ML INFOSEC

Project outline

Topics

- Basic Machine Learning techniques
- Coding (Python, NumPy, Scikit-Learn etc.)
- Use of open source and freely available ML tools
- Deep learning = ML with Deep Neural Networks
- ML-based malware classification and analysis

Tools

- Github
- Anaconda
- Jupyter notebooks
- CoCalc
- Python
- NumPy, Scikit-Learn
- Weka
- PyTorch, TensorFlow

Resources: ML

- Aurélien Géron: Hands-On Machine Learning with Scikit-Learn & Tensorflow, O'Reilly, 2017

- Tom M. Mitchell: Machine Learning, McGraw Hill, 1997

- Udacity MOOC: Deep Learning with PyTorch

- FutureLearn MOOC: Data Mining with Weka (U Waikato, NZ)

Resources: ML in INFOSEC

- Mark Stamp: Introduction to Machine Learning with Applications in Information Security, Chapman and Hall/CRC, 2017 (http://www.cs.sjsu.edu/~stamp/)

- Konrad Rieck (TU Braunschweig, DE): Machine Learning for Computer Security, 2018 (lecture notes, unpublished)

Resources: Deep Learning

- Michael A. Nielsen: "Neural Networks and Deep Learning", Determination Press, 2015. http://neuralnetworksanddeeplearning.com/

- Ian Goodfellow, Yoshua Bengio, Aaron Courville: Deep Learning, MIT Press, 2016. https://www.deeplearningbook.org/

Examples

Malware detection	SVM, PCA, HMM, PHMM
Masquerade detection	PHMM
Malware classification	K-Means, Naive Bayes
Malware analysis	EM, K-Means, DNN
Image spam detection	SVM, PCA
Intrusion detection	SVM, k-NN, k-Means

Feature engineering for malware

- N-grams on hexdumps/assembly files
- Opcodes
- API calls

Disassembling

- GNU Binary Utilities (e.g. objdump)
- radare2
- IDA Pro
- Capstone

First module

- Machine Learning landscape
- Elementary probability theory
- Bayes's theorem
- Naive Bayes
- Implementation of Naive Bayes
- Practical examples