

ISNN 2008, Beijing China


Turning Points on MLP's Error Surface

Hung-Han Chen, Ph.D.

Graphion LLC

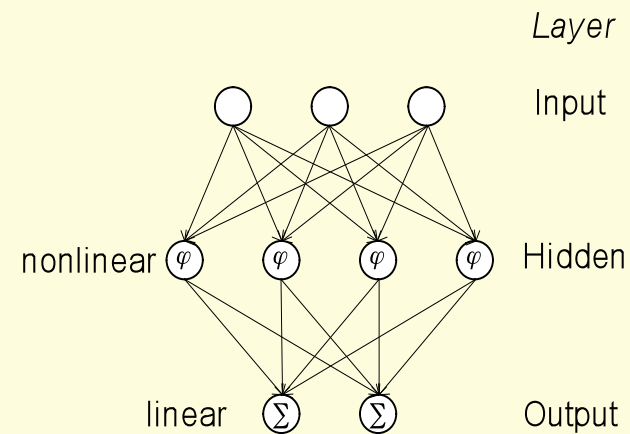


Outline

- MLP & BackPropagation
 - **Error Surface & Local Minima**
 - Search Method: **Retreat and Turn**
 - Examples
 - Conclusions
 - Q & A
- 

MLP

- Connectionism
- Feedforward Artificial Neural Networks
- Nonlinear activation function for hidden layer
- Backpropagation
- Supervised Learning
- Least Mean Square Algorithm
- Gradient Descents





MLP

3 Major Concerns

- Scale Up Problem
- Local Minima
- Generalization

Solutions

- **NeuroSequences**
(FFC conference, 2008)
 - **Retreat and Turn**
(ISNN, 2008)
 - **Monitoring Free Parameters**
(???, 2009)
- 

BackPropagation

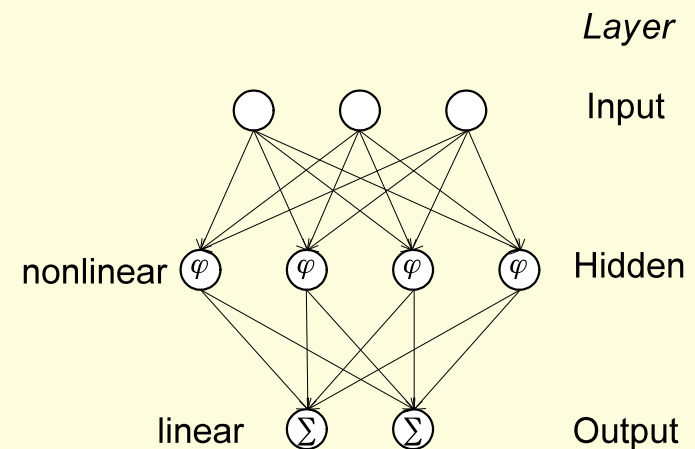
$$w(j,i) \leftarrow w(j,i) + \eta \cdot \frac{-\partial E}{\partial w(j,i)}$$

$$\frac{\partial E}{\partial w(j,i)} = -\delta(j) \cdot O(i)$$

$$\delta(j) = \frac{-\partial E}{\partial net(j)}$$

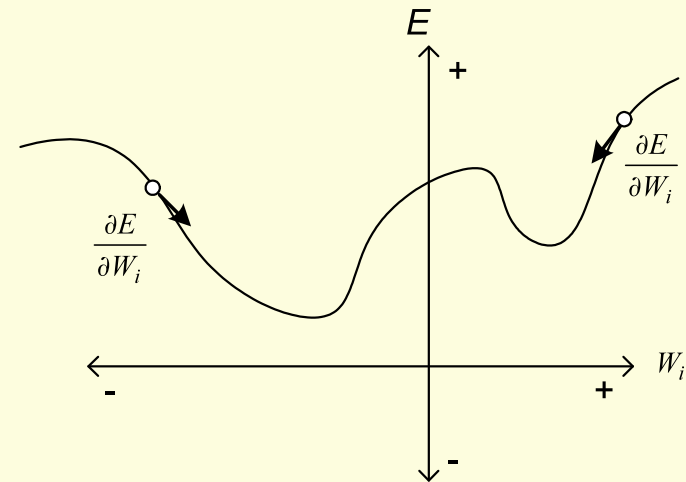
$$\delta(n) = f'(net(n)) \cdot (T(n) - O(n))$$

$$\delta(j) = f'(net(j)) \cdot \sum_n \delta(n) w(n,j) = O(j)(1 - O(j)) \cdot \sum_n \delta(n) w(n,j)$$



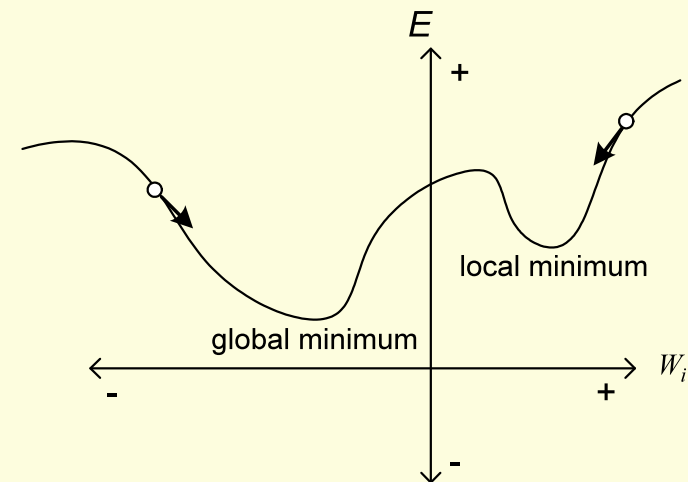
Gradient & Error Surface

- $E \geq 0$
- Gradient represents slope
- Multiple descents
- **Dimension**: Number of weights
- Error Surface depends on **Structure, Data, Functions** [5]
- Troughs [4], Ravines, Starfish [5]



Hill-Climbing Technologies

- Optimization Algorithm for local search
- Iteratively makes small changes to the solution
- No knowledge of global environment
- **Problems:**
 - Local Maxima (**Local Minima**)
 - Ridges (**Troughs, Ravines**)
 - Plateaus**






Search Algorithms

Hill Climbing

- Simple Hill Climbing
- Steepest Ascent
- Random-restart

Backpropagation

- Basic Line search
 - Scan through neighborhood
 - Random learning factor
- 



Local Minima

Two Types of Local Minima


- (1) Limitation of Search Algorithm
- (2) Minima on Topography of Error Surface

For (2),
$$\sum_i \eta \cdot \frac{-\partial E}{\partial W_i} = 0$$

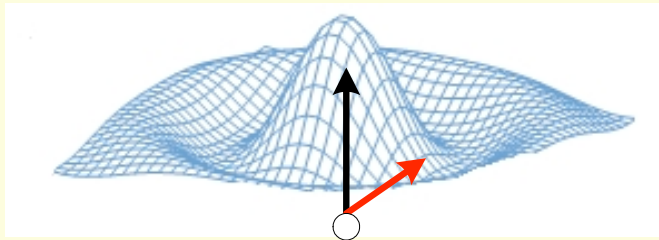
Remedies

- (1) Effective and Meaningful Search Algorithm
- (2) Component Removal to Reshape the Topography

Goal:
$$\sum_j \eta \cdot \frac{-\partial E}{\partial W_j} \neq 0$$

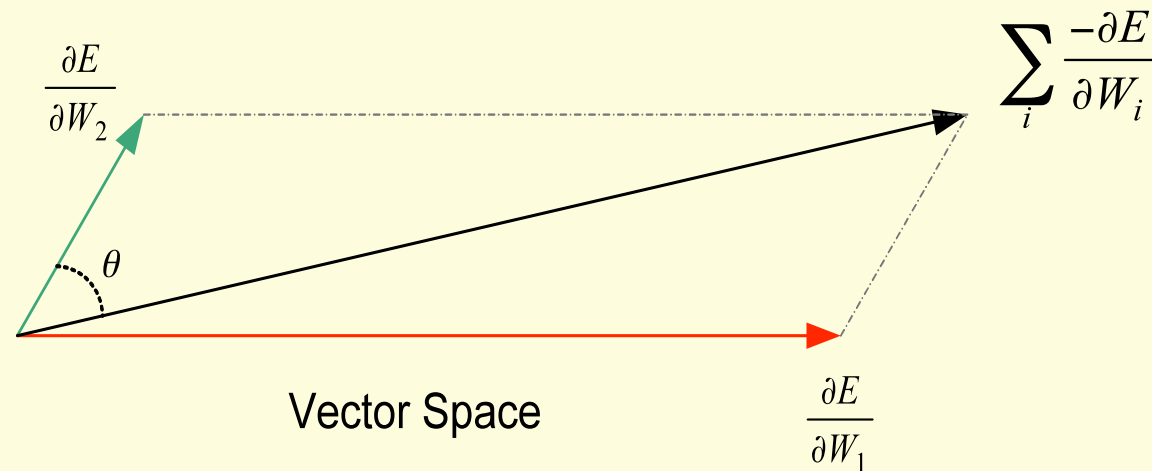


Retreat & Turn



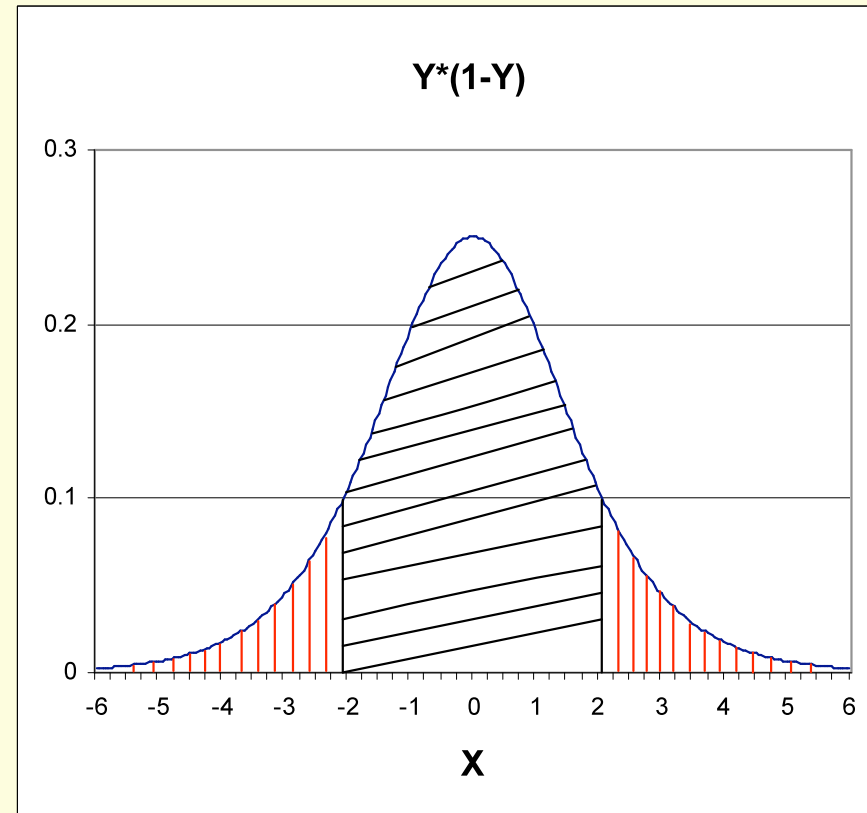
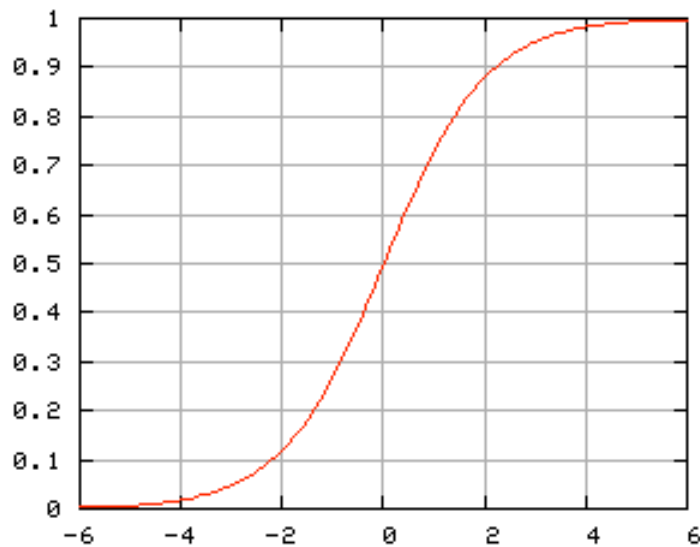
山不轉路轉

If we can't move the mountain, at least we can make the road turn



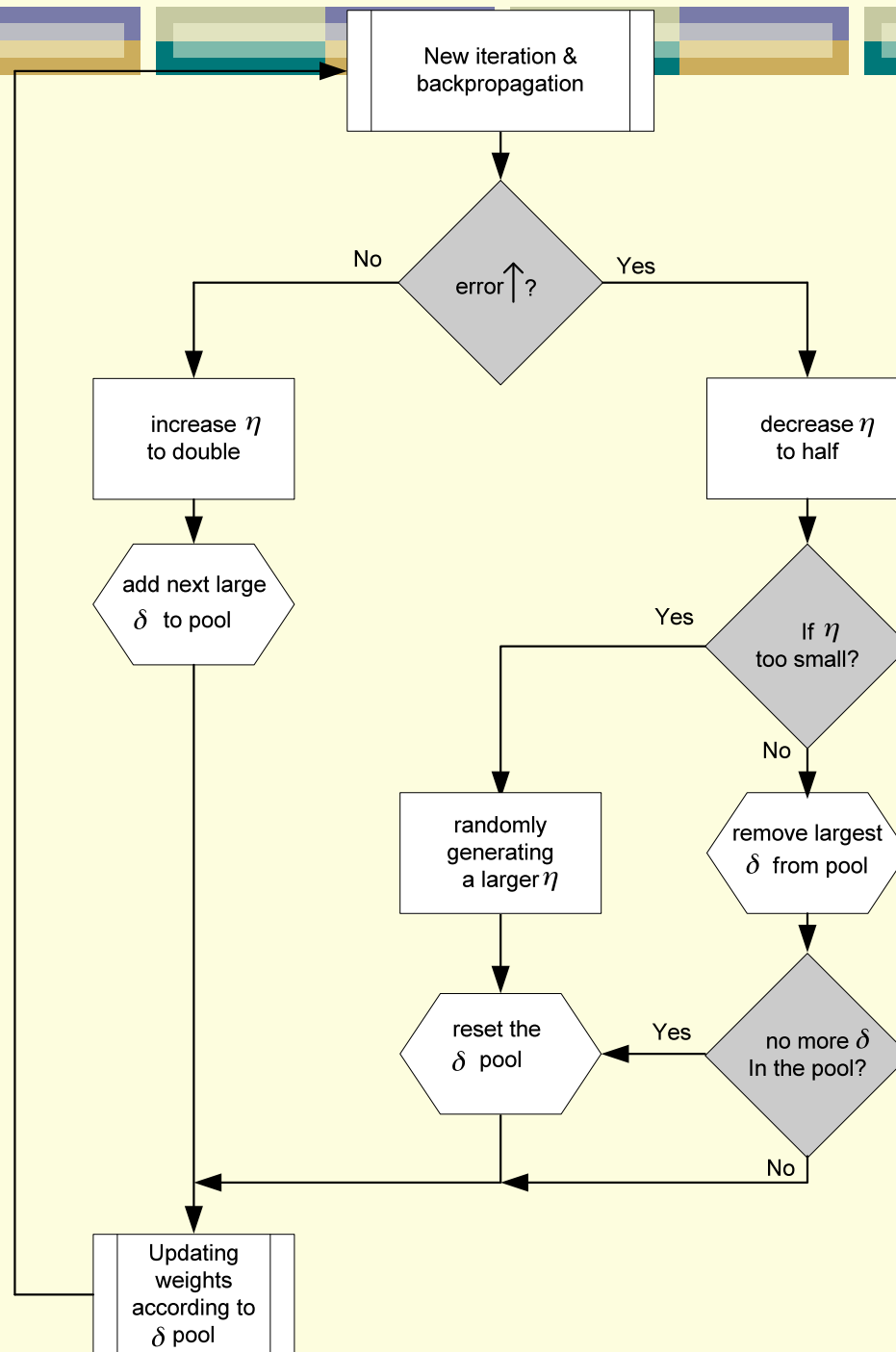
Neuron's Firing Status

- Sigmoid function
- Linear part
- Saturation part




$$\frac{\partial E}{\partial w(j,i)} \cong \delta(j) \cong O(j) * (1 - O(j))$$

Turning Mechanism





Examples

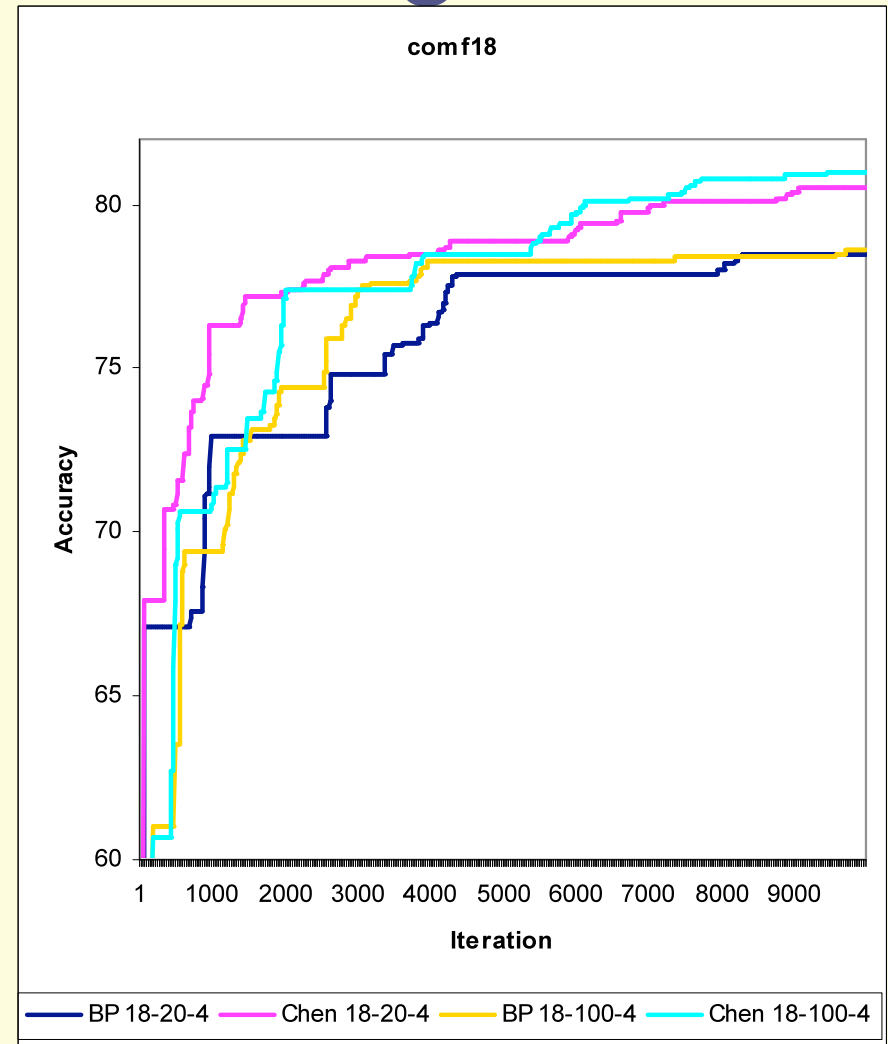
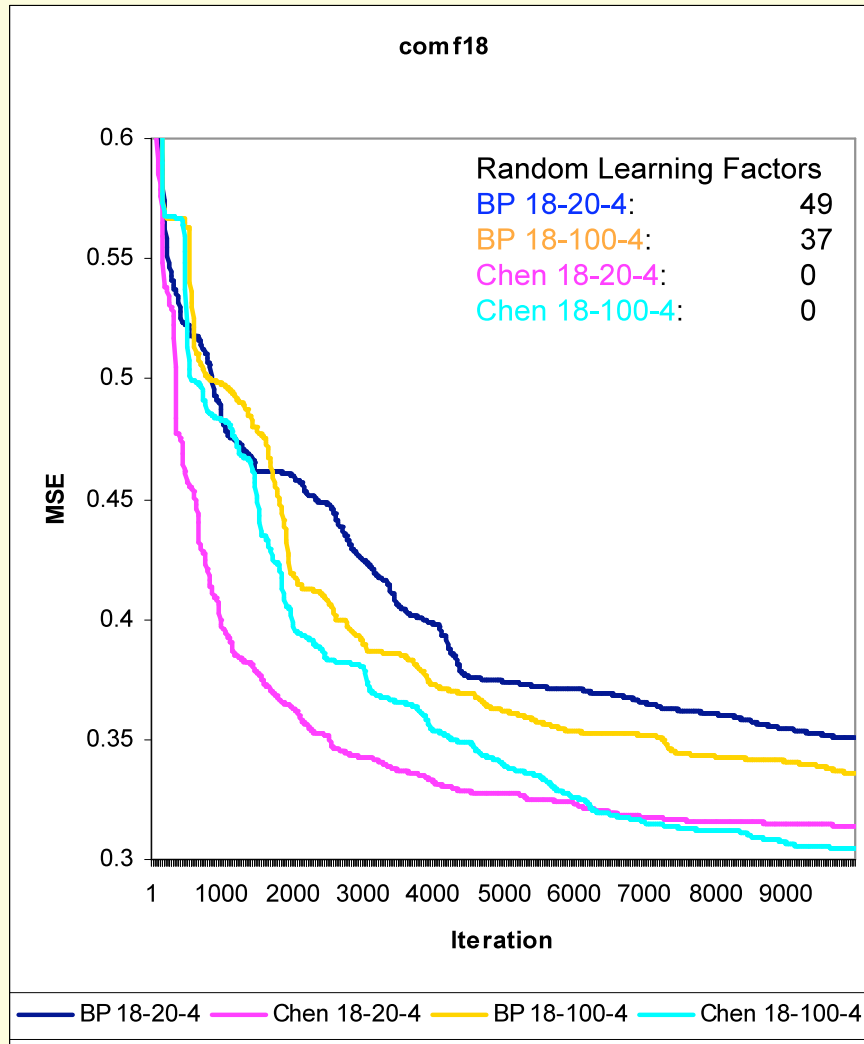
- To predict 3-month inpatient risk for 2.4 million insured members, a rare event from healthcare data [8]
 - USGS Land Use/Cover Categories, an even distribution data [9]
- 

3-Month Inpatient Risk

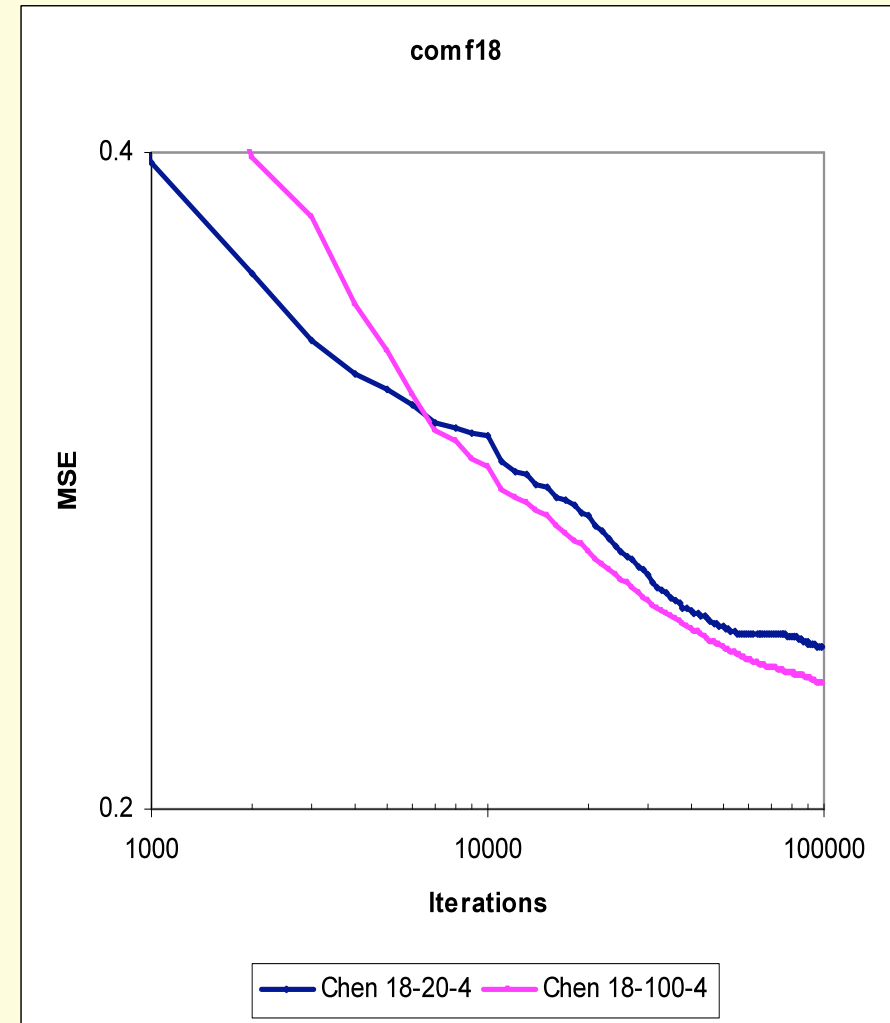
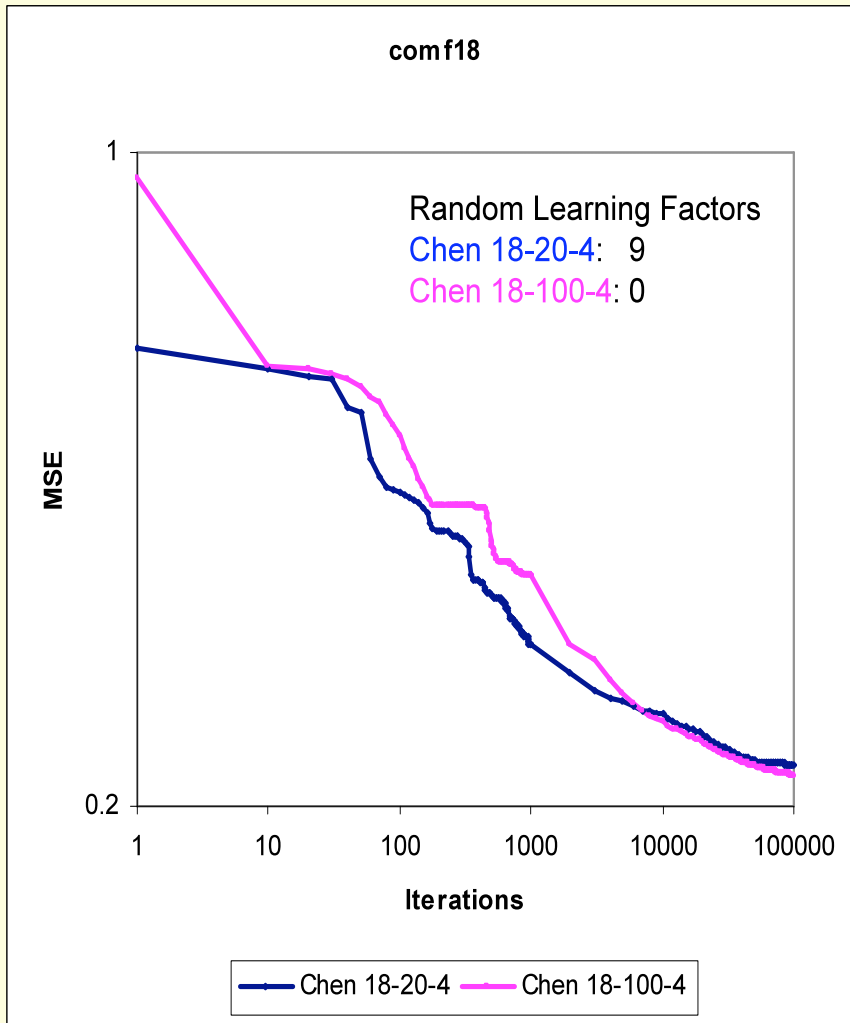
Table 1. Comparison with a commercial risk software

<i>Commercial Risk Score</i>	True Positives	False Positives	Total	Sensitivity	PPV
> 13	1,748	9,099	10,847	5.31%	16.12%
> 14	1,619	8,124	9,743	4.92%	16.62%
> 15	1,531	7,346	8,877	4.65%	17.25%
> 16	1,416	6,679	8,095	4.30%	17.49%
> 17	1,302	6,121	7,423	3.96%	17.54%
> 18	1,213	5,582	6,795	3.69%	17.85%
> 19	1,143	5,116	6,259	3.47%	18.26%
> 20	1,081	4,695	5,776	3.29%	18.72%
> 21	1,019	4,356	5,375	3.10%	18.96%
> 22	973	4,042	5,015	2.96%	19.40%
<i>Chen's model</i>					
5k model	1,778	2,708	4,486	5.40%	39.63%
10k model	2,412	5,913	8,325	7.33%	28.97%
15k model	3,004	10,336	13,340	9.13%	22.52%

Land Use/Cover Categories



Land Use/Cover Categories





Conclusions

- This proposed method incorporates the firing status of each hidden neurons to make a meaningful and efficient turn whenever it encounters an error increase.
 - The path for descending on error surface can be smooth with the least number of random learning factors.
- 