RSAConference2016

Abu Dhabi | 15–16 November | Emirates Palace

SESSION ID: SOP-W04

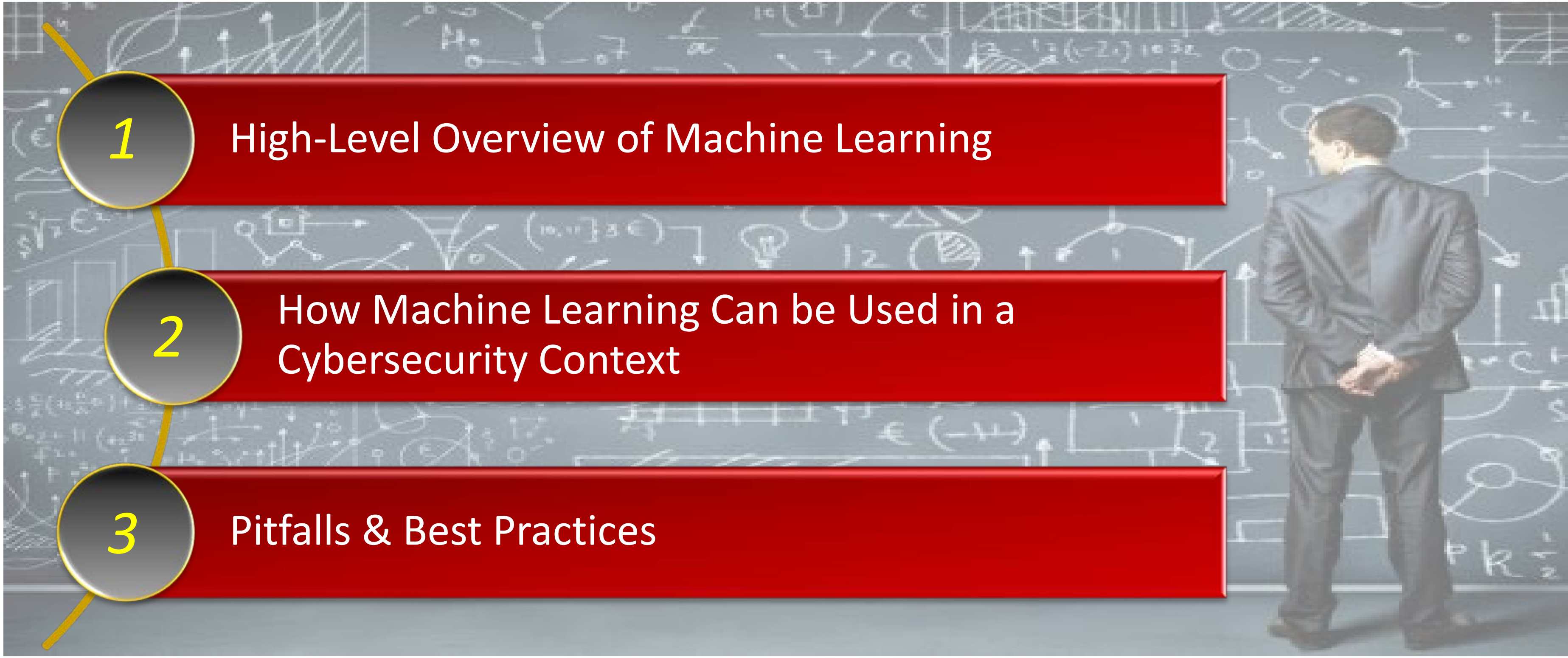
Machine Learning – Cybersecurity Boon or Boondoggle?

Zulfikar Ramzan

Chief Technology Officer
RSA
@zulfikar_ramzan

#RSAC

Outline



1 High-Level Overview of Machine Learning

2 How Machine Learning Can be Used in a Cybersecurity Context

3 Pitfalls & Best Practices

What is Machine Learning?

Machine Learning Overview

"Field of study that gives computers the ability to learn without being explicitly programmed" (Arthur Samuelson, 1959)

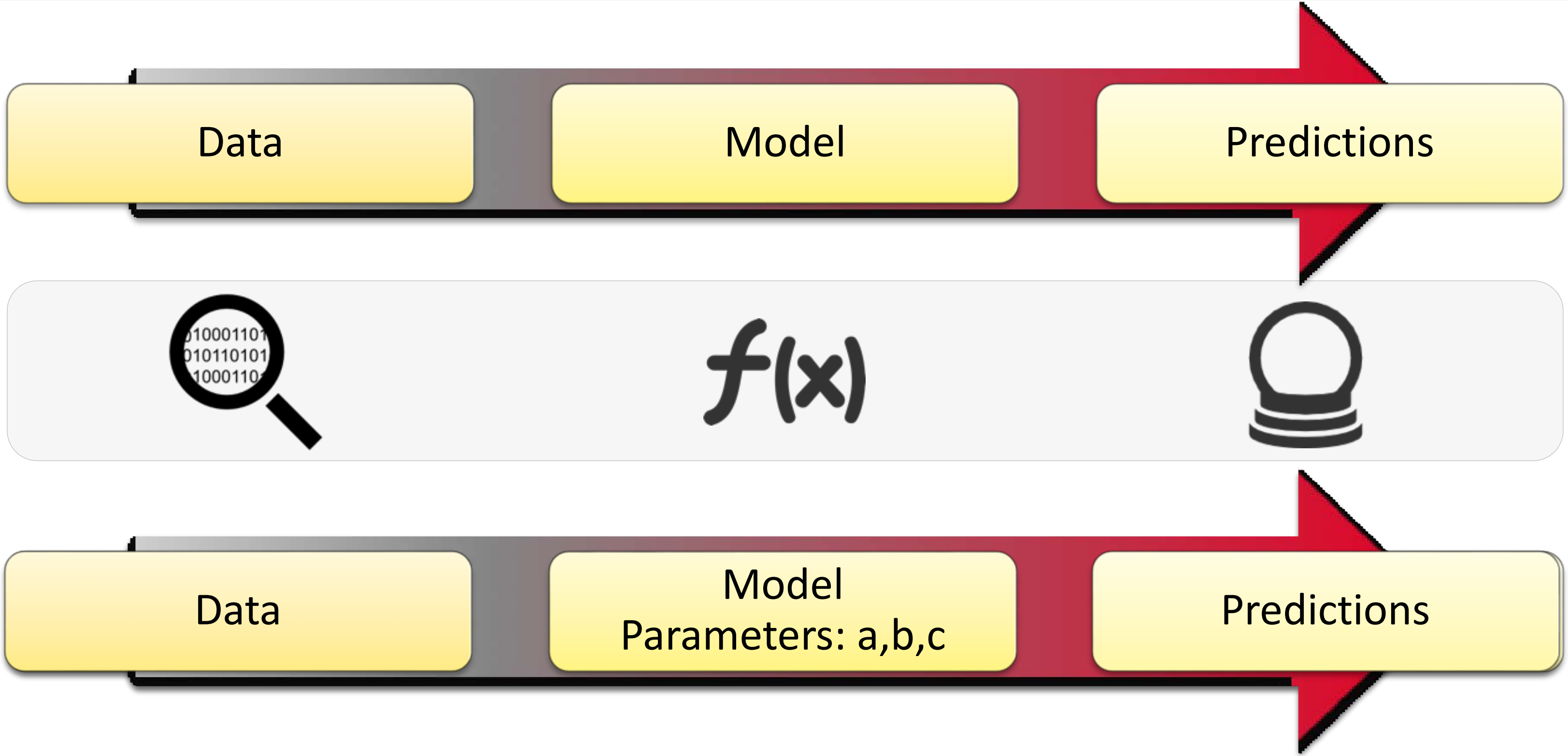
Development of algorithms that can draw inferences and make predictions based on data

Evolved from pattern recognition and computational learning theory; closely related to data mining and statistics

Benefits of machine learning include automation, being unbiased, and being able to improve over time

In cybersecurity applications, machine learning approaches can decrease time needed to find new threats and can evolve more-or-less automatically as new data becomes available

Machine Learning Process



Data

Model

Predictions

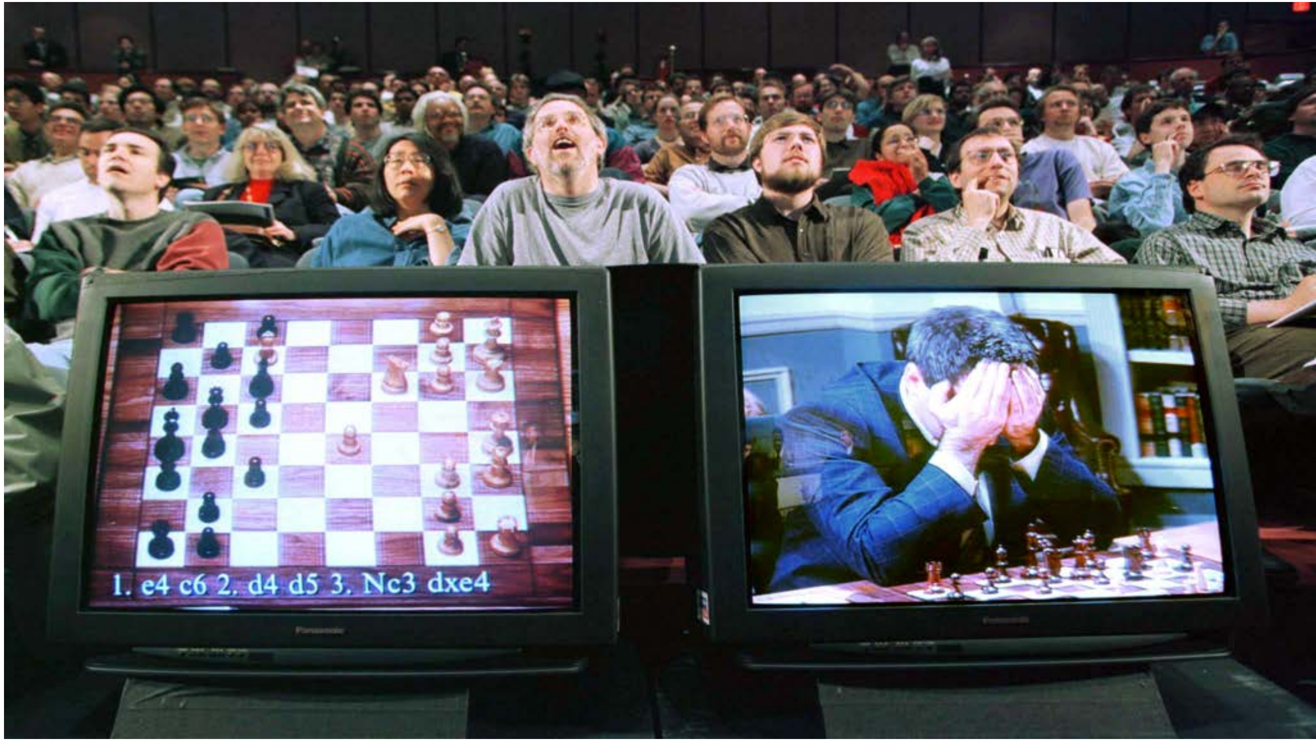
Data

Model
Parameters: a,b,c

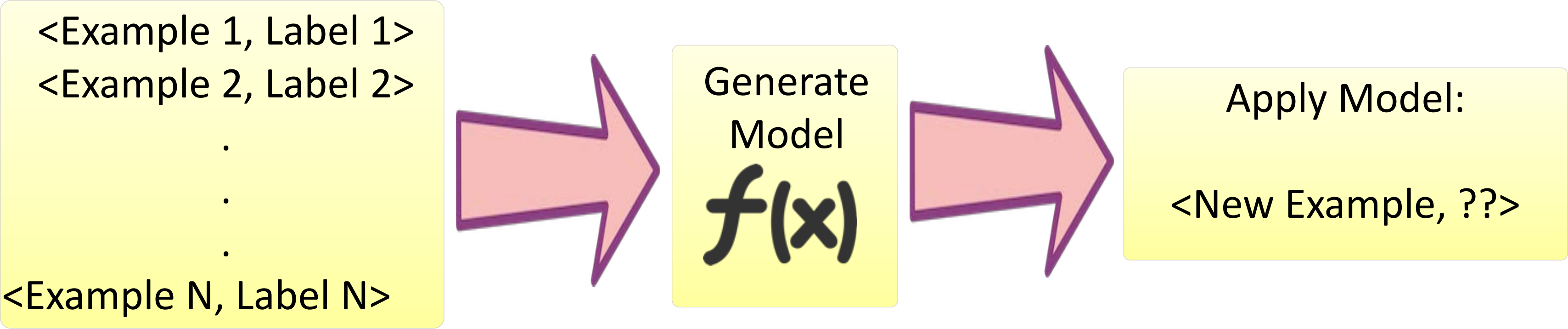
Predictions

What Machine Learning is Not...

In Kasparov vs. Deep Blue, the machine's programmers (with help from a team of strong chess players) explicitly entered parameters -- and even changed them between games in the match!



Supervised Learning



<Example 1, Label 1>
<Example 2, Label 2>
.
.
.
<Example N, Label N>

Generate Model

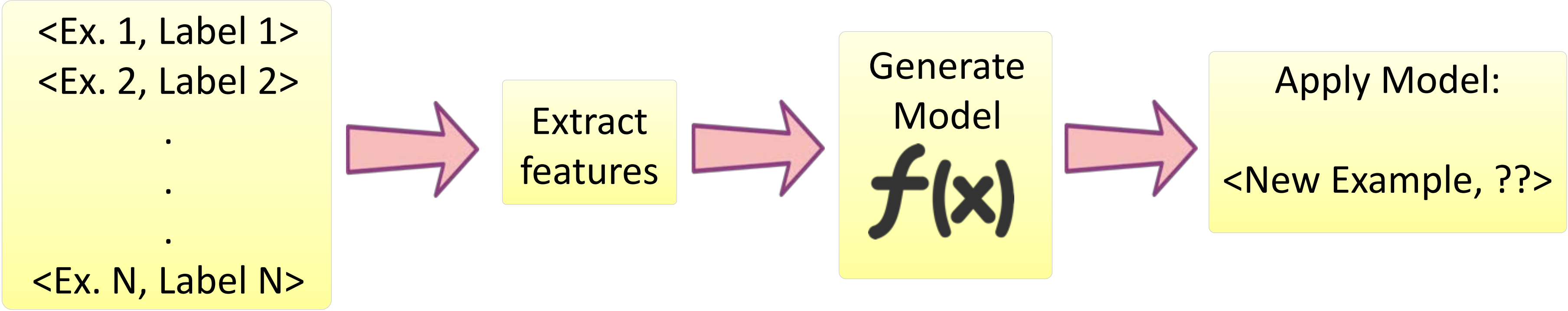
$f(x)$

Apply Model:

<New Example, ??>

Example:
Historical loan applications; labels represent whether applicant defaulted. Generate model that can predict whether new applicant will default.

Supervised Learning: More refined…



<Ex. 1, Label 1>
<Ex. 2, Label 2>
.
.
.
<Ex. N, Label N>

Extract features

Generate Model

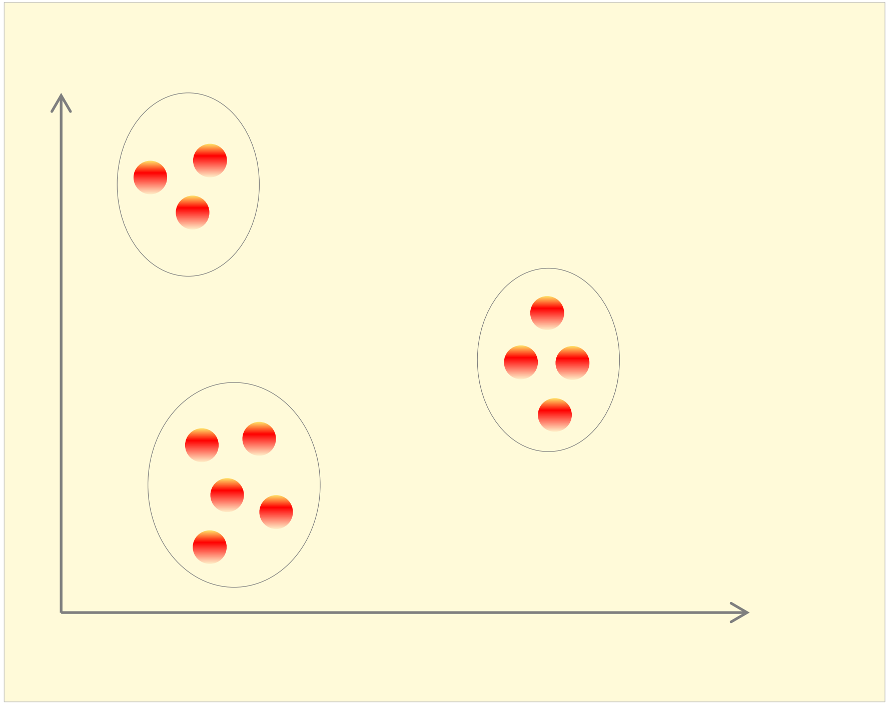
$f(x)$

Apply Model:

<New Example, ??>

Example:
Extract relevant attributes from loan application (e.g., years of employment, age, marital status, loan size, income). Generate model / classification algorithm from those features. (Feature identification is typically performed by a human domain expert.)

Unsupervised Learning



Learn from unlabeled examples

No labels, so no notion of right/wrong

Useful for identifying structure, patterns in data

How Can Machine Learning Be Used in a Cybersecurity Setting?

Applying Supervised Machine Learning to Security

Supervised machine learning naturally suited to classification problems; transactions can be viewed under a good / bad lens:

- **Spam:** Is a given email message spam?
- **Online Fraud:** Is a given financial transaction fraudulent?
- **Malware:** Is a given file malware?
- **Malicious URLs / domains / IPs:** Is a network connection to a given URL (resp. domain, IP address) associated with malicious activity?

Note these applications are not new; they have been studied in academia and/or have been implemented in commercial use for many years.

Unsupervised Machine Learning and Security

Unsupervised machine learning can be used to cluster data – which can help identify outliers; e.g., by baselining normal behavior and find instances outside the norm:

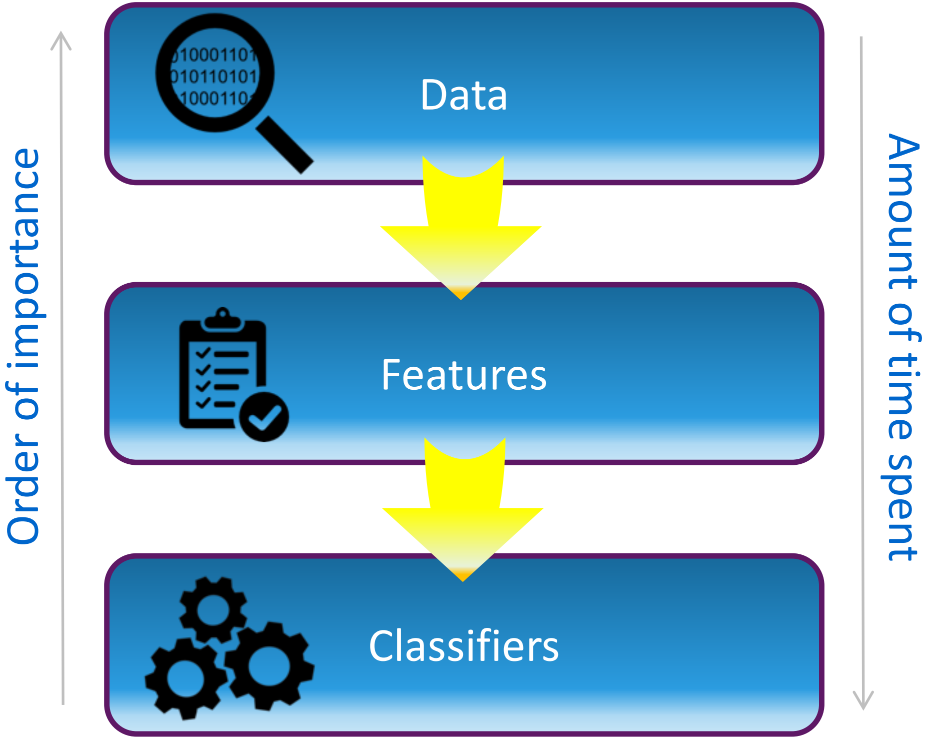
- Is there an abnormal amount of network traffic from a particular host?
- Is there a significant increase in failed log-in attempts?
- Is a user accessing resources he or she does not normally access (or that would not be normally accessed by his or her peer group)?
- Are there access patterns that are too "regular" to be associated with a human?
- Is a user working during hours outside his or her normal behavior?
- Is a user connecting from or to unusual geographic locations (or a set of geographic locations that does not make sense)?

What are the Pitfalls and Best Practices of Applying Machine Learning in Cybersecurity?

Challenge: Good Data is Critical

- Machine learning is garbage-in-garbage out
- Training set ideally has to be representative of what you will actually encounter in real life
- Ultimately, you have to ask the right questions of your data, otherwise you won’t get the right answers.

Pitfall: Becoming Obsessed with Classifier



Challenges: Class Imbalance in Security

Security scenarios: Class Imbalance Problem

# Good instances >> # Bad Instances

Must sidestep many landmines

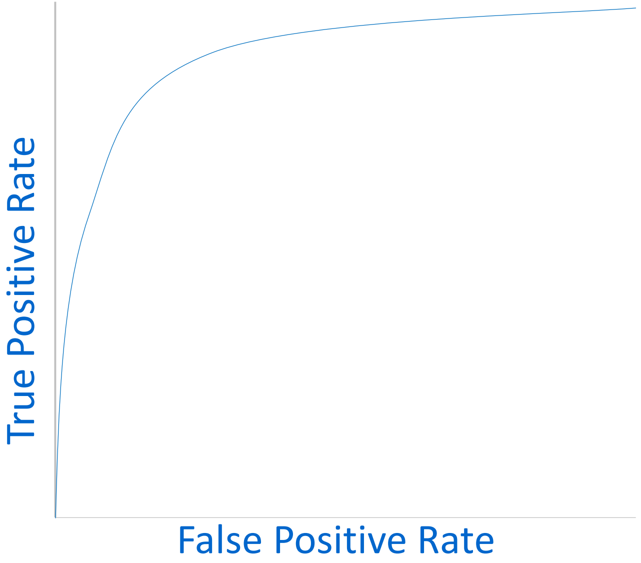
Typical Metrics

Classic efficacy metrics:

- True Positive Rate (TPR): % of actual malicious transactions are correctly identified
- False Positive Rate (FPR): % of legitimate transactions are labeled as malicious

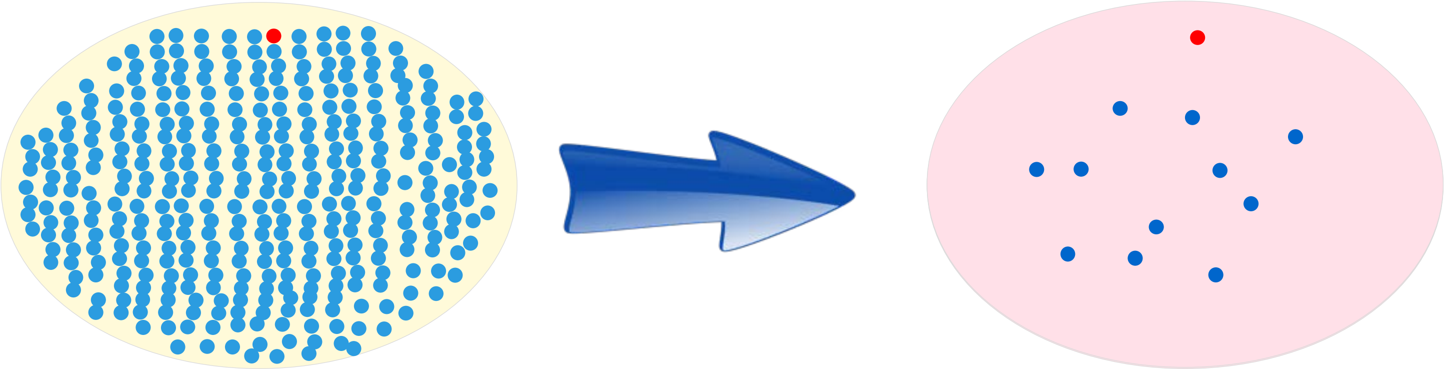
Tradeoff between these two metrics:

- Aggressive in calling activity malicious: high TPR, but likely also a high FPR
- Conservative in calling activity malicious: low FPR, but also likely low TPR
- Tradeoff often represented by Receiver Operating Characteristic (ROC) Curve



False Positives Don’t Tell the Whole Story

- Suppose that you have built a system with a false positive rate of %0.1 and a true positive rate of 100%. These numbers ostensibly seem amazing.
- However, imagine that for every 10,000 legitimate transactions, one is malicious.
- Then, out of these transactions, your system will have 11 alarms (10,000 * 0.001 + 1), of which only one is justified! (~90%+ of the time will be chasing false leads)



Think About Your Metrics Carefully

Not all false positives created equal (popular file vs. rare file)

Not all true positives created equal (pernicious threat vs. mild one)

Maliciousness is subjective (some people want adware and spam!)

If machine learning tech part of larger system; must consider additive benefit

Machine Learning Pitfalls: Adversaries Adapt

Threat actors change behaviors as needed; time to change is correlated with how big a threat you are to their operations

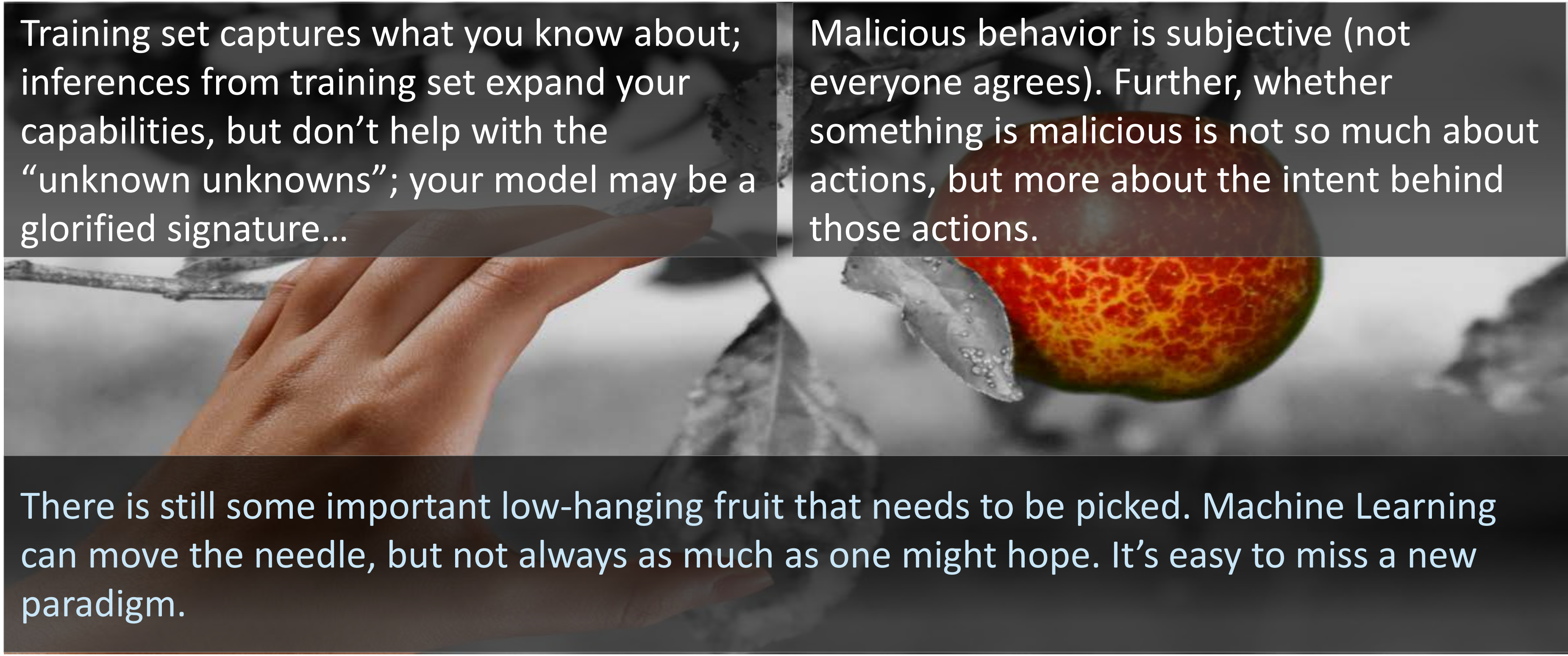
Machine learning algorithms typically don’t assume adversarial scenarios where threat is actively trying to sabotage the algorithm

Can address through careful weighting of recent data vs. older data, rapid retraining / relearning, online learning

Transfer learning or inductive transfer is area of research designed for developing ML algorithms that try to use knowledge gained from solving one problem towards different, but related problem

Much work remains to be done in this areas

Supervised Learning Pitfalls



Training set captures what you know about; inferences from training set expand your capabilities, but don’t help with the “unknown unknowns”; your model may be a glorified signature…

Malicious behavior is subjective (not everyone agrees). Further, whether something is malicious is not so much about actions, but more about the intent behind those actions.

There is still some important low-hanging fruit that needs to be picked. Machine Learning can move the needle, but not always as much as one might hope. It’s easy to miss a new paradigm.

The importance of line 23…

Unsupervised Learning Pitfalls

Unsupervised learning used to find abnormal behaviors, but that’s not the same as malicious behaviors

People often act abnormally for legitimate reasons (travel, deadlines, new project / role / promotion)

It’s not always easy to measure normal behavior (e.g., geolocation isn’t always accurate…)

Again, actions are not the same thing as the intent behind those actions….

Model Deployment

Key Concerns:

Machine Learning algorithms can produce models, but that model still has to be deployed successfully.

Need independent mechanism for determining whether model “makes sense” and will not cause issues in real life; i.e., real-world ROI is hard to measure.

Key Questions:

Did you overfit to the training set? Can someone with limited machine learning ability understand the model and how/why it works (e.g., when debugging in-field issue)? Are there fail-safe measures in place in case model is flawed?

Summary

Machine Learning is an excellent tool for security applications, but it's important to understand when hype fails to reflect reality

Best Practices:

- Data >> Features >> Classifier
- Know what success means (not necessarily high TP, low FP)
- Don't set it and forget it; adversaries adapt!
- Think about model deployment; not everyone touching system will have deep machine learning knowledge

Apply What You Have Learned (1)

Next week, you should:

- Identify what cybersecurity efficacy metrics matter to you
- Develop an efficacy testing plan that measures against these metrics and evaluates the tradeoffs among them

In the first three months following the presentation, you should

- Implement your efficacy testing plans
- Identify the data sources being fed into machine learning systems you use and determine how representative those sources are of real world scenarios.
- Determine how models are updated for future use

Apply What You Have Learned (2)

If you are considering procuring a third-party machine learning solution, develop a vendor questionnaire that asks (at least) the following questions:

- What data sources are being fed into their machine learning system?
- How is efficacy being measured prior to model deployment? (And is it being measured independently?)
- Do other vendors offer solutions to address the same problems (including those that do not leverage machine learning)? If so, are there rigorous third party testing reports to substantiate efficacy claims of ? (If not, why?)
- How are models deployed and updated?