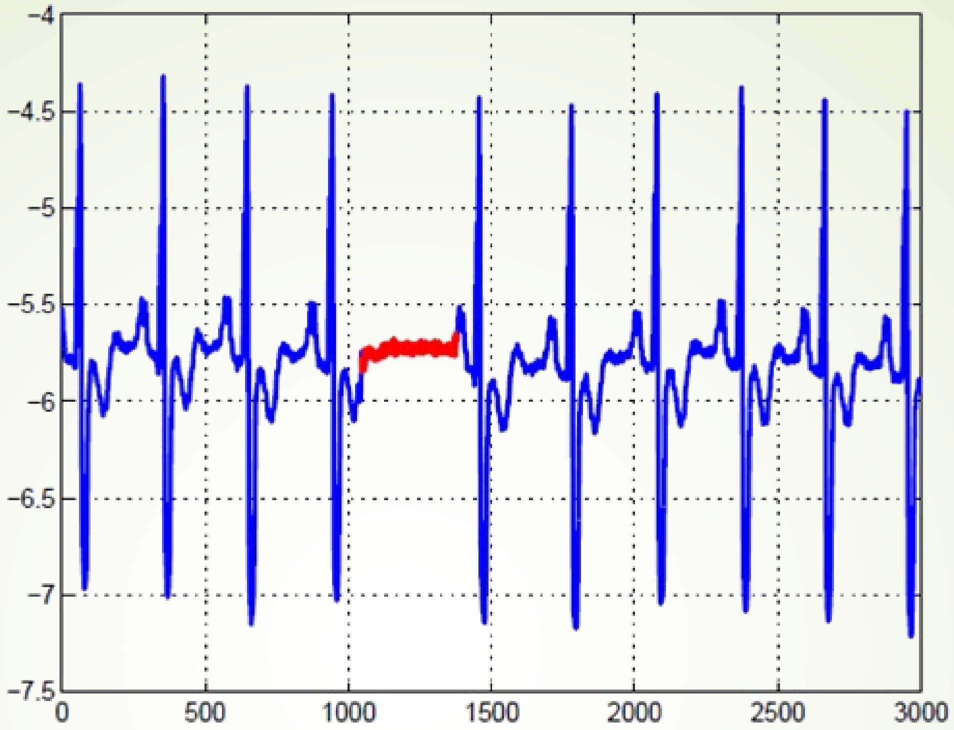
1. A coin was tossed 10 times. Got 7 heads
2. Tossed 100 times. Got 70 heads

Based on Statistical hypothesis testing

is it a fair coin after known 1? And after known 2?



Anomaly detection

Introduction to unsupervised learning – lecture 5

GUY SHTAR: SHTAR@POST.BGU.AC.IL

Based on slides by Prof. Lior Rokach

Agenda

- Introduction
  - Motivation
  - Taxonomy
- Point anomaly
  - Classifying single samples
- Contextual\collective
  - Anomalies in time series

Introduction

Motivation

- Anomaly detection: identification of rare items, events or observations which raise suspicions
- Detecting non-conforming or unexpected patterns in large data volumes is difficult in many application domains
  - many approaches coming from different research fields have been developed

?

Applications

- Fraud & misuse detection
  - credit card fraud
- Insurance claim fraud
- Telecommunication fraud
- Intrusion detection in IT security
- Fault detection in
  - safety critical systems
  - production processes
- Surveillance tasks
- Health care
- Data cleaning

Research areas

- Rule-based Systems
- Statistics
- Statistical Pattern Recognition
- Data Mining
- Machine Learning
- Probabilistic Reasoning
- Information Theory
- Expert Systems

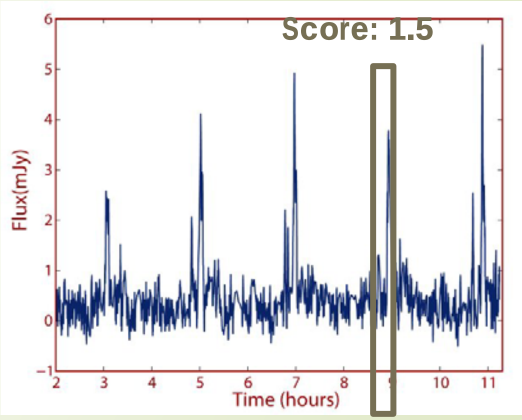
Dimensions of Anomaly Detection

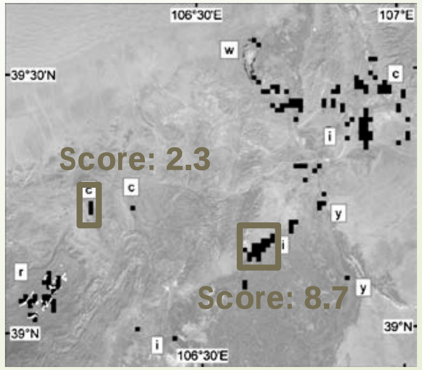
- **Nature of the input data (approach)**
  - Structure (#attributes, data types, relationships…)
  - Labels availability
- **Types of anomalies**
  - Point anomalies
  - Contextual anomalies
    - context (temporal\spatial) imply anomalies
  - Collective anomalies
    - co-occurrences of data records form anomalies
- **Expected output**
  - Scores vs. labels
  - Automation vs. assistance

?

	A	B	C	D	E
1	Year	Financial Period	Region	Salesperson	Sales
2	2008	1	North	Johnson R	12000
3	2008	1	Midlands	Smith A	9000
4	2008	1	Midlands	Herbert H	15100
5	2008	1	North	Jackson S	21000
6	2008	1	South	Newman P	13750
7	2008	1	South	Foster J	10810
8	2008	2	North	Johnson R	8100
9	2008	2	Midlands	Smith A	11430
10	2008	2	Midlands	Herbert H	12450
11	2008	2	North	Jackson S	17050
12	2008	2	South	Newman P	13000
13	2008	2	South	Foster J	7300

Score: 3.5





Anomaly detection approach

- Supervised
  - Requires fully labelled training set
  - Defined as machine learning classical problem
- Semi-supervised
  - Requires labeled training data only for normal (negative) instances
  - Detection of significant derivations from the normal instances in observed data
- Unsupervised
  - Does not require labeled training data at all
  - Assumes that abnormal situations occur rarely and are different in their features compared with normal situations

10

Types of anomalies

- Point anomalies
  - Single instance implies anomaly
  - *Snowden just copied 1M files to a remote server (instance=bulk)*
- Contextual anomaly
  - Context specific. Common in time series
  - *Snowden copied 1000 files on chrisms*
- Collective anomaly
  - A set of instances imply anomaly
  - *Copied 1,000 files every day in the last year*

?

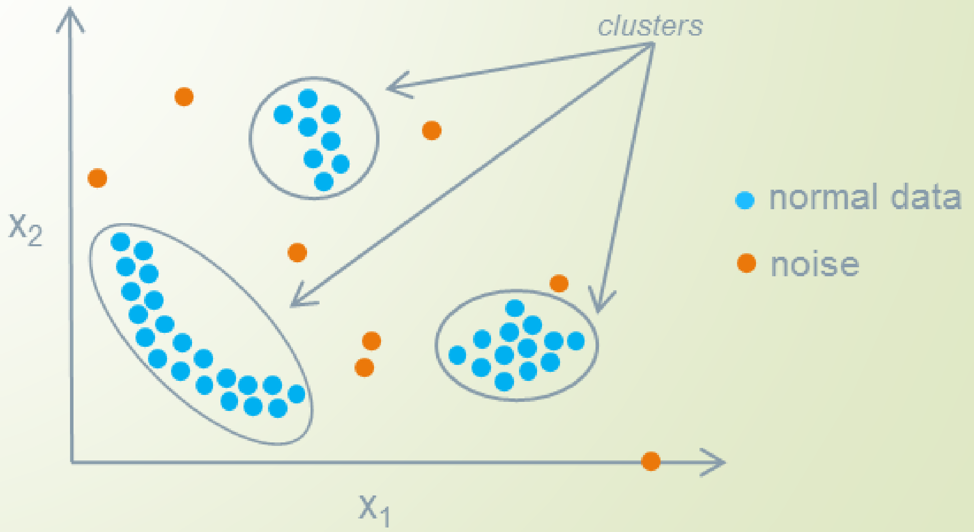
12

Notebook

Get to know our data for today

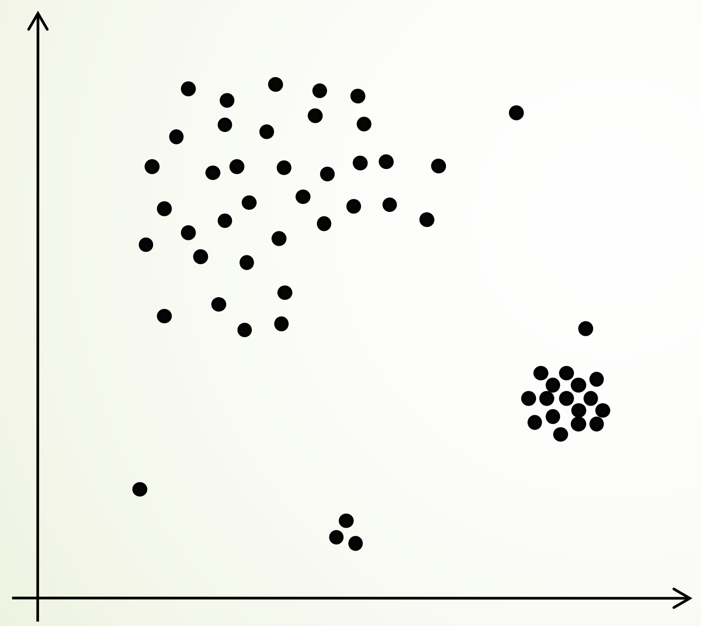
Point anomaly

Single instance implies anomaly



Point Anomaly Detection, Challenges

Point anomaly detection



Challenges

- High dimensional problem spaces
- Heterogeneous data types
- High data volumes
- Varying densities
- Microclusters
- Irrelevant attributes and noise
- Acquisition and exploitation of suitable training data

Many algorithms for identifying point anomalies have been developed.
Even if they are conceptionally similar, they apply different techniques.

Anomaly Detection Algorithms

- Based on the underlying principles, anomaly detection algorithms can be grouped into the following categories ([Hodge & Austin 2004], [Chandola et al. 2009]):
  - **Classification-based**
  - **Nearest Neighbor-based**
  - **Clustering-based**
  - Spectral-based
  - Probabilistic
  - Information theoretic
- These categories are not completely unique
  - E.g. Bayesian Networks or Decision Trees are traditional classification algorithms but have strong statistical / probabilistic foundations

16

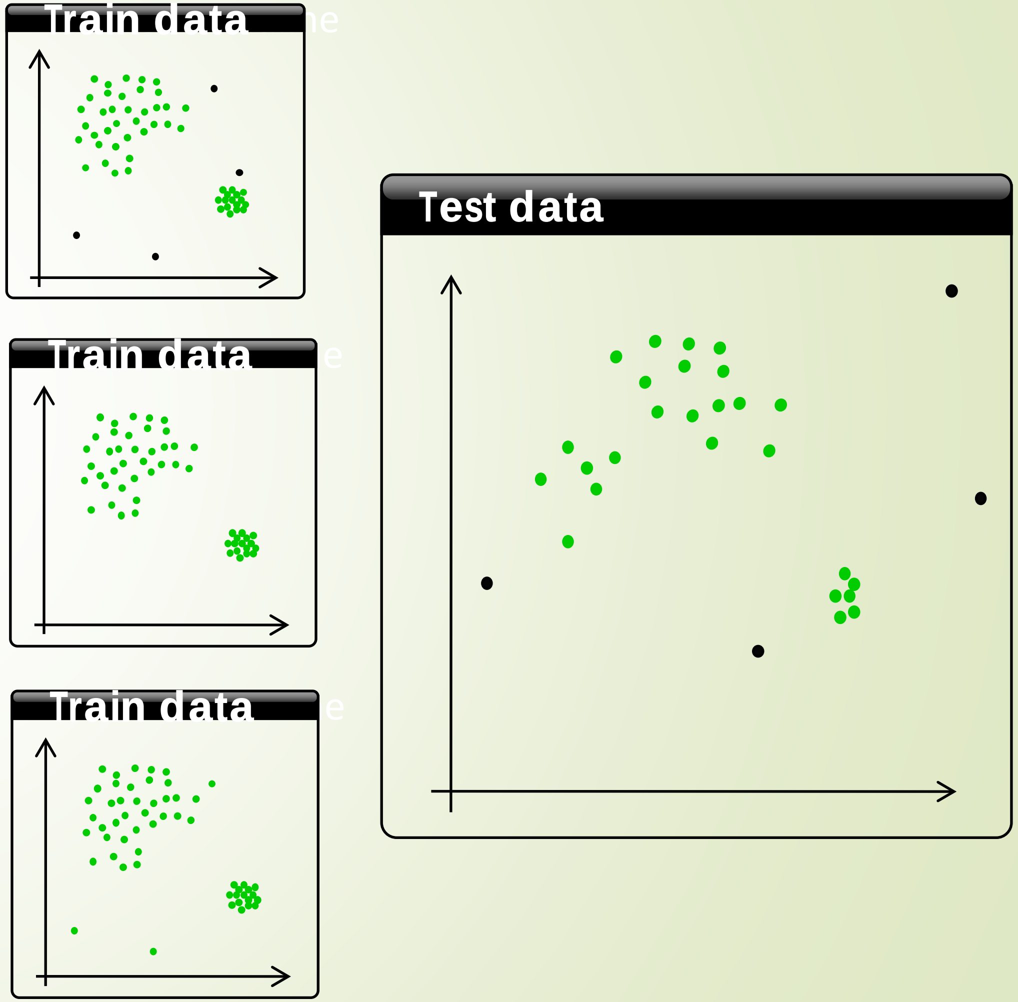
Point anomaly

Classification-based

Classification-based

- Classification-based methods - labels are known
  - ?
- Single class (semi-supervised learning)
  - Deep SAD
- Unsupervised
  - Isolation forest
  - One class SVM
  - One class neural network (2020)

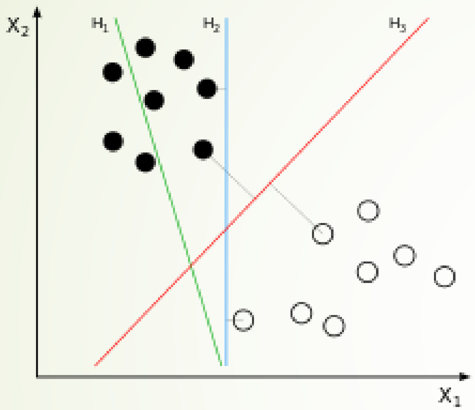
?



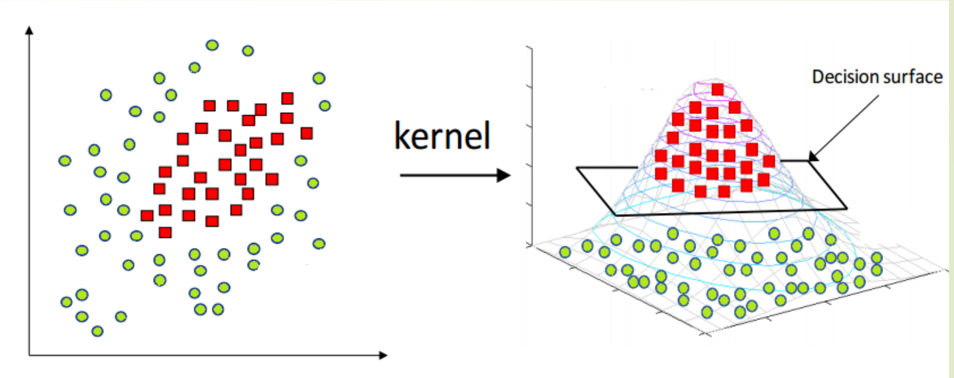
Classification-based Overview

- Use of standard Machine Learning algorithms, e.g.
  - An improved ensemble-based intrusion detection technique using XGBoost [Bhati et al. 2020]
  - Neural Networks [Moreau et al.1997]
  - Support Vector Machines [Ma & Perkins 2003]
  - Decision Trees [Reif et al 2008]
  - Rule-Based Systems [Rosset et al.1999], [Fawcett & Provost 1997]
- Advantages
  - May achieve high accuracy if accurate training data is available
  - Once the model is learnt, anomaly detection can be done very fast
    - Complexity of model generation varies depending on the algorithm
- Drawbacks
  - Requires labeled data
  - Not all algorithms are able to provide anomaly scores but only labels

SVM



H1 does not separate the classes. H2 does, but only with a small margin. H3 separates them with the maximal margin.

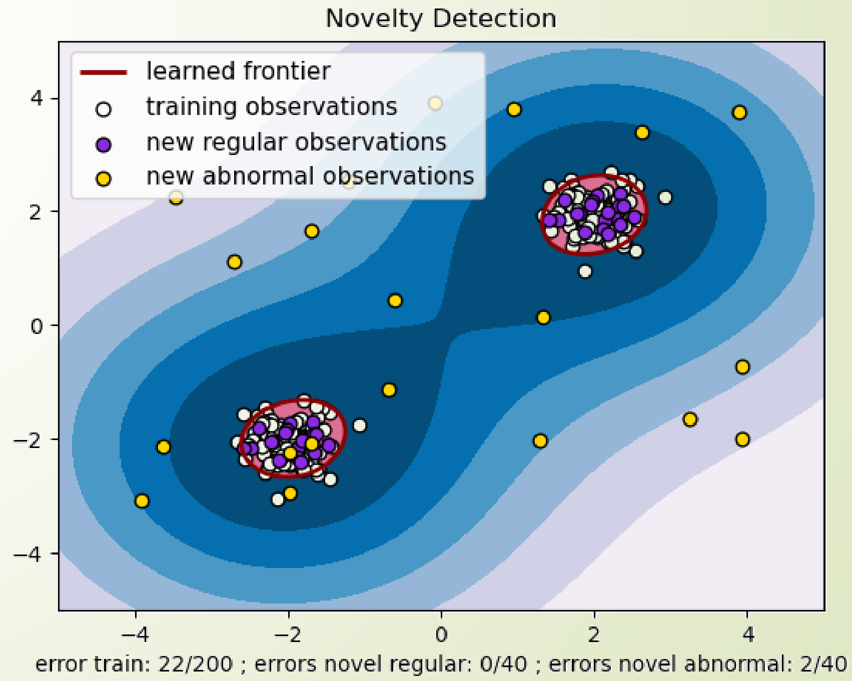


Using a kernel function improves results

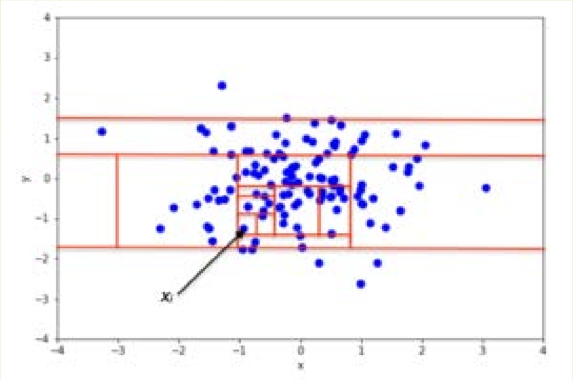
- One-class SVM has just one class

One class SVM

- Scholkopf et al. 2000. Link
- Unsupervised algorithm that learns a decision function
- Requires selecting a kernel
- **v (Greek nu),** the margin variable, corresponds to the probability of finding a new, but regular, observation outside the frontier



Isolation forest (unsupervised)

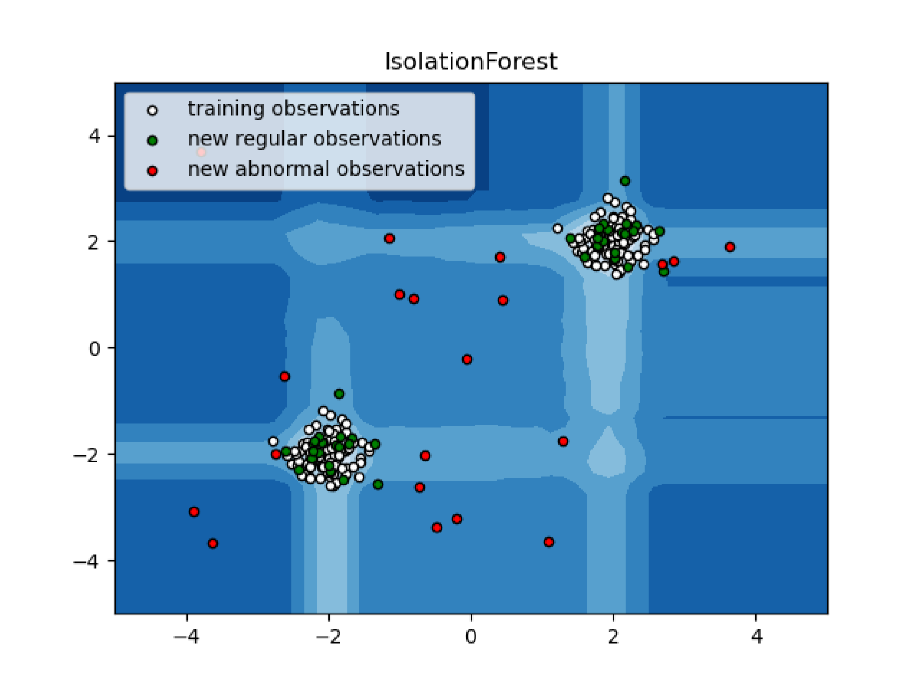


- Liu et al. 2008. Link
- Works by isolating anomalies and not profiling them
  - Isolates anomalous points in the dataset
- Anomaly assumption:
  - Few - they are the minority consisting of fewer instances and
  - Different - have attribute-values very different from those of normal instances
- Algorithm:
  - Randomly selecting an attribute, randomly selecting split value between the minimum and maximum of it. Repeat for *n* trees.

?

Isolation forest

- Shorter path length to leaf -> anomaly
- The path length is divided by the average expected path to find an anomaly score.



Notebook

One class SVM & Isolation

Point anomaly

Nearest Neighbor-Based

Anomaly Detection Algorithms. Nearest Neighbor Based.

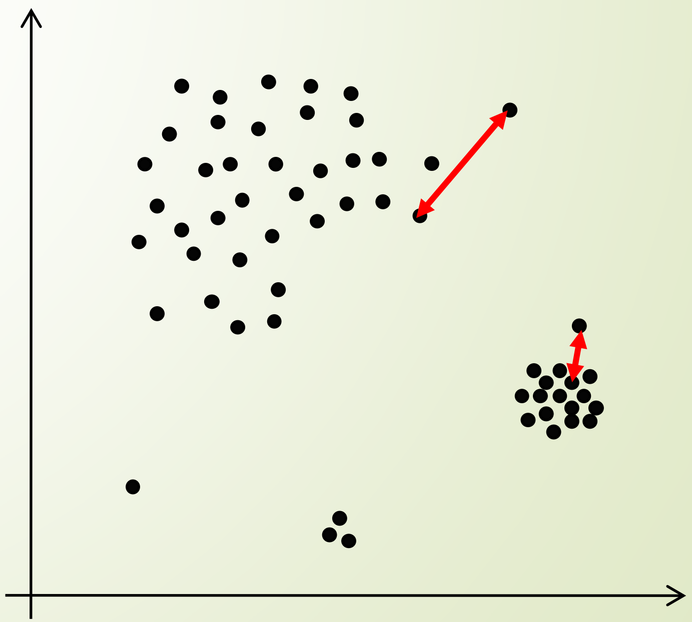
Assumption

- Normal data instances occur in dense neighborhoods, while anomalies occur far from their closest neighbors.

Basic approaches

- **Distance-based**
  - Distance to the k-th nearest neighbor determines an anomaly score
- **Density-based**
  - Ratio between the local density of the instance to the local density of the k-th nearest neighbors determines an anomaly score

Illustration



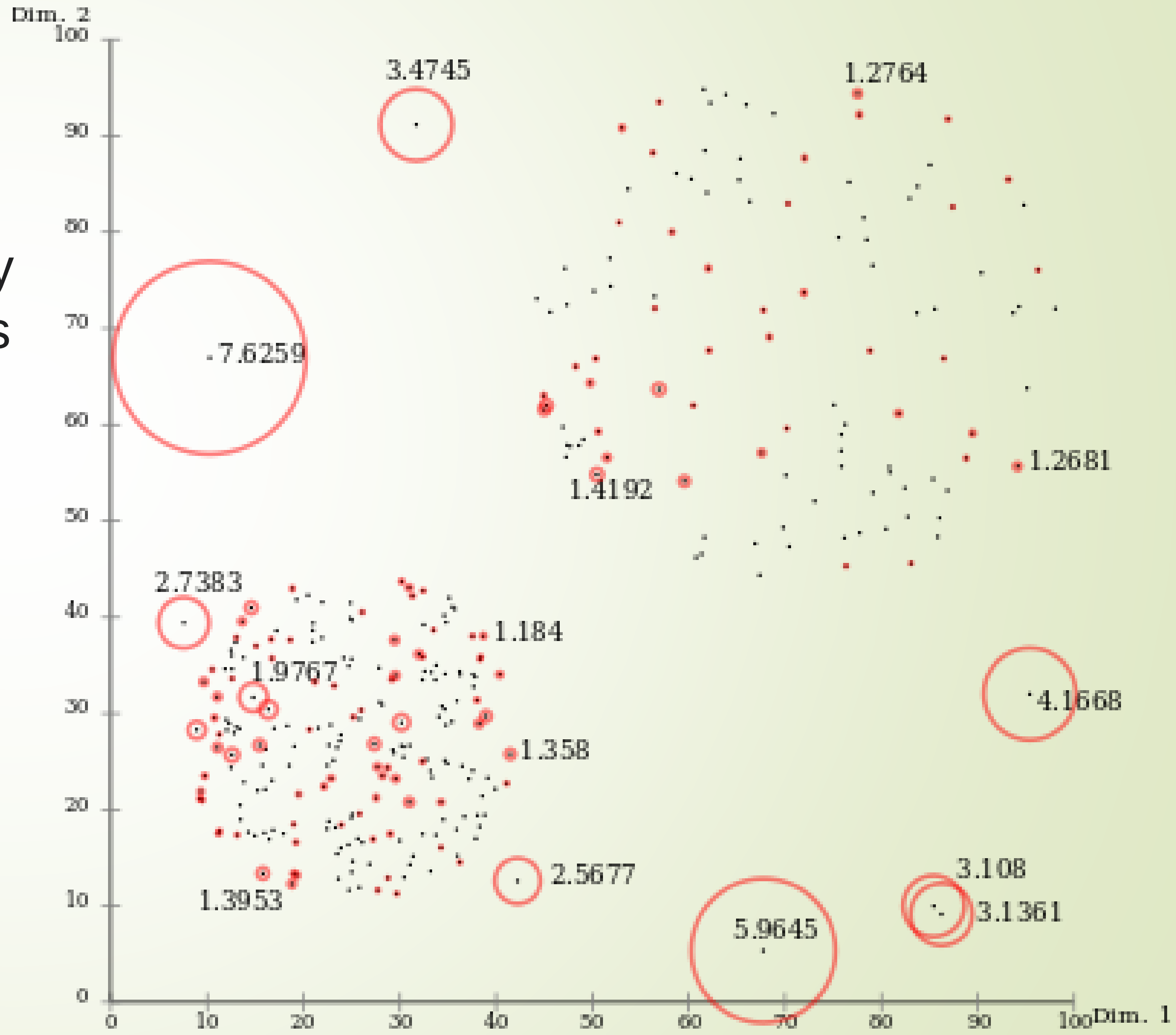
Anomaly Detection Algorithms. Nearest Neighbor Based.

- Many Variants exist
  - Different score calculations, e.g.
    - *Local Outlier Factor (LOF)* [Breunig et al. 2000]
    - *Connectivity-based Outlier Factor* (COF) [Tang et al.]
  - Different distance measures
- Advantages
  - Applicable for unsupervised anomaly detection
  - Pure data driven, i.e. do not make any assumptions regarding the underlying distributions
  - Flexible through the definition of domain specific distance measures
- Drawbacks
  - Computational expensive
  - May have problems to deal with microclusters correctly
  - Defining distance metrics, different data types

Local Outlier Factor (2000)

- Basic idea of LOF: comparing the local density of a point with the densities of its neighbors
  - Density is based on k-distance = distance to the k'th neighbor
- LOF scores as visualized by ELKI.

Wikipedia



Point anomaly

Clustering-based

Clustering Based

Assumption

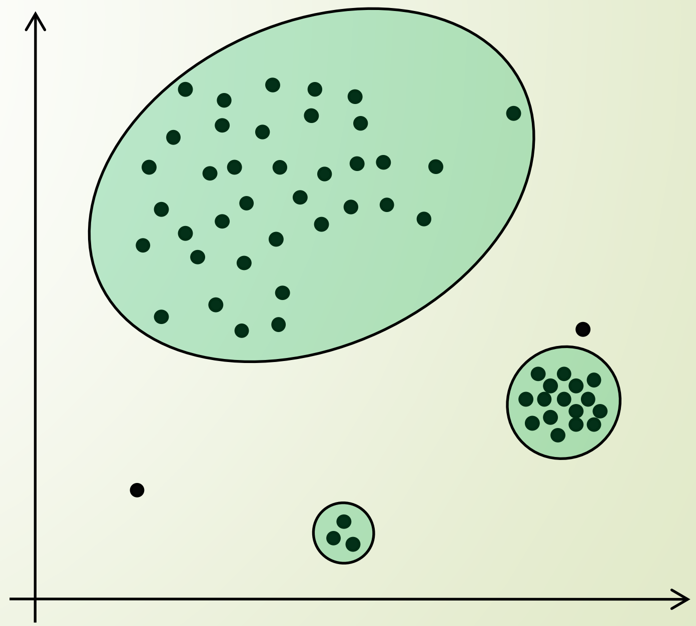
- The result of a clustering algorithm can be used to detect anomalous data records.

?

Basic approaches

- **Approach 1**: Data instances that do not belong to any cluster are considered as anomalies
- **Approach 2**: Properties of the corresponding cluster (size and density) determines the anomaly score
- **Approach 3**: Distance to cluster centroids determines the anomaly score

Illustration



Clustering-Based Overview

- Typical Clustering Algorithms
  - K-means, K-medoids, Self Organizing Maps, etc.
  - Many clustering algorithms require a distance measure
  - Hence, like NN approaches, the definition of a suitable distance measure is also essential here
- Advantages
  - Applicable for unsupervised anomaly detection
  - Once the clustering is computed, anomaly detection is fast
- Drawbacks
  - Runtime performance depends on the specific clustering algorithm
    - $O(n^2)$ for many algorithms
    - Faster algorithms (e.g. k-means) may require the predefinition of the number of clusters

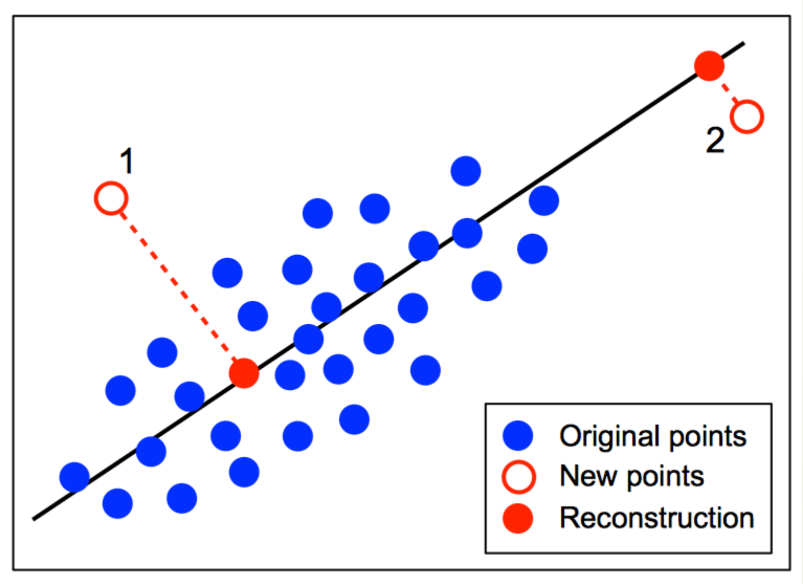
Dimensionality reduction for anomaly detection

- Spectral-based
- Dimensionality reduction can be used to find anomalies

?

Dimensionality reduction for anomaly detection

- Spectral-based
- Dimensionality reduction can be used to find anomalies
  - Reconstruction error is the score



Notebook

LOF & dimensionality reduction

Assumptions & requirements for applying point anomaly detection algorithms

Point Anomaly Detection Algorithms. Assumptions & Requirements.

Assumptions

- **Input data is typically multivariate record data (flatten, tabular)**
  - Data instances can be interpreted as points in a well-defined problem space
- **Distances between individual data instances are well-defined.**
- **Anomalies are individual data records**
  - Individual data instances have to be classified / scored as normal or abnormal
  - Complex dependencies between input data records are not directly considered

Requirements

- If required, transformation of raw input data into multivariate record data, e.g. by
  - database joins
  - feature extraction algorithms
- Specification of underlying data types
- Provided input data has to include all relevant information
- If required, transformation of more complex anomaly detection scenario into a point anomaly detection task
  - Specification of the purpose & target of the anomaly detection task
  - Applications of suitable preprocessing operators

Advantages and Limitations of Point Anomaly Detection Algorithms.

Limitations

- Focus on analysis of record data
- Cannot deal with collective anomalies directly
- Do not explicitly consider time information
- Are not specialized for specific anomaly detection problems like
  - Time series analysis
  - Sequence analysis
  - Graph analysis
  - Image analysis

Advantages

- Most algorithms can be adapted to more complex input data (e.g. graphs, images)
  - By defining a suitable distance measure
- Point anomaly detection algorithms are able to deal with collective anomalies and time-dependent data if appropriate pre-processing is applied
- They are the most generic anomaly detection algorithms
  - Can be applied to semi- and unsupervised scenarios
  - Given suitable data, they do not require deeper domain knowledge

Detecting Contextual and Collective Anomalies

Context specific

A set of instances imply anomaly

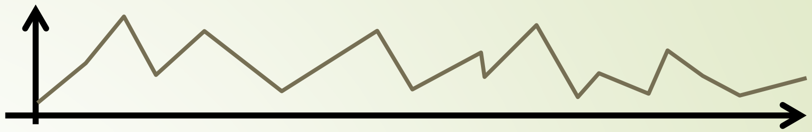
Common in time series

Reminder: Types of anomalies

- Point anomalies
  - Single instance implies anomaly
  - *Snowden copied 1M files to a remote server today*
- Contextual anomaly
  - Context specific. Common in time series
  - *Every day in the past month Snowden have copied 10,000 files*
- Collective anomaly
  - A set of instances imply anomaly
  - *copied 10,000 files each day for a year*

Can try manually converted

Time Series



- A *time series* is a set of observations ordered in time
  - Usually most helpful if collected at regular intervals
- In other words, a sequence of repeated measurements of the same concept over regular, consecutive time intervals

Why time series data different from other data?

- Data are not independent
  - Much of the statistical theory relies on the data being independent and identically distributed
- Large samples sizes are good, but long time series are not always the best
  - Series often change with time, so bigger isn't always better

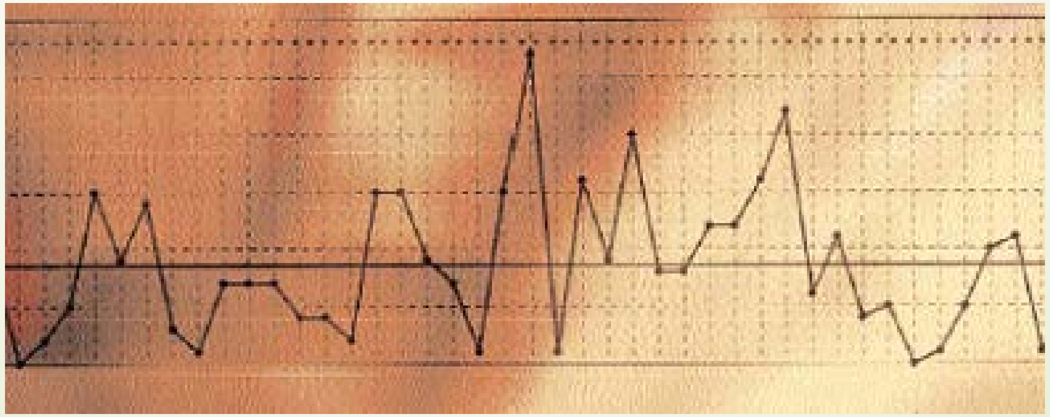
49

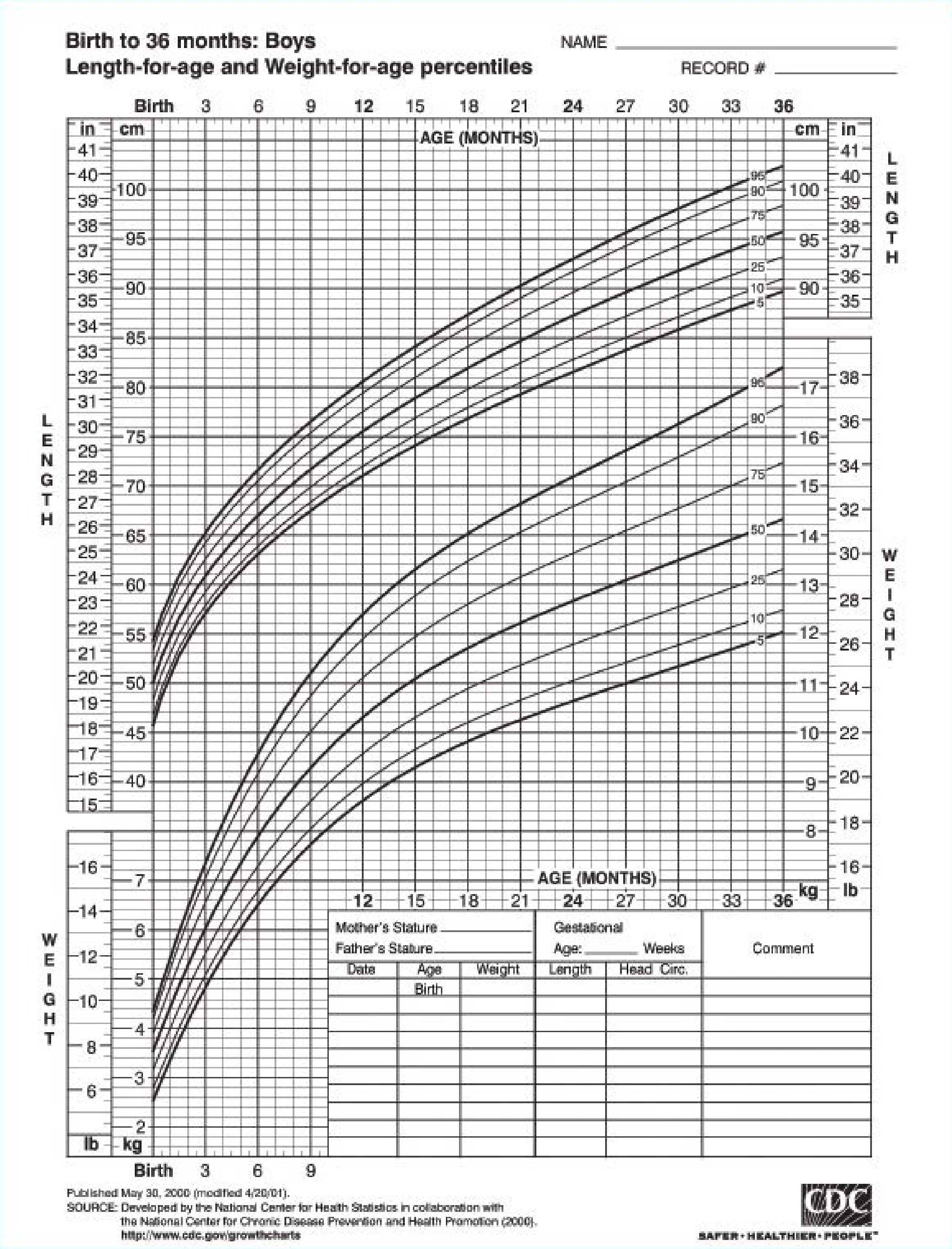
Notebook

Rethinking our Corona data

Control Charts

- Was developed in Bell Labs in the 1920s for improving the reliability of the telephony transmission systems
- Control charts are simple, robust tools for understanding process variability
- If the process is not in control, analysis of the chart can help determine the sources of variation, which can then be eliminated to bring the process back into control

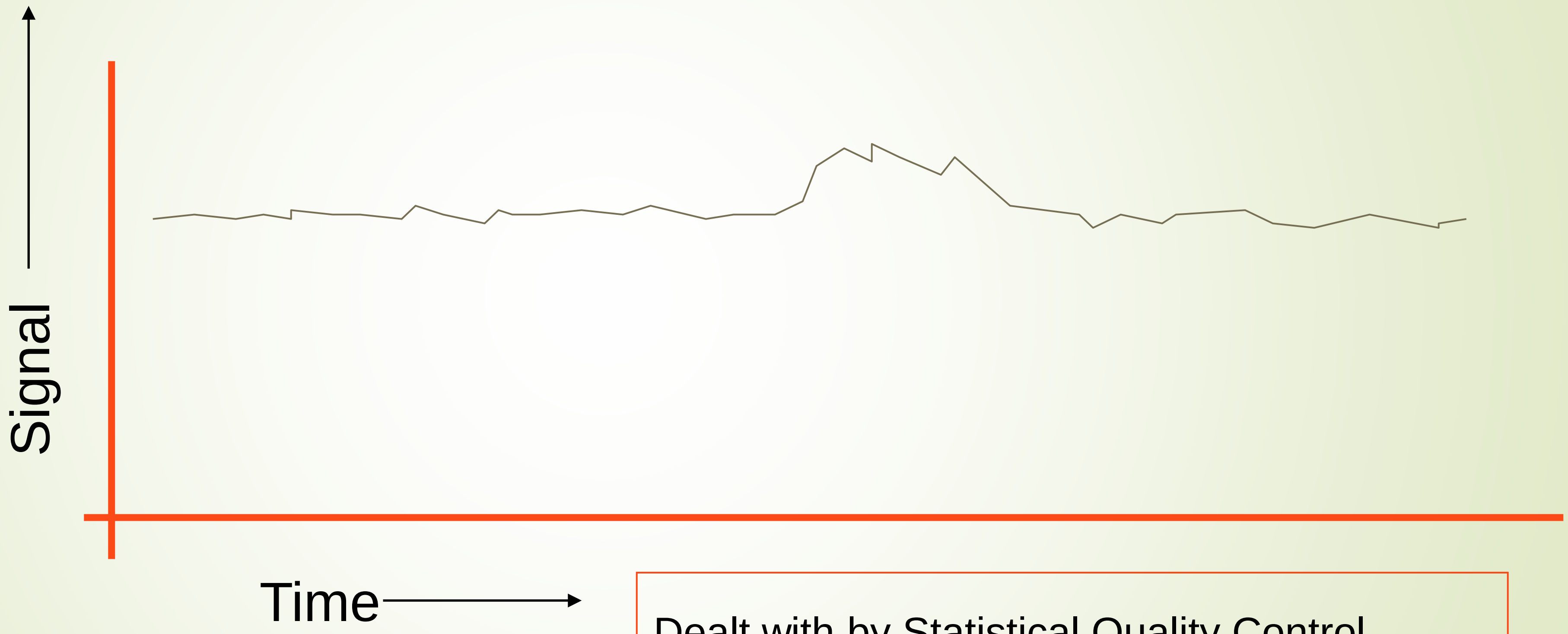



Birth to 36 months: Boys
Length-for-age and Weight-for-age percentiles
NAME
RECORD #
AGE (MONTHS)
LENGTH
WEIGHT
Mother's Stature
Father's Stature
Gestational Age: Weeks
Comment
Date
Age
Weight
Length
Head Circ.
Birth
Published May 30, 2000 (modified 4/20/01).
SOURCE: Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000).
http://www.cdc.gov/growthcharts
CDC
SAFER · HEALTHIER · PEOPLE

?

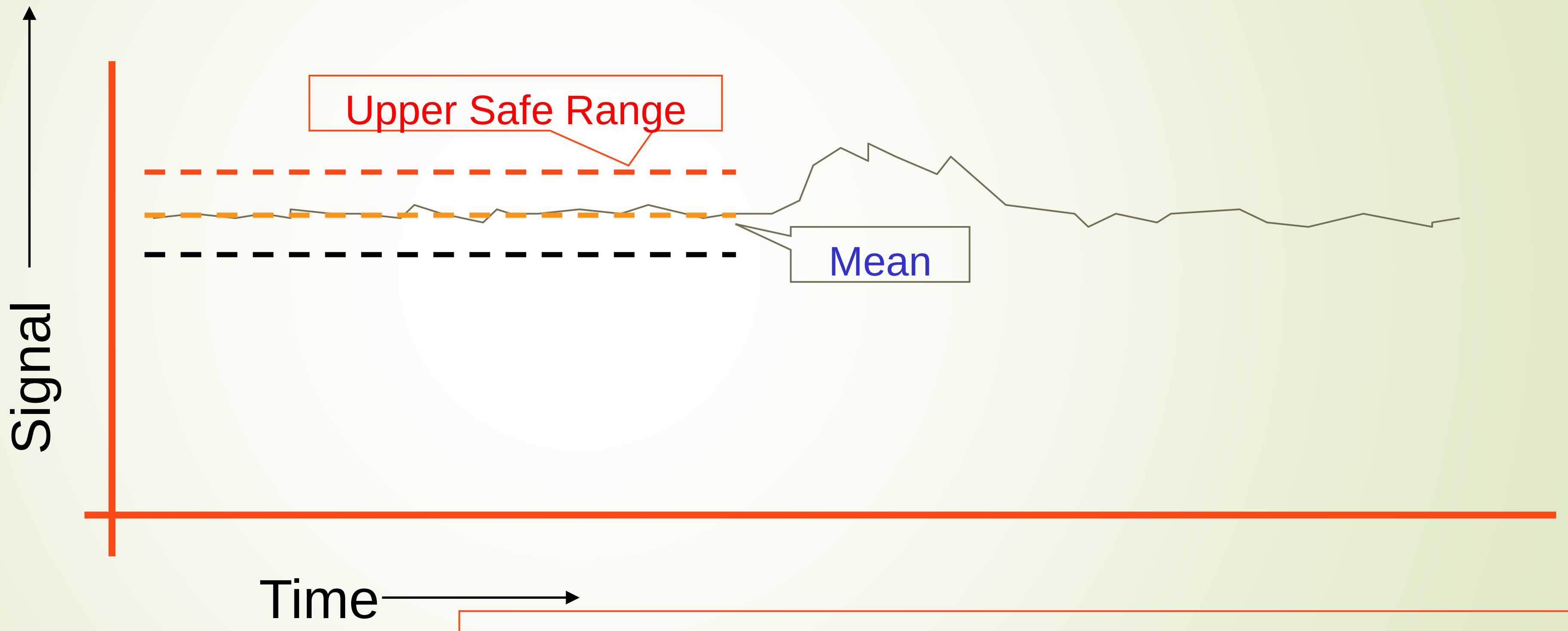
Univariate Time Series

An easy case (stationary)



Dealt with by Statistical Quality Control
Mean and std are constant
Signal an alarm if we go outside $\mu \pm 3\sigma$

An easy case: Control Charts

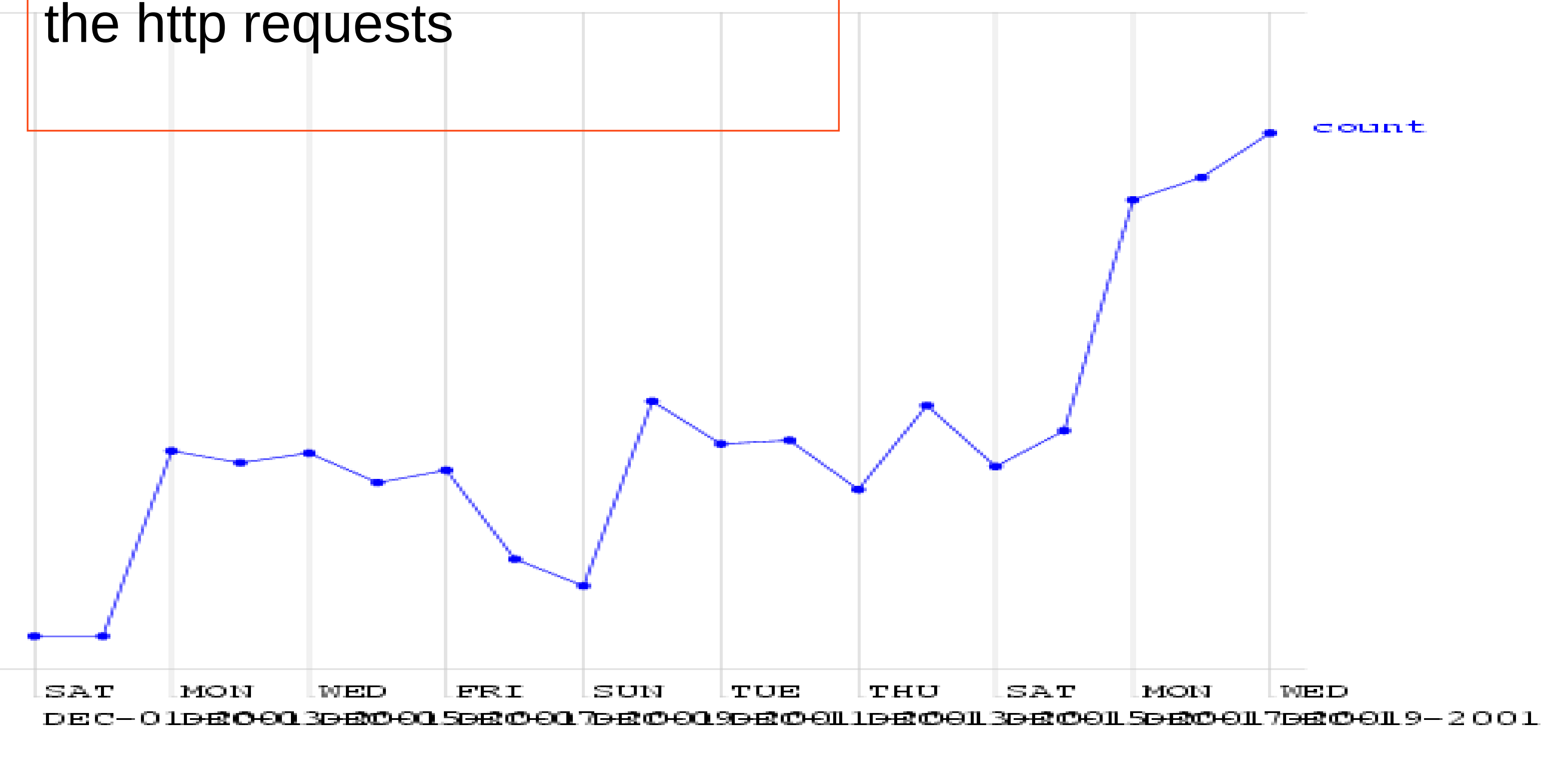


?

Dealt with by Statistical Quality Control
Signal an alarm if we go outside $\mu \pm 3\sigma$.

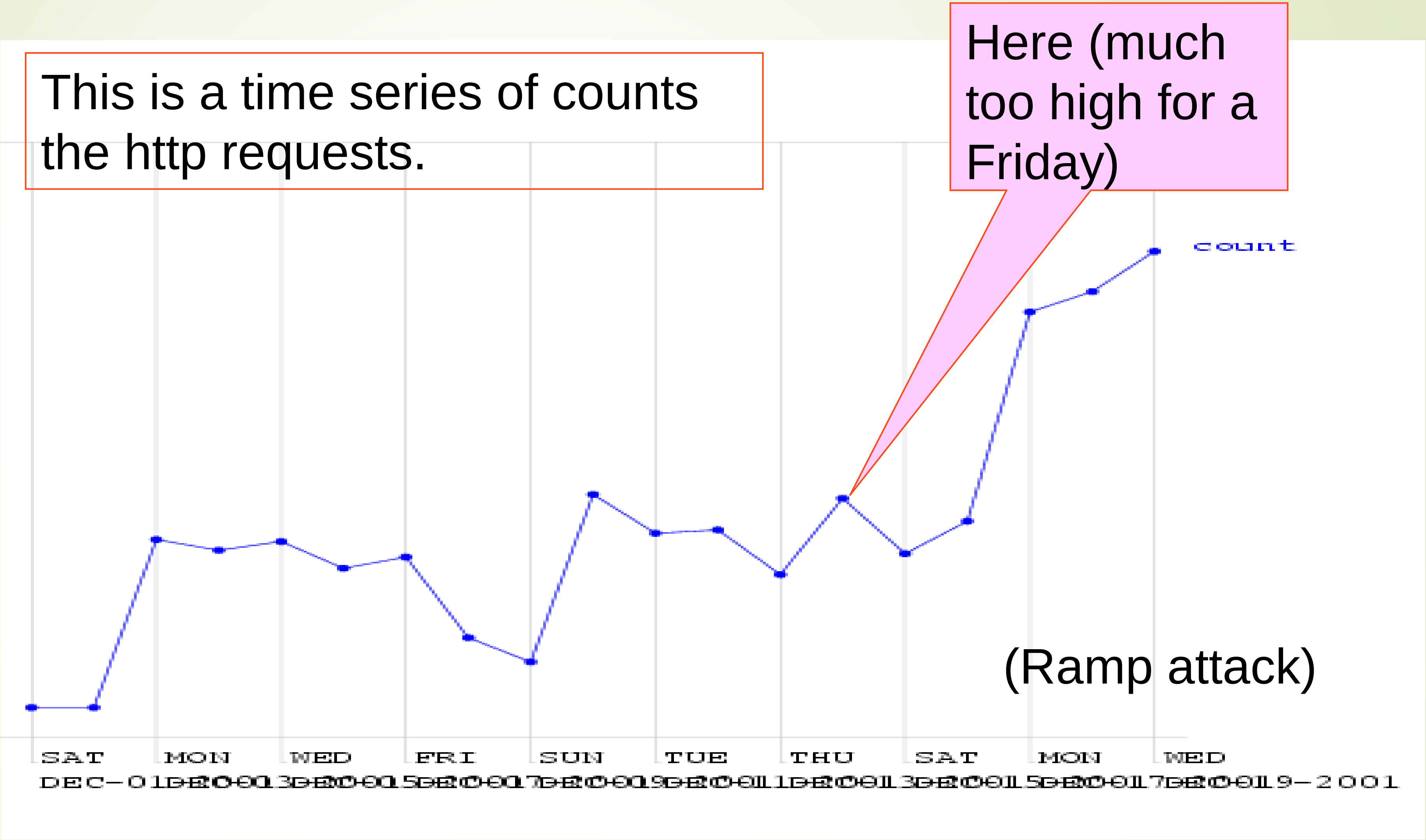
(When) is there an anomaly?

This is a time series of counts the http requests

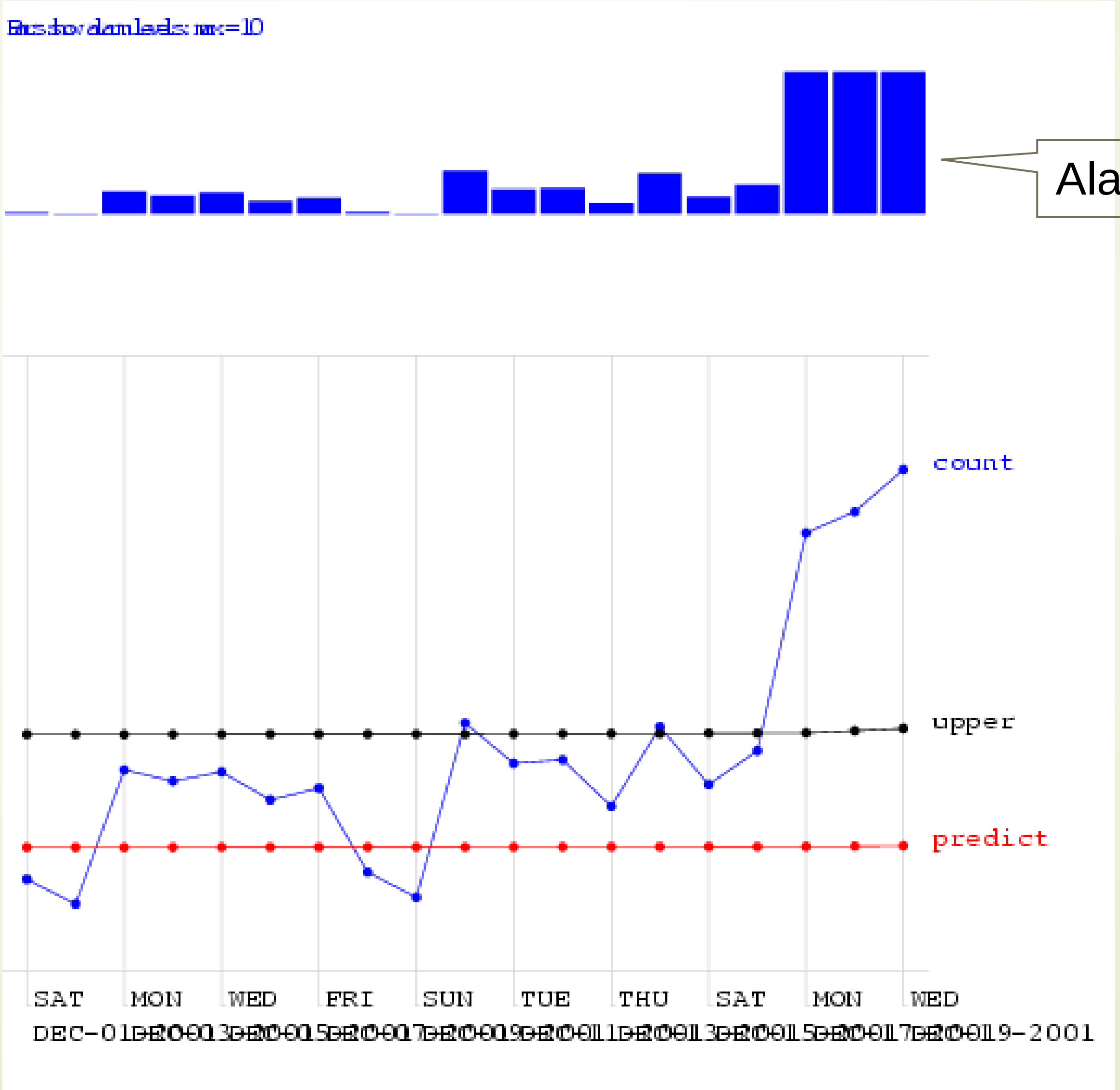


?

(When) is there an anomaly?



Control Charts on the Norfolk Data



Alarm Level

?

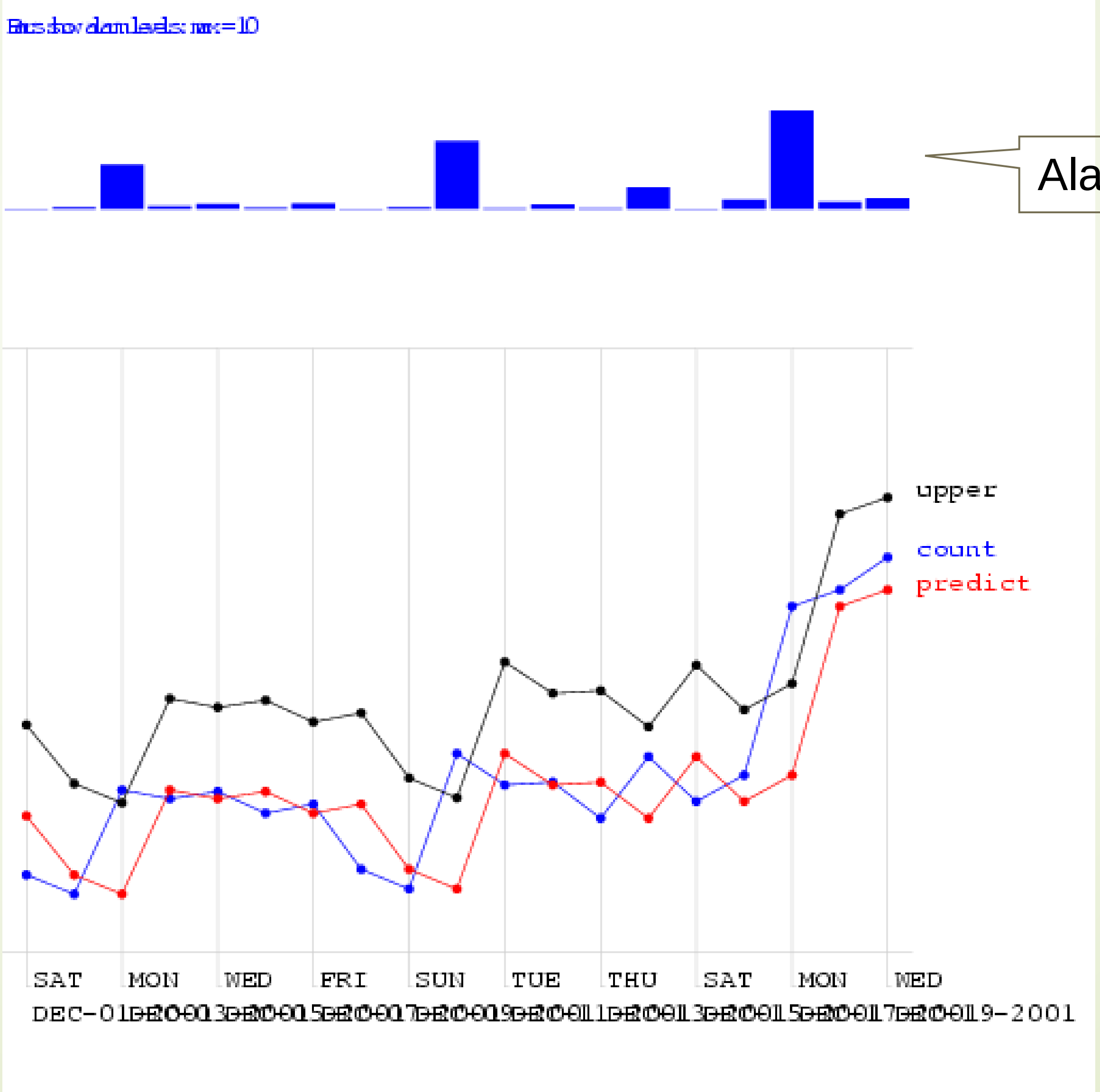
A simple rule of thumb for predicting the champion?

1987	***Los Angeles Lakers* (1)** (21, 10–11)	Pat Riley	4–2	Boston Celtics (1) (19, 16–3)	K. C. Jones
1988	***Los Angeles Lakers* (1)** (22, 11–11)	Pat Riley	4–3	Detroit Pistons (2) (3, 0–3)	Chuck Daly
1989	Los Angeles Lakers (1) (23, 11–12)	Pat Riley	0–4	***Detroit Pistons* (1)** (4, 1–3)	Chuck Daly
1990	Portland Trail Blazers (3) (2, 1–1)	Rick Adelman	1–4	***Detroit Pistons* (1)** (5, 2–3)	Chuck Daly
1991	Los Angeles Lakers (3) (24, 11–13)	Mike Dunleavy	1–4	***Chicago Bulls* (1)** (1, 1–0)	Phil Jackson
1992	Portland Trail Blazers (1) (3, 1–2)	Rick Adelman	2–4	***Chicago Bulls* (1)** (2, 2–0)	Phil Jackson
1993	*Phoenix Suns* (1) (2, 0–2)	Paul Westphal	2–4	**Chicago Bulls (2)** (3, 3–0)	Phil Jackson
1994	***Houston Rockets* (2)** (3, 1–2)	Rudy Tomjanovich	4–3	New York Knicks (2) (7, 2–5)	Pat Riley
1995	**Houston Rockets (6)** (4, 2–2)	Rudy Tomjanovich	4–0	*Orlando Magic* (1) (1, 0–1)	Brian Hill
1996	Seattle SuperSonics (1) (3, 1–2)	George Karl	2–4	***Chicago Bulls* (1)** (4, 4–0)	Phil Jackson
1997	Utah Jazz (1) (1, 0–1)	Jerry Sloan	2–4	***Chicago Bulls* (1)** (5, 5–0)	Phil Jackson
1998	*Utah Jazz* (1) (2, 0–2)	Jerry Sloan	2–4	**Chicago Bulls (1)** (6, 6–0)	Phil Jackson
1999[e]	***San Antonio Spurs* (1)** (1, 1–0)	Gregg Popovich	4–1	New York Knicks (8) (8, 2–6)	Jeff Van Gundy
2000	***Los Angeles Lakers* (1)** (25, 12–13)	Phil Jackson	4–2	Indiana Pacers (1) (1, 0–1)	Larry Bird
2001	***Los Angeles Lakers* (2)** (26, 13–13)	Phil Jackson	4–1	Philadelphia 76ers (1) (9, 3–6)	Larry Brown
2002	***Los Angeles Lakers* (3)** (27, 14–13)	Phil Jackson	4–0	New Jersey Nets (1) (1, 0–1)	Byron Scott
2003	***San Antonio Spurs* (1)** (2, 2–0)	Gregg Popovich	4–2	New Jersey Nets (2) (2, 0–2)	Byron Scott
2004	*Los Angeles Lakers* (2) (28, 14–14)	Phil Jackson	1–4	**Detroit Pistons (3)** (6, 3–3)	Larry Brown

West

East

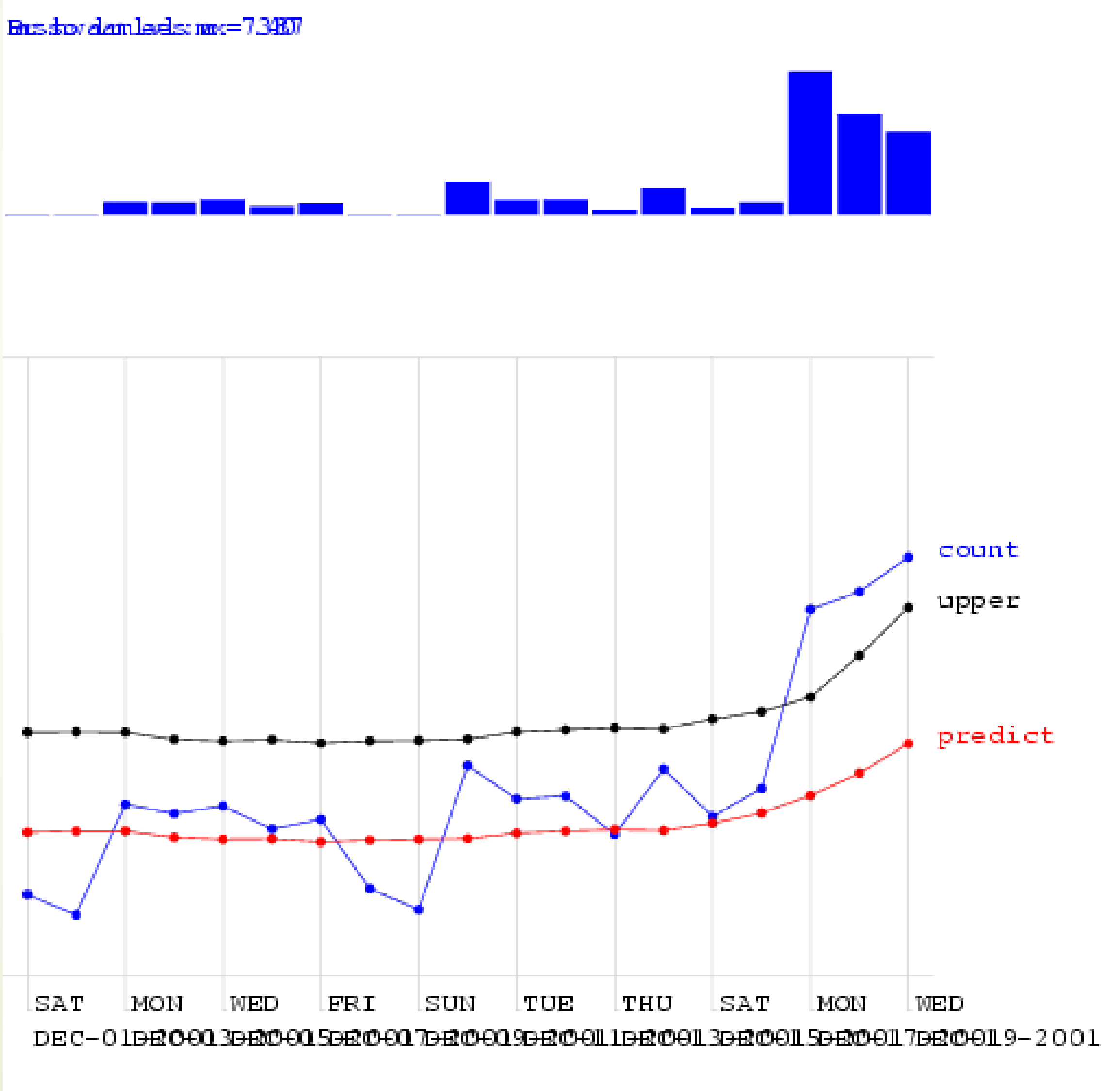
Looking at changes from yesterday



Alarm Level

?

Moving Average



Algorithm Performance

	Allowing one False Alarm per TWO weeks…		Allowing one False Alarm per SIX weeks…	
	Fraction of spikes detected	Days to detect a ramp attack	Fraction of spikes detected	Days to detect a ramp attack
standard control chart	0.39	3.47	0.22	4.13
using yesterday	0.14	3.83	0.1	4.7
Moving Average 3	0.36	3.45	0.33	3.79
Moving Average 7	0.58	2.79	0.51	3.31
Moving Average 56	0.54	2.72	0.44	3.54

Basic Idea contextual\collective anomaly detection

- Analyze the time series
- Make a prediction for the next point
- Evaluate the difference between the actual value and the predicted value
- Too big → anomaly
- It all relies on the quality of the prediction techniques

"It's Hard To Make Predictions, Especially About the Future" Attributed to Mark Twain and others

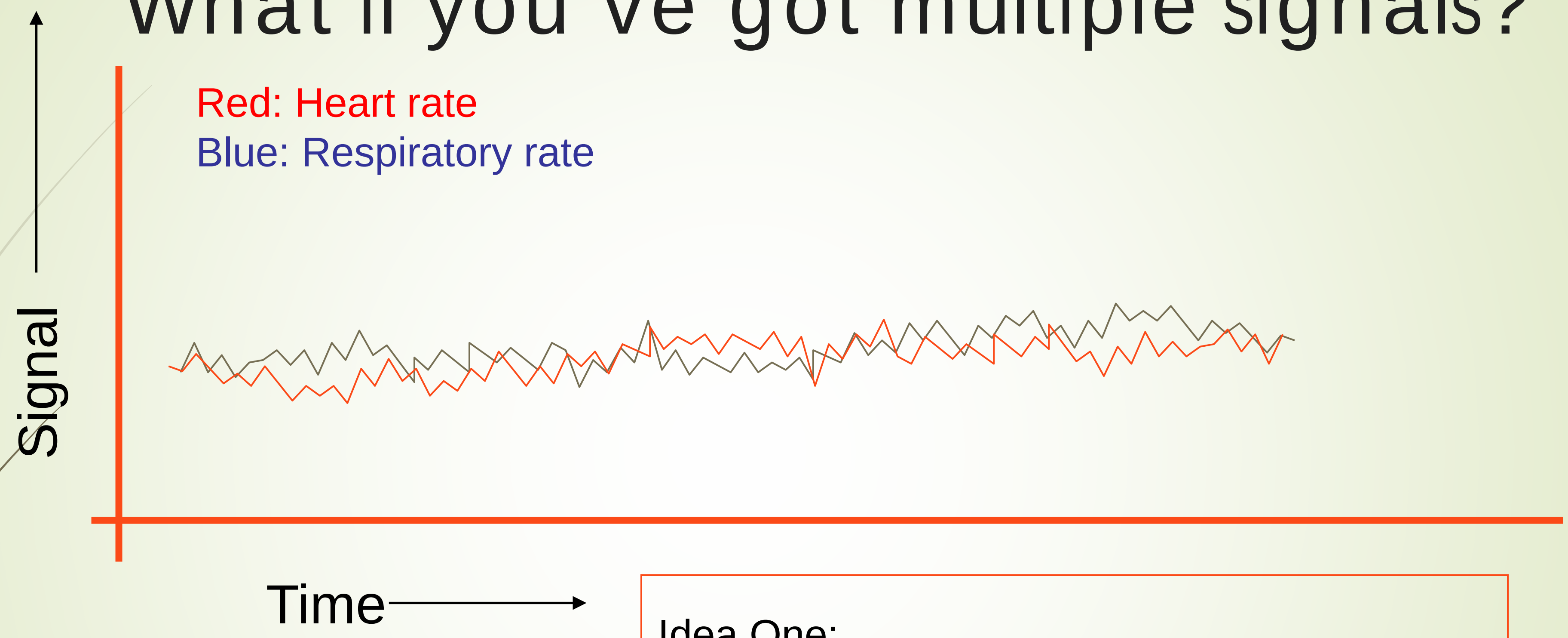
"Using the past to predict the future is akin to driving a car by looking into the rear-view mirror"

Notebook

Prediction using ARIMA

Multiple Signals

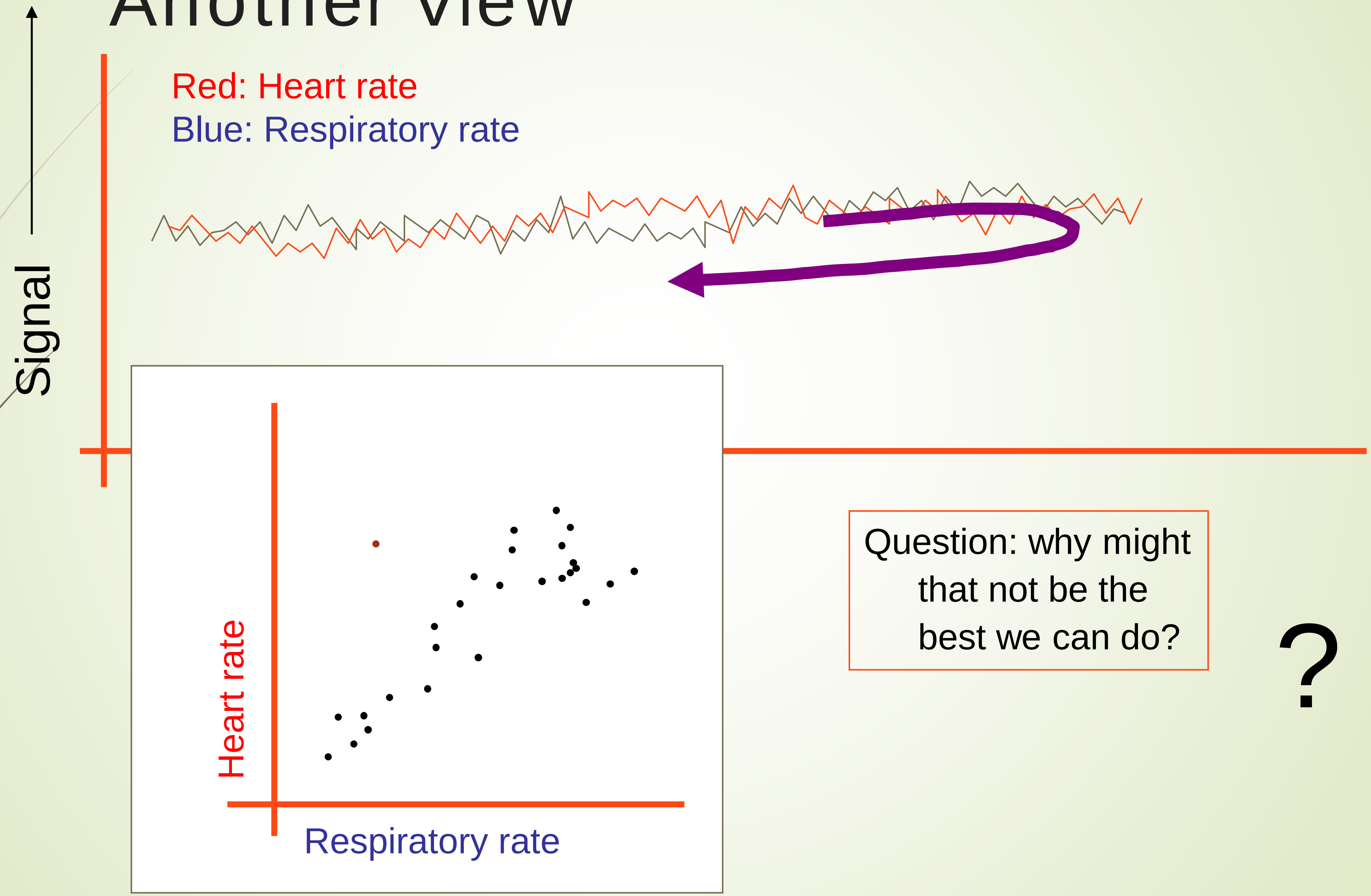
What if you've got multiple signals?



?

Idea One:
Simply treat it as two separate alarm-from-signal problems
…Question: why might that not be the best we can do?

Another View



Question: why might that not be the best we can do?

?

Aggregation

- There are two types of aggregation:
  - aggregation over series (or subject aggregation)
  - aggregation over time (or temporal aggregation)

Example in HW

Multivariate time series forecasting

- Recent review \ book
- An active research field
- Models were proposed to handle multiple features for prediction
- The subject can fill an entire academic course

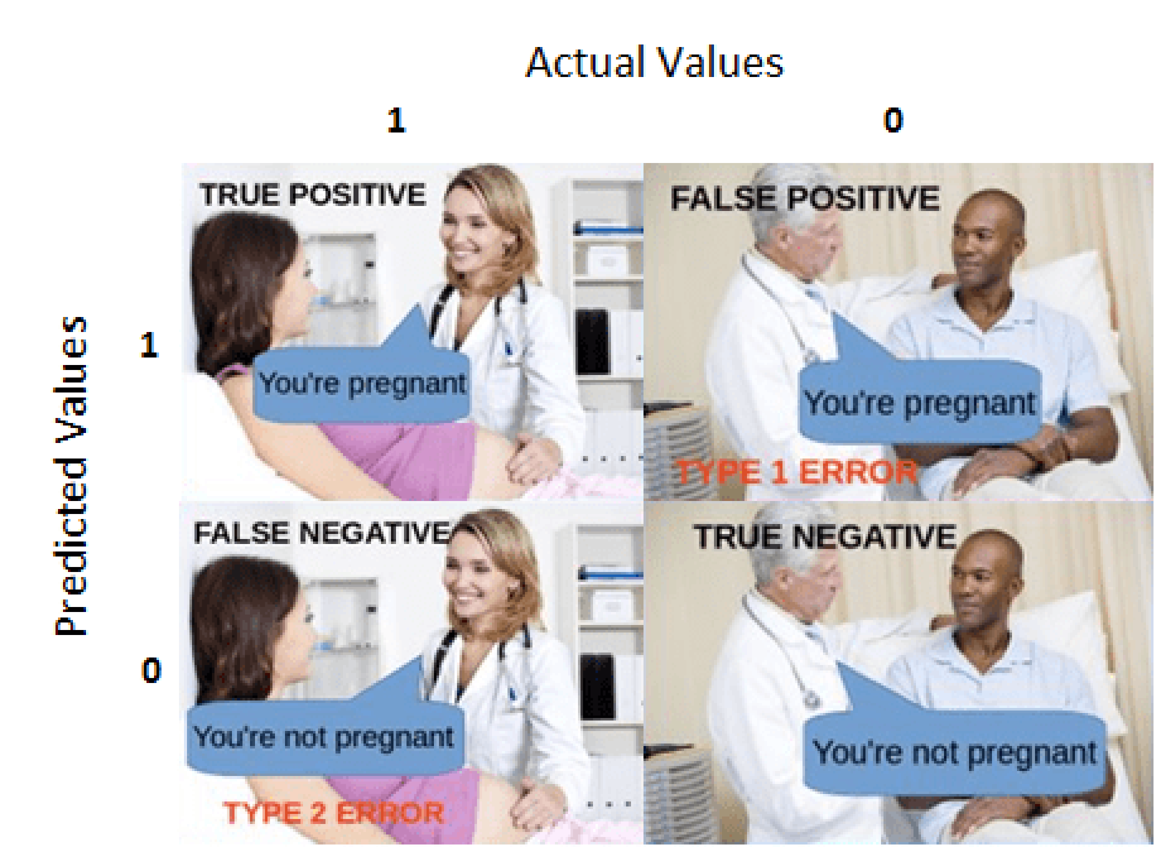
Forecasting: theory and practice

Fotios Petropoulos[1,*], Daniele Apiletti[2], Vassilios Assimakopoulos[3], Mohamed Zied Babai[4], Devon K. Barrow[5], Souhaib Ben Taieb[6], Christoph Bergmeir[7], Ricardo J. Bessa[8], Jakub Bijak[9], John E. Boylan[10], Jethro Browell[11], Claudio Carnevale[12], Jennifer L. Castle[13], Pasquale Cirillo[14], Michael P. Clements[15], Clara Cordeiro[16,17], Fernando Luiz Cyrino Oliveira[18], Shari De Baets[19], Alexander Dokumentov[20], Joanne Ellison[9], Piotr Fiszeder[21], Philip Hans Franses[22], David T. Frazier[23], Michael Gilliland[24], M. Sinan Gönül[25], Paul Goodwin[1], Luigi Grossi[26], Yael Grushka-Cockayne[27], Mariangela Guidolin[26], Massimo Guidolin[28], Ulrich Gunter[29], Xiaojia Guo[30], Renato Guseo[26], Nigel Harvey[31], David F. Hendry[32], Ross Hollyman[1], Tim Januschowski[33], Jooyoung Jeon[34], Victor Richmond R. Jose[35], Yanfei Kang[36], Anne B. Koehler[37], Stephan Kolassa[38,10], Nikolaos Kourentzes[39,10], Sonia Leva[40], Feng Li[41], Konstantia Litsiou[42], Spyros Makridakis[43], Gael M. Martin[23], Andrew B. Martinez[44,45], Sheik Meeran[1], Theodore Modis[46], Konstantinos Nikolopoulos[47], Dilek Önkal[25], Alessia Paccagnini[48,49], Anastasios Panagiotelis[50], Ioannis Panapakidis[51], Jose M. Pavía[52], Manuela Pedio[53,54], Diego J. Pedregal[55], Pierre Pinson[56], Patrícia Ramos[57], David E. Rapach[58], J. James Reade[59], Bahman Rostami-Tabar[60], Michał Rubaszek[61], Georgios Sermpinis[62], Han Lin Shang[63], Evangelos Spiliotis[3], Aris A. Syntetos[60], Priyanga Dilini Talagala[64], Thiyanga S. Talagala[65], Len Tashman[66], Dimitrios Thomakos[67], Thordis Thorarinsdottir[68], Ezio Todini[69,70], Juan Ramón Trapero Arenas[55], Xiaoqian Wang[36], Robert L. Winkler[71], Alisa Yusupova[10], Florian Ziel[72]

[1] School of Management, University of Bath, UK [2] Politecnico di Torino, Turin, Italy [3] Forecasting and Strategy Unit, School of Electrical and Computer Engineering, National Technical University of Athens, Greece [4] Kedge Business School, France [5] Department of Management, Birmingham Business School, University of Birmingham, UK [6] Big Data and Machine Learning Lab, Université de Mons (UMONS), Belgium [7] Faculty of Information Technology, Monash University, Melbourne, Australia [8] INESC TEC – Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal [9] Department of Social Statistics and Demography, University of Southampton, UK [10] Centre for Marketing Analytics and Forecasting, Lancaster University Management School, Lancaster University, UK [11] School of Mathematics and Statistics, University of Glasgow, UK [12] Department of Mechanical and Industrial Engineering, University of Brescia, Italy [13] Magdalen College, University of Oxford, UK [14] ZHAW School of Management and Law, Zurich University of Applied Sciences, Switzerland [15] ICMA Centre, Henley Business School, University of Reading, UK [16] Faculdade de Ciências e Tecnologia, Universidade do Algarve, Portugal [17] CEAUL, Faculdade de Ciências, Universidade de Lisboa, Portugal [18] Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil [19] Department of Business Informatics and Operations Management, Faculty of Economics and Business Administration, Universiteit Gent, Belgium [20] Let's Forecast, Australia [21] Faculty of Economic Sciences and Management, Nicolaus Copernicus University in Torun, Poland [22] Econometric Institute, Erasmus School of Economics, Rotterdam, The Netherlands [23] Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia [24] SAS, USA [25] Newcastle Business School, Northumbria University, Newcastle upon Tyne, UK [26] Department of Statistical Sciences, University of Padua, Italy [27] Darden School of Business, University of Virginia, USA [28] Finance Department, Bocconi University and Baffi-CAREFIN Centre, Milan, Italy [29] Department of Tourism and Service Management, MODUL University Vienna, Austria [30] Robert H. Smith School of Business, University of Maryland, USA [31] Department of Experimental Psychology, University College London, UK [32] Nuffield College and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK [33] Amazon Research, Germany [34] Korea Advanced Institute of Science and Technology, South Korea [35] McDonough School of Business, Georgetown University, USA [36] School of Economics and Management, Beihang University, Beijing, China [37] Miami University, Ohio, USA [38] SAP, Switzerland [39] Skövde Artificial Intelligence Lab, School of Informatics, University of Skövde, Sweden [40] Department of Energy, Politecnico di Milano, Italy [41] School of Statistics and Mathematics, Central University of Finance and Economics, Beijing, China [42] Manchester Metropolitan University Business School, UK [43] M Open Forecasting Center & Institute for the Future, University of Nicosia, Nicosia, Cyprus [44] Office of Macroeconomic Analysis, US Department of the Treasury, Washington DC, USA [45] GWU Research Program on Forecasting, Washington DC, USA [46] Growth Dynamics, Lugano, Switzerland [47] Durham University Business School, Durham University, UK [48] Michael Smurfit Business School, University College Dublin, Ireland [49] Centre for Applied Macroeconomic Analysis, Australia [50] Discipline of Business Analytics, The University of Sydney Business School, Australia [51] Department of Electrical and Computer Engineering, University of Thessaly, Volos, Greece [52] GIPEyOP, UMMICS, Department of Applied Economics, Universitat de Valencia, Spain [53] School of Accounting and Finance, University of Bristol, UK [54] Baffi-CAREFIN Centre, Bocconi University, Italy [55] ETSI Industrial, Universidad de Castilla-La Mancha, Ciudad Real, Spain [56] Department of Technology, Management and Economics, Technical University of Denmark, Denmark [57] Porto Accounting and Business School, Polytechnic of Porto, Portugal [58] Department of Economics, Saint Louis University, USA [59] Department of Economics, School of Politics, Economics and International Relations, University of Reading, UK [60] Cardiff Business School, Cardiff University, UK [61] SGH Warsaw School of Economics, Collegium of Economic Analysis, Poland [62] Adam Smith Business School, University of Glasgow, UK [63] Department of Actuarial Studies and Business Analytics, Macquarie University, Australia [64] Department of Computational Mathematics, University of Moratuwa, Sri Lanka [65] Department of Statistics, Faculty of Applied Sciences, University of Sri Jayewardenepura, Sri Lanka [66] Foresight, International Institute of Forecasters, USA [67] School of Economics and Political Science, National and Kapodistrian University of Athens, Greece [68] Norwegian Computing Center, Oslo, Norway [69] University of Bologna, Italy [70] Italian Hydrological Society, Bologna, Italy [71] Fuqua School of Business, Duke University, Durham, USA [72] House of Energy Markets and Finance, University of Duisburg-Essen, Germany

Alarm fatigue

- When one is exposed to many frequent alarms and consequently becomes desensitized
- Watch it with the false positives!



Conclusions

- Anomaly detection is a specific problem
- Machine learning techniques may be applied, but the special characteristics have to be considered
  - Point anomaly vs context\collective
- Many point-anomaly state-of-the-art algorithms are known
- Special algorithms for time series
- Usually supervised is preferred, but need data