A Mixed Integer Programming Approach for Verifying Properties of Binarized Neural Networks

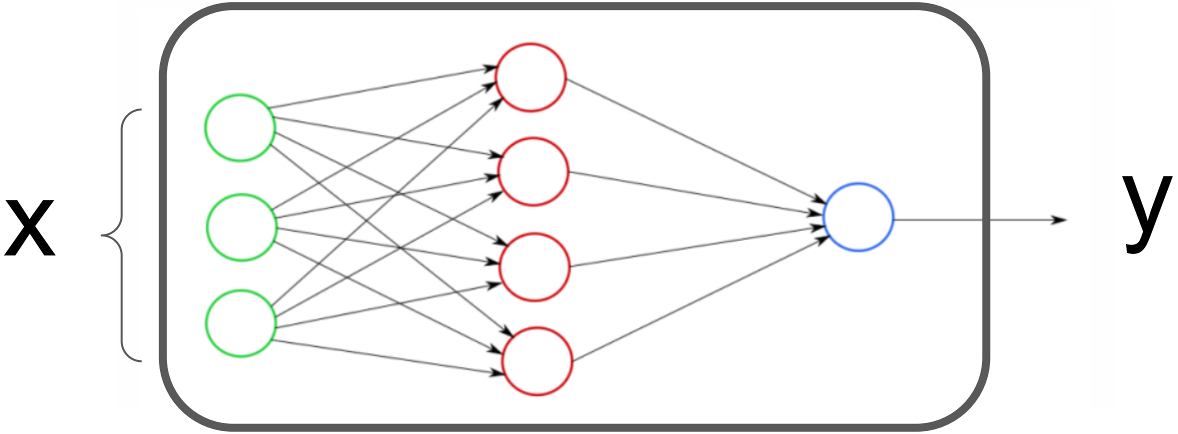
AISafety 2021 - IJCAI-21

Christopher Lazarus

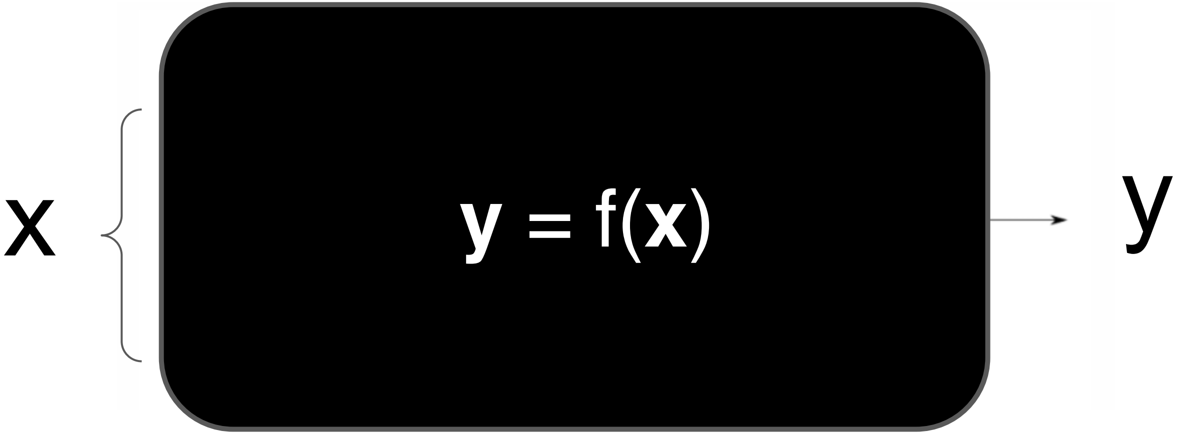
20 August 2021

SISL
Stanford Intelligent Systems Laboratory

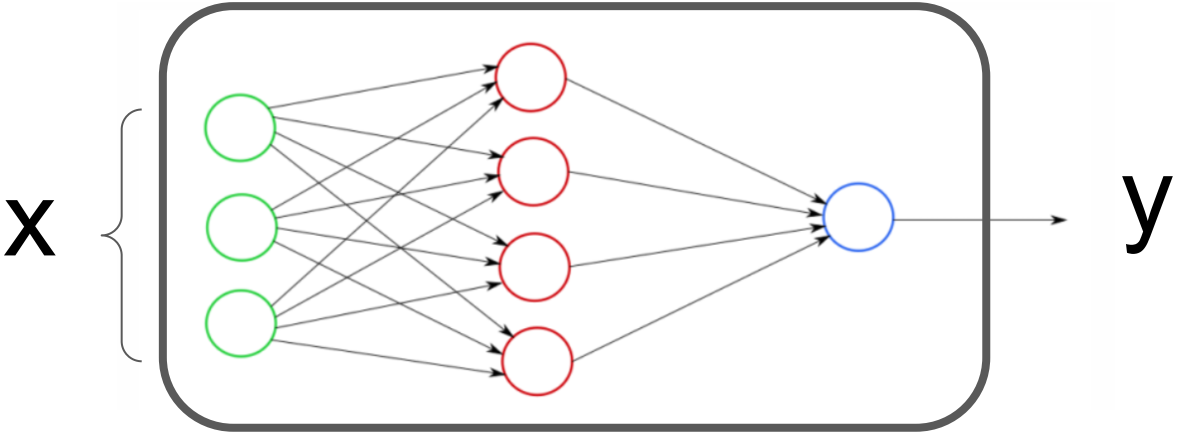
Neural Network Verification



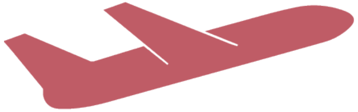
Neural Network Verification



Neural Network Verification



Car icon created by Alena Artemova from the Noun Project



Plane icon created by i cons from the Noun Project

Satellite icon created by Maxim Samos from the Noun Project

Testing Can Fail

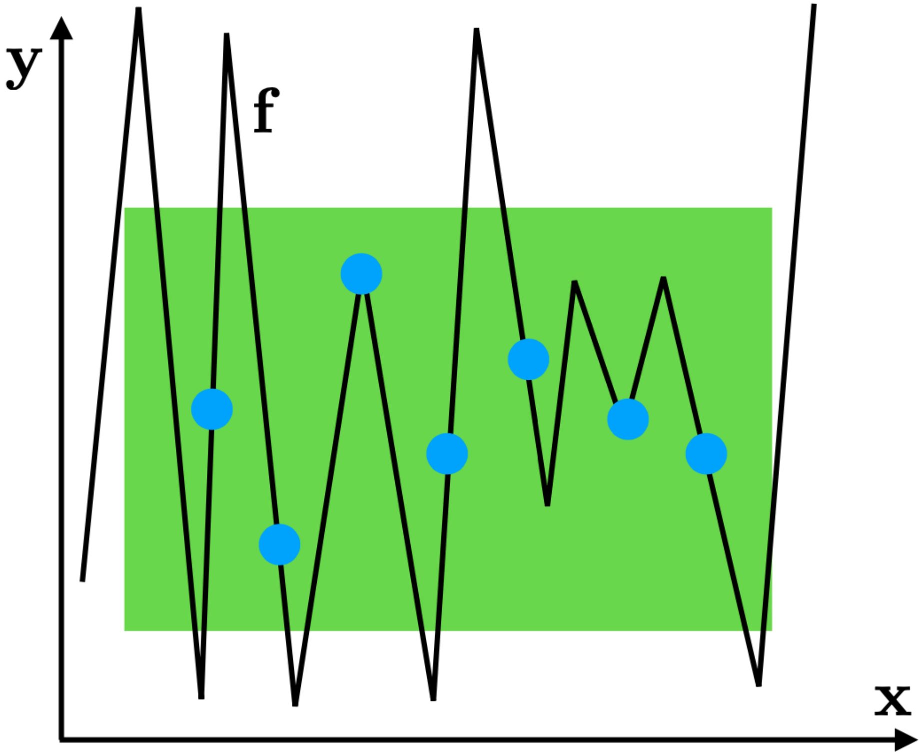
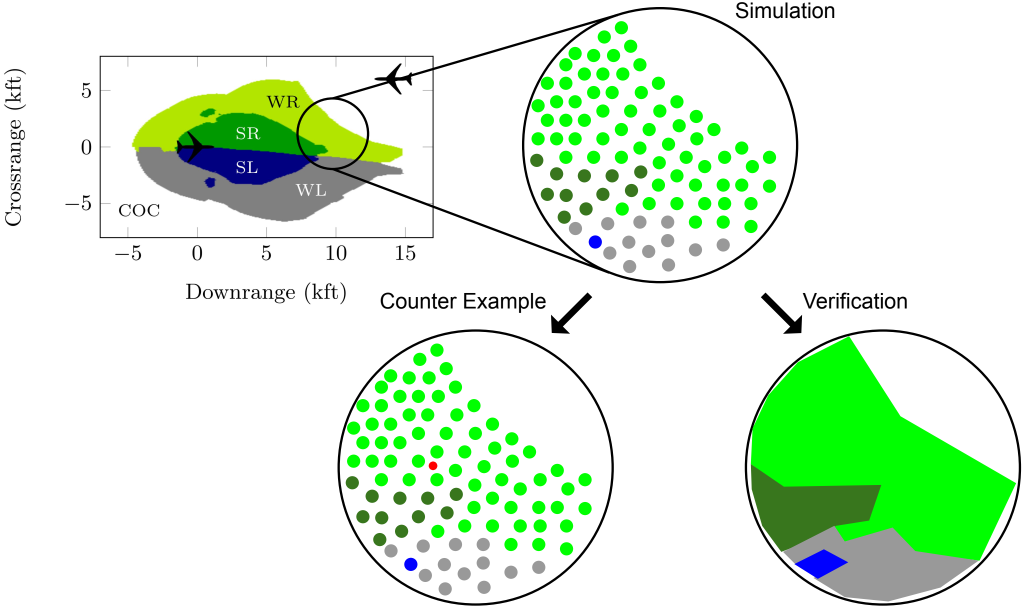
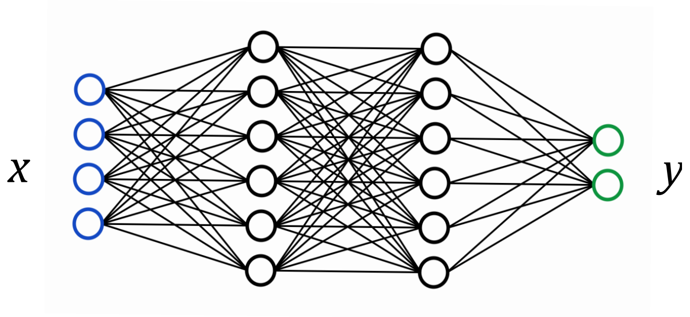


Image credit: Changliu Liu, http://www.cs.cmu.edu/~cliu6/files/NN-slides.pdf

Testing Can Fail



Verifying Neural Networks



$$(x \in \mathcal{X} \wedge y = NN(x)) \Rightarrow y \in \mathcal{Y}$$

Challenges

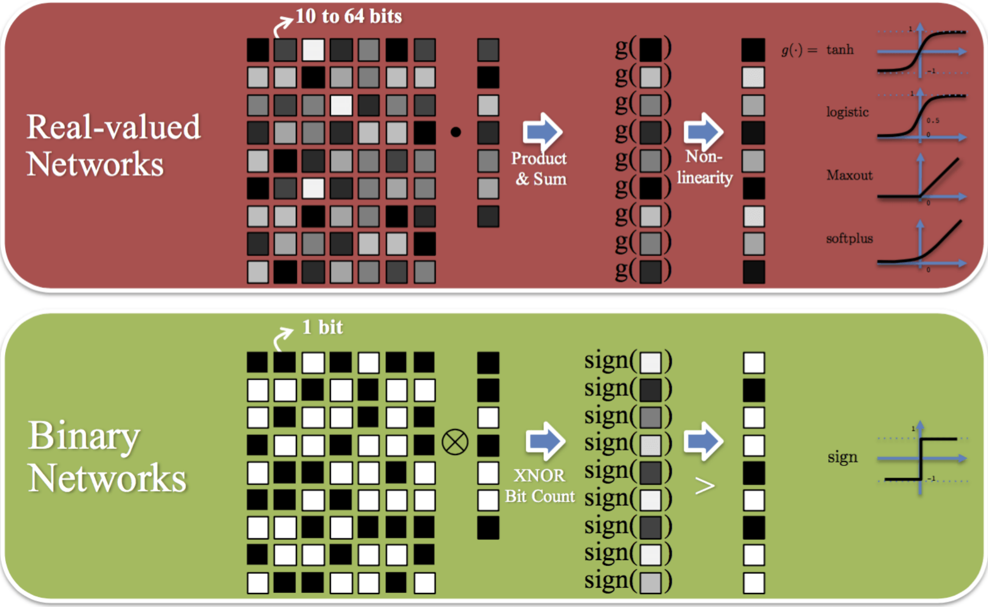
Testing

Pros: Scalable
Cons: Not Sound

Formal Verification

Pros: Sound
Cons: Not scalable

Binarized Neural Networks



[Figure: https://mohitjain.me/2018/07/14/bnn/]

Linear. $\hat{z}_i = Q_i z_{i-1}$ where $Q_i \in \{-1, 1\}^{k_{i+1} \times k_i}$

$$\hat{z}_{i,j} = q_j^T z_{i-1} \qquad j = 1, \ldots, k_{i+1}$$

ReLU. $z_i = \text{ReLU}(\hat{z}_i) = \max(0, \hat{z}_i)$ and given that $l \leq \hat{z}_i \leq u$ an can encode the block as:

$$z_i \leq \hat{z}_i - l(1-\beta)$$
$$\hat{z}_i \leq z_i$$
$$z_i \leq \beta \cdot u$$
$$0 \leq z_i$$
$$\beta \in \{0, 1\}$$

Sign. $z_i = \text{sign}(\hat{z}_i)$

$$\hat{z}_i \geq 0 \implies z_i = 1$$
$$\hat{z}_i < 0 \implies z_i = 0$$

but given bounds $l \leq z_i \leq u$, this can be formulated as

$$-1 \leq z_i$$
$$z_i \leq 1$$
$$l \cdot \beta \leq \hat{z}_i$$
$$\hat{z}_i \leq u(1-\beta)$$
$$z_i = 1 - 2 \cdot \beta$$
$$\beta \in \{0, 1\}$$

Experiments

ACAS

	ϵ	Time (s)			Accuracy (%)	
		Mean	Max	timeout	Verified	Data
BNN	0.1	0.223	3.21	0.00	88.24	95.6
DNN		5.47	28.12	0.05%	94.33	98.22
BNN	0.3	0.194	4.54	0.00	61.78	95.6
DNN		7.12	41.33	1.02%	80.68	98.22

MNIST

		Loss	time (s)	result
full precision	BNN	2174.43	2.37	violated
	DNN	1203.44	41.44	holds
8 bit	BNN	1634.25	5.73	holds

Formal Verification is Key for Safety Critical Systems