With the rise of Neural Networks, they are also applied in safety-critical systems (e.g. autonomous cars). It is important to prove their safety, however this definition may be. Since the verification is currently not even scalable to small NNs, we focus on abstraction, i.e. a method to reduce the size of the verification problem.

### Idea

- **Goal**: Verification of NNs
- **Problem**: Scalability
- **Solution**: Abstraction

Get semantic information of the neurons and feed them to the network. We then capture the activation values (=outputs) of the neurons. Based on an IO-set $X$ calculate the activation values of the neurons. We use the inputs $x \in X$ and find the network. We then capture the activation values (outputs) of the neurons.

### Hypothesis

Replacing neurons with linear combinations of other neurons and a fixed number of neurons and increasing number of layers.

**Error Bound Results**

Error of the neurons on the test-set of an MNIST network with $3 \times 100$ neurons, reduced by 30%.

**Theorem for Relation between Original and Abstraction**

The difference between the original $N^L(x)$ and the abstraction $\tilde{N}^L(x)$ can be bounded by

$$\|N^L(x) - \tilde{N}^L(x)\| \leq b(1 - a^{L-1})(1 - a)$$

assuming that for all layers $l \in \{1, \ldots, L\}$ and for all inputs $x \in X$, we have:

- for $i \in \{0\}$: $|\sigma_i(x) - \sigma_i(0)| \leq e_i(0)$
- $\sum_{i=0}^{L} W_i^L \sum_{i=0}^{L} a_i(0) \leq e_i(0)$

### Guarantees

The error of the neurons on the test set of an MNIST network with $3 \times 100$ neurons, reduced by 30%.

### Results

- **Mean Runtime [s]**
- **Accuracy**

### Syntactic VS Semantic

**DeepAbstract** [1]: Clustering of neurons based on the semantic information (many by one)

**Bisimulation** [2]: Replacing neurons based on the syntactic information, e.g. their weights and biases (many by many)

### Comparison

The greedy approach performs best in terms of accuracy.

**LINNA** has a higher accuracy than any other approach (even for the heuristic based approach).

Runtime is **magnitudes faster** than the predecessor DeepAbstract.

---