

# Outlier Detection

---

Rita P. Ribeiro

Machine Learning - 2021/2022

# Summary

## 1. Basic Concepts

Definition of Outlier

Application Domains

Challenges

Key Aspects

## 2. Outlier Detection Approaches

Unsupervised Learning Techniques

Semi-supervised Learning Techniques

Advanced Topics

## 3. Summary

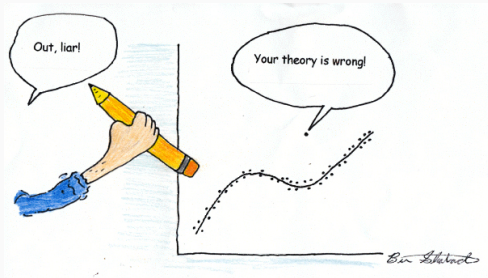
## Basic Concepts

---

- Most of data mining tasks focus on creating a model of the “normal” patterns in the data, extracting knowledge from what is common (e.g. frequent patterns).
- Still, rare patterns can also give us some important insights about data.
- Depending on the goal, those insights can be even more interesting/critical than the “normal” patterns.

# What is an Outlier?

- *“An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism” (Hawkins, 1980)*



## What is an Outlier? (cont.)

- Outliers represent patterns in data that do not conform to a well defined notion of normal.
- Initially, outliers were considered errors and their identification had data cleaning purposes.
- However, they can represent truthful deviation of data.
- For some applications, they represent critical information, which can trigger preventive or corrective actions.

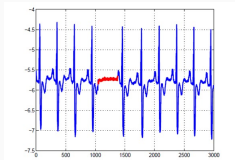


# Outliers and Anomalies

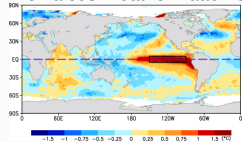
- Outlier and Anomaly detection are roughly related.
- Outliers can have a negative connotation being associated with data noise.
- Anomalies are often associated with unusual data that should be further investigated to identify the cause of occurrence.
- Anomaly can be considered as an outlier.
- But an outlier is not necessarily an anomaly.
- The following outlier detection application and methods involve outliers that can be seen as anomalies, i.e. meaningful outliers.

# Where can Outliers occur?

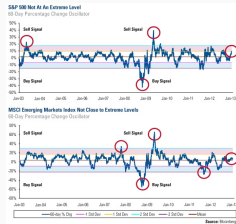
## Medical Analysis



## Anomalous Weather Patterns



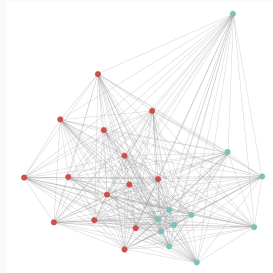
## Financial Markets



## Fraud Detection



## Social Network Analysis



## Event Detection in Text/Social Media





# Applications of Outlier Detection

- Quality Control and Fault Detection Applications
  - Quality Control
  - Fault Detection and Systems Diagnosis
  - Structure Defect Detection
- Financial Applications
  - Credit Card Fraud
  - Insurance Claim Fraud
  - Stock Market Anomalies
  - Financial Interaction Networks
- Intrusion and Security Applications
  - Host-based Intrusions
  - Network Intrusion Detection
- Web Log Analytics
  - Web Log Anomalies

# Applications of Outlier Detection (cont.)

- Market Basket Analysis
  - Outlier transactions in association patterns
- Medical Applications
  - Medical Sensor Diagnostics
  - Medical Imaging Diagnostics
- Text and Social Media Applications
  - Event Detection in Text and Social Media
  - Spam Email
  - Noisy and Spam Links
  - Anomalous Activity in Social Networks
- Earth Science Applications
  - Sea Surface Temperature Anomalies
  - Land Cover Anomalies
  - Harmful Algae Blooms

# Challenges of Outlier Detection

- Define every possible “normal” behaviour is hard.
- The boundary between normal and a outlying behaviour is often not precise.
- There is no general outlier definition; it depends on the application domain.
- It is difficult to distinguish real meaningful outliers from simple random noise in data.
- The outlier behaviour may evolve with time.
- Malicious actions adapt themselves to appear as normal.
- Inherent lack of known labeled outliers for training/validation of models.

# Key Aspects of Outlier Detection Problem

- Nature of Input Data
- Type of Outliers
- Intended Output
- Learning Task
- Performance Metrics

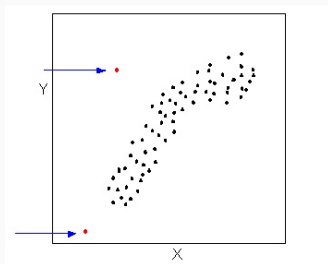
- Each data instance has:
  - One attribute (univariate)
  - Multiple attributes (multivariate)
- Relationship among data instances:
  - None
  - Sequential / Temporal
  - Spatial
  - Spatio-temporal
  - Graph
- Dimensionality of data

# Types of Outliers

- Point (or Global) Outlier
- Contextual Outlier
- Collective Outlier

### Point Outlier

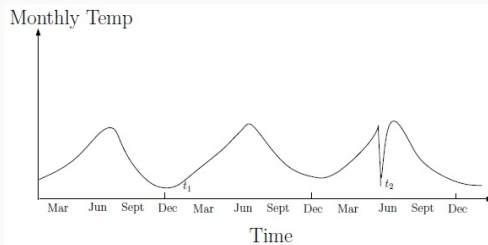
An instance that individually or in small groups is very different from the rest of the instances.



## Types of Outliers (cont.)

### Contextual Outlier

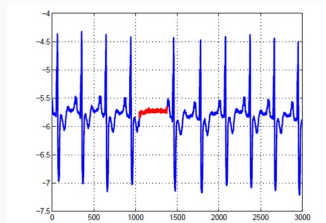
An instance that when considered within a context is very different from the rest of the instances.





### Collective Outlier

An instance that, even though individually may not be an outlier, inspected in conjunction with related instances and with respect to the entire data set is an outlier.



- Assign a **label/value**: identification normal or outlier instance.
- Assign a **score**: probability of being an outlier.
  - It allows the output to be **ranked**.
  - Requires the specification of a **threshold**.

## Unsupervised Outlier Detection

- data set has no information on the behaviour of each instance;
- it assumes that instances with normal behaviour are far more frequent;
- most common case in real-life applications.

## Semi-supervised Outlier Detection

- data set has a few instances of normal or outlier behaviour;
- some real-life applications, such as fault detection, provide such data.

## Supervised Outlier Detection

- data set has instances of both normal and outlier behaviour;
- hard to obtain such data in real-life applications.

## Inadequacy of Standard Performance Metrics

- Standard performance metrics (e.g. *accuracy*, *error rate*) assume that all instances are equally relevant for the model performance.
- These metrics would give a good performance estimation to a model that performs well on normal (frequent) cases and bad on outlier (rare) cases.

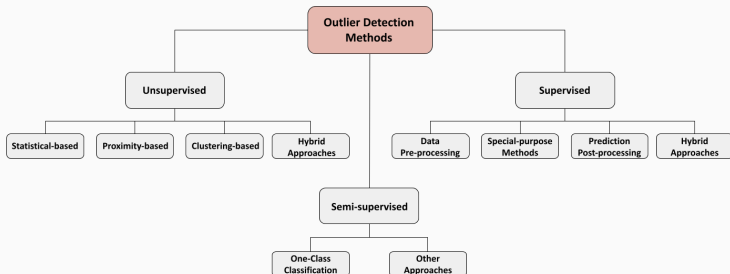
### Credit Card Fraud Detection:

- data set  $D$  with only 1% of fraudulent transactions;
- model  $M$  predicts all transactions as non-fraudulent;
- $M$  has a estimated accuracy of 99%;
- yet, all the fraudulent transactions were missed!

# Outlier Detection Approaches

---

# Taxonomy of Outlier Detection Methods

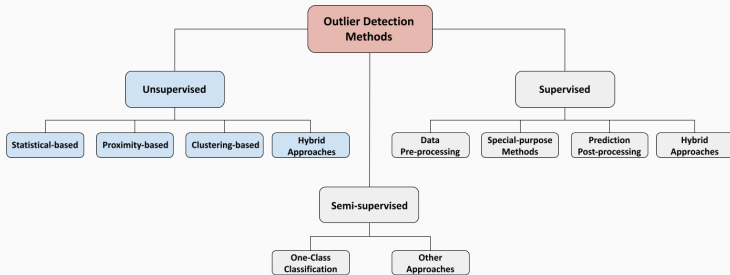


# Outlier Detection Approaches

---

## Unsupervised Learning Techniques

# Taxonomy of Anomaly Detection Methods





## Proposal

- All the points that satisfy a statistical discordance test for some statistical model are declared as outliers.

## Advantages

- If the assumptions of the statistical model hold true, these techniques provide a justifiable solution for outlier detection.
- The outlier score is associated with a confidence interval.

## Techniques

- Parametric
- Non-parametric

Assume one of the known probability distribution functions.

- *Grubbs' Test* (Grubbs, 1950)

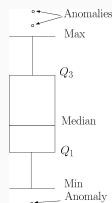
A statistical test used to detect outliers in a **univariate** data set assumed to come from a normally distributed population.

- *Boxplot* (Tukey, 1977)

It assumes a near-normal distribution of the values in a **univariate** data set, and identifies as outlier any value outside the interval

$$[Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR]$$

where  $Q_1$  ( $Q_3$ ) is the 1st (3rd) quartile and  $IQR$  is the interquartile range.



## Statistical-based Outlier Detection: Parametric Techniques (cont.)

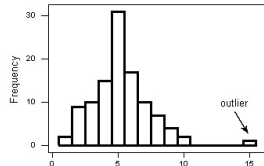
- *Mahalanobis distance* (Mahalanobis, 1936)
  - It assumes a multivariate normal distribution of data.
  - Incorporates dependencies between attributes by the covariance matrix.
  - Transforms a multivariate outlier detection task into a univariate outlier detection problem.
  - All the points with a large *Mahalanobis* distance are indicated as outliers.
- Mixture of parametric distributions
- etc.

# Statistical-based Outlier Detection: Non-parametric Techniques

The probability distribution function is not assumed, but estimated from data.

- Histograms

- Used for both univariate and multivariate data. For the later, the attribute-wise histograms are constructed and an aggregated score is obtained.
- Hard to choose the appropriate bin size.



- Kernel functions

- Adopt a kernel density estimation to estimate the probability density distribution of the data.
- Outliers are in regions of low density.

## Disadvantages

- The data does not always follow a statistical model.
- Choosing the best hypothesis test statistics is not straightforward.
- Capture interactions between attributes is not always possible.
- Estimating the parameters for some statistical models is hard.

## Proposal

- Normal instances occur in dense neighbourhoods, while outliers occur far from their closest neighbours.

## Advantages

- Purely data driven technique
- Does not make any assumptions regarding the underlying distribution of data.

## Some Techniques

- Distance-based
- Density-based

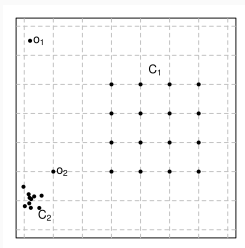
## Proximity-based Outlier Detection: Distance-based Techniques

A case  $c$  is an outlier if less than  $k$  cases are within a distance  $\lambda$  of  $c$   
[Knorr and Ng, 1998]

- Outliers are points far away from other points, thus given a distance metric there should not be a lot of other points in their neighborhood.
- Define proper distance metric (e.g euclidean distance)
  - The notion of distance between cases with many variables may be distorted by different scales, different importance, different types (numerical, nominal)
- Define a “reasonable” neighborhood ( $\lambda$ )
- Define what is “a lot of other points” ( $k$ )

## Proximity-based Outlier Detection: Distance-based Techniques (cont.)

- **Major cost:** for each point is calculated its distance to all the other points.
- The use of **global distance** measures poses difficulties in detecting outliers in data sets with different density regions.
- Example:



- $o_1$  and  $o_2$  are outliers
- but, for the point  $o_2$  to be identified as an outlier, all the points in  $C_1$  would have to be identified as outliers too.

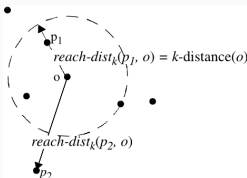


## Proximity-based Outlier Detection: Density-based Techniques

- Concept of outliers should be **locally** inspected.
- Compare points to their local neighborhood, instead of the global data distribution
- The density around an outlier is significantly different from the density around its neighbours.
- Use the relative density of a point against its neighbours as the indicator of the degree of the point being an outlier.
- Outliers are points in lower local density areas with respect to the density of its local neighbourhood.

# Proximity-based Outlier Detection: Density-based Techniques (cont.)

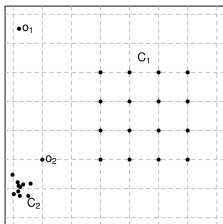
- LOF: Local Outlier Factor [Breunig et al., 2000]
  - *k-distance*: distance between  $p$  and its  $k$ -th nearest neighbour
  - *k-distance neighborhood*: all the points whose distance from  $p$  is not greater than the  $k$ -distance.
  - *reachability-distance* of  $p$  with respect to  $o$ : the maximum between their  $k$ -distance and their actual distance.



- intuition: high values of reachability-distance between two given points indicates that they may not be in the same cluster

# Proximity-based Outlier Detection: Density-based Techniques (cont.)

- LOF: Local Outlier Factor [Breunig et al., 2000] (cont.)
  - *local reachability-density* of a point is inversely proportional to the average reachability-distance of its  $k$  neighbourhood.
  - LOF assigns high values to the points that have a much lower *local reachability-density* in comparison to its  $k$ -neighbourhood.
  - Example:



- $o_2$  is assigned an higher LOF compared to the LOF values assigned to the points of  $C_1$  and  $C_2$
- This captures a local outlier whose local density is relatively low comparing to the local densities of its  $k$ -neighbourhood.

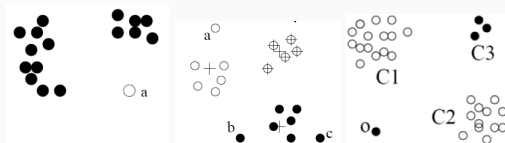
## Disadvantages

- True outliers and noisy regions of low density may be hard to distinguish.
- These methods need to combine global and local analysis.
- In high dimensional data, the contrast in the distances is lost.
- Computational complexity of the test phase.

# Clustering-based Outlier Detection

## Proposal

- Normal instances belong to large and dense clusters, while outlier instances are instances that:
  - do not belong to any of the clusters;
  - are far from its closest cluster;
  - form very small or low density clusters.



## Advantages

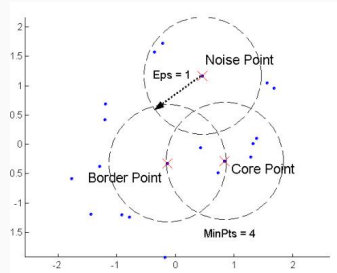
- Easily adaptable to on-line/incremental mode.
- Test phase is fast.

# Clustering-based Outlier Detection: Techniques

- DBSCAN [Ester et al., 1996]

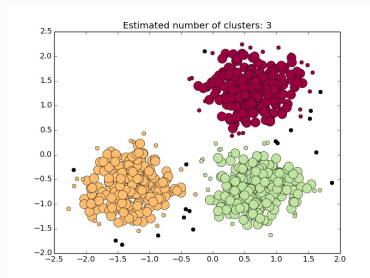
- Clustering method based on the notion of “density” of the points
- The density of a point is estimated by the number of points that are within a certain radius.
- Based on this idea, points can be classified as:

- *core points*: if the number of points within its radius are above a threshold
- *border points*: if the number of points within its radius are not above a threshold, but they are within a radius of a *core point*
- *noise points*: if do not have enough points within their radius, nor are sufficiently close to any *core point*.



# Clustering-based Outlier Detection: Techniques (cont.)

- DBSCAN [Ester et al., 1996] (cont.)
  - *noise points* are removed for the formation of clusters
  - all *core points* that are within a certain distance of each other are allocated to the same cluster
  - each *border point* is allocated to the cluster of the nearest *core points*
  - *noise points* are identified as outliers.



# Clustering-based Outlier Detection: Techniques (cont.)

- FindCBLOF [He et al., 2003]
  - To each point, assign a *cluster-based local outlier factor* (CBLOF)
  - The CBLOF score of a point  $p$  is determined by the size of the cluster to which  $p$  belongs, and the distance between  $p$  and
    - its cluster centroid, if  $p$  belongs to a large cluster
    - its closest large cluster centroid, if  $p$  belongs to a small cluster.
- $OR_H$  [Torgo, 2007]
  - Obtain an agglomerative hierarchical clustering of the data set
  - Use the information on the “path” of each point through the dendrogram as a form to determine its degree of outlyingness



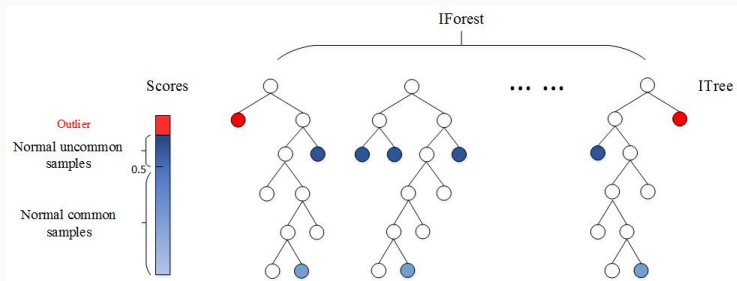
## Disadvantages

- Computationally expensive in the training phase.
- If normal points do not create any clusters, this technique may fail.
- In high dimensional spaces, clustering algorithms may not give any meaningful clusters.
- Some techniques detect outliers as a byproduct, i.e. they are not optimized to find outliers, their main aim is to find clusters.

- iForest [Liu et al., 2008] detects outliers purely based on the concept of isolation without employing any distance or density measure.
- Isolation: separating an instance from the rest of the instances
- A two-stage process.
  1. The first (training) stage builds an ensemble of data-induced random binary decision trees (isolation trees) using sub-samples of the given training set.
  2. The second (evaluation) stage passes test instances through isolation trees to obtain an outlier score for each instance.
- Parameters: number of trees and subsampling size

## Isolation Forest (cont.)

- The score is related to average path length
  - outliers are more likely to be isolated closer to the root
  - normal points are more likely to be isolated at the deeper levels



Source: <https://github.com/zmzhang/IOS/blob/master/images/IOS.jpg>

### Advantages

- No distance or density measures to detect anomalies;
- Eliminates a major computational cost of distance calculation in all distance-based and density-based methods;
- Scales up to handle extremely large data size and high-dimensional problems with a large number of irrelevant attributes.

### Disadvantages

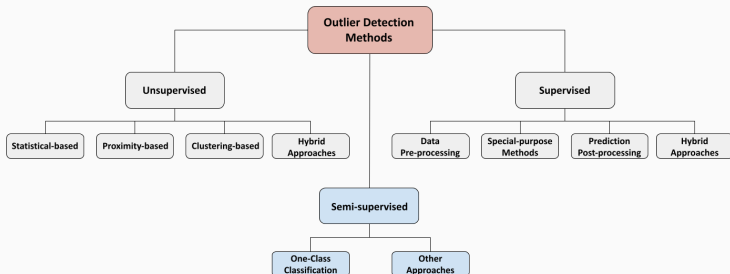
- Hyperparameters that must be tuned;
- Randomness component: different runs can give different results;
- Large sample sizes may cause masking or swamping.

# Outlier Detection Approaches

---

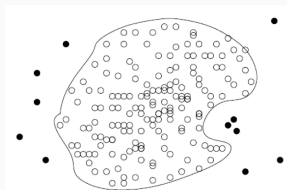
## Semi-supervised Learning Techniques

# Taxonomy of Outlier Detection Methods



## Proposal

- Build a prediction model to the normal behaviour and classify any deviations from this behaviour as outliers.

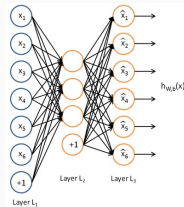


## Advantages

- Models are interpretable.
- Normal behaviour can be accurately learned.
- Can detect new outliers that may not appear close to any outlier points in the training set.

# One Class Classification: Techniques

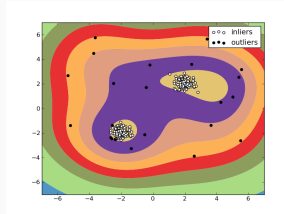
- Auto-associative neural networks [Japkowicz et al., 1995]
  - A feed-forward perceptron-based network is trained with normal data only.
  - The network has the same number of input and output nodes and a decreased number of hidden nodes to induce a bottleneck.
  - This bottleneck reduces the redundancies and focus on the key attributes of data.
  - After training, the output nodes recreate the example given as input nodes.
  - The network will successfully recreate normal data but will generate a high-recreation error for outlier data.





# One Class Classification: Techniques (cont.)

- One-class SVM [Tax and Duin, 2004]
  - It obtains a spherical boundary, in the feature space, around the normal data. The volume of this hypersphere is minimized, to minimize the effect of incorporating outliers in the solution.
  - The resulting hypersphere is characterized by a centre  $\mathbf{c}$  and a radius  $R$ .
  - The optimization problem consists of minimizing the volume of the hypersphere, so that includes all the training points.
  - Every point lying outside this hypersphere is an outlier.



## Disadvantages

- Requires previous labeled instances for normal behaviour.
- Possible high false alarm rate - previously unseen normal data may be identified as an outlier.

# Outlier Detection Approaches

---

## Advanced Topics

## Proposal

- If a data instance is an outlier in a specific context (but not otherwise), then it is considered as a contextual outlier.
- Each data instance is defined using two sets of attributes:
  - **Contextual attributes** used to determine the context (or neighbourhood) for that instance.
    - Sequential Context: position, time.
    - Spatial Context: latitude, longitude.
    - Graph Context: weights, edges.
  - **Behavioural attributes** which define the non-contextual characteristics of an instance.
- The outlier behaviour is determined using the values for the behavioural attributes within a specific context.

Example:

- Detect outlier customers in the context of customer groups
  - Contextual attributes: age group, postal code
  - Behavioural attributes: the number of transactions per year, annual total transaction amount

### **Advantages**

- Allow a natural definition of outlier in many real-life applications.
- Detects outliers that are hard to detect when analyzed in the global perspective.

## Techniques

- Reduction to point outlier detection
  - Segment data using contextual attributes.
  - Apply a traditional point outlier within each context using behavioural attributes.
  - Model “normal” behaviour with respect to contexts: an object is an outlier if its behavioural attributes significantly deviate from the values predicted by the model.
- Utilizing structure in data
  - Build models from the data using contextual attributes to predict the expected behaviour with respect to a given context.
  - Avoids explicit identification of specific contexts

### Disadvantages

- Identifying a set of good contextual attributes.
- It assumes that all normal instances within a context will be similar (in terms of behavioural attributes), while the outliers will be different.

## Proposal

- If a collection of related data instances is anomalous with respect to the entire data set, then it is considered a collective outlier.
- The individual data instances in a collective outlier may not be outliers by themselves, but their occurrence together as a collection is anomalous.

## Advantages

- Allow a natural definition of outlier in many real-life applications in which data instances are related.



### Techniques

- A collective outlier can also be a contextual outlier if analyzed with respect to a context.
- A collective outlier detection problem can be transformed to a contextual outlier detection problem by incorporating the context information.

### Disadvantages

- Contrary to contextual outliers, the structures are often not explicitly defined, and have to be discovered as part of the outlier detection process.
- Need to extract features by examining the structure of the dataset, i.e. the relationship among data instances for:
  - sequence data to detect anomalous sequences;
  - spatial data to detect anomalous sub-regions;
  - graph data to detect anomalous sub-graphs.
- The exploration of structures in data typically uses heuristics, and thus may be application dependent.
- The computational cost is often high due to the sophisticated mining process.

## Challenges

- Interpretation of outliers
  - Detecting outliers without saying why they are outliers is not very useful in high-D due to the many features (or dimensions) involved
  - Identify the subspaces that manifest the outliers
- Data sparsity
  - Data in high-D spaces is often sparse
  - The distance between objects becomes heavily dominated by noise as the dimensionality increases
- Data subspaces
  - Capturing the local behavior of data
- Scalable with respect to dimensionality
  - # of subspaces increases exponentially

## Techniques

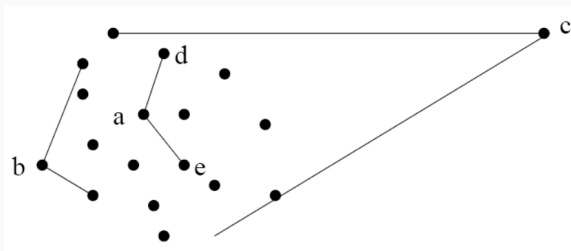
- Find distance-based outliers, but use the ranks of distance instead of the absolute distance in outlier detection.
- Dimensionality reduction: the principal components with low variance are preferred because, on such dimensions, normal objects are likely close to each other and outliers often deviate from the majority.
- Project data onto various subspaces to find an area whose density is much lower than average.

# Outlier Detection in High Dimensional Data (cont.)

## Techniques (cont.)

- Develop new models for high-dimensional outliers directly. Avoid proximity measures and adopt new heuristics that do not deteriorate in high-dimensional data.

E.g. Angle-based outliers.



# Summary

---

- Outliers are not necessarily random noise. They can represent critical information that can trigger preventive or corrective actions.
- The interpretability of an outlier detection method is extremely important.
- The nature of the outlier detection problem is dependent on the application domain.
- Different approaches to this problem are necessary.
- Contextual and collective outliers are having increasing applicability in several real-world domains.
- Online Outlier Detection and Distributed Outlier Detection are emerging topics.
- There is much space for the development of new techniques in this area.

# References

---



# References



Aggarwal, C. (2013).

***Outlier Analysis.***

Springer New York.



Aggarwal, C. C. (2015).

***Data Mining, The Textbook.***

Springer.



Aminian, E., Ribeiro, R. P., and Gama, J. (2021).

**Chebyshev approaches for imbalanced data streams regression models.**

*Data Min. Knowl. Discov.*, 35(6):2389–2466.



Branco, P., Torgo, L., and Ribeiro, R. P. (2016).

**A survey of predictive modeling on imbalanced domains.**

*ACM Comput. Surv.*, 49(2):31:1–31:50.



Branco, P., Torgo, L., and Ribeiro, R. P. (2018).

**Resampling with neighbourhood bias on imbalanced domains.**

*Expert Syst. J. Knowl. Eng.*, 35(4).



Branco, P., Torgo, L., and Ribeiro, R. P. (2019).

**Pre-processing approaches for imbalanced distributions in regression.**

*Neurocomputing*, 343:76–99.

## References (cont.)



Breunig, M. M., Kriegel, H. P., Ng, R., and Sander, J. (2000).

**Lof: Identifying density-based local outliers.**

*In Proceedings of ACM SIGMOD 2000 International Conference on Management of Data.*  
ACM Press.



Chandola, V., Banerjee, A., and Kumar, V. (2009).

**Anomaly detection: A survey.**

*ACM Computing Surveys (CSUR)*, 41(3):15.



Chawla, N. V., Bowyer, K. W., Hall, O. L., , and Kegelmeyer, W. P. (2002).

**Smote: Synthetic minority over-sampling technique.**

*Journal of Artificial Intelligence Research*, 16:321–357.  
AAAI Press.



Davari, N., Veloso, B., de Assis Costa, G., Pereira, P. M., Ribeiro, R. P., and Gama, J. (2021a).

**A survey on data-driven predictive maintenance for the railway industry.**

*Sensors*, 21(17):5739.



Davari, N., Veloso, B., Ribeiro, R. P., Pereira, P. M., and Gama, J. (2021b).

**Predictive maintenance based on anomaly detection using deep learning for air production unit in the railway industry.**

*In 8th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2021, Porto, Portugal, October 6-9, 2021*, pages 1–10. IEEE.

## References (cont.)



Ester, M., Peter Kriegel, H., S. J., and Xu, X. (1996).  
**A density-based algorithm for discovering clusters in large spatial databases with noise.**  
pages 226–231. AAAI Press.



Han, J., Kamber, M., and Pei, J. (2011).  
**Data Mining: Concepts and Techniques.**  
Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition.



Hawkins, D. M. (1980).  
**Identification of Outliers.**  
Chapman and Hall.



He, Z., Xu, X., and Deng, S. (2003).  
**Discovering cluster based local outliers.**  
*Pattern Recognition Letters*, 2003:9–10.



Hempstalk, K., Frank, E., and Witten, I. H. (2008).  
**One-class classification by combining density and class probability estimation.**  
In *ECML/PKDD (1)*, pages 505–519.



Hodge, V. J. and Austin, J. (2004).  
**A survey of outlier detection methodologies.**  
*Artificial Intelligence Review*, 22:2004.

## References (cont.)



Japkowicz, N., Myers, C., and Gluck, M. A. (1995).

**A novelty detection approach to classification.**

In *IJCAI*, pages 518–523. Morgan Kaufmann.



Joshi, M. V., Agarwal, R. C., and Kumar, V. (2002).

**Predicting rare classes: Comparing two-phase rule induction to cost-sensitive boosting.**

In *PKDD'02: Proceedings of the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases*, volume 2431 of *LNCS*, pages 237–249. Springer.



Joshi, M. V., Kumar, V., and Agarwal, R. C. (2001).

**Evaluating boosting algorithms to classify rare classes: Comparison and improvements.**

In *Proceedings of the 2001 IEEE International Conference on Data Mining, 29 November - 2 December 2001, San Jose, California, USA*, pages 257–264.



Knorr, E. M. and Ng, R. T. (1998).

**Algorithms for mining distance-based outliers in large datasets.**

In *VLDB'98: Proceedings of 24th International Conference on Very Large Data Bases*, pages 392–403. Morgan Kaufmann, San Francisco, CA.

## References (cont.)



Kubat, M. and Matwin, S. (1997).

**Addressing the curse of imbalanced training sets: one-sided selection.**

In *Proc. 14th International Conference on Machine Learning*, pages 179–186. Morgan Kaufmann.



Lazarevic, A. (2008).

**Anomaly detection: A tutorial.**

Tutorial Session on 2008 Siam Conference on Data Mining (SDM08).



Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008).

**Isolation forest.**

In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422.



Maloof, M. A. (2003).

**Learning when data sets are imbalanced and when costs are unequal and unknown.**

In *ICML-2003 workshop on learning from imbalanced data sets II*, volume 2, pages 2–1.



Moniz, N., Ribeiro, R. P., Cerqueira, V., and Chawla, N. V. (2018).

**Smoteboost for regression: Improving the prediction of extreme values.**

In *5th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2018*, pages 150–159. IEEE.

## References (cont.)



Portela, E., Ribeiro, R. P., and Gama, J. (2017).

**Outliers and the simpson's paradox.**

In *Advances in Soft Computing - 16th Mexican International Conference on Artificial Intelligence, MICAI 2017*, volume 10632 of *LNCS*, pages 267–278. Springer.



Portela, E., Ribeiro, R. P., and Gama, J. (2019).

**The search of conditional outliers.**

*Intell. Data Anal.*, 23(1):23–39.



Ribeiro, R. P. (2011).

***Utility-based Regression.***

PhD thesis, Dep. Computer Science, Faculty of Sciences - University of Porto.



Ribeiro, R. P. and Moniz, N. (2020).

**Imbalanced regression and extreme value prediction.**

*Mach. Learn.*, 109(9-10):1803–1835.



Ribeiro, R. P., Pereira, P. M., and Gama, J. (2016).

**Sequential anomalies: a study in the railway industry.**

*Mach. Learn.*, 105(1):127–153.



Tax, D. (2001).

***One-class classification: Concept learning in the absence of counter-examples.***

PhD thesis, Technische Universiteit Delft.

## References (cont.)



Tax, D. M. J. and Duin, R. P. W. (2004).

**Support vector data description.**

*Machine Learning*, 54(1):45–66.



Torgo, L. (2007).

**Resource-bounded fraud detection.**

*In Progress in Artificial Intelligence, 13th Portuguese Conference on Artificial Intelligence, EPIA 2007, Workshops*, pages 449–460.



Torgo, L. (2016).

**Outlier detection methods.**

Slides.



Torgo, L., Ribeiro, R. P., Pfahringer, B., and Branco, P. (2013).

**Smote for regression.**

*In Progress in Artificial Intelligence*, pages 378–389. Springer.



Weiss, G. M. (2004).

**Mining with rarity: a unifying framework.**

*SIGKDD Explorations Newsletter*, 6(1):7–19.



Zhang, Y., Meratnia, N., and Havinga, P. (2007).

**A taxonomy framework for unsupervised outlier detection techniques for multi-type data sets.**