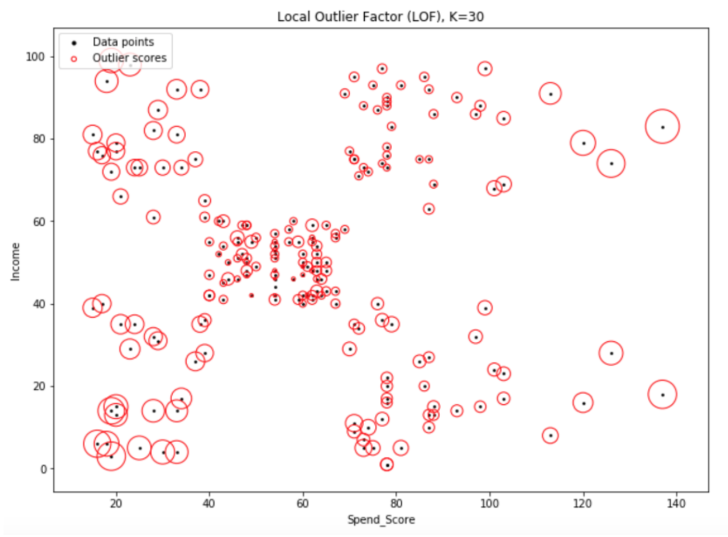
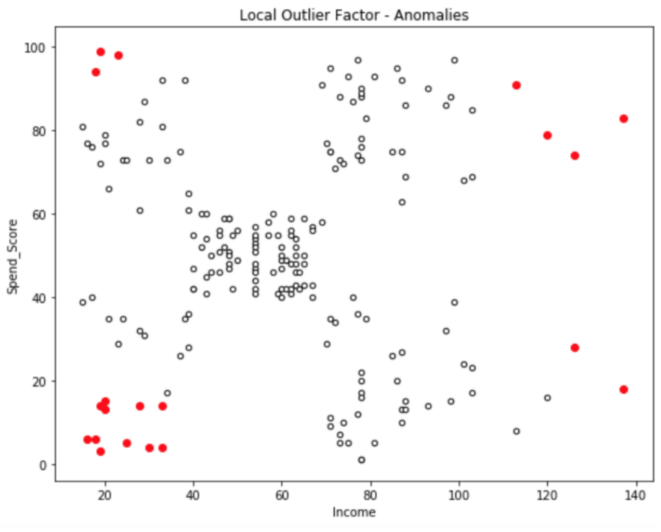



# Anomaly Detection: Local Outlier Factor (LOF)

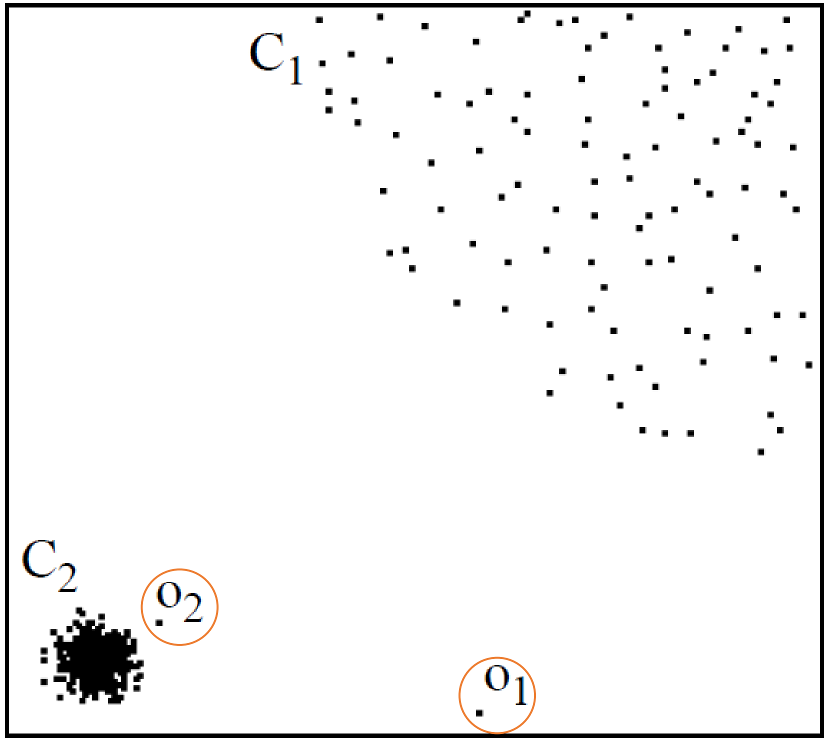
Pilsung Kang

School of Industrial Management Engineering

Korea University

# Local Outlier Factors (LOF)

- Motivation
  - ✓ Compute the novelty score of an instance by considering local density around it






고려대학교 KOREA UNIVERSITY

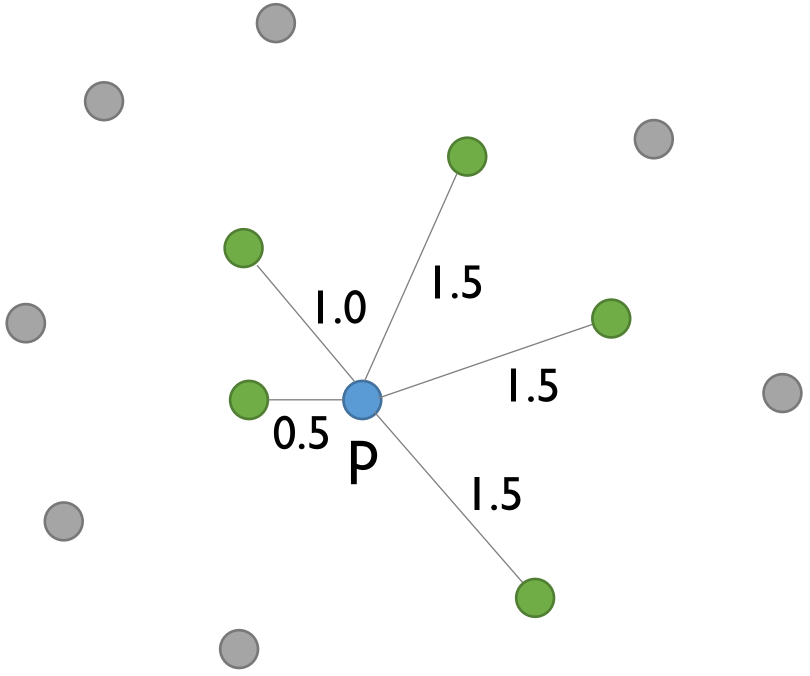
DSBA Data Science & Business Analytics

# Local Outlier Factors (LOF)

- Local Outlier Factors (LOF)
  - ✓ Definition 1: k-distance of an object p
    - For any positive integer k, the k-distance of object p, denoted as k-distance(p), is defined as the distance d(p,o) between p and an object o in D such that
    - for at least k objects o' in D\{p} it holds that $d(p, o') \leq d(p, o)$
    - for at most k-1 objects o' in D\{p} it holds that $d(p, o') < d(p, o)$
    - Simply it is the distance to the k-th nearest neighbor considering ties.

# Local Outlier Factors (LOF)

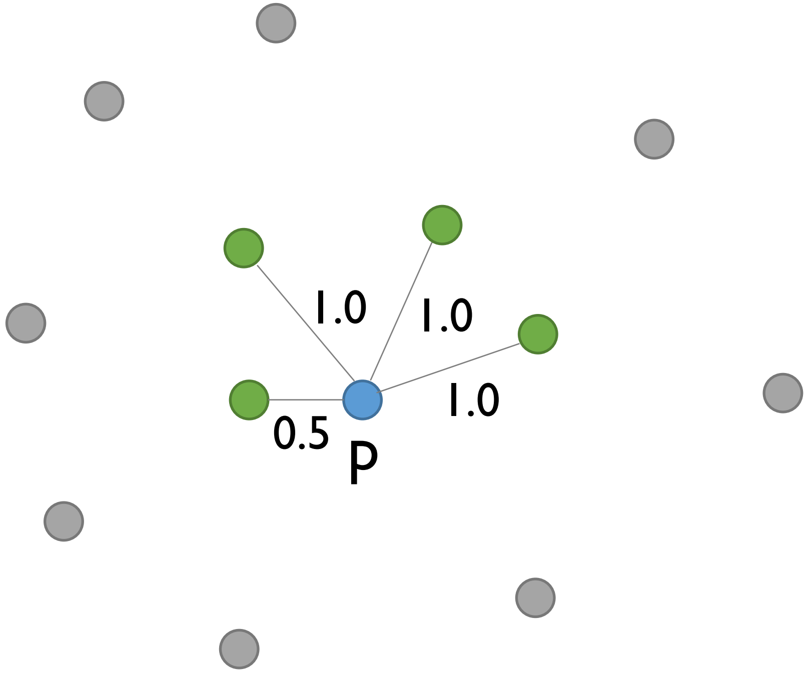
- Local Outlier Factors (LOF)
  - ✓ Definition 1: k-distance of an object p




3-distance of p = 1.5

# Local Outlier Factors (LOF)

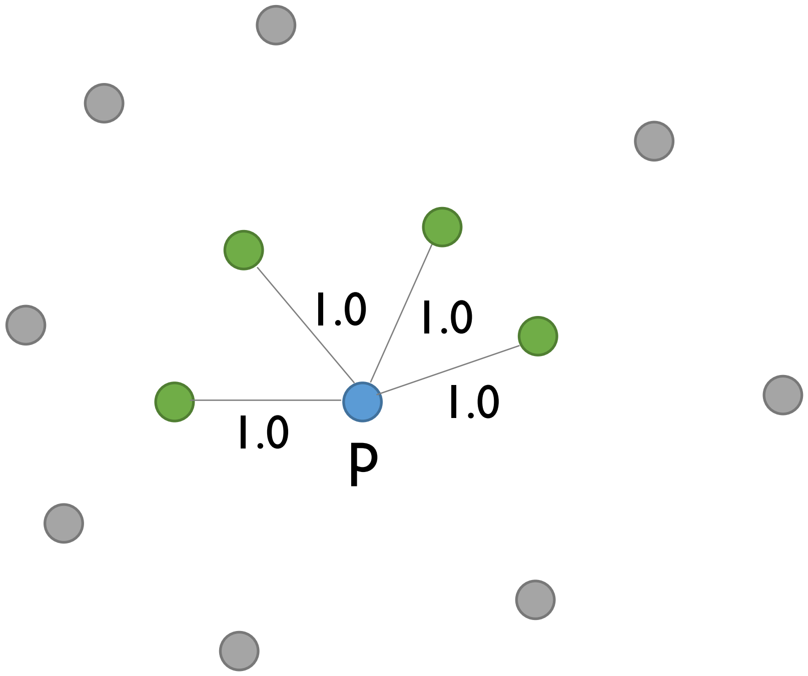
- Local Outlier Factors (LOF)
  - ✓ Definition 1: k-distance of an object p




3-distance of p = 1.0

# Local Outlier Factors (LOF)

- Local Outlier Factors (LOF)
  - ✓ Definition 1: k-distance of an object p




3-distance of p = 1.0

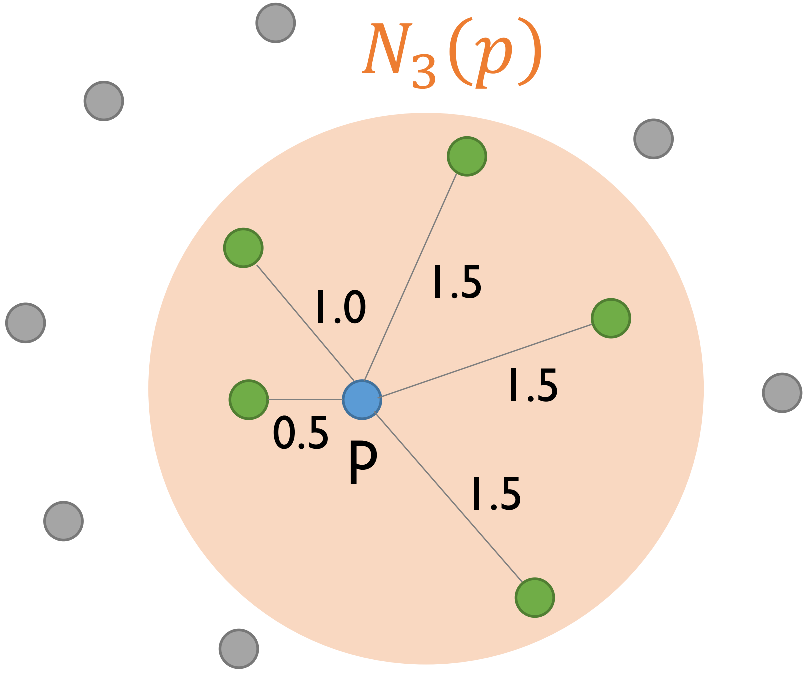
# Local Outlier Factors (LOF)

- Local Outlier Factors (LOF)
  - ✓ Definition 2: k-distance neighborhood of an object p
    - Given the k-distance of p, the k-distance neighborhood of p contains every object whose distance from p is not grater than the k-distance

$$N_k(p) = \{q \in D \backslash \{p\} \mid d(p, q) \leq k - distance(p)\}$$

# Local Outlier Factors (LOF)

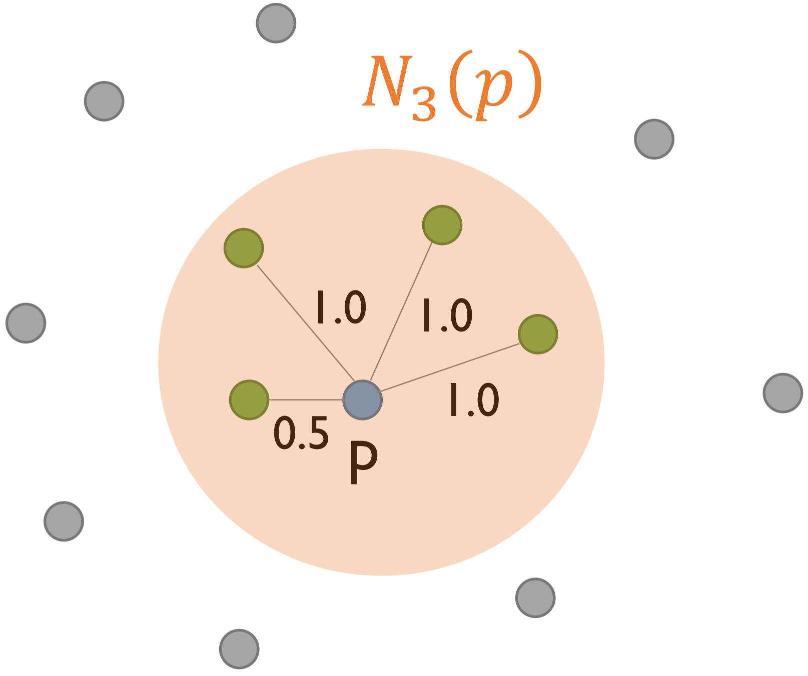
- Local Outlier Factors (LOF)
  - ✓ Definition 2: k-distance neighborhood of an object p




3-distance of p = 1.5

# Local Outlier Factors (LOF)

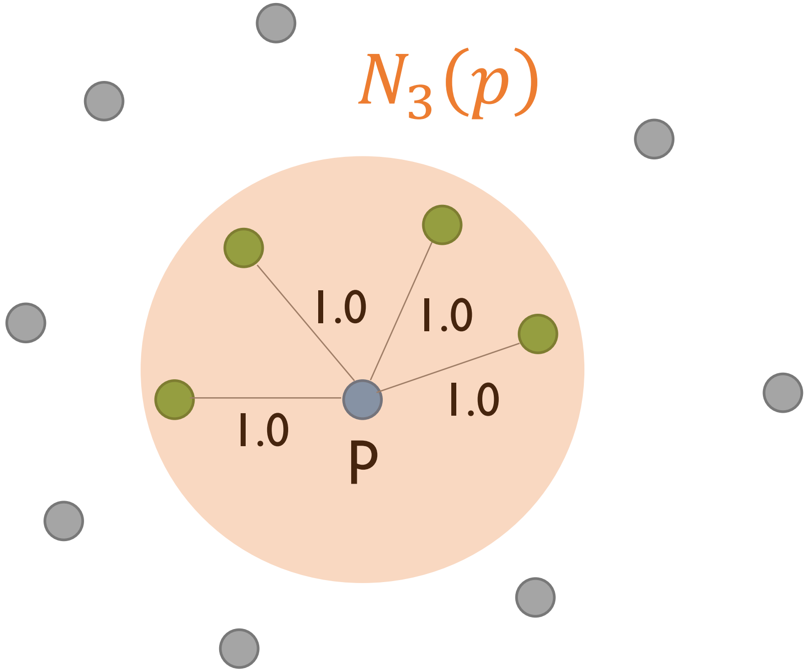
- Local Outlier Factors (LOF)
    - ✓ Definition 2: k-distance neighborhood of an object p




3-distance of p = 1.0

# Local Outlier Factors (LOF)

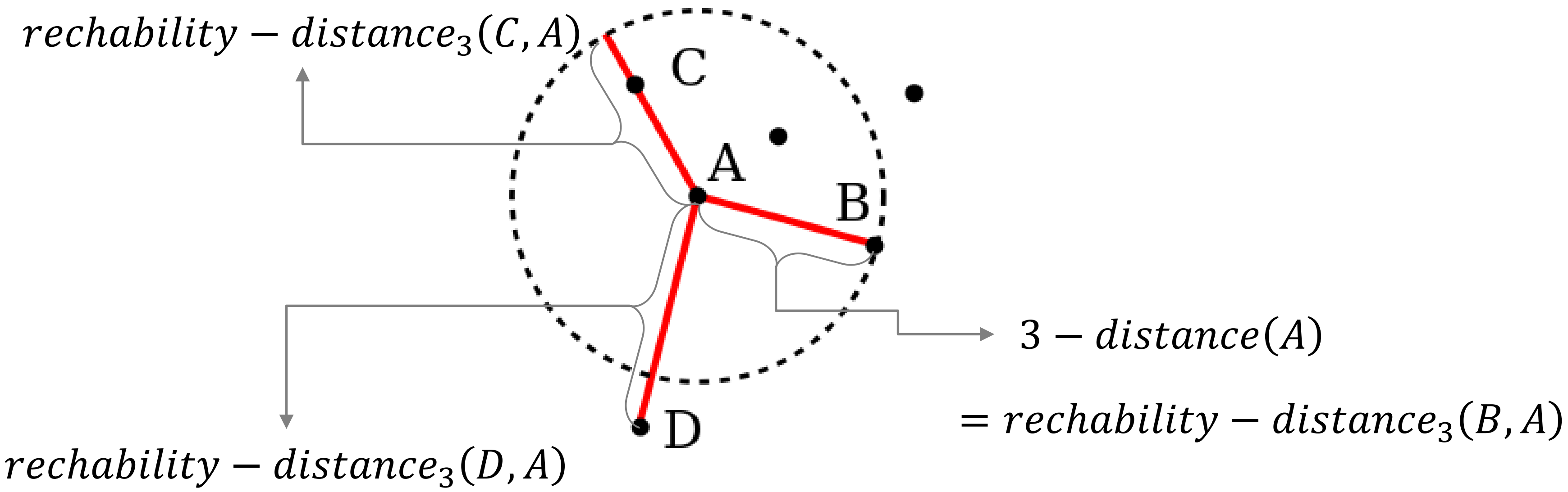
- Local Outlier Factors (LOF)
  - ✓ Definition 2: k-distance neighborhood of an object p




3-distance of p = 1.0

# Local Outlier Factors (LOF)

- Local Outlier Factors (LOF)
  - ✓ Definition 3: reachability distance
    - $rechability - distance_k(p, o) = \max\{k - distance(o), d(p, o)\}$
    - Examples



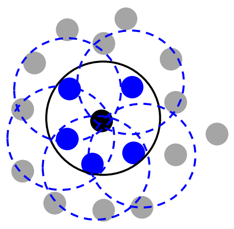
# Local Outlier Factors (LOF)

- Local Outlier Factors (LOF)
  - ✓ Definition 4: local reachability density of an object p

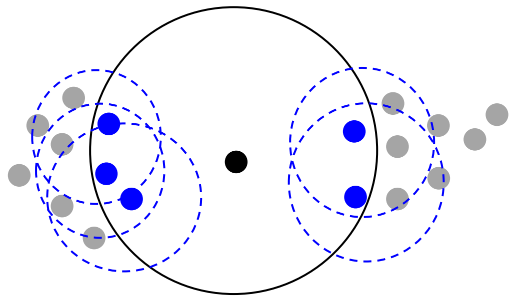
$$lrd_k(p) = \frac{|N_k(p)|}{\sum_{o \in N_k(p)} reachability - distance_k(p, o)}$$

- Case 1: p is located in the middle of a denser area: the denominator of $lrd_k(p)$ becomes small, which results in a large $lrd_k(p)$
- Case 2: p is located in a spare are between two dense data clusters: the denominator of $lrd_k(p)$ becomes large, which results in a small $lrd_k(p)$

Case 1



Case 2

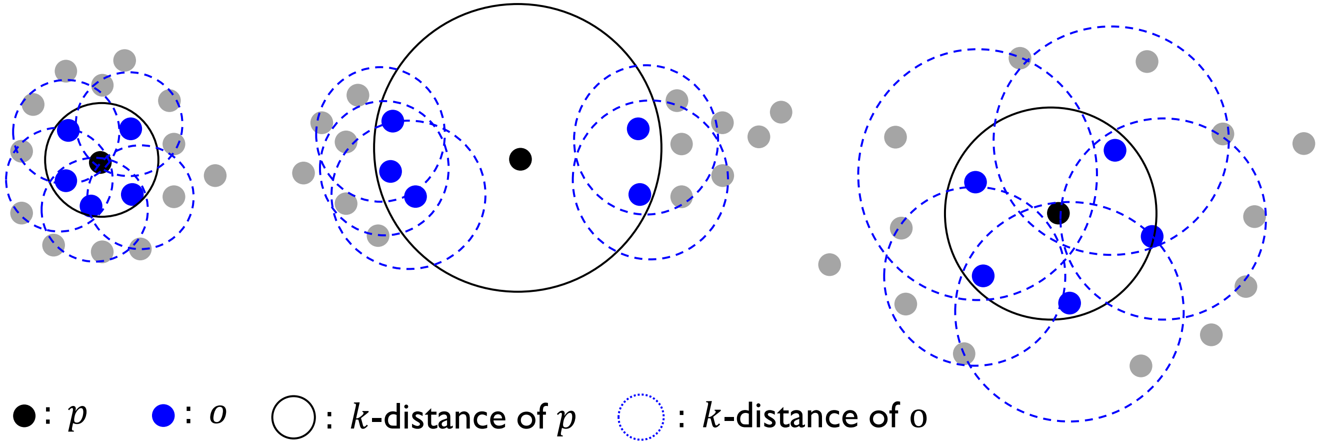


# Local Outlier Factors (LOF)

- Local Outlier Factors (LOF)
  - ✓ Definition 5: local outlier factor of an object p

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

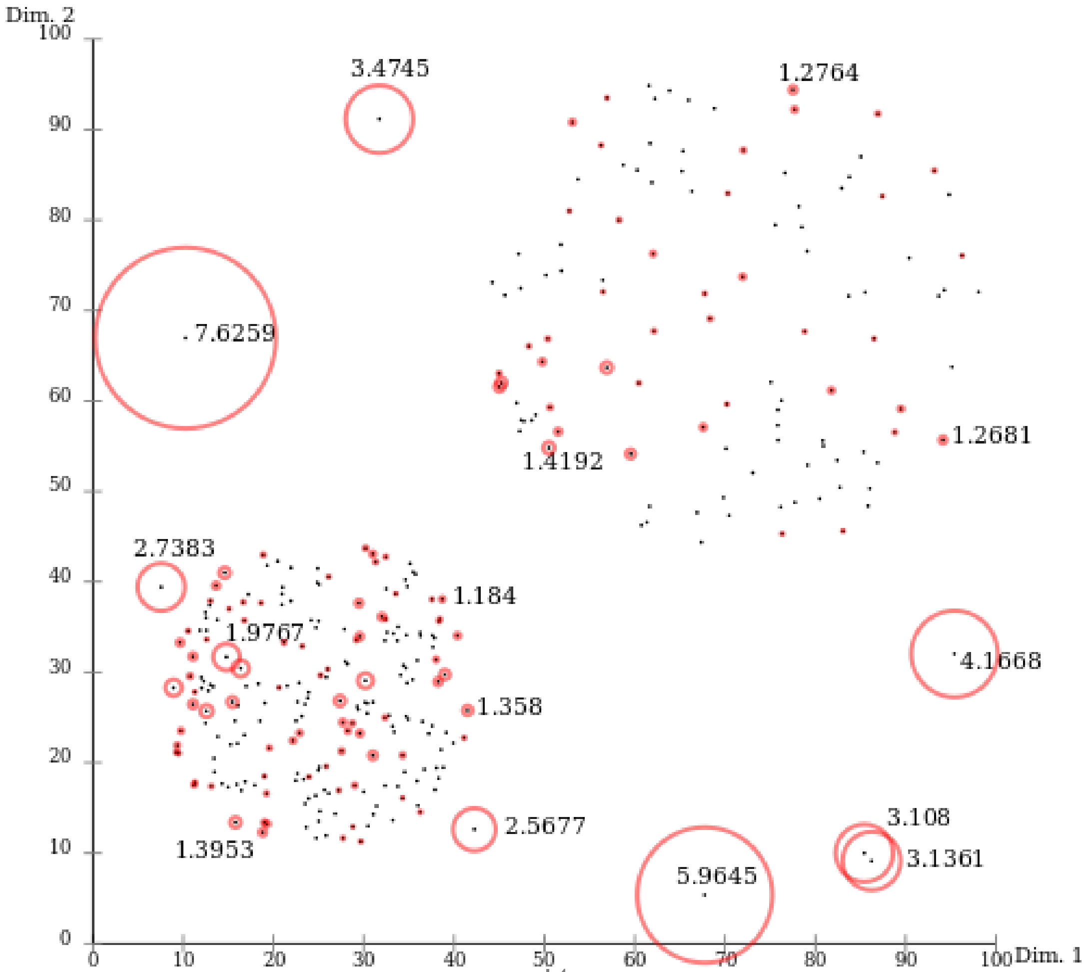
Case 1 Case 2 Case 3




| Case | $lrd_k(p)$ | $lrd_k(o)$ | $LOF_k(p)$ |
|---|---|---|---|
| Case 1 | Large | Large | Small |
| Case 2 | Small | Large | Large |
| Case 3 | Small | Small | Small |

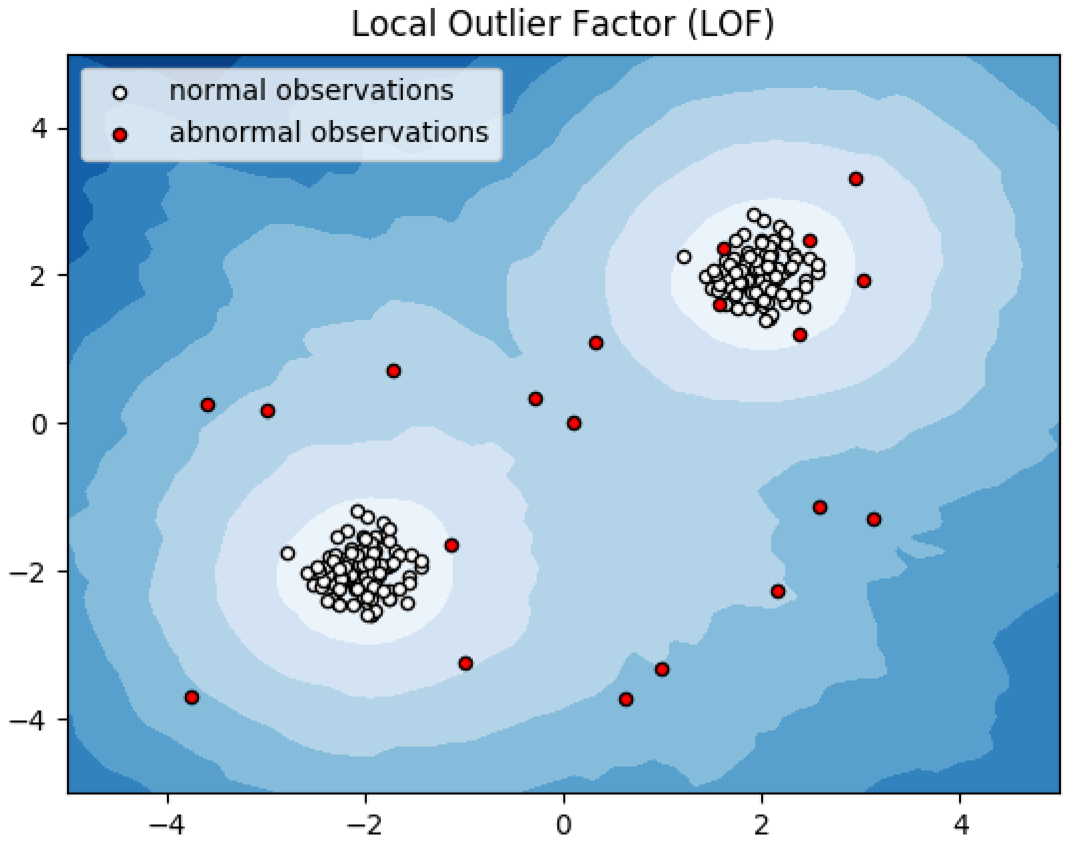
# Local Outlier Factors (LOF)

- Local Outlier Factors (LOF)
  - ✓ For each point, compute the density of its local neighborhood



# Local Outlier Factors (LOF)

- Local Outlier Factors (LOF)
  - ✓ LOF contour plot




http://scikit-learn.org/stable/auto_examples/neighbors/plot_lof.html


Q & A

# References

Research Papers


- Breunig, M.M., Kriegel, H.-P., Ng, R.T., and Sander, J. (2000). LOF: Identifying density-based local outliers. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data: 93-104.
- Chalapathy, R., Menon, A. K., & Chawla, S. (2018). Anomaly detection using one-class neural networks. arXiv preprint arXiv:1802.06360.
- Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. ACM computing surveys 41(3): 15.
- Harmeling, S. Dornhege, G., Tax, D. Meinecke, F., and Muller, K-R. (2006). From outliers to prototype: Ordering data. Neurocomputing 69(13-15): 1608-1618.
- Hariri, S., Kind, M. C., & Brunner, R. J. (2018). Extended Isolation Forest. arXiv preprint arXiv:1811.02141.
- Kang, P. and Cho, S. (2009). A hybrid novelty score and its use in keystroke dynamics-based user authentication. Pattern Recognition 42(11): 3115-3127.
- Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on (pp. 413-422). IEEE.
- Liu, F. T., Ting, K. M., & Zhou, Z. H. (2012). Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data (TKDD), 6(1), 3.
- Oza, P., & Patel, V. M. (2018). One-class convolutional neural network. IEEE Signal Processing Letters, 26(2), 277-281.
- Perera, P., Nallapati, R., & Xiang, B. (2019). Ocgan: One-class novelty detection using gans with constrained latent representations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2898-2906).

# References


Research Papers

- Pimentel, M.A.F., Clifton, D.A., Clifton, L., and Tarassenko, L. (2014). A review of novelty detection, Signal Processing 99: 215-249.
- Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., & Langs, G. (2017, June). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In International Conference on Information Processing in Medical Imaging (pp. 146-157). Springer, Cham.
- Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., and Williamson, R. C.(2001). Estimating the support of a high-dimensional distribution. Neural computation 13(7): 1443-1471.
- Tax, D.M. (2001). One-class classification, Ph.D. Thesis, Delft University of Technology, Netherlands.
- Tax, D.M. and Duin, R.P. (2004). Support vector data description. Machine learning 54(1): 45-66.

Other materials

- Pages 28-33 & 36: http://research.cs.tamu.edu/prism/lectures/pr/pr_l7.pdf
- Figures in Auto-encoder section: https://dl.dropboxusercontent.com/u/19557502/6_01_definition.pdf
- Gramfort, A. (2016). Anomaly/Novelty detection with scikit-learn: https://www.slideshare.net/agramfort/anomalynovelty-detection-with-scikitlearn