Anomaly Detection

Xiao-Zhi Gao

xiao-zhi.gao@uef.fi

Outline

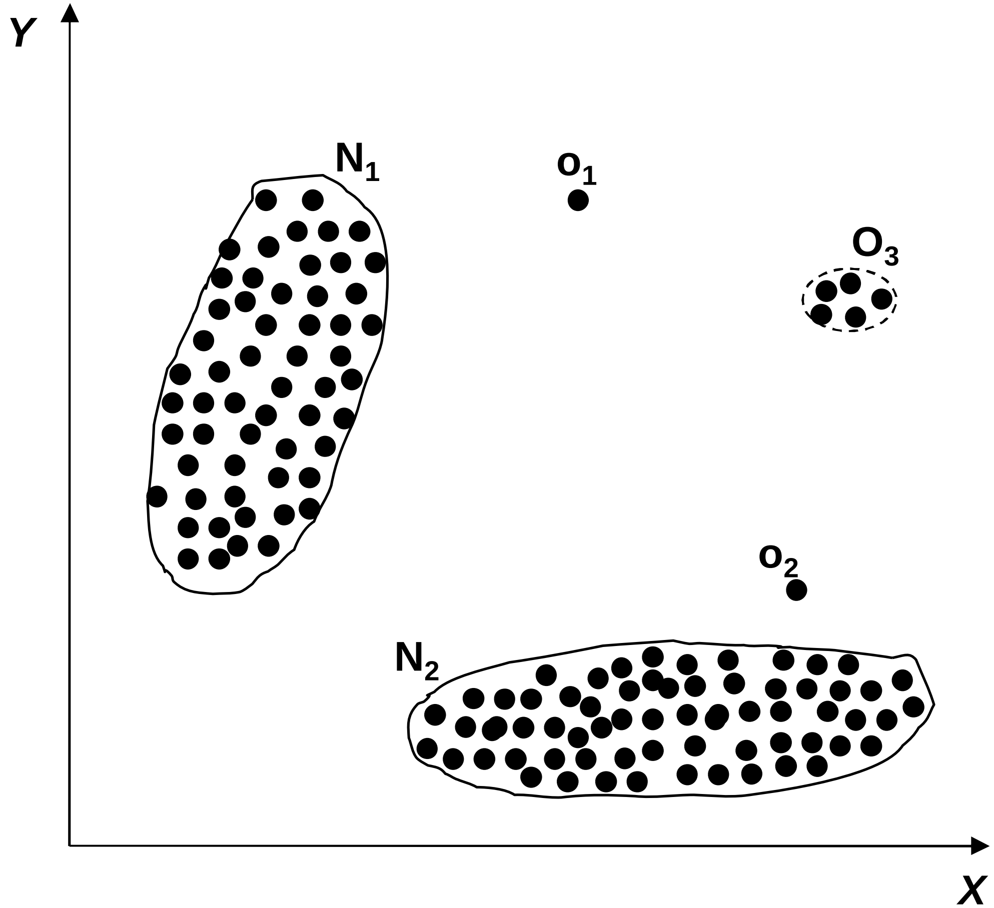
- Introduction
- Anomaly detection in data
- Negative Selection Algorithm (NSA)
- Application examples
- Conclusions

What are Anomalies?

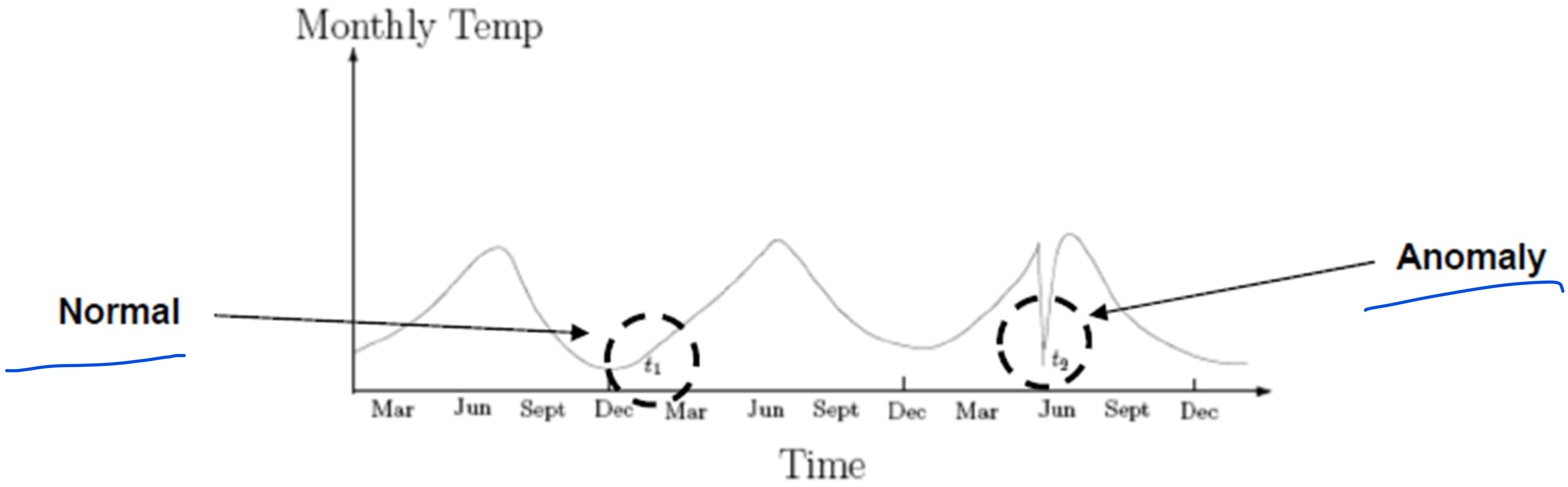
- Anomaly is a pattern in the data that does not conform to the expected behaviors.
- Anomaly is also referred to as outliers, exceptions, peculiarities, surprise, etc.
- Anomalies translate to significant (often critical) real life entities.
  - Credit card fraud
  - Cyber intrusions

Simple Example

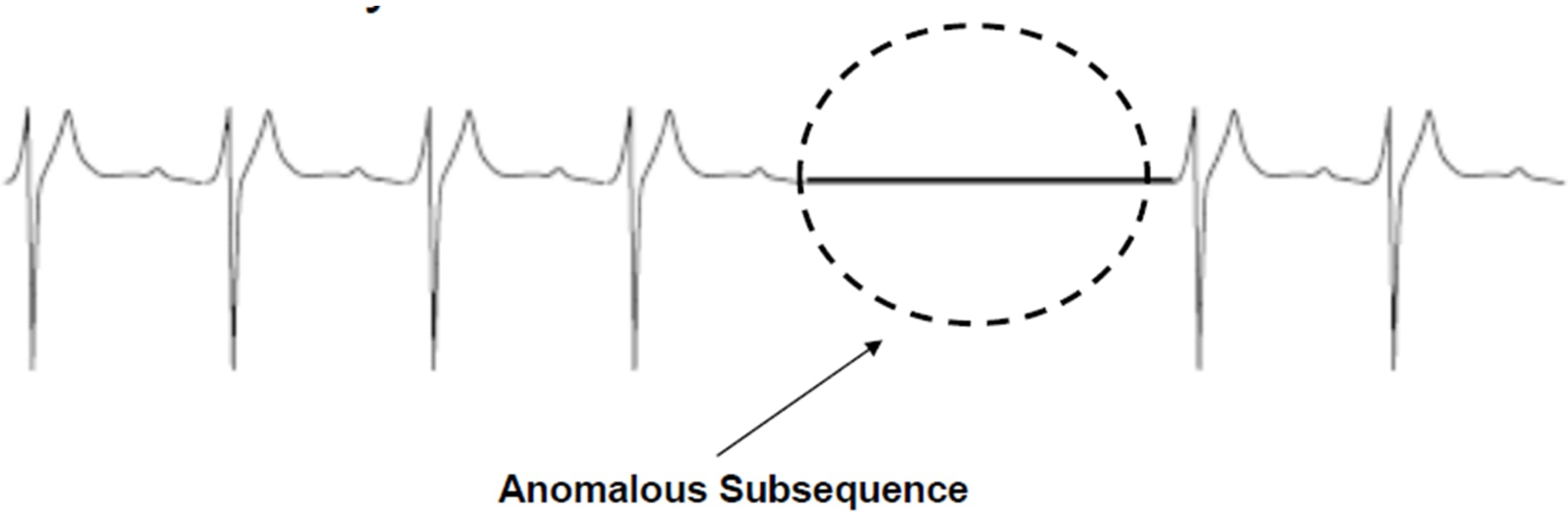
- N_1 and N_2 are regions of normal behavior.
- Points o_1 and o_2 are anomalies.
- Points in region O_3 are anomalies.



Simple Example



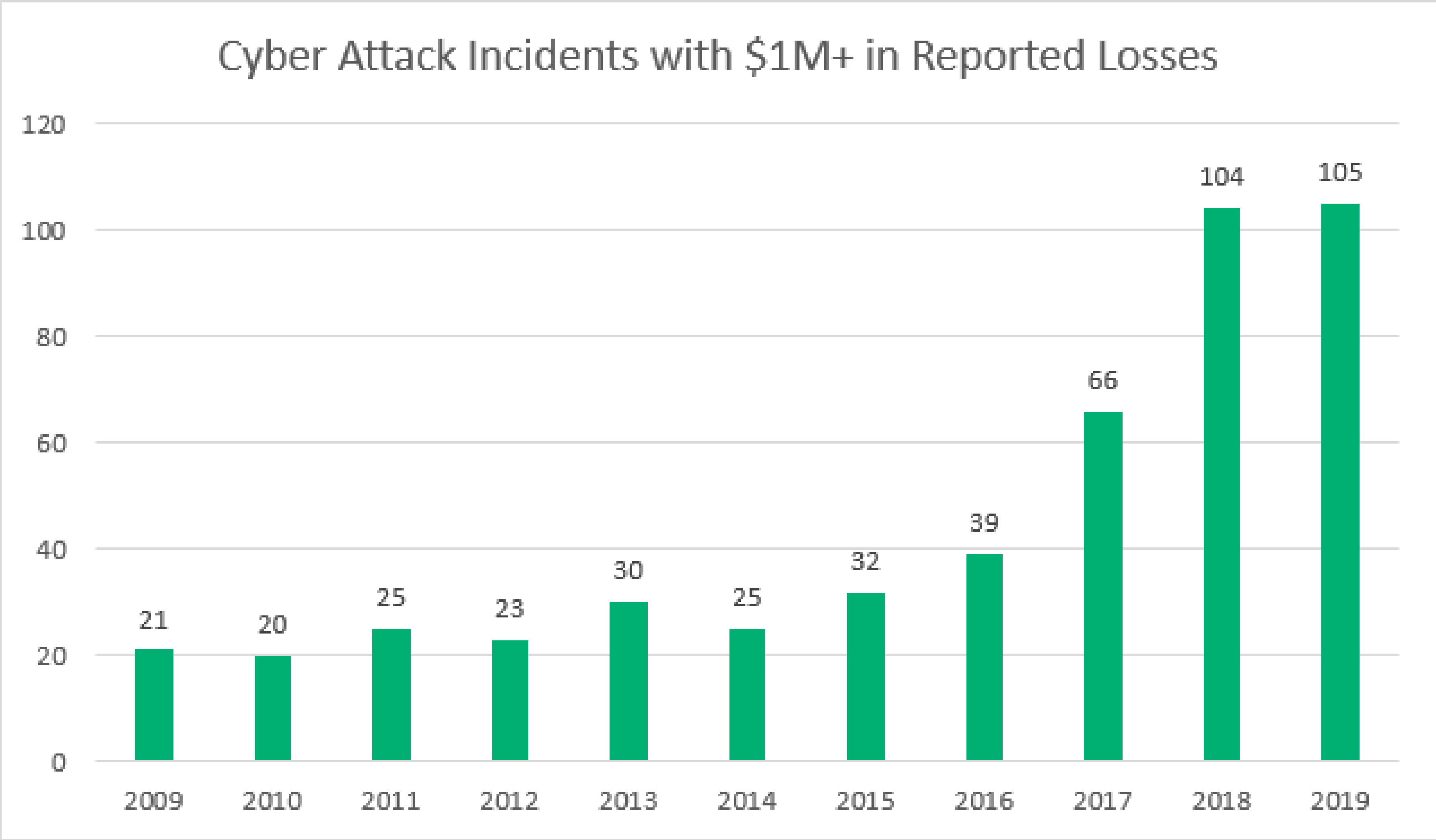
Simple Example



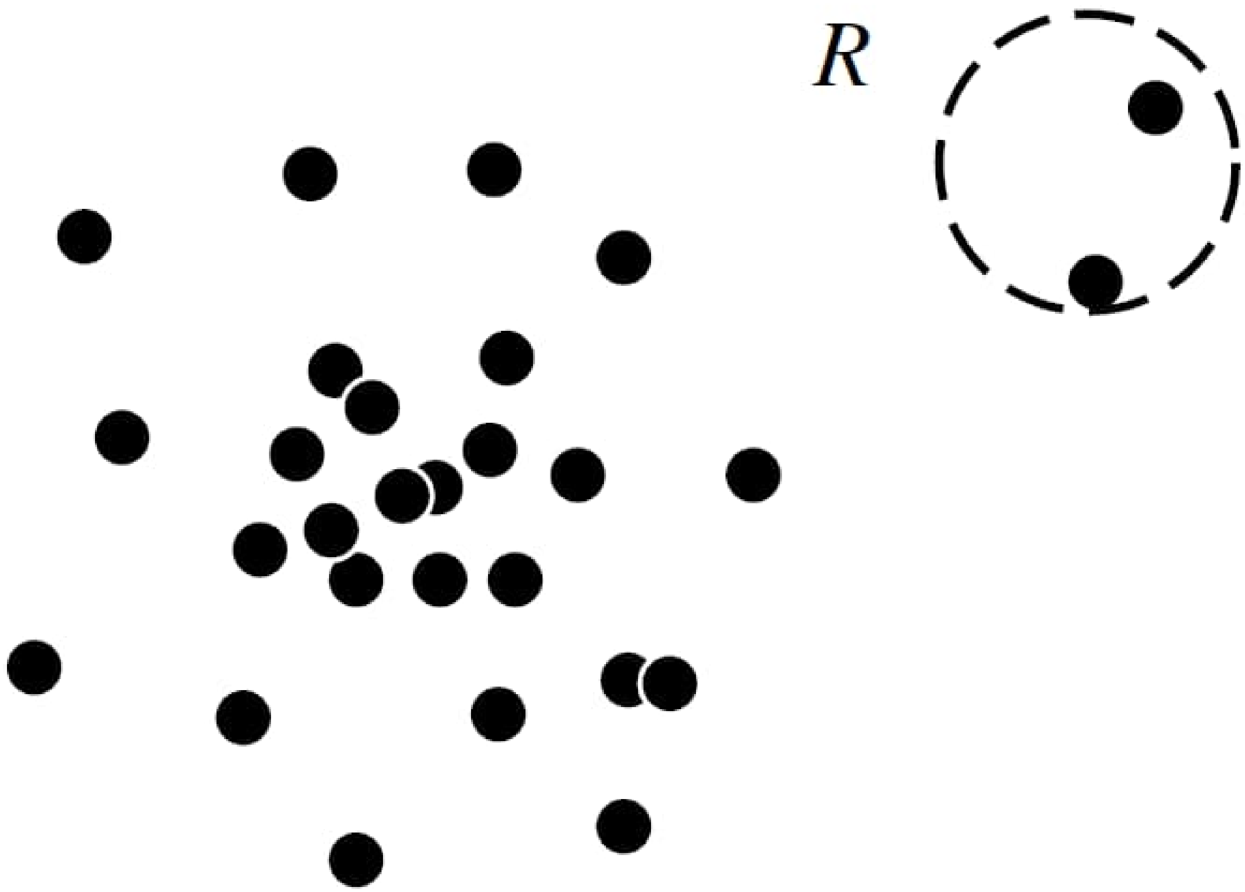
Real-world Anomalies

- Credit Card Fraud
  - An abnormally high purchase made on a credit card
- Cyber Intrusions
  - A web server involved in *ftp* traffic


Card Fraud Worldwide
CARD FRAUD IN BILLIONS
CARD FRAUD IN CENTS PER $100 OF TOTAL VOLUME
5.5¢
7.0¢
7.0¢
6.8¢
7.1¢
6.5¢
6.3¢
6.1¢
$13.70
$21.84
$23.97
$28.65
$32.04
$32.96
$35.31
$38.50
2013
2015
2017
2019
2021
2023
2025
2027
© 2020 Nilson Report

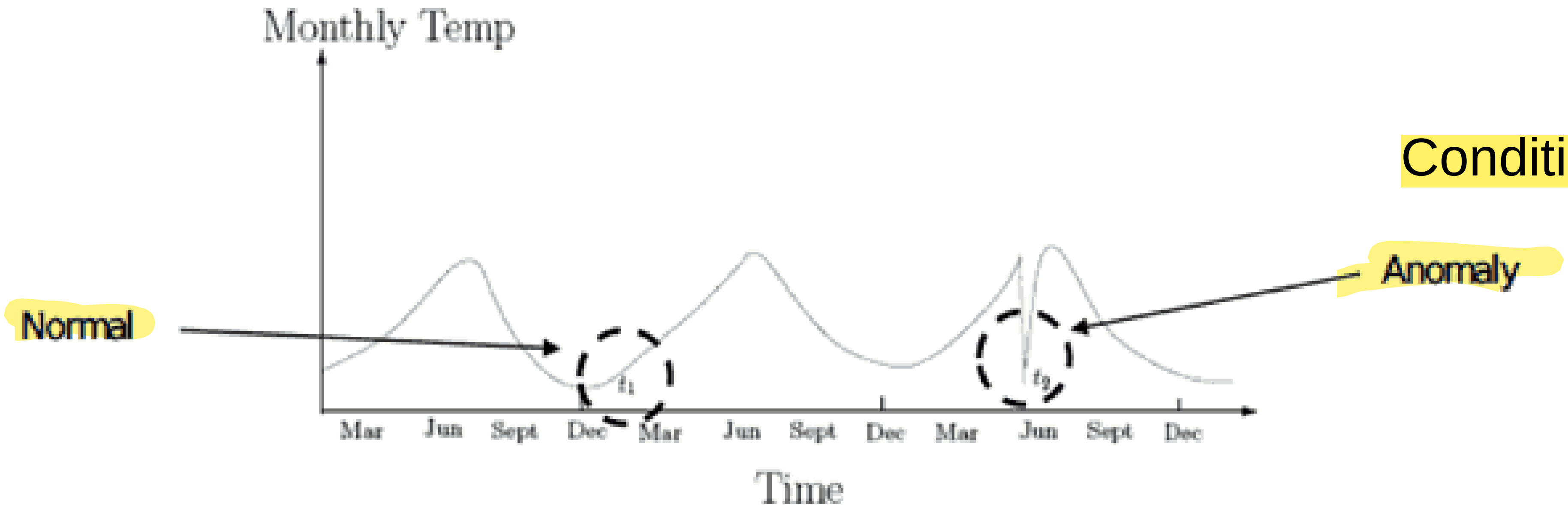

Cyber Attack Incidents with $1M+ in Reported Losses
120
100
80
60
40
20
0
21
20
25
23
30
25
32
39
66
104
105
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

Types of Outliers



Global outliers

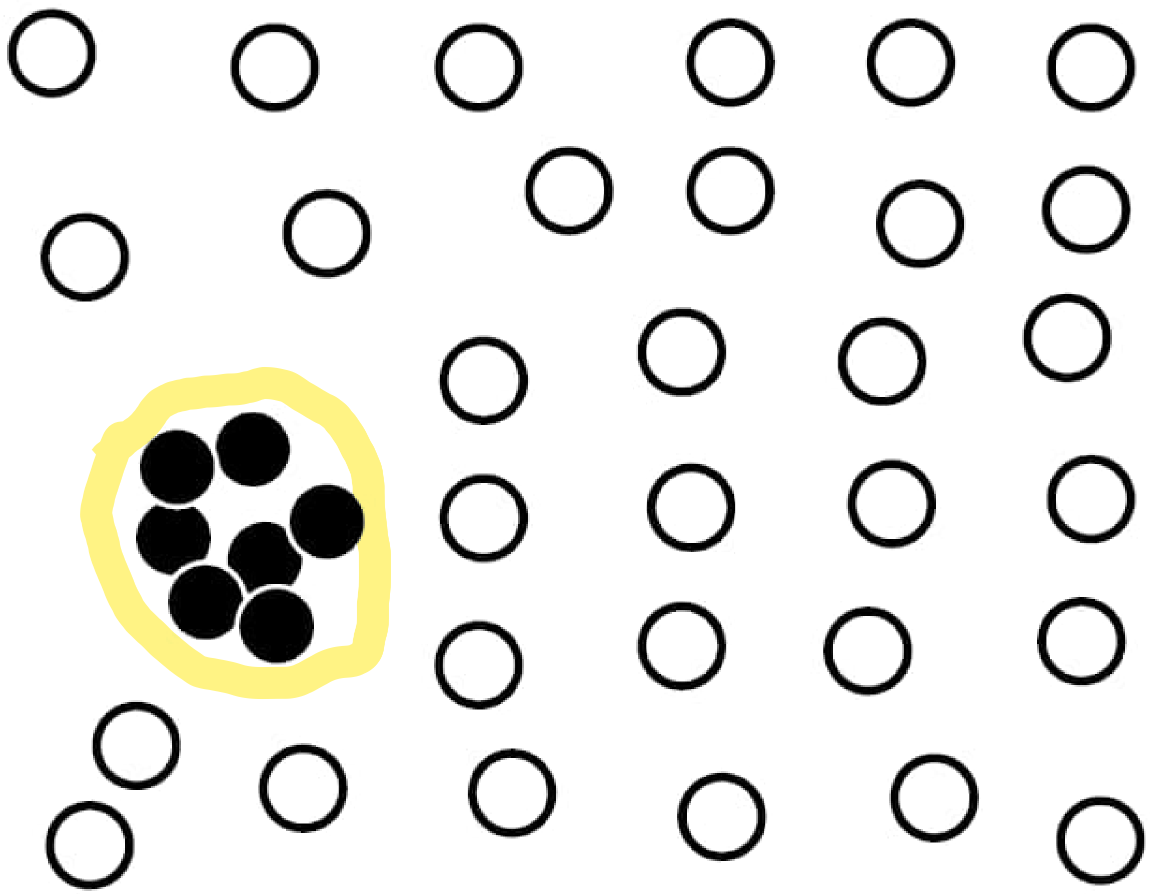
Types of Outliers



Conditional outliers

Types of Outliers

exam



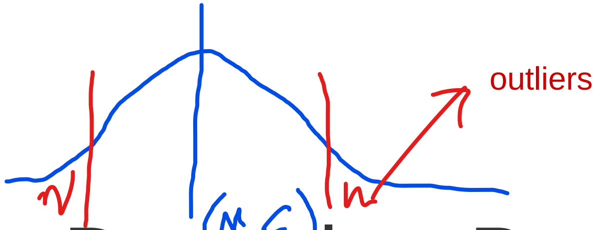
Collective outliers

Applications of Anomaly Detection

- Network intrusion detection
- Insurance/Credit card fraud detection
- Healthcare informatics/Medical diagnostics
- Industrial damage detection
- Image processing/video surveillance
- Novel topic detection in text mining
- ...

Key Challenges

- Defining a representative normal region is challenging.
- The boundary between normal and outlying behavior is often not precise.
- The exact notion of an outlier is different for different application domains.
- Data might contain noise.
- Normal behavior keeps evolving. time-varying

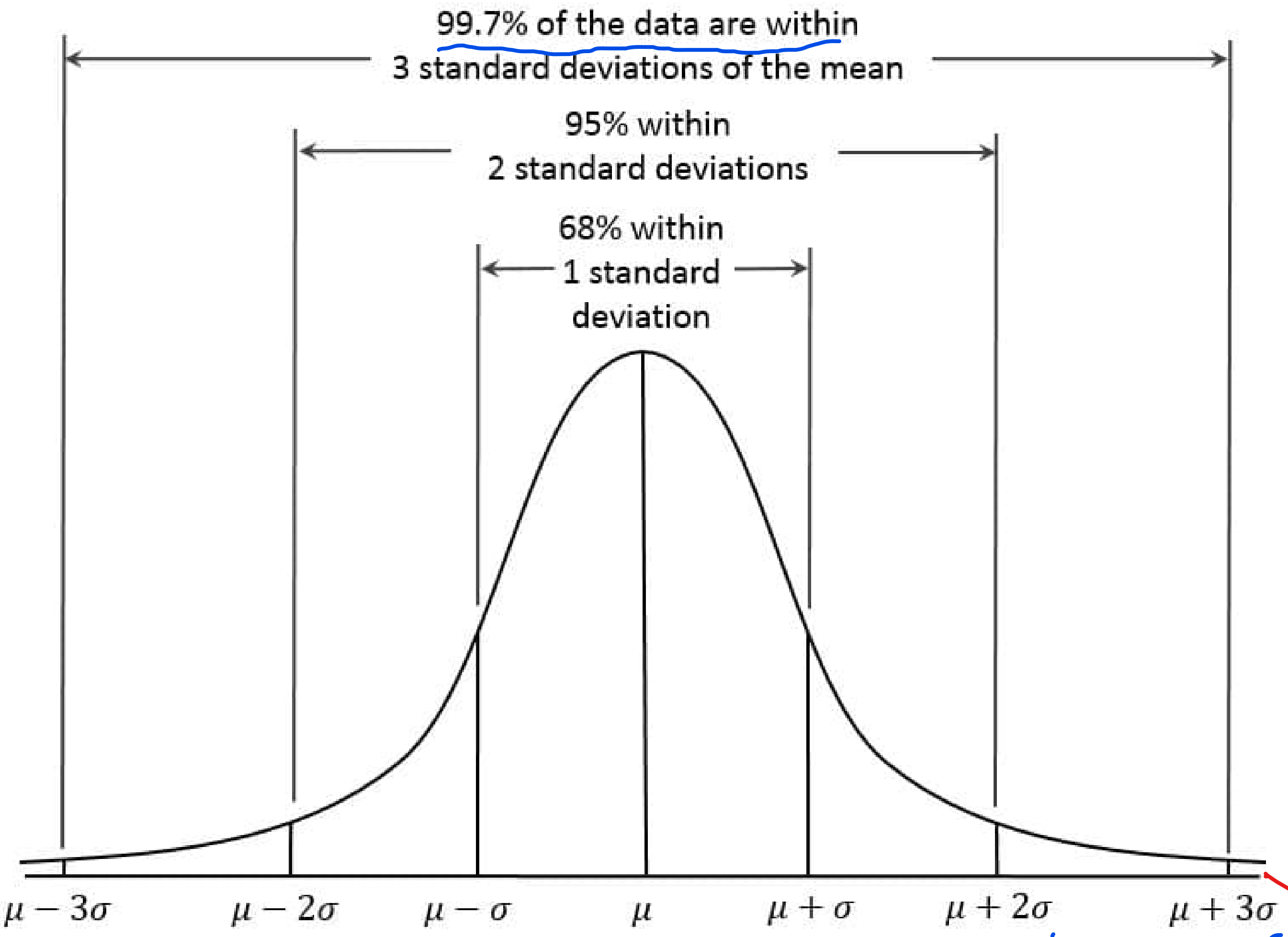


Outlier Detection Based on Normal Distribution

- Usually, we can assure that the data samples are generated from a normal distribution.
- The maximum likelihood method is used to estimate the parameters of the normal distribution based on the data samples.
- Outliers are always away from the mean of the normal distribution function.
- An example of outlier detection in city's average temperature in July.

Outlier Detection Based on Normal Distribution

0.3 left



0.15%

0.15%

Univariate outlier detection using maximum likelihood. Suppose a city's average temperature values in July in the last 10 years are, in value-ascending order, 24.0°C, 28.9°C, 28.9°C, 29.0°C, 29.1°C, 29.1°C, 29.2°C, 29.2°C, 29.3°C, and 29.4°C. Let's assume that the average temperature follows a normal distribution, which is determined by two parameters: the mean, μ, and the standard deviation, σ.

estimation

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

data samples

number of example

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

$$\hat{\mu} = \frac{24.0 + 28.9 + 28.9 + 29.0 + 29.1 + 29.1 + 29.2 + 29.2 + 29.3 + 29.4}{10} = 28.61$$

$$\hat{\sigma}^2 = ((24.1 - 28.61)^2 + (28.9 - 28.61)^2 + (28.9 - 28.61)^2 + (29.0 - 28.61)^2$$
$$+ (29.1 - 28.61)^2 + (29.1 - 28.61)^2 + (29.2 - 28.61)^2 + (29.2 - 28.61)^2$$
$$+ (29.3 - 28.61)^2 + (29.4 - 28.61)^2)/10 \backsimeq 2.29.$$

$$\hat{\sigma} = \sqrt{2.29} = 1.51$$

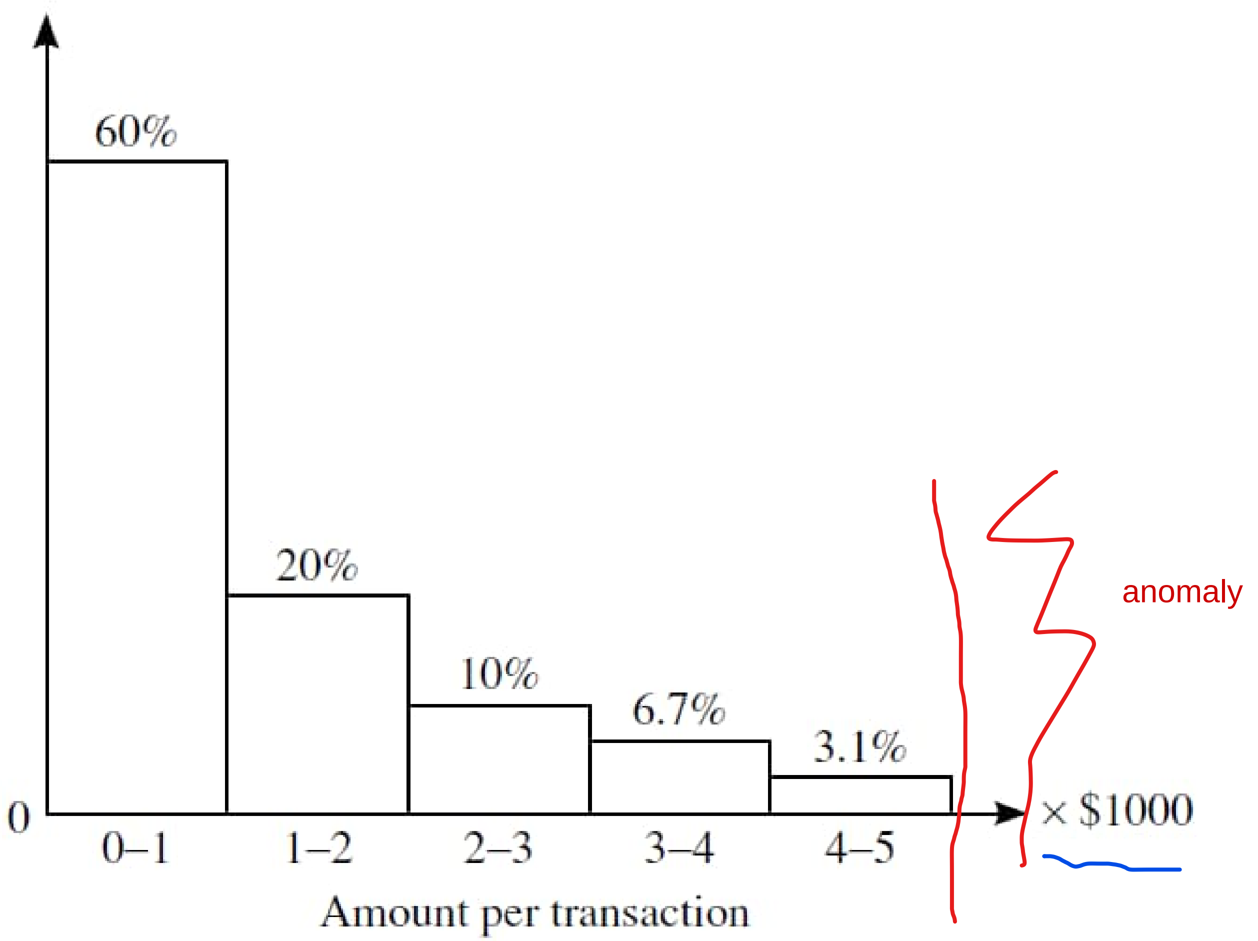
The most deviating value, 24.0°C, is 4.61°C away from the estimated mean. We know that the $\mu \pm 3\sigma$ region contains 99.7% data under the assumption of normal distribution. Because $\frac{4.61}{1.51} = 3.04 > 3$, the probability that the value 24.0°C is generated by the normal distribution is less than 0.15%, and thus can be identified as an outlier. ■

Outlier Detection Based on Histogram

Outlier detection using a histogram. *AllElectronics* records the purchase amount for every customer transaction. Figure 12.5 uses a histogram (refer to Chapters 2 and 3) to graph these amounts as percentages, given all transactions. For example, 60% of the transaction amounts are between \$0.00 and \$1000.

We can use the histogram as a nonparametric statistical model to capture outliers. For example, a transaction in the amount of \$7500 can be regarded as an outlier because only $1-(60\%+20\%+10\%+6.7\%+3.1\%)=0.2\%$ of transactions have an amount higher than \$5000. On the other hand, a transaction amount of \$385 can be treated as normal because it falls into the bin (or bucket) holding 60% of the transactions.

Outlier Detection Based on Histogram

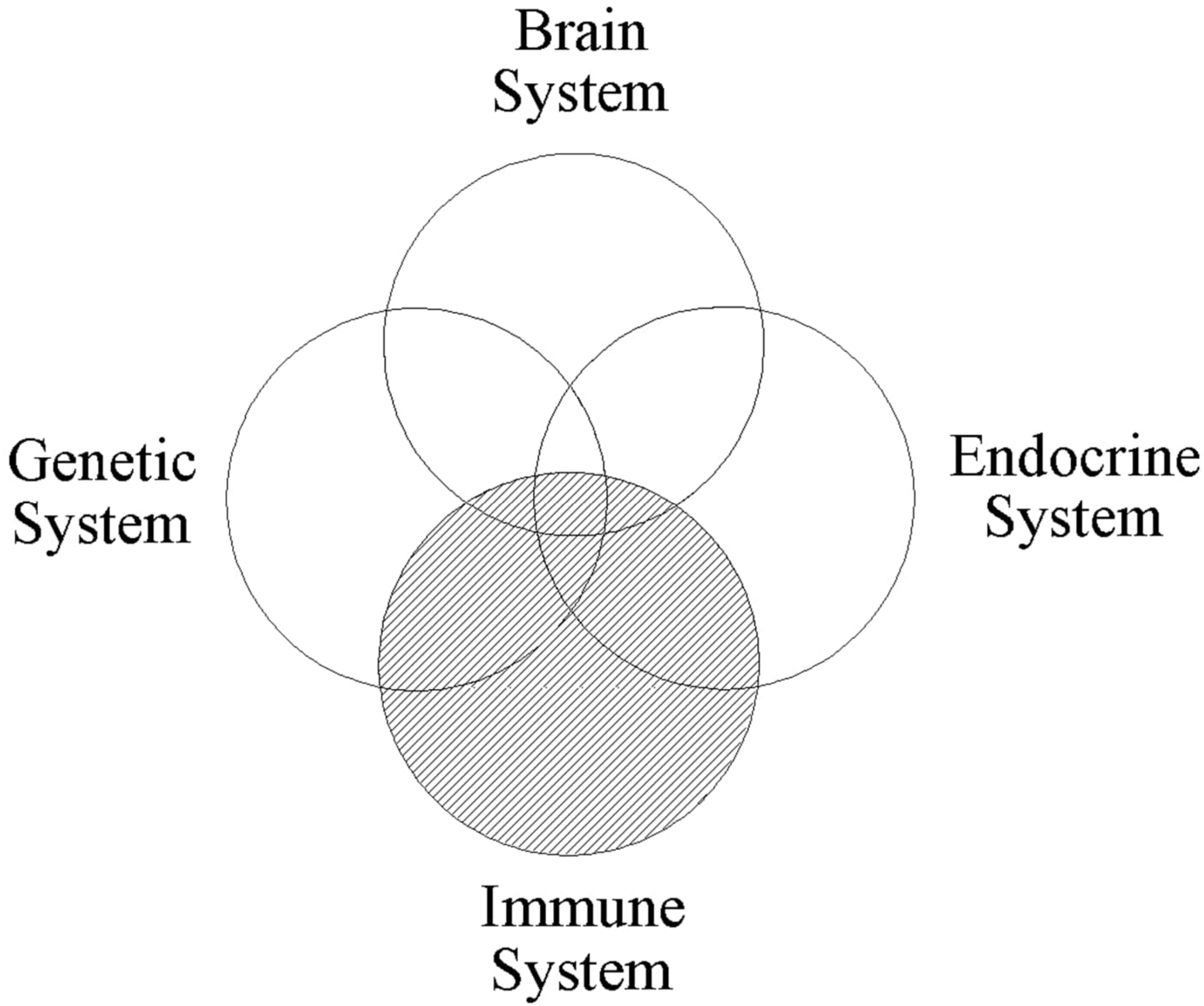


Histogram of purchase amounts in transactions.

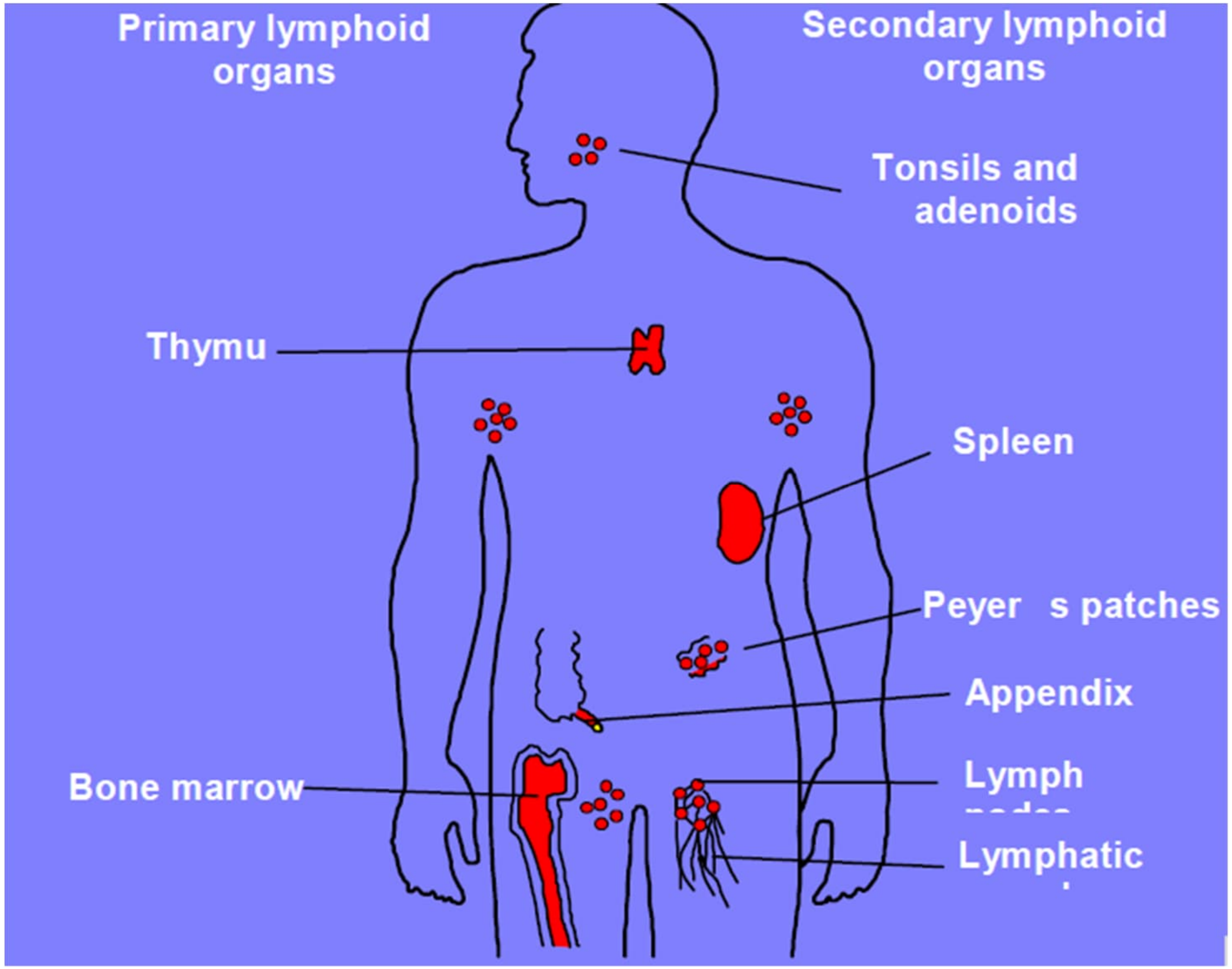
Artificial Immune Systems (AIS)

- Artificial Immune Systems (AIS) are an emerging kind of soft computing methods.
  - Inspired by natural immune systems
  - Features of pattern recognition, optimization, data analysis, machine learning, etc
- Negative Selection Algorithm (NSA) is an important partner of AIS.
  - Maturation of T cells and self/nonself discrimination
  - Developed by Forrest in 1994
- NSA has been widely applied to deal with anomaly detection in data.

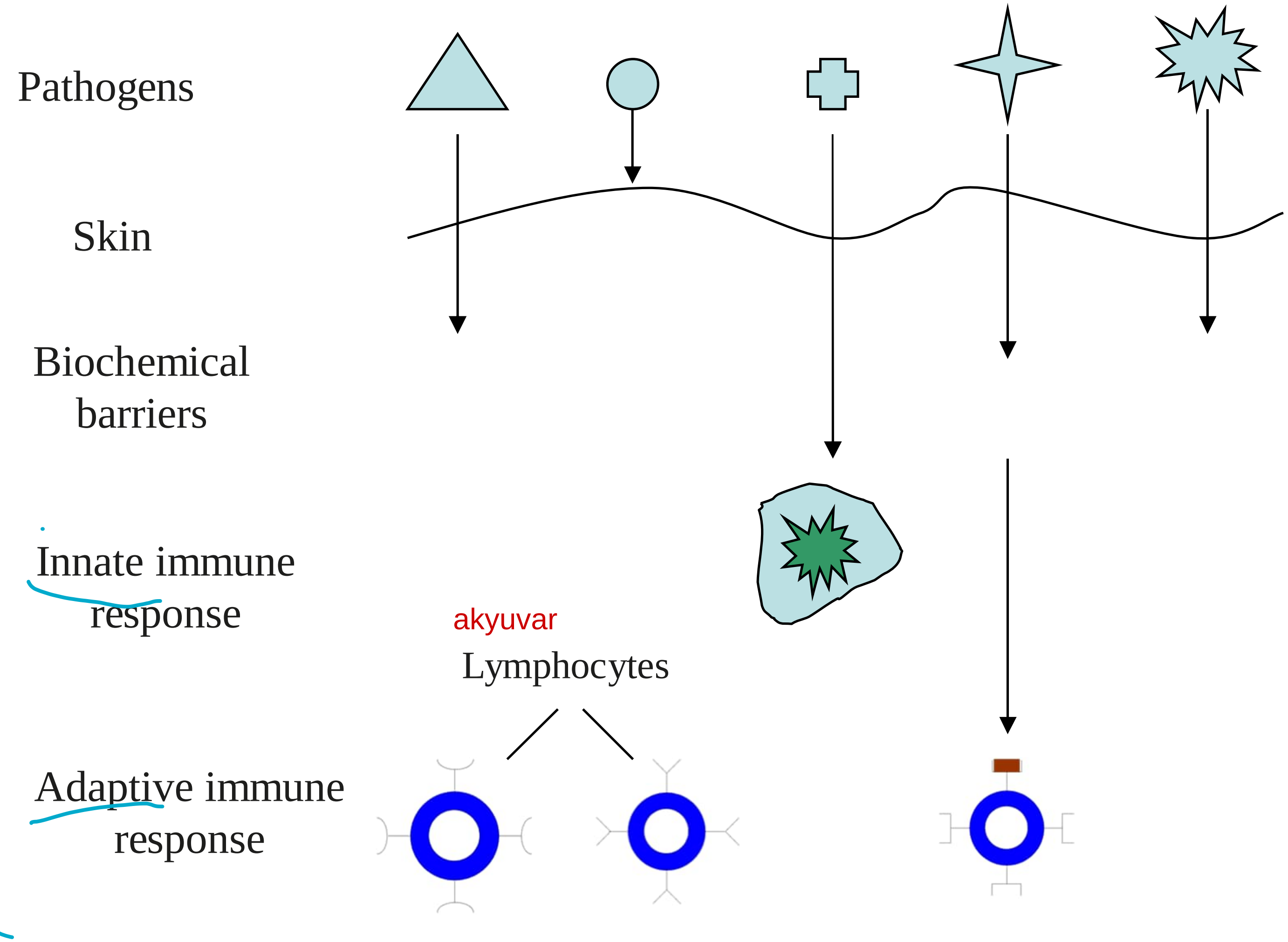
Biological Information Processing Systems



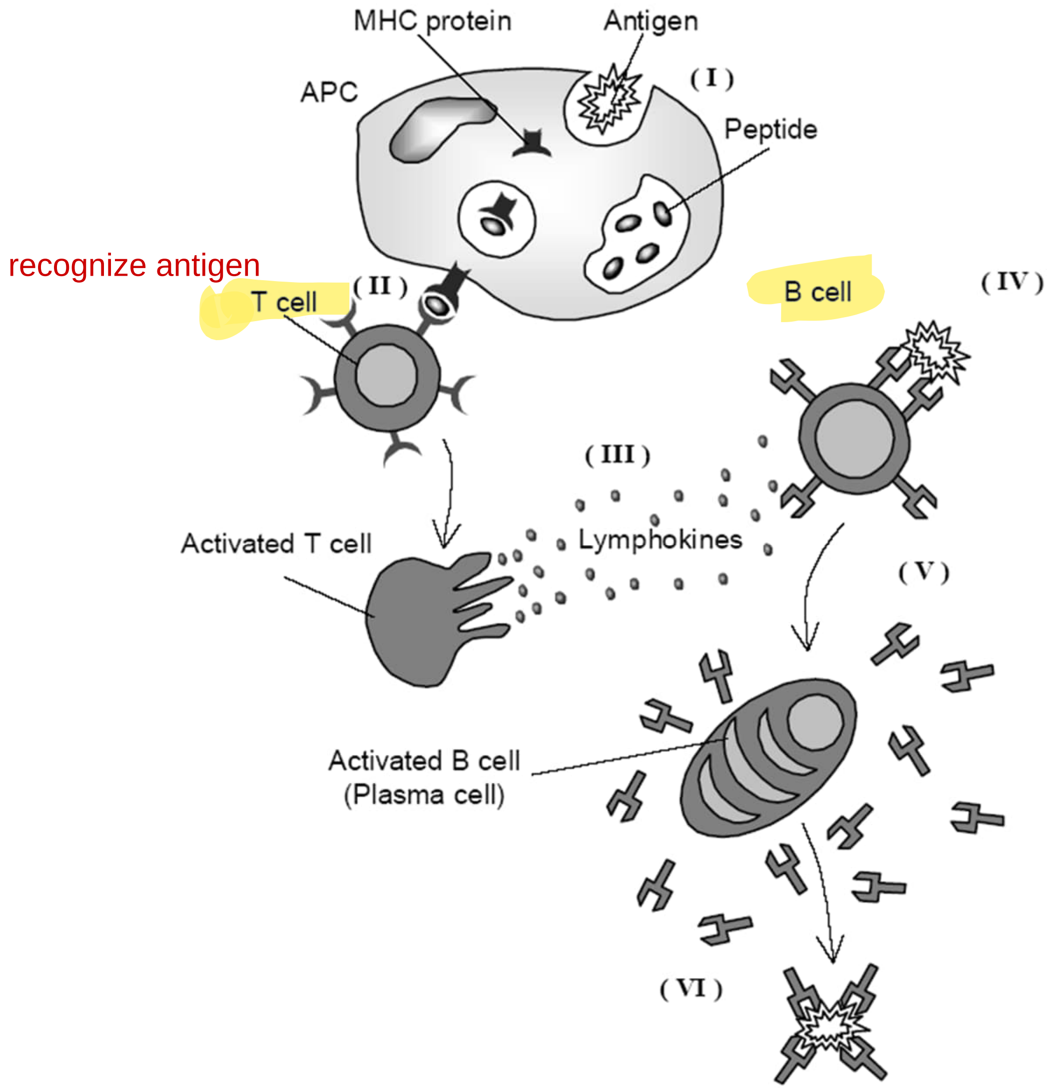
Natural Immune System



Natural Immune System



How Does Natural Immune System Work?



The NSA is used to detect anomalous data in a given dataset by modeling a set of self-patterns and detecting any data points that do not fit the model.

Negative Selection Algorithm (NSA)

- Immune system (B and T cells) is capable of distinguishing *self* from *nonself.*
  - Negative censoring of T cells in thymus
- Negative Selection Algorithm (NSA) mimics mechanism of immune system.
  1. Define *self samples* (representative samples)
  2. Generation of detectors (binary and real-valued)
  3. Negative selection of detectors
  4. Employment of detectors in anomaly detection

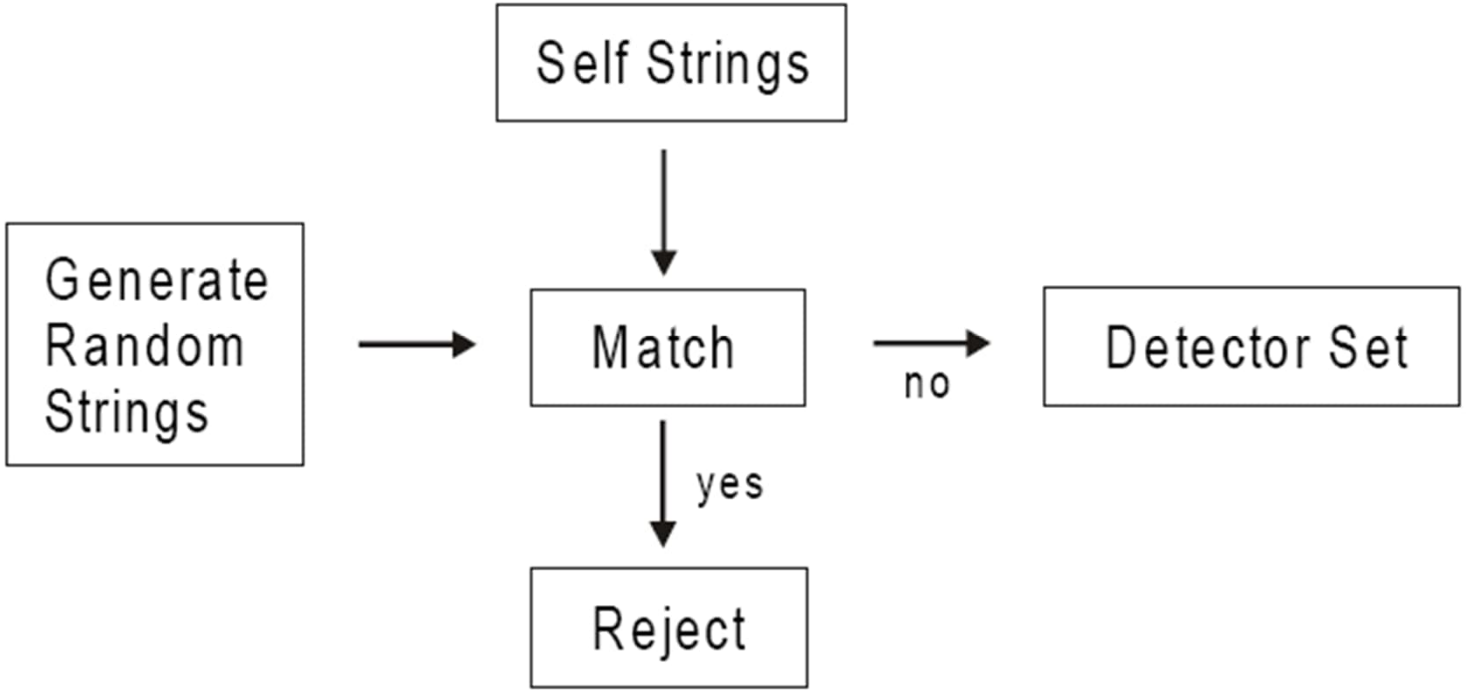
Define self-samples: In this step, a set of representative samples is chosen to represent what is considered "normal" or "self" in the data set.

Generation of detectors: Detectors are generated based on the self-samples. Detectors are binary or real-valued patterns that are used to detect anomalies.

Negative selection of detectors: In this step, detectors that match self-samples are eliminated, and only detectors that do not match any self-sample are kept. This mimics the process of negative selection in the immune system, where immune cells that recognize self antigens are eliminated.

Employment of detectors in anomaly detection: The remaining detectors are then employed to detect anomalies in the data how does it work? EXAM

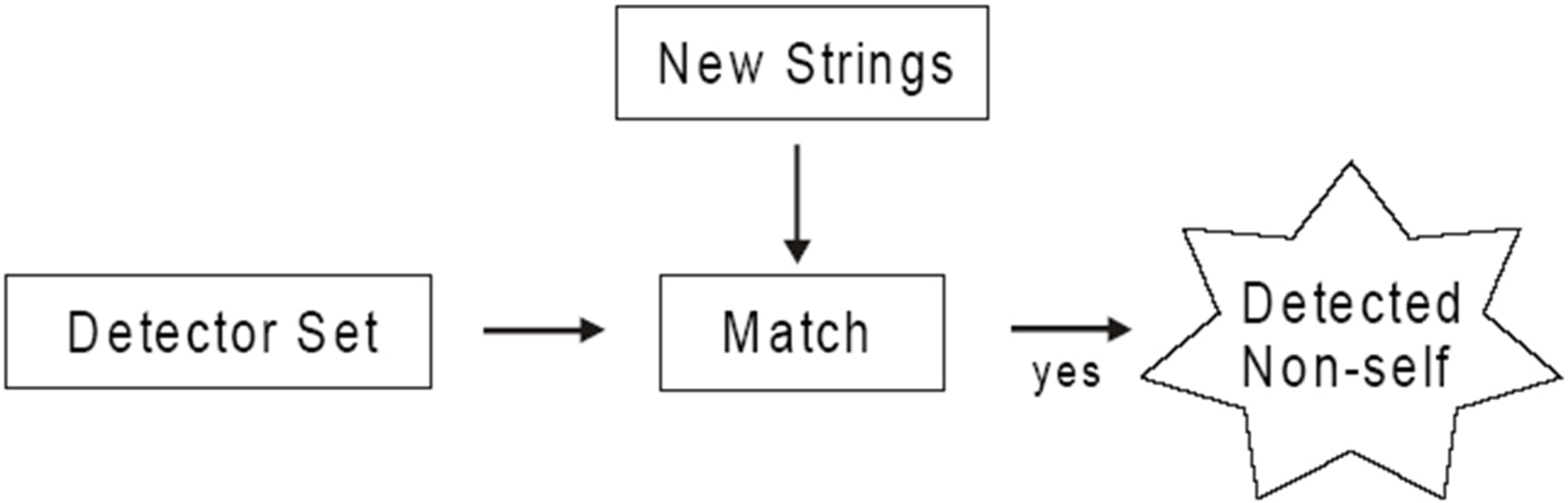
Negative Selection Algorithm (NSA)



Generation of NSA Detectors

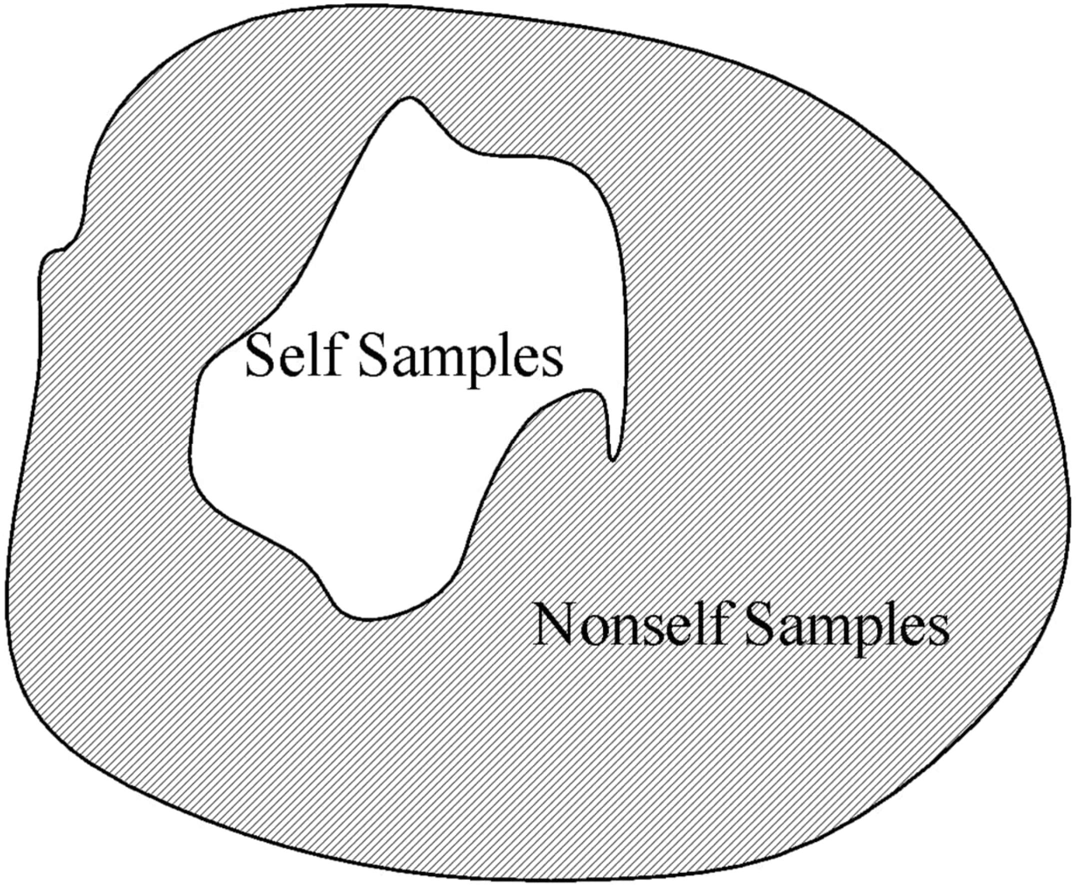
detector can be any shape

Negative Selection Algorithm (NSA)



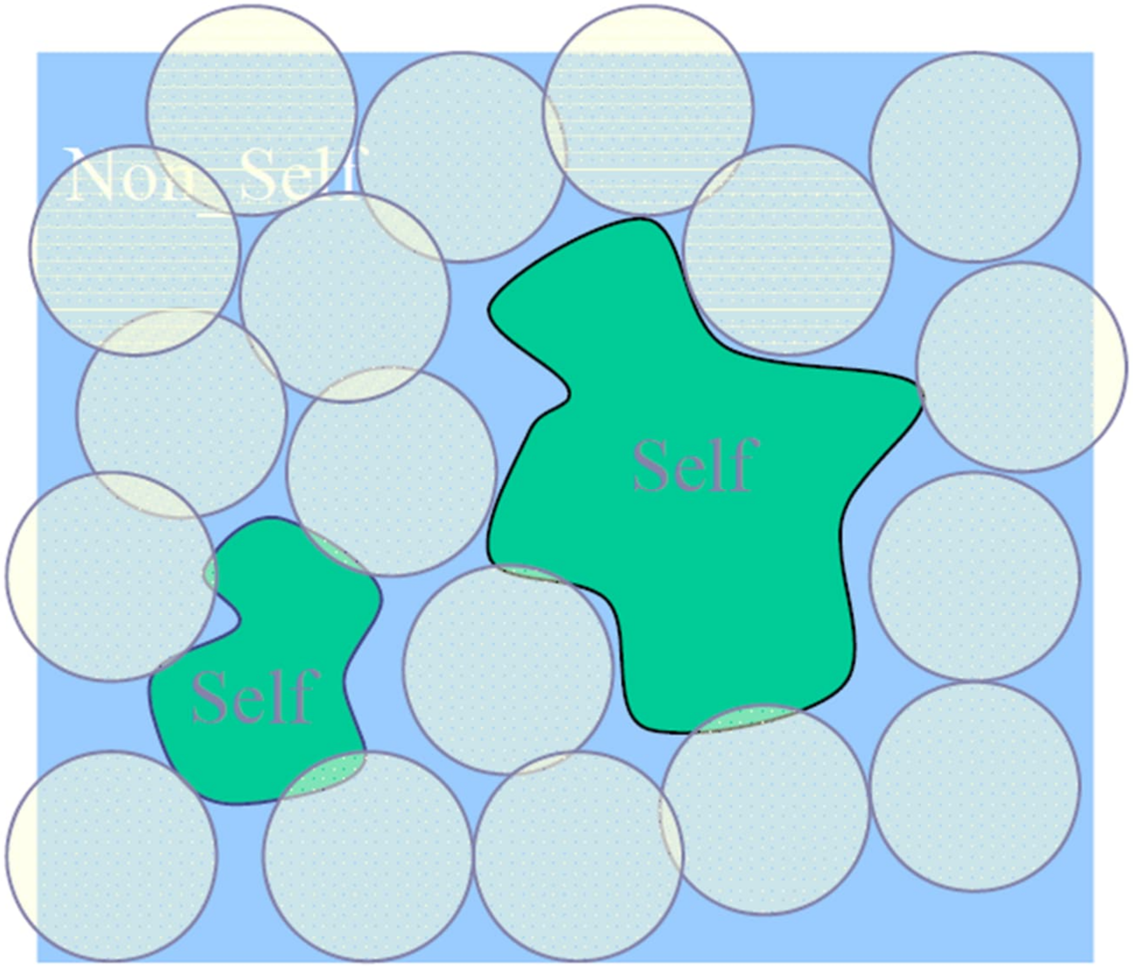
Anomaly Detection Using NSA

Self and Nonself Samples

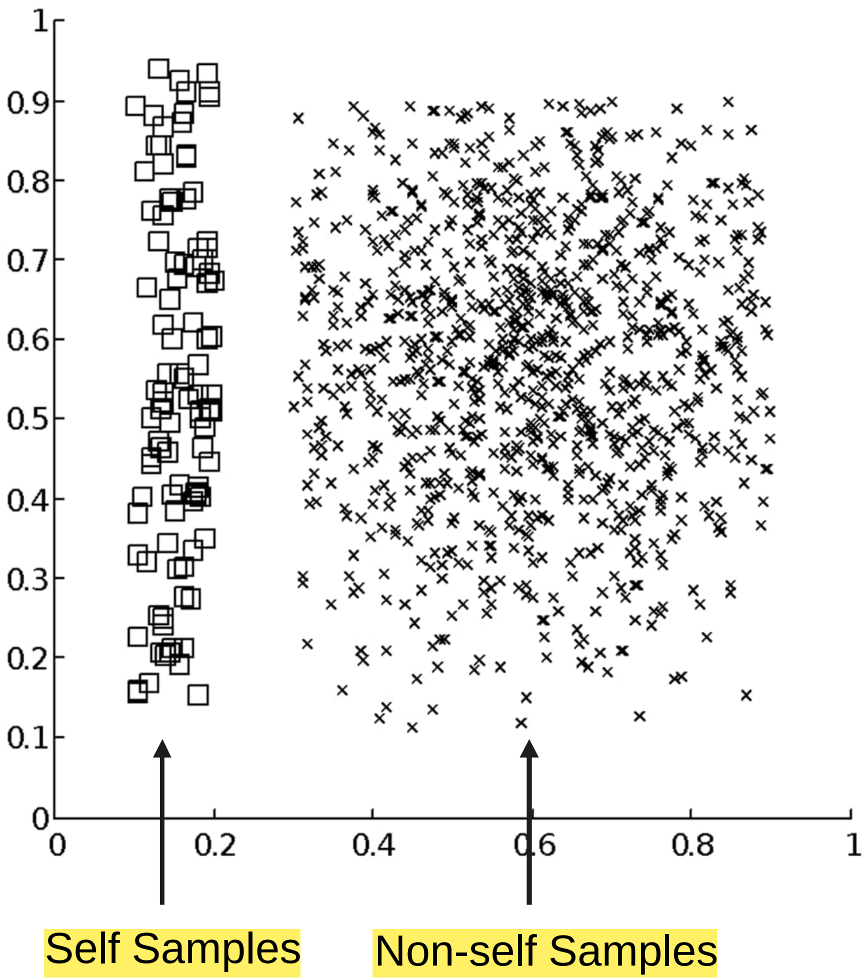


Self-coverage refers to the proportion of the self-patterns that are covered by the detectors in the feature space.

Self and Nonself Coverage in NSA

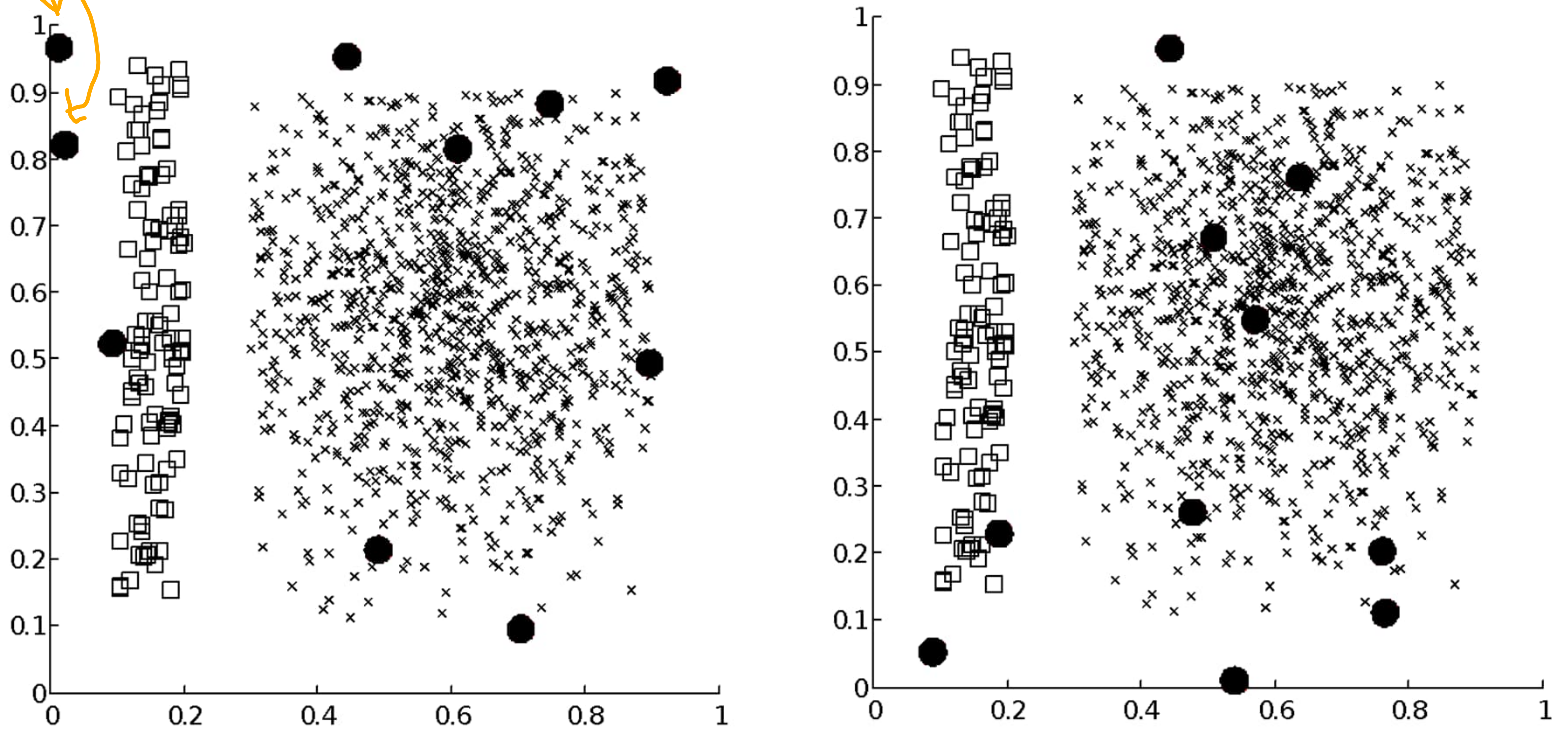


Self and Nonself Samples



some detector is useless(outliers)

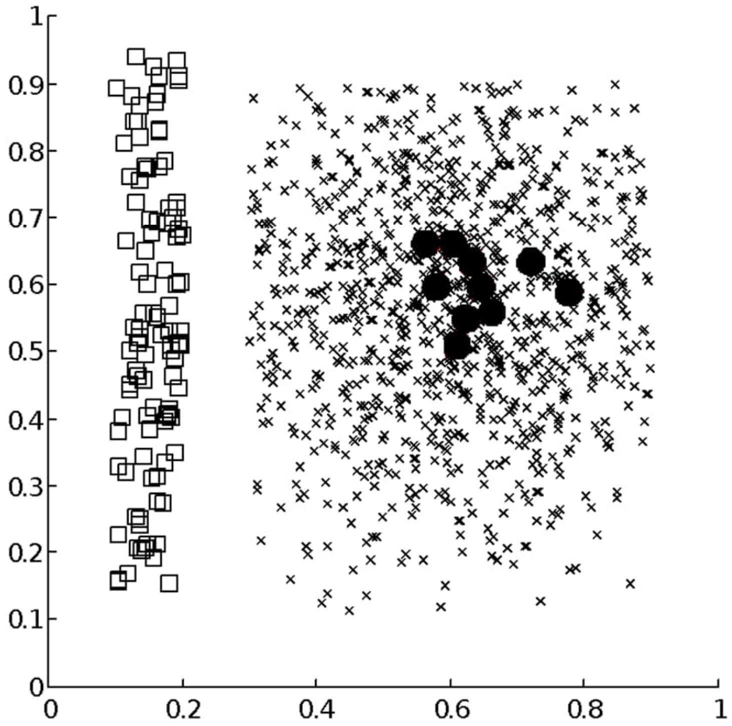
Two Examples of Random Detector Groups

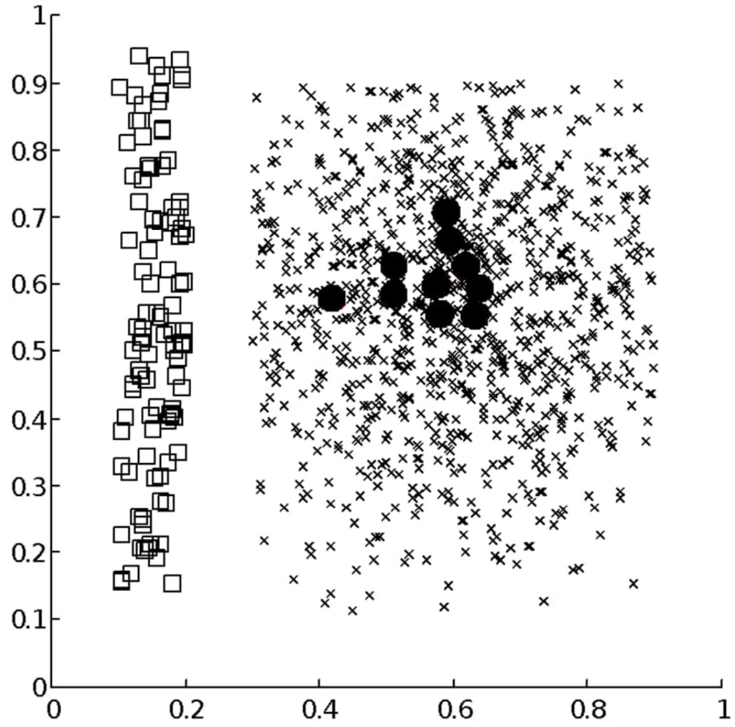


Two Examples of Optimized Detector Groups

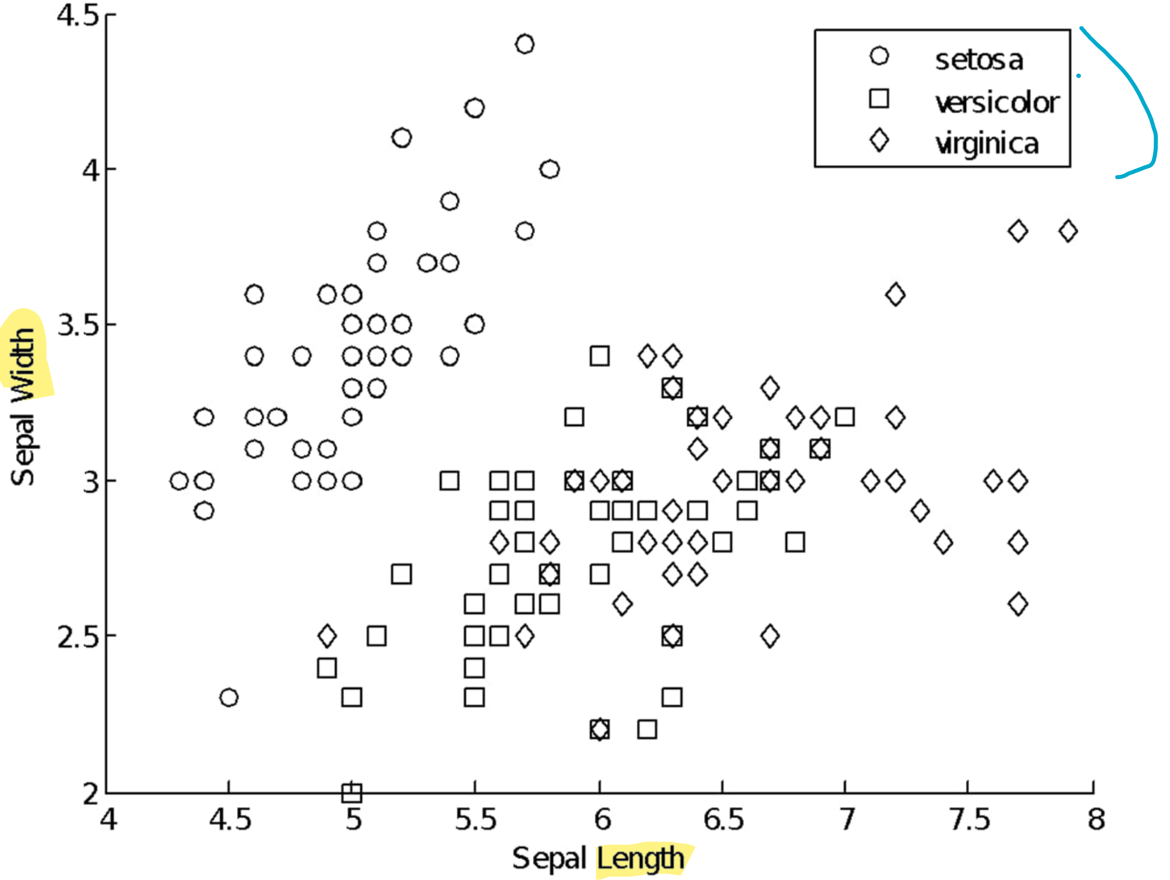
location

(Gao, 2004)

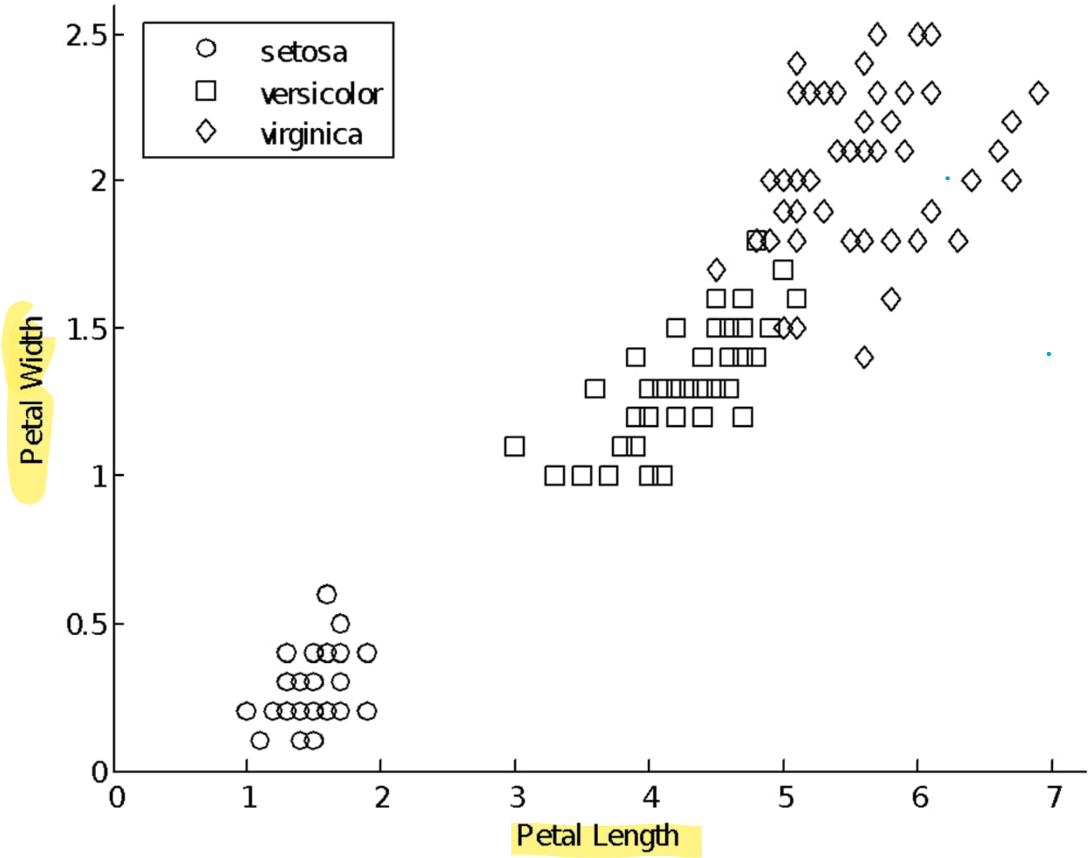




Distribution of Fisher's iris Data in Sepal Length-Sepal Width Dimensions

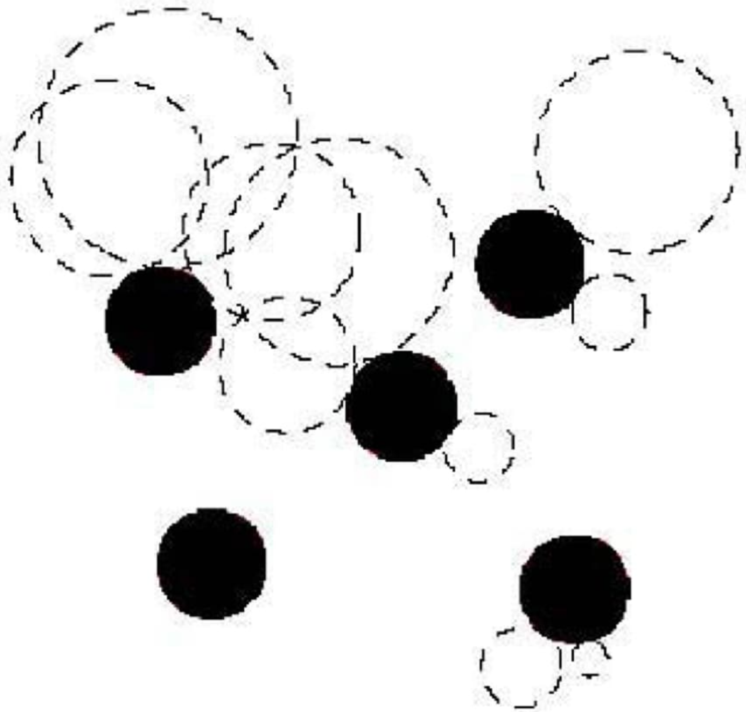


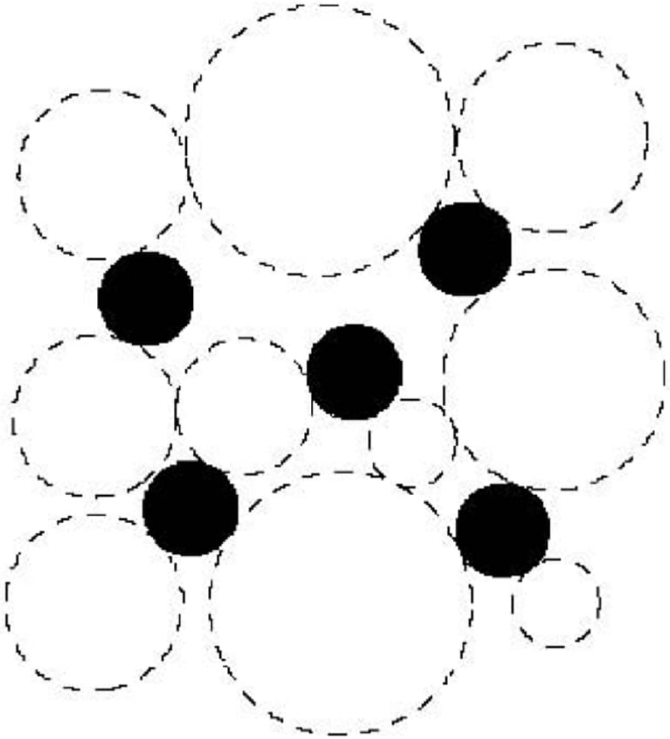
Distribution of Fisher’s iris Data in Petal Length-Petal Width Dimensions



Detector Generation in NSA (Gao, 2006)

sometimes we need that overlapping so it covers more area





Regular NSA Detectors

Optimized NSA Detectors

Anomaly Detection in Fisher's iris Data with NSA

Species	Detection Rates
Versicolor	82%
Setosa	86%

A chaotic time series is one that exhibits irregular and seemingly random fluctuations, making it difficult to distinguish normal behavior from anomalies.

Anomaly Detection in Chaotic Time Series

used for prediction

- Mackey-Glass chaotic time series: data sample changes in time

$$\frac{dx(t)}{dt} = \frac{\beta x(t-\tau)}{1+x^n(t-\tau)} + \gamma x(t)$$

delay

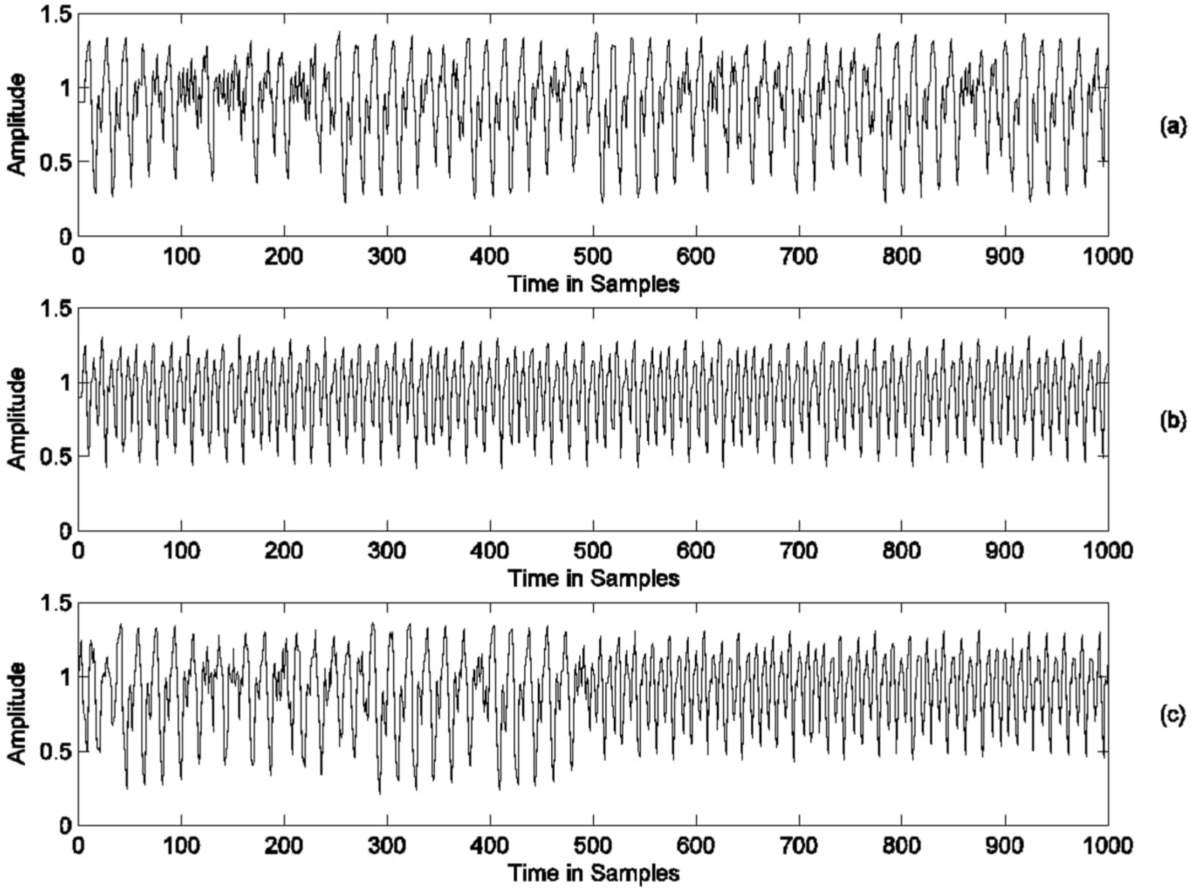
- τ controls the behaviors of Mackey-Glass time series. where is the time series, is time, is a time delay, is a non-linearity parameter, and is a scaling parameter that controls the amplitude of the oscillations in the time series.
  - $\tau = 30$ and $\tau = 17$
- Anomaly detection rate can be improved by neural networks-based NSA.

Mackey-Glass Time Series

$\tau = 30$ →

$\tau = 17$ →

Fresh Data →



Anomaly Detection in Mackey-Glass Time Series Using Adaptive NSA

Before Training

$M = 57 \quad L = 9$

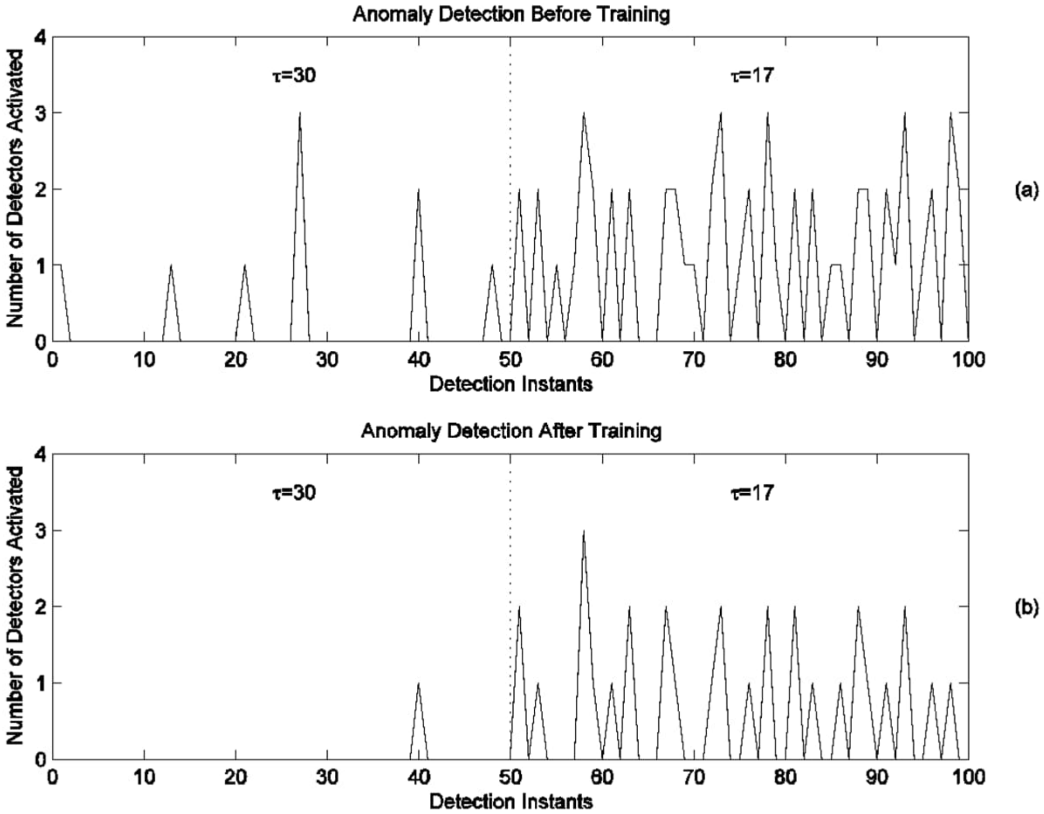
$\eta = 86\%$

detectiion rate

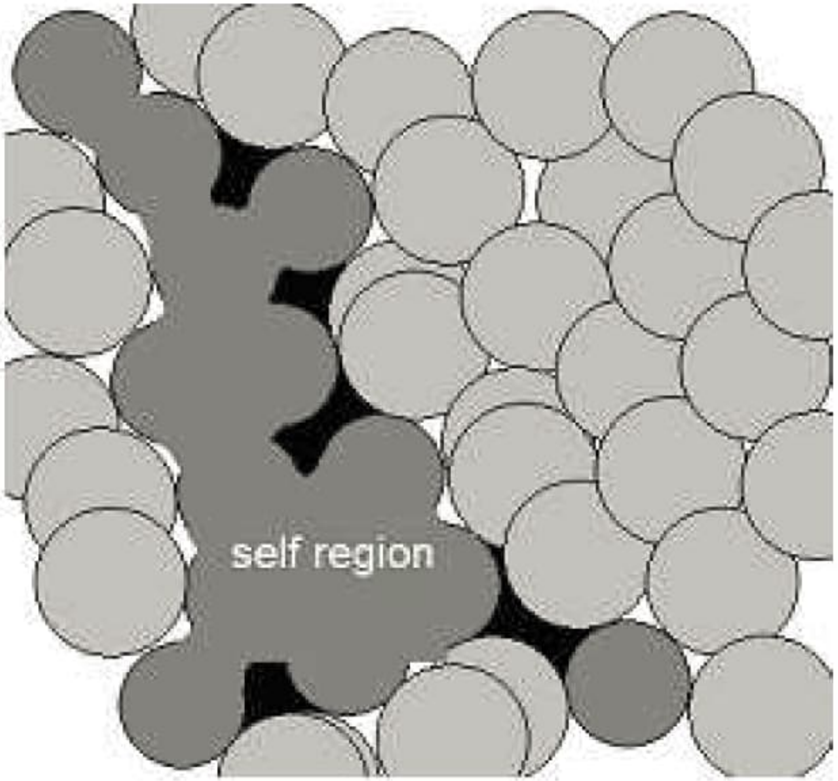
After Training

$M = 31 \quad L = 1$

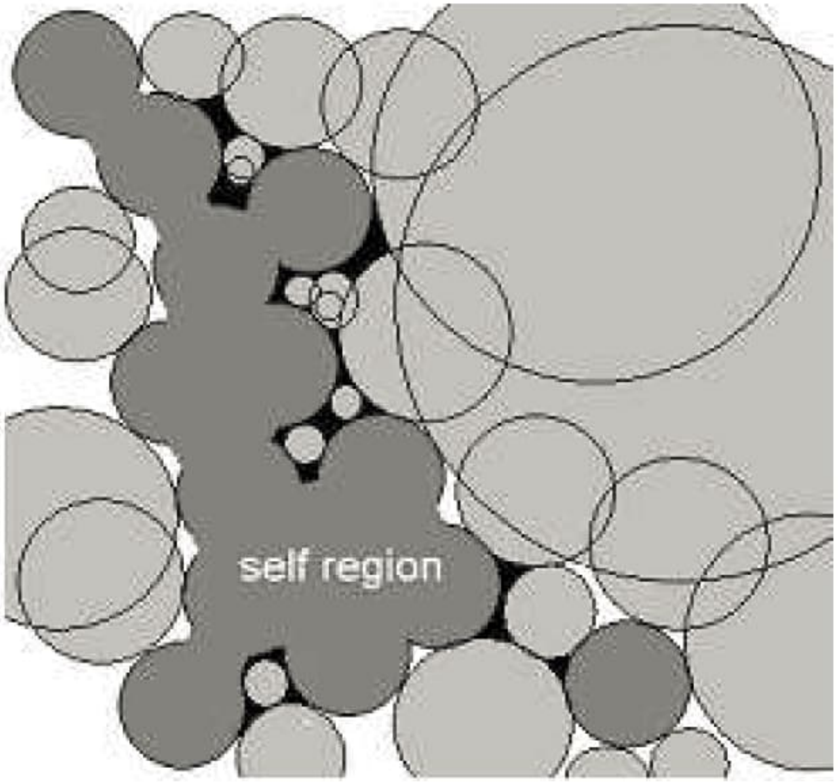
$\eta = 97\%$



NSA Detectors with Variable Properties (Ji and Dasgupta 2004)

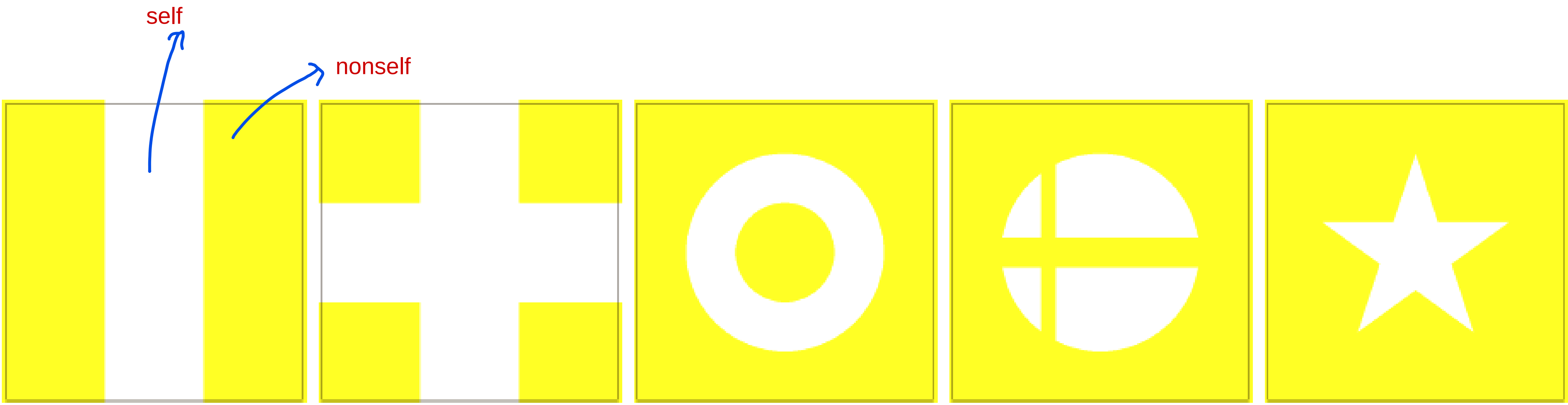


(a) Constant-sized detectors

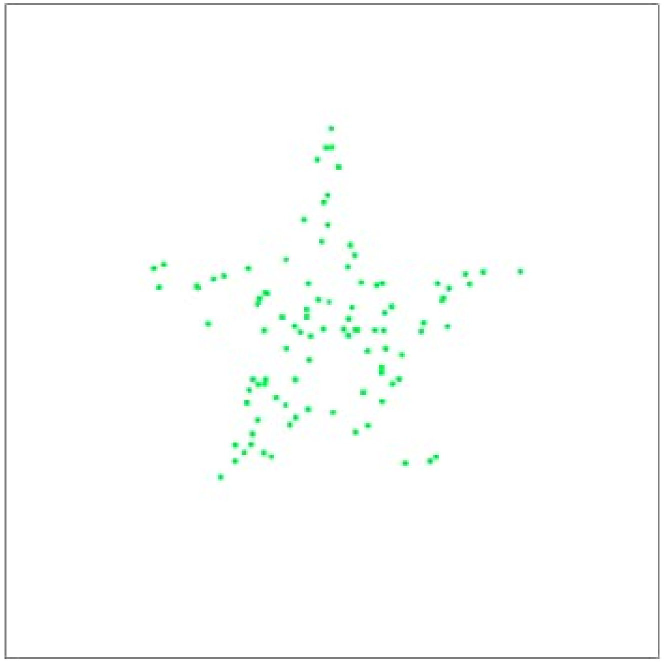


(b) Variable-sized detectors

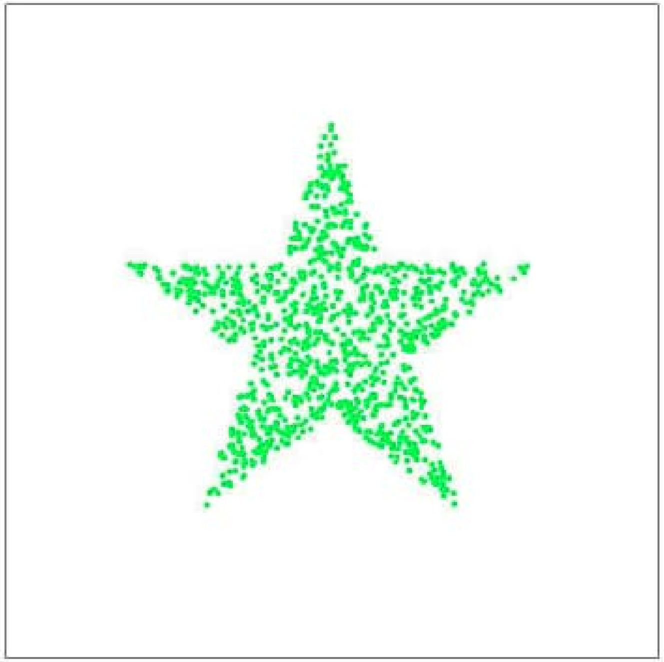
An Example of Variable NSA Detectors



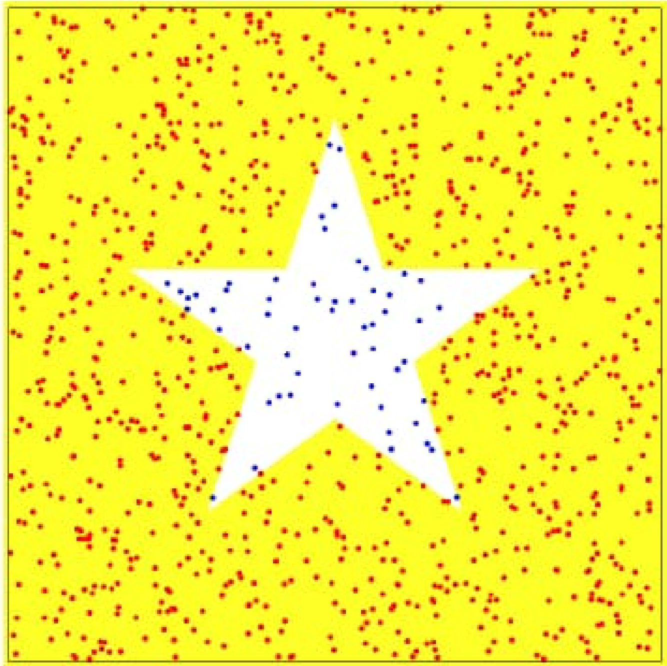
An Example of Variable NSA Detectors



(a) 100 points of self sample

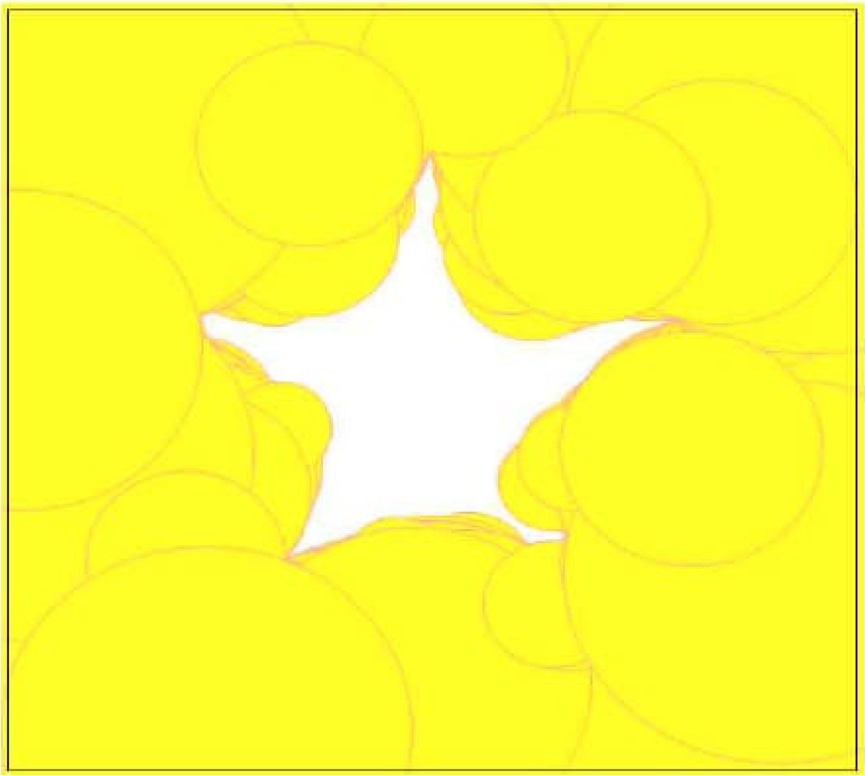


(b) 1000 points of self sample

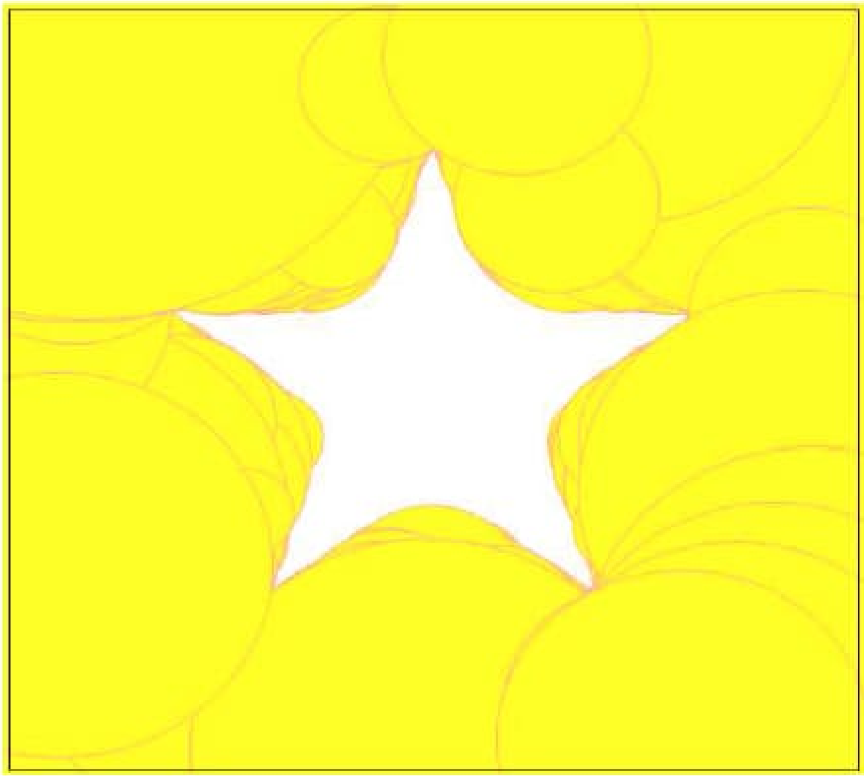


(c) 1000 points of test data

An Example of Variable NSA Detectors



(a) Trained with 100 points



(b) Trained with 1000 points

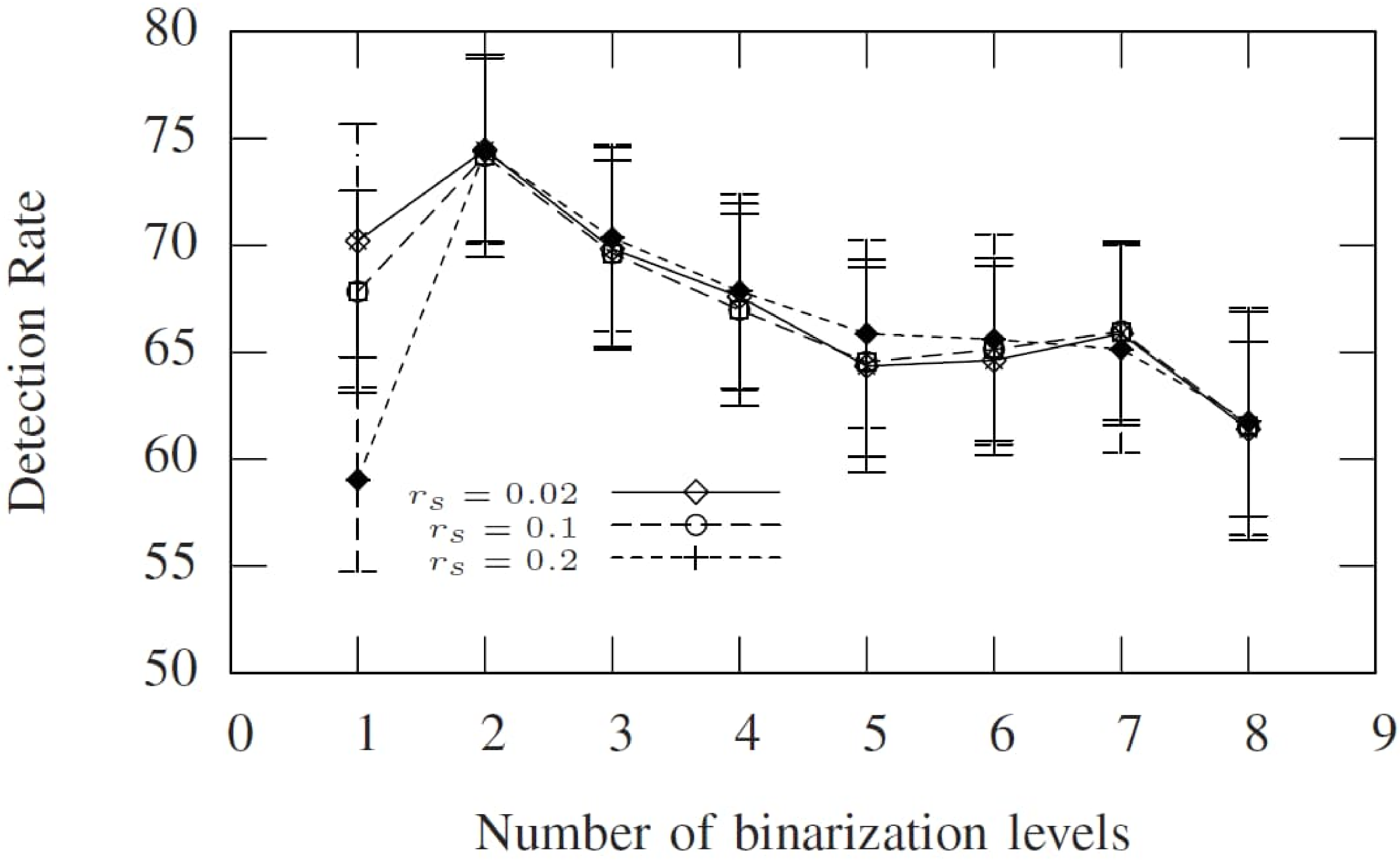
NSA in Dental Disease Diagnosis (Ji, *et. al*, 2006)

Normal

Malocclusion in Dental Images

Abnormal

Malocclusion Detection Rate



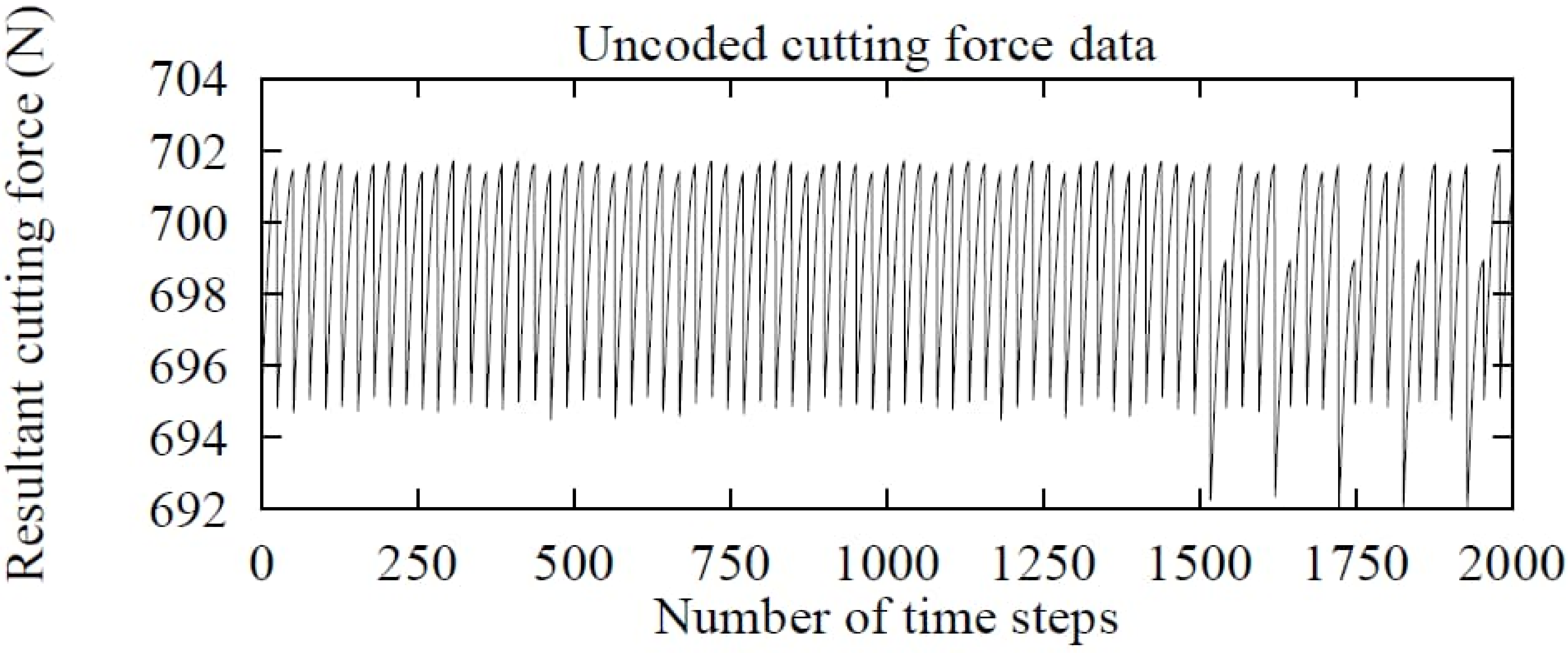
Conclusions

- Anomaly detection is an important pattern recognition topic.
- NSA can be used for anomaly detection based on only normality.
- A few application examples have demonstrated the effectiveness of the NSA in anomaly detection.
- Performance comparisons need to be made between the NSA and other anomaly detection methods, e.g., Support Vector Machine (SVM).

Appendix: Negative Selection Algorithm in Fault Detection

(Case study)

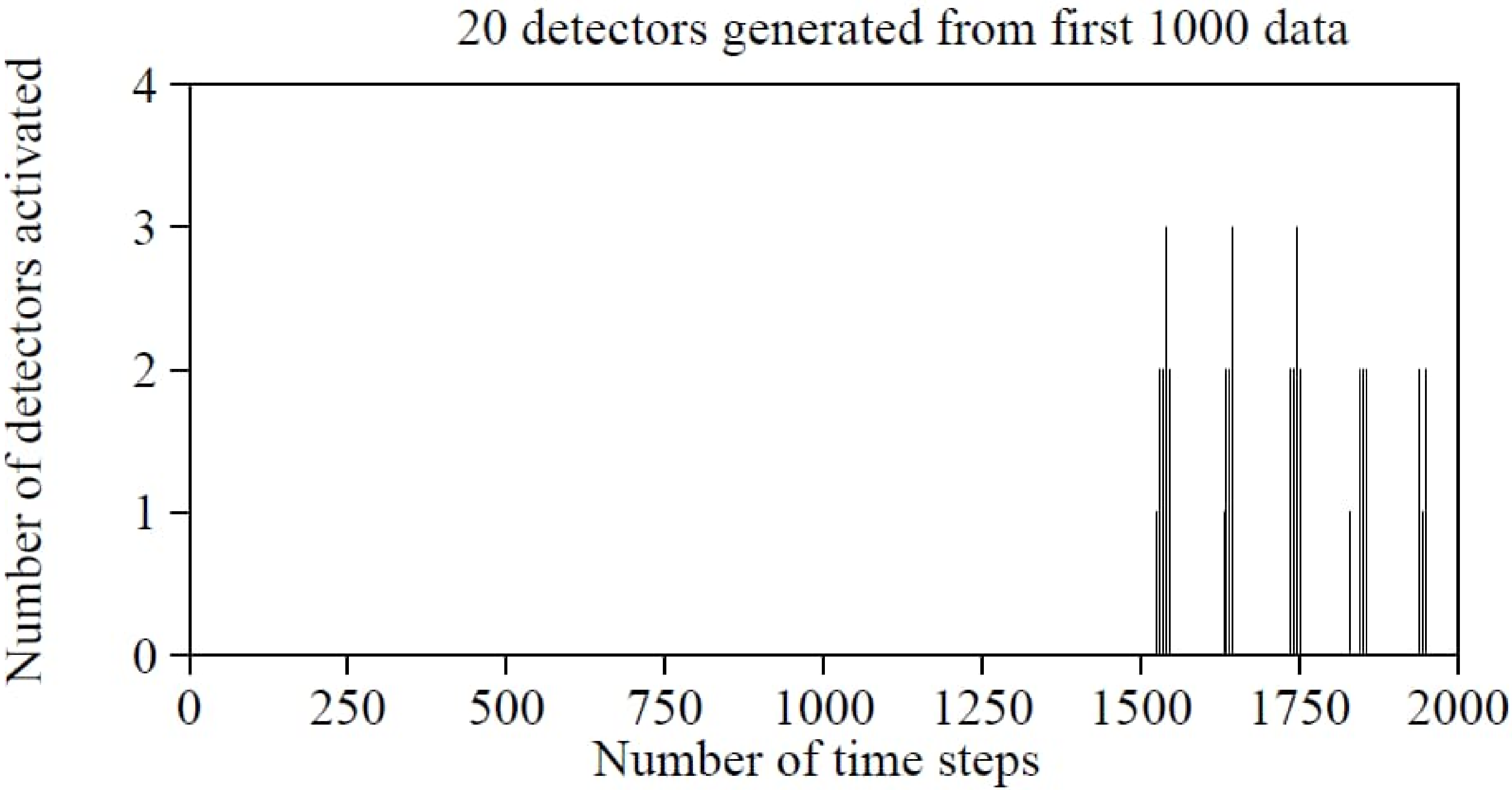
Tool Breakage Detection in Milling Operations (Dasgupta and Forrest, 1995)



Cutting Parameters

Experiment No.	Axial Depth cut(mm)	Feed rate (m/min)	Spindle Speed (rpm)	Spindle diameter (mm)
1	1.34	90.6	800	50
2	1.016	125.4	500	40
3	1.524	50.8	700	40

Detectors in Tool Breakage Detection



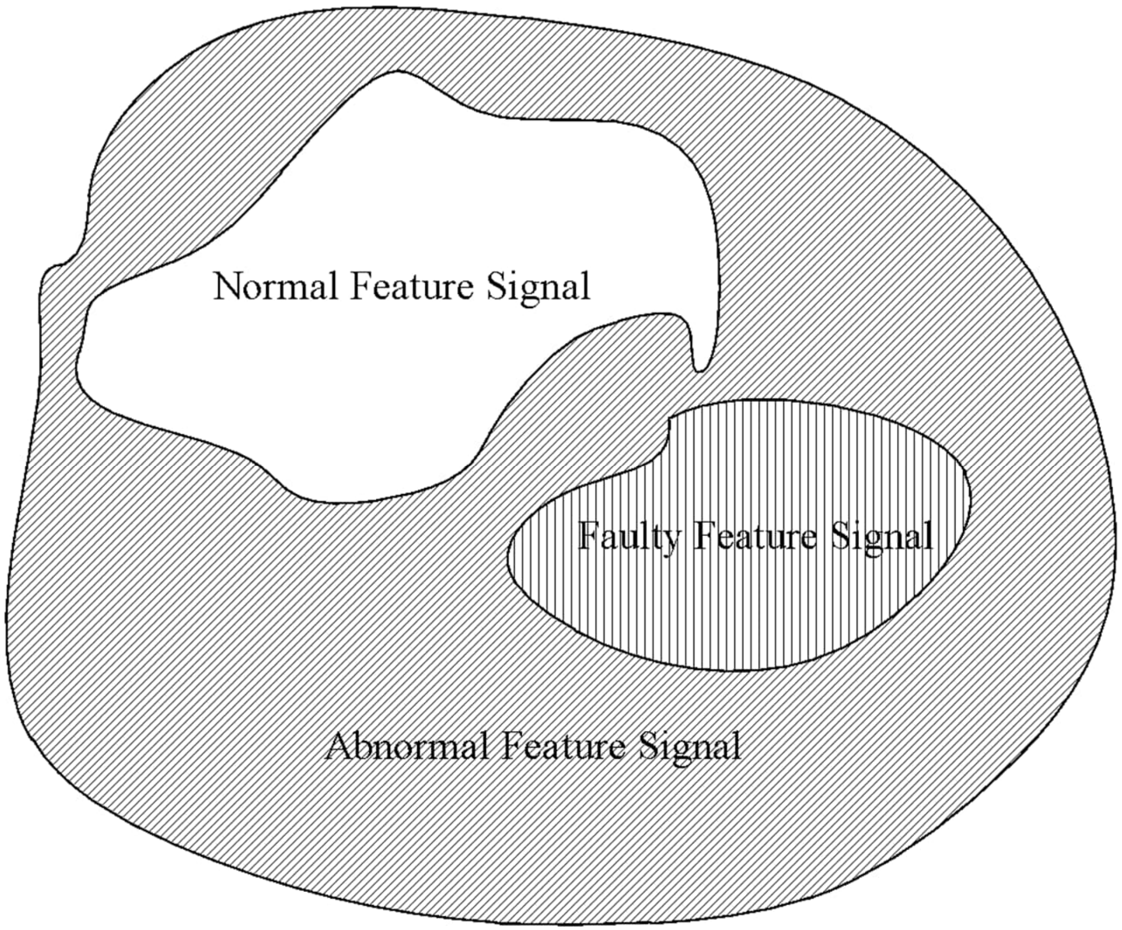
Detection Rates

Encoding parameters	Matching threshold(r)	Number of detectors(R)	Tool Breakage Detection	
			Mean(Std. dev)	Detection rate
Win_size = 5	10	50	14.30(2.32)	59.58%
Win_shift= 5	9	40	17.57(2.25)	74.32%
$l = 30$, $S = 200$	8	30	22.16(2.57)	91.64%
Win_size = 7	12	40	10.36(3.36)	62.78%
Win_shift= 7	10	30	20.38(5.57)	75.56%
$l = 42$, $S = 142$	9	20	30.75(7.91)	93.28%

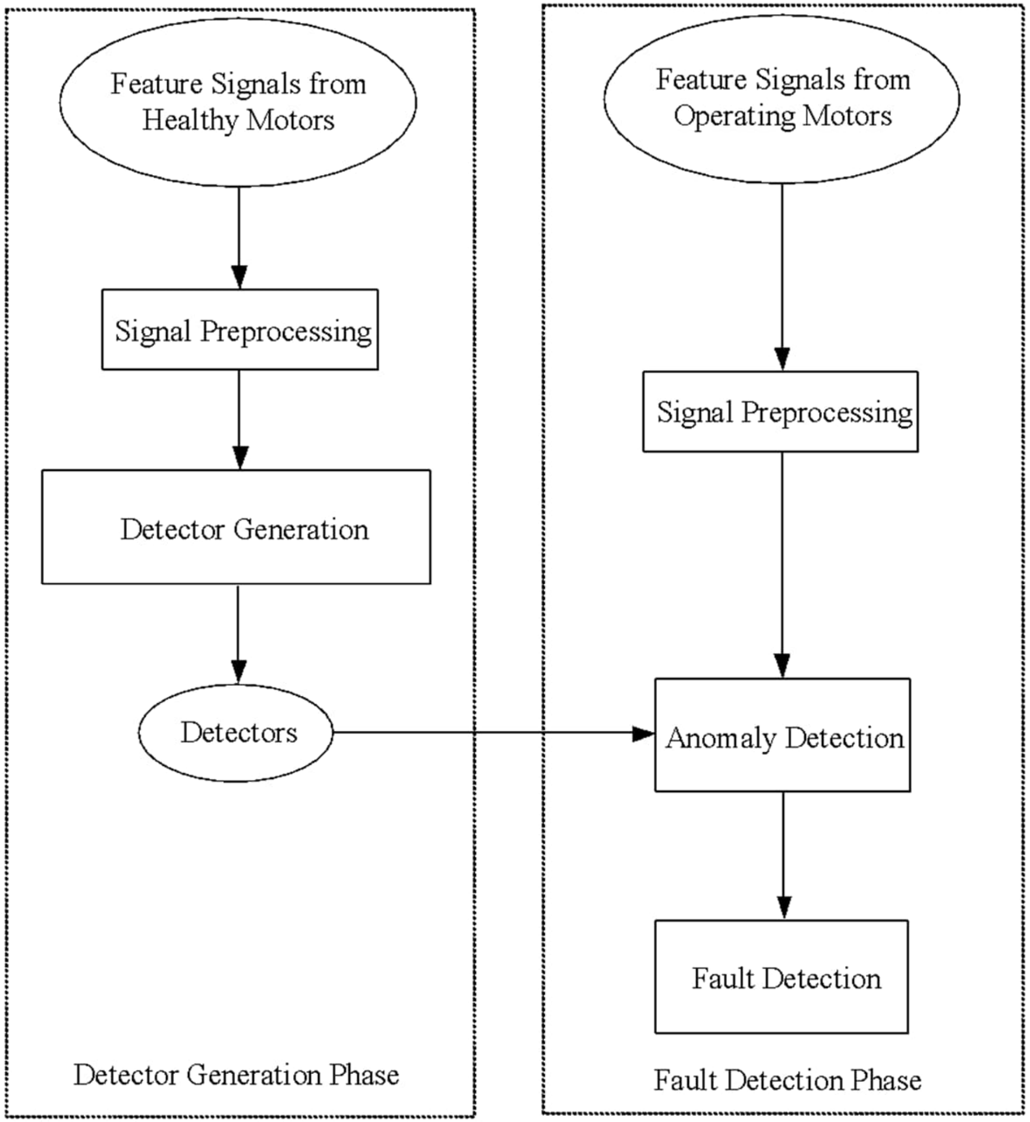
NSA-based Motor Fault Detection

- Monitoring of the working conditions of the running motors is necessary in maintaining their normal status.
- Anomaly in the feature signals acquired from the faulty motors is caused by faults.
- Motor fault detection is converted to a typical problem of data anomaly detection.

Normal, Abnormal, and Faulty Feature Signals of Motors



NSA in Motor Fault Detection



NSA-based Motor Fault Detection

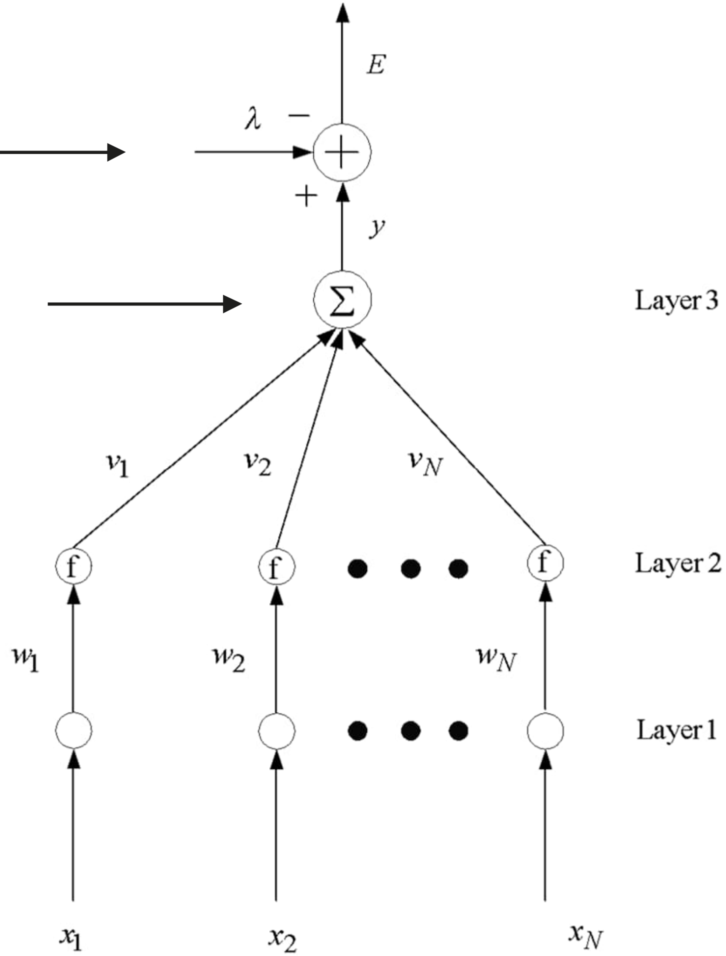
- Anomaly in feature signals is caused by faults.
  - Healthy feature signals: self
  - Faulty feature signals: nonself
- Neural networks are combined with NSA.
  - NSA detectors are built up on the structure of BP neural networks.
  - Neural networks training algorithm is applied.

Neural Networks-based NSA (Gao, 2010)

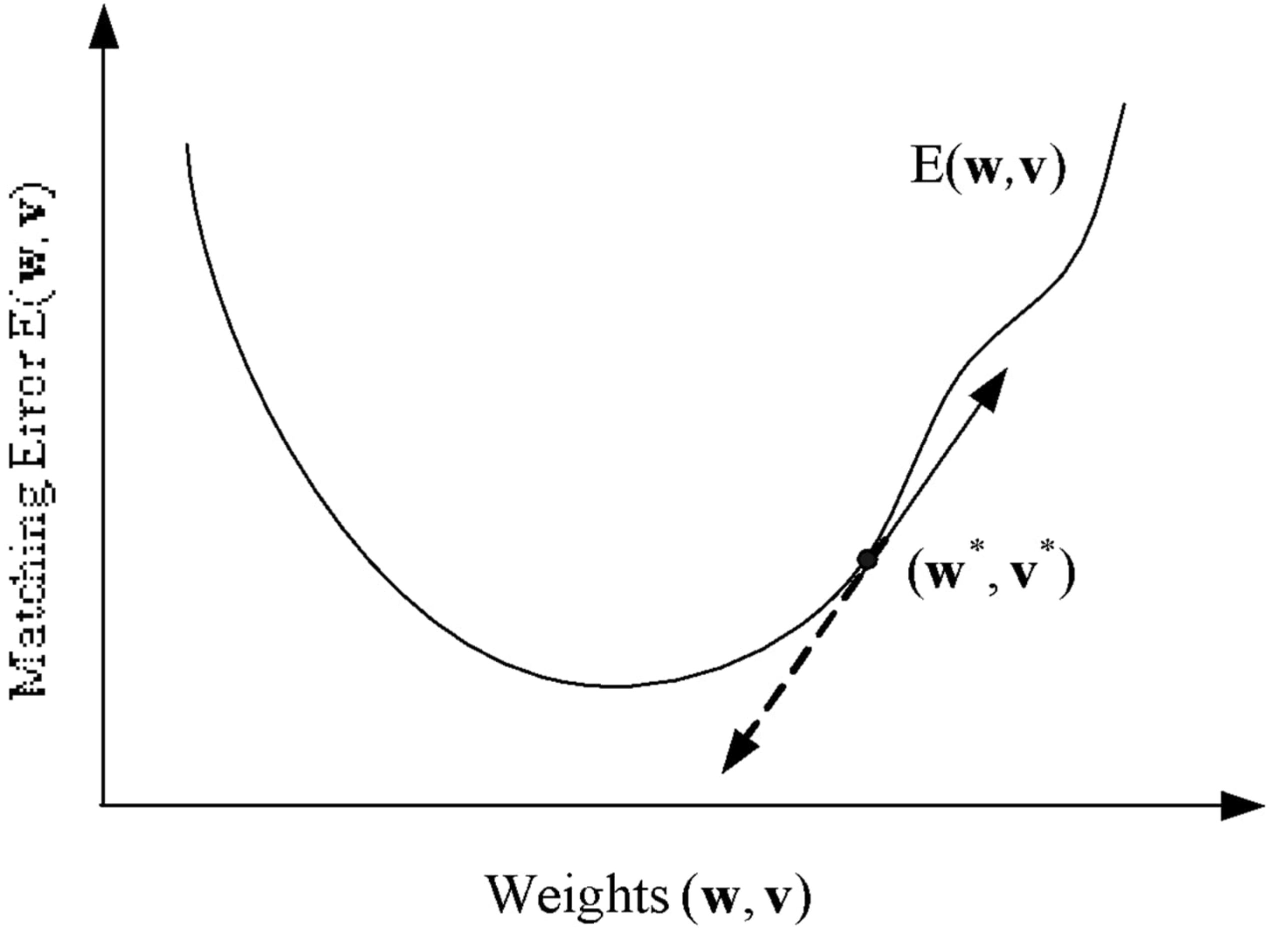
$$E = y - \lambda$$

$$y = \sum_{i=1}^{N} v_i \, \mathrm{f}(d_i) = \sum_{i=1}^{N} v_i \, \mathrm{f}\left[(x_i - w_i)^2\right]$$

$$d_i = (x_i - w_i)^2$$



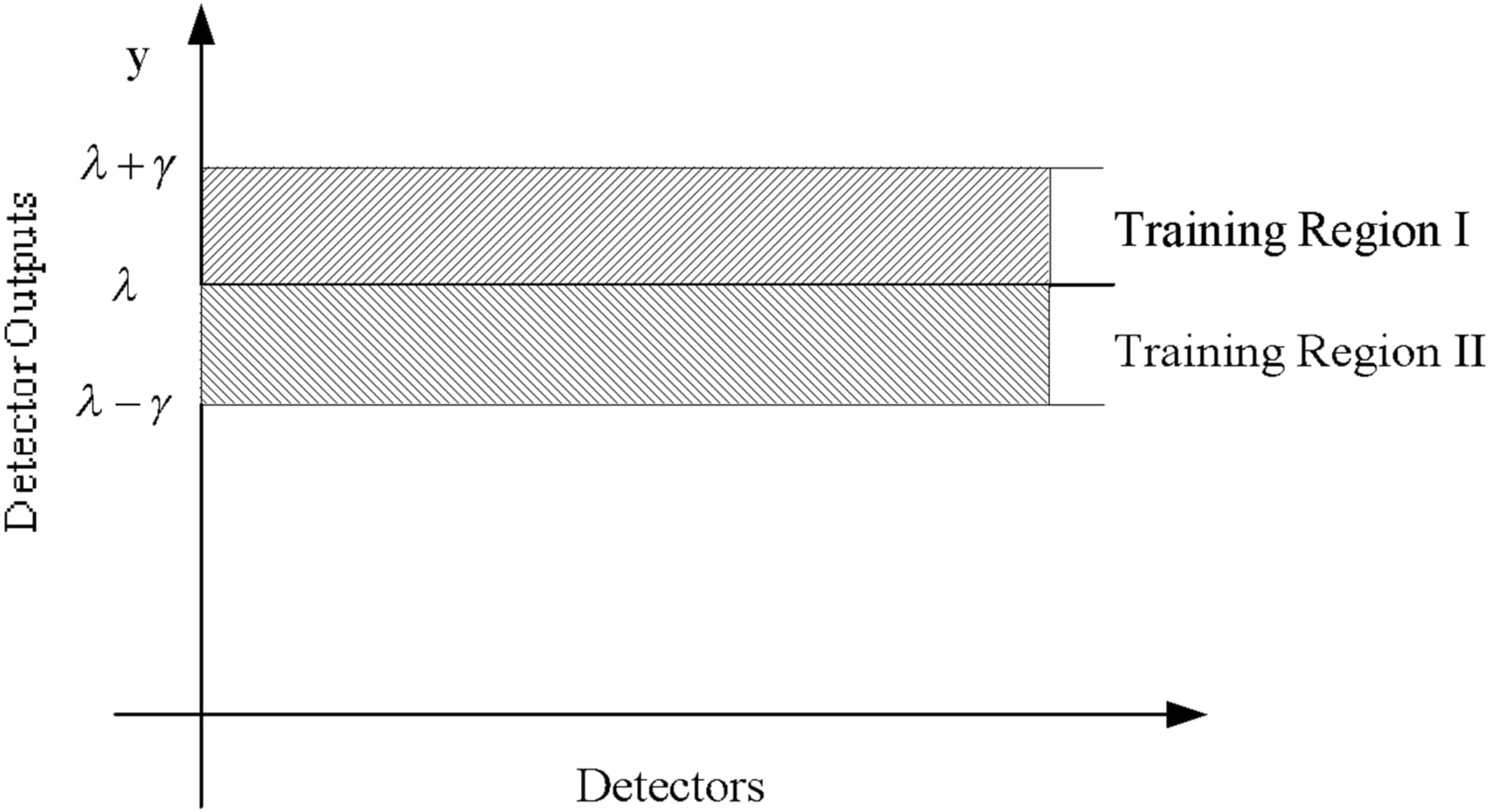
Training of Neural Networks-based NSA



Margin Training Strategy of NSA Detectors

- Case1: (for faulty plant feature signals only): if $0 < E < \gamma$, detectors are trained using normal BP learning algorithm to decrease *E*; otherwise, no training is employed.
- Case 2 (for healthy plant feature signals only): if $-\gamma < E < 0$, detectors are trained using 'positive' learning algorithm to increase *E*; otherwise, no training is employed.

Margin Training of NSA Detectors in Fault Detection



Inner Raceway Fault Detection of Bearings

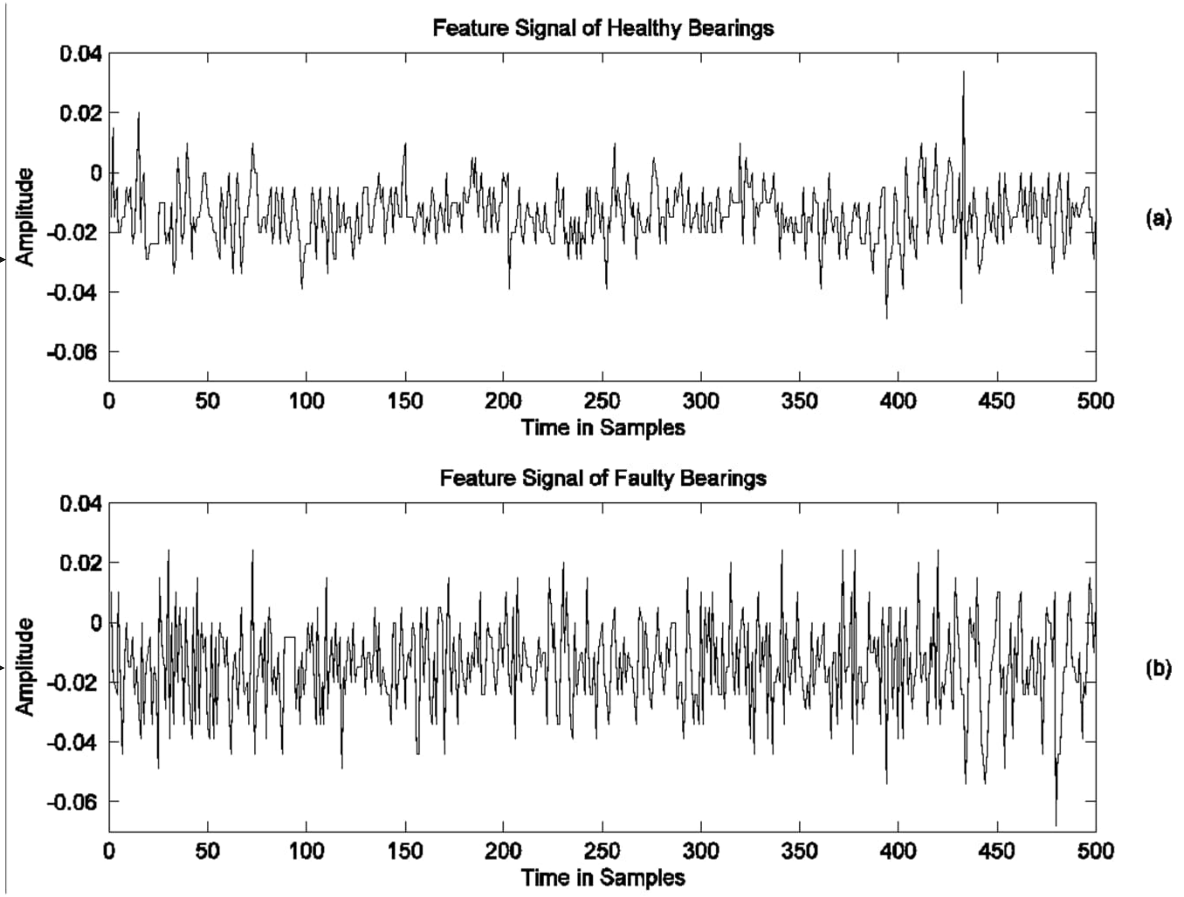
- Bearings are important components in rotating machinery.
- Defect on the inner raceway is a common but typical fault of bearings.
- Fault detection is based on vibration signals of bearings.
  - A sensor mounted on eight-ball bearings with a motor rotation speed at 1,782 rpm

Inner Raceway Fault of Bearings

Features Signals of Healthy and Faulty Bearings

Healthy Bearings

Faulty Bearings



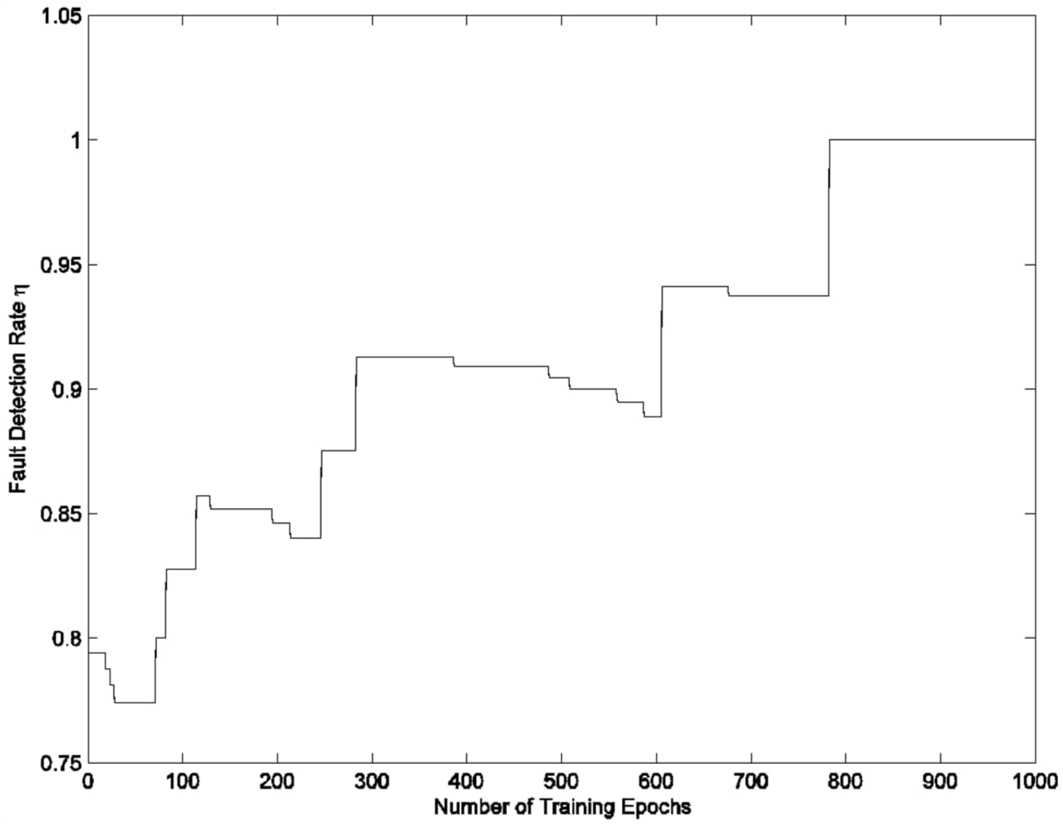
Fault Detection Rate of Neural Networks-based NSA

Before Training

$$M = 27$$

$$L = 7$$

$$\eta = 79\%$$



After Training

$$M = 15$$

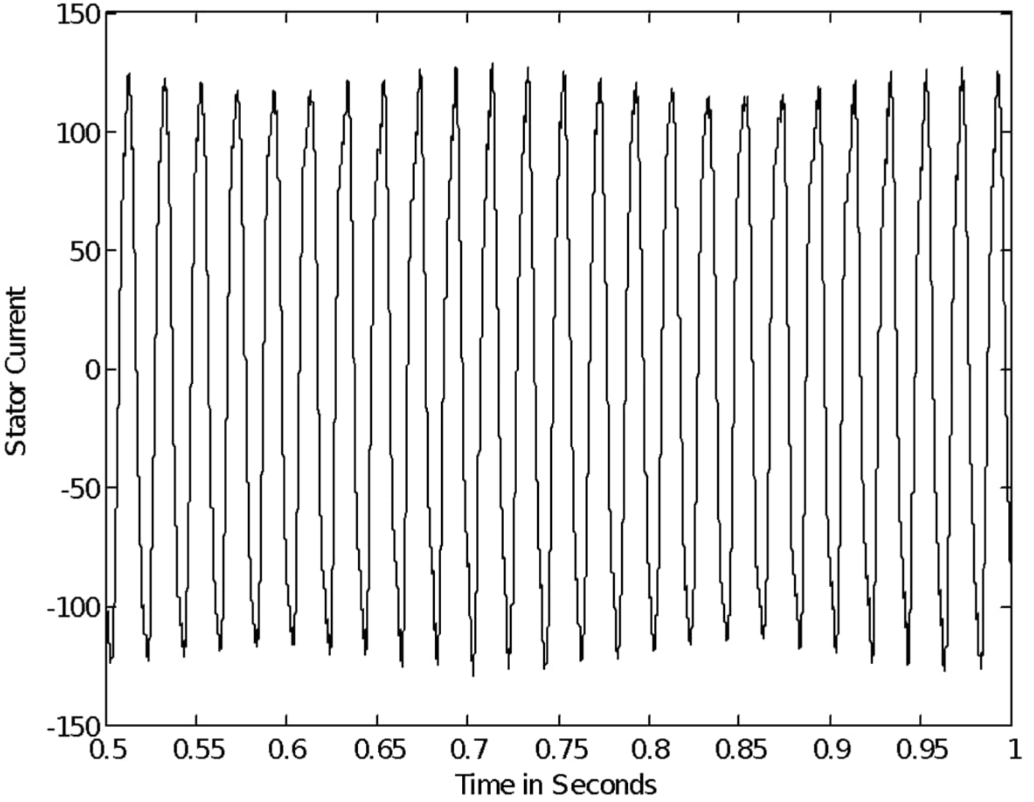
$$L = 0$$

$$\eta = 100\%$$

Motor Fault Detection using NSA (Gao, 2012)

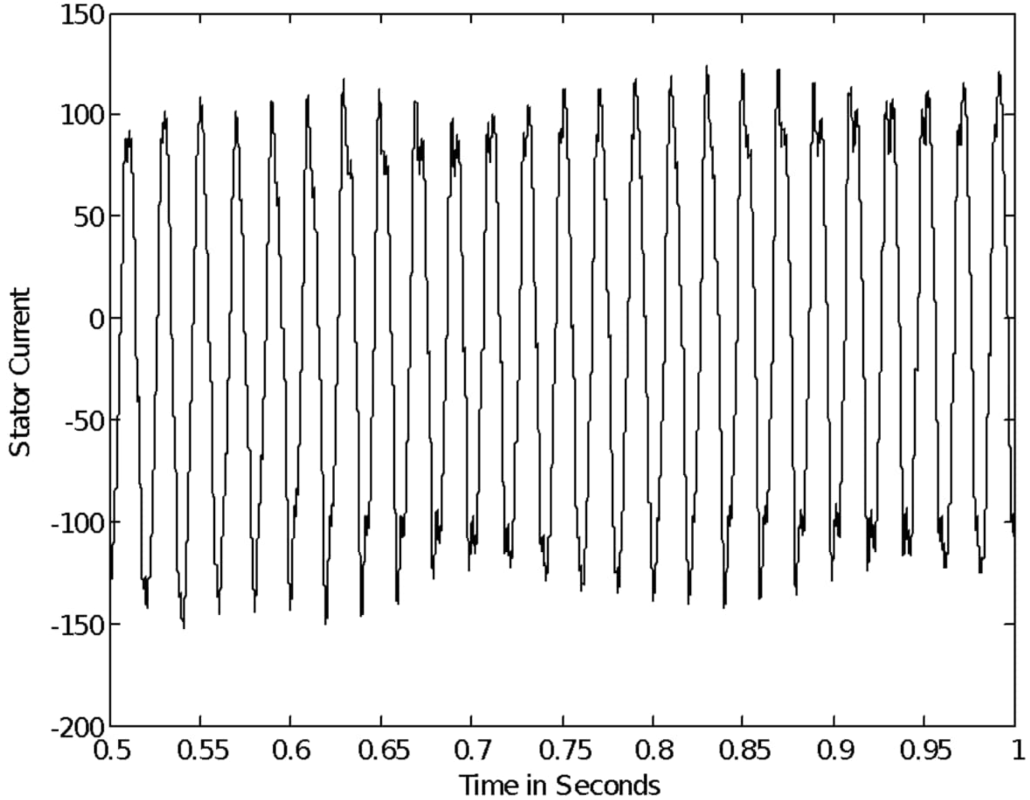
- Two kinds of motor faults are considered.
  - Rotor fault
  - Stator fault
- Stator current signals are used as feature signals.
- Both healthy and faulty motors are running with/without varying loads.
- Fault detection rate is $\eta = \dfrac{B}{A+B} \times 100\%$

Feature Signals for Motor Fault Detection



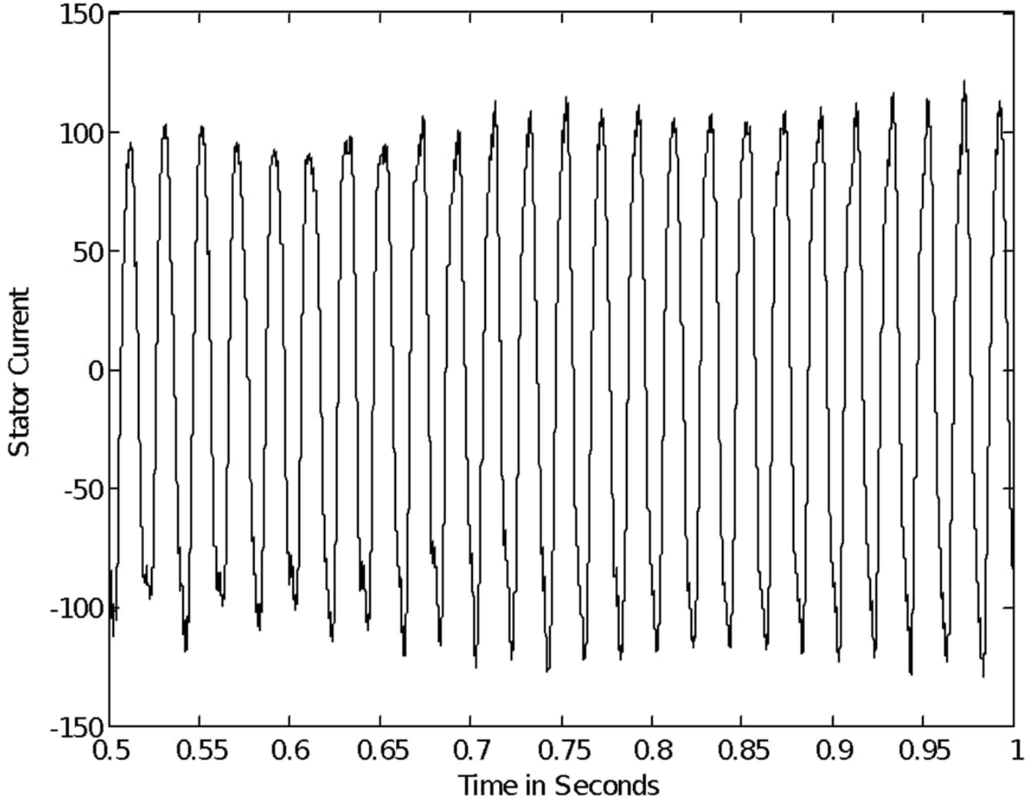
Healthy Motor

Feature Signals for Motor Fault Detection



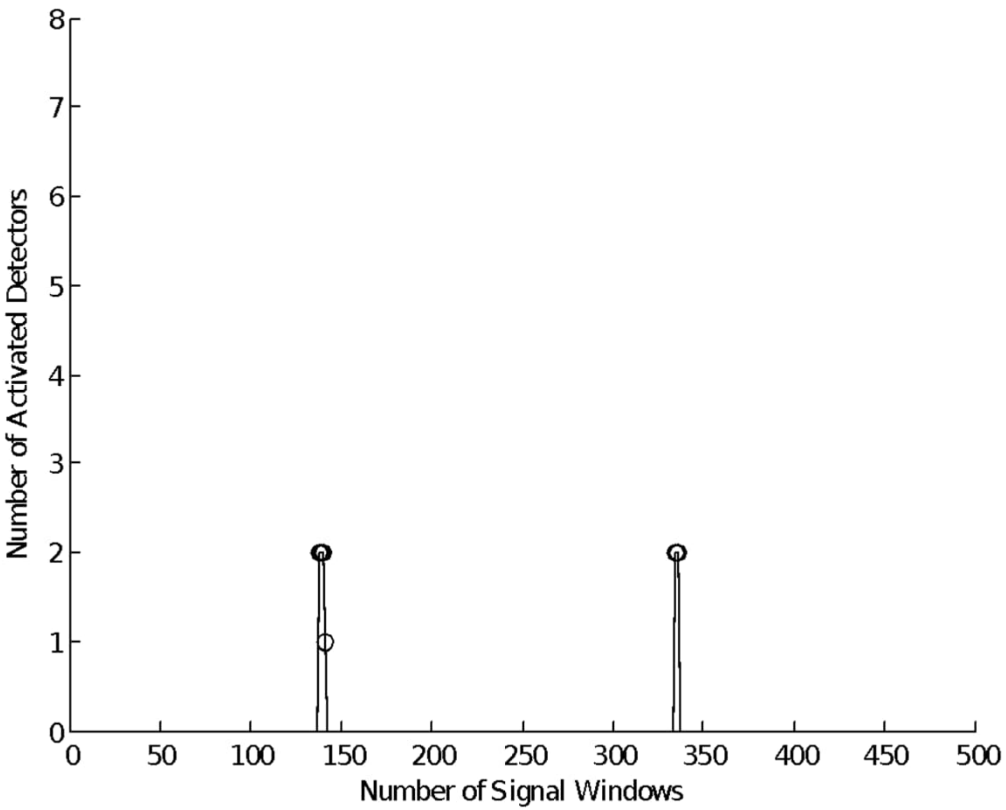
Broken Rotor

Feature Signals for Motor Fault Detection



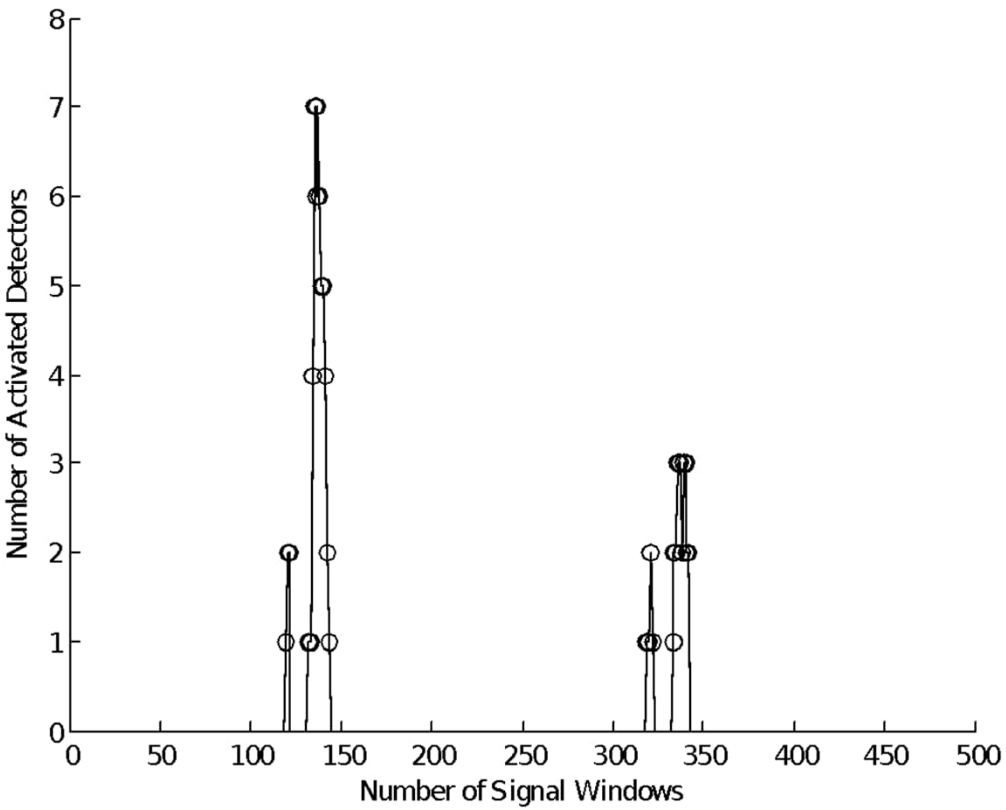
Broken Stator

Motor Fault Detection Results



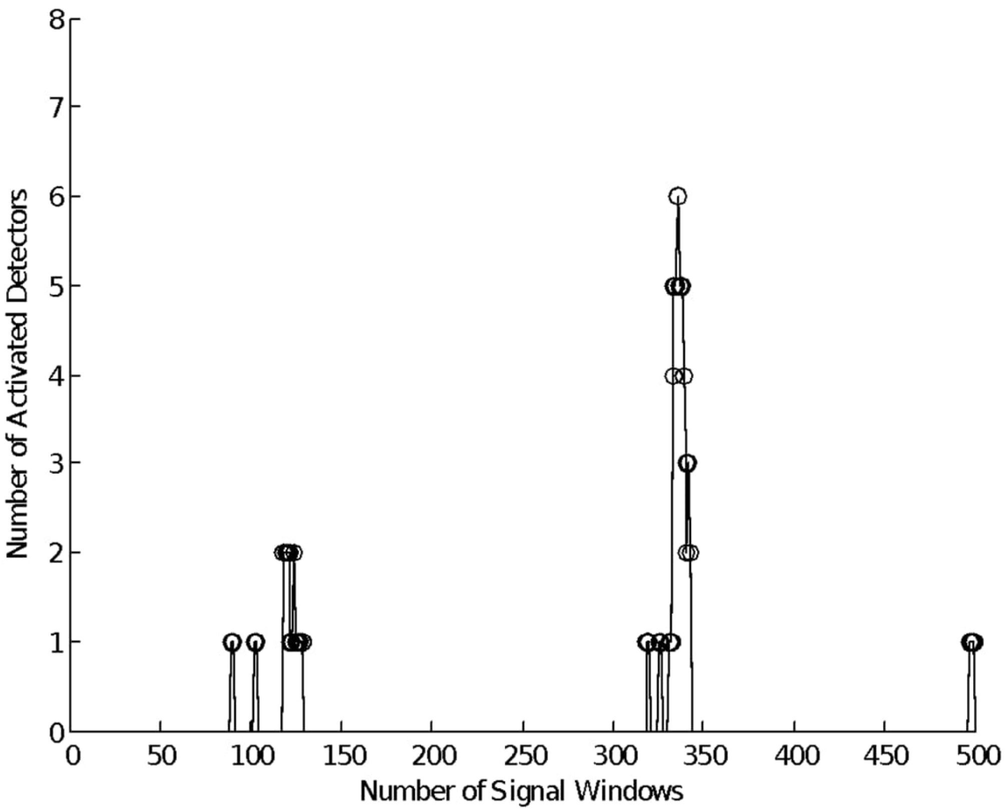
Healthy Motor

Motor Fault Detection Results



Broken Rotor

Motor Fault Detection Result

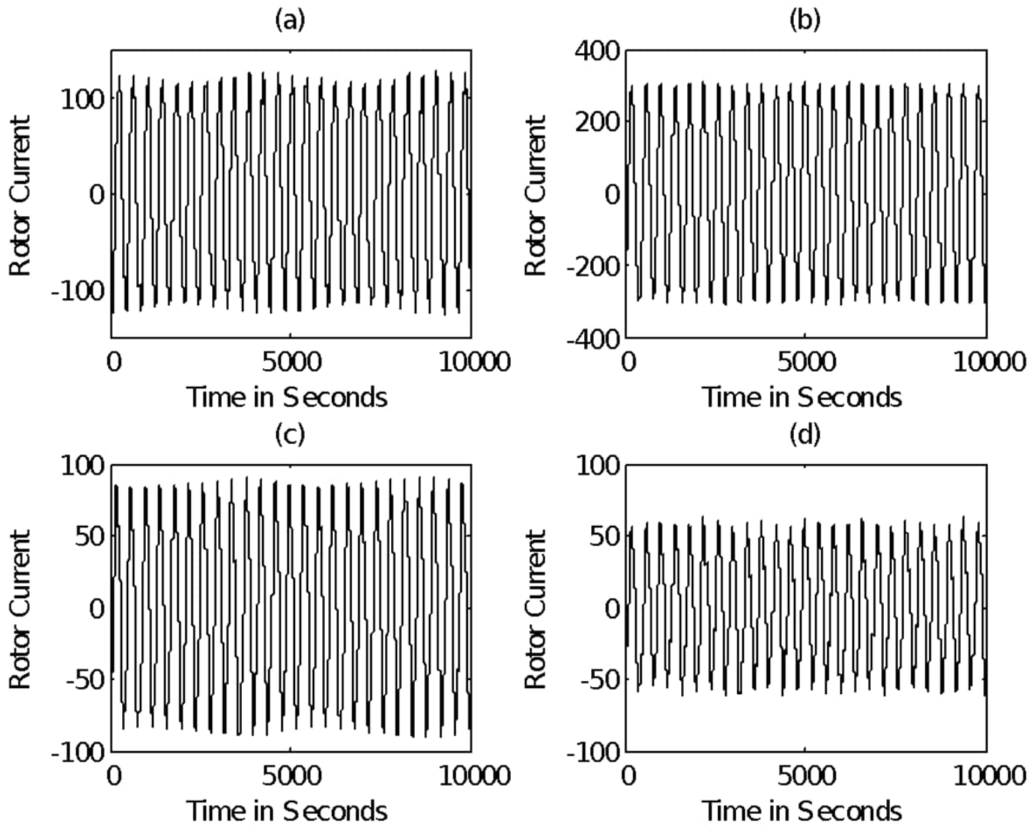


Broken Stator

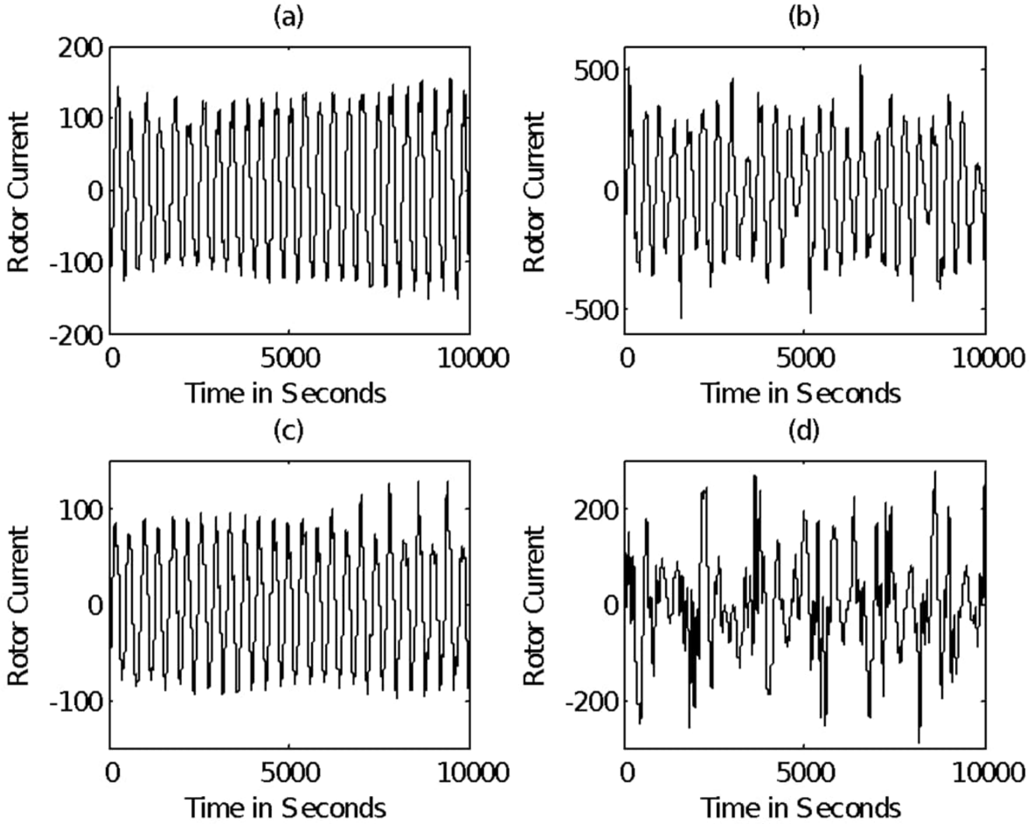
Detection Results of Rotor and Stator Faults

	A	B	η
Rotor Fault	11	84	88.42%
Stator Fault	11	76	87.36%

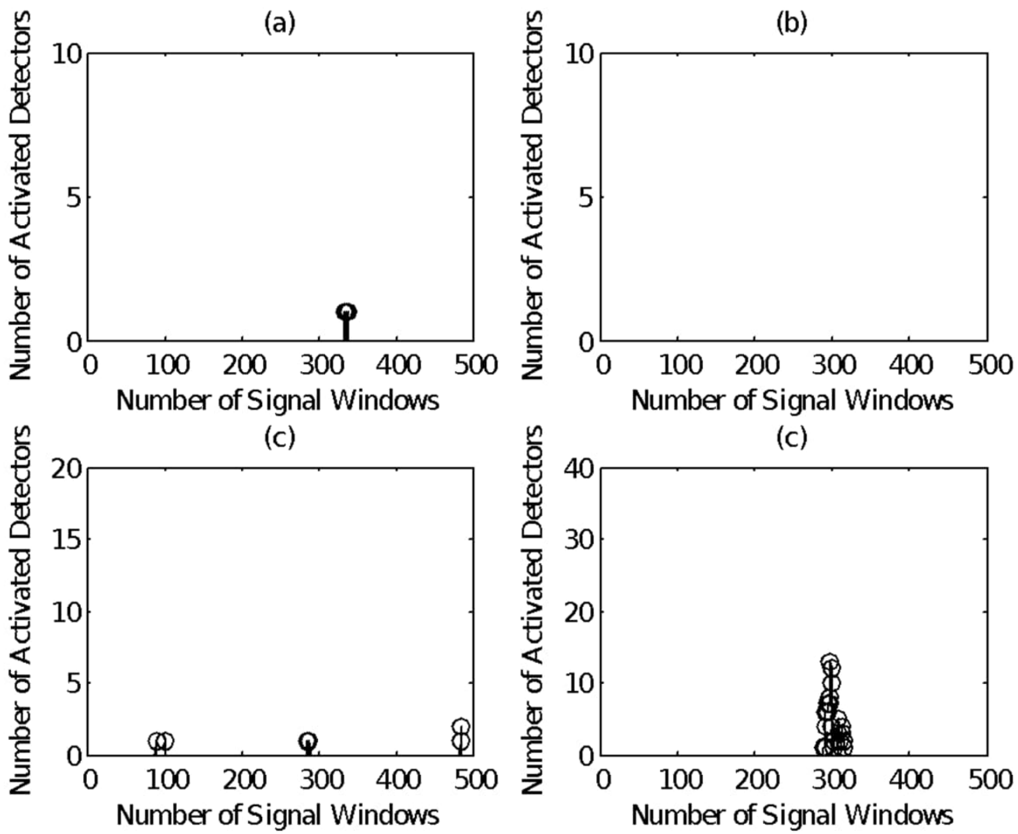
Current Signals of Healthy Motor with Four Different Loads



Current Signals of Faulty Motor with Four Different Loads (Gao, 2013)

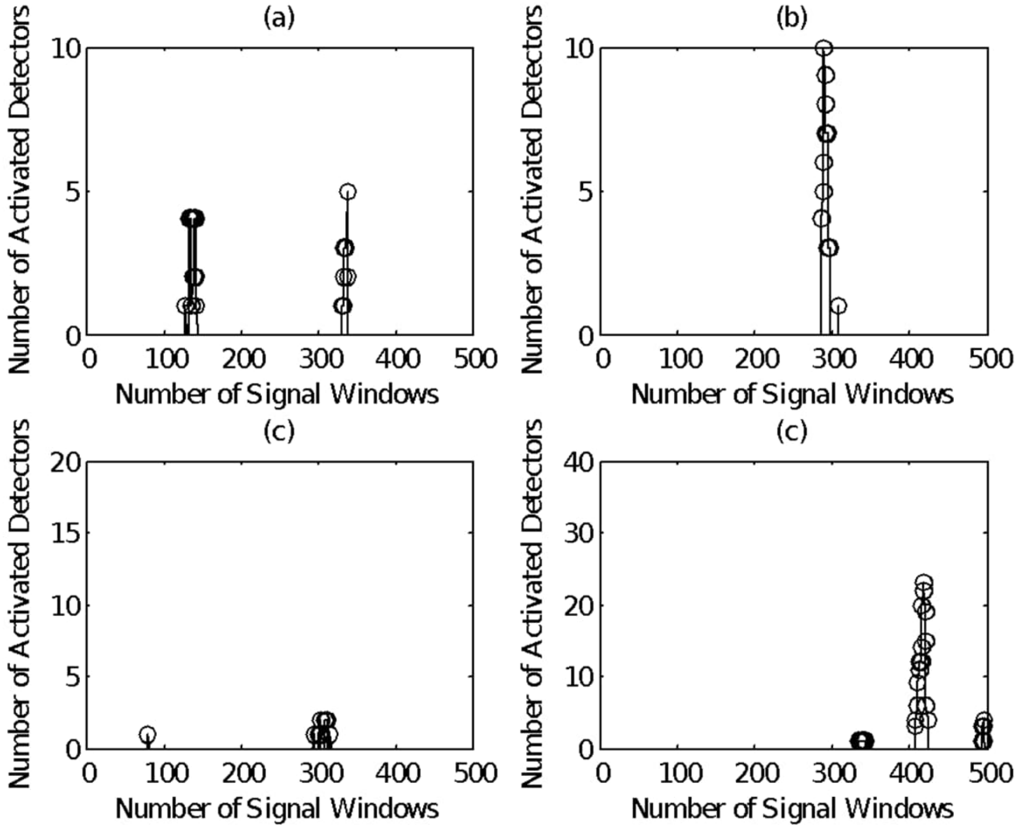


Numbers of Activated NSA Detectors for Healthy Motor



Healthy Motor

Numbers of Activated NSA Detectors for Faulty Motor



Faulty Motor

Fault Detection Rates of Faulty Motors with Different Loads

	A	B	η
Fault C	4	47	92.16%
Fault O	0	63	100%
Fault S	8	15	65.22%
Fault U	109	230	67.85%