Detección de Anomalías

(Anomaly detection)

© Juan Carlos Cubero

UNIVERSIDAD
DE GRANADA

Motivation and Introduction

- Motivation and Introduction
- Supervised Methods
- Semisupervised Methods
- Unsupervised Methods:
  - Graphical and Statistical Approaches
  - Nearest neighbor based approaches
  - Clustering-based
- Evaluation

Motivation and Introduction

Data Analysis/Data mining: Process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision making

→ Common oriented: Find patterns, trends, etc

→ Uncommon oriented: Identify anomalies

What are **outliers/anomalies**?

The set of data points that are considerably different than the remainder of the data

Working assumption: There are considerably more "normal" observations than "abnormal" observations (outliers/anomalies) in the data

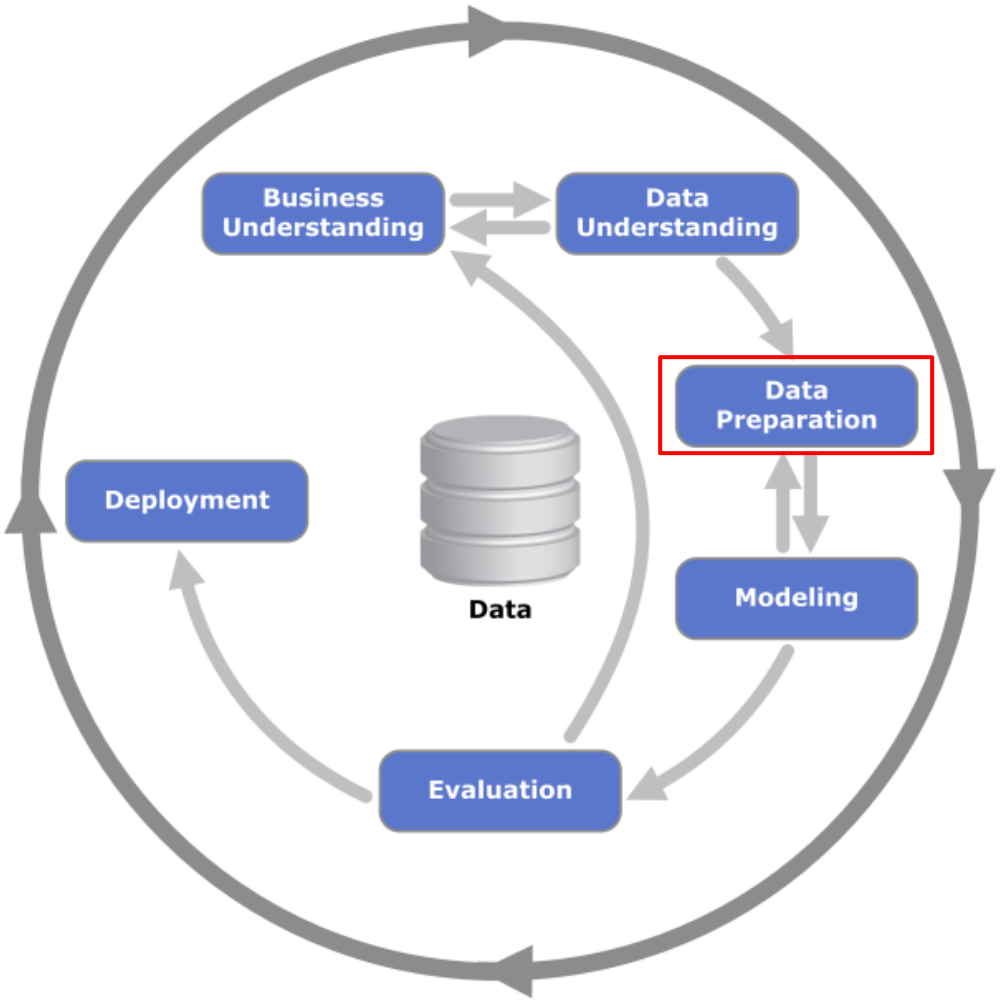
What to do with an anomaly?

Motivation and Introduction

What to do with an anomaly?

They may represent errors when capturing data. So, they should be removed from posterior studies

They may represent real data. So, they should be carefully analyzed

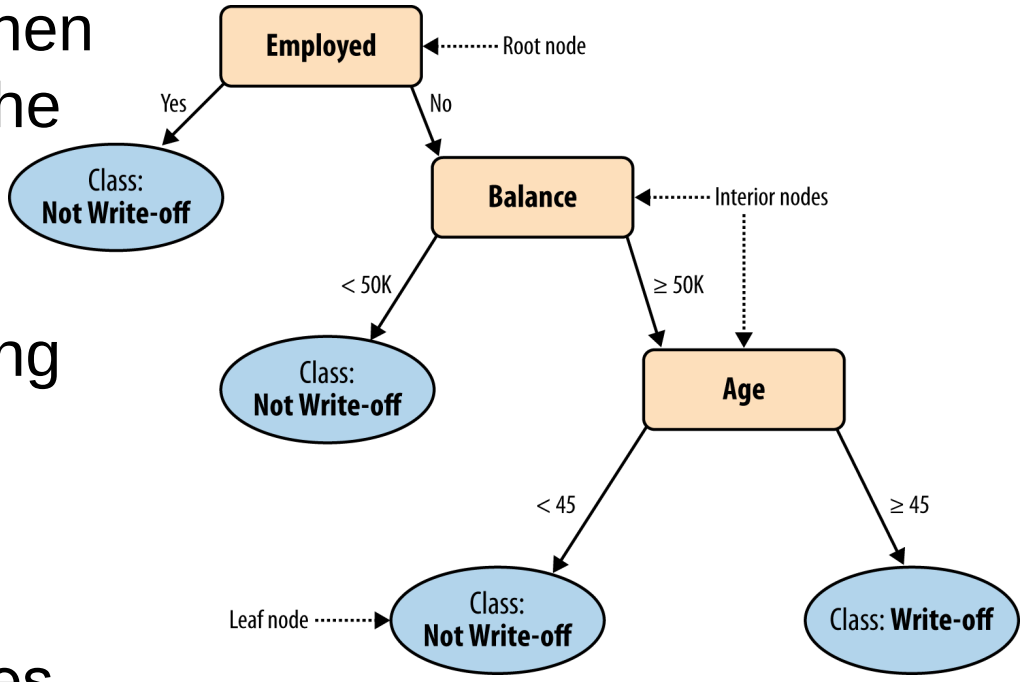


Motivation and Introduction

What to do with an anomaly?

Some techniques are quite robust to anomaly data, so it's presence is not so important:

- Classification methods, e.g, C4.5 behaves quite well when constructing the tree with the presence of anomalous values in interior nodes.
- They are good for classifying instances of "majority" classes.
- But they are not good for classifying "minority" classes

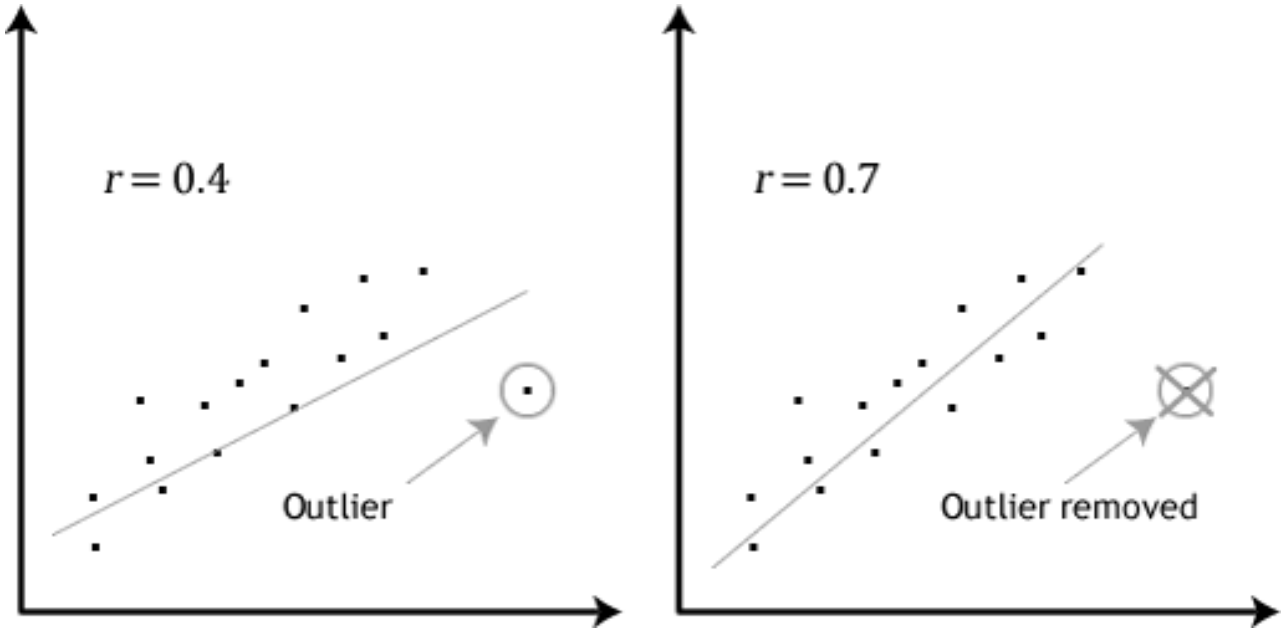


Motivation and Introduction

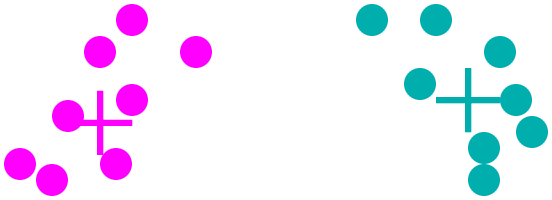
What to do with an anomaly?

- Association rules methods are based on support and confidence measures, which are robust to “exceptions”

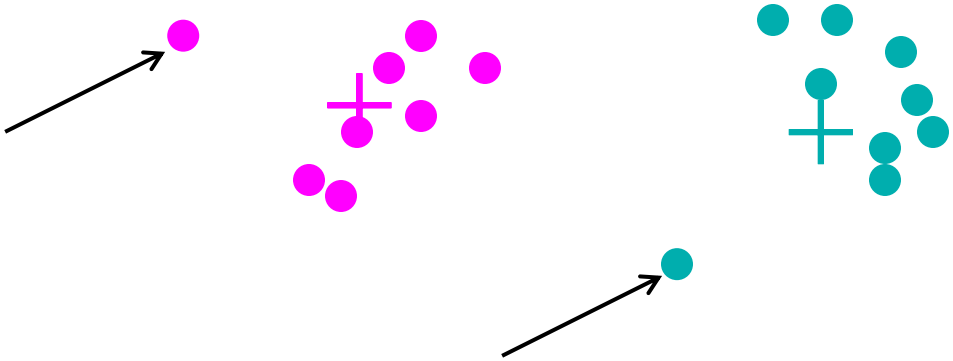
- Regression, Clustering, hypotheses tests, and any method based on averages are quite sensitive to abnormal values.



Centroids in a clustering with no abnormal data



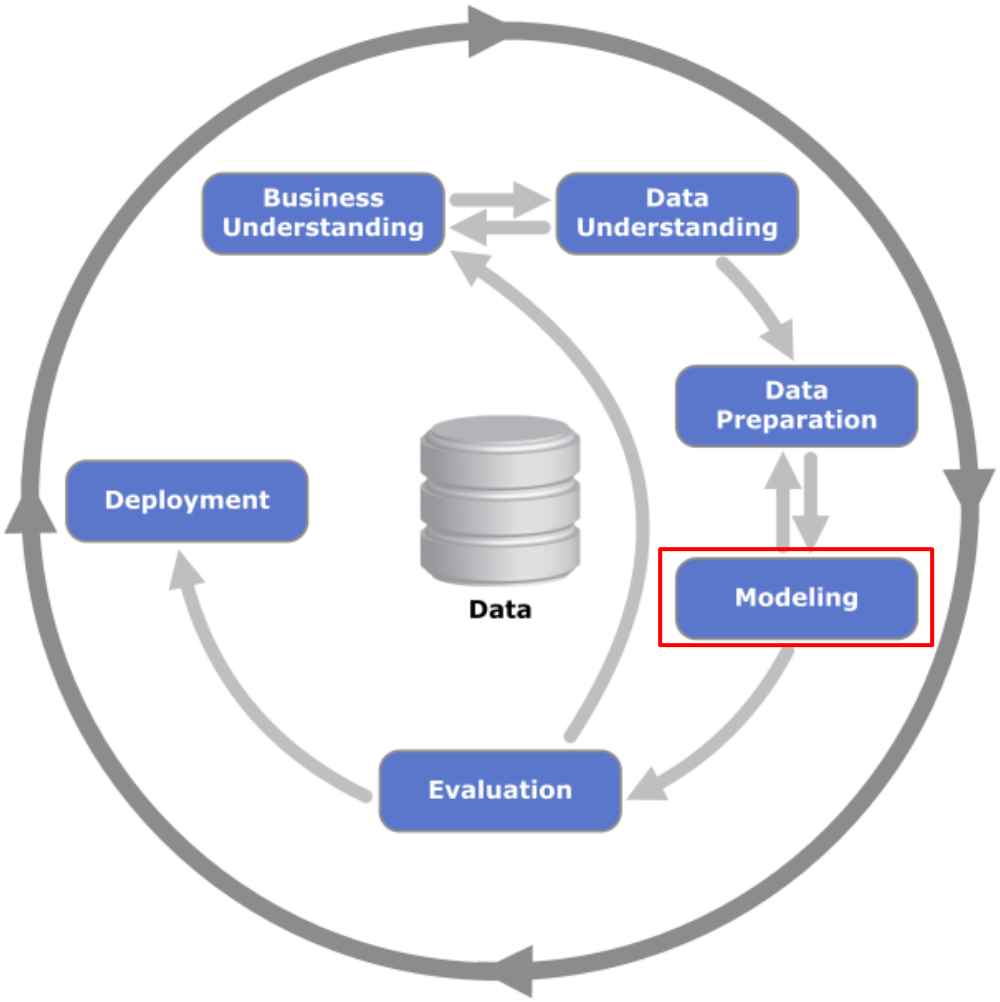
Centroids in a clustering with abnormal data. In this case, data with abnormal combinations of two variables.



Motivation and Introduction

What to do with an anomaly?

Some times we are explicitly interested on detecting anomalies because of the information they convey



Bacon, writing in *Novum Organum* about 400 years ago said:

"Errors of Nature, Sports and Monsters correct the understanding in regard to ordinary things, and reveal general forms. For whoever knows the ways of Nature will more easily notice her deviations; and, on the other hand, whoever knows her deviations will more accurately describe her ways."

Motivation and Introduction

What's the problem to tackle?

Finding a needle in a haystack

is not a correct phrase to refer to the problem of finding anomalies because I know what a needle looks like

What's the problem to tackle?

Do I know what I have to find?

I know what I have to find

I have a complete and accurate description of the anomalous entity to be found

I don't know what I have to find

An anomaly is an abnormal entity

Do I know what I have to find?

Example:

Network Intrusion Detection Systems (NIDS)

- NIDS Signature based:

  I know what I have to find.

- NIDS Anomaly detection based:

  I don't know what I have to find.

NIDS Signature based:

- I know what I have to find
- The system maintains a collection of known signatures (attacks) and compare them with the network data streams

Blaster virus

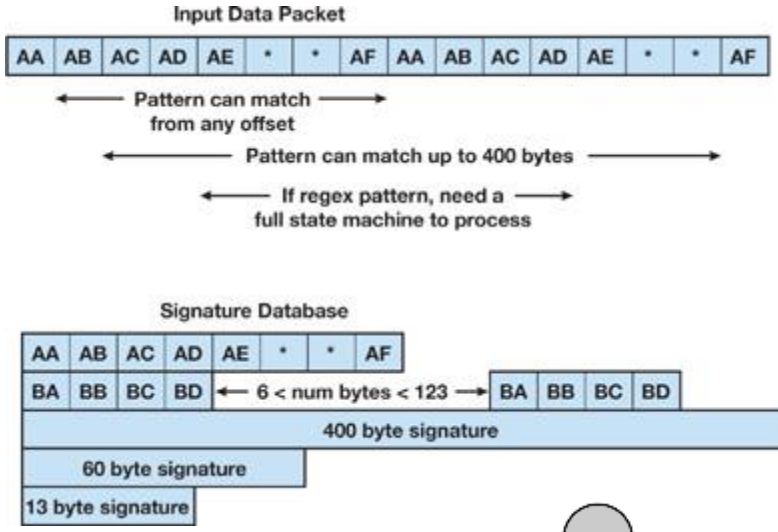
Image Copyright © F-Secure Corporation

Bagle virus

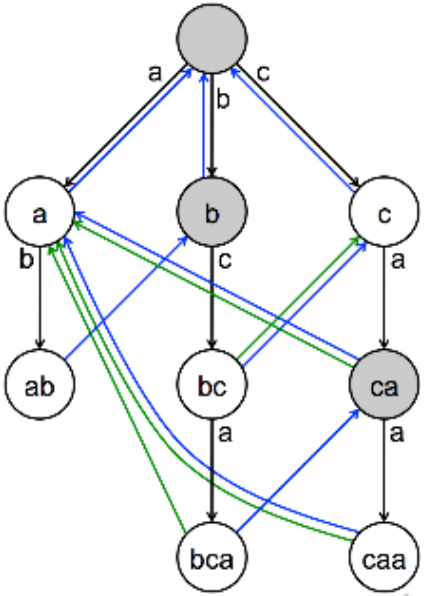
```
1   lea edi, ptr [ebp+0x4025]   // edi = mem[ebp+...]
2   mov ecx, 0x3ec5             // ecx = 0x3ec5
3   mov edx, 0xef4013a0         // edx = 0xef4013a0
    loop:
4   mov al, byte ptr ds[edi]    // al = mem[ds+edi]
5   sub al, dl                  // al = al - dl
6   sub al, dh                  // al = al - dh
7   xor al, cl                  // al = al ^ cl
8   rol edx, cl                 // rotate edx by cl bits
9   mov byte ptr ds[edi], al    // mem[ds+edi] = al
10  inc edi                     // edi = edi + 1
11  dec ecx                     // ecx = ecx - 1
12  jnz loop                    // jump
13  push edi                    // push args into stack
14  call 0x7c92a950             // call a lib func
```

NIDS Signature based:

- I know what I have to find.
- Main problem: string matching



- Aho-Corasick algorithm (and improvements)
  - A finite state machine is constructed
  - The dictionary is the virus database
  - A new entry is parsed



NIDS Signature based:

- I know what I have to find.

NIDS Anomaly detection based:

- I don't know what I have to find.

New types of attacks have not a signature.
An attack is an ***abnormal*** entry

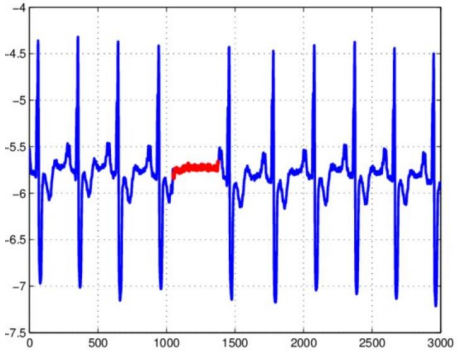
We'll focus on this kind of problems

Motivation and Introduction

- Network Intrusions
  - A web server involved in *ftp* traffic

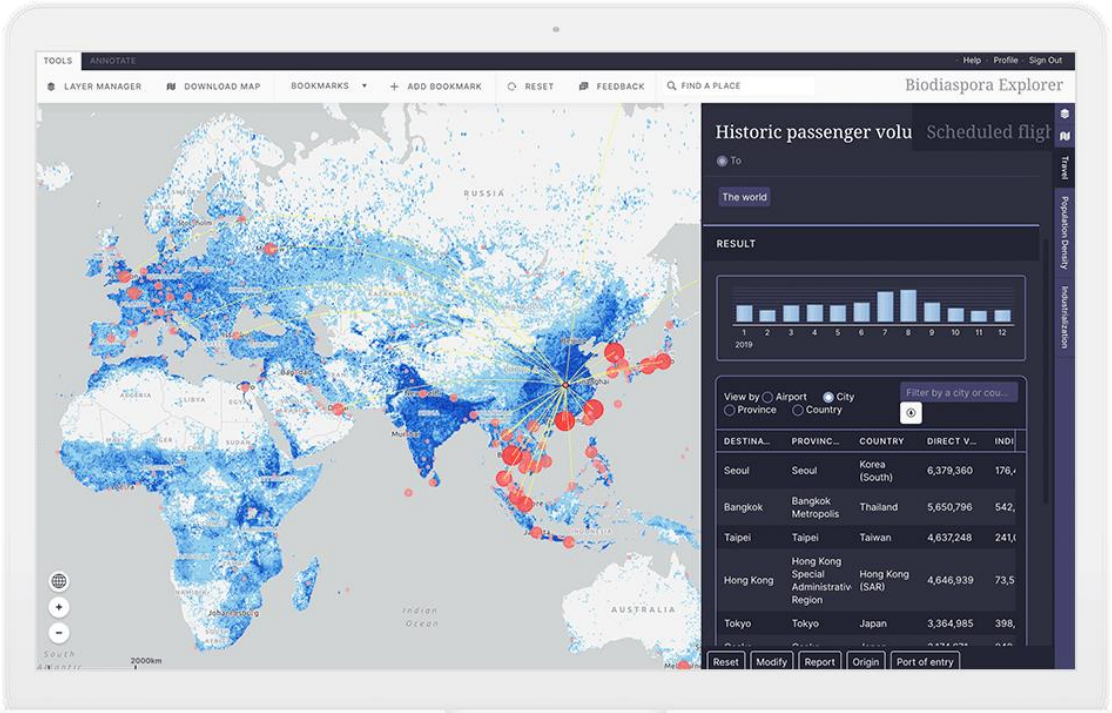
- Credit Card Fraud
  - An abnormally high purchase made on a credit card

- Anomalous patient records



Motivation and Introduction

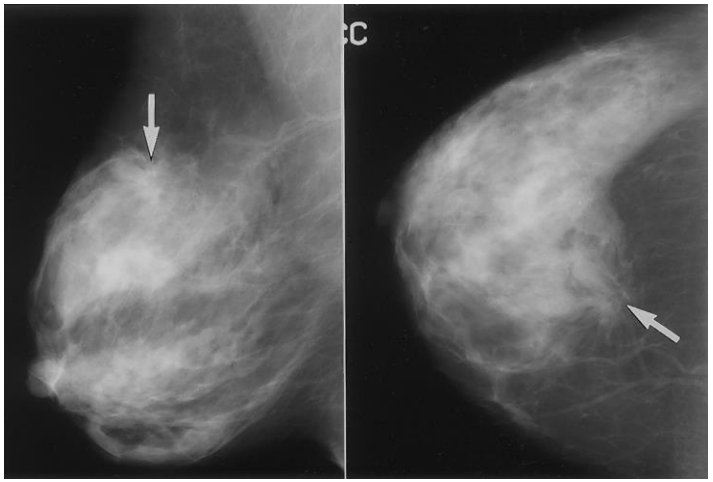
- Disease Outbreak detection



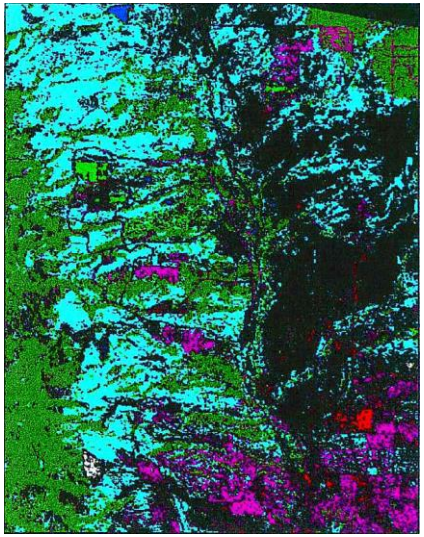
1st

BlueDot was among the **first in the world** to identify the emerging risk from COVID-19 in Hubei province and notify our clients via our Insights platform.

- Anomalous regions within an image
  - mammography image analysis



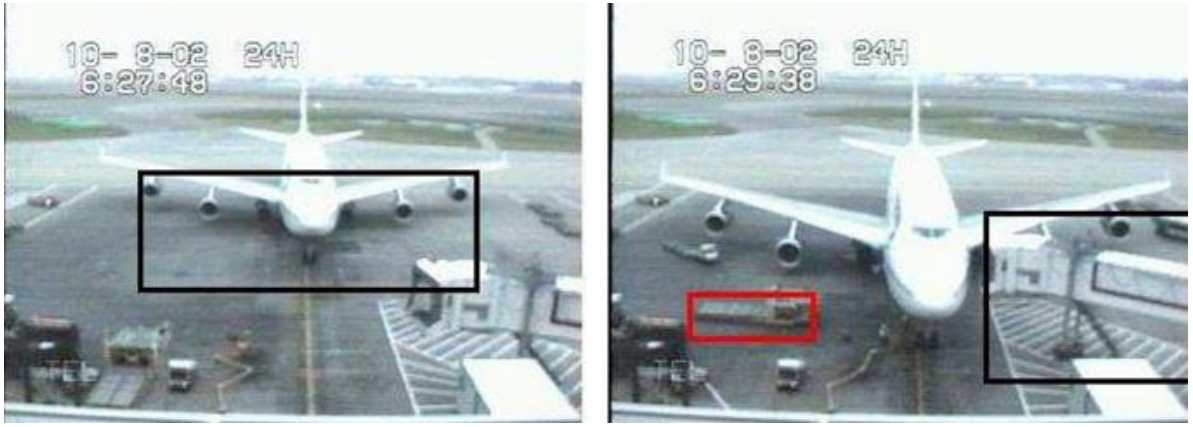
- satellite image analysis



Anomalies in red

- Detecting outliers in a image monitored over time

- Suspicious events in video surveillance



- Nature of input data
- Availability of supervision
- Output of anomaly detection
- Type of anomaly: point, contextual, structural
- Evaluation of anomaly detection techniques

- Tabular data.
- More complex data require adaptation or specific techniques:
  - Sequential data
  - Graph data
  - Spatial data
  - Time series

Motivation and Introduction

Key questions:

Do I have anomalies in my training set?

Do I know which are the anomalies in my training set?

Supervised Methods →

I have anomalies in my training set and they are labelled

A classification model (including the anomaly class) is built.

Tid	SrcIP	Start time	Dest IP	Dest Port	Number of bytes	Attack
1	206.135.38.95	11:07:20	160.94.179.223	139	192	No
2	206.163.37.95	11:13:56	160.94.179.219	139	195	No
3	206.163.37.95	11:14:29	160.94.179.217	139	180	No
4	206.163.37.95	11:14:30	160.94.179.255	139	199	No
5	206.163.37.95	11:14:32	160.94.179.254	139	19	Yes
6	206.163.37.95	11:14:35	160.94.179.253	139	177	No
7	206.163.37.95	11:14:36	160.94.179.252	139	172	No
8	206.163.37.95	11:14:38	160.94.179.251	139	285	Yes
9	206.163.37.95	11:14:41	160.94.179.250	139	195	No
10	206.163.37.95	11:14:44	160.94.179.249	139	163	Yes

Motivation and Introduction

Key questions:

Do I have anomalies in my training set?

Do I know which are the anomalies in my training set?

SemiSupervised Methods →

I do not have anomalies in my training set

Tid	SrcIP	Start time	Dest IP	Dest Port	Number of bytes	Attack
1	206.135.38.95	11:07:20	160.94.179.223	139	192	No
2	206.163.37.95	11:13:56	160.94.179.219	139	195	No
3	206.163.37.95	11:14:29	160.94.179.217	139	180	No
4	206.163.37.95	11:14:30	160.94.179.255	139	199	No
5	206.163.37.95	11:14:32	160.94.179.254	139	19	Yes
6	206.163.37.95	11:14:35	160.94.179.253	139	177	No
7	206.163.37.95	11:14:36	160.94.179.252	139	172	No
8	206.163.37.95	11:14:38	160.94.179.251	139	285	Yes
9	206.163.37.95	11:14:41	160.94.179.250	139	195	No
10	206.163.37.95	11:14:44	160.94.179.249	139	163	Yes

Motivation and Introduction

Key questions:

Do I have anomalies in my training set?

Do I know which are the anomalies in my training set?

Tid	SrcIP	Start time	Dest IP	Dest Port	Number of bytes	Attack
1	206.135.38.95	11:07:20	160.94.179.223	139	192	No
2	206.163.37.95	11:13:56	160.94.179.219	139	195	No
3	206.163.37.95	11:14:29	160.94.179.217	139	180	No
4	206.163.37.95	11:14:30	160.94.179.255	139	199	No
5	206.163.37.95	11:14:32	160.94.179.254	139	19	Yes
6	206.163.37.95	11:14:35	160.94.179.253	139	177	No
7	206.163.37.95	11:14:36	160.94.179.252	139	172	No
8	206.163.37.95	11:14:38	160.94.179.251	139	285	Yes
9	206.163.37.95	11:14:41	160.94.179.250	139	195	No
10	206.163.37.95	11:14:44	160.94.179.249	139	163	Yes

UnSupervised Methods →

I have anomalies in my training set but they are not labelled
I don't know if a record is an anomaly or not

- The output is a label or a score:

Label:

  - The procedure assigns to each test instance a label: *normal* or *anomaly*
  - This is especially true in classification-based approaches

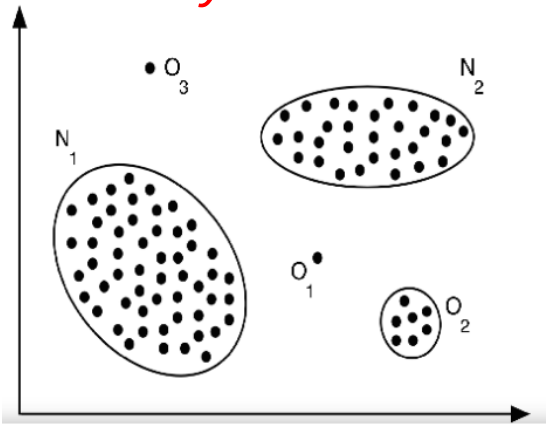
Score:

  - Each test instance is assigned an anomaly score (real number)
    - Allows the output to be ranked
    - Requires an additional threshold parameter

Motivation and Introduction

Output?

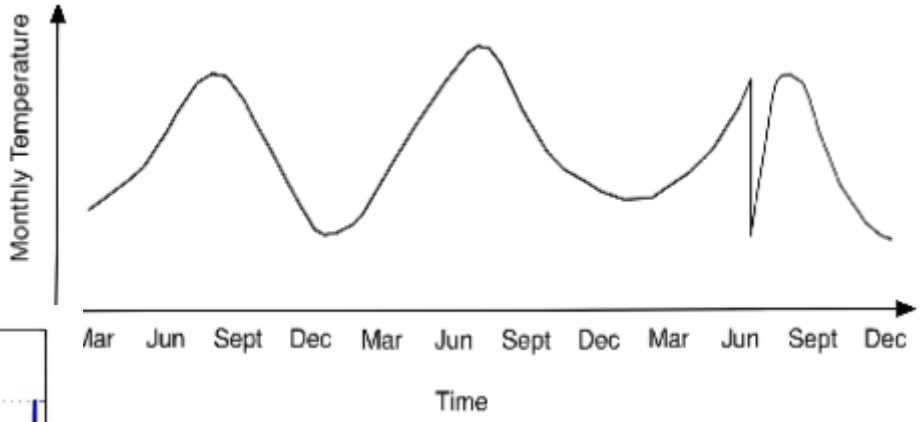
Point anomaly



Contextual Anomaly

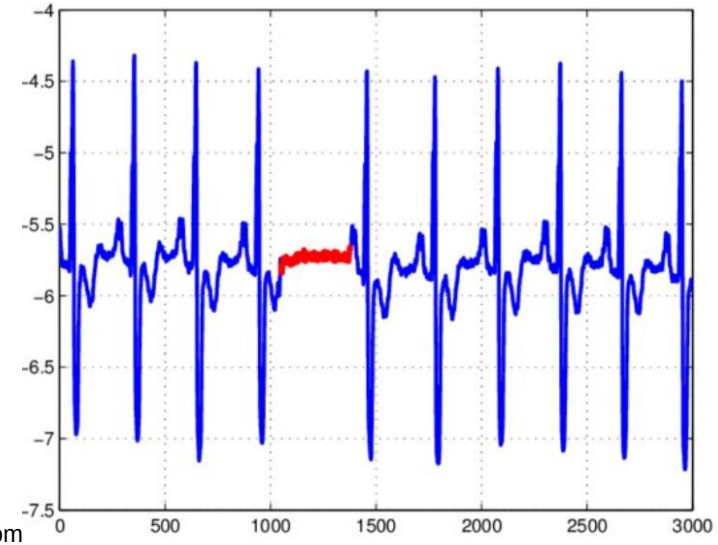
Typical contexts:

Time, location, economical status, etc



Collective Anomaly

Taking into account other data values



A good example to understand the differences:

https://stats.stackexchange.com/questions/323553/difference-between-contextual-anomaly-and-collective-anomaly