



Plausible Neural Networks ©

- An intelligent self-organized network system

Yuan Yan Chen, PhD

PNN Technologies
www.pnntech.com



Overview

- *Plausible Neural Network* (PNN) is a new computational model to mimic human learning and cognition.
- PNN performs clustering, classification, function approximation, associate memory and plausible inference with one of the fastest learning algorithm up-to-date.
- PNN connects information, fuzzy set, statistics, neural networks and statistical mechanic theories for the design of a *self-organized intelligent computing machine*.



Intelligent computing machine

- Response to an input with a proper output.
(Turing test)
- Self organize with incoming data information and extract data patterns.
- Link the patterns with existing knowledge and create inference rules.



Philosophical viewpoint and mathematical models

- “Probability theory as logic” — *E. T. Jaynes*
- Recognition is a *belief logic*.
- Intelligent computing machine is a *statistical inference machine*.

Chen, Y. (1993). Bernoulli trials: from a fuzzy measure point of view. *J. Math. Anal. Appl.* **175**, 392-404.

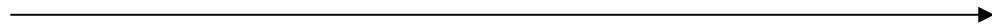
Chen, Y. (1995). Statistical inference based on the possibility and belief measures. *Trans. Amer. Math. Soc.* **347**, 1855-1863.

Chen, Y. and Chen, J. (2004). Neural networks and belief logic. *Proceeding of the Fourth International Conference on Hybrids Intelligent Systems (HIS'04)*, IEEE, 460-461.



Signal logic and computation device

complexity



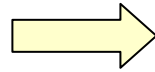
Computation device	Quantum	Analog	Digital
Signal	$ 0\rangle$ $ 1\rangle$ superposition	$[0, 1]$	0 1
Logic	Quantum logic	Fuzzy	Boolean

efficiency

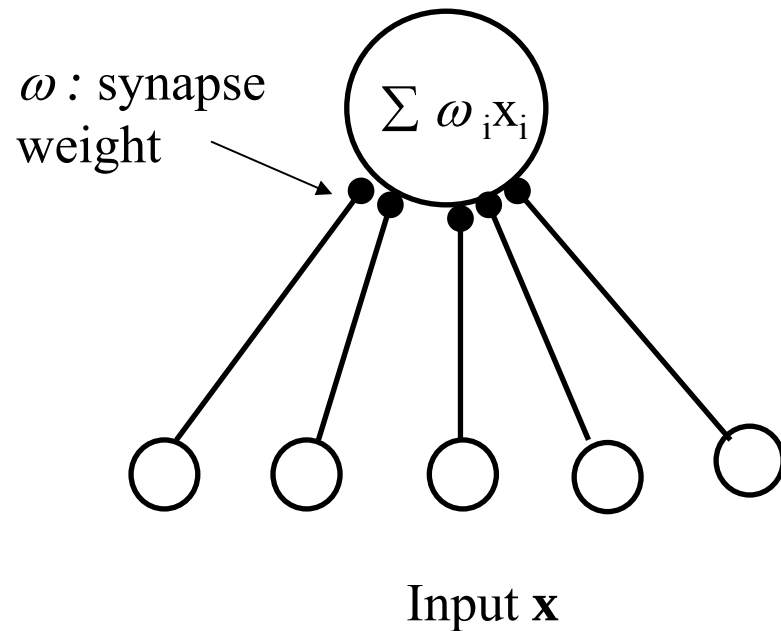


Neural networks model

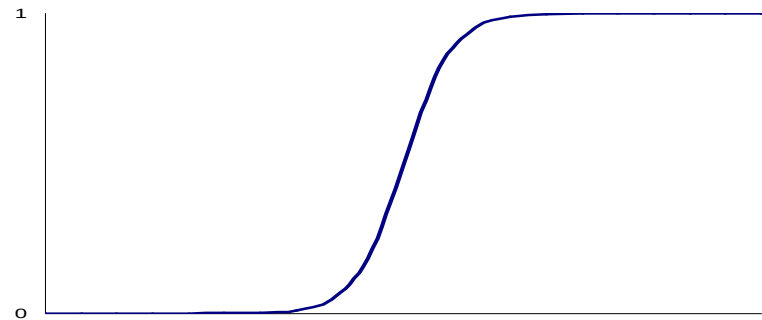
Input signal $x_i \in [0,1]$,
 $x_i = 0$, no input stimulus



Output signal
 $y = s(\sum \omega_i x_i)$



nonlinear signal function





Complex systems and learning models

- Complex systems

Dynamics systems with many basin of attraction, typically with subsystems and components of different interaction strength.

- Learning

Computing with weight configuration (gradient descent) – most ANN, e.g. backpropagation.

Computing with state configuration (attractor) – Boltzmann machine, Hopfield network, PNN.



Plausible neural networks model

- Neuron signals for a pair of connected neurons can be considered as the *generalized Bernoulli trials* of two dependent variables.
- Synapse weight is represented by *mutual information content*.
- Organization and activation is governed by *winner-take-all* (WTA) circuit.



Entropy and organization

- Shannon entropy

$$H(X) = - \sum p(x) \ln(p(x))$$

- Random – Entropy maximum



Organization

- Order – Entropy minimum
- Mutual information

$$I(X, Y) = \sum \sum p(x,y) \ln(p(x,y) / p(x) p(y))$$



Generalized Bernoulli trials

- $p(x | \theta) = c(\theta) \theta^x (1-\theta)^{1-x}, x \in [0,1]$
- pure state: 0 or 1 $\uparrow \downarrow$
- superposition: $x \in (0,1)$ $\nearrow$
- Joint probability

$$p(x_1, x_2, \dots, x_n | \theta) \propto \theta^{\sum x_i} (1-\theta)^{n - \sum x_i}$$

- Maximum likelihood estimate: $\hat{\theta} \approx \sum x_i / n$



Mutual information content

Weight connection

$$\omega = \ln (p(x,y) / p(x) p(y))$$

$\omega > 0$, positively associated (excitatory).

$\omega < 0$, negatively associated (inhibitory).

$\omega = 0$, statistically independent.

Maxim likelihood estimate

$$\hat{\omega} \approx \ln (n \Sigma x_i y_i / \Sigma x_i \Sigma y_i)$$

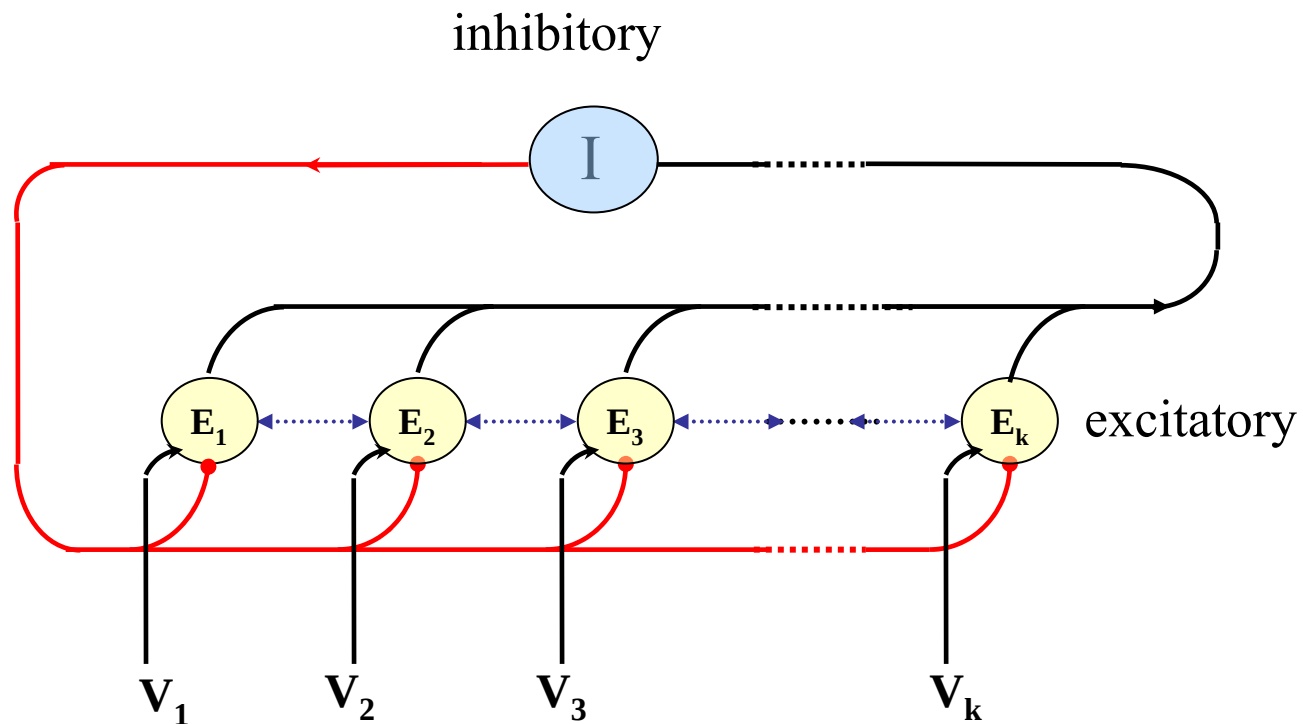


Neuron channels and Transistors

- Both depend on Boltzmann distributions.
- Current flow in transistors is exponentially dependent on voltage of terminals in the subthreshold regime of operation.
- Ionic conductance is exponentially dependent on voltage difference across neuron membrane.

Winner-Take-All (WTA) circuits

Fukui, and Tanaka, (1997). A simple neural network exhibiting selective activation of neural ensembles: from winner-take-all to winners-share-all. *Neural Computation*, **9**, 77-97.





WTA circuit

- Competition – winner take all or no winner

$$I_i = 0, \text{ if } I_i \ll \max I_j, \text{ or } I_i \approx I_j \quad \forall i, j.$$

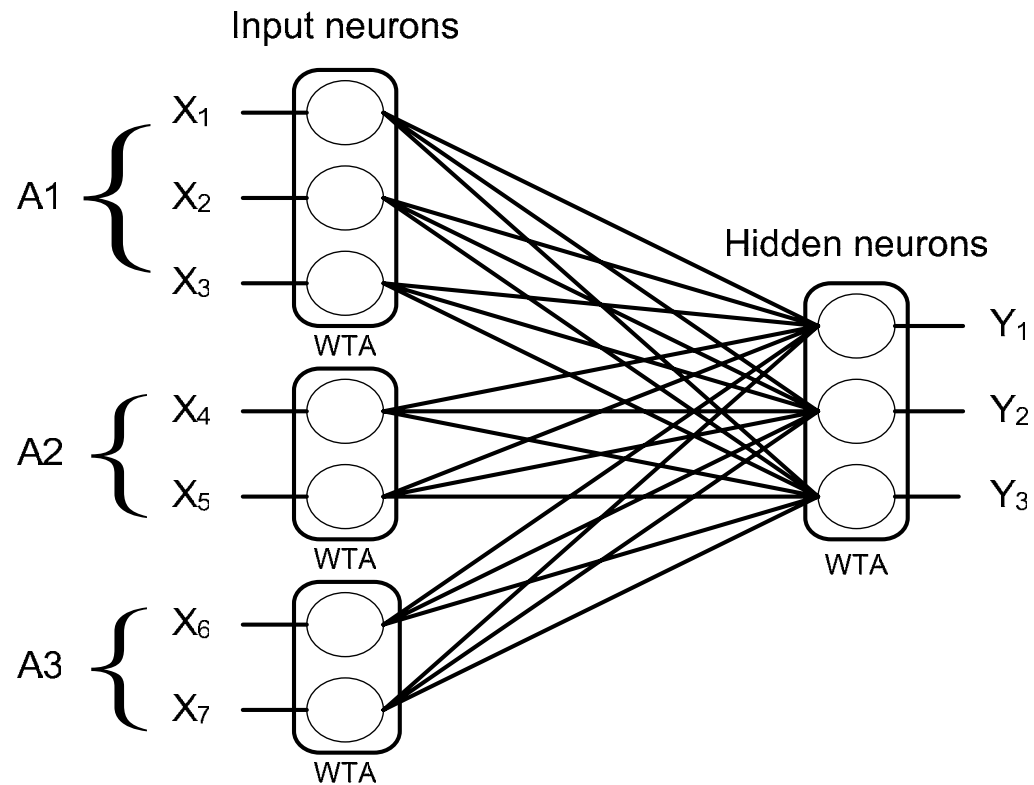
- Boltzmann distribution

$$I_i = \exp(\beta V_i) / \sum_{j=1, \dots, n} \exp(\beta V_j),$$

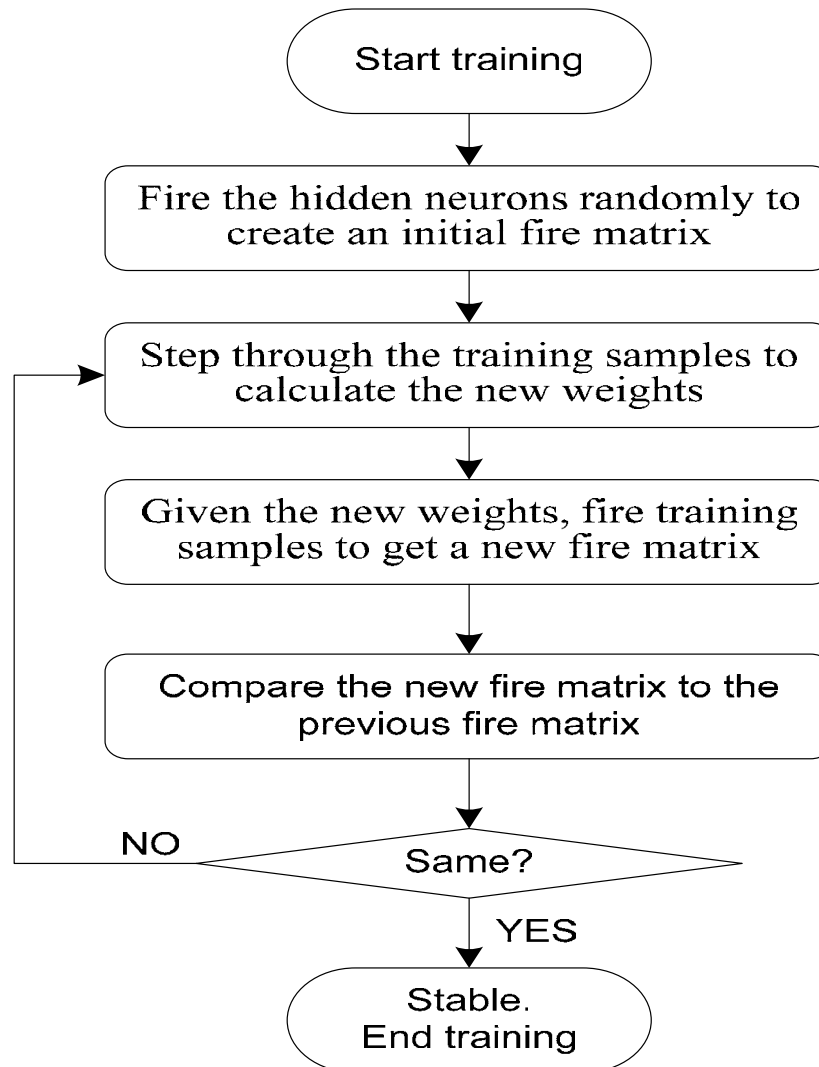
Maass, W. (2000). On the computational power with winner-take-all, *Neural Computation*, **12(11)**, 2519-2536.

Yuille and Geiger (2003) Winner-take-all networks. Arbib edited *The Handbook of Brain Theory and Neural Networks*. 1228-1231.

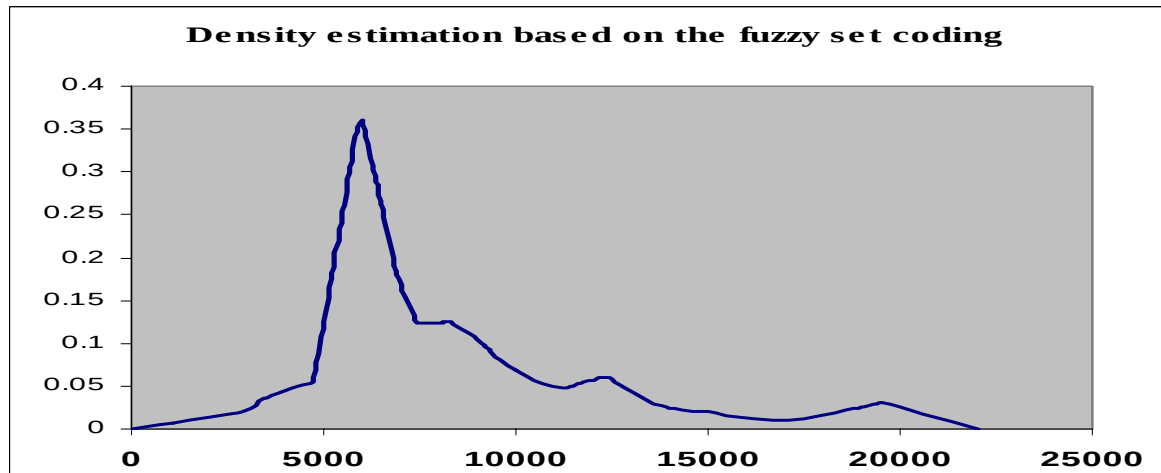
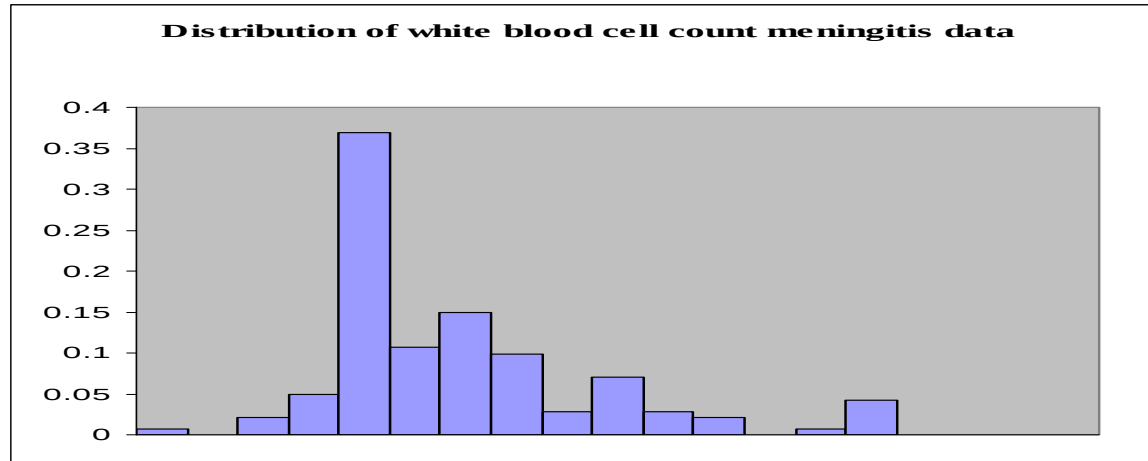
PNN architecture for unsupervised learning



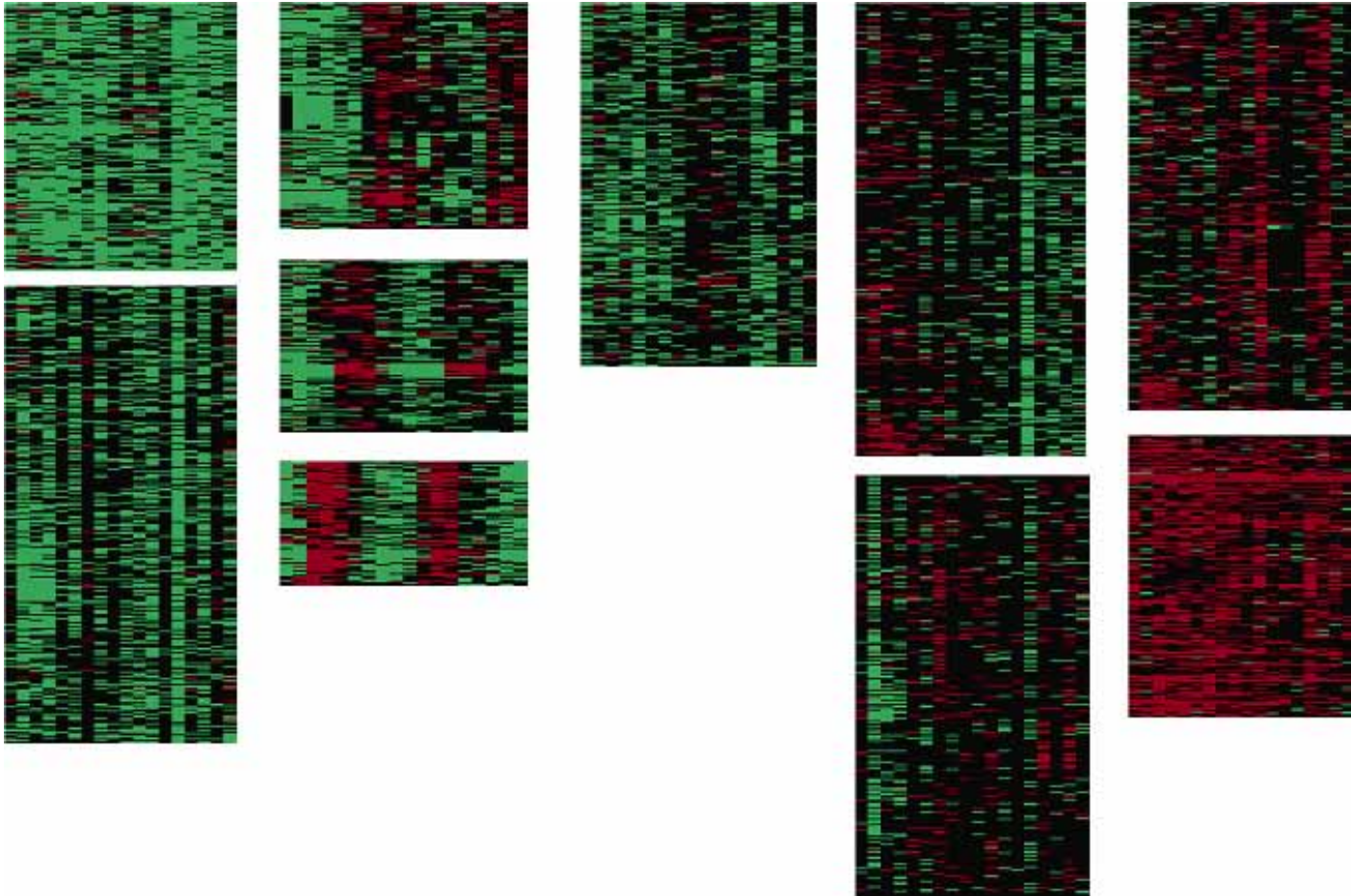
PNN learning algorithm



Probability function estimation



Clusters of gene expressions

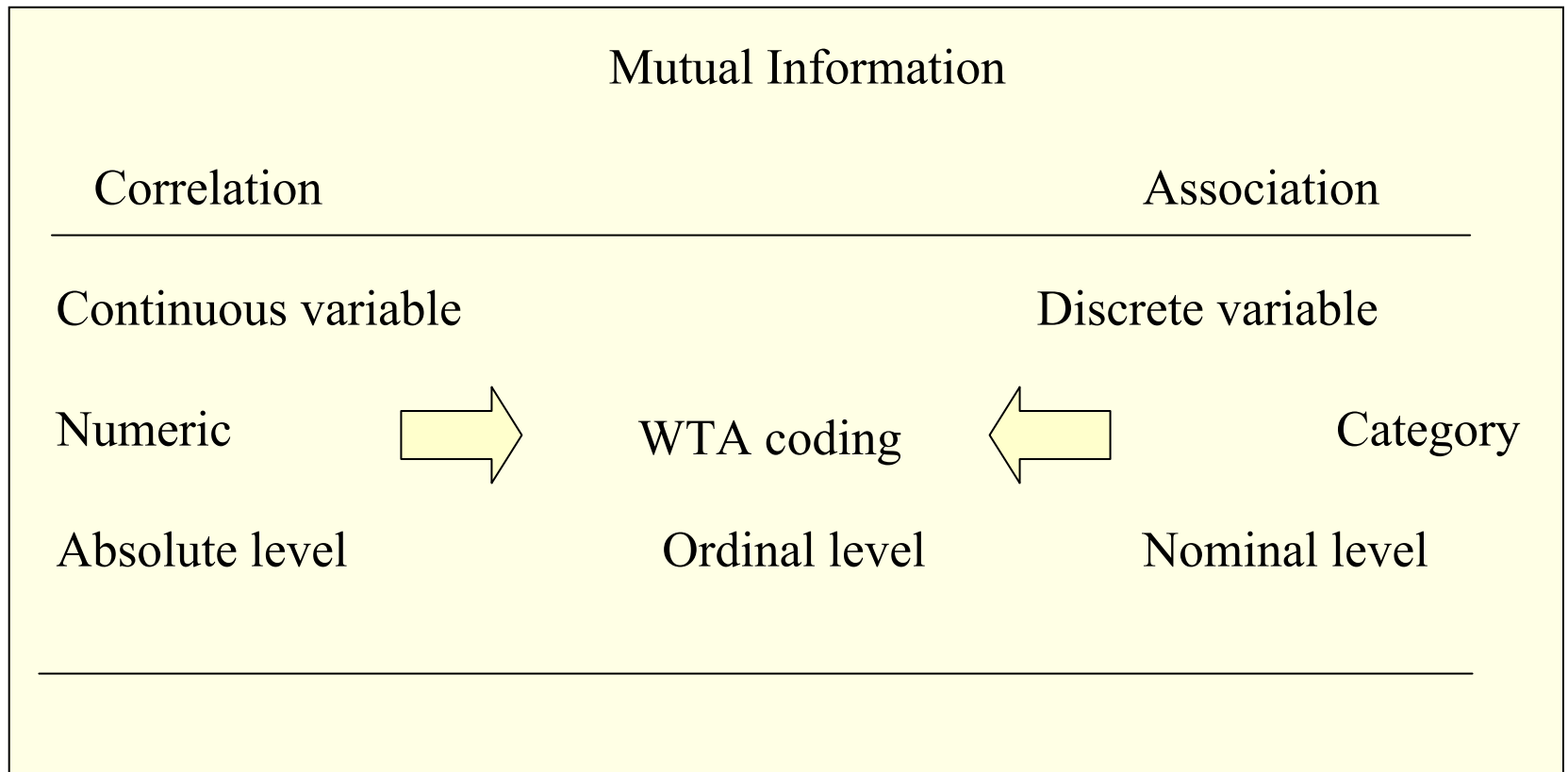


mitotic cell division cycle of yeast *S. cerevisiae* (2,467 genes)

Spellman *et al.*, (1998).

Measure of relationship

$$I(X, Y) = \sum \sum p(x,y) \ln (p(x,y) / p(x) p(y))$$



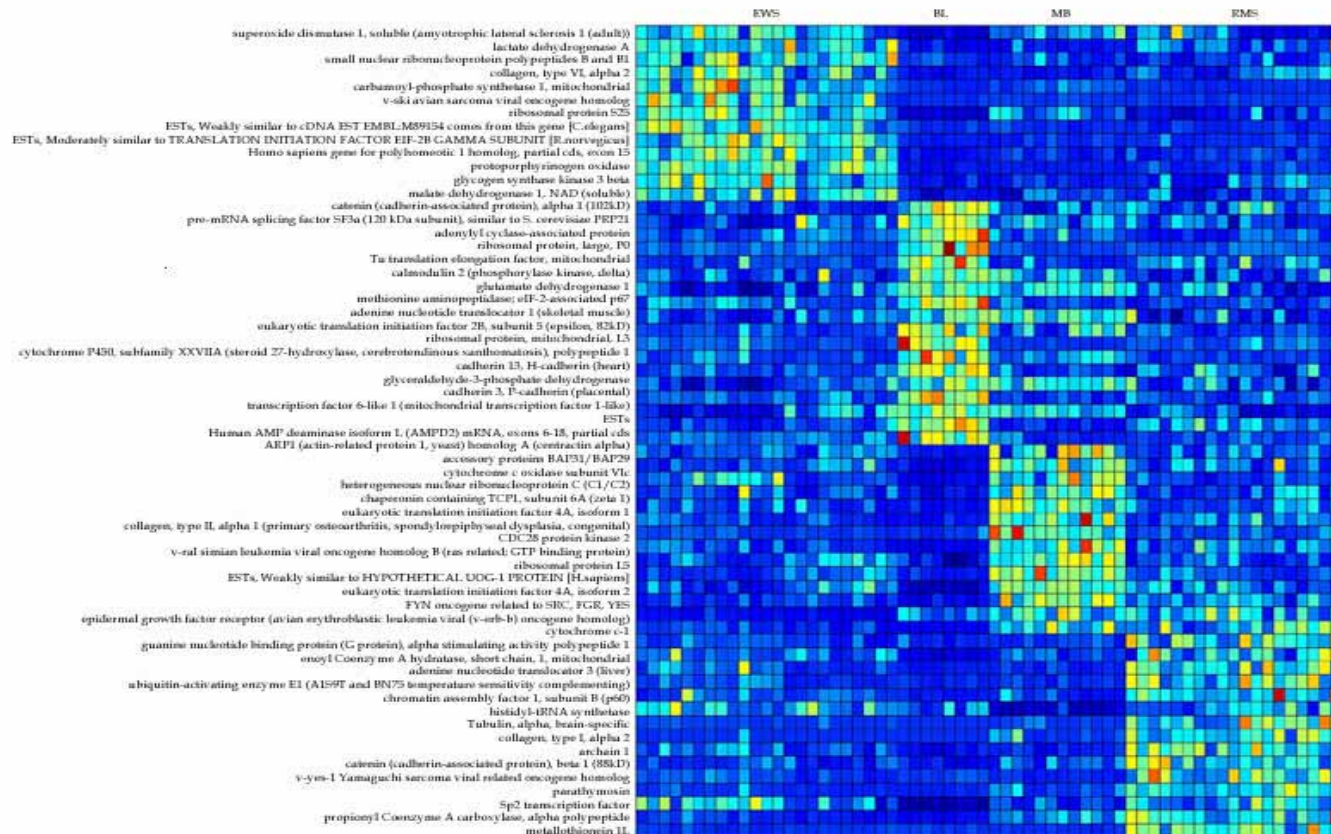


Features selection based on mutual information

Data: 43k gene expressions, 3 experimental conditions

GeneName	E1313	E1315	E1317	Information
4921530F17	1.2	6.6	-6.6	0.464
Ott	1.4	5.8	-6.1	0.458
TC1048789	-2.7	-3.4	5.4	0.444
L04961	0.3	6.6	-6.6	0.442
AK004663	-3.7	-5.0	3.1	0.440
Gpnmb	-2.3	3.8	-6.0	0.438
NAP106219-1	1.3	5.7	-4.9	0.437
Pro25G	-5.7	-4.1	-6.3	0.005
4932415G12Rik	-0.7	-0.8	-0.2	0.005
NAP104295-1	-0.7	0.01	-0.2	0.005
BC012438	-0.16	-0.12	0.06	0.00013
BC002100	0.1	0.14	0.2	0.00013

Cluster Analysis of SRBCT data



Dataset: 60 samples with 4 classes and 6,567 gene expressions.

Khan, et. al. (2001), Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, 7, 673-679.



Performance of PNN for classification

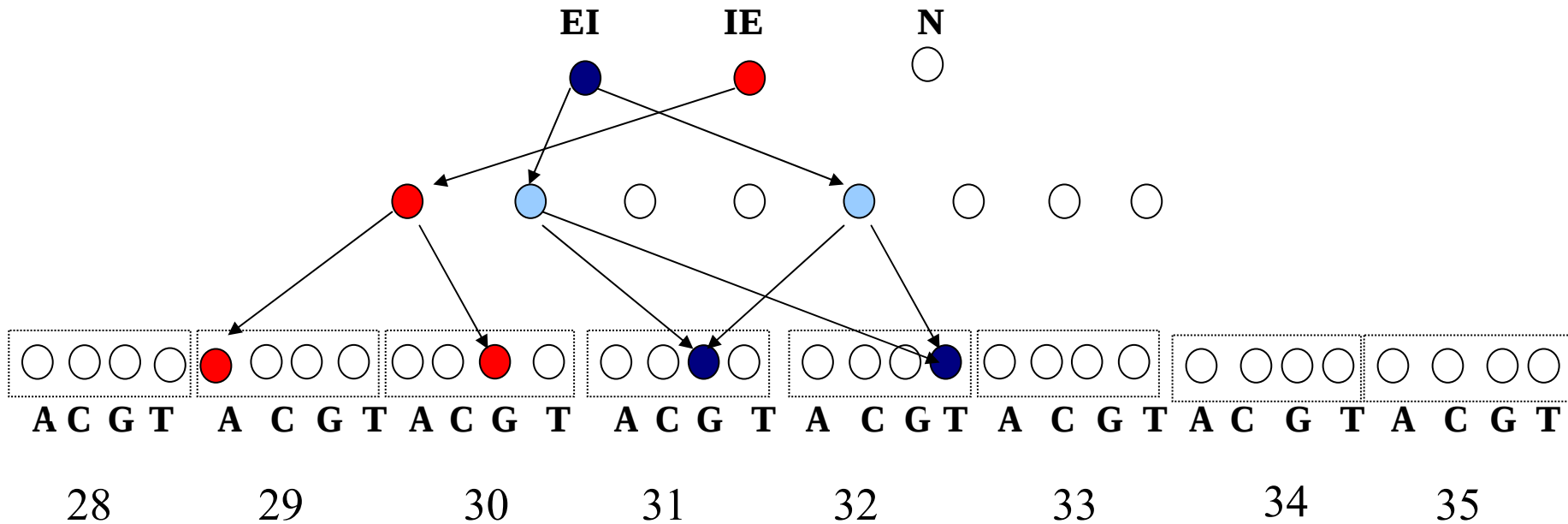
Data set	PNN	Best Alternative Methods
Hepatitis ¹	83.2%	83% (Assitant-86)
Glass ¹	85.5%	83% (NN)
Heart-disease (Cleveland) ¹	81.0%	77% (NT Growth)
Lung cancer ¹	81.3%	77% (KNN)
Pima Indians (diabetes) ¹	79.1%	76% (ADAP)
Primary tumors ¹	50.8%	48% (Naïve Bayes)
Wine ¹	100.0%	100% (RDA)
Leukemia ²	100.0%	98-100% (SVM)
Breast cancer ³	97.5%	97% (SVM)

¹UCI Machine Learning Repository

²Golub et al (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**: 531-537.

³Hedenfalk et al (2001) Gene Expression Profiles in Hereditary Breast Cancer. *New England Journal of Medicine*, **244**:539-548.

Knowledge extraction (Splice junction data)



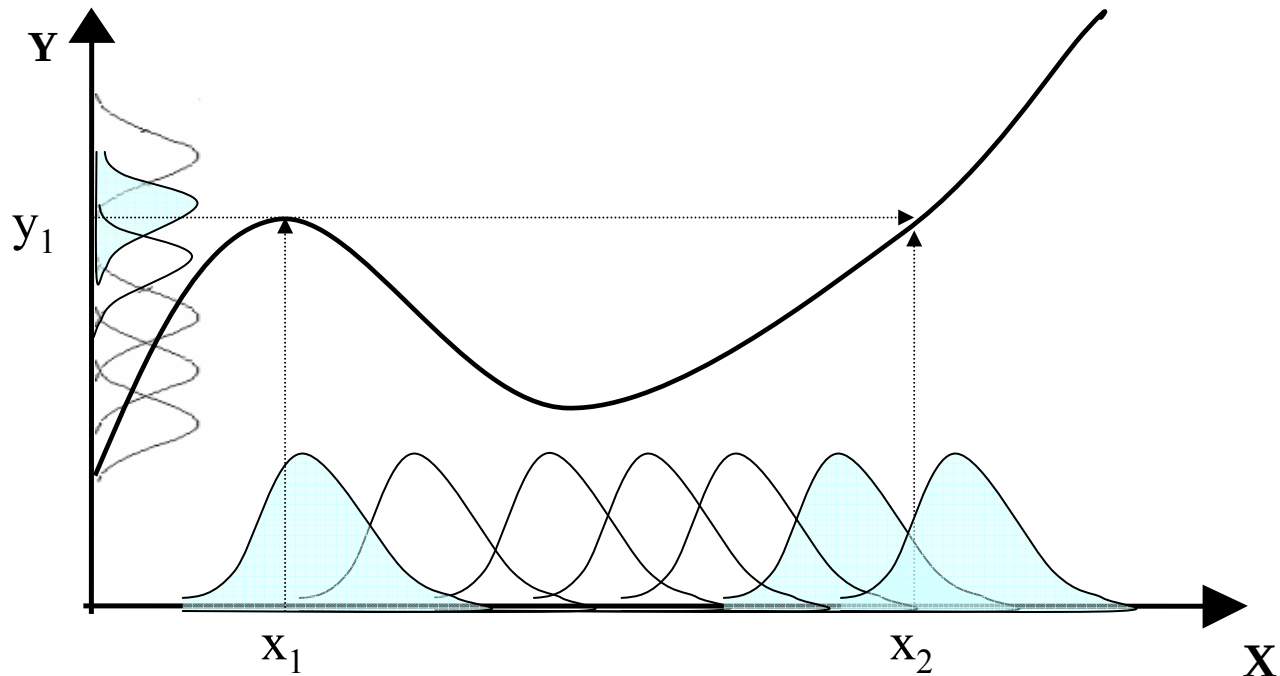
Acceptors (EI) -----GT-----

Donors (IE) -----AG-----

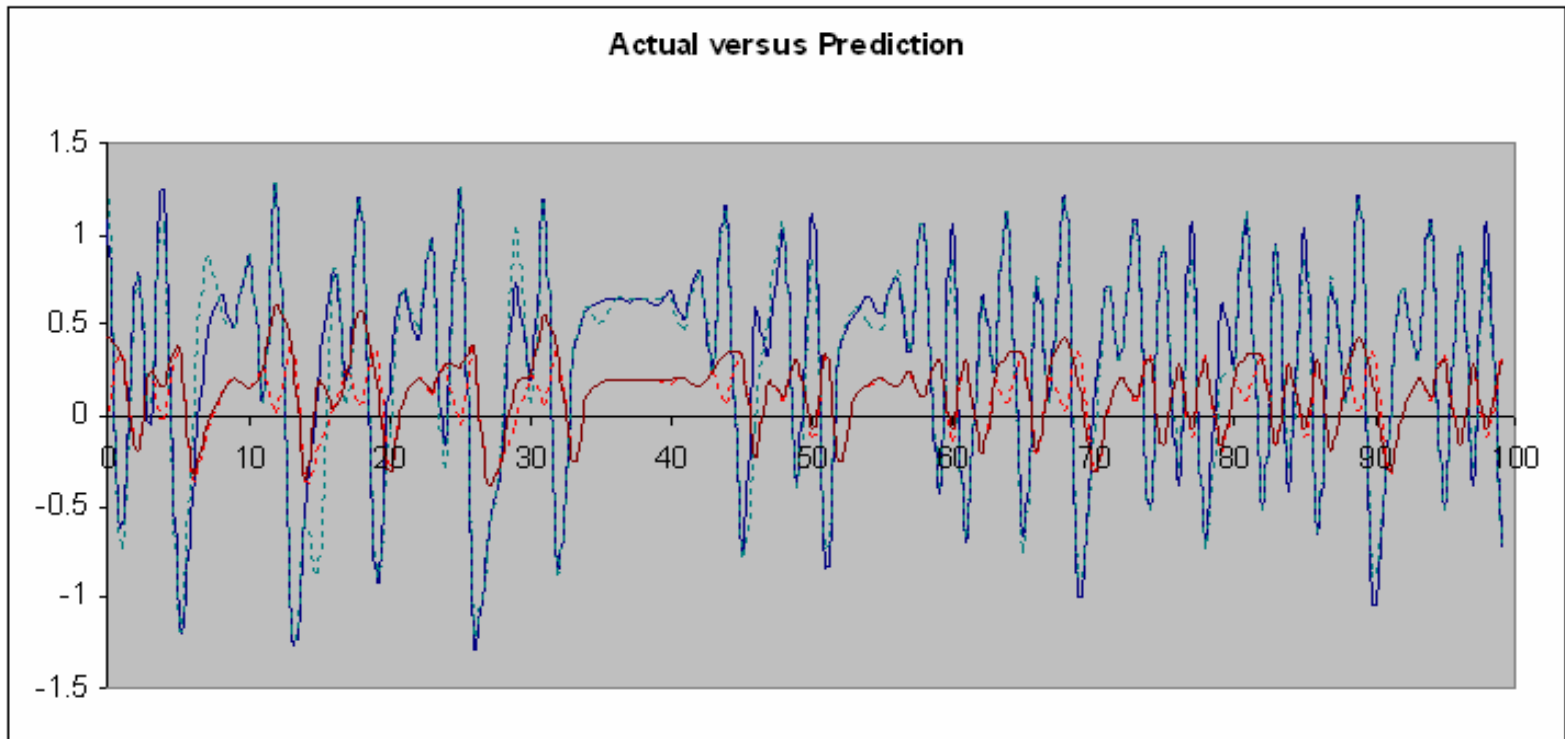
Function/relation estimation

Kosko B. (1994) Fuzzy Systems as Universal Approximators, *IEEE Transactions on Computers*, vol. **43**, no. **11**, 1329-1333.

Function : $f(x_1) = y_1$ Relation: $f^{-1}(y_1) = \{x_1, x_2\}$



Nonlinear time series - Hénon map (chaotic system)



$$x_{t+1} = 1 - 1.4x_t^2 + y_t, \quad y_{t+1} = .3 x_t$$

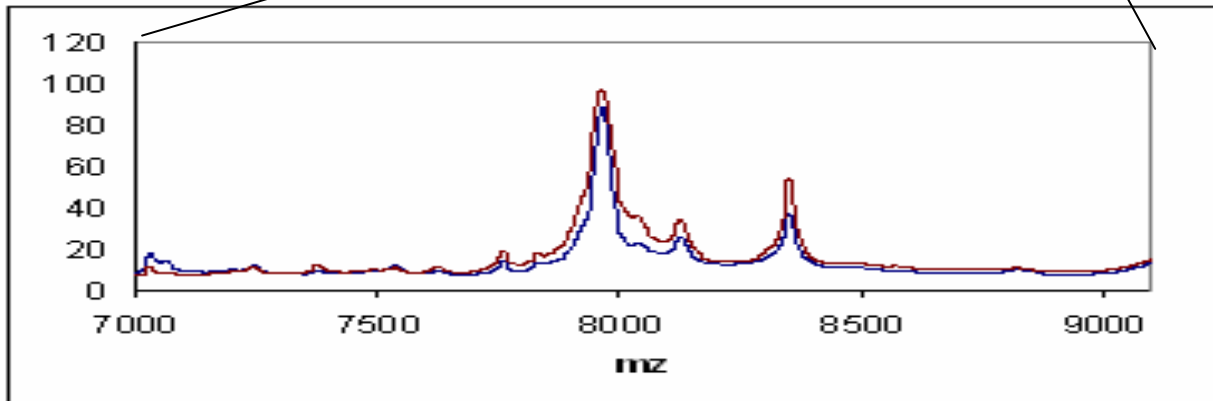
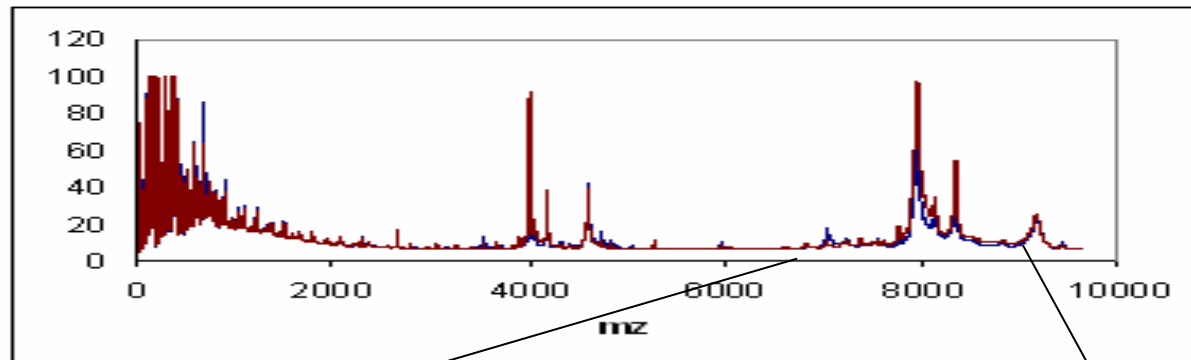
..... Actual
 — Prediction

Proteomic pattern recognition

Ovarian Cancer -161 Samples, Control- 92 Samples, 15k M/Z spectrum

Training: 131 samples, Test: 122 samples, Accuracy:98-100%

Petricoin et al.(2002) Use of proteomic patterns in serum to identify ovarian cancer. Lancet; **359**:572–7.





Prospects of plausible neural networks

- Large-scale data analysis and pattern recognition systems.
- Intelligent computing machine.
- “... spelled out some of the organic reasons why our species generate and maintains beliefs of all kinds. With that knowledge, perhaps all of those who possess belief at odds with the beliefs of others will, upon quiet consideration, become more tolerant.”

–*The Social Brain*, Gazzaniga (1985)