

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.

↓

Assume all the data are n -dimensional.

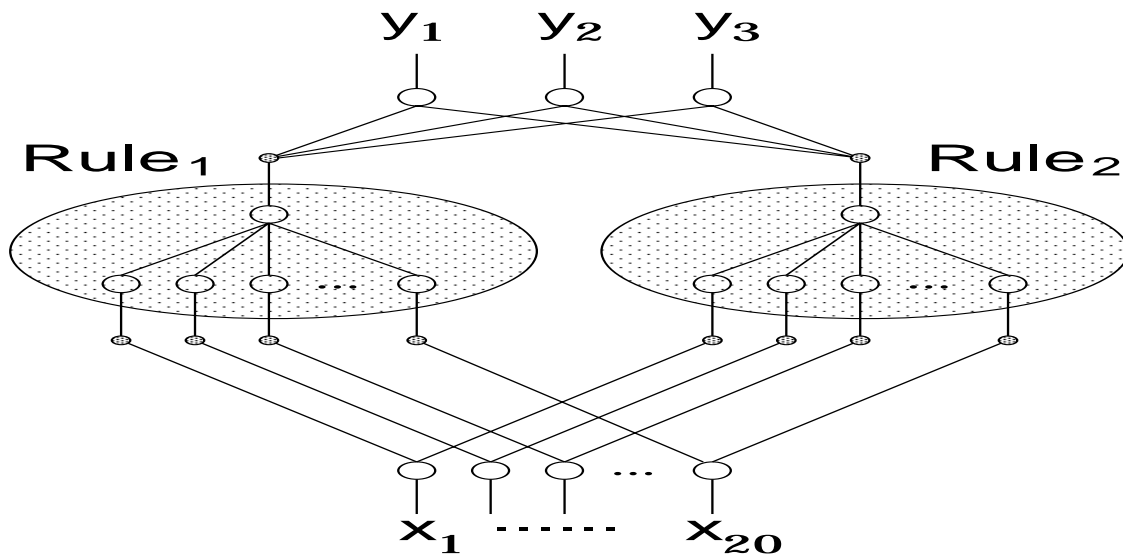
↓

Then rules are like

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

where y_i is a likliness of each of m classes.

□ A Fuzzy Neural Network Approach



Our Interest here



a NETWORK INTRUSION DETECTION SYSTEM

in which

we need a set of TEST-DATA

to TRAIN and TEST the system with.



a consideration on such a test-data.

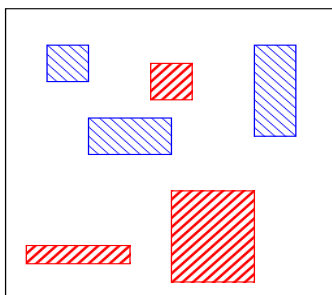
So many artificial data-samples have been proposed so far.



Let's categorize them

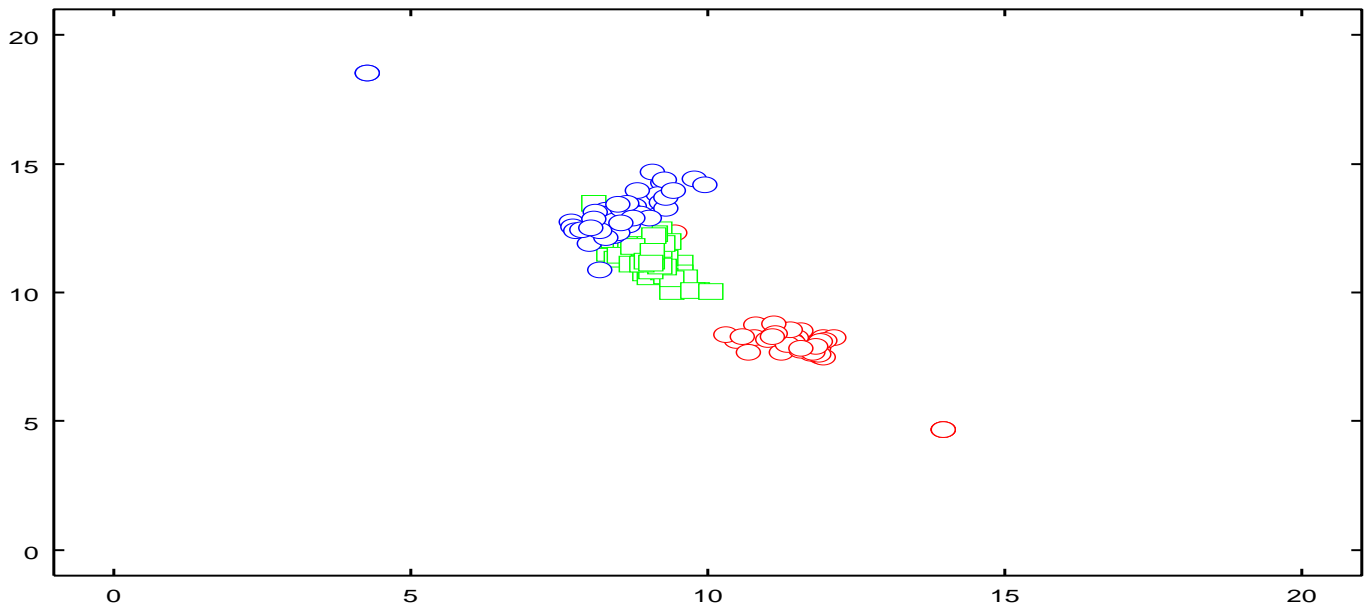
□ **Fictitious 2-D pictures of test-sample — Type I**

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 challenge cup 98**

4,920,210 data are given

↓

each is made up of 42 attributes of which

↓

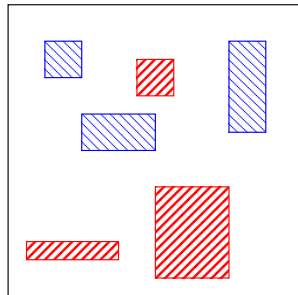
4-crisp + 17-binary + 6-integer + 15-real

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

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

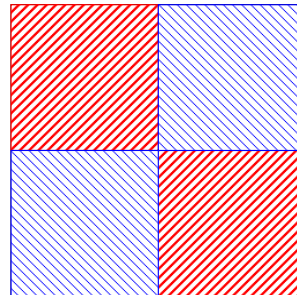
□ Fictitious 2-D pictures of test-sample — Type II

Do data cover
the whole universe?



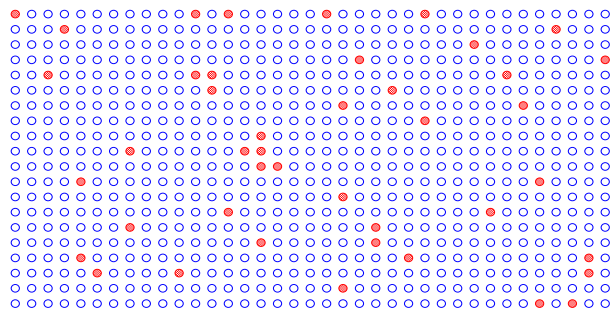
Fisher's IRIS Flower
KDD-cup 1999/2003
...

Is a training with both
normal & abnormal meaningful?



Ayara et al.
...

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



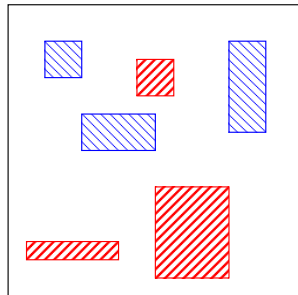
152 and 160 abnormal patterns out of $2^8 = 256$ search points.



Asserted that successfully trained.

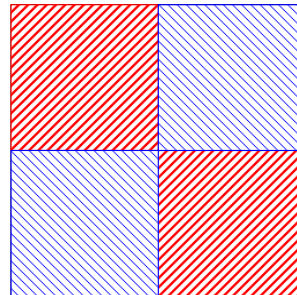
□ Fictitious 2-D pictures of test-sample — Type II

Do data cover
the whole universe?



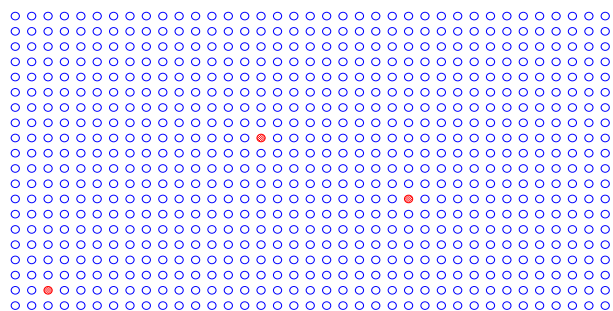
Fisher's IRIS Flower
KDD-cup 1999/2003
...

Is a training with both
normal & abnormal meaningful?



Ayara et al.
...

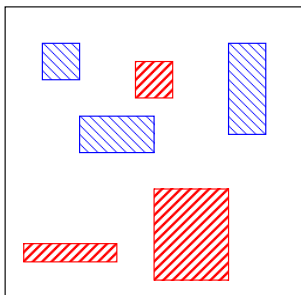
□ **What if abnormal sample are only a few?**



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

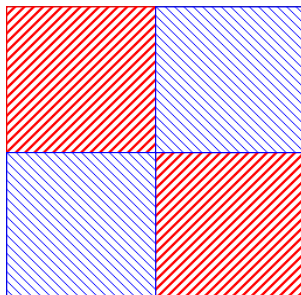
□ Fictitious 2-D pictures of test-sample — Type III

Do data cover
the whole universe?



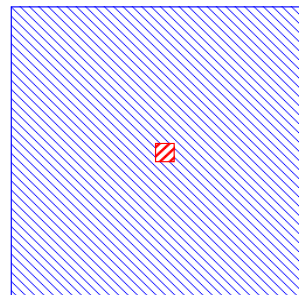
Fisher's IRIS Flower
KDD-cup 1999/2003
...

Is a training with both
normal & abnormal meaningful?



Ayara et al.
...

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



None so far

□ **Three of our claims.**

1. Data should cover the whole universe.

⇒ We could miss crucial abnormal in no-defined area.

2. Abnormal Sample should be assumed extremely tiny.

⇒ This is of usual case.

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

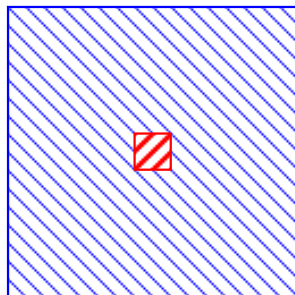
**A sommelier who is trained only by real champagne
can tell the difference when given a bootleg or other sparkling wine?**

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

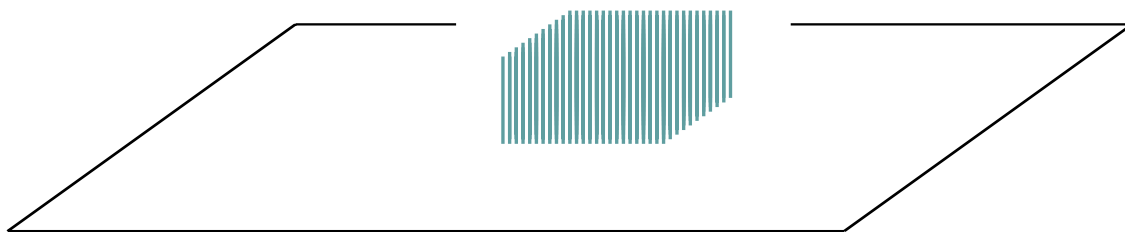
□ **A test-sample – A tiny-flat-island-in-a-huge-lake**



- Lake $\Rightarrow n$ -D Hypercube where $x_i \in [-1, 1]$ ($i = 1, \dots, n$)
- Island $\Rightarrow x_i \in [-a, a]$ ($a < 1$).

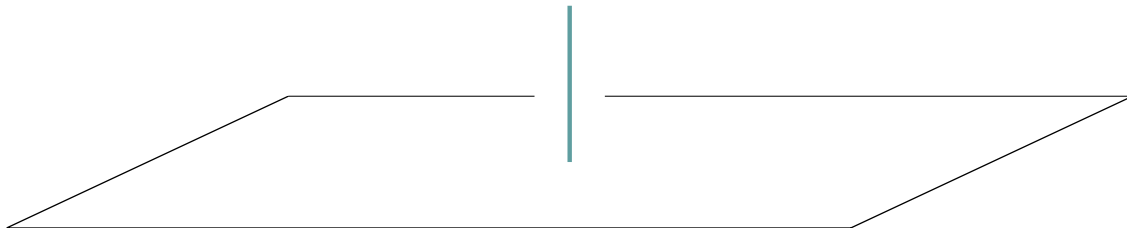
— We can control the complexity by changing the size.

□ **From a fitness landscape point of view**



□ A Needle in a Haystack

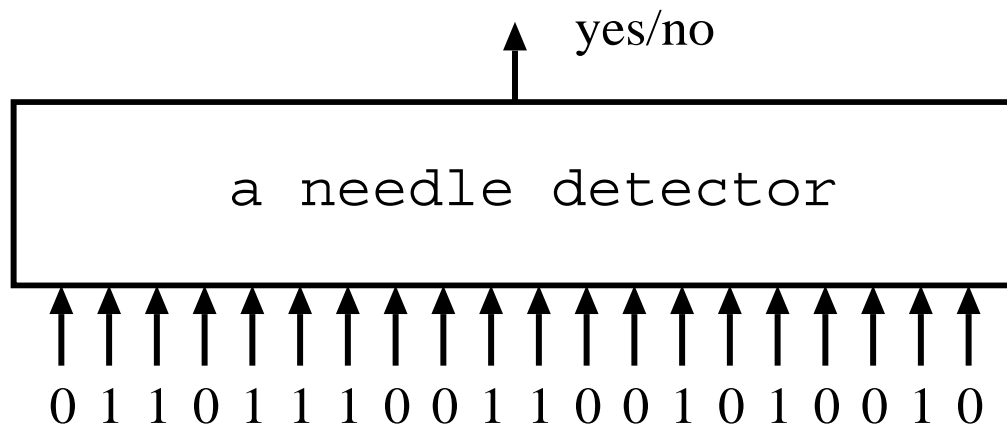
A schematic skeeth on fictitious 2D-space



The original Hinton & Nowlan's Needle:

- A-needle \Rightarrow Only one configuration of 20 bits of binary string.
 - ★ We don't know where the needle locates, but God knows.
- Haystack $\Rightarrow 2^{20} - 1$ search points

□ **How can we train the detector?**



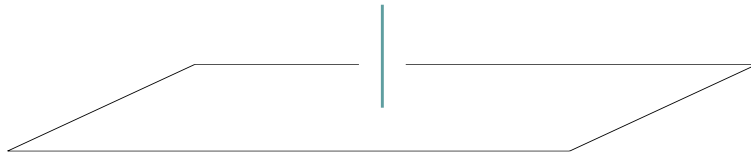
Can we train it with most likely haystack points?

□ **How difficult?**

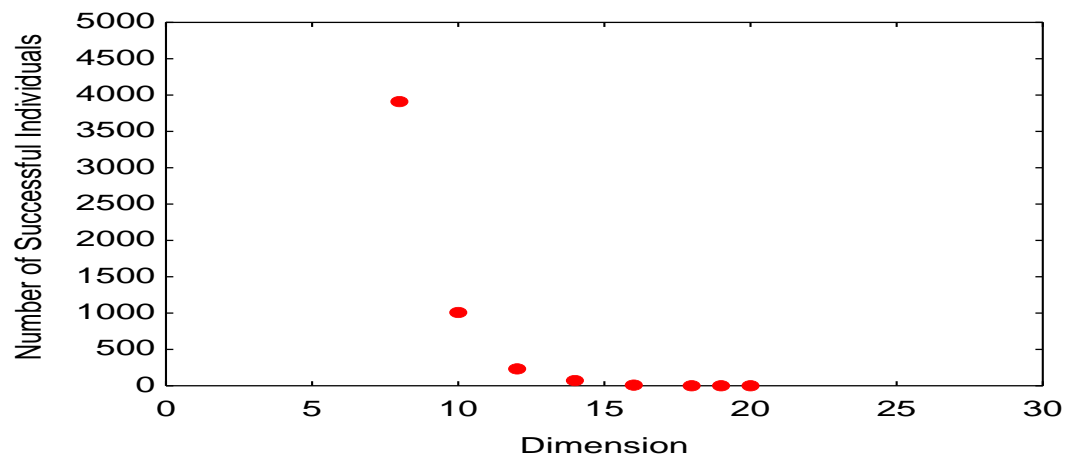
- Random Search

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

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



□ Random Search



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

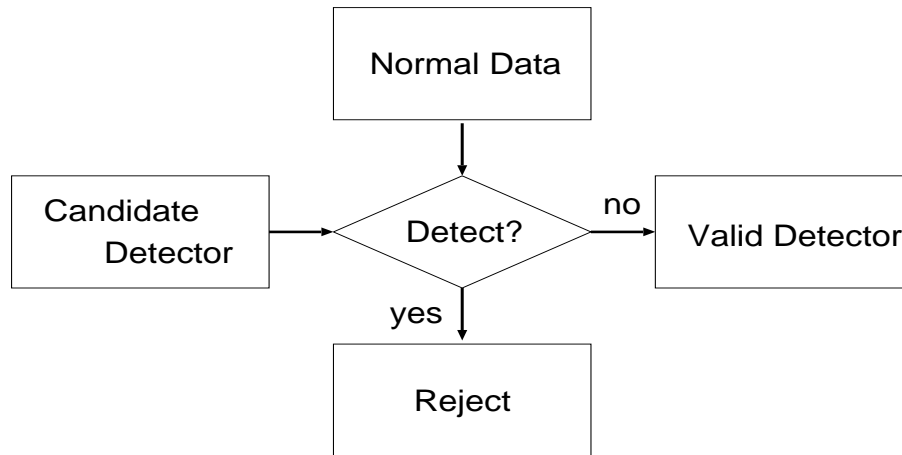
↓

TP (Successful Detection Rate) reached 100%

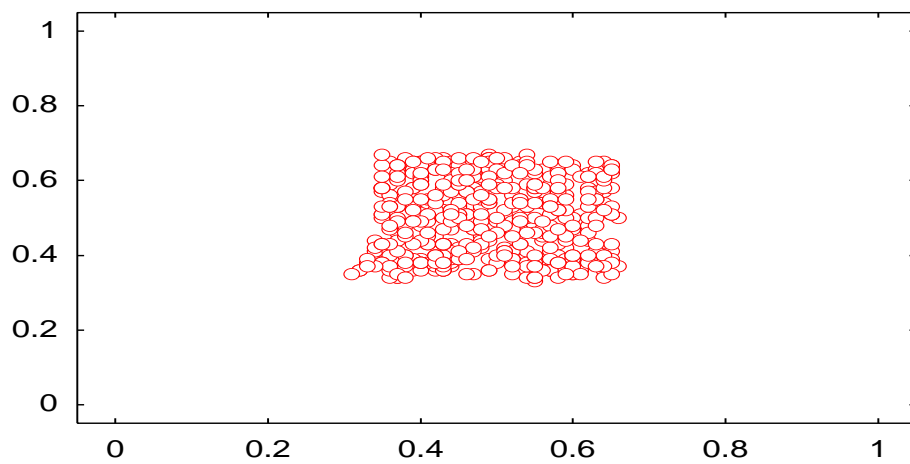
FP (False Alarm Rate) was only 1%.

□ A snapshot of our ongoing works

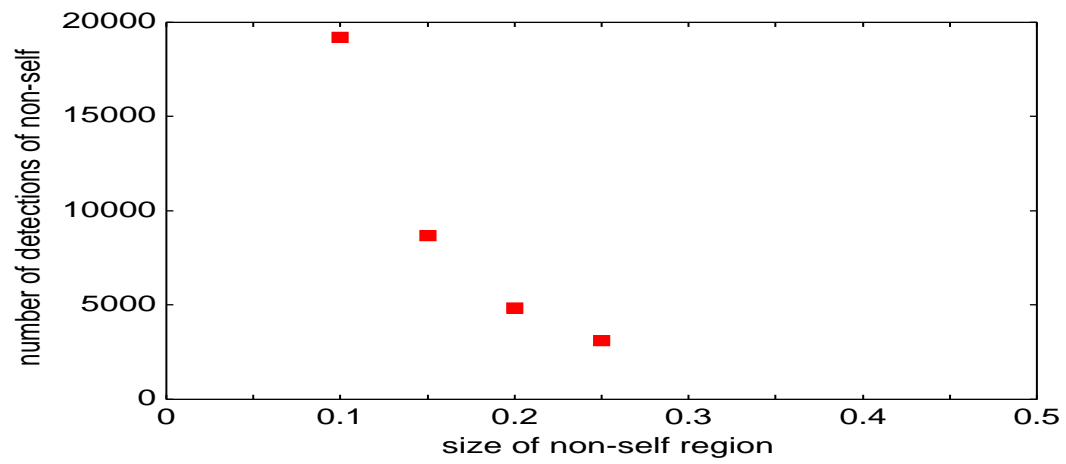
An immune approach — Constant-sized hyper-sphere Detectors



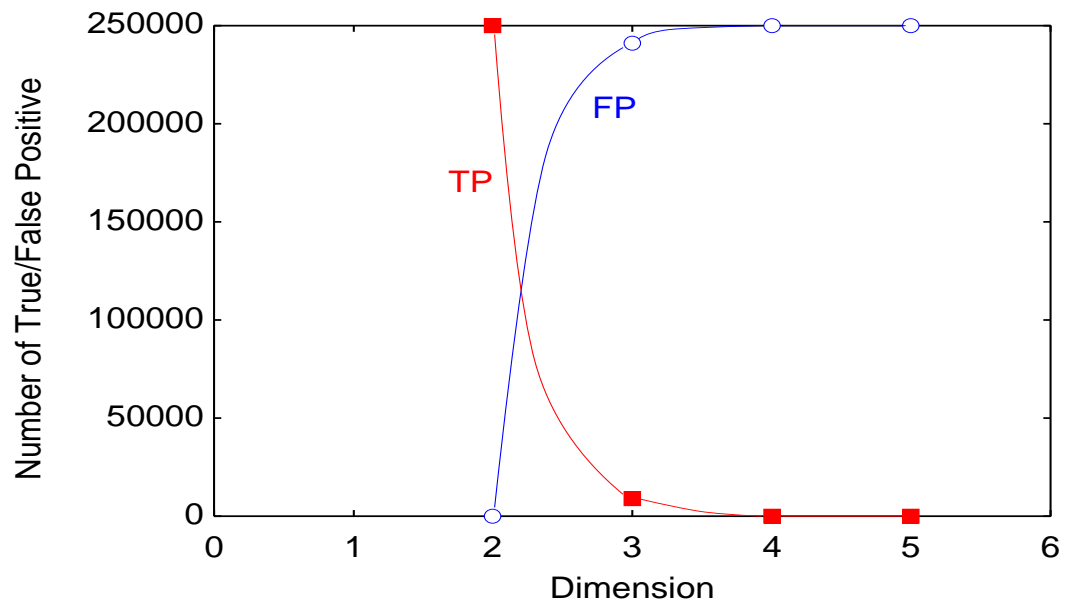
□ A result in 2-D space



□ What if top area shrinks to zero?

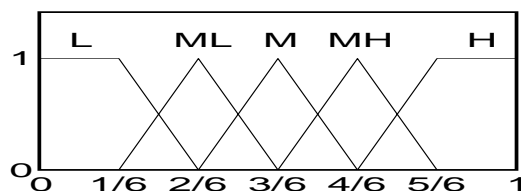


□ Alas! As dimension grows...



□ Yet another snapshot

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



How many rules we need?

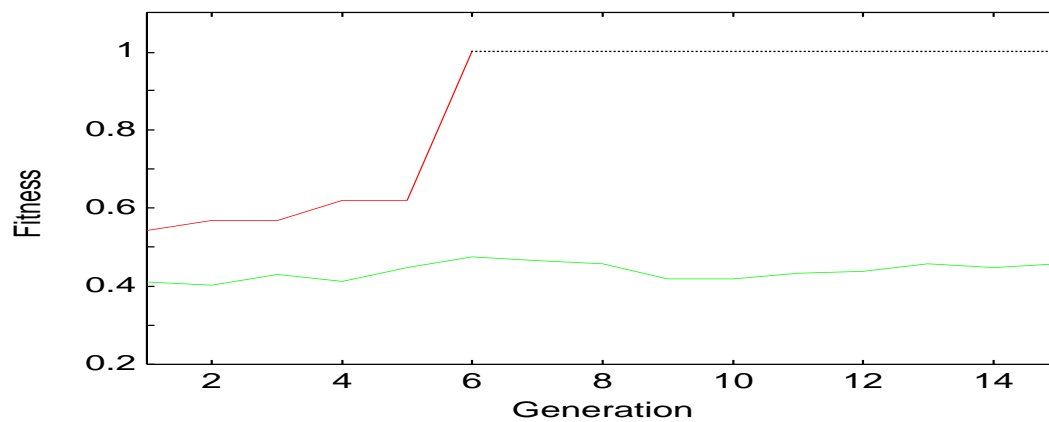
$(MMMMMM \cdots M)$



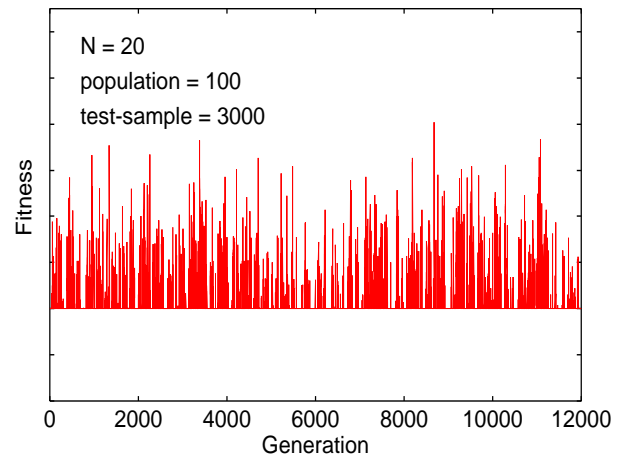
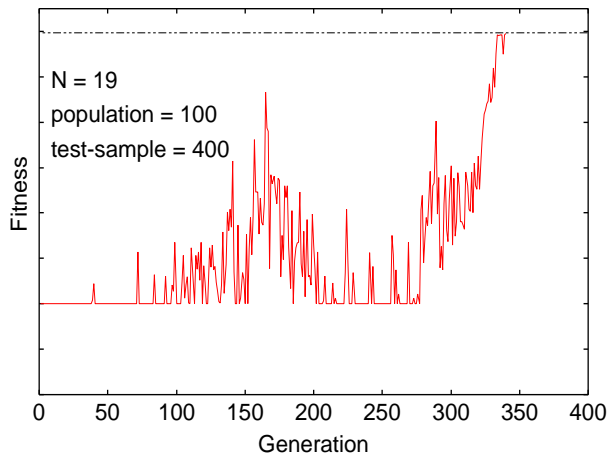
IF $\{x_1 \text{ is Middle}\}, \cdots$, and $\{x_{20} \text{ is Middle}\}$ THEN no-self.

□ Island in the 6-D lake

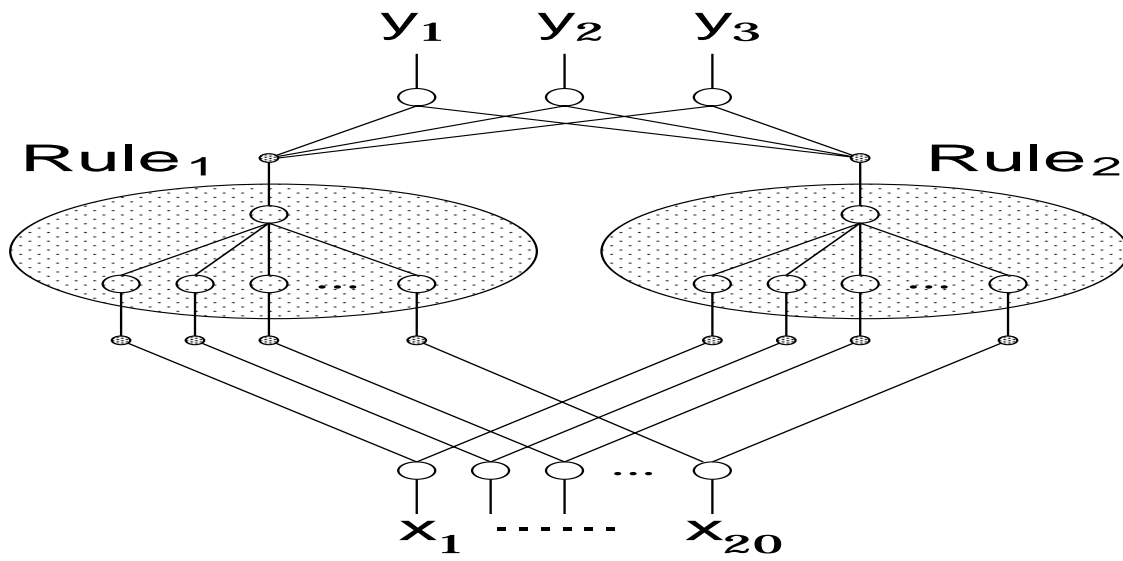
Fairly large island ($x_i \in [0.25, 0.75]$) vs. Small island ($x_i \in [0.45, 0.55]$)



□ Training with Non-self

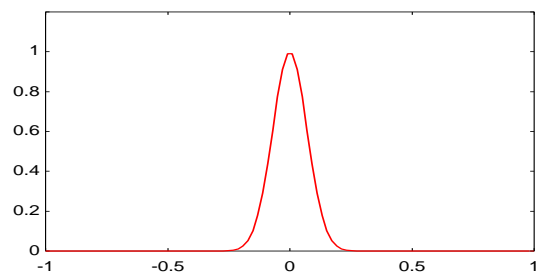


□ To detect island



A result of an evolution

M



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



Can training be performed only by using NORMAL samples?



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.



Worse than Placebo experiment?

Or, experiments have sometimes reversed our expectations.

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