Cyberattacks, machine learning, and data science

Michael Slawinski, Ph.D.
Staff Data Scientist

Agenda

Introduction to Cylance

Cyberattacks

Machine Learning applied to Cybersecurity

Our Data

Our Models

Challenges in Cybersecurity

Challenges in Data Science

Open Problems

Cylance – an overview

We are a machine learning based cybersecurity firm. We build machine learning models which prevent cyberattacks.

Founded in 2012 by Stuart McClure and Ryan Permeh

First cybersecurity firm to successfully apply machine learning to prevent file-based cyberattacks

The Data Science Teams

- Data Science Research
  - 85% Ph.D.’s
  - Backgrounds include chemistry, physics, mathematics, stats, computer science
- Data Science Engineering
  - Build and maintain infrastructure to support DSR
- Data Science Operations
  - Maintain current production models

CYLANCE®

Cyberattacks

What is a cyberattack?

An attempt by hackers to damage or destroy a computer network or system

Example: Ransomware is type of malware that prevents users from accessing their system or personal files and allows a malicious actor to demand ransom payment in order to regain access.

1. Distribution Campaign – attackers use social engineering to trick users to download a dropper
2. Malicious code infection – dropper downloads an executable which installs the ransomware
3. Malicious payload staging – ransomware establishes persistency to exist beyond reboot
4. Scanning – ransomware searches locally and on the network for files to encrypt
5. Encryption – discovered files encrypted
6. Payday – user is given instructions on how to pay the ransom

Example – ‘WannaCry’

7:44 UTC Friday May 12, 2017: Attack Begins.

Single machine running non-updated windows is infected with a ransomware worm (moves from machine to machine within a network) which searches and encrypts elements of the filesystem.

Affected machines present user with this screen demanding a bitcoin ransom under threat of file deletion.

230,000 computers in 150 countries affected within a single day

The Shadow Brokers leveraged an exploit thought to have been developed by the NSA, called Eternal Blue

Eternal blue: exploit of Server Message Block protocol

Damage: 100s of millions to billions USD

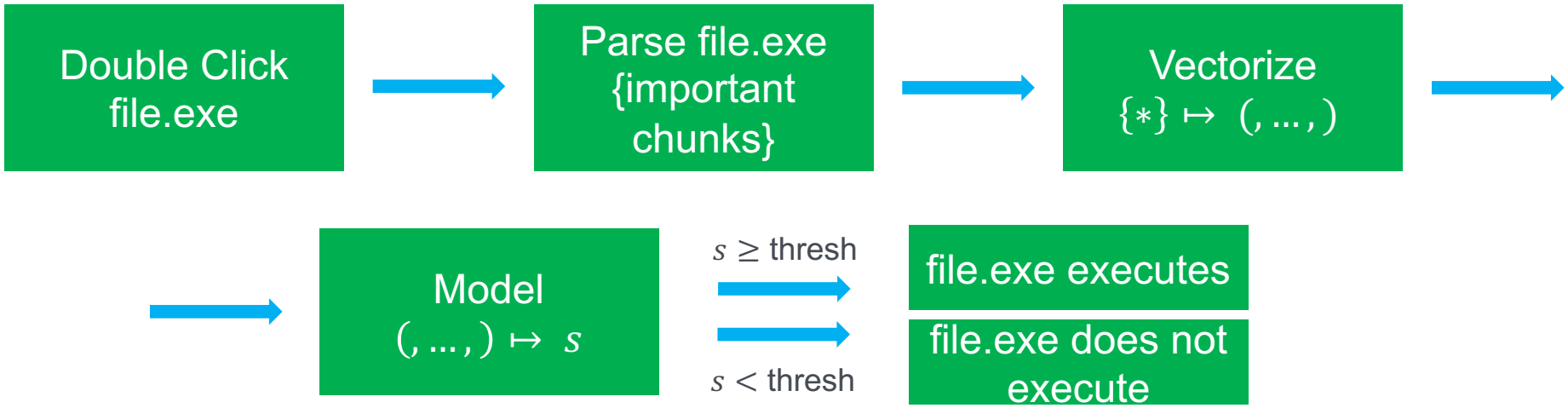
The defeat of ‘WannaCry’

Microsoft released security updates to patch the SMB exploit

Marcus Hutchins discovered the kill switch domain hardcoded in the malware

- The malware would look for a certain domain and encrypt the machine’s files only if it could not hit that domain.
- This slowed the spread and the attacks ceased after a brief cat and mouse battle between the attackers and responders.

What happened to machines running Cylance PROTECT™ ? Nothing.



Traditional Solution - signatures

Look for specific bytes or segments of code commonly associated with a given malicious file

Antivirus software discovers such a segment and prevents the file from running.

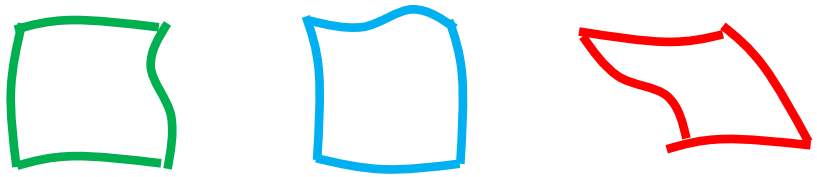
In machine learning parlance, this is a case of severe overfitting.

```
<?php
        @set_time_limit(9999);
        @ignore_user_abort(1);
        @ini_set('allow_url_fopen',1);
        @error_reporting(false);
        @ini_set('display_errors', false);
        @ini_set('error_reporting',  false);
        @ini_set('safe_mode', false);


        $apiKey = '98a91e769be784204e9568444*****';
        $campaignId = 'BMZ***';

        $requestTimeout = 5;
        $useCurl = 0;
        $new_request = new HttpRequest($useCurl, $requestTimeout);
```

Utility of Learning Algorithms: Why are learning algorithms necessary?

Extreme data variability resulting in a possibly infinite signature set

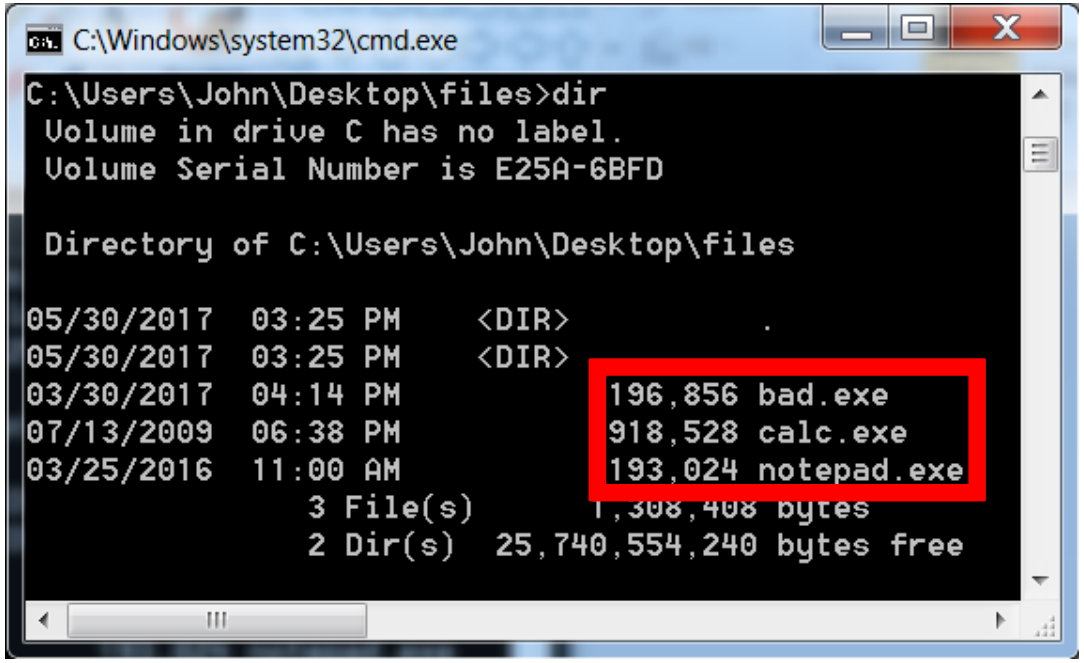
Example: Hand-drawn squares

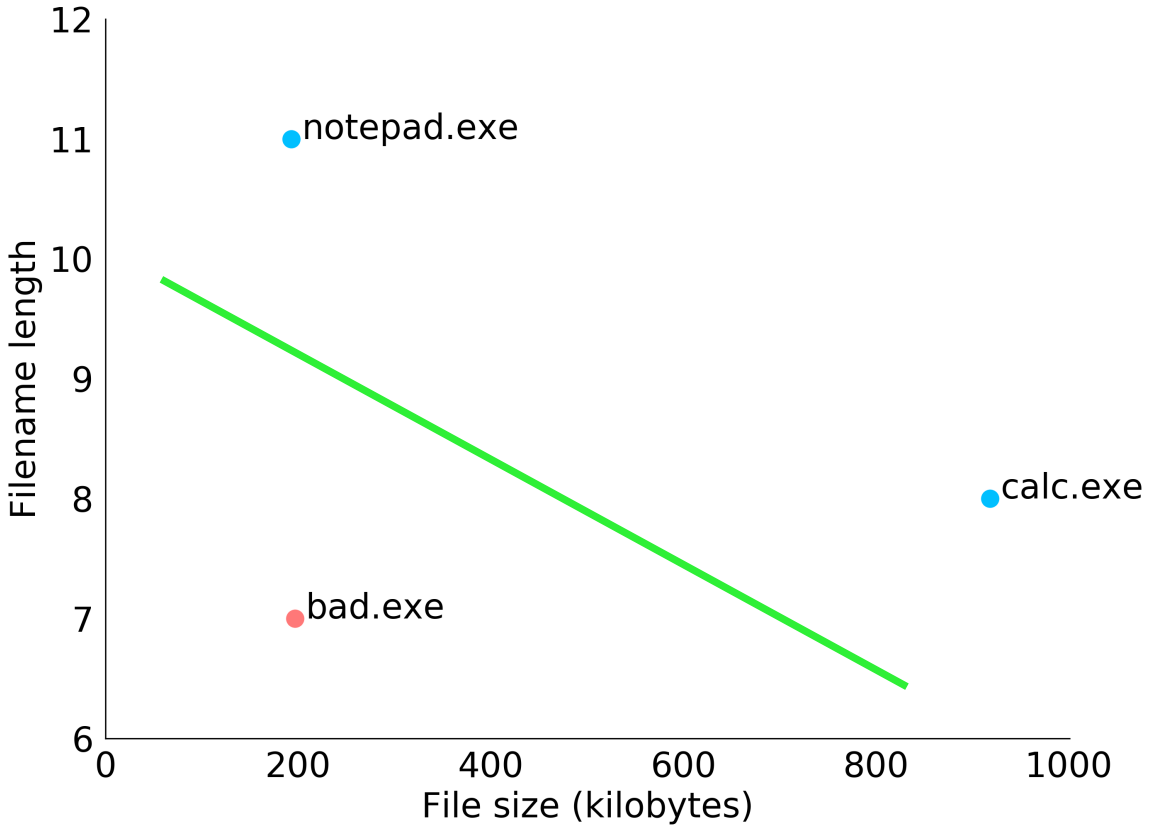


CYLANCE®

What does it take to prevent Cyberattacks?

Correct Solution – machine learning





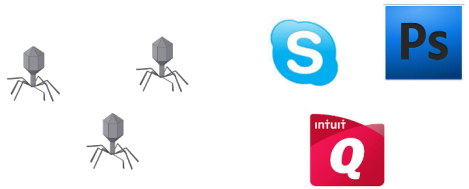
File	Filename length	File size(kB)
calc.exe	8	918.528
notepad.exe	11	193.024
malware.exe	11	193.024

CYLANCE

BlackBerry | CYLANCE

Building a Model

1. Gather Data



2. Vectorize (Choose a set of features)

$x_1, x_2, \ldots, x_N$

3. Choose a Learning Algorithm (a function f_W whose behavior depends on parameters W)

Classify( ) via $y = f_W(x_1, x_2, \ldots, x_N)$

4. Define a loss function $J(D, W)$ (measures extent to which classifier misclassifies your data)

$J(D, W)$ is high if f_W misclassifies many samples

5. Train the classifier (Solve an optimization problem: adjust W's to minimize J)

Adjust W in order to minimize $J(D, W)$ resulting in f_W few misclassifications

CYLANCE®

Our Data

File Types

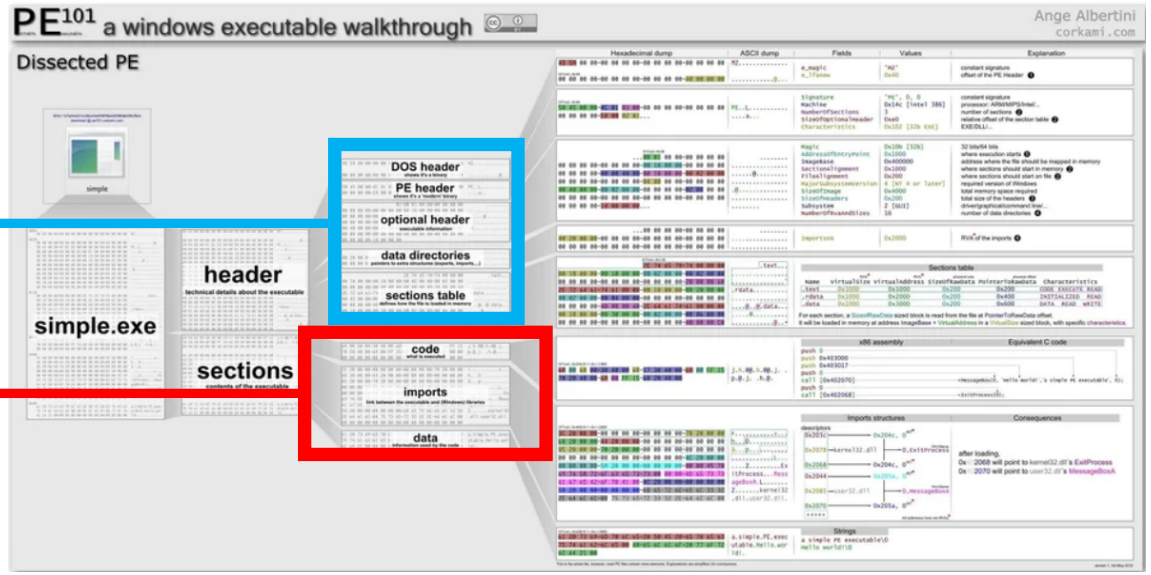
Header

dos header,PE header
sections table, data dir

Sections

code, imports, data

Portable Executable



Script

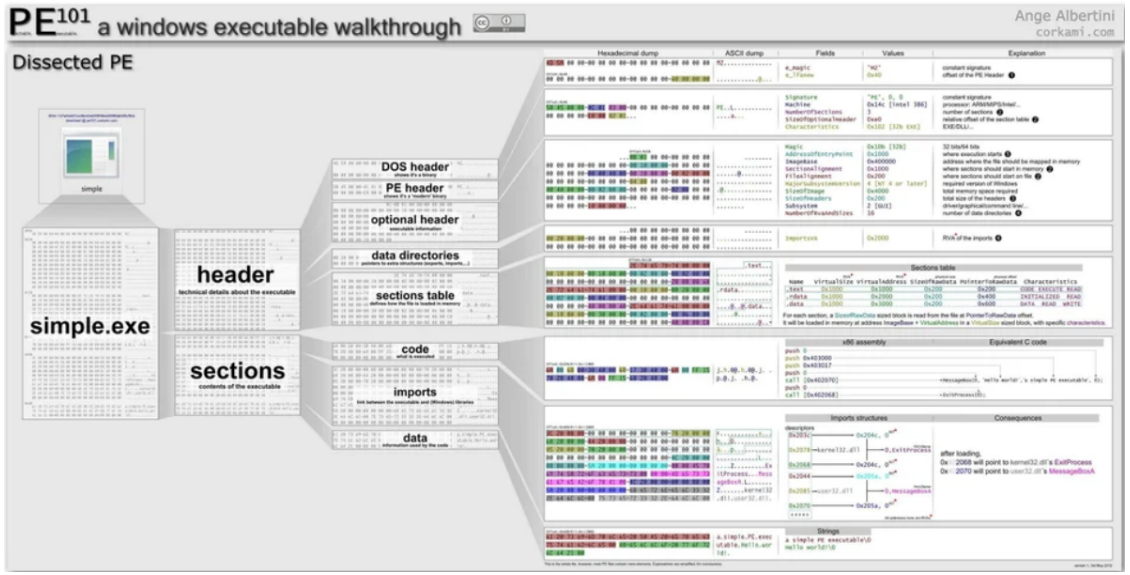
Classes

Functions

Other language-specific constructs

```
<?php

/* @WebProfiler/Profi er/base_js.html.twig */
class __TwigTemplate_ d6e518366d3824ef773afad89303bcb extends Twig_Template
{
    public function __construct(Twig_Environment $env)
    {
        parent::__construct($env);

        $this->parent = false;

        $this->blocks = array(
        );
    }
```

Vectorization




```
<?php

/* @WebProfiler/Profiler/base_js.html.twig */
class __TwigTemplate_0d6e518366d3824ef773afad89303bcb extends Twig_Template
{
    public function __construct(Twig_Environment $env)
    {
        parent::__construct($env);

        $this->parent = false;

        $this->blocks = array(
        );
    }
```

Expert, Statistical Features

Disassembly

Function Call Graphs

{ASTs}

Word Vectors

Bag of Words

AST

CYLANCE®

Our Models

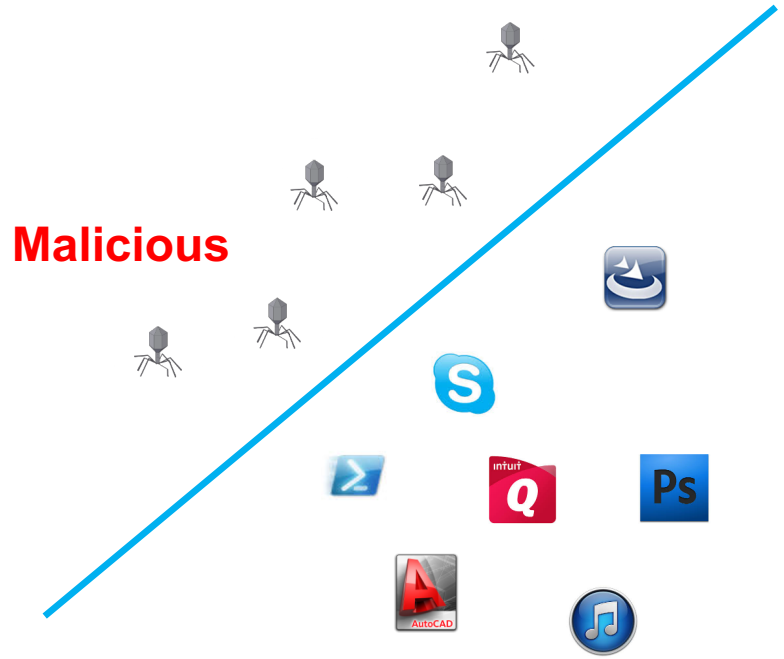
Learning Algorithms Used In-House

Supervised

Standard Neural Networks

LSTM/Transformer

Ensemble Methods

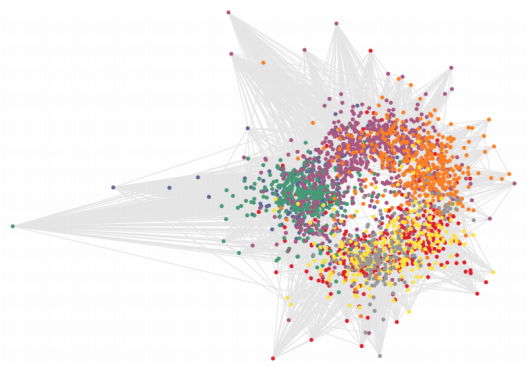


Unsupervised

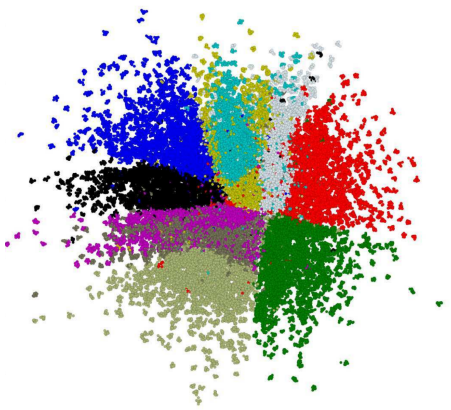
Clustering

VAE – standard and flow-based

Lite weight probabilistic models



Control flow embedding



Static embedding

CYLANCE®

Challenges and Solutions: Cybersecurity

Challenge: Vulnerability to Adversarial Attacks – ‘Snow Features’

Training Set

Model Says: ‘Wolf’ 99.9%

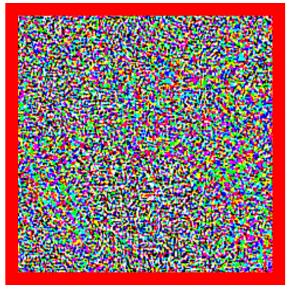
Upshot: Model overemphasizing white, i.e., snow pixels due to high wolf/snow co occurence

Adversarial Machine Learning

$\boldsymbol{x}$

"panda"
57.7% confidence

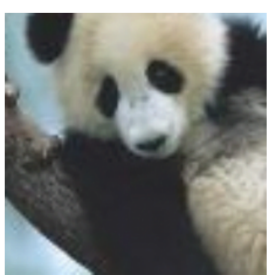
$+ .007 \times$



$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

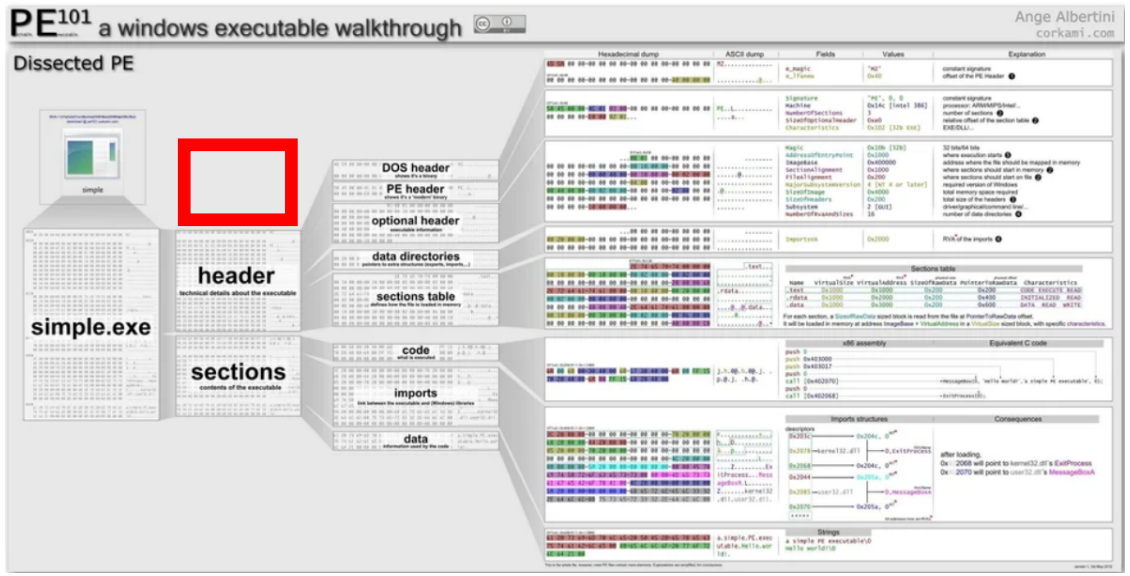
"nematode"
8.2% confidence

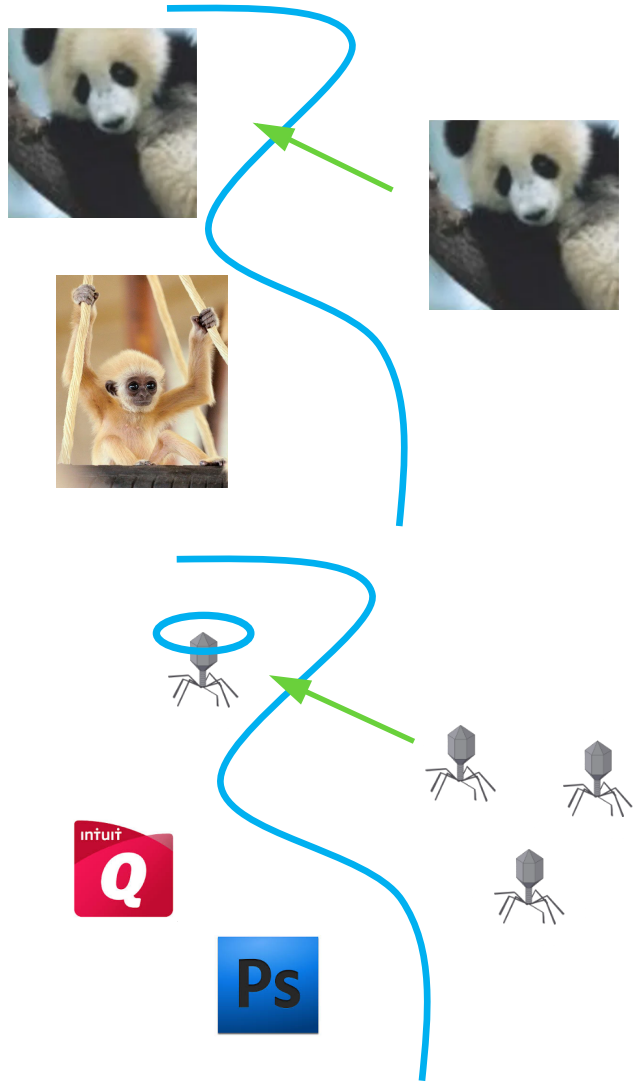
$=$



$\boldsymbol{x} + \epsilon\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"gibbon"
99.3 % confidence





BlackBerry | CYLANCE

Vulnerability: human/model paradigm shift

We make the mistake of assuming the model is judging as we judge.

In other words, we assume the machine learning model has a baked-in conceptual understanding of the objects being classified…

Example: Lie Detectors – what is a lie?

Human Point of View	Lie Detector Point of View
Lie is a statement believed to be false but offered as true	Heart rate above a threshold Perspiration above a threshold Body movement above a threshold

Adversarial Attacks – Defenses

1. **Reduce leveraging of ‘snow features’**
    - Features which when perturbed, change the nature of the original file

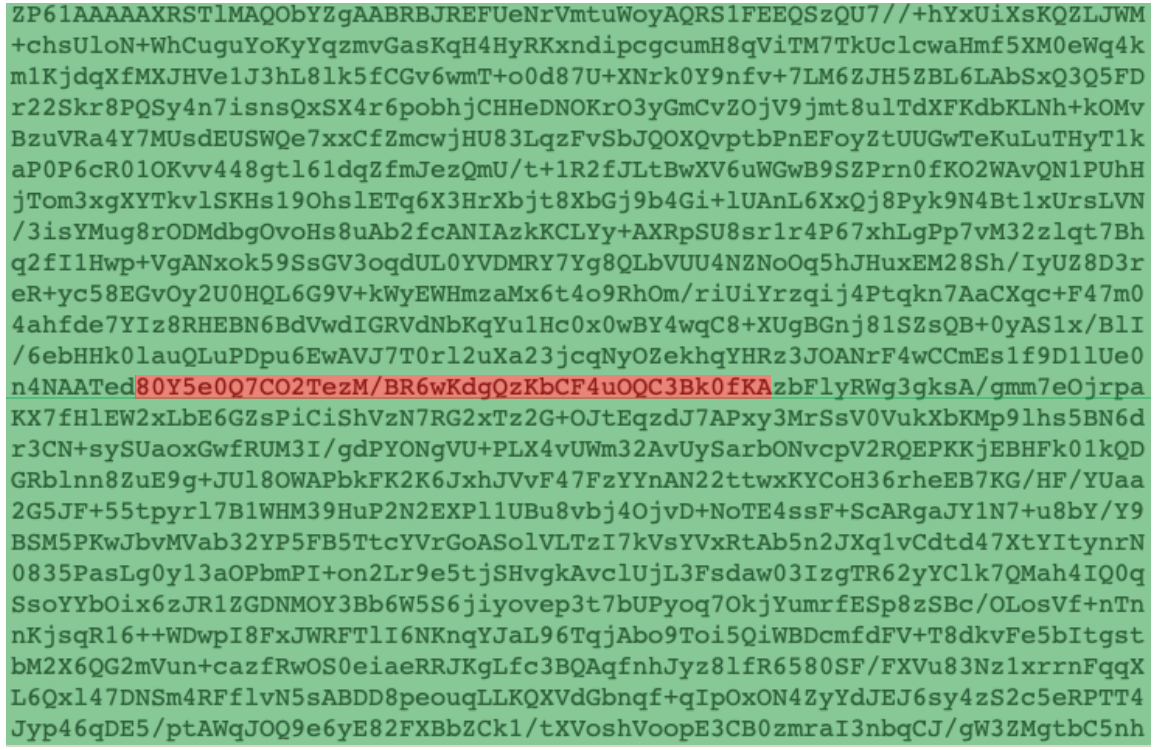
2. **Recognize what is meant by ‘model’, given context**
    - Effective classification does not imply a deep understanding of the objects being classified

3. **Anomaly Detection**
    - Learn probability distribution of the training set so that files altered to trick the classifier will be assigned very low probability

4. **Inclusion of Adversarial Examples in Training**
    - Construct adversarial examples at scale and add them to the training set to familiarize the model with such samples

Challenge: Needle in a Haystack

eval()



Sentiment Analysis on tweets

Loves the German bakeries in Sydney. Together with my imported honey it feels like home	Positive
@VivaLaLauren Mine is broken too! I miss my sidekick	Negative
Finished fixing my twitter...I had to unfollow and follow everyone again	Negative
@DinahLady I too, liked the movie! I want to buy the DVD when it comes out	Positive
@frugaldougal So sad to hear about @OscarTheCat	Negative
@Mofette briliant! May the fourth be with you #starwarsday #starwars	Positive
Good morning thespians a bright and sunny day in UK, Spring at last	Positive
@DowneyisDOWNEY Me neither! My laptop's new, has dvd burning/ripping software but I just can't copy the files somehow!	Negative

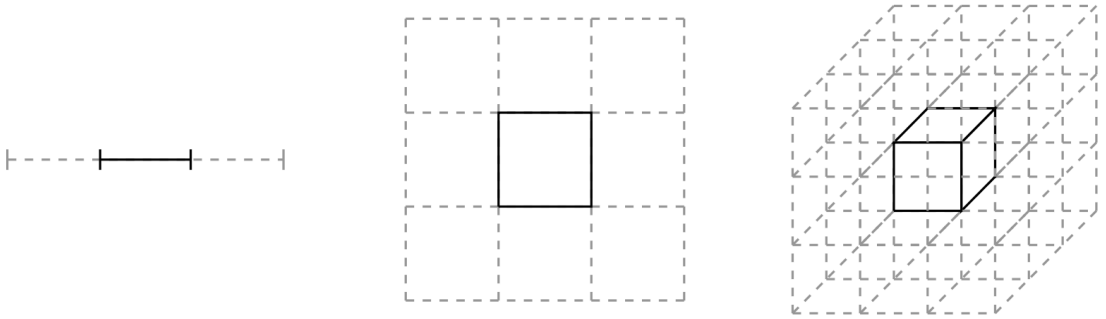
What do we mean by ‘malicious’?

- Encryption may be benign or malicious
- Obfuscation may be benign or malicious
- Packing may be benign or malicious

Solution: Allow the algorithm to *infer* what `maliciousness’ means

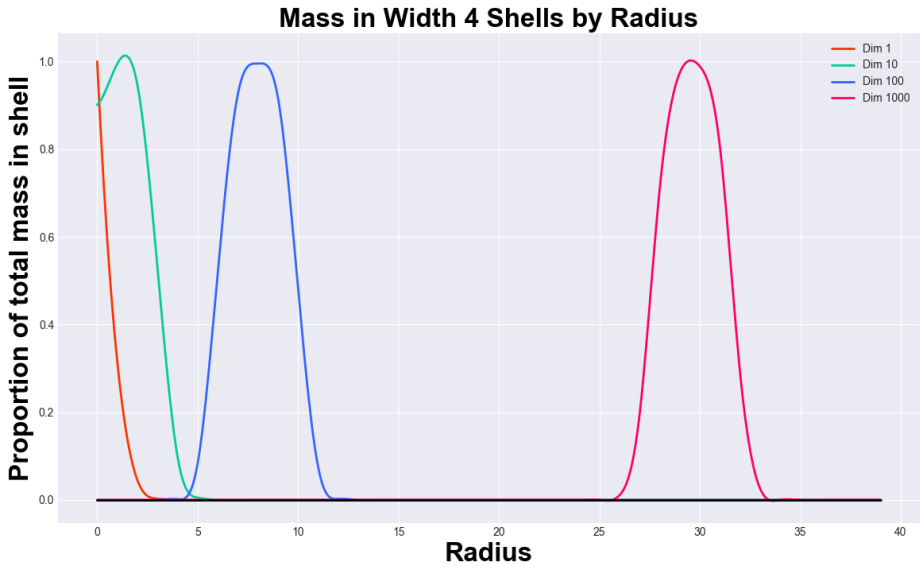
Challenge: High Dimensional Feature Space

Distribution of volume as a function of dimension:



$3^D - 1$ neighboring partitions in D – dimensional space

Singularity of probability mass in high dimensions



$d = 1$: $99.73\%\ of\ mass\ in\ S(0.1,3)$

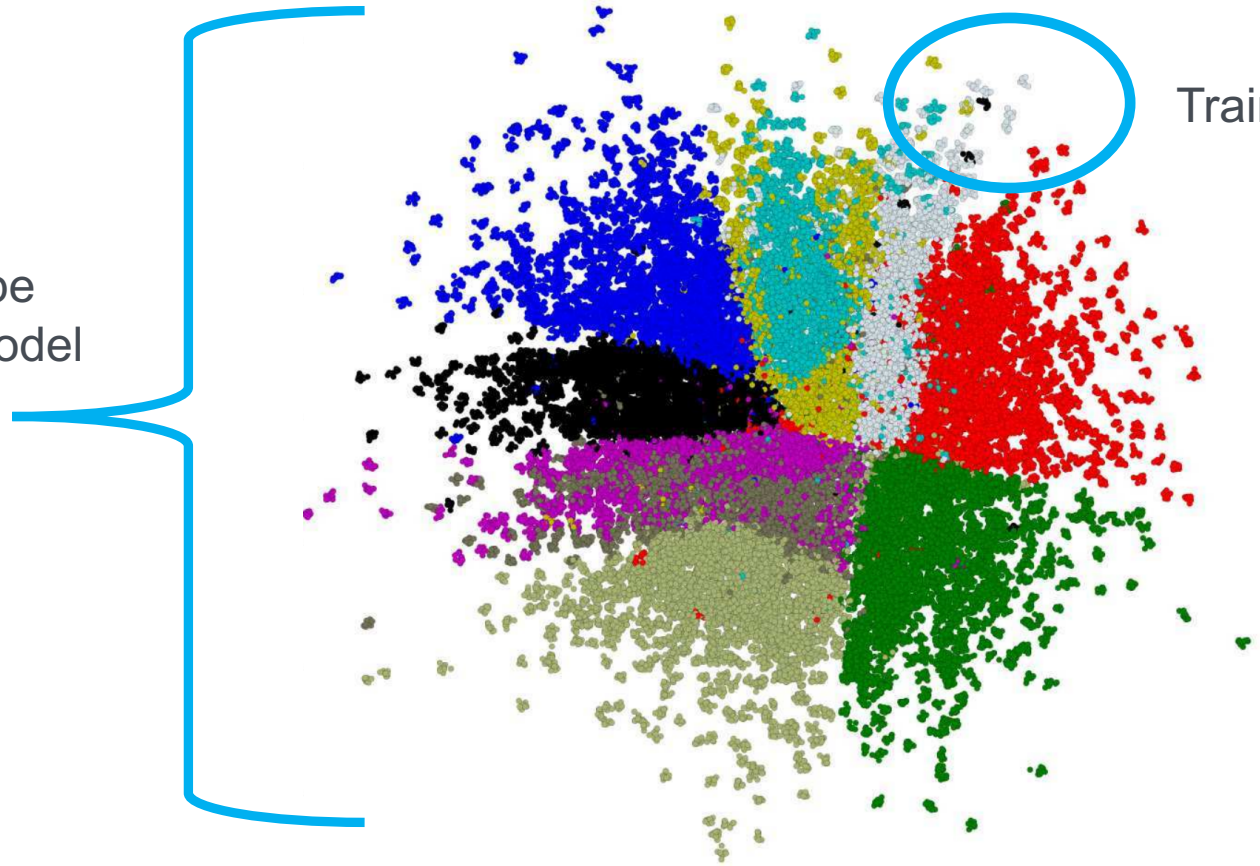
$d = 10$: $99.44\%\ of\ mass\ in\ S(1,5)$

$d = 100$: $99.53\%\ of\ mass\ in\ S(8,12)$

$d = 1000$: $99.00\%\ of\ mass\ in\ S(30,34)$

Challenge: Distribution of overall dataset vs data distribution to which we have access

Universe of possible samples that could be encountered by a model in production



Training Data

CYLANCE®

Challenges and Solutions: Professional Data Science

Challenge: Practical Constraints

Infrastructure Costs

AWS Costs – r4.8xlarge (32 cores, 244Gb RAM) $2.128/hr,
– 1TB EBS = $100/month

Model porting

python to C#, for example

Product Constraints

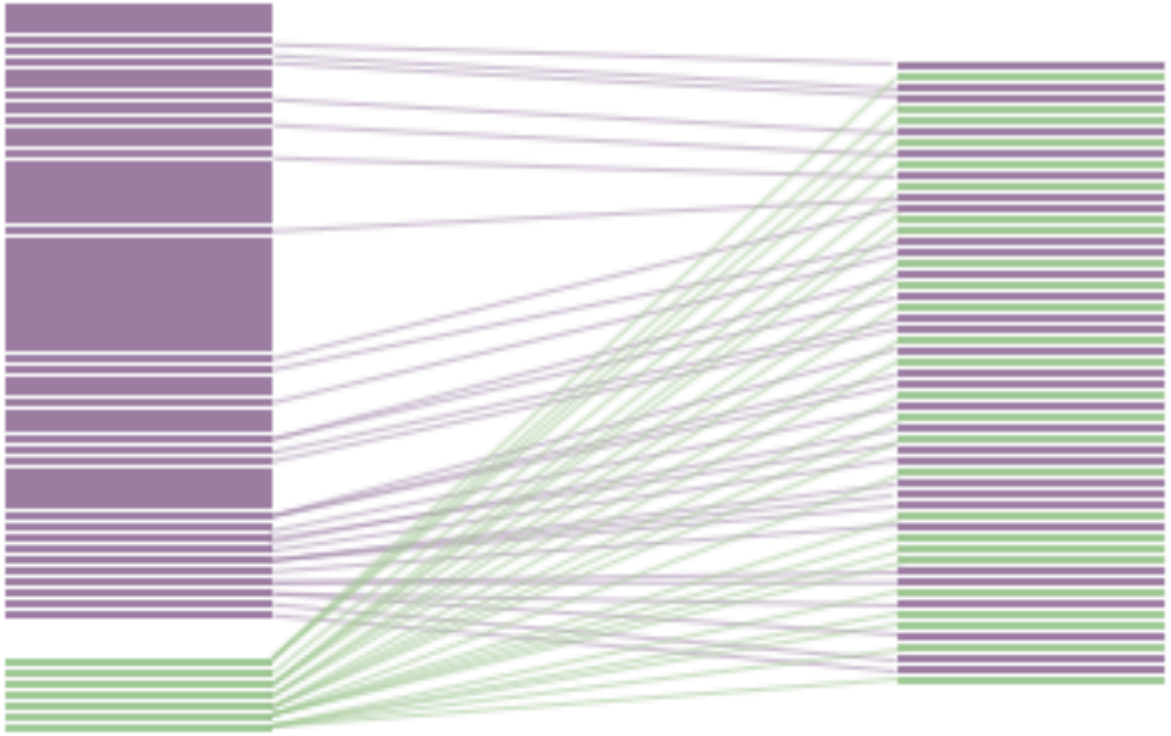
cpu/memory/FP/FN

Challenge: Class Imbalance

‘The model achieves 99.9% accuracy on test data’

This statement is essentially meaningless absent more information.

Weighted sampling



Class-weighted Loss

$$\text{loss}(x, class) = weight[class]\left(-x[class] + \log\left(\sum_j \exp(x[j])\right)\right)$$

Challenge: The `Prototype' Dilemma

Problem: A barely functional, not ready for prime time, model gets put into the production pipeline. Pain ensues.

How does this happen?

1. Sales and/or management will conceive of a product resulting from having solicited needs from customers.
2. Management will cobble together some small dataset
3. Data Science is charged with building a prototype with this inadequate dataset
4. Data Science objects after recognizing the futility of building a production-quality model on the given dataset
5. Management responds: 'No no, we understand. It's fine – just build a prototype which can be used in a demo to the board. We can always iterate later.'
6. Data Science reluctantly agrees because 'hey, we can always iterate…'.
7. After a successful demo, Engineering is tasked with productionalizing the prototype and Data Science is charged with adding features to the prototype.
8. Data Science, now constrained by decisions made by Marketing and Engineering, is relegated to the impossible task of meeting minimum product standards without adequate data.

CYLANCE®

Open Problems

Open Problems

Halting Problem

Learn the relationship between static code and execution behavior

Sandbox detonation

Develop better sandboxes, so that malicious files can be observed while executing

Manifold structure of PEs and other filetypes

Learn better latent space representations of files/control flow

CYLANCE®

Questions?