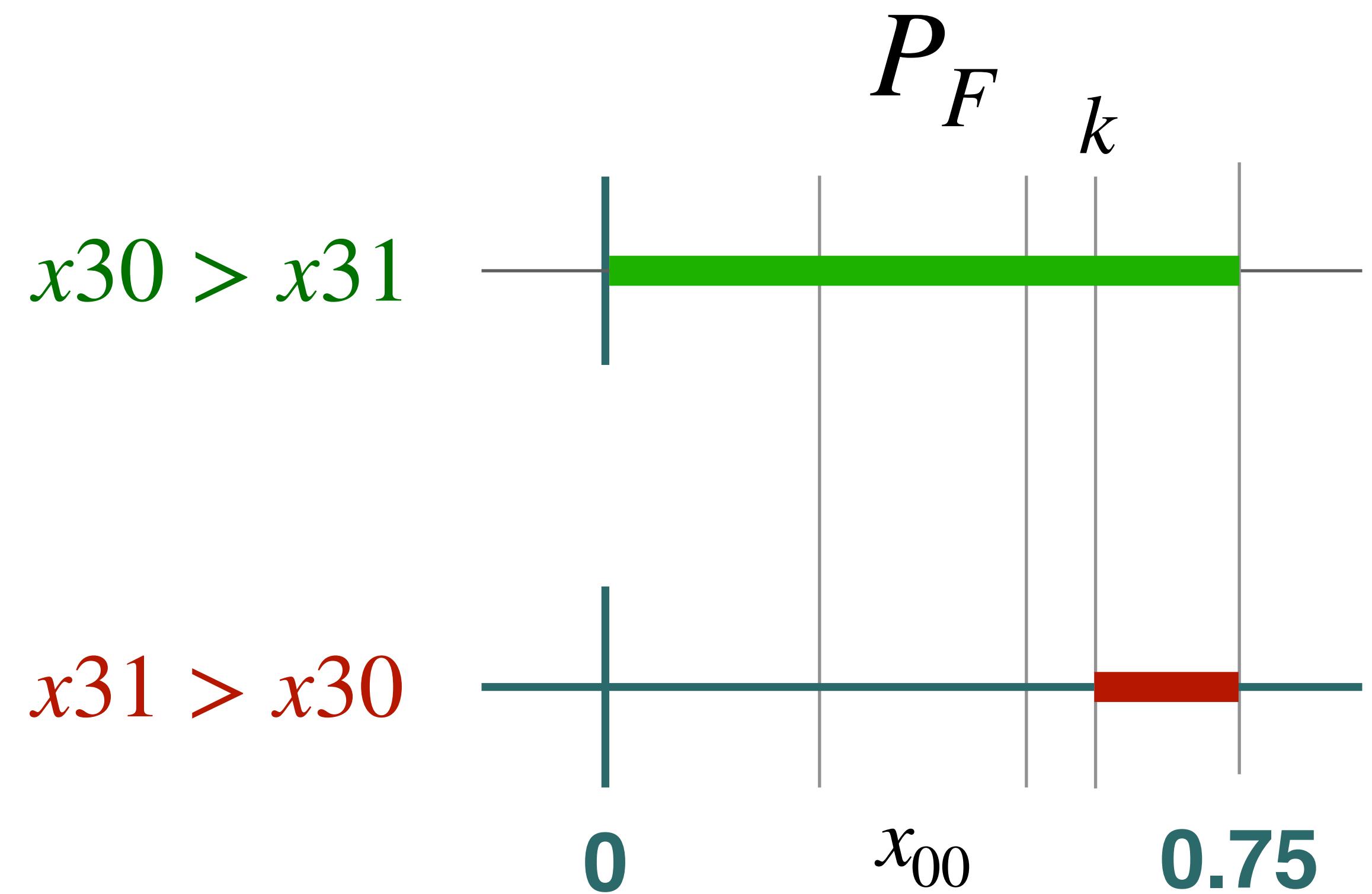
Analysis Result



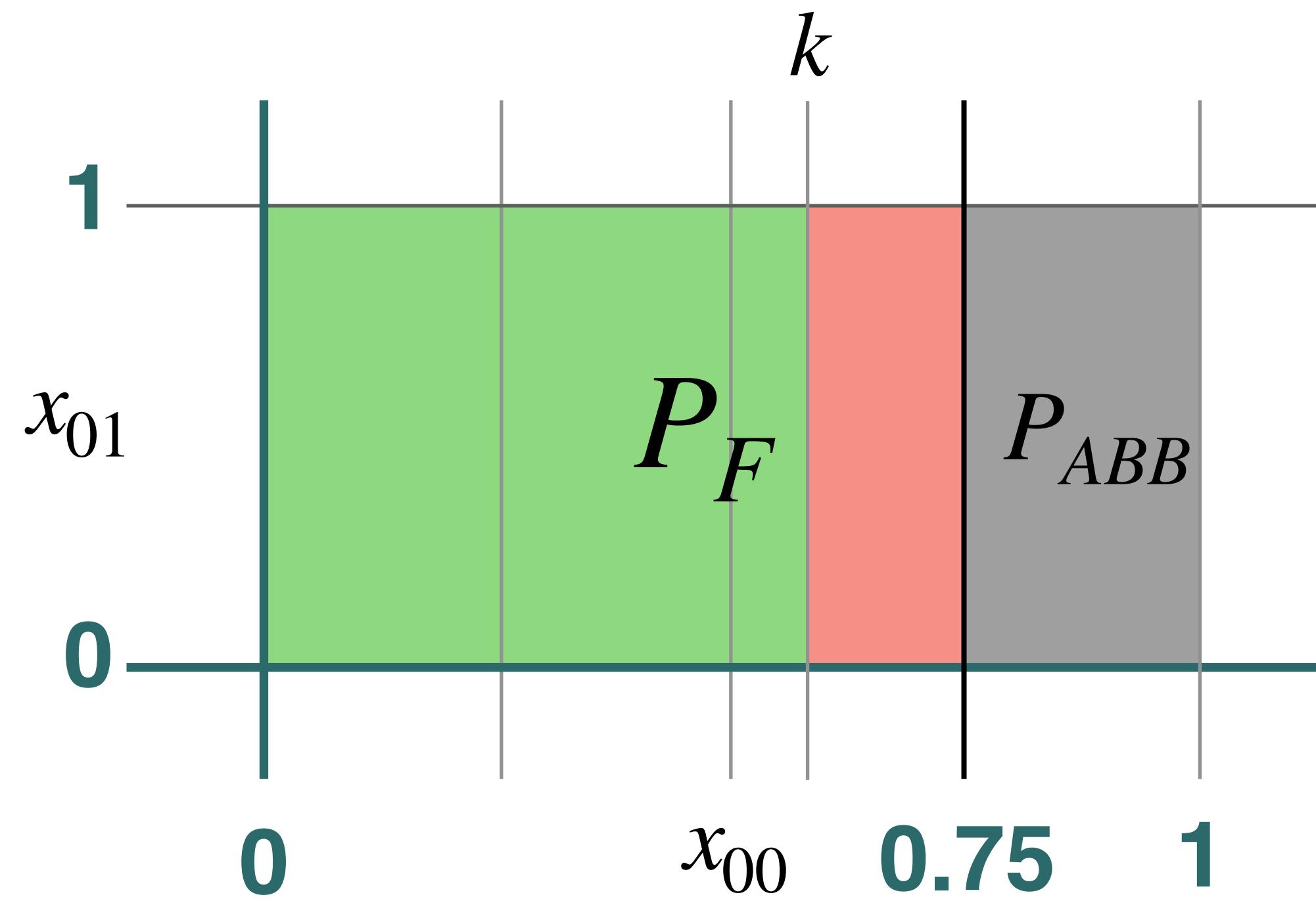
Analysis Result

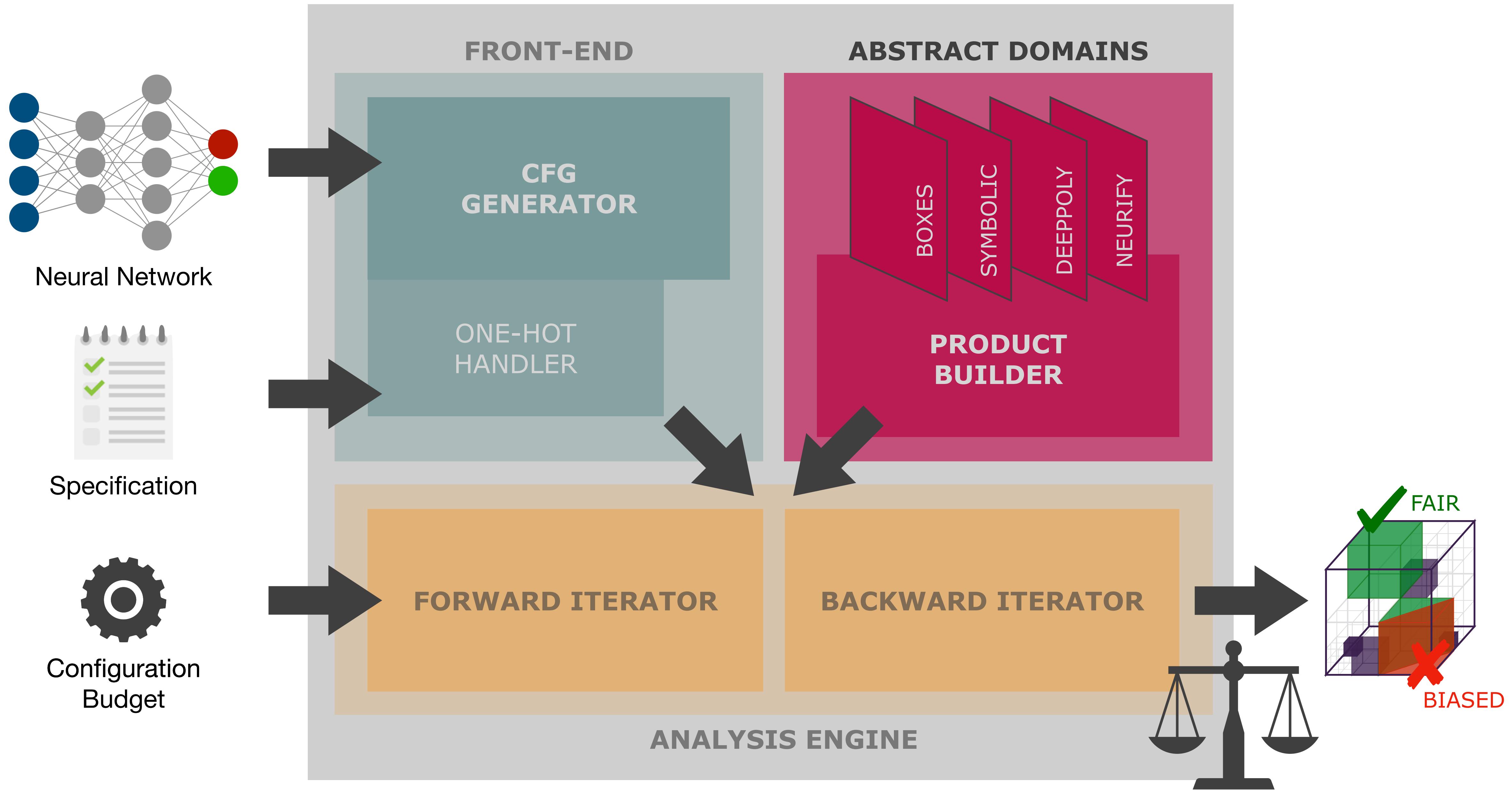


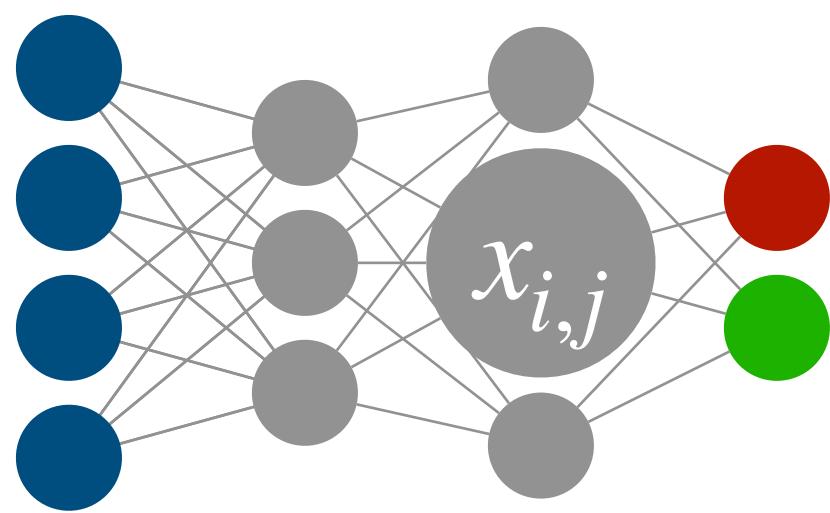
Analysis Result



Analysis Result





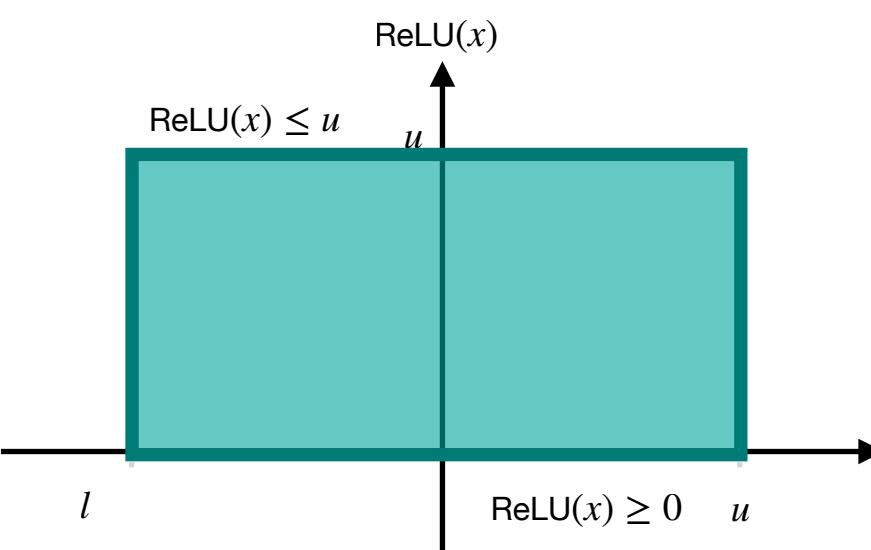
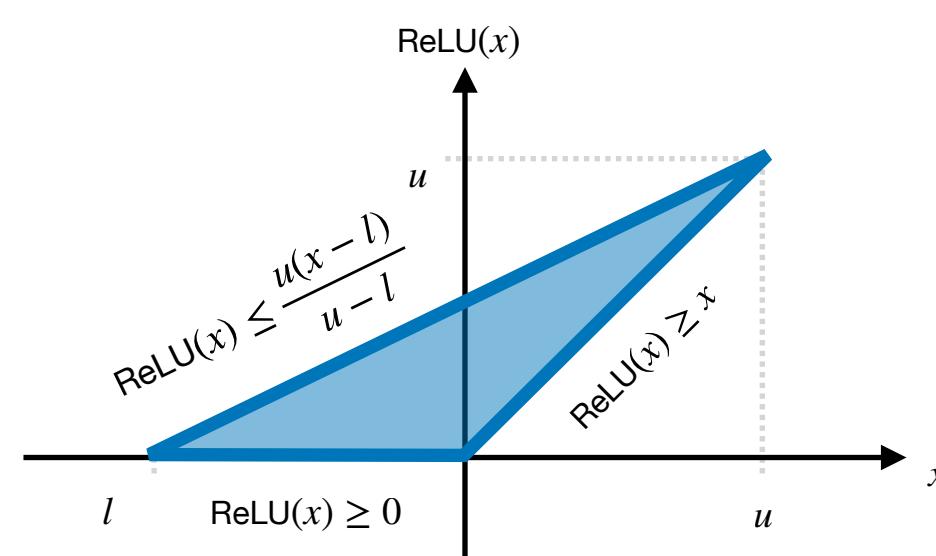


Symbolic

Li et al. @ SAS 2019

$$[l, u]$$

$$\sum_k m_k \cdot x_k + q$$

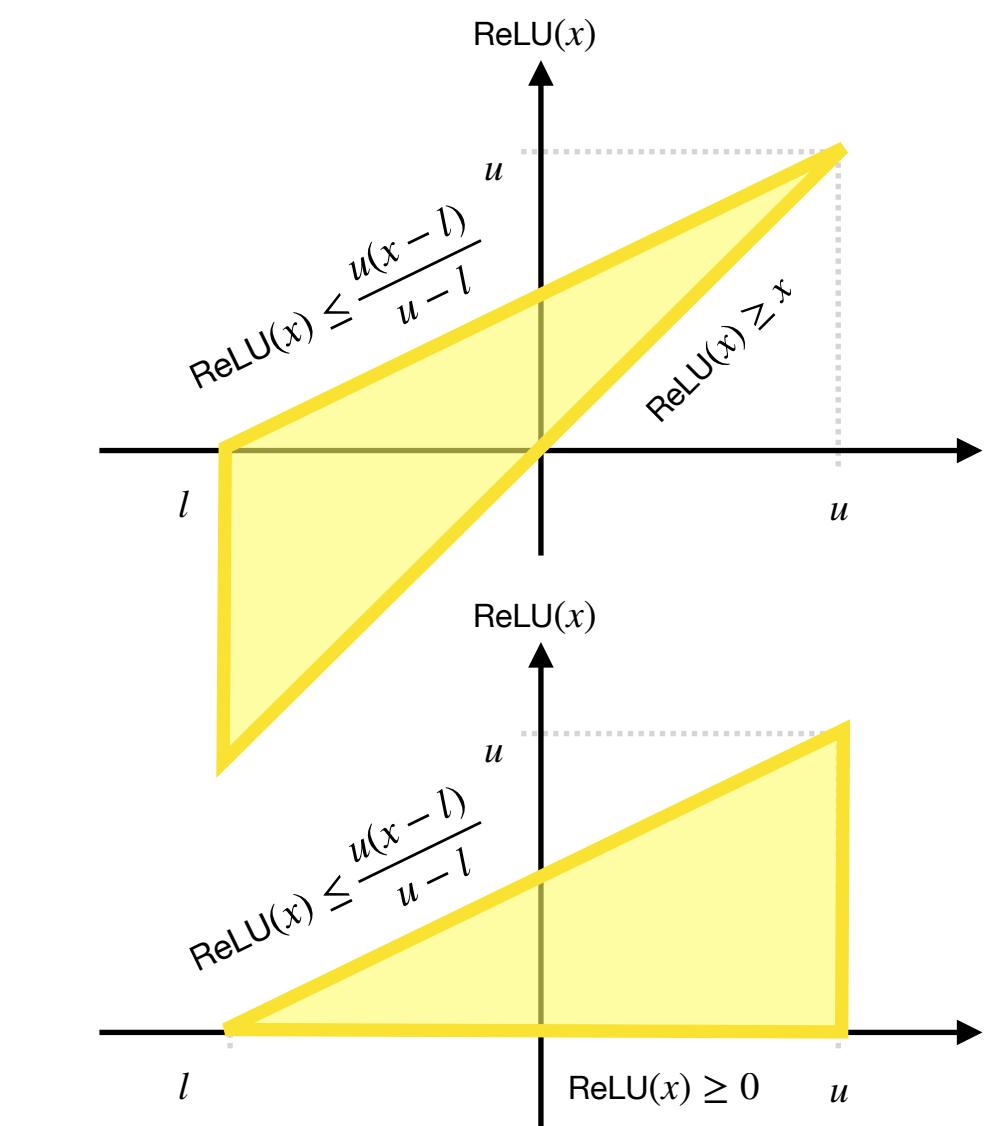


DeepPoly

Singh et al. @ POPL 2019

$$[l, u]$$

$$[\text{eq}_{\text{low}}, \text{eq}_{\text{up}}]$$

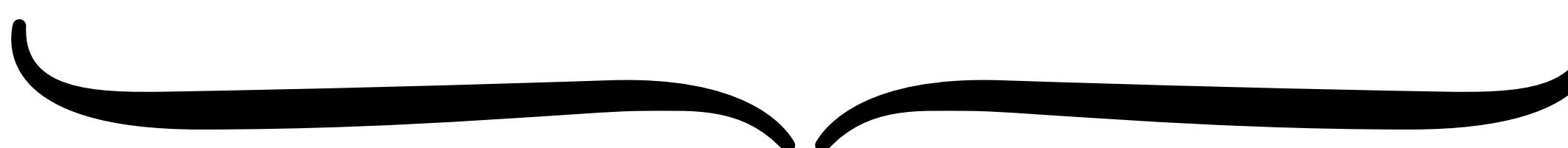
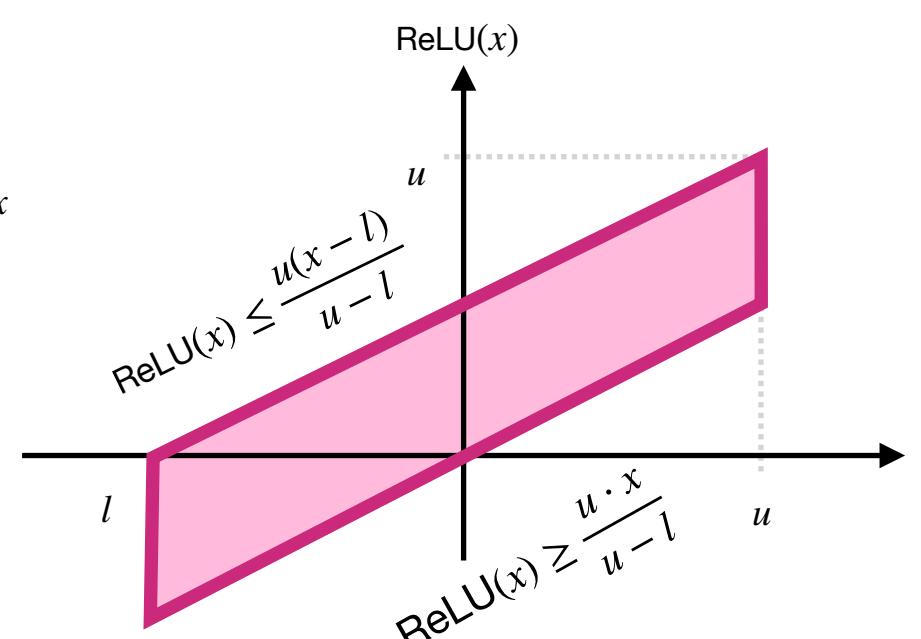


Neurify

Wang et al. @ NeurIPS 2018

$$[l_{\text{low}}, l_{\text{up}}, u_{\text{low}}, u_{\text{up}}]$$

$$[\text{eq}_{\text{low}}, \text{eq}_{\text{up}}]$$

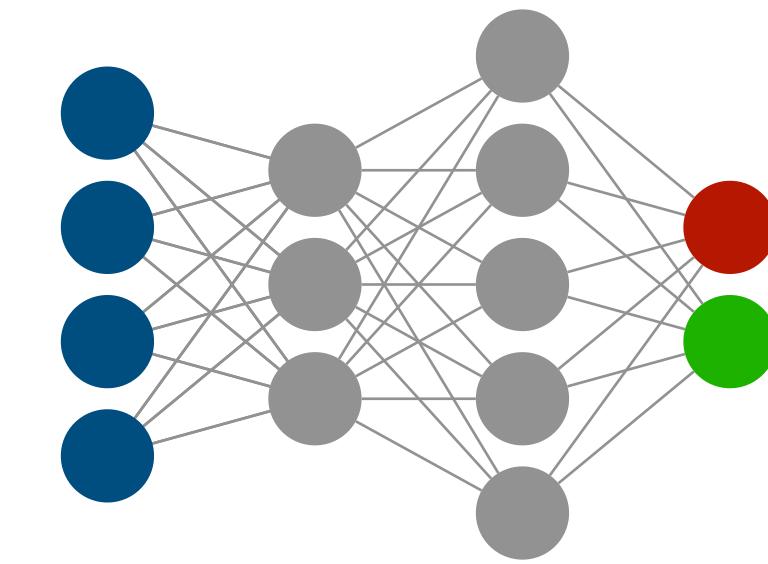


Reduced Product

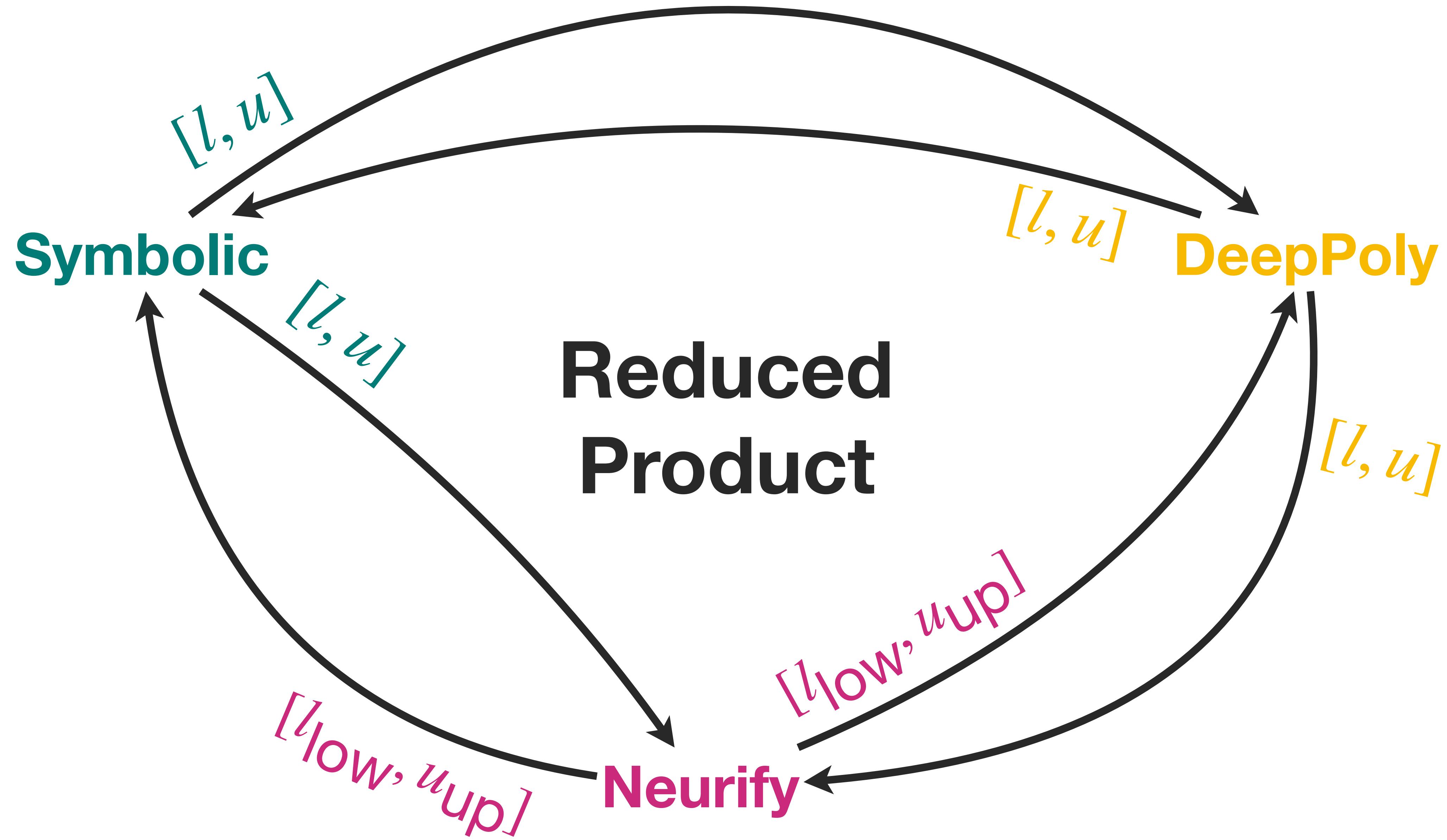
Mazzucato, Urban @ SAS 2021

Precision-vs-Scalability

L	U	Symbolic	DeepPoly	Neurify
0.5	3	48.78%	49.01%	46.49%
	5	56.11%	56.15%	53.06%
0.25	3	83.63%	81.82%	81.40%
	5	91.67%	91.58%	92.33%



- 4 Hidden Layers
- 5 Neuron per Layer
- 23 inputs $\in [0,1]$
- 2 Output classes

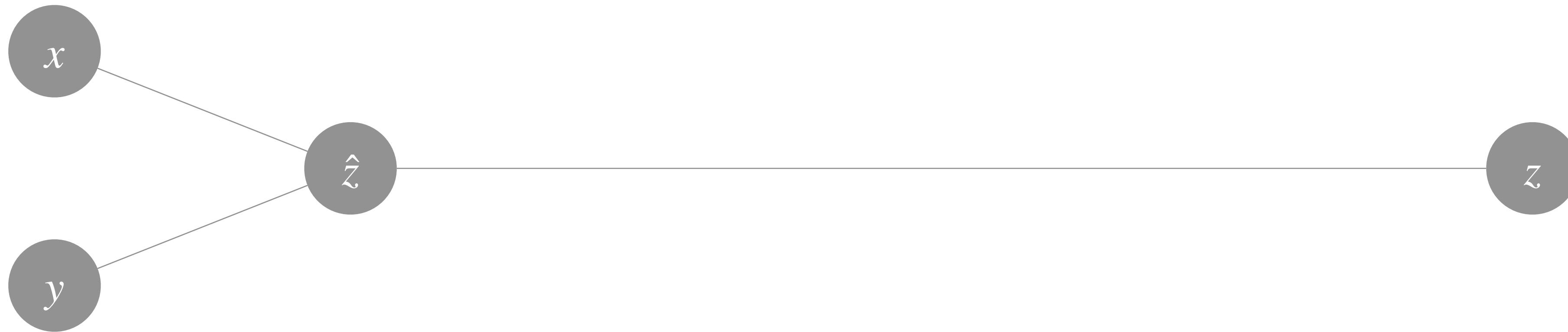


Reduced Product

Mazzucato, Urban @ SAS 2021

$$x \in \begin{cases} [l_x, u_x] \\ [l_x, u_x] \\ [l_{x\text{low}}, l_{x\text{up}}, u_{x\text{low}}, u_{x\text{up}}] \end{cases}$$

Symbolic
DeepPoly
Neurify

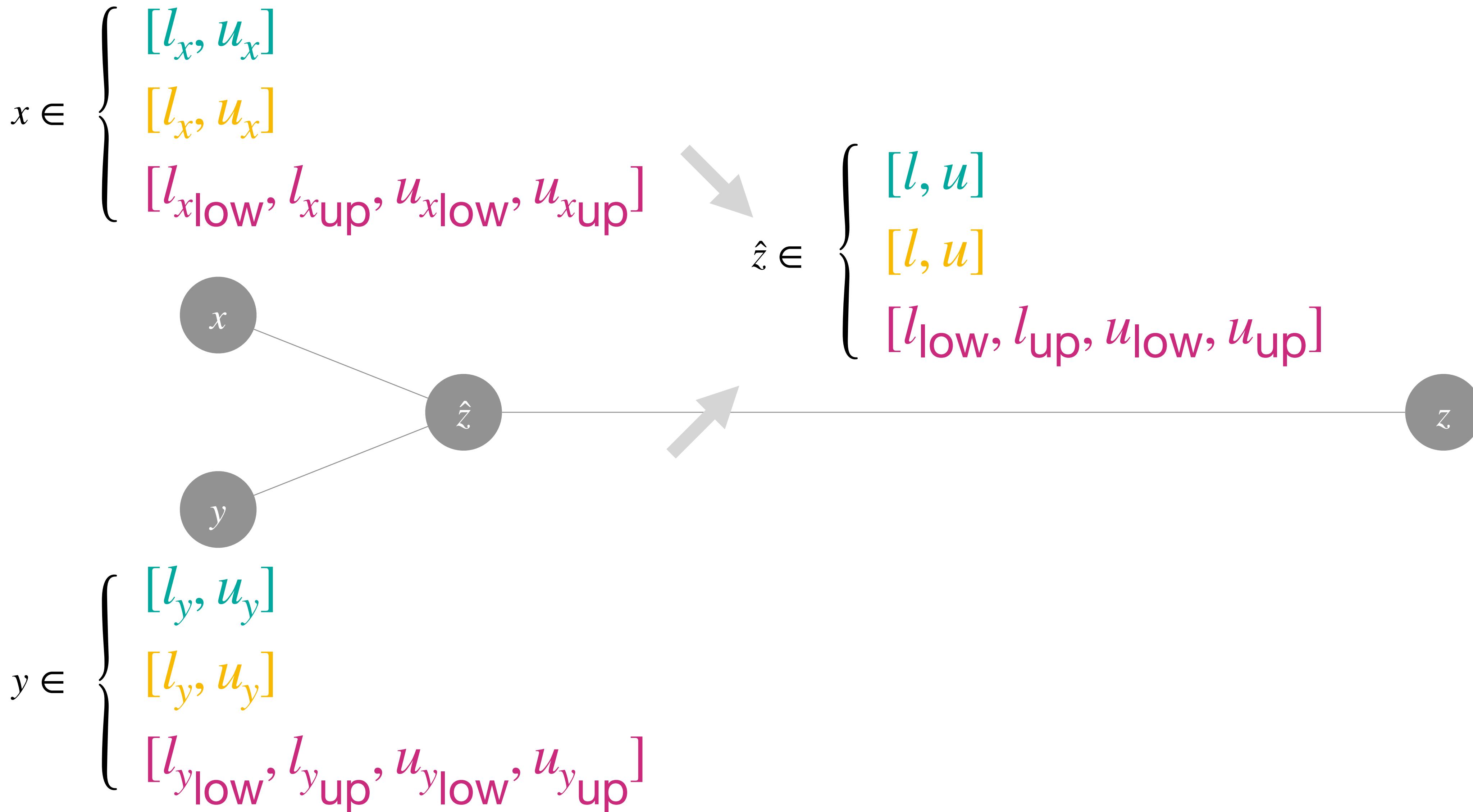


$$y \in \begin{cases} [l_y, u_y] \\ [l_y, u_y] \\ [l_{y\text{low}}, l_{y\text{up}}, u_{y\text{low}}, u_{y\text{up}}] \end{cases}$$

Symbolic
DeepPoly
Neurify

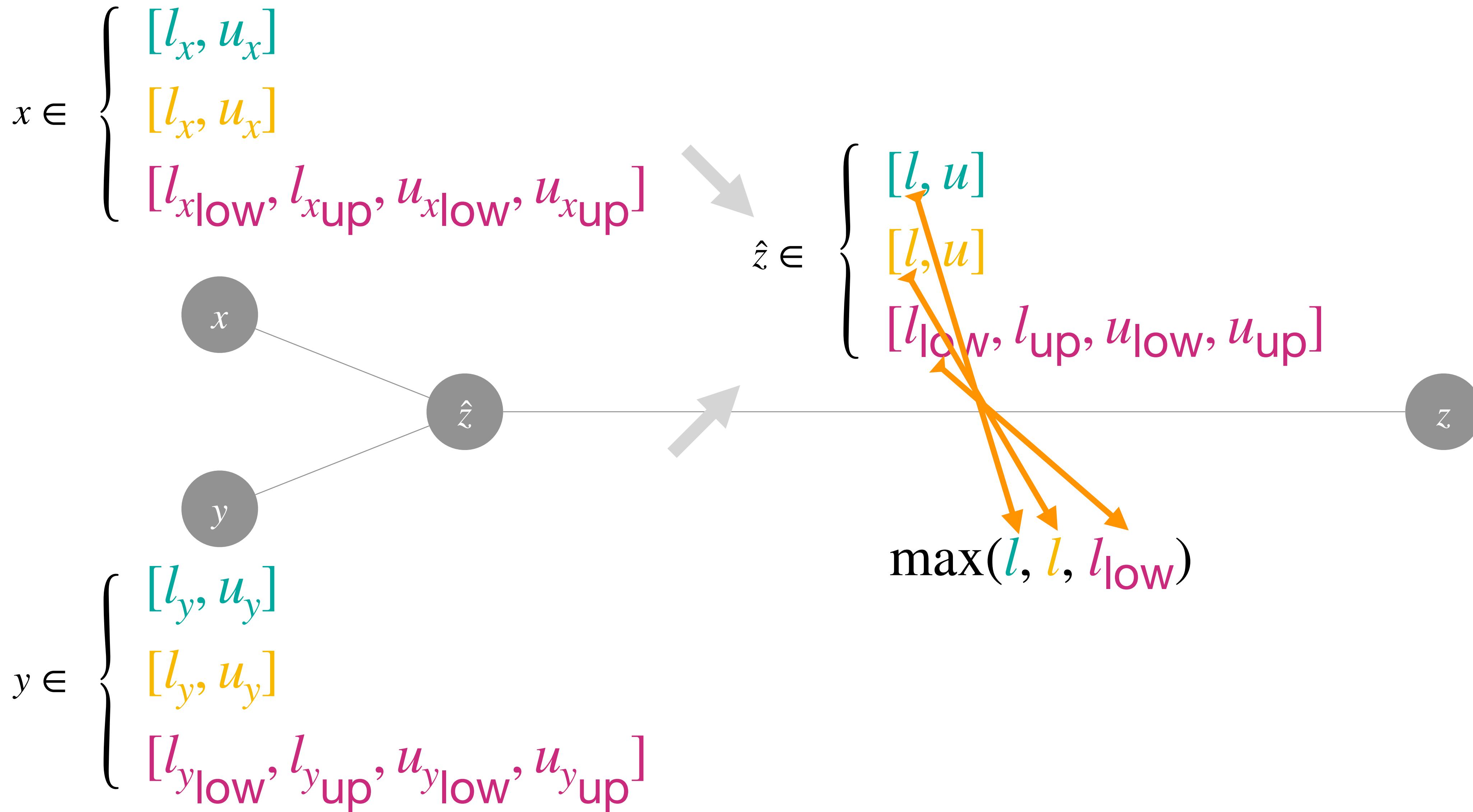
Reduced Product

Mazzucato, Urban @ SAS 2021



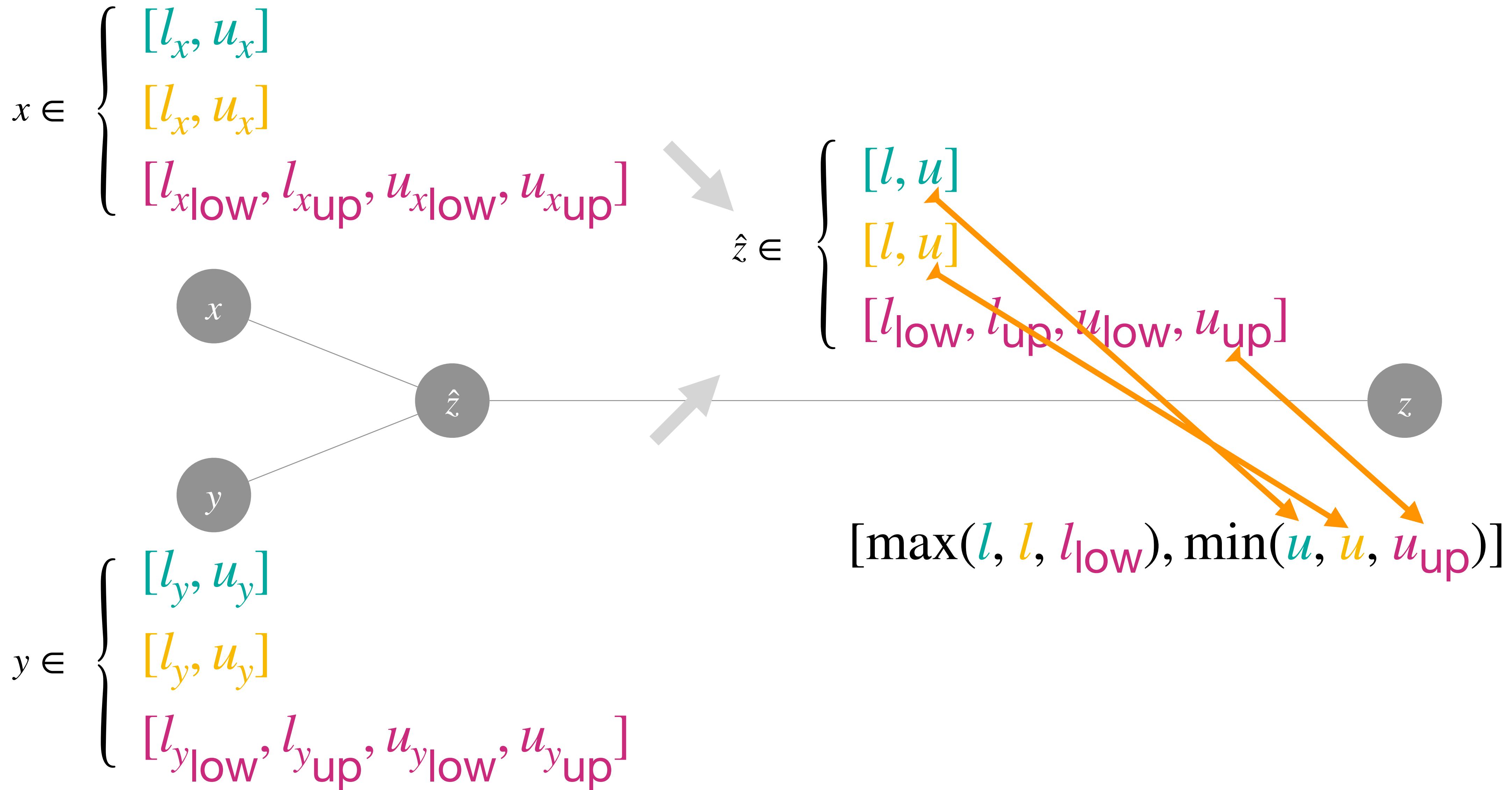
Reduced Product

Mazzucato, Urban @ SAS 2021



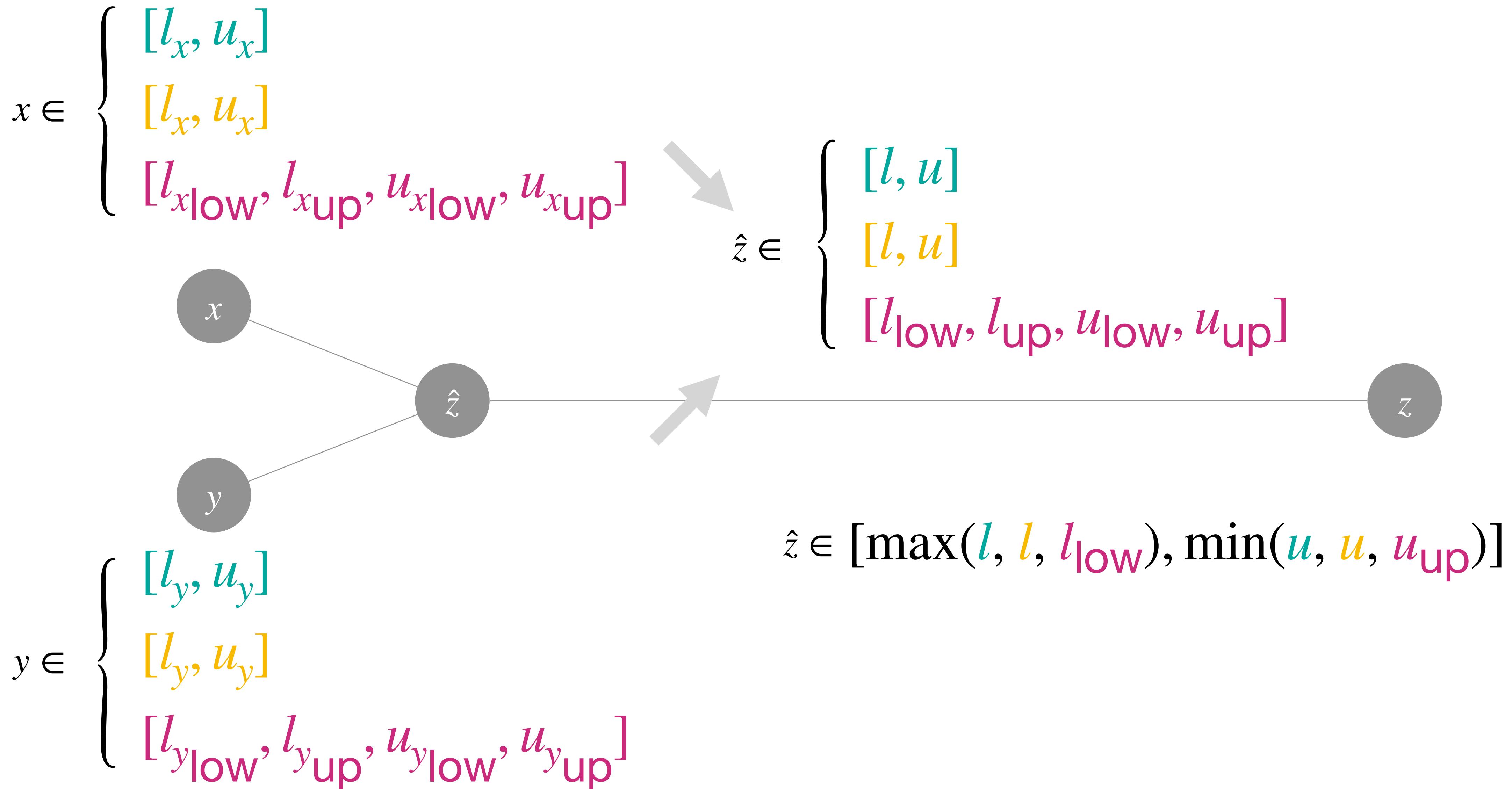
Reduced Product

Mazzucato, Urban @ SAS 2021



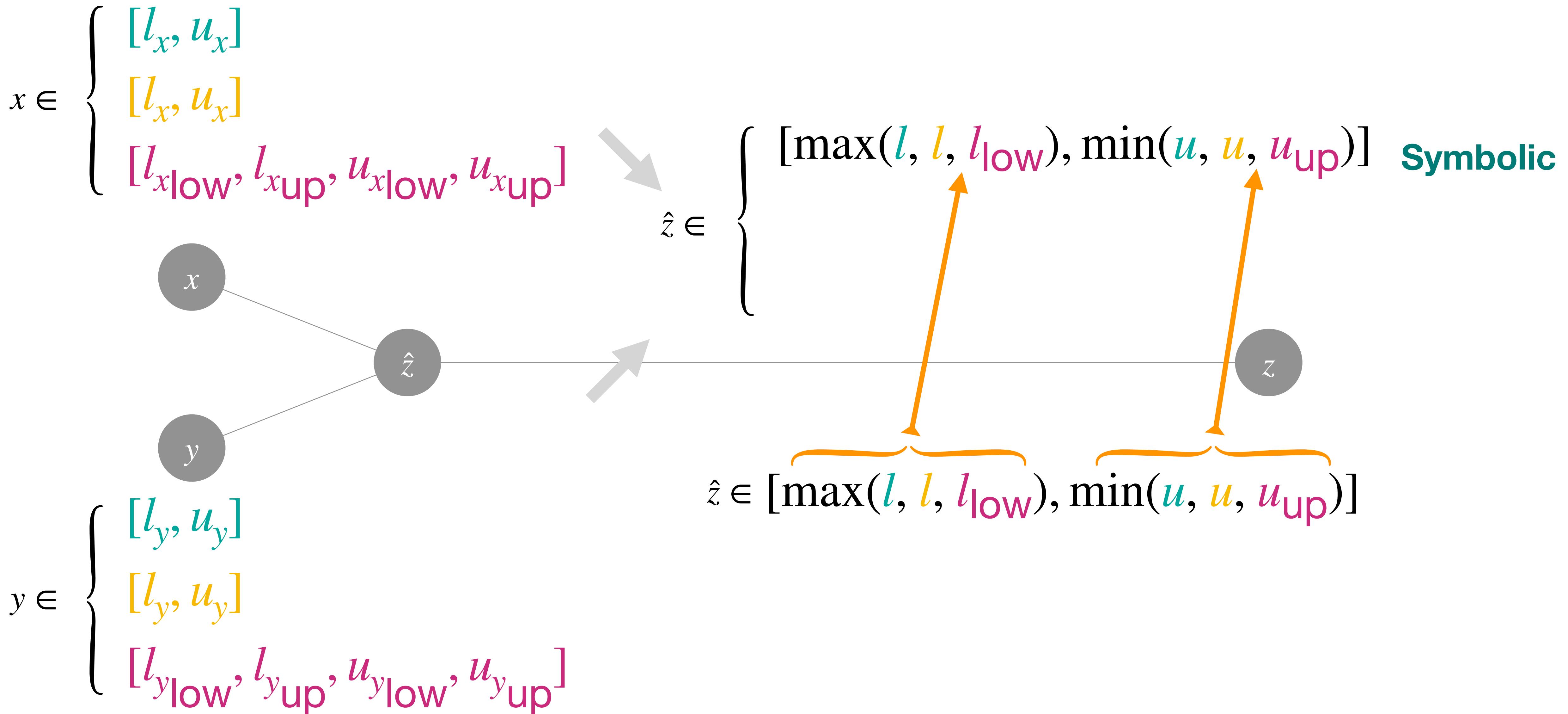
Reduced Product

Mazzucato, Urban @ SAS 2021



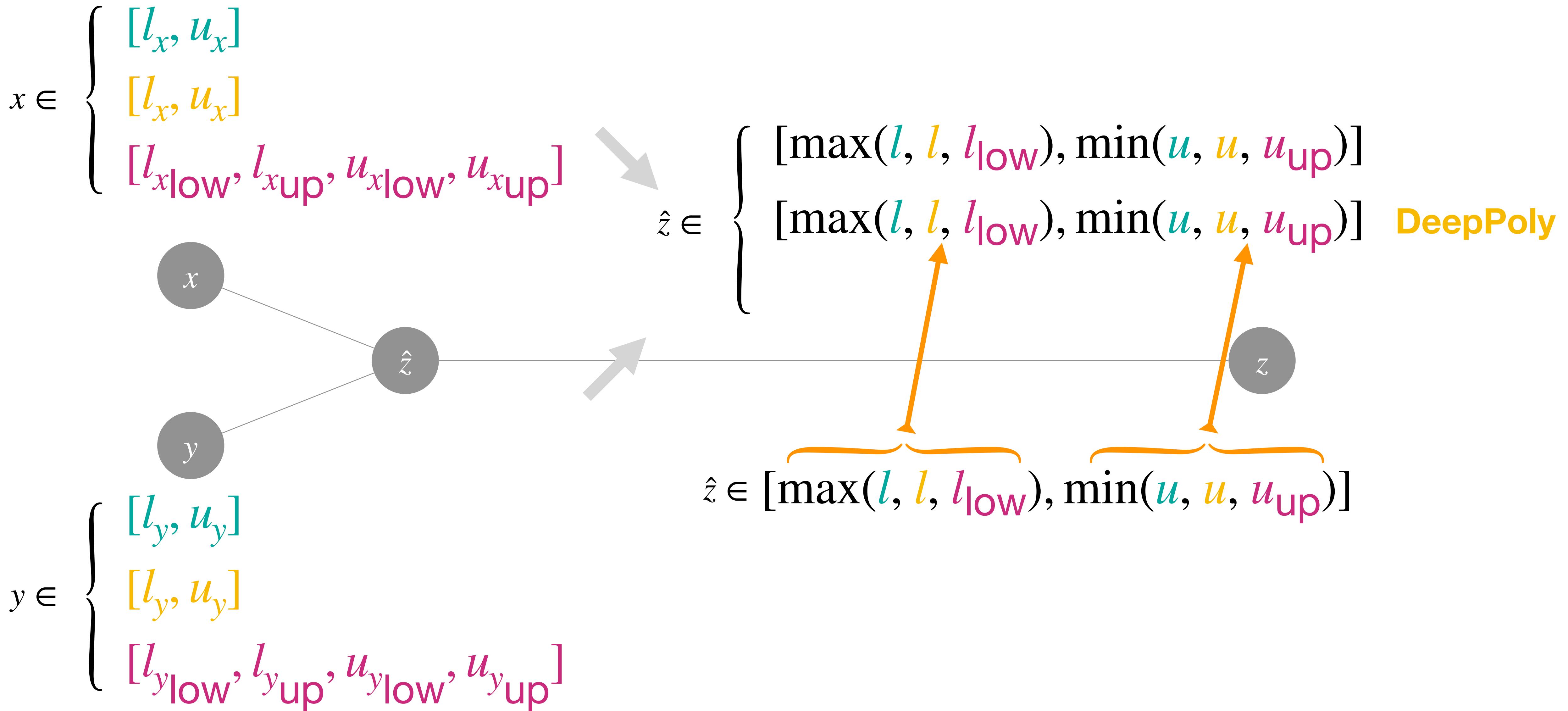
Reduced Product

Mazzucato, Urban @ SAS 2021



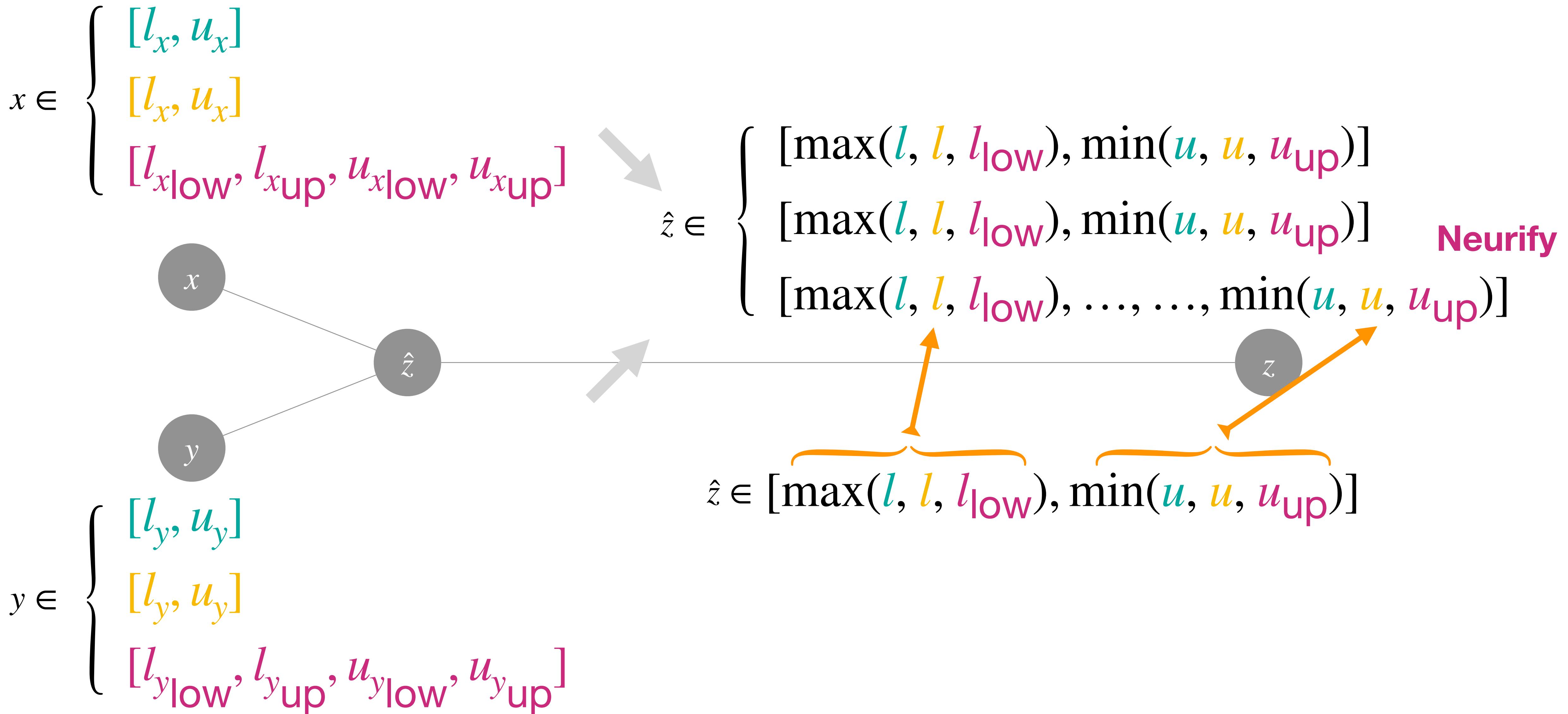
Reduced Product

Mazzucato, Urban @ SAS 2021



Reduced Product

Mazzucato, Urban @ SAS 2021



Reduced Product

Mazzucato, Urban @ SAS 2021

$$\hat{z} \in [l_{\text{low}}, l_{\text{up}}, u_{\text{low}}, u_{\text{up}}]$$

$$\hat{z} \in \left\{ \begin{array}{l} [\max(l, l, l_{\text{low}}), \min(u, u, u_{\text{up}})] \\ [\max(l, l, l_{\text{low}}), \min(u, u, u_{\text{up}})] \\ [\max(l, l, l_{\text{low}}), \dots, \dots, \min(u, u, u_{\text{up}})] \end{array} \right.$$

Neurify

$$\max(\max(l, l, l_{\text{low}}), l_{\text{up}})$$
$$\hat{z} \in [\max(l, l, l_{\text{low}}), \min(u, u, u_{\text{up}})]$$

Reduced Product

Mazzucato, Urban @ SAS 2021

$$\hat{z} \in [l_{\text{low}}, l_{\text{up}}, u_{\text{low}}, u_{\text{up}}]$$

$$\hat{z} \in \left\{ \begin{array}{l} [\max(l, \underline{l}, l_{\text{low}}), \min(u, \underline{u}, u_{\text{up}})] \\ [\max(l, \underline{l}, l_{\text{low}}), \min(u, \underline{u}, u_{\text{up}})] \\ [\max(l, \underline{l}, l_{\text{low}}), \dots, \dots, \min(u, \underline{u}, u_{\text{up}})] \end{array} \right.$$

Neurify

$$\min(\min(u, \underline{u}, u_{\text{up}}), \underline{u}_{\text{low}})$$
$$\hat{z} \in [\max(l, \underline{l}, l_{\text{low}}), \min(u, \underline{u}, u_{\text{up}})]$$

Reduced Product

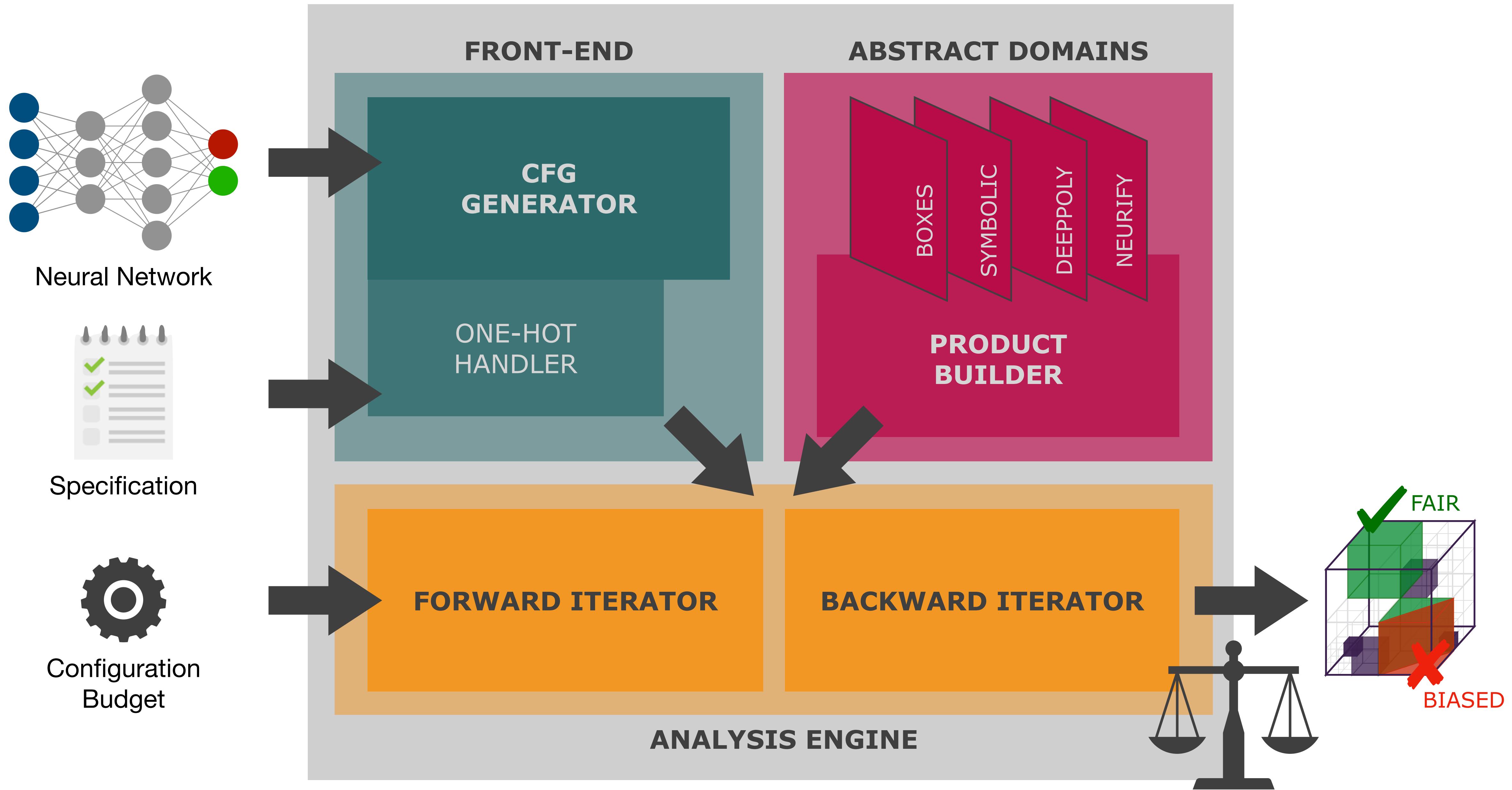
Mazzucato, Urban @ SAS 2021

$$\hat{z} \in \left\{ \begin{array}{ll} [\max(\textcolor{teal}{l}, \textcolor{orange}{l}, l_{\text{low}}), \min(\textcolor{teal}{u}, \textcolor{orange}{u}, u_{\text{up}})] & \textbf{Symbolic} \\ [\max(\textcolor{teal}{l}, \textcolor{orange}{l}, l_{\text{low}}), \min(\textcolor{teal}{u}, \textcolor{orange}{u}, u_{\text{up}})] & \textbf{DeepPoly} \\ [\max(\textcolor{teal}{l}, \textcolor{orange}{l}, l_{\text{low}}), \max(\max(\textcolor{teal}{l}, \textcolor{orange}{l}, l_{\text{low}}), l_{\text{up}}), \\ \min(\min(\textcolor{teal}{u}, \textcolor{orange}{u}, u_{\text{up}}), u_{\text{low}}), \min(\textcolor{teal}{u}, \textcolor{orange}{u}, u_{\text{up}})] & \textbf{Neurify} \end{array} \right.$$

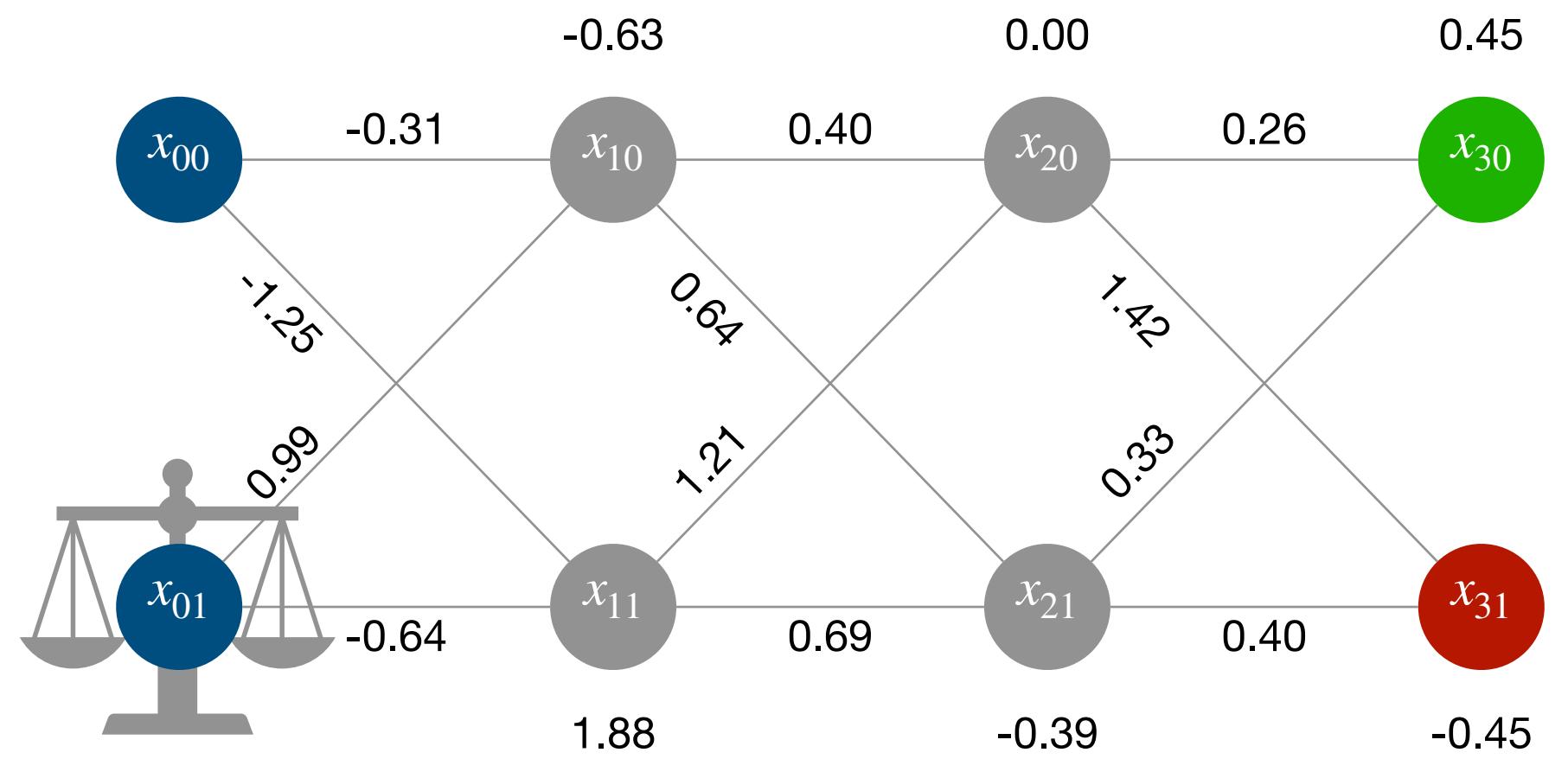
Precision-vs-Scalability

L	U	Symbolic	DeepPoly	Neurify	Product	
0.5	3	48.78%	49.01%	46.49%	59.20%	+10.3%
	5	56.11%	56.15%	53.06%	68.23%	+11.9%
0.25	3	83.63%	81.82%	81.40%	87.04%	+3.4%
	5	91.67%	91.58%	92.33%	95.48%	+3.2%

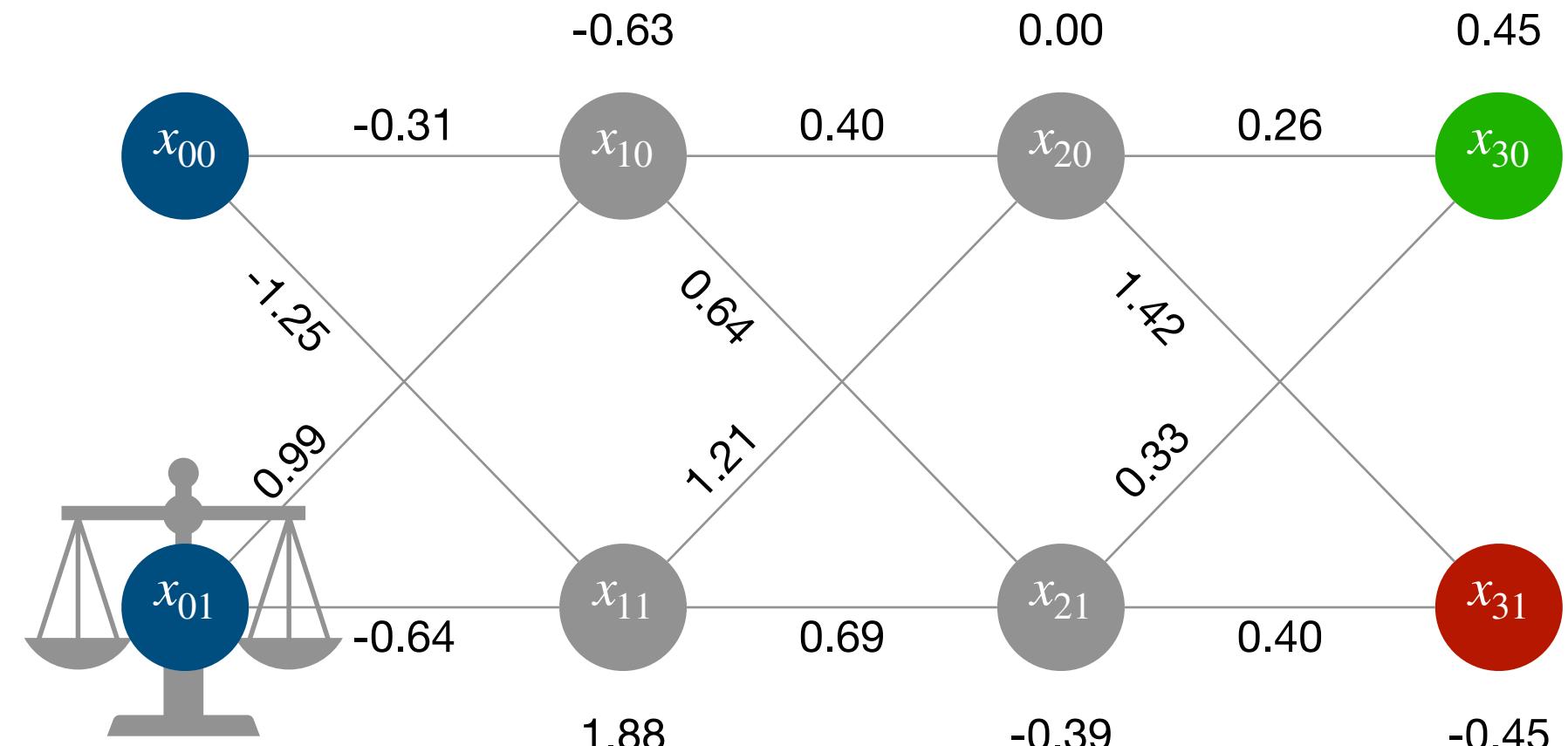




Ongoing work

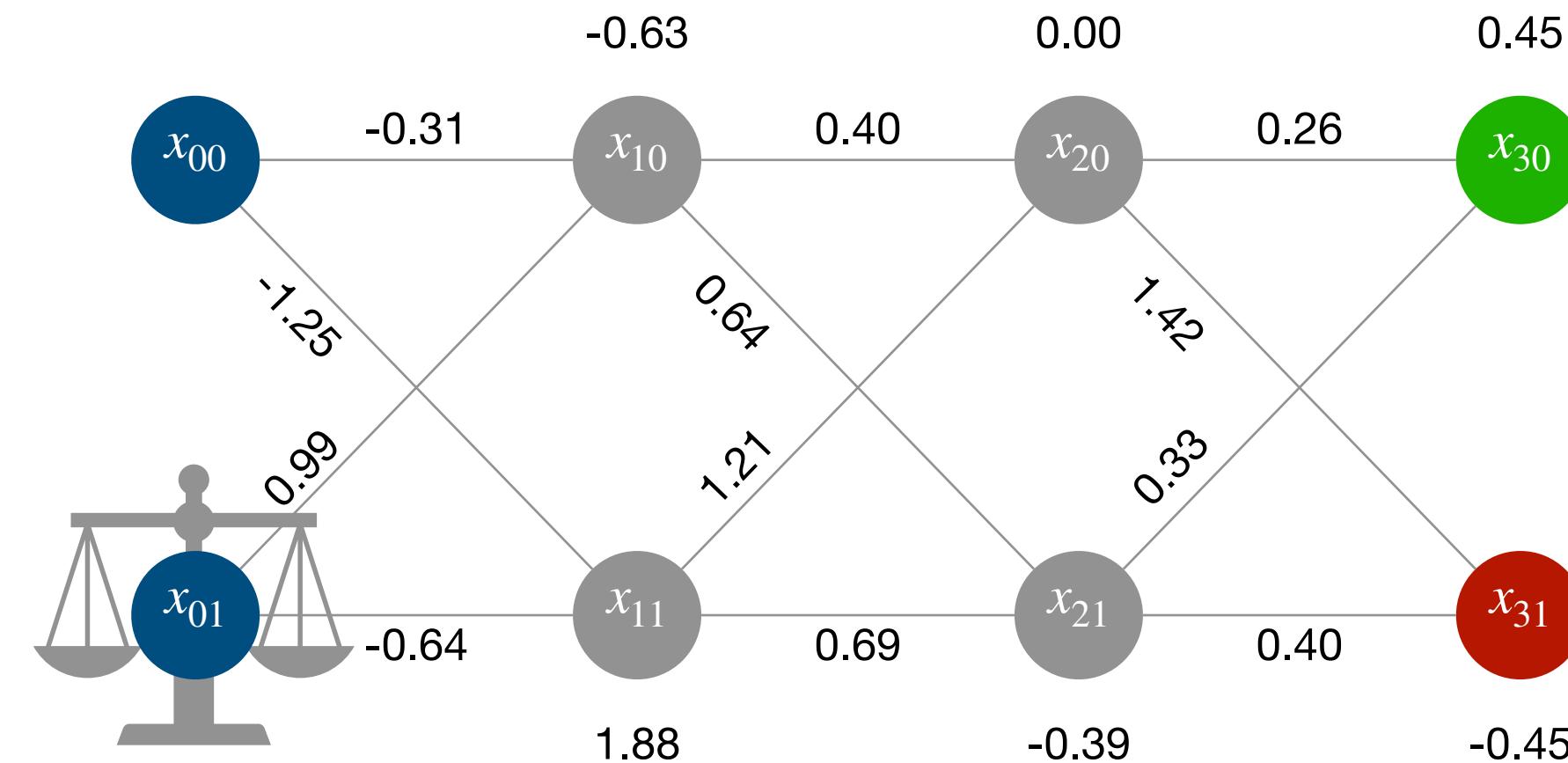


Ongoing work



```
1 | x00 = input()
2 | x01 = input()
3 |
4 | x10 = ReLU(-0.31*x00 +0.99*x01 -0.63)
5 | x11 = ReLU(-1.25*x00 -0.64*x01 +1.88)
6 |
7 | x20 = ReLU(0.40*x10 +1.21*x11)
8 | x21 = ReLU(0.64*x10 +0.69*x11 -0.39)
9 |
10 | x30 = 0.26*x20 +0.33*x21 +0.45
11 | x31 = 1.42*x20 +0.40*x21 -0.45
```

Ongoing work

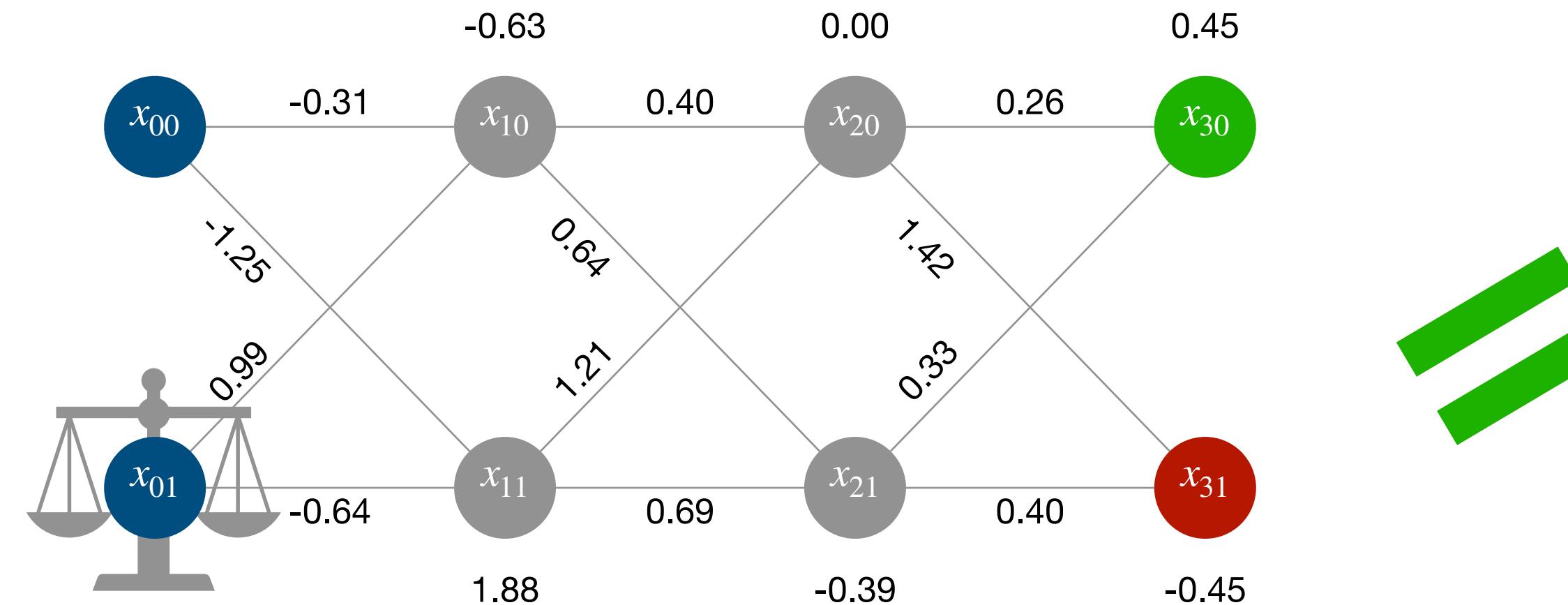


Input Data Usage Framework [Urban18]

```
1 | x00 = input()
2 | x01 = input() Unused
3 |
4 | x10 = ReLU(-0.31*x00 +0.99*x01 -0.63)
5 | x11 = ReLU(-1.25*x00 -0.64*x01 +1.88)
6 |
7 | x20 = ReLU(0.40*x10 +1.21*x11)
8 | x21 = ReLU(0.64*x10 +0.69*x11 -0.39)
9 |
10| x30 = 0.26*x20 +0.33*x21 +0.45
11| x31 = 1.42*x20 +0.40*x21 -0.45
```

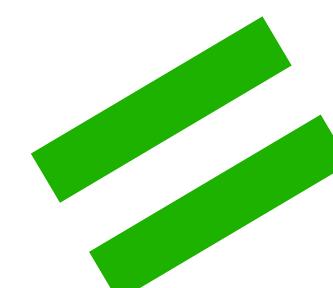


Ongoing work



Fair

Input Data Usage Framework [Urban18]



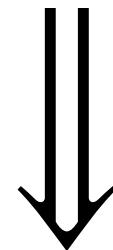
```
1 | x00 = input()
2 | x01 = input() Unused
3 |
4 | x10 = ReLU(-0.31*x00 + 0.99*x01 - 0.63)
5 | x11 = ReLU(-1.25*x00 - 0.64*x01 + 1.88)
6 |
7 | x20 = ReLU(0.40*x10 + 1.21*x11)
8 | x21 = ReLU(0.64*x10 + 0.69*x11 - 0.39)
9 |
10 | x30 = 0.26*x20 + 0.33*x21 + 0.45
11 | x31 = 1.42*x20 + 0.40*x21 - 0.45
12 |
13 | output(x30 > x31)
```

Ongoing work

Input Data Usage Framework [Urban18]

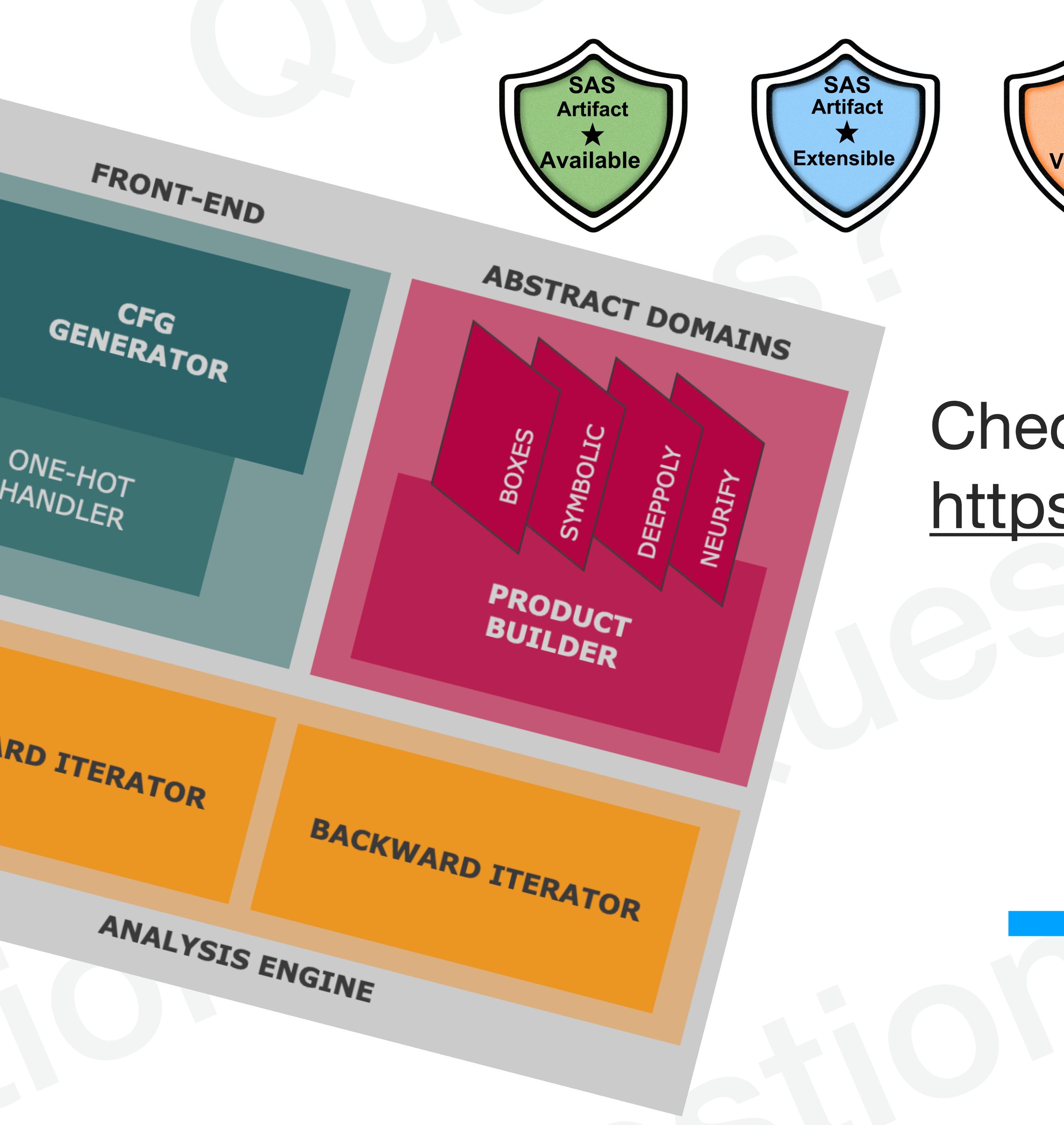
Quantitative Input Data Usage

Unused \implies no impact on the outcome

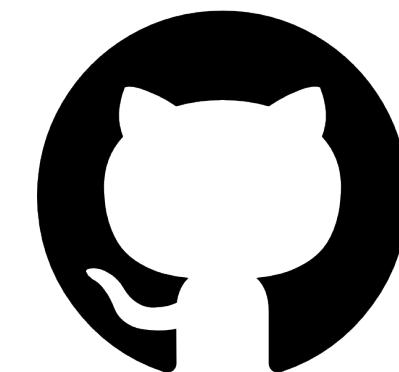


Used \implies how much?

```
1| x00 = input()
2| x01 = input()
3|
4| x10 = ReLU(-0.31*x00 +0.99*x01 -0.63)
5| x11 = ReLU(-1.25*x00 -0.64*x01 +1.88)
6|
7| x20 = ReLU(0.40*x10 +1.21*x11)
8| x21 = ReLU(0.64*x10 +0.69*x11 -0.39)
9|
10| x30 = 0.26*x20 +0.33*x21 +0.45
11| x31 = 1.42*x20 +0.40*x21 -0.45
12|
13| output(x30 > x31)
```

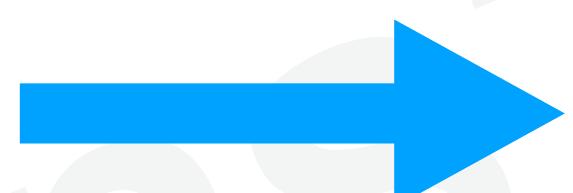


Ready-to-go Docker image at
<https://doi.org/10.5281/zenodo.4737450>

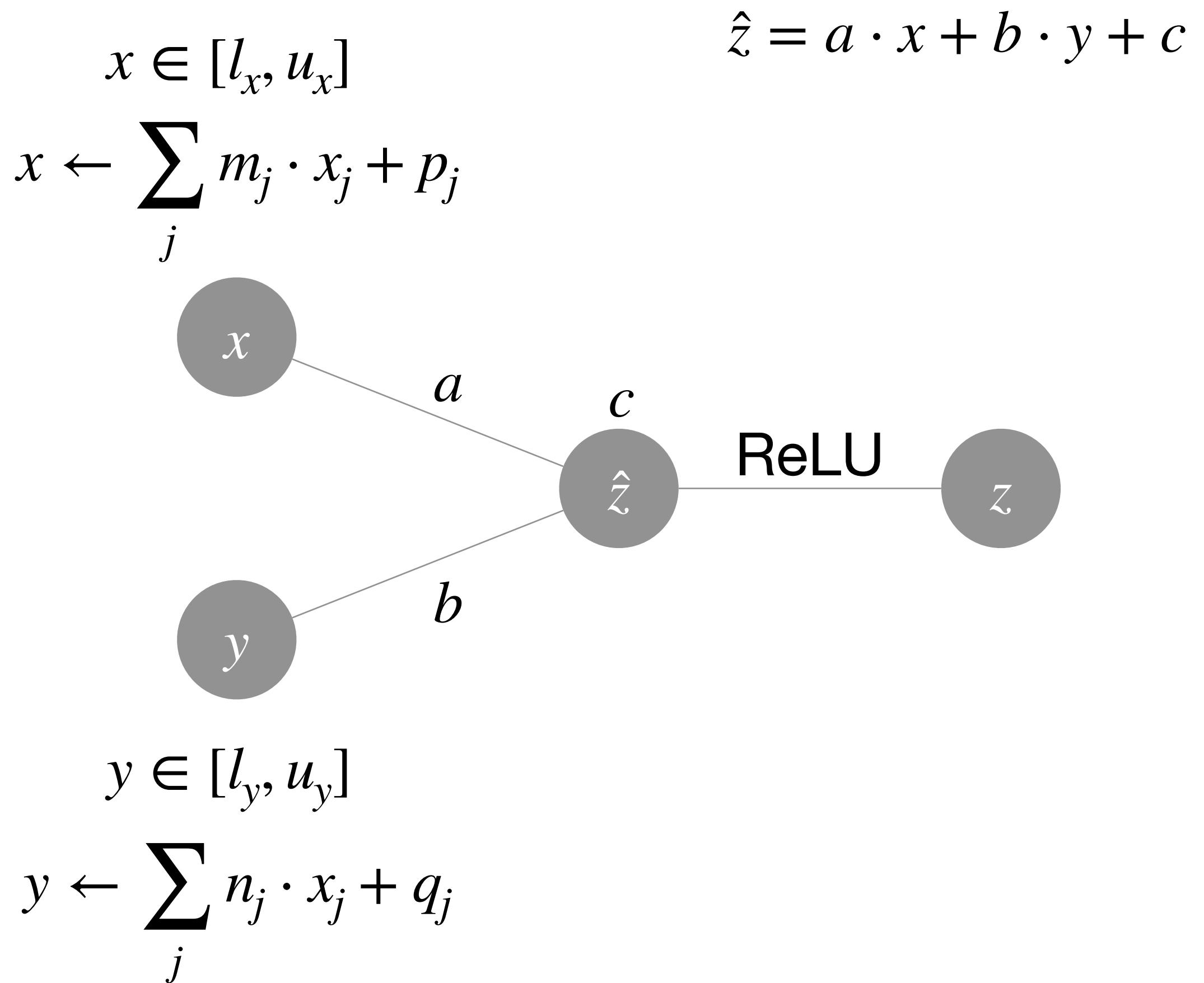


Check it out on GitHub!
<https://github.com/caterinaurban/libra>

Ongoing work
Quantitative Input Data Usage

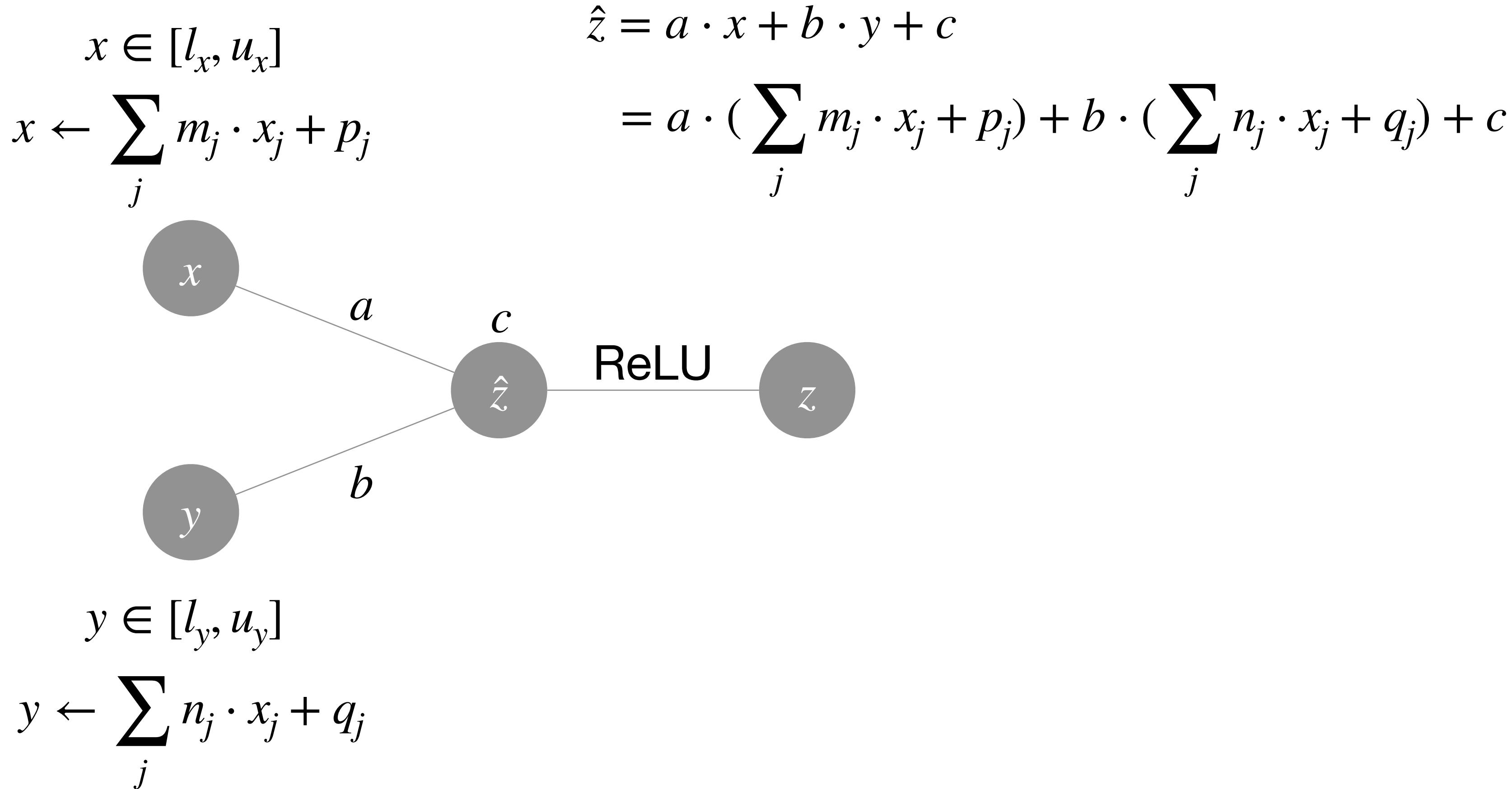


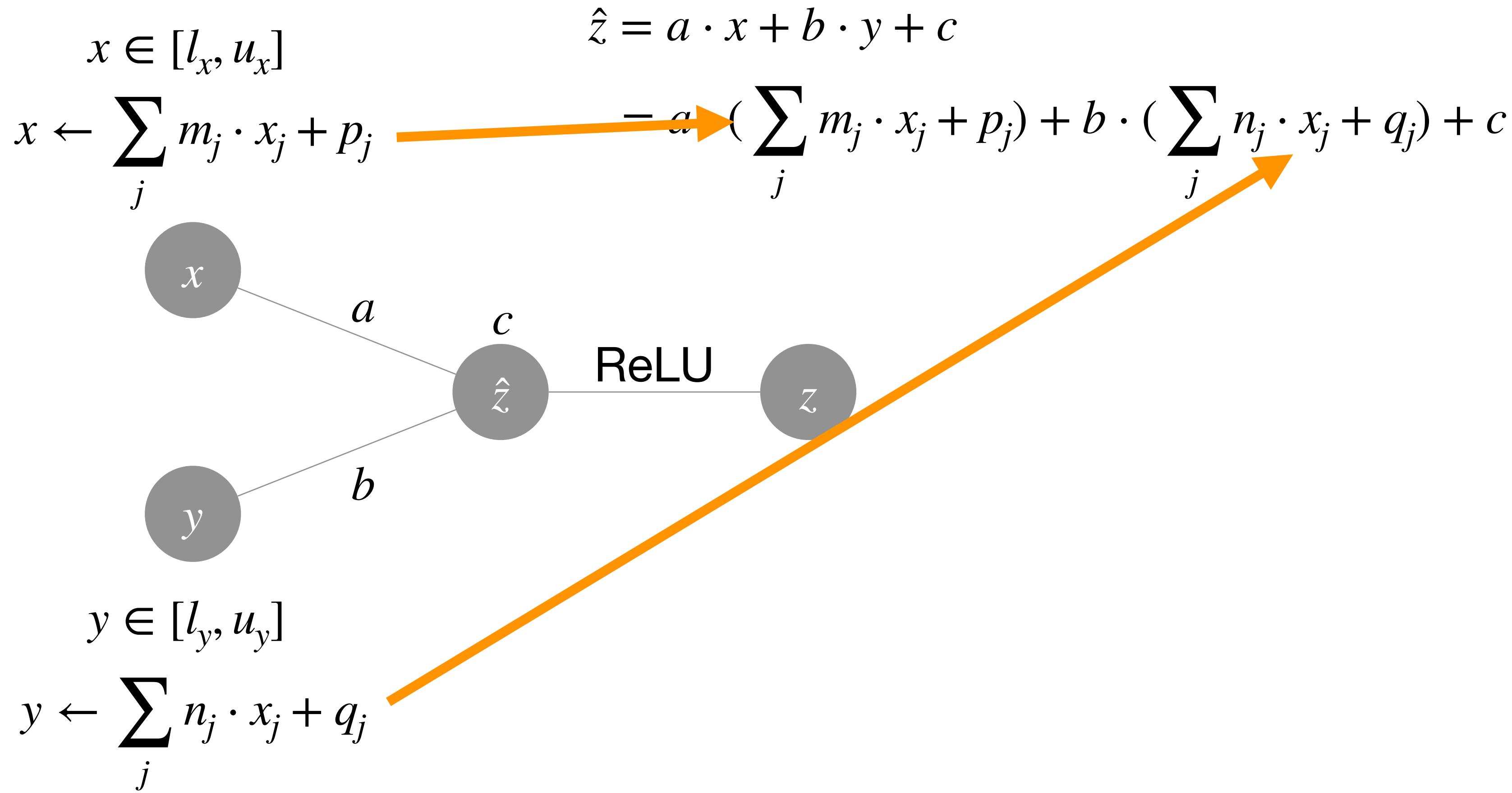
Backup



Symbolic

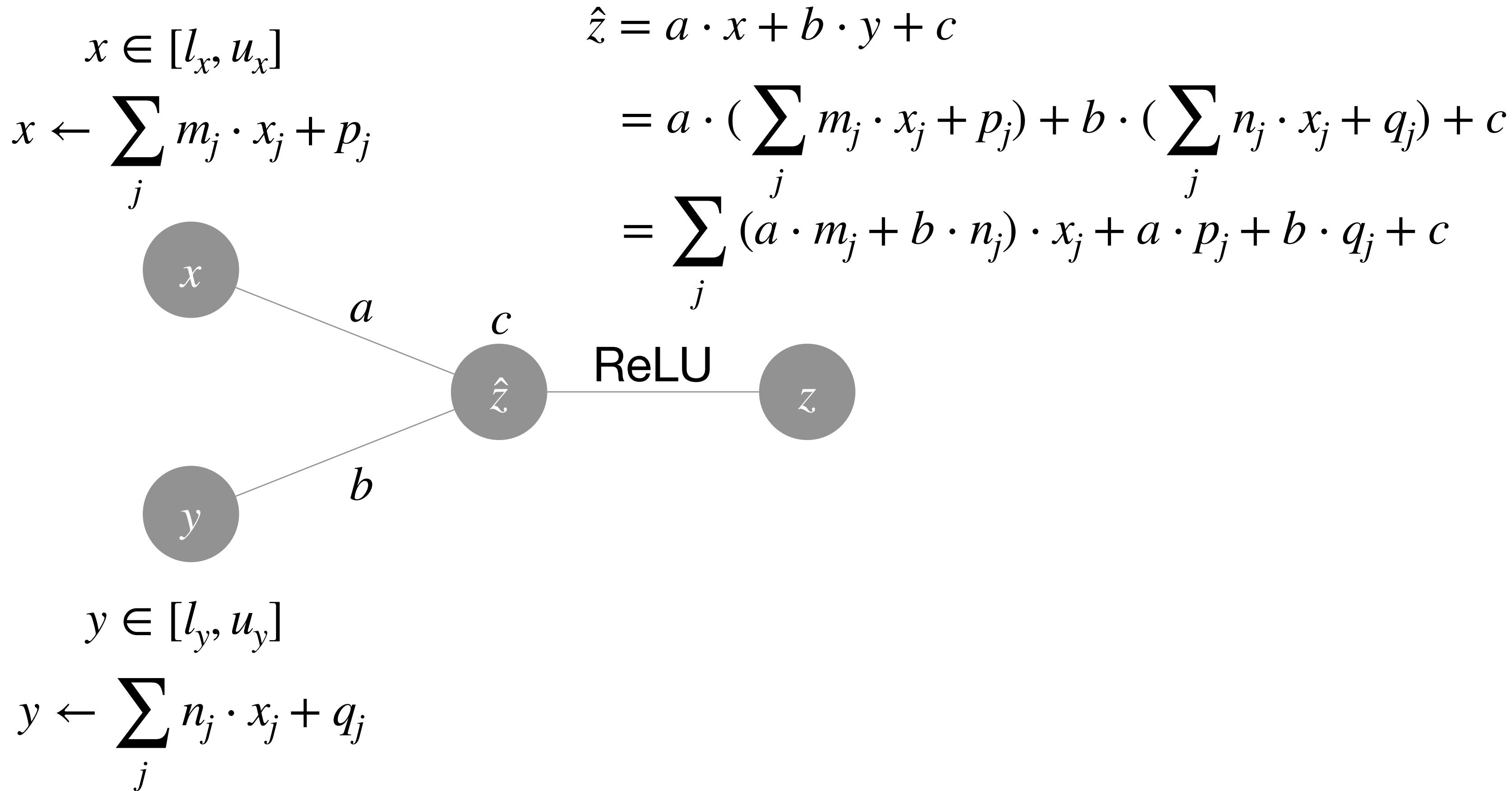
Li et al. @ SAS 2019





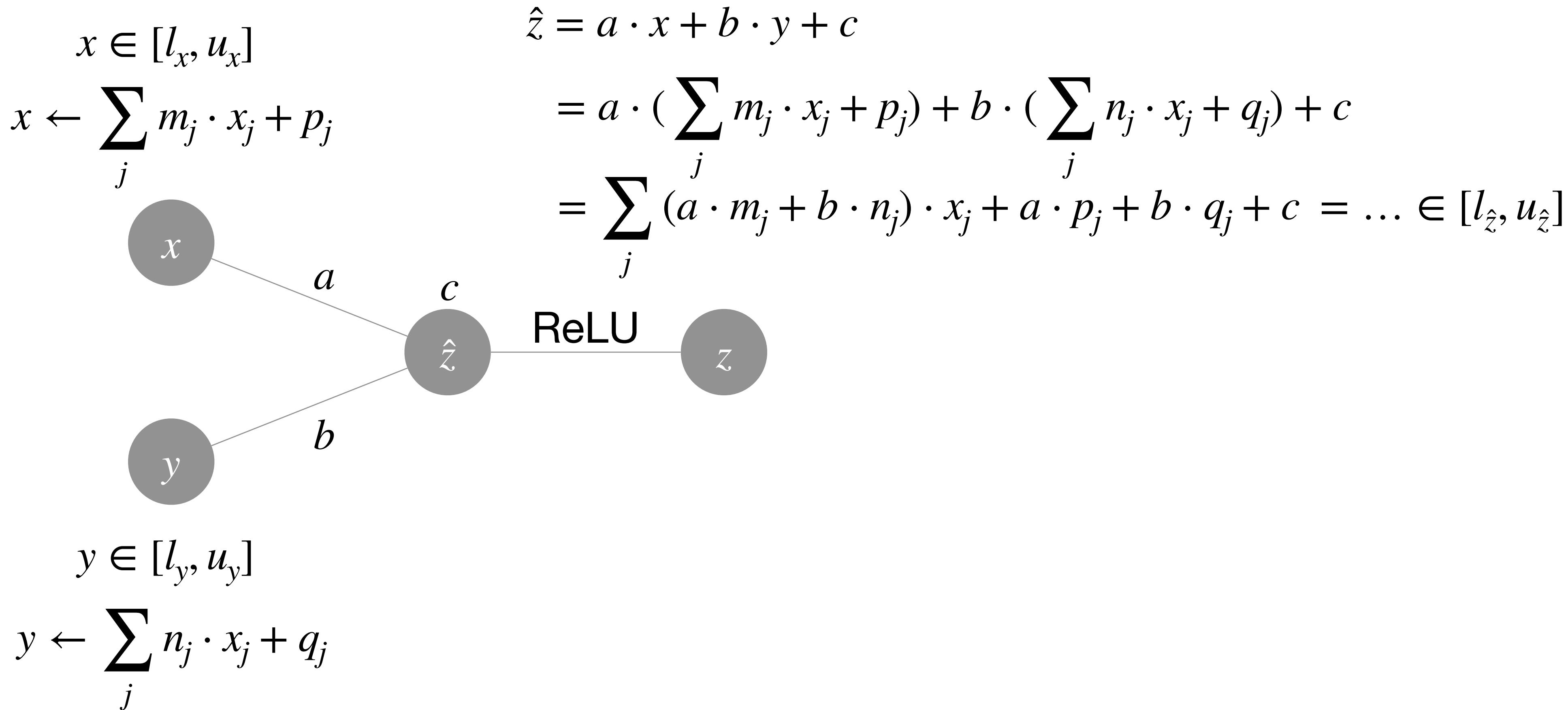
Symbolic

Li et al. @ SAS 2019



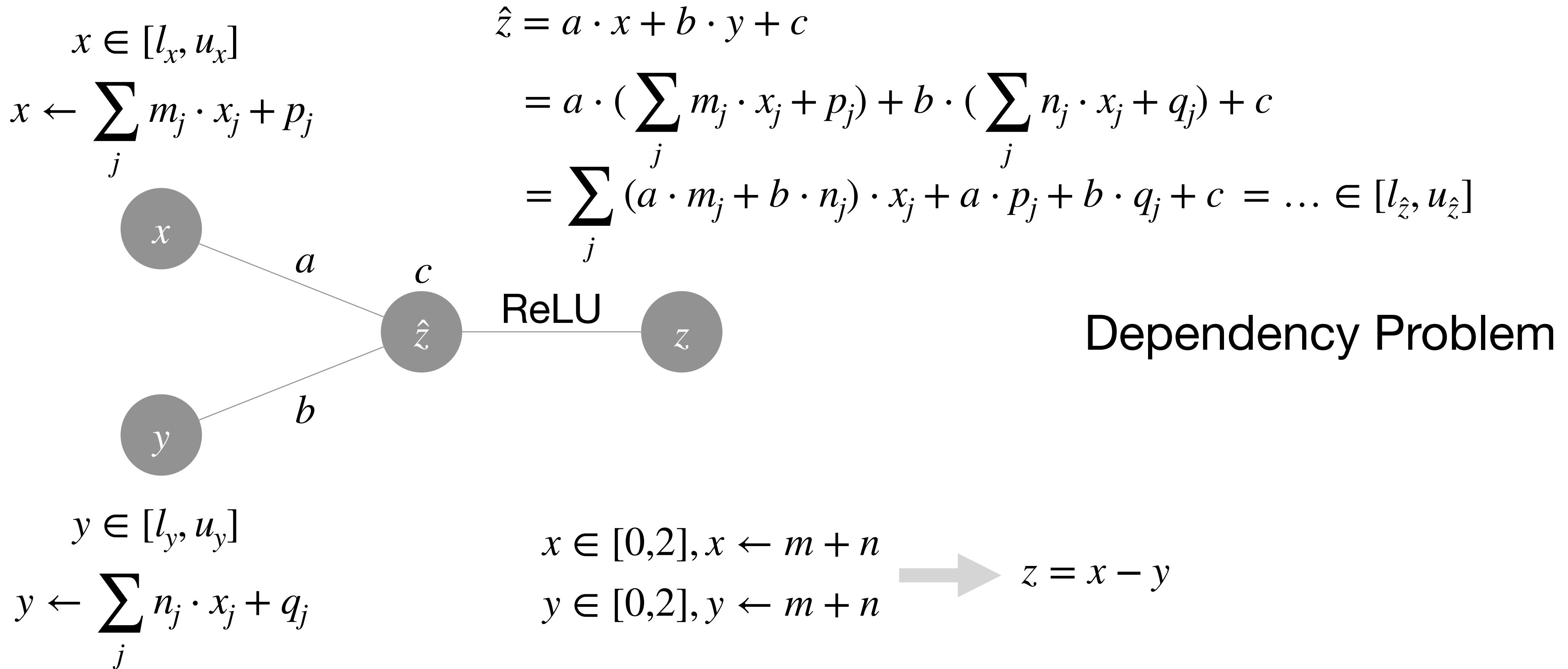
Symbolic

Li et al. @ SAS 2019



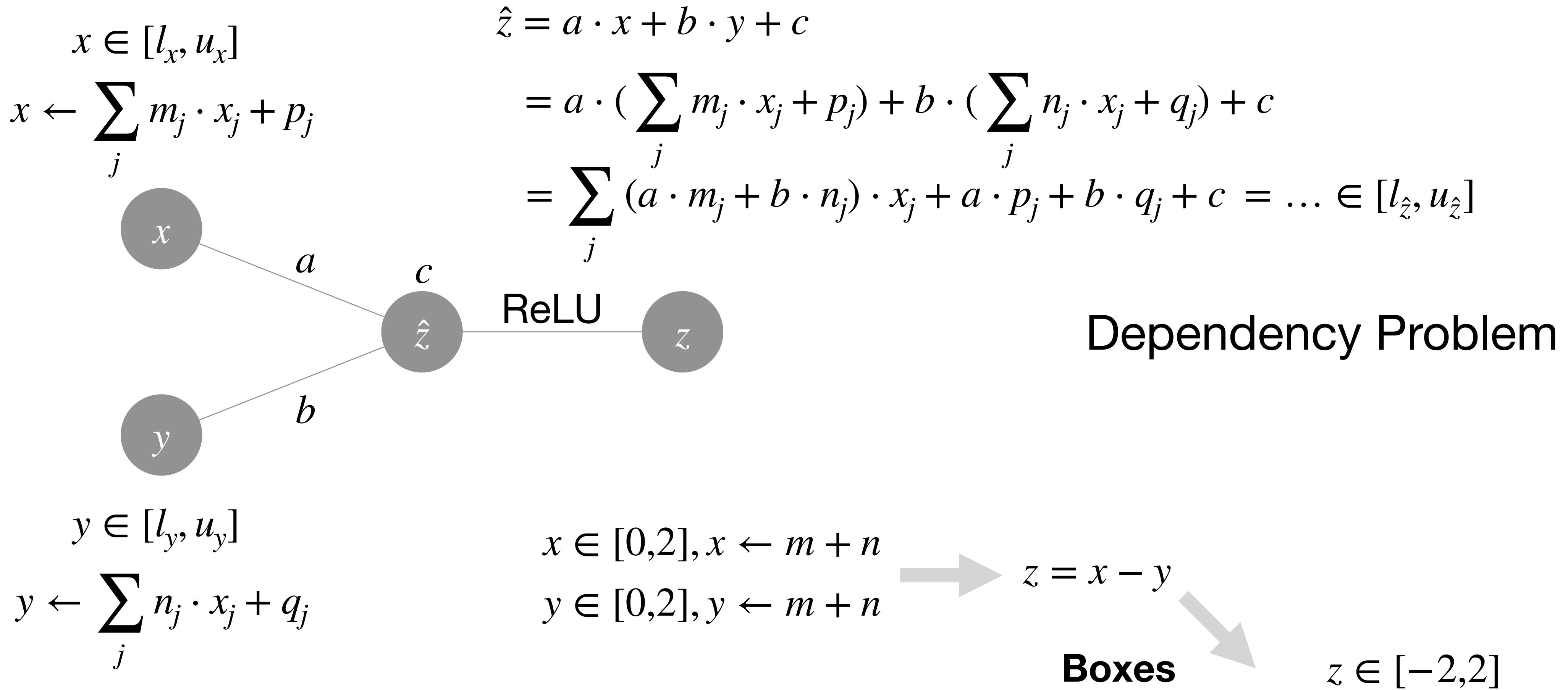
Symbolic

Li et al. @ SAS 2019



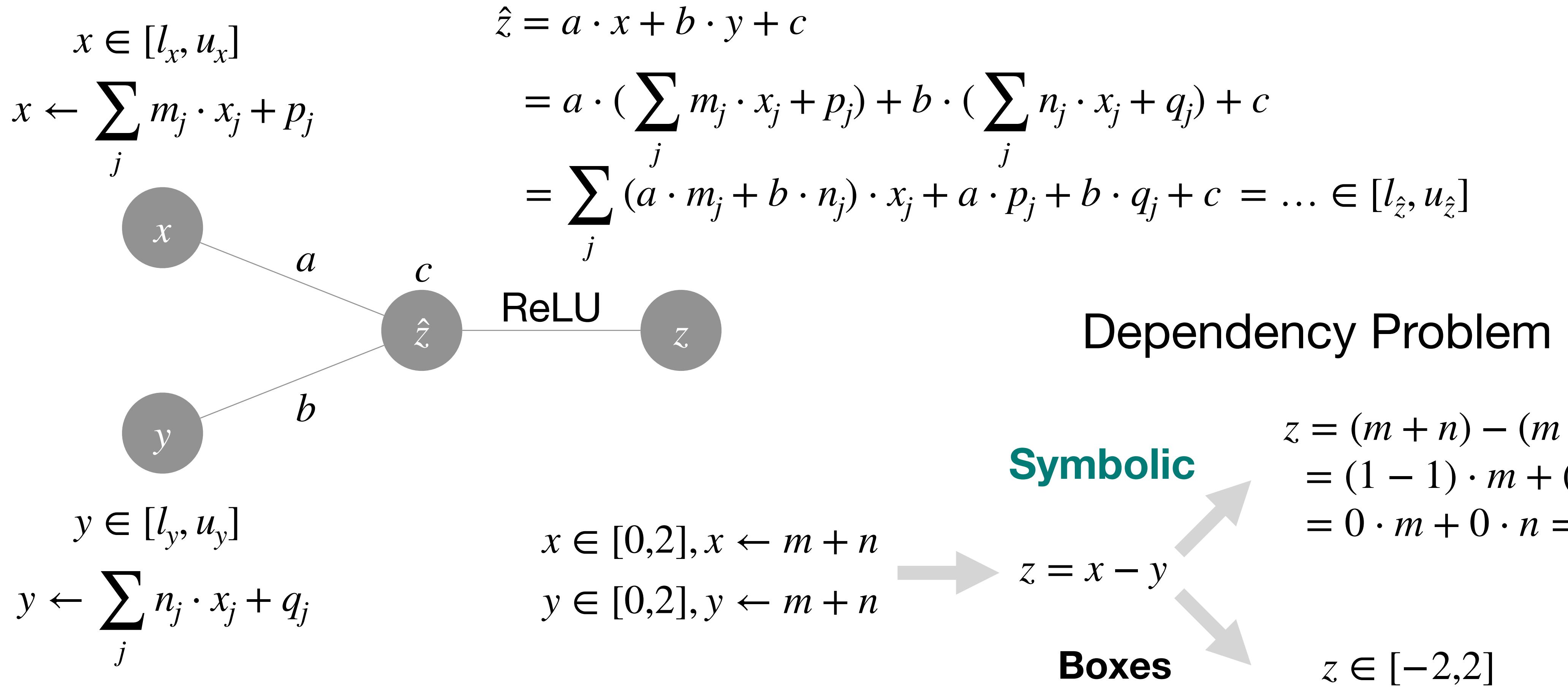
Symbolic

Li et al. @ SAS 2019



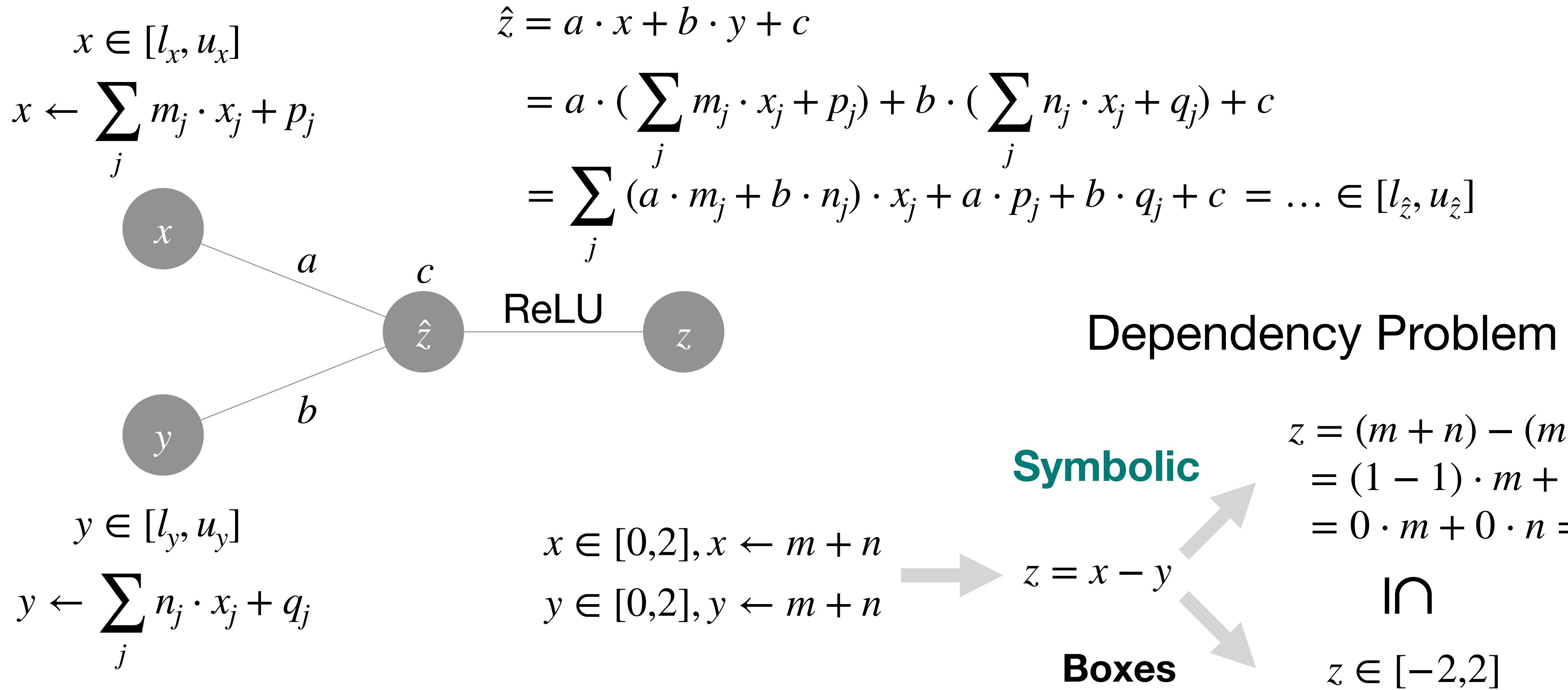
Symbolic

Li et al. @ SAS 2019



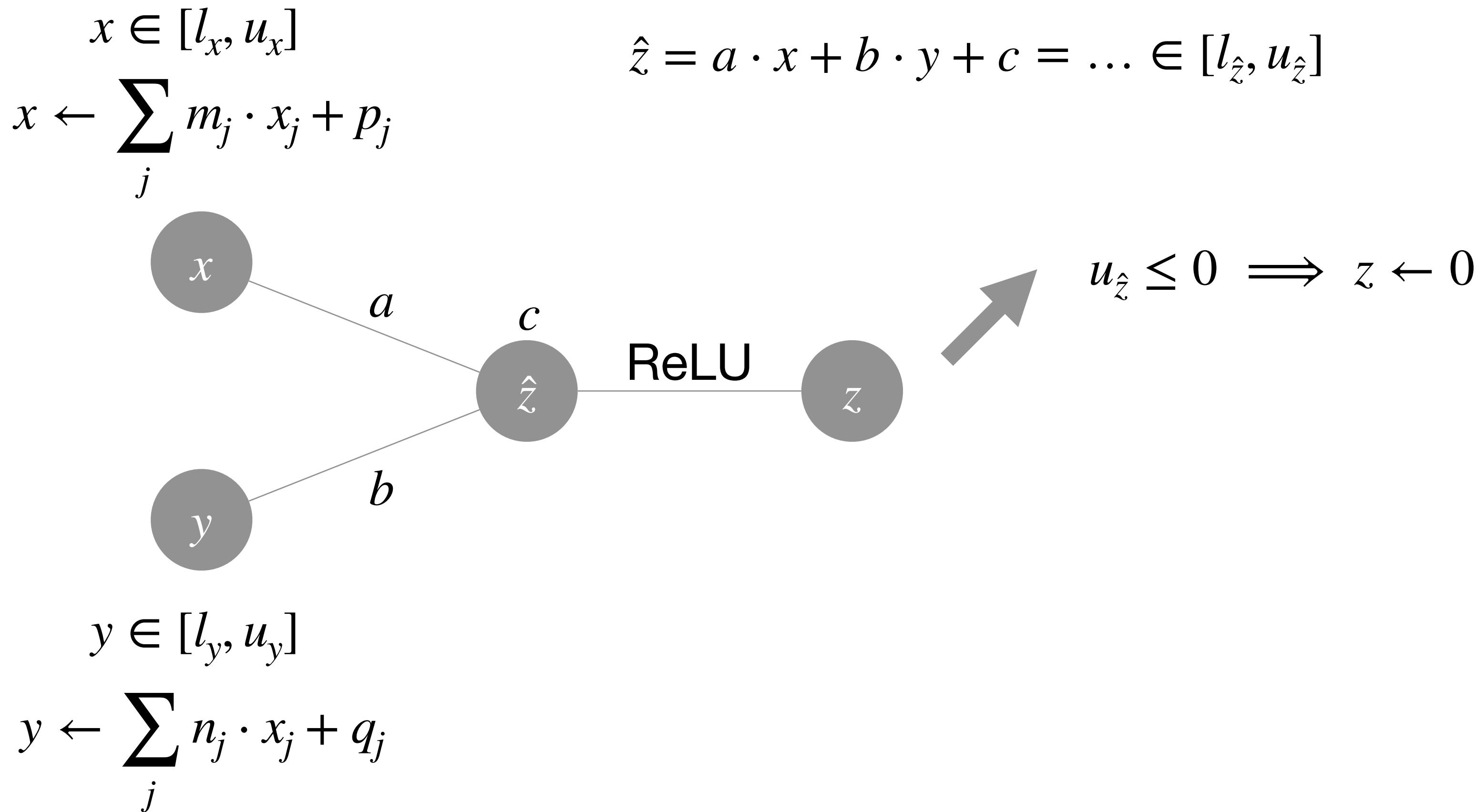
Symbolic

Li et al. @ SAS 2019



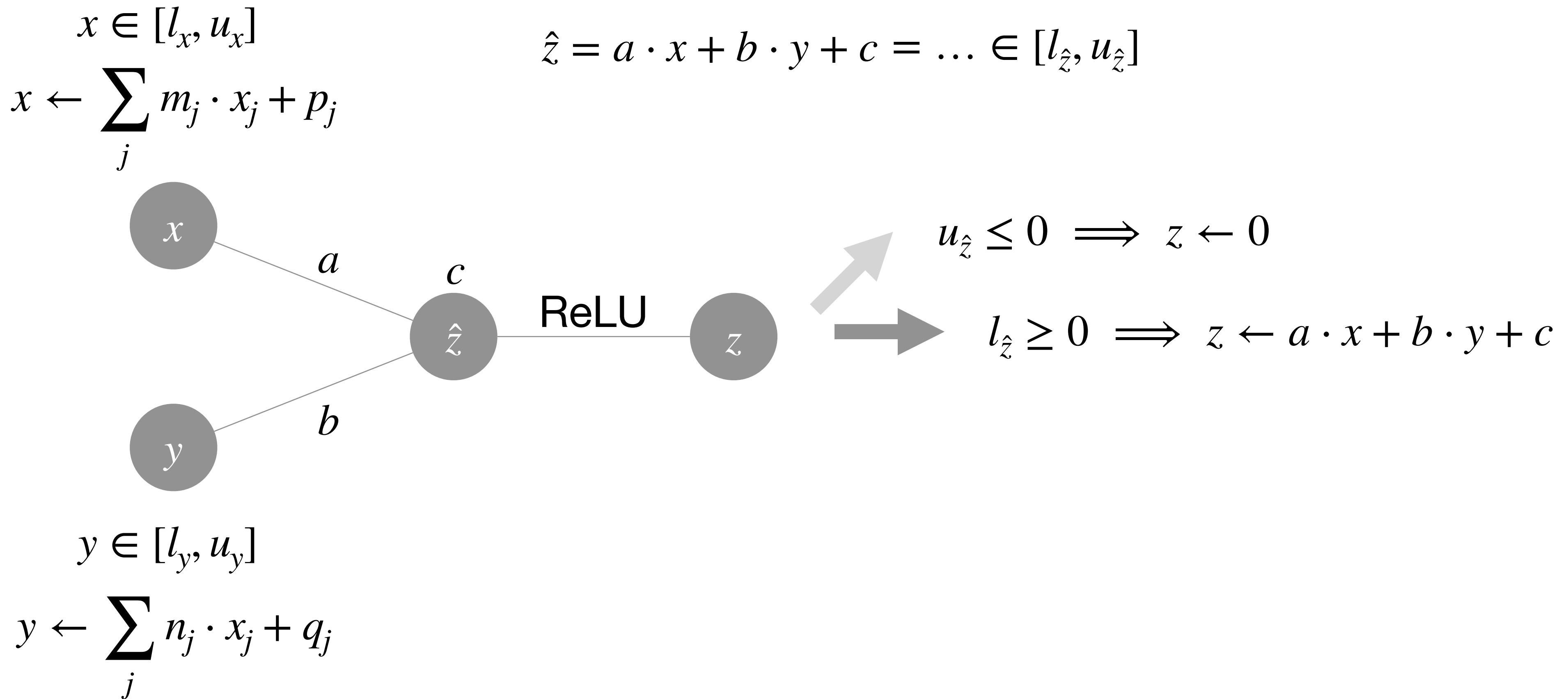
Symbolic

Li et al. @ SAS 2019



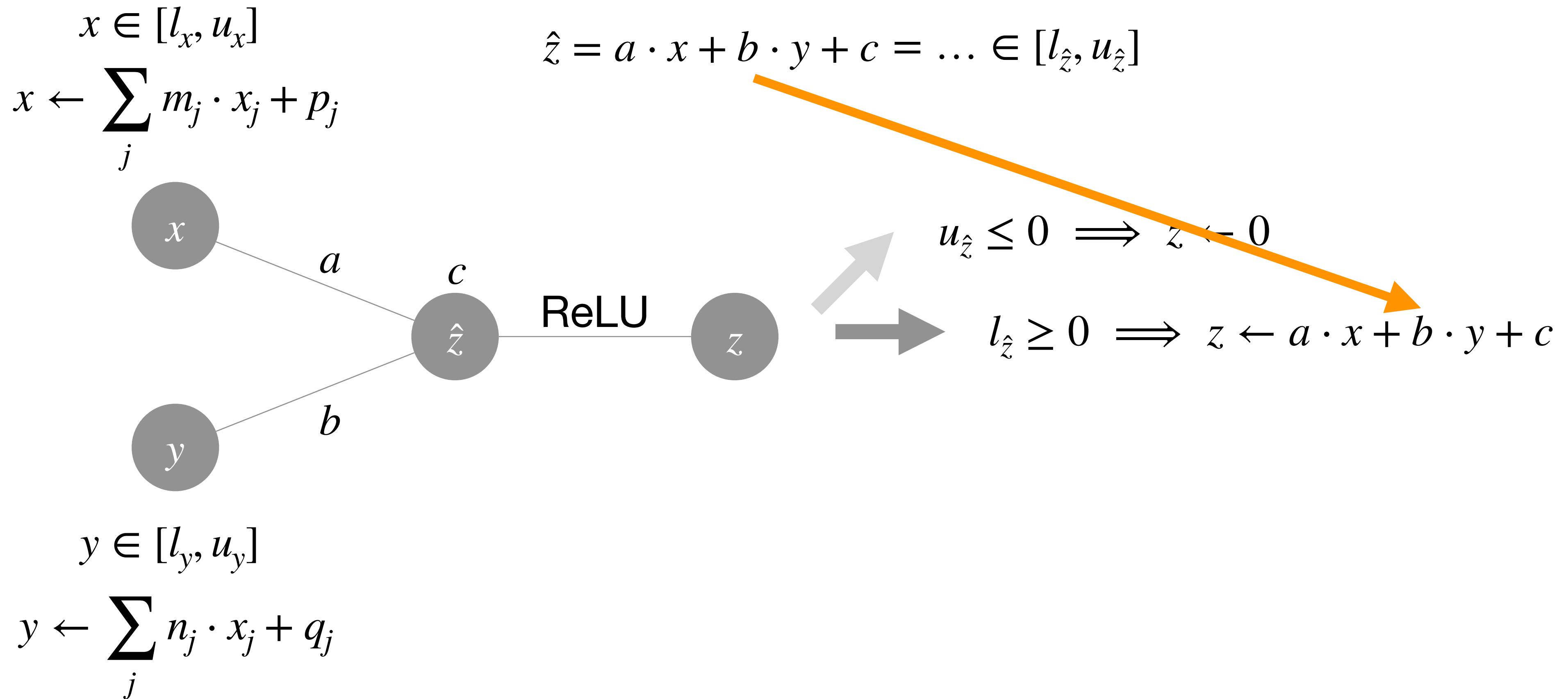
Symbolic

Li et al. @ SAS 2019



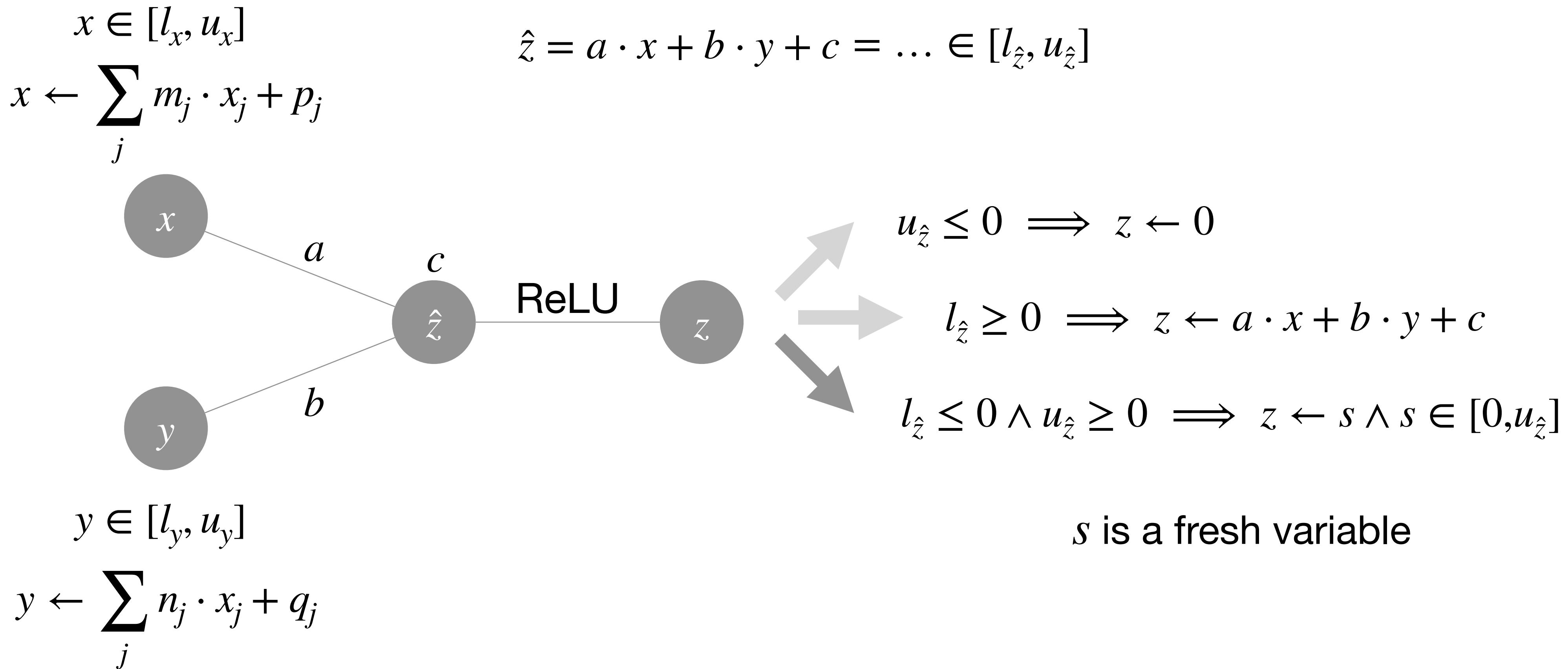
Symbolic

Li et al. @ SAS 2019



Symbolic

Li et al. @ SAS 2019



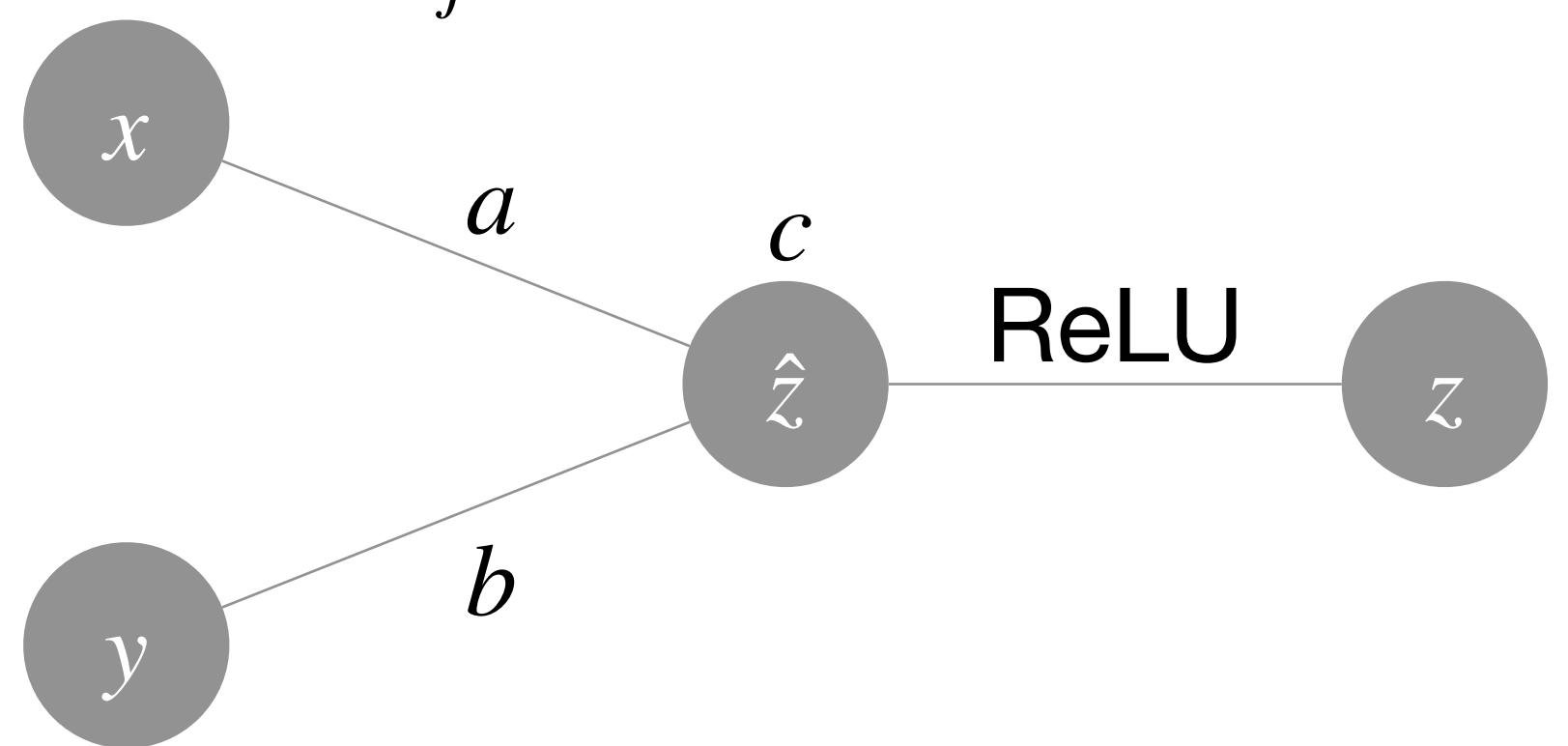
DeepPoly

Singh et al. @ POPL 2019

$$x \in [l_x, u_x]$$

$$\sum_j \underline{m}_j \cdot x_j + \underline{p}_j \leq x \leq \sum_j \bar{m}_j \cdot x_j + \bar{p}_j$$

$$\hat{z} = a \cdot x + b \cdot y + c$$



$$y \in [l_y, u_y]$$

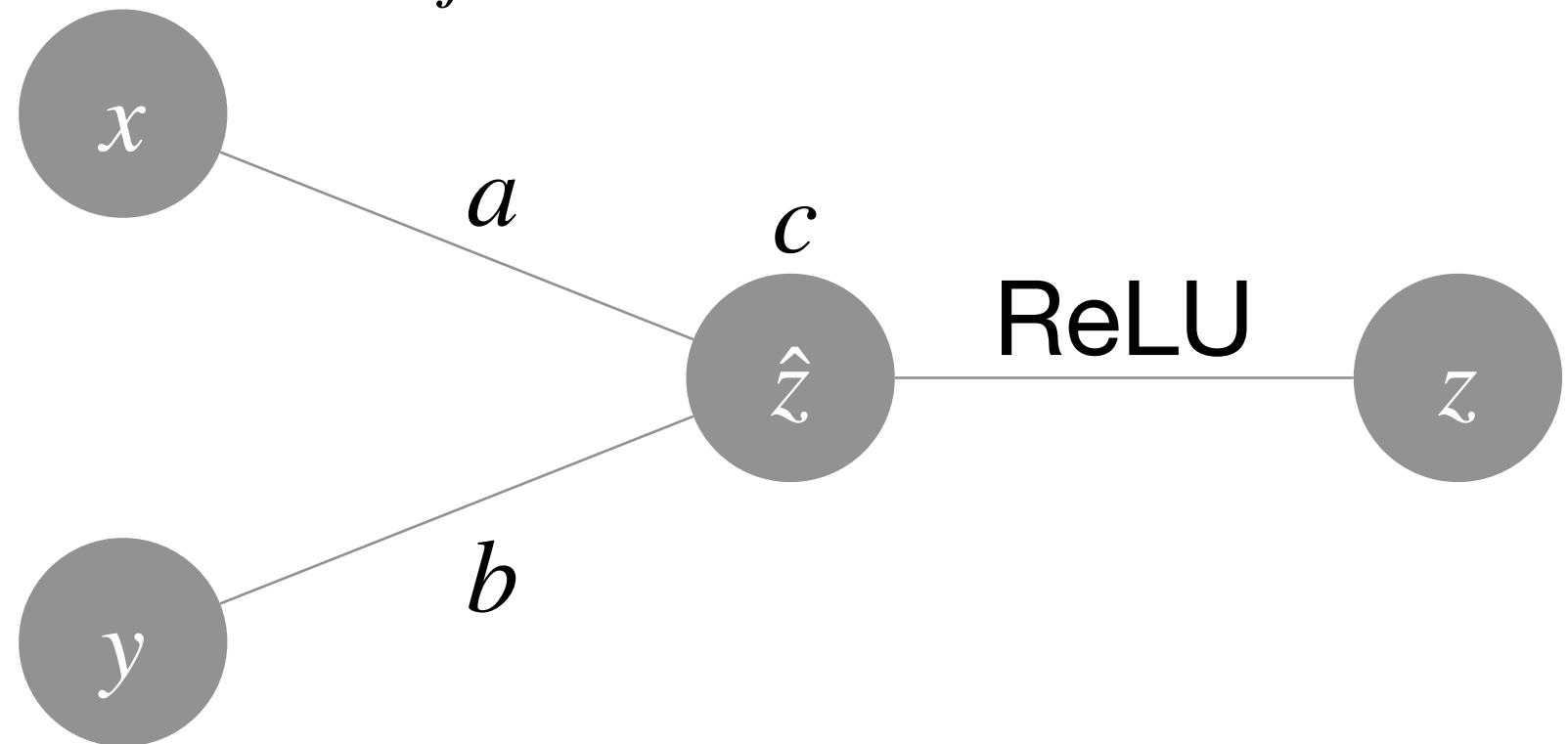
$$\sum_j \underline{n}_j \cdot y_j + \underline{q}_j \leq y \leq \sum_j \bar{n}_j \cdot y_j + \bar{q}_j$$

DeepPoly

Singh et al. @ POPL 2019

$$x \in [l_x, u_x]$$

$$\sum_j \underline{m}_j \cdot x_j + \underline{p}_j \leq x \leq \sum_j \bar{m}_j \cdot x_j + \bar{p}_j$$



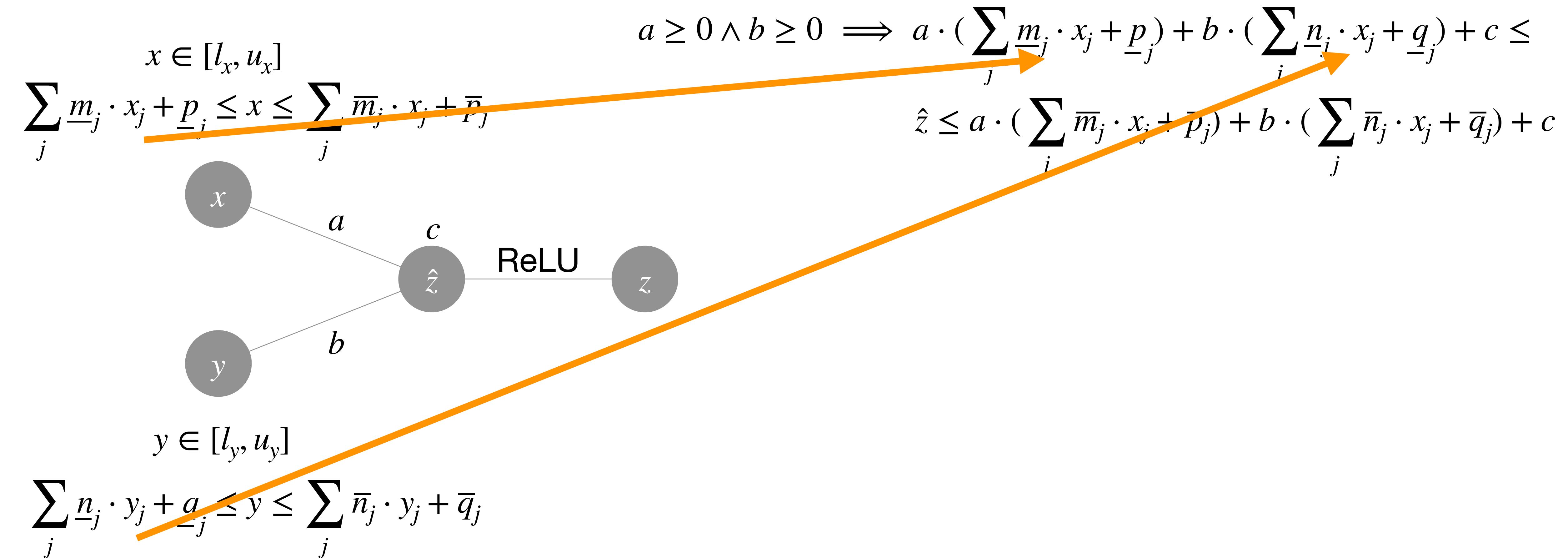
$$y \in [l_y, u_y]$$

$$\sum_j \underline{n}_j \cdot y_j + \underline{q}_j \leq y \leq \sum_j \bar{n}_j \cdot y_j + \bar{q}_j$$

$$a \geq 0 \wedge b \geq 0 \implies a \cdot \left(\sum_j \underline{m}_j \cdot x_j + \underline{p}_j \right) + b \cdot \left(\sum_j \bar{m}_j \cdot x_j + \bar{p}_j \right) + c \leq \hat{z} \leq a \cdot \left(\sum_j \bar{m}_j \cdot x_j + \bar{p}_j \right) + b \cdot \left(\sum_j \bar{n}_j \cdot x_j + \bar{q}_j \right) + c$$

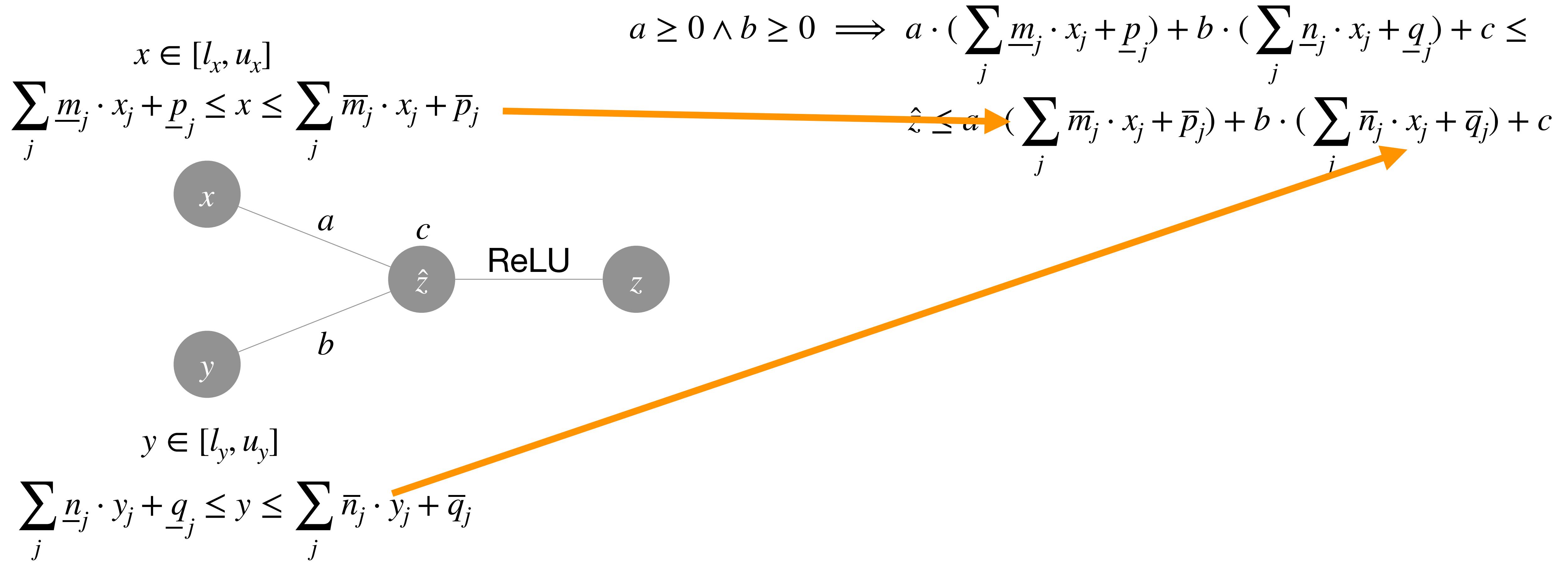
DeepPoly

Singh et al. @ POPL 2019



DeepPoly

Singh et al. @ POPL 2019

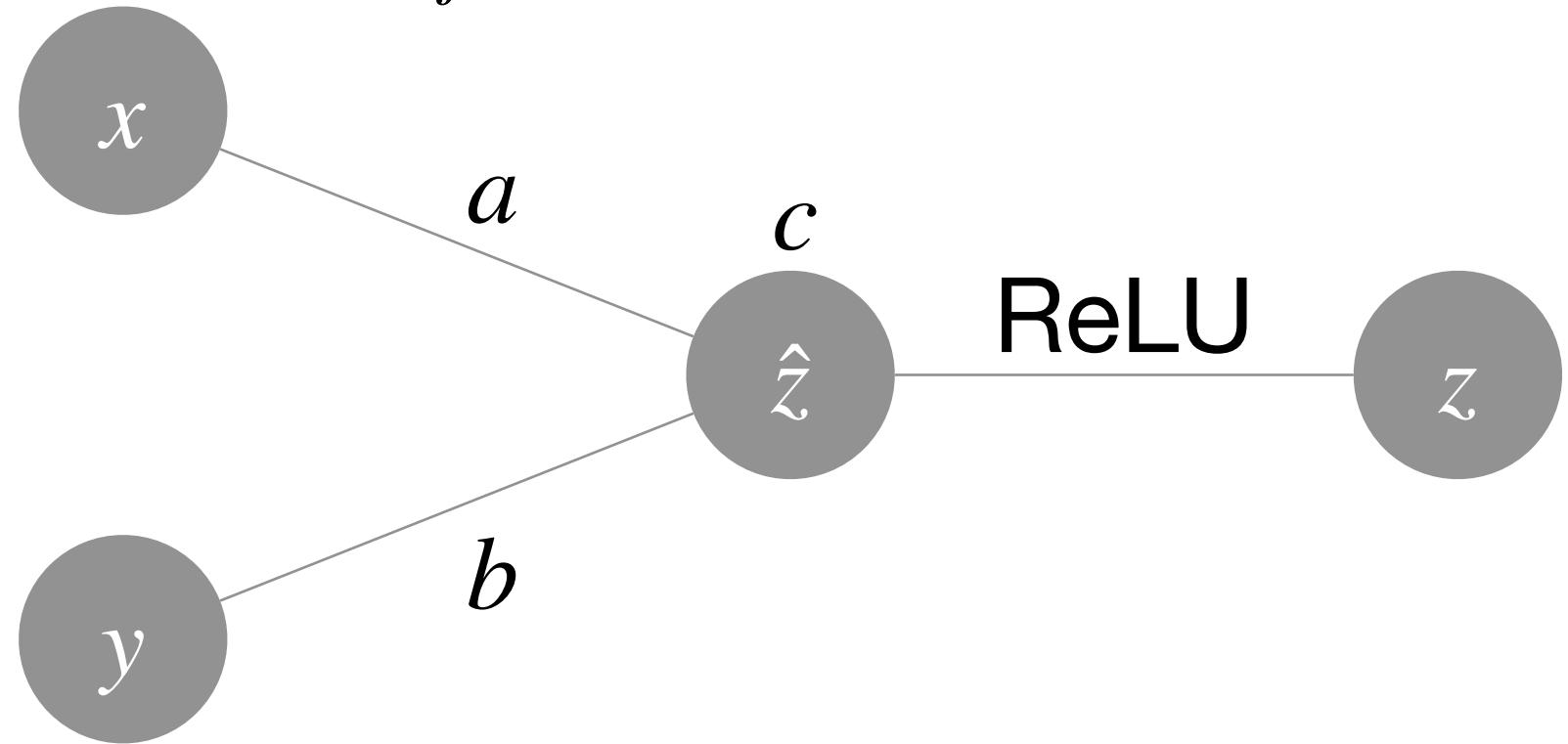


DeepPoly

Singh et al. @ POPL 2019

$$x \in [l_x, u_x]$$

$$\sum_j \underline{m}_j \cdot x_j + \underline{p}_j \leq x \leq \sum_j \bar{m}_j \cdot x_j + \bar{p}_j$$



$$y \in [l_y, u_y]$$

$$\sum_j \underline{n}_j \cdot y_j + \underline{q}_j \leq y \leq \sum_j \bar{n}_j \cdot y_j + \bar{q}_j$$

$$a \geq 0 \wedge b \geq 0 \implies a \cdot \left(\sum_j \underline{m}_j \cdot x_j + \underline{p}_j \right) + b \cdot \left(\sum_j \bar{n}_j \cdot x_j + \bar{q}_j \right) + c \leq$$

$$\hat{z} \leq a \cdot \left(\sum_j \bar{m}_j \cdot x_j + \bar{p}_j \right) + b \cdot \left(\sum_j \bar{n}_j \cdot x_j + \bar{q}_j \right) + c$$

$$\iff \sum_j (a \cdot \underline{m}_j + b \cdot \bar{n}_j) \cdot x_j + a \cdot \underline{p}_j + b \cdot \bar{q}_j + c \leq$$

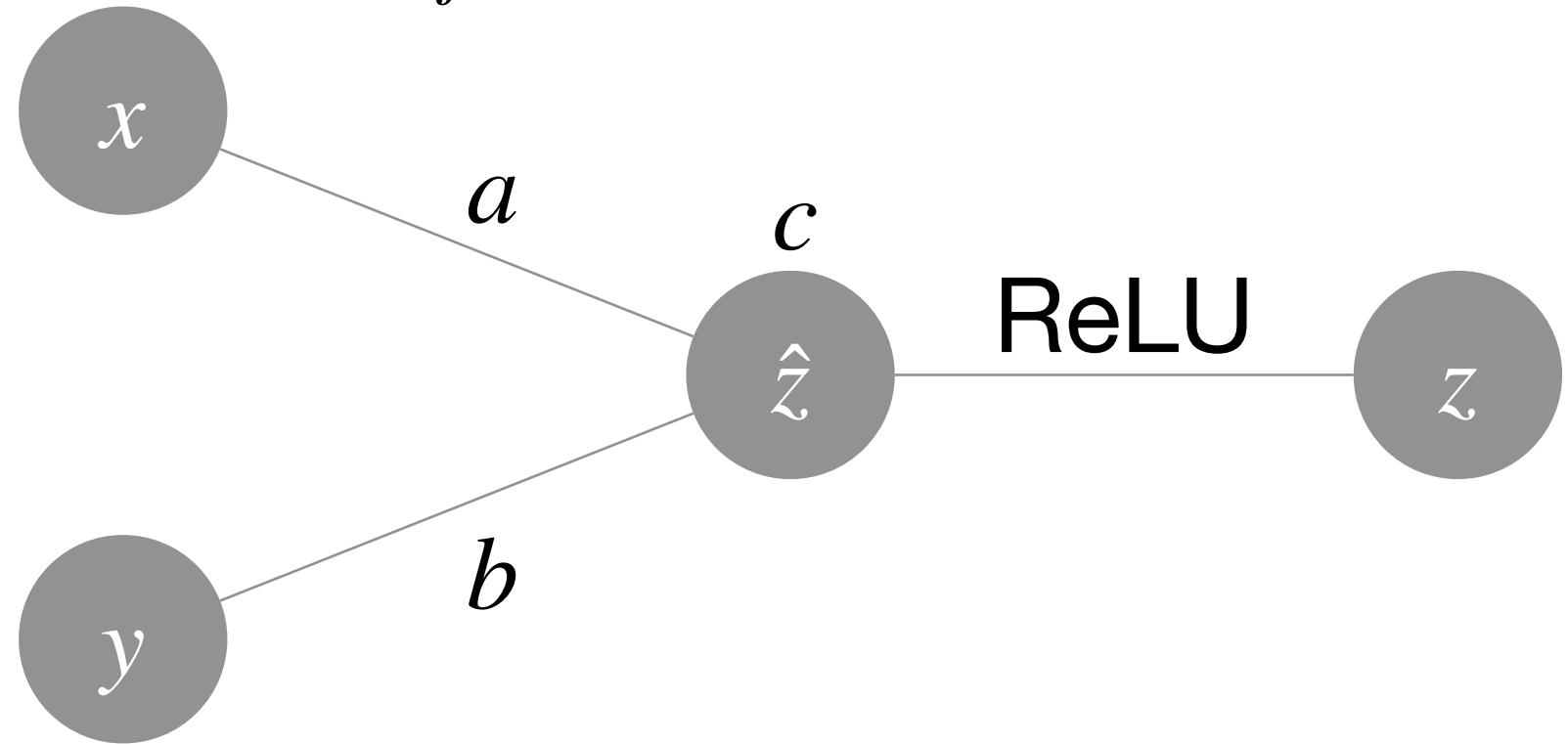
$$\hat{z} \leq \sum_j (a \cdot \bar{m}_j + b \cdot \bar{n}_j) \cdot x_j + a \cdot \bar{p}_j + b \cdot \bar{q}_j + c$$

DeepPoly

Singh et al. @ POPL 2019

$$x \in [l_x, u_x]$$

$$\sum_j \underline{m}_j \cdot x_j + \underline{p}_j \leq x \leq \sum_j \bar{m}_j \cdot x_j + \bar{p}_j$$



$$y \in [l_y, u_y]$$

$$\sum_j \underline{n}_j \cdot y_j + \underline{q}_j \leq y \leq \sum_j \bar{n}_j \cdot y_j + \bar{q}_j$$

$$a \geq 0 \wedge b \geq 0 \implies a \cdot \left(\sum_j \underline{m}_j \cdot x_j + \underline{p}_j \right) + b \cdot \left(\sum_j \bar{n}_j \cdot x_j + \bar{q}_j \right) + c \leq$$

$$\hat{z} \leq a \cdot \left(\sum_j \bar{m}_j \cdot x_j + \bar{p}_j \right) + b \cdot \left(\sum_j \bar{n}_j \cdot x_j + \bar{q}_j \right) + c$$

$$\iff \sum_j (a \cdot \underline{m}_j + b \cdot \bar{n}_j) \cdot x_j + a \cdot \underline{p}_j + b \cdot \bar{q}_j + c \leq$$

$$\hat{z} \leq \sum_j (a \cdot \bar{m}_j + b \cdot \bar{n}_j) \cdot x_j + a \cdot \bar{p}_j + b \cdot \bar{q}_j + c$$

$$\iff \dots \in [l_{\hat{z}}, u_{\hat{z}}]$$

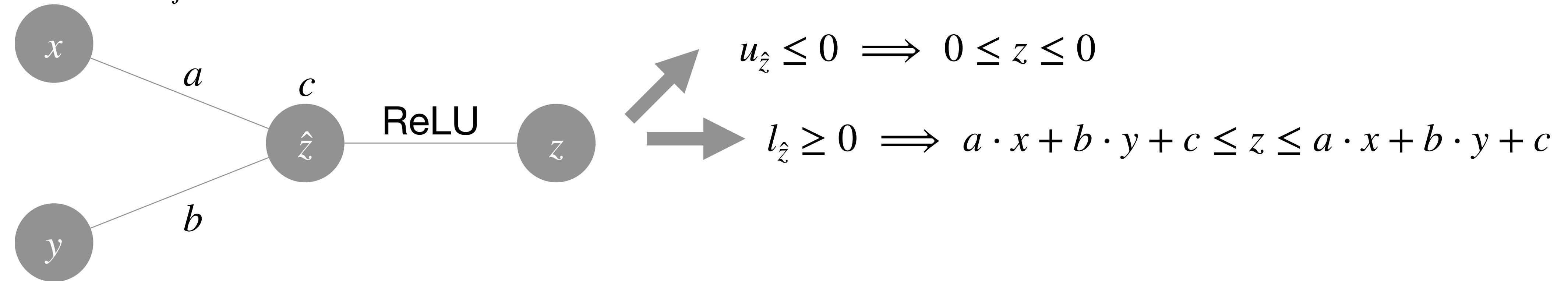
DeepPoly

Singh et al. @ POPL 2019

$$a \cdot x + b \cdot y + c \leq \hat{z} \leq a \cdot x + b \cdot y + c$$

$$x \in [l_x, u_x]$$

$$\sum_j \underline{m}_j \cdot x_j + \underline{p}_j \leq x \leq \sum_j \bar{m}_j \cdot x_j + \bar{p}_j$$



$$y \in [l_y, u_y]$$

$$\sum_j \underline{n}_j \cdot y_j + \underline{q}_j \leq y \leq \sum_j \bar{n}_j \cdot y_j + \bar{q}_j$$

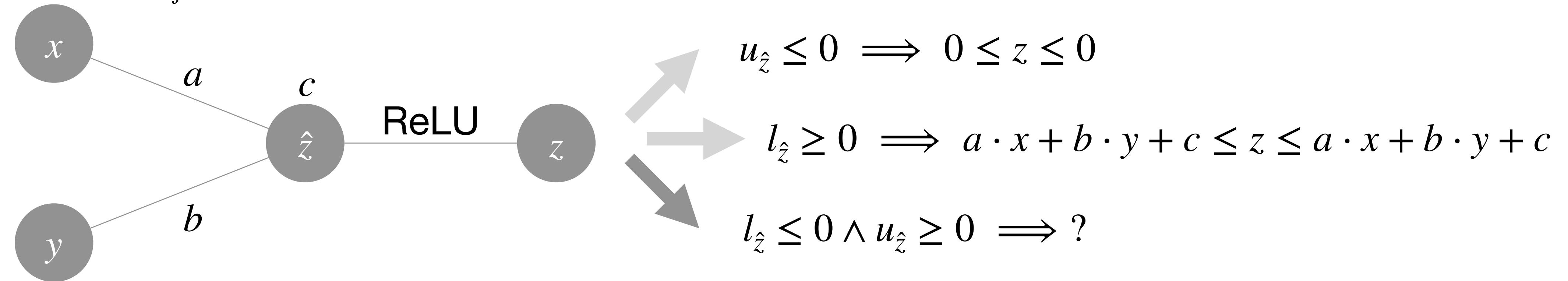
DeepPoly

Singh et al. @ POPL 2019

$$a \cdot x + b \cdot y + c \leq \hat{z} \leq a \cdot x + b \cdot y + c$$

$$x \in [l_x, u_x]$$

$$\sum_j \underline{m}_j \cdot x_j + \underline{p}_j \leq x \leq \sum_j \bar{m}_j \cdot x_j + \bar{p}_j$$

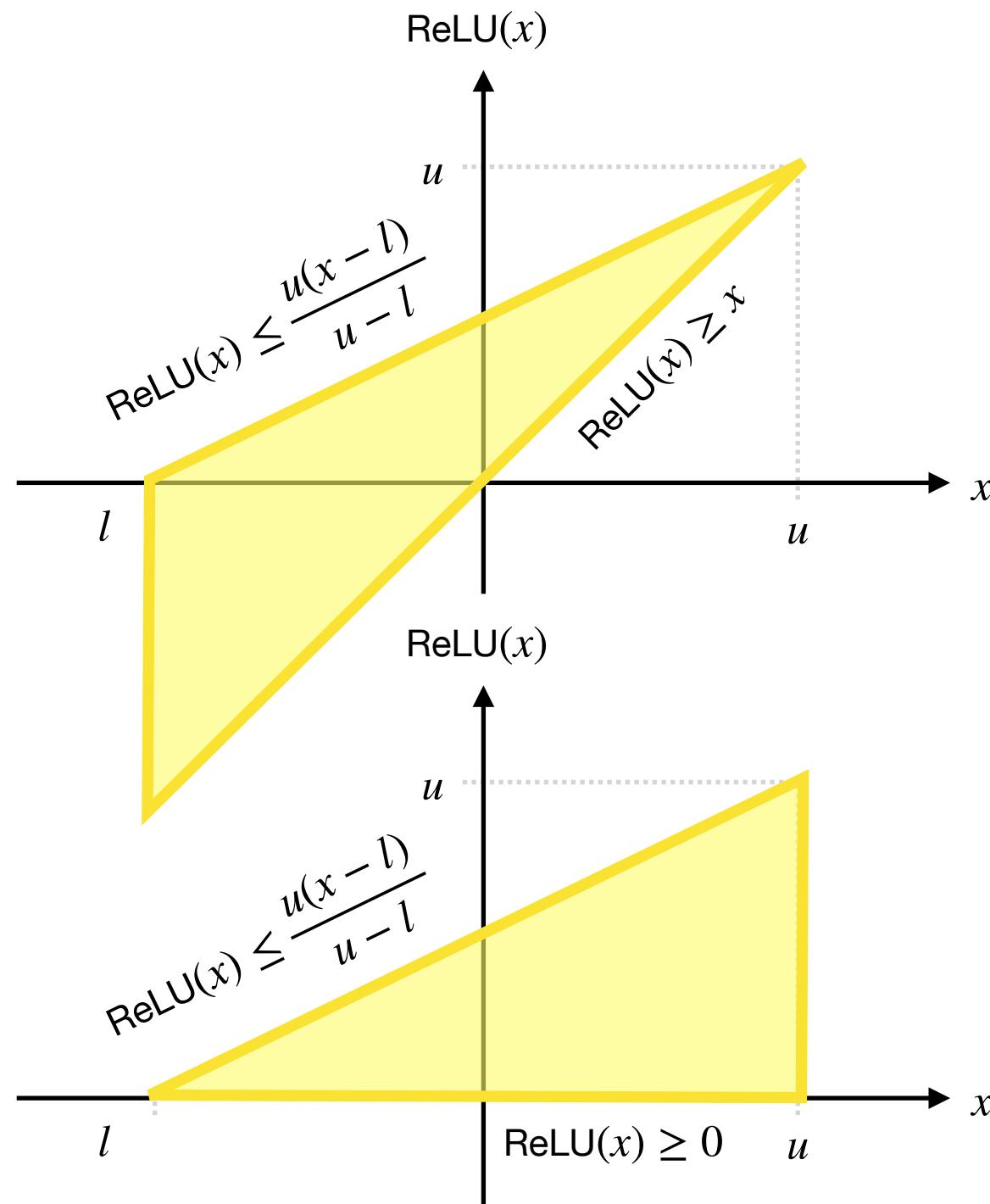


$$y \in [l_y, u_y]$$

$$\sum_j \underline{n}_j \cdot y_j + \underline{q}_j \leq y \leq \sum_j \bar{n}_j \cdot y_j + \bar{q}_j$$

DeepPoly

Singh et al. @ POPL 2019



$$l_{\hat{z}} \leq 0 \wedge u_{\hat{z}} \geq 0 \wedge -l_{\hat{z}} \leq u_{\hat{z}} \implies \begin{cases} z \leq \frac{u_{\hat{z}}(\hat{z} - l_{\hat{z}})}{u_{\hat{z}} - l_{\hat{z}}} \\ z \geq \hat{z} \end{cases}$$

$$l_{\hat{z}} \leq 0 \wedge u_{\hat{z}} \geq 0 \wedge -l_{\hat{z}} > u_{\hat{z}} \implies \begin{cases} z \leq \frac{u_{\hat{z}}(\hat{z} - l_{\hat{z}})}{u_{\hat{z}} - l_{\hat{z}}} \\ z \geq 0 \end{cases}$$

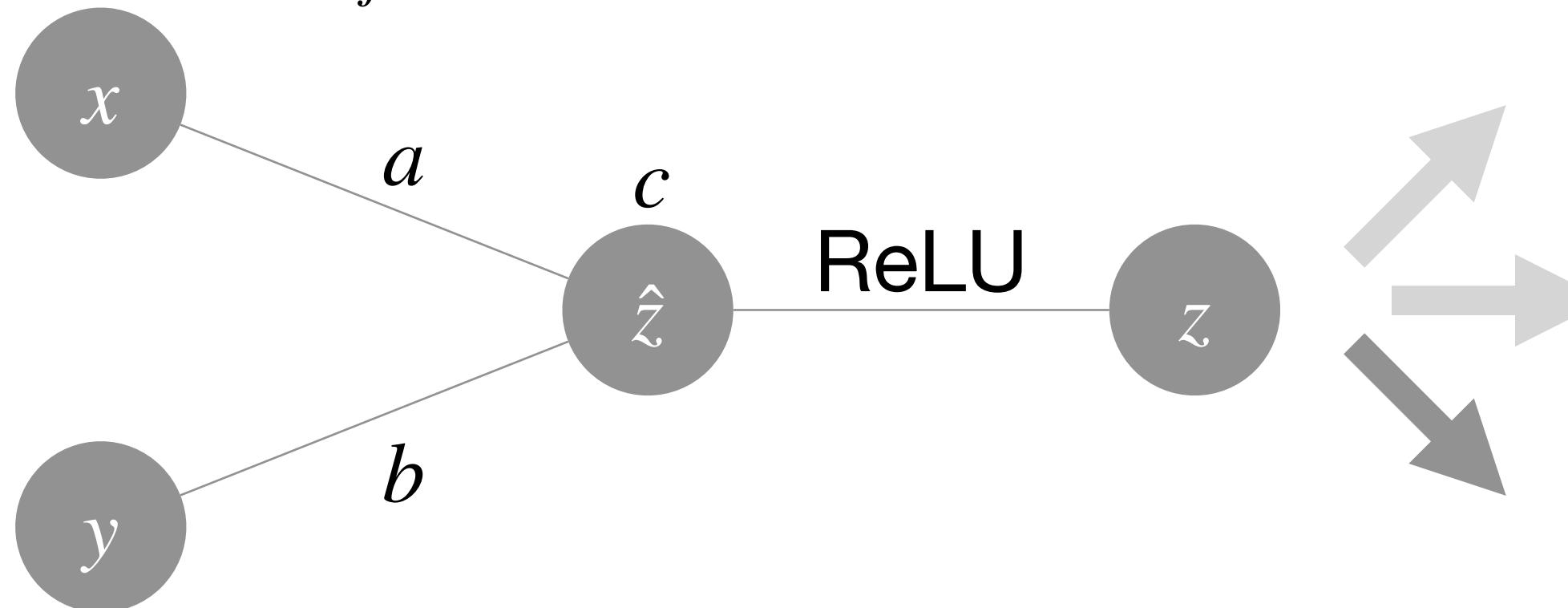
DeepPoly

Singh et al. @ POPL 2019

$$a \cdot x + b \cdot y + c \leq \hat{z} \leq a \cdot x + b \cdot y + c$$

$$x \in [l_x, u_x]$$

$$\sum_j \underline{m}_j \cdot x_j + \underline{p}_j \leq x \leq \sum_j \bar{m}_j \cdot x_j + \bar{p}_j$$



$$y \in [l_y, u_y]$$

$$\sum_j \underline{n}_j \cdot y_j + \underline{q}_j \leq y \leq \sum_j \bar{n}_j \cdot y_j + \bar{q}_j$$

$$u_{\hat{z}} \leq 0 \implies 0 \leq z \leq 0$$

$$l_{\hat{z}} \geq 0 \implies a \cdot x + b \cdot y + c \leq z \leq a \cdot x + b \cdot y + c$$

$$l_{\hat{z}} \leq 0 \wedge u_{\hat{z}} \geq 0 \implies \begin{cases} \hat{z} \leq z \leq \frac{u_{\hat{z}}(\hat{z} - l_{\hat{z}})}{u_{\hat{z}} - l_{\hat{z}}} & -l_{\hat{z}} \leq u_{\hat{z}} \\ 0 \leq z \leq \frac{u_{\hat{z}}(\hat{z} - l_{\hat{z}})}{u_{\hat{z}} - l_{\hat{z}}} & -l_{\hat{z}} > u_{\hat{z}} \end{cases}$$

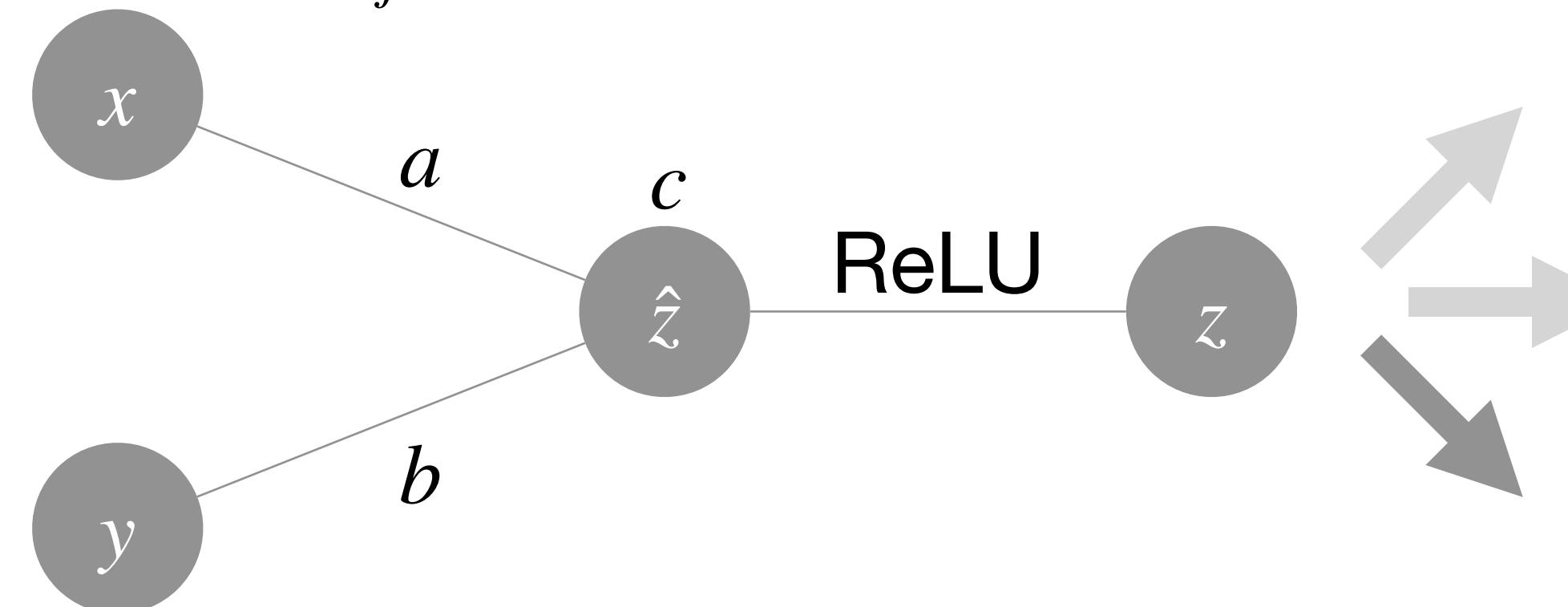
DeepPoly

Singh et al. @ POPL 2019

$$a \cdot x + b \cdot y + c \leq \hat{z} \leq a \cdot x + b \cdot y + c$$

$$x \in [l_x, u_x]$$

$$\sum_j \underline{m}_j \cdot x_j + \underline{p}_j \leq x \leq \sum_j \bar{m}_j \cdot x_j + \bar{p}_j$$



$$y \in [l_y, u_y]$$

$$\sum_j \underline{n}_j \cdot y_j + \underline{q}_j \leq y \leq \sum_j \bar{n}_j \cdot y_j + \bar{q}_j$$

$$u_{\hat{z}} \leq 0 \implies 0 \leq z \leq 0$$

$$l_{\hat{z}} \geq 0 \implies a \cdot x + b \cdot y + c \leq z \leq a \cdot x + b \cdot y + c$$

$$l_{\hat{z}} \leq 0 \wedge u_{\hat{z}} \geq 0 \implies \begin{cases} \hat{z} \leq z \leq \frac{u_{\hat{z}}(\hat{z} - l_{\hat{z}})}{u_{\hat{z}} - l_{\hat{z}}} & -l_{\hat{z}} \leq u_{\hat{z}} \\ 0 \leq z \leq \frac{u_{\hat{z}}(\hat{z} - l_{\hat{z}})}{u_{\hat{z}} - l_{\hat{z}}} & -l_{\hat{z}} > u_{\hat{z}} \end{cases}$$

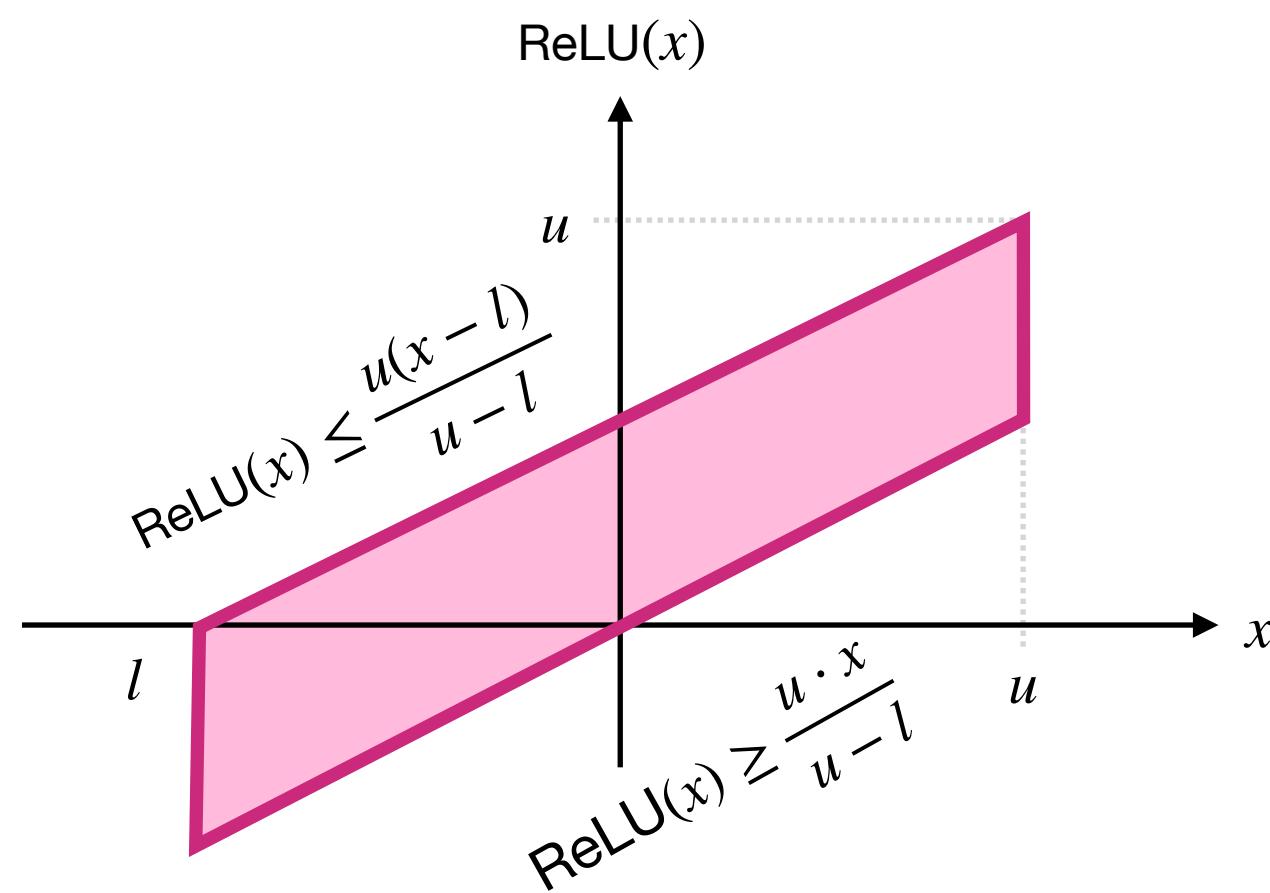
\cap

Symbolic $\implies z \leftarrow s$

Neurify

Wang et al. @ NeurIPS 2018

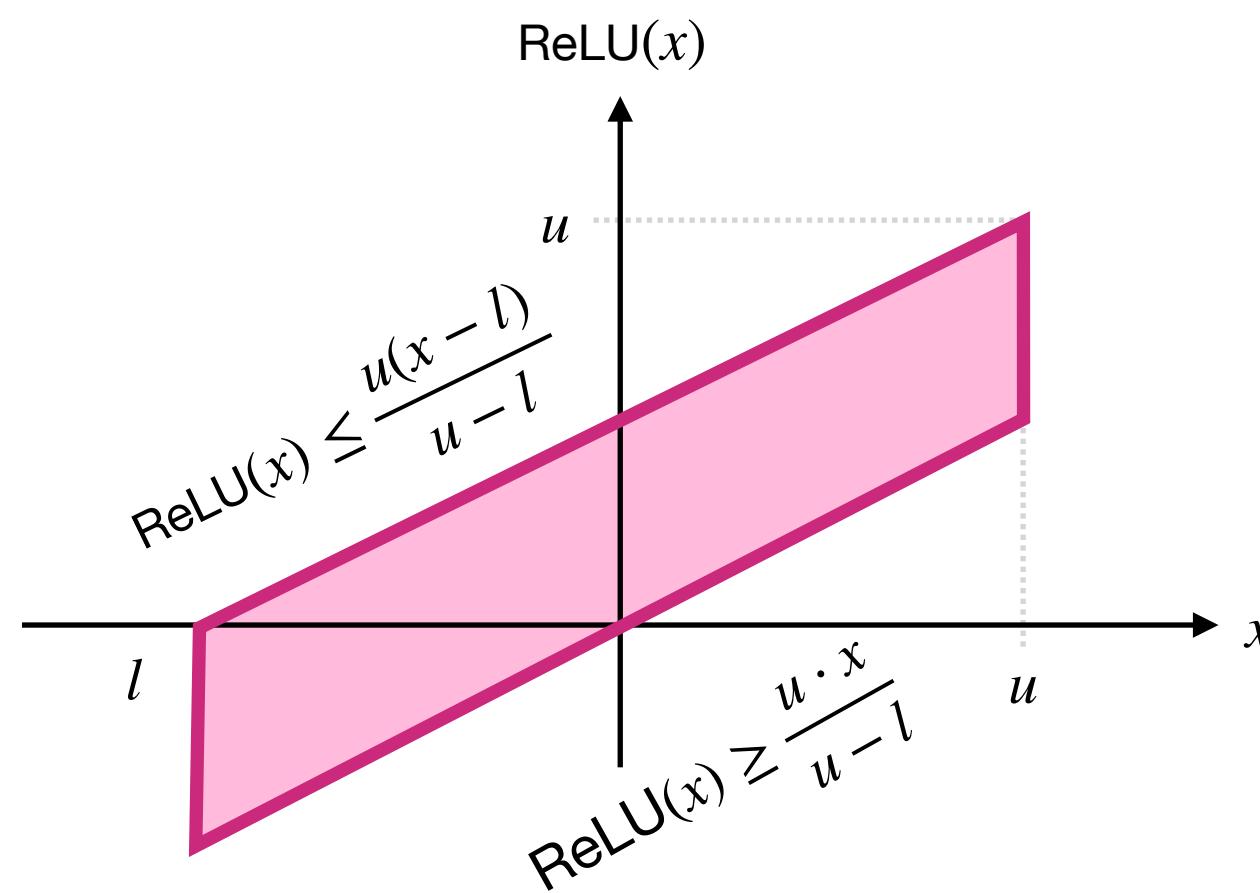
$$z \leftarrow \text{ReLU}(\hat{z}) \quad \hat{z} \in [l_{\text{low}}, l_{\text{up}}, u_{\text{low}}, u_{\text{up}}]$$



Neurify

Wang et al. @ NeurIPS 2018

$$z \leftarrow \text{ReLU}(\hat{z}) \quad \hat{z} \in [l_{\text{low}}, l_{\text{up}}, u_{\text{low}}, u_{\text{up}}]$$

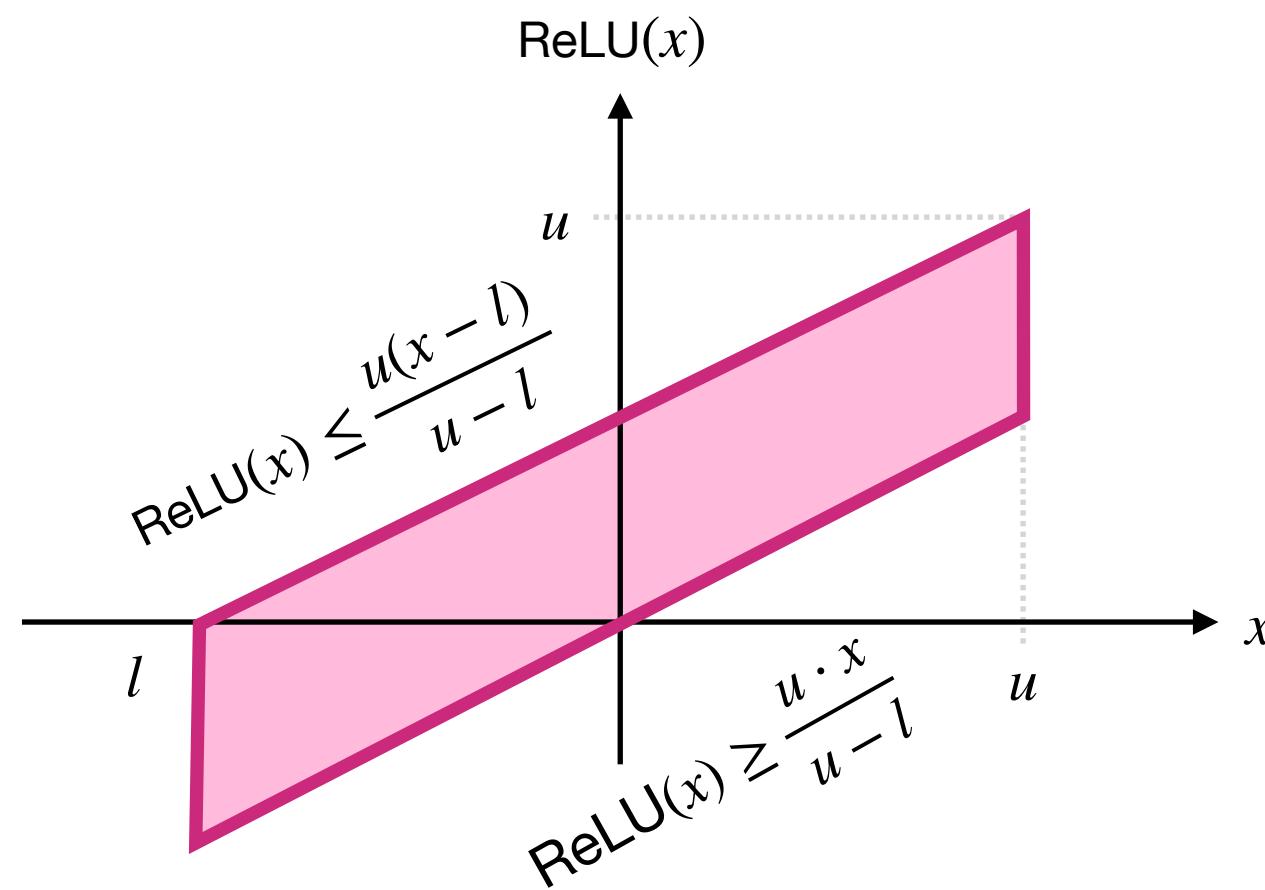


$$\underline{\text{ReLU}}(\hat{z}) \leq z \leq \overline{\text{ReLU}}(\hat{z})$$

Neurify

Wang et al. @ NeurIPS 2018

$$z \leftarrow \text{ReLU}(\hat{z}) \quad \hat{z} \in [l_{\text{low}}, l_{\text{up}}, u_{\text{low}}, u_{\text{up}}]$$



$$\underline{\text{ReLU}}(\hat{z}) \leq z \leq \overline{\text{ReLU}}(\hat{z})$$

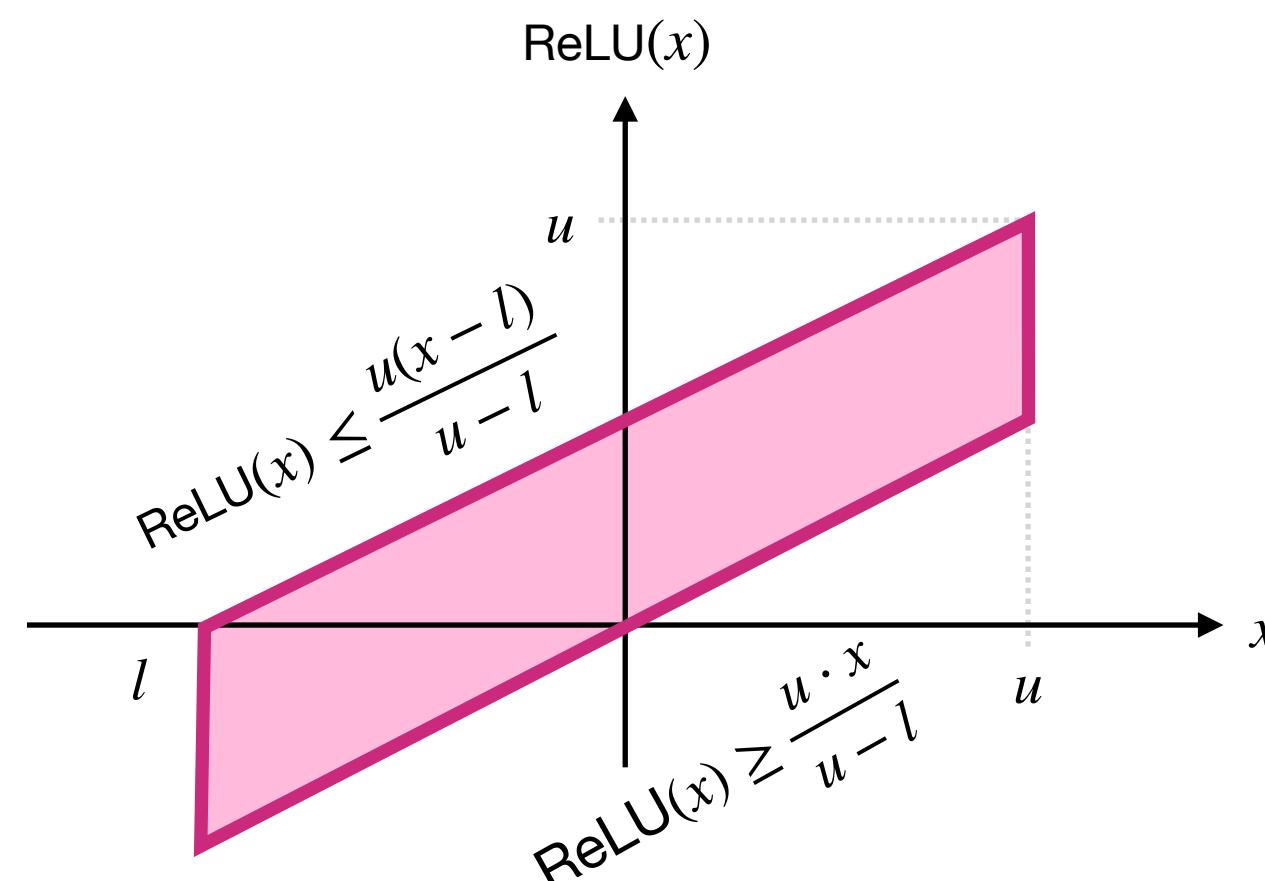
$$l_{\text{low}} \leq 0 \wedge l_{\text{up}} \geq 0 \implies \underline{\text{ReLU}}(\hat{z}) \leq z$$

$$l_{\text{low}} \leq 0 \wedge u_{\text{up}} \geq 0 \wedge$$

Neurify

Wang et al. @ NeurIPS 2018

$$z \leftarrow \text{ReLU}(\hat{z}) \quad \hat{z} \in [l_{\text{low}}, l_{\text{up}}, u_{\text{low}}, u_{\text{up}}]$$



$$l_{\text{low}} \leq 0 \wedge u_{\text{up}} \geq 0 \wedge$$

$$\underline{\text{ReLU}}(\hat{z}) \leq z \leq \overline{\text{ReLU}}(\hat{z})$$

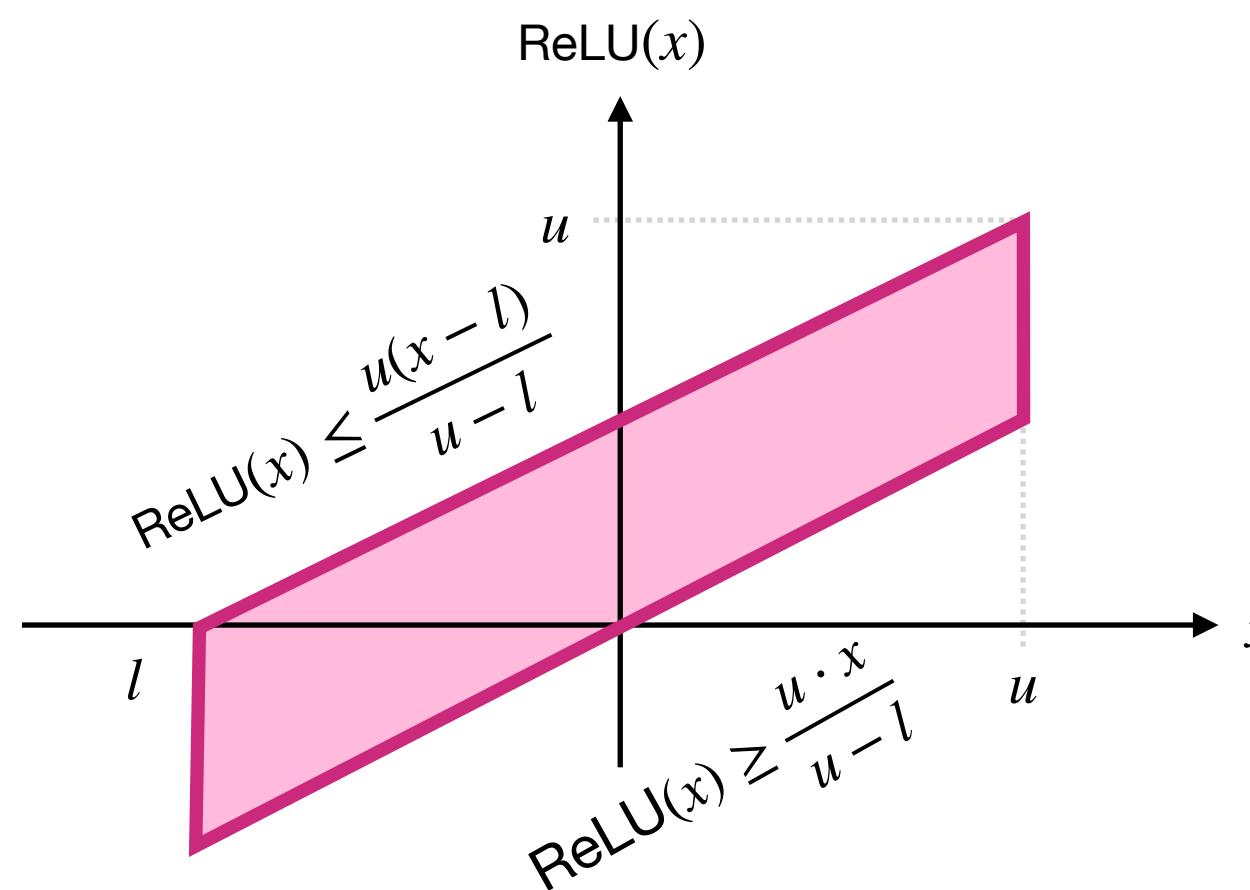
$$l_{\text{low}} \leq 0 \wedge l_{\text{up}} \geq 0 \implies \underline{\text{ReLU}}(\hat{z}) \leq z$$

$$l_{\text{low}} \leq 0 \wedge l_{\text{up}} \leq 0 \implies 0 \leq z$$

Neurify

Wang et al. @ NeurIPS 2018

$$z \leftarrow \text{ReLU}(\hat{z}) \quad \hat{z} \in [l_{\text{low}}, l_{\text{up}}, u_{\text{low}}, u_{\text{up}}]$$



$$l_{\text{low}} \leq 0 \wedge u_{\text{up}} \geq 0 \wedge$$

$$\underline{\text{ReLU}}(\hat{z}) \leq z \leq \overline{\text{ReLU}}(\hat{z})$$

$$l_{\text{low}} \leq 0 \wedge l_{\text{up}} \geq 0 \implies \underline{\text{ReLU}}(\hat{z}) \leq z$$

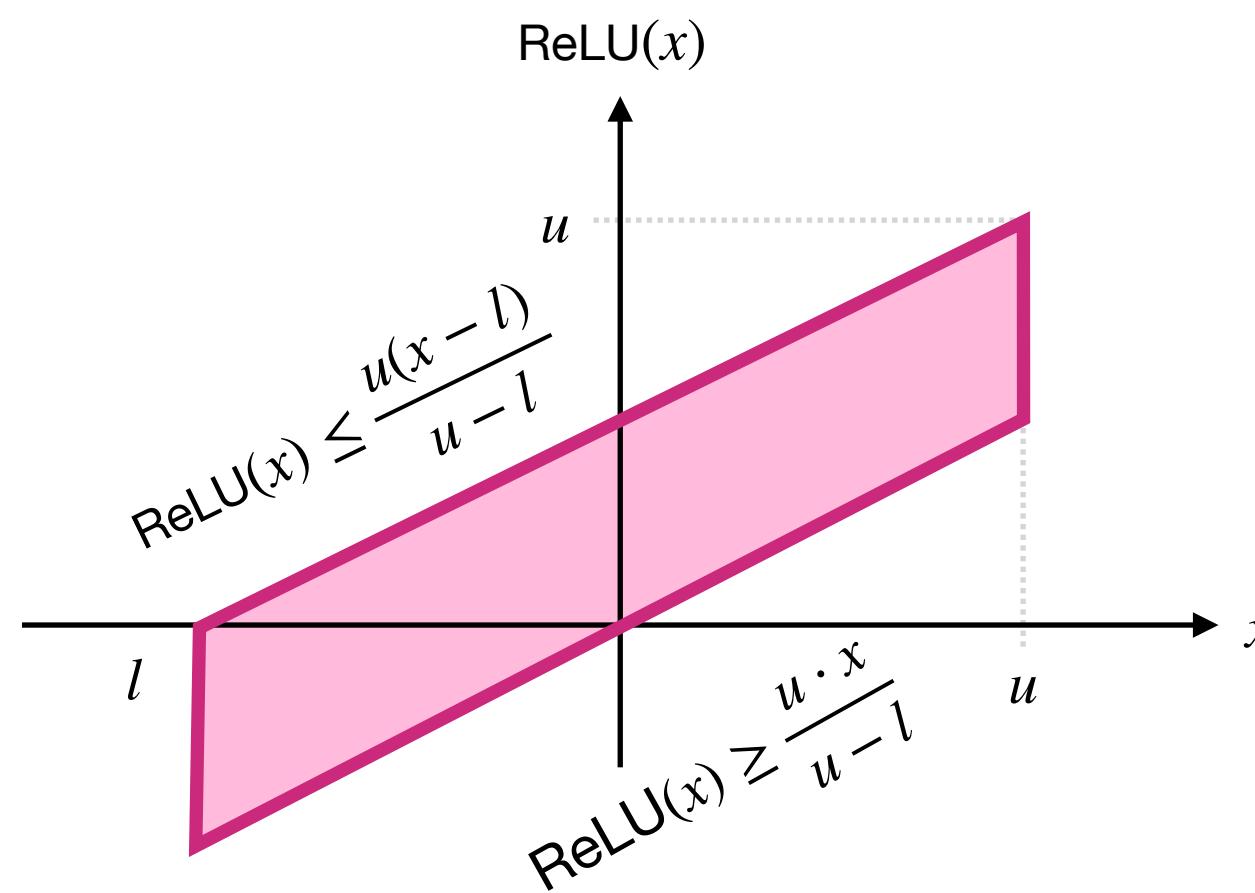
$$l_{\text{low}} \leq 0 \wedge l_{\text{up}} \leq 0 \implies 0 \leq z$$

$$u_{\text{low}} \leq 0 \wedge u_{\text{up}} \geq 0 \implies z \leq \overline{\text{ReLU}}(\hat{z})$$

Neurify

Wang et al. @ NeurIPS 2018

$$z \leftarrow \text{ReLU}(\hat{z}) \quad \hat{z} \in [l_{\text{low}}, l_{\text{up}}, u_{\text{low}}, u_{\text{up}}]$$



$$l_{\text{low}} \leq 0 \wedge u_{\text{up}} \geq 0 \wedge$$

$$\underline{\text{ReLU}}(\hat{z}) \leq z \leq \overline{\text{ReLU}}(\hat{z})$$

$$l_{\text{low}} \leq 0 \wedge l_{\text{up}} \geq 0 \implies \underline{\text{ReLU}}(\hat{z}) \leq z$$

$$l_{\text{low}} \leq 0 \wedge l_{\text{up}} \leq 0 \implies 0 \leq z$$

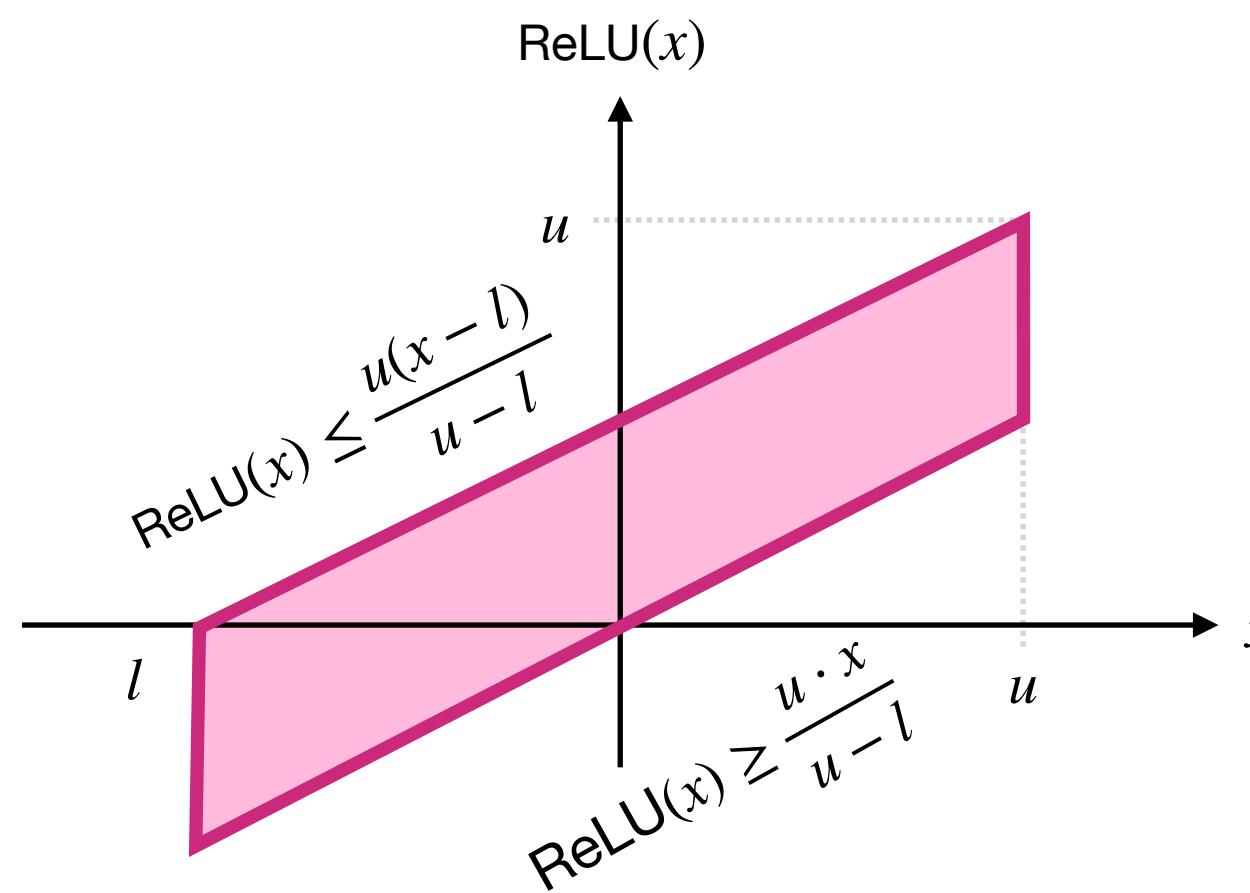
$$u_{\text{low}} \leq 0 \wedge u_{\text{up}} \geq 0 \implies z \leq \overline{\text{ReLU}}(\hat{z})$$

$$u_{\text{low}} \geq 0 \wedge u_{\text{up}} \geq 0 \implies z \leq \hat{z}$$

Neurify

Wang et al. @ NeurIPS 2018

$$z \leftarrow \text{ReLU}(\hat{z}) \quad \hat{z} \in [l_{\text{low}}, l_{\text{up}}, u_{\text{low}}, u_{\text{up}}]$$



$$l_{\text{low}} \leq 0 \wedge u_{\text{up}} \geq 0 \wedge$$

$$\underline{\text{ReLU}}(\hat{z}) \leq z \leq \overline{\text{ReLU}}(\hat{z})$$

$$l_{\text{low}} \leq 0 \wedge l_{\text{up}} \geq 0 \implies \underline{\text{ReLU}}(\hat{z}) \leq z$$

$$l_{\text{low}} \leq 0 \wedge l_{\text{up}} \leq 0 \implies 0 \leq z \quad \xleftarrow{\text{No relaxation}}$$

$$u_{\text{low}} \leq 0 \wedge u_{\text{up}} \geq 0 \implies z \leq \overline{\text{ReLU}}(\hat{z})$$

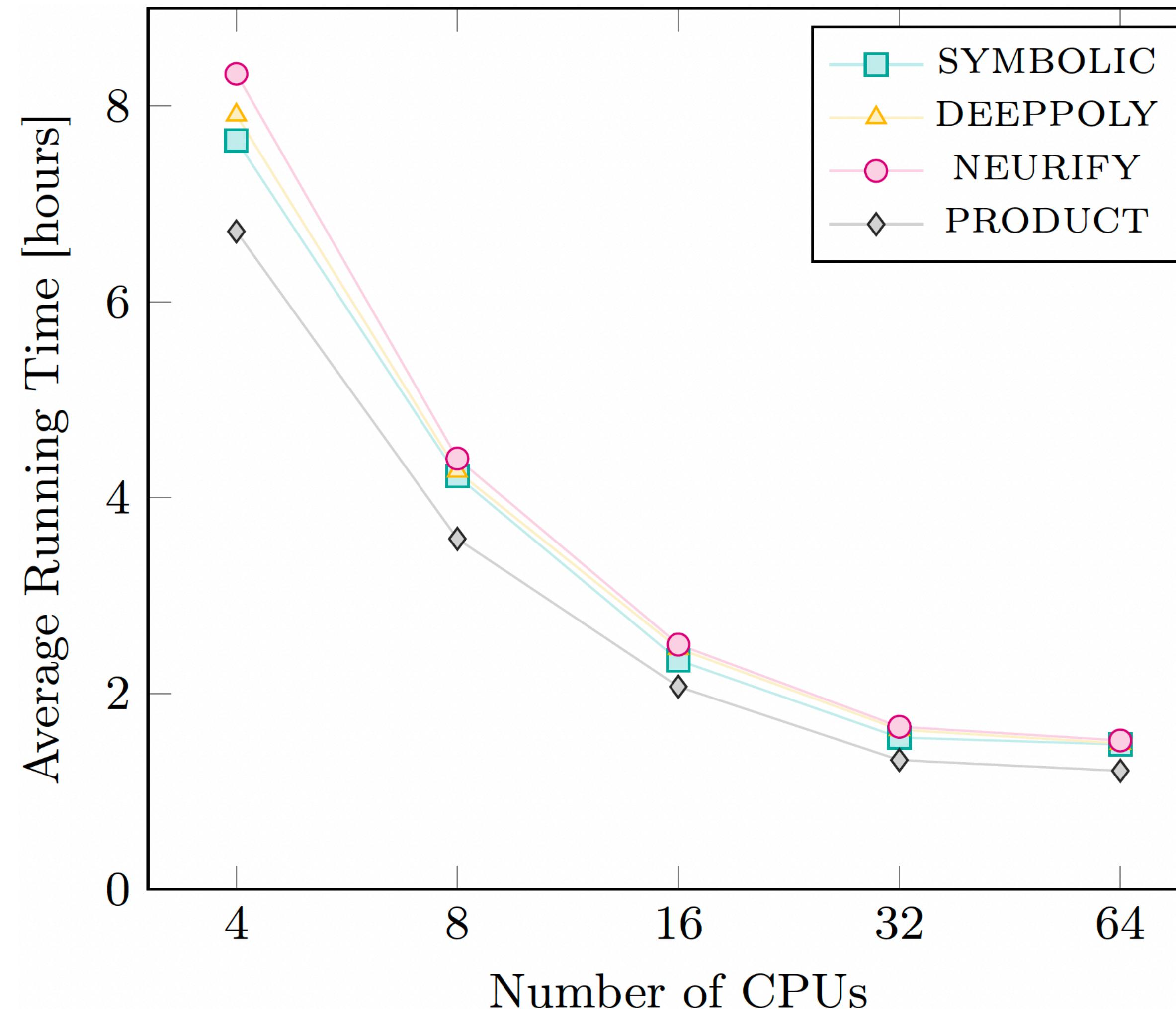
$$u_{\text{low}} \geq 0 \wedge u_{\text{up}} \geq 0 \implies z \leq \hat{z} \quad \xleftarrow{\text{No relaxation}}$$

Effect of Neural Network Structure

Size	Symbolic	DeepPoly	Neurify	Product	
10	98.72%	98.37%	98.51%	99.44%	+0.7%
12	76.70%	66.39%	64.58%	77.29%	+0.6%
20	56.11%	56.10%	53.06%	68.23%	+12.1%
40	34.72%	38.69%	41.22%	51.18%	+10%
45	43.78%	51.21%	50.59%	55.53%	+4.3%



Leveraging Multiple CPUs



Perfectly Parallelisation

