University of Hildesheim

International Master's Program on Data Analytics

Seminar III - Adversarial Machine Learning

Agenda of Paper Presentations

Vijaya Krishna Yalavarthi

November 5, 2019

- **19.11.19** at 14:00:
  Goodfellow et al., Explaining and Harnessing Adversarial Examples, ICLR 2015
- **19.11.19** at 15:00:
  Miyato et al., Distributional Smoothing with Adversarial Training, ICLR 2016
- **26.11.19** at 14:00:
  Dmoosavi-Dezfoli et al., DeepFool: a simple and accurate method to fool deep neural networks, CVPR 2016
- **26.11.19** at 15:00:
  Papernot et al., The Limitations of Deep Learning in Adversarial Settings, IEEE European Symposium on Security and Privacy 2016
- **03.12.19** at 14:00:
  Carlini and Wagner, Towards Evaluating the Robustness of Neural Networks, IEEE Symposium on Security and Privacy 2017
- **03.12.19** at 15:00:
  Brendel et al., Decision-Based Adversarial Attacks: Reliable Attacks Against Black-Box Machine Learning Models, ICLR 2018
- **10.12.19** at 14:00:
  Brown et al., Adversarial Patch, NIPS 2017
- **10.12.19** at 15:00:
  Chen et al., EAD: Elastic-Net Attacks to Deep Neural Networks via Adversarial Examples, AAAI 2018
- **17.12.19** at 14:00:
  Chen et al., ZOO: Zeroth Order Optimization based Black-box Attacks to Deep Neural Networks without Training Substitute Models, AISec 2017

- **17.12.19** at 15:00:
  Ilyas et al., Black-box Adversarial Attacks with Limited Queries and Information, ICML 2018
- **07.01.20** at 14:00:
  Liao et al., Defense against Adversarial Attacks Using High-Level Representation Guided Denoiser, CVPR 2018
- **07.01.20** at 15:00:
  Madry et al., Towards Deep Learning Models Resistant to Adversarial Attacks, ICLR 2018
- **14.01.20** at 14:00:
  Song et al., PixelDefend: Leveraging Generative Models to Understand and Defend against Adversarial Examples, ICLR 2018
- **14.01.20** at 15:00:
  Buckman et al., Thermometer Encoding: One Hot Way To Resist Adversarial Examples, ICLR 2018
- **21.01.20** at 14:00:
  Chen et al., Detecting Backdoor Attacks on Deep Neural Networks by Activation Clustering, Artificial Intelligence Safety Workshop @ AAAI 2019
- **21.01.20** at 15:00:
  Grosse et al., The Limitations of Model Uncertainty in Adversarial Settings, ArXiv 2018
- **28.01.20** at 14:00:
  Guo et al., Countering Adversarial Images using Input Transformations, ICLR 2018
- **28.01.20** at 15:00:
  Wang et al., MixTrain: Scalable Training of Verifiably Robust Neural Networks, ArXiv 2018
- **04.02.20** at 14:00:
  Weng et al., Evaluating the Robustness of Neural Networks: An Extreme Value Theory Approach, ICLR 2018
- **04.02.20** at 15:00:
  Xiao et al., Generating Adversarial Examples with Adversarial Networks, IJCAI 2018