Can a negative selection detect an extremely few non-self among enormous amount of self cells?

Akira Imada

Brest State Technical University
Belarus

Our Interest

$\Downarrow$

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.

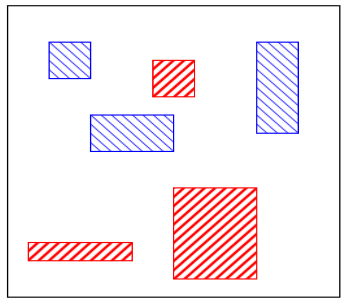
$\Downarrow$

Let's categorize them

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

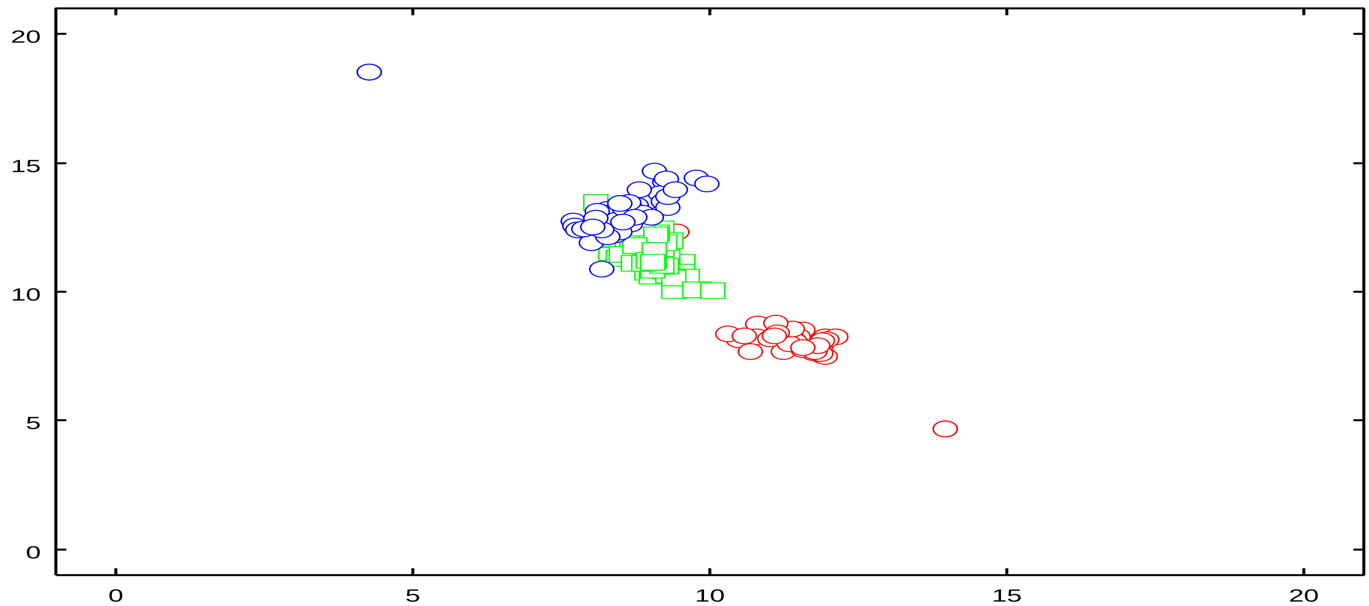
Do data cover

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

⇓

each is made up of 42 attributes of which

⇓

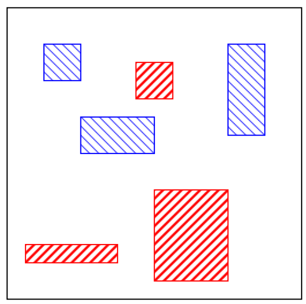
4-crisp + 17-binary + 6-integer + 15-real

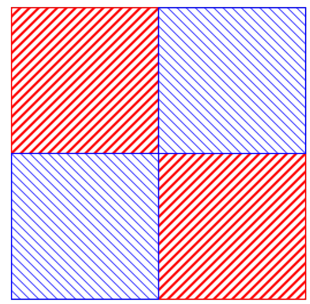
Still infinitely large not-defiened possible transactions remain!

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

Do data cover the whole universe?

Is a trainning with both normal & abnormal meaningful?

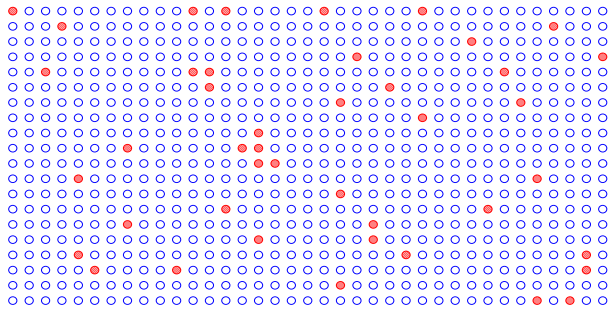




Fisher's IRIS Flower
KDD-cup 1999/2003
...

Ayara et al.
...

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

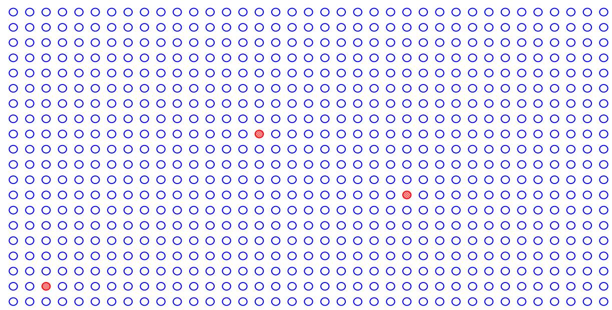


152 and 160 abnormal patterns out of $2^8 = 256$ search points.

$$\Downarrow$$

Asserted that successfully trained.

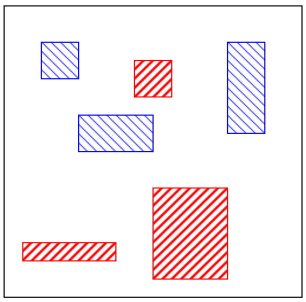
□ What if abnormal sample are only a few?



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

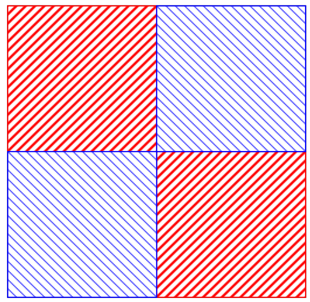
□ Fictitious 2-D pictures of test-sample — Type III

Do data cover the whole universe?



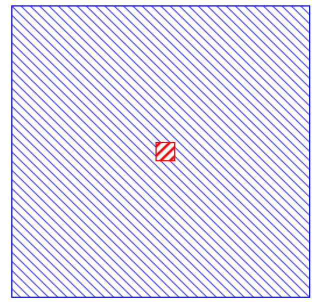
Fisher's IRIS Flower
KDD-cup 1999/2003
...

Is a trainning 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 forgery or other sparkling wine?

Our Goal is

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

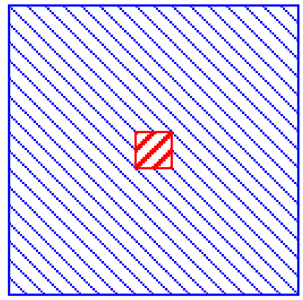
hidden in

an enormous amount of NORMAL (self) patterns

by

training using only NORMAL patterns

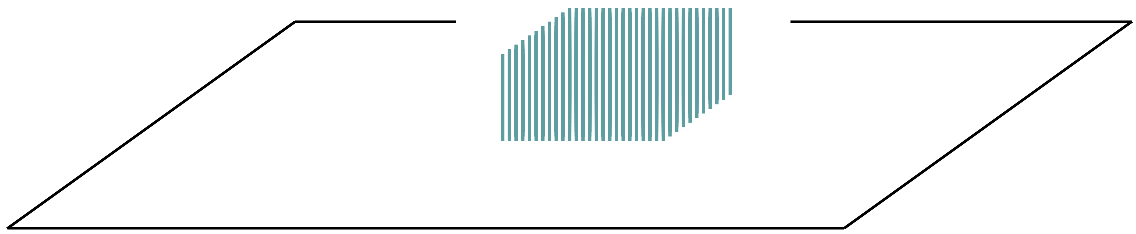
□ A test-sample – A tiny-flat-island-in-a-huge-lake



- Lake ⇒ $x_i \in [-1, 1] \quad (i = 1, \cdots, n)$
- Island ⇒ $x_i \in [-a, a] \quad (a < 1)$.

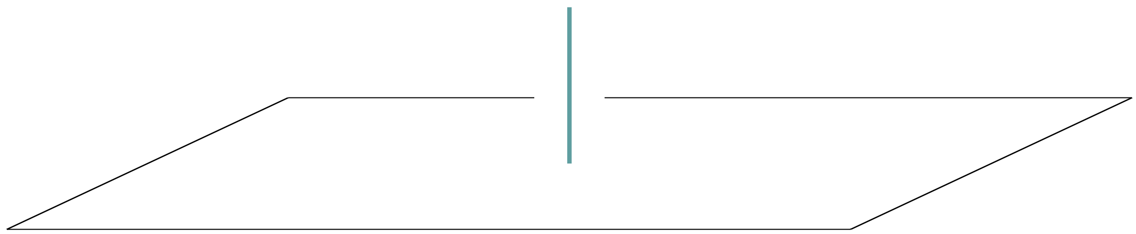
— We can control the complexity by changing the size.

□ From a fitness landscape point of view



□ A Needle in a Haystack

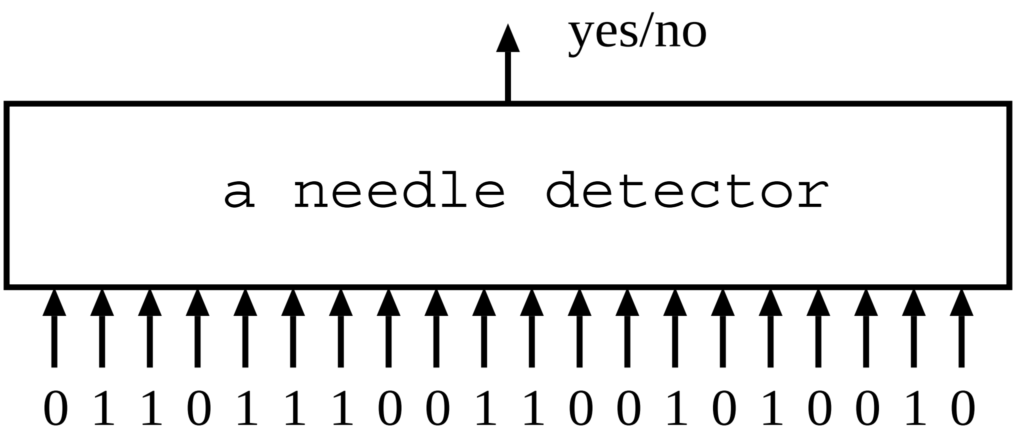
A schematic skecth on fictitious 2D-space



The original Hinton & Nowlan's Needle:

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

□ **How can we train the detector?**



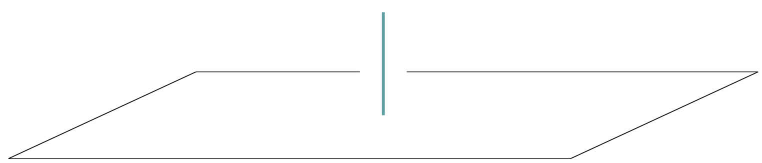
Can we train it with most likely haystack points?

□ How difficult?

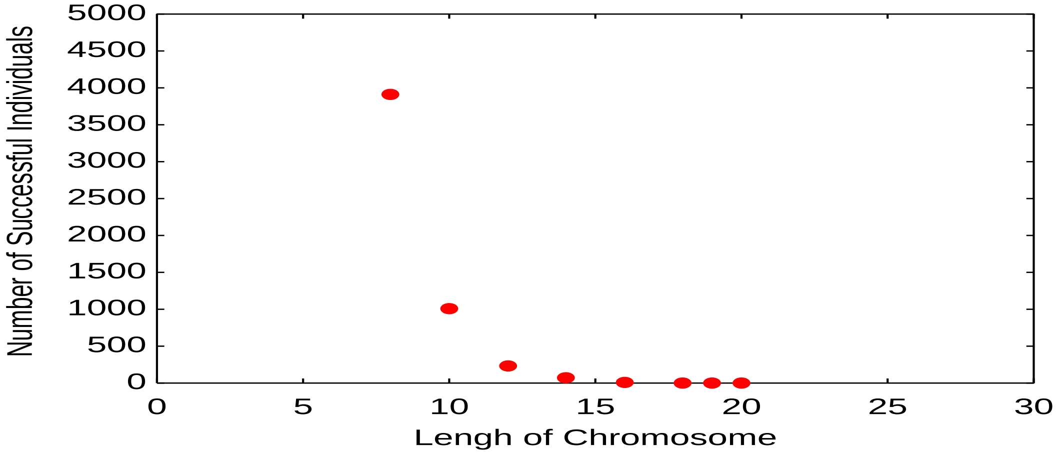
- Random Search

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

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



□ Random Search



□ **We have attacked this problem with lately reported approaches each of which claims very SUCCESSFUL.**

- Artificial Immune System
- Evolutionary Computation
- Fuzzy Rule
- Data-mining Technique
- etc

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

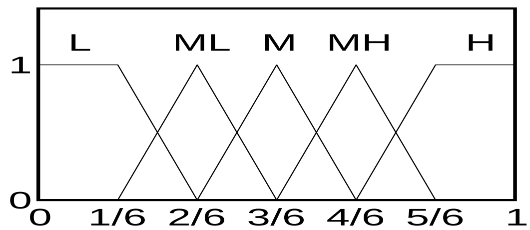
E.g., Kim & Bentley (2001) claimed

assuming one subspieces of IRIS is abnormal while other two normal

$\Downarrow$

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

□ A Fuzzy Rule approach — How many rules we need?



$$\Downarrow$$

$$(MMMMMM \cdots M)$$

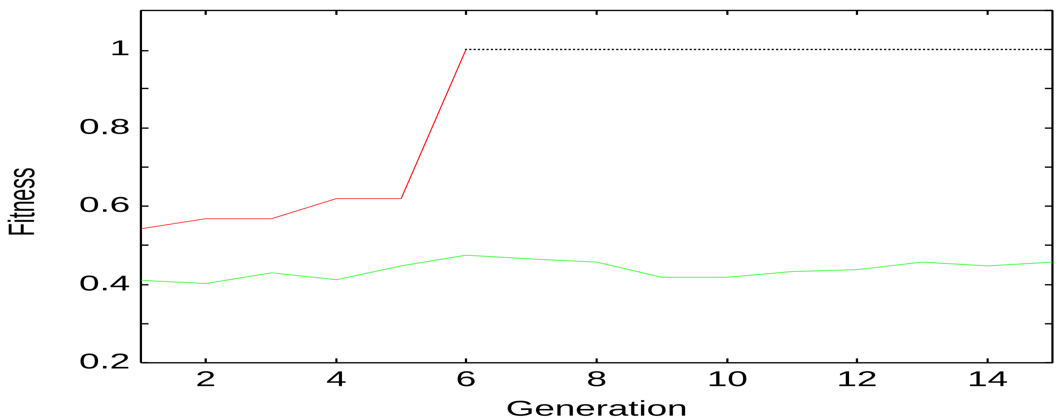
$$\Downarrow$$

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 ($x_i \in [0.45, 0.55]$)

$\Downarrow$



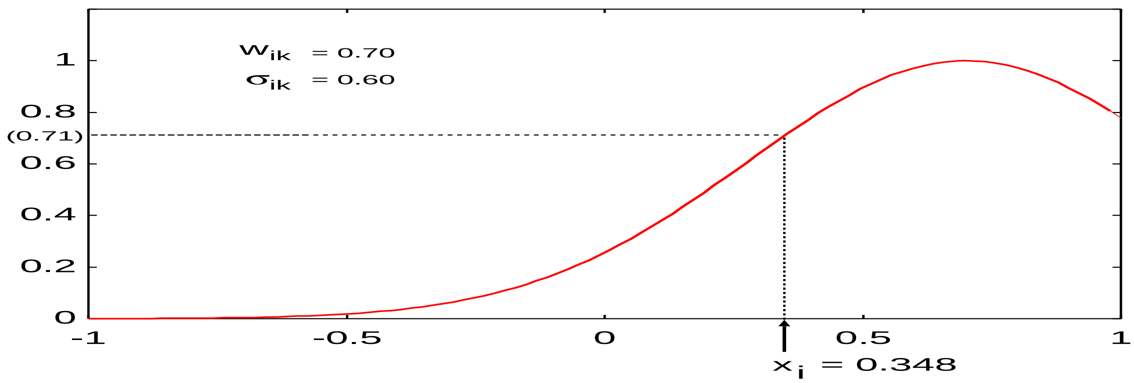
□ A curse of dimension

For the island $x_i \in [0.45, 0.55]$ in the 20-D lake

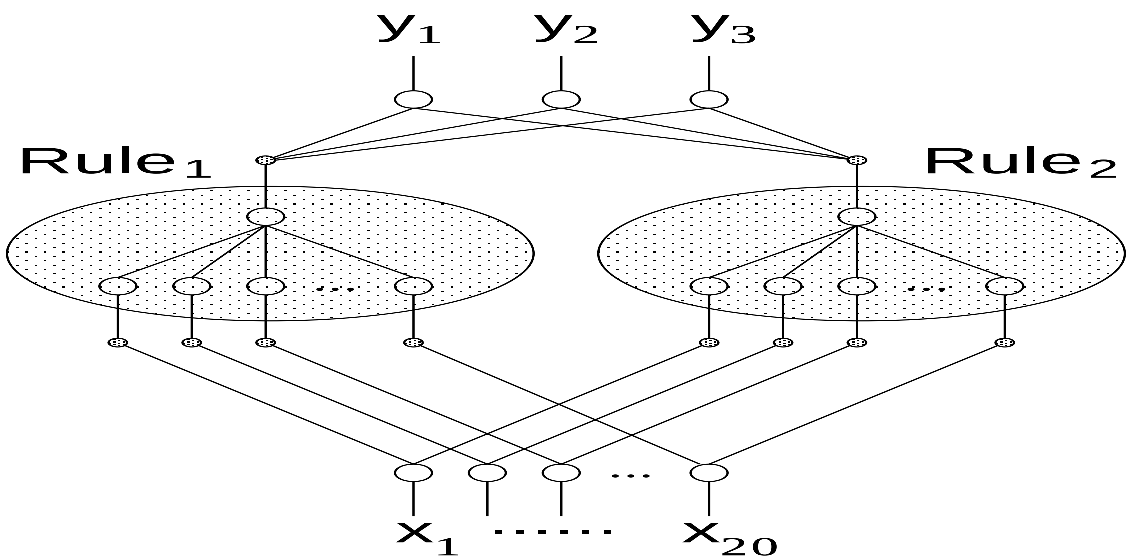
$$\Downarrow$$

$$(0\ 1\ 3\ 4\ 0\ 4\ 4\ 1\ 4\ 4\ 1\ 1\ 4\ 1\ 4\ 3\ 0\ 0\ 0\ 2)$$

Shape/Location adaptive membersip function



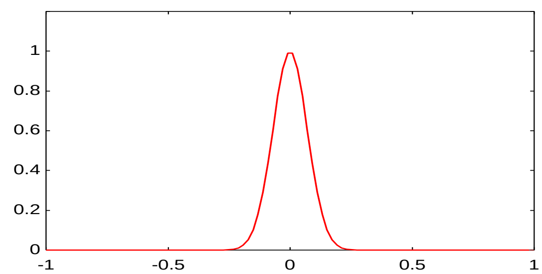
A Fuzzy Neural Network Approach



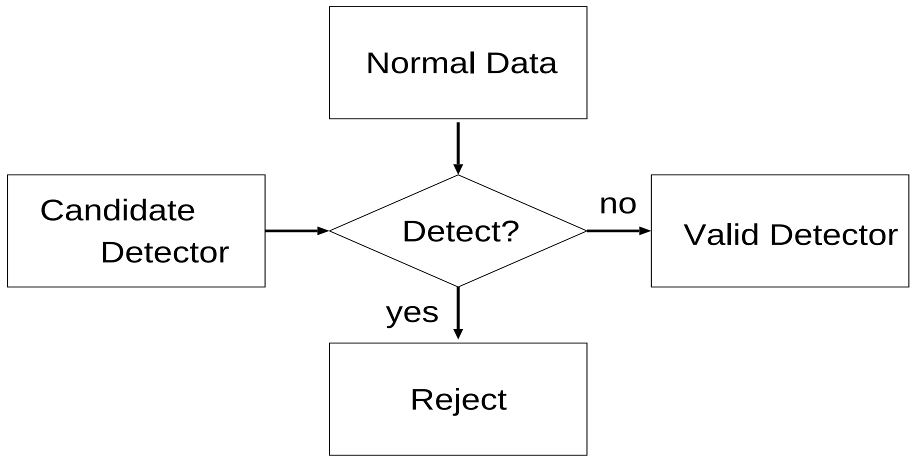
A result of an evolution

$$M$$

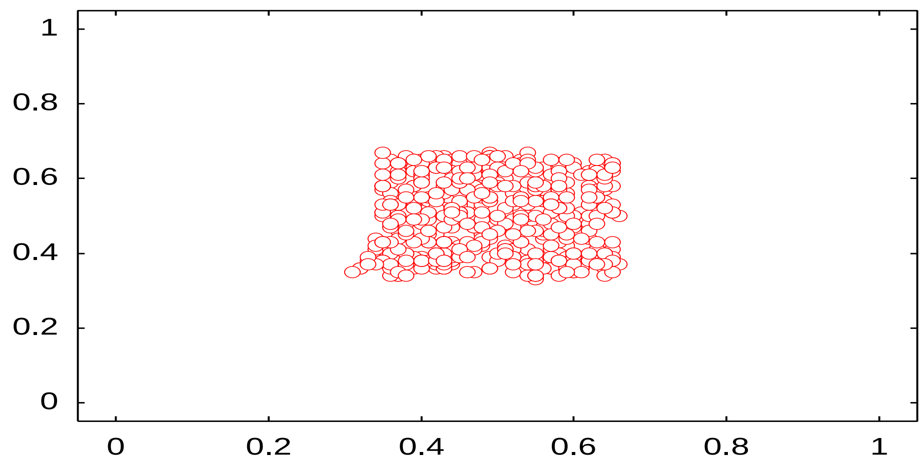
$$\Downarrow$$



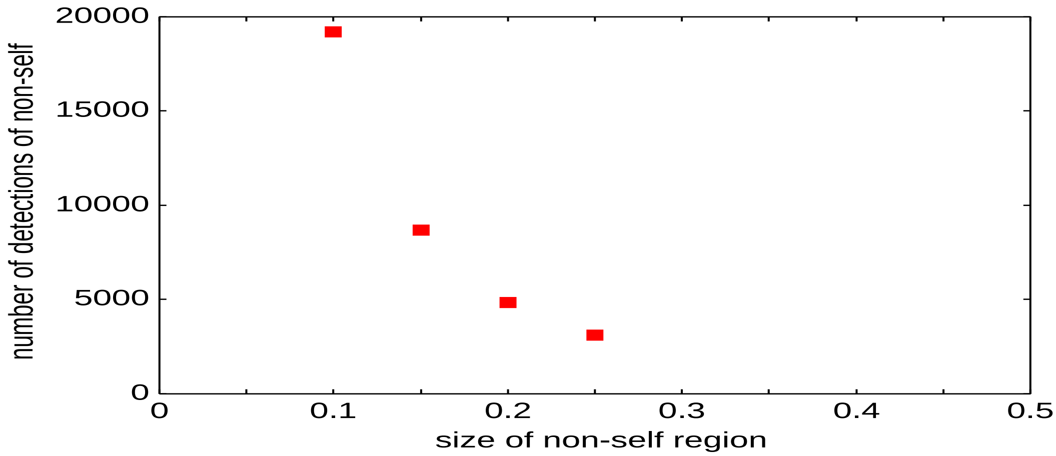
□ An immune approach — Constant-sized hyper-shpere Detectors



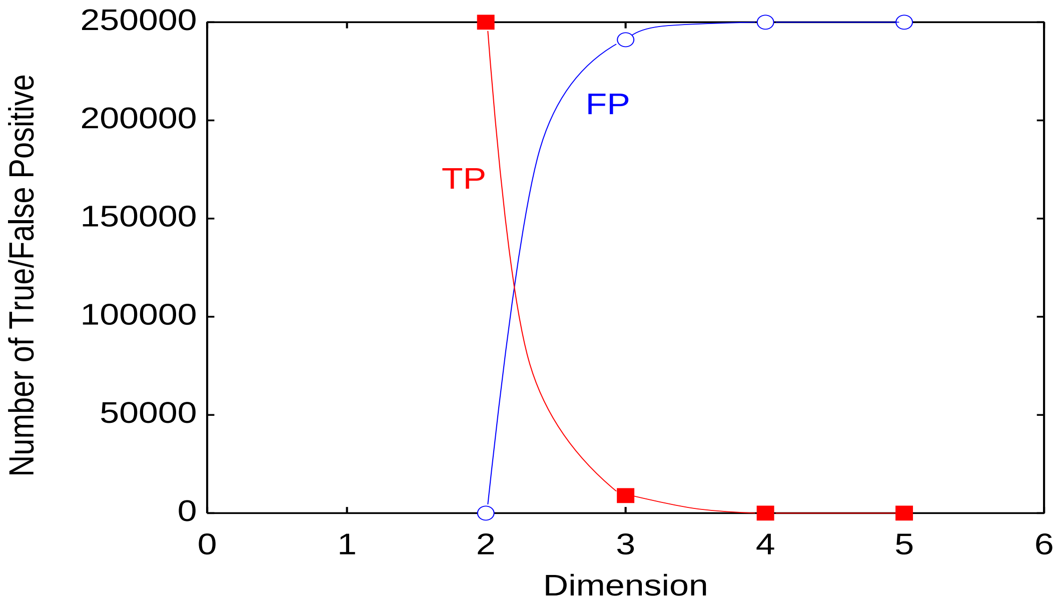
□ A result in 2-D space



□ **What if top area shrinks to zero?**



□ Alas! As dimension grows...



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

$\Downarrow$

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.

Though we now are negative more or less, still want to be neutral.