

Monitoring MLP's Free Parameters for Generalization

Graphion LLC.

Hung-Han Chen, Ph. D.

Graphion LLC. ©, all rights reserved.
<http://chenstory.com>

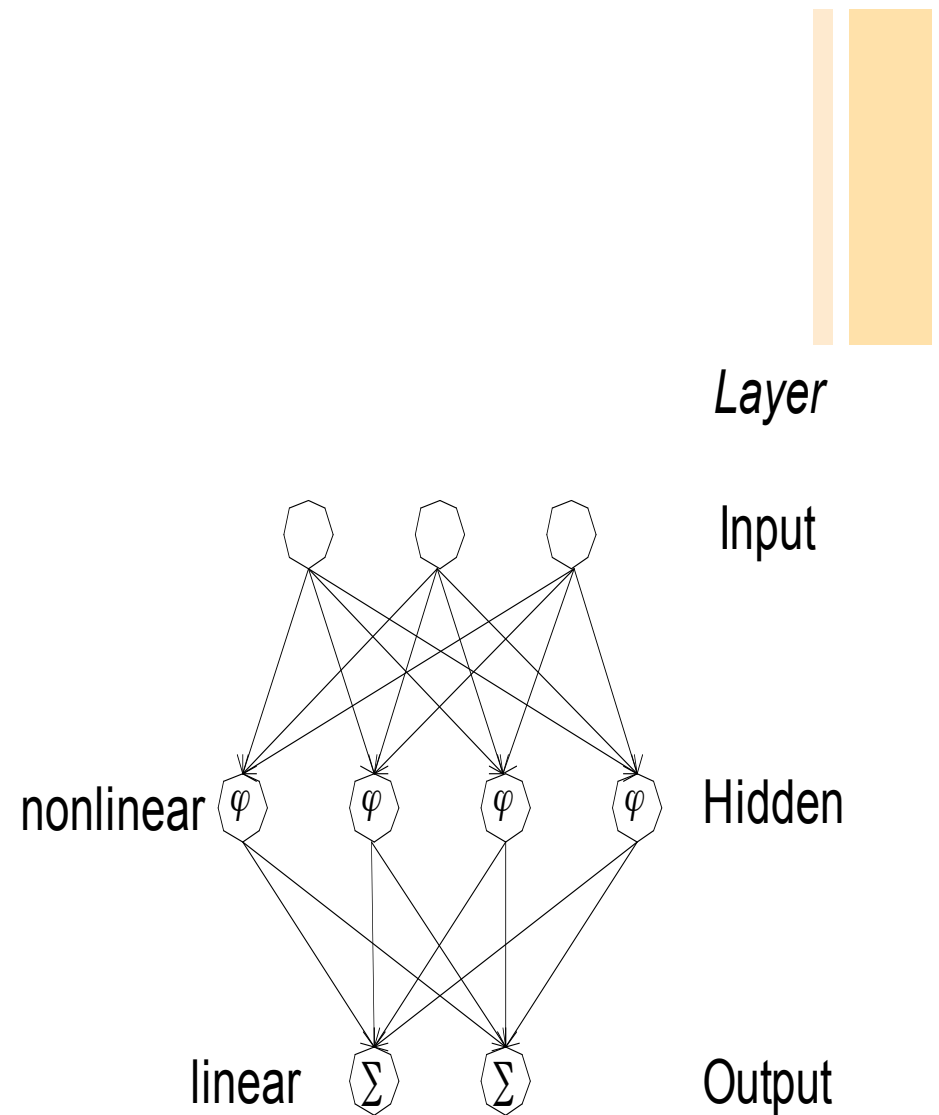
+ Outline

- MLP & Backpropagation
- Generalization
- MLP's Free Parameters
- How MLP's size matters?
- Automatic Pruning
- Conclusion
- Q & A



+ MLP

- Connectionism
- Feed Forward
- Nonlinear Activation
- Error Backpropagation
- Supervised Learning
- Least Square Algorithm
- Gradient Descents



+ MLP

4 major concerns

- **Big Scale/Rare Events**
- **Local Minima**
- **Generalization**
(Over Training)
- **Slow Speed**

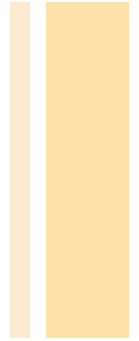
Solutions

- **NeuroSequences**
(FFC 2008, DC, USA)
- **Retreat & Turn Search**
(ISNN 2008, Beijing, China)
- **Monitoring Parameters**
(AIQED 2009, Cambridge, UK)
- **Lo-Shu Data Sampling**
(AIQED 2009, Cambridge, UK)

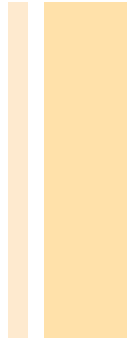
+ Generalization

Data:

- The inputs to the network contain sufficient information pertaining to the target.
- The training samples are sufficiently larger enough and representative to the set of all cases.



+ Generalization



Modeling Process:

- Too many parameters causes overfitting.
- When the degrees of freedom in parameter selection exceed the information content of the data, this leads to arbitrariness and reduces the ability of the model to generalize beyond the fitting data.
- Reducing the number of effective parameters

+ Generalization



Hypothesis: By reducing the magnitude of the weights, the number of effective weights is then reduced.

- Jittering (training with noise).
- Weight decay.
- Early Stop (with validation data).

+ MLP's Free Parameters

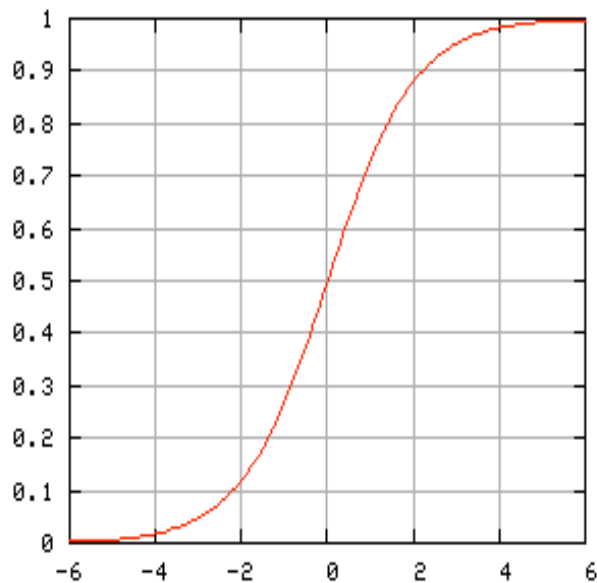
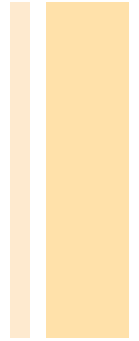


Fig. 1. Sigmoid function for hidden neuron

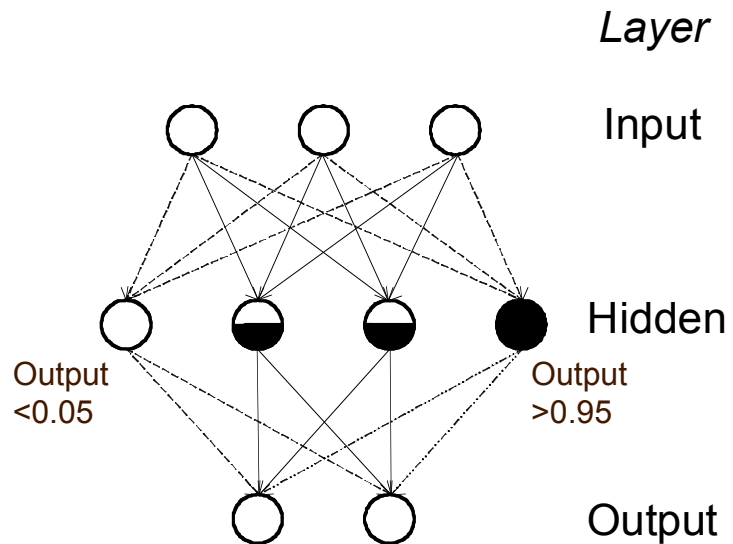


Fig. 2. The weights of saturated neurons are not effective

+ How MLP's Size matters

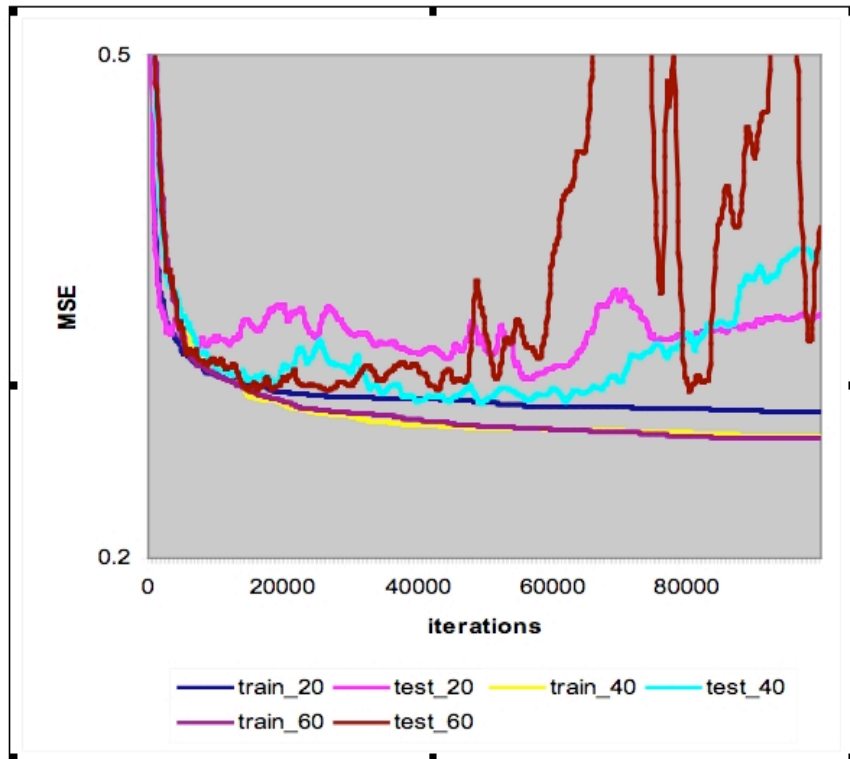


Fig. 3. 20, 40, and 60 hidden neurons

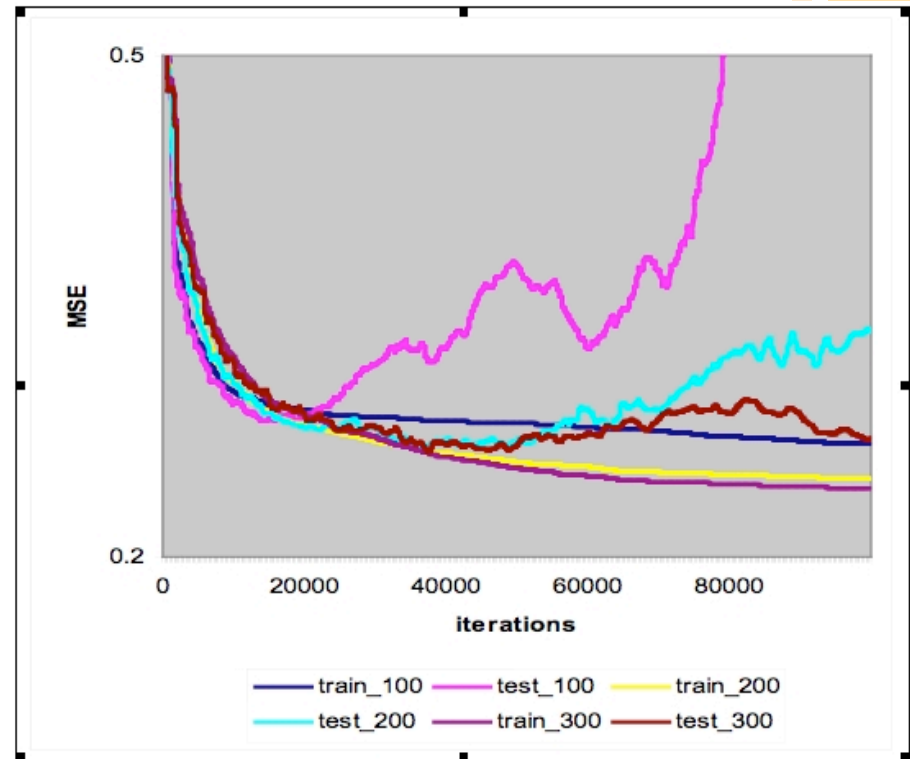


Fig. 4. 100, 200, and 300 hidden neurons

+ How MLP's Size matters

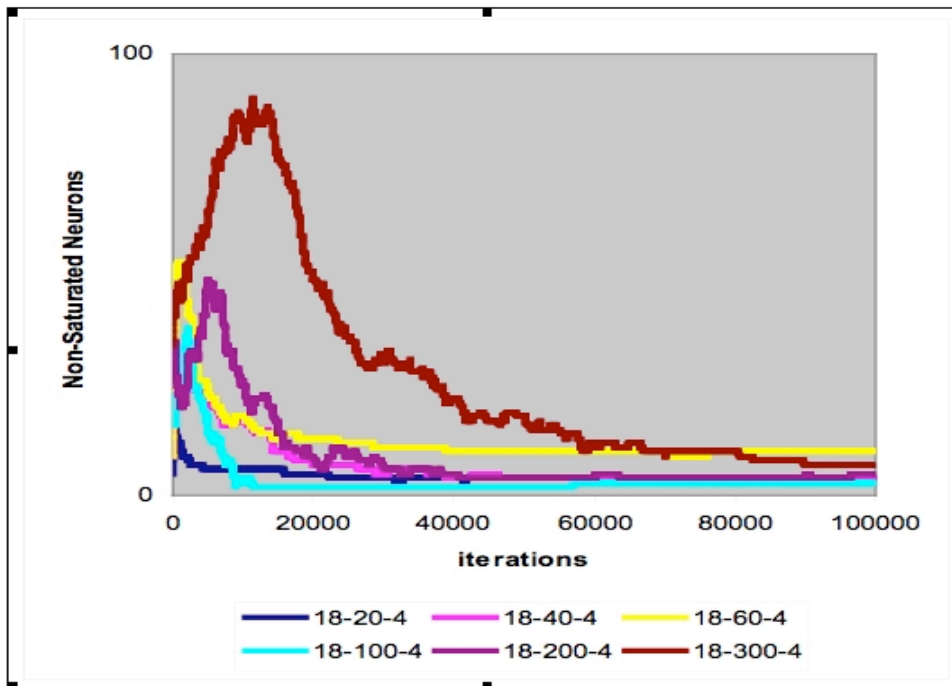
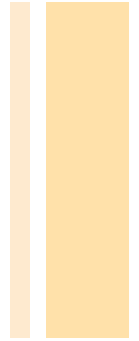


Table 2. Neuron Condition When Converged

<i># of hidden neurons</i>	Saturated low	Median low	Linear	Median high	Saturated high
20	10	0	4	0	6
40	15	0	4	0	21
60	37	0	9	1	13
100	2	0	2	0	96
200	0	0	3	1	196
300	1	0	4	3	292

Fig. 5. Number of non-saturated hidden neurons

+ Automatic Pruning

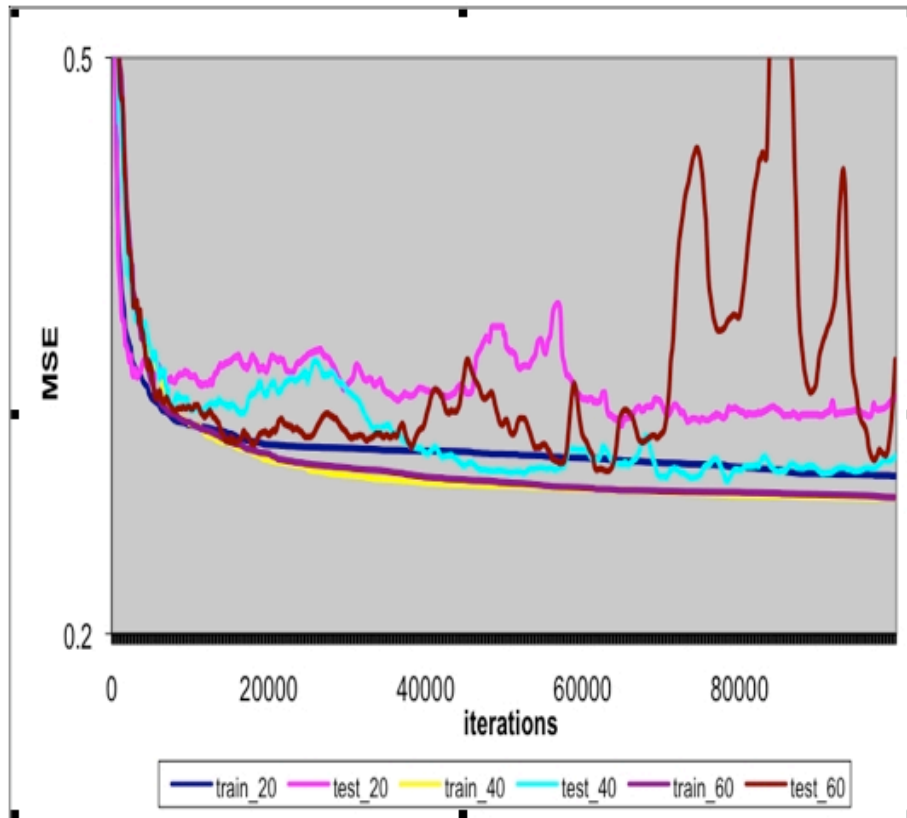


Fig. 6. 20, 40, and 60 hidden neurons

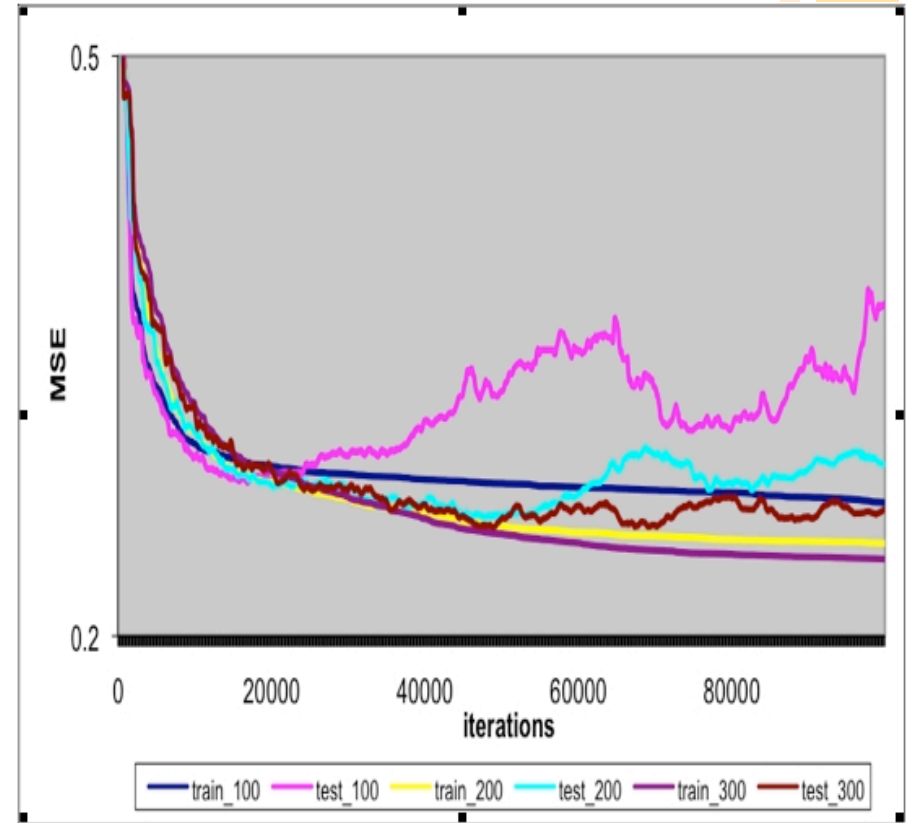


Fig. 7. 100, 200, and 300 hidden neurons

+ Automatic Pruning

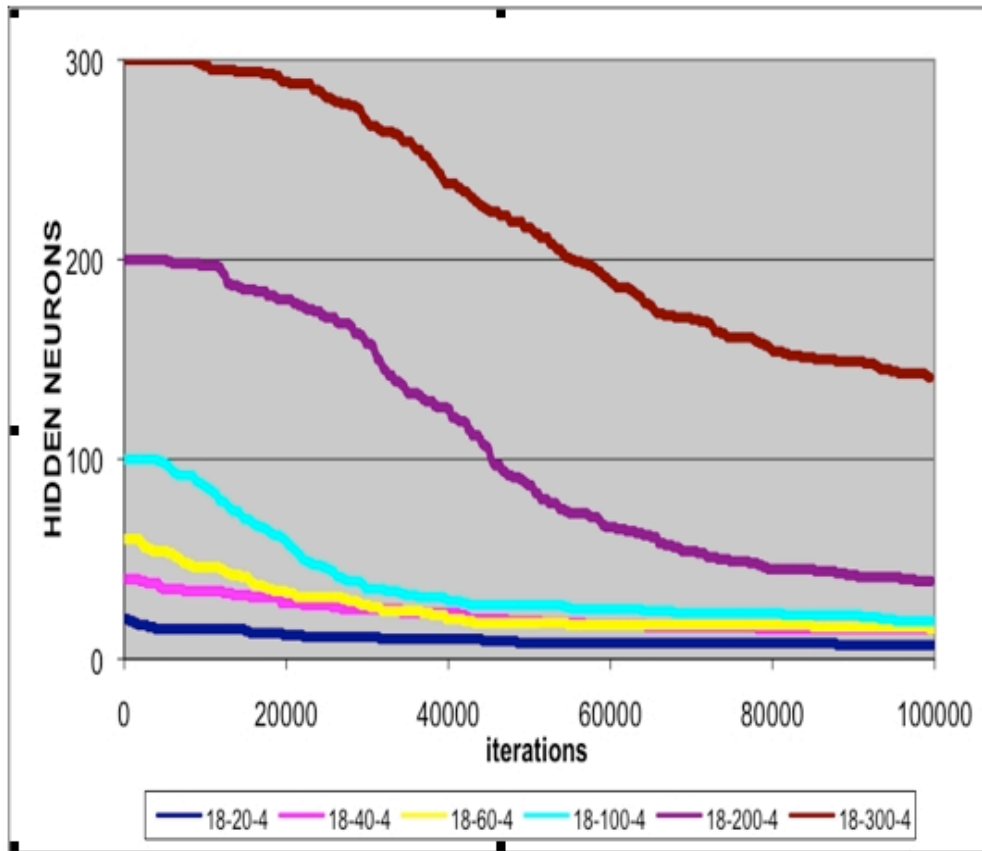
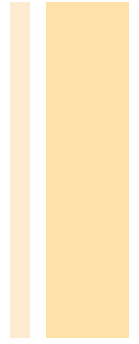
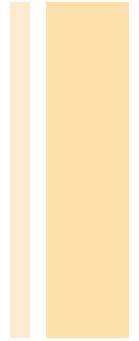


Fig. 8. Number of hidden neurons

Table 3. Converged Condition with Pruning

<i># of hidden neurons</i>	Saturated low	Median low	Linear	Median high	Saturated high
20	1	0	4	0	3
40	3	0	6	0	7
60	2	0	8	0	7
100	1	0	2	0	58
200	0	0	2	2	90
300	0	0	9	1	133

+ Conclusions



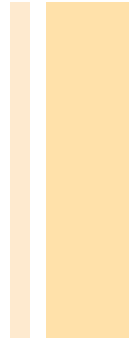
- **“Large MLP networks will overfit” is a myth:**
 - **Researchers:** Large MLPs sometimes generalize better.
 - **My Result:** If MLP is big enough, it will generalize better.
- **“All MLP’s weights are free parameters” is not always true:**
 - **Saturated neurons:** not effective as “free” parameters.
 - **Convergence:** The number of non-saturated neurons will converge and that is the best time to stop training.

+ Monitoring MLP's Free Parameters for Generalization



Q & A

+ Monitoring MLP's Free Parameters for Generalization



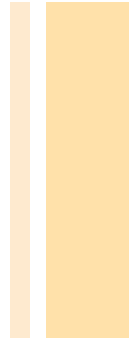
Appendix

+ Ancient philosophical observations - 1

- Before all the objects can be classified, stimuli from the objects to neurons must exist to cause the changes on neurons. And before those stimuli, neurons are in the state of primal chaos or emptiness (Laozi, Chapter 40, *Tao Te Ching*).
- 天下萬物生於有, 有生於無.
(老子, 四十章, 道德經)



Ancient philosophical observations - 2



- The changes on the neurons, through time, can form the path to classify all the objects. Along this path, first object can be classified, and then second, then third, and the rest of all objects too. Classifying these objects is embraced by the movement of neurons' outputs, between the state of inhibitory (or 0, Yin) and excitatory (or 1, Yang). This process can continue till a state of harmony has been reached (Laozi, Chapter 42, *Tao Te Ching*).
- 道生一，一生二，二生三，三生萬物，萬物負陰而抱陽，沖氣以為和。
(老子, 四十二章, 道德經)



Ancient philosophical observations - 3



- Moving toward 0 is Yin; moving toward 1 is Yang. Together they are the paths that neurons are designed to move. The ability of neurons to continue changing is essential for successful tasks of classification. The goal is achieved when a state of harmony is reached between the neurons and the characteristics of classified objects. Different characteristics of objects lead the solution to the state of harmony by different paths. All the neurons keep moving along the paths without knowing the overall solution till their traveling on the road of map found completed by reaching the state of harmony (Confucius, 10 commentaries, *I Ching*).
- 一陰一陽之謂道，繼之者善也，成之者性也，仁者見之謂之仁，知者見之謂之知，百姓日用不知，故君子之道鮮矣。
(孔子，繫辭上，易經)