Can a Fuzzy Rule Extraction Find an Extremely Tiny Non-self Region?

Akira Imada

Brest State Technical University
Belarus

□ **What is Fuzzy Rule Extraction?**

To extract knowledge (rules) from data given.

$$\Downarrow$$

Assume all the data are n-dimensional.

$$\Downarrow$$

Then rules are like

IF $(x_1$ is $A_1^k)$ and $\cdots$ and $(x_n$ is $A_n^k)$ THEN $(y_1$ is $\nu_{k1})$ and $\cdots$ and $(y_m$ is $\nu_{km})$

where y_i is a likliness of each of m classes.

A Zero-order Takagi-Sugeno's Fuzzy Model

$$\Downarrow$$

Membership functions

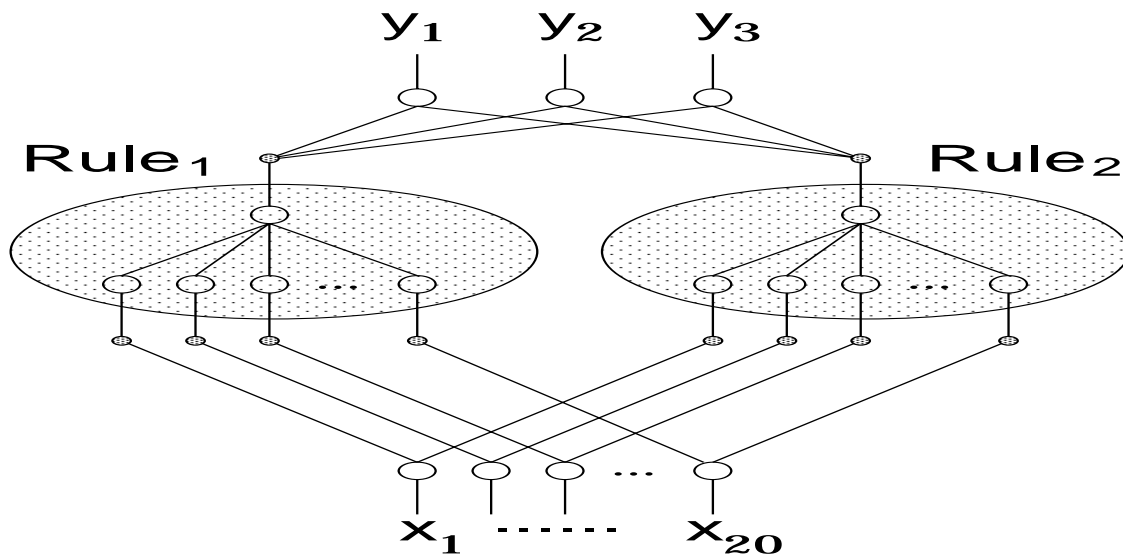$$\mu_{ik}(x_i) = \exp\{-(x_i - w_{ik})^2/\sigma_{ik}^2\}.$$

Defuzzification

$$y_j^0 = \sum_{k=1}^{H} \mu_k(\mathbf{x}^0) \cdot \nu_{kj} / \sum_{k=1}^{H} \mu_k(\mathbf{x}^0)$$

where

$$\mu_k(\mathbf{x}^0) = \prod_{i=1}^{n} \mu_{ik}(x_i^0)$$

□ A Fuzzy Neural Network Approach

— G. Castellano and A. Fanelli (2000)

Our Interest here

$\Downarrow$

a NETWORK INTRUSION DETECTION SYSTEM

in which

we need a set of TEST-DATA

to TRAIN and TEST the system with.

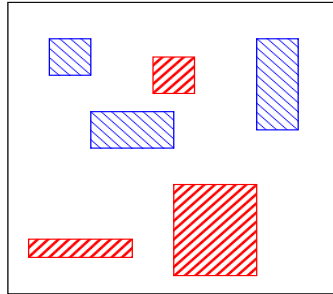$\Downarrow$

a consideration on such a test-data.

So many artificial data-samples have been proposed so far.

$\Downarrow$

Let's categorize them

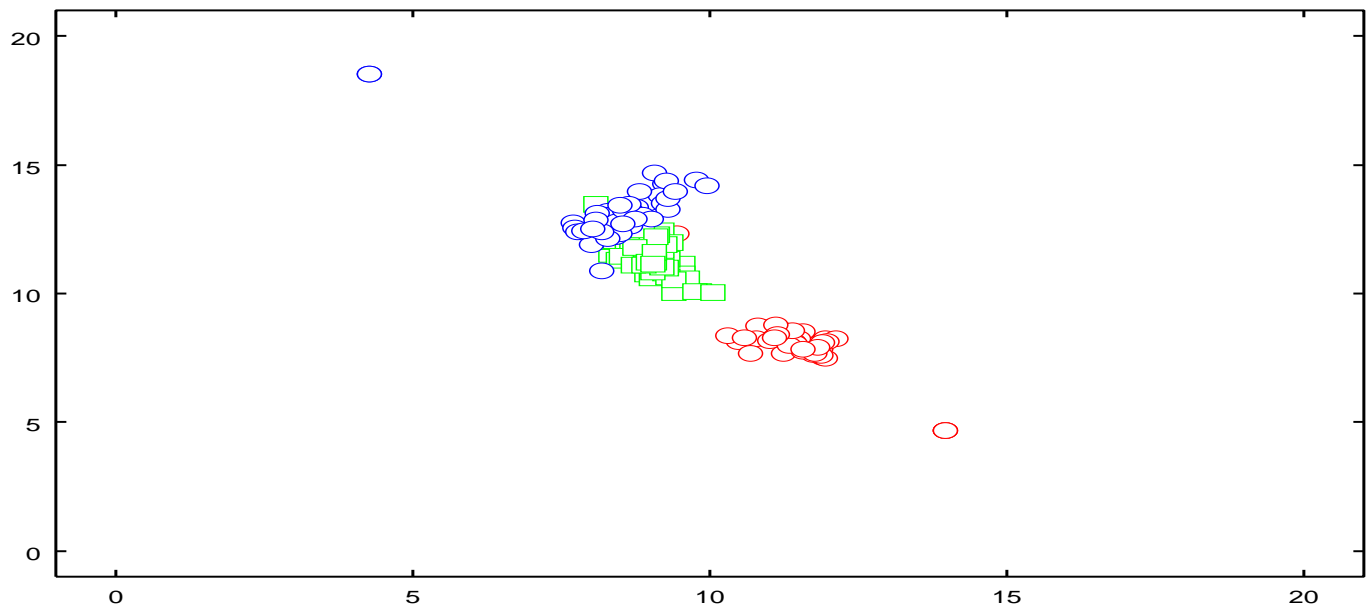# □ Fictitious 2-D pictures of test-sample — Type I

Do data cover

the whole universe?

E.g.,
Fisher's IRIS Flower
KDD-cup 1999/2003
· · ·

□ A visualization of IRIS data by Sammon Mapping

□ The data from KDD challange cup 98

4,920,210 data are given

$\Downarrow$

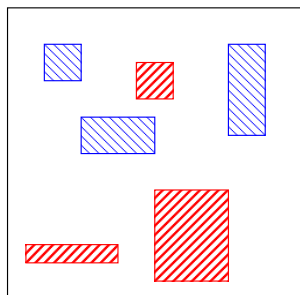each is made up of 42 attributes of which

$\Downarrow$

4-crisp + 17-binary + 6-integer + 15-real

Still infinitely large area of not-defiened possible transactions remain!

GUCCI vs. GUCCI-made-in-Hong-Kong
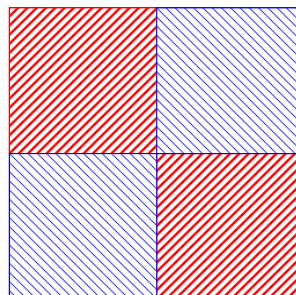**(Yet another from Istanbul, etc.)**

□ Fictitious 2-D pictures of test-sample — Type II
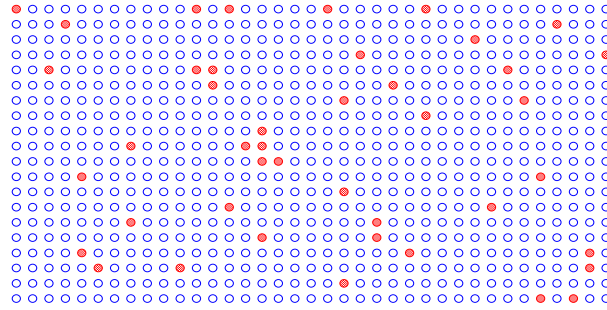
Do data cover

the whole universe?

Is a trainning with both

normal & abnormal meaningful?

Fisher's IRIS Flower

KDD-cup 1999/2003

· · ·

Ayara et al.

· · ·

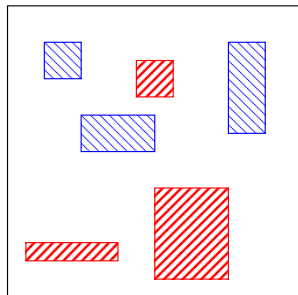## □ Ayara's Random Anomaly in 8-bit Binary Universe

152 and 160 abnormal patterns out of $2^8 = 256$ search points.
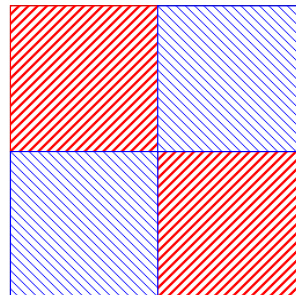
$$\Downarrow$$

Asserted that successfully trained.

□ Fictitious 2-D pictures of test-sample — Type II

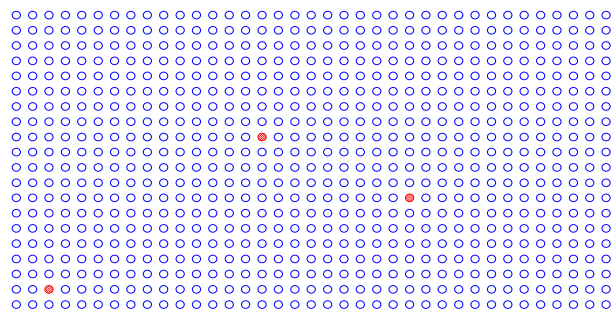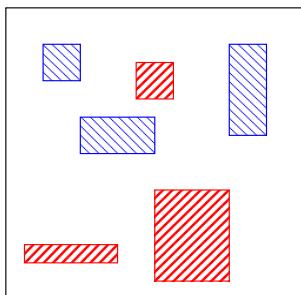| Do data cover the whole universe?	Is a trainning with both normal & abnormal meaningful?
Fisher's IRIS Flower KDD-cup 1999/2003 $\cdots$	Ayara et al. $\cdots$

□ **What if abnormal sample are only a few?**

Still can we train the system with normal and abnormal sample?

□ Fictitious 2-D pictures of test-sample — Type III
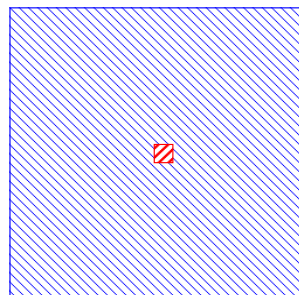
Do data cover
the whole universe?
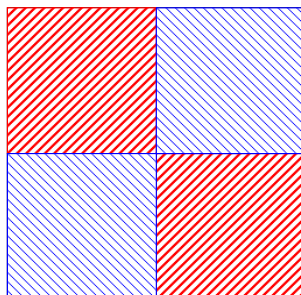
Is a trainning with both
normal & abnormal meaningful?

What if the size of known abnormal
sample is extremely tiny?

Fisher's IRIS Flower
KDD-cup 1999/2003
 · · ·

Ayara et al.
· · ·

None so far

□ **Three of our claims.**

1. Data should cover the whole universe.

    $\Rightarrow$ We could miss crucial abnormal in no-defined area.

2. Abnormal Sample should be assumed extremely tiny.

    $\Rightarrow$ This is of usual case.

3. Can we train the sytem only by NORMAL data?

**A sommelier who is trained only by real champagne
can tell the difference when given a bootleg or other sparkling wine?**

How about Caviar?

Forgery coins recognition?

Our Goal is

to search for only a few ABNORMAL (no-self) pattern

hidden in

an enormous amount of NORMAL (self) patterns
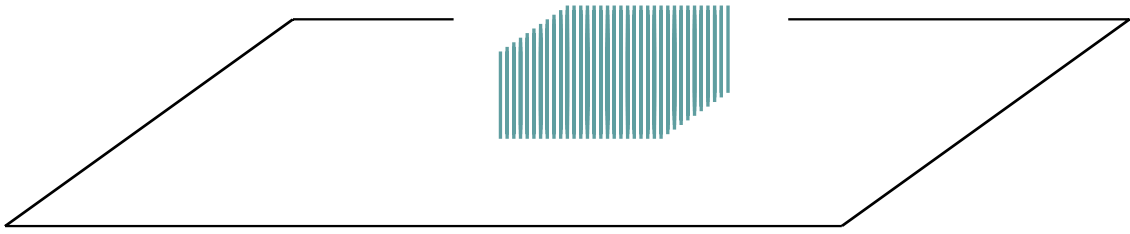
by

training using only NORMAL patterns

□ **A test-sample – A tiny-flat-island-in-a-huge-lake**
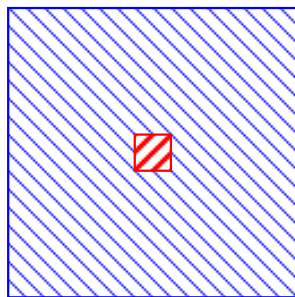
- Lake $\Rightarrow$ n-D Hypercube where $x_i \in [-1, 1]$ $(i = 1, ..., n)$

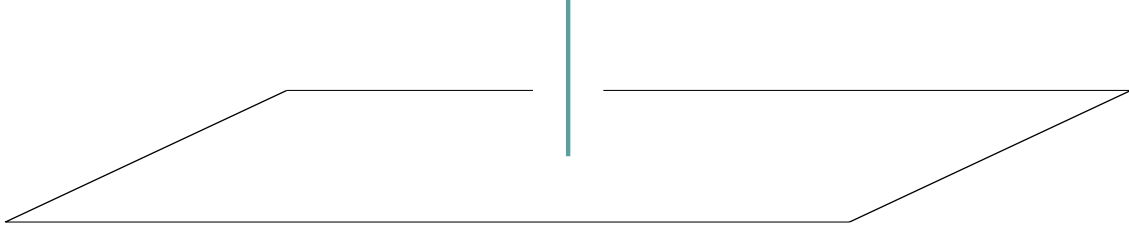- Island $\Rightarrow$ $x_i \in [-a, a]$ $(a < 1)$.

— We can control the complexity by changing the size.

□ From a fitness landscape point of view
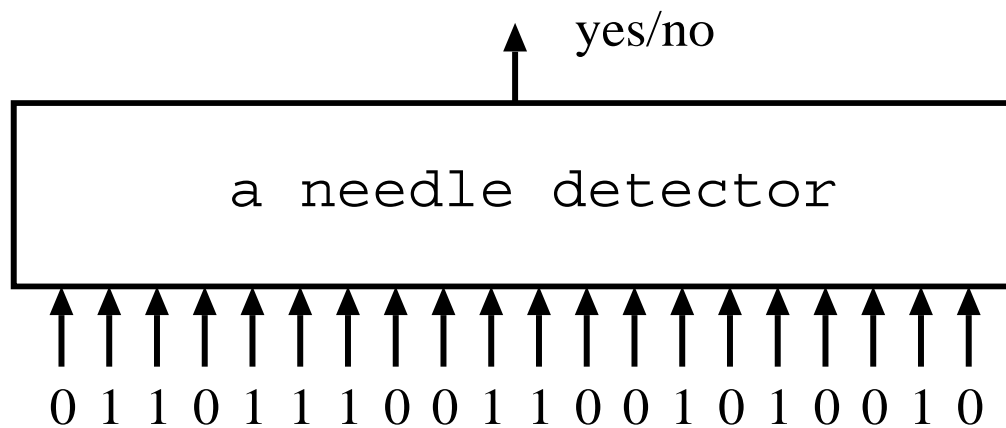
□ A Needle in a Haystack

A schematic skecth on fictitious 2D-space

The original Hinton & Nowlan's Needle:

- A-needle $\Rightarrow$ Only one configuration of 20 bits of binary string.

    $\star$ We don't know where the needle locates, but God knows.

- Haystack $\Rightarrow 2^{20} - 1$ search points

□ **How can we train the detector?**

yes/no

```
a needle detector
```
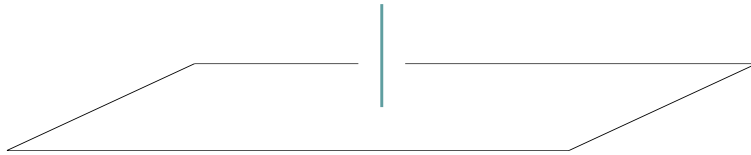
0 1 1 0 1 1 1 0 0 1 1 0 0 1 0 1 0 0 1 0

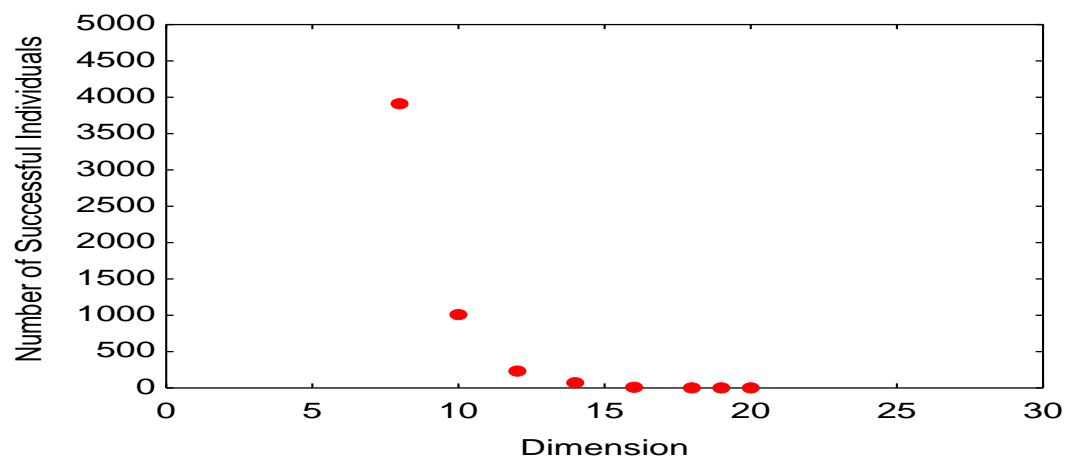Can we train it with most likely haystack points?

□ **How difficult?**

- Random Search

$$2^{20} = 1,048,576$$

- Lifetime Learning – Baldwin Effect (Hinton & Nowlan 1993)

□ **Random Search**

$$\Downarrow$$

Placebo $\Rightarrow$ Criteria of Comparison

□ **We have attacked this problem with lately reported approaches**

- Artificial Immune System

- Evolutionary Computation

- Fuzzy Rule

- Data-mining Technique

- etc

each of which claims very SUCCESSFUL.

□ **When a species of IRIS flower is normal then are others abnormal?**

E.g., Kim & Bentley (2001) claimed

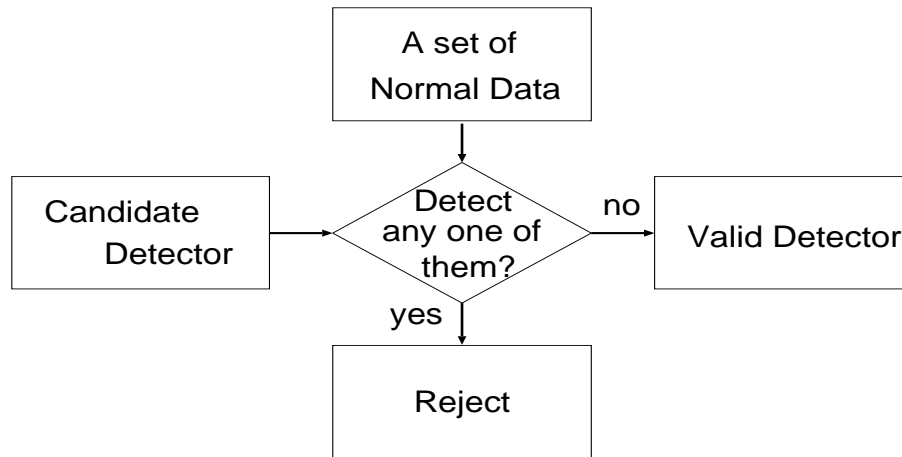**assuming one family of IRIS is abnormal while other two normal**

$$\Downarrow$$

TP (Successful Detection Rate) reached 100%
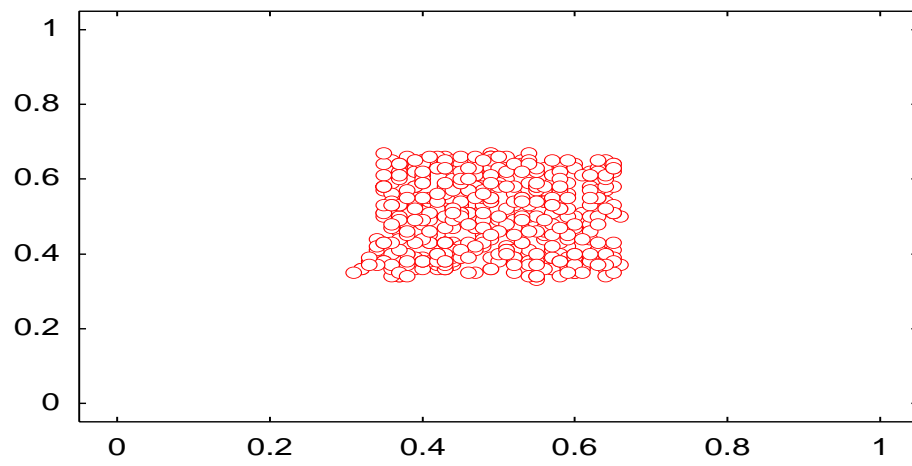**FP (False Alarm Rate) was only 1%.**

□ A snapshot of our ongoing works

An immune approach — Constant-sized hyper-shpere Detectors

```
                    ┌─────────────┐
                    │   A set of  │
                    │ Normal Data │
                    └─────────────┘
                            │
                            ▼
┌─────────────┐      ╱  Detect  ╲      no   ┌────────────────┐
│  Candidate  │ ──▶ ⟨ any one of  ⟩ ──────▶ │ Valid Detector │
│  Detector   │      ╲   them?   ╱           └────────────────┘
└─────────────┘            │
                     yes   │
                           ▼
                    ┌─────────────┐
                    │   Reject    │
                    └─────────────┘
```
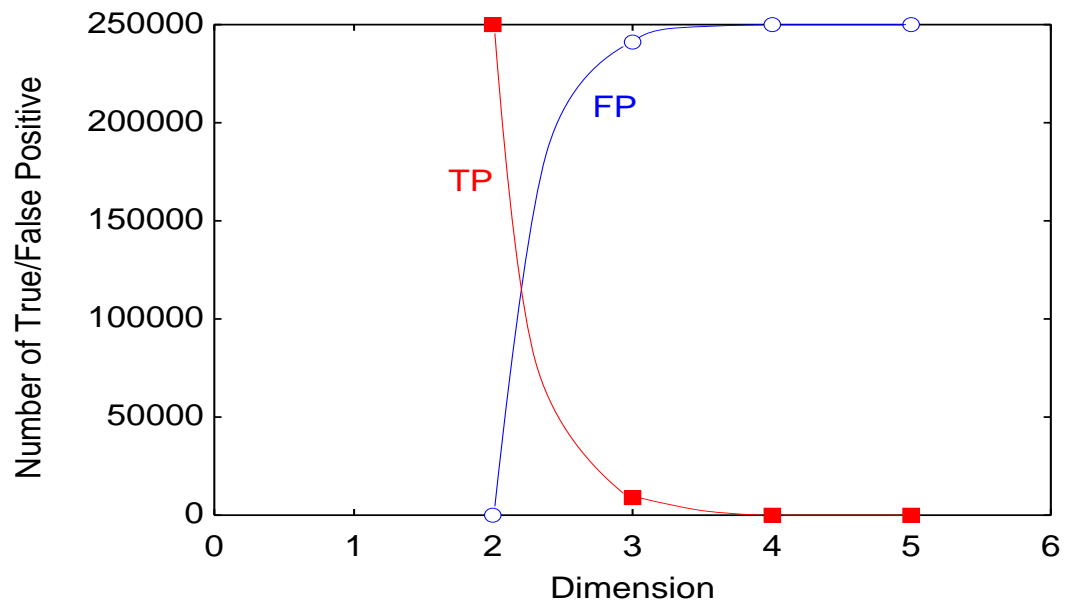
□ A result in 2-D space
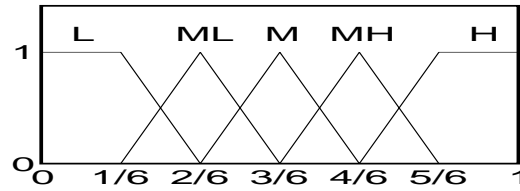
□ What if top area shrinks to zero?
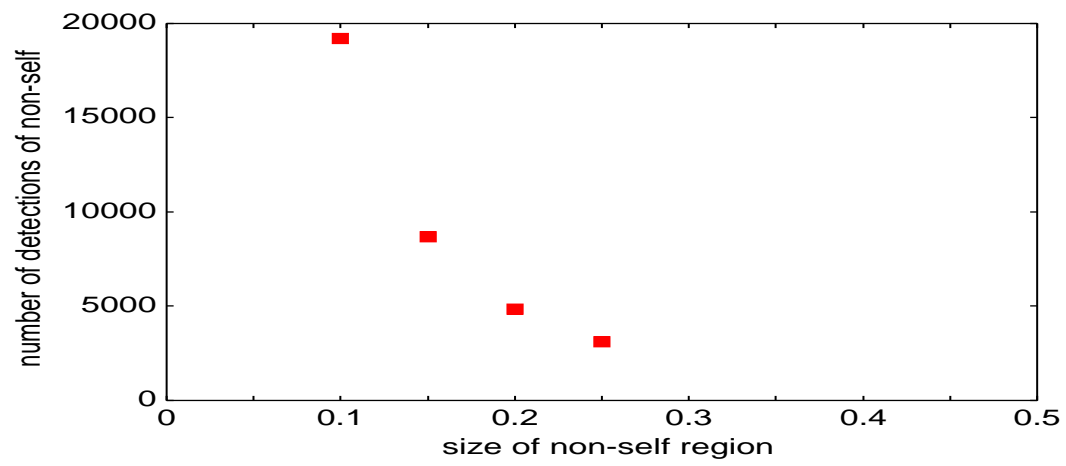
□ Alas! As dimension grows...

□ Yet another snapshot

A Fuzzy Rule approach — Can a fuzzy rule find an island?

$$\Downarrow$$
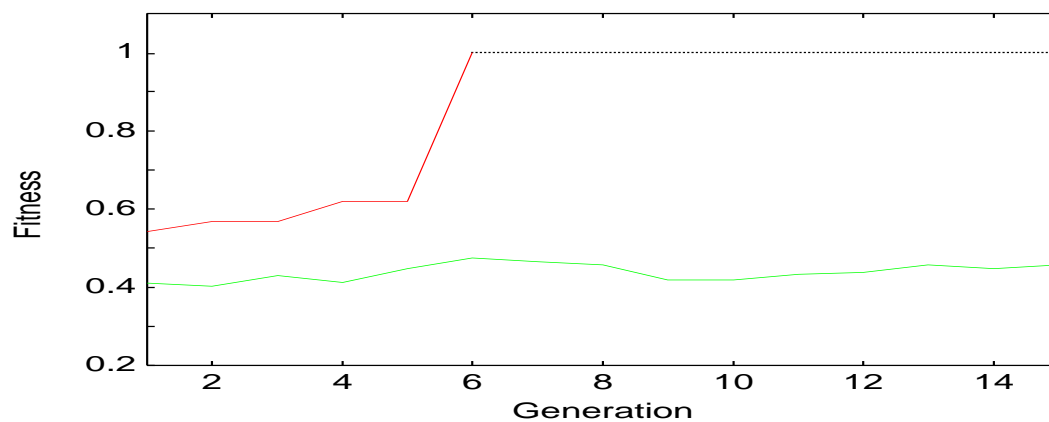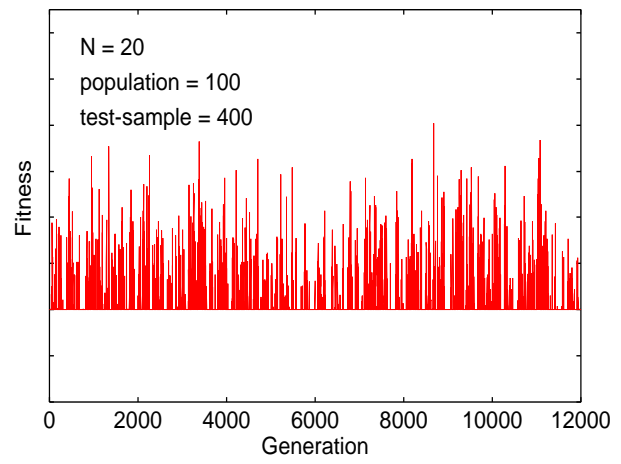
How many rules we need?

$$(MMMMMM\cdots M)$$

$$\Downarrow$$
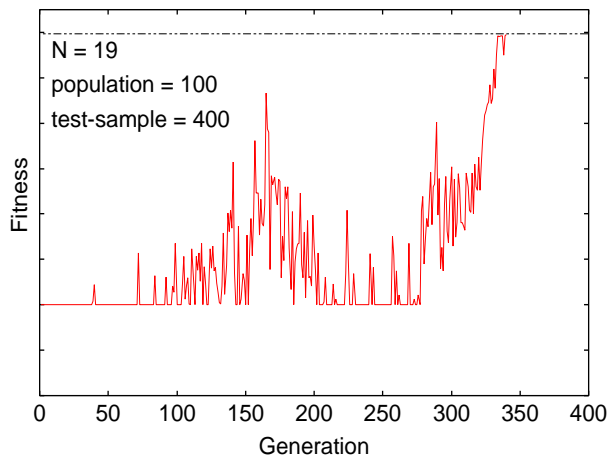
IF $\{x_1$ is Middle$\}$, $\cdots$, and $\{x_{20}$ is Middle$\}$ THEN no-self.
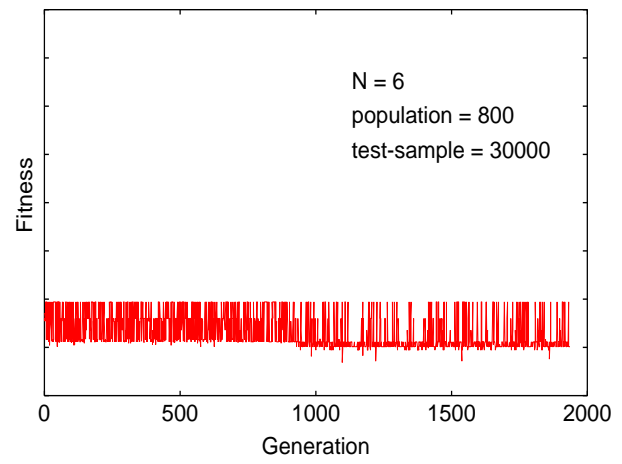
□ Island in the 6-D lake

Fairly large island ($x_i \in [0.25, 0.75]$)   vs.   Small island ($x_i \in [0.45, 0.55]$)
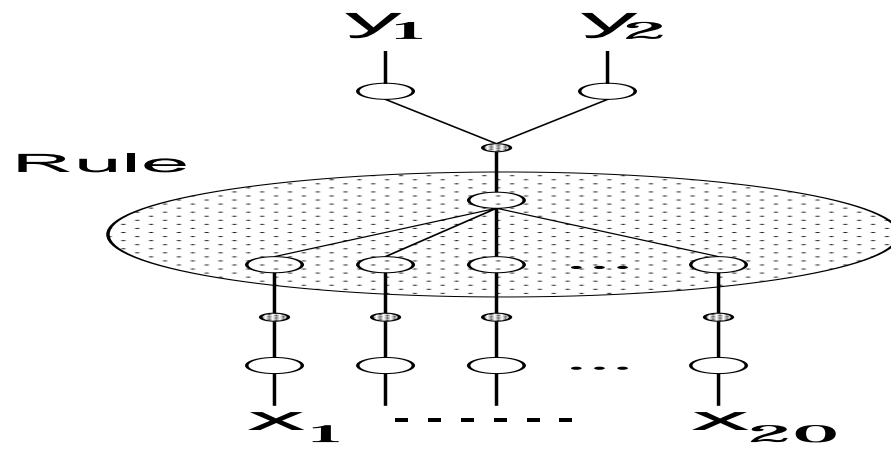
$\Downarrow$

□ Training with Non-self

N = 19
population = 100
test-sample = 400

Fitness

Generation

N = 20
population = 100
test-sample = 400

Fitness

Generation

□ Training with Self
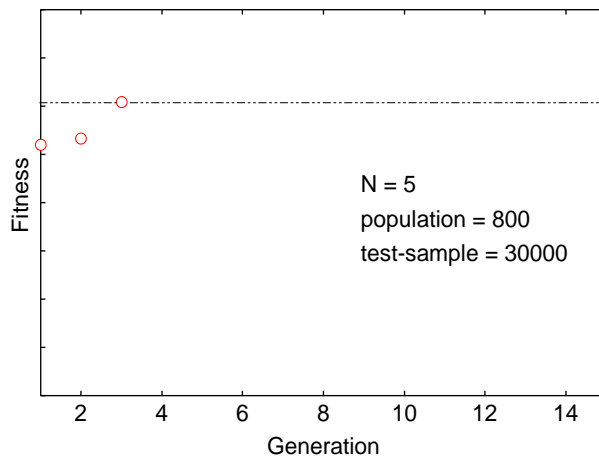
□ **To detct island**

A result of an evolution

$$M$$

$$\Downarrow$$
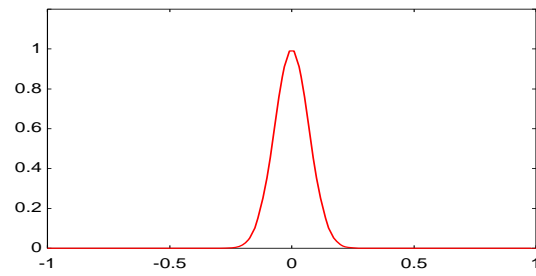
While usually we have huge NORMAL samples,
we don't know many ABNORMAL samples
(when we know them it's too late)

$\Downarrow$

Can training be performed only by using NORMAL samples?

$\Downarrow$

a-needle or tiny-island
as test-data to design a network intrusion detector.

□ Conclusion

Results have not been wonderful at all **AS THEY CLAIMED.**

$$\Downarrow$$

Worse than Placebo experiment?

Or, experiments have sometimes reversed our expectations.

We now are negative more or less, but still want to be neutral.