階層ネットワークと BP(Back propagation)

2021/10/16

バックプロパゲーション

バックプロパゲーション（BP）はいわゆるニューラルネットの代名詞として使われるほどに知られた階層型ニューラルネットの学習則である。1980 年代のニューラルネットブームに火をつけたのが BP の有用性であり、現在では BP は有用な信号処理の手法の一つの選択肢として多くの現場で使われている。しかしその利用にはそれなりの知識も必要であり、本実験ではその基本的な原理と動作の理解を目指す。

BP の対象である階層型ネットワークは、データが与えられる入力層、計算結果が出力されて教師信号との比較が行われる出力層、そして入力と出力の間で途中の計算を行う一層以上の中間層からなっている。中間層はシグモイド関数を典型例とする非線形の入出力関数を持っており、出力層も同様であるが、非線形を使わないという選択肢もある。

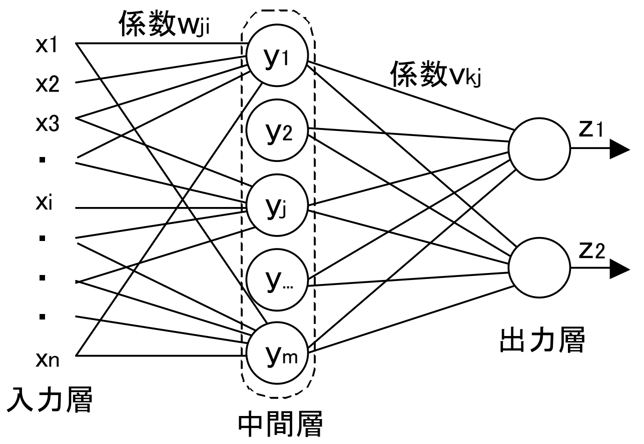
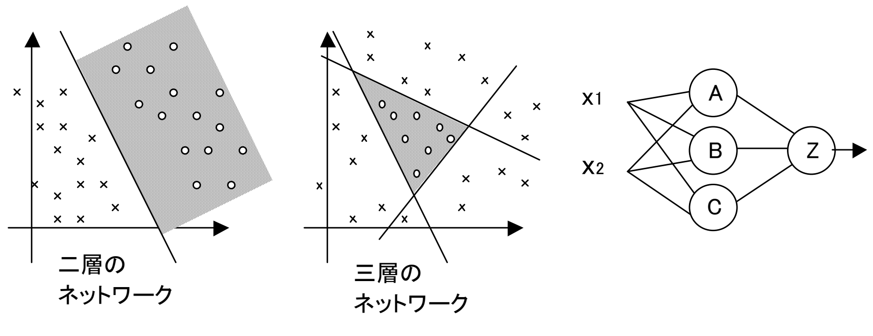


図4 階層型ネットワーク

階層ネットワークの理解

利用の形態からみたときのパーセプトロンの機能は分類であった。二層の構造しかもたないパーセプトロンは入力データの空間に一枚の超平面を張り、そのどちら側にあるかで認識と解釈した。多層型のネットワークでは、その超平面の数が増えると同時に、階層によって組み合わせることができるようになる。例えば三層（入力―中間―出力）のネットワークでは、中間層の個々の細胞が一つの超平面を表現し、出力層ではそれらの超平面の組み合わせによって凸領域を表現することができる。さらに四層以上になると、二つ以上の凸領域を組み合わせることが可能となり、より一般的なデータ分布の分類が可能となる。



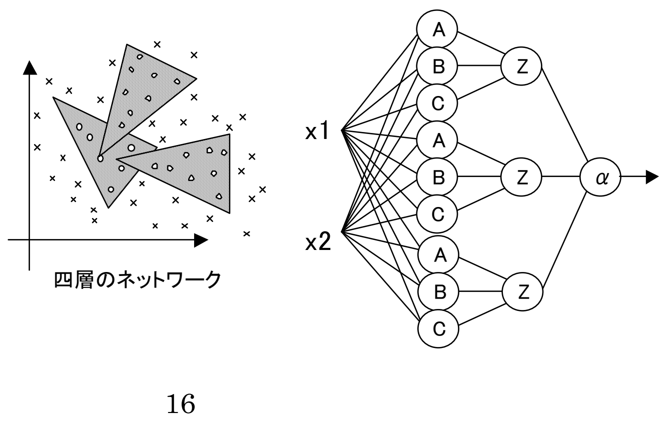


図 5　階層型ネットワークの分類能力の変化：
二層,三層,四層以上。

もう一つの解釈は、階層型ネットワークは入出力間の「関数近似」を実現すると考えることである。シグモイド関数は高次元の空間に配置されたなだらかな階段関数であり、それを多数組み合わせて線形和を取ることで、任意の関数の形を作ることができる。例えば $y = x^2$ という入出力関係は、階段関数をいくつか線形和として組み合わせることで実現できる。さらに、出力層に多数(例えば m 個)の細胞を用意すれば、m次元の出力を得ることも可能である。従って、階層型ネットワークは入力n次元から出力m次元への関数と考えることができる。実際、中間層の細胞数が十分あるなら、入力層から出力層へ向けて任意の精度の関数を実現できることが理論的に示されている。上記のパーセプトロンの分類は、分類対象のデータ群の間で閾値を横切る分類関数をデータから合成すると考えると、すっきり理解できる。

シグモイド関数：シグモイド関数とは以下の式で表される関数である。微分値の計算が簡単であることもあり、よく使われている。微分は自分で計算してみよ。

$$y = 1 \big/ \left(1 + \exp(-x)\right) \qquad (5)$$

図６　シグモイド関数

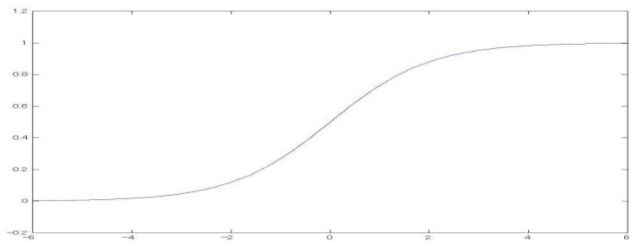
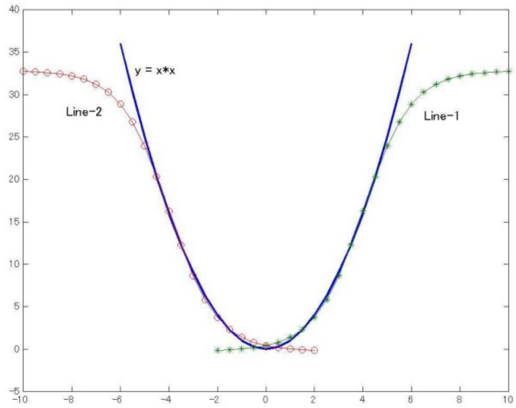


図６　$y = x^2$ の関数近似：
いくつかのシグモイド関数を
材料として、任意の関数を近似できる。



しかしながら、与えられた学習用のデータから適切な分類関数が学習できるかどうかは、そういう関数が階層型ネットワークで表現可能であるということとは別である。場合によっては学習が非常に困難な場合も存在する。そのような事例を学習を試みる前に検出することは非常に難しい。結局やってみたらうまくいくかどうかわかる、ということになる。

さらに、層の数や各層の細胞の個数は経験的に決定されるのが普通である。なぜなら対象とするデータの性質がわからないためである。ある入出力関係をデータから学習したいという場合はデータの性質がわからないということが前提としてあることが多い。データの性質が事前にわかっているなら、ニューラルネット以外の方法でも入出力関係を得る方法はいくらでもありうる。（もちろん、手間をかけたくないという動機も十分ありうる）

学習原理（δルール）: BP の学習では、パーセプトロンの学習に用いたδルールを階層ネットワークに適用する。すなわち、入力データ x^t に対してネットワークからほしい出力 T^t が同時に与えられる。ネットワークの出力 z は最終層から得られるので、誤差 $\delta=(z-T^t)$が定義される。

出力層の細胞のパラメータ w とθはパーセプトロンと同じ原理で以下の式となる。もちろん、これは誤差の最小化のための最急降下法の適用の結果である。

$$w = w - \alpha \delta x^P \quad (6a)$$ ：ただしαは小さい係数,例えば 0.05 程度。

$$\theta = \theta + \alpha \delta \quad (6b)$$

しかしこのままでは、教師信号が与えられない中間層の細胞群で使用する誤差δを定義することができない。そこで BP 法では、層の間のコネクションを使用して誤差を出力層から前の層に向けて逆伝播させることで、中間層の細胞に対する擬似的な誤差を定義する。これが error back propagation と呼ばれる手法で、BP 法の名前の由来となっている。例えば、M 層の細胞 m から次の N 層の細胞 n へのコネクションを w_{nm} とすると、細胞mの誤差 δm は N 層の細胞群の誤差 δ_n から次式で計算される。

$$\delta_m = \Sigma_n w_{nm} \delta_n \quad (7)$$

これは、誤差が N 層から M 層に逆向きにコネクションを伝って伝播している計算である。この考え方により、多層のネットワークでも出力層の誤差によって中間層の誤差を定義できるようになり、最急降下法による学習が可能となった。以上の結果として、BP 法の学習則は以下のようになる。

BP の学習則

０．学習データセットの用意

階層ネットワークで学習させたいデータ群 x^P（P＝1,,,N）と、それらのデータに対する望ましい出力値 T^P を用意する。シグモイド関数を出力関数として使う場合には、出力値はその最小値と最大値が 0.1 と 0.9 の間にはいるように正規化する。使わない場合は、適当でよい。

１．前向き計算

学習データ群から入力データ x^P と教師データ T^P を一セット選び、ネットワークに入力から出力まで前向きに計算させる。ネットワークのコネクションの初期値は小さな乱数で指定しておく。

$$y_j = f(\Sigma_i w_{ji} x_i^P - \theta_j) \quad (8a)$$

$$z_k = f(\Sigma_j v_{kj} y_j - \theta_k) \quad (8b)$$

２．出力層での誤差δの定義

前向き計算の出力 z^P と教師データ T^P の間で誤差 δ^P を計算する。出力層の細胞は一個とは限らない。

$$E_k = (\delta_k^P)^2 = (z_k^P - T^P)^2 \quad (9)$$

３．誤差の逆向き計算

出力層で定義された誤差 δ^P と、前向きのコネクションの値を逆向きに使用して、中間層の各細胞の誤差を順に計算していく。

$$E_j = (\delta_j)^2 = (\Sigma_k v_{kj} \delta_k)^2 \quad (10)$$

４．学習

各細胞に定義された誤差にもとづき、最急降下法で各細胞のコネクションと閾値を修正する。その際、非線型関数 f(s)の微分も忘れないこと。

$$d v_{kj}/dt = - \alpha\ dE_k/d v_{kj} = -2\alpha \delta_k^P df/ds\ y_j \quad (11a)$$

$$d\theta_k/dt = -\alpha\ dE_k/d\theta_k = 2\alpha\ \delta_k^P\ df/ds \qquad (11b)$$

$$d\,w_{ji}/dt = -\alpha\ dE_j/d\,w_{ji} = -2\alpha\ \delta_j\ df/ds\ x_i^P \qquad (12a)$$

$$d\theta_j/dt = -\alpha\ dE_j/d\theta_j = 2\alpha\ \delta_j\ df/ds \qquad (12b)$$

ただし　$s = \Sigma_j\, v_{kj}\, y_j - \theta_k$　(13)

$f(s) = 1/(1+\exp(-s))$　(14)

$\delta_j = \Sigma_k\, v_{kj}\ \delta_k$

５．反復

　すべての学習データに対して適切な出力が得られるか、
あるいは誤差の総和 $E = \Sigma_P|\delta^P|$ がある程度少なくなるまで、上記の 1－4 を反復する。

　階層型ネットワークの関数近似が有効に働く認識問題の例は exor 問題である。二次元（一般に多次元でもよいが）の入力データ／教師の関係が exor の形（(1,0)と(0,1)の近傍で出力 1 、(0,0)と(1,1)の近傍で 0 となる）の場合、それを一つの直線で表現することは困難であり、どうしても二本の直線を組み合わせて三層ネットワークにしなければならない。

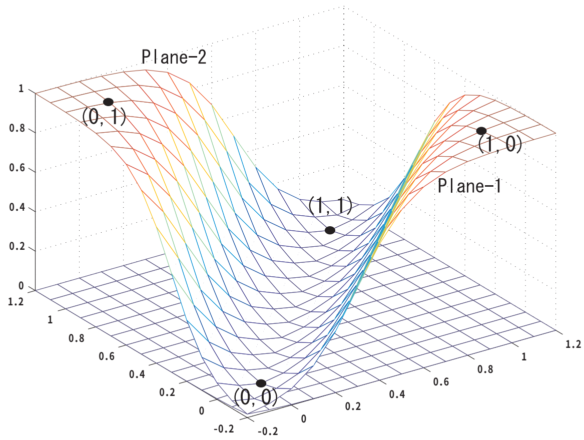


図 7　exor を学習した結果として得られるときに想定される関数

課題 4　Exor 問題

　Exor 的に分布したデータを 2 入力 2 中間細胞 1 出力の BP ネットワークで学習せよ．その結果として以下をグラフとしてプロットせよ．

　① 誤差の時間変化

　②中間層の境界線の時間変化+学習データ

OverFitting

CrossValidation

Soner 潜水艦

EXOR BP学習データ

R= 0.3

No.	X1	X2	T
1	1.211961	-0.17295	1
2	1.238369	0.94545	0
3	0.949147	-0.09687	1
4	0.999577	1.171578	0
5	1.200233	0.819478	0
6	1.2952	0.085471	1
7	0.049256	0.268667	0
8	0.037686	1.040169	1
9	0.830001	0.766374	0
10	-0.25907	-0.15944	0
11	1.21624	-0.18398	1
12	-0.2379	1.11853	1
13	0.940918	0.080239	1
14	0.864679	1.135235	0
15	0.015325	1.029233	1
16	0.813504	0.747619	0
17	-0.23696	-0.25376	0
18	0.796719	0.902244	0