CS344: INTRODUCTION TO ARTIFICIAL INTELLIGENCE

Pushpak Bhattacharyya
CSE Dept.,
IIT Bombay

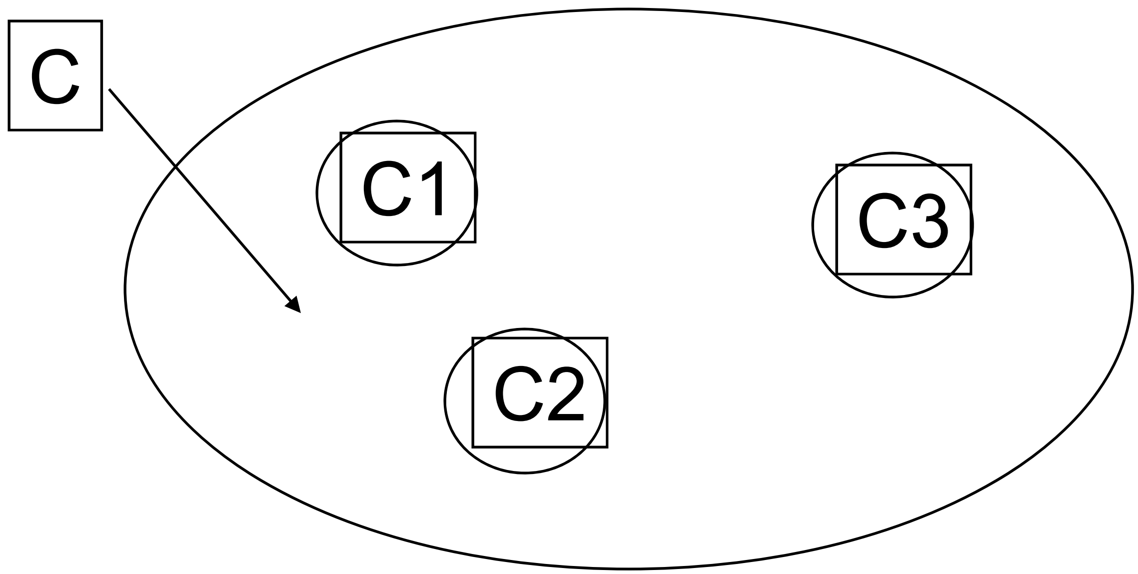
Lecture 38: PAC Learning, VC dimension; Self Organization

VC-dimension

Gives a necessary and sufficient condition for PAC learnability.

Def:-

Let C be a concept class, i.e., it has members c1,c2,c3,…… as concepts in it.



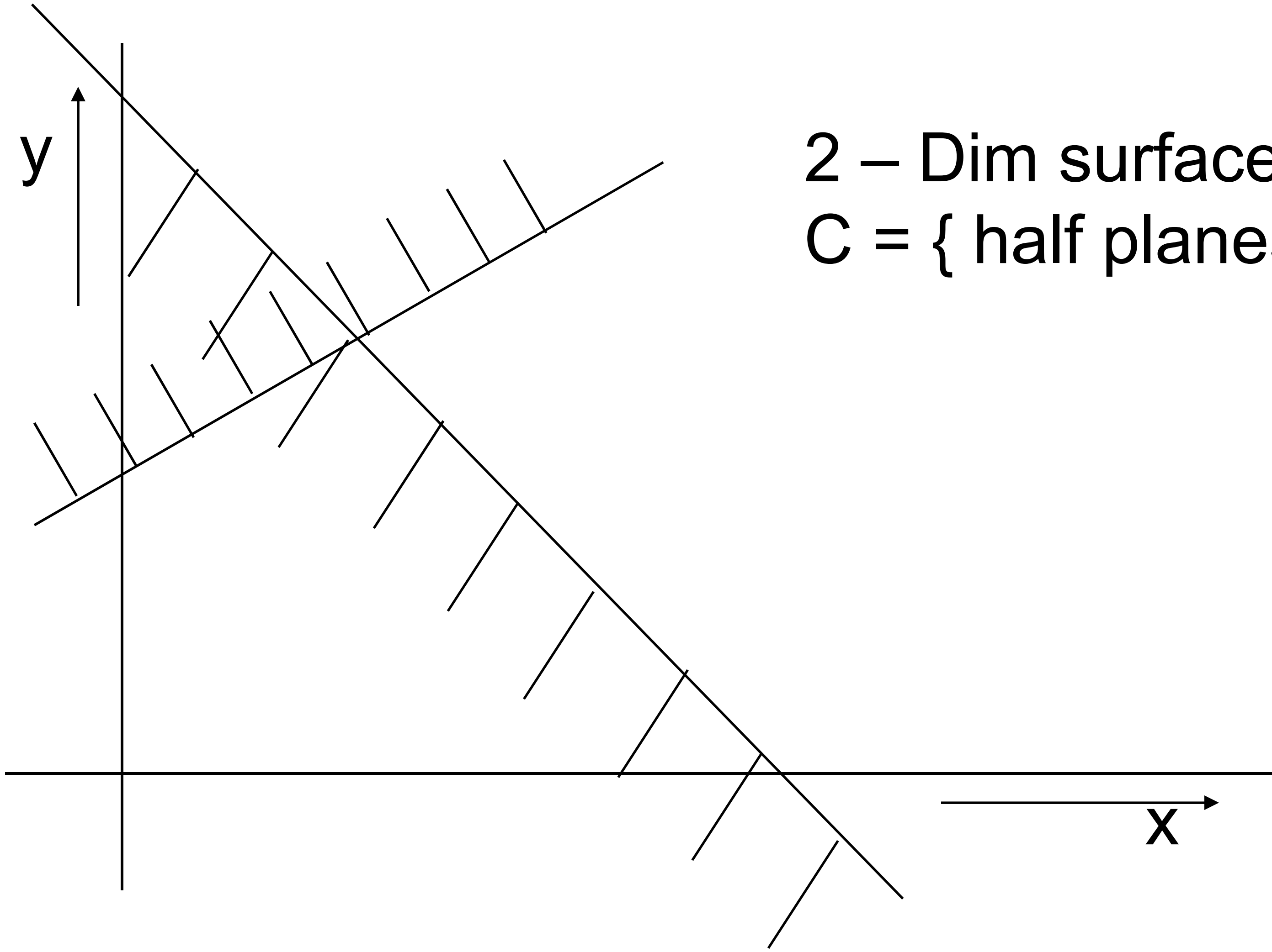
Let S be a subset of U (universe).

Now if all the subsets of S can be produced by intersecting with C_i^s, then we say C shatters S.

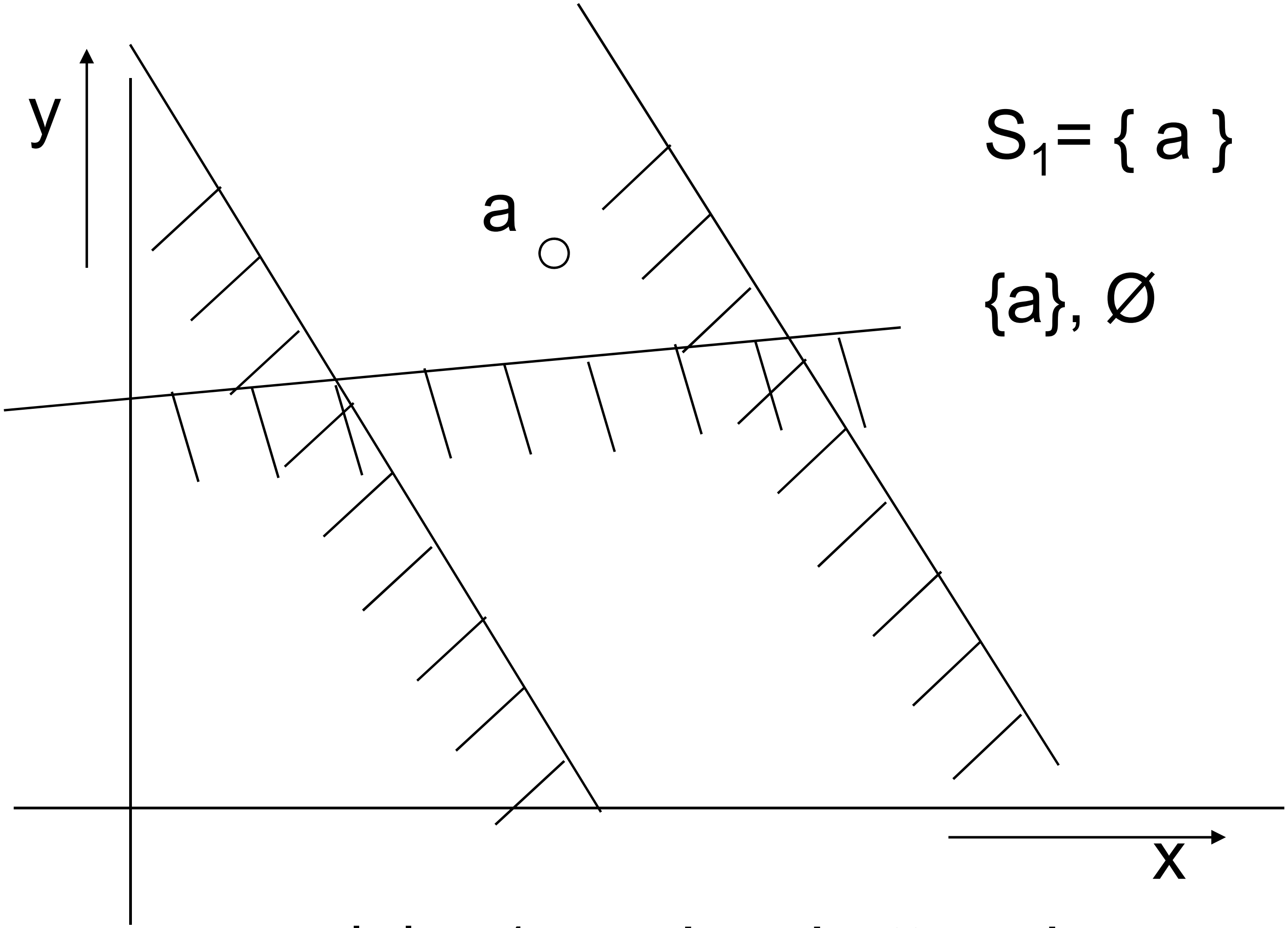
The highest cardinality set S that can be shattered gives the VC-dimension of C.

VC-dim(C)= $|S|$

VC-dim: Vapnik-Cherronenkis dimension.



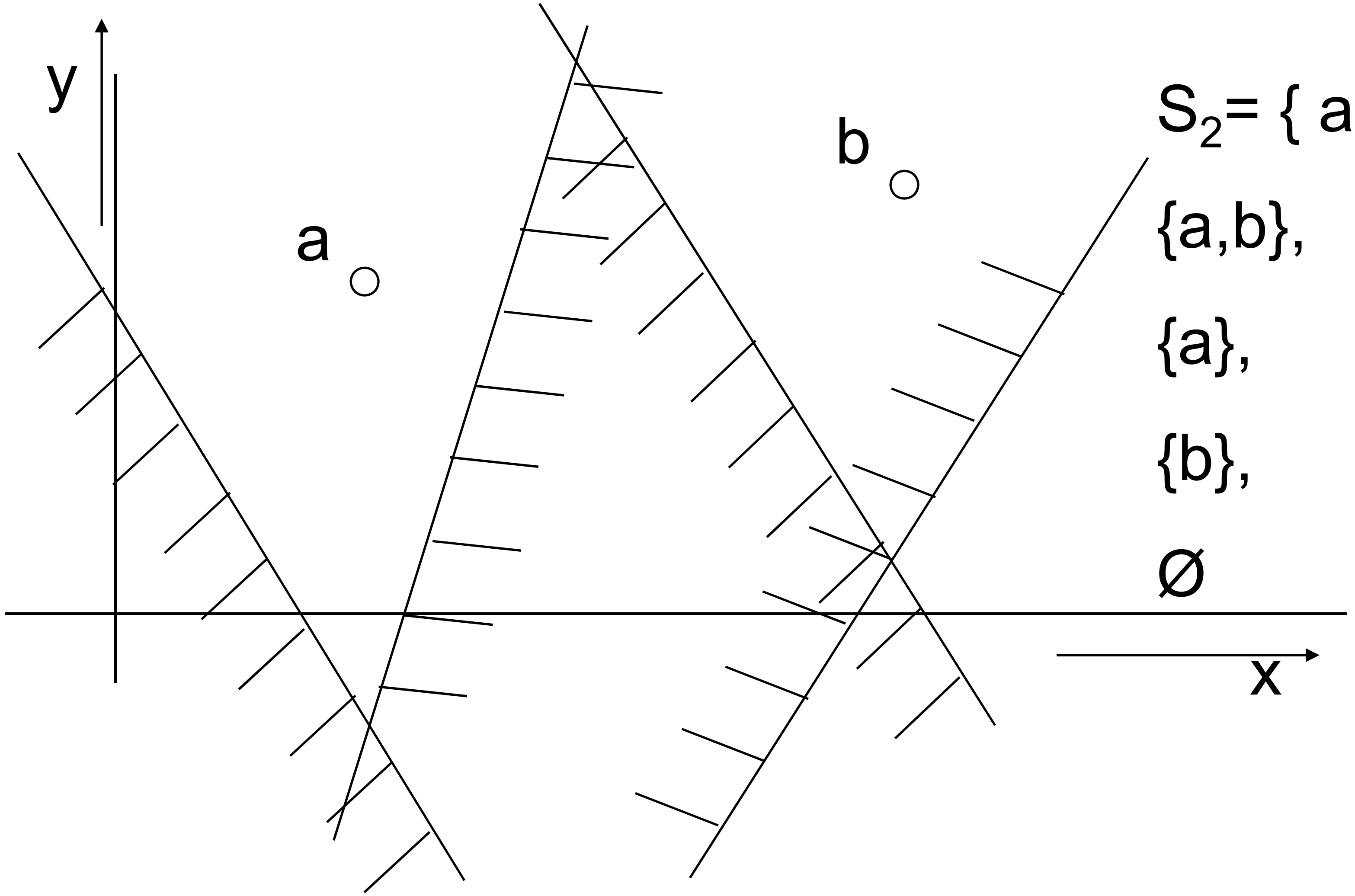
2 – Dim surface
C = { half planes}



$S_1 = \{ a \}$

$\{a\}, \emptyset$

|s| = 1 can be shattered



$S_2 = \{ a,b \}$

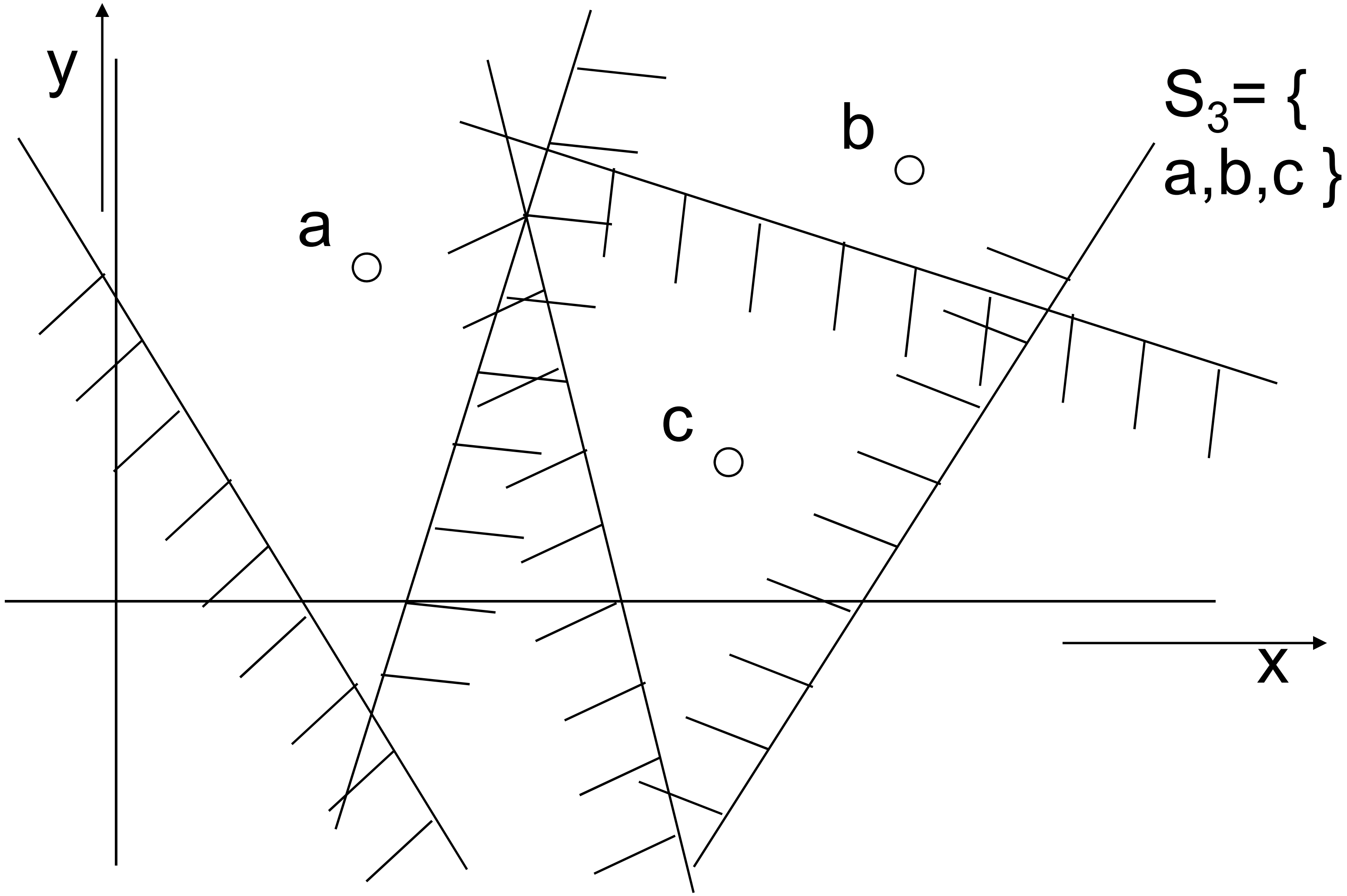
{a,b},

{a},

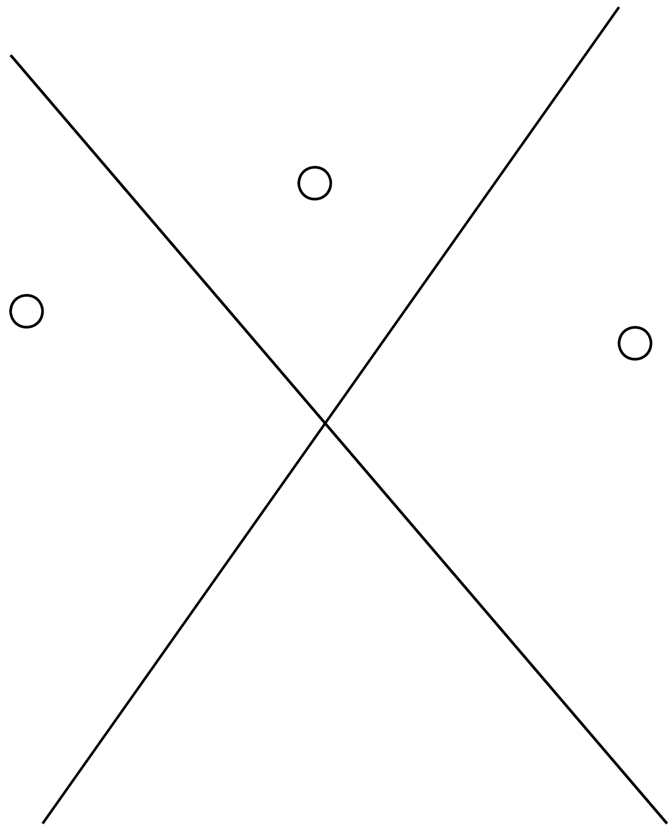
{b},

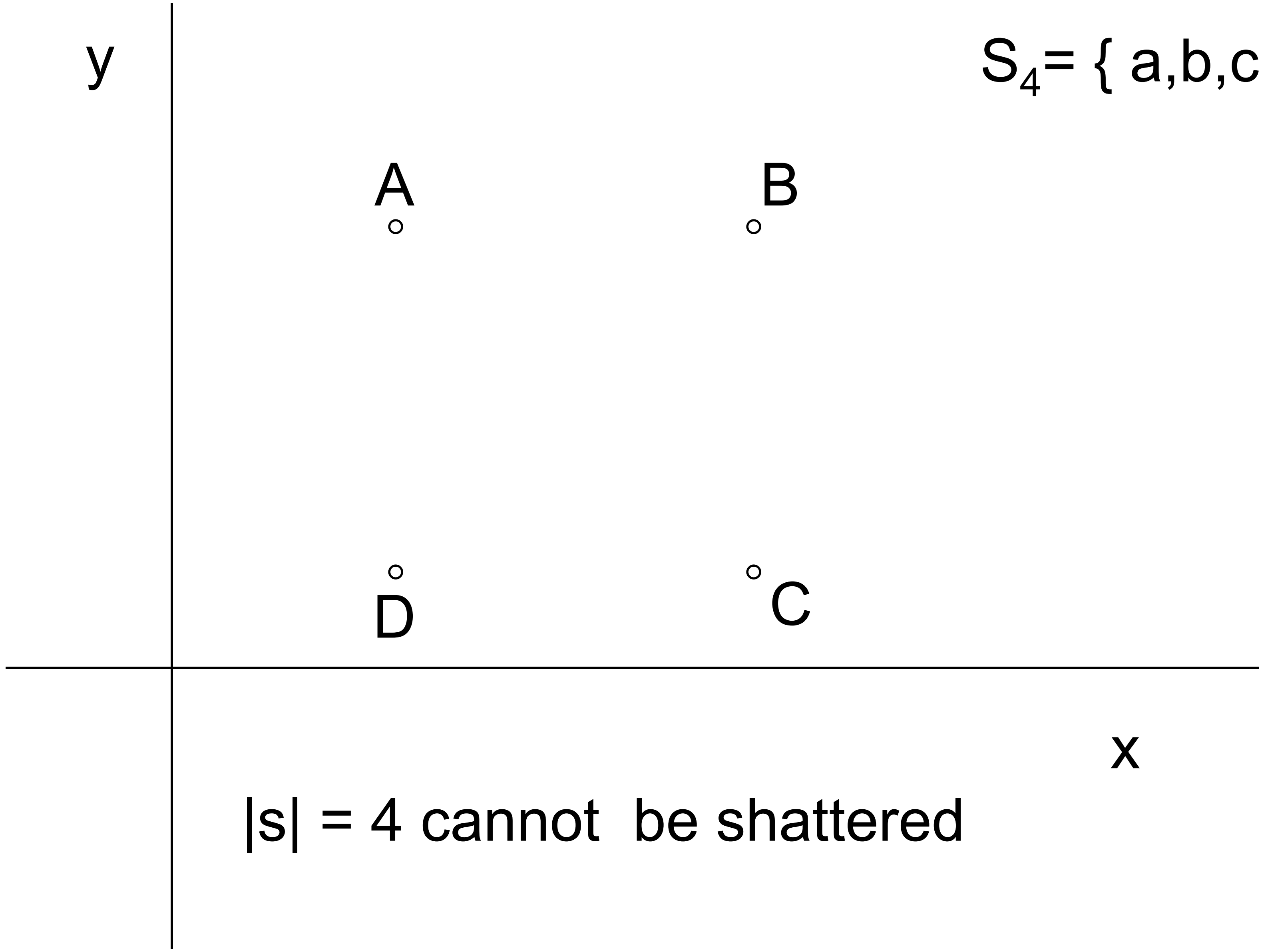
Ø

|s| = 2 can be shattered



|s| = 3 can be shattered





S_4= { a,b,c,d }

|s| = 4 cannot be shattered

Fundamental Theorem of PAC learning *(Ehrenfeuct et. al, 1989)*

- A Concept Class *C* is learnable for all probability distributions and all concepts in *C* if and only if the VC dimension of *C* is finite
- If the VC dimension of *C* is *d,* then…(next page)

Fundamental theorem (contd)

(a) for $0<\varepsilon<1$ and the sample size at least

$$max[(4/\varepsilon)log(2/\delta), (8d/\varepsilon)log(13/\varepsilon)]$$

any consistent function $A:S_c \rightarrow C$ is a learning function for C

(b) for $0<\varepsilon<1/2$ and sample size less than

$$max[((1-\varepsilon)/ \varepsilon)ln(1/ \delta), d(1-2(\varepsilon(1- \delta)+ \delta))]$$

No function $A:S_c \rightarrow H$, for any hypothesis space is a learning function for C.

Book

1. Computational Learning Theory, M. H. G. Anthony, N. Biggs, Cambridge Tracts in Theoretical Computer Science, 1997.

Paper's

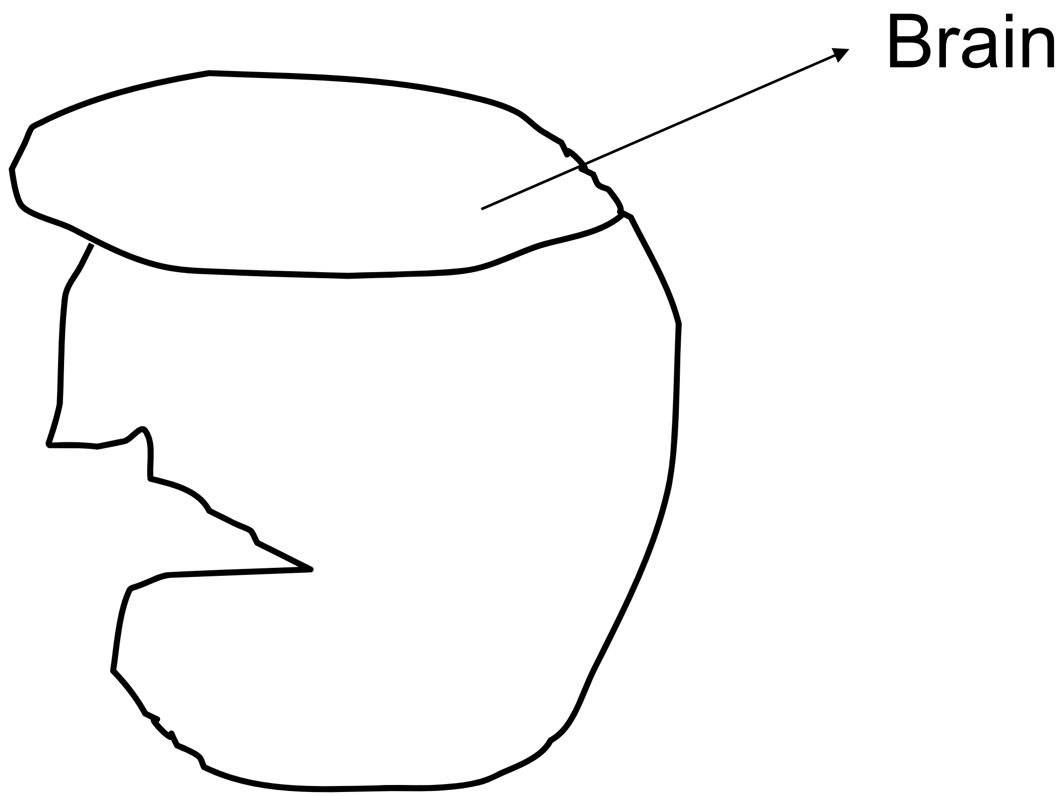
1. A theory of the learnable, Valiant, LG (1984), Communications of the ACM 27(11):1134 -1142.

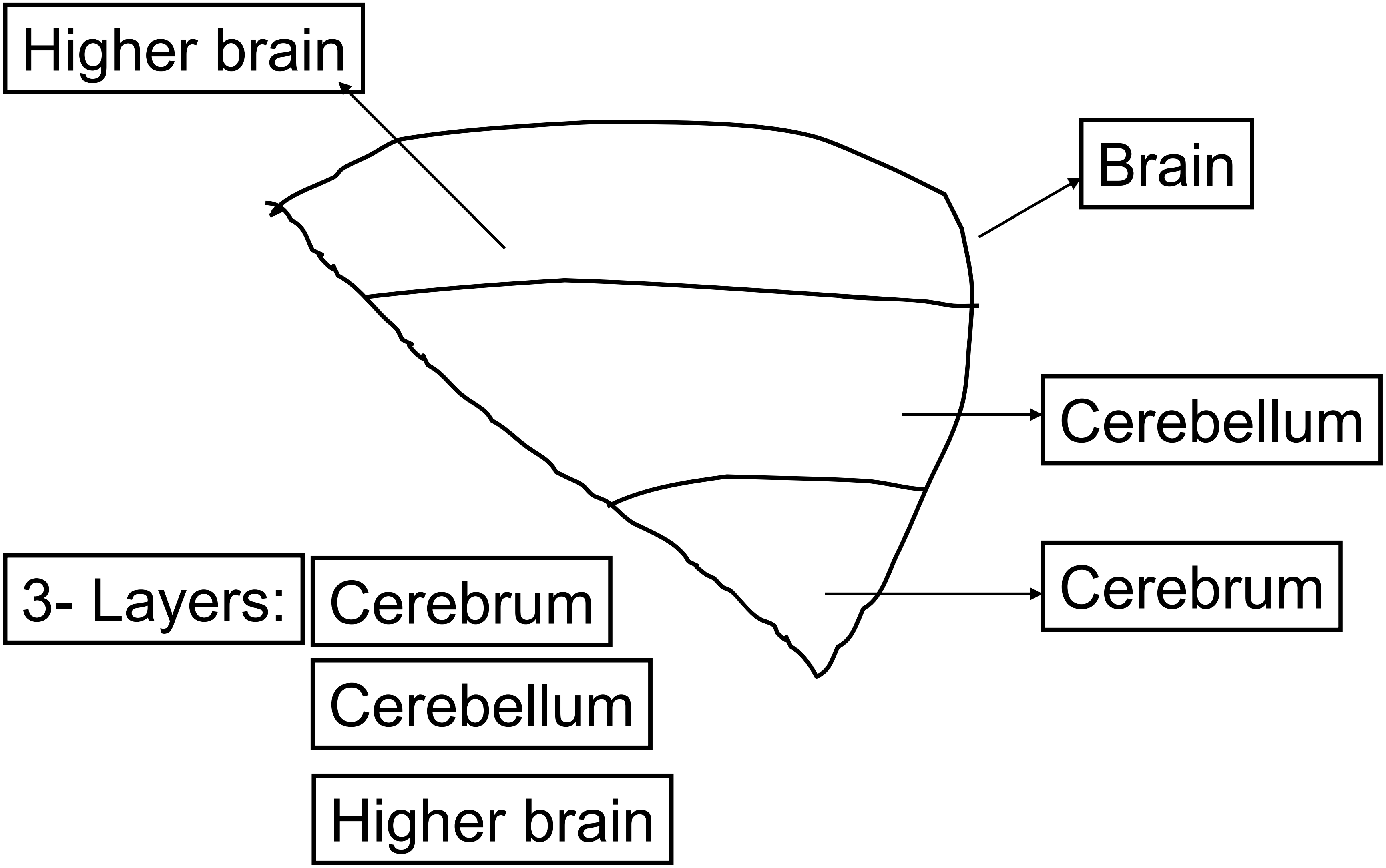
2. Learnability and the VC-dimension, A Blumer, A Ehrenfeucht, D Haussler, M Warmuth - Journal of the ACM, 1989.

SELF ORGANIZATION

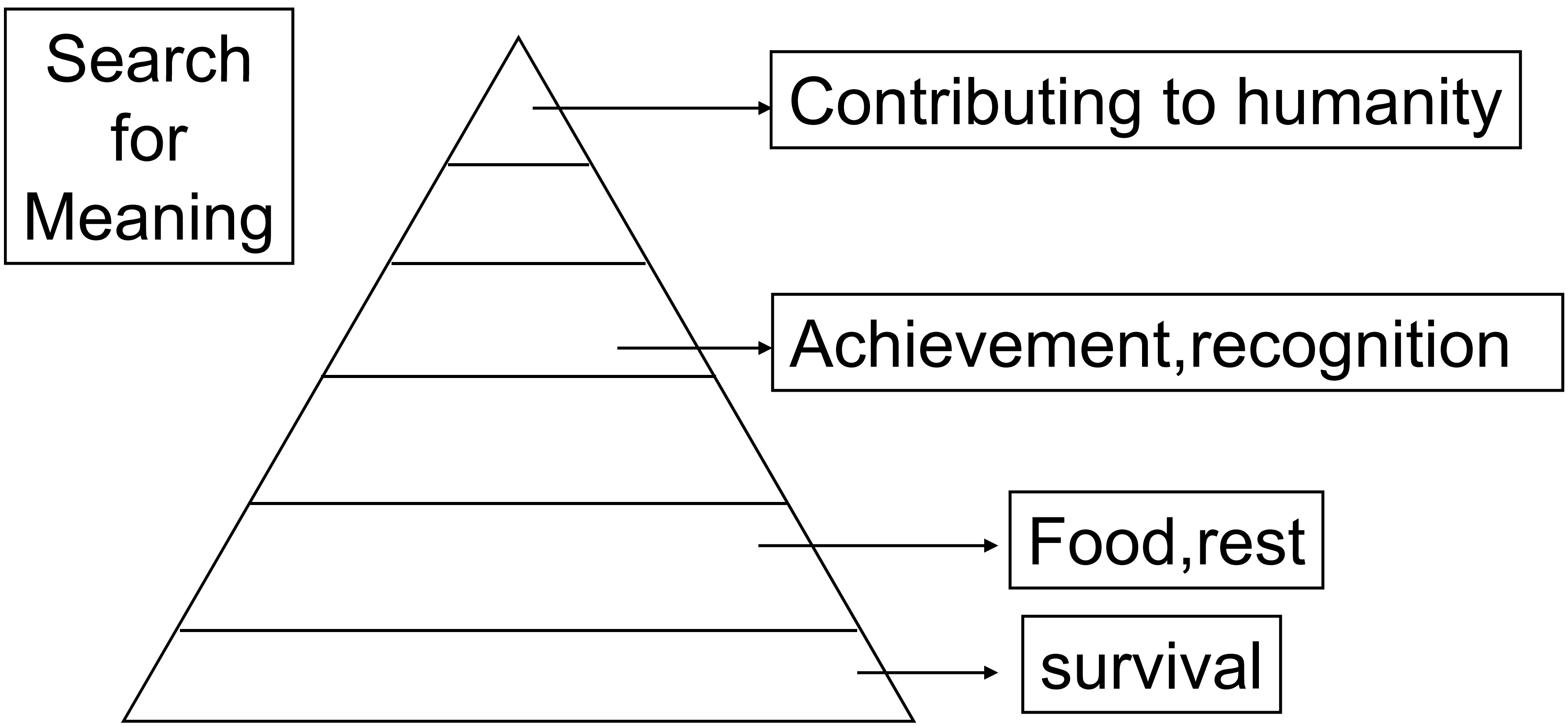
Self Organization

Biological Motivation

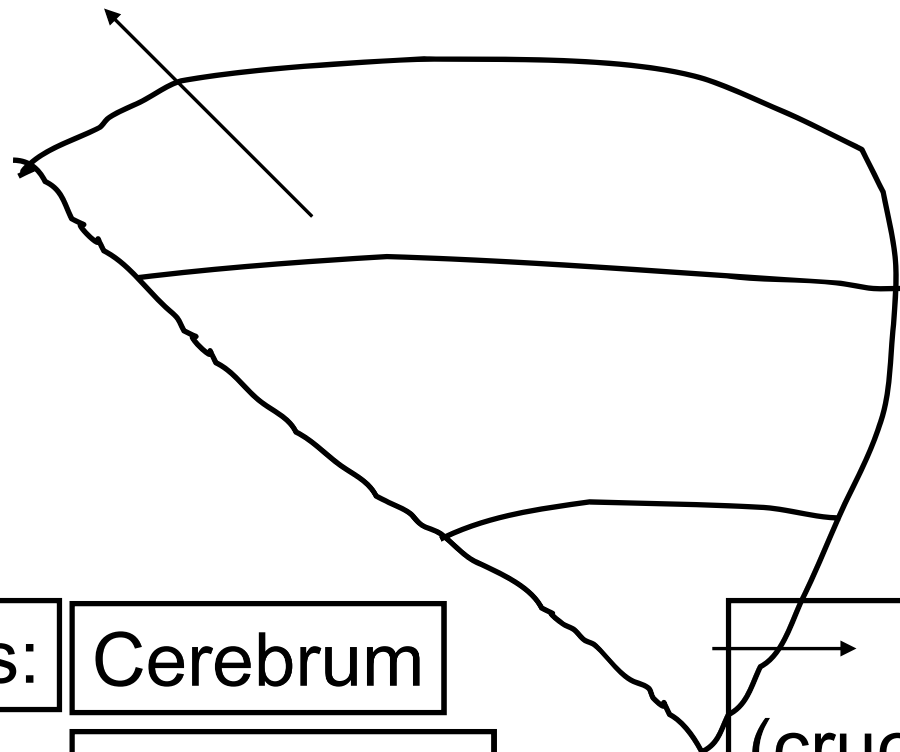



Higher brain
Brain
Cerebellum
Cerebrum
3- Layers:
Cerebrum
Cerebellum
Higher brain

Maslow's hierarchy



Higher brain (responsible for higher needs)



3- Layers:

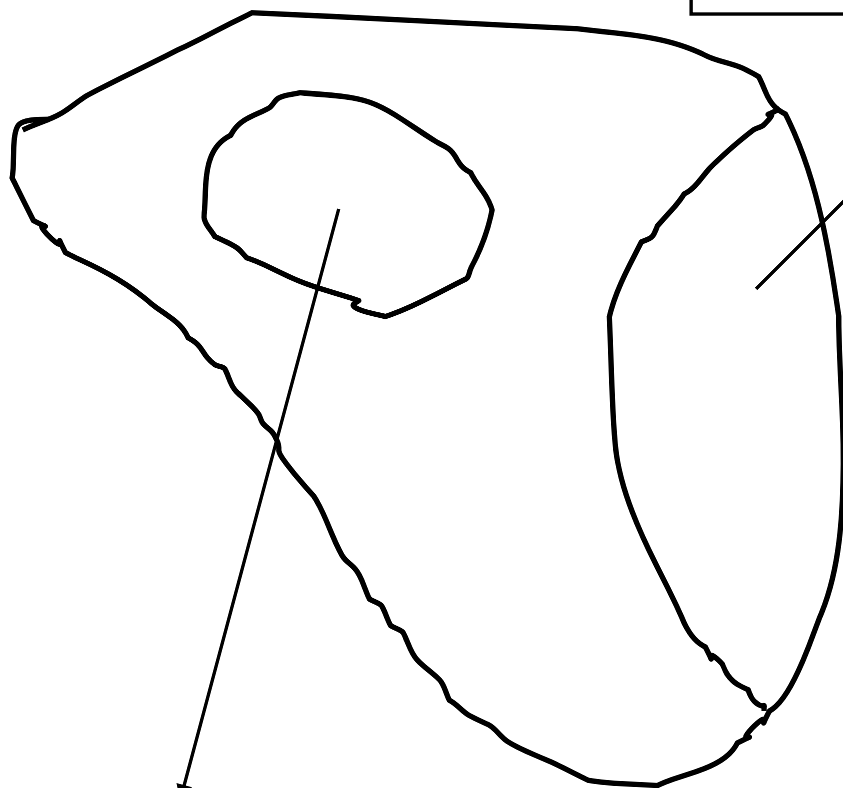
Cerebrum

Cerebellum

Higher brain

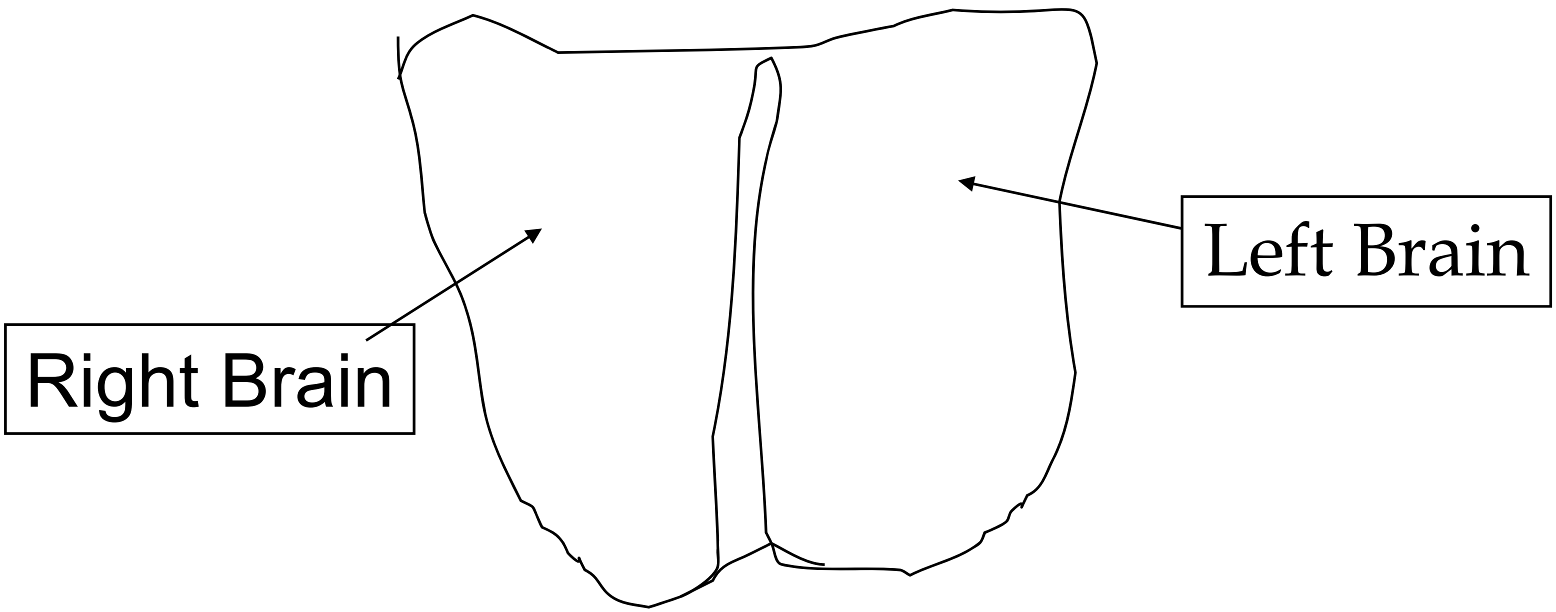
Cerebrum
(crucial for survival)

Mapping of Brain



Left Brain and Right Brain

Dichotomy



Left Brain – Logic, Reasoning, Verbal ability

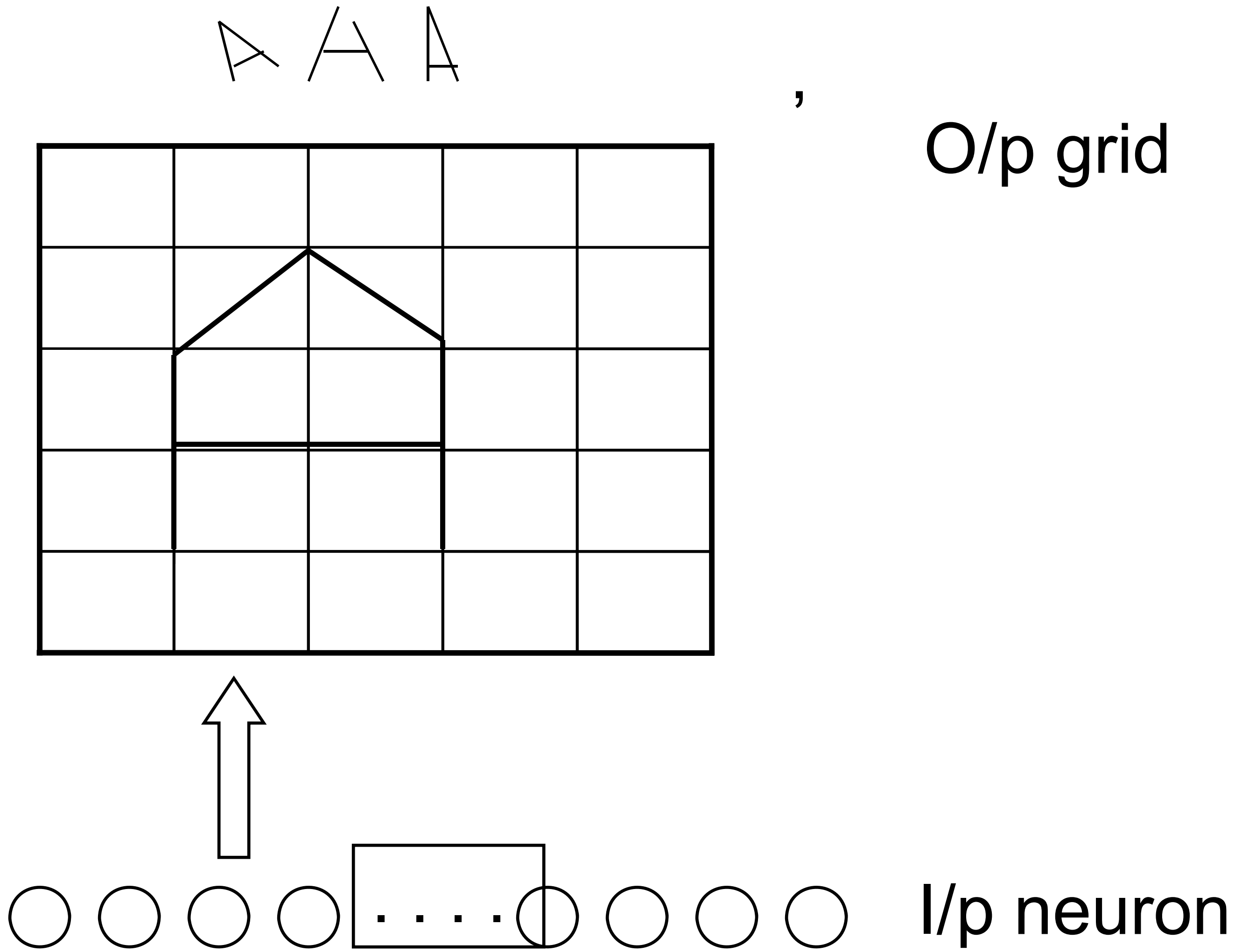
Right Brain – Emotion, Creativity

Music → Words – left Brain

Music → Tune – Right Brain

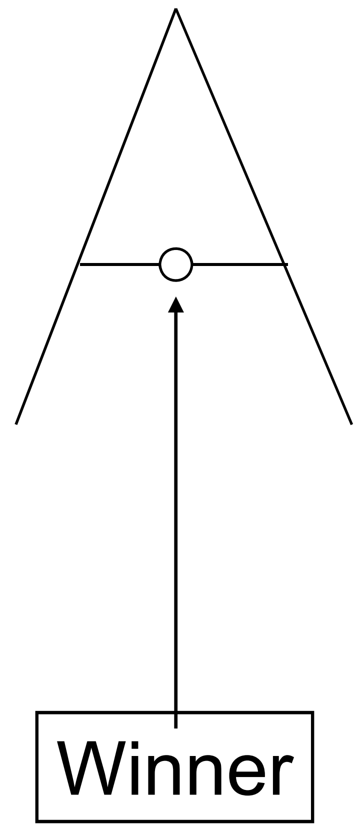
Maps in the brain. Limbs are mapped to brain

Character Reognition



KOHONEN NET

- Self Organization or Kohonen network fires a group of neurons instead of a single one.
- The group "some how" produces a "picture" of the cluster.
- Fundamentally SOM is competitive learning.
- But weight changes are incorporated on a neighborhood.
- Find the winner neuron, apply weight change for the winner and its "neighbors".



Neurons on the contour are the "neighborhood" neurons.

Weight change rule for SOM

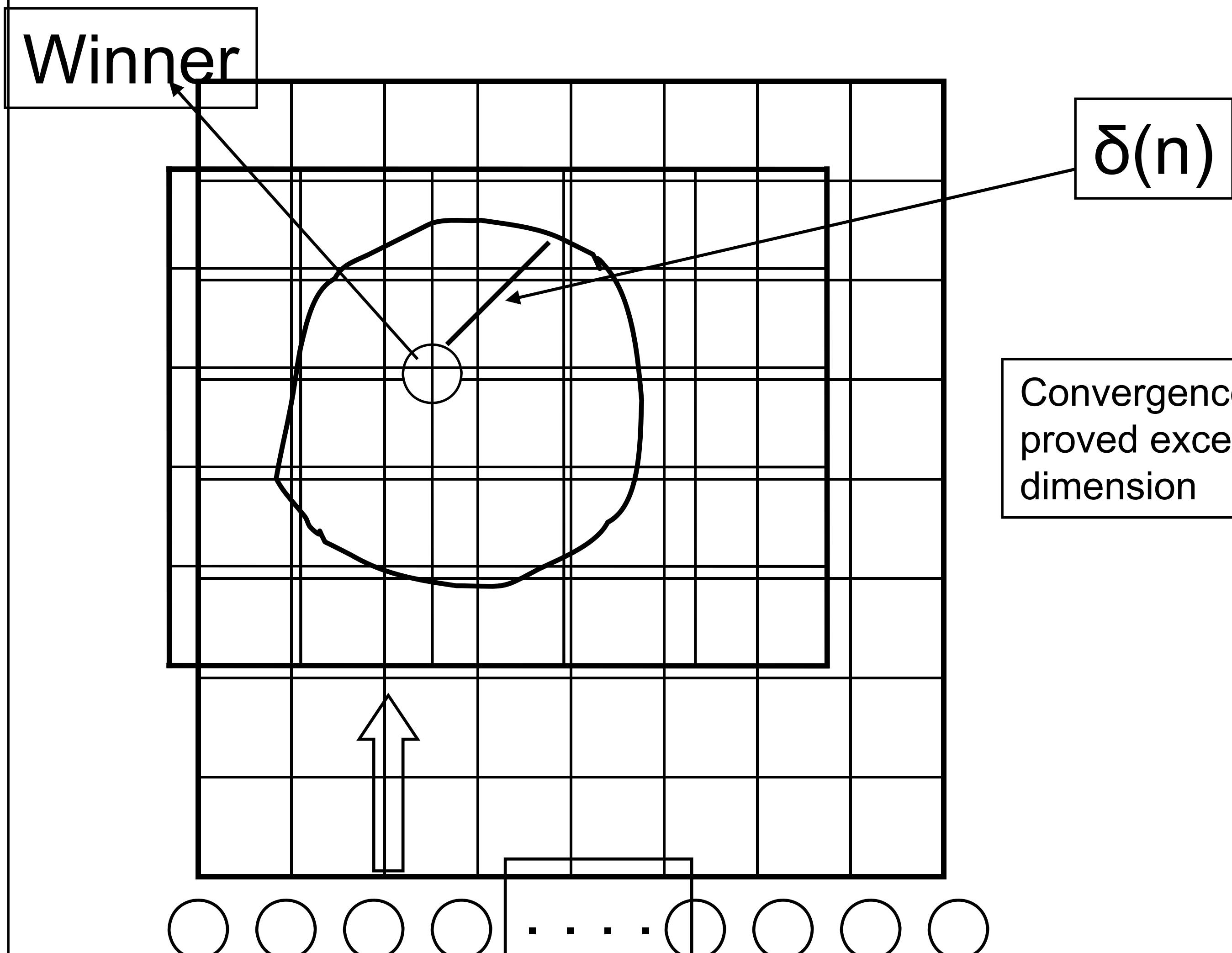
$$W^{(n+1)}_{P+\delta(n)} = W^{(n)}_{P+\delta(n)} + \eta^{(n)} (I^{(n)} - W^{(n)}_{P+\delta(n)})$$

Neighborhood: function of n

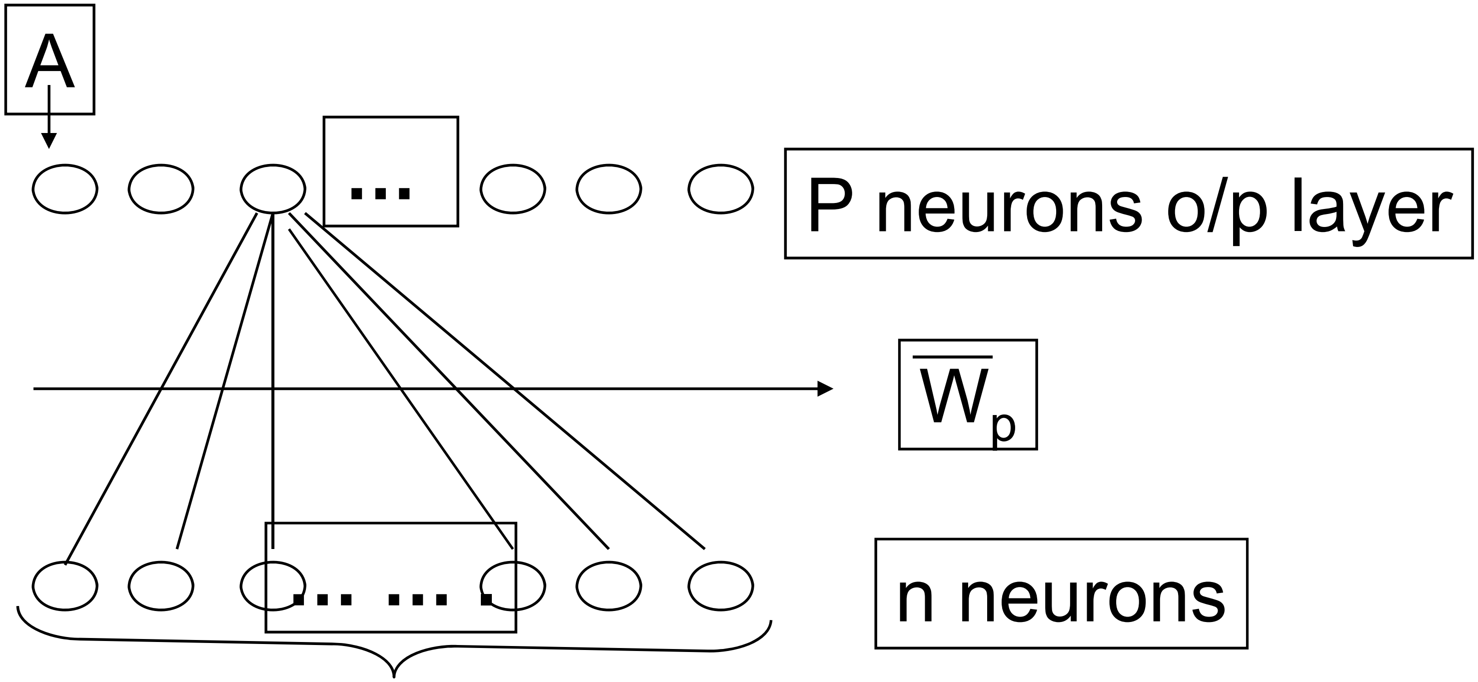
Learning rate: function of n

δ(n) is a decreasing function of n
η(n) learning rate is also a decreasing function of n
$0 < \eta(n) < \eta(n-1) <= 1$

Pictorially



Convergence for kohonen not proved except for uni-dimension



Clusters: