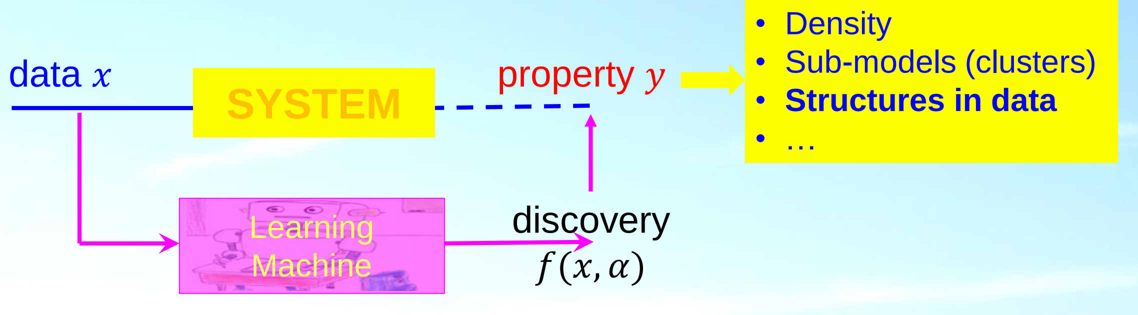
Chapter 16
Unsupervised Learning Neural Networks
(part 1)

Xuegong Zhang
Nov. 25, 2021

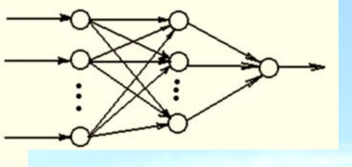
Unsupervised Learning



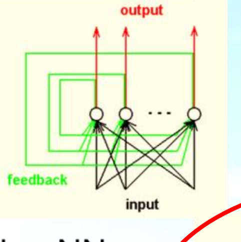
16.1
Competitive Neural Networks and Self-Organizing Map (SOM)

Three Major Types of ANN (in 1980s)

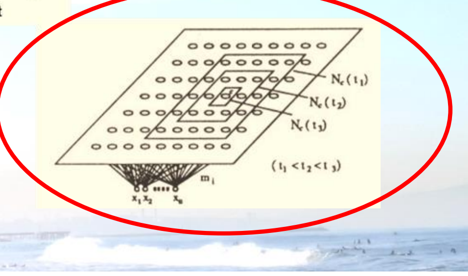
- Feedforward NN
  - Multi-Layer Perceptron



- Feedback NN
  - Hopfield NN



- Competitive Learning NN
  - Self-organizing map

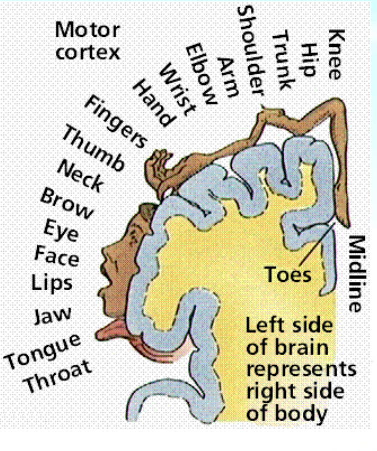


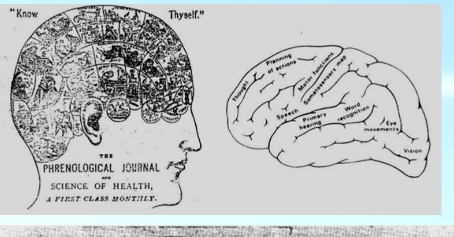
16.1.1
Self-Organizing Map (SOM)

Self-Organizing Maps (SOMs)

Known as SOFM in 1980s, proposed by Prof. Teuvo Kohonen

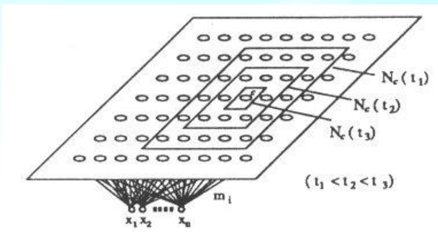
- Self-organizing phenomena in natural neural systems

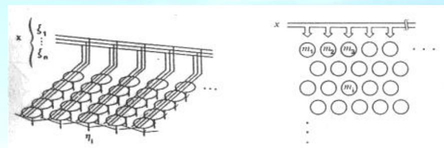




Structure of SOM

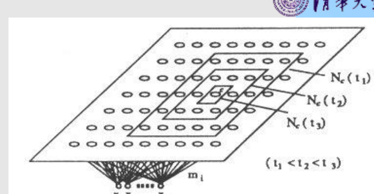
- An array of neurons arranged on a plane
- Inputs connected to all nodes
- Interactions between nodes according to positions





- Computation at neurons:
  - Matching between inputs and weights (distance or inner product)
  - Best matching node: winner

SOM Algorithm



(1) Initialization with small random weights

(2) Time t: input $x(t)$ (in sequential or random order)

(3) Find the winner c :

$$c: \quad \|x(t) - m_c(t)\| = \min_{i \in A}\{\|x(t) - m_i(t)\|\}$$

(if Euclidean distance is use)

(4) Competitive learning

$$m_i(t+1) = m_i(t) + \alpha(t) h_{ci}(t) d[x(t), m_i(t)], \quad \forall i \in A$$

where $\alpha(t)$ is the learning rate, $h_{ci}(t)$ is the neighborhood function between nodes i and c, $d[\cdot,\cdot]$ is the error function, usually the Euclidean distance.

For rectangular neighborhood, the learning rule can be re-written as

$$m_i(t+1) = \begin{cases} m_i(t) + \alpha(t)[x(t) - m_i(t)] & i \in N_c(t). \\ m_i(t) & i \notin N_c(t). \end{cases}$$

where $N_c(t)$ is the set off all nodes within the rectangular neighbor of c.

(5) Terminate if stopping criteria met; Otherwise update $\alpha(t)$ and $N_c(t)$ (decreasing and shrinking with time), set $t = t + 1$ and go to (2).

SOM Property

- After proper training, responses of nodes on inputs become organized, in a way that reflects intrinsic properties in the original data (densities and topological relations).
- A few terms:
  - *Image*: After learning, the winner node for each input sample will tend to be fixed. → The image of the input sample.
  - *Pre-image*: Image of x is node i ←→ x is a pre-image of node i
  - *Image density*: # of pre-images of a node
  - *Image density map*: image densities of all nodes shown as a map

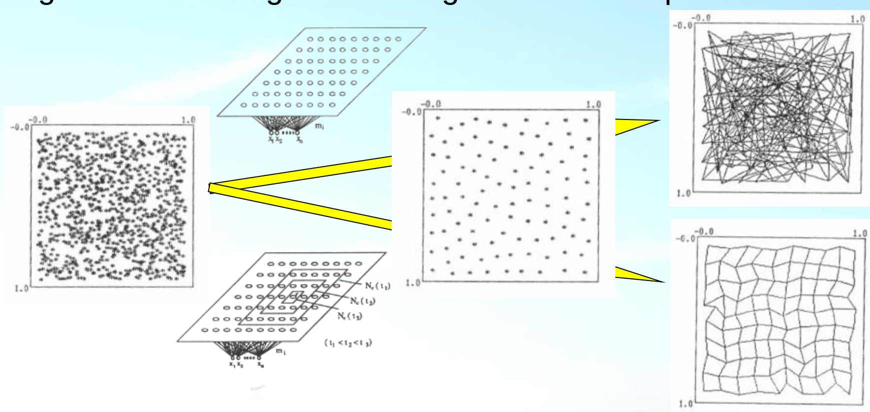
Xuegong Zhang & Yanda Li, Self-organizing map as a new method for clustering and data analysis, *IJCNN*, 1993

The Self-Organizing Property

- After proper training,
  - each sample has only one image node, but a node may have multiple pre-images,
  - samples close in the original space tend to map to same or close-by images, and
  - image density has monotone relation with the density in original space

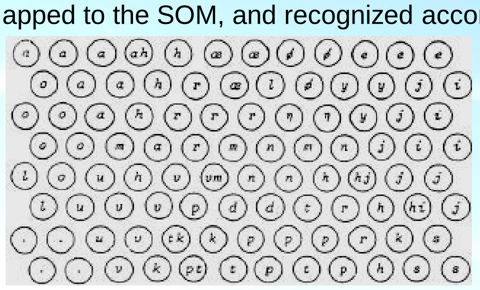
Xuegong Zhang & Yanda Li, Self-organizing map as a new method for clustering and data analysis, *IJCNN*, 1993

- SOM vs. C-means clustering
  - If no neighborhood interaction, SOM is equivalent to C-means or vector quantization (VQ).
  - Self-organization emerges from neighborhood competitive learning



Self-Organized Map

- Phonic typewriter by Kohonen (on Finnish)
  ① Use phonic samples to train SOM, to form the mapping from the phonic space to the SOM grid
  ② Use the known tags of the phonic samples to label the their image nodes
  ③ New samples are mapped to the SOM, and recognized according to the label of their image nodes.



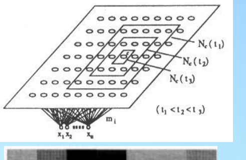
- Learning Vector Quantization (LVQ): VQ to form a set of representative vectors for all samples, and use them to classify.
- Unlike ordinary VQ (C-means), LVQ forms representations in an ordered manner.

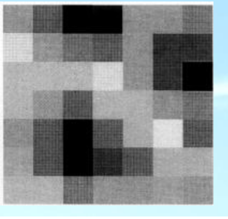
➔ **Unsupervised learning before supervised learning**

16.1.2
SOM Analysis (SOMA)

SOM for unsupervised learning --- SOMA

- Basic idea
  - Using unknown samples to train SOM
  - Compute the density map
  - Clustering on the density map
- Advantage
  - Not relying on shape of data distribution
  - No need to pre-set cluster number
  - Compressed 2D display for interactive clustering





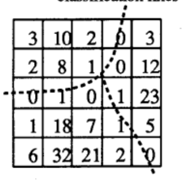


Fig.2 An example of SOM density map

Zhang & Li, Self-organizing map as a new method for clustering and data analysis, *Proc. IJCNN* 1993, pp. 2448-2451

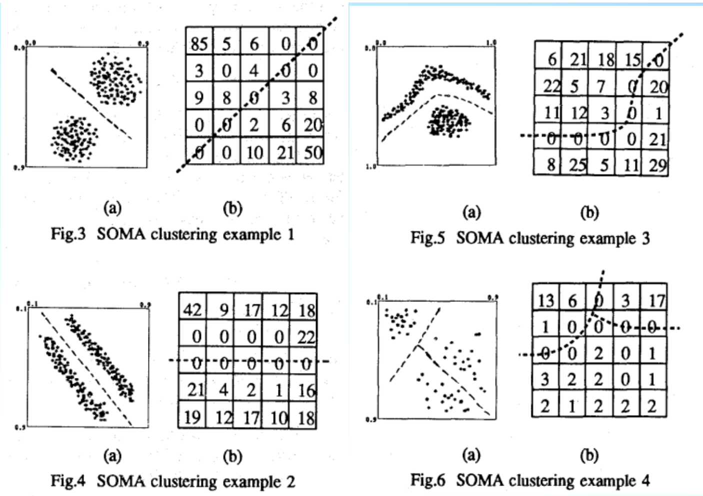


Fig.3 SOMA clustering example 1

Fig.5 SOMA clustering example 3

Fig.4 SOMA clustering example 2

Fig.6 SOMA clustering example 4

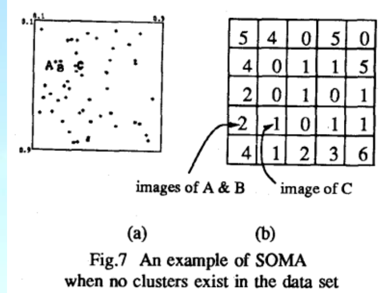
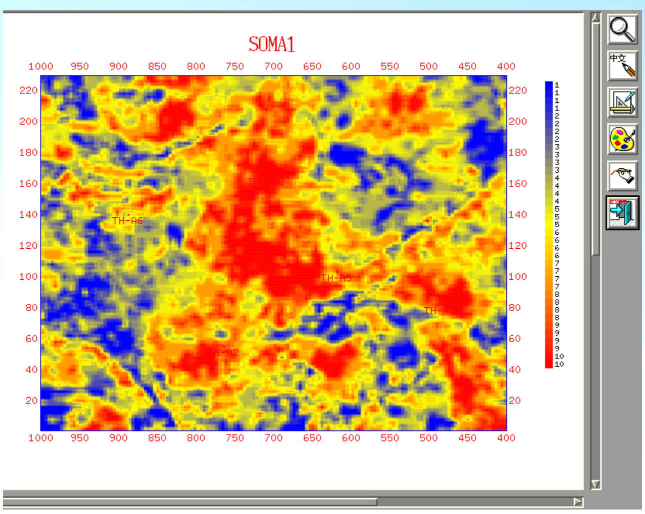


Fig.7 An example of SOMA when no clusters exist in the data set

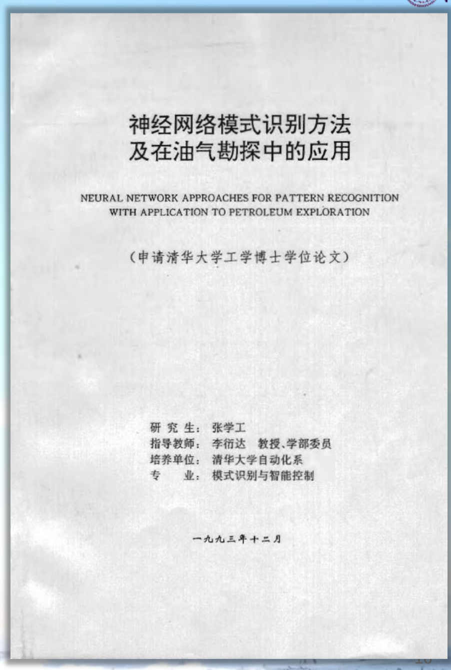
- Open questions
  - Strict analysis on relations of image density with original sample density
  - Strict analysis on topology relation preservation
  - How to do clustering on the density map?

Zhang & Li, Self-organizing map as a new method for clustering and data analysis, Proc. IJCNN 1993, pp. 2448-2451

Application Examples



SOMA for petroleum exploration


神经网络模式识别方法
及在油气勘探中的应用

NEURAL NETWORK APPROACHES FOR PATTERN RECOGNITION
WITH APPLICATION TO PETROLEUM EXPLORATION

（申请清华大学工学博士学位论文）

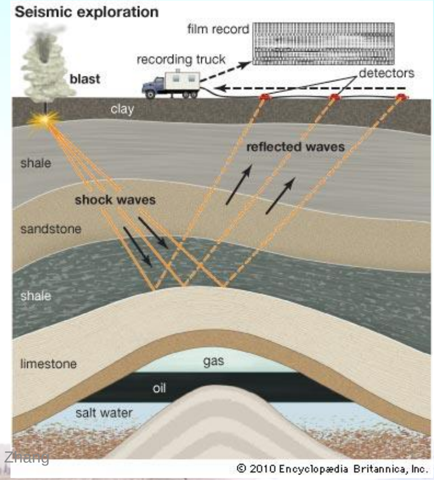
研 究 生：张学工
指导教师：李衍达 教授、学部委员
培养单位：清华大学自动化系
专 业：模式识别与智能控制

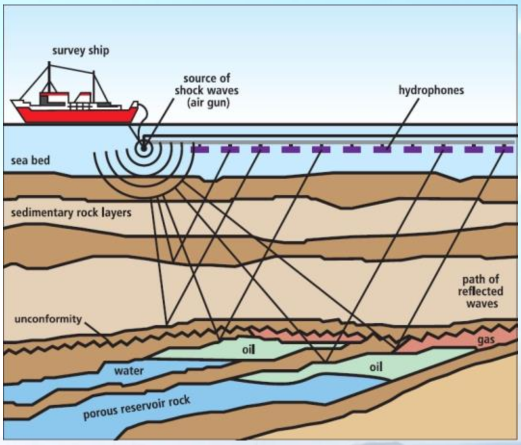
一九九三年十二月

X. Zhang, Tsinghua University PhD Thesis, 1993

清华大学

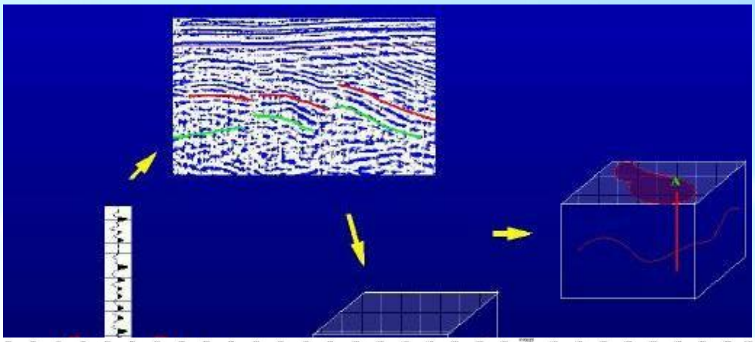
Example: Application of PR in Geophysical Exploration

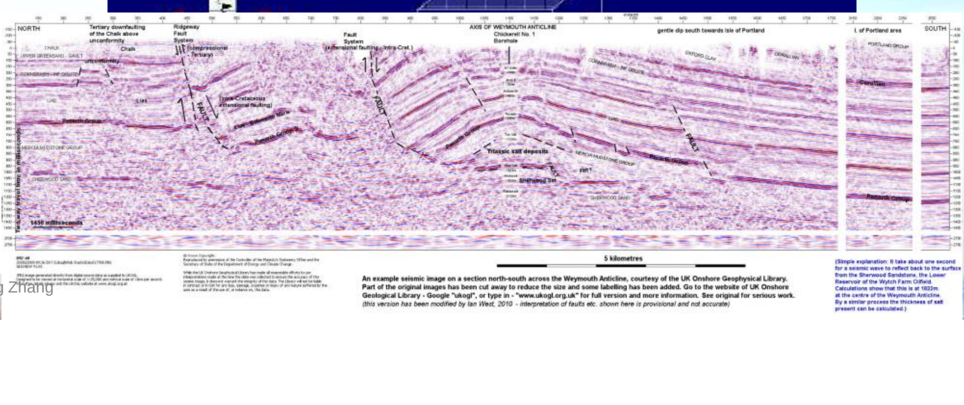




清华大学

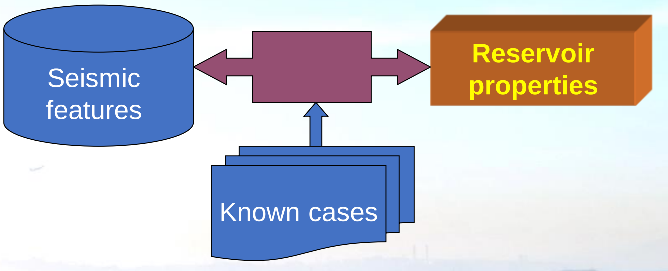
Seismic petroleum exploration

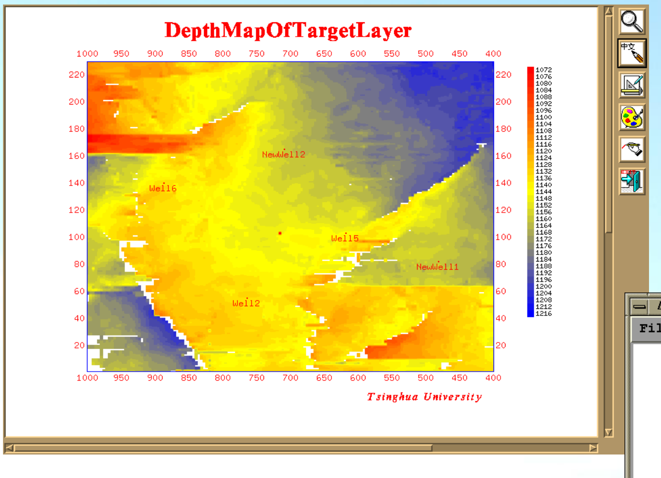


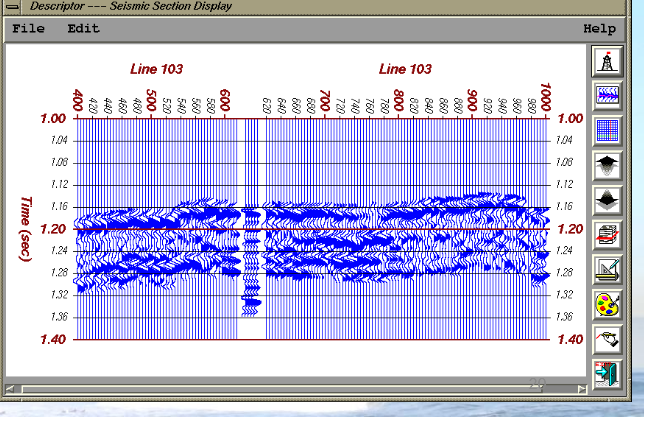


Reservoir analysis with seismic exploration data

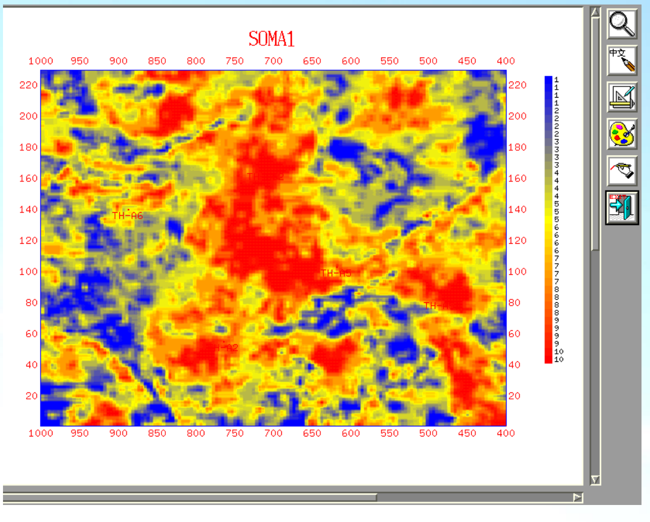
- Arrival time of seismic signals
  —— subsurface structures
- Information in seismic waves
  —— reservoir property information

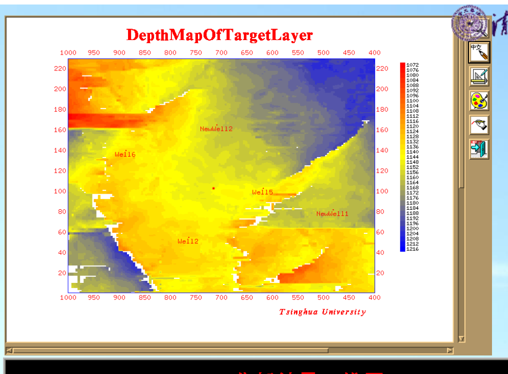


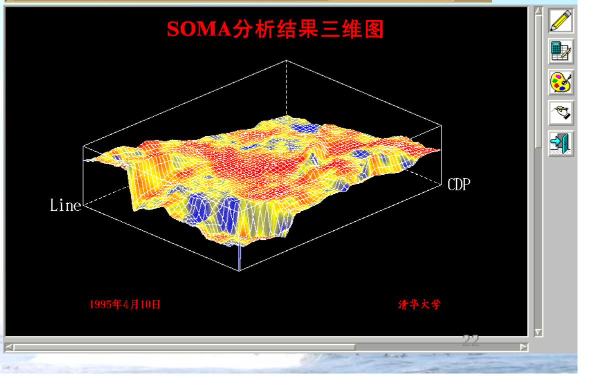




- Supervised approach
  - Insufficient training samples
  - Unbalanced and biased training samples
- Unsupervised approach
  - How to use information from known samples?
- Supervised + unsupervised
  - Unsupervised learning + knowledge-based annotation
  - Augmented training samples with help of unsupervised results
  - Supervised learning for more accurate study

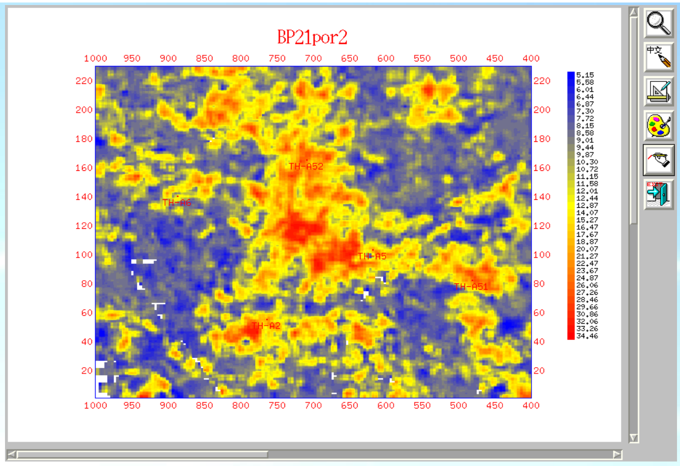






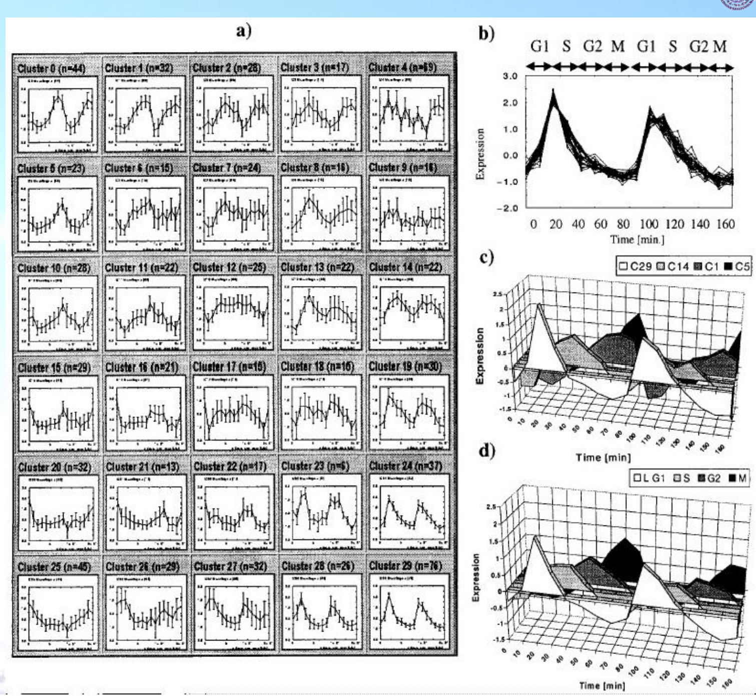
Unsupervised analysis results

Supervised prediction based on augmented training samples



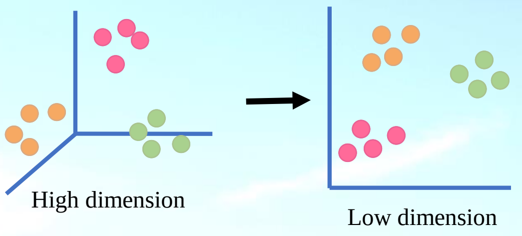
Wells	Depth(m)	Sand Thickness	
		Actual	predicted
TH-A2	1233-1410	35.7m	known
TH-A5	1242-1420	51.9m	known
TH-A6	1239-1416	21.8m	known
TH-A51	1264-1441	42.1m	42.4m
TH-A52	1261-1440	45.6m	43.8m

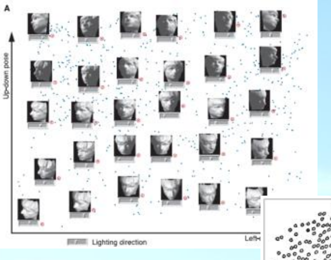
Application Examples

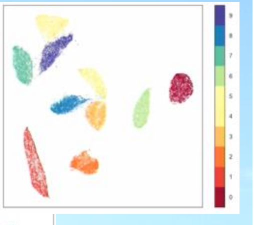


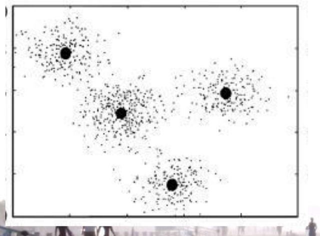
P. Tamayo, …, T.R. Golub, Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation, PNAS, 96:2907-2912, 1999

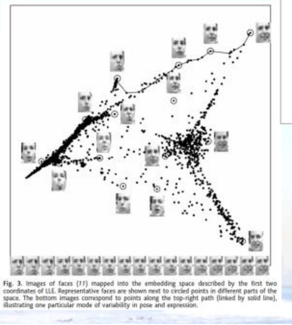
Learning Condensed Representations of Data

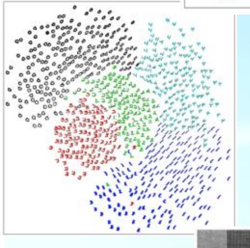


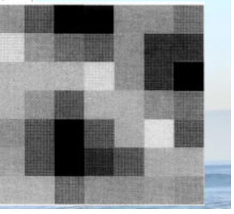












Representation is all they do.


See you next week
for
Hopfield Net, RBM and Deep Learning
Xuegong Zhang
27