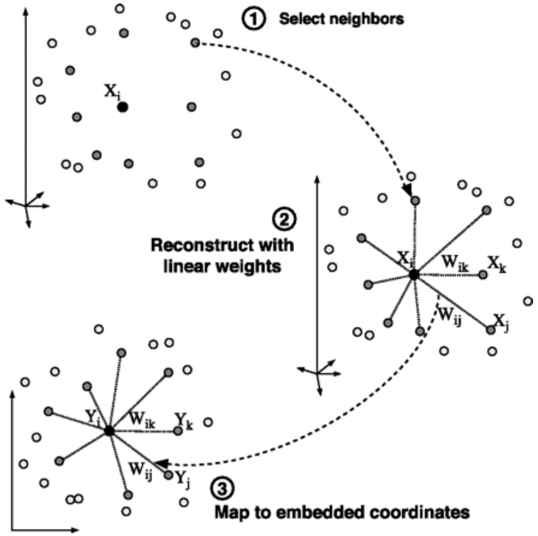
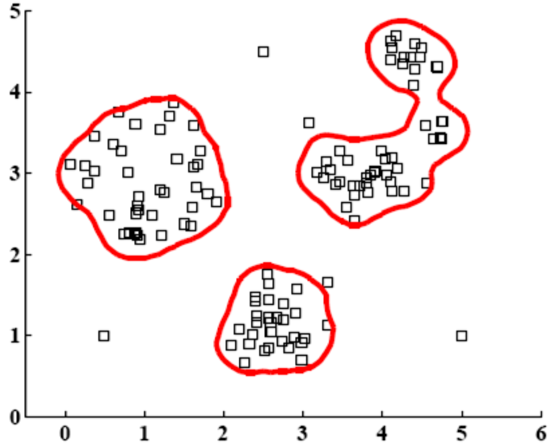




(f) d^5_{hybrid}

Anomaly Detection: Distance-based Methods

Pilsung Kang

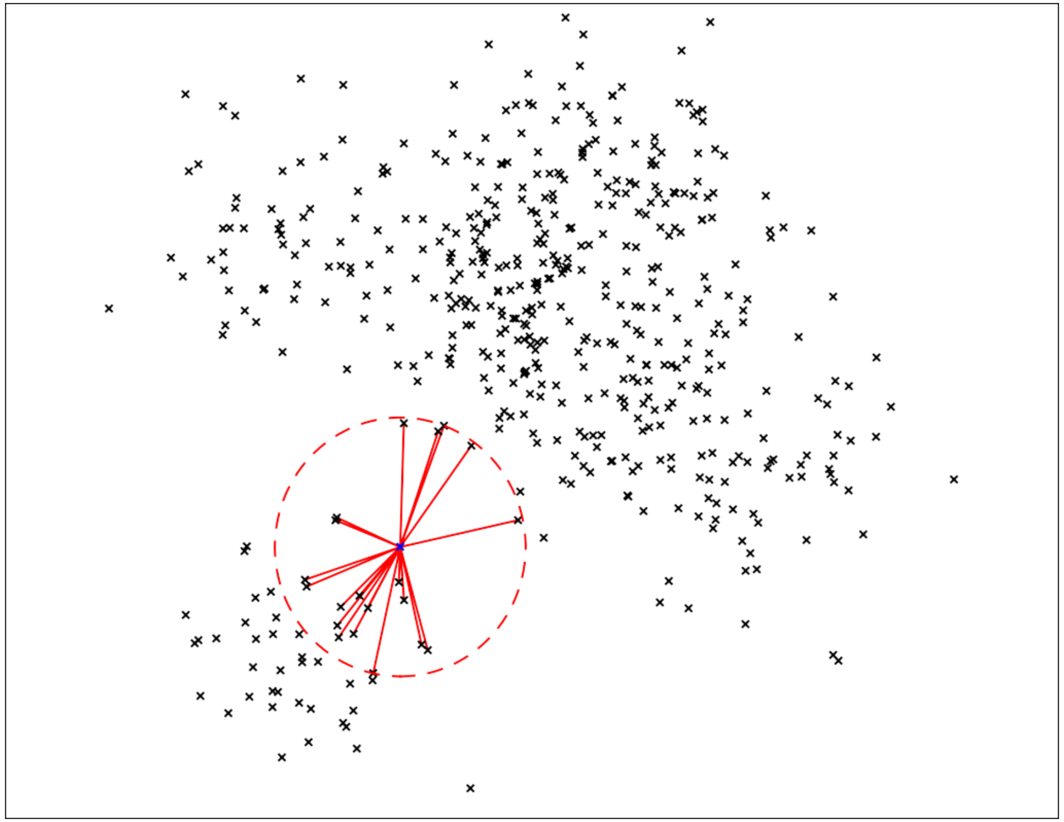
School of Industrial Management Engineering

Korea University

k-Nearest Neighbor-based Anomaly Detection

Harmeling et al. (2006)

- k-Nearest Neighbor-based Approach
  - ✓ Anomaly score of an instance is computed based on the distance information to k nearest neighbors
  - ✓ Does not assume any prior probability distribution for the normal class



https://erikbern.com/2015/09/24/nearest-neighbor-methods-vector-models-part-1.html

k-Nearest Neighbor-based Anomaly Detection

- Various distance information used for anomaly score
  - ✓ Maximum distance to the k-th nearest neighbor

$$d^k_{max} = \kappa(\mathbf{x}) = ||\mathbf{x} - z_k(\mathbf{x})||$$

  - ✓ Average distance to the k-nearest neighbors

$$d^k_{avg} = \gamma(\mathbf{x}) = \frac{1}{k}\sum_{j=1}^{k} ||\mathbf{x} - z_j(\mathbf{x})||$$

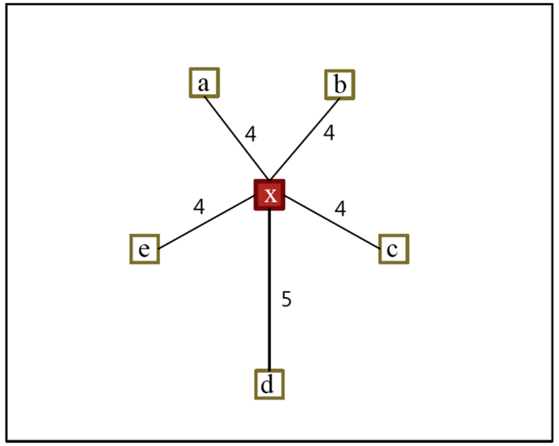
  - ✓ Distance to the mean of the k-nearest neighbors

$$d^k_{mean} = \delta(\mathbf{x}) = \left|\left|\mathbf{x} - \frac{1}{k}\sum_{j=1}^{k} z_j(\mathbf{x})\right|\right|$$

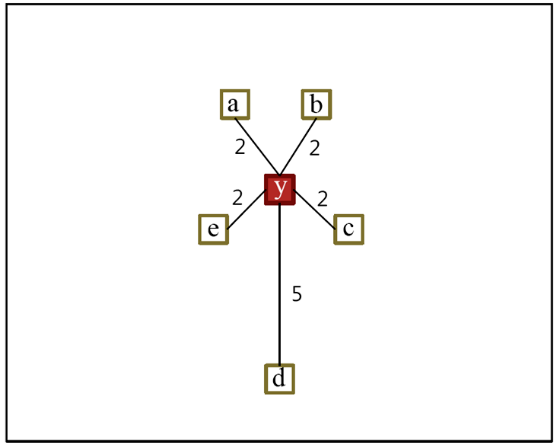
k-Nearest Neighbor-based Anomaly Detection

Kang and Cho (2009)

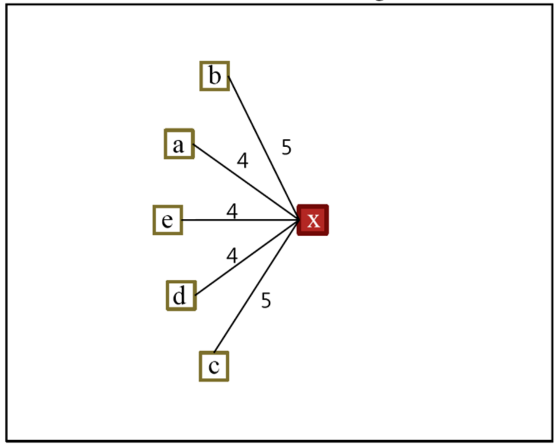
- Various distance information used for anomaly score
  - ✓ Comparison among the maximum, average, and mean distance



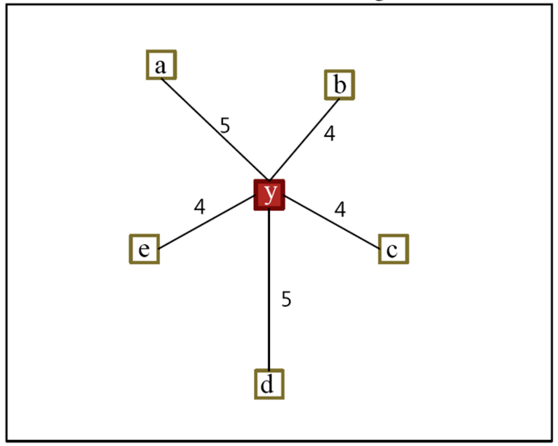
(a) $d^5_{max} = 5.0,\ d^5_{avg} = 4.2.$



(b) $d^5_{max} = 5.0,\ d^5_{avg} = 2.6.$



(c) $d^5_{avg} = 4.4,\ d^5_{mean} = 3.3.$

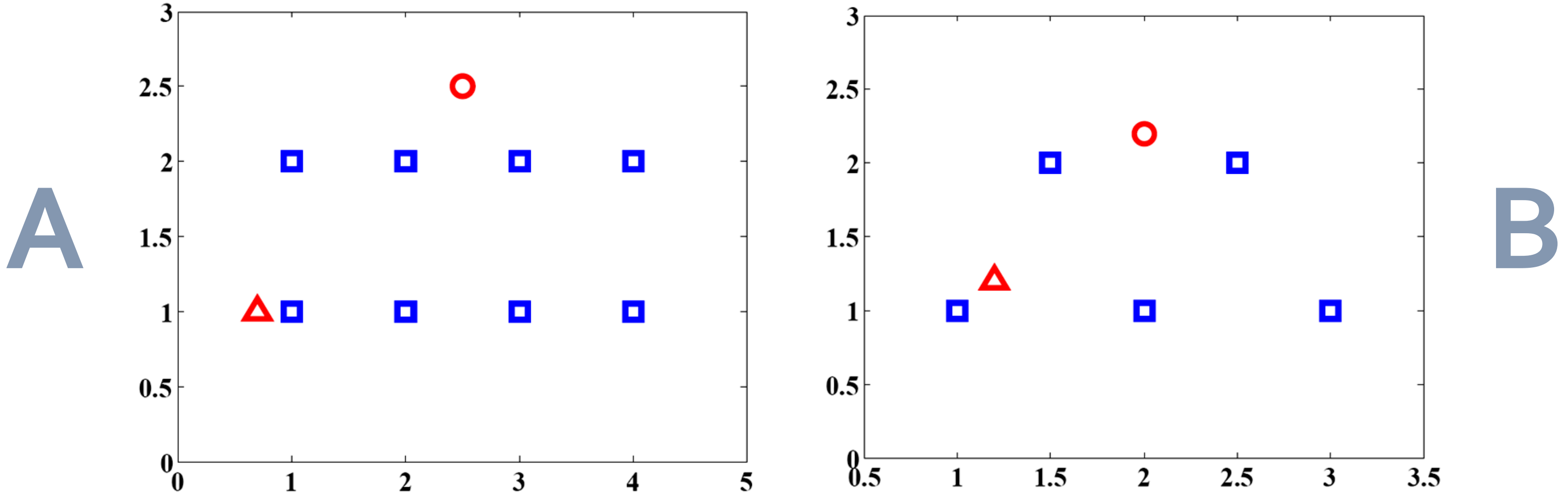


(d) $d^5_{avg} = 4.4,\ d^5_{mean} = 2.1.$

k-Nearest Neighbor-based Anomaly Detection

Kang and Cho (2009)

- Counter example of the previous anomaly scores
  - ✓ Which one should be identified as abnormal?



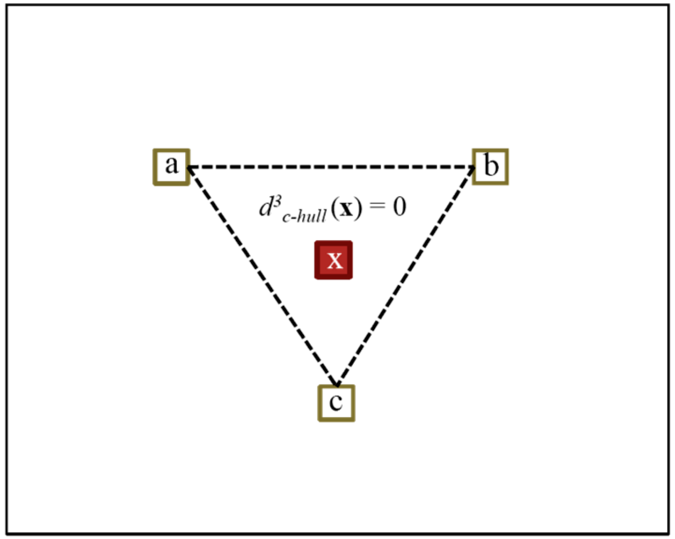
		d^k_{max}	d^k_{avg}	d^k_{mean}
A (k=4)	Circle	1.58	1.14	0.50
	Triangle	1.64	1.07	0.94
B (k=5)	Circle	1.56	1.08	0.80
	Triangle	1.86	1.09	0.88

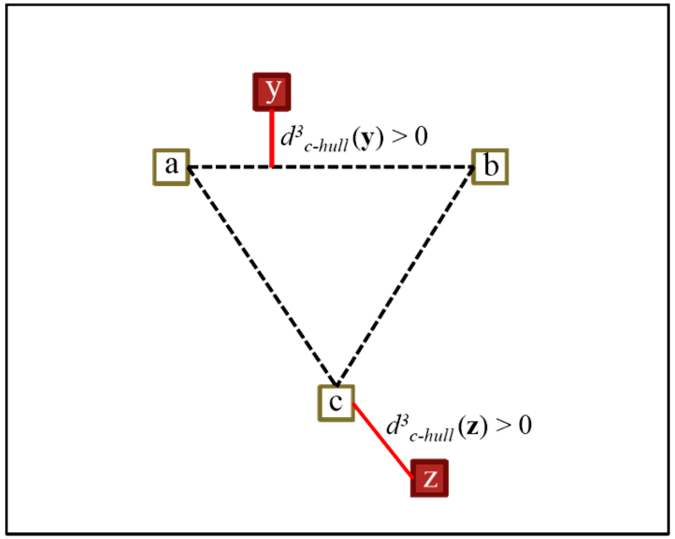
k-Nearest Neighbor-based Anomaly Detection

- Consider additional factor
  - ✓ whether the new instance is located inside the convex hull of its neighbors

$$\min_{\mathbf{w}} \left(d^k_{c-hull}(\mathbf{x})\right)^2 = \left|\left|\mathbf{x}_{new} - \sum_{j=1}^{k} \mathbf{w}_i z_j(\mathbf{x})\right|\right|^2$$

$$s.t. \sum_{i=1}^{k} \mathbf{w}_i = 1, \quad \mathbf{w}_i \geq 0, \ \forall i.$$





k-Nearest Neighbor-based Anomaly Detection

- Combine the average distance and convex distance
  - ✓ Average distance to the k-nearest neighbors

$$d^k_{avg} = \frac{1}{k}\sum_{j=1}^{k} ||\mathbf{x} - z_j(\mathbf{x})||$$

  - ✓ Convex distance to its k-nearest neighbors

$$d^k_{c-hull} = \left|\left|\mathbf{x} - \sum_{j=1}^{k} \mathbf{w}_i z_j(\mathbf{x})\right|\right|$$

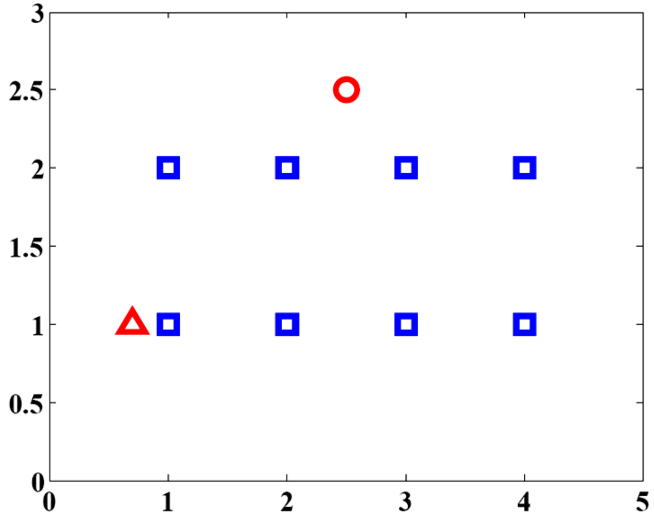
  - ✓ Put the penalty term using the convex distance for those instances located outside the convex hull of its k-nearest neighbors

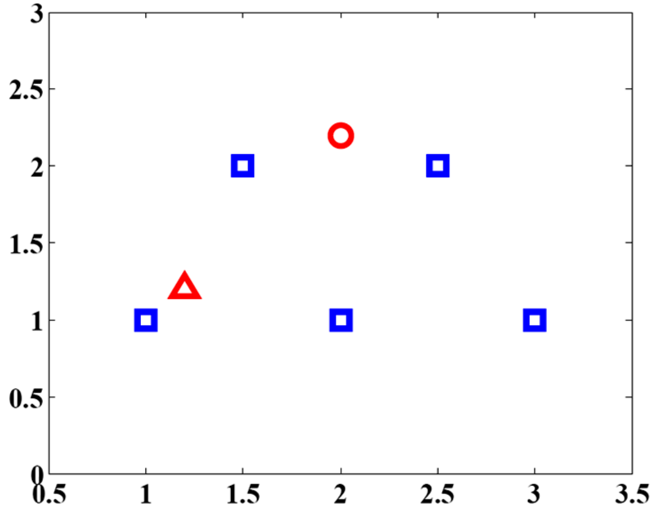
$$d^k_{hybrid} = d^k_{avg} \times \left(\frac{2}{1 + exp(-d^k_{c-hull})}\right)$$

k-Nearest Neighbor-based Anomaly Detection

- Counter example revisited

A

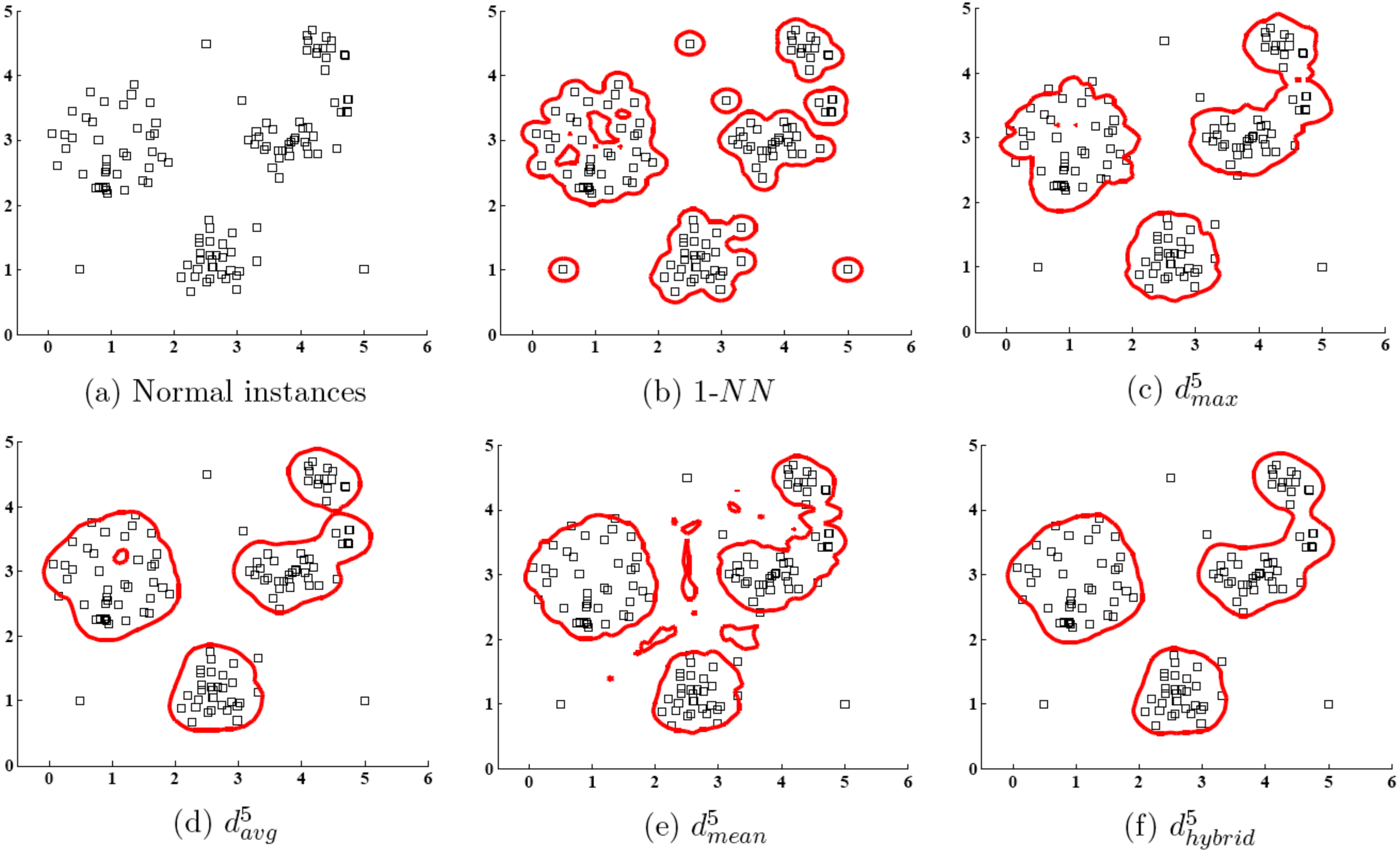




B

		d^k_{max}	d^k_{avg}	d^k_{mean}	d^k_{hybrid}
A (k=4)	Circle	1.58	1.14	0.50	1.42
	Triangle	1.64	1.07	0.94	1.18
B (k=5)	Circle	1.56	1.08	0.80	1.18
	Triangle	1.86	1.09	0.88	1.09

k-Nearest Neighbor-based Anomaly Detection



k-Nearest Neighbor-based Anomaly Detection

- Experiment
  - ✓ Datasets

No.	Name	Source	Class	Dim.	TrN_n	TsN_n	TsN_o
1	Banana	Rätch	-1	2	216	2,708	271
2	Titanic	Rätch	-1	3	100	1,390	139
3	Liver	UCI	healthy	6	73	72	7
4	Ecoli	UCI	cp	7	72	71	7
5	Yeast	UCI	0	8	232	231	23
6	Pima	UCI	0	8	250	250	25
7	Diabetes	Rätch	-1	8	304	196	20
8	Glass	UCI	1	9	35	35	4
9	Breast	Rätch	-1	9	142	54	5
10	Flare	Rätch	-1	9	300	178	18
11	Heart	Rätch	-1	13	94	56	6
12	Image	Rätch	-1	18	554	436	44
13	Twonorm	Rätch	-1	20	198	3,499	350
14	German	Rätch	-1	20	489	211	21
15	Waveform	Rätch	-1	21	268	3,085	308
16	Parkinsons	UCI	parkinsons	22	74	73	7
17	Ionosphere	UCI	0	33	113	112	11
18	Spectf	UCI	0	44	28	27	3
19	Sonar	UCI	mine	60	56	55	6
20	Ozone	UCI	non-ozone	72	29	28	3
21	Arrhythmia	UCI	normal	258	119	118	12

k-Nearest Neighbor-based Anomaly Detection

- Performance (in terms of the Integrated Error)

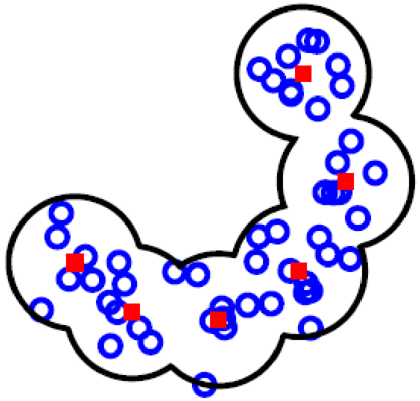
Data	Dim.	TrN$_n$	Gauss	MoG	Parzen	1-SVM	KMC	KCC	HC	PCA	d^k_{max}	d^k_{avg}	d^k_{mean}	1-NN	MST-CD	d^k_{hybrid}
Titanic	3	100	19.12	19.12	18.50	17.27	21.12	21.26	22.80	22.48	3.64	3.53	10.00	8.73	**1.33***	**1.33***
Liver	6	73	44.41	45.04	41.11	38.82	41.90	43.41	40.91	40.68	40.00	38.85	38.96	39.81	39.14	**38.18**
Ecoli	7	72	3.61	2.58	3.14	2.35	2.45	3.38	3.88	20.12	2.22	2.13	2.84	3.94	2.67	**2.11**
Glass	9	35	18.82	18.29	22.25	17.43	18.61	22.61	24.42	22.21	13.86	12.25	12.36	18.93	11.54	**11.39**
Breast	9	142	35.83	31.51	32.04	29.64	34.68	40.68	32.80	31.24	29.58	28.84	29.53	34.62	33.18	**26.99***
Banana	2	216	54.91	8.56	7.96	8.30	16.93	12.53	**7.65**	42.55	8.51	8.08	9.87	10.57	11.67	7.89
Yeast	8	232	31.55	28.98	27.99	26.58	28.78	33.25	27.32	31.47	25.81	24.54	25.87	27.79	26.50	**23.40**
Pima	8	250	29.86	33.55	26.04	27.50	29.60	32.69	28.46	33.28	24.82	24.57	27.68	28.17	27.72	**24.45**
Diabetes	8	304	30.61	34.68	27.35	26.60	28.80	35.31	26.45	31.32	23.70	23.29	25.68	26.66	28.21	**23.64**
Flare	9	300	23.19	23.19	24.82	15.40	29.67	26.65	25.57	26.76	10.49	9.74	17.09	5.62	**5.47**	6.14
Spectf	44	28	28.33	16.42	21.11	14.75	15.00	28.02	16.36	26.30	14.20	13.40	12.96	17.16	13.33	**11.67***
Sonar	60	56	41.59	37.27	34.32	33.80	41.55	40.06	40.07	42.32	39.65	34.67	32.12	33.73	**31.30**	32.62
Ozone	72	29	23.99	14.64	13.15	12.50	13.51	19.11	16.49	34.46	11.07	10.71	**9.76**	14.11	12.86	10.30
Arrhythmia	258	119	28.01	40.25	28.03	25.79	25.38	28.62	27.42	28.17	26.10	26.01	25.93	26.16	24.57	**23.93**
Heart	13	94	21.13	19.05	20.58	18.61	19.84	20.79	18.22	20.21	15.23	14.75	15.72	23.08	23.24	**14.30**
Image	18	554	13.11	13.61	11.79	10.38	23.22	30.13	31.21	15.87	13.80	11.99	10.32	10.53	**9.19***	11.04
Twonorm	20	198	9.83	11.80	11.62	9.48	8.82	12.38	**6.13***	9.62	9.62	10.02	10.79	12.87	12.62	10.11
German	20	489	38.16	36.99	37.24	35.72	39.23	42.79	40.56	37.73	34.72	33.95	34.59	37.88	36.35	**33.77**
Waveform	21	268	42.14	30.25	26.33	**22.65***	27.56	31.26	25.44	43.07	23.75	24.58	27.69	29.05	27.09	25.24
Parkinsons	22	74	32.71	46.35	32.14	**29.16**	32.65	32.66	32.79	34.60	36.95	32.54	31.63	34.22	30.68	30.30
Ionosphere	33	113	4.44	4.93	4.16	2.80	3.54	4.40	4.39	3.68	2.70	2.76	2.75	4.00	**2.70**	2.72

Clustering-based Approach

- K-Means clustering-based anomaly detection
  - ✓ anomaly score of an instance is computed based on the distance information to the nearest centroid
  - ✓ Does not assume any prior probability distribution for the normal class

$$\mathcal{X} = C_1 \cup C_2 \ldots \cup C_K, \quad C_i \cap C_j = \phi, \quad i \neq j.$$

$$\arg\min_{\mathbf{C}} \sum_{i=1}^{K} \sum_{\mathbf{x}_j \in C_i} ||\mathbf{x}_j - \mathbf{c}_i||^2$$

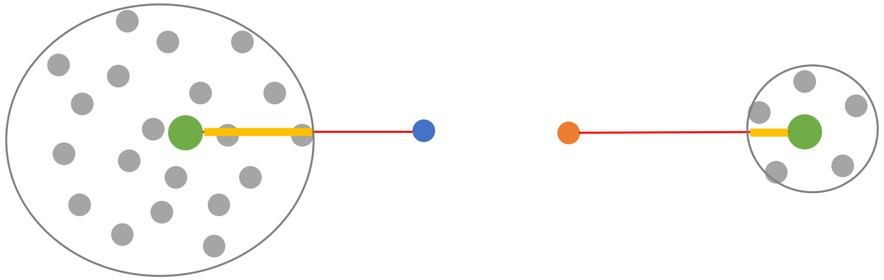


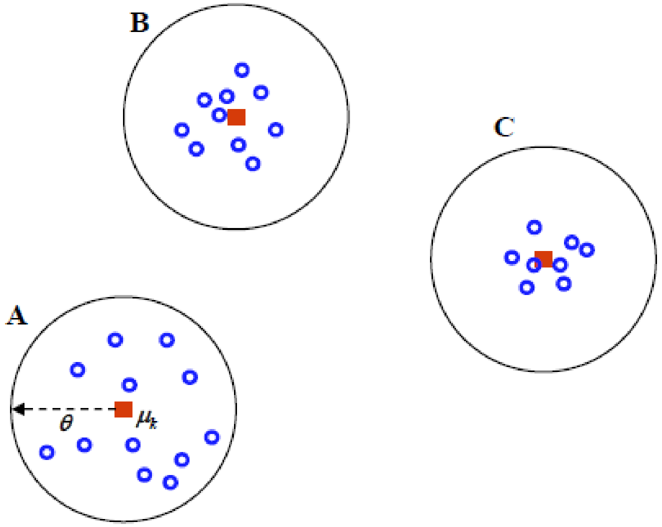
  - ✓ EM algorithm for K-Means clustering

1: Select K points as the initial centroids.
2: **repeat**
3: Form K clusters by assigning all points to the closest centroid.
4: Recompute the centroid of each cluster.
5: **until** The centroids don't change

Clustering-based Approach

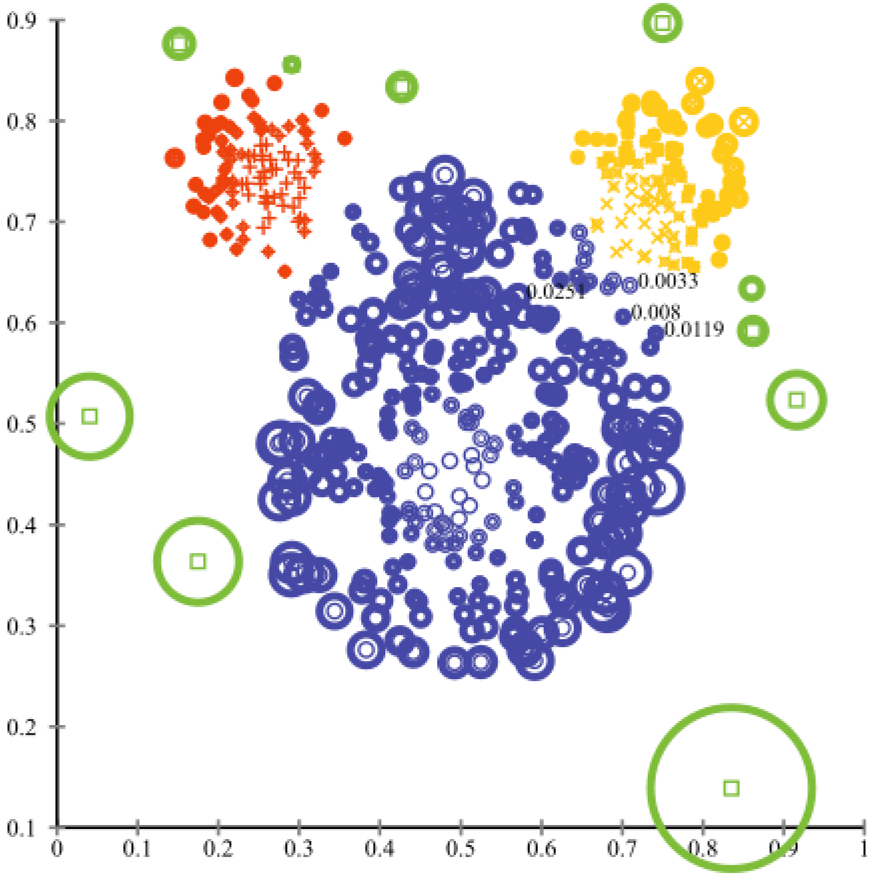
- Clustering-based Approach
  - ✓ Two anomaly scores by KMC
    - Absolute distance to the nearest centroid
    - Relative distance to the nearest centroid





Clustering-based Approach

- KMC-based anomaly score: Example



Principal Component Analysis-based Anomaly Detection

- PCA revisited
  - ✓ Purpose: maximize the variance after projection

$$\max \quad \mathbf{w}^T\mathbf{S}\mathbf{w}$$
$$s.t. \quad \mathbf{w}^T\mathbf{w} = 1$$

  - ✓ Solution

$$L = \mathbf{w}^T\mathbf{S}\mathbf{w} - \lambda(\mathbf{w}^T\mathbf{w} - 1)$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{S}\mathbf{w} - \lambda\mathbf{w} = 0 \Rightarrow (\mathbf{S} - \lambda\mathbf{I})\mathbf{w} = 0$$

Principal Component Analysis-based Anomaly Detection

- PCA as an anomaly detector
  - ✓ Anomaly score: the amount of reconstruction loss from the projected space into the original space

$$\underset{(d \text{ by } n)}{\mathbf{X}} \xrightarrow{\text{Projection}} \underset{(1 \text{ by } d)\,(d \text{ by } n)}{\mathbf{w}^T\mathbf{X}} \xrightarrow{\text{Reconstruction}} \underset{(d \text{ by } 1)(1 \text{ by } d)\,(d \text{ by } n)}{\mathbf{w}\mathbf{w}^T\mathbf{X}}$$

$\mathbf{x}_1$	0.69	-1.31	0.39	0.09	1.29	0.49	0.19	-0.81	-0.31	-0.71
$\mathbf{x}_2$	0.49	-1.21	0.99	0.29	1.09	0.79	-0.31	-0.81	-0.31	-1.01
$\mathbf{z}_1$	0.83	-1.78	0.99	0.27	1.68	0.91	-0.10	-1.14	-0.44	-1.22
$\mathbf{x}'_1$	0.56	-1.21	0.67	0.19	1.14	0.62	-0.07	-0.78	-0.30	-0.83
$\mathbf{x}'_2$	0.61	-1.31	0.73	0.20	1.23	0.67	-0.07	-0.84	-0.32	-0.90

$\mathbf{w}^T\mathbf{X}$

$\mathbf{w}\mathbf{w}^T\mathbf{X}$

고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics

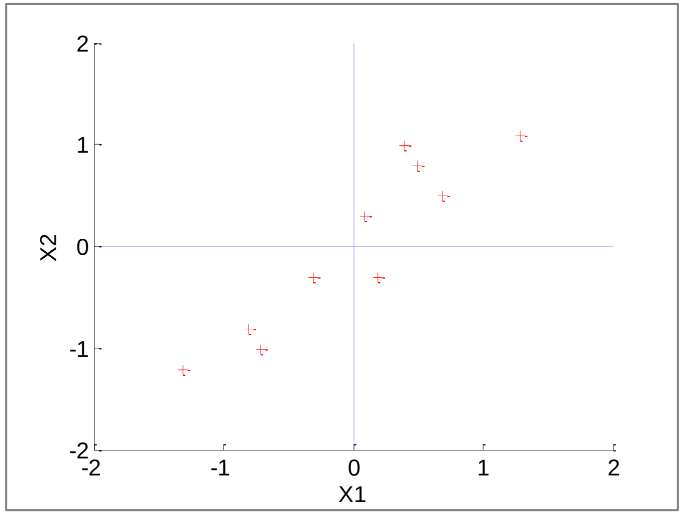
Principal Component Analysis-based Anomaly Detection

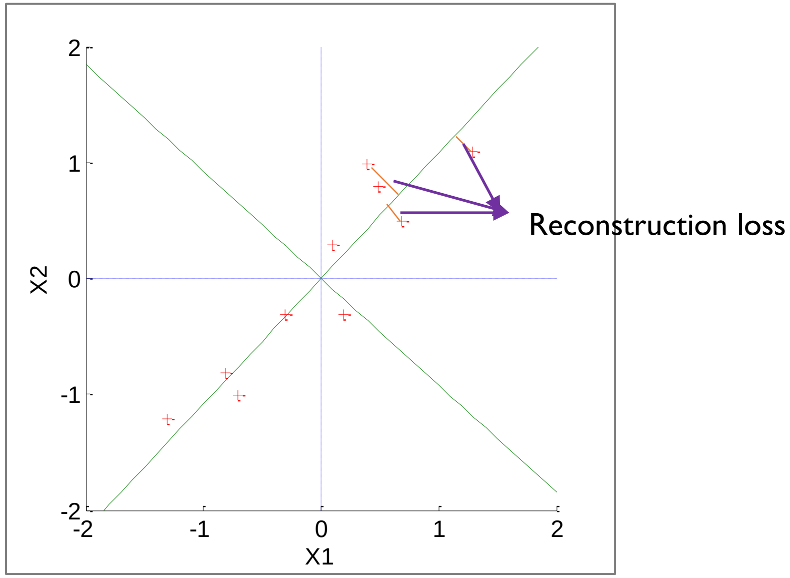
- PCA as an anomaly detector
  - ✓ Compute the reconstruction loss

$$\mathrm{error}(\boldsymbol{x}) = \left\|\mathbf{x} - \mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{x}\right\|^2 = \left(\mathbf{x} - \mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{x}\right)^{\mathrm{T}}\left(\mathbf{x} - \mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{x}\right)$$

$$= \mathbf{x}^{\mathrm{T}}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{x}$$

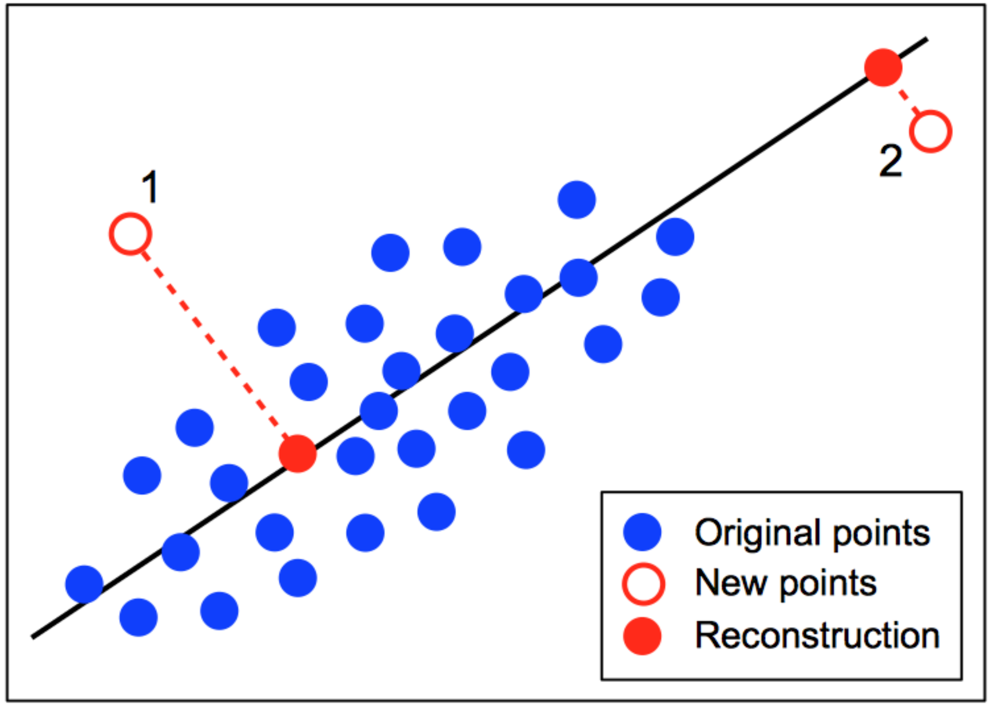
$$= \mathbf{x}^{\mathrm{T}}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{x} = \|\mathbf{x}\|^2 - \left\|\mathbf{w}^{\mathrm{T}}\mathbf{x}\right\|^2$$

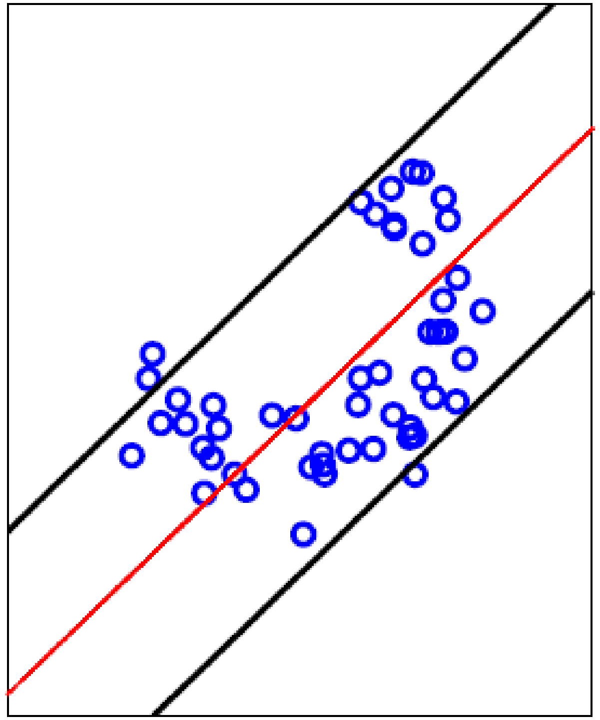




Principal Component Analysis-based Anomaly Detection

- PCA as an anomaly detector
  - ✓ Graphical interpretation





https://stats.stackexchange.com/questions/259806/anomaly-detection-using-pca-reconstruction-error


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