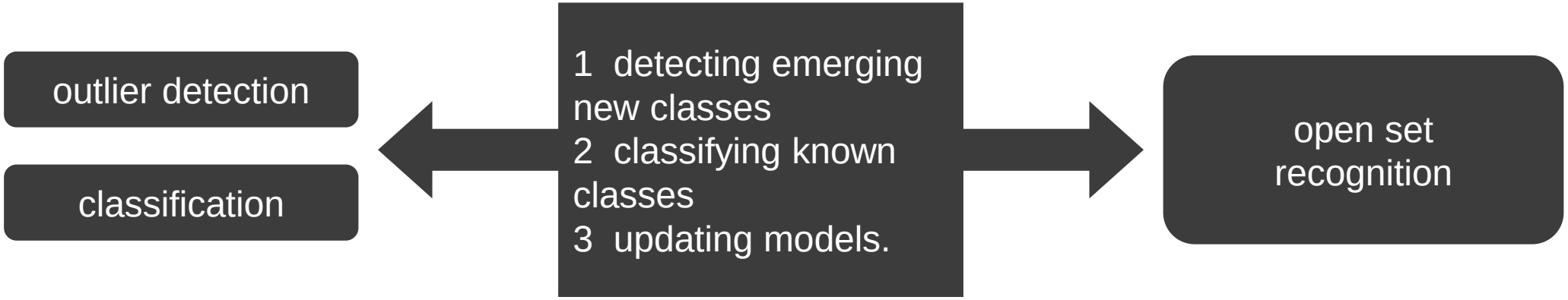
outlier

2018.4.26

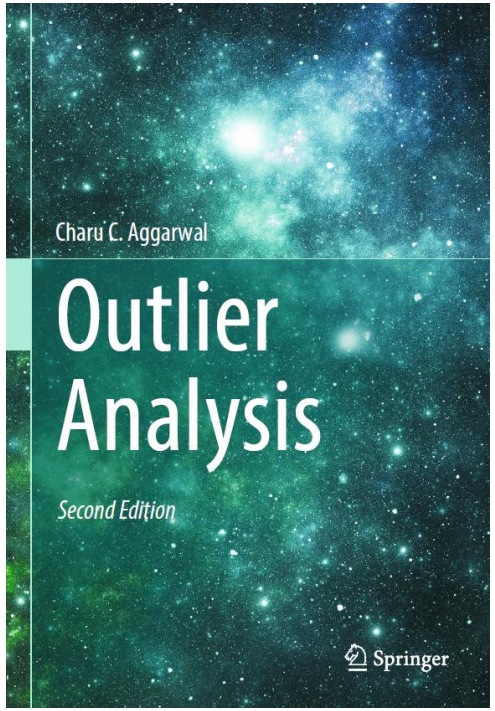
Question

Difference between outlier detection and open set recognition?



将已有异常检测方法“**”为既能检测离群值又同时可完成多分类任务

Example: Classification Under Streaming Emerging New Classes: A Solution Using Completely-Random Trees Xin Mu, Kai Ming Ting, and Zhi-Hua Zhou TKDE2017

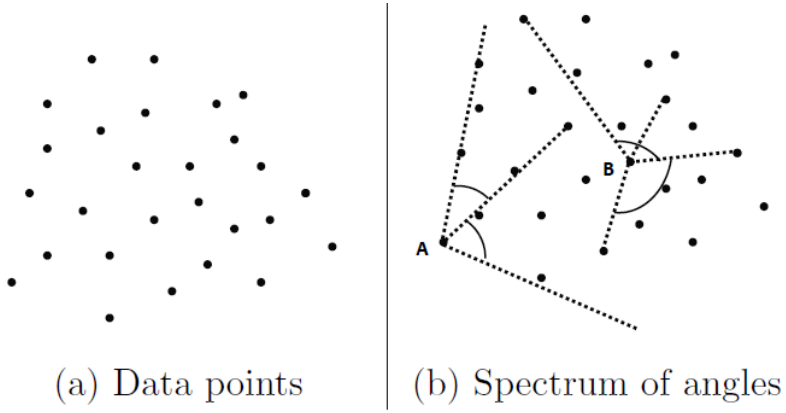


Semi-Supervised

Transductive

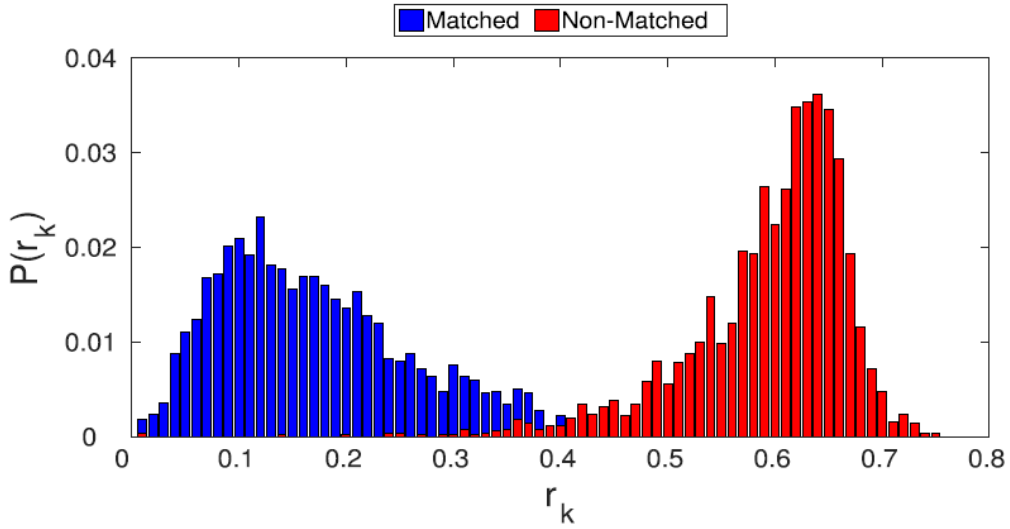
outlier

Extreme-Value Analysis

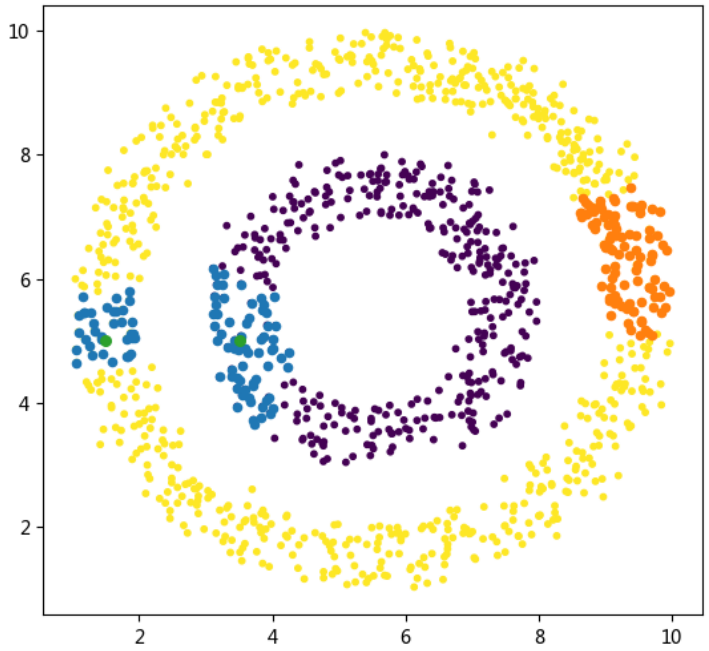


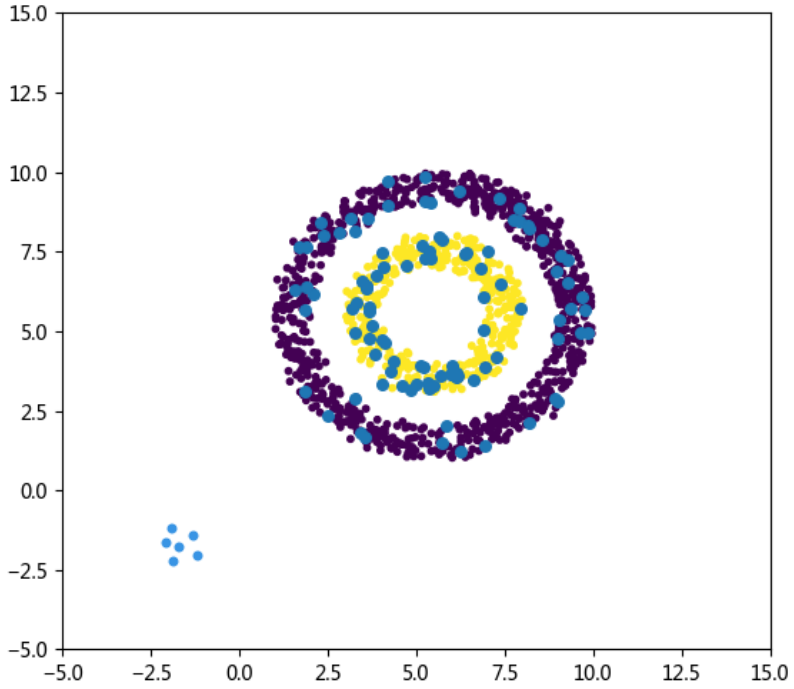
(a) Data points (b) Spectrum of angles

Figure 2.6: Angle-based outlier detection



(PAMI 2017)

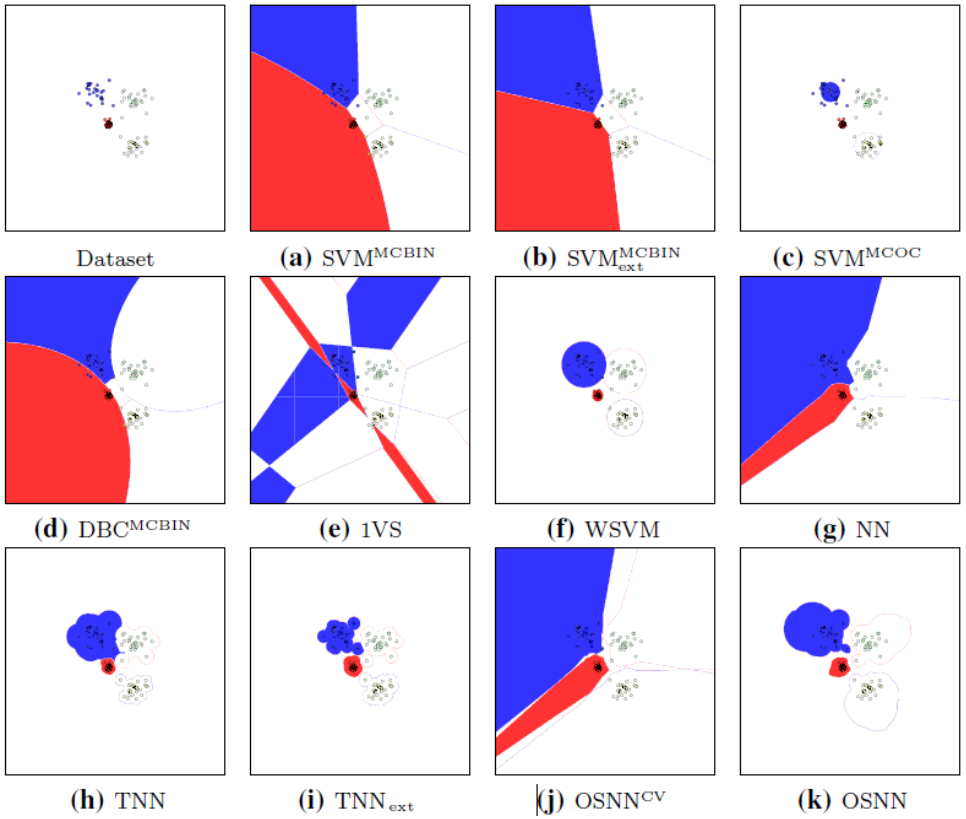


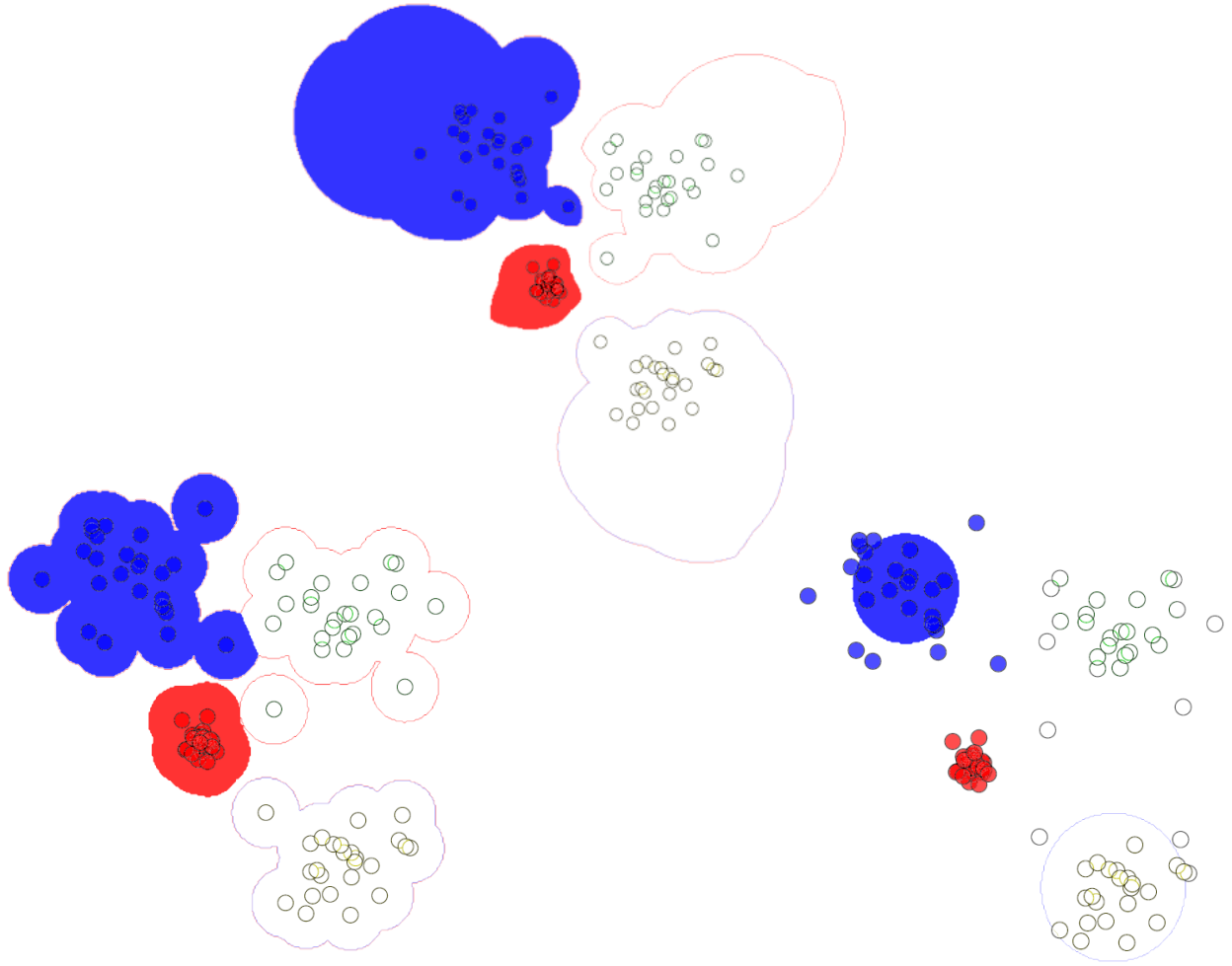


Two worries

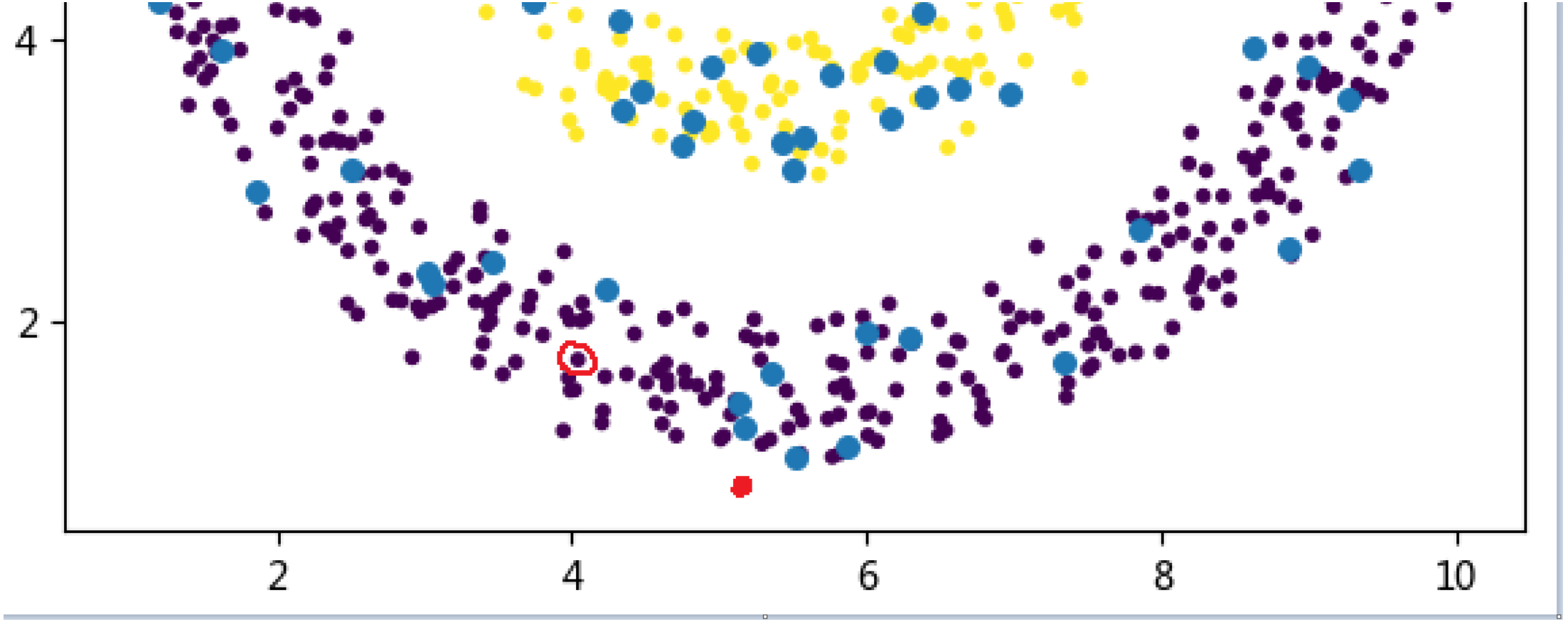
1. How to balance specific and generalization?

(Machine Learning 2017)

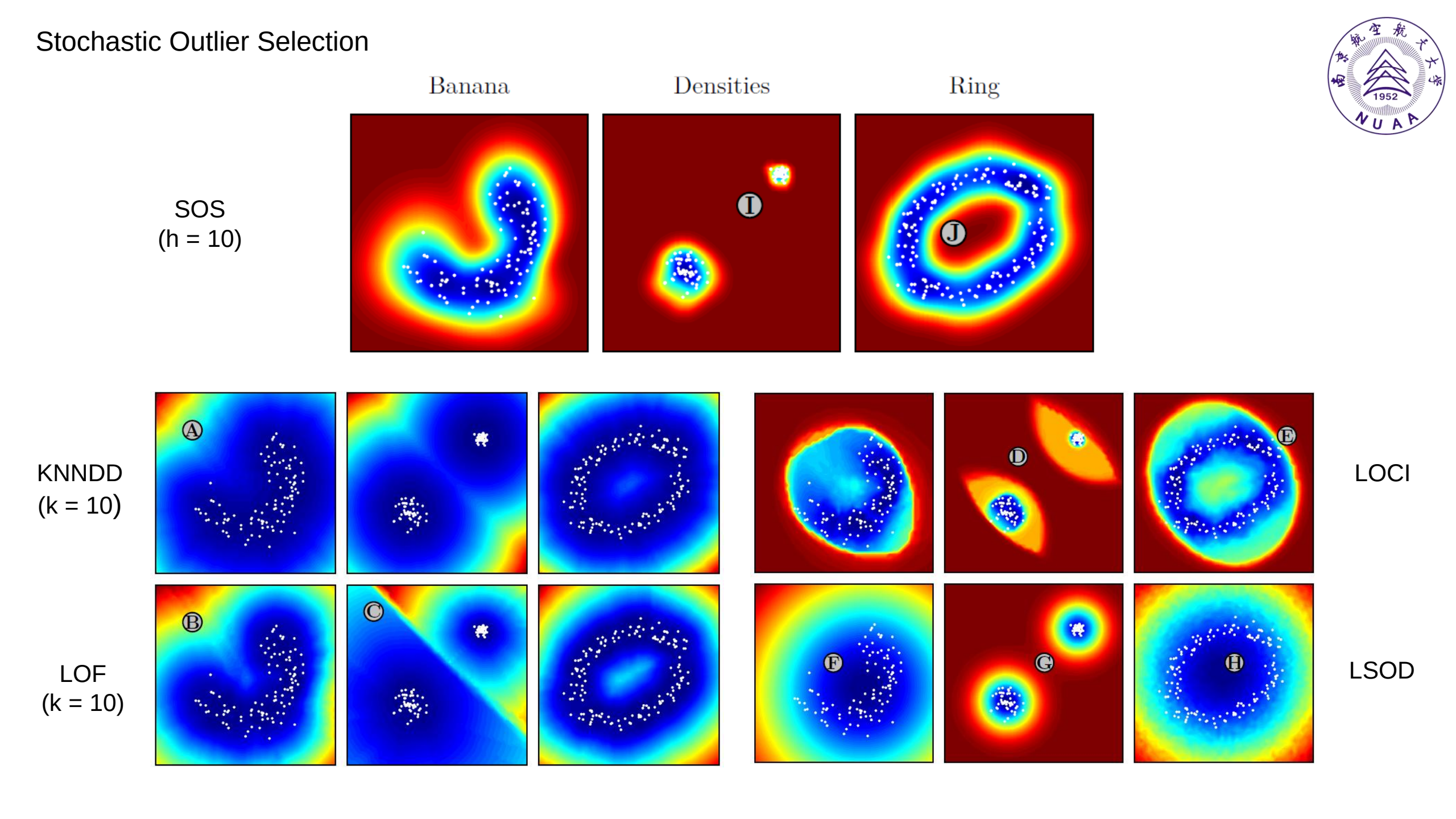


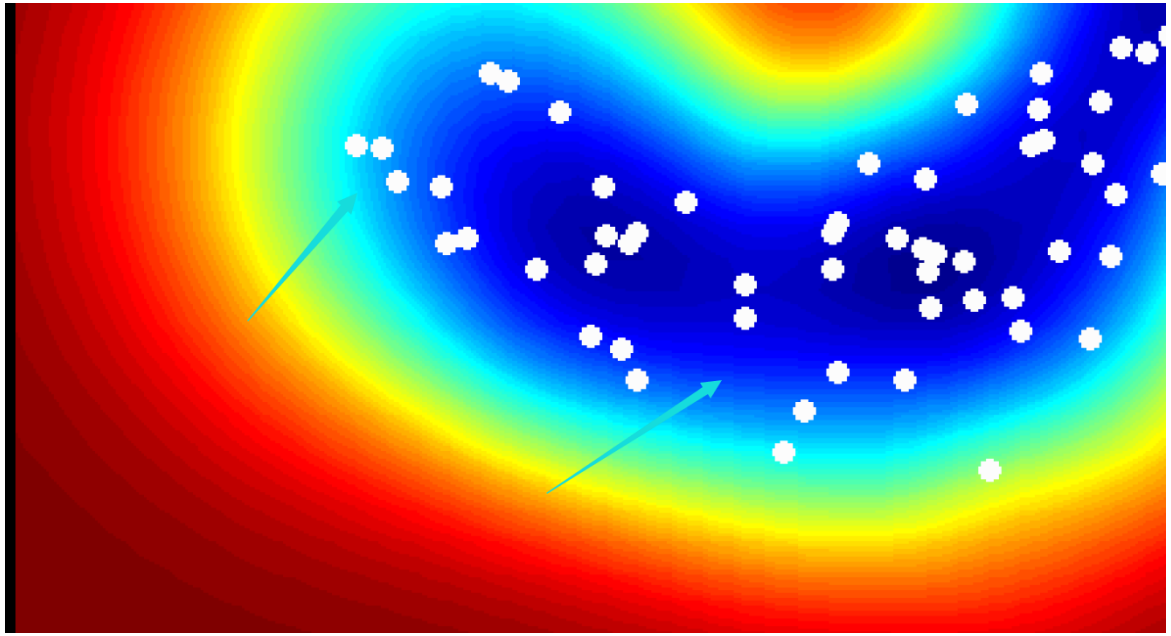


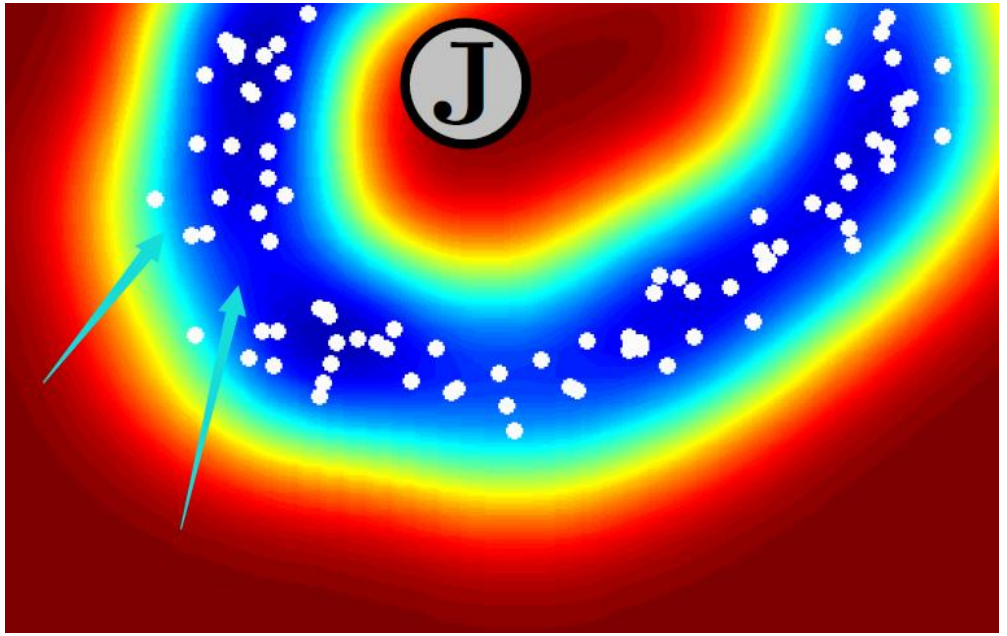
2. Smoothness assumption used in semi-supervised classification tasks, is often both local and global.



Stochastic Outlier Selection







Visualizing Data using t-SNE JMLR 2011

The similarity of datapoint x_j to datapoint x_i is the conditional probability, that $\boldsymbol{x_i}$ **would pick** $\boldsymbol{x_j}$ **as its neighbor** if neighbors were picked in proportion to their probability density under a Gaussian centered at x_i

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i - x_k\|^2 / 2\sigma_i^2\right)} \qquad p_{i|i} = 0 \qquad q_{j|i} = \frac{\exp\left(-\|y_i - y_j\|^2\right)}{\sum_{k \neq i} \exp\left(-\|y_i - y_k\|^2\right)}$$

$$C = \sum_i KL(P_i||Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}$$

$P_g(x) \to 0 \quad P_r \to 1 \qquad P_g(x)\log\frac{P_g(x)}{P_r(x)} \to 0$

$$\frac{\delta C}{\delta y_i} = 2\sum_j (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j})(y_i - y_j)$$

$P_g(x) \to 1 \quad P_r \to 0 \qquad P_g(x)\log\frac{P_g(x)}{P_r(x)} \to +\infty$

the SNE cost function focuses on retaining the local structure of the data in the map

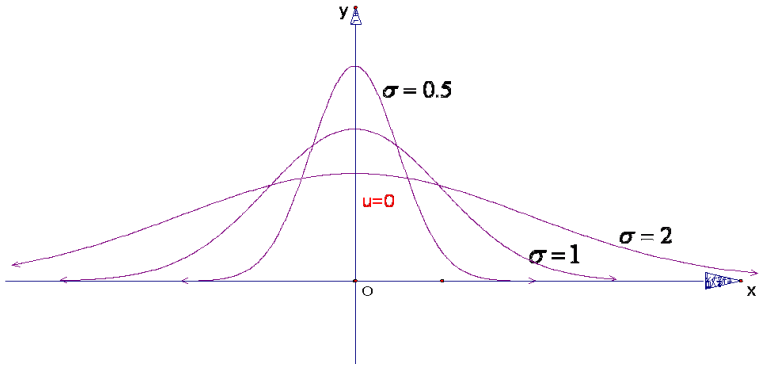
appropriate σ_i

Any particular value of σ_i induces a probability distribution, P_i, over all of the other datapoints. This distribution has an entropy which increases as σ_i increases

perplexity $\qquad Perp(P_i) = 2^{H(P_i)} \qquad H(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i}$

The perplexity can be interpreted as a smooth measure of the effective number of neighbors.

In dense regions, a smaller value of σ_i is usually more appropriate than in sparser regions.



short version

$$a_{ij} = \begin{cases} \exp\left(-d_{ij}^2 \,/\, 2\sigma_i^2\right) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

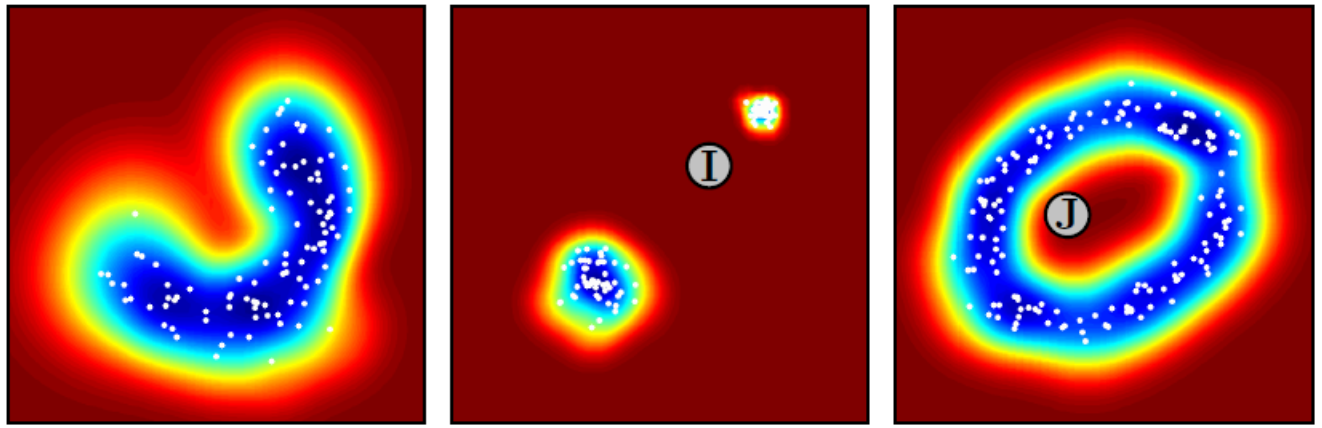
binding matrix **B**

$$b_{ij} = \frac{a_{ij}}{\sum_{k=1}^{n} a_{ik}}$$

$$p(\mathbf{x}_i \in \mathcal{C}_O) = \prod_{j \neq i} (1 - b_{ji})$$

$$f_{SOS}(\mathbf{x}) = \begin{cases} \text{outlier} & \text{if } \varphi_{SOS}(\mathbf{x}) > \theta, \\ \text{inlier} & \text{if } \varphi_{SOS}(\mathbf{x}) \leq \theta. \end{cases}$$

Banana Densities Ring

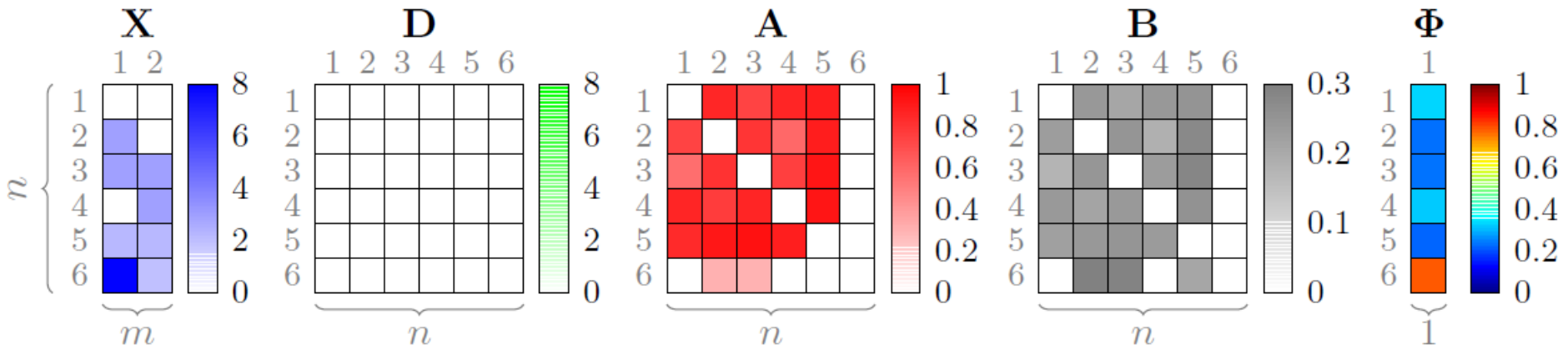


long version

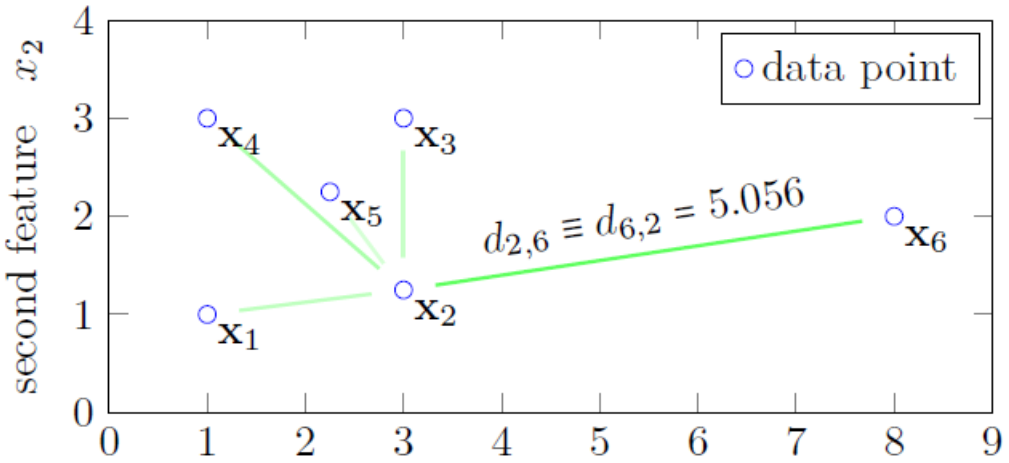
n. 密切关系；吸引力；姻亲关系；类同

SOS employs the concept of **affinity** to quantify the relationship from one data point to another data point. Affinity **is proportional to the similarity between two data points**.

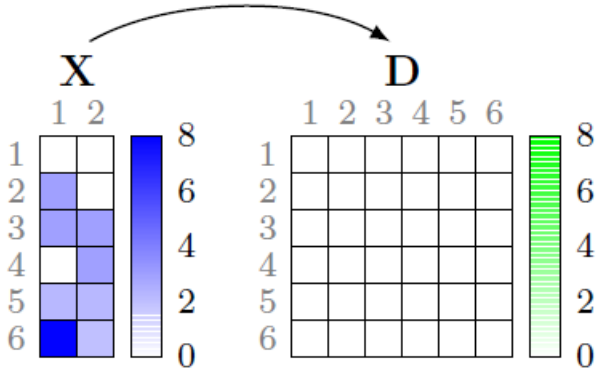
Affinity: the problem of clustering and dimension reduction



The features of the data points are used to measure the dissimilarity between pairs of data points.



Equation 4.1

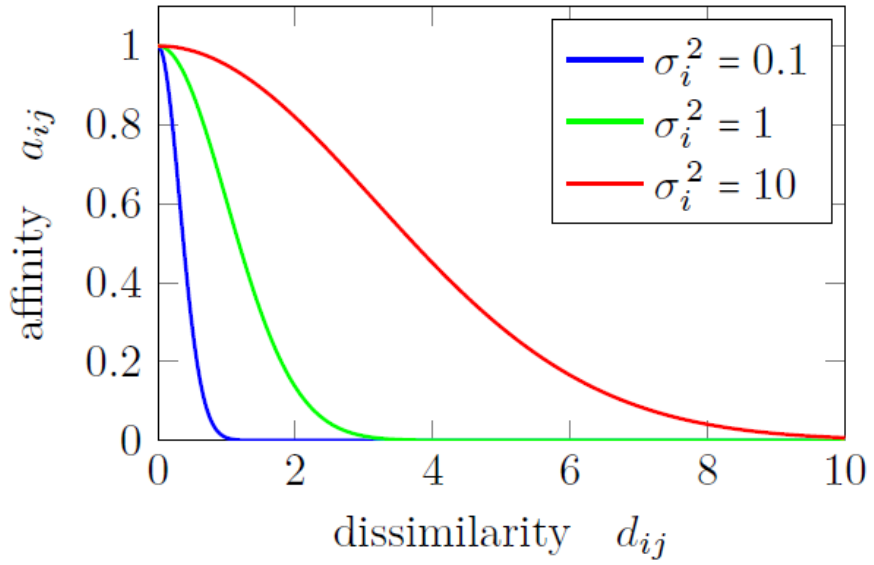


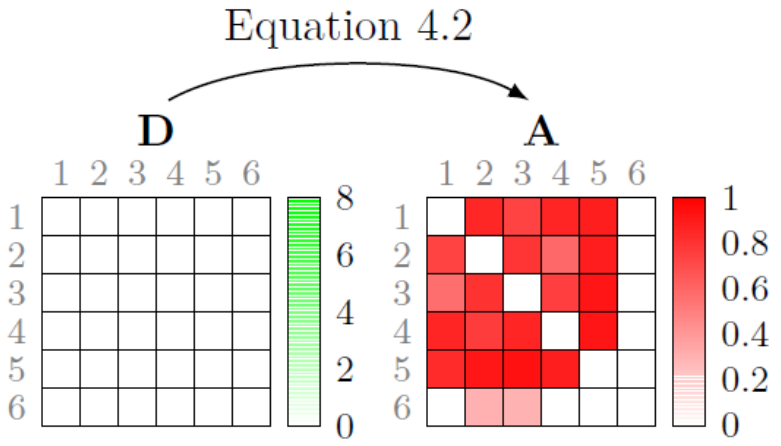
$$d_{ij} = \sqrt{\sum_{k=1}^{m} \left(x_{jk} - x_{ik}\right)^2}$$

Transforming dissimilarity into affinity

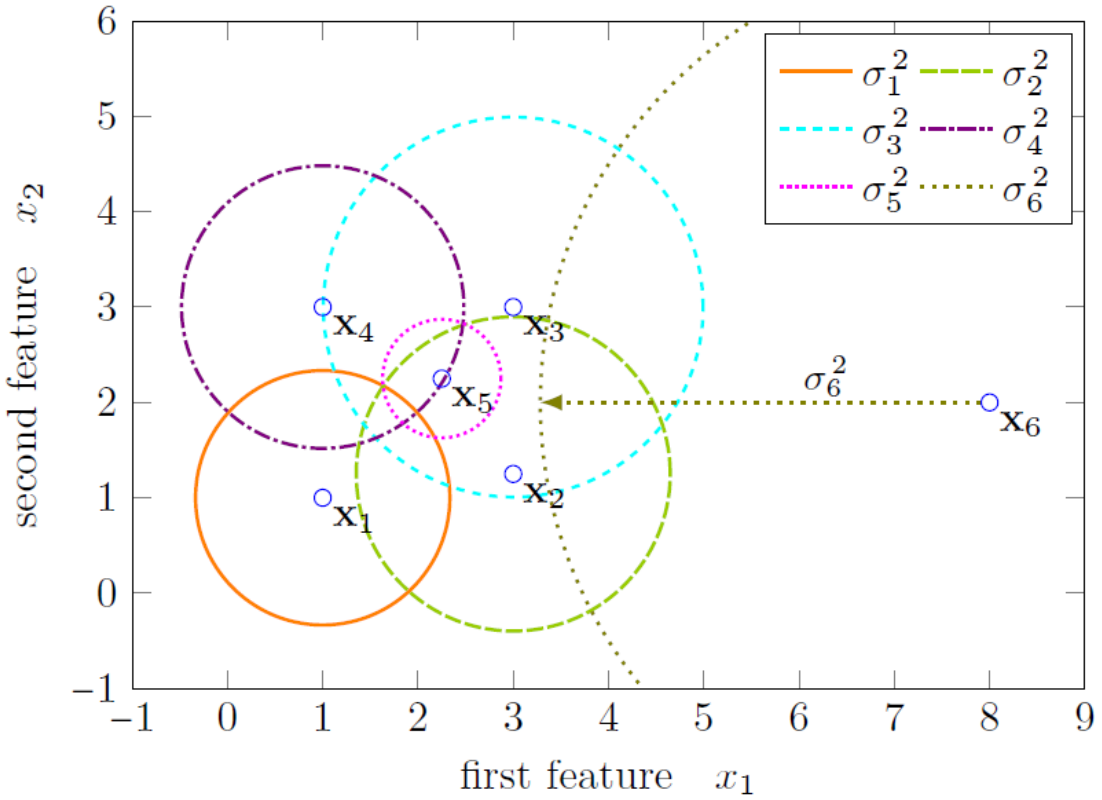
Definition (Affinity**)**

$$a_{ij} = \begin{cases} \exp\left(-d_{ij}^2 \;/\; 2\sigma_i^2\right) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$





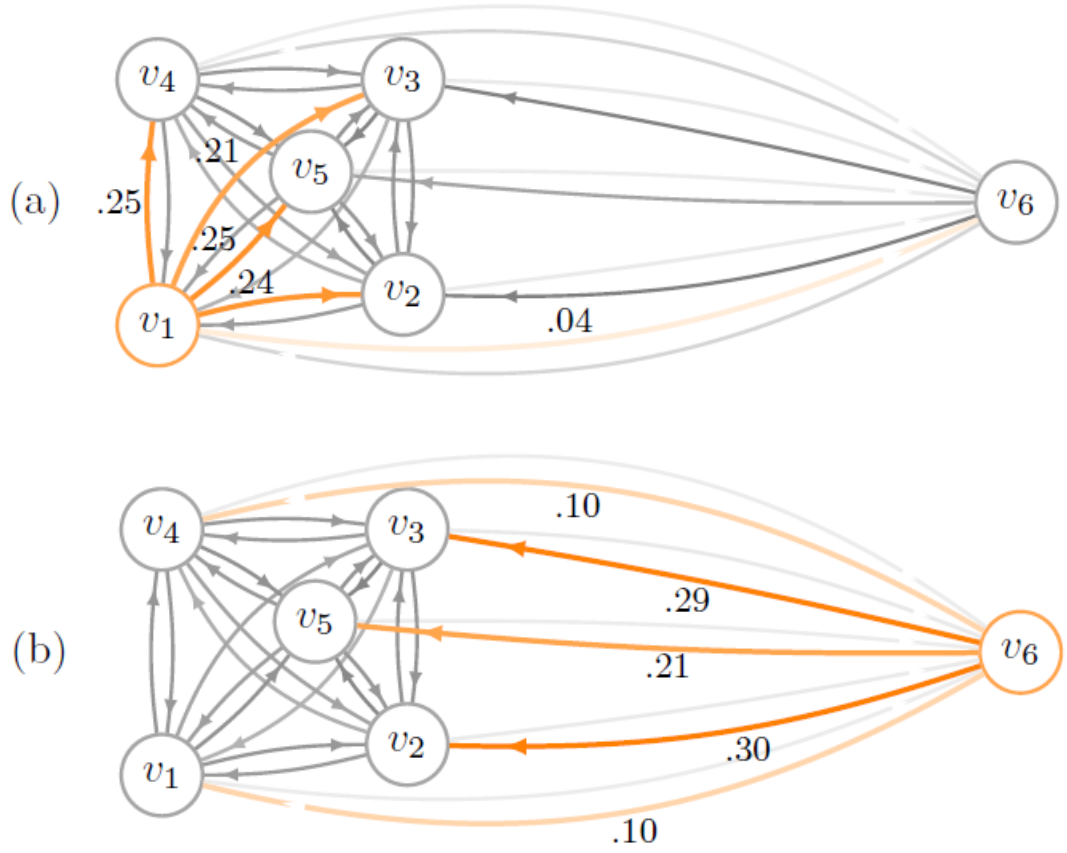
The radii of the circles correspond to the variance for the data points

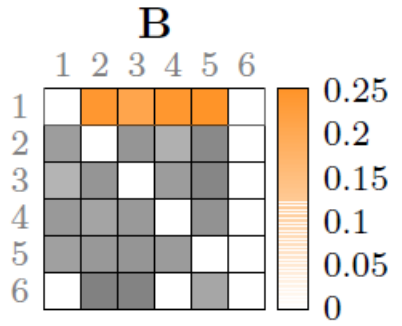


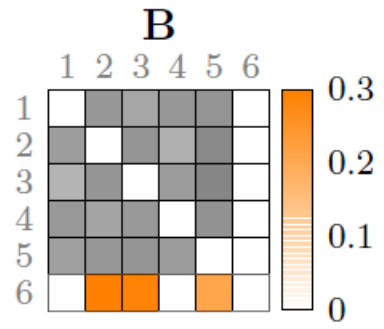
Stochastic **N**eighbour **G**raphs based on affinities

Binding probabilities

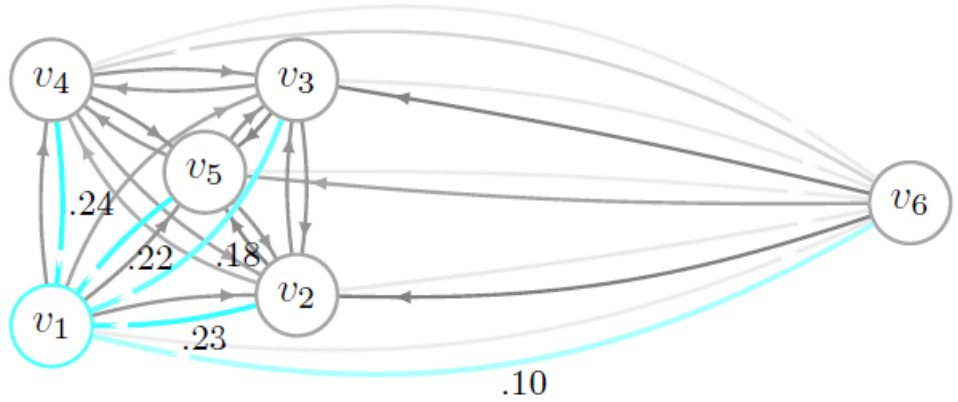
$$b_{ij} \equiv p\left(i \to j \in \mathcal{E}_G\right) \propto a_{ij} \qquad b_{ij} = \frac{a_{ij}}{\sum_{k=1}^{n} a_{ik}}$$

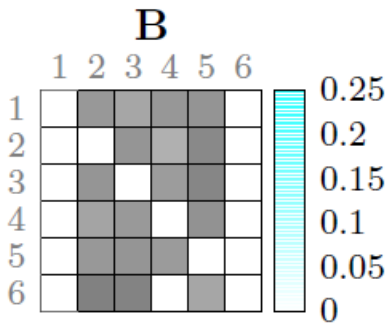




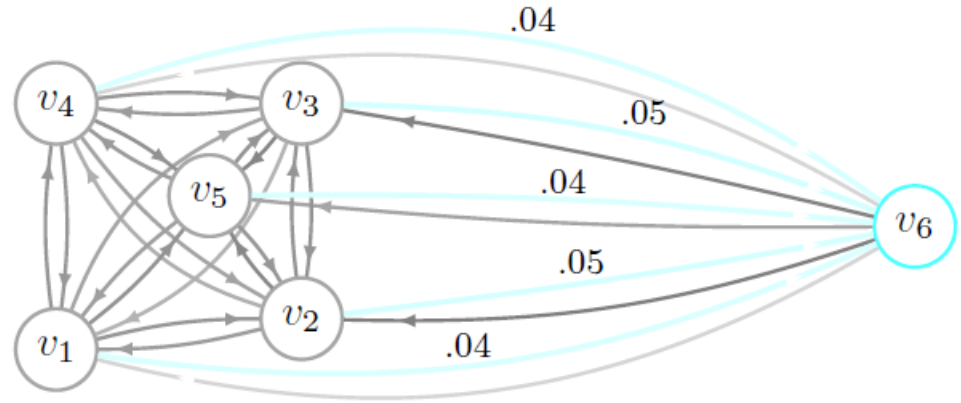


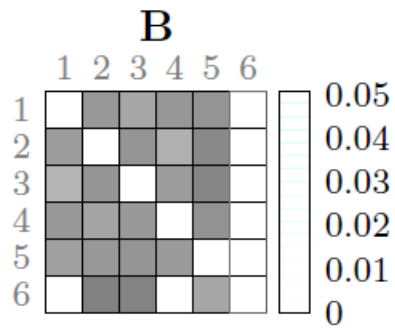
(c)


v_4
v_3
v_5
v_6
v_2
v_1
.24
.22
.18
.23
.10

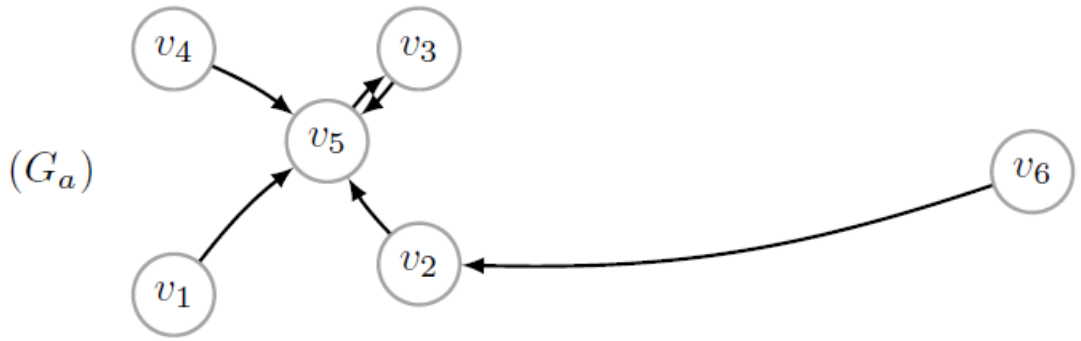

B
1 2 3 4 5 6
1
2
3
4
5
6
0.25
0.2
0.15
0.1
0.05
0

(d)


.04
v_4
v_3
.05
v_5
.04
v_6
.05
v_2
v_1
.04


B
1 2 3 4 5 6
1
2
3
4
5
6
0.05
0.04
0.03
0.02
0.01
0

Being an outlier given one SNG

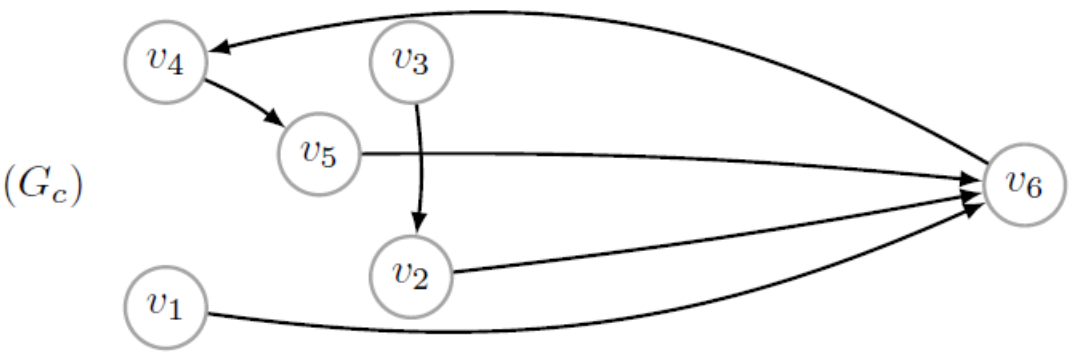


$$p(G_a) = 3.931 \cdot 10^{-4}$$

$$\mathcal{C}_O | G_a = \{\mathbf{x}_1, \mathbf{x}_4, \mathbf{x}_6\}$$

$$\mathcal{C}_{\mathrm{O}} \,|\, G = \{\mathbf{x}_i \in \mathbf{X} \mid \deg_G^-(v_i) = 0\}$$

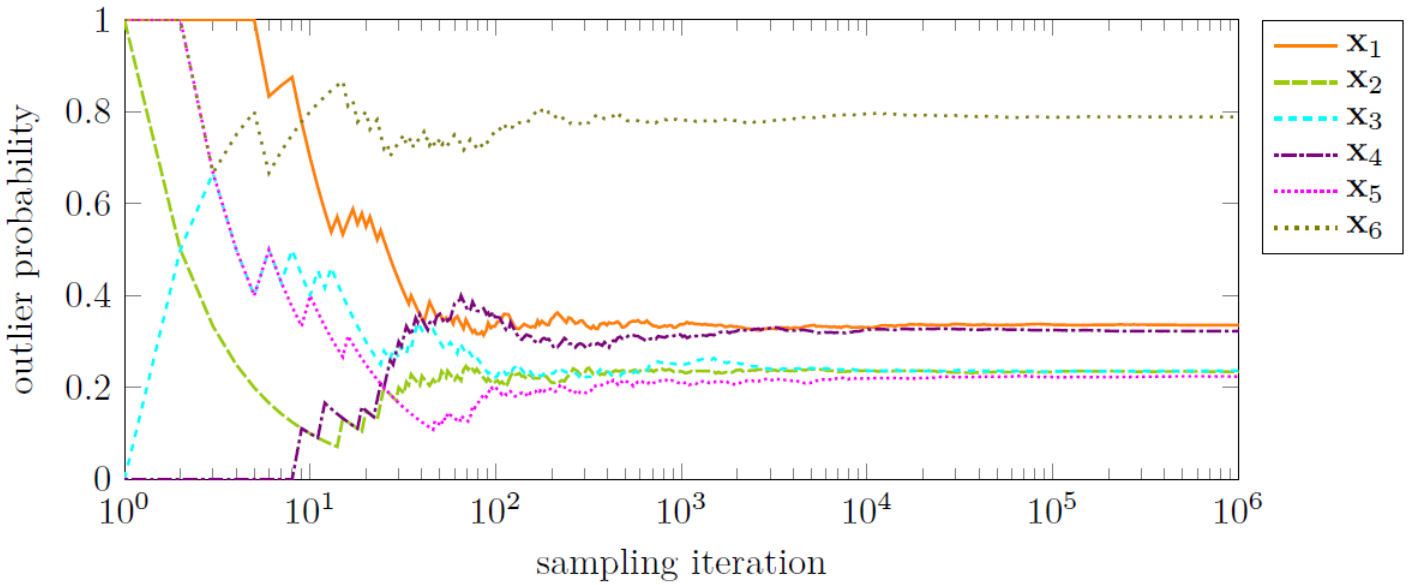
$$\mathcal{C}_{\mathrm{O}} \,|\, G = \{\mathbf{x}_i \in \mathbf{X} \mid \nexists v_j \in \mathcal{V} : j \to i \in \mathcal{E}_G\}$$



$$p(G_c) = 5.950 \cdot 10^{-7}$$

$$\mathcal{C}_O | G_c = \{\mathbf{x}_1, \mathbf{x}_3\}$$

$$p(\mathbf{x}_i \in \mathcal{C}_O) = \lim_{S\to\infty} \frac{1}{S} \sum_{s=1}^{S} \mathbb{1}\left\{\mathbf{x}_i \in \mathcal{C}_O \,|\, G^{(s)}\right\} \quad , \quad G^{(s)} \sim P(\mathcal{G}) \qquad p(G) = \prod_{i\to j \in \mathcal{E}_G} b_{ij}$$



$$p(\mathbf{x}_i \in \mathcal{C}_O) = \sum_{G\in\mathcal{G}} \mathbb{1}\{\mathbf{x}_i \in \mathcal{C}_O \,|\, G\} \cdot p(G)$$

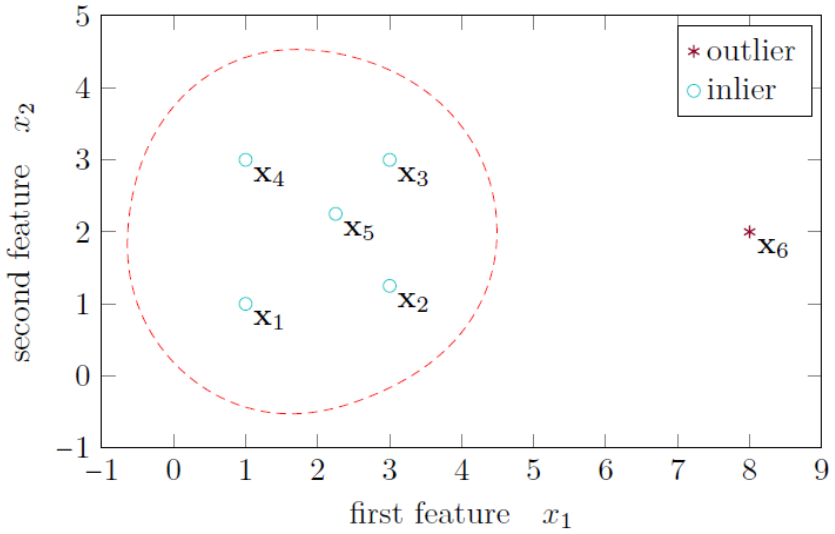
$$= \sum_{G\in\mathcal{G}} \mathbb{1}\{\mathbf{x}_i \in \mathcal{C}_O \,|\, G\} \cdot \prod_{q\to r \in \mathcal{E}_G} b_{qr}$$

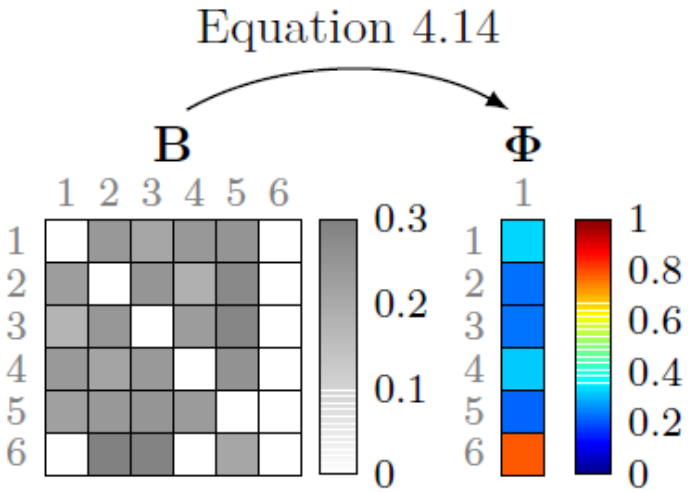
$5^6 = 15{,}625$ SNGs

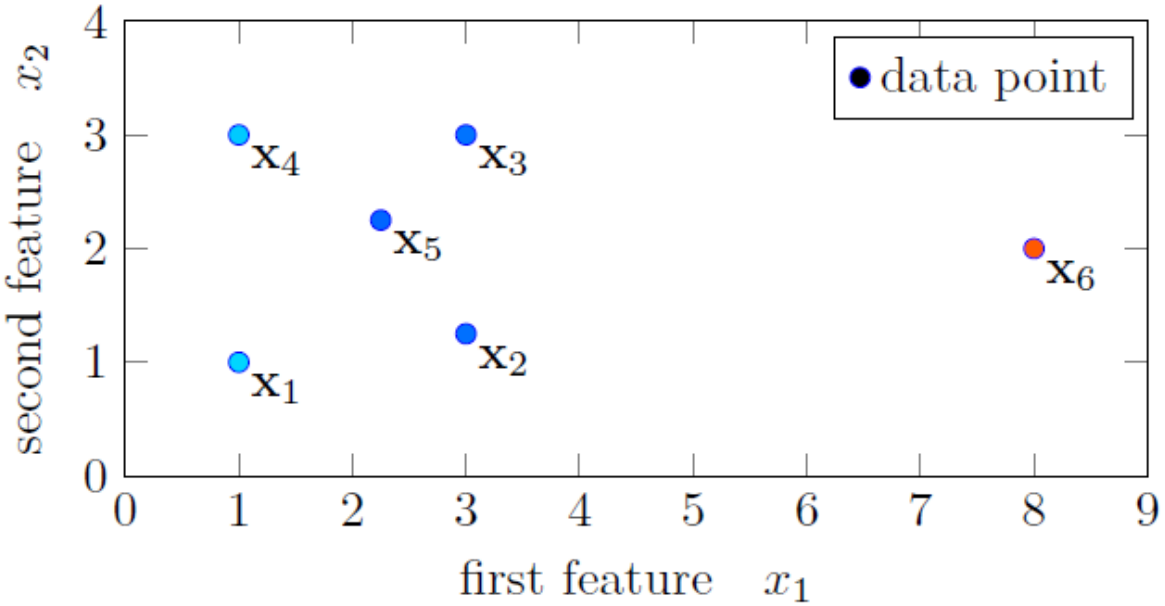
0:335; 0:235; 0:237; 0:323; 0:224, 0:788

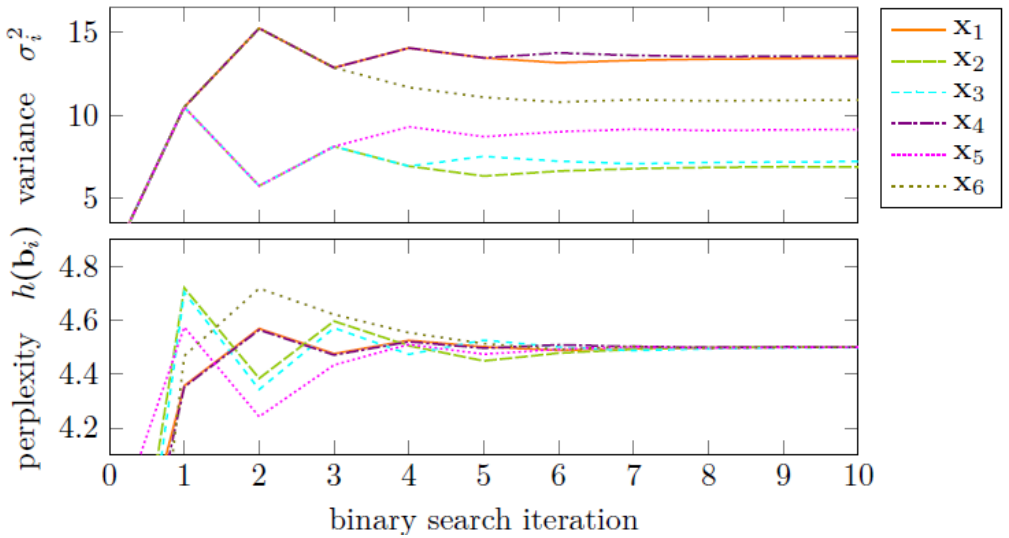
$$p(\mathbf{x}_i \in \mathcal{C}_O) = \prod_{j \neq i} (1 - b_{ji}) \qquad \varphi_{SOS}(\mathbf{x}_i) \equiv p(\mathbf{x}_i \in \mathcal{C}_O)$$

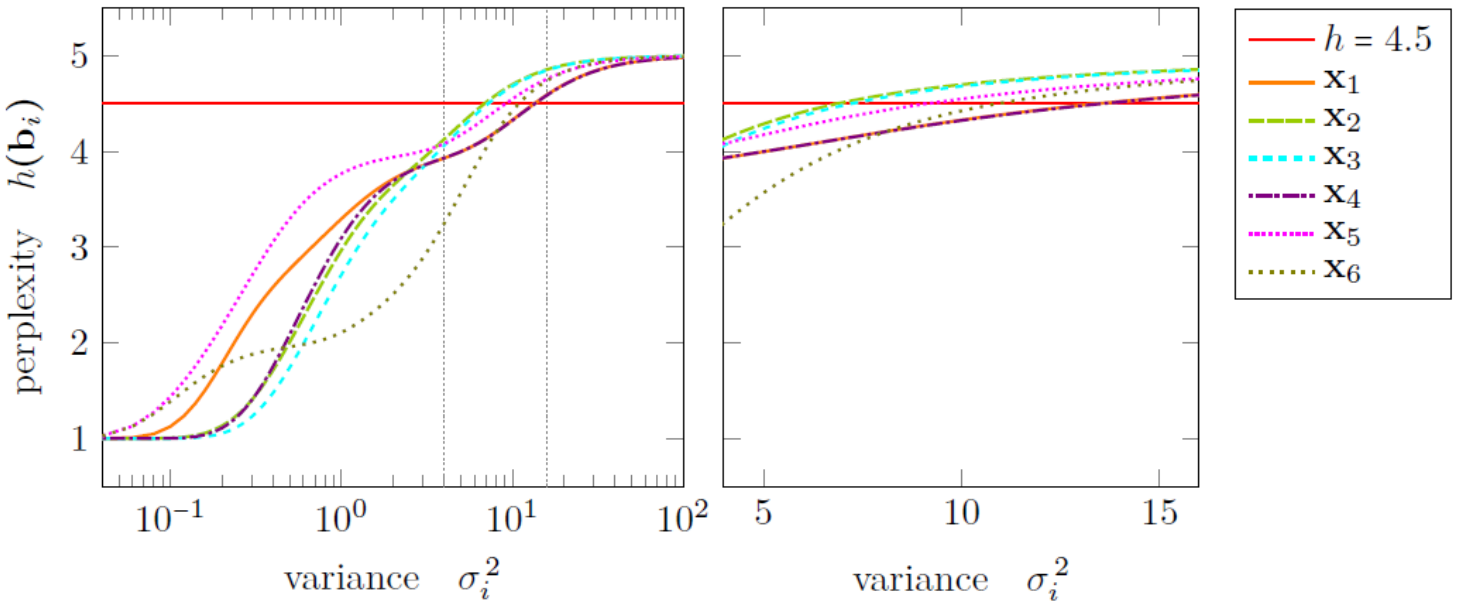
$$f_{SOS}(\mathbf{x}) = \begin{cases} \text{outlier} & \text{if } \varphi_{SOS}(\mathbf{x}) > \theta \\ \text{inlier} & \text{if } \varphi_{SOS}(\mathbf{x}) \leq \theta \end{cases}$$








variance σ_i^2
15
10
5
perplexity $h(\mathbf{b}_i)$
4.8
4.6
4.4
4.2
0
1
2
3
4
5
6
7
8
9
10
binary search iteration
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$


perplexity $h(\mathbf{b}_i)$
5
4
3
2
1
10^{-1}
10^{0}
10^{1}
10^{2}
variance σ_i^2
5
10
15
variance σ_i^2
$h = 4.5$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$

$$p(\mathbf{x}_i \in \mathcal{C}_O) = \prod_{j \neq i} (1 - b_{ji}) \qquad 0.9^{1000} = 1.748 \times 10^{-46}$$

classification

outlier detection is an unsupervised problem

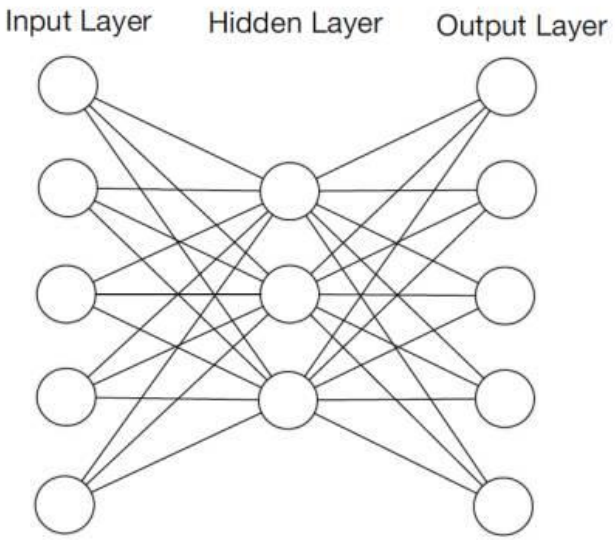
not only outliers

Matrix Sketch

better expression of the outlier probability?

Sparse Representation

RNN



PCA