Anomaly Detection

COMP90049

Semester 2, 2021

Qiuhong Ke, CIS

Acknowledgement: Jeremy Nicholson, Tim Baldwin Karin Verspoor

Roadmap

So far:

- Supervised learning
- Unsupervised learning
- Active learning
- Semi-supervised learning

Today: Anomaly Detection

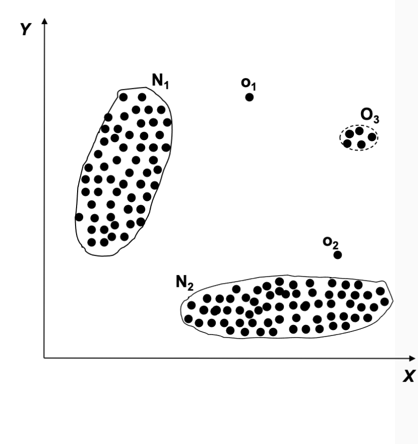
- Anomaly
  - Definition
  - Types
- Anomaly Detection Algorithms
  - Statistical
  - Proximity-based
  - Density-based
  - Clustering-based

THE UNIVERSITY OF MELBOURNE

Anomaly

What is Outlier/Anomaly?

A pattern in the data that does not conform to the normal/standard/expected behavior



Why Anomaly Detection?

Anomalous events are rare but can lead to dramatic (and often negative) consequence

Applications:

- Fraud Detection: odd credit card charges
- Ecosystem Disturbances: floods, droughts, heat waves
- Medicine and public health: influenza outbreaks
- Aviation Safety: abnormal pilot behavior or aircraft sequence of events

THE UNIVERSITY OF MELBOURNE

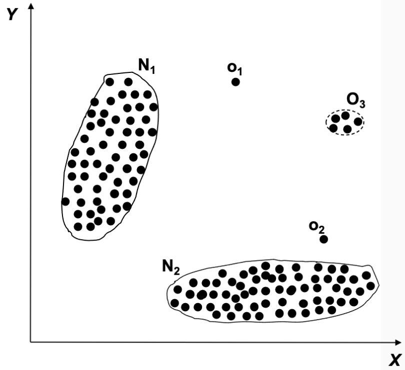
Type of Anomaly

- Point/global anomalies
- Contextual/conditional anomalies
- Collective anomalies

Point/global anomalies

An individual data instance is anomalous w.r.t. the data (deviate significantly the entirety of the data set)

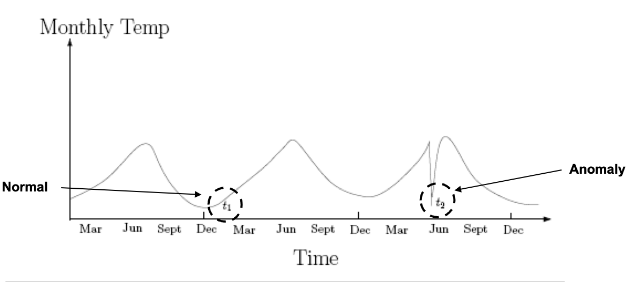
- Example: credit card fraud based on ”amount spent.”
- Detection: Find an appropriate measurement of deviation



Contextual/conditional anomalies

An individual data instance is anomalous within a context ((deviate significantly from the rest of data points in the same context)

- Example:
  - 150 heart rate is normal during exercise, but may be odd at rest.
  - Temperature in Paris:



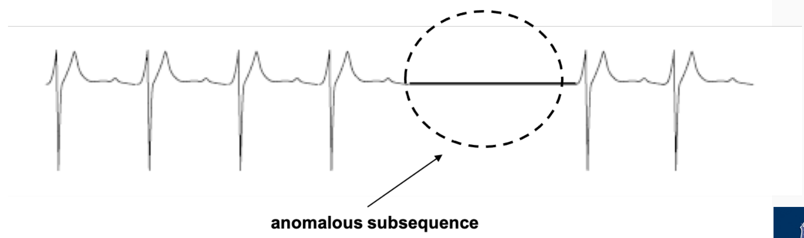
Contextual/conditional anomalies

- Attributes of data objects should be divided into two groups
  - Contextual attributes: defines the context, e.g., time
  - Behavioral attributes: characteristics of the object, used in anomaly evaluation, e.g., temperature
- Detection: How to define or formulate meaningful context?

Collective anomalies

A subset of data points is anomalous (deviate significantly from the entire data set)

- The individual instances within a collective anomaly are not anomalous by themselves
- Example:
  - cyber intrusion: Repeated failed login attempts
  - Heart rate signal:



Collective anomalies

Detection:

- Consider behavior of groups of objects
- Requires a relationship among data instances
  - Sequential data
  - Spatial data
  - Graph data

Anomaly vs Noise

- Anomalies are different from noise
  - Noise is random error
    - Label annotated incorrectly
    - Feature measured incorrectly
  - Noise is not necessarily interesting
  - Noise should be removed before anomaly detection
- Anomalies are interesting:
  - They violate the mechanism that generates the normal data
  - translate to significant (often critical) real life entities (e.g., cyber intrusions, credit card fraud)

Anomaly Detection Algorithms

Supervised Anomaly Detection

- Labels available for both normal data and anomalies
- Build classifier to distinguish between normal and known anomalies
- Challenges
  - Requires labels for both normal data and anomalies
  - Imbalanced classes,
  - Cannot detect unknown and emerging anomalies

Semi-supervised Anomaly Detection

- Labels available only for normal data
- Model normal objects and report those not matching the model as outliers
- Challenges:
  - Require labels from normal class
  - Possible high false alarm rate - previously unseen (yet legitimate) data records may be recognized as anomalies

Unsupervised point anomaly detection

- Statistical methods (model-based methods)
- Proximity-based: the nearest neighbors of outliers are far away
- Density-based: Outliers are objects in regions of low density
- Clustering-based Normal data belong to large and dense clusters

Statistical anomaly detection I

Anomalies are objects that are fit poorly by a statistical model.

- Assumption: normal data is generated by a parametric distribution
- Idea:
  - Estimate the parameters probability density function (PDF) of the distribution
  - Identify the instances in low probability regions of the distribution as anomalies
- Challenges of Statistical testing:
  - highly depends on whether the assumption of statistical model holds in the real data

Assumption: Gaussian distribution

$$PDF: \; f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$mean: \; \mu = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

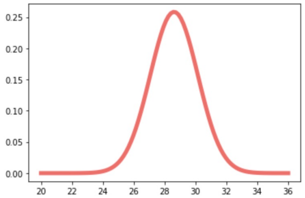
$$variance: \; \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

temp.: 24.0, 28.9, 28.9, 29.0, 29.1, 29.1, 29.2, 29.2, 29.3, 29.4

- Assumption: Gaussian distribution

$$\mu = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 28.61$$

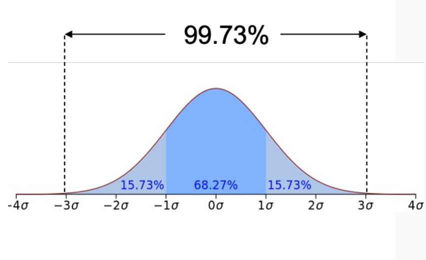
$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = 1.51$$



- Calculate probability using probability density function
- Outlier: low probabilities

temp.: 24.0, 28.9, 28.9, 29.0, 29.1, 29.1, 29.2, 29.2, 29.3, 29.4

- set a normal limit: $\mu \pm 3\sigma$ (the region contains 99.73% data)
- Then 24 is an outlier since: $(24–28.61)/1.51 = –3.04 < -3$



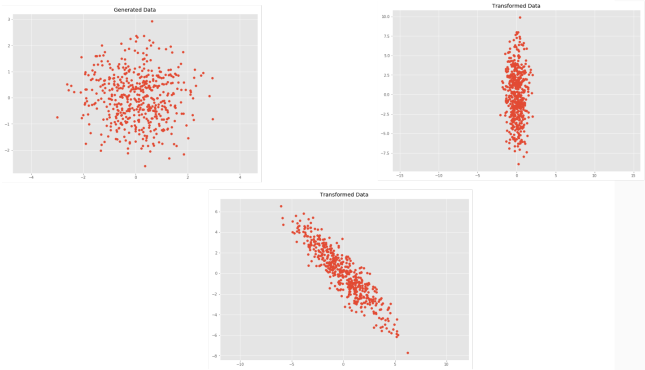
Multivariate Gaussian distribution

$$f(x) = \frac{1}{\sqrt{(2\pi)^k \det S}} \exp\left(-\frac{1}{2}(x-\mu)^T S^{-1}(x-\mu)\right)$$

μ: the mean.
k: dim of feature space.
S: covariance matrix.
For a 2-dimensional data:

$$S = \begin{bmatrix} \sigma^2(x,x) & \sigma^2(x,y) \\ \sigma^2(y,x) & \sigma^2(y,y) \end{bmatrix}$$

Multivariate Data II

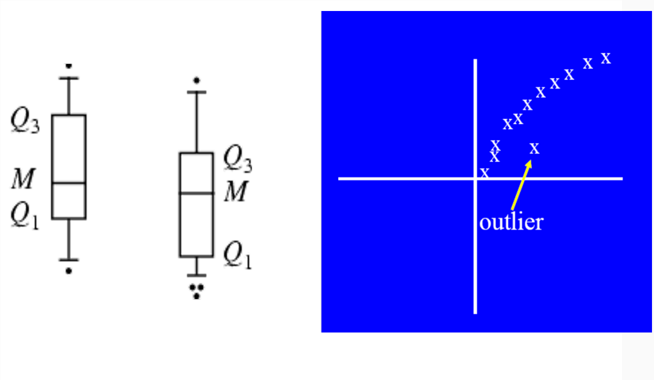


THE UNIVERSITY OF MELBOURNE

Statistical anomaly detection II

Graphical Approaches

- Boxplot (1-D), Scatter plot (2-D)
- Limitations
  - Time consuming
  - Subjective



Example

temp.: 24.0, 28.9, 28.9, 29.0, 29.1, 29.1, 29.2, 29.2, 29.3, 29.4

- Median: 29.1
- Q1: 28.9
- Q3: 29.2
- IQR: 29.2-28.9=0.4
- Minimum: Q1-1.5*IQR=28.3
- Maximum: Q3+1.5*IQR=29.8
- 24.0 $<$ Minimum: outlier

Proximity-based Anomaly detection

An object is an anomaly if the nearest neighbors of the object are far away,

- Compute the distance between every pair of data points
- To determine outliers:
  - Data points for which there are fewer than p neighboring points within a distance D
  - The top n data points whose distance to the kth nearest neighbor is greatest
  - The top n data points whose average distance to the k nearest neighbors is greatest

Proximity-based Anomaly detection

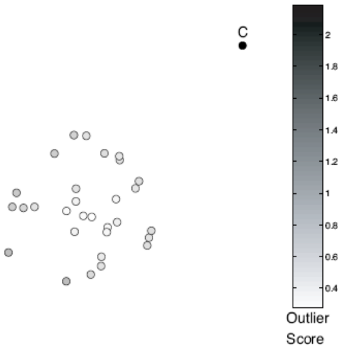


Figure 10.4. Outlier score based on the distance to fifth nearest neighbor.

Proximity-based (Nearest-Neighbor based) Anomaly detection

- Pros:
  - Easier to define a proximity measure for a dataset than determine its statistical distribution.
  - Quantitative measure of degree to which object is an outlier.
- Cons:
  - O(n2) complexity.
  - outlier score is sensitive to choice of k.
  - Does not work well if data has widely variable density.

Proximity-based Anomaly detection

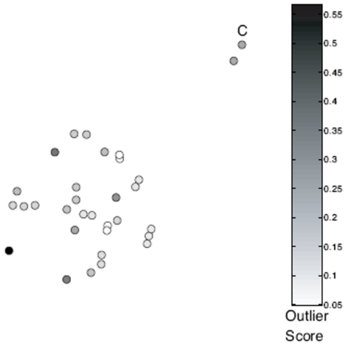


Figure 10.5. Outlier score based on the distance to the first nearest neighbor. Nearby outliers have low outlier scores.

Proximity-based Anomaly detection

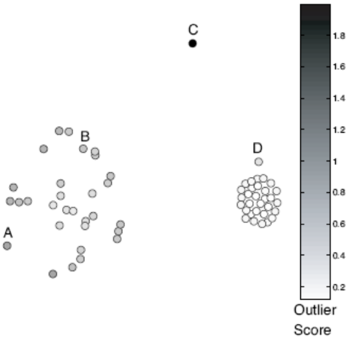


Figure 10.7. Outlier score based on the distance to the fifth nearest neighbor. Clusters of differing density.

Density-based outlier detection

Outliers are objects in regions of low density

- Outlier score is inverse of density around object.
- Density scores usually based on proximities. Example density scores:
  - Number of objects within fixed radius d.
  - inverse of average distance to k nearest neighbors:

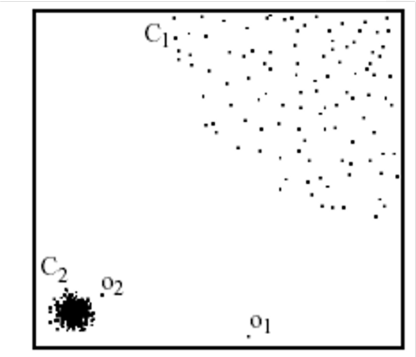
$$density(x, k) = \frac{1}{\frac{1}{k}\sum_{y \in N(x,k)} distance(x, y)}$$

  - These above two example scores work poorly if data has variable density.
  - Relative density outlier score (Local Outlier Factor, LOF):

$$relative\ density(x, k) = \frac{density(x, k)}{\frac{1}{k}\sum_{y \in N(x,k)} density(y, k)}$$

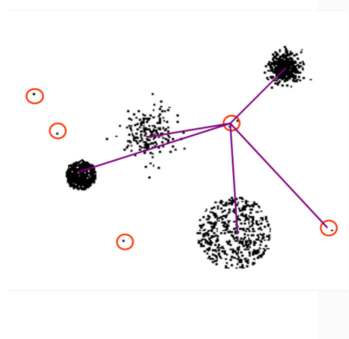
Example

How do you compare Proximity (Nearest-Neighbor) based and LOF in finding outliers?



Cluster-based Outlier Detection

Outliers are objects that do not belong strongly to any cluster



Approaches:

- Assess degree to which object belongs to any cluster.
- Eliminate object(s) to improve objective function.
- Discard small clusters far from other clusters.
- Issue: Outliers may affect initial formation of clusters.

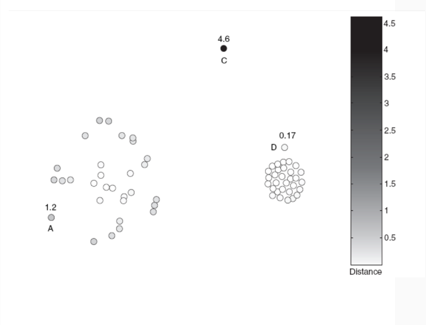
Cluster-based outlier detection

Assess degree to which object belongs to any cluster:

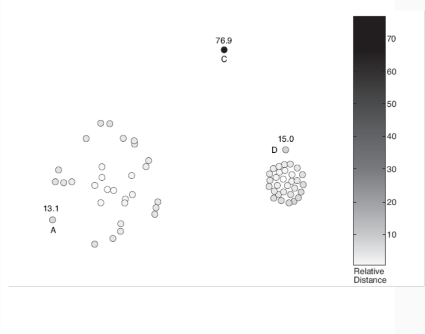
- For prototype-based clustering (e.g. k-means), use distance to cluster centers.
- To deal with variable density clusters, use relative distance:

$$\frac{\text{distance}(\mathbf{x}, \mathit{centroid}_C)}{\text{median}\left(\left\{\forall_{x' \in C}\ \text{distance}(\mathbf{x}', \mathit{centroid}_C)\right\}\right)}$$

Cluster-based outlier detection



Cluster-based outlier detection



Cluster-based outlier detection

Pro:

- Some clustering techniques have O(n) complexity.
- Extends concept of outlier from single objects to groups of objects.

Cons:

- Requires thresholds for the distance.
- Sensitive to number of clusters chosen.
- Outliers may affect initial formation of clusters.

Summary

Summary

- Types of outliers: global, contextual collective outliers
- Outlier detection: supervised, semi-supervised, or unsupervised
  - Statistical (or model-based) approaches
  - Proximity-base approaches
  - Density-based approaches
  - Clustering-base approaches

References

- Tan et al (2006) Introduction to Data Mining. Section 4.3, pp 150-171. (Chapter 10)
- V. Chandola, A. Banerjee, and V. Kumar, (2009). Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3), 1-58.
- A. Banerjee, et al (2008). Tutorial session on anomaly detection. The SIAM Data Mining Conference (SDM08)