Applications: Artificial Intelligence and Data Mining

COMP10001 Foundations of Computing

Week 11 Lecture 2

Artificial Intelligence

Definition: developing computer systems that can perform tasks that traditionally can only be done by a human

For example:

- Playing a good game of chess
- Self-driving car
- Translating spoken English into spoken Spanish in real-time
- Detecting that a user's account has been hacked

What types of intelligent behaviour are needed in these applications?

Today, we'll focus on one major area of Artificial Intelligence:

Data Mining and Machine Learning

Overview

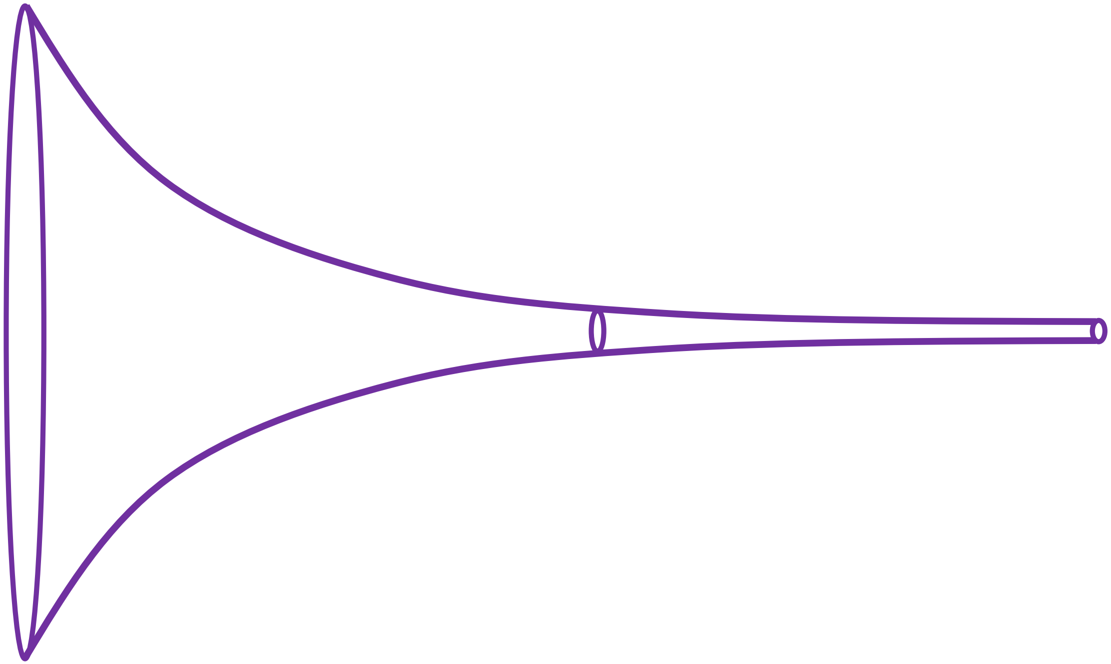
Data mining / Machine Learning aim to find useful patterns in large data sets

For example:

- Market segmentation studies
  - Find categories of customers with similar buying behaviour

- Predictive modelling
  - Find customers who are likely to commit fraud based on their transaction history

The Common Theme – Big Data

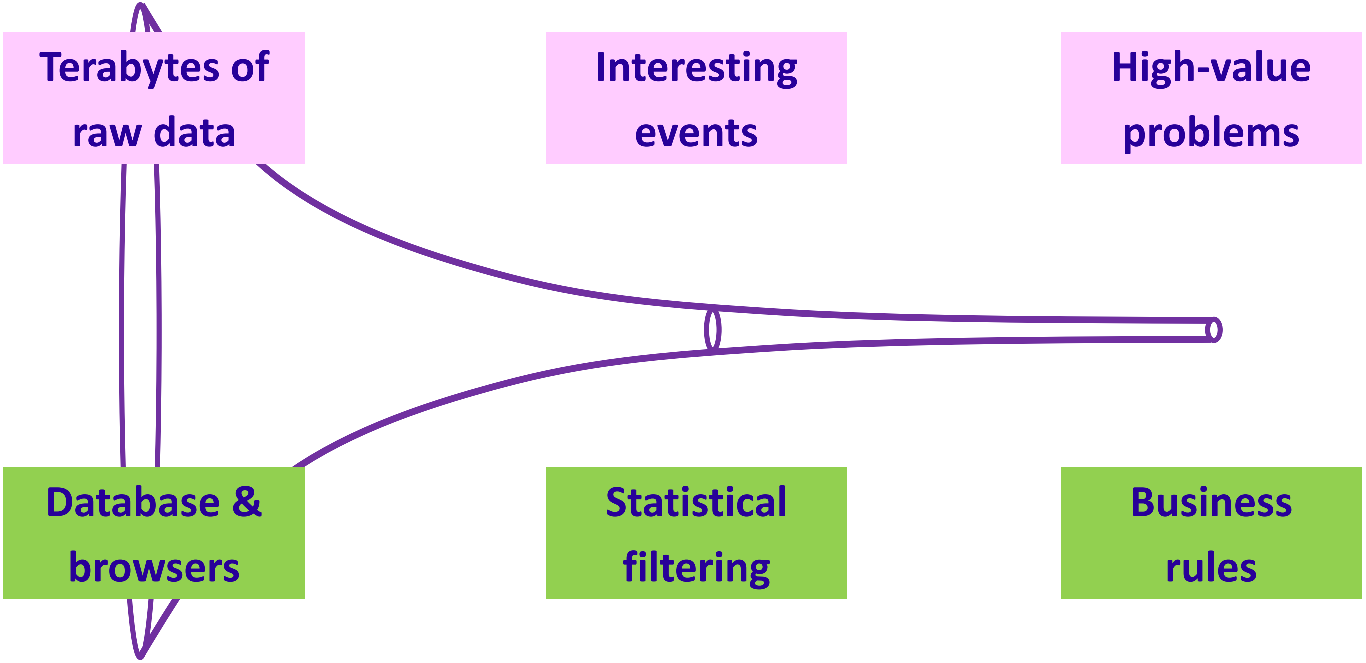


Terabytes of raw data

Interesting events

High-value problems

Automating the Data Analysis Pipeline



Part of the field of **data mining / machine learning**

Types of Learning Problems in Data Mining / Machine Learning

Supervised Learning:

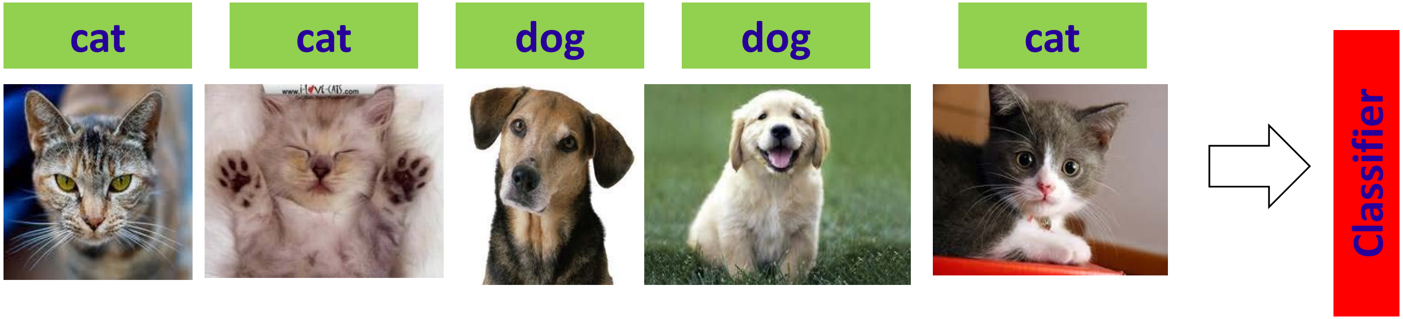
Learn a classifier from a set of labelled examples so that you can classify new unlabelled examples in the future

Unsupervised Learning:

Cluster a set of unlabelled examples to learn the natural categories or types of objects

Learning a Classifier (Supervised Learning)

Training a classifier



cat

cat

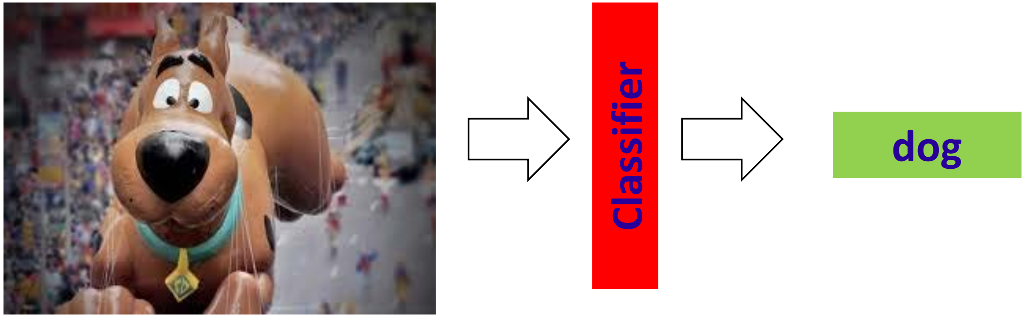
dog

dog

cat

Classifier

Classifying new examples



Classifier

dog

Clustering to Learn Categories (Unsupervised Learning)

What are the natural categories in a database?

Consider a database of animals.

How many different types of animals are there here?

Examples of Applications of Data Mining

Supervised Learning:

- Fraud detection from credit card transactions
- Face recognition in Facebook
- Diagnosing cancer from genetic test on blood samples

Unsupervised Learning:

- Modelling different types of network traffic (web, video, music, etc)
- Building an index of the types of documents on a web site
- Identifying different categories of customers on a retail website

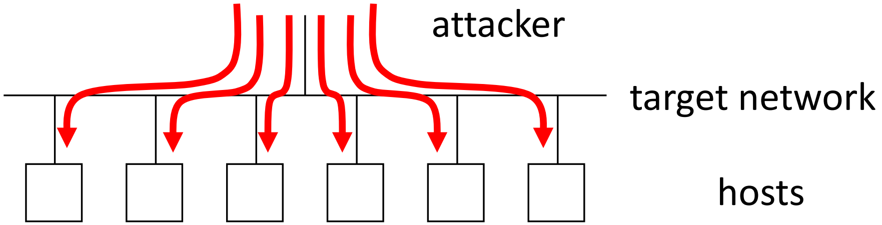
Learning Unusual Patterns (Anomaly Detection)

- Learn a model of “normal” database records
- Use this model to test new records for anomalies
- Any anomalies can be either interesting or errors

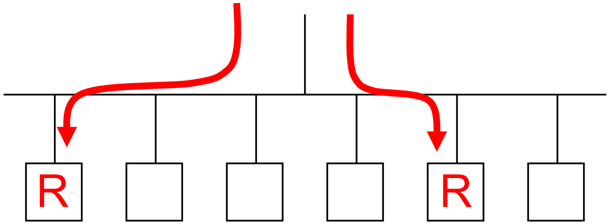
Example of Machine Learning in Cyber Security

Examples of Network Intrusion

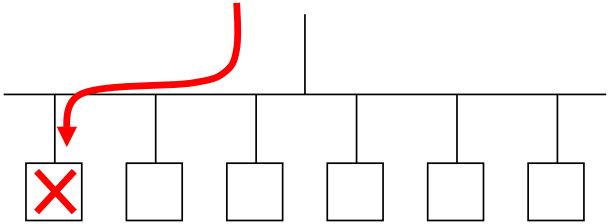
Probe for hosts with known weaknesses



Gain root access to hosts

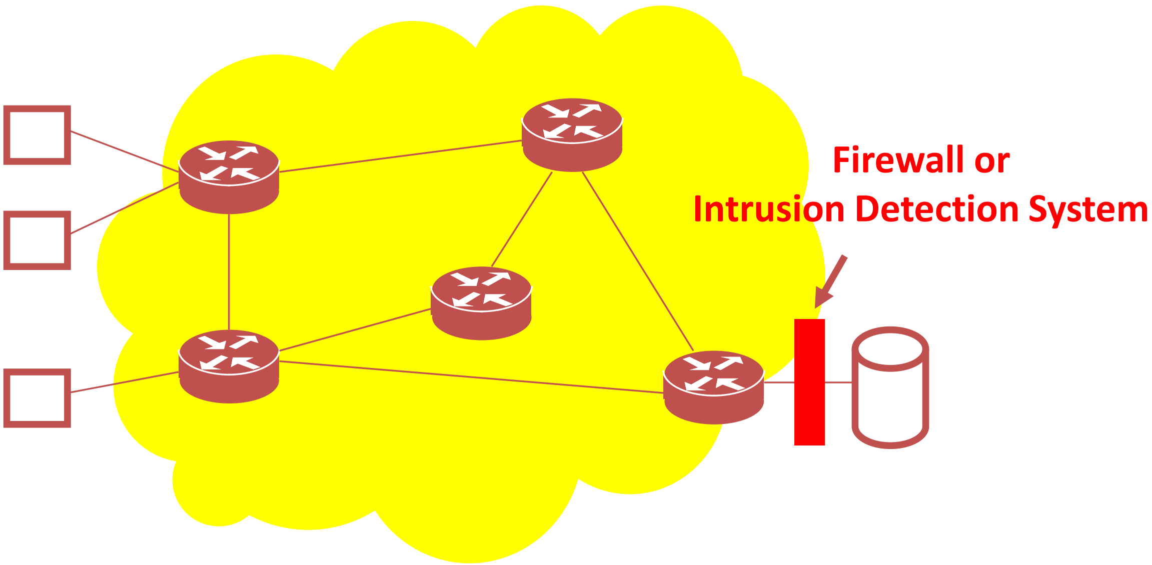


Denial-of-service attack using malformed packets



Existing Approaches to Defend Against Network Attacks

Write rules to detect *known* types of attack



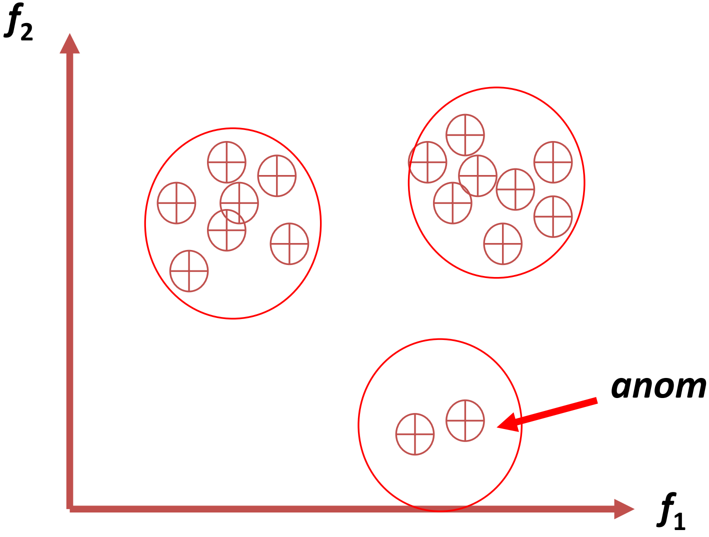
Drawback: unable to detect new attacks

Alternative Approach: Anomaly Detection

- Learn a model of “normal” traffic
- Use this model to test new traffic for anomalies
- Any anomalies are treated as an attack

Cluster-based Anomaly Detection

- **Map network connections into a feature space {f_1… f_k}**
- **Cluster similar connections**
- **Use large clusters to represent normal traffic**



Challenge: changing traffic patterns cause false alarms

Summary

How would you define Artificial Intelligence (AI)?

What are some example applications of AI?

What is the difference between supervised
and unsupervised learning in data mining / machine learning?

What are some example applications of data mining?

What is anomaly detection?

How can anomaly detection be used in network security?