3C−1

シグモイド関数の線形近似による バックプロパゲーション学習の初期値決定法

栗田　多喜夫
電子技術総合研究所

1. はじめに

バックプロパゲーション(BP)による学習法[1]が開発されたことにより，階層型のニューラルネットは，パターン認識等の多くの応用分野に適用され，その有効性が認められつつある．ニューラルネットを実際に応用する場合には，学習はなるべく速く収束することが望ましい．そうした観点から，学習の収束を加速するための BP の改良が試みられている（例えば，[2]参照）．しかし，学習時間に大きく影響するネットワークの初期重みをどうやって決定するかについては，あまり，関心が払われていないように思われる．普通，初期重みは乱数を用いて決められている．この問題に対して，二木は，パターン認識等の識別問題を解くネットワークの初期重みを，線形判別分析を用いて決定する方法を提案した[3]．ここでは，シグモイド関数を線形近似することによって，初期重みを決定する手法について述べる．

2. シグモイド関数の線形近似

階層型のネットワークを実際問題に応用する場合，各素子の入出力関数として，以下のようなシグモイド関数を用いることが多い．

$$f(x)=\frac{1}{1+\exp(-x)} \tag{1}$$

このシグモイド関数を，線形近似することを考えよう．

$$f(x)\approx\begin{cases}1 & \text{if } x\geq 1/2a\\ ax+1/2 & \text{if } |x|<1/2a\\ 0 & \text{if } x\leq -1/2a\end{cases} \tag{2}$$

のように近似すると，図1のように，比較的良い近似が得られる．

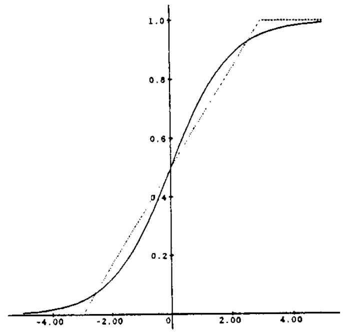

図 1: シグモイド関数の線形近似

3. 初期重みの決定

簡単のため3層のネットワークを考えよう．入力ベクトル x の次元を L，中間層の素子の数を M，出力ベクトル z の次元を N とする．このとき，ネットワークの出力 z は，

$$y = f(W_1'x-\theta_1) \tag{3}$$
$$z = f(W_2'y-\theta_2) \tag{4}$$

のように書ける．ここで，W_1, W_2 はそれぞれ入力層から中間層への結合の重み，中間層から出力層への結合の重みであり，θ_1, θ_2 は中間層および出力層のバイアスである．シグモイド関数 f を式(2)で線形近似すると，上式は

$$y = a(W_1'x-\theta_1)+1/2 1_M \tag{5}$$
$$z = a(W_2'y-\theta_2)+1/2 1_N \tag{6}$$

のようになる．ただし，比較的良い近似が得られるのは，中間層および出力層への入力ベクトル $u=W_1'x-\theta_1$ および $v=W_2'y-\theta_2$ の各要素の絶対値が $1/2a$ 以下の範囲である．

今，学習サンプル集合が $\{(x_i,t_i)|i=1,\ldots,m\}$ で与えられているとする．このとき，学習サンプルに対して平均2乗誤差の意味で最適な線形ネットワークの重みを BP による学習の際の初期重みとして用いる．

学習サンプルに対するシグモイド関数を線形近似したネットワークの出力と教師信号との平均2乗誤差は，

$$\varepsilon^2=\frac{1}{m}\sum_{i=1}^{m}||t_i-z_i||^2 \tag{7}$$

となる．これを最小とするような重みは以下のように解析的に求めることができる．ただし，中間層の出力 y に

[1]An Initial Weight Determination Method for Backpropagation Learning by Linear Approximation of Sigmoid Function
Takio KURITA (Electrotechnical Laboratory).

は座標軸の取り方に関して自由度が残るので，重み W_1, W_2 等は一意には定まらない．また，式 (6) のシグモイド関数の線形近似が成立するのは，中間層の入力 u の各要素の絶対値が $1/2a$ 以下のときである．従って，ここでは，中間層の入力 u の平均が 0 で，共分散行列が

$$\Sigma_U = 1/4 I_M \tag{8}$$

の解を求めよう．これは，中間層の入力の各要素の値の分布の分散が $1/2a$ になることを意味する．つまり，各要素の値の分布が正規分布であると仮定すると約 7 割が $1/2a$ 以下となる．以上の条件では，バイアス θ_1 は，

$$\theta_1 = W_1' \bar{x} \tag{9}$$

となる．また，初期重み W_1 は，固有値問題

$$\Sigma_{XT}\Sigma_{XT}' W_1 = \Sigma_X W_1 \Lambda \tag{10}$$

$$W_1' \Sigma_X W_1 = \frac{1}{4a^2} I_M \tag{11}$$

の解として求まる．一方，θ_2 および W_2 は，

$$W_2 = \frac{4}{a}\Sigma_{YT} \tag{12}$$

$$\theta_2 = W_2' \bar{y} + \frac{1}{a}(0.5 1_N - \bar{t}) \tag{13}$$

で与えられる．ここで，$\Sigma_{XT}, \Sigma_X, \Sigma_{YT}$ は，共分散行列であり，$\bar{x}, \bar{y}, \bar{t}$ は，平均ベクトルである．

ただし，式 (10) からわかるように，W_1 の意味のある解が得られるのは M が Σ_{XT} のランク以下の場合である．このことは，少なくとも中間層の素子の数が入力層あるいは出力層の素子の数より少ない場合（$\min(L, N)$）にしか初期重みが計算できないことを意味している．従って，中間層の素子の数が Σ_{XT} のランクよりも多い場合には，最初に，上記の方法で初期重みを求めて，その結果を中間層で情報がなるべく分散するように回転させるなどの方法を用いる必要がある．

4. 実験

4.1. アヤメの識別

フィシャーのアヤメ（3 種類）のデータ（入力ベクトルは 4 次元，各クラス 50 サンプル）に対して，乱数で初期重みを決めた場合と上記の方法で初期重みを決めた場合に対して，BP による学習を試みた．ただし，線形近似の傾きは，$a = 0.25$ とした．また，中間層の素子は 2 個とした．学習結果を図 2 に示す．

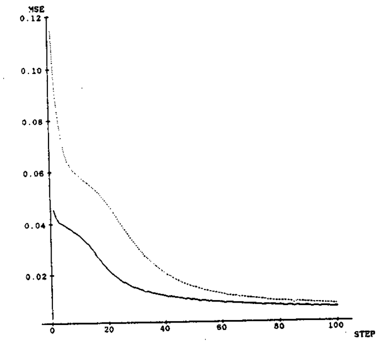

図 2: アヤメの識別の学習

4.2. 話者認識

音声を 17ch に帯域分割したフーリエスペクトルの平均パワーを正規化したデータ（4 単語，話者 5 人で各クラス 20 サンプル）[4] に対して，乱数で初期値を決めた場合と上記の方法で初期値を決めた場合に対して，BP による学習を試みた．線形近似の傾きは，同様に，$a = 0.25$ とした．また，中間層の素子数は 4 個とした．結果を図3に示す．

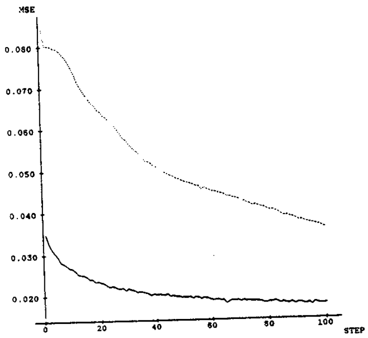

図 3: 話者認識の学習

5. まとめ

シグモイド関数を線形近似することによって，BP による学習の際の初期重みを決定する手法を示した．これは，中間層の素子の数が Σ_{XT} のランク以下でなければ，直接適用できないが，そうした条件を満足するネットワークに対しては，効果があるように思われる．

今後の課題としては，提案した手法のより詳細に評価すること，中間層の素子の数が多い場合の手法を開発することなどが考えられる．

参考文献

[1] D.E.Rumelhart et al."Parallel Distributed Processing," The MIT Press (1988)

[2] 麻生英樹，"誤差逆伝播学習の数理的性質，" 信学技報 PRU89-15 (1989)

[3] 二木徹，"バックプロパゲーションアルゴリズムの初期値決定法，" 信学技報 NC89-10 (1989)

[4] 海野他，"ニューラルネットワークによる話者認識，" テレビジョン学会技術報告, Vol.13, No.33 (1989)