How Many Parachutists will be Needed to Find a Needle in a Pastoral?

Akira Imada

Brest State Technical University
Belarus

Intrusion Detection by Soft-computing

CONTENTS

- Assuming *Supervised Learning*
  - ⋆ We need samples for *Training* & *Testing*
    - · *Iris-flower*
    - · *KDD-cup-99*

(cont'd)

CONTENTS

- *Conjecture 1*
  "Intrusion is like a-needle-in-a-hay-stack
  not like an iris family."

- *Conjecture 2*
  "No such algorithms
  to efficiently find a needle in a haystack."

(cont'd)

CONTENTS

- Warnings
  - ⋆ *Neither IRIS nor KDD-cup-99 are meaningful!*
  - ⋆ *Training should be only with normal samples!*
  - ⋆ *Don't assume any expectations a priori!*
- Concluding Remarks

How to design an intrusion detection system intelligently, if any?

- Two categories
  - ⋆ Detection of *known attack*.
  - ⋆ Detection of *unknown no-normal*.

To detect known-attacks, intelligence is not necessary but a lookup-table is enough

⇓

Our interest is on detecting *unknown no-normal*.

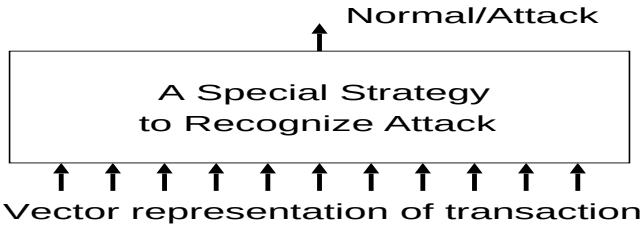
General assumption:

- TCP/IP connection to a network

  $\Rightarrow$ can be represented by an n-dimensional vector

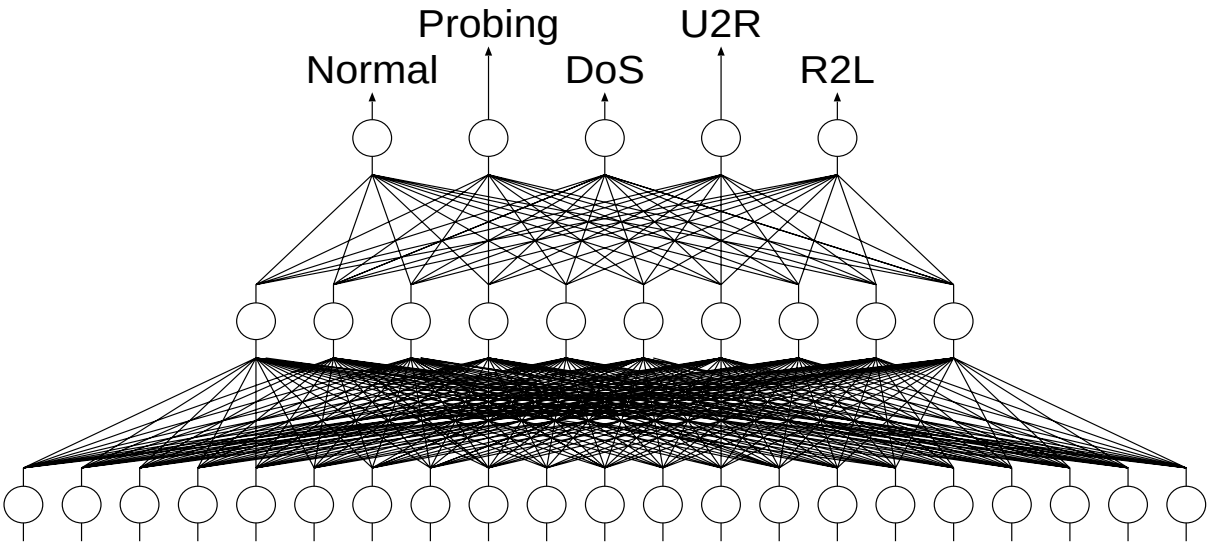
Our Goal

To design a machine which tells
the input is a *normal* or an *attack*.



Hopefully with
$Detection\ Rate > 90\%\ \&\ False\ Alarm\ Rate < 10\%$

An example of NN implementation



pre-proccessed 41 attributes in KDD cup 1999 data set

A question

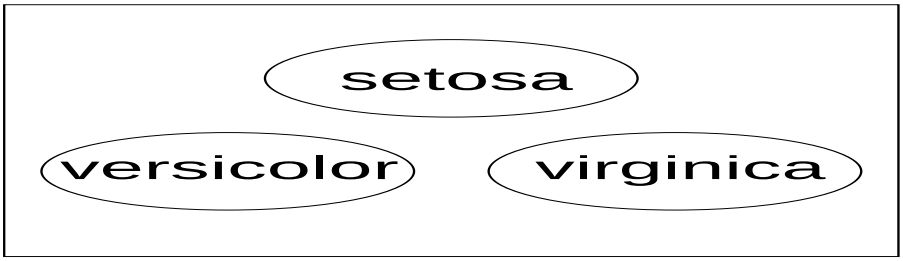
"Is it possible or not?"

Training & Testing

- Frequently used two public domain datasets.
  - ⋆ Iris-flower dataset.
  - ⋆ KDD-cup-1999 dataset

I. What is iris flower dataset

- 3 families of iris flower ⇒ 150 samples in total;
- Each family ⇒ 50 samples; Each sample ⇒ 4 features.



IRIS flower

Lots of successes have reported so far

Castellano et al. (2000)
assumed one to be normal while the other two abnormal.

$\Downarrow$

(by Fuzzy-NN with T-S model)

- abnormal detection rate = 96%
- false alarm rate = 0.6%

(cont'd)

But can we be so optimistic?

$\Downarrow$

When a family of iris is normal then are others abnormal?

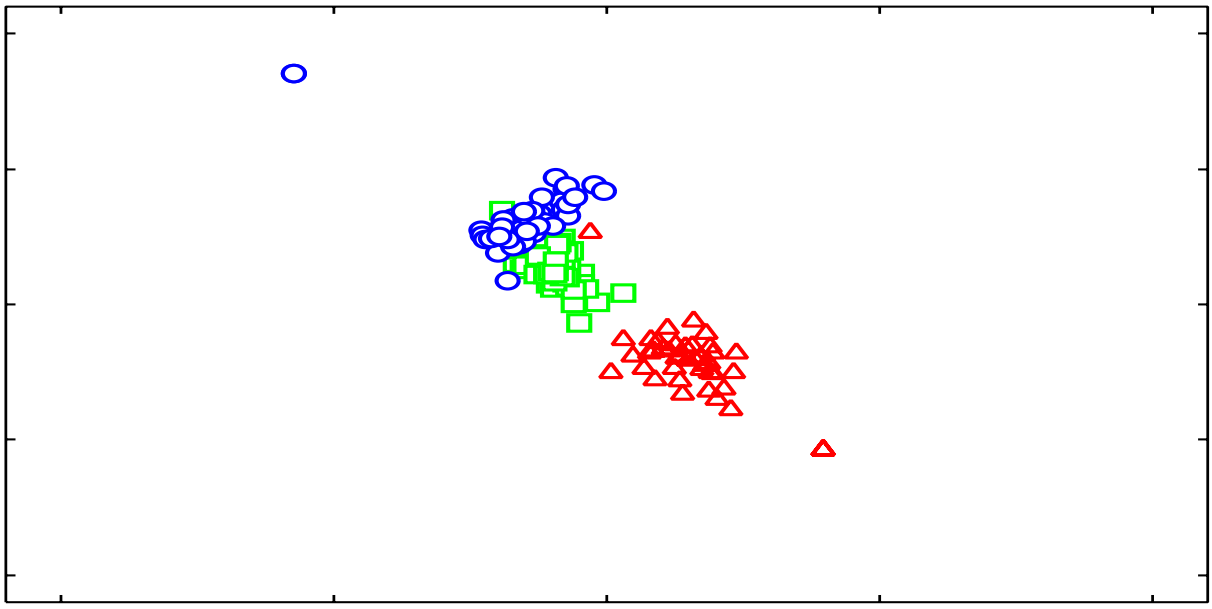
Let's visualize iris families — Sammon mapping

Sammon mapping maps points in a high-D space to 2-D

by

keeping distance relation preserved as much as possible
or
approximating distances in the n-D space
by distances in 2-D space with a minimal error.

A Sammon mapping of iris families



Where do outliers lie?

Not in the domain for the other families!!

$\Downarrow$

Not at the point at random, either.

$\Downarrow$

But outlier usually hides behind Normal.

□ Challenge 1

Attacks by Mutant

(1) Assume one family as $\{Normal\}$

(2) Mutate the normal samples and take them as $\{Attack\}$*.*

(3) Train with half of the $\{Normal + Mutant\}$

(4) Test with remaining $\{Normal + Mutant\}$

$\Rightarrow$ Will a successful result be possible?

The 1st warning

Iris is good to start with, but not of so useful.

II. What is the KDD-cup-1999 dataset?

- *Normal*
- *Attack* $\Rightarrow$ 32 attack types of the 4 categories:

(cont'd)

- *Attack* $\Rightarrow$ 32 attack types of the 4 categories:
  - $\star$ *Probing*

    by proving a vulnerability of the network;
  - $\star$ *Denial-of-Service (DoS)*

    by denying legitimate requests to a system;
  - $\star$ *User-to-Root (U2R)*

    an unauthorized access to local super-user or root;
  - $\star$ *Remote-to-Local (R2L)*

    an unauthorized local access from a remote machine.

KDD-cup-99 winner's result

Detection rate for 4 attack types

Probe	*DoS*	*U2R*	*R2L*
83.3%	*97.1%*	*13.2%*	*8.4%*

Detection rate by Sabhnani et al. (2003)

	Probe	*DoS*	*U2R*	*R2L*
Multi-layer Perceptron	*88.7*	*97.2*	*13.2*	*5.6*
Gaussian Classifier	*90.2*	*82.4*	*22.8*	*9.6*
K-mean Clustering	*87.6*	*97.3*	*29.8*	*6.4*
Nearest Cluster Algorithm	*88.8*	*97.1*	*2.2*	*3.4*
Radial Basis Function	*93.2*	*73.0*	*6.1*	*5.9*
Leader Algorithm	*83.8*	*97.2*	*6.6*	*1.0*
Hypersphere Algorithm	*84.8*	*97.2*	*8.3*	*1.0*
Fuzzy Art Map	*77.2*	*97.0*	*6.1*	*3.7*
C4.5 Decision Tree	*80.8*	*97.0*	*1.8*	*4.6*

We doubt some sort of the reported success

- Dimension Reduction by PCA – by Kuchimanchi et al.
  - ⋆ 41-D to 19-D
    - · (detection accuracy, false positive) = (99.92%, 0.26%)
  - ⋆ while the original being
    - · (detection accuracy, false positive) = (99.94%, 0.23%)

Yet another interesting claim (cont'd)

- Joshi et al. (2005)
  - ⋆ 41-D to 5-D (taking the first five from above)
    - · (detection accuracy, false positive) = (79%, 21%) (using Hidden Markov Process)

Let's not go into detail here, but rather focus on

Why U2R and R2L attacks resist to be detected?

How big is the KDD-cup-1999 dataset?

Labeled 4,898,430 records $\Rightarrow$ *Training*
Un-labeled 311,029 records $\Rightarrow$ *Testing*
and
Each record is 41-dimensional vector

$\Downarrow$

Sammon Mapping wouldn't work any more!

Ratios in the samples for testing

Normal	*Probe*	*DoS*	*U2R*	*R2L*
19.5%	*1.3%*	*73.9%*	*5.2%*	*0.1%*

Let's try a thought experiment!

- *always-return-random-answer-strategy*

  which returns either Normal, Probe, DoS, U2R, or R2L at random regardless of the input.

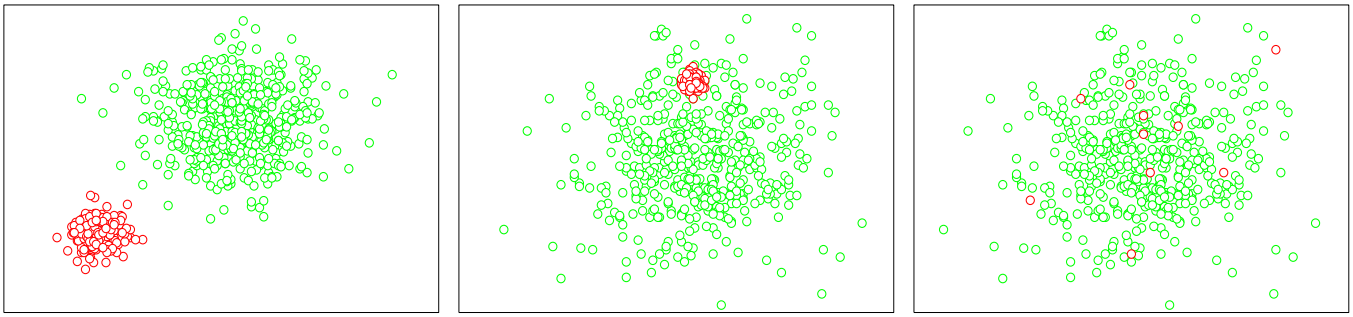
$\Rightarrow$ a high score to detect DoS attacks, like

Normal	*Probe*	*DoS*	*U2R*	*R2L*
19.5%	*1.3%*	*73.9%*	*5.2%*	*0.1%*

A-tiny-set-of-R2L vs. a-huge-set-of-normal

	Total	Normal	Probe	DoS	U2R	R2L
Labeled for Training	4,898,430	19.9%	0.8%	79.3%	≈ 0.0%	≈0.0%
Non-labeled for Testing	311,029	19.5%	1.3%	73.9%	5.2%	0.1%

Fictitious but possible distributions



□ Conjecture 1

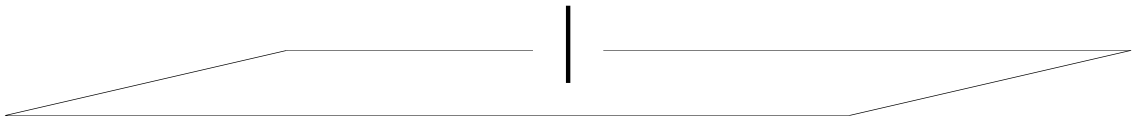
- *U2R, R2L and most of real attacks are like needles in a huge haystack of normals.*

What does a needle look like in a haystack?

The original Hinton & Nowlan's Needle:

- A-needle ⇒ Only one configuration of 20-bit binary string.
  like (11000 11010 11101 0100)
- Haystack ⇒ $2^{20} - 1$ search points

A fictitious needle in a haystack in 2D.



The reason of our title

a needle on pastoral & random fall of parachutists

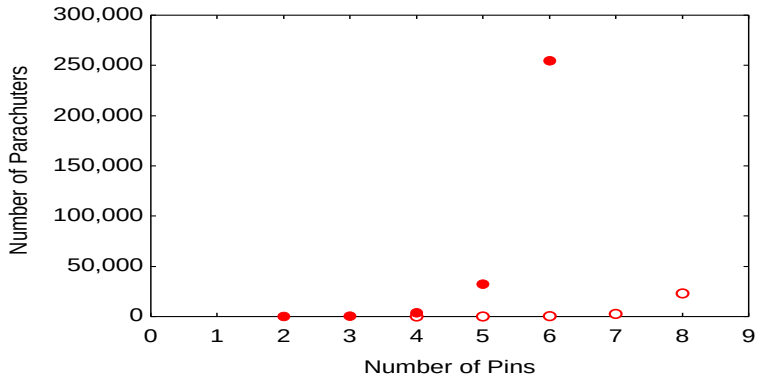
$\Downarrow$

How many parachutists will be needed to find
a needle in a pastoral?

Four experiments to look for a needle in pastoral

1. Random fall – a criterion for comparison
2. What if we permit walks after fall?
3. Neutral mutation
4. Mutation on intron

Random Fall & Walks after Fall



9-digit octal (= 27-bit binary) never stops a run

□ Conjecture 2

- *No such effective algorithm to look for a needle*

Yet another difficulty: The 2nd warning

(2) **Don't assume any expectations a priori!**

(3) *Training should be only with normal samples!*

"Artificial immune system detects a virus attack!"

How fantastic it sounds!

Forrest, Perelson et al. (1994)
"Self Nonself Discrimination in a Computer"

But more than a decade has passed since then,
and still an open issue.
⇓
We hope not, but isn't it just a fantasy?

□ Challenge 2

- *Always-return-random-answer-strategy*

vs.

- *An intelligently designed system*

Placebo Experiment

(1) Prepare a dataset uniformly from all attack data

(2) Test your intelligent system with dataset above

(2) Compare the result with the random-return-machine

The last warning

(2) *Don't assume any expectations a priori!*

(3) **Training should be only with normal samples!**

From papers

P. Laskov et al. (2005)

"In a real application, one can never be sure that a set of available labeled examples covers all possible attacks. If a new attack appears, examples of it may not have been seen in training data."

Colin M. Frayn (2006)

Real-world data are very rarely well-behaved in the sense required by conventional algorithms.

Me (2005)

"When we know the attack, it's too late"

With goal being training-only-by-normal

Gomez et al. (2003)

"A set of fuzzy rules characterized abnormal space using only normal samples."

with

10% dataset of KDD-cup-1999

$\Downarrow$

"It detects attacks with the detection rate 98.30% and false alarm rate 2.0%."

$\Downarrow$

Really satisfactory, if it's really true.

□ Challenge 3

Attack by dummy

(1) Prepare {Normal} and {Attack} from KDD-cup-99.

(2) Create {Dummy} at random

(3) Train with only {Normal}

(4) Test with {Attack} and {Dummy} one by one

⇓

Can a Sommelier be trained without bootlegs?

CONCLUSION

- *Neither IRIS nor KDD-cup-99 are useful.*
- *A hacker is a person who is extremely good at finding a pattern which is very close to the normal.*
- **So, we must regard attack as a mutant-of-normal.**

- *This is not to negate the possibility, but we hope this will be a start of debate for an emergence of innovation.*