ALGORITHMS FOR ANOMALY DETECTION

Michael Davis, Ph.D.

CERN, Meyrin

6 & 7 March 2017

Summary

- Introduction
- Global Anomaly Detection
  - Statistical Approaches
  - Classification-based Approaches
  - Clustering-based Approaches
- Local Anomaly Detection
  - Distance-based Approaches
  - Density-based Approaches
- Anomaly Detection in High-dimensional Data
- Other Approaches

Summary

- Introduction
- Global Anomaly Detection
  - Statistical Approaches
  - Classification-based Approaches
  - Clustering-based Approaches
- Local Anomaly Detection
  - Distance-based Approaches
  - Density-based Approaches
- Anomaly Detection in High-dimensional Data
- Other Approaches

Local Anomaly Detection Approaches

- Nearest Neighbours
- Distance-based Approaches
- Density-based Approaches

Local Anomaly Detection Approaches

Nearest Neighbour-based Approaches

Assumption

Normal data instances occur in dense neighbourhoods, while anomalies occur far from their closest neighbours

- **Distance-based:** Anomaly score is the distance of a data instance to its k^{th}-nearest neighbour
- **Density-based:** Anomaly score is the relative density of each data instance compared to its neighbourhood

Local Anomaly Detection Approaches

Nearest Neighbour-based Approaches

- All Nearest Neighbour-based approaches require a distance or similarity measure between pairs of data instances
- Distance measure is usually required to be:
  - **positive-definite** (can't have negative distances)
  - **symmetric**: d(a,b) = d(b,a)
  - but is not usually required to satisfy the **triangle inequality** $z \leq x + y$
- A **norm** is a function that assigns a strictly positive length or size to each vector in a vector space

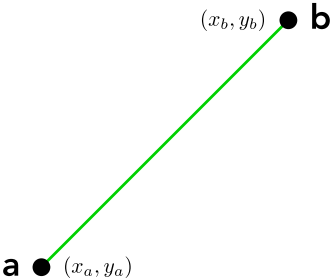
Local Anomaly Detection Approaches

Nearest Neighbour-based Approaches

- All Nearest Neighbour-based approaches require a distance or similarity measure between pairs of data instances
- Distance measure is usually required to be:
  - **positive-definite** (can't have negative distances)
  - **symmetric**: d(a,b) = d(b,a)
  - but is not usually required to satisfy the **triangle inequality** $z \leq x + y$
- A **norm** is a function that assigns a strictly positive length or size to each vector in a vector space

Nearest Neighbours

Euclidean distance



- "Ordinary" (straight-line) distance between two points in Euclidean space

Nearest Neighbours

Euclidean distance

- In 2D Cartesian space:

$$d(\mathbf{a},\mathbf{b}) = d(\mathbf{b},\mathbf{a}) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

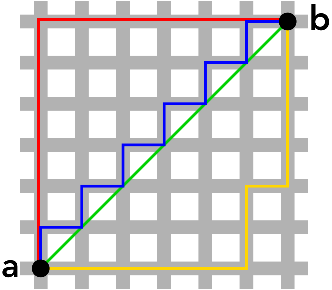
- In 3D Euclidean space:

$$d(\mathbf{a},\mathbf{b}) = d(\mathbf{b},\mathbf{a}) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

- Generalistion to any number of dimensions is the **Euclidean norm** or L^2 **norm**:

$$d(\mathbf{a},\mathbf{b}) = ||\mathbf{a} - \mathbf{b}|| = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$$

Nearest Neighbours
Manhattan Distance



- L^1 norm:

$$d(\mathbf{a},\mathbf{b}) = ||\mathbf{a}-\mathbf{b}||_1 = \sum_{i=1}^{n} |\mathbf{a}_i - \mathbf{b}_i|$$

Nearest Neighbours

Minkowski Distance

- Minkowski Distance is a generalisation of Manhattan distance and Euclidean distance
- L^p norm:

$$d(\mathbf{a}, \mathbf{b}) = ||\mathbf{a} - \mathbf{b}||_p = \left(\sum_{i=1}^{n} |\mathbf{a}_i - \mathbf{b}_i|^p \right)^{\frac{1}{p}}$$

- $p = 1$ gives Manhattan distance
- $p = 2$ gives Euclidean distance

Nearest Neighbours

Chebyshev Distance

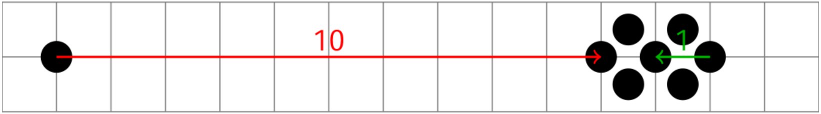
- In the limiting case of $p = \infty$, we get the **Chebyshev distance**
- L^∞ norm:

$$d(\mathbf{a},\mathbf{b}) = ||\mathbf{a}-\mathbf{b}||_\infty = \lim_{p\to\infty}\left(\sum_{i=1}^{n}|\mathbf{a}_i-\mathbf{b}_i|^p\right)^{\frac{1}{p}} = \max_{i=1}^{n}|\mathbf{a}_i-\mathbf{b}_i|$$

- Chebyshev distance has been used as a distance measure for high-dimensional data. Fast to compute and accuracy comparable to L^1 or L^2 norms.
- Distance measures with $0 < p < 1$ have also been effective for high-dimensional data

Distance-based Approaches

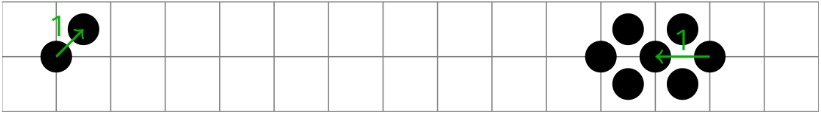
Distance to Nearest Neighbour



- Calculate distance of a data instance to its nearest neighbour
- **Labelling:** Threshold on the distance
- **Scoring:** Anomaly score = distance

Distance-based Approaches

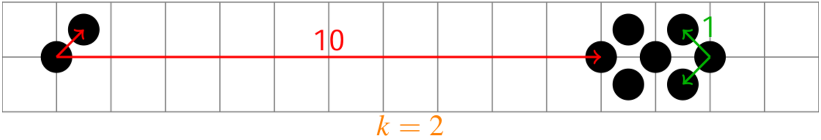
Distance to Nearest Neighbour



- Calculate distance of a data instance to its nearest neighbour
- **Labelling:** Threshold on the distance
- **Scoring:** Anomaly score = distance
- Misses paired outliers

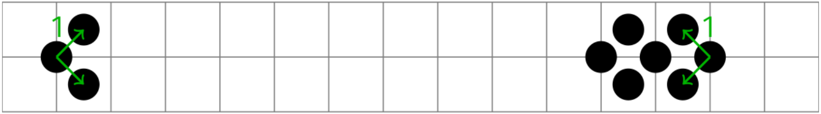
Distance-based Approaches

Distance to *k*-Nearest Neighbour



Distance-based Approaches

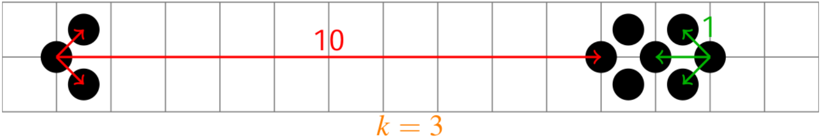
Distance to *k*-Nearest Neighbour



$k = 2$

Distance-based Approaches

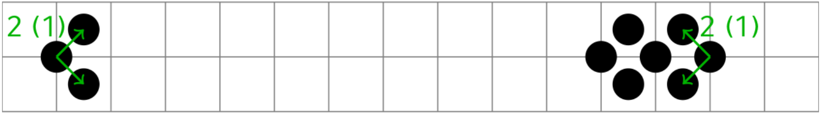
Distance to k-Nearest Neighbour



- Micro clusters ($|C| \leq k$) become outliers

Distance-based Approaches

Variations of k-NN

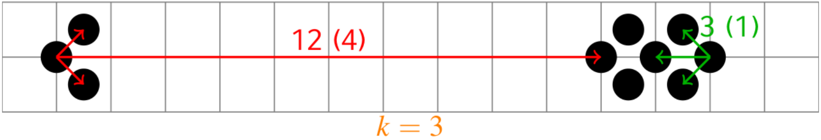


$k = 2$

- "Outliers are further away from the data"
- Anomaly score = sum (average) of distances to k-Nearest Neighbours
- More robust with respect to micro-clusters

Distance-based Approaches

Variations of k-NN



- "Outliers are further away from the data"
- Anomaly score = sum (average) of distances to k-Nearest Neighbours
- More robust with respect to micro-clusters

Distance-based Approaches

Global Density

- Count the number of neighbours in a hypersphere of radius *r*.
  - Fix radius *r* and use $\frac{1}{k}$ as the anomaly score
  - Fix *k* and use $\frac{1}{r}$ as the anomaly score
- The density of a data instance is:

$$\frac{k}{V(hypersphere)}$$

- For 2D data:

$$\frac{k}{\pi r^2}$$

Distance-based Approaches

Global Density

- Count the number of neighbours in a hypersphere of radius r.
  - Fix radius r and use $\frac{1}{k}$ as the anomaly score
  - Fix k and use $\frac{1}{r}$ as the anomaly score
- The density of a data instance is:

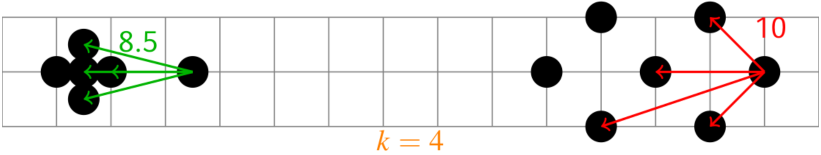
$$\frac{k}{V(hypersphere)}$$

- Anomaly score is the inverse of the density:

$$\frac{\pi r^2}{k}$$

Distance-based Approaches

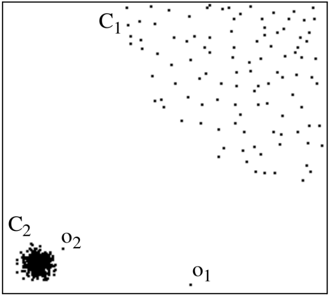
Global Density



- **However:** k-NN Cannot handle variations in density

Density-based Approaches

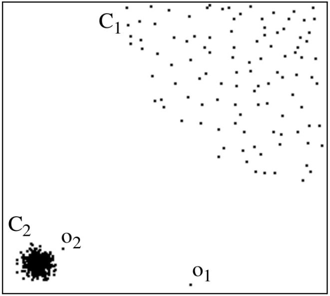
Relative Density



- **Global Density:** Outlier o_1 will be detected, o_2 will not.
- **Relative Density:** Anomalies are far from their neighbours, relative to the density of the local neighbourhood of each data point

Density-based Approaches

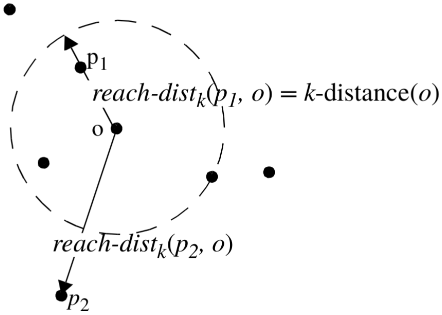
Relative Density



- **Global Density:** Outlier o_1 will be detected, o_2 will not.
- **Relative Density:** Anomalies are far from their neighbours, relative to the density of the local neighbourhood of each data point

Density-based Approaches

Reachability Distance



$\mathbf{k = 4}$

- **k-distance:** Distance to k^{th} Nearest Neighbour
- **Reachability Distance** (p, o):

$$\max\{d(p,o), \text{k-distance}(o)\}$$

- **Local Density:**

$$\frac{k}{V(hypersphere)}$$

Density-based Approaches

Local Outlier Factor (LOF)

- Choose parameter k
- Calculate the **local reachability density** of all data instances:

$$lrd_k(p) = \frac{1}{\text{average reachability distance of all points } o \in N_k(p)}$$

- The **Local Outlier Factor** (LOF) of a point p is the ratio:

$$\frac{\text{average local reachability density of all points } o \in N_k(p)}{\text{local reachability density of } p}$$

Density-based Approaches

Local Outlier Factor (LOF)

- Choose parameter k
- Calculate the **local reachability density** of all data instances:

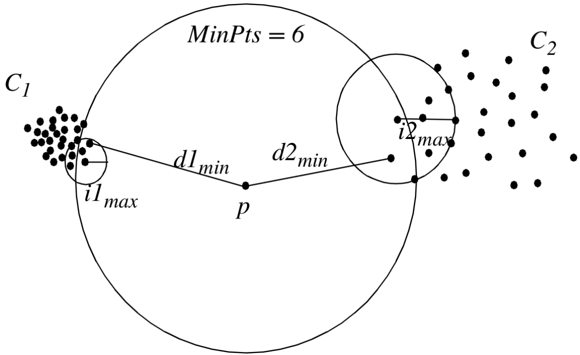
$$lrd_k(p) = \frac{1}{\text{average reachability distance of all points } o \in N_k(p)}$$

- The **Local Outlier Factor** (LOF) of a point p is the ratio:

$$LOF_k(p) = \frac{1}{|N_k(p)|} \sum_{o \in N_k(p)} \frac{lrd_k(o)}{lrd_k(p)}$$

Density-based Approaches

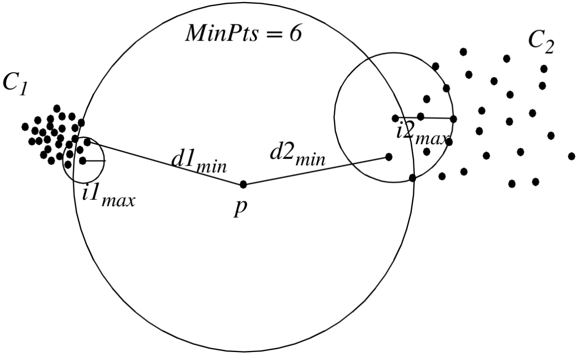
Local Outlier Factors (LOF)



- **Normal instance:** local density of p is similar to its neighbours
- **Outlier:** local density of p is lower than its neighbours

Density-based Approaches

Local Outlier Factors (LOF)



- Normal instance: $LOF_k(p) \leq 1.0$
- Outlier: $LOF_k(p) \gg 1.0$

Density-based Approaches

LOF Variants

- Outlier Detection using In-degree Number (ODIN)
  - ODIN score is the number of k-NNs of p which have p in their k-NN
- Connectivity-based Outlier Factor (COF)
  - Add the next-closest instance to the NN (rather than the closest to p)
  - Continue until we have k instances
- SLOM: LOF variant for detecting spatial anomalies in climate data
- LOF variant for categorical data using a similarity measure

Distance- and Density-based Approaches

Advantages

- Unsupervised, data-driven approach
- Does not make any assumptions about the generative process that created the data, or the statistical distribution of the data
- Adapting to other types of data is straightforward: define an appropriate distance measure

Distance- and Density-based Approaches

Disadvantages

- Risk of misclassification
- Performance greatly relies on the distance measure chosen
  - Euclidian distances perform well…
  - …but are expensive to compute
  - "Curse of Dimensionality"
- High $\mathcal{O}(N^2)$ computational complexity to calculate the neighbourhood

Distance- and Density-based Approaches

Computational Complexity

High $\mathcal{O}(N^2)$ computational complexity to calculate the neighbourhood can be mitigated by:

- **Indexing:** R-trees, R*-trees, X-trees can yield $\mathcal{O}(n \log n)$ complexity. But do not scale well in high dimensions.
- **Partitioning/Clustering:** Partion attribute space into a hypergrid. Linear in data size but exponential in number of attributes, so not suited to high-dimensional data.
- **Sampling/Pruning:** determine k-NN within small sample of dataset. Can result in incorrect anomaly scores if sample size is too small.

Anomaly Detection in High-dimensional Data

- "Curse of Dimensionality"
- Distance Concentration Effect
- Neighbourhood Selection
- Subspace Outlier Detection
- Outstanding Problems

Anomaly Detection in High-dimensional Data

Curse of Dimensionality

The term *dimensionality curse* is often used as a vague indication that high dimensionality causes problems in some situations.
The term was first used by Bellman in 1961 for combinatorial estimation of multivariate functions…
In the area of the nearest neighbors problem it is used for indicating that a query processing technique performs worse as the dimensionality increases.

—Beyer et al. 1999
When is "nearest neighbour" meaningful?

Curse of Dimensionality

"Curse of Dimensionality" is commonly used as a catch-all for three separate problems:

- Distance Concentration Effect
- Irrelevant attributes concealing relevant information
- Efficiency issues

Distance Concentration Effect

Assumption

The ratio of the variance of the length of any point vector (denoted by $||X_d||$) with the length of the mean point vector (denoted by $E[||X_d||]$) converges to zero with increasing data dimensionality.

Consequence

The proportional difference between the farthest-point distance D_{max} and the closest-point distance D_{min} (the relative contrast) vanishes.

—Beyer et al. 1999

Distance Concentration Effect

$$\text{If } \lim_{d \to \infty} \text{var}\left(\frac{||X_d||}{E[||X_d||]}\right) = 0, \text{ then } \frac{D_{max} - D_{min}}{D_{min}} \to 0.$$

- Relative contrast between near and far neighbours diminishes as the dimensionality increases
- This is known as the **concentration effect** of the distance measure
- It reduces the utility of the measure to discriminate between near and far neighbours

Distance Concentration Effect

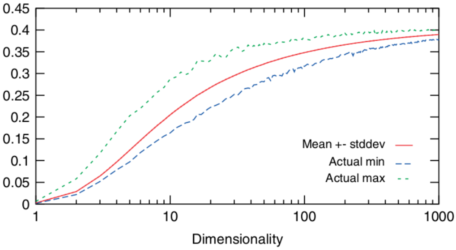
Distance Measures

- Covers a broad range of data distributions and distance measures (generally: all integer L^p norms with $p \geq 1$)
- *Hinnenburg et al.* show that L^1 and L^2 are the only integer norms useful for higher dimensions
- *Aggarwal et al.* show that fractional L^p norms can be used, but the result is only valid for uniformly distributed data
- Effect can be partially countered by rescaling to unit dimensions ($\forall x_i, x_i \in [-1, 1]$)

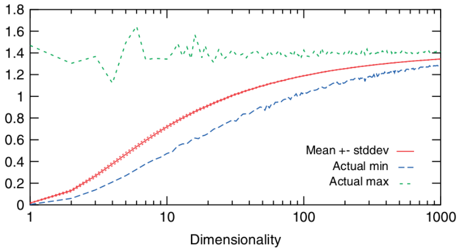
Distance Concentration Effect

Relative Contrast

- *k*-NN distances for uniformly distributed data and normally distributed data are very different at low dimensions
- They become almost the same at high dimensions



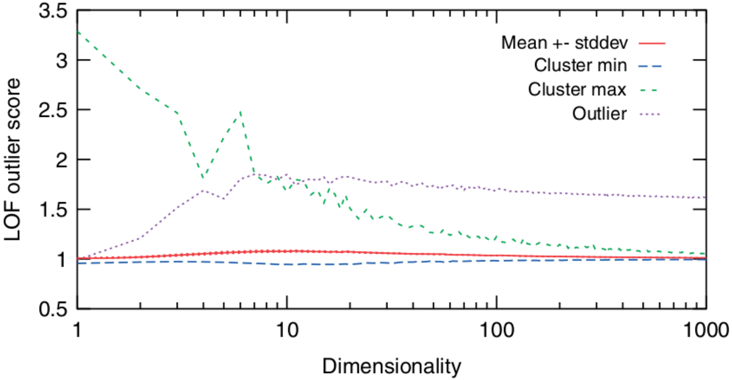
Normalized Euclidean 50NN-distance, uniform [0,1], logarithmic scale for dimensionality d.



Normalized Euclidean 50NN-distance, Gaussian [0,1], logarithmic scale for dimensionality d.

Distance Concentration Effect
Relative Contrast

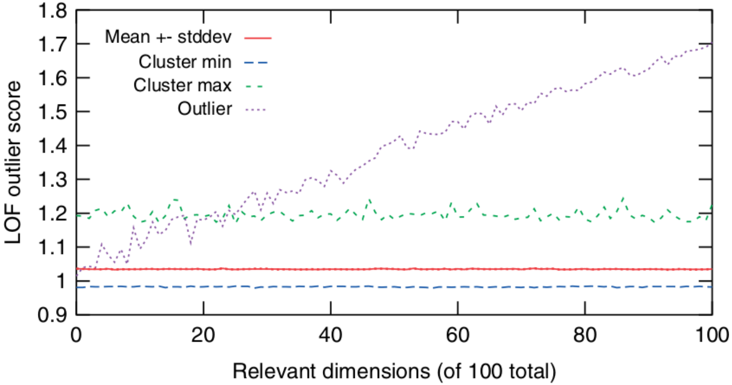
- However: for a constructed outlier ($x_i = 0.9$ in all dimensions), discrimination **increases** as dimensions increase:



Distance Concentration Effect

Relative Contrast

- This is because all dimensions add information
- Main problem for outlier detection in high-dimensional data is extra dimensions which **do not add information**



Neighbourhood Selection

What determines if the Nearest Neighbourhood is meaningful?

- **High signal-to-noise ratio:** irrelevant attributes mask the information in relevant attributes
- **"Self-similarity Blessing":** Latent correlation between the attributes results in an intrinsic dimensionality which is considerably lower than the representational dimensionality
- **Concentration effect** is less severe for clusters of points generated by the same mechanism

Neighbourhood Selection

What determines if the Nearest Neighbourhood is meaningful?

- **High signal-to-noise ratio:** irrelevant attributes mask the information in relevant attributes
- **"Self-similarity Blessing":** Latent correlation between the attributes results in an intrinsic dimensionality which is considerably lower than the representational dimensionality
- **Concentration effect** is less severe for clusters of points generated by the same mechanism

Neighbourhood Selection

What determines if the Nearest Neighbourhood is meaningful?

- **High signal-to-noise ratio:** irrelevant attributes mask the information in relevant attributes
- **"Self-similarity Blessing":** Latent correlation between the attributes results in an intrinsic dimensionality which is considerably lower than the representational dimensionality
- **Concentration effect** is less severe for clusters of points generated by the same mechanism

Neighbourhood Selection

Pairwise Stability

- **Pairwise stability** between clusters holds when:

mean distance between points of different clusters $\gg$ mean distance between points of the same cluster

- If clusters are pairwise stable, the NN of any point tends to belong to the same cluster
- NN queries on the order of the cluster size can still be meaningful, even if differentiation between neighbours within the same cluster is meaningless

Neighbourhood Selection

Pairwise Stability

- **Pairwise stability** between clusters holds when:

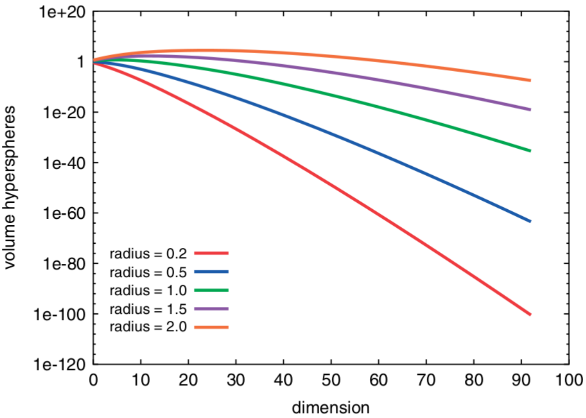
mean distance between points of different clusters	$\gg$	mean distance between points of the same cluster

- If clusters are pairwise stable, the NN of any point tends to belong to the same cluster
- NN queries on the order of the cluster size can still be meaningful, even if differentiation between neighbours within the same cluster is meaningless

Neighbourhood Selection

Absolute Distance vs. Distance Rank

- At high dimensions, a small change in radius *r* leads to big change in volume of a hypersphere



Neighbourhood Selection

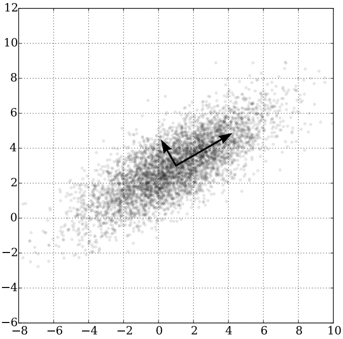
Absolute Distance vs. Distance Rank

- At high dimensions, a small change in radius r leads to big change in volume of a hypersphere
- Selecting neighbourhood using radius r is unstable
- Hard to select correct value for r
- Selecting k-NN is more stable as it relies on a **distance ranking** rather than absolute distances

Neighbourhood Selection

Approximate Neighbourhoods

- Anomaly detection in a reduced feature space:
  - **Step 1:** Global dimensionality reduction, *e.g.* by Principal Component Analysis
  - **Step 2:** Outlier detection in the reduced feature space



Neighbourhood Selection

Approximate Neighbourhoods

- Anomaly detection in a reduced feature space:
  - **Step 1:** Global dimensionality reduction, *e.g.* by Principal Component Analysis
  - **Step 2:** Outlier detection in the reduced feature space
- Can be effective in selecting the neighbourhood
- However, the subspace is usually insufficient to derive all outliers
- Outlier detection step is likely to fail

Neighbourhood Selection

Johnson-Lindenstrauss Transform

Johnson-Lindenstrauss Lemma

- Proves that when *n* objects are projected into a lower-dimensional space of dimensionality $\mathcal{O}(\frac{\log n}{\epsilon^2})$, the distances are preserved within a factor of $1 + \epsilon$
- Reduced dimensionality does not depend on the original dimensionality
- Choice of error bound ϵ gives a controlled trade-off between efficiency and precision
- Random projection is independent of the data and cheap to compute compared to PCA

Neighbourhood Selection

Projection Indexed Nearest Neighbours (PINN)

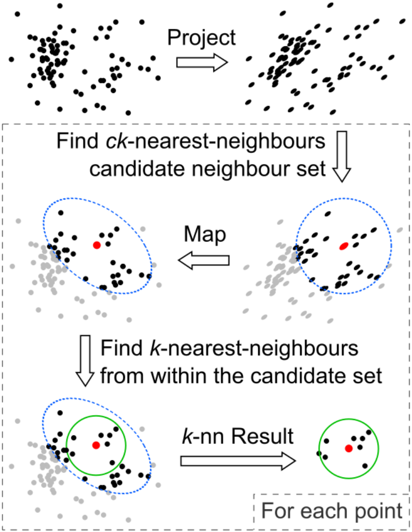
- High-dimensional variant of LOF
- Use random projections to find the approximate neighbourhood of each point
- Calculate LOF scores in the original space

Main Result

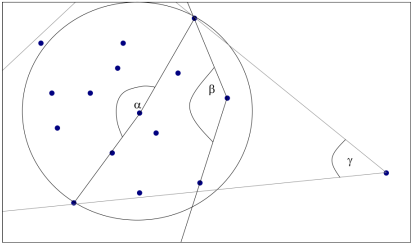
Proves that the outlier scores are preserved within the error bound ϵ of the random projection

Neighbourhood Selection

Projection Indexed Nearest Neighbours (PINN)



Angle-based Outlier Detection (ABOD)



- Distance measure based on Cosine Distance
- **Normal data:** most other data objects are distributed in all directions
- **Outlier:** most other data objects are distributed in a few directions
- Lower variance signifies higher outlierness

Subspace-based Outlier Detection

Data-snooping Bias

- Say we choose 3σ as the threshold, then likelihood of an outlier is 0.9973^d.
  - At $d = 10$, 97.33% of objects are within 3σ in every dimension
  - At $d = 100$, 76.31% of objects are within 3σ in every dimension
  - At $d = 1000$, 6.696% of objects are within 3σ in every dimension
- At high dimensions, virtually every object is extreme in at least one dimension
- Therefore, feature selection/searching many subspaces runs the risk of **data-snooping bias** (a kind of model overfitting)

Subspace-based Outlier Detection

OutRank

- Grid-based clustering approach
- Clusters (as opposed to outliers) are not rare objects and are recognisable at higher dimensionality
- Outlierness based on how often the object is recognised as part of a cluster
- Relies on clusters being well-separated
- By finding clusters first, avoids **data-snooping bias**…
- …but grid-based approach leads to **combinatorial explosion** of the search space

Problems of High-dimensional Data

- **Concentration of Scores:** distances of attribute-wise i.i.d. objects converge to a normal distribution with low variance
- **Noise Attributes:** irrelevant attributes can mask relevant attributes
- **Definition of Reference Sets:** Need to know the neighbours to choose the subspace; need to know the subspace to find the neighbours

Problems of High-dimensional Data

- **Data-snooping Bias:** Given enough subspaces, we can find at least one subspace in which the point is an outlier.
- **Exponential Search Space:** Number of possible subspaces grows exponentially with number of dimensions
- **Thresholding:** While ranking outlier scores may be valid, it may be impossible to find a threshold between inliers and outliers due to low contrast

Anomaly Detection in Structured Data

- Regression Models
  - Where we expect linearity in the data
- Time Series Data
  - Temporal sequence of data points, sometimes with cyclic patterns

Anomaly Detection in Structured Data

- Regression Models
  - Where we expect linearity in the data
- Time Series Data
  - Temporal sequence of data points, sometimes with cyclic patterns

Anomaly Detection in Structured Data

- Geographic/Directional Data
  - Spatial relationship between data points
  - Spatial Outlier Factor, a variant of LOF based on spatial neighborhoods
- Graph-based Data
  - Social networks, transport networks, computer networks, ecosystems, …
  - Analysis of the adjacency matrix
  - Frequent subgraph mining

Anomaly Detection in Structured Data

- Geographic/Directional Data
  - Spatial relationship between data points
  - Spatial Outlier Factor, a variant of LOF based on spatial neighborhoods
- Graph-based Data
  - Social networks, transport networks, computer networks, ecosystems, …
  - Analysis of the adjacency matrix
  - Frequent subgraph mining

Ensemble methods

- Sometimes results can be improved by using more than one method, and combining the results
- Methods must exhibit two properties:
  - Accuracy
  - Diversity
- How to combine scores in a principled way?
  - Normalisation
  - Greedy Ensemble algorithms

Summary

- There are many different approaches to anomaly detection
  - Statistical
  - Classification- and Clustering-based
  - Distance- and Density-based
  - Structured Data
- Which algorithm is best will depend on:
  - Nature of the problem we wish to solve
  - Availability of labelled data
  - Data type(s)
  - Data distribution
  - Data dimensionality

Tools for anomaly detection algorithms

ELKI Data Mining Toolkit

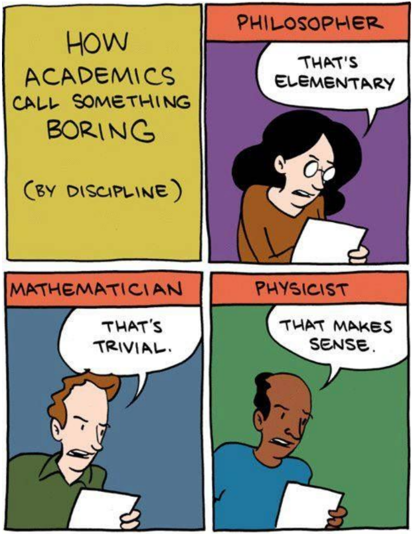
Environment for DeveLoping KDD-Applications Supported by Index-Structures

`https://elki-project.github.io/`

- Algorithms for clustering and outlier detection
- Emphasis on unsupervised methods
- Includes data index structures for performance and scalability (*e.g.* R*-tree)
- Designed to allow easy and fair evaluation and benchmarking of algorithms
- Extensible: written in Java, released under AGPLv3 license

The End

I Hope It Made Sense



References I

General Books and Papers on Anomaly Detection

1 V.Barnett and T.Lewis, *Outliers in Statistical Data*, 3rd edition, Wiley 1994

2 Markos Markou and Sameer Singh, *Novelty detection: a review—part 1: statistical approaches*, Signal Processing, Vol.83, 2003

3 Markos Markou and Sameer Singh, *Novelty detection: a review—part 2: neural network based approaches*, Signal Processing, Vol.83, 2003

4 Victoria J.Hodge and Jim Austin, *A Survey of Outlier Detection Methodologies*, Artificial Intelligence Review, Vol.22, No.2, Oct.2004

5 Varun Chandola, Arindam Banerjee and Vipin Kumar, *Anomaly Detection: A Survey*, ACM Computing Surveys, Vol.41, No.3, Jul.2009

6 Arthur Zimek, Erich Schubert and Hans-Peter Kriegel, *A Survey on Unsupervised Outlier Detection in High-Dimensional Numerical Data*, Statistical Analysis and Data Mining: The ASA Data Science Journal, Wiley 2012

References II

References for Specific Slides

Page	Reference
2	Reproduced from https://xkcd.com/1781/, under the Creative Commons Attribution–NonCommercial 2.5 License
7	Reproduced from V.Barnett and T.Lewis, *Outliers in Statistical Data*, 3rd edition, Wiley 1994
15	Reproduced from Varun Chandola, Arindam Banerjee and Vipin Kumar, *Anomaly Detection: A Survey*, ACM Computing Surveys, Vol. 41, No. 3, July 2009
40	Neal Stephenson, *The Cryptonomicon*, Arrow Books 1999
29	Reproduced from https://en.wikipedia.org/wiki/File:Warning_for_Gaussian_Distributions_(3539684489).jpg, under the Creative Commons Attribution 2.0 Generic License
34	Reproduced from https://en.wikipedia.org/wiki/File:William_Sealy_Gosset.jpg (public domain)

References III

34 Reproduced from https://en.wikipedia.org/wiki/File:Student_t_pdf.svg, under the Creative Commons Attribution 3.0 Unported License

36 Adapted from https://en.wikipedia.org/wiki/File:GaussianScatterPCA.svg, under the Creative Commons Attribution 4.0 International License

42 Adapted from https://en.wikipedia.org/wiki/File:Csv-2d-gaussian-multivarate-distributions.svg (public domain)

50 Alex Rogozhnikov, *Anomaly detection at CERN experiments*, blog post at Brilliantly Wrong, 10 May 2015

57 Reproduced from Chandola et al. 2009

58 Reproduced from Chandola et al. 2009

60 Reproduced from https://en.wikipedia.org/wiki/File:Kernel_Machine.svg, under the Creative Commons Attribution–Share Alike 4.0 International License

61 Reproduced from https://en.wikipedia.org/wiki/File:Svm_separating_hyperplanes_(SVG).svg, under the Creative Commons Attribution 3.0 Unported License

References IV

67 Adapted from https://en.wikipedia.org/wiki/File:K_Means_Example_Step_1.svg, https://en.wikipedia.org/wiki/File:K_Means_Example_Step_2.svg, https://en.wikipedia.org/wiki/File:K_Means_Example_Step_3.svg, https://en.wikipedia.org/wiki/File:K_Means_Example_Step_4.svg, under the Creative Commons Attribution 3.0 Unported License

68 Reproduced from https://en.wikipedia.org/wiki/File:ClusterAnalysis_Mouse.svg (public domain)

85–94 Adapted from Erich Schubert and Eirini Ntoutsi, *Data Mining Tutorial, Session 6: Outlier Detection*, Ludwig–Maximilians–Universität München, 2012

95–100 Markus M.Breunig, Hans-Peter Kriegel, Raymond T.Ng and Jörg Sander, *LOF: Identifying Density-based Local Outliers*, Proceedings of the 2nd ACM SIGMOD International Conference on the Management of Data—SIGMOD'00

102 V.Hautamaki, I.Karkkainen and P.Franti, *Outlier detection using k-nearest neighbour graph*, Proceedings of the 17th International Conference on Pattern Recognition, IEEE 2004

References V

J.Tang, Z.Chen, A.W.Chee Fu and D.W.Cheung, *Enhancing Effectiveness of Outlier Detections for Low Density Patterns*, Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2002

P.Sun and S.Chawla, *SLOM: A new measure for local spatial outliers*, Knowledge Information Systems 9, 4, 2006

J.X.Yu, W.Qian, H.Lu, A.Zhou, *Finding centric local outliers in categorical/numerical spaces*, Knowledge Information Systems 9, 3, 2006

107 Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft, *When Is "Nearest Neighbor" Meaningful?*, Proceedings of the 7th International Conference on Database Theory—ICDT'99

111 A.Hinneburg, C.C.Aggarwal and D.A.Keim, *What is the nearest neighbor in high dimensional spaces?*, Proceedings of the 26th International Conference on Very Large Data Bases—VLDB 2000

111 C.C.Aggarwal, A.Hinneburg and D.Keim, *On the surprising behavior of distance metrics in high dimensional space*, Proceedings of the 8th International Conference on Database Theory—ICDT 2001

References VI

112–114 Reproduced from Arthur Zimek, Erich Schubert and Hans-Peter Kriegel, *A Survey on Unsupervised Outlier Detection in High-Dimensional Numerical Data*, Statistical Analysis and Data Mining: The ASA Data Science Journal, Wiley 2012

123 Reproduced from `https://en.wikipedia.org/wiki/File:GaussianScatterPCA.svg`, under the Creative Commons Attribution 4.0 International License

124 W.B.Johnson and J.Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Conference in Modern Analysis and Probability, AMS 1984

125 Timothy de Vries and Sanjay Chawla, *Finding Local Anomalies in Very High Dimensional Space*, Proceedings of the IEEE International Conference on Data Mining—ICDM 2010

127 Hans-Peter Kriegel, Matthias Schubert and Arthur Zimek, *Angle-based Outlier Detection in High-dimensional Data*, Proceedings of the 14th SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD'08

References VII

129 Emmanuel Müller, Ira Assent, Uwe Steinhausen and Thomas Seidl, *OutRank: ranking outliers in high dimensional data*, Proceedings of the 24th International Conference on Data Engineering (ICDE) Workshop on Ranking in Databases (DBRank), 2008