Chap. 4 MLP 重點

- Simplified structure of interconnected neurons
- 2-2-1 for 16 boolean fctns
- No jump connection; no feedback connection

- 不等式 轉變成 邏輯 運作在輸入空間
- High-level abstractions (representations) of the front input patterns
  - ‘ linear algebra’s linear algebra

It has been shown

- that only one layer of hidden units sufices to approximate any function with finitely many discontinuities to arbitrary precision, provided the activation functions of the hidden units are non-linear (the universal approximation theorem).

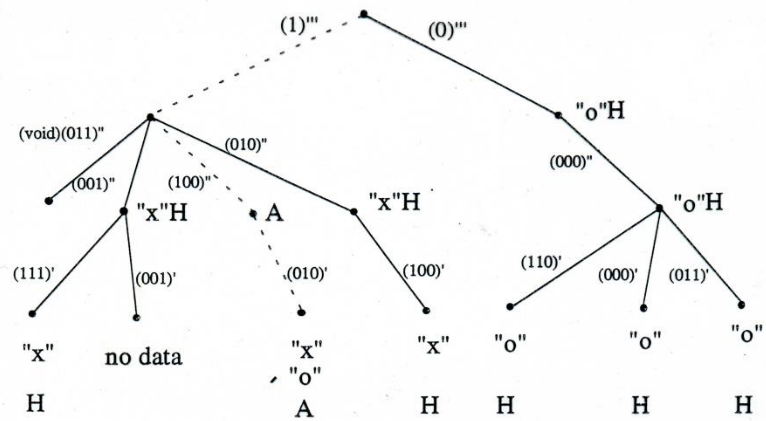
- (Hornik, Stinchcombe, & White, 1989; Funahashi, 1989; Cybenko, 1989; Hartman, Keeler, & Kowalski, 1990)

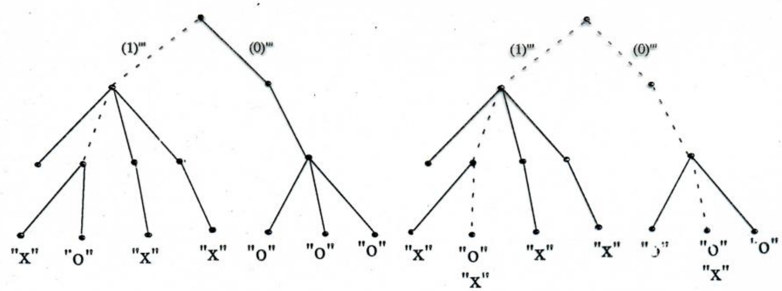
Chapter 4: Hidden tree in MLP

- Kolmogorov theorem
- Existence theorem

- He did not show how to implement it?
- An operational solution is lack.
- Even exists such sol, it will be too much complex to operate.

- The Hamming tree





可依此图消除整 neuron ③型

因任何 neuron 被去掉後仍能保有 homogeneous 结構
且不造成 ambiguous 现象皆消除掉

模糊區

MLP complex 不一樣 → MLP. 不易寫成代數式'
Boole 代數容易..

清除贅 neurons 的三种方法分辨

① 對所有 training pattern 反應一樣

② 兩個 neurons 的反應一樣

③ 選擇此 neuron 後並不造成 ambiguous 區域

idle neuron (void neuron) unchanged sign for all training patterns

Homogeneity unchanged when suppress this neuron

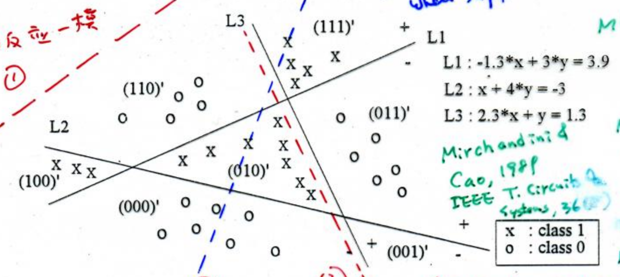


Figure 3 A typical training result of the first hidden layer
x is the first input unit
y is the second input unit

Mirchandani & Cao, 1989, IEEE T. Circuits & Systems, 36

$M(J,n)=\sum_{k=0}^{n}\binom{J}{k}$ complexity 非 2^n

$\binom{J}{k}=0$ for $J<k$

M: region 個數; n=2: dimension; J: hidden neurons: 3

$M=\binom{J}{0}+\binom{J}{1}+\dots+\binom{J}{J}=2^J$

L_3' neuron has same sign (or reversed sign) as neuron L^3

granularity

n=2: $J=\frac{1}{2}(\sqrt{8M-7}-1)$

- Hamming distance
  - Neighboring regions (010)' and (011)' have a common boundary L_3 and have a Hamming distance of only one bit.
  - All connected neighborhood regions have an exactly one bit Hamming neighborhood distance.
  - Far and unconnected regions may have maximum Hamming distance of all three bits.

MSB neuron
LSB

Local neurons for local data

note, ∵ when activated state vectors s <<< 2^N, N: # of neurons in that hidden layer

we can keep few neurons n, 2^n = s, to represent the states and prune, drop, all other neurons.

∴ We can always use 3 layers network to achieve arbitrary # of hidden layer training program.

Whenever a homogeneous layer is reached we augmented a new hidden layer and start the

Hidden tree

- Pruning neurons

1. two neurons both with the same or reverse responses to all patterns,
2. delete a neuron has a same response to all patterns. Delete it.
3. delete useless neurons 分割同類 data

4 delete a neuron will not generate any mixed ambiguity cells. 簡化結構

Day 6 / 17大進展 NetTalk 1987 wiki

- Perceptron 1943
- LMS learning = Widraw learning 1960
- Webro 1974 (Harvard)

  Backpropagation

  LMS 50Yr Celebration Presentation Paul Werbos part 1 (videos 2009/06/17 IJCNN)
- 1986 a “renaissance” in the field

  Book by Rumelhart, Hinton and Williams

1986多層 neurons 學習 大幅增加計算能力

NetTalk BP分配獎懲率各取所需 還是各取所值

- **BP** 分配獎懲率各取所需 還是各取所值
- BP : solve one type of credit assignment problem (舊的 homework)
- Chemistry application prediction of disulfide binding state
- http://www.uow.edu.au/~markus/teaching/CSCI323/Lecture_MLP.pdf

Day 6 / 17 NetTalk

- From where does its mysterious power come ?
- None of the existing methods or contemporary technologies can accomplish such pronunciation task 95% corrections.
- Discovers 規則變化 (regularity 80%) 並記住不規則變化

NetTalk

- Discovers 規則變化 (regularity 80%) 並記住不規則變化

- 例外使得 rule based 方法失效
- formal system 也失效
- logic 失效
- 集合論失效
- 機率失效 統計失效

NetTalk shows MLP by LMS

- Autonomously exploits useful hidden structures in training dataset; such as vowels and consonants.
- Discovers hidden structures (regularity 80%)
- Utilizes those implicit structure + 記住特例
- utilizes those structures to simplify the problem drastically.

Chapter 4 showed, suggested

- 世界是從底下自組(建築)起的 Boltzmann
-
- World not imposed from the above God,
- World merged from the low,

(33.00min in video)

Hidden tree

- Coding cell =\ 不是 binary number
- 不是 Minsky 所說的 數
- Neuron 有 MSB and LSB 性質
- Neuron does not understand number.
- Neurons use representations to solve problem.

  (101)’ =\ 1X2^2+0X2^1+1X2^0=7

Chap. 4: MLP tree

- Over simplified neural structure

  No jump connections;

  no feedback connections

  Capable of high-level logical abstraction

  May be called a kind of

  ‘鑑別區分 machine’ or ‘分類 machine’

  or ‘a self-tuning 分類 machine’

Finest areas of the first hidden layer

- are cells, 空間被 perceptrons 切到最細碎　每一cell 只含純一類 data
- or patches 多面體形狀 polyhedral
- or building blocks of succeeding layers
- Merge → merge → merge layer by layer
- Coding → coding → coding →
- Combining 結合 → combining → combining
- Piling → piling → piling →
- Grouping → grouping → grouping →

Finest areas of the first hidden layer

- 離散化 data space 的連續空間
- 簡化後續處理
- Cell 被編上符號後 第二層與更深(上)層 hidden layers 只用前一層 cell 的符號去 training 不需用原始 data training 可簡化後續處理
- 第二層與更深(上)層 hidden layers 持續簡化 最終符號數量 會愈來愈少

Each cell in 2-3-3-1 network

- Each cell has a polyhedron shape and also a convex shape (first hidden layer)
- Neighborhood cells are different in one bit, in one 區分 line
- Total number # of cells for J hyper-planes in n dimensional space is #(J,n)=Σ_{k=0}^{n} (J,k)

Equivalent Isomorphism

- 第一層hidden layer 異質同構 等價結構
- #(J,n)=Σ_{k=0}^{n}((j)/k) 的等價結構
- Xun Dong
- polyhedral complex
- The bounded complex of a hyperplane arrangement
- Xun Dong 的等價結構更複雜

Hidden tree 反映 cells 的組合結構

- There are so many rules among cells. 有各式各樣的編織方法編織這一群 cells.
- There exist equivalent parts in the hidden tree structure.
- This tree reveals all geometrical relations of cells in hyperspace 看不到.
- One can see the relations. 搬到眼前

Statistics and probability ways

- 機率會使 cells 成為不同程度的灰色 cell 內的 data 不可能為單純一類
- MLP perceptrons 區隔不同純類 沒有灰色類
- MLP 只處理 機率=0 機率=1 兩種類
- 不處理 機率=0.3 淺灰類
- 藉 MLP tree 可以挑出 ambiguity cells 機率=0..3 的 cells
- 遇到有 機率=0..3 灰類 cells 時 針對個別 cell 另行加工

Statistics and probability ways

- 如果非要加入機率假設
- 如孟德爾的雜交碗豆

- 淺灰色 <0.5 cells 與 深灰色 >0.5 cells 分別結合在不同的子樹上 可以做到 global minimum E.

Statistics and probability ways

- Cells 分三種情況 分別處理 (之後有時間再講)

- 1. 純黑白 cells 如 Chap.4 內容
- 2. 純黑白 cells + 灰色機率 cells
- 先區分純黑白 cells 再處理灰色機率 cells
- 3. 純灰色機率 cells 有機會再講
- 淺灰色 <0.5 cells 與 深灰色 >0.5 cells 分別結合在不同的子樹上 可以做到 global minimum E.

Minsky

- Neurons do not know 12 (數量) comes from,12=7+5? , see video 39.50 37.48
-
- Neurons do not know numbers.
- Neurons use representations to solve problem.
- 切割 input space 體重身高(數量)座標軸

MLP tree & logic

Can express any Boolean fctn by
Iterative logic fctn F1(F2(F3(X)))
邏輯 (and, or) 也需要 nest 運作

- Tree = logic relations in spatial space.編織架構
- Marriage 融合 logic and geometric relations
- Logic may have no spatial content,
- Conversely, space has no logic content
- 離散區割 X space into finest cells
- and coding them
- Combine coded cells into high-level codes

MLP

- Marriage of logic and geometric space
  將空間編織上 Logic 搭 Logic 便車 利用 Logic 輔助
- MLP tree 約略可看出 data 空間結構
- Piling → piling → piling →
- Grouping → grouping → grouping →

Hidden tree

- Codes of areas are symbols.
- Codes are not binary numbers.
- Neurons develop symbols to solve problem.
- Neurons do not understand numbers.
- Neurons do not understand probability.
- Neurons do not use probability to solve problem ← → Bayes.

MLP 將區割線轉變成 logic 內涵

- Even more,
  neurons do not understand logic.
- Neurons use group force (representations) collectively to solve 區分 分類 problems.
- 搭 Logic 便車 利用 Logic 輔助

MLP tree: constructive way

- MSB LSB neurons
- Redundant neurons
- Retrain locally and use local data
- Divide and conquer
- Constructive way
- BP errors tend to get lost in front layers

Tree nodes are representations

- 1-bit neighbors, 隔一條區別線,
- Tree support 支撐 dataset 空間結構 spatial structure
- 相鄰近的兩個 data 可能隔兩條區別線 見前圖
- Hamming dist.=2
- 2-bit neighbors,

  1-bit neighbors' neighbors

  excluding itself and 1-bit neighbors

Notes on BP and MLP

- 1. 分群 grouping of output vectors of first hidden layer by PCA 取代 binary tree in NetTalk 可用目視看出分群
- 2. Mark each sample in PCA with its error (深淺色),看出錯誤率大的資料
- 3.不同類用不同兩種顏色著色 (兩類)
- 4. 挑出相互矛盾的資料 samples 另行設法處理

Notes on BP and MLP

- 5.看出規則變化的特例 與 不規則變化共同交集激發的那一組 neurons (majority), see NetTalk
- 6.依錯誤程度訂出 MSB LSB neurons 重新訓練 LSB neurons (MSB 不更動)
- 7. 調整距離輸入資料最近的幾個 hyper-planes 三點共面 四點共體

Notes on BP and MLP

- 8. l weight l~~0 區別線平行該 weight 拒絕該輸入X
- 如果整層的 特定 l weight l~~0 可能由於該輸入是 noise 或 無關量 如病況與石頭的顏色
- 如果整層的 所有 l weight l~~0 代表 don't care 拒絕輸入資料 資料飽含矛盾 (BP 為降低 MSE 才造成 ~~0 現象) 應檢查輸入資料 有時整層都~~0 堵住輸入資料

Notes on BP and MLP

- 8. 找出 l weight l~~0 的所有 neurons

  l weight l~~0 代表 don’t care 輸入資料資料有矛盾 (BP 為降低 MSE 才造成 ~~0 現象) 應檢查輸入資料 有時整層都~~0 堵住輸入資料
- 9. 每一sample標上 各層的 label codes
- 10. 用 hard-limit 去算每一 sample 的 error error=0 的 sample BP不修正

MLP 經驗法則

- Practically, n_1>>n_2>>n_3
- The number of neurons in the first hidden layer is much larger than that in the second hidden layer.
- The number of neurons in the first hidden layer is estimated n_1=2n+1 . Komogorov theory 1957 Poggio(MIT)
- **Nash's Embedding Theorem on 2n+1**

Conclusions of hidden tree

- The most important conclusion of Chapter 4 on hidden tree is:

"To get perfect performance (100% correction; global solution) on training dataset, the MLP must be accomplished in a bottom-up manner."

Any BP algorithm will converge to a local minimum solution. BP errors will get lost in front layers.

Chapter 4

- LMS
- Front layers (下樑 input)
- Rear layers or deep layers (上樑 output)
- 上樑 provides logical (representation) track

- NetTalk learns
- 80%regular rules + 20% irregualr cases

LMS by Widrow

Minimizing probability expectation E(error**2) is a wrong direction. 應從原始 raw 著手
不必追求機率的完美

- Record errors for each pattern and for each neuron. One can develop various manipulations 策略 for the training sequence during training.
- Tune weights for large errors with priority.
- There is no need to introduce the
- assumption "stationary, …." E

舊的 Homework #1 Google cancer

- Write BP program for 2-3-3-1 network + online hidden tree.
- NetTalk updating eq. for w(t+1)
- Show how BP distributes corrections of error to each neuron. BP分配獎懲率, see NetTalk
- Record the 1-bit neighbors + 2-bit neighbors
- Hidden representations
  - + hidden tree by Sejnowski
- Generate artificial data set or use real dataset
- MSB & LSB neurons + pruning

MLP & math

- Widrow adaptive
- Kolmogorov learning NN theory
- Kolmogorov 泛化 大陸
- MIT open course ware
- Kolmogorov space filling 2n+1
- Kolmogorov Fr

Komogorov theory 1957

- *Debates*
- *Kolmogorov's Theorem* Is irrelevant
- An exact representation is hopeless
- *Kolmogorov's Theorem* Is Relevant