Chapter 1 - Introduction

1.1 Introduction to ANN Technology

＊Conventional computers > Human:

1. Scientific and mathematical computation
2. Creation and manipulation (操作) of databases
3. Control fuctions
4. Graphics
5. Word processing

106-2 第1題 (a) Give 3 example tasks that conventional computers outperform the human brain.
Answer: as above，選3個

＊Human > Conventional computers:
Learn, Analyze, Organize, Adapt, Comprehend, Associate, Recognize, Plan, Decide
考古： 106-2 第1題 (b) Give 3 example tasks that the human brain outperforms conventional computers.
Answer: as above，選3個

1.2 Neurophysiology

＊Three major components constructing the human nervous system:

1. Brain: multipolar
2. Spinal cord (脊隨神經): bipolar, multipolar
3. Periphery (周圍): unipolar, bipolar

106-2 第2題 (a) Three characteristcs of biological nervous systems:

1. Brain
2. Spinal cord
3. Periphery

註：這題不知道為啥這幾個答案被扣1分

106-2 第2題 (b) Three characteristics of artificial neural networks:

1. Dynamics
2. Adaptivity
3. Fault-tolerance

106-2 第2題 (c) Three basic functions of artificial neural networks:

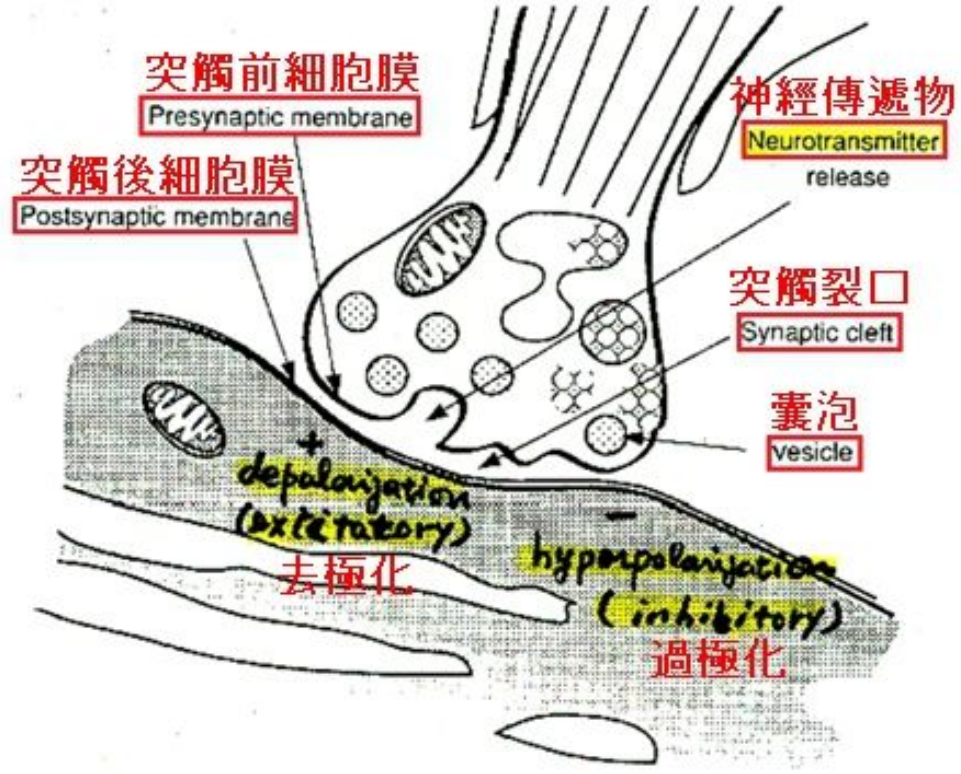
1. Analyze
2. Learn
3. Feedback

＊Single-neurons physiology’s three types of neurons:

1. Unipolar (單極): 末梢神經感覺器官
2. Bipolar (雙極): 聯絡神經脊隨
3. Multipolar (多極): 中樞神經腦

106-2 第3題 There are three major types of neurons. What are the types? Address their functions.
Answer: as above，記得寫功能

＊Synaptic Function



註: 去極化 (興奮)、過極化 (抑制)
106-2 第5題 Draw a figure to illustrate the synaptic function between two neurons.
Answer: (上方圖的簡化版)

1.3 Artificial Neural Networks

- Characteristics:

  nonlinearity, non-locality, non-algorithm, dynamics, adaptivity, fault-tolerance, input-output mapping, evidential response, self-organization,

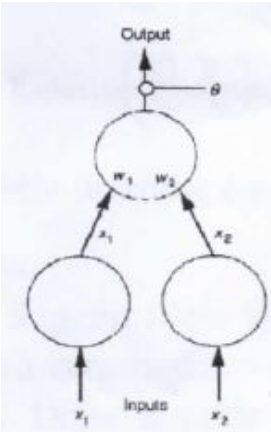
- Functions:

  learn, analyze, organize, comprehend, associate, recognize, plan, decide

106-2 第2題 (a)(b) [在文件第1頁]

＊XOR problem

106-2 第13題 Figure (a) shows a perceptron and figure (b) shows a truth table of the XOR problem. Illustrate graphically why the XOR problem cannot be solved by the perceptron.



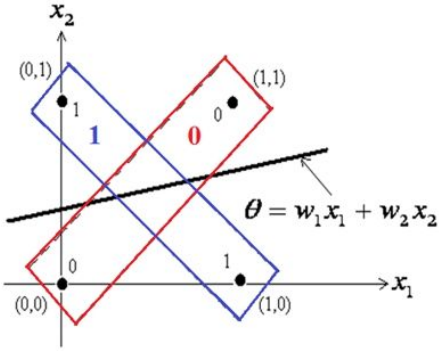
x_1	x_2	Output
0	0	0
0	1	1
1	0	1
1	1	0

(b)

Answer:

Reason:

Let $\theta = w_1 x_1 + w_2 x_2$ which is a line in the (x_1, x_2) plane.



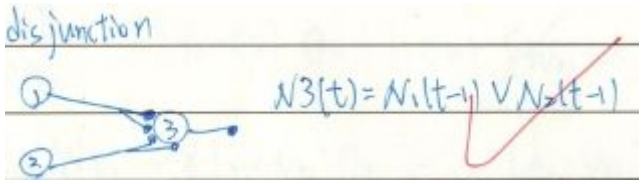
詳細可參考：機器學習- 神經網路(多層感知機 Multilayer perceptron, MLP)運作方式

疑似只在去年第1章的投影片出現，今年的找不到 (第6~8題)

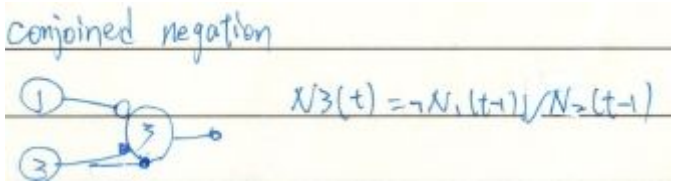
106-2 第6題 There are three basic neural circuits in the McCullock-Pitts Theory. What are the circuits?
Answer: Convergence, Divergence, Feedback

106-2 第7題 Give the propositional expressions and draw the neural circuits of (a) disjunction and (b) conjoined negation
Answer:

(a)



(b)



106-2 第8題 Write propsitional expressions for N3(t) and N4(t) in the follwing neural circuit.
Answer:

106-2 第9題 Graphically illustrate the Hebb's learn theory.
找不到出處

Chapter 2 - Adaline and Madaline

Adaline: Adaptive Linear neuron
Madaline: Multiple Adaline
考古： 106-2 第11題 What is the full names of adaline and madaline?
Answer: as above

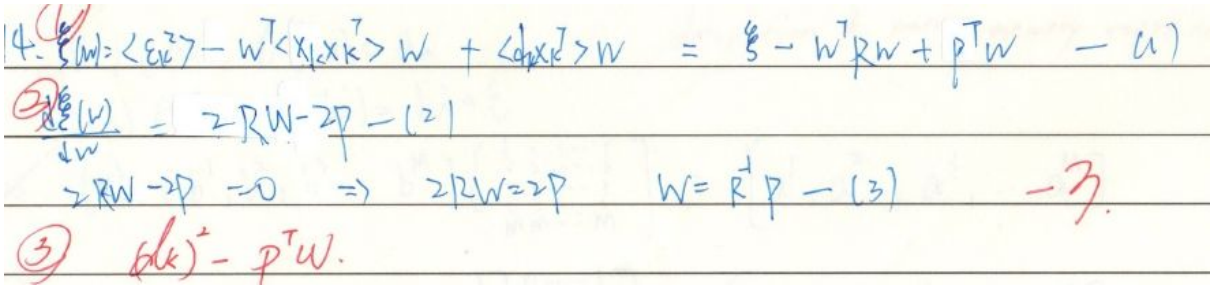
2.1 Adaline (Bernard Widrow, Stanford Univ.)

2.1.1 Least Mean Square (LMS) Learning

106-2 第14題

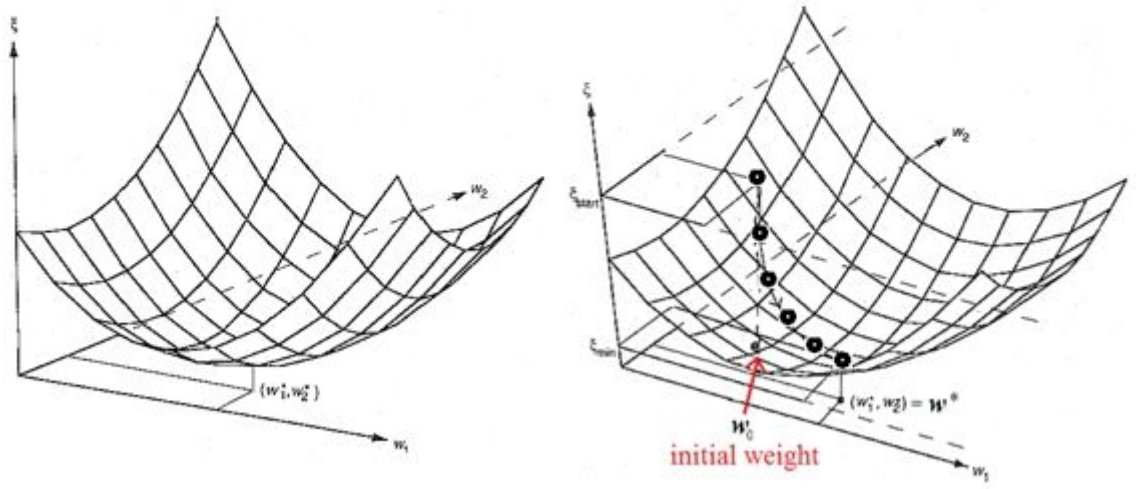
(10 pts) The objective of the least mean square learning rule has often been used to find a weight vector for an adaline using a set of training input and output pairs. Given training pairs: input vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_L\}$ and output values $\{d_1, d_2, \cdots, d_L\}$, suppose that the actual output values are $\{y_1, y_2, \cdots, y_L\}$. Let $\xi = \langle \varepsilon_k^2 \rangle$, $\boldsymbol{p} = \langle d_k \boldsymbol{x}_k \rangle$, $R = \langle \boldsymbol{x}_k \boldsymbol{x}_k^T \rangle$. (1) Formulate the mean squared error $\xi(\boldsymbol{w})$. (2) Derive the optimal weight vector $\boldsymbol{w}^*$ that minimizes $\xi(\boldsymbol{w})$. (3) Show that the minimum value of $\xi(\boldsymbol{w})$ can be written as $\xi_{\min}(\boldsymbol{w}^*) = \langle d_k^2 \rangle - \boldsymbol{p}^T \boldsymbol{w}^*$.

Answer:



2.1.2 Steepest Descent

The graph of $\xi(\boldsymbol{w}) = \left\langle d_k^2 \right\rangle + \boldsymbol{w}^T R\boldsymbol{w} - 2\boldsymbol{p}^T\boldsymbol{w}$ is a paraboloid.



Steps: 1. Initialize weight values $\boldsymbol{w}(t_0)$

2. Determine the steepest descent direction

$$-\nabla_{\boldsymbol{w}}\xi(\boldsymbol{w}(t)) = -\frac{d\xi(\boldsymbol{w}(t))}{d\boldsymbol{w}(t)} = 2(\boldsymbol{p} - R\boldsymbol{w}(t))$$

Let $\Delta\boldsymbol{w}(t) = -\nabla_{\boldsymbol{w}}\xi(\boldsymbol{w}(t)) = 2(\boldsymbol{p} - R\boldsymbol{w}(t))$

3. Modify weight values

$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) + \mu\Delta\boldsymbol{w}(t)$, μ: step size

4. Repeat 2~3.

106-2 第12題 Describe the steps of the steepest descent learning method for determining the weight vector of adaline using a set of traing input and output pairs.

Answer:

1. Initialize weight values w(t0)
2. Determine the steepest descent direction
3. Find the weight
4. Repeat 2~3

Drawbacks:

i) Need to calculate $\boldsymbol{p}$ and R,

ii) Steepest descent is a batch training method.

106-2 第15題 Address the drawbacks with the steepest descent learning approach?

Answer:

1. 必須計算 p 和 R
2. Steepest descent learning 是一個 batch training method

2.1.3 Stochastic Gradient Descent

2.1.4 Conjugate Gradient Descent

* Drawback: can only minimize quadratic functions

$$\text{e.g., } f(\boldsymbol{w}) = \frac{1}{2}\boldsymbol{w}^T A\boldsymbol{w} - \boldsymbol{b}^T\boldsymbol{w} + c$$

* Advantage: guarantees to find the optimum solution in at most n iterations, where n is the size of matrix A.

106-2 第16題 Address the advantages and disadvantages of the conjugate gradient descent learning method?

Answer:

1. Advantagea: 保證在最多n次迭代中找到最佳解，其中n是矩陣A的大小
2. Disadvantage: 只能最小化二次函數(quadratic functions)

2.3 Applications

2.3.1 Echo Cancellation in Telephone Circuits

2.3.2 Predict Signal

2.3.3 Reproduce Signal

2.3.4 Adaptive beam - forming antenna arrays

2.4 Madaline

2.4.2 Madaline Rule OO (MRII)

* Training algorithm: a trial-and-error procedure with a minimum disturbance principle (those nodes that can affect the output error while incurring the least change in their weights should have precedence in the learning process)

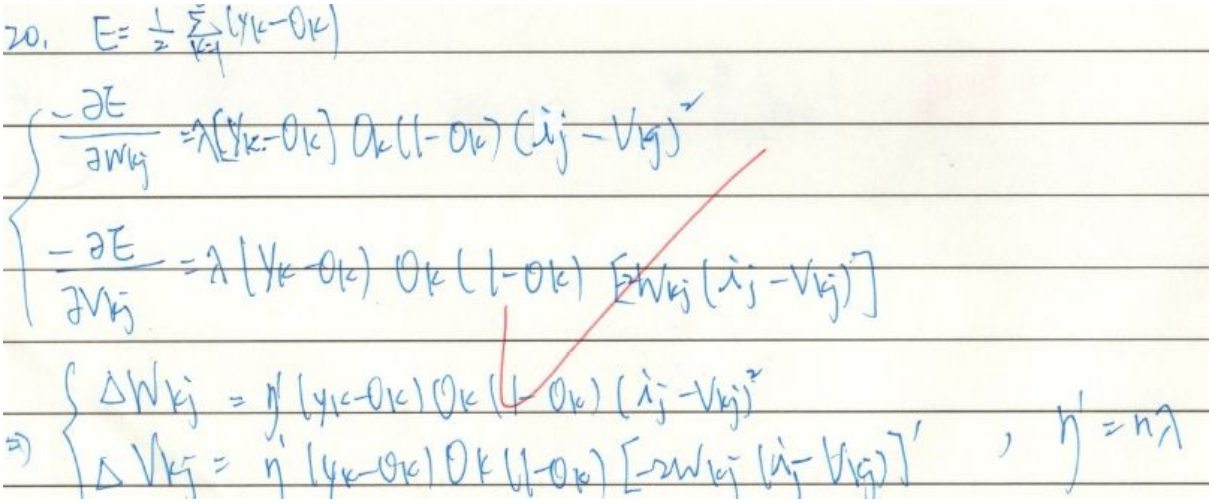
106-2 第17題 What does the minimum disturbance principle mean when training a madaline.
Answer: Those nodes that can affect the output error while incurring the least change in their weights should have precedence in the learning process.
翻譯米糕：儘管對 weights 的改變不大，那些可以影響 ouput error 的 nodes 應該在 learning process 有優先權

2.4.3 A Madaline for Traslation - Invariant Pattern Recognition

106-2 第20題 作業2

20. (15 pts) A network has quadratic output neurons. The net input to such a neuron is $net_k = \sum_j w_{kj}(i_j - v_{kj})^2$, where w_{kj} and v_{kj} are weights and are independent and i_j is the jth input value. The output function is sigmoidal $f_k^o(net_{jk}^o) = (1+e^{-\lambda net_{jk}})^{-1}$. Determine the weight updating equations of Δw_{kj}^o and Δv_{kj}^o for output-layer neurons.

Answer:



Chapter 4 - Backpropagation (BP)

* BP architecture characteristics: multilayer, feedforward, fully connected

106-2 第18題 Address the characteristics of back-propagation (BP) neural network.
Answer: 每層的輸出為下一層的輸入，而 BP 是由最末層把值送到第一層更新。

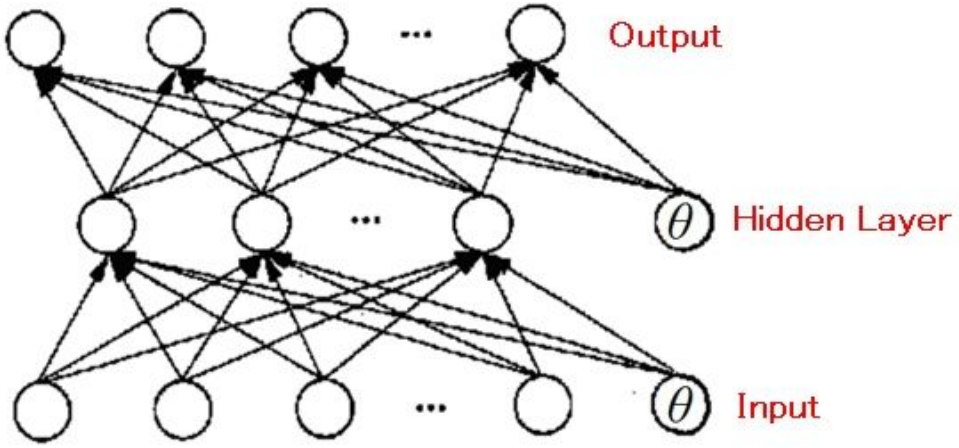
* Potential problems being solved by BP

1. Data trnslation
2. Best guess

* Suffer from:

1. Noise, distortion. incomplete
2. Time consuming

4.1 BP Neural Network



During training, self-organization of nodes on the intermediate layers s.t. different nodes recognize different features or their relationships. Noisy and incomplete patterns can thus be handled.

106-2 第22題 Draw a picture to illustrate how a BP performs image compression.
Answer: 畫出上圖

106-2 第19題 What learning rule and objective function are employed when training BP?
Answer: Self-organization of nodes on the intermediate layers s.t. different nodes recognize different features or their relationships. Noisy and incomplete patterns can thus be handled.
翻譯米糕：Nodes 的 Self-organization 在中間層上，使得不同 nodes 可辨識不同 features 或關係。如此一來，Noisy 和 不完整的 patterns 便可被處理。

4.1.2 BP NN Learning

4.2 Generalized Delta Rule (GDR)

4.3 Practicl Considerations

4.4 Applications

還沒解完 106-2 第21題 Suggest ideas for determining the numbers of hidden layers and their constituent nodes.
Answer:

Chapter 12 - Boltzmann Machine

12.1 Indtroduction

＊The major difference between BM and traditional NN is that the output function of PEs characterized by

BM: a nondeterministic function characterized by a stochastic function of inputs
NN: a deterministic function of inputs

＊Prerequisites for learning BM: (Boltzmann Machine Concepts) **記一下**

1. Information theory
2. Statistical dynamics
3. Simulated annealing
4. Energy function

12.1.1 Information Theory

12.1.2 Statistical Machanics

12.1.3 Simulated Annealing

12.1.4 Energy Function

＊Dynamic system: a system whose state changes with time.

106-2 第10題 What is a dynamic system?
Answer: The system evolve over time.

＊State: a collection of adaptable quantitative and qualitative items that characterizing the system, e.g., weights, data flows …..

＊Two types of dynamics in a neural network:

1. Training phase: iteratively update weights
2. Production phase: asymptotically converge to the solution patterns

106-2 第4題 There are two major neural operations or phase. What are they? Address their functions.
Answer:

1. Training: 不斷更新 weight
2. Production: recall the value

12.2 Boltzmann Machine

＊Three different types of BM: **記一下**

1. Completion network
2. I/O network
3. Restricted Network

12.2.1 Boltzmann Completion Network

12.2.2 Learning

＊Training the Boltzmann machine (ch.12 p.39)

12.2.3 Practical Considerations

12.3 Symptom-Diagnosis Application Diagnose why a car will not start

備註

好像沒在範圍內：6~9, 23~30

Chapter 6 - AM and BAM

我們懷念他.jpg