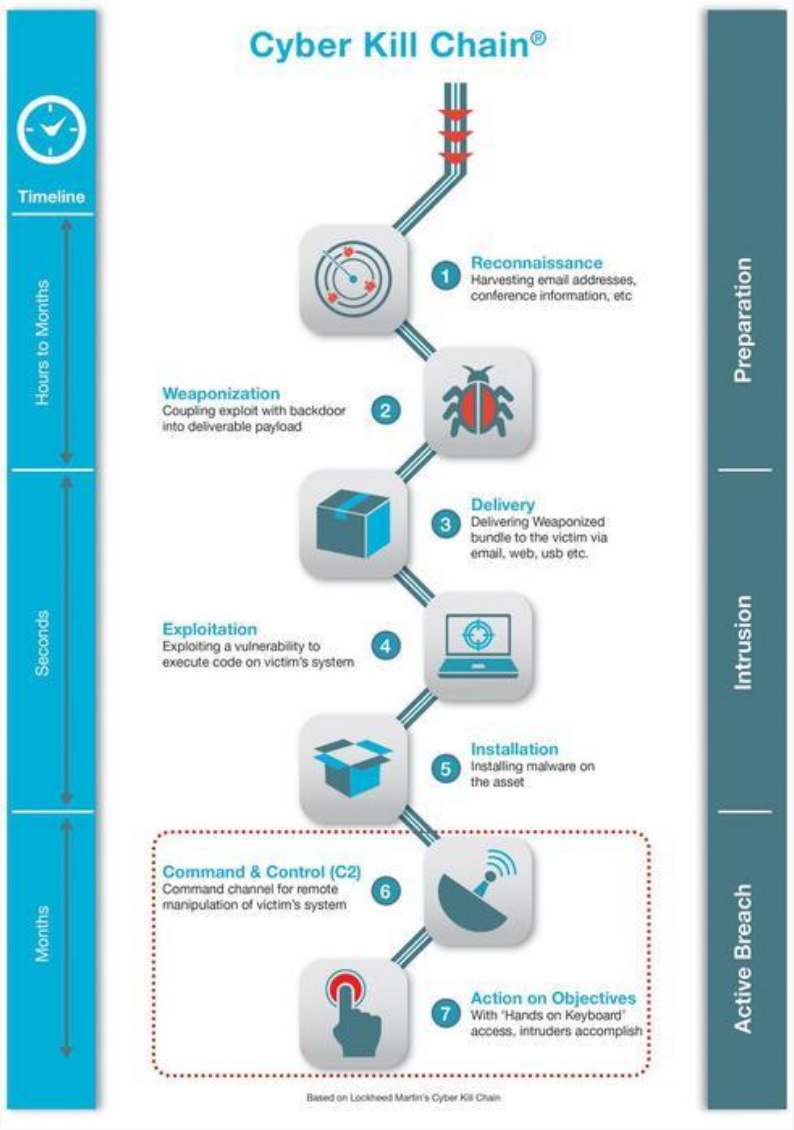
AI Considerations for an Automated Cyber Security Strategy

Ron Winward

Security Evangelist

May 2018

Cyber Kill Chain® by Lockheed Martin



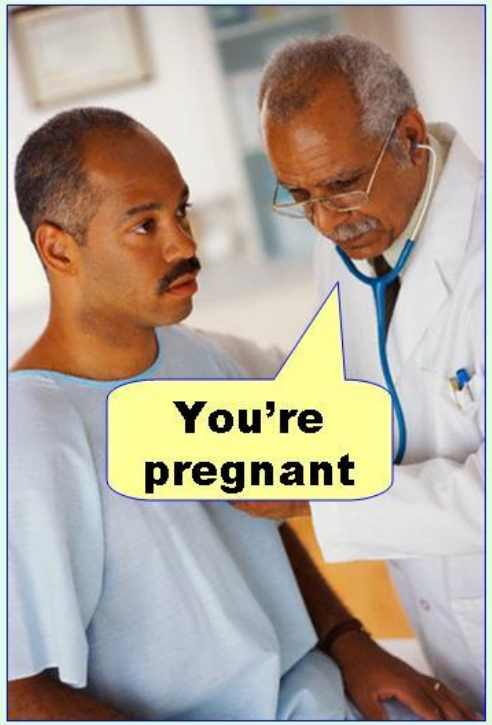
- Targeted attacks
- Plenty of opportunities to detect and block attacks before they cause actual damage

- So why organizations still getting breached and only find out (long) after the fact; by accident or through ransom ?
- Two reasons mainly:
  - Not enough events/visibility
  - Too many events

Source of Image: Dark Reading - Deconstructing The Cyber Kill Chain - Giora Engel

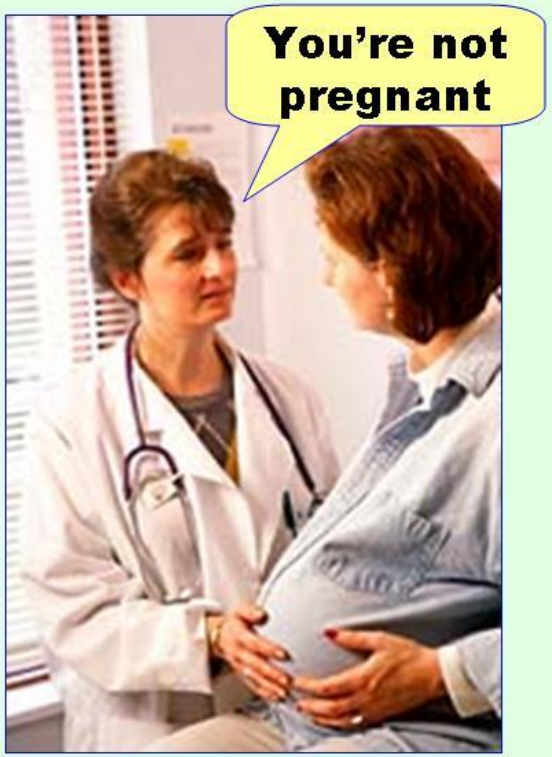
Minimizing False Positives & False Negatives

Type I error
(false positive)



Too many events

Type II error
(false negative)



Not enough events

Why minimize

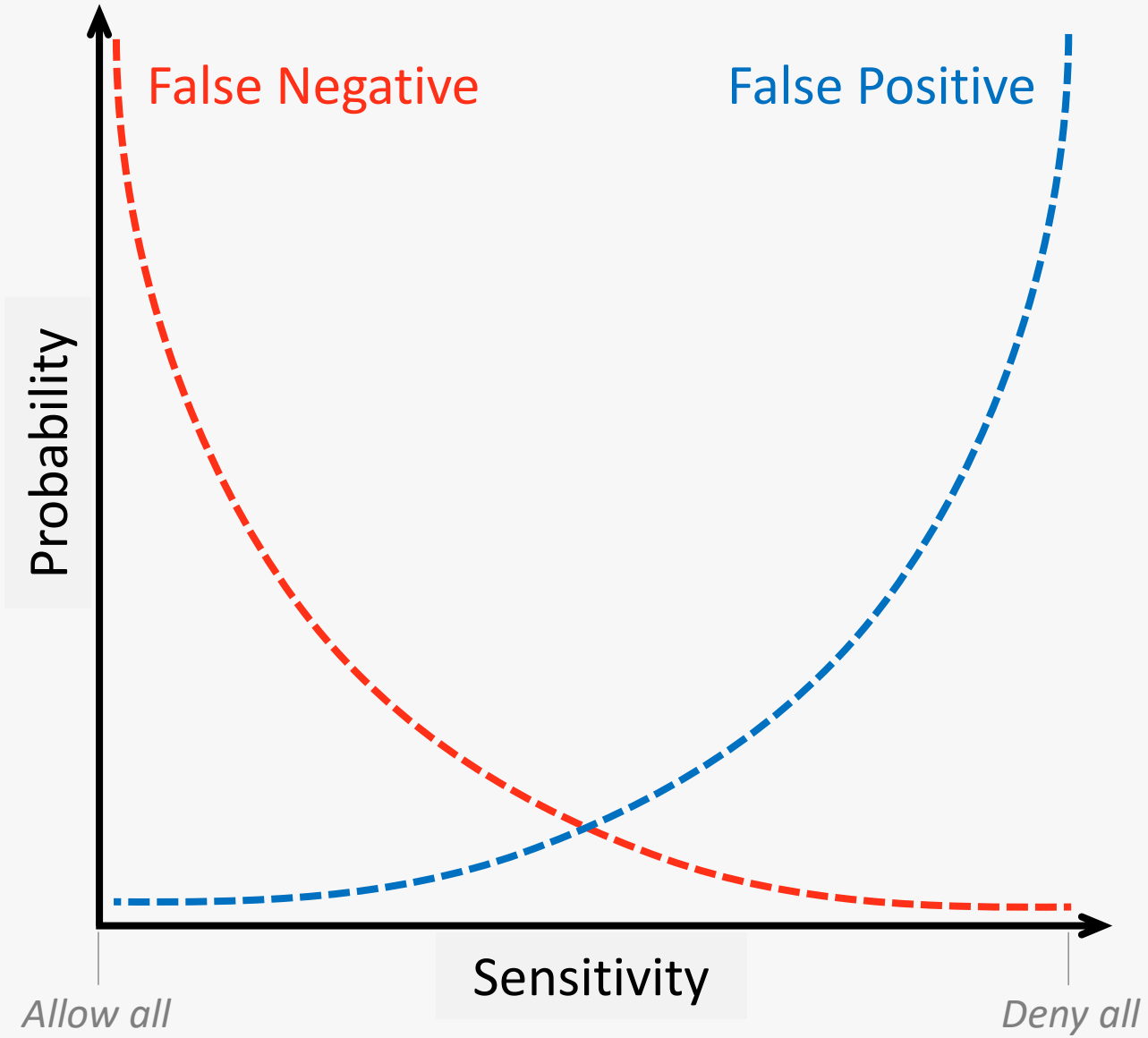
False Negatives?
Seriously?

False Positives?
How many incidents can your SOC investigate?

Can they give the right incidents the amount of time they deserve?

Source of Image: Effect Size FAQs by Paul Ellis

Detection Sensitivity in Positive Security Models



Anomaly Detection – Game On!

- Security threats growing faster than security teams and budgets, huge talent shortage
- Paradox: Proliferation of data from dozens of security products makes it harder to detect and investigate threats

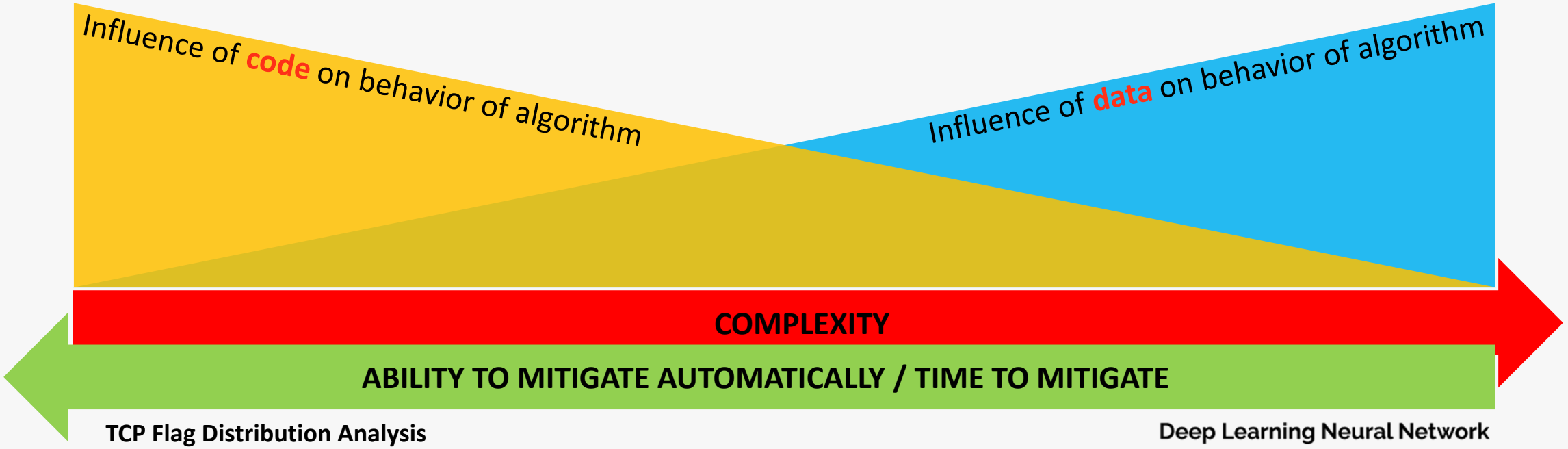
- Need for automation
- Rule based event correlation provides reduction from millions to thousands
- A good SOC can investigate maybe a couple of 100 incidents a day
- How to leverage previous work from the SOC to improve the future detection by automation?
- Need for automation that improves itself over time based on new data and user or researcher feedback

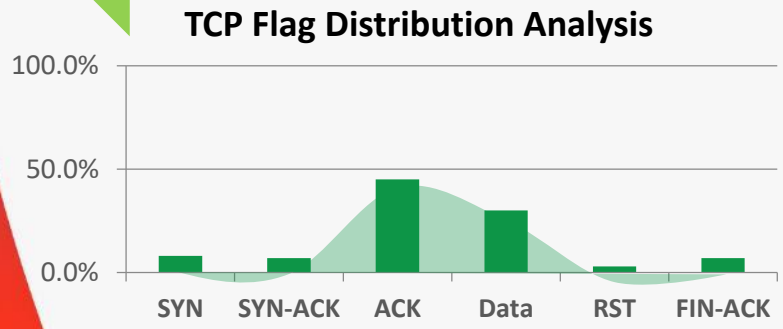
Machine Learning Challenges

Detection Algorithms & Machine Learning

Deterministic
Model Based
Transparent

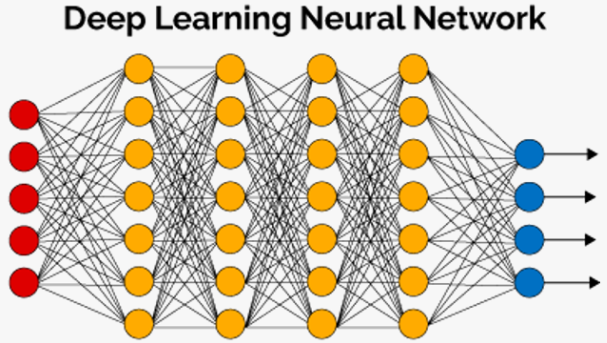
Too complex to code
Generalization
Opaque





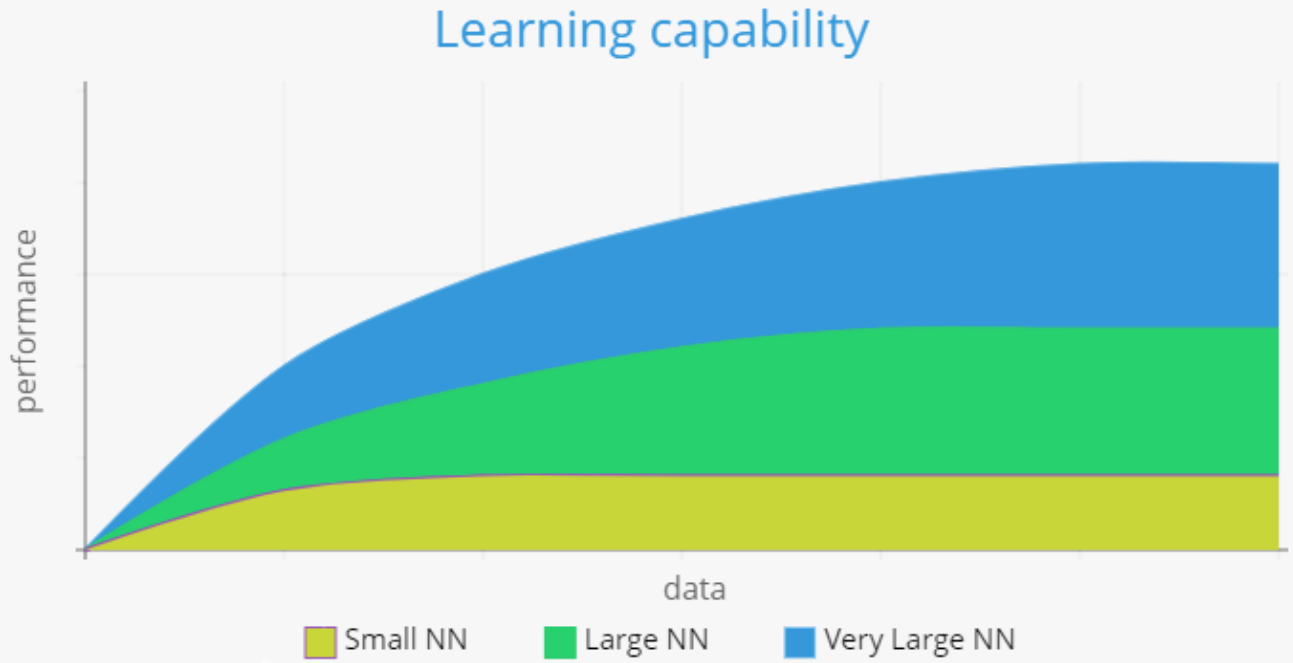
RFC based anomaly detection

K-means Clustering
Logistic Regression
Bayesian Linear Regression
Support Vector Machine
Principal Component Analysis



Neural Networks Need Data! Good Data and Lots of it…

- Larger networks have higher learning capability (memory)
- Performance is only as good as the amount of data put in
- Need extra data to evaluate the network's performance
- Quality of the network will on be as good as the quality of the data put in
- Synthetic data generation can be misleading, correlation between data points



Examples of Training Corpus Sizes:

Speech Recognition:
100,000 hrs of audio
~ 11 years of sound

Face recognition:
200 million images

Source: Andrew Ng, Baidu

Generalization: Size Matters

Bigger not always better!

Too small neural network

=

Underfitting

Too large neural network

=

Overfitting

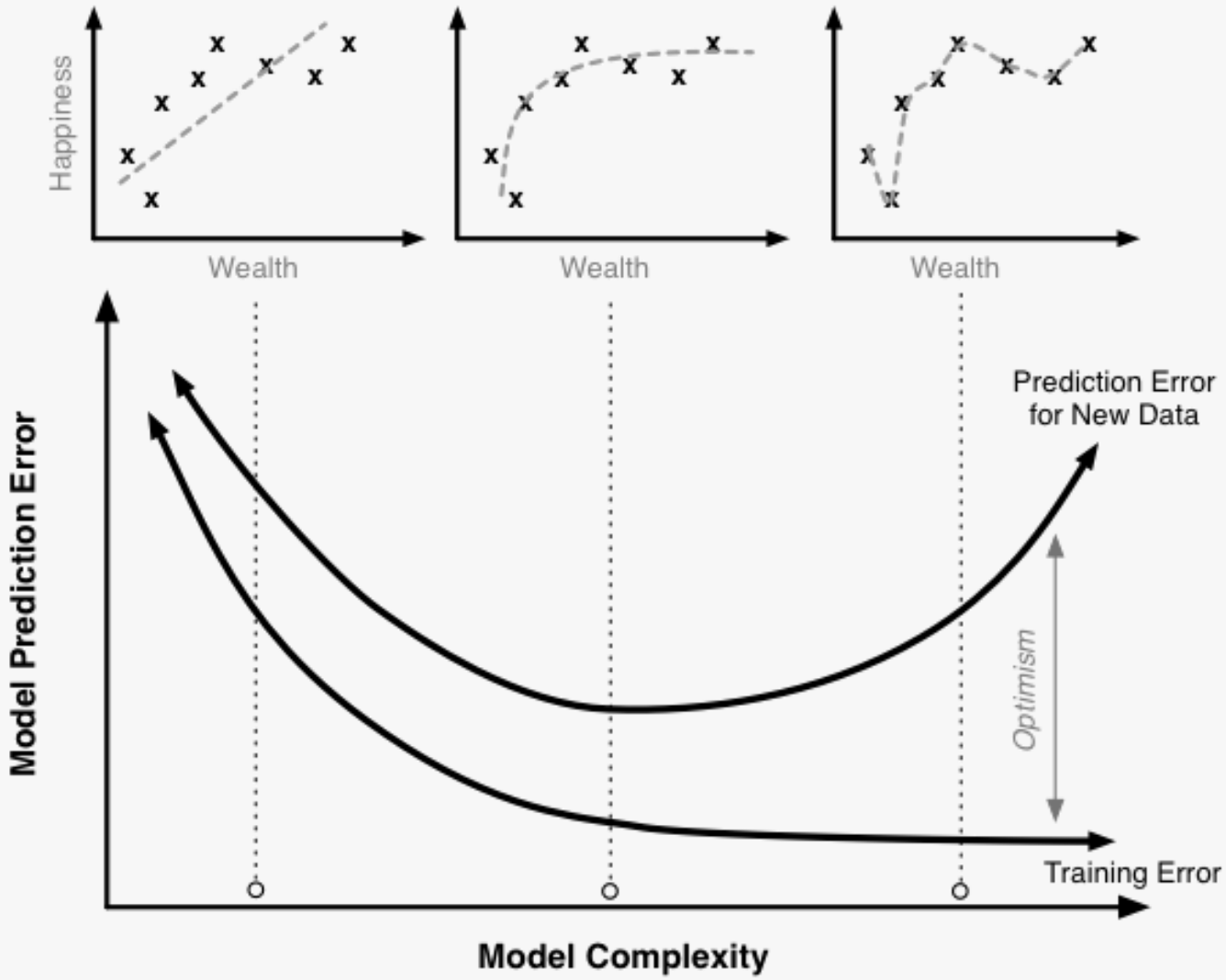


Image source: Quora

Changing environments

- Deep Learning systems are not good at handling changing and dynamic environments
- As networks grow, the Deep Learning system might need to be resized to prevent under-fitting
- As protocols change and device types are added/removed to the environments, models need to be re-trained to be effective
- Honeypots provide a good source of information, adds data labeled as “bad” or “attack” traffic – every source of data is useful for increasing the prediction quality and probability of correct output.

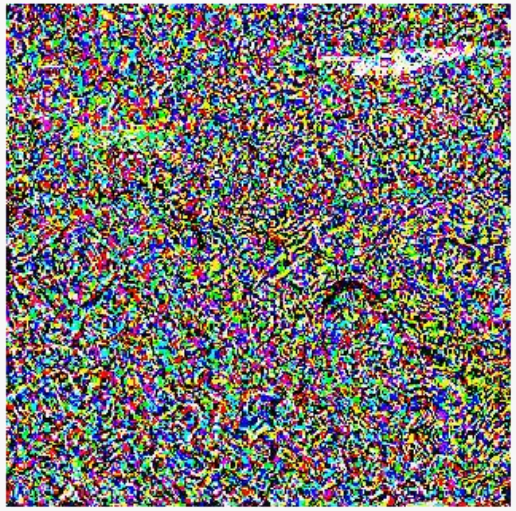
Attacking Deep Learning Systems

- Poisoning Attacks
  - Poisoned data to trick the learning system
- Evasion Attacks
  - Perturbations in the input data
- Studies in Adversarial Machine Learning
  - Learning in the presence of adversaries
  - Particularly important for security applications of Machine Learning

Adversarial Attack Examples

Original image: sports car

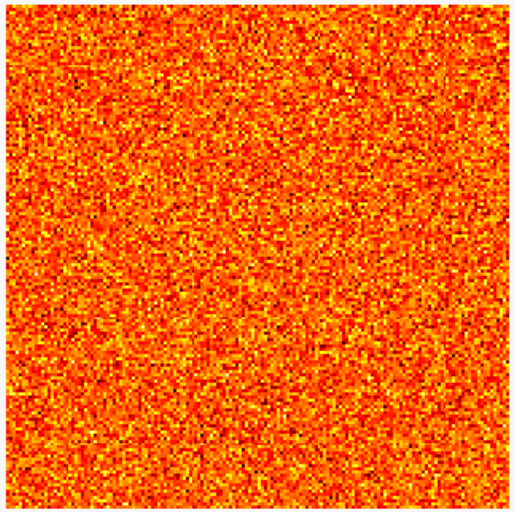
Attacking noise



Adversarial example: toaster

Sylvester Stallone

Adversarial noise



Keanu Reeves

Source: http://blog.ycombinator.com/how-adversarial-attacks-work/

Adversarial Attack Examples

Source: http://spectrum.ieee.org/cars-that-think/transportation/sensors/slight-street-sign-modifications-can-fool-machine-learning-algorithms

Weaponizing Machine Learning

Image: DARPA Cyber Grand Challenge

Ways hackers will use machine learning

- Increasingly Evasive Malware
  - Using a Generative Adversarial Network (GAN) algorithm
  - MalGAN [Feb 2017] generates adversarial malware samples
- Hivenets* and Swarmbots*
  - Smarter botnets using self-learning 'hivenets' and 'swarmbots'
  - BrickerBot: Autonomous PDOS botnet [Radware 2017]
- Advanced Spear Phishing at Scale
  - Using Natural Language Processing (NLP) algorithms for better social engineering
  - Training on genuine emails, scraping social networks/forums, stolen records…
- Raising the Noise Floor
  - Poisoning the model by flooding it with false positives, causing recalibration of the model

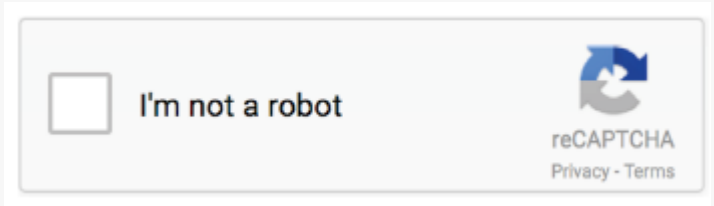
(*) Fortinet Predicts Highly Destructive and Self-learning "Swarm" Cyberattacks in 2018

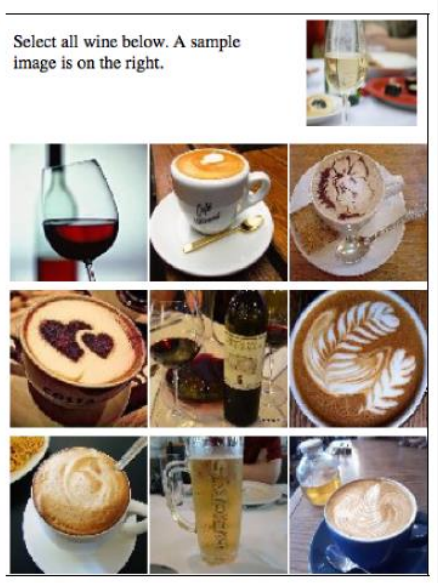
Breaking CAPTCHA

- 2012: Support Vector Machines (SVM) to break reCAPTCHA
  - 82% accuracy – Cruz, Uceda, Reyes

- 2016: Breaking simple-captcha using Deep Learning
  - 92% accuracy – How to break a captcha system using Torch

- 2016: I'm not Human - breaking the Google reCAPTCHA
  - 98% accuracy
  - Black Hat ASIA 2016 – Sivakorn, Polakis, Keromutis





SNAP_R – spear-phishing on Twitter

- SNAP_R sent simulated spear-phishing tweets to over 800 users at a rate of 6.75 tweets per minute, luring 275 victims
- Forbes staff writer Thomas Fox-Brewster, who participated in the experiment, was only able to pump out 1.075 tweets a minute, making just 129 attempts and luring in just 49 users

https://www.blackhat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-Social-Engineering-Automated-E2E-Spear-Phishing-On-Twitter.pdf

DeepHack – DEF CON 25

- Open-source hacking AI
- This bot learns how to break into web applications using a neural network, trial-and-error, and a frightening disregard for humankind.
- Github: https://github.com/BishopFox/deephack
- DeepHack can ruin your day without any prior knowledge of apps, databases – or really anything else. Using just one algorithm, it learns how to exploit multiple kinds of vulnerabilities, opening the door for a host of hacking artificial intelligence systems in the future.
- This is only the beginning of the end, though. AI-based hacking tools are emerging as a class of technology that pentesters have yet to fully explore. We guarantee that you'll be either writing machine learning hacking tools next year, or desperately attempting to defend against them.

Video: DEF CON 25 (2017) - Weaponizing Machine Learning - Petro, Morris - Stream - 30July2017

Concluding

Deep Learning Applicability

Learning in the presence of adversaries

"Good" Data >>> "Bad" Data

Applicability today: crowd sourcing or global community

Cloud applications and Threat Intelligence


radware
Thank You