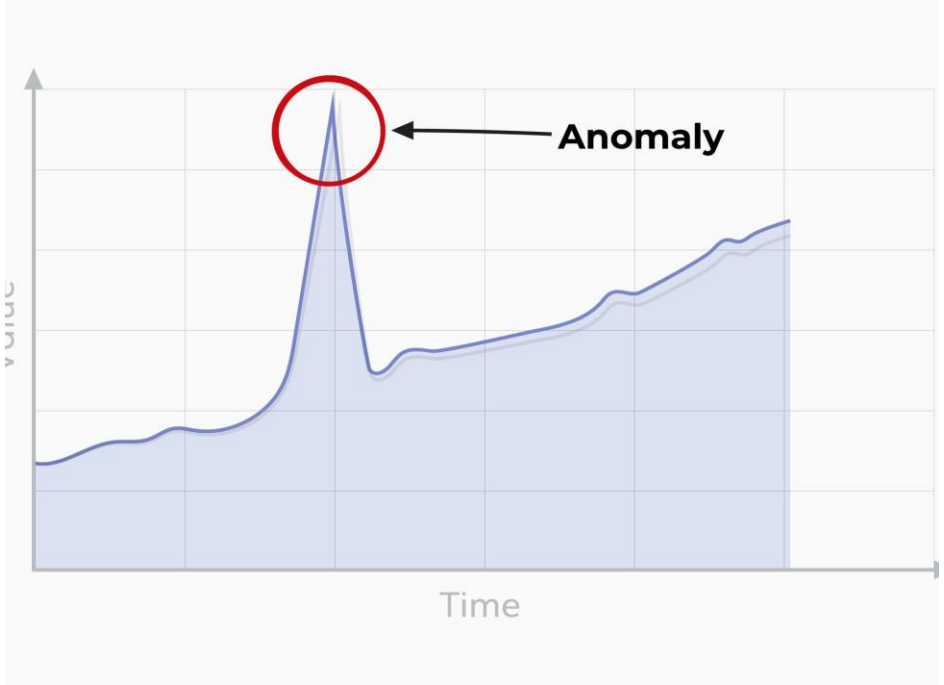
COMPUTER SCIENCE
& ENGINEERING
TEXAS A&M UNIVERSITY

Intrusion Detection - Denning

Dr. Martin “Doc” Carlisle

Anomaly

Merriam Webster: “something different, abnormal, peculiar, or not easily classified”



Why do we care?

"The model is based on the hypothesis that exploitation of a system's vulnerabilities involves abnormal use of the system; therefore, security violations could be detected from abnormal patterns of system usage." (Denning 1987)

Example Anomalies

1. High rate of password failures
2. Login times/locations
3. Access failures
4. Increased data access/rate
5. CPU time, I/O rate
6. System calls by viruses
7. IP addresses/ports

Intrusion Detection "Expert System" (IDES)

- 6 components
  - Subjects (usually users)
  - Objects (resources, files, commands, devices)
  - Audit records (logins, command executions, file access)
  - Profiles (auto generated statistical metrics)
  - Anomaly records (generated when abnormal behavior detected)
  - Activity rules (actions taken when condition met)

IDES Audit record

- 6 tuple
  - <Subject, Action, Object, Exception-Condition, Resource-Usage, Time-stamp>
    - Actions are things like login, read, execute
    - Resource usage – CPU time, pages printed, number of records read, CPU time, etc.

IDES Example Audit record

- 6 tuple
  - <Subject, Action, Object, Exception-Condition, Resource-Usage, Time-stamp>

COPY GAME.EXE TO <Library>GAME.EXE

issued by user Smith to copy an executable GAME file into the <Library> directory; the copy is aborted because Smith does not have write permission to <Library>:

(Smith, execute, <Library>COPY.EXE, 0,
CPU=00002, 11058521678)
(Smith, read, <Smith>GAME.EXE, 0,
RECORDS=0, 11058521679)
(Smith, write, <Library>GAME.EXE, write-viol,
RECORDS=0, 11058521680)

IDES Metrics

- Event counter (e.g. number of logins)
- Interval timer (e.g. length of time between logins, commands)
- Resource measure (CPU time, pages printed)

What's Abnormal? – Threshold Metric

- Microsoft Windows can lock a user account after n failed login attempts
- iPhone

Times	Screen Message	Action or Consequence
1th-5th	Notifications saying the passcode is wrong	Give another try
6th	iPhone is disabled, try again in 1 minute	Wait 1 minute and try again
7th	iPhone is disabled, try again in 5 minutes	Wait 5 minutes and try again
8th	iPhone is disabled, try again in 15 minutes	Wait 15 minutes and try again
9th	iPhone is disabled, try again in 60 minutes	Wait 60 minutes and try again
10th	iPhone is disabled. connect to iTunes	iPhone is completely disabled

What's Abnormal? – Statistical Methods

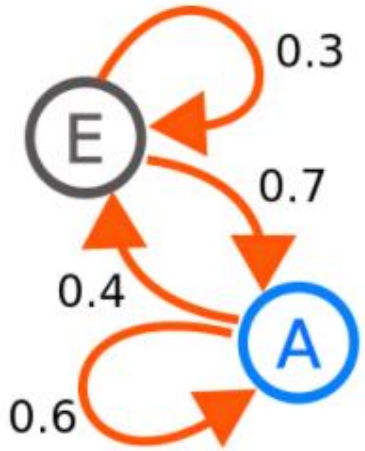
- Mean and Standard Deviation Model
  - A new observation is "abnormal" if it falls outside a confidence interval ($\mu \pm k\sigma$)
  - Chebyshev's inequality says the probability of being outside this is at most $1/d^2$.
    - D=4, 0.0625
    - D=5, 0.04
    - D=6, 0.0277
    - …
    - D=10, 0.01
  - Note we can get a much tighter bound if our data is normally distributed (D=5 is 0.000000574)

What's Abnormal? – Statistical Methods

- Multivariate Model
  - Similar to Mean & Stddev, but based on correlations between two or more metrics (e.g. CPU time and I/O units, login frequency, session elapsed time)

What's Abnormal? – Statistical Methods

- Markov Process Model
  - Applied to event counters
    - Each audit record is a state variable
    - Use a state transition matrix with frequencies
    - Abnormal if probability of transition is too low



What's Abnormal? – Statistical Methods

- Time Series Model
  - Uses interval timer with event counter/resource measure
  - Measures trends of behavior over time
  - Abnormal if probability for this event is "too low"

IDES Profile Structure

- 10 components
  - Variable-Name
  - Action-Pattern (regular expression)
  - Exception-Pattern (regex)
  - Resource-Usage-Pattern
  - Period
  - Variable-Type (model)
  - Threshold (upper/lower bound, # of std devs)
  - Subject-Pattern
  - Object-Pattern
  - Value (current statistical values e.g. mean, count, std dev)

IDES Suggested Profiles

- Login Frequency
- Location Frequency (login)
- Last Login (timer) – e.g. find break-in on “dead” account

IDES Profile Structure

Variable-Name:	SessionOutput
Action-Pattern:	'logout'
Exception-Pattern:	0
Resource-Usage-Pattern:	'SessionOutput=' # → Amount
Period:	
Variable-Type:	ResourceByActivity
Threshold:	4
Subject-Pattern:	'Smith'
Object-Pattern:	*
Value:	record of . . .

IDES Suggested Profiles

- Login and session activity
  - Login Frequency
  - Location Frequency (log in from unusual place)
  - Time between logins
  - Session time
  - Quantity of terminal output per session
  - Session CPU, IO, memory use
  - Password fails
  - Location fails

IDES Suggested Profiles

- Command or program execution
  - Frequency
  - CPU, IO, memory use
  - Execution denied
  - Resource exhaustion

IDES Suggested Profiles

- File Access Activity
  - Read/Write/Create frequency
  - Records read/written
  - Read/Write/Delete/Create fails
  - Resource exhaustion

Denning's Open Questions

- Soundness: Does the approach detect intrusions? Can we distinguish intrusion anomalies from others?
- Completeness: Do we miss a significant proportion of intrusions?
- Timeliness: Can this be done fast enough to matter?
- Choice of model
- Social implications: how are users affected? (e.g. consider valid charges being declined)

COMPUTER SCIENCE
& ENGINEERING
TEXAS A&M UNIVERSITY

Base Rate Fallacy - Axelsson

Dr. Martin “Doc” Carlisle

Axelsson Sec 2 – references James P Anderson - Masquerader

- Extra use of system by unauthorized user
  - Look for abnormal
    - Time
    - Frequency
    - Volume
    - Patterns of reference to programs or data

Later work - Questions about Intrusion Detection

- Effectiveness – how good is it? (false alarms, false negs)
- Efficiency – computing resources required
- Ease of use – esp. wrt false alarms
- Security – can the IDS be attacked?
- Interoperability – between IDS
- Transparency – how intrusive is IDS?

Base-Rate Fallacy-Axelsson

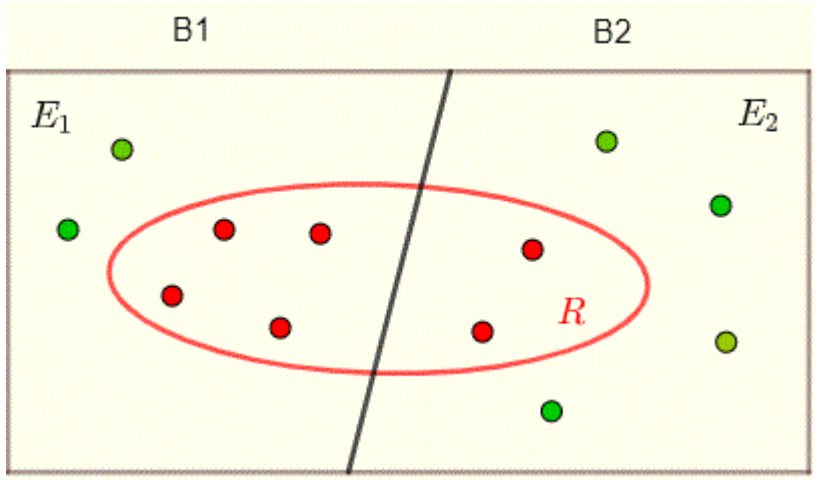
The base-rate fallacy is best described through example.[4] Suppose that your physician performs a test that is 99% accurate, i.e. when the test was administered to a test population all of which had the disease, 99% of the tests indicated disease, and likewise, when the test population was known to be 100% free of the disease, 99% of the test results were negative. Upon visiting your physician to learn of the results he tells you he has good news and bad news. The bad news is that indeed you tested positive for the disease. The good news however, is that out of the entire population the rate of incidence is only 1/10000, i.e. only 1 in 10000 people have this ailment. What, given the above information, is the probability of you having the disease?[6]

Bayes' Theorem

- $P(A|B)=\frac{P(A)P(B|A)}{P(B)}$
- Given a binary variable:
  - $P(A|B) = \frac{P(A)P(B|A)}{P(B|A)\,P(A)+P(B|not\ A)\,P(not\ A)}$

Bayes' Theorem example

$$P(A|B) = \frac{P(A)P(B|A)}{P(B|A)\,P(A) + P(B|not\ A)\,P(not\ A)}$$



P(B1)=0.5
P(B2)=0.5
P(R|B1)=2/3
P(R|B2)=1/3

If I draw a red ball, what's the probability it came from Box 1?

$$P(B1|R) = \frac{P(B1)P(R|B1)}{P(R|B1)\,P(B1)+P(R|B2)\,P(B2)} = \frac{0.5*2/3}{\frac{2}{3}*0.5+\frac{1}{3}*0.5} = 2/3$$

Base-Rate Fallacy

$$P(A|B) = \frac{P(A)P(B|A)}{P(B|A)\,P(A) + P(B|not\ A)\,P(not\ A)}$$

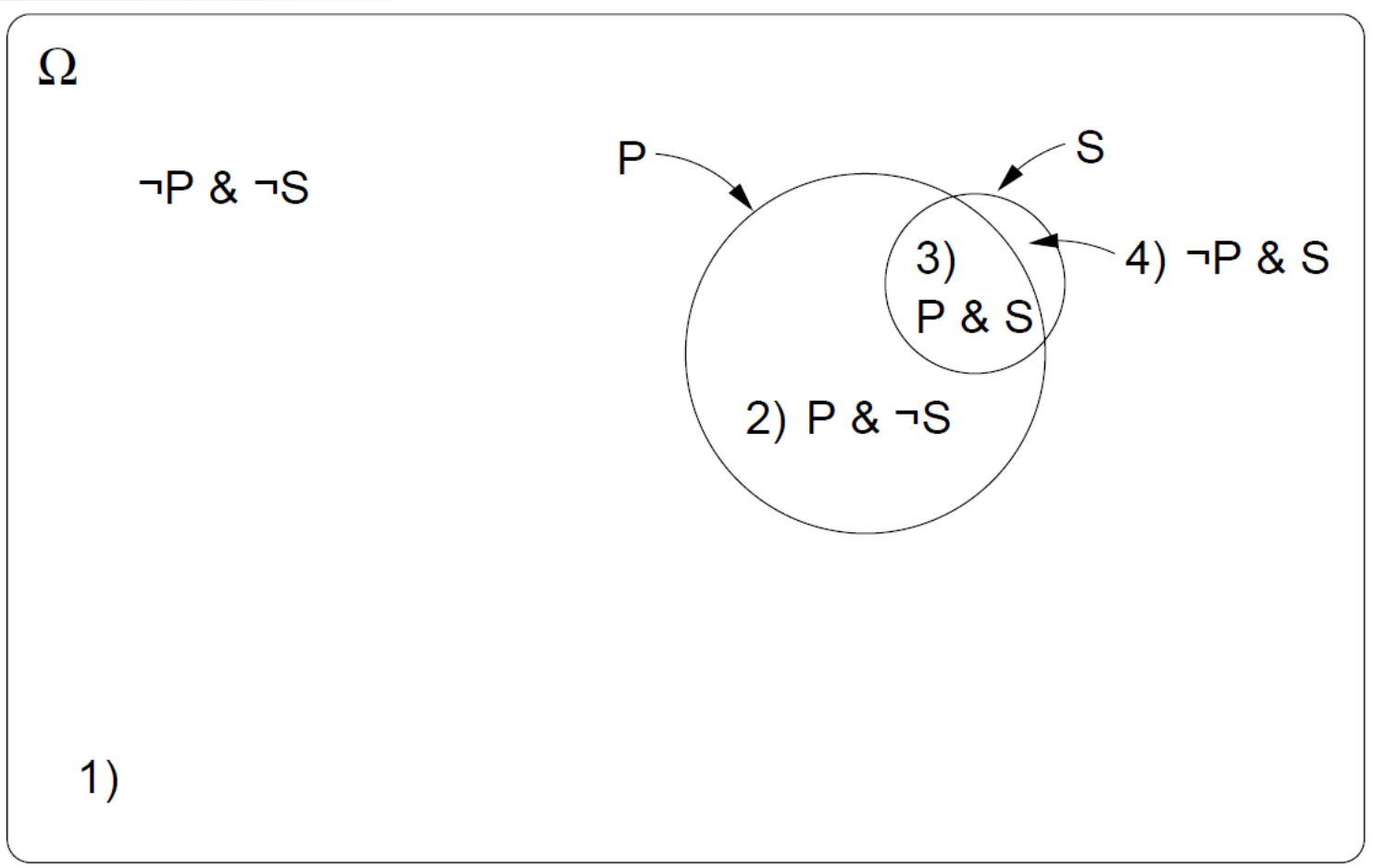
$$P(Disease|Pos) = \frac{\frac{1}{10000} * 0.99}{0.99 * \frac{1}{10000} + 0.01 * \frac{9999}{10000}} = 0.00980\ldots$$

– Only a ~1% chance!

The base-rate fallacy is best described through example.[4] Suppose that your physician performs a test that is 99% accurate, i.e. when the test was administered to a test population all of which had the disease, 99% of the tests indicated disease, and likewise, when the test population was known to be 100% free of the disease, 99% of the test results were negative. Upon visiting your physician to learn of the results he tells you he has good news and bad news. The bad news is that indeed you tested positive for the disease. The good news however, is that out of the entire population the rate of incidence is only 1/10000, i.e. only 1 in 10000 people have this ailment. What, given the above information, is the probability of you having the disease?[5]

Venn Diagram example

Not to scale



Definitions

- I=Intrusion, A=Alarm

- True positive – detection rate P(A|I)
- False positive – P(A|not I)
- False negative – P(not(A)|I) = 1-P(A|I)
- True negative – P(not(A)|not(I)) = 1-P(A|not(I))

Axelsson cont'd

- Suppose we get 1,000,000 audit records/day, 10 per intrusion and 2 intrusions per day.
- P(I)=20/10^6, P(not I)=0.99998

Dominant!

$$P(I|A) = \frac{2 \cdot 10^{-5} \cdot P(A|I)}{2 \cdot 10^{-5} \cdot P(A|I) + 0.99998 \cdot P(A|\neg I)}$$

Is it real?

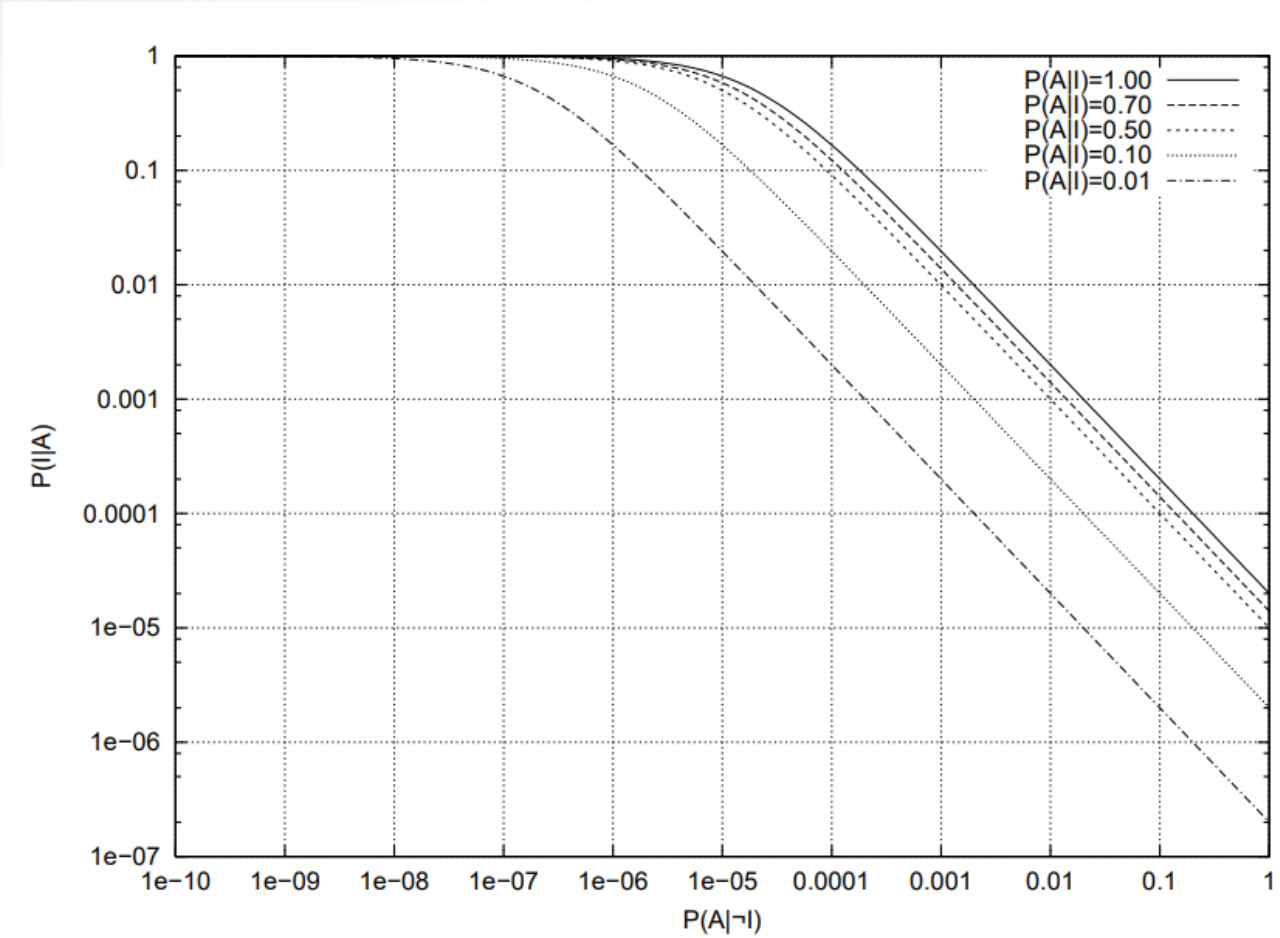
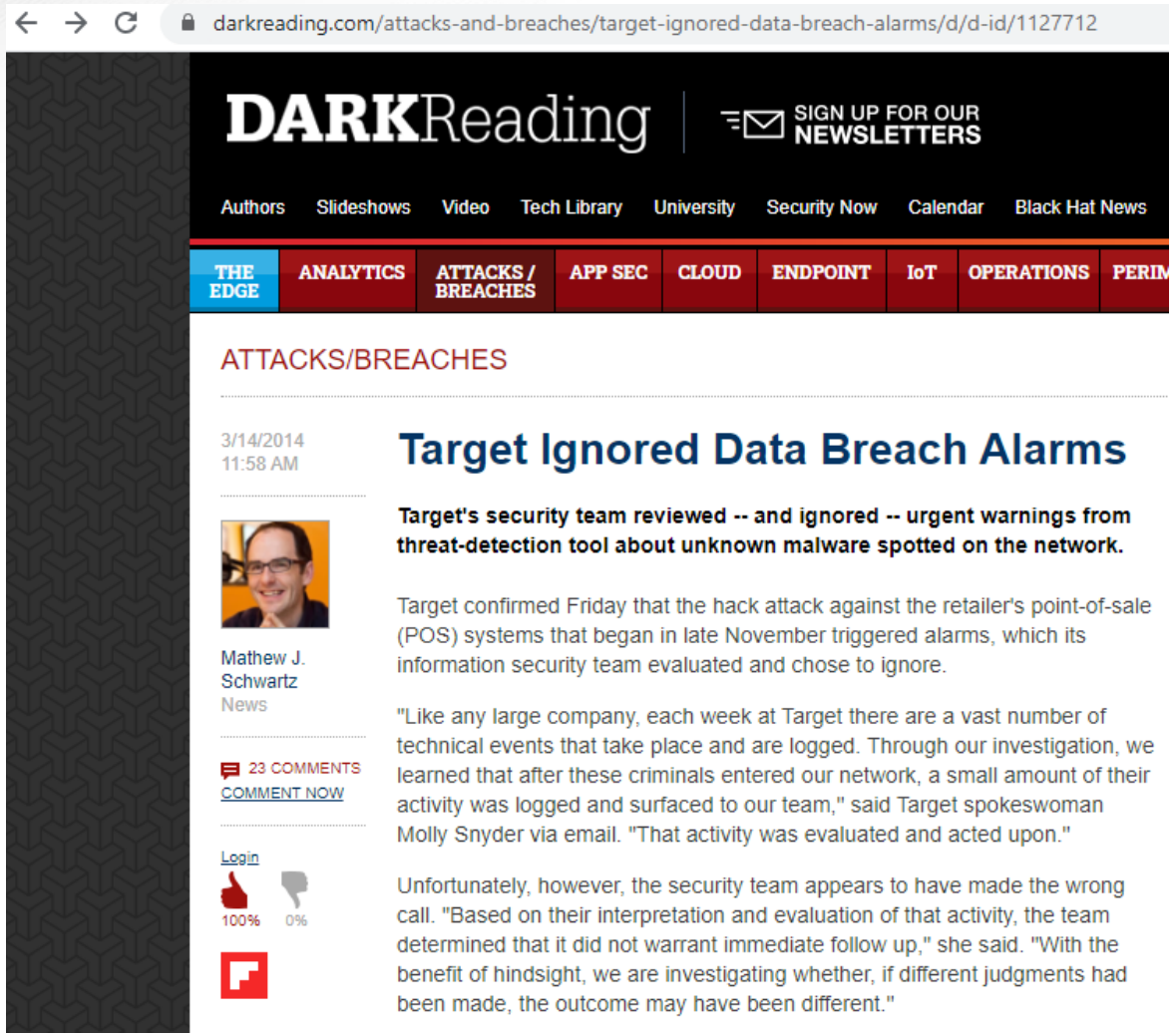


Figure 1: Plot of $P(I|A)$

What does this mean for Cyber?



Target Attack Timeline

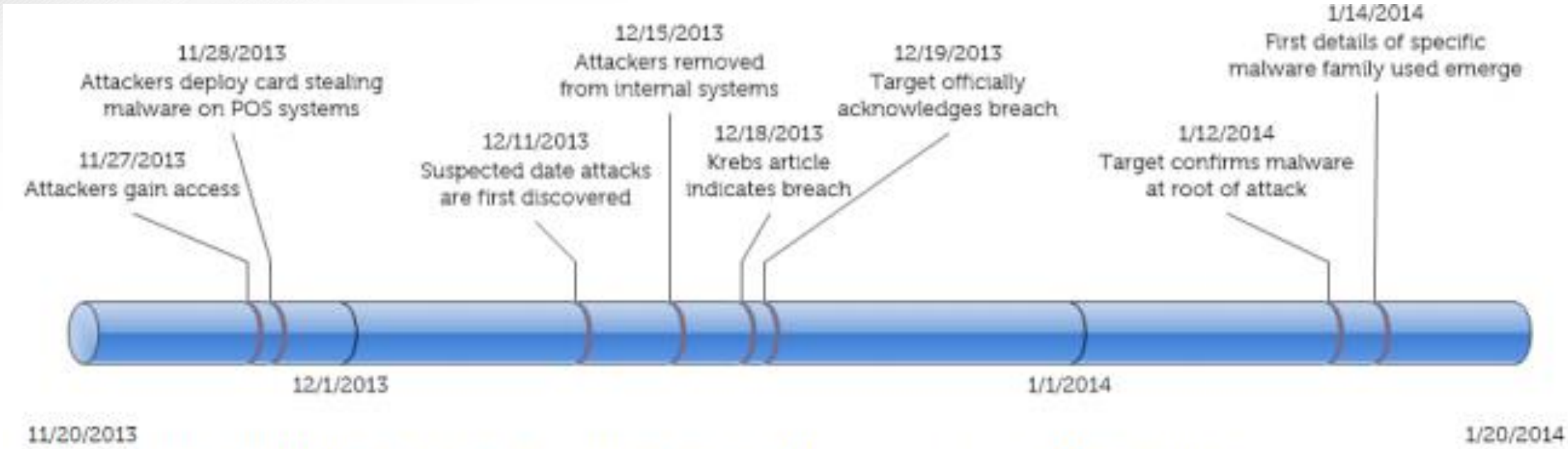


Figure 2. Approximate timeline of events in the Target data breach. (Source: Dell SecureWorks)

Axelsson Conclusion

- Must keep false alarm rate below 1/100,000
- This might be unattainable

COMPUTER SCIENCE
& ENGINEERING
TEXAS A&M UNIVERSITY

Anomaly Detection Survey

Dr. Martin “Doc” Carlisle

Motivation

- Anomalies may yield critical, actionable intelligence
  - Traffic pattern may indicate hacked computer
  - Anomaly in MRI might indicate malignant tumor
  - Anomaly in credit card data might mean theft
  - Spacecraft sensor anomaly might indicate component fault

Challenges

- Defining “normal” is hard
  - Boundary between “normal” and “anomalous” often not precise
- Malicious actors adapt to look “normal”
- “Normal” behavior keeps changing
- Techniques don’t transfer easily between domains
  - E.g. consider human temperature changes vs. stock price changes
- Labeled data is hard to get!
- Noise can be similar to “anomalies”

Anomaly Types (I)

- Point anomaly
  - Ex. Amount spent in credit card fraud detection

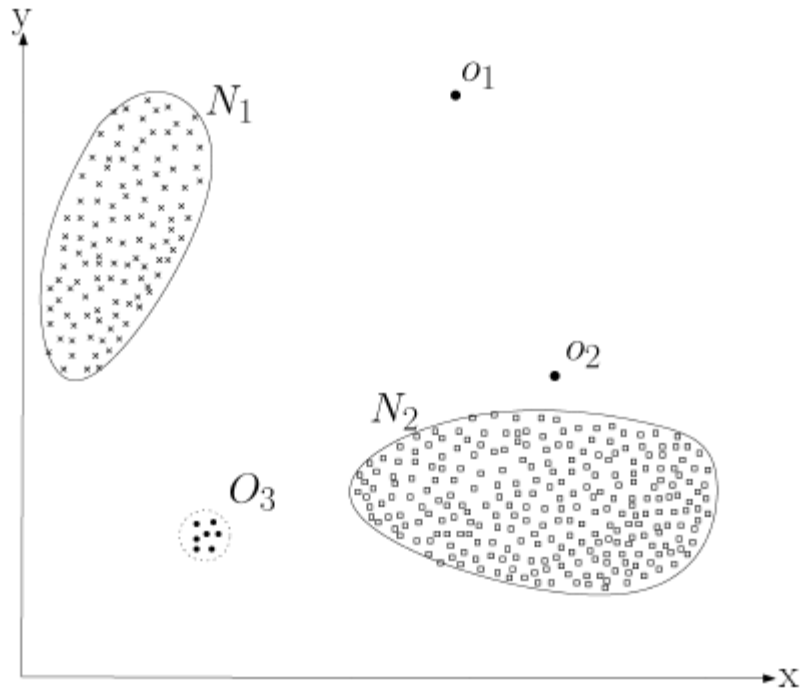
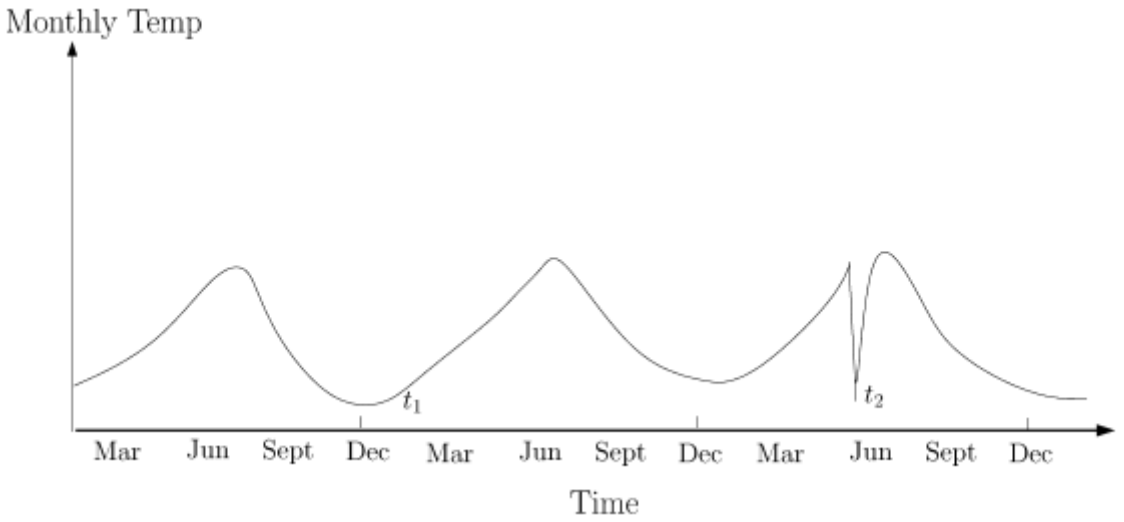


Fig. 1. A simple example of anomalies in a two-dimensional data set.

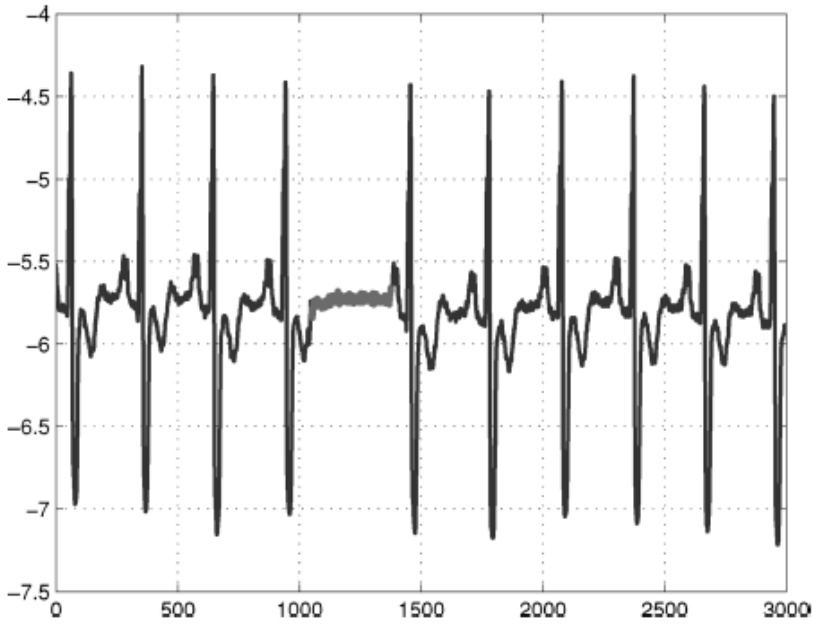
Anomaly Types (II)

- Contextual anomaly
  - Point is anomalous only in context (see June below)



Anomaly types (III)

- Collective Anomaly
  - Single points aren't anomalous, but collectively they are



Do we have labeled data?

- Supervised
  - Can train on data with labeled instances of normal vs anomaly classes
  - Not very common
- Semisupervised
  - Labeled instances for only normal data
- Unsupervised
  - No labeled data

Outputs

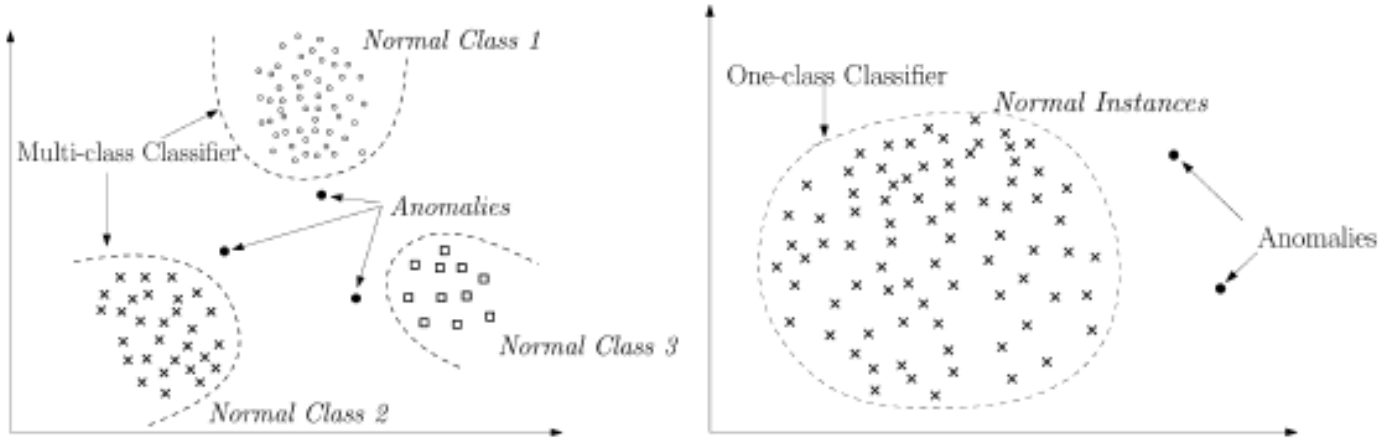
- Scores
  - E.g. a probability
- Labels
  - E.g. normal or anomalous

Applications

- Intrusion Detection (host and network based)
- Fraud Detection (e.g. ID theft, credit card, mobile phone, insurance claim, insider trading)
- Medical data (e.g. tumor detection, A-fib, etc.)
- Machine defects
- Image Processing
- Text data (e.g. terrorist threats)
- Sensor networks (e.g. gunshot detection)

Classification

- Semi-supervised
  - Train classifier with labeled normal data, use to identify anomalies

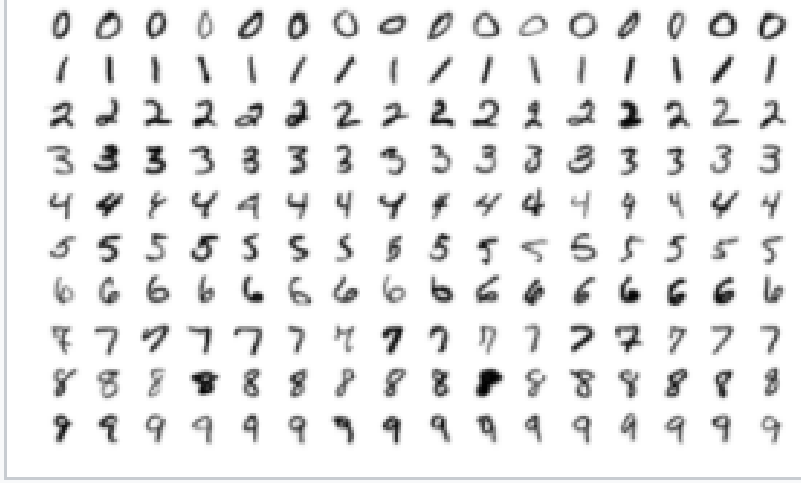


(a) Multi-class Anomaly Detection (b) One-class Anomaly Detection

Fig. 6. Using classification for anomaly detection.

Neural Network Classification

- Use NN to create classes, then see how new data is classified



Sample images from MNIST test dataset

Bayesian Networks

- Determine probability a particular data point is from each class using probability distributions obtained from training data
  - Select “most probable” class

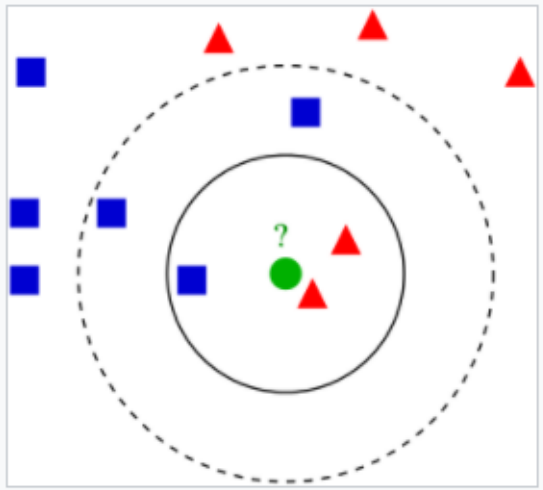
Support Vector Machine based

- Use hyperplanes to split data into regions with training data, and then select in which region datapoint falls.
  - “Kernels” can help map non-linear data to regular surface

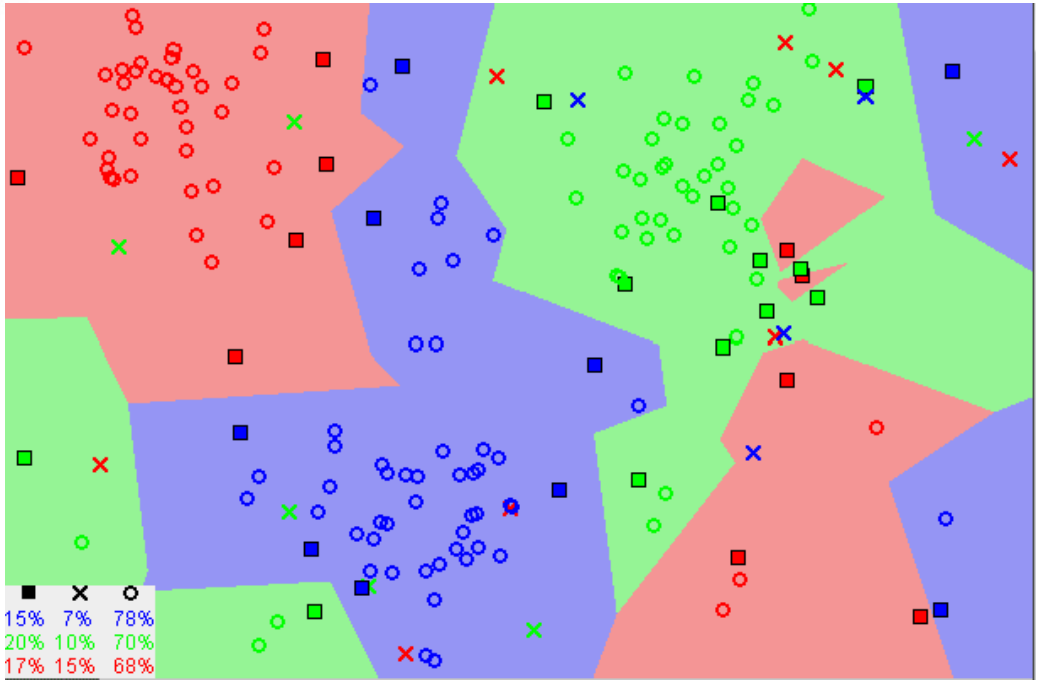
Rule based

- Create a set of rules (e.g. “users don’t log in more than once a day”) that capture normal behavior

k^{th} nearest neighbor

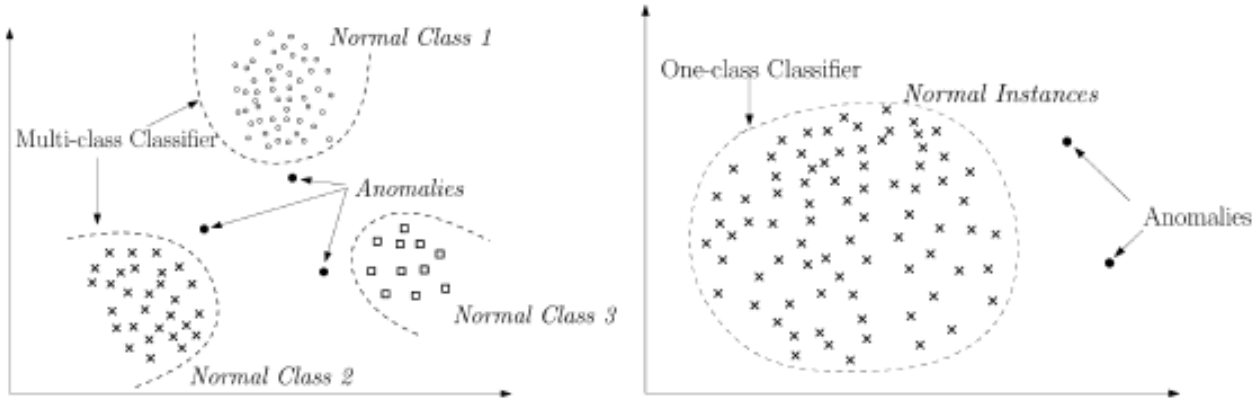


Example of *k*-NN classification. The test sample (green dot) should be classified either to blue squares or to red triangles. If *k* = *3* (solid line circle) it is assigned to the red triangles because there are 2 triangles and only 1 square inside the inner circle. If *k* = *5* (dashed line circle) it is assigned to the blue squares (3 squares vs. 2 triangles inside the outer circle).



Nearest Neighbor Techniques

- Distance to *kth* nearest neighbor is anomaly score
- Relative density of data instance is anomaly score



(a) Multi-class Anomaly Detection (b) One-class Anomaly Detection

Fig. 6. Using classification for anomaly detection.

k^{th} nearest neighbor

- Local density is better than global density

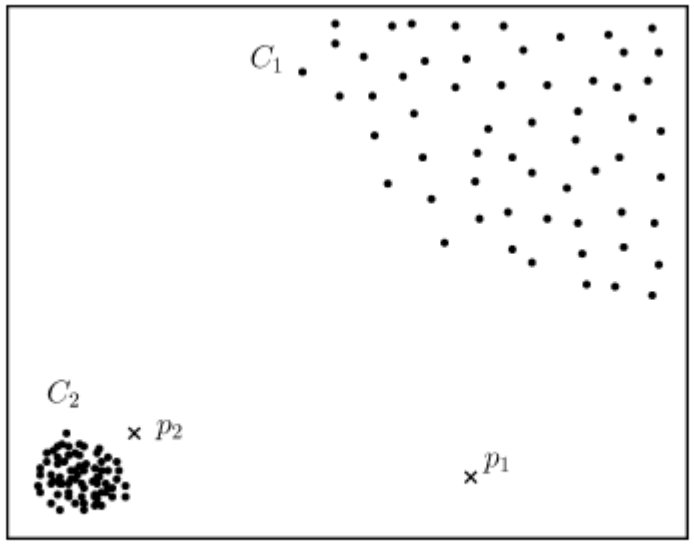


Fig. 7. Advantage of local density-based techniques over global density-based techniques.

k^{th} nearest neighbor

- Advantages
  - Unsupervised, data-driven
  - Can be improved with semi-supervised to catch more anomalies
  - Straight-forward to adapt to new datasets

k^{th} nearest neighbor

- Disadvantages
  - Unsupervised performs poorly if normal instances don't have enough neighbors, or anomalies have too many
  - Semi-supervised performs poorly if normal instances don't have enough neighbors
  - Computational complexity of testing is high
  - Defining distance may be difficult (think network packets, e.g.)

Clustering

- Similar to kth nearest neighbor, but define centroids of “normal”
- Anomaly score is distance to nearest centroid

Clustering Pros and Cons

- Advantages
  - Unsupervised
  - Can be adapted to other complex data types by plugging in a clustering algorithm
  - Testing phase is fast
- Disadvantages
  - Highly dependent on effectiveness of clustering algorithms in capturing structure
  - Many techniques detect anomalies as a byproduct of clustering and are not optimized for anomaly detection
  - Techniques may force anomalies to be assigned to some cluster
  - Some techniques only effective when anomalies don’t cluster
  - Clustering may be slow

Statistical Techniques

- Parametric
  - Assume data is generated by a parameterized distribution
    - Anomaly score is inverse of probability density function
  - Gaussian Model-Based
    - Assume data is generated from Gaussian distribution
      - 3 sigma rule (99.7%)
      - Box plot rule (use 1.5 * IQR – difference between lower and upper quartile), (99.3%)
      - Grubb's test: anomalous if difference from mean/std dev > $\frac{N-1}{\sqrt{N}}\sqrt{\frac{t^2_{\alpha/(2N),N-2}}{N-2+t^2_{\alpha/(2N),N-2}}},$

Statistical Techniques

- Parametric
  - Regression Model-Based
    - Use residual (part of test instance not explained by regression model)
- Non-Parametric
  - Histogram (does this fall in an empty or small bin?)
  - Kernel Function
    - E.g. parzen windows (use kernel function to approximate density)

Statistical Techniques

- Parametric
  - Regression Model-Based
    - Use residual (part of test instance not explained by regression model)
- Non-Parametric
  - Histogram (does this fall in an empty or small bin?)
  - Kernel Function
    - E.g. parzen windows (use kernel function to approximate density)

Statistical Pros and Cons

- Advantages
  - If assumptions regarding statistical distribution are true, very justifiable technique
  - Score corresponds to confidence interval, which can be used for tuning
  - If distribution estimate robust to anomalies, can be unsupervised
- Disadvantages
  - Many datasets don’t come from a particular distribution
  - Even if they do, choosing best test statistic isn’t straightforward
  - Histogram techniques don’t consider interactions between attributes
    - (each particular may not be rare, but combo is)

Information Theoretical Techniques

- Assumption: anomalies introduce irregularities in the information content of the dataset
  - (This is an oversimplification, but think about how easy it would be to compress the data if a particular element were removed)

Spectral Anomaly Detection

- Assumption: Can embed data in lower dimensional subspace where anomalies look very different than normal
  - E.g. Principal Component Analysis