

Formal Verification of Neural Networks

Tobias Ladner, Lukas Koller

Prof. Dr.-Ing. Matthias Althoff
Cyber-Physical Systems Group
Technische Universität München

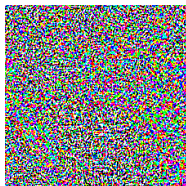
February 11th, 2025

Motivation



“panda” ✓

+ 0.007 ×



noise

=



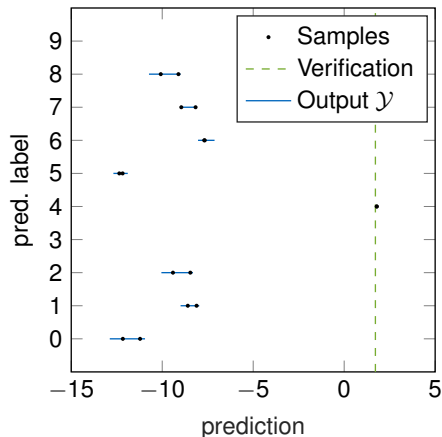
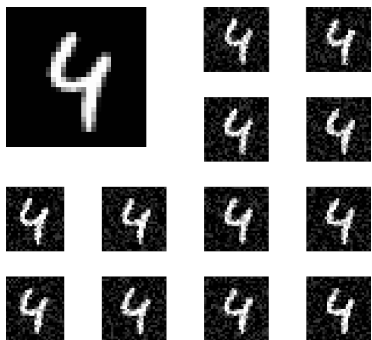
“gibbon” ✗

Adversarial attacks¹ limit the applicability of neural networks in cyber-physical systems!

¹ Ian Goodfellow, Jonathon Shlens, and Christian Szegedy. “Explaining and harnessing adversarial examples”. In: *International Conference on Learning Representations*. 2015

Motivation

Let us demonstrate the formal verification of neural networks by an example:



Topics

Possible topics:

- Design verification algorithms for general network architectures
- Enable easier verification by network design
- ...

Further examples can be found on the CORA website².

Interested? Contact us!

Tobias Ladner, Lukas Koller

`tobias.ladner@tum.de`, `lukas.koller@tum.de`

²CORA website: <https://cora.in.tum.de/pages/neural-networks>