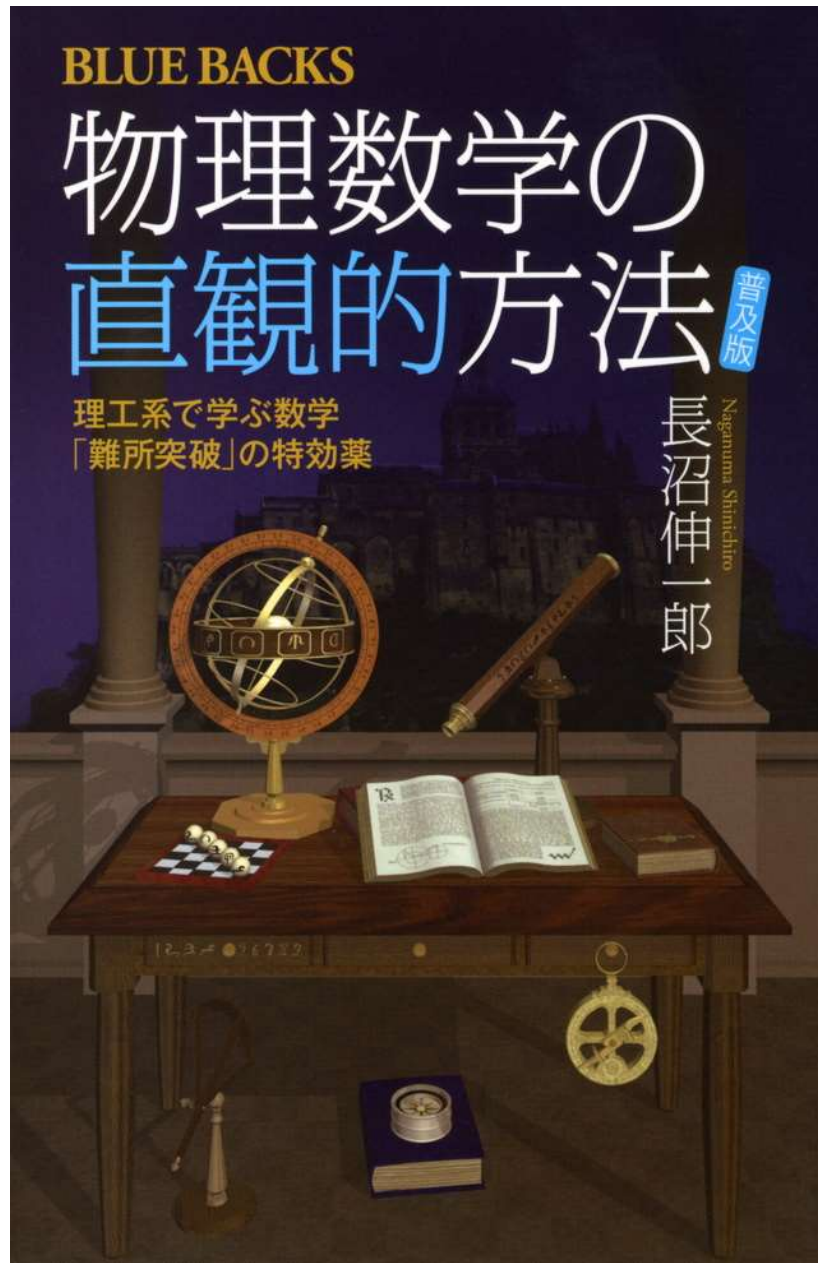


# 数学からのアプローチ

以上のことは、2011年当時、  
とてもよく分かっていた。

いわゆるディープラーニング元年は2012年

当時、ReLU関数に相当するものを、  
私たちは  
「**スイッチ演算子**」  
と呼んでいた。



電子版 ①

「対角化解法」で  
微分方程式は解けるか

## 問題意識

線形な行列では、非線形な問題は扱えない。  
非線形にどう立ち向かえばよいのか？

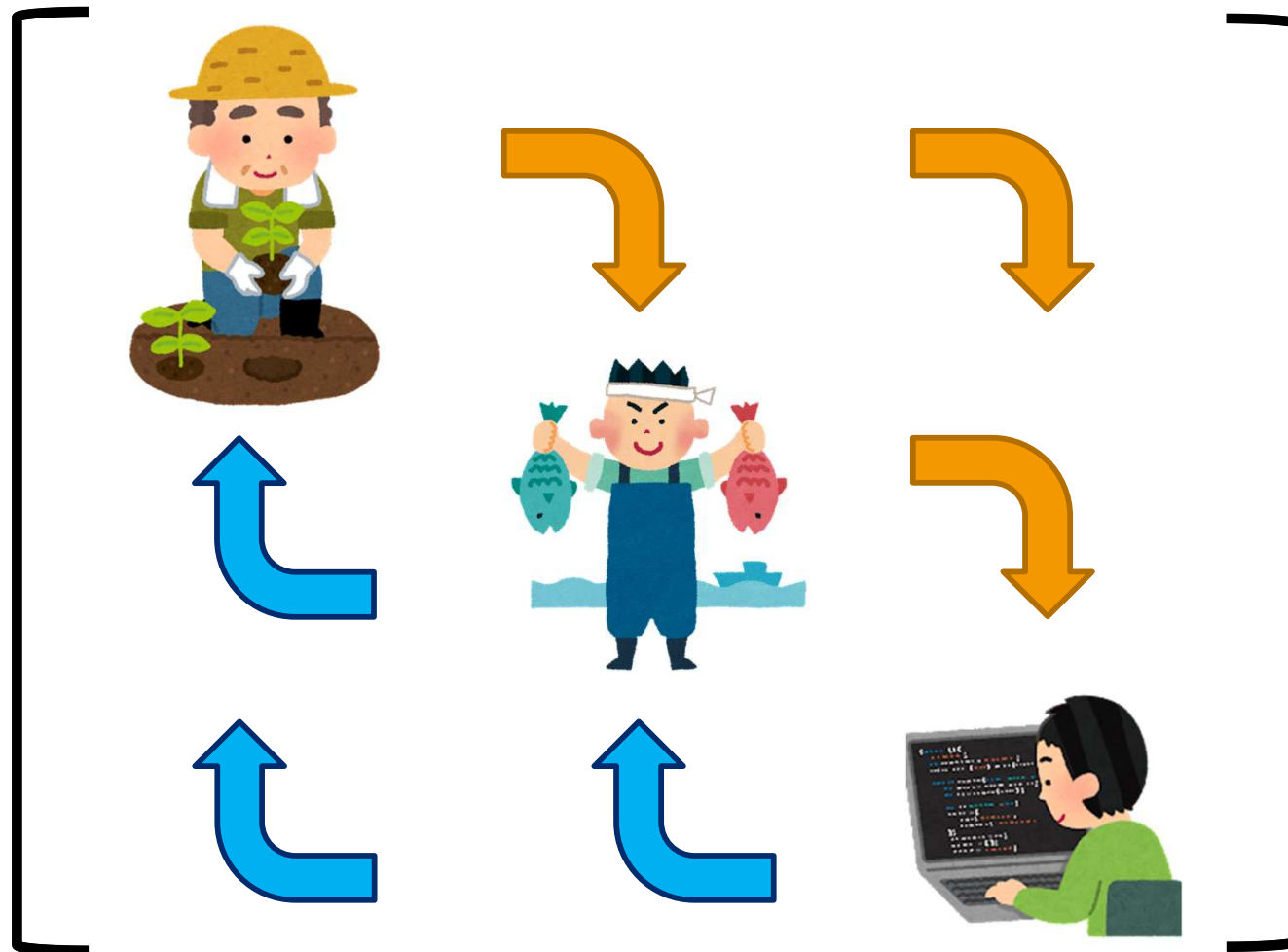
### 非線形の場合

しかしながらこのような固有値を用いる議論は、それが線形代数から来ている以上、基本的に「線形」の範囲内の問題しか扱えず、複雑系を把握するには不十分なものである。実際現実には線形方程式  $f'=f$  などよりも非線形方程式  $f'=f+f^2$  などのようなものが問題になることが多いのだが、今のままでは後者のように2乗の量が入り込んでいる問題はまだ作用マトリックスに表現すること自体が十分にできない。

そこで、このように変数自体が2乗や3乗になっている非線形の問題を作用マトリックスで具体的にはどう扱うかを、一応簡単に紹介しておかねばなるまい。

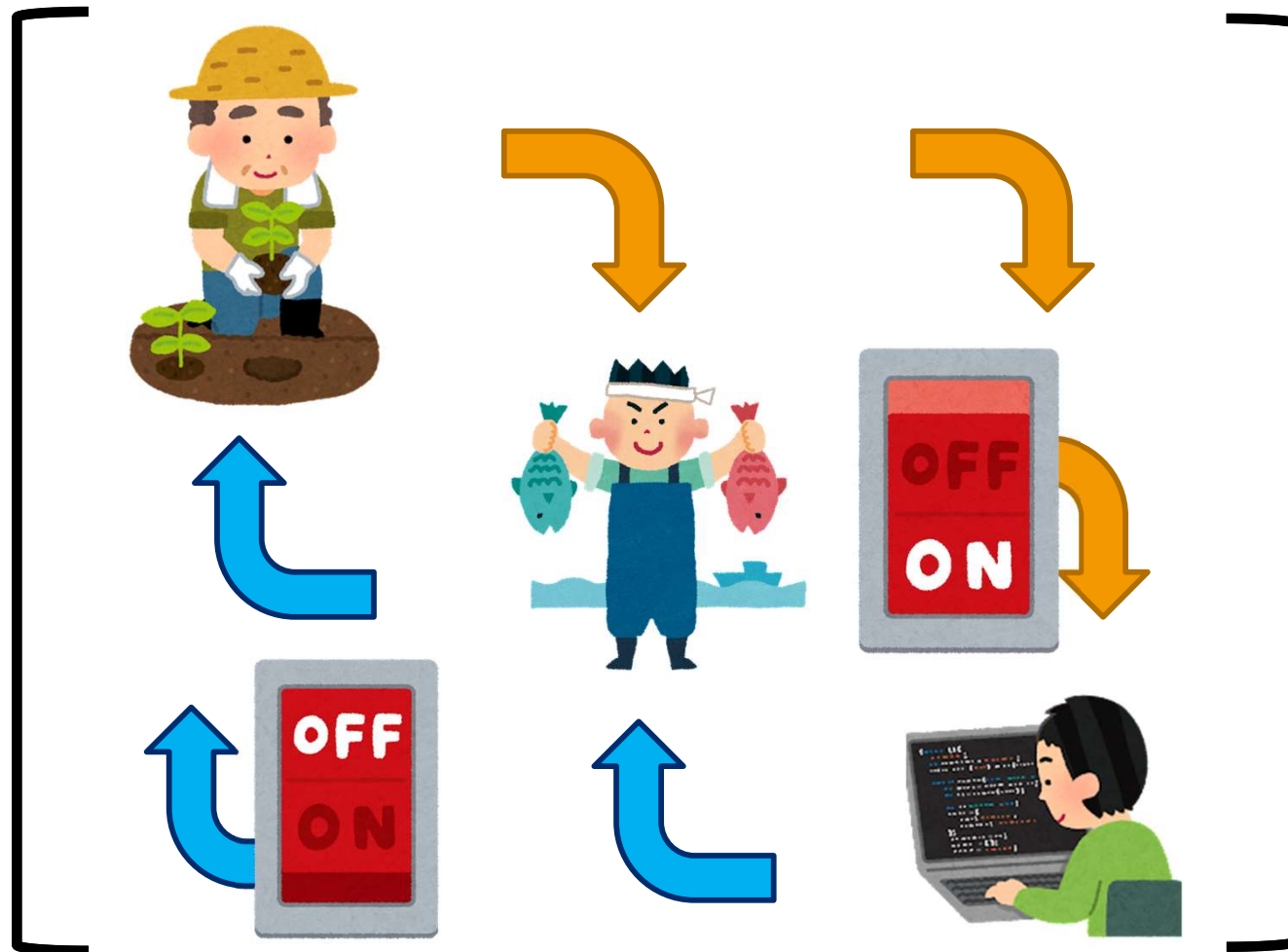
# “作用マトリックス”とは

世の中をざっくり直観的に把握するための道具立て。  
(この場だけの用語です。)



# “スイッチ演算子” のアイデア

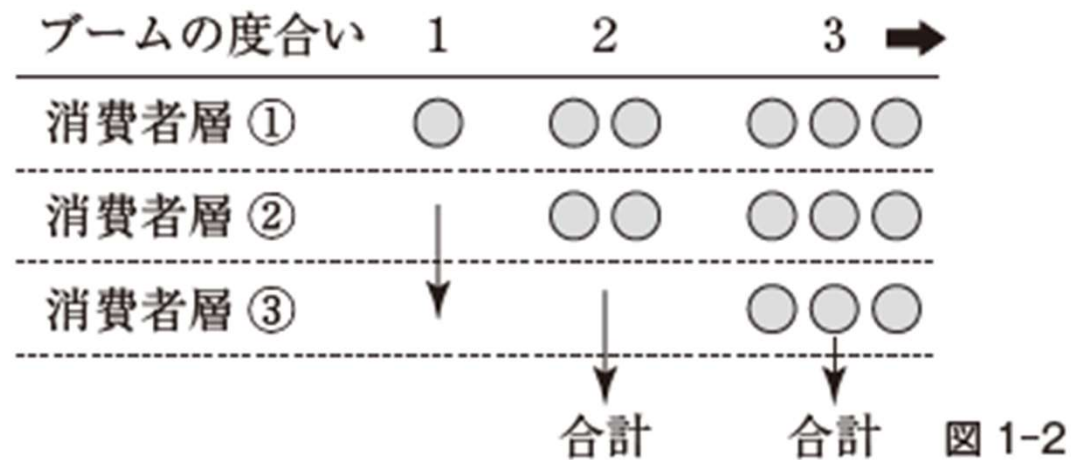
線形行列の中に ON/OFF のスイッチを組み込むことで、非線形現象が近似できる。



例えば何かブームが起こった時にその経済効果が2乗で出て来ることがある。つまりブームの度合いが2倍, 3倍となっていた時に, 関連商品の売上げ合計が4倍, 9倍となっていくことなどがしばしばあり, この現象が組み込まれた問題は立派な非線形問題だと言える。

ところがこの種の経済効果のメカニズムの場合, 良くみるとそれが線形現象の組み合わせになっていることが少なくないのである。実際先ほどのような場合, 個々の消費者の行動を見ると, 一人一人の購入額自体はブームに単純に比例して2倍, 3倍にしかになっていないことがしばしばある。

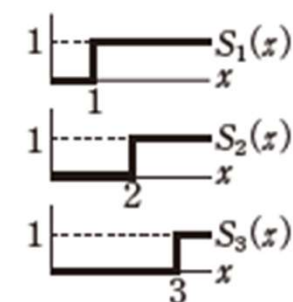
それがどうして2乗になってしまうかという点, ブームの到来によってそれまでそんな分野に全く関心をもっていなかった人々が, にわかマニアとして次々に消費に参加し始めるからである。つまり消費者の人数自体が2倍, 3倍に増えていくため, 社会全体での購入額は, 各自の購入額と人数の積で4倍, 9倍になっていくことになる。



ある意味でこれは非線形現象の本質を突いているとも言え、ほとんどの非線形系はその気になれば（無理にでも）このような形に置き換えて扱うことが十分可能である。



$$\begin{pmatrix} a \cdot S_1(x) & b \cdot S_2(x) & c \cdot S_3(x) \\ \vdots & & \end{pmatrix} \begin{pmatrix} \textcircled{1} \\ \textcircled{2} \\ \textcircled{3} \end{pmatrix} \text{グループ}$$

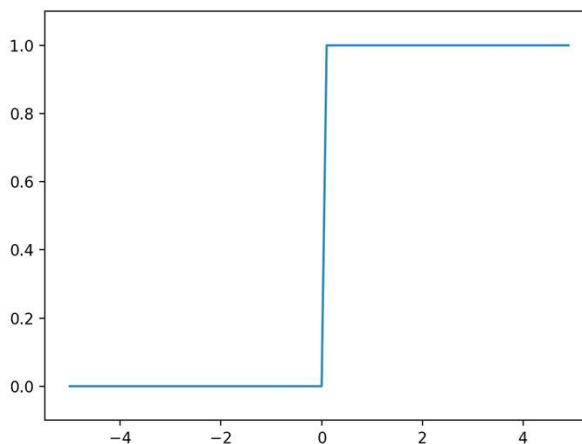
$$\begin{pmatrix} a & 0 & 0 \\ x=1 \end{pmatrix} \rightarrow \begin{pmatrix} a & b & 0 \\ x=2 \end{pmatrix} \rightarrow \begin{pmatrix} a & b & c \\ x=3 \end{pmatrix}$$


つまりこの場合、 $S_1$  は  $x$  (ブームの度合いを示す数値) が 1 でもオンになっているが、他の  $S_2$ ,  $S_3$  はこの時オフで 0 である。そして  $x$  が 2 の段階では  $S_1$  と  $S_2$  がオンで、消費額はその和となる。このようにしていくと、スイッチ演算子が次々にオンになっていくことで、消費者層の段階的参加が表現される。

そしてこれを細分化していくことで、ほぼ連続的に 2 乗で変化する量を扱えることになるわけである (3 乗以上は、これを入れ子式にしていけば原理的に表現可能である)。それらの具体的な細部については省略するが、ともあれこのようにすれば基本的に非線形現象を扱っていけるのだということはおわかりいただけたろう。

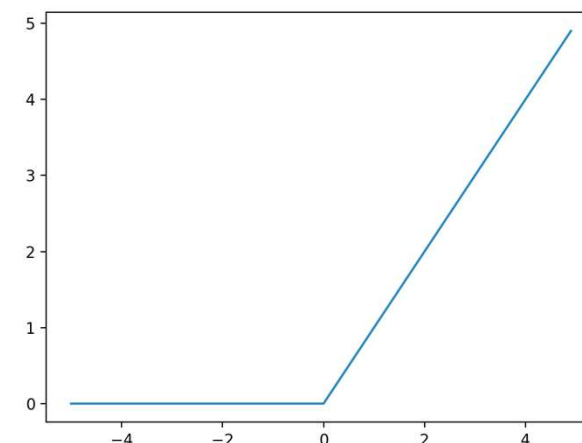
あらゆる実用的な非線形関数は、  
スイッチ演算子の重ね合せによって構成できる。

スイッチ演算子



微分'

ReLU関数



積分

概念的には  
同じもの

当時、私たちは  
ロジスティック曲線を  
“スイッチ演算子”で表す方法  
を考えていた。

つまり、ディープラーニングが  
やろうとしていたことを、  
目と鼻の先で考えていたことになる。

## 当時の状況を、試験問題風に書き下すと・・・

- (問1) シグモイド(S字関数)によって、任意の実用的(連続滑らか)な関数を近似する方法を説明せよ。
- (問2) 代表的なシグモイドであるロジスティック曲線は、スイッチ演算子の組合せで表せることを示せ。
- (問3) 今日のシグモイドを用いたニューラルネットワークは、勾配消失問題という限界を迎えている。この限界を突破する方法を考えなさい。

もし、このような形で誘導されたなら、  
できなかった学生は相当のうっかりさん。

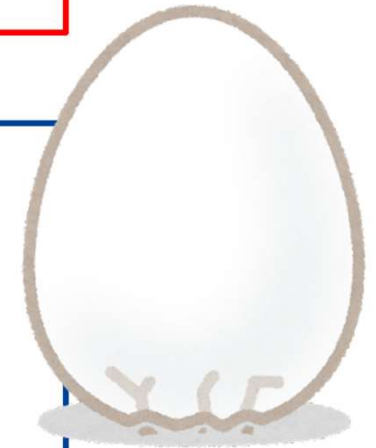


## ■ 数学からの見方

- ・世にある相互作用を行列のたたみかけと見る。
- ・非線形なふるまいを「スイッチ演算子」と見る。

## ■ ニューラルネットワークからの見方

- ・世にある相互作用を層の重ね合せで処理する。
- ・非線形なふるまいをReLU関数で近似する。



両者が一致することを、**本当に肌感覚で**理解しているかどうか、最後の一線を越えられるか否かを定める。

実は発見していました

- ・情報エントロピー：渡辺慧
  - ・ホップフィールドモデルなど：甘利俊一
  - ・オートエンコーダー：馬場則夫 et al.
  - ・ノイズ付き学習：久下哲郎
- ・・・きっと他にも、たくさん！

これを数学的に書いたのは1939年の私の論文です (Watanabe, 1939)

た. 詳しく申しますと, 核子のエントロピーというものはその核子間の相互組織の程度を表わすものであると同時に, 個々の核子の状態の不確定性を表わすものでもあるということを行ったのです. それから10年も経ったのちに, シャノン (Shannon, 1949) という人が情報理論ということ  
を言いだしまして, 情報理論においては, そのエントロピー  
というものが, 個々の, たとえばアルファベットの A と  
か B とかいうもののどれであるかの不確定性の程度を表  
わすると同時に, それの間の相互関係, 組織性を表わすも  
のであるということを行いだしたのです.

知ること 認識学序説  
(ちくま学芸文庫, 2011) より引用.

ところで, そういうことはすでに私が10年も前に原子核子の間の組織性ということについて言ったことです. それはどうしてエントロピー関数になるかということについてちょっと誤解を招きやすいことがありますので次に説明



ホップフィールドの考えたモデル自体は私がその10年ほど前に提案したものとまったく同じだったので、これを「ホップフィールド-甘利のモデル」と呼ぶ人もアメリカにいたが、多数の物理学者の勢いに押されてその声は掻き

消されていった。

ところが、21世紀に入って、この分野で活躍していたイタリアの物理学者アミット教授が私の部屋へきて、「これまで『ホップフィールドモデル』と呼んでいたが、文献を調べた結果、これは『甘利-ホップフィールドのモデル』と呼ぶべきであることがわかった。これまでそれを知らなくて申し訳ない。お詫びのしるしにイタリアのネクタイを贈る。明日の会議でお前がこのネクタイを着けていれば、お詫びは受け入れられたと思う」と述べてネクタイをくれたことがあった。

脳・心・人工知能 数理で脳を解き明かす (ブルーバックス新書, 2016)より引用.

学習することであり、これはいまの深層学習でも同じである。あまり知られていないが、これを最初に提案したのは私である。

## 📢... 世界初の多層パーセプトロンの確率降下学習

確率降下法による多層パーセプトロンの最初の学習例を示すことは歴史的な意味があると思うので、ここで紹介させていただきたい。

私の著書『情報理論Ⅱ——情報の幾何学的理論』（共立出版／1968年）から引用する（図5-2）。これは、NTTに入った教え子、斉藤庄司君の修士論文になったものである。当時のコンピュータは能力に乏しく、この計算にも苦労した。

※ SGD (stochastic gradient descent) として知られています。

じつは私が考えたのと同じ確率降下法のアイデアが、15年ほど遅れて第2次ニューロブームのときに発表され、花形の話の1つになる。こちらは「誤差逆伝搬法（バックプロパゲーション）」という格好のよい名前が付けられ、ニューロブームに乗って一世を風靡する。

これはアメリカの認知科学者ロンメルハルトらの仕事であるが、私はその論文のゲラ刷りを友人から送ってもらって、まったく同じ発想で議論が進んでいることに驚いた

脳・心・人工知能 数理で脳を解き明かす  
(ブルーバックス新書, 2016)より引用。

ところが、パーセプトロンは0, 1の値をとる離散素子、いわゆるマッカロック・ピッツの形式ニューロンを使っていた。だから、途中のシナプス効率を少し変えても、その変化が微弱であれば0, 1の答えは変わらない。先に述べたように、入力の線形和である $u$ がシナプス効率の変化によって少しぐらい変わっても、閾値を超えるほど大きく変化しなければ、出力の0, 1は元のままで何の変化も起こらない。だから、シナプスの効率を少しだけ変えて様子を見ようという戦略は、役に立たない。

それなら、ニューロンの動作を0, 1ではなくて、アナログ値にしてみればよいのではないかと私は考えた。

つまり、1つのニューロンの出力は、入力の重み付き和 $u$ の関数 $z=f(u)$ で決まるが、ここで出力 $z$ は0, 1の値をとるのではなくて、パルス頻度で表現されるアナログ値をとるとする。これをアナログニューロンという。

### 3.4 Encoder-Decoder 問題への応用

ニューラルネットの学習アルゴリズムの性能評価のためによく用いられる“ベンチマーク問題”としては、EX-OR 問題の他に Encoder-Decoder 問題が知られている。

この問題は、通信路に制約がある場合に、送信した信号をうまく復元するにはどうすればよいか、あるいは完全に復元可能かどうかといったことを考察する問題である。本節では、その一例として、0 または 1 のビット列からなる入力信号がニューラルネットにより復元できるかどうかという問題を考

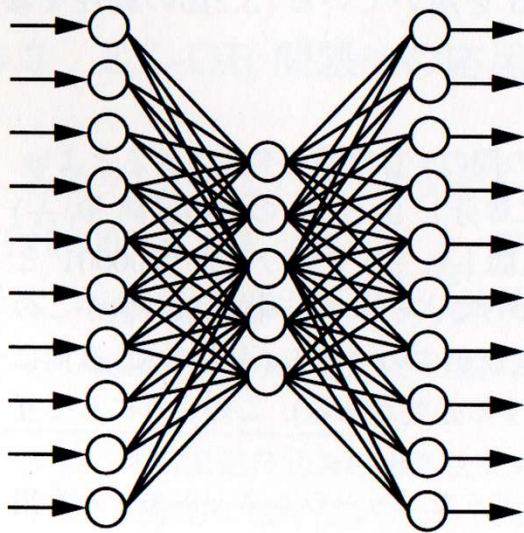


図 3.8 Encoder-Decoder 問題におけるニューラルネットモデルの一例

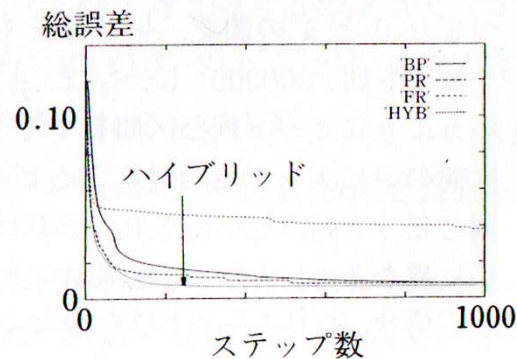


図 3.9 Encoder-Decoder 問題におけるニューラルネットの学習

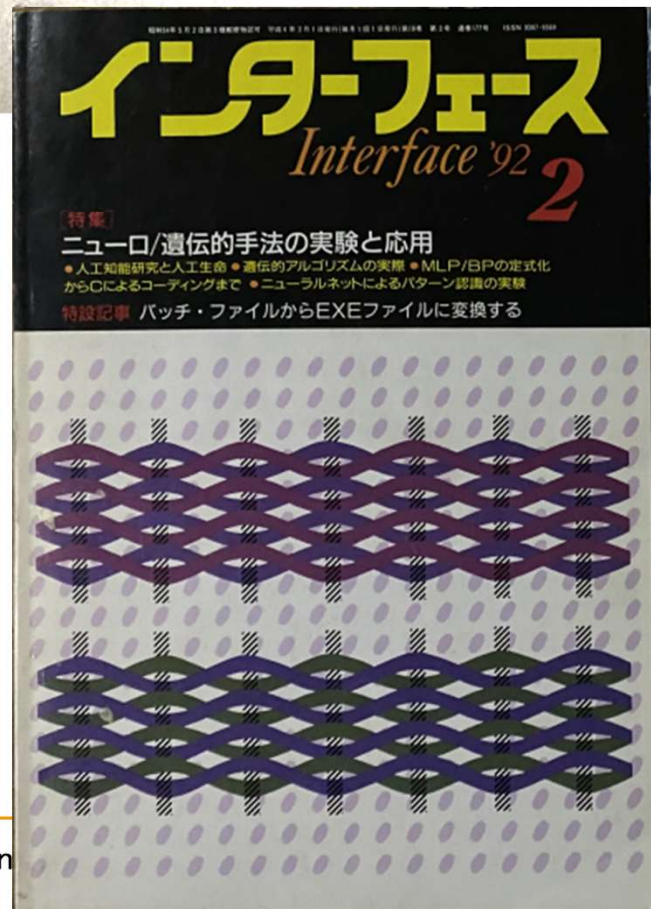
著者の一人馬場は、平成 5 年度の卒業研究のテーマの 1 つとしてニューラルネットによる入力信号復元問題を選び、各種問題を学生に行わせた<sup>42)</sup>。

ニューラルネットの基礎と応用  
[共立出版]  
馬場則夫(他) (1994)

2006年にジェフリー・ヒントン  
らが提案した。  
-- Wikipedia[オートエンコーダ]

## ●学習信号にノイズを加える

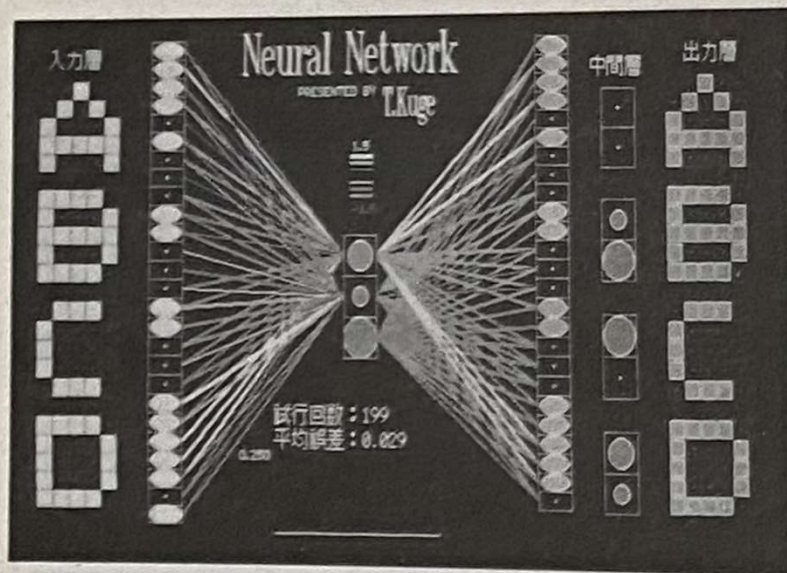
以上は、学習過程にノイズ・フリーな(ノイズを含まない)入力のみを用いたのであるが、つぎに、学習過程で入力にさまざまなレベルのノイズを加えて学習を行った結果を示す(この場合も教師信号はノイズ・フリー



## ●ノイズ付き学習はなぜ効果的か

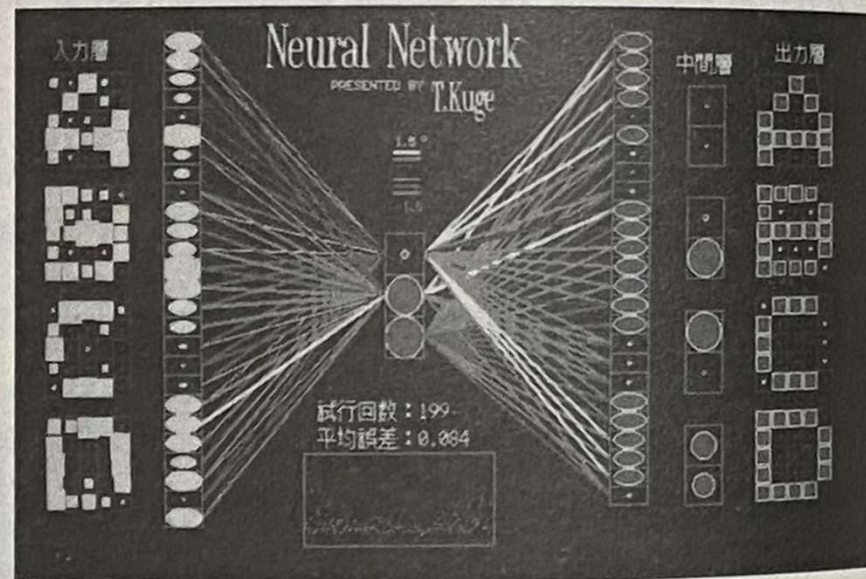
ところで、写真3.2 (a), 写真3.2 (b), 写真3.3, 写真3.4で示した画像符号化におけるノイズ付き学習の実験は非常に興味深い。ノイズ付き学習によってMLPの耐ノイズ性を向上させるという試みは、筆者の知るかぎり、これが最初である、と自負している(もしこの記事以前に同様の実験結果が発表されているのをご存じの方がおられれば、CQ出版(株)インターフェース編集部気付で筆者までお知らせいただければ幸いである)。

〔写真3.2〕 MLP(25-2-25)のノイズなし学習による  
画像符号化能力

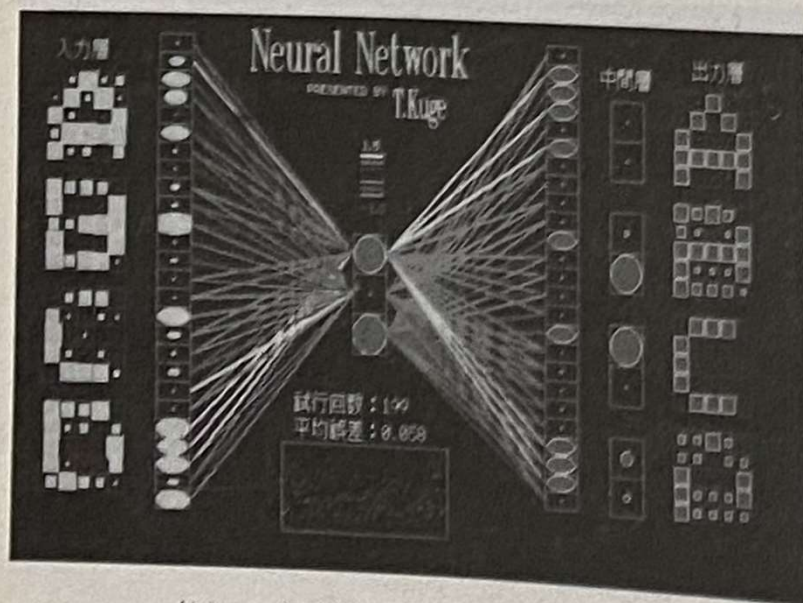


(a) ノイズなし文字入力提示

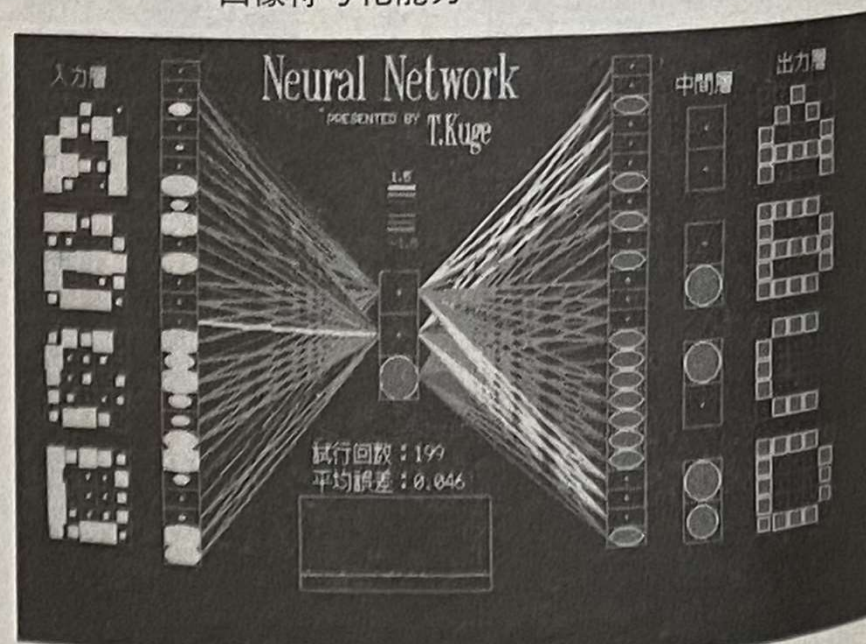
〔写真3.3〕 一様乱数  $[0, 0.1]$  ノイズあり学習による  
画像符号化能力



〔写真3.4〕 一様乱数  $[0, 0.5]$  ノイズあり学習による  
画像符号化能力



(b) ノイズ  $[0, 0.4]$  の文字入力提示



もちろん第一発見者・発明者は  
讃えられるべきだが、

実は到達していながら知られなかった人たち、  
到達まであと一歩だった人たち、  
2番手以降の人たちは  
無意味なのか？

同じ考えを持つ者がいるからこそ、  
普及するのだと思う。

今日、AIと呼ばれる潮流を  
支えているのは、  
一握りの主役だけではない。



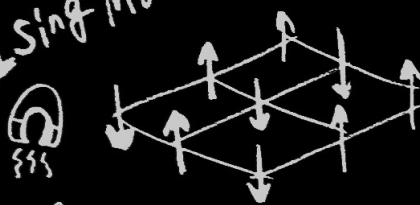
論文や書籍、文章の形で  
記録に残しておくことが大切。

たとえその場で流行らなかったとしても、  
いつか、誰かが  
見直す時がやってくる。

$\frac{dx}{dt}$  微分可能  
存世界  
人工知能

Rule Based  
 $\exists R = \{x | x \neq x\}$   
Prolog, Lisp...

Ising Model



第5世代...

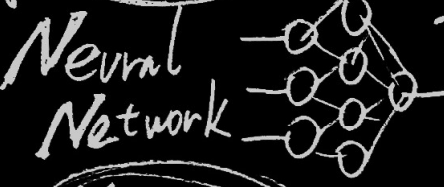
Perceptron



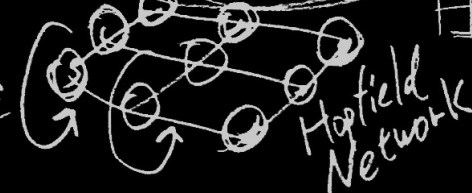
3次  
非線形解析

Logistic回歸

直列型

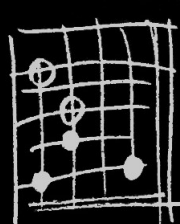
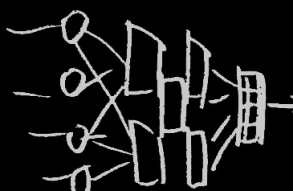


並列型



Hopfield Network

Combination Network



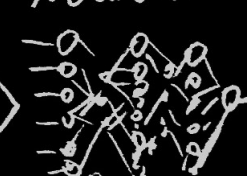
$\alpha$ -Go  
SL Policy Network  
Value-Net

Monte Carlo 法



Quantum  
Future...

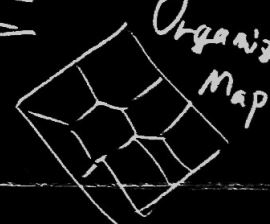
Deep Belief Network



Boltzman Machine

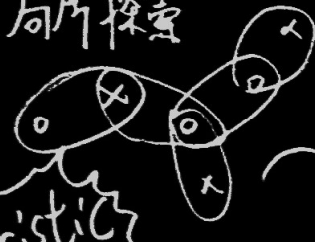


Self Organization Map



微分不能  
存世界  
Operations Research

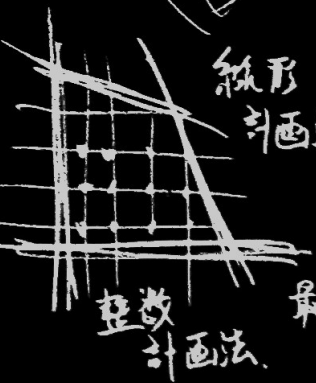
Local Search  
局所探索



多スタート法

焼きなまし法  
Simulated Annealing

遺伝的  
プログラム  
Genetic Algorithm



線形  
計画法

シフトワーク法

最大経路

整数  
計画法

MTS法

Artificial Life  
人工生命...

